Automatic seismic waveform identification using a Convolutional Neural Network

Jorge Alejandro Garcia[1], Lauren Waszek[1], Benoit Tauzin[2], and Nicholas Charles Schmerr[3]

[1]New Mexico State University
[2]Universite de Lyon
[3]University of Maryland, College Park

November 22, 2022

Abstract

Typical seismic waveform datasets comprise from hundreds of thousands to several millions records. Compilation is performed by time-consuming handpicking of phase arrival times, or signal processing algorithms such as cross-correlation. The latter generally underperform compared to handpicking. However, inconsistencies across and within handpicked datasets creates disagreement between observations and interpretation of Earth's structure. Here, we exploit the pattern recognition capabilities of Convolutional Neural Networks (CNN). Using a large global handpicked dataset, we train a CNN model to identify the seismic shear phase SS. This accelerates, automates, and makes consistent data compilation. The CNN model is then employed to identify precursors to SS generated by mantle discontinuities. The model identifies precursors in stacked and individual seismograms, producing new measurements of the mantle transition zone with quality comparable to handpicked data. The capability to rapidly obtain new, high-quality observations has implications for automation of future seismic tomography inversions and body wave studies.

Automatic seismic waveform identification using a Convolutional Neural Network

J. A. Garcia[*,1], **L. Waszek**[1,2], **B. Tauzin**[2,3], **N. Schmerr**[4]

[1]Department of Physics, New Mexico State University, Las Cruces, NM 88003, USA
[2]Research School of Earth Sciences, Australian National University, Acton, ACT 0200, Australia
[3]Laboratoire de Geologie de Lyon: Terre, Planates, Environnement, Universite de Lyon, Universite Lyon 1 and Ecole Normale Superieure de Lyon, UMR CNRS 5276, F-69622 Villeurbanne, France
[4]Department of Geology, University of Maryland, 8000 Regents Dr., College Park, MD 20742, USA

Key Points:

- We train a 1D Convolutional Neural Network to identify the arrival peaks of SS phases from a large data set of 58,567 handpicked waveforms.
- The model is used to predict for the arrival time of SS precursors relative to SS in stacked data and individual seismograms.
- New maps of the 410-km and 660-km discontinuities are generated using the model's picks, and show excellent agreement to maps from handpicked precursors.

Corresponding author: Jorge Garcia, `jorgeagr97@gmail.com`

Abstract

Typical seismic waveform datasets comprise from hundreds of thousands to several millions records. Compilation is performed by time-consuming handpicking of phase arrival times, or signal processing algorithms such as cross-correlation. The latter generally underperform compared to handpicking. However, inconsistencies across and within handpicked datasets creates disagreement between observations and interpretation of Earth's structure. Here, we exploit the pattern recognition capabilities of Convolutional Neural Networks (CNN). Using a large global handpicked dataset, we train a CNN model to identify the seismic shear phase SS. This accelerates, automates, and makes consistent data compilation. The CNN model is then employed to identify precursors to SS generated by mantle discontinuities. The model identifies precursors in stacked and individual seismograms, producing new measurements of the mantle transition zone with quality comparable to handpicked data. The capability to rapidly obtain new, high-quality observations has implications for automation of future seismic tomography inversions and body wave studies.

1 Introduction

Seismology is the major observational tool to map the structure and properties of Earth's interior. Global studies of the Earth benefit from hundreds of thousands of seismograms to make observations. The properties of seismic wave phase arrivals within seismograms (arrival time, amplitude, coda) provide measurements of Earth's velocity and attenuation structures. Although some studies use automated waveform processing to identify seismic phases (e.g., Earle & Shearer, 1994; Chambers et al., 2005; Houser et al., 2008), visual inspection of waveforms is used in many studies due to higher accuracy (e.g., Flanagan & Shearer, 1998; Schmerr & Garnero, 2006; Deuss, 2009; Waszek et al., 2018). However, handpicking is time-consuming, and susceptible to the decisions of the scientist. Inconsistencies across and within datasets propagate errors when determining geophysical models from the measurements, as evidenced by differences between global mantle discontinuity topography maps from the same data types (e.g., Flanagan & Shearer, 1998; Schmerr & Garnero, 2006; Deuss, 2009; Huang et al., 2019).

There are two possible approaches to create an accurate system capable of identifying the arrival of seismic phases. The ideal approach would attempt to find an accurate representation of the data by extracting useful features that describe time of the phase arrival. A more straightforward method allows the computer to perform this task, finding the necessary patterns through representation learning (LeCun et al., 2015). The use of deep learning trains a system that is capable of taking data, identifying characterizing features, and producing an informed prediction based on these signatures.

The task of picking seismic phases relies upon visual cues. Naturally, it is easier for the human eye to discern the correct peak associated with a particular seismic phase when trained to do so. This is the inspiration for the use of a Convolutional Neural Network (CNN) to perform this task. CNNs are the preferred deep learning algorithm for pattern recognition problems due to their ability to identify any set of objects given enough layers (Girshick et al., 2014; Simonyan & Zisserman, 2014; Krizhevsky et al., 2017). Within seismology, CNNs have proven capable of detecting and locating earthquakes (Perol et al., 2018), performing seismic arrival labeling (McBrearty et al., 2019), denoising data (Zhu et al., 2019), and picking the arrival time of compressional and shear wave phases (Ross et al., 2018; Zhu & Beroza, 2018).

Here, we apply CNNs to make new observations of mantle discontinuities. The two major global discontinuities at 410 km and 660 km depth ("410", "660") bound the mantle transition zone ("MTZ"). They result from mineral phase transitions in olivine as pressure and temperature increase with depth (Katsura & Ito, 1989; Ito & Takahashi, 1989). Due to their opposing Clapeyron slopes, the depths of the discontinuities respond oppositely to temperature. In cold regions the 410 becomes shallower and the 660 becomes deeper; vice versa in hot regions. Consequently, their separation acts as a first order thermometer for the MTZ.

Mapping of mantle discontinuities globally has been achieved through measuring shear-wave reflections from underneath these boundaries (e.g., P. M. Shearer, 1993; Flanagan & Shearer, 1998; Houser et al., 2008; Deuss, 2009; Waszek et al., 2018; Huang et al., 2019). SS is a seismic shear wave phase with two legs in the Earth's mantle and one reflection from Earth's surface (Fig. 1a). Reflections from mantle discontinuities generate precursors to SS ("SdS", where d is discontinuity depth), which arrive prior to the main phase. The SdS-SS travel time difference informs regarding the discontinuity depth.

We use a CNN to train a model capable of identifying SS in seismograms. We implement a duplication procedure on a large handpicked global dataset of 58,567 SS data (Waszek et al., 2018) to produce huge amounts of training data (316,262). Using the trained

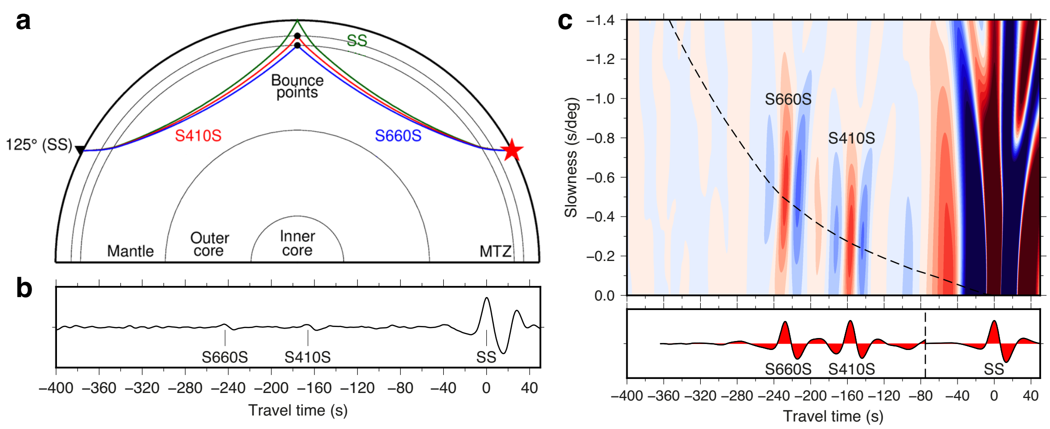

Figure 1. a. Ray paths of SS and its precursors, S410S and S660S. The red star denotes the location of the seismic event, and the black triangle a station to detect seismic waves. b. An example of a high-quality seismogram showing the SS, S410, and S660S arrivals. c. Global vespagram stack for all data and cross-section through the theoretical relative precursor time and slowness. The precursor amplitudes have been magnified and normalized to the SS phase amplitude; magnification factor is typically around 30.

model, a scanning algorithm quantifies the quality of a phase signal within a waveform. We then employ the algorithm to output the arrival times and quality of SS precursors, in both stacked and individual seismograms. Maps of the depths of the 410 and 660 discontinuities are generated, using the predictions to evaluate model performance. The study provides a new method to rapidly and automatically compile large high-quality seismic datasets and measurements, with implications for future seismic studies particularly global tomography.

2 Seismic Data and Processing

Our study employs a large, handpicked dataset of 58,567 SS waveforms (Waszek et al., 2018), aligned at the maximum peak in Fig. 1b. The seismograms are corrected for mantle and crustal structure using S40RTS (Ritsema et al., 2011) and Crust2.0 (Bassin et al., 2000). A full description of processing methods is provided in Waszek et al. (2018) and Waszek et al. (2020).

Precursors to SS are typically too small in amplitude to be identified on individual seismograms. Instead, the data are stacked in regional overlapping spherical caps partitioned by common reflection points, weighted according to the signal-to-noise ratio, where noise is the root-mean-square amplitude in the precursor window (-400 to -100 s). Vespagrams show stacked signals as a function of travel time and slowness relative to the main SS arrival (Figure 1c) (Davies et al., 1971). The cross-section taken along the dotted line is the predicted time and slowness of the precursors to SS in a standard reference model. These vespagrams are usually analyzed manually to measure the time and amplitude of the precursor signals. Bin radii used here are 5°, 7.5°, 10°, and 15°; these are selected to account for heterogenous data coverage in different regions: smaller bin sizes in areas of higher data density to obtain greater resolution.

3 Model Training and Evaluation

The SS dataset is divided into a training set of 90% of the seismograms, with the remaining 10% left as an unseen testing set to evaluate the model. For training, we use data uncorrected for crustal and mantle structure. Similar to Ross et al. (2018), a 40 s window of the 500 s waveform is considered for the model input, with the starting point being the theoretical onset time predicted by the 1D Earth model "PREM" (Dziewonski & Anderson, 1981). This smaller window permits for tractable computation time when training the network. Additionally, this enables us to augment the number of training records by creating variations of these segments, to obtain a more accurate model.

For each 40 s segment, we created five additional windows with a random time shifts of ± 5 s, increasing the training set by a factor of six. Although they are the same waveform shifted, to the network they appear as independent signals. This random shift allows the model to take into account the variability between the time of the onset and the peak, thereby enhancing the spatial invariance of the model. For the testing set, only the 40 s window from the theoretical onset time was used.

