Convolutional event embeddings for fast probabilistic earthquake assessment

Jannes Münchmeyer[1], Dino Bindi[1], Ulf Leser[2], and Frederik Tilmann[1]

[1]Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences
[2]Humboldt University of Berlin

November 22, 2022

Abstract

Recent research showed that machine learning, in particular deep learning, can be applied with great success to a multitude of seismological tasks, e.g. phase picking and earthquake localization. One reason is that neural networks can be used as feature extractors, generating generally applicable representations of complex data. We employ a convolutional network to condense earthquake waveforms from a varying set of stations into a high dimensional vector, which we call event embedding. For each event the embedding is calculated from instrument-corrected waveforms beginning at the first P pick and updated continuously with incoming data. We employ event embeddings for real time magnitude estimation, earthquake localization and ground motion prediction, which are central tasks for early warning and for guiding rapid disaster response. We evaluate our model on the IPOC catalog for Northern Chile, containing 100,000 events with low uncertainty hypocenters and magnitude estimates. We split the catalog sequentially into a training and a test set, with the 2014 Iquique event (Mw 8.1) and its fore- and aftershocks contained in the test set. Following preliminary results the system achieves a test RMSE of 0.28 magnitude units (m.u.) and 35 km hypocentral distance 1 s after the first P arrival at the nearest station, which improves to 0.17 m.u. and 22 km after 5 s and 0.11 m.u. and 15 km after 25 s. As applications in the hazard domain require proper uncertainty estimates, we propose a probabilistic model using Gaussian mixture density networks. By analyzing the predictions in terms of their calibration, we show that the model exhibits overconfidence i.e. overly optimistic confidence intervals. We show that deep ensembles substantially improve calibration. To assess the limitations of our model and elucidate the pitfalls of machine learning for early warning in general, we conduct an error analysis and discuss mitigation strategies. Despite the size of our catalog, we observe issues with two kinds of data sparsity. First, we observe increased residuals for the source parameters of the largest events, as training data for these events is scarce. Second, similar inaccuracies occur in areas without events of a certain size in the training catalog. We investigate the impact of these limitations on the Iquique fore- and aftershocks.

HEIBRiDS
Helmholtz Einstein International Berlin Research School in Data Science

HELMHOLTZ CENTRE POTSDAM
GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES

Convolutional event embeddings for fast probabilistic earthquake assessment

Jannes Münchmeyer[1,2], Dino Bindi[1], Ulf Leser[2], Frederik Tilmann[1,3]
[1] GeoForschungsZentrum Potsdam [2] Humboldt-Universität zu Berlin [3] Freie Universität Berlin

HUMBOLDT-UNIVERSITÄT ZU BERLIN
GFZ Helmholtz Centre POTSDAM

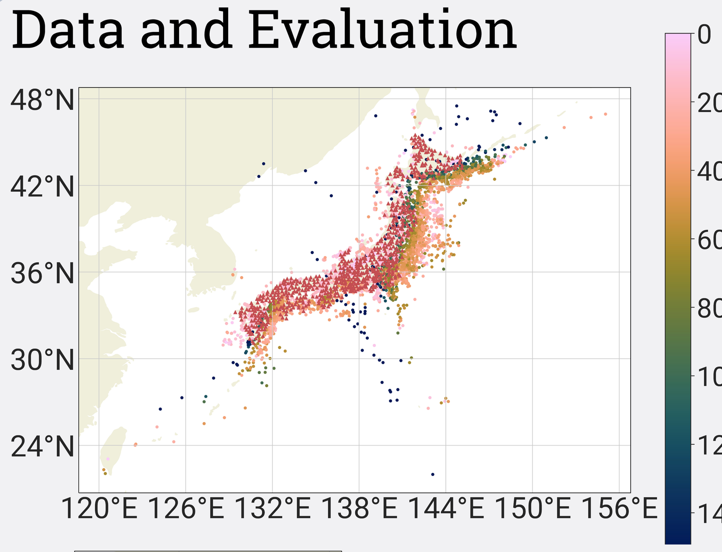
S43E-0698

Abstract

Timely and accurate earthquake source parameter estimates are essential for early warning. Classical parametric models suffer from simplified assumptions and discard information. We use a deep learning model directly on the waveforms to alleviate these issues. A key idea of our model is to represent events as vectors that are independent of the specific set of contributing stations and the time. We call these representations event embeddings.

