Geophysical Journal International

Geophys. J. Int. (2019) **219,** 1491–1513
Advance Access publication 2019 August 05
GJI Seismology

doi: 10.1093/gji/ggz364

A reappraisal of the H–κ stacking technique: implications for global crustal structure

C.S. Ogden [1], I.D. Bastow,[1] A. Gilligan[2] and S. Rondenay[3]

[1] *Department of Earth Science and Engineering, Imperial College London, London SW7 2BP, UK. E-mail: co1012@ic.ac.uk*
[2] *School of Geosciences, University of Aberdeen, Aberdeen AB24 3FX, UK*
[3] *Department of Earth Science, University of Bergen, Bergen 7803 N-5020, Norway*

Accepted 2019 August 4. Received 2019 June 24; in original form 2019 February 19

SUMMARY

H–κ stacking is used routinely to infer crustal thickness and bulk-crustal V_P/V_S ratio from teleseismic receiver functions. The method assumes that the largest amplitude *P*-to-*S* conversions beneath the seismograph station are generated at the Moho. This is reasonable where the crust is simple and the Moho marks a relatively abrupt transition from crust to mantle, but not if the crust–mantle transition is gradational and/or complex intracrustal structure exists. We demonstrate via synthetic seismogram analysis that H–κ results can be strongly dependent on the choice of stacking parameters (the relative weights assigned to the Moho *P*-to-*S* conversion and its subsequent reverberations, the choice of linear or phase-weighted stacking, input crustal *P*-wave velocity) and associated data parameters (receiver function frequency content and the sample of receiver functions analysed). To address this parameter sensitivity issue, we develop an H–κ approach in which cluster analysis selects a final solution from 1000 individual H–κ results, each calculated using randomly selected receiver functions, and H–κ input parameters. 10 quality control criteria that variously assess the final numerical result, the receiver function data set, and the extent to which the results are tightly clustered, are used to assess the reliability of H–κ stacking at a station. Analysis of synthetic data sets indicates H–κ works reliably when the Moho is sharp and intracrustal structure is lacking but is less successful when the Moho is gradational. Limiting the frequency content of receiver functions can improve the H–κ solutions in such settings, provided intracrustal structure is simple. In cratonic Canada, India and Australia, H–κ solutions generally cluster tightly, indicative of simple crust and a sharp Moho. In contrast, on the Ethiopian plateau, where Palaeogene flood-basalts overlie marine sediments, H–κ results are unstable and erroneous. For stations that lie on thinner flood-basalt outcrops, and/or in regions where Blue Nile river incision has eroded through to the sediments below, limiting the receiver function frequency content to longer periods improves the H–κ solution and reveals a 6–10 km gradational Moho, readily interpreted as a lower crustal intrusion layer at the base of a mafic ($V_P/V_S = 1.77$–1.87) crust. Moving off the flood-basalt province, H–κ results are reliable and the crust is thinner and more felsic ($V_P/V_S = 1.70$–1.77), indicating the lower crustal intrusion layer is confined to the region covered by flood-basaltic volcanism. Analysis of data from other tectonically complex settings (e.g. Japan, Cyprus) shows H–κ stacking results should be treated cautiously. Only in regions of relatively simple crust can H–κ stacking analysis be considered truly reliable.

Key words: Statistical methods; Body waves; Crustal imaging; Cratons; Crustal structure; Large igneous provinces.

1 OVERVIEW

The H–κ stacking technique (Zhu & Kanamori 2000) is a widely used method to obtain bulk-crustal information from teleseismic receiver functions by searching for the combination of Moho depth (H) and V_P/V_S ratio (κ) that maximizes the amplitude sum of *P*-to-*S* conversions from beneath a seismograph station. Central to the H–κ method is the assumption that the Moho is the sharpest substation velocity contrast, and that it produces the largest amplitude *P*-to-*S* conversions and reverberations in the receiver function (Fig. 1a). Previous studies have reduced the effect of noisy data in several ways: phase-weighted rather than linear stacking (Crotwell & Owens 2005) and varying the

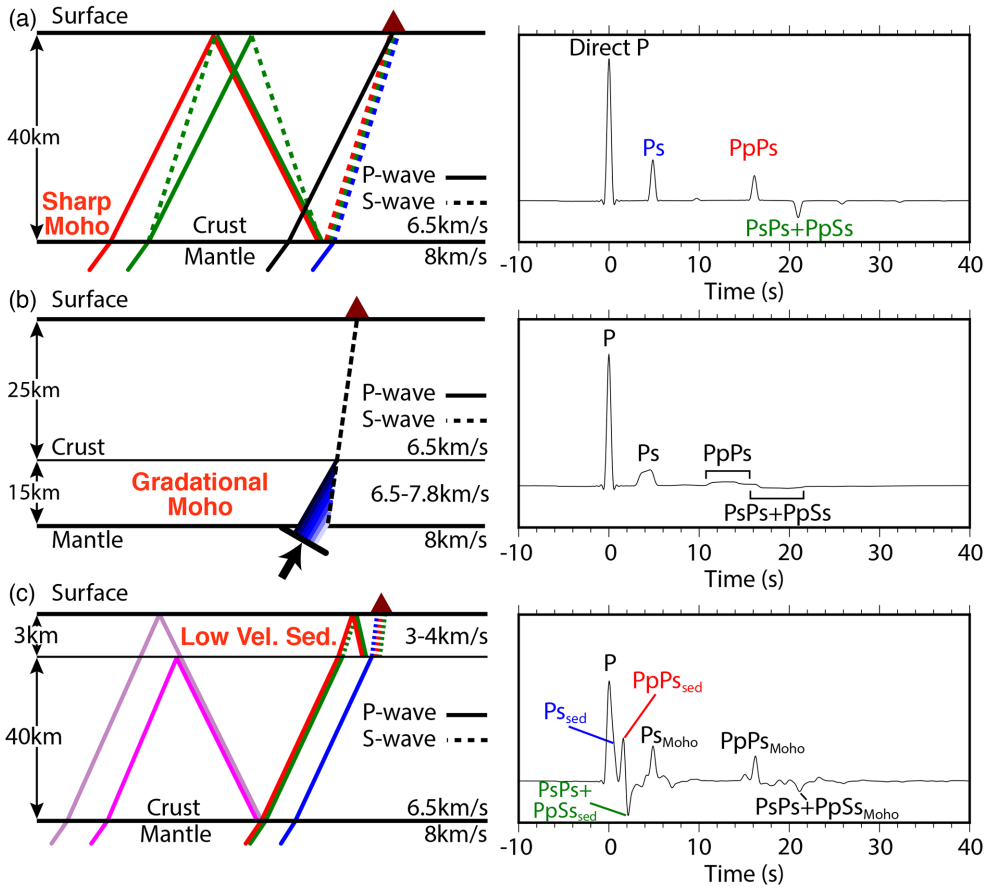

Figure 1. The impact of crustal structure on receiver functions. (a) Sharp Moho with high amplitude, impulsive *P*-to-*S* conversions. (b) A gradational Moho, for which *P*-to-*S* conversions occur over a large depth range. Receiver function signals weaker and more diffuse. (c) When near-surface layers exist, resulting *P*-to-*S* conversions can distort Moho signals.

weighting of the Moho *P*-to-*S* conversion relative to subsequent reverberations (e.g. Eaton *et al.* 2006; Thompson *et al.* 2010; Vanacore *et al.* 2013). Other studies have highlighted the importance of anisotropy (e.g. Levin & Park 2000) and backazimuthal variations in crustal structure (e.g. Dugda *et al.* 2005) when interpreting H–κ results. However, as far as we have been able to determine, very few studies have addressed the fundamental question of whether H–κ stacking should be used at all in some complex tectonic settings. For example, in areas where the crust–mantle transition is gradational (e.g. regions of lower crustal intrusions (Mackenzie *et al.* 2005), or subduction zones (Bostock *et al.* 2002)), Moho *P*-to-*S* converted energy will have low amplitude (e.g. Gallacher & Bastow 2012; Fig. 1b). In such settings, H–κ stacking will only be sensitive to the Moho using longer period receiver functions (e.g. Frassetto *et al.* 2011). Where complex shallow crustal structure exists, significant *P*-to-*S* converted energy may mask signals from the Moho (Fig. 1c). In such scenarios, the fundamental single layer over a half-space assumption that underpins H–κ stacking breaks down.

In this contribution, we first take a forward modelling approach to exploring the sensitivity of H–κ stacking to complex crustal structure. We examine the impact of varying the H–κ stacking input parameters, including the relative weights assigned to the Moho *Ps* arrival and its subsequent crustal reverberations, the style of stacking employed (linear versus phase-weight), and the *a priori* choice of crustal *P*-wave velocity. We also test how the frequency content of the receiver functions can be used to ascertain whether a station is underlain by a sharp or gradational Moho. We then develop a cluster analysis approach to H–κ stacking that rigorously explores its parameter space, including the frequency content of the receiver functions. In doing so, we assign a score to each station using 10 criteria that variously assess data signal-to-noise ratio (SNR), the ability of a single pair of H and κ values to explain the observations for a given station, and the likelihood that the Moho is gradational rather than sharp.

We test our new method on multiple synthetic data sets and several tectonic settings worldwide. While the H–κ method can often yield accurate bulk-crustal information in regions of simple crustal structure, it can fail completely in regions where these conditions are not met. Our new approach can provide the analyst with a strong indication for why the H–κ method fails in certain tectonic settings. In such circumstances, more sophisticated seismological inversion techniques are thus required, such as joint-inversion of receiver functions with surface waves for 1-D structure beneath the station (e.g. Julià *et al.* 2009; Gilligan *et al.* 2016) or Markov Chain Monte Carlo receiver function analyses (e.g. Piana Agostinetti & Malinverno 2010; Wirth *et al.* 2016).

2 REVIEW OF RECEIVER FUNCTIONS AND H–κ STACKING

Receiver functions are time-series calculated from three component seismograms that capture *P*-to-*S* conversions from velocity discontinuities below a seismograph station (e.g. Langston 1979). The H–κ stacking technique (Zhu & Kanamori 2000) utilizes the arrival times of the converted Moho arrivals *Ps*, *PpPs* and *PsPs* + *PpSs* (Fig. 1a) to determine H and κ, using a grid search of the plausible H and κ values to maximize the amplitudes of the three phases and, therefore, maximize the stacking function of the linear 'stack', $s(H, \kappa)$:

$$s(H,\kappa) = \sum_{j=1}^{N} w_1 r_j(t_1) + w_2 r_j(t_2) - w_3 r_j(t_3), \tag{1}$$

where w_1, w_2, w_3 are stacking weights (satisfying $\sum w_i = 1$) that govern the influence of each converted phase. $r_j(t_i)$ are the receiver function amplitudes at the predicted arrival times of the direct *P*-to-*S* conversion (*Ps*) and subsequent reverberations (*PpPs* and *PsPs* + *PpSs*), respectively, for the *j*th receiver function. *N* is the number of receiver functions stacked to improve the SNR. In this study, the $s(H, \kappa)$ grid search is performed using 100 values of both H and κ. The predicted traveltimes for each phase, t_i are given by eqs (2)–(4).

$$t_1 = H\left[\sqrt{\frac{1}{V_S^2} - p^2} - \sqrt{\frac{1}{V_P^2} - p^2}\right], \tag{2}$$

$$t_2 = H\left[\sqrt{\frac{1}{V_S^2} - p^2} + \sqrt{\frac{1}{V_P^2} - p^2}\right], \tag{3}$$

$$t_3 = 2H\sqrt{\frac{1}{V_S^2} - p^2}, \tag{4}$$

where *p* is the ray parameter.

Phase-weighted stacking (PWS) has been used to reduce the effect of incoherent noise (Schimmel & Paulssen 1997). This is particularly important where Moho signals are weak owing to complex Moho and crustal structure (e.g. Crotwell & Owens 2005). PWS modulates the linear stack with the coherency (*c*) of the instantaneous phases for each receiver function (eq. 5), amplifying coherent signals but damping incoherent noise:

$$c(H,\kappa) = \frac{1}{N}\sum_{j=1}^{N}\frac{\sum_{k=1}^{N} e^{i\,\Phi(t_k)}}{3}, \tag{5}$$

where Φ is the instantaneous phase at time *t*. Values of *c* range from 0–1 with 0 representing incoherent stacking and 1 representing a completely coherent stack (Schimmel & Paulssen 1997). This is applied to the linear stack as follows:

$$s(H,\kappa) = c^{\nu}\sum_{j=1}^{N} w_1 r_j(t_1) + w_2 r_j(t_2) - w_3 r_j(t_3), \tag{6}$$

where ν controls the sharpness of the PWS filtering. The linear stack is retrieved if $\nu = 0$, $\nu = 2$ represent PWS.

Stacking weights (w_1, w_2, w_3) are often picked on the assumption that *Ps* is the highest amplitude and clearest arrival and so should have highest weight; *PpPs* and *PsPs* + *PpSs* are lower amplitude so are generally assigned lower weights in the literature (e.g. Eaton *et al.* 2006). However, consensus on which values should be used is lacking. The stacking weights are often assigned in a 0.6:0.3:0.1 ratio or similar (e.g. Dugda *et al.* 2005; Thompson *et al.* 2010; Vanacore *et al.* 2013) but the precise choice is usually somewhat ad hoc. For H–κ stacking, *P*-wave velocity (V_P) is held constant for the whole crust and has to be known *a priori* or assumed. V_P is often unknown outside areas studied by wide-angle seismic reflection/refraction (e.g. Mackenzie *et al.* 2005), so the resulting uncertainties in H and κ must be borne in mind. Specific H–κ stacking input parameters are the input V_P, the stacking weights (w_1, w_2 and w_3) and the type of stacking applied (linear or phase weighted). Additionally, the receiver function frequency content and the subset of receiver functions for a given station used in H–κ analysis are data parameters that can be varied during H–κ stacking analysis.