We used the augmented dataset to train a 1D CNN through the Keras library (Chollet et al., 2015), using the "RossNet" model architecture employed in Ross et al. (2018). The overall configuration of the layers is visualized in Fig. 2a. The ReLU activation function (Nair & Hinton, 2010) was used in both the convolutional and fully-connected layers. Model cost was evaluated with the Huber loss function (Huber, 1964), and the Adam algorithm was used for layer weight optimization (Kingma & Ba, 2014). In order to account for variations in model convergence due to random initialization of weights, we trained five dif-

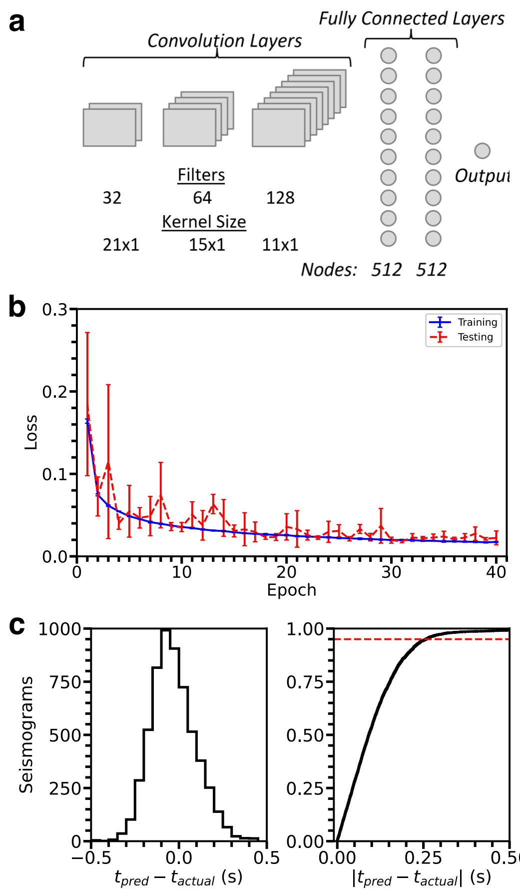

Figure 2. a. Diagram of the "RossNet" architecture used, from Ross et al. (2018). b. Average loss of five models from the architecture used across epochs during the training stage (blue line) and testing stage (red line). These represent the error in fit of each model to the data. The error bars correspond to one standard deviation of the average loss of the models on the unseen testing set. The instance that results in the lowest overall loss is the set of used weights. c. Histogram of prediction error (left) and cumulative histogram of absolute error (right) for the testing set, the red dashed line represents the 95th percentile of the data.

ferent models for 40 epochs. The models were trained on two NVIDIA Tesla P100 graphics processing units (GPUs); each epoch took approximately three minutes to train.

Figure 2b shows the average and standard deviation of the loss over the models and epochs; this corresponds to the error in fit of the models to the data. Results are shown for the training and test datasets. Despite the variability of the errors in the testing dataset, likely due to its relatively small size compared to the training set, there is an overall trend of decreasing loss as with increasing epoch. We use the best performing set of weights across all 200 instances for the remaining analysis. Summary statistics show an average prediction error on par with the sampling rate of the seismograms (0.1 s, Fig. 2c). The cumulative histogram confirms that 95% of the model's prediction for the test dataset are within 0.25 seconds of the picked arrival. This size of error is insignificant, since the picks are subsequently aligned to the maximum amplitude.

4 Phase Waveform Quality

We desire not only the maximum arrival time of each waveform, but also the quality of the phase. Quality of the waveform during handpicking is normally judged visually in a qualitative manner. Here we propose a scheme to assign a quantitative description of the quality using the trained model.

The CNN model was constrained to accept only 40 s of the original 500 s seismogram as its input. As such, we create a scanning algorithm that iteratively moves along the entire seismogram in 40 s windows to define the prediction quality through a statistical definition. The top three plots in Figure 3 provide an illustrative example of the scan algorithm. A 40 s window of data from time t to t+40 seconds is chosen and provided as input to the model to find the best matching shape to the ideal SS signal, giving an arrival time prediction for this window. The window from $t+\Delta t$ to $t+\Delta t+40$ is then analyzed; this process repeats for the entire seismogram. The sliding window moves in steps of the seismograms' sampling rate, i.e. $\Delta t = 0.1$ s. As the scanning iterates, the arrival time will be consistently identified if it is enclosed in the windows. In some cases the model can identify the onset of phases outside of the window (Supplementary Movie 1). If no recognizable features are present, the best prediction varies considerably as the scan iterates.

The obtained prediction times for a particular signal are not precisely the same throughout the scan. Due to slight differences in information within each window, the predicted time will vary by a value close to the sampling time of the data. We employ the DBSCAN algorithm (Ester et al., 1996) implementation in the Python scikit-learn library (Pedregosa et al., 2011) to perform density-based clustering of the predicted times. This way, a large amount of predictions that are close to each other form a tight cluster. Each prediction in a cluster is an approximate measure of a time $< t >$ of the signal, with a standard deviation corresponding to the error on the prediction ϵ.

We use the quantity of predictions to define a quality measure for each signal. Let T be the window size used in the model, and Δt the sampling time of the seismogram. An ideal arrival will therefore appear $T/\Delta t = 400$ times during the scanning process. The quality of a prediction q_{pick} is thus calculated:

$$q_{\text{pick}} = \frac{N_{\text{pred}} \Delta t}{T} \tag{1}$$

with N_{pred} the number of predictions within a cluster. We retain the prediction with the highest prediction frequency, or quality, as the SS "pick" for a particular seismogram. Correct identifications of SS result in higher quality of the main arrival compared to other features within the waveform (Fig. 3a).

This scheme of defining a quality also allows us to determine the correct polarity of the SS signal. Since the model is only trained on seismograms with positive polarity SS signals, running the scanning window on a seismogram with a negative polarity SS peak results in inconsistent predictions with lower quality around the time of the SS arrival (Supplementary Movie 2). In order to determine the polarity of an unknown seismogram, we employ the scanning algorithm on both the waveform and its inverse. For seismograms with an identifiable SS signal, the version with a positive polarity SS phase has the highest quality pick.

5 Prediction of SS Precursors in Stacked Data

Precursors ("SdS") may be approximated as lower amplitude versions of the main arrival with a similar shape. Thus, a model trained on the main arrival should be able to identify precursory signals in stacked waveforms due to their similarity, exploiting the pattern recognition capabilities of CNNs. We find that our scanning algorithm can indeed identify precursors as the highest quality predictions prior to the SS arrival (Fig. 3b; Supplementary Movie 3).

The handpicking quality criteria requires clear S410S and S660S signals in both the vespagrams and cross-sections, with no interfering phases or significant noise in the vespagram. The vespagrams are assigned qualities from "a" to "d". The "a" vespagrams have no noise and clear precursors with waveforms very similar to SS, while "d" bins have much noise and the precursor shape is dissimilar to SS, and are not retained for analysis of any precursors other than S410S and S660S (see Waszek et al. (2018) for a full description of methodology). Here, we use the CNN to obtain predictions of the S410S

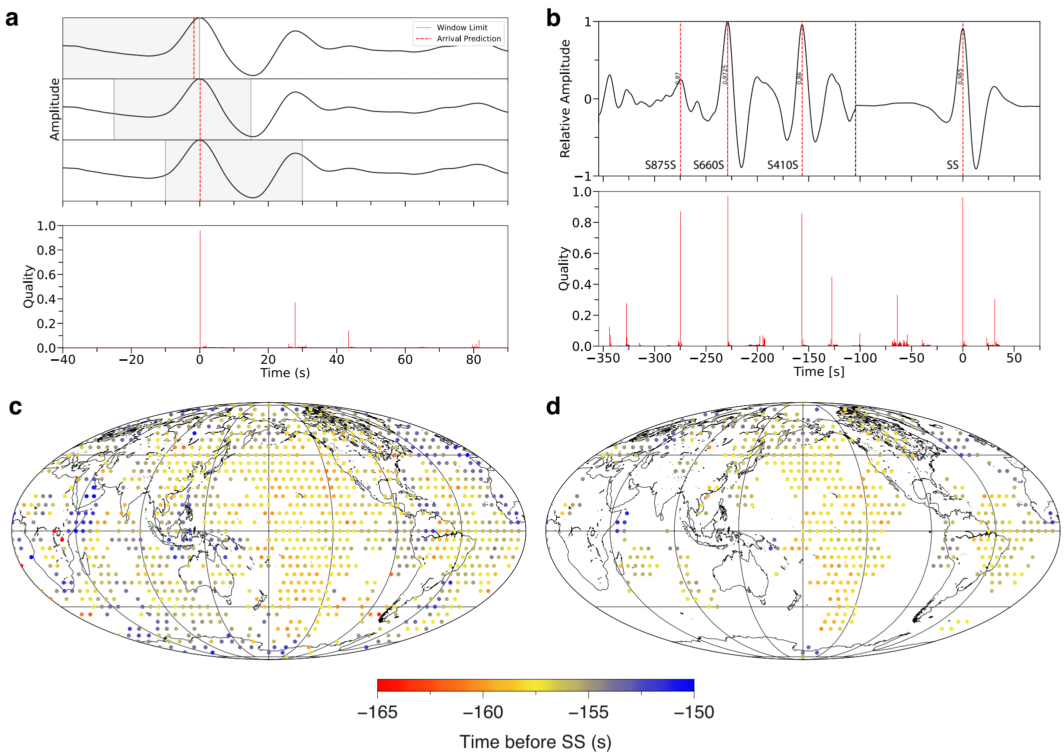

Figure 3. a. Example of iterative prediction for a seismogram, at 0, 15 s, and 30 s, with the histogram of prediction rate. The portion of the waveform within the shaded area is used as input for the model. The red line is the predicted arrival for the given input. The signal is predicted consistently when enclosed by the window, and the true SS arrival is at the time of highest prediction rate. b. Prediction of arrival times in a stacked cross-section and histogram. The four highest prediction times are marked on the cross-section. c. S410S precursor arrival times for 5° bin stacked data picked using the deep learning model. The minimum prediction quality of picks retained is 60%. d. Corresponding S410S arrival times measured using handpicking and visual quality checks.

and S660S times for all of the stacks from the bins (corrected for 3D mantle and crustal structure). We retain picks with quality 0.6 or higher; following visual inspection, this is the lowest quality for which precursors could be identified (Fig. S1). The resulting maps of S410S arrival times for 5° bins show good agreement in the measurements from the CNN (Figure 3c) and handpicking (Fig. 3d), with a correlation coefficient of 0.999. This indicates that, where both methods retain a bin, they measure the same relative arrival time for the precursor. This is true for both S410S and S660S picks in all bin sizes (Fig. S2).

The CNN picks retain significantly more precursor picks, which were removed by the handpicking quality procedure. The higher retention rates for the CNN is found for S410S and S660S measurements in all bin sizes (Fig. S3-S10). This suggests that stricter visual quality procedures may remove useful information, i.e. that the CNN can identify seismic signals in noisy data whereas handpicking cannot. Furthermore, the CNN model provides numerical measures of quality that the handpicking does not. Average quality of handpicked versus autopicked bins confirms that the bins removed by the CNN are indeed of lower quality than those retained by handpicking (Table S1). Furthermore, the average CNN quality also corresponds well to the handpicked quality, i.e. "a" quality bins have the highest CNN quality (Table S2). In order for the CNN method to retain the same number of bins as the handpicking, the minimum pick quality must be increased to as much as 0.86 for S660S in 5° bins (Table S3). This value drops as bin size increases, to 0.6125 for S410S in 15°, as the stacked signals become less similar to SS due to averaging over increasingly larger regions.

