

# A Bayesian Method for Real-time Earthquake Location Using Multi-Parameter Data

Aldo Zollo<sup>1</sup>, Alessandro Caruso<sup>1</sup>, Grazia De Landro<sup>1</sup>, Simona Colombelli<sup>1</sup>, and Luca Elia<sup>1</sup>

<sup>1</sup>University of Naples Federico II

November 22, 2022

## Abstract

A primary task of a network-based, earthquake early warning system is the prompt event detection and location, needed to assess the magnitude of the event and its potential damage through the predicted peak ground shaking amplitude using empirical attenuation relationships. Most of real-time, automatic earthquake location methods ground on the progressive measurement of the first P-wave arrival time at stations located at increasing distances from the source but recent approaches showed the feasibility to improve the accuracy and rapidity of the earthquake location by using the additional information carried by the P-wave polarization or amplitude, especially unfavorable seismic network lay-outs. Here we propose an evolutionary, Bayesian method for the real-time earthquake location which combines the information derived from the differential P-wave arrival times, amplitude ratios and back-azimuths measured at a minimum of two stations. As more distant stations record the P-wave the posterior pdf is updated and new earthquake location parameters are determined along with their uncertainty. To validate the location method we performed a retrospective analysis of mainshocks ( $M > 4.5$ ) occurred during the 2016-2017 Central Italy earthquake sequence by simulating the typical acquisition layouts of in-land, coastal and linear array of stations. Results show that with the combined use of the three parameters, 2-4 sec after the first P-wave detection, the method converges to stable and accurate determinations of epicentral coordinates and depth even with a non-optimal coverage of stations. The proposed methodology can be generalized and adapted to the off-line analysis of seismic records collected by standard local networks.

## **A Bayesian Method for Real-time Earthquake Location Using Multi-Parameter Data**

**Aldo Zollo, Alessandro Caruso, Grazia de Landro, Simona Colombelli and Luca Elia**

Dept. Physics E. Pancini, University of Naples Federico II.

Corresponding author: Aldo Zollo ([aldo.zollo@unina.it](mailto:aldo.zollo@unina.it))

### **Key Points:**

- A real-time earthquake location method for early warning
- An evolutionary and probabilistic approach which jointly uses P-arrival time, amplitude and polarization
- It quickly converges to reliable hypocenter coordinates even with a non-optimal coverage of stations

1 **Abstract**

2

3 A primary task of a network-based, earthquake early warning system is the prompt event  
4 detection and location, needed to assess the magnitude of the event and its potential damage  
5 through the predicted peak ground shaking amplitude using empirical attenuation  
6 relationships. Most of real-time, automatic earthquake location methods ground on the  
7 progressive measurement of the first P-wave arrival time at stations located at increasing  
8 distances from the source but recent approaches showed the feasibility to improve the  
9 accuracy and rapidity of the earthquake location by using the additional information carried by  
10 the P-wave polarization or amplitude, especially unfavorable seismic network lay-outs.

11 Here we propose an evolutionary, Bayesian method for the real-time earthquake location  
12 which combines the information derived from the differential P-wave arrival times, amplitude  
13 ratios and back-azimuths measured at a minimum of two stations. As more distant stations  
14 record the P-wave the posterior pdf is updated and new earthquake location parameters are  
15 determined along with their uncertainty. To validate the location method we performed a  
16 retrospective analysis of mainshocks ( $M > 4.5$ ) occurred during the 2016-2017 Central Italy  
17 earthquake sequence by simulating the typical acquisition layouts of in-land, coastal and linear  
18 array of stations.

19 Results show that with the combined use of the three parameters, 2-4 sec after the first P-wave  
20 detection, the method converges to stable and accurate determinations of epicentral  
21 coordinates and depth even with a non-optimal coverage of stations. The proposed

22 methodology can be generalized and adapted to the off-line analysis of seismic records  
23 collected by standard local networks.

## 24 **1 Introduction**

25  
26 When an earthquake happens, the determination of its hypocentral coordinates and origin time  
27 is a standard, routine operation for any near-fault seismological observatory, and is typically  
28 performed within a couple of minutes from the earthquake occurrence, when most or all the  
29 phase arrival times at the stations are available.

30 The earthquake location is the most common example of a non-linear inverse problem,  
31 requiring the use of multiple data, spatially distributed around the source, to provide a unique  
32 and well constrained solution. When included in an automatic, real-time process of earthquake  
33 source parameter determination, the constraint of achieving a fast and robust solution even  
34 using a poor initial arrival-time data-set represents a further complexity to be managed.

35 Some proposed location methods solve the related inverse problem within a probabilistic frame  
36 and the maximum likelihood solution with its uncertainty are provided in the form of an  
37 posterior probability density function (e.g. NLLoc, *Lomax et al.*, 2009; or NLDiffLoc, *De Landro et*  
38 *al.*, 2015). Prior constraints are also adopted to optimize the process and to rapidly converge to  
39 a unique solution.

40 When dealing with real-time applications for Earthquake Early Warning Systems (EEWS), which  
41 asks for very fast source parameter estimates (within a few seconds), the earthquake location  
42 procedure becomes a sensitive issue which requires the adoption of dedicated, non-trivial  
43 algorithmic solutions. These must account for the continuous and evolutionary waveform data

44 availability in real-time, depending on the geometry and distribution of the seismic stations  
45 around the epicenter, as well as on the velocity of propagation of seismic waves across the  
46 network.

47 Nevertheless, an effective early warning system must provide reliable estimates of the location  
48 and size of an ongoing event, in the shortest possible time. The correct determination of  
49 hypocentral coordinates and origin time is essential a) to identify the source area and the  
50 causative fault of the ongoing event, b) to assess the earthquake impact (together with the  
51 earthquake magnitude) and predict the expected ground shaking and potential damage in the  
52 target area and c) to estimate the available lead-time at sensitive target infrastructures to be  
53 protected in order to start emergency operations and security actions addressed to secure the  
54 population, building and industrial facilities.

55 Several approaches for the real-time location have been developed and various parameters  
56 have been proposed in order to gain constraint on the solution, when few observed data are  
57 available. In the Elarms methodology (*Allen, 2007*), for example, at the arrival of the first  
58 trigger, the event is positioned at that unique station and the depth is fixed to the typical depth  
59 of the events in the region. When two and three stations trigger the event, the epicenter  
60 location is fixed as the centroid position between the triggered stations. Finally, when four  
61 stations have recorded the P-wave, a grid search method is used to locate the event, searching  
62 for the minimum misfit between predicted and observed arrival times. *Horiuchi et al. (2005)*  
63 first introduced the concept of not-yet triggered stations to constrain the event location when  
64 only two stations are available. In their approach, the initial solution is constrained using the  
65 estimated Equal Differential Time (EDT) surface (*Font et al., 2004; Lomax, 2005*), i.e., the quasi-

66 hyperbolic surface on which the difference in calculated travel time to a pair of stations is equal  
67 to the difference in observed arrival times for the two stations. The EDT shape in the 3D  
68 medium is progressively updated as more stations record the P-wave arrival time. In this  
69 approach the not-triggered stations provide a constraint which allows to further delimitate the  
70 probabilistic volume containing the hypocenter. *Satriano et al. (2008)* then introduced an  
71 evolutionary, probabilistic approach for the real-time earthquake location, also based on the  
72 EDT formulation, on the concept of triggered and not-yet-triggered stations and on the use of  
73 the Voronoi cell associated to each available station, allowing to constrain the initial hypocenter  
74 location even with just one recorded P-wave arrival. More recently other authors (*Noda et al.,*  
75 *2012; Eiserman et al., 2015; Liu and Yamada, 2014*) introduced new observed parameters to  
76 constrain the real-time earthquake location in early warning applications.

77 Previous attempts of using single station back-azimuth (BAZ) determinations showed that these  
78 measures can be affected by large uncertainties, possibly preventing their use for EEW  
79 (*Lockman and Allen, 2005*).

80 *Noda et al. (2012)* have proposed a new approach to improve the accuracy of BAZ estimations  
81 with a variable-length time window which is determined by the first half cycle of the initial P-  
82 wave. Using the Japanese K-NET strong-motion dataset they showed that the estimation, using  
83 this new approach, can be significantly improved both in accuracy of BAZ estimation and speed.

84 *Eiserman et al. (2015)* evaluated the robustness of three independent real-time back-azimuth  
85 (BAZ) determination schemes, through the offline analysis of southern California earthquake  
86 records and found that the three methods provide equivalent levels of accuracy. After passing  
87 the P-wave signals through specifically designed algorithms for checking the signal coherency

88 and signal-to-noise quality they show that BAZ estimates can be achieved in real time, with an  
89 optimal error of less than 15°.

90 In their method for identification of multiple events for EEWS, *Liu and Yamada(2014)* and *Wu et*  
91 *al.(2015)* use both P- and S-wave travel-times and amplitudes to constrain the earthquake  
92 location and magnitude of events occurring in an aftershock sequence. In a Bayesian,  
93 probabilistic frame, they consider the possibility of having more than one event occurring at  
94 any given time, by introducing a new posterior probability density function which jointly uses  
95 time and amplitude information from triggered and not-triggered stations.

96 Here we propose a Bayesian, multiparametric approach for the real-time earthquake location  
97 (M-PLOC). The proposed methodology exploits the continuous waveform data streaming from  
98 dense three component networks deployed in the source zones of potential damaging  
99 earthquakes and is specifically conceived for real-time seismic hazard analysis and EEW  
100 applications. The approach combines three different observed parameters (differential arrival  
101 times, amplitude ratios and back-azimuth estimates) measured in progressive (or fixed) time  
102 windows after the first P-wave arrival. The most probable estimates of hypocenter coordinates  
103 and origin time are provided as soon as the first stations trigger the event and are progressively  
104 updated as the P-wavefront expands across the network and new portions of signals are  
105 acquired by more and more distant stations.

106 We first describe the details of the methodology and then present the results of its application  
107 to a set of events recorded during the 2016-2017 Central Italy seismic sequence. During the  
108 testing phase, we perform jackknife simulation experiments with optimal/non-optimal data

109 acquisition lay-outs, by changing the geometry, coverage and number of stations and discuss  
110 the performance of the method for the different situations.

## 111 **2 Method**

112  
113 Let us assume that the EEW seismic network is composed of  $N$  sensor probes, with the  
114 capability to detect the arrival of the first P-wave arrival and to measure the arrival time, the P-  
115 wave polarization direction (angle from the North) and peak displacement amplitude in  
116 progressively expanded time windows (unit window length of 1 sec, maximum window length  
117 of 5 sec).

118 We assume that the arrival time ( $t_i$ ), back-azimuth ( $\alpha_i$ ) and peak displacement amplitude ( $PD_i$ )  
119 are available at station  $i$  ( $i = 1, \dots, N$ ) in a fixed time window of 1 sec, after the first P-wave is  
120 detected at the station, although a minimum of two stations are required to get the first  
121 location estimate.

122 As the P-wavefront expands spatially from the hypocenter, more stations record the P-wave  
123 arrival and additional arrival times, polarization and amplitude data can be used to constrain  
124 the earthquake location. In this sense, the proposed location is evolutionary, by including more  
125 and more data as the time increases since the earthquake origin.

126 At any time after the recording of a 1sec P-wave time window at  $N$  stations (minimum  $N=2$ ), the  
127 multiple data-sets that will be used for earthquake location are:

- 128 - differential first P-wave arrival times  $\Delta t_{ij}$  at any couple of stations  $i, j$
- 129 - peak displacement amplitude ratios  $\Delta PD_{ij} = \log \frac{PD_i}{PD_j}$  at any couple of stations  $i, j$
- 130 - measurements of BAZ from P-wave polarizations at the available  $N$  stations

131 The methods for the real-time measurement of P-wave arrival times, polarization and peak  
132 displacement amplitude) for earthquake location are described in the following paragraph. We  
133 assume that measurements are available with the associated error estimate.

134

135 *Real-Time measurement of differential P-wave arrival times, polarizations and peak*  
136 *displacement amplitude ratios*

137 The algorithm processes the three-component, ground acceleration data streams recorded by  
138 an accelerometer seismic network. In its offline version, the real-time data acquisition of the  
139 vertical component of ground motion is simulated using local files (SAC - Seismic Analysis Code  
140 format) with the packetization of data-stream set at 0.5 seconds. A preliminary removal of the  
141 mean value and linear trend of the signal is operated when the first P-wave arrival is detected  
142 (automatically or provided from the header of SAC files).

143 In real-time mode, as soon as two stations have been triggered by the earthquake signal and  
144 the automatic P-phase picking is available, the differential P-wave arrival times is therefore  
145 computed as the difference of arrival times at each triggered station. For more than two  
146 available P-arrival times, all possible differential time combinations are evaluated and used for  
147 the event location.

148 First P-wave arrival times at each station are obtained through an automatic picking procedure  
149 based on a recursive STA/LTA trigger-based strategy, e.g. the FilterPicker method by *Lomax et*  
150 *al.(2012)*

151

152 The back-azimuth BAZ, i.e. the angle from the North of the epicenter-to-station direction, is  
 153 then estimated in a P-wave time window length of 0.5 seconds, this value having been set upon  
 154 preliminary tests with variable widow lengths. A polarization analysis is applied to the three  
 155 component P-amplitude signal, band-pass filtered in the frequency band 0.5-3 Hz (see d) and e)  
 156 in Figure 1). A Moving Average (MA) approach has been used as first proposed by *Nakamura*  
 157 (1988) and furtherly modified by *Eiserman et al.* (2015). In this approach, the BAZ evaluated at  
 158 the ending point  $n$  of the P-wave discrete-time series, is defined following a recursive formula:

$$159 \quad BAZ_n = g[\theta^n(R_{ZE}^n, R_{ZN}^n), \text{sign}(R_{ZN}^n)] \quad (1)$$

160

161 where:

$$162 \quad \theta^n = \arctan\left(\frac{R_{ZE}^n}{R_{ZN}^n}\right) + \pi$$

163 with:

$$164 \quad R_{ZE}^n = \alpha R_{ZE}^{n-1} + A_Z^n A_E^n$$

$$165 \quad R_{ZN}^n = \alpha R_{ZN}^{n-1} + A_Z^n A_E^n$$

166  $\alpha$  is a smoothing parameter smaller but close to the unity ( $\alpha = 0.99$ ),  $A_Z^n$ ,  $A_N^n$  and  $A_E^n$  are the  
 167 amplitudes of the Vertical, North and East component of n-th sample, respectively. The  
 168 recursive formula (1) provide the BAZ as a weighted average of the values estimated in a  
 169 progressively expanded P-wave time window, with weights given by the recorded vertical  
 170 amplitude. The factor  $\alpha$  ensure that series terms nearby to the n-th sample contribute more  
 171 than distant ones. The function  $g$  is defined as (*Eisermann et al.*, 2015):

$$172 \quad g(\varphi) = \begin{cases} \varphi + \pi & \text{if } R_{ZN}^n < 0 \\ \varphi & \text{otherwise} \end{cases}$$

173 This flip condition removes the 180° ambiguity in the BAZ definition. Finally, similarly to  
174 Eisermann et al. (2015), a muting condition has been applied to reject low signal-to-noise  
175 amplitudes in the weighted recursive formula (1).

176 In a time window of 2 seconds, the ground acceleration waveform is integrated once and the  
177 linear-trend of the signal is removed in order to get the peak velocity amplitude (Pv) within the  
178 considered time window. The parameter Pv at the first two stations is measured within a time  
179 window with the same length after the first P-arrival (see d) and e) in Figure 1) and used to  
180 compute the logarithm of their amplitude ratio to be used for the event location. As for the  
181 differential times, the peak velocity amplitude ratio is computed for any couple of stations for  
182 which Pv has been measured.

183

184 *Recursive use of the Bayesian method for model parameter estimation using multiple data sets*

185 Let us recall the general formulation of the Bayes formula for a general model parameter vector  
186 ( $\mathbf{m}$ ) to be determined using a single data-set vector ( $\mathbf{d}$ ):

187

$$188 \quad P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})}{p(\mathbf{d})} \quad (2)$$

189 Where

190  $P(\mathbf{m}|\mathbf{d})$  is the posterior probability density function (pdf) of parameters given the data;

191  $P(\mathbf{d}|\mathbf{m})$  is the conditional pdf of data given the model parameters;

192  $\rho(\mathbf{m})$  is the prior pdf on model parameters;

193  $p(\mathbf{d})$  is the data marginal likelihood ( $p(\mathbf{d}) = \int P(\mathbf{d}|\mathbf{m})\rho(\mathbf{m}) d\mathbf{m}$ ), e.g. the posterior pdf  
194 normalization factor.

