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)

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

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.

1 Introduction

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

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

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

When dealing with real-time applications for Earthquake Early Warning Systems (EEWS), which asks for very fast source parameter estimates (within a few seconds), the earthquake location procedure becomes a sensitive issue which requires the adoption of dedicated, non-trivial algorithmic solutions. These must account for the continuous and evolutionary waveform data availability in real-time, depending on the geometry and distribution of the seismic stations around the epicenter, as well as on the velocity of propagation of seismic waves across the network.

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

Several approaches for the real-time location have been developed and various parameters have been proposed in order to gain constraint on the solution, when few observed data are available. In the Elarms methodology (*Allen*, 2007), for example, at the arrival of the first trigger, the event is positioned at that unique station and the depth is fixed to the typical depth of the events in the region. When two and three stations trigger the event, the epicenter location is fixed as the centroid position between the triggered stations. Finally, when four stations have recorded the P-wave, a grid search method is used to locate the event, searching for the minimum misfit between predicted and observed arrival times. *Horiuchi et al.* (2005) first introduced the concept of not-yet triggered stations to constrain the event location when only two stations are available. In their approach, the initial solution is constrained using the estimated Equal Differential Time (EDT) surface (*Font et al.*, 2004; *Lomax*, 2005), i.e., the quasihyperbolic surface on which the difference in calculated travel time to a pair of stations is equal to the difference in observed arrival times for the two stations. The EDT shape in the 3D medium is progressively updated as more stations record the P-wave arrival time. In this approach the not-triggered stations provide a constraint which allows to further delimitate the probabilistic volume containing the hypocenter. *Satriano et al.* (2008) then introduced an evolutionary, probabilistic approach for the real-time earthquake location, also based on the EDT formulation, on the concept of triggered and not-yet-triggered stations and on the use of the Voronoi cell associated to each available station, allowing to constrain the initial hypocenter location even with just one recorded P-wave arrival. More recently other authors (*Noda et al.*, 2012; *Eiserman et al.*, 2015; *Liu and Yamada,* 2014) introduced new observed parameters to constrain the real-time earthquake location in early warning applications.

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

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

Eiserman et al. (2015) evaluated the robustness of three independent real-time back-azimuth (BAZ) determination schemes, through the offline analysis of southern California earthquake records and found that the three methods provide equivalent levels of accuracy. After passing the P-wave signals through specifically designed algorithms for checking the signal coherency and signal-to-noise quality they show that BAZ estimates can be achieved in real time, with an optimal error of less than 15°.

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

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

We first describe the details of the methodology and then present the results of its application to a set of events recorded during the 2016-2017 Central Italy seismic sequence. During the testing phase, we perform jackknife simulation experiments with optimal/non-optimal data acquisition lay-outs, by changing the geometry, coverage and number of stations and discuss the performance of the method for the different situations.

2 Method

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

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

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

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

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

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

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

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

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

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

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

$$BAZ_n = g[\theta^n(R^n_{ZE}, R^n_{ZN}), sign(R^n_{ZN})] \qquad (1)$$

where:

$$\theta^n = arctan\left(\frac{R^n_{ZE}}{R^n_{ZN}}\right) + \pi$$

with:

$$R^n_{ZE} = \alpha\, R^{n-1}_{ZE} + A^n_Z\, A^n_E$$

$$R^n_{ZN} = \alpha\, R^{n-1}_{ZN} + A^n_Z\, A^n_E$$

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

$$g(\varphi) = \begin{cases} \varphi + \pi \;\; if\; R^n_{ZN} < 0 \\ \;\; \varphi \quad otherwise \end{cases}$$

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

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

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

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

$$P(\boldsymbol{m}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{m})\rho(\boldsymbol{m})}{p(\boldsymbol{d})} \tag{2}$$

Where

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

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

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

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

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

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

$$\boldsymbol{d_1} \equiv (\Delta t_{12}, \dots, \Delta t_{N1})$$

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

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

The conditional probability $P(m)$ can be defined as the likelihood function for differential time residuals according (Tarantola & Valette, 1982):

$$P(\boldsymbol{d_1}|\boldsymbol{m}) = const\ e^{-\frac{\sum_1^{N-1}\sum_2^{N}\left(\Delta t_{ij-}\Delta\tau_{ij}(\boldsymbol{m})\right)^2}{2\sigma_T^2}} \tag{3}$$

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

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

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

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

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

$$To = \frac{\sum_i^N \left(t_i - t_{oi}(x,y,z)\right)}{N} \tag{4}$$

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

Where t_i are the measured arrival times at the N stations for which the P-picking is available and σ_{To} is the estimated uncertainty on To.

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

$$P(\boldsymbol{m}|\boldsymbol{d_1}) = const\ \ P(\boldsymbol{d_1}|\boldsymbol{m})\rho(\boldsymbol{m}) \tag{6}$$

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

In our recursive Bayesian method, (6) is set as the prior pdf for the posterior pdf of $\boldsymbol{m}$ given the differential P-amplitudes data-set $\boldsymbol{d_2} \equiv (\Delta PD_{12}, \dots, \Delta PD_{N1})$. In this case the conditional pdf of P-amplitudes is defined:

$$P(\boldsymbol{d_2}|\boldsymbol{m}) = const\ e^{-\frac{\sum_1^{N-1}\sum_2^N \left(\Delta PD_{ij} - \Delta PD'_{ij}(\boldsymbol{m})\right)^2}{2\sigma_A^2}} \tag{7}$$

Where ΔPD_{ij} is the theoretical differential P-amplitude at stations $i\ and\ j$, computed for a given model parameter vector $\boldsymbol{m}$, and σ_A^2 is a theoretical estimate of the variance for the log of the amplitude ratio. The theoretical values of this quantity are determined using the attenuation relations of the form:

$$\log PD = A + B\ M + \mathrm{C} \log R$$

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

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

Applying the Bayes' theorem and setting $\rho(\boldsymbol{m}) = P(\boldsymbol{m}|\boldsymbol{d_1})$, the posterior pdf for differential P-amplitudes can be written as:

$$P(\boldsymbol{m}|\boldsymbol{d_2}, \boldsymbol{d_1}) = const\ \ P(\boldsymbol{d_2}|\boldsymbol{m})P(\boldsymbol{m}|\boldsymbol{d_1})$$

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

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

$$P(\boldsymbol{d_3}|\boldsymbol{m}) = const\ e^{-\frac{\sum_1^N\left(\alpha_i-\alpha'_i\right)^2}{2\sigma_\alpha^2}} \tag{9}$$

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

$$P(\boldsymbol{m}|\boldsymbol{d_3}, \boldsymbol{d_1}, \boldsymbol{d_2}) = const\ P(\boldsymbol{d_3}|\boldsymbol{m})P(\boldsymbol{m}|\boldsymbol{d_2}, \boldsymbol{d_1}) =$$

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

Equation (10) provides the pdf for the model parameter $\boldsymbol{m}$, given the three-different data-sets. Its numerical computation requires the regular sampling of the discretized volume where earthquakes are expected to occur. The constant in eq. 10 has to be evaluated numerically, but setting the condition:

$$\int P(\boldsymbol{m}|\boldsymbol{d_3},\boldsymbol{d_1},\boldsymbol{d_2})\, d\boldsymbol{m} = 1$$

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

$$\boldsymbol{m}_{BEST} :\ P(\boldsymbol{m_{BEST}}|\boldsymbol{d_3},\boldsymbol{d_1},\boldsymbol{d_2}) = \max[P(\boldsymbol{m}|\boldsymbol{d_3},\boldsymbol{d_1},\boldsymbol{d_2})] \tag{11}$$

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

$$P\left(\boldsymbol{m}\middle|\boldsymbol{d_3},\boldsymbol{d_1^{best}},\boldsymbol{d_2^{best}}\right)$$

$$P\left(\boldsymbol{m}\middle|\boldsymbol{d_3^{best}},\boldsymbol{d_1},\boldsymbol{d_2^{best}}\right)$$

$$P\left(\boldsymbol{m}\middle|\boldsymbol{d_3^{best}},\boldsymbol{d_1^{best}},\boldsymbol{d_2}\right) \tag{12}$$

where $\boldsymbol{d_1^{best}},\boldsymbol{d_2^{best}},\boldsymbol{d_3^{best}}$ are the parameters of the maximum likelihood solution. These pdfs allow to measure the maximum likelihood model parameter vector and the interval of parameters associated with 31% and 68% significance levels (e.g. the parameter values at the 31% and 68% level of the cumulative pdf), which corresponds to $\pm 1\sigma$ case of normal pdfs.

