# Relative earthquake location procedure for clustered seismicity with a single station

Francesco Grigoli [1,2] William L. Ellsworth,[2] Miao Zhang [3] Mostafa Mousavi,[2] Simone Cesca,[4] Claudio Satriano,[5] Gregory C. Beroza[2] and Stefan Wiemer[1]

[1]*Swiss Seismological Service (SED), ETH-Zurich, Zurich,CH- 8092, 8092, Switzerland. E-mail: francesco.grigoli@sed.ethz.ch*
[2]*Department of Geohysics, Stanford University, Stanford, CA 94305, USA*
[3]*Department of Earth and Environmental Sciences, Dalhousie University, Halifax, NS B3H 4R2, Canada*
[4]*German Research Centre for Geosciences (GFZ), Potsdam, DE-14473, Germany*
[5]*Institute de Physique du Globe de Paris (IPGP), Université de Paris, Paris, F-75005, France*

## SUMMARY

Earthquake location is one of the oldest problems in seismology, yet remains an active research topic. With dense seismic monitoring networks, it is possible to obtain reliable locations for microearthquakes; however, in many cases dense networks are lacking, limiting the location accuracy, or preventing location when there are too few observations. For small events in all settings, recording may be sparse and location may be difficult due to low signal-to-noise ratio. In this work, we introduce a new, distance-geometry-based method to locate seismicity clusters using only one or two seismic stations. A distance geometry problem consists in determining the location of sets of points based only on the distances between member pairs. Applied to seismology, our approach allows earthquake location using the interevent distance between earthquake pairs, which can be estimated using only one or two seismic stations. We first validate the method with synthetic data that resemble common cluster shapes, and then test the method with two seismic sequences in California: the August 2014 $M_w$ 6.0 Napa earthquake and the July 2019 $M_w$ 6.4 Ridgecrest earthquake sequence. We demonstrate that our approach provides robust and reliable results even for a single station. When using two seismic stations, the results capture the same structures recovered with high-resolution double-difference locations based on multiple stations. The proposed method is particularly useful for poorly monitored areas, where only a limited number of stations are available.

**Key words:** Inverse theory; Earthquake source observation; Induced seismicity; Volcano seismology.

## 1 INTRODUCTION

Locating earthquakes is fundamental to observational seismology and despite being a very old problem, with the first methods introduced more than a century ago (Geiger 1912), the development of new techniques to improve the accuracy of seismic event locations is still an active topic of research. The exponential growth of computing power since the beginning of the digital era enables increasingly sophisticated techniques that improve our ability to locate earthquakes.

Earthquake location methods can be classified as absolute and relative methods. Absolute methods (e.g. Lomax *et al.* 2009) allow seismic event location with respect to a fixed (i.e. absolute) coordinate system. To obtain accurate locations requires a good knowledge of the velocity model and a sufficient number of observations (i.e. the arrival times of *P*, *S* or both seismic phases at different stations).

Absolute location approaches can be further extended using joint hypocentral and velocity inversion methods by locating groups of earthquakes using station corrections in a fixed velocity model (e.g. Douglas 1967; Pujol 1988, 2000) or by solving the inverse problem both for earthquake locations and velocity structure in 1-D or 3-D (e.g. Kissling *et al.* 1994). While absolute methods locate seismic events with respect to a fixed (i.e. absolute) coordinate system, relative methods locate seismic events with respect to the position of other seismic events (one or many), but they have the great advantage of being less dependent on the velocity model than are absolute location techniques (Waldhauser & Ellsworth 2000). These methods require a set of well-located events to bind the newly located events to an absolute reference frame. The master event technique introduced by Douglas (1967), the double-difference (DD) method (Waldhauser & Ellsworth 2000) and the location using the Source Specific Correction Term (Richards-Dinger & Shearer 2000) are

notable methods within this class. All these approaches, from the simplest to the most sophisticated, have poor performances when working with very sparse seismic networks and, in the worst case, become useless when only a single seismic station is available. In such cases, location of seismic events needs to be performed by using single station methods.

Classical single station location methods (e.g. Roberts *et al.* 1989; Saari 1991; Abercrombie 1995) are based on the analysis of three-component seismic recordings to estimate the backazimuth, incidence angle and source-station distance of a seismic event, thus providing its absolute location. Seismic event location using a single station is often a challenging task since the estimation of the backazimuth and the incidence angle is sensitive to noise and this lack of robustness makes difficult the development of automated methods. Promising results on single station location have been recently achieved by using deep learning approaches that have been applied to different seismic events across the globe and at different scales (Lomax *et al.* 2019; Mousavi & Beroza 2020).

In this study, we present a new relative single-station location method for clustered earthquakes that is suitable for the analysis of weak events, since it is not based on the backazimuth and incidence angle estimation. In addition, this approach does not require the use of a 3C station and locates seismicity clusters by using the interevent distances between each event pair of the cluster. Interevent distances can be estimated using the *P* and *S* arrival times that can be picked in single channel seismic recordings (Poupinet *et al.* 1984; Cauchie *et al.* 2020). Alternative methods for single station single channel interevent distance estimation based on coda wave interferometry have also been proposed (Snieder & Vrijlandt 2005; Rubinstein & Beroza 2007; Robinson *et al.* 2013; Zhao *et al.* 2017; Zhao & Curtis 2019). These methods are promising and can, in principle, lead to more accurate interevent distances; however, they require very high waveform similarity (i.e. very similar location and nearly identical focal mechanisms; Snieder & Vrijlandt 2005) a condition that is often not satisfied, hence we believe that the approach proposed in this paper (that only requires the arrival times of *P* and *S* phases) is suitable for a wide range of applications.

To locate seismic events using only the interevent distances between event pairs, we need to use distance geometry techniques. Distance geometry is a field of mathematics, introduced in the beginning of the 20th century by Karl Menger, consists of the characterization and study of sets of points based only on given values of the distances between member pairs (Liberti *et al.* 2014). A distance geometry problem (DGP) is one of finding the coordinates of a set of points by using the distances between some pairs of such points (Liberti *et al.* 2014). DGPs are applied to different scientific fields including biochemistry, where they are extensively used to determine protein structure based on nuclear magnetic resonance (NMR) spectroscopy data, which provide a set of interatomic distances for atoms pairs in a given protein (Souza *et al.* 2013). In this context, the so-called Molecular DGP (MDGP) consists of finding the position of the atoms of a molecule starting from a set of interatomic distances. Among the different methods used to solve the MDGP, the geometric build-up algorithms (Dong & Wu 2002, 2003; Wu & Wu 2007; Sit *et al.* 2009; Sit & Wu 2011) are the most popular. This class of algorithms allow determination of the atom coordinates in the molecule one atom at time by iteratively solving a system of equations (Voller & Wu 2013). There are different analogies between the protein structure determination problem and the location of clustered seismicity. The aim of this article is to apply distance geometry techniques to locate a group of seismic events by using their interevent distances estimated using the data available

at one or two seismic stations. In the following sections, we explain the methodology, test it on synthetic data sets, and apply it to two real data sets for the Napa and Ridgecrest seismic sequences in California.

## 2 METHODOLOGY

In this work we introduce a novel seismic event location method based on the molecular geometric build-up technique (Sit *et al.* 2009) adapted to solve the earthquake location problem. This approach requires as input data the interevent distance between earthquakes pairs within a cluster, that can be estimated from the *P* and *S* first arrival times $t_p$ and $t_s$, respectively. Our intent is to estimate interevent distances from source-station distances. Let us start with two seismic events *a* and *b* recorded at the station *s*, as illustrated in Fig. (1a). If the interevent distance $||\mathbf{r}_{ab}||$ between event *a* and event *b* is much smaller than the distance between these events and the station *s* then the interevent distance $||\mathbf{r}_{ab}||$ can be approximated by the following equation:

$$||\mathbf{r}_{ab}^s|| = |\ ||\mathbf{r}_{as}|| - ||\mathbf{r}_{bs}||\ |, \tag{1}$$

where $||\mathbf{r}_{ab}^s||$ is the approximated interevent distance and $||\mathbf{r}_{as}||$ and $||\mathbf{r}_{bs}||$ are the source-station distances of *a* and *b*, respectively, with respect to the station *s*. The interested reader can find the mathematical derivations of eq. (1) in the Appendix.

Source-station distances for the events *a* and *b* with respect to the station *s* can be estimated by using the *P* and *S* arrival times at the station *s*. If we consider a homogeneous half-space with seismic velocities $V_p$ and $V_s$ for *P* and *S* waves, respectively, then the source-station distance between the station *s* and the events *a* and *b* can be written as

$$||\mathbf{r}_{es}|| = k_v(t_s^e - t_p^e) \quad \text{with} \quad e = a, b \tag{2}$$

where $t_p^{a|b}$ and $t_s^{a|b}$ are respectively the first onset times of the *P* and *S* waves for the events *a* and *b* as recorded at the station *s* and $k_v = V_P V_S / (V_P - V_S)$. Finally, we can write the approximated interevent distance between *a* and *b* (eq. (1)) at the station *s* in the following way:

$$||\mathbf{r}_{ab}^s|| = k_v\ |\ (t_s^a - t_p^a) - (t_s^b - t_p^b)\ | \tag{3}$$

Eq. (3) can be used to estimate the approximated interevent distance $||\mathbf{r}_{ab}^s||$ by measuring the arrival times for both *P* and *S* waves of the events *a* and *b* recorded at the station *s*. From Fig. (1a), it is easy to see that $||\mathbf{r}_{ab}^s||$ is only an approximation of the true interevent distance $||\mathbf{r}_{ab}||$. If the interevent distance between the event *a* and the event *b* is much smaller than the distance between these events and the station *s*, then the value of the approximated interevent distance becomes very close to the true one. This condition is easily satisfied when the size of the seismicity cluster is much smaller than the distance between the station and the centre of mass of the cluster itself. It is possible to obtain a better approximation of the interevent distance by using two stations perpendicular to the barycentre of the cluster (Fig. 1b). In this case the main condition that must be satisfied is that both the stations and the events lie on the same plane. This condition can be approximated when the depth distribution of the seismicity cluster is small with respect to the distance between the cluster and the seismic stations. When such condition is satisfied a better approximation of the interevent distance between the event *a* and *b* can be obtained using the following equation:

$$||\mathbf{r}_{ab}|| \approx \sqrt{||\mathbf{r}_{ab}^{s1}||^2 + ||\mathbf{r}_{ab}^{s2}||^2}, \tag{4}$$

(a)



(b)
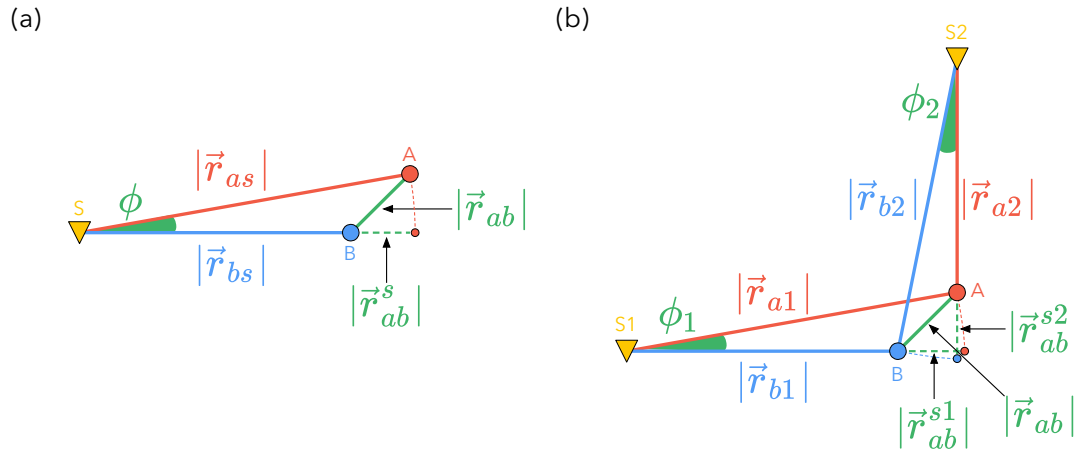


**Figure 1.** Geometric relations used for the calculation of the interevent distances between the events *a* and *b*, using the source-station distances at the station *s* (panel a) and the stations *s*1 and *s*2 (panel b) (in this second case stations and events are considered co-planar).