195 In our earthquake location problem using multiple data-sets, we propose the recursive use of  
 196 Bayes' formula, where the posterior pdf of  $\mathbf{m}$  given an initial data-set is used as prior  
 197 information for obtaining the posterior pdf of  $\mathbf{m}$  given the second data-set, which is in turn set  
 198 as the prior pdf for the final posterior pdf given the third data-set.

199 Let us consider N stations and define the differential P-times as the initial data-set for our  
 200 recursive Bayesian approach:

$$\mathbf{d}_1 \equiv (\Delta t_{12}, \dots, \Delta t_{N1})$$

201 The components of the model parameter vector are the cartesian coordinates of the  
 202 hypocenter location:  
 203

$$\mathbf{m} \equiv (x, y, z)$$

204 The conditional probability  $P(\mathbf{m})$  can be defined as the likelihood function for differential time  
 205 residuals according (Tarantola & Valette, 1982):  
 206

$$P(\mathbf{d}_1|\mathbf{m}) = \text{const } e^{-\frac{\sum_1^{N-1} \sum_2^N (\Delta t_{ij} - \Delta \tau_{ij}(\mathbf{m}))^2}{2\sigma_T^2}} \quad (3)$$

207  
 208  
 209 Where  $\Delta \tau_{ij}$  is the theoretical differential time, computed for a given model parameter vector  
 210  $\mathbf{m}$ , and  $\sigma_T^2$  is a theoretical estimate of the variance for differential times. In case the error  $\sigma_i$  on  
 211 single differential P-times is measured from data, its squared-inverse can be used in the above  
 212 formula as a weighting factor of the summation term.

213 Let us note that the differential arrival time between two stations  $i$  and  $j$  depends only on the  
 214 differential travel-times  $t_{oi} - t_{oj}$  and not on the event origin time  $t_o$ :

$$\Delta t_{ij} = (t_o + t_{oi}(x, y, z)) - (t_o + t_{oj}(x, y, z)) = t_{oi} - t_{oj}$$

216  $t_{oi}$  and  $t_{oj}$  are the travel times from the earthquake hypocenter to stations  $i$  and  $j$  ,  
 217 respectively.

218 The origin time corresponding to the hypocenter at  $(x, y, z)$  is given by:

$$219 \quad T_o = \frac{\sum_i^N (t_i - t_{oi}(x, y, z))}{N} \quad (4)$$

$$220 \quad \sigma_{T_o} = \sqrt{\frac{\sum_1^N (t_i - t_{oi}(x, y, z) - T_o)^2}{N}} \quad (5)$$

221

222 Where  $t_i$  are the measured arrival times at the  $N$  stations for which the P-picking is available  
 223 and  $\sigma_{T_o}$  is the estimated uncertainty on  $T_o$ .

224 According to the Bayes' theorem, the posterior pdf for P-times is therefore:

$$225 \quad P(\mathbf{m}|\mathbf{d}_1) = \text{const } P(\mathbf{d}_1|\mathbf{m})\rho(\mathbf{m}) \quad (6)$$

226 Lacking prior information about the most likely volumes of seismicity distribution,  $\rho(\mathbf{m})$  can be  
 227 set as the uniform pdf over the volume where earthquakes are expected to occur. This volume  
 228 should correspond to the grid volume for pdf computation.

229 In our recursive Bayesian method, (6) is set as the prior pdf for the posterior pdf of  $\mathbf{m}$  given the  
 230 differential P-amplitudes data-set  $\mathbf{d}_2 \equiv (\Delta PD_{12}, \dots, \Delta PD_{N1})$ . In this case the conditional pdf of  
 231 P-amplitudes is defined:

$$232 \quad P(\mathbf{d}_2|\mathbf{m}) = \text{const } e^{-\frac{\sum_1^{N-1} \sum_2^N (\Delta PD_{ij} - \Delta PD'_{ij}(\mathbf{m}))^2}{2\sigma_A^2}} \quad (7)$$

233 Where  $\Delta PD_{ij}$  is the theoretical differential P-amplitude at stations  $i$  and  $j$  , computed for a  
 234 given model parameter vector  $\mathbf{m}$ , and  $\sigma_A^2$  is a theoretical estimate of the variance for the log of

235 the amplitude ratio. The theoretical values of this quantity are determined using the  
 236 attenuation relations of the form:

$$237 \quad \log PD = A + B M + C \log R$$

238 where R is the source-to-receiver distance and M is the earthquake magnitude. The theoretical  
 239 differential amplitude for stations  $i$  and  $j$  is therefore:

$$240 \quad \Delta PD'_{ij} = \log PD_i - \log PD_j = C (\log R_i - \log R_j)$$

241 Applying the Bayes' theorem and setting  $\rho(\mathbf{m}) = P(\mathbf{m}|\mathbf{d}_1)$ , the posterior pdf for differential P-  
 242 amplitudes can be written as:

$$243 \quad P(\mathbf{m}|\mathbf{d}_2, \mathbf{d}_1) = \text{const } P(\mathbf{d}_2|\mathbf{m})P(\mathbf{m}|\mathbf{d}_1)$$

$$244 \quad = \text{const } e^{-\frac{\sum_1^{N-1} \sum_2^N (\Delta PD_{ij} - \Delta PD'_{ij}(\mathbf{m}))^2}{2\sigma_A^2}} e^{-\frac{\sum_1^{N-1} \sum_2^N (\Delta t_{ij} - \Delta \tau_{ij}(\mathbf{m}))^2}{2\sigma_T^2}} \quad (8)$$

245 Finally, we consider the third data-set, the P-polarization measurements at the N available  
 246 stations. In this case we define the conditional pdf as follows:

$$247 \quad P(\mathbf{d}_3|\mathbf{m}) = \text{const } e^{-\frac{\sum_1^N (\alpha_i - \alpha'_i)^2}{2\sigma_\alpha^2}} \quad (9)$$

248 Following the same approach used for the previous data-sets, we can define the posterior pdf  
 249 for P-polarization, which accounts for both differential arrival times and amplitudes:

$$250 \quad P(\mathbf{m}|\mathbf{d}_3, \mathbf{d}_1, \mathbf{d}_2) = \text{const } P(\mathbf{d}_3|\mathbf{m})P(\mathbf{m}|\mathbf{d}_2, \mathbf{d}_1) =$$

$$251 \quad = \text{const } e^{-\frac{\sum_1^N (\alpha_i - \alpha'_i)^2}{2\sigma_\alpha^2}} e^{-\frac{\sum_1^{N-1} \sum_2^N (\Delta PD_{ij} - \Delta PD'_{ij}(\mathbf{m}))^2}{2\sigma_A^2}} e^{-\frac{\sum_1^{N-1} \sum_2^N (\Delta t_{ij} - \Delta \tau_{ij}(\mathbf{m}))^2}{2\sigma_T^2}} \quad (10)$$

252 Equation (10) provides the pdf for the model parameter  $\mathbf{m}$ , given the three-different data-sets.

253 Its numerical computation requires the regular sampling of the discretized volume where

254 earthquakes are expected to occur. The constant in eq. 10 has to be evaluated numerically, but

255 setting the condition:

256 
$$\int P(\mathbf{m}|\mathbf{d}_3, \mathbf{d}_1, \mathbf{d}_2) d\mathbf{m} = 1$$

257 Once (10) is determined, the maximum likelihood solution can be obtained for the earthquake  
 258 location:

259 
$$\mathbf{m}_{BEST} : P(\mathbf{m}_{BEST}|\mathbf{d}_3, \mathbf{d}_1, \mathbf{d}_2) = \max[P(\mathbf{m}|\mathbf{d}_3, \mathbf{d}_1, \mathbf{d}_2)] \quad (11)$$

260  
 261

262 Errors on parameters can be estimated from the cross-section probabilities as defined below:

263 
$$P(\mathbf{m}|\mathbf{d}_3, \mathbf{d}_1^{best}, \mathbf{d}_2^{best})$$

264 
$$P(\mathbf{m}|\mathbf{d}_3^{best}, \mathbf{d}_1, \mathbf{d}_2^{best})$$

265 
$$P(\mathbf{m}|\mathbf{d}_3^{best}, \mathbf{d}_1^{best}, \mathbf{d}_2) \quad (12)$$

266 where  $\mathbf{d}_1^{best}, \mathbf{d}_2^{best}, \mathbf{d}_3^{best}$  are the parameters of the maximum likelihood solution. These pdfs  
 267 allow to measure the maximum likelihood model parameter vector and the interval of  
 268 parameters associated with 31% and 68% significance levels (e.g. the parameter values at the  
 269 31% and 68% level of the cumulative pdf), which corresponds to  $\pm 1\sigma$  case of normal pdfs.

270

271 **3 Inversion strategies for optimizing the real-time computation of posterior and marginal**  
 272 **pdfs**

273 We implemented a software platform written in Python (<https://www.python.org/>) that  
 274 manages the inversion code and is able to simulate the real-time data streaming. The  
 275 computational efficiency is optimized using a multi-parallel computational approach in order to  
 276 process each single station in parallel during the whole simulation. This approach guaranties a  
 277 rapidity in the solution estimation that is generally provided in a time less than 0.5 second  
 278 (source and network lay-outs of the application study illustrated in section 4), that is the usual  
 279

280 packet length during real-time transmission for modern data-loggers. The software pre-  
281 computes the theoretical travel time-table for a distributed grid of sources in the 3D medium  
282 (calculated using The *TauP* Toolkit by *Crotwell et al., 1999*), BAZ and P-amplitudes (through pre-  
283 existing empirical attenuation relationships) in order to minimize the computational cost during  
284 the software runs.

285 When new data are available, the code estimates a new term in the sum at the exponents of  
286 equations (3), (7) and (9). This event triggers a re-calculation of the total probability density  
287 function matrix (equation (10)). Finally, the pdf matrix is used to estimate the maximum  
288 likelihood solution and errors associated with the 31% and 68% level of the cumulative cross-  
289 section pdfs.

#### 290 **4 Retrospective analysis of mainshocks of the 2016-2017 Central Italy Earthquake sequence**

291 The events of the Central Italy seismic sequence that occurred between August 2016 and  
292 January 2017 have been used to test and demonstrate the algorithm performance. From the  
293 whole sequence (about 135 earthquakes) we selected 27 events with moment magnitude  
294 larger than 4.2, being this magnitude range of more interest for EEW applications. We  
295 considered a volume of  $80 \times 100 \times 20 \text{ km}^3$ , which contains the selected events and 63 stations  
296 belonging to the Italian Strong Motion Network (Rete Accelerometrica Nazionale - RAN),  
297 operated by the Dipartimento della Protezione Civile (DPC), and to the Italian National Seismic  
298 Network, operated by the Istituto Nazionale di Geofisica e Vulcanologia (INGV, Fig. 3a). The  
299 data-set includes the mainshocks of the sequence, the Mw 6 Amatrice event occurred on  
300 August, 24 2016, the two M5.9 and M5.4 events occurred on 26 October 2016, located near  
301 Visso (northeast of Norcia), the Mw 6.5 Norcia event on 30 October 2016, and the Mw 5.5,

302 January 18, 2017 earthquake located south of the town of Amatrice (*Chiaraluce et al., 2017*).  
303 The details about the event origin times, locations, magnitudes and number of recording  
304 stations can be found in table S1 in Supporting information (SI).

305 In this work we considered different station/event distributions in order to analyze different  
306 potential scenarios. We simulated the following configurations by downgrading the initial dense  
307 network configuration. In detail:

- 308 - a dense network of 63 stations, station inter-distance of about 20 km, deployed in the  
309 entire target area, and all the selected events (Fig. 3a) (“In-land” scenario);
- 310 - a network of 24 stations located in the western sector of the area, and 23 events  
311 located in the eastern sector of the area (Fig. 3b) (“Off-Shore” scenario);
- 312 - a network of 15 stations deployed along a linear configuration, and 22 events recorded  
313 by a minimum of 4 stations of the linear network (Fig. 3c) (“Linear array” scenario).

314 The first simulated scenario represents a standard network aimed at locating the seismicity  
315 within local distances (<100 km of aperture); the second scenario represents a case of a coastal  
316 network detecting and locating the seismicity occurring off-shore or outside-the-network as in  
317 the case of near-coastal seismicity in Japan or Mexico; the third scenario represents a linear  
318 seismic array aimed at locating the seismicity for early warning application using a set of sensor  
319 deployed following a “barrier” configuration (e.g Western Iberian Peninsula, Mexico coastline)  
320 or along an high speed train rail.

321 The INGV bulletin locations (<http://terremoti.ingv.it/>), obtained by considering the dense INGV-  
322 RSN network, has been chosen as the reference solution, to which we compared the solutions

323 obtained by the three network configurations. For the earthquake locations we used the 1D  
324 crustal velocity model obtained by Lii et al. (2007) for central Italy, parameterized on a 3D grid  
325 with a cell size of  $0.6 \times 0.6 \times 0.8 \text{ km}^3$ .

326 In order to simulate the real-time scenario, the P-arrival times have been obtained by an  
327 automatic picking procedure based on a recursive STA/LTA trigger-based strategy. We used a  
328 STA window of 0.5 s, a LTA window of 5 s and a threshold STA/LTA value of 10. We verified that,  
329 with the chosen picker parameters, the difference between manual and automatic picks were  
330 on average smaller than 0.2 s (see figure S1 of SI).

331 We performed an optimization analysis in order to set properly the standard deviations of the  
332 three variance factors in the probability distribution (i.e.  $\sigma_T$  for differential times,  $\sigma_\alpha$  for P-  
333 polarization and  $\sigma_A$  for P-amplitudes ratios in eq. 10) and the length of the time-windows to be  
334 used to measure the P-peak amplitude and the BAZ from the P-polarization. The choice of the  
335 time-window length has been done considering the requirement for a rapid but reliable  
336 estimate of the parameters. By considering the “in-land” configuration, we constructed the  
337 distributions of the difference between the calculated BAZ and the reference one (e.g the one  
338 obtained by considering the reference INGV bulletin location, see Fig. S2 of SI), and, similarly,  
339 the distribution of the calculated amplitude ratios and the reference one (see Fig. S3 of SI). We  
340 built these distributions by varying the window length between 0.5 s and 3 s and choosing the  
341 one for which the differences were minimized (i.e. 0.5 s for the BAZ and 2 s for the amplitude).  
342 The standard deviations of the chosen distributions were used to infer the variances  $\sigma_\alpha$  and  $\sigma_A$   
343 of the two probability distributions (i.e.  $60^\circ$  for BAZ and 0.4 for the log P-amplitude ratio). For

344 the  $\sigma_T$  of differential times we considered a value of twice (since we used the differential times)  
345 the mean of distribution of the difference between the automatic and the manual P-wave picks  
346 (i.e. 0.3 s, see Fig. S1 of SI).

347 During the simulations all three observed parameters (differential arrival times, back-azimuths  
348 and amplitude ratios) have been used to constrain the earthquake location parameters. Their  
349 measures are available at different times for each record and station. The differential arrival  
350 times are estimated when the time of the first P-reading is available at a minimum of two  
351 stations. The BAZ is estimated 0.5 sec after the P-wave pick, while the amplitude ratios are  
352 estimated 2 sec after that the P-wave picking time is available at a minimum of two stations.

353 As an example of the algorithm operation, Figure 4 shows the temporal evolution of the  
354 predicted hypocenter location (i.e. epicentral location, depth and origin time) of the Mw 4.2  
355 event occurred on 31 October 2016 in the station configuration “in-land”. Panels a-b-c show the  
356 evolution of the location accuracy, defined as the deviation of epicentral (Fig. 4a) and vertical  
357 (Fig. 4b) location and origin time (Fig. 4c) from the reference value of INGV revised bulletin,  
358 with the time measured from origin. Panel d displays the flow of information as a function of  
359 the time from the origin of the event, showing when arrival times, BAZ and amplitude ratios are  
360 available during the simulation. Once the first two picks are available, after about 2.6 s from the  
361 origin time, the first location is provided. The location accuracy improves with the time due to  
362 the addition of new data, but already after 4.1 s from the origin time, with only four available  
363 picks and the integration of 3 BAZs and 1 amplitude ratio, the predicted location is within a few  
364 kilometers and the origin time is within 0.2 s from the reference one.

