Supporting Information for "Comparison of deep learning techniques for the investigation of a seismic sequence: an application to the 2019, Mw 4.5 Mugello (Italy) earthquake"

S. Cianetti[1], R. Bruni[2,1], S. Gaviano[3,1], D. Keir[3,4], D. Piccinini[1], G. Saccorotti[1], and C. Giunchi[1]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Via Cesare Battisti 53, 56125, Pisa, Italy

[2]Università degli Studi di Pisa, Dipartimento di Scienze della Terra, Via Santa Maria 53, 50126, Pisa

[3]Università degli Studi di Firenze, Dipartimento di Scienze della Terra, Via G. La Pira 4, 50121, Florence, Italy

[4]University of Southampton, School of Ocean and Earth Science, European Way, SO14 3ZH, Southampton, UK

Contents of this file

1. GPD forecast probability threshold

2. EQT forecast probability thresholds

3. Local magnitude computation

Corresponding author: S. Cianetti, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Via Cesare Battisti 53, 56125, Pisa, Italy (spina.cianetti@ingv.it)

Corresponding author: D. Keir, Università degli Studi di Firenze, Dipartimento di Scienze della Terra, Via G. La Pira 4, 50121, Florence, Italy. University of Southampton, School of Ocean and Earth Science, European Way, SO14 3ZH, Southampton, UK (derekboswell.keir@unifi.it)

4. Figures S1 to S12

5. Table S1

GPD forecast probability threshold

Here we detail how the thresholds for GPD forecast probabilities have been determined. As shown in Fig. S1, the number of picks increases as the forecast probability threshold decreases, and the trend shows a linear growth. The number of S picks is almost twice the P count, and reaches 400,000 for a threshold of 0.95. In order to minimize false detections and limit computational costs, we tried to characterize the forecast probability values of the picks showing a match in the M_9M catalog. This is shown in Fig. S2 where we plot the percentage distribution of the forecast probability threshold for GPD picks matching M_9M picks. A pick is considered matched if $|t_{M\_9M} - t_{M\_9MGPD}| \leq 0.5s$. The percentage of P and S picks detected by GPD with threshold above 0.99 are the 91% and 96% of all those matched, respectively. Finally we plot the residuals between P and S M_9M and GPD detected picks as a function of the forecast probability threshold (Fig. S3). To highlight how the residuals are distributed we also add to the plot the histograms of the residuals and of the forecast probability for both P and S phases. The largest majority of matched picks lay above the 0.99 threshold. On the other hand there is no correlation between the forecast probability values and the time residuals, being them spread within a large interval above the 0.99 threshold.

EQT forecast probability thresholds

In Fig. S4 we show the detections for EQT as a function of the forecast probability detection threshold: the number of events linearly increase as the threshold value decreases; below 0.1 no further events are recognized; the total number of event detections (which

represents the upper bound for P and S detections) is at least one order of magnitude smaller than that retrieved for GPD, suggesting a higher precision of EQT method. We choose 0.2 as event detection threshold after verifying that, despite the increased computational costs, only a tiny amount of additional earthquakes below that threshold could be associated and located.

As for GPD, we show the residuals between P and S phases of M_9M catalog and those detected by EQT as a function of the forecast probability, and we add the histograms to show how they are distributed. The residuals between M_9M and EQT picks are less dispersed than GPD suggesting that EQT provides more accurate picks (Fig. S5). We fix the forecast probability threshold values to 0.1 for P and S phases since this allows us to include almost all the recognized phases.

Local magnitude computation

The preprocessing for ML computations includes the removal of mean and linear trend, and bandpass filtering in the [0.5,30] Hz frequency interval to remove signals related to microseisms. The deconvolution of the instrumental transfer function and the simulation of the Wood-Anderson response are then applied. Maximum amplitudes of S-waves at the two horizontal components are used to obtain two independent measurements of ML for each station. The final estimate of ML and associated uncertainties are determined by trimmed mean and standard deviation, respectively. We point out that the bandwidth frequency interval used to prefilter the waveform data is chosen to steepen the Wood-Anderson roll-off toward lower frequencies (Uhrhammer et al., 2011). This improves the SNR at low frequency (Bormann and Dewey, 2014) allowing a more accurate estimate of low magnitude events.

