

Automatic seismic waveform identification using a Convolutional Neural Network

J. A. Garcia^{*,1}, L. Waszek^{1,2}, B. Tauzin^{2,3}, N. Schmerr⁴

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

²Research School of Earth Sciences, Australian National University, Acton, ACT 0200, Australia

³Laboratoire de Geologie de Lyon: Terre, Planetes, Environnement, Universite de Lyon, Universite Lyon 1

and Ecole Normale Superieure de Lyon, UMR CNRS 5276, F-69622 Villeurbanne, France

⁴Department 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

48 phase arrival. A more straightforward method allows the computer to perform this task,
49 finding the necessary patterns through representation learning (LeCun et al., 2015). The
50 use of deep learning trains a system that is capable of taking data, identifying charac-
51 terizing features, and producing an informed prediction based on these signatures.

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

62 Here, we apply CNNs to make new observations of mantle discontinuities. The two
63 major global discontinuities at 410 km and 660 km depth (“410”, “660”) bound the man-
64 tle transition zone (“MTZ”). They result from mineral phase transitions in olivine as pres-
65 sure and temperature increase with depth (Katsura & Ito, 1989; Ito & Takahashi, 1989).
66 Due to their opposing Clapeyron slopes, the depths of the discontinuities respond op-
67 positively to temperature. In cold regions the 410 becomes shallower and the 660 becomes
68 deeper; vice versa in hot regions. Consequently, their separation acts as a first order ther-
69 mometer for the MTZ.

70 Mapping of mantle discontinuities globally has been achieved through measuring
71 shear-wave reflections from underneath these boundaries (e.g., P. M. Shearer, 1993; Flana-
72 gan & Shearer, 1998; Houser et al., 2008; Deuss, 2009; Waszek et al., 2018; Huang et al.,
73 2019). SS is a seismic shear wave phase with two legs in the Earth’s mantle and one re-
74 flection from Earth’s surface (Fig. 1a). Reflections from mantle discontinuities gener-
75 ate precursors to SS (“SdS”, where d is discontinuity depth), which arrive prior to the
76 main phase. The SdS-SS travel time difference informs regarding the discontinuity depth.

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

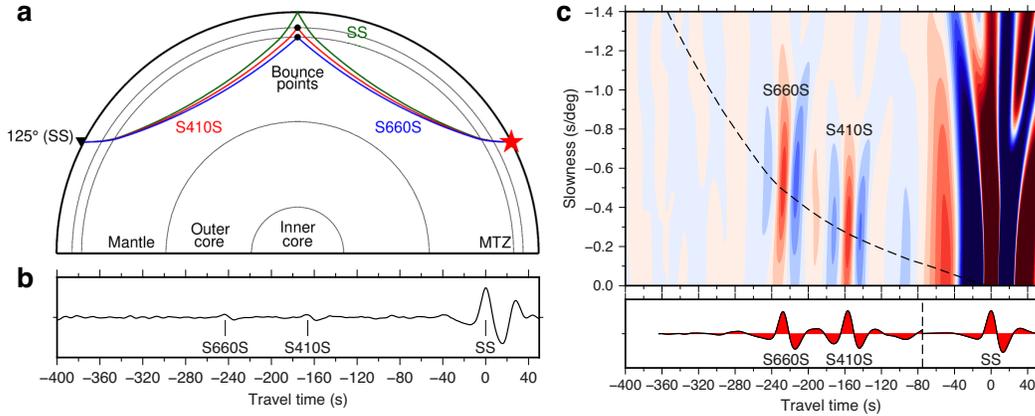


Figure 1. a. Ray paths of SS and its precursors, S410S and S660S. The red star denotes the location of the seismic event, and the black triangle a station to detect seismic waves. b. An example of a high-quality seismogram showing the SS, S410, and S660S arrivals. c. Global vespagram stack for all data 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.

