IL NUOVO CIMENTO **48 C** (2025) 225
DOI 10.1393/ncc/i2025-25225-8

COMMUNICATIONS: SIF Congress 2024

Seismic data analysis in the Campi Flegrei caldera with Self-Organizing Map (SOM) and Multiscale Entropy (MSE)

A. GRIMALDI([1]), O. AMOROSO([1]), F. NAPOLITANO([1]), S. SCARPETTA([1]), G. MESSUTI([1]), V. CONVERTITO([2]), D. GALLUZZO([2]), G. GAUDIOSI([2]), L. NARDONE([2]) and P. CAPUANO([1])

([1]) *Dipartimento di Fisica, Università degli Studi di Salerno - Fisciano, Italy*
([2]) *Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano - Napoli, Italy*

received 24 January 2025

Summary. — This work proposes an innovative technique for seismic signal analysis in the Campi Flegrei caldera. The goal is to combine the Multiscale Entropy (MSE) algorithm and the Self-Organizing Map (SOM) network to address the challenge of detecting small earthquakes, which is challenging due to low signal-to-noise ratio and large, complex datasets. MSE measures data complexity across different time scales, while SOM enables unsupervised clustering for efficient interpretation of seismic data with reduced human bias. Our approach employs a four-month dataset from the V0102 station in the Pisciarelli area. By integrating these techniques, seismic data can be effectively clustered and interpreted, reducing human bias and enhancing the identification of key seismic events. The findings suggest that this methodology could prove beneficial for earthquake detection and classification within a given dataset.

1. – Introduction

The increasing amount of data in seismology has led to the use of cutting-edge machine learning techniques [1]. The application of supervised and unsupervised learning methods is crucial for efficiently analyzing and interpreting seismic signals, allowing for more accurate classification and detection of seismic events. Among these, the Self-Organizing Map (SOM), an unsupervised learning algorithm, stands out for its ability to handle large volumes of data and uncover meaningful characterization through clustering [2]. This work evaluates the performance of an unsupervised strategy based on the SOM algorithm and Multiscale Entropy (MSE) features extraction technique for processing seismic signals recorded at Campi Flegrei, emphasizing its ability to catalog and cluster earthquakes. Multiscale Entropy (MSE) [3], a complexity measure widely used in biomedical analysis [4], is applied and evaluated as a feature for SOM training to enhance mapping accuracy. Using signals from station V0102 in the Agnano-Pisciarelli area located in the core of the Campi Flegrei caldera, features are extracted, dimensionality is highly reduced, and the data are subjected to unsupervised training with SOM to uncover valuable insights for both the pre-processing and the post-processing phases.

Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0)

2. – Data and methods

2˙1. *Dataset.* – For this study, we utilized the continuous recordings from seismic station V0102 located in Pisciarelli, in the Campi Flegrei Caldera, from July 20th to November 12th, 2023, a period of intense seismic activity in the region. The data, sampled at a rate of 200 Hz, was segmented into 1440 one-minute windows per day, resulting in a dataset of 165000 seismic windows. A comprehensive catalog of seismic events from the wider Campi Flegrei area during this timeframe has been used as a reference for validating the seismic signals extracted from the continuous data.

2˙2. *Self-Organizing Map.* – Unsupervised machine learning algorithms have emerged as a fundamental tool in seismology, particularly for clustering and event classification [5], data visualization [6] and mislabeled polarity detection [7,8]. Unlike supervised methods, which require large sets of labeled data, unsupervised learning methods analyze seismic data to uncover similarities or correlations without prior labels or classifications. This approach is particularly useful for managing the large, complex, and high-dimensional volumes of seismic data that are typical of modern research in Seismology. Self-Organizing Map (SOM) [2] is an automated, unsupervised, and competitive data analysis technique widely applied to clustering problems and data exploration. It is a tool for representing an input data distribution using a finite set of prototypes or models. The input data-vectors are associated with the nodes of a regular grid in an ordered way, following the idea that the most similar data is associated with adjacent nodes in the grid. Since SOM compresses information while preserving the main topological and metric relationships, it is a powerful visualization, projection, and abstraction tool [9, 10].

