Originally published as:

Karam Zadeh Toularoud, N., Heimann, S., Dahm, T., Krüger, F. (2020): Earthquake source arrays: optimal configuration and applications in crustal structure studies. - Geophysical Journal International, 221, 1, 352-370.

<https://doi.org/10.1093/gji/ggaa002>

Earthquake source arrays: optimal configuration and applications in crustal structure studies

N. Karamzadeh,^{1,2} S. Heimann,¹ T. Dahm^{1,2} and F. Krüger²

¹*Physics of Earthquakes and Volcanoes, GFZ German Research Centre for Geosciences, Potsdam, Germany. E-mail: karamzad@gfz-potsdam.de*

²*Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany*

Accepted 2020 January 3. Received 2019 November 7; in original form 2019 August 5

SUMMARY

A collection of earthquake sources recorded at a single station, under specific conditions, are considered as a source array (SA), that is interpreted as if earthquake sources originate at the station location and are recorded at the source location. Then, array processing methods, that is array beamforming, are applicable to analyse the recorded signals. A possible application is to use source array multiple event techniques to locate and characterize near-source scatterers and structural interfaces. In this work the aim is to facilitate the use of earthquake source arrays by presenting an automatic search algorithm to configure the source array elements. We developed a procedure to search for an optimal source array element distribution given an earthquake catalogue including accurate origin time and hypocentre locations. The objective function of the optimization process can be flexibly defined for each application to ensure the prerequisites (criteria) of making a source array. We formulated four quantitative criteria as subfunctions and used the weighted sum technique to combine them in one single scalar function. The criteria are: (1) to control the accuracy of the slowness vector estimation using the time domain beamforming method, (2) to measure the waveform coherency of the array elements, (3) to select events with lower location error and (4) to select traces with high energy of specific phases, that is, sp- or ps-phases. The proposed procedure is verified using synthetic data as well as real examples for the Vogtland region in Northwest Bohemia. We discussed the possible application of the optimized source arrays to identify the location of scatterers in the velocity model by presenting a synthetic test and an example using real waveforms.

Key words: Location of scatterers; Optimization; Source array design.

1 INTRODUCTION

In seismology the word ‘array’ implicitly refers to a ‘receiver array’, that is a number of sensors deployed in a special geometry with common precise timing, acquisition parameters and instrument types. A receiver array signal processing provides signal-to-noise ratio (SNR) improvement by enhancing coherent signal arrivals while suppressing the incoherent background, using a time-domain or frequency-domain stacking procedure (Rost & Thomas 2002, 2009; Schweitzer *et al.* 2012). Using the so called ‘array beamforming’ method, slowness vectors of the incoming waves are estimated, which yields to seismic phase identification and further seismological findings, that is event location and rupture front tracking (Krüger & Ohrnberger 2005; Ishii 2011). Based on the reciprocity theorem of the Green’s function, under particular conditions, a number of seismic sources recorded at a single station are called a ‘source array’ (SA), in an analogy of an array of sensors (Spudich & Bostwick 1987; Scherbaum *et al.* 1991; Krüger *et al.* 1996). The reciprocity

theorem implies that for a vectorial force and receiver, if the positions of source and receiver in a seismic experiment are exchanged, the observed seismograms remain identical. Thus, after corrections due to different origin times and radiation pattern, signals of an SA received at a surface station are interpreted as if they originated at the location of the station and are recorded at the location of sources. This basic idea provides a unique tool to study the seismic structure of the areas with as few as a single recording instrument using array processing methods. In addition, by simultaneous use of an RA and an SA, double array method (Krüger *et al.* 1993, 1995; Scherbaum *et al.* 1997) is applicable to achieve a further SNR improvement which makes it possible to study low amplitude scattered phases and to locate scatterers out of the great circle path more precisely. To benefit from SA requires that the coordinate, origin time and depth of each source are precisely known. In addition, akin to the waveform coherency requirement for an RA, waveforms of an SA should be coherent. This criterion implicitly indicates that the travel path and source mechanisms of events are similar. If they are not

Inputs: Catalog, E ; K ; n_{start} ; n_{select} ; n_{new} ; $n_{iteration}$; n_{try}
Output: Optimized source array (m_o)

procedure SOURCE ARRAY OPTIMIZATION

- 1: Operate initial models $\rightarrow$ model population $M = \{m_i, i : 1, n_{start}\}$.
- 2: Create objective function population $O = \{o_i, i : 1, n_{start}\}$.
- 3: Create guiding models population $G = \{g_i \in M, i : 1, n_{select}\}$.
- 4: **for each iteration do:**
- 5: **for each model g in G do:** $\triangleright$ This is to update models.
- 6: Divide the E into K clusters, C , by Voronoi tessellation algorithm (Fig. 2).
- 7: **for n_{new} times do:** $\triangleright n_{new}$: No. of new models per each g .
- 8: **for each e in g do:** $\triangleright$ each element (event location) of g is updated.
- 9: Update e by random walk in the related C .
- 10: Accept or Reject the updated models based on inter-event distance comparison
- 11: with
- 12: models in G .
- 13: Add new accepted models into the M $\triangleright$ The procedure to find n_{new} accepted models is tried n_{try} times.
- 14: Update objective function population O .
- 15: Update guiding model population G . $\triangleright$ The no. of models in G : n_{select} .
- 16: Model with minimum value in $O \rightarrow m_o$

Figure 1. Source array optimization algorithm. A model m is defined as a source array with K events. The value of input parameters in the test examples in this study are selected as: E the given catalogue of 570 events; $K = 20$ number of events in each model; $n_{start} = 200$ number of initial random models; $n_{select} = 20$ number of guiding models in each iteration; $n_{new} = 20$ number of new accepted models per each guiding model; $n_{try} = 10$ number of attempts to make n_{new} accepted models per each guiding model; $n_{iteration} = 15$ number of iterations. After the last iteration a model m_o with the least value of objective function is selected as the final optimized source array.

Figure 2. Example of Voronoi diagrams (black) used to partition a 2-D distribution of earthquake locations (grey circles) by a set of irregular cells. Each cell includes events closest to its centre (red star) and so the shape of the cells is entirely defined by the location of the centre. Arbitrary source array (SA) elements are used as central points of Voronoi cells. The radius of the red circle drawn around one of the SA elements indicates the maximum distance for random walk and the events inside the circle (in grey area) can be replaced by the Voronoi centre. For each iteration a new set of Voronoi centres is selected.

identical, it is necessary to remove the source mechanisms from the recorded waveforms. Due to these pre-requisites, either nuclear explosions or big earthquakes are most often used to define an SA. The former due to more simple waveforms and well controlled source parameters, and the latter due to the possibility to remove a well-known source mechanism. The use of SA technique and the double array method are mostly focused on the study of properties of the velocity interfaces in different scales using converted or reflected phases, such as crust-upper mantle boundary structure (Niazi 1969; Goldstein *et al.* 1992), the distribution of heterogeneities in

the Earth's mantle and subduction zones (Weber & Wicks 1996), S -wave coda composition of microearthquakes (Scherbaum *et al.* 1991), origin of S -phase coda from earthquakes (Dodge & Beroza 1997) and the imaging of crust and upper mantle structure using scattered energy within the coda of the teleseismic P -phases (Revenaugh 1995). Inhomogeneities in the lowermost mantle was studied by reflected phases using double beam forming techniques (Krüger *et al.* 1993, 1995, 2001; Scherbaum *et al.* 1997; Rietbrock & Scherbaum 1999). In addition, recently the double array method is used to extract body and surface waves in volcano ambient noise processing (Nakata *et al.* 2016). The double array technique was also used to identify and separate low-amplitude body waves from high-amplitude dispersive surface waves at the exploration geophysics scale (Boué *et al.* 2013) and for acoustic tomography in a shallow ultrasonic waveguide (Roux *et al.* 2008).

SA allows for full slowness field calculation of the phases emerging from the source region. Thus, the positive attribute of using an SA is due to the possibility of constructing a 3-D array to study the near source structure and the travel path of the wavefield. In addition, providing the prerequisite conditions are fulfilled, without bearing the extra cost of sensor installation array processing methods are applicable and higher SNRs are achievable. Accordingly the number of SA elements can be increased, without paying extra cost. On the other hand, the negative attributes are due to the prerequisite conditions, which can be hard to achieve, such as similarity of sources, similarity of near source structure and earthquake location and origin time determination with high precision.

