



Quantifying location uncertainties in seismicity catalogues: application to the Pyrenees

Antoine L. Turquet  · Thomas Bodin ·
Pierre Arroucau · Matthieu Sylvander ·
Kevin Manchuel

Received: 4 January 2019 / Accepted: 25 July 2019
© Springer Nature B.V. 2019

Abstract Linearised least-square inversions are commonly used to locate small-magnitude earthquakes, as they are fast and simple to implement. These methods are based on minimising the root-mean-square (RMS) of travel time residuals to find the best-fitting location coordinates and origin time. There are two well-known problems that affect location estimates: (1) the linearisation of the inverse problem causes dependence on the initial guess; (2) regularisation produces solutions that depend on the chosen damping coefficient and biased uncertainty estimates. In this work, we propose a method to quantify unbiased

uncertainties with a series of synthetic tests. We first generate travel times for events from all possible coordinates on a 3D grid and then locate each synthetic event by using HYPOCENTER software (this can be applied to any location method). We show that the uncertainties estimated from the standard linearised inversion are strongly underestimated, and we propose another method to compute uncertainties. We produce a 3D error map, where at each grid point we plot the location error, defined as the distance between the event at the given grid point and its inverted location. Moreover, we show how this error map varies with the quantity and quality of the data, and with user-defined parameters such as maximum event–station distance or station corrections. We also provide a methodology to tune the seismic location parameters and calculate the corresponding uncertainties for users who are using similar earthquake location software. Finally, we present an application to the Pyrenean region.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10950-019-09857-8>) contains supplementary material, which is available to authorized users.

A. L. Turquet (✉) · T. Bodin
UCBL, CNRS, LGLTPE, Université de Lyon,
69622 Villeurbanne, France
e-mail: turquetal@gmail.com

P. Arroucau · K. Manchuel
EDF-Direction Industrielle, Aix-en-Provence, France

M. Sylvander
Observatoire Midi Pyrénées, IRAP, CNRS UMR 5277,
Universit Paul Sabatier, Toulouse, France

Present Address:
A. L. Turquet
SFF PoreLab, Njord Centre, Department of Physics,
University of Oslo, Oslo, Norway

Keywords Earthquake location · Seismicity catalog · Uncertainty quantification · Velocity model

1 Introduction

Location of seismic sources is necessary for many different geological and geophysical studies (Lee et al. 2014). These studies require information such as location, origin time and magnitude of past earthquakes, which are available in earthquake catalogues produced

by seismological observatories. Even though there are a number of location algorithms that are available today (see Lomax et al. 2000 for a review), linearised iterative methods (LIM) are still being widely used for constructing seismicity catalogues. Several studies have been conducted on the comparison of different earthquake location methods (e.g. Thurber 1986; Lomax et al. 2009; Richards et al. 2006; Turkaya et al. 2016; Turquet et al. 2019) but an unbiased quantification of error in source location estimations done using LIMs remains challenging (Pavlis 1986, 1992). It is clear that linearised iterative methods require better uncertainty quantification since they are widely used by different communities for different purposes. For example, earthquake catalogues are used as an input for seismic hazard assessment. A poor seismicity catalogue lacking uncertainty estimates may lead to an underestimated hazard assessment, and may have a direct societal impact (Abrahamson and Bommer 2005). Thus, in this work, we propose a simple method to quantify uncertainties in earthquake source estimations obtained from LIMs (Lee and Lahr 1972; Lahr et al. 1994, 1999; Kissling et al. 1995; Lienert et al. 1986). We focus our work on the Pyrenean region. The Pyrenees is a slowly deforming mountain range having continuous low magnitude seismicity, which makes it a good choice for our study. The uncertainty in earthquake source locations—estimated via LIMs—is present due to three main factors as given in Pavlis (1986): (1) seismic measurement errors, (2) errors in the modelled arrival times and (3) non-linearity of the problem. First of all, seismic measurement errors depend on the level of signal to noise ratio of the observed waveforms and on the method used to pick arrival times. Secondly, the errors in modelled arrival times depend on the theory and velocity model used for computation. It is well known that the errors in the model can lead to systematic biases in the estimations (Jordan and Sverdrup 1981; Thurber 1992; Billings et al. 1994). Pavlis (1986) has shown that the conventional methods of obtaining error ellipses from LIMs do not include the systematic errors in the velocity model. The inverse problem of earthquake location can be defined as:

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) \tag{1}$$

where \mathbf{d} is the data available to solve the problem, and \mathbf{g} is the non-linear function linking the model \mathbf{m}

and the data. The earthquake location problem is non-linear and thus, to solve this equation, we need at first to linearise it. This is done around a reference model m_i in Eq. 1 and we obtain:

$$\begin{aligned} \Delta \mathbf{d} &= G \Delta \mathbf{m} \\ \mathbf{d}_{obs} - \mathbf{d}_i &= G(\mathbf{m}_{i+1} - \mathbf{m}_i) \end{aligned} \tag{2}$$

The linearisation followed by an iterative procedure allows us to solve Eq. 2. The solution may vary based on the initial guess m_0 which is required for the least squares solution. After linearising the equation, the solution can be obtained with a generalised inverse as given in Tarantola and Valette (1982):

$$\Delta \mathbf{m} = (G^T G)^{-1} G^T \Delta \mathbf{d}. \tag{3}$$

$(G^T G)$ can be a singular (i.e. non-invertible) matrix. A common way to regularise the inverse of this matrix is to introduce damping into the system as in Tarantola and Valette (1982):

$$\Delta \mathbf{m} = (G^T G + \lambda \mathbf{I})^{-1} G^T \Delta \mathbf{d} \tag{4}$$

where λ is the level of damping applied to the system and \mathbf{I} is the identity matrix. Even though the regularisation enables us to find a stable solution, it biases the estimation of the error for the solution (Tarantola and Valette 1982). To estimate the error for the solution, first we define a $G^* = (G^T G)^{-1} G^T$ to calculate covariance matrix C which represents uncertainties in the model. Then, C can be written as:

$$C = G^{*T} C_d G^* \tag{5}$$

where C_d is the covariance matrix for data errors, which is set to 0.1 s in this study. Finally, we can express the estimation error E as:

$$\begin{aligned} E_h &= \sqrt{C_{1,1}^2 + C_{2,2}^2} \\ E_z &= C_{3,3} \\ E &= \sqrt{E_h^2 + E_z^2} \end{aligned} \tag{6}$$

where E_h and E_z are the estimation errors in horizontal and vertical directions respectively. In this work, the likelihood of the event falling into the area defined by E_h and E_z is taken as 68%.

The usual way in the LIM is to use the covariance matrix C to determine location errors (Peters and Crosson 1972; De Natale et al. 1984; Tramelli et al. 2013) as given in Eqs. 5 and 6. It should be noted that C can only be estimated in the case where $G^T G$ can be inverted to compute G^* . Otherwise, a damping parameter in G^* will stabilise the inverse but will bias the

error estimate (an increase in the damping parameter will artificially decrease the size of the error ellipse).

In Section 2 of this article, we explain our methodology. We describe our synthetic study, the algorithm used for source location estimation and the proposed scheme to compute maps of location errors. In Section 3, we present results of this synthetic study for the Pyrenees. This study allows us to quantify the pre-existing problem of underestimation of the uncertainties in LIMs. Then, we explore the effect of different parameters. For example, we quantify the contribution of errors in the velocity models (3D variations) in estimation errors. We explore and discuss the effect of *S* arrivals in the input data, the importance of the spatial coverage, the maximum event station distance, subregional velocity models, station corrections and the effect of the number of phases used in the estimation.

2 Method

In this study, we compute the distance between a given source location and the estimated location obtained from synthetic travel times produced for this given source. We compute this error for sources located over a fine 3D grid and compare it with the estimated uncertainties given by Eq. 6. Event location is performed by means of HYPOCENTER, a commonly used open source algorithm (Lienert and Havskov 1995), but the error quantification procedure proposed here can be applied to any location technique (e.g. HYPO71 (Lee and Lahr 1972), VELEST (Kissling et al. 1995), HYPODD (Waldhauser 2001)).

2.1 Constraints based on OMP methodology

There are 24,500 event locations available from the OMP (Observatoire Midi-Pyrenees) catalogues. Total number of P phases (P_g , P_n) recorded is 235,261 and S phases is 235,172 for these 24,500 events. This gives an average 10 P and S arrivals for each event in the catalogue. The mean uncertainties estimated for this catalogue are $E_h = 1.3$ km and $E_z = 2.7$ km (Theunissen et al. 2017). These uncertainties are computed using Eq. 6.

The results from linearised iterative methods depend on the initial location. The standard OMP procedure is to choose the coordinates of the station

having the smallest travel time as the initial location. Next, the inversion is carried out 20 times with 20 different initial depths, eventually obtaining 20 estimated locations for an event. The spatial mean of the 20 locations is computed and chosen as the final earthquake location.

The contribution of different observations to the solution can be weighted. In OMP for earthquake location, they use a station weighting based on the event–station distance, which helps limit 3D effects that are accumulating as we get farther from the source. Travel time picks for stations farther than 30 km have a linearly decreasing weighting until 120 km where the weights diminish to 0. The stations farther than 120 km have no effect on epicentre location procedure.