365 Figure 5 shows, for the same event, the normalized probability and its cross-sections in  
366 correspondence of the maximum, for four different times indicated by the red numbers in  
367 panels a-b-c of Fig. 4. The decreasing in time of location errors (error bars in Fig. 4a-b) indicates  
368 that the probability distribution is increasingly narrow and peaked around the reference  
369 location (Fig.5). Figure 6 shows, as an aggregate plot, the location accuracy for all the analyzed  
370 events in each tested configuration (In-land, Off-shore and Linear array) as a function of the  
371 number of available stations. The grey dashed lines in each panel are drawn in correspondence  
372 of the 16th and the 84th percentile of the distribution (i.e. within  $1\sigma$ ). From this figure it is  
373 possible to understand how the system can produce stable and reliable estimates of the  
374 earthquake parameters as a function of the amount of data in the different scenarios. The  
375 availability of data as a function of time from the event origin strongly depends on the  
376 station/source configuration.

377 With reference to results obtained for the In-land configuration, the hypocenter locations of all  
378 the considered events are well constrained (i.e. within 5 km from the reference location)  
379 starting from the very first estimates, with less than 6 stations, within the first 5-6 seconds after  
380 the event origin time (see Fig. 6a-c and Fig. S4 in SI). The epicentral and vertical errors (Fig. 6b-  
381 d) decrease in accordance with the decrease of location deviation from the reference value.  
382 Finally, the origin times are within 1.5 s from the reference ones with at least 5-6 stations for  
383 most of the events (Fig. 6e).

384 Considering the “Off-Shore” network lay-out of configuration, the location accuracy is smaller  
385 than 5 km with about 6-7 stations for the most of events (Fig. 6f-h), despite of the worst

386 azimuthal coverage of the stations compared to the one in In-land configuration. For 4 events,  
387 data from at least 9 stations must be available in order to have a well constrained epicentral  
388 estimate. As expected, for the considered events the depth is less constrained using the Off-  
389 shore configuration, but with about 7 stations the depth accuracy is within 5 km. Concerning  
390 the origin time, its deviation from the reference one is on average smaller than 1 s with 3-4  
391 stations, i.e. within 4 s from the event origin time. On average, 6-7 P arrival-time readings are  
392 available within 7 s from the origin time (see Fig. S4 in SI).

393 In the “Linear array” scenario, the epicenter and depth locations are within 5 km from the  
394 reference ones with about 7-8 stations for all the considered events. Concerning the origin  
395 time, its median deviation from the reference one is lower than 1 s with at least 4 stations for  
396 all the considered events. On average, P-data from 4 stations are available within 2-3 s and 7-8  
397 stations within 10 s from the origin time (see Fig. S4 in SI).

398 Due to the time-delay at which the different observed parameters are available at each station,  
399 we expect that differential arrival times (early information) are predominant in constraining the  
400 earthquake location at the very beginning of the analysis and the weight of the other  
401 parameters starts to be relevant as soon as few observations are available at more than two  
402 stations. In order to understand the influence on the retrieved solution of the different  
403 parameter data-sets we compare in Fig. 7 and Fig. 8 the accuracy in retrieving the epicenter  
404 location and depth in two specific time windows (i.e. 2 and 4 sec from the first P-wave picking  
405 at the network) by using only differential arrival times (panels a-d-g), the differential arrival  
406 times and back-azimuths (panels b-c-h), and all three observed parameters together (panel c).

407 In order to have a homogeneous metrics to measure the accuracy in the different cases we  
408 computed the *cumulative normalized frequency* of the observed parameter residual distribution  
409 and report the parameter residual value associated with the 68% (red values and dashed lines  
410 in each panel of Fig. 7 and 8) and 95% (blue values and dashed lines in each panel of Fig. 7 and  
411 8) levels of the distribution.

412 This analysis shows that the location accuracy, especially for the epicenter parameter,  
413 significantly improves the solution obtained with differential times only when integrated by  
414 BAZs and amplitude ratios, for all three tested configurations. After 2 seconds from the first P  
415 arrival (Fig. 7), for the network lay-out In-land, the 68% residual epicenter value decreases from  
416 16 km to 4 km, and the value associated with the 95% level decreases from 60 km to 44 km. For  
417 the Off-shore network configuration, the decrease concerned only the 95% residual value, that  
418 passed from 44 km to 24 km. Finally, concerning the Linear array configuration, a clear  
419 improvement in the epicenter location is shown by the decreasing of both the 68% residual  
420 value, which passed from 25 km to 10-12 km, and the 95% value, which passed from 83 km to  
421 44 km. Concerning the depth parameter, the integration of BAZs and amplitude parameters  
422 does not affect the depth accuracy. For the In-land configuration, the 68% and 95% residual  
423 values are, respectively, 5-6 km and 9 km. While, for the Off-shore configuration, the 68% and  
424 95% residual values slightly increase from 7 km to 8 km and from 9 km to 10 km, respectively;  
425 for the Linear array configuration, the 68% residual values passed from 2.5 km to 7 km and the  
426 95% residual values is 10 km.

427 The results of this comparative analysis with the addition of BAZs and amplitude ratios to  
428 differential time were very similar to the ones obtained by the integration of the BAZ alone,  
429 which indicates a relatively high weight of BAZ with respect to amplitude ratio for the real-time  
430 location in the short window of 2 seconds from the first P-pick. Indeed, after 2 seconds from  
431 the first P pick, only few amplitude ratio data are generally available in all the considered  
432 network configurations.

433 Four seconds after the first P-wave arrival time, the differential times are the most influential  
434 parameter for the location in the different network configurations. In fact, the histograms  
435 obtained by using only differential arrival times (Fig. 8a and d) are very similar to the ones  
436 obtained by using the differential arrival times and BAZs (Fig. 8b and e) and by using all three  
437 observed parameters together (Fig. 8c and f). But in the least favorable configuration (i.e., the  
438 “Linear array”), BAZs and amplitude ratios show to be relevant to reduce the uncertainty on the  
439 epicentral location. With the additional use of BAZs and amplitude ratios the 68% epicenter  
440 residual value passes from 13 km to 8 km, while the 95% value from 28 km to 15 km. Beside, in  
441 terms of depth accuracy, the results clearly indicated that, as it is expected, the BAZ usage may  
442 improve the epicenter location but it does not affect the depth. In fact, the histograms of depth  
443 residuals obtained by using only the differential times (Fig. 8a-d-g, left panel) are very similar to  
444 the ones obtained by using the integrated data-set (Fig. 8b-c-h and c-f-i, left panels). Despite  
445 this, the histograms of depth residuals at 4 seconds after the first pick show significant  
446 improvements respect to the ones at 2 seconds after first pick. In details, for the “In-land”  
447 configuration, the 68% residual value decreases from 6 km to 4 km; for the “Off-shore”  
448 configuration the 68% and 95% residual values decrease from 7-8 km to 2 km and from 9-10 km

449 to 5-6 km; for the “Linear array” configuration the 68% and 95% residual values decrease from  
450 7 km to 6 km and from 10-11 km to 9 km.

451 Finally, we compared the performance of the proposed algorithm with that of another method  
452 for real-time earthquake location in regional EEWS, the RTLOC method (Satriano et al., 2008)  
453 implemented in the PRESto EW platform (Satriano et al., 2011). RTLOC is based on the real-  
454 time measures of P-wave differential times at a dense seismic network and uses an  
455 evolutionary and probabilistic approach to provide the maximum likelihood hypocenter  
456 solution as a function of the time from the first recorded P-wave arrival time. It has been tested  
457 with a dense seismic network (ISNet network, 28 stations with average inter-distance of about  
458 10-15 km), providing reliable estimates of earthquake location within 5-6 s from the event  
459 origin (Satriano et al. 2008).

460 We compared the performance of RTLOC and our location method in the case of the “Off-  
461 shore” and “Linear array” network configurations, i.e. the ones in which we expect that the  
462 integration of BAZs and amplitude ratios could improve the location accuracy. We chose for the  
463 comparison the location accuracy at 3 second after the first P pick, so to guarantee at least 3 P  
464 picks, 2-3 BAZs and 1-2 amplitude ratios for each location. The results of the comparison, in  
465 terms of epicentral location and depth accuracy are shown in Fig. 9 for the “Off-shore” (a-b  
466 panels) and “Linear array” (c-d panels) configurations. As for the previous figures, we computed  
467 the *cumulative normalized frequency* of the observed parameter residual distribution and  
468 report the parameter residual value associated with the 68% (red values and dashed lines in  
469 each panel of Fig. 9) and 95% (blue values and dashed lines in each panel of Fig. 9) levels of the

470 distribution. From these values it can be inferred that the presented location method is more  
471 suitable than RTLOC in cases of unfavorable station/event distribution. In fact, with reference to  
472 the “Off-shore” configuration, the value associated with the 68% of epicenter residual decrease  
473 from 5 km of RTLOC (red value in b, left panel) to 3 km of our M-PLOC (red value in a, left  
474 panel), while the depth residual decreases from 4 km (red value in b, right panel) to 3 km (red  
475 value in b, right panel), respectively. The values associated with the 95% of epicenter and depth  
476 residuals increase of 1 km with our technique (blue values in a-b panels). For the “Linear array”  
477 configuration, the value associated with the 68% of epicenter residual decreases from 11 km of  
478 RTLOC (red value in d, left panel) to 6 km of M-PLOC (red value in c, left panel), while the depth  
479 residual remains at 6 km (red value in d-c, right panels) with the two methods. The values  
480 associated with the 95% of epicenter and depth residuals decrease of 1 km with M-PLOC (blue  
481 values in c-d panels).

#### 482 **4 Discussion**

483 The proposed methodology is a real-time location technique suitable to constrain the  
484 hypocenter coordinates and origin time in Earthquake Early Warning applications. The  
485 approach is based on the probabilistic, Bayesian combination of differential arrival times,  
486 amplitude ratios and back-azimuth estimates, which are continuously measured on the  
487 recorded P-wave signals and updated with the passing of time, as new portions of seismograms  
488 and more recording stations in the source area become available.

489 Dedicated algorithms, suitable to work in real-time, have been developed to measure the three  
490 parameters on limited portions of the P-wave signals, when no other source information is

491 available. In principle, the measurement of P-wave arrival times, amplitudes and signal  
492 polarization are relatively simple and do not require sophisticated approaches. When dealing  
493 with real-time applications, however, these measurements become non-trivial and their  
494 accuracy may critically depend on a number of factors, such as the quality of recorded signals  
495 and unknown contaminating effects of the propagation medium. In this context, the combined  
496 use of three parameters can be strongly advantageous to constrain the source location, if these  
497 parameters are correctly measured, as well as largely inconvenient, if incorrect real-time  
498 estimates are used. For example, in the case of a poor signal quality with low signal-to-noise  
499 ratio and in the absence of any other source information allowing to properly set the suitable  
500 parameters (i.e., filters and threshold levels), the real-time, automatic P-wave picking operation  
501 may generate erroneous phase detections, with consequent bias for the whole location  
502 method. Furthermore, a reliable (1D or 3D), pre-defined velocity model is needed for the  
503 computation of theoretical P-wave travel times at the available stations, to be compared with  
504 observed phase arrivals when solving the inverse problem. The real-time measurement of  
505 amplitudes is ideally straightforward, although it is critically dependent on the correct  
506 knowledge of attenuation relationships with distance, used to compare the observed  
507 amplitudes at pairs of stations. Finally, both amplitude and polarization measurements are  
508 sensitive to high frequency heterogeneities and local site amplifications, which are not  
509 accounted for the simplified assumption of a 1D attenuating medium. It is therefore relevant to  
510 get reliable estimates of the uncertainty on real-time measured quantities so to weigh them  
511 when used for location parameter estimation. Our proposed probabilistic approach accounts  
512 for the different uncertainties related to the estimates of differential times, amplitude ratios

513 and back-azimuths, which are taken into account through the variance factors  $\sigma_T^2$ ,  $\sigma_A^2$  and  $\sigma_\alpha^2$  of  
514 the pdf in eq.6. Although we assumed constant variance factors in our analysis, more in  
515 general, these factors could be replaced by single data variances, as inferred from real-time  
516 measurements.

517 The proposed approach is Bayesian in the sense that it provides as output a multi-dimensional  
518 Probability Density Function, evaluated at each time step, starting from the first P-wave  
519 detection. This allows to estimate the maximum likelihood parameters (i.e., the most probable  
520 solution for the hypocentral coordinates and origin time of the event) along with their  
521 uncertainty, that can be used to monitor the progressive convergence of the real-time solutions  
522 toward the final estimates.

523 The combination of different observed quantities ensures redundancy and robustness to the  
524 approach, so that reliable location solutions are retrieved even with a limited number of  
525 available data. Furthermore, one of the key features of the multi-parametric approach used  
526 here is the possibility of assigning a relative weight to each of the 3 parameters through the  
527 variance factors of the pdf (eq.6). A high uncertainty parameter is associated with a nearly flat  
528 and smooth pdf, while a high accurate parameter shows a peaked pdf concentrated around the  
529 most likely parameter value. The variance factors are set from the statistical uncertainty on  
530 times, amplitudes and BAZ, separately, that can be prior estimated through data-driven  
531 analyses. This probabilistic framework has the main advantage of combining different  
532 observables into a single estimator, while letting the best parameter (i.e., the one with smaller  
533 statistical uncertainty) drive the search for the optimal solution.

534 The location method proposed here works with differential observables, which are jointly  
535 measured at pairs of stations at each time. This allows to determine reliable earthquake  
536 locations as soon as two stations have recorded the P-wave arrival and to achieve accurate  
537 solutions when a few seconds of P-wave signals are recorded at few stations (3 to 5).  
538 This is confirmed by the retrospective analysis of mainshocks of the 2016-2017 Central Italy  
539 earthquake sequence, whose results are summarized in Figures 6-9. Overall, after few iterations  
540 the method converges to stable solutions, in terms of both epicentral coordinates and source  
541 depth, as it can be seen from Figure 6. By considering the “In-land” configuration, the epicentral  
542 locations indeed are well constrained (i.e. within 5 km from the reference location) with about  
543 5-6 stations (Fig. 6a), typically 5-6 sec after the event origin in the analyzed cases. With the  
544 same number of stations, the difference in depth estimate with INGV catalogue is nearly stable  
545 around zero, varying between  $\pm 5$  km from the reference estimate (Fig. 6c) and the origin times  
546 are within 1.5 s from the reference ones (Fig. 6e).  
547 Similar results are observed even with a non-optimal coverage of stations (“Off-shore”  
548 configuration), in which, about 6-7 stations are necessary to converge to stable solutions, with  
549 epicentral and depth error smaller than 5 km and a deviation from the reference origin time of  
550 about 1s (Figure 6 f-j). On average, 7 sec after the event origin in the analyzed cases.  
551 The major strength of the proposed approach is the ability of providing correct location  
552 solutions, even in non-optimal network geometries and in unfavorable station distributions.  
553 This is the case of events outside the area covered by the stations, which are distributed only by  
554 one side of the epicenter, or the case of linear arrays. In this last case, for example, standard  
555 location techniques using phase arrival times are often not suitable to constrain the hypocenter

556 position, or provide strongly undetermined solutions, with large uncertainties in both epicentral  
557 position and depth. The combined use of times, amplitudes and BAZ makes the proposed  
558 method suitable to work in disadvantageous conditions of sparse networks, with a limited  
559 number of recording nodes and/or poor azimuthal coverage. Indeed, in the linear array  
560 configuration, the majority of the analyzed events require 7-8 stations, available on average  
561 after 10 s from origin time, to constrain the location solution, both in terms of epicentral  
562 estimates and of source depth and origin time (Fig. 6k-o).

