Deliverable

[D2.8 Progress of new generation catalogues for public dissemination]

Deliverable information	
Work package	[WP2]
Lead	[INGV, OGS]
Authors	[L. Chiaraluce[1], M. Michele[1], R. Di Stefano[1], B. Castello[1] –D. Latorre[1], A. Vuan[2], S. Campanella[2], M. Sugan[2]]
Reviewers	[Ian Main, WP2 Lead]
Approval	[Management Board]
Status	[Final]
Dissemination level	[Public/internal]
Will the data supporting this document be made open access? (Y/N)	[Yes]
If No Open Access, provide reasons	
Delivery deadline	[31.08.2021]
Submission date	[31.08.2021]
Intranet path	[DOCUMENTS/DELIVERABLES/File Name]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821115

Table of contents

1. Post-processing existing data to improve the baseline for OEF and predictability research 3

2. Template matching 7

Summary

This deliverable summarizes the work carried out under the Task 2.4.

TASK 2.4 - Advancing observational capabilities

1. Post-processing existing data to improve the baseline for OEF and predictability research

In order to produce a higher-resolution earthquake catalog of the Italian peninsula, including homogeneous local magnitudes (ML), we started from the Italian seismicity catalogue named CLASS (1.0; Latorre et al., in prep.), consisting of absolute hypocentre locations, for events occurring in the time period between 1981 and 2018 (Figure 1). CLASS locations have been obtained by applying a non-linear inversion location method (NonLinLoc; Lomax et al., 2000) in a 3D regionally fine tomographic velocity model [from Di Stefano and Ciaccio, 2014].

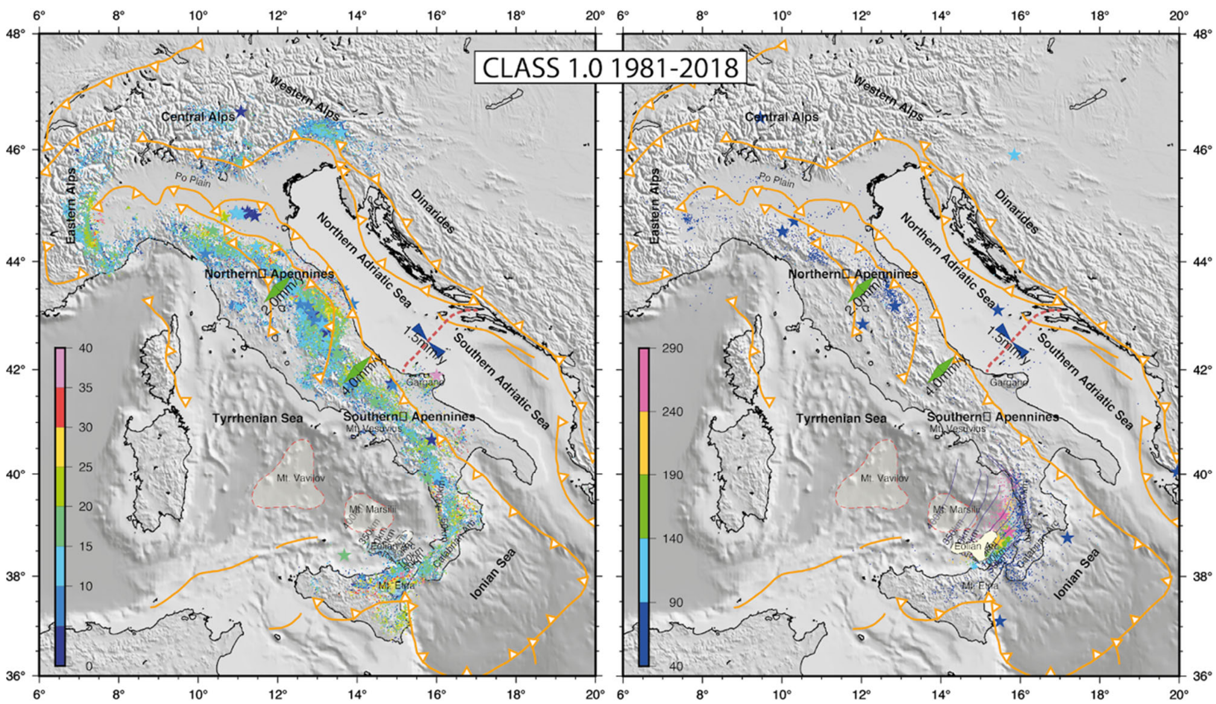

Figure 1 – Shallow (left; <40km) and deep (right; 40-200km) seismicity distribution of the Italian peninsula from CLASS 1.0, in the period 1981-2018. Colours are related to depth in kilometres' scale.