2.2 Algorithm HYPOCENTER

HYPOCENTER is an algorithm to locate the seismic events as seen in Fig. 1. First, the software reads the initial parameters (e.g. number of maximum iterations, damping, maximum event distance from stations), station coordinates and the 1D velocity model provided with the input file. Second, the software reads the data related to the particular event. After a quick consistency check on the arrival time set to find obvious bad arrivals (miswritten phases, large differences between adjacent stations, etc.), it computes the modelled arrival times. We use a 1D horizontally layered velocity model to calculate the arrival times. This type of velocity model is suitable for local studies as we have in the Pyrenean Region. Finally, after a number of iterations, we obtain the output solution corresponding to the minimum misfit. Next, the software continues with calculating the uncertainty as described in Lienert and Havskov (1995). To choose the optimal damping—which minimises the travel time residuals—the software proposes an “Adaptive Damped Least Squares Method” described in Lienert et al. (1986).

2.3 Map of mean error

We first build a simple system consisting of a homogeneous half-space velocity model with a station network made of 9 regularly distributed stations. As seen in Fig. 2, 9 stations are placed symmetrically. The distance from one side of the array to the other is 1° (around 111 km).

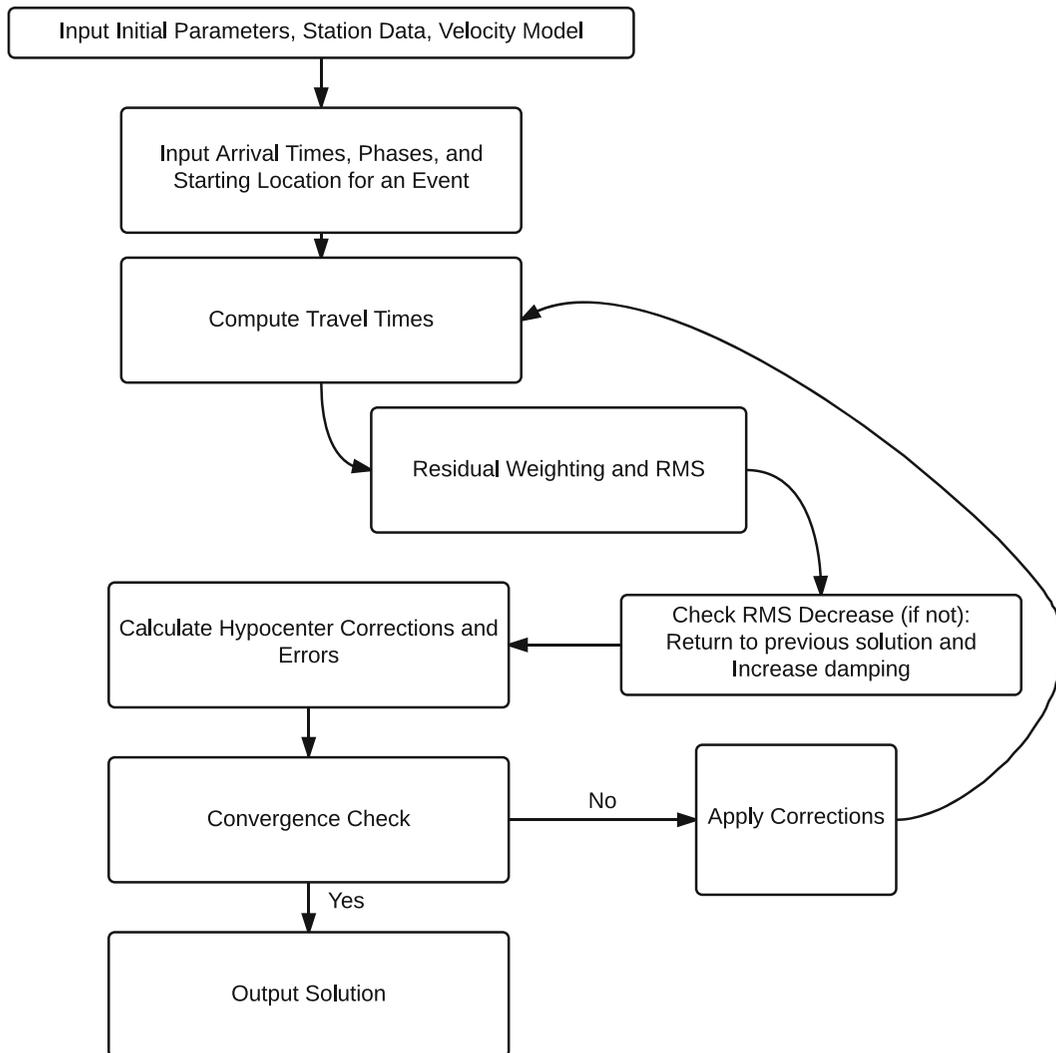


Fig. 1 Flowchart showing the algorithm of earthquake location in HYPOCENTER

After defining the grid where events are generated, we produce a map showing the location errors ε for the set of events shown in Fig. 2. We start with a case in which we have a fixed initial point m_0 (see the red square in Fig. 2) having the coordinates α_0 , β_0 , and z_0 that are the initial longitude, latitude and depth respectively. We locate the synthetic events using HYPOCENTER. Then, we compute the distance between the estimated position and the real event location as:

$$\varepsilon(\alpha, \beta, z) = |\mathbf{m}_R(\alpha, \beta, z) - \mathbf{m}_E(\alpha_0, \beta_0, z_0)| \quad (7)$$

where m_R is the exact location of the event having coordinates α , β , and z . In Fig. 3, we see the location

error (7) of the events that are located using the initial guess marked with a red square ($\alpha_0 = 3^\circ$ E longitude, $\beta_0 = 3^\circ$ N latitude and $z_0 = 5$ km depth). One of the reasons to see the high error zones even below the station network is the instability of the matrix $(G^T G)$. This numerical instability can cause an increase in the error over the whole region.

We mentioned two problems that need to be considered for a better quality estimation: (1) dependence on the initial guess, (2) biased error estimations. Dependence on the initial guess has been treated by using the procedure of OMP. Then, Eq. 7 can be updated as:

$$\varepsilon(\alpha, \beta, z) = |\mathbf{m}_R(\alpha, \beta, z) - \langle \mathbf{m}_E(\alpha_0, \beta_0, z_0(q)) \rangle| \quad (8)$$

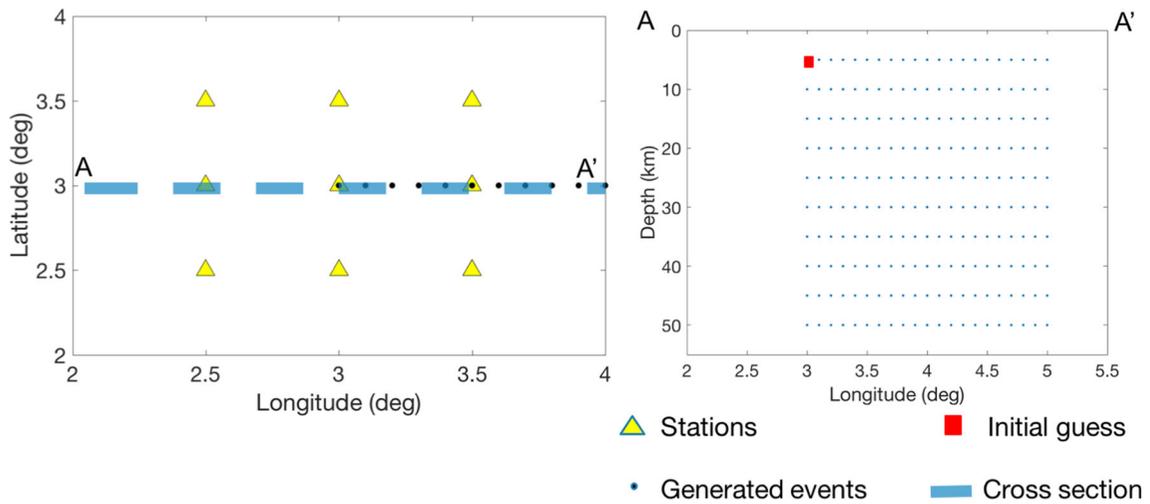


Fig. 2 Setup of the synthetic station network. Top view of the station array: triangles indicate the station locations (left). A–A’ cross section; dots indicate the synthetic event locations and red square show the initial guess m_0 (right)

where $\langle \mathbf{m}_E \rangle$ shows $\mathbf{m}_E(q)$ having minimum RMS of arrival time residuals over a set of initial depths q . To make these analyses more realistic, we added Gaussian noise with a standard deviation of 0.1 s in our initial datasets which will be discussed further in the next section. We created $T = 10$ datasets having

different noise realisations and computed the mean error $\hat{\epsilon}$ as:

$$\hat{\epsilon}(\alpha, \beta, z) = \frac{1}{T} \sum_{\tau=1}^T |\mathbf{m}_R(\alpha, \beta, z) - \langle \mathbf{m}_E^\tau(\alpha_0, \beta_0, z_0(q)) \rangle| \tag{9}$$

