

**Near-Surface Compressional and Shear Wave Speeds Constrained by Body-Wave
Polarisation Analysis**

Sunyoung Park

Department of Earth and Planetary Sciences,
Harvard University
20 Oxford Street, Cambridge, MA 02138
sunyoungpark01@fas.harvard.edu

Miaki Ishii

Department of Earth and Planetary Sciences,
Harvard University
20 Oxford Street, Cambridge, MA 02138
ishii@eps.harvard.edu

SUMMARY

A new technique to constrain near-surface seismic structure that relates body-wave polarisation direction to the wave speed immediately beneath a seismic station is presented. The P-wave polarisation direction is only sensitive to shear wave speed but not to compressional wave speed, while the S-wave polarisation direction is sensitive to both wave speeds. The technique is applied to data from the High-Sensitivity Seismograph Network in Japan, and the results show that the wave speed estimates obtained from polarisation analysis are compatible with those from borehole measurements. The lateral variations in wave speeds correlate with geological and physical features such as topography and volcanoes. The technique requires minimal computation resources, and can be used on any number of three-component teleseismic recordings, opening opportunities for non-invasive and inexpensive study of the shallowest (~ 100 m) crustal structures.

Key words: Body waves; Crustal imaging; Earthquake ground motions; Earthquake hazards; Site effects; Japan

1 INTRODUCTION

Near-surface structure and its regional variation influence the level of ground shaking at different sites, and thus, are important in assessing earthquake hazards (e.g. Shearer & Orcutt 1987; Cranswick et al. 1990; Semblat & Pecker 2009). One of the main quantities considered in the seismic hazard analysis is V_S^{30} , the average shear wave speed from the surface to 30-m depth (e.g. Borchardt & Glassmoyer 1992; Martin & Dobry 1994; Rinne 1994). This parameter is widely used as an input to ground-motion prediction equations (e.g. Boore et al. 1997; Chiou & Youngs 2008; Laurendeau et al. 2013), and several geophysical techniques have been established to estimate the quantity (e.g. Roser & Gosar 2010; Odum et al. 2013).

There have been various methodologies developed to improve constraints of V_S^{30} . A direct approach is to drill boreholes and obtain wave speed logs (e.g. Levander et al. 1994; Wu et al. 1994; Kneib 1995; Holliger 1996; Milkereit & Eaton 1998; Boerner et al. 2000), or measure the

1 elastic properties of the drill core samples in laboratories (e.g. Salisbury et al. 1994). Vertical
 2 seismic profiling also takes advantage of boreholes by recording seismic waves at various depths
 3 from a surface source (e.g. Anstey & Geyer 1987; Hardage 2000). However, drilling is expensive
 4 and requires specialised equipment. Alternatively, non-invasive approaches exist, such as seismic
 5 reflection or refraction studies using vibroseis or explosives as seismic sources (e.g. Choukroune
 6 1989; Mooney et al. 1998; Yilmaz 2001; Biondi 2006). Earthquakes also provide important sources
 7 to image the subsurface, and allow construction of regional tomographic models based on body
 8 waves such as P, S, Pn, and Sn waves (e.g. Ritzwoller et al. 2002; Liang et al. 2004; Schmandt
 9 & Lin 2014) or surface waves (e.g. Mitchell & Yu 1980; Laske & Masters 1996; Ritzwoller &
 10 Levshin 1998). Noise interferometry is another powerful tool to obtain near-surface wave speed,
 11 and is well-suited to detect changes in wave speeds over time (e.g. Lobkis & Weaver 2001; Shapiro
 12 & Campillo 2004; Draganov et al. 2007; Brenguier et al. 2008; Stehly et al. 2008; Nakata & Snieder
 13 2012). Wald & Allen (2007) have also made an effort to approximate V_S^{30} based upon topography
 14 of the region of interest.

15 We introduce an alternative approach to obtain near-surface wave speeds using body-wave
 16 polarisation. Note that the term polarisation has been used in the literature to describe two different
 17 directions, one associated with the particle motion, and another associated with the incident wave
 18 direction. In this paper, we use the terminology to indicate the particle motion. Polarisation of
 19 body waves is typically used to study anisotropy, for example, Schulte-Pelkum et al. (2001) and
 20 Fontaine et al. (2009) have used P-wave polarisation to constrain upper mantle anisotropy, while
 21 Cristiano et al. (2016) have examined similar data sets to study upper crustal anisotropy. On the
 22 other hand, Hu et al. (1994) have investigated shallow isotropic structure based upon perturbations
 23 in the incident wave direction. Finally, the amplitude of P receiver functions at zero time is directly
 24 related to the polarisation of P waves, and has been used to infer near-surface shear wave speed
 25 (Svenningsen & Jacobsen 2007; Hannemann et al. 2016).

26 The polarisation analysis presented in this paper has multiple advantages compared to previous
 27 methods. It requires minimal computation resources and can be applied wherever three-component

seismometers are installed. The technique is non-invasive, and thus, opens a new path to a reliable earthquake hazard assessment in any environment where drilling or a field experiment using vibro-trucks or explosives is not a practical option for measuring the near-surface seismic wave speeds. We apply the new approach to the High-Sensitivity Seismograph Network in Japan (Hi-net; Okada et al. 2004), where the results are benchmarked against the borehole well data available at most stations.

2 METHOD

We first seek a relationship between the seismic wave speed and the polarisation directions of incoming body waves. We show that the polarisation directions of teleseismic P waves observed at the surface are sensitive only to shear wave speed, while those of S waves are sensitive to both compressional and shear wave speeds. Based on this formulation, the best-fitting compressional and shear wave speeds near the surface can be obtained by a grid search, and the bootstrapping of the data allows their uncertainties to be quantified.

2.1 Polarisation of P and S Waves

When a body wave arrives at a seismic station located at the Earth's surface, the interaction of the incident wave with the free surface generates reflected and converted waves, and the combined particle motions of both incident and reflected/converted waves are recorded by the seismic instrument. The expression of particle motion can be derived for P and S wave incidences using free-surface boundary conditions (Aki & Richards 2002). We can then retrieve the relationship between the observed polarisation directions and the near-surface wave speed.

2.1.1 P-Wave Polarisation

The total displacement, $\mathbf{u}^P$, arising from the up-going (incident) P wave travelling in the x - z plane ($z = 0$ at the free surface and z increases downward; Figure 1), is a function of x , z , and time t ,

1 and can be written as,

$$\begin{aligned}
 \mathbf{u}^P(x, z, t) = & U^P(\sin \theta \hat{\mathbf{x}} - \cos \theta \hat{\mathbf{z}}) \exp \left[i\omega \left(\frac{\sin \theta}{\alpha} x - \frac{\cos \theta}{\alpha} z - t \right) \right] \\
 & + U^P(\sin \theta \hat{\mathbf{x}} + \cos \theta \hat{\mathbf{z}}) \dot{P}\dot{P} \exp \left[i\omega \left(\frac{\sin \theta}{\alpha} x + \frac{\cos \theta}{\alpha} z - t \right) \right] \\
 & + U^P(\cos \phi \hat{\mathbf{x}} - \sin \phi \hat{\mathbf{z}}) \dot{P}\dot{S} \exp \left[i\omega \left(\frac{\sin \phi}{\beta} x + \frac{\cos \phi}{\beta} z - t \right) \right],
 \end{aligned} \tag{1}$$

2 where U^P is the incident amplitude, θ and ϕ are the angles of the P and S ray paths from the vertical,
 3 respectively, i is the imaginary number, ω is the angular frequency, α and β are the compressional
 4 and shear wave speeds of the surface layer, respectively, and $\dot{P}\dot{P}$ and $\dot{P}\dot{S}$ are the P-to-P and P-to-S
 5 reflection coefficients, respectively, at the free surface (Figure 1a). The symbols $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ denote
 6 unit vectors in the x and z directions. The first term on the right hand side of Equation 1 describes
 7 the motion due to the incoming P wave while the second and third terms correspond to the reflected
 8 P and S waves.

9 A seismic station at the surface ($z = 0$) reduces Equation 1 to

$$\mathbf{u}^P(x, 0, t) = U^P \left\{ \left[(1 + \dot{P}\dot{P}) \sin \theta + \dot{P}\dot{S} \cos \phi \right] \hat{\mathbf{x}} + \left[(-1 + \dot{P}\dot{P}) \cos \theta - \dot{P}\dot{S} \sin \phi \right] \hat{\mathbf{z}} \right\} \exp [i\omega (px - t)],$$

10 where p is the ray parameter, and the relationship between the angles θ and ϕ , the speeds α and β ,
 11 and the ray parameter p is described by the Snell's law, $\frac{\sin \theta}{\alpha} = \frac{\sin \phi}{\beta} = p$. Based upon the above
 12 equation, the angle of the apparent polarisation with respect to the vertical, $\bar{\theta}$ (Figure 1a), can be
 13 derived as

$$\tan \bar{\theta} = \frac{(1 + \dot{P}\dot{P}) \sin \theta + \dot{P}\dot{S} \cos \phi}{(1 - \dot{P}\dot{P}) \cos \theta + \dot{P}\dot{S} \sin \phi}. \tag{2}$$

14 The reflection coefficients $\dot{P}\dot{P}$ and $\dot{P}\dot{S}$ are determined using the free-surface boundary condition
 15 so that (Aki & Richards 2002),

$$\dot{P}\dot{P} = \frac{-A + B}{A + B} \quad \text{and} \quad \dot{P}\dot{S} = \frac{C}{A + B}, \tag{3}$$

1 where

$$A = \left(\frac{1}{\beta^2} - 2p^2 \right)^2, \quad B = 4p^2 \frac{\cos \theta}{\alpha} \frac{\cos \phi}{\beta}, \quad \text{and} \quad C = 4p \frac{\cos \theta}{\beta} \left(\frac{1}{\beta^2} - 2p^2 \right).$$

2 Substituting Equation 3 simplifies Equation 2 as

$$\tan \bar{\theta} = \tan 2\phi, \tag{4}$$

3 which leads to a remarkable outcome,

$$\bar{\theta} = 2\phi. \tag{5}$$

4 Equation 5 implies that the observed polarisation direction of P wave is not a function of the
 5 incident angle θ , but exactly two times the angle ϕ of reflected S wave. This result was derived
 6 by Wiechert (1907) where he emphasised that the typical approximation of $\bar{\theta}$ as θ is not valid. He
 7 showed that a P wave with 90° incidence, i.e., horizontal incidence, is polarised at 70° ($\bar{\theta} = 70^\circ$),
 8 for surface speeds with $\beta^2 = \alpha^2/3$ or Poisson's ratio of 0.25, and demonstrated that this can lead
 9 to mislocation of earthquakes, i.e., locating the earthquakes at depth rather than near the surface.

10 This result can be extended to analyse the near-surface wave speed if the ray parameter of the
 11 incident wave is known. Using the Snell's law, Equation 5 can be written as

$$\bar{\theta} = 2 \arcsin(\beta p). \tag{6}$$

12 The apparent polarisation angle of the P wave, therefore, is purely controlled by the shear wave
 13 speed beneath the station, not the compressional wave speed. The shear wave speed can be mea-
 14 sured with the apparent polarisation angle of the P wave as

$$\beta = \frac{\sin \phi}{p} = \frac{\sin \left(\frac{\bar{\theta}}{2} \right)}{p}, \tag{7}$$

15 which was also recognised by Sverdrup & Jacobsen (2007).