Given a well-located catalogue, which can be available by advanced relocation procedures and a well-configured monitoring system, the main challenge to make use of an SA is to eliminate the effect of waveforms dissimilarity and search for coherent arrivals. If events are big enough to calculate the related focal mechanisms and source time functions, one can deconvolve source effects from the signals. However, this is often not possible for smaller events. But smaller events are more numerous and we can alternatively search for similar waveforms indicating similar source mechanisms and travel paths. This can be feasible when a large number of well-located events in a relatively small volume exist in the catalogue. In analogy to an array of receivers, the distribution of source array

Inputs: Optimized source array (m_0 and Waveforms); Velocity model;
Phase name (i.e: sp or ps)
Output: Location of the scatterer

1: **procedure** SEMBLANCE
2: Grid the 3-D travel path $\rightarrow N_g$ Grids.
3: **for** each grid point g in Grids **do**:
4: **for** each source e in m_0 **do**:
5: Estimate the travel time of the scattered phase using the velocity model
6: assuming g is the scatterer point (eq:15).
7: Select the phase waveform using the estimated arrival time.
8: Calculate the semblance (Eq. 17) for the selected waveforms and assign to grid point g_i .
9: Maximum value of the semblance $\rightarrow$ location of the scatterer g_f .

Figure 3. Algorithm to find the location of the scatterer.

Figure 4. The Vogtland area which is an intraplate swarm region in German-Czechia border is shown. The green triangle shows the location of Rohrbach receiver array, and blue triangles represent the WEBNET stations. Red circles are epicenters of 2008 Nový Kostel swarm activity, and the black rectangle shows the search area for an optimized source array. Two predominant focal mechanisms are shown. The inset depicts the P -phase velocity model used to generate synthetic seismograms.

elements is also imperative aside from the signals coherency, in order to get the precise slowness vector components (Rost & Thomas 2009; Schweitzer *et al.* 2012). To fulfill all these criteria, our solution is to search for similar events in a specific geometry that imposes enough time delays in signal arrival times to estimate the slowness vector precisely based on coherent waveforms of interested phases.

In this study, the aim is to facilitate the use of SAs by automatic search of array elements, given a catalogue of events and associated waveforms. We propose the use of an optimization technique to find the best source array configuration. Using an optimization technique allows to define one or more objective functions to handle necessary criteria of setting the SA elements. We suggest four subfunctions, but in general the number of subfunctions can be adapted according to the specific goals of the studies: (1) to ensure high resolution of SA beamforming in 3-D slowness space, that can be evaluated using either synthetic or real data. This criterion controls the SA geometry, (2) to check the P -phase waveforms similarity over the array elements, (3) to select events with low location error and (4) to check the presence of reflected/converted phases. Then the weighting sum technique (Caramia & Dell' Olmo 2008) is used to linearly combine the defined subfunctions in one scalar objective

function. The proposed procedure can promote a variety of interesting applications of SA techniques related to the study of the near source structural complexity, for example to localize the position of a velocity or density contrast within the source region of an earthquake swarm. The double array technique further allows to study the velocity interfaces in different scales.

2 THEORETICAL APPROACH

We define a ‘model’, $\mathbf{m}$, as a source array with a specified number of elements, K . Each element of $\mathbf{m}$, that is m_k is the hypocentre of an earthquake which occurred within a confined epicentral area. A random model has a 3-D spatial configuration, every element of which is drawn from a given event’s catalogue E with J events. It is assumed that the given catalogue, includes precise event locations and origin times as well as related errors, but the source mechanisms of the events are not necessarily available. The ‘model space’ accommodates all possible K -element combinations of the J given events. The aim is to find the best model, that is a member of model space, that optimizes (minimizes) a well-defined scalar objective function.

Figure 5. Distribution of the events relative locations (stars) and their location errors (colour). The geometrical centre is set to zero in all planes, and the positive direction of the depth axis directs to the shallower events.

Figure 6. Location precision for different magnitudes. The total error is calculated assuming the latitude, longitude and depth error reported for each individual event in the catalogue.

Table 1. The maximum and minimum value of each subobjective function for the given catalogue (including 570 events) used to define the normalization coefficients.

Type	f_1	f_2	f_3	f_4
Synthetic	0.33–0.5	0.02–0.1	12.0–32.6	1.89–5.83
Real	0.36–0.65	0.15–0.61	12.0–31.3	0.36–2.7

In this section, the objective function formation as well as the implemented optimization algorithm are introduced. The objective function is evaluated during the optimization process employing the data associated to a trial model, such as the waveforms and location error in the catalogue. A key point of our approach is to automatize

the elaborate task of the SA set-up by utilizing an optimization process.

As an example application of the source arrays in crustal structure studies, a grid search algorithm to locate a scatterer is described in this section as well.

Table 2. The value of subfunctions used to define weighting factors.

	Type	f_1	f_2	f_3	f_4
p_1	Synthetic	0.33	0.11	14.8	4.5
p_2	Synthetic	0.51	0.02	14.1	4.5
p_3	Synthetic	0.43	0.121	12.1	4.7
p_4	Synthetic	0.5	0.2	14.6	1.9
p_1	Real	0.36	0.39	15.2	1.15
p_2	Real	0.51	0.15	15.7	2.65
p_3	Real	0.63	0.56	12.0	1.33
p_4	Real	0.53	0.45	14.2	0.36

2.1 Objective function formation

We introduced and evaluated four criteria to search for the best model. Each criterion is to target one specific property of the desired SA, and is formulated in one individual subfunction. The weighted sum technique is used to estimate the relative weights of the subfunctions and to combine them in one single scalar function. The single final objective function is used in the final optimization process, which results in the optimized SA. In the following subsections, first, individual subfunctions are introduced, then the final objective function formation is presented.

2.1.1 3-D array slowness resolution

Assuming the signal coherency is valid and the sources have similar mechanisms, the 3-D source array beam trace is calculated as the sum of all time-shifted traces recorded at one single station:

$$B(t) = \frac{1}{K} \sum_{j=1}^K Y_j(t - \tau_j + dT_j) \quad \text{with} \quad dT_j = s_x \delta x_j + s_y \delta y_j + s_z \delta z_j, \quad (1)$$

where Y_j is the normalized trace of source j and K is the number of sources. τ_j is the origin time of the source j . Assuming the plane wave approximation is valid, the time-shift dT_j for the source j , depends on the slowness components of the leaving wavefront s_x , s_y and s_z , and the relative distance to the SA reference point (centre) δx_j , δy_j and δz_j . Accordingly, the time-shift needed for array beamforming depends on the slowness vector components and source array elements distance vector from the array centre. The precision of the estimated slowness vector using array beamforming depends on the aperture and the 3-D intersource distances, that is 3-D geometry of the elements. Thus, by defining an objective function based on the source array beam power, we can take into account the SA geometry. We measure the average amount of relative power of the 3-D array beam which is calculated in 3-D slowness space using the given sources and one station location. Ideally, array beam power in slowness space contains a concentrated sharp main peak, like a delta function and no other localized secondary peaks. The maximum value of the array beam relates to the resolved slowness vector. The sharpness of the main peak shows the resolution of the estimated slowness vector. A sharp main peak causes a relative low plateau surrounding the peak, resulting in a small average of the beam power in s_x - s_y - s_z space. To control the accuracy of slowness vector estimation and achieve the highest possible slowness resolution using the SA, we set our objective to minimize the average value of the array beam power. An SA beam is calculated for all 3-D gridpoints of the slowness space, considering appropriate time windows of the traces around arrival times of the selected phases. Then, three 2-D cross-sections of the 3-D beam pattern, crossing

the beam power maximum value, are extracted from the full 3-D pattern and the related average beam power is calculated as:

$$f_{xy} = \frac{1}{n_x n_y} \sum_k^{n_y} \sum_l^{n_x} A_{x_l, y_k}^2, \quad (2)$$

$$f_{xz} = \frac{1}{n_x n_z} \sum_k^{n_z} \sum_l^{n_x} A_{x_l, z_k}^2, \quad (3)$$

$$f_{zy} = \frac{1}{n_z n_y} \sum_k^{n_y} \sum_l^{n_z} A_{z_l, y_k}^2, \quad (4)$$

where n_x , n_y and n_z are number of gridpoints in the predefined slowness ranges for s_x , s_y and s_z ; and A^2 is the value of beam power which is normalized to the global maximum of all tested gridpoints, so it is called relative beam power. The average relative beam power, as a measure of 3-D array resolution, is defined as f_1 to combine above equations in one formula and to define the first objective function as:

$$f_1 = \frac{1}{3} (f_{xy} + f_{xz} + f_{zy}). \quad (5)$$

A larger value of f_1 , indicates higher value of the average beam power in the slowness space for the SA, which can happen either because of high or numerous side lobes or a wide main lobe in each of the 2-D slowness maps, both of which are undesirable.

