SYSTEMATIC REVIEW OF SIMILARITY CRITERIA FOR IDENTIFYING SIMILAR SEISMIC EVENTS

REVISÃO SISTEMÁTICA DOS CRITÉRIOS DE SIMILARIDADE NA IDENTIFICAÇÃO DE EVENTOS SÍSMICOS SEMELHANTES

REVISIÓN SISTEMÁTICA DE LOS CRITERIOS DE SIMILITUD EN LA IDENTIFICACIÓN DE EVENTOS SÍSMICOS SIMILARES

https://doi.org/10.56238/sevened2026.008-019

Jaqueline Alexsandra Azevedo Ferreira[1]

ABSTRACT

The identification of similar seismic events constitutes an important aspect for understanding seismic dynamics, identifying the recurrence of aftershocks and seismic swarms, and assessing geological hazards. This work presents a systematic literature review on the computational methods employed for identifying similarity between seismic waveforms, with an emphasis on decision criteria and signal processing techniques. The review was conducted based on searches in scientific databases, utilizing specific strings related to seismic event similarity measures. The analysis of the selected studies evidences that, despite the recurrent limitations associated with the dependence on empirical thresholds, sensitivity to noise, and the difficulty of ensuring consistency and reproducibility in event classification within large seismic catalogs, cross-correlation-based methods are widely adopted. It is concluded that the literature points to conceptual gaps in the modeling of similarity between seismic events, indicating the need for future mathematical and computational investigations aimed at developing more consistent and well-founded criteria for the automated curation of seismic data.

Keywords: Similar Seismic Events. Waveform Similarity. Systematic Review.

RESUMO

A identificação de eventos sísmicos semelhantes constitui um aspecto importante entender a dinâmica sísmica, e identificar recorrência de réplicas, enxames sísmicos e avaliar riscos geológicos. Este trabalho apresenta uma revisão sistemática da literatura sobre os métodos computacionais empregados na identificação de similaridade entre formas de onda sísmicas, com ênfase nos critérios de decisão e nas técnicas de processamento de sinais. A revisão foi conduzida a partir de buscas em bases de dados científicas, utilizando strings específicas relacionadas à medidas de similaridade de eventos sísmicos. A análise dos estudos selecionados evidencia que, apesar das limitações recorrentes associadas à dependência de limiares empíricos, da sensibilidade ao ruído e da dificuldade de garantir consistência e reprodutibilidade na classificação de eventos em grandes catálogos sísmicos, os métodos

[1] Master's degree in Mathematics. Universidade Federal do Recôncavo da Bahia (UFRB).
E-mail: Jaqueline_azevedo@ufrb.edu.br Orcid: https://orcid.org/0000-0002-3184-6135
Lattes: https://lattes.cnpq.br/0964632956964410

baseados em correlação cruzada são amplamente adotados. Conclui-se que a literatura aponta lacunas conceituais na modelagem da similaridade entre eventos sísmicos, indicando a necessidade de investigações matemáticas e computacionais futuras voltadas ao desenvolvimento de critérios mais consistentes e fundamentados para a curadoria automatizada de dados sísmicos.

Palavras-chave: Eventos Sísmicos Semelhantes. Similaridade de Formas de Onda. Revisão Sistemática.

RESUMEN

La identificación de eventos sísmicos similares constituye un aspecto importante para comprender la dinámica sísmica, e identificar la recurrencia de réplicas, enjambres sísmicos y evaluar riesgos geológicos. Este trabajo presenta una revisión sistemática de la literatura sobre los métodos computacionales empleados en la identificación de similitud entre formas de onda sísmicas, con énfasis en los criterios de decisión y las técnicas de procesamiento de señales. La revisión se llevó a cabo a partir de búsquedas en bases de datos científicas, utilizando cadenas de búsqueda específicas relacionadas con medidas de similitud de eventos sísmicos. El análisis de los estudios seleccionados evidencia que, a pesar de las limitaciones recurrentes asociadas a la dependencia de umbrales empíricos, la sensibilidad al ruido y la dificultad de garantizar consistencia y reproducibilidad en la clasificación de eventos en grandes catálogos sísmicos, los métodos basados en correlación cruzada son ampliamente adoptados. Se concluye que la literatura señala lagunas conceptuales en el modelado de la similitud entre eventos sísmicos, indicando la necesidad de futuras investigaciones matemáticas y computacionales orientadas al desarrollo de criterios más consistentes y fundamentados para la curación automatizada de datos sísmicos.

