Bulletin of Geophysics and Oceanography Vol. XX, n. X, pp. XXX-XXX; Xxxxx 20xx
DOI 10.4430/bgo00518

NESTORE algorithm: a machine learning approach for strong aftershock forecasting. A comparison of California, Italy, western Slovenia, Greece, and Japan results

S. Gentili[1], P. Brondi[1], G.D. Chiappetta[1], G. Petrillo[2], J. Zhuang[3], E.-A. Anyfadi[4], F. Vallianatos[4], L. Caravella[1], E. Magrin[1], P. Comelli[1] and R. Di Giovambattista[5]

[1] *National Institute of Oceanography and Applied Geophysics - OGS, Udine, Italy*
[2] *Earth Observatory of Singapore (EOS), Nanyang Technological University (NTU), Singapore*
[3] *The Institute of Statistical Mathematics – ISM, Tokyo, Japan*
[4] *Section of Geophysics-Geothermics, Department of Geology and Geoenvironment, National and Kapodistrian University, Athens, Greece*
[5] *Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy*

(Received: 31 July 2025; accepted: 25 November 2025; published online: 13 February 2026)

ABSTRACT Large earthquakes are often followed by aftershocks, which can cause further damage and cost lives. A pattern recognition approach called Next Strong Related Earthquake (NESTORE) has been developed to forecast whether one of these subsequent large events is to be expected in an occurring seismicity cluster. This method, already successfully applied in Italy, Slovenia, California, Greece, and Japan, has been optimised in the NESTOREv1.0 software written in MATLAB. Using machine learning, NESTOREv1.0 provides a probabilistic forecast of earthquake clusters where a mainshock is followed by a significant aftershock. It classifies clusters as type A (mainshock and strongest aftershock differ by ≤ 1 magnitude unit) or type B (larger difference). NESTOREv1.0 adapts to specific regions through supervised training. It trains one-node decision trees on individual features at increasing time intervals, selects the best classifiers, and combines them using a Bayesian method to forecast type A clusters. Recent improvements to the algorithm added a new approach for identifying clusters based on Epidemic-Type Aftershock Sequence (ETAS) and an innovative method for detecting outliers before training. This study compares results from Greece, Italy, western Slovenia, California, and Japan, highlighting the performance on independent test sets and seismicity features in different regions and interpreting the differences between the regions.

Key words: machine learning, cluster identification, strong aftershock, earthquake forecasting, outlier detection.

1. Introduction

The increasing availability of high-quality seismic data and the advancement of machine learning techniques applied to seismology (DeVries *et al.*, 2018; Karimzadeh *et al.*, 2019; Wang *et al.*, 2019; Mignan and Broccardo, 2020; Zhao *et al.*, 2022; Stockman *et al.*, 2023; Liu *et al.*, 2024; Schimmenti *et al.*, 2024) have opened new roads for improving aftershock forecasting, a crucial component of seismic hazard assessment. In this context, we developed the Next Strong

Related Earthquake (NESTORE) algorithm, a data-driven approach designed to identify and classify clusters with strong aftershocks using supervised learning methods. This study presents the outcomes resulting from the application of NESTORE in five tectonically diverse regions, California, Italy, western Slovenia, Greece, and Japan, enabling a comparative analysis of its performance and generalisation capability.

In this paper, we summarise, integrate, and coherently compare the results of previous NESTORE applications, which have been presented individually in papers, conferences, workshops, and theses (Gentili and Di Giovambattista, 2017, 2020, 2022; Anyfadi, 2023; Anyfadi *et al*., 2023; Gentili *et al*., 2023a, 2023b, 2024, 2025; Brondi *et al*., 2025a).

Here, we aim to jointly exploit these results to improve the understanding of feature performance. By comparing the behaviour of features and threshold values in different seismotectonic domains, we seek to identify patterns and evaluate how different tectonic settings may influence predictive capabilities, providing clues for future methodological and operational developments.

As NESTORE is also applied to new clusters, some results are shown here for the first time.

Over the years, the method has been modified and improved to make it more robust and to meet the challenges posed by the analyses of new regions. Section 2 describes the regions where NESTORE has been applied from a seismotectonic perspective, highlighting the differences within and between regions whereas Section 3 provides a detailed description of the newer version of the method. Section 4 illustrates how the results differ between regions both in terms of the performance of the algorithm and the parameters obtained through training. Finally, in Section 5, the differences observed in the different studied regions are discussed and interpreted, and the limitations and advantages of the method are explored together with the research approaches pursued to address these limitations.

2. Analysed regions: seismotectonic and available data

The NESTORE algorithm has been successfully applied in several regions (see Fig. 1) characterised by different productivity and characteristics of seismic clusters. In the following sections we summarise the seismotectonic regimes of the analysed regions.

It is important to point out that the quality of NESTORE's performance is highly dependent on the selected area and the quality and consistency of the catalogue. For this reason, preliminary analyses of the available catalogues were conducted in each territory, taking into account location accuracy and magnitude homogeneity. We excluded some offshore regions because of their poorer location and magnitude accuracy, as well as their higher completeness magnitude. In addition, we excluded volcanic regions and regions with deep seismicity, since they behave differently from shallow tectonic seismicity regions.

In one case, after a preliminary analysis, we also removed a sub-region (the small orange area in Fig. 1, located in northern Italy), where the seismicity characteristics, in terms of event frequency, differed from those of the rest of the analysed territory (Brondi *et al*., 2025a). In the corresponding paper, we proposed some hypotheses regarding the physical meaning of this behaviour. From a statistical perspective, the clusters in this area belong to a region-specific population distinct from the other clusters, so including them in a joint analysis would introduce bias. Particular care was taken in defining magnitude homogeneity. For example, we used the Southern California Earthquake Catalogue (Hutton *et al*., 2010; SCEDC, 2013), which lists local magnitudes for all earthquakes. However, we did not use the Northern California Earthquake Catalogue (NCEDC, 2014), despite its similar quality, because it lists different types

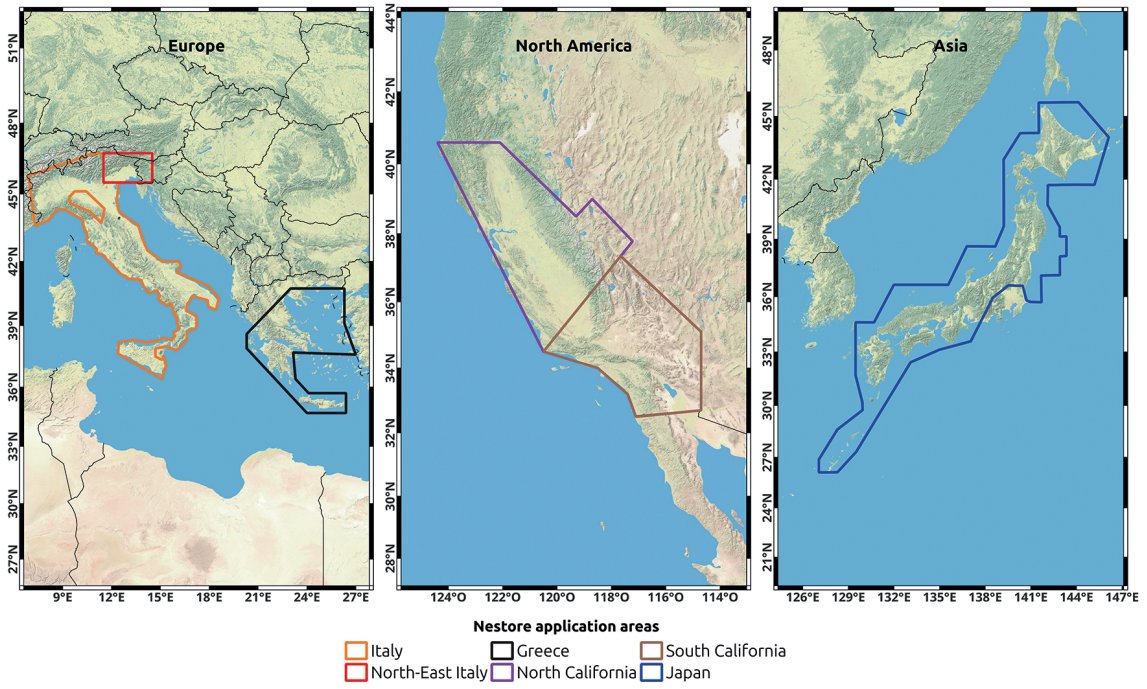


Fig. 1 - Areas analysed in past NESTORE applications.

of magnitude for different earthquakes. For northern California and the limited data available in western Nevada along the California border, we used the Advanced National Seismic System (ANSS) California Comprehensive Earthquake Catalogue (U.S. Geological Survey, n.d.) when local magnitude data were available as its magnitude assessment aligns well with the Southern California Earthquake Catalogue. For the same reason, we used the revised Lolli and Gasperini (2006) catalogue prior to the start of the ISIDE (Italian Seismological Instrumental and Parametric Database) catalogue in 2005, as it was explicitly proposed to provide compatible magnitude assessment. Catalogues adopted in Italy and NE Italy and western Slovenia after 2020 have been used for applying our methodologies to clusters different from those used for the calibration of the method; these are the real-time versions of the catalogues used for the previous years. They have been used for the near-real-time (NRT) application of the algorithm.

Table 1 lists the catalogues used for each region.

The following sections summarise the tectonic framework across analysed regions, providing context for interpreting how local structures may affect the occurrence of the largest aftershocks within seismic sequences and how these regional tectonic differences may influence the performance of features used for forecasting strong subsequent earthquakes.

2.1. Italy

One of the regions where NESTORE has been more widely used is Italy (Gentili and Di Giovambattista, 2017; Brondi *et al*., 2025a) (indicated in orange in Fig. 1).

Italian seismicity results from the interaction between the African and Eurasian plates and can be divided into five main zones (Gentili and Di Giovambattista, 2017; Brondi *et al*., 2025a):

Table 1 - Catalogues adopted in different regions and time periods for which they are used.

Region	Time interval	Reference
Italy	1980–2004 2005–2020 2021–2025	Lolli and Gasperini (2006) Italian Seismological Instrumental and Parametric Database (ISIDe Working Group, 2007) https://terremoti.ingv.it/
NE Italy – western Slovenia	1977–2020 2021–2025	Snidarcig *et al.* (2020) https://terremoti.ogs.it/
California	1980–2023 1980–2020	Southern California Seismic Network (SCSN) earthquake catalogue (Hutton *et al.*, 2010) ANSS ComCat
Greece	1995–2023	Aristotle University of Thessaloniki (AUTH) earthquake catalogue
Japan	1973–2024	Japan Meteorological Agency (2024)

- Alps and Dinarides. This sector was formed by the collision between the Eurasian plate and the Adriatic microplate, which led to crustal compression and uplift. Seismicity is shallow and clustered in the eastern sectors (Friuli, eastern Alps) and the western regions (Piedmont, Ligurian Alps), while the central Alps are quieter. Historical events include the 1511 Idrija (M_w ~6.8) and 1976 Friuli (M_w 6.4) earthquakes, as well as modern M_w 4–5 earthquakes along thrusts and strike-slip faults. Global positioning system data indicate ongoing deformation consistent with these seismogenic structures (Devoti *et al*., 2008);
- Apennine Range. This range is characterised by ongoing extension due to slab retreat, with frequent earthquakes occurring along NW–SE normal faults (http://emidius.mi.ingv.it/CPTI/; Malinverno and Ryan, 1986; Doglioni, 1991, 1995). Major historical and recent events include earthquakes of 1688, 1805, 1962, and 1980 Irpinia, 1984 Abruzzo-Lazio (Milano and Di Giovambattista, 2011), 1990 Potenza (Azzara *et al*., 1993), 1997 Umbria-Marche (Miller *et al*., 2004), and 2016 Amatrice (Calderoni *et al*., 2017). This range presents seismicity clusters along the active fault network with increasing deformation rates southwards;
- Calabrian Arc and southern Tyrrhenian Sea. This is an active subduction margin where the Ionian plate subducts beneath the Calabrian Arc, generating shallow and deep earthquakes (Doglioni *et al*., 2007; Chiarabba *et al*., 2008). Notable events include earthquakes of 1783 and 1905 Calabria, and 1908 Messina-Reggio Calabria (M_w ~7.1). Deep-focus quakes reach ~500 km beneath Calabria and the Aeolian Islands, with seismicity reflecting slab rollback and crustal stretching (Polonia *et al*., 2016). Studies also suggest the presence of a slab tear at the southern edge of the subducting Ionian lithosphere (Polonia *et al*., 2016), which had been imaged by Godano *et al*. (2025) using b-more-positive method (Lippiello and Petrillo, 2024) tomography;
- Southern Tyrrhenian basin. A back-arc basin formed by slab retreat, with crustal thinning, extension, and volcanic activity (Aeolian Islands). Seismicity spans from shallow extensional and volcanic earthquakes to deep events marking the subducting Ionian plate, highlighting the interplay of extension, magmatism, and subduction (Chiarabba *et al*., 2008);
- Northern Sicily. Located at the junction of compressional and strike-slip tectonics between the Sicilian thrust belt and Calabrian Arc. Moderate to strong earthquakes occur inland and offshore, with the 1908 Messina earthquake as a key destructive event. Seismicity

is focused along the offshore northern Sicily fault system and inland thrust faults, with seismic risk driven by compression, strike-slip motion, and proximity to subduction (Totaro *et al*., 2016).

