Geophysical Journal International

Geophys. J. Int. (2020) **221,** 1941–1958
Advance Access publication 2020 March 18
GJI Seismology

doi: 10.1093/gji/ggaa111

Migration-based near real-time detection and location of microearthquakes with parallel computing

Mariangela Guidarelli, Peter Klin and Enrico Priolo

National Institute of Oceanography and Applied Geophysics - , 34010 Sgonico, Trieste, Italy. E-mail: mguidarelli@inogs.it

Accepted 2020 March 11. Received 2020 January 22; in original form 2019 February 11

SUMMARY

Prompt detection and accurate location of microseismic events are of great importance in seismic monitoring at local scale and become essential steps in monitoring underground activities, such as oil and gas production, geothermal exploitation and underground gas storage, for implementing effective control procedures to limit the induced seismicity hazard. In this study, we describe an automatic and robust earthquake detection and location procedure that exploits high-performance computing and allows the analysis of microseismic events in near real-time using the full waveforms recorded by a local seismic network. The implemented technique, called MigraLoc, is based on the space–time migration of continuous waveform data and consists of the following steps: (1) enhancement of *P* and *S* arrivals in noisy signals through a characteristic function, by means of the time–frequency analysis of the seismic records; (2) blind event location based on delay-and-sum approach systematically scanning the volume of potential hypocentres; (3) detection notification according to the information content of the hypocentre probability distribution obtained in the previous step. The technique implies that theoretical arrival times are pre-calculated for each station and all potential hypocentres as a solution of the seismic-ray equation in a given 3-D medium. As a test case, we apply MigraLoc to two, low-magnitude, earthquake swarms recorded by the Collalto Seismic Network in the area of the Veneto Alpine foothills (Italy) in 2014 and 2017, respectively. Thanks to MigraLoc, we can increase the number of events reported in the network catalogue by more than 25 per cent. The automatically determined locations prove to be consistent with, and overall more accurate than, those obtained by classical methods using manual time-arrival picks. The proposed method works preferably with dense networks that provide signals with some degree of coherency. It shows the following advantages compared to other classical location methods: it works on the continuous stream of data as well as on selected intervals of waveforms; it detects more microevents owing to the increased signal-to-noise ratio of the stacked signal that feeds the characteristic function; it works with any complex 3-D model with no additional effort; it is completely automatic, once calibrated, and it does not need any manual picking.

Key words: Time-series analysis; Computational seismology; Earthquake monitoring and test-ban treaty verification; Earthquake source observations.

1 INTRODUCTION

Microseismicity detection and location has become an increasingly important research topic in the last years because it provides important information on active processes in the subsurface (Grigoli *et al.* 2017). It is especially important in connection with the monitoring of the seismicity possibly induced by several industrial activities (oil and gas reservoirs, mines, water reservoirs, geothermal systems, etc.) that cause stress field perturbations in the seismogenic layer. Improving the event detection sensitivity will not only benefit the understanding of the geomechanical processes that drive induced seismicity but also allow the prompt application of safety procedures (Grigoli *et al.* 2017). Microseismic monitoring requires continuous data acquisition and therefore the processing of many events for picking (i.e. determination of phase onset arrivals) and location in near real-time. Automated and robust procedures for systematically detecting and locating microevents are mandatory (Cesca & Grigoli 2015; Grigoli *et al.* 2017). The main difficulty in the implementation of these procedures consists in dealing with the low signal-to-noise ratio (SNR), which characterizes the traces of the microevents collected by local seismic networks. Most of the standard automated location routines rely on a detection step in

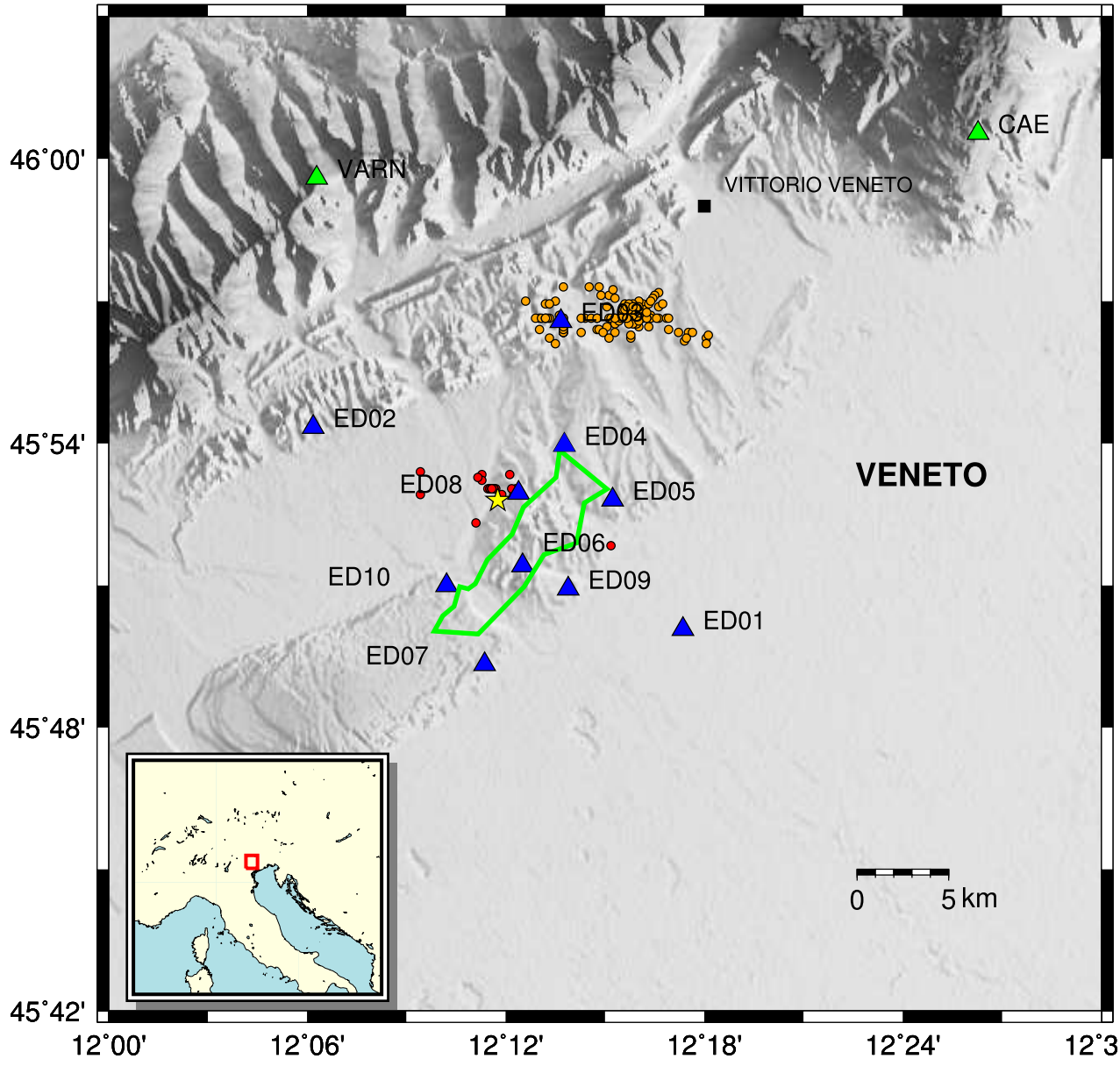

Figure 1. Location of the Collalto Seismic Network (blue triangles) and the stations of the Northeast Italy Seismic Network used in this study (green triangles). The yellow star marks an event occurred in the study area used to illustrate the methodology (01.05.2014, 13:26:24.01UTC, Lat = 45.880, Lon = 12.196, depth = 9.6 km, after manual location). The irregular area marked by a green line corresponds to the surface projection of the Collalto gas storage reservoir. The red dots show the location of the seismic events from the May 2014 swarm as reported in the OGS catalogue, while orange dots show the locations from the February 2017 swarm.

which the identification and association of the *P*- and *S*-wave first arrivals are performed. This part includes the precise estimation of the *P*- and *S*-phase arrival time, that is, automatic picking.

The most commonly used algorithm of signal detection is the STA/LTA detector proposed by Allen (1982), which is based on the ratio of a short-term average (STA) to a long-term average (LTA) calculated on sliding windows over a characteristic function (CF) extracted from the trace. Baer & Kradolfer (1987) introduced the envelope function as the CF and used a dynamic threshold to detect signals buried in noise. Sleeman & van Eck (1999) and Leonard & Kennett (1999) proposed auto-regressive (AR) techniques, based on the Akaike Information Criterion (AIC; Akaike 1974), to pick the phase onset. Saragiotis *et al.* (2002) suggested the application of higher order statistical functions (skewness and kurtosis) to seismic traces for phase picking. Küperkoch *et al.* (2010) added the AIC to Saragiotis' method and developed a quality-weighting scheme for picks. For most of the picking algorithms, the *P* onsets can be accurately picked, while for *S* and later seismic phases the performance is limited especially by both the presence of noise in the data and the overlapping of phases (Cesca & Grigoli 2015). A positive detection starts the location step, which is based on the minimization of the difference between the observed and predicted arrival times of the considered seismic phases. If we exclude the joint hypocentre and velocity inversion approach, which requires a dense array of receivers and a large number of events (e.g. Davenport *et al.* 2015), the prediction of arrival times requires a previously determined velocity model. Even though homogeneous horizontal layers are routinely used for travel-time calculation, the application of 3-D

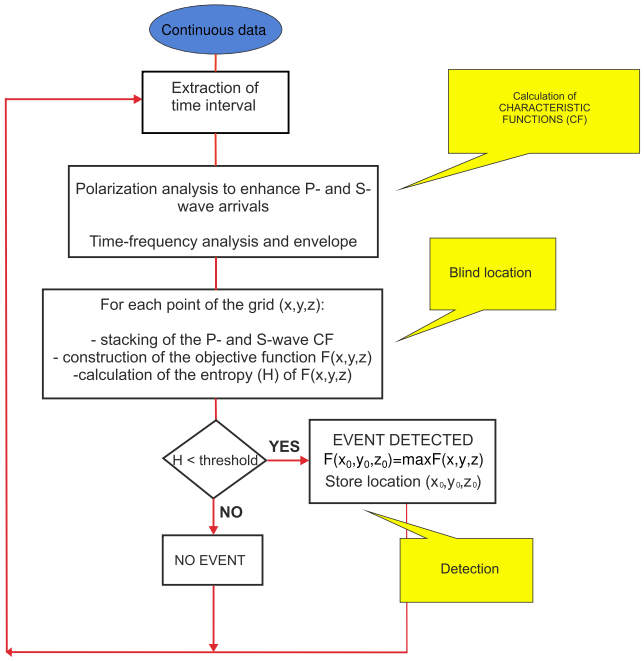

Figure 2. Flow diagram of the detection and location algorithm.

velocity models may improve the earthquake location in case of significant lateral structural variations (e.g. Chen *et al.* 2006).

In the recent years, alternative techniques for automated seismic event location, similar to the migration techniques used in reflection seismics, have been developed. These methods are based on the concept of delay and stack of seismic waveforms and do not need prior phase picking nor phase identification. Kao & Shan (2004, 2007) proposed the Source Scanning Algorithm (SSA) that localizes seismic events on regional scale making use of the maxima of a brightness function defined in space and time. The value of the brightness function at a given time is defined in a volume of potential hypocentre locations as the sum of the absolute value of the normalized seismogram sample recorded at different stations and anticipated by the predicted traveltime from the potential hypocentre location to the station. In SSA, the predicted traveltime refers to the *S*-wave velocity model, since it is assumed that *S* waves produce the largest amplitudes. Other authors proposed approaches that follow the same principle as SSA: Baker *et al.* (2005) used the term image instead of brightness function in an approach which consists in summing up the envelopes of the three components, considering *P* traveltimes for the ray parallel component and *S* traveltimes for the other two components, respectively. Gharti *et al.* (2010) proposed to find the maxima of the objective function (as they renamed the brightness function) with a search algorithm instead of the systematic full grid search. Grigoli *et al.* (2013a,b) substituted the waveform envelopes with STA/LTA traces, using the waveform energy as CF. Their approach enhances the contribution of the arrivals recognizable by STA/LTA to the coherence function (their equivalent of the brightness function) and uses polarization analysis to enhance *P*- and *S*-wave arrivals. Grigoli *et al.* (2016) recently improved their previous methodology combining the waveform stacking approach with some features of the relative location methods, in order to reduce the dependency of the results on the crustal velocity model. The previous methods work on already identified seismic events and they do not include any event detection scheme, but in the recent years some methodologies have been introduced for the simultaneous detection and location of seismic events. Langet *et al.* (2014) presented an automatic earthquake detection and location technique also based on the migration of continuous waveform data, pre-processed using a kurtosis estimator instead of STA/LTA. Higher order statistics was also used by Poiata *et al.* (2016). They developed a method for simultaneous detection and location of seismic sources based on multiband filtering and higher-order statistic characterization of the signal. Then, back-projection and stacking of time-delay functions, evaluated by cross-correlation of transformed signals, are used to produce 3-D spatial images representing the likelihood of each point to be part of a seismic source. Grigoli *et al.* (2017) introduced a detection method based on a modified version of the waveforms stacking location method developed by Grigoli *et al.* (2013a). Most of these studies are based on the information provided by *P*-wave arrivals only (e.g. Langet *et al.* 2014) and ignore the contribution from *S* waves. They also perform earthquake detection using simple triggering algorithms applied on time-series, generally coherency functions, and simplified strategies for error or resolution estimation.