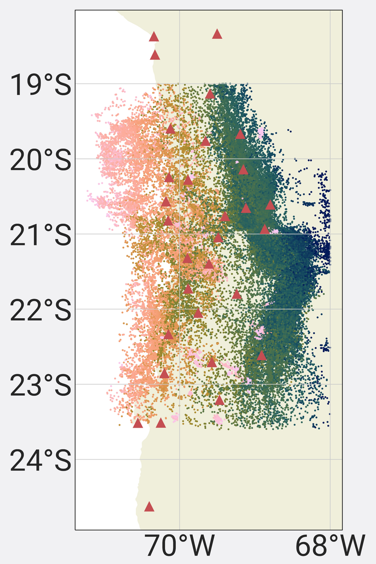
We compare our model to a Bayesian peak displacement baseline on two catalogs from Japan and Chile. On both catalogs our model achieves a higher precision 2 s after the first P arrival than the baseline after 8 s. After 8 s our model has a 50 % lower RMSE.

Data and Evaluation



	Chile	Japan
Events	96 132	8 258
Records	1 785 434	322 205
Spatial extent	4.5° x 3.5°	25.5° x 34.5°
Magnitude range	M_A 1.2 - 8.1	M_W 3.5 - 8.7
Stations	29	696
Station density	17 km	37 km
Years	2007 - 2014	1997 - 2018
Sensors	Broadband & Strong motion	Strong motion
Correction	Instrument response	Gain
Recordings	Continuous	Trigger based
Borehole sensors	None	All stations

Table 1: Characteristics of the two data sets used for evaluation. Station density is given as median inter-station distance.



We consider two test data sets in Northern Chile and Japan. We split both data sets into training, development and test sets with a ratio of 60:10:30. The test set for Chile consists of all events after 07/2012, for Japan after 02/2013. We use the development set for model selection. All results reported on this poster are test set results.

Figure 1: Events (dots) and stations (triangles) in the Japan (top) and Chile (bottom) catalogs.

Catalogs from Sippl et al. (2018)/Münchmeyer et al. (2019) and NIED F-net. Waveforms from NIED KIK-net, IPOC, GEOFON, CSN, WestFissure, Iquique and Minas networks.

Parametric magnitude estimation

Most source based early warning methods rely on waveform features and parametric models. Typical features are peak displacement (Lancieri and Zollo 2008), energy (Festa et al. 2008), or dominant period (Allen 2007). We use the Bayesian peak displacement method by Lancieri and Zollo (2008) as baseline, as it performed best on our data sets.

Parametric models suffer from imprecise assumptions, saturation effects for large magnitudes and errors in location estimates. Location errors are especially pronounced for estimates with only few contributing stations. For the baseline experiments we use the final hypocenter estimate, which is overly optimistic as it is only avaible after P wave detections at a sufficient number of stations.