Palabras clave: Eventos Sísmicos Similares. Similitud de Formas de Onda. Revisión Sistemática.

1 INTRODUCTION

The identification of similar seismic events (such as aftershocks and earthquake swarms) plays an important role in the assessment of geological risks, with regard to the evolution of faults, and in the understanding of seismological dynamics, with respect to the failure mechanism. In this sense, the comparative analysis of seismic waveforms is a fundamental tool for the investigation of the similarity between events and for the construction of more consistent seismic catalogs.

Culturally, Geoscience specialists have been identifying similar seismic events through visual inspections or methods based on cross-correlation between seismic records. Although these approaches have proven effective in different applications, they rely on empirical criteria, variations in wave propagation, macroseismic surveying, and exhaustive analysis of seismic event pairs. This imposes limitations when applied to environments with high noise levels or large volumes of data.

In recent decades, the increased availability of seismic data has driven the development of automated methods for detecting similar events. Despite the progress observed, the literature still presents a diversity of criteria and decision strategies, often not very comparable to each other, which evidences the absence of a conceptual consensus on the definition and modeling of the similarity between seismic events.

In view of this scenario, it is relevant to conduct a systematic review of the literature that allows a synthesis of existing methods, identifies their main advantages and limitations, and evaluates the need for new approaches. Thus, the objective of this work is to present a systematic review of the approaches used in the identification of similar seismic events, highlighting the methods used and the recurrent challenges pointed out in the literature. By gathering and organizing this knowledge, it seeks to provide subsidies for the evaluation of the need (or not) for the development of new criteria or conceptual structures for the analysis of similarity between seismic waveforms.

2 METHODOLOGY

This study was conducted as a systematic review of the literature (SAMPAIO; MANCINI, 2007), with the objective of identifying, analyzing and comparing the main methods used in the literature for the identification of similar seismic events from the comparison of waveforms.

2.1 TYPE OF REVIEW

This research is structured as a systematic review of the literature (SAMPAIO;

MANCINI, 2007) which evaluates the state of the art in a qualitative way and discusses the sufficiency of the approaches currently employed, with emphasis on the similarity criteria used, their advantages and limitations.

2.2 DATABASES CONSULTED

The bibliographic search was carried out in the following scientific databases, selected for their relevance and comprehensiveness in the area of seismology and signal processing:

- Web of Science
- Scopus
- Google Scholar

In addition to these sources, technical documents and institutional materials from official bodies, such as the United States Geological Survey (USGS), were considered when relevant for conceptual contextualization.

2.3 SEARCH STRATEGY AND STRINGS

The following search strings have been applied, with adaptations according to the syntax of each database:

("seismic event similarity" OR "waveform similarity" OR "repeating earthquakes")
AND ("cross-correlation" OR "template matching" OR "matched filter")

The searches included publications between **2000 and 2025**, a period that covers the development and consolidation of the main approaches used in the area.

2.4 INCLUSION CRITERIA

Studies that met the following criteria were included in the review:

- Peer-reviewed articles;
- Works that use real seismic data;
- Studies that employ waveform comparison to identify similar events;
- Publications that explicitly describe the similarity criterion or metric adopted;
- Articles published in English.

