Deep-learning phase-onset picker for deep Earth seismology: PKIKP waves

Jiarun Zhou, Thanh-Son Phạm and Hrvoje Tkalčić

Corresponding author: J. Zhou (jiarun.zhou@anu.edu.au)

Research School of Earth Science, The Australian National University, Canberra ACT, Australia

Key points:

- We employed a Convolutional Neural Network to automatically pick the onsets of inner-core sensitive PKIKP waves.
- Our automatic picker approaches near human-level precision and initially reproduces established inner-core anisotropic models.
- Automatic phase onset measurements would open unprecedented avenues for studying the Earth's deep interior, including the inner core.

Key words: Automatic phase picking; Earth’s deep interior; Inner core; Convolutional Neural Network

Abstract

Body waves traversing the Earth's interior from a seismic source to receivers on the surface carry rich information about its internal structures. Their travel time measurements have been widely used in seismology to constrain Earth's interior at the global scale by mapping the time anomaly along their ray paths. However, picking the travel time of global seismic waves, suitable for studying Earth's fine-scale structures, requires highly skilled personnel and is often fairly subjective. Here, we report the development of an automatic picker for PKIKP waves, traversing the Earth nearly along its diameters and through the inner core, based on the latest advances in supervised deep learning. A convolutional neural network (CNN) we developed automatically determines the PKIKP onset on vertical seismograms near its theoretical prediction of cataloged earthquakes. As high-quality manual onset picks of global seismic phases are limited, we employed a scheme to generate a synthetic supervised training dataset containing 300,000 waveforms. The PKIKP onsets picked by our trained CNN automatic picker exhibit a mean absolute error of ~0.5 s compared to 1,503 manual picks, comparable to the estimated human-picking error. In an integration test, the CNN automatic picks obtained from an extended waveform dataset yield a cylindrically anisotropic inner core model that agrees well with the models inferred from manual picks, which illustrates the success of this pilot model. This is a significant step closer to harvesting an unprecedented volume of travel time measurements for studying the inner core or other regions of the Earth's deep interior.

Plain language summary

Seismic body waves traversing the Earth's interior provide critical constraints on structures and dynamics of the Earth's deep interior, including the solid inner core. The onset time of compressional waves from a seismic source passing through the inner core, known as PKIKP waves, has been widely used to study the Earth's deepest shell. However, the collection of manual onset time picks meticulously analyzed by experienced analysts is scarce because it is laborious. At the same time, large collections compiled by multiple data centers, such as the International Seismological Center's PKIKP onset dataset, are inhomogeneous and deemed less reliable for inner-core research than the data collected by individual researchers. Here, we develop an automatic PKIKP onset picker based on recent advancements in machine learning in computer vision, the Convolutional Neural network (CNN). We trained the network with synthetic waveforms, mimicking the influence of the Earth's structure on the initial waveform shape. Our comprehensive tests benchmark the consistency between the automatically picked and the researcher-examined datasets. The automatic picker enables the further exploration of the vast seismic archive for unprecedentedly large datasets devoted to the study of the Earth's deep interior with greater details.

1 Introduction

Seismic energy radiated by earthquakes travels through the Earth's interior, carrying information to seismic receivers on the Earth's surfaces. Seismological tools for studying the Earth's deep interior can be broadly categorized into three main groups, depending on their frequency characteristics: body waves, global coda-correlation wavefield, and normal modes. Teleseismic body waves are high-frequency seismic signals (periods from around 0.1 to ~10 seconds) whose sensitivity to the structures can be mapped along their ray paths, thanks to the infinite frequency approximation (Bullen, 1961; Kennett, 2009; Kennett et al., 1995; Shearer, 2019). There are numerous applications of the methods based on body waves, from discovering the main boundaries within the Earth (Gutenberg, 1914; Inge Lehmann, 1936; Mohorovičić, 1910) to constructing 1D or 3D Earth models (Aki & Lee, 1976; Kennett et al., 1995; Obayashi et al., 2013). The normal mode data, or spectra, of long-period standing waves excited by large earthquakes at hundreds of seconds, are important to constrain larger-scale structures (Dahlen & Tromp, 1998). Normal modes have also been used to construct 1D models of the Earth (Dziewonski & Anderson, 1981) and illuminate the IC (e.g., Deuss et al., 2010; Romanowicz & Bréger, 2000; Woodhouse et al., 1986). Correlation wavefield is an emerging concept that uses features formed by the similarity between weak seismic signals at mid-range periods (~ tens of seconds) (Tkalčić et al., 2020). It has been used to constrain a new 1D Earth model (Ma & Tkalčić, 2021) and study the Earth's IC (e.g., Costa de Lima et al., 2022; Tkalčić & Phạm, 2018; Wang & Tkalčić, 2021).

As the Earth's deepest and most mysterious layer, the Earth's IC plays an essential role in the Earth's dynamics and geomagnetic field (Tkalčić, 2017). In studying the IC and the Earth's deep interior, utilizing body waves, with their high-frequency nature, is unparalleled among seismological tools. Body wave travel times are the most commonly measured property because the data can be simulated efficiently using the ray theory, yet full waveform simulation is computationally expensive. The IC is challenging to study because of the limited sampling coverage and data quality of IC-sensitive waves, such as PKIKP traversing the IC and PKiKP reflecting off the inner core boundary (ICB) (**Figure 1**). The expansion of global seismic networks brings a valuable opportunity to place additional constraints and details on the Earth's deep interior, including the IC, via array-based observations of exotic seismic phases (e.g., Burdick et al., 2019; Phạm & Tkalčić, 2023; Waszek & Deuss, 2015).

However, collecting the arrival onsets of body waves for global studies is laborious. The largest collection of manually picked onsets is available at the International Seismological

Centre (ISC; Bondár & Storchak, 2011), where seismological agencies worldwide report their waveform data and time picks of the main seismic phases. Although several researchers have used the mass dataset to study IC deep structures in the early years (e.g., Ishii & Dziewoński, 2002; Shearer, 1994), the ISC dataset (www.isc.ac.uk/iscbulletin/search/arrivals/) has often been criticized for significant data scatter because the onset picks are performed by multiple analysts with few measures to ensure picking qualities across the data centers (Stephenson et al., 2021; Su & Dziewonski, 1995). Deep-Earth seismologists often need to manually pick a small portion of available waveforms using their methods to ensure the homogeneity of the input data. However, due to the large volume of seismic data, it is challenging to hand-pick all waveform datasets consistently. Thus, developing automatic tools is crucial in advancing studies of Earth's deep interior at the global scale. This will not only ensure consistency in the onset picks but also harness the full capacity of the global seismic network.

Automatic determination of seismic arrivals for local, small earthquakes has been developed with the short-temporal average over long-temporal average (STA/LTA)-type algorithms for several decades (e.g., Allen, 1982; Baer & Kradolfer, 1987; Hildyard et al., 2008; Sleeman & van Eck, 1999). The last few years have witnessed the emergence of machine learning algorithms to support automatic data processing as a more reliable technique, with many recent breakthroughs. Recent advances in deep learning algorithms include Convolutional Neural Networks (CNNs) in computer vision (LeCun et al., 2015) and Transformers in natural language processing (Vaswani et al., 2017). They have been widely deployed to detect local microearthquakes and discriminate seismic signals (S. M. Mousavi et al., 2020; Saad et al., 2022; Zhu et al., 2019), determine first-motion polarity (Ross, Meier, & Hauksson, 2018), and pick the onsets of local P and S waves (e.g., S. M. Mousavi et al., 2020; Ross, Meier, Hauksson, et al., 2018; Zhu & Beroza, 2019). It is broadly agreed that deep learning can achieve higher accuracy and precision than engineered automatic picking based on the STA/LTA approach (S. M. Mousavi et al., 2020; Ross, Meier, & Hauksson, 2018; Zhu & Beroza, 2019).

The strategies and purposes of automatic picking are different on local and global scales. Ultimately, local microearthquake pickers scan through continuous seismic waveforms to detect earthquakes by determining their phase arrivals, such as P or S waves. On a global scale, earthquakes of global significance are routinely documented and reported in earthquake catalogs such as the Global Centroid Moment Tensor (GCMT; Dziewonski et al., 1981; Ekström et al., 2005), the U.S. Geological Survey National Earthquake Information Center

(USGS NEIC; Guy et al., 2015), or the ISC (Bondár & Storchak, 2011). We can predict the arrivals of their seismic phases utilizing existing 1D Earth models, such as PREM (Dziewonski & Anderson, 1981), CCREM (Ma & Tkalčić, 2021), IASP91 (Kennett & Engdahl, 1991), or ak135 (Kennett et al., 1995). Therefore, the main task of an automatic global picker is to pick the actual onsets of seismic phases precisely above the background noise level in pre-windowed waveforms based on theoretical predictions.

Despite the successful application of automatic local seismic phase pickers in communities targeting shallow Earth structures, it slowly makes its way into global earthquakes and the study of deep Earth interior, perhaps due to the community's relatively small size. However, the benefit of having an automatic tool for data collection is significant. In recent years, deep neural networks have been used to detect PmKP waves reflecting multiple times at the core-mantle boundary (Dong et al., 2024), and SS signals used to study upper mantle structures (Garcia et al., 2021). This paper presents a pivotal effort to develop a tool for picking IC-sensitive PKIKP onsets from global seismic networks. It demonstrates the need for high-quality training datasets and provides new datasets of absolute PKIKP wave travel times for studying the Earth's IC.

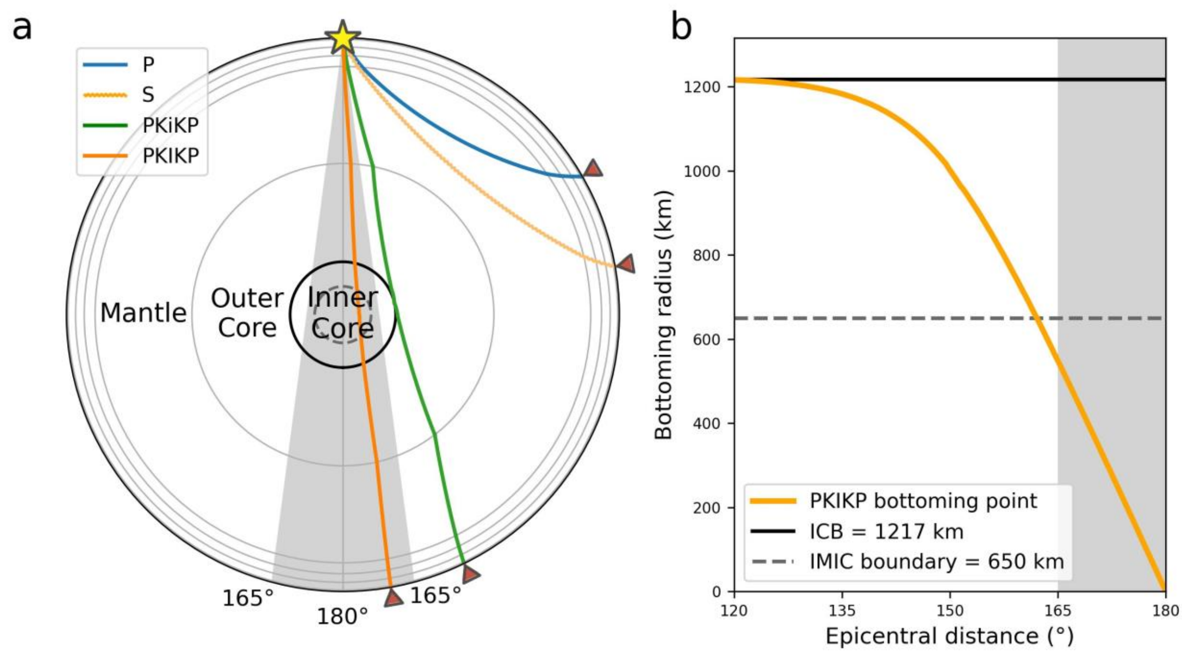

Figure 1 Illustration of the ray paths of IC-sensitive waves. *In (a), the orange line denotes the PKIKP ray path traversing the inner core (IC), and the green line denotes the PKiKP ray path reflected off the inner core boundary (ICB). The star and triangles denote the location of the earthquake and the receiving stations, respectively. The ray paths of P and S waves are*

also denoted with blue and light orange lines. (b) bottoming radius of PKIKP waves relative to the epicentral distances. The bottom radius = 0 indicates the center of the Earth. The solid and dashed lines denote the radii of the ICB and the innermost inner core (IMIC) transition (Phạm & Tkalčić, 2023; Stephenson et al., 2021), respectively. The shaded areas (165° - 180°) in a) and b) denote the epicentral range of the PKIKP dataset in this study.