Previous studies have calculated measurement errors using the shape of $s(H, \kappa)$ (e.g. Zhu & Kanamori 2000; Eaton *et al.* 2006), however, Crotwell & Owens (2005) found this sometimes produced implausibly low errors. Instead, they used a bootstrapping algorithm that resampled the receiver functions multiple times for a given station and used the associated standard deviations in H and κ as error estimates. We calculate both data set derived and $s(H, \kappa)$ derived errors in this study. We require measurement errors in both H (eq. 7) and κ (eq. 8) for each H–κ stacking attempt, which are calculated using the maximum bounds of the 95 per cent contour of $s(H, \kappa)$ (Fig. 2). The generally elliptical nature of the contour provides an uncertainty about the exact $s(H, \kappa)$ maxima. A tighter peak, resulting from a more certain stack will, therefore, have a smaller 95 per cent contour, and smaller errors:

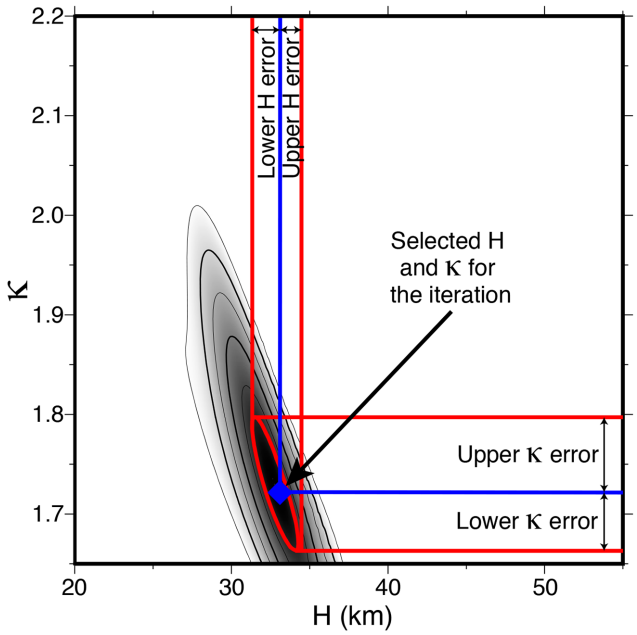

Figure 2. Errors in H and κ are calculated using the 95 per cent contour of $s(H, \kappa)$ for each repetition. In both H and κ, the mean average of the minimum and maximum errors is calculated to obtain the final errors in H and κ.

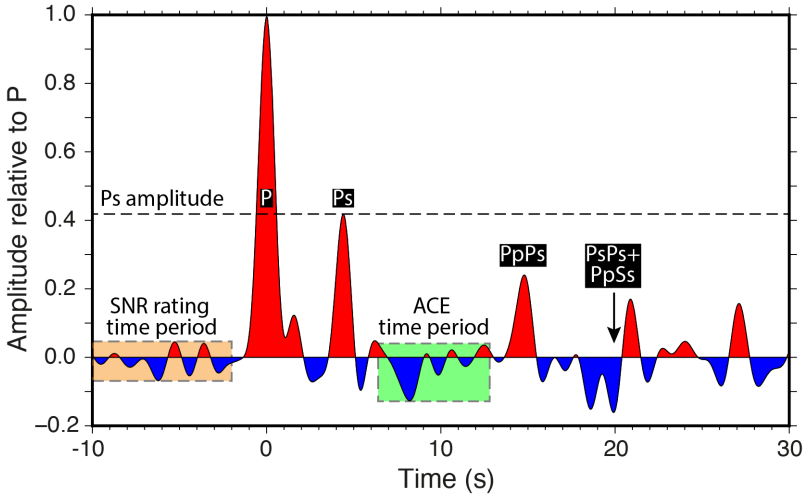

Figure 3. ACE value (demonstrated for a receiver function recorded at station HYB in Hyderabad, India) is determined by dividing the *Ps* amplitude by the RMS value of amplitude within the green shaded region. Similarly, the SNR value is calculated by dividing the *Ps* amplitude by the RMS amplitude of the orange shaded region prior to the *P* arrival.

$$H_{\text{error}} = \frac{(\text{Upper } H \text{ 95 per cent contour} - \text{Lower } H \text{ 95 per cent contour})}{2}, \tag{7}$$

$$\kappa_{\text{error}} = \frac{(\text{Upper } \kappa \text{ 95 per cent contour} - \text{Lower } \kappa \text{ 95 per cent contour})}{2}. \tag{8}$$

To address the issue of noise in receiver functions and to quantify the extent to which a receiver function carries information that cannot be described by a single layer over a half-space, we introduce the amplitude comparison estimate (ACE), the SNR and cross-correlation coefficient (CCC) measures. ACE (eq. 9) compares the amplitude at the predicted t_1 arrival time for each receiver function, assuming that H–κ stacking has correctly identified the Moho *Ps* arrival, with the root-mean-square (RMS) amplitude of the receiver function between $t_1 + 2\,s$ and $t_2 - 2\,s$ (Fig. 3):

$$\text{ACE} = \frac{1}{N}\sum_{j=1}^{N} r_j(t_1)\left(\frac{\sum_{t=t_{1+2s}}^{t_{2-2s}} r_j(t)}{\text{rate}(t_{2-2s} - t_{1+2s}) + 1}\right)^{-1/2}, \tag{9}$$

where *rate* is the sample rate of the receiver function ($r_j(t)$). A simple layer over half-space model with a sharp Moho theoretically has a larger amplitude *Ps* phase compared to the general signal of the receiver function. A gradational Moho produces a lower amplitude *Ps* phase than a sharp Moho, and thus a lower ACE. Similarly, a model with complex intracrustal structure will have additional *P*-to-*S* conversions between $t_1 + 2$ s and $t_2 - 2$ s, which lower the ACE. Using the predicted t_1 time from the chosen final H–κ solution, ACE becomes a measure of how prominently the *Ps* arrival stands out from the rest of the receiver function. The SNR (eq. 10) compares the amplitude of the predicted *Ps* phase (defined by the H–κ solution for that station) with the RMS amplitude of 8 s of pre-*P*-arrival noise (Fig. 3). A larger *Ps* amplitude, indicative of a sharper Moho, will produce a larger SNR than for a gradational Moho:

$$\mathrm{SNR} = \frac{1}{N}\sum_{j=1}^{N} r_j(t_1)\left(\frac{\sum_{t=t_{P-10\mathrm{s}}}^{t_{P-2\mathrm{s}}} r_j(t)}{\mathrm{rate}(t_{P-2\mathrm{s}} - t_{P-10\mathrm{s}}) + 1}\right)^{-1/2}. \tag{10}$$

Finally, the CCC tests the effect of noise and backazimuthal variations at a station by measuring the mean CCC of all possible pairs of receiver functions calculated with the same frequency, for each different frequency of receiver function. Stations with highly correlated receiver functions will yield more stable H–κ estimates. These three receiver function analytics (ACE, SNR and CCC) supplement the overall stacking approach from eqs (1) to (6) by providing direct information about how the final H–κ solution relates to the receiver functions used in the stack.

3 THE SENSITIVITY OF H–κ SOLUTIONS TO INPUT PARAMETER SELECTION

To determine the effect of varying each H–κ input (V_P, w_1, w_2, w_3 and stacking type) and data (receiver function frequency content and the subset of receiver functions chosen) parameter on the final result, we conduct tests using synthetic receiver functions that represent a variety of tectonic scenarios. Synthetic seismograms for flat-layered models are generated using the ray tracing program respknt (Randall 1989) with a 1 s Gaussian pulse and no noise added to demonstrate the purely seismological challenges that complex structures present during H–κ stacking. To test the effect of crustal anisotropy and a dipping Moho, we calculate seismograms using the Raysum method of Frederiksen & Bostock (2000). Horizontal component seismograms are rotated into radial and tangential components and receiver functions are computed using the extended-time multitaper frequency domain cross correlation receiver function estimation method (ETMTRF; Helffrich 2006). ETMTRF computes receiver functions using a low-pass $\cos^2$ taper with the maximum frequency chosen by the user. We automatically set the receiver function window as 10 s before the *P* arrival, and 100 s after the *P* arrival. Other receiver function calculation strategies exist (e.g. Langston 1979; Ligorría & Ammon 1999; Park & Levin 2000), and for moderate-to-high-quality seismograms they yield similarly robust results (Rondenay *et al.* 2016).

We present a sharp Moho model with an abrupt V_P change from mantle (8.0 km s^{-1}) to continental crust (6.5 km s^{-1}; Figs 4a, c and 6a). The second model replaces the sharp Moho with a gradual velocity change over a depth range of 15 km represented by a series of finite steps (Figs 4b, d and 6g). Conceptually, the steps represent a zone of lower crustal mafic intrusions (e.g. Mackenzie *et al.* 2005). In each test, H–κ input parameters are varied systematically between plausible limits defined in Table 1, with only one parameter varied for each test while the rest remain constant. We test the effect of stacking weights but retain the limit $w_3 \leq 0.5$ in line with the observation that *PsPs* + *PpSs* is usually a low-amplitude signal compared to *Ps*. w_1 and w_2 in most cases will thus have highest weights. However, we do test the effect of w_3 having the largest weight to allow thorough examination of the parameter space. This encompasses tectonic scenarios where, for example, dipping layers may produce larger *PsPs* + *PpSs* conversions than *PpPs* (Frederiksen & Bostock 2000). Because $\sum w_i = 1$, two of the stacking weights must change synchronously in the weight tests; 21 combinations of the stacking weights satisfy this condition. V_P is varied since it is often unknown *a priori*. To identify the influence of linear stacking and PWS, we re-run all the tests for both stacking strategies. Finally, previous studies have varied receiver function frequency content to investigate Moho sharpness (e.g. Frassetto *et al.* 2011), so we test this also. Lower frequencies are required to detect a gradational Moho to which higher frequency H–κ analysis is blind.

For the sharp Moho model (Figs 4a and c), the final H–κ solution is independent of frequency and stacking weights, and the grid search reliably returns the correct H–κ solution for both stacking types. For the gradational model (Figs 4b and d), the final result is highly dependent on the choice of both input and data parameters, and the input model is not identified correctly. For both models, as V_P increases, H increases and κ decreases systematically: a 0.1 km s^{-1} change in V_P translates to a ∼0.71 km variation in H and a ∼0.002 change in κ.

Fig. 4 demonstrates that the H–κ input parameters can dictate the H–κ solution obtained; it also indicates that the frequency content of receiver functions influences the H–κ solutions. To ascertain if receiver functions can resolve a crust–mantle boundary that is manifest over an increasingly larger depth range, we produce a suite of synthetic models with Moho thicknesses ranging from 0 to 15 km. In each model, the crustal V_P and V_S transition towards mantle values over a series of small steps to simulate a gradient; crustal V_P/V_S is 1.765 in all models. If H–κ stacking is reliable for a model, it will identify the centre of the Moho depth range (40 km) successfully. From the sharp Moho synthetic tests (Figs 4a and c), a 6.2–6.8 km s^{-1} change in V_P-produced ranges of 4.2 km and 0.013 in H and κ, respectively. Accounting for the individual measurement errors of those H and κ solutions, all subsequent H–κ solutions are expected to fit within the H range 37.1–42.9 km and the κ range 1.723–1.807.

For each Moho model, we sample the H–κ stacking input and data parameters randomly and repeat 1000 times (according to parameter ranges defined in Table 1). In each test, 80 per cent of the available synthetic receiver functions are selected randomly with no duplicates. Each of the 21 possible combinations of stacking weight has an equal chance of selection. We extract the H–κ stacking solutions and their

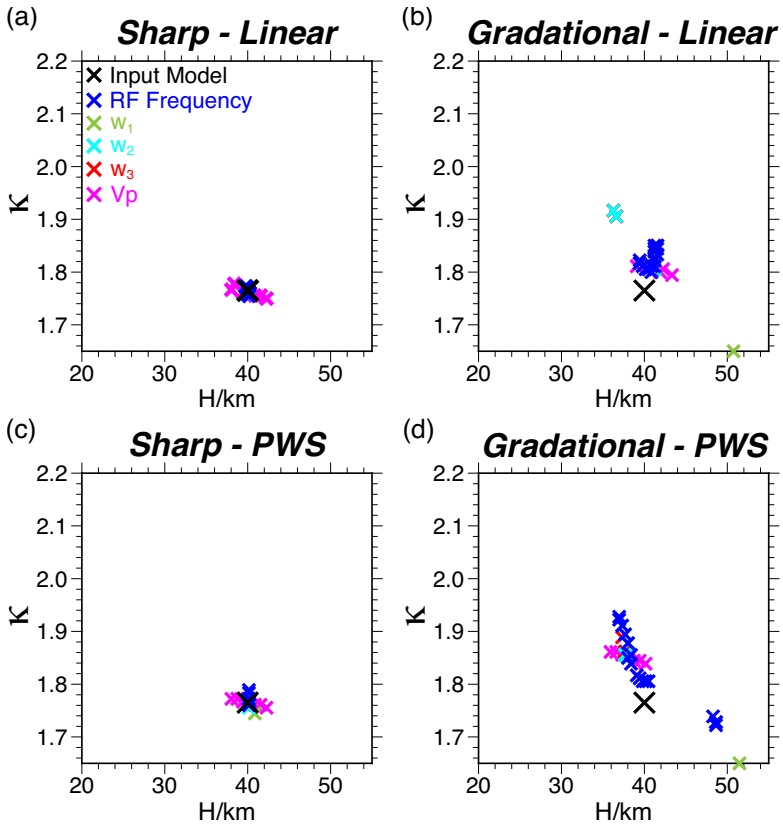

Figure 4. Distribution of H–κ results from (a) sharp Moho with linear stacking, (b) gradational Moho with linear stacking, (c) sharp Moho with phase-weight stacking, (d) gradational Moho with phase-weight stacking, as individual H–κ input parameters are varied. The black cross indicates the input model values of H and κ. The colour indicates the parameter being varied while the rest are held constant. n.b. in the case of w_1, w_2 and w_3, two are varied synchronously to satisfy $\sum w_i = 1$.

Table 1. H–κ input parameter ranges for analysis, resulting in 4998 possible unique combinations. There are 21 unique possible combinations of the stacking weights that satisfy $\sum w_i = 1$.