2.5 EXCLUSION CRITERIA

The following were excluded from the analysis:

- Purely theoretical work without application to seismic data;
- Studies that do not present a clear description of the decision criterion or similarity;

- Duplicate articles among the databases consulted;

2.6 STUDY SELECTION PROCESS

The study selection process was conducted in three successive stages:

(i) reading of the titles;

(ii) reading of the abstracts;

In the final stage, only studies that fully met the inclusion criteria were considered for qualitative analysis. The information extracted from each article included the type of data analyzed, the method used, the similarity metric, the decision criterion, and the main advantages and limitations identified.

3 LITERATURE REVIEW

3.1 CROSS-CORRELATION

Cross-correlation is the most traditional and widely used method in the identification of similar seismic events, being used to quantify the degree of similarity between waveforms recorded by one or more seismographic stations. In this context, events are considered similar when the maximum coefficient of cross-correlation between their signs exceeds a certain previously established threshold. Classical studies have shown that this approach is effective for the identification of repeated microearthquakes and for the analysis of seismic clusters associated with the same source or geological fault (CHENG et al., 2007; LI et al., 2007).

Some studies have successfully applied this technique in different geological contexts, evidencing its ability to reveal seismic repetition patterns and estimate relevant physical parameters, such as slip rates on active faults (LI et al., 2011). However, despite its wide adoption, classical cross-correlation has important limitations, especially with regard to noise sensitivity, dependence on the temporal alignment between signals, and the empirical choice of decision threshold, which can influence the number of events classified as similar (BACHMANN et al., 2007; ZHU et al., 2021).

3.2 REPEATING SIGNAL DETECTORS

In addition to correlation, other studies have employed template-based approaches for the identification of similar seismic events. Seismic waveforms previously detected by seismographs are used as templates, being systematically compared with continuous online seismic records or with sets of events, by means of cross-correlation measurements or

Euclidean distances. This strategy has proven to be particularly efficient for the detection of seismic swarms, allowing the automatic identification of occurrences with high temporal and spectral similarity (SKOUMAL; BRUDZINSKI; CURRIE, 2016).

The use of specific detectors, such as the *Repeating Signal Detector*, expands the applicability of the *template* by enabling the analysis of large volumes of data with greater computational efficiency. Recent work demonstrates that this approach is capable of revealing sequences of repeated events that would not be easily identified by manual inspection (TEPP, 2018; CHAMBERLAIN et al., 2020). However, template-based methods and tools also have limitations related to the choice of *templates*, to the definition of correlation thresholds, since events with characteristics different from the previously selected waveforms may not be identified, due to environmental issues external to the event at that moment, restricting the scope of the analysis.

In addition to the methods described, specialists use *templates* in computer tools for the automated identification of similar seismic events. PyMPA (Python Matching Phase Algorithm) is a package aimed at automatic seismic phase matching (PYMPA DEVELOPERS, 2025), while SCDetect consists of a module of the SeisComP environment based on cross-correlation for the detection of earthquakes from waveforms (SCDETECT DEVELOPERS, 2025). These tools show the predominance of approaches based on cross-correlation and empirical thresholds.

3.3 CLUSTERING METHODS APPLIED TO SEISMIC SIMILARITY

In addition to approaches based on direct comparisons between pairs of events, some studies have employed clustering methods to organize sets of similar seismic events based on measures of similarity between waveforms. From this perspective, the cross-correlation metric is applied in clustering algorithms with the objective of identifying groups of events associated with the same seismic source or similar rupture mechanisms (WALDHAUSER; SCHAFF, 2007; BACHMANN et al., 2007).

These approaches allow for a more global analysis of the relationships between events, reducing the reliance on one-off comparisons. Recent studies have shown that clustering techniques can reveal internal structures in seismic data that are not evident through methods based exclusively on correlation thresholds, contributing to the understanding of the spatial and temporal organization of seismicity (WALDHAUSER; SCHAFF, 2021; GAO; KAO; LIU, 2023).