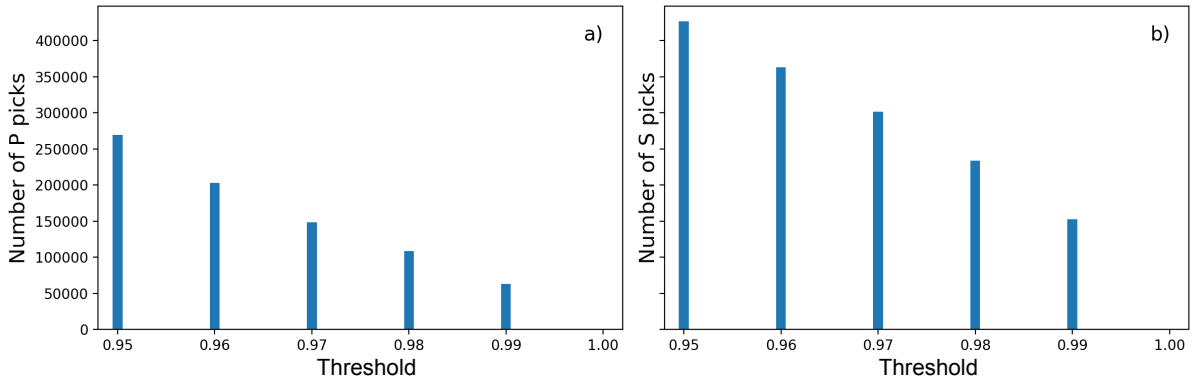

Figure S1. Number of P (a) and S (b) picks detected by GPD as a function of the threshold

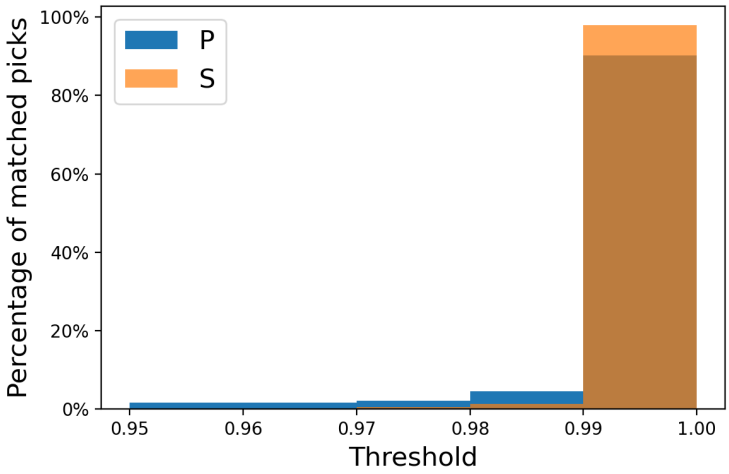

Figure S2. Percentage distribution of GPD picks matching M_9M picks as a function of the forecast probability threshold. A pick is considered matched if $|t_{M\_9M} - t_{M\_9MGPD}| \leq 0.5s$. The percentage of P and S picks detected by GPD with threshold above 0.99 are the 91% and 96% of all those matched, respectively.