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

We implemented a software platform written in Python (https://www.python.org/) that manages the inversion code and is able to simulate the real-time data streaming. The computational efficiency is optimized using a multi-parallel computational approach in order to process each single station in parallel during the whole simulation. This approach guaranties a rapidity in the solution estimation that is generally provided in a time less than 0.5 second (source and network lay-outs of the application study illustrated in section 4), that is the usual packet length during real-time transmission for modern data-loggers. The software pre-computes the theoretical travel time-table for a distributed grid of sources in the 3D medium (calculated using The *TauP* Toolkit by *Crotwell et al.,*1999), BAZ and P-amplitudes (through pre-existing empirical attenuation relationships) in order to minimize the computational cost during the software runs.

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

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

The events of the Central Italy seismic sequence that occurred between August 2016 and January 2017 have been used to test and demonstrate the algorithm performance. From the whole sequence (about 135 earthquakes) we selected 27 events with moment magnitude larger than 4.2, being this magnitude range of more interest for EEW applications. We considered a volume of 80x100x20 km^3, which contains the selected events and 63 stations belonging to the Italian Strong Motion Network (Rete Accelerometrica Nazionale - RAN), operated by the Dipartimento della Protezione Civile (DPC), and to the Italian National Seismic Network, operated by the Istituto Nazionale di Geofisica e Vulcanologia (INGV, Fig. 3a). The data-set includes the mainshocks of the sequence, the Mw 6 Amatrice event occurred on August, 24 2016, the two M5.9 and M5.4 events occurred on 26 October 2016, located near Visso (northeast of Norcia), the Mw 6.5 Norcia event on 30 October 2016, and the Mw 5.5, January 18, 2017 earthquake located south of the town of Amatrice (*Chiaraluce et al.,* 2017). The details about the event origin times, locations, magnitudes and number of recording stations can be found in table S1 in Supporting information (SI).

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

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

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

The INGV bulletin locations (http://terremoti.ingv.it/), obtained by considering the dense INGV-RSN network, has been chosen as the reference solution, to which we compared the solutions obtained by the three network configurations. For the earthquake locations we used the 1D crustal velocity model obtained by Lii et al. (2007) for central Italy, parameterized on a 3D grid with a cell size of 0.6x0.6x0.8 km^3.

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

We performed an optimization analysis in order to set properly the standard deviations of the three variance factors in the probability distribution (i.e. σ_T for differential times, σ_α for P-polarization and σ_A for P-amplitudes ratios in eq. 10) and the length of the time-windows to be used to measure the P-peak amplitude and the BAZ from the P-polarization. The choice of the time-window length has been done considering the requirement for a rapid but reliable estimate of the parameters. By considering the "in-land" configuration, we constructed the distributions of the difference between the calculated BAZ and the reference one (e.g the one obtained by considering the reference INGV bulletin location, see Fig. S2 of SI), and, similarly, the distribution of the calculated amplitude ratios and the reference one (see Fig. S3 of SI). We built these distributions by varying the window length between 0.5 s and 3 s and choosing the one for which the differences were minimized (i.e. 0.5 s for the BAZ and 2 s for the amplitude). The standard deviations of the chosen distributions were used to infer the variances σ_α and σ_A of the two probability distributions (i.e. 60° for BAZ and 0.4 for the log P-amplitude ratio). For the σ_T of differential times we considered a value of twice (since we used the differential times) the mean of distribution of the difference between the automatic and the manual P-wave picks (i.e. 0.3 s, see Fig. S1 of SI).

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

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

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

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

Considering the “Off-Shore” network lay-out of configuration, the location accuracy is smaller than 5 km with about 6-7 stations for the most of events (Fig. 6f-h), despite of the worst azimuthal coverage of the stations compared to the one in In-land configuration. For 4 events, data from at least 9 stations must be available in order to have a well constrained epicentral estimate. As expected, for the considered events the depth is less constrained using the Off-shore configuration, but with about 7 stations the depth accuracy is within 5 km. Concerning the origin time, its deviation from the reference one is on average smaller than 1 s with 3-4 stations, i.e. within 4 s from the event origin time. On average, 6-7 P arrival-time readings are available within 7 s from the origin time (see Fig. S4 in SI).

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

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

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

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

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

Four seconds after the first P-wave arrival time, the differential times are the most influential parameter for the location in the different network configurations. In fact, the histograms obtained by using only differential arrival times (Fig. 8a and d) are very similar to the ones obtained by using the differential arrival times and BAZs (Fig. 8b and e) and by using all three observed parameters together (Fig. 8c and f). But in the least favorable configuration (i.e., the “Linear array”), BAZs and amplitude ratios show to be relevant to reduce the uncertainty on the epicentral location. With the additional use of BAZs and amplitude ratios the 68% epicenter residual value passes from 13 km to 8 km, while the 95% value from 28 km to 15 km. Beside, in terms of depth accuracy, the results clearly indicated that, as it is expected, the BAZ usage may improve the epicenter location but it does not affect the depth. In fact, the histograms of depth residuals obtained by using only the differential times (Fig. 8a-d-g, left panel) are very similar to the ones obtained by using the integrated data-set (Fig. 8b-c-h and c-f-i, left panels). Despite this, the histograms of depth residuals at 4 seconds after the first pick show significant improvements respect to the ones at 2 seconds after first pick. In details, for the “In-land” configuration, the 68% residual value decreases from 6 km to 4 km; for the “Off-shore” configuration the 68% and 95% residual values decrease from 7-8 km to 2 km and from 9-10 km to 5-6 km; for the “Linear array” configuration the 68% and 95% residual values decrease from 7 km to 6 km and from 10-11 km to 9 km.

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

We compared the performance of RTLOC and our location method in the case of the “Off-shore” and “Linear array” network configurations, i.e. the ones in which we expect that the integration of BAZs and amplitude ratios could improve the location accuracy. We chose for the comparison the location accuracy at 3 second after the first P pick, so to guarantee at least 3 P picks, 2-3 BAZs and 1-2 amplitude ratios for each location. The results of the comparison, in terms of epicentral location and depth accuracy are shown in Fig. 9 for the “Off-shore” (a-b panels) and “Linear array” (c-d panels) configurations. As for the previous figures, we computed the *cumulative normalized frequency* of the observed parameter residual distribution and report the parameter residual value associated with the 68% (red values and dashed lines in each panel of Fig. 9) and 95% (blue values and dashed lines in each panel of Fig. 9) levels of the distribution. From these values it can be inferred that the presented location method is more suitable than RTLOC in cases of unfavorable station/event distribution. In fact, with reference to the “Off-shore” configuration, the value associated with the 68% of epicenter residual decrease from 5 km of RTLOC (red value in b, left panel) to 3 km of our M-PLOC (red value in a, left panel), while the depth residual decreases from 4 km (red value in b, right panel) to 3 km (red value in b, right panel), respectively. The values associated with the 95% of epicenter and depth residuals increase of 1 km with our technique (blue values in a-b panels). For the “Linear array” configuration, the value associated with the 68% of epicenter residual decreases from 11 km of RTLOC (red value in d, left panel) to 6 km of M-PLOC (red value in c, left panel), while the depth residual remains at 6 km (red value in d-c, right panels) with the two methods. The values associated with the 95% of epicenter and depth residuals decrease of 1 km with M-PLOC (blue values in c-d panels).

