ORNL/LTR-2024/3670

Model Card for PickerXL

Chengping Chai
Derek Rose
Scott Stewart
Nathan Martindale
Mark Adams
Lisa Linville
Christopher Stanley
Anibely Torres Polanco
Philip Bingham

December 2024

DOCUMENT AVAILABILITY

Online Access: US Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via https://www.osti.gov.

The public may also search the National Technical Information Service's National Technical Reports Library (NTRL) for reports not available in digital format.

DOE and DOE contractors should contact DOE's Office of Scientific and Technical Information (OSTI) for reports not currently available in digital format:

US Department of Energy
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831-0062
Telephone: (865) 576-8401
Fax: (865) 576-5728
Email: reports@osti.gov
Website: www.osti.gov

ORNL/LTR-2024/3670

Electrification and Energy Infrastructures Division

MODEL CARD FOR PICKERXL

Chengping Chai
Derek Rose
Scott Stewart
Nathan Martindale
Mark Adams
Lisa Linville
Christopher Stanley
Anibely Torres Polanco
Philip Bingham

December 2024

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

1. MODEL DETAILS

- Model name:
  - PickerXL
- Developed by:
  - Chengping Chai
  - Derek Rose
  - Scott Stewart
  - Nathan Martindale
  - Mark Adams
- Contributed by:
  - Lisa Linville
  - Christopher Stanley
  - Anibely Torres Polanco
  - Philip Bingham
- Model date:
  - June 26, 2024
- Model version:
  - 0.1.0
- Model short description:
  - PickerXL is a large deep learning model to measure arrival times from noisy seismic signals. Measuring arrival times accurately and quickly is challenging for conventional methods. Our model contains considerably more model parameters than published models. PickerXL outperformed a state-of-art model for the same task.
- Model description:
  - This model was trained on the Stanford Earthquake Dataset (STEAD) for primary (P) and secondary (S) wave arrival picking.
- Model type:
  - UNet

- Inputs and outputs:
  - Time series (ground motion) data
  - **Input**—This model was trained with a three-channel time series, each of which contained 5,700 data points with a sampling rate of 100 Hz.
  - **Output**—The output is a batch of probabilities as a function of time. For each input time series, the output probabilities have three channels (5,700 data points). The output probability channels correspond to P waves, S waves, and noise. These probabilities can be used to infer P and S wave arrival times.
- Compute Infrastructure:
  - This model was trained on a workstation.
  - Hardware—This model was trained on one NVIDIA A100 GPU.
  - Software—This model was trained with Lightning (https://github.com/Lightning-AI/pytorch-lightning) and PyTorch (https://github.com/pytorch/pytorch).
- Paper:
  - Title—"PickerXL: A Large Deep Learning Model to Measure Arrival Times from Noisy Seismic Signals"
  - Citation—TBD
  - Link to the paper—TBD
- License:
  - GNU General Public License v3
- Contact:
  - Chengping Chai, chaic@ornl.gov

2. INTENDED USES

The model is intended for measuring P and S waves from seismic waveforms. The model was trained using earthquake data at local distances (0–350 km). The model may not perform well for earthquake data at larger distances or for nonearthquake sources.

- Intended Use:
  - The model can be applied to three-component seismic data recorded by seismic stations at local distances and can measure P and S wave arrival times.
- Primary Intended Users:
  - Data analysts working with seismic data

- Out-of-Scope Use Cases:
    - Applying the model to data other than seismic time series, earthquake data at distances larger than 350 km from the source, and nonearthquake sources

3. HOW TO USE

A software package, PickerXL, is provided for using the trained model.

- Installation Instructions:
    - Use the following command in a terminal to install the package:

      pip install pickerxl

- Training Configuration:
    - NA
- Inference Configuration:
    - Follow the code snippet below for inferencing. An example data file is provided along with the software package.
- Code Snippets of How to Use the Model:

```
import h5py
import numpy as np
from pickerxl.pickerxl import Picker
#
if __name__ == "__main__":
    model = Picker()
    #
    fid = h5py.File("example_waveforms.h5", "r")
    data_group = fid["data"]
    example_data = []
    true_p_index = []
    true_s_index = []
    for akey in data_group.keys():
        dataset = data_group[akey]
        example_data.append(dataset[...])
        true_p_index.append(float(dataset.attrs["p_arrival_sample"]))
        true_s_index.append(float(dataset.attrs["s_arrival_sample"]))
    fid.close()
    #
    example_data = np.array(example_data)
    preds = model.predict_probability(example_data)
    p_index, s_index = model.predict_arrivals(example_data)
    print("True P-wave arrival index:", true_p_index)
    print("Predicted P-wave arrival index:", p_index)
    print("True S-wave arrival index:", true_s_index)
```

```
print("Predicted S-wave arrival index:", s_index)
```

4. BIAS, RISKS, AND LIMITATIONS

The model may have a less-than-optimal performance for earthquake data outside of a source-receiver distance range of 10–110 km and a magnitude range of 0–4.5 because of biases in the training data.

- Known Biases in the Training Data:
  - The source-receiver distance and earthquake magnitude of the training data are not uniformly distributed. Most earthquake data have a source-receiver distance between 10-110 km and a magnitude between 0 and 4.5. Transfer learning has been successfully applied to adapt a pretrained model for different scales [1]. We suggest finetuning the PickerXL model for out-of-distribution earthquake data.
- Examples of Biased Predictions:
  - NA
- Limitations:
  - The model was trained for a sampling rate of 100 Hz and may not perform ideally for other sampling rates.
- Trust Calibration Maturity Model (TCMM) Score:
  - Performance: Level 2. The performance of the model has been evaluated.
  - Bias & Robustness: Level 2. This model was solely pretrained on data from the STEAD dataset. This dataset is not representative of all seismic source types expected in a real-world application scenario; therefore, the statistical patterns that this model learned will not apply equally well to previously unseen source types and to data of rare sources.
  - Transparency: Level 3. Model weights are accessible, and intermediate results can be visualized.
  - Safety & Security: Level 1. The system is designed to be deployed locally. User inputs cannot be accessed by others.
  - Usability: Level 2. User interface exists. Example data and script are provided. User document is included.