563 A tangible confirmation of the convenient use of three parameters is provided in Fig. 7-8. Here  
564 the differences of epicentral position and depth with respect to the reference solutions, are  
565 compared when using only times (panels a-d-g), times and amplitudes (panels b-e-h) and times  
566 plus amplitudes and BAZs (panels c-f-i), for the three network configurations. The *cumulative*  
567 *normalized frequency* of the residual distributions is characterized by the 68% and 95% levels  
568 (red and blue dashed line, respectively) and the associated difference to the reference  
569 parameter at these levels is also reported in each panel. While for the in-land and for the off-  
570 shore configurations comparable results are obtained with different input parameters, in the  
571 case of a linear array, the joint use of times, amplitudes and BAZ significantly improves the  
572 convergence to the real solution at very short times (Fig.7-8).

573 A relevant result for all the tested network configurations is that the decrease of uncertainties  
574 in real-time estimates (panels b, d, g, i, l, n in Fig. 6) is associated to the convergence of the  
575 solution toward the reference parameters (panels a, c, f, h, k, m in Fig. 6) from INGV catalogue  
576 obtained in optimal distance and azimuth coverage conditions. This suggests the possibility to

577 use in real-time the estimated parameter uncertainty vs the station number (or time from the  
578 event origin) to assess the reached convergence to the final solution and stop further iterations.  
579 As compared to another real-time location method, RTLOC (Satriano et al, 2008), which uses  
580 only the P-wave arrival times, the proposed multi-parametric approach turns out to provide  
581 better constrained location solutions, since the very first available data. This is especially true in  
582 the unfavorable case of the “Linear array” distribution, where the joint use of three parameters  
583 strongly reduces the difference to the reference solutions, as it can be seen from the  
584 *cumulative normalized frequency* (and its 68-95% levels) of Fig. 9.  
585 From the computational point of view, the proposed approach is efficient and optimized for  
586 running in real-time applications, where the earthquake location has to be retrieved in a very  
587 short time (around 1 sec) after data acquisition. The methodology proposed here does not  
588 require complex computational structures and can be easily integrated in other regional,  
589 network-based EEW approaches.  
590

591 **5 Conclusions**

592 In this article we propose a new method for earthquake location to be implemented in  
593 network-based earthquake early warning systems. The main conclusions of our study are:

- 594 - the method combines in a Bayesian probabilistic framework three observed quantities,  
595 measured at a minimum of two stations, in a time window of 0.5-2 sec width, the P-  
596 wave differential arrival time, the P-wave amplitude ratio and the back-azimuth  
597 orientation;
- 598 - the method is evolutionary since it updates the estimates of the earthquake  
599 coordinates, depth and origin time along with their uncertainties as the P-wavefront  
600 propagates through a dense network of receivers;
- 601 - the relative weighting of the different parameters is implicitly accounted by their  
602 conditional pdf where the variance factors are set from the statistical uncertainty on  
603 times, amplitudes and BAZ, separately, that can be prior estimated through data-driven  
604 analyses;
- 605 - the method has been validated through a retrospective analysis of the mainshocks of  
606 the 2016-2017 Central Italy sequence, considering three different sub-networks that  
607 simulated the typical “In-land”, “Off-shore” and “Linear array” network lay-outs;
- 608 - Results show that precise solutions are obtained within 2-4 sec from the first recorded  
609 P-wave and that the integration of the three observed quantity allow to improve the  
610 accuracy of the solution, relative to the use of arrival times only, especially in non-  
611 optimal and unfavorable network configurations;

612 - As compared to other EW location method, (e.g. RTLoc in PRESTo platform), which uses  
613 only the P-wave arrival times, the proposed multi-parameteric approach turns out to  
614 provide better constrained location solutions, since the very first available data.

## 615 **Acknowledgments and Data**

616 Accelerograms used in this study were collected from the Italian Accelerometric Archive (ITACA) 2.0  
617 (Pacor et al., 2011) at <http://itaca.mi.ingv.it>.  
618 The Istituto Nazionale di Geofisica e Vulcanologia [INGV] catalog is available at <http://cnt.rm.ingv.it>. For  
619 information on the INGV network, see <http://cnt.rm.ingv.it/instruments/network/IV>; for information on  
620 the Engineering Strong Motion (ESM) database, see <http://esm.mi.ingv.it/>.  
621 We are beginning to archive the data derived from our analyses in an appropriate repository (Figshare)  
622 but the process is not complete.

623

624

## 625 **References**

626

627 Allen, R. M., 2007: The ElarmS earthquake warning methodology and application across  
628 California. In: Gasparini P, Zschau J (eds.), *Seismic Early Warning*. New York: Springer, pp. 21–  
629 44.

630

631 Chiaraluce, L., Di Stefano, R., Tinti, E., Scognamiglio, L., Michele, M., Casarotti, E., Cattaneo, M.,  
632 De Gori, P., Chiarabba, C., Monachesi, G., Lombardi, A., Valoroso, L., Latorre, D. and S. Marzorati  
633 (2017), The 2016 Central Italy Seismic Sequence: A First Look at the Mainshocks, Aftershocks,  
634 and Source Models, *Seism. Res. Lett.* ; 88 (3): 757–771. doi:  
635 <https://doi.org/10.1785/0220160221>.

636

637 Crotwell, H. P., Owens, T.J. and J. Ritsema (1999), The TauP Toolkit: Flexible Seismic Travel-time  
638 and Ray-path Utilities, *Seism. Res. Lett.*, 70 (2): 154–160. <https://doi.org/10.1785/gssrl.70.2.154> )

639

640 De Landro, G., Amoroso, O., Stabile, T.A., Matrullo, E., Lomax, A. and A. Zollo (2015), High-  
641 precision differential earthquake location in 3-D models: evidence for a rheological barrier  
642 controlling the microseismicity at the Irpinia fault zone in southern Apennines, *Geophys. J. Int.*,  
643 203, 3, 1821–1831, <https://doi.org/10.1093/gji/ggv397>.

644

645 Eisermann, A.S, Ziv, A. and G. H. Wust-Bloch (2015), Real-Time Back Azimuth for Earthquake  
646 Early Warning, *Bull. Seism. Soc. Am.*, 105 (4): 2274–2285,  
647 doi:<https://doi.org/10.1785/0120140298>

648

- 649 Font, Y., Kao, H., Lallemand, S., Liu, C-S. and L-Y Chiao (2004), Hypocentral determination  
650 offshore Eastern Taiwan using the Maximum Intersection method. *Geophys. J. Int.*, 158, 655-  
651 675.  
652
- 653 Horiuchi, S., Negishi, H., Abe, K., Kamimura, A. and Y. Fujinawa (2005), An Automatic Processing  
654 System for Broadcasting Earthquake Alarms. *Bull. Seism. Soc. Am.*; 95 (2): 708–718.  
655 doi:<https://doi.org/10.1785/0120030133>.  
656
- 657 Lii, H., Michelini, A., Zhu, L., Bernardi, F. and M. Spada, Crustal Velocity Structure in Italy from  
658 Analysis of Regional Seismic Waveforms (2007), *Bull. Seism. Soc. Am.* , 97 (6): 2024–2039. doi:  
659 <https://doi.org/10.1785/0120070071>.  
660
- 661 Liu, A. and M. Yamada (2014), Bayesian Approach for Identification of Multiple Events in an  
662 Early Warning System. *Bull. Seism. Soc. Am.*, 104 (3). pp. 1111-1121. ISSN 0037-1106.  
663
- 664 Lockman, A.B., and R.M. Allen (2005). Single-Station Earthquake Characterization for Early  
665 Warning. *Bull. Seism. Soc. Am.* , 95(6):2029-2039 DOI: 10.1785/0120040241  
666
- 667 Lomax, A., Virieux, J., Volant, P. and C. Berge (2000), Probabilistic earthquake location in 3D and  
668 layered models: Introduction of a Metropolis-Gibbs method and comparison with linear  
669 locations, in *Advances in Seismic Event Location* , Thurber, C.H., and N. Rabinowitz (eds.),  
670 Kluwer, Amsterdam, 101-134.  
671
- 672 Lomax, A., Michelini, A. and A. Curtis (2009), Earthquake Location, Direct, Global-Search  
673 Methods, in *Complexity In Encyclopedia of Complexity and System Science*, Part 5, Springer,  
674 New York, pp. 2449-2473, doi:10.1007/978-0-387-30440-3.  
675
- 676 Lomax, A (2005), A Reanalysis of the Hypocentral Location and Related Observations for the  
677 Great 1906 California Earthquake. *Bulletin of the Seismological Society of America* ; 95 (3): 861–  
678 877. doi: <https://doi.org/10.1785/0120040141>.  
679
- 680 [Lomax, A., Satriano, C.](#) and M. Vassallo (2012), Automatic picker developments and  
681 optimization: FilterPicker - a robust, broadband picker for real-time seismic monitoring and  
682 earthquake early-warning, *Seism. Res. Lett.* , 83, 531-540, doi: 10.1785/gssrl.83.3.531.  
683
- 684 Nakamura, Y. (1988), On the urgent earthquake detection and alarm system (UrEDAS), in  
685 *Proceedings of Ninth World Conference on Earthquake Engineering*, Paper# 13-2-12, vol. VII,  
686 pp. 673–678, Tokyo-Kyoto, 2–9 Aug.  
687
- 688 Noda, S., Yamamoto, S., Sato, S. et al. (2012) Improvement of back-azimuth estimation in real-  
689 time by using a single station record. *Earth Pla. Space* 64, 305–308 (2012).  
690 <https://doi.org/10.5047/eps.2011.10.005>  
691

692 Satriano, C., Lomax, A. and A. Zollo (2008), Real-time evolutionary earthquake location for  
693 seismic early warning. *Bull. Seism. Soc. Am.* 2008; 98(3):1482–94, doi:10.1785/0120060159.

694

695 Satriano, C., Elia L., Martino C., Lancieri M., Zollo A. and Iannaccone G. (2011), PRESTo, the  
696 earthquake early warning system for Southern Italy: concepts, capabilities and future  
697 perspectives. *Soil Dyn. Earthq. Eng.* 31 (2), 137–153, doi: 10.1016/j.soildyn.2010.06.008.

698

699 Wu, S., Yamada, M., Tamaribuchi, K. and J. L. Beck, Multi-events earthquake early warning  
700 algorithm using a Bayesian approach, *Geophys. J. Int.*, Volume 200, Issue 2, February, 2015,  
701 Pages 791–808, <https://doi.org/10.1093/gji/ggu437>

702

### 703 **Figure captions**

704

705 **Figure 1.** Figure shows an example of parameters estimation on records of the same  
706 earthquake at station RTL and FOS. a) The BAZ was estimated as mean value of  
707 estimations in a time window of 0.5 sec after the P-wave arrival time (red signal in  
708 panels b and c) only for the samples that exceed a prefixed signal-to-noise threshold of  
709 3. In this case, the Best estimation of BAZ is about 51°, very close to the real one of 47°.  
710 d) and e) are the vertical components of velocity, derived from the integration of  
711 acceleration, where two different P wave arrivals are detected and the  $\Delta T$  estimation is  
712 provided. After 2 sec of P picks two different estimation of  $P_v$  (red circles) are provide to  
713 evaluate the differential amplitude in order to integrate the information of  $\Delta T$  and Baz  
714 in the inversion algorithm.

715 **Figure 2** A block diagram of software platform that represents the workflow of  
716 algorithm from parameters estimation to the final solution.

717 **Figure 3.** Map of the station/event distributions used in the analyzed scenarios. a) A  
718 dense network of 63 stations (grey triangles), station inter-distance of about 20 km,  
719 deployed in the entire target area, and all the selected 27 events (black stars). The Mw  
720 6.5, 2016 Norcia earthquake in Central Italy is included (red star). b) A network of 24  
721 stations (grey triangles) located in the western sector of the area, and 23 events located  
722 in the eastern sector of the area (black stars). The Norcia earthquake was included (red  
723 star). c) A network of 15 stations (gray triangles) deployed along a linear configuration,  
724 and 22 events recorded by a minimum of 4 stations of the linear network (black stars).  
725 The Norcia earthquake is included (red star).

726 **Figure 4.** M-PLOC location example. a) Temporal evolution of M-PLOC performance in  
727 terms of difference between obtained and reference epicentral location (a), depth  
728 location (b) and origin time determination (c). The d panel indicated the number of  
729 parameters available for the correspondent location. The grey curve is representative of  
730 the P picks availability, the red curve of the BAZs and the turquoise of the amplitudes.  
731 The red numbers indicate the time at which are obtained the correspondent probability  
732 distributions in Figure 5.

733 **Figure 5.** Relative location probability distribution after 2.6 s (a), 3.1 s (b), 4.1 s (c) and  
734 9.1 s (d) from origin time. For each snapshot, was shown the normalized location  
735 probability, the stations used for the location (black triangles), the reference event  
736 location (white star) and the predicted hypocenter (black star). The dashed vertical and  
737 horizontal lines represent the uncertainty intervals in the three directions. Moreover, in  
738 the other three panels was shown the location probability in the cross-sections in  
739 correspondence of the optimal solution (maximum of probability).

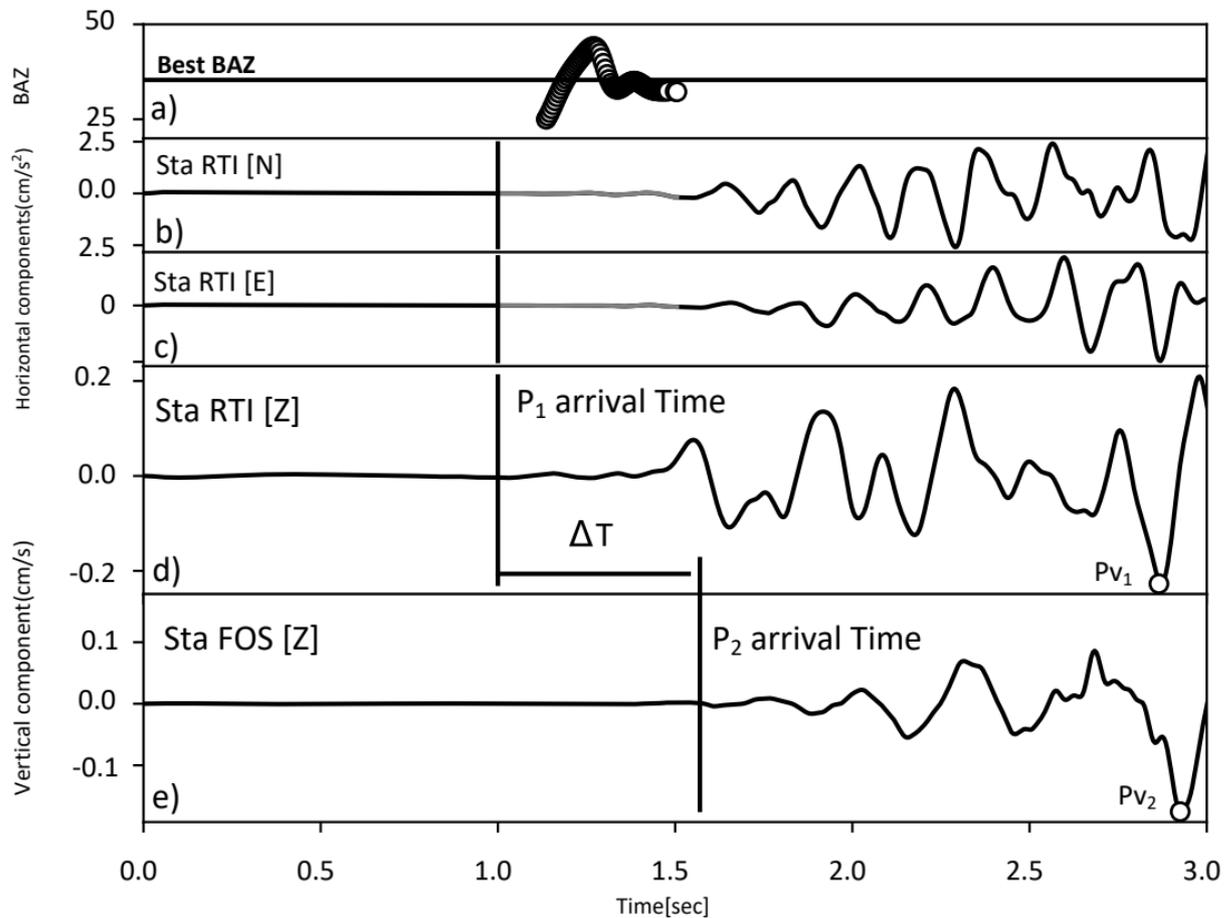
740 **Figure 6.** Aggregate plot of the evolution of M-PLOC location accuracy as a function of  
741 the number of stations for the three configurations: "In-land" (a-e); "Off-shore" (f-j) and  
742 "Linear array" (k-o). The panels a-f-k represent the epicentral location accuracy, the  
743 panels b-g-l the epicentral error evolution, the pane c-h-m the depth location accuracy,  
744 the panels d-i-n the depth error evolution and the panels e-j-o the origin time  
745 estimation accuracy. The light grey area in each panel represents the curve dispersion.  
746 The grey dashed lines in each panel are drawn in correspondence of the 16th and the  
747 84th percentile of the distribution (i.e. within one sigma).