4 Discussion

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

Dedicated algorithms, suitable to work in real-time, have been developed to measure the three parameters on limited portions of the P-wave signals, when no other source information is available. In principle, the measurement of P-wave arrival times, amplitudes and signal polarization are relatively simple and do not require sophisticated approaches. When dealing with real-time applications, however, these measurements become non-trivial and their accuracy may critically depend on a number of factors, such as the quality of recorded signals and unknown contaminating effects of the propagation medium. In this context, the combined use of three parameters can be strongly advantageous to constrain the source location, if these parameters are correctly measured, as well as largely inconvenient, if incorrect real-time estimates are used. For example, in the case of a poor signal quality with low signal-to-noise ratio and in the absence of any other source information allowing to properly set the suitable parameters (i.e., filters and threshold levels), the real-time, automatic P-wave picking operation may generate erroneous phase detections, with consequent bias for the whole location method. Furthermore, a reliable (1D or 3D), pre-defined velocity model is needed for the computation of theoretical P-wave travel times at the available stations, to be compared with observed phase arrivals when solving the inverse problem. The real-time measurement of amplitudes is ideally straightforward, although it is critically dependent on the correct knowledge of attenuation relationships with distance, used to compare the observed amplitudes at pairs of stations. Finally, both amplitude and polarization measurements are sensitive to high frequency heterogeneities and local site amplifications, which are not accounted for the simplified assumption of a 1D attenuating medium. It is therefore relevant to get reliable estimates of the uncertainty on real-time measured quantities so to weigh them when used for location parameter estimation. Our proposed probabilistic approach accounts for the different uncertainties related to the estimates of differential times, amplitude ratios and back-azimuths, which are taken into account through the variance factors σ_T^2, σ_A^2 and σ_α^2 of the pdf in eq.6. Although we assumed constant variance factors in our analysis, more in general, these factors could be replaced by single data variances, as inferred from real-time measurements.

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

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

The location method proposed here works with differential observables, which are jointly measured at pairs of stations at each time. This allows to determine reliable earthquake locations as soon as two stations have recorded the P-wave arrival and to achieve accurate solutions when a few seconds of P-wave signals are recorded at few stations (3 to 5).

This is confirmed by the retrospective analysis of mainshocks of the 2016-2017 Central Italy earthquake sequence, whose results are summarized in Figures 6-9. Overall, after few iterations the method converges to stable solutions, in terms of both epicentral coordinates and source depth, as it can be seen from Figure 6. By considering the "In-land" configuration, the epicentral locations indeed are well constrained (i.e. within 5 km from the reference location) with about 5-6 stations (Fig. 6a), typically 5-6 sec after the event origin in the analyzed cases. With the same number of stations, the difference in depth estimate with INGV catalogue is nearly stable around zero, varying between $\pm$5 km from the reference estimate (Fig. 6c) and the origin times are within 1.5 s from the reference ones (Fig. 6e).

Similar results are observed even with a non-optimal coverage of stations ("Off-shore" configuration), in which, about 6-7 stations are necessary to converge to stable solutions, with epicentral and depth error smaller than 5 km and a deviation from the reference origin time of about 1s (Figure 6 f-j). On average, 7 sec after the event origin in the analyzed cases.

The major strength of the proposed approach is the ability of providing correct location solutions, even in non-optimal network geometries and in unfavorable station distributions. This is the case of events outside the area covered by the stations, which are distributed only by one side of the epicenter, or the case of linear arrays. In this last case, for example, standard location techniques using phase arrival times are often not suitable to constrain the hypocenter position, or provide strongly undetermined solutions, with large uncertainties in both epicentral position and depth. The combined use of times, amplitudes and BAZ makes the proposed method suitable to work in disadvantageous conditions of sparse networks, with a limited number of recording nodes and/or poor azimuthal coverage. Indeed, in the linear array configuration, the majority of the analyzed events require 7-8 stations, available on average after 10 s from origin time, to constrain the location solution, both in terms of epicentral estimates and of source depth and origin time (Fig. 6k-o).

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

A relevant result for all the tested network configurations is that the decrease of uncertainties in real-time estimates (panels b, d, g, i, l, n in Fig. 6) is associated to the convergence of the solution toward the reference parameters (panels a, c, f, h, k, m in Fig. 6) from INGV catalogue obtained in optimal distance and azimuth coverage conditions. This suggests the possibility to use in real-time the estimated parameter uncertainty vs the station number (or time from the event origin) to assess the reached convergence to the final solution and stop further iterations. As compared to another real-time location method, RTLOC (Satriano et al, 2008), which uses only the P-wave arrival times, the proposed multi-parametric approach turns out to provide better constrained location solutions, since the very first available data. This is especially true in the unfavorable case of the "Linear array" distribution, where the joint use of three parameters strongly reduces the difference to the reference solutions, as it can be seen from the *cumulative normalized frequency* (and its 68-95% levels) of Fig. 9.

From the computational point of view, the proposed approach is efficient and optimized for running in real-time applications, where the earthquake location has to be retrieved in a very short time (around 1 sec) after data acquisition. The methodology proposed here does not require complex computational structures and can be easily integrated in other regional, network-based EEW approaches.

5 Conclusions

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

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

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

Acknowledgments and Data

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

References

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

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

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

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

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

Font, Y., Kao, H., Lallemand, S., Liu, C-S. and L-Y Chiao (2004), Hypocentral determination offshore Eastern Taiwan using the Maximum Intersection method. Geophys. J. Int., 158, 655-675.

Horiuchi, S., Negishi, H., Abe, K., Kamimura, A. and Y. Fujinawa (2005), An Automatic Processing System for Broadcasting Earthquake Alarms. Bull. Seism. Soc. Am.; 95 (2): 708–718. doi:https://doi.org/10.1785/0120030133.

Lii, H., Michelini,A., Zhu,L., Bernardi, F. and M. Spada, Crustal Velocity Structure in Italy from Analysis of Regional Seismic Waveforms (2007), Bull. Seism. Soc. Am. , 97 (6): 2024–2039. doi: https://doi.org/10.1785/0120070071.

Liu, A. and M. Yamada (2014), Bayesian Approach for Identification of Multiple Events in an Early Warning System. Bull. Seism. Soc. Am., 104 (3). pp. 1111-1121. ISSN 0037-1106.

Lockman, A.B., and R.M. Allen (2005). Single-Station Earthquake Characterization for Early Warning. Bull. Seism. Soc. Am. , 95(6):2029-2039 DOI: 10.1785/0120040241

Lomax, A., Virieux,J., Volant,P. and C. Berge (2000), Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations, in Advances in Seismic Event Location , Thurber, C.H., and N. Rabinowitz (eds.), Kluwer, Amsterdam, 101-134.

Lomax, A., Michelini, A. and A. Curtis (2009), Earthquake Location, Direct, Global-Search Methods, in Complexity In Encyclopedia of Complexity and System Science, Part 5, Springer, New York, pp. 2449-2473, doi:10.1007/978-0-387-30440-3.

