Data-driven constraints on earthquake modeling and rupture segmentation from teleseismic multi-array backprojection and InSAR

Andreas Steinberg[1], Henriette Sudhaus[2], and Frank Krüger[3]

[1]Federal Institute for Geosciences and Natural Resources
[2]Christian Albrechts Universität zu Kiel
[3]University of Potsdam

November 23, 2022

Abstract

Earthquakes have been observed to rupture in segments. A good understanding of rupture segmentation is important to characterize fault geometries at depth for follow-up tectonic, stress-field or other analyses. Earthquakes with magnitudes Mw<7 are however often modeled with simple source models. We propose a data-driven strategy and develop pre-optimization methods for a segmentation-sensitive source modeling analysis.

The first method we develop is a time-domain, multi-array backprojection of teleseismic data to infer the spatio-temporal evolution of the rupture, including a potential occurrence of rupture segmentation. We calibrate the backprojection using empirical traveltime corrections and we provide robust precision estimates based on bootstrapping of the travel-time models and array weights. Secondly we apply image analysis methods on InSAR surface displacement maps to infer modeling constraints on rupture characteristics (e.g. strike and length) and the number of potential segments.

Both methods can provide model-independent constraints on fault location, dimension, orientation and rupture timing, applicable to form prior probabilities of model parameters before modeling.

We use the model-independent constrains delivered by these two newly developed methods to inform a kinematic earthquake source optimization about parameter prior probability estimates.

We demonstrate and test our methods based on synthetic tests and an application to the 25.11.2016 Muji Mw 6.6 earthquake. Our results indicate segmentation and bilateral rupturing for the 2016 Muji earthquake. The results of the backprojection using high-frequency filtered teleseismic wavforms in particular shows the capability to illuminate the rupture history with the potential to resolve the start and stop phases of individual fault segments.

Data-driven constraints on earthquake modeling and rupture segmentation from teleseismic multi-array backprojection and InSAR

Andreas Steinberg[1,2], **Henriette Sudhaus**[1] **and Frank Krüger**[3]

[1]Christian Albrechts University Kiel, Otto-Hahn-Platz 1, Germany
[2]Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany
[3]Institute for Earth and Environmental Sciences, University of Potsdam, Potsdam, Germany

Key Points:

- We develop a teleseismic multi-array backprojection method to constrain the spatio-temporal rupture evolution and segmentation occurrence
- We use image segmentation methods to analyse InSAR displacement gradients and obtain modeling constraints
- Joint data modeling for the 2016 Muji earthquake with the model independent constraints finds a bilateral and segmented rupture

Corresponding author: Andreas Steinberg, andreas.steinberg@ifg.uni-kiel.de

Abstract

Earthquakes have been observed to rupture in segments. A good understanding of rupture segmentation is important to characterize fault geometries at depth for follow-up tectonic, stress-field or other analyses. Earthquakes with magnitudes Mw<7 are however often modeled with simple source models. We propose a data-driven strategy and develop pre-optimization methods for a segmentation-sensitive source modeling analysis. The first method we develop is a time-domain, multi-array backprojection of teleseismic data to infer the spatio-temporal evolution of the rupture, including a potential occurrence of rupture segmentation. We calibrate the backprojection using empirical traveltime corrections and we provide robust precision estimates based on bootstrapping of the travel-time models and array weights. Secondly we apply image analysis methods on InSAR surface displacement maps to infer modeling constraints on rupture characteristics (e.g. strike and length) and the number of potential segments. Both methods can provide model-independent constraints on fault location, dimension, orientation and rupture timing, applicable to form prior pseudo-probabilities of model parameters before modeling. We use the model-independent constrains delivered by these two newly developed methods to inform a kinematic earthquake source optimization about parameter prior pseudo-probability estimates. We demonstrate and test our methods based on synthetic tests and an application to the 25.11.2016 Muji Mw 6.6 earthquake. Our results indicate segmentation and bilateral rupturing for the 2016 Muji earthquake. The results of the backprojection using high-frequency filtered teleseismic wavforms in particular shows the capability to illuminate the rupture history with the potential to resolve the start and stop phases of individual fault segments.

1 Introduction

The accuracy in estimating earthquake source characteristics is limited by many factors. Among them are a limited data resolution, non-linear dependencies between observations and some of the sought source parameters as well as simplifications applied to a model representation compared to the real rupture process (Steinberg et al., 2020). Also, based on surface observations alone, uncertainties in the earth structure influence the accuracy of earthquakes source estimation strongly (Weston et al., 2012) and some earthquake properties can not be resolved independently from others. Continuous progress is made regarding the data resolution, because the density of global sensors is increasing steadily. This enables more detailed studies of shallow crustal earthquakes of moderate magnitude and allows applying more realistic earthquake models that represent better potentially common source complexities such as segmentation into sub-sources and slip heterogeneities. The challenges of solving the non-linear problem and dealing with parameter dependencies remain, in particular for source analyses with complex segmented models that involve the estimation of a large number of model parameters (Weston et al., 2012; Ragon et al., 2018; Lohman & Simons, 2005; Razafindrakoto et al., 2015).

Most current operational earthquake analysis frameworks (Dziewonski et al., 1981; Hanka & Kind, 1994) providing earthquake catalogues, only consider point-source models to represent any given earthquake (e.g. a single Double-Couple or moment tensor). It has been shown that in presence of significant source complexity also the apparent earthquake characteristics based on point-source or single-source kinematic models can be significantly biased (Steinberg et al., 2020). As these earthquake catalogues form the basis for many statistical studies on earthquake characteristics (i.e. (Heidbach et al., 2018; Woessner et al., 2015) and inferred dependent properties like spatio-temporal aftershock patterns (McCloskey & Nalbant, 2009), the observational bias from single earthquakes could introduce a bias in the currently possibly incomplete earthquake statistics. Inferred source behaviour used in dynamic modeling and empirical analysis is often based on statistics derived from kinematic modeling. As mentioned above, these statistics however might be influenced by observational bias. Rupture segmentation is an important source characteristic which is often based on expert judgement in case studies of larger earthquakes.

An objective and data-driven study of the segmentation of shallow crustal moderate magnitude earthquakes is difficult but necessary to increase these statistics on source complexity. Such a study approach should be data driven and minimize expert bias (e. g. the choice of the model and complexity). This undertaking is challenged by a strongly enlarged model space to be sampled and by increased parameter trade-offs compared to point-source or single-source kinematic inversion. Very slow converging or even non-convergent optimizations can be the consequence.

In this study we demonstrate how model-independent and data-based methods can be employed to inform kinematic modeling of earthquake sources objectively. We put a special focus on the minimum modelled segmentation required to meaningfully represent earthquake ruptures. Our here presented methods are designed to enable the investigation of rupture segmentation with globally available datasets, e. g. space-borne InSAR data of co-seismic static near-field displacements and broadband recordings at distant seismological stations. The suggested methods extract information on the earthquake source in a pseudo-probabilistic way. This information can be used to judge on the occurrence of fault segmentation independent from inverse modeling and to enable enriched statistical analyses of medium-sized earthquakes in an effort to reduce potential observational bias. We use this source information further on to set up the model parametrization of earthquake source optimizations, which includes the number of relevant model parameters and their prior pseudo-probabilities.

We present a multi-array backprojection (BP) approach based on teleseismic waveform data to image the location and dynamics of a rupture. From the evolution of the rupture dynamics we aim to detect the number of significant sub-sources. Seismological backprojection takes advantage of source-receiver reciprocity and has proven to be a reliable tool to image the dynamic rupture process by mapping coherent seismic radiators in space and time (Kiser & Ishii, 2017). The principal idea of the seismic backprojection method here used is the alignment and then stacking of the seismic waveforms to the predicted P-wave and SH-wave onsets of potential point sources located on a 2D or 3D grid in sliding time windows. If energy is coherently emitted in a certain time window from a certain grid point, the time-shifted waveforms should stack constructively. The grid point is therefore a potential source of the signal at that time. The waveforms should stack destructively if the grid point is not a source of seismic energy during the given time window. Backprojection of teleseismic data has first been used to investigate the 26.12.2004 Mw 9.1 Sumatra earthquake (Krüger & Ohrnberger, 2005; Ishii et al., 2005) and is usually carried out for larger earthquakes (Mw >7) (Bao et al., 2019; Meng et al., 2016; Kiser & Ishii, 2017; Hicks et al., 2020).

Seismological backprojection is applied in different frequency bands of the seismic waves. From frequencies below the corner frequency we gain prior information on the fault location and potentially also the number of sub-sources. High-frequency energy radiation is concentrated near the hypocenter and the asperities rupture initiation points, representing start/stop phases (Ide, 2002; Madariaga, 1977; Okuwaki & Yagi, 2017). The mapping of higher frequency coherent energy release can therefore potentially be used as prior information on the ruptures nucleation position, rupture velocity and the number of sub-sources for a kinematic fault model optimization. Seismological backprojection is an ideal tool to inform our modeling about rupture segmentation and validate that modelled rupture segmentation is not only a requirement to fit our models better but really representative of an actual physical process. Because traditional seismological backprojection uses a single array and is known to produce "swimming" artifacts, we implemented a new multi-array backprojection method resistant to this effect (Kiser & Ishii, 2017) based on an earlier approach (Rössler et al., 2010).

Static surface displacement as measured through the InSAR technique can reveal an earthquake source location by apparent significant displacement. To the eye of an expert the pattern of the displacement potentially reveals more characteristics directly, such as the approximate rupture dimension, the fault orientation and the mechanism. We mimic, formalize and automate a similar extraction of information prior to modeling by employing image analysis methods like edge detection on the gradient of the displacement. Using the presented method we estimate the source location, size and the number of sources from the gradient of displacement maps among other source features.

We first present the two data-driven analysis methods developed for far-field and near-field data. The methods are implemented in python-based open-source software codes. We test the methods with synthetic data first. We then present a framework with a focus on moderate and larger sized shallow crustal earthquakes in mind in which we use the extracted information for estimating model parameter prior pseudo-probabilities to guide a finite fault optimization and constrain the modeling of segmented ruptures. We finally apply the presented methods in an investigation of the 25.11.2016 Muji Mw 6.6 earthquake to better inform a joint optimization of teleseismic and static near-field data.

2 Methods

2.1 Time-domain backprojection using multiple virtual arrays

The reported applications of teleseismic backprojection (BP) enclose only a few studies dealing with shallow crustal intermediate-sized earthquakes of magnitudes between $M_W 6$ and $M_W 7$ (Kiser & Ishii, 2013; Fan & Shearer, 2017; Yin & Denolle, 2019). A likely reason is that the spatial precision of traditional time-domain teleseismic BP by using large arrays are similar to the size of the rupture area of Mw<7 earthquakes (Fan & Shearer, 2017).

2.1.1 Introduction to the backprojection method

Traditional time-domain BP involves an alignment of seismic recordings within an array and a subsequent stacking (Krüger & Ohrnberger, 2005; Ishii et al., 2005). Phase arrivals of earthquakes stack constructively to high amplitudes if the trace alignments correspond well to the actual source-receiver configuration. Different phase arrivals are separated by moving time windows along the waveform on which the BP is applied. Then, mapping the source locations that lead to high-amplitude stacks for the corresponding time window provides images of the seismic energy release of a rupture. This energy originates from abrupt relative and spatially variable changes in the fault slip or abrupt changes in rupture velocity (Okuwaki et al., 2018; Yin & Denolle, 2019; Madariaga, 1977).

The main assumption of time-domain BP is that wave traveltimes from the source to global receivers correspond well to those of commonly used 1-D velocity Earth models. Unwanted effects of this strong assumption for real data applications can be weakened by applying empirical traveltime corrections (Section 2.1.4). Other common assumptions are that the waveforms of phases are coherent within an array, e. g. no occurrences of polarity changes as across a nodal plane of the focal mechanism, and furthermore that noise is uncorrelated. Constructive stacking of coherent coda waves can create secondary sources and introduce a bias in the time-domain BP imaging.

We implemented the time-domain BP in the following way. We stack P- and SH-phases separately, using their respective slowness values. Other phases with different slowness values inherently stack destructively. The depth phases of P and SH-phases, pP, pPP and sS, however, have similar slowness values as the corresponding direct phase for shallow events and will therefore also stack constructively. Depth phases can have relatively large amplitudes compared to the direct phase and, for shallow earthquakes, follow them very close in time. Therefore they generally significantly influence the stack of the direct phase. With higher frequencies the importance of depth phases decreases, because these are more strongly influenced by topography and shallow structure at the surface reflection point, which results in less coherent high-frequency waveforms with reduced constructive stacking.

We use the phase-weighted stacking method (Schimmel & Paulssen, 1997) to increase the signal-to-noise ratio of the stacks, which basically realizes a trace weighting based on the phase coherence within the array. Specifically, the phase-weighted stacking is a non-linear stacking method, where each sample in a linear stack is weighted by an amplitude-unbiased coherence measure. In this way, phase-weighted stacking sharpens up signals, reduces signal artifacts and suppresses noise. Phase-weighted stacking comes at the cost of loss of absolute amplitude information (Fan & Shearer, 2017; Schimmel & Gallart, 2007) and of a strong relative enhancement of the dominant period. However, the advantages outweigh the disadvantages of the method.

We first calculate the coherence based on complex traces in a phase stack and then multiply this coherence with the linear stack, sample by sample. Therefore we first calculate the phase stack $c(t)$ for all N waveforms:

$$c(t) = \| \frac{1}{N} \sum_{j=1}^{N} e^{i\Phi_j(t)} \|. \tag{1}$$

$c(t)$ is based on the similarity of the phases $\Phi_j(t)$ of the complex signals of the N traces $u_j(t)$ at time t (Bracewell & Bracewell, 1986). The amplitudes of the phase stack are coherence measures and range between 0 for non-coherent and 1 for coherent signals.

We carry out the BP for point locations that form a horizontal grid of source points. We stack the waveforms for each of these grid points according to Eq. 2, with the specific expected arrival time from a grid point source to each station. Each waveform $u(t_\mathrm{r})$ of the linear stack is multiplied with the phase coherence (Schimmel & Paulssen, 1997) to calculate for each grid point the phase-weighted stack $\hat{S}_k(t_\mathrm{r})$ of an array:

$$\hat{S}_k(t_\mathrm{r}) = \frac{1}{N} \sum_{j=1}^{N} u_j(t_\mathrm{r} + t_{kj}) \| \frac{1}{N} \sum_{j=1}^{N} e^{i\Phi_j(t_\mathrm{r}+t_{kj})} \|^{\nu}, \tag{2}$$

with t_r being the rupture onset time and t_{kj} the source-receiver traveltime. The coherence weighting here is tuned with the parameter ν for an adaptable transition between coherent and less coherent signal summation. ν=0 realizes a linear stack, while we use ν=2 to increase the signal-to-noise ratio. t_{kj} is the traveltime for the respective grid point k of the waveform record j.

Waveforms that get stacked in this way form the semblance S_k of the array for the respective source grid point k by normalization:

$$S_k(t_\mathrm{r}) = \frac{\hat{S}_k(t_\mathrm{r})}{\sum_{k=1}^{K} \hat{S}_k(t_\mathrm{r})}. \tag{3}$$

The semblance $S_k(t_\mathrm{r})$ can be seen as in terms of a (pseudo-)probability of coherent radiation of seismic energy from a given source point k at a time t_r (Rössler et al., 2010; Douze & Laster, 1979). We can form maps of spatial semblance for single time steps or of the cumulative semblance, to which we refer to as incremental or cumulative semblance maps. The semblance spatial resolution is described by the frequency- and azimuth-dependent beam pattern and is an analogue to the *array response* or *array transfer function* (Rost & Thomas, 2002), defined by stacking with respect to slowness (Johnson & Dudgeon, 1993). The spatial resolution of a seismic array increases with array aperture as well as with frequency and aliasing is decreased with increasing station coverage (Rost & Thomas, 2009). Therefore, large and dense arrays are desirable to image rupture evolution, but there are limits. The use of very large arrays has been found to result in relatively low resolution of the semblance (Xu et al., 2009). The reason is that the waveform recordings from a very large range of source distances and source azimuths resemble each other less and less and loose their coherence. This coherence loss is stronger for high frequencies and leads to a decrease of the upper frequency that remains coherent (Rost & Thomas, 2009). Less high frequency content in the semblance decreases the spatio-temporal resolution as mentioned above.

Additionally, we calculate the beampower $E(t_\mathrm{r})$, which is an absolute measure of the amplitudes at the *i*th array and a time window centered around *t* for waveforms **u** recorded at N stations of an array. Beampower is the sum of the energy at all arrays:

$$E(t_\mathrm{r}) = \sum_{k=1}^{K} \frac{1}{L+1} \sum_{l}^{L} \left| \frac{1}{N} \sum_{j=1}^{N} u_j(t_\mathrm{r} + l + t_{kj}) \right|^2, \tag{4}$$

where l is the sample index of the waveform in the time window with total number of samples L, Δt is the duration of the time window. The beampower time trace in our case is closely related to the source-time-function (STF) as it is a stack of body waves. This should be similar to a STF but scaled by a factor depending on the radiation pattern, the source-receiver distance and the elastic medium properties (Vallée & Douet, 2016).

2.1.2 Multi-array BP method

The combination of several arrays subdues side lobes of the array response compared to a single-array BP. It also minimizes the effect of azimuth-dependent "smear" or “swimming” artifacts, which are systematic apparent drifts of the energy towards the array (Meng et al., 2016). The reasons are that the sidelobes of the single-array response functions are at different positions for each array, while the central lobe is always at the same position in the slowness plane. Migration artifacts "swim" in different directions with different apparent velocities. Additionally the combination of P- and SH-phases BPs suppresses sidelobes and migration artifacts, because of the different delays between the P-phase depth phases (pP, sP) and the SH-phase depth phases (sS) (Hong & Fujita, 1981). Multi-array BP results in more certain and better resolved spatio-temporal imaging.

In our multi-array BP we cluster all globally available stations at reasonable teleseismic distances to form a multitude of small virtual arrays using the k-means algorithm (Steinhaus, 1956). The combination of many small virtual arrays has the advantage of minimizing the effect of velocity differences between stations in the array as well as the effect of radiation patterns and source directivity across arrays (Rössler et al., 2010). Virtual arrays are formed assuming a lower limit for the number of array stations, distance between stations and a maximum aperture. Included stations are part of one array only. We multiply the single-array semblance maps instead of adding them which further suppresses sidelobes in the multi-array response function and is related to the interpretation of semblances as relative, non-normalized pseudo-probability (Rössler et al., 2010). The multiplication of the array responses also corresponds to a multiplication of the transfer functions of the arrays (Rössler et al., 2010).

We calculate the multi-array semblance from the product of the semblances from M arrays:

$$S_k(t_\mathrm{r}) = \prod_{m=1}^{M} S_{km}(t_\mathrm{r}). \tag{5}$$

The global distribution of virtual arrays may have gaps. To avoid an azimuthal bias in the multi-array response function we subdivide the azimuth into 12 sectors and, based on the azimuth of the earthquake epicenter to the array center, each array is assigned to a corresponding azimuth sector. The semblance from all virtual arrays in each azimuth sector is normalized to 1 for each time window, so that each azimuth sector has the same influence on the combined semblance. The azimuth weight $w_{\mathrm{azi},m}$ for the mth virtual array is then:

$$w_{\mathrm{azi},m} = \frac{S_{mk}(t_\mathrm{r})}{\sum_{m=1}^{M} \max\limits_{k..K,t..T}(S_m(t,k))}. \tag{6}$$

The weighted semblance becomes:

$$S_k(t_\mathrm{r}) = \prod_{m=1}^{M} w_{\mathrm{azi},m} \cdot S_{km}(t_\mathrm{r}). \tag{7}$$

Multi-array BP is associated with uncertainties, particularly for locating the source of energy, that we want to account for. Several studies have investigated these limitations. Yin and Denolle (2019) found from theoretical considerations that the minimum resolvable feature in a semblance map should have a dimension of at least twice the P-wave wavelength. The resolution length of beamforming, which is the minimum distance between sources that can be distinguished, is estimated as the width at half-peak amplitude of the main lobe of the array response function (Meng, Ampuero, Sladen, & Rendon, 2012). The array's spatial accuracy is the error in the estimation of the true source location (Meng, Ampuero, Sladen, & Rendon, 2012). (Fan & Shearer, 2017) found a median location error of around 25 km for traditional time-domain BP using large arrays. They also found, from the methods they considered, that the best sub-event resolution is achieved if a global phase-weighted stack is used (Fan & Shearer, 2017). To estimate the spatial precision and accuracy of a BP for individual cases, bootstrapping methods have been used (Yao et al., 2012; Shearer, 1997; D. Wang et al., 2016; Meng, Ampuero, Stock, et al., 2012).

We use Bayesian bootstrapping (Rubin, 1981) to quantify the spatial and temporal precision of the multi-array BP results. The bootstrapping is applied to the weights of the combined virtual arrays in the multi-array semblance at each timestep. For each timestep we create a set of 100 differently weighted BP stack. This bootstrapping of the weights is then further combined with traveltime perturbations to asses the impact of the velocity model choice. The array weight controls how strongly each virtual array contributes to the multi-array semblance (see Eq. 10). We draw n_{clusters} random real numbers $r \in [0, n_{\mathrm{clusters}}]$ from a uniform distribution. We then sort the obtained random values in an ascending order and ensure $r_1 = 0$. The ith bootstrap weight $w_{\mathrm{boot},i_{\mathrm{boot}}}$ for a virtual array is then defined as:

$$w_{\mathrm{boot},i_{\mathrm{boot}}} = r_{i_{\mathrm{boot}}+1} - r_{i_{\mathrm{boot}}}. \tag{8}$$

The traveltime perturbation simulates the effect of model errors in the semblance that are introduced by assuming a 1-D velocity model in the phase alignment before stacking. We assume these traveltime errors to be random and normally distributed, with a standard-deviation of 2 s for the P-phase arrivals and with 4 s twice as large for SH-phase arrivals. So in each bootstrap set, we apply these traveltime shifts to the waveforms before stacking. The semblance of each array is therefore affected by the bootstrapping of the traveltimes and we arrive at 100 sets of semblances for each virtual array. From the chosen 100 bootstrapping realizations we get from each bootstrap the weight $w_{\mathrm{boot},m}$ and also take into account the azimuthal balance weights $w_{\mathrm{azi},m}$ to calculate the combined weighted semblance:

$$S(t_{\mathrm{r}}) = \prod_{m=1}^{\mathrm{M}} \frac{1}{100} \sum_{i=1}^{100} w_{\mathrm{boot},mi} \cdot w_{\mathrm{azi},mi} \cdot S_{mi}(t_{\mathrm{r}}). \tag{9}$$

We combine the results for each timestep from the individual P- and SH-phase BP to phase-combined BP. The phase-combined multi-array semblance $S_{\mathrm{comb}}(t_{\mathrm{r}})$ from P- and SH-phases BP is derived by the multiplication of the semblances at each timestep t_{r} (after Eq. 9):

$$S_{\mathrm{comb}}(t_{\mathrm{r}}) = S_{\mathrm{P}}(t_{\mathrm{r}}) \cdot S_{\mathrm{S}}(t_{\mathrm{r}}). \tag{10}$$

The final result of the multi-array BP obtained is a phase-combined multi-array semblance map for each timestep of the BP, which we call time-incremental semblance maps. The earliest mapped occurrence of coherent energy release is likely located close to the nucleation point, indicating the start phase. The latest semblance peak is likely to represent the stop phase. We also combine the semblance from all timesteps in a single cumulative semblance map.

2.1.3 Backprojection settings

We consider waveforms from broadband stations between 28 degree and 93 degree distance from the source, to avoid phase triplications and having P- and SH-wave arrivals clearly separated from later arriving bodywaves with significant amplitudes. After removing the instrument response from velocity seismograms through restitution, the waveforms are rotated from an ENZ (east, north, vertical) into the RTZ (radial, transversal, vertical) coordinate system, and they get downsampled to 10 Hz. We select the Z-component of the records for the P-wave BP and the T-component for the SH-wave BP. The virtual arrays have a maximum aperture of 5° and a minimum number of 5 stations. If in any of the 12 azimuth sectors no virtual array can be formed, to increase the azimuthal coverage, we allow for successively larger array apertures up to $10°$ and a smaller number of stations down to 3. Generally, we perform separate low-frequency BPs (LF BP) and high-frequency BPs (HF BP), within a total frequency range of 0.003 Hz up to about 1.5 Hz. The LF BP and HF BP frequency bands are separated by the expected corner frequency f_c of the earthquake studied. We estimate the corner frequency f_c of the seismic radiation based on the rupture duration T_r, which we take from the GCMT catalog, with $f_\mathrm{c} = 2/T_\mathrm{r}$ (Aki & Richards, 2002). In other words, f_c is the upper frequency limit of the LF BP records and the lower frequency limit of the HF BP records. We filter with a butterworth bandpass filter of fourth order.

The BP is carried out with fixed-length time windows over the recorded waveforms, which are moved with small timesteps of a few seconds. The window length depends on the longest period at which the data is filtered. This results in longer time windows for the LF BP of 20 s to 30 s and shorter time windows for the HF BP of around 10 s. At this stage we discard seemingly poor quality records. As a measure we calculate the cross-correlation between the P- and SH-wave records of an array station with the corresponding records of the center-most station within the array. We only include waveforms with a cross-correlation coefficient of at least 0.6. By chance, this center-most station can have serious quality problems itself, which would lead to low cross-correlations for all stations and in effect to an exclusion of all waveforms of the array. In such a case the center-most station is excluded and the cross-correlation coefficients are recalculated for a randomly chosen new reference station. Before the stacking of the time windows (Eq. 2), the waveforms in each array are aligned based on pre-calculated traveltime tables using the AK-135 1-D velocity model.

We multiply the responses of all virtual arrays at each timestep, which gives the response of the global array (Eq. 10). We account for the effect of unmodelled site effects close to the stations by including traveltime perturbations per station (traveltime shift bootstrapping). The effect of systematic traveltime shifts due to unmodelled large-scale 3-D path effects can be reduced with empirical traveltime corrections.

2.1.4 Empirical traveltime corrections

Large-scale 3-D velocity structures affect the waveform paths and the traveltimes in systematic ways for stations within an array and across neighboring arrays. By using traveltime predictions based on 1-D velocity models only, residual systematic traveltime shifts persist.

If these time shifts remain unaccounted for, they may produce spatially and temporally, significantly shifted and defocused BP results. We can reduce this bias by automatically calibrating the traveltimes for each station based on empirical traveltime shifts (Palo et al., 2014; Ishii et al., 2007; Meng et al., 2016; Fan & Shearer, 2017).

We estimate arrival time shifts empirically by selecting a cataloged reference event from single fore- or aftershocks, which occurred close to the investigated earthquake. For this reference event we assume that catalog location and time are accurate and fix them to a single grid point and a single time window. Unknown is a set of traveltime shifts, maximizing the semblance for this setup. Per array and for each station individually we vary the traveltime shifts of P- and SH-waves such that the single-array semblance at the reference location is maximized. For this optimization problem we use the differential evolution algorithm (Storn & Price, 1997). We allow the traveltime to vary by +/- 2 s for the P-wave and +/- 4 s for the SH-wave, with respect to the theoretical onset, assuming the same 1-D velocity model as for the main event BP.

For a successful traveltime correction, the reference event has to be near the studied earthquake, small enough such that the point-source approximation holds well, but also large enough that its phase arrivals have a high signal-to-noise ratio globally. Preferably, the location of the reference event is very well known from local or regional station data analysis. Ideally, the reference event has a similar focal mechanism compared to the main event. An error in the location of the reference event will cause a wrongly estimated global time shift for the phases of the main studied earthquake. Typical mislocation errors that have to be expected for smaller earthquakes in remote areas are of the order of tens of kilometers (Fan & Shearer, 2017; Palo et al., 2014). An important gain of the empirical traveltime correction is an increased phase coherence, because unmodelled path effects are generally well compensated (Palo et al., 2014).

The rupture dimensions of intermediate-sized earthquakes, which we primarily want to investigate, are of the order of several tens of kilometers. When applying timing corrections for larger earthquakes however the validity of the timing corrections is spatially limited to a spatial extent of several tens of kilometers (Fan & Shearer, 2017). For earthquakes that rupture larger areas therefore traveltime corrections from multiple fore- or aftershocks along the potential rupture area should be considered (Palo et al., 2014).

2.1.5 Defining the model space based on backprojection results

We propagate the information on the source obtained using multi-array BP to define the model space for a direct search optimization. These BP results are the low-frequency (LF) and high-frequency (HF) time-incremental semblance maps, which include 100 bootstrap realisations of the semblance for each time step. From these semblance maps we extract information on the location of the rupture, rupture size and orientation as well as the rupture evolution in time.

The centroid location and dimension of the rupture are outlined by significant cumulative LF semblance, which maps the area of significant seismic energy release. The semblance values can be related to relative pseudo-probability (Rössler et al., 2010). So based on the LF semblance grid, we construct discrete pseudo-probability functions for the longitude and latitude parameters of the source centroid location.

For the strike and length source parameters we estimate the prior distributions using different approaches on LF and HF semblances, and finally combine their results afterwards. Based on the LF cumulative semblance map, we fit arbitrarily one or more oriented minimum bounding boxes and minimum bounding ellipses to neighboring grid point values with semblance values that exceed 1% of the maximum semblance. The bounding box orientations and lengths of the major axis enter as single values in the estimation of the strike and length prior distributions, respectively. We calculate azimuth and distance from consecutive HF increments between their semblance maxima, for all bootstrap realisations. This produces an ensemble of azimuth and distance values, which translate directly into probable strike and length values. We simply merge the single estimates for strike and length from the cumulative LF semblance and the ensemble of estimates from the incremental HF semblances to construct Gaussian prior distributions for these values.

Rupture velocity can be inferred from the calculated distance and time separation of subsequent time-incremental maxima in the HF semblance maps (Ishii et al., 2005; Meng, Ampuero, Sladen, & Rendon, 2012). We assume here that between semblance maxima the rupture velocity does not change. We estimate the rupture velocity by dividing the collected apparent distances through the difference of the occurring times. Similar to (Rössler et al., 2010) we select the start, stop and duration of the rupture using a STA/LTA trigger algorithm on the obtained combined HF semblance map but also on the individual bootstrap semblance maps before combining as in Eq. 9. We divide the largest distance between two semblance maxima by their time separation and obtain another set of average rupture velocity estimates. From this procedure one could pick up the ensemble of apparent source durations as well.

The location of the earthquake hypocenter, or nucleation point, and the relative onset time of the rupture are also parameters we want to estimate in an optimization. To retrieve prior pseudo-probability distributions of the horizontal nucleus position we use incremental HF semblance maps normalized to a two-dimensional, discrete pseudo-probability function. We extract the horizontal location of the first excitation of coherent energy. With there being only horizontal information in the HF semblance maps, we can not retrieve information on the depth of the nucleation point. To infer the source parameter of time we assign for each grid point the first time when significant semblance is mapped in this grid point. This time is given in relative seconds after the first semblance peak. Potentially, the time of the first semblance mapping onto a grid point is different in each bootstrap realisation of the multi-array BP. The result is an ensemble of potentially different semblance values for potentially different time steps for each grid point from each bootstrap. We choose one of the times associated to the drawn nucleation point at random as the source parameter time.

Potentially also the number of significant sub-sources, or segments, involved in the rupture can be estimated from the BP results, e.g. to define the initial number of segments in modeling or the range of possible significant segments in an multi-dimensional modeling framework. In LF semblance maps individual regions of high semblance can mark segments and these regions could be analysed individually. In HF semblance maps we can estimate the number of segments based on the number of significant high-frequency semblance peaks. For the simplest case, assuming smooth unilateral rupture along a single segment, two high-frequency energy emissions, one from a start phase and one from a stopping phase, should occur for each segment. In case of significant segmentation, HF semblance can be used to also estimate sub-source nucleation point positions and sub-source onset times. In our current application we do that, only at a later stage of the analysis. We make these sub-source nucleation point estimations dependent on other geometrical source parameters that defined the source's outline based on position, length, width and strike. These parameters are estimated not only from the BP results as outlined, but also from results of the surface displacement map segmentation introduced below (Section 2.2).

2.1.6 Synthetic tests with multi-array backprojection

We evaluate the introduced multi-array BP method carrying out several synthetic tests. With these tests we assess the spatial and temporal resolution of the method and the capabilities to recover input models of earthquake sources. The station distribution that we use in these synthetic tests is identical to the one that was available for an investigation of the intermediate-sized M_w 6.6 Muji earthquake on 25 November of 2016, 14:24:30 (USGS) in the Pamir region (Fig. 1).

For the synthetic tests we use kinematic model parameters for a moment tensor point source or a finite rectangular source (see also section 2.3 and Fig. S1), equivalent to a Mw 6.6 earthquake. To model a segmented source with two sub-sources, we divide this seismic moment equally between the two sources. The medium model is based on the AK-135 global velocity model and we use 4 Hz Green's functions based on the QSSP code by (R. Wang et al., 2017) to calculate synthetic waveforms. The Green's functions have been pre-calculated and stored (Heimann et al., 2019a). We carry out LF BP and HF BP with the frequency bands from 0.003 Hz to the corner frequency f_c of 0.25 Hz and from f_c =0.25 Hz to 1.5 Hz, respectively. The array weights have been bootstrapped 100 times and azimuthal array weights have been applied. To each synthetic waveform real pre-event noise from the corresponding real waveform record before the 2016 Muji earthquake is added.

In a first test Test 1 we estimate the ability of the multi-array BP method to recover the position of a single point source. The source is pure a double-couple with a triangular source-time function that has a duration of 3 s. The source is located at 8.7 km depth and the BP is calculated for a source point grid at that same depth.