Keeping CLASS 1.0 as a reference, within RISE we are working to generate a new catalogue composed of relative earthquake locations (CARS 1.0), retrieved by using the Double Differences (DD; Waldhauser and Ellsworth, 2000) location algorithm. The dataset before 2005 will comprise only absolute arrival times from handmade pickings, while delay times from picks will be combined with cross-correlation derived relative arrival times measurements, for the subsequent period.
The CLASS 1.0 starting catalogue (Figure 1) consists of 422,557 hypocentral locations (~280k in very high-quality and robust location classes, following Michele et al., 2019), retrieved starting

from ~5M P + ~3.5M S-readings either from several merged bulletins (1981-2008) or from INGV bulletin only (2009-2018).

ZONATION AND VELOCITY MODELS

To enhance the resolution of the starting catalogue, we post-process CLASS 1.0 by re-locating all the events, always with the same location code NonLinLoc, with both a smaller scale regionalized 1D velocity models retrieved from the starting tomographic 3D model and by applying locally specific station corrections. There are many reasons for using local 1D models; first of all, these are the same models we will use in the HypoDD relocation procedure, second, the 3D model can be locally too coarse thus generating local artefacts. We show in Figure 2 an example of the improvement we gain by using a local 1D velocity model.

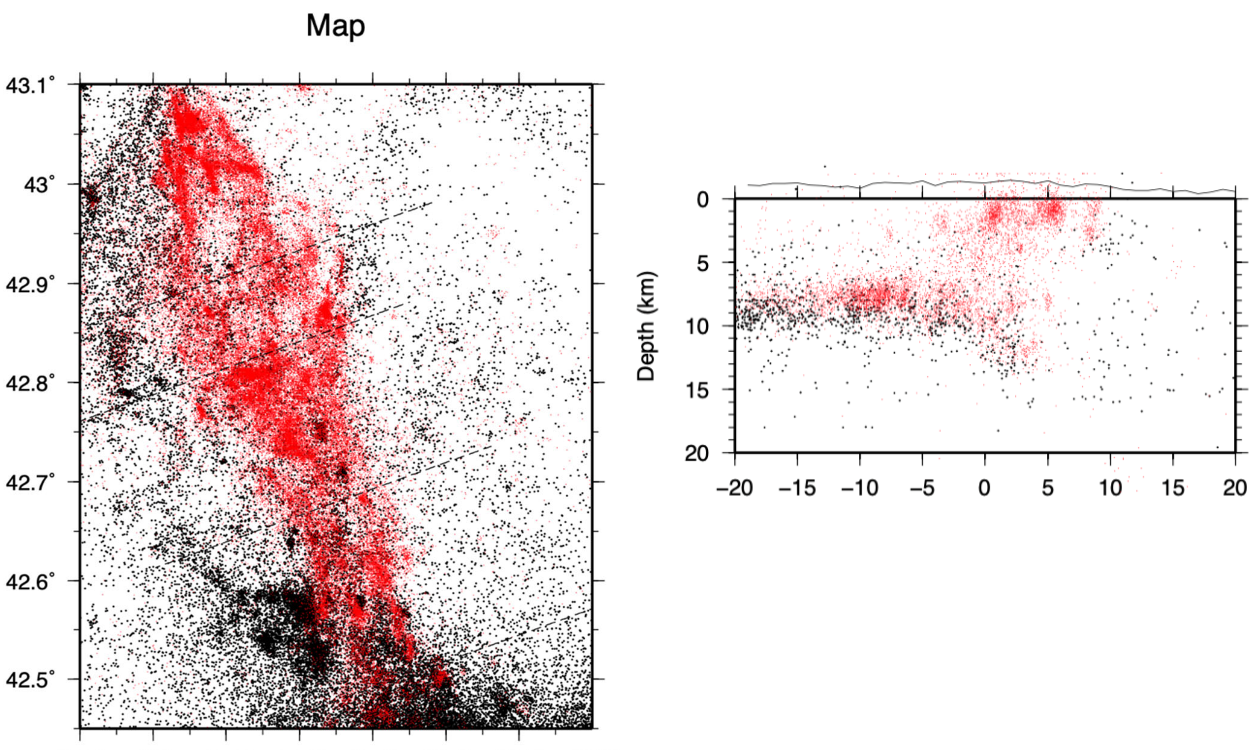

Figure 2: Map view and (northernmost) vertical section of the 2010-2016 (back) and 2016-17 (red) seismic activity of Central Italy. NLL locations have been retrieved by using a local (Zone 5b in Figure 3) 1D velocity model derived from the 3D tomographic one.

