Geophysical Journal International

Geophys. J. Int. (2023) **232,** 1219–1235
Advance Access publication 2022 October 06
GJI Seismology

https://doi.org/10.1093/gji/ggac385

Towards fast machine-learning-assisted Bayesian posterior inference of microseismic event location and source mechanism

D. Piras [1], A. Spurio Mancini [1,2,3], A. M. G. Ferreira,[4] B. Joachimi[1] and M. P. Hobson[3]

[1] *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK. E-mail: dr.davide.piras@gmail.com*
[2] *Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK*
[3] *Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK*
[4] *Deptartment of Earth Sciences, Faculty of Mathematical & Physical Sciences, University College London, London WC1E 6BT, UK*

Accepted 2022 October 3. Received 2022 August 8; in original form 2021 January 14

SUMMARY

Bayesian inference applied to microseismic activity monitoring allows the accurate location of microseismic events from recorded seismograms and the estimation of the associated uncertainties. However, the forward modelling of these microseismic events, which is necessary to perform Bayesian source inversion, can be prohibitively expensive in terms of computational resources. A viable solution is to train a surrogate model based on machine learning techniques to emulate the forward model and thus accelerate Bayesian inference. In this paper, we substantially enhance previous work, which considered only sources with isotropic moment tensors. We train a machine learning algorithm on the power spectrum of the recorded pressure wave and show that the trained emulator allows complete and fast event locations for *any* source mechanism. Moreover, we show that our approach is computationally inexpensive, as it can be run in less than 1 hr on a commercial laptop, while yielding accurate results using less than 10^4 training seismograms. We additionally demonstrate how the trained emulators can be used to identify the source mechanism through the estimation of the Bayesian evidence. Finally, we demonstrate that our approach is robust to real noise as measured in field data. This work lays the foundations for efficient, accurate future joint determinations of event location and moment tensor, and associated uncertainties, which are ultimately key for accurately characterizing human-induced and natural earthquakes, and for enhanced quantitative seismic hazard assessments.

Key words: Machine learning; Statistical methods; Induced seismicity; Waveform inversion.

1 INTRODUCTION

Underground human activity, including fluid injection in rocks and mining operations, can cause microseismic events (Majer *et al.* 2007; Ellsworth 2013). The monitoring of both human-induced and natural microseismicity is critical for understanding seismic hazard (Brueckl *et al.* 2008; Shapiro *et al.* 2010; Mukuhira *et al.* 2016; Das *et al.* 2017, and references therein). Accurate seismic event locations in space and in time are of paramount importance for reliable seismic monitoring efforts, and are mainly obtained from seismograms recorded on land and/or at the seafloor.

Various methods for locating seismic events are available in the literature, dating back to the work of Geiger (1910), and up to today (see e.g. Vasco *et al.* 2019, and references therein, for a recent review). One of the most common approaches relies on using the Eikonal equation to determine the theoretical travel time of first seismic arrivals (see e.g. Noack & Clark 2017; Smith *et al.* 2020), which is compared with real travel times through a direct grid search or more sophisticated inverse modelling techniques (Wuestefeld *et al.* 2018). More accurate source location estimations can be obtained exploiting methods that use the full waveform, even though they generally require heavier computational resources (Song & Toksöz 2011; Li 2013; Angus *et al.* 2014; Li *et al.* 2016; Vasco *et al.* 2019; Willacy *et al.* 2019; see also Li *et al.* 2020 for a recent review of waveform-based inversion methods).

A Bayesian approach can also be adopted to solve the location inverse problem (Lomax *et al.* 2000; Tarantola 2005; Stähler & Sigloch 2014, 2016; Pugh *et al.* 2016). In this framework, a posterior distribution of the model parameters (i.e. the event's location and/or the moment tensor) is estimated and used to determine the optimal model parameters and their associated uncertainty. Markov chain Monte Carlo (MCMC; see e.g. Craiu & Rosenthal 2014, for a review) and nested sampling (Skilling 2006) techniques are among those employed to sample the posterior distribution. However, these approaches become prohibitive when dealing with a high number of parameters, or when the forward model is computationally expensive to simulate (see e.g. Rajaratnam & Sparks 2015; Conrad *et al.* 2016; Alsing *et al.* 2018). For these reasons, being able to cheaply

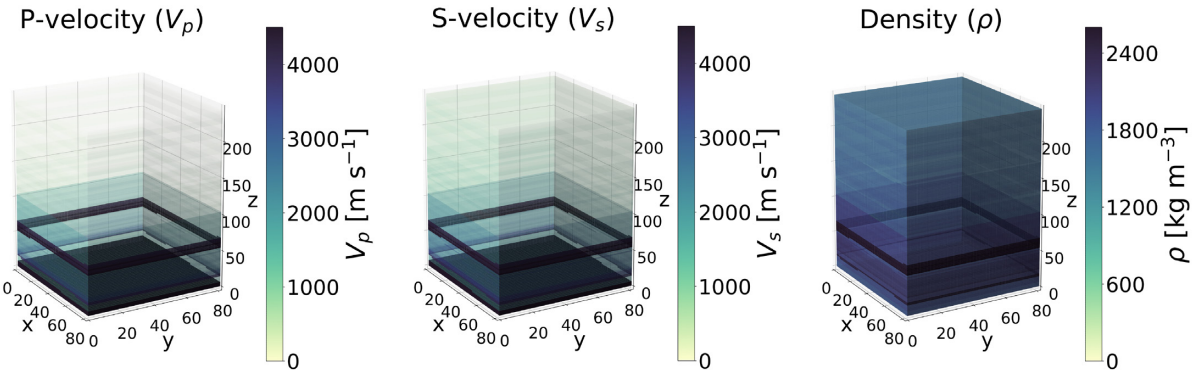

Figure 1. *P*-wave velocity (V_p), S-wave velocity (V_s) and density (ρ) models of the simulated domain we consider in this work. The models are specified as 3-D grids of voxels, with size 81 × 81 × 301 points, corresponding to a real geological model of size 1 km × 1 km × 3 km. We observe that our model has a layered structure, with variation along the vertical dimension more marked than along horizontal planes. The plots were adapted from fig. 1 in Spurio Mancini *et al.* (2021).

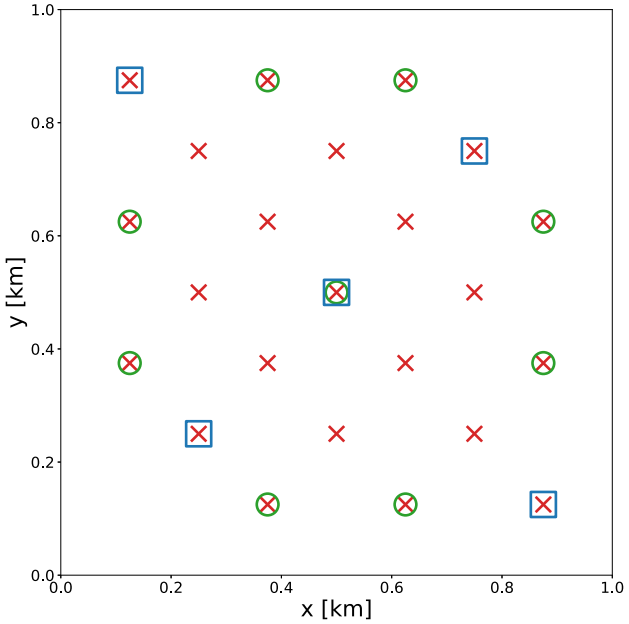

Figure 2. Projection of the positions of the 23 receivers on the *x*–*y* plane; $z = 2.43$ km corresponds to the seabed, where all sensors lie. Each receiver records the acoustic pressure wave and particle velocity generated by a microseismic event below the seabed. The red crosses indicate all 23 receivers, while green circles and blue squares (9 and 5 receivers, respectively) refer to subsets of receivers we used to test the robustness of our method, as shown in Section 4.3.

and accurately simulate the theoretical waveforms of microseismic events given their location has become paramount in recent years. The forward modelling of microseismic events typically involves solving the elastic wave equation in a wide frequency range given a 3-D heterogeneous density and velocity model for the propagating medium (Das *et al.* 2017), which can be prohibitively expensive.

Machine learning generative models have gained considerable attention in recent years, with applications to many fields ranging from computer vision (Goodfellow *et al.* 2014; Gulrajani *et al.* 2016) to astrophysics (Auld *et al.* 2007, 2008; Spurio Mancini *et al.* 2022), as well as climate science, nuclear physics and drug selection (see e.g. Chenthamarakshan *et al.* 2020; Kasim *et al.* 2020). These advances have been enabled by both an increased accessibility to computational resources, as well as by a significant growth in the amount of available data.

In seismology, machine learning techniques have been successfully applied to a wide range of problems (see e.g. Bergen *et al.* 2019, for a recent review). For instance, Zheng *et al.* (2018) used a recurrent neural network to pick the arrival times of microseismic (or acoustic) events, while Zhu & Beroza (2018), Ross *et al.* (2018) and Zhu *et al.* (2019) trained convolutional neural networks (CNNs) to measure *P*- and *S*-wave arrival times and determine high-level features with high precision, often outperforming the measurements performed manually or semi-automatically by expert seismologists. Mousavi *et al.* (2020) further proposed a deep-learning model for simultaneous earthquake detection and phase picking based on an attention mechanism, while Li *et al.* (2022) trained multiple CNNs both to detect whether there is an acoustic emission data event in some recorded data, and to pick the arrival time of the *P* wave.

Looking in particular at generative models, Das *et al.* (2017) developed an optimized approach to simulate microseismic event propagation that, for each event location and given a physical model for the propagating medium, produces the corresponding seismogram in $\mathcal{O}(1\text{ h})$ using a Tesla graphics processing unit (GPU) and the software K-WAVE (Treeby *et al.* 2014). Subsequently, Das *et al.* (2018) and Spurio Mancini *et al.* (2021) (D18 and SM20 hereafter, respectively) showed the limitations of this direct approach, and presented an alternative whereby the mapping is learnt using machine learning techniques. In particular, D18 showed how Gaussian processes (GPs; Rasmussen & Williams 2005) can be used to learn an accurate surrogate model, while SM20 demonstrated the effectiveness of a variety of machine learning algorithms as emulators, and showed how their surrogate model yields an accurate estimate of the posterior distribution of an event's location in a fraction of the time required by the D18 method.

D18 and SM20, however, only applied their methodologies to microseismic events with an isotropic source mechanism. It is well known that any source mechanism can be mathematically decomposed into three components: isotropic (ISO), double couple (DC) and compensated linear vector dipole (CLVD) (see e.g. Knopoff & Randall 1970; Vavryčuk 2001, 2005, 2015). The pure ISO source is associated with implosive or explosive forces, while the pure DC source is caused by shear faulting.

