

JGR Solid Earth

RESEARCH ARTICLE

10.1029/2021JB023405

Special Section:

Machine learning for Solid Earth observation, modeling and understanding

Comparison of Deep Learning Techniques for the Investigation of a Seismic Sequence: An Application to the 2019, Mw 4.5 Mugello (Italy) Earthquake

S. Cianetti¹, R. Bruni^{2,1}, S. Gaviano^{3,1}, D. Keir^{3,4}, D. Piccinini¹, G. Saccorotti¹, and C. Giunchi¹

¹Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Pisa, Italy, ²Dipartimento di Scienze della Terra, Università degli Studi di Pisa, Pisa, Italy, ³Dipartimento di Scienze della Terra, Università degli Studi di Firenze, Florence, Italy, ⁴School of Ocean and Earth Science, University of Southampton, Southampton, UK

Key Points:

- We analyze the Mw 4.5 Mugello (Italy) earthquakes sequence, December 2019, using deep learning algorithms, and compare with bulletin data
- Deep learning performances on detections and phase picking works very well even without ad-hoc retraining
- Deep learning based earthquake catalogs show differences due to different algorithm designs

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Cianetti and D. Keir,
spina.cianetti@ingv.it;
derekboswell.keir@unifi.it

Citation:

Cianetti, S., Bruni, R., Gaviano, S., Keir, D., Piccinini, D., Saccorotti, G., & Giunchi, C. (2021). Comparison of deep learning techniques for the investigation of a seismic sequence: An application to the 2019, Mw 4.5 Mugello (Italy) earthquake. *Journal of Geophysical Research: Solid Earth*, 126, e2021JB023405. <https://doi.org/10.1029/2021JB023405>

Received 12 OCT 2021

Accepted 13 DEC 2021

Author Contributions:

Conceptualization: S. Cianetti, D. Keir, C. Giunchi

Data curation: S. Cianetti, R. Bruni, S. Gaviano, D. Piccinini, C. Giunchi

Funding acquisition: D. Keir

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract The increase of available seismic data prompts the need for automatic processing procedures to fully exploit them. A good example is aftershock sequences recorded by temporary seismic networks, whose thorough analysis is challenging because of the high seismicity rate and station density. Here, we test the performance of two recent Deep Learning algorithms, the Generalized Phase Detection and Earthquake Transformer, for automatic seismic phases identification. We use data from the December 2019 Mugello basin (Northern Apennines, Italy) swarm, recorded on 13 permanent and nine temporary stations, applying these automatic procedures under different network configurations. As a benchmark, we use a catalog of 279 manually repicked earthquakes reported by the Italian National Seismic Network. Due to the ability of deep learning techniques to identify earthquakes under poor signal-to-noise-ratio (SNR) conditions, we obtain: (a) a factor 3 increase in the number of locations with respect to INGV bulletin and (b) a factor 4 increase when stations from the temporary network are added. Comparison between deep learning and manually picked arrival times shows a mean difference of 0.02–0.04 s and a variance in the range 0.02–0.07 s. The improvement in magnitude completeness is ~0.5 units. The deep learning algorithms were originally trained using data sets from different regions of the world: our results indicate that these can be successfully applied in our case, without any significant modification. Deep learning algorithms are efficient and accurate tools for data reprocessing in order to better understand the space-time evolution of earthquake sequences.

Plain Language Summary We focus on the testing and comparison of recently published frameworks for seismic events and seismic waves detection based on deep learning. As an example we use data from an earthquake swarm that occurred in December 2019 in the Mugello basin (N of Florence, Italy). Earthquake swarms analysis is particularly challenging because of the high seismic rates, with events closely occurring in time and space. The results of the automatic analysis are compared with results reviewed by human analysts. The deep learning algorithms used in this application were originally trained on data collected from different regions of the world: Our results indicate their capability to apply the knowledge gained in different contexts to totally unseen data. However, some differences, due to the different design of the algorithms and to training data sets, arise and suggest that careful tuning and benchmarking are required to achieve stable and reliable results in seismic data analysis workflows.

1. Introduction

Over the last few decades, the techniques used to automatically detect and locate earthquakes have dramatically evolved with the improvement of computing power. Such computer-based approaches to earthquake detection are gradually replacing traditional methods of manual and visual inspection of the seismograms, and have become increasingly reliable through time.

Amongst the first implemented and still most widely used techniques to automatically pick seismograms is comparing the short term average (STA) to a longer term average (LTA) of the seismic signal (or of a characteristic function of it) in sliding windows through the continuous data stream to pick the arrival of a seismic phase (Allen, 1978, 1982; Baer & Kradolfer, 1987). Another class of methods are those based on autoregressive models, which go beyond the bare analysis of the time and frequency domain variations of the seismic signal and instead try to fit the statistical properties of noise and signal segments of the waveforms. An example of such an approach

Investigation: D. Piccinini, G. Saccorotti
Methodology: S. Cianetti, S. Gaviano, C. Giunchi
Software: S. Cianetti, S. Gaviano
Supervision: D. Keir, G. Saccorotti, C. Giunchi
Validation: D. Piccinini
Visualization: S. Cianetti, R. Bruni, S. Gaviano
Writing – original draft: S. Cianetti, G. Saccorotti, C. Giunchi
Writing – review & editing: S. Cianetti, D. Keir, D. Piccinini, G. Saccorotti, C. Giunchi

is the AIC autoregressive picker proposed by Sleeman and van Eck (1999). The application of these procedures in general provide better accuracy with respect to the simple STA/LTA criteria. However, they are only applied after an arrival has been recognized (usually with an STA/LTA based method).

The abrupt changes in the seismic signal due to arrivals of earthquake waves can also be characterized using variance and higher order statistics parameters such as skewness and kurtosis (Ross & Ben-Zion, 2014; Saragiotis et al., 2002), which further increase sensitivity with respect to the aforementioned techniques. All the previous approaches, however, suffer from two fundamental drawbacks. The first is that they require a non trivial and accurate tuning of the parameters in order to provide reliable results. As an example, the implementation of the ubiquitous Allen Picker (Allen, 1978) in Earthworm (Johnson et al., 1995), a widespread tool to process real time data, requires 18 parameters to be defined and most of them are highly affected by the type of sensor used, the geometry of the network, the noise level, and other factors (Mele et al., 2010). The second reason is that, being based on energy variations within the time series, it may be non trivial to isolate the earthquake signals from other disturbances. On the other hand, these techniques provide valuable tools for general investigations of time series anomalies.

An alternative approach is inherited from controlled-source seismic imaging, in which some characteristic function parameterizing the seismic time series at the different stations is back-propagated through the propagation medium. The optimal subsurface location and origin time of the earthquake are then given by the space-time coordinates at which all the back-propagated wavefronts interfere constructively (e.g., Drew et al., 2013; Grigoli et al., 2014; Kao & Shan, 2004; Langet et al., 2014). This class of techniques has an advantage when applied to noisy data sets since the direct stacking process enhances the SNR filtering out incoherences in the time series.

Algorithms based on the similarity of waveforms between similarly located earthquakes such as the matched filter (e.g., Peng & Zhao, 2009) partially overcome the limitations of the previously described picking approaches. Indeed, the reprocessing of continuous data streams using cataloged events as a template may improve the detection of P and S waves arrivals by up to an order of magnitude (e.g., Allen, 1982; Ross et al., 2018). The drawback of these methods is that the new found events are duplicates of the original templates. This means that other seismic signals strongly different from the templates still remain undetected, resulting in the new catalogs potentially still being incomplete. The recent introduction of the FAST algorithm (Yoon et al., 2015) has minimized this bias by demonstrating the ability to match the general characteristics between matched waveforms with respect to the specific shape of the template waveform. Additional drawback of the template matching is the potential spatial bias in detection ability which depends on the density of events, and also the character of events, with some regions more prone to have similar events than others.

In recent years there has been a rapid growth in the application of Machine Learning and Deep Learning (DL) techniques to solid Earth geoscience including seismology (Bergen et al., 2019; Kong et al., 2019). Due to the large data volumes and many ubiquitous properties of seismic waveforms independent of setting, seismic phase detections have been an attractive target for DL applications (García et al., 2016; Mousavi et al., 2020; Perol et al., 2018; Ross et al., 2018; Titos et al., 2019; Woollam et al., 2019; Zhu & Beroza, 2018). The extensive work already done by analysts for seismic catalog compilation can be easily converted into high quality labeled data sets suited for the setup of supervised DL methods. The labels are used to train the algorithm to learn relevant features in the seismograms, provided that a sufficient amount of data is available. Once this process is complete, the trained DL model is able to map new and unseen seismic waveforms into target representations that, in the case of phase detection, encode the P or S wave arrivals with a continuous function peaked at the arrival time. Among the various flavors of DL, convolutional neural networks (CNN) and attention-based transformer network have been successfully applied to event detection and seismic phase recognition (Mousavi et al., 2020; Perol et al., 2018; Ross et al., 2018). The picking task is accomplished by a succession of layers, performing first feature extraction and then classification of the targets. The power of DL methods is that the features upon which the target recognition is based are learned implicitly. This reflects the agnostic character of DL as a data driven discipline. In real world applications this results in the possibility of feeding data with minimal pre-processing, which results in very efficient analysis pipelines from the computational point of view. Further, providing the full information included in the waveforms minimizes potential biases due to a priori selections of particular features. The key role in the success of the learning process is played by the quality, completeness and size of the training data set to ensure the network is learning correctly and without any bias. In seismology this task is particularly facilitated by the fact that a fair amount of information is made publicly available by data providers (e.g., IRIS,

FDSN, Ingate, 2008) and also by recent publications of extensive data sets ready for DL applications such as STEAD (Mousavi et al., 2019), LEN-DB (Magrini et al., 2020) and INSTANCE (Michellini et al., 2021).

Among the different efforts in the application of DL to earthquake detection, ConvNetQuake (Perol et al., 2018) represents the first CNN for phase detection and earthquake location. This network classifies the input data into different clusters as a function of earthquake location. Woollam et al. (2019) investigated the performances of that approach to classify seismic phases, when only a relatively small set of events is available for training the network. More recently, Ross et al. (2018) developed Generalized Phase Detection (GPD), a CNN algorithm to identify seismic phases, while Mousavi et al. (2020) designed Earthquake Transformer, an attention-based model to detect events and picks. Both these frameworks take as input 3 components of seismic waveform recordings and return time series of forecast probabilities for P, S waves and noise for GPD or event detection, P and S waves for EQT.