The LF BP and HF BP results of Test 1 (Fig. 2) show that the source position can be well recovered within 2 km using LF BP and within 0.2 km using HF BP. Test 1 results with separate P- and SH-wave BPs can be found in the Supplement (Fig. S2). They give similar results as the combined P- and SH semblance results, but show systematically lower spatial precision compared to the phase-combined semblance results.

In a second test Test 2 we keep all parameters as they are in Test 1, apart from the size of the virtual arrays, which may increase from an average aperture of 3.5 degrees in Test 1 to much larger apertures of up to 30 degrees in Test 2 (Fig. S3). The results of Test 2 (Fig. S4)

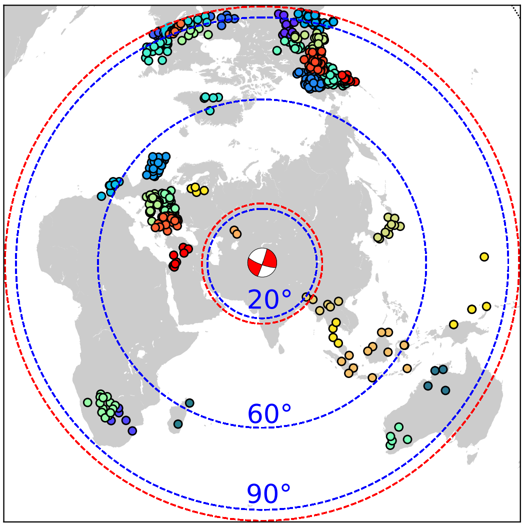

Figure 1: The stations used for the BP with the virtual clusters of the 2016 Muji earthquake plotted on a world map. The stations of the same virtual array have the same color.

show larger uncertainties in the precision for the position of the mapped semblance of about 50-60% in comparison to the smaller virtual arrays used in Test 1.

In Test 3 we further test the recovery of signals from a line source of 80 km length that ruptures unilaterally from the eastern edge with a rupture velocity of 4000 m/s. LF semblance shows a broader distribution (Fig. 2c) that well matches with the extent of the source. In HF semblance maps (Fig. 2d) two regions of high-energy release are recovered, which show well localized start and stop phases. The corresponding rupture velocity is derived from the distance between the first and last semblance maxima, which is here approximately 78 km and their time difference of 20 s. The resulting rupture velocity estimate is 3900 m/s (Fig. S6). The small difference of 100 m/s between input rupture velocity and recovered rupture velocity can be attributed to discretizations in the semblance calculation, both in space by the choice of the semblance calculation grid, and in time by the choice of time window sizes and time steps.

In a Test 4 we use two point sources with the same moment, duration and timing that are spatially separated by 50 km (Fig. S5). The individual locations of both sources are successfully recovered. The spatial precision for each source is about 20-30 km, estimated through bootstrapping and velocity model pertubations.

We also carry out all synthetic tests based on a different and more sparse station distribution that resembles the station situation at the time of the Mw 6.3 Ahar earthquake doublet

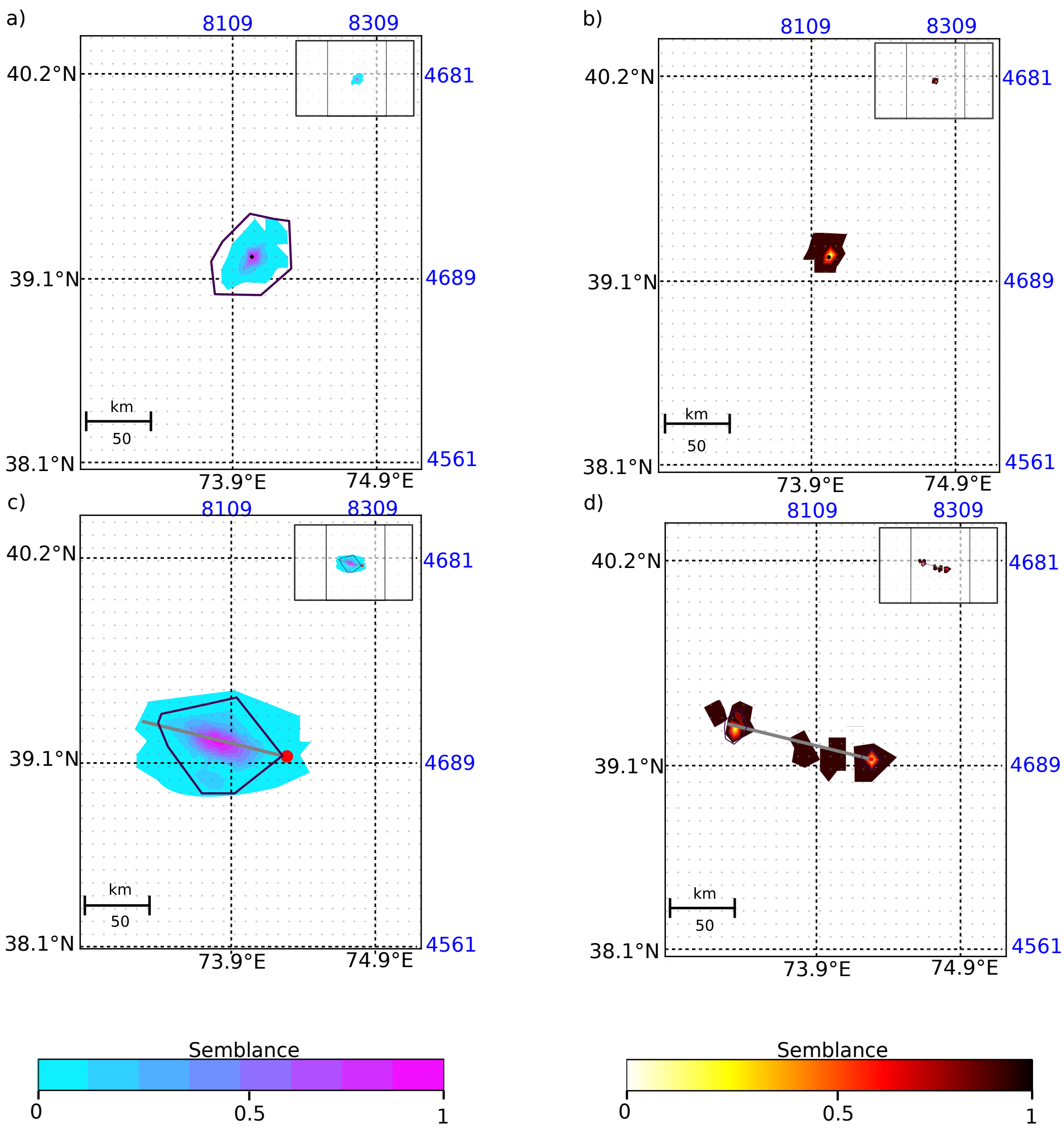

Figure 2: Cumulative LF and HF semblance for Test 1 (a,b) and Test 2 (c,d) as color-coded contours. a) LF and b) HF semblance of a double-couple source (Test 1). The black dot shows the model input position. c) LF and d) HF cumulative semblance of a line-source source (Test 3). Model outline and input nucleation point are indicated by a grey line and a red circle, respectively. In all subplots, the black outlines around high semblance values represent the 98 %spatial precision of the semblance maxima estimated from bootstrapping. Top-right insets in each subplot show the extent of the whole search grid. Gray background dots mask the BP source point grid. Coordinates are given in Latitude/Longitude (black labels) and UTM (blue labels).

on August 8, 2012. We use the real noise from before the Mw 6.3 Ahar earthquake to perturb the corresponding synthetic waveforms. The semblance resolutions in the results of these additional tests (see figures in Supplement 2.1), compared to the tests based on the Muji 2016 earthquake setup, suffer from the combined effect of mainly two factors. First, the Ahar earthquake has a smaller signal-to-noise ratio because of the smaller earthquake magnitude of Mw 6.3 compared to Mw 6.6 in the earlier tests. Secondly, the sparser station coverage at that time leads to a lower number of virtual arrays and therefore creates azimuth and distance gaps.

2.2 Pseudo-probability of source location from image segmentation on InSAR displacement maps

The spatial pattern of coseismic surface displacement is to some extent characteristic for the properties of the source. It can therefore provide valuable source information before any inverse modeling. These apparent characteristics of the surface displacement pattern are that 1) the highest displacement gradients usually occur very close to the rupture, 2) loss of interferometric coherence, producing InSAR data gaps, can be caused by very high displacement gradients, surface rupture or near-fault landslides, 3) elongation of significant displacement is parallel to the strike direction of the causative fault and 4) sign changes of the displacement separate footwall and hanging wall of the faulting. Furthermore, complexity in these displacement characteristics hint at the occurrence of significant changes in source properties, e.g. distinctly separated regions of relatively high displacement gradients point to rupture segmentation. We formalize the extraction of displacement pattern characteristics by using image segmentation methods on the surface displacement maps. Based on the results we form a pseudo-probability map of the rupture location from which we then derive other first-order rupture properties in an automated framework as described in Section 2.2.1 below.

The here proposed image segmentation of InSAR surface displacement maps includes phase coherence evaluation, displacement gradient calculation, sign change tracing in the displacement amplitudes and combination of the resulting gradient maps. We describe and illustrate the steps in detail in the following and test them, based on synthetic displacement maps and a real-data example. The tests include the analysis of synthetic data of two vertical, two-segments strike-slip faults, one 6 km deep (top edge) EW-striking and one 1 km deep NS-striking faults, and real data of the 2009 L'Aquila earthquake in Italy, a shallow unsegmented normal-faulting earthquake. The first example resembles our application to the 2016 Muji earthquake (Pamir) presented below, and the second and third tests are set up to show a variety of mechanisms with different imprints on InSAR displacement maps. Only the third, real-data example contains data gaps due to interferometric phase incoherence.

InSAR displacement maps show the three-dimensional surface displacement projected into the line-of-sight of the satellite. Different satellite look directions lead to different displacement projections, which cause apparent shifts of the surface displacement signals up to several kilometres between. We mitigate the projection effects by combining at least two different look directions, one from ascending and one from descending satellite tracks. For the synthetic tests the forward modeling was done using a layered 1-D velocity model (Xu et al., 2006; W. Li et al., 2018) and the PSGRN/PSCMP code (R. Wang et al., 2006) to produce a Green's functions database (Heimann et al., 2019a). We add synthetic correlated noise generated from real-data noise power spectra (Sudhaus & Jónsson, 2009).

The first step in our displacement gradient segmentation is treating incoherent areas that are either marked by no-data values or get marked, e.g. based on a coherence map and using a coherence threshold. Pre-processing of the displacement data should include deramping to minimise potential bias on results.

We assign a zero displacement value to incoherent pixels to enable numeric calculations and to contrast incoherent areas to areas showing large displacements. We then calculate the absolute displacement gradients for each pixel pair. For the displacement gradient map we apply a moving average across the pixels with Gaussian weighting using a window that spans about 500 m by 500 m in the examples (Fig. 3b and e). Next we trace sign changes of the displacement amplitudes. This is done based on a binary image that distinguishes positive and negative displacements, on which we apply the same gradient calculation with the same settings as described above. Non-zero gradients are normalized to 1 and effectively provide the sign change traces (Fig. 3c and f).

We determine the area of interest (AOI) as a minimum bounding box which comprises the 95% highest displacement values. For a pseudo-probability map of rupture location, we first combine the displacement gradient maps and the sign change traces for each data set individually by multiplying the two maps pixel-wise, with a relative weight in place based on the data set signal-to-noise ratios. Signal-to-noise ratio is evaluated between the signal in the area of interest and the noise from surrounding areas. The following processing steps include summing up the combined gradient maps of all available data sets, masking values of less than 1% of the maximum combined gradient, and applying a normalization (Fig. 3g). In other words, we keep the non-negligible gradient information only in places where there is displacement sign change in one of the data sets. In these remaining areas, the pseudo-probability of rupture location scales with sum of the displacement gradients from all data sets.

As a final processing step we aim to clean the simple pseudo-probability map of spurious pixels with non-zeros probabilities. A simple threshold for distinguishing signal and noise in the pseudo-probabilities seems inappropriate given the variety of displacement patterns. To best outline areas with significant pseudo-probability, we therefore apply the Otsu's method

(Otsu, 1979) thresholding to classify pixel-wise the pseudo-probabilities into signal and noise. The Otsu's method is an iterative and exhaustive approach that seeks to minimize the differences of the pixel values in two distinct classes (Otsu, 1979; Shaus & Turkel, 2016).

In our first two tests with shallow and deep (top edge depth 0.5 km and 6 km, respectively), EW and NS striking strike-slip faults, the area of significant surface displacement and near-field of the rupture are highlighted well with relatively high gradients (Figs. 3 and S15, b and e). Together with sign change traces, the location of the causative fault is marked in the pseudo-probability map of rupture location. In the first test with a deeper source and a consequently lower signal-to-noise ratio the pseudo-probability is more scattered than in the second test. Still, the highest pseudo-probabilities occur very close to the input fault in both cases (Figs. 3g and S15g).

For vertically dipping strike-slip faults, the displacement exactly above the fault is zero and coincides with high displacement gradients. For inclined and blind faults the so-called hinge line of largest displacement gradients will be offset in the direction of a projected surfacing of the fault, as is the displacement sign change.

A small signal-to-noise ratio of around 1 or less is in our experience challenging for the described simple gradient based sign-change tracing. In such cases, sign changes are abundant which results in very wiggly sign-change traces. Therefore, in these cases we substitute the gradient-based sign change tracing with less scattered contours of Chan-Vese image classes. The well established iterative Chan-Vese segmentation method (Chan & Vese, 2001; Getreuer, 2012) divides an image into two classes of minimum intra-class variance. These classes represent the topological changes of an image (Chan & Vese, 2001). The Chan-Vese image segmentation was applied in the real-data test to the 2009 L'Aquila earthquake displacements (Fig. S18).

In the displacement gradient maps the large window size used for averaging the gradients across pixels (500 m by 500 m) has the desired effect of a smeared-out gradient estimation. This estimation may therefore bridge over high gradient values surrounding incoherent areas. It also somewhat reflects the slightly ambiguous relationship between fault location and high displacement gradient location.

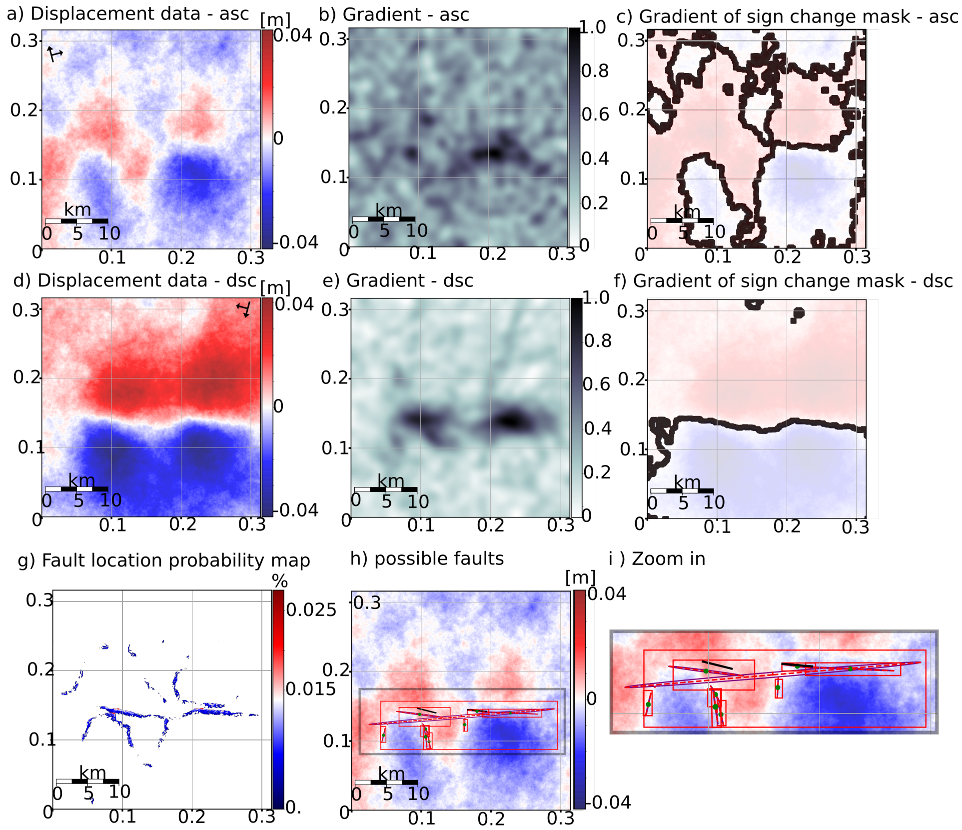

Figure 3: Image segmentation applied to synthetic displacement maps two-segments strike-slip source at 6 km depth (top edge). a) and d) show ascending and descending displacements, b) and e) the corresponding displacement gradients, c) and f) the corresponding sign-change trace (black) over the displacement. g) Pseudo-probability map of fault location. h)
Bounding boxes and ellipses applied on g). The green box surrounds the area of interest, also enlarged in i). The red dashed lines indicate the major axes of the ellipses containing the highest pseudo-probability values in each region found as described above. The outline of the synthetic source(s) is indicated in the figures with black lines. The ellipse (purple outline) is centered at the centroid of each region.

2.2.1 Defining the model space based on displacement map segmentation results

The displacement pattern analysis using image segmentation methods (Section 2.2) provides a pseudo-probability map of rupture locations. Similarly to the methods we apply to the semblance maps, we use this pseudo-probability map to derive arbitrarily-oriented minimum bounding boxes and ellipses that enclose highly probable rupture locations, provide information on the probable number of rupture segments, and based on the ellipses individual segment dimensions and orientations.

On the pseudo-probability map of the rupture location we mark regions using the python scikit-image pack (Van der Walt et al., 2014) by evaluating the neighbourhood of each pixel to find connected pixels of any value, i.e. pixels share an edge or a corner. Connected pixels form regions, which potentially correspond to individual fault segments. A minimum size for a region can be given. A single connected region at this stage points to an unsegmented rupture. If, however, two or more of separated regions with extents larger than about 300 m in any direction occur, we apply minimum arbitrarily-oriented bounding boxes and ellipses to these regions to define the properties of those for potential sub-sources used in multi-segment or multi-dimensional fault modeling. We also apply a minimum arbitrarily-oriented bounding box and ellipse encompassing all regions.

Based on a single region or several, an arbitrarily-oriented minimum bounding box and an ellipse are defined each using the image processing algorithms provided by the scikit-image project (Van der Walt et al., 2014). The minimum bounding box length and width provide estimates of the fault length or segment length. The center of the best-fitting ellipse is defined by the focal point of the pixels within a region, with pixels of high pseudo-probability value having a large weight in this calculation. The major axis of the ellipse is likely a good indicator for the strike direction and is used as a prior. By slightly changing the threshold settings and minimum separation distance between patches to connect them to regions we retrieve a set of length and strike values for each region.

We can now construct Gaussian distributed continuous functions for the source parameters strike and length for each source from these estimates. The prior pseudo-probability functions are guiding the first stage of sampling the model space during the actual optimizations as described below.

2.3 Earthquake source optimization implementing data-driven model parameter prior distributions.

To characterize earthquakes we carry out a joint kinematic source modeling applying a non-linear, randomized direct-search optimization. We represent the co-seismic faulting with rectangular dislocations (see model sketch in Fig. S1b) embedded in a horizontally layered elastic medium. We optimize for each such dislocation the following model parameters: the fault location (north, east, depth), the fault dimension (length, width), the fault orientation (strike, dip), the slip, the rupture velocity, the relative position of rupture nucleation on the fault plane and rupture onset time. We define the fault location at the center of the top edge of the rectangular plane. An earthquake can be represented by more than one of these dislocations, e.g. if segmentation plays a significant role. For such segmented sources, the optimization setup enforces non-overlapping and non-intersecting dislocations.

The boundaries of the model parameter space in which the direct search optimization is applied and the model parameter prior distributions, commonly called *priors*, have to be predefined. Here we set these based on the pseudo-probabilities of fault locations and the time evolution of the rupture estimated in displacement map segmentation and multi-array back-projection (Sections 2.2.1 & 2.1.5). We point out that the choice of the specific optimization method, with its objective function and model space sampling strategy, is independent from the presented approach. We use and extend the open-source optimization code Grond (Heimann et al., 2018), which has the capabilities to estimate model uncertainties through the use of Bayesian bootstrapping.

Our optimization procedure works in adaptable sampling phases. The first sampling phase usually involves uniformly random sampling. It is followed by "directed" sampling phases that become more and more directed to good-fit models in the course of sampling. Each drawn model is evaluated against a set of different data weights, based on Bayesian random station weighting for the seismological data, and different noise perturbations for the static InSAR data. Source models are collected in a fixed-size highscore list for each of these sets of weights, forming a so called *bootstrap chain*. A detailed description of the method can be found in Section 1 and the online documentation (Heimann et al., 2018). In the optimization we are seeking the minimum of the L2-norm between observed data $\mathbf{d}_{\mathrm{obs}}$ and predicted data $\mathbf{d}_{\mathrm{pred}}$. The general form of this objective or *misfit* function is:

$$||e|| = \sqrt{\sum (\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_{\mathrm{pred}})^2}. \quad (11)$$

The prior information of source characteristics is taken into account from the start of the optimization by setting corresponding model parameter bounds and for some model parameters non-uniform non-normalized prior pseudo-probabilities. In this way we replace in our optimization a usually much more exploratory first phase of model space sampling, i.e. within wide bounds for the model parameters and with uniform random sampling therein, with a more focused and guided sampling. After a defined amount of sampling and based on a selection of low misfit models, the model space is reshaped and defined by the multi-dimensional distribution of low-misfit model parameters, the highscore list, which is constantly updated. In this way the start set of parameter bounds and prior distributions is dropped such that the optimization is entirely driven by the objective function. Wide fall-back parameter bounds can be used to facilitate exploration.

2.3.1 Guided optimization

We call our first optimization phase that uses source parameter priors *guided* optimization phase to reflect the narrowed model space settings. We describe how we form model parameter prior distributions from either method, the multi-array BP (2.1.5) and the displacement map segmentation (2.2.1), individually. If both methods are used complementary, we first combine their pseudo-probability maps of rupture locations. We re-sample the grid of the BP location pseudo-probabilities to the grid spacing of the fault pseudo-probability map using a nearest-neighbor interpolation. We then combine these two prior distributions by multiplication. This procedure inherently gives weight to the better resolved prior parameter distribution of each method. From the defined joint discrete pseudo-probability functions the source models locations are sampled for a given number of models.

The horizontal location parameters in the optimization are relative east and north shifts in a metric coordinate system with the chosen reference location at the origin. If more than one source segment is considered, parameters of a single source segment are drawn first, which define the first segment's outline. A second source is then drawn from the prior distributions and accepted if its outline is not intersecting the outline of the first source. A redrawing of models is necessary until this condition is fulfilled. This sampling scheme can be extended to an arbitrary number of sources.

For strike and length we also combine the Gaussian-distributed continuous pseudo-probabilities obtained from the multi-array BP and the displacement map segmentation methods by multiplication. In most cases this gives more weight to the displacement map segmentation methods result. The source parameters time and velocity are only drawn from prior distributions inferred from the multi-array BP.

2.3.2 Settings for the modeling of the far-field waveform data

For the forward modeling of seismic waveforms we make use of pre-calculated Green's function stores (Heimann et al., 2019b) to speed up calculation. The Green's functions stores (Heimann et al., 2017) are calculated for up to 0.5 Hz using the QSSP code by (R. Wang, 1999) and are based on the AK-135 1-D velocity model (Kennett & Engdahl, 1991). The Green's functions are calculated with spatial sampling of 4 by 4 km and we enable continuous source-receiver distances by multilinear interpolation in space between the grid points.

Before the optimization we determine waveform balancing weights after Heimann (2011) in an empirical way from 1.000 $(k, ..., K)$ uniformly random models. This balancing corrects amplitude differences due to geometrical spreading, amplitude differences between P and S phases and different length of the cut-out windows. At each station $(i, ..., N)$ and for each component (phase) $(j, ..., M)$ we determine the balancing weights $\mathbf{r}_{\text{balance},ij}$ as:

$$\mathbf{r}_{\text{balance},ij} = \frac{1}{\frac{1}{K}\sum_k^K |\mathbf{d}_{\text{pred}ijk}|}. \tag{12}$$

The objective function for waveforms that needs to be minimized is defined with Eq. 11 and Eq. 12 as:

$$||e|| = \frac{\sqrt{(\sum |\mathbf{r}_{\text{balance}} \cdot (\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}}|))^{\mathbf{2}}}}{\sqrt{\sum |\mathbf{r}_{\text{balance}} \cdot \mathbf{d}_{\text{obs}}|^2}}. \tag{13}$$

We calculate the misfit according to Eq. 11 for each waveform individually and thereby allow for an individual, fit-maximizing time shift from -4 s to +4 s. With those time shifts we account for traveltime deviations due to 3D velocity variation not represented in the AK-135 1-D velocity model.

2.3.3 Settings for the modeling of static near-field data

We use PSGRN/PSCMP (R. Wang et al., 2006) to calculate static Green's function stores for the forward modeling (Heimann et al., 2019b).

We combine ascending and descending scenes into one data vector $\mathbf{d}_{\mathrm{obs}}$. The data error is considered in Eq. 11 by a weighting matrix $\mathbf{R}$ that derives from the data error variance-covariance matrix Σ:

$$||e|| = \sqrt{[\mathbf{R}(\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_{\mathrm{pred}})]^{\mathbf{T}} \mathbf{R}(\mathbf{d}_{\mathrm{obs}} - \mathbf{d}_{\mathrm{pred}})}, \tag{14}$$

with $\mathbf{R} = \sqrt{\mathbf{\Sigma}^{-1}}$.

Following the "Randomize-then-Optimize" (Bardsley et al., 2014) procedure we add synthetic noise, which modifies the objective function to:

$$\begin{aligned} ||e|| = \Big\{ & [\mathbf{R}(\mathbf{d}_{\mathrm{obs}} + \epsilon_{\mathrm{syn},\mathbf{i}} - \mathbf{d}_{\mathrm{pred}})]^{\mathbf{T}} \\ & \mathbf{R}(\mathbf{d}_{\mathrm{obs}} + \epsilon_{\mathrm{syn},i} - \mathbf{d}_{\mathrm{pred}}) \Big\}^{1/2}, \end{aligned} \tag{15}$$

with $\mathbf{R} = \sqrt{\mathbf{\Sigma}^{-1}}$.

The seismic moment is calculated using $M_0 = \mu AD$, with shear modulus μ, fault area *A* and the fault slip *D*. We use μ based on the layered 1-d velocity for the region based on Xu et al. (2006) and W. Li et al. (2018). Additionally to earthquake source model parameters, three data ambiguity model parameters are used for each InSAR data set to remove any residual average data offset and a linear phase ramp in east and north direction.

3 Application to the 2016 Muji earthquake

3.1 The 2016 Muji earthquake

The Muji earthquake struck in north-eastern Pamir in the Chinese county Aketao on the 25 November of 2016 at 14:24:30 (UTC) and is sometimes also called Aketao earthquake in the literature after the region. It had a moment magnitude of M_w 6.6. The rupture occurred along the Kongur Extensional System (Chevalier et al., 2011; T. Li et al., 2019; Chevalier et al., 2015), located between the Tarim basin and the Muji–Tashkorgan basin. The 2016 Muji earthquake is the first instrumentally recorded earthquake of $M_w > 6$ to have ruptured the transpressional Muji fault (Fig. 4). This fault bounds the south side of the Muji range and the northern margin of the Muji graben. In the east the Muji fault starts at the eastern side of the impact crater lake Karakul (Gurov & Yamnichenko, 1995) and extends south-eastwards until connecting with the perpendicularly running Kongur Shan fault (Chevalier et al., 2011). Farther south the Kongur Shan fault ultimately connects with the major Karakoram fault.

The Muji fault accommodates EW extension due to the northward indentation of the Pamir salient (Chevalier et al., 2011). Fluvial terraces cover parts of the surface expression of the fault (Chevalier et al., 2011). Geological markers in the western part of Muji fault indicate right-lateral fault movement, while the eastern part of the fault displays mostly evidence of normal faulting that is associated with a small component of right-lateral movement (Chevalier et al., 2016, 2011). In field investigations Chen et al. (2016) found some surface breaks that appear to have formed co-seismically during the 2016 Muji earthquake.

The 2016 Muji earthquake has been studied by several authors who used InSAR, GNSS and/or seismic waveform data in earthquake source inversions (J. Li et al., 2019; Feng et al., 2017; Bie et al., 2018; He et al., 2018; Ma et al., 2018), compiled in Table S2. They unanimously suggest a complex faulting mechanism that involves more than one fault segment. Feng et al. (2017) found the coseismic displacement signal to be consistent with two spatially separated segments.

He et al. (2018) first assumed for the 2016 Muji earthquake a listric geometry based on the aftershock distribution, but using geodetic data found a better fit to the data using a planar fault geometry. Bie et al. (2018) modelled the rupture using regional waveform data and estimated a sub-shear rupture velocity of 3.7 km/s as the most plausible scenario. Bie et al. (2018) also found a significant overlap of the modelled source-time functions (STFs) from the two sub-events, indicating a near simultaneous rupture of the two segments. In their study, the eastern sub-event displays a temporally more compact STF. However, they relate that the modeling of the STFs of the two sub-events proved difficult. Furthermore, Bie et al. (2018) state that they could not distinguish the rise and fall times for each sub-event. Bie et al. (2018) concluded that the 2016 Muji earthquake, being an intermediate-sized earthquake, has the smallest reported temporal gap between two sub-events upon publication date.

3.2 Waveform data processing and multiarray BP setup

In our analyses of the 2016 Muji earthquake we use seismic waveforms from broadband stations with sampling rates of at least 10 Hz and with locations at teleseismic epicentral distances between 23° and 93°. The data are accessed via the FDSN services IRIS and Geofon, and additionally RESIF and ORFEUS for the multi-array BP. For the seismological waveform processing we use the Pyrocko software (Heimann et al., 2017; Heimann, 2011; Cesca et al., 2010). We use a layered 1-d regional velocity model (Fig. S19) based on Xu et al. (2006) and W. Li et al. (2018).

For the teleseismic BP we can use the downloaded data without further manual data checks and/or selection. Through the stacking process for the BP noisy data and faulty response functions of singular stations have a comparatively small impact. The method strongly benefits from more stations and hence more virtual arrays. We resample the waveforms to a common 10 Hz, rotate the seismogram components into the source-centred RTZ coordinate system, and restitute the data to ground velocity by removing the instrument answer. While we use the frequency range from 0.003 Hz up to 1.5 Hz for the BP, we separate within that band a low-frequency and a high-frequency band at the estimated corner frequency, here 0.16 Hz, through bandpass-filtering as described in Section 2.1.3. We obtain two LF and HF waveform sets using the Z-components for P-wave BP and the T-components for the SH-wave BP. We show exemplary normalised waveform data and spectra of P-waves and SH-waves from an array with stations located between epicentral distances of 5633 km and 6243 km in Figures S21 and S22.

In our multi-array BP of the 2016 Muji earthquake we form 34 virtual arrays from 563 stations in total (Fig. 1). The virtual arrays have a maximum aperture of 5° and at least 5 stations. To form an array in the Pacific we allowed for larger array apertures of up to 10° and decrease the number of required stations to 4, to increase coverage. The resulting average number of stations per array is 9. Most stations are located in North America and Europe. Only

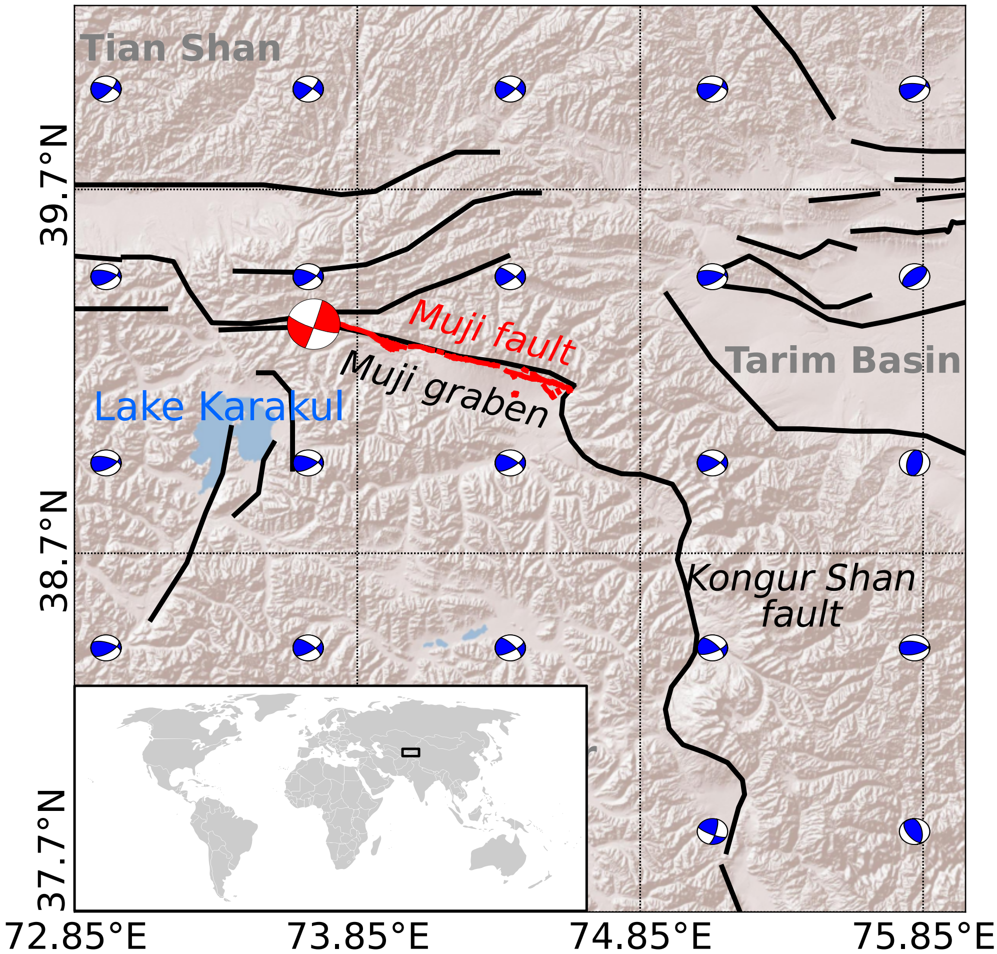

Figure 4: Setting of the Muji 2016 earthquake. Map of the region around the area of the 2016 Muji earthquake. Black lines indicate regional faults as mapped in the GEM fault database (Styron, 2019). Red lines are reported co-seismic surface ruptures (T. Li et al., 2019; Chen et al., 2016). The red beachball indicates the USGS hypocenter and body-phase determined focal mechanism. Other beachballs are representing the potential focal mechanisms from the World Strain Map (Kreemer et al., 2014). Inset shows the location of the area on the world map.

waveforms with a cross-correlation coefficient above 0.6 to the center-most station of each array are taken into account for further processing. The cross-correlation coefficient is calculated after shifting the waveforms with regard to the theoretical onset time given by the USGS hypocenter location and the velocity model. The horizontal grid of locations for which the BP is performed is at 9 km depth and extends 1.5 degrees around the USGS hypocenter. The grid spacing is about 0.018 degree or 2 km.

