Exploration of the data space via trans-dimensional sampling: the case study of seismic double-difference data

Nicola Piana Agostinetti[1,2] and Giulia Sgattoni[3]

[1]Department of Earth and Environmental Sciences, Universitá di Milano Bicocca, Milano, Italy
[2]Department of Geology, Universitat Wien, Althanstrasse 14, 1090, Wien, Austria
[3]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, Italia

Correspondence: Nicola Piana Agostinetti (nicola.pianaagostinetti@unimib.it)

Abstract.

Double-difference (DD) seismic data are widely used to define elasticity distribution in the Earth's interior, and its variation in time. DD data are often pre-processed from earthquakes recordings through expert opinion, where pairs of earthquakes are selected based on some user-defined criteria, and DD data are computed from the selected pairs. We develop a novel methodology for preparing DD seismic data based on a trans-dimensional algorithm, without imposing pre-defined criteria on the selection of event pairs. We apply it to a seismic database recorded on the flank of Katla volcano (Iceland), where elasticity variations in time have been indicated. Our approach quantitatively defines the presence of changepoints that separate the seismic events in time windows. Within each time window, the DD data are consistent with the hypothesis of time-invariant elasticity in the subsurface, and DD data can be safely used in subsequent analysis. Due to the parsimonious behavior of the trans-dimensional algorithm, only changepoints supported by the data are retrieved. Our results indicate that: (a) retrieved changepoints are consistent with first-order variations in the data (i.e. most striking changes in the amplitude of DD data are correctly reproduced in the changepoint distribution in time); (b) changepoint locations in time do correlate neither with changes in seismicity rate, nor with changes in waveforms similarity (measured through the cross-correlation coefficients); and (c) noteworthy, the changepoint distribution in time seems to be insensitive to variations in the seismic network geometry during the experiment. Our results demonstrate that trans-dimensional algorithms can be effectively applied to pre-processing of geophysical data before the application of standard routines (e.g. before using them to solve standard geophysical inverse problems).

1 Introduction

Data preparation is a daily routine in the worklife of geoscientists. Before using data to get insights into the Earth system, geoscientists try to deeply understand their data sets, to avoid introducing, e.g. instrumental issues, redundant data, un-wanted structures such as data density anomalies, and many others (Yin and Pillet, 2006; Berardino et al., 2002; Lohman and Simons, 2005). All the activities for preliminary data analysis can be considered as exploration of the "data space" (Tarantola, 2005) and are mainly based on *expert opinion*. Previous experience drives scientists in selecting the most trustable portion of their experiments, cleaning data sets before using them for getting new knowledge about Earth model parameters. There are two main

reasons for moving a step forward from expert opinion. First, the huge amount of (often multidisciplinary) data, accumulating in geosciences in the last decade, requires more and more data screening and preparation, sometimes involving multidisciplinary expertise. Research activities could greatly benefit from a more automated exploration of the data space able to ease preparatory tasks. Second, expert opinion is a human activity and is mainly based on dual categories, e.g. good/bad data, and can not easily handle a continuous probability distribution over the data (i.e. expert opinion can not easily associate a continuous "confidence" measure to each data-point).

In recent years, in the framework of Bayesian inference, exploration of the data space has been introduced in a few cases to "explore" unknown features of the data sets. For example, the so called *Hierarchical Bayes* approach has been introduced to estimate data uncertainties from the data themselves (Malinverno and Briggs, 2004). More complex Hierarchical Bayes approaches have been developed to measure the data correlation as well (e.g. Bodin et al., 2012a; Galetti et al., 2016) or to evaluate an error model (e.g. Dettmer and Dosso, 2012). The exploration of the data space, in all these studies, implies considering some additional unknowns (e.g. data uncertainties or error correlation length), so called *hyper-parameters* or *nuisance parameters*, and to estimate them directly from the data. A step forward in the exploration of the data space has been presented by Steininger et al. (2013) and Xiang et al. (2018), who used a data space exploration approach to evaluate the performance of two different error models directly from the data. In such studies, the number of hyper-parameters considered is not fixed, but can assume two different values (1 or 2), depending on the error model considered. Another interesting, recent case of exploration of the data space is represented by the work of Tilmann et al. (2020), in which the authors used Bayesian inference to separate the data in two sets: "outliers" and "regular". In this case, the data themselves are probabilistically evaluated to understand their contribution to the final solution as "regular" data or "outlier", i.e. the data are classified in two different families, according to their coherence with the hypothesis of being "regular" data or not.

In this study, we push the exploration of data space in a new direction. We develop an algorithm for computing Bayesian inference specifically for the exploration of the data space. Exploration of the data space is performed through a trans-dimensional algorithm (e.g. Malinverno, 2002; Sambridge et al., 2006) so that the number of hyper-parameters is neither fixed nor limited to 1 or 2. We represent *data structure* as partitions of the covariance matrix of uncerties, i.e. changepoints that create sub-matrices of the covariance matrix with homogeneous characteristics, where the number of partitions is not dictated by the user, but it is derived by the data themselves, in a Bayesian sense (i.e. we obtain a posterior probability distribution, PPD, of the number of partitions). In this way, similar to Tilmann et al. (2020), portions of data can be classified and used differently in the subsequent steps of the analysis.

We apply our algorithm to prepare a widely exploited type of seismic data set, the seismic double-difference (DD) data-set, that has been used as input in seismic tomography for defining subsurface elasticity (e.g. Zhang and Thurber, 2003) and its variation in time (e.g. so called "time-lapse tomography", Caló et al., 2011; Zhang and Zhang, 2015). DD data need to be re-constructed from specific partitions of the original data (i.e. seismic events). Subjective choices have a great impact on the definition of DD data. In particular, such choices can be used to limit the number of DD data itself and the selection, in turn, could introduce biases in the subsequent definition of the elastic model and its variations in time. We apply our algorithm to

statistically define, in a more objective way, the distribution of partitions in the DD data. We show how a more data-driven approach can obviate expert-driven data selection and can be used as a preliminary tool for, e.g., time-lapse seismic tomography.

1.1 Double-difference data in seismology

Double-difference seismic data are widely used for relocating seismic events and imaging the subsurface (e.g. Waldhauser and Ellsworth, 2000; Roecker et al., 2021). DD data rely on the assumption of (nearly) co-located events for which seismic recordings have been obtained from the same station (Zhang and Thurber, 2003) or for the same pair of stations (e.g. Guo and Zhang, 2016). The concept of co-located events relies on expert opinion. It is generally assumed a-priori as a maximum distance between hypocenters in order to consider a pair of events to be included in the DD data, together with a high value of cross-correlation for their waveforms. A DD datum is the differential travel-time for the selected pair of events. The same scheme has been applied to more complex analyses, such as full waveform inversion (Lin and Huang, 2015). Based on the assumption of nearly co-located events, the information contained in the DD datum can be used to refine event locations (e.g. small events referred to a master event, Waldhauser and Ellsworth, 2000) or the seismic properties of the rocks in the area where events are clustered (e.g. Zhang and Thurber, 2003). In recent years, seismic monitoring of subsurface processes has also been realized through seismic tomography (e.g. Chiarabba et al., 2020) and in particular with the analysis of DD data: rock weakening due to mining activities (Qian et al., 2018; Ma et al., 2020; Luxbacher et al., 2008), granite fracturing during geothermal well stimulation (Caló et al., 2011; Caló and Dorbath, 2013) and oil & gas operations (Zhang et al., 2006). For monitoring purposes, an additional assumption is considered during DD data preparation: elastic properties of the media traversed by the seismic waves should not change between the occurrence of the selected event pairs. This fact implies the computation of the so called *time-lapse* analysis, where pre-defined time windows are considered and static images of the subsurface (Caló et al., 2011), or differential elastic models (Qian et al., 2018), are reconstructed for each time window. In any case, the most relevant issue in time-lapse tomography remains how to define the time windows, which artificially separate events and prevent their coupling to obtain DD data. How many time windows are meaningful to construct DD data? And which should be their time-lengths? This issue is critical due to the dependence of the number of DD data on the number of paired events and, thus, from the number of time windows, as schematically shown in Figure 1.

The definition of the set of time windows, on which the sequence of 3D time-lapse tomographic inversions should be computed, demands expert opinion. There are three main possibilities in time-lapse tomography: (a) imposing time windows based on known seismic history (before and after a known, relevant seismic event: Young and Maxwell, 1992; Chiarabba et al., 2020); (b) keeping the same length for all time windows (e.g. one day, Qian et al., 2018), or (c) trying to have the same amount of data in all the time windows (e.g. Patanè et al., 2006; Kerr, 2011; Zhang and Zhang, 2015). In other cases, the lengths of the time windows vary based on research needs (e.g. Caló et al., 2011). A human-defined set of time windows might mask the real variations of the physical properties, in which case the time-evolution of the elastic model found could be not associated with the investigated geophysical process.

Here, we tackle the issue of defining the number and time length of the time windows in DD data preparation though a novel approach. To simplify the experiment, we focus on closely associated events recorded on a volcanic edifice in Iceland. Such
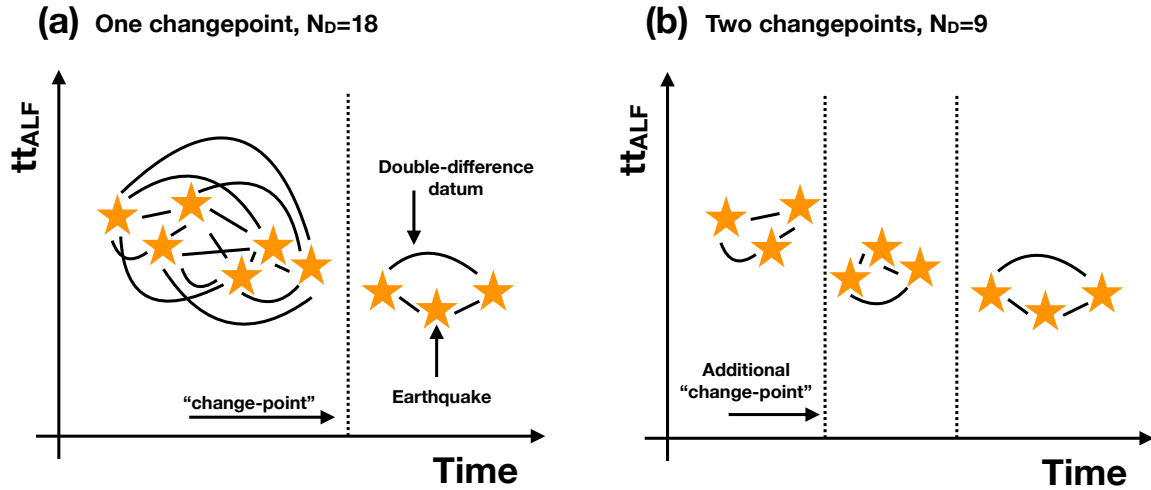
Figure 1. Schematic example of standard preparation of DD data in different time windows. Time windows are defined by changepoints (also called "hard partitions"). Here, for sake of simplicity, we represented the travel-time to station ALF for each seismic events (yellow stars) as a function of origin time. A DD datum (curved black line) is prepared for each pair of events not separated by a changepoint. (a) Here, only one changepoint is present, so $N_D = 18$ DD data can be prepared. (b) In case of two changepoints, only $N_D = 9$ DD data can be prepared.

a cluster of events, which spans no more than 100 meters in diameter, is considered as a source of repeating events that are recorded from a seismic station 6 km away for more than two years continuously. In this way, we assume perfectly co-located

95   events and we can focus on time variations of DD data. More generally, the novel approach can be applied to both temporal and spatial associations (i.e. to define both time windows and spatial length for pairing events, and composing DD data).