In this paper we select a seismic data set of earthquakes, recorded by a dense network of 22 seismic stations, associated with the Mw 4.5, 9 December 2019 mainshock in the Mugello basin (Northern Apennines, Italy). According to the INGV bulletin (Pagliuca et al., 2020) the sequence consists of about 280 events occurring between 8 and 26 December 2019. However, it is likely that a large number of low magnitude events, preceding and following the mainshock, and potentially overlapping, may have been missed by a standard analysis for seismic monitoring purposes. This data set can be considered a valuable and challenging test case of automatic event detection and seismic phase recognition because of the difficulty to separate earthquakes occurring close in time during an intense aftershocks sequence. Further, a catalog of manually revised seismic phases and earthquake locations is available making possible to compare with advanced phase detection such as those based on DL. Our goal is to analyze this data set, extended in time from 1 to 31 December 2019, using template matching and DL frameworks such as GPD and EQT. We provide insights about the performance of the different algorithms through the study of the resulting catalogs, earthquake statistics, locations, and pick quality.

2. Location, Instruments, and Data

We study an earthquake sequence that occurred in the Mugello Basin (Figure 1a), located ~30 km north of Florence. This is a WNW-ESE striking intra-mountain basin that is 25 km long and 15 km wide, and it is part of the Northern Apennines, a NW-SE striking fold-and-thrust belt composed of a pile of NE-verging tectonic units that developed during Cenozoic collision between the European plate (Corso–Sardinian block) and the Adria plate (e.g., Carminati et al., 2012). Seismicity of the Northern Apennines is mostly concentrated within a ~40 km wide, NW striking swath along the mountain chain. Earthquakes are generally shallow (<20 km deep), and exhibit normal-fault focal mechanisms on the SW side of the chain (e.g., Piccinini et al., 2014). Larger focal depths (up to 40 km) and dominant thrust-faulting mechanisms are instead observed on the NE side of the chain (Piccinini et al., 2014). This spatial change in seismicity pattern across the Apennines is one of the primary pieces of evidence for the current spatial transition in tectonic style. The Mugello basin has experienced recent NW-striking seismic swarms in 2008 and 2009, near the western margin of the basin (Bonini et al., 2016; Piccinini et al., 2014). The most significant historical earthquakes occurred in 1542 and 1919, with macroseismic magnitudes of $M = 5.9$ and $M = 6.3$, respectively (Rovida et al., 2016).

As is the case for the entirety of the Italian territory, seismicity of the study area is monitored through the Italian National Seismic Network (INSN), a permanent infrastructure operated by the Istituto Nazionale di Geofisica e Vulcanologia (INGV hereinafter (INGV Seismological Data Centre, 1997)). In the summer of 2019, nine additional temporary broadband stations were added as part of a collaboration between INGV, the Universities of Southampton (UK) and Florence (IT), and the Italian National Research Council (CNR). These stations, on lease from the SEIS UK facility, were deployed in order to achieve improved coverage of the Mugello basin and adjoining areas (Figure 1a), with the ultimate goal of accurately identifying faults from seismicity and doing high resolution seismic imaging of the crust in this portion of the Northern Apennines. Three of these stations (RINC, RONT and VISG) became part of the INSN (IV network) shortly after their installation, while the remaining have been registered as seismic network 9M (Bruni et al., 2019). A full list of the stations is available in Table S1 of Supporting Information S1.

This new deployment recorded the Mw = 4.5 earthquake that struck the NW margin of the basin on 9 December 2019 (Figure 1b). The mainshock and its aftershocks were clustered along a previously identified

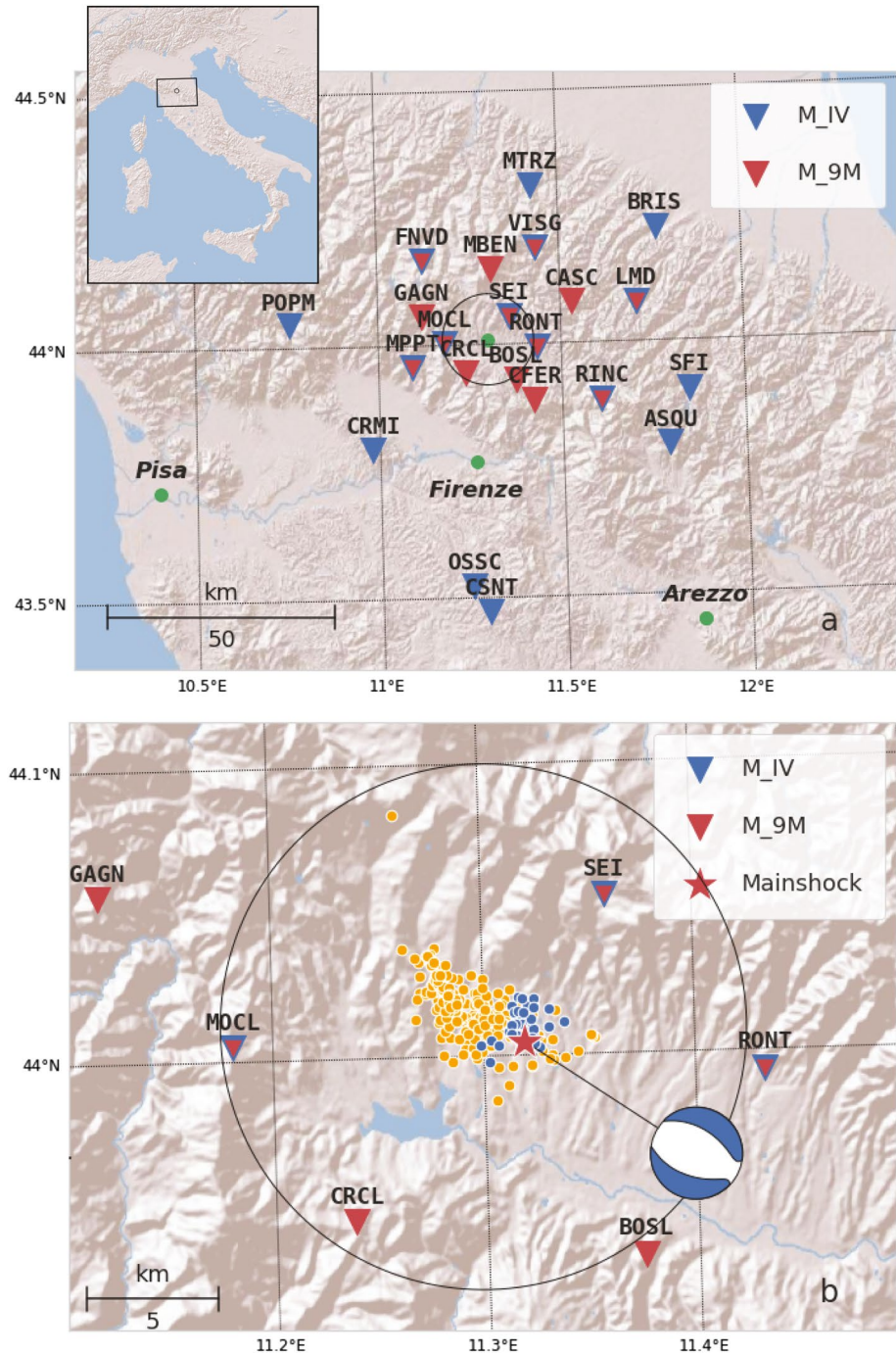


Figure 1. (a) Seismic stations belonging to the INSN permanent (IV) and to the additional temporary network (9M). Blue and red colors denote which stations are being used in catalogs M_IV and M_9M, respectively. See text for details. (b) Mw = 4.5, 9 December 2019 mainshock focal mechanism (<http://cnt.rm.ingv.it/event/23558121>, Scognamiglio et al. [2006]) and associated seismicity from M_IV catalog (1–31 December 2019). Blue and orange dots show foreshocks and aftershocks, respectively. The 10 km radius circle delimits the area of interest of the earthquake sequence.

WNW-ESE-striking lineament (ISIDe Working Group, 2007; Pagliuca et al., 2020), and the focal mechanism of the mainshock suggests a normal fault striking $\sim N105^\circ E$ (<http://cnt.rm.ingv.it/event/23558121>, Scognamiglio et al. [2006]). Since the deployment of the temporary seismic network preceded the occurrence of the mainshock we are able to study this seismic sequence for its entire development including some foreshocks occurred few days before the main event. The mainshock prompted unexpectedly large accelerations (PGA ~ 0.24 g at ~ 7 km

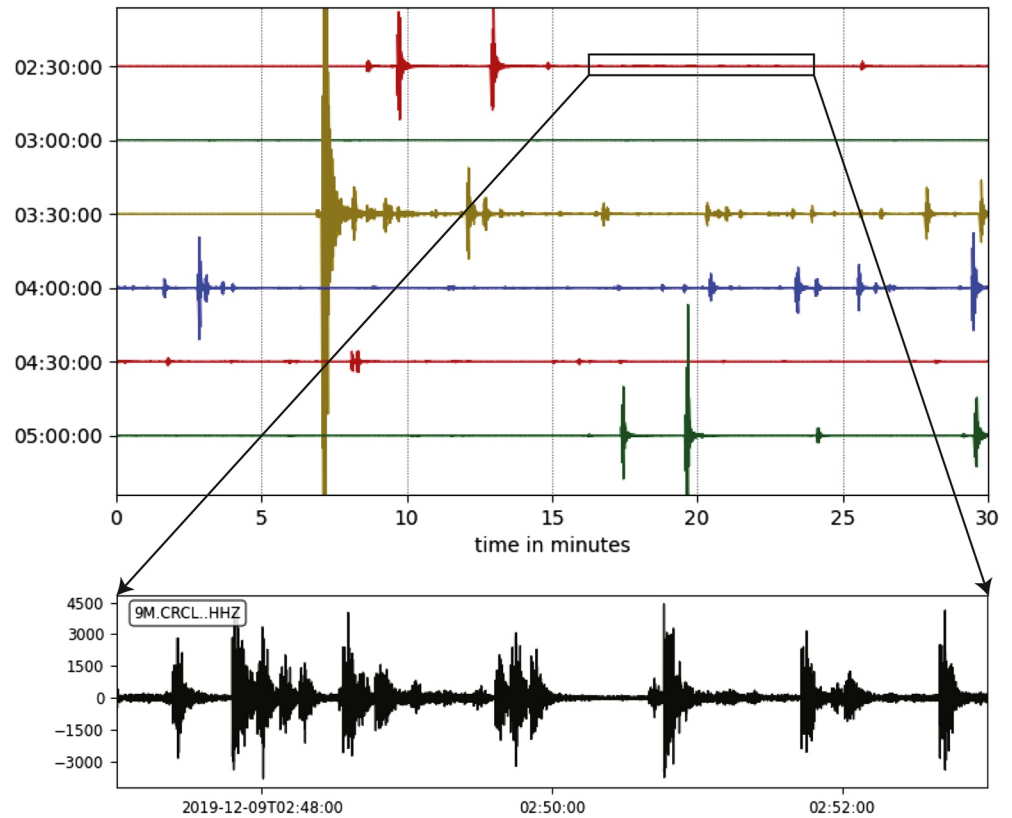


Figure 2. Example of 2.5 hr of recordings before and after the mainshock at 9M.CRCL, vertical channel. The inset shows a zoom of the foreshocks sequence.

epicentral range), causing damage that resulted in the evacuation of more than 150 residents and economic losses of several million. The whole sequence lasted less than one month. During 1–31 December 2019, the INGV catalog (ISIDe Working Group, 2007) reports 279 earthquakes, with local magnitudes (M_L) in between 0.5 and 4.5. We will refer to this catalog as M_IV (Figure 1b).