Lomax, A (2005), A Reanalysis of the Hypocentral Location and Related Observations for the Great 1906 California Earthquake. Bulletin of the Seismological Society of America ; 95 (3): 861–877. doi: https://doi.org/10.1785/0120040141.

Lomax, A., Satriano, C. and M. Vassallo (2012), Automatic picker developments and optimization: FilterPicker - a robust, broadband picker for real-time seismic monitoring and earthquake early-warning, Seism. Res. Lett. , 83, 531-540, doi: 10.1785/gssrl.83.3.531.

Nakamura, Y. (1988), On the urgent earthquake detection and alarm system (UrEDAS), in Proceedings of Ninth World Conference on Earthquake Engineering, Paper# 13-2-12, vol. VII, pp. 673–678, Tokyo-Kyoto, 2–9 Aug.

Noda, S., Yamamoto, S., Sato, S. et al. (2012) Improvement of back-azimuth estimation in real-time by using a single station record. Earth Pla. Space 64, 305–308 (2012). https://doi.org/10.5047/eps.2011.10.005

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

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

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

Figure captions

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

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

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

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

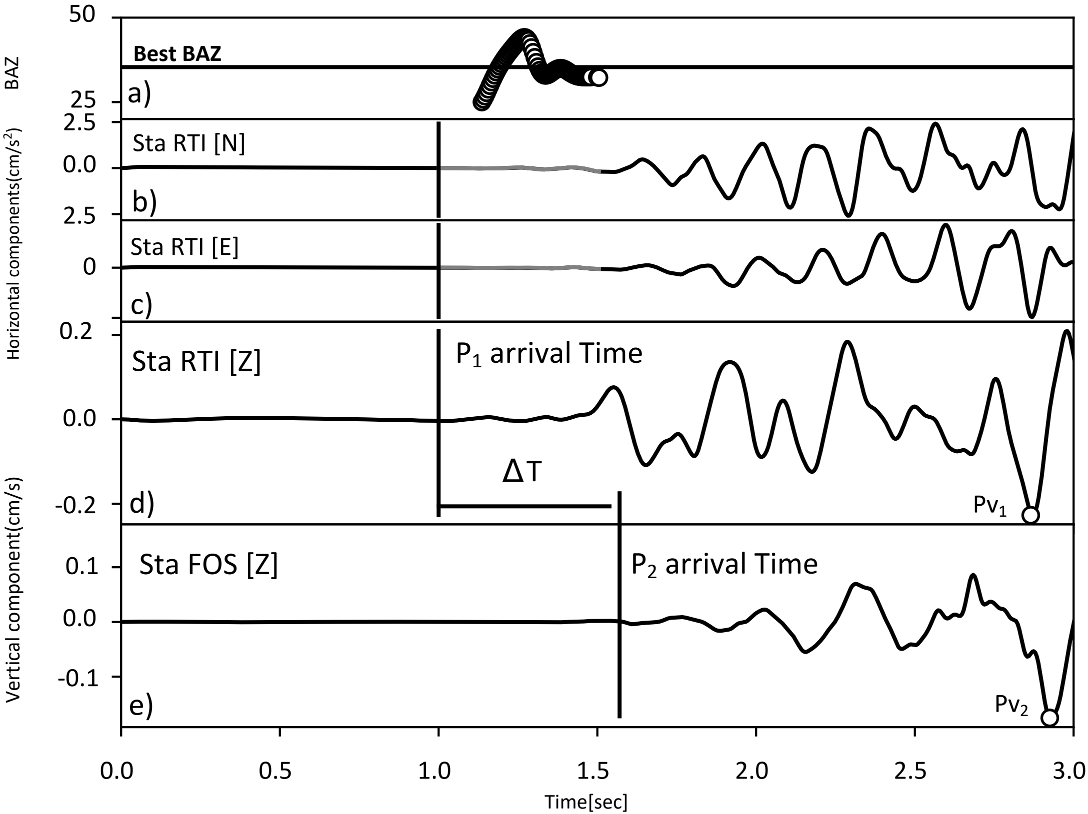
Figure 5. Relative location probability distribution after 2.6 s (a), 3.1 s (b), 4.1 s (c) and 9.1 s (d) from origin time. For each snapshot, was shown the normalized location probability, the stations used for the location (black triangles), the reference event location (white star) and the predicted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

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

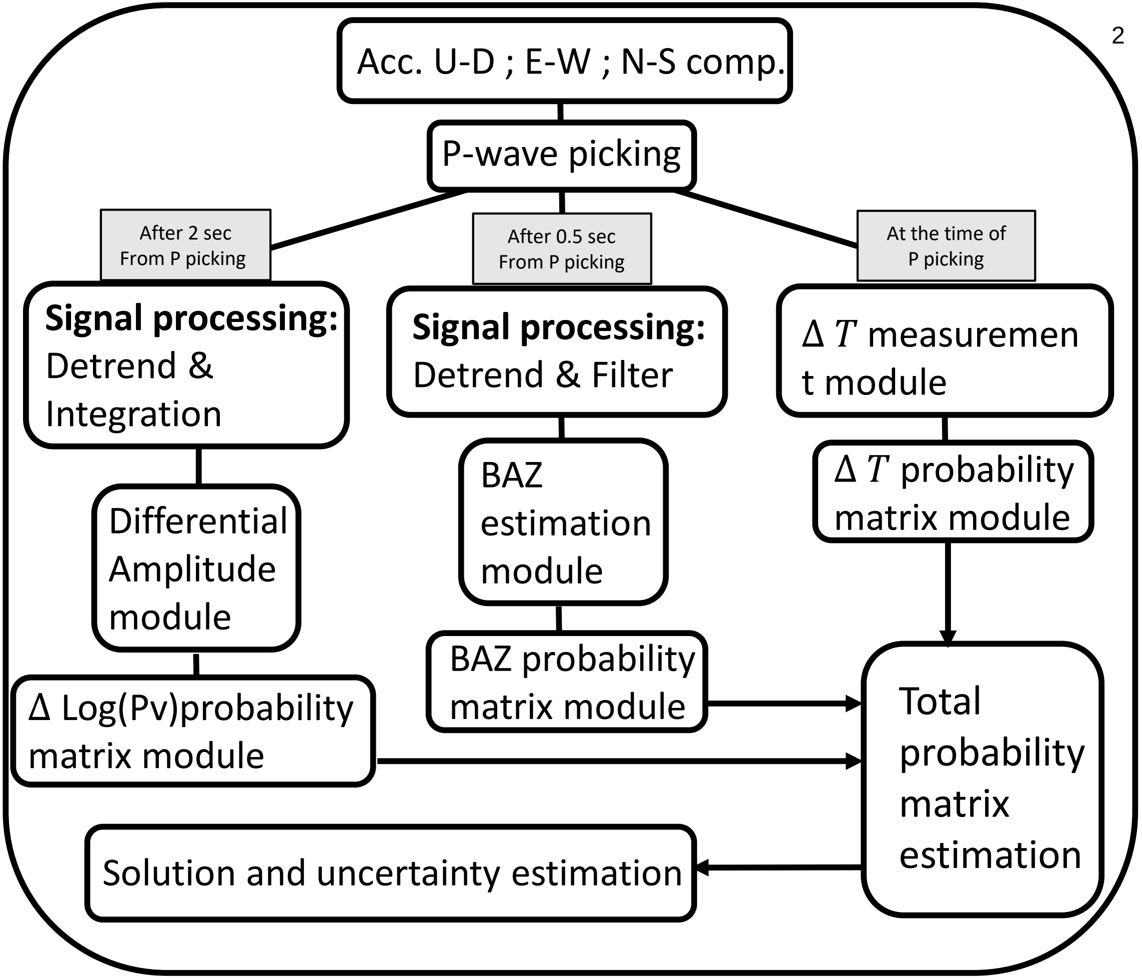
Figure 7.Comparison between the algorithm performance after 2 second from the first P pick by using different data type combinations: only differential times (a, d, g), differential times and BAZs (b, e, h) and differential times plus differential amplitude and BAZs (c, f, i). The epicentral and in-depth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