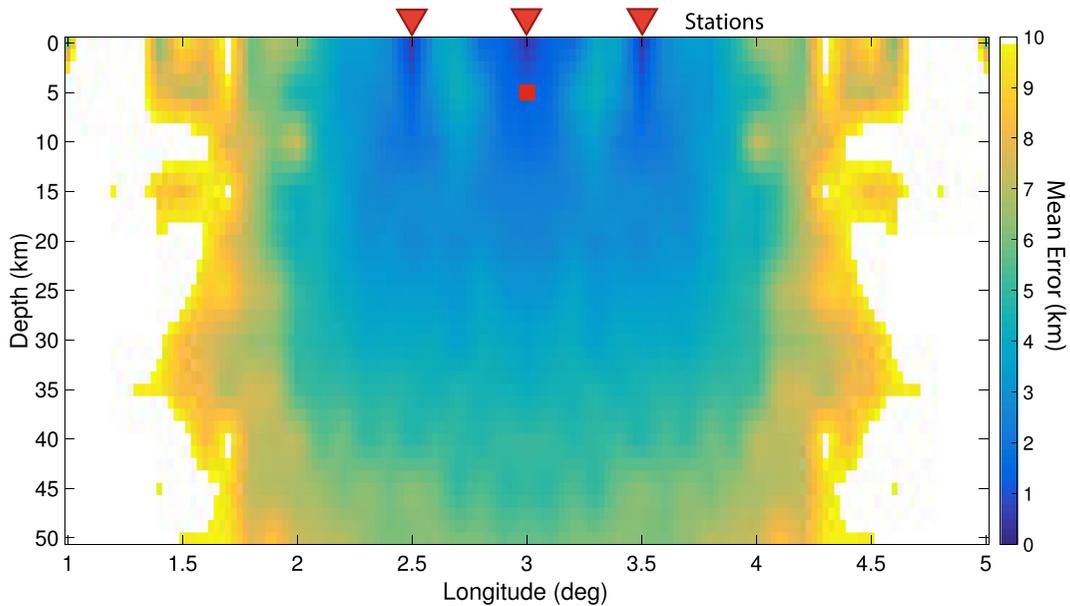


Fig. 3 An error map obtained using a fixed initial guess (red square). Red triangles show the station locations. The colour bar indicates the offset from the real source in km. Error is plotted over a grid based on true source location

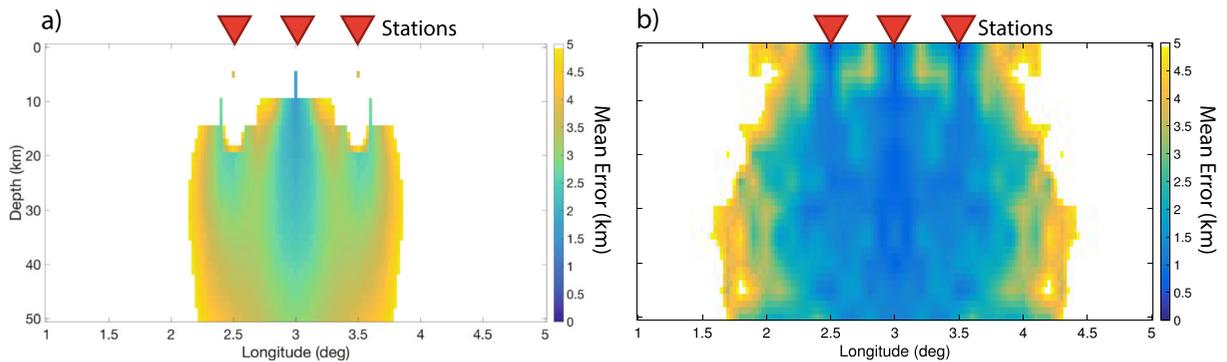


Fig. 4 Error maps obtained using different methods. **a** Conventional error map obtained using Eq. 6. **b** Proposed error map obtained using Eq. 9 for the same setup

where τ indicates the noise set. By computing $\hat{\varepsilon}$ for every point on the grid, we obtain a map of mean error and eventually compare with the uncertainty obtained from conventional formulation, i.e. Eq. 6. In Fig. 4, we see these two maps for the same dataset. Even for this simple system, the conventional results are quite different from the exact mean error.

2.4 Noise in the initial dataset

A common way to take into account the noise in the recorded seismic data in a numerical study is to add Gaussian noise to the synthetic data. We checked the evolution of error with different levels (with a standard deviation 0.1 and 0.2 s) of noise in the arrival time dataset on the mean error maps. We saw that an increase in the noise level increases the error in almost every location. A 100% increase in the standard deviation of the Gaussian noise causes a 67% increase of the maximum of the mean error map. In

Supplementary Figure S1, we see that the error increase is almost constant everywhere in the map and it does not have a particular signature.

Furthermore, we are comparing two types of noise functions in a simple synthetic model in Fig. 5: (a) noise having 0 s mean and 0.1 s standard deviation for all the events around the grid; and (b) noise having 0 s mean and linearly varying standard deviation from 0.05 s for the events closest to the stations and 0.15 s for the events farthest to the stations. When we compare these two cases, we clearly see that the signature of the error map that we are interested in during this study does not change significantly based on the type of noise used on the input data.

2.5 Effect of 3D velocity model

In this section, we present an application to the Pyrenean region using a 3D velocity model. We used the station coordinates presented in Theunissen et al.

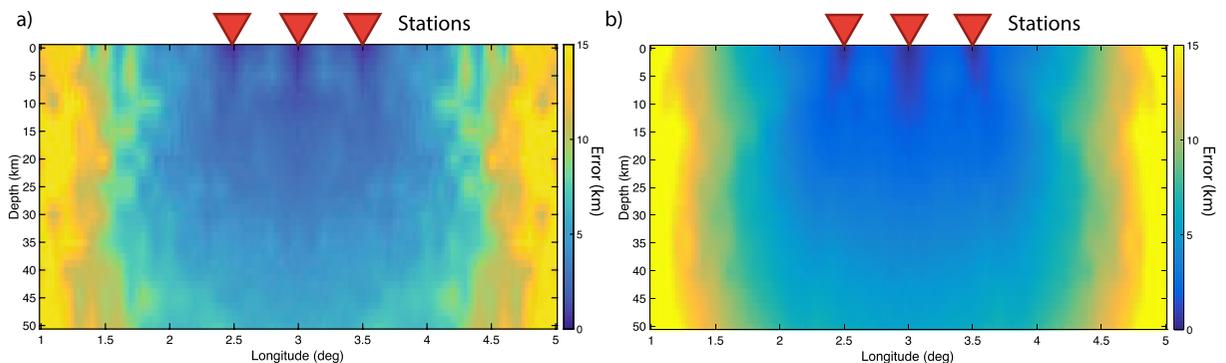


Fig. 5 Mean error map of the simple region using different noise functions: **a** $\sigma_N = 0.1$ s and **b** $\sigma_N = 0.05 - 0.15$ s (linearly increasing with distance between the true location of the event and the station) noise are added

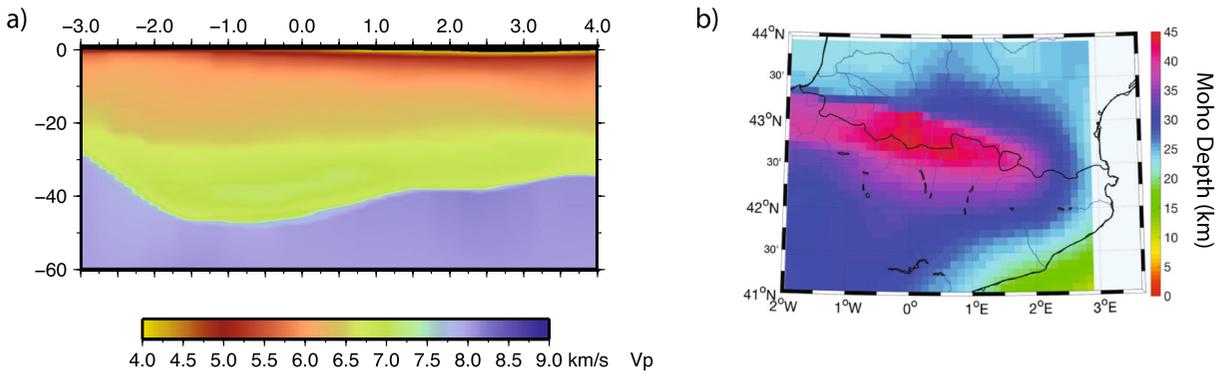


Fig. 6 Details of the 3D velocity model. **a** A slice on the latitude 42.5° N of the 3D P wave velocity profile. The colour bar shows the P wave velocity in km/s. **b** A map of Moho depth in the region. The colour bar shows the Moho depth in km

(2017). Our study covers the region between longitude 2°W–3.4°E and latitude 42°–43.6°N. We have generated 6300 events positioned at each node of a regular grid (0.5° spacing over longitude and latitude and 1 km spacing in depth, down to 25 km in depth). We use the Fast Marching Method (FMM) to compute synthetic arrival times in a 3D model (Rawlinson and Sambridge 2005). A brief explanation of the method is provided in the Supplementary Material A - Fast Marching Method.

A cross section from the 3D velocity model showing the P wave velocity profile between longitudes 3°W–4°E and latitude 42.5° N is given in Fig. 6a. Further cross sections can be found in the Supplementary Fig. S2 - 3D Velocity Model. In Fig. 6b, we provided a map of Moho depth in the region.

