NESTORE v1.0: A MATLAB based code to forecast strong aftershocks applied to Italian seismicity

GENTILI, Stefania[(1)]; BRONDI, Piero[(1)]; DI GIOVAMBATTISTA, Rita [(2)];

[(1)] National Institute of Oceanography and Applied Geophysics - OGS, Via Treviso 55, Udine, Italy, sgentili@ogs.it, pbrondi@ogs.it

[(2)] Istituto Nazionale di Geofisica e Vulcanologia (INGV), Via di Vigna Murata 605, Roma, Italy, rita.digiovamabattista@ingv.it

SUMMARY

NESTOREv1.0 is a MATLAB-based algorithm that uses a machine learning approach to provide a probabilistic forecasting of clusters in which a strong mainshock is followed by at least one subsequent earthquake of comparable magnitude. This objective is critical to mitigate seismic risk during strong seismic sequences because building already damaged by the mainshock occurrence can collapse, increasing damages and death toll. In particular, the algorithm distinguishes clusters in which the magnitude difference between the mainshock and the strongest aftershock is less than or equal to 1 (type A) and the other cases (type B). NESTOREv1.0 identifies clusters and is trained to distinguish the two typologies by using cluster seismicity parameters (features) on a training dataset. It is then able to produce a forecasting of A-type cluster for retrospective analysis on a test database or for ongoing clusters. For the application of NESTOREv1.0 to the Italian data, we focused our analysis on two areas covering the northeast of Italy and most of the complementary part of Italy. For these two areas, we used the seismic data recorded by the INGV and OGS seismic networks over the last 40 years. In particular, we trained NESTOREv1.0 for the clusters that occurred in the first 30 years of the catalogues and evaluated its performance in the period 2010-2021. We found for both areas that the percentage of correct forecasting of the cluster type a few hours after the occurrence of the mainshock is over 85%.

KEY WORDS

Spanish-Portuguese Assembly, Geophysics, Seismology, Machine Learning, Statistical Seismology

INTRODUCTION

After a strong earthquake, forecasting strong subsequent events is strategic for civil defense purposes because already weakened structures may suffer further damage, increasing the risk of collapse and fatalities. NESTORE (Next STrOng Related Earthquake) is a multi-parameter machine learning-based algorithm for probabilistic forecasting of the clusters in which a strong mainshock is followed by at least one following earthquake of comparable magnitude. More specifically, NESTORE labels clusters as type A if the magnitude difference between the mainshock and its strongest aftershock is less than or equal to 1, and Type B otherwise. The goal of NESTORE is to provide a near real-time estimation of the probability that the ongoing cluster is of type A. Some previous versions of the algorithm have been successfully applied to seismicity in California and Italy (Gentili & Di Giovambattista 2017, 2020, 2022).

The main issues concerning this type of application are essentially two:

1. The number of examples (clusters) available in good quality catalogs is typically an order of magnitude smaller than that typically required for machine learning applications.

2. The classes are unbalanced, because a subsequent strong earthquake is recorded in a percentage of less than 50% of the observed clusters.

Several algorithm improvements, as well as a renovation of the old version of the software, resulted in the current version of the software, NESTOREv1.0, which is now mature enough to be made available to the scientific community for use and testing in new seismo-tectonic domains, and will be available on GitHub in the coming months. It analyzes nine selected seismicity features at increasing time intervals after the occurrence of the mainshock, through a multiparameter pattern recognition approach based on one-node decision trees. These features are related to the aftershocks whose magnitude value is between the mainshock value Mm and Mm-2, and concern their number, spatial distribution, source area, magnitude and seismic energy over time.

NESTOREv1.0 consists of four modules:

1) Cluster Identification Module: identifies clusters from a seismic catalog

2) Training module: Determines appropriate thresholds for features to distinguish cluster types in a training set.

3) Testing Module: uses the result of the previous module to verify the performance of the trained algorithm in an independent test set

4) Near Real Time classification Module: allows near real-time application to clusters as they occur.