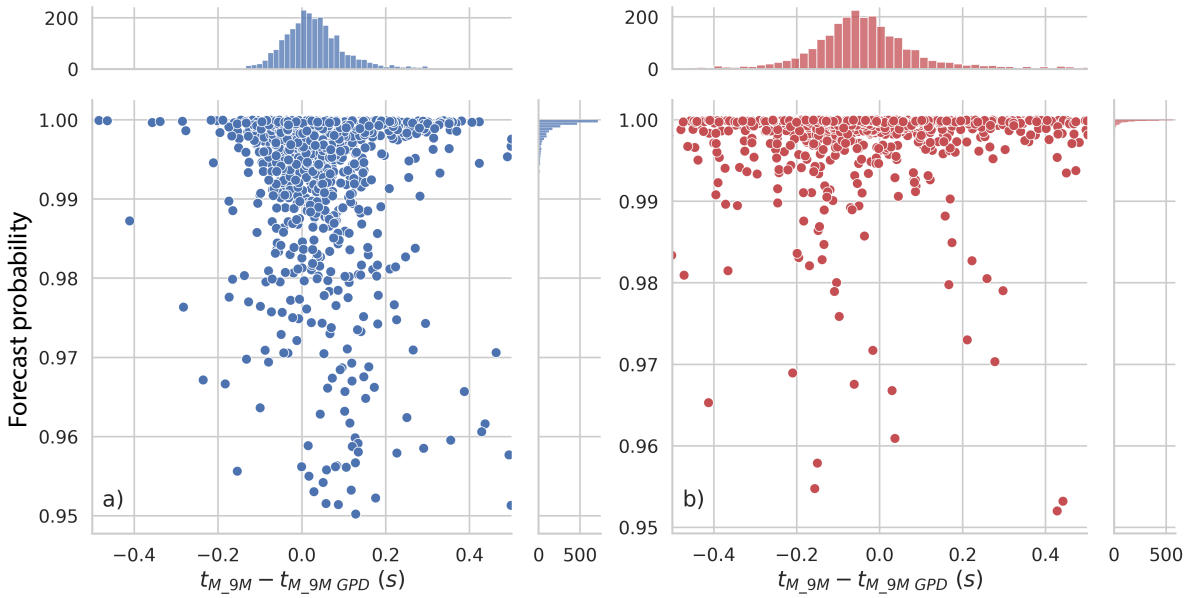

Figure S3. Scatter plots of pick time residuals (s) between M_9M and GPD detected P (a) and S (b) phases as a function of the forecast probability. Their distributions are also shown as histograms.

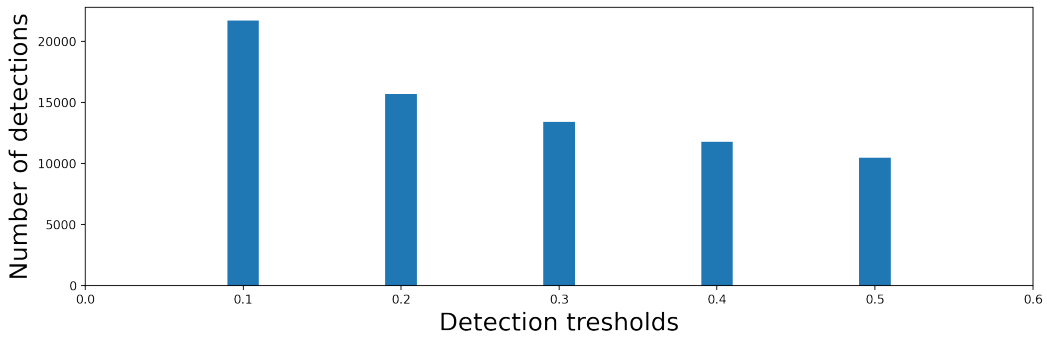

Figure S4. Number of event detections by EQT as a function of the forecast probability thresholds

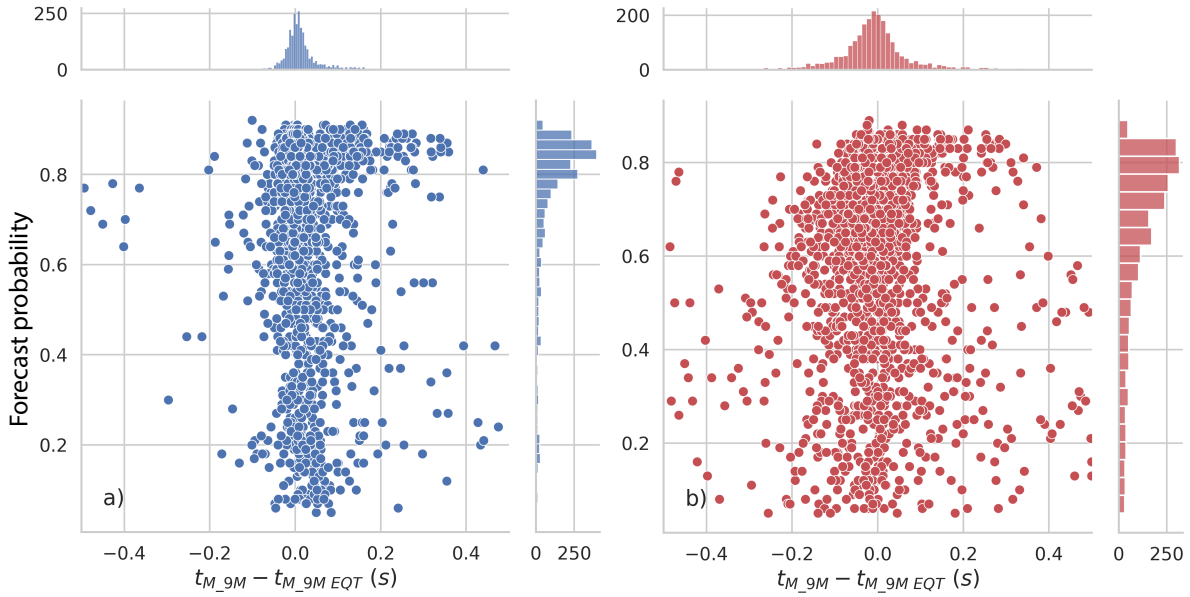

Figure S5. As in Fig. S3 but for EQT detected P (a) and S (b) phases for event detection threshold = 0.2.