Due to inhomogeneities in the 3D velocity structure, one cannot locate earthquakes without error by using a 1D velocity model-based earthquake location software. In Fig. 7, we quantify these errors. In this figure, both maps are prepared using Eq. 9 showing an arithmetic mean of $\hat{\epsilon}$ over 25 km depth. In Fig. 7a, we present an error map in which the arrival times are generated with a 1D velocity model (see Table 1, see Supplementary Material B - Choice of 1D Velocity Model for the Pyrenees for more details) and the synthetic events are located using the same velocity model. However, in Fig. 7, we generate the arrival times using a 3D velocity model presented in Fig. 6 and we use the 1D velocity model given in Table 1 to locate the events. We see in the colour bars of Fig. 7a and b that 3D velocity variations in the model

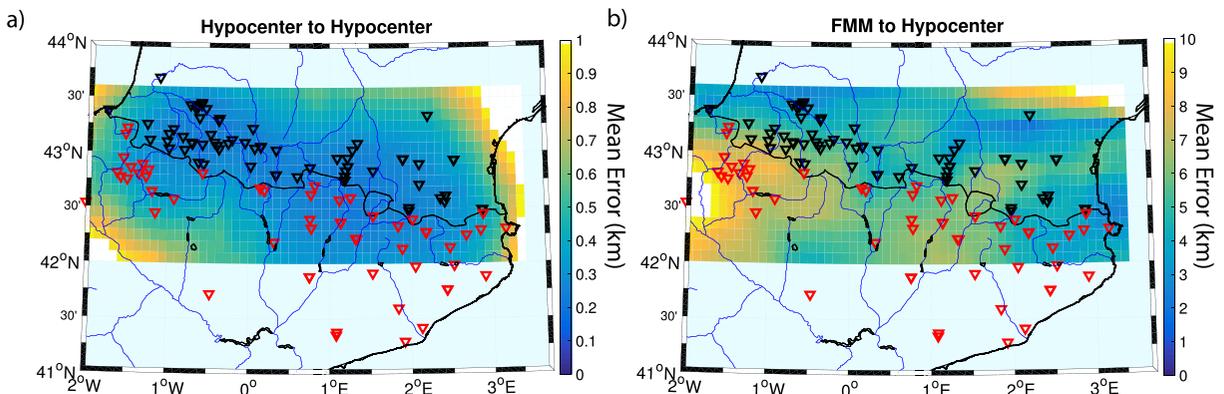


Fig. 7 Mean error map of the Pyrenean Region using **a** HYPOCENTER and **b** Fast Marching Method to generate arrival times. The colour bar shows the mean of the error (in

km) of the all events over 25 km depth. Triangle points show the stations in France; circles show the stations in Spain

Table 1 1D velocity model of the Pyrenees

| Depth (km) | V_p (km/s) |
|------------|--------------|
| 0 | 5.0 |
| 5 | 5.0 |
| 10 | 5.5 |
| 20 | 6.0 |
| 30 | 7.0 |
| 40 | 7.5 |
| 60 | 8.0 |

increase the mean error observed on the map up to 10 times.

3 Results and discussion

3.1 Comparison of synthetic work with the catalogue

In this study, first we compare the synthetic results with the relocation of earthquakes recorded by the observatory. To locate events, we use the constraining parameters that we have defined earlier. For a fully controlled synthetic test, we identify the differences between the uncertainties computed using Eq. 6 and the exact distance (i.e. Eq. 9) between estimated location and exact location of the event. The mean uncertainty obtained from relocation (i.e. $E=1.5$ km) is of the same order as the one given in Theunissen

et al. (2017) (i.e. $E=3$ km). This difference occurs because the parameters, velocity model and software used in Theunissen et al. (2017) are not exactly the same as the ones used in this study.

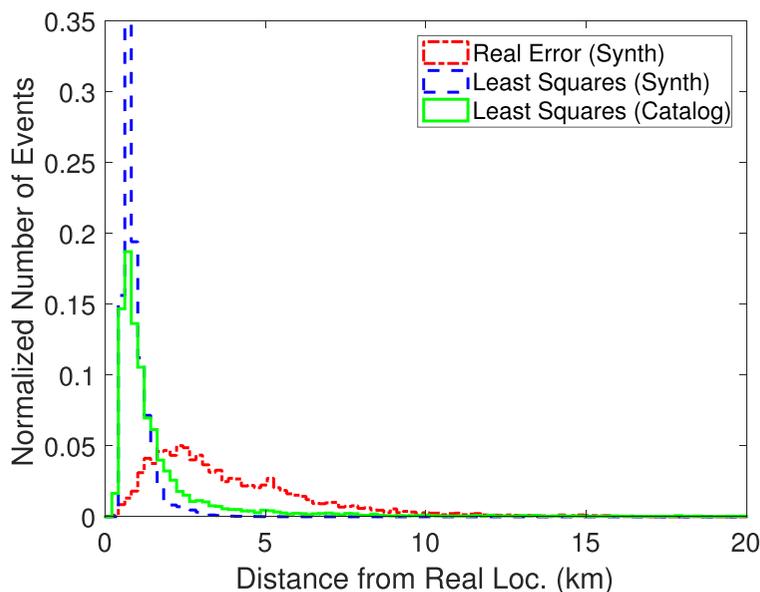
As we see in Fig. 8, the blue and green histograms are very similar. However, the red curve which shows the exact error is overall larger, more spread than the other histograms. Equation 6 can estimate the error which is 3–4 times smaller than the average of the error of 6300 events and 10 times smaller than the maximum of error of these events.

3.2 Effect of S arrivals in the input data

After quantifying the effect of 3D velocity model in Section 2.5, we want to understand the effect of P and S pairs. For the analysis presented in Fig. 7, we used P and S arrivals together. To quantify the effect of S arrivals on the quality of estimation, we removed S arrivals and located the events using the same setup. This effect depends on the velocity model used in the study, but we can quantify it for this particular setup.

In Fig. 9, we clearly see how the quality of the estimation increases with the use of S wave arrivals. When we use P and S arrivals together, we see that the maximum error recorded in the region is decreased approximately by 40%. The use of S wave arrivals does not particularly ameliorate the results of well-located events but it increases the quality of poorly located events. This can particularly be seen at

Fig. 8 Comparison of computed uncertainties with measured exact error. Blue and green curves are obtained using Eq. 6 and red curve is obtained using Eq. 9



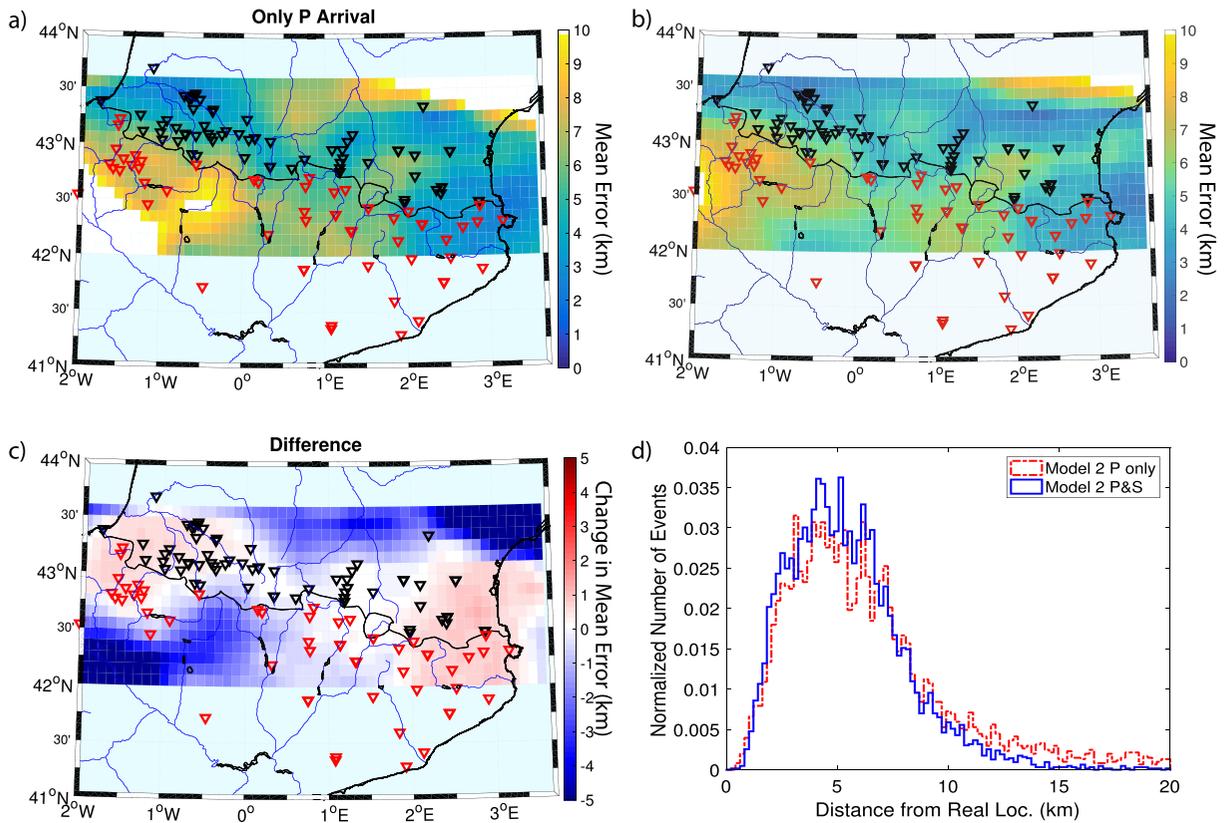


Fig. 9 Estimation quality increases by using S arrivals. In the subfigures, error maps of the region without (a) and with (b) S wave arrivals are given. c The difference of these maps is

provided by subtracting image (b) from (a). d Histogram of the error sets corresponding to (a) and (b)

the top-right and bottom-left corners of the region shown with saturated white pixels in Fig. 9a.

smaller RMS). The following formula is used to compute the station corrections for our synthetic study:

3.3 Effect of the station corrections

$$\bar{\delta t}^x = \frac{1}{N} \sum_{i=1}^N (t_o^x(i) - t_e^x(i)) \tag{10}$$

A common technique to increase the quality of estimations obtained using a 1D velocity model is to integrate 1D station corrections. This method helps reduce the systematic errors due to the differences between the 3D velocity model used to generate events (or the events recorded in the observatory) and the 1D model used in estimation. A constant value is added to the arrival times recorded at a particular station to account for the abrupt velocity variations in the vicinity of that station. In our synthetic tests, these station corrections strictly depend on the chosen underlying 3D velocity model. HYPOCENTER is one of the publicly available software that can take into account corrections to reduce the residual between the observed arrival time t_o and the estimated arrival time t_e for a better fit (i.e.

where $\bar{\delta t}^x$ is the correction for station x ; i is the number of an event over N events.