16 2.1.2 S-Wave Polarisation

17 While the incident P wave always generates both reflected P and S waves regardless of the incident
 18 angle θ or the ray parameter p , this is not the case for the incident S wave. Only when the ray

parameter is smaller than the critical value ($p = \frac{1}{\alpha}$) can conversions of S to P wave occur. Beyond the critical value, total internal reflection generates only a reflected S wave. Here, we consider the case of a small ray parameter below the critical value, which is applicable to observations of teleseismic S wave arrivals.

Similar to Equation 1, the total displacement, $\mathbf{u}^S$, arising from the incident SV wave in the x - z plane is written as,

$$\begin{aligned} \mathbf{u}^S(x, z, t) = & U^S(\cos \phi \hat{\mathbf{x}} + \sin \phi \hat{\mathbf{z}}) \exp \left[i\omega \left(\frac{\sin \phi}{\beta} x - \frac{\cos \phi}{\beta} z - t \right) \right] \\ & + U^S(\sin \theta \hat{\mathbf{x}} + \cos \theta \hat{\mathbf{z}}) \dot{S}\dot{P} \exp \left[i\omega \left(\frac{\sin \theta}{\alpha} x + \frac{\cos \theta}{\alpha} z - t \right) \right] \\ & + U^S(\cos \phi \hat{\mathbf{x}} - \sin \phi \hat{\mathbf{z}}) \dot{S}\dot{S} \exp \left[i\omega \left(\frac{\sin \phi}{\beta} x + \frac{\cos \phi}{\beta} z - t \right) \right], \end{aligned} \quad (8)$$

where U^S is the amplitude of the incident SV wave, $\dot{S}\dot{P}$ and $\dot{S}\dot{S}$ are the S-to-P and S-to-S reflection coefficients at the free surface, respectively (Figure 1b). The particle motion at the surface becomes

$$\mathbf{u}^S(x, 0, t) = U^S \left\{ \left[(1 + \dot{S}\dot{S}) \cos \phi + \dot{S}\dot{P} \sin \theta \right] \hat{\mathbf{x}} + \left[(1 - \dot{S}\dot{S}) \sin \phi + \dot{S}\dot{P} \cos \theta \right] \hat{\mathbf{z}} \right\} \exp [i\omega (px - t)],$$

and the apparent polarisation can be characterised with an angle $\bar{\phi}$, defined to be perpendicular to the particle motion (Figure 1b). Using the ratio of vertical to horizontal motion,

$$\tan \bar{\phi} = \frac{(1 - \dot{S}\dot{S}) \sin \phi + \dot{S}\dot{P} \cos \theta}{(1 + \dot{S}\dot{S}) \cos \phi + \dot{S}\dot{P} \sin \theta}. \quad (9)$$

The reflection coefficients $\dot{S}\dot{S}$ and $\dot{S}\dot{P}$ can be computed using the wave speeds α and β , the angles θ and ϕ , and the ray parameter p such that

$$\dot{S}\dot{S} = \frac{A - B}{A + B} \quad \text{and} \quad \dot{S}\dot{P} = \frac{D}{A + B}, \quad (10)$$

where

$$D = 4p \frac{\cos \phi}{\alpha} \left(\frac{1}{\beta^2} - 2p^2 \right),$$

with expressions of A and B given in Equation 3 (Aki & Richards 2002).

Substituting Equation 10 simplifies Equation 9 as

$$\tan \bar{\phi} = \frac{\tan \phi \tan 2\phi}{\tan \theta}. \quad (11)$$

This expression shows that the apparent S-wave polarisation depends on both angles θ and ϕ , implying that the angle $\bar{\phi}$ is controlled by both the compressional and shear wave speeds beneath the station. For a given ray parameter p , Equation 11 can be rewritten as

$$\tan \bar{\phi} = \frac{2\beta^2 p \sqrt{1 - \alpha^2 p^2}}{\alpha(1 - 2\beta^2 p^2)}, \quad \text{and thus,} \quad \bar{\phi} = \arctan \left[\frac{2\beta^2 p \sqrt{1 - \alpha^2 p^2}}{\alpha(1 - 2\beta^2 p^2)} \right]. \quad (12)$$

2.2 Measurement of Polarisation

Polarisation, the direction or particle motion, is a robust quantity, since it is relatively insensitive to errors in source location and origin time compared to more conventionally measured quantities such as travel time. For example, a 50-km error in the source location leads to only about 0.1% error in the polarisation, while causing about 5% error in the travel time (Hu et al. 1994). Likewise, a few seconds error in the origin time does not affect the polarisation as long as the arrival is identified, whereas it maps entirely into the travel time.

A variety of time and frequency domain algorithms have been developed to measure the polarisation from three-component seismic recordings (e.g. Flinn 1965; Montalbetti & Kanasevich 1970; Samson & Olson 1981; Christoffersson et al. 1985; Vidale 1986; Magotra et al. 1987; Lilly & Park 1995; Wagner 1997). We use the Principal Component Analysis (PCA; Pearson 1901) in the time domain to calculate the apparent polarisation angle. Three-component seismograms of the body-wave arrival projected onto the vertical-radial plane can be written as an N by 2 matrix $\mathbf{X} = [\mathbf{q}, \mathbf{r}]$ where the column vectors $\mathbf{q} = (q_1, \dots, q_N)^T$ and $\mathbf{r} = (r_1, \dots, r_N)^T$ contain demeaned and discretised vertical and radial time series data, respectively, with the subscript and superscript T indicating the time and the transpose, respectively. The matrix $\mathbf{X}$ represents the data within the time window of length $(N - 1)dt$, where dt is the sampling interval. The particle motion is given

22 by the eigenvectors of the covariance matrix defined as

$$\mathbf{C} = \frac{\mathbf{X}^T \mathbf{X}}{N} = \frac{1}{N} \begin{bmatrix} \mathbf{q}^T \mathbf{q} & \mathbf{q}^T \mathbf{r} \\ \mathbf{r}^T \mathbf{q} & \mathbf{r}^T \mathbf{r} \end{bmatrix}.$$

1 The two eigenvalues λ_1 and λ_2 ($\lambda_1 > \lambda_2$) and eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are determined by

$$(\mathbf{C} - \lambda_l \mathbf{I}) \mathbf{v}_l = \mathbf{0}, \quad \text{with} \quad l = 1, 2, \quad (13)$$

2 where $\mathbf{I}$ and $\mathbf{0}$ are a 2 by 2 identity matrix and a 2 by 1 zero column vector, respectively. The
 3 eigenvector associated with the largest eigenvalue, $\mathbf{v}_1 = (v_{11}, v_{12})^T$, represents the major particle
 4 motion direction, i.e., the polarisation direction, for P waves. For the S-wave incidence, $\mathbf{v}_2 =$
 5 $(v_{21}, v_{22})^T$ represents the normal vector to the particle motion (Figure 1b). Consequently, the angles
 6 $\bar{\theta}$ and $\bar{\phi}$ are measured as,

$$\bar{\theta} = \arccos \left(\frac{v_{11}}{|\mathbf{v}_1|} \right) \quad \text{and} \quad \bar{\phi} = \arccos \left(\frac{v_{21}}{|\mathbf{v}_2|} \right).$$

7 The covariance matrix $\mathbf{C}$ is usually well-conditioned, since the noise in each of the three compo-
 8 nents is not correlated (Jurkevics 1988). PCA is, therefore, extremely robust and efficient.

9 **2.3 Estimation of Near-Surface Compressional and Shear Wave Speeds**

10 The ray parameter of P or S wave can be computed for a given station and earthquake geometry
 11 based upon, for example, a one-dimensional Earth model, and polarisation of P and S waves at the
 12 station can be modelled in terms of compressional and shear wave speeds using Equations 6 and
 13 12. In order to find the compressional and shear wave speeds that best-fit the observed polarisation
 14 data, we define misfit as

$$f(\alpha, \beta) = \frac{\sum_{i=1}^M w_i^P [\bar{\theta}_i(\alpha, \beta) - \bar{\theta}_i^{obs}]^2 + w_i^S [\bar{\phi}_i(\alpha, \beta) - \bar{\phi}_i^{obs}]^2}{\sum_{i=1}^M w_i^P + w_i^S}, \quad (14)$$

15 where $\bar{\theta}_i$ and $\bar{\theta}_i^{obs}$ are predicted and observed apparent incidence angles of the P wave from the
 16 i th earthquake, respectively, $\bar{\phi}_i$ and $\bar{\phi}_i^{obs}$ are the i th prediction and measurement of S polarisations,

17 respectively, and the summation is over all M earthquakes. The weighting factors w_i^P and w_i^S are
 1 assigned based on the quality of the measurement.

2 Since there are only two parameters, i.e., compressional and shear wave speeds, and the calcu-
 3 lation of the misfit (Equation 14) is fast, a grid search is suitable for finding the best-fitting wave
 4 speeds. Moreover, the effectiveness in the computation enables one to perform the grid search for
 5 hundreds of randomly resampled data sets, i.e., bootstrapping (Efron 1992). This provides a col-
 6 lection of parameter values with good fits to the observations, from which the final estimates of the
 7 wave speeds and their uncertainties are obtained.

8 **3 DATA**

9 In order to demonstrate the robustness and effectiveness of the new technique for obtaining the
 10 near-surface wave speeds, we apply the method to the Hi-net recordings of teleseismic earthquakes.
 11 A unique advantage of the Hi-net is its dense coverage with more than 700 three-component seis-
 12 mic stations located throughout Japan (Figure 2a). Comparison of the wave speed estimates be-
 13 tween the stations provides constraints on the lateral variation in near-surface wave speeds that can
 14 be examined against features such as topography. Another advantage of the Hi-net is that the local
 15 wave speeds at most of the stations have been documented by vertical seismic profiling (Obara et
 16 al. 2005), which can be used as benchmarks.

17 The Hi-net data are available through the National Research Institute for Earth Science and
 18 Disaster Resilience (NIED) of Japan (<http://www.hinet.bosai.go.jp>). For the analysis, intermediate
 19 and deep earthquakes, i.e., events deeper than 60 km, with M_w greater than 6, which occurred
 20 between 2004 January and 2016 April in the teleseismic distance range of 30 to 90 degrees are
 21 considered (Figure 2b). The depth criterion is set to ensure that the depth phases do not arrive
 22 within the direct phase time window. Regional events are excluded since the error in the earthquake
 23 location can introduce significant uncertainty in the predicted ray parameter values. Based on the
 24 USGS National Earthquake Information Center (NEIC) catalogue, these selection criteria result in

234 events. The data are corrected for the amplification factor, as provided by NIED as the sensor sensitivity, so that each of the three components has the same scaling, and for the misorientation if there is any. Data with signal-to-noise ratio (SNR) less than 2 are discarded, i.e., the weight w_i^P or w_i^S is set to zero, and the signal window is defined to be 5 seconds from the P and S onsets and the noise window is set to be between 10 and 5 seconds before the onset (Figure 3). There is no filtering of the data, and the onset is automatically selected using the continuous wavelet transform algorithm (Bogiatzis & Ishii 2015). Robustness of each polarisation measurement, w_i^P or w_i^S , is calculated by the amount of the total variance in the data explained by the major particle motion direction $\mathbf{v}_1$, i.e., $\frac{\lambda_1}{\lambda_1 + \lambda_2}$. Note that the data used in this study are velocity seismograms and not displacement, but the polarisation direction remains the same for either case.