5. TRAINING DETAILS

5.1 TRAINING DATA

This study used STEAD [2] to train and evaluate this model because STEAD is among the best benchmark datasets for local to regional data. STEAD is a global dataset that contains over 1 million 60 s long seismic waveforms originated from approximately 450,000 earthquakes and background noise captured by more than 2,500 seismic stations. Each waveform in STEAD was attached with metadata such as earthquake locations, station locations, and signal arrival times when available. This study divided the waveforms into three sets: train, development, and test. The test set consisted exclusively of seismograms originated from earthquakes in the

Oklahoma region or recorded by stations in the same area. The Oklahoma region was selected for the test set because of the considerable number of earthquake signals available. The remaining seismograms were distributed between the train or development sets. Earthquake seismograms and noise waveforms were randomly apportioned such that 95% of waveforms belong to the train set. The train, development, and test sets comprised approximately 950,000, 50,000, and 30,000 earthquake seismograms, respectively, as well as 220,000, 11,000, and 2,000 noise waveforms, respectively.

5.2 TRAINING PROCEDURE

This study trained the model to measure P and S wave arrival times from noisy signals. Noise waveforms have been integrated into earthquake signals as a data augmentation technique for training machine learning pickers. This study adapted this data augmentation technique and generated noisy signals by combining clean signals (original seismograms) with noise waveforms. The amplitudes of clean signals were normalized by the maximum value of each three-component seismogram. This study added 21 randomly selected and amplitude-scaled noise waveforms to each clean signal to increase dataset diversity. Furthermore, a random time shift was applied to each noisy signal to enhance data diversity. The models process 57 s long three-component seismograms at a sampling rate of 100 Hz. The models output P wave, S wave, and noise probabilities for each sampling point on the seismogram. Noisy signals and associated catalog P and S wave arrival times were used to train the models. The noisy signals were constructed by adding a scaled version of noise waveforms to clean signals. Target P wave and S wave probabilities were constructed using truncated Gaussian functions with a width of 0.06 s centered at the P and S wave arrival times. The development set was used to determine model size and training parameters. The optimal model was selected to have the maximum accuracy for the development set. This study used a learning rate of 10^{-6}, a batch size of 256, 50 training epochs, and the Adam optimizer.

6. EVALUATION DETAILS

6.1 EVALUATION DATA

The test set consists exclusively of seismograms originating from earthquakes in the Oklahoma region or recorded by stations in the same area. The test set was extracted from STEAD [2]. The processing speed of the model depends heavily on the hardware and data loading processes. It took around 2 minutes to evaluate the entire test set on our system (see section 7 for hardware specifications), which contains around 164,000 seismograms.

6.2 EVALUATION PROCEDURE

A model prediction is considered a true positive if the absolute difference between the predicted and ground truth arrival times is less than or equal to 0.5 s. In comparison, the absolute difference for the original PhaseNet model was 0.24 s [3]. Mean absolute error (MAE) is calculated as the average of the absolute differences between the predicted signal arrival times using a convolutional neural network (CNN) model and the ground truth arrival times for true positives. If the absolute difference is larger than 0.5 s, the model prediction is considered a false positive. The model predictions are considered false negatives when no arrival times are predicted by a CNN model. Accuracy is calculated as the ratio between the number of true positives to the total waveforms. Recall is computed by dividing the number of true positives by the sum of the number of true positives and false negatives. Precision is calculated by dividing the number of true positives by the sum of the number of true positives and false positives. The F1 score is the harmonic mean of recall and precision.

The model reached a P wave MAE of 0.048 s, an S wave MAE of 0.063 s, a P wave accuracy of 0.94, an S wave accuracy of 0.96, a P wave recall of 1.00, an S wave recall of 1.00, a P wave precision of 0.95, an S wave precision of 0.96, a P wave F1 of 0.97, and an S wave F1 of 0.98.

7. ENVIRONMENTAL IMPACT

- Where the Model was Trained:
  - Oak Ridge, Tennessee
- Hardware Used:
  - One GPU
- Training Type:
  - Pretraining
- Estimated Carbon Footprint
  - **Hardware Type**—One NVIDIA A100 GPU (80G)
  - **Hours Used—**Training one model took 125 h, or roughly 5 days
  - **Cloud Provider**—Standard compute
  - **Compute Region—**Standard energy grid

8. ETHICAL CONSIDERATIONS

- Were any sensitive data used?
  - No
- Any implications for health or safety?
  - See the Disclaimer of Warranty and Limitation of Liability sections of GNU General Public License v3 for more details.

9. MODEL CARD CONTACT

- Chengping Chai, chaic@ornl.gov

10. REFERENCES

1. Chai, C., M. Maceira, H. J. Santos-Villalobos, S. V. Venkatakrishnan, M. Schoenball, W. Zhu, et al. (2020). Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking. *Geophysical Research Letters*, 47(16), e2020GL088651. https://doi.org/10.1029/2020GL088651
2. Mousavi, S. M., Y. Sheng, W. Zhu, and G. C. Beroza. 2019. "STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI." *IEEE Access* 7: 179464–179476. https://doi.org/10.1109/ACCESS.2019.2947848.

3. Zhu, W., and G. C. Beroza. 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.