where $||\mathbf{r}_{ab}^{s1}||$ and $||\mathbf{r}_{ab}^{s2}||$ are the approximated interevent distances estimated using eq. (3) at the stations *s*1 and *s*2, respectively. In this case, we first estimate the interevent distance using eq. (3) separately at each station (*s*1 and *s*2) and then use eq. (4) to obtain a better approximation. To locate seismicity clusters we use a class of iterative methods used in biochemistry to solve the MDGP. More specifically we use the Geometric Build-Up algorithm first introduced by Dong & Wu (2002, 2003) and successively improved by Sit *et al.* (2009). The original method requires exact interatomic distances and suffers from error propagation after each iteration (Sit *et al.* 2009; Sit & Wu 2011; Voller & Wu 2013). This is due to the fact that the coordinates of the atoms are calculated by using the coordinates of previously determined atoms, and the errors in the previously determined atoms propagates to the later determined atoms. As a result, the coordinates for later determined atoms can be affected by large errors, especially when dealing with large molecules (Sit *et al.* 2009). In order to fix this problem an improved version of this algorithm has been proposed by Sit *et al.* (2009). This approach is robust against inexact interatomic distances and does not suffer of the error propagation problem, since the determination of the location of a particular atom do not depend on the coordinates of the previously determined atoms (Sit *et al.* 2009; Sit & Wu 2011). Here we use the improved version of geometric build-up algorithm proposed by Sit *et al.* (2009), adapted to solve the earthquake location problem and modified to obtain an estimation of the location uncertainties. The location method we propose requires (at least) only four non-coplanar seismic events for which the absolute hypocentral coordinates are known. We call these events *master events*. This is a mathematical requirement to ensure the uniqueness of the inverse problem. In order to describe how the method works let us consider a seismicity cluster consisting of $k + 1$ events with $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{k+1}$ the hypocentral coordinates (defined in $\mathbb{R}^3$) of each seismic event $1, 2, ..., k + 1$. If we know the Euclidian distance between each pair of events *i* and *j*, we can write

$$||\mathbf{r}_{ij}||^2 = ||\mathbf{x}_i - \mathbf{x}_j||^2 = ||\mathbf{x}_i||^2 - 2\mathbf{x}_i^T\mathbf{x}_j + ||\mathbf{x}_j||^2, \quad (5)$$

where $||\mathbf{r}_{ij}||$ is the interevent distance between *i* and *j*, and $||\mathbf{x}_i||$ and $||\mathbf{x}_j||$ are respectively the distances between the events *i* and *j* and the origin of the reference system (with $i, j = 1, 2, .., k + 1$). We can consider the internal structure of a seismicity cluster invariant to translation and rotation (i.e. we can translate and rotate it but the internal configuration of the events does not change), thus we can

arbitrarily consider the event $k + 1$ at the origin of the reference system (i.e. $\mathbf{x}_{k+1} = (0, 0, 0)$ ). In this case $||\mathbf{r}_{ik+1}||^2 = ||\mathbf{x}_i||^2$ and $||\mathbf{r}_{jk+1}||^2 = ||\mathbf{x}_j||^2$; thus, we can write the eq. (5) as

$$||\mathbf{r}_{ij}||^2 = ||\mathbf{r}_{ik+1}||^2 - 2\mathbf{x}_i^T\mathbf{x}_j + ||\mathbf{r}_{jk+1}||^2 \quad \text{with} \quad i, j = 1, 2, ..., k.$$
$$(6)$$

We define a $k \times 3$ matrix of coordinates $\mathbf{X}$ (i.e. the coordinates of hypocentres):

$$\mathbf{X} = \mathbf{x}_i \quad \text{with} \quad i = 1, 2, ..., k \quad (7)$$

and a $k \times k$ distance matrix $\mathbf{R}$:

$$\mathbf{R} = \left( ||\mathbf{r}_{ik+1}||^2 - ||\mathbf{r}_{ij}||^2 + ||\mathbf{r}_{jk+1}||^2 \right)/2 \quad \text{with} \quad i, j = 1, 2, ..., k.$$
$$(8)$$

The elements of the matrix $\mathbf{R}$ (i.e. the interevent distances between each pair of events) are estimated by using the approach described in the appendix. From eq. (6), it follows that

$$\mathbf{R} = \mathbf{X}\mathbf{X}^T \quad \text{in} \quad \mathbb{R}^{k \times k}, \quad (9)$$

where $\mathbf{X}^T$ is the transpose of $\mathbf{X}$. We can now write the singular value decomposition (SVD) of $\mathbf{R}$ as

$$\mathbf{R} = \mathbf{V}\mathbf{S}\mathbf{V}^T, \quad (10)$$

where $\mathbf{V}$ is an orthogonal matrix and $\mathbf{S}$ is the diagonal matrix of singular values of $\mathbf{R}$. Then, we can find $\mathbf{X}$ by solving

$$\mathbf{X} = \mathbf{G}\Sigma^{1/2} \quad (11)$$

with $\mathbf{G} = \mathbf{V}(1 : k, 1 : 3)$ and $\Sigma = \mathbf{S}(1 : 3, 1 : 3)$ (where the notation *l* : *m* means that we are extracting the elements of the matrix along one axis from the *l*th to the *m*th positions). The interested reader can find a more formal mathematical description in Havel (1998). Essentially, if we know all the distances between each events pair, we can calculate $\mathbf{R}$ and retrieve the coordinates of the *k* seismic events by finding $\mathbf{X}$, considering the event $k + 1$ fixed at the origin of the reference system. If we have *n* events within a cluster, we need to apply this approach iteratively. At each step *k*, the coordinates of the *k* (already located) events $\mathbf{X}$ are mapped (i.e. relocated) in a reference system with the $k + 1$ event is fixed at the origin (i.e. $\mathbf{x}_{k+1} = (0, 0, 0)$). This means that the first *k* events must be moved back to their initial positions by rotations and translations, so that the difference between the initial and the new locations for the first *k* events is minimized in a least-squares sense.

We describe here how to do it in more detail. Let $\mathbf{X}_A$ be the $k$ matrix with the original hypocentral coordinates of the first $k$ events, and let be $\mathbf{X}_B$ the $k$ matrix with the hypocentral coordinates of the events in the new reference frame where $\mathbf{x}_{k+1} = (0, 0, 0)$. We first need to translate $\mathbf{X}_B$ to $\mathbf{X}_A$ with a translation vector $\mathbf{x}_c^A - \mathbf{x}_c^B$, where $\mathbf{x}_c^A$ and $\mathbf{x}_c^B$ are the position of the geometric centres of clusters in the original and the new reference system respectively. After the translation operation, we need to perform a rotation by using the matrix $\mathbf{Q} = \mathbf{U}\mathbf{V}^T$ which is obtained from the SVD of $\mathbf{X}_B^T\mathbf{X}_A = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. If $\mathbf{x}_i^B$ are the coordinates of the seismic events $i = 1, 2, 3, ....k, k + 1$, then $\mathbf{Q}\mathbf{x}_i^B$ produces the location of the $i = 1, 2, 3, ....k, k + 1$ events in the original reference system, with the event $k + 1$ now located. Essentially the rotation matrix $\mathbf{Q}$ allows to minimize the Euclidian norm $||\mathbf{X}_A - \mathbf{X}_B\mathbf{Q}||$ (Golub & Van Loan 1989). Finally, it is possible to locate $n$ seismic events within a cluster, by starting with four non co-planar seismic events with known absolute locations and by iteratively applying this approach to the remaining $n - 4$ seismic events.

We perform uncertainty estimation following a bootstrap-like approach. It is well known that the largest source of error in earthquake location is often related to uncertainties in the velocity model. Location uncertainties are thus estimated by randomly perturbing the $V_P$ and $V_S$ (i.e. the $k_v$ in eq. 3) and relocating the same cluster $N_{boot}$ times. Then for each event we have a population of $N_{boot}$ relocations whose standard deviation can be used as a proxy of the uncertainty. Finally, after the resampling process, we apply a Gaussian kernel density estimator (Gaussian KDE) to the whole population of relocated events (i.e. for cluster of $n$ seismic events, the entire population after the resampling will be $nN_{boot}$) to have a visual proxy of locations stability and estimate the size of the cluster by taking into account location uncertainties and mislocated events (Lopez-Comino *et al.* 2017). In this context, the Gaussian KDE can be viewed as a powerful tool for the visualization of large seismicity data sets that allow to couple the main geometrical information related of a seismic cluster with location uncertainties (including systematically mislocated events).

# 3 APPLICATION TO SYNTHETIC DATA

We tested our location approach with synthetic data resembling different conditions, including the number of reference events, the shape of the cluster and the position of the seismic station used for the location process. We run two series of synthetic tests. In the first one we use, as input data, the Euclidian interevent distances directly calculated from the events coordinates (results shown in Figs 2 and 3), while in the second test we use approximated interevent distances (results shown in Figs 4–7) obtained by using eqs (3) and (4) for one station and two stations cases, respectively, as illustrated in Figs 1(a) and (b).

## 3.1 Synthetic test using Euclidian interevent distances

We focus the first test on the evaluation the effectiveness of the method in reconstructing a seismicity cluster by calculating the interevent distances directly from event coordinates. The synthetic data set consists of 200 events randomly distributed within a sphere of 500 m radius centred at the origin of the reference coordinate system. For each event pair, we first compute the Euclidian interevent distance and, then we add a random error assuming a uniform distribution between $-100$ and $100$ m. We additionally evaluate how

the number of reference events influences the quality of the reconstructed shape of the seismic cluster. We show location results for a spherical seismicity cluster using four master events in Figs 2 and 3; here, about 80 per cent of the estimated locations are within 75 m of the true locations (Fig. 2, in the lower right panel). The location error for the same fraction of events reduces to about 35 m, when using 8 events (Fig. 3) and does not improve significantly by using 16 master events (Supporting Information Fig. S1). This finding illustrates the relation between the number of reference events used to locate the seismic events, if homogeneously distributed around within the cluster, and the average location error.

## 3.2 Synthetic test using approximated interevent distances

A more realistic synthetic test requires the computation of the approximated interevent distances using eq. (1), which only needs the knowledge of the source-station distances of each event as illustrate in Fig. (1). While in real-world cases, source-station distances are estimated by using the *P* and *S* arrival times, in this test we calculate the source-station distances directly from the station and event coordinates. Theoretical source-station distances are then used to calculate the interevent distance of each event pair by using eq. (1) (and eq. 4 for the two stations applications). As described in the methodology section, the estimation of the interevent distance of an event pair based on the use of the source-station distances depends on the position of the station with respect to the seismic cluster and on the cluster shape.

The first synthetic data set consists of 200 events randomly distributed within a sphere of 500 m radius centred at $x = 0$, $y = 0$ and $z = 5$ km (z-axis is considered positive downwards). The interevent distances of each event pair are calculated by considering a surface ($z = 0$ km) seismic station located at an epicentral distance of 10 km from the origin of the reference system and $45°$ with respect to the *x*-axis. Prior to computation of the interevent distances, we perturbed each source-station distance with random errors (uniform distribution between $-100$ and $100$ m). We then performed the location of the synthetic events by using four and eight master events (Fig. 4). A first important finding from this test is that the location errors are about one order of magnitude larger than the ones obtained in the previous test, despite the same noise level. Location errors using four master events are within 800 m of the true locations for about 80 per cent of the events (Fig. 4a). These errors reduce to 600 m, when using eight master events (Fig. 4b). In order to test the sensitivity to the station azimuth, we applied our location method for the same data set (including the reference events) but now considering the station located at $-45°$ with respect to the *x*-axis (same epicentral distance). Location results (Supporting Information Fig. S2) are comparable to the previous results (shown in Fig. 4), indicating a negligible station azimuth effect. This is mainly due to the isotropy of a spherical shape of the cluster, which makes the results less sensitive to the azimuthal location of the station.