For these events we proceeded with manual repicking of P and S wave arrival times (to achieve uniformity in picking made by a single operator) at all the stations from the temporary network and a selected subset of the closest stations from INSN (namely, stations SEI, LMD, MOCL, MPPT and FNVD in Figure 1a). This new set of arrival times was then inverted for source location using the NonLinLoc software package (Lomax et al., 2000, 2014) with an improved velocity model specifically derived for the area (Piccinini et al., 2014). Local magnitudes (M_L) of earthquakes are finally computed following Di Bona (2016) and Mele and Quintiliani (2020), who provide parameters and station correction terms for INSN. Details about M_L computations are described in Supporting Information S1. Throughout the following, we refer to this catalog as M_9M.

As generally occurs during the early stages of aftershock sequences, earthquakes repeat at tight temporal intervals and often overlap each other (Figure 2). Under such conditions the classical automatic discrimination procedures such the STA/LTA algorithm (Allen, 1982), and the subsequent phase picking and association, generally fail and a significant portion of events may pass undetected or not correctly associated and located. For this reason the waveform data collected during December 2019 are here processed with template matching and DL based algorithms to explore the efficiency of these methods in finding new events and to better characterize the sequence.

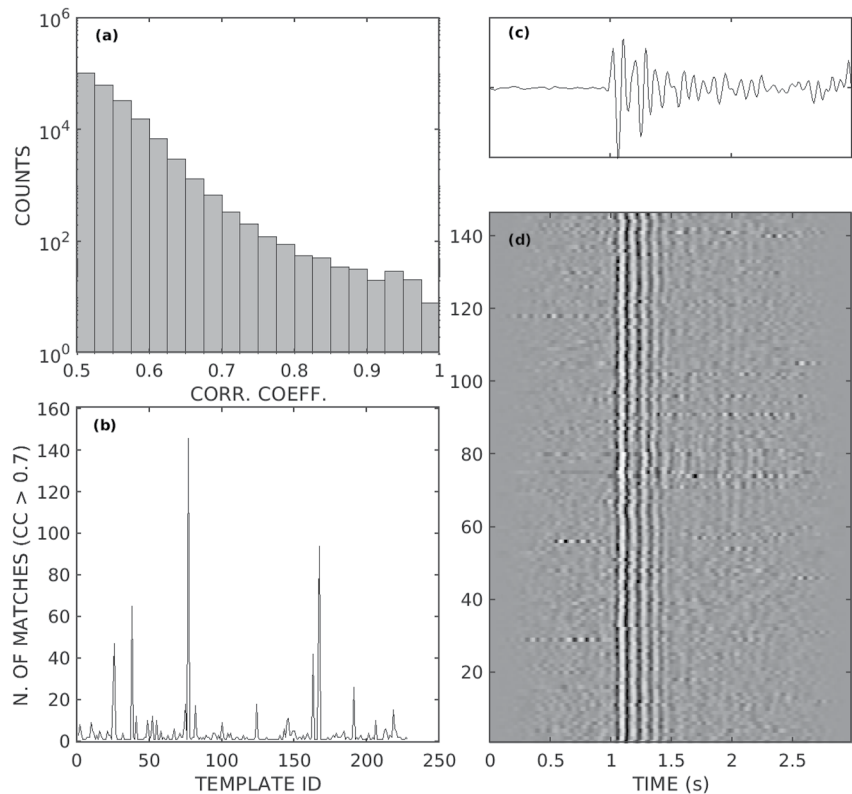


Figure 3. (a) Histogram count of the number of P-wave matches at station RONT, as a function of CC. (b) The number of matches with $CC \geq 0.7$ for each template at the same station. (c) P-wave template wavelet having the largest number of associated detections; (d) the corresponding matched waveforms.

3. Methods

3.1. Matched Filter

Matched filter is an earthquake detection method based on waveform cross-correlation for identifying replicas of a known signal in continuous time series data (e.g., Gibbons & Ringdal, 2006). The method takes advantage of the high waveform similarity observed within seismic sequences (Geller & Mueller, 1980), when seismic waves originating from sources of similar location and mechanism travel a similar path resulting in similar waveform shapes at a receiver (Herrmann et al., 2019). This technique facilitates detection of new events characterized by low SNR yielding a better sensitivity than that offered by the common detection techniques such as STA/LTA (e.g., Gibbons & Ringdal, 2006; Schaff, 2008).

We use up to 278 earthquakes reported in catalog M_9M during 1–31 December 2019 as template earthquakes and performed a detection procedure on the continuous seismic data at six stations that are close to the epicentral area and have the most continuous records (namely, stations CRCL, GAGN, MBEN, MOCL, RONT, and SEI; see Table S1 in Supporting Information S1). The exact number of template events does vary (227–278) between stations, since not all the events were picked at all the stations. The matched filter procedure is conducted separately for P- and S-waves. P-waves are scanned over the continuous vertical-component recordings, using 3 s long template waveforms starting 1 s before the picked arrival time. For S-waves, the procedure is iterated over the N and E components, using 5 s long template waveforms starting 1 s before the picked arrival time. Both templates and continuous data streams are band-pass filtered within the [2–15] Hz frequency band. In a first stage, we retained all the matching wavelets having maximum cross-correlation (CC) larger than 0.5. Then, we removed duplicate matches derived from different templates with similar waveforms, keeping only the template-match pair with the highest CC. At all the stations, the number of matches decreases with increasing CC; the rate of decrease, however, is not constant, exhibiting an evident change of slope at $CC \approx 0.7$ (see example in Figure 3a). Hence, we choose $CC = 0.7$ as a threshold for selecting the matched phases to be subsequently used for the location

procedure. A sample P-wave template and corresponding matching phases are shown in Figure 3, panels (b and c), respectively.

P- and S-waves detected via matched filter are associated by first sorting in ascending order the entire set of arrival times, then declaring an event whenever the time difference between two consecutive times is larger than 5 s; such simple procedure is justified by the narrowness of the epicentral region, and the restricted time interval spanned by the travel times to the different stations. The set of associated arrival times is finally located and M_L are computed using the same procedure described for the manual catalog M_9M. The resulting catalog will be referred in the following as M_MATCH.

3.2. Deep Learning Analysis

The waveform data from 22 stations from the IV and 9M networks (Figure 1a) are processed in the time domain using GPD and EQT. Here we use these frameworks as published and the model weights as provided in supplements or data publications without any further training or modification, apart from evaluating the detection threshold values most suitable for our data set. These algorithms have been selected for their high sensitivity in detecting P and S phases with low SNRs, for example, when arrivals overlap the coda of previous events, or in case of small magnitude earthquakes, as usually occurs during seismic sequences.

GPD (Ross et al., 2018) is based on a CNN trained on 4.5 million three-component, hand-labeled seismic phases from southern California earthquakes (for a detailed description of the data set, see Ross et al. [2018]). The algorithm requires as input 4 s segments of the three-components ground velocities, which are preprocessed by removing the mean and linear trend, resampled at 100 Hz and band-pass filtered within the [3–20] Hz frequency interval. The sliding time windows are overlapped by 97.5%, corresponding to a step width of 0.1 s. The method returns three time series corresponding to the forecast probability associated with the detection of noise, P and S wave picks. The picks are declared if the forecast probability is greater than a user-defined threshold. The chosen value for the threshold is 0.99. This was determined by studying (a) how the total number of picks changes with different threshold values and (b) the forecast probabilities values for the GPD picks which have a match in the manual catalog. We came up with the value of 0.99 which represents the best compromise to retrieving more than 90% of manual catalog picks. In Supporting Information S1 we describe how we select this value in the GPD threshold choice section. The picking error is formally set to 0.1 s for all the detections, resulting from the step width for the sliding window.

EQT (Mousavi et al., 2020) is the second algorithm we selected for the analysis. It consists of an event detector and picker based on a hierarchical attention mechanism. The neural network has a multi-task structure consisting of one very-deep encoder and three separate decoders mapping the output into three time series of forecast probabilities associated with the earthquake detection, a P phase, and an S phase (Mousavi et al., 2020). First, it performs earthquake detection in a 60 s window of three component seismic data. If an event is flagged, P and S phases are recognized within the analyzed window. The network has been trained on a subset of STEAD, a large-scale global data set of 1.2 million hand-labeled waveforms of earthquakes and noise (Mousavi et al., 2019). Events, P and S wave picks are declared upon the exceedance of predefined thresholds, that we fixed to 0.2 for event detection and 0.1 for both P and S waves picking, respectively. In Supporting Information S1 we describe how we select these values in the EQT threshold choice section. For each available station, waveform data are pre-processed by filling gaps, removing mean and linear trend, resampling to 100 Hz and band-pass filtering within the [1–45] Hz frequency band. Data are then fed to EQT in blocks of 60 s with 30% overlap. EQT also provides picking error estimates based on a combination of the forecast probability and the model uncertainty evaluation, accomplished by a procedure equivalent to a Bayesian sampling of the posterior parameters distribution. This uncertainty translates into weights associated with the P and S phases for the subsequent location procedure.