We apply phase-weighted stacking of the waveform sets in virtual arrays to calculate the multi-array semblance, as described in Section 2.1.1. For comparison, we show an example of single-array semblance formed with phase-weighted stacking together with the semblance formed with linear stacking in Fig. S20. To investigate the time evolution of the rupture we carry out BPs with moving times windows. In LF BP these time windows have a duration of 24 s and are moved by 8 s in each step. In the HF BP the time windows and step sizes are shorter with only 10 s and 2 s, respectively.

For the finite-rupture optimization we resample the waveforms to 0.5 Hz, apply a bandpass-filter from 0.01 Hz to 0.13 Hz and restitute the waveforms to ground displacement. For the P-wave we only use the Z-component of the waveforms and evaluate the full-waveform misfit in a time window from 15 s before to 25 s after the theoretical onset of the P-wave. For the SH-wave we use the T-component of the waveforms and evaluate the misfit in a time window from 25 s before to 35 s after the theoretical onset.

3.2.1 Empirical traveltime correction on the Muji earthquake waveform data

We apply empirical traveltime corrections (see also Section 2.1.4) to the processed waveforms of the Muji earthquake. For the estimation of the corresponding traveltime shifts we use as the reference event the Mw 5.2 November 25 earthquake in 2016, which occurred at 14:18:59, so 5:30 minutes before, and about 10 km south-east of the main shock. Its mechanism is similar to the main earthquake (USGS catalogue) but the source is about 10 km deeper. To estimate traveltime shifts we use the exact same set up of stations, filters and array forming as for the main earthquake and maximise the semblance of the reference event for each waveform set independently (Fig. S23). For this operation we use a single time window of 32 s for the LF BP and 24 s for the HF BP, which begins 4 s and 6 s before the theoretical onset of the P- and SH-phases, respectively.

We find strong azimuthal correlations of traveltime shifts between the stations (Fig. S23). In general, stations north of the event display negative time shifts and stations south of it positive shifts, for both P- and SH-waves. Also the differences between the empirically estimated time shifts found for the LF and HF BPs are generally small and in good agreement for P- and SH-phases. Only a few individual stations display significantly different time shifts to other stations of the same array and/or show a sign change in the time shifts between LF and HF BPs. For the P-waves time shifts are in the range of +/- 1.5 s and increase for SH-waves to +/- 3.5 s.

3.3 Near-field data

For the estimation of the Muji fault location based on the gradient of the surface displacement data we employ an ascending and a descending SAR interferogram, based on Sentinel-1 interferometric wide-swath satellite data in VV polarization. The SAR data were downloaded from the Copernicus Open Access Hub (`https://scihub.copernicus.eu/`). Primary and secondary image dates are 2016/10/20 and 2016/12/07 for the ascending data, and 2016/11/25 and 2016/12/19 for the descending data. The differential interferograms are processed using the ESA SNAP Sentinel-1 toolbox (s1tbx) and the SRTM elevation model (Farr & Kobrick, 2000). The interferograms have been filtered using an adaptive Goldstein filter with a window size of 16 and a filter factor of 0.8. Unwrapping was conducted using the tree-branch-cut algorithm (Goldstein et al., 1988), with a coherence threshold of 0.1. We account for the presence of correlated data errors in the displacement maps in the optimization. We empirically estimate the variance-covariance functions of the data error, assuming that they resemble Gaussian random field and stationarity (Hanssen, 2001). This estimation takes place in areas of the displacement map that show no apparent surface movement. Before the kinematic source modeling using these data, their number is reduced through irregular data subsampling with the quadtree algorithm (Jónsson et al., 2002). Data error estimation, data subsampling and the implementation of the variance-covariance functions to build variance-covariance matrices for the subsampled data (Sudhaus & Jónsson, 2009) are done using the Kite software package (Isken et al., 2017). For forward modeling the near-field displacement, we calculate Green's function based on the layered 1-d regional velocity model Xu et al. (2006) and W. Li et al. (2018).

3.4 Results

3.4.1 Spatio-temporal evolution of the 2016 Muji earthquake

The LF (0.003-0.16 Hz) BP results of the 2016 Muji earthquake display two, in east-west direction spatially separated, high semblance regions (Fig. 5a). This pattern also appears in the individual BP results of P- and SH-phases (Figs. S27 and S28), with the P-phase BP providing somewhat better resolution. This semblance pattern points to a segmented rupture. The temporal evolution retrieved from LF BP in moving time windows suggests that the earliest coherent energy release took place in the western region before seismic energy excitation occurred in the eastern region (Fig. S24). The western region seems to remain activated throughout the duration of the rupture (Fig. S24). The LF BP results are used quantitatively to inform about the model space for the parameters onset time and rupture velocity (Fig. 8)in the optimization, and will be used as well to set model parameter priors for strike, length and position in combination with the results from the surface displacement image segmentation method results.

The HF BP results show spatially more localized areas of high semblance (Fig. 5b) compared to the LF BP results, while the location and orientation of LF and HF semblances agree very well. Also the time evolution revealed in HF BP is similar to the LF BP results, with some more detail (Fig. 6 and 1 s steps in Fig. S26). The first HF semblance peak occurs in the western corner of the Muji basin, close but slightly west of the Muji fault centre (Fig. 6b). All BP-derived semblance times are relative to this first occurrence of coherent energy mapping. This first semblance peak is associated with the strongest beampower of the sequence. The rupture then propagates simultaneously west- and eastward along the Muji fault (Fig. 6b-f). In the time from 6 s to 10 s seismic energy is continuously radiated in the onset area, in the area within 15 km east of it and also slightly west of it. At the latest stage of the rupture, between 12-14 s, the second strongest semblance peak within a somewhat widespread semblance high is found, located almost 30 km west of the onset peak. The location precision estimates based on bootstrapping (Section 2.1.1) range from 5 km to 15 km (Fig. 6). From the HF BP spatio-temporal semblance results of the 2016 Muji earthquake we estimate a rupture velocity, considering fault segmentation, and the length of segments. For each segment we take the distance and time between the first and last occurrence of HF semblance in the time-incremental semblance maps as nucleation and stop phase, respectively. For the western segment these estimates deliver rupture velocities between 1.8 km/s and 2.1 km/s (Fig. S25) and for the eastern segment between

2.1 km/s and 2.6 km/s. Using straight-line distances between the semblance peaks and ignoring the first peak as potential nucleation point, because of the indications that the rupture is likely bilateral our method estimates for the western segment a length of 25 km to 30 km and for the eastern segment a length of 10 km to 15 km. The source parameter estimates for length, rupture velocity and the locations of the nucleation points for each fault are used as described as prior information (Fig. 8) for the guided optimization.

Based on InSAR data of the 2016 Muji earthquake we create a pseudo-probability fault location map applying image segmentation methods (Section 2.2). As detailed in the given section, we use the interferometric phase coherence, the displacement gradients and sign changes of the displacement to get information on the deformation source (Fig. 7). Based on the pseudo-probability fault location map (Fig. 7,g), we surround all areas that mark a high pseudo-probability of fault activation with a single minimum bounding box (Fig. 7,h) to estimate the dimension of the entire fault. Furthermore, we identify two distinct areas with of high pseudo-probability of fault activation and enclose these with bounding boxes and ellipses, respectively (Fig. 7,h). We interpret these separated areas as markers for two distinguishable fault segments, which we represent with two kinematic sources in the optimization. For each segment we estimate independent parameter priors. The source parameter priors for length, strike and position obtained agree well with literature values (Fig. 8, Tab. S2).

3.5 2016 Muji earthquake two-segment rupture optimization results

3.5.1 Exploratory and guided optimizations of the 2016 Muji rupture using the same data

We carry out two independent non-linear kinematic source optimizations for a two-segment fault model, without and with including prior information from data analyses as described in the method section (2.3). From the Bayesian bootstrapping of the data we realize 100 sets of different combinations of objective functions and realize 100 bootstrap chains, each based on a different combination of target weights and different realizations of noise-perturbed data. We use the same random seeds in both optimization to create the same random weights and noise-perturbations. The highscore list of models, on which the statistics for new model samples are generated during the direct search, keeps a fixed number of $4 \cdot n_{\mathrm{par}} + 1$ low-misfit models.

For the exploratory optimization we choose parameter bounds as could be chosen by an informed and cautious, conservative investigator, who has had access to the BP results and the displacement maps. For the source locations this results in 20 km wide ranges for north and

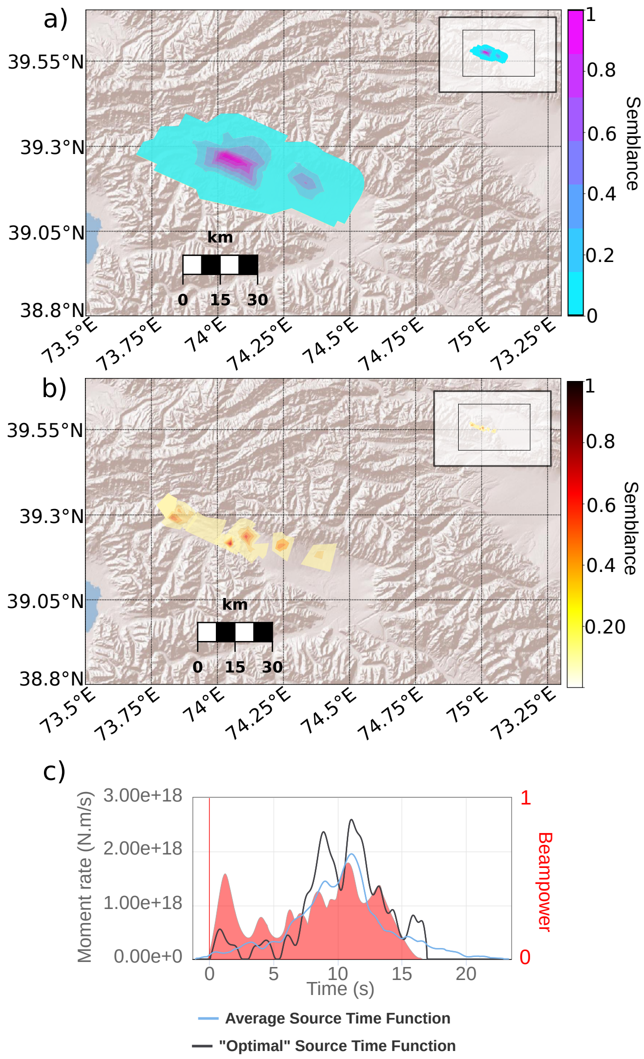

Figure 5: Cumulative spatial semblance map for the (a) low- and (b) high-frequency BPs. Contour lines are colored after the cumulative semblance. The figures are a zoom in upon the area of interest from the main grid. The inset window in the top right shows the extent of the grid. C) Beampower of the high-frequency BP as a function over time as a red and filled function of time together with the optimal (black line) and average (blue line) source time functions from the SCARDEC catalog (Vallée & Douet, 2016).

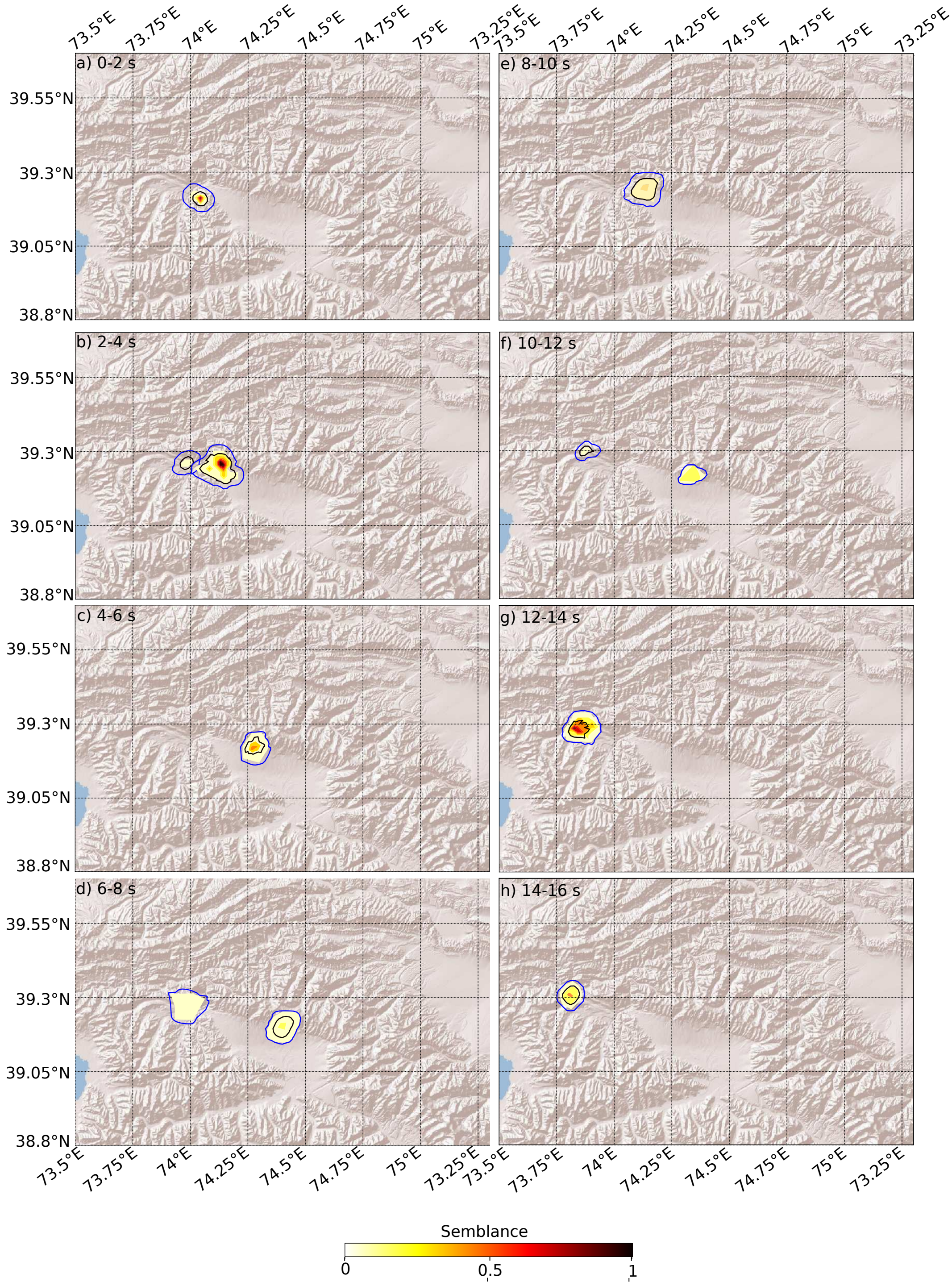

Figure 6: HF BP incremental spatial semblance map for 10 s time windows moving with 2 s time steps. Solid outlines mark 95% (black) and 68% (blue) of all maximum semblance locations from bootstrapping.

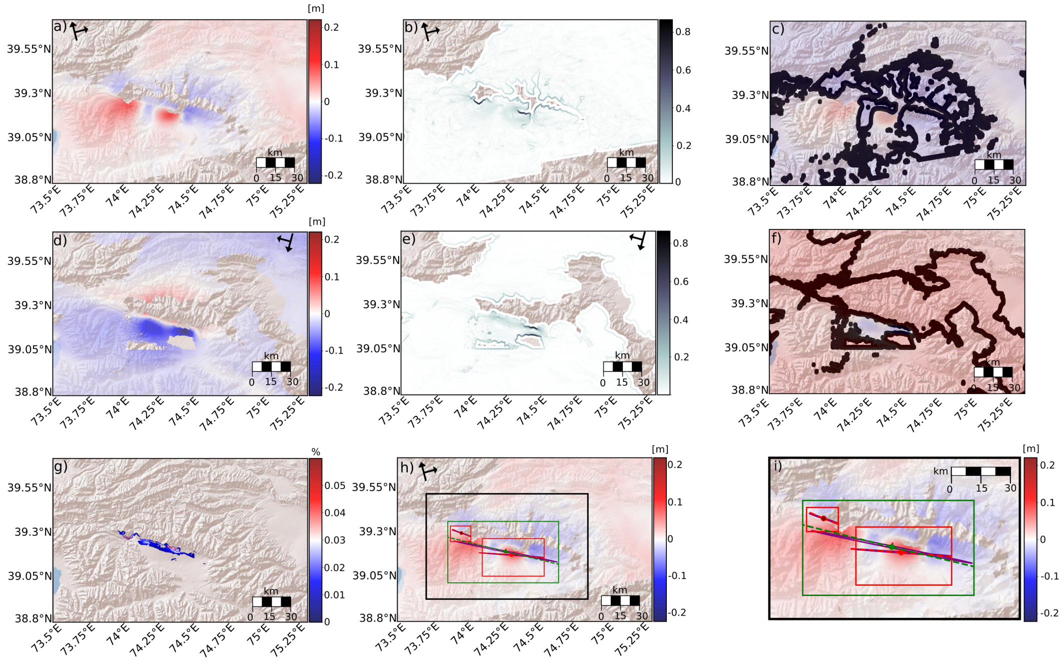

Figure 7: Displacement map segmentation results for the 2016 Muji earthquake. a) shows the ascending line-of-sight displacement data, (b) the corresponding gradient map and c) the gradient of the sign change mask, superimposed on the displacement data. d) shows the descending line-of-sight displacement data, e) the corresponding gradient and f) the gradient of the sign change, superimposed on the displacement data. g) Combined pseudo-probability map of fault location adding ascending and descending pseudo-probabilities. h) Minimum bounding boxes and ellipses on the pseudo-probability maps in i), enclosing the automatically determined area of interest (black box), enclosing all high pseudo-probability values (green boxes, long purple ellipse), and the separated areas of high pseudo-probability (red boxes, small purple ellipses). Major axis of the ellipses and centroids are shown as dashed lines and dots the corresponding colors for single and two segments estimations. The background shows the ascending displacement map for visual reference. i) Zoom in on the area of interest (black box in h).

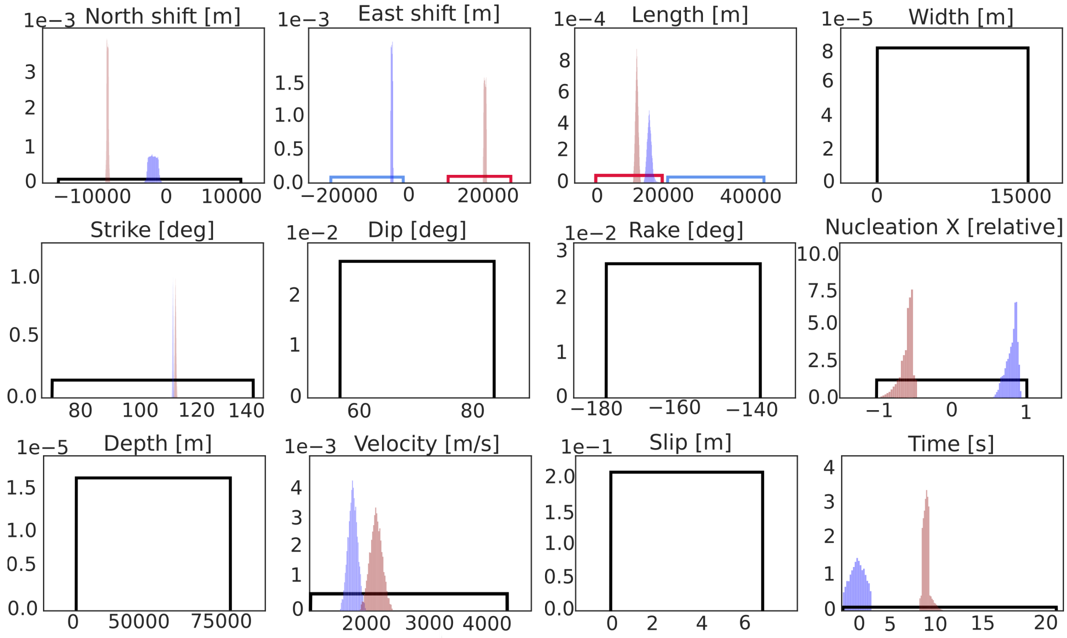

Figure 8: Model parameter space for the guided optimization of the Muji 2016 earthquake as specified through BP and displacement map segmentation or otherwise assumed. Black box outlines mark assumed uniform prior probabilities without any data analyses, while colored boxes and histograms show uniform or non-uniform inferred prior pseudo-probability functions for the parameters. Colored pseudo-probability functions mark priors for the two distinguished source segments in the western (light blue) and eastern (light red) part of the fault.

east source locations around the approximate center of the signals. For the onset time of each source the parameter range is set from 0 s to 20 s for both sources, with 20 s roughly being the rupture duration as given by the SCARDEC catalog (Vallée & Douet, 2016) for the 2016 Muji earthquake. To pre-constrain the source mechanism parameters (strike, dip, rake), we set 80 degrees wide parameter bounds, centered around the expected focal mechanisms from the World Strain Map (Kreemer et al., 2014). For each source the model parameter slip can range from 0 m to 4.5 m, the parameter width from 0 m to 15000 m and the depth (top edge) from 0 m to 7500 m. The prior probability is uniform for all these parameters.

For the optimization starting with the guided phase we base the model parameter prior distributions on the available estimates of the BP results and/or the displacement map segmentation, or the same priors as in the exploratory optimization. The aim is to well constrain the 2016 Muji faulting and to compare the results of the two different optimization runs in terms of source results and performance.

In general, both optimizations converge to very close locations within the high-dimensional model space such that parameter marginals mostly cover the same parameter ranges (Fig. 9). The spatio-temporal evolution of the best model of the guided optimization more closely resembles the inferred spatio-temporal evolution of the BP result, in contrast to the best model of the exploratory optimization (Fig. 10). The inferred result from the guided optimization and the inferred BP result is however part of the ensemble of the exploratory optimization. The exploratory optimization needs more sampling to converge compared to the optimization that starts with the guided sampling (Fig. S35). The corresponding posterior probabilities have not always the same shape and also, the best-performing source models from both these optimizations are not very similar (Fig. 9 and Figs. S36 and S37). The best model of the guided optimization is a subset of the exploratory optimization source parameter estimates, but is not the best performing model in that ensemble. The misfit of the best fitting model from the exploratory optimization is lower than from the best fit model of the guided optimization (Fig. S35). We note that for the guided optimization several source parameters estimates converge, e.g. nucleation x and time, which do not converge in the exploratory optimization. The source parameter estimates and especially the best model of the guided optimization also represent the results of the backprojection much better (Fig. 10). Fits for the static displacement data can be found in Figure S31 and for the waveforms in Figure S32 with trace weights at stations shown in the Figures S34 and S33. The best fit model of the exploratory optimization produces better fits for some waveforms but performs worse for the static displacement fits in comparison to the best fit model of the guided optimization.

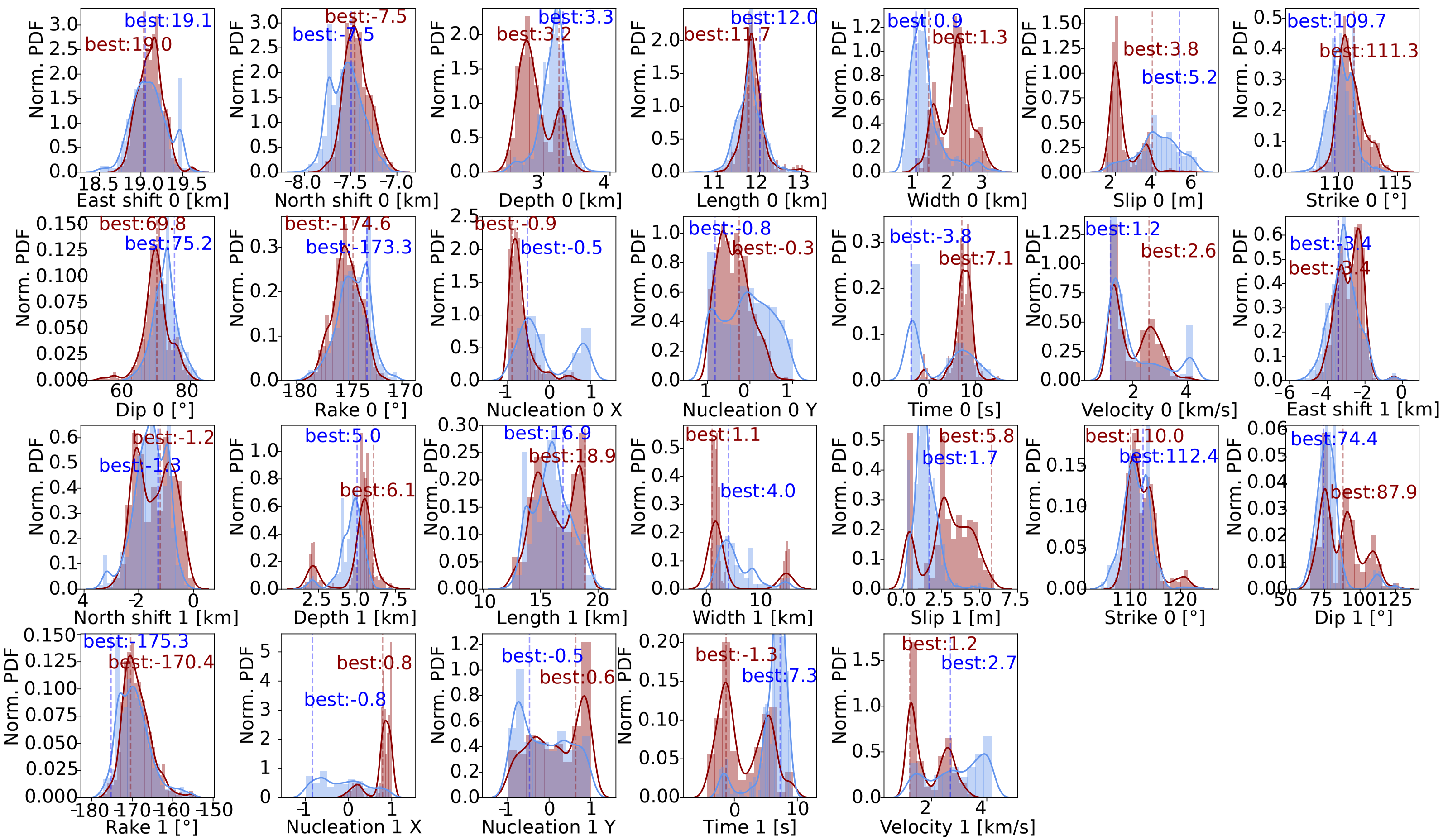

Figure 9: Histograms of the source parameter estimates for the exploratory (blue) and guided (red) optimizations. The best models are indicated by a dashed line. Source parameter names are indexed with 0 for the eastern segment (e.g. "Depth 0") and with 1 are for the western segment.

3.5.2 Backprojection of synthetic waveforms from the 2016 Muji minimum-misfit kinematic source model

We test if the waveforms of our best-fit two-segment source model of the 2016 Muji earthquake lead to similar spatio-temporal semblance results in a multi-array BP as the observed waveforms. To synthesize waveforms we use the same assumptions for the medium model as in all other analyses and calculate synthetic waveforms up to frequencies of 8 Hz for the same stations that we used in the real-data BP and we apply the same BP settings (Section 5). We add no noise to the synthetic waveforms.

We obtain LF and HF BP results for synthetic P- and SH-phases shown as cumulative semblance maps in Figure 11 and as time-incremental semblance maps in Figure S30). The semblance maps strongly resemble the real-data semblance maps (Fig. S24 and Fig. 6). They show very similar locations and numbers of high-semblance peaks. The synthetic semblance is spatially somewhat more focused, particularly for the LF BP. We also carry out a synthetic BP for a best-fit single-segment source model of the Muji 2016 earthquake (Fig. S29). Overall the synthetic BP results of the two-segment source model match the real-data semblance pattern more closely than the BP results of the single-segment source model. The synthetic LF semblance map for the single-segment source model shows a single high-semblance peak only, which is further east compared to the real-data LF semblance map. The HF synthetic semblance is missing the particularly strong central semblance peak apparent in the real-data HF semblance map. Both these features and the timing of semblance peaks are well reproduced using the two-segment source model.

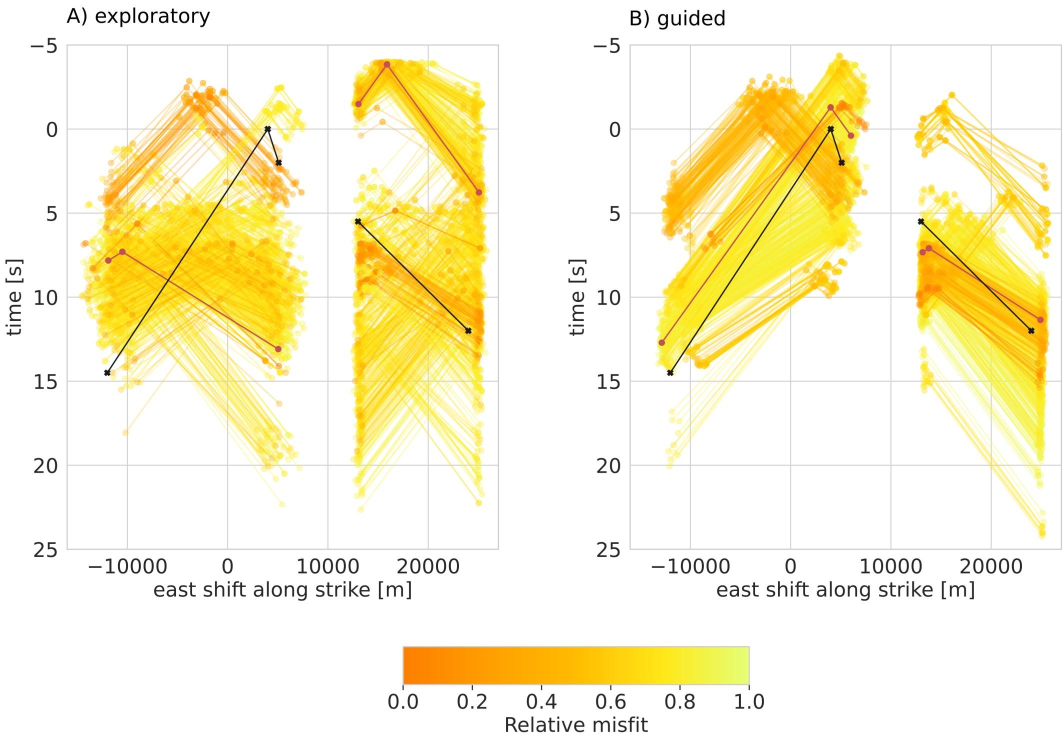

Figure 10: Rupture nucleation and termination times plotted as a function of distance along strike for each source segment for both exploratory (a) and guided (b) optimization ensembles (b). Single models are colored by misfit with the best model drawn in black. The corresponding unweighted HF BP result is shown with a red line.