Parameter	Minimum	Maximum	Interval	Possible combinations
V_P (km s^{-1})	6.2	6.8	0.1	7
w_1	0.4	0.9	0.1	–
w_2	0.1	0.6	0.1	21
w_3	0	0.5	0.1	–
Stack type	Linear	PWS	–	2
Fmax (Hz)	0.4	2.0	0.1	17

associated measurement errors for each receiver function frequency value separately and determine how many individual solutions for that frequency fit within the aforementioned ranges of H and κ. The percentage of solutions within these limits for each Moho thickness/frequency combination is shown in Fig. 5. When the Moho thickness is $\leq$5 km, all frequencies identify the input model correctly. For Moho thicknesses $\geq$14 km, H–κ stacking fails across all frequencies. At intermediate Moho thicknesses (6–13 km), lower frequency receiver functions identify the Moho most consistently (Fig. 5). The general decrease in frequency required to accurately resolve a Moho of increasing thickness provides us with a useful, if imprecise, proxy for diagnosing Moho architecture.

We next examine two models with complex upper crustal structure, with a sharp (Fig. 6g) and a gradational (Fig. 6m) Moho, respectively. Both models fail at all frequencies, indicating that complex near-surface structure can preclude H–κ from working at all. Estimating Moho thickness by varying receiver function frequency content is thus only feasible when intracrustal structure is relatively simple.

Having demonstrated that full exploration of H–κ input parameter space, including the frequency content of the receiver function data set, is essential for robust crustal study, we now seek a semi-automated means of (i) determining whether or not H–κ analysis works for a given station, and (ii) gleaning a preferred H–κ solution, where appropriate.

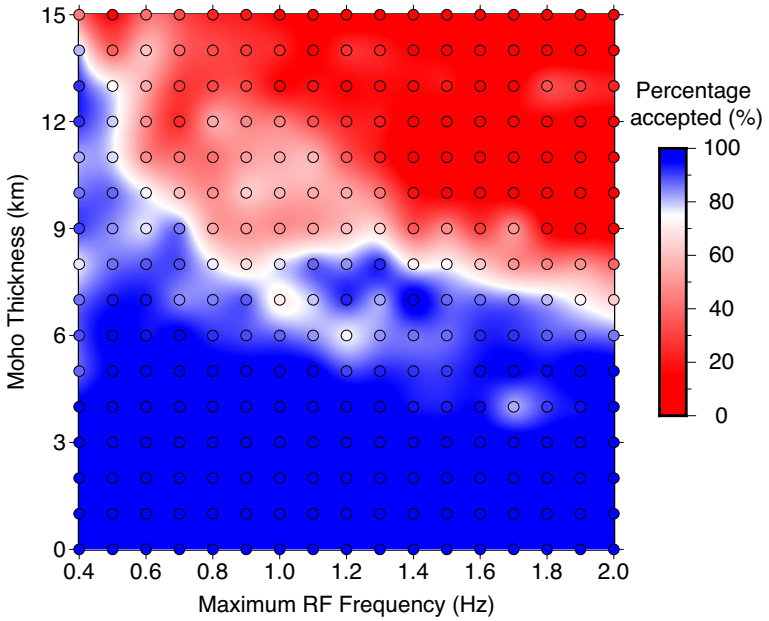

Figure 5. The effect of varying Moho thickness and receiver function frequency content on the accuracy of H–κ solutions. At Moho thickness $\leq$5 km, H–κ stacking is reliable and accurately constrains the input model for all frequency bands. Between 6 and 9 km, H–κ stacking becomes unreliable at high frequencies and results should be interpreted cautiously. H–κ stacking maintains reliability at very low frequencies until the Moho thickness is $\sim$13 km. However, when the Moho is $\geq$14 km thick, H–κ stacking becomes unreliable for all frequencies of receiver functions.

4 PARAMETER SEARCH APPROACH TO H–κ STACKING

Our new approach to H–κ stacking repeats the standard H–κ method 1000 times per station, randomly selecting input and data parameters each time. The challenge now is to obtain a preferred solution from this data set in a quantitative, semi-automated manner. To this end, we pursue a hierarchical cluster analysis (e.g. Everitt *et al.* 2001) approach. A methodological summary is documented by Everitt *et al.* (2001), and an analogous workflow to ours is presented by Teanby *et al.* (2004). Appendix A provides a more detailed description of the mathematical steps followed here.

Since H and κ have different orders of magnitude (20–55 km for H and 1.65–2.2 for κ), they are re-scaled between 0 and 1 using their respective ranges (eqs A1 and A2) to avoid the much larger Euclidean distances in H dominating the clustering algorithm. We begin with $N =$ 1000 scaled H–κ solutions, and split the data into M clusters, each containing one solution. The grid-search nature of H–κ stacking produces discrete data, which can cause hierarchical clustering to falter (e.g. Teanby *et al.* 2004). To counter this, each initial cluster is assigned a scaled numerical error in H and κ corresponding to the interval size of the grid search with a value of 1/99 used because 100 values of H and κ are used in the grid search. The intercluster Euclidean distances are calculated for all possible pairs of clusters and the closest two combined into one, reducing the number of clusters by one. The number of points per cluster is calculated, alongside the mean centroid position of each cluster centre according to eqs (A3) and (A4) for H and κ, respectively. Using the position of the newly merged cluster, Euclidean distances are re-calculated between all remaining clusters and the process is repeated until one cluster of 1000 data points remains. This produces the so-called hierarchical structure of clusters where the optimum number of clusters lies between 1 and 1000 (Everitt *et al.* 2001).

We desire a method to automatically choose the optimum number of clusters that best represents the 1000 individual H–κ measurements. There are several methods to achieve this (see Milligan & Cooper 1985, for a review) but the criteria of Caliński & Harabasz (1974) and Duda *et al.* (1973) are used here. Following Caliński & Harabasz (1974),

$$c(M) = \frac{(N - M)\text{trace}(B)}{(M - 1)\text{trace}(W)}, \tag{11}$$

where B is the between-cluster covariance (eq. A6) and W is the within-cluster covariance (eq. A7), both calculated at each cluster step. When $c(M)$ peaks, the between-cluster variance is maximized compared to the within-cluster variance indicating individual clusters are well spaced but that data points within each cluster are tightly distributed. The optimum number of clusters is, therefore, found when $c(M)$ is a maximum. The criterion of Duda *et al.* (1973) uses the ratio of within-cluster variances before (σ_2^2, eq. A10) and after (σ_1^2, eq. A11) the clusters are combined into a single cluster, assuming the two clusters will always be combined into one. This is rejected when

$$\left(1 - \frac{\sigma_2^2}{\sigma_1^2} - \frac{2}{\pi p}\right)\left(\frac{N_j p}{2[1 - 8/(\pi^2 p)]}\right)^{1/2} > c_{\text{critical}}, \tag{12}$$

where $p = 2$ and is the number of parameters in the analysis, and $c_{\text{critical}} = 3.20$ and is the typical value assumed if the data points within a cluster are normally distributed (Milligan & Cooper 1985). The optimum number of clusters is found when eq. (12) is invalidated as the

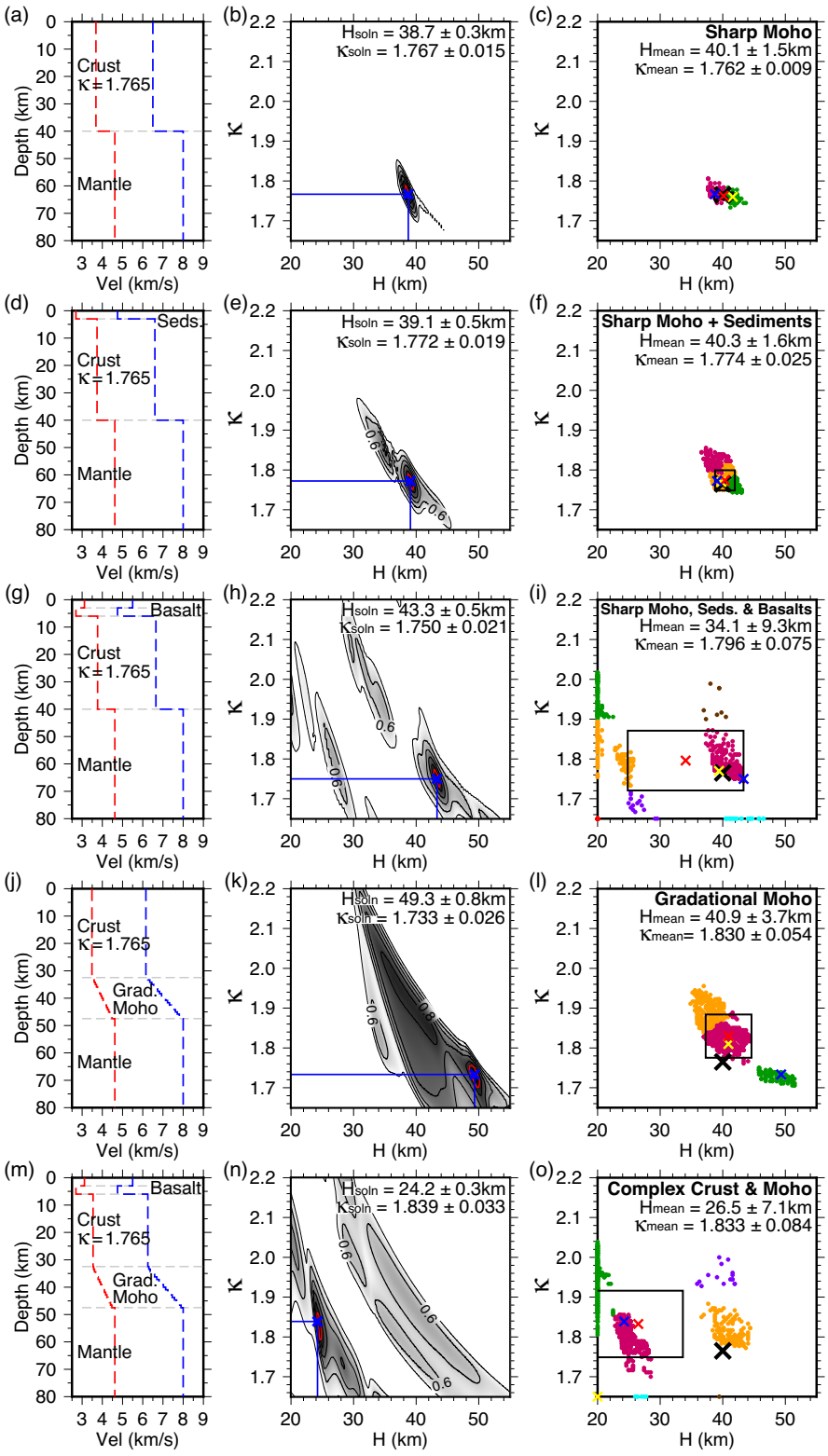

Figure 6. H–κ results for various synthetic models. Left-hand column: The blue-dashed line indicates V_P with depth, the red-dashed line indicates V_S with depth. Central column: Final H–κ stack solution for the model. Right-hand column: The distribution of H and κ solutions for the model. The dot colour indicates the cluster that a result belongs to in the hierarchical cluster analysis. The black cross is the expected H and κ from the input model, the red cross is the mean of H and κ from the 1000 repeats, the yellow cross is the mode combination of H and κ, the blue cross is the combination of H and κ selected by the cluster analysis. The black box marks one standard deviation in H and κ calculated for the 1000 repeat results. (a–c) Sharp Moho synthetic model. (d–f) Model with a sharp Moho and 3 km-thick, low velocity, near-surface sediments. (g–i) Model with a sharp Moho, and 3 km of high-velocity basalts overlying 3 km of low-velocity sediments. (j–l) Model with a 15 km gradational Moho centred at 40 km depth. (m–o) Model with a gradational Moho, and 3 km of high-velocity basalts overlying 3 km of low-velocity sediments.

Table 2. The 10 criteria used for determining result quality. RF data set: refers to the *N* receiver functions for the station.

Number	Criterion	Criteria assesses
1	Chosen H and κ solutions lie within the boundaries of the H–κ grid space.	Numerical H–κ solution
2	H and κ errors for the chosen result are $<\pm 2.5$ km in H and ± 0.042 in κ.	Numerical H–κ solution
3	Standard deviation in H after all 1000 repetitions is $<\pm 2.5$ km.	Cluster analysis
4	Standard deviation in κ after all 1000 repetitions is $<\pm 0.042$.	Cluster analysis
5	Mean average ACE of the 1000 repetitions is >3.	RF data set
6	Mode H and κ lie within the same cluster as the mean H and κ.	Cluster analysis
7	Summed amplitudes of *Ps*, *PpPs* and *PsPs* + *PpSs* for all stacked receiver functions, are positive, positive and negative, respectively.	RF data set
8	Mean average CCC of all receiver functions at each individual frequency is >0.6.	RF data set
9	Mean average SNR of the 1000 repetitions is >5.	RF data set
10	Overall mean H and κ for the linearly stacked repetitions lie within one standard deviation of H and κ for PWS repetitions, and vice versa.	Cluster analysis

number of clusters is reduced progressively. The criteria that produces the larger number of clusters is taken to be the optimum number of clusters, up to a maximum $M = 7$; larger numbers are considered indicative of poor clustering (Teanby *et al.* 2004).

With the optimum number of clusters calculated, the most representative cluster must be identified, from which the final individual H–κ solution will be chosen. The within-cluster variance ($\sigma_{c_j}^2$, eq. A13) and error variance ($\sigma_{d_j}^2$, eq. A14) are calculated for all clusters with $N_j >$ 15 (Teanby *et al.* 2004). The within-cluster variance measures cluster tightness; the error variance quantifies the measurement errors within each cluster. A diffuse cluster with small individual measurement errors will have large within-cluster variance but small error variance; a tight cluster with large measurement errors will have small within-cluster variance but large error variance (Teanby *et al.* 2004). Clusters containing <15 data points are rejected because they could have un-representatively small within-cluster variances (indicative of a tight cluster). We desire the cluster that optimizes these two variances. The overall variance ($\sigma_{o_j}^2$, eq. A15) finds the maximum of the within-cluster variance and error variance for each cluster with the best cluster having the minimum $\sigma_{o_j}^2$. From this cluster, we define the final H–κ solution to be the measurement with smallest combined rescaled errors in H and κ. If the 1000 H–κ stacking solutions do not form one cluster, H–κ stacking is not consistently identifying the Moho arrival and respective reverberations. Multiple clusters result from arrivals from intracrustal velocity discontinuities, particularly when the Moho arrivals are indistinguishable from the trace (e.g. a gradational Moho).