1.2   Background on Bayesian inference, Markov chain Monte Carlo sampling and trans-dimensional algorithms

Geophysical inverse problems have been solved for a long time following direct search or linearized inversion schemes, due to the limited number of computations needed to obtain a solution. Such solutions have been given in the form of a single "final"

100   model, presented as representative of the Earth's physical properties. Thanks to the computational resources now available, such approaches are outdated for more sophisticated and CPU-time consuming workflows, in which multiple models are evaluated and compared, to obtain a wider view of the Earth's physical properties. Algorithms based on Bayesian inference belong to this second category, for which the "solution" is no longer a single model but a distribution of probability on the possible value of the investigated parameters, following Bayes theorem (Bayes, 1763):

$$105 \quad p(\mathbf{m} \mid \mathbf{d}) = \frac{p(\mathbf{m})p(\mathbf{d} \mid \mathbf{m})}{p(\mathbf{d})} \tag{1}$$

where $p(\mathbf{m} \mid \mathbf{d})$ represents the information obtained on the model parameters $\mathbf{m}$ through the data $\mathbf{d}$, the so called "posterior probability distribution (PPD)", or simply "posterior". Such information is obtained combining the prior knowledge on the model: $p(\mathbf{m})$, with the likelihood of the model given the data: $p(\mathbf{d} \mid \mathbf{m})$. The denominator of the right term is called "evidence" and represents the the probability of the data in the model space:

$$p(\mathbf{d}) = \int p(\mathbf{m})p(\mathbf{d} \mid \mathbf{m})d\mathbf{m} . \tag{2}$$

The evidence is a high-dimensional integral that normalizes the PPD. It is generally difficult to compute and, thus, methods that do not require its computation (such as Markov chain Monte Carlo, McMC, see below) are widely used in Bayesian inference.

The likelihood of the data for a given model is necessary to evaluate and compare different sets of model parameters. It is generally expressed as

$$L(\mathbf{m}) = p(\mathbf{d} \mid \mathbf{m}) = \frac{1}{(2\pi|C_e|)^{1/2}} e^{-\frac{1}{2}\phi} \tag{3}$$

where ϕ represents the fit between model prediction p_i and the actual value of the i-th observation o_i, i.e. the residuals $\mathbf{e}_i = (o_i - p_i)$, through the covariance matrix of the uncertainties $C_\mathbf{e}$:

$$\phi = \mathbf{e}^T C_\mathbf{e}^{-1} \mathbf{e} . \tag{4}$$

Due to the difficulties in computing the evidence and the analytic solution of Equation 1, and thanks to the improved computational resources, in the last two decades the emerging trend in Bayesian inference is represented by "sampling methods", in which the direct computation of Equation 1 is replaced by the sampling of the model space according to the PPD (Sambridge and Mosegaard, 2002). One of the most famous sampling methods is called Markov chain Monte Carlo, for which the chain samples the model space according to probability rules, such as Gibbs sampler or Metropolis rule (Metropolis et al., 1953; Gelman et al., 1996). Briefly, starting from a given point in the model space, called the *current model*, a new point of the model space, called the *candidate model* is proposed and evaluated according to some rules based on the PPD. In particular, the Metropolis rule coupled to the approach developed in Mosegaard and Tarantola (1995), which is the workflow adopted in this study, accepts or rejects to move from a current model to a candidate model according to the ratio of their likelihoods, i.e.:

$$\alpha = L(\mathbf{m}_{cand})/L(\mathbf{m}_{cur}). \tag{5}$$

This is a simplified version of a more general formulation of the acceptance probability in Metropolis-based McMC (Gallagher et al., 2009). It is worth noting that our workflow does not directly specify the dimensionality of the model space. In fact, following the recent advancements in the solution of geophysical inverse problems, we do not consider models with a fixed number of parameters, but we make use of the so called trans-dimensional (trans-D) algorithm and propose candidate models with a different number of dimensions with respect to the current models. This approach is called trans-D sampling and it has been widely used for the solution of geophysical inverse problems in the last decades (Malinverno, 2002; Sambridge et al., 2006; Bodin et al., 2012a; Dettmer et al., 2014; Mandolesi et al., 2018; Poggiali et al., 2019). Trans-D algorithms have been

proven to be intrinsically "parsimonious" (Malinverno, 2002) and, thus, they preferably sample simpler models with respect to complex ones. This is one of the most important characteristics of trans-D algorithms, enabling a fully data-driven solution for the model parameters.

The Covariance matrix $C_{\mathbf{e}}$ plays an important role in Equation 5. In fact, it (inversely) scales the differences between the observations and the predictions (vector $\mathbf{e}$). Having larger values in the entries of $C_{\mathbf{e}}$ means that differences between observations and predictions are less relevant and more candidate models can be accepted during the McMC sampling through the Metropolis' rule. As a consequence, if the entries in $C_{\mathbf{e}}$ related to a certain observation are all larger with respect to the entries related to other observations, such observation will have limited importance in the McMC sampling. Defining an appropriate $C_{\mathbf{e}}$ becomes fundamental for correctly driving the McMC sampling of the PPD.

2   An algorithm for exploration of double-difference data space

What happens to the DD data-set if we create a hard-partition in time, i.e., if we artificially separate some events from the others? As clearly illustrated in Figure 1, the number of data N_D in the DD data set varies, decreasing for increasing number of hard partitions. From a Bayesian point of view, this is not admissible, because Equations 3 and 5 need to consider the same number of data points in two models to allow their comparison (see also Tilmann et al., 2020).

Our novel approach to solve this issue relies on the introduction of a family of "hyper-parameters", which represent the partitions of the events, and such hyper-parameters are used for scaling the different entries in the covariance matrix $\mathbf{C_e}$. In our approach, the number of "hyper-parameters" in the family is not fixed, but it is directly derived from the data themselves. Hyper-parameters have been introduced in geophysical inversions for estimating the data uncertainties, expressed, for example, as the variance of a Gaussian distribution (Malinverno and Briggs, 2004). Hyper-parameters are generally part of the model vector together with physical parameters. As stated in Bodin et al. (2012b), estimated hyper-parameters do not only account for measurement uncertainties, but include other contributions that build up the uncertainty in the geophysical inversions, such as simplification of the physics included in the forward solutions, or simplified model parameterization. Hyper-parameters have been used to estimate uncertainty models (Dettmer and Dosso, 2012; Galetti et al., 2016) or to discriminate between two different families of uncertainty models (Steininger et al., 2013; Xiang et al., 2018). In this last case, the number of hyper-parameters belonging to a model vector is not constant, but can be one or two, depending on the family. More recently, a nuisance parameter has been introduced to evaluate the probability for each datum to belong to the "regular data" or to the "outliers" (Tilmann et al., 2020).

For the DD case, we introduce a family of hyper-parameters to estimate which portions of the DD data violate our initial assumptions. In fact, in our assumptions, double-difference data are computed from pairs of seismic events that occurred in the same rock volume, recorded at the same seismic station. For perfectly co-located events, and in absence of any change in the rock seismic velocity field between the first and the second event, double-difference measurements should have mean 0 and should be distributed following a simple Gaussian error model, which can be a represented by the (diagonal) covariance matrix $\mathbf{C_e^*}$, with the DD uncertainties along the diagonal. Here we assume no correlation between uncertainties computed for

two different event pairs (see Appendix B1 for the definition of $\mathbf{C}_\mathbf{e}^*$ from our data). In this case, the value of the fit $\tilde{\phi}$, expressed as:

$$\tilde{\phi} = \mathbf{d}^T \mathbf{C}_\mathbf{e}^{*-1} \mathbf{d} \tag{6}$$

with $\mathbf{d}$ the DD data vector, should be close to N_D (N_D is the number of DD data, i.e. the length of the DD data vector).

When the value of $\tilde{\phi}$ significantly deviates from N_D, a modified covariance matrix $\mathbf{C}_\mathbf{e}(\mathbf{m})$ should be adopted, where the portion of the data inconsistent with the hypotheses are considered differently from the portions of DD data that do not violate the hypotheses. The new modified covariance matrix $\mathbf{C}_\mathbf{e}(\mathbf{m})$ is obtained as

$$\mathbf{C}_\mathbf{e}(\mathbf{m}) = W^{-1}(\mathbf{m}) \ \mathbf{C}_\mathbf{e}^* \ W^{-1}(\mathbf{m}) \tag{7}$$

where the matrix $W(\mathbf{m})$ is a diagonal matrix that depends on the weight for each DD datum, based on the hyper-parameters. Noteworthy, if we use Equation 7 in Equation 3, we see that in our case the dependence of the Likelihood function on the model does not reside anymore in the residuals, as is generally the case in geophysical inverse problems, but only in the Covariance matrix. However, for a simple case such as ours, we highlight that this dependence could in principle be moved back to the residuals, if we allow the physical assumptions to be variable in time (i.e., if we allow the elastic model to change in time, which in our assumption can not).