6 Prediction of SS Precursors in Individual Seismograms

Following the success of the CNN model for identifying precursors in the stacked data, we next apply it to precursors in individual seismograms. Normally, these can only be visually identified in the highest-quality waveforms due to their small amplitudes (e.g. Fig. 1b). We scan the corrected data set, and consider the top 10 predictions before the main arrival (Supplementary Movie 4). Predictions with a quality below 0.6 are discarded, retaining a total of 38,985 measurements. This corresponds to multiple picks in some seismograms, and none in others. Examining the predictions as a function of epicentral distance (Figure 4a) reveals clusters corresponding to the 410 and 660, in addition to regional-scale discontinuities at 300-km and 520-km depth. The gaps with different slowness to the precursors (particularly between $100-120°$ distance) are interfering phases that the model does not pick, namely SdiffS660S which has a negative polarity, highlighting its success to discard non-SdS signals.

The linear trends for both global discontinuities are calculated using the DBSCAN algorithm for density-based clustering, to determine statistically the predictions most likely to correspond to S410S and S660S. We select arrival time bounds of -185 to -135 s before the main arrival for S410S, and -250 to -200 s for S660S. These are selected to fully enclose the observed data trends, while excluding theoretical arrival times for other discontinuities, to ensure that the most dense cluster corresponds to robust picks. A linear fit, with the data weighted by pick quality, is applied as an initial estimate for the trends. Predictions within ±10 seconds of this fit are considered to also be correct measurements for the discontinuity in question. The weighted linear model is then fit to this new set of data points (Fig. 4b). Maps of the uncorrected and corrected relative travel time measurements are included in the Supplement (Fig. S11-S13).

7 Discussion

The task of pattern recognition in seismology is not new. Cross-correlation has previously been used to generate SS datasets (Houser et al., 2008), measure precursor arrival times in stacked data for the mid-mantle (Waszek et al., 2018), and identify precursor signals in individual data (P. Shearer, 1991). It performs well when the two signals are noise and defect-free, but the majority of real data does not fulfil these criteria. Setting the cross-correlation approach as our benchmark, we repeat the clustering analysis to identify 410-km and 660-km measurements from cross-correlation predictions, and compare to the CNN picks. A cutoff cross-correlation score of 0.9379 is required to obtain an equal number of precursor signals when using the cross-correlation method as compared to the CNN model (i.e. 38,985 picks), significantly higher than the 0.6 typically used for automatic cross-correlation picking (Chambers et al., 2005).

The histograms in Figures 5a and b are the number of predictions made between epicentral distances of $120 - 130°$ in time bins of 1 s. The two large Gaussian distributions correspond to predictions from the discontinuities, with the fraction of seismograms within the bin associated to that cluster shown. The CNN produces roughly twice as many predictions at this epicentral distance range, and identifies over 50% more precursors overall than using cross-correlation; e.g. 410 picks are found for 28% of seismograms using the CNN compared to 19% from cross-correlation, demonstrating its greater predictive capabilities.

Plotted in Figure 5c are a random selection of precursor picks from the CNN model with various epicentral distance and phase quality. These picks were considered by the clustering analysis to be true identification of precursors. A corresponding examination of picks with a range of qualities confirms the marked improvement in waveform shape with increasing quality (Fig. S1), and justifies our lower quality limit of 0.6. We note,

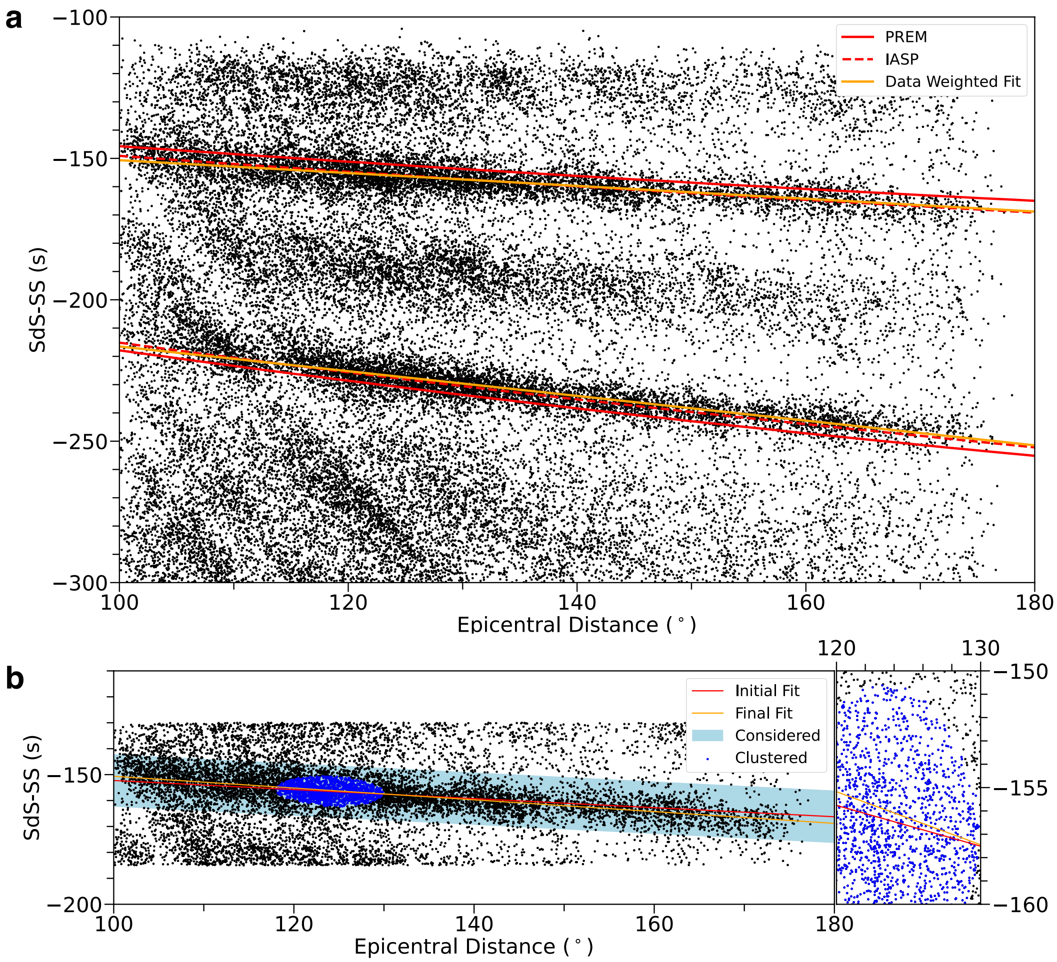

Figure 4. a. Predicted precursor relative arrival time as a function of epicentral distance for the individual seismograms, with theoretical (IASP and PREM) and fitted trends for the S410S and S660S measurements. Picks retained have prediction quality of 60% or higher. Note that these measurements have been corrected for S40RTS (Ritsema et al., 2011) and Crust2.0 (Bassin et al., 2000). b. Visualization of the procedure for determining real measurements for a discontinuity. We first consider a subset of the data that encloses the discontinuity in question. By using density-based clustering, the most dense cluster will consist of points that correspond to the observed trend, shown in blue. An initial linear fit is done using these points to have a guess at the trend, shown in red. An uncertainty cutoff is established, and points within that boundary are now considered to be real measurements, show in light blue. A final linear fit is performed on this new set of points to correct the trend, show in yellow. Notice the small difference between the initial and final linear models.

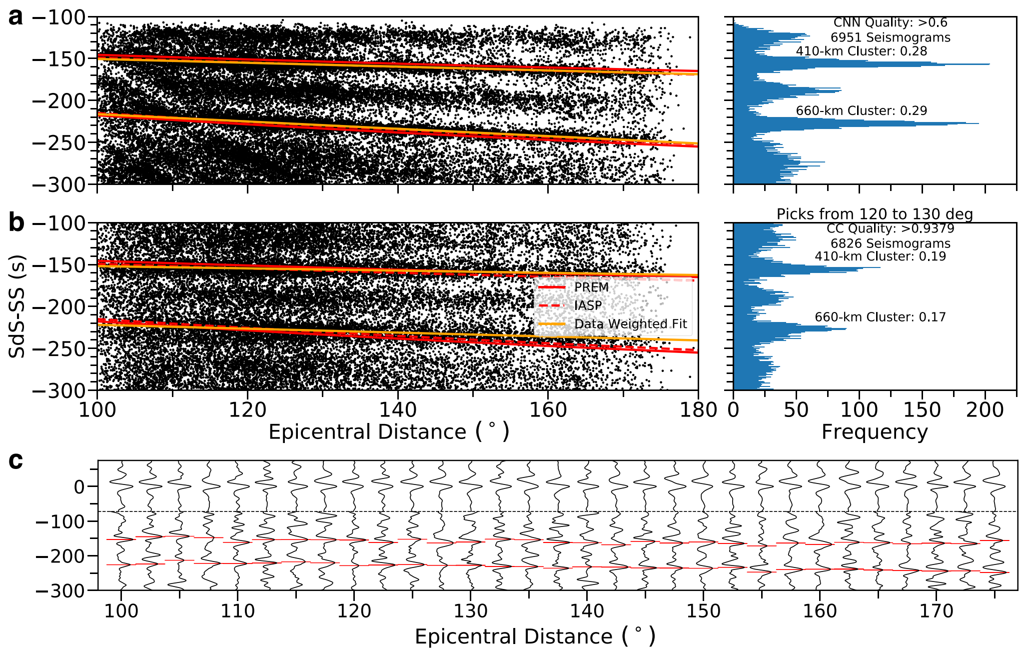

Figure 5. a. Predicted differential time as a function of epicentral distance for individual seismograms, with a histogram of picks between 120 – 130° epicentral distance range, for the CNN models. Lines correspond to predictions from IASP91 (Kennett & Engdahl, 1991) (red dotted), PREM (Dziewonski & Anderson, 1981) (orange solid), and the best fit (red solid). Cluster quantities refer to the proportion of picks in each cluster. b. As in (a), but for cross-correlation picks. c. A random selection of seismograms and their respective precursor picks from the CNN model. The width of the pick (red line) is proportional to 2σ of the predicted arrival.

however, that the CNN occasionally picks signals that appear to be sidelobes from negative amplitude interfering phases (Fig. 4a). This is because the code picks the best-matching signal in a window regardless of shape, relying on the moving window to produce quality. A future goal is the implementation of a null output. In the meantime, the DBSCAN clustering analysis could be applied to remove interfering signals and their sidelobes. This would be particularly useful for mid-mantle precursors which have both positive and negative polarities (Waszek et al., 2018). The cross-correlation picks do not pick the interfering negative signal gap, instead showing significant noise, highlighting its poorer performance.