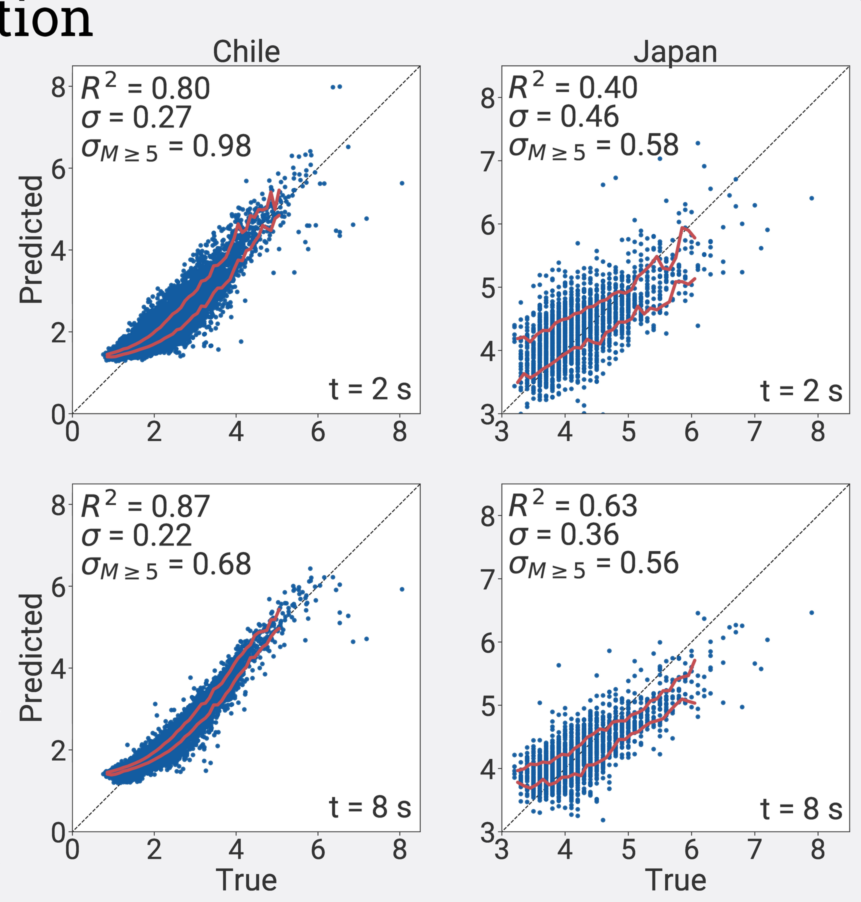


Figure 2: Early magnitude estimates at 2 s and 8 s after the first P arrival using the Bayesian peak displacement method from Lancieri and Zollo (2008). The lines indicate moving 20[th] and 80[th] percentiles.

Method

Our model consists of a feature extraction, a feature combination and an estimation step for magnitude and location. The feature extraction is a convolutional neural network independently applied to each station. The feature combination is a transformer model (Vaswani et al. 2017), integrating information from a variable set of stations. The estimators are mixture density networks, outputting probabilistic estimates for magnitude and location. We train the model end-to-end.