Using the station corrections, we obtain maximum error $\epsilon_{max} = 24.2$ km, mean error $\bar{\epsilon} = 4.6$ km (which is the arithmetic mean of $\hat{\epsilon}$ over the region), and minimum error $\epsilon_{min} = 0.22$ km over the entire zone of study. These corrections increase the average quality by 25% and decrease the minimum error by 78%. In Fig. 10a and b, we see the error maps without station corrections and with station corrections, respectively. By subtracting these two maps, we obtain a map which indicates the zones where we decreased the error (positive values mean a decrease

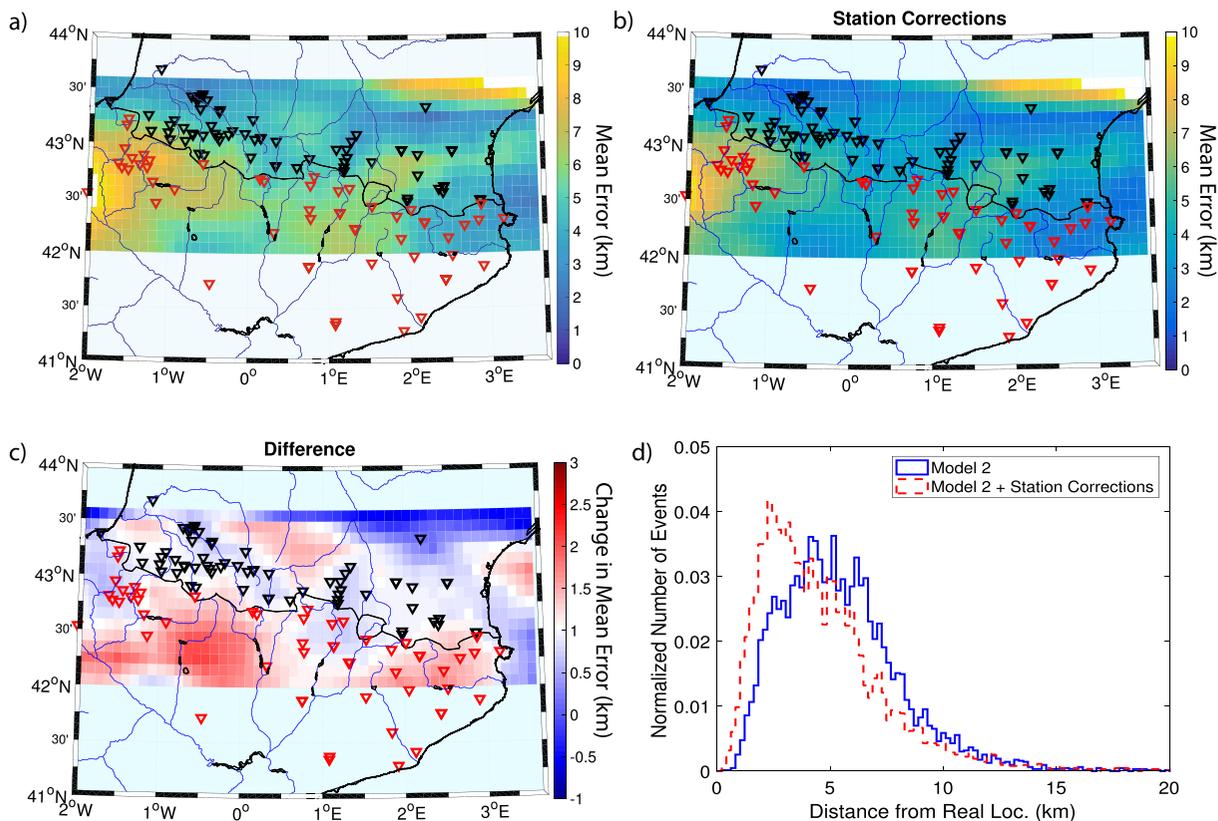


Fig. 10 Estimation quality increase by using the station corrections. In the subfigures, two error maps of the region without (a) and with (b) station corrections are given. c The difference

of these maps is provided by subtracting image (b) from (a). d Histogram of the error sets correspond to (a) and (b)

in mean error $\hat{\epsilon}$) in Fig. 10c. We also compare the histogram of these two mean error sets in Fig. 10d which shows the general stabilising effect of station corrections. We observe in Fig. 10 that there is a general tendency to have smaller errors when station corrections are applied. We see particularly that the use of station corrections has almost no negative effects on the error map and thus we strongly recommend such a procedure.

3.4 Effect of spatial coverage

The Pyrenean region covers the border between France and Spain, and both countries have seismic stations in the region. Figure 7 shows the distribution of stations used in our synthetic study. 56 of the stations belong to Spain and 66 of them belong to France. The stations are assumed to be identical. In the simulations, the only difference is the spatial spreading of the station network.

In Fig. 11a and b, we see the important effect of station networks. When we mark the stations on this map, we see that as density of the stations increases, the error measured in estimations decreases. To analyse this point further, we prepare two error maps using only French stations and only Spanish stations. In Fig. 11a and b, we see that if we remove the stations in distance, we obtain a mean error lower than 5 km in the zones in the vicinity of the station network. Furthermore, we can quantify this variation with respect to the initial error map where we used all of the stations. In Fig. 11c and d, we identify the necessity to optimise the distance between the event and the stations. The farther away we get from the event, the more important reflected and refracted phases become. Instead of direct arrivals, we start to have indirect phases in our recordings (see Supplementary Material C: Effect of the station aperture for a detailed analysis on a simplified case). These indirect arrivals bear higher average uncertainties, which means the

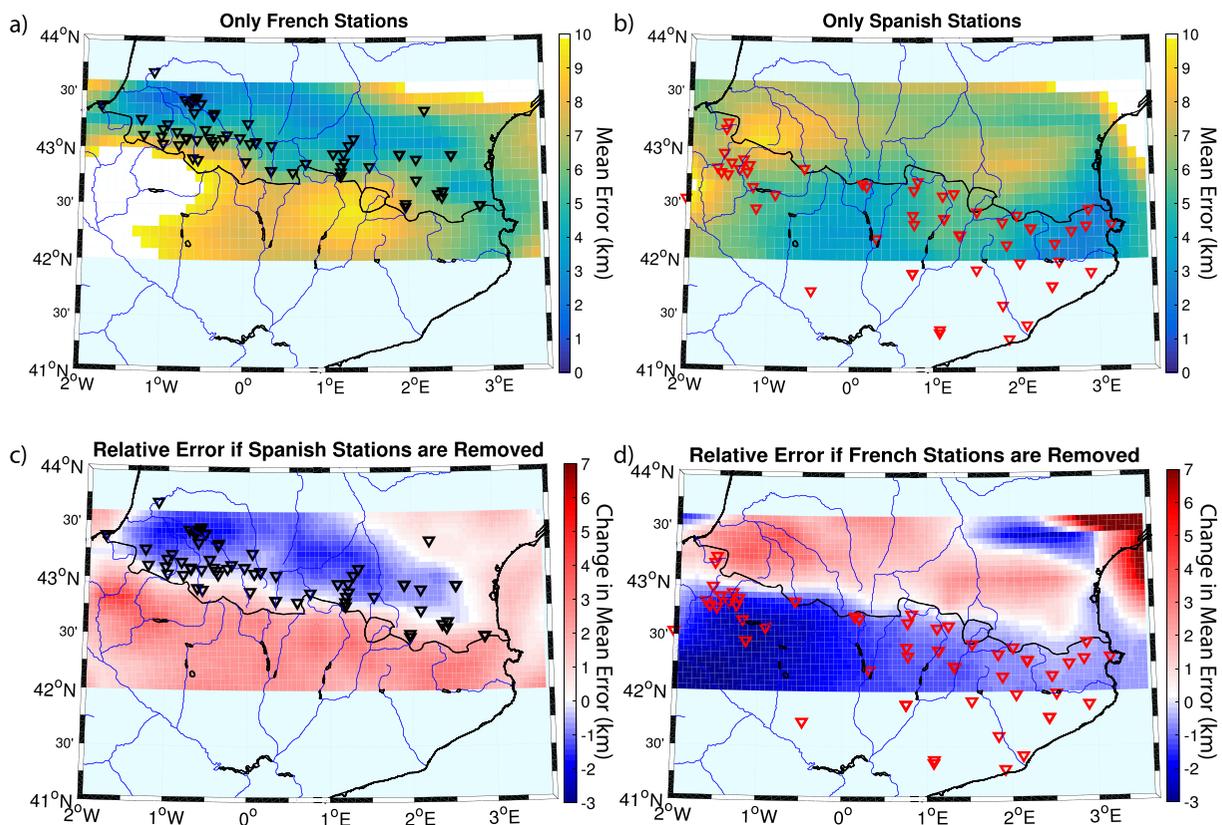


Fig. 11 Mean error map of the Pyrenean region computed with different station sets. The map is computed using only the stations in France (a) and in Spain (b). Then, the difference with the map computed using all the stations is calculated respectively in (c) and (d)

residuals between the recorded and modelled arrival times are higher at distant stations than closer stations. This happens due to the differences between the 1D velocity model used to estimate the location and the 3D velocity model used to generate the seismic dataset. We are going to investigate this further in the next section.