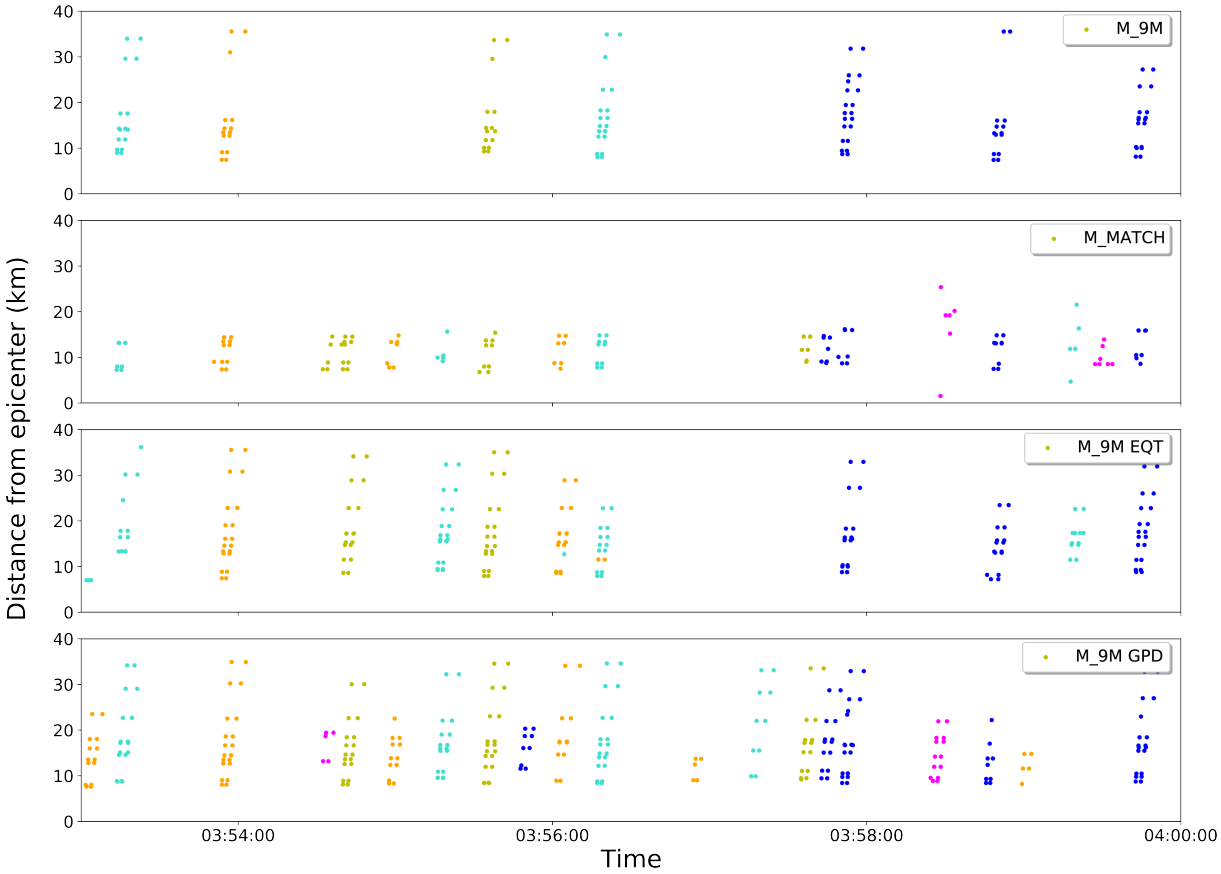

Figure S6. Example of P and S picks association into events. The picks are ordered as a function of station distance from the epicenter. Colors are assigned according to the event origin time and change every 10 seconds; the same color repeats after 1 minute.