The numbers of P and S measurements made at each Hi-net station differ, depending on the quality of the data (Figure 4). At most stations, more P measurements are available with higher quality than S measurements, since P arrivals are often clearer than S arrivals. As the ray parameter increases, the apparent P-wave angle, $\bar{\theta}$, increases (Equation 6), while the apparent S-wave angle, $\bar{\phi}$, either increases or decreases depending on the compressional and shear wave speeds (Equation 12). One of the main source of the scatter around the trend is the noise in the data, which accounts for about 4 and 8° scatter for P and S measurements, respectively. Also note that for S waves, high SNR does not necessarily guarantee a good quality SV signal, since the S signal window consists not only of the SV wave but also of P coda and SH wave. This contributes to the larger scatter in S measurements compared to P measurements.

4 RESULTS

For each Hi-net station, a grid search is performed using Equation 14 with ray parameters computed from the 1-D reference model IASP91 (Kennett & Engdahl 1991). Compressional and shear wave speeds are searched from 50 m/s to 7 and 5 km/s, respectively, with 50-m/s increment, while imposing an additional constraint $\beta \leq \sqrt{3}\alpha/2$ (Figure 5a). This condition ensures the shear and

bulk moduli to be positive such that dilatational and shear energy are positive (Aki & Richards
 2002). This grid search takes only about a couple of seconds with a single 2.66-GHz core, which
 permits one to execute a bootstrapping analysis, and repeating the grid search for 500 randomly
 resampled data sets allows the examination of the distribution of each wave speed (Figure 5b and
 c). The mean and standard deviation from bootstrapping are taken to be the final compressional
 and shear wave speed estimates and their uncertainties. Note that stations that do not have a global
 minimum solution within the search area, i.e., the estimates at the lower or upper bound (50 m/s
 or 7 and 5 km/s for compressional and shear wave speeds, respectively), are not included in the
 results shown below.

The shear wave speed is better constrained than the compressional wave speed, and its estimates
 for the Hi-net stations range from 0.1 to 4 km/s, with the average of 1.7 ± 0.1 km/s (Figure 6a).
 These are tightly-constrained values, where about 71% of them have uncertainties below 0.1 km/s
 and 96% below 0.2 km/s (Figure 6b). In contrast, the average uncertainty of the compressional
 wave speed estimates is 0.7 km/s, markedly larger than that of the shear wave speed estimates
 (Figure 7b). The compressional wave speed estimates range from 0.5 to 6.9 km/s with the average
 of 3.2 ± 0.1 km/s (Figure 7a), which, when combined with the average shear wave speed estimate,
 corresponds to a Poisson's ratio of 0.30 ± 0.04 . This value is higher than the typical value of about
 0.25 (e.g. Christensen 1996; Gercek 2007), suggesting that the estimates are sampling the shal-
 lowest and least consolidated parts of the crust (e.g. Stokoe & Woods 1972). Both compressional
 and shear wave speed estimates exhibit a similar spatial pattern, with a correlation coefficient of
 0.50. The corresponding p -value of the correlation, i.e., the probability of getting a correlation as
 large as 0.50 by random chance, when there is no correlation, is practically zero, confirming the
 statistical significance of the correlation between the two quantities, despite the large uncertainty
 in the compressional wave speed estimates.

24 5 DISCUSSION

1 5.1 Length of the Time Window

2 A factor that can affect the estimates of the wave speeds is the length of the time window used
 3 for the PCA. In order to investigate the effect, we repeat PCA with a 1-second window using the
 4 recordings from the station KAWH and compare against the polarisation measurements with the
 5 5-second window presented in Figure 4. The 1-second window improves the quality of each mea-
 6 surement; the means of measurement robustness w^P and w^S increase from 0.90 to 0.95 and 0.88
 7 to 0.91, respectively. This is expected since there are fewer scattered or converted phases arriving
 8 within the shorter time window. Even though each measurement has higher quality, the number of
 9 total measurements decreases significantly, from 214 to 156 for the P angle and from 102 to 83
 10 for the S angle. The events that have been discarded have emergent arrivals where the amplitude is
 11 small within the 1-second window leading to SNR lower than 2. Furthermore, some emergent ar-
 12 rivals having SNR over 2 give inconsistent angles, resulting in larger scatter than with the 5-second
 13 measurements (Figure S1). This is particularly evident for the S waves, where the measured angles
 14 cover almost all possible values, i.e., -90 to 90° . This large scatter in S measurements is attributed
 15 to other phenomena, such as shear wave splitting due to anisotropy. This effect has significant
 16 consequences for the S arrival time picks; since the S arrivals are picked using both horizontal
 17 components, if the SH phase arrives earlier, the 1-second window would miss a significant part of
 18 the SV phase arrival, and result in an apparent polarisation angle that deviates substantially from
 19 the correct value.

20 The large scatter in the collection of S-wave polarisation angles makes it difficult to identify
 21 any trend as a function of ray parameter, and leads to unreliable estimation of compressional wave
 22 speed; the value inferred from these measurements is 4.7 km/s, while the estimate using the 5-
 23 second analysis window is 4.1 km/s. On the other hand, the shear wave speed estimate, which is
 24 constrained by both P and S measurements, remains the same as 2.7 km/s regardless of the analysis
 25 window duration. This demonstrates that the P measurements are not affected by the choice of the

time window, and effects due to scattered or converted phases arising from the structure beneath or around the station does not contribute significantly to the measured P angles. One of the largest among these phases is the P to S converted wave from the Moho, which arrives at about 3 or 4 seconds after the P onset time for a typical crustal structure relevant for most stations in Japan. However, its amplitude relative to the main phase is small, and its particle motion direction is nearly orthogonal to the main phase. As a result, it decreases the weighting factor, w^P , but does not alter the principal component direction. Note that S to P converted phases do not affect the S measurements either, since they arrive earlier than the main phase and most of their energy is not captured in the analysis time window that starts from the S onset. Even if they partially arrive within the time window, the motion is orthogonal to the main S motion and does not affect the S angle measurements significantly. We choose the longer time window that stabilises the analysis and ensures the main arrivals to be present for the PCA.

5.2 Ray Parameter Assumption

Next, we examine the assumption that the ray parameter is a theoretical value based on IASP91 (Kennett & Engdahl 1991). There are two aspects that can be tested: the dependence of the ray parameter on the chosen 1-D reference model and the deviation from the reference 1-D model. To address the former, we compare results from two widely used 1-D reference models: IASP91 and PREM (Dziewoński & Anderson 1981). The maximum absolute differences in the ray parameter values for P and S waves are 0.02 and 0.10°/second, respectively. This translates to the differences of about 0.1° for both P and S measured angles, which are negligible compared to the scatter in the observed angles and their uncertainties.

The other aspect is the difference in the predicted ray parameter due to the local 1-D structure, i.e., the near-surface wave speeds are not described well by global 1-D models. Theoretically, if the wave speed at the top layer differs from the reference 1-D model value, the ray parameter should be altered for given earthquake depth and distance. The magnitude of the difference increases as the deviation of the wave speed structure from the 1-D model increases. In order to understand

the upper limit of the change in the ray parameter, we test for an unrealistic case of extremely thick low speed layer, where the surface layer has compressional and shear wave speeds of 0.10 and 0.05 km/s, respectively, with the thickness of the layer reaching 10 km. For all possible values of earthquake depths and distances, the maximum absolute difference in the ray parameter for P and S waves are 0.003 and 0.010°/second, respectively, which are an order of magnitude smaller than the effect due to different reference models. Hence, the dependence of the ray parameter on the local structure is insignificant. It is important, however, to note that the ray parameter can also be affected by the wave speed structure along the path deviating from the 1-D model. Combining data from different distances and azimuths helps averaging out such 3-D effect. Nevertheless, we acknowledge that with uneven data coverage, the issue requires further study.

5.3 Effect of Anisotropy

Existence of faults, fractures, and tectonic stresses in the crust induces azimuthal anisotropy, i.e., wave speeds vary as sinusoidal functions of backazimuth (e.g. Backus 1965; Crampin et al. 1982). In the presence of anisotropy, the near-surface wave speed estimates would be robust average isotropic wave speeds, if they are based on polarisation measurements from all backazimuths. In the case of uneven backazimuthal coverage as shown in Figure 2(b), however, effect due to anisotropy can bias wave speed estimates. We examine this potential bias using the average compressional and shear wave speed estimates, i.e., 3.2 and 1.7 km/s, respectively, and the range of azimuthal anisotropy in the crust of 5 to 20% (e.g. Daley & McEvilly 1990; Aster & Shearer 1991; Coutant 1996; Peng & Ben-Zion 2004; Liu et al. 2005; Boness & Zoback 2006). The fast direction is assumed to be in the east-west direction (e.g. Kaneshima 1990), and the predicted compressional and shear wave speeds at each backazimuth are calculated based on the sinusoidal function with the 180° periodicity (e.g. Backus 1965). By averaging these wave speeds for the uneven backazimuthal coverage, “azimuthally biased” wave speed estimates are obtained and compared with the input average isotropic wave speeds. The biases in the wave speeds range from 0.02 to 0.08 km/s for P and 0.01 to 0.04 km/s for S wave, smaller than the average uncertainties of 0.7 km/s for P and

26 and 0.1 km/s for S wave speed estimates. Thus, the effect from anisotropy is within the uncertainty
 1 of wave speed estimates, even if a strong (20%) anisotropy is assumed.

2 Existence of local anisotropy can also be investigated by examining the backazimuthal de-
 3 pendence of the apparent body-wave polarisation measurements. When the residuals, i.e., the
 4 difference between the observed and the predicted polarisations, are examined as a function of
 5 backazimuth, however, we do not observe systematic variation (Figure S2), which confirms that
 6 anisotropy is difficult to detect from the given data. Nevertheless, with better quantity and qual-
 7 ity of data, the technique presented in this manuscript can be extended to examine local shallow
 8 anisotropy structure.

9 **5.4 Comparison to the Benchmark Measurements**

10 In order to check the validity of the estimates obtained by the method presented in this manuscript,
 11 the results are compared against the well measurements at the instrument depth (inset of Fig-
 12 ure 6c). The shear wave speed estimates using teleseismic wave arrivals are in good agreement
 13 with the borehole measurements with a correlation coefficient of 0.62 (Figure 6c and d), and the
 14 corresponding p -value of nearly zero ($< 10^{-60}$), verifying the statistical significance of the corre-
 15 lation between the two quantities. Our estimates are slightly lower than the well values such that
 16 the mean of the difference, i.e., $\langle \text{estimate} - \text{well} \rangle$, is -0.1 ± 0.1 km/s, and the magnitude of the
 17 underestimation is consistent with that predicted from the synthetic test for instruments at some
 18 depths (Figure A1). The underestimation is more pronounced at wave speeds below 1.5 km/s, re-
 19 sulting in the best fitting linear regression with a positive intercept, 0.61 ± 0.10 km/s, and the slope
 20 of 0.67 ± 0.06 . Given that the 95% of the benchmarked wave speeds are measured at depths less
 21 than 500 m, and more than half at about 100 m, this comparison demonstrates the efficacy of our
 22 technique in recovering the shear wave speed at these shallow depths, i.e., less than a kilometre.

