OPTIMIZING NEURAL NETWORK ARCHITECTURES AND USING CLUSTERING TO DETECT SEISMIC EVENTS IN NOISY OCEAN BOTTOM SEISMOMETER DATA

EGU General Assembly 2025

Abstract   Related work

Alicia Ximena Cortés-Barajas (ximenacortesba@gmail.com)[1], Marco Calò[2], Erik Molino Minero Re[3], Francesca Di Luccio[4]

1. *PCIC, UNAM, Mexico; 2. Instituto de Geofísica, UNAM, Mexico; 3. IIMAS, UNAM. Mexico, 4. INGV, Italy*

Introduction

In this work we focus on the TYrrhenian Deep-sea Experiment (TYDE), carried out from December 2000 to May 2001 in the southern Tyrrhenian Sea with 14 Guralp CMG40T OBS/H stations and differential pressure gauges [1]

Although valuable, much of the TYDE data remained underexploited at the time due to technological limitations.

Our approach uses evolutionary optimization to design an optimal neural network architecture tailored to this OBS data. The optimized network rapidly scans large amounts of raw seismic recordings, selecting a smaller set of candidate windows likely to contain seismic events. With only 61 manually picked events available for training, we employ PickBlue[2] to accurately pick seismic phases from these candidates. Finally, a Self-Organizing Map (SOM) clusters the events, helping to distinguish different types of seismic activity and reduce false positives.

This integrated method improves detection accuracy, reduces manual workload, and highlights the potential of reanalyzing legacy datasets like TYDE using modern machine learning techniques.

Objectives

General Objective

Develop an automated system for detecting and classifying seismic events in noisy OBS data, combining optimized neural networks, precise phase picking, and unsupervised clustering.

Specific Goals

- 🧠 Optimize a neural network architecture using evolutionary algorithms to adapt to noisy seismic data.
- 🌀 Preprocess OBS time series with wavelet transform and PCA to extract relevant features efficiently.
- ⚡ Filter candidate event windows using the optimized network to reduce the dataset by ~80%.
- 📍 Refine event detection with PhaseNet as a second filter for verifying seismic phases in the candidate windows.
- 🔍 Cluster seismic events with a Self-Organizing Map (SOM) to identify different event types and reduce false positives.

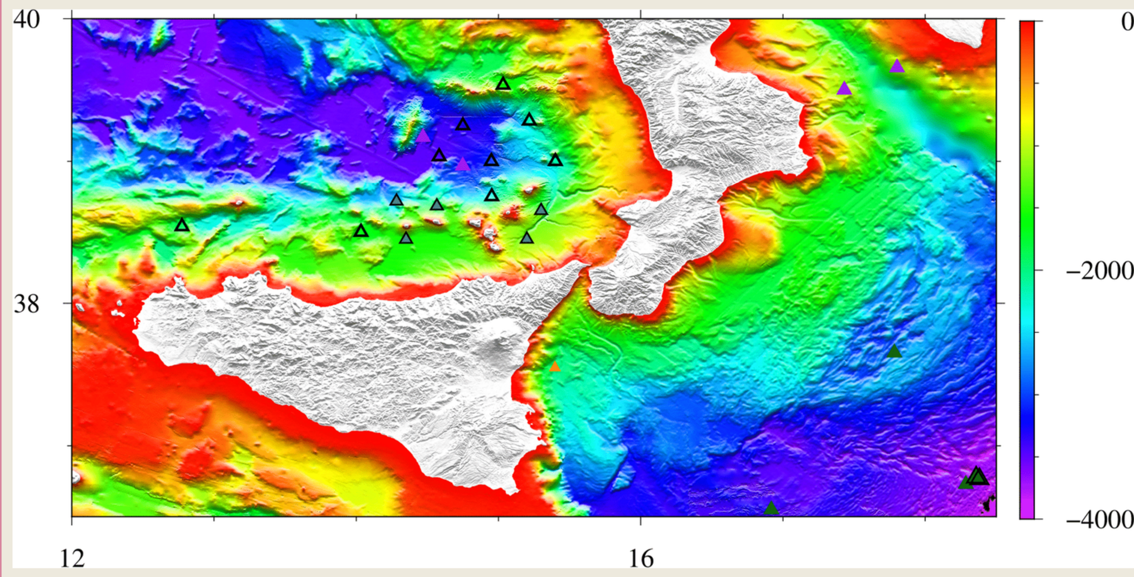


Figure 1. Bathymetric map of the southern Tyrrhenian Sea showing the location of the OBS and OBH stations deployed around the Aeolian Islands for the TYDE experiment.