2.1.2 Waveform similarity

Similar to the RA, waveforms of an ideal SA should be identical in full measure. This condition is fulfilled if the source parameters of all events are identical and the travel paths between the sources and the station are similar. The given event catalogue can include heterogeneous source mechanisms, accordingly for a randomly selected model, including a number of random sources, dissimilarity of the related waveforms is likely to happen. Nevertheless, the desire is to have maximum waveform coherency of the direct and the reflected/converted phases for the events of the optimized SA. Waveform similarity over the full length of the recording is not needed and cannot be ensured, because due to different sources-receiver and sources-reflector paths the time delay of the secondary phases from the first arrival P -phases may not be identical for all sources. In addition, we presume that events of similar P -phase waveform, aside from similar source mechanisms, might travel along similar source-receiver paths, consequently the waveforms of the secondary phases, that is P to S converted phases ps , may be similar as well. Thus, an objective function f_2 , is introduced to measure P -phases waveform similarity of a given SA, by calculating the waveform cross-correlation matrix, where each element is the cross-correlation of a pair of P -phase waveforms measured in time domain. Accordingly, a K^2 element symmetric the cross-correlation matrix, is associated to each model.

Figure 7. Variation of subfunctions and final objective function, for random models (20-element source arrays) selected from the given catalogue and tested during the optimization process in synthetic test. Panels (a)–(d) are related to the f_1 array beam power, f_2 waveform coherency, f_3 location error and f_4 energy of the coda phases, respectively and panel (e) shows the final objective function defined by (18). The first 300 models are randomly chosen, while the later models are generated by random walk in the neighbourhood of the guiding models as introduced in Section (2.2).

Given two signals x_k and $x_{k'}$ are related to the sources k and k' , respectively, we can delay x_k by m samples and then calculate the cross-covariance between the pair of signals, that is:

$$\sigma_{kk'}(m) = \frac{1}{N-1} \sum_{n=1}^N (x_k(n-m) - \mu_{x_k})(x_{k'}(n) - \mu_{x_{k'}}), \quad (6)$$

where μ_{x_k} and $\mu_{x_{k'}}$ are the means of each signal and there are N samples in each. n is the sample's number; assuming t is the time, $n = \frac{t}{\delta t}$ and δt is the sampling rate of the signal. The function $\sigma_{xy'}(m)$ is the cross-covariance function. The cross-correlation function is a

Figure 8. Geometry of an optimized 20-element source array obtained from synthetic test. See the Fig. (5) caption for further explanations.

normalized version of the cross-covariance function:

$$\sigma'_{kk'}(m) = \frac{\sigma_{kk'}(m)}{\sqrt{\sigma_{kk'}(0)\sigma_{k'k'}(0)}}. \quad (7)$$

The $\sigma'_{kk'}(m)$ is calculated for a possible range of valid m , which is defined based on the signal's length. Then the maximum calculated value $\hat{\sigma}$, is used to define the cross-correlation matrix:

$$C = [\hat{\sigma}_{kk'}], k = 1 : K, k' = 1 : K, \quad (8)$$

where C is the cross-correlation matrix. Assuming the given signals are P -phase waveforms, then f_2 , as a second subobjective function is defined as:

$$f_2 = \left(\sum_{k=1}^{K-1} \sum_{k'=k+1}^K (1 - C_{kk'})^2 \right)^{\frac{1}{2}}. \quad (9)$$

2.1.3 Location error

Although we assumed that the given catalogue includes well-located events, any event location solution comes with a specific location error, regardless of the type of implemented location method. We defined another criterion, f_3 , to ensure the selection of the events with the lowest location error. The total location error for one event is measured as:

$$e = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}, \quad (10)$$

where σ_x , σ_y and σ_z are the errors in latitude, longitude and depth of each earthquake, respectively. f_3 for an assumed SA is defined as:

$$f_3 = \sqrt{\sum_{k=1}^K e_k^2}. \quad (11)$$

2.1.4 Presence of scattered phases

Arrival time and amplitude of the near-source scattered waves on seismograms, depends on the geometry and position of source, receiver and the scatterer, while their amplitude is strongly dependent on the focal mechanisms and the radiation pattern (e.g. Hrubcová et al. 2016). The scatterer, that is velocity interface, can be localized or laterally extended in part of the propagation media. Thus,

depending on the incident angle of the phases impinging on the interface, the relative distances between the sources and the interface, and the velocity contrast, the scattered phases can be strong or weak or not existing at all. Furthermore, if the media is highly scattering or more than one scatterer exists, deconstructive interference of the scattered phases is likely to happen. Accordingly, the waveforms are inspected to verify if such phases are present with considerable energy.

The ideal is to select events with the strongest reflected/converted phases produced from the same reflector. Accordingly, we suggest to measure the average kinetic energy of the waveforms, in a specified time window between the P - and S -phase arrivals, for each element of model after rectifying the magnitude differences between events. To rectify the magnitude differences, the traces are scaled by the maximum/minimum amplitude of P -phases, so that the maximum/minimum amplitude of P -phases are normalized to 1 or -1 . So, for P to S and S to P phase scatterings, the fourth objective function is defined as:

$$f_4 = \frac{w}{\sum_{n_i}^{n_i+w} x_n^2}, \quad (12)$$

where w is the number of samples in the specified time window and n_i is its beginning sample. f_4 is defined as the inverse of the mean energy of the normalized traces to contribute in the overall objective function. In our tested example we assumed the segment of the trace after the P -phase and before the S -phase arrival time. However, in application, the time window can be adjusted according to the theoretical traveltime of the interested secondary phases.

2.1.5 Final objective function

Each subfunction introduced before, has different unit and range of values, so before combining, they should be scaled and made dimensionless. The range of the subfunctions are normalized using the general formula given by:

$$\hat{P} = \frac{P - P_{\min}}{P_{\max} - P_{\min}}, \quad (13)$$

where P is the original value and $\hat{P}$ is the normalized value, assuming P has the limited range between $P_{\min}$ and $P_{\max}$. Accordingly, we estimated the possible range of each subfunction, by finding optimized models of individual functions. In addition, the values of other objective functions for individual optimized models are used to estimate relative weights of the pair of objective functions as described in Karamzadeh et al. (2018). We employed a procedure to define weighting coefficients between two subfunctions so that both subfunctions could contribute almost equally into the final solution (Karamzadeh et al. 2018). The method is extendable to define the weighting coefficients of more subfunctions (the Appendix). To combine four subfunctions, first the weighting coefficients between $\hat{f}_1$ and $\hat{f}_2$ are estimated and a combined function, f_{12} is made, then following the same procedure the weighting coefficients between $\hat{f}_3$ and f_{12} are calculated and a combined function is named f_{123} . Finally, the coefficients between $\hat{f}_4$ and f_{123} are calculated and The final scalar objective function is defined as:

$$F = \alpha \hat{f}_1 + \beta \hat{f}_2 + \gamma \hat{f}_3 + \lambda \hat{f}_4, \quad (14)$$

where $\hat{f}_i$ are scaled subfunctions, and weighting coefficients α , β , γ and λ are defined so that: $\alpha + \beta + \gamma + \lambda = 1$.

Figure 9. (a) Full synthetic waveforms of the optimized source array (see Fig. 8) aligned according to the *P*-phases arrival times. Arrival times of directed *p*- and *s*-phases and *sp*- and *ps*-converted (scattered) phases are indicated by markers. The pair of numbers on each trace are epicentral distance and depth given in km. (b) *P*-phase waveforms of the optimized source array. (c) *P*-phase waveforms of all events in the used catalogue. All traces are bandpass filtered (2–20 Hz).

2.2 Optimization algorithm

The aim of the optimization process is to find the best possible model, m_o , from the model space. Assuming a specific receiver, it gives the best performance of the defined scalar objective function (eq. 14) as described in Section 2.1.5. Since the objective function calculation can be computationally expensive (it takes about 1 min for one function evaluation using a 8-core processor), we used a fast converging procedure (Karamzadeh *et al.* 2018) which is a modified version of the simulated annealing procedure that benefits from a sampling technique based on the neighbourhood concept (Sambridge 1999; Wathelet 2008) and an objective function approximation using Voronoi cells (Okabe *et al.* 1992).