In this paper, we present a modified implementation of the SSA approach, which allows the automatic detection and location of microseismic events in near real-time using data recorded by regional and local seismic networks. We make use of information provided by both *P*- and *S*-wave arrivals and introduce the definition of entropy to investigate the presence of a seismic event within the CFs obtained from the analysed seismic data. Our procedure is based on the following steps:

(1) enhancement of possible *P* and *S* arrivals in noisy signals by transforming waveforms into CFs, obtained by a time–frequency (TF) polarization analysis of the signals;

(2) delay-and-sum of the CFs for a set of points elected as hypocentre candidates in a pre-defined investigated volume (typically a grid). Parallel computing is adopted in order to manage huge numbers of points. The product of the stacked CFs for *P* and *S* waves defines an objective function (*F*) over the points;

(3) notification of the event detection based on the information content in *F*.

In parallel computing, we denote a problem as 'perfectly parallel' or 'embarrassingly parallel' if we can solve it by decomposing the workload in a number of parallel (concurrent) tasks which are independent of each other. This is the case of the evaluation of the objective function *F* in our procedure: we can assign the calculation of *F* on subsets of points to different tasks, which do not need to communicate among them during the calculations. The perfect parallelism of the *F* evaluation problem at step 2 allows a straightforward application of high-performance computing (HPC) approaches in order to reduce the execution time of the code below the duration of the analysed signals, making it possible the near real-time application of the procedure.

In the following sections, we first describe our methodology for automatic detection and location of microseismic events in detail. Next, we apply the procedure to synthetic waveforms in order to evaluate its robustness and performance. Finally, we describe two real case applications, corresponding to two swarms of low-magnitude earthquakes recorded in the area monitored by the Collalto Seismic Network (northeast Italy, see Fig. 1; Priolo *et al.* 2015; Moratto *et al.* 2019). First, we use the manual picks and locations (www.rete-collalto.crs.inogs.it) of the first swarm to finely tune our method; then, we apply our method to the second swarm, and compare the automatic detections and locations with those obtained manually by the monitoring operators

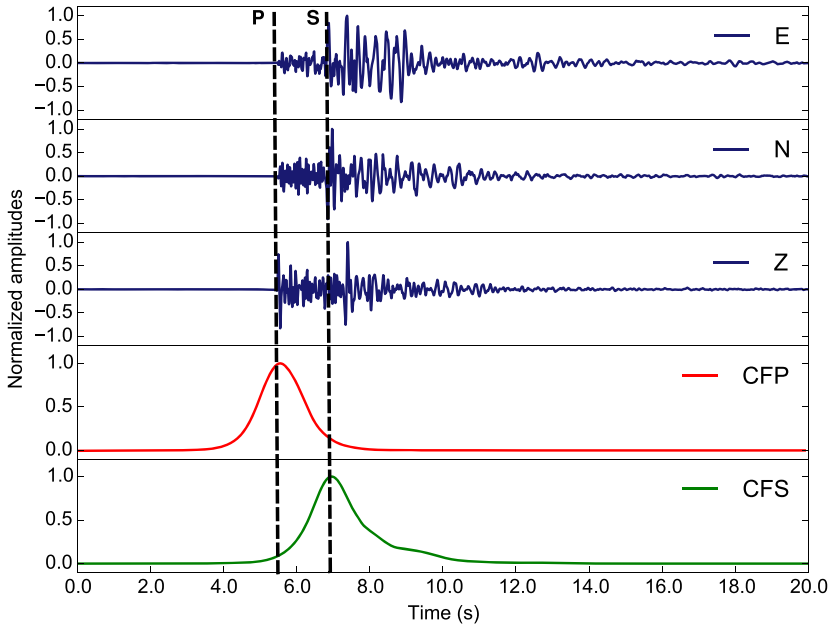

Figure 3. Waveforms related to the M_L1.4 seismic event denoted by the yellow star in Fig. 1 and recorded by station ED08. We show the three component seismograms and the characteristic function for the *P* waves and for the *S* waves.

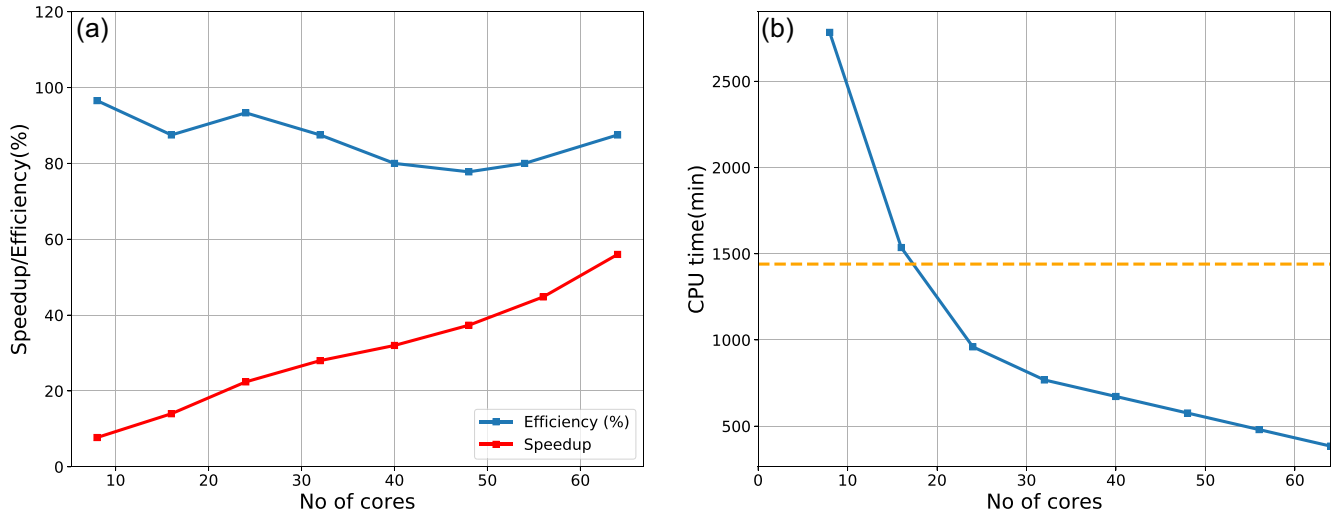

Figure 4. (a) Speedup and efficiency as a function of number of cores used for the code runs on the HPC resources of the CINECA consortium, Bologna. (b) Running time as a function of number of cores: the orange dashed line corresponds to one-day time interval.

(www.rete-collalto.crs.inogs.it). The comparisons indicate that the procedure is at least as accurate as a trained analyst. MigraLoc runs in near real-time on the HPC resources available at the CINECA consortium, Bologna, Italy.

2 METHODOLOGY

The first step of our approach consists in the partition of the data flow received from the seismic network into overlapping time windows that will be processed sequentially. The window length is chosen according to the maximum possible delay in the *S*-wave arrival among the recording stations that could be expected for a hypocentre in the investigated volume. The overlap is estimated as the maximum expected difference between *P*- and *S*-wave arrival times.

The procedure for the detection and localization is repeated on each time window (as illustrated schematically in Fig. 2). Each iteration can be divided in three steps: (i) transformation of waveforms in CFs, (ii) 'blind' localization with the estimation of the hypocentre objective function and (iii) identification of the event. Below we describe the three steps of the proposed methodology.

2.1 Characteristic functions

Since the localization procedure depends on the individuation of *P*- and *S*-phase onsets in the recorded waveforms, we process the waveforms in order to transform them in adequate time-series, called CFs, in which these onsets are enhanced.

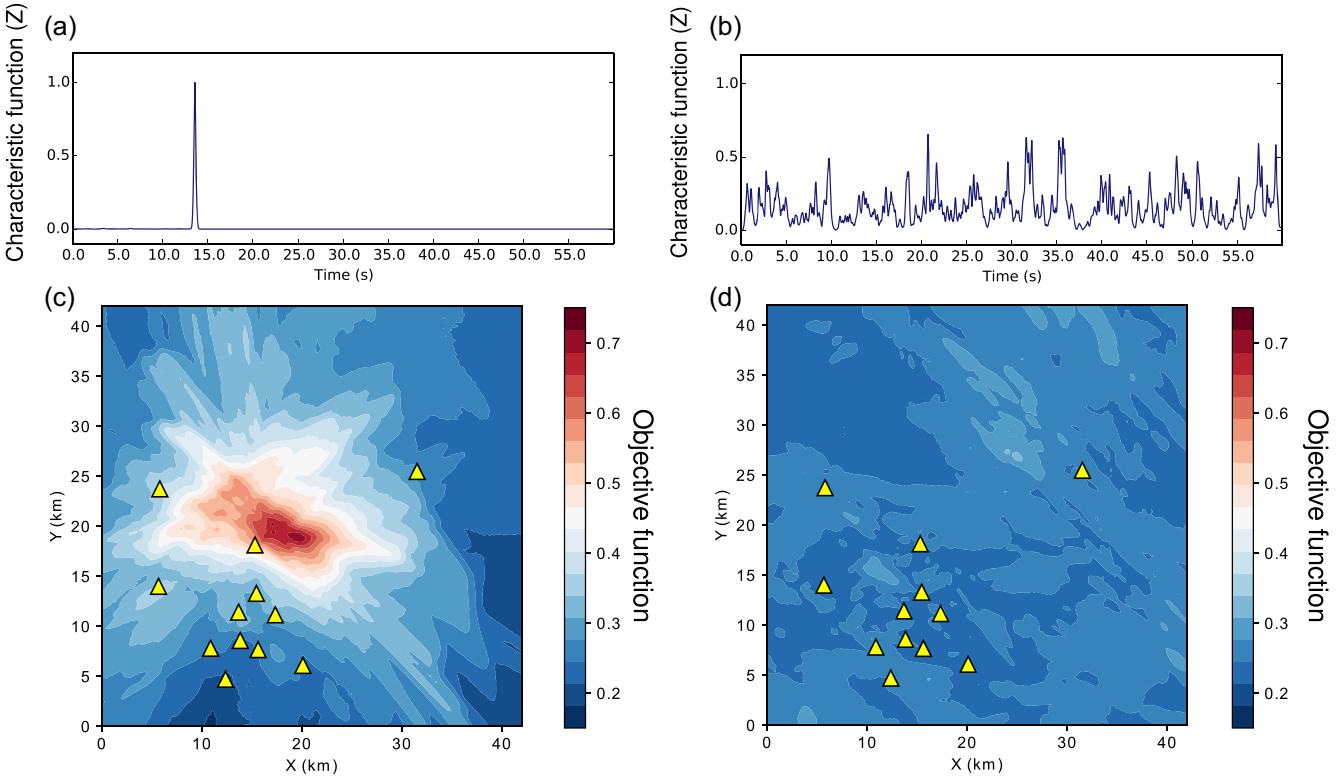

Figure 5. (a and b) Examples of a characteristic function for the vertical component of a signal containing a seismic event and random noise, respectively; (c and d) examples of the projection on the horizontal plane of the maximum values of the objective function F along the Z-direction obtained from signals containing a seismic event and random noise, respectively.

Table 1. Seismic properties of the horizontally layered medium used to compute synthetic seismograms and arrival times (Model 1).

Model 1	V_P (km s^{-1})	V_P/V_S	ρ (kg m^{-3})
Layer 1	5.85	1.78	2.7
Layer 2	6.80	1.78	2.9
Half-space	8.00	1.78	3.3

Table 2. Location results after synthetic tests and comparison with true source location.

	Source location					
	x (km)	y (km)	z (km)	t_0 (s)	F	H
True	12.4	12.4	1.0	0.0		
0% noise	12.4	12.4	1.0	0.025	0.997	0.9887
10% noise	12.4	12.4	1.0	0.025	0.992	0.9917
30% noise	12.4	12.4	1.0	0.045	0.981	0.9934
70% noise	12.4	12.4	1.0	0.045	0.962	0.9910
Real noise	12.4	12.4	1.0	0.045	0.9980	0.9932
Wrong model	12.2	12.3	0.8	-0.036	0.716	0.9713
Event not in the region[a]	12.1	0.0	16.8	0.114	0.632	0.9961

[a]The true location for the event is outside of the study region: $(x, y, z) = (13.4, -5.4, 1.0)$. Units in km.

A key factor in the construction of our CFs is the polarization analysis. The analysis of particle motion recorded in three component seismographs can be used to determine the seismic wave type as shown by Flinn (1965), Montalbetti & Kanasewich (1970), Rosenberger (2010) and Baillard *et al.* (2014). Like Ross & Ben-Zion (2014), Ross *et al.* (2016) and Jurkevics (1988), we use the covariance matrix eigenvectors and eigenvalues to compute a set of polarization attributes. These quantities can be applied to identify body wave phases because compressional and shear waves exhibit a high degree of linear polarization in contrast to any Rayleigh-type wave (Flinn 1965).

