Check for updates

ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL102983

Weighing Geophysical Data With Trans-Dimensional Algorithms: An Earthquake Location Case Study

Nicola Piana Agostinetti[1,2], Alberto Malinverno[3], Thomas Bodin[4], Christina Dahner[5], Savka Dineva[5,6], and Eduard Kissling[7]

[1]Department of Earth and Environmental Sciences, Universitadi Milano Bicocca, Milano, Italy, [2]Department of Geology, University of Vienna, Vienna, Austria, [3]Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA, [4]University of Lyon, University of Lyon 1, ENSL, CNRS, LGL-TPE, Villeurbanne, France, [5]Luossavaara-Kiirunavaara AB, Kiruna, Sweden, [6]Department of Civil, Environmental and Natural Resources Engineering, Lulea, Sweden, [7]Department of Earth Sciences, ETH Zurich, Zürich, Switzerland

Key Points:

- We develop a novel approach for automatic weighting of data in geophysical inverse problems, based on a trans-dimensional algorithm
- We apply the novel approach to seismic event location in mines, obtaining consistent results compared to a more standard method
- Our approach outperforms standard seismic monitoring approaches, when limited information are available on local seismic structure

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

N. Piana Agostinetti,
nicola.pianaagostinetti@unimib.it

Citation:

Piana Agostinetti, N., Malinverno, A., Bodin, T., Dahner, C., Dineva, S., & Kissling, E. (2023). Weighing geophysical data with trans-dimensional algorithms: An earthquake location case study. *Geophysical Research Letters*, *50*, e2023GL102983. https://doi.org/10.1029/2023GL102983

Received 22 JAN 2023
Accepted 19 AUG 2023

Author Contributions:

Conceptualization: Nicola Piana Agostinetti, Alberto Malinverno, Thomas Bodin
Data curation: Christina Dahner, Savka Dineva
Methodology: Nicola Piana Agostinetti
Software: Nicola Piana Agostinetti
Supervision: Alberto Malinverno, Eduard Kissling
Validation: Christina Dahner, Savka Dineva, Eduard Kissling
Writing – original draft: Nicola Piana Agostinetti, Alberto Malinverno

© 2023. The Authors.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract In geophysical inverse problems, the distribution of physical properties in an Earth model is inferred from a set of measured data. A necessary step is to select data that are best suited to the problem at hand. This step is performed ahead of solving the inverse problem, generally on the basis of expert knowledge. However, expert-opinion can introduce bias based on pre-conceptions. Here we apply a trans-dimensional algorithm to automatically weigh data on the basis of how consistent they are with the fundamental hypotheses made to solve the inverse problem. We demonstrate this approach by inverting arrival times for the location of a seismic source in an elastic half-space, assuming a point-source and uniform weights in concentric shells. The key advantage is that the data do no longer need to be selected by an expert, but they are assigned varying weights during the inversion procedure.

Plain Language Summary In the Big data era, automated approaches to data evaluation are needed for two main reasons: to be able to process a large amount of data in a limited time, and to avoid bias introduced by data analysists. In this study we present a novel approach to data analysis, where the data themselves measure their consistency with our hypotheses. The approach is applied to earthquake location in mines, where millions of seismic events occur every year, and automatic processing of seismic data is mandatory. We demonstrate that our approach outperforms standard ones when almost nothing is known about the data and their measurement errors.

1. Introduction

Measured scientific data make possible a quantitative analysis of observations (e.g., a seismometer can record seismic waves, which are only felt by humans as transient phenomena). Scientific data are routinely processed before making inferences on the spatio-temporal distribution of physical quantities and/or physical processes (e.g., arrival times for seismic P-waves are extracted from continuous seismic recordings to infer the position of a seismic source). Processing steps can be necessary to remove spurious data (e.g., arrival times from seismic sensors that are not synchronized), but also to enhance data to better represent the most relevant signal for the problem being investigated (e.g., seismic waveforms may be filtered in the frequency domain before picking relative arrival times by cross-correlation (VanDecar & Crosson, 1990), for a clear identification of phases and for removing noise-site-effect interferences with targeted signal wavelet).

Geo-scientific data are especially challenging, because they are generally used to make inferences on physical quantities which are not directly measurable, but need to be estimated by solving an inverse problem (Tarantola, 2005), where processed measurements (e.g., P-wave arrival times or maximum wavelet amplitudes) are combined with hypotheses about the physics of the system (e.g., models of seismic wave propagation in the rock volume or seismic energy released by source). In this case, data processing typically includes selecting a subset of the data that is most relevant for the problem at hand (e.g., by removing arrival times for P-waves that do not travel directly from source to receiver). Additionally, seemingly less accurate data are often excluded or apriori downweighted to make them less influential in the final solution (e.g., arrival times recorded at distant seismic sensors that are likely to show larger effect of influence by attenuation or scattering along the ray-path). These data processing steps are usually based on expert opinion, but expert decisions made a priori before solving the inverse problem can be somewhat arbitrary and bias the inversion results.

Writing – review & editing: Thomas Bodin, Eduard Kissling

Here we propose a novel approach to incorporate the choice of weights for the data in the inversion process (or, more precisely, the variance of data noise). Our approach is based on trans-dimensional Markov chain Monte Carlo (McMC) sampling (Piana Agostinetti et al., 2021; Piana Agostinetti & Sgattoni, 2021) and works by proposing and accepting/rejecting data weighing schemes following the Metropolis algorithm (Sambridge & Mosegaard, 2002) where the data weighing schemes have a variable number of parameters (Malinverno, 2002; Sambridge et al., 2006). The complexity of the weighing scheme is dictated by the data themselves, rather than by user-defined choices made during pre-processing. The assigned weights depend on how closely different data match the fundamental assumptions made in solving the inverse problem.