3.4 MACHINE LEARNING-BASED APPROACHES IN IDENTIFYING SIMILAR SEISMIC EVENTS

Machine learning-based approaches have been explored for the identification and classification of similar seismic events, especially in contexts characterized by complex signals and with large volumes of seismic data. These methods use supervised learning techniques to extract relevant characteristics from seismic waveforms, allowing automatic comparison between events from representations that are learned from the data, and in this case, are not restricted to explicitly defined similarity metrics (LI et al., 2018).

Machine learning models can be more robust in the face of noise and variations in seismic wave propagation conditions, when compared to traditional methods based exclusively on cross-correlation. In addition, these approaches have been successfully applied in the classification and grouping of seismic events, expanding the ability to detect complex and nonlinear patterns in the data (SI et al., 2023). However, despite their potential, machine learning-based methods have relevant limitations, such as the dependence on large labeled datasets, the lower physical interpretability of the results, and the difficulty of generalizing to regions or geological contexts other than those used in the training, which imposes challenges to their systematic application in seismic similarity studies.

4 RESULTS

The application of the search strategy defined in this systematic review initially resulted in a broad set of studies related to the identification of similar seismic events. After applying the inclusion and exclusion criteria and removing duplicates, a final set of materials considered relevant for the analysis was obtained. These studies cover classical methods based on cross-correlation, Euclidean distance, repeating signal detection techniques, and newer machine learning-based approaches.

The selected studies predominantly use methods based on the direct comparison of seismic waveforms, with emphasis on cross-correlation as a measure of similarity. Classical and widely cited works employ this coefficient to identify repeated or similar events, exploiting the temporal alignment of waves and maximizing the correlation value. In addition, approaches were identified that combine cross-correlation with pre-processing methods, such as spectral filtering, in order to reduce the influence of noise and external variations.

Some studies have used sliding windows to identify possible events, while others propose specific detectors for recurrent signals, capable of identifying multiple occurrences of similar events over large volumes of seismic data.

In addition to the classical methods, part of the analyzed works explores clustering techniques and machine learning algorithms for the identification of similar seismic events. Approaches include unsupervised methods, such as distance-based clustering, as well as supervised models aimed at classifying previously observed seismic events. In general, these methods use characteristics extracted from waveforms or representations in the time and frequency domains, seeking to automate the process of identifying recurring patterns in seismic catalog logos extracted from different seismic stations.

The analysis of the studies shows that the similarity criteria adopted in the literature vary considerably, being predominantly based on quantitative measures between pairs of events, such as correlation coefficients, distances in space of features or metrics derived from spectral transforms. Although these criteria allow the identification of similar events in different contexts, it is observed that there is no consensus on a unified model for representing the similarity between multiple events, which results in the coexistence of diverse approaches, adapted to specific objectives and data sets.

Table 1

Comparative synthesis of the main methods used in the identification of similar seismic events in the literature

Author(s)	**Year**	**Main Method**	**Similarity criterion**	**Type of data analyzed**
Bachmann et al.	2007	Cross-correlation	Maximum correlation coefficient between time-aligned waveforms	Seismic records of natural events
Li et al.	2007	Cross-correlation and spatial clustering	Waveform similarity associated with spatial proximity	Microearthquakes along active faults
Cheng et al.	2007	Cross-correlation	High correlation between repeated events	Microearthquakes recorded in Japan
Zhu et al.	2021	Re-evaluation of cross-correlation	Cross-correlation with pre-processing adjustments	Regional seismic catalogs

Skoumal et al.	2016	Repeating Signal Detector (RSD)	Correlation-based similarity and temporal recurrence	Seismic swarms
Tepp	2018	Repeating Signal Detector	Identifying recurring patterns in the signal	Continuous seismic data
Li et al.	2011	Cross-correlation	Repeated Event Similarity for Estimation of Slip Rates	Deep microearthquakes
Gonzales & Wintz	1987	Digital signal processing	Euclidean Distance	Discreet digital images
Li et al.	2018	Machine learning	Similarity in the Feature Space	Automatically classified seismic data
USGS	2024	Observational analysis	Physical and temporal criteria	Seismic events and aftershocks

Source: Authors.