Overall, these diverse tectonic settings (collision zones, extensional belts, and active subduction) make Italy a highly seismically active region in the Mediterranean area, with a long history of destructive earthquakes and significant ongoing hazard.

2.2. NE Italy and western Slovenia

In addition to the national level, NESTORE was also applied to a region at the edge between Italy and Slovenia (red region in Fig. 1) (Gentili and Di Giovambattista, 2020; Brondi *et al*., 2025a) using the regional OGS (National Institute of Oceanography and Applied Geophysics) catalogue (Snidarcig *et al*., 2020). This area is characterised by the reduced spatio-temporal extent of the clusters and a lower productivity in terms of cumulative moment and number of earthquakes (for a given mainshock magnitude) with respect to the Italian seismicity (Brondi *et al*., 2025a). The predominant deformation regime varies from compressional in the western sector to strike-slip in the eastern sector. Several tectonic phases have reactivated old fault structures, leading to a fragmentation of the crust into different tectonic blocks that define the current seismotectonic zonation (Bressan *et al*., 2003). Seismic activity is mostly concentrated in two main fault systems: the Alpine system, with an approximate E-W orientation, and the Dinaric system, with a NW-SE orientation, respectively in the northern and eastern parts of the region. The earthquakes are shallow and typically occur at a depth of 6–12 km in the western area (Bressan *et al*., 2012, 2018) and 2–7 km in the eastern area (Bressan *et al*., 2009; Gentili and Gentile, 2015). Focal mechanism solutions indicate that thrust faulting prevails in the western part, while strike-slip faulting mechanisms are more common in the eastern part, with occasional normal faulting (Bressan *et al*., 2016 and references therein). The area experienced several strong earthquakes in the past [11 clusters of seismicity with earthquakes of intensity IX-X from 1117 to 1976, see e.g. Gentili and Franceschina (2011)] causing death and destruction. After the M_L = 6.4 earthquake in 1976, lower magnitudes followed, with 10 stronger events with latitudes between 45°–46.5° N and longitudes between 11.5°–14.5° E (a rectangle approximately 230 km in longitude and 170 km in latitude) in the magnitude range of 4.5–5.6 in the almost 50 years from 1977 to 2025 (Gentili and Di Giovambattista, 2020; Brondi *et al*., 2025a).

2.3. California

The works on seismicity in California (Gentili and Di Giovambattista, 2022; Gentili *et al*., 2023a, 2023b) represented a significant development of NESTORE. These papers introduced several improvements to the algorithm and code, resulting in NESTOREv1.0, now publicly available on GitHub (Gentili *et al*., 2023a).

California is among the most seismically active regions in the world, located along the complex transform boundary between the Pacific and North American plates (Hauksson and Shearer, 2005). The seismicity of the region is dominated by the right-lateral San Andreas fault system, the eastern California shear zone, and subsidiary faults such as the San Jacinto and the Hayward, creating distinct seismotectonic domains with different stress regimes and seismic productivity (Trugman and Ben-Zion, 2023).

Southern California, in particular, exhibits high seismicity rates due to the interaction of these major fault systems with a network of secondary faults, resulting in frequent moderate-

to-large earthquakes and diverse focal mechanisms (Hauksson and Shearer, 2005). This complexity drives the cluster behaviour and aftershock sequences observed across the region and has implications for seismic hazard assessment and operational forecasting (Trugman and Ben-Zion, 2023).

2.4. Greece

The newly developed NESTOREv1.0 algorithm was also applied to the seismicity of Greece (black region in Fig. 1) (Anyfadi, 2023; Anyfadi *et al*., 2023). Greece lies at the boundary between the Eurasian and African tectonic plates, where the complex and sometimes powerful seismicity is driven by various geological structures, such as the Hellenic subduction zone, stretching from the Ionian Islands to Crete (Le Pichon *et al*., 1995), while the North Anatolian Fault, a major strike-slip fault extending from Turkey into the Aegean region, transfers deformation westwards. The Hellenic subduction zone is one of the dominant geodynamic features in the eastern Mediterranean region. The Mediterranean lithosphere subducts beneath the back-arc area of the southern Aegean Sea, the tectonics of which are mainly driven by a regional field of extension (McKenzie, 1978). The complex geologic setting of the Aegean is dominated by two tectonic processes. The first is the northward movement and subduction of the African plate beneath the Aegean at a rate of 0.9 cm/year (McClusky *et al*., 2000; Reilinger *et al*., 2010), resulting in the formation of the Hellenic subduction zone, while the second is the southward migration of the Aegean plate due to the African slab rollback. The subduction of the African plate formed the Wadati-Benioff zone in the southern Aegean which extends down to about 170 km (Papazachos *et al*., 2000). The southern Aegean itself is a region where high seismicity occurs along the fore-arc. On the other hand, the back-arc region is characterised by moderate seismicity except from its eastern part with stronger seismicity. Dimitriadis *et al*., 2010). In front of the Hellenic Arc, Crete endures intense seismic activity due to active subduction and regional extension. Fault systems on Crete align in multiple directions, with the most recent and active normal faults trending WNW–ESE and NNE–SSW. The interplay of arc-parallel and arc-perpendicular extension contributes to significant vertical and horizontal displacements (Delibasis *et al*., 1999; Vallianatos *et al*., 2022). Furthermore, in the western part of Greece the Cephalonia transform fault marks the transition between the subduction zone of the Hellenic Arc and the zone of western Greece. In central Greece, the Corinth rift is rapidly extending and is a highly active break-up zone (Ambraseys and Jackson, 1997; Sakellariou *et al*., 2007).

In addition, the Peloponnese region experiences extensional tectonics with major active normal faults trending NE–SW and WNW–ESE, influenced by the subduction zone to the SW and the Corinthian Gulf to the north (Koukouvelas *et al*., 2010). Thessaly in the main land of Greece, shaped during the Neogene–Quaternary period, is characterised by a network of normal faults, grabens, and horsts trending NW–SE and E–W. These faults result from ongoing N–S and NE–SW extensional forces and demonstrate a mix of active normal, reverse, and thrust faulting, especially along the edges of major mountain ranges like the Pindos (Caputo and Pavlides, 1993). Southern Thessaly is seismically active with several large earthquakes occurring over the past century, while northern Thessaly, despite similar geological conditions, shows significantly less seismicity (Caputo, 1995; Vallianatos *et al*., 2021).

In northern Greece, the central Macedonia region lies within the inner Hellenic orogeny and exhibits complex tectonics. The extension of the N-S to NE-SW high-angle (50–80° dips) (Gkarlaouni *et al*., 2015) normal faults and rift basins formed since the Miocene and are consistent with the Aegean back-arc tectonics. The Mygdonian graben, aligned NW–SE, is especially active,

reflecting a broader N–S extensional regime tied to back-arc spreading behind the Hellenic subduction zone. In the Halkidiki peninsula, faults trend NW–SE, E–W, and N–S, with notable systems like the Stratoni-Varvara and Gomatius faults (Alatza, 2013). Overall, the seismic activity in mainland Greece and its islands is shaped by a combination of extensional and compressional forces, leading to a variety of seismotectonic regimes which make it one of the most interesting areas for the implementation of the NESTORE algorithm.

2.5. Japan

The last published application in chronological order of NESTORE is Japan seismicity (blue region in Fig. 1) (Gentili *et al*., 2025). Japan's seismicity is mainly driven by the convergence of multiple tectonic plates. The Pacific plate subducts beneath the Eurasian plate, along the Japan trench, producing powerful megathrust earthquakes. In the SE, the Philippine Sea plate also subducts beneath the Eurasian plate, generating significant interplate seismicity along the Nankai trough. Subduction rates are high, approximately 9–11 cm/year for the Pacific plate and 6–11 cm/year for the Philippine Sea plate (DeMets *et al*., 2010; Argus *et al*., 2011; Wu *et al*., 2016). In the north, interactions between the Eurasian and North American plates contribute to complex seismic patterns. Japan's earthquake distribution is highly heterogeneous and reflects the complex interplay between these tectonic systems. The eastern offshore region, encompassing the Japan trench, is characterised by frequent intermediate to large subduction earthquakes. The south-western margin, particularly along the Nankai trough, has a well-documented history of great earthquakes occurring in segmented patterns, e.g. the 1944 Tonankai and 1946 Nankai events (Kanamori, 1972). Inland Japan is also seismically active due to crustal deformation resulting from ongoing plate convergence. Numerous active faults, such as the Itoigawa-Shizuoka tectonic line and the Japan median tectonic line, accommodate interplate stress and can generate damaging shallow earthquakes, but these are just examples of regional-scale structures. Nonetheless, the seismicity is not only confined to the plate boundaries: very strong earthquakes also occur in central and western Honshu, Kyushu, and the Niigata and Nagano area with reverse or strike-slip faulting. Volcanic activity is another important source of seismicity in Japan, with hundreds of volcanoes distributed all along the archipelago islands with most of them still active. These volcanoes are associated with magma-induced seismic swarms and shallow volcano-tectonic earthquakes. In March 2011, the Tōhoku earthquake sequence occurred, starting with a M_w 7.3 foreshock on 9 March, followed by the devastating M_w 9.1 mainshock on 11 March. This event triggered a massive tsunami, resulting in significant human and environmental losses, including the Fukushima Dai-ichi nuclear disaster. The mainshock ruptured a ~400 km long segment of the subduction interface with an estimated slip of ~50 m (Lay, 2018), highlighting the destructive potential of subduction zone earthquakes. It was one of the largest earthquakes ever recorded. While events of this magnitude are rare, over 80 earthquakes with $M \geq 7$ were recorded between 1973 and 2023 in Japan and surrounding areas belonging to the Asia-Pacific region. To define the study area, we considered factors such as completeness magnitude, seismic network coverage, and the distribution of volcanic seismicity. Mount Fuji and Miyake Island, for instance, were excluded from our study.

3. Methods

NESTOREv1.0 consists of four main modules: the cluster identification module, training module, testing module, and NRT classification module. During these years, the key advancements of NESTORE development lay in the enhanced cluster identification methodology and implementation of stringent quality criteria for the training dataset and result validation, aimed at improving the algorithm's generalisation capability (Brondi *et al*., 2025a; Gentili *et al*., 2025). The modularity of the approach enabled us to substitute the cluster identification module if it was not effective for the current dataset, as in the Japan seismicity case (Gentili *et al*., 2025).

In all NESTORE versions, the analysis starts with the occurrence of an earthquake exceeding a certain magnitude threshold, provided it is not part of any previously identified cluster. This event is termed the "operative mainshock" (or o-mainshock). This concept is used because NESTORE aims to produce NRT forecasts during the early evolution of a seismic cluster, assessing the probability of a strong event occurrence after a moderate to intense o-mainshock. It is important to note that the operative mainshock may not be the actual mainshock of the cluster, as the true mainshock might not have occurred yet at the time of the forecast.

3.1. Cluster identification and selection

The easiest way to select clusters corresponds to methods known as the window-based methods, which consider all events within a given radius and time window from the stronger one(s) as part of the same cluster. The NESTOREv1.0 code (Gentili *et al*., 2023a) enables to select clusters by applying user-defined windows based on the magnitude of the stronger event(s). This method (see Section 3.1.1) supplied good results for most clusters in Italy, western Slovenia, California, and Greece (Gentili and Di Giovambattista, 2017, 2020, 2022; Anyfadi *et al*., 2023; Brondi *et al*., 2025a; Spassiani *et al*., 2025), while its application to Japan required a further, more complex approach. In fact, identifying earthquake clusters in Japan poses significant challenges due to the region's high seismicity rate and spatial-temporal overlap of nearby sequences (Gentili *et al*., 2025).

3.1.1. Window-based methods

In the window-based methods, any event with magnitude M_m greater than a fixed threshold M_{th} is considered as a potential main earthquake. The corresponding cluster aftershocks are detected in space and time by using a circular region centred on the main earthquake epicentre and a temporal window, with both the circle radius and the time duration depending on M_m. If a potential mainshock is in the cluster spatio-temporal window of a previous one, the two clusters are merged, and only one of the two is considered as the mainshock. The radius and duration of the cluster are defined based on some equations, generally referred to as <Proposer's name> (<year>) law. We selected different laws for different analysed areas by comparing different laws existing in literature with real data (further details in the corresponding papers). In detail, we used the following laws for each area:

- Uhrhammer (1986) and Lolli and Gasperini (2003) laws for the spatial and temporal selection of aftershocks, respectively, in Italy;
- Gentili and Bressan (2008) laws for both space and time for NE Italy - western Slovenia; in order to take into account the inaccuracies in localisation in the first years of the catalogue, 2 km have been added to the radius;

- Kagan (2002) law for space and Gardner and Knopoff (1974) law for time in California;
- Uhrhammer (1986) for both space and time windows in Greece.

3.1.2. Graph-based approach to the epidemic-type aftershock sequence: a method based on stochastic declustering

Traditional window-based methods may be inadequate because clusters from different sources often occur close in both time and space. This can lead to incorrect associations and contamination of clusters, negatively affecting the performance of classification algorithms such as NESTORE.

To overcome these limitations, we implemented a hybrid cluster identification method that integrates probabilistic and deterministic elements. The core of the method is based on the stochastic declustering algorithm developed by Zhuang *et al*. (2002), which models seismicity using the Epidemic-Type Aftershock Sequence (ETAS) model (Ogata, 1988). In this framework, each event may be either a background (independent) earthquake or may be triggered by a previous one, with associated probabilities estimated through maximum likelihood.