In this paper, we present an approach that aims at learning the direct mapping between event locations and seismograms for any microseismic source type (ISO, DC and CLVD). We show that it is

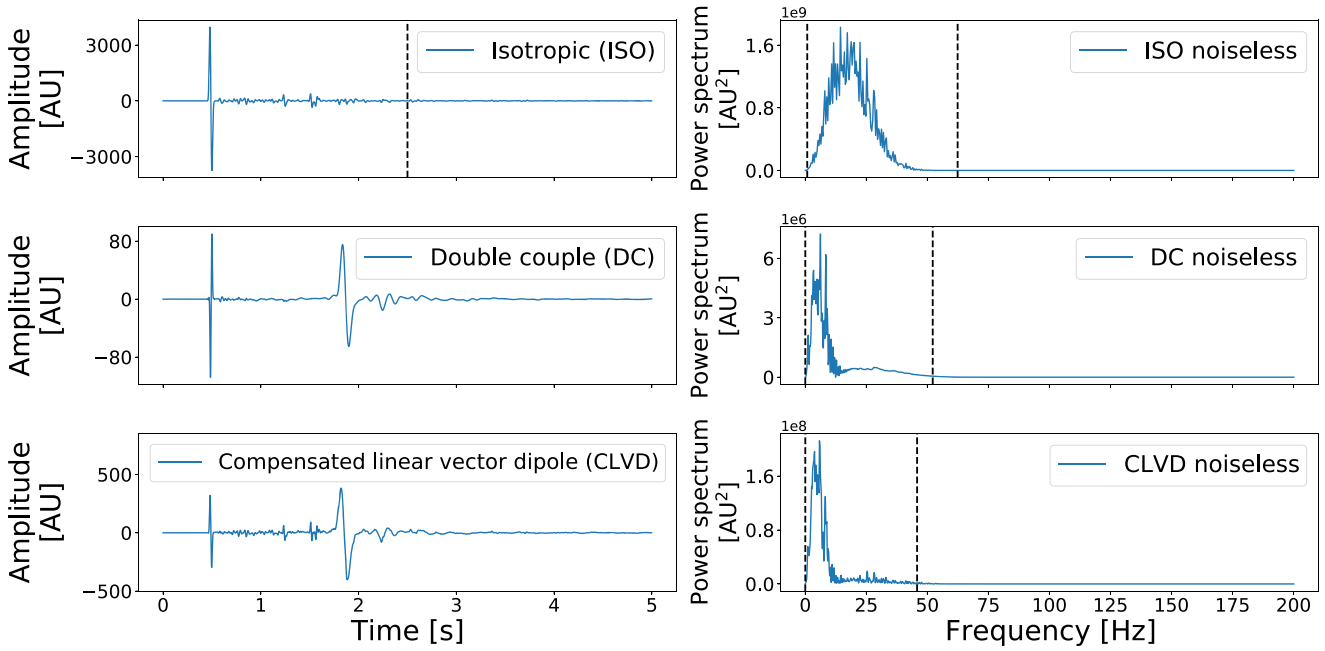

Figure 3. Left-hand column: example acoustic pressure wave for each different type of moment tensor: ISO, DC and CLVD. These seismograms correspond to a source location of $(x, y, z) = (0.55, 0.73, 1.8$ km) as recorded by a receiver in $(x, y, z) = (0.13$ km, 0.38 km, 2.43 km). The seismograms' amplitude is measured in arbitrary units of pressure. Note the different scales for each source mechanism. The vertical dashed black line in the top panel indicates a cut we perform for the isotropic sources only, based on the left-hand panel of Fig. 6. Right-hand column: the corresponding power spectra, calculated as described in Section 3.2. No noise is added when training the emulator, while some noise is introduced when doing inference, as described in Section 3.1. The vertical dashed lines indicate a frequency cut we perform to further reduce the number of features and to be robust to noise, based on the right-hand panel of Fig. 6.

sufficient to consider the power spectrum of the recorded pressure waves in order to distinguish different seismograms (as already investigated e.g. in Pratt 1999; Pratt & Shipp 1999; Tao & Sen 2013; Jakobsen & Ursin 2015), and we train a simple machine learning algorithm to learn this mapping efficiently. Moreover, we demonstrate how our method allows the accurate inference of the posterior distribution of the coordinates of a single source in $\mathcal{O}(0.1$ h) on a commercial laptop, thus paving the way for fast and computationally cheap joint determinations of event locations and seismic moment tensors. Finally, we show how we can use the trained emulators to identify the source mechanism through Bayesian evidence estimation, thus demonstrating the versatility of our Bayesian approach.

Regarding the structure of this paper, in Section 2, we describe the data we use in this work. In Section 3, we explain our inversion approach, describe what pre-processing steps we perform and recall the details of the generative method we employ. We show its performance at both training and inference time in Section 4, where we also compare our approach to standard arrival time analysis (Lomax *et al.* 2009) and include an experiment with real noise from Groningen field data (Smith 2019), thus showing the robustness of our method. Finally, in Section 5, we discuss our results and provide an outlook on possible extensions of this work.

2 DATA

In this work, we consider the same data framework as D18 and SM20, starting from 3-D heterogeneous density and velocity models for the propagating medium, which we show in Fig. 1. The model, which is discretized on a 3-D grid of voxels, specifies the values of the density ρ of the propagating medium, as well as the propagation velocities for P and S waves (V_p, V_s). We assume that sensors are placed at the seabed, and that they record both pressure and three-component particle velocity of the propagating medium (even though we will use only the former, as we explain later on). As anticipated, our aim is to apply our method to any source mechanism, so we will consider a more general generation procedure than previous work. Unlike D18 and SM20, who only considered isotropic sources, we take the microseismic moment tensor to be one of three types, which we denote as $\mathbf{M}_{\rm ISO}$, $\mathbf{M}_{\rm DC}$ and $\mathbf{M}_{\rm CLVD}$. Following Vavryčuk (2005) and Li *et al.* (2015), we define these quantities as

$$\mathbf{M}_{\rm ISO} = \begin{bmatrix} M_{11} & 0 & 0 \\ 0 & M_{22} & 0 \\ 0 & 0 & M_{33} \end{bmatrix}, \tag{1}$$

$$\mathbf{M}_{\rm DC} = \begin{bmatrix} 0 & M_{12} & 0 \\ M_{21} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{2}$$

$$\mathbf{M}_{\rm CLVD} = \begin{bmatrix} M_{11} & 0 & 0 \\ 0 & M_{22} & 0 \\ 0 & 0 & -2M_{33} \end{bmatrix}, \tag{3}$$

where each M_{ij} represents a different couple of forces. We additionally assume $M_{11} = M_{22} = M_{33} = M_{12} = M_{21} = 1$ MPa, which is a realistic assumption following Collettini & Barchi (2002); we anticipate that while in this work we fix the M_{ij} coefficients, we will explore the joint determination of moment tensor components and event coordinates in future work. In practice, ISO events are characterized by a single (explosive or implosive) P wave, while DC events are linked to shear stress and are characterized by both a P wave and an S wave, with comparable amplitudes. Similarly to ISO events, CLVD events display an often dominant principal wave, whose amplitude is however much smaller than the ISO one, and a non-negligible S wave.

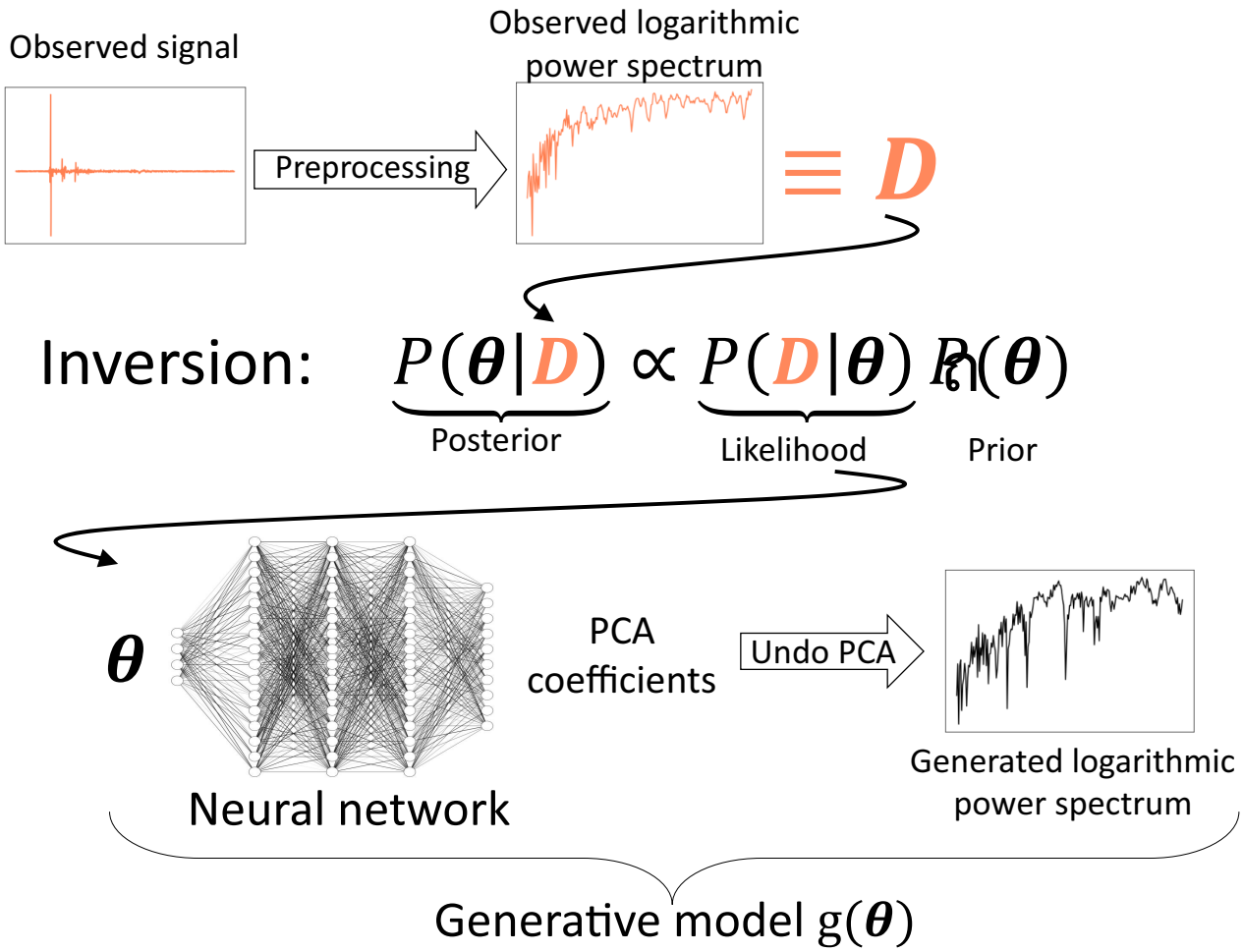

Figure 4. Workflow of our inversion approach. The observed event, indicated in coral in the top left corner, is pre-processed according to the steps described in Section 3.2 in order to obtain a logarithmic power spectrum. Our goal is to obtain the posterior distribution of the coordinates θ of the event's location beneath the surface, as detailed in Section 3.1. In order to sample from the posterior distribution, we specify a data likelihood, and use a generative model $g(\boldsymbol{\theta})$ to accelerate the evaluation of the likelihood. As our generative model, which we describe in Section 3.3, we employ a feedforward neural network that maps source coordinates to the principal components of the logarithmic power spectra. The neural network sketch has been drawn using NN-SVG (LeNail 2019). Note that the form of Bayes' theorem in this picture is a simplified version of eq. (4).

We generate forward simulations using the GPU implementation of Das *et al.* (2017), employing different types of GPUs with memory size ranging from 5 GB to 11 GB. The precise GPU cards used to generate the data are Tesla K20, Tesla C2075, GeForce RTX 2080 Ti and GeForce GTX 1080 Ti, which carry a different number of CUDA cores each, ranging from 500 to 5000. We observe that the speed of the generation scales linearly with the number of cores available; however, given that we produce the data on a shared cluster, we cannot always choose which card to employ. Note that the software with which we work is optimized for GPUs, and therefore cannot be expected to scale similarly when running on central processing units (CPUs) or tensor processing units (TPUs).

For each source type, we produce 10 000 events corresponding to different source locations, which are randomly sampled using Latin Hypercube Sampling on a 3-D grid of 81 × 81 × 301 points, corresponding to a real geological model of size 1 km × 1 km × 3 km. The total time to generate these data using the hardware specified above is about 150 hr for each source type. We consider 23 receivers in total, whose position is depicted in Fig. 2, even though we stress that we will not focus on finding the optimal geometry of the sensors in this work. Each receiver records each of the 10 000 events independently, thus producing a set of 10 000 waveforms for each source type at each receiver. In Section 4.3, we will present the results of a full analysis of the dependence of the posterior distribution on the number of training points, the number of receivers and the noise level, in order to demonstrate the robustness of our approach.