23 For the compressional wave speed estimates, the comparison to the well data provides a lin-
 24 ear trend with a slope of 0.13 ± 0.10 (Figure 7). The slope is considerably lower than unity be-
 25 cause the variance in the estimates is larger than that of the well measurements, also resulting

in a low correlation coefficient of 0.12. However, the correlation between the estimates and the well values is statistically significant, where the p -value for the correlation for this number of data is 0.002. The difference between the compressional wave speed estimates and the well values, i.e., (estimate—well), has an average of -0.4 ± 0.1 km/s. The overall underestimation can be explained by the borehole effect (See the discussion in Appendix A), which is also compatible with the magnitude of the underestimation being larger for the compressional than shear wave speed. Nonetheless, the comparison to the benchmark demonstrates that the polarisation measurements are effective in recovering the compressional wave speeds in the top few hundreds of metres or less.

5.5 Correlation with Geological Features

Comparison of the lateral variations of the shear wave speed against topographic slope on a log-log scale shows a positive correlation as reported by Wald & Allen (2007) (Figure 8a). This is a manifestation of the fact that materials of higher rigidity or wave speed can support steeper slopes while the sedimentation process forms more horizontal layers. The correlation implies that the use of the topographic gradient as a proxy to approximate the shear wave speed in the top 30 m can be extended to deeper depths. There are, however, high topographic areas with slow wave speeds, for example, in the southwestern part of Japan. These regions correspond to volcanic zones (Global Volcanism Program 2013), suggesting that the high heat flow (Tanaka et al. 2004) or the presence of magma results in the low speeds. Stations within 25 km of a volcano have wave speeds that are not statistically correlated with the gradient of topography (p -value of 0.105; Figure 8a). Without the stations near volcanoes, however, the correlation becomes more significant with the p -value of 0.001, and the regression analysis gives a slope of 0.20 ± 0.12 . These results imply that the approximation of shear wave speed based on the topography should take volcano locations into account for reliable hazard assessments (Matsuoka et al. 2006).

The shear wave speed estimates can also be compared to the basin thickness, another important parameter for seismic hazards. We use the Japan Integrated Velocity Structure Model (JIVSM;

26 Koketsu et al. 2012), a 3-D seismic wave speed model that has been constrained by various data
 1 sets such as refraction/reflection experiments, gravity surveys, and surface geology. Even though
 2 its depth resolution is lower than the well data, it provides wave speed estimates to greater depths,
 3 down to the upper mantle. This allows comparisons of the wave speed estimates against the thick-
 4 ness of basins, the depth at which shear wave speed reaches the speed of the bedrock, i.e., about
 5 3 km/s. The basin thickness can be calculated for each Hi-net station location using the nearest grid
 6 point of the JIVSM, which is as close as 1 km. When the thickness is compared to the polarisation-
 7 based shear wave speeds, there is a significant negative correlation with the regression slope of
 8 -0.27 ± 0.09 and the p -value of about 10^{-8} ; thicker the sediment layer, the lower the wave speed
 9 (Figure 8b). This is consistent with the expectation that the thicker and more prominent basins
 10 contain finer grains with low speeds. The correlation suggests that the shear wave speed and the
 11 basin depth share common information, i.e., ground motion prediction for sites without the basin
 12 depth information can still be achieved reasonably as long as there is good knowledge of the shear
 13 wave speeds.

14 6 CONCLUSIONS

15 We have introduced and implemented a new technique to estimate near-surface seismic wave
 16 speeds based on body-wave polarisation. P-wave polarisation direction has no sensitivity to sub-
 17 surface compressional wave speed but only to shear wave speed. S-wave polarisation direction, on
 18 the other hand, is sensitive to both compressional and shear wave speeds. Combining the P- and
 19 S-polarisation directions measured by principal component analysis, therefore, provides estimates
 20 of both P- and S-wave speeds at shallow depths, e.g., the top hundred metres. The polarisation
 21 measurement and the wave speed estimation are computationally efficient, providing tremendous
 22 opportunities to study near-surface seismic structures of different parts of the world, some of which
 23 may be difficult to obtain using other computationally-intensive approaches such as noise corre-

lation (e.g. Shapiro & Campillo 2004) or invasive and expensive approaches such as well logging (e.g. Wu et al. 1994).

Application of the technique to the dense Hi-net array produces distributions of near-surface compressional and shear wave speeds in Japan. The wave speed estimates are consistent with the well data, demonstrating the effectiveness of the technique and near-surface sensitivity. We find that our wave speed estimates correlate with the gradient of topography, confirming that the topography can be used as a proxy for regions without sufficient seismic data. However, it is important to note that the sites near volcanoes require a separate treatment. Furthermore, the basin thickness is also shown to correlate with the shear wave speed, which can help predict the ground motion in regions without constraints on the basin thickness. For future studies, the technique introduced here can be expanded to examine more complex seismic structure, such as anisotropy beneath a station. Depth dependent wave speeds can also be studied by investigating the frequency-dependence of the body-wave polarisation measurements, where the higher frequency data are sensitive to shallower depths. Furthermore, analyses of the data from different time periods can help monitor the changes in subsurface seismic wave speed over time.

ACKNOWLEDGEMENTS

The authors thank Michael Ritzwoller and two anonymous reviewers for comments and suggestions that improved the manuscript. We also thank National Research Institute for Earth Science and Disaster Prevention in Japan for providing the data, and Loïc Viens for the Japan Integrated Velocity Structure Model. Some figures have been generated using the Generic Mapping Tools (Wessel & Smith 1991). This research is supported by NSF grant EAR-1735960. S.P. is supported by the Samsung Scholarship.

22 References

- 1 Aki, K. & Richards, P. G., 2002. *Quantitative Seismology*, University Science Books, Sausalito,
2 California.
- 3 Anstey, N. A. & Geyer, R. L., 1987. Borehole Velocity Measurements and the Synthetic Seismo-
4 gram, *Boston, MA, IHRDC*.
- 5 Aster, R.C. & Shearer, P.M., 1991. High-frequency borehole seismograms recorded in the San
6 Jacinto fault zone, southern California. Part 1. Polarizations, *Bull. Seismol. Soc. Am.*, **81**, 1057–
7 1080.
- 8 Backus G., 1965. Possible forms of seismic anisotropy of the uppermost mantle under oceans, *J.*
9 *Geophys. Res.*, **70**, 3429–3439.
- 10 Biondi, B. L., 2006. *3D Seismic Imaging*, Society of Exploration Geophysicists.
- 11 Boerner, D. E., Milkereit, B. & Davidson, A., 2000. Geoscience Impact: A Synthesis of Studies of
12 the Sudbury Structure, *Can. J. Earth Sci.*, **37**, 477–501.
- 13 Bogiatzis, P. & Ishii, M., 2015. Continuous Wavelet Decomposition Algorithms for Automatic De-
14 tection of Compressional- and Shear-Wave Arrival Times *Bull. Seismol. Soc. Am.*, **105** 1628–
15 1641, doi:10.1785/0120140267.
- 16 Boness, N.L. & Zoback, M.D., 2006. A multiscale study of the mechanisms controlling shear
17 velocity anisotropy in the San Andreas Fault Observatory at Depth. *Geophys.*, **71**, F131–F146.
- 18 Boore, D. M., Joyner, W. B. & Fumal, T. E., 1997. Equations for Estimating Horizontal Response
19 Spectra and Peak Acceleration from Western North American Earthquakes: A Summary of
20 Recent Work, *Seismol. Res. Lett.*, **68**, 128–153.
- 21 Borchardt, R. D. & Glassmoyer, G., 1992. On the Characteristics of Local Geology and Their
22 Influence on Ground Motions Generated by the Loma Prieta Earthquake in the San Francisco
23 Bay Region, California, *Bull. Seismol. Soc. Am.*, **82**, 603–641.
- 24 Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M. & Larose, E., 2008.

- 25 Postseismic Relaxation Along the San Andreas Fault at Parkfield from Continuous Seismolog-
 1 ical Observations, *Science*, **321**, 1478–1481.
- 2 Chiou, B.-J. & Youngs, R. R., 2008. An NGA Model for the Average Horizontal Component of
 3 Peak Ground Motion and Response Spectra, *Earthq. Spectra*, **24**, 173–215.
- 4 Choukroune, P., 1989. The ECORS Pyrenean Deep Seismic Profile Reflection Data and the Overall
 5 Structure of an Orogenic Belt, *Tectonics*, **8**, 23–39.
- 6 Christensen, N. I., 1996. Poisson's ratio and crustal seismology, *J. Geophys. Res.*, **101**, 3139–3156.
- 7 Christoffersson, A., Husebye, E. S. & Ingate, S. F., 1985. A New Technique for 3-Component Seis-
 8 mogram Analysis, *Semiannual Technical Summary, Oct. 1984-Mar. 1985, NORSAR Scientific*
 9 *Report*, 2- 85/86, Kjeller, Norway.
- 10 Coutant, O., 1996. Observation of shallow anisotropy on local earthquake records at the Garner
 11 Valley, Southern California, downhole array, *Bull. Seismol. Soc. Am.*, **86**, 477–488.
- 12 Crampin S., Stephen R., & McGonigle R., 1982. The polarization of P-waves in anisotropic media,
 13 *Geophys. J. R. astr. Soc.*, **68**, 477–485.
- 14 Cranswick, E., King, K., Carver, D., Worley, D., Williams, R., Spudich, P. & Banfill, R., 1990. Site
 15 Response Across Downtown Santa Cruz, California, *Geophys. Res. Lett.*, **17**, 1793–1796.
- 16 Cristiano, L., Meier, T., Krüger, F., Keers, H. & Weidle, C., 2016. Teleseismic P-Wave Polarization
 17 Analysis at the Gräfenberg array, *Geophys. J. Int.*, **207**, 1456–1471, doi:10.1093/gji/ggw339.
- 18 Daley, T.M. & McEvilly, T.V., 1990. Shear-wave anisotropy in the Parkfield Varian well VSP, *Bull.*
 19 *Seismol. Soc. Am.*, **80**, 857–869.
- 20 Draganov, D., Wapenaar, K., Mulder, W., Singer, J. & Verdel, A., 2007. Retrieval of Reflections
 21 from Seismic Background-Noise Measurements, *Geophys. Res. Lett.*, **34**, L04305, doi:10.1029/2006GL028
- 22 Dziewoński, A.M. & Anderson, D.L., 1981. Preliminary Reference Earth Model, *Phys. Earth*
 23 *Planet. Inter.*, **25**, 297–356.
- 24 Efron, B., 1992. Bootstrap methods: another look at the jackknife. *Breakthroughs in statistics*,
 25 Springer New York, 569–593.