Following a Bayesian inference approach, we reconstruct the statistical distribution of the hyper-parameters (i.e. event partitions) in time through trans-dimensional McMC sampling. The fully novel idea in our algorithm resides in the trans-dimensional behaviour of our exploration of the data space. In fact, the number of hyper-parameters in the model (and, thus, the number of partitions of $\mathbf{C}_\mathbf{e}^*$) is not fixed and can change along the McMC sampling. At the end of the McMC sampling, we can compute a PPD of the number of partitions in the problem, an information fully prescribed from the data and priors. In Figure 2, we present the flow-chart of our algorithm indicating the main elements. The algorithm follows a standard McMC sampling scheme based on the Metropolis' rule. First of all, we need to define the model parameterization, i.e. how we formalize our family of hyper-parameters (Section 2.1). Thus, an explicit definition of the prior probability distribution on the parameters which compose the models is needed (Section 2.2). Such parameters are perturbed along the McMC sampling following prescribed rules which are randomly applied to propose a new candidate model (Section 2.3). Finally, we present the DD data which will be used to test our algorithm, and how we obtain them (Section 2.4 and Appendix B).

2.1 Model parameterization

In our algorithm, a model is described by a set of k changepoints that define the partitions of $\mathbf{C}_\mathbf{e}^*$, and their associated quantities, that is: $\mathbf{m} = (k, T_k, \pi_k)$. The k-vector T_k represents the time-occurrence of the k changepoints, whereas the k-vector π_k contains the weights associated with each changepoint. We assume that a DD datum d_{ij}, associated with events i and j,
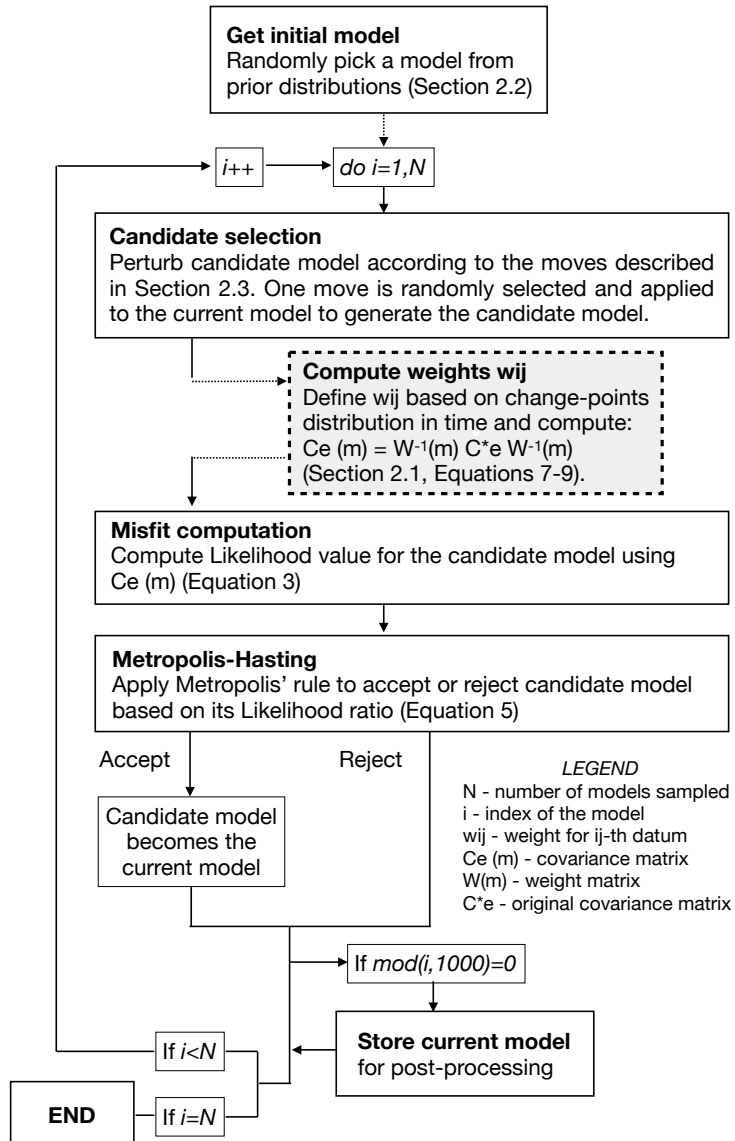
Figure 2. Flowchart of the algorithm. The grey-shaded box indicates the element which can be shared on multiple CPUs on a cluster.

retains its original variance σ_{ij}^2 if no changepoint occurs between OT_i and OT_j. Otherwise, its importance is modified with weight $\mathbf{W_{ij}(m)}$:

$$\mathbf{W_{ij}(m)} = 10^{-w_{ij}(\mathbf{m})},\tag{8}$$

where w_{ij} is computed as:

$$w_{ij}(\mathbf{m}) = \sum_{p=1}^{k}\pi_p \quad if \quad OT_i < T_p < OT_j \quad ,\tag{9}$$

recalling that $\mathbf{W}$ is a $N_D \times N_D$ diagonal matrix and W_{ij} represents the element along the diagonal associated with DD datum d_{ij}. Following our approach for defining the $w_{ij}(\mathbf{m})$, specifically the sum of the values associated with the relevant changepoints, we assume that a pair of "distant" events in time has more probability of being "separated" by one or more changepoints and, thus, of having a lower weight $\mathbf{W_{ij}(m)}$. This assumption reflects the standard process of DD data, in which distant (in space and/or time) events are almost never paired in DD data-sets. However, if no changepoints are present between distant events, our trans-dimensional approach still works, reducing the number of changepoints to the minimum. Several synthetic tests, shown in the Appendix A, demonstrate that non-necessary changepoints are removed from the family due to the parsimoniosity of the trans-dimensional algorithm, and new ones have limited probability of being accepted in the family.

2.2 Prior information

Uniform prior probability distributions are selected for our inverse problem. Here, the number of changepoints is bounded between 0 and 100. Changepoints can be distributed everywhere in time between 2011.5 and 2013.7. To make the algorithm more efficient, we set a minimum distance between two changepoints as large as 0.5 day (Malinverno, 2002). Changepoint weights π_k follow a uniform prior probability distribution between 0.0 and 1.0.

2.3 Candidate selection

Having an efficient workflow for carrying out the McMC sampling is fundamental for keeping the CPU-time within acceptable limits. From a theoretical point of view, any recipe can be implemented at the core of the McMC due to the fact that results (i.e. Equation 1) do not depend on the McMC details[1], i.e. the same prior information and the same data will give the same PPD, whatever recipe is selected for the McMC sampling. However, inefficient recipes can take too long to sample adequately the PPD and, thus, from a practical point of view, the users should spend some time in defining a proper recipe. In our case, to perturb the current model and propose a new candidate model, we randomly select one of the following four "moves":

1. (This move is randomly selected with probability $P_1 = 0.4$) The i-th changepoint is moved from its time-position T_i. There are two equally-possible perturbations: the changepoint time-position T_i is randomly selected from the prior, or

---

[1]as long as the recipe follows the necessary probabilistic rules (Sambridge and Mosegaard, 2002; Mosegaard and Sambridge, 2002)

the changepoint time-position T_i is slightly perturbed from the original value in the current model with a micro-McMC approach (see Appendix A2, Piana Agostinetti and Malinverno, 2010, for the details of the micro-McMC)

2. ($P_2 = 0.4$) The weight π_i of the i-th changepoint is perturbed with a micro-McMC approach (Appendix A2, Piana Agostinetti and Malinverno, 2010)

3. ($P_3 = 0.1$) "Birth" of a changepoint: a new changepoint is added to the current model

4. ($P_4 = 0.1$) "Death" of a changepoint: a changepoint is removed from the current model.

To increase the acceptance probability (Equation 5), the candidate model is generated randomly selecting only one of the possible moves (selected with the prescribed probability). In this way, the difference between candidate and current models should be limited and their Likelihood values should be close one to each other. The last two moves represent the trans-dimensional moves, in which the dimensionality of the model is changed from the current model to the candidate. For move (3), we follow the approach described in Mosegaard and Tarantola (1995) and we propose a completely new changepoint with T_{k+1} and π_{k+1} randomly sampled from their prior distributions. For move (4), we simply randomly select one changepoint and remove it from the model.

2.4 Data

We test our novel methodology for preparing DD data to a single station data set recorded on Katla volcano in Iceland, during a two year monitoring experiment. We collected all the waveforms recorded by one seismic stations on the volcano flack, which is far enough from a source of repeating earthquakes (i.e. all waveforms have a high degree of similarity) to be considered a point-wise seismic source. Details about the data preparation are given in Appendix B. The DD data are represented by the vector $\mathbf{d} = d_{ij}$ with $i = 1, ..., N_e - 1$, $j = i+1, ..., N_e$ and N_e the number of events. Starting with $N_e = 1119$ events, we obtain $N_e \times (N_e - 1)/2$ DD data. The total number of DD data is $N_D = 625521$. DD data values d_{ij} and uncertainties σ_{ij}, associated with events i and j, are reported in Figure 3. Striking changes in DD values suggest the presence of clustering of data in time, but the exact number and positions of such clusters are not easy to define by visual inspection. Moreover, some of those changes could be due to modifications of the seismic network, in principle mapped in our DD data. In fact, our DD data depend on the Origin Time (OT) computed exploiting all the recordings from the seismic network. Thus, a change in the seismic network configuration could influence the quality of seismic network detections and locations, which in turn could introduce a bias in our data, as a shift in the DD value or an increase in DD uncertainties. Given the independent processes used to define each single DD datum, as a first approximation we consider our final covariance matrix of the uncertainties in the DD data $\mathbf{C}_e^*$ as a 625521×625521 diagonal matrix, with the square of the uncertainties presented in Figure 3b along the diagonal, omitting the correlation between uncertanties given by, e.g., biases in OT determination.
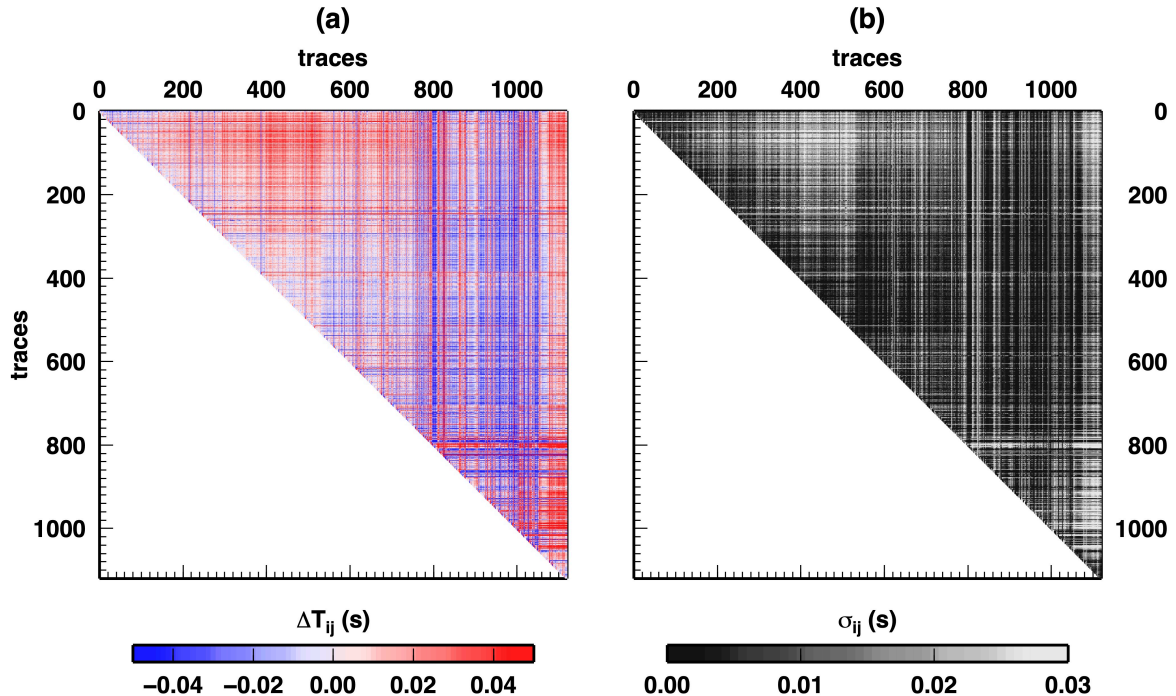
Figure 3. DD data presented as a matrix of ij-pairs: (a) values; and (b) uncertainties.