The novel approach is developed in a fully consistent Bayesian framework, which means that Posterior Probability Distributions (PPD) for the weights are retrieved at the end of the McMC sampling. From a Bayesian point of view, what we call "weights" are data variances (i.e., data uncertainties). In Hierarchical Bayes, estimated data uncertainties is known to incorporate both measurement errors, modeling errors and other sources of uncertainty (e.g., Bodin et al., 2012). Along these lines, our approach can be considered a step further in the context of the algorithms for error covariance estimation (Bodin et al., 2012; Dettmer & Dosso, 2012; Galetti et al., 2016; Kolb & Lekić, 2014; Mustać & Tkalčić, 2015; Piana Agostinetti et al., 2021; Piana Agostinetti & Sgattoni, 2021; Steininger et al., 2013; Xiang et al., 2018). In particular, Ghalenoei et al. (2022) developed an approach, where two separate trans-D samplings are applied to both error and geophysical models.

We test our approach in the geophysical inverse problem of locating a seismic point source using P- and S-wave arrival times recorded by sensors in a seismic network. In this inverse problem, data are generally downweighted with distance of the sensor from the seismic source or are removed in a pre-processing step if the sensors are farther than a chosen distance from the source. In our novel approach, we define a set of spherical shells centered on the source (Figure 1a). All sensors within a shell are assigned the same weight (Figure 1b), but more complex weight assignments can be made (e.g., weights that vary linearly with distance from the source within each shell; see Figure 1c). The number of shells, their radii and weights are unknown, and will be defined by the McMC sampling. The stations that receive the largest weights will be those that measure arrival times consistent with the fundamental assumptions made in the inverse problem (namely, a point-wise seismic source and constant P- and S-wave velocities in the rock volume).

Our natural laboratory is Kiirunavaara mine (Sweden), a 6 km-long active mine with more than 200 seismic sensors in a 3D configuration that spans along the exploited rock volume (Dineva et al., 2022). Given such an extensive seismic network, events can be well located in three dimensions. We selected two seismic events. The first is a man-made blast, used to calibrate the seismic network (Figure 1d). The actual location of this seismic source is known within <1 m and can be immediately used to evaluate our results. The second is a natural M_w 4.2 multi-phase seismic event that occurred on 18 May 2020 (Dineva et al., 2022) and it was recorded on all working sensors in the mine (Figure 1e). Our experiment is structured as follows. We first compute a reference solution for the calibration blast by applying a standard McMC algorithm (Riva et al., 2023). In this reference solution, we do not use our novel approach, but we solve for the source location by removing data from sensors at a range of distances from a preliminary location of the seismic source, as done in standard seismological workflows in mines. This is intended to simulate a range of possible expert opinions on the distance threshold for data selection (here we assume that the hypocentral distance is of such utmost importance that observational quality differences may be neglected, which is not the case in crustal studies). We then apply our novel approach to the complete data set for the calibration blast and compare the results with those in the reference solution. Finally, we apply our methodology to the natural seismic event, to observe its performance for a larger and more complex event.

2. Data and Methods

In our study, we performed three different experiments based on two data sets of seismic measurements. Raw seismic waveforms were recorded in the Kiirunavaara mine (Sweden) by a seismic network (253 seismic sensors spanning the 3D volume around the ore body (Dineva et al., 2022)). From the raw seismic waveforms, P- and S-wave arrival times were automatically retrieved and manually revised by the local seismic system provider (Institute of Mine Seismology, IMS). These arrival times are the data used to solve the geophysical inverse problem of locating the seismic source. In the first data set, the seismic source is a man-made blast used to calibrate the seismic system. The position of the seismic source is known with an accuracy less than 1 m. This data set can be used as the input of a robust field test, as the source location is known with an accuracy higher than the half-width of the characteristic