The time interval (sampling rate) for the solution of the elastic wave equation is 0.5 ms (2 kHz), and the total length of each seismic event is 5 s. After generation, all seismograms are downsampled to a time resolution of 2.5 ms (400 Hz) to reduce computational storage. In this way, each seismic trace is ultimately a time-series composed of $N_t = 2001$ time samples: we show an example for each source mechanism in the left-hand column of Fig. 3. Finally, note

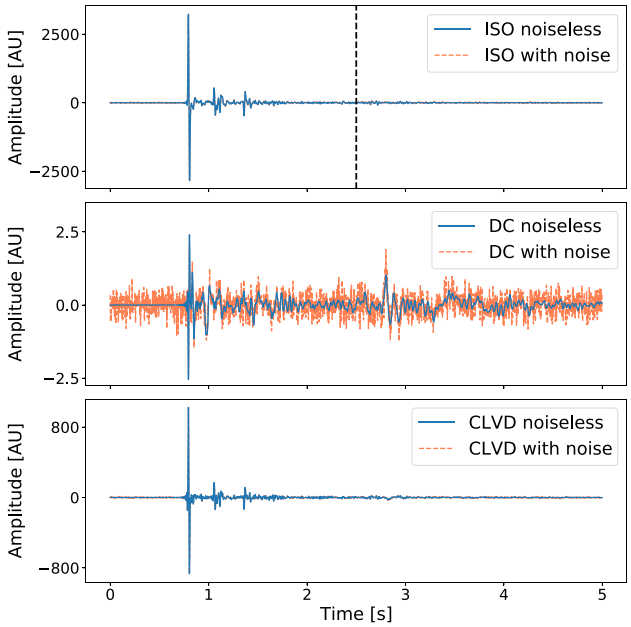

Figure 5. Comparison of the signal without (solid blue) and with (dashed coral) noise for each source mechanism: isotropic (ISO), double couple (DC) and compensated linear vector dipole (CLVD). As explained in Section 3.1, after training the emulator on the noiseless traces, we add Gaussian noise to the observed signal to infer its coordinates. In this case, the seismogram corresponds to event 3 in Table 1 as recorded by a receiver in $(x, y, z) =$ (0.13 km, 0.38 km, 2.43 km), and the signal-to-noise ratio (SNR) is 33 dB; we explore higher and lower SNR values in Section 4.3 and Fig. 11. The vertical dashed black line in the top panel indicates a cut we perform for the isotropic sources only, based on the left-hand panel of Fig. 6.

that we consider the seismograms to be noiseless at training time, while some noise is added to the simulated recorded event when performing inference on the coordinates' posterior distribution, as we are working in a Bayesian framework, detailed in Section 3.1. We will discuss and show the effect of the noise level in Sections 3.2 and 4.3, respectively.

3 INVERSION APPROACH

Our goal is to perform Bayesian inference on the source location of a microseismic event. In order to do so, we will employ a neural network (NN) as the forward model to parametrize the mapping between the coordinates and the principal components of the power spectra of the seismograms. Our inversion approach is summarized in Fig. 4.

3.1 Inference

We recall here the basic assumptions of our Bayesian analysis. Given a set of parameters $\boldsymbol{\theta}$ (the coordinates, in our case), their posterior distribution given some data $\boldsymbol{D}$ and some hypothesis $\mathcal{H}$ can be written using Bayes' theorem (see e.g. Bishop 2006):

$$\Pr(\boldsymbol{\theta}|\boldsymbol{D}, \mathcal{H}) = \frac{\Pr(\boldsymbol{D}|\boldsymbol{\theta}, \mathcal{H}) \Pr(\boldsymbol{\theta}|\mathcal{H})}{\Pr(\boldsymbol{D}|\mathcal{H})}, \tag{4}$$

which expresses the posterior distribution $\Pr(\boldsymbol{\theta}|\boldsymbol{D}, \mathcal{H})$ as the product of the likelihood $\Pr(\boldsymbol{D}|\boldsymbol{\theta}, \mathcal{H})$ and the parameters' prior $\Pr(\boldsymbol{\theta}|\mathcal{H})$, divided by the evidence $\Pr(\boldsymbol{D}|\mathcal{H})$. For the purposes of inference, we will ignore this last term, as it is just a normalization factor independent of $\boldsymbol{\theta}$; however, in Section 3.4 we show how the evidence can be used to perform model selection, which is another advantage of working in a Bayesian framework. In order to sample the posterior distribution of the source coordinates, we employ nested sampling (Skilling 2006), as implemented in PyMultiNest[1] (Buchner *et al.* 2014), the Python interface to MultiNest (Feroz *et al.* 2009). We choose nested sampling over Metropolis–Hastings sampling or other MCMC techniques as it generally converges faster (Allison & Dunkley 2014) and provides an estimate of the evidence. For the prior, we assume a uniform distribution in the range of the physical model ([0, 1] × [0, 1] × [0, 2.43], with units in kilometres).

To perform Bayesian inference on a given seismogram, we first randomly choose an event's coordinates from the test set. For this set of coordinates, we simulate the observation of a microseismic event for each source mechanism, and generate the noiseless trace as it would be recorded by each of the 23 receivers. We add random Gaussian noise to each component of the noiseless trace. We define the signal-to-noise ratio (SNR) as (Li *et al.* 2018; Zhang *et al.* 2020):

$$\mathrm{SNR} = 10 \log_{10} \frac{\sum_i^{\mathrm{N}} \sum_j^{\mathrm{N_{keep}}} s_{ij}^2}{\sum_i^{\mathrm{N}} \sum_j^{\mathrm{N_{keep}}} (s_{ij} - \tilde{s}_{ij})^2}, \tag{5}$$

where s_{ij} refers to the j-th sample of the i-th trace and $\tilde{s}_{ij}$ to the corresponding noisy trace, $\mathrm{N} = 8000$ is the number of training data and $\mathrm{N_{keep}} = 2001$ (1000 in the ISO case) is the number of time samples. As we explain in Section 3.3, we consider 8000 events out of the total 10 000 for training, and reserve the remainder for validation and testing purposes. Following Li *et al.* (2018), we set SNR=33 dB, which corresponds to a standard deviation of the Gaussian noise of $\sigma = 10.0$, 0.3 and 3.5 for ISO, DC and CLVD, respectively, in the same arbitrary units as the seismograms' amplitude. We show examples of noiseless and noisy signals in Fig. 5.

We note that the choice of Gaussian noise has a quantifiable consequence on the power spectra of the signals, as white noise has an expected constant power in Fourier space (see e.g. Haykin 2001; Papoulis *et al.* 2002). This is reflected on the right-hand side of Fig. 6, where the mean noisy signal is shifted up with respect to the noiseless signals due to the noise addition. We argue that in general any information about the noise power (even if more complicated than Gaussian noise) can be easily accounted for when pre-processing the data: knowing the behaviour of the noise signal in the frequency domain allows one to fully take into account its effect on the relevant signal. For this reason, we decide to transform our data into the Fourier space, as we describe in Section 3.2. We will show in Section 4.3 that this choice makes our proposed approach robust to noise; moreover, our approach lends itself to the extension to coloured noise, which is most likely in realistic seismic data, especially for events with a low SNR (Liu *et al.* 2017).

The noisy seismogram is further pre-processed as described in Section 3.2. At each likelihood evaluation of PyMultiNest, the proposed coordinates are mapped to the predicted pre-processed seismograms by means of the generative model $g(\boldsymbol{\theta})$ described in Section 3.3: by evaluating the likelihood in multiple points of the

[1] https://github.com/JohannesBuchner/PyMultiNest

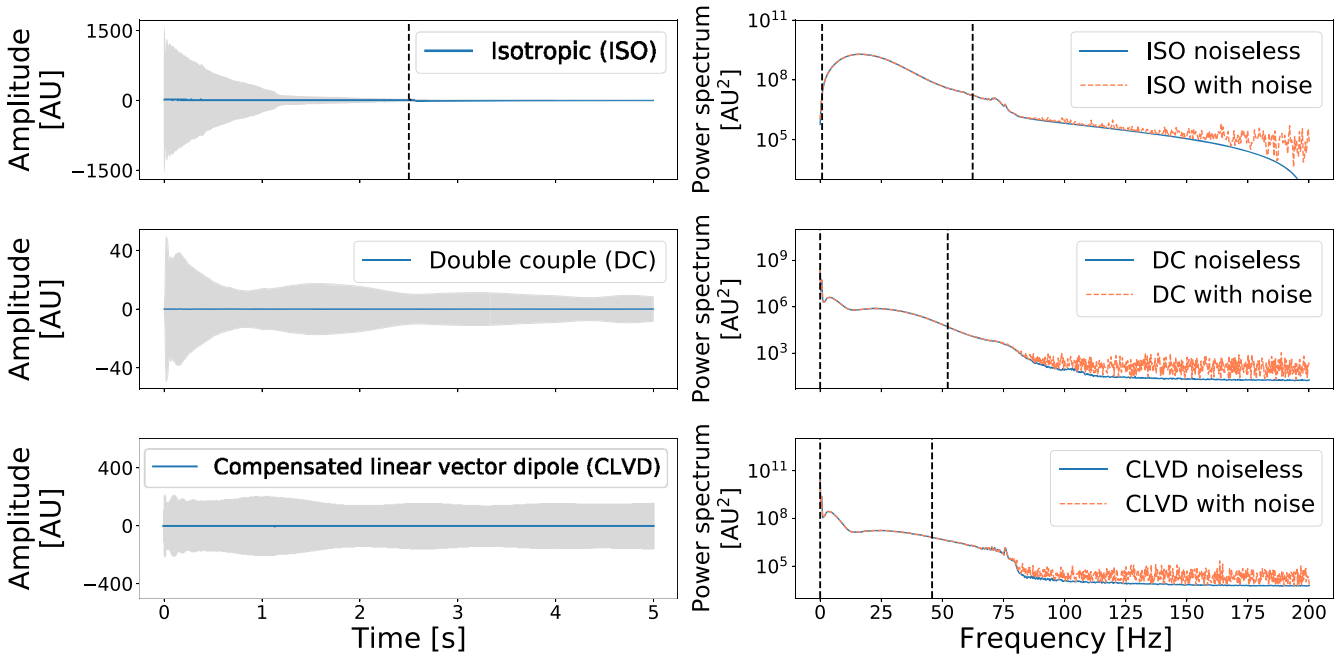

Figure 6. Left-hand column: mean (blue line) and standard deviation (grey area) of all the seismic traces in the training set, for each source mechanism: ISO, DC and CLVD. The seismograms' amplitude is measured in arbitrary units of pressure. The training set is made of 8000 traces for each source mechanism. We cut the isotropic sources at 2.5 s, as indicated by the vertical dashed black line. Right-hand column: noiseless (solid blue) and noisy (dashed coral) mean of the corresponding power spectra, calculated as described in Section 3.2. We consider a signal-to-noise ratio of 33 dB, as described in Section 3.1. We filter the power spectra between the vertical dashed lines selecting only the frequencies where the ratio between the noiseless and noisy signals is more than 99 per cent.

Table 1. Prior range and marginalized mean and 68 per cent credibility intervals on the coordinates (x, y, z) for each source mechanism—ISO, DC and CLVD. The three events are randomly sampled from the test set. These results are obtained by considering all 23 receivers, and training on 8000 simulated events. The noise level is set to 10.0, 0.3 and 3.5, respectively, which corresponds to a signal-to-noise ratio of 33 dB.