The optimization process (see Fig. 1) works according to the following steps: (1) Operate initial models: The n_{start} (input parameter) initial models are generated as the model population, M , which is a subset of the model space. Each model is created by a random selection of K (input parameter) events from the given catalogue. However, a minimum interevent distance threshold of 100 m is considered to avoid spatial overlap of array elements. (2) Update objective function: the objective function is evaluated for models in M to make, at the first iteration, and update, at the later iterations, the objective function population, O . (3) Select best models: M is sorted according to the corresponding values in O , and n_{select} (input parameter) models with the lowest value in O are selected as the ‘best models’ to produce the ‘guiding models’ population, G . The

Figure 10. The 3-D array beam pattern for the optimized SA calculated using synthetic waveforms (see Fig. 8 for array geometry).

Figure 11. The 3-D cross-sections of the semblance values including its maximum value. Semblance is calculated using the time windows of 0.1 s length, extracted from the source array synthetic recordings, according to the sp-phase theoretical arrival time. White stars show the location of the sources and the white triangle on (a) is the projection of the receiver at the surface. The horizontal white lines on depth axes (b and c) indicate the depth of the velocity interface at 5 km. The yellow arrow in (b) and (c) plotted from the location of the maximum semblance towards the location of the station at the surface. The traces are bandpass filtered: 2–20 Hz.

best model selection is repeated in the subsequent iterations since M and O are updated. (4) Perturb randomly (update models): n_{new} (input parameter) new models are generated using each of the models g , out of the G . All events in the input catalogue are divided into the K clusters using the euclidean distances of hypocentres using the Voronoi clustering algorithm implemented by SciPy-Spatial algorithms and data structures package (Virtanen *et al.* 2019, see Fig. 2). Each cluster is associated to one event $g_l \in g$, and all events in that cluster are closer to the g_l compared to the other g_j , $j \neq l$. New models are generated by replacement of each element g_l with another element of the same cluster. The maximum possible distance between two events for random replacement is reduced by

increasing the iteration steps using a cooling scheme. In this way, an ensemble of new candidate models are created using each guiding model. (5) Accept or reject the updated models: Before a candidate model is accepted as a new model, the inter-event distances of that is compared with the corresponding values of all models in M . For each model, the inter-event distance matrix I is evaluated, that is a squared matrix of K^2 elements. All values above the main diagonal of I are sorted to make a quantitative accept or reject criterion. A candidate model is accepted as a new model, if it is in the closest neighbourhood of one of the best models in M , that is models in G . By applying this condition, for instance, models with small aperture (largest interevent distance) compared to the aperture of the

Figure 12. The 3-D cross-section of the semblance values including its maximum value, calculated using synthetic traces and according to the ps-phase traveltimes, see the Fig. 11 caption for more explanations.

Figure 13. Schematic ray path of sp- and ps-phases calculated using the given 1-D velocity model used for the synthetic test in this study. The ray paths are calculated by using the Cake tool implemented in Pyrocko package (Heimann *et al.* 2017). The horizontal lines represent the velocity layers. The source is at 8 km depth and epicentral distance is 10 km.

guiding models and models that are not well distributed in space are rejected. In every iteration, new models are added to the M , and O and subsequently G are updated. After the last iteration, the model with lowest value in O is reported as the final optimized model.

In this algorithm, the reason to employ the Voronoi tessellation algorithm is to cluster the given catalogue into a fix number of cells, where the centre of each cell is predefined by the guiding model's elements (event's location in the guiding model). Any other clustering algorithm that can do such a division is applicable as well.

2.3 Location of the scatterer

A possible application of an SA is to locate the scatterer which generates strong coherent scattered waves. We investigated the SA application by applying a grid search procedure to find the location of the scatterer. The procedure (see Fig. 3) requires an estimation

of the travel path and the arrival time of scattered phases such as sp-phase and ps-phases for a single receiver and works according to the following steps: (1) A 3-D volume which includes the travel path of the scattered phases originated at the source array and recorded at the single receiver, is divided into small size 3-D grids. (2) Assuming each gridpoint, g_i is the trial position of the scatterer, for all sources, trial traveltimes of the phases are calculated by summing the s- and p-phase segments:

$$\tau_{sp}(i, k) = \tau_s(i, k) + \tau_p(i, k), \quad k = 1 : K, \quad (15)$$

where K is the size of the source array. (3) The trial sp-phase waveforms are extracted from the related waveforms, using the trial traveltimes values:

$$z_k^{sp}(i) = z_k[\tau_{sp}(i, k) + t_{\text{origin}}(k) : \tau_{sp}(i, k) + w + t_{\text{origin}}(k)], \quad (16)$$

Figure 14. Source array beam traces calculated using time delays according to the gridpoints with maximum semblance values. Beam 1 and Beam 2 are computed using synthetic traces and Beam 3 and Beam 4 are related to the real data.

where w is the length of the time window, and t_{origin} is the origin time of each source. Then the semblance coefficient is calculated for the selected waveforms. The gridpoint g_i , where the related semblance value is maximum, is assumed to be the location of the scatterer. It should be noted that for gridpoints, whose related $\tau_{sp}(i, k)$ were too large, and close to the $\tau_{ps}(i, k)$ (ps-phase arrival time), the semblance is not calculated and the value set to a constant value calculated from the background noise semblance. Semblance is calculated from the following formula and assigned to each g_i .

$$S = \frac{\sum_{t=0}^{t=w} (\sum_{k=1}^{k=K} z_k(t))^2}{K \sum_{t=0}^{t=w} z_k(t)^2}. \quad (17)$$

This approach does not require the plane wave approximation condition to be hold, so it can be applied for any source–receiver distances.

3 DATA AND EXPERIMENTS

The presented method is evaluated by performing two experiments, (1) a synthetic test using realistic synthetic seismograms based on a real catalogue and (2) using real seismograms. In both experiments, subject of the study is Northwest Bohemia/Vogtland, in the border region between Germany and Czech Republic, that is well known for the repeated occurrence of earthquake swarms. Earthquake swarms are a vast number of weak events occurring in a spatio-temporal cluster over a period of weeks or months and are not generally associated with a typical main shock–aftershock sequence of earthquakes. The causes of earthquake swarms are considered to be either magmatic activities (Morita *et al.* 2006; Dahm *et al.* 2008), fluid-migration (Hensch *et al.* 2008; Hainzl *et al.* 2016) or aseismic creep on faults (Neunhöfer & Hemmann 2005; Passarelli *et al.* 2015).

The Vogtland area has experienced various swarm activities during 1985–1986, 1997, 2000, 2008, 2011, 2014, 2017 and 2018 (Fischer *et al.* 2014; Hainzl *et al.* 2016; Krentz 2019). The region has been subject of many seismological studies, mainly based on the observations of the West Bohemia seismic network (WEBNET) operating in Czechia territory, and is selected for developing an interdisciplinary observatory using shallow drilling and small aperture seismic arrays (Dahm *et al.* 2013).

Our application is focused on a swarm in 2008 which occurred in the Nový Kostel (NK) fault zone in Vogtland (see Fig. 4), a planar structure steeply dipping westward (Fischer & Horálek 2003; Fischer *et al.* 2010), oriented nearly S–N, and reactivated at a depth

between 6 and 11 km. According to (Horálek & Šílený 2013), the focal mechanisms of the swarm activity which occurred close to Nový Kostel are mostly oblique-normal and oblique-thrust types but the oblique-normal faulting predominates. The oblique-normal events have predominant strikes of 160° – 170° , dips of 72° – 80° and rakes of -28° to -38° whereas the oblique-thrust events show mainly strikes of 355° – 360° , dips of 80° – 85° and rakes of 35° – 40° .

The waveforms of swarm earthquakes in Vogtland typically display distinct direct P and S waves followed by high frequency coda waves generated at crustal interfaces and at small-scale inhomogeneities. The secondary phases are proved to be sensitive to the focal mechanisms and are different for each station of the local network and can be used to identify and image the prominent crustal discontinuities (Hrubcová *et al.* 2013, 2016, see Fig. 17a for examples of waveforms). The 2008 swarm activity was also recorded by a small-aperture seismic array (Rohrbach array) operated by the University of Potsdam from 19 October 2008 until 18 March 2009 (see Fig. 4, Hiemer *et al.* 2012). The analysis of high-quality data recorded in Rohrbach array indicated near-vertical ray incidence of P - and S -phases, while calculated back azimuths showed 30° deviation from theoretical values, due to the structural inhomogeneities in the propagating zone (Hiemer *et al.* 2012). A well-located catalogue based on relative master event location using precise arrival-time picking of WEBNET seismic network is used (Fischer *et al.* 2010) for both experiments. The original catalogue includes 5679 events in the magnitude range of $M = -1$ up to $M = 3.5$. For our test 570 events of the total catalogue are selected based on the availability of waveforms in station V02 of the Rohrbach array. Fig. 5 shows the relative location of hypocentral parameters of those selected events. The overall reported location error in the catalogue versus magnitude of events is plotted in Fig. 6. The overall error is calculated from root mean square of errors in latitude, longitude and depth estimation. To calculate synthetic waveforms, we assigned realistic hypothetical focal mechanisms to each event in the catalogue. The hypothetical mechanisms are generated by assuming 5° variation in each of the rake, strike and dip values, reported for two possible predominant clusters in NK zone. We used a velocity model with a strong contrast at depth of 5 km, to produce converted phases (such as ps- and sp-phases). The velocity model is shown in Fig. 4. Random uniform noise is also considered for the individual synthetic traces. Both synthetic and real waveforms are bandpass filtered in a frequency range which is selected to ensure good resolution and reasonable small aliasing of final results, given the magnitude and epicentral distance range of events and the visibility of the scattered phases on signals.