Moreover, these are the same 1D models that are going to be used in INGV monitoring system. This implies that we will have the option to use the work flow we are constructing now for relocating the seismicity coming out from the INGV monitoring system.

The regional mean 1D velocity models, extracted from the national scales 3D tomographic models (Di Stefano and Ciaccio, 2020 and Magnoni et al., 2021), are the result of an upgrading process of some of the analysis tools of the INGV monitoring system. The Italian region is divided into 18 sub-regions possessing homogeneous geological and seismotectonic characteristics (red polygons in Figure 3).

To construct the 1D local velocity modes representing the 18 zones, we use the grid nodes included in the green boxes (in Figure 3) of the tomographic models, around the centre of each area (yellow symbol). Where higher precision 1D local velocity models are, available and documented from literature, these will be preferred.

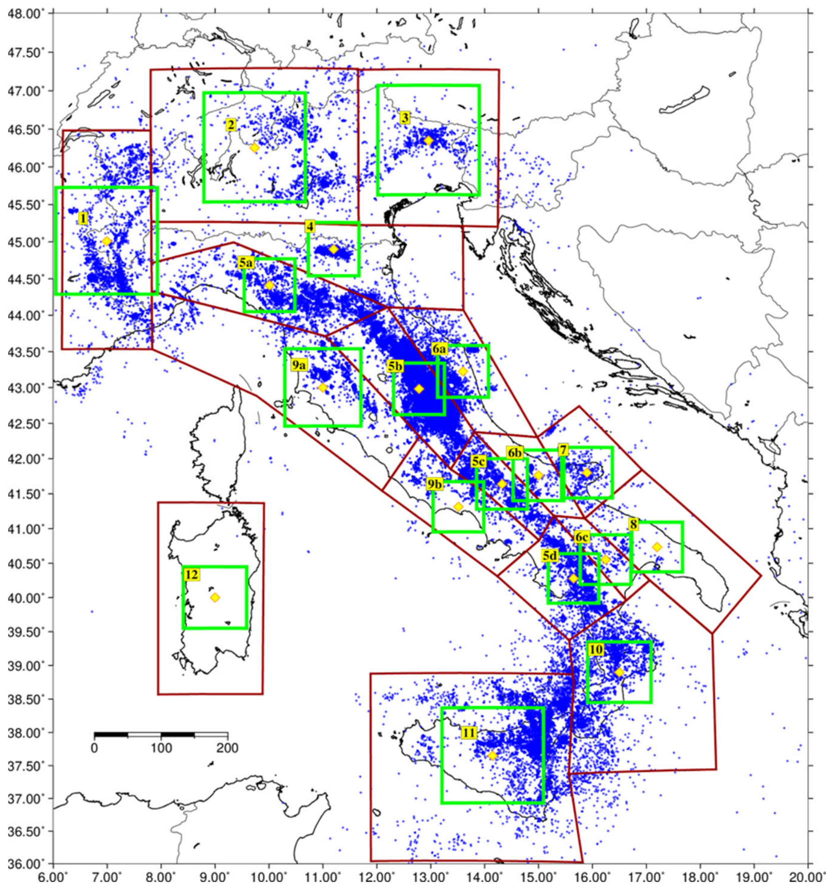

Figure 3 - Map of Italy with highlighted the 18 areas.

Finally, to mitigate the oversimplification of the 1D models, we compute specific stations corrections for all the stations in the whole area.

SEISMIC WAVEFORMS DATABASE

Based on CLASS 1.0 catalogue of events and related picks, we generated a dataset of waveform cuts designed to perform cross-correlation analysis that will allow us to include additional delay times in the double-difference relocation process. We selected only the events in the 2005–2018-

time window because the oldest were recorded at only a few stations, and not all were recorded digitally.

We included hypocentre, station location information, origin time, and P- and S- onsets in the waveforms' header to favour different applications. We also include 30s of pre-event noise allowing the possibility of a variety of subsequent signal analysis approaches, for example to enhance the signal to noise ratio.