The ETAS model defines the conditional intensity function as the sum of a background term and a triggering term that accounts for the contributions of all preceding events. This includes temporal decay (Omori-Utsu law), spatial decay, and productivity scaling with magnitude. The model assumes magnitude independence between triggering and triggered events (Petrillo and Zhuang, 2022, 2023). ETAS model parameters are estimated using the original iterative algorithm by Zhuang *et al*. (2002) and Ogata and Zhuang (2006), ensuring reproducibility with prior studies.

Each event in the catalogue is assigned a probability of being background or being triggered by a specific preceding event. Stochastic declustering proceeds by generating a random number and assigning each event to the background or to a parent event probabilistically. Because of its stochastic nature, the clustering output varies across realisations.

The new algorithm proposed in Gentili *et al*. (2025), here referred to as the graph-based approach to ETAS (GRETAS), treats the ETAS output from the graph approach point of view. Different runs of the stochastic declustering algorithm supply a set of clusters that can be represented by oriented graphs (trees) that are different from one run to the other depending on the values of the random numbers generated by the algorithm. The nodes of the trees are the earthquakes, while the weight of the arcs between events *n* (parent) and *j* (descendent) is related to the probability for event *j* of being the background and to the sum of probabilities that *j* is the descendent of events up to *n*. The arc is generated only if its weight is larger than the random variable. The events that are strongly connected within a cluster, i.e. the ones that are close in space and time to several previous events, tend to remain grouped together across all runs of the stochastic declustering method. These events typically persist in the cluster because they are linked to multiple other events, not just one. For this reason, we chose to define cluster belonging based only on events that show strong and consistent connections to others. In order to do this, we ran stochastic declustering several times [in Gentili *et al*. (2025) we chose 10 times]. For each mainshock (defined as the largest event in the cluster), we identified the corresponding cluster members by intersecting the outcomes of the different runs. Only events consistently associated with the same mainshock across multiple runs were retained. This ensures that the final clusters include only the most stably connected events, avoiding the inclusion of loosely connected or ambiguous events that may belong to overlapping sequences.

To avoid excessive spatial spreading of clusters, the last part of the GRETAS algorithm applies a final spatial filter selecting only earthquakes within a given distance from the mainshock. In

Gentili *et al.* (2025), we selected the empirical law of Uhrhammer (1986) for Japan. However, this constraint had a marginal effect on the results, removing less than 1% of the events.

This robust identification process was applied uniformly to both the training and testing datasets, ensuring consistency in cluster definition throughout the study.

3.2. NESTORE training, testing, and near-real-time modules

3.2.1. Training

The clusters identified in the cluster identification module are the inputs for the training module. Nine features related to the seismic productivity and spatio-temporal evolution, in terms of radiated energy, magnitude, number of events, and their spatial distribution, are computed at different time intervals starting soon after the operative mainshock [for further details on the time start see Gentili *et al*. (2023a)]. A threshold is estimated for each of these features to distinguish type A clusters from type B clusters. Typically, most type A clusters have feature values above this threshold, whereas, in contrast, most type B clusters have values below this threshold. The training routine consists of four steps: 1) feature extraction, 2) decision tree training, 3) selection of the time intervals characterised by good performance for each feature (good interval), and 4) inheritance and validation. Since several improvements have been made over time, the training for versions of the algorithm starting with Brondi *et al*. (2025a) is described herein.

Feature extraction. The features, detailed in Section 3.3, are extracted from events with magnitude greater than, or equal to, M_m-2 in 10 time intervals *Ti*, starting s_1 = 1 minute after the occurrence of the o-mainshock and ending at steps of six hours on the first day and 24 hours during the first week.

Decision tree training. For each feature, a one-node decision tree is trained in each time interval so that, if the tree converges, it finds a threshold above which the majority of the feature values of the type A clusters lie. If no threshold value can be found, the feature is discarded for this time interval.

Selection of good interval. The reliability of the determined threshold values is tested with a *k*-fold cross-validation method with *k* = 1, i.e. the Leave-One-Out (LOO) method, which is particularly suitable for small training datasets (i.e. the order is of the dozens of cases). The decision tree performance is evaluated based on four statistical properties: accuracy, precision, recall (ranging between 0 and 1), and informedness (ranging from -1 to 1). These evaluators are used for evaluating binary classifier performance. The two classes are defined as positive and negative. In our case we considered the type A cluster as positive and the type B as negative. Accuracy is defined as the percentage of correct classifications. Precision is defined as the percentage of samples classified as positive that are actually positive. Recall is defined as the percentage of correct classification of positives; vice-versa, inverse recall, is the percentage of correct classification of negatives. All these percentages are normalised to 1. Informedness is the sum of recall and inverse recall -1 and is a measure of how informed the classifier is about the classes.

A feature is used by NESTORE in a certain interval, which is defined as “good interval”, if the evaluators fulfill certain criteria. More specifically:

- accuracy, precision, and recall must be strictly above 0.5;
- accuracy must be equal to or greater than that of a tree classifying each cluster in the more populated class;
- informedness must be strictly greater than 0.

The good interval ends at a time less than, or equal to, s_2 when informedness is maximal. The feature value and the threshold value for s_2 are 'inherited' into the subsequent time intervals.

Inheritance and validation. The constancy of the calculated and inherited feature values and thresholds is reinforced by reapplying the LOO method to calculate the percentage of correct classification of type A clusters (recall or true positive rate) and the percentage of incorrect classification of type B clusters (false positive rate) for each *Ti*. The feature thresholds of the *Ti* intervals, in which the false alarm rate is greater than the recall value, are discarded. For the selected features and time intervals, the probability that the cluster is of type A when the feature value is above or below the threshold is estimated from the training set as the percentage of A clusters with this feature above or below the threshold. Since the method is applied to real data, the probability may be less than one above the threshold and greater than zero below the threshold due to noise or imperfect performance of the feature in discriminating the cluster typology.

3.2.2. Testing

The testing module uses the values determined in the training module to forecast the class type of the clusters of an independent test set. For each test set cluster, the features selected in the training phase are calculated in the 10 time intervals *Ti* from the operative mainshocks of this new dataset and, depending on whether they are above or below the training threshold, the probability that the cluster belongs to type A is estimated for each feature and each time interval *Ti*.

Let $p_{n,i} = P(A|F_{n,i})$ be the probability that the cluster is of type A for the n^{th} feature F_n in the i^{th} time interval. In more recent versions of NESTORE [first used in Gentili and Di Giovambattista (2020)], these probabilities are combined using an approach based on the Bayes theorem to calculate the overall probability *P* that the cluster is of type A for each time interval *Ti*:

$$P_i\left(A \mid F_{1,i} \dots F_{N,i}\right) = \frac{[N(B)_i]^{N-1} \prod_{n=1}^{N} p_{n,i}}{[N(B)_i]^{N-1} \prod_{n=1}^{N} p_{n,i} + [N(A)_i]^{N-1} \prod_{n=1}^{N} (1-p_{n,i})} \quad (1)$$

where $N(A)_i$ and $N(B)_i$ are the number of clusters A and B at time *Ti* in the training dataset. It is important to know the number of samples for each class to account for the possible imbalance between classes. Eq. (1) was obtained from equations by Bailer-Jones and Smith (2011) for combining probabilities from different independent features, $D_1 \dots, D_n$:

$$P(C|D_1 \dots D_N) = a \frac{\prod_{n=1}^{N} P(C|D_n)}{P(C)^{\{N-1\}}} \quad (2)$$

where $P(C)$ is the probability of having class *C*, $P(C|D_n)$ is the posterior probability that class is *C* given D_n, and *a* is a normalisation factor so that $\Sigma_k P(C_k|D_1 \dots, D_N) = 1$.

In our case, we assumed that the probability of class type A and type B corresponds to the fraction of instances of such class in the training set [for further details, see Gentili and Di Giovambattista (2020)]. The output of the NESTOREv1.0 testing module is the estimation of the cluster type, expressed as probability $P(A)$.

In order to estimate the performance of the method, the forecasting about type A or type B clusters is provided by comparing the resulting probability $P(A)$ with the threshold equal to 0.5: if $P(A) \geq 0.5$ the cluster is classified as type A, otherwise as type B. The comparison of the output classification with the actual cluster class provides an estimate of the algorithm performance,

which is described by the receiving operating characteristics (ROC) and the precision-recall diagrams. Examples of the result diagrams can be found in Section 4.1.

It is important to note that, due to the small dataset available, validation is performed during training using the LOO method and for the individual features. The merging of the feature performance is only carried out in the testing phase, and there is no validation set to determine the best interval for each analysed region. For simplicity, in this paper, only the best results are provided in the dedicated section, but all data can be found in the corresponding articles.

3.2.3. Near-real-time application

The NRT module is the latest development of NESTOREv1.0, which aims to forecast the class type of an ongoing seismic cluster. The first description of this module is in Gentili *et al*. (2023a). According to the NESTORE general procedure, the cluster is defined with a window-based, magnitude-dependent method from the mainshock, which also determines the time and area for the forecasting. The features are calculated for increasing time intervals, *Ti*, until the one for which the testing procedure estimated the classification performances are the best, and compared with the corresponding values and thresholds obtained in the training for the study area. The NRT output consists of a map of the cluster in each time interval and the class forecast, given in the form of an AFtershock-based Traffic Light System (AFTLS) (Gentili *et al*., 2023a). The AFTLS colours are related to the probability of occurrence of a strong aftershock: red for high probability (type A), green for low probability (type B), and yellow for uncertain assessment.

Before applying the NRT module, training is carried out once again with all the available data (training set and test set) in order to obtain more stable thresholds. Like in the case of previous training, validation is performed by means of the LOO method, without using a dedicated validation set. In Brondi *et al*. (2025a) and Gentili *et al*. (2025), the thresholds obtained with the larger dataset are compared with the previous ones, detailing which are stable and which are improved. Since the NRT module needs an estimate of which are the more reliable time intervals for the forecasting, but no proper validation set is available, this information is taken from the output of the testing phase (see Section 5 for further details).

3.3. NESTORE features

The features used in NESTORE (see Gentili *et al*., 2025) are related to the spatio-temporal distribution of the aftershocks, source area, magnitude trends, and radiated energy. They are calculated for events with magnitude in the range [M_m-2, M_m], chosen in order to account for the possible incompleteness of the catalogue. They are:

a) *S,* the normalised event source area:

$$S(i) = \sum_i 10^{(m_i - M_m)} \tag{3}$$

where m_i is the i^{th} event magnitude;

b) *Z,* the linear concentration of events:

$$Z(i) = \frac{mean(10^{0.69 m_i - 3.22})}{mean(r_{ij})} \tag{4}$$

with r_{ij} being the distance between the i^{th} and j^{th} events;

c) *Q,* the normal radiated energy:

$$Q(i) = \frac{\sum_i E_i}{E_m} \tag{5}$$

being E_m the mainshock energy, and E_i the energy of the i^{th} event;

d) *SLCum,* the cumulative deviation of *S* from the long-term trend on increasing length windows:

$$SLCum(i) = \sum_i abs[S(t_i) - S(t_{i-1})] \frac{i \cdot dt}{(i-1) \cdot dt}; \tag{6}$$

e) *SLCum2,* the cumulative deviation of *S* from the long-term trend on sliding windows:

$$SLCum2(i) = \sum_i abs\{S[s_1 + (i-1) \cdot dt, (s_1 + 1) \cdot dt] - S[s_1 + (i-1) \cdot dt, s_1 + (i-1) \cdot dt + d\tau]\} \frac{dt}{d\tau}; \tag{7}$$

f) *QLCum,* the cumulative deviation of *Q* from a long-term trend on increasing length windows:

$$QLCum(i) = \sum_i abs[Q(t_i) - Q(t_{i-1})] \frac{i \cdot dt}{(i-1) \cdot dt}; \tag{8}$$

g) *QLCum2,* the cumulative deviation of *Q* from the long-term trend on sliding windows:

$$QLCum2(i) = \sum_i abs[Q([s_1 + (i-1) \cdot dt,\ s_1 + i \cdot dt]) - Q([s_1 + (i-1) \cdot dt, s_1 + (i-1) \cdot dt + d\tau]) \frac{dt}{d\tau}]; \tag{9}$$

h) V_m, the cumulative variation of magnitude from event to event:

$$V_m(i) = \sum_i |m_i - m_{i-1}|; \tag{10}$$

i) N_2, the number of events with magnitude ≥M_m-2:

$$N_2(i) = \sum_i H[m_i - (M_m - 2)] \tag{11}$$

where *H* is the Heaviside step function.
In the case of the application of NESTORE to Japan (Gentili *et al*., 2025), we substituted N_2 with:

i') $$N_{2S}(i) = N_2(i) + 110 \cdot S(i) \tag{12}$$

which performed better in that area.

3.4. Data cleaning by outlier removal – REPENESE algorithm

The biggest challenge of NESTORE is the classification of small and often very unbalanced datasets such as, for example, in the Japanese seismicity case. For this reason, starting with the paper of Gentili *et al*. (2025), we developed the RElevant features, class imbalance PErcentage, NEighbourhood detection, SElection (REPENESE) algorithm. A preliminary quality assessment of the training dataset is essential to achieve reliable results and to establish more appropriate thresholds for discriminating between cluster populations. The detection of outliers must fulfil three important requirements: 1) the selection must be solely based on relevant features, 2) the imbalance between classes must be taken into account (removing a rare cluster of type A is more critical than removing a frequent cluster of type B), and 3) the standard centroid distance methods such as *Z*-score (Rousseeuw and Hubert, 2011) must be replaced by a threshold-based approach, since correct classification depends on the position relative to a feature threshold rather than the proximity to a centroid.