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

87 2 Seismic Data and Processing

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

93 Precursors to SS are typically too small in amplitude to be identified on individ-
 94 ual seismograms. Instead, the data are stacked in regional overlapping spherical caps par-

95 tioned by common reflection points, weighted according to the signal-to-noise ratio, where
 96 noise is the root-mean-square amplitude in the precursor window (-400 to -100 s). Ves-
 97 pograms show stacked signals as a function of travel time and slowness relative to the
 98 main SS arrival (Figure 1c) (Davies et al., 1971). The cross-section taken along the dot-
 99 ted line is the predicted time and slowness of the precursors to SS in a standard refer-
 100 ence model. These vespograms are usually analyzed manually to measure the time and
 101 amplitude of the precursor signals. Bin radii used here are 5° , 7.5° , 10° , and 15° ; these
 102 are selected to account for heterogenous data coverage in different regions: smaller bin
 103 sizes in areas of higher data density to obtain greater resolution.

104 **3 Model Training and Evaluation**

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

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

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

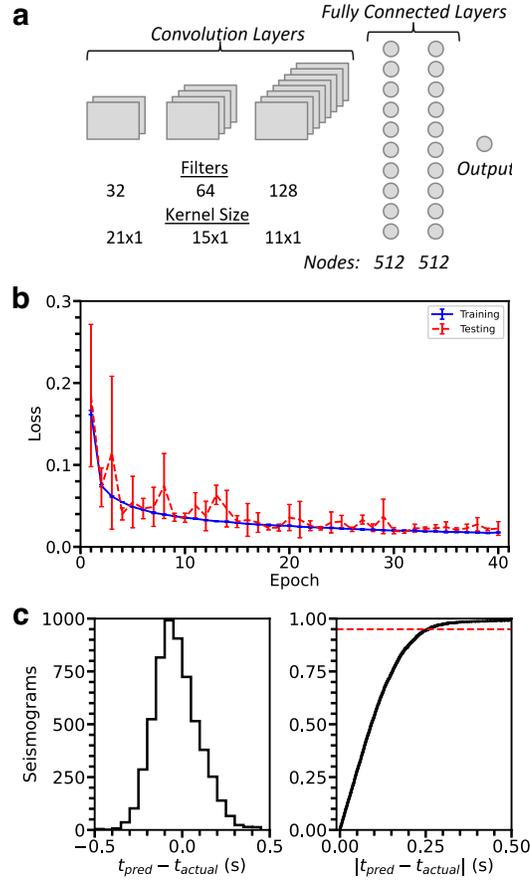


Figure 2. a. Diagram of the “RossNet” architecture used, from Ross et al. (2018). b. Average loss of five models from the architecture used across epochs during the training stage (blue line) and testing stage (red line). These represent the error in fit of each model to the data. The error bars correspond 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.

126 different models for 40 epochs. The models were trained on two NVIDIA Tesla P100 graph-
 127 ics processing units (GPUs); each epoch took approximately three minutes to train.

128 Figure 2b shows the average and standard deviation of the loss over the models and
 129 epochs; this corresponds to the error in fit of the models to the data. Results are shown
 130 for the training and test datasets. Despite the variability of the errors in the testing dataset,
 131 likely due to its relatively small size compared to the training set, there is an overall trend
 132 of decreasing loss as with increasing epoch. We use the best performing set of weights

133 across all 200 instances for the remaining analysis. Summary statistics show an average
 134 prediction error on par with the sampling rate of the seismograms (0.1 s, Fig. 2c). The
 135 cumulative histogram confirms that 95% of the model’s prediction for the test dataset
 136 are within 0.25 seconds of the picked arrival. This size of error is insignificant, since the
 137 picks are subsequently aligned to the maximum amplitude.

138 4 Phase Waveform Quality

139 We desire not only the maximum arrival time of each waveform, but also the qual-
 140 ity of the phase. Quality of the waveform during handpicking is normally judged visu-
 141 ally in a qualitative manner. Here we propose a scheme to assign a quantitative descrip-
 142 tion of the quality using the trained model.