To test the algorithm, we considered for the Italian and western Slovenia territory the seismicity (Figure 1) recorded in the last 40 years by the INGV and OGS networks, the latter benefiting from a dense network in the northeast of Italy.

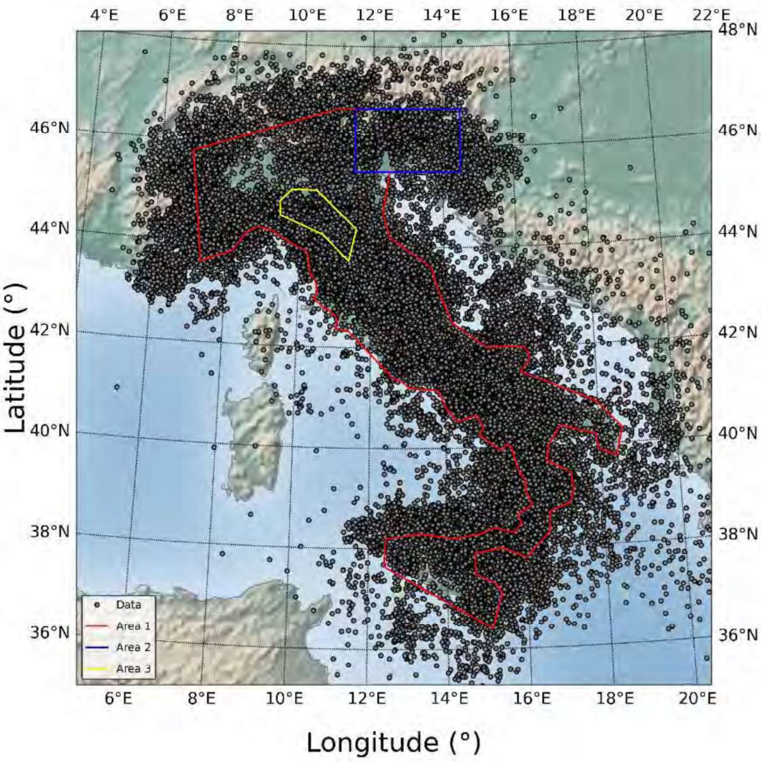

Figure 1: Seismicity recorded in Italy from 1980 to 2021. Area 1, Area 2 and Area3 are outlined in red, blue and yellow, respectively.

The analysis revealed three regions with different seismicity characteristics.

1. Area 1: Most of Italy, dominated by Apennines seismicity (Figure 1).

2. Area 2: The region of northeastern Italy and western Slovenia, which generally has lower productivity than Area 1.

3. Area 3: A region located in a smaller area between northwestern Tuscany and the southwestern part of Emilia-Romagna, with anomalous seismicity characterized by short-duration bursts with high productivity. Bonini et al. (2016) found that clusters occurring in that area release a big amount of seismic energy in short time, suggesting a complex interaction between fault seismicity.

While Area 3 cannot be analyzed, because the available training set for this region consists almost exclusively of type B clusters, we analyzed Areas 1 and 2 by a separate training of NESTOREv1.0.

In the case of Area 1, we analyzed the INGV catalogs on the time interval 1980-2021, using the Lolli & Gasperini (2006) catalogue for the years [1980-2004] and ISIDE catalogue (ISIDe Working Group, 2007) for the years [2005-2021]. In particular, we trained NESTOREv1.0 on 22 clusters occurring in the period 1980-2009 (Figure 2) and we tested its forecasting capability on 14 clusters occurring between 2010 and 2021 (Figure 3).