We use this region as an example and ask the question if the swarm catalogue can be used to systematically search for the source array elements to be optimally used for further seismological studies. Using the WEBNET stations distributed at different backazimuths to the sources can help to study lateral variations of the crustal discontinuities, while Rohrbach array can help to apply the double array method.

4 SYNTHETIC TEST

Using the real catalogue of swarm events (including 570 events), assuming realistic random focal mechanisms and an appropriate velocity model, realistic synthetic seismograms at one station of the Rohrbach array are simulated by means of pre-calculated Green's function databases and efficient storage and accessing tools (Heimann *et al.* 2017). The generated synthetic signals are used

Figure 15. Variation of the subfunctions f_1 – f_4 and the final objective function (F), for random models selected from the given catalogue and tested during the optimization process using the real waveforms. See the Fig. 7 caption for more explanations.

during the optimization process to evaluate the SA beamforming, P -phase similarity and the energy of coda phases. The search algorithm is initiated to find a model, that is an SA including 20 sources. The process starts by defining the weighting coefficients of the subobjective functions (eq. 14) and formulating the final scalar function. Following the procedure described in (Karamzadeh *et al.* 2018), we have to run the optimization algorithm using the individual objective functions to calculate the normalization and relative weighting coefficients. So, the optimization algorithm is operated

given the defined objective functions, f_1, f_2, f_3 and f_4 , while in each operation for the tested models, the values of the other functions are evaluated as well. The maximum and minimum value of each objective function used for normalization are listed in Table 1, and the values of objective functions to define the weighting coefficients are given in Table 2. For instance, p_1 is a model obtained by optimizing f_1 , and the optimal/minimal value of f_1 is 0.33, the values of other three objective functions for p_1 are evaluated and listed in the first row of the table. Using values of Table 1, values of Table 2

Figure 16. Geometry of an optimized 20-element source array obtained from real data. See the Fig. 5 caption for further explanations.

are scaled and normalized, then following the procedure described in Karamzadeh *et al.* (2018, see the Appendix) the final objective function is formulated as:

$$F = 0.14\hat{f}_1 + 0.2\hat{f}_2 + 0.18\hat{f}_3 + 0.48\hat{f}_4, \quad (18)$$

where $\hat{f}_i$ indicates the normalized and scaled version of f_i .

4.1 Results of synthetic data test

Using the defined final objective function (eq. 18), the optimization process gives us the final optimized SA which is a model that is supposed to fulfill all 4 defined criteria. In the ideal case, it is expected that the final solution minimizes all subfunctions f_1 – f_4 simultaneously. However, in practice there might be an inherent trade-off between subfunctions, and such an ideal solution might not exist. For example, a model with minimum overall location error is not necessarily the model with highest waveform similarity and a model with highest waveform similarity might not have well distributed elements in 3-D space to ensure the highest possible slowness vector estimation resolution. However, by using the final objective function (eq. 18), which is a weighted sum of all subfunctions, we try to find a model as close as possible to an ideal solution for each subfunction.

Fig. 7 shows variation of individual subfunctions (a)–(d) as well as the final function (e) during the optimization process. In this example, the first 300 models are selected in random and show higher variation range in all five plots. The subfunction defined based on the array beam power f_1 (Fig. 7a) shows steady improvement by increasing the iteration number, as it depends on the geometry and relative location of a tested model's element. In the applied optimization method, the radius of the random walk decreases by the iteration, so variation of the geometry gets steadily smaller during the process. The other subfunctions experience local fluctuations as the corresponding properties can even change from one event to the adjacent event. For example, according to Fig. 5 the location error is distributed heterogeneously. Most of events have similar location error (20–25 m), but events with large errors up to 60 m are existing in the catalogue and can be included in the trial models during the optimization process. Fig. 8 shows the geometry of

the final optimized SA. The waveforms of the SA are depicted in Fig. 9(a), where the theoretical ps- and sp-phases onset times are marked on individual traces. The waveforms are sorted according to the depth. The values of depth and source–receiver horizontal distances are written on each trace as well. In Fig. 9(b) *P*-phases waveforms are plotted in the shared frame to compare the similarity and quality of the waveforms visually. Fig. 9(c) shows the *P*-phases waveforms of all events in the catalogue, including the waveforms of the optimized SA. Comparing Figs 9(b) and (c) indicates that the optimization algorithm could successfully select a similar subset of events. In Fig. 10 the related array beam power is depicted. Fault geometry imposes a narrower distribution of events in east–west direction, accordingly the maximum achievable source array apertures in east–north and east–depth plane are smaller than the maximum achievable aperture in north–depth plane, which results in the lower slowness resolution in both planes compared to the east–depth plane.

Using the final optimized source array, the procedure described in Section 2.3 is employed to locate possible scatterers in the wave propagation media. The calculated semblance value for all trial gridpoints, using the sp- and ps-phase arrival times, are plotted in the cross-section plots shown in Figs 11 and 12, respectively. In these plots, white stars represent the location of the source array, the white triangle in the horizontal Northing–Easting plane (plot a) shows the relative location of the receiver at the surface. In plots (b) and (c) the horizontal white lines represent the location of the interface at 6 km depth in the velocity model, the yellow arrows are plotted from the gridpoint showing the maximum semblance and direct to the relative position of the receiver at the surface. In each cross-section plot, the gridpoint with maximum value of the semblance shows the location of the point scatterer related to the converted phase origin.

The calculated depth and location of the scatterers in horizontal plane are in accordance with the theoretical ray paths of sp- and ps-phases for this geometry. Fig. 13 illustrates the ray path of p to s and s to p converted phases at the velocity interface at 5 km depth for the source–receiver geometry similar to our example, that is a source located at 8 km depth and a receiver at 10 km epicentral distance. For this setup, the impinging points of the ps-phase ray on the interface is closer to the source than to the receiver, while the impinging point of sp-phase ray is closer to the receiver than to the source. This point is visible in the semblance patterns shown in Figs 11 and 12, while the depth of the maximum semblance is very close to the 5 km depth, that is 4.95 km for the sp-phase and 4.85 km for the ps-phase. For the sp-phase, in the horizontal cross-section plot, that is Fig. 11a, another peak (label B) is appeared very close to the main peak (label A). It is due to the fact that the ray path of each event can be slightly different from the others and the impinging points of all rays to the interface surface may not be at the same location. The reason for another localized high semblance in this plane (label C) can be due to the existence of other low energy coherent waves (e.g. ps-phase) with similar slowness vector as the sp-phase which is targeted in this example.

The sp- and ps-phase beam traces are plotted in Fig. 14, labelled as ‘Beam 1’ and ‘Beam 2’, respectively.

5 APPLICATION ON REAL DATA

Using the given catalogue and the related waveforms, the algorithm is initiated to search for an optimized 20-element SA at one station of the Rohrbach array, the same as used for the synthetic test. Similar

Figure 17. (a) Full real waveforms of the optimized source array (16) aligned according to *P*-phases arrival times. The pair of numbers on each trace are epicentral distance and depth. (b) *P*-phase waveforms of the optimized source array. (c) *P*-phase waveforms of all events in the used catalogue. All traces are bandpass filtered (2–20 Hz).

to the experience with the synthetic data in Section 4, we use the values reported in Tables 1 and 2, and define the final objective function as:

$$F = 0.17\hat{f}_1 + 0.175\hat{f}_2 + 0.34\hat{f}_3 + 0.315\hat{f}_4, \quad (19)$$

where $\hat{f}_i$ indicates the normalized and scaled version of f_i . The dynamic range of the array beam power f_1 , is comparable for both experiments (Table 1). However, the waveform similarity f_2 is scaled up to 7.5 and 6.1 times for the lower and upper bands of real data compared to the synthetic data, which indicates that the *P*-phase waveforms of the real test are more diverse compared to the synthetic test. The fact that a uniform magnitude ($M_L = 0-3.2$) has been used in the synthetic tests to generate the earthquake catalogue might partially explain this observation. Additionally, the noise level in the synthetic test has less variation for different events compared to the

real waveforms. We have not excluded larger magnitude events from the catalogue to minimize the effect of later source extension on waveform dissimilarity, but preferred to keep them in the catalogue, and let the optimization algorithm to avoid automatically the larger events from the optimized SA.