5 DISCUSSIONS

The analysis of the studies selected in this systematic review shows the predominance of techniques based on the cross-correlation metric. This confirms the relevance of this metric as a fundamental tool for the comparison of waveforms, especially in contexts involving repeated events or seismic aftershocks. However, the way this metric is employed varies significantly between studies, both in terms of pre-processing and the definition of decision thresholds.

The use of empirical parameters is a recurrent aspect identified in the literature. They are often adjusted according to the local geology. Although this flexibility allows the methods to be adapted to different geological contexts, it introduces subjectivity into the process of identifying similar events, making it difficult to directly compare studies and reproduce results in different scenarios. This limitation is very relevant in analyses involving large seismic catalogs, where small variations in the parameters can result in very different classifications.

It was also observed that most approaches deal with the similarity between events by pairs, without considering the global structure of the relationships established between multiple events in the same database. This characteristic can lead to details that may not be well observed regarding comparisons between seismic waveforms, such as the absence of the transitivity relationship of the similarity between seismic events, which makes it difficult to

interpret the groupings. Although clustering techniques and machine learning-based methods have been proposed as alternatives to deal with this issue, such approaches often present additional challenges related to the choice of appropriate parameters.

Another relevant point refers to the sensitivity of methods for identifying seismic events similar to the conditions of data acquisition and the presence of external noise. Differences in the location of the installation of the stations and in the propagation characteristics of the waves can significantly affect the similarity measurements, even between events originating from the same seismic source. In order to mitigate these effects, several studies have adopted pre-processing strategies, including spectral filtering for noise reduction and temporal alignment of waveforms before calculating similarity metrics (BACHMANN et al., 2007; BARANI; FERRETTI, 2007). More recent approaches, such as multichannel waveform clustering methods, also incorporate these filtering, normalization, and alignment steps as a way to improve the robustness of cross-event comparisons (CASTELLANOS; VAN DER BAAN, 2015). Despite these advances, it was identified that there is still no consensus in the literature regarding a standard procedure capable of ensuring consistent performance in different seismic scenarios, especially when considering large catalogs and variable acquisition and noise conditions (ZHU et al., 2021). In general, the results of this review indicate that, despite the advances achieved in recent decades, methodological gaps still persist in modeling the similarity between seismic events. The absence of unified criteria and formal structures that adequately capture the relationships between multiple events suggests the need for further investigations aimed at developing more consistent approaches. In this context, this critical analysis contributes to the understanding of the state of the art and establishes a solid basis for future research that seeks to improve the criteria for identifying similar seismic waveforms.

6 CONCLUSION

This work presents a systematic review of the literature focused on techniques for identifying similar seismic events. The synthesis of the studies showed that, although there are a variety of approaches available, most methods are based on direct comparisons between pairs of events, often based on cross-correlation measures and empirical decision criteria that depend on the local geology.

The results seen in this systematic review indicated that the approaches have recurrent limitations related to noise sensitivity, dependence on data acquisition conditions, and inconsistency in the classification of similar events regarding the form of seismic waves throughout large seismic catalogs. In addition, the absence of unified conceptual frameworks

for modeling the similarity between multiple events contributes to inconsistencies in the interpretation of clusters and in the comparison between different studies. It is observed that the identification of similar seismic events remains dependent on the complexity of the data analyzed. Thus, future investigations may benefit from the incorporation of interdisciplinary approaches that explore mathematical concepts applied to the normalization, interpolation and noise filtering of seismic waveforms, in order to promote a more consistent standardization of data and enable fairer comparisons between events. This integration between Geosciences, Mathematics and Computing represents a promising path for the improvement of similarity criteria and for the development of more robust computational tools in seismic analysis.