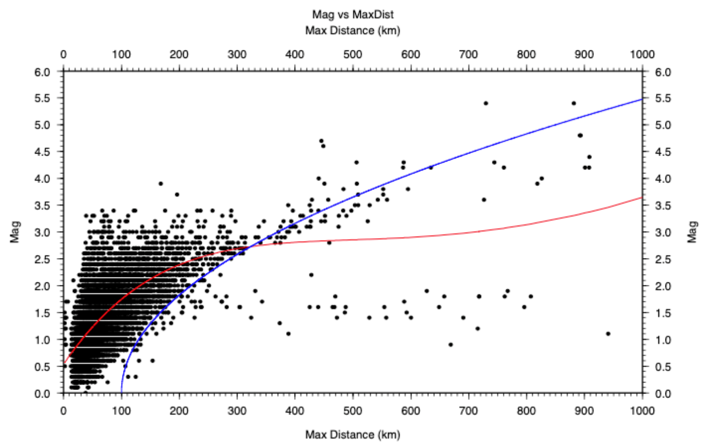

Figure 4 - Magnitude as function of distance. Red line is the not selected polynomial best fit, while the blue line is the empirical relation we used.

We use SAC as the preferred waveforms format, both for compatibility with existing correlations tools and to include all the needed information into one single file. For each event we cut signals from all the stations within a distance compatible according to an empirical function linking magnitude of the event to the distance of the most faraway station, based on bulletin data over the last 40 years. We set the maximum distance for the smaller (ML<1.0) events to 100 km (Figure 4).

AMPLITUDE RE-ESTIMATION AND ML CALCULATION

To produce a homogeneous catalogue of local magnitude (ML) for the 2005–2018-time window we are working to build a dataset of maximum amplitudes, performing the analysis on the waveform's cuts we produced for the cross-correlation analysis.

Since homogeneity starts with producing homogeneous maximum amplitudes estimations, we adopted the same deconvolution-convolution, filtering and signal peaks search across the whole dataset. While this was not always possible in the runtime production of monitoring data over the evolving analysis systems, it is possible a-posteriori. To this end we developed a specific tool to:

remove disturbing frequencies maximizing the signal to noise ratio by performing spectral analysis to define the corner frequencies of an adaptive band-pass filter;
convert filtered signals into an equivalent Wood-Anderson seismogram, used in the definition of the ML scale;
determine maximum and minimum amplitudes of the signal from the P onset to after the S onset or a synthetic S-travel time on the NS and EW components and, additionally for future regression studies to the past, on the vertical component;
store max and min peaks and related metadata (e.g., corner frequencies, time distance between min and max) in a database table.

These tables will allow us to calculate local magnitude, applying both different selection criteria on the amplitudes and different attenuation laws. In the first phase, we will apply Di Bona [2016] derived for the Italian region including station correction terms for the Italian seismic networks [Mele and Quintiliani, 2020].

ML re-estimation for strong earthquakes at close stations (within 50/100) is likely to be hampered by the effect of signal saturation, where the signal exceeds the dynamic range of the instrument, and hence might introduce high errors in the final ML value. This is also more likely to be an issue for velocimeter records, and for recordings on older instruments with a lower dynamic range. To minimise this effect, we implemented a pre-processing procedure able to automatically identify and reject the majority of the saturated waveforms, thereby minimizing the mis-identification of good waveforms. This procedure, written in python, merges the results of two approaches: 1) sorting the normalized waveform for growing amplitudes and consequent statistical analysis of the best exponential fit parameters; 2) considering the number of almost consecutive counts above and below maximum and minimum amplitude values. The procedure was first educated on a waveforms subset of manually categorized data (not saturated vs. saturated), and then positively tested on an independent set of waveforms from other M5+ earthquakes.

In the remaining months of the project, once we have finalized the stations correction computation and the cross-correlation measurements (for the most recent events), we will be ready to relatively relocate the entire catalogue. At the same time, once we have finished the creation of the maximum amplitude database, we will compute the local magnitude dataset.

2. Template matching

Template matching was run to search 8 years (2009-2016) of continuous data exploiting 48,000 well-located earthquakes in Central Italy, provided by Task 2.4.1. Codes are rewritten to improve the performance and scalability. Scientific results will be provided within the next 3 months and will concern a background seismicity analysis and clustering coalescence prior to the 2016-2017 sequence in Central Italy.