To investigate the effect of the station-cluster geometry, we generate an additional synthetic data set consisting of 200 events randomly distributed within a rectangular volume of size 800 m × 200 m × 200 m (along *x*, *y* and *z*, respectively) and centred at $x = 0$, $y = 0$ and $z = 5$ km (Fig. 5). Source-station distances are perturbed as before. We relocate the seismic cluster assuming two station locations. In the first case, we use an optimally oriented station (i.e. when the major elongation axis of the seismicity lies on line between the station and the centre of the cluster). Location results with four and eight master events (Fig. 5 in panels a and b, respectively) show
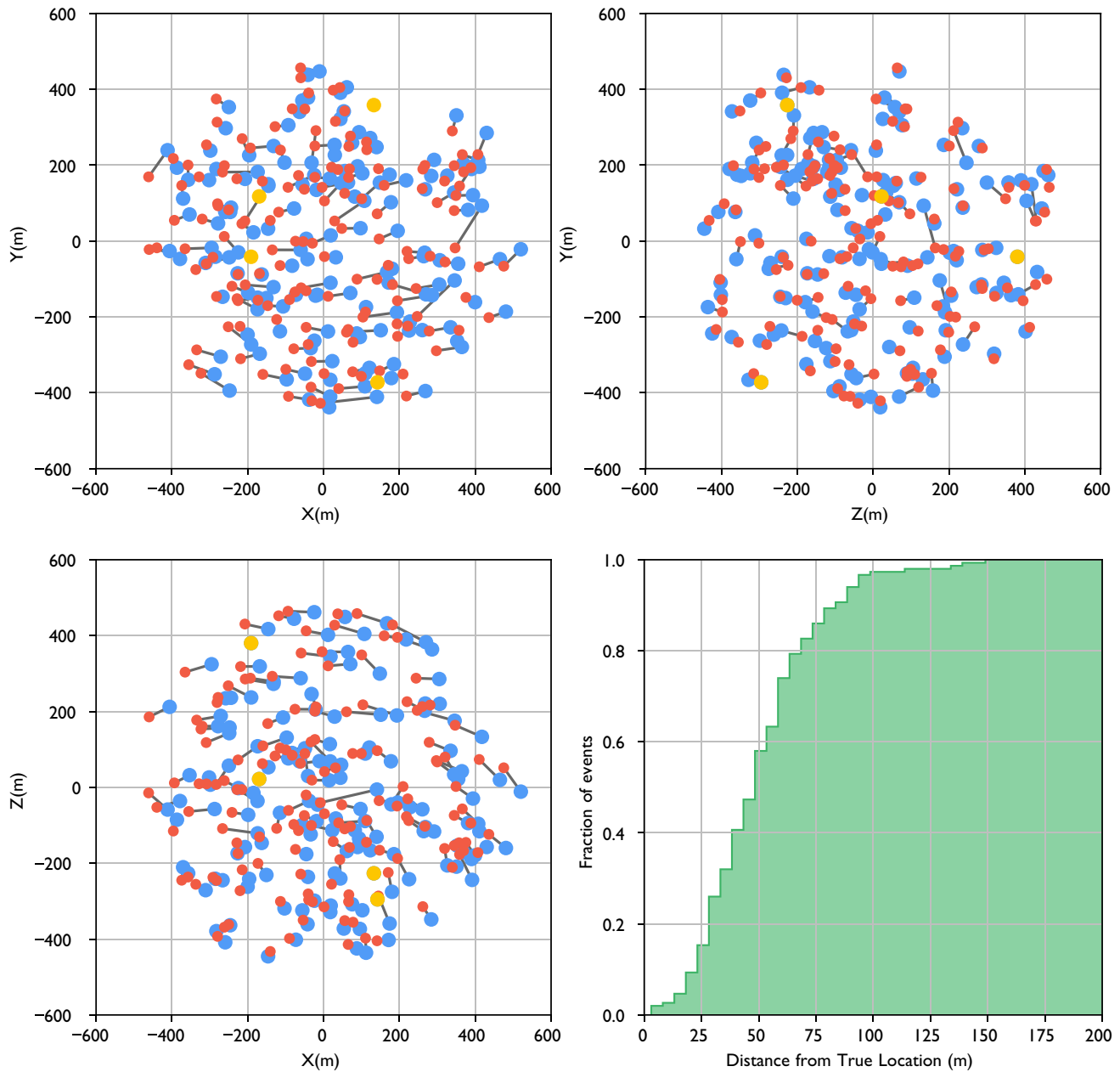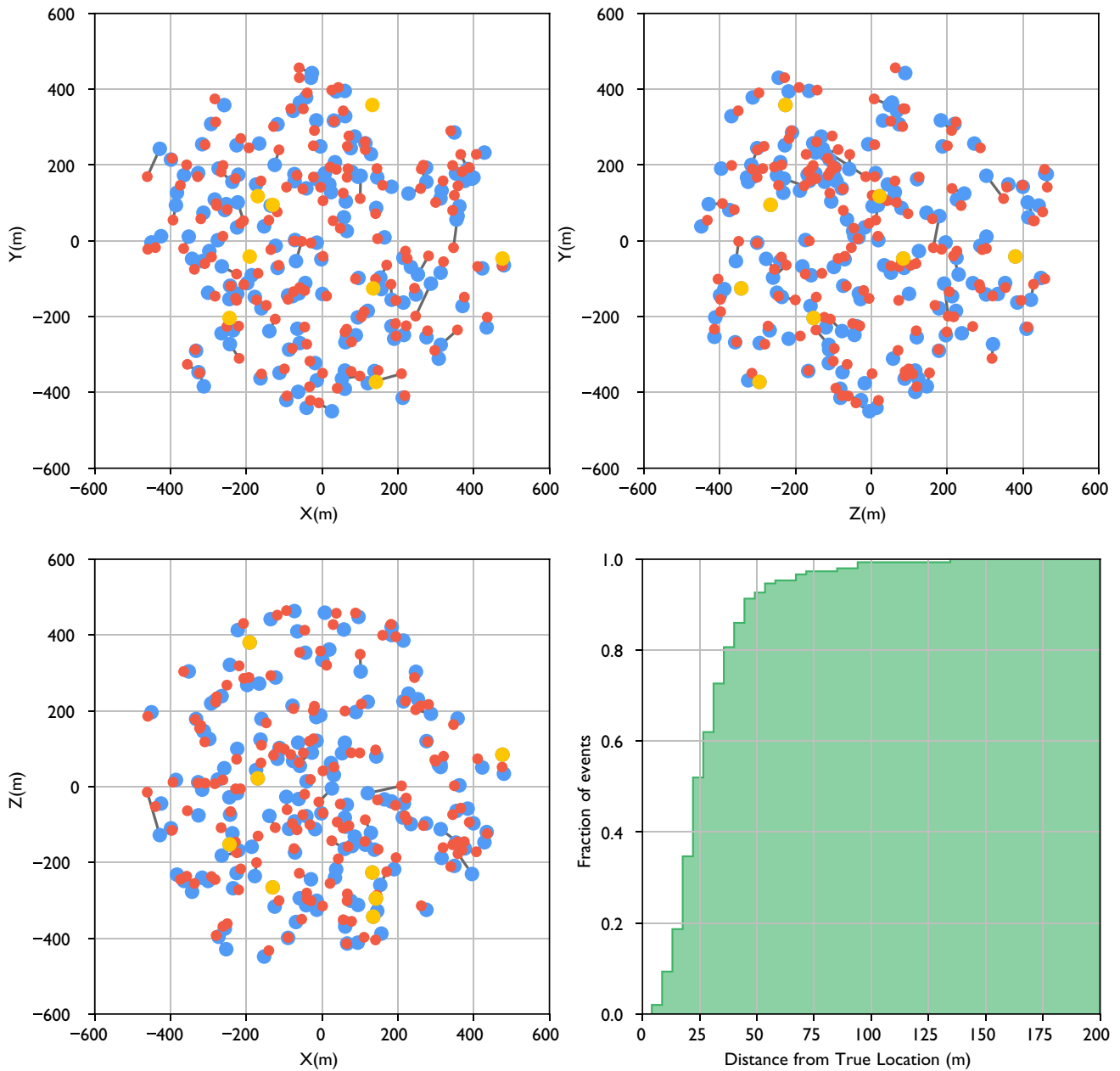
**Figure 2.** Synthetic test performed using four master events. Red dots represent true locations while blue dots represent the estimated ones using noise contaminated Euclidian interevent distances. In green, we show the cumulative plot showing the fraction of events located within a given (hypocentral) distance from the true event location.

similar location errors with about 80 per cent of the events located within 400 m of the true locations. In this case, our method is also able to reconstruct the main patterns of the cluster. We then test the worst possible seismic station orientation with respect to the cluster (i.e. when the major elongation axis of the seismicity is perpendicular to the line between the station and the centre of the cluster). In this case the location performance was strongly degraded and the algorithm was not able to correctly reconstruct the shape of the cluster (Fig. 6). Location results with four and eight master events (Fig. 6 in panels a and b, respectively) show larger location errors with about 80 per cent of the events located within 600 m from the true locations (with respect to 400 m from the previous test).

The location performance strongly depends on the accuracy of the estimated interevent distances of each event pair. When using

data from a single station, however, obtaining reliable interevent distances is challenging. Using two stations to estimate the interevent distance between two seismic events (eq. 4) leads to a significant improvement in individual event locations and reconstruction of the shape of the seismic cluster. We used the two data sets introduced in the previous examples (and shown in Figs 4 and 5) and tested the location algorithm using interevent distances estimated with two seismic stations and only four reference events. We also tested two different network geometries: in the first case, the two stations are located along the $x$ and $y$ axes (as shown in Fig. 7) while in the second case, the stations are rotated counter-clockwise by 45° (as shown in Supporting Information Fig. S3). It is interesting to note that, in this case, the quality of the location strongly improved, with the errors reduced by about 60–80 per cent with respect to the results

**Figure 3.** Synthetic test performed using eight master events. Red dots represent true locations while blue dots the estimated ones using noise contaminated Euclidian interevent distances. In green we show the cumulative plot showing the fraction of events located within a given (hypocentral) distance from the true event location.

obtained using a single station (Figs 4–6). The location of seismic clusters using two stations yields proper reconstruction of the shape of the cluster. For the spherical shaped cluster (Fig. 7a), we find that about 80 per cent of the events are located within 400 m from the true locations, while for the planar shaped cluster (Fig. 7b), this error reduces to 200–250 m. Additional tests considering different station azimuths led to similar results (as shown in Supporting Information Fig. S3).

# 4 APPLICATION TO REAL DATA

To validate our location method we tested it with real data. In particular, we applied it to data for the August 2014 $M_w$ 6.0 Napa

and the July 2019 $M_w$ 7.1 Ridgecrest earthquakes. Both seismic sequences occurred in California (USA).