The covariance matrix for a set of n points taken over three-component seismic waveform is computed as

$$\sigma = \begin{bmatrix} \mathrm{cov}(u_1, u_1) & \mathrm{cov}(u_1, u_2) & \mathrm{cov}(u_1, u_3) \\ \mathrm{cov}(u_1, u_2) & \mathrm{cov}(u_2, u_2) & \mathrm{cov}(u_2, u_3) \\ \mathrm{cov}(u_1, u_3) & \mathrm{cov}(u_2, u_3) & \mathrm{cov}(u_3, u_3) \end{bmatrix}, \quad (1)$$

where u_1 and u_2 are the horizontal components and u_3 is the vertical component and cov is the covariance operator. As in Ross & Ben-Zion (2014), we compute σ using a sliding window with appropriate length τ. Ross & Ben-Zion (2014) and Baillard *et al.* (2014) found that 3–4 s long sliding windows are mostly appropriate to evaluate the covariance matrices; however, in our application, dealing with small magnitude events on short distances, sliding windows with shorter lengths can provide better results. Some tuning is therefore required in order to find the optimal value for τ. As described more in detail in Sections 3 and 4, the tuning consists in iterating the procedure with different values of τ on a smaller set of waveforms, which contains a number of known events, and selecting the τ value, which allows the highest number of earthquake detections. We use the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and the corresponding orthogonal unitary eigenvectors $\boldsymbol{e_1}$, $\boldsymbol{e_2}$ and $\boldsymbol{e_3}$ of σ to estimate the rectilinearity

$$r = 1 - \left(\frac{\lambda_2 + \lambda_3}{\lambda_1}\right) \quad (2)$$

and the cosine of the apparent vertical incidence angle

$$\cos(\varphi) = e_{13}. \quad (3)$$

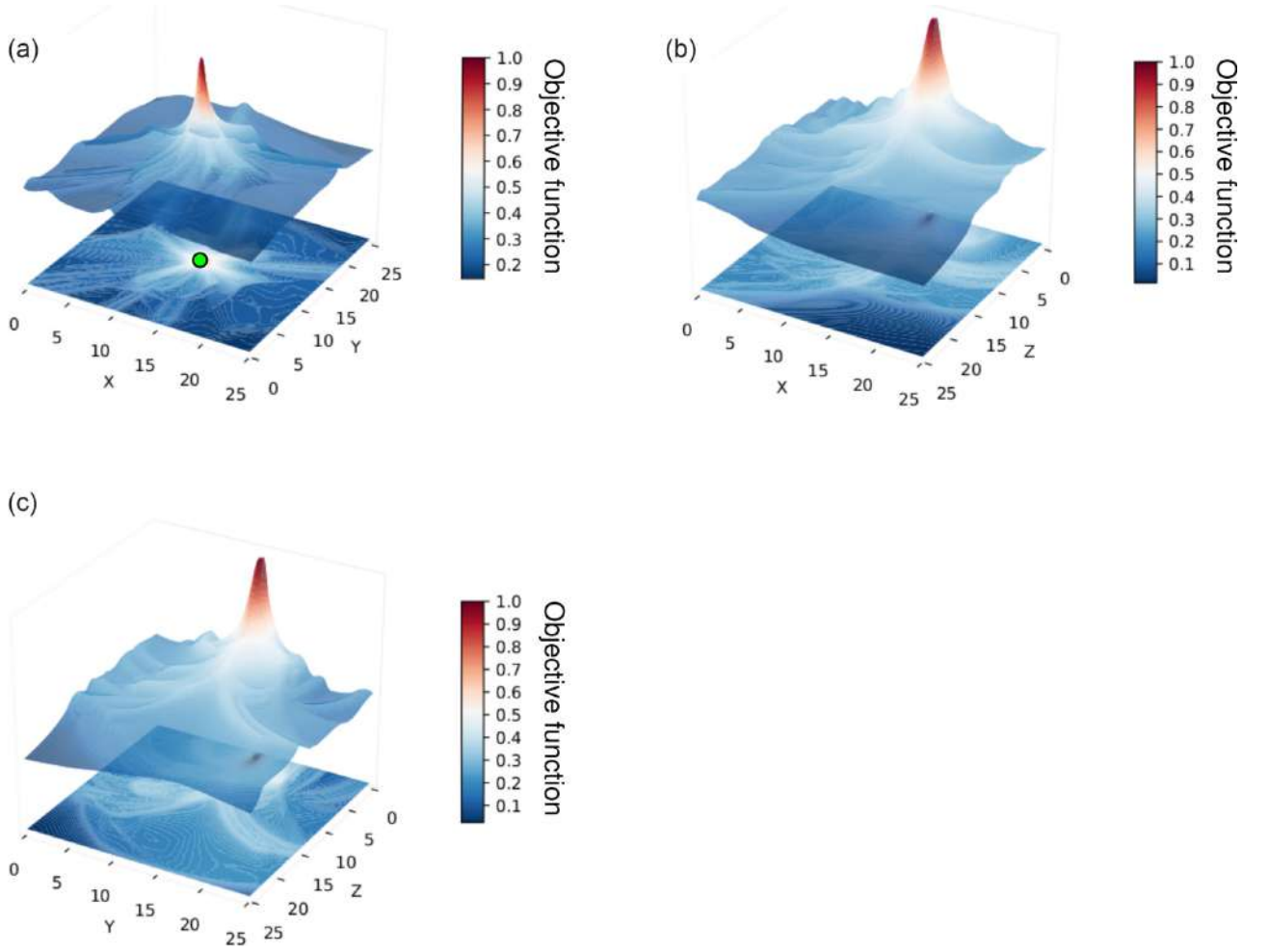

Figure 6. *F* resulting from the synthetic test, projected on cross-sections through the true source location: (a) *xy*-plane, (b) *xz*-plane and (c) *yz*-plane. The true source location is shown by the green dot.

With these quantities, we modulate the three-component data (e.g. Rosenberger 2010) to enhance the *P* waves in the vertical component and the *S* waves in both horizontal components:

$$\begin{aligned} u_P &= u_3\, r \cos(\varphi) \\ u_{S1} &= u_1\, r\,(1 - \cos(\varphi)) \\ u_{S2} &= u_2\, r\,(1 - \cos(\varphi))\,. \end{aligned} \tag{4}$$

As an alternative to the polarization analysis based on the real covariance matrix, it is possible to use the complex covariance matrix as computed by Vidale (1986).

To further enhance the first arrivals, we perform the TF analysis by applying a bank of *N* Gaussian narrow-band frequency filters

$$G_i = \exp\left[-\alpha_i \left(\frac{\omega - \omega_i}{\omega_i}\right)^2\right] \quad i = 1, \ldots, N, \tag{5}$$

where ω is the circular frequency and ω_i and α_i are parameters controlling the central frequency and relative width of filters, respectively. We consider 64 Gaussian filters with central frequencies ω_i, logarithmically distributed over a frequency interval to build a TF representation $U_P(\omega_i,t)$, $U_{S1}(\omega_i,t)$, $U_{S2}(\omega_i,t)$ of the records modulated as in eq. (4). The multiple filter technique aims at minimizing the effect of background noise (e.g. Alvarez *et al.* 2013), processing separately each band that can be affected by noise in different ways. We use Gaussian filters because they do not affect the phase of the signal and they can be easily applied in the frequency domain.

The characteristic function is calculated from the TF representation by stacking the envelopes obtained in each frequency band. Finally, to further reduce the noise in each channel, a threshold is applied to the envelope

$$\mathrm{env}(U_i) = \begin{cases} 1 & \text{if } \mathrm{env}(U_i) \le \mathrm{thr} \\ \mathrm{env}(U_i)/\mathrm{thr} & \text{if } \mathrm{env}(U_i) > \mathrm{thr} \end{cases}. \tag{6}$$

The threshold 'thr' is chosen as a percentile of the amplitude of env(U_i) for each component. In our application, we used the 90 per cent percentile after some preliminary tests, as described in the fourth paragraph. Then env(U_i) *i*s normalized to the peak value and we finally obtain

$$\mathrm{CF}_P(t) = \left(\frac{\sum_{i=1}^{N} \mathrm{env}(U_P(\omega_i, t))}{\left\| \sum_{i=1}^{N} \mathrm{env}(U_P(\omega_i, t)) \right\|_\infty}\right)^2 \tag{7}$$

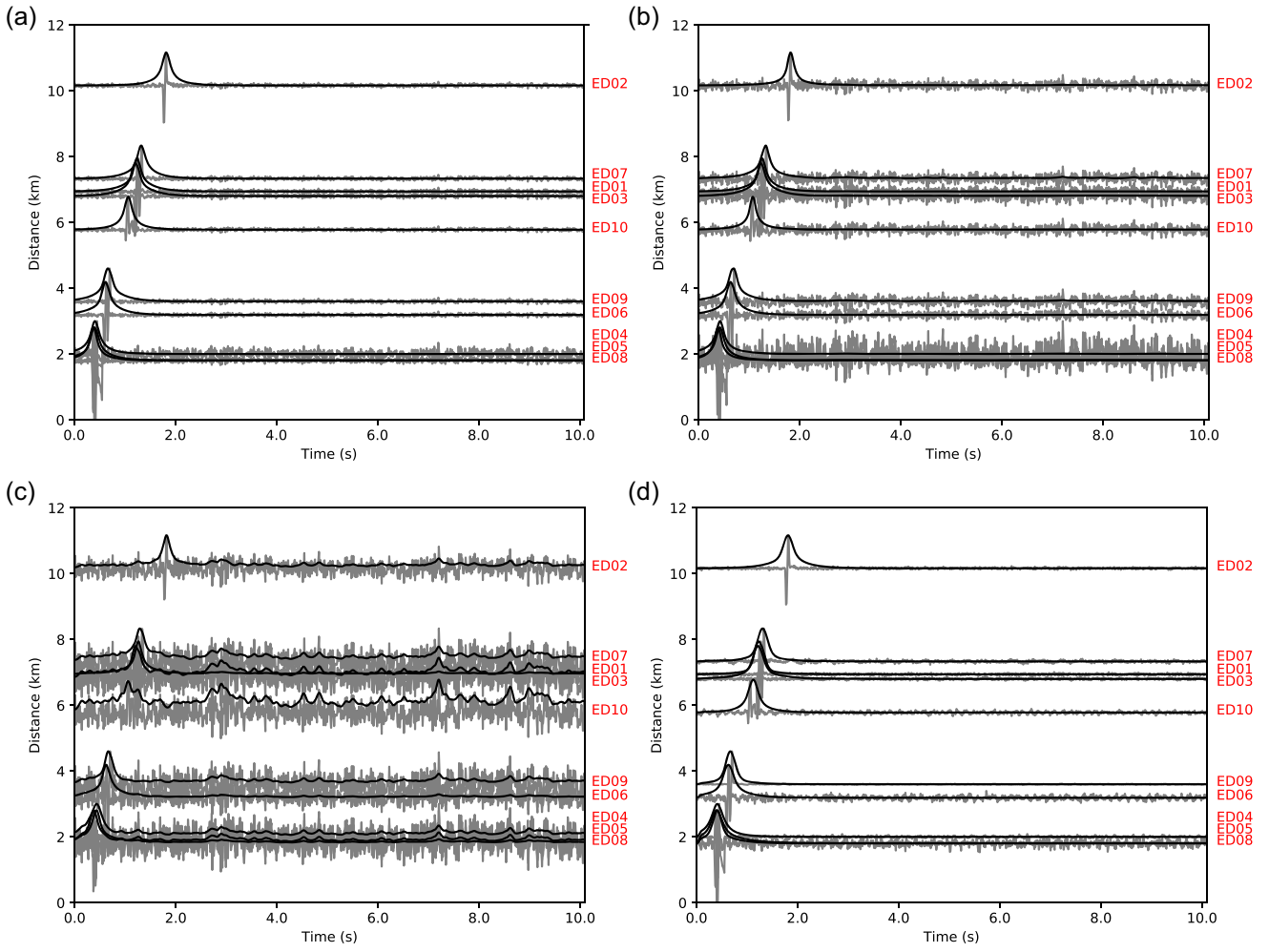

Figure 7. Synthetic seismograms (vertical component) used for synthetic tests, with different noise levels: (a) 10 per cent of the maximum amplitude; (b) 30 per cent of the maximum amplitude; (c) 70 per cent of the maximum amplitude; (d) real noise conditions.

and

$$\mathrm{CF}_S(t) = \frac{1}{2}\left[\left(\frac{\sum_{i=1}^{N}\mathrm{env}\left(U_{S1}(\omega_i,t)\right)}{\left\|\sum_{i=1}^{N}\mathrm{env}\left(U_{S1}(\omega_i,t)\right)\right\|_\infty}\right)^2 + \left(\frac{\sum_{i=1}^{N}\mathrm{env}\left(U_{S2}(\omega_i,t)\right)}{\left\|\sum_{i=1}^{N}\mathrm{env}\left(U_{S2}(\omega_i,t)\right)\right\|_\infty}\right)^2\right] \quad (8)$$

where env() denotes the envelope and ||-||$_\infty$ denotes the infinity norm that corresponds to the maximum absolute value. In Fig. 3, we show an example of characteristic functions calculated for a seismic event (the yellow star in Fig. 1) at station ED02.

