

EGU21-12218

<https://doi.org/10.5194/egusphere-egu21-12218>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



## SeisBench: A toolbox for benchmarking and applying machine learning in seismology.

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

<sup>1</sup>Geophysical Institute (GPI), Karlsruhe Institute of Technology, Karlsruhe, Germany (jack.woollam@kit.edu)

<sup>2</sup>Deutsches GeoForschungZentrum (GFZ), Potsdam, Germany

<sup>3</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Rome, Italy

<sup>4</sup>Institute of Geophysics, ETH Zurich, Zurich, Switzerland

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

Machine learning methods have seen widespread adoption within the seismological community in recent years due to their ability to effectively process large amounts of data, while equalling or surpassing the performance of human analysts or classic algorithms. In the wider machine learning world, for example in imaging applications, the open availability of extensive high-quality datasets for training, validation, and the benchmarking of competing algorithms is seen as a vital ingredient to the rapid progress observed throughout the last decade. Within seismology, vast catalogues of labelled data are readily available, but collecting the waveform data for millions of records and assessing the quality of training examples is a time-consuming, tedious process. The natural variability in source processes and seismic wave propagation also presents a critical problem during training. The performance of models trained on different regions, distance and magnitude ranges are not easily comparable. The inability to easily compare and contrast state-of-the-art machine learning-based detection techniques on varying seismic data sets is currently a barrier to further progress within this emerging field. We present SeisBench, an extensible open-source framework for training, benchmarking, and applying machine learning algorithms. SeisBench provides access to various benchmark data sets and models from literature, along with pre-trained model weights, through a unified API. Built to be extensible, and modular, SeisBench allows for the simple addition of new models and data sets, which can be easily interchanged with existing pre-trained models and benchmark data. Standardising the access of varying quality data, and metadata simplifies comparison workflows, enabling the development of more robust machine learning algorithms. We initially focus on phase detection, identification and picking, but the framework is designed to be extended for other purposes, for example direct estimation of event parameters. Users will be able to contribute their own benchmarks and (trained) models. In the future, it will thus be much easier to compare both the performance of new algorithms against published machine learning models/architectures and to check the performance of established algorithms against new data sets. We hope that the ease of validation and inter-model comparison enabled by SeisBench will serve as a catalyst for the development of the next generation of

machine learning techniques within the seismological community. The SeisBench source code will be published with an open license and explicitly encourages community involvement.