3 Finding data-driven time-variations of rock elasticity during Katla's seismic swarm

We apply our novel methodology for the definition of the changepoints in DD data to the data set recorded on Katla volcano in Iceland, during a two year monitoring experiment. Based on the observations of the limited dimension of the cluster with respect to the events-station distance (100 m versus 6.0 km) and the overall high similarity of the waveforms (correlation coefficient always larger than 0.9), our algorithm is able to map out which portions of the data violate our underlying hypotheses: co-located events and constant elasticity field in time. Separating those two effects with a single station would be challenging, but here we want to illustrate in detail how the time-occurrence of the changepoints is defined and compare them to other potential approaches for the definition of changepoints in DD data, namely, variations in seismicity rate and waveforms cross-correlation.

As shown in Figure 1, defining changepoints for DD data using expert opinion is a dangerous task, due to the limitation in the number of data available for subsequent uses. For example, seismologists could be tempted to test if using fewer data could give better images in a subsequent tomographic inversion, based on pre-defined ideas about the subsurfaces structures.
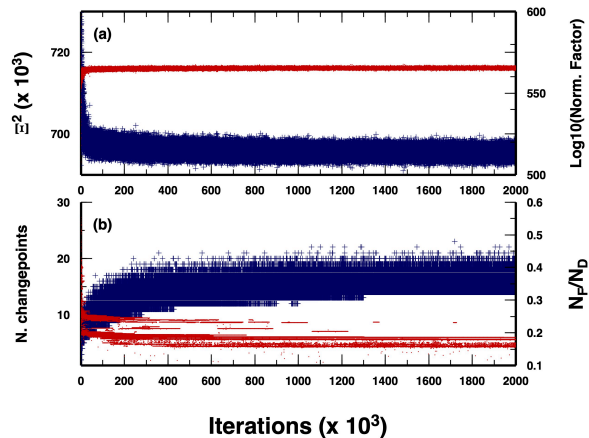
Figure 4. Evolution of some parameters along the McMC sampling. (a) Chi-square value (blue crosses) and logarithmic value of the normalizing factor in the likelihood function (red dots). (b) Number of changepoints in the sampled models (blue crosses) and ratio between the number of un-affected data (i.e. DD data where the two events are not separated by any changepoint) and the total number of DD data, N_F/N_D (red dots).

In fact, some changes in DD data are obviously present in the observations (Figure 3, between close to event 550 and 1050 for example), but others are more subtle to be defined.

We compute the data-space exploration running 100 independent McMC samplings, for which each chain sampled 2 million of changepoint models. We discarded the first million models and collected one model every 1000 in the subsequent million models. Our final pool of models used for reconstructing the PPD is composed by 100 000 models. The full CPU-time for running the algorithm is about 19 hour on a 100-CPU cluster. The value of chi-square decreases in the first half-million models, together with the logarithmic value of the normalizing factor in Equation 3 (Figure 4a). The number of changepoints reaches a stable value around 15 to 20 after 1 million models confirming the length of the burn-in period used (Figure 4b). The ratio between the number of event pairs not separated from any chagepoint (N_F) and the total number of DD data (N_D) is also stable around 0.2 after 1 million models, indicating that no relevant changepoint is added in the second half of the McMC sampling. It is worth noting that N_F should approximately indicate the number of DD data to be used in any subsequent analysis.

Looking at the full details of the PPD reconstructed from the McMC samplings, we observe the presence of long time windows completely without any changepoint (e.g. between 2011.7 and 2012.4), demonstrating the parsimoniosity of the trans-D approach: if changepoints are not supported by the data they are removed during the sampling and do not appear in the final PPD (Figure 5e). Moreover, the most probable (relevant) changepoint (changepoint number 3 in Figure 5e) perfectly aligns with one of the most striking changes in the DD data as shown in Figure 5a, confirming the goodness of the approach. The distribution of the weights clearly defines the partition of the $\mathbf{C_e(m)}$, in which initial data (i.e., DD data related to events occurring at the initial stages of the swarm, in 2011) slowly release their "connection" to later events and, thus, indicate that they
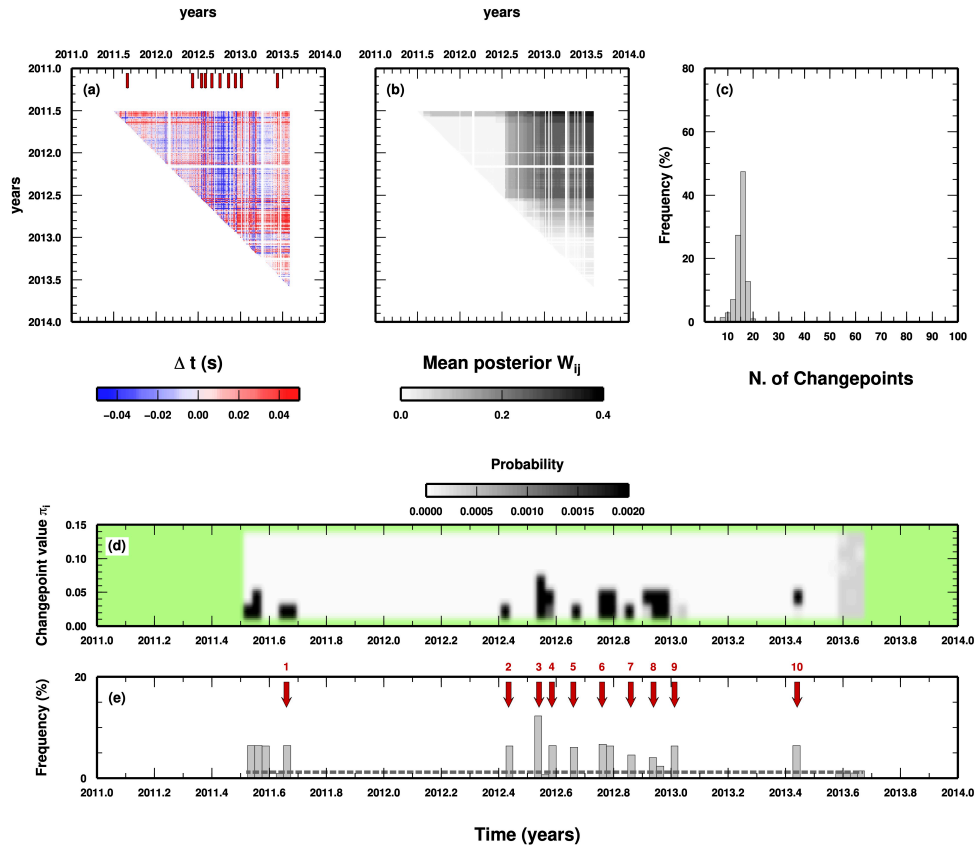
Figure 5. Results of the application of the algorithm to the Katla data set. (a) DD data and position of the most probable changepoints, see panel (e). Changepoints occurrence in time is indicated by red arrows on top. (b) Mean posterior values for the weights associated with each DD datum. (c) Histogram of the number of changepoints in the sampled models. (d) 1D marginal PPD for the values of the changepoints in time. (e) Histogram of the distribution of changepoints in time. Red arrows and numbers indicate the most probable occurrence in time for a changepoint.

