Unsupervised probabilistic machine learning applied to seismicity declustering: a new approach to represent earthquake catalogues with fewer assumptions

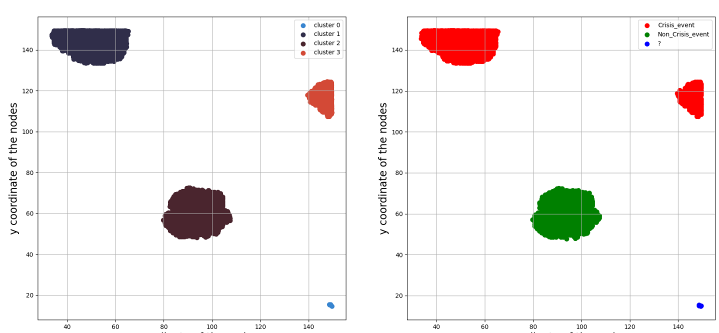
Antoine Septier[1], Alexandra Renouard[2], Jacques Déverchère[3], and Julie Perrot[4]

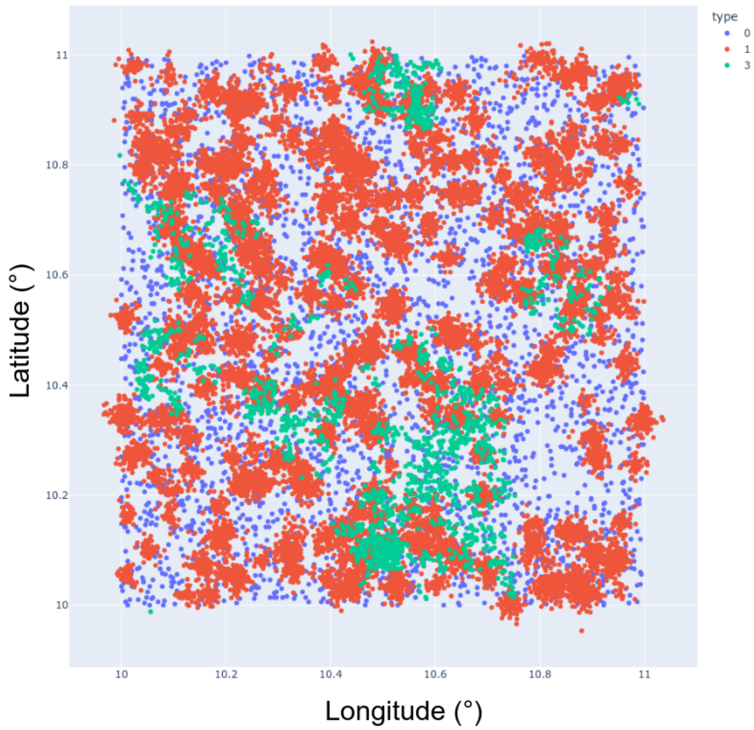
[1]Université de Brest
[2]Imperial College
[3]Université de Bretagne Occidentale
[4]IUEM-CNRS

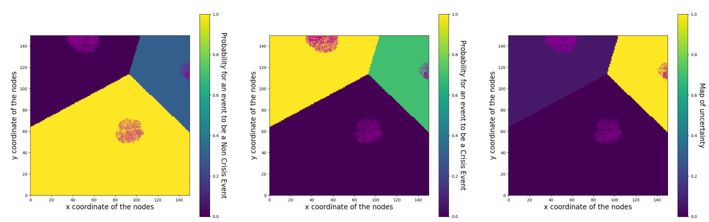
April 16, 2023

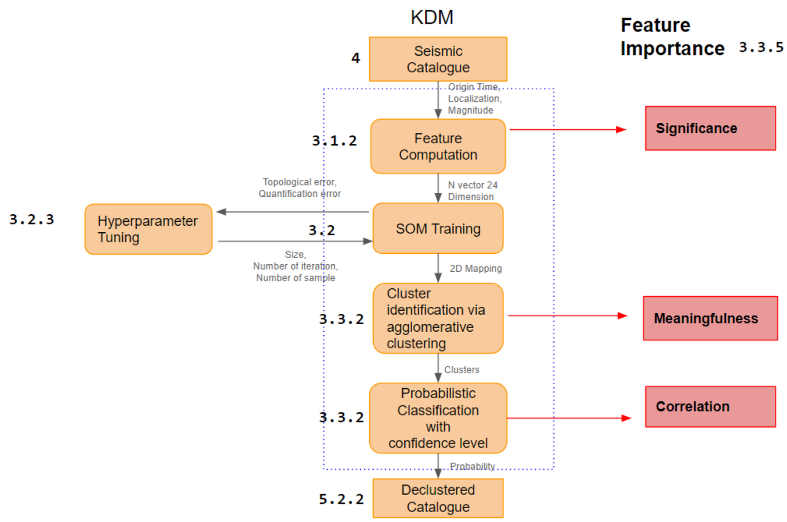
Abstract

Many applications in seismology require to isolate earthquake clusters from a background activity. Relative declustering methods essentially find a 2D representation of an earthquake catalogue that distinguishes between two classes of events: crisis and non-crisis events. However, the number of statistical and/or physical parameters to be used is often limited due to the difficulty of concatenating the information onto a physically meaningful 2D grid. In this study, we propose to alleviate the declustering task by using the ability of unsupervised artificial intelligence to model complex spatio-temporal relationships directly from data. Through a data-driven approach, we define an easily transferable declustering model that provides declustering results with fewer assumptions and no prior selection of thresholds. We first obtain this model by training a self-organising neural network (SOM) that learns to cluster data points according to their feature similarity on a 2D map. We then assign each SOM cluster a label (crisis or non-crisis class) using an agglomerative clustering procedure. We quantify the classification uncertainty by developing a probabilistic function based on the projection learned by SOM. Our method is applied to a synthetic dataset and to real catalogues from the Gulf of Corinth, Central Italy and Taiwan. We discuss the validity of the method by estimating its classification accuracy. For real data, we qualitatively compare our results to previous declustering attempts. We show that our approach is easy to handle, provides a fairly new representation of earthquake catalogues and has the potential to reduce classification ambiguities between nearby events.

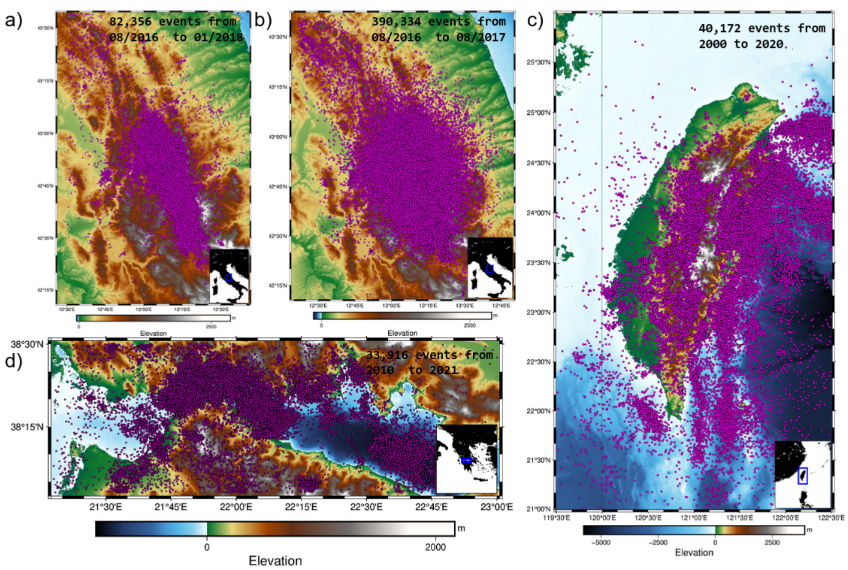



type
0
1
3
Latitude (°)
Longitude (°)
11
10.8
10.6
10.4
10.2
10

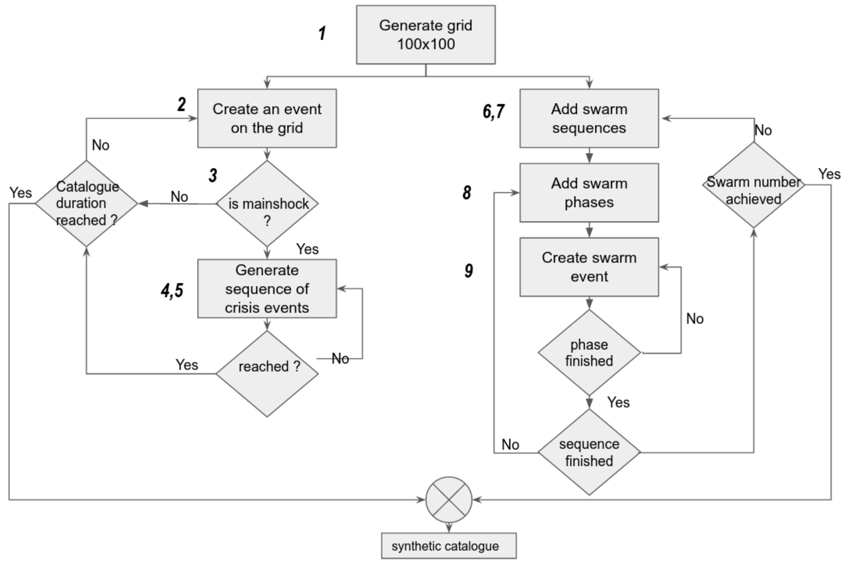

y coordinate of the nodes
x coordinate of the nodes
Probability for an event to be a Non Crisis Event
Probability for an event to be a Crisis Event
Map of uncertainty



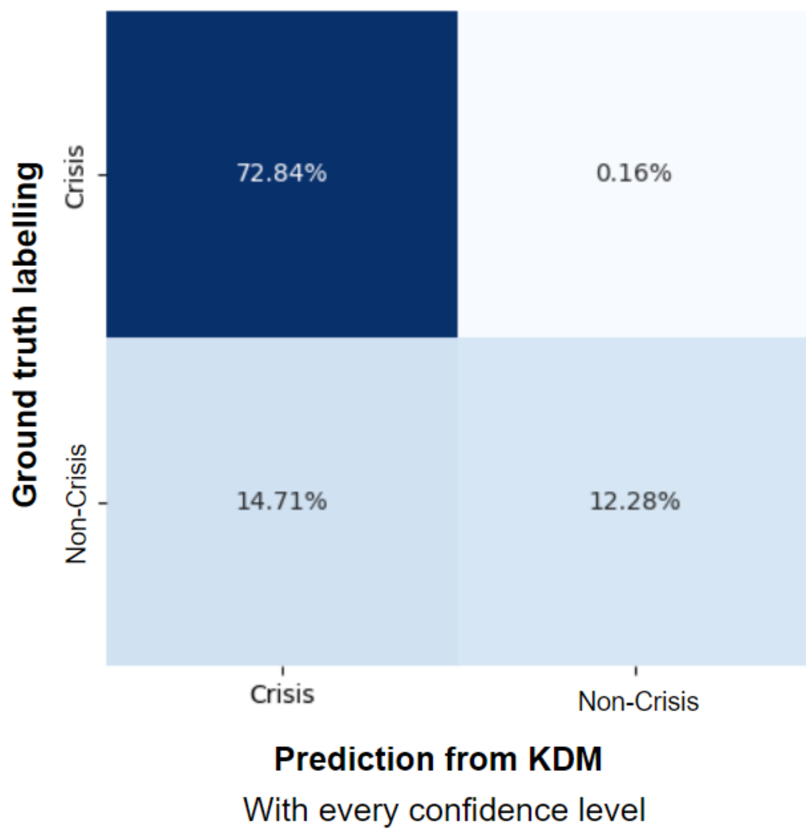
KDM
Feature Importance 3.3.5
4
Seismic Catalogue
Origin Time, Localization, Magnitude
3.1.2
Feature Computation
Significance
N vector 24 Dimension
Topological error, Quantification error
3.2.3
Hyperparameter Tuning
3.2
SOM Training
Size, Number of iteration, Number of sample
2D Mapping
3.3.2
Cluster identification via agglomerative clustering
Meaningfulness
Clusters
3.3.2
Probabilistic Classification with confidence level
Correlation
Probability
5.2.2
Declustered Catalogue



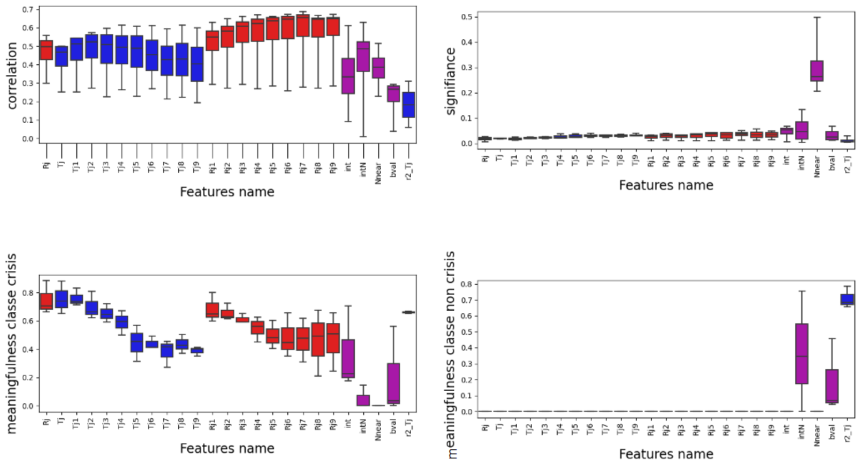
a)
82,356 events from 08/2016 to 01/2018
b)
390,334 events from 08/2016 to 08/2017
c)
40,172 events from 2000 to 2020.
d)
33,916 events from 2010 to 2021
38°30'N
38°15'N
21°30'E
21°45'E
22°00'E
22°15'E
22°30'E
22°45'E
23°00'E
Elevation

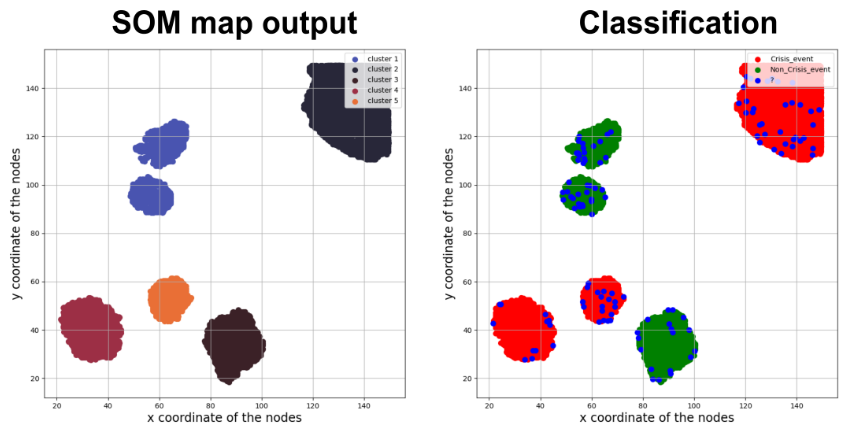


1
Generate grid 100x100
2
Create an event on the grid
3
is mainshock ?
Catalogue duration reached ?
No
Yes
No
Yes
4,5
Generate sequence of crisis events
reached ?
Yes
No
6,7
Add swarm sequences
8
Add swarm phases
9
Create swarm event
phase finished
No
Yes
sequence finished
No
Swarm number achieved
No
Yes
synthetic catalogue