3.5 Maximum event–station distance

After observing the fact that using subsets of the big station network increases the estimation quality, we want to quantify this quality variation with respect to maximum event–station distance. We choose 5 events in the Pyrenean region. Next, we locate these synthetic events with maximum event–station ranges starting from 10 up to 1000 km. We observed that the error ϵ decreases with the distance until 130 km and increases beyond this threshold. We decided to apply linearly decaying weighting, starting from 130 down to 0 km

at distance 230 km. This weighting function is very similar to that used by OMP as described previously in Section 2.1.

As we can see by comparing Fig. 12a and b—which are without and with range correction, respectively—the range correction increases the quality significantly on the zones having a high station density. However, we also observe that, in the zones with a lower station density, estimation error increases since there are not enough stations to ensure a high-quality location estimation. In particular in Fig. 12d, we see the significant shift towards the left on the histogram just by adding the range correction to the data, which means that the overall quality of estimation increases significantly. Approximately 50% reduction in mean error can be achieved by taking into account the maximum event–station range weighting. It should be noted that the parameters that we use for this weighting are specific to this very region and to the velocity models that we have used to generate and locate the events. For

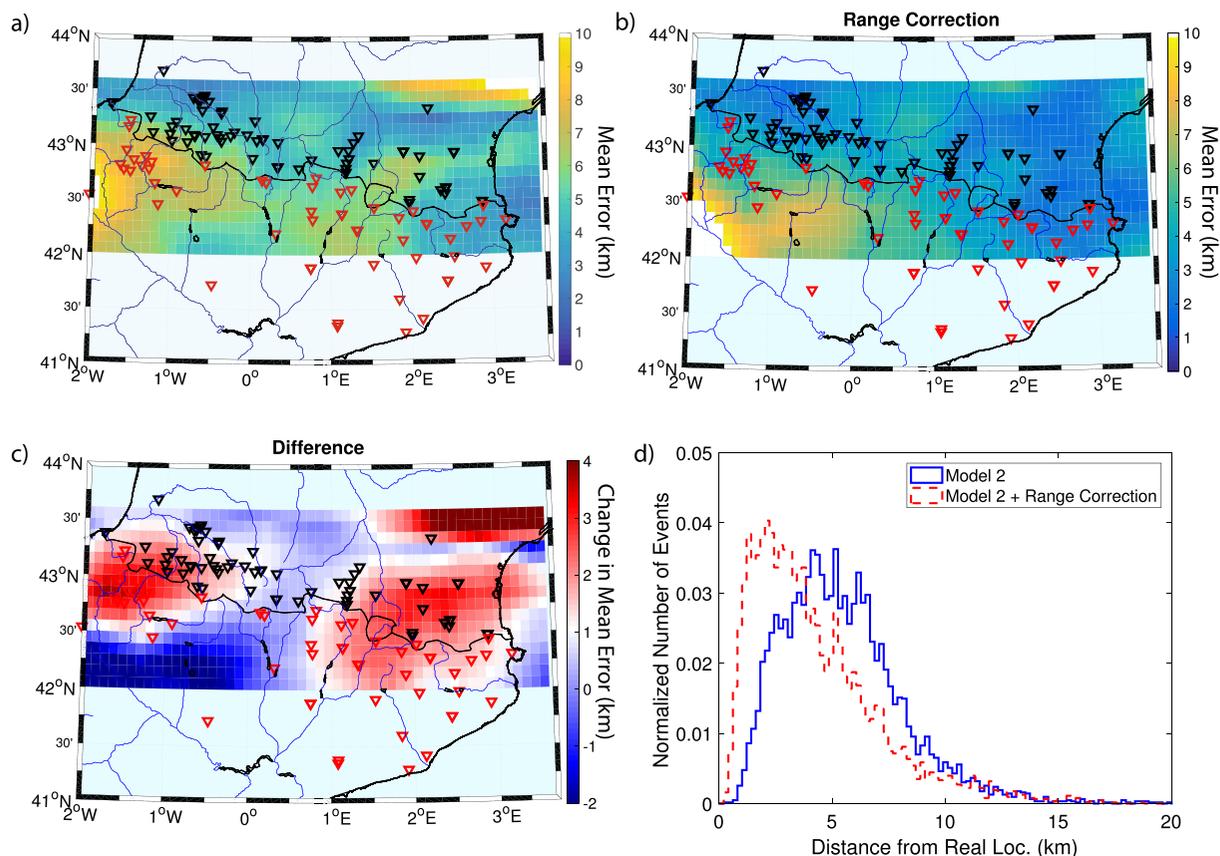


Fig. 12 Mean error map of the Pyrenean region computed with maximum event station distance weighting. The map is computed without any weighting (**a**) and with weighting (**b**). Then,

the difference with the map computed using all the stations is calculated respectively in (**c**), and the histogram of these two maps is provided in (**d**)

another region or another velocity model, one should optimise their own weighting parameters via similar synthetic tests. Furthermore, the maximum range and station corrections can be applied together to obtain even better quality estimations. In Fig. 13, we present the case where we apply both of these corrections and we see that the number of events having lower than a 5-km error increases more than 25% compared to the cases where these corrections are applied separately.

3.6 Subregional velocity models

In earthquake location, there are many studies advising use of 3D velocity models for better estimation quality (Font et al. 2004; Flanagan et al. 2007; Simmons et al. 2012; Theunissen et al. 2012). If 1D models are used, it has been mentioned that in complex heterogeneous zones, the use of a subregional 1D

velocity model can increase the quality of estimation (Husen et al. 2011).

In this section, we aim at quantifying the effect of subregional 1D velocity models. Instead of using a unique 1D velocity model in the zone of interest, the area is divided into subregions and a 1D velocity model is defined for each region. To calculate the mean velocity model in an area closest to an event, we need to know the location of the event. Since we do not know the location of an event, we consider the closest station and calculate the mean velocity model around it. To define the models, we stack the 3D velocity model within a 230-km radius (corresponding to the optimal range computed in a previous section) of each station so that we obtain one velocity model per station. In other words, we build one velocity model corresponding to the area around the initial guess for one event.

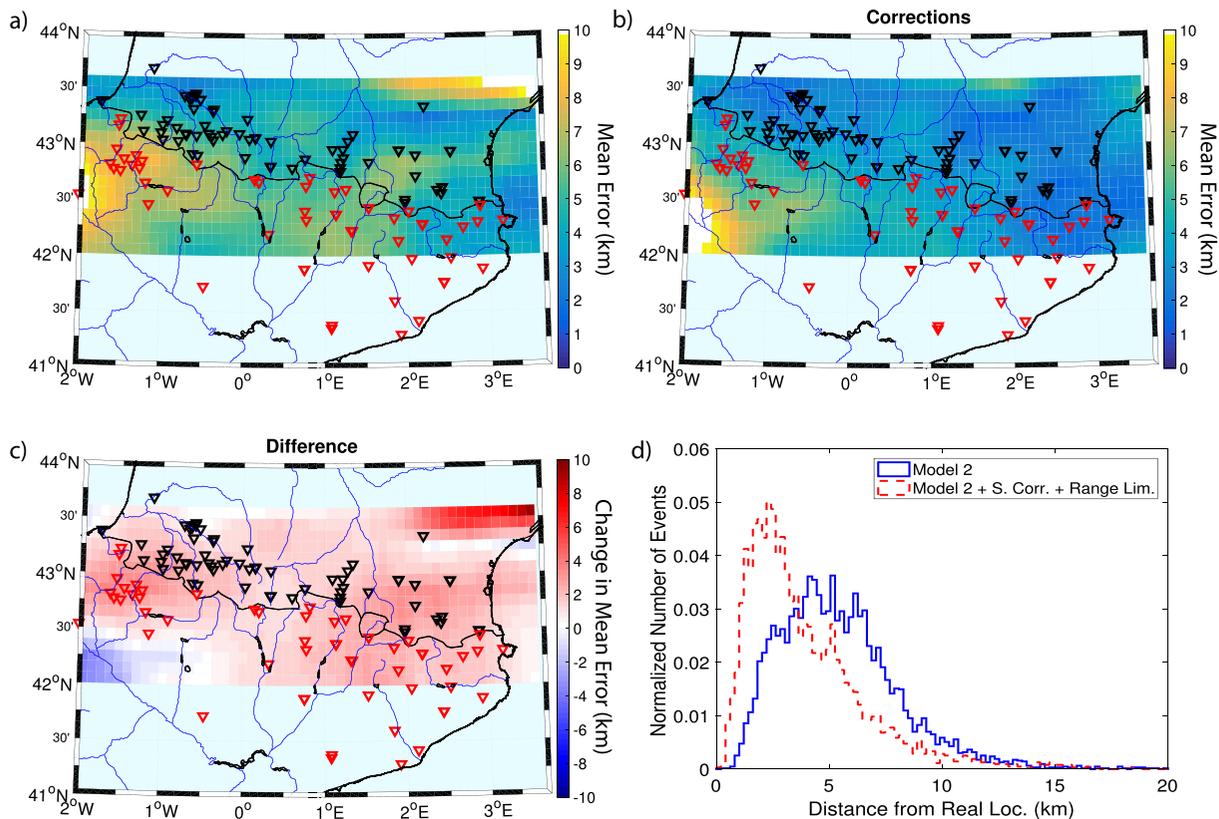


Fig. 13 Mean error map of the Pyrenean region computed using distance weighting and station corrections. Then, the difference with the map computed using all the stations is calculated respectively in (c) and the histogram of these two maps is provided in (d)