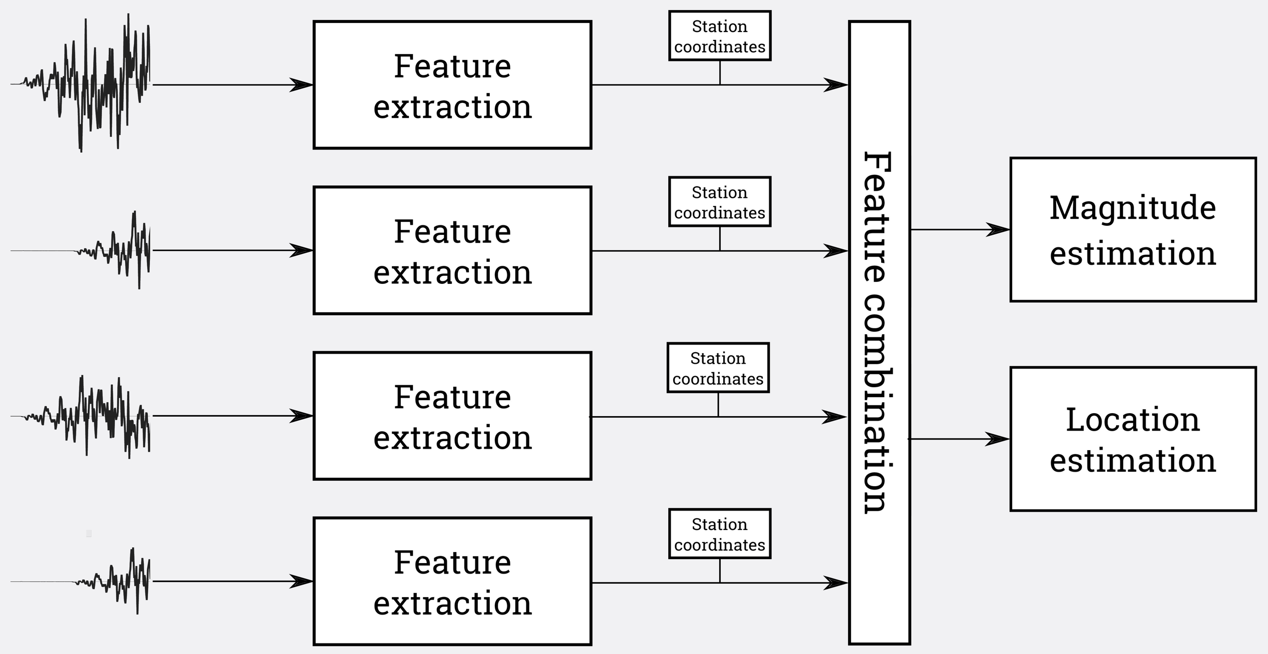


Figure 3: Schematic overview of the fast assessment model. Each row corresponds to one station. Boxes represent neural networks, arrows represent data vectors.

The output of the feature combination is a vector representation of the event, which we call event embedding. It is independent of the contributing stations and the time. Therefore event embeddings can be used as features for predicting event properties like magnitude and location.

Probabilistic predictions

At each time after the first P arrival (t = 0 s) our model outputs a distribution over magnitudes. This allows to evaluate the model's response to real time data.

The model input always consists of 30 s long waveform samples at 100 Hz. The samples start 5 seconds before the first P arrival at any station. To simulate real time data, all measurements after a given time are set to zero. This data augmentation is applied both in training and evaluation of the model. As the Japanese data set is trigger based, there we always blind out stations until their respective trigger time.