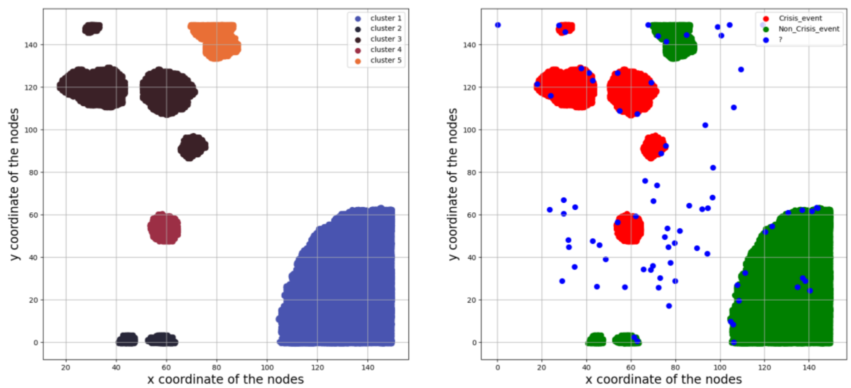


Ground truth labelling
Crisis
Non-Crisis
72.84%
0.16%
14.71%
12.28%
Crisis
Non-Crisis
Prediction from KDM
With every confidence level

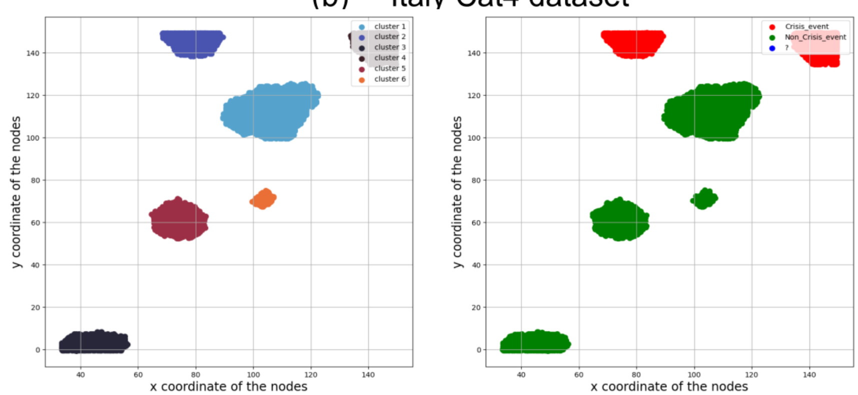

correlation
Features name
signifiance
Features name
meaningfulness classe crisis
Features name
meaningfulness classe non crisis
Features name
Rj
Tj
Tj1
Tj2
Tj3
Tj4
Tj5
Tj6
Tj7
Tj8
Tj9
Rj1
Rj2
Rj3
Rj4
Rj5
Rj6
Rj7
Rj8
Rj9
int
intN
Nnear
bval
r2_Tj



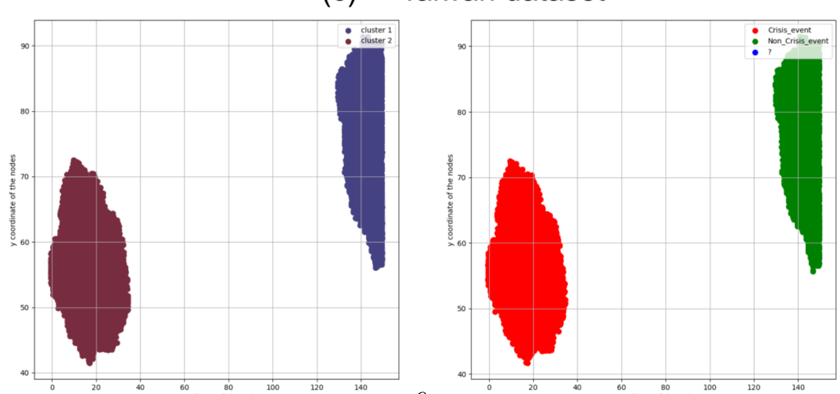
(a) Italy Cat1 dataset



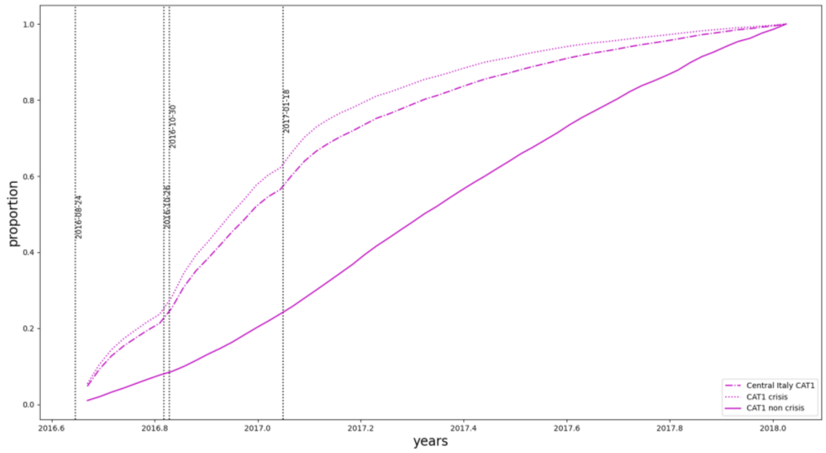
(b) Italy Cat4 dataset



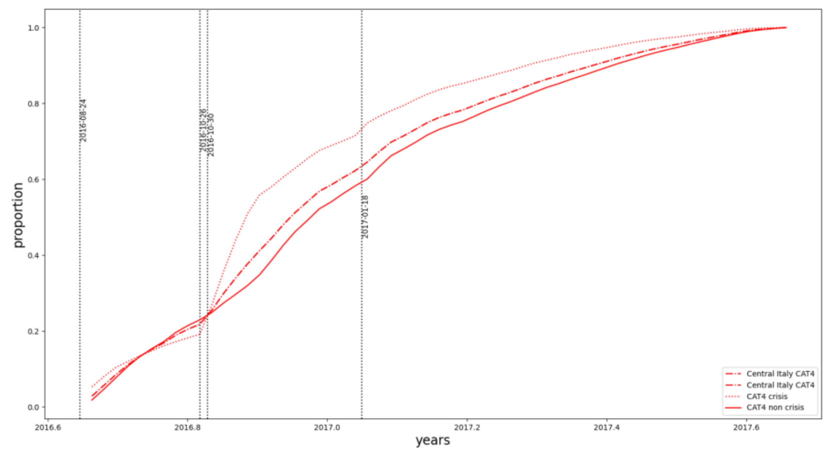
(c) Taiwan dataset



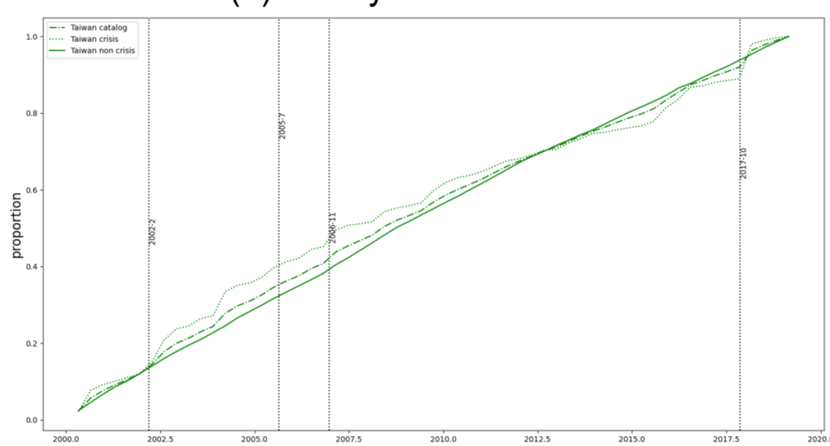
(d) GOC dataset



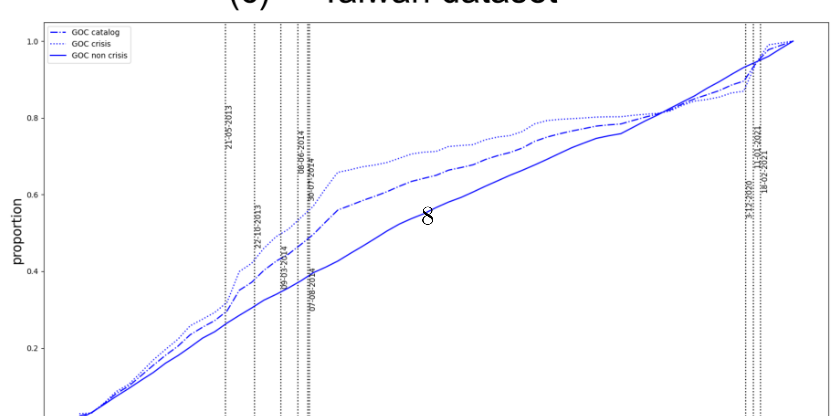
(a) Italy Cat1 dataset

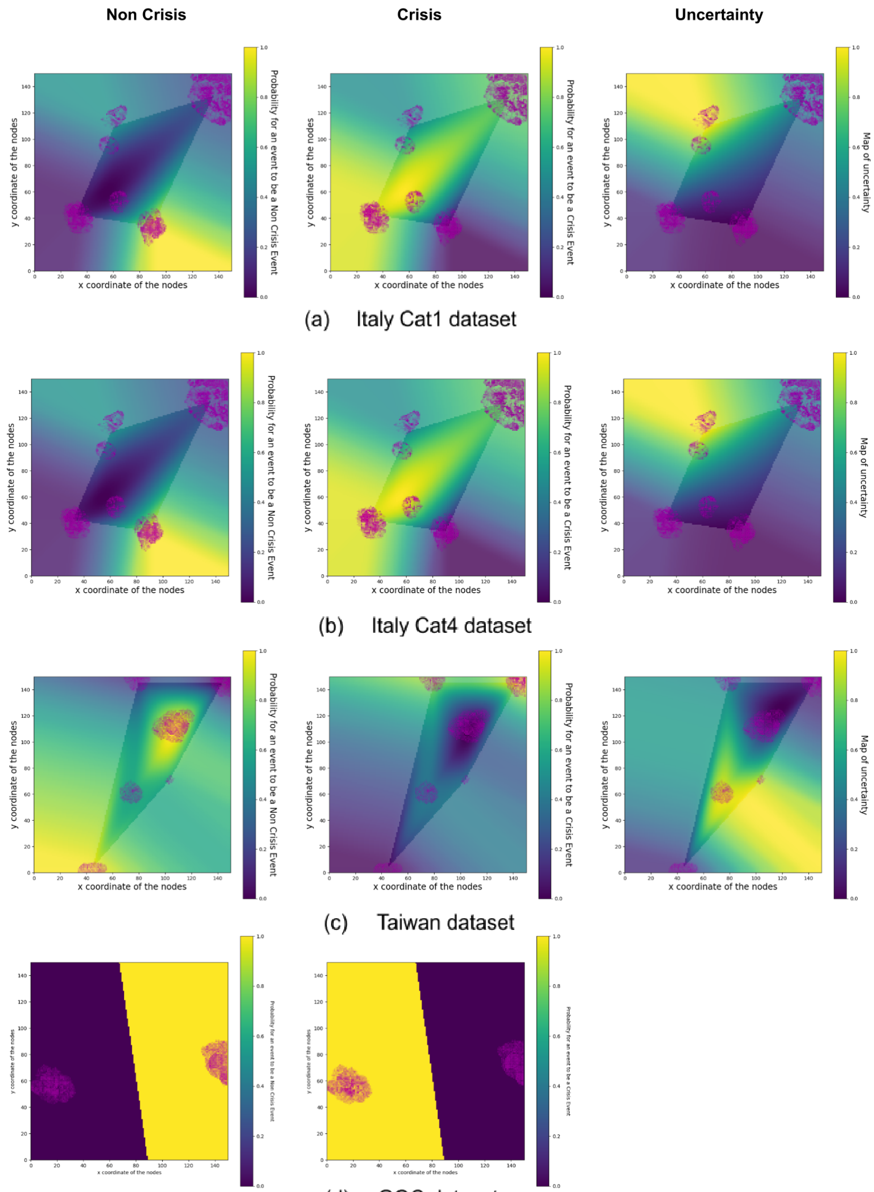


(b) Italy Cat4 dataset



(c) Taiwan dataset





(a) Italy Cat1 dataset

(b) Italy Cat4 dataset

(c) Taiwan dataset

(d) GOC dataset

Unsupervised probabilistic machine learning applied to seismicity declustering: a new approach to represent earthquake catalogues with fewer assumptions

Antoine Septier[1], Alexandra Renouard[2], Jacques Déverchère[1], Julie Perrot[1]

[1]Geo-Ocean, Univ Brest, CNRS, Ifremer, UMR6538, Plouzané, France
[2]Department of Civil & Environmental Engineering, Imperial College London, UK

Key Points:

- We present a flexible declustering methodology that handles large input parameter spaces with fewer assumptions and no threshold effects
- We probabilistically interpret the data mapping modelled by the self-organising neural network to classify non-crisis and crisis events
- The method is 85% accurate on synthetic data and is used to examine previous attempts to decluster data from Greece, Italy and Taiwan

Corresponding author: A. Septier, antoine.septier@univ-brest.fr

Corresponding author: A. Renouard, alexandra.renouard@gmail.com

Corresponding author: J. Déverchère, Jacques.Deverchere@univ-brest.fr

Corresponding author: J. Perrot, jperrot@univ-brest.fr

Abstract

Many applications in seismology require to isolate earthquake clusters from a background activity. Relative declustering methods essentially find a 2D representation of an earthquake catalogue that distinguishes between two classes of events: crisis and non-crisis events. However, the number of statistical and/or physical parameters to be used is often limited due to the difficulty of concatenating the information onto a physically meaningful 2D grid. In this study, we propose to alleviate the declustering task by using the ability of unsupervised artificial intelligence to model complex spatio-temporal relationships directly from data. Through a data-driven approach, we define an easily transferable declustering model that provides declustering results with fewer assumptions and no prior selection of thresholds. We first obtain this model by training a self-organising neural network (SOM) that learns to cluster data points according to their feature similarity on a 2D map. We then assign each SOM cluster a label (crisis or non-crisis class) using an agglomerative clustering procedure. We quantify the classification uncertainty by developing a probabilistic function based on the projection learned by SOM. Our method is applied to a synthetic dataset and to real catalogues from the Gulf of Corinth, Central Italy and Taiwan. We discuss the validity of the method by estimating its classification accuracy. For real data, we qualitatively compare our results to previous declustering attempts. We show that our approach is easy to handle, provides a fairly new representation of earthquake catalogues and has the potential to reduce classification ambiguities between nearby events.

Plain Language Summary

One of the main approaches to removing some of the biases from earthquake catalogues and facilitating the decoding of the information they contain is to decluster them. There are many declustering methods in the literature, each producing significant differences in the resulting declustered catalogues. The reason why there are so many methods is that each of them takes into account new or additional statistical and/or physical features that may better describe the behaviour of earthquakes in the specific seismotectonic context for which they are applied.