As we expect, because the location error for both cases are identical, the dynamic range of f_3 is almost the same as well (Table 1). In particular the lower band of the dynamic range is identical for both tests, indicating that the optimization algorithm successfully could select the best event ensembles in terms of the location error.

In the case of the f_4 criterion, which is to measure the energy of the coda waves, the values of both limits of dynamic range are measured lower for the real test compared to the synthetic test. This indicates that the SNR of coda phases in the synthetic test is higher than for the real data, whereas the ratio of upper-to-lower value in real case

Figure 18. Array beam pattern for an example SA using real traces (see Fig. 16 for the array geometry).

Figure 19. (a) Waveforms recorded from source S11 (Fig. 17a) at all 10 stations of the Rohrbach array. (b) *P*-phase waveforms.

is 2.5 times higher than the same value measured in the synthetic test. This may be explained by pairs of time overlapping events, for which the *P*-phase of the second event could be misinterpreted as a secondary phase of the first event in each pair. In synthetic traces, such pairs of events are stored in separated waveform files, and the related waveforms are not overlapping in time.

It is noteworthy that the effect of some different dynamic ranges for real and synthetic tests are eliminated by the normalization process, as each criterion is normalized regarding to its individual measured dynamic range.

5.1 Results of real data application

Having defined the final objective function for the real waveforms recorded at one single station of the Rohrbach array (see Fig. 4), the optimized real SA is specified. Fig. 15 shows the values of all subfunctions in panels (a)–(d) and the final function in (e), which are evaluated during the optimization process. Fig. 16 shows the geometry of the final optimized SA. The waveforms of the optimized

SA are depicted in Fig. 17(a) and are sorted according to the depth. The values of depth and source–receiver horizontal distances are indicated on each traces. In Fig. 17(b) the *P*-phases are plotted in a shared frame to compare the similarity and quality of the waveforms. The *P*-phase waveforms of all events in the catalogue are depicted in Fig. 17(c). The array beam power in 3-D slowness space for the SA geometry is shown in Fig. 18.

Fig. 19(a) shows waveforms collected using a single source and 10 stations of the receiver array (Rohrbach Array), and the *P*-phases are extracted and plotted in a single frame in Fig. 19(b). Comparing the *P*-phase similarity of a source and receiver array (Figs 17b and 19b) we can conclude that waveforms of an SA recorded in a single station can show higher similarity compared to the waveforms of a single source recorded in a receiver array. In other words near receiver structure differences are in this case stronger than the near source structure differences. The value of *P*-phase cross-correlation for the receiver array is 0.36, whereas for the source array waveforms shown in Fig. 17(c) this value is 0.87.

Using the final optimized source array, according to the procedure described in Section 2.3, similar to the synthetic test case, we tried to locate possible scatterers existing in the wave propagation media. The same velocity model as was used in the synthetic test case, employed in this test to predict arrival times of ps- and sp-phases assuming trial point scatterers in the velocity model. The calculated semblance value for all trial gridpoints are plotted in the cross-section plots shown in Figs 20 and 21 for sp- and ps-phases, respectively. In these plots, white stars are locations of source array elements. The white triangle on horizontal Northing–Easting plane, that is plot (a), shows the relative location of the receiver at the surface. In plots (b) and (c) the yellow arrows are plotted from the gridpoints showing the maximum semblance and direct to the position of the receiver at the surface. The maximum value of the semblance in each cross-section plot, indicates the location of the

Figure 20. Semblance patterns calculated using the time windows extracted from the real source array recordings, according to the sp-phase theoretical arrival times. The location of maximum semblance in each plot indicates the location of the scatterer (see the Fig. 11 caption for more explanations.). Array beam trace for the time-shifts related to the maximum semblance location is plotted in Fig. 14 and the related phase is highlighted on the trace named ‘Beam 3’.

Figure 21. Semblance patterns calculated using the time windows extracted from the real source array recordings, according to the ps-phase theoretical arrival times. The location of maximum semblance in each plot indicates the location of the scatterer (see the Fig. 11 caption for more explanations.). Array beam trace for the time-shifts related to the maximum semblance location is plotted in Fig. 14, and the related phase is highlighted on the trace named ‘Beam 4’.

scatterer. In both cases, Figs 20 and 21, clear peaks in semblance map are visible, however the indicated scatterers are localized in different location. The sp- and ps-beam traces are plotted in Fig. 14, labelled as ‘Beam 3’ and ‘Beam 4’, respectively.

6 DISCUSSION AND CONCLUSIONS

In this paper, we presented a method to search for seismic source array elements, given a well-located earthquake swarm catalogue. The proposed method benefits from an optimization scheme, which allows to evaluate a number of customizable objective functions quantitatively. We formulated the preconditions to make an SA, by defining 4 objective functions, f_1 – f_4 .

The synthetic test and the test with real data proved that the proposed method by using the suggested optimization approach is capable to find an optimized SA according to the defined objective functions. In addition, the values of the individual subfunctions at the final optimized SA, indicates that our weighting strategy performs well.

One of the key conditions to define an SA is the accuracy of source locations, which is formulated in f_3 . A test is performed to see how the location error propagates into the final outcomes of the source array beamforming analysis, that is slowness vector estimation.

First, 3-D slowness vector elements s_x , s_y , s_z , for a 20-element source array are estimated using real waveforms according to eq. (1) assuming the exact locations of the sources are given. The corresponding 3-D array beam patterns are shown in Fig. 22 and the slowness vector is indicated by red stars on 2-D slowness planes. Then, assuming arbitrary error vectors for elements of SA, they are shifted to new locations. Afterwards, the time-shifts (eq. 1) are recalculated for the new locations and the slowness vector is updated and compared with the value obtained initially without considering the location errors. The circles shown in Fig. 22 indicate the updated slowness values for the perturbed SAs. Components of the tested error vectors are determined randomly, assuming an individual Gaussian distribution for each component, that is latitude, longitude and depth, with a specific standard deviation. The colour bar in Fig. 22 shows the maximum value of the assumed standard deviation for each perturbed model. Histograms depicted in Fig. 23 illustrate the values of deviations obtained for each component. According to the results, for the tested source–receiver geometry and SA configuration, as it is expected, by increasing the value of location error, the error of estimated slowness vector increases, so that assuming a maximum 75 m location error, up to 0.15 s km^{-1} deviation in S_x component of slowness vector is likely to happen,

Figure 22. Location error propagation in slowness vector components. The array beam power depicted in the background is calculated assuming zero error. Each circle represents a slowness vector assuming a specific error in source array elements location. The colourbar shows the maximum error in each tested case.

Figure 23. Deviation of slowness vector components calculated after perturbation of the array elements. Values of the maximum errors are shown in Fig. 22.

while for the other components, the value of deviation is less than 0.07 s km^{-1} . It is interesting that the deviating slowness values are distributed around the main peaks and show correlation with the array beam power patterns of the original error-less SA. Accordingly, in the s_x-s_z plane, where we find the sharpest central peak, the

lowest deviation of slowness vector component is observed. In other words, for the smaller aperture source arrays, more precise locations are necessary to produce reliable final results. Larger aperture source arrays are more robust against location errors.

The results of the performed synthetic test to localize the point scatterers related to the observed sp- and ps-phases, demonstrated that if the coherency of the scattered phases are persistent over the SA elements, it is possible to image the scatterer location by measuring the semblance of the phases. However, even in a synthetic test, resolution of the resolved scatterers depends on the geometry of the experiment, that is the source array and receiver relative location and ray path of the phase. For instant, the semblance pattern calculated using ps-phase (Fig. 12) shows one distinct peak, but the peak is wider compared to the main peak observed for sp-phase (Fig. 11), that implies to a lower resolution in determining the scatterer location using ps-phase which is closer to the receiver than to the source region. This simple experiment demonstrated that to image different parts of an extended scatterer, or if there are many local scatterers, different phases and variable SA and receiver set ups are required. However in such cases, phase interference can violate the coherency of the phases which arrive in a specific time window.