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

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

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

$$q_{\text{pick}} = \frac{N_{\text{pred}}\Delta t}{T} \quad (1)$$

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

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

180 5 Prediction of SS Precursors in Stacked Data

181 Precursors (“SdS”) may be approximated as lower amplitude versions of the main
 182 arrival with a similar shape. Thus, a model trained on the main arrival should be able
 183 to identify precursory signals in stacked waveforms due to their similarity, exploiting the
 184 pattern recognition capabilities of CNNs. We find that our scanning algorithm can in-
 185 deed identify precursors as the highest quality predictions prior to the SS arrival (Fig.
 186 3b; Supplementary Movie 3).

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

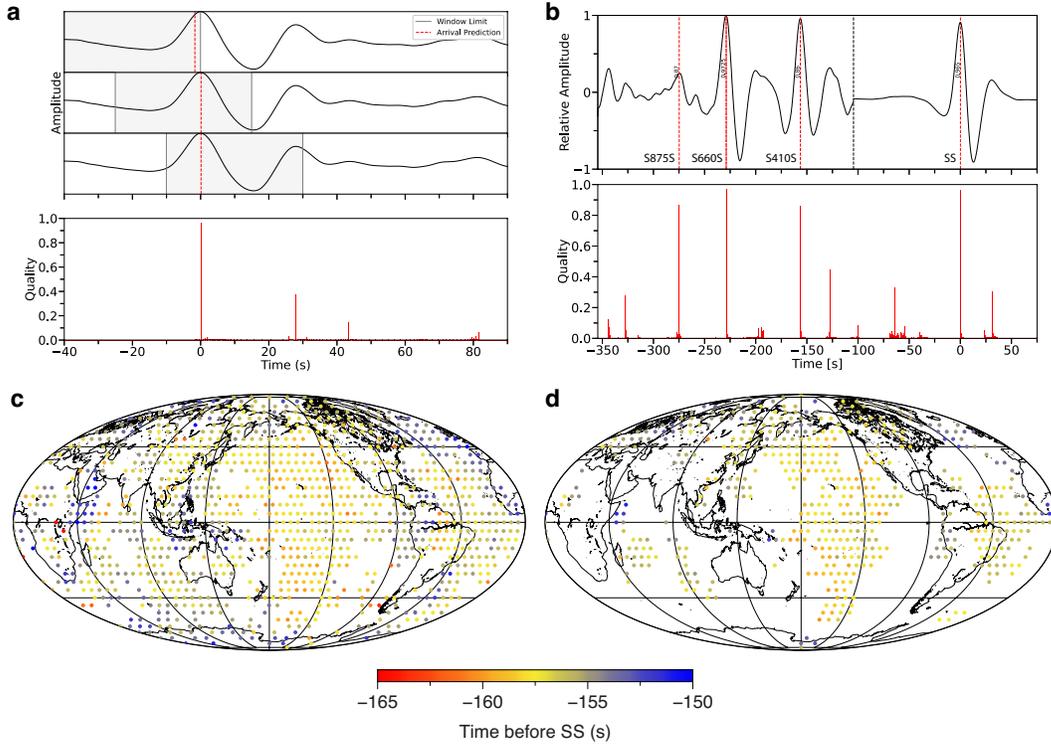


Figure 3. a. Example of iterative prediction for a seismogram, at 0, 15 s, and 30 s, with the histogram of prediction rate. The portion of the waveform within the shaded area is used as input for the model. The red line is the predicted arrival for the given input. The signal is predicted consistently 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.

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

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

216 **6 Prediction of SS Precursors in Individual Seismograms**

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

229 The linear trends for both global discontinuities are calculated using the DBSCAN
 230 algorithm for density-based clustering, to determine statistically the predictions most
 231 likely to correspond to S410S and S660S. We select arrival time bounds of -185 to -135 s
 232 before the main arrival for S410S, and -250 to -200 s for S660S. These are selected to fully
 233 enclose the observed data trends, while excluding theoretical arrival times for other dis-

234 continuities, to ensure that the most dense cluster corresponds to robust picks. A lin-
235 ear fit, with the data weighted by pick quality, is applied as an initial estimate for the
236 trends. Predictions within ± 10 seconds of this fit are considered to also be correct mea-
237 surements for the discontinuity in question. The weighted linear model is then fit to this
238 new set of data points (Fig. 4b). Maps of the uncorrected and corrected relative travel
239 time measurements are included in the Supplement (Fig. S11-S13).