In this study, we propose a flexible relative declustering methodology capable of handling all desired seismic features while reducing subjective assumptions and threshold effects. This declustering procedure is based on an unsupervised machine learning approach that uses an artificial neural network called a self-organising map (SOM). Through a clustering process, the SOM neural network is able to non-linearly map large input spaces onto a 2D grid, which hopefully preserves the topological and metric relationships of the data. Thanks to this reduction in dimensionality, high-dimensional datasets of seismic features can be easily visualised and interpreted in a 2D representation, as shown here with synthetic data and real earthquakes catalogues from Greece, central Italy and Taiwan.

1 Introduction

Earthquake catalogues are key datasets widely used by the scientific community for understanding the statistical behaviour of earthquakes, their spatio-temporal evolution and their triggering factors. They can also highlight the 3D geometry of seismically active structures, contribute to the quantification of seismic hazard and improve earthquake forecasting (Zhu et al., 2023). In addition, new generations of high-definition seismic catalogues are being built with more powerful detection procedures. Unprecedented levels of details can then be achieved to reveal finer spatio-temporal seismic patterns that were previously undetectable (Beroza et al., 2021; Herrmann et al., 2022; Mancini et al., 2022).

However, the exploration of all these earthquake catalogues remains actually difficult to operate due to their high dimensionality and intrinsic heterogeneity (e.g. spatio-temporal evolution of seismological networks, changes in recording and/or processing procedures). The representation of fundamental earthquake properties through these datasets is therefore challenging and affected by many biases (Weatherill et al., 2016).

One of the main approaches to remove some of these biases and to facilitate the decoding of information from earthquake catalogues is to decluster them (Zaliapin & Ben-Zion, 2022). Seismicity declustering is indeed commonly used in seismological analyses to extract recurrent seismic features and to solve complex problems such as estimating the evolution of seismic locations prior to large earthquakes (Zaliapin & Ben-Zion, 2022) or relating earthquake depth distributions to the mechanical strength properties of the crust (Scholz, 2002; Albaric et al., 2009; Cheng & Ben-Zion, 2019).

Declustering methods usually provide distinct sub-catalogues containing two categories of seismic events: "independent" events, which are related to long-term deformation processes and referred as background seismicity, and "dependent", transient events (swarms, foreshock or aftershock sequences), which are wholly or partly triggered by previous events and exhibit clustered spatio-temporal behaviours (Pisarenko & Rodkin, 2019). However, there are many different methods of declustering, each creating dissimilarities in their resulting declustered catalogues (van Stiphout et al., 2012), (Gurjar & Basu, 2022). We may cite for instance those based on the Epidemic Type Aftershock Sequence (ETAS) model (Iacoletti et al., 2022; Zhang & Huang, 2022; Mizrahi et al., 2022; Field et al., 2021, 2022; Hainzl, 2022), on nearest-neighbour distances (Zaliapin et al., 2008; Zhuang et al., 2002) or on supervised machine learning (Aden-Antoniow et al., 2022; Pavez O & Estay H, 2021; Seydoux et al., 2020). The reason why there are so many methods is that each of them takes into account new or additional statistical and/or physical features that are assumed to better describe the behaviour of earthquakes in the specific seismotectonic context for which they are applied (Zaliapin & Ben-Zion, 2021).

A more homogeneous and less subjective approach is therefore needed for more in-depth analyses of earthquake clustering with complex and heterogeneous catalogues. Among the available declustering methods, relative declustering, as opposed to declustering based on stochastic models such as the ETAS model (Ogata, 1988, 1998, 2004; Zhuang et al., 2004), creates a two-dimensional (2D) representation of the dataset, assuming the existence of two classes in a catalogue: dependent and independent events. To obtain a human-interpretable 2D space of a two-event class seismic catalogue, these relative methods must perform a physically meaningful concatenation of all the seismic features used, which limits the number of seismic features to be taken into account.

In this study, we propose a more flexible relative declustering methodology that is able to handle all desired seismic features while reducing the number of subjective assumptions and threshold effects. This declustering procedure is based on an unsupervised machine learning approach that uses an artificial neural network called a self-organising map (SOM). A SOM neural network is indeed capable of non-linearly mapping large input spaces onto a 2D grid through a clustering process, which hopefully preserves the topological and metric relationships of the data. Through this reduction in dimensionality, high-dimensional datasets of seismic features can easily be visualised and interpreted in a 2D representation.

We therefore first train a SOM neural network to produce a data representation with as many seismic feature inputs as desired. We then use hierarchical agglomerative clustering to identify clusters in the 2D SOM grid. We finally classify them as containing background events, aftershocks or swarms, using a probabilistic approach based on the seismic features we select to train the SOM network (inter-event space-time distances and b-value, average magnitude, density of events). To estimate the classification uncertainty and confidence level of our declustering approach, we develop a probabilistic function based on the projection learned by the SOM. To evaluate the reliability and potential of our machine learning approach, we apply our SOM declustering method to several datasets: first, a synthetic seismic dataset and second, real earthquake catalogues from the Corinth Rift (RESIF, 1995; Evangelidis et al., 2021), Central Italy (Chiaraluce et al., 2022) and Taiwan (Peng et al., 2021). The real data were selected to represent a wide range of criteria such as the size of the study area, the tectonic regime, the degree of magnitude completeness, the duration and the detection and location procedures used. The consistent declustering results obtained with these datasets show that our machine learning-based declustering approach has a strong generalisation capability, even when using only information contained in standard catalogues.

2 Towards a Spatio-Temporal Declustering of Complex and Heterogeneous Catalogues using Self-Organising Maps

The two categories of events we seek to identify through the declustering process are the so-called crisis and non-crisis events. We define a crisis event as an event that is directly triggered by another event (e.g. aftershocks and swarms) and a non-crisis event as an event that is seemingly uncorrelated to the neighbouring seismic activity (e.g. background events).

2.1 First Approach: Spatial Representation of Seismic Events

The first and simplest way to represent a seismicity catalogue is through a 2D geographical map (longitude and latitude). This representation allows a quick visual identification of areas with a denser number of seismic events as well as earthquake propagation patterns in the same direction or around a same location. A first declustering approach could be carried out on the basis of this information. However, it would not take into account the temporal dimension, which would result in the loss of many background events in the process.

2.2 Second Approach: Spatio-Temporal Representation of Seismic Events Using Cumulative Curves

The third temporal dimension is needed to improve the declustering of seismic catalogues, as they reflect the natural occurrence of multiple distinct seismic sequences over time. For example, the cumulative number of seismic events over time can provide additional information about event productivity. In combination with 2D maps, space-time windows can be created around a seismic crisis, reducing loss and missing classification on the background event class during declustering.

Single-Link cluster analysis is an example of declustering approach (Frohlich & Davis, 1990) that exploits the spatio-temporal information included in the seismic catalogues by calculating temporal ($\Delta t_{i,j}$) and spatial ($r_{i,j}$) distances:

$$d_{i,j} = \sqrt{r_{i,j} + Cst^2(\Delta t_{i,j})^2} \tag{1}$$

Equation (1) is applied for each event i and j of a given catalogue to obtain an inter-event distance metric. An empirical distance threshold (see Equation 1), applied on the smallest values of $d_{i,j}$ found for each event i in the catalogue, is used to obtain the final declustering. However, the use of a single threshold is a limitation, as there is no guarantee that the best threshold is the same over time or space. Furthermore, this space-time declustering approach requires a uniformly sampled catalogue in order to maintain the same relative distance distribution in each window. However, in reality, the density of detected events in time and space can change as a function of inter-station and epicentral distances, especially when operational changes are made (e.g. detection system or instrumentation).

2.3 Third Approach: Time-Magnitude Representation of Seismic Events

Magnitude information from seismic catalogues can also be used to more accurately identify the onset of mainshock-aftershock sequences. According to Omori's law (Utsu et al., 1995), these sequences theoretically all start with a large mainshock, followed by aftershocks whose number and magnitude decrease with time. Plotting the magnitude distribution of events as a function of their origin time can be useful to distinguish aftershocks from mainshocks. However, we must assume that the spatial windowing chosen is good and includes all necessary information.

To find the tail of aftershock sequences, it is possible to weight the probability of an event being in the tail of a sequence by a factor depending on the magnitude of the event (m_i) and the b-value of the given sequence, assuming a pure Omori's law (Utsu et al., 1995):

$$factor = 10^{-b*m_i} \tag{2}$$

2.4 Fourth Approach: Nearest-Neighbour Approach in a Space-Time-Magnitude domain

Combining the above-mentioned features, some authors (e.g. Zaliapin and Ben-Zion (2021)) have developed a declustering method using relative spatial and temporal distances, weighted by a b-value and a magnitude factor function that describes the distribution of events relative to their first neighbours:

$$T_j = \Delta t_{i,j} * 10^{-b*mi/2} \tag{3}$$

$$R_j = r_{i,j}^{Cst} * 10^{-b*mi/2} \tag{4}$$

$$\eta_j = R_j * T_j \tag{5}$$

The graphical representation of all the temporal T_j and spatial R_j distances shows two distinct lobes: the first lobe, described by the smallest average values of T_j and R_j, corresponds to crisis events and the other lobe to non-crisis events. This approach is very robust when relatively homogeneous spatial and temporal calculation windows are selected. However, it becomes more unstable as the diversity of crises in a catalogue increases: a b-value must be calculated for each crisis to obtain a rigorous result, each crisis having a specific b-value (Mesimeri et al., 2019). In addition, the systematic search for an optimal space-time window is necessary to correctly differentiate the two lobes.

2.5 Fifth Approach: Rupture Process of Swarms and Aftershocks

The methods described above correctly identify the sequences of mainshocks and aftershocks observed in the catalogues, but have difficulties in identifying swarms which have a different distribution in space, time and magnitude. Moreover, swarm activity can occur in the vicinity of mainshock-aftershock sequences, making difficult to distinguish between all these sequences. We therefore need to add dimensions that can provide new information on the physics of nucleation, such as a b-value, a finer temporal or spatial distribution with neighbouring events (i.e. more than one nearest neighbour) or other criteria such as waveform similarity (Barani et al., 2007; Seydoux et al., 2020).

A combination of several seismic features is therefore needed to efficiently solve the declustering problem: relative space-time distances, magnitude values, magnitude distribution (e.g. b-value), or features linked to the physics of nucleation.

However, the difficulty of correctly extracting information from a seismic catalogue increases with the number of seismic features to be concatenated into a human-interpretable

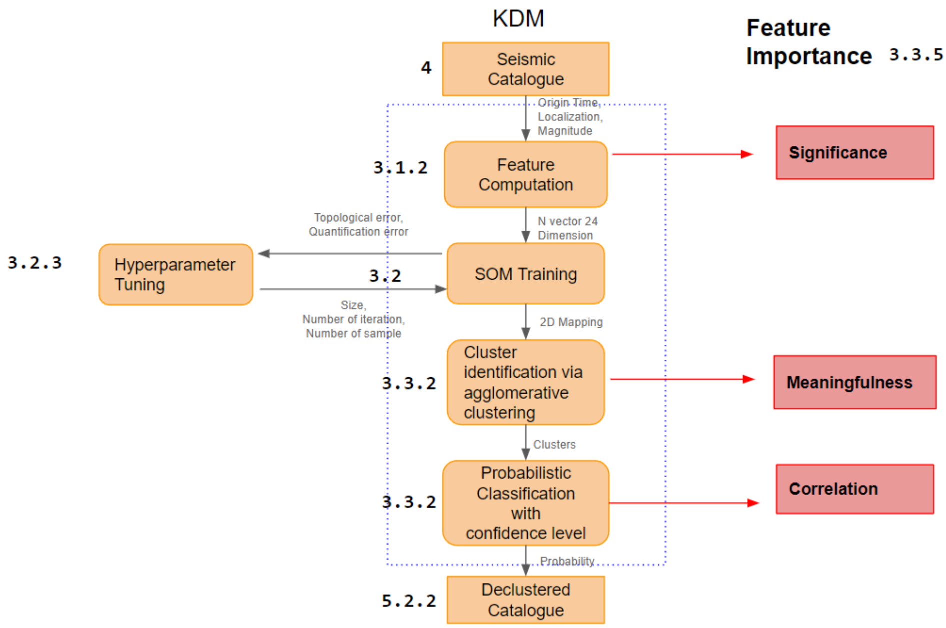


Figure 1: Architecture of our declustering methodology summarised in different steps. The numbers in bold refer to the different sections of the article. The acronym KDM is the name given to the whole workflow of our method: KDM stands for Kohonen Map Declustering Method.

2D space. The axes of the resulting 2D representation must be physically meaningful to allow a more objective assignment of the correct class of events to each cluster represented in 2D space.

3 Declustering Methodology Based on Self-Organised Maps and Agglomerative Clustering

In this section, we present step by step the machine learning methodology we use to solve the declustering problem. This methodology is based on a set of higher dimensional seismic features in order to obtain a robust and interpretable 2D representation of a seismic catalogue. The different steps are summarised in Figure 1. To achieve this goal, we perform a SOM dimensionality reduction, followed by an agglomerative clustering performed on the SOM generated map.

3.1 Learning Architecture of Self-Organising Maps

3.1.1 Definition of Self-Organising Map

SOM (Vesanto, J. & Alhoniemi, E., 2000) is an unsupervised neural network-based dimensionality reduction algorithm used to represent a high-dimensional dataset as a low-dimensional (usually 2D) discretised pattern. The dimensionality reduction is performed while maintaining the topological structure of the input data. The neural network is trained by competitive learning, as opposed to error-correction learning (e.g. back-propagation with gradient descent). After dimensionality reduction by SOM, each dataset used, represented by vectors of p features measured in n observations, is visualised on a 2D SOM map by clusters of observations. Observations in the proximal clusters have more similar feature values than observations in the distal clusters.

The SOM neural network is based on a purely mathematical process that aims to find a new topological space to represent the hidden distribution of input features. This process is comparable to the Principal Component Analysis (PCA), which is often used to analyse datasets with a large number of dimensions. As with SOM, PCA reduces the output space of the dataset while retaining as much of its properties as possible to provide the best representation of each class in the dataset. However, PCA only linearly projects the dataset onto the best principal component, while SOM creates a complete new topological space. Unlike PCA, SOM is an injection (multiple inputs give the same output) that projects the input vectors of the dataset into a new space that uses each component of the input space.

3.1.2 Self-Organising Map Learning Process

The learning process of the SOM is a repetition of a few steps :

1. The SOM algorithm models an input space with a fixed grid of nodes.
2. Each node in the grid has the same dimensions (i.e. the same values) as the input vectors. Random scalars are assigned to nodes in the input vector value range.
3. For each input vector, the algorithm searches for the Best Matching Unit (BMU), which is equivalent to finding the smallest Euclidean distance between the input vector and the nodes.
4. The BMU and its neighbouring nodes within a certain radius are modified, so that the nodes values are slightly adjusted to reduce the Euclidean distance to the input vector.
5. The last two steps are repeated in the learning process: with each new iteration, the radius and the maximum allowed change in node values decrease.

By running through all the input vectors in the dataset, the entire grid of nodes ends up reaching the shortest distance between the nodes and the dataset, with similar nodes (i.e. inputs to the dataset) being grouped together in one area, and dissimilar nodes being separated. The dataset can then be visualised on a 2D map where each input vector is assigned to its best matching nodes.

3.2 Self-Organising Map Training Process

We aim to explore the dataset by calculating the relative distances of each data point to its neighbours as multiple features. The first step is to define the distance scoring function between two input feature vectors.