In our experiments we did not include the error in origin time, as it was not available in our catalogue. However, in case of origin time error, aligning the waveforms based on *P*-phase beamforming will eliminate the timing errors. In this case, usually the value of the relative slowness of the reflected phases are measurable.

According to the test using real data, there are indications for scatterers in the velocity structure in the tested area, Northwest Bohemia/Vogtland. The scatterer location revealed by using ps-phases in this study (Fig. 21) is in agreement with findings of Rößler *et al.* (2008). Using the receiver array analysis (Rohrbach array) Rößler *et al.* (2008) concluded that backazimuth angle of

the converted phases can be deviated towards the north direction compared to the backazimuth angle of the direct phases, indicating the heterogeneity and inclined discontinuities along the ray path. Nevertheless, to get reliable image of the scatterers for such an application, we suggest using a collection of the optimal source arrays, by repeating the SA optimization algorithm several times. Considering other single receivers is also useful, to take into account variable SA to receiver geometries. Then, the search algorithm to discover the strong scatterers, using different SAs, should reveal consistent results in overall, to make reliable interpretation of the velocity structure.

For the examples given in this paper, the number of elements in SA is chosen to be 20 elements. In general more high quality array elements can increase the SNR of the coherent scattered phases, and subsequently increase the resolution of the scatterer location. Larger aperture helps in slowness resolution, however the coherency of the scattered phase signals might decrease because of heterogeneous model and different ray path. The optimal number therefore depends on the quality of the available data, the wavelengths the study wants to resolve and the heterogeneity in and around the source volume. To define the optimal number of a source array for the specific catalogue and region, the algorithm can be run with different numbers and the misfit functions can be evaluated and compared to find an optimal number.

SA technique can be applied whenever earthquake clusters are accurately located and contain events with similar waveforms. Often relative localizations are already available by applying a waveform cross-correlation and double difference location techniques. Effects due to varying source mechanisms can possibly be taken into account by deconvolving the source mechanism from measured waveforms. Examples for potential data sets include, beside midcrustal swarms as in NW Bohemia, deep seismic nests, as often observed in slabs beneath orogens as in Buccaramanga in Colombia (e.g. Prieto *et al.* 2012) or the Hindu Kush (e.g. Kufner *et al.* 2017). Near-surface swarms of earthquakes are often observed for induced seismicity (e.g. Cesca *et al.* 2014) and offer another possible future application of SA. Volcanic intrusions induced seismicity (e.g. Cesca 2019); is a third target of interest to apply the SA technique. The SA application examined here was aimed at the detection of scattered bodies near the source. However, SA is also interesting for illuminating structures at a greater distance from the source, for example using the so-called double beam technique, in which receiver and source arrays are processed simultaneously. For example, the slab location in a subduction zone with *S*-to-*P* phases (Kaneshima 2019) could be investigated, the Moho depth determination with PmP phases (Hrubcová *et al.* 2013), or low shear wave provinces with P (Pdiff) phases (Frost & Rost 2014).

The method may be further developed for future double array applications from vertical seismic profiles (Boué *et al.* 2013) using passive sources, or to define discrete, optimal source arrays in linear continuous seismic sensors as will be available for Distributed Acoustic Sensors (e.g. Jousset *et al.* 2018) and Ocean-Bottom Distributed Acoustic Sensing arrays (e.g. Williams *et al.* 2019).

ACKNOWLEDGEMENTS

We appreciate the anonymous reviewers and the handling editor for their thorough review and constructive comments and suggestions that significantly improved our paper. We acknowledge the University of Potsdam especially Daniel Vollmer and Matthias Ohrnberger for the data at the temporary array installed in Vogtland. We also acknowledge Potsdam Graduate School (PoGS) that supported the first author to finish this work.

REFERENCES

- Boué, P., Roux, P., Campillo, M. & de Cacqueray, B., 2013. Double beam-forming processing in a seismic prospecting context, *Geophysics*, **78**(3), V101–V108.
- Caramia, M. & Dell’ Olmo, P., 2008. Multi-objective optimization, in *Multi-Objective Management in Freight Logistics: Increasing Capacity, Service Level and Safety with Optimization Algorithms*, pp. 11–36, Springer-Verlag London.
- Cesca, S., 2019. Drainage of a deep magma reservoir near mayotte inferred from seismicity and deformation, *Nat. Geosci.*, **19**(2), 941–953.
- Cesca, S., Grigoli, F., Heimann, S., González, A., Bufo, E., Maghsoudi, S., Blanch, E. & Dahm, T., 2014. The 2013 September–October seismic sequence offshore Spain: a case of seismicity triggered by gas injection?, *Geophys. J. Int.*, **198**(2), 941–953.
- Dahm, T., Fischer, T. & Hainzl, S., 2008. Mechanical intrusion models and their implications for the possibility of magma-driven swarms in NW Bohemia Region, *Stud. Geophys. Geod.*, **52**(4), 529–548.
- Dahm, T., Hrubcová, P., Fischer, T., Horálek, J., Korn, M., Buske, S. & Wagner, D., 2013. Eger Rift ICDP: an observatory for study of non-volcanic, mid-crustal earthquake swarms and accompanying phenomena, *Scientific Drilling*, **16**, 93–99.
- Dodge, D.A. & Beroza, G.C., 1997. Source array analysis of coda waves near the 1989 Loma Prieta, California, mainshock: implications for the mechanism of coseismic velocity changes, *J. geophys. Res.*, **102**(B11), 24 437–24 458.
- Fischer, T. & Horálek, J., 2003. Space-time distribution of earthquake swarms in the principal focal zone of the NW Bohemia/Vogtland seismic region: period 1985–2001, *J. Geodyn.*, **35**(1–2), 125–144.
- Fischer, T., Horálek, J., Michálek, J. & Boušková, A., 2010. The 2008 West Bohemia earthquake swarm in the light of the WEBNET network, *J. Seismol.*, **14**(4), 665–682.
- Fischer, T., Horálek, J., Hrubcová, P., Vavryčuk, V., Bräuer, K. & Kämpf, H., 2014. Intra-continental earthquake swarms in west-bohemia and vogtland: a review, *Tectonophysics*, **611**, 1–27.
- Frost, D.A. & Rost, S., 2014. The p-wave boundary of the large-low shear velocity province beneath the Pacific, *Earth planet. Sci. Lett.*, **403**, 380–392.
- Goldstein, P., Walter, W.R. & Zandt, G., 1992. Upper mantle structure beneath central Eurasia using a source array of nuclear explosions and waveforms at regional distances, *J. geophys. Res.*, **97**(B10), 14 097–14 113.
- Hainzl, S., Fischer, T., Čermáková, H., Bachura, M. & Vlíček, J., 2016. Aftershocks triggered by fluid intrusion: evidence for the aftershock sequence occurred 2014 in west bohemia/vogtland, *J. geophys. Res.*, **121**(4), 2575–2590.
- Heimann, S. *et al.*, 2017. Pyrocko—an open-source seismology toolbox and library. V. 0.3. GFZ Data Services, <http://doi.org/10.5880/GFZ.2.1.2017.001>.
- Hensch, M., Riedel, C., Reinhardt, J. & Dahm, T., 2008. Hypocenter migration of fluid-induced earthquake swarms in the Tjörnes Fracture Zone (North Iceland), *Tectonophysics*, **447**(1–4), 80–94.
- Hiemer, S., Rüssler, D. & Scherbaum, F., 2012. Monitoring the West Bohemian earthquake swarm in 2008/2009 by a temporary small-aperture seismic array, *J. Seismol.*, **16**, 169–182.
- Horálek, J. & Šílený, J., 2013. Source mechanisms of the 2000 earthquake swarm in the West Bohemia/Vogtland region (Central Europe), *Geophys. J. Int.*, **194**(2), 979–999.
- Hrubcová, P., Vavryčuk, V., Boušková, A. & Horálek, J., 2013. Moho depth determination from waveforms of microearthquakes in the West Bohemia/Vogtland swarm area, *J. geophys. Res.*, **118**(1), 120–137.
- Hrubcová, P., Vavryčuk, V., Boušková, A. & Bohnhoff, M., 2016. Shallow crustal discontinuities inferred from waveforms of microearthquakes: method and application to ktb drill site and west bohemia swarm area, *J. geophys. Res.*, **121**(2), 881–902.
- Ishii, M., 2011. High-frequency rupture properties of the Mw 9.0 off the Pacific coast of Tohoku Earthquake, *Earth, Planets Space*, **63**, 609–614.
- Jousset, P. *et al.*, 2018. Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features, *Nat. Commun.*, **9**, doi:10.1038/s41467-018-04860-y.