The picks retrieved using the DL techniques are associated into four catalogs, two for GPD and two for EQT, according to the seismic stations defined for catalogs M_IV and M_9M (Figure 1). For both GPD and EQT results, the association of the seismic phases is performed according to a coincidence criterion, requiring that at least three P wave picks are detected within an 8 s long time interval and considering a 2 s trigger-off extension to lengthen the research window in the coincidence sum computation. This window is long enough to include the picks from the furthest station (CSNT ~60 km away from the seismic swarm). Once the association of P phases has been performed, we search at all the available stations for the S picks in the time window ranging from the

Table 1

For the Stations Used in Matched Filter Computation We Report the Total (Indicated as T) Number of Picks (Sum of Templates and Matches With $CC \geq 0.7$), the Number of Arrivals (as A) and Percentages of Picks Retrieved as Arrivals (as % A/T), for Both P and S Phases, for Matched Filter (MATCH), GPD, and EQT, Respectively

		RONT	GAGN	CRCL	MBEN	SEI	MOCL
MATCH P	T	1,008	736	3,632	577	760	19,380
	A	534	598	352	520	333	795
	% A/T	53%	81%	10%	90%	44%	4%
MATCH S	T	1,676	1,449	880	1,243	1,319	1,457
	A	966	908	710	851	560	779
	% A/T	58%	63%	81%	68%	42%	53%
GPD P	T	2,217	1,899	2,156	370	2,194	4,041
	A	1,086	943	799	206	340	1,037
	% A/T	49%	50%	37%	56%	16%	26%
GPD S	T	9,265	13,345	4,818	6,381	5,054	5,872
	A	1,078	932	588	200	317	1,022
	% A/T	12%	7%	12%	3%	6%	17%
EQT P	T	1,001	1,051	593	554	1,475	1,184
	A	753	786	455	483	405	789
	% A/T	75%	75%	77%	87%	28%	67%
EQT S	T	1,019	1,064	586	551	1,462	1,189
	A	749	778	451	476	409	759
	% A/T	73%	73%	77%	86%	28%	64%

first P onset and a maximum duration computed as 4 times the difference between the first and the last P trigger times. Two cases may occur. If a P pick is available for the station under investigation we search the S pick closest to the P onset, otherwise we add the first available S pick for the time interval considered. If two P arrivals are detected for the same station, a second separate event is created. Despite the simplicity of this phase association method, we are confident in its effectiveness because of the limited spatial extent of the sequence and the high level of accuracy of the DL pickers (Ross et al., 2018). For this purpose, Figure S6 in Supporting Information S1 we show an example of how the picks are associated into events.

The associated P and S phases are inverted to identify the source location using the same code and velocity structure adopted for the locations of catalog M_9M. We require at least three P phases, and a total of 5 phases (P or S) to locate the earthquake. Finally we computed the M_L for the located events. From all the located events, we select those whose epicenters lie within a 10 km radius of Figure 1b. The final catalogs are named M_IV GPD, M_9M GPD, M_IV EQT, and M_9M EQT, based on whether they have been assembled using the M_IV and M_9M stations and GDP or ETQ methods, respectively.

4. Results and Discussion

4.1. Matched Filter

The detection performances of the different methods should ideally be carried out with a comparison to the ground truth, which is not established in an unexploited seismic data set. In Table 1 we show, for the six stations used to determine the matched filter catalog, the total number of picks before associating (T), the arrivals (A, i.e., picks associated and located into events within our study area) and the percentages of picks retrieved as arrivals (% A/T) in

the seismic catalogs. This is an indirect way to assess the presence of false positive picks, that is, picks that do not contribute to event location. However, this approach has some limitations since it does not allow us to take into account true positive picks belonging to very low magnitude events that cannot be associated due to low phase count or picks associated to events located outside the selected area (within 10 km radius). Table 1 summarizes and compares the results obtained for Matched Filter and DL methods. On average the percentages obtained for M_MATCH span between 40% and 90%. Two exceptions are constituted by stations CRCL and MOCL, which reported a huge number of detections, of which only 10% and 4%, respectively, are successfully associated and located. This occurrence is related to electronic disturbances and short-duration gaps in the data streams from those two stations, which provoked numerous false detections, particularly for the most impulsive P-wave arrivals. The M_MATCH catalog finally has 930 locations. Taking into account the high cross-correlation thresholding, the reduced number of used stations and their proximity to the hypocentral area, we consider this catalog to be rather robust, thus providing a lower bound on the actual number of detectable events in the sequence.

4.2. Deep Learning

The most evident feature in Table 1 is that both DL methods show fairly strong variations among the stations considered. For example, the percentage of GPD P picks turned into arrivals decreases from 56% at MBEN to only 16% at SEI and that for EQT P phase ranges between 87% at MBEN to 28% at SEI. These variations may be partially related to the large number of false detection triggered by data stream problems (as is the case for SEI) and eventually mitigated with an *ad-hoc* training of the DL models for each station: as an example at MBEN, where few picks are detected by GPD with respect to EQT. Overall, EQT provides the best performance since the arrivals percentage for both P and S phases are consistently above 65%.

Table 2 reports the number of events we detect and locate in the time interval 1–31 December 2019, the mean of the event-wise root-mean-square travel time residuals (\overline{RMS}) and its associated standard deviation (σ) for the

Table 2
Catalogs Name, Number of Events, Location Mean of the Event-wise RMS Value (\overline{RMS}) and Associated Standard Deviation (σ)

Catalog	Earthquake count	$\overline{RMS}(s)$	$\sigma (s)$
M_IV	279	0.236	0.071
M_IV GPD	833	0.088	0.051
M_IV EQT	740	0.085	0.061
M_9M	279	0.101	0.026
M_MATCH	930	0.026	0.016
M_9M GPD	1,156	0.087	0.036
M_9M EQT	983	0.080	0.053

seven catalogs discussed in the previous section. \overline{RMS} is here reported to characterize each single catalog but not for the purpose of providing a metric to compare the various catalogs since they are obtained using different sets of stations and picks. The most evident achievement attained by DL methods is the large increase in the number of detected earthquakes.

The results, after earthquake location and selection, show that the number of events in the whole sequence increases significantly: while M_IV reports 279 earthquakes, DL based detections increase by about a factor of 3 (833 for M_IV GPD and 740 for M_IV EQT). We remark, however, that INSN is designed to detect events either damaging or possibly felt by population ($M_L \geq 2.0$) and not to exploit microseismic activity. The effect of enhancing seismic station density in the proximity of the sequence can be appreciated considering the set of M_9M catalogs: the number of located events increases by a factor of about 4 (1,156 for M_9M GPD and 983 for M_9M EQT, in

agreement to that attained by M_MATCH). This is due to the denser network, providing better coverage and more detailed information needed to reveal low magnitude earthquakes.

The time evolution of the swarm (Figure 4) shows a synchronous step-like behavior in all the catalogs, corresponding to the occurrence of mainshock and following major events.

The first events of the sequence start on 8 December 2019. Before the mainshock, matched filter and DL methods detect between 50 (Figure 4a) and 100 (Figure 4b) more earthquakes than reported on the M_IV catalog. Three hours of inactivity precede the reactivation of the seismic sequence that culminates with the Mw 4.5 event at UTC 03:37:03 on 9 December 2019. 90%–95% of the earthquakes are included in five clusters occurring before 17 December 2019. The swarm slowly fades by the end of the year. Figure 4 suggests that the increased number of earthquakes in the matched filter and DL catalogs may be interpreted as small magnitude events clustered around (and possibly triggered by) the larger events already present in M_IV and M_9M.

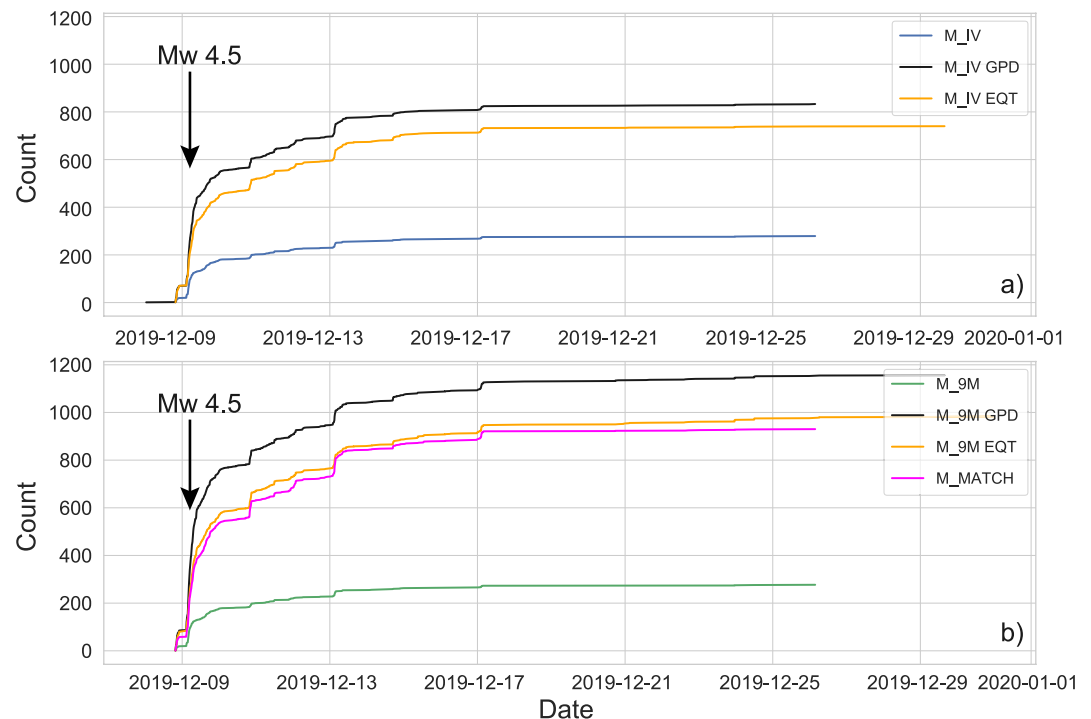


Figure 4. Cumulative number of earthquakes as a function of time for the seven catalogs analyzed. Catalogs attained using the stations set denoted as M_IV and M_9M plus M_MATCH are plotted in panels (a and b), respectively.

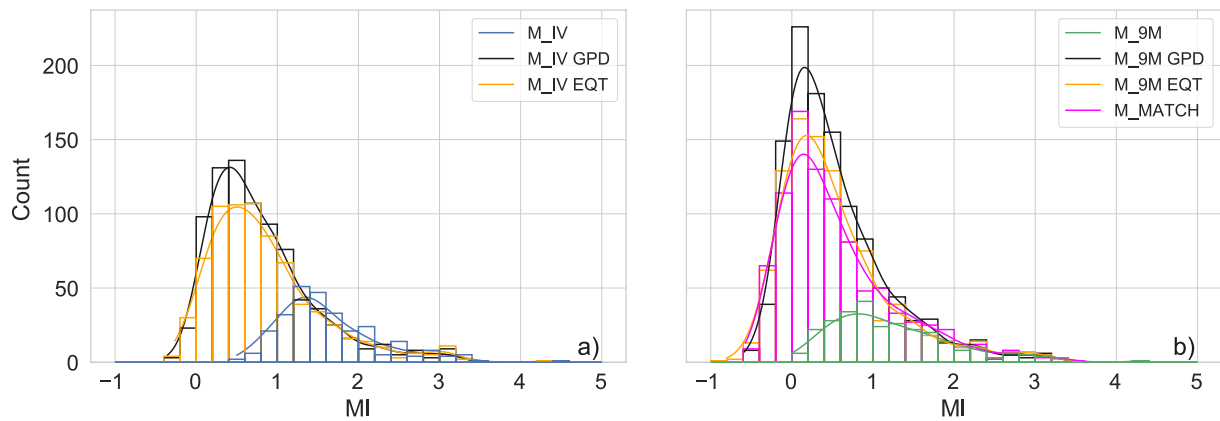


Figure 5. M_L frequency histogram for (a) M_{IV} series catalogs and (b) M_{9M} and M_{MATCH} , respectively.