3.2.1 Neighbourhood Function

To find the nearest neighbours j of each seismic event i in the catalogue, we develop the following neighbourhood function:

$$D_{i,j} = \sqrt{\left(\frac{haversine(event_i, event_j)}{D}\right)^2 + \left(\frac{\Delta T(event_i, event_j)}{T}\right)^2} \tag{6}$$

In Equation 6, T and D are constants that define the third quartile of all temporal (T) and spatial distances (D) between events in the catalogue. We design these constants to make the temporal and spatial dimensions comparable: each inter-event distance is normalised by the third quartile, so that the resulting statistical parameters are less dependent on catalogue size and length.

3.2.2 Feature Input Vectors

We note here that it is possible to add as many coordinates as possible to each input vector (i.e. the feature values that define the problem to be solved), allowing customisation of features adapted to each study.

For our study, we use 25 features. The first 20 features are the spatial R_jn and temporal T_jn distances between an event j and its n nearest neighbours (n = 10). The following four features are calculated over a sliding window centred on the event, whose length is proportional to the duration T and the distance D (T and D are constants defined in section 3.2.1). In a window of duration T and distance D centred on the event j, we calculate the number of events and the average magnitude. We normalise the magnitude and the number of events obtained by an average of the same quantity measured over a larger window of duration $2T$ and distance $2D$. We thus obtain a magnitude ratio using an approach equivalent to the calculation of the signal amplitude ratio between a Short-Term Average window (STA) and a Long-Term Average window (LTA) that is used to classically detect seismic events (Trnkoczy, 2009). We also use the average magnitude value without normalisation. We finally calculate the b-value over a window of $10T$ and distance of $10D$.

The last feature is the coefficient of determination R^2 of the ten closest temporal distances (in ascending order) with a linear extrapolation, to check whether they follow an increasing law or not. This feature measures the linear relationship between the spatial and temporal distances of each event j from its neighbours i. If event j is independent of its neighbours i, the possibility of such a linear space-time relationship is less likely. In the case of a very dense catalogue, we use a spatio-temporal window of 2 days for T and 2 km for D to ensure a significant number of events.

3.2.3 Hyperparameter tuning

To optimise the SOM learning process, three hyperparameters are fine-tuned : the number of grid nodes, the number of iterations to achieve optimal clustering results, and the number of training samples used to converge to a good learning performance. To find the best values for these hyperparameters (Table 1), we use two scoring metrics: Topological Error (TE) and Quantisation Error (QE) (Tsai et al., 2017):

- → **QE** measures the mean distance error of each input vector from its associated neuron. Its values range from 0 to ∞, with smaller values of QE corresponding to the definition of a model that fits the dataset perfectly.
- → **TE** is a global indicator that measures how well the structure of the input space is modelled by the map. More precisely, it evaluates the local discontinuities of the mapping. Thus, if two input vectors are neighbours in the dataset (their feature values are close) and they are neighbours on the map, then TE is 0, otherwise TE is 1. Taking the average value of TE for each input vector gives a value between 0 and 1, with TE close to 0 indicating a model that preserves the topology of the dataset.

The search for the best hyperparameters is therefore equivalent to finding a local minimum for TE and QE. To obtain these local minima, we perform several tests by recursively setting two hyperparameters to a fixed value and studying the variations of the third. To do this, we proceed in the following three steps until the values of the hyperparameters values converge:

1. We first set the number of training samples to the maximum available and the number of iterations to the largest possible based on the dataset, and then iteratively search for the best number of nodes needed for optimal clustering (maximum distance between clusters).

2. Once we have obtained the optimal number of nodes, we keep the number of training samples at the previous value, and we find the best number of iterations that will make TE and QE converge to a flat growth.
3. We finally use the optimal number of nodes and iterations found in the previous steps to find the best number of training samples that no longer increases the QE value without too much cost on the TE value.

Maintaining the trade-off between TE and QE ensures good learning of the neural network, since TE is a global parameter that quantitatively measures the degree of preservation of the original topology of the dataset, while QE is a relative parameter that measures the average Euclidean distance between an input vector and its best matching nodes.

Hyperparameter	Taiwan	Synthetic	GOC	Italy Cat1	Italy Cat2
Size (in nodes)	150x150	150x150	150x150	150x150	150x150
Samples for training	7500	10000	10000	28000	140000
Iterations	10000	15000	15000	40000	200000

Table 1: *Optimal hyperparameters used in the SOM training process.*

3.3 Post-hoc Analysis of the Trained Self-Organising Map

3.3.1 Identification of SOM Clusters Through Agglomerative Clustering

We train the SOM with a 25-dimensional training dataset. Each seismic event is described by an input vector containing the values of the 25 features described above. The SOM learning process leads to the creation of a reduced 2D space representing the high-dimensional dataset. We exploit the 2D SOM space by identifying each cluster displayed on the SOM map with an agglomerative clustering procedure (Pedregosa et al., 2011; Hubert & Arabie, 1985).

Agglomerative clustering is a type of hierarchical clustering used to group objects into clusters based on their similarity. Each cluster identified on the 2D SOM map should therefore contain only seismic events that share similar feature values.

3.3.2 Probabilistic Classification of SOM Clusters Identified by Agglomerative Clustering

3.3.3 Probabilistic Approach

Once the clusters have been identified by agglomerative clustering (Pedregosa et al., 2011; Hubert & Arabie, 1985), we classify each SOM cluster. This interpretation of the SOM output gives a new representation of the studied catalogue by assigning each event to a class: crisis class or non-crisis class.

To obtain a relevant classification of each event class, we develop a centroid-based probabilistic approach. For each event class (crisis and non-crisis classes), we define a reference centroid which corresponds to the centre of mass of an imaginary cluster. Whether the location centroid is real or imaginary, its coordinates are usually defined as the average feature values of all points in the cluster.

We assume that a seismic event j belongs to the crisis class if it has a high number of close neighbours i, if it is associated with a high magnitude ratio and a high b-value (larger proportion of small events), and if it is close in space and time to its neighbours i. Conversely, an event j belongs to the non-crisis class if it has a low number of close neighbours i, if it is associated with a low magnitude ratio and a low b-value (fewer proportion of small events), and if it is distant in space and time from its neighbours i.

For the crisis or non-crisis class, the coordinates of the reference centroid are therefore the feature values that will best define each class. These feature values are selected from all the coordinates of the real centroid clusters identified in the 2D SOM map.

Thus, for the non-crisis (crisis) class, the best feature values correspond to the lowest (highest) possible number of nearest neighbours, the lowest (highest) possible magnitude ratio and the highest (lowest) possible average space-time inter-event distance. For the b-value, we consider that the best feature value is 1, which is the classical b-value encountered during a quiet seismic period in a given area.

We then compare each real k-centroid identified in the 2D SOM map to each of the two reference centroids by calculating a relative deviation from each reference centroid coordinate (Equations 8 and 9). For all features, if a given k-centroid is further away from the reference centroid corresponding to the crisis class than from the reference centroid corresponding to the non-crisis class, it is classified as belonging to the non-crisis class. Conversely, if this k-centroid is closer, then it is classified as belonging to the crisis class.

We develop the probabilistic function according to the previously explained centroid-based approach. This function is presented in the following. In Equation 8, the variable EC_{max} is the relative deviation of each k-centroid coordinate (i.e. number of nearest neighbours, magnitude ratio, spatial or temporal distances between events) from the corresponding maximum coordinate found among the two reference centroids, while the variable EC_{min} is the relative deviation of each k-centroid coordinate from the corresponding minimum coordinate found among the two reference centroids. In Equation 9, the variable EC_1 corresponds to the relative deviation between the b-value of a given k-cluster and the reference value of 1. The variations from a b-value of 1 is supposed to be associated with the ability of an earthquake rupture to propagate (b-value lower than 1) or not (b-value higher than 1) once nucleated (Taroni & Akinci, 2020; Narteau et al., 2009; Mesimeri et al., 2019).

$$EC_{max}(Y,k) = \frac{|max(Y)-Y_k|}{max(Y)} \tag{7}$$

$$EC_{min}(Y,k) = \frac{|min(Y)-Y_k|}{min(Y)} \tag{8}$$

$$EC_1(Y,k) = \frac{|1-Y_k|}{1} \tag{9}$$

For each SOM cluster k, we define A_k and B_k (A and B for for crisis and non-crisis events respectively, see Equations 10 and 11). The variable A_k (or B_k) represents the sum of the relative distances between the coordinates of a given centroid k and the coordinates of the reference centroid corresponding to the crisis (or non crisis) class.

A given SOM cluster k is then classified according to the highest value of A_k or B_k obtained. The values of A_k and B_k are between $[0, \inf]$. If, for a given cluster of the SOM grid, the value of A_k is the highest, then this cluster is classified as belonging to the crisis class. Conversely, if B_k has the highest value, the cluster in question is classified as belonging to the non-crisis class.

$$\begin{aligned} A_k = EC_{max}(\bar{R_j}, k) + EC_{max}(\bar{T_j}, k) + EC_{min}(Nn, k) + EC_{min}(Mn, k) \\ +EC_1(Bval, k) + EC_{max}(r^2T_j, k) \end{aligned} \tag{10}$$

$$\begin{aligned} B_k = EC_{min}(\bar{R_j}, k) + EC_{min}(\bar{T_j}, k) + EC_{max}(Nn, k) + EC_{max}(Mn, k) \\ -EC_1(Bval, k) + EC_{min}(r^2T_j, k) \end{aligned} \tag{11}$$

In Equations 10 and 11, any $\bar{y}$ is the arithmetic mean equal to $\sum_j^N \frac{y_j}{N}$. We recall that the coordinates of each cluster k are equivalent to the average feature values of all points in that cluster. Therefore, Tj_k, Rj_k, Nn_k, Mn_k, $Bval_k$ and r^2Tj_k denote respectively the average temporal distances, the average spatial distances, the average number of neighbours, the average magnitude ratios, the average b-value and the average coefficient of determination of the 10 closest temporal distances of all points in a given cluster k.

We use a softmax function σ to interpret the values of A_k and B_k as probabilities, since this function is designed to transform the values into values between 0 and 1 (see Equation 12). In Equation 12, e is the exponential function, β a weighting factor (fixed to one in our study) and z_i the coordinates i of the vector z (in our case z = (A_k, B_k)) :

$$\sigma(\mathbf{z})_i = \frac{e^{\beta z_i}}{\sum_{j=1}^K e^{\beta z_j}} \tag{12}$$

Applying Equation 12 to Equations 10 and 11 yields two probability equations (13 and 14) for each SOM cluster k:

$$P_{crisis}(k) = \frac{e^{A_k}}{e^{A_k}+e^{B_k}} \tag{13}$$

$$P_{non\_crisis}(k) = \frac{e^{B_k}}{e^{A_k}+e^{B_k}} \tag{14}$$

To discretise our probability values on the entire 2D SOM space (i.e. the SOM node grid), we extrapolate the probability values from the centroid of each cluster k (see Equation 15).

$$C = \frac{\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k}{k} \tag{15}$$

In Equation 15, C is the centroid of a given cluster k containing a set of k points x. The centroid corresponds to the point that minimises the Euclidean distance to every point in the set of k points x.

This extrapolation allows us to obtain two probability values for each node of the SOM grid: $P_{crisis}(k)$ defines the probability that an event belongs to the class of crisis events and $P_{non_crisis}(k))$ the probability that the event is a non-crisis event. We do not choose to calculate the two probability values for each node of the SOM grid individually, as we want to keep the continuity of our 2D space and select the most informative values from the nearby points.

$T_{j,i}$	Temporal distances between each event j and its ten nearest neighbours i
$R_{j,i}$	Spatial distances between each event j and its ten nearest neighbours i
$Concentration_of_Energy$	Average magnitude in a window of duration T and width D
$Concentration\_of\_Energy\_norm$ or Mn	Average Magnitude in a window of duration $2T$ and width $2D$ divided by $Concentration\_of\_Energy$
$Nnear$ or Nn	Number of events normalised in a window of duration T and width D
$bval$	The b-value in a window of duration T and width D
$r2_Tj$	R-Squared with a linear regression of the ten T_j

Table 2: *Names and meaning of the features used.*

3.3.4 Confidence Level of the Probabilistic Classification

We calculate a confidence level of the probabilistic classification obtained for each node (see Equation 16). This confidence level represents the decisiveness of the classification.

$$Confidence = \frac{|0.5 - max(P_{crisis}, P_{non\_crisis})|}{0.5} \tag{16}$$

The main advantage of using a probabilistic approach for the classification of each SOM cluster is that, in the case of a complex catalogue with many clusters highlighted by the agglomerative clustering procedure, we can deduce the class of each SOM cluster by comparing the coordinates of the centroid of each cluster with the coordinates of the reference centroids calculated for each class (crisis and non-crisis).

3.3.5 Estimation of Feature Importance

After obtaining a probabilistic classification of each SOM cluster, we analyse the impact of the input features on the resulting classification using three complementary scoring metrics:

1. The significance provides an intrinsic spatial measure of feature importance on the 2D SOM map. The significance is the variance of the features divided by the number of features used.
2. The meaningfulness provides a class-specific measure of the importance of a feature in the classification, allowing the distinctive features (with a specific range of values) to be deduced for one of the classes. The meaningfulness is calculated as the maximum value of features minus the difference between the maximum and minimum value of features in one class divided by the maximum value of features.
3. The correlation of feature values to the final class gives a relative measure of the importance of features in the overall classification.

By using these feature importance measures, we want to better understand the overall SOM decision process and the relative importance of each dimension (feature) used on the 2D SOM projection. As the process is unsupervised, this is the only way to understand what information determines the position of the input feature vector on the SOM map and thus the inferred event class. In addition, studying how feature importance changes as we add new ones or remove others help us to select the best features for our application, with the objective of obtaining a global classification of crisis and non-crisis events that works for multiple catalogues and geographical areas.

4 Datasets Used

4.1 Synthetic Catalogue

To measure the absolute accuracy of our method, we use a "ground truth" dataset (i.e. a dataset labelled with 100% accuracy). However, such a dataset does not exist with real data, since the labelling is obtained from a preliminary declustering of the catalogue. Although the variations in the declustering results produced by different methods are actually small (see for example the variations in the ETAS model Mizrahi et al. (2022)), the fact that they are highly dependent on subjective choices of declustering parameters adds considerable uncertainty to the event class labelling.

Therefore, in order to avoid arbitrary and model-dependent relative comparisons, we create a synthetic dataset by generating classes of known events: seismicity crises (i.e. aftershock sequences and swarms) and non-crisis events (i.e. background events).

4.1.1 Catalogue Generation

We generate a deliberately simple catalogue in order to better analyse the 2D SOM map and to more easily highlight the limitations of the SOM learning process. In the following, we summarise the different steps and assumptions we use to produce our synthetic dataset (see also Figure 2).