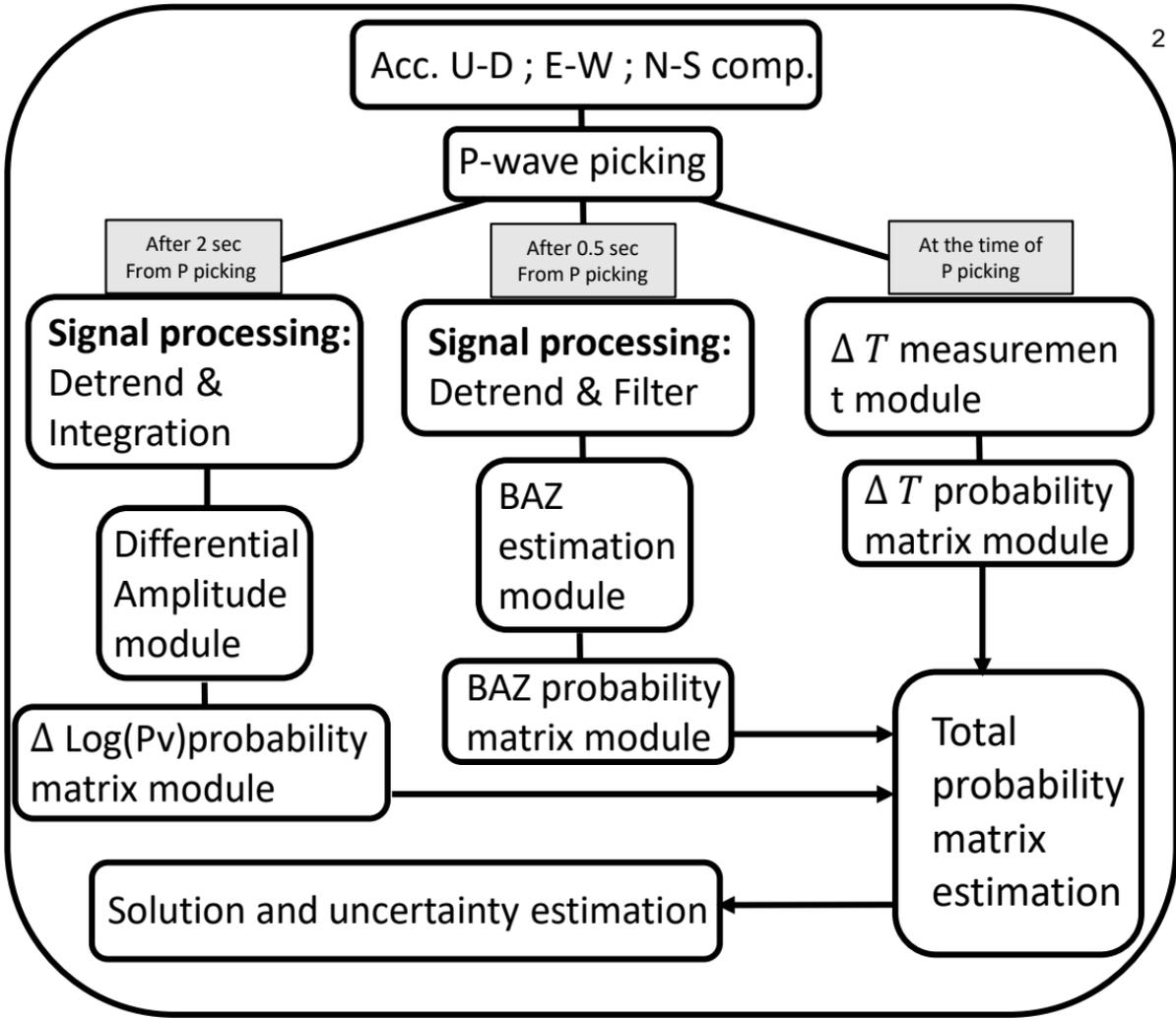
748 **Figure 7.** Comparison between the algorithm performance after 2 second from the first P  
749 pick by using different data type combinations: only differential times (a, d, g),  
750 differential times and BAZs (b, e, h) and differential times plus differential amplitude  
751 and BAZs (c, f, i). The epicentral and in-depth location accuracy (difference between the  
752 estimated and the reference one) is shown also for the different station/event  
753 configurations: "In-land" (a, b, c), "Off-shore" (d, e, f) and "Linear array" (g, h, i). In each  
754 panel, the dark grey curve is the cumulative histogram of the distribution, and the  
755 dashed vertical lines represent the values correspondent to the 68% (red) and the 95%  
756 (blue) of the cumulative histogram.

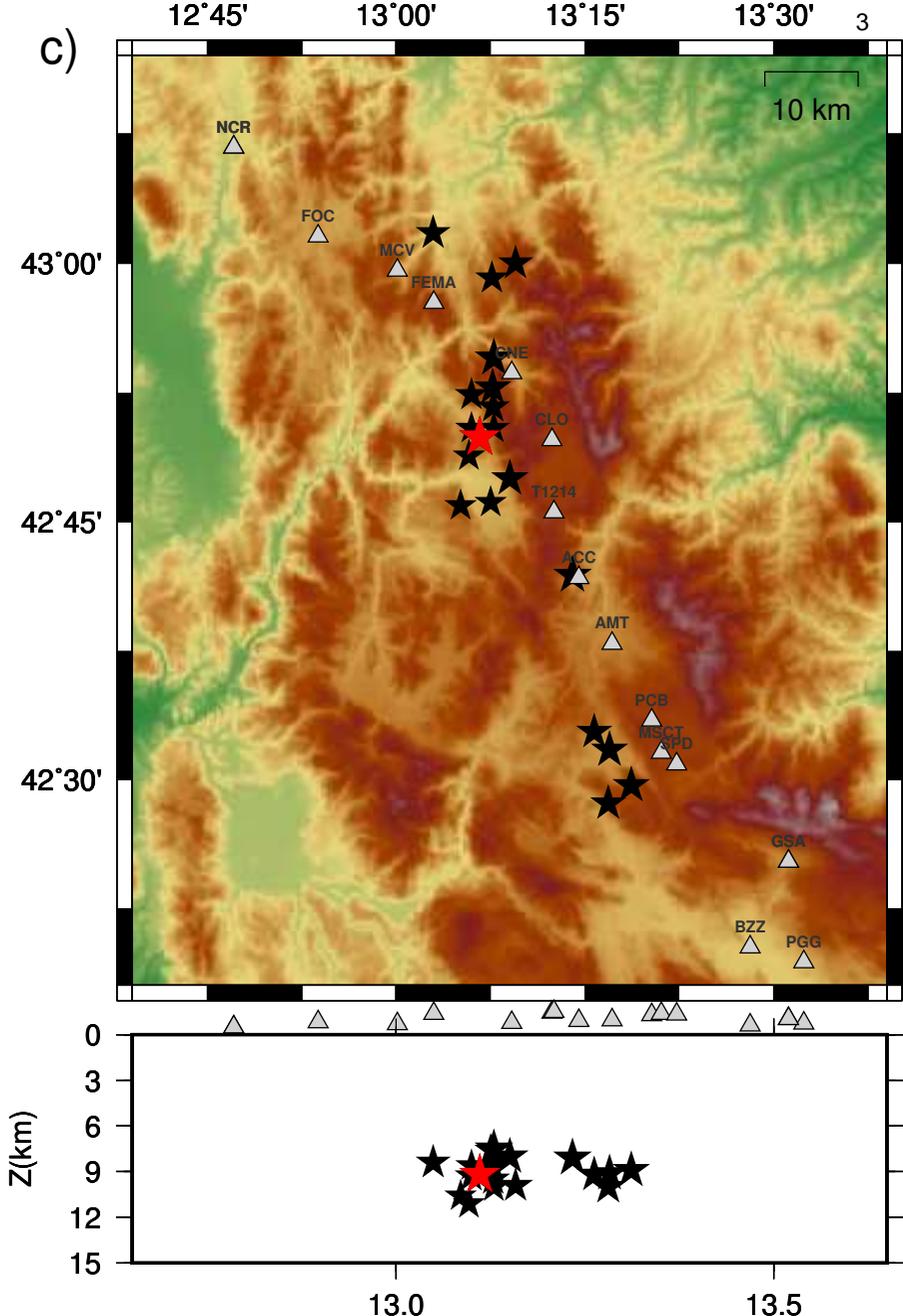
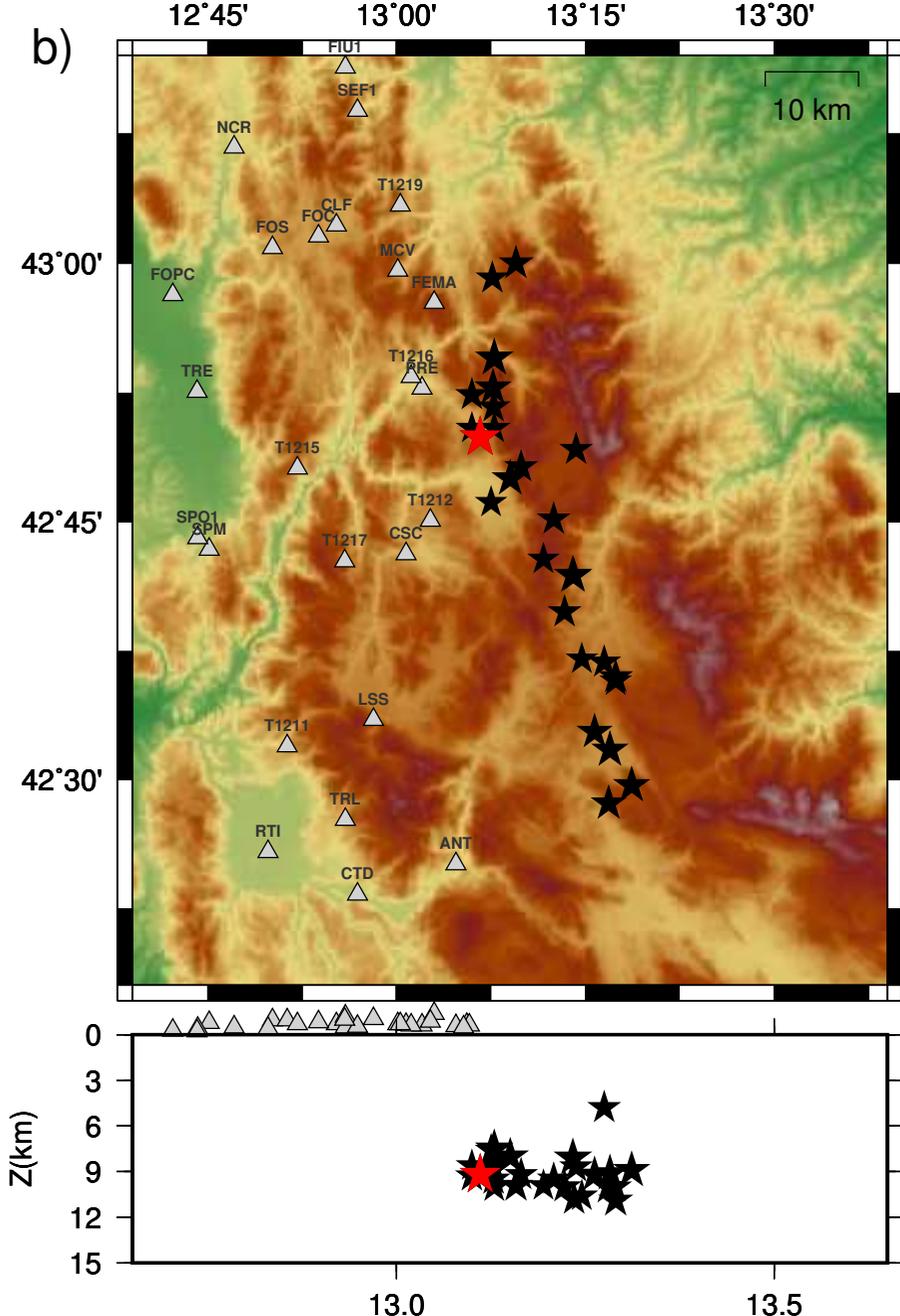
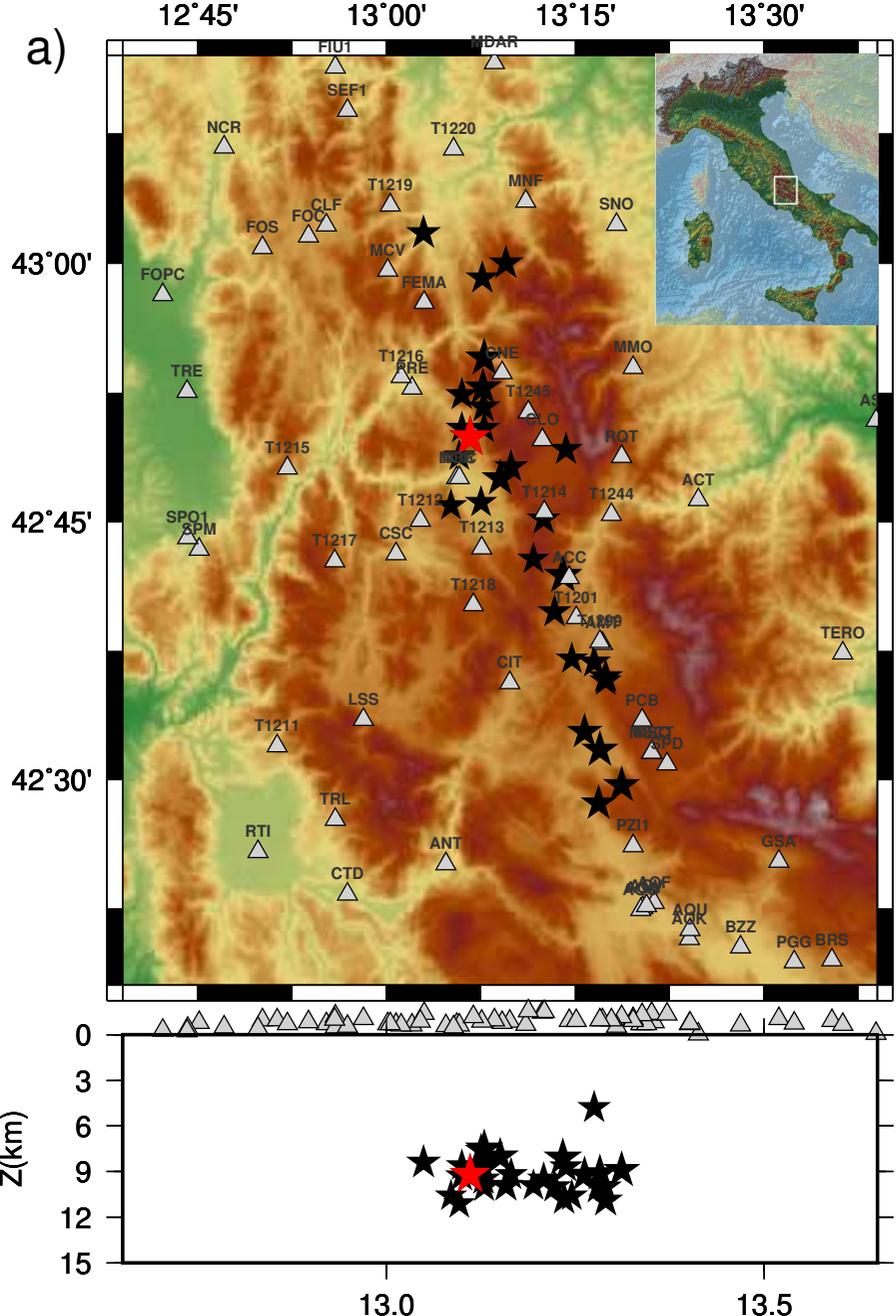
757 **Figure 8.** Comparison between the algorithm performance after 4 second from the first  
758 P pick by using different data type combinations: only differential times (a, d, g),  
759 differential times and BAZs (b, e, h) and differential times plus differential amplitude  
760 and BAZs (c, f, i). The epicentral and in-depth location accuracy (difference between the  
761 estimated and the reference one) is shown also for the different station/event  
762 configurations: "In-land" (a, b, c), "Off-shore" (d, e, f) and "Linear array" (g, h, i). In each  
763 panel, the dark grey curve is the cumulative histogram of the distribution, and the  
764 dashed vertical lines represent the values correspondent to the 68% (red) and the 95%  
765 (blue) of the cumulative histogram.