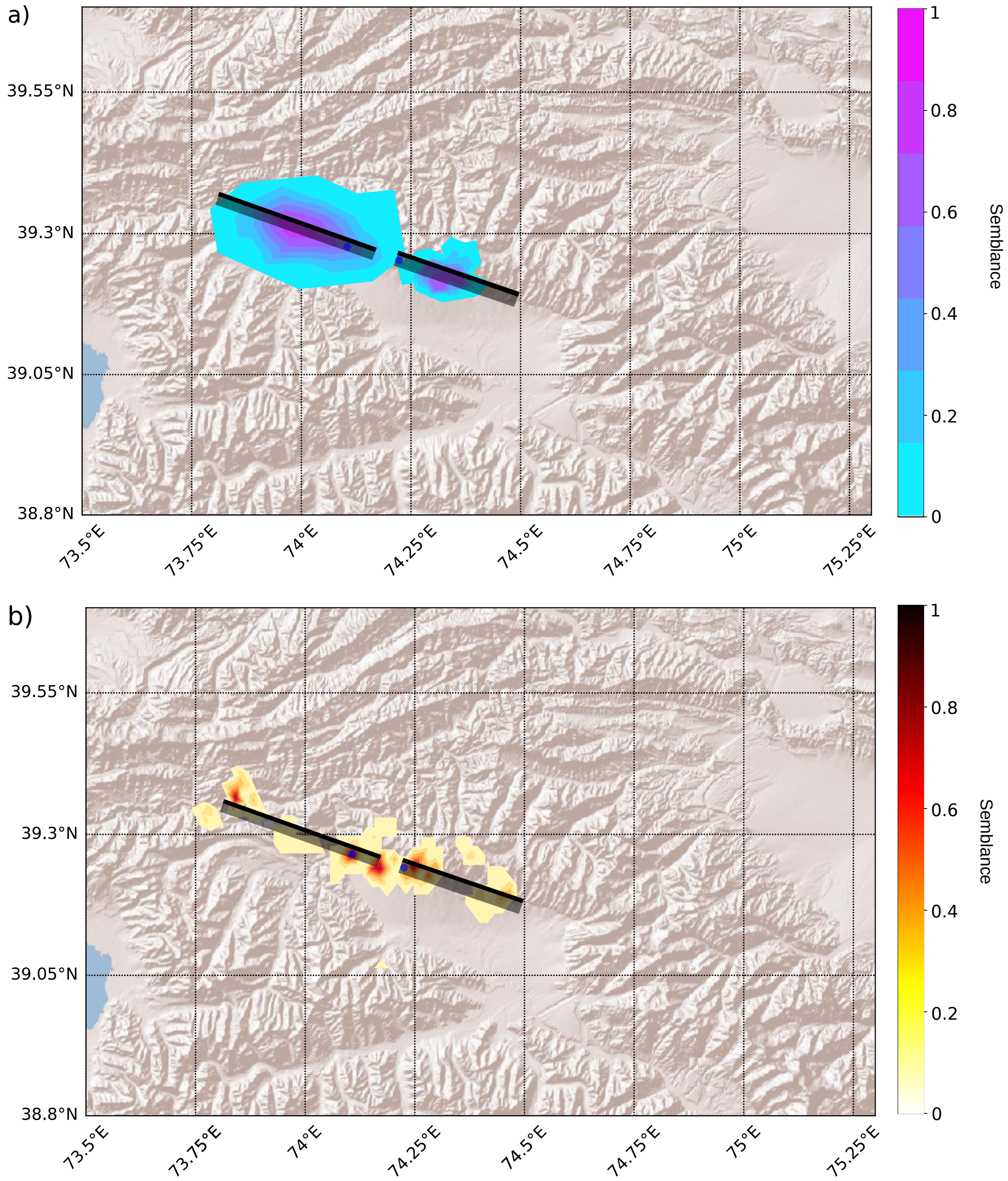

Figure 11: Contour plots of the LF (a) and HF (b) cumulative semblance from BP of synthetic waveforms of the minimum-misfit, two-segments kinematic source model. The source segment outlines are shown as gray-shaded rectangles, with the thick lines indicating the upper edges. Blue circles give the positions of the rupture nucleation points on each segment.

4 Discussion

4.1 Discussion of multi-array backprojection

The presented multi-array BP shows in synthetics tests a pleasingly high performance in recovering the horizontal location, the time, and rupture history with high accuracy. In comparison to BP using data from a large array, our multi-array BP with many small-aperture arrays clearly achieves stronger spatial focusing of seismic energy (Fig. 2b and Fig. S4). The presented multi-array BP has been applied successfully to other earthquakes already, e.g. the 2016 Mw 7.1 Romanche transform-fault earthquake (Hicks et al., 2020) and the 2008-2009 Qaidam earthquake sequence (Daout et al., 2020).

The focusing ability of our multi-array BP in depth direction is less precise compared to the horizontal resolution. The reason is that the depth direction is subparallel to the dominant path of wave propagation from shallow earthquake sources to far-field stations of a global network. Therefore, multi-array BP shares the generally relatively poor depth resolution for Mw 6-7 earthquake studies based on the global network of seismic station (Engdahl et al., 1998; Maggi et al., 2002), which may only be improved using more sophisticated methods (Craig, 2019). To account for poor depth resolution, our multi-array BP uses a purely horizontal grid of source points at a fixed depth. Seismic energy that is emitted at depths below or above to the chosen grid depth may appear horizontally shifted to the real horizontal location in the corresponding semblance maps. The accuracy of the location of the semblance therefore depends on an appropriate grid depth compared to the seismic source and may generally vary across the semblance maps for ruptures with a large depth extent.

Potentially an inclined source grid representing a known fault could be implemented in our multi-array BP to circumvent such a bias. However, the fault geometry should be well known, since wrong assumptions on the plane location and orientation will again lead to shifts in the backprojected seismic energy. For many applications, in particular those similar to the here presented earthquake case studies, fault location and fault geometry are unknowns to be constrained in the problem. Using a volume of grid points for the BP is possible as well and could be implemented in the presented framework. However, this requires a source-station configuration with many near-enough stations providing sufficient resolution in depth to enable focusing in this direction. In the general case, using a horizontal grid is in our eyes the least strict assumption.

From the LF and HF BP results of the 2016 Muji earthquake (Figs. 5, S24 and 6) we infer the spatio-temporal evolution of the rupture. The 2016 Muji rupture starts at the eastern end of the western segment and from there propagates simultaneously eastwards and westwards. The time-incremental high-frequency semblance maps show five peaks within the duration of 15 s. We interpret the first semblance peak as a representation of the rupture nucleation or start phase. We then observe slow rupture propagation from the nucleation point to both the east and west from that potential nucleation point. A second mapped semblance peak occurs east of that location, which we interpret as a rupture stop phase at the eastern end of the western segment. A third semblance peak occurs seconds later some five kilometers to the east, which likely represents the start phase on the eastern segment. No coherent seismic energy emission is mapped between the locality of these two mapped semblance peaks in either the low-frequency or the high-frequency semblance maps. Another high-frequency semblance peak occurs several seconds later to the east of the third peak, possibly representing the stop phase on the eastern segment. The last high-frequency semblance peaks is mapped 30 km west of the first semblance occurrence and potentially indicates the western-end stop phase of the western segment. We observe two distinct and separated patches of significant semblance in the low-frequency semblance map, which indicate a rupture segmentation. We interpret the time-incremental LF and HF semblance map as a bilateral rupture and the rupture jumping from the western segment to the eastern segment without emitting coherent seismic energy in between. This agrees with a previously postulated slip gap between the two segments (Feng et al., 2017). The area where the rupture segmentation and slip gap occurs coincides with mapped fluvial terraces that show a right-lateral offset across the Muji fault (Chevalier et al., 2011) and lies at the outlet of the longest glacial valley in the Muji range. The termination of the rupture on the western segment is located at a previously mapped discontinuity in the surface fault traces (Chevalier et al., 2011). We find a co-location of significant static surface displacement and the cumulative LF semblance map in the near field of the 2016 Muji earthquake. Such an agreement is to be expected and it has been observed before (Okuwaki et al., 2018; Yin & Denolle, 2019). Since static surface displacement correlates strongly with moment and therefore with fault slip, it is in close neighborhood to the excitation of seismic waves. A similarly good agreement between static InSAR surface displacements and the semblance from multi-array LF BP has been found for the 2008 and 2009 Qaidam earthquakes (Daout et al., 2020).

We estimate the rupture velocity of the 2016 Muji earthquake based on the HF spatio-temporal semblance to be within the range from 1.8 km/s to 2.1 km/s for the larger western segment and from 2.1 km/s to 2.6 km/s for the eastern segment (Fig. S25). These velocities are much slower than the rupture velocity of 3.7 km/s found in the Muji source analysis by Bie et al. (2018) for their most plausible scenario, based on the inferred source time functions and the rupture geometry. Bie et al. (2018) see a significant overlap of the two modelled subevent source-time functions that seems to indicate a near simultaneous rupture of both rupture segments. HF semblance peaks in the BP results appear co-located with boundaries of high gradients in static InSAR surface displacements. This is similar to results found for the 2008 and 2009 Qaidam earthquakes (Daout et al., 2020). Furthermore, no HF semblance is mapped in the area of an apparent slip gap between two regions of high static surface displacement. The first HF high-semblance peak that is close to the eastern edge of the western segment and the rupture seems to propagate through or jump the area which has been identified in previous studies as a slip gap (Bie et al., 2018; Feng et al., 2017), without emitting significant coherent seismic energy within the frequency bands considered in this study. The rupture on the western segment appears to start slow (Fig. 6b). The estimate of the rupture velocity on the western segment is likely representing an average between the initial and late stage rupture velocities (Fig S25). The total duration of the rupture as inferred from the beampower of our multi-array BP agrees well with the duration of the optimal SCARDEC source-time functions (Vallée and Douet (2016), Fig. 5), but is shorter by 1 s or 2 s in comparison to the average SCARDEC STF. Further comparisons between STFs and beampower, e.g. in shape, are not meaningful, since they represent different measures of the rupture process.

Intriguing is the strong resemblance of the real-data multi-array BP semblance with the semblance based on synthetic waveforms of the best-fit two-segment source model (Fig. 11), despite the fact that the kinematic source model is rather simple. It consists of only two rectangular source models with uniform slip and constant rupture velocity. Already such first-order source characteristics appear to describe the source well enough to well predict the waveforms up to a frequency of at least 1.5 Hz. It proves that our multi-array BP can reveal source geometry properties as well as other first-order rupture parameters for $M<7$ earthquakes. Multi-array BP shows a high potential to add value in future inverse source modeling problems.

4.2 Surface displacement map segmentation method

The image segmentation methods that we apply on surface displacement maps to extract probable fault traces prove to work well in synthetic tests (Figs. 3, S15 and S18). They enable recovering of the true source position with an accuracy of 100 m and the true length with an accuracy of 500 m. The inferred fault traces are closely located to the well studied fault traces for the 2016 Muji and the 2009 L'Aquila earthquakes. However, we caution against the direct use of the inferred fault trace location in fault mapping or as a fixed position in source optimizations, for a number of reasons. First, it is debated how well observed surface ruptures and surface deformation represent the slip and fault geometry at depth (Dolan & Haravitch, 2014; Soliva et al., 2008). Second, we observe biases in the fault trace location estimates for deeper earthquake sources and due to the line-of-sight projection of the three-dimensional surface displacement in InSAR data. Using more than one dataset that have different line-of-sight vectors will reduce this bias to some degree. We also note that the method might be susceptible to a very heterogeneous slip distribution. We underline again that the aim of the method is not to find the true fault line but rather derive pseudo-probabilities of the fault location for prior model parameter distributions. The assumption of the here used very simple source geometries, e.g. a planar fault, is suitable for low-parametric source modeling.

Our application of surface displacement map segmentation was very successful in synthetic tests (Figs. 2,S29,S5,S2 and Figures in Supplement 2.1) and on the InSAR data of the 2016 Muji earthquake. The co-seismic fault trace identified in previous studies (J. Li et al., 2019) agrees well with the result of our obtained fault pseudo-probability map (Fig. 7) and we note an agreement of the inferred fault traces (Fig. 7) with the field mapping of the Muji fault trace (Chevalier et al., 2016, 2011) and the reported co-seismic surface ruptures (T. Li et al., 2019; Chen et al., 2016).

The presented set of methods is straightforward. We believe that with few modifications only they can be applied to pixel-offset maps from optical images and to wrapped-phase interferograms. The image segmentation can be profitable also as a stand-alone signal detection method that in an automated way is used to detect significant deformation signals at specific sites and in big-data catalogues. As we show, the method is suited to produce fast estimates of source parameters. The method can be applied to big-data catalogues of InSAR surface displacements to automatically identify and characterize first-order source parameters.

4.3 Guided optimization

In our guided optimization we use the source parameter pseudo-probabilities which we estimated beforehand based on the multi-array backprojection and the image segmentation of surface displacement maps. With these source parameter prior distributions we succeed in automatically tailoring the model space for an efficient start of the direct parameter search. The guided optimization, in comparison with a more exploratory optimization, needs significantly less sampling to converge. It is not possible to give a simple indicator on performance gain here, because the gain strongly depends on the model parameter space that is chosen for the exploratory optimization, which in itself strongly depends on the problem. Usually, the model space of exploratory optimizations is either based on parameter bounds chosen by the researcher based on earthquake information, data visuals, experience and else, or on very wide parameter ranges that allow for almost all possible solutions. In the first case the benefit of the here proposed methods stems less from the potentially reduced optimization cost, but rather from the reduced need of supervision by a human researcher. In the latter case the gain will definitely be largely reduced computational cost, while the implementation of multi-array backprojection and image segmentation comes at its own cost. From our point of view, the main advantage of including prior information from multi-array backprojection method into kinematic modeling is that it gives physics-based evidence to model rupture segmentation with distinct sources.

Tailoring of the model space potentially excludes the global minimum model, which is a serious risk. While the chosen priors include extra margins from bootstrapping, we further reduce this risk by enlarging the model space after the initial tailored phase in the optimization. From this point of only the parameter distributions of low-misfit models drive the selection of new models in the widened model space.

We used some soft model space tailoring in the exploratory optimization as well, which also form the enlarged model space of the guided optimization after its initial phase. For the source parameters strike, dip and rake we based this soft tailoring on the expected focal mechanisms from the World Strain Map (Kreemer et al., 2014). Similarly we employed information from source time functions in the SCARDEC catalog (Vallée & Douet, 2016) for the onset time of 2016 Muji earthquake in the exploratory optimization. Both these data sources could also inform future operational source parameter optimizations in an automated fashion.

We demonstrate the practicality of the guided optimization in its application to the 2016 Muji earthquake. Here, it decreased the parameter ranges of the model space to be searched significantly. As a consequence, the guided optimization arrived comparatively early at low-misfit models (Fig. S36). The final source parameters for length, strike and position are close to the prior distributions (Fig. 8) determined by the surface displacement map segmentation method. The estimated prior distributions of source parameters compare well to the kinematic source model parameter estimates. This is also true for the source parameters nucleation position and time as inferred from the multi-array backprojection. Including prior parameter distributions in this way not only speeds up the convergence, it also helps resolving common parameter trade-offs in kinematic source modeling, e.g. between the onset times and positions of the nucleation points in case of two sub-sources. We note that the best-performing models of the guided and exploratory optimizations differ for the onset time, nucleation position and rupture velocity. Here, the source model ensembles of the guided optimization form a subset of the ensembles of the exploratory optimization.

5 Conclusions

We present a multi-array backprojection (BP) method and image segmentation applied to InSAR surface displacement measurements to improve the imaging of the spatio-temporal evolution of the rupture process of an earthquake. The information that we assemble based on these methods not only boosts follow-up non-linear kinematic source optimizations. They also enable an beforehand objectively informed setting of the number of source segments for a rupture. The multi-array BP method uses many small virtual arrays instead of a single large array to form a combined semblance maps from many single-array responses. We realize a large number of arrays with good coverage across azimuth and distance by clustering the globally available broadband stations into virtual arrays. In the combined semblance unwanted side lobes are suppressed that may result from e.g. depth phases. Additionally, we combine P- and SH-wave semblances to further increase the resolution of the semblance. Furthermore, our multi-array BP allows for an estimation of the semblance precision by using Bayesian bootstrapping of the single array contribution. In this bootstrapping we account for modeling errors due to uncertainties in the earth velocity structure by randomly perturbing traveltimes. Our synthetic tests with realistic station distributions and real noise proved the method to be robust. We show that it is capable of resolving the location of synthetic sources with a location error of less than 5 km in low-frequency semblance maps and with less than 2 km in high-frequency semblance maps. Included in our presentation of the multi-array BP is a novel approach for obtaining empirical travel time corrections. It is based on a semi-automatic search of a set of traveltime corrections maximizing the semblance of an fore- or aftershock.

We apply the multi-array BP method successfully to the real data of the 2016 Mw 6.6 Muji earthquake. For the semblance maps from the 2016 Mw 6.6 Muji earthquake we find a spatial precision of maximum 30 km and 7 km for the low-frequency and high-frequency semblance maps, respectively. We note that significant cumulative semblance, especially in the low-frequency results, corresponds well with significant surface displacement measured with InSAR. This in turn strongly supports the high accuracy that we estimated in the synthetic tests. From the BP results we infer a bilateral and segmented rupture starting close to the eastern end of the western segment, jumping a seismic gap to the eastern segment and propagating on both segments simultaneously. The rupture terminates first on the smaller eastern segment. We find a higher average rupture velocity of 2.1 km/s to 2.6 km/s for the eastern segment in comparison to the average 1.8 km/s to 2.1 km/s for the larger western segment.

The second presented method, an image segmentation approach applied to surface displacement maps as measured with InSAR that, performs well in estimating the location, orientation and the number of segments of a rupture. We show how this information can be cast into prior model parameter distributions for a multi-segment finite kinematic source model. In synthetic tests we demonstrate that with image segmentation we successfully recover fault strike, length, horizontal position and number of input sources. In the application of this method to the InSAR data of the 2016 Muji earthquake we find very good agreements of the results with the results of the multi-array BP, field mappings of the fault trace and also estimated fault characteristics in other source studies of this earthquake.

Both developed methods can be used separately and as stand-alone methods, to provide useful information about the rupture process. They could become regular parts in future operative frameworks. All developed algorithms are available as open-source software. In our work here we implemented them to ultimately infer prior model parameter distributions to be used in a guided joint-data two-sources non-linear optimization of the 2016 Muji earthquake. The resulting two-segment kinematic rupture model is not only consistent with the seismic waveforms and surface displacement data used in the inverse modeling, but also with the rupture evolution as inferred through the multi-array backprojection. Additionally the this guided optimization converged faster compared to the exploratory optimization without such prior source information.

Our results supports previous reports that 2016 Muji earthquake has been a bilateral rupture, with the rupture starting on the western segment to propagate eastward and westward on this segment. After an initiation phase the rupture appears to jump to the eastern segment.

The presented methods ease the detection of significant rupture segmentation, in particular for shallow crustal earthquakes with Mw$<$7, and are suitable to be applied in an automated fashion. A better and more frequent imaging of rupture complexity can be crucial for a better mapping of crustal faults and understanding of crustal faulting.

Data availability statement

Data from regional seismometers are available via FDSN services from GEOFON and IRIS. SAR images used are openly available from the Copernicus Open Access Hub at `https://scihub.copernicus.eu`. The dynamic Green's function store is uploaded on the Pyrocko project Green's mill repository `https://greens-mill.pyrocko.org/` as "$global\_4hz$" for reproduction. The static Green's function store is uploaded for reproduction as "$pamir\_static$". The software and algorithms for the displacement map segmentation have been released on Zenodo under DOI: 10.5281/zenodo.4465169 and for the multi-array backprojection have been released on Zenodo under DOI: 10.5281/zenodo.4465171 .

Author contribution statement

A.S. wrote the original draft, developed and conceptualized the method and software, visualized and processed the data. H.S. reviewed and edited the draft, provided supervision, project administration and conceptualized the methods. F.K. reviewed and edited the draft, provided supervision and conceptualized the backprojection method.

All authors have read, agreed, and participated to the published version of the manuscript.

Acknowledgments

This research was funded by the German Research Foundation DFG through an Emmy-Noether Young-Researcher-Grant (Number 276464525, "Bridging Geodesy and Seismology for improved and automated estimation of faulting events").

Contains modified Copernicus Sentinel data (2016) for Sentinel data.

The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to waveforms, related metadata, and/or derived products used in this study. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience and EarthScope (SAGE) Proposal of the National Science Foundation under Cooperative Agreement EAR-1261681.

We thank Dr. Simon Daout and Dr. Hannes Vasyura-Bathke for helpful discussions on the method and results.

References

Aki, K., & Richards, P. G. (2002). *Quantitative seismology*.

Bao, H., Ampuero, J.-P., Meng, L., Fielding, E. J., Liang, C., Milliner, C., … Huang, H. (2019). Early and persistent supershear rupture of the 2018 magnitude 7.5 palu earthquake. *Nature Geoscience*, *12*(3), 200–205.

Bardsley, J. M., Solonen, A., Haario, H., & Laine, M. (2014). Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Scientific Computing*, *36*(4), A1895–A1910.

Bie, L., Hicks, S., Garth, T., Gonzalez, P., & Rietbrock, A. (2018). 'two go together': Near-simultaneous moment release of two asperities during the 2016 mw 6.6 muji, china earthquake. *Earth and Planetary Science Letters*, *491*, 34–42.

Bracewell, R. N., & Bracewell, R. N. (1986). *The fourier transform and its applications* (Vol. 31999). McGraw-Hill New York.

Cesca, S., Heimann, S., Stammler, K., & Dahm, T. (2010, jun). Automated procedure for point and kinematic source inversion at regional distances. *Journal of Geophysical Research*, *115*(B6), B06304. Retrieved from `http://doi.wiley.com/10.1029/2009JB006450` doi: 10.1029/2009JB006450

Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on image processing*, *10*(2), 266–277.

Chen, J., Li, T., Sun, J., Fang, L., Yao, Y., Li, Y., … Fu, B. (2016). Coseismic surface rupture and seismogenic muji fault of the 25 november 2016 aketao mw 6.6 earthquake in northern pamir. *Seismol. Geol*, *38*, 1160–1174.

Chevalier, M.-L., Leloup, P. H., & Li, H. (2016). Comment on "no late quaternary strike-slip motion along the northern karakoram fault" published by robinson et al. in epsl, 2015. *Earth and Planetary Science Letters*, *443*, 216–219.

Chevalier, M.-L., Li, H., Pan, J., Pei, J., Wu, F., Xu, W., … Liu, D. (2011). Fast slip-rate along the northern end of the karakorum fault system, western tibet. *Geophysical Research Letters*, *38*(22).

Chevalier, M.-L., Pan, J., Li, H., Liu, D., & Wang, M. (2015). Quantification of both normal and right-lateral late quaternary activity along the kongur shan extensional system, chinese pamir. *Terra Nova*, *27*(5), 379–391.

Craig, T. (2019). Accurate depth determination for moderate-magnitude earthquakes using global teleseismic data. *Journal of Geophysical Research: Solid Earth*, *124*(2), 1759–1780.

Daout, S., Steinberg, A., Isken, M. P., Heimann, S., & Sudhaus, H. (2020). Illuminating the spatio-temporal evolution of the 2008–2009 qaidam earthquake sequence with the joint use of insar time series and teleseismic data. *Remote Sensing*, *12*(17), 2850.

Dolan, J. F., & Haravitch, B. D. (2014). *How well do surface slip measurements track slip at depth in large strike-slip earthquakes? The importance of fault structural maturity in controlling on-fault slip versus off-fault surface deformation* (Vol. 388). doi: 10.1016/j.epsl.2013.11.043

Douze, E., & Laster, S. (1979). Statistics of semblance. *Geophysics*, *44*(12), 1999–2003.

Dziewonski, A., Chou, T.-A., & Woodhouse, J. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research: Solid Earth*, *86*(B4), 2825–2852.

Engdahl, E. R., van der Hilst, R., & Buland, R. (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America*, *88*(3), 722–743.

Fan, W., & Shearer, P. M. (2017). Investigation of backprojection uncertainties with m6 earthquakes. *Journal of Geophysical Research: Solid Earth*, *122*(10), 7966–7986.

Farr, T. G., & Kobrick, M. (2000). Shuttle radar topography mission produces a wealth of data. *Eos, Transactions American Geophysical Union*, *81*(48), 583–585.

Feng, W., Tian, Y., Zhang, Y., Samsonov, S., Almeida, R., & Liu, P. (2017). A slip gap of the 2016 m w 6.6 muji, xinjiang, china, earthquake inferred from sentinel-1 tops interferometry. *Seismological Research Letters*, *88*(4), 1054–1064.

Getreuer, P. (2012). Chan-vese segmentation. *Image Processing On Line*, *2*, 214–224.

Goldstein, R. M., Zebker, H. A., & Werner, C. L. (1988). Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio science*, *23*(4), 713–720.

Gurov, E., & Yamnichenko, A. Y. (1995). Morphology of rim of complex terrestrial craters. In *Lunar and planetary science conference* (Vol. 26).

Hanka, W., & Kind, R. (1994). The geofon program. *Annals of Geophysics*, *37*(5).

Hanssen, R. F. (2001). *Radar interferometry: data interpretation and error analysis* (Vol. 2). Springer Science & Business Media.

He, P., Ding, K., & Xu, C. (2018). The 2016 mw 6.7 aketao earthquake in muji range, northern pamir: Rupture on a strike-slip fault constrained by sentinel-1 radar interferometry and gps. *International journal of applied earth observation and geoinformation*, *73*, 99–106.

Heidbach, O., Rajabi, M., Cui, X., Fuchs, K., Müller, B., Reinecker, J., … others (2018).

The World Stress Map database release 2016: Crustal stress pattern across scales. *Tectonophysics*, *744*, 484–498.

Heimann, S. (2011). A Robust Method To Estimate Kinematic Earthquake Source Parameters. Retrieved from `http://ediss.sub.uni-hamburg.de/volltexte/2011/5357/`

Heimann, S., Isken, M., Kühn, D., Sudhaus, H., Steinberg, A., Vasyura-Bathke, H., ... Dahm, T. (2018). *Grond - A probabilistic earthquake source inversion framework.* Retrieved 2018-08-27, from `http://pyrocko.org/grond/docs/current/` doi: 10.5880/GFZ.2.1.2018.003

Heimann, S., Kriegerowski, M., Isken, M., Cesca, S., Daout, S., Grigoli, F., ... Dahm, T. (2017). *Pyrocko - An open-source seismology toolbox and library.* GFZ Data Services. doi: 10.5880/gfz.2.1.2017.001

Heimann, S., Vasyura-Bathke, H., Sudhaus, H., Isken, M. P., Kriegerowski, M., Steinberg, A., & Dahm, T. (2019a). A python framework for efficient use of pre-computed green's functions in seismological and other physical forward and inverse source problems. *Solid Earth*, *10*(6), 1921–1935. Retrieved from `https://se.copernicus.org/articles/10/1921/2019/` doi: 10.5194/se-10-1921-2019

Heimann, S., Vasyura-Bathke, H., Sudhaus, H., Isken, M. P., Kriegerowski, M., Steinberg, A., & Dahm, T. (2019b). A python framework for efficient use of pre-computed green's functions in seismological and other physical forward and inverse source problems.

Hicks, S. P., Okuwaki, R., Steinberg, A., Rychert, C. A., Harmon, N., Abercrombie, R. E., ... others (2020). Back-propagating supershear rupture in the 2016 m w 7.1 romanche transform fault earthquake. *Nature Geoscience*, *13*(9), 647–653.

Hong, T.-L., & Fujita, K. (1981). Modelling of depth phases and source processes of some central aleutian earthquakes. *Earth and Planetary Science Letters*, *53*(3), 333–342.

Ide, S. (2002). Estimation of radiated energy of finite-source earthquake models. *Bulletin of the Seismological Society of America*, *92*(8), 2994–3005.

Ishii, M., Shearer, P. M., Houston, H., & Vidale, J. E. (2005). Extent, duration and speed of the 2004 Sumatra-Andaman earthquake imaged by the Hi-Net array. *Nature*, *435*(7044), 933–936. doi: 10.1038/nature03675

Ishii, M., Shearer, P. M., Houston, H., & Vidale, J. E. (2007). Teleseismic p wave imaging of the 26 december 2004 sumatra-andaman and 28 march 2005 sumatra earthquake ruptures using the hi-net array. *Journal of Geophysical Research: Solid Earth*, *112*(B11).

Isken, M., Sudhaus, H., Heimann, S., Steinberg, A., Daout, S., & Vasyura-Bathke, H. (2017). *Kite - Software for Rapid Earthquake Source Optimisation from InSAR Surface Displacement.* GFZ Data Services. doi: 10.5880/gfz.2.1.2017.001

Johnson, D. H., & Dudgeon, D. E. (1993). *Array signal processing: concepts and techniques*. PTR Prentice Hall Englewood Cliffs.

Jónsson, S., Zebker, H., Segall, P., & Amelung, F. (2002). Fault slip distribution of the 1999 Mw 7.1 Hector Mine, California, earthquake, estimated from satellite radar and GPS measurements. *Bulletin of the Seismological Society of America*, *92*(4), 1377–1389.

Kennett, B., & Engdahl, E. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, *105*(2), 429–465.

Kiser, E., & Ishii, M. (2013). Hidden aftershocks of the 2011 mw 9.0 tohoku, japan earthquake imaged with the backprojection method. *Journal of Geophysical Research: Solid Earth*, *118*(10), 5564–5576.

Kiser, E., & Ishii, M. (2017). Back-Projection Imaging of Earthquakes. *Annual Review of Earth and Planetary Sciences*, *45*, 271–299.

Kreemer, C., Blewitt, G., & Klein, E. C. (2014). A geodetic plate motion and global strain rate model. *Geochemistry, Geophysics, Geosystems*, *15*(10), 3849–3889.

Krüger, F., & Ohrnberger, M. (2005). Tracking the rupture of the m w= 9.3 sumatra earthquake over 1,150 km at teleseismic distance. *Nature*, *435*(7044), 937.

Li, J., Liu, G., Qiao, X., Xiong, W., Liu, D., Sun, J., … others (2019). Rupture characteristics of the 25 november 2016 aketao earthquake (m w 6.6) in eastern pamir revealed by gps and teleseismic data. In *Earthquakes and multi-hazards around the pacific rim, vol. ii* (pp. 49–61). Springer.

Li, T., Schoenbohm, L. M., Chen, J., Yuan, Z., Feng, W., Li, W., … others (2019). Cumulative and coseismic (during the 2016 mw 6.6 aketao earthquake) deformation of the dextral-slip muji fault, northeastern pamir orogen. *Tectonics*, *38*(11), 3975–3989.

Li, W., Chen, Y., Yuan, X., Schurr, B., Mechie, J., Oimahmadov, I., & Fu, B. (2018). Continental lithospheric subduction and intermediate-depth seismicity: Constraints from s-wave velocity structures in the pamir and hindu kush. *Earth and Planetary Science Letters*, *482*, 478–489.

Lohman, R. B., & Simons, M. (2005). Some thoughts on the use of insar data to constrain models of surface deformation: Noise structure and data downsampling. *Geochem-*

istry, Geophysics, Geosystems, *6*(1).

Ma, Y., Qiao, X., Chen, W., & Zhou, Y. (2018). Source model of 2016 mw6. 6 aketao earthquake, xinjiang derived from sentinel-1 insar observation. *Geodesy and Geodynamics*, *9*(5), 372–377.

Madariaga, R. (1977). High-frequency radiation from crack (stress drop) models of earthquake faulting. *Geophysical Journal International*, *51*(3), 625–651.

Maggi, A., Priestley, K., & Jackson, J. (2002). Focal depths of moderate and large size earthquakes in iran.

McCloskey, J., & Nalbant, S. S. (2009). Near-real-time aftershock hazard maps. *Nature Geoscience*, *2*(3), 154–155.

Meng, L., Ampuero, J.-P., Sladen, A., & Rendon, H. (2012, apr). High-resolution backprojection at regional distance: Application to the Haiti M 7.0 earthquake and comparisons with finite source studies. *Journal of Geophysical Research: Solid Earth*, *117*(B4), n/a–n/a. Retrieved from `http://doi.wiley.com/10.1029/2011JB008702` doi: 10.1029/2011JB008702

Meng, L., Ampuero, J.-P., Stock, J., Duputel, Z., Luo, Y., & Tsai, V. (2012). Earthquake in a maze: Compressional rupture branching during the 2012 mw 8.6 sumatra earthquake. *Science*, *337*(6095), 724–726.

Meng, L., Zhang, A., & Yagi, Y. (2016). Improving back projection imaging with a novel physics-based aftershock calibration approach: A case study of the 2015 gorkha earthquake. *Geophysical Research Letters*, *43*(2), 628–636.

Okuwaki, R., Kasahara, A., Yagi, Y., Hirano, S., & Fukahata, Y. (2018). Backprojection to image slip. *Geophysical Journal International*, *216*(3), 1529–1537.

Okuwaki, R., & Yagi, Y. (2017). Role of geometric barriers in irregular-rupture evolution during the 2008 Wenchuan earthquake. *Geophysical Journal International*.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, *9*(1), 62–66.

Palo, M., Tilmann, F., Krueger, F., Ehlert, L., & Lange, D. (2014). High-frequency seismic radiation from maule earthquake (m w 8.8, 2010 february 27) inferred from high-resolution backprojection analysis. *Geophysical Journal International*, *199*(2), 1058–1077.

Ragon, T., Sladen, A., & Simons, M. (2018). Accounting for uncertain fault geometry in earthquake source inversions–I: theory and simplified application. *Geophysical Jour-*

nal International, *214*(2), 1174–1190.

Razafindrakoto, H. N., Mai, P. M., Genton, M. G., Zhang, L., & Thingbaijam, K. K. (2015). Quantifying variability in earthquake rupture models using multidimensional scaling: Application to the 2011 tohoku earthquake. *Geophysical Journal International*, *202*(1), 17–40.

Rössler, D., Krueger, F., Ohrnberger, M., & Ehlert, L. (2010). Rapid characterisation of large earthquakes by multiple seismic broadband arrays. *Natural Hazards and Earth System Sciences*, *10*(4), 923–932.

Rost, S., & Thomas, C. (2002). Array seismology: Methods and applications. *Reviews of geophysics*, *40*(3), 2–1.

Rost, S., & Thomas, C. (2009). Improving seismic resolution through array processing techniques. *Surveys in Geophysics*, *30*(4-5), 271–299.

Rubin, D. B. (1981). The bayesian bootstrap. *The annals of statistics*, 130–134.

Schimmel, M., & Gallart, J. (2007). Frequency-dependent phase coherence for noise suppression in seismic array data. *Journal of Geophysical Research: Solid Earth*, *112*(B4).

Schimmel, M., & Paulssen, H. (1997). Noise reduction and detection of weak, coherent signals through phase-weighted stacks. *Geophysical Journal International*, *130*(2), 497–505.

Shaus, A., & Turkel, E. (2016). Chan-vese revisited: relation to otsu's method and a parameter-free non-pde solution via morphological framework. In *International symposium on visual computing* (pp. 203–212).

Shearer, P. M. (1997). Improving local earthquake locations using the l1 norm and waveform cross correlation: Application to the whittier narrows, california, aftershock sequence. *Journal of Geophysical Research: Solid Earth*, *102*(B4), 8269–8283.