Comparing the origin times of DL catalogs with those of M_{9M} catalogs we find that origin times differ by less than 0.5 s for the 95% of the events. We analyze the remaining 5% of events that have origin times that differ by more than 0.5 s. They are mainly clustered in the hours immediately following the mainshock and major bursts of the seismic sequence on 9, 13, and 17 December. A closer view of the picks for the missing events suggests that the main problems arise with earthquakes occurring close in time, at intervals comparable to the travel time, especially if the second event is more energetic than the first one. In these cases the P wave onset of the second event is masked by the S wave coda of the first event. The most common cases are the following: First, the forecast probabilities of the P wave of the second event does not rise above the chosen threshold, so the P onset is not detected (Figure S7 in Supporting Information S1). Second, the DL method correctly picks the two P phases and separates the two events (Figure S8 in Supporting Information S1). In both cases the operator incorrectly declared the S onset of the first event as the P onset of the second one. This error causes the subsequent discrepancy in the origin time computation between manual and DL methods and hypocentral location. In other cases the differences may arise when the DL methods associate a larger number of phases to an event with respect to the manual picking, resulting in a better constrained location but different from that in the M_{9M} catalog. Other factors that may also contribute to the quality of DL locations are: (a) the choice of a high probability threshold to declare a pick (0.99 for GPD and 0.1 for EQT) that may prevent some detections of true positive picks, (b) the use of pre-trained models on data collected in a different geological environment, and (c) the simple event association approach implemented. Despite these differences, the quality of DL catalogs is supported by the low values of the RMS (Table 2).

The M_L computed for all the catalogs except M_{IV} , whose values are taken from INGV bulletin, confirms (Figure 5) that the matched filter and DL techniques are very effective in detecting low magnitude earthquakes ($M_L \leq 1.5$) with respect to those included in M_{IV} and M_{9M} . Indeed the frequency histograms in Figure 5 almost overlap for $M_L \geq 1.5$ while the number of events strongly increases below that value, especially when stations from the temporary network are used. Figure 5a clearly shows that there are few differences in the distributions of the two DL and the IV catalogs especially for events with $M_L \geq 2.0$. These discrepancies are due to the larger number of stations used in the INGV bulletin which takes data from the whole INSN network while M_{IV} GPD and M_{IV} EQT are limited to the subset of stations used in our analysis (see Figure 1a and Table S1 in Supporting Information S1). This implies that the most energetic earthquakes are recorded by a larger number of stations, resulting in a slightly different hypocentral location and, more important, into a larger number of amplitude estimates obtained over a much larger distance range.

We also plot the Gutenberg-Richter distribution of the catalogs (Figure 6) as a further control of the performance of the DL techniques. Again, the DL methods confirm their effectiveness in detection and characterization of microearthquakes. Using the Goodness-of-Fit technique (Wiemer & Wyss, 2000), we estimate the magnitudes of completeness (M_c) for M_{IV} , M_{9M} , M_{MATCH} , and DL based catalogs. The DL and matched filter catalogs show a significant reduction of M_c . The corresponding b-value is around ≈ 0.6 , very low but consistent with that retrieved from the INSN (M_{IV}). The low b-value is likely due to the short time span, and limited spatial range,

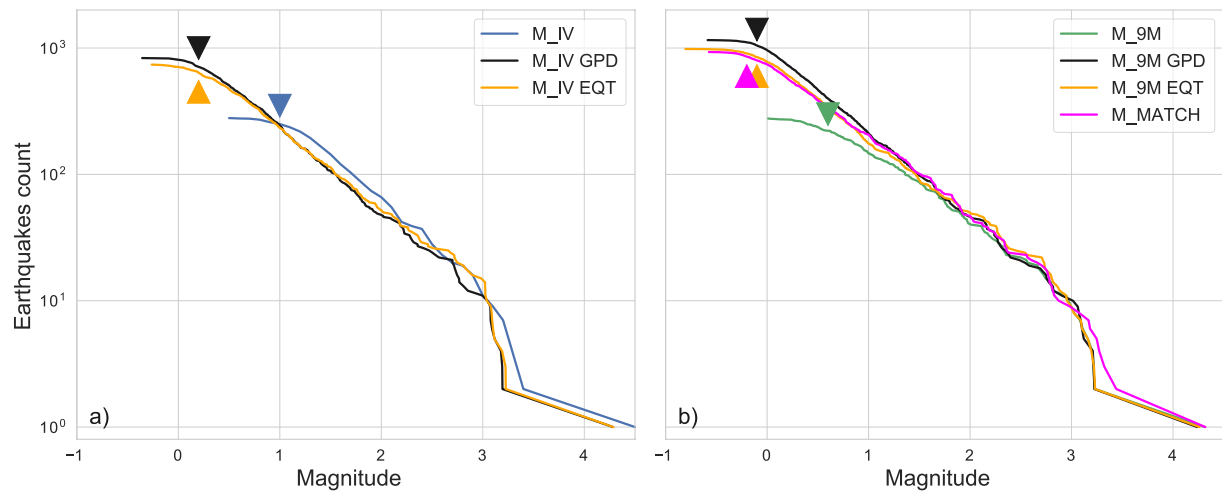


Figure 6. Gutenberg-Richter M_L distributions. Completeness magnitude (M_c) from Table 3 are shown for each catalog as triangles.

spanned by our catalogs. Standard error (SE) of the frequency-magnitude distribution b -value is computed using the Shi and Bolt (1982) approach. Results are summarized in Table 3.

In Figure 7 we plot the 2D projections and vertical cross sections of earthquake hypocenters for catalogs M_9M, M_MATCH, M_9M GPD, and M_9M EQT, respectively: the color shades reflect the kernel density estimates of the hypocenter distribution, with lighter color indicating higher density; in the cross sections only earthquakes less than 1 km away from the profile are plotted; red stars indicate the mainshock locations. The locations from the four different catalogs all show a main NW-SE alignment, common to the original M_IV catalog (Figure 1b), that suggests a fault plane with a strike of about 120° and steeply dipping (about 70°) SW. This is consistent with one of the two fault planes derived from the Time Domain Moment Tensor solution (see Figure 1), in particular with the plane striking $N308^\circ$ and dipping 46° (<http://cnt.rm.ingv.it/event/23558121>). In all the catalogs, the mainshock is located at the SE tip of the epicentral alignment, and most of the hypocenters are clustered within the 5–10 km depth interval. The region characterized by highest density of earthquakes shows different geometry according to the methods used to retrieve the catalogs. For M_MATCH most of the earthquakes are located in the westernmost part of the sequence. This may be due to the fact that three out of the six stations used for the matched filter procedure (stations MOCL, CRCL, and GAGN) are positioned in the western part and are closer to the sequence. For this reason the resulting catalog may have a slight spatial bias toward the west. The M_9M GPD and M_9M EQT catalogs show a more consistent distribution of events with an elongated shape in the NW-SE direction. A further, common feature which is particularly evident in the M_MATCH and M_9M EQT catalogs is a secondary alignment at the NW margin of the sequence. This secondary cluster, NNW-SSE oriented, exhibits hypocentral depths concentrated in the shallowest portion (5–8 km) of the main sequence. One question that

arises from the comparison of the results of GPD and EQT (Figures 7e, 7f, 7g, and 7h) is whether the application of different methods could lead to different interpretation of the active structures. If we assume that both GPD and EQT are affected in the same way by the network density and geometry of the Mugello area, the differences shown can be attributed to the different implementation of phase picking in the two algorithms. In one case (GPD) only small waveforms windows of 4 s length are considered, while for EQT a more complex, hierarchical attention mechanisms is pursued in 60 s windows. Despite these different approaches, some common features of the seismicity are shown in the results, such as the continuity of the steeply dipping NW-SW fault plane located NW of the mainshock. Nonetheless, the overall picture of the seismicity obtained from EQT (Figure 7g) is less smeared with respect of the GPD (Figure 7e), showing more details and suggesting the activation of a complex fault system. We expect that these differences may eventually be

Table 3
Completeness Magnitude (M_c), b -Value, and Standard Errors (SE) for the Seven Catalogs

Catalog	M_c (90%)	b -Value	SE
M_IV	1	0.59	0.03
M_IV GPD	0.2	0.59	0.02
M_IV EQT	0.2	0.55	0.02
M_9M	0.6	0.49	0.03
M_MATCH	-0.2	0.51	0.01
M_9M GPD	-0.1	0.59	0.02
M_9M EQT	-0.1	0.56	0.02

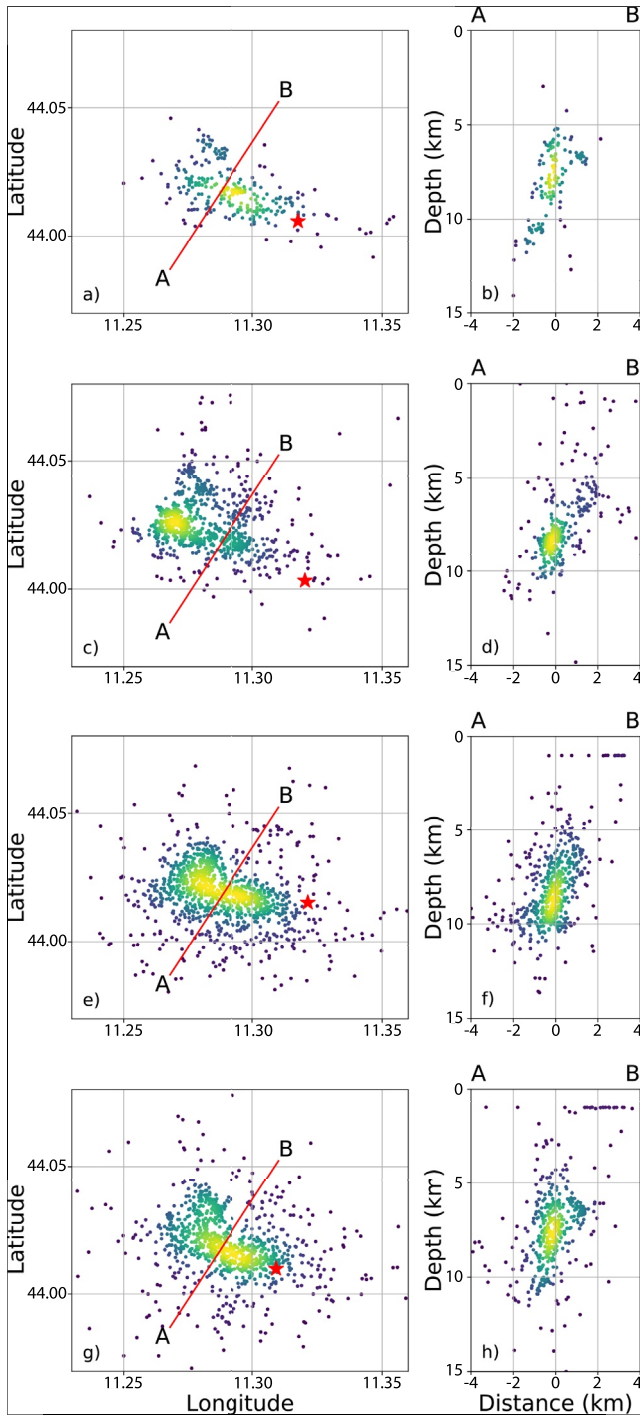


Figure 7. 2D projection of earthquake locations (left panels) and AB vertical cross section where only earthquakes whose distance from AB is below 1 km are plotted (right panels). From top to bottom we show earthquakes from (a and b) M_9M, (c and d) M_MATCH, (e and f) M_9M GPD, and (g and h) M_9M EQT catalogs, respectively. The color shades reflect the density of the earthquakes distribution, with lighter color indicating higher density. The relative location of the mainshock is shown as red star.

reduced by the retraining of GPD and EQT with specific data set reflecting the peculiarities of the seismic activity along the Apennines.