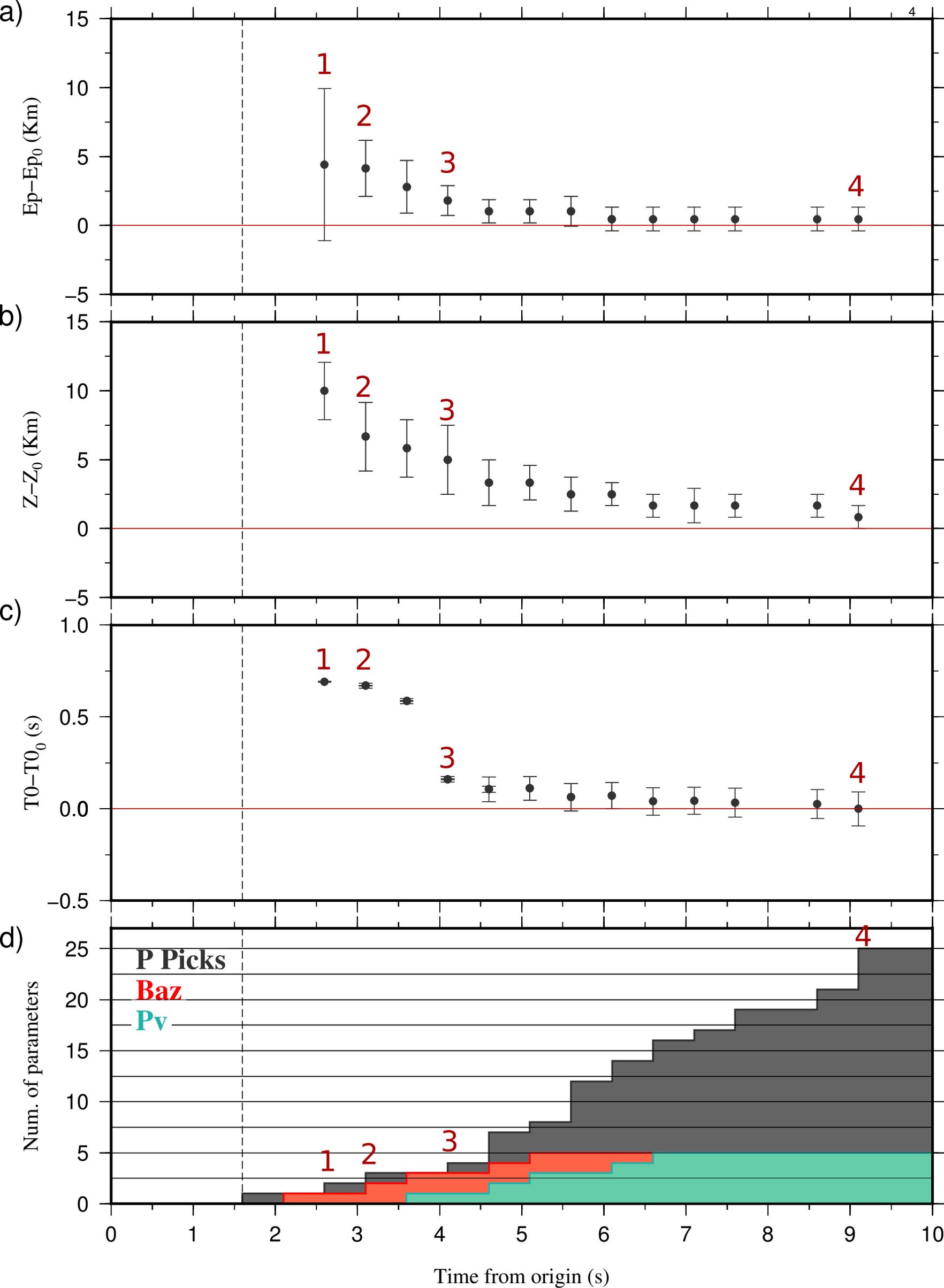
766

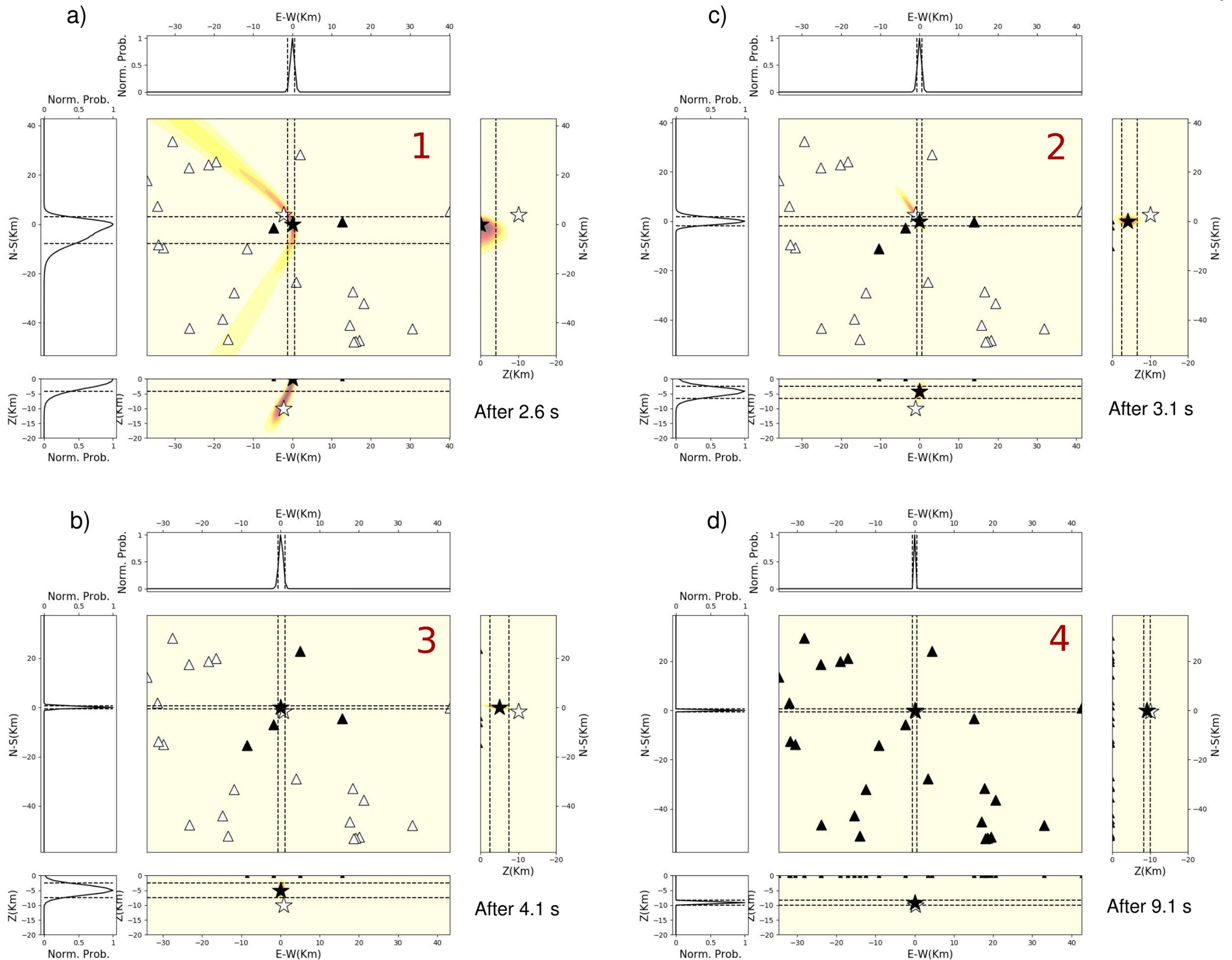
767 **Figure 9.** Comparison between the performance at 3 seconds after the first P pick of M-  
768 PLOC (a, b) and RTLOC (c, d). The epicentral and depth location accuracy (difference  
769 between the estimated and the reference one) is shown for the "Off-shore" (a-c) and  
770 "Linear array" (b-d) configurations. In each panel, the dark grey curve is the cumulative  
771 histogram of the distribution, and the dashed vertical lines represent the values  
772 correspondent to the 68% (red) and the 95% (blue) of the cumulative histogram.

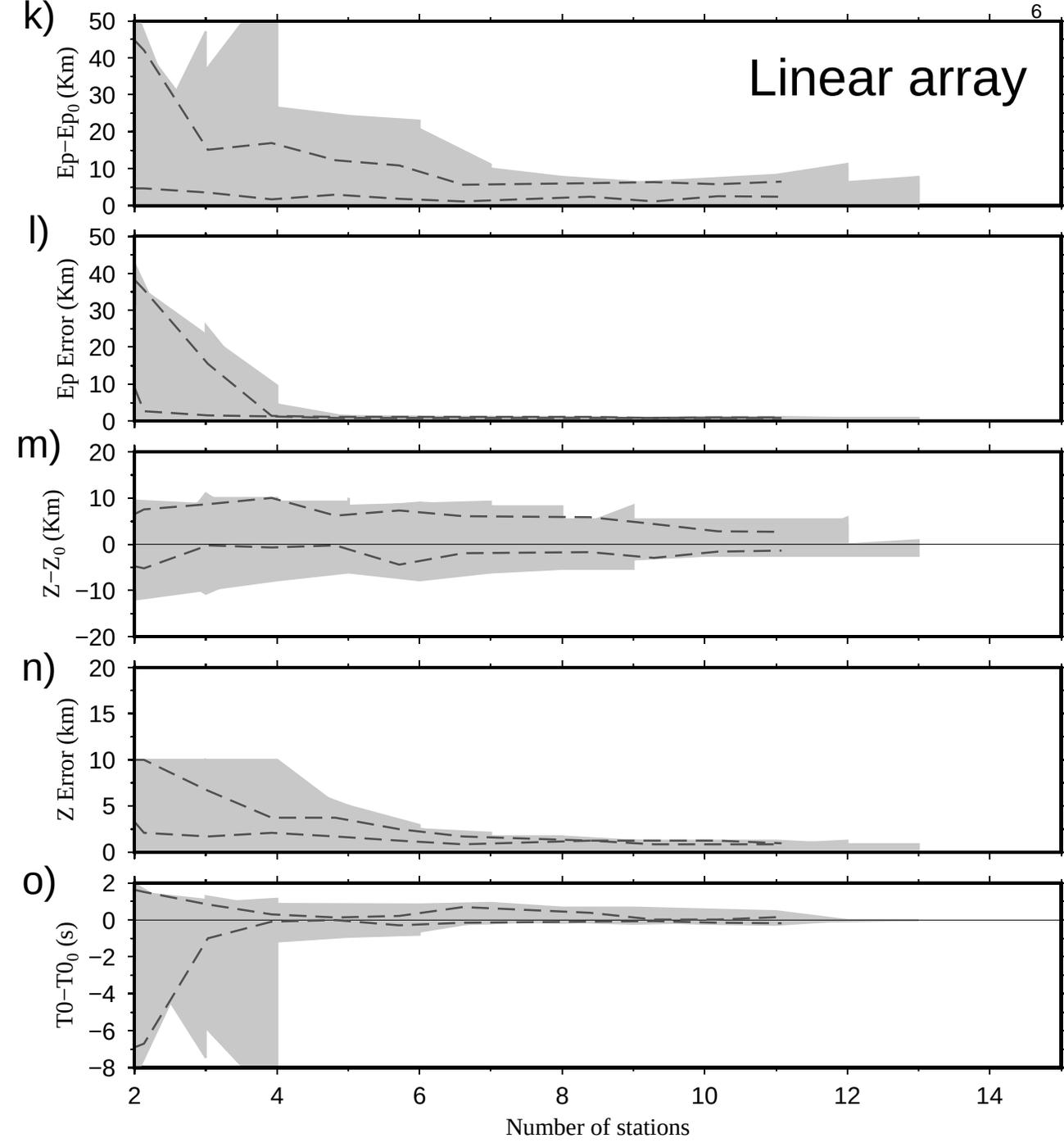
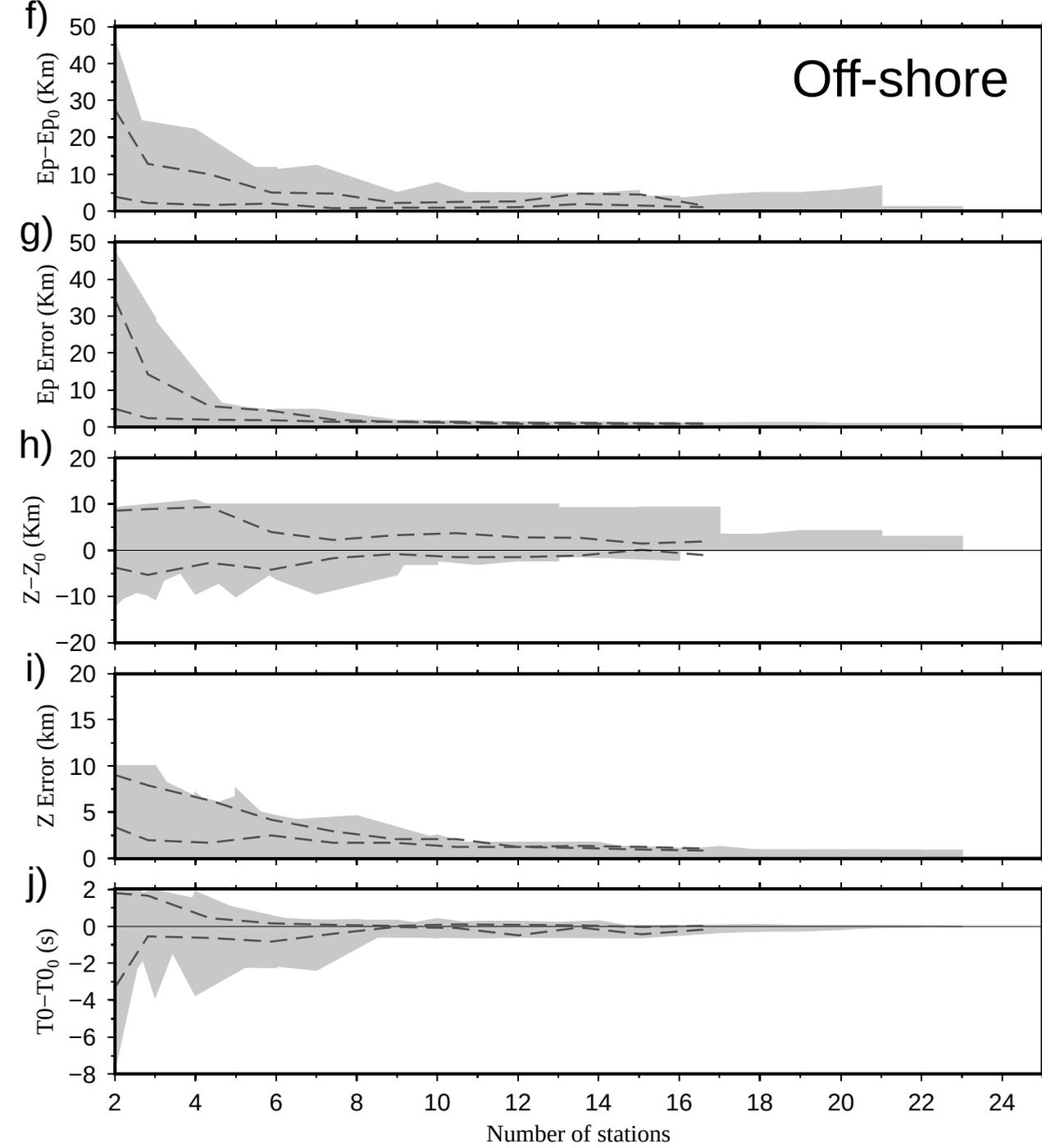
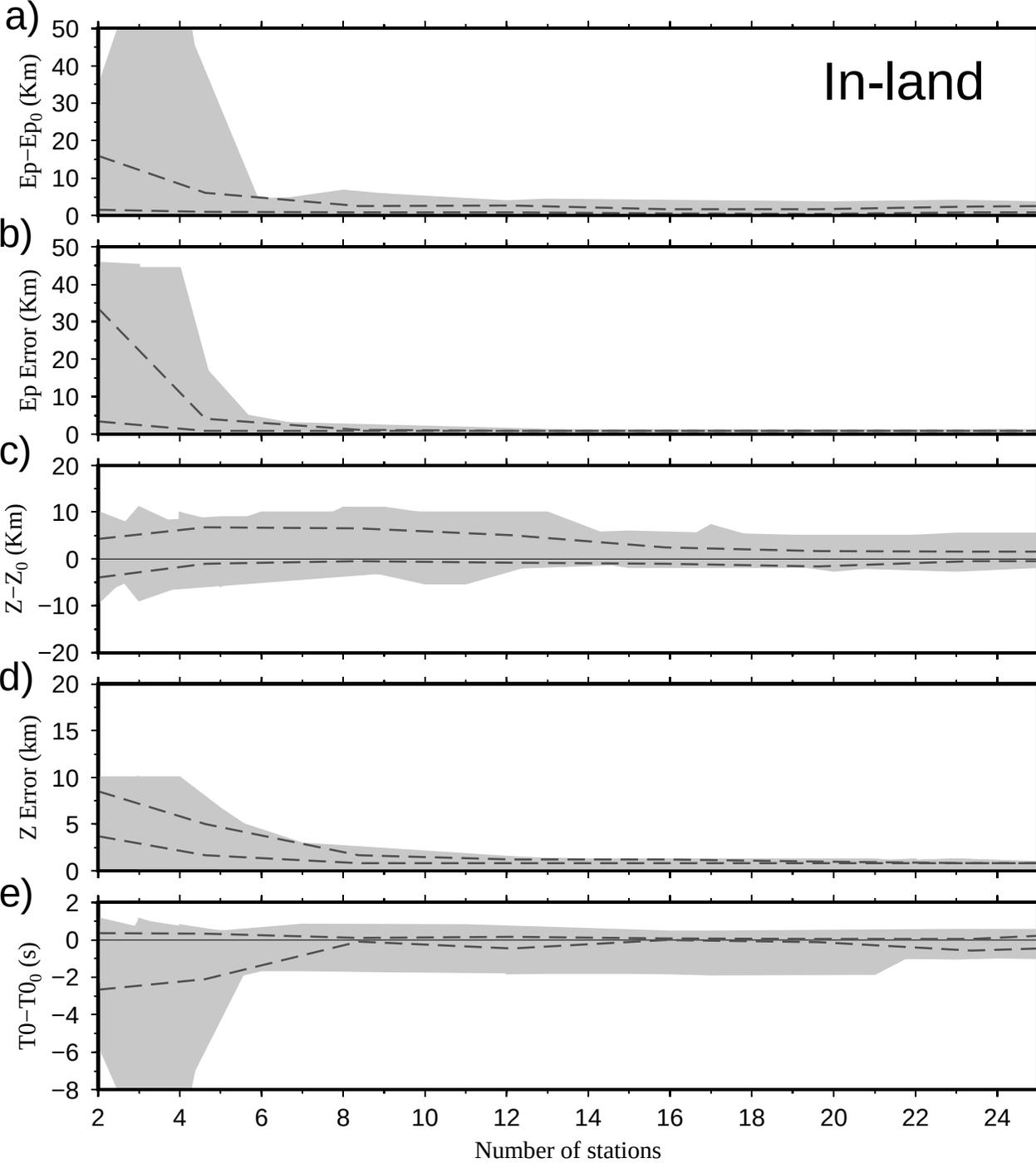