Event	Coordinate	Prior range [km]	Ground truth [km]	ISO [km]	DC [km]	CLVD [km]
1	x	[0, 1]	0.71	$0.63^{+0.13}_{-0.12}$	$0.72^{+0.13}_{-0.11}$	$0.71^{+0.11}_{-0.10}$
	y	[0, 1]	0.25	$0.28^{+0.13}_{-0.14}$	$0.25^{+0.10}_{-0.11}$	$0.24^{+0.11}_{-0.11}$
	z	[0,2.43]	2.10	$2.04^{+0.21}_{-0.47}$	$2.09^{+0.11}_{-0.13}$	$2.10^{+0.22}_{-0.25}$
2	x	[0, 1]	0.46	$0.48^{+0.19}_{-0.19}$	$0.43^{+0.12}_{-0.13}$	$0.46^{+0.15}_{-0.12}$
	y	[0, 1]	0.34	$0.32^{+0.16}_{-0.19}$	$0.33^{+0.09}_{-0.08}$	$0.33^{+0.15}_{-0.14}$
	z	[0, 2.43]	1.48	$1.60^{+0.23}_{-0.30}$	$1.54^{+0.18}_{-0.15}$	$1.44^{+0.22}_{-0.24}$
3	x	[0, 1]	0.20	$0.28^{+0.20}_{-0.15}$	$0.20^{+0.10}_{-0.10}$	$0.22^{+0.25}_{-0.13}$
	y	[0, 1]	0.43	$0.46^{+0.18}_{-0.16}$	$0.41^{+0.10}_{-0.13}$	$0.54^{+0.17}_{-0.21}$
	z	[0, 2.43]	0.99	$0.93^{+0.17}_{-0.19}$	$1.09^{+0.28}_{-0.17}$	$0.98^{+0.21}_{-0.20}$

Table 2. Natural logarithm of the evidence for a source located at $(x, y, z) = (0.71$ km, 0.25 km, 2.10 km) and source mechanism ISO, DC and CLVD, as described in Sections 3.4 and 4.5. The hypotheses correspond to an ISO ($\mathcal{H}_{\rm ISO}$), DC ($\mathcal{H}_{\rm DC}$) and CLVD ($\mathcal{H}_{\rm CLVD}$) source mechanism, respectively. We highlighted in bold the highest evidence in each line, which correctly corresponds to the known source mechanism. Note that the natural logarithm of the evidence is returned by PyMultiNest, and in this instance we ignored its associated error (which is very small).

Event type	$\ln \Pr(\boldsymbol{D}\|\mathcal{H}_{\rm ISO})$	$\ln \Pr(\boldsymbol{D}\|\mathcal{H}_{\rm DC})$	$\ln \Pr(\boldsymbol{D}\|\mathcal{H}_{\rm CLVD})$
ISO	**2601**	−608	−79
DC	−17688	**−1165**	−2232
CLVD	−6620	−1052	**−394**

prior space, PyMultiNest can sample from the posterior distribution of the coordinates, thus yielding the required credibility regions in parameter space. Note that, similarly to D18 and SM20, we assume a Gaussian likelihood, i.e. we write

$$\Pr(\boldsymbol{D}|\boldsymbol{\theta}, \mathcal{H}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{D} - g(\boldsymbol{\theta}))^T \boldsymbol{C}^{-1} (\boldsymbol{D} - g(\boldsymbol{\theta}))\right), \quad (6)$$

where $\boldsymbol{C}$ indicates the covariance matrix of the pre-processed seismograms, estimated from the training data.

We note that our choice of a Gaussian likelihood comes without loss of generality as our method can be easily extended to more complicated likelihood models. It is worth stressing that adding Gaussian noise to the seismograms does not necessarily imply that the distribution of the pre-processed seismograms will also be Gaussian; however, we verified experimentally that the distribution of each pre-processed seismogram component is unimodal and symmetric, thus supporting our assumption.

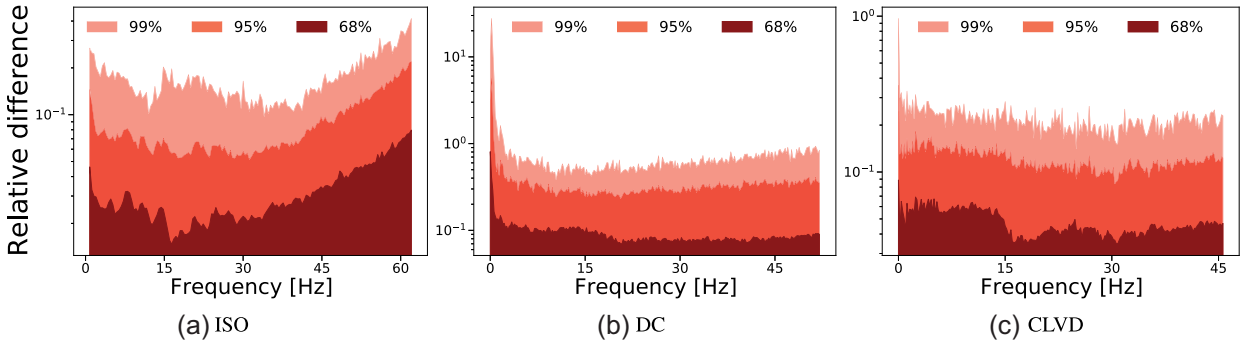

Figure 7. Accuracy of the generative model described in Section 3.3, for ISO, DC and CLVD. The dark red, red and salmon areas enclose the 68, 95 and 99 percentiles of the absolute value of the relative error between the predictions of the model and the target waveforms from the test set.

3.2 Pre-processing

Learning a mapping between coordinates and seismograms directly would be hard for at least two reasons. First, each signal has features with different amplitudes: this means that e.g. a neural network (which will be described in detail in Section 3.3) would likely just focus on the main peak and ignore the other components, thus losing useful information for the source location purpose. Secondly, given the complexity of the seismograms and the high number of features, the amount of data required to train an accurate emulator without overfitting would be at least an order of magnitude higher than what we consider in this work (see e.g. Bishop 2006; Zhu *et al.* 2015, and references therein).

In this sense, we *have to* pre-process the seismograms in order to extract only the relevant information that is needed to locate an event, while discarding all the noisy or redundant features of the signal. Both D18 and SM20 showed the importance of pre-processing, employing GPs in order to select only the components of each seismogram that are essential for inference. However, their methods fail on more complicated sources like the ones we consider in this work. While it could be argued that employing more training data could improve the results, it is also well known that GPs do not scale well with the number of training points (Liu *et al.* 2018), thus it is likely that the D18 method would struggle to generalize to more complicated source mechanisms. Additionally, the method proposed in SM20 is applied directly to the complicated seismic traces like the ones in the left-hand panel of Fig. 3, thus making it more difficult for any algorithm to capture the most useful features for event location. Therefore, we follow a different pre-processing procedure, based on translating the data to the Fourier domain, and we outline the steps in the next paragraphs.

The left-hand panel of Fig. 6 shows the mean and standard deviation of all seismograms in the training set (8000 seismic traces): based on these distributions, we keep all samples of the DC and CLVD signals, and only keep the first half of the ISO traces since they sharply vanish after about one-third of the trace. In other words, we keep only $N_{\rm keep} = 1000$ time samples for the ISO traces, and $N_{\rm keep} = N_{\rm t} = 2001$ time samples for the other mechanisms.

The next pre-processing step we implement is applying the 1-D discrete Fourier transform (Cooley & Tukey 1965) to each seismogram, using the version of NUMPY[2]. Since the amplitudes at each time sample are real numbers, the Fourier Transform returns $(\lfloor N_{\rm keep}/2 \rfloor + 1)$ frequency components: this means that for DC and CLVD sources we are left with 1001 components (501 in the ISO case) in the Fourier domain. We then take the square of the absolute value of these complex numbers: this is usually referred to as a power spectrum. In the right-hand column of Fig. 3 we report the power spectra corresponding to the individual seismograms in the left-hand column of the same figure. We further take the decimal logarithm of the power spectra at each frequency value, and refer to it as logarithmic power spectra in the rest of the paper.

As anticipated in Section 3.1, we shall add some noise to the observed seismogram whose source location coordinates will be inferred. In the right-hand panel of Fig. 6 we show the mean power spectra for each source mechanism with and without noise. We calculate the ratio between the noiseless and the noisy signals, and filter out the frequencies for which this ratio is less than 0.99. We experimented with different thresholds, and chose 0.99 as a good balance between retaining enough features to locate a seismogram and being insensitive to noise. In other words, we additionally cut each power spectrum in the ranges [1 Hz, 62.4 Hz], [0 Hz, 52.2 Hz] and [0 Hz, 45.8 Hz] for ISO, DC and CLVD, respectively, to keep the parts of each signal that are less affected by noise. We observe that translating the seismograms to the Fourier domain has allowed us to obtain smoother signals, as well as to reduce the number of features by a factor of 10. Moreover, this allows our proposed method to be robust to noise: any effect due to noise can be translated into some information of the noise power, and hence readily accounted for in the analysis, for example by selecting a band-limited frequency signal as we discussed in Section 3.1.

To further reduce the number of features, we apply principal component analysis (PCA). PCA is a standard linear compression technique where the data are projected along the eigenvectors of the data covariance matrix. Considering only the components that carry more variance (the so-called 'principal components', corresponding to the largest eigenvalues), it is possible to reduce the number of features while maintaining the relevant information for inference (Bishop 2006). We fit PCA to the training data, and use it to compress the whole data set. After applying PCA to the logarithmic power spectra, we retain 10 principal components for each signal when training the generative model; note that inference is done at the level of the logarithmic power spectra instead. We verified experimentally that varying the number of retained PCA components does not impact the final results significantly.

3.3 Generative model

As our generative model $g(\boldsymbol{\theta})$, we choose to employ a feedforward neural network (NN). A feedforward NN is a set of subsequent layers, each made of a certain number of neurons, that allow for the

[2] https://numpy.org/doc/stable/reference/routines.fft.html

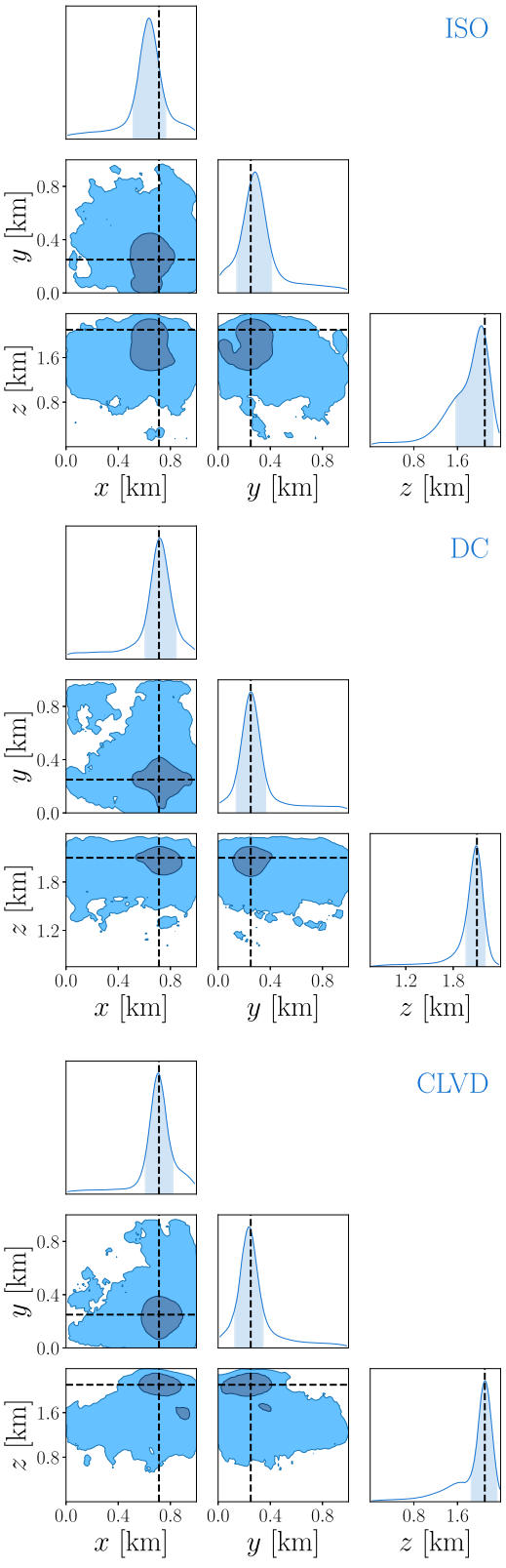

Figure 8. Marginalized 68 and 95 per cent credibility contours obtained with our method for a source located at $(x, y, z) = (0.71$ km, 0.25 km, 2.10 km), indicated by the dashed black lines. We compare three source mechanisms: ISO, DC CLVD. The event corresponds to event 1 in Table 1. Note that we are considering 23 receivers, the signal-to-noise ratio is 33 dB and the emulator was trained on 8000 training points.