Despite the differences discussed above, the main contribution of the enhanced catalogs based on DL is a better definition of the spatial extent of the structure activated after the mainshock. Both M_9M GPD and M_9M EQT catalogs clearly image a rectangular fault patch of about 5.5 km length by 5 km width at depth, which is confined on the E by the location of the mainshock and on the W by a sharp, NW-SE trending, line (parallel to the shown AB cross section in Figure 7). We finally remark that a significant amount of additional information has been gained with the DL analysis, as can be seen comparing the starting bulletin data from Figure 1b and the results of Figure 7.

4.3. Picking Performance Comparison

In this section, our aim is to quantitatively compare the performance of the different methods in the task of phase picking. Further details are available in Supporting Information S1 where we add a set of Figures S9–S12 in Supporting Information S1 showing an example of P and S phase arrival detection at CRCL station in the time interval shown in Figure 2, characterized by repeated foreshock activity.

4.3.1. M_9M Versus DL

We compare the time differences between P and S wave arrival times for events in catalog M_9M and the corresponding arrivals (same event, same stations) from M_9M GPD and EQT catalogs (Figure 8). The results are shown as frequency histograms: in each panel are reported the mean value (μ) and standard deviation (σ) of the corresponding distribution.

Both DL methods are more accurate in determining P wave arrivals with respect to S wave ones; this is likely due to the lower SNR of S wave arrivals, whose onset is contaminated by the P wave coda, or to the difference in the frequency content. The μ and σ values suggest that EQT picks have better quality than those attained with GPD for both P and S waves (e.g., Mousavi et al., 2020, Figure 6). The non zero mean values and the slightly asymmetrical distributions of frequency histogram for P phase detections (Figures 8a and 8c) indicates that both GPD and EQT tend to anticipate manual picks, while the opposite occurs for S phase detection (Figures 8b and 8d). This behavior is likely ascribed to both the bandpass filter used to maximize the SNR and to a feature of the procedures themselves.

4.3.2. EQT Versus GPD

The M_9M GPD and M_9M EQT catalogs have 769 events whose origin times differ by less than 0.5 s. In addition to what is shown in Figure 8, we now examine the arrival time differences for both the P and S phases picked by the two methods at common stations and common events (Figure 9). Both P and S wave differential times exhibit a symmetric distribution, but while there are no systematic differences in P wave timing, GPD detections of the S wave anticipate the EQT ones by a mean value of ~ 0.03 s. This slight difference may derive from the different frequency bands in which waveform data are filtered: [3–20] Hz for GPD and [1–45] Hz for EQT, which are the same used in the pre-processing of the respective training data sets.

We also compare how GPD and EQT arrivals are distributed among the stations in shared events. In Figure 10 the number of P and S arrivals are plotted

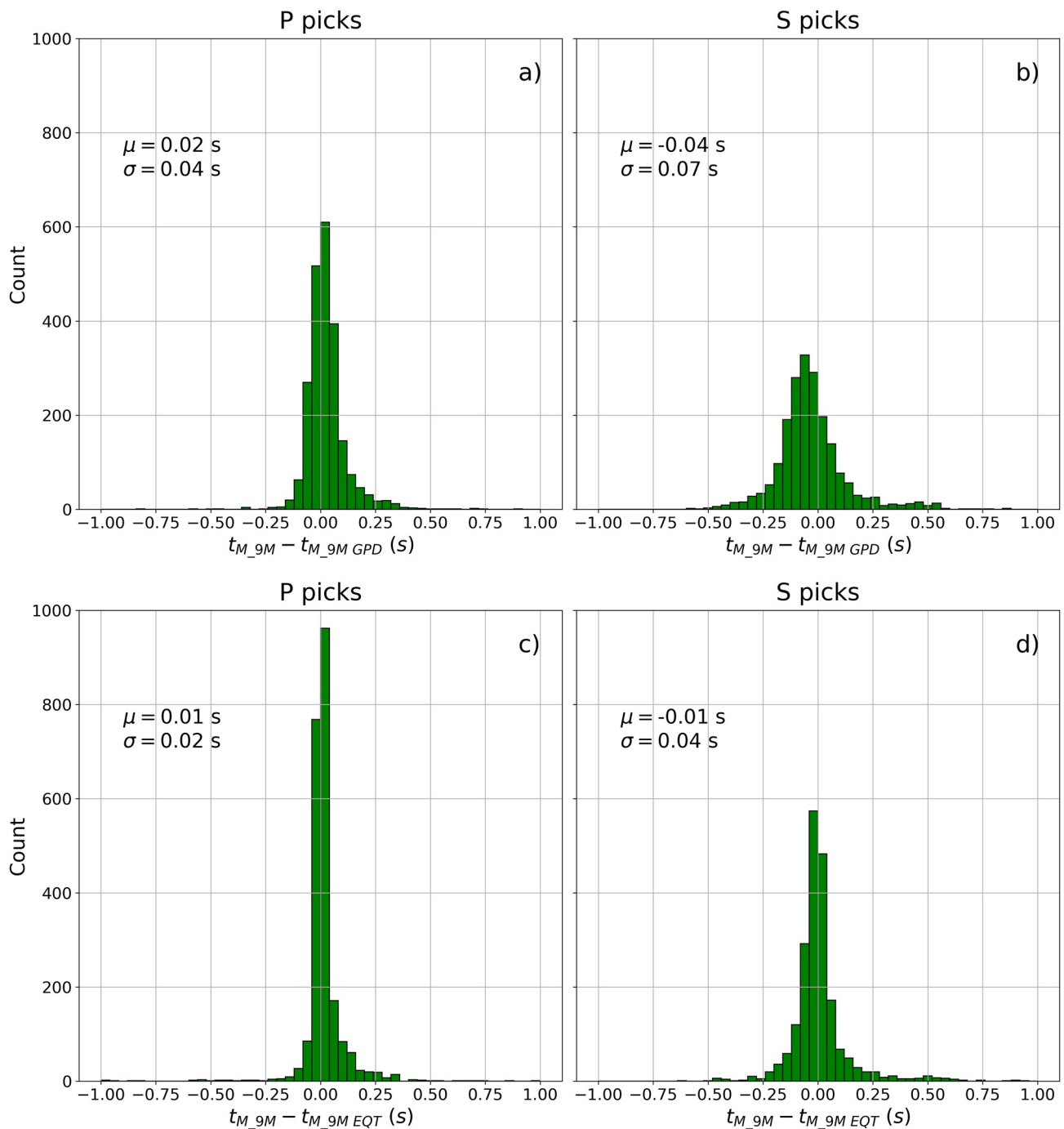


Figure 8. (a and c) Distributions of P and (b and d) S arrival-time differences (in second) between common earthquakes in (a and b) M_{9M} and M_{9M} GPD and (c and d) M_{9M} and M_{9M} EQT catalogs. μ and σ are mean and standard deviation in seconds, respectively. The number of picks is 2,246 for P wave and 1,956 for S wave per GPD and 2,275 for P wave and 2,011 for S wave for EQT.

as a function of the seismic stations on which they are observed in the M_{9M} catalog set. The stations are organized according to increasing distance from the sequence center.

In many cases, the two DL methods identify almost the same number of arrivals for both P and S phases. Despite this generalized similarity there are some relevant exceptions such as the large difference observed at station MBEN: here GPD model detects only 30% of both P and S phase arrivals with respect to EQT. EQT performance is also significantly better for stations CASC and SEI. The opposite is found for the P phases at CRCL (EQT

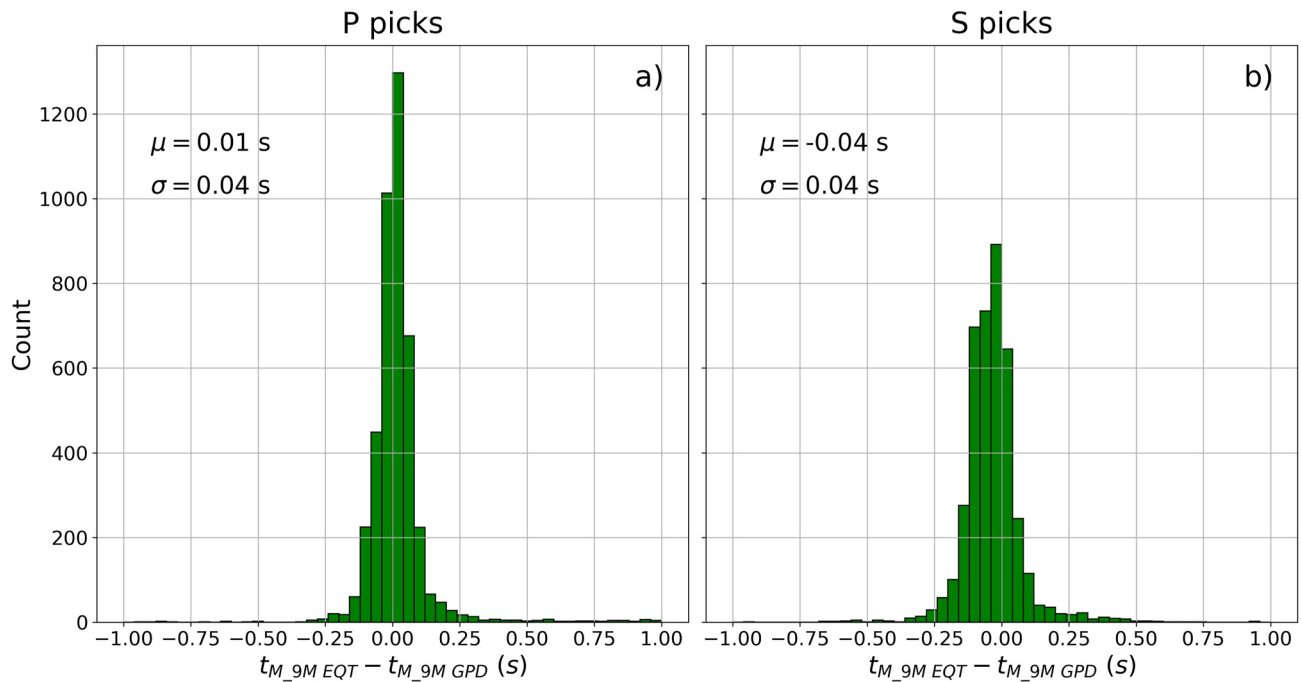


Figure 9. (a) Distributions of P and (b) S arrival-time differences between common earthquakes in M_{9M} GPD and M_{9M} EQT catalogs. μ and σ are mean and standard deviation in seconds, respectively. The total number of picks are 4,245 for P wave and 4,026 for S wave.