The result for the sharp Moho synthetic model (Figs 6a–c) has all 1000 repeat solutions clustering tightly around the true H and κ values. This is clear evidence that a sharp Moho is insensitive to changes in the H–κ input parameters, and is, therefore, the ideal example of H–κ stacking. When a 3 km-thick surface layer of low-velocity sediment is added, H–κ still retrieves the input model, albeit with a larger spread of results (Figs 6d–f). Adding a surface layer of higher velocity basalt above the sediment (Figs 6g–i) disperses the solutions, such that the mean H and κ values no longer match the input model. This implies complex upper crustal structure can cause H–κ stacking to fail. For the gradational model (Figs 6j–l), the results are dispersed into a much larger cluster and H–κ stacking fails to identify a single reliable answer. Furthermore, adding the high-velocity basalt layer overlying a low-velocity sediment layer (Figs 6m–o), the results become further dispersed into a number of clusters indicating a complete failure of H–κ stacking. The effect of 5 per cent crustal anisotropy with fast axis directed northwards (with 0° dip) is to again disperse the results (Supporting Information Figs S1a–c) but the input model is largely returned. A 15° dipping Moho is also successful in identifying the Moho (Supporting Information Figs S2a–c). We thus demonstrate that only by varying the input parameters of H–κ stacking can the reliability of the technique be truly tested. This is not possible to detect when only one set of parameters is chosen and the method performed using these alone.

We characterize and visualize the overall quality of a station using 10 pass/fail criteria (where *C* is the number of passed criteria; Table 2) and a diagnostic result figure (e.g. Fig. 7). Criteria 1 and 2 assess the numerical quality of the final H–κ solution, which are failed, respectively, if the solution lies on the edge of the H–κ grid space, and if errors exceed ± 2.5 km in H (defined by the resolvable Moho thickness in Fig. 5) and ± 0.042 in κ (limit defined by the sharp Moho synthetic test in Fig. 4). Criteria 3, 4 and 6 assess how well clustered the solutions of the 1000 H–κ repetitions are. Criteria 5, 7 and 9 analyse the receiver functions directly and assess whether or not the *Ps* phase and its reverberations are impulsive in nature. Criteria 8 assesses whether the receiver functions are coherent, or reflective of a noisy data set and/or strong backazimuthal variations in crustal structure. Criteria 10 compares linear and phase weighted stacking H–κ strategies.

A seismograph station that overlies a sharp Moho, with no near surface structure and little backazimuthal variation, passes ≥ 9 of the criteria, indicating a successful H–κ stacking result from which bulk-crustal properties can be reliably inferred. Stations passing <6 criteria have an unreliable H–κ result and their results should not be trusted. Stations with an intermediate number of passed criteria must be analysed carefully to ascertain result reliability prior to interpreting Moho depth and V_P/V_S ratio.

In the event that criteria are failed solely because the Moho below a given station is gradational, not sharp, we next investigate if the frequency content of the receiver functions can be limited to improve the H–κ solutions at stations that previously failed. We detect the frequency above which the H–κ solutions start to disperse. The maximum frequency is chosen to be the one above which the standard deviation of the solutions exceed 2.5 km in H or 0.042 in κ. The cluster analysis is then repeated using only the longer period solutions, removing the solutions calculated using higher frequency receiver functions (identified as causing H–κ stacking failure in Fig. 5). Limiting the analysis to longer period data may resolve a gradational Moho, but will not necessarily improve the solution when complex intracrustal structure exists (e.g. Figs 6g and m). In such instances, it is recommended that the analyst pursue more sophisticated analysis techniques

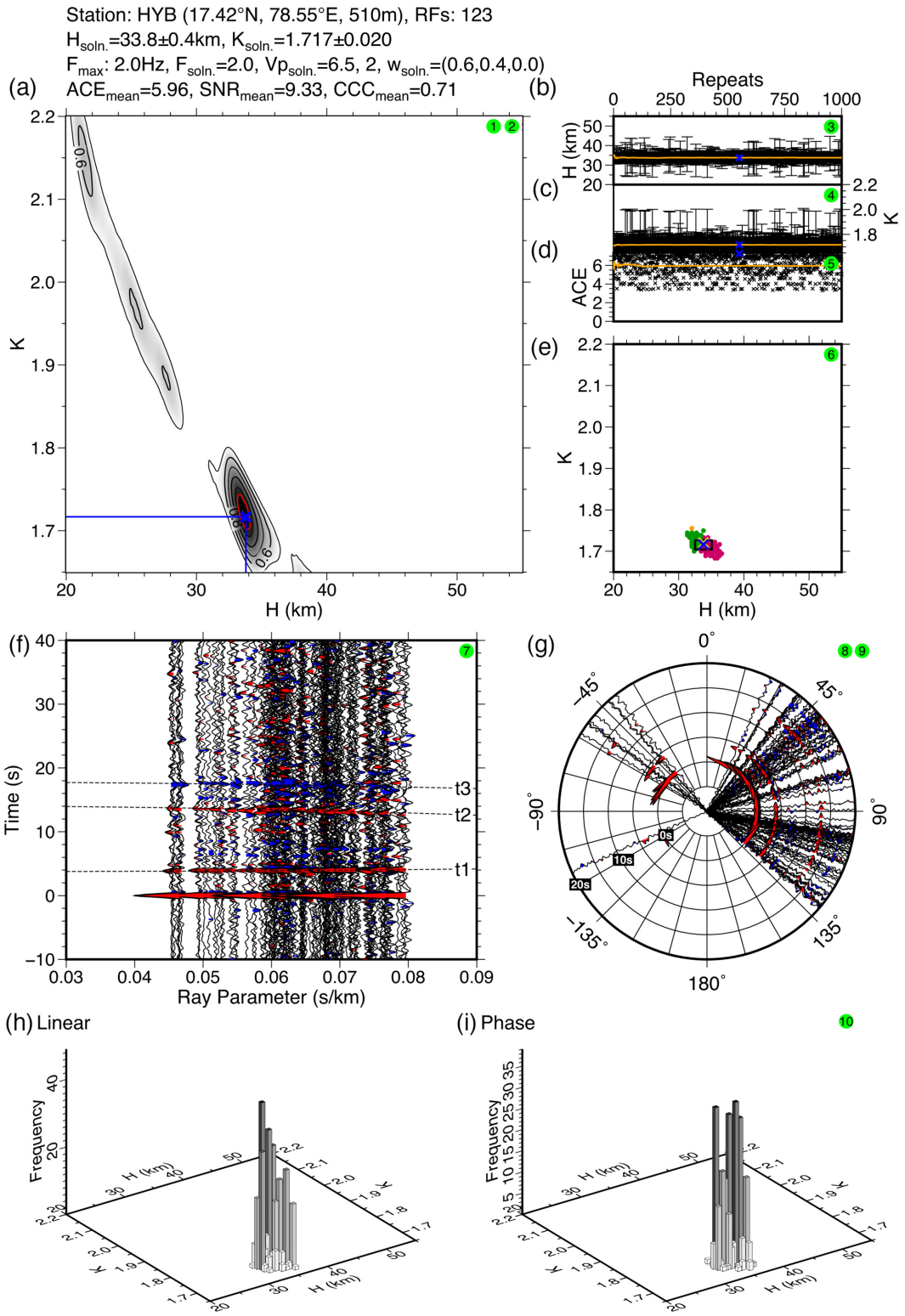

Figure 7. Final result diagram for station HYB on the East Dharwar craton. (a) H–κ stack of the selected repetition. The red ellipse outlines the 95 per cent amplitude contour, the blue cross is the final H and κ solution. (b) H result, (c) κ result and (d) ACE result for each repetition, with running means marked by the orange lines. (e) All 1000 H–κ solutions with colour representing the cluster that a result is assigned to. The red cross is the mean of H and κ from the 1000 repeats, the yellow cross is the mode combination of H and κ, the blue cross is the combination of H and κ selected by the cluster analysis, the black box marks one standard deviation in H and κ. (f) Accepted receiver functions plotted by horizontal slowness, t_1, t_2 and t_3 denote the predicted arrival times from the chosen H and κ solution. The peaks/troughs are coloured when their amplitude is >10 per cent of *P* arrival. (g) Receiver functions arranged by backazimuth. (h) Distribution of linear stacking results. (i) Distribution of phase-weighted stacking results.

such as Markov Chain Monte Carlo receiver function analysis (e.g. Piana Agostinetti & Malinverno 2010; Wirth *et al.* 2016) or 1-D joint inversions of surface waves and receiver functions (e.g. Julià *et al.* 2009; Gilligan *et al.* 2016). We next examine the applicability of our new H–κ strategy to a number of tectonic settings worldwide.

5 DATA PROCESSING

We extract three-component seismograms from the IRIS and ORFEUS data centres for all mb $\geq$ 5.5 teleseismic earthquakes listed in the NEIC earthquake catalogue in the epicentral distance range 30–90° from each individual station. Seismograms are Butterworth bandpass filtered with 0.04 and 3 Hz corner frequencies to reduce noise and visually inspected for pre-*P*-wave noise to determine if they are suitable for analysis. Receiver functions are again extracted using ETMTRF (Helffrich 2006). All receiver functions with pre-*P*-arrival amplitudes $\geq$25 per cent of the P-arrival amplitude at any time in the preceding 10 s are automatically removed, those remaining are visually inspected to ensure quality. Stations with fewer than eight acceptable receiver functions are not analysed because they are deemed unsuitable for stacking in our parameter search approach. A detailed list of stations used in this study and their respective results can be found in Supporting Information Table S1.

6 CASE STUDY 1—SIMPLE CRUSTAL STRUCTURE

To test our H–κ approach on regions of simple crustal structure, we analyse receiver functions from station HYB on the East Dharwar craton in India (e.g. Haggerty & Birkett 2004), station KMBL on the Yilgarn craton in western Australia (e.g. Swager *et al.* 1997) and on the POLARIS/HUBLE seismic networks (Eaton *et al.* 2005; Bastow *et al.* 2015, respectively) from northern Canada. We choose these stations because (i) they are installed on crystalline basement rocks, avoiding the effect of sedimentary layers; (ii) we have *a priori* constraints on Moho depth at these locations from wide-angle seismic refraction and/or joint inversion of surface waves with receiver functions; (iii) they have undergone no major tectonic activity since the Precambrian.

Station HYB, situated on the East Dharwar craton in India, is expected to display reliable results owing to a simple crustal structure imaged by both long period *P* waves (Singh & Rastogi 1978) and joint inversions of surface wave group velocities with receiver functions (e.g. Rai *et al.* 2003; Julià *et al.* 2009; Borah *et al.* 2014). HYB exhibits extremely tight clustering of results in H–κ space (Fig. 7e) and the crustal thickness of 33.8 km agrees with previous estimates (36 km from Singh & Rastogi 1978, 34 km from Rai *et al.* 2003, 35 km from Julià *et al.* 2009 and 34 km from Borah *et al.* 2014). Station KMBL, on the Yilgarn craton in western Australia, also returns well-clustered results and a crustal thickness value (37.3 km) is in agreement with the 36 km crustal thickness extracted from the nearby deep-crustal seismic reflection line of Swager *et al.* (1997) and the 36 km value derived by Collins *et al.* (2003).

Northern Canada is an ideal study locale to test H–κ stacking owing to its lack of surface sediments, removed by billions of years of erosion (e.g. St-Onge *et al.* 2006). We analyse several stations from the Canadian National Seismograph Network (Geological Survey of Canada 1989), POLARIS (Eaton *et al.* 2005) and HUBLE (Bastow *et al.* 2015) seismic networks using our broad parameter search approach (Fig. 8a). In 29 of the 37 stations deemed suitable for analysis, H–κ stacking is reliable at all frequencies, and final estimates for H and κ are recorded for these locations (Figs 8b and c). Station ILON produces a reliable H–κ result, with H = 36.6 km in close agreement with Thompson *et al.* (2010; they found H = 37.7 km for linear, H = 38.4 km for PWS), although our lower value of crustal thickness is simply due to a smaller V_P value selected by our analysis. Across northern Hudson Bay, bulk-crustal V_P/V_S is low, consistent with the region's felsic tonalite–trondhjemite–granodiorite Precambrian geology (Thompson *et al.* 2010).

Station SCHQ in Quebec (Fig. 8a) allows examination of the suitability of H–κ stacking to constrain crustal properties in a region lacking near-surface complexity (St-Onge *et al.* 2002) but where the Moho is known to be gradational (Gilligan *et al.* 2016). Using joint inversion of surface waves and receiver functions, Gilligan *et al.* (2016) observed an increase in shear-wave velocity from 3.8 to 4.5 km s^{-1} over a 20 km depth range (see Supporting Information Fig. S3). Our result for SCHQ exhibits a large data spread, very weak and incoherent Ps_{Moho} arrivals (Fig. 9f), and reverberations are completely incoherent despite little pre-*P*-arrival noise. Surface sediments are lacking, hence scattering in H–κ space must be produced by a gradational crust–mantle transition. SCHQ, therefore, provides an opportunity to investigate whether limiting the frequency content of the receiver functions can be used to identify the Moho. At every frequency band, the standard deviations in H and κ exceed the allowed limits, perhaps indicating the Moho is >13 km thick beneath SCHQ (Fig. 5). However, re-analysing the H–κ stacking solutions with frequencies $\leq$1.2 Hz (Supporting Information Fig. S4), we obtain an improved, but suboptimal solution. The crustal thickness of 46.5 km and κ of 1.750 are consistent with other stations in the Canadian shield and corroborate the 40–50 km Moho depth observed by Gilligan *et al.* (2016; Supporting Information Fig. S3). Station SCHQ, therefore, acts as a cautionary warning that H–κ cannot always be relied upon, even in cratonic areas and our parameter search approach is necessary to decipher this.

ACE, SNR and CCC values for the shields all generally pass the minimum cut-off limits for their respective criteria (Fig. 10), with the highest values of ACE (HYB: 5.96), SNR (ILON: 12.31) and CCC (SRLN: 0.80) all associated with Precambrian terranes. This supports the view that shields have generally simple, laterally homogeneous crustal structure, with a sharp Moho. There are exceptions, however: the high CCC value (0.72) at the Canadian station SCHQ suggests it is a low-noise station with relatively laterally homogeneous crust, but its low ACE (2.07) and SNR (3.71) values imply low-amplitude *Ps* Moho arrivals from a gradational Moho, consistent with the conclusions of Gilligan *et al.* (2016).