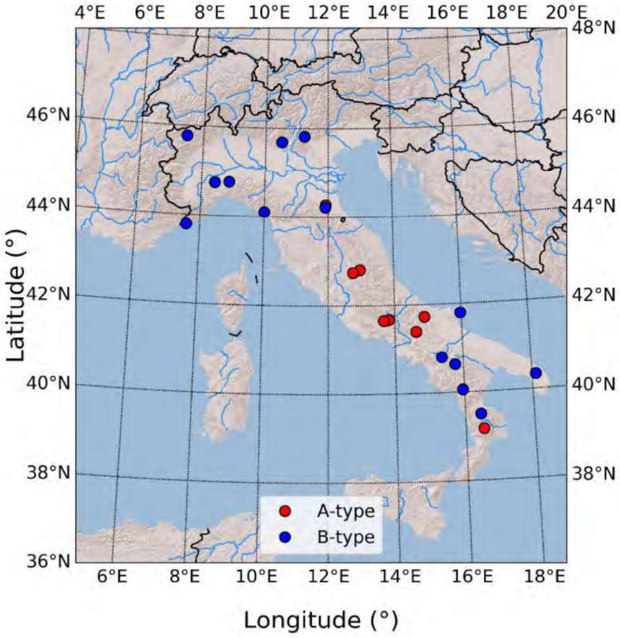

Figure 2: Training database used for Area 1. The circles correspond to the mainshock position of the clusters, the color to their type.

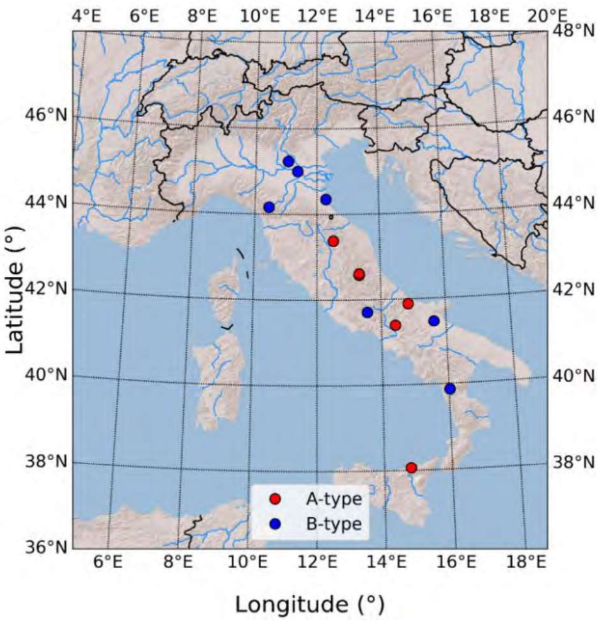

Figure 3: Testing database used for Area 1. The circles correspond to the mainshock position of the clusters, the color to their type.

For Area 2, we used OGS catalog (Snidarcig, et al., 2021; Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS., 2016; Friuli Venezia Giulia Seismometric Network Bulletin, 2021) between 1977 and 2021. Specifically, we used 13 clusters occurring in the 1977-2009 period (Figure 4) for training and we evaluated the forecasting performance on 18 clusters correspondent to the 2010-2021 period (Figure 5).

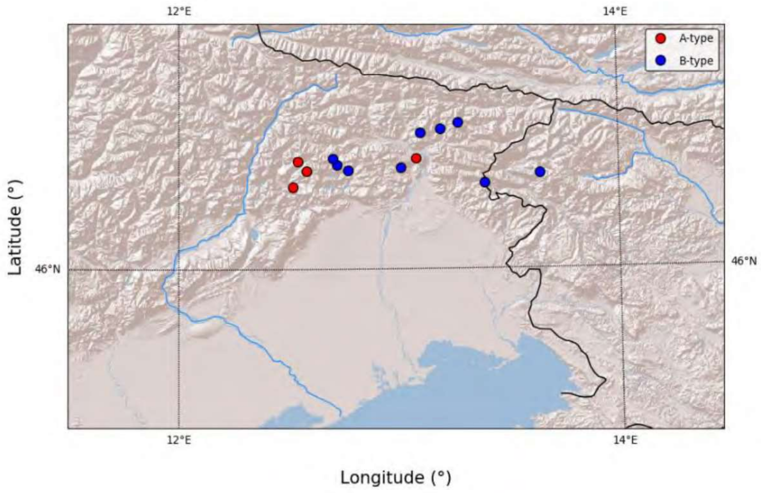

Figure 4: Training database used for Area 2. The circles correspond to the mainshock position of the clusters, the color to their type.