2 Methods

2.1 PKIKP arrival dataset with manual picks

We employ 8208 PKIKP arrival records of 419 globally distributed earthquakes (**Figure 3**) between March 1990 – June 2019. Most events have a magnitude range of 7.5 ≥ Mw ≥ 5.7 to avoid introducing uncertainties from too large or too small earthquakes in picking PKIKP onsets. All PKIKP records are collected from a near-antipodal distance (>165°) to place additional constraints on the Earth's innermost inner core (IMIC) (**Figure 1**) and sampled at 40 Hz. In automatic picking by the deep-learning network, the input waveforms are filtered between 0.5–2.0 Hz, a common frequency band visualizing most PKIKP arrivals in our dataset. Tkalčić et al. (2023) presented 1503 manual onset picks of PKIKP arrivals from this dataset. They did not analyze all events along the quasi-equatorial paths due to source-receiver ray path geometry saturation. The hand-picked waveform dataset size is much smaller than the typical size to train a deep learning network in seismological applications (**Figure 2**). Consequently, the picked waveforms (or labeled waveforms in deep-learning terms) are used to test the performance of our trained CNN only. The rest of the unpicked waveforms are also used for an integrated test of the network performance in Section 3.2.

The manual determination of PKIKP onsets involves the collection of seismic waveforms around the predicted arrivals based on the Earth models (Tkalčić et al., 2023). Firstly, 60-second-long waveforms around the arrivals are visually inspected to determine if the anticipated seismic phase is visible and the time of its onset. Several filter bands assist in visually determining arrival onsets when unfiltered waveforms are unclear. To ensure the signal is from a teleseismic event, waveforms from multiple stations recording the expected PKIKP phase are visualized in a gathered plot. The human picker may also analyze a large number of recordings from a single event, sort them as functions of epicentral distance or back azimuth, and compare PKIKP with waveforms of other PKP branches, PKPbc and PKPab to recognize the PKIKP arrivals. The waveforms of P waves recorded at epicentral distances between 30° and 90° but falling within the same azimuthal corridor to the PKIKP waves around 10° are also of interest because they inform the human picker of the earthquake's source time function, i.e., the emergence of the arrival. In general, the error of manual picking of PKIKP absolute travel times, often due to the emergent onsets of the PKIKP waves, the attenuation by the IC structure, or due to waveform change introduced through bandpass filtering, is around 0.5 s (Tkalčić et al., 2023).

On the contrary, in the automatic picking of the CNN network, we applied a single filter of 0.5–2.0 Hz. 50-s waveform segments around ak135 predicted PKIKP onsets are used as the network input. The automatic picker does not accept any supplementary information used in manual picking.

2.2 Design of synthetic training datasets

Our automatic picker is based on a supervised learning model (Rumelhart et al., 1986), which longs for a large, labeled dataset to determine an object's features. **Figure 2a** compares the training dataset sizes of several local phase pickers. Due to the lack of high-quality labeled seismic data corresponding to the Earth's IC, we adopted a synthetically generated training dataset. This approach has been widely used in deep-learning seismology, mostly in denoising and interpreting seismic imaging (S. M. Mousavi & Beroza, 2022).

Following the procedures earlier introduced by Phạm and Tkalčić (2017), we generated synthetic waveforms to simulate the key properties of antipodal PKIKP arrivals. Firstly, Telewavesim (Audet et al., 2019), a Python package implementing the matrix propagation method (Kennett, 2009), generates the local crustal responses of teleseismic plane wave arrival. The structural responses are then convolved with randomized source time functions, which are lowpass filtered at a random corner frequency between 0.2 and 1.5 Hz. The choice of the corner frequency simulates the Earth's natural filtering effects due to its attenuation, including the IC effects, of the global phase. The local crustal conditions are randomly selected from the global CRUST1.0 model (Laske et al., 2013) to enhance the generated waveform's variability. For each waveform generated, its label indicating the first phase onset is calculated analytically. Thus, the sets of labeled waveforms can be synthesized at large volumes.

Furthermore, real noise is added to the synthesized waveforms to simulate the ambient noise conditions of PKIKP waves. We take noise segments 5 s before P arrivals from real P seismograms collected from globally distributed seismic stations in epicentral distances of

60° – 90° and within 10° deviation of azimuth with respect to the stations used to collect the PKIKP dataset. Both synthetic waveforms and noise records are normalized to their maximum amplitudes. The noise amplitude is scaled by multiplying a randomized ratio in the normal distribution of 0.2 standard deviation. Noise waveforms are also used in model training as an individual set to make the CNN model able to distinguish seismic signals from ambient noise. Such generated datasets form our so-called "regular P-wave" dataset and "noise" dataset (**Figure 2b**).

Synthetic waveforms generated by the scheme above consider the response of crustal variability and introduce the attenuation from the IC implicitly through the choice between 0.2 to 1.5 Hz corner frequencies for the synthetic source time functions, but they are still inadequate. The "regular P-wave" training set often has sharp onsets, so an automatic picker trained on it tends to pick the most significant onset of the arrivals. However, actual PKIKP arrivals might have an emergent onset due to the gradual release of earthquake energy. Experienced analysts can realize the pattern by learning the accompanying P-arrival waveform shapes (Tkalčić et al., 2023). For the automatic picker, we explicitly generated an additional dataset of emergent arrival for training, called the "targeted P-wave" dataset. A random amplitude ratio between 0.2 – 0.5 is multiplied by the first 4-s of the originally synthesized P arrival to create the emergent onset. The targeted P set is the key to improving the precision of the CNN in picking PKIKP onsets because enhanced learning on the targeted P dataset enables our picker to identify emergent arrivals from noise and return PKIKP onset picks with lower error (see demonstration in Section 3.1).

In summary, our training dataset consists of three subsets: a regular set of synthetic P waveforms, a real noise set, and a targeted P-wave set (**Figure 2b**), each containing 100K 20-s waveform samples to make the total data amount comparable to the training datasets used in other studies (**Figure 2a**). We demonstrate the influence of each dataset on onset picking later in Section 3.1. In the training dataset, we randomly cut a 20-s segment around the onset label from each synthetic waveform so that the onset can be at any location within the window. Thus, labels for positive signals are linear P phase onsets between 0 and 20, and the pure noise record labels are always 0.

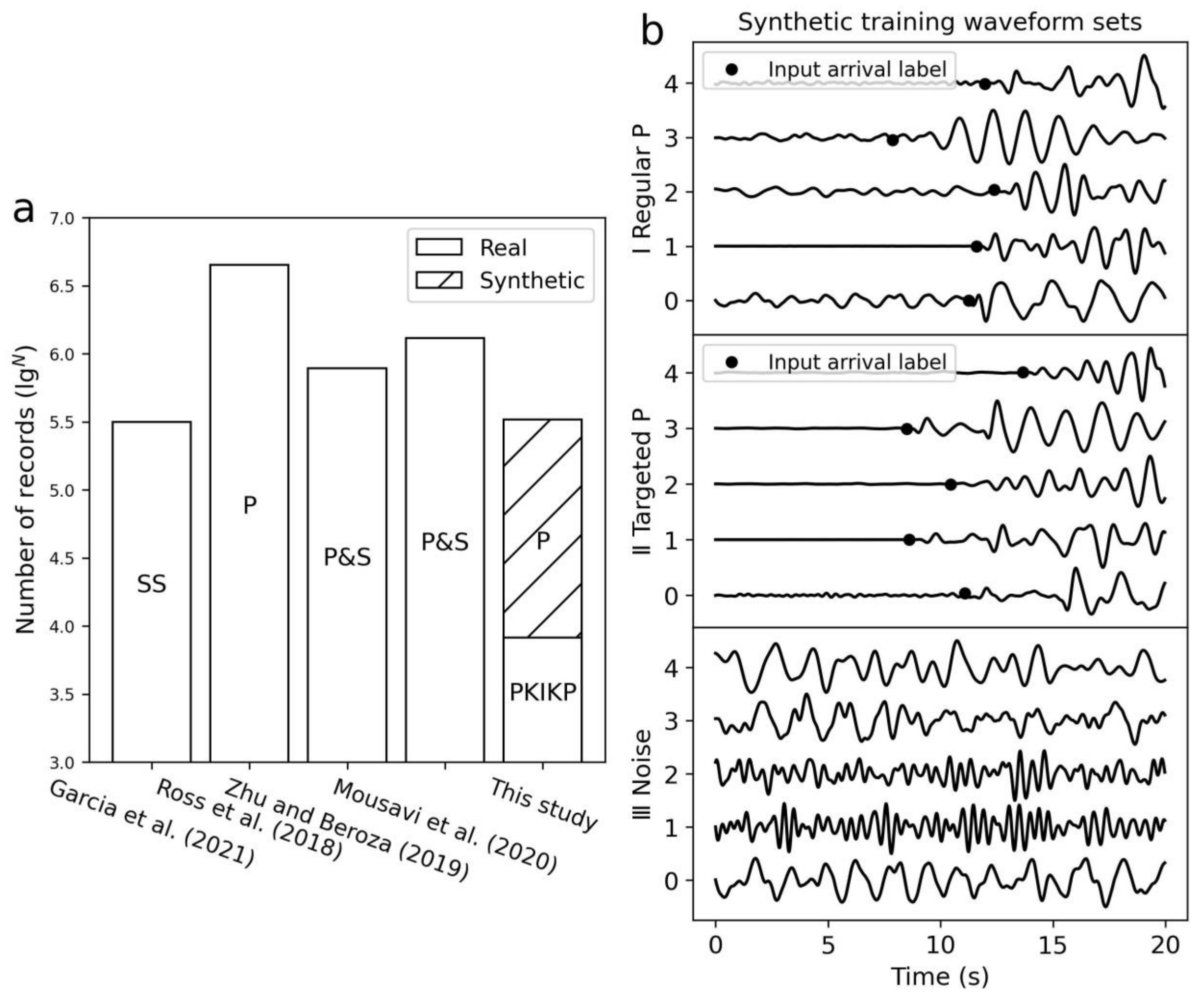

Figure 2 Datasets used in this study. *(a) compares the number of seismic records used in this study with four previous representative studies (SS phases in Garcia et al. (2021), P in Ross, Meier, and Hauksson (2018), P and S in S. M. Mousavi et al. (2020) and Zhu and Beroza (2019)). The record numbers also count noise samples and are taken the common logarithm in y, e.g., the training set of S. M. Mousavi et al. (2020) is approximately 100 times bigger than our PKIKP dataset. This study uses a specialized synthetic waveform dataset for model training and a real PKP waveform dataset with manually picked PKIKP onsets for testing. (b) shows several waveform examples in each training subset: regular P, targeted P, and noise. Each subset consists of 100K 20-s long waveforms. The dots denote the location of true P onsets in the presented samples. Targeted P waveforms have relatively small amplitude onsets compared to the regular ones to simulate and approach the features of real emergent PKIKP arrivals. The main text explains how these synthetic waveforms are generated.*