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2023GL102983

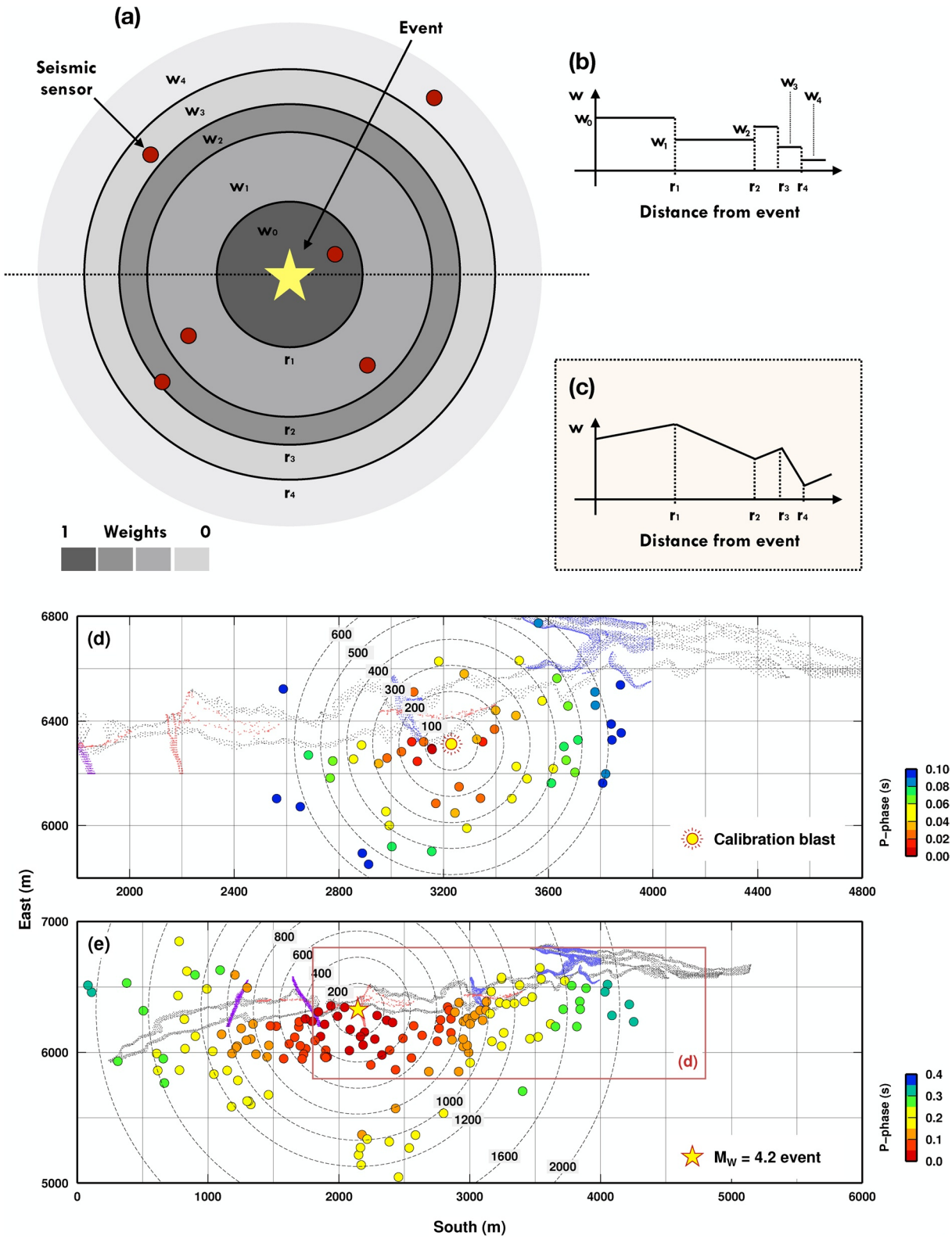

Figure 1.

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

wavelength of the P-wave (given a blast corner frequency close to 1,000 Hz, if we assume Vp = 6,600 m/s, e.g., the half-wavelength is 3.3 m). Due to the limited amount of explosive used (5 kg), seismic waves have been clearly recorded at 57 seismic sensors only (with 57 P- and 9 S-wave arrival times), which translates in a minimum and maximum sensor distance from the source of 83 and 710 m, respectively. The second data set is based on P- and S-waves generated during a M_w 4.2 event that occurred on 18 May 2020. This large event partially destroyed the mine infrastructure in a section about 1,300 m wide (Dineva et al., 2022). The resulting P- and S-waves were recorded at 151 seismic sensors (151 P-wave and 81 S-wave arrival times) as far as about 2,130 m from the preliminary seismic source location. Our preliminary seismic source is the official location (Dineva et al., 2022), even though the damaged area may be as large as 700 × 250 m. For the present study, we specifically revised all P- and S-wave arrival times associated to this event to include as many seismic sensors as possible. It is worth noting that some seismic sensors close to the rupture area were destroyed during the event, limiting the availability of data near the source. The closest sensor to the preliminary seismic source is about 140 m away.

2.1. The Reference Solution: McMC Location of a Seismic Event

Our fist experiment obtains a reference solution that consists of estimated source locations obtained using only sensors within a maximum distance from the calibration blast (which is equivalent to have equal weight for all stations within the max distance and ignore stations that are beyond the maximum distance). By considering a range of possible maximum sensor distances, we aim to reproduce the results that would be obtained by different expert opinions.

The observed arrival time data in a vector $\mathbf{d}$ are $t^{\text{obs}}_{x,i}$, where $x = P$ or S, $i = 1, \ldots, N$, and $N = N_P + N_S = 66$, with N_P and N_S being number of P- and S-wave arrival times, respectively. We apply a standard Markov chain Monte Carlo approach to solve the inverse problem of locating the source. We make four simplifying assumptions: (a) The rock volume is a homogeneous half-space; (b) The seismic event is a point source; (c) The arrival times have associated uncertainties equal to the sampling rate (i.e., σ_0 = 1/6,000 s); and (d) The covariance matrix of the data errors $\mathbf{C}^*_e$ is diagonal, that is, $\mathbf{C}^*_e = \sigma_0^2\mathbf{I}$, with $\mathbf{I}$ being the identity matrix (see Riva et al., 2023) for details on the methodology).

The model vector $\mathbf{m}$ in the inverse problem contains eight parameters. Six parameters are related to the physical model: the coordinates of the seismic source (X_s, Y_s, Z_s), the origin time measured with respect to the first P-wave arrival time (OT), the P-wave velocity of the half-space (V_P), and the ratio between P-wave velocity and S-wave velocity in the half-space (V_P/V_S). Two additional "hyperparameters" (Malinverno & Briggs, 2004) π_P and π_S, defined below, control the data uncertainties. Thus:

$$\mathbf{m} = (X_s, Y_s, Z_s, OT_s, V_P, V_P/V_S, \pi_P, \pi_S)$$

In our experiment, all prior probability distributions are uniform within the minimum and maximum values given in Table S1 in Supporting Information S1.

For the given assumptions, the seismic ray paths for model $\mathbf{m}$ are straight lines from the seismic source in (X_s, Y_s, Z_s) to the known position of each seismic sensor. Thus, predicted arrival times $t^{\text{pred}}_{x,i}$ (with $x = P$, S and $i = 1, \ldots, N$) can be easily computed from the source-sensor distance, V_P, and V_P/V_S. The vector $\mathbf{e} = t^{\text{obs}}_{x,i} - t^{\text{pred}}_{x,i}$ contains the residual differences between predicted and observed arrival times. The hyperparameters π_P and π_S multiply the error variances that define the error covariance matrix as follows: $\sigma_x^2(\mathbf{m}) = \sigma_0^2 \cdot 10^{2\pi_x}$, where $x = P$ or S. Therefore, the error covariance $\mathbf{C}_e(\mathbf{m})$ is a diagonal matrix that contains the values of $\sigma_x^2(\mathbf{m})$ for each P- or S-wave arrival time.