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

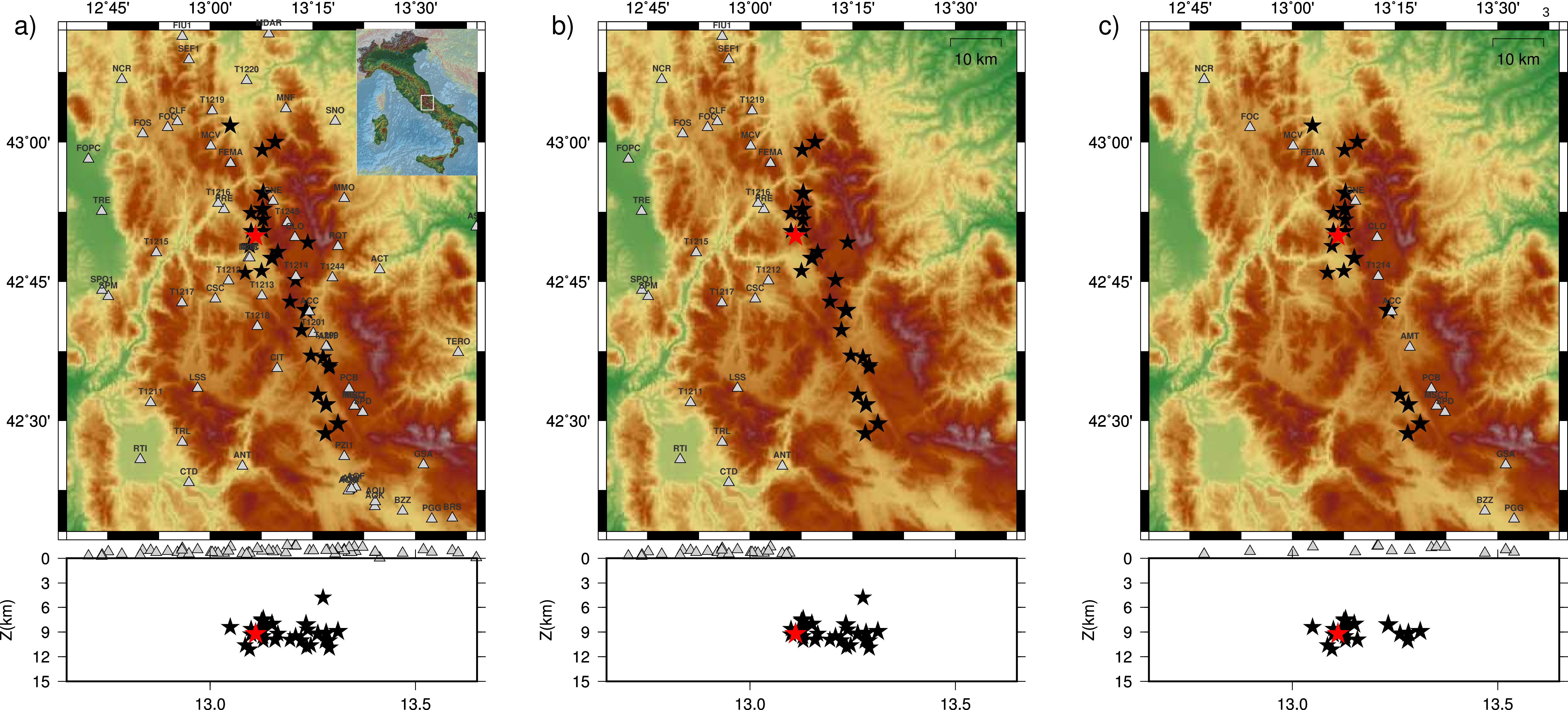
Figure 9. Comparison between the performance at 3 seconds after the first P pick of M-PLOC (a, b) and RTLOC (c, d). The epicentral and depth location accuracy (difference between the estimated and the reference one) is shown for the “Off-shore” (a-c) and “Linear array” (b-d) configurations. In each panel,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

