



## Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers

Jannes Münchmeyer<sup>1</sup>, Jack Woollam<sup>2</sup>, Andreas Rietbrock<sup>2</sup>, Frederik Tilmann<sup>1</sup>, Dietrich Lange<sup>5</sup>, Thomas Bornstein<sup>1</sup>, Tobias Diehl<sup>3</sup>, Carlo Giunchi<sup>4</sup>, Florian Haslinger<sup>3</sup>, Dario Jozinović<sup>6</sup>, Alberto Michelini<sup>6</sup>, Joachim Saul<sup>1</sup>, and Hugo Soto<sup>1</sup>

<sup>1</sup>GFZ German Research Centre for Geosciences, Potsdam, Germany

<sup>2</sup>Geophysical Institute (GPI), Karlsruhe Institute of Technology, Karlsruhe, Germany

<sup>3</sup>Swiss Seismological Service, ETH Zurich, Zurich, Switzerland

<sup>4</sup>Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Pisa, Italy

<sup>5</sup>GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

<sup>6</sup>Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy

Seismic event detection and phase picking are the base of many seismological workflows. In recent years, several publications demonstrated that deep learning approaches significantly outperform classical approaches, achieving human-like performance under certain circumstances. However, as studies differ in the datasets and evaluation tasks, it is yet unclear how the different approaches compare to each other. Furthermore, there are no systematic studies about model performance in cross-domain scenarios, i.e., when applied to data with different characteristics.

Here, we present the results from a large-scale benchmark to address these questions. We compare six previously published deep learning models on eight datasets covering local to teleseismic distances and on three tasks: event detection, phase identification and onset time picking. Furthermore, we compare the results to a classical Baer-Kradolfer picker.

Overall, we observe the best performance for EQTransformer, GPD and PhaseNet, with a small advantage for EQTransformer on teleseismic data. Furthermore, we conduct a cross-domain study, analyzing model performance on datasets they were not trained on. We show that trained models can be transferred between regions with only mild performance degradation, but models trained on regional data do not transfer well to teleseismic data.

As deep learning for detection and picking is a rapidly evolving field, we ensured extensibility of our benchmark by building our code on standardized frameworks and making it openly accessible. This allows model developers to easily evaluate new models or performance on new datasets. Furthermore, we make all trained models available through the SeisBench framework, giving end-users an easy way to apply these models.

T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J. & Soto, H. (2021). Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*. doi.org/10.1029/2021JB023499