SdS-SS differential travel time measurements from individual seismograms are a new type of measurement that is not yet widely used, primarily due to the difficulty in detection of the precursors. The measurements provide new high resolution observations of the MTZ discontinuities (Fig. S11-13), allowing for refinement of existing global and regional-scale seismic velocity models. For example, our preliminary global analysis presented here reveals that IASP91 (Kennett & Engdahl, 1991) provides a better fit to both the 410 and 660 (Fig. 5). PREM uses 400 and 670 km for the discontinuity depths, and our measurements here are deeper and shallower than these values respectively. In consequence, the outputs and future developments from our algorithm represent a critical contribution to global seismology, in particular for tomography modelling efforts which require measurements from millions of seismograms.

In addition to consistency of picking, and extraction of seismic signals from noise, the CNN technique provides a remarkable time saver in its capability to automatically process and pick seismic phases. Once a model is trained, the methods developed here allow for very rapid acquisition of new seismic datasets. The scanning algorithm picks a 140 s subset of a seismogram in approximately six seconds, which is similar to handpicking times, however the computer will continue to pick data constantly. Using a high performance computer, the scanning algorithm picked the entire dataset of 58,567 signals in 10 hours. In comparison, the same dataset required several months for compilation via handpicking (Waszek et al., 2018). Naturally, any automation represents a time saver compared to handpicking, and this method requires a significantly larger computational time then basic automatic algorithms (cross-correlation). However, it provides a performance comparable to the former; significantly better than the latter.

8 Conclusions

We have demonstrated the significant capabilities of CNNs in the task of picking seismic phases, exploiting the pattern recognition capabilities of these deep learning models. A trained model picks new data accurately and efficiently. It is able to identify other phases with similar features, and extract small-amplitude signals that typically appear masked by noise to the human eye. Thus, a model trained on SS data can produce a dataset of directly measured travel times for SS precursors, providing a new dataset to constrain Earth's upper mantle. Further consideration of deep learning models and potential applications to seismology could revolutionize the field by automatically picking waveforms as they become available. We encourage the use of and welcome contributions to our open-source Autopicker code.

Acknowledgments

The material is based upon work supported by NSF grants EAR-1661985 and EAR-1853662. J.A.G. was supported by a New Mexico Alliance for Minority Participation scholarship. L.W. is the recipient of a Discovery Early Career Research Award (project number DE170100329) funded by the Australian Research Council. B.T. and this work are supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement 793824. N.C.S. and L.W. were supported by NSF grant EAR-1361325. Waveform data were obtained from the IRIS Data Management Center (NSF grant EAR-1063471). The processed data are available from L.W. upon request. Travel time measurements will be made available from the ISC repository. The Autopicker code is available from https://github.com/JorgeAGR/neuralpick.

References

Bassin, C., Laske, G., & Masters, G. (2000). The current limits of resolution for surface wave tomography in north america. *EOS Trans AGU*, *81*(F897).

Chambers, K., Deuss, A., & Woodhouse, J. (2005). Reflectivity of the 410-km discontinuity from pp and ss precursors. *Journal of Geophysical Research*, *110*(B02301).

Chollet, F., et al. (2015). *Keras.* `https://keras.io`.

Davies, D., Kelly, E., & Filson, J. (1971). Vespa process for analysis of seismic signals. *Nat. Phys. Sci.*, *232*, 8-13.

Deuss, A. (2009, Oct 01). Global observations of mantle discontinuities using ss and pp precursors. *Surveys in Geophysics*, *30*(4), 301-326. Retrieved from `https://doi.org/10.1007/s10712-009-9078-y` doi: 10.1007/s10712-009-9078-y

Dziewonski, A. M., & Anderson, D. L. (1981). Preliminary reference earth model. *Physics of the Earth and Planetary Interiors*, *25*(4), 297 - 356. Retrieved from http://www.sciencedirect.com/science/article/pii/0031920181900467 doi: 10.1016/0031-9201(81)90046-7

Earle, P., & Shearer, P. (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, *84*(2), 366-376.

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd.*

Flanagan, M., & Shearer, P. (1998). Global mapping of topography on transition zone discontinuities by stacking ss precursors. *Journal of Geophysical Research*, *103*(B2), 2673-2692.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 ieee conference on computer vision and pattern recognition* (p. 580-587).

Houser, C., Masters, G., Flanagan, M., & Shearer, P. (2008). Determination and analysis of long-wavelength transition zone structure using ss precursors. *Geophysical Journal International*, *174*(1), 178-194.

Huang, Q., Schmerr, N., Waszek, L., & Beghein, C. (2019). Constraints on seismic anisotropy in the mantle transition zone from long-period ss precursors. *Journal of Geophysical Research: Solid Earth*, *124*(7), 6779-6800. doi: 10.1029/2019JB017307

Huber, P. J. (1964, 03). Robust estimation of a location parameter. *Ann. Math. Statist.*, *35*(1), 73–101. doi: 10.1214/aoms/1177703732

Ito, E., & Takahashi, E. (1989). Postspinel transformations in the system Mg2SiO4-Fe2SiO4 and some geophysical implications. *Journal of Geophysical Research: Solid Earth*, *94*(B8), 10637-10646. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB094iB08p10637 doi: 10.1029/JB094iB08p10637

Katsura, T., & Ito, E. (1989). The system Mg2SiO4-Fe2SiO4 at high pressures and temperatures: Precise determination of stabilities of olivine, modified spinel, and spinel. *Journal of Geophysical Research: Solid Earth*, *94*(B11), 15663-15670. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB094iB11p15663 doi: 10.1029/JB094iB11p15663

Kennett, B., & Engdahl, E. (1991). Travel times for global earthquake location and phase association. *Geophysical Journal International*, *105*, 429-465.

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization.*

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017, May). Imagenet classification with deep convolutional neural networks. *Commun. ACM*, *60*(6), 8490. Retrieved from `https://doi.org/10.1145/3065386` doi: 10.1145/3065386

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. doi: 10.1038/nature14539

McBrearty, I. W., Delorey, A. A., & Johnson, P. A. (2019, 01). Pairwise Association of Seismic Arrivals with Convolutional Neural Networks. *Seismological Research Letters*, *90*(2A), 503-509. Retrieved from `https://doi.org/10.1785/0220180326` doi: 10.1785/0220180326

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Icml.*

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2). Retrieved from `https://advances.sciencemag.org/content/4/2/e1700578` doi: 10.1126/sciadv.1700578

Ritsema, J., Deuss, A., van Heijst, H. J., & Woodhouse, J. H. (2011). S40rts: a degree-40 shear-velocity model for the mantle from new rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophysical Journal International*, *184*(3), 1223-1236. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.2010.04884.x` doi: 10.1111/j.1365-246X.2010.04884.x

Ross, Z. E., Meier, M., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120-5129. doi: 10.1029/2017JB015251

Schmerr, N., & Garnero, E. (2006). Investigation of upper mantle discontinuity structure beneath the central pacific using ss precursors. *Journal of Geophysical Research: Solid Earth*, *111*(B8). doi: 10.1029/2005JB004197

Shearer, P. (1991). Constraints on upper mantle discontinuities from observations of long-period reflected and converted phases. *Journal of Geophysical Research: Solid Earth*, *96*(B11), 18147-18182. doi: 10.1029/91JB01592

Shearer, P. M. (1993, 12). Global mapping of upper mantle reflectors from long-period SS precursors. *Geophysical Journal International*, *115*(3), 878-904. Retrieved from `https://doi.org/10.1111/j.1365-246X.1993.tb01499.x` doi: 10.1111/j.1365-246X.1993.tb01499.x

Simonyan, K., & Zisserman, A. (2014, 09). Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*.

Waszek, L., Schmerr, N. C., & Ballmer, M. D. (2018). Global observations of reflectors in the mid-mantle with implications for mantle structure and dynamics. *Nature Communications*, *9*(1), 385. doi: 10.1038/s41467-017-02709-4

Waszek, L., Tauzin, B., Schmerr, N. C., Ballmer, M. D., & Afonso, J. C. (2020). A poorly mixed mantle and its thermal state inferred from seismic waves. *In review*.

Zhu, W., & Beroza, G. C. (2018, 10). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261-273. Retrieved from `https://doi.org/10.1093/gji/ggy423` doi: 10.1093/gji/ggy423

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9476-9488.

Supporting Information for "Automatic seismic waveform identification using a Convolutional Neural Network"

J. A. Garcia*,1, L. Waszek[2,1], Benoit Tauzin[3,2], Nicholas Schmerr[4]

[1]Department of Physics, New Mexico State University, Las Cruces, NM 88003, USA

[2]Research School of Earth Sciences, Australian National University, Acton, ACT 2601, Australia

[3]Laboratoire de Geologie de Lyon: Terre, Planastes, Environnement, Universite de Lyon, Universite Lyon 1 and Ecole Normale Superieure de Lyon, UMR CNRS 5276, F-69622 Villeurbanne, France

[4]Department of Geology, University of Maryland, 8000 Regents Dr., College Park, MD 20742, USA

Contents of this file

1. Figures S1 to S13
2. Tables S1 to S3

Additional Supporting Information (Files uploaded separately)

1. Captions for Movies S1 to S4

Introduction

Corresponding author: J.A. Garcia, Department of Physics, New Mexico State University, Las Cruces, NM 88003, USA. jorgeagr97@gmail.com)

This document contains examples of precursors picks for different CNN quality values (Fig. S1), and plots of S410S-SS and S660S-SS relative travel times for stacked data in all bin sizes, comparing autopicked to handpicked results (Figs. S2-S10). We also present plots of S410S-SS and S660S-SS relative travel times generated by autopicked results for individual seismograms (Fig. S11-S13).

Movies S1 to S4 show visually how the autopicker iteratively scans stacked data or seismograms to identify signals, and determines the polarity of the signals by calculating quality values.

Table S1 lists average pick quality for stacked data, comparing handpicked and autopicked. Table S2 is the average CNN pick quality for handpicked data, separated by handpicked quality. Table S3 shows the minimum CNN quality pick values required to retain the same number of bins as the handpicking quality check procedure.

Movie S1. Example of the CNN model scanning for and identifying an SS phase signal.

Movie S2. Determining phase polarity by scanning on the positive and negative polarity versions of a seismogram.

Movie S3. Identifying SS precursors in a stack.

Movie S4. Scanning for SS precursors in an individual seismogram.

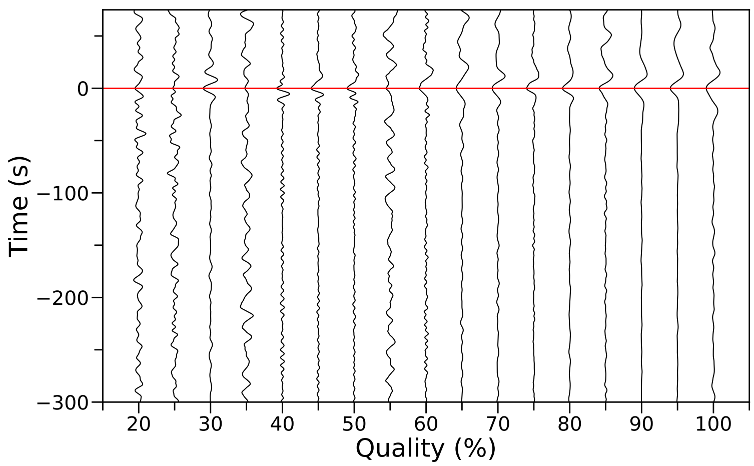

Figure S1. Examples of SS picks of various qualities.