In this context, the likelihood function can be written as

$$L(\mathbf{m}) = P(\mathbf{d}|\mathbf{m}) = \frac{1}{\left[(2\pi)^N |\mathbf{C}_e(\mathbf{m})|\right]^{1/2}} \exp\left(-\frac{1}{2}\mathbf{e}^T \mathbf{C}_\mathbf{e}(\mathbf{m})^{-1}\mathbf{e}\right), \tag{1}$$

Figure 1. The arrival time weights $\mathbf{w}_k$ are associated with a set of k concentric 3D spherical shells with radii $\mathbf{r}_k$, centered on a preliminary event location. (a) 2D representation of the spherical shells. (b) Constant weights within each shell. (c) Alternative parameterization with linearly varying weights within each shell. (d)–(e) Seismic data used in this study: sensor locations projected on a horizontal plane (circles) and arrival times for P-waves relative to the earliest recorded arrival time (circle colors). Dashed circles show the distance in meters from the preliminary event location. Important geological features are depicted with colored dots: gray = ore body; blue = clay zones; red/purple = diapir/diabase. Panel (d) shows seismic arrival times for a calibration blast (yellow Sun symbol), and (e) for a M_w 4.2 event (yellow star). The red box in (e) shows the smaller area plotted in (d).

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where the determinant of the diagonal covariance matrix is the product of its diagonal entries, that is, $|\mathbf{C}_e(\mathbf{m})| = \sigma_P^{2N_P}(\mathbf{m})\,\sigma_S^{2N_S}(\mathbf{m})$. The complete results for one choice of the maximum distance are reported in Figure S1 in Supporting Information S1 for references.

2.2. Assigning Data Weights With a Trans-Dimensional Algorithm

The novel approach applied in our second and third experiments (using arrival times from the calibration blast and the natural event) follows the algorithm presented in Piana Agostinetti and Sgattoni (2021), except that here we use spherical shells in space rather than change points in time. In our case, the spherical shells are centered at an approximate preliminary location of the source (see in Supporting Information S1 for a discussion on determining a preliminary source location). The models sampled by McMC are composed by six physical parameters, as done earlier, plus a variable number of parameters related to the spherical shells: the number of shells k, the k-vector of shell radii $\mathbf{r}_k$, and the $k+1$-vectors of the weights for P- and S-wave arrivals in each shell $\mathbf{w}_{k,P}$ and $\mathbf{w}_{k,S}$. Thus:

$$\mathbf{m} = (X_s, Y_s, Z_s, OT_s, V_P, V_P/V_S, k, \mathbf{r}_k, \mathbf{w}_{k,P}, \mathbf{w}_{k,S})$$

As for the previous reference solution, all Priors for the spherical shell parameters are uniform probability distributions (see Table S2 in Supporting Information S1). Given the model vector $\mathbf{m}$, we compute a modified version of the covariance matrix of the data errors $\mathbf{C_e}(\mathbf{m})$ that accounts for the data weights as follows:

$$\mathbf{C_e}(\mathbf{m}) = \mathbf{W}^{-1}(\mathbf{m})\,\mathbf{C_e^*}\,\mathbf{W}^{-1}(\mathbf{m}) \tag{2}$$

where $\mathbf{W}(\mathbf{m})$ is a diagonal matrix that contains the weight assigned to each arrival time; as $\mathbf{C_e^*}$ is diagonal, $\mathbf{C_e}(\mathbf{m})$ is also diagonal. We remember here, as stated in Section 1, that an equivalent view is that the algorithm samples variances, rather than weights, in each of the spherical shells as in a hierarchical Bayes strategy (Malinverno & Briggs, 2004). In fact, in a Bayesian framework, "data noise" includes observational and theoretical errors. For example, here, as the source-station distance increases, the theoretical "noise," due to a wrong velocity model hypothesis, increases. The resulting variances in the diagonal matrix $\mathbf{C_e}(\mathbf{m})$ will be consistent with the size of the residual differences between predicted and observed arrival times in the vector $\mathbf{e} = t_{x,i}^{\text{obs}} - t_{x,i}^{\text{pred}}$.

Writing δ_i as the distance from the source of the sensor recording the i-th arrival time, the entries of $\mathbf{W}$ are

$$\mathbf{W}_{ii} = 10^{-w_i(\mathbf{m})}, \tag{3}$$

where $w_i(\mathbf{m})$ is

$$w_i(\mathbf{m}) = \begin{cases} w_{1,x} & \text{if} \quad \delta_i < \tilde{r}_1 \\ w_{j,x} & \text{if} \quad \tilde{r}_j < \delta_i < \tilde{r}_{j+1} \qquad 1 < j < k, \\ w_{k,x} & \text{if} \quad \tilde{r}_k < \delta_i, \end{cases} \tag{4}$$

where $x = P$ or S depending on whether the i-th arrival time is for a P- or S-wave and $\tilde{\mathbf{r}}_k$ is the ordered version of the vector $\mathbf{r}_k$ (i.e., $\tilde{r}_1 < \ldots < \tilde{r}_k$). The likelihood function is as in Equation 1 with the covariance matrix $\mathbf{C_e}(\mathbf{m})$ of Equation 2.

Composing the covariance matrix $\mathbf{C_e}(\mathbf{m})$ needs a preliminary source location to compute the source-sensor distances δ_i. In general, an approximate preliminary location is available soon after the event takes place and can be safely used. If a preliminary event location was not available, the position of the sensor that receives the earliest P-wave arrival can be used as well. Here we use as preliminary locations the actual location of the calibration blast and the location of the natural event published in Dineva et al. (2022); the coordinates of these preliminary locations are in Table S3 in Supporting Information S1.