240 7 Discussion

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

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

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

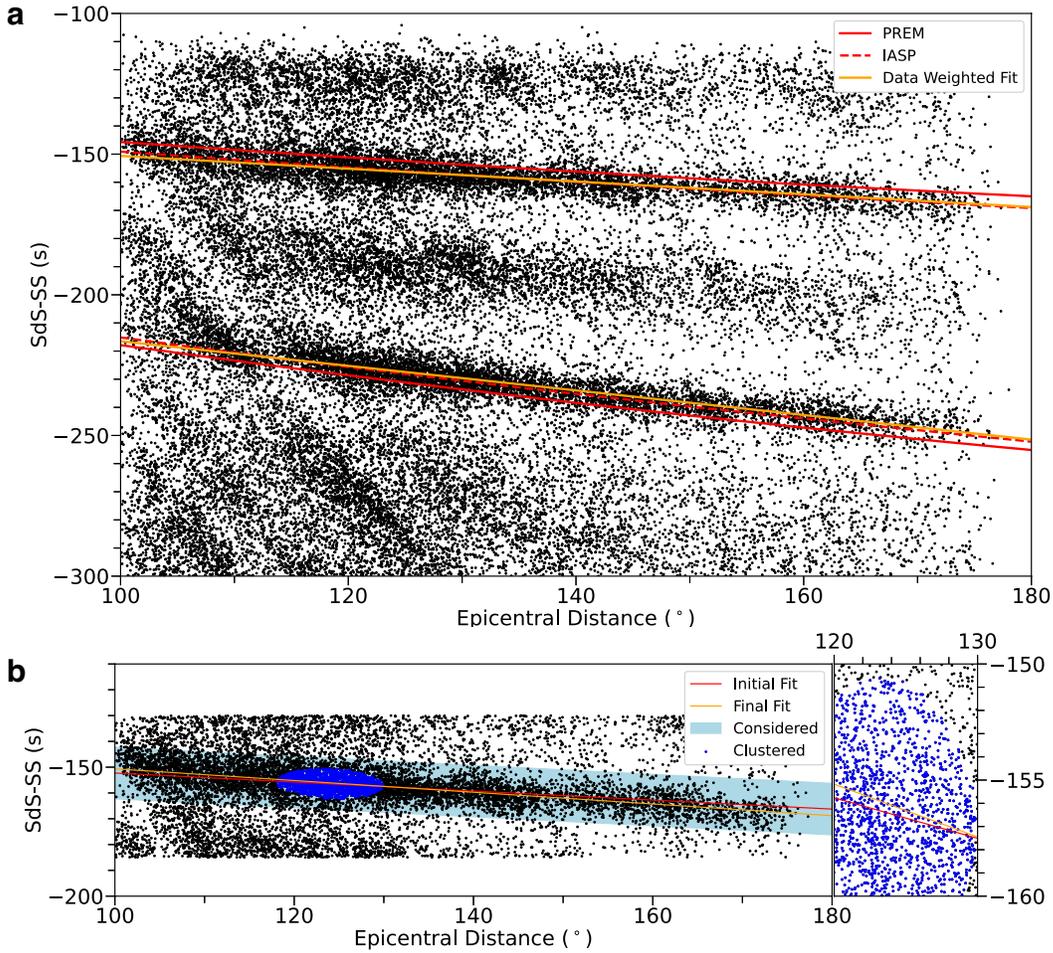


Figure 4. a. Predicted precursor relative arrival time as a function of epicentral distance for the individual seismograms, with theoretical (IASP and PREM) and fitted trends for the S410S and S660S measurements. Picks retained have prediction quality of 60% or higher. Note that these measurements have 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, shown in light blue. A final linear fit is performed on this new set of points to correct the trend, shown in yellow. Notice the small difference between the initial and final linear models.