Soliva, R., Benedicto, A., Schultz, R., Maerten, L., & Micarelli, L. (2008). Displacement and interaction of normal fault segments branched at depth: Implications for fault growth and potential earthquake rupture size. *Journal of Structural Geology*, *30*(10), 1288–1299.

Steinberg, A., Sudhaus, H., Heimann, S., & Krüger, F. (2020). Sensitivity of insar and teleseismic observations to earthquake rupture segmentation. *Geophysical Journal International*.

Steinhaus, H. (1956). Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci*, *1*(804), 801.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341–359.

Styron, R. (2019, August). *GEMScienceTools/gem-global-active-faults: First release of 2019.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.3376300` doi: 10.5281/zenodo.3376300

Sudhaus, H., & Jónsson, S. (2009, feb). Improved source modelling through combined use of InSAR and GPS under consideration of correlated data errors: application to the June 2000 Kleifarvatn earthquake, Iceland. *Geophysical Journal International*, *176*(2), 389–404. Retrieved from `https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.2008.03989.x` doi: 10.1111/j.1365-246X.2008.03989.x

Vallée, M., & Douet, V. (2016). A new database of source time functions (STFs) extracted from the SCARDEC method. *Physics of the Earth and Planetary Interiors*, *257*, 149–157.

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... Yu, T. (2014). scikit-image: image processing in python. *PeerJ*, *2*, e453.

Wang, D., Kawakatsu, H., Mori, J., Ali, B., Ren, Z., & Shen, X. (2016). Backprojection analyses from four regional arrays for rupture over a curved dipping fault: The mw 7.7 24 september 2013 pakistan earthquake. *Journal of Geophysical Research: Solid Earth*, *121*(3), 1948–1961.

Wang, R. (1999). A Simple Orthonormalization Method for Stable and Efficient Computation of Green's Functions. *Bulletin of the Seismological Society of America*, *89*(3), 733–741.

Wang, R., Heimann, S., Zhang, Y., Wang, H., & Dahm, T. (2017). Complete synthetic seismograms based on a spherical self-gravitating earth model with an atmosphere–ocean–mantle–core structure. *Geophysical Journal International*, *210*(3), 1739–1764.

Wang, R., Lorenzo-Martín, F., & Roth, F. (2006). Psgrn/pscmp—a new code for calculating co-and post-seismic deformation, geoid and gravity changes based on the viscoelastic-gravitational dislocation theory. *Computers & Geosciences*, *32*(4), 527–541.

Weston, J., Ferreira, A. M., & Funning, G. J. (2012). Systematic comparisons of earthquake source models determined using InSAR and seismic data. *Tectonophysics*, *532*, 61–81.

doi: 10.1016/j.tecto.2012.02.001

Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., … others (2015). The 2013 european seismic hazard model: key components and results. *Bulletin of Earthquake Engineering*, *13*(12), 3553–3596.

Xu, Y., Koper, K. D., Sufri, O., Zhu, L., & Hutko, A. R. (2009). Rupture imaging of the mw 7.9 12 may 2008 wenchuan earthquake from back projection of teleseismic p waves. *Geochemistry, Geophysics, Geosystems*, *10*(4).

Xu, Y., LIU, J.-H., LIU, F.-T., ZHU, L.-R., LONG, H.-Y., & Wei, B. (2006). Crustal velocity structure and seismic activity in the tianshan-pamir conjunctive zone. *Chinese Journal of Geophysics*, *49*(2), 417–425.

Yao, H., Shearer, P. M., & Gerstoft, P. (2012). Subevent location and rupture imaging using iterative backprojection for the 2011 tohoku m w 9.0 earthquake. *Geophysical Journal International*, *190*(2), 1152–1168.

Yin, J., & Denolle, M. A. (2019). Relating teleseismic backprojection images to earthquake kinematics. *Geophysical Journal International*, *217*(2), 729–747.

Supplement to: Data-driven constraints on earthquake modeling and segmentation from teleseismic multi-array backprojection and InSAR

Andreas Steinberg[1,2], Henriette Sudhaus[1]and Frank Krger[3]
[1]Christian Albrechts University Kiel, Otto-Hahn-Platz 1, Germany
[2]Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany
[3]Institute for Earth and Environmental Sciences, University of Potsdam, Potsdam, Germany

1. Exploratory Optimization with Bayesian Bootstrapping

We estimate model parameter uncertainties alongside an optimization by Bayesian bootstrapping. Here bootstrapping is realized through Bayesian random weighting (Rubin, 1981) of the seismic waveforms and through residual bootstrapping with synthetic correlated noise on the InSAR data to form multiple objective functions for a single forward-model realization. The misfit weighting of the waveforms respects the uncorrelated data error between stations caused by e. g. site effects. The synthetic noise $\epsilon_{\mathrm{syn},i}$ used in the residual bootstrapping is generated based on the estimated variance-covariance functions of the data error (Sudhaus & Jónsson, 2009) and reflects the apparent data error. We use a large number of different sets with random weights and synthetic noise for these multiple misfit calculations, usually above 100, and achieve as many different bootstrap optimization chains. Once the optimizations converge, the best-fit models of each bootstrap chain may start to diverge, when the data error becomes significant with respect to the difference in model fit, and form model ensembles. In this way, which is very similar to the so called "Randomize-then-Optimize" procedure (Bardsley et al., 2014), we retrieve source parameter distributions similar to a Markov Chain Monte Carlo sampling of the model space (Jonsson et al., 2014).

The optimization that involves a large number of bootstrap chains works in the following way. Each bootstrap chain shares the same sampled models, but because of the different weighting, the misfit of a model is different in each bootstrap chain. A source model may perform well in one bootstrap chain, but poorly in another. Throughout the optimization we monitor a given number of best-fit models of each chain, to which we refer to as the highscore list of the chain. The number of models in the highscore lists is defined dependent on the number of model parameters N_{par}. The highscore list acts as a memory of past visited models, which allows the sampler to retain several good models and explore multiple minima, which is especially important for optimizing models with several earthquake sources. The highscore list of each bootstrap chain will therefore differ and converge differently. The differences between the performance of the models in each bootstrap chain represent the uncertainty of the models with respect to the data error.

The optimization is a direct-search optimization and has two distinct phases. The first phase is a random sampling (uniform distribution) of the model parameter space, constrained by given upper and lower parameter bounds. Here this first phase samples 20.000 models. The uniform distribution as prior probability of the earthquake source parameters for the initial sampling is well justified if the parameter space is large and the solution unknown. This creates a unbiased set of sampled initial starting solutions. At the end of the first phase, the best-performing models are determined for each bootstrap chain with its specific objective function and collected in the corresponding highscore lists for each bootstrap chain.

The bootstrap chain highscore lists are playing a vital role in the second optimization phase, the "directed sampling". The bootstrap chain that shares the least number of models in its highscore list with other highscore lists determines the sampling of the next model. This ensures that also the directed phase is still exploring the model space. The new model is drawn from a multivariate normal distribution based on the variance-covariance matrix $\mathbf{R}$ of the source model parameters from all models currently in the respective highscore list. We use the excentricity compensated method to give models with less neighbors a higher probability to be drawn and considered as the center of the search space. The search space is scaled by a factor a. This scale factor is logarithmically decreasing from the first sampling $a_{\mathrm{start}} = 2$ to the last sampling $a_{\mathrm{end}} = 1$ of this second optimization phase. In other words, the search space is an ellipsis in the model space, around a highscore list model, which is shrinking with increasing number of models sampled. Each newly drawn model is ranked in all bootstrap chains. They enter a highscore list if they outperform any of the current highscore list models. The formerly largest-misfit model in the highscore list is removed from it. With each new model in the highscore lists their statistics change and with it a search radius for a new model. At the start of the optimization the different highscore lists likely contain the same models. Only when the misfit starts to differ subtly between models, the data errors reflected in the different objtective functions start to play a role in the ranking of the well-performing models.

0148-0227/21/$9.00

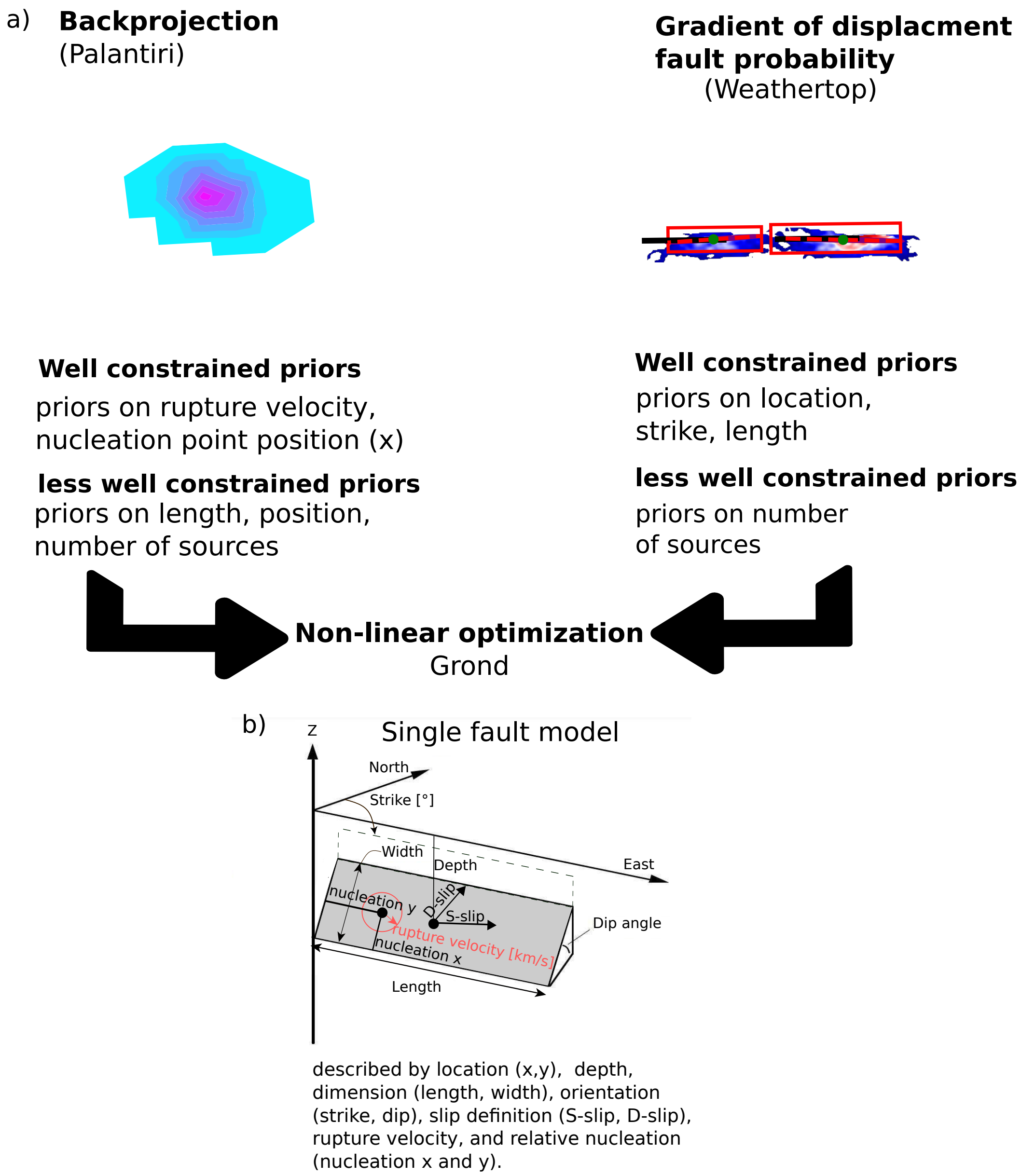

Figure S1: a) Scheme of how the proposed methods, the multi-array backprojection and the displacement map segmentation feed prior information into a non-linear optimization. b) Sketch of finite source model used as for forward model and its source parameters. modeling.

2. Additional synthetic tests of the multi array backprojection

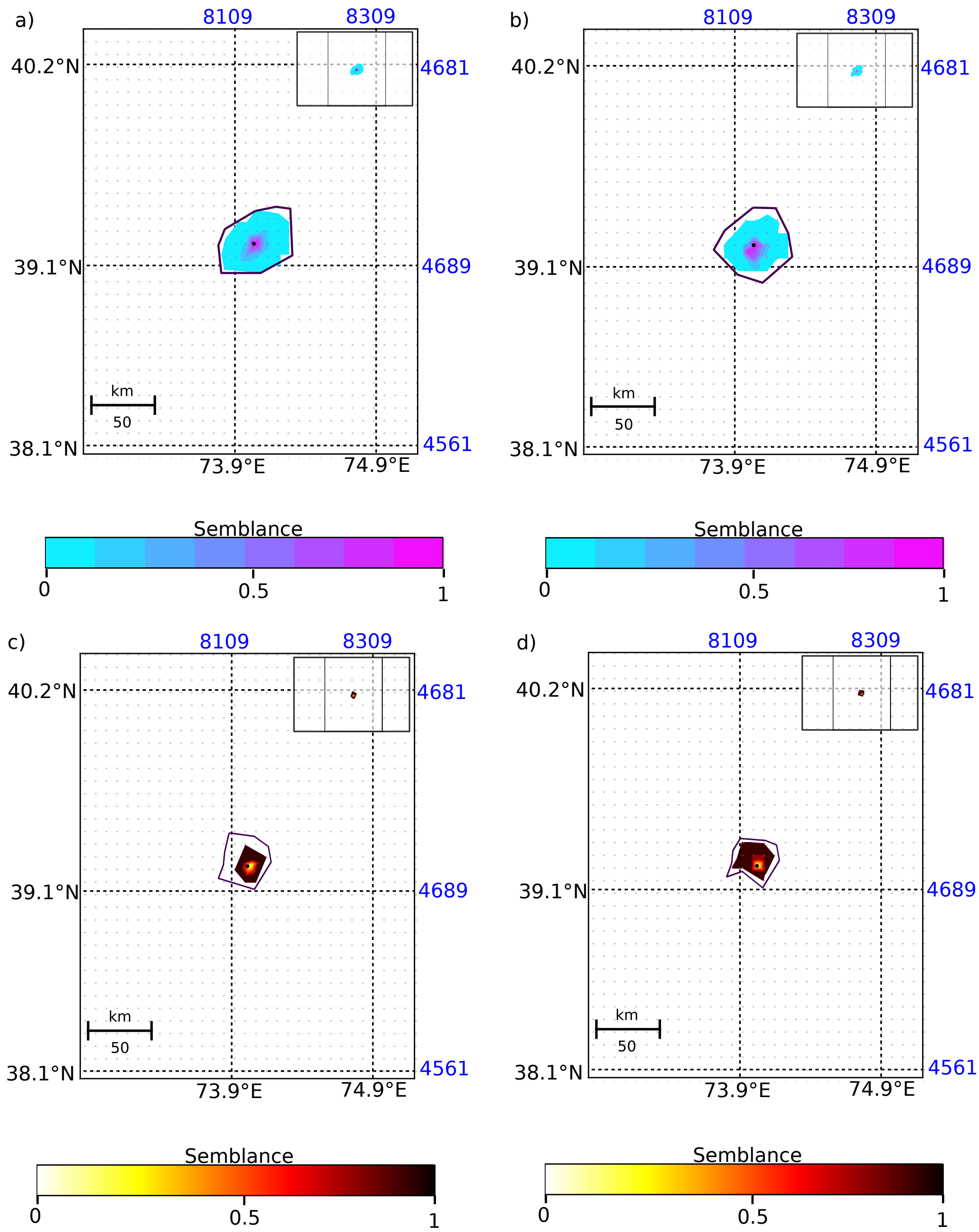

Figure S2: Cumulative semblance from the backprojection of a synthetic DC source backprojection (Test 1) using the Muji 2016 earthquake array setup for a) LF and P-wave only, b) LF and S-wave only, c) HF and P-wave only and d) HF and S-wave only. Semblance is plotted as contour color plot. The black outlines represent the 68% precision estimate from bootstrapping on the semblance maxima location. They are drawn as a minimum bounding outlines for the locations of the maxima from 100 bootstraps. The image is a zoom in and the extent of the whole search grid is given in the top right. The travel-time grid points are indicated as gray dots in the background. The black dot indicates the true position of the synthetic source. Coordinates are given in Latitude/Longitude (black) and UTM (blue).

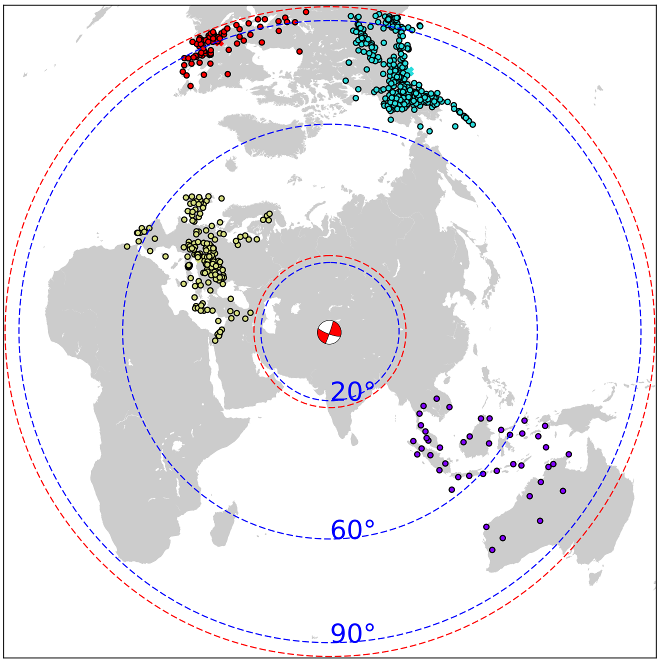

Figure S3: Stations combined to large arrays used for the synthetic backprojection Test 2 , (Sec. 2.1.6) for the 2016 Muji earthquake plotted on a world map. The stations belonging to the same array share the same color.

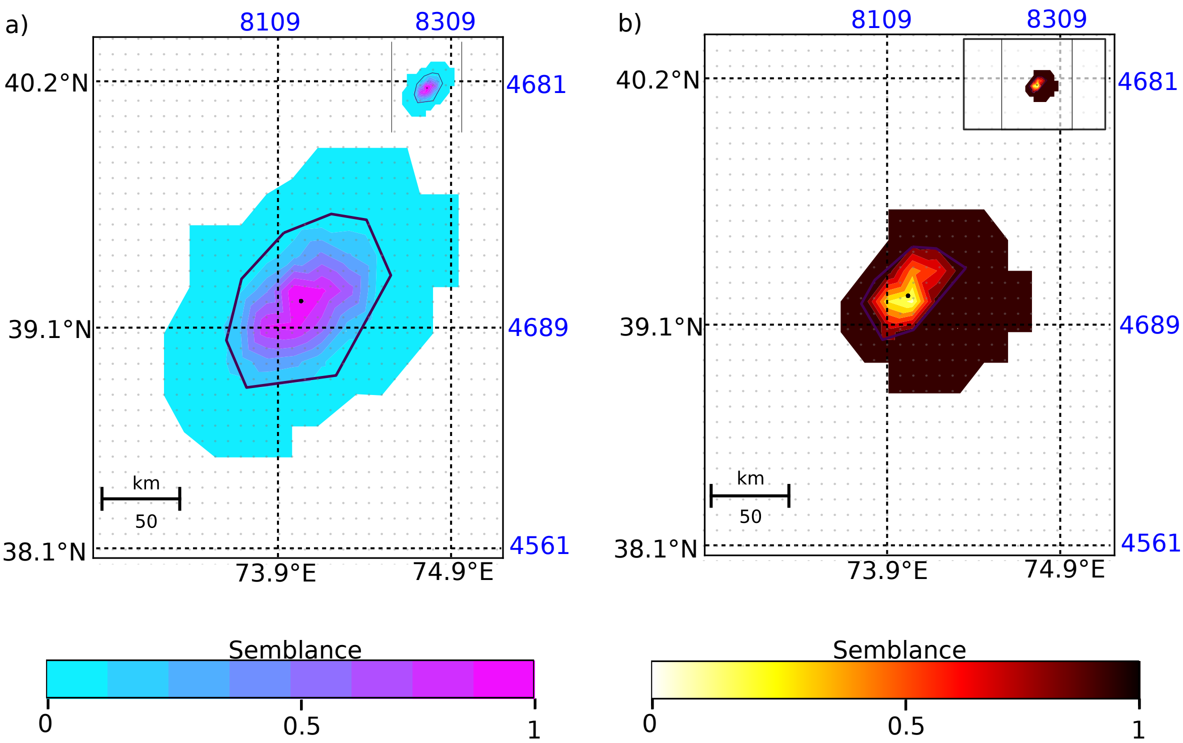

Figure S4: Cumulative semblance from the backprojection of a synthetic DC source (Test 2, Sec. 2.1.6) for P- and SH-waves using the large array setup for the 2016 Muji earthquake (Fig. S3) for a) LF and b) HF. Shown is the cumulative semblance from all timesteps from the non-bootstrapped LF synthetic single DC source backprojection using large arrays. The outlines in black for the LF and in red for the HF indicates the uncertainty from the bootstrapped semblance. Other details as in Fig. S2.

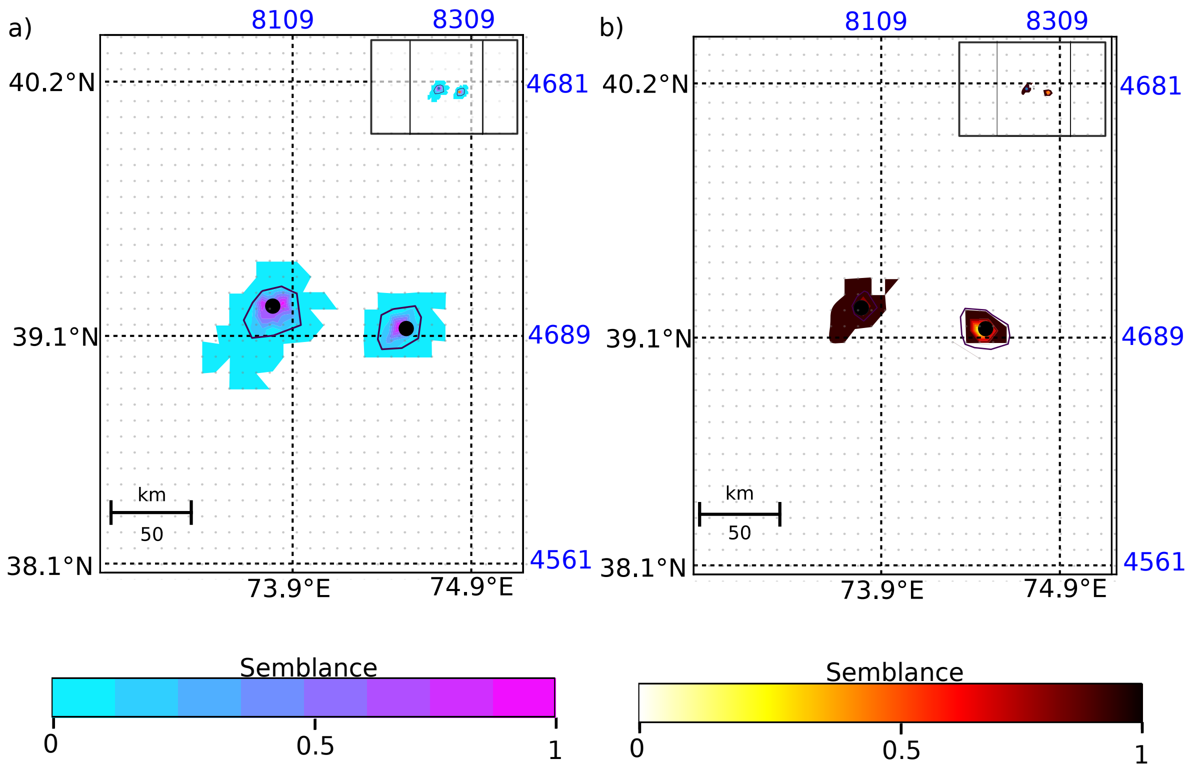

Figure S5: Cumulative semblance from the backprojection of two synthetic DC sources (Test 4, Sec. 2.1.6) from P- and SH-waves using the Muji 2016 earthquake array setup for a) LF and b) HF. The blue and orange dots indicate the true position of the two input sources used for forward calculation. The outlines in black for the LF and in red for the HF indicates the uncertainty from the bootstrapped semblance. Other details as in Fig. S2.

2.1. Additional backprojection synthetic tests based on Ahar

We carry out additional synthetic backprojection tests based on a another set of stations, mimicking the situation for the Mw 6.3 2012 Ahar earthquake, resulting in a different azimuthal coverage and distance distribution. Again, we backproject two differently filtered datasets, one at high frequencies, 0.25-1.5 Hz and one at low frequencies, 0.01-0.24 Hz. In all cases the source is set to be equivalent of a Mw 6.3 earthquake. The waveforms have been randomly shifted by up to +/- 2 s to simulate model errors. The array weights have been bootstrapped 100 times and the semblance is weighted by azimuth. For each synthetic waveform real pre-event noise from the corresponding waveform real record from before the 2012 Ahar earthquakes is added. We use a 4 Hz Green's function store to calculate the synthetics based on the QSSP code by (Wang, 1999) and use the AK-135 traveltime model.

We test for the recovery of the position of a single point-source using the clustering of stations into small virtual arrays. The station and array map can be found in Fig. S7. The source is defined with a triangular source-time function of 3 s duration. Backprojection results are shown for low frequencies in Fig. S8a and for high frequencies, 0.25-1.5 Hz, in Fig. S8b. At both frequencies the source position can be recovered. The source is set at 8.7 km depth and the traveltime grid is calculated the same depth.

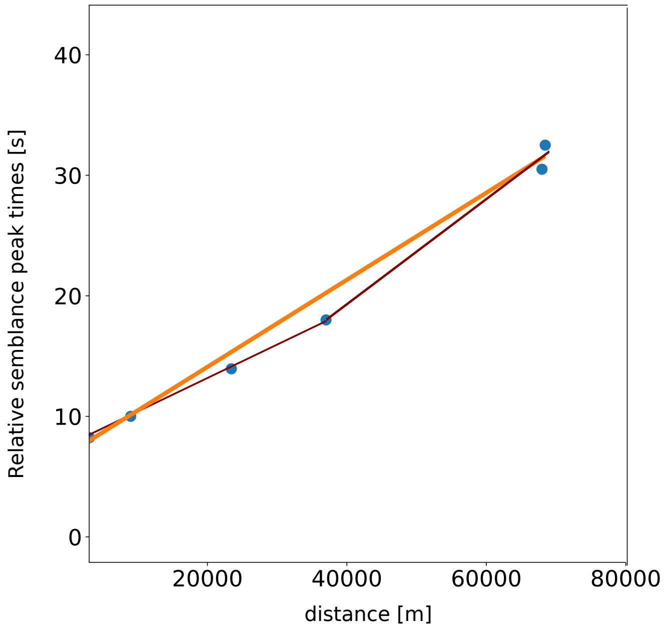

Figure S6: Time-Distance plot for the line source. Time is relative to the first window with semblance. Blue dots indicate the first and last maxima of the high-frequency BP, the orange line the estimated velocity (4000 m/s) and the red line the true velocity.

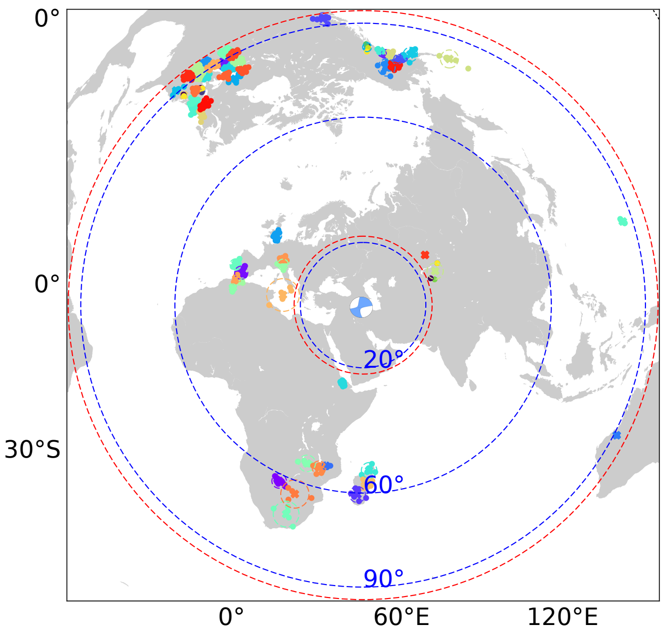

Figure S7: The stations used for the synthetic backprojections based on the 2012 Ahar earthquake with multi-array clusters. The stations belonging to the same array share the same color.

Similar to the synthetic test of the Muji 2016 earthquake we repeat the same synthetic test (we keep all parameters the same as before) but use large arrays S9 instead of the smaller virtual arrays used before. The results (Lf and HF, Figs. S10a and S10b) shows broader distributed semblance mappings in comparison to the smaller virtual arrays.

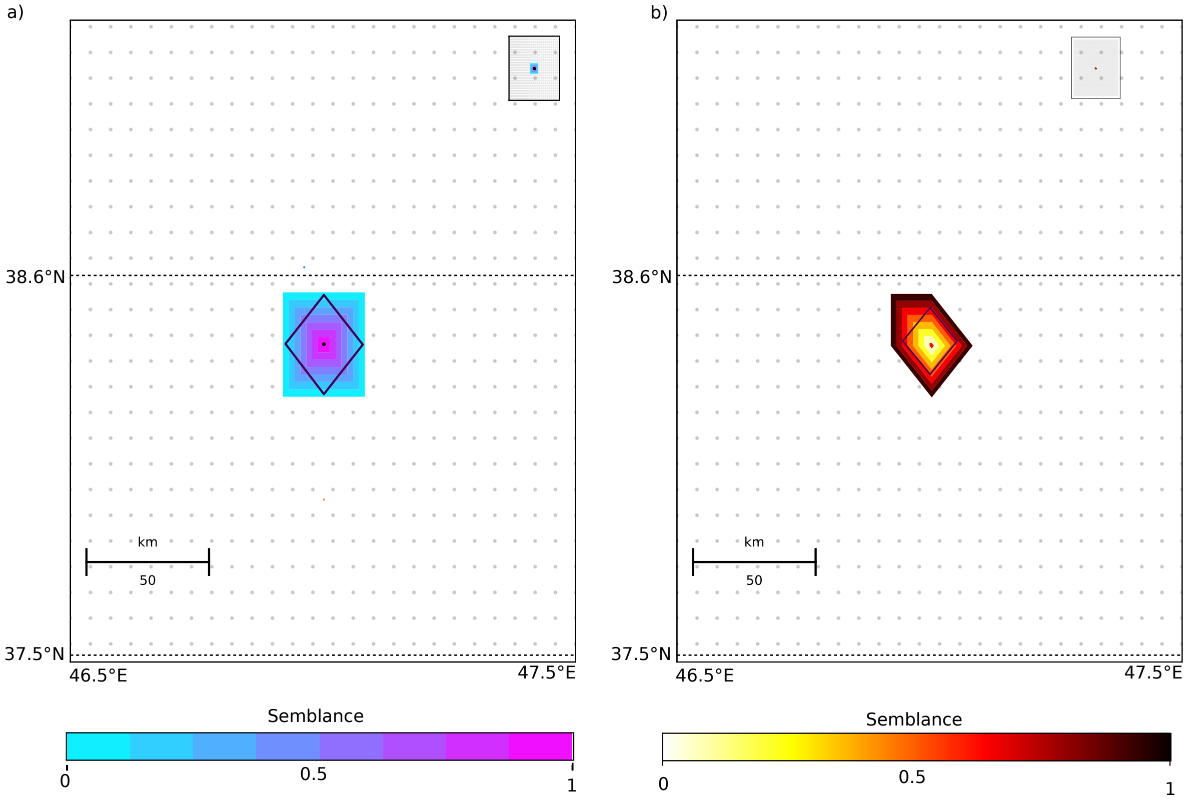

Figure S8: Cumulative semblance from the backprojection of a synthetic DC source (comparable to Test 1, Sec. 2.1.6) for P- and SH-waves using the more sparse array setup of the 2012 Ahar earthquake (Fig. S7) of a) LF and b) HF. The source is located approximately at the location of the 2012 Ahar earthquake. Shown is the cumulative semblance from all timesteps from the non-bootstrapped LF synthetic single DC source backprojection using large arrays. Other details as in Fig. S2. The outlines in black indicate the uncertainty from the bootstrapped semblance.

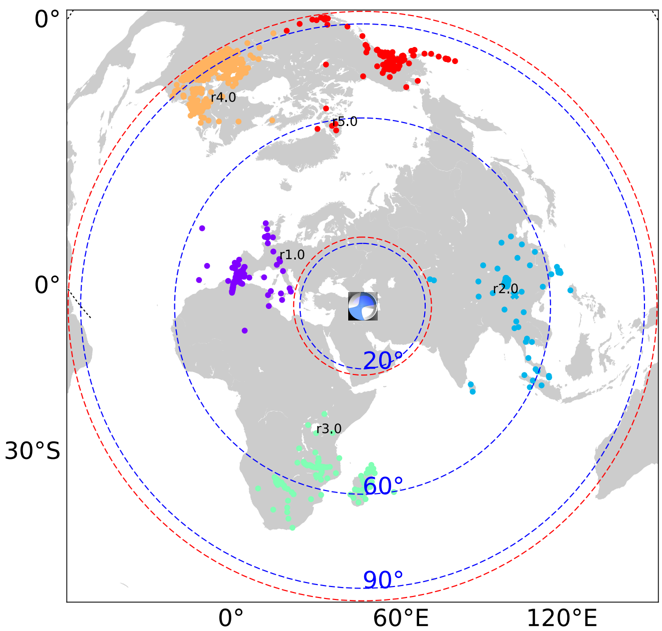

Figure S9: The stations used for the synthetic backprojections based on the 2012 Ahar earthquake with large arrays. The stations belonging to the same array share the same color.