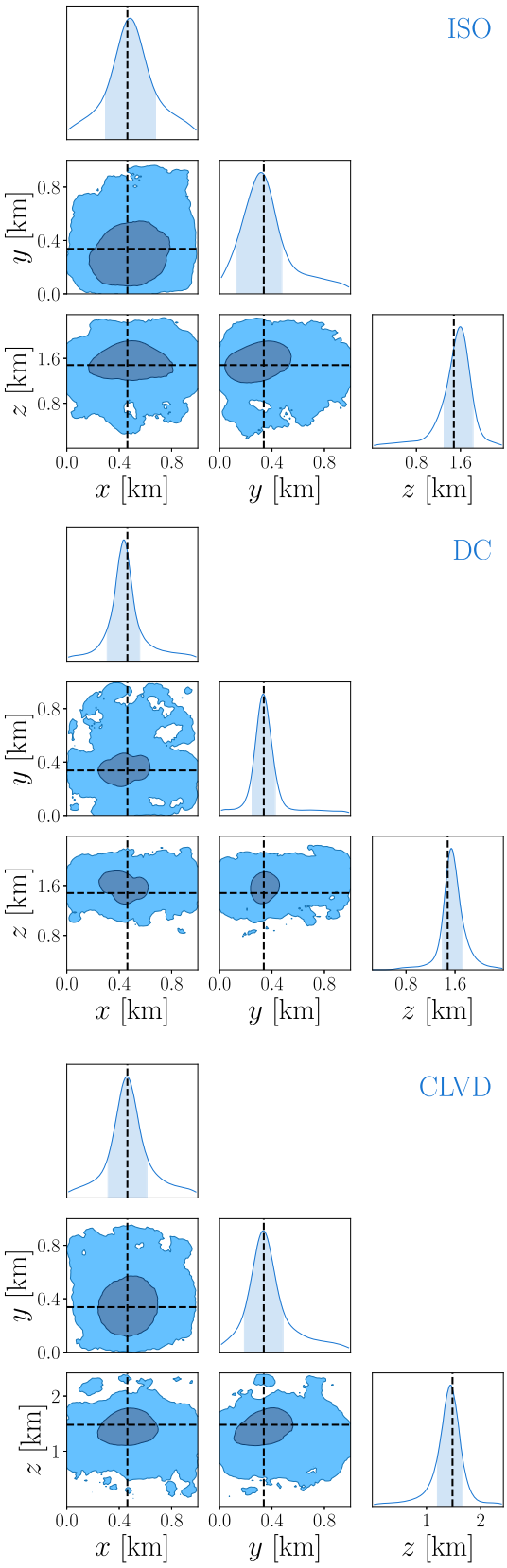

Figure 9. Same as Fig. 8 for a source located at $(x, y, z) = (0.46$ km, 0.34 km, 1.48 km). The event corresponds to event 2 in Table 1.

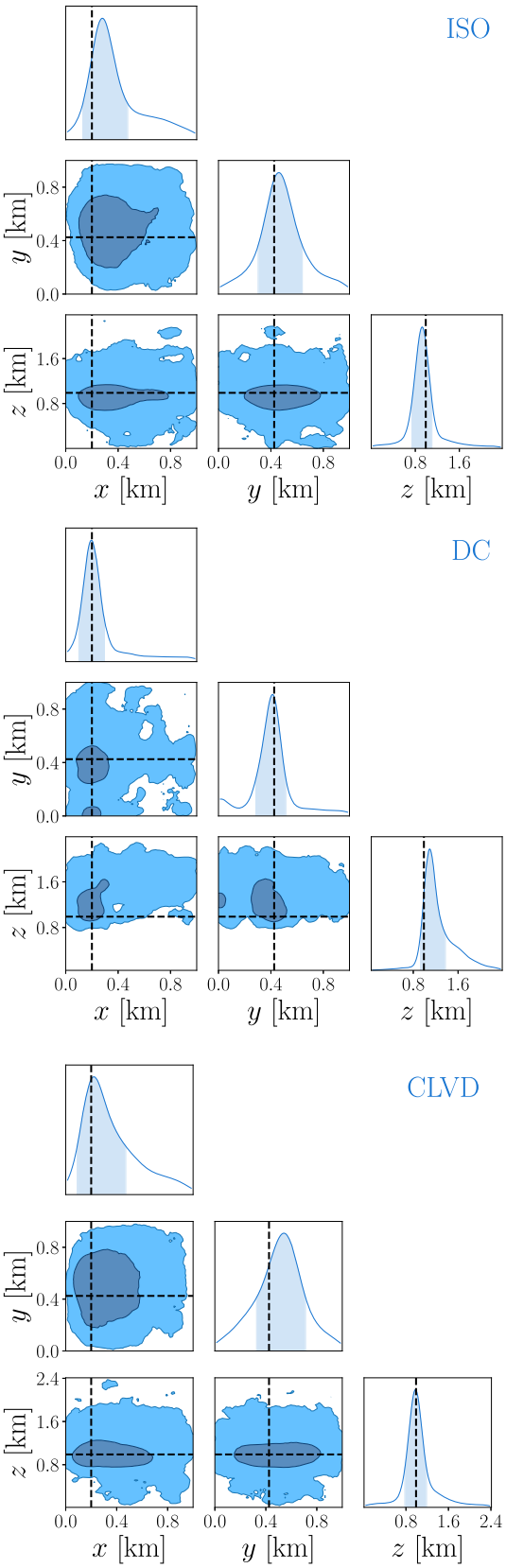

Figure 10. Same as Fig. 8 for a source located at $(x, y, z) = (0.20\ \text{km}, 0.43\ \text{km}, 0.99\ \text{km})$. The event corresponds to event 3 in Table 1.

parametrization of any measurable function between finite dimensional spaces (Hornik *et al.* 1989). Each neuron is associated to a weight, and each layer is additionally associated to a bias (i.e. an offset): weights and biases constitute the parameters that we wish to learn. Additionally, activation functions can be introduced after each layer to model nonlinear mappings. Training the neural network consists of feeding some data through all the layers, and then updating the value of the parameters in order to optimize a chosen loss function.

We employ a neural network made of three layers with 256 neurons each, to provide enough flexibility to the parametrization without consuming too much memory. We set Leaky ReLU (Maas *et al.* 2013) as the activation function for all layers except the last one, where we keep a linear activation function. We recall here that Leaky ReLU acts on the output of each layer $\mathbf{O}$ as follows:

$$\text{LeakyReLU}(\mathbf{O}) = \begin{cases} \mathbf{O} & \text{if } \mathbf{O} > 0 \\ \alpha\mathbf{O} & \text{otherwise} \end{cases}, \tag{7}$$

where we set the hyperparameter $\alpha = 0.3$; Leaky ReLU is usually preferred over the standard ReLU (rectified linear unit) because of non-vanishing gradients (Kolen & Kremer 2001). We also experimented with the ELU (exponential linear unit; Clevert *et al.* 2015) activation function, and found no significant improvements in the overall results with respect to using Leaky ReLU. We choose the Mean Squared Error (MSE) between the network output and the principal components of the training data as our loss function to minimize.

For each source mechanism and each receiver, we train the emulator using 8000 traces; we reserve 1000 seismograms for validation purposes and 1000 seismograms for testing purposes. We remark that we train a single neural network for each receiver: for a given source mechanism and underground location, the training data for each receiver includes the waveforms as observed by that particular station. In this way, we are able to include the information on the receiver's position in our generative model.

To train the neural network, we use the Adam optimizer (Kingma & Ba 2014) with default parameters; moreover, we choose a learning rate of 0.001 and a batch size of 256: the former controls the step size of the parameters' update, while the latter indicates the number of training points that are fed through the network at each iteration. We additionally set a patience of 50 to early-stop (Yao *et al.* 2007) based on the validation loss: this means that if the loss calculated on the validation set has not decreased in the last 50 epochs, we stop training and take the model corresponding to the minimum validation loss as the best model, as we interpret the model to have achieved its minimum error on unseen data[3]. We report the typical trend for the training and validation loss curves in Supporting Information Fig. S1.

The test set is used to randomly sample events on which we perform our Bayesian inversion analysis, as described in Section 3.1. We also explore the behaviour of the posterior distribution as a function of the number of training events, number of receivers and noise scale, which we show in Section 4.3. However, we do not perform a full grid search amongst the hyperparameters (e.g. number of layers, number of neurons, activation function and learning rate),

[3]We also tested a dynamic learning rate decreasing by a factor of 10 every time the validation loss did not decrease for 50 epochs, but found no significant improvement over a constant learning rate of 0.001, which we therefore chose for our analysis.