The technical improvements needed to address the massive computations involved include: a) performance: ≈200% speedup in single-threaded mode, linear speedup using multiple threads. GPU support with further performance improvements: 50 templates per second per node with 4 GPU (NVIDIA V100), and higher speedups possible using longer signals. Faster post-processing thanks to AVRO data serialization; b) algorithm development: new detections based on maximum cross-correlation, irrespective of time lag between channels and Median Absolute Deviation (MAD) of the stacked cross-correlation: the maximum cross-correlation is precomputed at all times and a robust threshold on cross-correlation is used. Filtering of detections is based on minimum distance between peaks of cross-correlation (spurious/duplicate events); c) usability: CLI, logging, input & output handling (scan all file system for valid inputs, skip missing/incomplete data, output single structured binary output with optional compression), arbitrary signal length, template duration (per channel basis), and sampling rate; d) robustness & correctness: better error handling, among fixed bugs: negative normalization in Obspy cross-correlation routine, drop multiple detections within template length, more stable magnitude estimation (missing data, template data and other bug fixes), detections at beginning or end of the signal, full usage of data (template/signal traces matching before processing); e) maintainability: less code (-80%) and dependencies, enhanced readability (refactoring into functions, meaningful variable nomenclature), modularity, easier deployment via registered PyPI package.

Figure 5 shows new detections in time for each template used before the 2016-2017 sequence (first 23500 templates).

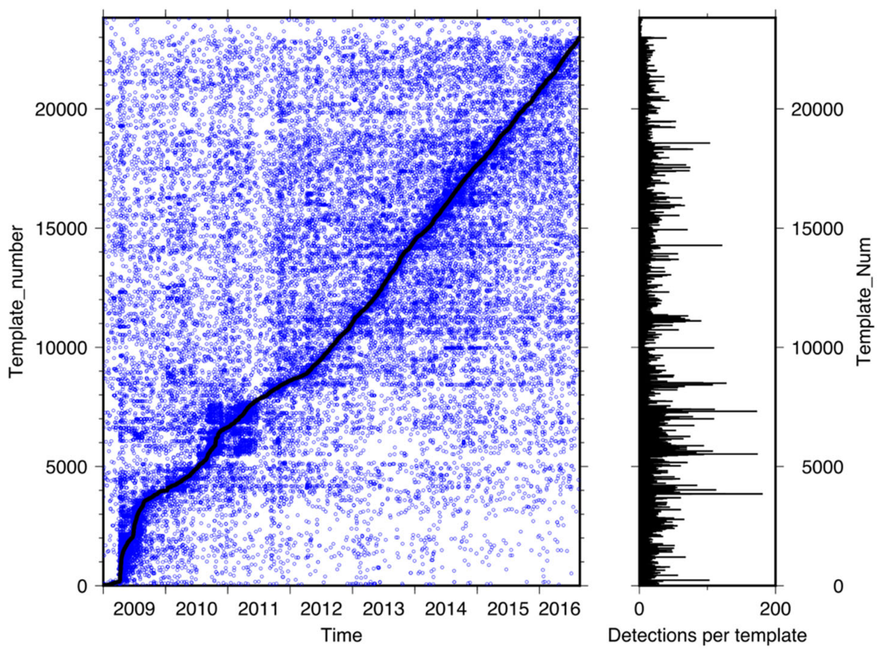

Figure 5 – 2009-2016 detections obtained by template matching using ~ 24000 templates before the 2016-2017 Central Italy seismic sequence (black dots on the left). The template number is sorted over time and new detections increase close to the templates.

In the same long-time frame, other detections are found by using ML>1.5 aftershocks of the 2016-2017 seismic sequence (Figure 6). Preliminary results confirm seismicity rate variations due to tectonic loading of the segmented sub-horizontal shear zone (e.g. Vuan et al., 2017) prior to the 2016-2017 Central Italy sequence.