To overcome these challenges, REPENESE (Gentili *et al*., 2025) involves four steps:

- RE (RElevant features). Features are selected as relevant if a threshold exists and the true positive rate is greater than 0.5, the false positive rate is smaller than 0.5, and the precision is greater than 0.5;
- PE (PErcentage class weighting). The imbalance between the classes is quantified using probabilities P_A and P_B calculated from the relative frequencies of clusters of type A and B in the training set;
- NE (NEighbourhood detection). Neighbourhoods are defined for each relevant feature based on the sorted feature values. Type A samples consider the nearest N_1 (set to 5) larger values, type B considers the N_1 smallest values;
- SE (SElection). A sample of a given class (A or B) is labelled as an outlier if the number of samples of the same class in its neighbourhood does not exceed $P{\cdot}Sn$, where Sn is the size of the neighbourhood and P is probability P_A or P_B. Only samples that are consistently identified as outliers across all relevant features and thresholds are removed from the training set.

This method increases the robustness of the training set by ensuring a better representation of minority classes and reducing noise, which ultimately improves the generalisation capability of the model.

4. Results

This section summarises the more recent results for Italy, NE Italy and western Slovenia, California, Greece, and Japan. For California, reference is made to the results available on GitHub and Zenodo (Gentili *et al*., 2023a), where a version of the code, more recent than the one in the paper by Gentili and Di Giovambattista (2022), is available.

In Section 4.1., training and testing performances in different regions are compared. Section 4.2 provides results on the NESTORE thresholds in different regions, outlining their differences. Section 4.3 yields further statistical information on the evolution of the clusters in space-time and the seismic moment, which can be provided together with the cluster class at the time of forecasting. Finally, the NRT application of NESTORE is presented in Section 4.4. In the final section, the outcome of the results is discussed.

4.1. Training and testing

The clustering module identifies seismic clusters from the input catalogues. However, not all clusters are suitable for NESTORE. Clusters are discarded under various conditions: 1) if the mainshock is outside of the definition areas (see Fig. 1); 2) if the strongest aftershock occurs within the first six hours after the mainshock, limiting the availability of statistical data; 3) if the magnitude of the strongest aftershock falls within the range of M_m−1±0.2, to avoid incorrect class assignments due to inaccuracies in magnitude estimation; 4) if the completeness magnitude is greater than M_m-2.

The cardinality of the refined training set used for the first six-hour time interval of each analysed area is shown in Table 2.

Table 2 - Training set and test set cardinality for the analysed areas.

Region	Dataset	Period	No. of clusters	No. of A clusters	No. of B clusters
California	Training	1981–2019	50	21	29
	Testing	1981–2023	13	4	9
Greece	Training	1995–2015	46	6	40
	Testing	2016–2022	27	6	21
Italy	Training	1980–2009	24	7	17
	Testing	2010–2020	14	6	8
NE Italy - western Slovenia	Training	1977–2009	13	4	9
	Testing	2010–2020	18	3	14
Japan	Training	1973–2004	50	7	43
	Testing	2005–2023	31	4	27

The table shows how the ratio between the number of type A and type B clusters changes from region to region and how, in Japan and Greece, the classes are highly unbalanced, with type B cardinality being six to seven times greater than that of type A. In order to account for the class imbalance and the presence of outliers, we introduced in our work on Japan the REPENESE algorithm (see Section 3.4). After its application, the number of type A clusters in the training set decreased from 11 to 7 and of type B from 45 to 43 (the final number is the one used for training and shown in the table).

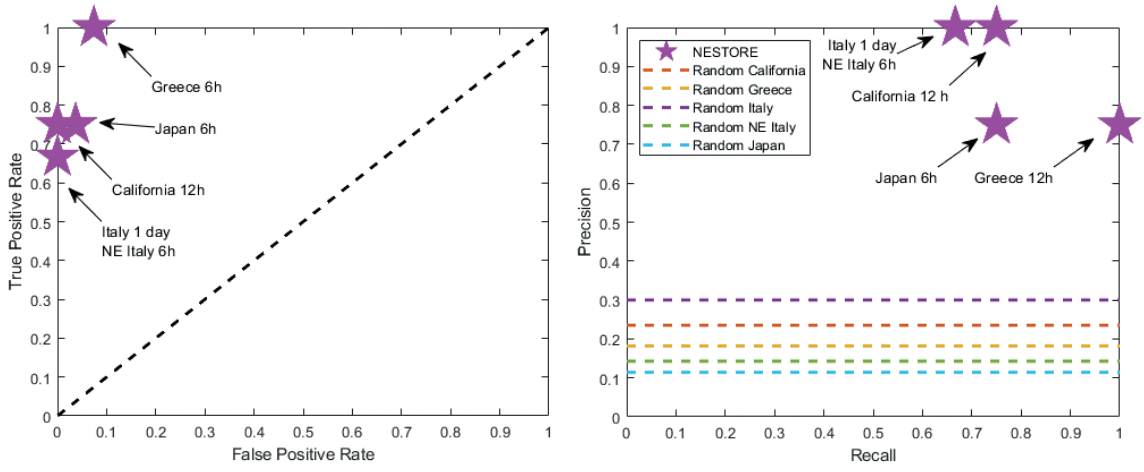
The cluster selection process described above (points 1 to 4) is also applied to the clusters of the test set, but the REPENESE algorithm is not applied because the outliers should be taken into account in the performance evaluation. See Table 2 for the cardinality testing datasets of all the analysed regions.

The performance of NESTORE in the different regions varies according to the region characteristics, the catalogue quality, the number of available clusters, the balance of classes A and B, and so on. In articles on NESTORE, the testing module estimates and visualises the performances through the ROC and precision-recall diagrams, suitable for binary classifications. Fig. 2 shows the results corresponding to the time interval *Ti* for which the algorithm provided the best performances in the analysed regions. The ROC diagram (Fig. 2a) shows, on the ordinate axis, the true positive rate (or recall), i.e. the percentage of correctly classified positives that in our case is the percentage of correctly forecasted as type A clusters (normalised to one). On the abscissa axis, it shows the false positive rate, which in our case corresponds to the percentage of

type B clusters incorrectly classified as type A (normalised to one). The dashed line corresponds to a random response. All applications are clearly above the random response and close to the ideal classification that is the one in which the true positive rate is equal to one and the false positive rate is equal to zero.

If the negative class (in our case type B) has a much greater cardinality than the positive one, even if a small percentage of the negative class is misclassified as positive, a large percentage of the events classified as positive can be misclassified.

A crucial parameter to evaluate in the case of a heavily imbalanced class distribution is, therefore, precision, which in our case is the percentage of clusters classified as type A that are actually type A. This information is shown in the precision-recall diagram (Fig. 2b). Horizontal dashed lines show the response of a random classifier, i.e. a value of precision equal to the normalised percentage of type A clusters in the dataset for different regions in the test set. Also, in this case, the performance in different regions is close to the ideal precision=recall=1 classification, corresponding to the upper right corner in the diagram.



(a) (b)

Fig. 2 - NESTORE performances in different regions.

The large unbalancing of the classes of the test set in Japan and Greece (corresponding to a low value of random classifier precision) slightly affects precision in the two regions, which is however greater than 70%. In the case of Italy and NE Italy, no type-B cluster is misclassified, so that precision is one.

As a further analysis, in order to validate our results, we estimated probability α of obtaining h or more hits (type A clusters correctly classified) by chance accordingly with Zechar (2010) as:

$$\alpha = \sum_{i=h}^{N} \binom{N}{i} \tau^i (1-\tau)^{N-i} \,. \tag{13}$$

In this case, τ, the fraction of the time-space occupied by alarms, is simply the ratio between the number of clusters classified as A and the number of all the clusters; unlike applications in evaluating mainshock forecasting, in which the measure of time-space may be not uniquely

defined, in this particular application τ is uniquely defined, avoiding ambiguous results; in this case N is the number of observed type A clusters. A small α corresponds to an alarm with high skill.

Table 3 shows the value of α for the different analysed regions. Note that for all the cases analysed, the number of type A clusters in the training set was unchanged with respect to the one of Table 2, even when longer time intervals were considered. For this reason, such values are not specified in the table.

Table 3 - Values of α for the different regions.

Region	Time interval	α
California	12 hours	4.06 %
Greece	6 hours	0.07 %
Italy	24 hours	5.97%
NE Italy - western Slovenia	6 hours	3.41 %
Japan	6 hours	3.36 %

In all cases, the probability α of obtaining h or more hits by chance is very low, lower than 6%. The highest value is obtained for Italy, where two type A clusters were wrongly classified. However, a detailed analysis (Gentili *et al*., 2024) of one of the two clusters which occurred in Molise (southern Italy) in 2018 showed how its misclassification may be due to an incorrect cluster identification caused by effects related to fluid diffusion. Considering the proposed cluster definition by Gentili *et al*. (2024), α is reduced to 2.45% and recall (true positive rate) passes from 0.67 (Fig. 2) to 0.83.

4.2. Comparison between feature thresholds

In this section, the feature thresholds encountered in different regions (Italy, NE Italy and western Slovenia, California, and Japan) are compared by analysing the results of papers from 2023 (Anyfadi *et al*., 2023; Gentili *et al*., 2023a, 2025; Brondi *et al*., 2025a).

For this purpose, Fig. 3 shows, when available (Italy, NE Italy and western Slovenia, Japan), the thresholds obtained from the whole dataset after outlier removal, being this the most stable characterisation of the areas. Since for all cases the best performances were obtained with time intervals smaller or equal to one day, only these thresholds are compared. It should be noted that the values for Italy and NE Italy - western Slovenia may differ, as Italian seismicity is dominated by Apennine seismicity (Brondi *et al*., 2025a), which is different from Alpine-Dinaric seismicity in NE Italy. The thresholds inherited from previous intervals are not shown. For details on the thresholds obtained during the original training, reference should be made to the corresponding papers.

As shown in Fig. 3, all feature threshold values increase or remain constant with increasing length of time interval *Ti*. This is due to the fact that most features (e.g. the number of events or radiated energy) are cumulative and, therefore, they cannot decrease, and that the larger productivity, especially for type B clusters, occurs in the first hours after the operative mainshock, so some thresholds do not need to increase. When available, the threshold values for the cumulative source area (*S*), radiated energy (*Q*), their time variation (*SLCum*, *SLCum2*, *QLCum*,

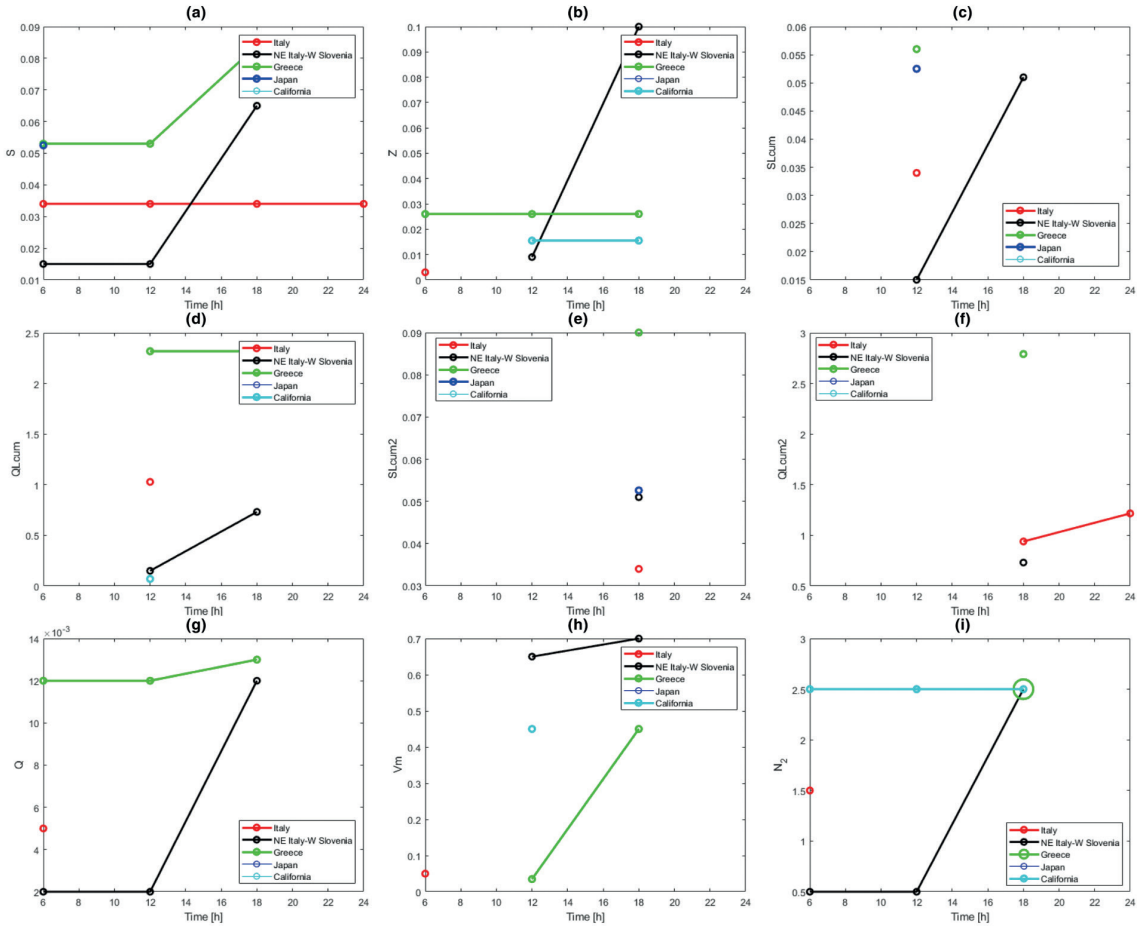


Fig. 3 - Thresholds for each feature and time interval (when available) for the analysed regions.

and *QLCum2*), the number of events (N_2), and the earthquake concentration (Z) generally show higher values for Greece and Japan compared to Italy and NE Italy, while California does not show a general trend with respect to the other regions. The *Vm* feature, which refers to the fluctuation in magnitude from one event to the next, does not show the region-dependent trend of the other features.