should not be included in the subsequent analysis. From the histogram of the number of changepoints in each sampled model, we can see how the trans-D algorithm works: no less than 10 and no more than 20 changepoints are generally considered, even if we allow the number to increase to 100. Combining this information with the distribution of chagepoints in time given in Figure 5e, we define eleven relevant changepoints (red arrows). We acknowledge that this number could be, again, a subjective choice, however, looking at Figure 5d, we see that changepoints can be "ranked" in some sense given their mean PPD weights. For example, changepoints 2, 5, and 9 have clearly associated lower weights compared to the others and, thus, should be considered as less relevant. Our methodology does not solve all issues connected to the preparation of DD data, but, at least, it can be used to quantify the occurrence of changepoints and their importance, and such quantification can be exploited in a later stage depending on the subsequent analysis planned.
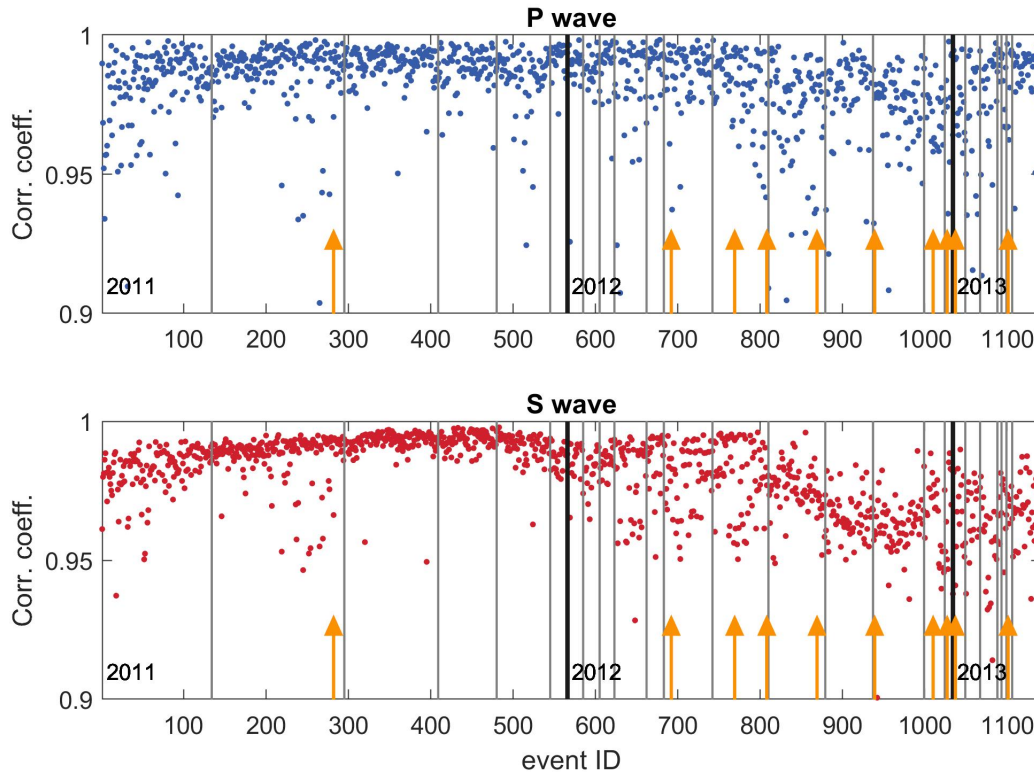
Figure 6. Correlation coefficients for P-waves (a) and S-waves (b) for each event with respect to the largest event (see Sgattoni et al., 2016a, for details). In each plot, the most probable time-occurrences for the 11 relevant changepoints are also reported (yellow arrows).

We compare the occurrence in time of the resulting 10 changepoints with the cross-correlation coefficients between each event in the seismic swarm and the largest one (see Sgattoni et al., 2016a, for details). Both P-wave and S-wave cross-correlation coefficients display some degree of variability and some well-defined patterns in the time window used in this study (Figure 6), even if we consider that the smaller values are always larger than 0.9. We observe that there is no clear correlation between changepoint position in time and cross-correlation values. In mid 2011, around event 300, and early 2012, around event 700, we have two changepoints for which cross-correlation seems stable for both P and S waves. At the beginning of 2012, when the seismic network was redefined, the cross-correlation for S waves changes dramatically, whereas the change in cross-correlation for P waves is less evident, and no changepoint is found at all. Our results seem to indicate that variations in cross-correlation coefficients (for example, computed for repeating earthquakes) could indicate unrealistic variations in elasticity and could be a problematic choice for a monitoring system of the subsurface.

We also compare the position of the retrieved changepoints with the seismicity rate, another parameter usually associated with time variations of the subsurface properties (e.g. Dou et al., 2018). In Figure 7, we report the seismicity rate every two
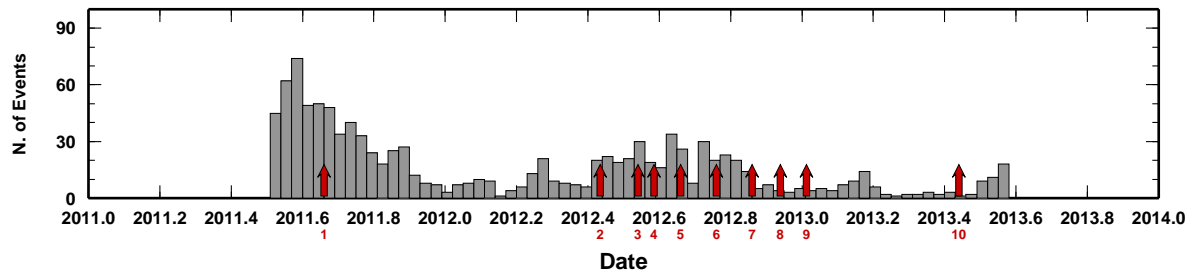
Figure 7. Comparison between the seismicity rate (grey histogram) and the time-occurrences for each changepoint (red arrows and numbers).

weeks. The rate of events is highly variable along the studied time window, with values ranging between a few and more
305   than 30 events/week. The seismicity rate decreases in 2012, with some bursts up to 15 events/week in late 2012. As for the
cross-correlation coefficients, the position of the retrieved changepoints does not simply correlate with the time-history of the
seismicity rate. We have found changepoints for which the seismicity rate is very high (2011.6) and very low (e.g. 2013.4). The
most probable changepoint (changepoint number 3) occurs in a period of sustained seismicity rate that starts 5-6 weeks before.
If our changepoints indicate variations of subsurface elasticity, the time-history of seismicity rate should be carefully evaluated
310   before using it for tracking elasticity changes in time. Nevertheless, we point out that the differences in the time-evolution of
the two indicators could be attributed to other geophysical processes, such as stress variations which does not strictly imply
elasticity variations, and we suggest that an integrated analysis could be necessary.

4   Discussion

The DD data recorded on Katla volcano and the results presented here clearly indicate that time variations in elastic properties
315   occurred between 2011 and 2014 on the south flank of the volcanic edifice. Thus, data-driven time windows can be found using
our approach to define when to apply standard DD analysis for retrieving elasticity variations, with no need to prepare the DD
data following subjective choices on the paired events. With the algorithm being naturally parsimonious, there is almost no
possibility of having "no DD data" (i.e. one partition per event). Changepoints are always limited in number, even if, strictly
speaking, their number should be given by the user because, as final output, we have the full PPD and not just one set of
320   best-fit changepoints. Defining the exact number of changepoints to use in subsequent analysis depends on the analysis itself.
Our approach quantifies the presence and the relevance of the changepoints. Using such information could be straightforward
in some cases (e.g. if we look for one most probable changepoint only) or more complex (e.g. if we also wish to appreciate
correlation between changepoints, which can be measured using the PPD, such as the occurrenc of a changepoint in time with
respect to the occurrence of another changepoint). It is worth noticing that, in simple cases, our algorithm generally performs
325   as expert opinion (e.g. in the case of the search for one most probable changepoint), and this confirms the overall performance

of our methodology. In more complex cases, the weights associated with the changepoints should be used to classify the changepoints themselves, and this allows selecting the most relevant changepoints using quantified information.

The network of seismic stations deployed around Katla volcano changed in time. This fact has been previously indicated as a potential "bias" in the analysis of the seismic data themselves, as the location uncertainties increased after major network operations (January 2012). Our results point out that the changepoints found do not correlate with such change in the seismic network. Being a statistical analysis, our methodology seems to be insensitive to changes in the acquisition system. Alternatively, the changes in location uncertainties may not be large enough to affect our procedure. In both cases, our approach appears to be well-suited for handling long-lived databases, in which changes in the spatial distribution of observational points is likely to occur from time to time.

Finally, we investigate how our changepoints relate to more commonly-used indicators of sub-surface variations of elasticity, such as time-series of cross-correlation coefficients and seismicity rate. In both cases, we found poor correlation between our results and the time-series of the two quantities. Although this observation is not totally unexpected since the two time-series are based on different seismological observables, it suggests that care should taken when investigating time-variations of elasticity retrieved from methodologies based on cross-correlation, and to re-asses approaches based on variations of the seismicity rate as a proxy of "rock instabilities" (Dou et al., 2018).

5   Conclusions

We developed an algorithm for defining data-driven partitions in a seismic database, for a more objective definition of double-difference data. The algorithm is based on the trans-dimensional sampling of data-structures, here represented as partitions of the covariance matrix. The algorithm has been tested in the case of a seismic database acquired in a volcanic setting, where subsurface variations of rock elasticity have probably occurred over a period of two years. Our results indicate that:

1. trans-dimesional algorithms can be efficiently used to map data-structures in the case of double-difference data, namely separating events with a number of changepoints that define time windows consistent with the underlying hypothesis (here a constant-in-time elasticity field between station and event cluster);

2. changepoints are quantitatively defined and, thus, can be ranked based on their relevance (i.e. weights) and probability of occurrence at a given time;

3. the results obtained are insensitive to changes in network geometry during the seismic experiment.

The applications of our algorithm are not limited to the standard preparation of DD data, but could be adopted for a more complex workflow, such as separting time windows for demeaning of DD data (Roecker et al., 2021). Future development and testing will provide additional insights into the use of trans-dimensional algorithms for the exploration of the data-space. For example, in this specific case, our algorithm can be applied to the joint inversion of both P-wave and S-wave databases following the approach described in Piana Agostinetti and Bodin (2018), to reconstruct a set of changepoints based on P-wave data and a set of changepoints based on S-wave data. Comparing the two sets of changepoints, "decoupled changepoints" (i.e.

changepoints occurring for one set of waves and not for the other) would properly map out elasticity variations, resolving the trade-off (still existent now) between elasticity changes and changes in event locations. In fact, variations in event location would be indicated by "coupled changepoints", i.e., changepoints occuring in both sets (Piana Agostinetti and Bodin, 2018).

Data availability. Waveform data used in this study come from the Icelandic permanent seismic network run by the Icelandic Meteorological Office (IMO). The data are available upon request to IMO.

Competing interests. The authors declare that they have no conflict of interest.

Author contributions. N.P.A. planned and conduced the data analysis. G.S. provided the seismic data and the relevant data pre-processing. N.P.A. and G.S. equally contributed to the discussion of the results and the implication for monitoring the subsurface. N.P.A. and G.S. wrote the draft of the manuscript and drew the relevant figures.

Appendix A: Synthetic tests

We perform two simple synthetic tests to illustrate the "parsimonious" behaviour of our trans-dimensional approach. In a first test, we make use of synthetic DD data created without imposing any changepoint in the data. Basically, the data are distributed according to the uncertainty statistics introduced in Equation 6. Moreover, we also add a random shift to the origin time (OT) of the events, to closely mimic the field measurements. OT shifts are randomly picked from a uniform distribution between ± 0.02 s. Synthetic data used as input can be seen in Figure A1a. In the second test, we introduce three changepoints in the same DD data used in the previous test. The three changepoints introduce a DD value as large as ± 0.04 s for some time windows, affecting some of the event pairs, which is overimposed to the OT shift. Synthetic data used as input for this second test can be seen in Figure A2a. The time-occurrence of the three changepoints is easily recognized in Figure A2a, however the presence of the OT shifts introduces an additional source of noise which seems to create fake changepoints.