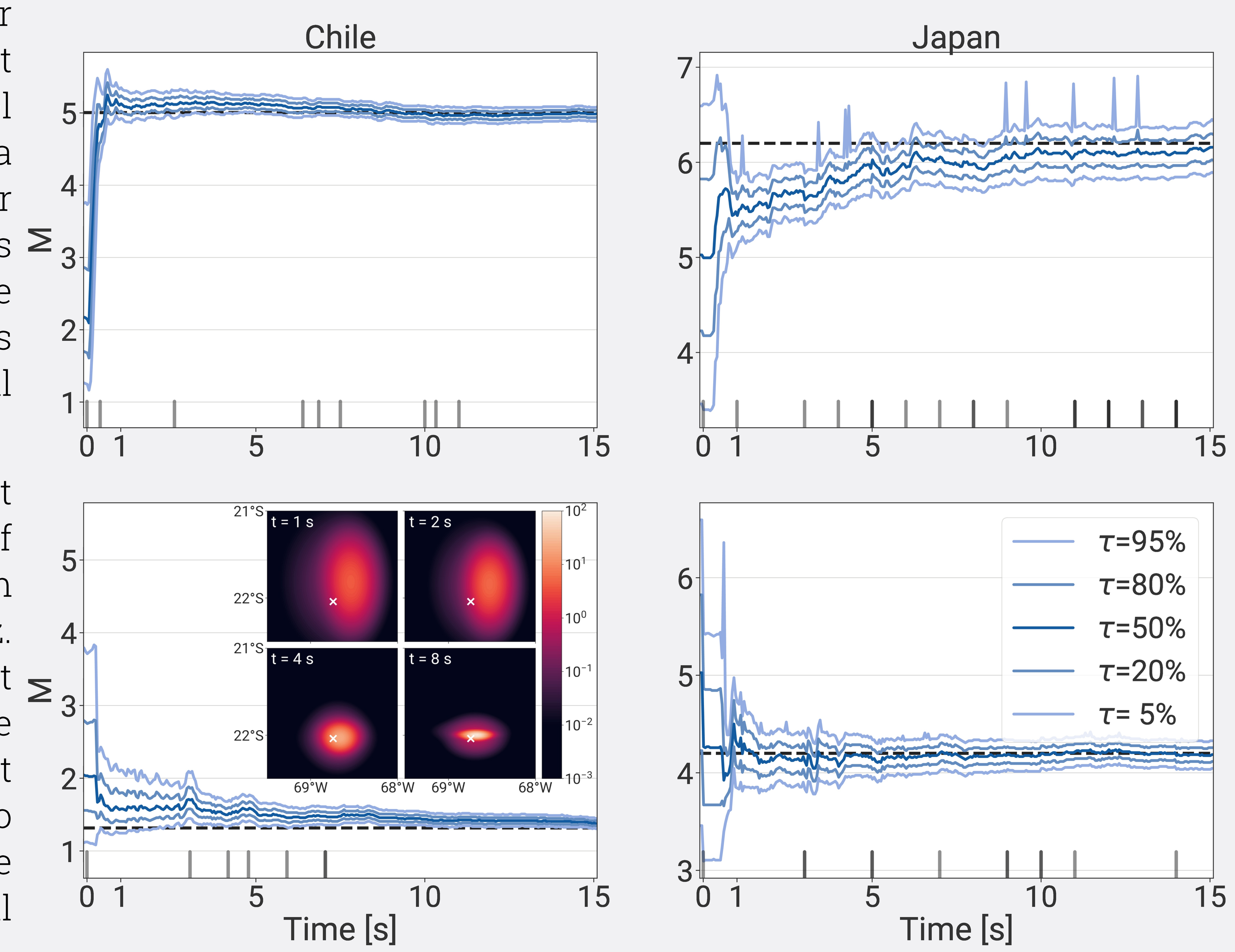


Figure 4: Predicted magnitude distributions for four sample events indicated by quantile lines. The additional ticks above the x axis show approximate P arrivals at contributing stations for interpretation purposes. These picks were not used by the algorithm. The events are: a large (M_A = 5.0), deep (150 km) on-shore event in Chile (top left); a deep (110 km) but small (M_A = 1.3) event colocated with the one before (bottom left); a large (M_w = 6.2) shallow (20 km) event offshore (100 km) the Japanese east coast (top right); an intermediate size (M_w = 4.2) event underneath central Honshu at 47 km depth (bottom right). The inset shows the evolution of the location prediction over time through its likelihood.

Results

The proposed model improves precision and timeliness in both data sets compared to the parametric estimates. This holds true in terms of both R^2 and RMSE. The quality of predictions is generally better for the Chile dataset, likely caused by the more homogeneous data set and lower amount of offshore events.

Large magnitudes are generally underestimated. For events with a magnitude above 5.0 we see an increased RMSE for the Chile catalog compared to the baseline. For the Japan catalog, we still see a decrease in RMSE compared to the baseline, but it is lower than the average decrease. Underestimation occurs already for smaller magnitudes in the Chile catalog than in the Japan one.