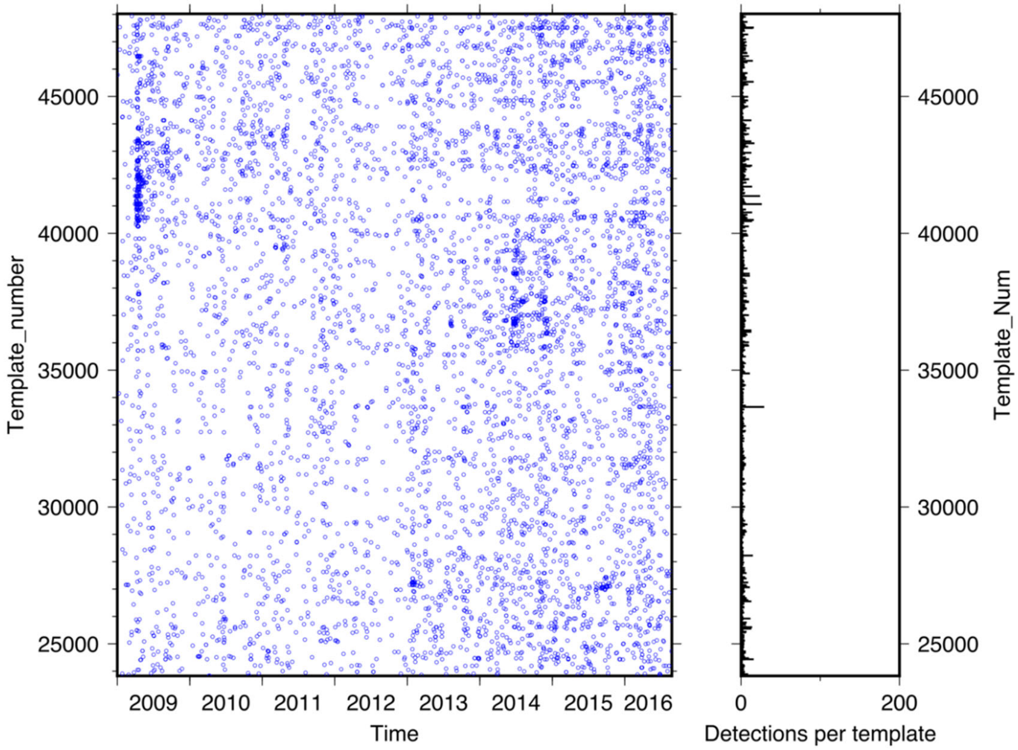

Figure 6 – 2009-2016 detections obtained by template matching using ~ 24000 aftershocks (M>1.5) of the 2016-2017 Central Italy seismic sequence. The template number is sorted over time. The new events are significantly lower than those shown in Figure 1.

REFERENCES

Di Bona M., (2016). A Local Magnitude Scale for Crustal Earthquakes in Italy. Bull. Seismol. Soc. Am., 106 (1), pp. 242–258. doi: 10.1785/0120150155.

Di Stefano R. and M. G. Ciaccio (2020). Seismic Velocity Model of P- and S-waves for the Italian Lithosphere (Version 1.0) [Data set]. http://doi.org/10.13127/tomorama.1.

Latorre, D., Di Stefano, R., Castello, S., Michele, M., and L. Chiaraluce. A complete catalog of 3D probabilistic hypocenter locations of the Italian seismicity recorded from 1981 to 2018. In prep.

Lomax, A., J. Virieux, P. Volant, and C. Berge-Thierry (2000). Probabilistic earthquake location in 3D
and layered models: introduction of a Metropolis–Gibbs method and comparison with linear locations. In: Advances in seismic event location, ed. C. H. Thurber and N. Rabinowitz, 101–134. Dordrecht and Boston: Kluwer Academic Publishers.

Magnoni, F., Casarotti, E., Komatitsch, D., Di Stefano, R., Ciaccio, M.G., Tape, C., Melini, D., Michelini, A., Piersanti, A., Tromp, J. Adjoint Tomography of the Italian Lithosphere. 2021, submitted.

Mele F.M. e Quintiliani M., (2020). Correzioni di stazione per il calcolo della Magnitudo Locale in Italia (aggiornate alla configurazione della Rete Sismica Nazionale di giugno 2018). Rapp. Tec. INGV, 412: 1-60.

Michele, M., Latorre, D., Emolo, A. (2019). An Empirical Formula to Classify the Quality of Earthquake
Locations. Bulletin of the Seismological Society of America. Vol. 109, No. 6, pp. 2755–2761, December 2019, doi: 10.1785/0120190144.

Vuan, A., Sugan, M., Chiaraluce, L., & Di Stefano, R. (2017). Loading rate variations along a midcrustal shear zone preceding the Mw6.0 earthquake of 24 August 2016 in Central Italy. Geophysical Research Letters, 44, 12,170– 12,180. https://doi.org/10.1002/2017GL076223.