3. Results

The reference solution results are shown in Figure 2. Starting with all the available data (all 57 seismic sensors that recorded the blast to a maximum distance of 800 m from the source), we get a posterior mean event location

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(a)

Max R (m)	N	X (m)	Y (m)	Z (m)	dX / Xe (m)	dY / Ye (m)	dZ /Ze (m)	D (m)
200.0	5	3227	6313	-1075	-4 / 4	1 / 7	3 / 6	**6**
300.0	19	3227	6312	-1080	-4 / 2	0 / 2	-2 / 2	**4**
400.0	27	3227	6311	-1081	-4 / 2	-1 / 2	-3 / 2	**6**
500.0	40	3225	6312	-1081	-6 / 2	0 / 3	-3 / 3	**7**
600.0	46	3224	6310	-1082	-7 / 2	-2 / 3	-4 / 3	**9**
700.0	56	3224	6305	-1086	-7 / 2	-7 / 4	-8 / 4	**13**
800.0	57	3224	6305	-1085	-7 / 2	-7 / 4	-7 / 4	**13**

(b)

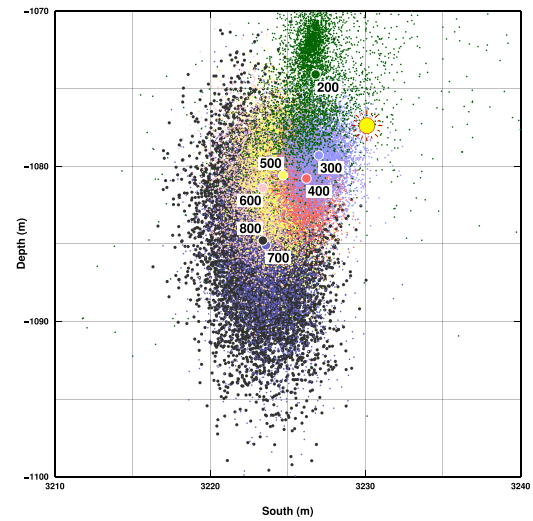

Figure 2. (a) Blast locations obtained in the reference solution with data for sensors that are within different maximum distances from the actual source location. The last column reports the distance *D* between the posterior mean and true location of the source. The best location (where *D* is minimum) is obtained with data from sensors up to 300 m from the blast. The *X* (South), *Y* (East), and *Z* (depth) columns list the posterior mean value for the location coordinates. (b) Source locations sampled by the McMC algorithm for different maximum source-sensor distances projected on the *X*-*Z* vertical plane (dots). The maximum source-sensor distances are 800 m (black dots), 700 m (dark blue), 600 m (pink), 500 m (yellow), 400 m (red), 300 m (light blue), and 200 m (green). Colored circles are posterior mean locations. The yellow sun indicates the true position of the calibration blast.

which is about 12 m away from the blast, with estimated uncertainties as large as 7 m. We then start removing data from sensors farther than 700 m, 600 m, etc., in steps of 100 m (see Figure 2 and "Materials and Methods"). The event location uncertainties and the differences with the actual blast position reach a minimum for a maximum sensor distance of 300 m (19 sensors). Considering sensors closer to the source (200 m, 5 sensors) results in an increase in uncertainties and location error.

Our novel approach applied to the blast data gives results that are consistent with those obtained in the reference solution (Figure 3 and Figure S2 in Supporting Information S1). The variation of weights with distance for both P- and S-wave arrival times follows a simple pattern, with a single step decrease at about 380 ± 30 m from the source (Figures 3b and 3c). The weights for S-wave arrival times decrease much more sharply than those for P-waves. This main step is well defined, as seen from the histogram of the sampled shell radii (Figure 3d), although the histogram of the number of shells has a maximum between 5 and 7 (Figure 3a). The sampled weights result in a cloud of event locations that closely reproduces what was found in the reference solution for a maximum distance of 400 m (red vs. black dots in Figure 3e).

In crustal studies, it has been observed that event location uncertainties depend on the azimuthal coverage (Husen et al., 2012). Here we computed the azimuthal coverage of the 3D distribution of seismic sensors (see "Materials and Methods"). Azimuthal coverage reaches a nearly stable value at a distance of ca. 300 m from the source, and it does not change substantially at greater distances (Figure 3d). The best reference solution was found when selecting stations only within 300 m from the source, which is also close to the distance where the weights obtained in our new method decrease substantially.

We apply our data-space exploration algorithm to the arrival times of the natural event (Figure 4). This event has a magnitude M_w 4.2, it is composed of several subsequent processes, where the extent of the very first sub-event S1 is likely ca. 100–200 m (Dineva et al., 2022)). The final posterior distribution of the source location is close to that initially estimated (Figures 4a and 4b). The pattern of weights with distance is more complex compared to that obtained for the blast. There is a main step at about 1,230 ± 70 m, but also three other maxima in the histogram of shell radii (marked with colored arrows in Figure 4c). The weights for the P-wave arrival times slightly increase from the origin to 150 ± 50 m (gray arrow) and remain near a maximum value between 150 ± 50 and 500 ± 60 m (red dashed arrow). At greater distances, the weight decrease slightly to a nearly constant value out to 1,230 ± 70 m (red arrow), where there is a sharp decrease of almost one order of magnitude. The weights increase again at about 1,900 ± 60 m (blue arrow).

