geoconvention
Calgary • Canada • May 11-13 2026

Importance of parameter tuning for earthquake detectors

Serafim Grubas, Mirko van der Baan
Department of Physics, University of Alberta.

Summary

Parameters of any algorithm in any problem must be properly tuned to achieve optimal performance. The conventional trial-and-error approach is inefficient and biased. The exhaustive grid search is computationally expensive. We propose a parameter autotuning strategy based on the Optuna framework and multi-objective optimization to efficiently explore the parameter space. We focus on earthquake detection and compare multiple traditional and deep learning detectors on the TXED benchmark dataset, which contains over half a million 3C waveforms. Each detector is autotuned with the proposed strategy on a few TXED waveforms to ensure optimal performance and a fair comparison. The performance evaluation on the full TXED revealed that PhaseNet achieved the highest F1 score, followed by CATS with a small difference, then PSD, STA/LTA, and EQTransformer. An important observation is that the simple and very fast STA/LTA outperformed the sophisticated and slow EQTransformer. Thus, the comparison highlights that simple traditional algorithms are no worse than sophisticated deep-learning algorithms for earthquake detection, especially when the parameters are properly tuned. The autotuning strategy can be applied to any other geophysical problem with objective performance criteria.

Method

We use the *Optuna* framework for autotuning (Akiba et al., 2019) because it supports multi-objective optimization algorithms. *Optuna* is efficient as it leverages a *define-by-run* paradigm, enabling dynamic search spaces and requiring minimal code.

Autotuning is akin to training in deep learning: it requires a tuning dataset and performance metrics. An algorithm is evaluated on the tuning dataset iteratively for *N* iterations to update its parameters. The parameters are updated by a parameter sampler (optimization algorithm). The key steps of the autotuning include:

1. Randomly initialize the algorithm's parameters.
2. Apply the algorithm to the tuning data.
3. Evaluate the results by calculating a set of metrics (e.g. error and runtime).
4. Update the parameters using a parameter sampler.
5. Repeat steps 2-4 *N* times.

We consider the earthquake detection problem and compare five detectors: PhaseNet (Zhu & Beroza, 2019), EQTransformer (Mousavi et al., 2020), STA/LTA (Allen, 1978), PSD (Vaezi & Van der Baan, 2015), and CATS (Grubas et al., 2023). The parameters of each detector will be autotuned using the outlined strategy above.

In step 2, we use ten waveforms (6 earthquakes and 4 noise waveforms) from the TXED dataset (Chen et al., 2023) as the tuning data. These waveforms will be contaminated with white noise for the data augmentation. In step 3, we focus on optimizing recall, precision, and runtime (all three simultaneously). In step 4, we use NSGA-III (Deb & Jain, 2019; Jain & Deb, 2019), which supports multi-objective optimization to optimize recall, precision, and runtime. In step 5, we run the autouning for *N*=2000 iterations.

Results

The result of the autotuning is a set of parameter combinations (the Pareto front) for each detector, optimizing all three metrics. While recall and precision should be maximized (1 is the maximum), runtime should be minimized. The Pareto front contains various parameter combinations: some maximize only precision, others only recall, and some provide a trade-off between the metrics. To select the best parameters (best trade-off), we manually reviewed and validated the Pareto front for each detector. The validation on the full TXED (a half-million waveforms) showed that the best performance is achieved when precision is prioritized. Only EQTransformer performed better when recall and precision are balanced.

The final evaluation metric is F_β score:

$$F_\beta = (1 + \beta^2)\frac{Pr \cdot Re}{\beta^2 \cdot Pr + Re},$$

where Pr, Re are precision and recall, respectively, and β is a weighting factor. F_1 represents the score when recall and precision are considered equally, where $F_{0.5}$ assumes that precision is twice as important as recall.

We evaluated F_1 and $F_{0.5}$ scores for each detector on the full TXED data with the best selected parameters. The results are shown in Fig. 1. One can see that PhaseNet has the highest F_1 score (leftmost orange bar), whereas PSD has the largest $F_{0.5}$ (middle green bar). We use F_1 as the primary metric because it balances the recall and precision. Thus, the most accurate detector is PhaseNet ($F_1 = 0.936$), followed by CATS with a small difference ($F_1 = 0.932$). Then, PSD ($F_1 = 0.916$), STA/LTA ($F_1 = 0.912$), and finally EQTransformer ($F_1 = 0.887$).

Discussion and Conclusions

The proposed autotuning strategy helped select optimal parameters. However, it did not give the best parameters blindly, and we had to review the autotuning results to identify optimal trade-offs. Despite that, the autotuning strategy remains less labour-intensive and more objective than trial-and-error for large datasets such as TXED.