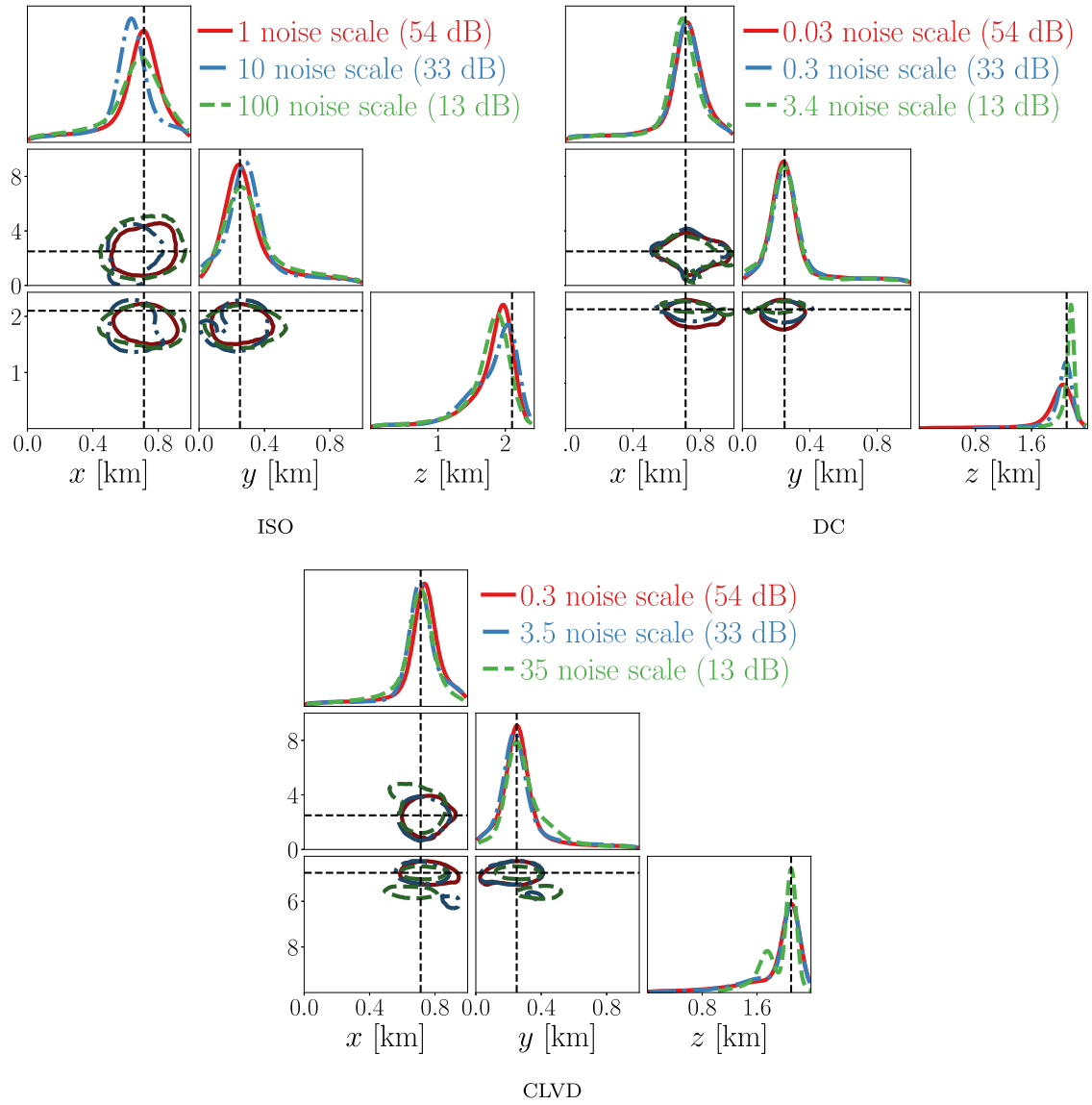

Figure 11. Marginalized 68 per cent credibility contours obtained with our method for a source located at $(x, y, z) = (0.71\ \text{km}, 0.25\ \text{km}, 2.10\ \text{km})$, indicated by the dashed black lines, comparing different levels of noise. In particular, in each panel we show signal-to-noise ratios of 13, 33 and 54 dB for the three source mechanisms: ISO, DC and CLVD. Note that we are considering 8000 training data for the emulator and 23 receivers. This figure is best viewed in colour.

as we observe the results are not significantly affected by them; we defer a more complete grid search to future work.

3.4 Model selection

As detailed in Section 3.1, we ignore the denominator in eq. (4) when inferring the coordinates of a microseismic event. However, we can use the evidence $\Pr(\boldsymbol{D}|\mathcal{H})$ to perform model selection, thus showing another advantage of our proposed Bayesian approach (see e.g. Knuth *et al.* 2015). The quantity $\Pr(\boldsymbol{D}|\mathcal{H})$ can be interpreted as the likelihood of a given signal under a certain hypothesis, with the constraint that the hypotheses form a set of $n_{\rm hyp}$ pairwise disjoint events whose union is the entire possibility space—that is$\Pr(\mathcal{H}_i \cap \mathcal{H}_j) = 0$ for $i \neq j, \forall i, j = \{1, \ldots, n_{\rm hyp}\}$, and $\sum_{i=1}^{n_{\rm hyp}} \Pr(\mathcal{H}_i) = 1$.

To compare two hypotheses, and thus perform model selection, we define the Bayes factor BF as

$$\text{BF} \equiv \frac{\Pr(\boldsymbol{D}|\mathcal{H}_i)}{\Pr(\boldsymbol{D}|\mathcal{H}_j)} = \frac{\Pr(\mathcal{H}_i|\boldsymbol{D})\Pr(\mathcal{H}_j)}{\Pr(\mathcal{H}_j|\boldsymbol{D})\Pr(\mathcal{H}_i)}, \tag{8}$$

where the second equality has been obtained using Bayes' theorem, $\Pr(\mathcal{H}_i)$ is the prior distribution of the hypothesis $\mathcal{H}_i$, and $\Pr(\mathcal{H}_i|\boldsymbol{D})$ is the posterior distribution of hypothesis $\mathcal{H}_i$ given $\boldsymbol{D}$. If the hypotheses are equiprobable *a priori*, we can write $\Pr(\mathcal{H}_i) = \Pr(\mathcal{H}_j)$, which allows us to express the Bayes factor as the ratio of the posterior distribution of one hypothesis over the other. Hence, if the Bayes factor as defined in eq. (8) is greater than 1, we can interpret it as hypothesis $\mathcal{H}_i$ being more favoured than hypothesis $\mathcal{H}_j$ under the observed data $\boldsymbol{D}$ (Knuth *et al.* 2015).

Translating this into practice, after training the emulators, given an observed signal $\boldsymbol{D}$ as described in Section 3.1, we can compare the three following equiprobable hypotheses: the source mechanism

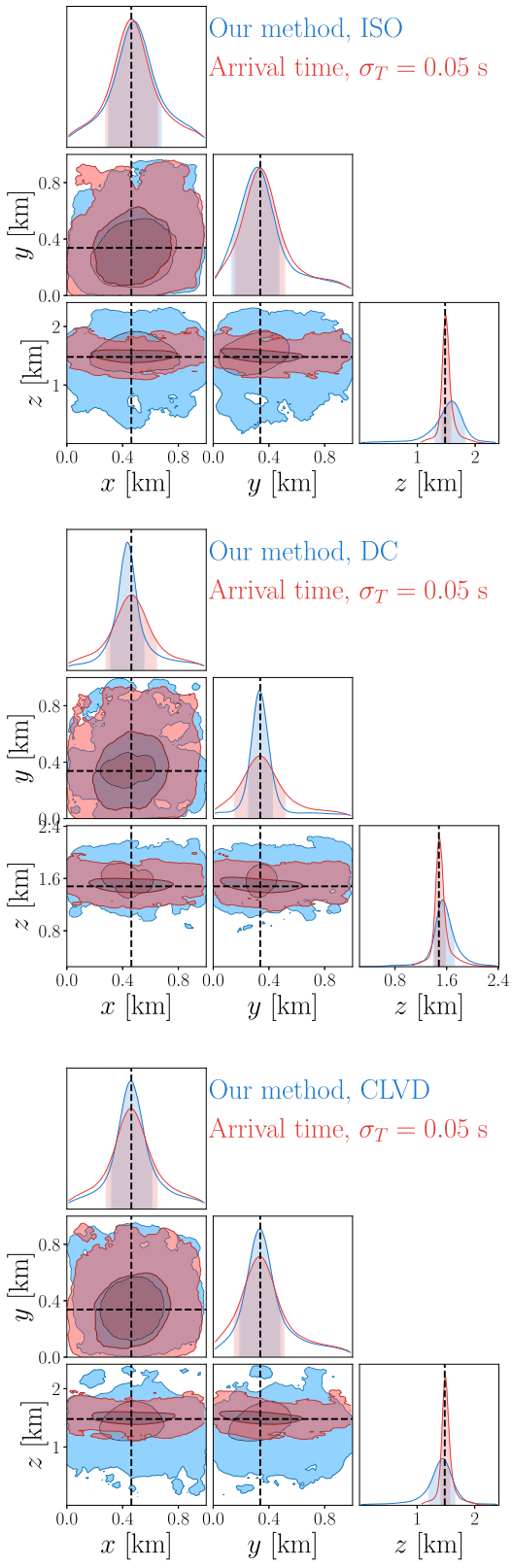

Figure 12. Comparison of the marginalized 68 and 95 per cent credibility contours obtained with our method and with the arrival time approach described in Section 4.4. The event corresponds to event 2 in Table 1. Note that we are considering 23 receivers in both cases, and for our approach the signal-to-noise ratio is 33 dB and the emulator was trained on 8000 training points.

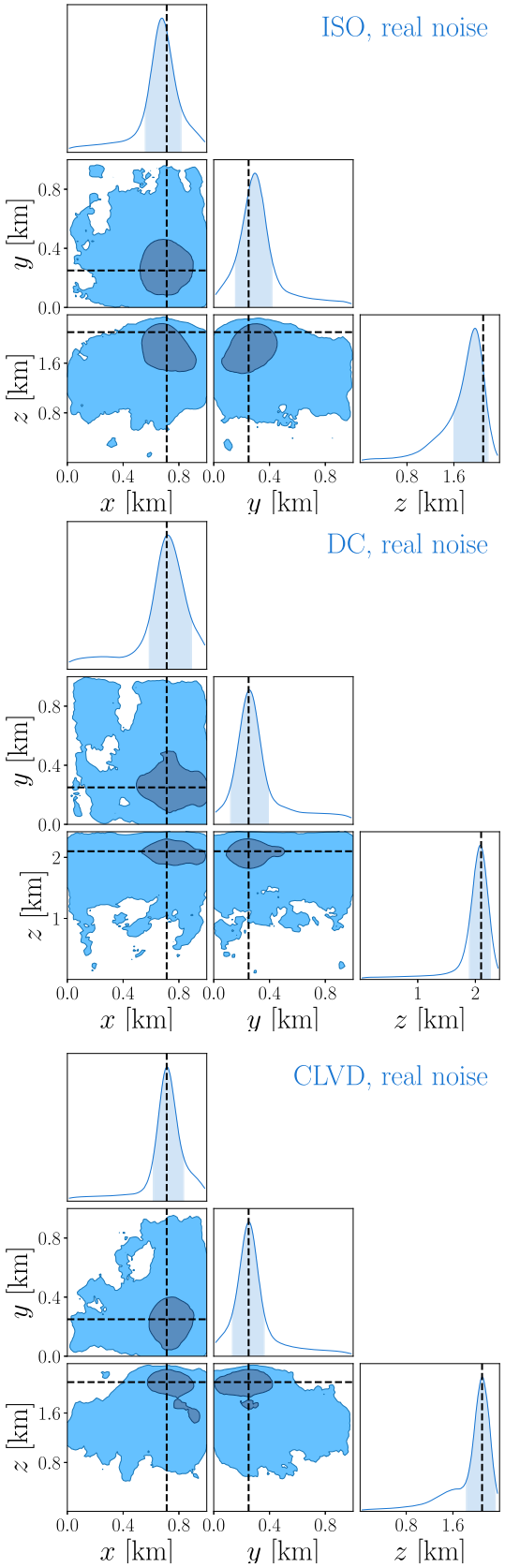

Figure 13. Same as Fig. 8, considering real noise from Groningen field data instead of Gaussian noise. As in our fiducial analysis, the signal-to-noise ratio is 33 dB, and we consider 23 receivers.

is isotropic ($\mathcal{H}_{\rm ISO}$), the source mechanism is double couple ($\mathcal{H}_{\rm DC}$), or the source mechanism is compensated linear vector dipole ($\mathcal{H}_{\rm CLVD}$). The advantage of using nested sampling is that the evidence is calculated while sampling the posterior distribution. Consequently, by feeding the observation $\boldsymbol{D}$ to each emulator it is straightforward to obtain the evidences $\Pr(\boldsymbol{D}|\mathcal{H}_{\rm ISO})$, $\Pr(\boldsymbol{D}|\mathcal{H}_{\rm DC})$ and $\Pr(\boldsymbol{D}|\mathcal{H}_{\rm CLVD})$. By looking at the hypothesis that maximizes the evidence, we can select the model that best describes the given observation, thus identifying the source type for a given observation. We show the results in Section 4.5 and Table 2.

4 RESULTS

4.1 Speed performance

We first report on the speed performance of our method. We recall here that if we were to solve the elastic wave equation at each likelihood evaluation, inference would be severely compromised, as a single event's source inversion would take thousands of hours on a high performance computing (HPC) cluster, if at all possible. In contrast, our method requires only $\sim 10^4$ simulations to be produced once, and the emulator to be trained once—an overhead of $\mathcal{O}(100\,\mathrm{h})$ and $\mathcal{O}(1\,\mathrm{h})$, respectively—and then it allows for the complete source inversion of any event in $\mathcal{O}(0.1\,\mathrm{h})$ on a commercial laptop. In other words, most of the time taken by our approach is spent training the emulator, which needs to be done only once, provided the density and velocity models remain unchanged[4]; after the training is complete, performing inference on a given recorded seismogram takes less than 10 min on a commercial laptop. As a reference, conventional full-waveform inversion techniques can take several hours on CPUs, and a comparable amount of time to our method when running on GPUs (see e.g. Abreo-Carrillo *et al.* 2015).