Thus, the critical synthesis conducted in this systematic review reinforces the importance of future investigations aimed at the development and evaluation of more consistent, interpretable and theoretically grounded similarity criteria, capable of dealing with the heterogeneity of seismic waveforms collected by seismological stations. In this sense, this work provides an overview of the state of the art and establishes a conceptual basis for subsequent research that seeks to advance the curation and automated analysis of seismic data.

REFERENCES

Bachmann, C., Wiemer, S., Woessner, J., & Hainzl, S. (2007). Analysis of seismic event similarity using cross-correlation. Geophysical Journal International, 168(2), 459–473. https://doi.org/10.1111/j.1365-246X.2006.03199.x

Barani, S., & Ferretti, G. (2007). The waveform similarity approach to identify dependent events in instrumental seismic catalogues. Geophysical Journal International, 168(1), 100–108. https://doi.org/10.1111/j.1365-246X.2006.03193.x

Castellanos, J., & van der Baan, M. (2015). Multi-channel waveform clustering for seismic event analysis. Journal of Seismology, 19(2), 273–289. https://doi.org/10.1007/s10950-014-9465-7

Cheng, X., Niu, F., Silver, P. G., Horiuchi, S., Takai, K., Iio, Y., & Ito, H. (2007). Similar microearthquakes observed in western Nagano, Japan, and implications for rupture mechanics. Journal of Geophysical Research: Solid Earth, 112(B4), Article B04306. https://doi.org/10.1029/2006JB004416

Gao, D., Kao, H., & Liu, J. (2023). Identification of repeating earthquakes: Controversy and rectification. Seismological Research Letters, 94(6), 2655–2665. https://doi.org/10.1785/0220230124

Kanasewich, E. R. (1981). Time sequence analysis in geophysics. University of Alberta Press.

Li, L., Chen, Q. F., Cheng, X., & Niu, F. (2007). Spatial clustering and repeating of seismic events observed along the 1976 Tangshan fault, North China. Geophysical Research Letters, 34(23), Article L23309. https://doi.org/10.1029/2007GL031594

Li, L., Chen, Q. F., Niu, F., & Su, J. (2011). Deep slip rates along the Longmen Shan fault zone estimated from repeating microearthquakes. Journal of Geophysical Research: Solid Earth, 116(B9), Article B09310. https://doi.org/10.1029/2011JB008406

Li, W., Nakshatra, Narvekar, N., Raut, N., Sirkeci, B., & Gao, J. (2018). Seismic data classification using machine learning. In 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (pp. 1–8). IEEE. https://doi.org/10.1109/BigDataService.2018.00009

PyMPA Developers. (n.d.). PyMPA — Python Matching Phase Algorithm. https://pympa37.readthedocs.io/en/latest/index.html

Sampaio, R. F., & Mancini, M. C. (2007). Estudos de revisão sistemática: Um guia para síntese criteriosa da evidência científica. Revista Brasileira de Fisioterapia, 11(1), 83–89.

SCDetect Developers. (n.d.). SCDetect: SeisComP package for waveform cross-correlation based earthquake detection. https://scdetect.readthedocs.io/en/stable/

Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2016). An efficient repeating signal detector to investigate earthquake swarms. Journal of Geophysical Research: Solid Earth, 121(8), 5880–5897. https://doi.org/10.1002/2016JB012981

Smith, S. W. (1997). The scientist and engineer's guide to digital signal processing. California Technical Publishing.

Tepp, B. (2018). Repeating signal detection using the repeating signal detector algorithm. Seismic Detection Methods, 15(1), 15–29.

U.S. Geological Survey. (n.d.). What is an earthquake and what causes them to happen? https://www.usgs.gov/faqs/what-earthquake-and-what-causes-them-happen

van Etten, W. C. (2006). Introduction to random signals and noise. John Wiley & Sons.

Wikipedia Contributors. (2025). Cross-correlation. In Wikipedia. https://en.wikipedia.org/wiki/Cross-correlation

Zhu, T., et al. (2021). Reassessing cross-correlation for seismic event identification. Journal of Seismology, 25(3), 789–802.