In Figure A1, we report the result of the first test. The parsimonious behaviour of our trans-dimensional approach limits the number of chagepoints to the minimum (1). Such changepoint appears to be at the very beginning of our time-series of DD data (Figure A1e), where few event pairs are affected by its presence. Hence, we suggest that the occurence of changepoints close to the boundary of the time-series should be carefully evaluated, due to the limited information on which their presence is based on.

In Figure A2, we report the result obtained during the second test. Here, the three changepoints are clearly retrieved in the mid of our time-series (Figure A2e). However, also in this case, few additional changepoints are inserted at the beginning and at the end of the time-series, reinforcing the suggestion of the first synthetic test to carefully evaluate changepoints which affect a limited number of data (in our case, changepoints occuring at the very beginning of the time interval considered here).
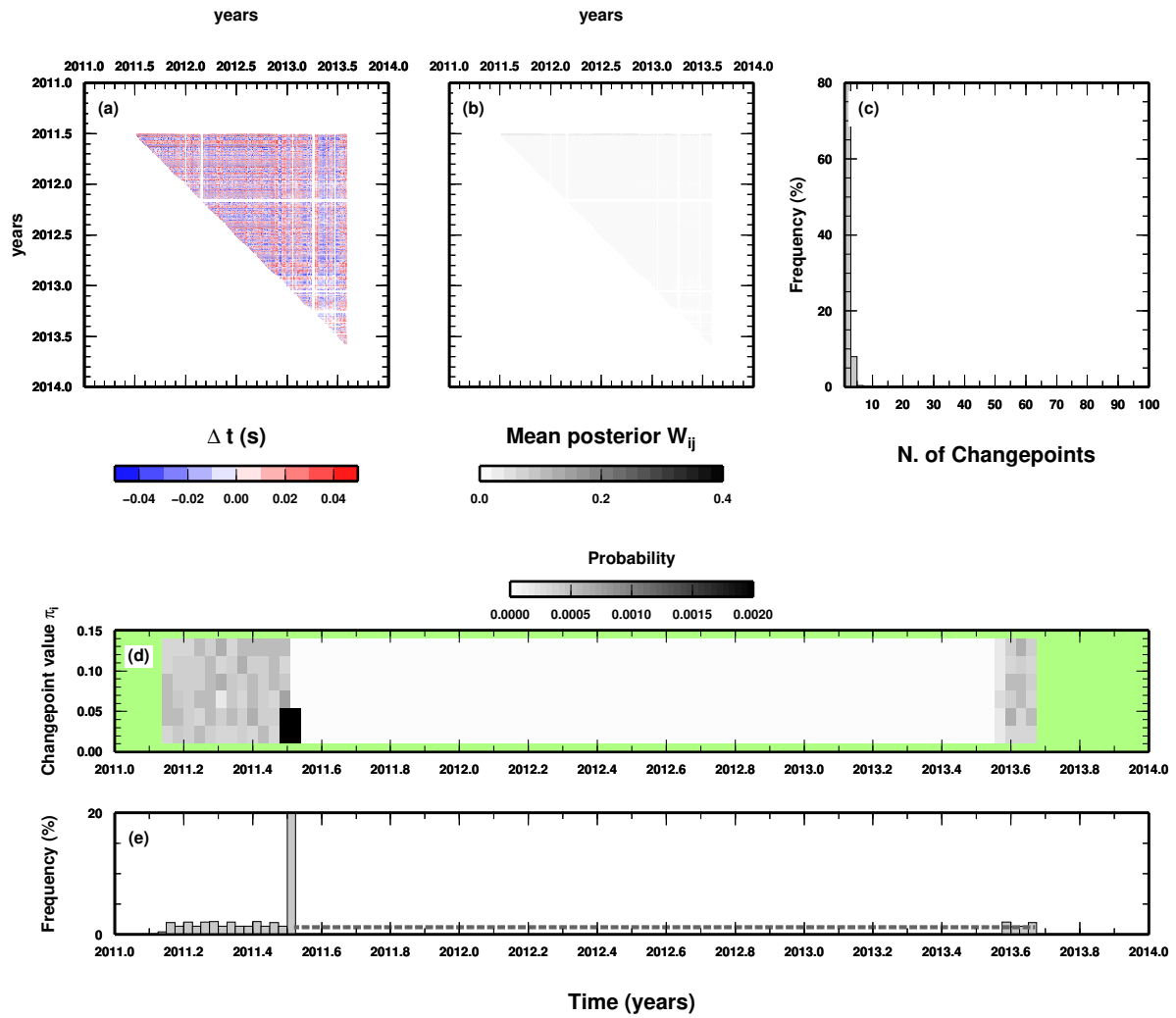
Figure A1. Results of the synthetic test where "observed" data have been created without changepoints. Figure details are the same as in Figure 5. In panel (a) we show the synthetic data used as input to the algorithm.
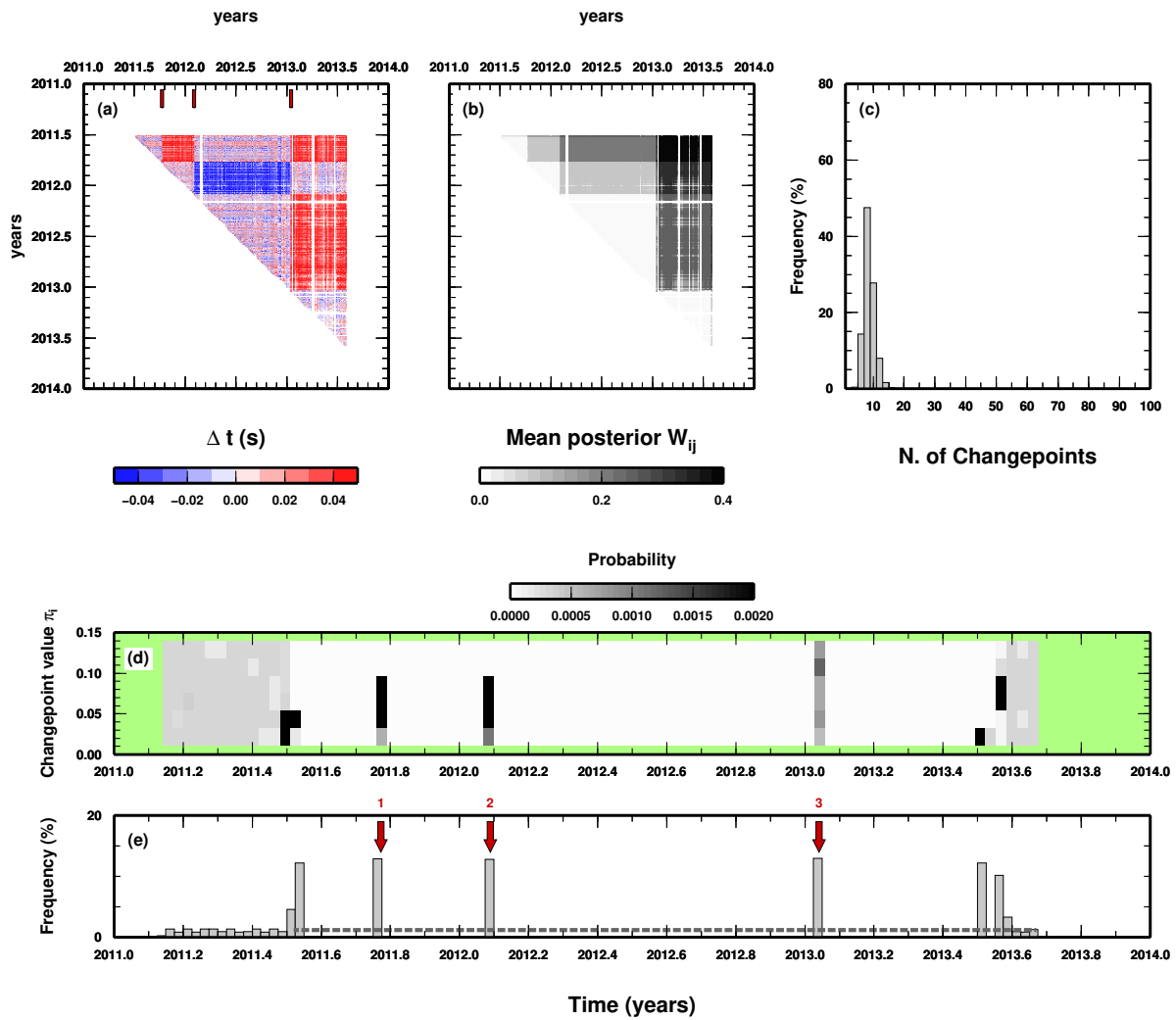
Figure A2. Results of the synthetic test where "observed" data have been created with three changepoints. Figure details are the same as in Figure 5. In panel (a) we show the synthetic data used as input to the algorithm.

Appendix B: Data prepartion

We use data from a cluster of repeating earthquakes located on the southern flank of Katla volcano (Iceland; Figure B1a). This seismic activity initiated in July 2011 following an unrest episode of the volcano (Sgattoni et al., 2017) and continued for several years with remarkably similar waveform features over time. The cluster is located at very shallow depth (< 1 km) and consists of small magnitude events (~ -0.5 - 1.2 ML) characterized by emergent P waves and unclear S waves, a narrow-band frequency content around 3 Hz at most stations, and correlation coefficients well above 0.9 at the nearest stations during the entire sequence. The temporal behavior is also peculiar, with a regular event rate of about 6 events per day during warm seasons gradually decreasing to one event every 1-2 days during cold seasons (Sgattoni et al., 2016b). Sgattoni et al. (2016a) obtained relative locations of 1141 events recorded between July 2011 and July 2013 by designing a method optimized for very small clusters that includes the effects of 3D heterogeneities and tracks uncertainties throughout the calculation. The number of relocated events depends on a selection of the best events among a total of > 1800, based on thresholds on correlation coefficient and amount of detected P and S phases. The resulting size of the cluster is on the order of $25 \times 50 \times 100$ m^3 (easting, northing, depth), with estimated uncertainties on the order of few tens of meters (Figure B1b). Changes in the station network configuration around the cluster occurred due to technical problems, with the greater loss of data in the second part of the sequence, after January 2012. This coincides with a clear increase in relative location uncertainties, which correlates also with a decrease in correlation coefficients, mainly for S phases. Other temporal changes in waveform correlation were identified by Sgattoni et al. (2016a) in August 2012 and January 2013. In this study we focus on P-wave data recorded at station ALF (part of the Icelandic Meteorological Office seismic network), which is located about 6 km away from the cluster (Figure B1a) and is the only nearby station that has been continuously operating during the entire time. The similarity of the waveforms recorded at ALF is remarkable, with correlation coefficients of the largest events above 0.99 throughout the entire period of study (Figure B1c). To compute the DD data set, we use the origin times (OT$_i$) of $N_e = 1119$ relocated events from Sgattoni et al. (2016a). We remark that the increased location uncertainties due to the network geometry change in January 2012 may affect the quality of the locations of the events and, consequently, the determination of their origin times, which is relevant for computing uncertainties in the DD data (see Section B1).