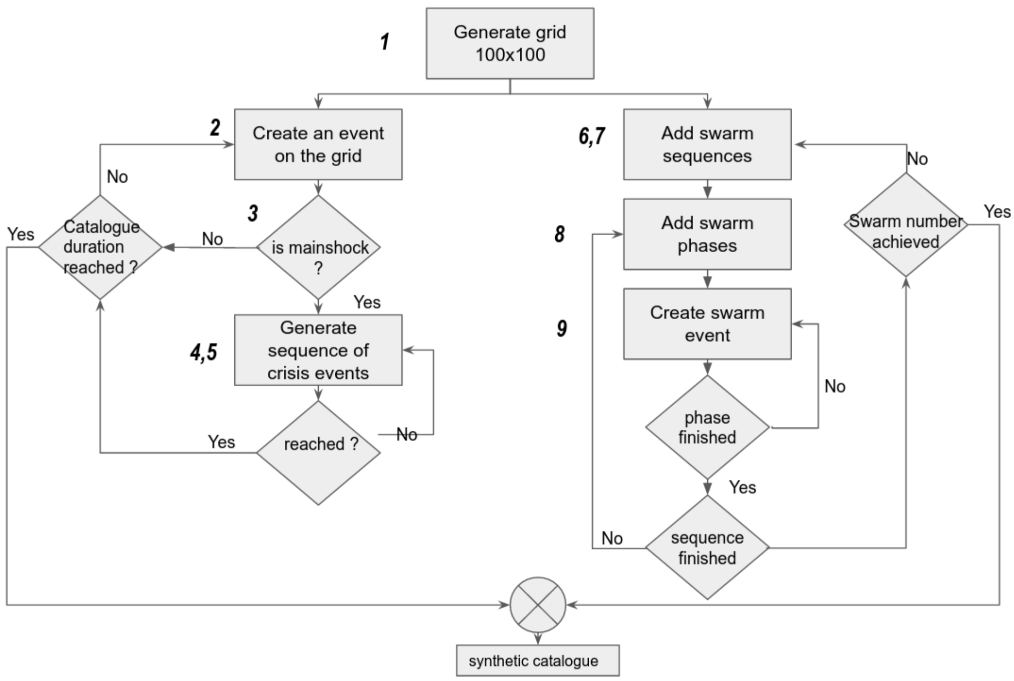


Figure 2: Pseudo-code used to generate the synthetic dataset. The numbers in bold refer to the numbers in the list summarising all the steps leading to the synthetic catalogue and described in the section 4.

We first create a 20-year catalogue containing events belonging to the non-crisis class.

1. These events are generated in a 2D map space of 1°x1° degree. Their origin times (in decimal years) and locations are evenly distributed over the entire time inter-

val. The origin times follow a uniform law bounded by 2000 and 2020 and their latitude and longitude follow a uniform law bounded by 0 and 1 degree.

2. We assign each event a moment magnitude (Mw) using an exponential distribution of rate $\lambda = 0.7$.

We then generate aftershock sequences for each identified mainshock:

3. We assign to each non-crisis event a probability of triggering a sequence of aftershocks (i.e. of being the first event in a new crisis sequence) equal to $0.1 * Mw$ if Mw < 5 otherwise 1.
4. Each aftershock sequence has a random duration that is a function of the magnitude of the mainshock.
5. The magnitudes of the aftershock sequence decrease exponentially with time and follow an exponential distribution law of rate $\lambda = 0.8$.
6. Each aftershock sequence has a longitude and a latitude that follow a normal distribution, with a mean μ equal to the latitude or longitude of the mainshock and a variance σ^2 equal to $4 + \epsilon$, where ϵ is a random value between -2 and 2.

We finally add 5 swarm sequences:

6. We assume that the swarms can occur uniformly over the 20-year catalogue.
7. Their magnitude Mw follows an exponential law of rate $\lambda = 0.8$.
8. The swarms are generated in N random phases (a uniform law bounded by 20 and 400) which produce a number of events according to a uniform distribution bounded by 1 and 10. The phases represent "bursts" of activity within a swarm crisis. Swarms are then represented by a succession of bursts in a spatially and temporally shifted location.
9. Each swarm is spaced in time by a random interval Dt of 0 to 3 days from the last swarm produced. Their spatial coordinates follow a normal distribution with a mean μ equal to the centroid (equation 15) of the previous swarming phase and a variance σ^2 equal to $10 + \epsilon$, ϵ being a random value between 0 and 4.

When generating the synthetic catalogue, we ensure that each seismic sequence is unique in terms of spatial distribution (i.e. inter-event distances) and event density (i.e. number of events per km^2). Each random variable used is therefore renewed at each new sequence (for swarms and aftershocks) and phase (for swarms).

4.2 Real Data: Study Areas

In the following, we briefly present the four earthquake times series we selected to test the ability of our approach to accommodate different deformation regimes and seismogenic patterns (Figure 4). These datasets are defined by wide time ranges, heterogeneous completeness magnitude and contrasting tectonic settings.

4.2.1 Gulf of Corinth (GOC)

The Gulf of Corinth is a continental rift with high seismicity rates and extensional deformation (Zelt et al., 2005). Numerous swarm sequences and frequent aftershocks following earthquakes of low to moderate magnitude (magnitudes of 5 and above are rare in recent catalogues) are recurrently recorded (Mesimeri et al., 2019).

The data used consist of 33,916 manually picked seismic events detected between 2010 and 2021 by a well-covered network of 46 three-component broadband seismometers maintained by the National Observatory of Athens (NOA) (Evangelidis et al., 2021) and the French Seismological and Geodetic Network (RESIF, 1995). The moment mag-

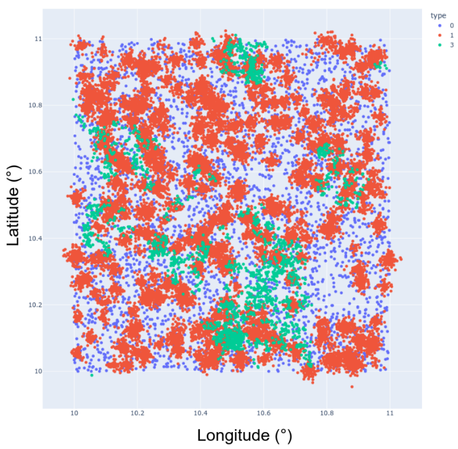


Figure 3: Spatial distribution of synthetic events generated using the procedure described in the section 4. In the legend, 0 corresponds to non-crisis events, 1 to aftershock sequences and 3 to swarms.

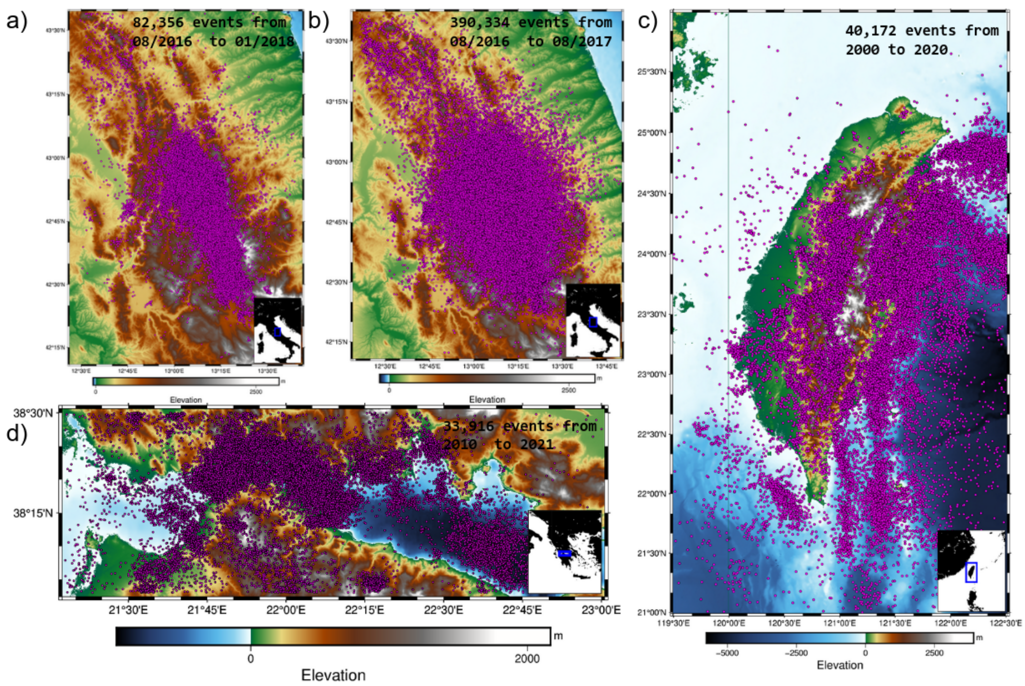


Figure 4: Maps of the four datasets used (before declustering). Each pink cricle indicates an earthquake of any magnitude. a) Italian catalogue CAT1 (Chiaraluce et al., 2022) b) Italian catalogue CAT4 (Chiaraluce et al., 2022) c) Taiwan catalogue (Peng et al., 2021) d) Corinth rift (GOC) catalogue (Evangelidis et al., 2021; RESIF, 1995).

nitude (Mw) of these events range from 0 to 5. The magnitude of completeness is equal to 1.2.

4.2.2 Taiwan

The island of Taiwan is the result of the collision between the Chinese continental margin and the Luzon volcanic arc. Due to the rapid subduction systems to the south and north of Taiwan, deformation rates across the island are extremely high, producing a large number of earthquakes in a wide range of magnitudes (Dadson et al., 2003). As in the GOC, this region has many swarms that are thought to be triggered by phenomena other than inter-earthquake triggering (such as fluid migration), as evidenced by earthquake clusters that deviate from Omori's law (Nishikawa & Ide, 2017).

The Taiwanese seismic data come from a recent study published by (Peng et al., 2021) who worked with a catalogue mainly from the Taiwan Central Weather Bureau Seismic Network. For consistency, we only use data from 2000 to 2020 in the entire Taiwan region (between 21.5°—25.5° longitude and 119.5°—122.9° latitude), including the nearest subduction zones. The maximum event depth is 50 km to ensure that most earthquakes occur either in the thickened continental crust or the upper oceanic lithosphere. The completeness magnitude and the minimum magnitude of the catalogue is 3.

4.2.3 Central Italy

The Italian peninsula is a fold-and-thrust belt undergoing a recent post-orogenic extension. Intense seismicity is recorded with low to moderate magnitude events and sometimes strong earthquakes. The Central Apennines have experienced numerous historical and instrumental earthquakes, mainly normal fault earthquakes (at least 16 events of magnitude greater than 6 before 2016), highlighting the predominance of the current extensional tectonic regime (Falcucci, E et al., 2016).

For this study, we use the two Italian seismic catalogues (called CAT1 and CAT4) provided by (Chiaraluce et al., 2022). Both catalogues (CAT1, CAT4) are published in the study area between $[12.5, 14]$ degrees longitude and $[42, 44]$ degrees latitude. The monitored sequence belongs to a 150-km long normal fault system. (Papadopoulos et al., 2017).

CAT1 covers the period between $2016-08-24$ and $2018-01-17$, and contains 82,356 manually reviewed events. This catalogue has a completeness magnitude of 1.5. CAT4 covers the period between 2016−08−24 and 2017−08−31, and contains 390,334 events detected shortly after the first mainshock of the Amatrice sequence of August 24, 2016 reaching magnitude 6. Its minimum completeness magnitude is estimated at 0.4.

5 Results

5.1 SOM methodology applied to synthetic data

For simplicity, we call "KDM" (see figure 1) our exploratory classification methodology based on the exploitation of SOM maps.

5.1.1 Classification Performance of the SOM-Based Method

After training the SOM network with the synthetic dataset, we obtain a 2D SOM map represented in Figure 5. A total of three SOM clusters are identified by the agglomerative procedure, each cluster representing similar feature input vector characteristics. These clusters are classified using the probabilistic approach we previously described in section 3.3.2. One cluster is classified as containing non-crisis events with high certainty and high confidence, while the other two are classified as containing crisis events, one with high confidence and the other one with low confidence (Figure 7). As we only have three clusters in our SOM map, we have extrapolated the probability values and confidence level using nearest neighbour interpolation to represent the boundary between

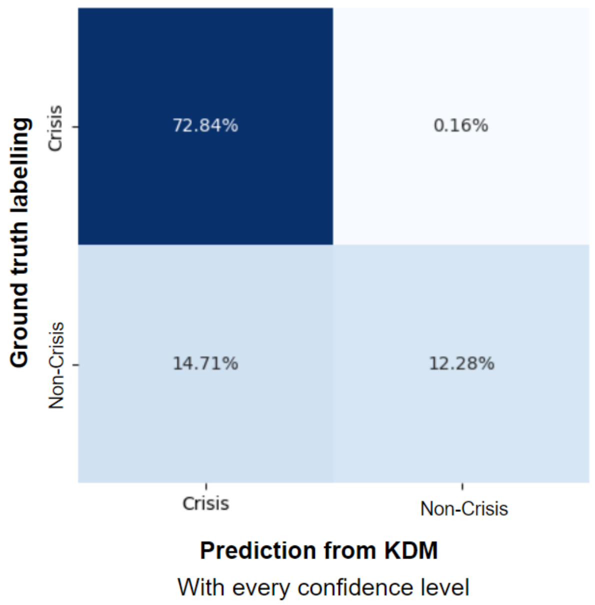


Figure 5: Confusion matrix obtained by comparing KDM predictions with ground truth labelling of synthetic events.

the two classes (four points are needed to interpolate linearly using the Qhull algorithm (Barber et al., 1996)).

In order to assess the classification accuracy of our method, we compare the classification results obtained with the "ground truth" labelling of each event class. As shown by the confusion matrix presented in Figure 5, our method presents a good average classification accuracy: 85% of events are correctly classified. While only 0.16% of non-crisis events are misclassified, our procedure seems to have more difficulties in classifying crisis events: about 15% of them are misclassified.

Our synthetic data contain two types of seismic sequences: mainshock-aftershock sequences and swarms. Looking at the classification accuracy for both sequences, we find that most of them are correctly classified (85% of accuracy) by our method. The errors are primary on swarms and can be explained by the nature of the crises we generated: swarm events are less concentrated in space and show a large variation in spatial and temporal inter-event distances. To better classify this type of events, it would probably be necessary to use a criterion other than their spatio-temporal distribution to relate them (for instance, the inter-correlations between waveforms).

The other factor causing misclassification of crisis events concerns events that occur in the vicinity of dense seismic clusters. Our method has some difficulty in determining whether an event close to a cluster in time and space is part of that cluster or not. This limitation actually stems from the choice of whether a non-crisis event can occur during a crisis period. Based on the assumptions chosen to generate our synthetic data (non-crisis events are equiprobable in time and space, swarms are episodic and randomly shifted, and aftershock sequences decrease exponentially in time and magnitude), we accept the occurrence of non-crisis events along with crisis events. However, based on the features we use to decluster our catalogue, these events are actually classified as crisis events. Our KDM method considers that in a crisis period, the conditional probability that an event close to a crisis is a non-crisis event is quite low. A rigorous distinction would require additional information that is not contained in the catalogues so far, such as fault plane solutions or the stress field.

The SOM 2D map shows three clusters that can be classified either according to the type of events encountered in the catalogues (i.e. background events, aftershocks, swarms) or according to the class of events that our study aims to identify (non-crisis and crisis classes). In the latter case, the third cluster could be defined as an indeterminate class. In fact, we observe that 90% of the non-crisis events belong to cluster 2 and 95% of the crisis events belong to cluster 1 (aftershocks and swarms classified with good confidence, see Figure 7). In addition, most of the swarms (63%), which are defined by inter-event space-time distances that can match both classes, belong to cluster 3. This observation could explain why the classification confidence of cluster 3 is low. Therefore, these results can invalidate the cluster classification based on event type and confirm that the SOM declustering approach is better suited to a classification based on two event classes: crisis events and non-crisis events.