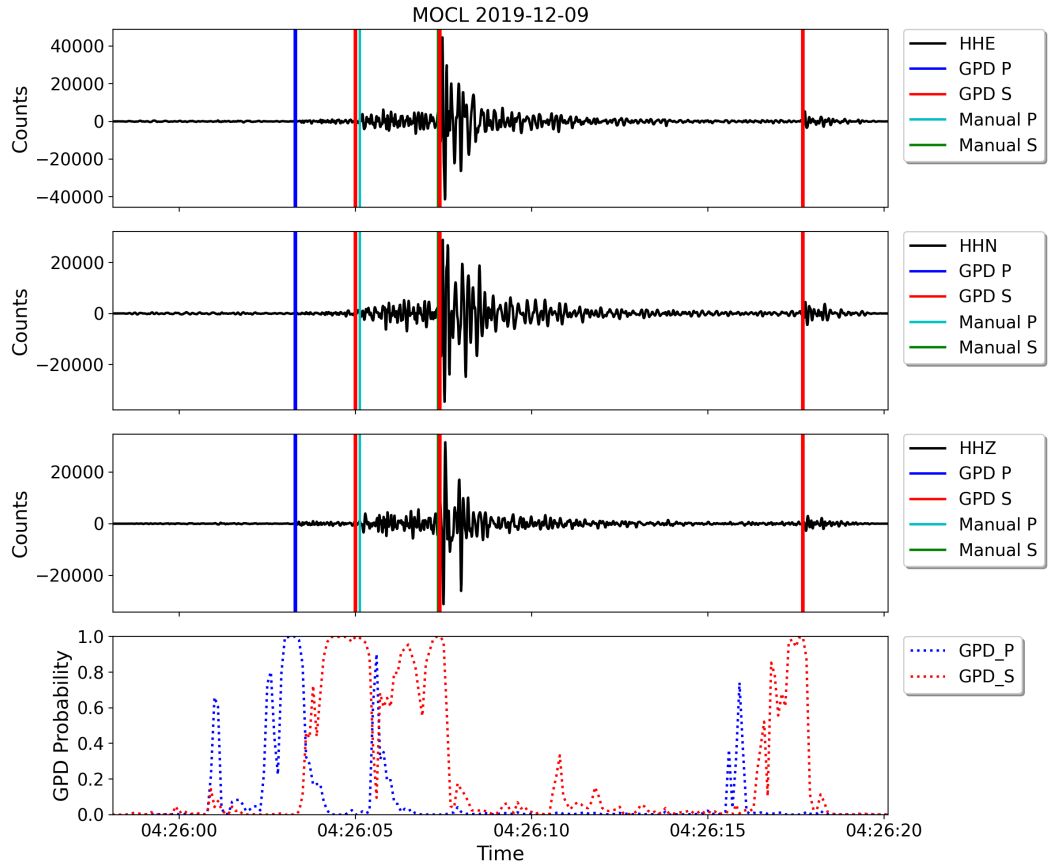

Figure S7. Waveform traces with manual (M_9M) and GPD determined picks (three top panels) and GPD P and S probability (bottom panel) at MOCL for one of the 14 events in the M_9M catalog whose origin time differs by more than 0.5 s with respect to the corresponding event in the M_9M GPD catalog. In this case two events occur in a time range comparable to $t_S - t_P$: GPD does not detect the onset of the P wave belonging to the second event because it is hidden inside the S wave train of the previous event. We indeed observe an increase in the P forecast probability (approximately at 04:26:06) that does not rise above the threshold of 0.99. Moreover, the operator here wrongly identify the onset of the S phase of the first event as the P onset of the second one.