Focusing on the partially overlapping areas of Italy and NE Italy - western Slovenia, the threshold values in NE Italy - western Slovenia are found to be lower at short time intervals after the mainshock. The trend is inverted for the time interval ending 18 hours after the mainshock (Brondi *et al.*, 2025a). One reason for this inversion could be related to the change in the size of the dataset. Since we only consider type A clusters that have not yet had the strongest aftershock at the time of the analysis, the data set shrinks over time, possibly increasing noise in the data. Another explanation could be a selection effect in the population of type A clusters, if the type A clusters with the strongest aftershock at a later point in time are also characterised by higher feature values.

4.3. Further information based on classification

When NESTOREv1.0 was applied to the catalogues of California (Gentili and Di Giovambattista, 2022), Greece (Anyfadi *et al*., 2023), Italy (Gentili and Di Giovambattista, 2017, 2020; Brondi *et al*., 2025a), and Japan (Gentili *et al*., 2025), type A and type B clusters were found to differ in terms of various seismicity characteristics in the first hours or days after the operative mainshock (Gentili *et al*., 2023a, 2025).

Recently, it has been observed that for Italy, NE Italy - western Slovenia (Brondi *et al*., 2025a) and Japan (Gentili *et al*., 2025), type A and type B clusters also differ with regards to three parameters measured over the entire duration of the cluster: the effective cluster duration normalised by the cluster duration derived from the window method, the number of aftershocks with a magnitude greater than, or equal to, M_m-2, and the cumulative seismic moment associated with previous aftershocks normalised by the seismic moment of the operative mainshock. In most cases, the type A clusters show greater values than the type B clusters in all three areas mentioned. This corresponds to what was observed with the NESTOREv1.0 features in the first hours after the o-mainshock.

Comparing the distribution of these characteristics for the whole clusters (Brondi *et al*., 2025a; Gentili *et al*., 2025) in different regions, we found the following:

1. Japan and Italy are characterised by a higher number of events with $M \geq M_m$-2 for type A with respect to NE Italy - western Slovenia (no type A clusters with less than three events, while this is the case in NE Italy - western Slovenia in one third of the cases);
2. although the minimum value of normalised cumulative seismic moment associated with type A clusters is higher in the case of NE Italy - western Slovenia, for all three areas most type A clusters are above 0.2;
3. for type B clusters, the normalised cumulative seismic moment is always confined between 0 and 0.2 and, more specifically, always below 0.1 in the case of NE Italy - western Slovenia and Japan. Therefore, on average, Italy shows a slightly higher type B cluster productivity compared to NE Italy - western Slovenia and Japan;
4. normalised duration of type A clusters in Italy and NE Italy - western Slovenia are comparable; data on Japan cannot be compared because of the different cluster identification method adopted. However, in all cases, the normalised duration of type A clusters is generally higher than that of type B ones.

4.4. NESTORE application in near-real time

The main goal of the NESTORE algorithm is to provide an alarm when in the first hours/ days after a strong earthquake, the values of seismicity feature are consistent with the ones measured for a type A cluster. For this purpose, we proposed a NRT module. In the following sections, the performance of the NRT module is shown for the time intervals corresponding to the best performances in the testing phase. The application of this module to the various regions analysed by NESTORE is at different stages of development. The NRT module is functional for Greece and California, but is not currently being used. In Japan, automatic cluster identification for NRT is still under development, as the window-based methods are not suitable for the seismicity of this area. In Italy and NE Italy - western Slovenia, the NRT module is operational and routinely employed.

4.4.1. Italy

Fig. 4 shows the performances of the NRT method in Italy from 2021 to 2024. With this method, the analysed clusters should meet the following requirements:

1. the location must be within the analysis area (see Fig. 1);
2. the mainshock magnitude must be greater than, or equal to, 4 (as the completeness magnitude was set to two for the whole Italian territory after 2020);
3. no strong aftershock must be recorded in the first six hours after the operative mainshock.

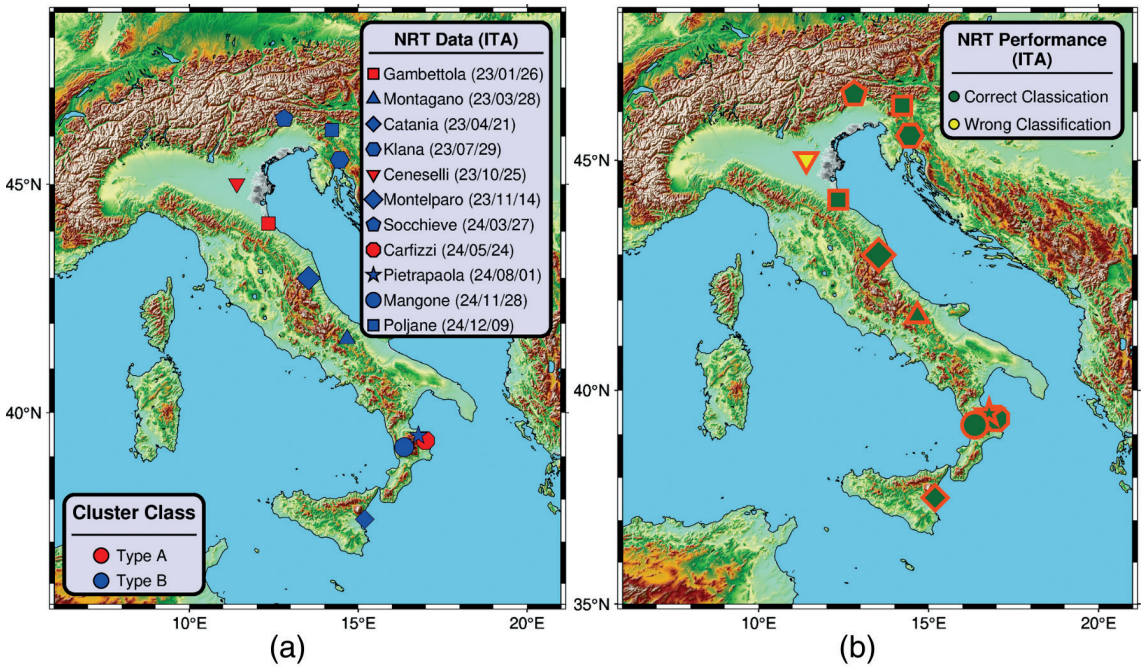


Fig. 4 - NRT module performances on the INGV catalogue.

The code is manually launched when the previous conditions are satisfied. Even though the dataset is not very large, due to the short duration of the NRT module evaluation (11 available clusters), the performances of the method are in good agreement with those obtained in the testing phase, with all type B clusters correctly forecasted (false positive rate = 0), two thirds of type A clusters correctly forecasted (true positive rate = recall = 0.66), and all clusters classified as A actually being of type A (precision = 1).

4.4.2. NE Italy and western Slovenia

Fig. 5 shows the performance in NE Italy and western Slovenia using the OGS catalogue (Snidarcig *et al*., 2021; Brondi *et al*., 2024b, 2025b; https://terremoti.ogs.it/). The limited size of the area leads to a dataset that is smaller than the one of the entire Italian territory. Originally, the NRT module was started manually after the occurrence of an earthquake with magnitude M_L ≥ 3.7 that was not already part of an analysed cluster. Since March 2024, the analysis has been carried out automatically (see next section).

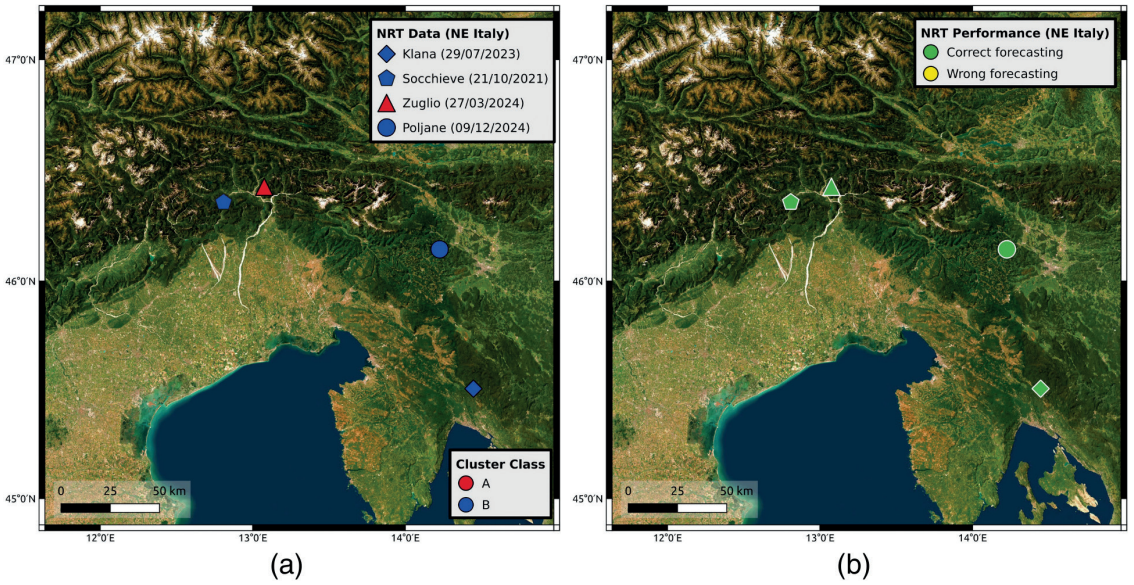


Fig. 5 - NRT module performances on the OGS catalogue.

All clusters analysed through the NRT module were correctly forecasted. As with training and testing, the eastern part of the area is characterised by type B clusters. From 2024, the NRT algorithm is automatically applied to the monitoring system managed by the department Seismological Research Centre (CRS) of OGS. For further details see the Appendix.

4.4.3. Greece

In Greece the NRT module has not been applied in a systematic way, but was applied in the thesis work by Anyfadi (2023). A total of four clusters, including data from 1995 to 2022, were analysed and 75% were correctly classified (Fig. 6).

Among the three type B clusters, one was misclassified as type A. The M_m = 4.9 type A cluster, located in the Evia area, was correctly classified by NESTORE. The more relevant features are *S*, *Z*, and *Q* at the six hours. The probabilities for the majority of these features exceed the threshold indicating that these features voted for the classification of this cluster as type A and a strong aftershock with magnitude larger than, or equal to, 3.9 was correctly expected to occur within a circular region with radius of 19 km, in the 25 days after the o-mainshock. The number of clusters considered in this analysis is too small to enable a fully robust statistical evaluation. Nevertheless, the method's performance is consistent with the testing phase: the type A cluster was correctly classified (as were all type A clusters in the testing phase), while the only misclassifications occurred among the type B clusters, which also produced some false alarms during testing. A routine application of NESTORE to Greek seismicity would enable a more reliable performance assessment. The analysis of the misclassified type B clusters did not reveal any region-specific characteristics. In Anyfadi *et al*. (2023), the two misclassified type B clusters were located in eastern Crete and near the Anatolian peninsula coast, far from the location of the B cluster misclassified by the NRT algorithm.

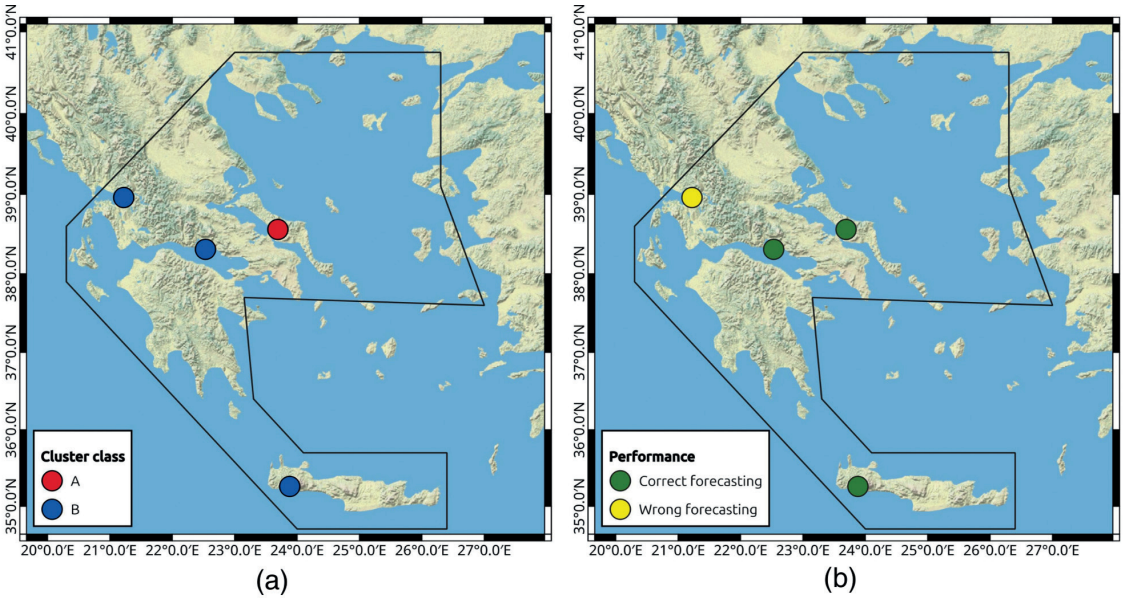


Fig. 6 - NRT module performances on the AUTH catalogue. **Editor note: please cite Fig. 6 in the text**

5. Discussion and conclusions

In this paper, we have shown the performance of the NESTORE algorithm in classifying seismic clusters based on the characteristics of the first hours of seismicity after the o-mainshock. NESTORE is based on a set of features extracted from seismic catalogues. It assumes that type A clusters (those with the strongest aftershocks) have large feature values. In contrast, type B clusters typically show smaller values. The classification is based on threshold feature values determined in the training phase. In all analysed regions, feature *S*, which refers to the normalised cumulative source area, has shown good results shortly after the o-mainshock (six hours).