Our final ranking of the detectors indicates that STA/LTA outperformed the pre-trained EQTransformer, which may seem controversial at first glance. However, the authors of EQTransformer (Mousavi et al., 2020) observed a similar pattern: STA/LTA outperformed other pre-trained deep learning detectors in their work. Additionally, it is known that pre-trained models may exhibit significant performance variation in cross-domain tests: when trained on one dataset and tested on another one (Münchmeyer et al., 2022). Thus, our comparison does not contradict previous studies and shows that simpler methods, such as PSD and especially STA/LTA, are on par with deep-learning detectors, whereas CATS demonstrates the accuracy of the top-performing PhaseNet.

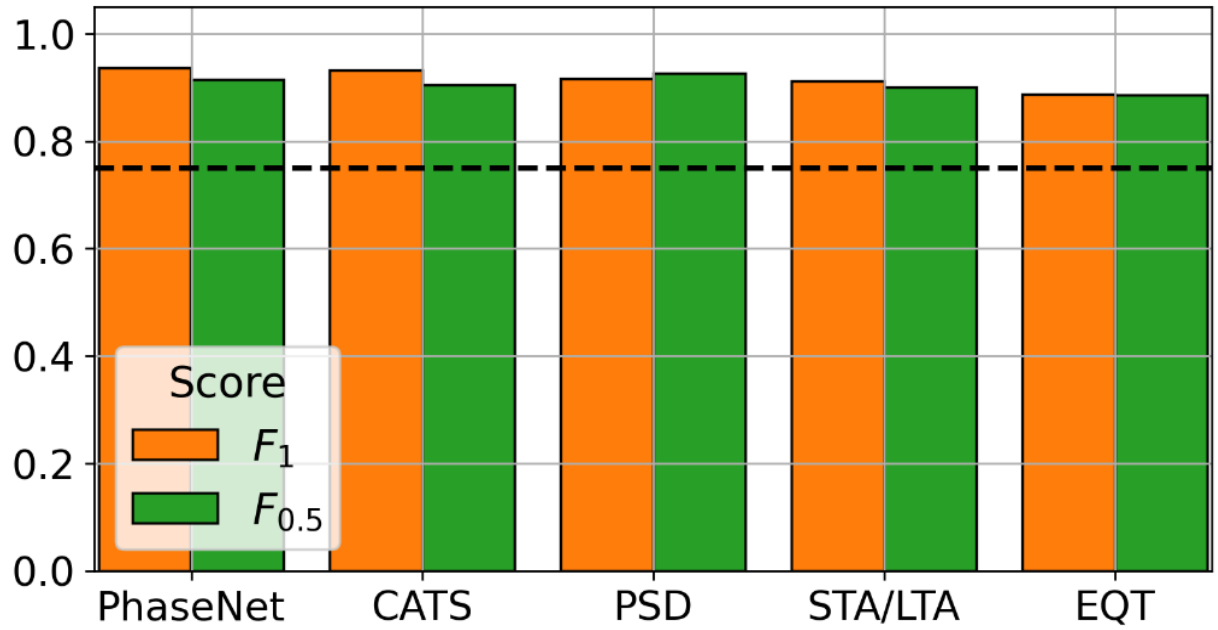


Figure 1. The bar chart with evaluated F_1 (orange) and $F_{0.5}$ (green) scores on the full TXED. The detectors are sorted horizontally from left to right by decreasing F_1 score. The vertical axis is the score value. The black-dashed line shows the baseline performance when all TXED waveforms are treated as earthquakes.

Novel Information

Simple, traditional algorithms are no worse than sophisticated deep learning algorithms for earthquake detection, especially when the parameters are properly tuned.

Acknowledgements

The authors would like to thank the sponsors of the Consortium for Distributed and Passive Sensing for supporting this research.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, July). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623-2631).

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the seismological society of America*, *68*(5), 1521-1532.

Deb, K., & Jain, H. (2013). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. *IEEE transactions on evolutionary computation*, *18*(4), 577-601.

Jain, H., & Deb, K. (2013). An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: Handling constraints and extending to an adaptive approach. *IEEE Transactions on evolutionary computation*, *18*(4), 602-622.

Grubas S., Van der Baan M. and Dokht R. (2023) Seismic event detection based on clustering of trimmed spectrograms. Int. Mtg. Applied Geoscience and Energy, IMAGE 23, SEG and AAPG, Houston.

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.

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., ... & Soto, H. (2022). Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, *127*(1), e2021JB023499.

Vaezi, Y., & Van der Baan, M. (2015). Comparison of the STA/LTA and power spectral density methods for microseismic event detection. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, *203*(3), 1896-1908.

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