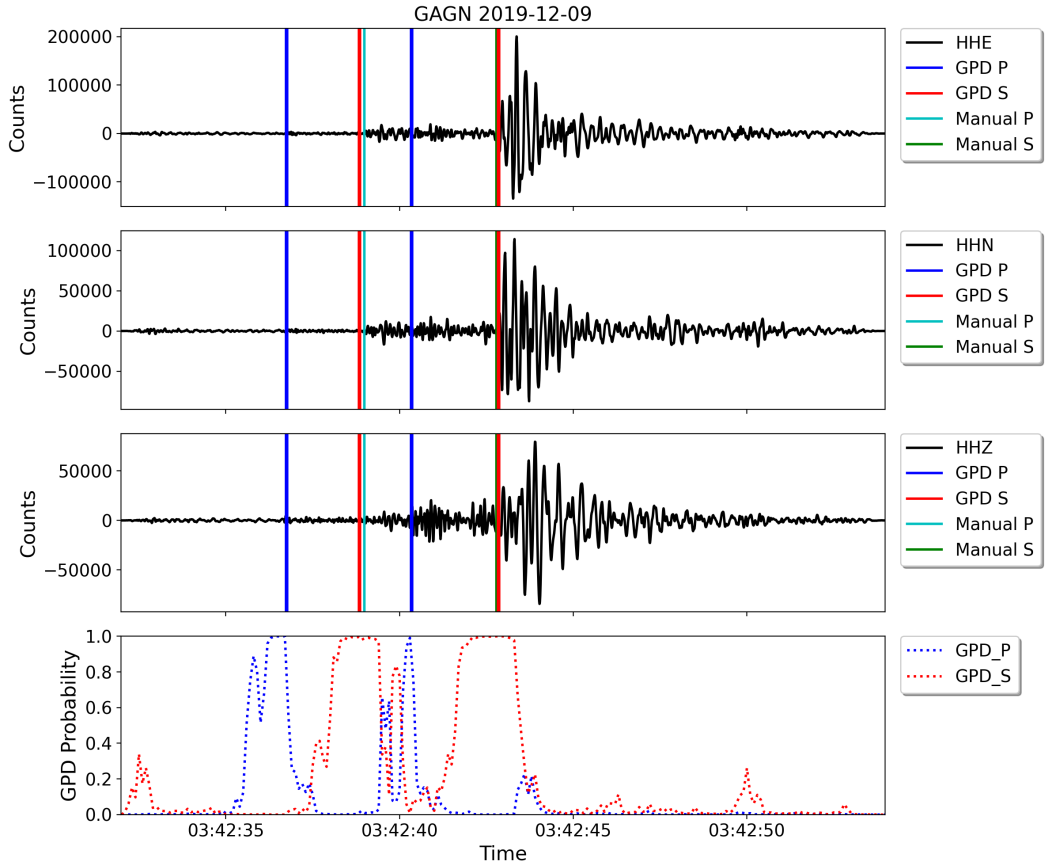

Figure S8. Another example as in S7 but for GAGN station. In this case GPD identity P and S phases for two events. Also in this case the operator wrongly identify the onset of the S phase of the first event as the P of the second one.

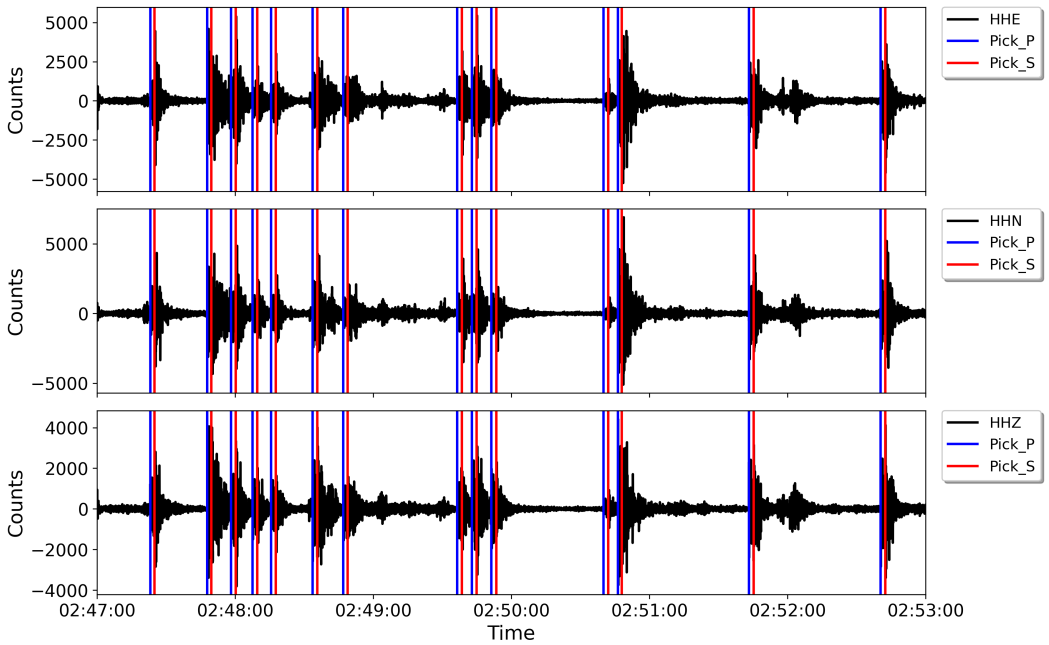

Figure S9. From top to bottom: seismic waveforms at CRCL for the three channels with P and S triggers detected by manual inspection of the waveforms. In the M_IV and M_9M catalogs only 4 of these events were included.