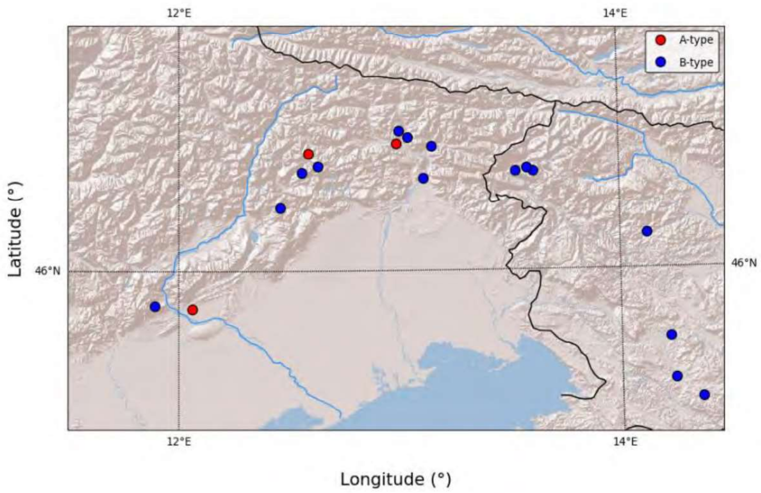

Figure 5: Testing database used for Area 2. The circles correspond to the mainshock position of the clusters, the color to their type.

The software test resulted in a percentage of correct cluster type forecasting equal to 86% at 12 hours for Area 1 (Figure 6) and to 94% at 6 hours after mainshock for Area 2 (Figure 7). The misclassifications obtained for Area 1 refer to two A type clusters that occurred in Central Italy (2017) and Molise (2018) region (Figure 7). In the first case, we assumed an anomalous seismic behavior due to the occurrence of the sequence in central Italy in 2016, while in the second case a more detailed study was carried out to better investigate the characteristics of the cluster. The only wrong forecasting given in Area 2 is related to the Valdobbiadene A type cluster (2015), characterized by a lack of events between the mainshock and its strong aftershock (Figure 8). We concluded that our results are encouraging for the application of NESTOREv1.0 in Italian territory.

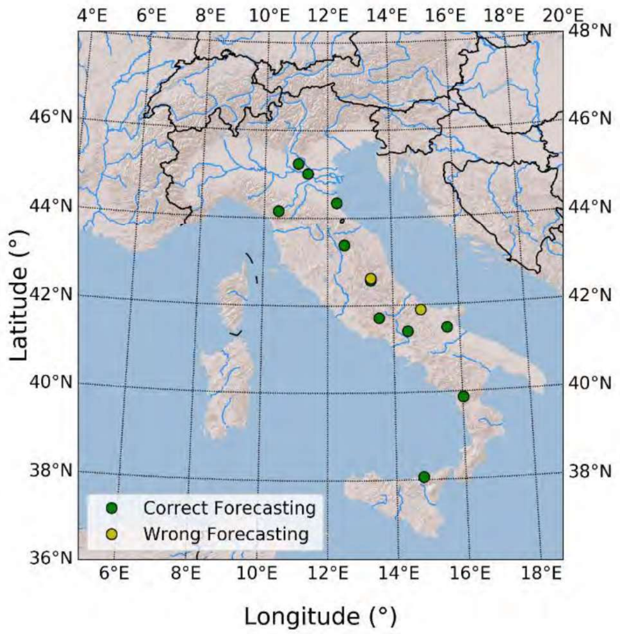

Figure 6: NESTOREv1.0 performance for Area 1 at 12 hours after mainshock. The circles correspond to the mainshock position of the clusters, the color to the quality of the classification.