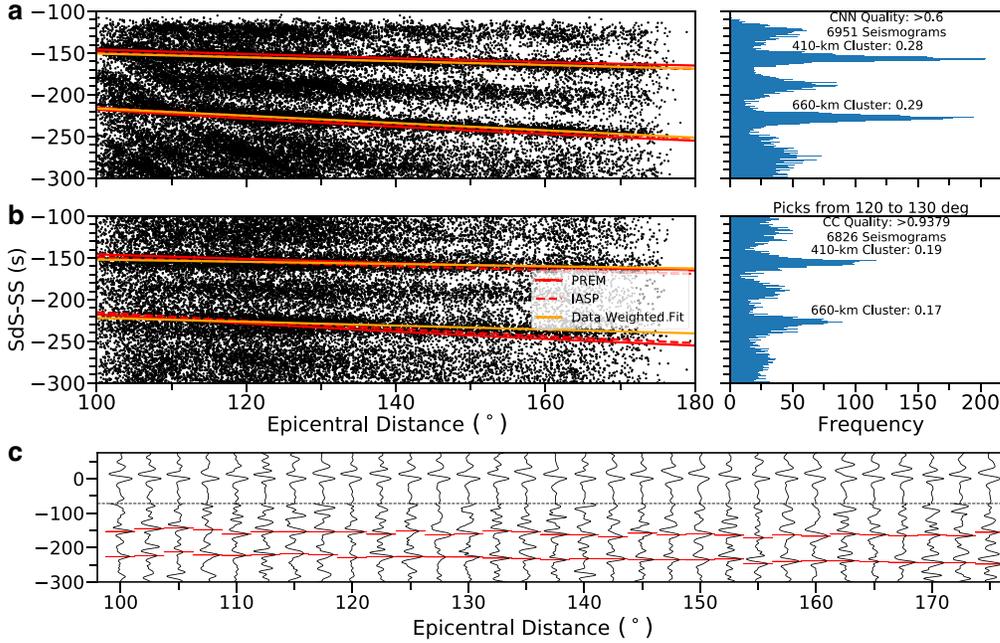


Figure 5. a. Predicted differential time as a function of epicentral distance for individual seismograms, with a histogram of picks between 120 – 130° epicentral distance range, for the CNN models. Lines correspond to predictions from IASP91 (Kennett & Engdahl, 1991) (red dotted), PREM (Dziewonski & 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.

265 however, that the CNN occasionally picks signals that appear to be sidelobes from neg-
 266 ative amplitude interfering phases (Fig. 4a). This is because the code picks the best-matching
 267 signal in a window regardless of shape, relying on the moving window to produce qual-
 268 ity. A future goal is the implementation of a null output. In the meantime, the DBSCAN
 269 clustering analysis could be applied to remove interfering signals and their sidelobes. This
 270 would be particularly useful for mid-mantle precursors which have both positive and neg-
 271 ative polarities (Waszek et al., 2018). The cross-correlation picks do not pick the inter-

272 fering negative signal gap, instead showing significant noise, highlighting its poorer per-
273 formance.

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

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

297 **8 Conclusions**

298 We have demonstrated the significant capabilities of CNNs in the task of picking
299 seismic phases, exploiting the pattern recognition capabilities of these deep learning mod-
300 els. A trained model picks new data accurately and efficiently. It is able to identify other
301 phases with similar features, and extract small-amplitude signals that typically appear
302 masked by noise to the human eye. Thus, a model trained on SS data can produce a dataset

303 of directly measured travel times for SS precursors, providing a new dataset to constrain
 304 Earth's upper mantle. Further consideration of deep learning models and potential ap-
 305 plications to seismology could revolutionize the field by automatically picking waveforms
 306 as they become available. We encourage the use of and welcome contributions to our open-
 307 source Autopicker code.