B1 Data uncertainties from full-waveform investigation

To apply our novel Bayesian approach, we need to estimate a covariance matrix of the uncertainties in the DD data. Having an origin time OT_i for the i-th event (given by the location obtained in Sgattoni et al. (2016a) using the full seismic network), we derive the DD data and their uncertainties directly comparing the raw waveforms and finding the absolute delays between each pair of events. From the absolute delays of the P arrivals, the subtraction of the time differences in the OTs of two events gives the DD datum for each pair. We estimate the absolute time-delay between two events following the Bayesian approach described in Piana Agostinetti and Martini (2019). Briefly, (a) we collected a 20-s record of each event $w_i(t)$ with $i = 1, ..., N_e$, centred on the approximate P-wave arrival time; (b) we stack all the waveforms of the events and obtain the so called "beam waveform" $b(t) = \frac{1}{N_e} \sum_i w_i(t)$ (Figure B2a); (c) from the beam waveform, we compute the residuals for each
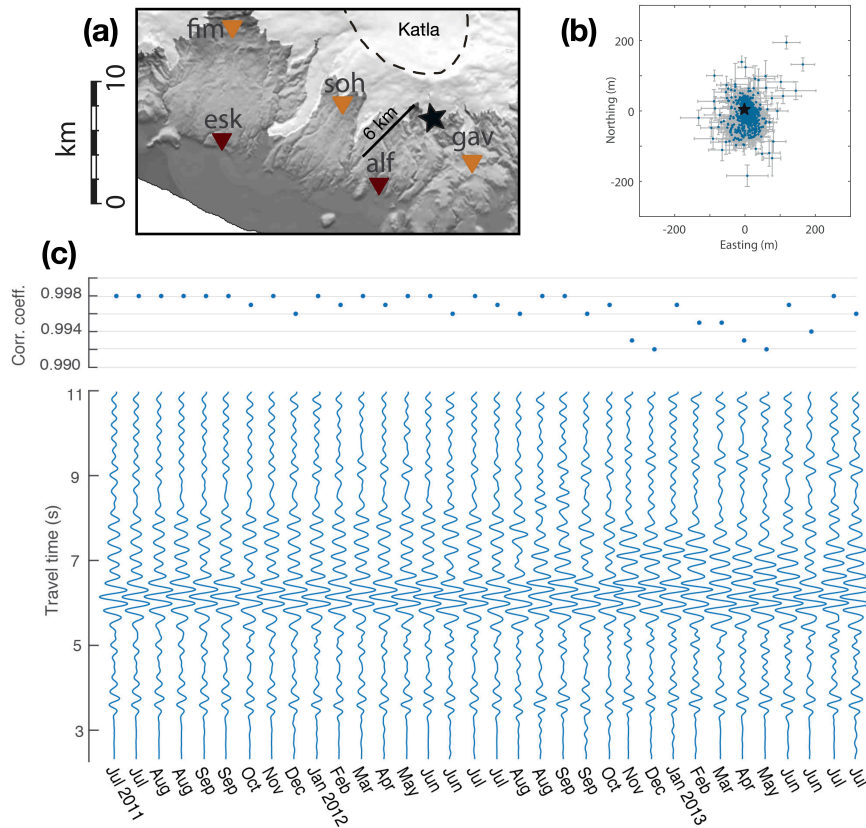
Figure B1. (a) Map of the southern flank of Katla volcano (Iceland; topography information from the National Land Survey of Iceland). The caldera rim is outlined by the black dashed line. White areas are glaciers. The star marks the location of the seismic cluster. Dark brown triangles: permanent Icelandic Meteorological Office (IMO) seismic stations. Orange triangles: temporary Uppsala University seismic stations operating between May 2011 and August 2013. (b) Relative locations (blue points) and uncertainties ($\pm$ std; grey lines) from Sgattoni et al. (2016a) (c) Example waveforms of the Z component recorded at station ALF throughout the entire period investigated and correlation coefficients of the P waves with respect to the master event used for the relative locations shown in (b). Panels (a) and (b) have been modified from Sgattoni et al. (2016a). Panel (c) has been modified from Sgattoni et al. (2016b).

event $r_i(t) = b(t) - w_i(t)$ (Figure B2b); (d) event residuals define a standard deviation function $\sigma(t) = \frac{1}{N_e}\sqrt{\sum_i r_i(t)^2}$ for the 20s record; (e) event residuals are also auto-correlated to obtain an averaged auto-correlation function $c(t)$. The standard deviation and the auto-correlation function are used to define a covariance matrix $\mathbf{\Sigma_{e,w}}$ (the same for all the waveforms, Piana Agostinetti and Malinverno, 2018) using the equation:

$$\mathbf{\Sigma_{e,w}} = \mathbf{S\,R\,S}\,, \tag{B1}$$

where $\mathbf{\Sigma_{e,w}}$ is the covariance matrix of the waveform uncertanties; $\mathbf{S}$ is a diagonal matrix containing the standard deviation $\sigma(t)$ computed from the residuals; and $\mathbf{R}$ is a symmetric Toeplitz matrix whose rows and columns contain the auto-correlation function $c(t)$ with $t = 0$ on the diagonal. In this way, we reconstruct a full covariance matrix, which can be used to obtain realistic uncertainty estimates for our DD data. Noteworthy, the use of a diagonal covariance matrix instead of the full $\Sigma_{e,w}$ covariance matrix would risk to underestimate the uncertainties, biasing the subsequent analysis for defining the DD time windows. Having the uncertainties model for the waveforms for each pair of waveforms associated with i-th and j-th events we perform a Markov chain Monte Carlo sampling (Mosegaard and Tarantola, 1995) to reconstruct the PPD of the time-shift t_{ij} between the two waveforms. Following Sgattoni et al. (2016a), we use a 1s-long time window to compute the likelihood of the waveforms, centred on the approximate P-wave arrival time. From the reconstructed PPD, we use the mean posterior value of the time-shift $\tilde{t}_{ij}$, its standard deviation $\tilde{\sigma}_{ij}$ and the origin times OT_i to produce the DD data, namely: $d_{ij} = \tilde{t}_{ij} - (OT_j - OT_i)$ and $\sigma_{ij} = \tilde{\sigma}_{ij}$, with $i = 1, ..., N_e - 1$, $j = i + 1, ..., N_e$ and $j > i$.

Acknowledgements. We thanks Cliff Thurber, Jiaqi Li and an anomymous reviewer for their constructive comments on the first verion of the manuscript. This study has been published with the financial support of the Austrian Science Fund (FWF), project number: M2218-N29. NPA and GS thank the Icelandic Meteorological Office for providing the original seismic data. The Generic Mapping Tools software has been used for plotting the figures of this manuscript (Wessel and Smith, 1998).
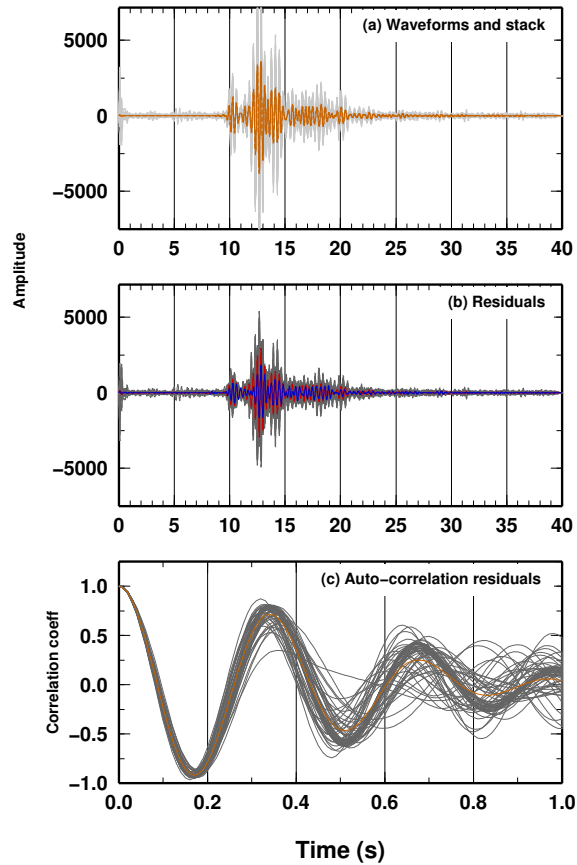
Figure B2. (a) Original waveforms $w_i(t)$ (grey lines) and their stack, called "beam waveform" $b(t)$ (orange line). (b) Residual of each single waveform $r_i(t)$ (grey lines) with respect to the "beam waveform". As a reference, the residual for the first (last) trace is shown as a blue (red) line. (c) Auto-correlation of the single residuals (grey line) and averaged value of the autocorrelation $c(t)$ (orange line).

References

440   Bayes, T.: An essay towards solving a problem in the doctrine of chances, Philos. Trans. R Soc. London,, 53, 370–418, 1763.

Berardino, P., Fornaro, G., Lanari, R., and Sansosti, E.: A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms, IEEE Transactions on Geoscience and Remote Sensing, 40, 2375–2383, https://doi.org/10.1109/TGRS.2002.803792, 2002.

Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P.: Transdimensional tomography with unknown data noise, Geophys. J. Int, doi:

445   10.1111/j.1365-246X.2012.05414.x, 2012a.