## 4.1 The August 2014 $M_w$ 6.0 Napa (California, USA) seismic sequence

The first real data set we used for testing purposes is from the $M_w$ 6.0 South Napa earthquake sequence, which occurred on 2014 August 24, and was the largest earthquake to strike the greater San Francisco Bay area since the $M_w$ 6.9 1989 Loma Prieta earthquake (Brocher *et al.* 2015). The data set consists of 81 events ranging from magnitude 1.1 to 6.0 that occurred between August 24th and 26th; for comparison, we extracted locations from the DD (Waldhauser & Ellsworth 2000) catalogue from the Northern California Earthquake
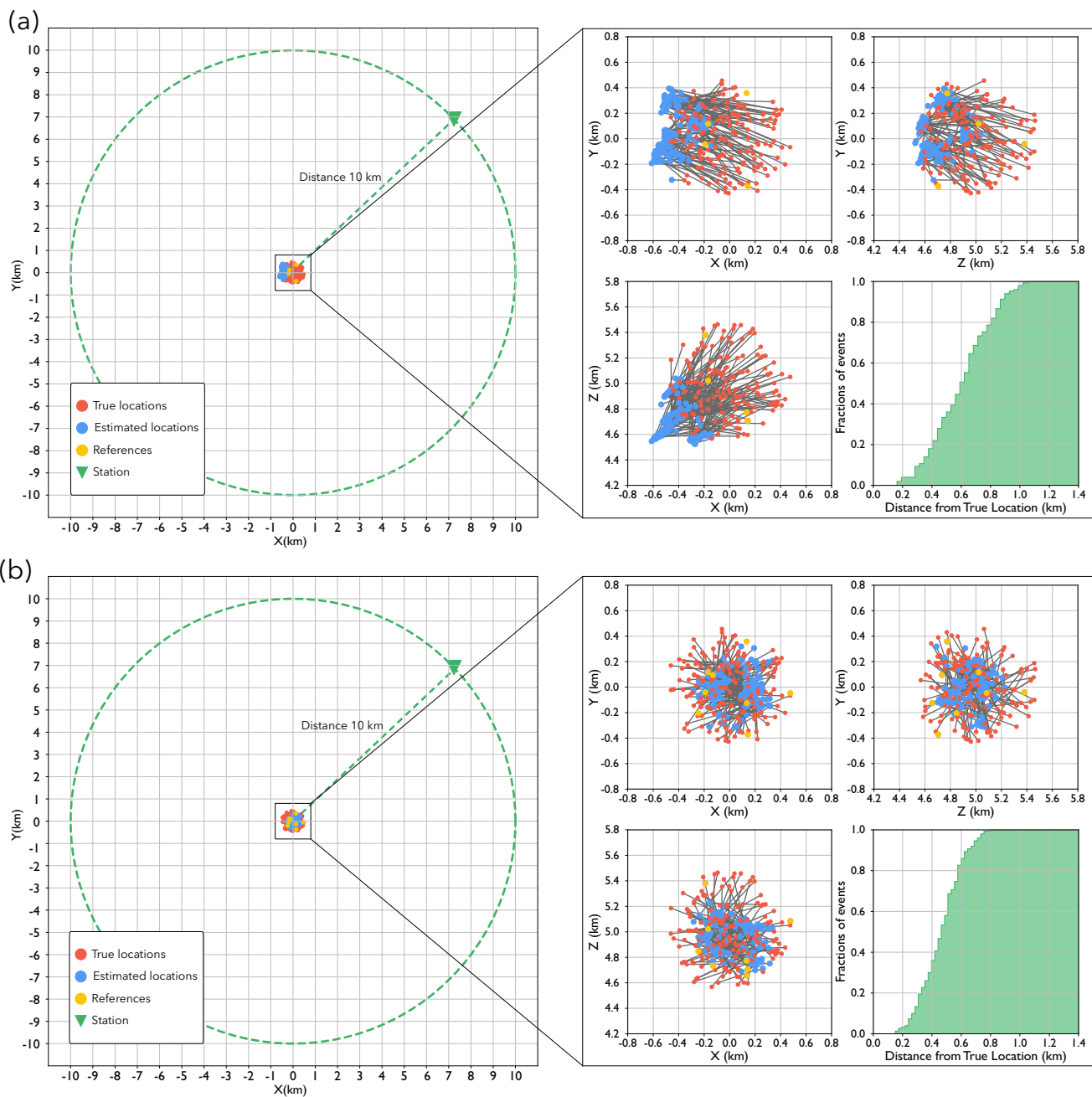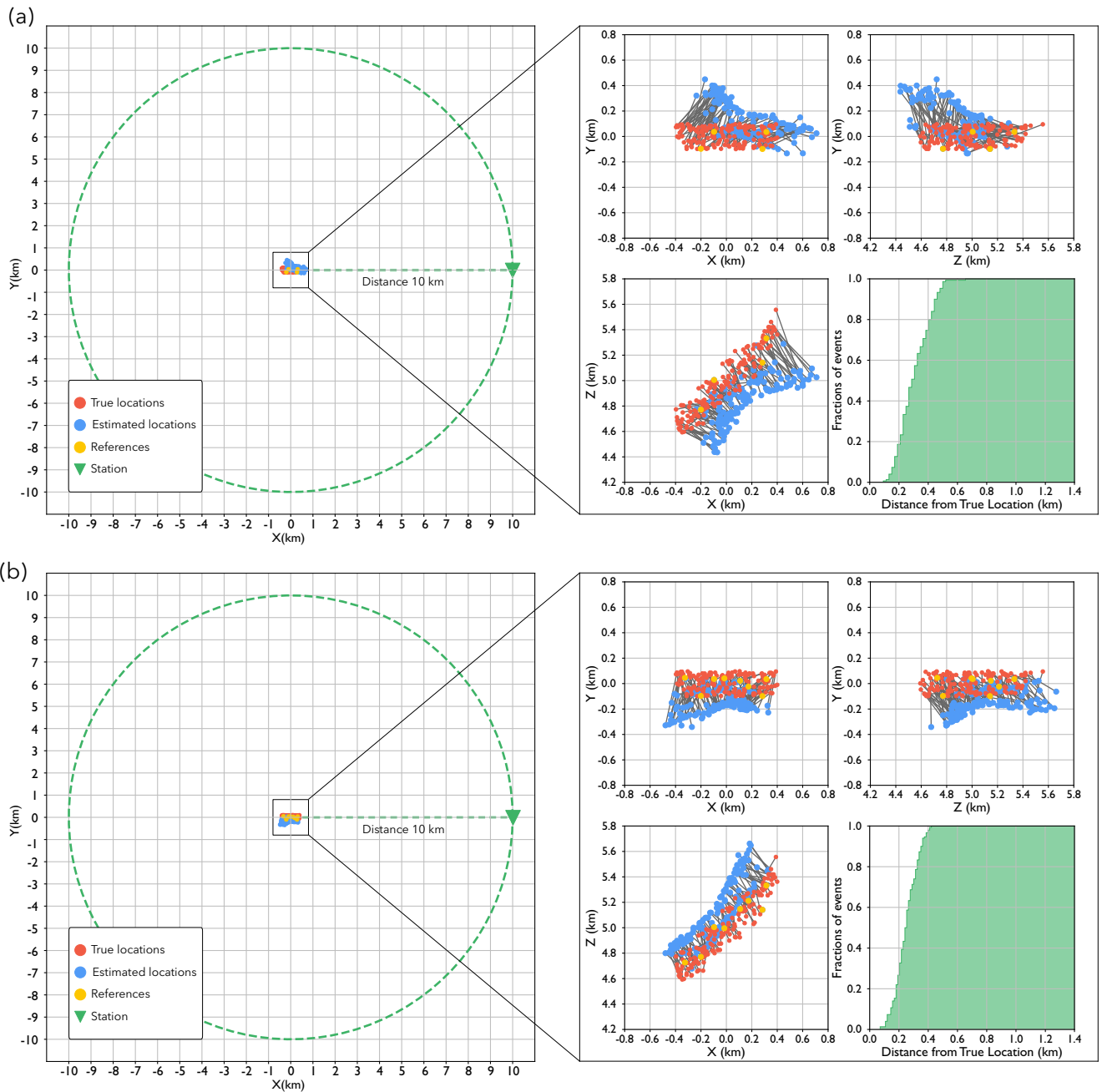
**Figure 4.** Synthetic test with a spherical shaped seismic cluster performed using 4 (panel a) and 8 (panel b) master events. The plots on the left show the position of the station (green triangle) with respect to the cluster. The right plots show a detail of the location results. Red dots represent true locations while blue dots the estimated ones using approximated interevent distances (eq. 1) contaminated with random noise. In green, we show the cumulative plot showing the fraction of events located within a given (hypocentral) distance from the true event location.

Data Center (Fig. 8). To calculate the interevent distances for each event pair we used manual *P* and *S* picks for a single station (station CVS, Fig. 8) and a homogeneous velocity model with $V_P = 6000$ m s$^{-1}$ and $V_S = V_P/\sqrt{3}$.

The Napa earthquake sequence exhibits an N–S elongation pattern (red dots in Fig. 9), while by using a single station approach we are not able to reproduce this pattern (blue dots in Fig. 9). Location results with four reference events show that about 80 per cent of the events are located within 4 km of the DD locations (Fig. 9). Although these results can be considered of good quality for a single

station method (Mousavi & Beroza 2020), in this case, our approach fails to fully reproduce the geometry of the seismicity cluster. Our results show that seismicity is apparently distributed along the E–W direction (blue dots in Fig. 9), while the cluster geometry estimated using the DD method shows an elongation in the N–S direction (red dots in Fig. 9). In these cases, however, additional source of information such as focal mechanisms, source directivity and knowledge of the faults distribution in the area could help to avoid wrong interpretations. Similar location results are obtained by using 8 and 16 master events (see Supporting Information Figs S4 and S5).
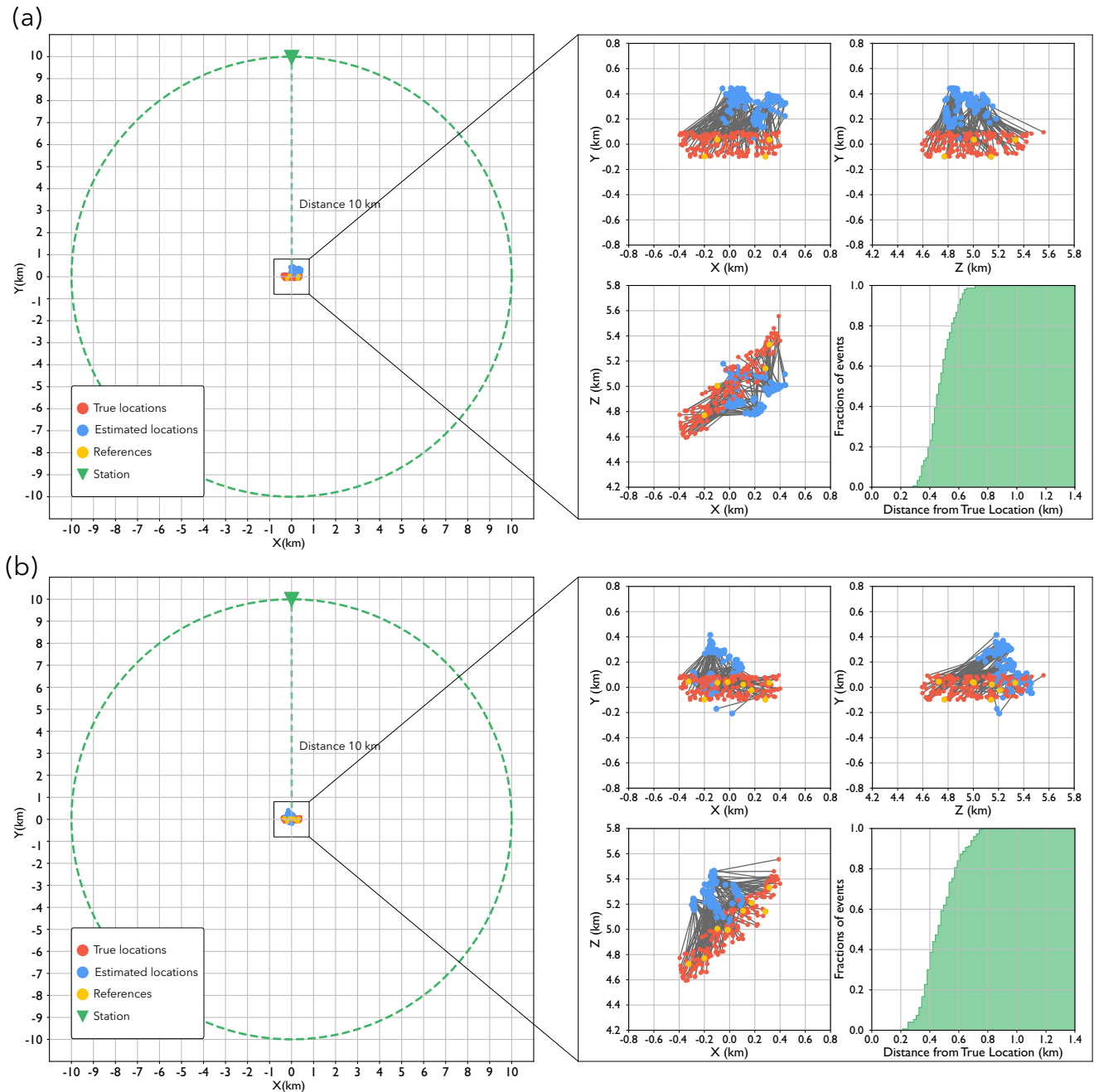
**Figure 5.** Synthetic test with a planar shaped seismic cluster performed using four (panel a), eight (panel b) master events and using a station aligned to the maximum elongation direction of the cluster. The plots on the left show the position of the station (green triangle) with respect to the cluster. The right plots show the details of the location results. Red dots represent true locations while blue dots represent the estimated ones using approximated interevent distances (eq. 1) contaminated with random noise. In green, we show the cumulative plot showing the fraction of events located within a given (hypocentral) distance from the true event location.

The locations obtained using 16 master events are less clustered if compared with the ones obtained using 4 and 8 master events. In this case, an increased number of master events did not lead to better location results. This may be related to the inclusion of badly located master events that have a pernicious effect in the location process.

We estimated uncertainties by perturbing the velocity model used for the interevent distance calculation within the range 5500–6500 m s$^{-1}$ for the $V_P$, while keeping the $V_P/V_S$ ratio fixed at $\sqrt{3}$. At each iteration (i.e. a perturbation in the velocity model), we recalculate

the interevent distances and relocate all the events. In this case, we relocate each event of the cluster 25 times and measure the location uncertainty as the sample standard deviation. We found that location uncertainties are in the order of 2 km. Finally, we plot all the location results for each iteration using the Gaussian KDE that allows visualization of the results, while taking into account location uncertainties and mislocated events (Lopez-Comino *et al.* 2017). The KDE results show that the cluster depth is well constrained (Fig. 10).

**Figure 6.** Synthetic test with a planar shaped seismic cluster performed using four (panel a), eight (panel b) master events and using a station perpendicular to the maximum elongation direction of the cluster. The plots on the left show the position of the station (green triangle) with respect to the cluster. The right plots show the details of the location results. Red dots represent true locations while blue dots the estimated ones using approximated interevent distances (eq. 1) contaminated with random noise. In green, we show the cumulative plot showing the fraction of events located within a given (hypocentral) distance from the true event location.