- 26 Flinn, E. A., 1965. Signal Analysis Using Rectilinearity and Direction of Particle Motion, *Proc.*
1 *IEEE*, **53**, 1874–1876, doi:10.1109/PROC.1965.4462.
- 2 Fontaine, F.R., Barruol, G., Kennett, B.L.N., Bokelmann, G.H.R. & Reymond, D., 2009. Upper
3 mantle Anisotropy Beneath Australia and Tahiti from P Wave Polarization: Implications for
4 Real-Time Earthquake Location, *J. Geophys. Res.*, **114**, B03306, doi:10.1029/2008JB005709.
- 5 Gercek, H., 2007. Poisson's ratio values for rocks, *Int. J. Rock Mech. Min. Sci.*, **44**, 1–13.
- 6 Global Volcanism Program, 2013. *Volcanoes of the World*, v. 4.6.0. Venzke, E (ed.). Smithsonian
7 Institution. Downloaded 03 Dec 2016. <http://dx.doi.org/10.5479/si.GVP.VOTW4-2013>.
- 8 Hannemann, K., Krüger, F., Dahm, T. & Lange, D., 2016. Oceanic Lithospheric S-Wave Velocities
9 from the Analysis of P-Wave Polarization at the Ocean Floor, *Geophys. J. Int.*, **207**, 1796–1817,
10 doi:10.1093/gji/ggw342.
- 11 Hardage, B. A., 2000. *Vertical Seismic Profiling*, Pergamon
- 12 Holliger, K., 1996. Upper-Crustal Seismic Velocity Heterogeneity as Derived from a Variety of
13 P-Wave Sonic Logs, *Geophys. J. Int.*, **125**, 813–829.
- 14 Hu, G., Menke, W. & Powell, C., 1994. Polarization Tomography for P Wave Velocity Structure in
15 Southern California, *J. Geophys. Res.*, **99**, 15245, doi:10.1029/93JB01572.
- 16 Jurkevics, A., 1988. Polarization Analysis of Three-Component Array Data, *Bull. Seismol. Soc.*
17 *Am.*, **78**, 1725–1743.
- 18 Kaneshima, S., 1990. Origin of crustal anisotropy: shear wave splitting studies in Japan, *J. Geo-*
19 *phys. Res.*, **95**, 11121–11133.
- 20 Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for Global Earthquake Location and Phase
21 Identification, *Geophys. J. Int.*, **105**, 429–465.
- 22 Kneib, G., 1995. The Statistical Nature of The Upper Continental Crystalline Crust Derived from
23 in Situ Seismic Measurements, *Geophys. J. Int.*, **122**, 594–616.
- 24 Koketsu, K., Miyake, H. & Suzuki, H., 2012. Japan Integrated Velocity Structure Model Version
25 1. *Proceedings of the 15th World Conference on Earthquake Engineering*, 3–6.

- 26 Laske, G. & Masters, G., 1996. Constraints on Global Phase Velocity Maps from Long-Period
1 Polarization Data, *J. Geophys. Res.*, **101**, 16059–16075.
- 2 Laurendeau, A., Cotton, F., Ktenidou, O.-J., Bonilla, L.-F. & Hollender, F., 2013. Rock and Stiff-
3 Soil Site Amplification: Dependency on Vs30 and Kappa (κ_0), *Bull. Seismol. Soc. Am.*, **103**,
4 3131–3148.
- 5 Levander, A., Hobbs, R.W., Smith, S.K., England, R.W., Snyder, D. B. & Holliger, K., 1994. The
6 Crust as a Heterogeneous “Optical” Medium, or “Crocodiles in The Mist,” *Tectonophys.*, **232**,
7 281–297.
- 8 Liang, C., Song, X. & Huang, J., 2004. Tomographic Inversion of Pn Travel Times in China, *J.*
9 *Geophys. Res.*, **109**, B11304, doi:10.1029/2003JB002789.
- 10 Lilly, J. M. & Park, J., 1995. Multiwavelet Spectral and Polarization Analyses of Seismic Records,
11 *Geophys. J. Int.*, **122**, 1001–1021, doi:10.1111/j.1365-246X.1995.tb06852.x.
- 12 Liu, Y., Teng, T.L. & Ben-Zion, Y., 2005. Near-surface seismic anisotropy, attenuation and disper-
13 sion in the aftershock region of the 1999 Chi-Chi earthquake, *Geophys. J. Int.*, **160**, 695–706.
- 14 Lobkis, O. I. & Weaver, R. L., 2001. On the Emergence of the Green’s Function in the Correlations
15 of a Diffuse Field, *J. Acoust. Soc. Am.*, **110**, 3011–3017.
- 16 Magotra, N., Ahmed, N. & Chael, E., 1987. Seismic Event Detection and Source Location Using
17 Single-Station (Three-Component) Data, *Bull. Seismol. Soc. Am.*, **77**, 958–971.
- 18 Martin, G. R. & Dobry, R., 1994. Earthquake Site Response and Seismic Code Provisions, *NCEER*
19 *Bull.*, **8**, 1–6.
- 20 Matsuoka, M., Wakamatsu, K., Fujimoto, K. & Midorikawa, S., 2006. Average shear-wave velocity
21 mapping using Japan engineering geomorphologic classification map, *J. Struc. Mech. Earthq.*
22 *Eng.*, **23**, 57s–68s.
- 23 Milkereit, B. & Eaton, D., 1998. Imaging and Interpreting the Shallow Crystalline Crust, *Tectono-*
24 *phys.*, **286**, 5–18.
- 25 Mitchell, B. J. & Yu, G., 1980. Surface Wave Dispersion, Regionalized Velocity Models, and
26 Anisotropy of the Pacific Crust and Upper Mantle, *Geophys. J. Int.*, **63**, 497–514.

- Montalbetti, J. F. & Kanasewich, E. R., 1970. Enhancement of Teleseismic Body Phases with a Polarization Filter, *Geophys. J. Int.*, **21**, 119–129, doi:10.1111/j.1365-246X.1970.tb01771.x.
- Mooney, W. D., Laske, G. & Masters, T.G., 1998. CRUST 5.1: A Global Crustal Model at 5 x 5, *J. Geophys. Res.*, **103**, 727–747.
- Nakata, N. & Snieder, R., 2012. Estimating Near-Surface Shear Wave Velocities in Japan by Applying Seismic Interferometry to KiK-net Data, *J. Geophys. Res.*, **117**, B01308, doi:10.1029/2011JB008595.
- National Geophysical Data Center, 2006. 2-minute Gridded Global Relief Data (ETOPO2) v2. National Geophysical Data Center, NOAA. doi:10.7289/V5J1012Q [accessed 01 Nov 2012].
- Obara, K., Kasahara, K., Hori, S. & Okada, Y., 2005. A Densely Distributed High-Sensitivity Seismograph Network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention, *Rev. Sci. Instrum.*, **76**, 21301, doi:10.1063/1.1854197.
- Odum, J. K., Stephenson, W. J., Williams, R. A. & von Hillebrandt-Andrade, C., 2013. Vs30 and Spectral Response from Collocated Shallow, Active-, and Passive-Source Vs Data at 27 Sites in Puerto Rico, *Bull. Seismol. Soc. Am.*, **103**, 2709–2728.
- Okada, Y., Kasahara, K., Hori, S., Obara, K., Sekiguchi, S., Fujiwara, H. & Yamamoto, A., 2004. Recent progress of seismic observation networks in Japan: Hi-net, F-net, K-NET and KiK-net, *Earth, Planets, Space*, **56**, xv–xxviii.
- Pearson, K., 1901. Principal Components Analysis, London, Edinburgh Dublin, *Philos. Mag. J.*, **6**, 566.
- Peng, Z. & Ben-Zion, Y., 2004. Systematic analysis of crustal anisotropy along the KaradereDüzce branch of the North Anatolian fault. *Geophys. J. Int.*, **159**, 253–274.
- Rinne, E. E., 1994. Development of New Site Coefficients For Building Codes, *Proc. of the 5th US Nat. Conf. on Earthq. Eng.*, **3**, 69–78.
- Ritzwoller, M. H. & Levshin, A. L., 1998. Eurasian Surface Wave Tomography: Group Velocities, *JGR* **103**, 4839–4878.
- Ritzwoller, M. H., Barmin, M. P., Villasenor, A., Levshin, A. L. & Engdahl, E. R., 2002. Pn and

- 26 Sn Tomography Across Eurasia To Improve Regional Seismic Event Locations, *Tectonophys.*,
 1 **358**, 39–55.
- 2 Roser, J. & Gosar, A., 2010. Determination of Vs30 for Seismic Ground Classification in the
 3 Ljubljana Area, Slovenia, *Acta Geotech. Slov.*, **7**, 61–76.
- 4 Salisbury, M. H., Iuliucci, R. & Long, C., 1994. Velocity and Reflection Structure of the Sudbury
 5 Structure from Laboratory Measurements, *Geophys. Res. Lett.*, **21**, 923–926.
- 6 Samson, J. C. & Olson, J. V., 1981. DataAdaptive Polarization Filters for Multichannel Geophysi-
 7 cal Data, *Geophysics*, **46**, 1423–1431, doi:10.1190/1.1441149.
- 8 Schmandt, B. & Lin, F. C., 2014. P and S Wave Tomography of the Mantle Beneath the United
 9 States, *Geophys. Res. Lett.*, **41**, 6342–6349.
- 10 Schulte-Pelkum, V., Masters, G. & Shearer, P. M., 2001. Upper Mantle Anisotropy from Long-
 11 Period P Polarization, *J. Geophys. Res.*, **106**, 21917–21934, doi:10.1029/2001JB000346.
- 12 Semblat, J. F. & Pecker, A., 2009. *Waves and Vibrations in Soils: Earthquakes, Traffic, Shocks,*
 13 *Construction Works*, IUSS Press.
- 14 Shapiro, N. M. & Campillo, M., 2004. Emergence of Broadband Rayleigh Waves From Correla-
 15 tions of the Ambient Seismic Noise, *Geophys. Res. Lett.*, **31**, L07614, doi:10.1029/2004GL019491.
- 16 Shearer, P. & Orcutt, J., 1987. Surface and Near-Surface Effects on Seismic WavesTheory and
 17 Borehole Seismometer Results, *Bull. Seismol. Soc. Am.*, **77**, 1168–1196.
- 18 Stehly, L., Campillo, M., Froment, B. & Weaver, R. L., 2008. Reconstructing Green’s Function by
 19 Correlation of the Coda of the Correlation (C3) of Ambient Seismic Noise, *J. Geophys. Res.*,
 20 **113**, B11306, doi:10.1029/2008JB005693.
- 21 Stokoe, K. H. & Woods, R. D., 1972. In situ shear wave velocity by cross-hole method, *J. Soil*
 22 *Mech. Found. Div.*, **ASCE SM5**, 443–460.
- 23 Svenningsen, L. & Jacobsen, B. H., 2007. Absolute S -Velocity Estimation from Receiver Func-
 24 tions, *Geophys. J. Int.*, **170**, 1089–1094, doi:10.1111/j.1365-246X.2006.03505.x.
- 25 Tanaka, A., Yamano, M., Yano, Y. & Sasada, M., 2004. Geothermal gradient and heat flow data in
 26 and around Japan (I), *Earth, Planets, Space*, **56**, 1191–1194.