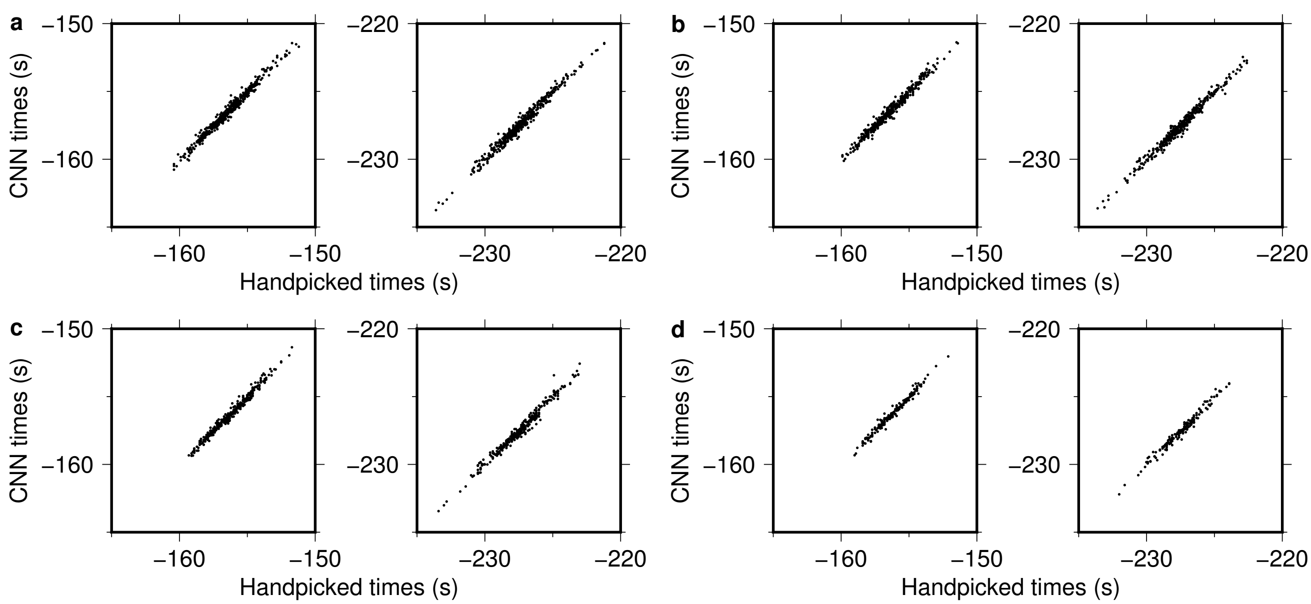

Figure S2. Comparison of measured travel times in handpicked versus autopicked stacked data, for all bin sizes. a. 5°. b. 7.5°. c. 10°. d. 15°.

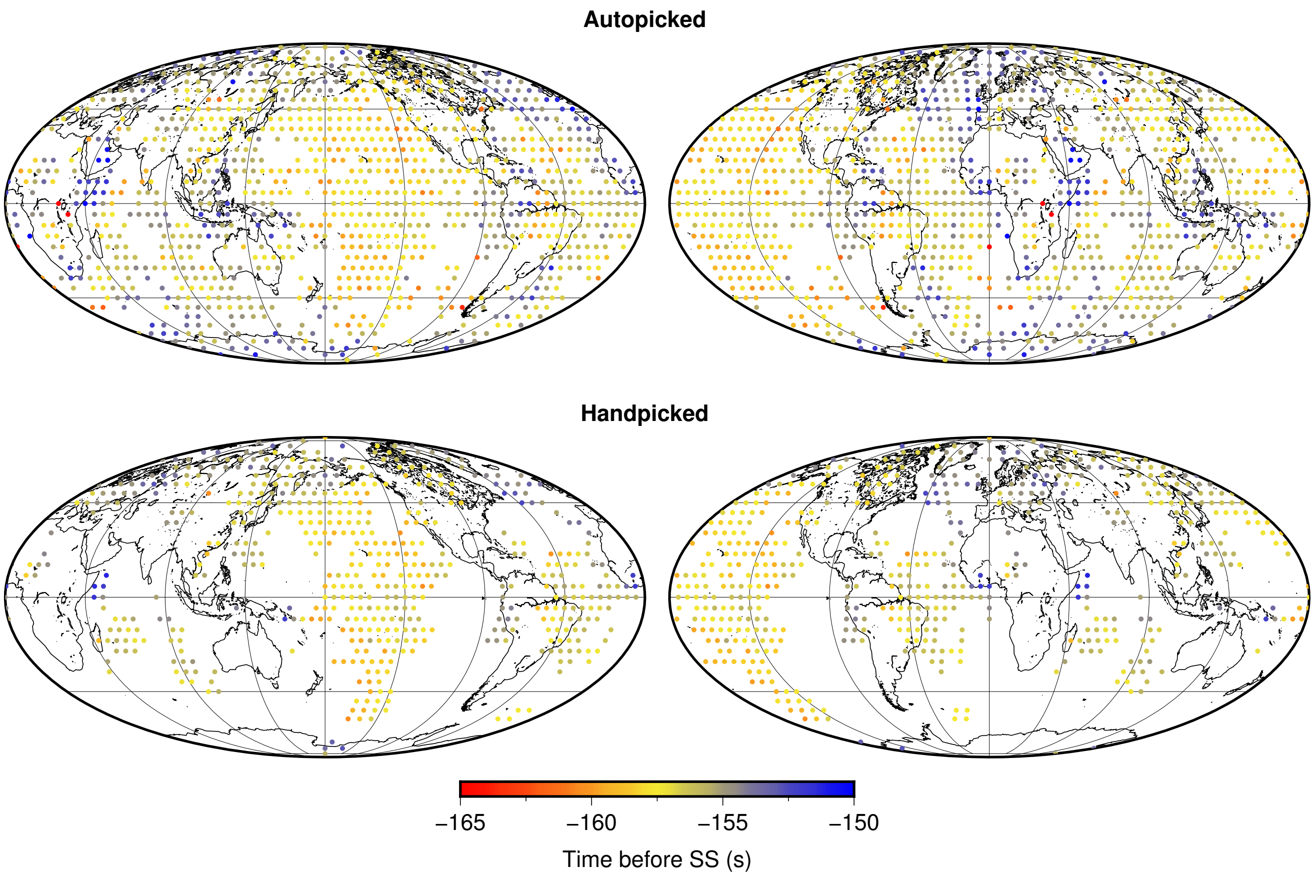

Figure S3. Maps of autopicked (top) and handpicked (bottom) S410S-SS travel time measurements in stacked data, 5° radius caps.

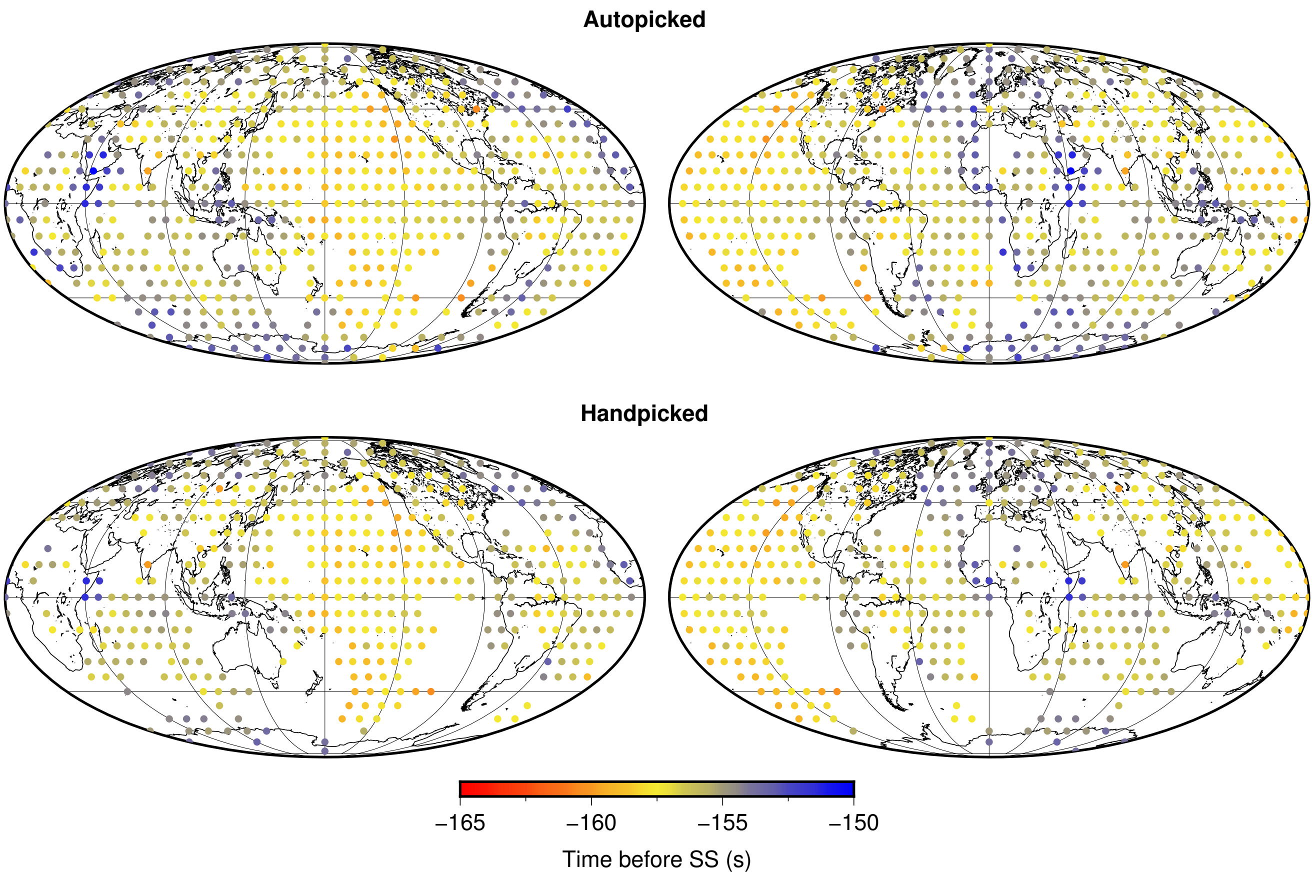

Figure S4. Maps of autopicked (top) and handpicked (bottom) S410S-SS travel time measurements in stacked data, 7.5° radius caps.