308 **Acknowledgments**

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

319 **References**

- 320 Bassin, C., Laske, G., & Masters, G. (2000). The current limits of resolution for sur-
 321 face wave tomography in north america. *EOS Trans AGU*, *81*(F897).
- 322 Chambers, K., Deuss, A., & Woodhouse, J. (2005). Reflectivity of the 410-km
 323 discontinuity from pp and ss precursors. *Journal of Geophysical Research*,
 324 *110*(B02301).
- 325 Chollet, F., et al. (2015). *Keras*. <https://keras.io>.
- 326 Davies, D., Kelly, E., & Filson, J. (1971). Vespa process for analysis of seismic sig-
 327 nals. *Nat. Phys. Sci.*, *232*, 8-13.
- 328 Deuss, A. (2009, Oct 01). Global observations of mantle discontinuities us-
 329 ing ss and pp precursors. *Surveys in Geophysics*, *30*(4), 301-326. Re-
 330 trieved from <https://doi.org/10.1007/s10712-009-9078-y> doi:
 331 10.1007/s10712-009-9078-y
- 332 Dziewonski, A. M., & Anderson, D. L. (1981). Preliminary reference earth model.
 333 *Physics of the Earth and Planetary Interiors*, *25*(4), 297 - 356. Retrieved from

- 334 <http://www.sciencedirect.com/science/article/pii/0031920181900467>
335 doi: 10.1016/0031-9201(81)90046-7
- 336 Earle, P., & Shearer, P. (1994). Characterization of global seismograms using an
337 automatic-picking algorithm. *Bulletin of the Seismological Society of America*,
338 *84*(2), 366-376.
- 339 Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for
340 discovering clusters in large spatial databases with noise. In *Kdd*.
- 341 Flanagan, M., & Shearer, P. (1998). Global mapping of topography on transition
342 zone discontinuities by stacking ss precursors. *Journal of Geophysical Research*,
343 *103*(B2), 2673-2692.
- 344 Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies
345 for accurate object detection and semantic segmentation. In *2014 ieee confer-*
346 *ence on computer vision and pattern recognition* (p. 580-587).
- 347 Houser, C., Masters, G., Flanagan, M., & Shearer, P. (2008). Determination and
348 analysis of long-wavelength transition zone structure using ss precursors. *Geo-*
349 *physical Journal International*, *174*(1), 178-194.
- 350 Huang, Q., Schmerr, N., Waszek, L., & Beghein, C. (2019). Constraints on seis-
351 mic anisotropy in the mantle transition zone from long-period ss precur-
352 sors. *Journal of Geophysical Research: Solid Earth*, *124*(7), 6779-6800. doi:
353 10.1029/2019JB017307
- 354 Huber, P. J. (1964, 03). Robust estimation of a location parameter. *Ann. Math.*
355 *Statist.*, *35*(1), 73-101. doi: 10.1214/aoms/1177703732
- 356 Ito, E., & Takahashi, E. (1989). Postspinel transformations in the system
357 Mg₂SiO₄-Fe₂SiO₄ and some geophysical implications. *Journal of Geophys-*
358 *ical Research: Solid Earth*, *94*(B8), 10637-10646. Retrieved from [https://](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB094iB08p10637)
359 agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB094iB08p10637
360 doi: 10.1029/JB094iB08p10637
- 361 Katsura, T., & Ito, E. (1989). The system Mg₂SiO₄-Fe₂SiO₄ at high pressures and
362 temperatures: Precise determination of stabilities of olivine, modified spinel,
363 and spinel. *Journal of Geophysical Research: Solid Earth*, *94*(B11), 15663-
364 15670. Retrieved from [https://agupubs.onlinelibrary.wiley.com/doi/](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB094iB11p15663)
365 [abs/10.1029/JB094iB11p15663](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB094iB11p15663) doi: 10.1029/JB094iB11p15663
- 366 Kennett, B., & Engdahl, E. (1991). Travel times for global earthquake location and