## 4.2 The July 2019 $M_w$ 6.4 Ridgecrest (California, USA) seismic sequence

The second real data set is for the July 2019 $M_w$ 6.4 Ridgecrest (California) seismic sequence, which occurred on 2019 July 4 and followed by an $M_w$ 7.1 earthquake on 2019 July 6 (Liu *et al.* 2020). The $M_w$ 6.4 event and its aftershocks activated an 'L' shape fault system (Ross *et al.* 2019). The particular shape of this cluster makes this data set a particularly compelling test of the ability of our method to reconstruct complex seismic sequences using one or two stations. We select a subset of 320 aftershocks of the $M_w$ 6.4 earthquake

from the DD catalogue generated by Liu *et al.* (2020) as our target events, and whose phases were picked by a deep-learning phase picker (Zhu & Beroza 2019). We selected only the events that have *P* and *S* picks at both the WMF and WBS stations (Figs 11). The magnitude ($M_l$) range is between 1.5 and 6.4. Interevent distances for each event pair of the cluster were calculated using a homogeneous velocity model with $V_P = 6000$ m s$^{-1}$ and $V_S = V_P/\sqrt{3}$. Due to the extension and geometric complexity of the seismic sequence, it is important to find a set of reference events that homogeneously sample the entire seismic cluster. In this application, we found this
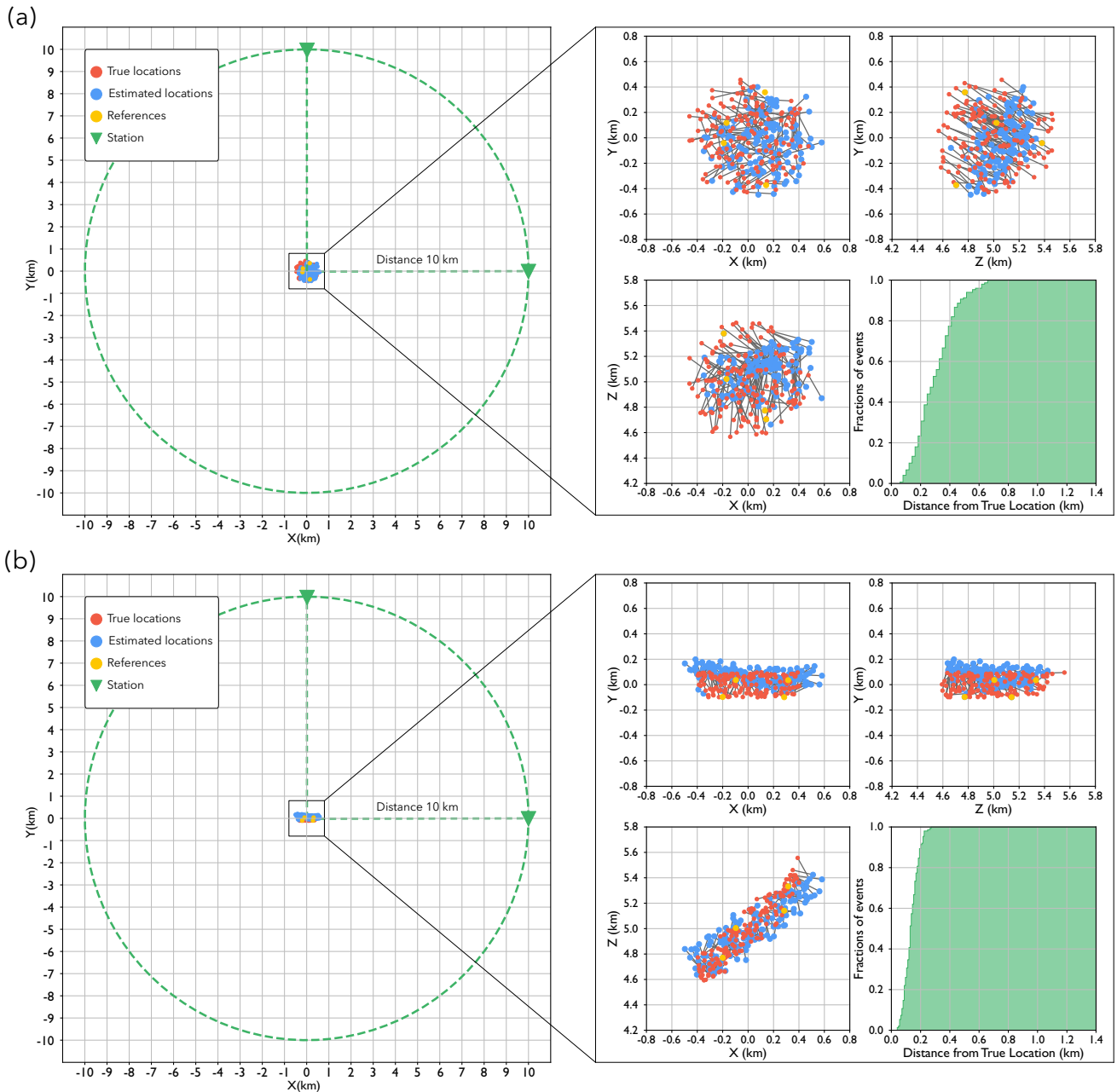
**Figure 7.** Synthetic test performed using four master events and two seismic stations with two different cluster shapes: spherical (panels a) and the locations of events located with the planar (panel b). The plots on the left show the position of the two stations (green triangles) with respect to the cluster. The right plots show a detail of the location results. Red dots represent true locations while blue dots represent the estimated ones using the approximated interevent distances based on two stations (eq. 4) and contaminated with random noise. In green, we show the cumulative plot showing the fraction of events located within a given (hypocentral) distance from the true event location.

condition satisfied by using 20 early aftershocks occurred in the first hour of the sequence (Fig. 13).

We first tested our methodology using a single seismic station located about 60 km from the centre of the cluster to the north–northeast (station WMS, Fig. 11) and then another station located at about 60 km from the centre of the cluster to the east-southeast (station WBF, Fig. 11). In both cases, about 90 per cent of the events are located within 15 km from the DD locations but these results show that the use of interevent distances estimated with a single station do not allow to reconstruct the shape of the cluster (Figs 12 and 13). Hypocentral depths are better constrained, however, and

distributed within 7.5 km from the average depth of the cluster (Figs 12 and 13). This sequence, due to its particular geometrical shape, represents the worst scenarios for the estimation of interevent distances with a single station (as explained in the methodological section).

We also tested the location performance using interevent distances estimated with two seismic stations and using eq. (4) (Fig. 14). In this last case, with about 80 per cent of the events located within 5 km of the DD locations, we achieved better performances and we successfully reproduced the shape of the cluster (Fig. 14). The use of two stations for the estimation of the interevent
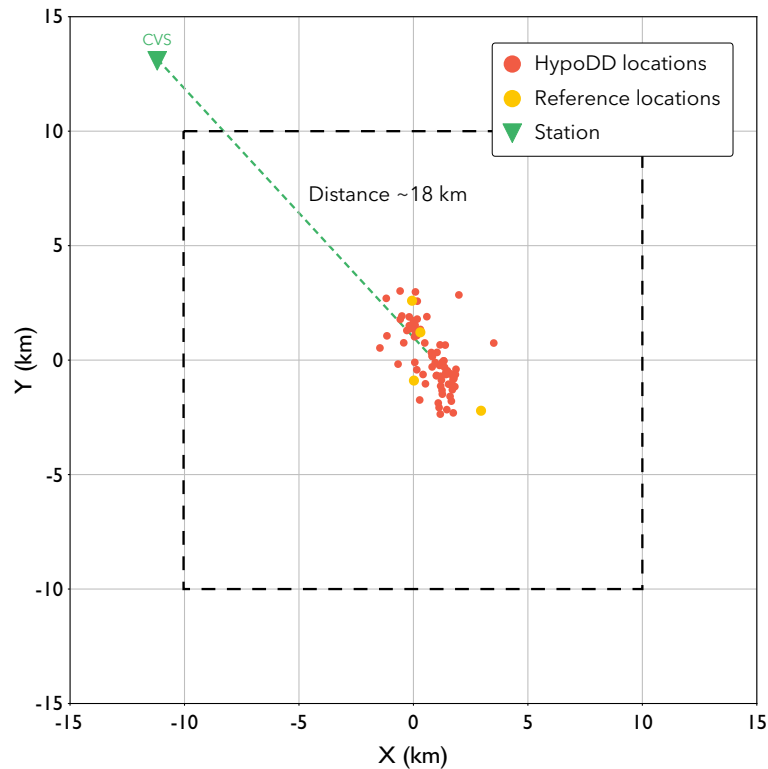
**Figure 8.** Napa earthquake sequence located using the DD method. The four master events used to locate the other events are represented by yellow dots. The green triangle at about 18 km distance from the centre of the cluster shows the position of the seismic station CVS, used to estimate the interevent distances. Location results presented in the next figures are referred to the region defined by the dashed square.

distances has also the important advantage to reduce the number of master events needed for the location of the seismic events. By using two stations for the location of the Ridgecrest earthquake sequence, the minimum number of master events required to correctly reproduce the shape of the cluster decreased to 8 (Supporting Information Fig. S6), although a not optimal distribution of master events with respect to the depth also reduced the accuracy in depth estimation (Supporting Information Fig. S6). However, if we increase the number of master events to 15 (Supporting Information Fig. S7), we obtain results which are very similar to those illustrated in Fig. 14, where we use 20 master events. These last results illustrate the potential of this location method in critical monitoring conditions when only one or two stations are available. In analogy with the application to the Napa sequence, to estimate the location uncertainties (for the two station case), we randomly perturbed (25 times) the velocity model within the range 5500–6500 m s$^{-1}$ for the $V_P$, and keeping fixed the $V_P/V_S$ ratio to $\sqrt{3}$. Estimated location uncertainties are in the range of 1–4 km for most the events. The KDE plot (Fig. 15) shows that solutions are very stable and the shape of the cluster is still clearly visible, which highlights the effectiveness of this location method even when dealing with very complex sequences.

## 5  DISCUSSION AND CONCLUSIONS

Earthquake location using a single seismic station or with a very sparse network is a challenging problem and remains an active topic of research. In this paper we introduced a novel technique based on distance geometry solvers that allow location of clustered seismicity using one or two seismic stations. Synthetic tests show that the

location results strongly depend on the accuracy of the interevent distances, this can be shown by comparing the location results obtained using the Euclidian interevent distances and the approximated ones (eq. 1). While the use of Euclidian interevent distances, even if contaminated with noise, allows a proper reconstruction of the shape of the cluster (Figs 2 and 3), the use of approximated interevent distances strongly reduces the performance of the location method (Figs 4 and 5).

These findings highlight the fact that the accuracy of the interevent distance estimations is one of the main factors controlling the performance of the location method. The use of two seismic stations at perpendicular azimuths to the direction of the cluster produces more accurate interevent distances (eq. 4, Fig. 1b). This explains why the location results obtained using two stations will always produce better results than single station locations, even if these ones are obtained with a larger number of well-located master events. Tests both with both synthetic (Fig. 7) and real (Fig. 14) data show that the use of interevent distances estimated with two stations strongly improves the recovery of the cluster geometry.

The performance of the location method also depends on the number and the quality of the master events used to locate the other events of the cluster. As a general rule, an increasing number of well-located master events improves location results (Figs 4 and 5); however, this is not always the case since the insertion of mislocated master events in an existing set of well-located master events does not improve location results and worse, reduces the overall quality of the results. The Napa earthquake sequence is an interesting case where we can observe this kind of problem. If we compare the locations obtained using 4 and 8 master events (Figs 9 and Supporting Information Fig. S4) with those obtained using 16 master events (see Supporting Information Fig. S5), we find that the increased number
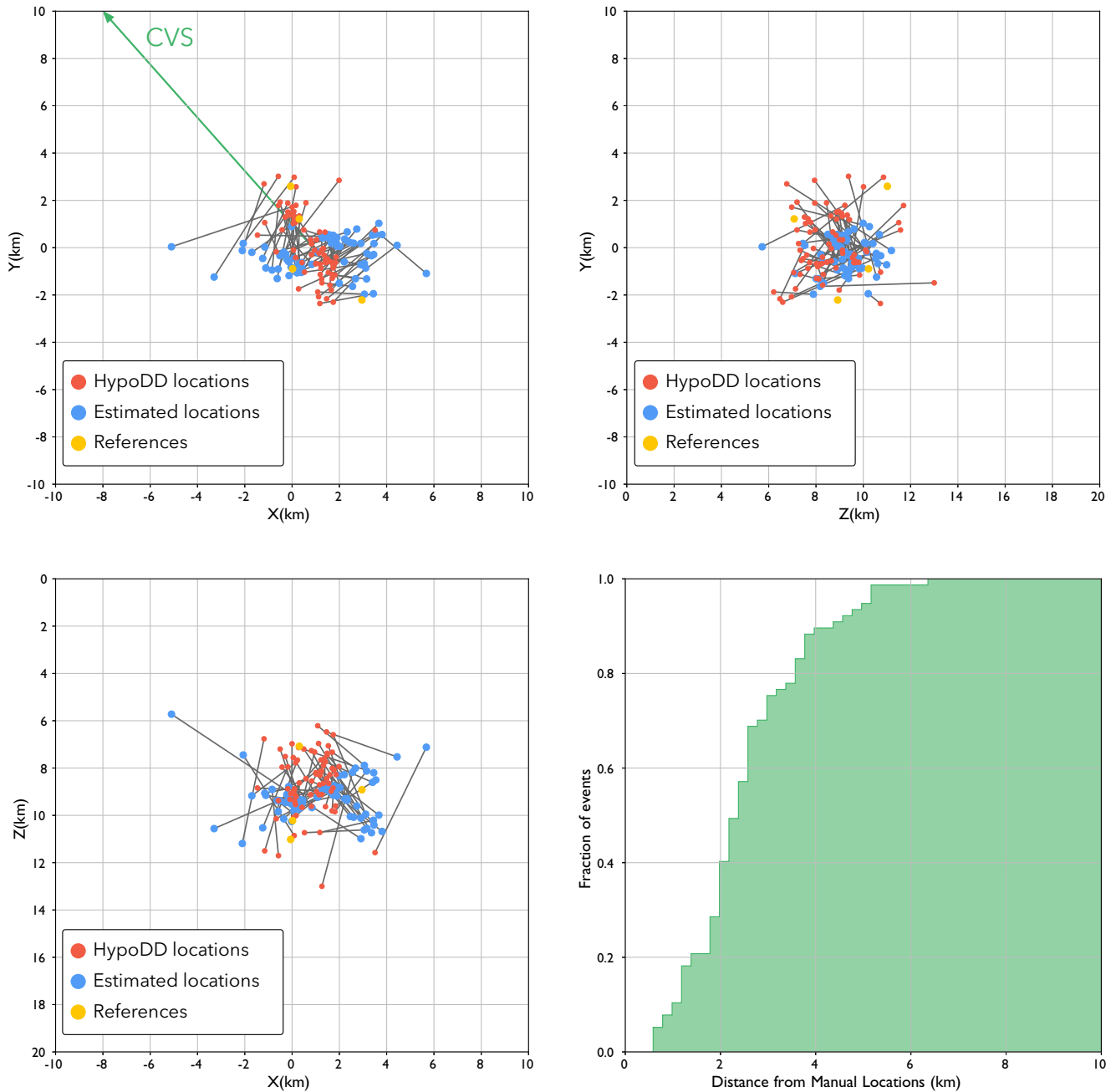
**Figure 9.** Location results of the Napa earthquake sequence obtained by using four master events (yellow dots) and the station CVS (direction indicated by the green arrow). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

of master events did not improve location results (i.e. locations are less clustered). This may be related to the inclusion of badly located master events having a pernicious effect on the location process.

