*Journal of Geophysical Research: Solid Earth*

Supporting Information for

# Comparing and integrating artificial intelligence and similarity search detection techniques for seismic sequences in Southern Italy

F. Scotto di Uccio[1], A. Scala [1,2], G. Festa[1,2], M. Picozzi[1], G.C. Beroza[3]

[1] Department of Physics "Ettore Pancini", University of Napoli Federico II, 80126 Napoli, Italy

[2] Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Roma 1, Rome, Italy.

[3] Department of Geophysics, Stanford University, Stanford, CA 94305, USA.

## Contents of this file



## Introduction

Here we report a parametric study on the machine learning detector EQTransformer, some additional figures and one table. The Figures compare the picks of EQTransformer to the manual ones (Figure S1), show the cumulative distributions of similarity parameters for EQCorrScan (Figure S2) and FAST (Figure S3), illustrate two seismic records that show the events detected by EQCorrScan and not by FAST (Figure S4) and viceversa (Figure S5). Figure S6 represents the scaling between the logarithm of the seismic moment and the local magnitude for the area. Table S1, finally, contains the magnitude of completeness and the *b*-value retrieved by the three methods for all analyzed sequences.

**Text S1.**

To better understand how the input parameters influence the detection performance of EQTransformer, we performed a parametric study on a specific sequence (Rocca San Felice sequence, ID1) starting from the following values: det_thresh=0.3, P_thresh=S_thresh=0.1, overlap=30%, batch size =500.
We systematically explored the batch size value, fixing the threshold parameters to the default values and reported an almost constant number of detections within the range 100-600. Reducing the detection threshold parameter from 0.3 to 0.15, to be more permissive in the selection of transients, we did not find significant changes in the number of events. When we lower the threshold for all the parameters (det_thresh=0.1, P_thresh=S_thresh=0.08), the final catalog increases by only about 10%. This latter increase, however, is not considered significant with respect to the catalog obtained with the initial parameterization, while it increases the risk of false detections. Therefore, we decided to keep the initial values of the parameters as reference for this study. We only changed the batch size to 100, to reduce memory demand**.**

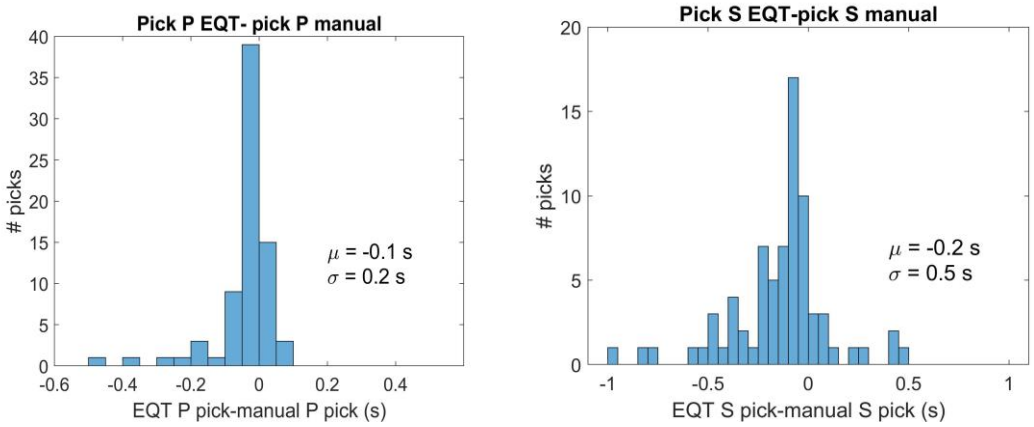


**Figure S1.** Comparison between the manual and EQTransformer automatic picks for 13 events (left panel: P picks; right panel: S picks). Automatic picks tend to anticipate manual ones, with a larger lag for S waves**.**

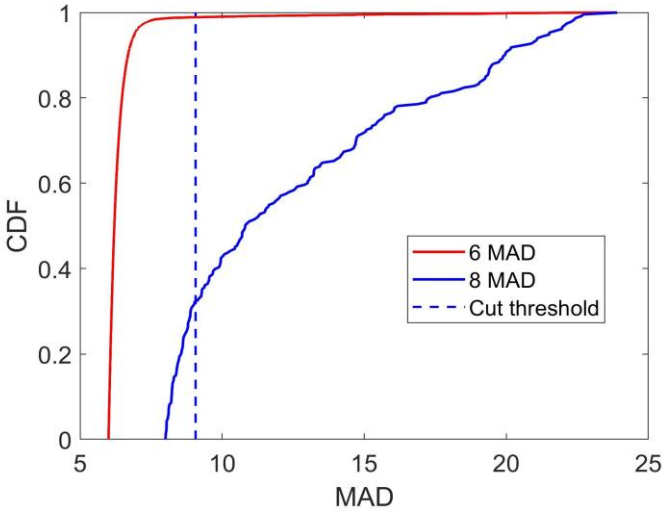


**Figure S2**. Cumulative distribution of the ratio *CCsum/*MAD for the catalog obtained with 6 MAD (red curve) and 8 MAD (blue curve) as thresholds.