2.2 Estimation of the objective function

At each iteration, we use the characteristic functions CF_P and CF_S in a location procedure without any prior attempt to recognize whether they are describing an earthquake signal or not; we therefore refer to this step as *blind location*. The location approach we use is based on the delay-and-sum process introduced by Kao & Shan (2004) and successively modified by several authors (Baker *et al.* 2005; Kao & Shan 2007; Gharti *et al.* 2010; Grigoli *et al.* 2013a,b). The scheme we use in this study is similar to that described by Grigoli *et al.* (2013a,b). We sample the investigated volume of possible hypocentres with a regular 3-D grid. The spatial sampling is chosen according to the extension of the study area and the time sampling step of the seismic input data. The theoretical *P* and *S* first arrival times at each station, for each trial location in the grid, are computed using the velocity model before the execution of the procedure. To this aim, we use the NonLinLoc software (Lomax *et al.* 2000) based on the finite-difference solution of the eikonal equation by Podvin & Lecomte (1991). The CF_P and CF_S are then shifted in time according to the theoretical *P*- and *S*-wave arrival times at each station and stacked. The product of the stacked CF_P and CF_S defines our objective function (F) at each grid point, similarly to the coherence matrix used by Grigoli *et al.* (2013a,b):

$$F\left(x_j,y_k,z_l\right) = \left[\left(\sum_{m=1}^{M}\mathrm{CF}_P^{(m)}\left(t-t_P^{(m)}\left(x_j,y_k,z_l\right)\right)\right) \times \left(\sum_{m=1}^{M}\mathrm{CF}_S^{(m)}\left(t-t_S^{(m)}\left(x_j,y_k,z_l\right)\right)\right)\right]_\infty, \quad (9)$$

where M is the number of stations, $t_P{}^{(m)}(x_j,y_k,z_l)$ and $t_S{}^{(n)}(x_j,y_k,z_l)$ are the theoretical arrival times and ($j,k,l, j = 1,\ldots,nx, k = 1,\ldots,ny, l = 1,\ldots,nz$) are the grid point indexes. The term objective function was already used by Gharti *et al.* (2010) to indicate an analogous function. In theory, if the velocity model is correct, an event with

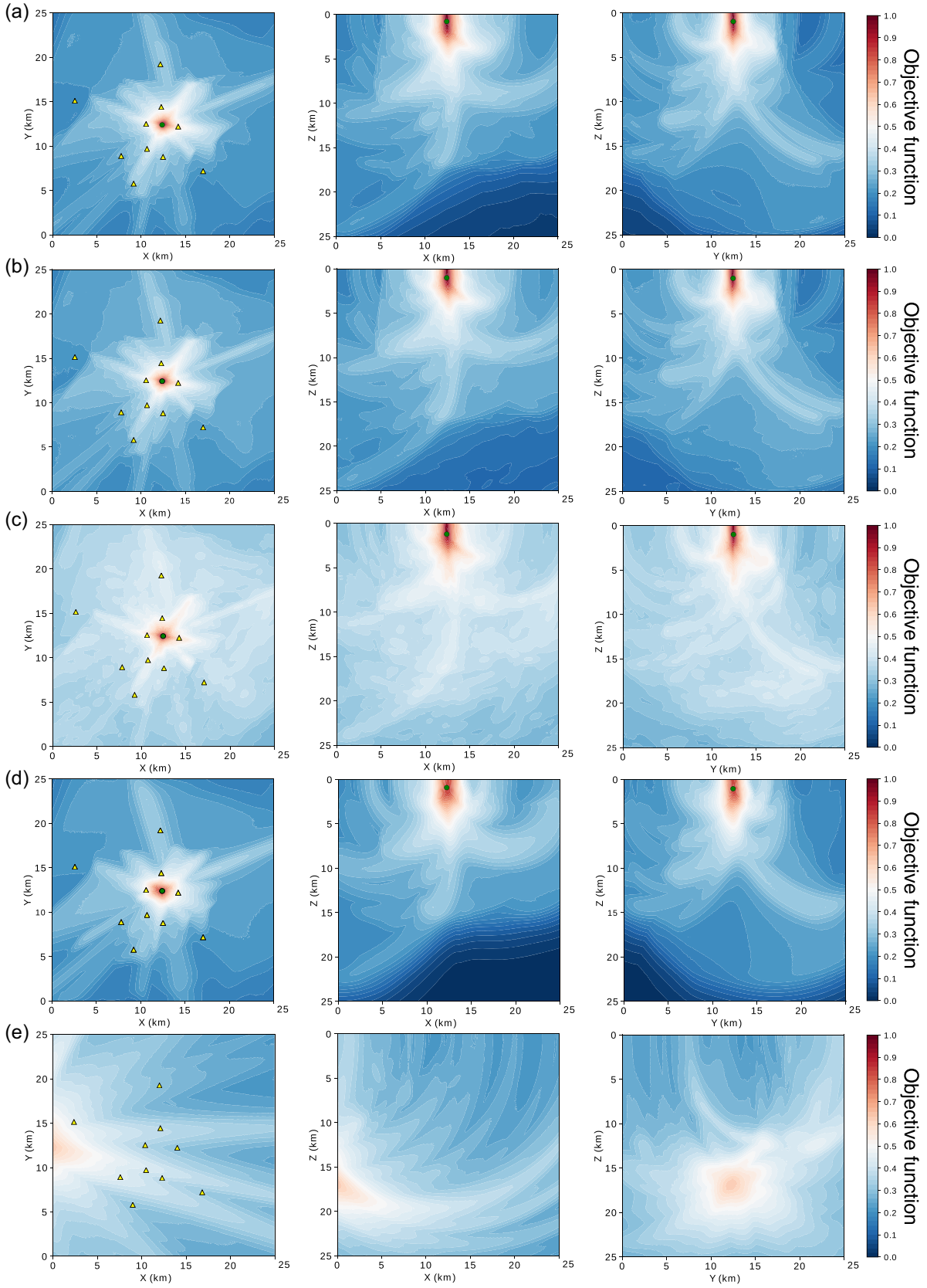

Figure 8. *F* resulting from the synthetic test, projected on cross-sections through the true source location: (left) *xy*-plane, (centre) *xz*-plane and (right) *yz*-plane; the yellow triangles represent the stations used for synthetic tests; the green dot represents the location of the event obtained by the automated procedure. (a) Noise level at 10 per cent; (b) noise level at 30 per cent; (c) noise level at 70 per cent; (d) wrong crustal model; (e) event occurred outside of the 3-D computation grid. *F* for the synthetic test with real noise conditions is not reported, being similar to the case with noise level at 30 per cent.

Table 3. Seismic properties of the horizontally layered medium used to compute synthetic seismograms (Model 2).

Model 2	V_P (km s^{-1})	V_P/V_S	ρ (kg m^{-3})
Layer 1	5.03	1.86	2.7
Layer 2	5.14	1.85	2.7
Layer 3	5.90	1.82	2.7
Layer 4	5.78	1.83	2.7
Layer 5	6.14	1.83	2.7
Layer 6	6.12	1.83	2.7
Half-space	8.00	1.78	3.3

hypocentre in the investigated volume will produce an F with a sharp peak at the point corresponding to the hypocentre. Nevertheless, the generally used velocity models are only approximations of the real velocity structure, hence we expect at best that F shows a smooth peak in the neighbourhood of the hypocentre. On the other hand, in case of no events, we expect an F with a random pattern. We discuss how to discriminate the presence/absence of an event within F in Section 2.3.

The discrete domain of F may be considerably large (several tens of million points). For example, to test our application we used a grid of 250 × 250 × 250 points. To accomplish the location in near real-time, the wall-clock computational time requested by F should be kept shorter than the time window under analysis. In order to achieve this performance in the case of large computational domains, we parallelize the algorithm following the data decomposition approach, that is, we subdivide the volume of potential source locations among the available computational cores of the computer. The parallel algorithm is coded using the Message Passing Interface (MPI) paradigm (Message Passing Interface Forum, 2015) .

Tests performed with the HPC resources available at the Italian Consortium for supercomputing (CINECA) demonstrate that we can estimate F for 10 three-component station data, with 0.01 sampling interval, on a 250 × 250 × 250 points grid over 24 hr of seismic data in 14 hr 46 min using 32 computational cores on Intel Broadwell processors. We recall that in parallel computing the speedup measures the relative performance of a parallel algorithm in respect to the best sequential algorithm for the same problem and it is evaluated as the ratio of the runtime of the serial code to the runtime of the parallel code on a given number of computational cores. In the ideal case, we expect a speedup equal to the number of computational cores. The parallel efficiency is defined as the ratio of parallel speedup to the number of processors and in the ideal case it should remain 100 per cent but in practice lower values are typically observed. In our case, we observe an efficiency of above 80 per cent, which we consider an acceptable value.

In Fig. 4(a), the code speedup and efficiency as a function of the used cores are plotted. As shown in Fig. 4(b), the present version of the code requires a minimum of 20 cores to work in near real-time with the volume of potential sources corresponding to our tested application.

The main purpose of this test is to check the scalability of the implemented parallelization of the code and therefore to infer the feasibility of near real-time location considering the available computational resources and the given number of potential hypocentres. Of course, the same code would require significantly less computational resources if a smaller number of potential hypocentres was considered. However, we do not discuss here possible strategies one can adopt in order to optimize the number of potential hypocentres in respect to the investigated volume, the given velocity model and network geometry. Apart from the raw parallelization, we do not discuss here other possible optimizations in the procedure, either.

2.3 Detection and location

The objective function $F(x,y,z)$ is computed for subsequent time windows without prior knowledge about whether an earthquake has occurred within the corresponding time interval. In the following paragraphs we illustrate how we use the information contained in $F(x,y,z)$ to detect an earthquake. We denote the points that sample our investigated volume with $\boldsymbol{x_i}$ and for the sake of simplicity $F_i = F(\boldsymbol{x}_i)$.

In the case an event happened in the investigated volume during the analysed time window the function described in eq. (9) implies that values of F are higher in a restricted neighbourhood of the hypocentre location than in rest of the volume. If N is the number of points in which the investigated volume is discretized, we propose to describe the normalized F:

$$p(\boldsymbol{x}_i) = F(\boldsymbol{x}_i) \Big/ \sum_i^N F(\boldsymbol{x}_i) \tag{10}$$

as a weight we assign to the hypothesis that the hypocentre is in $\boldsymbol{x_i}$, on the basis of the adopted velocity model and the considered CFs. We can also describe p_i as the probability that an independent experiment (ideally error-less) would individuate the presence of an event in $\boldsymbol{x_i}$.

The weight p_i determines how much information will be added to our knowledge if we find out that $\boldsymbol{x_i}$ is in fact a hypocentre, thanks to the independent experiment. The amount of added information will be high if the calculated p_i is low (we underestimated the probability that the ith point could be the hypocentre from our calculations) and on the contrary, the gained knowledge will be low when p_i is high (we already calculated that $\boldsymbol{x_i}$ is a probable hypocentre). Following Shannon (1948), we can assign the information value to each point as

$$s_i = -\log_2 p_i . \tag{11}$$

The weighted average of s_i is called information entropy

$$H = -\sum_i^N p_i \log_2 p_i \tag{12}$$

which ranges continuously between the extremes $H_{\min} = 0$ and $H_{\max} = \log_2 N$. As we can infer from the previous reasoning, if the objective function returns a low H, then an independent experiment telling us there is an event in the volume will not increase significantly our knowledge because the event is already evident from the objective function by construction. We can therefore declare a detection as soon as the information entropy of the objective function drops below a threshold value H_{thr}. In that case we can associate the 'expected' hypocentre position as

$$\boldsymbol{x}_h = \sum_i^N p_i \boldsymbol{x}_i . \tag{13}$$

The origin time is then calculated from the P- and S-wave time-shifts associated with the point x_h. The point x_h should not be interpreted as 'the' hypocentre but only as a 'guess' for it in the case p_i is sufficiently 'regular'. On the contrary, a high value of H indicates that the objective function does not provide us any evidence for an event in the volume. Of course, a high H does not imply that there was no event, but only that we are unable to detect it from the objective function.