2˙3. *Multiscale entropy.* – Each one-minute seismic trace is employed as an input vector for SOM training. However, utilizing the entire seismic record as input is computationally inefficient. To address this issue, it is useful to reduce the input dimensionality by extracting specific features. In this study, we chose Multiscale Entropy [3], a metric that quantifies the entropy, or complexity, of a waveform across multiple time scales. Specifically, Multiscale Entropy consists of two steps: first, different time scales are obtained by coarse-graining the original time series into non-overlapping time windows and averaging the data points within each window. Second, Sample Entropy (SE) is calculated iteratively for each coarse-grained sequence, generating an SE vector across various time scales. SE is defined as the negative natural logarithm of the conditional probability that two sequences of m consecutive samples will remain similar within a certain tolerance when an additional sample is added to the sequence. We computed 20 scale-factors, generating 20 SE values for each one-minute trace, which form the input vector of the SOM training.

3. – Results

For the SOM training, we selected a 6×6 hexagonal map. The analysis of the SOM trained on seismic traces encoded with MSE reveals notable findings. Specifically, the seismic waveform compression through MSE values shows the capability to differentiate between noise-dominated seismic traces and those containing seismic events, highlighting its utility in preprocessing seismic data for neural network training. Moreover, there is evidence of a partial separation of the noise signals divided into two large time bands, one corresponding to daytime and the other to nighttime.

3˙1. *Seismic events clustering.* – The SOM displayed a group of nodes, located primarily on the right side of the network, characterized by significant distances from the other nodes (fig. 1). This cluster contains mainly seismic waveforms corresponding to events that match the ones listed in the referenced event catalog for the analysis period (fig. 2(a)). The clustering of cataloged seismic events illustrates the effectiveness of the combined MSE and SOM approach in isolating meaningful seismic signals from background noise. Interestingly, some traces within this event cluster do not match with any event in the catalog. These include potential uncataloged local events or signals that may represent false positives. This indicates that an MSE approach may also enhance the detection of previously unidentified events, although further validation is needed.

3˙2. *Daytime vs. Nighttime signals.* – In addition to the event cluster, the SOM revealed two distinct large groups of noise-dominated traces, one primarily containing signals recorded during daytime hours and the other composed of signals from nighttime recordings (fig. 2(b)). However, the separation between daytime and nighttime noise clusters is not well defined. A transition zone exists, with nodes containing traces that do not clearly belong to either cluster, indicating a gradual separation between the two noise types. This overlap suggests that while the SOM trained with MSE values captures broad temporal patterns, finer distinctions may require additional feature refinement or preprocessing.

4. – Conclusion

This study underlines the potential of the MSE-SOM approach in seismic data analysis. By using MSE as a feature extraction method, we have effectively reduced the dimensionality of seismic traces, while retaining essential information about waveform complexity. The SOM clustering proved to be meaningful, showing a distinct separation between seismic events and noise-dominated traces, leading to the identification of potential uncataloged events. The approach also highlighted distinctions in noise characteristics, with daytime and nighttime signals grouping in separate clusters, albeit with some overlap in transitional zones. These results emphasize the ability of the MSE-SOM