The capability of the *S* feature, in discriminating type A and type B clusters in most datasets for a short time after the mainshock, may be interpreted from the physical point of view. Similarly to the analysis by Smirnov and Petruchov (2025) on precursors of strong earthquakes, we can state that, since the value of the normalised cumulative source area *S* is obtained from the aftershock magnitude and, since the size of the sources is related, from a physical perspective, to the size of lithospheric heterogeneity in fracturing (Scholz, 2019), the dimensional distribution of heterogeneities plays a crucial role in determining the distribution of aftershock magnitudes (Aki, 1981). The development of fracturing and the subsequent stress redistribution lead to the widening of cracks as they merge. The tendency of cracks to coalesce into larger cracks, leading to the development of strong aftershocks, depends on the fracturing state of the medium, the stress field, the orientation of cracks relative to the stress field, and the presence of fluids, and varies in time and space. Clusters in which this tendency is most pronounced are those where strong aftershocks (high *S*), with magnitude comparable to that of the mainshock, are most likely to occur (type A clusters). This result is independent of the analysed region. Type A clusters, on the other hand, are also the ones characterised by strong *S* fluctuations (as shown by *SLCum* and *SLcum2*) and higher magnitude variations from event to event (as shown by *Vm*). This shift can be interpreted as a manifestation of instability within the nonlinear system of earthquake-generating faults (Vorobieva, 1999), consistent with observations preceding major earthquakes

and stronger events within seismic clusters (Keilis-Borok and Kossobokov, 1990; Keilis-Borok and Rotwain, 1990; Vorobieva and Panza, 1993).

From the machine learning point of view, the variability in feature performance across seismotectonic regimes highlights an important point: including a variety of features during training helps the model to adapt to regional differences. Some features show different performance depending on the analysed area: the *Q* feature, referring to the cumulative radiated energy, works well in Greece, Italy, and NE Italy - western Slovenia, but not in California and Japan; feature N_2, based on the earthquake number, gives good results in California, Italy and NE Italy - western Slovenia but not in Greece and Japan.

Overall, NESTORE correctly classified between 66% and 100% of type A clusters, and between 90% and 100% of type B clusters. Precision, corresponding to the percentage of clusters classified as type A that actually were type A, ranged between 75% and 100%, depending on the class imbalance (lower precision if type B clusters are more than type A clusters). Most of these results are obtained within the first 6–12 hours or, at the most, within one day after the o-mainshock. Therefore, the method shows a great potential for application in seismic risk mitigation.

A recent and interesting result is that type A and type B clusters differ not only in terms of the strongest aftershock magnitude, but also in terms of the characteristics of the entire cluster, such as the cumulative seismic moment, duration, and number of earthquakes. The analyses for Italy, NE Italy - western Slovenia, and Japan have shown that type A clusters have higher values for seismic moment, duration, and number of earthquakes compared to type B clusters with the same o-mainshock magnitude. Although the clusters in Italy and Japan depend on different seismotectonic contexts, seismic network characteristics, clustering laws, and magnitude scales, it is interesting to note that, to a first approximation, the different behaviour between type A and type B clusters is the same with respect to the three analysed characteristics. Shortly after the o-mainshock, the results on these three seismicity characteristics provide further information on how the ongoing cluster will behave.

Another interesting outcome of NESTORE is its capability of estimating feature thresholds robust enough to enable class forecasting in a whole region, even when multiple and diverse seismotectonic regimes coexist. The next step in our research investigation will be to understand whether the different threshold values from region to region are related to seismicity characteristics. We, therefore, compared it with the geometric mean of the annual seismic rate obtained by analysing the same catalogues. Recently, van der Elst and Page (2023) proposed a new method to estimate the seismic rate in the presence of incomplete data. This method, called a-positive (a+), is based on measuring interevent time interval not between all earthquakes, as in the standard definition above, but only between pairs of earthquakes where the second event has a larger magnitude than the first. Specifically, the difference in magnitude between the two earthquakes in the pair must exceed threshold value dM_{th}. By rescaling Δt_i with , and the expected seismic rate can be estimated as if the catalogue were complete.

Table 4 shows the geometric mean for parameter a+. We used the geometric mean because the distribution of the a+ values is very skewed and long-tailed, due to the considerably heterogeneous and clustered seismicity nature. Greece and Japan have the highest productivity value; Italy and California are in the middle range, while NE Italy - western Slovenia has the lowest values. We hypothesise that the threshold values of the features may be related to the seismic rate. When the rate is higher, both the type A and type B clusters show higher values for the features and, thus, higher thresholds while, conversely, a lower rate leads to lower thresholds for both type A and type B clusters.

Table 4 - Geometric mean of annual seismic rate a+.

Region	Geometric mean a+
Greece	$2.767282 \cdot 10^6$
Italy	$3.727444 \cdot 10^5$
Japan	$1.058114 \cdot 10^6$
California	$4.393627 \cdot 10^5$
NE Italy - western Slovenia	$3.128294 \cdot 10^3$

While the relationship between the seismicity rate of the whole catalogue and, for example, the cumulative radiated energy or the cumulative source area of the clusters is not obvious, the a+ value in the whole catalogue can be more easily related to the number of events with a magnitude larger than, or equal to, M_m-2 in the clusters. The seismicity rate simply behaves the same way in the whole catalogue and in the clusters, and regions with higher a+ values are also those with a higher number of events with magnitude larger than, or equal to, M_m-2 in both the type A and type B clusters. As shown in Section 4.3, Japan and Italy are characterised by a higher number of events with $M \geq M_m$-2 for the A type compared to NE Italy - western Slovenia. This difference is also reflected in the threshold on N_2, which only covers the first hours of the cluster and is smaller for NE Italy - western Slovenia than for Italy; for Japan, the decision tree for N_2 does not converge.

Understanding how differences in fault geometry, slip rates, and tectonic loading can influence seismic patterns is essential for the future improvement of machine learning methods such as NESTORE, which aims to forecast strong aftershocks in different tectonic settings. Trugman and Ben-Zion (2023) conducted a comprehensive analysis of the productivity of earthquake sequences in California and Nevada using a 40-year seismicity catalogue and nearest-neighbour clustering. They showed that productivity variations are strongly related to geophysical parameters such as heat flow, fault complexity, and deformation velocity. Highly productive sequences are found in areas such as the Walker Lane, while lower productivity sequences are found along the San Andreas fault. Using XGBoost and SHAP, they showed how these physical conditions influence the clustering and triggering potential of earthquakes.

Their results provide a useful framework for interpreting the classification of type A and type B clusters in NESTORE. Since high heat flow, high fault density, and high complexity are associated with high productivity, these parameters could help explain why certain clusters evolve into type A clusters while others remain type B.

Both Gentili and Di Giovambattista (2022) and Trugman and Ben-Zion (2023) use interpretable machine learning approaches and focus on seismicity in California seismicity, showing potential from the combination of seismic and geophysical features to improve NESTORE forecasts.

NESTORE is still under development and several studies are in progress, among which the application to New Zealand (Caravella and Gentili, 2025) and the use of new features in improved catalogues such as that of the Rapid Assessment of Moment and Energy Service, which also includes the energy and seismic moment (Brondi *et al*., 2024). Furthermore, routine testing has been conducted in Italy and in the NE Italy–western Slovenia region since 2021. The analysis is based on the real-time version of the catalogues, which may have reduced accuracy in event location and detection performance. Although the proposed results require validation on a larger test set, they show performance levels consistent with those achieved during the testing phase, confirming that NESTORE also performs well on unrevised catalogues.

Newer performance assessments are planned where the best performance interval for NESTORE, as shown in Fig. 2, is estimated not only in the predefined test set but by a *k*-fold approach (Caravella and Gentili, 2025). This approach could be particularly important for the NRT module, where NESTORE is applied to new clusters after training with both the old training set and the old test set, while the window in which performances are considered optimal is estimated based on the old test set only.

Acknowledgments. Following the presentation at the recent GNGTS conference (Gentili *et al.*, 2025), the work was recognised by the conveners as worthy of publication in this special issue of the Bulletin of Geophysics and Oceanography. This work is co-funded within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan - NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005) and by the grant "Progetto INGV Pianeta Dinamico: Near real-time results of Physical and Statistical Seismology for earthquakes observations, modelling and forecasting (NEMESIS)" - code CUP D53J19000170001 - funded by the Italian Ministry of Education, Universities, and Research ("Fondo Finalizzato al rilancio degli investimenti delle amministrazioni centrali dello Stato e allo sviluppo del Paese", Fund for the relaunch of investments by central State administrations and for the development of the country, established by Italian Law 145/2018). G.P. would like to acknowledge the Earth Observatory of Singapore, and the Singapore Ministry of Education Tier 3b project "Investigating Volcano and Earthquake Science and Technology (InVEST)". G.P. and J.Z. are also partially supported by the Japan Ministry of Education, Culture, Sports, Science and Technology (MEXT) project for seismology Toward Research innovation with data of earthquakes (STAR-E), Grant Number JPJ010217. Another project, currently ended, strongly contributed to the results used in this analysis: the bilateral project Italy-Japan funded by a grant from the Italian Ministry of Foreign Affairs and International Cooperation was pivotal for NESTORE development. E.-A.A. was partially supported by the Erasmus+ programme of the European Union. Map figures have been realised using QGIS (http://www.qgis.org), while other analyses have been performed using ZMAP (http://www.seismo.ethz.ch/en/research-and-teaching/products-software/software/ZMAP). The NESTOREv1.0 toolbox is available for free download from GitHub at the address https://github.com/StefaniaGentili/NESTORE and its reproducibility package is available at the Zenodo website https://zenodo.org/account/settings/github/repository/StefaniaGentili/NESTORE.

REFERENCES

Aki K.; 1981: *A probabilistic synthesis of precursory phenomena*. In: Simpson D.W. and Richards P.G. (eds), Earthquake prediction: an international review, Am. Geophys. Union, Washington, DC, USA, Vol. 4, pp. 556-574, doi: 10.1029/ME004p0566.

Alatza S.; 2013: *Geodetic determination of tectonic displacements in Chalkidiki*. Bachelor's thesis in Geological Sciences, National Technical University of Athens School of Surveying and Surveying Engineering, Athens, Greece.

Ambraseys N.N. and Jackson J.A.; 1997: *Seismicity and strain in the Gulf of Corinth (Greece) since 1694*. J. Earthquake Eng., 1, 433-474, doi: 10.1080/13632469708962374.

Anyfadi E.-A.; 2023: *Forecasting strong aftershocks in Greek earthquake clusters using the NESTORE machine learning algorithm*. PH.D. Thesis in Geological Sciences, Hellenic Republic, National and Kapodistrian University of Athens, Department of Geology and Geoenvironment, Sector of Geophysics and Geothermy, Athens, Greece, 177 pp.

Anyfadi E.-A., Gentili S., Brondi P. and Vallianatos F.; 2023: *Forecasting strong subsequent earthquakes in Greece with the machine learning algorithm NESTORE*. Entropy, 25, 797, doi: 10.3390/e25050797.

Argus D.F., Gordon R.G. and DeMets C.; 2011: *Geologically current motion of 56 plates relative to the no-net-rotation reference frame*. Geochem. Geophys. Geosyst., 12, Q11001, doi: 10.1029/2011GC003751.

Azzara R., Basili A., Beranzoli L., Chiarabba C., Di Giovambattista R. and Selvaggi G.; 1993: *The seismic sequence of Potenza (May 1990)*. Ann. Geofis., 36, 237-243.

Bailer-Jones C.A.L. and Smith K.; 2011: *Combining probabilities*. Gaia DPAC Data Processing & Analysis Consortium, 14 pp., <www.mpia.de/3432751/probcomb_TN.pdf>.

Bressan G., Bragato P.L. and Venturini C.; 2003: *Stress and strain tensors based on focal mechanisms in the seismotectonic framework of the Friuli-Venezia Giulia region (northeastern Italy)*. Bull. Seismol. Soc. Am., 93, 1280-1297.

Bressan G., Gentile G.F., Perniola B. and Urban S.; 2009: *The 1998 and 2004 Bovec-Krn (Slovenia) seismic sequences: aftershock pattern, focal mechanisms and static stress changes*. Geophys. J. Int., 179, 231-253.

Bressan G., Gentile G.F., Tondi R., de Franco R. and Urban S.; 2012: *Sequential integrated inversion of tomographic images and gravity data: an application to the Friuli area (north-eastern Italy)*. Boll. Geof. Teor. Appl., 53, 191-212.

Bressan G., Ponton M., Rossi G. and Urban S.; 2016: *Spatial organization of seismicity and fracture pattern in NE-Italy and W-Slovenia*. J. Seismol., 20, 511-534.