4.2 Generative model accuracy

In this section, we briefly look at the performance of the generative model that we described in Section 3.3. We consider the 1000 events in the test set, and measure the relative difference between the prediction of our model and the target waveforms. We show the 68, 95 and 99 percentiles of the relative difference in Fig. 7 for all source mechanisms. We observe that there are some significant discrepancies, especially at low frequency for the DC and CLVD, which however we attribute to low values in the logarithmic power spectra, and which do not seem to compromise inference, as we show in the next section.

4.3 Inference results

We then turn our attention to the accuracy of the inferred posterior distribution of the coordinates. For each source mechanism, we report the inference results for three different coordinates in Figs 8–10: each shows the posterior contour plots obtained with our methodology, considering all 23 receivers, SNR = 33 dB and using 8000 seismograms as the training set for the emulator. The numerical results are summarized in Table 1, reporting the prior ranges and marginalized mean and 68 per cent credibility interval on the coordinates. We additionally report the mean and the standard deviation of the absolute value of the difference between the ground truth and the maximum of the retrieved posterior distribution for 100 randomly picked test events: this is (0.17 ± 0.29) km. This difference indicates good agreement, despite being skewed by a small fraction (less than 5 per cent) of test cases for which the inference results were totally unconstrained, or for which the width of the posterior distribution was particularly broad: we remark that these occurrences generally corresponded to events for which the ground truth had at least one coordinate lying at the border of the prior space, thus complicating the sampling procedure. We argue that it is safe to ignore these limit cases, given their particular location; in a real scenario, one would train the neural network on events coming from a larger prior volume than the one that is being investigated here, thus avoiding this problem. We therefore conclude that with our method we can accurately retrieve the correct value of the coordinates across almost the entire prior parameter space.

Additionally, we observe that the x and y source coordinates are usually less constrained than the z coordinate for ISO and CLVD, while this behaviour is less prominent in the DC case, for which all coordinates are always tightly constrained. We attribute this effect to two possible causes. On one hand, it can be related to the specific density model we are considering in this work, which has a layered structure. On the other hand, we note that when translating the seismic traces to the Fourier domain we ignored the phase signal (since we only considered the amplitude power spectra), thus possibly losing useful information for the source location purpose (e.g. Ferreira & Woodhouse 2007); we argue that this also results in larger uncertainties in the retrieved posterior contours with respect to a full-waveform approach, which however would be too computationally expensive to run. We additionally note that retaining the phase information would yield a generative model, as by combining the power spectra with the phase one could in principle reconstruct a full seismogram from the source coordinates (after training the emulator). We defer the study of phase information to future work, as we anticipate that given the oscillatory behaviour of the phase signals it will be harder to train an emulator on them.

We then study the dependence of the posterior contours on the SNR, the number of training data and the number of receivers used. In Fig. 11, we first show the effect of different noise levels; in particular, for each source mechanism we vary the SNR from 13 to 54 dB. We note that, in order to increase the robustness to noise, we cut different frequency windows based on the noise levels: given the Gaussian noise model we assume in this work, a smaller SNR corresponds to a higher noise power, and hence to a smaller number of retained power spectrum principal components. We observe that while higher noise levels can lead to small biases in some of the 1-D marginalized coordinates' distributions, in general no significant variations in the shape of the 2-D posterior contours are present.

In Supporting Information Fig. S2, we show the posterior contours when training the emulator with 2000, 5000 and 8000 data points. Again, we observe no significant differences for the ISO and CLVD sources, while very small deviations appear when using fewer training data in the DC case. In general, $\mathcal{O}(10^3)$ training data are enough to obtain accurate posterior contours for all source mechanisms. Finally, in Supporting Information Fig. S3, we vary the geometry of the receivers used for recording the microseismic traces. While we are not interested in a full study of the optimal geometry of the receivers, we observe that using fewer receivers leads to broader posterior contours, while still allowing accurate event locations for all source mechanisms.

[4]In general, the stability of a given velocity model is not well known—see e.g. Thornton (2013), Usher *et al.* (2013), Gesret *et al.* (2013, 2014), Das *et al.* (2018) and references therein for a discussion on the uncertainties of velocity models, and their consequences on location errors.

In summary, our results are very robust to the noise injected into the observed seismogram. Additionally, very few receivers—$\mathcal{O}(10)$—are needed to obtain accurate results, and a number of training points of order $\mathcal{O}(10^3)$ is sufficient to locate any event.

4.4 Comparison with arrival time techniques

In order to establish a connection with existing standard earthquake location methods, we compare our results with a nonlinear probabilistic location technique based on arrival times. We consider the NonLinLoc algorithm (Lomax *et al.* 2000), which implements the LS-L2 approach (Tarantola & Valette 1982; Moser *et al.* 1992; Wittlinger *et al.* 1993). In this framework, the likelihood of the arrival time for a single observed event for a single receiver is

$$\mathcal{L}_{\text{arrival time}} \propto \exp\left(-\frac{1}{2}\frac{(T_{\text{obs}} - T_{\text{calc}})^2}{\sigma_T^2}\right) \tag{9}$$

where T_{obs} is the observed arrival time, T_{calc} is the theoretical estimate for the travel time and σ_T is the error on the manual picking of arrival times. We use a fast marching method (Sethian 1996) as implemented in Pykonal (White *et al.* 2020) to estimate theoretical arrival times given a set of coordinates for the geological model described in Section 2. We further replace the posterior sampler available in NonLinLoc with a PyMultiNest implementation for the likelihood in eq. (9).

We compare the location inference results for event 2 in Fig. 12. In this instance, we are interested in showing that there exists a reasonable value of the picking error σ_T that yields comparable constraints to our method; this error is usually arbitrary and heavily depends on the SNR and prior information (Smith 2019; Abakumov *et al.* 2020). We remark that we are not interested in showing that the constraints obtained with our method are tighter than the ones obtained with the time-arrival information: while we start from the full waveform, we subsequently discard the phase information and part of the power spectrum, as we described in Section 3.2. We additionally note that the arrival time estimates are typically obtained an order of magnitude faster than our method, taking $\mathcal{O}(0.01\text{ h})$ on the same commercial laptop used in Section 4.1; this is not surprising, given the simple form of the likelihood in eq. (9) and the efficiency of Pykonal.

Nevertheless, we show that with a picking error of $\sigma_T = 0.05$ s, we obtain results that are generally in agreement with our approach. In particular, the constraints on the z source coordinate are tighter in the arrival-time case, while the x and y coordinates are equally or better constrained by our approach; moreover, as in our method, the arrival-time analysis yields z coordinates that are more constrained than the x and y ones. We speculate that this is due to the particular layered structure of the geological model described in Section 2, and, as anticipated in Section 4.3, we expect the inference power of our approach to improve significantly by including the phase information into the analysis. We stress that our work is a step towards the fast joint inversion of coordinates and moment tensor components, for which arrival time techniques alone are not sufficient (Dahm & Krüger 2014; Pugh *et al.* 2016; Alvizuri *et al.* 2018); therefore, we leave a full comparison to other techniques to future work. However, we further discuss the joint inversion approach and how to deal with more complicated noise types in Section 5.

4.5 Model selection results

In general, a microseismic event can be described as a linear combination of the three source mechanisms described in Section 2 (Vavryčuk 2015). While we considered the three sources separately in this work, we show how the proposed Bayesian approach additionally allows for the identification of a source type given an observed seismogram. We consider event 1 as reported in Table 1, and produce an observation for each source mechanism (ISO, DC and CLVD). As described in Section 3.4, we can run nested sampling for each event and for every trained emulator, and obtain 9 evidence values in total, whose logarithm we report in Table 2. As expected, the evidence is maximal in correspondence of the source type that generated the given event, which indicates that we are capable of correctly identifying the source type for a given observation. What is more, this selection is also very fast, as after training the emulators each evidence calculation takes less than 10 min on a commercial laptop.

4.6 Realistic noise and network configuration

In this section, we explore the application of our approach to a more realistic scenario. We consider the same 23-receiver configuration, which is uniformly spread on the plane at $z = 2.43$ km; however, instead of Gaussian noise, we add real noise traces extracted from the freely available Groningen field data set (Smith 2019).

Our approach to obtain real noise traces is as follows. We cluster the receivers and synthetic traces in cubic volumes of side 0.5 km; this returns 100 clusters in total. We therefore take 100 different real waveforms among Groningen data, and cut the first 5 s window, well before the first wave arrival. We interpolate the signal in these windows in order to match the sampling rate of the synthetic signals; then, in order to reach the required SNR of 33 dB, we tune the amplitude of the noise signal we add to each synthetic waveform, and subtract the average signal so that the resulting noisy waveform has zero mean. We found that in some cases the SNR of the resulting noisy signals is lower than 33 dB: these correspond to waveforms originating at the edge of our simulated field, which, however, do not impact the inference analysis, as we show in this section.

Real noise is added to event 1 in Table 1, as well as to all training waveforms to calculate the covariance matrix in eq. (6). We re-train the neural networks in the DC and CLVD cases removing the zero-frequency value of the power spectrum, in order to be insensitive to the scale of the noisy signals in the time domain. In Fig. 13, we show the posterior contours of our inference analysis using real noise: the results show little difference with respect to the case with Gaussian noise in Fig 8, thus further demonstrating the robustness of our approach.

5 CONCLUSIONS

In this paper, we proposed a method that allows for the fast and accurate retrieval of the source coordinates for any microseismic source mechanism: ISO, DC and CLVD. This offers an efficient technique to both locate an event and identify its source type, exploiting the power of machine learning and Bayesian tools to extract the information contained in seismic waveforms.

Our proposed method is based on a physically motivated preprocessing of the raw signals, using Fourier analysis and principal component compression, followed by the use of a neural network to learn the mapping between coordinates and principal components.

Using the learnt forward model in combination with Bayesian techniques, we showed that we can retrieve an accurate estimate of any microseismic event coordinates, for any source mechanism, in less than 10 min on a commercial laptop. Therefore, we demonstrated for the first time that machine learning techniques allow for a fast and accurate Bayesian analysis on microseismic traces, yielding competitive results on ISO sources and state-of-the-art results on DC and CLVD sources.

We showed that $\mathcal{O}(10^3)$ events for each source mechanism are enough to train a representative emulator, when using the data coming from $\mathcal{O}(10)$ receivers placed at the seabed as indicated in Fig. 2. We explored the effect of the noise level, and how the number of receivers and the number of training data for the emulator impact the accuracy of the coordinates' posterior distribution, demonstrating the robustness of our approach, which was also compared with a standard arrival-time inversion technique, showing good agreement for an appropriate choice of the picking error. We also showed that our method works in the presence of realistic noise statistics and network configuration. Finally, we demonstrated the utility of our Bayesian approach by calculating the Bayesian evidence for a given observation and three hypotheses, and showed that this correctly identifies the source type of any given event.

In conclusion, our work lays the foundations for the fast and reliable location of microseismic events with any source mechanism, given a minimal amount of computing resources. We also foresee that recent improvements in solving the forward model even outside the boundary of the training data, like physics-informed neural networks (PINNs; Costa Nogueira Junior *et al.* 2019; Raissi *et al.* 2019; Xu *et al.* 2019; Moseley *et al.* 2020), could be combined with our proposed approach to make it even more robust, especially when different velocity models have to be employed.