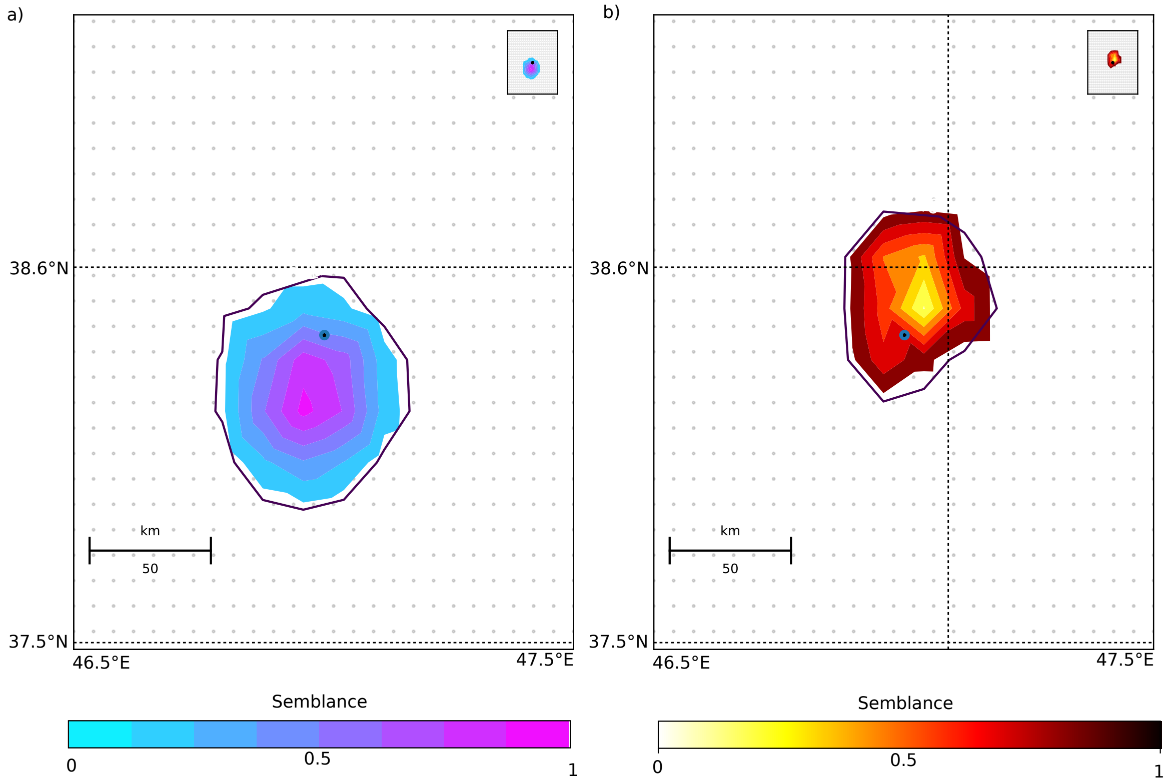

Figure S10: Cumulative semblance from the backprojection of a synthetic DC source (comparable to Test 2, Sec. 2.1.6) for P- and SH-waves using the large-array setup of the 2012 Ahar earthquake (Fig. S9) and for a) LF and b) HF. The source is located approximately at the location of the 2012 Ahar earthquake. Other details as in Fig. S2.

We also tested the recovery of signals from a backprojection of a synthetic forward modelled line source of 80 km length (a finite rectangular source with very small width of 0.1 m and a dip of 90°) with nucleation at the eastern edge. The low-frequency backprojection shows a broader distribution of significant semblance (Fig. S11a). For the high-frequency backprojection (Fig. S11b) the start and stop phases can be recovered. The rupture speed on the fault was set to 4000 m/s and approximately recovered by taking the distance and time between the first and last semblance maxima (Fig. S12).

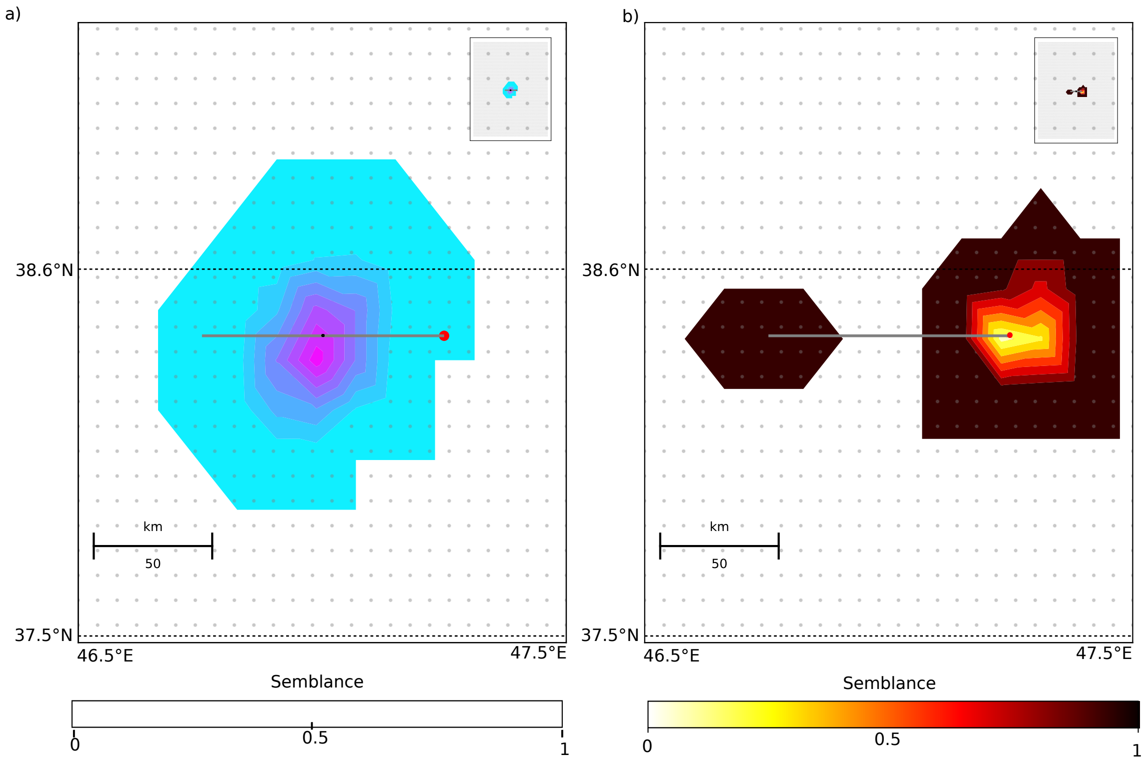

Figure S11: Cumulative semblance from the backprojection of a synthetic horizontal line source (comparable to Test 3, Sec. 2.1.6) for P- and SH-waves using the array setup of the 2012 Ahar earthquake (Fig. S7) for a) LF and b) HF. The source is located approximately at the location of the 2012 Ahar earthquake. Shown is the cumulative semblance from all timesteps from the non-bootstrapped LF synthetic backprojection. Other details as in Fig. S2.

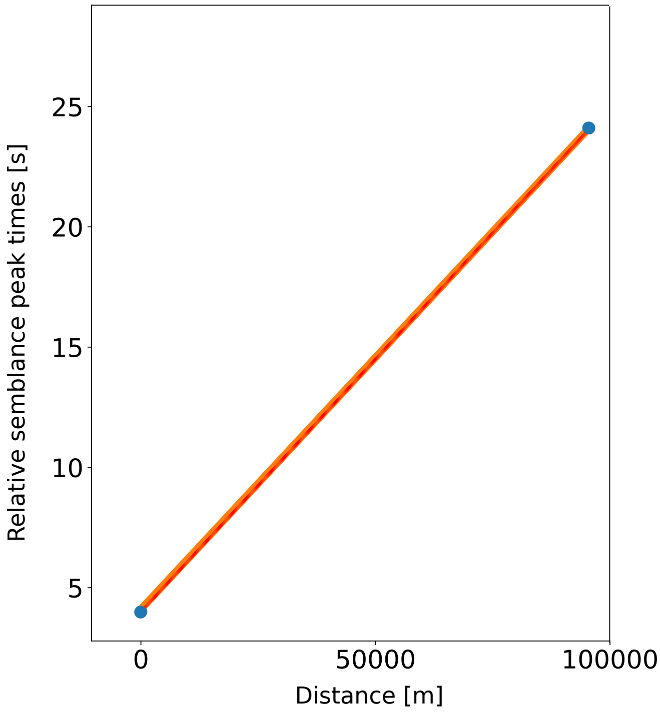

Figure S12: Time-Distance plot for the synthetic line source HF backprojection of the 2012 Ahar earthquake as seen in Fig. Time is relative to the first window with semblance. S11b. Blue dots indicate the first and last maxima of the high-frequency backprojection, the orange line the estimated velocity (4000 m/s). The red line indicates the true velocity.

Another synthetic test is conducted for a vertical line source with top depth 1 km and bottom depth 21 km (20 km length), dip 90° and very small width of 0.1 m. The nucleation starts at the bottom. Again we carry out the tests for low-frequency backprojections (Fig. S13a) and for high-frequency backprojection (Fig. S13b). The rupture speed on the fault was set to 4000 m/s. The start and stop phase spatially overlay each other.

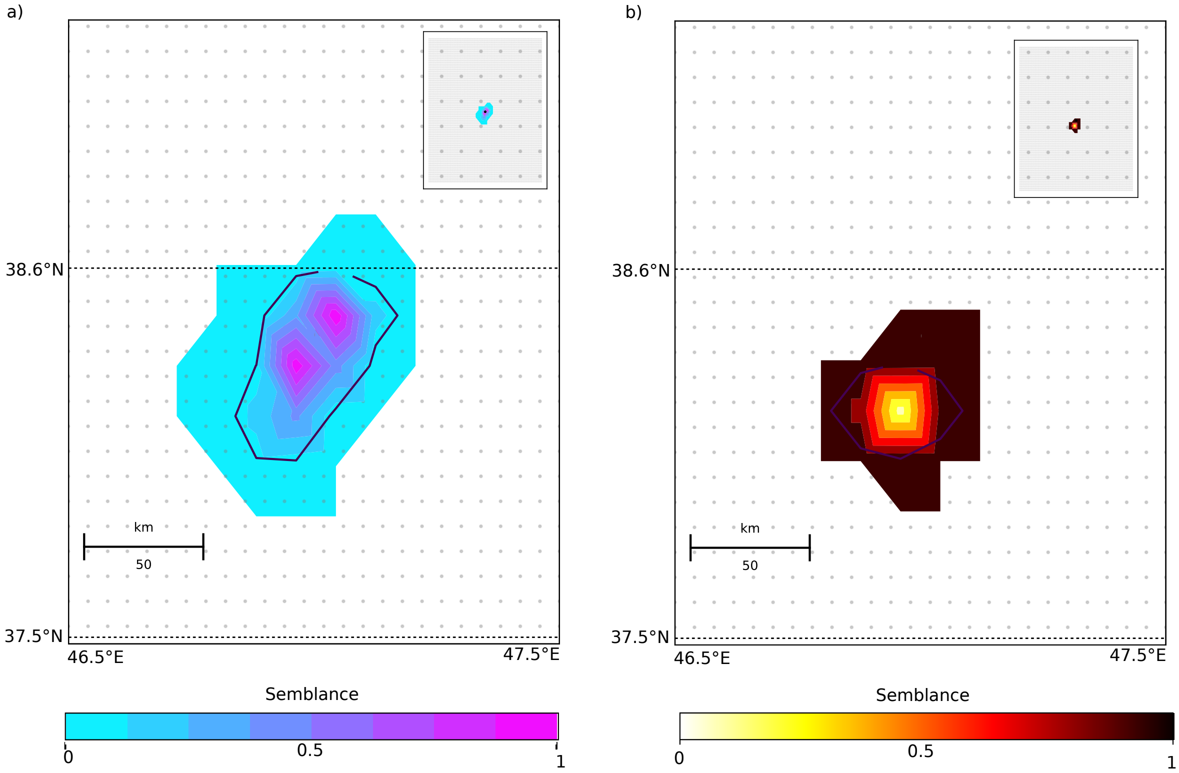

Figure S13: Cumulative semblance from the backprojection of a synthetic vertical line source for P- and SH-waves using the array setup of the 2012 Ahar earthquake (Fig. S7) of a) LF and b) HF. The source is located approximately at the location of the 2012 Ahar earthquake. Shown is the cumulative semblance from all timesteps from the non-bootstrapped LF synthetic backprojection. The gray dot indicates the true position of the synthetic line source (vertical). Other details as in Fig. S2.

We also carry out a synthetic test for two point-sources of same moment, duration and timing, but which are spatially separated by 50 km (Fig. S14a, Fig. S14b).

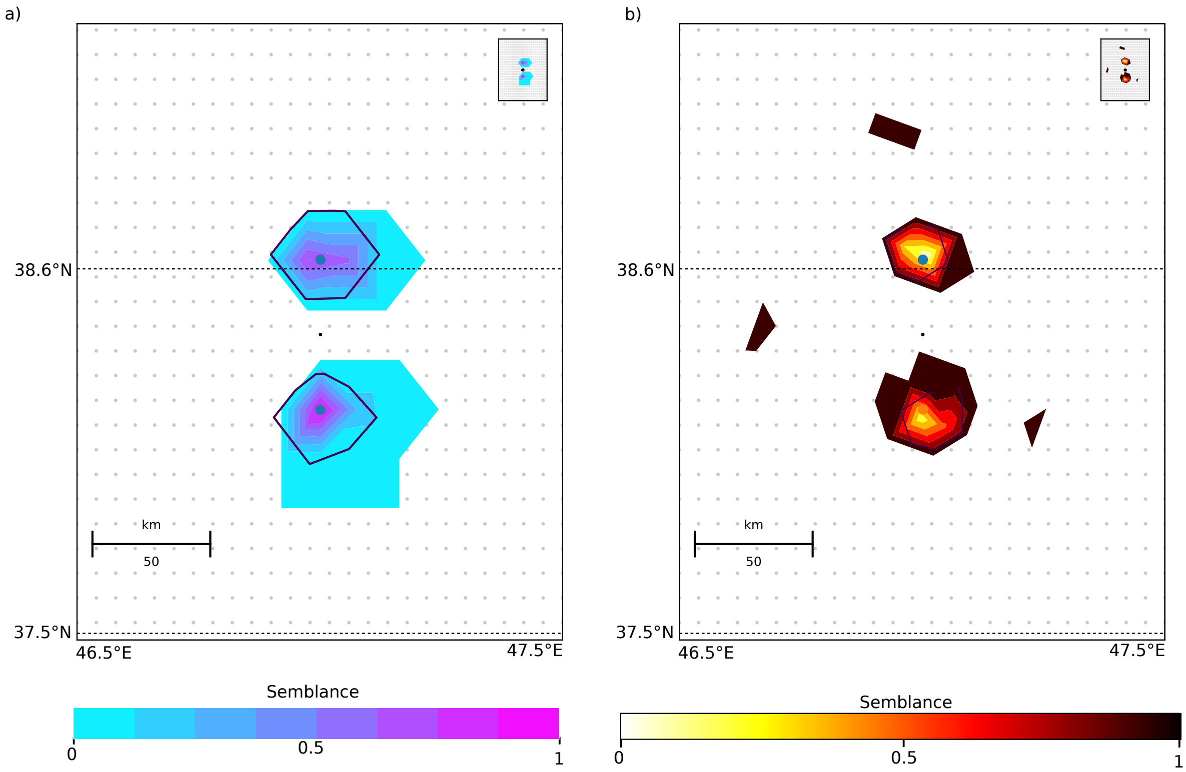

Figure S14: Synthetic backprojection of P- and SH-waves for two DC sources (comparable to Test 4, Sec. 2.1.6), using the array setup of the 2012 Ahar earthquake (Fig. S7) for a) LF and b) HF. The source is located approximately at the location of the 2012 Ahar earthquake. Shown is the cumulative semblance from all timesteps from the non-bootstrapped LF synthetic of the two DC sources backprojection. The blue circles indicate the true position of the synthetic sources. Other details as in Fig. S2.

3. Additional synthetic tests of displacement map segmentation

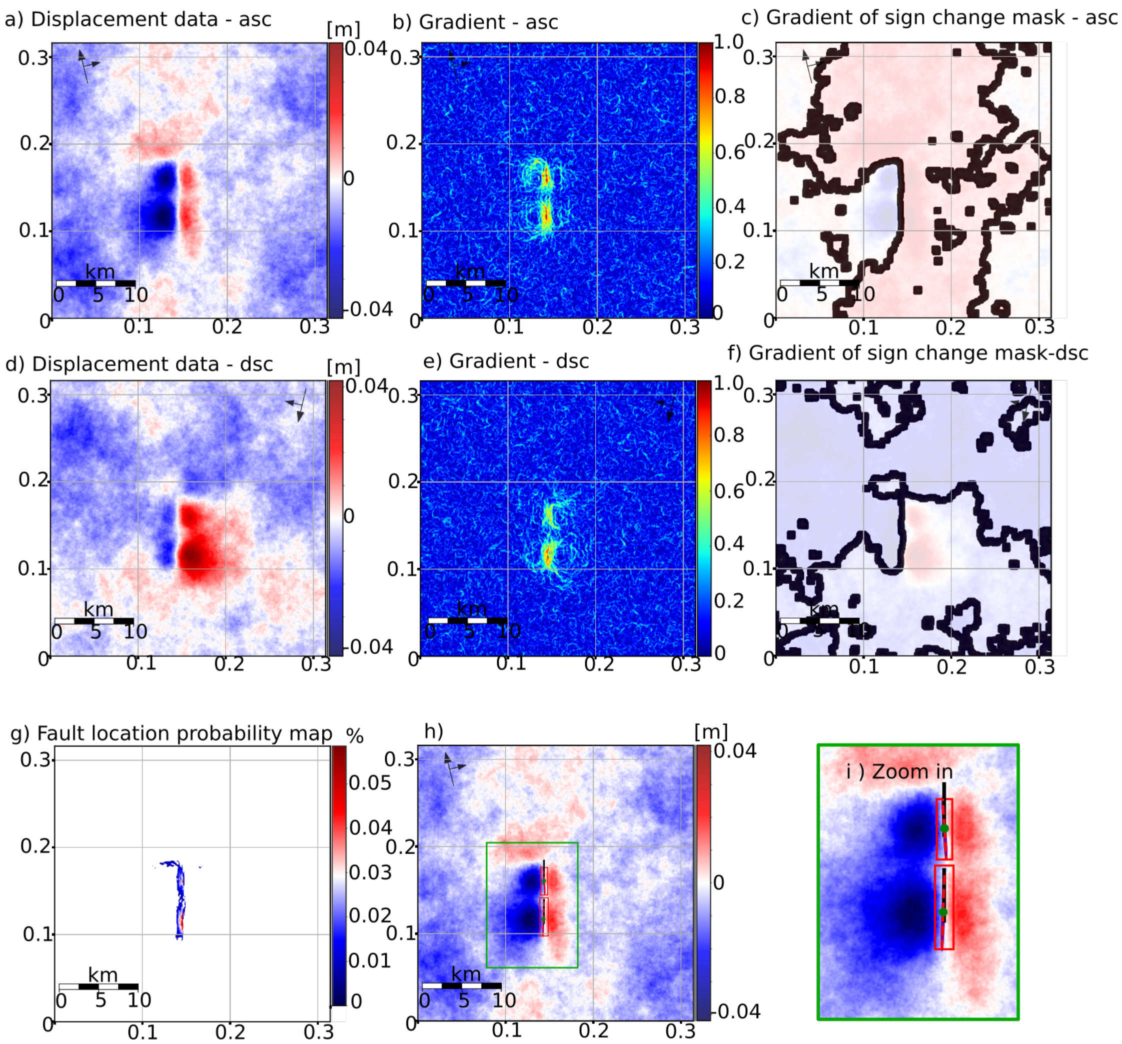

Figure S15: Synthetic test of the displacement map segmentation method for two strike-slip sources trending north-south at 0.5 km top edge depth. a), b) and c) contain the displacement data, the gradient and the gradient of the sign change mask (superimposed on the displacement data), respectively, for the ascending dataset and d), e) and f) accordingly for the descending data. g) shows the normalized combined product of the gradient sign change mask with the gradient from ascending and descending InSAR data. Values below 1% of the maximum value are masked out. This map is used as a pseudo-probability estimate for the position of the fault(s) location centroid. h) shows the bounding boxes and ellipses applied on the product shown in g). The green box is the area of interest, zoomed into in i). The red dashed line indicates the major axis of the ellipses containing the highest values for each region found as described above. The outline of the synthetic source(s) is indicated in the figures with black lines that are thicker for the top edge. The ellipses (indicated by the purple outline) is centered at the centroid of each region.

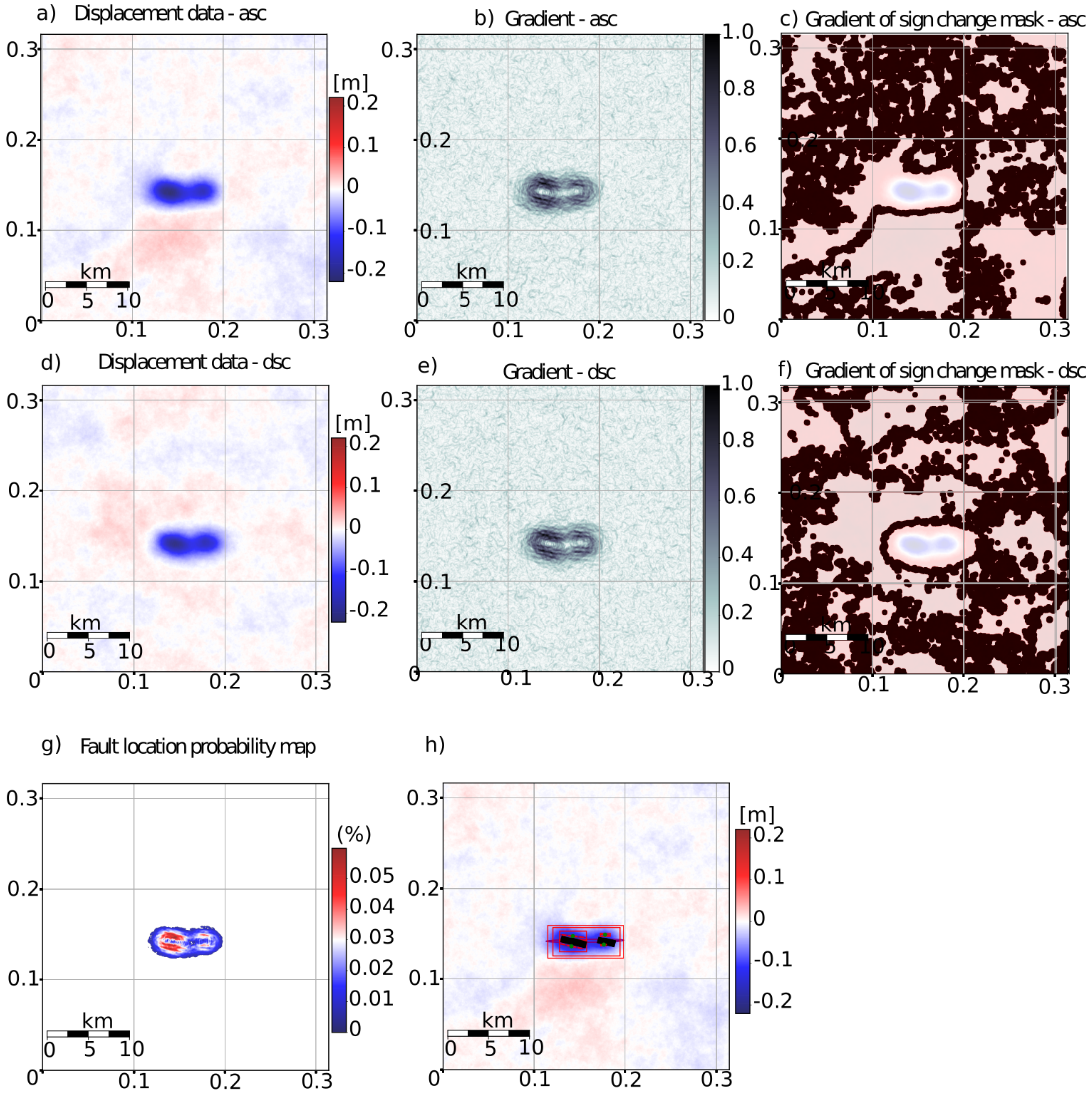

Figure S16: Synthetic test of the displacement map segmentation method for a two normal dip-slip earthquakes at a depth of 0.5 km. a), b) and c) contain the displacement data, the gradient and the gradient of the sign change mask (superimposed on the displacement data), respectively, for the ascending dataset and d), e) and f) accordingly for the descending data. g) shows the normalized combined product of the gradient sign change mask with the gradient from ascending and descending InSAR data. Values below 1% of the maximum value are masked out. This map is used as a pseudo-probability estimate for the position of the fault(s) location centroid. h) shows the bounding boxes and ellipses applied on the product shown in g). The green box is the area of interest, zoomed into in i). The red dashed line indicates the major axis of the ellipses containing the highest values for each region found as described above. The outline of the synthetic source(s) is indicated in the figures with black lines that are thicker for the top edge. The ellipses (indicated by the purple outline) is centered at the centroid of each region.

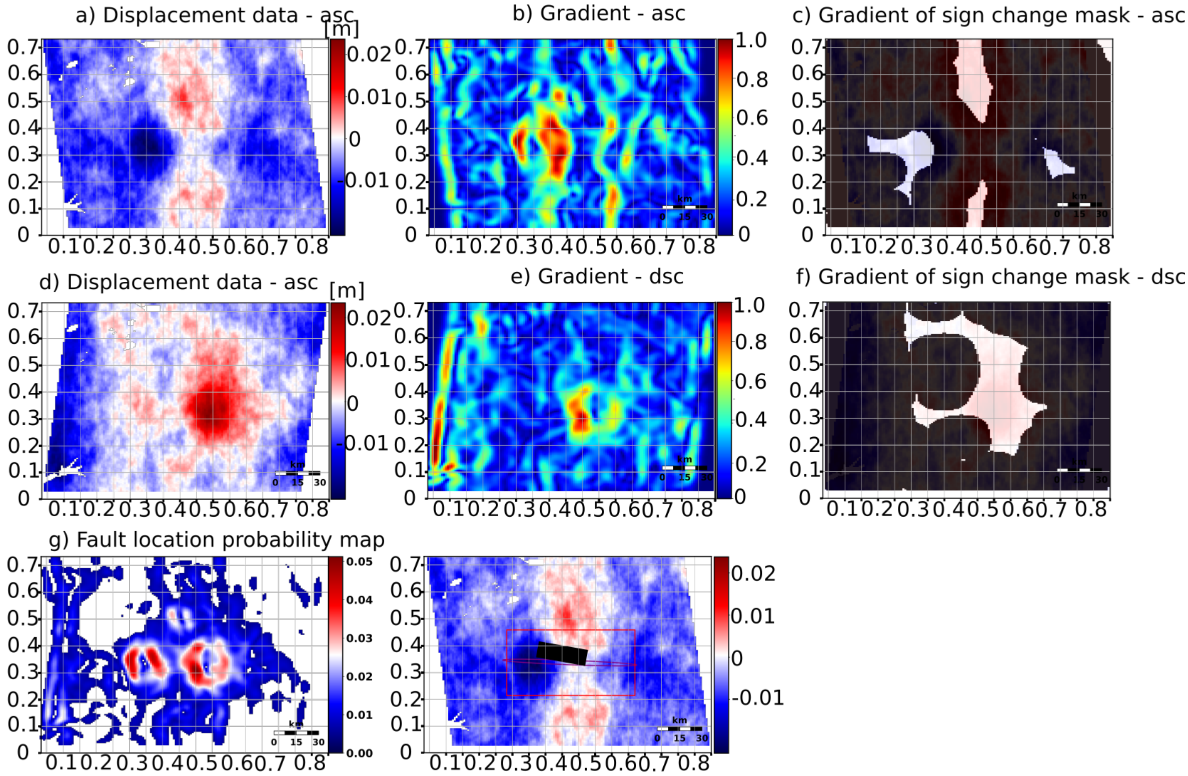

Figure S17: Synthetic test of the displacement map segmentation method for a single normal fault at a depth of 6 km. a), b) and c) contain the displacement data, the gradient and the gradient of the sign change mask (superimposed on the displacement data), respectively, for the ascending dataset and d), e) and f) accordingly for the descending data. g) shows the normalized combined product of the gradient sign change mask with the gradient from ascending and descending InSAR data. Values below 1% of the maximum value are masked out. This map is used as a pseudo-probability estimate for the position of the fault(s) location centroid. h) shows the bounding boxes and ellipses applied on the product shown in g). The green box is the area of interest, zoomed into in i). The red dashed line indicates the major axis of the ellipses containing the highest values for each region found as described above. The outline of the synthetic source(s) is indicated in the figures with black lines that are thicker for the top edge. The ellipses (indicated by the purple outline) is centered at the centroid of each region.

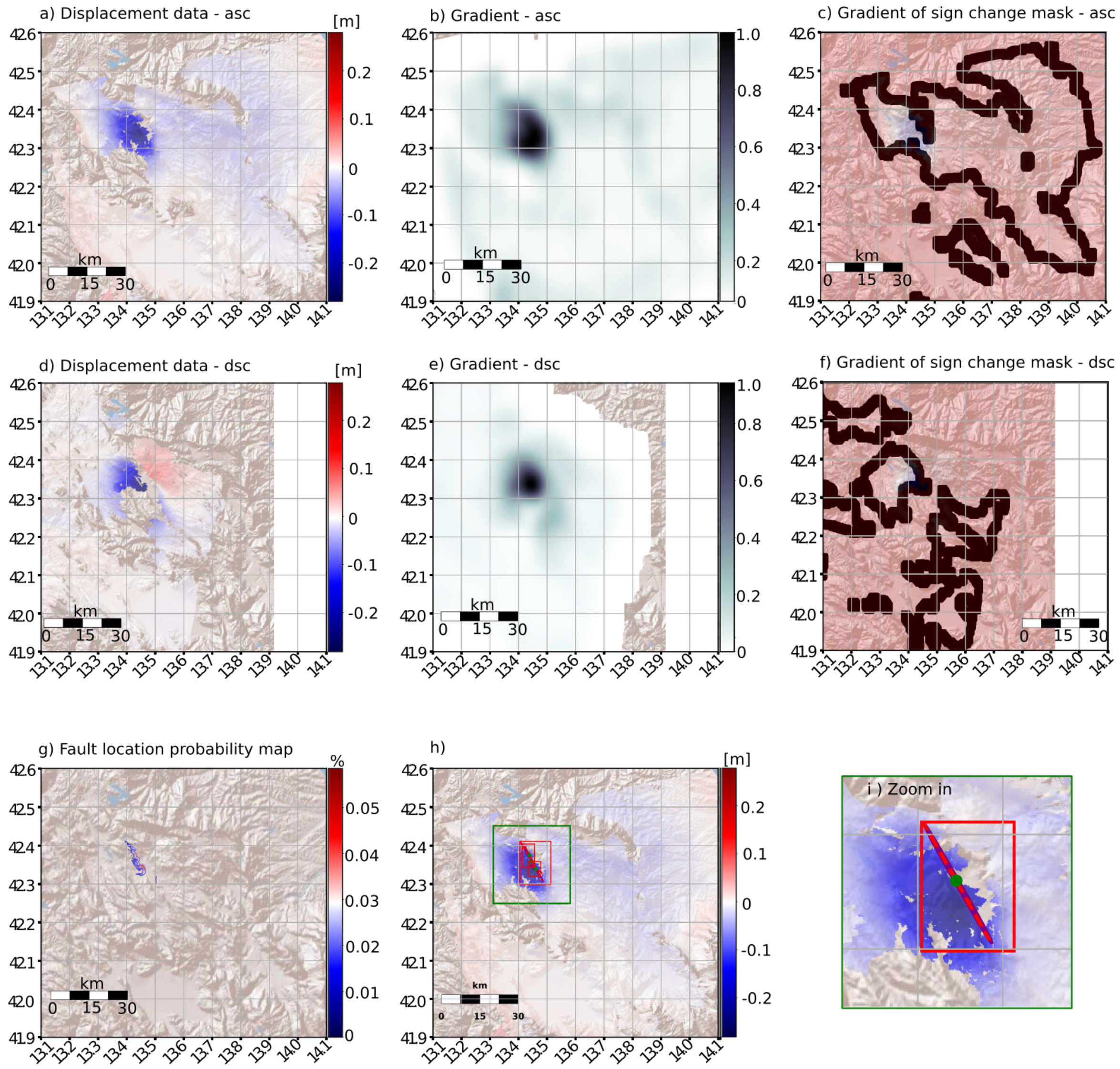

Figure S18: Source characteristics estimation from segmentation of InSAR displacement maps applied to the real InSAR Enivsat data of the 2009 L'Aquila earthquake (Steinberg et al., 2020). a), b) and c) contain the displacement data, the gradient and the gradient of the sign change mask (superimposed on the displacement data), respectively, for the ascending dataset and d), e) and f) accordingly for the descending data. g) shows the normalized combined product of the gradient sign change mask with the gradient from ascending and descending InSAR data. Values below 1% of the maximum value are masked out. This map is used as a pseudo-probability estimate for the position of the fault(s) location centroid. h) shows the bounding boxes and ellipses applied on the product shown in g). The green box is the area of interest, zoomed into in i). The red dashed line indicates the major axis of the ellipses containing the highest values for each region found as described above. The outline of the synthetic source(s) is indicated in the figures with black lines that are thicker for the top edge. The ellipses (indicated by the purple outline) is centered at the centroid of each region.