Bressan G., Barnaba C., Magrin A. and Rossi G.; 2018: *A study on off-fault aftershock pattern at N-Adria microplate*. J. Seismol., 22, 863-881.

Brondi P., Gentili S., Picozzi M., Spallarossa D. and Di Giovambattista R.; 2024a: *Characterizing clusters with strong subsequent events in central Italy using RAMONES*. In: Proc. EGU General Assembly 2024, Vienna, Austria, EGU24-6346, doi: 10.5194/egusphere-egu24-6346.

Brondi P., Snidarcig A., Bernardi P., Bragato P.L. and Di Bartolomeo P.; 2024b: *Bollettino della Rete Sismometrica dell'Italia Nord Orientale (RSINO)*. Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Trieste, Italy, doi: 10.13120/w1vp-b578.

Brondi P., Gentili S. and Di Giovambattista R.; 2025a: *Forecasting strong subsequent events in the Italian territory: a national and regional application for NESTOREv1.0*. Nat. Hazards, 121, 3499-3531, doi: 10.1007/s11069-024-06913-6.

Brondi P., Snidarcig A., Di Bartolomeo P., Magrin A., Barnaba C., Poggi V., Pettenati F., Romanelli M., Rebez A., Moratto L., Sandron D., Plasencia M., Pesaresi D., Compagno A., Del Negro E., Comelli P., Magrin E., Zuliani D., Bertoni M., Fabris P. and Bernardi P.; 2025b: *Bollettino della Rete Sismometrica dell'Italia Nord Orientale (RSINO), Anno 2023*. Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Trieste, Italy, doi: 10.13120/m4pk-nd81

Calderoni G., Rovelli A. and Di Giovambattista R.; 2017: *Rupture directivity of the strongest 2016-2017 central Italy earthquakes*. J. Geophys. Res., 122, 9118-9131.

Caputo R.; 1995: *Inference of a seismic gap from geological data: Thessaly (central Greece) as a case study*. Ann. Geophys., 38, 1-19.

Caputo R. and Pavlides S.; 1993: *Late Cainozoic geodynamic evolution of Thessaly and surroundings (central-northern Greece)*. Tectonophys., 223, 339-362, doi: 10.1016/0040-1951(93)90144-9.

Caravella L. and Gentili S.; 2025: *Forecasting strong subsequent aftershocks in New Zealand: preliminary results*. In: Proc. 21st International Conference Applied Statistic 2025, Koper/Capodistria, Slovenia, p. 40.

Chiarabba C., De Gori P. and Speranza F.; 2008: *The southern Tyrrhenian subduction zone: deep geometry, magmatism and Plio-Pleistocene evolution*. Earth Planet. Sci. Lett., 268, 408-423.

Delibasis N., Ziazia M., Voulgaris N., Papadopoulos T., Stavrakakis G., Papanastassiou D. and Drakatos G.; 1999: *Microseismic activity and seismotectonics of Heraklion area (central Crete Island, Greece)*. Tectonophys., 308, 237-248.

DeMets C., Gordon R.G. and Argus D.F.; 2010: *Geologically current plate motions*. Geophys. J. Int., 181, 1-80, doi: 10.1111/j.1365-246X.2009.04491.x.

Devoti R., Riguzzi F., Cuffaro M. and Doglioni C.; 2008: *New GPS constraints on the kinematics of the Apennines subduction*. Earth Planet. Sci. Lett., 273, 163-174, doi: 10.1016/j.epsl.2008.06.031.

DeVries P.M.R., Viégas F., Wattenberg M. and Meade B.I.; 2018: *Deep learning of aftershock patterns following large earthquakes*. Nature, 560, 632-634, doi: 10.1038/s41586-018-0438-y.

Dimitriadis I., Papazachos C., Panagiotopoulos D., Hatzidimitriou P., Bohnhoff M., Rische M. and Meier T.; 2010: *P and S velocity structures of the Santorini-Coloumbo volcanic system (Aegean Sea, Greece) obtained by non-linear inversion of travel times and its tectonic implications* J. Volcanol. Geotherm. Res., 195, 13-30, doi: 10.1016/j.jvolgeores.2010.05.013.

Doglioni C.; 1991: *A proposal of kinematic modelling for W-dipping subductions - possible applications to the Tyrrhenian-Apennines system*. Terra Nova, 3, 423-434.

Doglioni C.; 1995: *Geological remarks on the relationships between extension and convergent geodynamic settings*. Tectonophys., 252, 253-268.

Doglioni C., Carminati E., Cuffaro M. and Scrocca D.; 2007: *Subduction kinematics and dynamic constraints*. Earth Sci. Rev., 83, 125-175.

Gardner J.K. and Knopoff L.; 1974: *Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?* Bull. Seismol. Soc. Am., 64, 1363-1367, doi: 10.1785/BSSA0640051363.

Gentili S. and Bressan G.; 2008: *The partitioning of radiated energy and the largest aftershock of seismic sequences occurred in the northeastern Italy and western Slovenia*. J. Seismol., 12, 343-354, doi: 10.1007/s10950-007-9075-x.

Gentili S. and Franceschina G.; 2011: *High frequency attenuation of shear waves in the southeastern Alps and northern Dinarides*. Geophys. J. Int., 185, 1393-1416.

Gentili S. and Gentile G.F.; 2015: *High frequency attenuation k parameter and QS 3D model for south-eastern Alps and north-western Dinarides*. Boll. Geof. Teor. Appl., 56, 383-406.

Gentili S. and Di Giovambattista R.; 2017: *Pattern recognition approach to the subsequent event of damaging earthquakes in Italy*. Phys. Earth Planet. Inter., 266, 1-17, doi: 10.1016/j.pepi.2017.02.011.

Gentili S. and Di Giovambattista R.; 2020: *Forecasting strong aftershocks in earthquake clusters from northeastern Italy and western Slovenia*. Phys. Earth Planet. Inter., 303, 106483, doi: 10.1016/j.pepi.2020.106483.

Gentili S. and Di Giovambattista R.; 2022: *Forecasting strong subsequent earthquakes in California clusters by machine learning*. Phys. Earth Planet. Inter., 327, 106879, doi: 10.1016/j.pepi.2022.106879.

Gentili S., Brondi P. and Di Giovambattista R.; 2023a: *NESTOREv1.0: a MATLAB package for strong forthcoming earthquake forecasting*. Seismol. Res. Lett., 94, 2003-2013, doi: 10.1785/0220220327.

Gentili S., Brondi P. and Di Giovambattista R.; 2023b: *Machine learning for forecasting strong aftershocks in California: the NESTOREv1.0 application*. In: Poster, AGU Fall Meeting 2023, San Francisco, CA, USA, Session: Seismology / Statistical Seismology Expanded: Observation, Modeling, Testing, and Forecasting II Poster, 0499.

Gentili S., Brondi P., Rossi G., Sugan M., Petrillo G., Zhuang J. and Campanella S.; 2024: *Seismic clusters and fluids diffusion: a lesson from the 2018 Molise (southern Italy) earthquake sequence*. Earth Planets Space, 76, 157, doi: 10.1186/s40623-024-02096-3.

Gentili S., Chiappetta G.D., Petrillo G., Brondi P. and Zhuang J.; 2025: *Forecasting strong subsequent earthquakes in Japan using an improved version of NESTORE machine learning algorithm*. Geosci. Front., 16, 102016, doi: 10.1016/j.gsf.2025.102016.

Gkarlaouni C., Papadimitriou E., Karakostas V., Kilias A. and Lasocki S.; 2015: *Fault population recognition through microseismicity in Mygdonia region (northern Greece)*. Boll. Geof. Teor. Appl., 56, 367-382, doi: 10.4430/bgta0153.

Godano C., Petrillo G., Tramelli A. and Convertito V.; 2025: *The b-value tomography of the Calabrian Arc*. Earth Space Sci., 12, e2024EA004065, 7 pp.

Hauksson E. and Shearer P.M.; 2005: *Southern California hypocenter relocation with waveform cross-correlation, part 1: results using the double-difference method*. Bull. Seismol. Soc. Am., 95, 896-903, doi: 10.1785/0120040167.

Hutton K., Woessner J. and Hauksson E.; 2010: *Earthquake monitoring in southern California for seventy-seven years (1932-2008)*. Bull. Seismol. Soc. Am., 100, 423-446.

ISIDe Working Group; 2007: *Italian seismological instrumental and parametric database (ISIDe)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV), Roma, Italy, doi: 10.13127/ISIDE.

Japan Meteorological Agency; 2024: *The seismological bulletin of Japan*. <www.data.jma.go.jp/svd/eqev/data/bulletin/index_e.html>.

Kagan Y.Y.; 2002: *Aftershock zone scaling*. Bull. Seismol. Soc. Am., 92, 641–655, doi: 10.1785/0120010172.

Kanamori H.; 1972: *Tectonic implications of the 1944 Tonankai and the 1946 Nankaido earthquakes*. Phys. Earth Planet. Inter., 5, 129-139, doi: 10.1016/0031-9201(72)90082-9.

Karimzadeh S., Matsuoka M., Kuang J. and Ge L.; 2019: *Spatial prediction of aftershocks triggered by a major earthquake: a binary machine learning perspective*. ISPRS Int. J. Geo-Inf. (IJGI), 8, 462, doi: 10.3390/ijgi8100462.

Keilis-Borok V.I. and Kossobokov V.G.; 1990: *Premonitory activation of earthquake flow: algorithm M8.* Phys. Earth Planet. Inter., 61, 73-83.

Keilis-Borok V.I. and Rotwain I.M.; 1990: *Diagnosis of time of increased probability of large earthquakes in different regions of the world: algorithm CN.* Phys. Earth Planet. Inter., 61, 57-72.

Koukouvelas I.K., Kokkalas S. and Xypolias P.; 2010: *Surface deformation during the Mw 6.4 (8 June 2008) Movri Mountain earthquake in the Peloponnese, and its implications for the seismotectonics of western Greece*. Int. Geol. Rev., 52, 249-268.

Lay T.; 2018: *A review of the rupture characteristics of the 2011 Tohoku-oki Mw 9.1 earthquake*. Tectonophys., 733, 4-36, doi: 10.1016/j.tecto.2017.09.022.

Le Pichon X., Chamot-Rooke N., Lallemant S., Noomen R. and Veis G.; 1995: *Geodetic determination of the kinematics of central Greece with respect to Europe: implications for eastern Mediterranean tectonics.* J. Geophys. Res., 100, 12675-12690, doi: 10.1029/95JB00317.

Lippiello E. and Petrillo G.; 2024: *B-more-incomplete and b-more-positive: insights on a robust estimator of magnitude distribution*. J. Geophys. Res.: Solid Earth, 129, e2023JB027849.

Liu B., Wen H., Di M., Huang J., Liao M., Yu J. and Xiang Y.; 2024: *Mapping and interpretability of aftershock hazards using hybrid machine learning algorithms*. J. Rock Mech. Geotech. Eng., 17, 4908-4932, doi: 10.1016/j.jrmge.2024.09.015.

Lolli B. and Gasperini P.; 2003: *Aftershocks hazard in Italy Part I: estimation of time-magnitude distribution model parameters and computation of probabilities of occurrence*. J. Seismol., 7, 235-257.

Lolli B. and Gasperini P.; 2006: *Comparing different models of aftershock rate decay: the role of catalog incompleteness in the first times after mainshock*. Tectonophys., 423, 43-59, doi: 10.1016/j.tecto.2006.03.025.

Malinverno A. and Ryan W.B.F.; 1986: *Extension in the Tyrrhenian Sea and shortening in the Apennines as a result of arc migration driven by sinking of the lithosphere*. Tectonics, 5, 227-245.

McClusky S., Balassanian S., Barka A., Demir C., Ergintav S., Georgiev I., Gurkan O., Hamburger M., Hurst K.J., Kahle H.G., Kastens K.A., Kekelidze G., King R.W., Kotzev V., Lenk O., Mahmoud S., Mishin A., Nadariya M., Ouzounis A., Paradissis D., Peter Y., Prilepin M., Reilinger R., Şanlı I., Seeger H., Tealeb A., Toksöz M.N. and Veis G.; 2000: Global *Positioning System constraints on plate kinematics and dynamics in the eastern Mediterranean and Caucasus.* J. Geophys. Res. Atmos., 105, 5695-5719, doi: 10.1029/1996JB900351.

McKenzie D.; 1978: *Active tectonics of the Alpine-Himalayan belt: the Aegean Sea and surrounding regions*. Geophys. J. R. Astron. Soc., 55, 217-254.

Mignan A. and Broccardo M.; 2020: *Neural network applications in earthquake prediction (1994-2019): meta-analytic and statistical insights on their limitations*. Seismol. Res. Lett., 91, 2330-2342, doi: 10.1785/0220200021.

Milano G. and Di Giovambattista R.; 2011: *Seismicity at the border between central and southern Apennines (Italy): re-evaluation of the early 1984 instrumental earthquake*. Tectonophys., 499, 92-104, doi: 10.1016/j.tecto.2010.12.008.

Miller S.A., Collettini C., Chiaraluce L., Cocco M., Barchi M. and Kaus B.J.P.; 2004: *Aftershocks driven by a high-pressure CO2 source at depth*. Nature, 427, 724-727.

NCEDC; 2014: *Northern California earthquake data center.* UC Berkeley Seismological Laboratory, Dataset, doi: 10.7932/NCEDC.