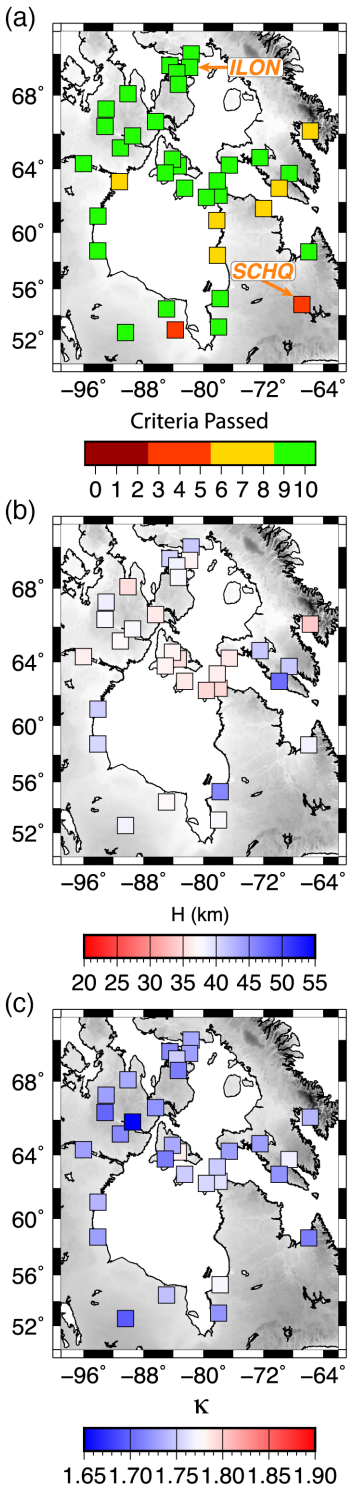

Figure 8. Results from stations in northern Canada. (a) Number of criteria passed. (b) Crustal thickness (H). (c) Bulk-crustal V_P/V_S (κ).

Overall, of the 55 stations that fail the ACE criteria (criteria 5), 51 pass ≤8 criteria overall. Similarly, 42 of 52 stations that pass the ACE criteria, also pass ≥9 criteria overall. ACE, therefore, identifies erroneous H–κ analysis 87 per cent of the time. The equivalent predictive success rates for SNR and CCC are 77 and 78 per cent, respectively, demonstrating the collective utility of our three receiver function analytics when carrying out H–κ analysis.

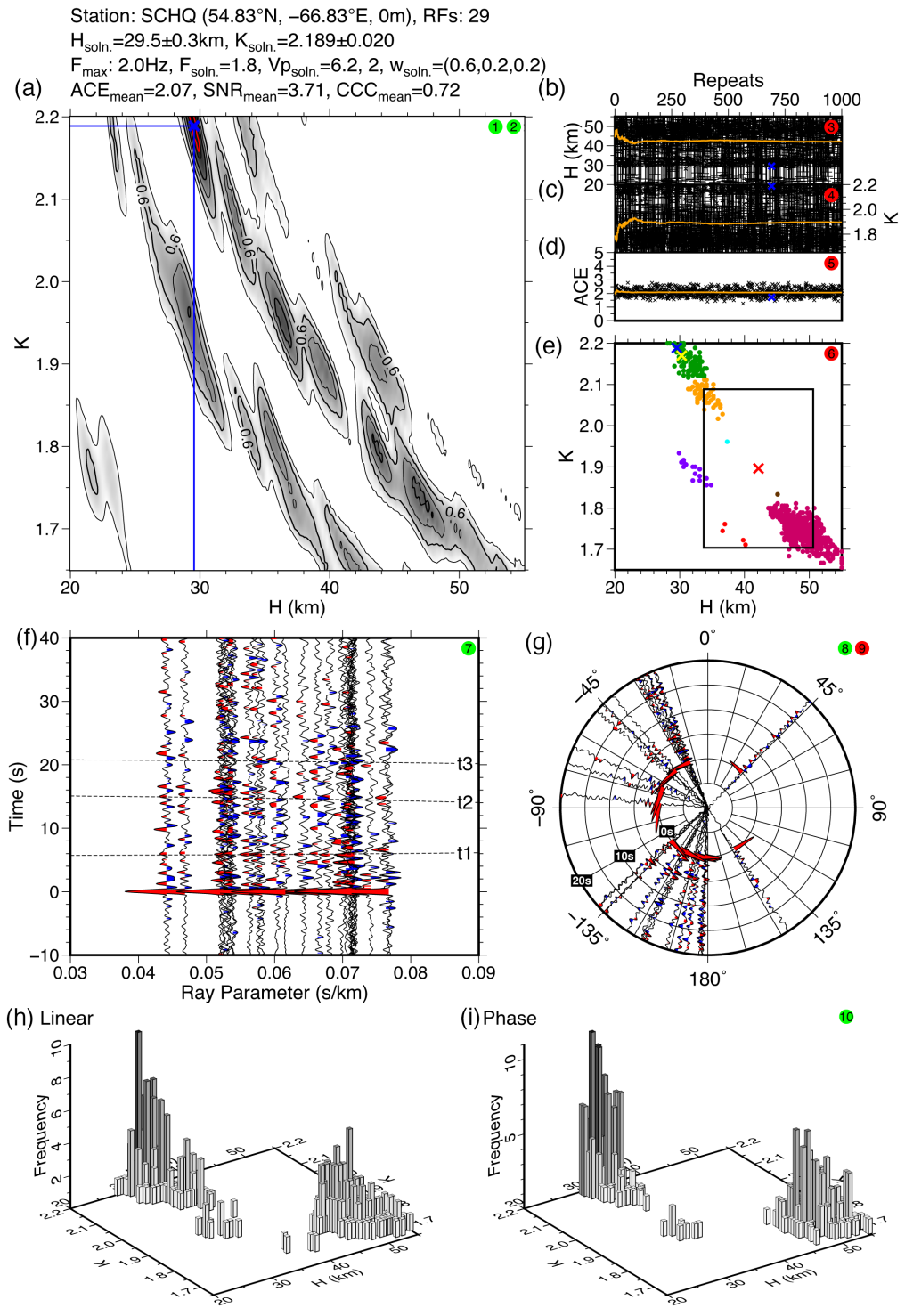

Figure 9. Final result diagram for station SCHQ in the Canadian shield (Fig. 8), which displays incoherent *P*-to-*S* conversions and an unreliable H–κ result. Figure details are as per Fig. 7.

7 CASE STUDY 2—THE ETHIOPIAN TRAPS: NEAR-SURFACE COMPLEXITY AND A GRADATIONAL MOHO

The Ethiopian Traps largely formed at 30 Ma, with 2–3 km flood-basalts erupting atop marine sediments during the development of the Red Sea rift (e.g. Hofmann *et al.* 1997; Rooney *et al.* 2012, 2018). Wide-angle seismic (Mackenzie *et al.* 2005) and gravity surveys (Cornwell *et al.* 2006) reveal a 8–12-km thick, lower crustal intrusion layer below the Ethiopian plateau. Thus, low-amplitude, diffuse Moho *P*-to-*S* conversions (Fig. 1b), coupled with arrivals/reverberations from the near surface are expected to render H–κ stacking unreliable. To explore this hypothesis, we analyse stations from several permanent and temporary networks in Ethiopia (e.g. Nyblade 2000; Bastow *et al.* 2011;

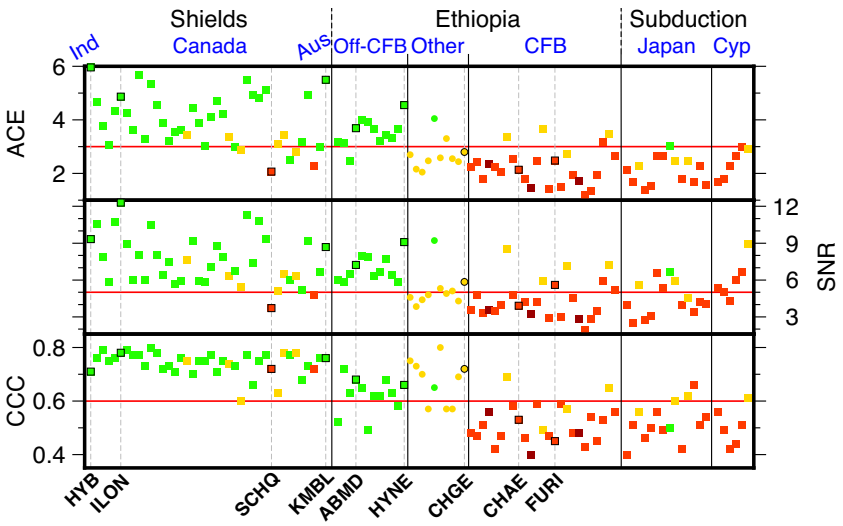

Figure 10. A comparison of the mean ACE, SNR and CCC analytics for stations used in the study. The squares are results from stations analysed at 2 Hz, the circles are stations analysed with limited frequencies. The colours indicate the number of criteria passed according to scale in Fig. 8(a). The red lines indicate the cut-off limit for each respective analytic. CFB: continental flood-basalt.

Keranen 2013; Ebinger *et al.* 2017). Fig. 11(a) demonstrates that most stations located directly on the flood-basalts produce unreliable results (e.g. CHAE, Fig. 12; FURI, Supporting Information Fig. S5). 11 of 12 off-flood-basalt stations yield reliable results, for example, stations ABMD (Fig. 13) and HYNE exhibit tight clustering similar to those from cratonic Canada. The slight increase in the spread of data compared to cratonic settings (Fig. 7e) is due to the sediments (Fig. 6f) on which these stations sit. Our result at HYNE ($H = 34.8$ km, $\kappa = 1.761$) matches closely the 35 km and $\kappa = 1.74$ obtained by Hammond *et al.* (2011). V_P/V_S ratios of 1.71–1.76 at stations where H–κ works well in Ethiopia are lower than the global average of 1.765. This implies that the crust beneath these stations is predominantly Precambrian in age, and lacking in modification by hotspot-related mafic magmatism in the form of dyke intrusions and/or lower crustal intrusions. ACE and SNR values at these off-flood-basalt stations (Fig. 10) are generally high, consistent with the conclusion that the *Ps* arrival at these stations is high amplitude. The mean CCC values are lower than for shields, indicative of a more variable backazimuthal structure.

Towards the western edge of the flood-basalt province and/or where the flood-basalts have been eroded by Blue Nile river incision, stations pass an intermediate number of criteria (e.g. CHGE; $C = 6$, Supporting Information Fig. S6). Frequency-limited analysis improves the results (Supporting Information Table S2) significantly (e.g. CHGE; $C =$, Fig. 14) and reduces the spread of the remaining H–κ solutions. The Moho is ∼4 km deeper for these stations than below the adjacent off-flood-basalt stations. Noting that H–κ analysis selects the centre of the Moho velocity gradient, this can be interpreted as a crust that is ∼8 km thicker. Corroborating this, the 0.6–1.4 Hz frequency limit for these stations implies a gradational Moho of thickness 6–10 km (Figs 5 and 11b), in close agreement with the 8–12 km-thick fast *P*-wave speed (∼7.38 km s^{-1}) layer found by Mackenzie *et al.* (2005). κ also increases for these stations (Fig. 11c), indicative of a more mafic bulk-crustal composition, which when interpreted in light of the thicker crust suggests lower crustal intrusions exist below the western extent of the flood-basalt province. Mean ACE and SNR values for these stations (Fig. 10) are generally low (e.g. GIDA; ACE = 2.05, SNR = 4.39), suggesting that although the frequency-limited analysis has improved the solutions, these stations have genuinely low-*Ps* amplitudes indicative of a gradational Moho.

Stations in the centre of the flood-basalts fail across all frequencies, indicating the Moho is >13 km thick and/or that arrivals from the basalt-sediment contact are causing H–κ stacking to fail. The ACE, SNR and CCC values from stations on the flood-basalts (Fig. 10) consistently fail their respective criteria, supporting the hypothesis that the *Ps* arrival is low amplitude and/or not always discernible from arrivals produced by intracrustal structure. Our observations may thus imply that the lower crustal intrusion layer and flood-basalts are thickest in the centre of the Ethiopian traps, near the major Palaeogene eruptive centres and thinner (≤5 km) or non-existent at off-flood-basalt stations, consistent with petrological studies of the Ethiopian traps (Rooney *et al.* 2016; Rooney 2017; Rooney *et al.* 2018).

8 CASE STUDY 3—SUBDUCTION ZONES

Another tectonic regime on which to test our new H–κ stacking approach is subduction zones, where crustal complexity is expected owing to the presence of two tectonic plates (and, therefore, two Moho discontinuities). Magmatism and a thick mantle wedge are also expected to add complexity (e.g. Bostock *et al.* 2002).