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

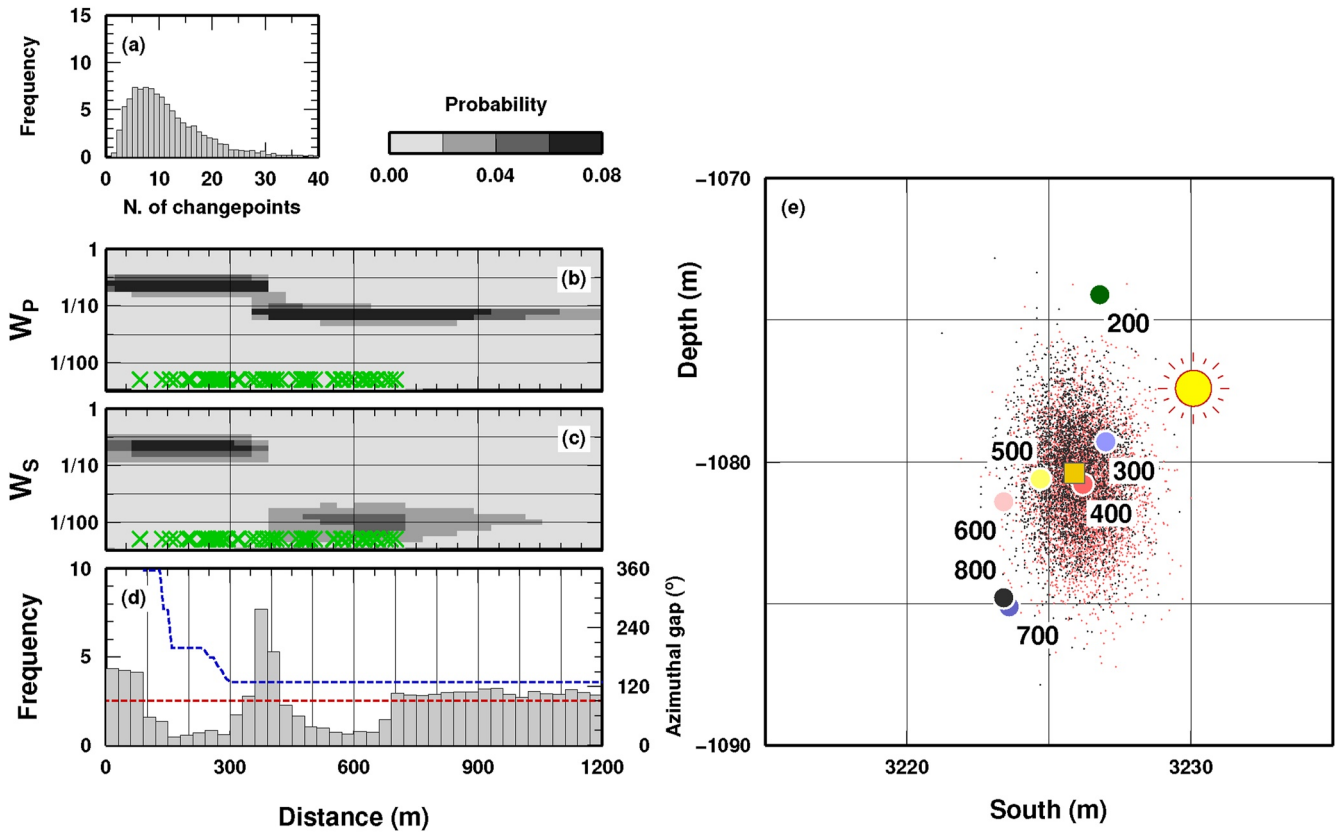

Figure 3. Application of the novel data-weighing method to blast recordings. (a) Posterior probability density function (PDF) of the number of spherical shells, approximated by the histogram obtained by McMC sampling. (b) Posterior PDF of the weights assigned to P-wave arrival times as a function of source-sensor distance. Green crosses indicate the distance of each sensor from the source. (c) As in (b) for S-wave arrival times. (d) Posterior PDF of shell radii. The blue dashed line indicates the azimuthal gap as a function of distance from the source (see "Materials and Methods" for a definition). The red dashed line indicate the prior probability distribution for the shell distance. (e) Sampled source locations projected onto a *X*-*Z* vertical plane (black dots) compared to source locations in the reference solution for a maximum source-sensor distance of 400 m (red dots; see Figure 2b). The yellow square shows the posterior mean source location obtained with the data-weighing method. Colored circles are the posterior mean source locations obtained in the reference solution (same as in Figure 2b). The yellow sun indicates the true position of the calibration blast.

We also conducted a test to check whether the overall pattern of weights with distance is significantly affected by the simple parameterization of constant weights in each spherical shell. To this end, we implemented an alternative parameterization where weights are defined at the shell boundaries and vary linearly within each shell (Figure 1c). The pattern of weights with distance obtained with linearly varying weights is very similar to that obtained with constant weights (see yellow contours in Figure 4c, Figures S3 and S4 in Supporting Information S1). The choice of parameterization does not seem to strongly control the variation of weights with distance.

4. Discussion

In our first test with a controlled blast, the reference solution found with all seismic stations within 300 m from the source could seem to outperform our novel approach, because the mean PPD event location is slightly closer to the blast true position with respect to the mean PPD solution found with the novel approach. However, this difference is small (less than 2 m) and the pattern of the weights found with the novel approach closely mimics a step function (with a step at about 380 m). In case another maximum distance is picked (200, 400, 500 …meters), that is, in case of a different expert opinion, the solution found by our novel approach is better than the reference solution. Thus, we can affirm that our approach outperforms and is more robust than standard approaches in case of limited information about the data.

Comparing the results obtained in the two tests carried out with our novel approach, we note that the pattern of weights with distance seems to be event-dependent and is not a constant in a particular sensor network. While