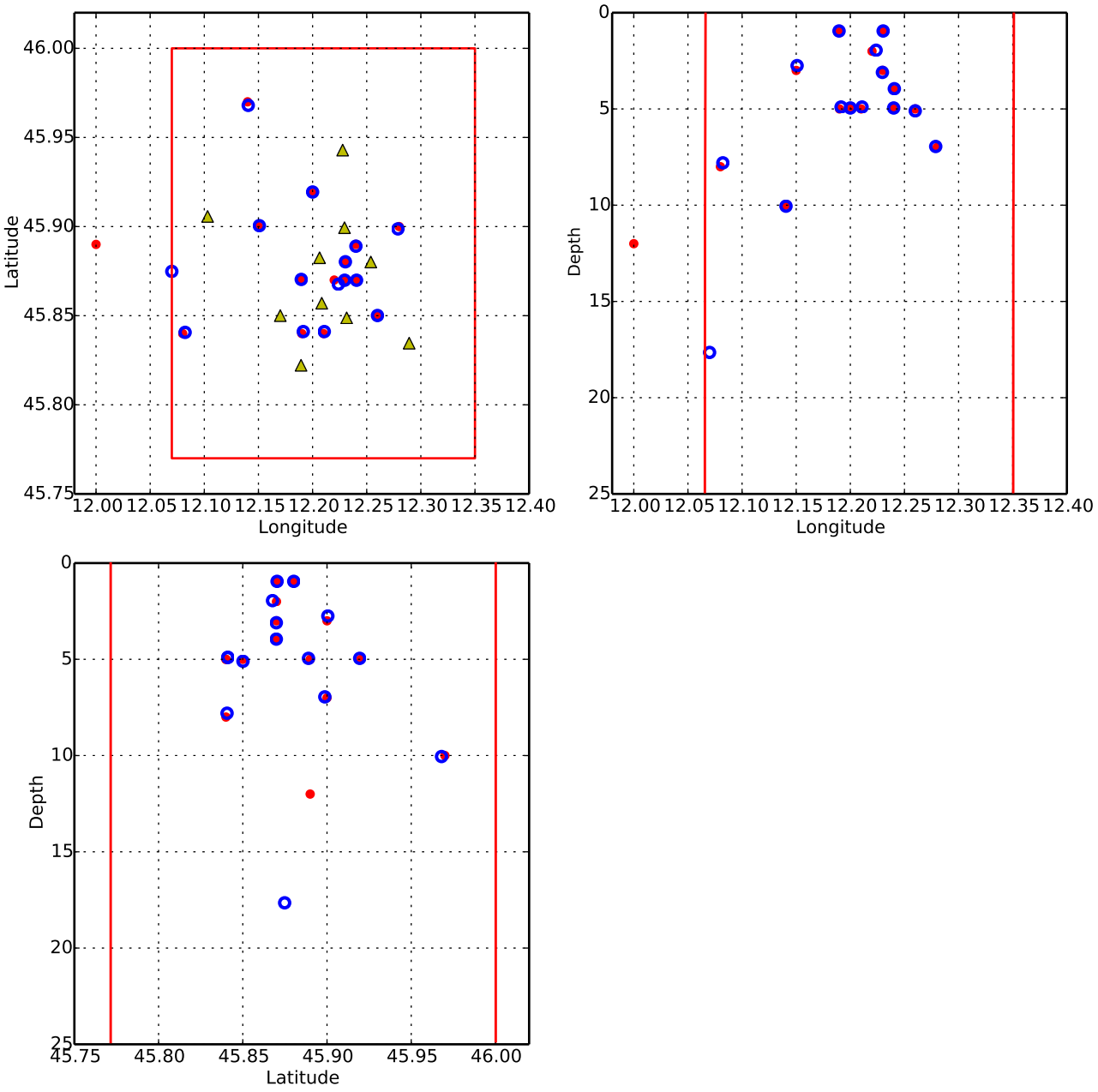

Figure 9. Results of the location for a synthetic test, considering synthetic waveforms corresponding to 15 sources located in different parts of the volume, which mimic a real case scenario for the RSC network. The yellow triangles represent the seismic stations; the red dots represent the true seismic sources; the blue open circles represent the calculated locations. The red lines represent the boundary of the grid of test hypocentres for which traveltimes are calculated.

We can estimate H_{thr}, by applying to the daily sequence of entropy values a low-pass Butterworth filter with a corner frequency 40 per cent of the sampling frequency of the time sequence; this value was determined empirically considering the sampling period of the entropy sequence and its signal-to-noise ratio. This provides filtered time-series that represents the baseline, or threshold, for the earthquake detection. When the entropy value in the original time-series is larger than the value in the corresponding point of the baseline, we consider a positive detection in the signal time window for which that entropy values was obtained. By considering a time dependent baseline as threshold we account for noise fluctuations that affect the event detection throughout the day.

To give a first idea of the uncertainty associated to the location results, we calculate the covariance matrix from the F function; the eigenvalues of the covariance matrix can be used to define the semi-axes of a covariance ellipsoid. The 3-D distribution of the objective function has the potential to provide a formal framework for defining confidence regions, but we leave the development as subject for future research.

Fig. 5 shows the example of two F images. One corresponds to seismic traces that contain a seismic event while the second example corresponds to seismic traces without seismic events.

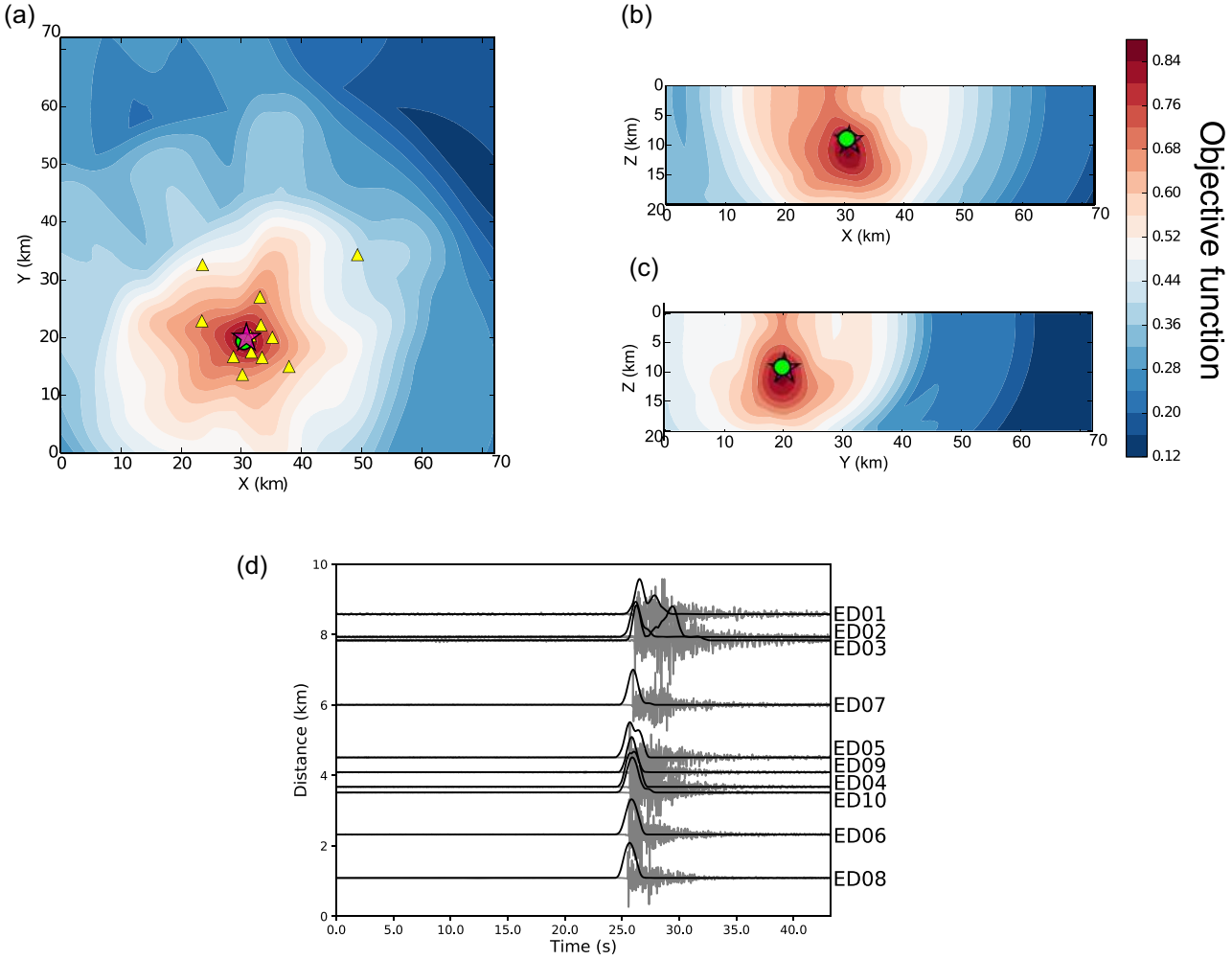

Figure 10. (a–c) F resulting from the location procedure of a microseismic event occurred on 2014 May 1 at 13:26:24; (d) vertical component traces representative of the event and corresponding characteristic functions. The yellow triangles represent the seismic stations used to locate the event; the green circle is the location of the event obtained by the automated procedure while the purple star is the location obtained by the manual procedure.

3 SYNTHETIC TESTS

To evaluate the performance of MigraLoc, we apply it to a synthetic data set of seismograms for a given seismic network and a set of hypocentre locations. First, we tune the procedure on the ideal noise-free and exact velocity model case, then we investigate the effects of noise in the waveforms and the adoption of an incorrect velocity model.

We compute the seismograms with the wavenumber integration method (Herrmann 2013) for the horizontally layered medium listed in Table 1 (Model 1) and consider the receiver locations corresponding to 10 three-component seismic stations resembling the existing Collalto Seismic Network (Fig. 1), that is, the seismic monitoring infrastructure of the natural, underground gas storage located in Collalto, at the foothills of Southern Alps, in northeastern Italy (Priolo *et al.* 2015). We compute synthetic seismograms for a strike slip point source at depth = 1.0 km, with moment magnitude $M_w = 2.0$. The 3-D grid of possible source points has an extension of $25 \times 25 \times 25$ km^3, and 100 m grid spacing. As a first experiment, we use synthetic tests to tune the parameters of the signal enhancement part of our procedure, that is, the sliding window τ for the covariance matrix, the width of the Gaussian filters α, and the frequency pass-band Δf. Considering that the maximum possible delay in the arrival times at the stations is 12 s, the length of the investigated time windows was set to 15 s. We detect the event and retrieve exactly the location used to calculate the synthetic seismograms with $\tau = 0.1$ s, $\alpha = 50$ and $\Delta f = 1$–50 Hz; the resulting F and entropy values are listed in Table 2. The F reveals a clear global maximum, as shown in Fig. 6. In a second experiment, we contaminate the synthetic waveforms with white noise, using different noise levels (10, 30 and 70 per cent of the maximum waveform amplitude, respectively; Fig. 7) and with real noise as in Grigoli *et al.* (2018). Even with high noise levels, our approach retrieves both the location and the origin time used for the simulated event, with time shifts between 0.035 s and 0.060 s from the true origin time. The F images relative to these tests show a clear maximum for all the three noise levels (Figs 8a–c), with a reduction of the F maximum and an increase of the entropy at increasing noise levels. The test demonstrates the effectiveness of our procedure in identifying the events and retrieving the correct source location parameters with noise levels up to 70 per cent of the maximum waveform amplitude. On the other hand, as we use entropy values as a criterion to detect seismic events, it is likely that a larger level of noise in seismic signals, may degrade the process of event detection. We therefore tested MigraLoc with increasing

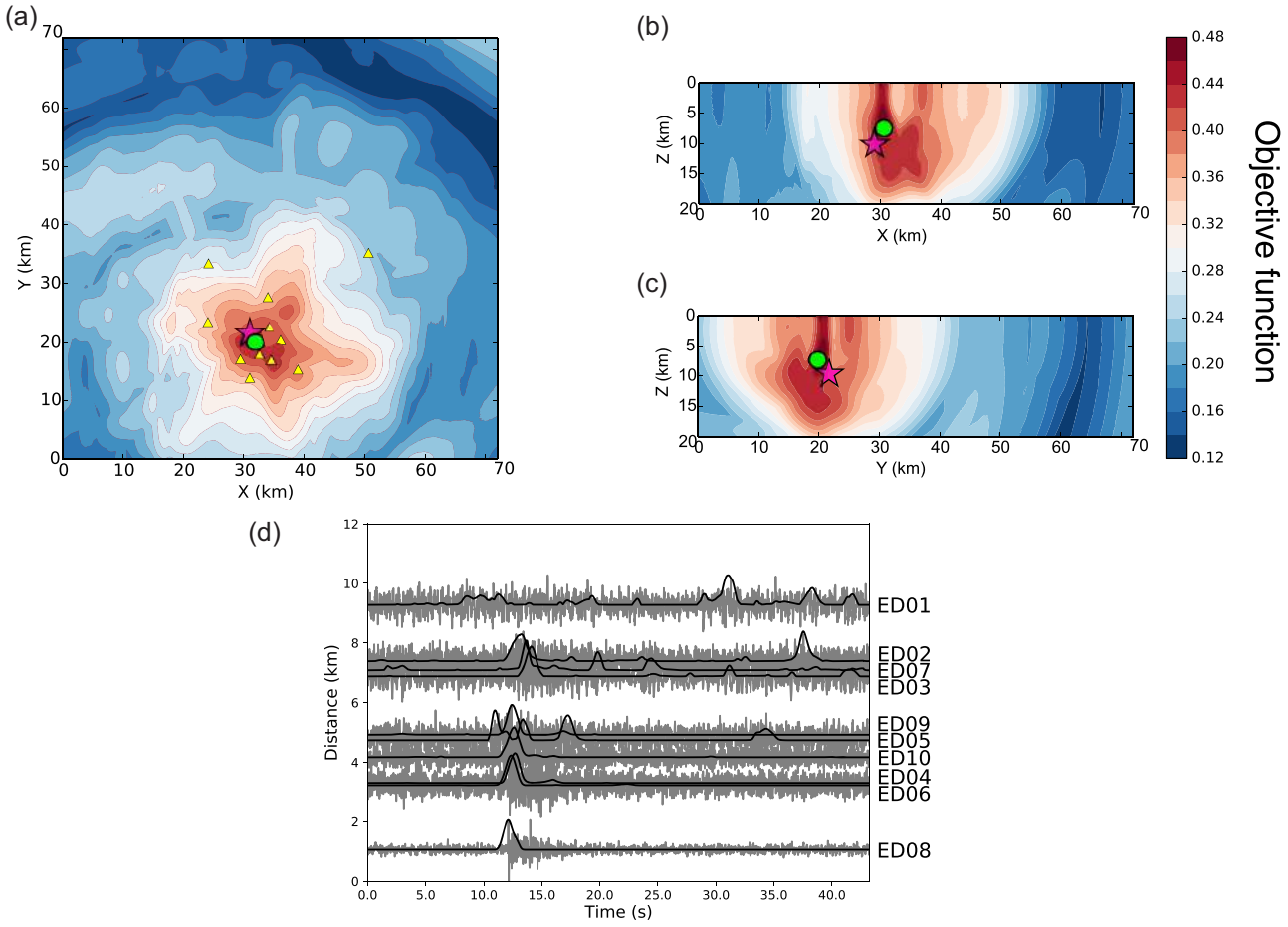

Figure 11. (a–c) F resulting from the location procedure of a microseismic event occurred on 2014 May 1 at 22:22:57; (d) vertical component traces representative of the event and corresponding characteristic functions. The yellow triangles represent the seismic stations used to locate the event; the green circle is the location of the event obtained by the automated procedure while the purple star is the location obtained by the manual procedure.