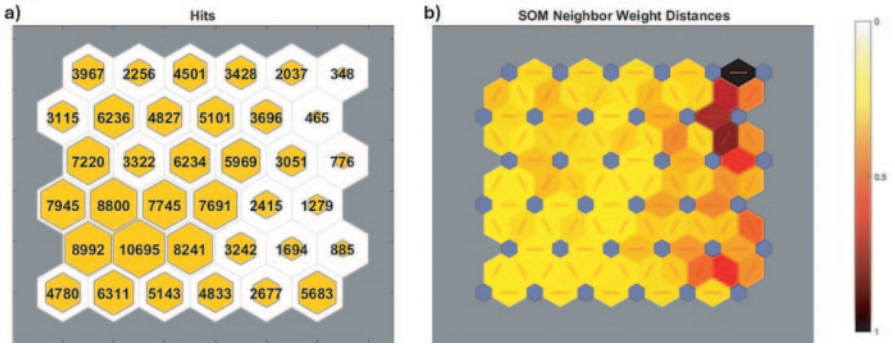


Fig. 1. – SOM trained on the one-minute seismic waveforms dataset recorded from July 20th, 2023 to November 12th, 2023, and encoded using the Multiscale Entropy values. (a) Number of seismic waveforms belonging to each of the 6×6 nodes in the SOM (hits on each node); (b) Euclidean distances between the SOM node prototypes. The chromatic scale indicates that darker colors are associated with greater distances among the nodes.

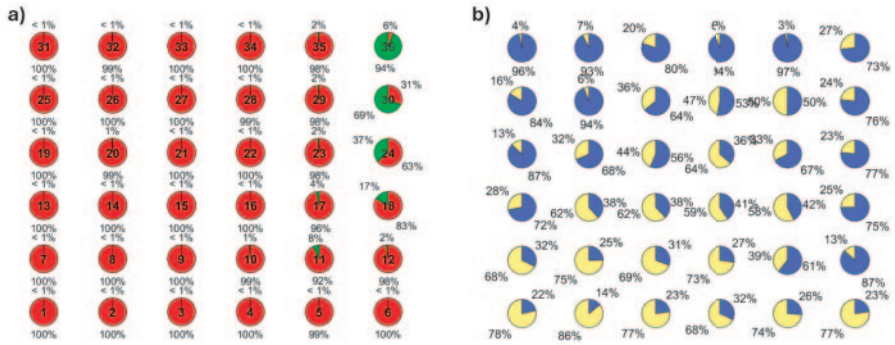


Fig. 2. – (a) Percentage of cataloged seismic events (green section) and of the remaining waveforms (red section) collected in each node of the SOM. (b) Percentage of daytime recordings (6:00 AM–6:00 PM, yellow section) *vs.* nighttime recordings (6:00 PM–6:00 AM, blue section) collected in each node of the SOM.

methodology to capture fine temporal and spatial variations in seismic data, offering a valuable tool for improving the resolution and interpretability of large seismic datasets. Future work will focus on validating uncataloged traces and refining feature selection to improve clustering fidelity and minimize noise separation ambiguity.

REFERENCES

[1] Mostafa Mousavi S. and Beroza Gregory C., *Science*, **377** (2022) eabm4470.
[2] Kohonen T., *Biol. Cybern.*, **43** (1982) 59.
[3] Costa M., Goldberger A. and Peng C.-K., *Phys. Rev. Lett.*, **89** (2002) 068102.
[4] Wolfson E. J., Fekete T., Loewenstein Y. and Shriki O., *Sci Rep.*, **14** (2024) 14680.
[5] Romano P., Di Lieto B., Scarpetta S., Apicella I., Linde A. T. and Scarpa R., *Front. Earth Sci.*, **10** (2022) 862086.
[6] Meyer S., Reading A. and Bassom A., *Geophys. J. Int.*, **231** (2022) 2156.
[7] Messuti G., Scarpetta S., Amoroso O., Napolitano F., Falanga M. and Capuano P., *Front. Earth Sci.*, **11** (2023) 1223686.
[8] Messuti G., Napolitano F., Falanga M., Capuano P. and Scarpetta S., *Uncertainty estimation via ensembles of deep learning models and dropout layers for seismic traces*, arXiv:2410.06120 (2024).
[9] Liu Z., Cao J., Chen S., Lu Y. and Tan F., *IEEE Geosci. Remote Sens. Lett.*, **18** (2021) 1491.
[10] Esposito A. M., Giudicepietro F., D'Auria L., Scarpetta S., Martini M. G., Coltelli M. and Marinaro M., *Bull. Seismol. Soc. Am.*, **98** (2008) 2449.