Methodology

Preprocessing

The raw OBS signal is segmented into overlapping 120-second windows. We apply a Discrete Wavelet Transform (DWT)[3] to extract time–frequency features. These features are normalized and then reduced in dimensionality using Principal Component Analysis (PCA)[4], retaining 90% of the variance while significantly compressing the input size for the neural network.

Evolutionary Neural Architecture Search

We built a custom ENAS system using NSGA-III [5] to design neural networks tailored to our data. It automatically adjusts the **number of layers, neurons, activation functions and connections** in order to maximize true positives and negatives, while minimizing complexity.

Once trained, the selected network scans large datasets quickly to identify candidate events.

Phase Picking and Clustering

Candidate windows selected by the neural network are processed with PickBue to refine the detection by picking seismic phases.

We then apply a Self-Organizing Map (SOM) to cluster similar seismic signals. This step helps identify different types of events and reduces false positives by grouping outliers separately.

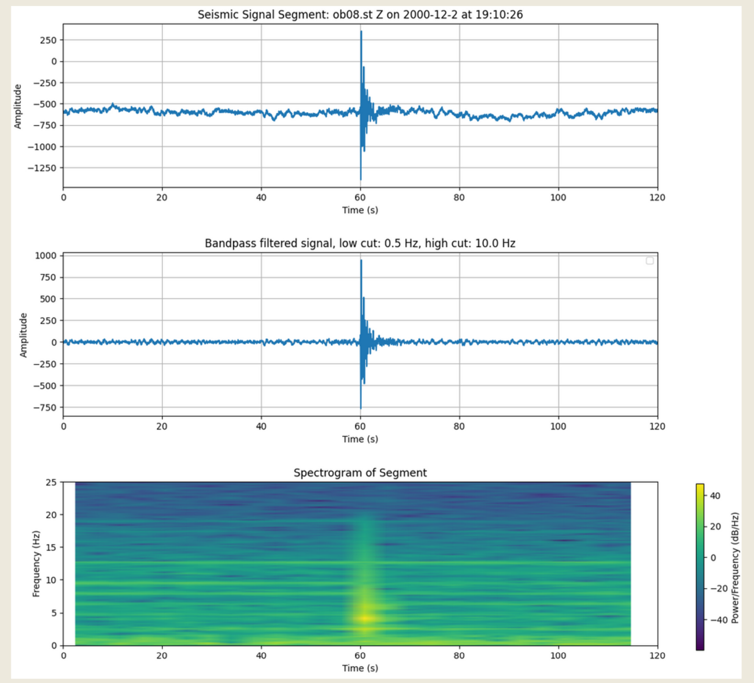
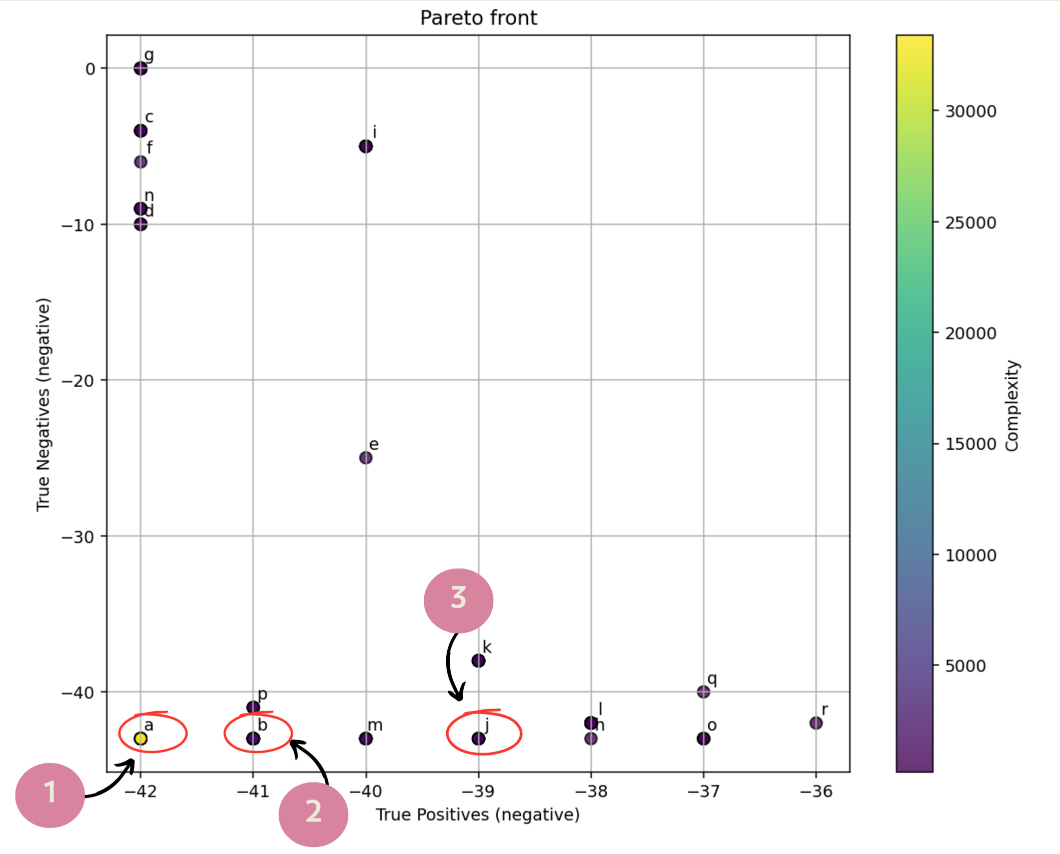