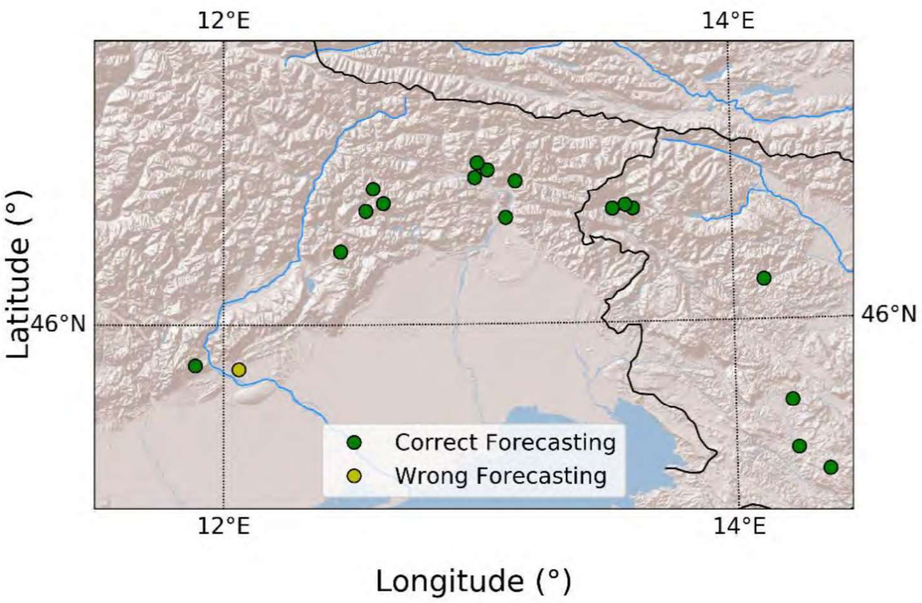

Figure 7: NESTOREv1.0 performance for Area 2 at 6 hours after mainshock. The circles correspond to the mainshock position of the clusters, the color to the quality of the classification.

ACKNOWLEDGEMENTS

Funded by a grant from the Italian Ministry of Foreign Affairs and International Cooperation

REFERENCES

[1] Bonini, M., G. Corti, D. Delle Donne, F. Sani, L. Piccardi, G. Vannucci, R. Genco, L. Martelli and M. Ripepe (2016): "Seismic sources and stress transfer interaction among axial normal faults and external thrust fronts in the Northern Apennines (Italy): A working hypothesis based on the 1916–1920 time–space cluster of earthquakes". *Tectonophysics*, 680, 67-89.

[2] Friuli Venezia Giulia Seismometric Network Bulletin (2021): http://www.crs.inogs.it/bollettino/RSFVG/RSFVG.en.html

[3] Gentili, S. and R. Di Giovambattista (2017): "Pattern recognition approach to the subsequent event of damaging earthquakes in Italy". *Physics of the Earth and Planetary Interiors*, 266, 1-17.

[4] Gentili, S. and R. Di Giovambattista (2020): "Forecasting strong aftershocks in earthquake clusters from northeastern Italy and western Slovenia". *Physics of the Earth and Planetary Interiors*, 303, 106483.

[5] Gentili, S. y R. Di Giovambattista (2022): "Forecasting strong subsequent earthquakes in California clusters by machine learning". *Physics of the Earth and Planetary Interiors*. 327, 106879.

[6] ISIDe Working Group. (2007): "Italian Seismological Instrumental and Parametric Database (ISIDe)." Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/ISIDE

[7] Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS. (2016). *North-East Italy Seismic Network* [Data set]. FDSN. https://doi.org/10.7914/SN/OX

[8] Lolli, B. and P. Gasperini (2006): "Comparing different models of aftershock rate decay: The role of catalog incompleteness in the first times after main shock." *Tectonophysics* 423, 43–59.

[9] Snidarcig, A., P. Bernardi, P.L. Bragato, P. Di Bartolomeo, M. Garbin, S. Urban, (2020): "Bollettino della Rete Sismometrica dell'Italia Nord Orientale (RSINO)" [Data set]. Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS. DOI: 10.13120/108b8d94-361a-45f3-8195-fc4e8f73d264