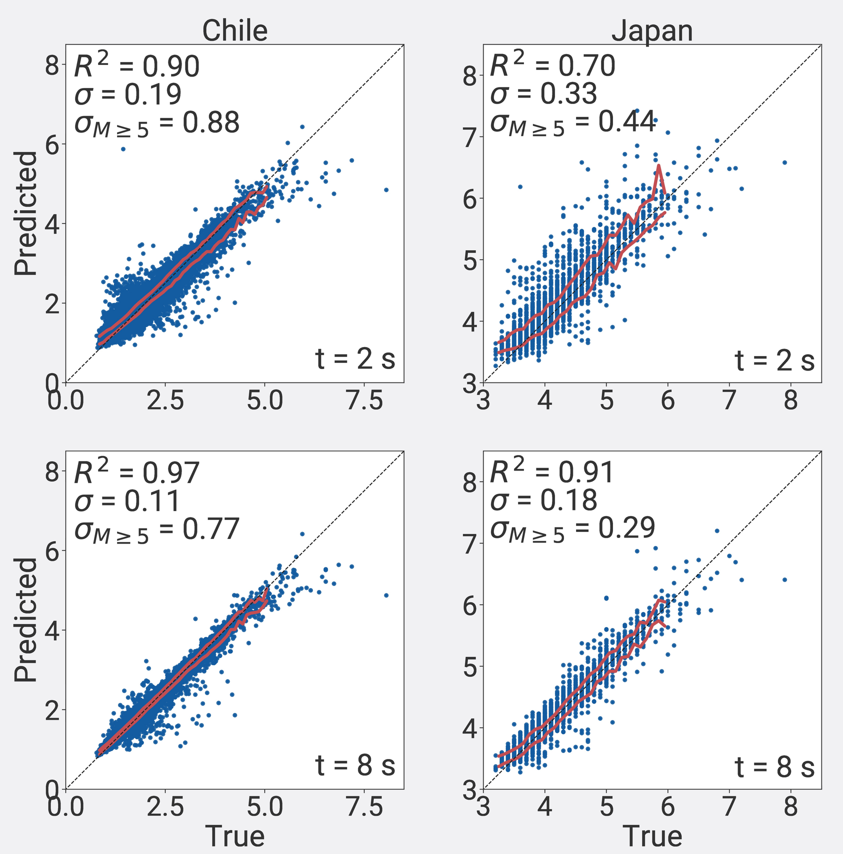


Figure 5: Mean magnitude predictions at different times after the first P arrival (t = 0 s). The lines indicate moving 20[th] and 80[th] percentiles.

Transfer learning

We assume the underestimation of large magnitude events is caused by data sparsity. Therefore we trained an additional transfer learning model for the Chile data set by first training on the combination of both catalogs and then finetuning on the Chile one. This increases the number of large events available in the training set and reduces the severity of underestimation. On the other hand we see a slight degradation for the predictive performance at small magnitudes, while still clearly outperforming the baseline.

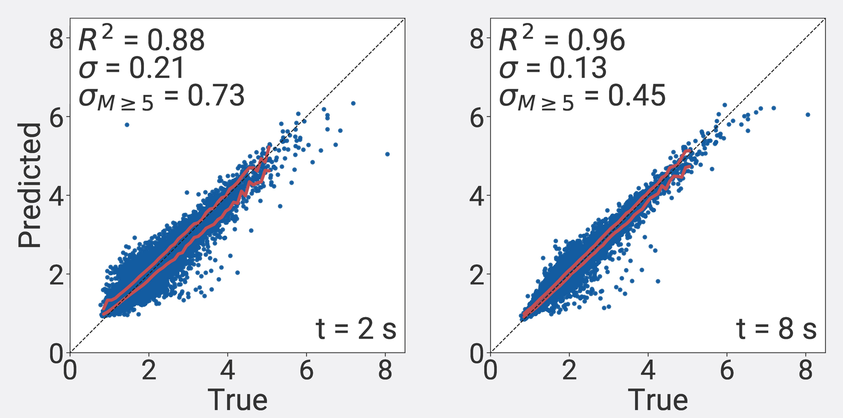


Figure 6: Mean predictions for the Chile catalog using transfer learning. The lines indicate moving 20[th] and 80[th] percentiles.

Conclusion

- We built an end to end model to estimate magnitude and location in real time after the first P arrival. The model provides probabilistic estimates.
- Our model achieves the same precision 2 s after the first P arrival as a baseline method after 8 s.
- Transfer learning form the Japan catalog to the Chile catalog reduces RMSE for large magnitude events (M>5) by 40 %.
- We propose the event embedding, a vector representation for events that is independent of time and the specific set of contributing stations.

References

Allen, R. M. (2007). The ElarmS earthquake early warning methodology and application across California. In Earthquake Early Warning Systems.
Festa, G., Zollo, A., & Lancieri, M. (2008). Earthquake magnitude estimation from early radiated energy. Geophysical research letters.
Lancieri, M., & Zollo, A. (2008). A Bayesian approach to the real-time estimation of magnitude from the early P and S wave displacement peaks. Journal of Geophysical Research: Solid Earth.
Münchmeyer, J., Bindi, D., Sippl, C., Leser, U., & Tilmann, F. (2019). Low uncertainty multifeature magnitude estimation with 3-D corrections and boosting tree regression: application to North Chile. Geophysical Journal International.
Sippl, C., Schurr, B., Asch, G., & Kummerow, J. (2018). Seismicity structure of the Northern Chile forearc from> 100,000 double-difference relocated hypocenters. Journal of Geophysical Research: Solid Earth.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł, & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems.

Acknowledgements

Jannes Münchmeyer acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS). We thank IPOC and NIED for providing the data sets.

Contact: munchmej@gfz-potsdam.de munchmeyer.de github.com/yetinam

www.gfz-potsdam.de

HELMHOLTZ