- 27 Vidale, J. E., 1986. Complex Polarization Analysis of Particle Motion, *Bull. Seismol. Soc. Am.*, **76**,
1 1393–1405.
- 2 Wagner, G. S., 1997. Regional Wave Propagation in Southern California and Nevada: Observations
3 from a Three-Component Seismic Array, *J. Geophys. Res.*, **102**, 8285–8311, doi:10.1029/96JB02460.
- 4 Wald, D. J. & Allen, T. I., 2007. Topographic Slope as a Proxy for Seismic Site Conditions and
5 Amplification, *Bull. Seismol. Soc. Am.*, **97**, 1379–1395, doi:10.1785/0120060267.
- 6 Wessel, P. & Smith, W.H., 1991. Free software helps map and display data, *EOS, Trans. Am.*
7 *geophys. Un.*, **72**, 441–446.
- 8 Wiechert, E., 1907. Über Erdbebenwellen. Part I: Theoretisches über die Ausbreitung der Erd-
9 bebenwellen, Nachrichten von der K. Gesellschaft der Wissenschaften zu Göttingen, *Math.*
10 *Phys. Klasse*, 415–529.
- 11 Wu, R.-S., Xu, Z. & Li, X.-P., 1994. Heterogeneity Spectrum and Scale-Anisotropy in the Upper
12 Crust Revealed by the German Continental Deep-Drilling (KTB) Holes, *Geophys. Res. Lett.*,
13 **21**, 911–914.
- 14 Yilmaz, Ö, 2001. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic*
15 *Data*, Society of exploration geophysicists.

16 **Appendix A: Effect of the Borehole**

1 One complication for the Hi-net data arises from the fact that the instruments are located in a
 2 borehole. The relationship between the seismic wave speed and the wave polarisation described
 3 in the Method section is based on the instrument at the free surface such that the incident and
 4 reflected waves superimpose to generate the observed ground motion. At a borehole station, the
 5 incident and reflected waves arrive at distinct times, potentially requiring a different framework for
 6 their analyses.

7 We have performed synthetic tests to examine the effect of different borehole depths on po-
 8 larisation measurements and the inferred wave speeds with the surface-station assumption (Fig-
 9 ure A1). The synthetic seismograms at different depths are generated by convolving Equations 1
 10 and 8 with the incoming P and S waveforms, respectively. In order to perform the test with realistic
 11 waveforms, we use normalised P and S seismograms from an Mw 6.6 earthquake that occurred
 12 in Molucca Sea on August 26, 2012 as incoming P and S waves, respectively. The ray parameter
 13 of 7 and 13 s/deg for P and S data, respectively, and the compressional and shear wave speeds of
 14 3.2 and 1.7 km/s, respectively, are used as input values, which corresponds to incident angles of
 15 11.6° for P and 11.5° for S. The instrument depths of 0 (surface), 0.1, 0.2, 0.3, 0.5, 1, 2 km, are
 16 considered where 0.1 and 0.2 km are the depths at which nearly 90% of the Hi-net instruments
 17 are located. Another 5% of the instruments are situated between 0.3- and 0.5-km depths, and the
 18 remaining 5% are at deeper depths, as deep as 2 km.

19 At each depth, the P- and S-polarisation angles are measured with uncertainties associated with
 20 the linearity of the particle motions. Using these measurements, compressional and shear wave
 21 speeds are estimated for comparison against the input values. In order to investigate the effect of
 22 the noise in the data, we repeat the polarisation measurement and wave speed estimation using
 23 noisy synthetic waveforms. The waveforms are obtained by adding noise signal from 500 different
 24 noise time windows of real data to the synthetics with the median SNR of our data, i.e., 7.5 for P

25 and 3 for S wave. They result in 500 wave speed estimates, where their standard deviation becomes
 1 the uncertainty in the estimate representing the effect of noise.

2 The synthetic waveforms for a station at the surface have maximum amplitude of twice the
 3 incident P wave, as expected from the free surface effect. The particle motions are perfectly linear,
 4 with the polarisation measured as 12.3° and 12.0° for P and S incidences, respectively, where the
 5 S angle is measured for the direction orthogonal to the particle motion. Note that these angles are
 6 different from the incident angles, demonstrating the effects due to the reflected waves. There is
 7 zero uncertainty for both P and S measurements based upon the linearity of noise-free synthetics,
 8 since the signal is perfectly linear ($w^P = w^S = 1$). Using the measurements of P and S apparent
 9 angles and the 500 noise-added synthetics, the corresponding wave speed estimates are 3.2 ± 1.6
 10 and 1.7 ± 0.3 km/s for compressional and shear wave speeds, respectively, values which agree with
 11 the input compressional and shear wave speeds. Note that the uncertainties are large since, effec-
 12 tively, only one P and one S measurements are used to search for wave speeds, and should not be
 13 confused with the uncertainties arising from fitting multiple P and S measurements.

14 For instruments at depth, the same calculation is performed to obtain the angle, wave speeds,
 15 and their uncertainties. The P- and S-polarisation angles decrease and increase, respectively, down
 16 to about 0.5 km, which results from the later arrival of the reflected S wave that reduces the relative
 17 amplitude of the horizontal component. Also note that since S waves are longer period than P
 18 waves, the differences in the arrival times of incident and reflected waves have smaller effect in
 19 changing the total wavefields and measured angles. Therefore, the inferred S angles remain more
 20 stable than P angles at shallow depths, e.g., less than 2° change for S compared to about 4° change
 21 for P in the top 0.5 km. This is also reflected in the linearity of the particle motion, where the
 22 uncertainties increase as a function of depth, with S maintaining higher linearity than the P motion.
 23 On the other hand, the inferred wave speeds decrease for both compressional and shear waves with
 24 depth; for example, at 0.1 km, the compressional and shear wave speeds are calculated to be 2.9 and
 25 1.6 km/s, respectively, 0.3 and 0.1 km/s less than the true values. It suggests that the wave speeds
 26 may be underestimated at Hi-net borehole depths, and the magnitude of the underestimation is

larger for the compressional wave speed. After about 0.5 km, the polarisation measurements and the corresponding wave speeds either increase or decrease, depending on the waveform of the incident wave, which implies that interpreting the wave speed estimates for stations deeper than 0.5 km requires caution. However, only a small fraction of stations are at such depths, and their wave speed estimates do not affect the overall conclusions of the study.

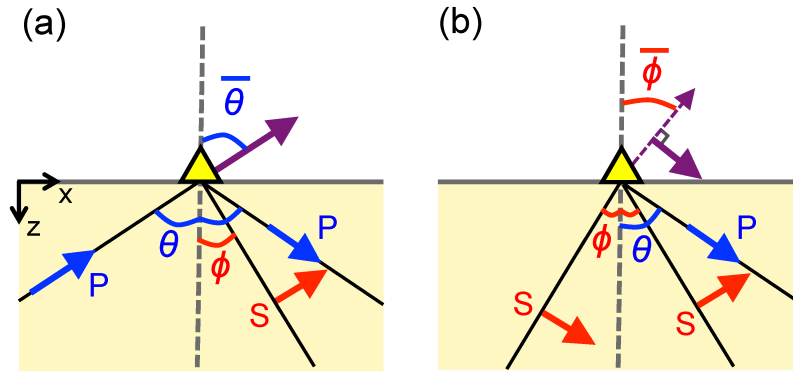


Figure 1. Geometry of the incident, reflected, and observed particle motions for incident (a) P and (b) S waves at a station (triangle). The particle motion or polarization directions are shown by arrows for P (blue), S (red), and apparent direction (purple). The angles θ and $\bar{\theta}$ correspond to the incident and the apparent polarisation directions of the P wave, respectively, measured from the vertical (dashed line). Similarly, the angles ϕ and $\bar{\phi}$ correspond to the wavevector directions perpendicular to the particle motions of the incident and the apparent polarisation of S wave, respectively, measured from the vertical.

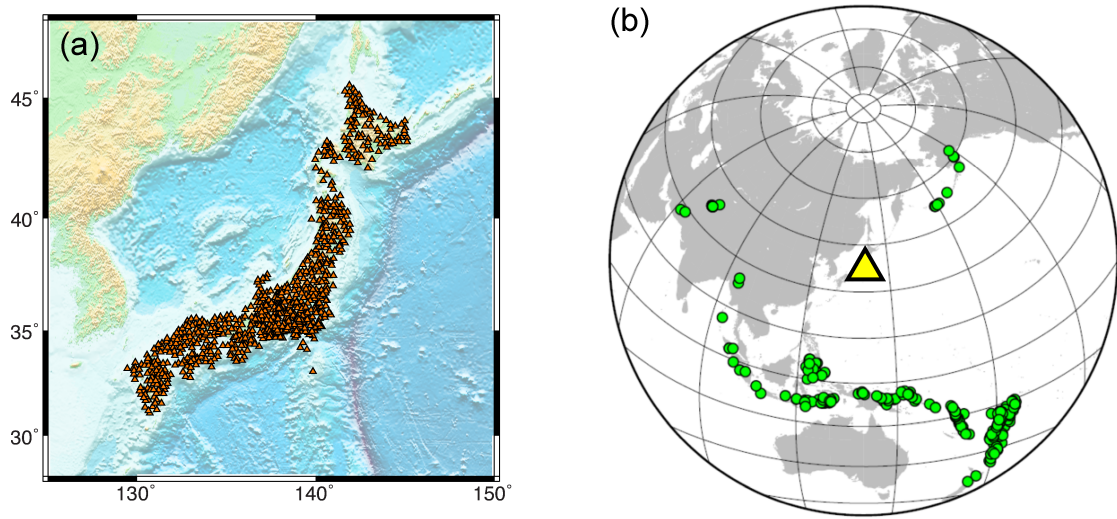


Figure 2. (a) Distribution of the Hi-net stations (orange triangles) plotted on a topography map based on ETOPO2 (National Geophysical Data Center 2006). (b) Distribution of the intermediate and deep events (green circles) at teleseismic distances with respect to the Hi-net (yellow triangle).

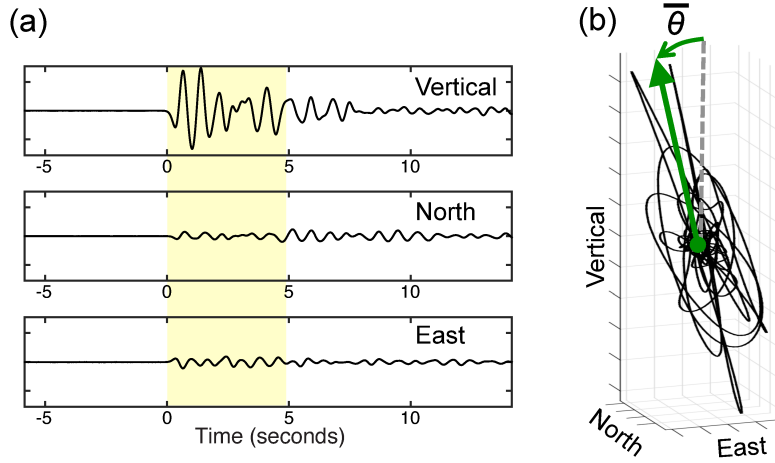


Figure 3. (a) Three-component velocity seismograms recorded at the SGOH station, showing the teleseismic P-wave arrival from an Mw 6.9 event that occurred on July 21, 2014 at 71.2° from the station. The yellow shade indicates the 5-second time window used for the PCA. (b) The particle motion during the 5-second window shown in (a). The PCA allows identification of the polarisation direction (thick green arrow) and measurement of the angle $\bar{\theta}$ from the vertical (dashed line).