3.1. Additional information for the 2016 Muji earthquake

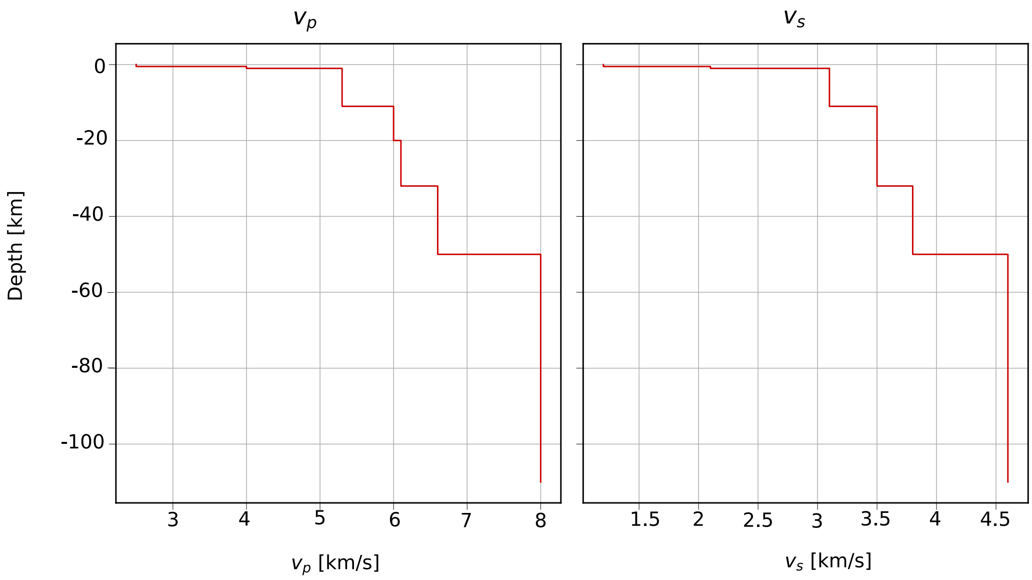

Figure S19: The layered 1-d velocity model for the static displacement modeling, based on (Xu et al., 2006) and (Li et al., 2018).

Table S1: Details of the Sentinel-1 SAR Data used in the study. Data are acquired in interferometric wide swath mode by Terrain Observation with Progressive Scans (TOPS) in VV polarization. The single look complex SAR images were downloaded from the Copernicus Open Access Hub.

rel. orbit (track)	primary date	secondary date	$\perp$ baseline [m]
107 (dsc)	2016/11/25	2016/12/19	78.2
27 (asc)	2016/10/20	2016/12/07	98.6

Table S2: Earthquake source model parameters for the 2016 Muji earthquake from published point and finite source models. Models from (Bie et al., 2018) for InSAR and seismology, and for seismology only from USGS and GCMT (Dziewonski et al., 1981) catalogs.

	Time	Lat	Lon	Depth	Strike	Dip	Rake	M_o	Length	Width	Slip
		°	°	km	°	°	°	$10^{18}\ N \cdot m$	km	km	m
Bie 1. source seis.	+7.88s	39.2313	74.1428	14	108/198	78/88	178/12	5.07			
Bie 2. source seis.	+10.52s	39.1681	74.4208	10.1	108/198	78/88	178/12	1.905			
Bie 1. source InSAR		39.2261	74.11165	8.5	106.4	70	-176	5.420			0.9
Bie 2. source InSAR		39.1754	74.3869	4.7	106.4	70	-176	2.847			1.31
USGS (body-wave)		39.273	73.978	17	19/288	86/86	4/176	7.5			
USGS (W-phase)		39.273	73.978	11.5	107/199	76/84	174/14	8.746			
USGS (Centroid)		39.273	73.978	16.7	113/18	63/81	-170/-28	10.5			
GCMT		39.27	74.14	19.1	110/19	78/87	-177/-12	11.3			
Feng InSAR	39.226	74.219	<15	105.5	80 (+/-4)	-161 (+/-12)	9.87	55	20		
He InSAR/GNSS		39.21	74.254	-	110.7 $\pm$ 0.5	83.7 $\pm$ 1.0	167 $\pm$ 1.0	12.03	38.4	18.3	0.56 $\pm$ 0.3

3.2. Additional Muji 2016 earthquake backprojection results

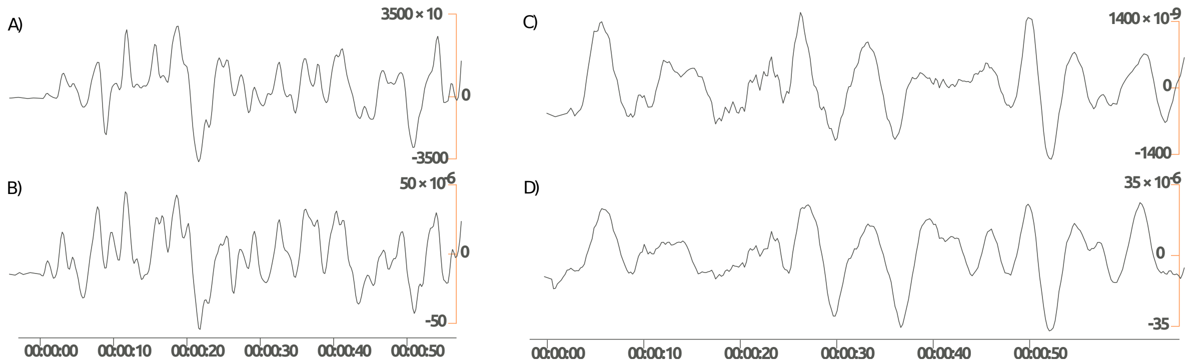

Figure S20: Linear stacking for both P- and SH-phases for the grid point closest to the hypocenter, compared to phase-weighted stacked waveforms. P-wave phases stacks for A) linear and B) phase-weighted methods. SH-wave phase stack for C) linear and D) phase-weighted methods. Note the difference in the scaling of the amplitude between diagrams.

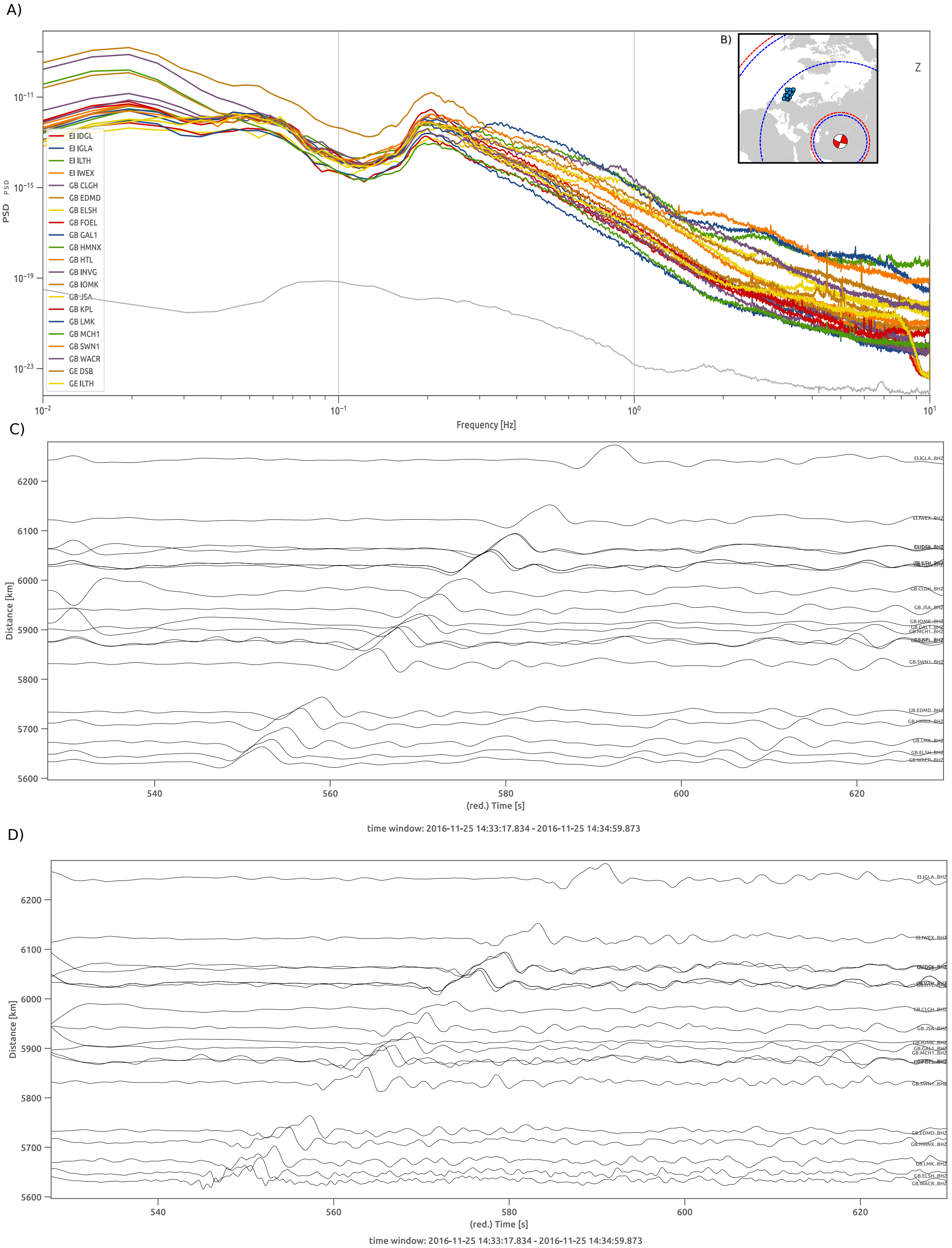

Figure S21: Example waveform data from an exemplary array used in the backprojection of the Muji 2016 earthquake (array number 16, located in central Europe). A) shows the array's waveform spectra of the z-component, color-coded for each station. The gray shaded spectrum shows the average noise spectrum from all stations immediately before the event. Inset B) shows the array location and stations. C) and D) show normalised waveforms with the P-wave onset for C) the low-frequency filtered data (0.003-0.16 Hz) and D) the high-frequency filtered data (0.16-1.5 Hz).

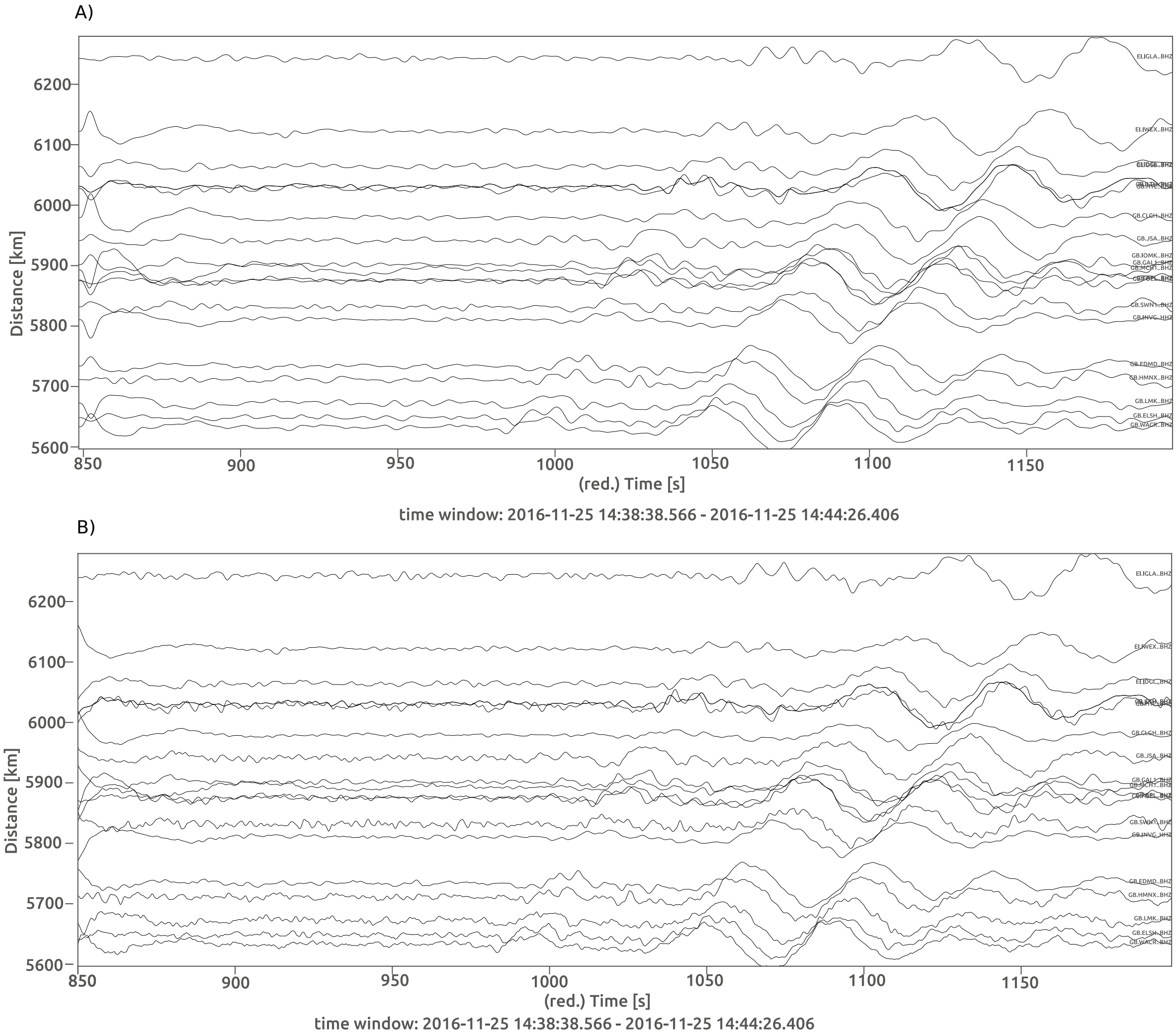

Figure S22: SH-wave onset waveforms from the exemplary array in central Europe (see also Fig. S21). A) the low-frequency filtered data (0.003-0.16 Hz) and B) the high-frequency filtered data (0.16-1.5 Hz).

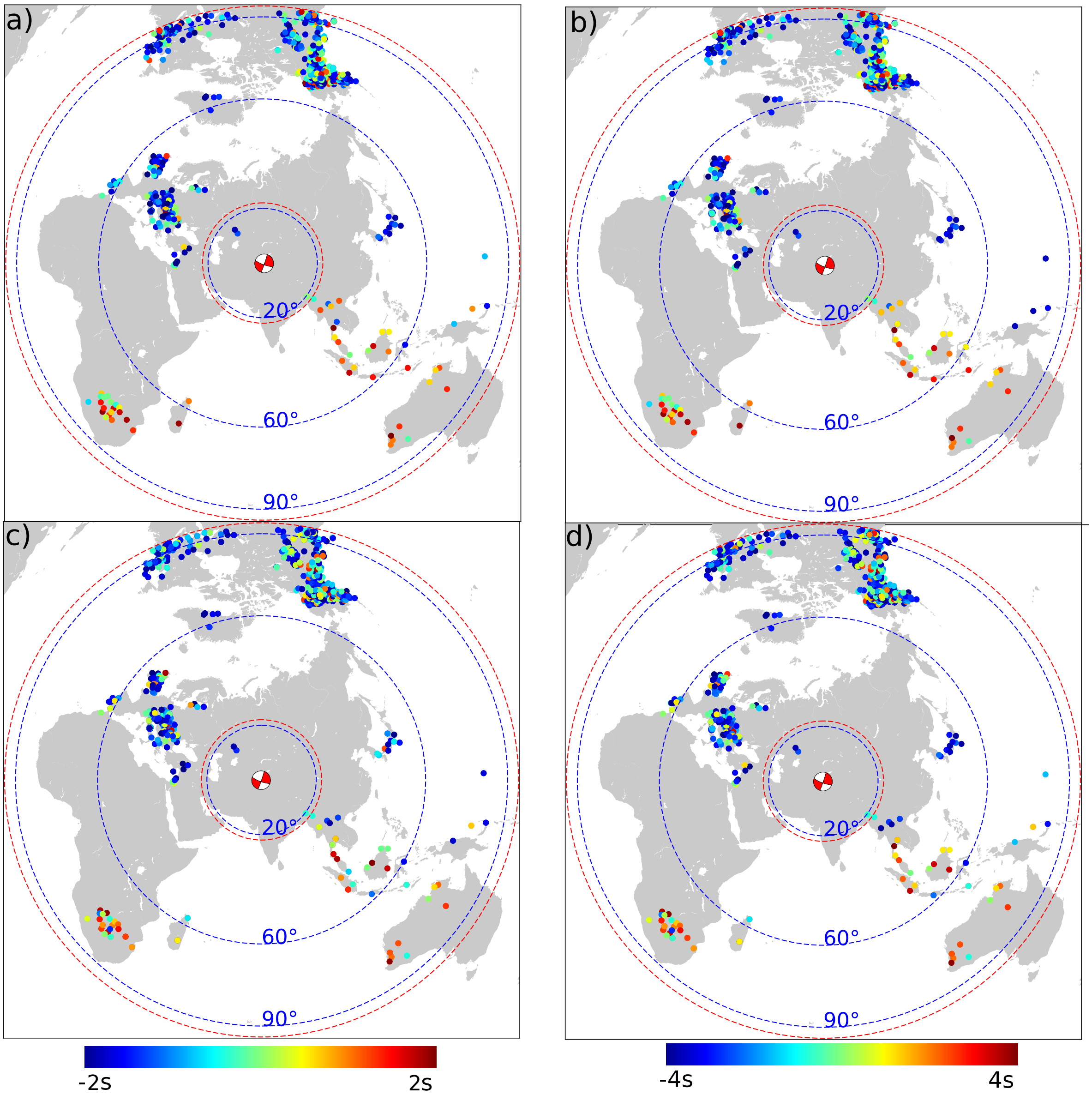

Figure S23: Empirical time shifts for different phases at the stations used in the backprojection that maximize the semblance of the reference event and are used for the BP of the 2016 Muji earthquake. Shown timeshifts for the low-frequency backprojection are in a) of the P-phase and in b) for the SH-phase, while the high-frequency backprojections are shown in c) for the P-phase and in d) for the SH-phase. Timeshifts are given relative to the gCMT onset time.

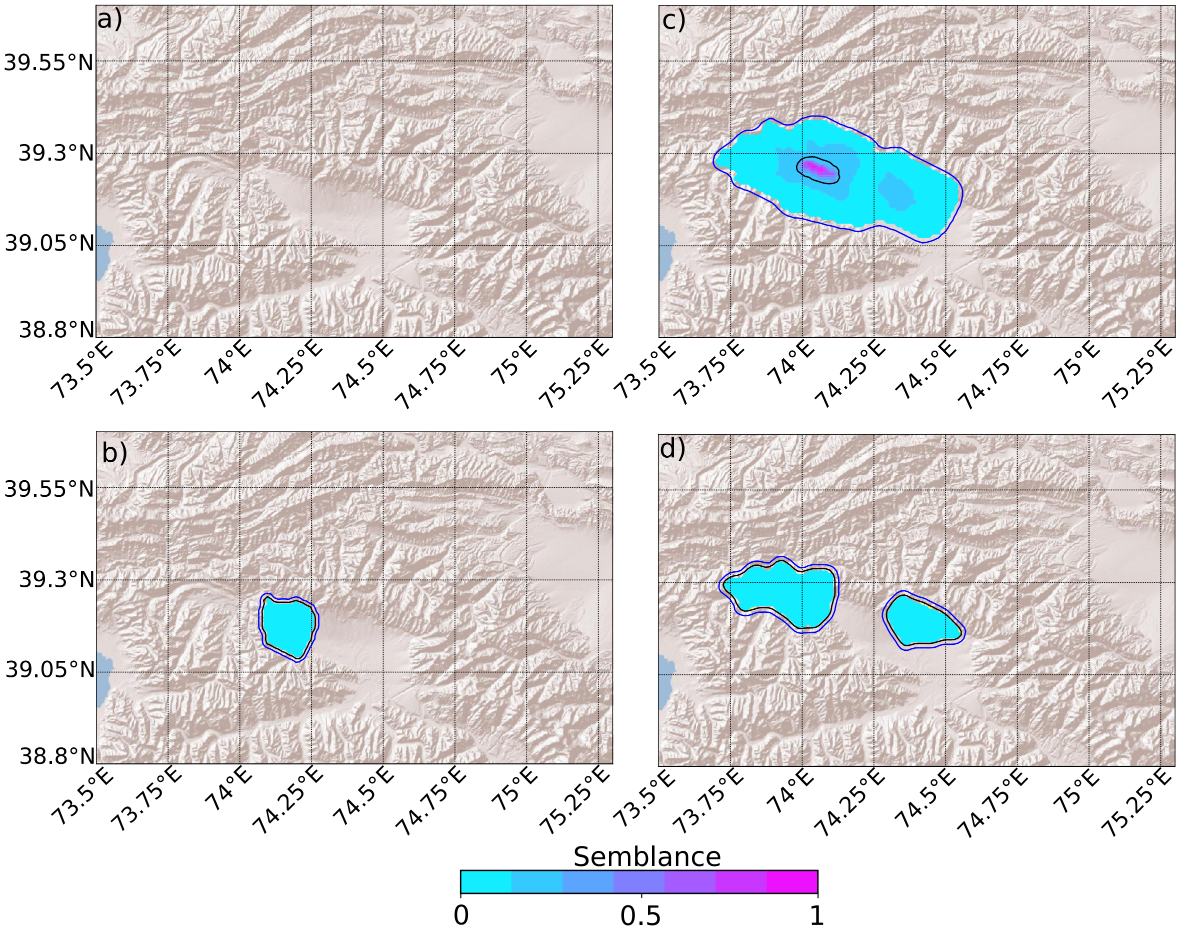

Figure S24: Time-incremental low-frequency semblance maps from the backprojection of the 2016 Muji earthquake for every timestep of 8 s individually in a) to d). The time given is relative to the onset of the first occurrence of significant semblance.

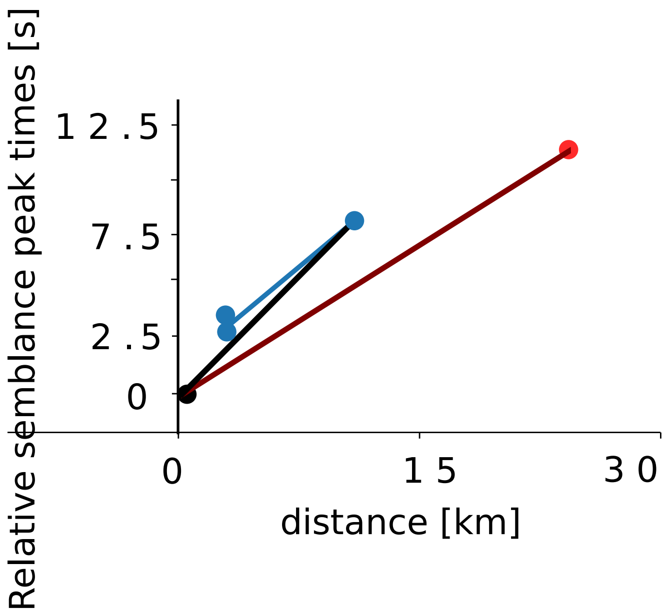

Figure S25: Rupture velocity estimate for the 2016 Muji earthquake from the high-frequency BP at the western segment (red line) and at the eastern segment (black line), measured from the nucleation point to the last respective semblance mapping. Time is relative to the first window with semblance. The blue line shows the rupture velocity estimate for the eastern segment, from its rupture start of the eastern segment only to the respective end of rupture on each segment, indicated by the two blue dots at the beginning.

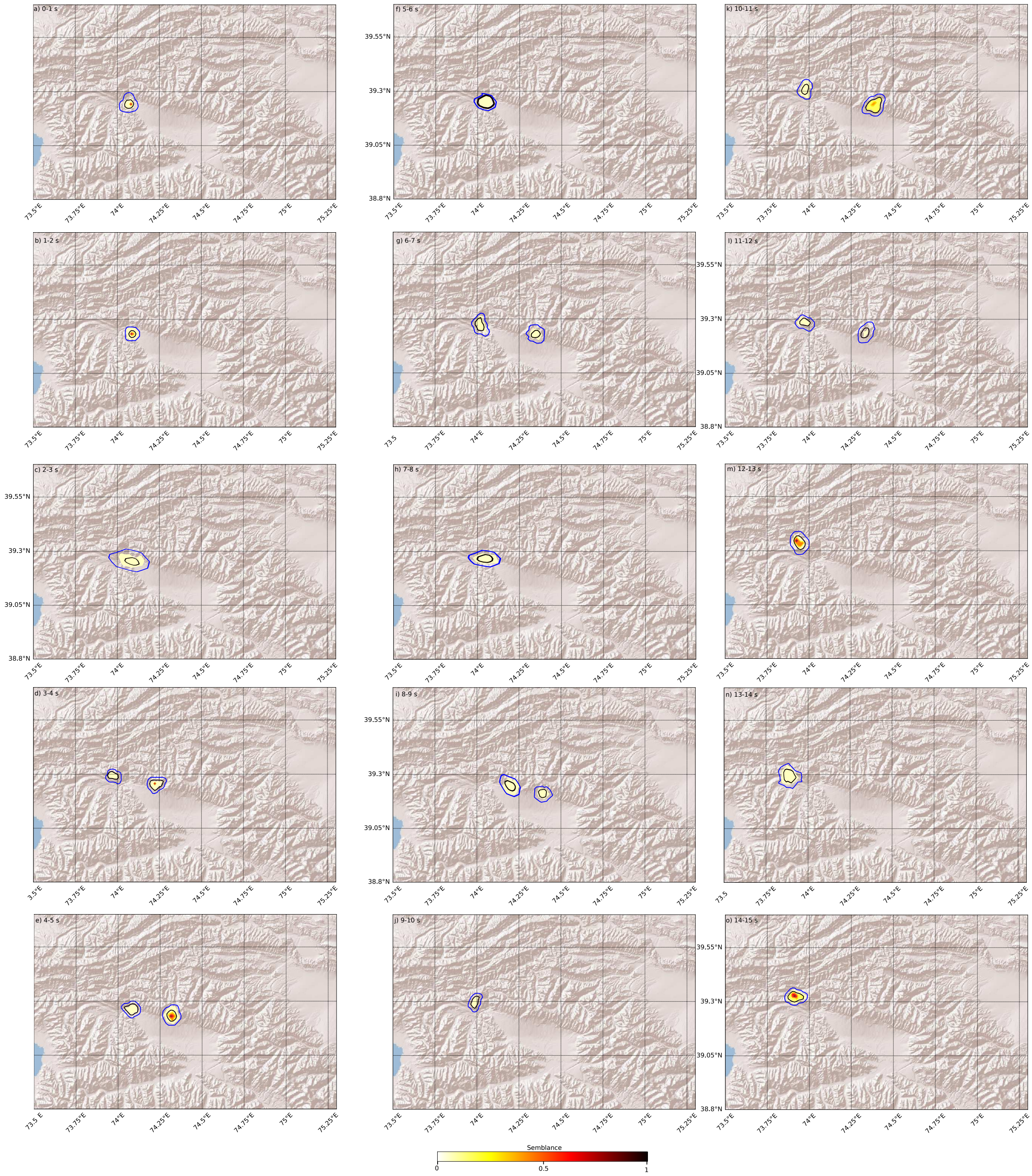

Figure S26: Time-incremental high-frequency semblance maps for the backprojection of the 2016 Muji earthquake for every timestep of 1 s from 0 s in a) to 15 s in o). The time given is relative to the onset of the first occurrence of significant semblance.

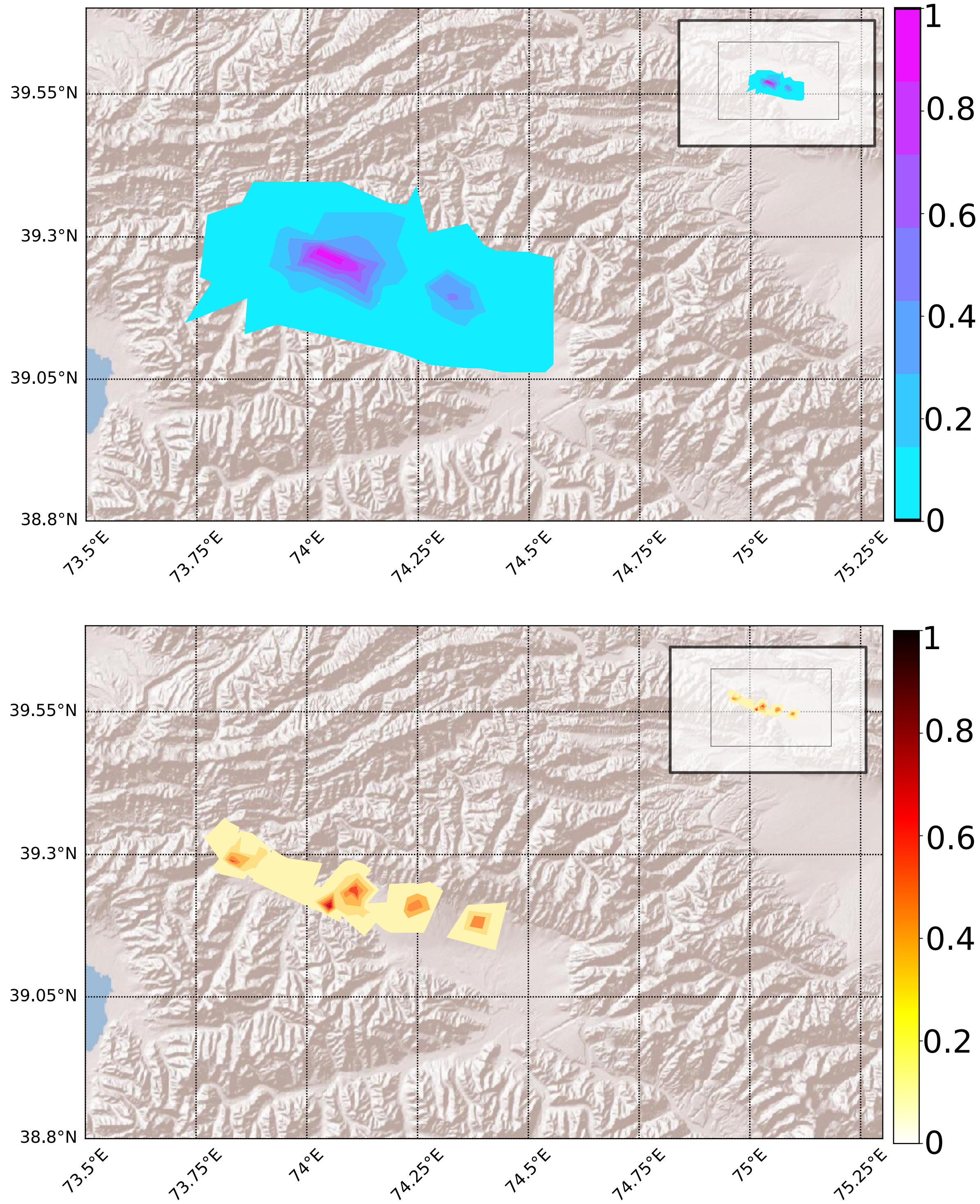

Figure S27: Cumulative low-frequency P-phase semblance map, from all timesteps of the backprojection of the 2016 Muji earthquake.

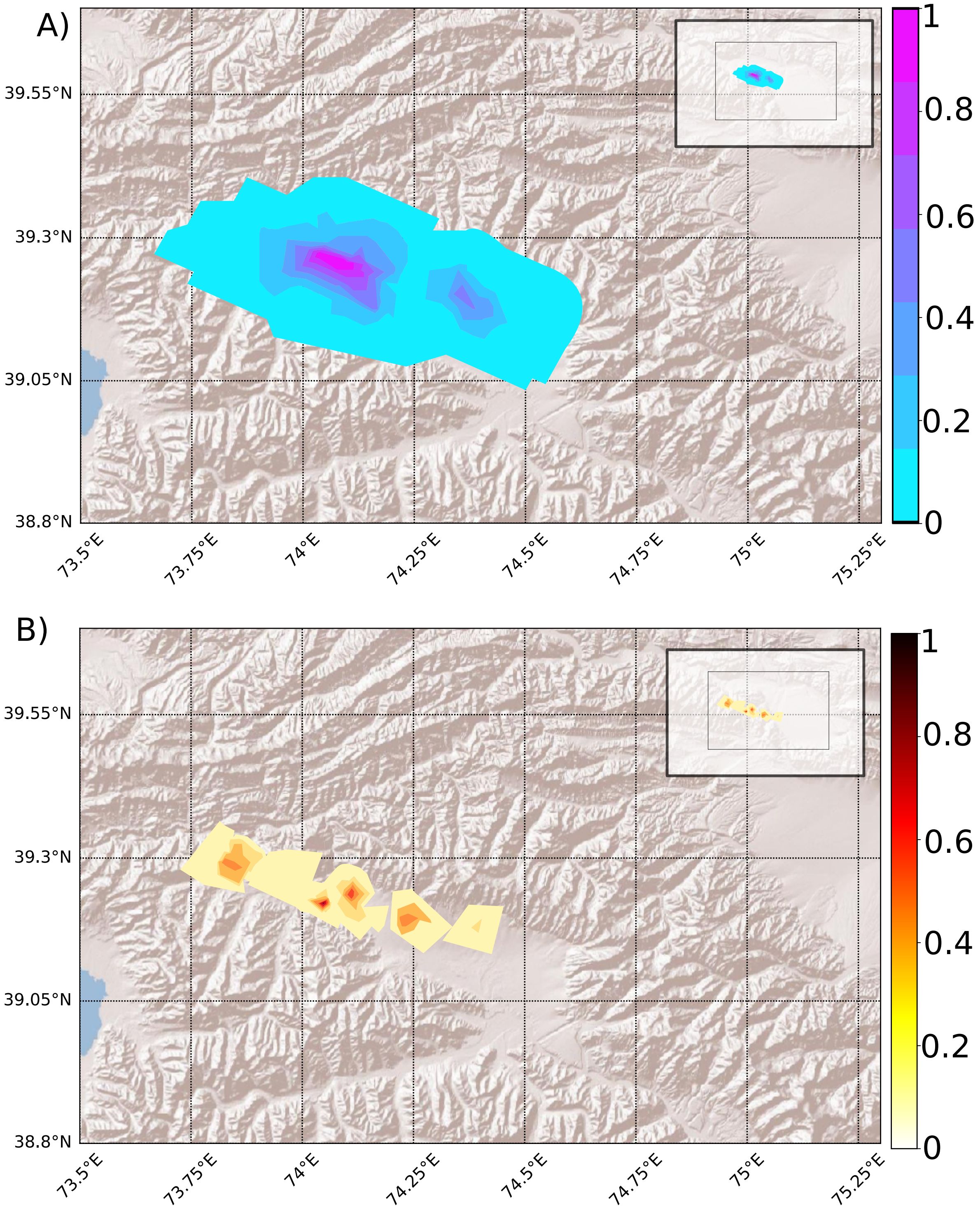

Figure S28: Cumulative low-frequency SH-phase semblance map, from all timesteps of the backprojection of 2016 Muji earthquake.