Japan is an archetypal example of a subduction zone, where the Pacific plate is subducting below Eurasia. In addition to the expected presence of multiple velocity discontinuities associated with two tectonic plates, a thick mantle wedge and voluminous crustal magmatism (e.g. Nakajima *et al.* 2005) are expected to render H–κ stacking challenging on the island. To test this hypothesis, we analyse seismograms

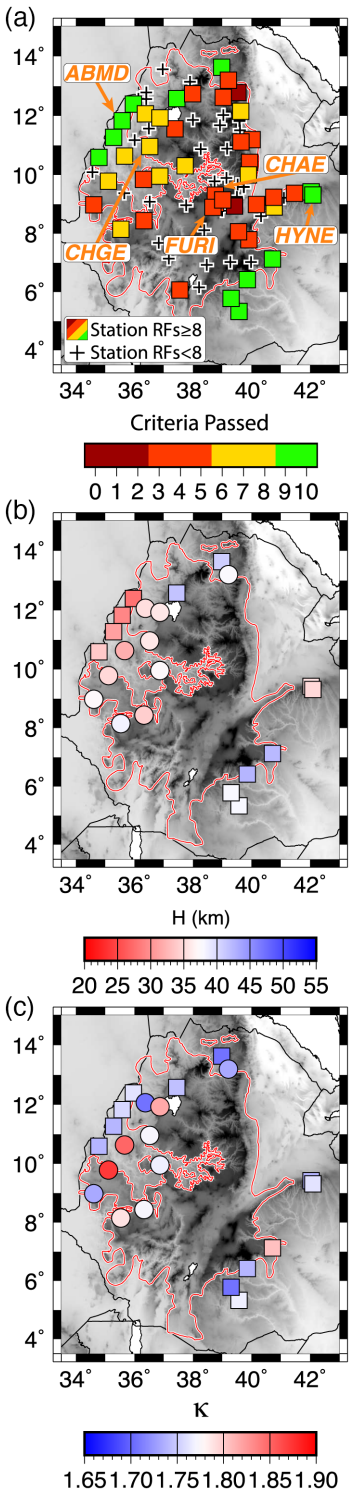

Figure 11. Results from stations in Ethiopia. (a) Number of criteria passed. The black crosses indicate analysed stations where fewer than eight suitable receiver functions were calculated. The red line delineates the spacial extent of the 30 Ma Ethiopian flood-basalt province (after Rooney 2017). (b) Crustal thickness (H). The squares are results from stations that passed ≥8 criteria with all frequencies, the circles are solutions for stations from which the frequency-dependent analysis produced a reliable result. (c) Bulk-crustal V_P/V_S (κ).

from a selection of Japan Meteorological Agency Seismic Network (Tatehata 1997) and Global Seismograph Network (ASL/USGS 1988) stations. Predictably, our H–κ approach generally obtains unreliable results (Fig. 15b) across the island.

The Cyprus arc in the eastern Mediterranean has developed due to subduction of the African Plate beneath the Anatolian Plate (e.g. Robertson & Mountrakis 2006). Unlike Japan, the ∼8 mm yr^{-1} convergence rate (Gripp & Gordon 2002) is slower and arc magmatism is lacking, perhaps leading to the expectation of a simpler crustal structure. However, a velocity discontinuity between the high-velocity Troodos

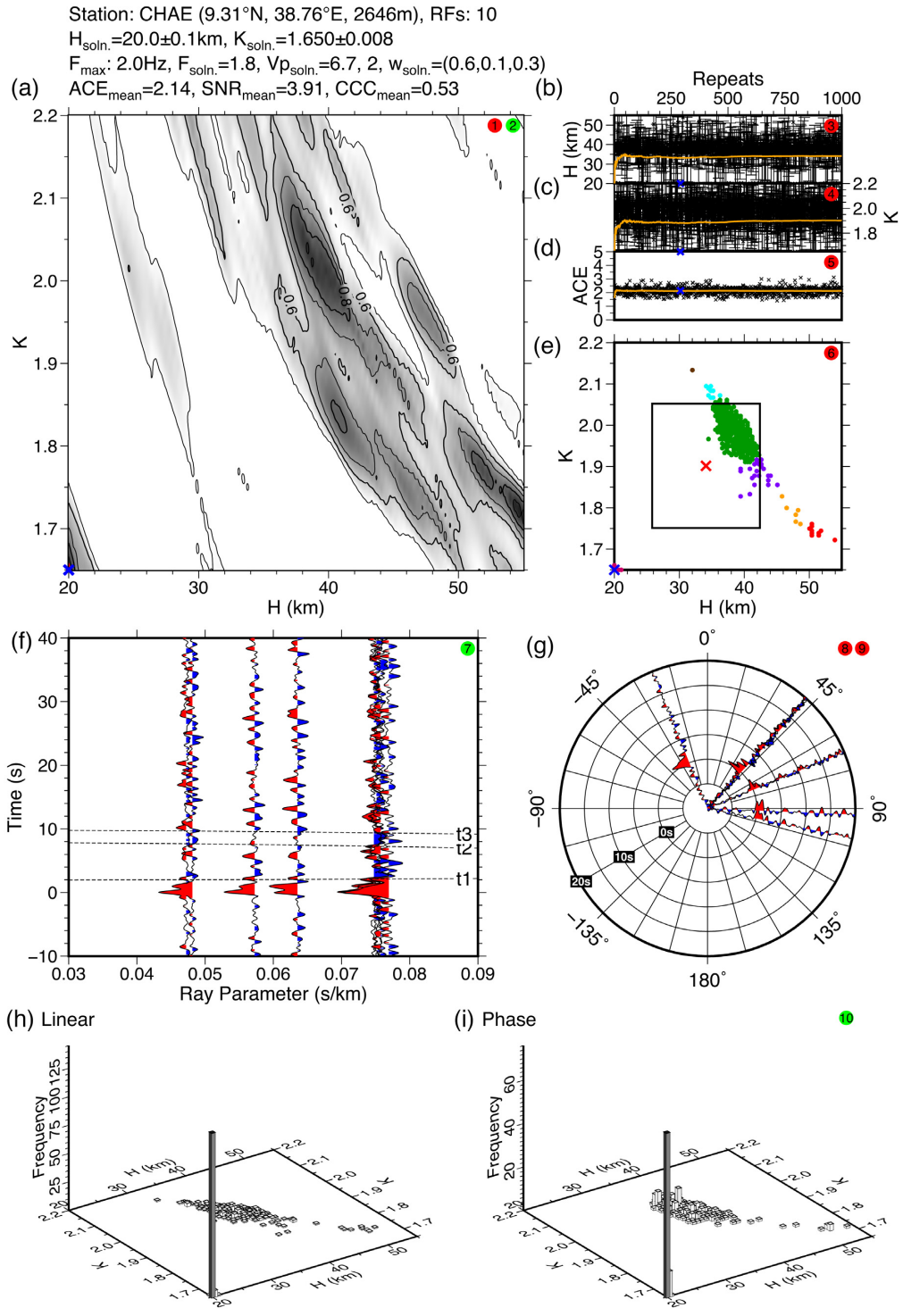

Figure 12. Final result diagram for station CHAE on the Ethiopian Plateau (Fig. 11). Figure details are as per Fig. 7.

Ophiolite and slower wave speed underlying Anatolian continental crust (Mackenzie *et al.* 2006; Feld *et al.* 2017) may introduce receiver function complexity. Terrestrial sediments of thickness ∼3 km (Harrison *et al.* 2008) surrounding the ophiolite may also generate additional *P*-to-*S* conversions (Fig. 1c). To explore these issues, we analysed data from the Cyprus Broadband Seismological Network (Cyprus Geol. Survey Dept. 2013), the TROODOS temporary broadband deployment (Bastow *et al.* 2017) and the Kandilli Observatory Broadband Network (BUKO 2001). Of 18 stations on the island, only six have eight or more acceptable-quality receiver functions (Fig. 15a) and all of these lie on ophiolitic material. Nowhere on Cyprus is H–κ stacking deemed reliable.

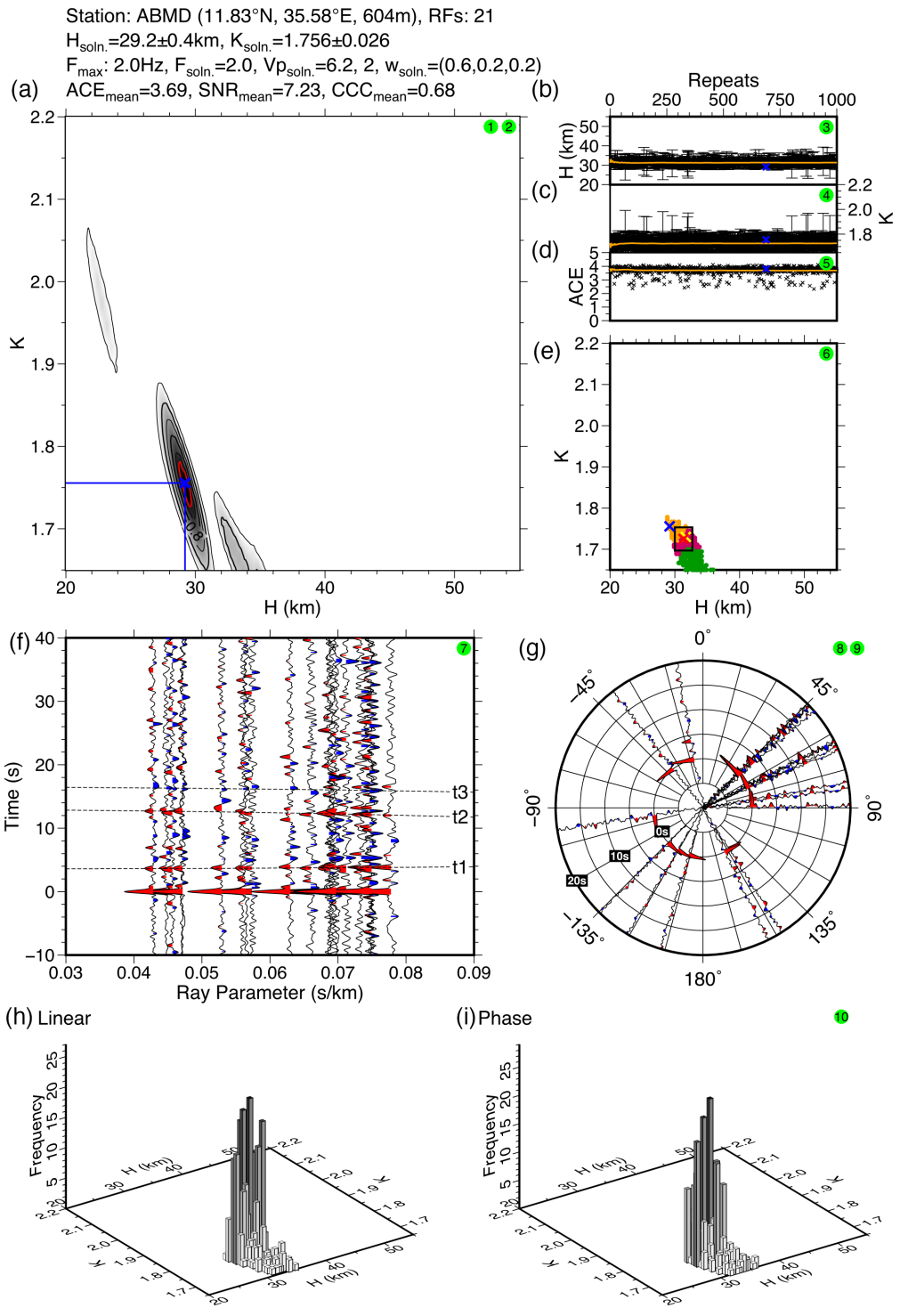

Figure 13. Final result diagram for station ABMD, just off the westernmost extent of the flood-basalts (Fig. 11). Figure details are as per Fig. 7.

Our H–κ stacking analyses in Japan and Cyprus are by no means a complete analysis of the global subduction zone system. However, we contend that the failure of H–κ stacking in both regions suggests all H–κ stacking results in subduction zone settings should be treated with extreme caution.

9 CONCLUSIONS

We have demonstrated via analysis of synthetic seismograms that key to resolving where the H–κ stacking method succeeds and fails is a rigorous search of the H–κ stacking parameter space (including the relative weights assigned to the Moho *P*-to-*S* conversion and its subsequent

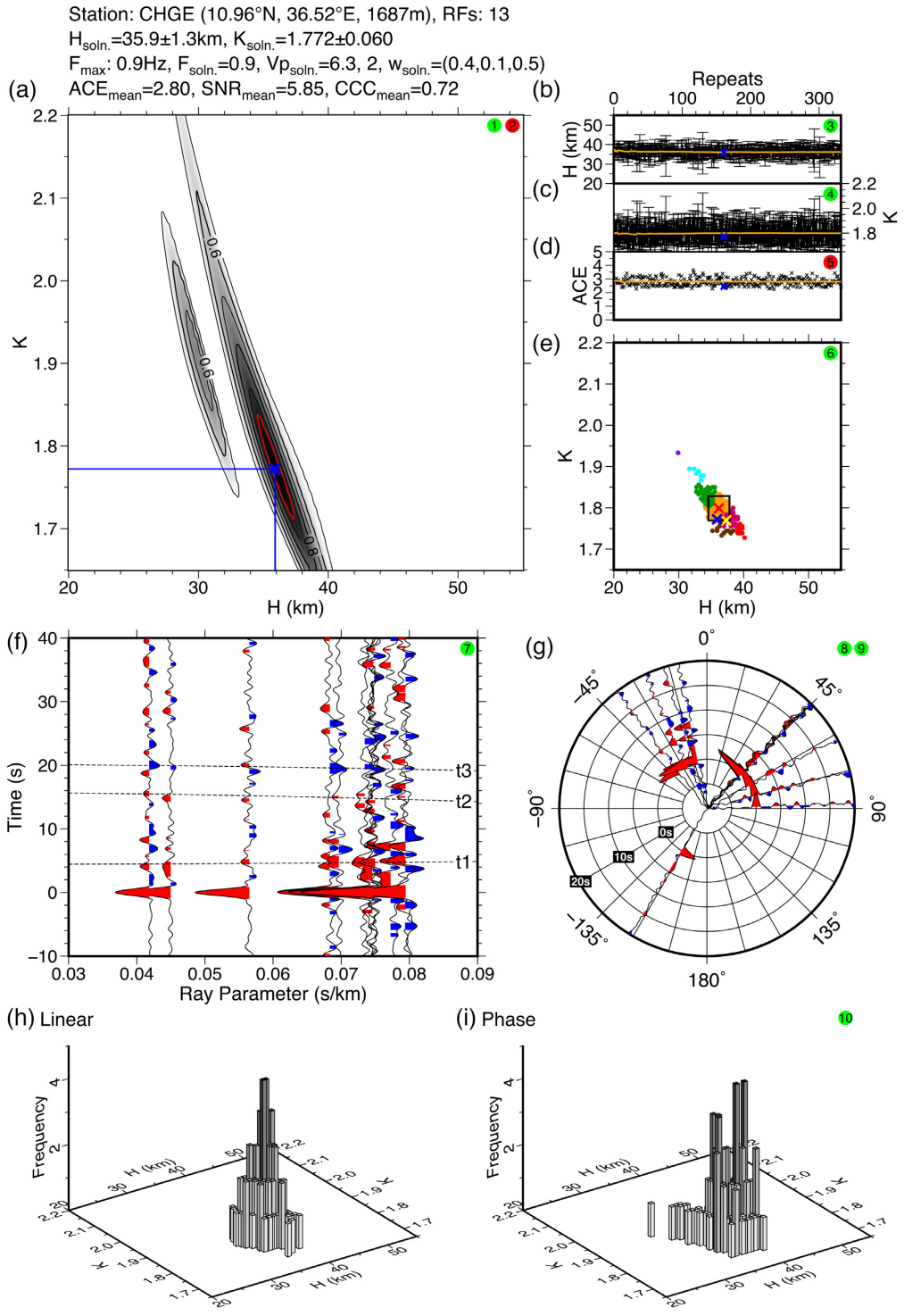

Figure 14. Final result diagram for station CHGE on the Ethiopian Plateau (Fig. 11), where only solutions using receiver functions with frequencies ≤0.9 Hz are sampled. Figure details are as per Fig. 7.

reverberations, the choice of linear or PWS and *P*-wave velocity). Data parameters including the receiver function frequency content and the subset of receiver functions selected for analysis must also be explored thoroughly.