Table 4. Location results for two events of the 2014 May 1 swarm.

Origin time	Latitude automated	Longitude automated	Depth automated	Latitude manual	Longitude manual	Depth manual
13:26:24.13	45.876	12.196	10.2	45.880	12.196	9.6
22:22:57.23	45.882	12.192	4.2	45.888	12.186	10.0

noise level. The results show that with noise level corresponding to 90 per cent of the maximum amplitude the algorithm fails the detection; however, this is a case of very low SNR with the signal almost completely masked by seismic noise.

Crustal models used to locate earthquakes are described by a limited number of parameters and include some approximation with respect to the true Earth. It is therefore important to consider the possible presence of errors in velocity models and possibly estimate and test the sensitivity of the locations to those errors. We tested the methodology against the presence of errors in the crustal model using different velocity models for the computation of the synthetic waveforms (Model 2 in Table 3) and arrival times (Model 1 in Table 1). Model 2 is the result of a tomographic study for a larger area surrounding that of the Collalto underground gas storage and including the Montello hills (Anselmi *et al.* 2011). The results for hypocentral coordinates show errors up to 0.2 km in x, y and z, and the maximum of F is lower than the previous case. We also consider the case of a seismic event whose source lies outside of the volume of possible source points, used to calculate the theoretical arrival times. The results for hypocentral coordinates are clearly not correct, because the procedure can scan only the points in the predefined 3-D search volume, F is sensibly lower than in the previous synthetic tests, and the event is attributed to a source located within the 3-D search volume.

In the final experiment, we test possible location misfits due to network geometry by simulating a real case scenario seismic network with sources arranged in different locations with respect to it in the 3-D search volume. In particular, we use synthetic waveforms, with no noise added, corresponding to 15 sources located in different zones of the volume; the results are plotted in Fig. 9. While all sources internal to the 3-D search volume are correctly detected and located, it is evident that

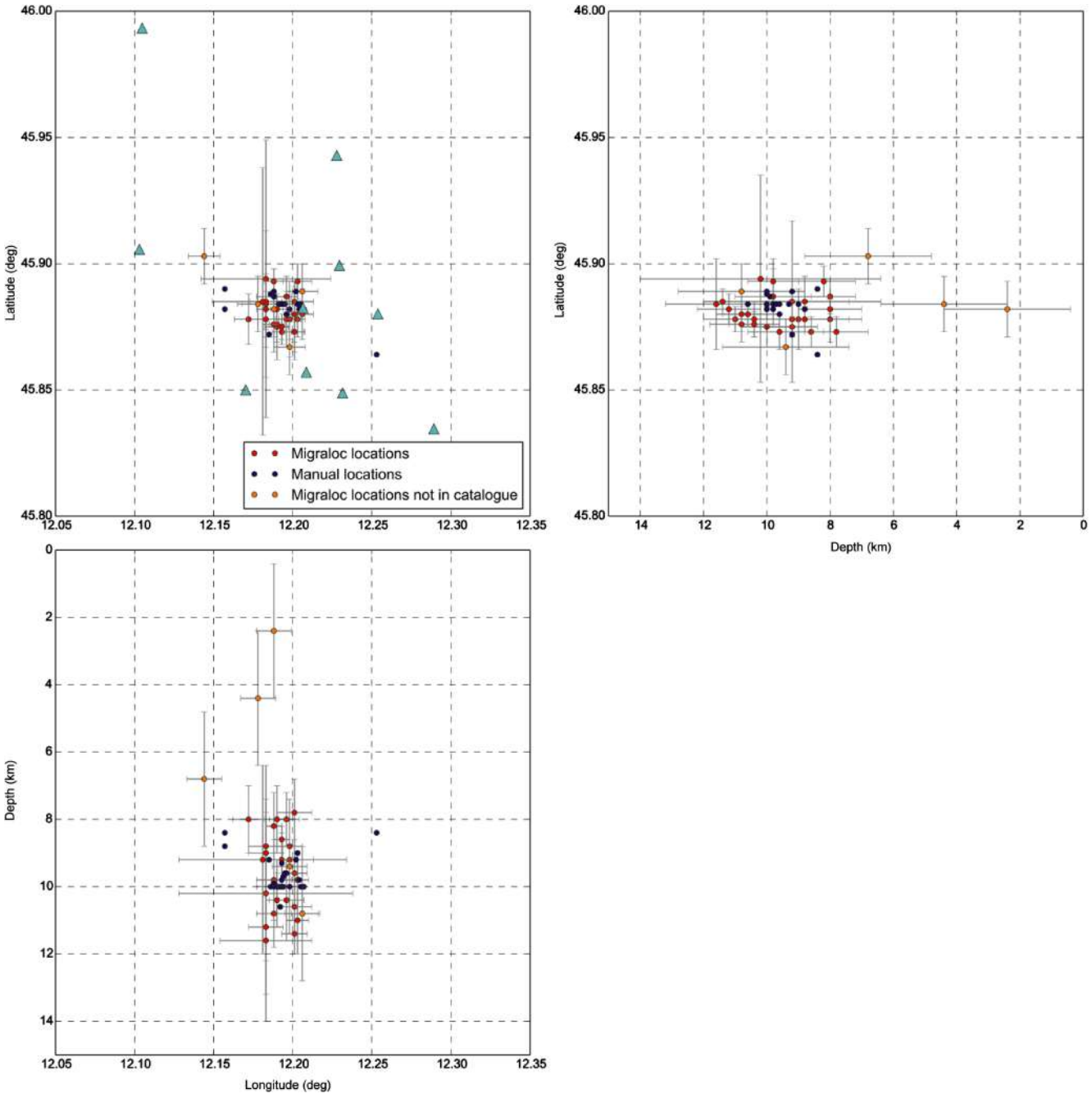

Figure 12. Location of the events of the 2014 May 1 seismic swarm. The figure compares the manual locations of the RSC catalogue (blue dots) to those obtained automatically by MigraLoc (red and orange dots), respectively. Locations are projected on the three Cartesian planes. The cyan triangles represent the stations. The red and orange dots represent MigraLoc locations for events present and not present in the RSC catalogue, respectively; error bars correspond to the confidence ellipsoids' major axes.

the location results deteriorate close to the boundaries of the volume.

The results of synthetic tests confirm the robustness of our approach: MigraLoc can retrieve the correct location in both cases of noise-free data and data contaminated with different noise levels, using a station configuration that mimics an existing seismic network. As it is expected with any location procedure, results deteriorate when the crustal model is not correct, or events occur at the periphery of the network.

4 APPLICATION TO REAL CASES

We test MigraLoc on two seismic swarms recorded by the Collalto Seismic Network (Priolo *et al.* 2015; Moratto *et al.* 2019), hereafter called RSC according to its Italian name (Fig. 1). The RSC is the infrastructure used to monitor the natural and induced seismicity of the natural gas storage concession known as Collalto Stoccaggio. The Collalto gas storage concession, held by Edison Stoccaggio S.p.A., is located at the foothills of the southeastern Alps, at the northern margin of the Venetian Plain. The reservoir is a geologic

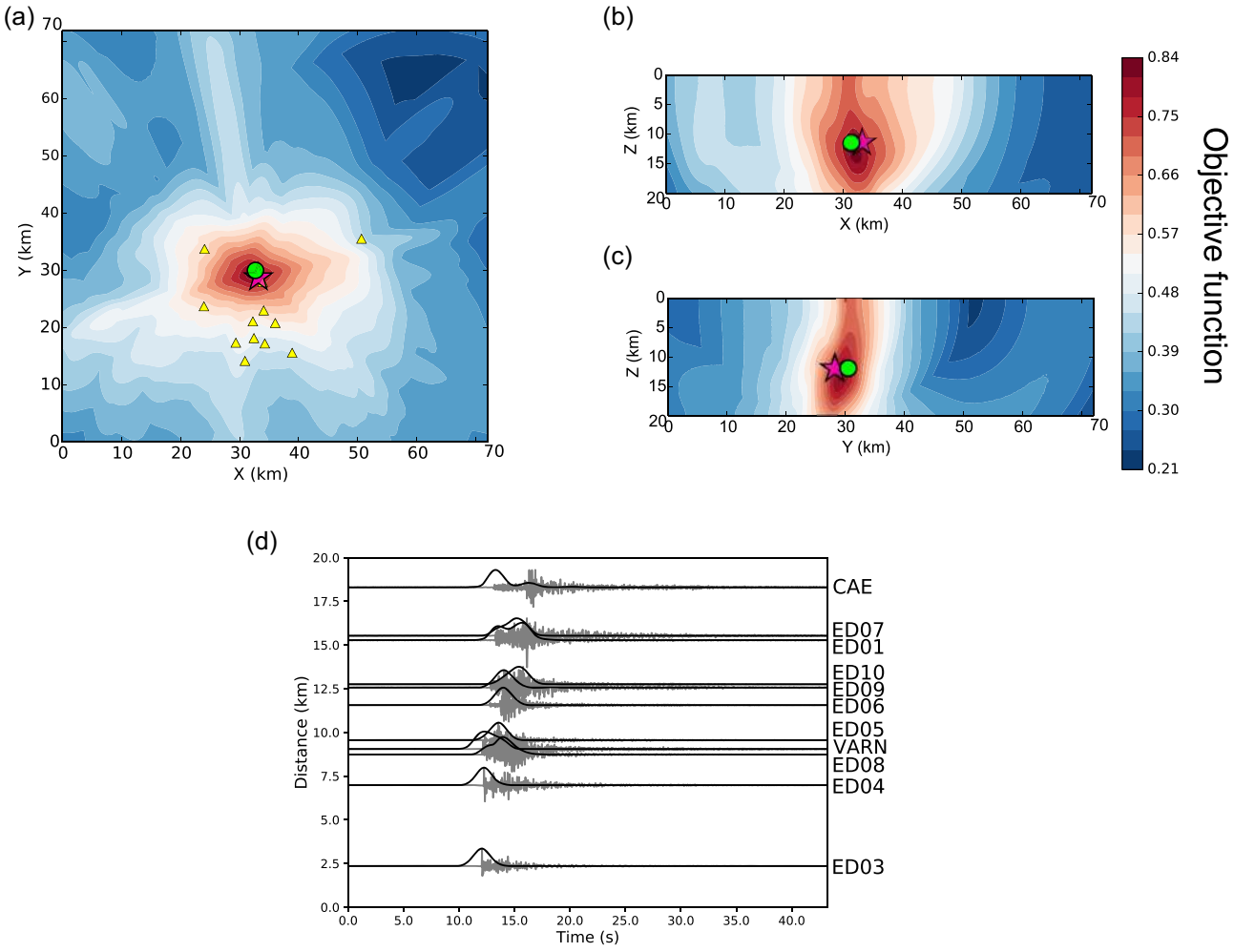

Figure 13. (a–c) F resulting from the location procedure of a microseismic event occurred on 2017 February 6 at 13:23:40; (d) vertical component traces representative of the event and corresponding characteristic functions. The yellow triangles represent the seismic stations used to locate the event; the green circle is the location of the event obtained by the automated procedure while the purple star is the location obtained by the manual procedure.

trap, with porous and permeable rock layers a few meters thick, sealed by impermeable formations and located at 1500–1600 m depth. It extends approximately over a 10×4 km^2 area (Fig. 1). The RSC was developed by the OGS—National Institute of Oceanography and Applied Geophysics following the recommendation of the Italian Ministry of the Environment, as Edison Stoccaggio S.p.A. was authorized to increase the storage pressure to the original confining pressure at which the gas was found in the reservoir. Since the region is characterized by medium–high seismic hazard, seismic monitoring is a key not only to identify the microseismicity possibly induced by gas storage operations, but also to understand whether and how this activity may interact with the tectonic structures surrounding the reservoir. The RSC started operating on 2012 January 1. It consists of 10 stations equipped with extended-band borehole seismometers ($T \geq 10$ s), and it is integrated by some other stations of the North-East Italy Seismic Network managed by the OGS (rts.inogs.it). Data are acquired with sampling frequency set at 200 Hz. The RSC has a fully automatic real-time detection and location system. All automatic locations are revised manually by the OGS seismologists and relocated by Hypo71 (Lee & Lahr 1975). The RSC catalogue reaches a completeness magnitude of 0.2 in the target area for the microseismicity (named area A in Priolo *et al.* 2015). Full information and data of the RSC (OGS 2012) are available at www.rete-collalto.crs.inogs.it.