The quality of location results also depends on the position of stations with respect to the cluster (e.g. compare Figs 5 and 6); however, this is not a limitation of the location method itself but is related to the interevent distance estimation procedure which strongly depends on the azimuth of the station with respect to the seismic cluster. For fixed offsets (i.e. source-station distances), the estimation of the interevent distances using a single station is more accurate when the major elongation axis of the seismicity lies along the same line between the station and the centroid of the cluster,

while the worst geometry occurs when they are perpendicular. For this reason, when dealing with complex seismic sequences, such as the Ridgecrest earthquake, the location performance strongly decreases and the shape of the cluster cannot be reconstructed properly (Figs 12 and 13). Due to its 'L' shape fault system, the Ridgecrest earthquake sequence represents a worst-case scenario for the application of this location approach using interevent distances estimated with a single station. As shown in Fig. 14, this limitation can be overcome by using (at least) two seismic stations that allow to estimate more accurate interevent distances.

The application to the Ridgecrest seismic sequence shows that the location performance strongly improves when using interevent
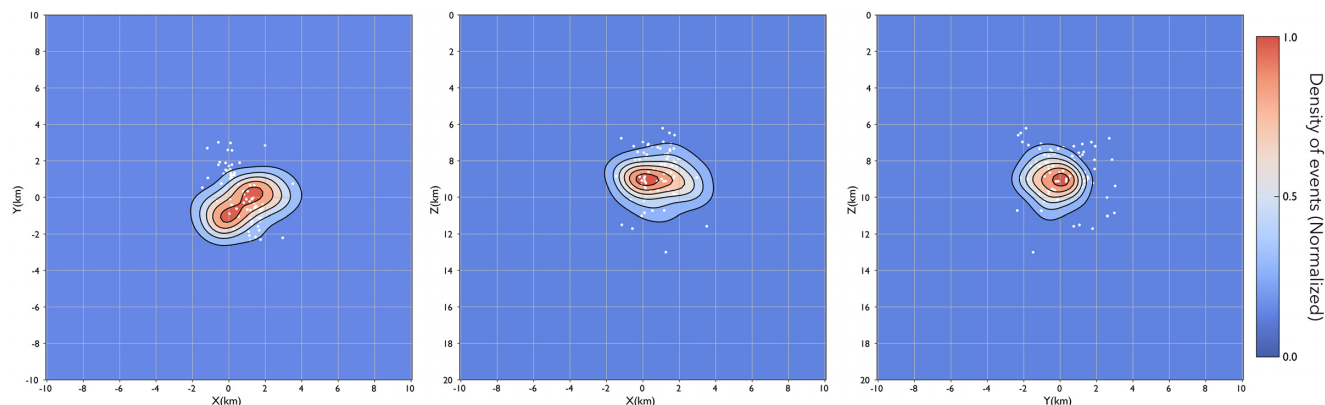
**Figure 10.** Gaussian KDE of the Napa earthquake sequence resampled 25 times by perturbing the velocity model within the range 5500–6500 m s$^{-1}$ for the $V_P$, and keeping the $V_P/V_S$ ratio fixed at $\sqrt{3}$. White dots denote the locations of events located with the DD technique. The colour scale (normalized to 1) moves from cyan to the red colour in the region containing the maximum density of samples.
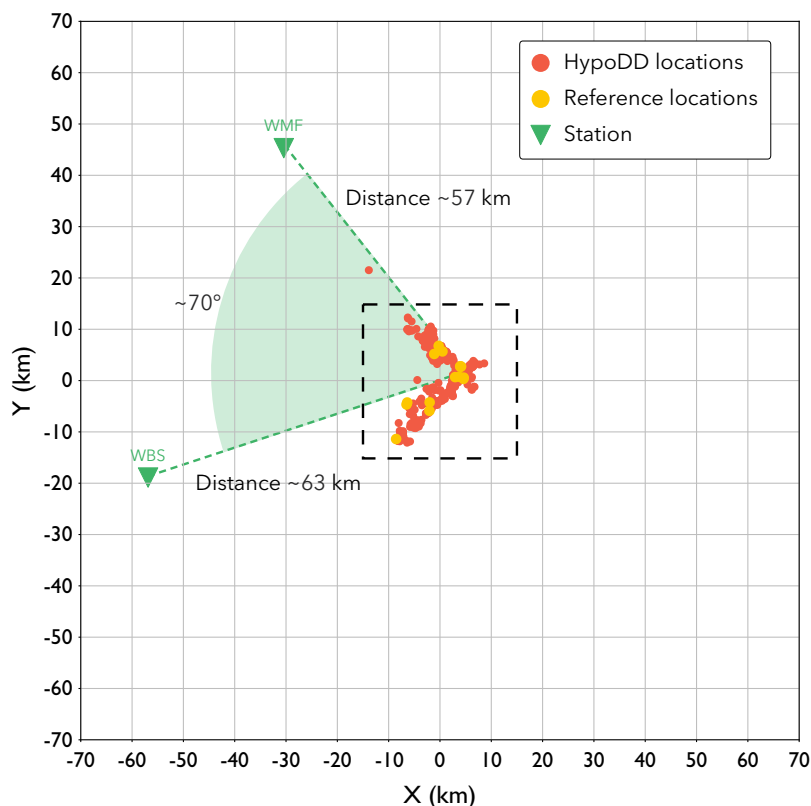


**Figure 11.** Ridgecrest earthquake sequence located using the DD method. Master events are represented by the 20 yellow dots. The green triangles at about 60 km distance from the centre of the cluster shows the position of the two seismic stations WMF and WBS, used for the location process. The two stations with respect to the centre of the cluster form an angle of about 70°. Location results presented in the next figures are referred to the region defined by the dashed square.

distances estimated with two stations (Fig. 14). Location results are robust (Fig. 15) and comparable with high-resolution DD locations obtained with a larger number of stations (more than 10); however, our relocated cluster shows a small distortion due to the fact that the two stations are not perfectly perpendicular with respect to the barycentre of the cluster or in other words, the angle between the lines connecting the two stations with the barycentre of the cluster is not 90° (Fig. 11). The applications to real data show that the algorithm is able to constrain the depth of the sequence, which is often a critical consideration for interpretation, and is a

daunting challenge when locating events with a one or two stations. However, this result strongly depends both on the location accuracy of the master events and on their spatial distribution. An illustrative example is provided by the application to the Ridgecrest sequence relocated by using eight master events (Supporting Information Fig. S6). In this case, the selected master events are mainly distributed in the depth range 6–10 km and, for this reason, although the cluster shape is properly reconstructed in terms of epicentres, the shallower events (i.e. with depth <5 km) are systematically located at a larger depth (with a shift of the order of 5–10 km). This result explains
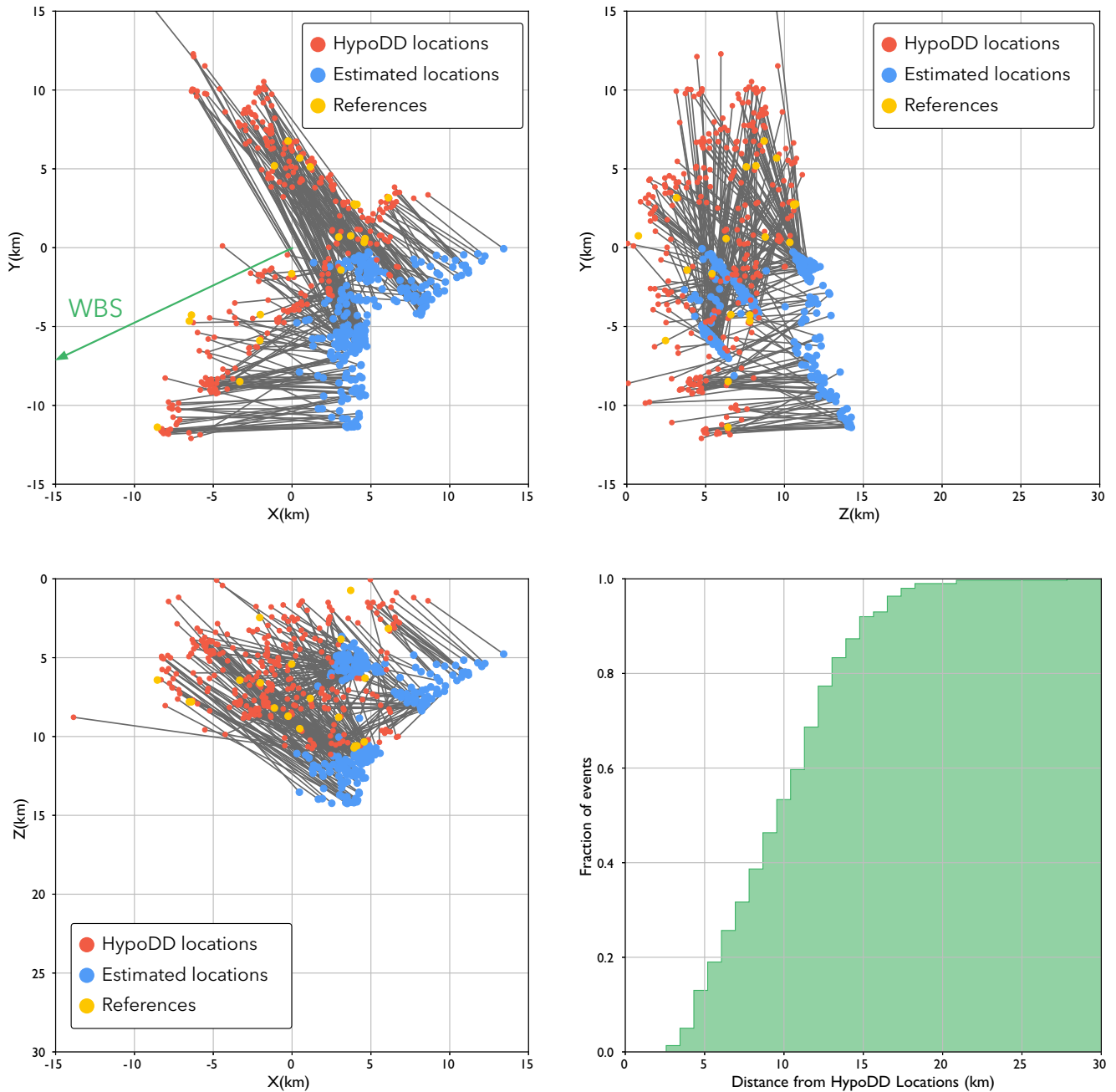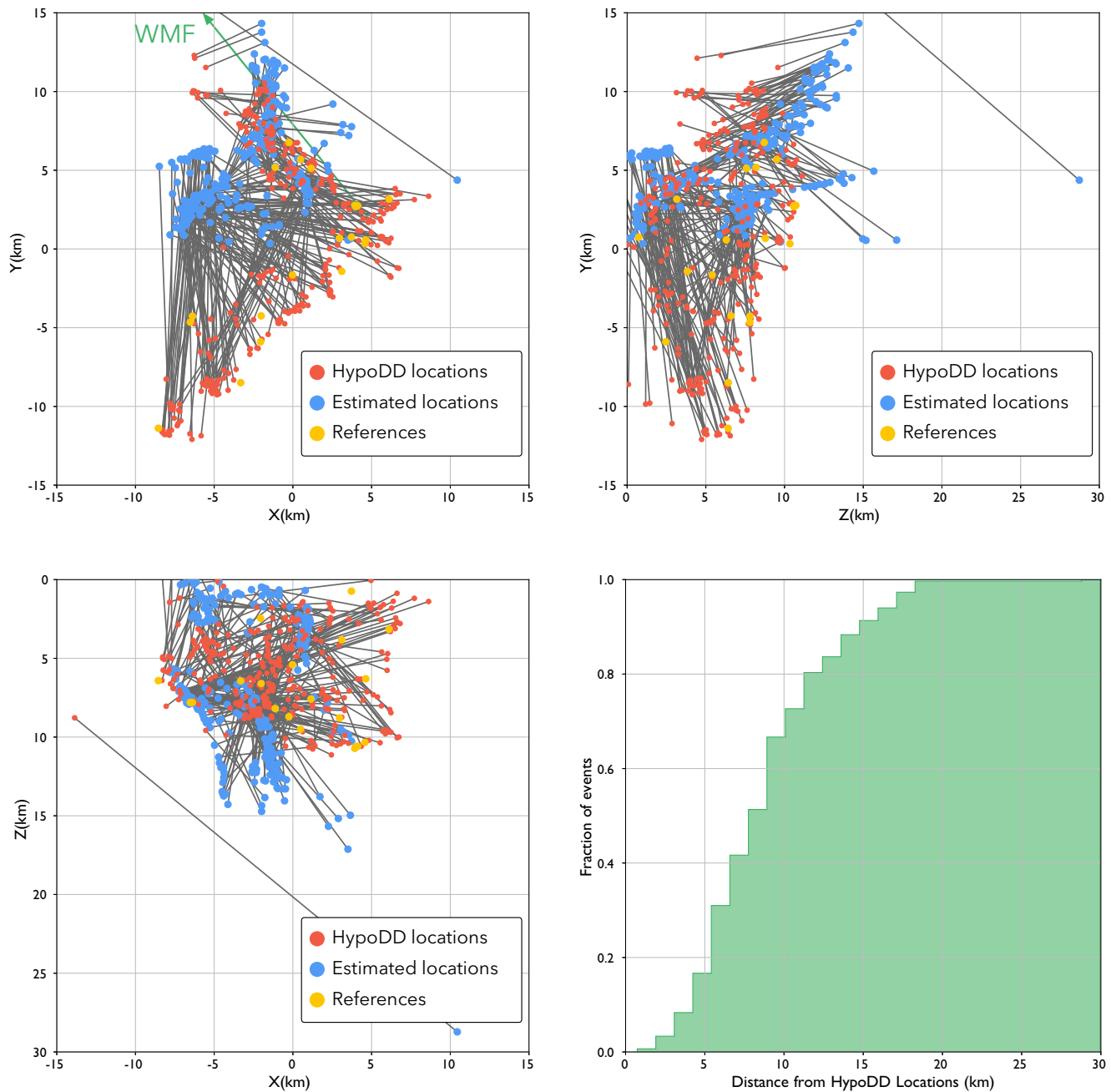
**Figure 12.** Location results of the Ridgecrest earthquake sequence obtained by using 20 master events (yellow dots) and the station WBS (direction indicated by the green arrow). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