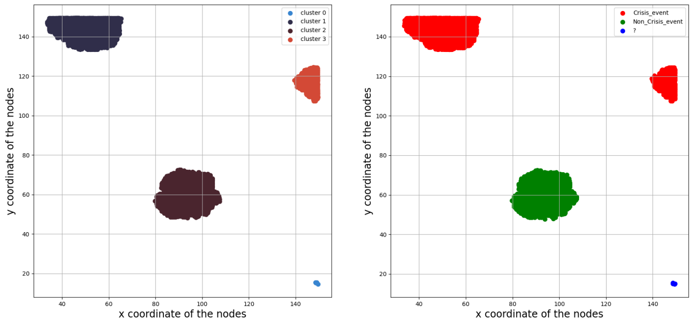


Figure 6: (left) 2D SOM map output for the synthetic dataset, each point is a vector and each colour is a SOM cluster (right) Classification of the resulting clusters using agglomerative clustering.

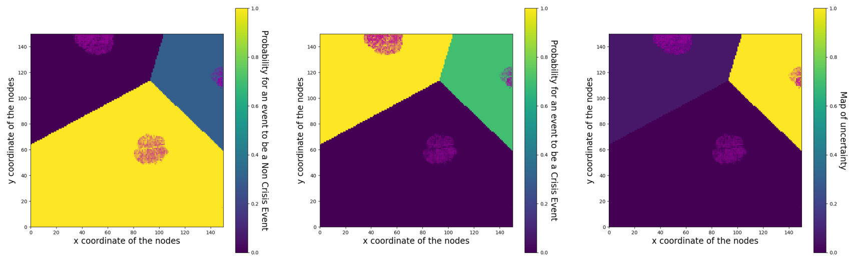


Figure 7: Probabilistic classification and confidence level for synthetic data: (left) probability of an event being a non-crisis event, (middle) probability of an event being a crisis event, and (right) classification confidence

5.2 Application to Real Data

5.2.1 SOM Representation

Unlike the 2D SOM map obtained from the synthetic data, the 2D SOM maps resulting from the real data (Gulf of Corinth, central Italy and Taiwan) contain more than three clusters. Each 2D SOM map gives a unique representation of SOM cluster patterns for each dataset (see Figure 8). The number of clusters obtained in the 2D SOM maps depends on the intrinsic complexity of the dataset, i.e. the size of the study area, the duration of the catalogues, the quality of the event locations, the number and density of the seismic sequences. However, in all cases, the 2D space manages to represent each dataset with clusters that can be easily classified as non-crisis or crisis events.

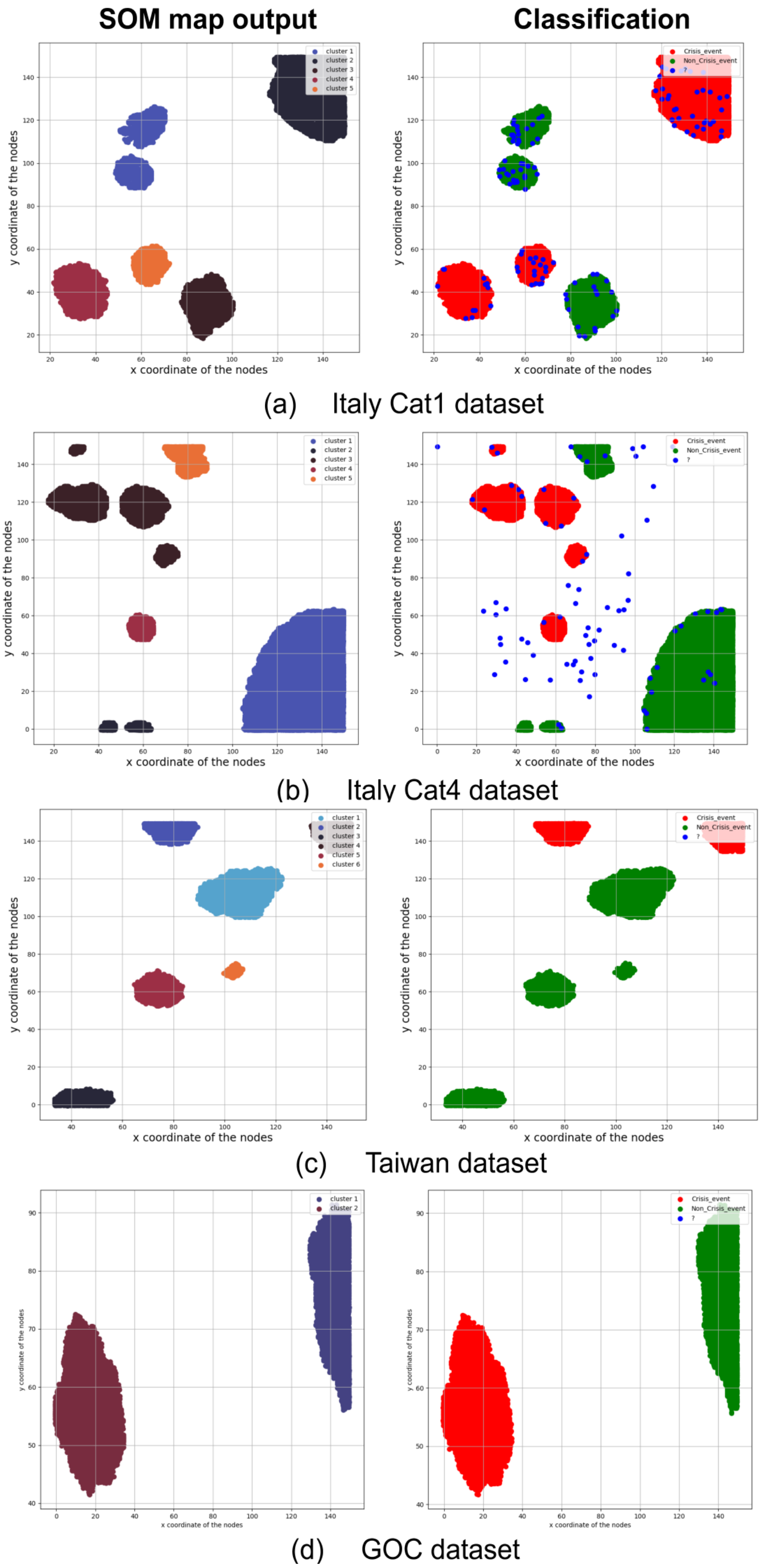


Figure 8: (left) 2D SOM maps obtained for the real data, each point is a vector and each colour is a SOM cluster (right) Resulting classification of identified clusters.

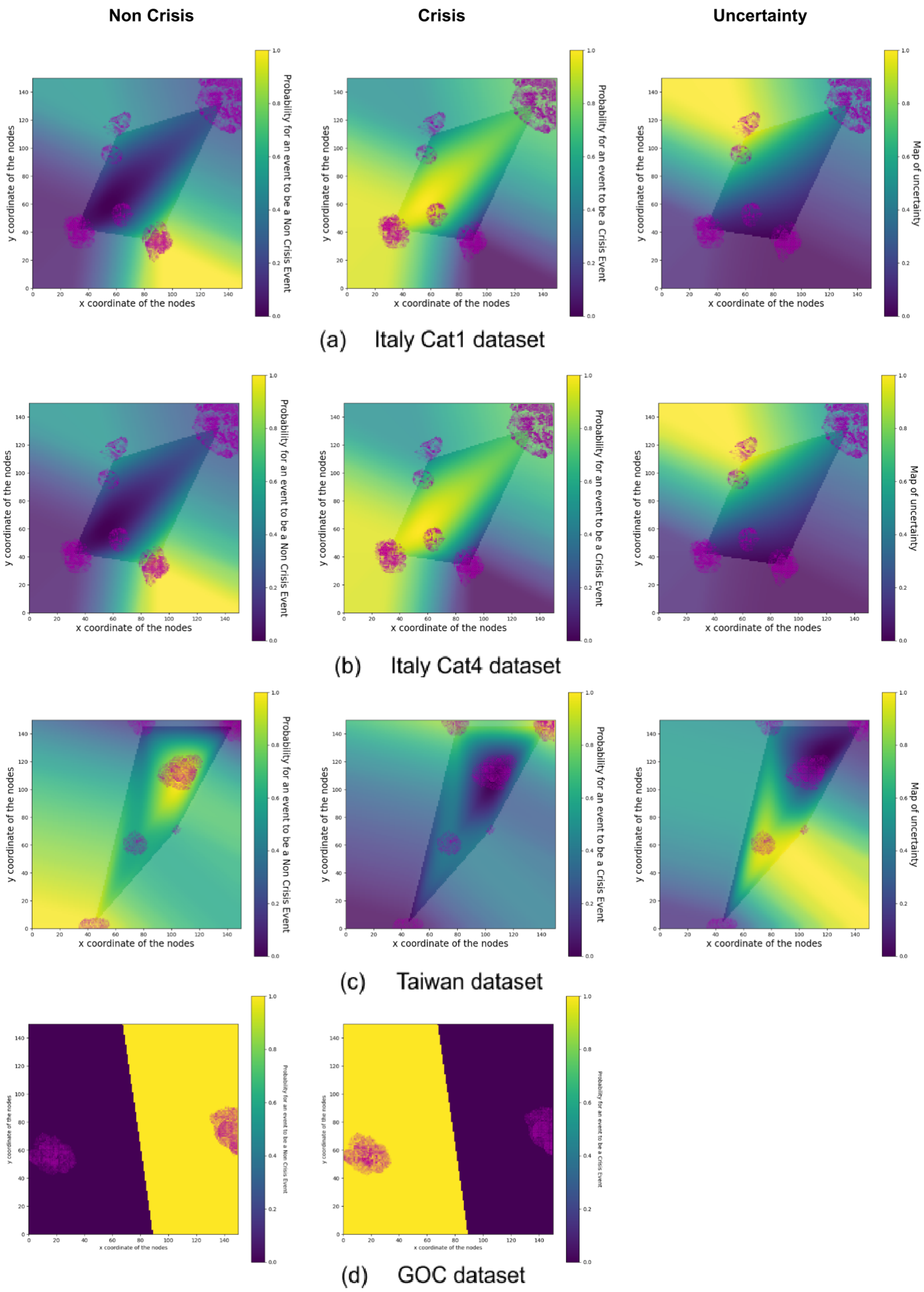
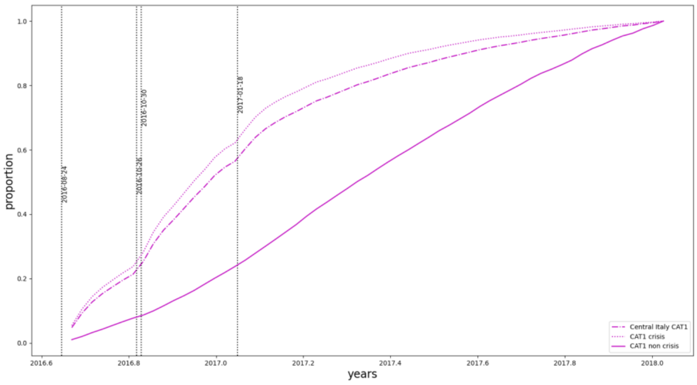
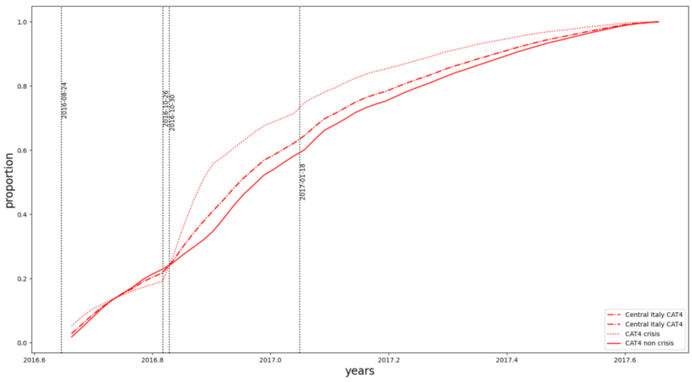


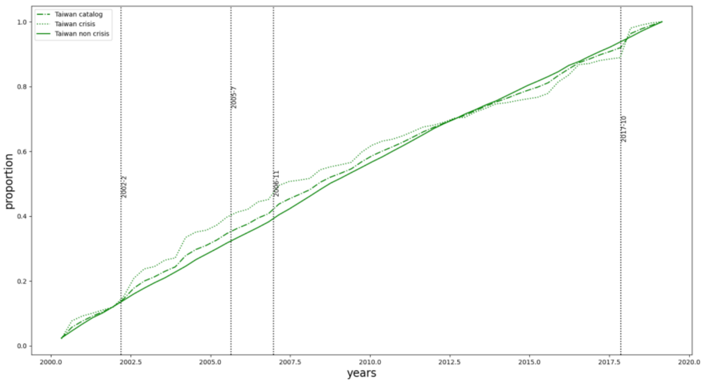
Figure 9: Probabilistic classification and confidence level for real data: (left) probability of an event being a non-crisis event, (middle) probability of an event being a crisis event, (right) confidence in the classification. Figures a,b,c are made using linear interpolation on the cluster centroid, figure d is made using nearest interpolation; this figure has only 2 cluster centroids.



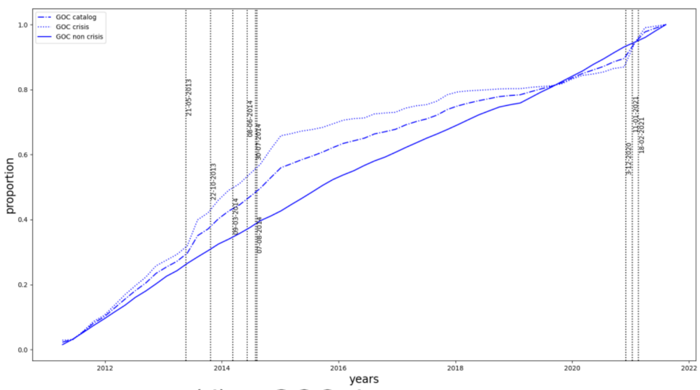
(a) Italy Cat1 dataset



(b) Italy Cat4 dataset



(c) Taiwan dataset



(d) GOC dataset

Figure 10: Cumulative curves obtained for the real datasets after applying our KDM methodology. For each dataset, the dashed line corresponds to the whole catalogue, the dotted and solid lines to the crisis and non-crisis events respectively.(a) (b) The vertical dotted lines refer to the date of the mainshocks Michele et al. (2020) (c) The vertical dotted lines refer to the beginning of the largest crises considering the number of events Peng et al. (2021) (d) The vertical dotted lines refer to the start of the seismic crises according to Papadimitriou et al. (2022) and Bountzis et al. (2020)

5.2.2 Cumulative Curves

By analysing the cumulative curves of the number of events versus time for each study area (figure 10), we observe that our declustering method leads to a classification of events with staircase behaviour for crisis events, as expected, while the temporal evolution of the number of non-crisis events does not seem to be correlated with the steps.