In our test, we consider the two seismic swarms occurred on May 2014 and February 2017 in the area monitored by the RSC, and compare the results obtained by MigraLoc to those contained in the RSC catalogue. For the purpose of convenience, we decided to use waveform data recorded by the 10 RSC stations and 2 stations of the North-East Italy Seismic Network (OGS 2016). We applied MigraLoc to one-day long continuous seismic traces. Each detection was visually confirmed and compared to the results obtained manually. Before processing the continuous seismic data for earthquake detection and location, we apply a basic pre-processing by removing the mean and band-pass filtering (1–25 Hz) in order to remove the long period noise that is usually present in the data. Similar operation is performed routinely by the real-time automatic processing. The continuous waveforms are partitioned into overlapping time windows 30 s long, with overlapping time of 10 s; those values are estimated from the calculated arrival times, as described in paragraph 2.

The 2014 May 1 swarm features 25 events, located in the central part of the RSC and with local magnitudes M_L in the range -0.7 to 1.4. As a first step, we tune MigraLoc parameters (τ and α, Δf, and threshold) on the continuous waveform data set. We test five frequency ranges (1–40, 5–12, 15–25, 25–40, 1–25 and 5–25 Hz), three τ values (0.5, 1.0 and 3.0 s), four α values (5, 50, 100 and 1000) and four threshold levels (95, 90, 75 and 50 per cent). The

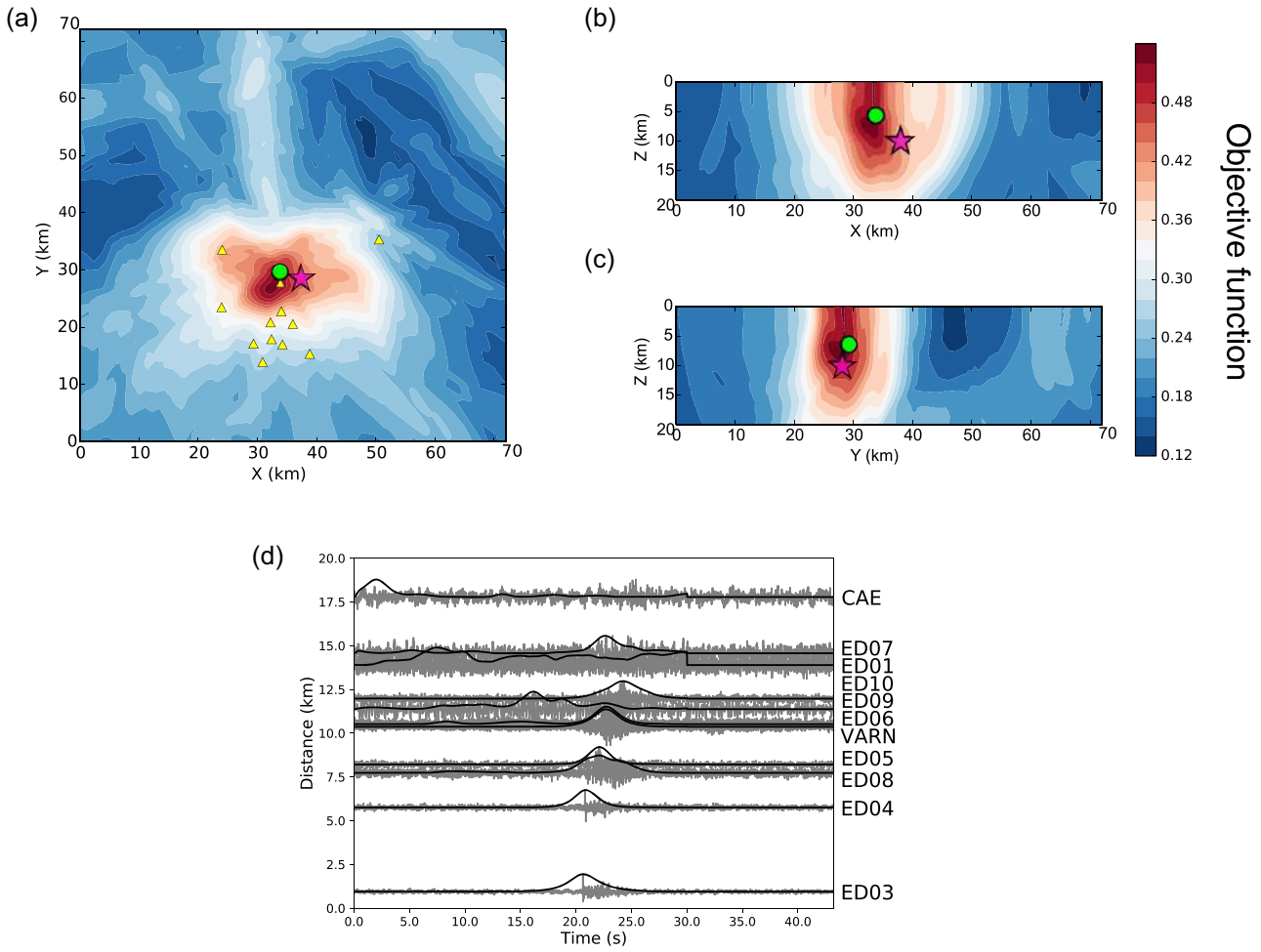

Figure 14. (a–c) *F* resulting from the location procedure of a microseismic event occurred on 2017 February 7 at 22:45:50; (d) vertical component traces representative of the event and corresponding characteristic functions. The yellow triangles represent the seismic stations used to locate the event; the green circle is the location of the event obtained by the automated procedure while the purple star is the location obtained by the manual procedure.

Table 5. Location results for two events of the 2017 February 6 and 7 swarm.

Date	Origin time	Latitude automated	Longitude automated	Depth automated	Latitude manual	Longitude manual	Depth manual
6.2.2017	13:23:40.18	45.963	12.214	11.3	45.944	12.218	11.3
7.2.2017	22:22:57.23	45.951	12.224	4.9	45.946	12.271	10.3

grid of possible source points used for direct search has an extension of 36 × 36 × 25 km^3 and grid spacing of 200 m; *P* and *S* traveltimes were computed using the same velocity model used for the manual location of the RSC catalogue (Model 1 in Table 1). As a result of the tuning, the parameters that allow the detection of the largest number of microseismic events are: 1–25 Hz for frequency range, 0.5 s for τ, 1000 for α, and 90 per cent for threshold level. Considering that the maximum possible delay in the arrival times at the stations is 27 s, the length of the investigated time windows was set to 30 s. The procedure then computes the objective function F for the 24 h continuous waveforms and, for each F, the corresponding entropy H at 30 s time intervals. The sequence of entropy values is filtered according to paragraph 2.3 to obtain a time-series that represents the threshold for event detection.

For the 2014 May 1 swarm, the methodology recovered all the 25 events present in the RSC catalogue plus 5 more events and 5 false positives. As an example of the results obtained for a weak earthquake well detected by the whole network, Fig. 10 shows the waveforms and the *F* images calculated for the event of 2014 May 1 (13:26:24), with magnitude M_L 1.4. Fig. 11 shows instead the results for the smallest microearthquake located by the RSC, that is, the event 2014 May 1 (22:22:57) with magnitude M_L −0.7. The hypocentre parameters are reported in Table 4. In both cases the *F* shows a clear maximum, its value being lower for the 22:22 microearthquake, which also features a large smearing, especially in depth. The hypocentral locations retrieved by MigraLoc are close to those obtained manually for the entire data set, the largest difference being for the depth of the 22:22 microearthquake owing to the lower SNR of the signals.

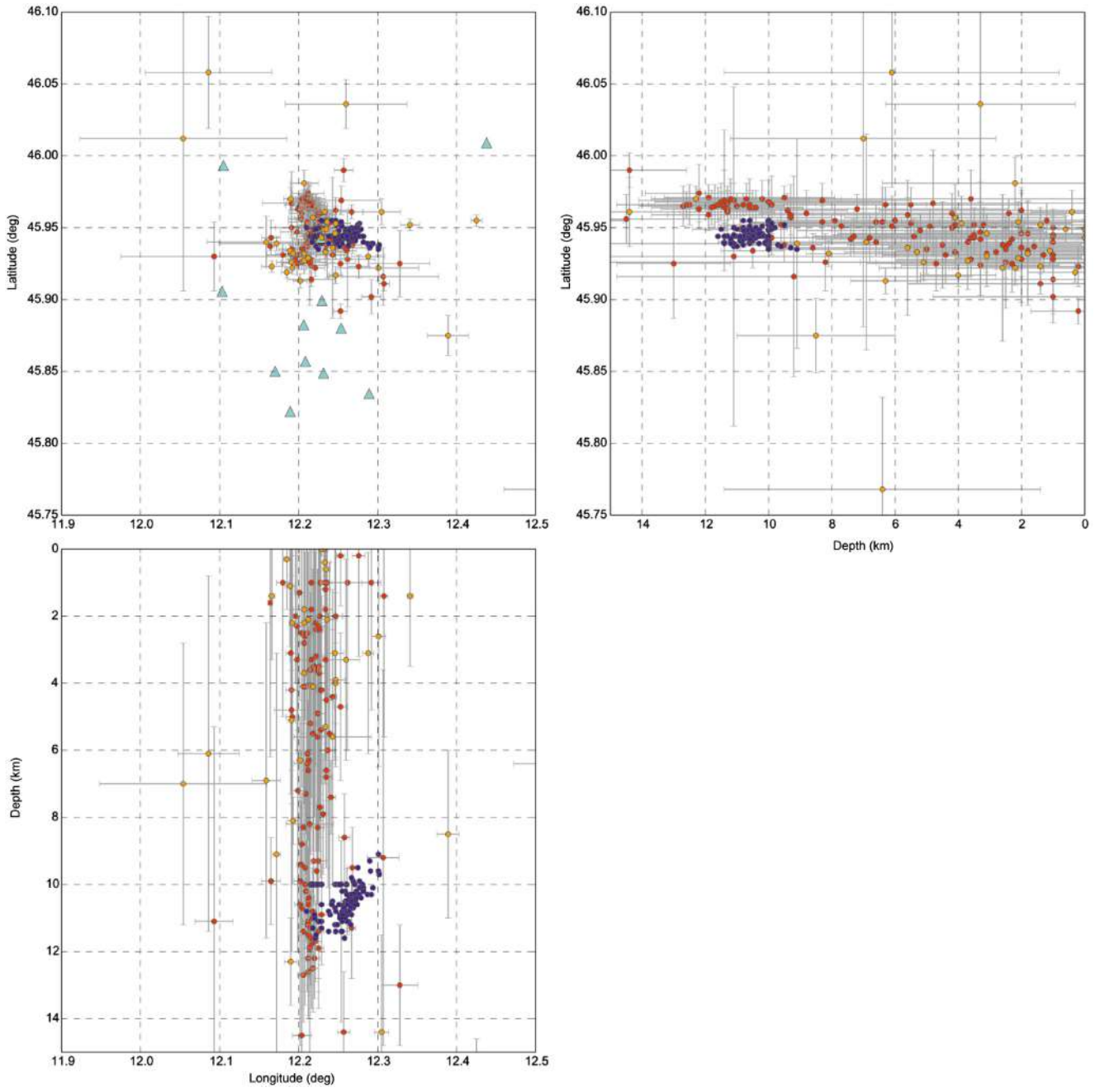

Figure 15. Location of all events of the 2017 February 6 and 7 seismic swarm. Other details as in Fig. 12.

Fig. 12 shows the location of all events obtained by MigraLoc and manually, respectively. They are very similar to each other, with horizontal distances < 1 km for most cases. The comparison of the hypocentral parameters obtained by manual and automated locations is given in Fig. S1 (Supporting Information). Especially for events with very low magnitude ($M_L < 0.0$) manual locations were obtained using a small number of seismic phases (only four in some cases) while with MigraLoc we use the full set of stations and exploit all the possible information contained in the three component seismic signals.

In general, location methods based on waveform stacking are designed to work with many stations well distributed around the seismic source area (e.g. Grigoli *et al.* 2013b). With a view to the application of MigraLoc to real-time data analysis, we test its performance in the worst-case scenario, with seismic events that occur at the boundary of the RSC target area, that is in the case of bad azimuth coverage of the stations, in particular for what concerns depth estimation.

The second test with real data was then performed on the seismic swarm occurred on 2017 February 6 and 7 near the town of Vittorio Veneto, about 12–15 km northeast of the RSC, and with magnitudes M_L in the range −0.3 to 2.5. The location was performed using the same input parameters used for the May 2014 case. We decided to apply the MigraLoc code to one-day long seismic traces, one for each day of seismic recordings. The RSC catalogue reports 122 events for this swarm. MigraLoc identifies all the 122 mi-

croearthquakes reported in the catalogue and 35 unreported events. The detection process is as described for the May 2014 swarm, the only difference being that the baseline was estimated for each day separately.