the need for a larger number of master events in this application and raises the question of how to select optimal master events to optimize location accuracy.

The main condition required by the location method is the availability of at least four non-coplanar master events, which is one of the mathematical requirements to achieve a unique solution to the location problem. In principle, any accurately located event can be used as master, however, in order to obtain reliable location results, master events should have the following characteristics: (1) sample different depths, (2) have a spatial homogeneous distribution within the cluster and (3) have location uncertainties much smaller than

the expected size of the cluster. In the case of $N > 4$ potential master events, then a possible way to choose the optimal set of events is to relocate the target cluster with all possible combinations of references and choose the combination leading to the most stable solution in terms of uncertainty and/or stability of results (e.g. the result of the bootstrap analysis).

Among the different applications in which this location approach could be particularly useful, induced seismicity applications may be the most relevant. In such applications, the estimation of reliable depths and tracking seismicity evolution are extremely important to ensure the safety of industrial operations. For similar reasons this method could be useful in monitoring operations with ocean

**Figure 13.** Location results of the Ridgecrest earthquake sequence obtained by using 20 master events (yellow dots) and the station WMF (direction indicated by the green arrow). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

bottom seismometers, in volcano monitoring operations and, more in general, in seismicity monitoring of remote areas, since in all these applications we often deal with very sparse seismic monitoring networks. In this context, the new location method proposed in this paper complements traditional location approaches, especially when working with unfavourable monitoring conditions. From a computational point of view our location approach allows location of about 300 events in less than a minute with a laptop computer. The computation time, however, strongly depends on the size of the distance matrix and, for large data sets (i.e. more than 1000 events), a parallelization of the algorithm (including the SVD) will be required.

We conclude that the main limitations of the method are: (1) the need for high-quality master events with good absolute locations that are used to locate the other events in the cluster (but this is also valid for many other relative location methods; (2) the availability of *P* and *S* first arrival times picks that, for very small events, may not be easy to pick (but this is in common with other single station location methods); and (3) the need for accurate estimation of interevent distances, especially when working in single-station mode. Concerning this last point, a higher accuracy in interevent distance estimation could be achieved by combining both arrival time information (i.e., *P* and *S* arrival times) with waveform parameters, such as the maximum value of the cross-correlation coefficient between
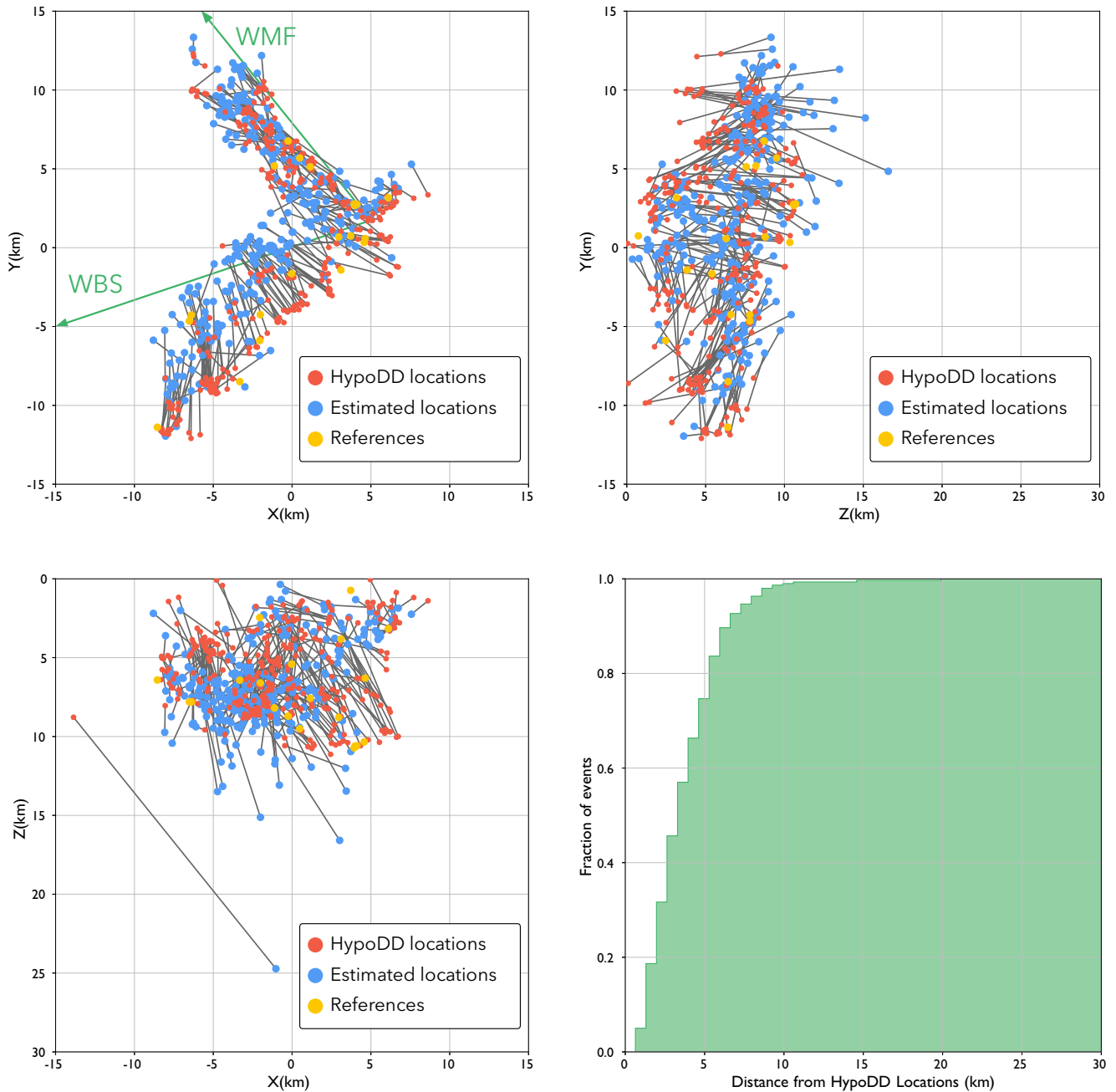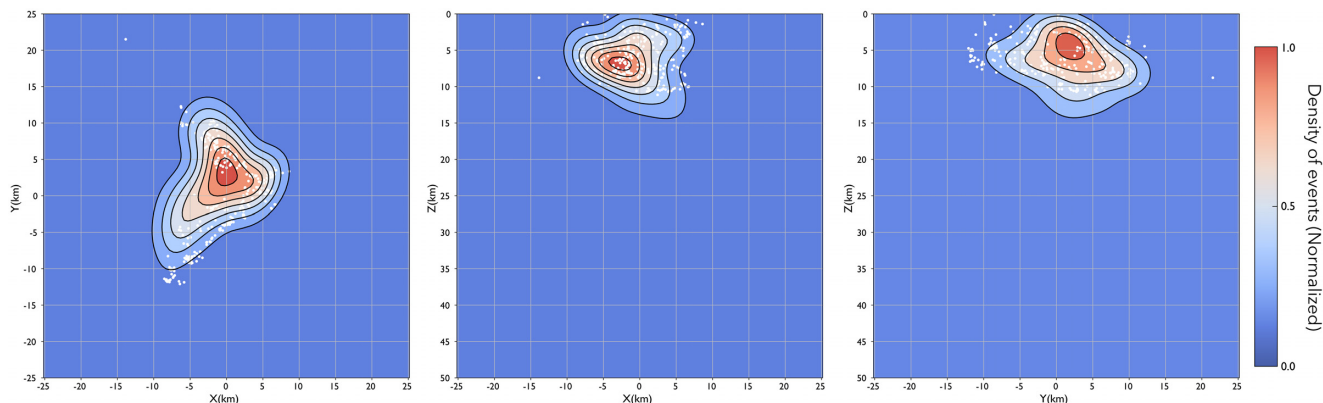
**Figure 14.** Location results of the Ridgecrest earthquake sequence obtained by using 20 master events (yellow dots) and two stations, WMF and WBF (directions indicated by the green arrows). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

waveforms of each event pair (Geller & Mueller 1980; Menke 1999) or cross-correlation of coda waves (Snieder & Vrijlandt 2005; Rubinstein & Beroza 2007; Robinson *et al.* 2013; Zhao *et al.* 2017; Zhao & Curtis 2019). Further investigation on the performance of these methods is, however, outside the scope of this work.

**Figure 15.** Gaussian kernel density estimation of the Ridgecrest earthquake sequence resampled 25 times by perturbing the velocity model within the range 5500–6500 m s$^{-1}$ for the $V_P$, and keeping the $V_P/V_S$ ratio fixed at $\sqrt{3}$. White dots denote the locations of events located with the DD technique. The colour scale (normalized to 1) moves from cyan to the red colour in the region containing the maximum density of samples.

paper have been generated using the python library MATPLOTLIB (Hunter 2007).

## REFERENCES

Abercrombie, R.E., 1995. Earthquake locations using single-station deep borehole recordings: implications for microseismicity on the San Andreas fault in southern California, *J. geophys. Res.,* **100**(B12), 24 003–24 014.

Brocher, T.M. *et al.*, 2015. The $M_w$ 6.0 24 August 2014 South Napa earthquake, *Seismol. Res. Lett.,* **86**(2A), 309–326.

Cauchie, L., Lengline, O. & Schmittbuhl, J., 2020. Seismic asperity size evolution during fluid injection: case study of the 1993 Soultz-sous-Forêts injection, *Geophys. J. Int.,* **221**(2), 968–980.

Dong, Q. & Wu, Z., 2002. A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances, *J. Glob. Optim.,* **22**(1–4), 365–375.

Dong, Q. & Wu, Z., 2003. A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data, *J. Glob. Optim.,* **26**(3), 321–333.

Douglas, A., 1967. Joint epicentre determination, *Nature,* **215**(5096), 47–48.

Geiger, L., 1912. Probability method for the determination of earthquake epicenters from the arrival time only, *Bull. St. Louis Univ.,* **8**(1), 56–71.

Geller, R.J. & Mueller, C.S., 1980. Four similar earthquakes in central California, *Geophys. Res. Lett.,* **7**(10), 821–824.