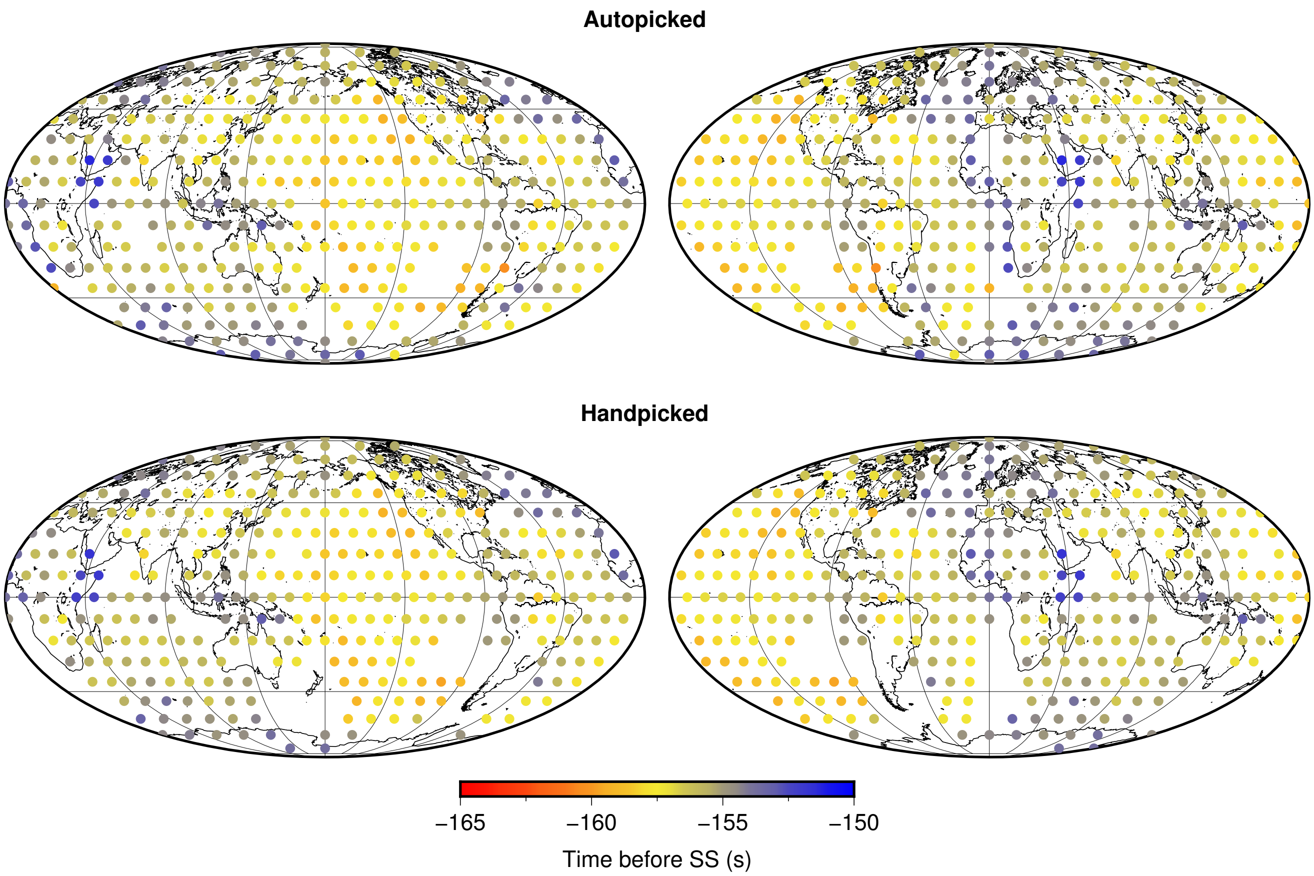

Figure S5. Maps of autopicked (top) and handpicked (bottom) S410S-SS travel time measurements in stacked data, 10° radius caps.

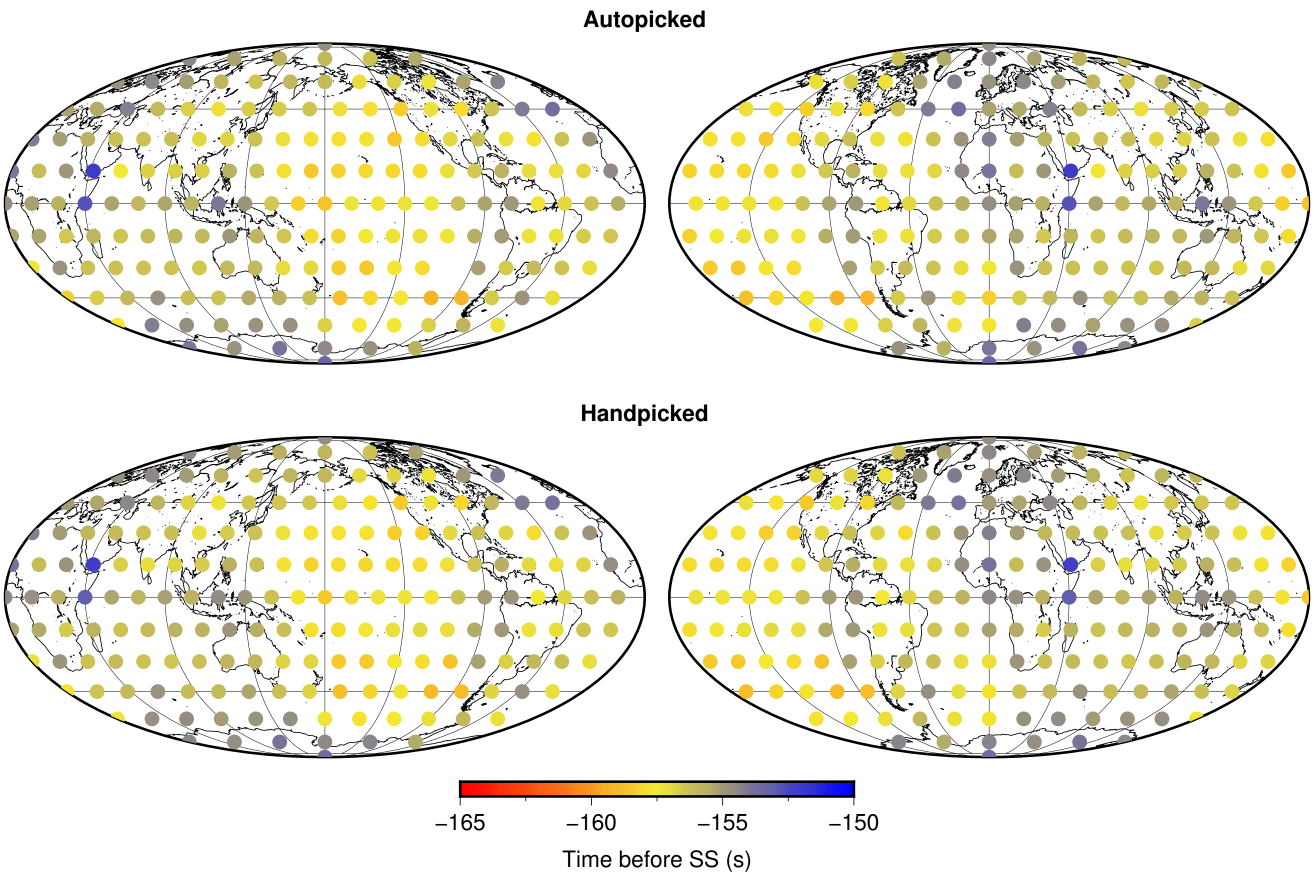

Figure S6. Maps of autopicked (top) and handpicked (bottom) S410S-SS travel time measurements in stacked data, 15° radius caps.

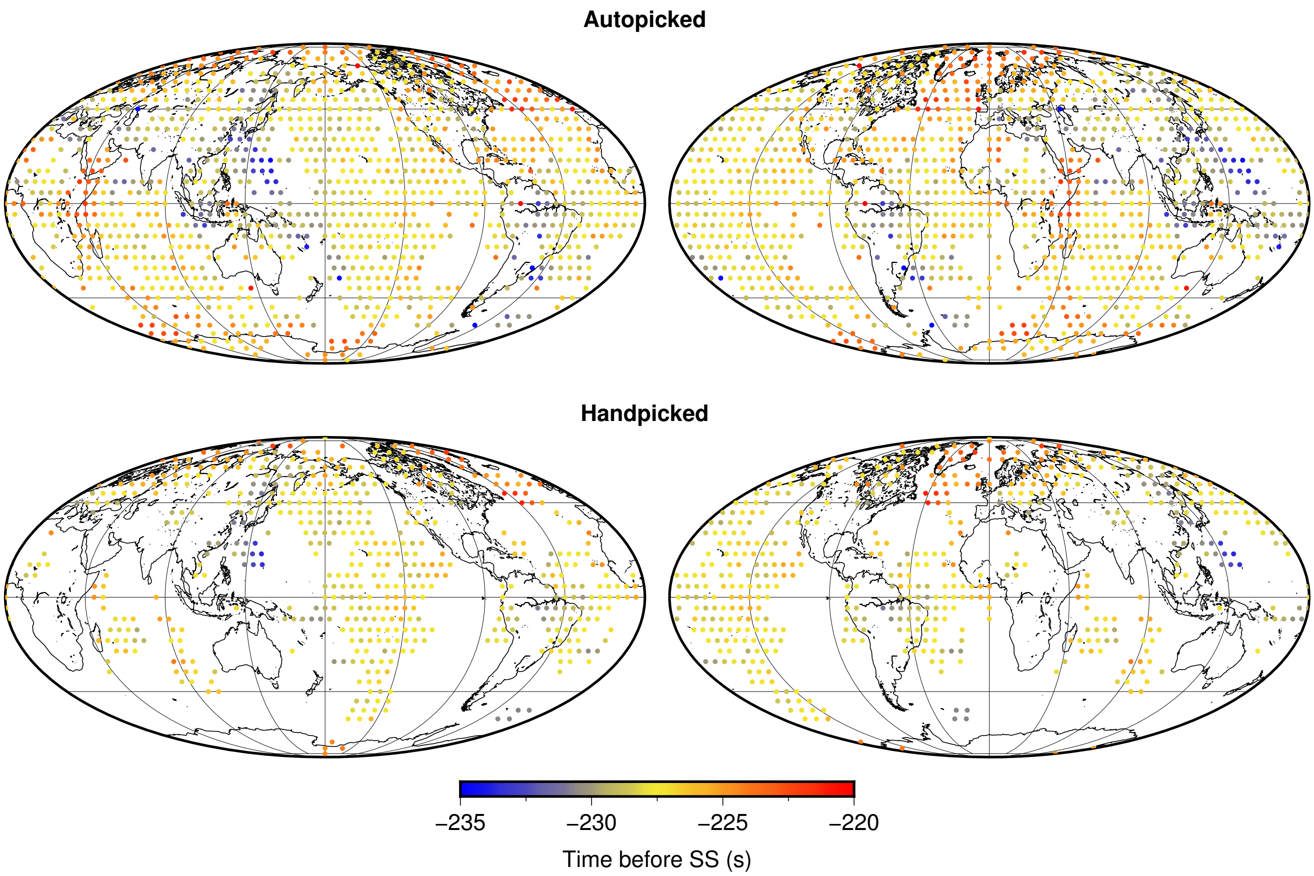

Figure S7. Maps of autopicked (top) and handpicked (bottom) S660S-SS travel time measurements in stacked data, 5° radius caps.