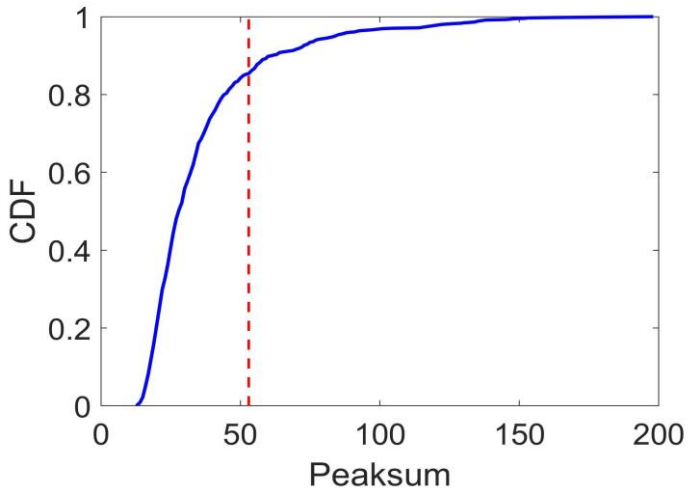


**Figure S3**. An example of cumulative distribution of the *peaksum* parameter, for events declared at 2 stations with FAST. We clearly identify a change in the slope of the distribution, around the 90% percentile, which is considered as the threshold to discriminate real events from false positives.

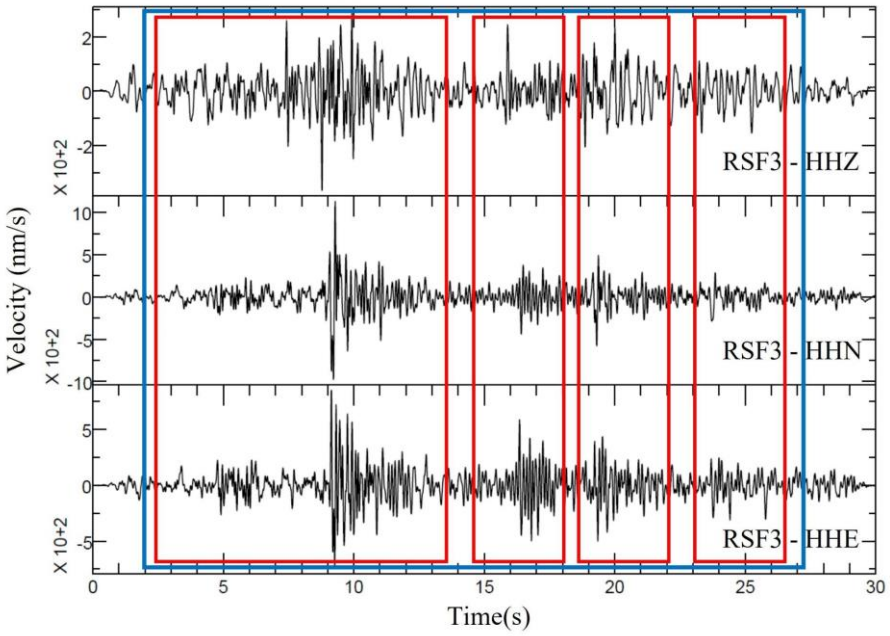


**Figure S4.** Events detected from TM (red boxes) and FAST (blue box): FAST is not able to discriminate events occurring closely in time.