To further validate our results for the Corinth rift region, we perform a qualitative comparison with previous studies that have already described the major seismicity crises of 2021, 2017 and 2013-14. Our results are consistent with what was found in these studies (e.g. (Michas et al., 2021), (Bountzis et al., 2020), (Papadimitriou et al., 2022): each step observed in our cumulative curves are indeed identified after the start of each crisis (Figure 10). We also compare our results to seismic clusters described in (Mesimeri et al., 2019) which contain a total of 1560 crisis events. We find less than 1% differences. All the crises presented in their study and occurring during the period covered by the catalogue we extracted are identified by our method.

For the Taiwan region, we compare our results with those published by Peng et al. (2021). We find eighty-three percent similarity. Our method is consistent with the authors' classification of non-crisis events. However, when we compare the SOM classification of crisis events to the authors' classification, we find that our classification results are closer to the results obtained with their composite model than with their pure ETAS model. Their composite model combines three distinct declustering approaches (a modified ETAS model of Marsan et al. (2013), a nearest-neighbour method of Zaliapin and Ben-Zion (2013) and the classical approach of Reasenberg (1985)) and is used to improve swarm detection.

When our classification results disagree with those of Peng et al. (2021), our method often tends to classify mainshocks as crisis events if a low-magnitude event occurs nearby, whereas the approach of Peng et al. (2021) labels them as non-crisis events. Indeed, the magnitude of the precursor influences the number and magnitude of the hypothetical aftershocks in ETAS-based models (Console et al., 2010).

For the Central Italy region (Chiaraluce et al., 2022), each step observed on the cumulative curves corresponding to the two catalogues CAT1 and CAT4 is correlated with the occurrence of the mainshocks of the Amatrice sequence described in Michele et al. (2020) (see Figures 10 and 10). Overall, we observe that the cumulative curves of the non-crisis events corresponding to CAT1 and CAT4 are non-stationary and show a slightly variable growth rate. The non-stationarity observed reflects the absence of a quiet period before the Amatrice sequence in the catalogue. This curve shape confirms that our method does not alter the inherent properties of the dataset. For example, by forcing a linear background rate, in some cases, this non-stationarity may also indicate a change in the seismic productivity of the region.

5.3 Overall Feature Importance

Finally, we need to examine what are the most important features in the classification process (Figure 11). The average normalised magnitude appears to be quite significant: however, this only means that this dimension is dominant in the 2D SOM space. The correlation metrics show that the classification is mainly correlated with the relative spatial and temporal distances between events with a decrease in importance after the five nearest neighbours. The magnitude features, the coefficient of determination of the ten nearest temporal distances (in ascending order) and the local b-value feature also remain important, mainly for the classification of background events as shown by the meaningfulness values (Figure 11). These features are useful for distinguishing between nearby and related events.

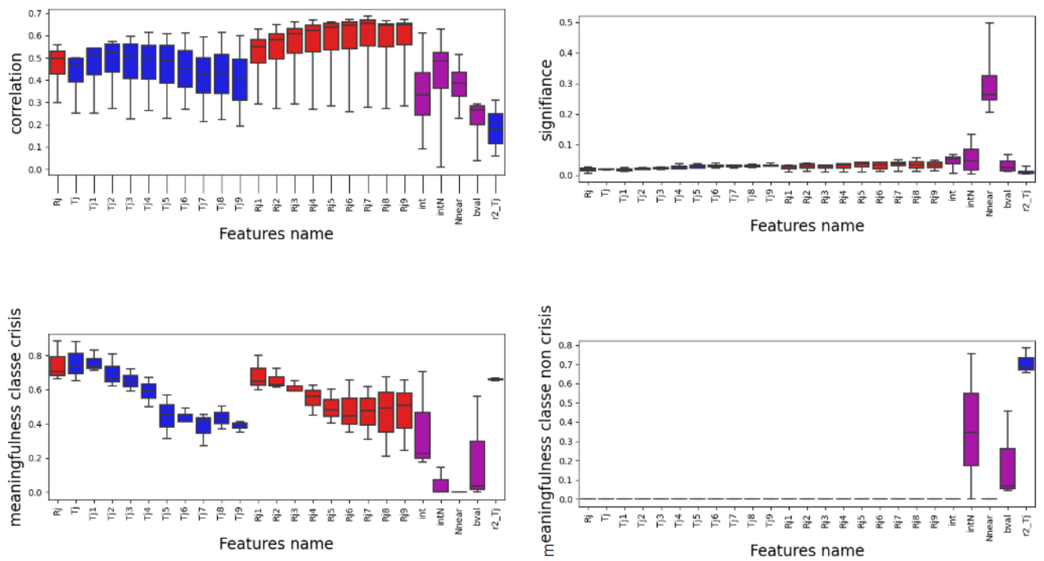


Figure 11: Estimation of feature importance using three metrics: meaningfulness, significance and correlation. The results are calculated for all datasets (Central Italy, Gulf of Corinth, Taiwan, and synthetic data). The blue boxes correspond to the temporal distance features, the red boxes to spatial distance features, and the purple boxes to the windowed features (number of nearest neighbours, magnitude ratios, b-value, coefficient of determination of the ten nearest temporal distances).

6 Discussion

6.1 Comparative Analysis of Declustering Results for Real Data

The datasets used in this study are very different, both in their geodynamic context and in the completeness of their magnitude. These differences likely explain why we obtain different ratios of crisis and non-crisis events.

In this study, we considered all classifications independently of their confidence level. Thus, depending on the complexity of the catalogue (i.e. the complexity of the seismic sequences that occurred), the ratio between crisis and non-crisis events could be influenced by the confidence threshold. For example, it is more difficult to detect background events than crisis events (Figure 5) because the spontaneity criteria classically used to describe background events can be ambiguous and variable in time and space, especially between very heterogeneous catalogues.

In addition, the ratio of growth rates between the cumulative non-crisis curve and the cumulative crisis curve observed for each dataset (Figure 10) strongly depends on the initial selection of the catalogues by their authors. For example, the Taiwanese catalogue lacks low magnitude events, while the Italian catalogues are dominated by small events. It is therefore expected that a greater number of crisis events will be detected in the Italian catalogues than in the Taiwanese catalogue.

The GOC dataset contains 10 times less recorded events, but has a duration 10 times longer than the Italian datasets. As a result, the 2D SOM map for the Corinth region clearly identifies two clusters, each representing a class of events, either crisis or non-crisis (Figure 8). These two clusters are scattered due to the diversity of seismic sequences recorded in the GOC catalogue over 10 years. Despite the lower overall number of events, this cat-

Dataset	Number of Crisis Events	Number of Non-crisis Events	T (days)	D (km)
Taiwan	8036	13839	9.20	19.66
SyntheticData	21901	8099	4.37	15.0
GOC	18392	13608	2.53	8.79
Italy_CAT1	69538	12122	1.03	0.38
Italy_CAT4	117745	272257	0.47	0.05

Table 3: *Classification of event classes for each dataset used. The lower probability threshold for the classification of events as crisis events is 0.5*

.

alogue covers a quiet seismological period (in terms of seismic activity). With this dataset, our KDM method has the ability to learn more efficiently the relationships between features and labels of non-crisis targets. The duration of the catalogue therefore seems to be a more critical factor than its size for a successful declustering.

Finally, the Taiwan dataset, with a minimal magnitude of 3 but a duration of 20 years and a more complex geodynamic setting, is difficult to interpret. The 2D SOM map shows many clusters compared to the number of events recorded in the catalogue. These multiple clusters suggest the existence of several types of crisis sequences with varying inter-event relationship characteristics (spatial distances, temporal distances, magnitude distribution). Moreover, the absence of low magnitude events in the catalogue makes the crisis sequences incomplete, artificially increasing the spatial and temporal distances between events in the same sequence. Although the distinction between non-crisis events and crisis events is difficult to manage in terms of spatial and temporal distribution, we can clearly highlight the different crisis sequences in the cumulative curves, with a background curve that increases with the average evolution of the number of events. Again, the duration of the catalogue determines the learning quality of the SOM network, because this quality is improved with a greater diversity of data distributions in time, space and magnitude. Catalogue duration is therefore a key factor in obtaining the most accurate classification, although classification uncertainty is highly variable.

However, regardless of the dataset used, cumulative curves should be interpreted with caution. The non-crisis and crisis curves cannot be completely independent from each other owing to the relaxation and reloading process that occurs between and at the same time as the crisis sequences. Therefore, the non-linear behaviour of the cumulative curves corresponding to the Italian catalogues cannot be interpreted as mere errors. Indeed, it remains an open question whether a linear trend in the number of non-crisis events over time should actually be expected, even more so around periods of occurrence of swarms, foreshocks and aftershocks, i.e. before or after a crisis sequence (Lombardi et al., 2010; Llenos & Michael, 2019).

6.2 Potential Future Applications of the Method

The method developed here uses little memory and works quite quickly, even on a laptop. For a dataset of 100,000 events, it takes about 20 minutes. This makes it an easily accessible tool, even for non-specialists.

Our KDM workflow, from input features to probabilistic formula, is very flexible: all users can add their own features or weights without any additional research work. The method we propose can even be applied to the classification of more specific events depending on the user's classification goals.

As the selection of neighbours is only done backwards in time when calculating the inter-event distances, the procedure is applicable in real-time, which increases the applicability of this method.

The method makes only relative use of the catalogue information, so that spatial features related to uncertain event locations do not bias the SOM training. In addition, no preliminary threshold is required for classification, allowing users to have interpretable crisis and non-crisis classes without subjective assumptions or instabilities in the classification results that could be obtained by changing the threshold.

Our method does not require manual post-windowing. On the other hand, the larger and more geodynamically diverse the area, the better the SOM is able to learn.

With our method, we first explore the datasets by calculating the relative distances in time, space, magnitude variations, without having to assume any type of distribution for any of the event classes.

However, the classification accuracy of the method depends on the length of the dataset (e.g. time period and spatial coverage) to achieve statistical robustness of the SOM decision response. For shorter datasets, this limitation could be resolved by manually inspecting the clusters highlighted by the SOM and determining for each the probability of being linked to crisis events.

Another shortcoming of the method is the difficulty in detecting background events that are close in space and time to extended space-time seismic clusters or swarms. To improve the method, further research on potential features that can measure the link between rupture physics and earthquake propagation is underway. We propose to use waveform inter-correlations as an indicator. This would not really increase the computational time as many catalogues are relocated using cross-correlation approaches, so this dormant information would be readily available.

7 Conclusions/Perspectives

In this study, we sought to build a more homogeneous and less subjective declustering approach than previous declustering attempts in order to improve catalogue analyses. The KDM method we propose is an unsupervised process that learns directly from input features without the need for a human-labelled dataset. This unsupervised machine learning approach can therefore reveal new hidden patterns from datasets that are less biased by human input.

As KDM does not learn from the posterior labelling of events established by another existing declustering method, it offers the possibility of declustering catalogues with fewer assumptions (no spatial distribution or productivity rate is assumed), and hopefully new insights. For example, our method does not impose an initial background rate or productivity rate for swarms, since it relies only on a relative comparison of parameters with respect to spatial and temporal neighbours. Furthermore, the SOM approach used here greatly increases the "distances" on its map representation, providing an easy-to-read distribution figure. As shown by the results obtained with synthetic data and real catalogues from Greece, Italy and Taiwan, 2D SOM maps provide a fairly new representation of the spatio-temporal distribution of earthquakes, useful for identifying and discussing the different modes (Zaliapin & Ben-Zion, 2022) present in a catalogue.

Our KDM declustering method taught us that the space-time distances between events are the most important features, not only for the first neighbours, but also for the other ones, as the probability of being a crisis event increases with the number of nearby events. However, we still need additional features that are not a function of space and time to better classify crisis events. In particular, the addition of new features will reduce classification ambiguity between nearby events that are not crisis events and events that are actually part of a crisis, especially in the tail of crisis sequences.

Our systematic way of interpreting the 2D representation provided by the SOM network is based on a probabilistic approach that allows users to decide on the degree of accuracy they wish to achieve depending on their use. This method can be applied at any scale, as it has been designed to work on datasets of different sizes. Finally, this method does not rely on strong assumptions, so that it is possible to compare the background rate or the productivity rate without the bias of commonly used declustering approaches.

Open Research Section

For this study, we use the SOM python libraries from V (2018), and Pedregosa et al. (2011), McKinney (2010), (Harris et al., 2020) for the data management, useful interpolating function and random number generator.

The Greek catalogue used in this paper is available from Evangelidis et al. (2021) and RESIF (1995) via `https://eida.gein.noa.gr/webdc3/,https://seismology.resif.fr/fr/constructeur-de-requetes-dataselect/#/`. The catalogues of Central Italy (Cat1 and Cat4) are freely available in Chiaraluce et al. (2022). The catalogue of Taiwan was obtained by contacting the corresponding author (see (Peng et al., 2021)).

Acknowledgments
We are indebted to Guy Ouillon for his helpful advice during this study. We also warmly thank Wei Peng for providing us with the Taiwan seismicity dataset. This work was supported by ISblue project, Interdisciplinary graduate school for the blue planet (ANR-17-EURE-0015) and co-funded by a grant from the French government under the program "Investissements d'Avenir" embedded in France 2030.

References

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, *11*(3), 586–600. doi: 10.1109/72.846731

Aden-Antoniow, F., Frank, W., & Seydoux, L. (2022). An Adaptable Random Forest Model for the Declustering of Earthquake Catalogs. *Journal of Geophysical Research: Solid Earth*, *127*. doi: 10.1029/2021JB023254

Albaric, J., Déverchère, J., Petit, C., Perrot, J., & Le Gall, B. (2009). Crustal rheology and depth distribution of earthquakes: Insights from the central and southern east african rift system. *Tectonophysics*, *468*(1), 28-41. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0040195108002205` (Role of magmatism in continental lithosphere extension continental lithosphere extension) doi: https://doi.org/10.1016/j.tecto.2008.05.021

Barani, S., Ferretti, G., Massa, M., & Spallarossa, D. (2007). The waveform similarity approach to identify dependent events in instrumental seismic catalogues. *Geophysical Journal International*, *168*(1), 100–108. Retrieved 2022-11-28, from `https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.2006.03207.x` doi: 10.1111/j.1365-246X.2006.03207.x

Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.*, *22*(4), 469–483. Retrieved from `https://doi.org/10.1145/235815.235821` doi: 10.1145/235815.235821

Beroza, G. C., Segou, M., & Mostafa Mousavi, S. (2021). Machine learning and earthquake forecasting—next steps. *Nature communications*, *12*(1), 4761.