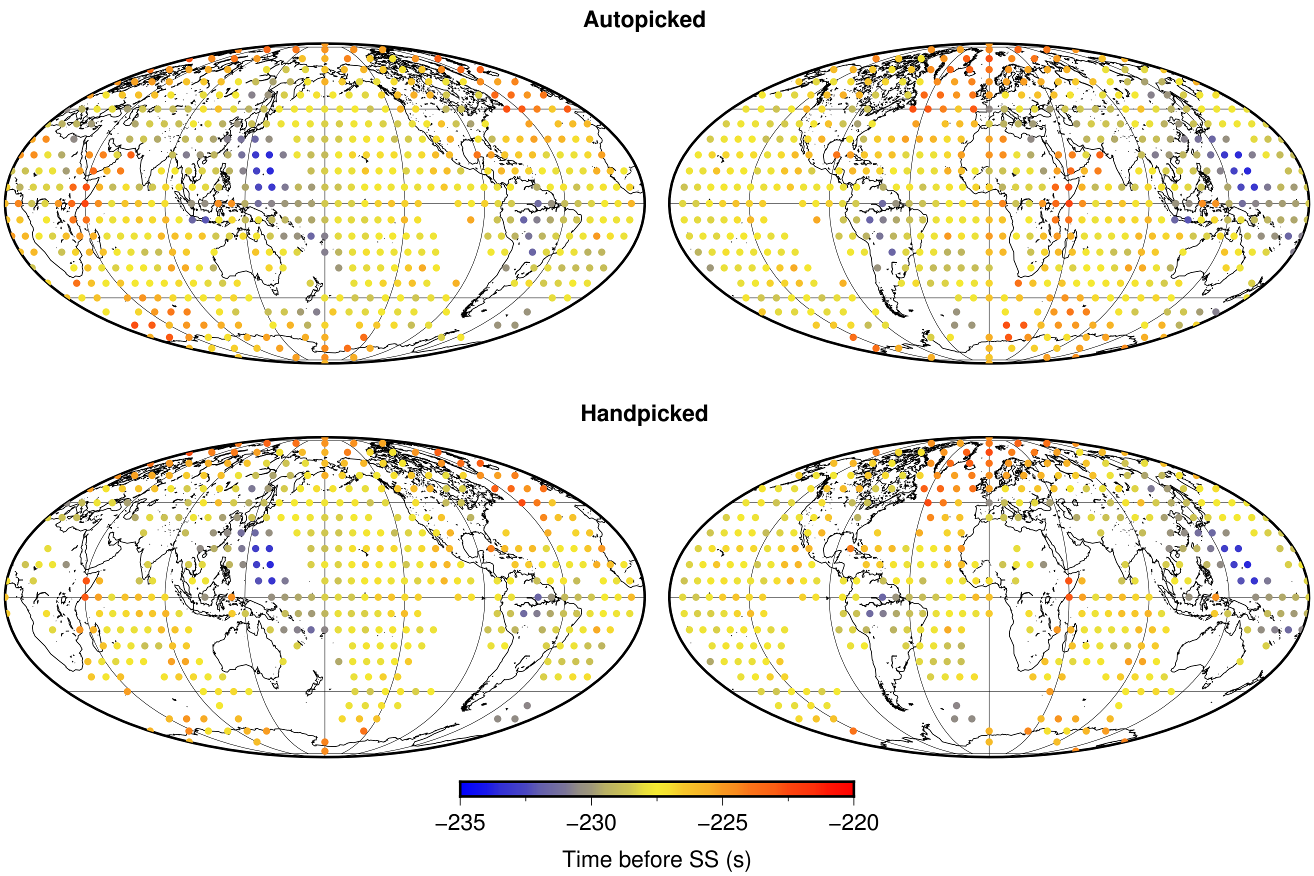

Figure S8. Maps of autopicked (top) and handpicked (bottom) S660S-SS travel time measurements in stacked data, 7.5° radius caps.

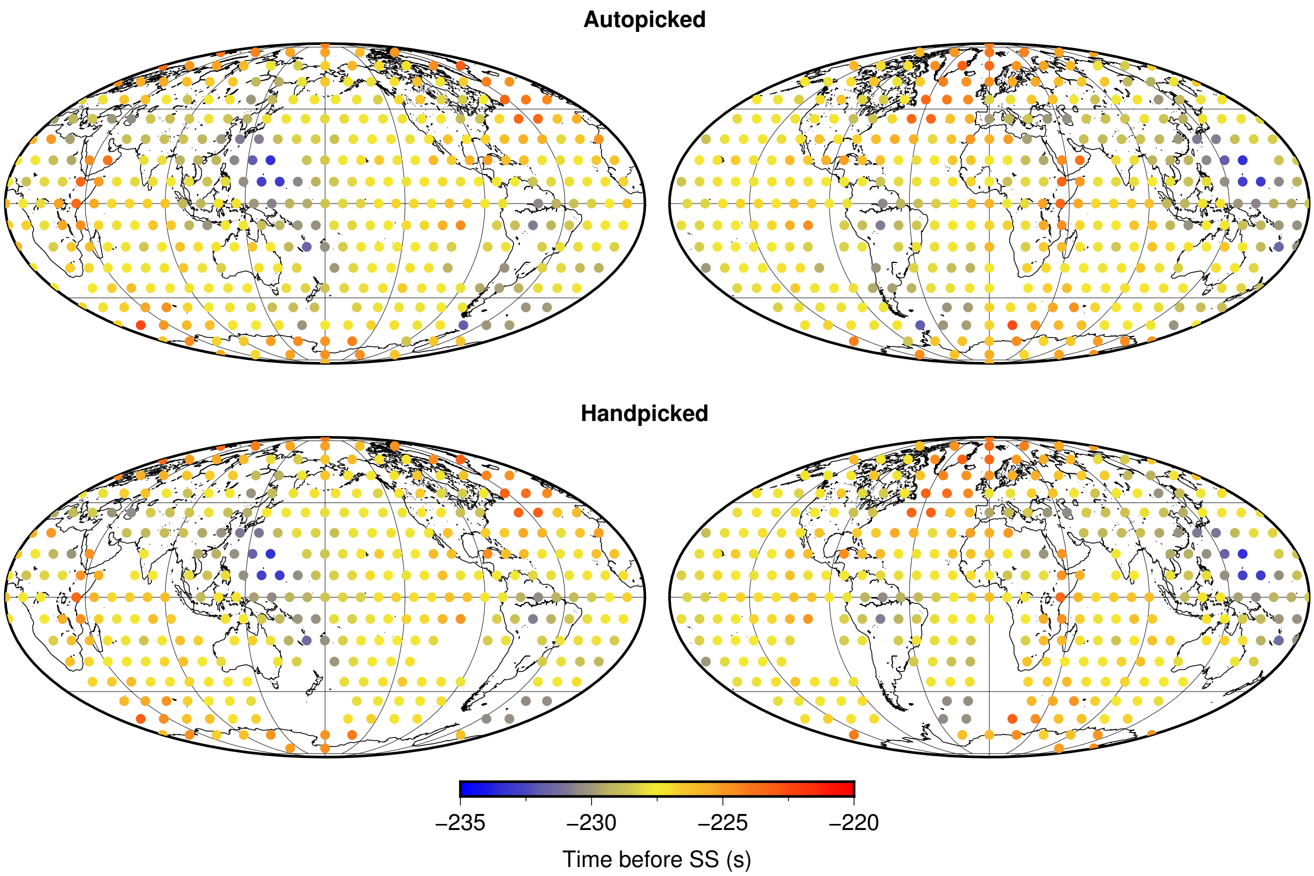

Figure S9. Maps of autopicked (top) and handpicked (bottom) S660S-SS travel time measurements in stacked data, 10° radius caps.

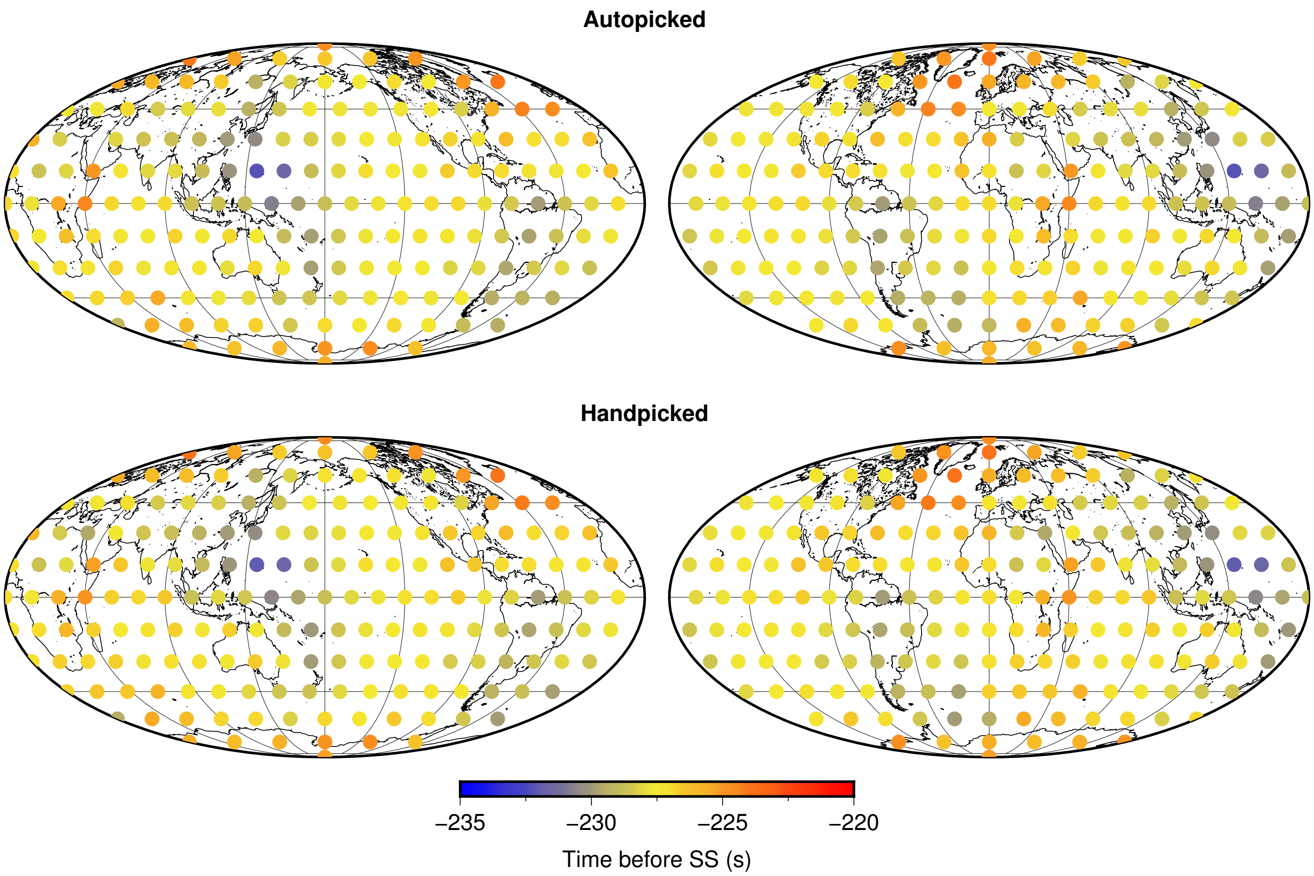

Figure S10. Maps of autopicked (top) and handpicked (bottom) S660S-SS travel time measurements in stacked data, 15° radius caps.