–30% with respect to GPD), and for both P and S phases at stations FNVD, VISG and RONT. We have verified that the difference in the number of picks detected at MBEN could be attributed to the value of the probability threshold defined to declare a pick in GPD algorithm. Lowering this value from 0.99 to 0.96 would allow us to triplicate the number of picks found. This suggests that a fine tuning of the probability threshold at different stations could yield better performance from the DL algorithms.

4.4. Summary of Results and Discussion

The main results can be summarized as follows:

1. The DL frameworks used throughout this study, EQT and GPD, show a marked ability to detect low magnitude earthquakes similar to that achieved using matched filter. The DL methods increase by a factor of ~ 3 the number of earthquakes with respect to benchmark permanent station derived M_{IV} catalog. The use of the denser temporary network increases the number of earthquakes by a factor of 4 with respect to the benchmark catalog.
2. The DL algorithm's performance in picking seismic phases is comparable to manual analysis. Both EQT and GPD pick P and S wave onset with high accuracy. However, GPD S picks slightly anticipate both those obtained by EQT and manually determined picks.
3. GPD has a better overall performance in finding low magnitude earthquakes occurring very close in time with respect to EQT. However, EQT shows the larger number of detected phases turned into arrivals suggesting a higher precision of the algorithm.
4. The distributions of the hypocenters from all catalogs show that the aftershocks align with a W-NW direction from the mainshock.
5. DL catalogs definitely improved the definition of the spatial extent of the fault structures activated after the mainshock, suggesting an area of 5.5 km length and 5 km width.

In this paper, we analyze a single seismic sequence limited in time and space. This provides a “controlled environment” to test the effectiveness and the applicability of DL approaches we used to exploit the data set. Due to the lack of previous comparison studies and with the aim of testing how transportable these DL methods are to new data we use the GPD and EQT models out-of-the-box, simply tuning the forecast probability thresholds.

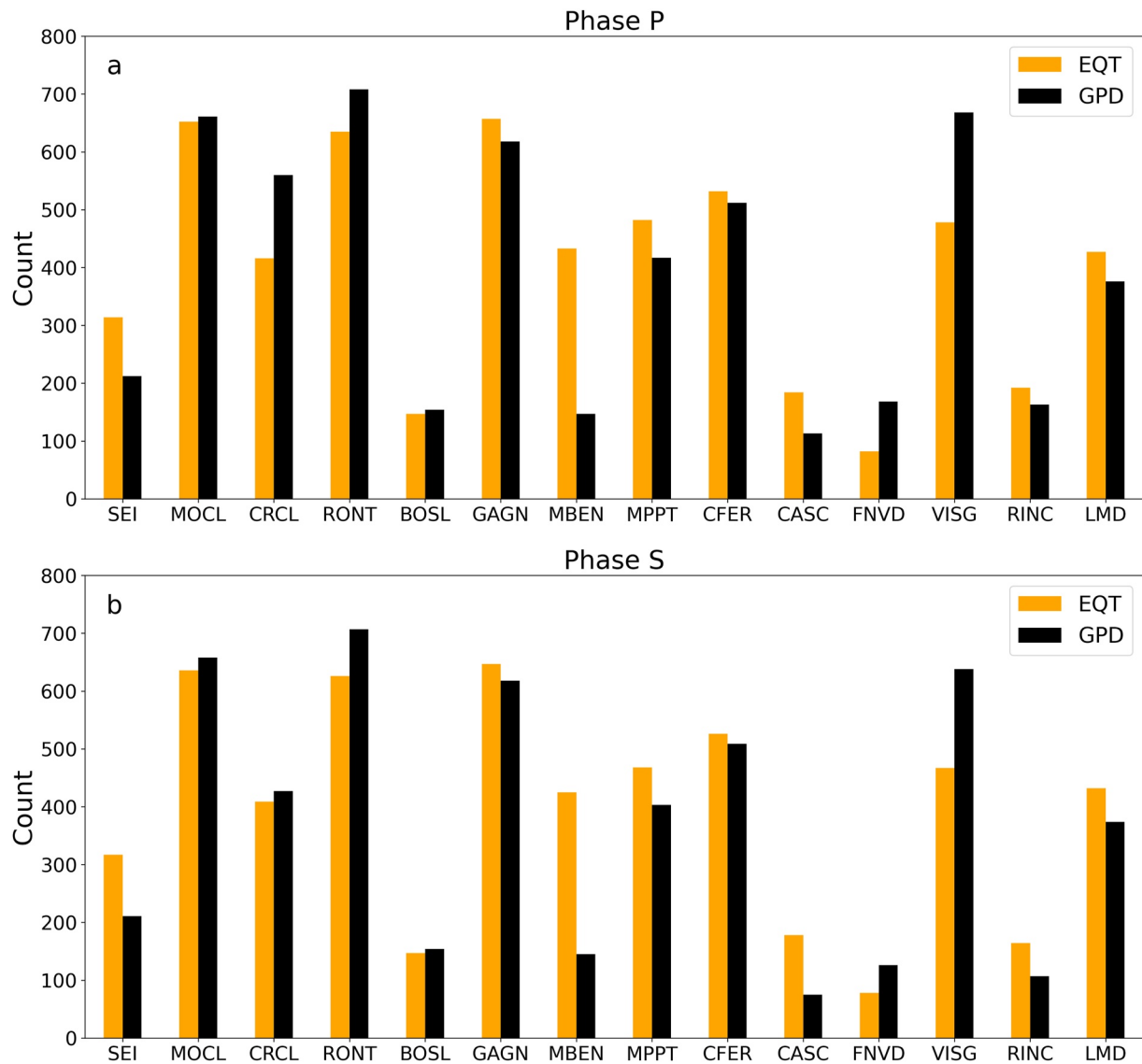


Figure 10. Comparison of the number of P and S arrivals detected at single stations, for the earthquakes in M_{9M} EQT and M_{9M} GPD catalogs (only events with origin time differing by less than 0.5 s are shown). The stations are listed according to increasing distance from the seismic sequence center.

While our results show that these methods work very well on our data set we expect to get better results if the DL models are trained on good quality data reflecting the peculiarities of the seismic release in the Apennines (Michelini et al., 2021). DL is a relatively new technique for the analysis of seismic data, and there is a need for a quantitative comparison of different approaches and DL schemes. To this end our paper is a first attempt to provide some insights, but there are ongoing efforts in developing open source frameworks, such as SeisBench (Woollam et al., 2021), whose aim is to facilitate the comparison of different DL algorithms by a standardized approach (Münchmeyer et al., 2021).

5. Conclusions

In this paper, we focus on testing the performance of recent DL algorithms for event detection and seismic phase picking against manually derived picks. As a case study we analyze the seismic swarm associated to the Mw 4.5 occurred on 9 December 2019 in the Mugello basin, which lasted about one month, using seismic data from a dense network of 13 permanent and 9 temporary stations. The DL based analyses implemented on the permanent stations alone detect a factor of 3 more earthquakes than the 279 achieved manually by the INGV bulletin, due

mainly to the better capability to identify seismic phases of low magnitude earthquakes with low SNR. When data from the denser temporary network are included a factor of 4 more earthquakes than the INGV bulletin are obtained. The DL algorithms detect phase arrivals with uncertainty similar to the manual picks, leading to better definition of microseismicity in the swarm and providing constraints on the geometry of activated structures. The DL models, generalizing what is learned from different data sets, can be effectively applied in different areas in an accurate, fast way. The adoption of DL methods in reprocessing existing data can lead to significant advances in the study of earthquake sequences, shed light on a variety of processes such as the time evolution of seismic sequences, or for more complex cases of fault interaction.

Data Availability Statement

Data from INGV stations (network IV) are publicly available by FDSN web service implemented on the Italian node of EIDA (<http://eida.ingv.it/>). Data from 9M network (<https://doi.org/10.7914/SN/9M2019>, <http://www.fdsn.org/networks/detail/9M2019/>) are destined for open access after an embargo period expiring on 7 January 2023. In the meantime, the subset of recordings used in this paper is available from the authors upon request. Seismic data processing has been carried out using open source software ObsPy (Beyreuther et al., 2010) and Pandas (The Pandas Development Team, 2020). Figures have been drawn using Matplotlib (Hunter, 2007). GPD and EQT are available from the Github repository (<https://github.com/interseismic/generalized-phase-detection> and <https://github.com/smousavi05/EQTransformer>, respectively). NonLinLoc is available from Anthony Lomax website <http://alomax.free.fr/nlloc/>.