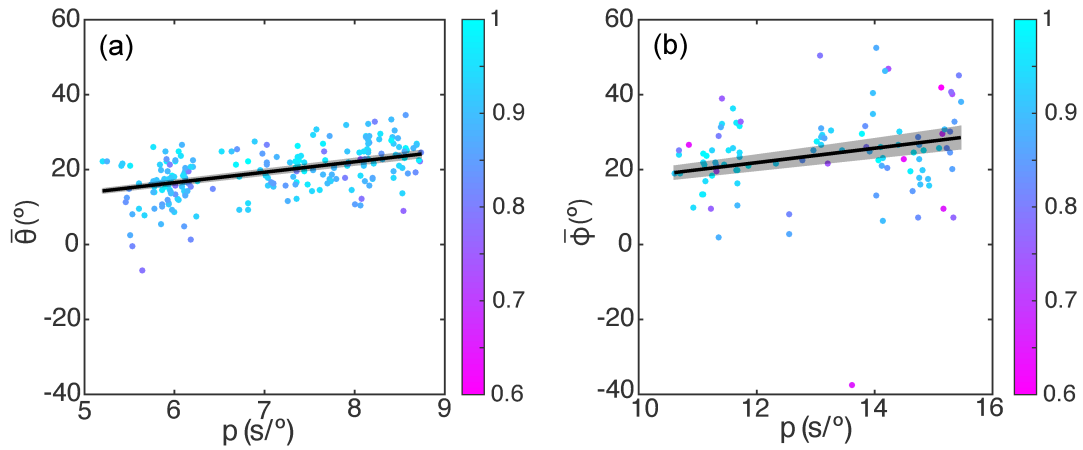


Figure 4. Measured (a) P- and (b) S-polarisation angles (coloured circles) at station KAWH. The data are shown as a function of theoretical ray parameter calculated using IASP91 (Kennett & Engdahl 1991). The colours represent the quality of each measurement, $\frac{\lambda_1}{\lambda_1 + \lambda_2}$. The black solid line and the shade are the predicted angles using the wave speeds of 4.1 ± 0.2 km/s for P and 2.7 ± 0.1 km/s for S based on Equations 6 and 12.

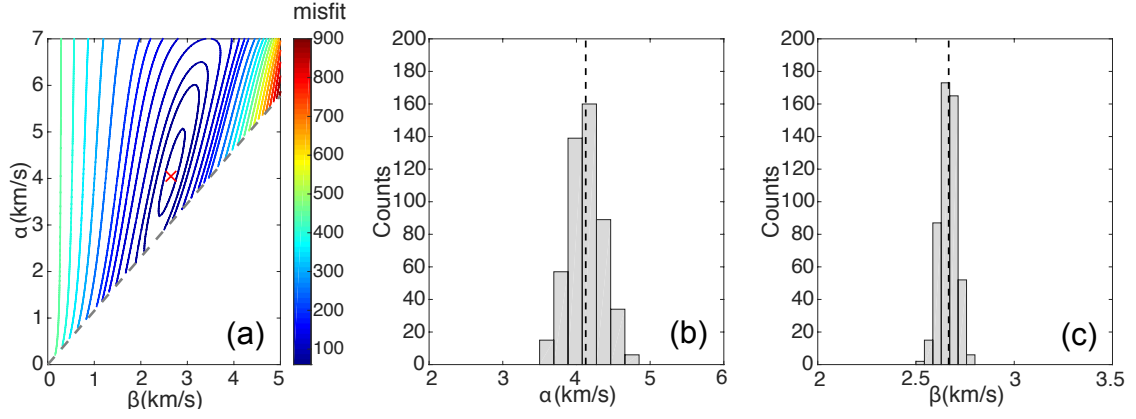


Figure 5. (a) Grid search result for station KAWH. The grey dashed line marks the constraint $\beta \leq \sqrt{3}\alpha/2$, and the upper trapezoid region above this line is the search area with the contours showing the misfit. A global minimum is marked by a red cross. (b) and (c) The distribution of compressional (b) and shear (c) wave speeds obtained by bootstrapping. The grid search is performed on 500 randomly resampled data sets. The mean value (dashed line) is taken as the final wave speed estimate and used to obtain the black solid lines in Figure 4(a) and (b). The standard deviation is used as the uncertainty estimate.

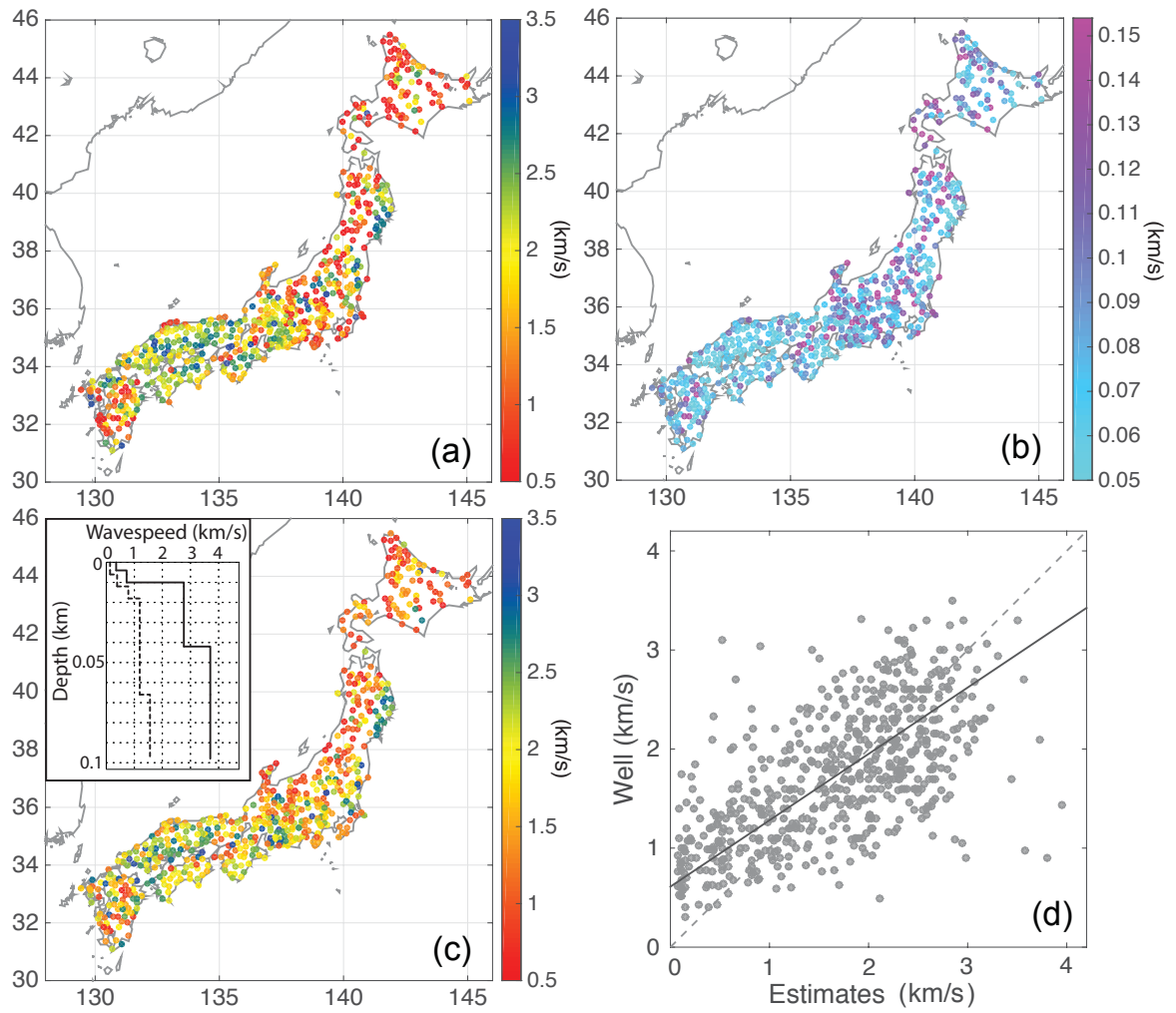


Figure 6. (a) Shear wave speed estimates where each coloured circle corresponds to an estimate for one of the Hi-net stations. Red is slow and blue is fast, and the estimates range from 0.5 to 3.5 km/s. (b) Same as (a) except that the colour of each circle represents the uncertainty estimate at each station. (c) Same as (a) except that the map is for the well measurement (Obara et al. 2005). The colour bar has the same scale as in (a). The inset shows an example well data, i.e., the compressional (solid line) and shear (dashed line) wave speeds from the surface to the borehole instrument at station HYSH (modified from Obara et al. 2005). (d) The scatter plot between the shear wave estimates and the well values. Black solid line is the linear regression line and the grey dashed line is the 1:1 line.

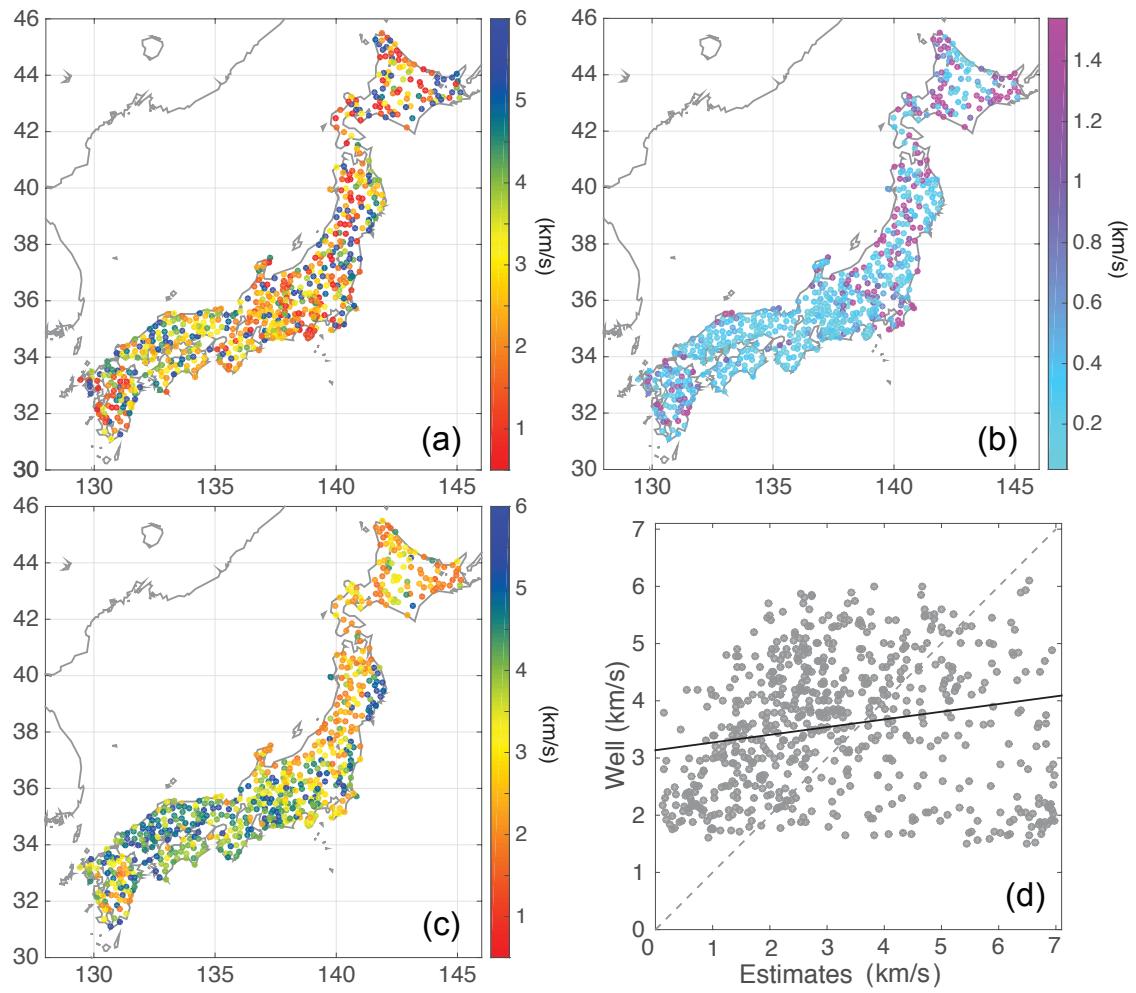


Figure 7. Same as Figure 6, except that the panels are for the compressional wave speed. (a) and (c) have the same colour bar. Note that the colour scale for (b) is significantly different from Figure 6(b), showing that the shear wave speeds are better constrained than the compressional wave speeds.