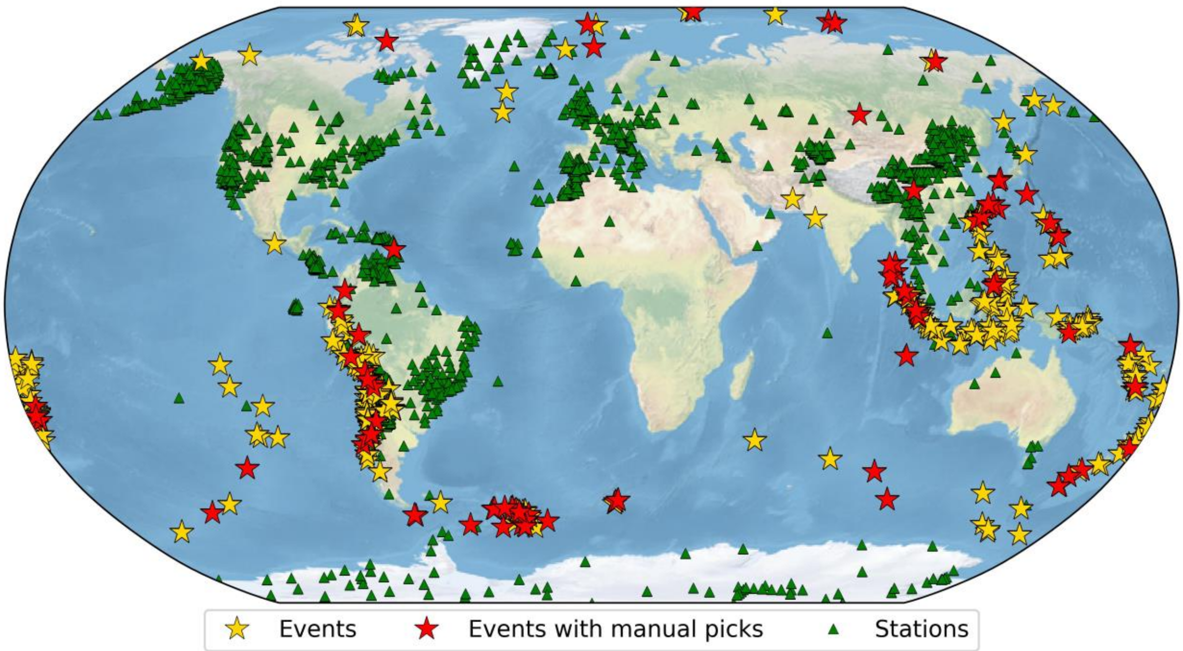

Figure 3 Location of earthquake events and antipodal stations used for collecting the PKIKP waveform dataset in this study. *Green triangles denote receiver stations. They all have epicentral distances greater than 165° from the source events. Stars denote events, whereas ones with manual picks are labeled in red. The selected events are between March 1990 and June 2019 and have 7.5 ≥ Mw ≥ 5.7.*

2.3 Model architecture and training

We employed a convolutional neural network (CNN) model in this study, inspired by the work of Ross, Meier, Hauksson, et al. (2018) and Garcia et al. (2021). The network's architecture (**Figure 4**) consists of four 1D convolutional layers, constructing the feature-extraction module. The convolution operation is performed between the input and a trainable filter in a convolutional layer. During the training process, the filters keep updated by matching the network's output and given data labels (phase onsets in this study) to allow the CNN to extract and return more sophisticated and more correct features of input data. Each convolutional layer is followed by a max-pooling layer that downsizes the input waveforms by retaining the maximum value in every two consecutive values of the inputs to extract the highlighted features and a batch-normalization layer that normalizes the previous layer's output within mini-batches. The convolutional output is then flattened and fed into fully connected layers with a one-to-one point connection to their previous layers. The network output is a directly weighted linear of the final layer, referring to CNN-picked phase onsets. Except for the final layer, all the other layers are activated by a rectified linear unit (ReLU) function, a popular activation function in deep learning that returns the input if positive and zero if negative. The sizes of the convolution kernels and dense layers are customized considering the signal frequency content of the PKIKP waveforms recorded at the global scale.

The Huber loss function (Huber, 1964) was employed, and the network was trained using the Adam optimization algorithm (Kingma & Ba, 2014). We experimented with different learning rates representing the model weights' updating speed during training, from 0.01 to 0.0001, and empirically found that the default one, 0.001, produced the most desirable results. The training process on 300K samples took about 10 minutes on a graphical processing unit at the Australian National Computational Infrastructures' (NCI) Gadi cluster. An early stopping method was applied to monitor the loss on the validation dataset during model training, allowing the model to restore the optimal weights as the validation loss no longer decreases within five epochs, in which case we considered the network well-trained. Our network's training process lasted nine epochs. The model weights obtained at the fourth epoch were restored as the optimal (**Figure S1**). Further training would not lead to significant improvements in picking precision by our test.

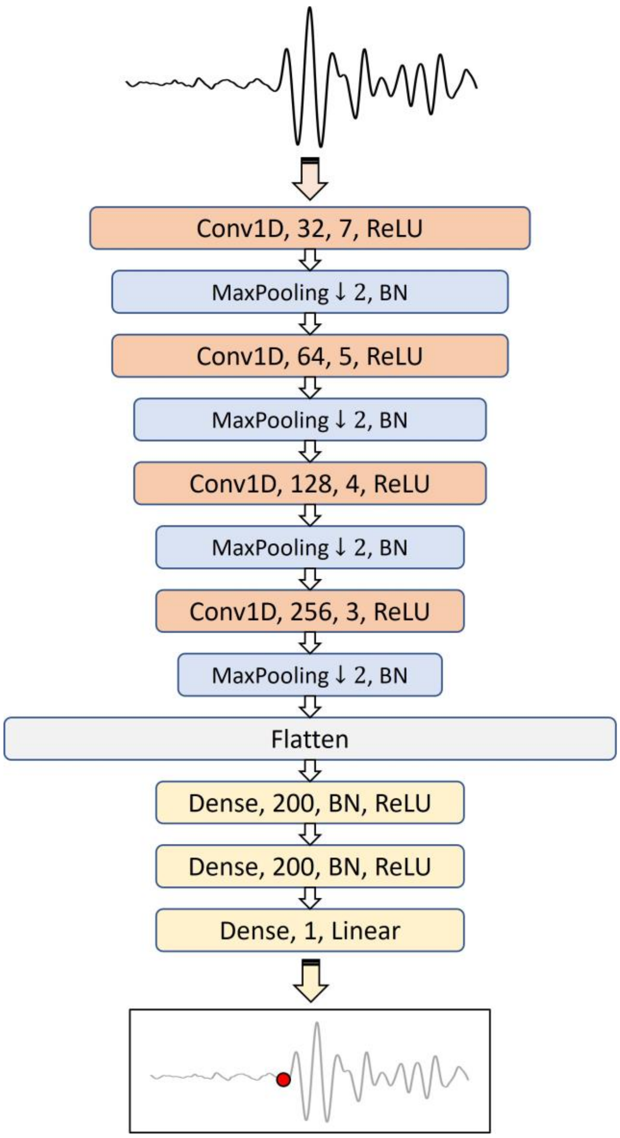

Figure 4 Architecture of the Convolutional Neural Network (CNN) used in this study. *The network consists of four convolutional layers, pooling layers, and three fully connected layers. The rectified linear unit (ReLU) is used as the activation function in all layers except the last output layer using linear activation. The network input is a 1D time series (i.e., a seismic waveform) whose label is a time pick of the phase onset (represented by the red dot; see the bottom waveform). The boxes representing layers in the network are labeled by the layer types and their corresponding parameters. Conv1D represents a convolutional layer in 1D, followed by the numbers of filters and kernel size. Max-Pooling is a pooling layer with a stride of 2 that halves the size of the input array by retaining the maximum value in every two consecutive values of the inputs. All inputs are connected to outputs in a fully connected layer (Dense layer in this network). Where specified, layers' outputs are normalized within their mini-batches, known as batch normalization (BN).*

2.4 Picking PKIKP onsets

We followed the approach proposed by Garcia et al. (2021) to repeat picking on a 20-s window sliding along a long input seismogram at every sample. At each 20-s window, the CNN network instantly picks a wave onset or detects the waveform segment as noise (Animation S1). We rely on the consistency of a pick over multiple windows as a measurement of quality for the pick, which avoids the potential randomness of the CNN picker. In an ideal condition, once an onset appears in the sliding window, it will be picked immediately and kept picked until the sliding window passes it. Since the picking may wander slightly around the onset, a clustering method, DBSCAN, is introduced to group the picks with small variations (Ester et al., 1996; Garcia et al., 2021). If there are more than five adjacent picked points in which the interval between any two neighbors is less than or equal to the sampling interval 0.1 s, they are regarded as a single cluster, and, in the meantime, the picks in this cluster are averaged to return a unique onset pick. Thus, the quality of a cluster-returned pick is defined as the number of actual picks (picks in its cluster) over the ideal maximum number of picks:

$$Q_{pick} = \frac{N_{actual}}{N_{max}} = \frac{N_{actual}\Delta t}{T} \tag{1}$$

In this study, the window's sliding step Δt is set to 0.1 s and length T is 20 s. Thus, N_{max}, the theoretical maximum counts of a cluster-returned pick, is 200. For example, if a time point is repeatedly picked 100 times by the CNN picker during sliding-window picking, its returned quality is 0.5.

In practice, we chose 50-s waveform segments around ak135 predictions as input. This waveform length allows an onset pick's quality to reach the maximum by a 20-s sliding window if the deviation between actual PKIKP onsets and ak135 predictions is less than 5 s, ideal for covering most realistic situations. Among the cluster-returned picks, only the pick with the highest quality is chosen as the final CNN picked PKIKP onset. Sliding-window picking provides a simple way to classify the quality of automatic onset picks. We empirically regard 0.2 as the quality threshold of credible picks in this study, but in practice the value depends on the desired picking precision. High-quality picks obtained by our CNN picker always have high precision.