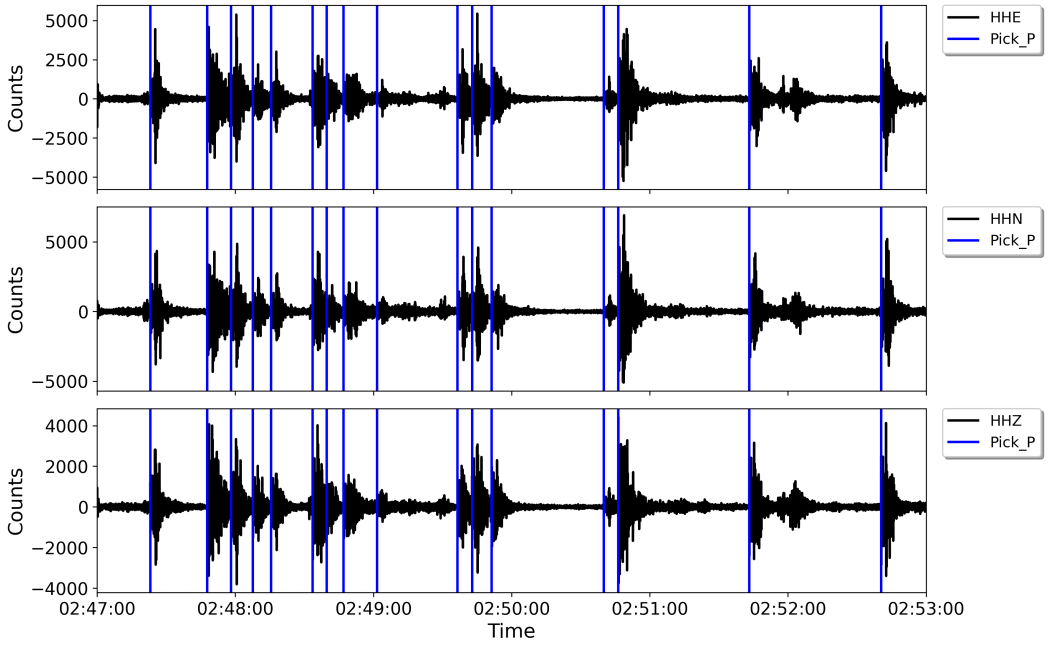

Figure S10. From top to bottom: seismic waveforms at CRCL for the three channels with P triggers detected by Match Filter assuming CC $\geq$ 0.7.

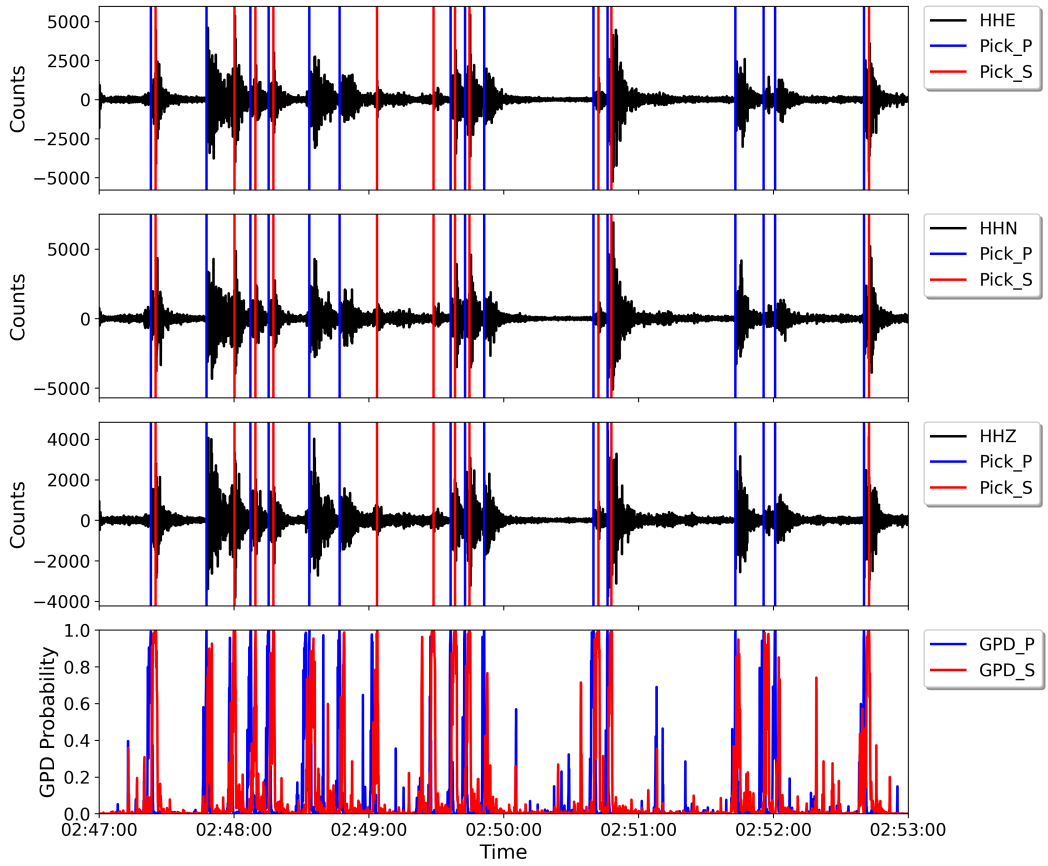

Figure S11. From top to bottom: seismic waveforms at CRCL for the three channels with P and S triggers detected by GPD assuming a minimum probability of 0.99 In In the last panel the P and S probability.