Ogata Y.; 1988: *Statistical models for earthquake occurrences and residual analysis for point processes*. J. Am. Stat. Assoc., 83, 9-27, doi: 10.2307/2288914.

Ogata Y. and Zhuang J.; 2006: *Space-time ETAS model and an improved extension*. Tectonophys., 413, 13-23, doi: 10.1016/j.tecto.2005.10.016.

Papazachos B., Karakostas V., Papazachos C. and Scordilis E.; 2000: *The geometry of the Wadati-Benioff zone and lithospheric kinematics in the Hellenic arc*. Tectonophys., 319, 275-300, doi: 10.1016/S0040-1951(99)00299-1.

Petrillo G. and Zhuang J.; 2022: *The debate on the earthquake magnitude correlations: a meta-analysis*. Sci. Rep., 12, 20683.

Petrillo G. and Zhuang J.; 2023: *Verifying the magnitude dependence in earthquake occurrence*. Phys. Rev. Lett., 131, 154101.

Picozzi M., Abdi F., Barnaba C., Bertoni M., Bragato P.L., Brondi P., Caravella L., Capotosti G., Cataldi L., Comelli P., Compagno A., Chiappetta G.D., Gentili S., Fabris P., Fangqing D., Magrin A., Moratto L., Pesaresi D., Peruzza L., Poggi V., Romano M.A., Rossi G., Sandron D., Santulin M., Saraò A., Spallarossa D., Sugan M. and Zuccolo E.; 2025: *Rapporto di sintesi sul terremoto di Raveo (Udine) ML 3.9 del 12 Gennaio 2025*. Centro di Ricerche Sismologiche - Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (CRS-OGS) - Udine, Italy, doi: 10.5281/zenodo.15168324.

Polonia A., Torelli L., Artoni A., Carlini M., Faccenna C., Ferranti L., Gasperini L., Govers R., Klaeschen D., Monaco C., Neri G., Nijholt N., Orecchio B. and Wortel R.; 2016: *The Ionian and Alfeo-Etna fault zones: new segments of an evolving plate boundary in the central Mediterranean Sea?* Tectonophys., 675, 69-90, doi: 10.1016/j.tecto.2016.03.016.

Reilinger R., McClusky S., Paradisis D., Ergintav S. and Vernant P.; 2010: *Geodetic constraints on the tectonic evolution of the Aegean region and strain accumulation along the Hellenic subduction zone.* Tectonophys., 488, 22-30, doi: 10.1016/j.tecto.2009.05.027.

Rousseeuw P.J. and Hubert M.; 2011: *Robust statistics for outlier detection*. Wiley Interdiscip. Rev.: Data Mining Knowl. Discov., 1, 73-79, doi: 10.1002/widm.2.

Sakellariou D., Lykousis V., Alexandri S., Kaberi H., Rousakis G., Nomikou P., Georgiou P. and Ballas D.; 2007: *Faulting, seismic-stratigraphic architecture and Late Quaternary evolution of the Gulf of Alkyonides Basin-east Gulf of Corinth, central Greece*. Basin Res., 19, 273-295, doi: 10.1111/j.1365-2117.2007.00322.x.

SCEDC; 2013: *Southern California earthquake center*. Caltech, Dataset, Pasadena, CA, USA, doi: 10.7909/C3WD3xH1.

Schimmenti V.M., Petrillo G., Rosso A. and Landes F.P.; 2024: *Assessing the predictive power of GPS-based ground deformation data for aftershock forecasting*. Seismol. Res. Lett., 95, 3243-3249, doi: 10.1785/0220240008.

Scholz C.H.; 2019: *The mechanics of earthquakes and faulting 3rd ed*. Cambridge Univ. Press, Cambridge, UK, 493 pp., doi: 10.1017/CBO9780511818516.

Smirnov V.B. and Petruchov A.A.; 2025: *On the relationship between RTL and b-value anomalies of seismicity.* Izvestiya Phys. Solid Earth, 61, 539-552, doi: 10.1134/S1069351325700545.

Snidarcig A., Bernardi P., Bragato P.L., Di Bartolomeo P., Garbin M. and Urban S.; 2020: *Bollettino della Rete Sismometrica dell'Italia Nord Orientale (RSINO)*. Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Trieste, Italy, doi: 10.13120/108b8d94-361a-45f3-8195-fc4e8f73d264.

Snidarcig A., Bernardi P., Bragato P.L., Di Bartolomeo P., Garbin M. and Urban S.; 2021: *Bollettino della Rete Sismometrica dell'Italia Nord Orientale (RSINO)*. Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Trieste, Italy, doi: 10.13120/8b252b09-314f-456f-812a-b05268ecd001.

Spassiani I., Gentili S., Console R., Murru M., Taroni M. and Falcone G.; 2025: *Reconciling the irreconcilable: window-based versus stochastic declustering algorithms*. Geophys. J. Int., 240, 1009-1027, doi: 10.1093/gji/ggae425.

Stockman S., Lawson D.J. and Werner M.J.; 2023: *Forecasting the 2016-2017 central Apennines earthquake sequence with a neural point process*. Earth's Future, 11, e2023EF003777, doi: 10.1029/2023EF003777.

Totaro C., Orecchio B., Presti D., Scolaro S. and Neri G.; 2016: *Seismogenic stress field estimation in the Calabrian Arc region (south Italy) from a Bayesian approach*. Geophys. Res. Lett., 43, 8960-8969, doi: 10.1002/2016GL070107.

Trugman D.T. and Ben-Zion Y.; 2023: *Coherent spatial variations in the productivity of earthquake sequences in California and Nevada*. Seism. Rec., 3, 322-331, doi: 10.1785/0320230039.

Uhrhammer R.A.; 1986: *Characteristics of northern and central California seismicity*. Earthquake Notes, 57, 21-37.

U.S. Geological Survey; n.d.: *ANSS Comprehensive Earthquake Catalog (ComCat) documentation.* U.S. Department of the Interior, accessed November 2025, https://usgs.gov/programs/earthquake-hazards/anss-comprehensive-earthquake-catalog-comcat-documentation.

Vallianatos F., Michas G. and Hloupis G.; 2021: *Seismicity patterns prior to the Thessaly (Mw 6.3) strong earthquake on 3 March 2021 in terms of multiresolution wavelets and natural time analysis*. Geosci., 11, 379, doi: 10.3390/geosciences11090379.

Vallianatos F., Karakonstantis A., Michas G., Pavlou K., Kouli M. and Sakkas V.; 2022: *On the patterns and scaling properties of the 2021-2022 Arkalochori earthquake sequence (central Crete, Greece) based on seismological, geophysical and satellite observations*. Appl. Sci., 12, 7716, doi: 10.3390/app12157716.

van der Elst N.J. and Page M.T.; 2023: *A-positive: a robust estimator of the earthquake rate in incomplete or saturated catalogs*. J. Geophys. Res.: Solid Earth, 128, e2023JB027089, doi: 10.1029/2023JB027089.

Vorobieva I.A.; 1999: *Prediction of a subsequent large earthquake*. Phys. Earth Planet. Inter., 111, 197-206.

Vorobieva I.A. and Panza G.F.; 1993: *Prediction of the occurrence of related strong earthquakes in Italy.* Pageoph, 141, 25-41, doi: 10.1007/BF00876232.

Wang M., Shen J., Pan Z.A. and Han D.L.; 2019: *An improved supported vector regression algorithm with application to predict aftershocks*. J. Seismol., 23, 983-993.

Wu J., Suppe J., Lu R. and Kanda R.; 2016: *Philippine Sea and east Asian plate tectonics since 52 Ma constrained by new subducted slab reconstruction methods*. J. Geophys. Res.: Solid Earth, 121, 4670-4741, doi: 10.1002/2016JB012923.

Zechar J.D.; 2010: *Evaluating earthquake predictions and earthquake forecasts: a guide for students and new researchers*. Community Online Resour. Stat. Seismicity Anal., 126, doi: 10.5078/corssa-77337879, <www.corssa.org>.

Zhao S., Wang H., Xue Y., Wang Y., Li S., Liu J. and Pan G.; 2022: *What are more important for aftershock spatial distribution prediction, features, or models? A case study in China*. J. Seismol., 26, 181-196, doi: 10.1007/s10950-021-10044-x.

Zhuang J., Ogata Y. and Vere-Jones D.; 2002: *Stochastic declustering of space-time earthquake occurrences*. J. Am. Stat. Assoc., 97, 369-380, doi: 10.1198/016214502760046925.

Corresponding author: Stefania Gentili
National Institute of Oceanography and Applied Geophysics – OGS
Via Treviso 55, 33100 Udine, Italy
Phone: +39 040 2140234; e-mail: sgentili@ogs.it

Appendix: Online NRT software and web page on OGS data

The OGS automatic seismic monitoring system managed by its department Seismological Research Centre (CRS) performs several real-time procedures as part of its 24/7 monitoring activity, which includes localising events, estimating their magnitude, and forwarding warnings to civil protection authorities. We integrated the NESTORE NRT module among these processes. To support this functionality, a dedicated server has been set up to continuously monitor new events that exceed the magnitude threshold and execute the NRT module. This module, like the rest of the NESTORE codebase, is written in MATLAB. For our purpose, it is wrapped in a shell script that handles the interaction with the Linux operating system. When an event is detected, the server first evaluates whether the event fulfils conditions 1 and 2 for the NESTORE NRT analysis. If these criteria are satisfied, the system automatically prepares the required environment (generating input files and directory structures) and launches the NRT analysis in a temporary workspace (see Fig. A1). NESTORE stores, as potential aftershocks, all events that occur within the six hours

following the mainshock and within a radius that depends on *Mm* according to the window-based law defined in the training and testing (see Section 3.2). If the magnitude of the strongest aftershock is greater than, or equal to, M_m−1, there are two possibilities:

1. if the magnitude is greater than M_m, the aftershock becomes the new operative mainshock and the procedure restarts;
2. if the magnitude is smaller than M_m, the cluster is already classified as type A and no further classification is performed.

This minor change to the original algorithm (the branch in point 1 was not present in the original algorithm) was added taking into account the low threshold (M_L = 3.7) for the o-mainshock in the region, to allow analysis even in cases where the mainshock is preceded by a small (e.g. M_L = 3.7 or similar) foreshock a few hours earlier.

If during the first six hours after the mainshock the aftershocks have magnitudes smaller than M_m−1, the seismic features are analysed and the system calculates the probability that the cluster is a type A based on prior machine learning training. The output includes the forecasted cluster class, its probability, and the forecasting spatiotemporal window.

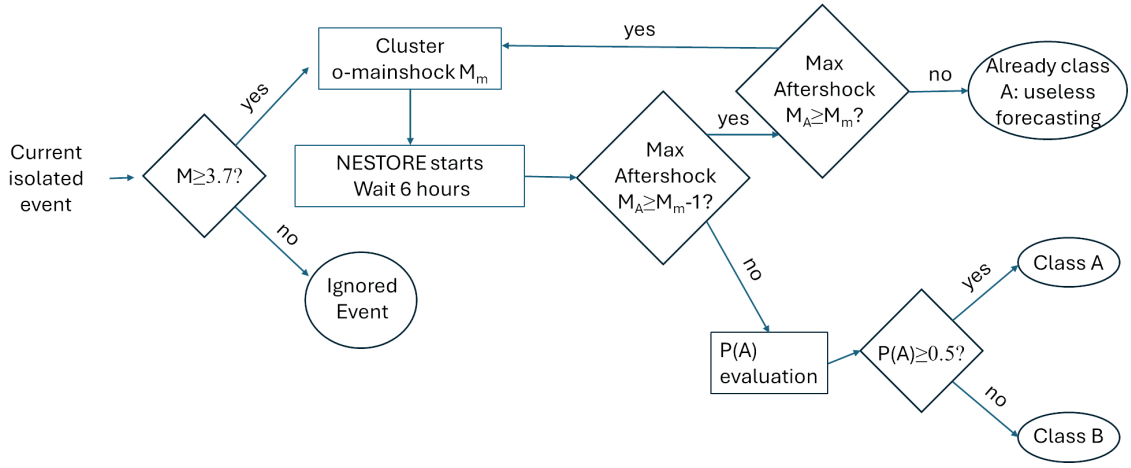


Fig. A1 - Block diagram of the NRT algorithm.

Since the best results of the test phase in the region were obtained for a period of six hours after the main o-mainshock, according to the NRT procedure, the NRT performance should only be considered for six hours. In order to consider the possible future trainings on larger database, which may show better results for longer time intervals, the procedure is performed for all the intervals with training, i.e. 6, 12, and 18 hours after the o-mainshock. The o-mainshock is kept unchanged for longer time intervals, to avoid confusion with the six-hour analysis.

Since alerts are typically received by the NESTORE server within a few minutes after the event origin time, the automated pipeline is configured to wait for the minimum time window (six hours) before querying the CRS server for the updated aftershock catalogue (if available). This approach ensures that the analysis is based on the most up-to-date information (e.g. location, magnitude), which may have been refined by an operator in the meantime, and allows the system to deliver an informed classification (type A or B). The final output is published at https://nestore.crs.ogs.it. The site is under development and is password protected. At present, the algorithm has correctly processed two sequences in 2024 (see Fig. 5). On 12 January 2025 an M_L

3.9 event occurred close to the village of Raveo (Picozzi *et al*., 2025), followed by an event of magnitude M_L = 3.3 in the first six hours after the mainshock. According to the procedure shown in the block diagram in Fig. A1, the cluster already resulted to be a type A cluster in the first six hours, as the second event was not strong enough to become a new operative mainshock. The cluster did not require any further forecasting and was therefore ignored.