3 Test case results

3.1 Test on labeled waveforms

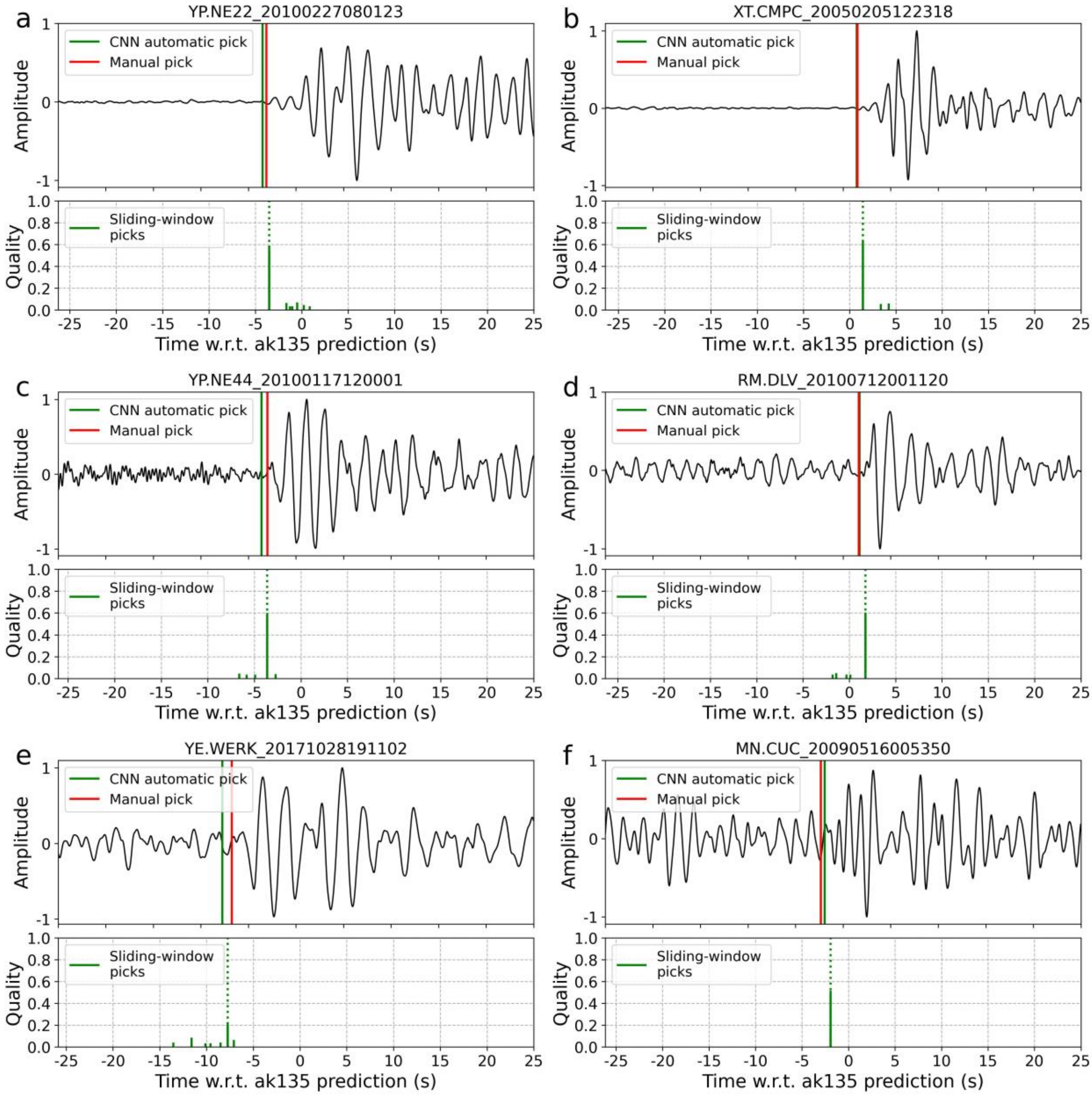

Figure 5 Representative CNN-picked PKIKP onsets. *Quality plots list the quality of all potential picks within the window by the CNN automatic picker (how the quality is defined is explained in Section 2.4). For each waveform, the pick with the highest quality is chosen as the CNN-picked PKIKP onset (labeled on the waveforms above in green), and the others are abandoned. Human-picked PKIKP onsets (as per Tkalčić et al. (2023)) are labeled in red. Note that the labeled manual picks were made by using multiple frequency filters and multiple stations and looking at P onsets (see detailed explanations in Section 2.1), while the CNN picker looks at the presented waveforms under the single frequency band of 0.5–2.0 Hz when picking. (a) and (b) are two waveforms with emergent PKIKP onsets. (c) and (d) are*

two waveforms with clear PKIKP arrivals, and (e) and (f) have higher noise. Our automatic picker returns high-quality (i.e., over 0.2 by sliding-window picking) onset picks in all cases.

Figure 5 demonstrates the result of the automatic picking procedure for six waveform examples of various data qualities in our PKIKP dataset. Animation S1 displays the sliding-window picking process dynamically. Examples of waveforms with emergent onsets are shown in panels a) and b), sharper onset waveforms are shown in panels c) and d), while waveforms at low signal-to-noise ratio are shown in panels e) and f). The clustered pick counts and their proportions to the total picks are shown in the bottom panels (see Section 2.4 for more details). The automatic picks of the onsets, corresponding to the most consistently picked time by the CNN picker while sliding the window, are close to the manual counterparts in the examples.

We first applied automatic picking to the entire set of manually picked data. There are 1457 automatic picks returned (**Figure 6a**), which include 1300 picks with empirically determined high qualities of over 0.2 from sliding-window picking (**Figure 6b**). The improvement of the automatic picks' precision attests to the necessity of the sliding-window picking scheme. The picks' precision, indicated by the mean average error (MAE), reduces markedly from 0.94 s to 0.55 s by removing the 157 low-quality picks. A small portion (46 waveforms) of samples is picked by human analysts but classified as noise by our picker. This is attributed to humans' more comprehensive picking process, as explained in Section 2.1 and the next. These CNN-classified noise waveforms do not show a recognizable onset in 0.5–2 Hz, the frequency band we adopt in automatic picking (**Figure S2**).

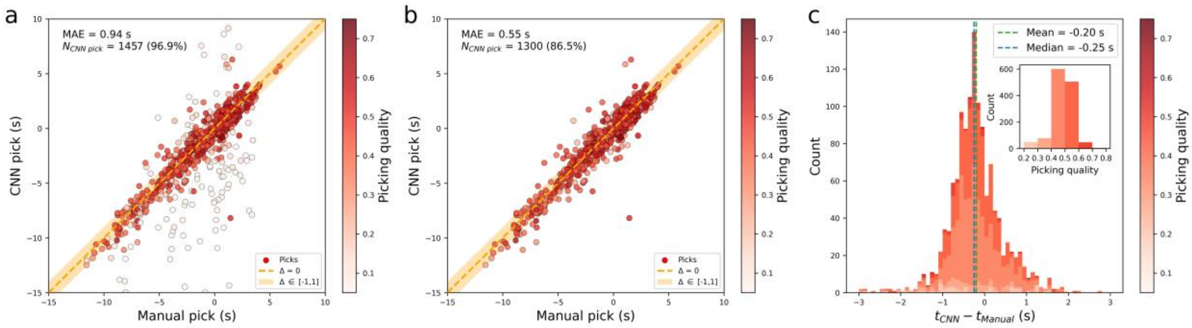

Figure 6 Comparison between human-picked and CNN-picked PKIKP onsets. *(a): all non-zero CNN picks obtained on 1503 manually picked PKIKP waveforms; (b) and (c): the filtered high-quality picks with a picking quality greater than 0.2. (a) and (b) compare the CNN-picked onsets to available manual picks, both of which are with respect to ak135*

predictions. The dot colors represent the quality of the CNN picks determined by sliding-window picking. Darker colors mean more confidence that the pick is credible. This study's CNN picking error Δ *refers to the bias of CNN picking to manual picking. The orange shadow areas indicate the CNN picks with* $\Delta <= 1$ *s. Mean absolute error (MAE), the number of plotted CNN picks, and its percentage to the number of manual picks are labeled in the upper left corner of (a) and (b). The orange dashed lines in (a) and (b) indicate the ideal result that a CNN pick is identical to the human counterpart. The picking error distribution of high-quality CNN picks is plotted in (c) with the labels of mean and median errors. The small panel in (c) shows the number of CNN picks in each quality interval from 0.2 to 0.8.*

We demonstrate the necessity of using a targeted training dataset of emerging P onsets in **Figure 7**. As introduced in Section 2.1, human analysts can get the onset picks in **Figure 7a** because they can utilize the teleseismic P onset at shorter distances to learn about the source time function and single-event sorted recordings to identify and deduce the PKIKP onset. Furthermore, they can observe how different frequency content changes the look of the PKIKP waveform. Thus, the manual picks in **Figure 7a** should not be associated with immediately recognizable PKIKP signals and are not the results of instant decisions based on individual waveforms but of a cognitive process that involves processing and analyzing many waveforms, supplementary information, and disciplinary experience.

An automatic picker makes decisions by analyzing the shape of PKIKP waveforms only. The network trained on the regular P-wave set works well on clear arrivals (**Figure 7b**) but consistently misses emergent arrivals (**Figure 7a**). As explained in Section 2.2, a targeted training P-wave dataset is our solution to enhance learning for targeted features, which refer to the emergent onsets of PKIKP resulting from extended earthquake energy release in this study. Our CNN picker trained with the new targeted dataset returned results consistent with manual picking, demonstrating that it owns the ability to identify emergent onsets and now works well on actual PKIKP waves.

The global distribution and extended period of the earthquake events selected in this study result in significant differences in quality, quantity, and shape of PKP waveforms between the events. With this in mind, we take a closer look at individual events. The events with more than 20 records are selected to compare the precision of PKIKP onset picking between our CNN picker and humans (**Figure 8**). The MAE of obtained high-quality CNN picks on each event, except for a few poorly picked ones, is approximately 0.5 s, consistent with the overall

MAE and experiential human-picking error. Since the process of CNN picking only focuses on the waveforms themselves, we can use the CNN picker to pick PKIKP onsets from any newly obtained data, independent of the actual geographical conditions behind them. However, it is worth noting that there are several events where the CNN picks very few PKIKP onsets because the waveforms are too noisy. Humans will likely obtain more credible picks for these events because of the cognitive process described above.