Figure 2. Example of a raw seismic window, bandpass-filtered signal (0.5–10 Hz), and its corresponding spectrogram used for feature extraction.

Figure 3. Pareto front obtained during evolutionary optimization. Each point represents a candidate network, balancing true positive and true negative detections against model complexity. The signaled points are the ones corresponding to the respective architectures on the results section.



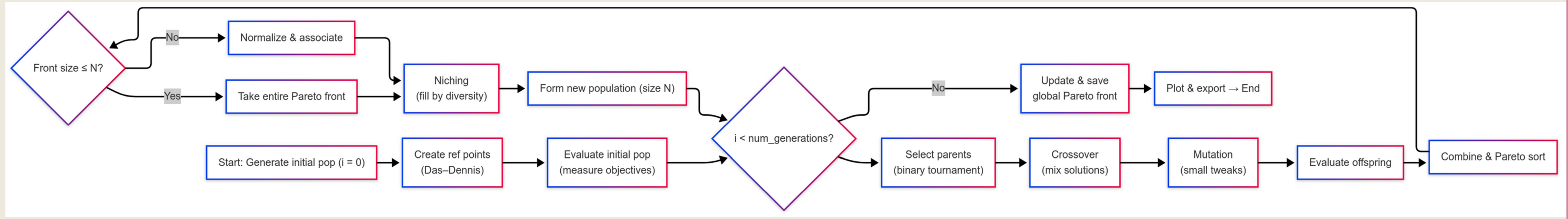
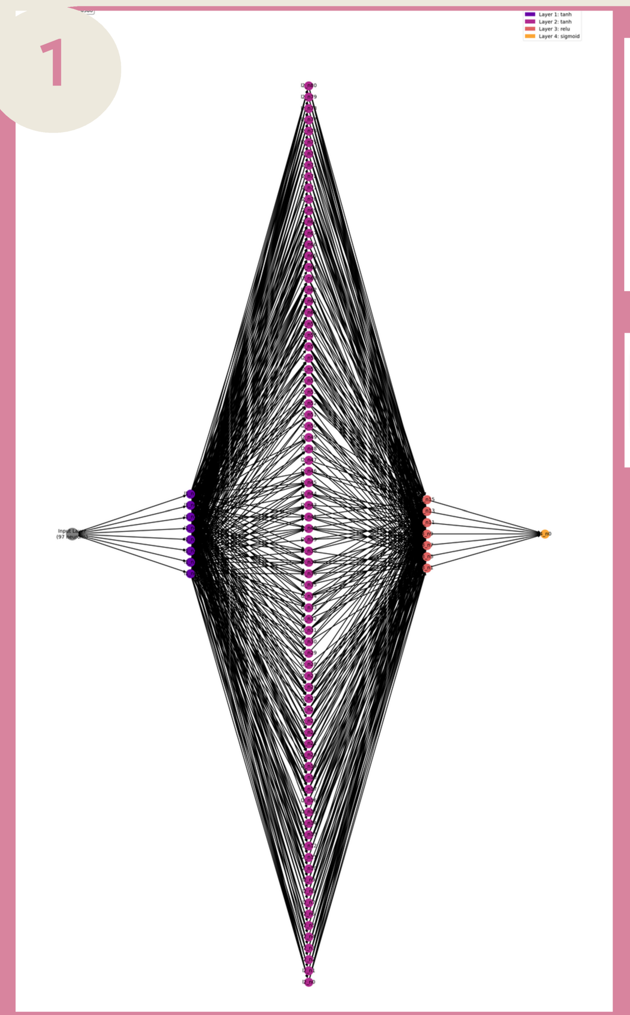


Figure 4. Flowchart of the implemented ENAS using NSGA-III, showing (1) population initialization with Das–Dennis reference points, (2) iterative loop of tournament selection, crossover and mutation, (3) non-dominated sorting and NSGA-III's normalization + niching to preserve diversity, and (4) final Pareto-front update and export.

Preliminary Results



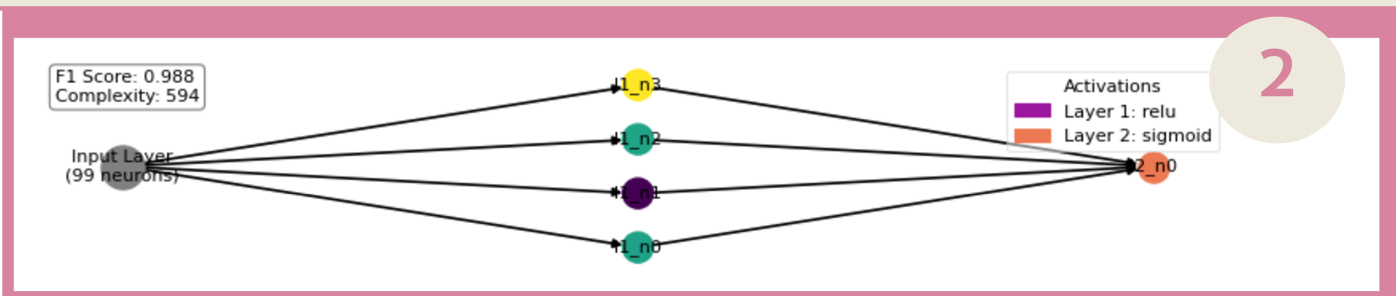


figure 5. Examples of optimized architectures: (left) most complex, (top right) most balanced, (bottom right) simplest