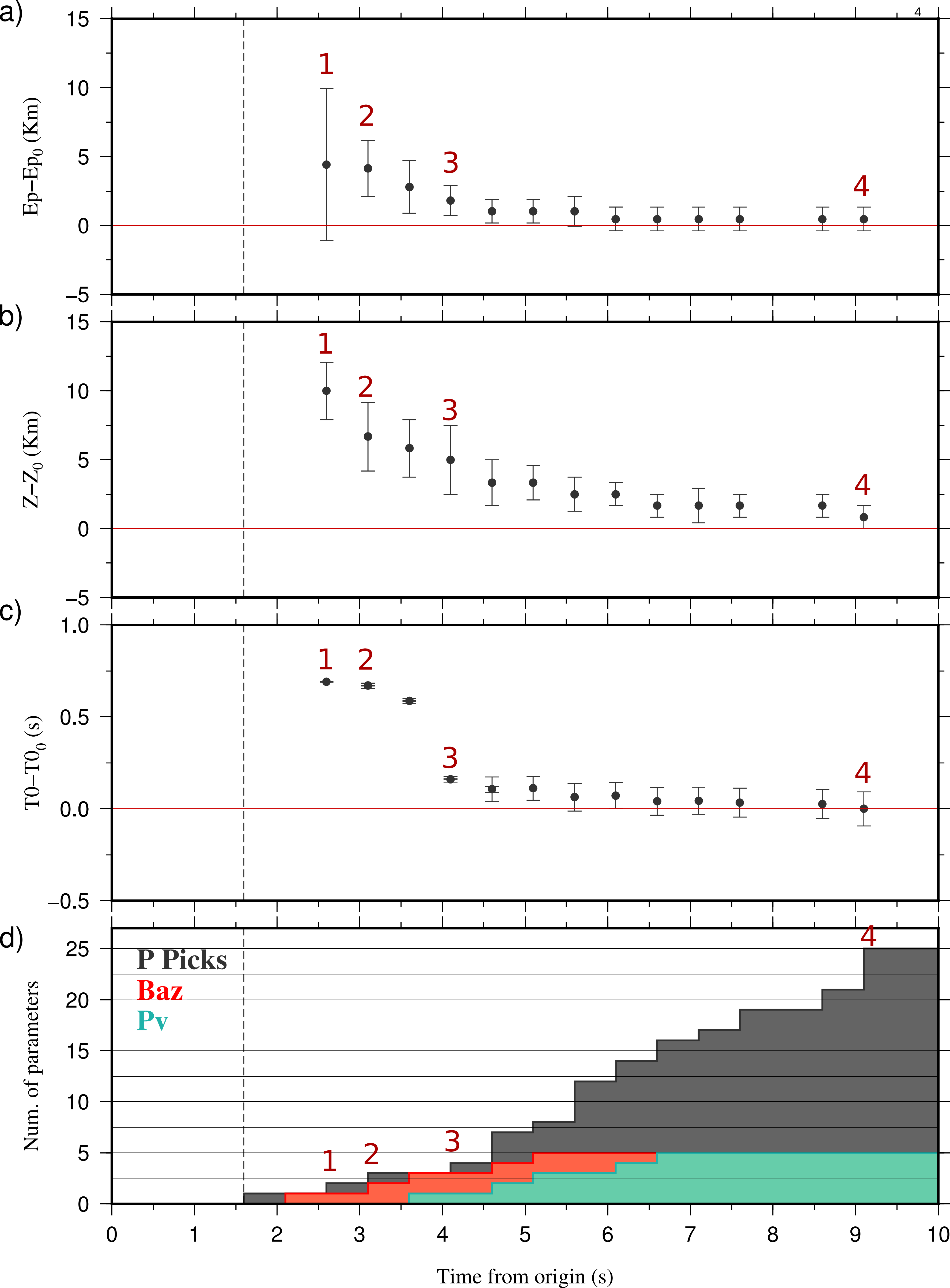


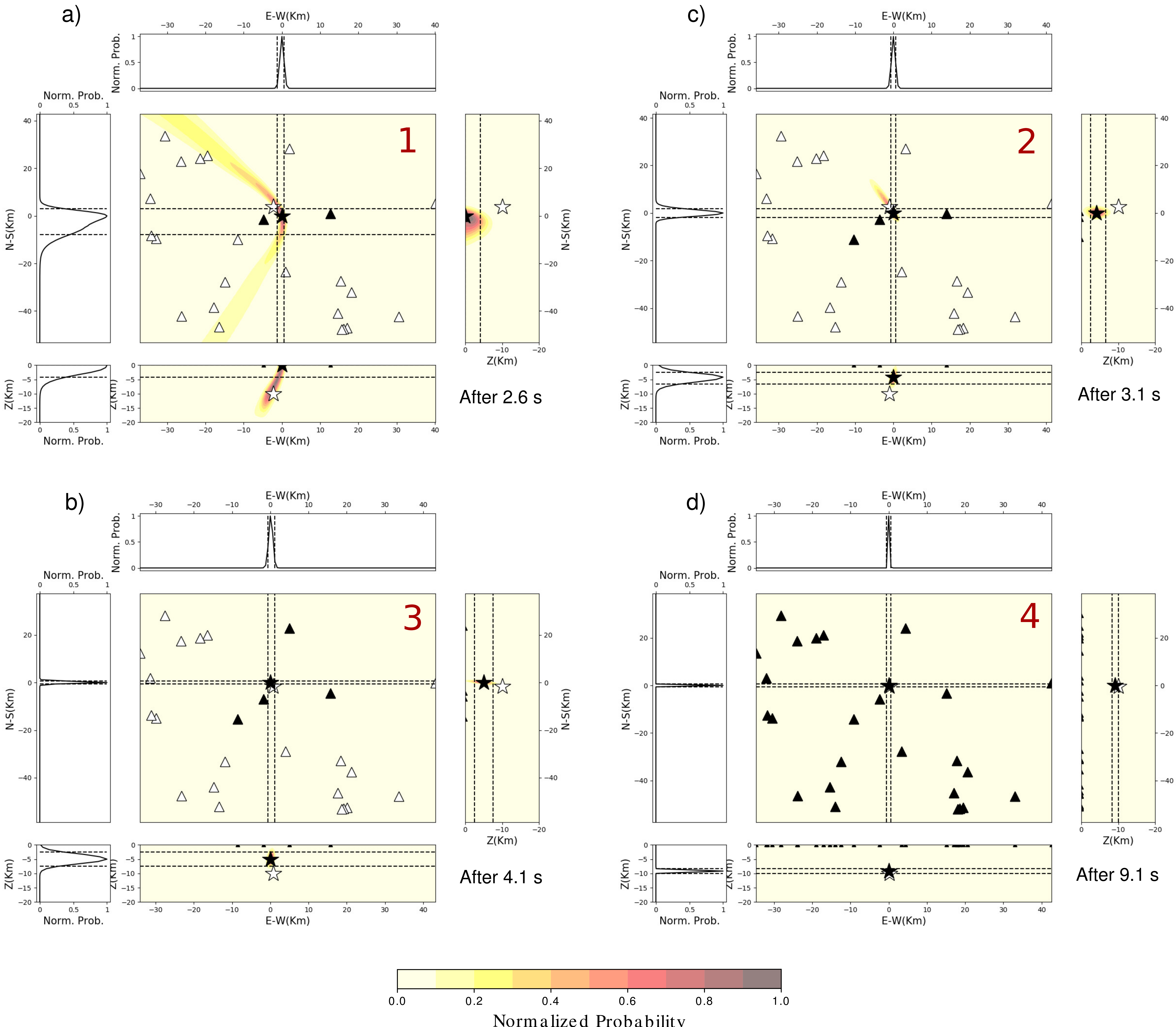
50
25
BAZ
Best BAZ
a)
2.5
0.0
2.5
Sta RTI [N]
b)
2.5
0
Sta RTI [E]
c)
Horizontal components(cm/s2)
0.2
0.0
-0.2
Sta RTI [Z]
P1 arrival Time
ΔT
d)
Pv1
0.1
0.0
-0.1
Sta FOS [Z]
P2 arrival Time
e)
Pv2
Vertical component(cm/s)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time[sec]

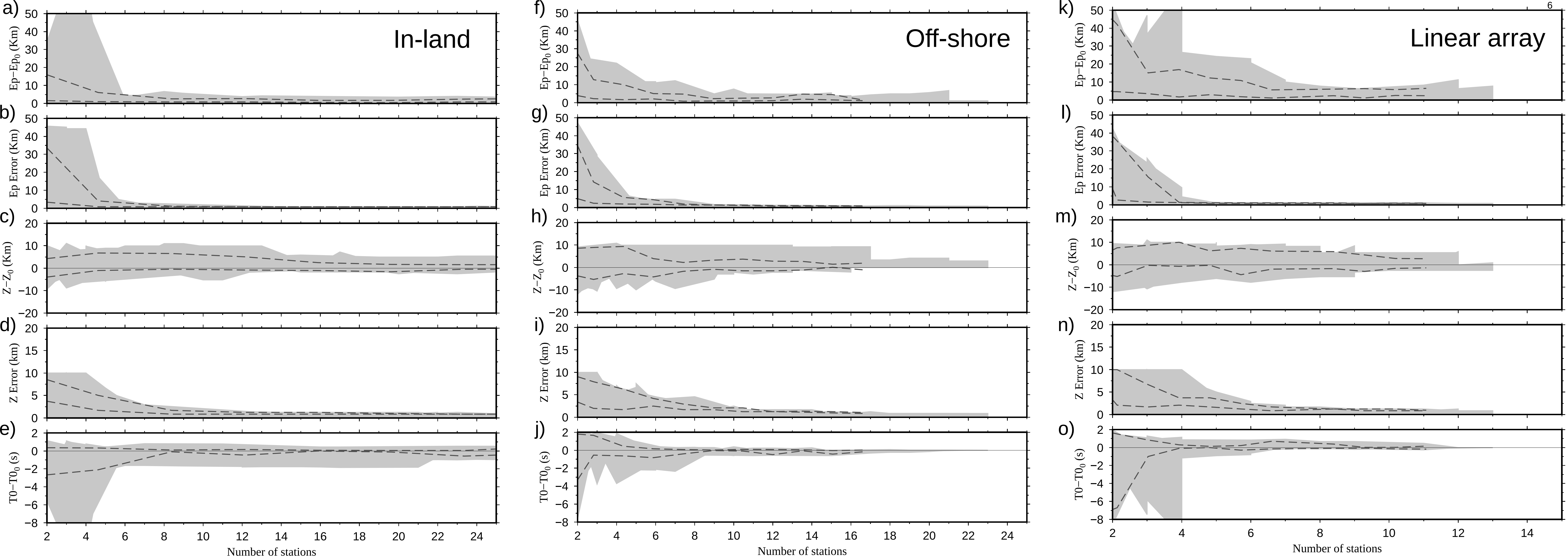


2
Acc. U-D ; E-W ; N-S comp.
P-wave picking
After 2 sec
From P picking
After 0.5 sec
From P picking
At the time of
P picking
Signal processing:
Detrend &
Integration
Signal processing:
Detrend & Filter
Δ T measuremen
t module
Differential
Amplitude
module
BAZ
estimation
module
Δ T probability
matrix module
BAZ probability
matrix module
Δ Log(Pv)probability
matrix module
Total
probability
matrix
estimation
Solution and uncertainty estimation