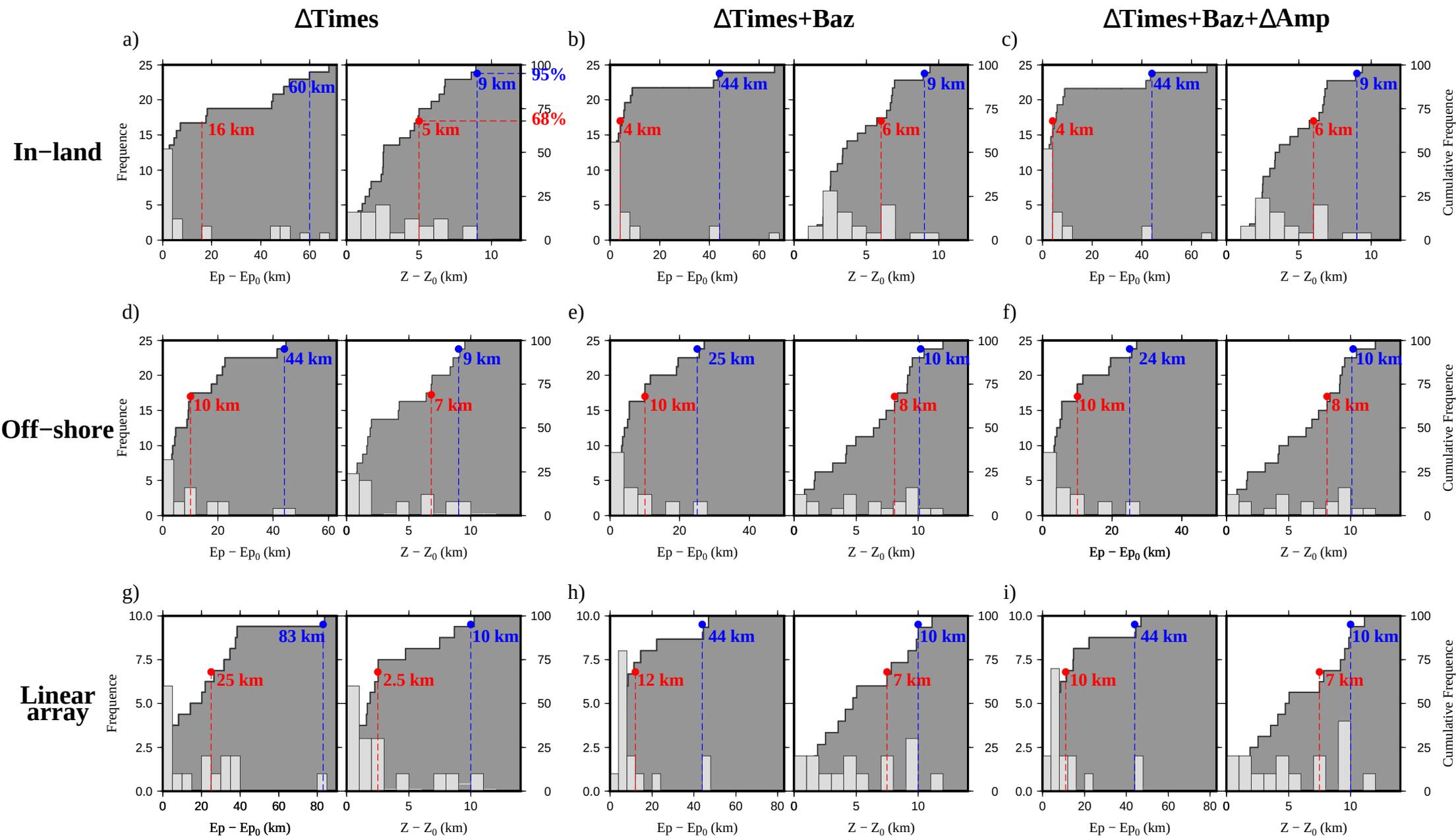




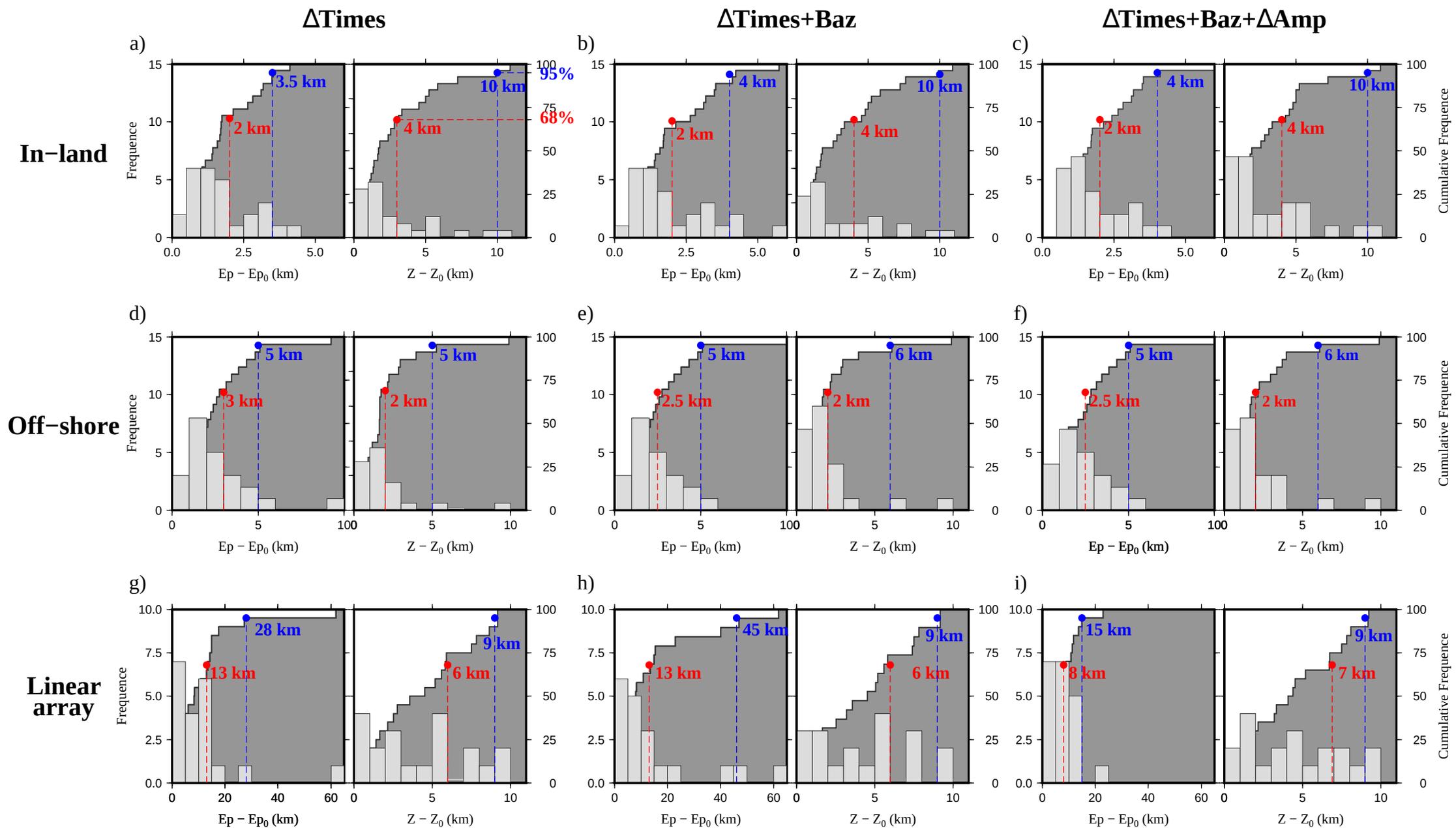


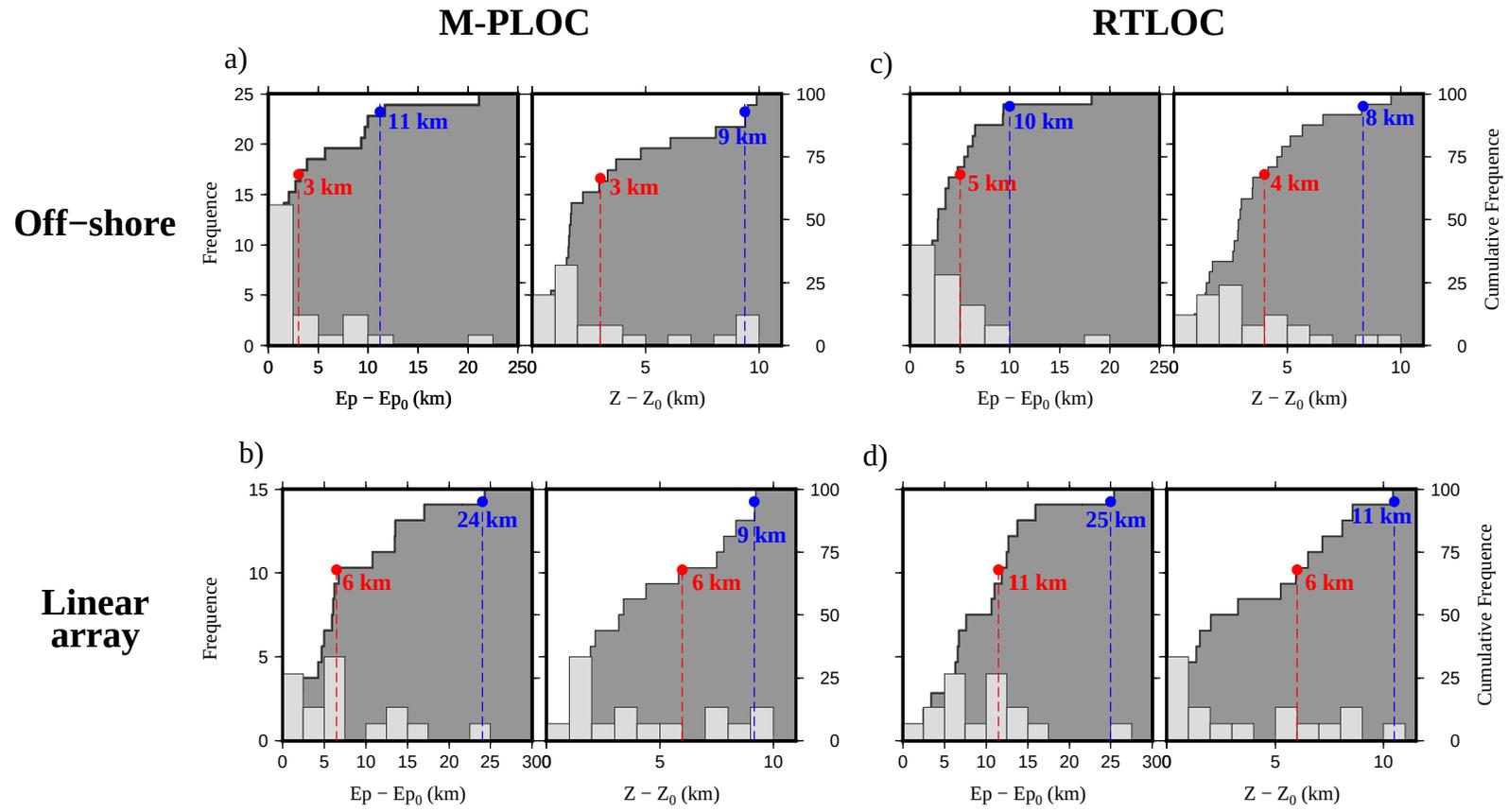


## After 2 s



# After 4 s





**Supporting Information of  
“A Bayesian Method for Real-time Earthquake Location  
Using Multi-Parameter Data”**

**Aldo Zollo, Alessandro Caruso, Grazia de Landro, Simona Colombelli and Luca**

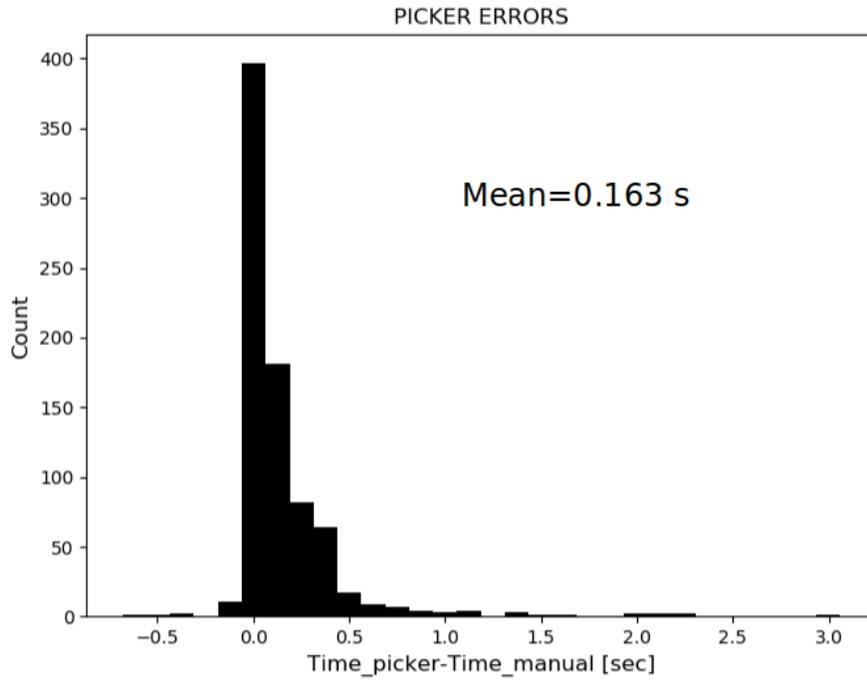
**Elia**

Dept. Physics E. Pancini, University of Naples Federico II.