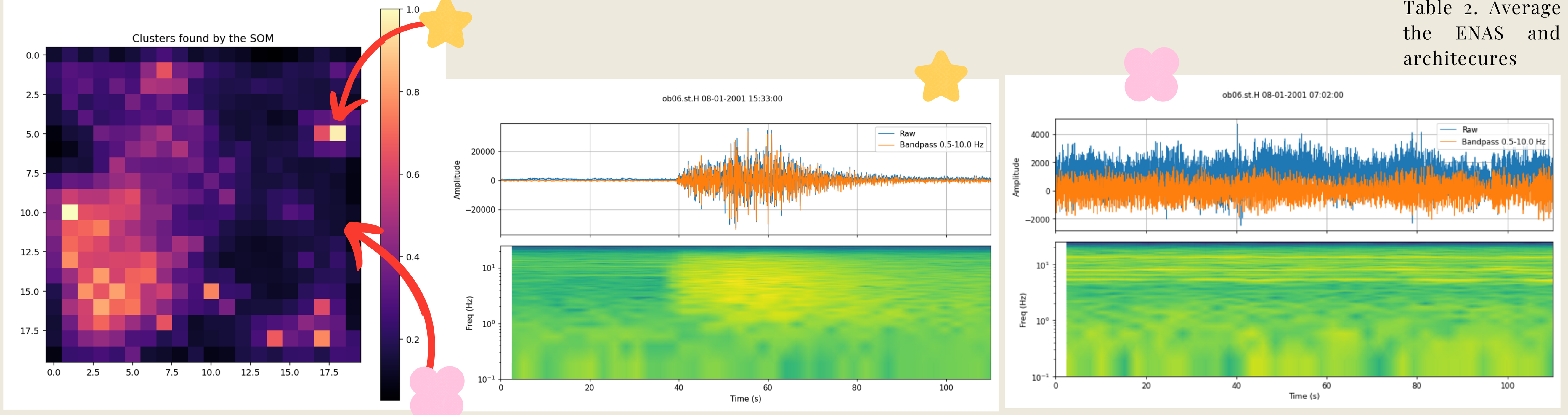
Table 1. Performance of 3 different architectures given by the ENAS

	F1 score	Complexity	Training time [s]	Reduced Data
1	100%	18580	0.1126	15.76%
2	98.8%	594	0.1136	7.77%
3	91.1%	186	0.0352	20.26%

Metric	Result
Windows (5 days)	222941
Windows selected	14.55% (85.45% reduction)
True positives (avg)	44.87 / 61
False positives (avg)	16.13 / 61
NAS optimization time	57.7 ± 38.2 s
Training time (selected net)	0.29 ± 0.28 s
Avg. network complexity	1503.76 ± 1528.97 connections

Table 2. Average performance of the ENAS and it's resulting architecures

Figure 6 (below): Left: U-Matrix visualization of the Self-Organizing Map (SOM), highlighting similarity patterns among candidate seismic events. Middle: Raw signal, filtered signas, and spectrogram of an event grouped in cluster (18,4). Left:Raw signal, filtered signas, and spectrogram of raw signal grouped in cluster (17,10)



Conclusions

Automated neural architecture search combined with clustering techniques enabled efficient detection of seismic events in noisy OBS data, even with limited labeled examples.

This approach reduced manual workload, improved detection reliability, and offers a complementary tool for traditional seismic monitoring.

References

[1] Sgroi, T., Braun, T., Dahm, T., & Frugoni, F. (2006). An improved seismicity picture of the Southern Tyrrhenian area by the use of OBS and land-based networks: the TYDE experiment. Annals of Geophysics, 49(2/3), 765–778.[

[2] Bornstein, T., Lange, D., Münchmeyer, J., Woollam, J., Rietbrock, A., Barcheck, G., Grevemeyer, I., & Tilmann, F. (2023). PickBlue: Seismic phase picking for ocean bottom seismometers with deep learning. arXiv preprint, arXiv:2304.06635. https://arxiv.org/abs/2304.06635

[3] Daubechies, I. (1992). Ten Lectures on Wavelets. SIAM. https://doi.org/10.1137/1.9781611970104

[4]Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

[5] Deb, K., & Jain, H. (2014). 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. https://doi.org/10.1109/tevc.2013.2281535

[6] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43(1), 59–69. https://doi.org/10.1007/BF00337288