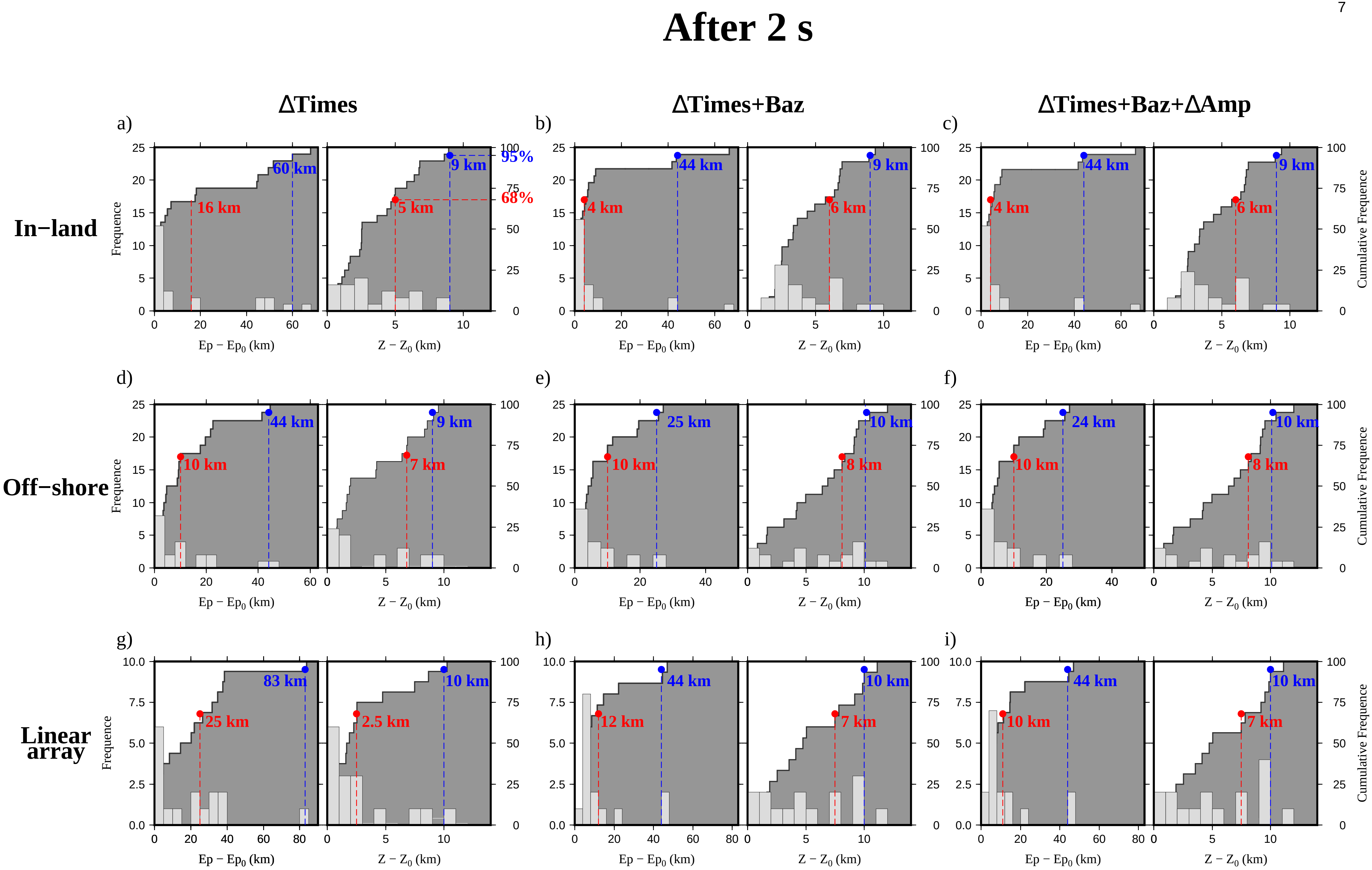

a)
b)
c)
12˚45'
13˚00'
13˚15'
13˚30'
43˚00'
42˚45'
42˚30'
10 km
FIU1
MDAR
SEF1
NCR
T1220
T1219
MNF
SNO
FOS
FOC
CLF
MCV
FEMA
FOPC
T1216
PRE
CNE
MMO
TRE
T1245
CLO
ROT
T1215
T1212
T1214
T1244
ACT
SPO1
SPM
T1217
CSC
T1213
ACC
T1218
T1201
AMT
TERO
CIT
LSS
PCB
MSCT
SPD
T1211
TRL
RTI
ANT
PZI1
GSA
CTD
AQU
AQK
BZZ
PGG
BRS
Z(km)
0
3
6
9
12
15
13.0
13.5


4
a)
b)
c)
d)
Ep−Ep$_0$ (Km)
Z−Z$_0$ (Km)
T0−T0$_0$ (s)
Num. of parameters
Time from origin (s)
P Picks
Baz
Pv
1
2
3
4