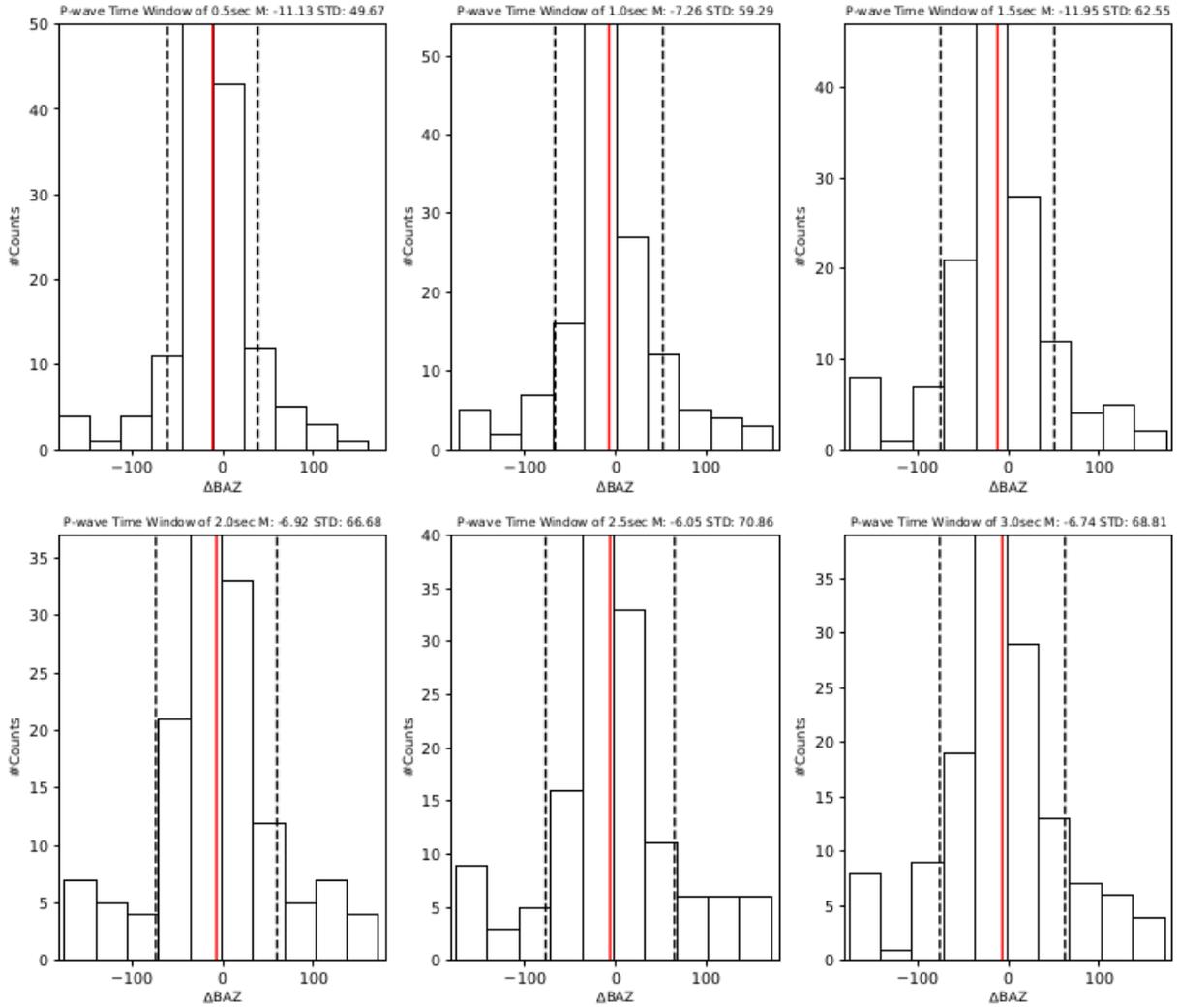
Corresponding author: Aldo Zollo ([aldo.zollo@unina.it](mailto:aldo.zollo@unina.it))

Date	Time	Lat.	Lon.	Depth (Km)	M	# Stat. A	# Stat. B	# Stat. C
20160824	01:36:32	42.698	13.233	8.1	6.0	38	15	10
20160824	01:56:00	42.614	13.275	4.8	4.4	15	8	5
20160824	02:33:29	42.792	13.151	8.0	5.4	36	16	9
20160824	03:40:11	42.617	13.245	10.6	4.2	17	11	4
20160824	04:06:50	42.769	13.125	7.6	4.3	20	14	/
20160824	17:46:09	42.663	13.222	10.0	4.4	17	10	5
20160825	03:17:16	42.753	13.208	9.5	4.5	16	10	5
20160825	12:36:05	42.596	13.290	10.0	4.3	20	11	/
20160826	04:28:25	42.600	13.290	10.9	4.8	19	10	5
20160828	15:55:35	42.820	13.238	8.7	4.4	17	10	/
20161026	17:10:36	42.880	13.127	8.7	5.4	57	23	14
20161026	19:18:05	42.909	13.129	7.5	5.9	55	21	13
20161026	21:42:01	42.861	13.128	9.5	4.6	27	13	7
20161027	03:50:24	42.986	13.127	8.9	4.4	19	10	/
20161027	08:21:45	42.873	13.100	9.3	4.4	24	10	6
20161029	16:24:33	42.814	13.096	11.1	4.2	27	/	7
20161030	06:40:17	42.832	13.111	9.2	6.5	63	25	15
20161030	07:13:05	42.698	13.235	10.8	4.5	26	10	/
20161030	13:34:54	42.803	13.165	9.2	4.5	20	12	5
20161031	03:27:40	42.766	13.085	10.6	4.2	28	/	7
20161031	07:05:44	42.841	13.129	10.0	4.2	25	13	5
20161101	07:56:39	43.000	13.158	9.9	4.7	34	13	8
20161103	00:35:01	43.029	13.049	8.4	4.8	38	/	9
20170118	09:25:40	42.547	13.262	9.2	5.1	46	20	13
20170118	10:14:09	42.529	13.282	9.1	5.5	48	20	13
20170118	10:25:23	42.494	13.311	8.9	5.4	45	20	12
20170118	13:33:36	42.477	13.281	10.0	5.0	49	20	13

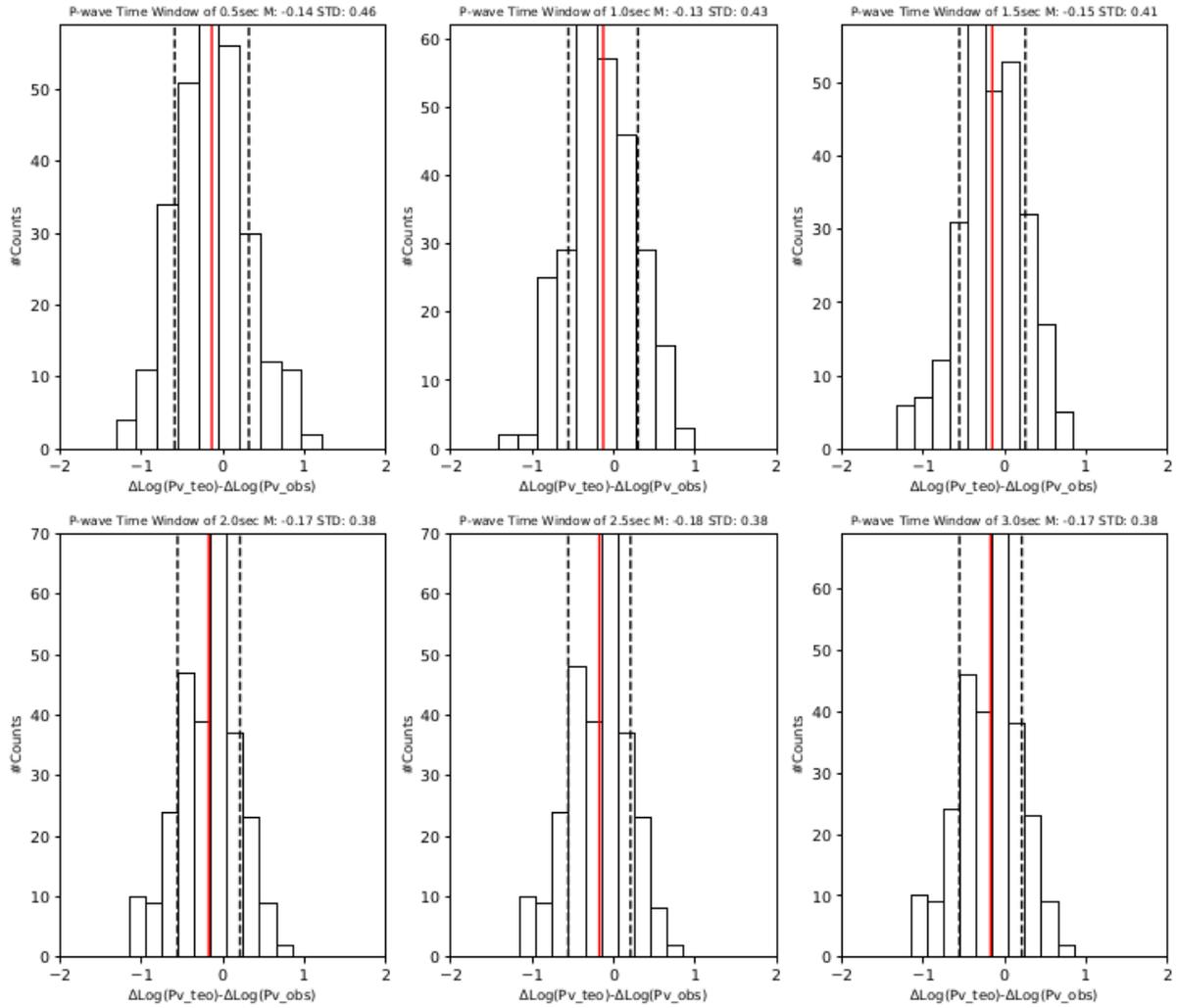
**Table S1.** Table that contains the detailed information about the earthquakes of the Central Italy sequence used in the different station/source configurations. The event locations and magnitudes are taken from the INGV bulletin (<http://terremoti.ingv.it/>). The last three columns refer to the number of stations that recorded the event in each simulated configuration A, B and C. The “/” symbol indicates that the event was not used in the configuration.



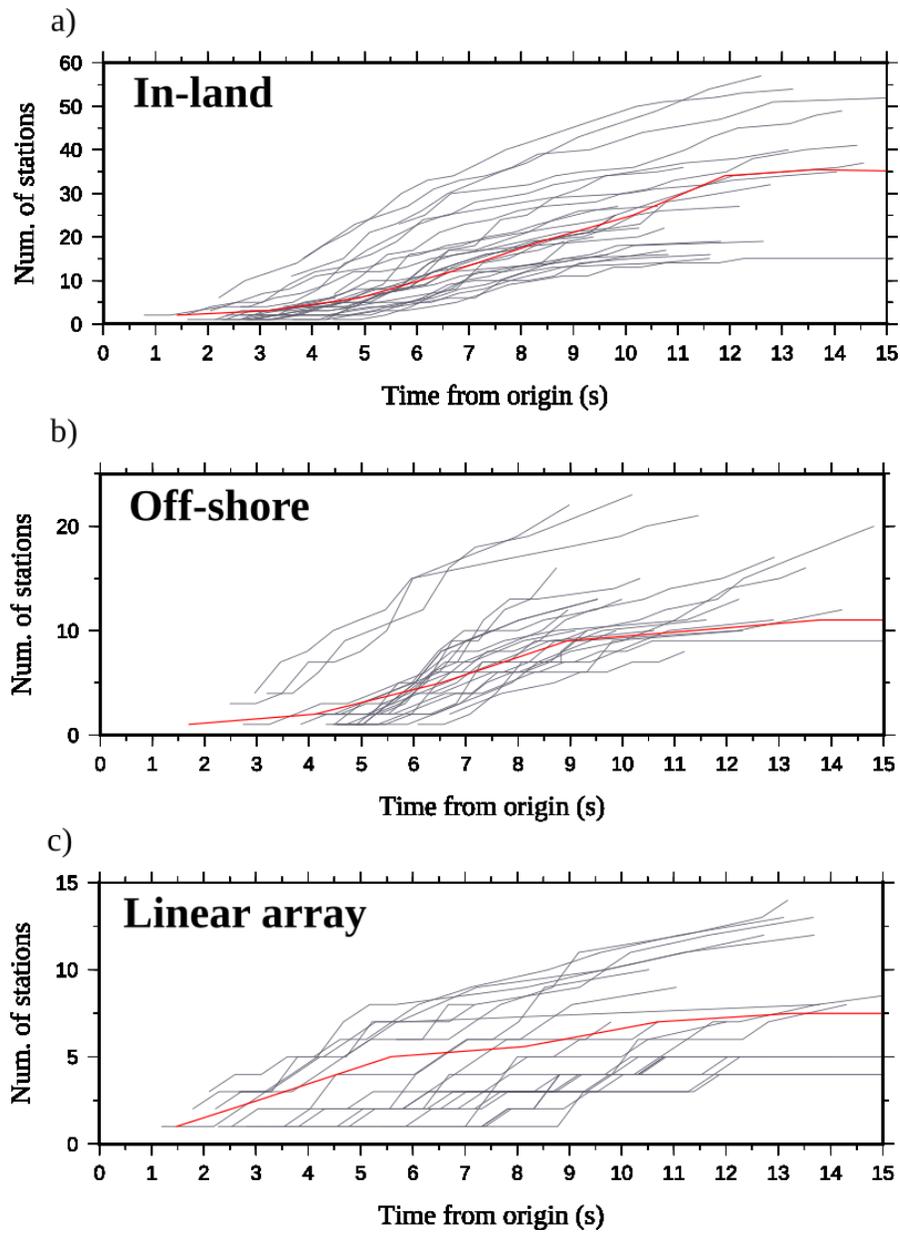
**Figure S1.** Histograms of the distribution of the difference between the automatic and the manual P picks.



**Figure S2.** Histograms of the distribution of the difference between the BAZazimuth calculated for different window length between 0.5 s and 3 s, and the reference one (obtained from the reference INGV location). In each panel were shown the mean value (red vertical line) and the standard deviation of the distribution (dashed black vertical lines).



**Figure S3.** Histograms of the distribution of the difference between the differential P-wave amplitude ratio calculated for different window length between 0.5 s and 3 s, and the reference one (obtained from the reference INGV location). In each panel were shown the mean value (red vertical line) and the standard deviation of the distribution (dashed black vertical lines).



**Figure S4.** Temporal evolution of station number from event origin time, for the different simulated configurations: In-land (a); Off-shore (b) and Linear array (c) . The red lines represent the median curves.