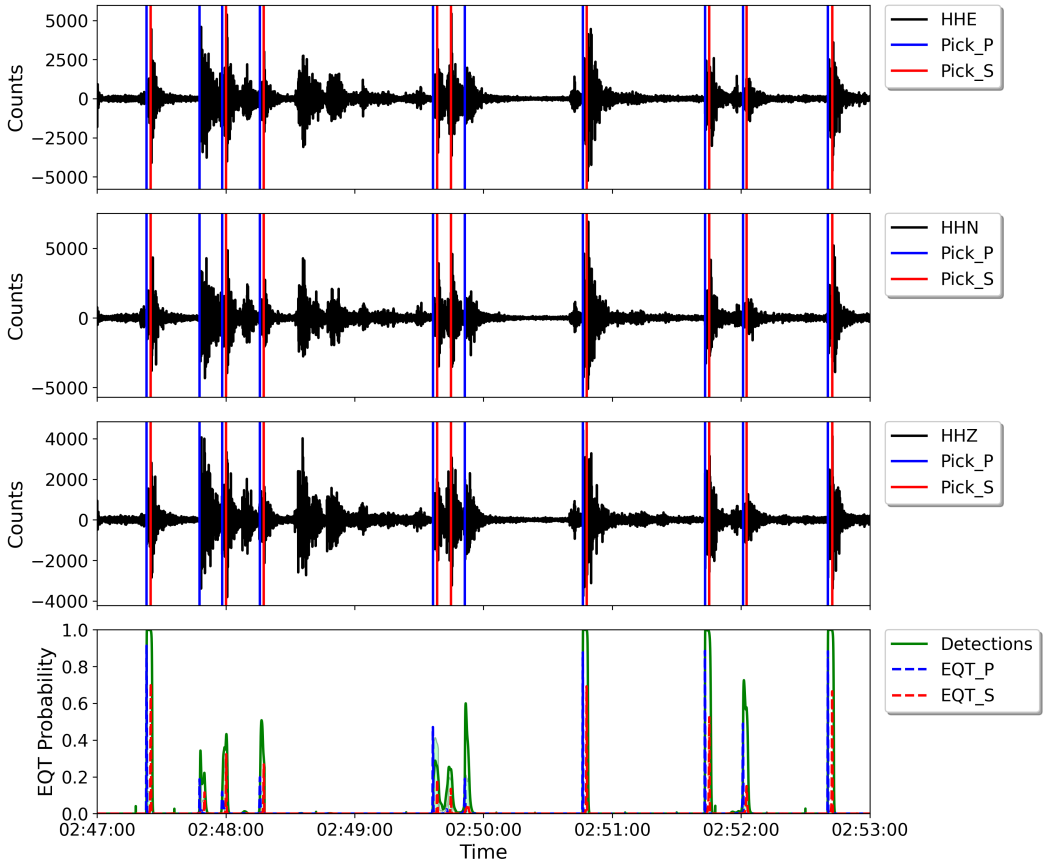

Figure S12. EQT probability. Earthquake detection threshold: 0.2 , P and S arrivals detection threshold 0.1

Network	Station	Channel	Location	Instrument type
IV	CRMI	HH	Carmignano	broadband
IV	OSSC	HH	Osservatorio del Chianti	broadband
IV	CSNT	HH	Castellina in Chianti	broadband
IV	ASQU	HH	Asqua	broadband
IV	SFI	EH	Santa Sofia	short period
IV	BRIS	HH	Brisighella	broadband
IV	MTRZ	HH	Monterenzio	broadband
IV	FNVD	HH	Fontana Vidola	broadband
IV	LMD	HH	Lutirano	broadband
IV	SEI	HH	Sant'Agata	broadband
IV	MPPT	EH	Montemurlo	short period
IV	MOCL	EH	Montecuccoli	short period
GU	POPM	HH	Popiglio, Piteglio	broadband
9M	BOSL	HH	Borgo San Lorenzo	broadband
9M	CASC	HH	Cascheta	broadband
9M	CFER	HH	Colleferro	broadband
9M	GAGN	HH	Gagnaia	broadband
9M	MBEN	HH	Monte Beni	broadband
9M	RINC	HH	Rincine	broadband
9M	RONT	HH	Ronta	broadband
9M	VISG	HH	Visignano	broadband
9M	CRCL	HH	Croci di Calenzano	broadband

Table S1. List of stations considered in the paper, from permanent (IV, GU) and temporary (9M) networks.