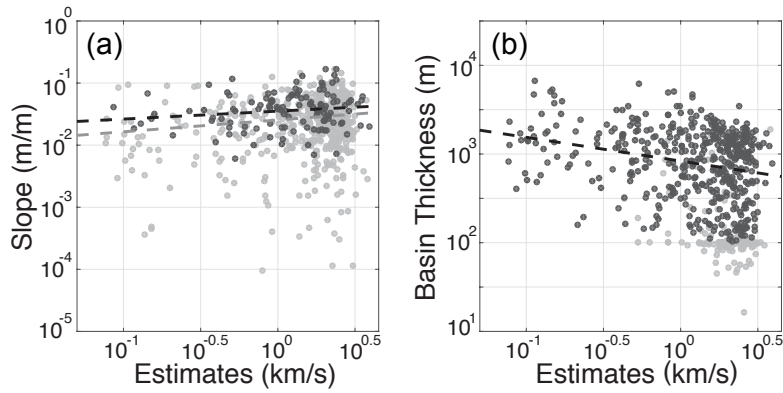


Figure 8. (a) Comparison between the shear wave speed estimates and the topographic slope based on ETOPO2 (distributed by National Centers for Environmental Information) in log scale. The stations within 25 km from any volcano locations are shown by black circles while the others are plotted as grey circles. The black and grey dashed lines result from linear regression of black and grey data points, respectively. (b) Comparison between the shear wave speed estimates and the basin depth derived from the Japan Integrated Velocity Structure Model (Koketsu et al. 2012). Most instrument depths are within the basin (black circles), while some are deeper than the estimated basin depths (grey circles). The black dashed line shows the linear trend in the black data points.

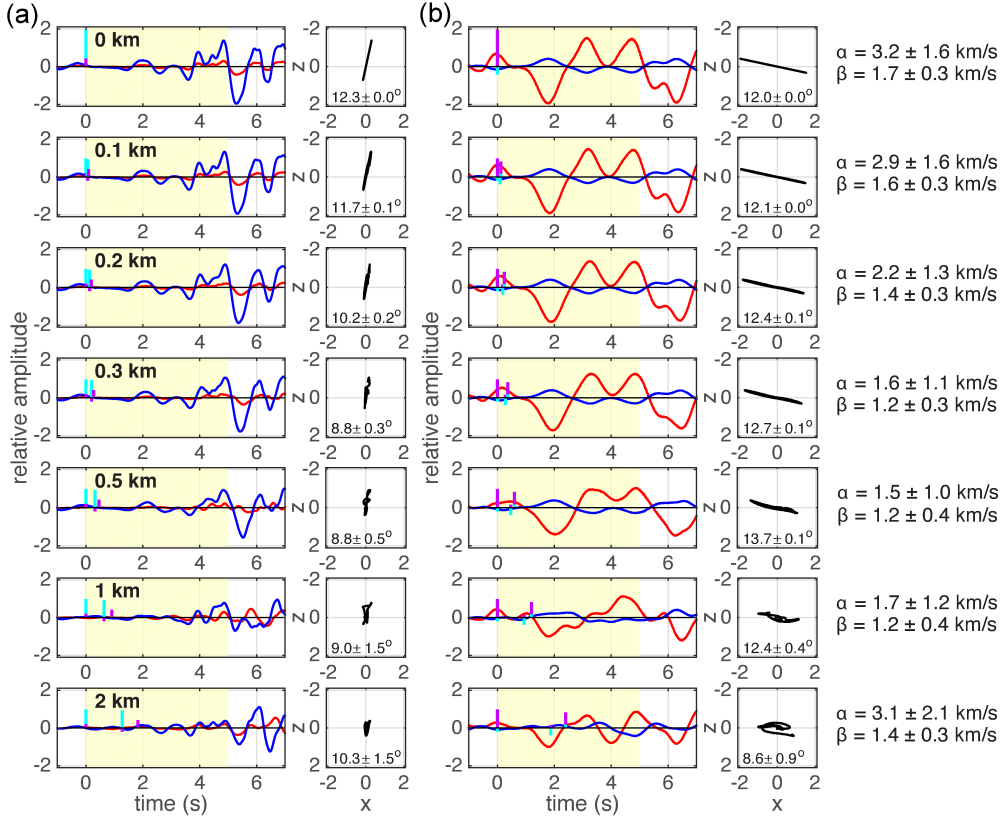


Figure A1. Synthetic analyses showing the effect of the borehole depth. (a) Each row represents different borehole depth ranging from 0 to 2 km. The left panel shows the radial (red line) and vertical (blue line) motion resulting from an incoming P wave combined with reflected P and S waves computed based on Equation 1. The incident P waveform is taken from the vertical P waveform of 2012/08/26 Mw 6.6 earthquake in Molucca Sea region recorded at TOBH station, and is identical to the waveform computed at the surface, i.e., 0 km. The theoretical ray parameter is 7 s/deg and the input compressional and shear wave speeds are 3.2 and 1.7 km/s, respectively. The thick cyan and purple sticks represent the predicted vertical and radial amplitude of incoming P, reflected P, and reflected S motion arising from an impulsive source with unit amplitude. Only for the surface instrument, the three phases arrive at the same time, and with increasing depth, the arrivals become more separated. The right panel shows the particle motion in the radial (x)-vertical (z) plane for the first 5 seconds corresponding to the yellow shaded time window in the left panel. The angle at the bottom is the apparent polarisation angle measured using PCA, where the uncertainty estimates are obtained from the linearity of the particle motion. (b) Same as (a) except that the analysis is for the S wave polarisation using the radial S waveform from the same earthquake and the theoretical ray parameter of 13 s/deg. For each depth, the compressional (α) and shear (β) wave speeds corresponding to the measured P and S angles are indicated on the right, where uncertainties are obtained using 500 different noise-added synthetic data.

Supplementary Materials

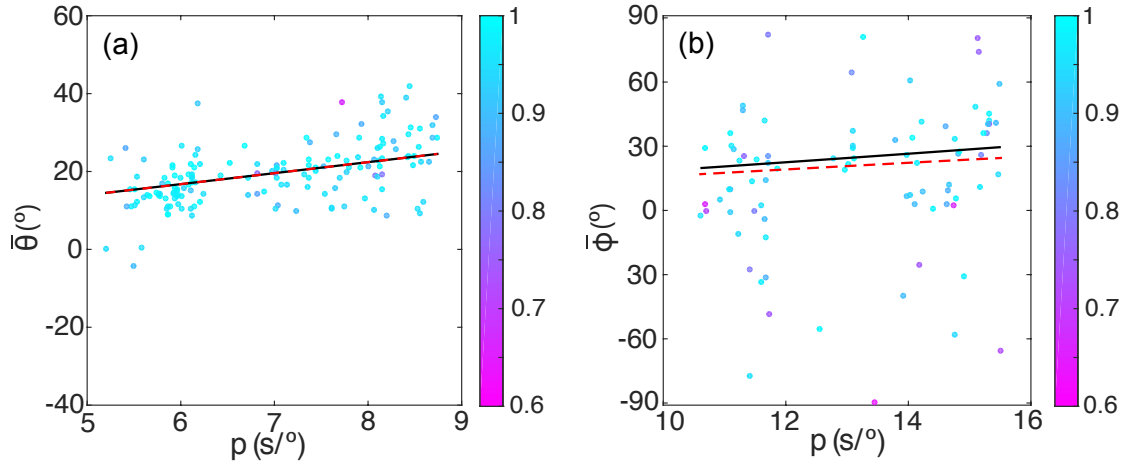


Figure S1. Same as Figure 4 but the length of the time window for PCA is 1 second. The red dashed lines are the predicted angles using the best fit wave speeds of 4.7 and 2.1 km/s for P and S waves, respectively. The black solid lines are the same as the black solid lines from Figure 4. Note that the range of the vertical axis in (b) is significantly larger than that of Figure 4(b).

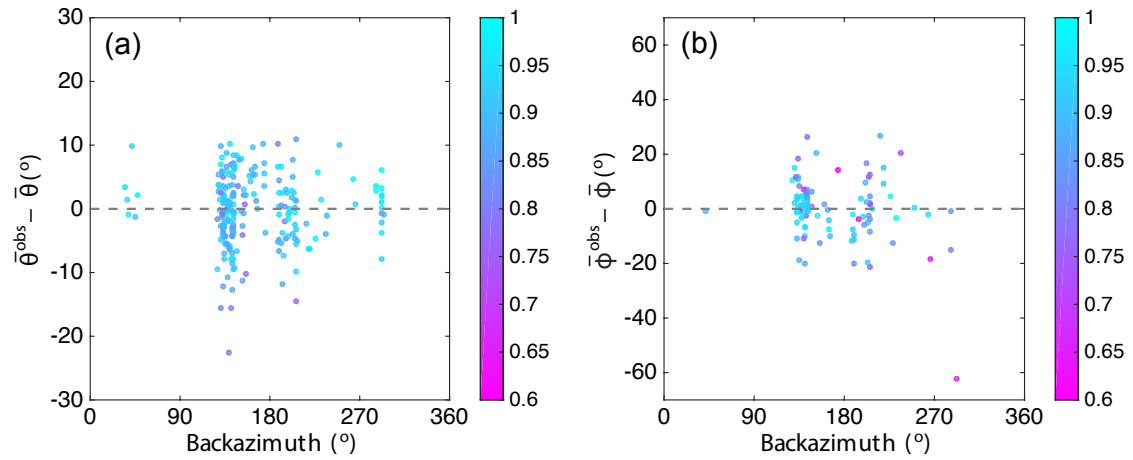


Figure S2. Residuals of (a) P- and (b) S-polarisation angles (coloured circles) at station KAWH. Similar to Figure 4 but the residuals are plotted against the backazimuth. The grey dashed line marks zero degree.