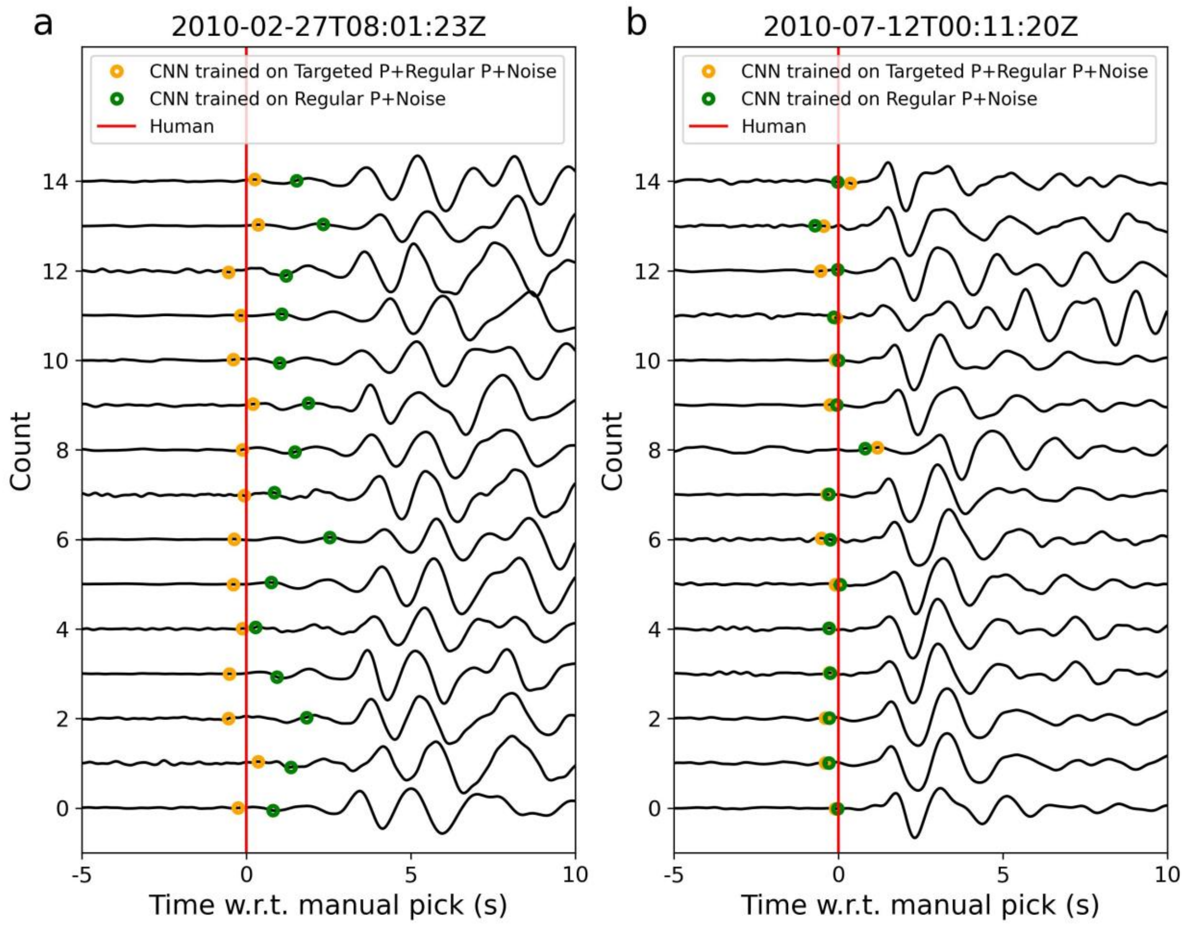

Figure 7. Demonstration of targeted training datasets. *(a) and (b) show waveform samples from two events with differential picking results, respectively. The waveforms in (a) have representative emergent PKIKP onsets compared with (b). The plotted 15-s waveform segments are centered on human-picked PKIKP onsets. The yellow circles denote the automatic picks by our chosen CNN model trained on all three training datasets, regular P, noise and targeted P, and the green circles denote the ones from a CNN model trained only on regular P and noise datasets.*

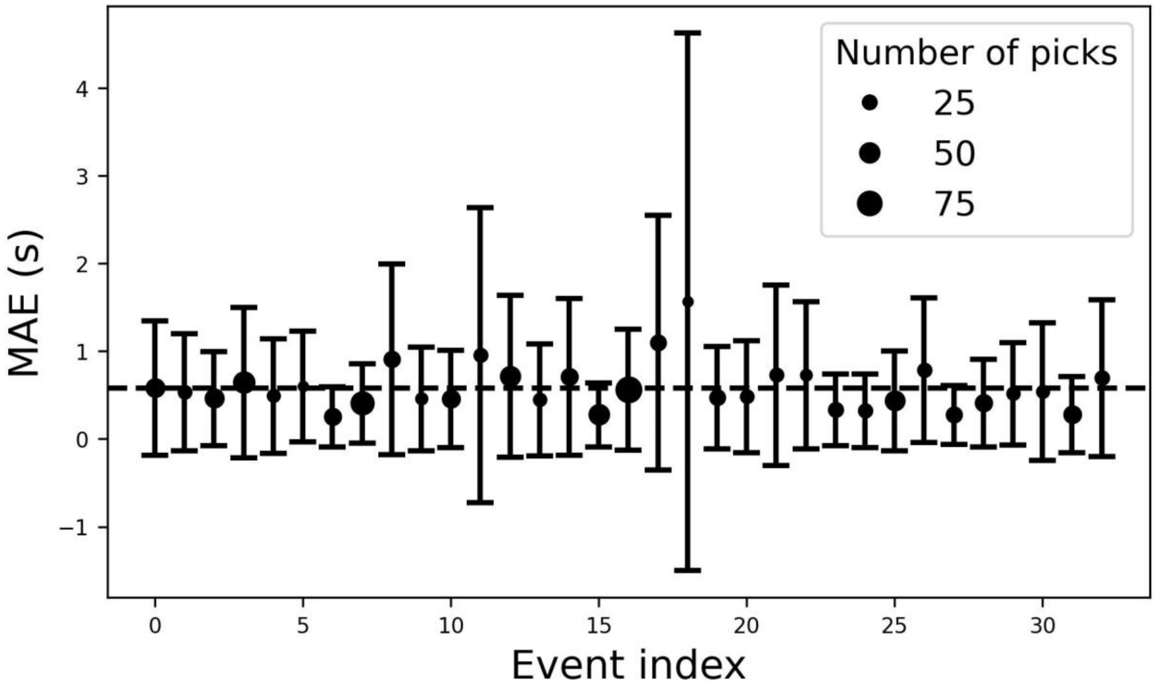

Figure 8 Precision of CNN picking compared to hand picking for individual events. *The dots and error bars denote the non-negative mean absolute error (MAE) and the root mean square error (RMSE) of CNN picks relative to their human counterparts. The size of the dots indicates the number of high-quality CNN picks obtained that are used to calculate MAE and RMSE. Only the events with over 20 samples are taken into the comparison to avoid randomness. The dashed line denotes the mean MAE over all plotted events.*

3.2 Integration test of fitting IC's anisotropic model

Given that only a small portion of our PKIKP dataset is labeled manually, we are relatively limited in testing the automatic picker's performance on the labeled data. To further showcase, we applied it to the whole dataset, including the remaining unlabeled waveforms. We conducted an integration test by comparing our inferences on the IC anisotropy by manually and CNN automatically picked PKIKP travel times (**Figure 9**). The picking results visualized on all waveforms are shown in File S1. Typically, the absolute PKIKP travel time residuals with respect to a spherically symmetric Earth model (ak135) are utilized to characterize cylindrical anisotropy in the IC. We compare the generated IC's anisotropic models from automatic picks and available manual picks, respectively. In a cylindrically anisotropic model of the Earth's IC, PKIKP travel time residuals in the IC are a function of the sampling angle ξ (Creager, 1992):

$$\frac{\Delta v}{v} = \gamma + \varepsilon cos^2\xi + \sigma sin^2\xi cos^2\xi \quad (6)$$

where the fractional velocity $\Delta v/v$ can be expressed as:

$$\frac{\Delta v}{v} = -\frac{\Delta T}{\tau_{ak135}^{IC}} \quad (7)$$

where ΔT is the residual of a CNN or human picked PKIKP travel time relative to the ak135 prediction. The absolute travel time measurements are corrected for Earth's ellipticity (Kennett & Gudmundsson, 1996) and mantle heterogeneity model DETOX-P3 (Hosseini et al., 2020). τ_{ak135}^{IC} is the ak135 predicted PKIKP travel time in the IC under 5153.9 km. $\Delta v/v$ indicates the PKIKP's travelling speed residual in the IC. ξ is the angle between the PKIKP wave's ray path in its bottoming point and the Earth's rotation axis.

We used the hierarchical Bayesian method to compare CNN and human obtained PKIKP residuals and the Markov chain Monte Carlo (McMC) to sample the anisotropy parameters, ε, σ, and γ as done by Tkalčić et al. (2023). Here we used the CNN picks with a picking quality of over 0.4, higher than the value of 0.2 used in onset-picking comparison (**Figure 6**), because the data quality of unlabeled waveforms is disparate, and we need a higher threshold to filter out uncertain automatic picks. The estimated values of model parameters are:

CNN: $\varepsilon = 3.2 \pm 0.1, \sigma = -4.5 \pm 0.3, \gamma = -0.2 \pm 0.0, \delta = 2.1 \pm 0.0$ (8)

Human: $\varepsilon = 3.0 \pm 0.1, \sigma = -4.4 \pm 0.3, \gamma = 0.0 \pm 0.0, \delta = 1.7 \pm 0.0$ (9)

δ is a hyperparameter in the Bayesian inversion inferring the amplitude of data noise expressed in seconds. It returned close values in both datasets, however, a few CNN picks significantly away from the main body led to the bias (**Figure 9a**). In order to show the most general results with minimal manual interference, we did not choose to isolate them. The stricter threshold of automatic picking, 0.4, results in a lower noise amplitude; however, it exacerbates the imbalance in the distribution of picks among latitudes because the automatic picker returns less stable onset picks on part of high-ξ waveforms with lower signal-to-noise ratios (**Figure 9b**). In addition, for the IC anisotropy models from automatic and manual picks, the values of their anisotropy strength are approximately 3.3% and 3.1% and the lowest PKIKP travelling speed in the IC appears at around 67.5° and 66.0°, respectively. The values of parameters corrected for other mantle models are listed in Table S1. The fitting results shown in **Figure 9** demonstrate that our CNN-based automatic picking method can reproduce the established knowledge of the IC anisotropy established in literature by painstaking analysts in the past with similar level of confidence. 4318 automatic picks with high picking quality are returned in minutes, which are three times larger than the archive of manual picks. However, the CNN picker shows no improvement in picking high-latitude PKIKP arrivals (low ξ) compared to human analysts. In addition, as automatic picks tend to be slightly prior to their manual counterparts, their fitting curves show an observable bias at high ξ-angles with concentrated picks.