Bodin, T., Sambridge, M., Tkalcic, H., Arroucau, P., Gallagher, K., and Rawlinson, N.: Transdimensional inversion of receiver functions and surface wave dispersion, J. Geophys. Res., 117, doi:10.1029/2011JB008560, 2012b.

Caló, M. and Dorbath, C.: Different behaviours of the seismic velocity field at Soultz-sous-Forets revealed by 4-D seismic tomography: case study of GPK3 and GPK2 injection tests, Geophysical Journal International, 194, 1119–1137, https://doi.org/10.1093/gji/ggt153, 2013.

Caló, M., Dorbath, C., Cornet, F., and Cuenot, N.: Large-scale aseismic motion identified through 4-D P-wave tomography, Geophysical Journal International, 186, 1295–1314, https://doi.org/10.1111/j.1365-246X.2011.05108.x, 2011.

Chiarabba, C., De Gori, P., Segou, M., and Cattaneo, M.: Seismic velocity precursors to the 2016 Mw 6.5 Norcia (Italy) earthquake, Geology, 48, 924–928, doi: https://doi.org/10.1130/G47048.1, 2020.

Dettmer, J. and Dosso, S. E.: Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains, J. Acoust. Soc. Am., 132, 2239–2250, 2012.

Dettmer, J., Benavente, R., Cummins, P. R., and Sambridge, M.: Trans-dimensional finite-fault inversion, Geophysical Journal International, 199, 735–751, https://doi.org/10.1093/gji/ggu280, 2014.

Dou, L., Cai, W., Cao, A., and Guo, W.: Comprehensive early warning of rock burst utilizing microseismic multi-parameter indices, International Journal of Mining Science and Technology, 28, 767–774, https://doi.org/https://doi.org/10.1016/j.ijmst.2018.08.007, sI: Dynamic failure in rock masses, 2018.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D., and Nicolson, H.: 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, 36–58, https://doi.org/10.1093/gji/ggw286, 2016.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M., and Stephenson, J.: Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, Marine and Petroleum Geology, 26, 525–535, 2009.

Gelman, A., Roberts, G. O., and Gilks, W. R.: Efficient Metropolis jumping rules, pp. 599–607, Oxford University Press, 1996.

Guo, H. and Zhang, H.: Development of double-pair double difference earthquake location algorithm for improving earthquake locations, Geophysical Journal International, 208, 333–348, https://doi.org/10.1093/gji/ggw397, 2016.

Kerr, J.: Applications of Double-Difference Tomography for a Deep Hard Rock Mine, Master's thesis, Virginia Polytechnic Institute and State University, https://vtechworks.lib.vt.edu/bitstream/handle/10919/35850/Kerr_JB_T_2011.pdf, 2011.

Lin, Y. and Huang, L.: Quantifying subsurface geophysical properties changes using double-difference seismic-waveform inversion with a modified total-variation regularization scheme, Geophysical Journal International, 203, 2125–2149, https://doi.org/10.1093/gji/ggv429, 2015.

Lohman, R. B. and Simons, M.: Some thoughts on the use of InSAR data to constrain models of surface deformation: Noise structure and data downsampling, Geochemistry, Geophysics, Geosystems, 6, https://doi.org/https://doi.org/10.1029/2004GC000841, 2005.

Luxbacher, K., Westman, E., Swanson, P., and Karfakis, M.: Three-dimensional time-lapse velocity tomography of an underground longwall panel, International Journal of Rock Mechanics and Mining Sciences, 45, 478 – 485, https://doi.org/https://doi.org/10.1016/j.ijrmms.2007.07.015, 2008.

Ma, X., Westman, E., and Counter, D.: Passive Seismic Imaging of Stress Evolution with Mining-Induced Seismicity at Hard-Rock Deep Mines., Rock Mech Rock Eng, 53, 2789â<80><93>2804, https://doi.org/10.1007/s00603-020-02076-5, 2020.

Malinverno, A.: Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, Geophys. J. Int., 151, 675–688, 2002.

Malinverno, A. and Briggs, V. A.: Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, Geophysics, 69, 1005–1016, doi:10.1190/1.1778243, 2004.

Mandolesi, E., Ogaya, X., Campanya, J., and Piana Agostinetti, N.: A reversible-jump Markov chain Monte Carlo algorithm for 1D inversion of magnetotelluric data, Computers and Geosciences, 113, 94 – 105, https://doi.org/10.1016/j.cageo.2018.01.011, 2018.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, N. M., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, J. Chem. Phys., 1, 1087–1092, 1953.

490 Mosegaard, K. and Sambridge, M.: Monte Carlo analysis of inverse problems, Inverse Problems, 18, R29–R54, 2002.

Mosegaard, K. and Tarantola, A.: Monte Carlo sampling of solutions to inverse problems, J. Geophys. Res., 100, 12 431–12 447, 1995.

Patanè, D., Barberi, G., Cocina, O., De Gori, P., and Chiarabba, C.: Time-Resolved Seismic Tomography Detects Magma Intrusions at Mount Etna, Science, 313, 821–823, https://doi.org/10.1126/science.1127724, 2006.

Piana Agostinetti, N. and Bodin, T.: Flexible Coupling in Joint Inversions: A Bayesian Structure Decoupling Algorithm, Journal of Geophys-
495 ical Research: Solid Earth, 123, 8798–8826, https://doi.org/10.1029/2018JB016079, 2018.

Piana Agostinetti, N. and Malinverno, A.: Receiver Function inversion by trans-dimensional Monte Carlo sampling, Geophys. J. Int, 181, doi:10.1111/j.1365-246X.2010.04530.x, 2010.

Piana Agostinetti, N. and Malinverno, A.: Assessing uncertainties in high-resolution, multi-frequency receiver function inversion: a compar-
ison with borehole data, Geophysics, 83, KS11–KS22, doi: 10.1190/geo2017-0350.1, 2018.

500 Piana Agostinetti, N. and Martini, F.: Sedimentary basins investigation using teleseismic P-wave time delays, Geophysical Prospecting, 67, 1676–1685, https://doi.org/10.1111/1365-2478.12747, 2019.

Poggiali, G., Chiaraluce, L., Di Stefano, R., and Piana Agostinetti, N.: Change-point analysis of V_P/V_S ratio time-series using a trans-
dimensional McMC algorithm: applied to the Alto Tiberina Near Fault Observatory seismic network (Northern Apennines, Italy), Geo-
physical Journal International, 217, 1217–1231, https://doi.org/10.1093/gji/ggz078, 2019.

505 Qian, J., Zhang, H., and Westman, E.: New time-lapse seismic tomographic scheme based on double-difference tomography and its applica-
tion in monitoring temporal velocity variations caused by underground coal mining, Geophysical Journal International, 215, 2093–2104, https://doi.org/10.1093/gji/ggy404, 2018.

Roecker, S., Maharaj, A., Meyers, S., and Comte, D.: Double Differencing by Demeaning: Applications to Hypocenter Location and Wavespeed Tomography, Bulletin of the Seismological Society of America, 111, 1234–1247, https://doi.org/10.1785/0120210007, 2021.

510 Sambridge, M. and Mosegaard, K.: Monte Carlo methods in geophysical inverse problems, Rev. Geophys., 40, doi:10.1029/2000RG000 089, 2002.

Sambridge, M., Gallagher, K., Jackson, A., and Rickwood, P.: Trans-dimensional inverse problems, model comparison and the evidence, Geophys. J. Int., 167, 528–542, doi:10.1111/j.1365-246X.2006.03155.x., 2006.

Sgattoni, G., Gudmundsson, O., Einarsson, P., and Lucchi, F.: Joint relative location of earthquakes without a pre-defined velocity model:
515 an example from a peculiar seismic cluster on Katla volcano's south-flank (Iceland), Geophysical Journal International, 207, 1244–1257, https://doi.org/10.1093/gji/ggw331, 2016a.

Sgattoni, G., Jeddi, Z., Gudmundsson, O., Einarsson, P., Tryggvason, A., Lund, B., and Lucchi, F.: Long-period seismic events with strikingly regular temporal patterns on Katla volcano's south flank (Iceland), J. Volcanol. Geotherm. Res., 324, 28–40, doi:10.1016/j.jvolgeores.2016.05.017, 2016b.

520 Sgattoni, G., Gudmundsson, O., Einarsson, P., Lucchi, F., Li, K., Sadeghisorkhani, H., Roberts, R., and Tryggvason, A.: The 2011 unrest at Katla volcano: characterization and interpretation of the tremor source, J. Volcanol. Geotherm. Res., 338, 63–78, doi:10.1016/j.jvolgeores.2017.03.028, 2017.

Steininger, G., Dettmer, J., Dosso, J., and Holland, S.: Transdimensional joint inversion of seabed scattering and reflection data, J. acoust. Soc. Am., 133, 1347–1357, 2013.

525 Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, 2005.

Tilmann, F. J., Sadeghisorkhani, H., and Mauerberger, A.: Another look at the treatment of data uncertainty in Markov chain Monte Carlo inversion and other probabilistic methods, Geophysical Journal International, 222, 388–405, https://doi.org/10.1093/gji/ggaa168, 2020.

Waldhauser, F. and Ellsworth, W.: A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, California, Bulletin of the Seismological Society of America, 90, 1353–1368, 2000.

530 Wessel, P. and Smith, W. H. F.: New, improved version of the Generic Mapping Tools released, EOS Trans. AGU, 79, 579, 1998.

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

Yin, Y. and Pillet, W.: Seismic data preparation for improved elastic inversion of angle stacks, pp. 2042–2046, https://doi.org/10.1190/1.2369937, 2006.

535 Young, R. P. and Maxwell, S. C.: Seismic characterization of a highly stressed rock mass using tomographic imaging and induced seismicity, Journal of Geophysical Research: Solid Earth, 97, 12 361–12 373, https://doi.org/10.1029/92JB00678, 1992.

Zhang, H. and Thurber, C. H.: Double-difference tomography; the method and its application to the Hayward Fault, California, Bull. Seism. Soc. Am., 93, 1875–1889, 2003.

Zhang, H., Sarkar, S., Toksoz, M. N., Kuleli, S., and Al-Kindy, F.: Passive seismic tomography using induced seismicity at a petroleum field
540 in Oman, Geophysics, 74, 2006.

Zhang, X. and Zhang, H.: Wavelet-based time-dependent travel time tomography method and its application in imaging the Etna volcano in Italy, J. Geophys. Res. Solid Earth, 120, 7068–7084, doi:10.1002/2015JB012182, 2015.