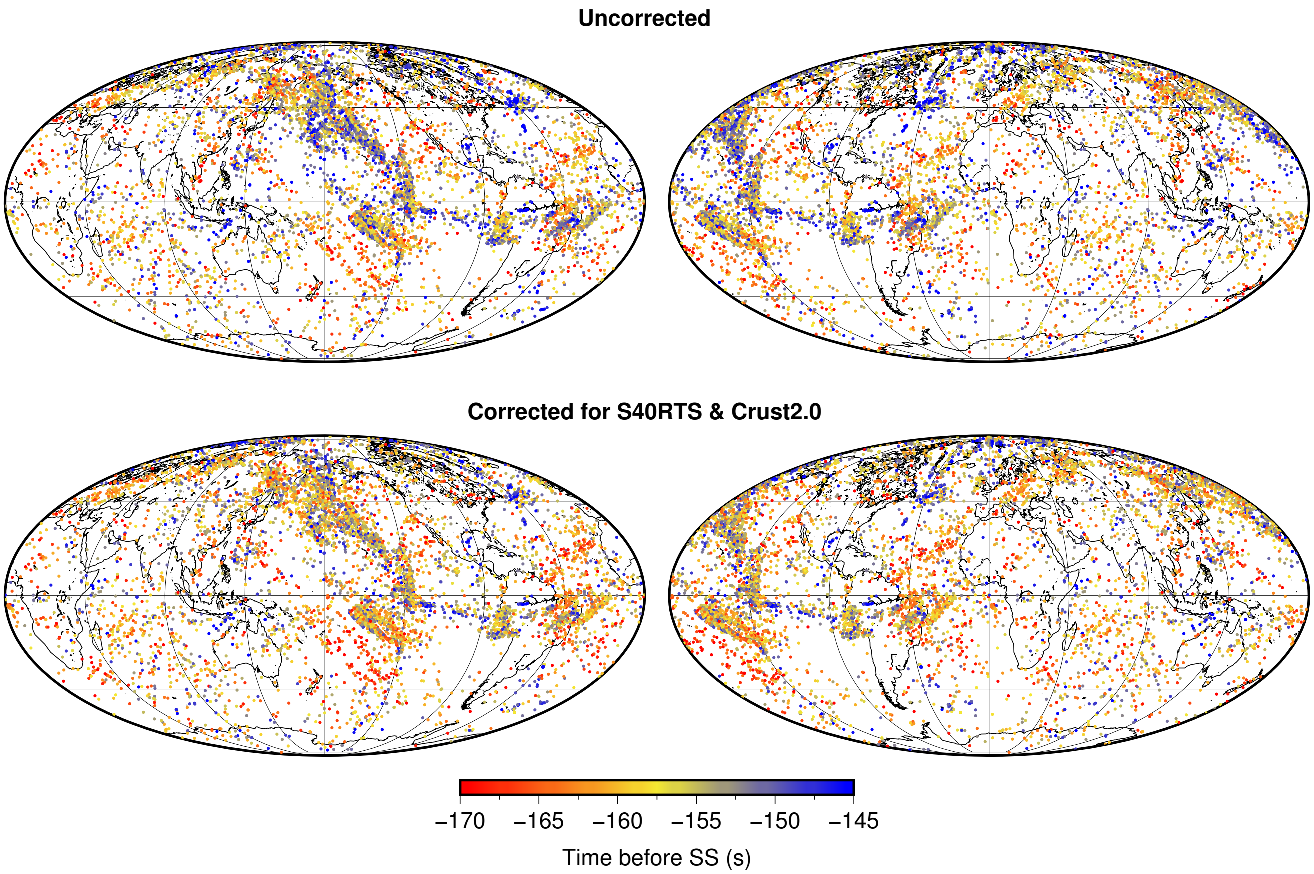

Figure S11. Maps of autopicked S410S-SS travel time measurements in individual seismograms, uncorrected and corrected for S40RTS and Crust2.0.

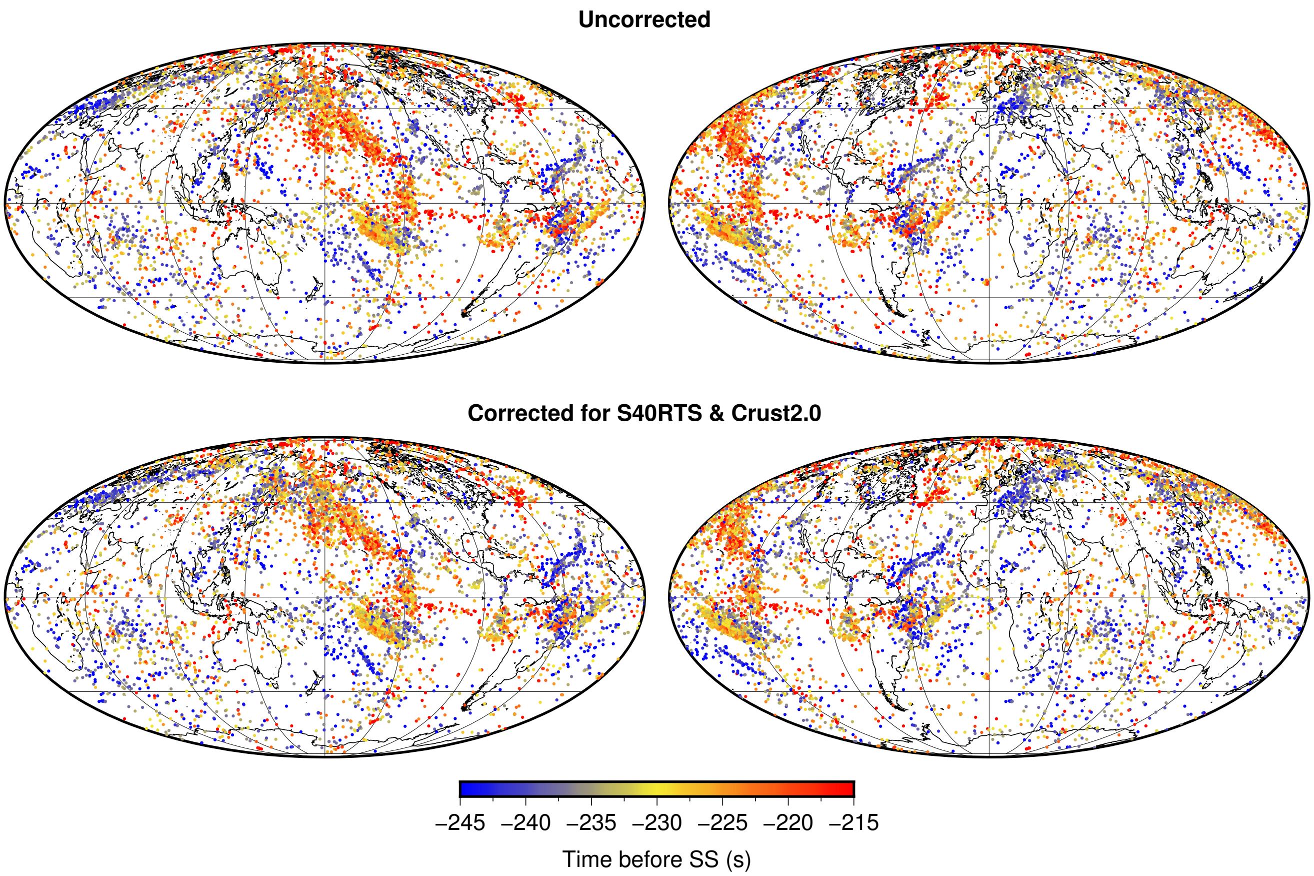

Figure S12. Maps of autopicked S660S-SS travel time measurements in individual seismograms, uncorrected and corrected for S40RTS and Crust2.0.

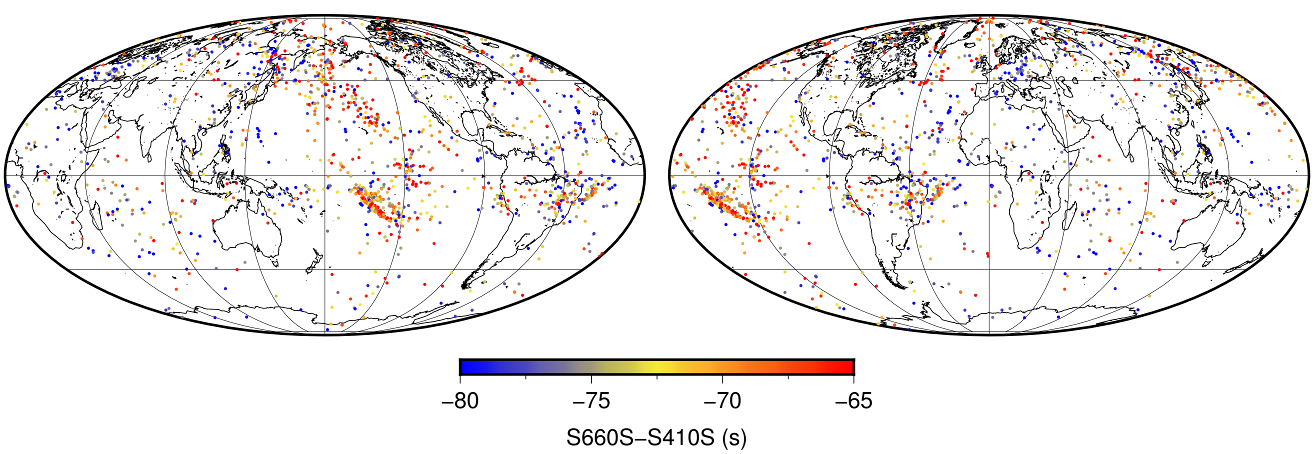

Figure S13. Maps of autopicked S660S-S410S travel time measurements in individual seismograms, corrected for S40RTS and Crust2.0.

Table S1. Average pick quality for stacked data.

Data	Handpicked	Autopicked (all)	Autopicked only
S410S, 5°	0.816	0.773	0.740
S410S, 7.5°	0.816	0.800	0.765
S410S, 10°	0.839	0.829	0.762
S410S, 15°	0.856	0.856	N/A
S660S, 5°	0.881	0.822	0.772
S660S, 7.5°	0.886	0.854	0.771
S660S, 10°	0.895	0.881	0.778
S660S, 15°	0.911	0.911	N/A

Table S2. Average pick quality for stacked data, separated by handpicked quality.

Data	a	b	c	d
S410S, 5°	0.859	0.837	0.803	0.762
S410S, 7.5°	0.878	0.799	0.803	0.773
S410S, 10°	0.874	0.838	0.813	0.787
S410S, 15°	0.885	0.820	0.808	0.730
S660S, 5°	0.919	0.887	0.866	0.863
S660S, 7.5°	0.918	0.900	0.870	0.838
S660S, 10°	0.913	0.906	0.881	0.845
S660S, 15°	0.919	0.908	0.889	0.839

Table S3. Minimum pick quality for stacked autopicked data to achieve the same quantity of picks as handpicked.

Data	S410S quality	S660S quality
5°	0.798	0.860
7.5°	0.698	0.825
10°	0.660	0.773
15°	0.613	0.693