Bountzis, P., Papadimitriou, E., & Tsaklidis, G. (2020). Earthquake clusters identification through a markovian arrival process (map): Application in corinth gulf (greece). *Physica A: Statistical Mechanics and its Applications*, *545*, 123655. Retrieved from `https://www.sciencedirect.com/science/article/pii/`

S0378437119320394 doi: https://doi.org/10.1016/j.physa.2019.123655

Cheng, Y., & Ben-Zion, Y. (2019). Transient brittle-ductile transition depth induced by moderate-large earthquakes in southern and baja california. *Geophysical Research Letters*, *46*(20), 11109-11117. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL084315 doi: https://doi.org/10.1029/2019GL084315

Chiaraluce, L., Michele, M., Waldhauser, F., Tan, Y., Herrmann, M., Spallarossa, D., ... Segou, M. (2022). A comprehensive suite of earthquake catalogues for the 2016-2017 Central Italy seismic sequence. *Scientific Data*, *9*(1), 710. Retrieved 2022-11-30, from https://www.nature.com/articles/s41597-022-01827-z doi: 10.1038/s41597-022-01827-z

Console, R., Jackson, D. D., & Kagan, Y. Y. (2010). Using the ETAS Model for Catalog Declustering and Seismic Background Assessment. *Pure and Applied Geophysics*, *167*(6-7), 819–830. Retrieved 2023-01-31, from http://link.springer.com/10.1007/s00024-010-0065-5 doi: 10.1007/s00024-010-0065-5

Dadson, S. J., Hovius, N., Chen, H., Dade, W. B., Hsieh, M.-L., Willett, S. D., ... Lin, J.-C. (2003). Links between erosion, runoff variability and seismicity in the Taiwan orogen. *Nature*, *426*(6967), 648–651. Retrieved from https://doi.org/10.1038/nature02150 doi: 10.1038/nature02150

Evangelidis, C. P., Triantafyllis, N., Samios, M., Boukouras, K., Kontakos, K., Ktenidou, O., ... Tselentis, G. (2021). Seismic Waveform Data from Greece and Cyprus: Integration, Archival, and Open Access. *Seismological Research Letters*, *92*(3), 1672-1684. Retrieved from https://doi.org/10.1785/0220200408 doi: 10.1785/0220200408

Falcucci, E, Gori, S, Galadini, F, Fubelli, G, Moro, M, & Saroli, M. (2016). Active faults in the epicentral and mesoseismal Ml 6.0 24, 2016 Amatrice earthquake region, central Italy. Methodological and seismotectonic issues. *Annals of Geophysics*, *59*, 5. Retrieved 2023-01-20, from http://www.annalsofgeophysics.eu/index.php/annals/article/view/7266 doi: 10.4401/ag-7266

Field, E. H., Milner, K. R., & Luco, N. (2022). The Seismic Hazard Implications of Declustering and Poisson Assumptions Inferred from a Fully Time-Dependent Model. *Bulletin of the Seismological Society of America*, *112*(1), 527–537. Retrieved 2023-01-31, from https://pubs.geoscienceworld.org/bssa/article/112/1/527/608112/The-Seismic-Hazard-Implications-of-Declustering doi: 10.1785/0120210027

Field, E. H., Milner, K. R., Page, M. T., Savran, W. H., & van der Elst. (2021). Improvements to the Third Uniform California Earthquake Rupture Forecast ETAS Model (UCERF3-ETAS). *The Seismic Record*, *1*(2), 117–125. Retrieved 2022-06-12, from https://doi.org/10.1785/0320210017 doi: 10.1785/0320210017

Frohlich, C., & Davis, S. D. (1990). Single-Link Cluster Analysis as a Method to Evaluate Spatial and Temporal Properties of Earthquake Catalogues. *Geophysical Journal International*, *100*(1), 19-32. Retrieved from https://doi.org/10.1111/j.1365-246X.1990.tb04564.x doi: 10.1111/j.1365-246X.1990.tb04564.x

Gurjar, N., & Basu, D. (2022). On the declustering methods of seismic catalogue — an application over Indian subcontinent. *Journal of Seismology*, *26*(5), 1077–1103. Retrieved 2022-12-12, from https://link.springer.com/10.1007/s10950-022-10105-9 doi: 10.1007/s10950-022-10105-9

Hainzl, S. (2022). ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake Catalogs. *Bulletin of the Seismological Society of America*, *112*(1), 494–507. Retrieved 2022-08-31, from https://pubs.geoscienceworld.org/bssa/article/112/1/494/607584/ETAS-Approach-Accounting-for-Short

-Term doi: 10.1785/0120210146

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. Retrieved from https://doi.org/10.1038/s41586-020-2649-2 doi: 10.1038/s41586-020-2649-2

Herrmann, M., Piegari, E., & Marzocchi, W. (2022). Revealing the spatiotemporal complexity of the magnitude distribution and b-value during an earthquake sequence. *Nature Communications*, *13*(1), 5087.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218. Retrieved 2023-02-03, from http://link.springer.com/10.1007/BF01908075 doi: 10.1007/BF01908075

Iacoletti, S., Cremen, G., & Galasso, C. (2022). Validation of the Epidemic-Type Aftershock Sequence (ETAS) Models for Simulation-Based Seismic Hazard Assessments. *Seismological Research Letters*, *93*(3), 1601–1618. Retrieved 2022-11-28, from https://pubs.geoscienceworld.org/srl/article/93/3/1601/611678/Validation-of-the-Epidemic-Type-Aftershock doi: 10.1785/0220210134

Llenos, A. L., & Michael, A. J. (2019). Ensembles of ETAS Models Provide Optimal Operational Earthquake Forecasting During Swarms: Insights from the 2015 San Ramon, California Swarm. *Bulletin of the Seismological Society of America*, *109*(6), 2145–2158. Retrieved 2023-02-07, from https://pubs.geoscienceworld.org/ssa/bssa/article/109/6/2145/573839/Ensembles-of-ETAS-Models-Provide-Optimal doi: 10.1785/0120190020

Lombardi, A. M., Cocco, M., & Marzocchi, W. (2010). On the Increase of Background Seismicity Rate during the 1997-1998 Umbria-Marche, Central Italy, Sequence: Apparent Variation or Fluid-Driven Triggering? *Bulletin of the Seismological Society of America*, *100*(3), 1138–1152. Retrieved 2023-02-07, from https://pubs.geoscienceworld.org/bssa/article/100/3/1138-1152/349415 doi: 10.1785/0120090077

Mancini, S., Segou, M., Werner, M., Parsons, T., Beroza, G., & Chiaraluce, L. (2022). On the use of high-resolution and deep-learning seismic catalogs for short-term earthquake forecasts: Potential benefits and current limitations. *Journal of Geophysical Research: Solid Earth*, *127*(11), e2022JB025202.

Marsan, D., Reverso, T., Helmstetter, A., & Enescu, B. (2013). Slow slip and aseismic deformation episodes associated with the subducting pacific plate offshore japan, revealed by changes in seismicity. *Journal of Geophysical Research: Solid Earth*, *118*(9), 4900–4909.

McKinney, W. (2010). Data structures for statistical computing in python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (p. 51 - 56).

Mesimeri, M., Karakostas, V., Papadimitriou, E., & Tsaklidis, G. (2019). Characteristics of earthquake clusters: Application to western Corinth Gulf (Greece). *Tectonophysics*, *767*, 228160. Retrieved 2022-09-15, from https://linkinghub.elsevier.com/retrieve/pii/S0040195119302616 doi: 10.1016/j.tecto.2019.228160

Michas, G., Kapetanidis, V., Kaviris, G., & Vallianatos, F. (2021). Earthquake Diffusion Variations in the Western Gulf of Corinth (Greece). *Pure and Applied Geophysics*, *178*(8), 2855–2870. Retrieved 2022-10-10, from https://link.springer.com/10.1007/s00024-021-02769-0 doi: 10.1007/s00024-021-02769-0

Michele, M., Chiaraluce, L., Di Stefano, R., & Waldhauser, F. (2020). Fine-Scale Structure of the 2016–2017 Central Italy Seismic Sequence From Data Recorded at the Italian National Network. *Journal of Geophysical Research: Solid Earth*, *125*(4). Retrieved 2022-12-05, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JB018440 doi: 10.1029/2019JB018440

Mizrahi, L., Nandan, S., Savran, W., Wiemer, S., & Ben-Zion, Y. (2022). *Question-driven ensembles of flexible ETAS models.* arXiv. Retrieved 2022-08-31, from http://arxiv.org/abs/2207.06247 (arXiv:2207.06247 [physics])

Narteau, C., Byrdina, S., Shebalin, P., & Schorlemmer, D. (2009). Common dependence on stress for the two fundamental laws of statistical seismology. *Nature*, *462*(7273), 642–645. Retrieved 2023-03-24, from http://www.nature.com/articles/nature08553 doi: 10.1038/nature08553

Nishikawa, T., & Ide, S. (2017). Detection of earthquake swarms at subduction zones globally: Insights into tectonic controls on swarm activity: Detection of Earthquake Swarms. *Journal of Geophysical Research: Solid Earth*, *122*(7), 5325–5343. Retrieved 2023-01-20, from http://doi.wiley.com/10.1002/2017JB014188 doi: 10.1002/2017JB014188

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, *83*(401), 9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*, 379–402.

Ogata, Y. (2004). Space-time model for regional seismicity and detection of crustal stress changes. *Journal of Geophysical Research: Solid Earth*, *109*(B3).

Papadimitriou, E., Bonatis, P., Bountzis, P., Kostoglou, A., Kourouklas, C., & Karakostas, V. (2022). The Intense 2020–2021 Earthquake Swarm in Corinth Gulf: Cluster Analysis and Seismotectonic Implications from High Resolution Microseismicity. *Pure and Applied Geophysics*. Retrieved 2022-10-10, from https://link.springer.com/10.1007/s00024-022-03135-4 doi: 10.1007/s00024-022-03135-4

Papadopoulos, G. A., Ganas, A., Agalos, A., Papageorgiou, A., Triantafyllou, I., Kontoes, C., ... Diakogianni, G. (2017). Earthquake Triggering Inferred from Rupture Histories, DInSAR Ground Deformation and Stress-Transfer Modelling: The Case of Central Italy During August 2016–January 2017. *Pure and Applied Geophysics*, *174*(10), 3689–3711. Retrieved 2023-01-20, from http://link.springer.com/10.1007/s00024-017-1609-8 doi: 10.1007/s00024-017-1609-8

Pavez O, C., & Estay H, R. (2021). Seismic declustering and seismic events pattern recognition in Norway: Preliminary results using the Wolfram Mathematica machine learning tools. In *Agu fall meeting abstracts* (Vol. 2021, p. IN25A-0450).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peng, W., Marsan, D., Chen, K. H., & Pathier, E. (2021). Earthquake swarms in Taiwan: A composite declustering method for detection and their spatial characteristics. *Earth and Planetary Science Letters*, *574*, 117160. Retrieved 2022-11-15, from https://linkinghub.elsevier.com/retrieve/pii/S0012821X21004155 doi: 10.1016/j.epsl.2021.117160

Pisarenko, V., & Rodkin, M. (2019). Declustering of seismicity flow: statistical analysis. *Izvestiya, Physics of the Solid Earth*, *55*, 733–745.

Reasenberg, P. (1985). Second-order moment of central california seismicity, 1969–1982. *Journal of Geophysical Research: Solid Earth*, *90*(B7), 5479–5495.

RESIF. (1995). *Resif-rlbp french broad-band network, resif-rap strong motion network and other seismic stations in metropolitan france.* RESIF - Réseau Sismologique et géodésique Français. Retrieved from https://seismology.resif.fr/networks/#/FR doi: 10.15778/RESIF.FR

Scholz, C. H. (2002). *The mechanics of earthquakes and faulting* (2nd ed.). Cambridge University Press. doi: 10.1017/CBO9780511818516

Seydoux, L., Balestriero, R., Poli, P., Hoop, M. d., Campillo, M., & Baraniuk, R.

(2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature Communications*, *11*(1), 3972. Retrieved 2023-02-07, from https://www.nature.com/articles/s41467-020-17841-x doi: 10.1038/s41467-020-17841-x

Taroni, M., & Akinci, A. (2020). Good practices in PSHA: declustering, b-value estimation, foreshocks and aftershocks inclusion; a case study in Italy. *Geophysical Journal International*, *224*(2), 1174–1187. Retrieved 2023-03-24, from https://academic.oup.com/gji/article/224/2/1174/5911581 doi: 10.1093/gji/ggaa462

Trnkoczy, A. (2009). Understanding and parameter setting of sta/lta trigger algorithm. In *New manual of seismological observatory practice (nmsop)* (pp. 1–20). Deutsches GeoForschungsZentrum GFZ.

Tsai, W., Huang, S., Cheng, S., Shao, K., & Chang, F. (2017). A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map. *Science of The Total Environment*, *579*, 474–483. Retrieved 2022-10-05, from https://linkinghub.elsevier.com/retrieve/pii/S0048969716325104 doi: 10.1016/j.scitotenv.2016.11.071

Utsu, T., Ogata, Y., Ritsuko, S., & Matsu'ura. (1995). The Centenary of the Omori Formula for a Decay Law of Aftershock Activity. *Journal of Physics of the Earth*, *43*(1), 1–33. Retrieved 2022-11-02, from http://www.jstage.jst.go.jp/article/jpe1952/43/1/43_1_1/_article doi: 10.4294/jpe1952.43.1

V, G. (2018). *MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map.* Retrieved from https://github.com/JustGlowing/minisom/

van Stiphout, T., Zhuan, J., & D, M. (2012). Theme V–Models and Techniques for Analyzing Seismicity. *Corssa.org*. doi: 10.5078/corssa52382934

Weatherill, G., Pagani, M., & Garcia, J. (2016). Exploring earthquake databases for the creation of magnitude-homogeneous catalogues: tools for application on a regional and global scale. *Geophysical Journal International*, *206*(3), 1652–1676.

Zaliapin, I., & Ben-Zion, Y. (2013). Earthquake clusters in southern california i: Identification and stability. *Journal of Geophysical Research: Solid Earth*, *118*(6), 2847–2864.

Zaliapin, I., & Ben-Zion, Y. (2022). Perspectives on Clustering and Declustering of Earthquakes. *Seismological Research Letters*, *93*(1), 386–401. Retrieved 2023-01-31, from https://pubs.geoscienceworld.org/srl/article/93/1/386/607809/Perspectives-on-Clustering-and-Declustering-of doi: 10.1785/0220210127

Zaliapin, I., & Ben-Zion, Y. (2021). Perspectives on Clustering and Declustering of Earthquakes. *Seismological Research Letters*, *93*(1), 386-401. Retrieved from https://doi.org/10.1785/0220210127 doi: 10.1785/0220210127

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., & Wong, H. (2008). Clustering Analysis of Seismicity and Aftershock Identification. *Physical review letters*, *101*, 018501. doi: 10.1103/PHYSREVLETT.101.018501

Zelt, B. C., Taylor, B., Sachpazi, M., & Hirn, A. (2005). Crustal velocity and Moho structure beneath the Gulf of Corinth, Greece. *Geophysical Journal International*, *162*(1), 257–268. Retrieved from https://doi.org/10.1111/j.1365-246X.2005.02640.x (_eprint: https://academic.oup.com/gji/article-pdf/162/1/257/5903026/162-1-257.pdf) doi: 10.1111/j.1365-246X.2005.02640.x

Zhang, Y., & Huang, Q. (2022). Seismicity Changes before Major Earthquakes in Sichuan, China, Revealed by a Combination of the RTL Algorithm and ETAS Model. *Seismological Research Letters*. Retrieved 2023-01-31, from https://pubs.geoscienceworld.org/srl/article/doi/10.1785/

0220220282/619381/Seismicity-Changes-before-Major-Earthquakes-in doi: 10.1785/0220220282

Zhu, C., Cotton, F., Kawase, H., & Nakano, K. (2023). How well can we predict earthquake site response so far? machine learning vs physics-based modeling. *Earthquake Spectra*, *39*(1), 478–504.

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic Declustering of Space-Time Earthquake Occurrences. *Journal of the American Statistical Association*, *97*(458), 369–380. Retrieved from https://doi.org/10.1198/016214502760046925 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/016214502760046925) doi: 10.1198/016214502760046925

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, *109*(B5).