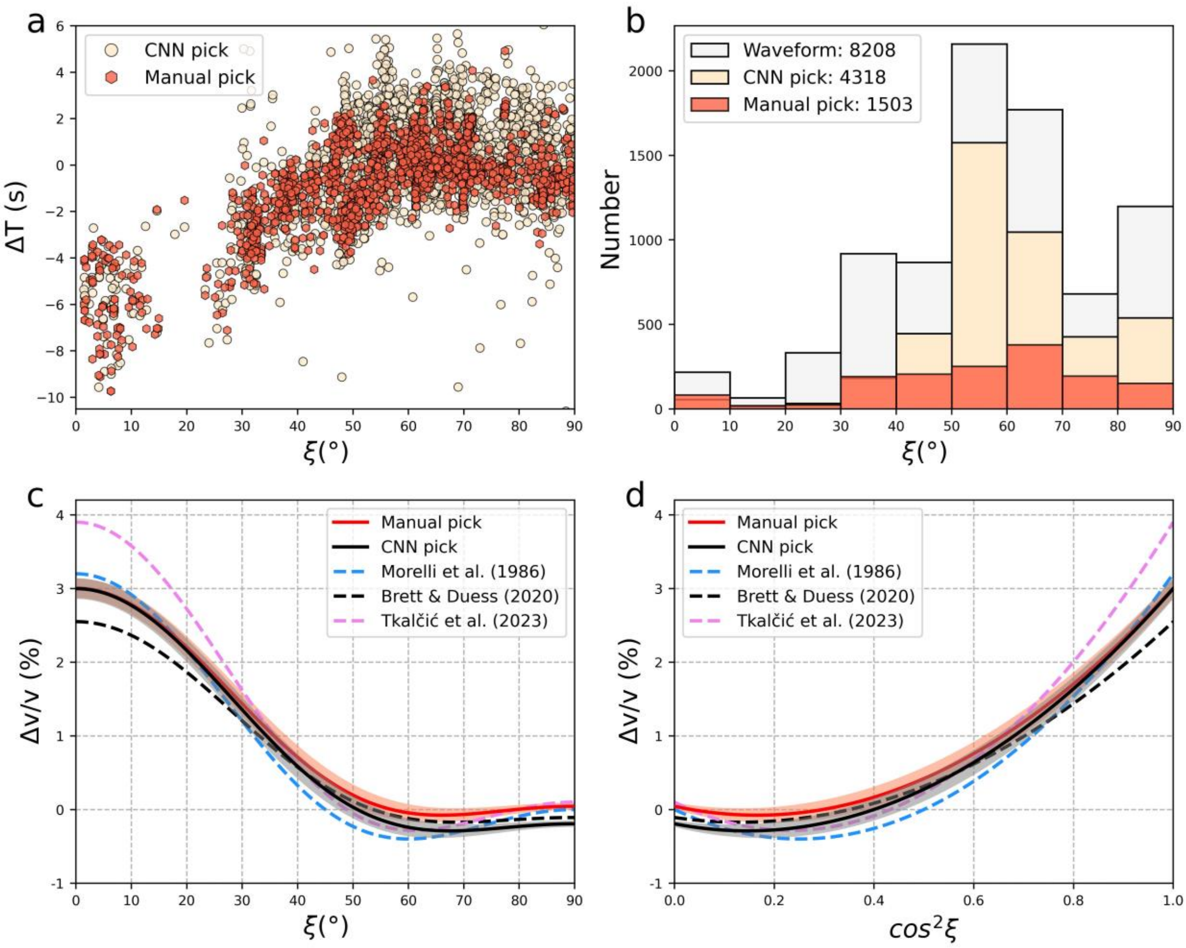

Figure 9 PKIKP residuals obtained from the CNN picker and the fitting IC anisotropy models compared to manual picking. *(a) plots all available manual picks and the CNN picks with a picking quality of over 0.4. ξ is the angle between the Earth's rotation axis and the bottoming point of PKIKP wave's ray path, and ΔT represents the residual between picked PKIKP onsets and the ak135 predictions corrected for Earth's ellipticity (Kennett & Gudmundsson, 1996) and mantle heterogeneity model DETOX-P3 (Hosseini et al., 2020). (b) shows the distribution of waveform data, manual picks, and high-quality CNN picks relative to* ξ*. Their respective numbers are annotated in the legend. The curves of fractional velocity as a function of* ξ *(c) and* $\cos^2\xi$ *(d) fitted from the two category picks using the hierarchical Bayesian method are denoted in black and red, respectively. See more details about the fitting*

in the main text. The shadow zones denote the uncertainties of the fitting. For reference, a few previously published IC anisotropic models (Brett & Deuss, 2020; Morelli et al., 1986; Tkalčić et al., 2023) are plotted using dashed curves in different colors.

4 Discussion

It has been demonstrated that our CNN model picks PKIKP onsets nearly as well as experienced analysts. However, a MAE of approximately 0.5 s consistently exists in overall and individual event picking, higher than the picking error of current leading automatic pickers in the local scale, low to 0.1 s (Münchmeyer et al., 2022). It should be noted that we work with lower-frequency signals of 0.5-2.0 Hz in global events than over 5 Hz in local events and sample them at 40 Hz (Δt=0.025 s) rather than the usual 100 Hz ((Δt=0.01 s), leaving higher systematic errors in our picks. In addition, we generated a synthetic waveform dataset while training our network, expecting to simulate the IC attenuation due to the lack of manually picked PKIKP arrival data. Regarding the complexity and diversity of geological conditions and datasets on the global scale, it is challenging for our global picker to achieve the same high precision based on it as previous local-scale pickers using real data.

It is also reasonable to see a minor deviation between automatic and manual picking as we consider its sources from two aspects. On the one hand, experienced human analysts can identify emergent seismic signals on noisy waveforms precisely by inspecting multiple frequencies and observing the source time functions using P wave arrivals at shorter epicentral distances. These procedures are necessary for measuring global and antipodal datasets (Tkalčić et al., 2023). Lacking supplementary information and limited in a fixed frequency band, the automatic picker struggles more to locate the emergent onsets than humans. On the other hand, however, the shortage of published labeled PKIKP waves cannot be ignored. Humans observe waveforms with their eyes and pick the onsets subjectively. The precision of manual picks varies among their authors. The automatic picker can ensure that onsets are picked immediately and consistently once they appear. Instead, humans can only pick them up until they become recognizable by eyes. This explains why the automatic picks tend to be earlier than their manual counterparts, with a mean deviation of 0.2 s (**Figure 6c**).

In this study, we used a CNN architecture with a modest size. This network containing four convolutional layers is shallower in depths than recent state-of-the-art developments in local earthquake detection and phase picking, for example, PhaseNet with an up-sampling module

(Zhu & Beroza, 2019) and EQTransformer consisting of 56 layers (S. M. Mousavi et al., 2020). Also, we used less training data, 300K synthetic waveforms (1.2M for EQTransformer and 780K for PhaseNet), to achieve our task. This is because our strategy is different from the local pickers above (see the description in Introduction) and our task is less complicated. We deal with large earthquakes that are documented in global earthquake catalogs (such as NEIC and GMTs). The global earth model can determine waveform windows that certainly contain PKIKP arrivals. Thus, the sole purpose of our network is to determine whether it can pick the seismic arrival onset already present in the windowed waveforms, not to determine their presence without any prior knowledge. This explains the satisfactory performance of the network, given its modest size.

Besides picking individual waveforms, we also experimented with using our CNN to determine arrival onsets on stacked waveforms with higher signal-to-noise ratios. Firstly, the adaptive stacking method aligns similar waveforms by minimizing a misfit function defined as the overall difference with the simultaneous stack (Rawlinson & Kennett, 2004). Thanks to its great performance in waveform stacking, our picker can be applied to the stacked waveforms from stations in a regional network, potentially improving the precision. However, automatic picking on stacked waveforms is subjected to the precision of adaptive stacking and the risk of introducing systematic bias in the onset pick of every individual waveform of the network. Theoretically, individual waveform picking using a fine-trained network can prevent these issues. Hence, we prefer not to apply the picker on adaptively stacked waveforms.

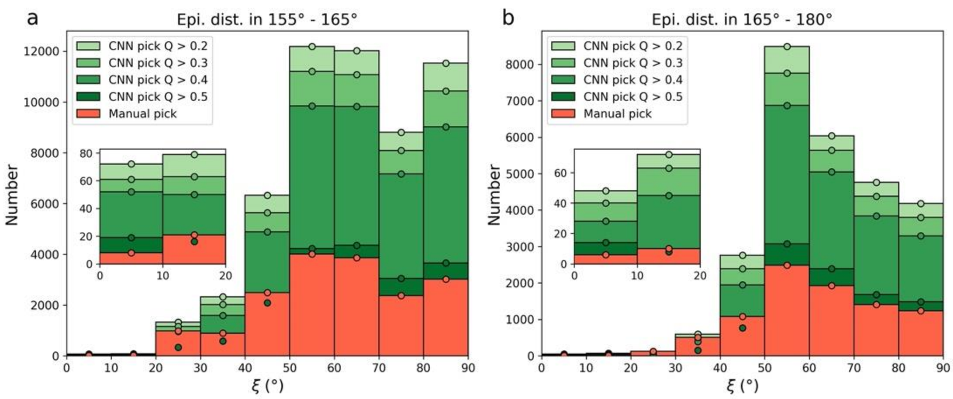

Figure 10 Distribution of manual picks and CNN picks under multiple quality levels relative to the sampling angle ξ on a global PKIKP waveform dataset collected between 2001-2020. *We choose all m_b >5.8 earthquakes occurring in the time interval 2001-2020 and download the corresponding PKIKP waveforms in the epicentral distance range of (a) 155° -*

165° and (b) 165° - 180° from the IRIS Data Management Center. The catalogs of events and manual picks of PKIKP onsets are obtained from the ISC (Bondár & Storchak, 2011). We increase the quality (Q) threshold of CNN picks from 0.2, the criteria of precise picks chosen by experience, to 0.5, at which the number of CNN picks is comparable to manual picks. The upper margin of each quality interval bar is marked with dots.

Next, to demonstrate the potential of our automatic picker in expanding the dataset measurements, we collected new PKP waveforms in two epicentral distance ranges, 165°-180°, the range of antipodal stations as same as the present dataset, and 155°-165° to cover the whole IC, in all m_b5.8+ earthquake events during the 20 years from 2001 to 2020 from the IRIS Data Management Center. Record numbers in the two new datasets are ~57K and ~123K. Meanwhile, we searched their associated manual picks in the ISC database, which were reported by multiple authors with various tags. As implemented earlier in Section 3.2, our CNN network picked over half of the new large datasets, greatly expanding the PKIKP onset archive near the equator (Figure 10). The picking quality do provide a one-step strategy that allows us to filter the automatic picks according to research purposes. We can acquire more precise and credible onset picks close to human picking by setting a high threshold of quality as Section 3.2 or catch as many picks as possible with medium quality. Interestingly, both humans and the CNN automatic picker obtain relatively few picks from polar events. Despite being limited by the number of earthquakes and receivers at high latitudes and their data quality (Tkalčić, 2017), this illustrates the need for more careful consideration of regional differences in improving the picker. Further developed automatic pickers are expected to extend the usable data for studying the Earth's IC with potentially unlocked datasets.

We acknowledge the potential drawbacks when using the synthetic training dataset. On a positive note, the automatic picker performs well on the labeled waveforms (Figure 6) and the integration task (Figure 9). On the other hand, the gap between synthetic waveforms and real PKIKP waves cannot be filled. We introduced emergent arrivals in training by experience to consider the IC attenuation, which is inadequate and not what we expected when constructing an automatic picker. Collecting more labeled real waveforms sensitive to the IC and the Earth's deep interior for training and testing the next iteration's networks becomes more efficient thanks to the successful application of the automatic picker and the mechanism of picking quality control in this study. It is in our imminent plan to inspect manually and define the potentially unambiguous PKIKP onsets, which are initially picked by the present

CNN picker. This could save considerable time and effort as we will not need to visualize tens of thousands of noisy waveforms.