References

- Allen, R. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5), 1521–1532. <https://doi.org/10.1785/BSSA0680051521>
- Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B), S225–S242. <https://doi.org/10.1785/BSSA07206B0225>
- Baer, M., & Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77(4), 1437–1445. <https://doi.org/10.1785/BSSA0770041437>
- Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433), eaau0323. <https://doi.org/10.1126/science.aau0323>
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). Obspy: A python toolbox for seismology [Software]. *Seismological Research Letters*, 81(3), 530–533. <https://doi.org/10.1785/gssrl.81.3.530>
- Bonini, M., Corti, G., Donne, D. D., Sani, F., Piccardi, L., Vannucci, G., et al. (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. <https://doi.org/10.1016/j.tecto.2016.04.045>
- Bruni, R., Saccorotti, G., Keir, D., Giunchi, C., & Fiaschi, A. (2019). Mugello temporary seismic deployment [Dataset]. *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/9M_2019
- Carminati, E., Lustrino, M., & Doglioni, C. (2012). Geodynamic evolution of the central and western Mediterranean: Tectonics vs. igneous petrology constraints. *Tectonophysics*, 579, 173–192. <https://doi.org/10.1016/j.tecto.2012.01.026>
- Di Bona, M. (2016). A local magnitude scale for crustal earthquakes in Italy. *Bulletin of the Seismological Society of America*, 106(1), 242–258. <https://doi.org/10.1785/0120150155>
- Drew, J., White, R. S., Tilmann, F., & Tarasewicz, J. (2013). Coalescence microseismic mapping. *Geophysical Journal International*, 195(3), 1773–1785. <https://doi.org/10.1093/gji/ggt331>
- García, L., Álvarez, I., Benítez, C., Titos, M., Bueno, A., Mota, S., et al. (2016). Advances on the automatic estimation of the P-wave onset time. *Annals of Geophysics*, 59(4), S0434. <https://doi.org/10.4401/ag-7087>
- Geller, R. J., & Mueller, C. S. (1980). Four similar earthquakes in central California. *Geophysical Research Letters*, 7(10), 821–824. <https://doi.org/10.1029/GL007i010p00821>
- Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1), 149–166. <https://doi.org/10.1111/j.1365-246X.2006.02865.x>
- Grigoli, F., Cesca, S., Amoroso, O., Emolo, A., Zollo, A., & Dahm, T. (2014). Automated seismic event location by waveform coherence analysis. *Geophysical Journal International*, 196(3), 1742–1753. <https://doi.org/10.1093/gji/ggt477>
- Herrmann, M., Kraft, T., Tormann, T., Scarabello, L., & Wiemer, S. (2019). A consistent high-resolution catalog of induced seismicity in Basel based on matched filter detection and tailored post-processing. *Journal of Geophysical Research: Solid Earth*, 124(8), 8449–8477. <https://doi.org/10.1029/2019JB017468>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment [Software]. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Ingate, S. (2008). The IRIS Consortium: Community based facilities and data management for seismology. In E. S. Husebye (Ed.), *Earthquake Monitoring and Seismic Hazard Mitigation in Balkan Countries* (Vol. 81, pp. 121–132). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6815-7_8
- INGV Seismological Data Centre. (1997). Rete Sismica Nazionale (RSN) [Dataset]. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/SD/X0FXNH7QFY>
- ISIDe Working Group. (2007). *Italian seismological instrumental and parametric database (ISIDe)*. <https://doi.org/10.13127/ISIDE>

Acknowledgments

The authors thank Federico Sani and Giacomo Corti for fruitful discussion and encouragement. Marco Capello, Damiano Biagini, Michele D’Ambrosio, and Andrea Fiaschi are acknowledged for their great field work during the installation and maintenance of the 9M network whose seismic instruments were loaned by SEIS UK. The facilities of SEIS UK are supported by the Natural Environment Research Council (NERC) under agreement R8/H10/64. This work has been partially supported by the project INGV Pianeta Dinamico 2021 Tema 8 SOME (CUP D53J1900017001) funded by Italian Ministry of University and Research “Fondo finalizzato al rilancio degli investimenti delle amministrazioni centrali dello Stato e allo sviluppo del Paese, legge 145/2018.” S. Gaviano benefited from DL training schools organized by Cost Action 17137 G2NET coordinated by Elena Cuoco. D. Keir acknowledges support from NERC grant NE/L013932/1 and from the University of Florence funds ATEN 2019, 2020, and 2021. PhD scholarships to R. Bruni and S. Gaviano were funded by Pegaso fellowship awarded by Regione Toscana, and by INGV during the XXXV PhD cycle at the Università di Firenze on the topic “Geophysics and Volcanology,” respectively. Four anonymous reviewers are thanked for their fruitful comments and observations that helped us to improve our study.

- Johnson, C., Bittenbinder, A., Bogaert, B., Dietz, L., & Kohler, W. (1995). Earthworm: A flexible approach to seismic network processing. *IRIS Newsletter*, 14, 1–4.
- Kao, H., & Shan, S.-J. (2004). The Source-Scanning Algorithm: Mapping the distribution of seismic sources in time and space. *Geophysical Journal International*, 157(2), 589–594. <https://doi.org/10.1111/j.1365-246X.2004.02276.x>
- Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, 90(1), 3–14. <https://doi.org/10.1785/0220180259>
- Langet, N., Maggi, A., Michelini, A., & Brenguier, F. (2014). Continuous kurtosis-based migration for seismic event detection and location, with application to Piton de la Fournaise Volcano, La Reunion. *Bulletin of the Seismological Society of America*, 104(1), 229–246. <https://doi.org/10.1785/0120130107>
- Lomax, A., Michelini, A., & Curtis, A. (2014). Earthquake location, direct, global-search methods. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (pp. 1–33). Springer New York. https://doi.org/10.1007/978-3-642-27737-5_150-2
- Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C. (2000). Probabilistic earthquake location in 3D and layered models. In C. H. Thurber & N. Rabinowitz (Eds.), *Advances in seismic event location* (pp. 101–134). Springer Netherlands. https://doi.org/10.1007/978-94-015-9536-0_5
- Magrini, F., Jozinović, D., Cammarano, F., Michelini, A., & Boschi, L. (2020). Local earthquakes detection: A benchmark dataset of 3-component seismograms built on a global scale. *Artificial Intelligence in Geosciences*, 1, 1–10. <https://doi.org/10.1016/j.aiig.2020.04.001>
- Mele, F., Bono, A., Lauciani, V., Mandiello, A., Marcocci, C., Pintore, S., et al. (2010). *Tuning an Earthworm phase picker: Some considerations on the PICK_EW parameters* (Vol. 164, pp. 1–27). Rapporti Tecnici INGV. http://editoria.rm.ingv.it/archivio_pdf/rapporti/163/rapporti_t164.pdf
- Mele, F. M., & Quintiliani, M. (2020). *Correzioni di stazione per il calcolo della Magnitudo Locale in Italia (aggiornate alla configurazione della Rete Sismica Nazionale di giugno 2018)* (Vol. 413, pp. 1–60). Rapporti Tecnici INGV. Retrieved from <http://editoria.rm.ingv.it/rapporti/2020/rapporto413/>
- Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., & Lauciani, V. (2021). Instance – The Italian seismic dataset for machine learning. *Earth System Science Data*, 13(12), 5509–5544. <https://doi.org/10.5194/essd-13-5509-2021>
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1), 3952. <https://doi.org/10.1038/s41467-020-17591-w>
- Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). Stanford earthquake dataset (STEAD): A global data set of seismic signals for AI. *IEEE Access*, 7, 179464–179476. <https://doi.org/10.1109/ACCESS.2019.2947848>
- Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., et al. (2021). Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *preprint*. <https://arxiv.org/pdf/2110.13671v1>
- Pagliuca, N. M., Battelli, P., Berardi, M., Modica, G., Castellano, C., Melorio, C., et al. (2020). Bollettino sismico italiano (BSI), III quadrimestre 2019 [Dataset]. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/BSI/201903>
- Peng, Z., & Zhao, P. (2009). Migration of early aftershocks following the 2004 Parkfield earthquake. *Nature Geoscience*, 2(12), 877–881. <https://doi.org/10.1038/ngeo697>
- Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2), e1700578. <https://doi.org/10.1126/sciadv.1700578>
- Piccinini, D., Piana Agostinetti, N., Saccorotti, G., Fiaschi, A., Matassoni, L., & Morelli, M. (2014). Orogen-parallel variability in 3D seismicity distribution, Northern Apennines (Italy): Evidence for a slab tear fault? *Journal of Geodynamics*, 82, 110–117. <https://doi.org/10.1016/j.jog.2014.09.005>
- Ross, Z. E., & Ben-Zion, Y. (2014). Automatic picking of direct P, S seismic phases and fault zone head waves. *Geophysical Journal International*, 199(1), 368–381. <https://doi.org/10.1093/gji/ggu267>
- Ross, Z. E., Meier, M., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A), 2894–2901. <https://doi.org/10.1785/0120180080>
- Rovida, A., Locati, M., Camassi, R., Lolli, B., & Gasperini, P. (2016). Catalogo Parametrico dei Terremoti Italiani (CPTI15) [Dataset]. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.6092/INGV.IT-CPTI15>
- Saragiotis, C., Hadjileontiadis, L., & Panas, S. (2002). PAI-S/K: A robust automatic seismic P phase arrival identification scheme. *IEEE Transactions on Geoscience and Remote Sensing*, 40(6), 1395–1404. <https://doi.org/10.1109/TGRS.2002.800438>
- Schaff, D. P. (2008). Semiempirical statistics of correlation-detector performance. *Bulletin of the Seismological Society of America*, 98(3), 1495–1507. <https://doi.org/10.1785/0120060263>
- Scognamiglio, L., Tinti, E., & Quintiliani, M. (2006). Time domain moment tensor (TDMT) [Dataset]. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/TDMT>
- Shi, Y., & Bolt, B. A. (1982). The standard error of the magnitude-frequency *b* value. *Bulletin of the Seismological Society of America*, 72(5), 1677–1687. <https://doi.org/10.1785/BSSA0720051677>
- Sleeman, R., & van Eck, T. (1999). Robust automatic P-phase picking: An on-line implementation in the analysis of broadband seismogram recordings. *Physics of the Earth and Planetary Interiors*, 113(1–4), 265–275. [https://doi.org/10.1016/S0031-9201\(99\)00007-2](https://doi.org/10.1016/S0031-9201(99)00007-2)
- The Pandas Development Team. (2020). pandas-dev/pandas: Pandas [Computer software manual] [Software]. Zenodo. <https://doi.org/10.5281/zenodo.3509134>
- Titos, M., Bueno, A., García, L., Benítez, M. C., & Ibañez, J. (2019). Detection and classification of continuous volcano-seismic signals with recurrent neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4), 1936–1948. <https://doi.org/10.1109/TGRS.2018.2870202>
- Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4), 859–869. <https://doi.org/10.1785/0119990114>
- Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., et al. (2021). SeisBench – A Toolbox for Machine Learning in Seismology. arXiv:2111.00786 [preprint]. <http://arxiv.org/abs/2111.00786>
- Woollam, J., Rietbrock, A., Bueno, A., & De Angelis, S. (2019). Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network. *Seismological Research Letters*, 90(2A), 491–502. <https://doi.org/10.1785/0220180312>
- Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, 1(11), e1501057. <https://doi.org/10.1126/sciadv.1501057>
- Zhu, W., & Beroza, G. C. (2018). PhaseNet: A deep-neural-network-based seismic arrival time picking method. *Geophysical Journal International*, 216(1), 261–273. <https://doi.org/10.1093/gji/ggy423>