To address these issues, we have developed an H–κ stacking approach in which cluster analysis selects a final solution from 1000 repeat results, each calculated using randomly selected input and data parameters. We define 10 criteria that variously assess the final numerical result, the receiver function data set and the extent to which the results are tightly clustered. If a station passes ≥9 criteria, H–κ stacking is reliable and crustal structure can be considered simple. If a station passes ≤5 criteria, its H–κ results cannot be interpreted reliably and more sophisticated seismological techniques (e.g. Julià *et al.* 2009; Wirth *et al.* 2016) are required to characterize crustal architecture. Synthetic

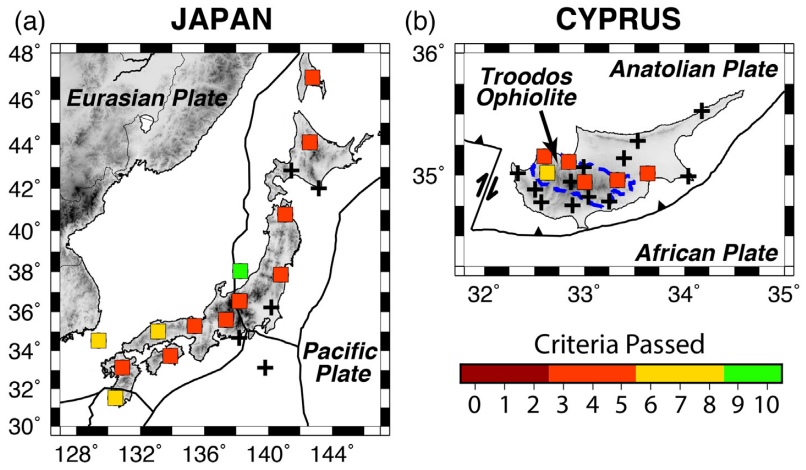

Figure 15. Results from stations in (a) Japan and (b) Cyprus with stations plotted by number of criteria passed. The black crosses indicate analysed stations where fewer than eight suitable receiver functions were calculated. The blue dashed line outlines the surface extent of the Troodos Ophiolite.

testing of our new approach shows that when the Moho is sharp, H–κ solutions cluster tightly at all frequencies and return the input model; in areas of more complex crustal structure, H–κ stacking yields erratic results and cannot be trusted. Limiting the frequency content of the receiver functions can allow an estimation of the thickness of a gradational Moho, provided that complex intracrustal structure is lacking.

Applying our H–κ cluster analysis method to the East Dharwar craton, Yilgarn craton and Canadian shield demonstrate the suitability of the H–κ method in regions where the crust is simple and the Moho sharp. Our three new receiver function analytics (ACE, SNR and CCC), have generally high values in these shields, supportive of a sharp Moho at the base of simple crust, and little backazimuthal variation in crustal structure. In contrast, on the younger recently volcanically active Ethiopian plateau where 2 km-thick flood-basalts overlie marine sediments, and the Moho is known *a priori* to be a gradational feature due to an 8–12-km thick layer of lower crustal mafic intrusions, H–κ stacking is particularly unreliable. By limiting the frequency content of receiver functions to longer periods, at stations where the flood-basalts are thinner and/or have been eroded by Blue Nile incision, the quality of H–κ solutions improves. These stations have elevated κ (1.77–1.87) values, with ACE and SNR values lower than on the shields, evidence that these stations overlie a gradational Moho of ∼6–10 km thickness. Moving just a few kilometers off the western extent of the flood-basalt province, solutions cluster tightly at all frequencies and crustal thicknesses are ∼4 km thinner than the adjacent flood-basalt stations. Bulk-crustal V_P/V_S ratios are low (∼1.73) at these stations compared to the global average of 1.765. Unreliable H–κ results at the Cyprus and Japan subduction zones are an inevitable consequence of their complex Moho and crustal architectures. H–κ stacking is, therefore, a valuable tool to infer crustal thickness and V_P/V_S ratio in locations where the crust is relatively simple. However, the technique should be used with extreme caution where crustal structure is complex.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers and editor Michael Ritzwoller for insightful comments, which have improved this manuscript. We also thank H. Meek for hard work during the early stages of this project and S. Pilidou, I. Dimitriadis, P. Iosif and their colleagues at the Geological Survey Department of Cyprus for their help establishing the TROODOS network (Bastow *et al.* 2017). V. Lane and D. Daly (both of SEIS-UK), A. Boyce, M. Liddell and R. Kounoudis were all excellent field assistants in Cyprus. SAC (Helffrich *et al.* 2013) and GMT (Wessel & Smith 1991) software were used to process and image seismic data, which were sourced from IRIS DMC and ORFEUS. CSO is funded by the Natural Environment Research Council (NERC) Doctoral Training Partnership: Science and Solutions for a Changing Planet, Grant NE/L002515/1. SR's contribution to this work was supported by Career Integration Grant 321871—GLImER from the FP7 Marie Curie Actions of the European Commission, and by the Research Council of Norway FRINATEK programme through SwaMMIS project 231354. This work was supported by the Royal Society International Exchanges Grant IE160688.

REFERENCES

ASL/USGS, 1988. Global Seismograph Network (GSN—IRIS/USGS), International Federation of Digital Seismograph Networks, Other/Seismic Network, doi:10.7914/SN/IU.

Bastow, I.D., Keir, D. & Daly, E., 2011, The Ethiopia Afar Geoscientific Lithospheric Experiment (EAGLE): probing the transition from continental rifting to incipient seafloor spreading, *Geol. Soc. Am. Spec. Pap.,* **478,** 51–76.

Bastow, I.D. *et al.*, 2015. The Hudson Bay Lithospheric Experiment (HuBLE): insights into Precambrian plate tectonics and the development of mantle keels, *Geol. Soc. Lond. Spec. Publ.,* **389**(1), 41–67.

Bastow, I.D., Ogden, C., Pilidou, S., Dimitriadis, I., Constantinou, C. & Iosif, P., 2017. TROODOS: tomography and receiver function observations of an Ophiolite using data obtained from seismology. International federation of digital seismograph networks, Other/Seismic Network, doi:10.7914/SN/XQ_2017.

Borah, K., Rai, S., Prakasam, K., Gupta, S., Priestley, K. & Gaur, V., 2014. Seismic imaging of crust beneath the Dharwar Craton, India, from ambient noise and teleseismic receiver function modelling, *Geophys. J. Int.,* **197,** 2, 748–767.

Bostock, M., Hyndman, R., Rondenay, S. & Peacock, S., 2002. An inverted continential Moho and serpentinization of the forearc mantle, *Nature,* **417,** 536–538.

BUKO, 2001. Bogazici University Kandilli Observatory And Earthquake Research Institute, International Federation of Digital Seismograph Networks, Other/Seismic Network, doi:10.7914/SN/KO.

Caliński, T. & Harabasz, J., 1974. A dendrite method for cluster analysis, *Commun. Stat. - Theory Methods,* **3**(1), 1–27.

Collins, C., Drummond, B. & Nicoll, M., 2003. Crustal thickness patterns in the Australian continent, *Geol. Soc. Am. Spec. Pap.,* **372,** 121–128.

Cornwell, D., Mackenzie, G., England, R., Maguire, P., Asfaw, L. & Oluma, B., 2006. Northern main Ethiopian rift crustal structure from new high-precision gravity data, *Geol. Soc. Lond. Spec. Publ.,* **259**(1), 307–321.

Crotwell, H. & Owens, T., 2005. Automated receiver function processing, *Seismol. Res. Lett.,* **76,** 702–708.

Cyprus Geol. Survey Dept., 2013. Cyprus Broadband Seismological Network, International Federation of Digital Seismograph Networks, Other/Seismic Network, doi:10.7914/SN/CQ.

Duda, R., Hart, P. & Stork, D., 1973. *Pattern Classification,* Wiley.

Dugda, M., Nyblade, A., Julià, J., Langston, C., Ammon, C. & Simiyu, S., 2005. Crustal structure in Ethiopia and Kenya from receiver function analysis, *J. geophys. Res.,* **110**(B1).

Eaton, D. *et al.*, 2005. Investigating Canada's lithosphere and earthquake hazards with portable arrays, *EOS, Trans. Am. geophys. Un.,* **86,** 169–173.

Eaton, D., Savka, D. & Robert, M., 2006. Crustal thickness and Vp/Vs variations in the Grenville orogen (Ontario, Canada) from analysis of teleseismic receiver functions, *Tectonophysics,* **420,** 223–238.

Ebinger, C., Keir, D., Bastow, I., Whaler, K., Hammond, J., Ayele, A., Miller, M. & Tiberi, C., 2017. Crustal structure of active deformation zones in Africa: implications for global crustal processes, *Tectonics,* **36**(12), 3298–3332.

Everitt, B., Landau, S. & Leese, M., 2001. *Cluster Analysis,* 4th edn, Arnold.

Feld, C. *et al.*, 2017. Crustal structure of the Eratosthenes Seamount, Cyprus and S. Turkey from an amphibian wide-angle seismic profile, *Tectonophysics,* **700–701,** 32–59.

Frassetto, A., Zandt, G., Gilbert, H., Owens, T. & Jones, C., 2011. Structure of the Sierra Nevada from receiver functions and implications for lithospheric foundering, *Geosphere,* **7**(4), 898–921.

Frederiksen, A. & Bostock, M., 2000. Modelling teleseismic waves in dipping anisotropic structures, *Geophys. J. Int.,* **141**(2), 401–412.

Gallacher, R. & Bastow, I., 2012. The development of magmatism along the Cameroon volcanic line: evidence from teleseismic receiver functions, *Tectonics,* **31,** doi:10.1029/2011TC003028.

Geological Survey of Canada, 1989. Canadian National Seismograph Network, International Federation of Digital Seismograph Networks, Other/Seismic Network, doi:10.7914/SN/CN.

Gilligan, A., Bastow, I. & Darbyshire, F., 2016. Seismological structure of the 1.8 Ga Trans-Hudson Orogen of North America, *Geochem. Geophys. Geosyst.,* **17**(6), 2421–2433.

Gripp, A. & Gordon, R., 2002. Young tracks of hotspots and current plate velocities, *Geophys. J. Int.,* **150**(2), 321–361.

Haggerty, S. & Birkett, T., 2004. Geological setting and chemistry of kimberlite clan rocks in the Dharwar craton, India, *Lithos,* **76**(1–4), 535–549.

Hammond, J., Kendall, J.-M., Stuart, G., Keir, D., Ebinger, C., Ayele, A. & Belachew, M., 2011. The nature of the crust beneath the Afar triple junction: evidence from receiver functions, *Geochem. Geophys. Geosyst.,* **12**(12).

Harrison, R. *et al.*, 2008. Bedrock geologic map of the greater Lefkosia area, Cyprus, U.S. Geological Survey Scientific Investigations Map 3046, 1 map, scale 1:25,000, 1–36, doi:10.3133/sim3046.

Helffrich, G., 2006. Extended-time multitaper frequency domain cross-correlation receiver-function estimation, *Bull. seism. Soc. Am.,* **96**(1), 344–347.

Helffrich, G., Wookey, J. & I.D.B., 2013. *The Seismic Analysis Code: A Primer and User's Guide,* 1st edn, Cambridge Univ. Press.

Hofmann, C., Courtillot, V., Feraud, G., Rochette, P., Yirgu, G., Ketefo, E. & Pik, R., 1997. Timing of the Ethiopian flood basalt event and implications for plume birth and global change, *Nature,* **389,** 838–841.

Julià, J., Jagadeesh, S., Rai, S. & Owens, T., 2009. Deep crustal structure of the Indian shield from joint inversion of P wave receiver functions and Rayleigh wave group velocities: implications for Precambrian crustal evolution, *J. geophys. Res.,* **114**(B10).

Keranen, K., 2013. Exploring extensional tectonics beyond the Ethiopian Rift, International Federation of Digital Seismograph Networks, Other/Seismic Network, doi:10.7914/SN/YY 2013.

Langston, C., 1979. Structure under Mount Rainer, Washington, inferred from teleseismic body waves, *J. geophys. Res.,* **84,** 4749–4762.

Levin, V. & Park, J., 2000. Shear zones in the Proterozoic lithosphere of the Arabian Shield and the nature of the Hales discontinuity, *Tectonophysics,* **323**(3–4), 131–148.

Ligorrìa, J. & Ammon, C., 1999. Iterative deconvolution and receiver-function estimation, *Bull. seism. Soc. Am.,* **89**(5), 1395–1400.

Mackenzie, G., Thybo, H. & Maguire, P., 2005. Crustal velocity structure across the main Ethiopian rift: results from 2-dimensional wide-angle seismic modelling, *Geophys. J. Int.,* **162,** 994–1006.

Mackenzie, G., Maguire, P., Coogan, L., Khan, M., Eaton, M. & Petrides, G., 2006. Geophysical constraints on the crustal architecture of the Troodos ophiolite: results from the IANGASS project, *Geophys. J. Int.,* **167**(3), 1385–1401.

Milligan, G. & Cooper, M., 1985. An examination of procedures for determining the number of clusters in a data set, *Psychometrika,* **50**(2), 159–179.