Figure 14 illustrates the effect of the subregional velocity models applied on the Pyrenean region. In Fig. 14a, we see the mean error map obtained with a regional 1D velocity model and Fig. 14b shows the same map obtained using subregional 1D velocity models. There is a significant increase in the quality (around 50% decrease in estimation error) in the zone where the station network is quite dense; however, there is a significant decrease of estimation quality (around 80–90% increase in estimation error) in the regions away from the station network. A map of mean error difference is provided in Fig. 14c to better identify the affected zones. The red colour shows a decrease in estimation error and the blue colour shows an increase in estimation error. A histogram of the estimation error over the events generated of the region is given in Fig. 14d. We clearly see in the histogram that when we use subregional velocity models, the events having the location error smaller than 10 km in Fig. 14a have their error further decreased in Fig. 14b. However, events estimated with a higher error have

their estimation quality decreased even further. For the earthquakes located under the station network, the use of subregional velocity models is a very powerful way to increase source location quality.

3.7 Effect of number of phases

We made a detailed analysis of the effect of the number of recorded P and S pairs on the absolute error for the region of interest. For 63 selected grid points spread over the region, we generated synthetic events with different P and S pairs. The number of input arrival time pairs is weighted with the distance as described in Section 2.1 to be able to quantify the mean error when we follow the standard protocol of the observatory. For location, we are taking into account all the PS pairs closer than 30 km to the station. Following this, we are picking stations randomly (with a uniform picking probability for all stations) up to the P and S pair limit. We are doing 10 repetitions for each event with different PS sets to be

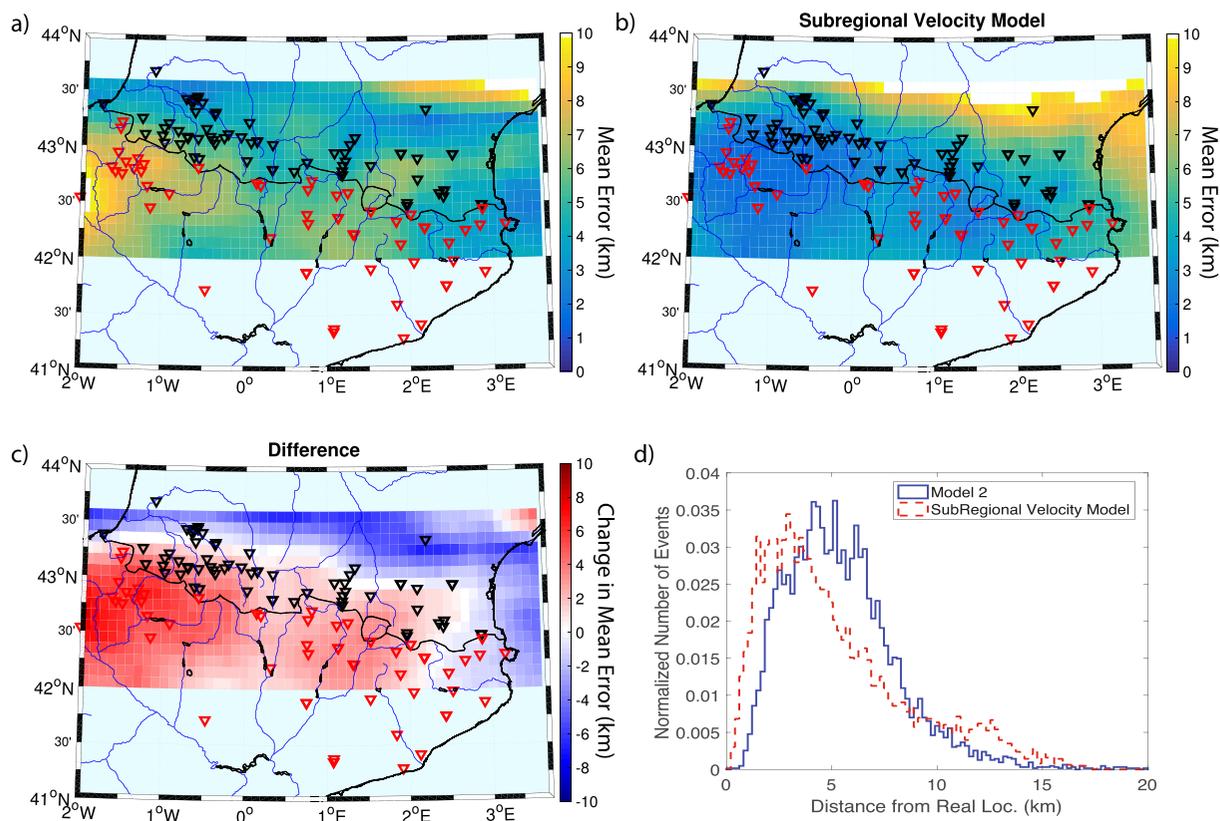


Fig. 14 Mean error map of the Pyrenean region computed using subregional velocity models. The map is computed using velocity model given in Table 1 (a) and using mean of the 3D velocity model 230 km vicinity of the initial guess for the event

more realistic. We saw that the mean error for the region reduces asymptotically from 7.8 km towards 5.5 km with increasing number of P and S pairs from 5 towards 25. In Fig. 15, we present how the estimation error decreases with an increasing number of pairs. Up to 20 pairs, the estimation quality increases as the number of PS pairs increases. Starting from 20 pairs, the quality does not vary too much for an event based on the number of PS pairs.

3.8 Discussion

The mean error map proposed in this article is a tool for earthquake location error estimation. Knowing that the conventional tools assessing the earthquake location error are limited, one can use a mean error map produced for a region to directly find the associated error for a located earthquake. In Sections 2.3 and 3.1 of this article, we have shown this limitation clearly.

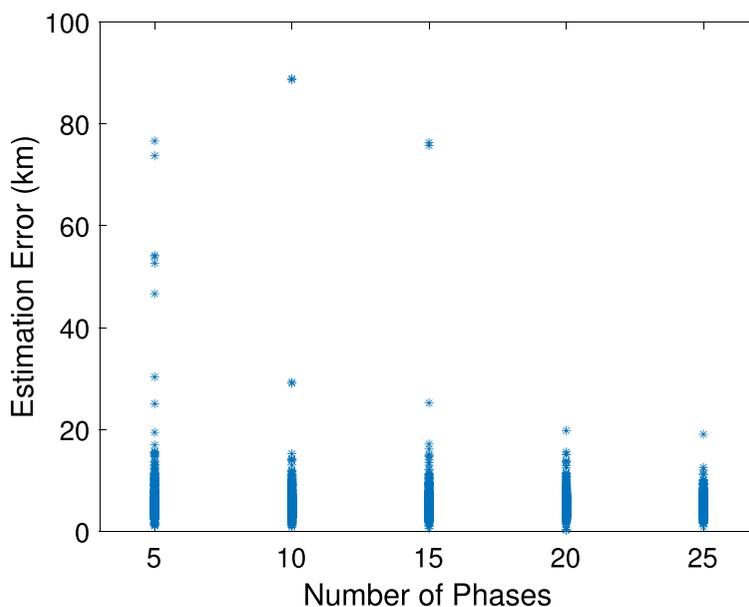
(b). Then, the difference with the map computed using all the stations is calculated respectively in (c) and the histogram of these two maps is provided in (d)

Furthermore, since the mean error map requires generating many synthetic events on a regular grid over the region, it can be used to identify the zones in which the location error is systematically higher than the overall. Covering these weak zones with more stations can increase the seismic resolution significantly.

The proposed error estimation procedure can also be used to calibrate the temporary (or weak) networks. Using a synthetic analysis, the resolution of temporary networks can be adjusted and sensor placements can be optimised. This will help increase the quality of the subsurface monitoring and risk assessment. For example, this technique can be used to increase the control during industrial operations such as carbon storage or geothermal energy.

Furthermore, a mean error map can be used to optimise the software parameters. By checking the variation of the mean error on a region as function of certain software parameters as we conducted in this

Fig. 15 The variation of the error in function of number of phases. Synthetic events are located with 10 different phase pair sets. As the number of PS pairs increases, the outliers disappear and location estimation gives more robust solutions



study, one can maximise the quality of the seismic locations and minimise the overall error on the region of interest.

Apart from the software parameters, one other important issue is understanding the effect of complementary techniques to obtain better quality estimations. We recommend using station corrections as a tool to reduce the effect of un-modelled 3D variations in the velocity model. It is a very practical way to overcome the constant differences between the modelled and recorded arrival times on a specific station. These constant differences can occur naturally since 1D velocity model is not enough to explain local velocity fluctuations. The second most important thing is weighting with respect to the event-station distance. Correctly optimising this distance may decrease the average error significantly. Because the wavefront travelling longer distances will be more affected by the 3D effects and eventually can be harder to model with a 1D velocity model. When we limit this distance, we limit the 3D effects and achieve higher quality event locations.