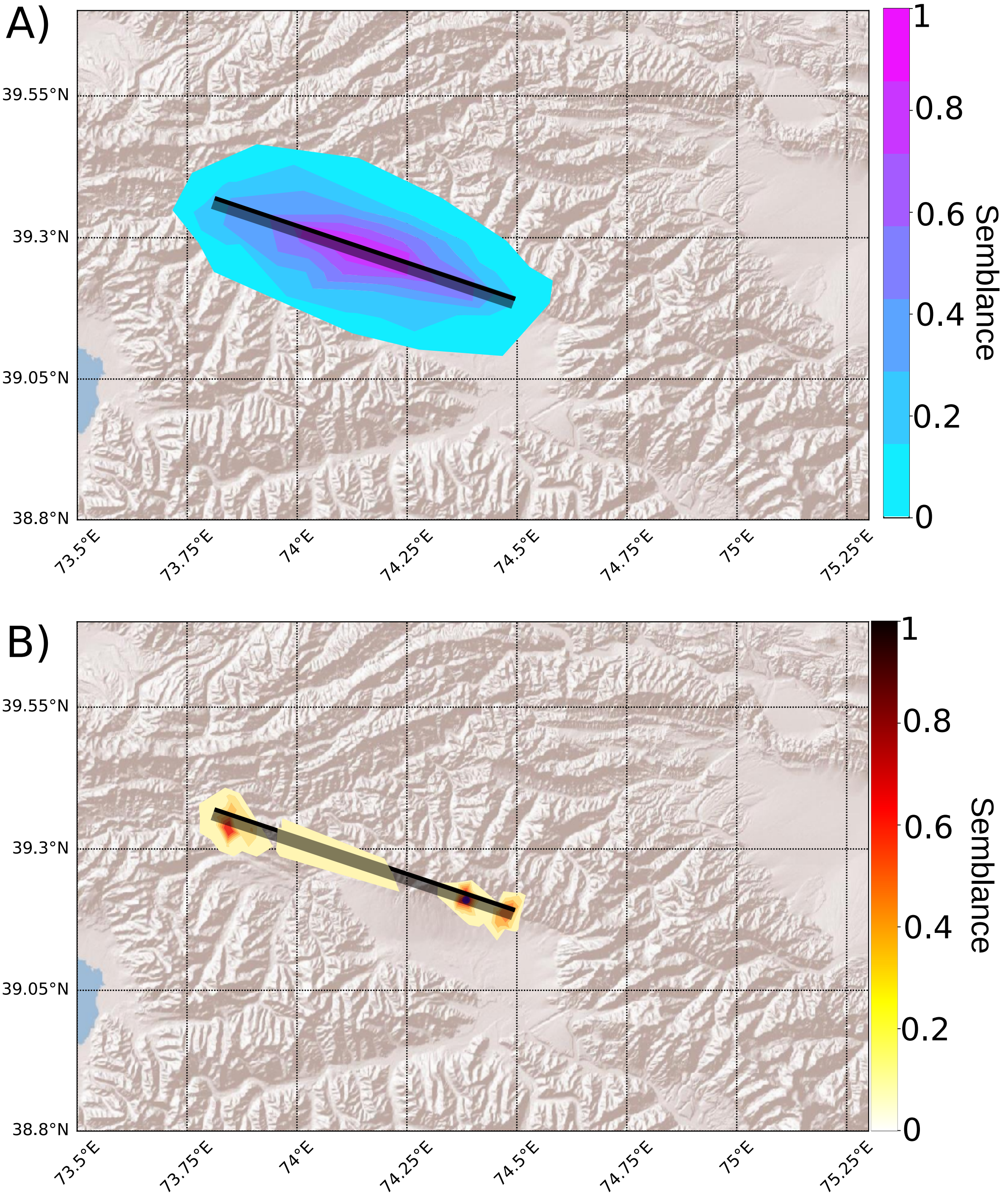

Figure S29: Cumulative semblance maps from a synthetic backprojection of a single-segment kinematic source model representing the 2016 Muji earthquake for a) high-frequency and b) low-frequency waveforms. The thick black line indicates the upper edge of the fault and the gray-shaded area the fault projection to the surface. The blue dot indicates the rupture nucleation point.

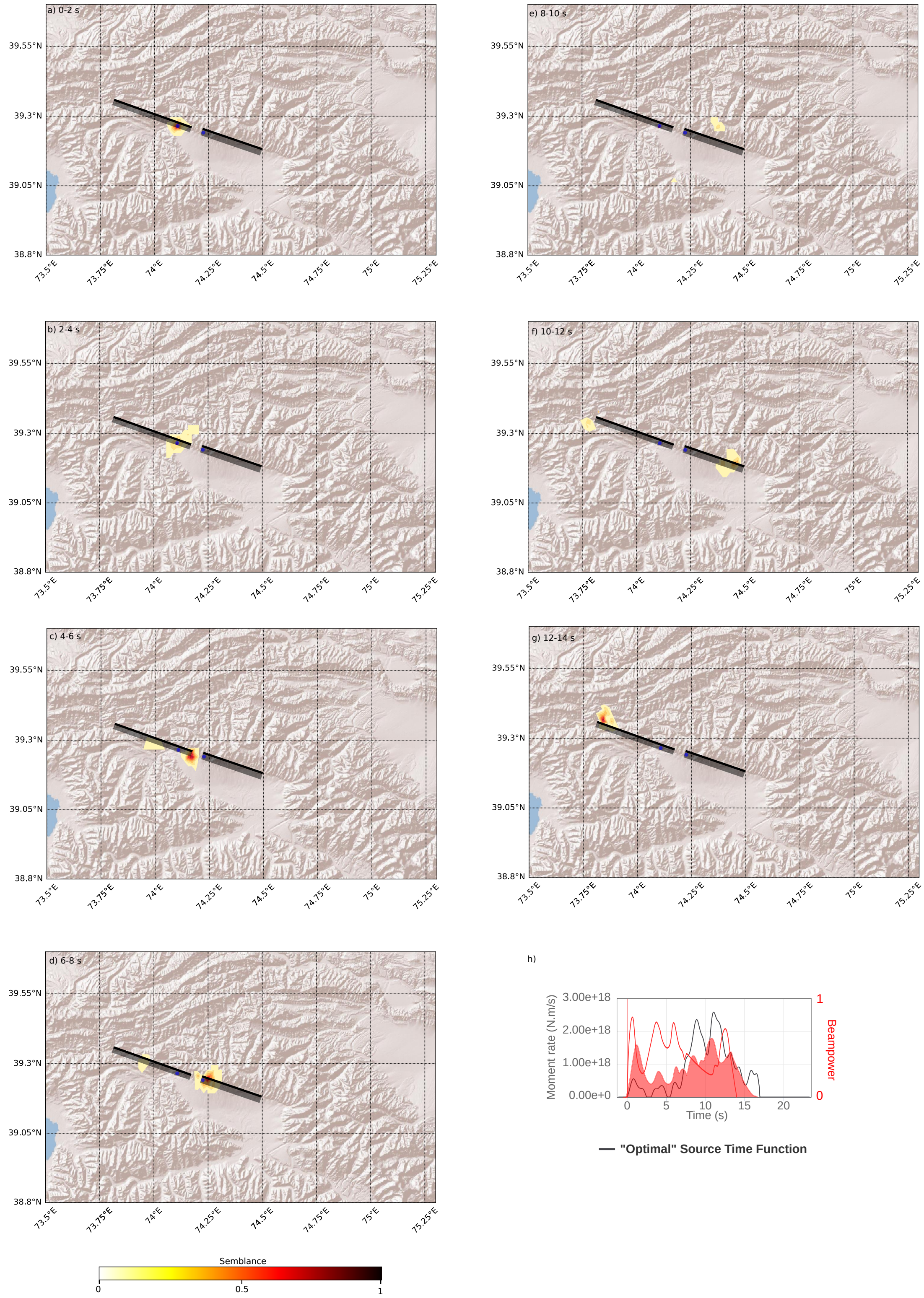

Figure S30: Time-incremental high-frequency semblance mappings for all timesteps in a) to d) from a synthetic backprojection of a single-segment kinematic source model representing the 2016 Muji earthquake. The thick black line indicates the upper edge of the fault and the gray-shaded area the fault projection to the surface. The blue dot indicates the nucleation point. h) Beampower of the high-frequency BP as a function over time as a red and filled function of time together with the optimal (black line) source time functions from the SCARDEC catalog (Vallée & Douet, 2016). Additionally shown is the beampower from using the single large array aperture backprojection as a red line.

3.3. Additional optimization results for the 2016 Muji

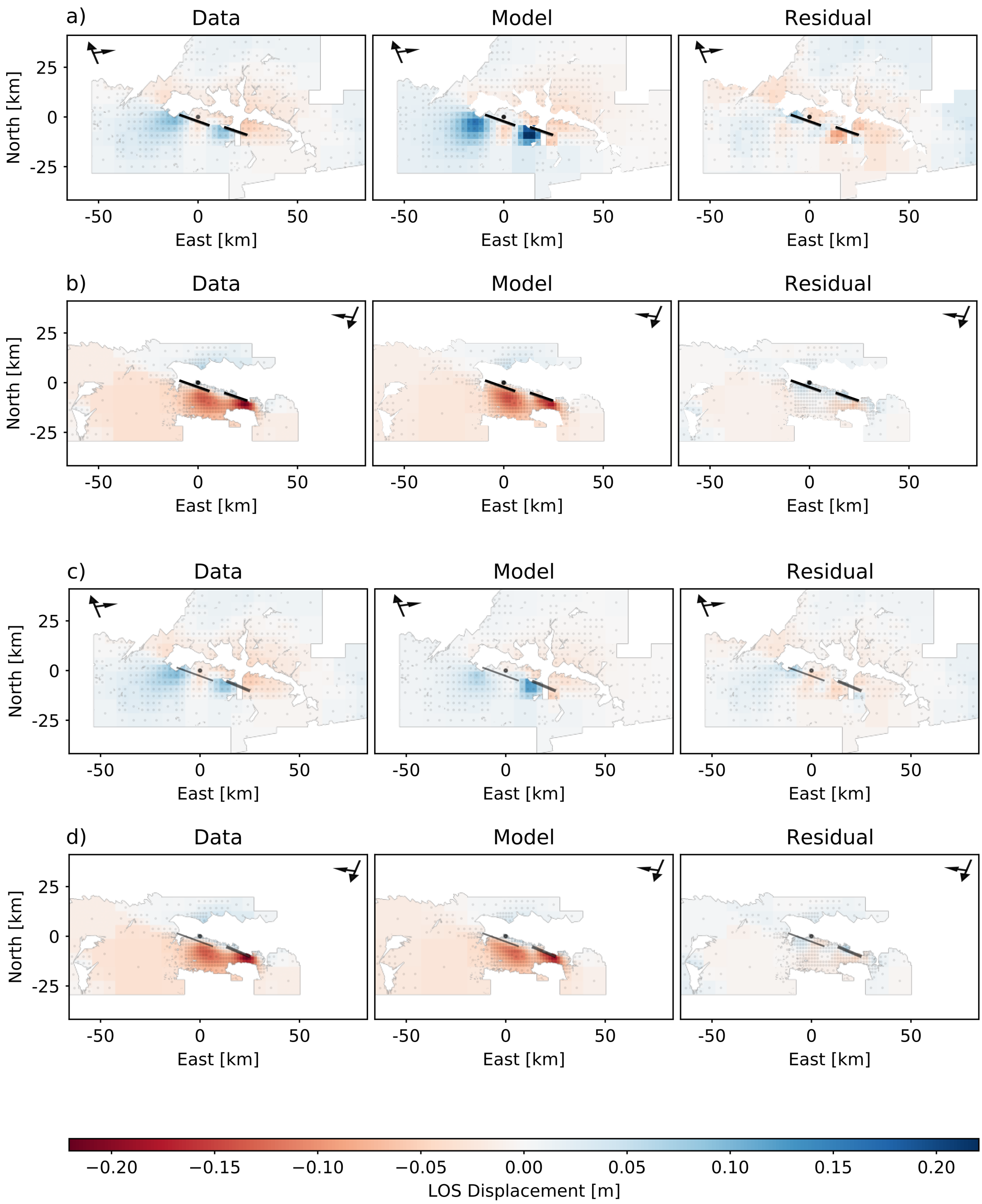

Figure S31: Data, model and residual for the InSAR line-of-sight displacements for the best-performing model from the exploratory optimization for a) ascending data and b) descending data as well as from the guided optimization for c) ascending data and d) descending data.

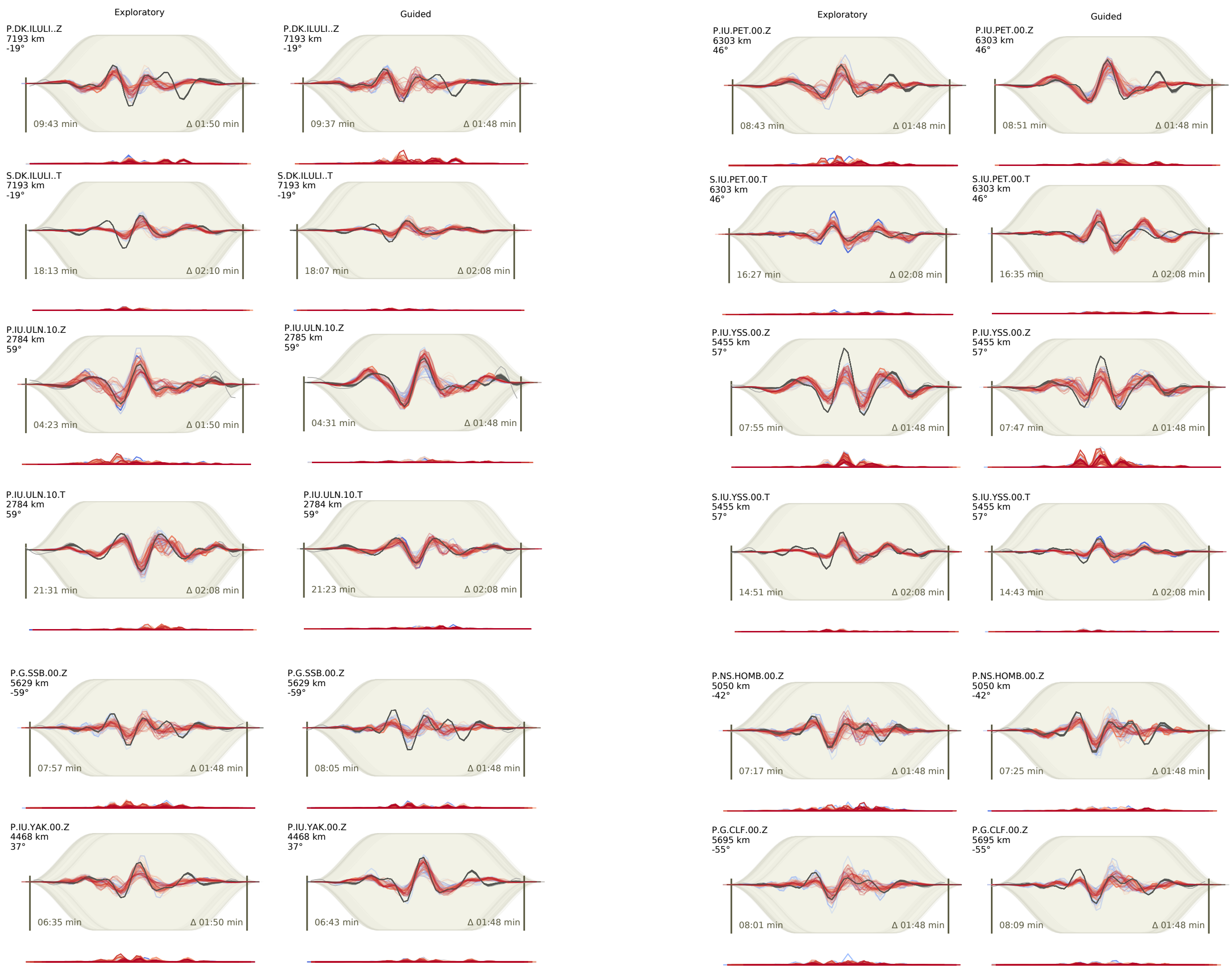

Figure S32: Waveform fits for the ensemble of the exploratory and guided optimizations side-by-side for selected stations. Left rows show the exploratory and right rows the guided optimization fits. Z-components and for some stations also the T-components are shown. In each subplot the black lines show the original waveforms data, and colored waveforms show the modelled synthetic waveforms with blue to red showing decreasing misfits (with blue poor and red good misfit. The light yellow shading shows the applied waveform taper. At the bottom of each panel the absolute waveform misfit with time is plotted in red.

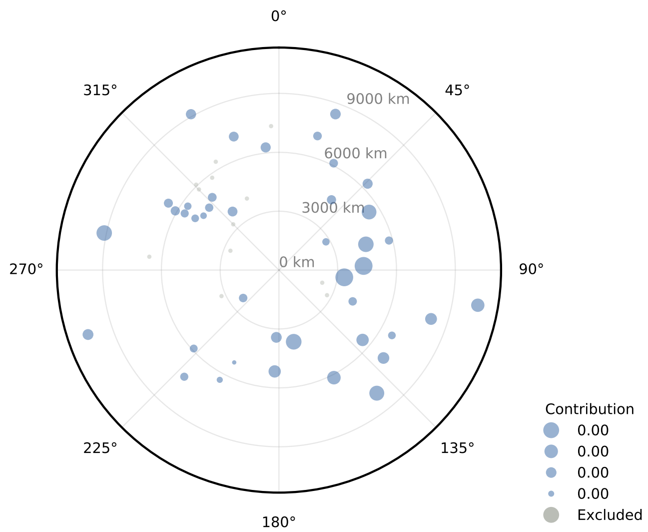

Figure S33: Station map indicating trace weights in the non-linear optimization for the Z-component (P-phase).

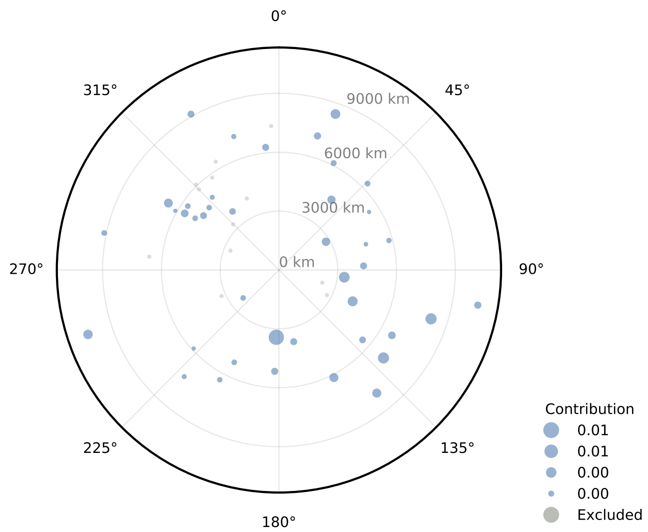

Figure S34: Station map indicating trace weights in the non-linear optimization for the T-component (SH-phase).

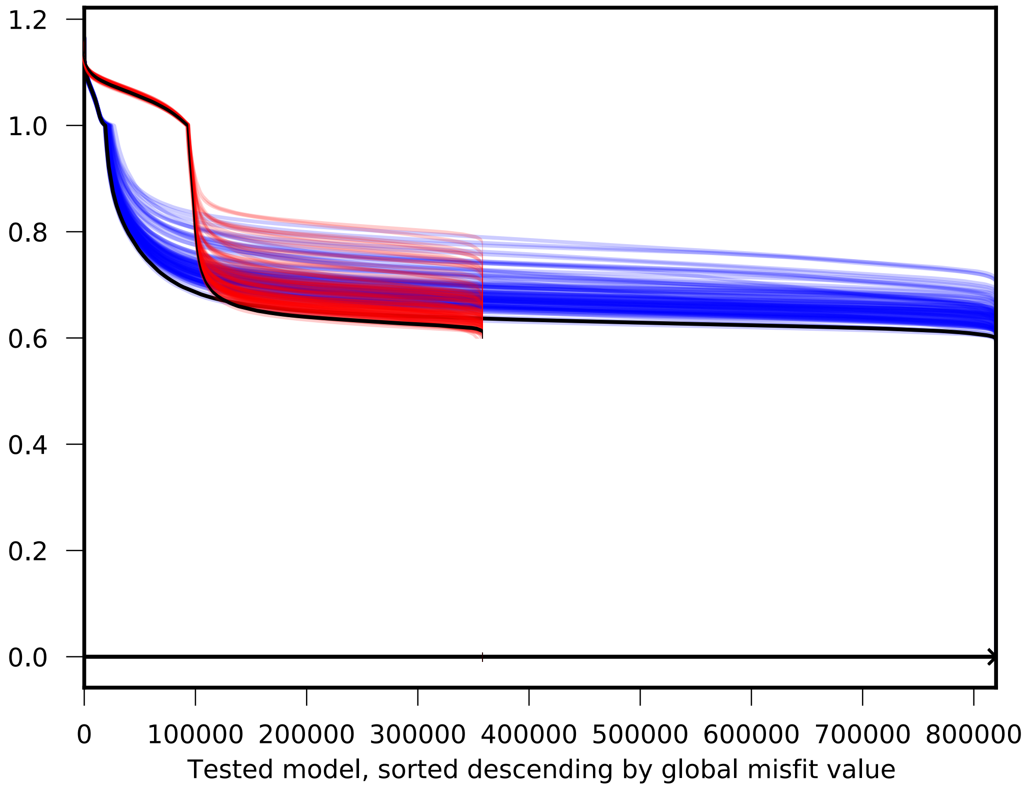

Figure S35: Bootstrap chain misfits (ensemble) as a function of the sample number for the guided optimization (red) compared to the exploratory optimization (blue).

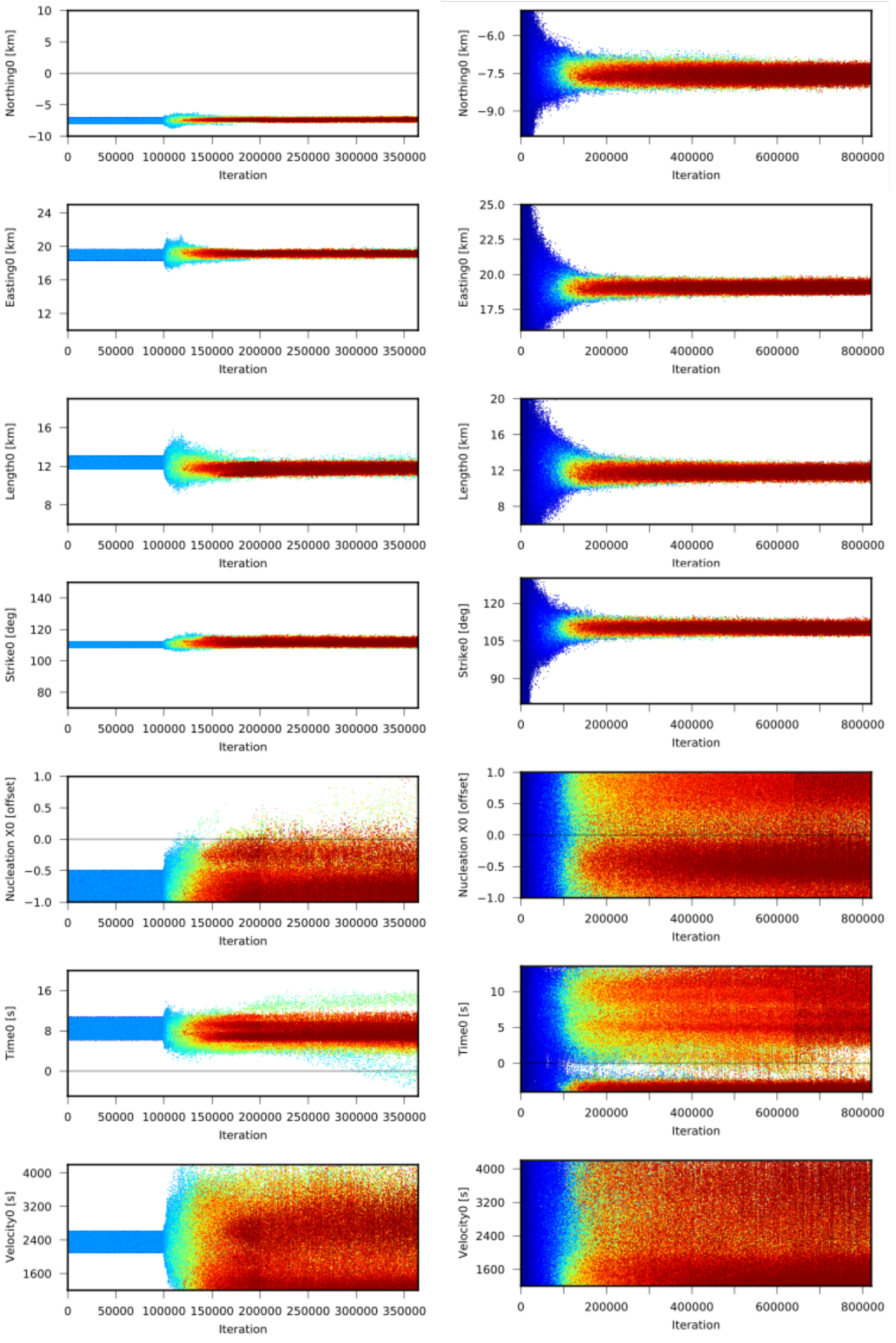

Figure S36: Sampled parameter values for the eastern source segment as a function of sample number, color-coded according to misfit, with warmer colors showing lower misfits. Shown are the source parameters sampled for the eastern source segment from the guided (right column) and exploratory optimizations (left column) in comparison. Shown are only source parameters with different priors in the two optimizations.

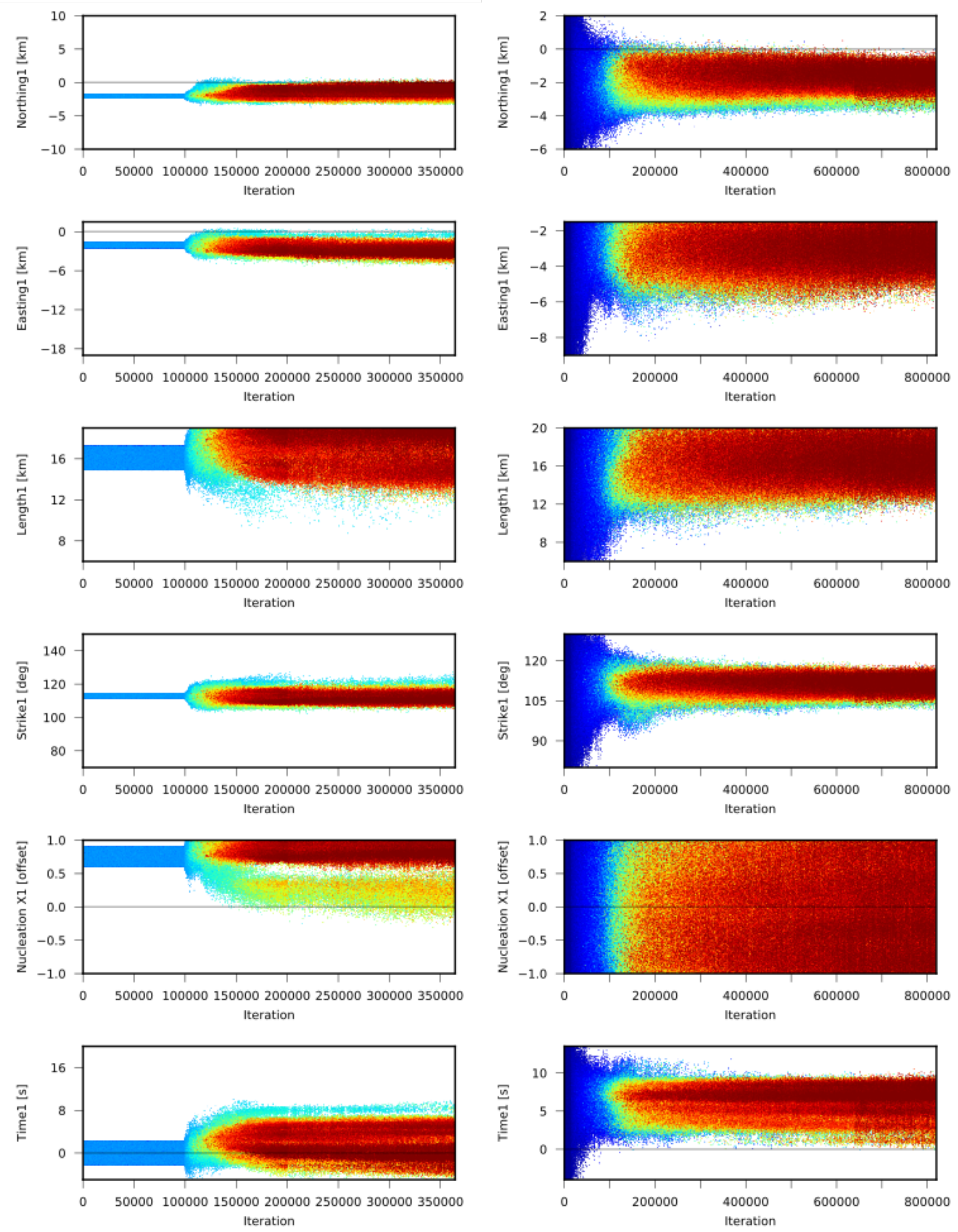

Figure S37: Sampled parameter values for the western source segment as a function of sample number, color-coded according to misfit, with warmer colors showing lower misfits. Shown are the source parameters sampled for the western source segment from the guided (right column) and exploratory optimizations (left column) in comparison. Shown are only source parameters with different priors in the two optimizations.

References

Bardsley, J. M., Solonen, A., Haario, H., & Laine, M. (2014). Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Scientific Computing*, *36*(4), A1895–A1910.

Bie, L., Hicks, S., Garth, T., Gonzalez, P., & Rietbrock, A. (2018). 'two go together': Near-simultaneous moment release of two asperities during the 2016 mw 6.6 muji, china earthquake. *Earth and Planetary Science Letters*, *491*, 34–42.

Dziewonski, A., Chou, T.-A., & Woodhouse, J. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research: Solid Earth*, *86*(B4), 2825–2852.

Jonsson, S., Metzger, S., Sudhaus, H., Dutta, R., & Xu, W. (2014). Bayesian Estimation and Deterministic Optimization of Earthquake and Interseismic Model Parameters from InSAR and GPS Data. In *AGU Fall Meeting Abstracts.*

Li, W., Chen, Y., Yuan, X., Schurr, B., Mechie, J., Oimahmadov, I., & Fu, B. (2018). Continental lithospheric subduction and intermediate-depth seismicity: Constraints from s-wave velocity structures in the pamir and hindu kush. *Earth and Planetary Science Letters*, *482*, 478–489.

Rubin, D. B. (1981). The bayesian bootstrap. *The annals of statistics*, 130–134.

Steinberg, A., Sudhaus, H., Heimann, S., & Krüger, F. (2020). Sensitivity of insar and teleseismic observations to earthquake rupture segmentation. *Geophysical Journal International*.

Sudhaus, H., & Jónsson, S. (2009, feb). Improved source modelling through combined use of InSAR and GPS under consideration of correlated data errors: application to the June 2000 Kleifarvatn earthquake, Iceland. *Geophysical Journal International*, *176*(2), 389–404. Retrieved from `https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.2008.03989.x` doi: 10.1111/j.1365-246X.2008.03989.x

Vallée, M., & Douet, V. (2016). A new database of source time functions (STFs) extracted from the SCARDEC method. *Physics of the Earth and Planetary Interiors*, *257*, 149–157.

Wang, R. (1999). A Simple Orthonormalization Method for Stable and Efficient Computation of Green's Functions. *Bulletin of the Seismological Society of America*, *89*(3), 733–741.

Xu, Y., LIU, J.-H., LIU, F.-T., ZHU, L.-R., LONG, H.-Y., & Wei, B. (2006). Crustal velocity structure and seismic activity in the tianshan-pamir conjunctive zone. *Chinese Journal of Geophysics*, *49*(2), 417–425.