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

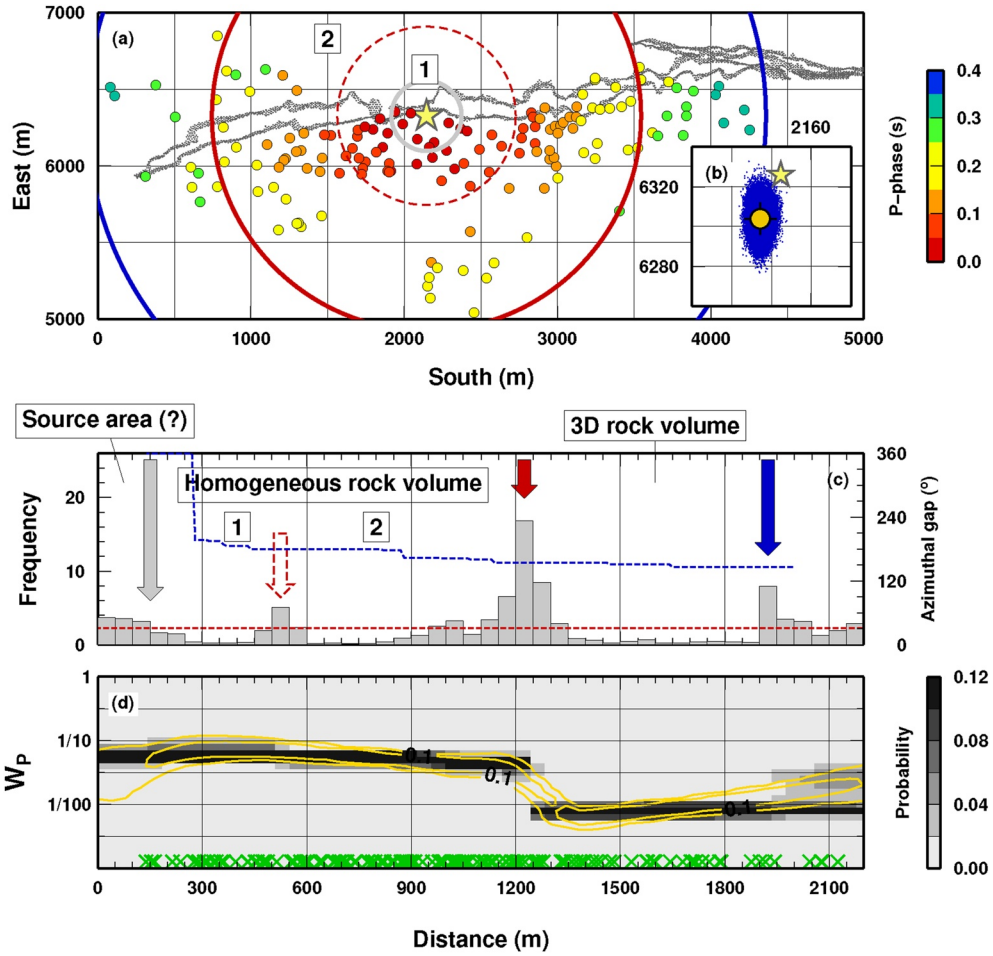

Figure 4. Application of the novel data-weighing method to recordings of the M_w 4.2 natural event. (a) Seismic network geometry (same as in Figure 1e). Colored circles report the position of the main modes in the histogram of sampled shell radii, indicated with colored arrows in panel (c). The inset (b) plots the sampled source locations projected onto a *X-Z* vertical plane (blue dots) compared to the preliminary location of the event (yellow star). (c) Posterior probability density function (PDF) of shell radii, approximated by the histogram obtained by McMC sampling. The colored arrows indicate the main modes in the posterior PDF, corresponding to the boundaries of the source area (gray arrow), homogeneous rock volume with all sensors on the same side of the ore body (dashed red), homogeneous rock volume (red), heterogeneous rock volume (blue). (d) Posterior PDF of the weights assigned to P-wave arrival times as a function of source-sensor distance. Green crosses indicate the distance of each sensor from the source. The yellow contours display the posterior PDF of the data weights obtained with the alternative parameterization in Figure 1c (see also Figures in Supporting Information S1).

further research would be necessary to determine which event parameters (e.g., magnitude, location) affect the weight pattern, the results indicate that a static workflow for all events would probably introduce artifacts and underestimate the actual uncertainties. In contrast, our approach is adapted to each single event, giving a solution that is statistically consistent and parsimonious (in terms of complexity of the weight pattern parameterization). Alternatively, a more detailed approach to data weighting (e.g., considering a more general 2D/3D spatial distribution of weights, instead of concentric shells, as done in Ghalenoei et al., 2022) could help in reducing the difference in the results obtained for different events.

The relationship between azimuthal coverage of the event and our results is not straightforward. In the controlled blast, the main decrease in the weights we obtains is near the distance where the azimuthal coverage increases substantially (Figures 3d and 4c). On the other hand, there is no clear correspondence between weight patterns and azimuthal coverage in the test with a natural event. This suggests that azimuthal coverage is only one of the factors affecting the reliability of the inverted source location. A workflow based on this parameter (e.g., where distant seismic sensors are removed once the azimuthal gap decreases below a certain threshold) may not give

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

optimal results. In fact, if the gap is larger than 180° with stations in the epicenter near vicinity, a moderately distant station closing this gap may be very useful if the real subsurface velocities are not perfectly well known (which is almost never the case). On the other hand, closing a gap to significantly less than 180° with a single very distant station is at least questionable (if not useless) when considering the uncertainties of phase identification and frequency difference in first arriving/visible wavelets.

The pattern of weights allows us to interpret the results in terms of specific properties of the rock volumes at different distances from the source. We suggest that the seismic sensors closest to the natural event (at distances <150 m, first gray circle in Figure 4a), very likely are in the source area, where the assumption of a point-wise seismic source is not realistic for such a large event. Between the gray and the dashed red circle (distances of 150–500 m) the weights reach their highest values, indicating where the inverse problem assumptions should be valid. Indeed, all sensors within the red dashed circle in Figure 4a are located on the same side of the ore body, where the rock volume is expected to be comparatively homogeneous. Between the dashed and solid red circles in Figure 4a (distances of 500–1,200 m) the weights are still high, but less than in the previous interval. This is likely due to some ray-paths partially crossing the ore body and thus violating the homogeneous rock assumption. Farther than 1,200 m from the source (red circle in Figure 4a), the seismic rays start to densely sample the ore body and the surrounding rocks on both sides of the ore body itself. Here we can expect that the assumption of a homogeneous rock finally breaks down, and the weights decrease significantly. Further investigations are needed to confirm our hypothesis and to check how complex pattern in weights could be related to a less circularity in the data distribution around the seismic source in the case of the natural event than in the case of the blast.