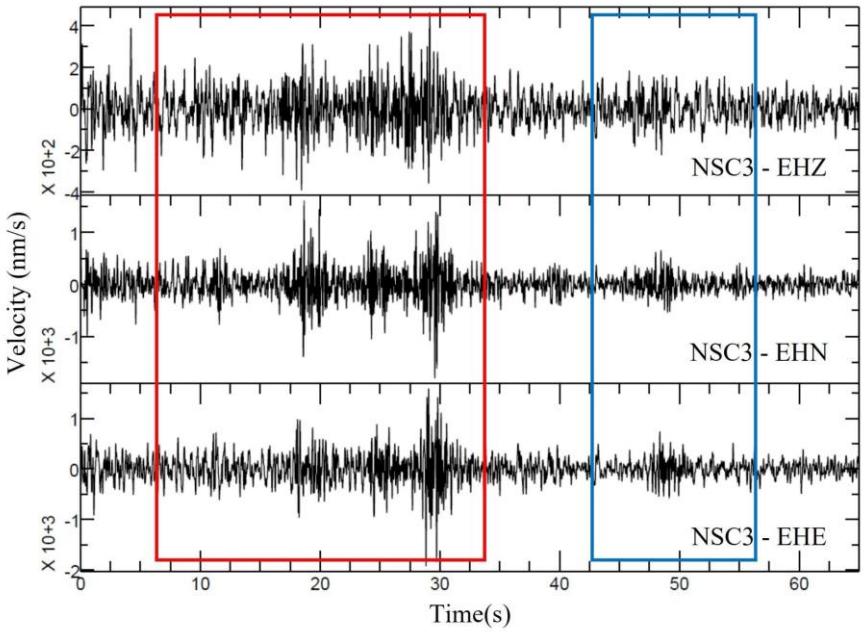


**Figure S5.** An example of a low SNR event declared by FAST and missed by TM. The three events close in time in the red box are correctly separately detected by TM.

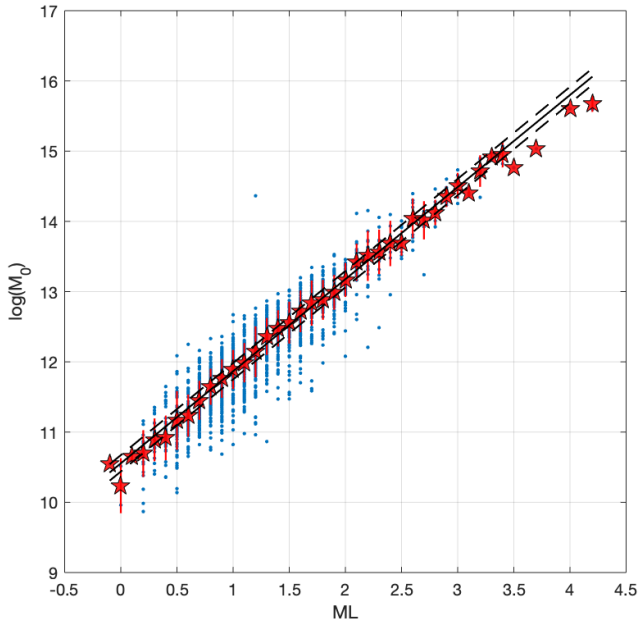


**Figure S6.** Linear regression between $\log M_0$ and $Ml$ using the historical earthquakes recorded in the Irpinia region. We obtain $\log M_0 = 1.31\, Ml + 10.55$, with $\sigma = 0.12$.

| ID | Ml main event | FAST | EQT | TM |
|---|---|---|---|---|
| 1 | 3.0 | Mc 0.0<br>$b = 0.71 \pm 0.05$ | Mc 0.2<br>$b = 0.54 \pm 0.04$ | Mc -0.3<br>$b = 0.71 \pm 0.03$ |
| 2 | 2.7 | Mc -0.3<br>$b = 0.60 \pm 0.05$ | Mc -0.2<br>$b = 0.51 \pm 0.07$ | Mc -0.3<br>$b = 0.68 \pm 0.06$ |
| 3 | 2.8 | Mc 0.3<br>$b = 0.62 \pm 0.11$ | N.A | Mc 0.4<br>$b = 0.76 \pm 0.13$ |
| 4 | 3.7 | Mc -0.1<br>$b = 0.76 \pm 0.08$ | Mc 0.1<br>$b = 0.54 \pm 0.08$ | Mc -0.2<br>$b = 0.65 \pm 0.07$ |
| 5 | 1.8 | Mc 0.3<br>$b = 0.83 \pm 0.10$ | Mc 0.3<br>$b = 0.75 \pm 0.10$ | Mc -0.3<br>$b = 0.73 \pm 0.05$ |
| 6 | 3.0 | Mc 0.3<br>$b = 0.75 \pm 0.18$ | Mc 0.4<br>$b = 0.53 \pm 0.13$ | Mc 0.2<br>$b = 0.69 \pm 0.11$ |
| 7 | 2.7 | Mc 0.4<br>$b = 0.66 \pm 0.13$ | N.A | Mc 0.3<br>$b = 0.78 \pm 0.14$ |
| 8 | 2.8 | Mc 0.1<br>$b = 1.16 \pm 0.11$ | Mc 0.9<br>$b = 1.00 \pm 0.14$ | Mc 0.2<br>$b = 1.26 \pm 0.13$ |
| 9 | 2.9 | N.A | N.A | Mc -0.3<br>$b = 0.60 \pm 0.09$ |
| 10 | 3.1 | Mc 0.2<br>$b = 0.73 \pm 0.16$ | Mc 0.2<br>$b = 0.53 \pm 0.09$ | Mc 0.1<br>$b = 0.76 \pm 0.13$ |

**Table S1.** Magnitude of completeness, b-values and magnitude of the main event in the sequence, using the different detection techniques for the 10 analyzed sequences