The application of our CNN model in picking PKIKP onsets demonstrates the potential of using deep learning algorithms to facilitate IC studies. In addition to optimizing the automatic picker, we plan to develop new deep-learning-based models to predict PKIKP-PKPbc and PKIKP-PKPab differential travel times widely used in IC studies for different purposes (e.g., Attanayake et al., 2014; Creager, 1992; Niu & Chen, 2008; Shearer & Toy, 1991; Song & Helmberger, 1993). Most recently, a physics-informed neural network (PINN) was developed to model P wave travel times between any source-receiver pair in a global mantle model (Taufik et al., 2023). It shows significant advantages in saving storage and computing resources compared with traditional 3-D Earth velocity models. The high picking efficiency of our automatic picker demonstrated in this study provides strong support for obtaining a much larger PKP-wave differential travel time dataset than ever before. This helps us construct a new automated tool that can generate differential travel times based on any eligible coordinate pair input directly, thereby avoiding the processes of phase onset picking and computing Earth structures along the ray path.

5 Conclusion

The application of deep-learning algorithms has expanded from local earthquake seismology to structural seismology on global scales. In this work, however, we focused on the PKIKP waves traversing near the Earth’s center. The global dataset of PKIKP waves is crucial for exploring the Earth’s deep interior, particularly the Earth’s inner core. To expand the PKIKP onset archive and ensure its consistent quality across datasets, we employed a CNN network to pick the onsets of PKIKP waves automatically. Our CNN automatic picker, though simple in architecture, picked the majority of human-picked PKIKP onsets achieving human-level precision, thanks to the well-designed synthetic training dataset considering the features of PKIKP waves. Automatic picks show a consistent precision across earthquake events. Our automatic picker obtained 4,318 high-quality picks, three times the manual picks, out of 8,208 PKIKP waveforms in just several minutes, demonstrating its efficiency in harnessing big global datasets compared to human analysts.

The deep learning expanded PKIKP travel time dataset is expected to increase the current sampling coverage of the IC in places where data are available by at least an order of magnitude. Thus, it could reveal more details of the Earth's interior. In an integration test, we selected IC anisotropy as one of the most prominent and well-documented features of the IC. The IC anisotropy model produced by the CNN-picked PKIKP travel times is similar to the existing ones based on previous meticulously hand-picked datasets. However, it should be clear that there are still notable advantages in experienced human analysts picking the arrivals from noisy waveforms, e.g., where the sampling paths are rare, particularly those originating from events and stations at high latitudes. They can use multiple frequency filters and look at supplementary information to distill information from a few valuable records.

The performance of our initial deep-learning-based automatic picker on PKIKP waves and the previous applications on SS waves (Garcia et al., 2021) and PmKP waves (Dong et al., 2024) shed light on the path forward for deep Earth seismology harnessing large datasets of existing and new waveforms and information therein. These could trigger the need for more comprehensive analysis to support deep Earth models with adequate uncertainty estimates. Our future work will apply the automatic picker to new datasets and provide a more in-depth analysis of the results obtained from it to improve current IC anisotropy models, though the picker has to be further improved to learn IC characteristics better by introducing more real PKIKP data into training. Furthermore, we will explore the deep-learning approaches to other laborious tasks in deep Earth seismology, such as measuring differential travel times between first arrivals of phase pairs, such as ScS-S (Houser et al., 2008; S. Mousavi et al., 2021) or PcP-P waves sensitive to the lowermost mantle (Muir et al., 2022; Tkalčić & Romanowicz, 2002).

Data availability statement

Synthetic teleseismic waveforms used for network training in this study are generated using Teleswavesim software package (Audet et al., 2019) available at https://zenodo.org/badge/latestdoi/204565459. All real records used in this study, including PKIKP phase and real noise waveforms, are downloaded from the Incorporated Research Institution for Seismology Data Management Center (IRIS DMC; https://ds.iris.edu/ds/nodes/dmc/) using the ObsPy package (Beyreuther et al., 2010). The PKIKP waveform dataset with hand-picked PKIKP onsets described in Section 2.1 is prepared by Tkalčić et al. (2023). The picked waveforms used for model testing include the following seismic networks: 2H (10.7914/SN/2H_2016), 3D (10.7914/SN/3D_2010), AI

(10.7914/SN/AI), AK (10.7914/SN/AK), AT (10.7914/SN/AT), AU (10.26186/144675), AV (10.7914/SN/AV), BE (10.7914/SN/BE), BL (https://www.fdsn.org/networks/detail/BL), BR (https://www.fdsn.org/networks/detail/BR), C (https://www.fdsn.org/networks/detail/C), CB (10.7914/SN/CB), CH (10.12686/sed/networks/ch), CN (10.7914/SN/CN), CU (10.7914/SN/CU), CZ (10.7914/SN/CZ), ET (https://www.fdsn.org/networks/detail/ET), G (10.18715/GEOSCOPE.G), GE (10.14470/TR560404), GR (10.25928/mbx6-hr74), GT (10.7914/SN/GT), HK (https://www.fdsn.org/networks/detail/HK), IC (10.7914/SN/IC), II (10.7914/SN/II), IM (10.7914/vefq-vh75), IU (10.7914/SN/IU), JP (https://www.fdsn.org/networks/detail/JP), KN (10.7914/SN/KN), KZ (10.7914/SN/KZ), LD (10.7914/SN/LD), MN (10.13127/SD/fBBBtDtd6q), MY (https://www.fdsn.org/networks/detail/MY), NB (https://www.fdsn.org/networks/detail/NB), NE (10.7914/SN/NE), NL (10.21944/e970fd34-23b9-3411-b366-e4f72877d2c5), NM (https://www.fdsn.org/networks/detail/NM), NU (10.7914/SN/NU), ON (10.7914/SN/ON), PA (10.7914/SN/PA), PM (10.7914/SN/PM), PN (https://www.fdsn.org/networks/detail/PN), PS (https://www.fdsn.org/networks/detail/PS), RM (10.7914/SN/RM), TA (10.7914/SN/TA), TM (https://www.fdsn.org/networks/detail/TM), TO (10.7909/C3RN35SP), TW (10.7914/SN/TW), US (10.7914/SN/US), X4 (10.7914/SN/X4_2007), X5 (https://www.fdsn.org/networks/detail/X5_2007), X6 (10.7914/SN/X6_2007), XB (10.7914/SN/XB_2009), XC (10.7914/SN/XC_2012), XD (10.7914/SN/XD_2002; 10.7914/SN/XD_2007), XE (10.7914/SN/XE_2009), XG (10.7914/SN/XG_1999), XH (10.7914/SN/XH_2008), XJ (10.15778/RESIF.XJ2009), XN (10.7914/SN/XN_2008), XT (10.7914/SN/XT_2003), XW (10.7914/SN/XW_1997), YC (10.7914/SN/YC_2000; 10.7914/SN/YC_2006), YE (10.7914/SN/YE_2011), YG (10.7914/SN/YG_2016), YM (10.7914/SN/YM_2006), YP (10.7914/SN/YP_2009), YS (10.7914/SN/YS_2009), YT (10.7914/SN/YT_2007), YZ (10.7914/SN/YZ_2009), Z8 (10.7914/SN/Z8_2006), ZI (10.7914/SN/ZI_2011), ZL (10.7914/SN/ZL_2007), ZM (10.7914/SN/ZM_2007), ZQ (10.7914/SN/ZQ_2001), ZV (10.7914/SN/ZV_2008). The manual picks of the expanded datasets in Discussion are from the International Seismological Centre (ISC) Bulletin (https://www.isc.ac.uk/iscbulletin/search/arrivals/). The model of the convolutional neural network is built using the TensorFlow package (Abadi et al., 2016; https://www.tensorflow.org/), and the figures are made using the matplotlib package (Hunter, 2007). The codes, trained model parameters, and training datasets will be made available when the paper is considered for later stages.

Acknowledgments

The model training was performed with the assistance of resources and services from the National Computational Infrastructure (NCI) facility supported by the Australian Government. We are grateful for the suggestions and discussion from the RSES group. The Australian Research Council supported this work through a Discovery Project (DP220102815) and a Discovery Early Career Research Award (DE230100025).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). TensorFlow: a system for large-scale machine learning. In *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation* (pp. 265–283). USA: USENIX Association.

Aki, K., & Lee, W. H. K. (1976). Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model. *Journal of Geophysical Research (1896-1977)*, *81*(23), 4381–4399. https://doi.org/10.1029/JB081i023p04381

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, *72*(6B), S225–S242. https://doi.org/10.1785/BSSA07206B0225

Attanayake, J., Cormier, V. F., & de Silva, S. M. (2014). Uppermost inner core seismic structure – new insights from body waveform inversion. *Earth and Planetary Science Letters*, *385*, 49–58. https://doi.org/10.1016/j.epsl.2013.10.025

Audet, P., Thomson, C., Bostock, M., & Eulenfeld, T. (2019). Telewavesim: Python software for teleseismic body wave modeling. *Journal of Open Source Software*, *4*, 1818. https://doi.org/10.21105/joss.01818

Baer, M., & Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, *77*(4), 1437–1445. https://doi.org/10.1785/BSSA0770041437

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, *81*(3), 530–533. https://doi.org/10.1785/gssrl.81.3.530

Bondár, I., & Storchak, D. (2011). Improved location procedures at the International Seismological Centre. *Geophysical Journal International*, *186*(3), 1220–1244.

https://doi.org/10.1111/j.1365-246X.2011.05107.x

Brett, H., & Deuss, A. (2020). Inner core anisotropy measured using new ultra-polar PKIKP paths. *Geophysical Journal International*, *223*, 1230–1246. https://doi.org/10.1093/gji/ggaa348

Bullen, K. E. (1961). Seismic Ray Theory. *Geophysical Journal International*, *4*(Supplement_1), 93–105. https://doi.org/10.1111/j.1365-246X.1961.tb06806.x

Burdick, S., Waszek, L., & Lekić, V. (2019). Seismic tomography of the uppermost inner core. *Earth and Planetary Science Letters*, *528*, 115789. https://doi.org/10.1016/j.epsl.2019.115789

Costa de Lima, T., Tkalčić, H., & Waszek, L. (2022). A New Probe Into the Innermost Inner Core Anisotropy via the Global Coda-Correlation Wavefield. *Journal of Geophysical Research: Solid Earth*, *127*(4), e2021JB023540. https://doi.org/10.1029/2021JB023540

Creager, K. C. (1992). Anisotropy of the inner core from differential travel times of the phases PKP and PKIKP. *Nature*, *356*(6367), 309–314. https://doi.org/10.1038/356309a0

Dahlen, F. A., & Tromp, J. (1998). *Theoretical Global Seismology*. Princeton University Press. https://doi.org/10.2307/j.ctv131bvfd

Deuss, A., Irving, J. C. E., & Woodhouse, J. H. (2010). Regional Variation of Inner Core Anisotropy from Seismic Normal Mode Observations. *Science*, *328*(5981), 1018–1020. https://doi.org/10.1126/science.1188596

Dong, S., Chen, Y., Zhang, B., Ni, S., Chen, X., & Wang, Y. (2024). Identification of Rare Multiple Core-Mantle Boundary Reflections PmKP Up To P7KP With Deep Learning. *Geophysical Research Letters*, *51*(2), e2023GL105464. https://doi.org/10.1029/2023GL105464

Dziewonski, A. M., & Anderson, D. L. (1981). Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, *25*(4), 297–356. https://doi.org/10.1016/0031-9201(81)90046-7

Dziewonski, A. M., Chou, T.-A., & Woodhouse, J. H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research: Solid Earth*, *86*(B4), 2825–2852. https://doi.org/10.1029/JB086iB04p02825

Ekström, G., Dziewoński, A. M., Maternovskaya, N. N., & Nettles, M. (2005). Global seismicity of 2003: centroid–moment-tensor solutions for 1087 earthquakes. *Physics of the Earth and Planetary Interiors*, *148*(2), 327–351. https://doi.org/10.1016/j.pepi.2004.09.006

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). Portland, Oregon: AAAI Press.