In a more general context, our novel approach can be applied to most of the scientific inference problems, where huge amount of data need to be pre-processed in some way, without introducing bias related to preconceptions of the data-analysts. We mention that our approach only works if data can be ordered or clusterized in some way. Here for example, they are "ordered" with regards to the source-sensor distance. In this case, "ordering" is necessary, but it is not the only way of performing the trans-dimensional data-space exploration. To apply our approach, we need either a metric to be used to "measure" some kind of data-point distance in the data-space, or, equivalently, some kind of data characterization which enables data clustering, where the trans-dimensional approach is used to define the number of data cluster from the data themselves.

Data Availability Statement

Software and data (i.e., P- and S- arrival times for the blast occurred in the mine) can be freely accessed on Mendeley Data Repository (Piana Agostinetti, 2023).

Acknowledgments
We thank the Editor Daoyuan Sun, and the reviewers Jan Dettmer and Andreas Fichtner for their constructive comments. We thanks Mirjana Boskovic for discussion on calibration blasts in Kiruna mine. N.P.A. research is conducted with the financial support of the Ministero dell'Istruzione e del Merito in the framework of the programme "Dipartimenti di eccellenza (2018–2022)." The Generic Mapping Tools software has been used for plotting the figures of this manuscript (Wessel & Smith, 1998).

References

Bodin, T., Sambridge, M., Rawlinson, N., & Arroucau, P. (2012). Transdimensional tomography with unknown data noise. *Geophysical Journal International*. https://doi.org/10.1111/j.1365-246X.2012.05414x

Dettmer, J., & Dosso, S. E. (2012). Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains. *Journal of the Acoustical Society of America*, *132*(4), 2239–2250. https://doi.org/10.1121/1.4746016

Dineva, S., Dahner, C., Malovichko, D., Lund, B., Gospodinov, D., Piana Agostinetti, N., & Rudzinski, L. (2022). *Analysis of the magnitude 4.2 seismic event on 18 may 2020 in the Kiirunavaara mine*. RaSiM Conference.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D., & Nicolson, H. (2016). Transdimensional love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry. *Geophysical Journal International*, *208*(1), 36–58. https://doi.org/10.1093/gji/ggw286

Ghalenoei, E., Dettmer, J., Ali, M. Y., & Kim, J. W. (2022). Trans-dimensional gravity and magnetic joint inversion for 3-D earth models. *Geophysical Journal International*, *230*(1), 363–376. https://doi.org/10.1093/gji/ggac083

Husen, S., Kissling, E., & von Deschwanden, A. (2012). Induced seismicity during the construction of the Gotthard Base Tunnel, Switzerland: Hypocenter locations and source dimensions. *Journal of Seismology*, *16*(2), 195–213. https://doi.org/10.1007/s10950-011-9261-8

Kolb, J. M., & Lekić, V. (2014). Receiver function deconvolution using transdimensional hierarchical Bayesian inference. *Geophysical Journal International*, *197*(3), 1719–1735. https://doi.org/10.1093/gji/ggu079

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, *151*(3), 675–688. https://doi.org/10.1046/j.1365-246x.2002.01847.x

Malinverno, A., & Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, *69*(4), 1005–1016. https://doi.org/10.1190/1.1778243

Mustać, M., & Tkalčić, H. (2015). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, *204*(1), 311–323. https://doi.org/10.1093/gji/ggv458

Piana Agostinetti, N. (2023). Weighing geophysical data with trans-dimensional algorithms: An earthquake location case study. *Mendeley Data*. https://doi.org/10.17632/zjmcrjcx8p.1

Piana Agostinetti, N., Kotsi, M., & Malcolm, A. (2021). Exploration of data space through trans-dimensional sampling: A case study of 4D seismics. *Journal of Geophysical Research*, *126*(12), e2021JB022343. https://doi.org/10.1029/2021JB022343

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Piana Agostinetti, N., & Sgattoni, G. (2021). Changepoint detection in seismic double-difference data: Application of a trans-dimensional algorithm to data-space exploration. *Solid Earth*, *12*, 2717–2733. https://doi.org/10.5194/se-2021-79

Riva, F., Piana Agostinetti, N., Marzoratii, S., & Horan, C. (2023). The micro-seismicity of Co. Donegal (Ireland): Defining baseline seismicity in a region of slow lithospheric deformation. *Terra Nova*, 1–15. https://doi.org/10.1111/ter.12691

Sambridge, M., Gallagher, K., Jackson, A., & Rickwood, P. (2006). Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, *167*(2), 528–542. https://doi.org/10.1111/j.1365-246X.2006.03155.x

Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, *40*(3), 3-1–3-29. https://doi.org/10.1029/2000RG000089

Steininger, G., Dettmer, J., Dosso, J., & Holland, S. (2013). Transdimensional joint inversion of seabed scattering and reflection data. *Journal of the Acoustical Society of America*, *133*(3), 1347–1357. https://doi.org/10.1121/1.4789930

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

VanDecar, J. C., & Crosson, R. S. (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares. *Bulletin of the Seismological Society of America*, *80*(1), 150.

Wessel, P., & Smith, W. H. F. (1998). New, improved version of the generic mapping tools released. *EOS Trans. AGU*, *79*(47), 579. https://doi.org/10.1029/98eo00426

Xiang, E., Guo, R., Dosso, S. E., Liu, J., Dong, H., & Ren, Z. (2018). Efficient hierarchical trans-dimensional Bayesian inversion of magnetotelluric data. *Geophysical Journal International*, *213*(3), 1751–1767. https://doi.org/10.1093/gji/ggy071

19448007, 2023, 22, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL102983 by Cochrane France, Wiley Online Library on [08/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License