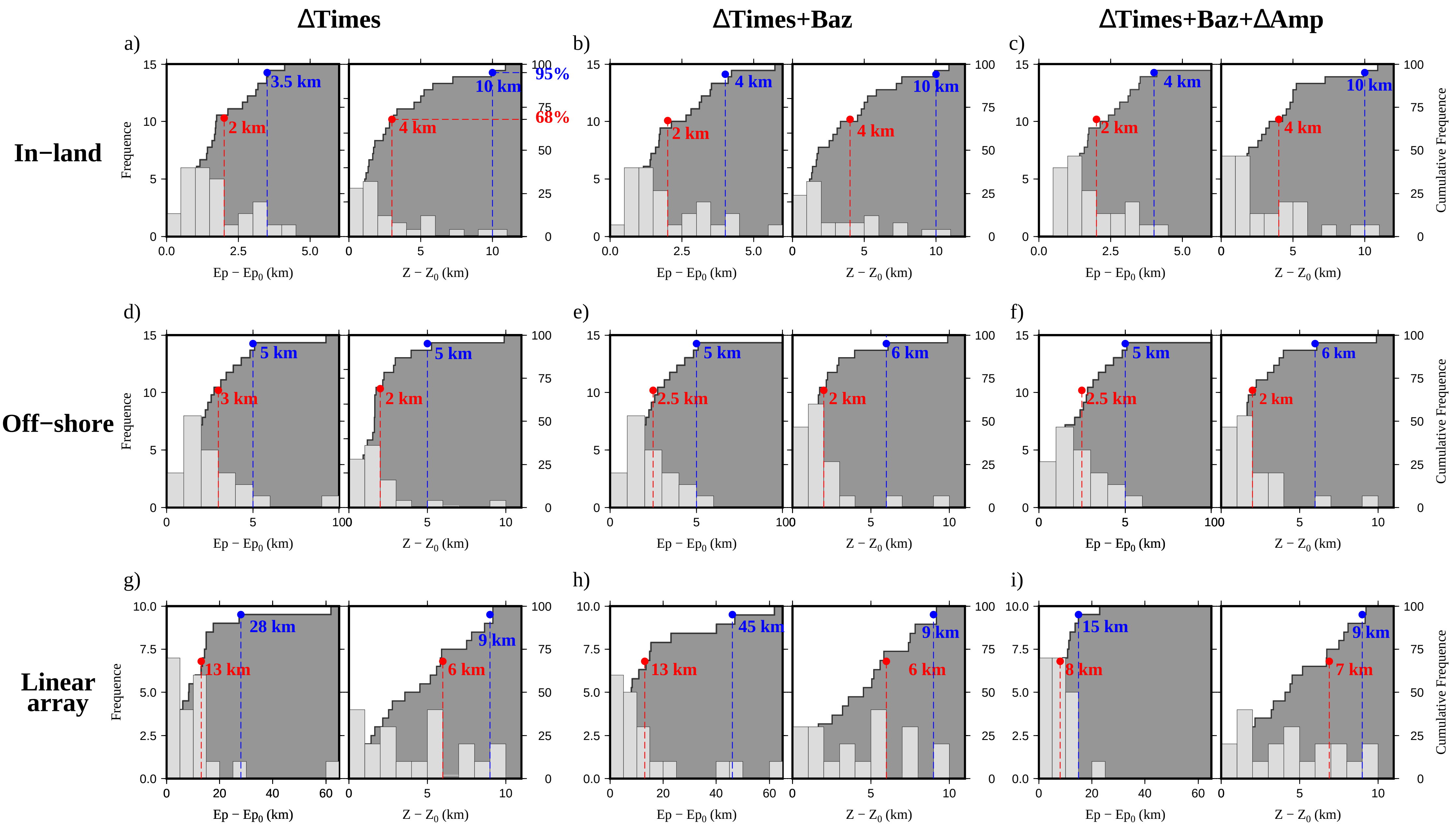

a)
b)
c)
d)
1
2
3
4
After 2.6 s
After 3.1 s
After 4.1 s
After 9.1 s
E-W(Km)
N-S(Km)
Z(Km)
Norm. Prob.
Normalized Probability



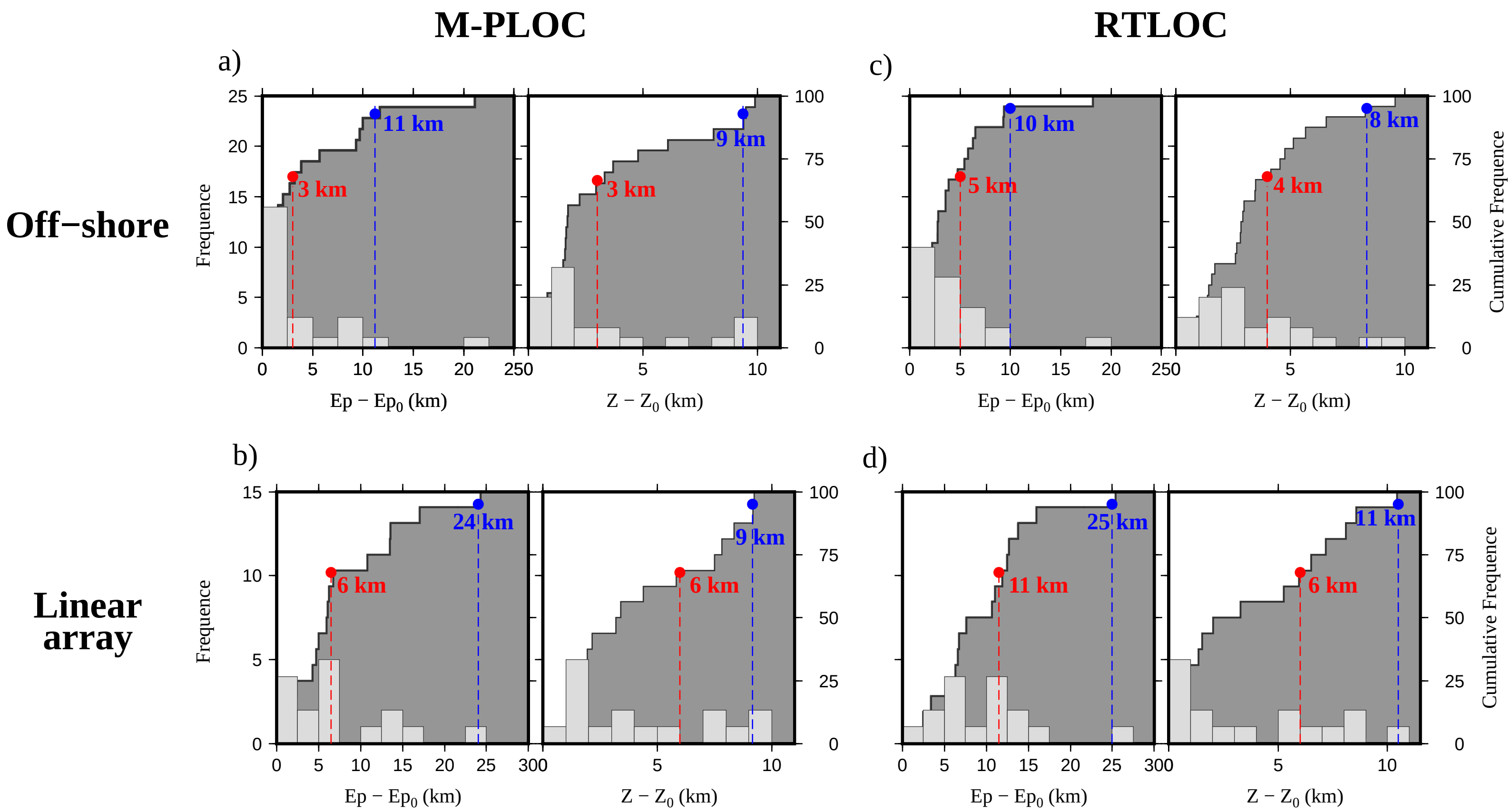
6
In-land
Off-shore
Linear array
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
o)
Ep−Ep$_0$ (Km)
Ep Error (Km)
Z−Z$_0$ (Km)
Z Error (km)
T0−T0$_0$ (s)
Number of stations


After 2 s
ΔTimes
ΔTimes+Baz
ΔTimes+Baz+ΔAmp
a)
b)
c)
d)
e)
f)
g)
h)
i)
In−land
Off−shore
Linear array
Frequence
Cumulative Frequence
Ep − Ep0 (km)
Z − Z0 (km)
95%
68%
60 km
16 km
9 km
5 km
44 km
4 km
6 km
10 km
7 km
25 km
8 km
24 km
83 km
2.5 km
12 km

After 4 s


ΔTimes
ΔTimes+Baz
ΔTimes+Baz+ΔAmp
In−land
Off−shore
Linear array
a)
b)
c)
d)
e)
f)
g)
h)
i)
Frequence
Cumulative Frequence
Ep − Ep0 (km)
Z − Z0 (km)
95%
68%
3.5 km
2 km
10 km
4 km
5 km
3 km
2.5 km
6 km
28 km
13 km
9 km
6 km
45 km
15 km
8 km
7 km

After 3 s


M-PLOC
RTLOC
a)
b)
c)
d)
Off−shore
Linear array
Frequence
Cumulative Frequence
3 km
11 km
9 km
5 km
10 km
4 km
8 km
6 km
24 km
11 km
25 km
Ep − Ep0 (km)
Z − Z0 (km)