- 367 phase association. *Geophysical Journal International*, *105*, 429-465.
- 368 Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*.
- 369 Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017, May). Imagenet classification
370 with deep convolutional neural networks. *Commun. ACM*, *60*(6), 8490. Re-
371 trieved from <https://doi.org/10.1145/3065386> doi: 10.1145/3065386
- 372 LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-
373 444. doi: 10.1038/nature14539
- 374 McBrearty, I. W., Delorey, A. A., & Johnson, P. A. (2019, 01). Pairwise Associa-
375 tion of Seismic Arrivals with Convolutional Neural Networks. *Seismological Re-*
376 *search Letters*, *90*(2A), 503-509. Retrieved from [https://doi.org/10.1785/](https://doi.org/10.1785/0220180326)
377 [0220180326](https://doi.org/10.1785/0220180326) doi: 10.1785/0220180326
- 378 Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltz-
379 mann machines. In *Icml*.
- 380 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ...
381 Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of*
382 *Machine Learning Research*, *12*, 2825-2830.
- 383 Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for
384 earthquake detection and location. *Science Advances*, *4*(2). Retrieved from
385 <https://advances.sciencemag.org/content/4/2/e1700578> doi: 10.1126/
386 [sciadv.1700578](https://advances.sciencemag.org/content/4/2/e1700578)
- 387 Ritsema, J., Deuss, A., van Heijst, H. J., & Woodhouse, J. H. (2011). S40rts:
388 a degree-40 shear-velocity model for the mantle from new rayleigh wave
389 dispersion, teleseismic travelttime and normal-mode splitting function mea-
390 surements. *Geophysical Journal International*, *184*(3), 1223-1236. Re-
391 trieved from [https://onlinelibrary.wiley.com/doi/abs/10.1111/](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.2010.04884.x)
392 [j.1365-246X.2010.04884.x](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.2010.04884.x) doi: 10.1111/j.1365-246X.2010.04884.x
- 393 Ross, Z. E., Meier, M., & Hauksson, E. (2018). P wave arrival picking and first-
394 motion polarity determination with deep learning. *Journal of Geophysical Re-*
395 *search: Solid Earth*, *123*(6), 5120-5129. doi: 10.1029/2017JB015251
- 396 Schmerr, N., & Garnero, E. (2006). Investigation of upper mantle discontinuity
397 structure beneath the central pacific using ss precursors. *Journal of Geophysi-*
398 *cal Research: Solid Earth*, *111*(B8). doi: 10.1029/2005JB004197
- 399 Shearer, P. (1991). Constraints on upper mantle discontinuities from observations

- 400 of long-period reflected and converted phases. *Journal of Geophysical Research:*
401 *Solid Earth*, 96(B11), 18147-18182. doi: 10.1029/91JB01592
- 402 Shearer, P. M. (1993, 12). Global mapping of upper mantle reflectors from long-
403 period SS precursors. *Geophysical Journal International*, 115(3), 878-904. Re-
404 trieved from <https://doi.org/10.1111/j.1365-246X.1993.tb01499.x> doi:
405 10.1111/j.1365-246X.1993.tb01499.x
- 406 Simonyan, K., & Zisserman, A. (2014, 09). Very deep convolutional networks for
407 large-scale image recognition. *arXiv 1409.1556*.
- 408 Waszek, L., Schmerr, N. C., & Ballmer, M. D. (2018). Global observations of reflec-
409 tors in the mid-mantle with implications for mantle structure and dynamics.
410 *Nature Communications*, 9(1), 385. doi: 10.1038/s41467-017-02709-4
- 411 Waszek, L., Tauzin, B., Schmerr, N. C., Ballmer, M. D., & Afonso, J. C. (2020). A
412 poorly mixed mantle and its thermal state inferred from seismic waves. *In re-*
413 *view*.
- 414 Zhu, W., & Beroza, G. C. (2018, 10). PhaseNet: a deep-neural-network-based seis-
415 mic arrival-time picking method. *Geophysical Journal International*, 216(1),
416 261-273. Retrieved from <https://doi.org/10.1093/gji/ggy423> doi: 10
417 .1093/gji/ggy423
- 418 Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and
419 decomposition using deep neural networks. *IEEE Transactions on Geoscience*
420 *and Remote Sensing*, 57(11), 9476-9488.