Figs 13 and 14 show an example of the waveforms and F images for the events occurred on 2017 February 6 (13:23:40) with magnitude M_L 2.0 and on 2017 February 7 (22:45:50) with magnitude M_L −0.1, respectively. The hypocentre parameters of the two events are reported in Table 5. The differences between MigraLoc and catalogue locations are larger in this case than those obtained for the May 2014 swarm. While there still is a good control on epicentral coordinates and depth for the M_L 2.0 earthquake, the automatic location is about 3 km away from the manual one and the depth estimation is loose for the M_L −0.1 event. The reason can be understood from Fig. 14(d): only the three closest stations could record the event while the signal is hardly recognizable in the other traces (e.g. ED07 in Fig. 14d). The comparison of the hypocentral parameters obtained by manual and automated locations is given in Fig. S2 (Supporting Information). Fig. 15 represents the map of the locations of February 2017 swarm. The map shows that the automated locations are characterized by quite large errors bars in this case, especially in depth. Nevertheless, the results are encouraging: MigraLoc has the potential to detect and locate, at least with a certain approximation, more events than classical manual methodologies. Even if the seismic swarm occurred at the boundary of the RSC area, we could detect more events than the standard procedures and obtain a first overview of the swarm location.

5 CONCLUSIONS

We have presented MigraLoc, a new procedure for automated detection and location of microseismic events in near real-time using waveforms recorded by local seismic networks. Our procedure is based on space–time migration of characteristic functions computed by a TF polarization analysis, and it is designed to process continuous traces recorded by a dense seismic network on moderate number of computer cores.

We tested the procedure on both synthetic seismograms and real data. The results obtained by the synthetics show that MigraLoc is reliable and allows the detection and location of events also with noisy traces, since it exploits all the information contained in the three component seismic signals. The application to the real data recorded by the RSC reveals that our procedure detects and localizes with a good accuracy not only all microseismic events recognized by the standard manual location implemented in the RSC (several of which with negative magnitudes), but it finds more events. As expected, the location accuracy decreases for very weak events occurring at the periphery of the network, especially for depth. In general, we proved that MigraLoc is sensitive and robust in detecting and locating events, and especially microevents.

It is important to note that in its present version MigraLoc uses a time window with fixed length, so it is possible that it fails in the detection of multiple events contained in the same time window. The success of the detection process is also a function of the entropy threshold chosen. An elevated detection threshold may compromise detection sensitivity while a too low threshold may introduce many false positives events.

The proposed methodology is not limited to microseismic events, but it can be applied in principle to earthquakes in other ranges of magnitude and distance as well. Regional networks will require a larger grid for hypocentre search, while large magnitude events will require longer time windows and appropriate TF analysis parameters. The application of the methodology to different-scale earthquakes must take into account that, in general, waveform-based algorithms work more efficiently when used with dense networks, with stations well distributed around the seismic source area (Grigoli *et al.* 2013b, 2017).

6 DATA AND RESOURCES

Information and data on the RSC can be found at http://rete-collalto.crs.inogs.it (last accessed October 2019). The full data set of continuous waveforms of the RSC is freely available at the National Institute of Oceanography and Applied Geophysics (OGS) Archive System of Instrumental Seismology (OASIS; http://oasis.crs.inogs.it/, last accessed February 201), which is the database that archives the instrumental seismological data of the OGS.

ACKNOWLEDGEMENTS

This work was supported by the program ‘HPC Training and Research for Earth Sciences’ (HPC-TRES) of the National Institute of Oceanography and Applied Geophysics (OGS). We acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources and support. We also acknowledge the support provided by the Italian Ministry for the Economic Development (MiSE), General Directorate for the Mining Resources (DGRME, http://unmig.sviluppoeconomico.gov.it/), within the Program Agreement with OGS (Record Number 3835 of the Corte dei Conti) under Activity 1, entitled ‘Development of analysis methods for a rapid correlation of the detected seismicity to the underground exploitation for energy production’. The Collalto Seismic Network (RSC) was developed and is managed by the National Institute of Oceanography and Applied Geophysics (OGS) on behalf of Edison Stoccaggio S.p.A. (i.e. Italian Societa` per Azioni) under the requirements of the Italian Ministry for the Environment and Land and Sea Protection and in agreement with the Veneto Region. We thank the Editor, S. Ni, F. Grigoli and one anonymous reviewer for useful comments that improved the original manuscript.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification, *IEEE Trans. Autom. Control,* **19**(6), 716–723.

Allen, R., 1982. Automatic phase pickers: their present use and future prospects, *Bull. seism. Soc. Am.,* **72,** S225–S242.

Alvarez, I., Garcia, L., Mota, S., Cortes, G., Benitez, C. & De la Torre, A., 2013. An automatic *P*-phase picking algorithm based on adaptive multiband processing, *IEEE Geosci. Remote Sens. Lett.,* **10,** 1488–1492.

Anselmi, M., Govoni, A., De Gori, P. & Chiarabba, C., 2011. Seismicity and velocity structures along the south-Alpine thrust front of the Venetian Alps (NE-Italy), *Tectonophysics,* **513,** 37–48.

Baer, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.,* **77,** 1437–1445.

Baillard, C., Crawford, W.C., Ballu, V., Hilbert, C. & Mangeney, A., 2014. An automatic kurtosis-based *P*- and *S*-phase picker designed for local seismic networks, *Bull. seism. Soc. Am.,* **104,** 394–409.

Baker, T., Granat, R. & Clayton, R.W., 2005. Real-time earthquake location using Kirchhoff reconstruction, *Bull. seism. Soc. Am.,* **95,** 699–707.

Cesca, S. & Grigoli, F., 2015. Chapter two—full waveform seismological advances for microseismic monitoring, *Adv. Geophys.,* **56,** 169–228.

Chen, H., Chiu, J., Pujol, J., Kim, K., Chen, K., Huang, B., Yeh, Y. & Chiu, S., 2006. A simple algorithm for earthquake location using 3D V_P and V_S

models: test examples in the Central United States and in Central Eastern Taiwan, *Bull. seism. Soc. Am.,* **96,** 288–305.

Davenport, K.K., Hole, J.A., Quiros, D.A., Brown, L.D., Chapman, M.C., Han, L. & Mooney, W.D., 2015. Aftershock imaging using a dense seismometer array (AIDA) after the 2011 Mineral, Virginia, earthquake, in *The 2011 Mineral, Virginia, Earthquake, and Its Significance for Seismic Hazards in Eastern North America: Geological Society of America Special Paper 509,* eds Horton, J.W., Jr., Chapman, M.C. & Green, R.A., doi:10.1130/2015.2509(15).

Flinn, E.A., 1965. Signal analysis using rectilinearity and direction of particle motion, *Proc. IEEE,* **56,** 1874–1876.

Gharti, H.N., Oye, V., Roth, M. & Kühn, D., 2010. Automated microearthquake location using envelope stacking and robust global optimization, *Geophysics,* **75**(4), MA27–MA46.

Grigoli, F., Cesca, S., Amoroso, O., Emolo, A., Zollo, A. & Dahm, T., 2013b. Automated seismic event location by waveform coherence analysis, *Geophys. J. Int.,* **196,** 1742–1753.

Grigoli, F., Cesca, S., Krieger, L., Kriegerowski, M., Gammaldi, S., Horalek, J., Priolo, E. & Dahm, T., 2016. Automated microseismic event location using master-event waveforms stacking, *Sci. Rep.,* **6,** 25744, doi:10.1038/srep25744.

Grigoli, F., Cesca, S., Vassallo, M. & Dahm, T., 2013a. Automated seismic event location by travel-time stacking: an application to mining induced seismicity, *Seismol. Res. Lett.,* **84**(4), 666–677.

Grigoli, F., Scarabello, L., Böse, M., Weber, B., Wiemer, S. & Clinton, J.F., 2018. Pick- and waveform-based techniques for real-time detection of induced seismicity, *Geophys. J. Int.,* **213,** 868–884.

Grigoli, F. *et al.*, 2017. Current challenges in monitoring, discrimination, and management of induced industrial activities: a European Perspective, *Rev. Geophys.,* **55,** 310–340.

Herrmann, R.B., 2013. Computer programs in seismology: an evolving tool for instruction and research, *Seism. Res. Lettr.,* **84,** 1081–1088.

Jurkevics, A., 1988. Polarization analysis of three-component array data, *Bull. seism. Soc. Am.,* **78,** 1725–1743.

Kao, H. & Shan, S.J., 2004. The source-scanning algorithm: mapping the distribution of seismic sources in time and space, *Geophys. J. Int.,* **157,** 589–594.

Kao, H. & Shan, S.J., 2007. Rapid identification of earthquake rupture plane using source-scanning algorithm, *Geophys. J. Int.,* **168,** 1011–1020.

Ku¨perkoch, L., Meier, T. & Friederich, W., and EGELADOS working group, 2010. Automated *P*-wave arrival time determination using higher order statistics. *Geophys. J. Int.,* **181**(2), 1159–1170.

Langet, N., Maggi, A., Michelini, A. & Brenguier, F., 2014. Continuous kurtosis-based migration for seismic event detection and location, with application to Piton de la Fournaise Volcano, La Réunion, *Bull. seism. Soc. Am.,* **104,** 229–246.

Lee, W.H.K. & Lahr, J.C., 1975. Hypo71 (revised): a computer program for determining hypocenter, magnitude and first motion pattern of local earthquakes, U.S. Geol. Surv. Open-File Rep. 75–311, 113 pp.

Leonard, M. & Kennett, B.L.N., 1999. Multi-component autoregressive techniques for the analysis of seismograms, *Phys. Earth planet Inter.,* **113,** 247–263.

Lomax, A., Virieux, P & Berge, C., 2000. Probabilistic earthquake location in 3D layered models: introduction of a Metropolis–Gibbs method and comparison with linear locations, in *Advances in Seismic Event Location,* pp. 101–134, Kluver.

Message Passing Interface ForumMessage Passing Interface Forum, 2015. MPI: a message-passing interface standard, version 3.1.

Montalbetti, J.F. & Kanasewich, E.R., 1970. Enhancement of teleseismic body phases with a polarization filter, *Geophys. J. R. astr. Soc.,* **21,** 119–129.

Moratto, L., Romano, M.A., Laurenzano, G., Colombelli, S., Priolo, E., Zollo, A. & Picozzi, M., 2019. Source parameters analysis of microearthquakes recorded around the underground gas storage in the Montello-Collalto Area (Southeastern Alps, Italy), *Tectonophysics,* **762,** 159–168.

OGS, 2012 . National Institute of Oceanography and Applied Geophysics Collalto Seismic Network - Rete Sismica di Collalto. International Federation of Digital Seismograph Networks. Dataset/Seismic Network,

OGS, 2016. National Institute of Oceanography and Applied Geophysics - OGS: North-East Italy Seismic Network. International Federation of Digital Seismograph Networks. Dataset/Seismic Network.,

Podvin, P. & Lecomte, L., 1991. Finite difference computations of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int.,* **105,** 271–284.

Poiata, N., Satriano, C., Villotte, J.-P., Bernard, P & Okabara, K., 2016. Multi-band array detection and location of seismic sources recorded by dense seismic networks, *Geophys. J. Int.,* **205**(3), 1548–1573.

Priolo, E. *et al.*, 2015. Seismic monitoring of an underground natural gas storage facility: the Collalto Seismic Network, *Seismol. Res. Lett.,* **86**(1), 109–123.

Rosenberger, A., 2010. Real-time ground-motion analysis: distinguishing *P* and *S* arrivals in a noisy environment, *Bull. seism. Soc. Am.,* **100,** 1252–1262.

Ross, Z.E. & Ben-Zion, Y., 2014. Automatic picking of direct *P*, *S* seismic phases and fault zone head waves, *Geophys. J. Int.,* **199,** 368–381.

Ross, Z.E., White, M.C., Vernon, F.L. & Ben-Zion, Y., 2016. An improved algorithm for real-time *S*-wave picking with application to the (augmented) ANZA network in Southern California. *Bull. seism. Soc. Am.,* **106,** 2013–2022.

Saragiotis, C.D., Hadjileontiadis, L.J. & Panas, S.M., 2002. PAI-S/K: a robust automatic seismic *P* phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.,* **40,** 1395–1404.

Shannon, C.E., 1948. A mathematical theory of communication, *Bell Syst. Tech. J.,* **27,** 379–423, and 623–656.

Sleeman, R. & van Eck, T., 1999. Robust automatic P-phase picking: an online implementation in the analysis of broadband seismogram recordings, *Phys. Earth planet. Inter.,* **113,** 265–275.

Vidale, J.E, 1986. Complex polarization analysis of particle motion, *Bull. seism. Soc. Am.,* **76,** 1393–1405.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Histograms showing the comparison of the hypocentral parameters obtained by Migraloc and the manual location for the seismic swarm of 2014 May 1.

Figure S2. Histograms showing the comparison of the hypocentral parameters obtained by Migraloc and the manual location for the seismic swarm of 2017 February 6 and 7.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.