We studied the effect of the use of subregional velocity models as well. Varying 1D velocity model for each station is a more tedious procedure than using station corrections. However, for this tedious procedure the final estimation quality is not better than using only station corrections. The maximum error

reduction achieved is around 75% when using station corrections compared with 50% when using subregional velocity models. Therefore, the authors prefer using station corrections for future studies on this region. Subregional velocity models can be more useful in the regions having more abrupt changes in their 3D velocity model than the Pyrenees. However, in Pyrenean region, velocity fluctuations are not high enough to make subregional velocity models preferable. Another question that we wanted to answer is the effect of the number of phases in the catalogues. In seismic recordings, the number of phases recorded varies from event to event. We compared the evolution of average error with respect to number of phases included. We observed that in up to 20 PS pairs mean error decreases and becomes stable for pair numbers higher than 20. This number can be changed with respect to the other parameters chosen such as the maximum event-station distance or the 3D velocity model. A future analysis can be conducted to create a correction function linking the number of events taken into account in the location and the mean error for that event to correct the events recorded with a number of pairs lower than 25.

The results presented in this paper can vary as a function of the velocity model chosen for the numerical modelling. Therefore, it is recommended to use

high-quality velocity models to reach better quality results on the mean error maps. Also, the mean error is related to the noise level on the numerically generated data. In this study, the standard deviation of the noise in the input data is set to 0.1 s but this can be increased to better analyse the regions where the recorded data has overall lower quality.

4 Conclusion

We have conducted a study to quantify the error in earthquake location obtained from an iterative linearised earthquake location method. We proposed a method to compute mean error maps. These maps show the mean distance between the estimated location and the exact location of an event on a grid (based on true event locations). We showed that the effects of including S arrival times are mostly noticeable when the events are farther away from the station network. We have also observed that it is possible to increase the quality of estimations using station corrections for arrival times. For Pyrenean region, station corrections are more efficient than using 1D subregional velocity models. We have observed that the maximum event - station distance should be taken into account in an optimised way to increase estimation quality. As we include farther stations to the solution procedure, after a certain threshold around 100 km, the mean error increases. We have also shown that up to 20 PS pairs—respecting the maximum event–station distance—increasing the number of pairs, increases the location quality.

Acknowledgements The authors thank the editor and the reviewers for their precious comments to make the article more reader friendly. The authors would like to thank, for the fruitful discussions, Renaud Toussaint and Francois Renard. Furthermore, the authors would like to thank Thomas Theunissen for providing the seismicity dataset of the Pyrenean Region and his remarks on the project. We also thank Cedric Twardzik from Geoazur (Nice, France), Stephane Rondenay from University of Bergen, Knut Jorgen Maloy from PoreLab (Oslo, Norway), Sophie Lambotte from IPG (Strasbourg, France) and Francois Renard from ISTerre (Grenoble, France) for providing the opportunity for laboratory visits and discussions with their teams during the preparation of this manuscript.

Funding information This project is financed by Electricity of France (EDF) via the project SINAPS@.

References

- Abrahamson NA, Bommer JJ (2005) Probability and uncertainty in seismic hazard analysis. *Earthquake Spectra* 21(2):603–607. <https://doi.org/10.1193/1.1899158>
- Billings SD, Sambridge MS, Kennett BLN (1994) Errors in hypocenter location: picking, model, and magnitude dependence. *Bullet Seismol Soc Amer* 84(6):1978
- De Natale G, Zollo A, Del Gaudio C, Ricciardi G, Martini M (1984) Error analysis in hypocentral locations at Phlegraean Fields. *Bull Volcanol* 47(2):209–218
- Flanagan MP, Myers SC, Koper KD (2007) Regional travel-time uncertainty and seismic location improvement using a three-dimensional a priori velocity model. *Bull Seismol Soc Am* 97(3):804–825
- Font Y, Kao H, Lallemand S, Liu CS, Chiao LY (2004) Hypocentre determination offshore of eastern Taiwan using the maximum intersection method. *Geophys J Int* 158(2):655–675
- Husen S, Kissling E, Clinton JF (2011) Local and regional minimum 1d models for earthquake location and data quality assessment in complex tectonic regions: application to Switzerland. *Swiss J Geosci* 104(3):455–469
- Jordan TH, Sverdrup KA (1981) Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific. *Bull Seismol Soc Am* 71(4):1105–1130
- Kissling E, Kradolfer U, Maurer H (1995) Program velest user's guide-short introduction. Institute of Geophysics, ETH Zurich
- Lahr J, Chouet B, Stephens C, Power J, Page R (1994) Earthquake classification, location, and error analysis in a volcanic environment: implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska. *J Volcanol Geotherm Res* 62(1–4):137–151
- Lahr JC (1999) HYPOELLIPSE: A computer program for determining local earthquake hypocentral parameters, magnitude and first motion pattern. US Department of the Interior, US Geological Survey
- Lee WHK, Lahr JC (1972) Hypo71: a computer program for determining hypocenter, magnitude and first motion pattern of local earthquakes. Technical report, US Geological Survey
- Lee Y, Taylor C, Hu Z, Graf WP, Huyck CK (2014) Uncertainty estimates for earthquake hazard analysis through robust simulation. In: Tenth US National Conference on Earthquake Engineering Frontiers of Earthquake Engineering, pp 21–25
- Lienert BR, Berg E, Frazer LN (1986) Hypocenter: an earthquake location method using centered, scaled, and adaptively damped least squares. *Bull Seismol Soc Am* 76(3):771–783
- Lienert BR, Havskov J (1995) A computer program for locating earthquakes both locally and globally. *Seismol Res Lett* 66(5):26–36
- Lomax A, Virieux J, Volant P, Berge-Thierry C (2000) Probabilistic earthquake location in 3d and layered models. In: *Advances in seismic event location*. Springer, pp 101–134

- Lomax A, Michelini A, Curtis A (2009) Earthquake location, direct, global-search/global-search methods. In: Encyclopedia of complexity and systems science. Springer, pp 2449–2473
- Pavlis GL (1986) Appraising earthquake hypocenter location errors: a complete, practical approach for single-event locations. *Bullet Seismol Soc Amer* 76(6):1699
- Pavlis GL (1992) Appraising relative earthquake location errors. *Bullet Seismol Soc Amer* 82(2):836
- Peters DC, Crosson RS (1972) Application of prediction analysis to hypocenter determination using a local array. *Bull Seismol Soc Am* 62(3):775–788
- Rawlinson N, Sambridge M (2005) The fast marching method: an effective tool for tomographic imaging and tracking multiple phases in complex layered media. *Explor Geophys* 36(4):341–350
- Richards PG, Waldhauser F, Schaff D, Kim WY (2006) The applicability of modern methods of earthquake location. *Pure Appl Geophys* 163(2):351–372. <https://doi.org/10.1007/s00024-005-0019-5>
- Simmons NA, Myers SC, Johannesson G, Matzel E (2012) Llnl-g3dv3: Global p wave tomography model for improved regional and teleseismic travel time prediction. *Journal of Geophysical Research: Solid Earth* 117(B10). <https://doi.org/10.1029/2012JB009525>
- Tarantola A, Valette B (1982) Generalized nonlinear inverse problems solved using the least squares criterion. *Rev Geophys* 20(2):219–232
- Theunissen T, Lallemand S, Font Y, Gautier S, Lee CS, Liang WT, Wu F, Berthet T (2012) Crustal deformation at the southernmost part of the Ryukyu subduction (East Taiwan) as revealed by new marine seismic experiments. *Tectonophysics* 578:10–30
- Theunissen T, Chevrot S, Sylvander M, Monteiller V, Calvet M, Villaseñor A, Benahmed S, Pauchet H, Grimaud F (2017) Absolute earthquake locations using 3-d versus 1-d velocity models below a local seismic network: example from the pyrenees. *Geophys J Int* 212(3):1806–1828
- Thurber CH (1986) Analysis methods for kinematic data from local earthquakes. *Rev Geophys* 24(4):793–805. <https://doi.org/10.1029/RG024i004p00793>
- Thurber CH (1992) Hypocenter-velocity structure coupling in local earthquake tomography. *Phys Earth Planet Inter* 75(1–3):55–62
- Tramelli A, Troise C, De Natale G, Orazi M (2013) A new method for optimization and testing of microseismic networks: an application to Campi Flegrei (southern Italy). *Bull Seismol Soc Am* 103(3):1679–1691
- Turkaya S, Toussaint R, Eriksen FK, Lengliné O, Daniel G, Flekkøy EG, Måløy KJ (2016) Note: Localization based on estimated source energy homogeneity. *Rev Sci Instrum* 87(9):096101. <https://doi.org/10.1063/1.4962407>
- Turquet AL, Toussaint R, Eriksen FK, Daniel G, Lengline O, Flekkoy EG, Maloy KJ (2019) Source localization of microseismic emissions during pneumatic fracturing. *Geophys Res Lett* 46(7):3726–3733. <https://doi.org/10.1029/2019GL082198>. <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL082198>
- Waldhauser F (2001) hypodd—a program to compute double-difference hypocenter locations (hypodd version 1.0-03/2001). US Geol Surv Open-File Rept 01 113

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.