Golub, G.H. & Van Loan, C.F., 1989. *Matrix Computations,* The Johns Hopkins University Press.

Havel, T.F., 1998. Distance geometry: theory, algorithms, and chemical applications , in *Encyclopedia of Computational Chemistry,* Vol. 120, pp. 723–742, eds von Ragué Schleyer P., Allinger N.L., Clark T., Gasteiger J., Kollman P.A., Schaefer H.F., Schreiner P.R., doi:10.1002/0470845015.cda018.

Hunter, J., 2007. Matplotlib: a 2D graphics environment, *Comput. Sci. Eng.,* **9**, 90–95.

Kissling, E., Ellsworth, W.L., Eberhart-Phillips, D. & Kradolfer, U., 1994. Initial reference models in local earthquake tomography, *J. geophys. Res.,* **99**(B10), 19 635–19 646.

Liberti, L., Lavor, C., Maculan, N. & Mucherino, A., 2014. Euclidean distance geometry and applications, *SIAM Rev.,* **56**(1), 3–69.

Liu, M., Zhang, M., Zhu, W., Ellsworth, W.L. & Li, H., 2020. Rapid characterization of the July 2019 Ridgecrest, California earthquake sequence from raw seismic data using machine learning phase picker, *Geophys. Res. Lett.,* **42**, e2019GL086189.

Lomax, A., Michelini, A. & Curtis, A., 2009. Earthquake location, direct, global-search methods, in *Encyclopedia of Complexity and Systems Science,* Vol. 5, pp. 2449–2473, ed. Meyers, R., Springer.

Lomax, A., Michelini, A. & Jozinovic, D., 2019. An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network, *Seismol. Res. Lett.,* **90**(2A), 517–529.

Lopez-Comino, J.A., Cesca, S., Heimann, S., Grigoli, F., Milkereit, C., Dahm, T. & Zang, A., 2017. Characterization of hydraulic fractures growth during the Aspo Hard Rock Laboratory experiment (Sweden), *Rock Mech. Rock Eng.,* **50**(11), 2985–3001.

Menke, W., 1999. Using waveform similarity to constrain earthquake locations, *Bull. seism. Soc. Am.,* **89**(4), 1143–1146.

Mousavi, S.M. & Beroza, G.C., 2020. Bayesian-deep-learning estimation of earthquake location from single-station observations, *IEEE Trans. Geosci. Remote Sens.,* **58**, 8211–8224.

Poupinet, G., Ellsworth, W.L. & Frechet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California, *J. geophys. Res.,* **89**(B7), 5719–5731.

Pujol, J., 1988. Comments on the joint determination of hypocenters and station corrections, *Bull. seism. Soc. Am.,* **78**(3), 1179–1189.

Pujol, J., 2000. Joint event location—the JHD technique and applications to data from local seismic networks, in *Advances in seismic event location,* pp. 163–204, eds Thurber C.H., Rabinowitz N., Springer.

Richards-Dinger, K.B. & Shearer, P.M., 2000. *Earthquake locations in southern California obtained using source-specific station terms*, *J. geophys. Res.,* **105**(B5), 10 939–10 960.

Roberts, R.G., Christoffersson, A. & Cassidy, F., 1989. Real-time event detection, phase identification and source location estimation using single station three-component seismic data, *Geophys. J. Int.,* **97**(3), 471–480.

Robinson, D.J., Sambridge, M., Snieder, R. & Hauser, J., 2013. Relocating a cluster of earthquakes using a single seismic station, *Bull. seism. Soc. Am.,* **103**(6), 3057–3072.

Ross, Z.E. *et al.*, 2019. Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence, *Science,* **366**(6463), 346–351.

Rubinstein, J.L. & Beroza, G.C., 2007. Full waveform earthquake location: application to seismic streaks on the Calaveras Fault, California, *J. geophys. Res.,* **112**(B5), doi:10.1029/2006JB004463.

Saari, J., 1991. Automated phase picker and source location algorithm for local distances using a single three-component seismic station, *Tectonophysics,* **189**(1–4), 307–315.

Sit, A. & Wu, Z., 2011. Solving a generalized distance geometry problem for protein structure determination, *Bull. Math. Biol.,* **73**(12), 2809–2836.

Sit, A., Wu, Z. & Yuan, Y., 2009. A geometric buildup algorithm for the solution of the distance geometry problem using least-squares approximation, *Bull. Math. Biol.,* **71**(8), 1914–1933.

Snieder, R. & Vrijlandt, M., 2005. Constraining the source separation with coda wave interferometry: Theory and application to earthquake doublets in the Hayward fault, California, *J. geophys. Res.,* **110**(B4), doi:10.1029/2004JB003317.

Souza, M., Lavor, C., Muritiba, A. & Maculan, N., 2013. Solving the molecular distance geometry problem with inaccurate distance data, *BMC Bioinformatics,* **14**(S9), S7, doi:10.1186/1471-2105-14-S9-S7.

Voller, Z. & Wu, Z., 2013. Distance geometry methods for protein structure determination, in *Distance Geometry,* pp. 139–159, eds Mucherino, A., Lavor, C., Liberti, L. & Maculan, N., Springer.

Waldhauser, F. & Ellsworth, W.L., 2000. A double-difference earthquake location algorithm: method and application to the northern Hayward fault, California, *Bull. seism. Soc. Am.,* **90**(6), 1353–1368.

Wu, D. & Wu, Z., 2007. An updated geometric build-up algorithm for solving the molecular distance geometry problems with sparse distance data, *J. Glob. Optim.,* **37**(4), 661–673.

Zhao, Y. & Curtis, A., 2019. Relative source location using coda-wave interferometry: method, code package, and application to mining-induced earthquakes, *Geophysics,* **84**(3), F73–F84.

Zhao, Y., Curtis, A. & Baptie, B., 2017. Locating microseismic sources with a single seismometer channel using coda wave interferometry, *Geophysics,* **82**(3), A19–A24.

Zhu, W. & Beroza, G.C., 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.,* **216**(1), 261–273.

## SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

**Figure S1**. Synthetic test performed using 16 master events. Red dots represent true locations while blue dots represent the estimated ones using noise contaminated Euclidian interevent distances. In green, we show the cumulative plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the true event location.

**Figure S2**. Synthetic test with a spherical shaped seismic cluster performed using four (panel a) and eight (panel b) master events. The plots on the left show the position of the station (green triangle) with respect to the cluster. The right plots show the details of the location results. Red dots represent true locations while blue dots represent the estimated ones using approximated interevent distances contaminated with random noise. In green, we show the cumulative plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the true event location.

**Figure S3**. Synthetic test performed using four master events and two seismic stations with two different cluster shapes, spherical (panel a) and planar (panel b). The plots on the left show the position of the two stations (green triangles) with respect to the cluster. The right plots show the details of the location results. Red dots represent true locations while blue dots represent the estimated ones using the approximated interevent distances based on two stations and contaminated with random noise. In green, we show the cumulative plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the true event location.

**Figure S4**. Location results of the Napa earthquake sequence obtained by using eight master events (yellow dots) and the station CVS (direction indicated by the green arrow). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

**Figure S5**. Location results of the Napa earthquake sequence obtained by using 16 master events (yellow dots) and the station CVS (direction indicated by the green arrow). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

**Figure S6**. Location results of the Ridgecrest earthquake sequence obtained by using eight master events (yellow dots) and two stations, WMF and WBF (directions indicated by the green arrows). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

**Figure S7**. Location results of the Ridgecrest earthquake sequence obtained by using 15 master events (yellow dots) and two stations, WMF and WBF (directions indicated by the green arrows). Red dots represent DD locations while blue dots represent the estimated ones. The lower-right subfigure (in green) represents the cumulative distribution plot showing the fraction of events (in per cent) located within a given (hypocentral) distance from the DD location.

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.

## APPENDIX: INTEREVENT DISTANCE CALCULATION

The Euclidian interevent distance between pairs of earthquakes within a seismic cluster is the basic input data used to solve the earthquake location problem with Distance Geometry techniques. Consider two seismic events $a$ and $b$ with coordinates $\mathbf{x}_a = x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k}$ and $\mathbf{x}_b = x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}$ respectively (defined in $\mathbb{R}^3$), and a seismic station $s$ with coordinates $\mathbf{x}_s = x_s\mathbf{i} + y_s\mathbf{j} + z_s\mathbf{k}$ (Fig. 1.a), where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are the unit vectors along the $x$, $y$ and $z$ axis, respectively. The position of the events $a$ and $b$ with respect to the station $s$ is

$$\mathbf{r}_{es} = \mathbf{x}_e - \mathbf{x}_s = (x_e - x_s)\mathbf{i} + (y_e - y_s)\mathbf{j} + (z_e - z_s)\mathbf{k} \quad \text{(A1)}$$

with $e = a$, $b$, then the Euclidian distance between the events $a$, $b$ and the station $s$ is

$$||\mathbf{r}_{es}|| = ||\mathbf{x}_e - \mathbf{x}_s|| = \sqrt{(x_e - x_s)^2 + (y_e - y_s)^2 + (z_e - z_s)^2} \quad \text{(A2)}$$

while the Euclidian distance between the event $a$ and the event $b$ is

$$||\mathbf{r}_{ab}|| = ||\mathbf{x}_a - \mathbf{x}_b|| = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2} \quad \text{(A3)}$$

and we can write $||\mathbf{r}_{ab}||^2$ as

$$||\mathbf{r}_{ab}||^2 = ||\mathbf{x}_a||^2 - 2\mathbf{x}_a^T\mathbf{x}_b + ||\mathbf{x}_b||^2, \quad \text{(A4)}$$

where $\mathbf{x}_a^T\mathbf{x}_b$ is the scalar product of $\mathbf{x}_a$ and $\mathbf{x}_b$.

Our goal is to estimate $||\mathbf{r}_{ab}||$ by using data recorded at a single seismic station (i.e. $P$ and $S$ first arrival times). From this data, it is only possible to estimate $||\mathbf{r}_{as}||$ and $||\mathbf{r}_{bs}||$ (i.e. the source station distances for events $a$ and $b$). We need to estimate $||\mathbf{r}_{ab}||$ by knowing $||\mathbf{r}_{as}||$ and $||\mathbf{r}_{bs}||$. It is easy to prove that $||\mathbf{r}_{ab}|| = ||\mathbf{r}_{as} - \mathbf{r}_{bs}||$; thus, we can write

$$||\mathbf{r}_{ab}||^2 = ||\mathbf{r}_{as}||^2 - 2\mathbf{r}_{as}^T\mathbf{r}_{bs} + ||\mathbf{r}_{bs}||^2, \quad \text{(A5)}$$

where $\mathbf{r}_a^T\mathbf{r}_b$ is the scalar product of $\mathbf{r}_a$ and $\mathbf{r}_b$, which can be written as

$$\mathbf{r}_{as}^T\mathbf{r}_{bs} = ||\mathbf{r}_{as}|| \, ||\mathbf{r}_{bs}|| \cos\phi, \quad \text{(A6)}$$

where $\phi$ is the angle between the vectors $\mathbf{r}_a$ and $\mathbf{r}_b$, lying on the plane passing through the station $s$ and the events $a$ and $b$ (as depicted in Fig. 1a). By substituting eq. (A6) into eq. (A5), we obtain

$$||\mathbf{r}_{ab}||^2 = ||\mathbf{r}_{as}||^2 + ||\mathbf{r}_{bs}||^2 - 2\,||\mathbf{r}_{as}||\,||\mathbf{r}_{bs}|| \cos\phi. \quad \text{(A7)}$$

If the interevent distance between the event $a$ and the event $b$ is much smaller than the distance between these events and the station $s$, then $\phi \approx 0$ and $\cos \phi \approx 1$. This condition is easily satisfied when the size of the seismicity cluster is much smaller than the distance between the station and the centre of mass of the cluster itself. If this condition is satisfied, we calculate an approximation of the interevent distance $\left\|\mathbf{r}_{ab}^s\right\|$ between the events $a$ and $b$ as follows:

$$\left\|\mathbf{r}_{ab}^s\right\|^2 = \|\mathbf{r}_{as}\|^2 + \|\mathbf{r}_{bs}\|^2 - 2\,\|\mathbf{r}_{as}\|\,\|\mathbf{r}_{bs}\| = (\|\mathbf{r}_{as}\| - \|\mathbf{r}_{bs}\|)^2,$$
(A8)

which becomes

$$\left\|\mathbf{r}_{ab}^s\right\| = |\,\|\mathbf{r}_{as}\| - \|\mathbf{r}_{bs}\|\,| \,.$$
(A9)

This last equation leads to an approximate estimate of the interevent distance $\left\|\mathbf{r}_{ab}^s\right\|$, as the absolute value of the difference between $\|\mathbf{r}_{as}\|$ and $\|\mathbf{r}_{bs}\|$. Eq. (A9) provides the exact interevent distance when the station and both events lie on the same line (i.e. when $\cos \phi = 1$ in eq. A7). It provides the worst results when the station and both events lie on lines that are perpendicular.