Garcia, J. A., Waszek, L., Tauzin, B., & Schmerr, N. (2021). Automatic Identification of Mantle Seismic Phases Using a Convolutional Neural Network. *Geophysical Research Letters*, *48*(18), e2020GL091658. https://doi.org/10.1029/2020GL091658

Gutenberg, B. (1914). Ueber Erdbebenwellen. VII A. Beobachtungen an Registrierungen von Fernbeben in Göttingen und Folgerung über die Konstitution des Erdkörpers (mit Tafel). *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, *1914*, 125–176.

Guy, M. R., Patton, J. M., Fee, J., Hearne, M., Martinez, E., Ketchum, D., et al. (2015). *National Earthquake Information Center systems overview and integration* (Report No. 2015–1120) (p. 28). Reston, VA. https://doi.org/10.3133/ofr20151120

Hildyard, M. W., Nippress, S. E. J., & Rietbrock, A. (2008). Event Detection and Phase Picking Using a Time-Domain Estimate of Predominate Period Tpd. *Bulletin of the Seismological Society of America*, *98*(6), 3025–3032. https://doi.org/10.1785/0120070272

Hosseini, K., Sigloch, K., Tsekhmistrenko, M., Zaheri, A., Nissen-Meyer, T., & Igel, H. (2020). Global mantle structure from multifrequency tomography using P, PP and P-diffracted waves. *Geophysical Journal International*, *220*(1), 96–141. https://doi.org/10.1093/gji/ggz394

Houser, C., Masters, G., Shearer, P., & Laske, G. (2008). Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms. *Geophysical Journal International*, *174*(1), 195–212. https://doi.org/10.1111/j.1365-246X.2008.03763.x

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Inge Lehmann. (1936). P’. *Publications Du Bureau Central Séismologique International*, *A14*(3).

Ishii, M., & Dziewoński, A. M. (2002). The innermost inner core of the earth: Evidence for a change in anisotropic behavior at the radius of about 300 km. *Proceedings of the National Academy of Sciences*, *99*(22), 14026–14030. https://doi.org/10.1073/pnas.172508499

Kennett, B. L. N. (2009). *Seismic Wave Propagation in Stratified Media*. ANU Press. Retrieved from https://www.jstor.org/stable/j.ctt24h2zr

Kennett, B. L. N., & Engdahl, E. R. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, *105*(2), 429–465. https://doi.org/10.1111/j.1365-246X.1991.tb06724.x

Kennett, B. L. N., & Gudmundsson, O. (1996). Ellipticity corrections for seismic phases. *Geophysical Journal International*, *127*(1), 40–48. https://doi.org/10.1111/j.1365-246X.1996.tb01533.x

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

Kingma, D. P., & Ba, J. (2014, December 1). Adam: A Method for Stochastic Optimization. *arXiv E-Prints*. https://doi.org/10.48550/arXiv.1412.6980

Laske, G., Masters, G., Ma, Z., & Pasyanos, M. (2013). Update on CRUST1.0 - A 1-degree Global Model of Earth's Crust, EGU2013-2658. Presented at the EGU General Assembly Conference Abstracts.

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

Ma, X., & Tkalčić, H. (2021). CCREM: New Reference Earth Model From the Global Coda-Correlation Wavefield. *Journal of Geophysical Research: Solid Earth*, *126*(9), e2021JB022515. https://doi.org/10.1029/2021JB022515

Mohorovičić, A. (1910). *Potres od 8. X. 1909 = Das Beben vom 8. X. 1909*. Zagreb: Albrechts.

Morelli, A., Dziewonski, A. M., & Woodhouse, J. H. (1986). Anisotropy of the inner core inferred from PKIKP travel times. *Geophysical Research Letters*, *13*(13), 1545–1548. https://doi.org/10.1029/GL013i013p01545

Mousavi, S., Tkalčić, H., Hawkins, R., & Sambridge, M. (2021). Lowermost Mantle Shear-Velocity Structure From Hierarchical Trans-Dimensional Bayesian Tomography.

Journal of Geophysical Research: Solid Earth, *126*(9), e2020JB021557. https://doi.org/10.1029/2020JB021557

Mousavi, S. M., & Beroza, G. C. (2022). Deep-learning seismology. *Science*, *377*(6607), eabm4470. https://doi.org/10.1126/science.abm4470

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 3952. https://doi.org/10.1038/s41467-020-17591-w

Muir, J. B., Tanaka, S., & Tkalčić, H. (2022). Long-Wavelength Topography and Multiscale Velocity Heterogeneity at the Core-Mantle Boundary. *Geophysical Research Letters*, *49*(17), e2022GL099943. https://doi.org/10.1029/2022GL099943

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., et al. (2022). Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers. *Journal of Geophysical Research: Solid Earth*, *127*(1), e2021JB023499. https://doi.org/10.1029/2021JB023499

Niu, F., & Chen, Q.-F. (2008). Seismic evidence for distinct anisotropy in the innermost inner core. *Nature Geoscience*, *1*(10), 692–696. https://doi.org/10.1038/ngeo314

Obayashi, M., Yoshimitsu, J., Nolet, G., Fukao, Y., Shiobara, H., Sugioka, H., et al. (2013). Finite frequency whole mantle P wave tomography: Improvement of subducted slab images. *Geophysical Research Letters*, *40*(21), 5652–5657. https://doi.org/10.1002/2013GL057401

Phạm, T.-S., & Tkalčić, H. (2017). On the feasibility and use of teleseismic P wave coda autocorrelation for mapping shallow seismic discontinuities. *Journal of Geophysical Research: Solid Earth*, *122*(5), 3776–3791. https://doi.org/10.1002/2017JB013975

Phạm, T.-S., & Tkalčić, H. (2023). Up-to-fivefold reverberating waves through the Earth's center and distinctly anisotropic innermost inner core. *Nature Communications*, *14*(1), 754. https://doi.org/10.1038/s41467-023-36074-2

Rawlinson, N., & Kennett, B. L. N. (2004). Rapid estimation of relative and absolute delay times across a network by adaptive stacking. *Geophysical Journal International*, *157*(1), 332–340. https://doi.org/10.1111/j.1365-246X.2004.02188.x

Romanowicz, B., & Bréger, L. (2000). Anomalous splitting of free oscillations: A reevaluation of possible interpretations. *Journal of Geophysical Research: Solid Earth*, *105*(B9), 21559–21578. https://doi.org/10.1029/2000JB900144

Ross, Z. E., Meier, M., Hauksson, E., & Heaton, T. H. (2018). Generalized Seismic Phase Detection with Deep Learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894–2901. https://doi.org/10.1785/0120180080

Ross, Z. E., Meier, M.-A., & 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. https://doi.org/10.1029/2017JB015251

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Saad, O. M., Soliman, M. S., Chen, Y., Amin, A. A., & Abdelhafiez, H. E. (2022). Discriminating Earthquakes From Quarry Blasts Using Capsule Neural Network. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5. https://doi.org/10.1109/LGRS.2022.3207238

Shearer, P. M. (1994). Constraints on inner core anisotropy from PKP(DF) travel times. *Journal of Geophysical Research: Solid Earth*, *99*(B10), 19647–19659. https://doi.org/10.1029/94JB01470

Shearer, P. M. (2019). *Introduction to Seismology* (3rd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316877111

Shearer, P. M., & Toy, K. M. (1991). PKP(BC) versus PKP(DF) differential travel times and aspherical structure in the Earth's inner core. *Journal of Geophysical Research: Solid Earth*, *96*(B2), 2233–2247. https://doi.org/10.1029/90JB02370

Sleeman, R., & van Eck, T. (1999). Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the Earth and Planetary Interiors*, *113*(1), 265–275. https://doi.org/10.1016/S0031-9201(99)00007-2

Song, X., & Helmberger, D. V. (1993). anisotropy of Earth's inner core. *Geophysical Research Letters*, *20*(23), 2591–2594. https://doi.org/10.1029/93GL02812

Stephenson, J., Tkalčić, H., & Sambridge, M. (2021). Evidence for the Innermost Inner Core: Robust Parameter Search for Radially Varying Anisotropy Using the Neighborhood Algorithm. *Journal of Geophysical Research: Solid Earth*, *126*(1), e2020JB020545. https://doi.org/10.1029/2020JB020545

Su, W., & Dziewonski, A. M. (1995). Inner core anisotropy in three dimensions. *Journal of Geophysical Research: Solid Earth*, *100*(B6), 9831–9852. https://doi.org/10.1029/95JB00746

Taufik, M. H., Waheed, U. bin, & Alkhalifah, T. A. (2023). A neural network based global traveltime function (GlobeNN). *Scientific Reports*, *13*(1), 7179. https://doi.org/10.1038/s41598-023-33203-1

Tkalčić, H. (2017). *The Earth's Inner Core: Revealed by Observational Seismology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139583954

Tkalčić, H., & Phạm, T.-S. (2018). Shear properties of Earth's inner core constrained by a detection of J waves in global correlation wavefield. *Science*, *362*(6412), 329–332. https://doi.org/10.1126/science.aau7649

Tkalčić, H., & Romanowicz, B. (2002). Short scale heterogeneity in the lowermost mantle:

insights from PcP-P and ScS-S data. *Earth and Planetary Science Letters*, *201*(1), 57–68. https://doi.org/10.1016/S0012-821X(02)00657-X

Tkalčić, H., Phạm, T.-S., & Wang, S. (2020). The Earth's coda correlation wavefield: Rise of the new paradigm and recent advances. *Earth-Science Reviews*, *208*, 103285. https://doi.org/10.1016/j.earscirev.2020.103285

Tkalčić, H., Costa de Lima, T. P., Phạm, T.-S., & Tanaka, S. (2023). Inner Core Anisotropy from Antipodal PKIKP Traveltimes. In *Core-Mantle Co-Evolution* (pp. 165–189). American Geophysical Union (AGU). https://doi.org/10.1002/9781119526919.ch10

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017, June 1). Attention Is All You Need. *arXiv E-Prints*. https://doi.org/10.48550/arXiv.1706.03762

Wang, S., & Tkalčić, H. (2021). Shear-Wave Anisotropy in the Earth's Inner Core. *Geophysical Research Letters*, *48*(19), e2021GL094784. https://doi.org/10.1029/2021GL094784

Waszek, L., & Deuss, A. (2015). Observations of exotic inner core waves. *Geophysical Journal International*, *200*(3), 1636–1650. https://doi.org/10.1093/gji/ggu497

Woodhouse, J. H., Giardini, D., & Li, X.-D. (1986). Evidence for inner core anisotropy from free oscillations. *Geophysical Research Letters*, *13*(13), 1549–1552. https://doi.org/10.1029/GL013i013p01549

Zhu, W., & Beroza, G. C. (2019). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273. https://doi.org/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. https://doi.org/10.1109/TGRS.2019.2926772

898