- Kaneshima, S., 2019. Seismic scatterers in the lower mantle near subduction zones, *Geophys. J. Int.*, **218**(3), 1873–1891.
- Karamzadeh, N., Heimann, S., Dahm, T. & Krüger, F., 2018. Application based seismological array design by seismicity scenario modelling, *Geophys. J. Int.*, **216**(3), 1711–1727.
- Krentz, O., 2019. Saxon state ministry of environment and agriculture, <http://www.geologie.sachsen.de/schwarmbeben-im-vogtland-12796.html#a-26016>, Accessed: 2019-07-29.
- Krüger, F. & Ohrnberger, M., 2005. Tracking the rupture of the Mw = 9.3 Sumatra earthquake over 1,150 km at teleseismic distance, **435** (7044), 937–939.
- Krüger, F., Weber, M., Scherbaum, F. & Schlittenhardt, J., 1993. Double beam analysis of anomalies in the core-mantle boundary region, *Geophys. Res. Lett.*, **20**(14), 1475–1478.
- Krüger, F., Weber, M., Scherbaum, F. & Schlittenhardt, J., 1995. Evidence for normal and inhomogeneous lowermost mantle and core-mantle boundary structure under the arctic and northern Canada, *Geophys. J. Int.*, **122**(2), 637–657.
- Krüger, F., Scherbaum, F., Weber, M. & Schlittenhardt, J., 1996. Analysis of asymmetric multipathing with a generalization of the double-beam method, *Bull. seism. Soc. Am.*, **86**(3), 737–749.
- Krüger, F., Baumann, M., Scherbaum, F. & Weber, M., 2001. Mid mantle scatterers near the Mariana slab detected with a double array method, *Geophys. Res. Lett.*, **28**(4), 667–670.
- Kufner, S.-K., Schurr, B., Haberland, C., Zhang, Y., Saul, J., Ischuk, A. & Oimahmadov, I., 2017. Zooming into the Hindu Kush slab break-off: a rare glimpse on the terminal stage of subduction, *Earth planet. Sci. Lett.*, **461**, 127–140.
- Morita, Y., Nakao, S. & Hayashi, Y., 2006. A quantitative approach to the dike intrusion process inferred from a joint analysis of geodetic and seismological data for the 1998 earthquake swarm off the east coast of Izu Peninsula, central Japan, *J. geophys. Res.*, **111**(6), 1–18.
- Nakata, N., Boué, P., Brenguier, F., Roux, P., Ferrazzini, V. & Campillo, M., 2016. Body and surface wave reconstruction from seismic noise correlations between arrays at piton de la fournaise volcano, *Geophys. Res. Lett.*, **43**(3), 1047–1054.
- Neunhöfer, H. & Hemmann, A., 2005. Earthquake swarms in the Vogtland/Western Bohemia region: spatial distribution and magnitude-frequency distribution as an indication of the genesis of swarms?, *J. Geodyn.*, **39**(4), 361–385.
- Niazi, M., 1969. Use of source arrays in studies of regional structure, *Bull. seism. Soc. Am.*, **59**(4), 1631.
- Okabe, A., Boots, B. & Sugihara, K., 1992. *Spatial Tessellation: Concepts and Applications of Voronoi Diagrams*, John Wiley and Sons.
- Passarelli, L. et al., 2015. Aseismic transient driving the swarm-like seismic sequence in the Pollino range, Southern Italy, *Geophys. J. Int.*, **201**(3), 1553–1567.
- Prieto, G.A., Beroza, G.C., Barrett, S.A., López, G.A. & Florez, M., 2012. Earthquake nests as natural laboratories for the study of intermediate-depth earthquake mechanics, *Tectonophysics*, **570**, 42–56.
- Revenaugh, J., 1995. A scattered-wave image of subduction beneath the transverse ranges, *Science*, **268**(5219), 1888–1892.
- Rietbrock, A. & Scherbaum, F., 1999. Crustal scattering at the ktb from a combined microearthquake and receiver analysis, *Geophys. J. Int.*, **136**(1), 57–67.
- Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.*, **40**(3), 1–2.
- Rost, S. & Thomas, C., 2009. Improving seismic resolution through array processing techniques, *Surv. Geophys.*, **30**(4-5), 271–299.
- Roux, P., Cornuelle, B.D., Kuperman, W. & Hodgkiss, W., 2008. The structure of raylike arrivals in a shallow-water waveguide, *J. acoust. Soc. Am.*, **124**(6), 3430–3439.
- Rößler, D., Hiemer, S., Bach, Ch., Delavoud, E., Krüger, K., Ohrnberger, M., Sauer, D., Scherbaum, F. & Vollmer, D., 2008. Rohrbach/Vogtland Seismic Array 2008, Universität Potsdam.
- Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.*, **138**(2), 479–494.
- Scherbaum, F., Gillard, D. & Deichmann, N., 1991. Slowness power spectrum analysis of the coda composition of two microearthquake clusters in northern Switzerland, *Phys. Earth planet. Inter.*, **67**(1–2), 137–161.
- Scherbaum, F., Krüger, F. & Weber, M., 1997. Double beam imaging: mapping lower mantle heterogeneities using combinations of source and receiver arrays, *J. geophys. Res.*, **102**(B1), 507–522.
- Schweitzer, J., Fyen, J., Mykkeltveit, S. & Kväerna, T., 2012. Seismic arrays, New Manual of Seismological Observatory Practice 2 (NMSOP-2), pp. 1–80.
- Spudich, P. & Bostwick, T., 1987. Studies of the seismic coda using an earthquake cluster as a deeply buried seismograph array, *J. geophys. Res.*, **92**(B10), 10 526–10 546.
- Virtanen, P., 2019. SciPy 1.0—fundamental algorithms for scientific computing in python, arXiv:1907.10121.
- Wathelet, M., 2008. An improved neighbourhood algorithm: parameter conditions and dynamic scaling, *Geophys. Res. Lett.*, **35**(9), 1–5.
- Weber, M. & Wicks, C.W., 1996. Reflections from a distant subduction zone, *Geophys. Res. Lett.*, **23**(12), 1453–1456.
- Williams, E., Fernandez-Ruiz, M.R., Magalhaes, R., Vanthillo, R., Zhan, Z., Gonzalez-Herrera, M. & Martins, H.F., 2019. Teleseisms and microseisms on an ocean-bottom distributed acoustic sensing array, EarthArXiv, doi:10.31223/osf.io/kg7q4.

APPENDIX: WEIGHTING COEFFICIENTS CALCULATION

A multi-objective optimization problem is defined in mathematical terms as:

$$\min [f_n(\mathbf{m})]; \quad \mathbf{m} \in S, \quad (\text{A1})$$

where $\mathbf{m}$ is the model space, $f_n(\mathbf{m})$ is a set of objective functions with index n in the number and S is a set of constraints. One strategy to solve a multi-objective problem is combine all objective functions in a single scalar function using the weighting sum technique. To define the weighting coefficients, we can start by evaluating them between two functions according to (Karamzadeh et al. 2018), and then increasing the number of functions step by step:

$$F_{1,2} = \gamma_1 f_1 + (1 - \gamma_1) f_2, \quad (\text{A2})$$

where $0 < \gamma_1 < 1$ is the weighting factor, $F_{1,2}$ is the combined scalar function for f_1 and f_2 . We can proceed the procedure using $F_{1,2}$ and f_3 , to make $F_{1,2,3}$:

$$F_{1,2,3} = \gamma_2 F_{1,2} + (1 - \gamma_2) f_3, \quad (\text{A3})$$

and, to include f_4 :

$$F_{1,2,3,4} = \gamma_3 F_{1,2,3} + (1 - \gamma_3) f_4, \quad (\text{A4})$$

$F_{1,2,3,4}$ can be written as:

$$F_{1,2,3,4} = \gamma_3 \gamma_2 \gamma_1 f_1 + \gamma_3 \gamma_2 (1 - \gamma_1) f_2 + \gamma_3 (1 - \gamma_2) f_3 + (1 - \gamma_3) f_4. \quad (\text{A5})$$

Finally, for n objective functions the final combined scalar function is formulated as:

$$F_{1,2,\dots,n} = \gamma_{n-1} \gamma_{n-2} \dots \gamma_1 f_1 + \gamma_{n-2} \gamma_{n-3} \dots \gamma_2 (1 - \gamma_1) f_2 + \dots + \gamma_{n-1} (1 - \gamma_{n-2}) f_{n-1} + (1 - \gamma_{n-1}) f_n. \quad (\text{A6})$$