Nakajima, J., Takei, Y. & Hasegawaa, A., 2005. Quantitative analysis of the inclined low-velocity zone in the mantle wedge of northeastern Japan: a systematic change of melt-filled pore shapes with depth and its implications for melt migration, *Earth planet. Sci. Lett.,* **234,** 59–70.

Nyblade, A., 2000. Seismic investigation of deep structure beneath the Ethiopian plateau and afar depression, international federation of digital seismograph networks, Other/Seismic Network, doi:10.7914/SN/XI_2000.

Park, J. & Levin, V., 2000. Receiver functions from multiple-taper spectral correlation estimates, *Bull. seism. Soc. Am.,* **90**(6), 1507–1520.

Piana Agostinetti, N. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181**(2), 858–872.

Rai, S., Priestley, K., Suryaprakasam, K., Srinagesh, D., Gaur, V. & Du, Z., 2003. Crustal shear velocity structure of the south Indian shield, *J. geophys. Res.,* **108**(B2).

Randall, G., 1989. Efficient calculation of differential seismograms for lithospheric receiver functions, *Geophys. J. Int.,* **99**(3), 469–481.

Robertson, A. & Mountrakis, D., 2006. Tectonic development of the eastern mediterranean region: an introduction, *Geol. Soc. Lond. Spec. Publ.,* **260,** 1–9.

Rondenay, S., Spieker, K., Sawade, L., Halpaap, F. & Farestveit, M., 2016. Glimer: a new global database of teleseismic receiver functions for imaging earth structure, *Seismol. Res. Lett.,* **88**(1), 39–48.

Rooney, T., 2017. The Cenozoic magmatism of East-Africa: part I—flood basalts and pulsed magmatism, *Lithos,* **286,** 264–301.

Rooney, T., Hanan, B., Graham, D., Furman, T., Blichert-Toft, J. & Schilling, J., 2012. Upper mantle pollution during Afar Plume–Continental rift interaction, *J. Petrol.,* **53**(2), 365–389.

Rooney, T., Lavigne, A., Svoboda, C., Girard, G., Yirgu, G., Ayalew, D. & Kappelman, J., 2016. The making of an underplate: pyroxenites from the Ethiopian lithosphere, *Chem. Geol.,* **455,** 264–281.

Rooney, T., Krans, S., Mège, D., Arnaud, N., Korme, T., Kappelman, J. & Yirgu, G., 2018. Constraining the magmatic plumbing system in a zoned continental flood basalt province, *Geochem. Geophys. Geosyst.,* **19**(10), 3917–3944.

Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signal through phase-weighted stacks, *Geophys. J. Int.,* **130,** 497–505.

Singh, D. & Rastogi, B., 1978. Crustal structure of the peninsular shield beneath Hyderabad (India) from the spectral characteristics of long-period P-waves, *Tectonophysics,* **51**(3-4), 127–137.

St-Onge, M., Scott, D. & Wodicka, N., 2002. Review of crustal architecture and evolution in the Ungava Peninsula-Baffin Island area: connection to the Lithoprobe ECSOOT transect, *Can. J. Earth Sci.,* **39**(5), 589–610.

St-Onge, M., Searle, M. & Wodicka, N., 2006. Trans-Hudson Orogen of North America and Himalaya-Karakoram-Tibetan Orogen of Asia: structural and thermal characteristics of the lower and upper plates, *Tectonics,* **25**(4).

Swager, C., Goleby, B., Drummond, B., Rattenbury, M. & Williams, P., 1997. Crustal structure of granite-greenstone terranes in the Eastern Goldfields, Yilgarn Craton, as revealed by seismic reflection profiling, *Precambrian Res.,* **83**(1–3), 43–56.

Tatehata, H., 1997. *The New Tsunami Warning System of the Japan Meteorological Agency,* pp. 175–188, Springer.

Teanby, N., Kendall, J.-M. & Van der Baan, M., 2004. Automation of shear-wave splitting measurements using cluster analysis, *Bull. seism. Soc. Am.,* **94**(2), 453–463.

Thompson, D., Bastow, I., Helffrich, G., Kendall, J.-M., Wookey, J., Snyder, D. & Eaton, D., 2010. Precambrian crustal evolution: seismic constraints from the Canadian shield, *Earth planet. Sci. Lett.,* **297,** 655–666.

Vanacore, E., Taymaz, T. & Saygin, E., 2013. Moho structure of the Anatolian plate from receiver function analysis, *Geophys. J. Int.,* **193**(1), 329–337.

Wessel, P. & Smith, W. H.F., 1991. Free software helps map and display data, *EOS, Trans. Am. geophys. Un.,* **72**(41), 441–446, http://gmt.soest.hawaii.edu.

Wirth, E., Long, M. & Moriarty, J., 2016. A Markov chain Monte Carlo with Gibbs sampling approach to anisotropic receiver function forward modeling, *Geophys. J. Int.,* **208**(1), 10–23.

Zhu, L. & Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions, *J. geophys. Res.,* **105**(B2), 2969–2980.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Table S1 H–κ results for stations with more than N $\geq$ 8 suitable receiver functions. N: number of acceptable receiver functions; C.P.: criteria passed; H_{soln} and κ_{soln}: H and κ for final solution, $\pm$ the measurement error; H_{mean} and κ_{mean}: Mean H and κ for the 1000 repeats, $\pm$ the standard deviation. ACE_s and SNR_s: ACE and SNR of H–κ solution; ACE_m and SNR_m: Mean ACE and SNR for the 1000 repeats. CCC: mean cross-correlation coefficient; Freq: receiver function frequency of the solution.

Table S2 H–κ results for stations that are analysed with limited frequencies denoted by F_{max}. The other table headings are as per Supporting Information Table S1.

Figure S1 Synthetic model with a sharp Moho and 5 per cent whole crustal anisotropy with a fast axis oriented north. (a) The blue-dashed line indicates V_P with depth, and the red-dashed line indicates V_S with depth. (b) Final H–κ stack solution for the model. (c) Distribution of H and κ solutions for the model. The dot colour indicates the cluster that a result is assigned to in the hierarchical cluster analysis. The black cross is the expected H and κ from the input model, the red cross is the mean of H and κ from the 1000 repeats, the yellow cross is the mode combination of H and κ, the blue cross is the combination of H and κ selected by the cluster analysis. The black box marks one standard deviation in H and κ calculated for the 1000 repeat results.

Figure S2 Synthetic model with a sharp Moho and 15° dipping slab (with 5 per cent down-dip anisotropy) at 80 km depth to partially replicate the model of Frederiksen & Bostock (2000). Figure details as per Supporting Information Fig. S1.

Figure S3 The results from the study by Gilligan *et al.* (2016) using joint inversion of surface waves with receiver functions. The Rae domain (where ILON is situated) is found to have the sharpest Moho; the Ungava domain (where SCHQ is located) has a very gradual Moho transition.

Figure S4 Result diagram for station SCHQ on the Canadian shield (Fig. 8), where only solutions using receiver functions with frequencies $\leq$1.2 Hz are sampled. Figure details are as per Fig. 7.

Figure S5 Final result diagram for station FURI on the Ethiopian Plateau (Fig. 11). Figure details are as per Fig. 7.

Figure S6 Final result diagram for station CHGE on the Ethiopian Plateau (Fig. 11), using solutions from all frequencies. Figure details are as per Fig. 7.

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

APPENDIX A: CLUSTER ANALYSIS

Before hierarchical clustering can be performed, H and κ must span normalized ranges to avoid the much larger Euclidean distances in H dominating the clustering algorithm. H and κ values are thus re-scaled between 0 and 1 using their respective ranges:

$$H_{\text{scaled}} = \frac{H - H_{\text{min}}}{H_{\text{max}} - H_{\text{min}}}, \tag{A1}$$

$$\kappa_{\text{scaled}} = \frac{\kappa - \kappa_{\text{min}}}{\kappa_{\text{max}} - \kappa_{\text{min}}}, \tag{A2}$$

where H and κ are the original values and H_{min}, κ_{min}, H_{max} and κ_{max} are the minimum and maximum grid-search values of H and κ, respectively. Henceforth, H and κ represent the re-scaled measurements and distances are between re-scaled measurements.

Initially, there are $N = 1000$ scaled measurement pairs (H_i, κ_i) with variances ($\sigma^2_{H_i}$, $\sigma^2_{\kappa_i}$) where $i = 1...N$. The data are divided into M clusters with each cluster, C_j, containing N_j data points, where $j = 1...M$. To reduce the effect of discrete data, each initial cluster is assigned an initial error of 1/99 in H and κ that corresponds to the 99 intervals used in the grid search.

Starting with $M = N$ clusters, the intercluster Euclidean distances are calculated for all possible cluster pairs and the closest two clusters are combined into one, reducing the number of clusters by one. The number of points (N_j) per cluster C_j is calculated, and the mean centroid position of each cluster centre ($\bar{H}_j$, $\bar{\kappa}_j$) is calculated as

$$\bar{H}_j = \frac{\sum_{i=1}^{N_j} H_i^{(j)}}{N_j}, \tag{A3}$$

$$\bar{\kappa}_j = \frac{\sum_{i=1}^{N_j} \kappa_i^{(j)}}{N_j}, \tag{A4}$$

where $H_i^{(j)}$ and $\kappa_i^{(j)}$ refer to the i number of data points within cluster j.

Euclidean distances are re-calculated between all remaining clusters, including the newly merged cluster, and the process is repeated until one cluster remains containing all N data points. The optimum number of clusters is between $1 \leq M \leq N$ (Everitt *et al.* 2001). The optimum number of clusters is chosen automatically using the criterion of Caliński & Harabasz (1974) and Duda *et al.* (1973). The criteria of Caliński & Harabasz (1974) are defined by

$$c(M) = \frac{(N - M)\text{trace}(B)}{(M - 1)\text{trace}(W)}, \tag{A5}$$

where B is the between-cluster covariance and W is the within-cluster covariance, both calculated at each cluster step ($M = 1...N$). The optimum number of clusters is found when $c(M)$ is a maximized. The between-cluster covariance (B) and within-cluster covariance (W) are

$$B = \begin{bmatrix} \sum_{j=1}^{M}(\bar{H}_j - \bar{H})^2 & \sum_{j=1}^{M}(\bar{H}_j - \bar{H})(\bar{\kappa}_j - \bar{\kappa}) \\ \sum_{j=1}^{M}(\bar{H}_j - \bar{H})(\bar{\kappa}_j - \bar{\kappa}) & \sum_{j=1}^{M}(\bar{\kappa}_j - \bar{\kappa})^2 \end{bmatrix}, \tag{A6}$$

$$W = \begin{bmatrix} \sum_{j=1}^{M}\sum_{i=1}^{N_j}(H_i^{(j)} - \bar{H}_j)^2 & \sum_{j=1}^{M}\sum_{i=1}^{N_j}(H_i^{(j)} - \bar{H}_j)(\kappa_i^{(j)} - \bar{\kappa}_j) \\ \sum_{j=1}^{M}\sum_{i=1}^{N_j}(H_i^{(j)} - \bar{H}_j)(\kappa_i^{(j)} - \bar{\kappa}_j) & \sum_{j=1}^{M}\sum_{i=1}^{N_j}(\kappa_i^{(j)} - \bar{\kappa}_j)^2 \end{bmatrix}, \tag{A7}$$

where $H_i^{(j)}$ and $\kappa_i^{(j)}$ are the H and κ values of each measurement (i) in each cluster (j), and $\bar{H}$ and $\bar{\kappa}$ are the mean H and κ values for the entire data set:

$$\bar{H} = \frac{\sum_{i=1}^{N} H_i}{N}, \tag{A8}$$

$$\bar{\kappa} = \frac{\sum_{i=1}^{N} \kappa_i}{N}. \tag{A9}$$

The criteria of Duda *et al.* (1973) use the ratio of within-cluster variances before and after the two clusters are combined into a single cluster. The within-cluster variance of the two clusters prior to being combined is

$$\sigma_2^2 = \sum_{j=1}^{2}\sum_{i=1}^{N_j}[(H_i^{(j)} - \bar{H}_j)^2 + (\kappa_i^{(j)} - \bar{\kappa}_j)^2], \tag{A10}$$

and the within-cluster variance once the two clusters are combined is

$$\sigma_1^2 = \sum_{i=1}^{N_1}[(H_i^{(1)} - \bar{H}_1)^2 + (\kappa_i^{(1)} - \bar{\kappa}_1)^2]. \tag{A11}$$

The assumption is that the two clusters will be combined into one cluster that is rejected when

$$\left(1 - \frac{\sigma_2^2}{\sigma_1^2} - \frac{2}{\pi p}\right)\left(\frac{N_j p}{2[1 - 8/(\pi^2 p)]}\right)^{1/2} > c_{\text{critical}}, \tag{A12}$$

where $p = 2$ and is the number of parameters in the analysis and $c_{\text{critical}} = 3.20$, the value assumed when the data points within a cluster are normally distributed (Milligan & Cooper 1985). The optimum number of clusters is found when eq. (A12) is invalidated as the number of clusters is reduced.

The optimum number of clusters is taken to be the criteria that indicates the larger number of clusters up to a maximum of $M = 7$ clusters. To find the most suitable cluster, we define the within-cluster variance ($\sigma^2_{c_j}$) and error variance ($\sigma^2_{d_j}$) according to Teanby *et al.* (2004)

for each cluster containing >15 points.

$$\sigma_{c_j}^2 = \frac{\sum_{i=1}^{N_j}(H_i^{(j)} - \bar{H}_j)^2 + (\kappa_i^{(j)} - \bar{\kappa}_j)^2}{N_j}, \tag{A13}$$

$$\sigma_{d_j}^2 = \left[\sum_{i=1}^{N_j} \frac{1}{(\sigma_{H_i}^{(j)})^2}\right]^{-1} + \left[\sum_{i=1}^{N_j} \frac{1}{(\sigma_{\kappa_i}^{(j)})^2}\right]^{-1}. \tag{A14}$$

We define the overall variance ($\sigma_{o_j}^2$) for each remaining cluster as

$$\sigma_{o_j}^2 = \max(\sigma_{c_j}^2, \sigma_{d_j}^2), \tag{A15}$$

The best overall cluster has the minimum $\sigma_{o_j}^2$ and the final H–κ solution is measurement with smallest combined rescaled errors in H and κ from within this chosen cluster.