Our work represents a step forward towards the fast Bayesian characterization of microseismic events: after training our models on noiseless simulation data, it is possible to apply our method to each individual seismic trace as recorded by the receivers in order to obtain a fast and accurate source location estimate. We plan to integrate our approach into a joint inversion analysis, where both moment tensor components and coordinates are inferred: such analysis is usually slow and computationally expensive (e.g. Pugh *et al.* 2016), and we therefore expect our approach to significantly accelerate it. Some straightforward extensions of our approach then include the following three points. First, a microseismic event is in general described by a linear mixture of components of the moment tensor, which we considered separately in this work. As a consequence, this implies that a larger set of parameters, including two more free parameters measuring the relative strength of the source types, has to be considered. We anticipate that this increase in the total number of parameters to infer will require a larger data set to train the emulator; however, we note that our proposed method scales well with the number of training data, and therefore we anticipate that performing a Bayesian analysis with a larger parameter space is an attainable goal using our approach. Second we note that in order for this method to be deployed in a realistic scenario, the noise associated with the recorded seismograms should be modelled more carefully: a lower SNR may have to be considered, a more complicated likelihood distribution might have to be implemented, or a 'likelihood-free' approach should be investigated (Sunnåker *et al.* 2013). Last, the errors in the 3-D density and velocity models should be incorporated into the analysis, in order to account for all sources of uncertainty (Gesret *et al.* 2013, 2014). This will be addressed in future work.

ACKNOWLEDGMENTS

We thank Saptarshi Das for useful discussions, as well as Jonathan Smith and the anonymous reviewers for their helpful comments. DP was supported by the STFC UCL Centre for Doctoral Training in Data Intensive Science. Generation of the synthetic data used in this work has been performed in part on the Wilkes High Performance GPU computer cluster at the University of Cambridge, and in part on the Beaker cluster at UCL. This work was partially enabled by funding from Royal Dutch Shell PLC, and used facilities provided by the UCL Cosmoparticle Initiative. We acknowledge the use of NumPy (Harris *et al.* 2020), Matplotlib (Hunter 2007), TensorFlow (Abadi *et al.* 2015), SciPy (Virtanen *et al.* 2020) and ChainConsumer (Hinton 2016).

DATA AVAILABILITY

The code to reproduce this work is publicly available in the GitHub repository at this link: https://github.com/alessiospuriomancini/seismoML. The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abadi, M. *et al.*, 2015. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,* Software available from tensorflow.org, https://www.tensorflow.org/about/bib.

Abakumov, I., Roeser, A. & Shapiro, S.A., 2020. Arrival-time picking uncertainty: theoretical estimations and their application to microseismic data, *Geophysics,* **85**(4), U65–U76.

Abreo-Carrillo, S.A., Ramirez, A.B., Reyes, O., Abreo-Carrillo, D.L. & González-Alvarez, H., 2015. A practical implementation of acoustic full waveform inversion on graphical processing units, *CT&F—Cienc. Tecnol. Futuro,* **6(2),** 5–16.

Allison, R. & Dunkley, J., 2014. Comparison of sampling techniques for Bayesian parameter estimation, *Mon. Not. R. astr. Soc.,* **437**(4), 3918–3928.

Alsing, J., Wandelt, B. & Feeney, S., 2018. Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology, *Mon. Not. R. astr. Soc.,* **477**(3), 2874–2885.

Alvizuri, C., Silwal, V., Krischer, L. & Tape, C., 2018. Estimation of full moment tensors, including uncertainties, for nuclear explosions, volcanic events, and earthquakes, *J. geophys. Res.,* **123**(6), 5099–5119.

Angus, D., Aljaafari, A., Usher, P. & Verdon, J., 2014. Seismic waveforms and velocity model heterogeneity: Towards a full-waveform microseismic location algorithm, *J. Appl. Geophys.,* **111,** 228–233.

Auld, T., Bridges, M., Hobson, M. & Gull, S., 2007. Fast cosmological parameter estimation using neural networks, *Mon. Not. R. astr. Soc.,* **376,** L11–L15.

Auld, T., Bridges, M. & Hobson, M.P., 2008. cosmonet: fast cosmological parameter estimation in non-flat models using neural networks, *Mon. Not. R. astr. Soc.,* **387**(4), 1575–1582.

Bergen, K.J., Johnson, P.A., de Hoop, M.V. & Beroza, G.C., 2019. Machine learning for data-driven discovery in solid earth geoscience, *Science,* **363**(6433), eaau0323 , doi:10.1126/science.aau0323.

Bishop, C.M., 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics),* Springer-Verlag.

Brueckl, E., Binder, D., Hausmann, H. & Mertl, S., 2008. Hazard estimation of deep seated mass movements by microseismic monitoring, in *International Strategy for Disaster Reduction*.Austrian Academy of Sciences

Buchner, J. *et al.*, 2014. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue, *Astron. Astrophys.,* **564,** A125 , doi:10.1051/0004-6361/201322971.

Chenthamarakshan, V. *et al.*, 2020. *CogMol: Target-Specific and Selective Drug Design for COVID-19 Using Deep Generative Models,* preprint (arXiv:2004.01215).

Clevert, D.-A., Unterthiner, T. & Hochreiter, S., 2015. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs),* preprint (arXiv:1511.07289).

Collettini, C. & Barchi, M.R., 2002. A low-angle normal fault in the Umbria region (Central Italy): a mechanical model for the related microseismicity, *Tectonophysics,* **359**(1), 97–115.

Conrad, P.R., Marzouk, Y.M., Pillai, N.S. & Smith, A., 2016. Accelerating asymptotically exact MCMC for computationally intensive models via local approximations, *J. Am. Stat. Assoc.,* **111**(516), 1591–1607.

Cooley, J.W. & Tukey, J.W., 1965. An algorithm for the machine calculation of complex Fourier series, *Math. Comput.,* **19,** 297–301.

Costa Nogueira Junior, A., De Sousa Almeida, J., Paredes Quiñones, M. & de Albuquerque Martins, L.S., 2019. Physics-based machine learning inversion of subsurface elastic properties, in *Conference Proceedings, 81st EAGE Conference and Exhibition,* Vol. 2019, pp. 1–5, doi:10.3997/2214-4609.201901147.

Craiu, R.V. & Rosenthal, J.S., 2014. Bayesian computation via Markov chain Monte Carlo, *Annu. Rev. Stat. Its Appl.,* **1**(1), 179–201.

Dahm, T. & F., Krüger. 2014. Moment tensor inversion and moment tensor interpretation, (Ed.) Bormann, P. , *New Manual of Seismological Observatory Practice 2 (NMSOP-2),* Potsdam : Deutsches GeoForschungsZentrum GFZ, 1-37.

Das, S., Chen, X. & Hobson, M.P., 2017. Fast GPU-based seismogram simulation from microseismic events in marine environments using heterogeneous velocity models, *IEEE Trans. Comput. Imaging,* **3**(2), 316–329.

Das, S., Chen, X., Hobson, M.P., Phadke, S., van Beest, B., Goudswaard, J. & Hohl, D., 2018. Surrogate regression modelling for fast seismogram generation and detection of microseismic events in heterogeneous velocity models, *Geophys. J. Int.,* **215**(2), 1257–1290.

Ellsworth, W.L., 2013. Injection-induced earthquakes, *Science,* **341**(6142) , doi:10.1126/science.1225942.

Feroz, F., Hobson, M. & Bridges, M., 2009. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics, *Mon. Not. R. astr. Soc.,* **398,** 1601–1614.

Ferreira, A. M.G. & Woodhouse, J.H., 2007. Source, path and receiver effects on seismic surface waves, *Geophys. J. Int.,* **168**(1), 109–132.

Geiger, L., 1910. Herdbestimmung bei Erdbeben aus den Ankunftszeiten, Nachrichten der K. Gesellschaft der Wissenschaften zu Gottingen, *Math.-Phys. Klasse,* **1910,** 331–349, https://eudml.org/doc/58769.

Gesret, A., Noble, M., Desassis, N. & Romary, T., 2013. Microseismic monitoring - consequences of velocity model uncertainties on event location uncertainties, in *Proceedings of the Third Passive Seismic Workshop,* European Association of Geoscientists and Engineers, Athens, Greece , doi:10.3997/2214-4609.20131027.

Gesret, A., Desassis, N., Noble, M., Romary, T. & Maisons, C., 2014. Propagation of the velocity model uncertainties to the seismic event location, *Geophys. J. Int.,* **200**(1), 52–66.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y., 2014. Generative adversarial nets, in *Advances in Neural Information Processing Systems 27,* pp. 2672–2680, eds Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D. & Weinberger, K.Q., Curran Associates, Inc.

Gulrajani, I., Kumar, K., Ahmed, F., Taiga, A.A., Visin, F., Vazquez, D. & Courville, A., 2016. *PixelVAE: A Latent Variable Model for Natural Images,* preprint (arXiv:1611.05013).

Harris, C.R. *et al.*, 2020. Array programming with NumPy, *Nature,* **585**(7825), 357–362.

Haykin, S., 2001. *Communication Systems,* Wiley.

Hinton, S.R., 2016. ChainConsumer, *J. Open Source Softw.,* **1,** 00045 , doi:10.21105/joss.00045.

Hornik, K., Stinchcombe, M. & White, H., 1989. Multilayer feedforward networks are universal approximators, *Neural Netw.,* **2**(5), 359–366.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment, *Comput. Sci. Eng.,* **9**(3), 90–95.

Jakobsen, M. & Ursin, B., 2015. Full waveform inversion in the frequency domain using direct iterative T-matrix methods, *J. Geophys. Eng.,* **12,** 400–418.

Kasim, M.F. *et al.*, 2020. *Building high accuracy emulators for scientific simulations with deep neural architecture search,* preprint (arXiv:2001.08055).

Kingma, D.P. & Ba, J., 2014. *Adam: A Method for Stochastic Optimization,* preprint (arXiv:1412.6980).

Knopoff, L. & Randall, M.J., 1970. The compensated linear-vector dipole: a possible mechanism for deep earthquakes, *J. geophys. Res.,* **75**(26), 4957–4963.

Knuth, K.H., Habeck, M., Malakar, N.K., Mubeen, A.M. & Placek, B., 2015. Bayesian evidence and model selection, *Digit. Signal Process.,* **47,** 50–67 , Special Issue in Honour of William J. (Bill) Fitzgerald, doi:10.1016/j.dsp.2015.06.012.

Kolen, J.F. & Kremer, S.C., 2001. *Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies,* pp. 237–243, Wiley-IEEE Press.

LeNail, A., 2019. NN-SVG: publication-ready neural network architecture schematics, *J. Open Source Softw.,* **4**(33), 747 , doi:10.21105/joss.00747.

Li, H., Wang, R. & Cao, S., 2015. Microseismic forward modeling based on different focal mechanisms used by the seismic moment tensor and elastic wave equation, *J. Geophys. Eng.,* **12**(2), 155–166.

Li, J., 2013. Study of induced seismicity for reservoir characterization, *PhD thesis*, Massachusetts Institute of Technology, Department of Earth, Atmospheric, and Planetary Sciences.

Li, J., Kuehl, H., Droujinine, A. & Blokland, J.-W., 2016. *Microseismic and Induced Seismicity Simultaneous Location and Moment Tensor Inversion: Moving beyond Picks with a Robust Full-Waveform Method,* pp. 2535–2539, Society of Exploration Geophysicists.

Li, J., Ji, S., Li, Y., Qian, Z. & Lu, W., 2018. Downhole microseismic signal-to-noise ratio enhancement via strip matching shearlet transform, *J. Geophys. Eng.,* **15**(2), 330–337.

Li, L. *et al.*, 2020. Recent advances and challenges of waveform-based seismic location methods at multiple scales, *Rev. Geophys.,* **58**(1), e2019RG000667 , doi:10.1029/2019RG000667.

Li, Z., Zhu, L., Officer, T., Shi, F., Yu, T. & Wang, Y., 2022. A machine-learning-based method of detecting and picking the first P-wave arrivals of acoustic emission events in laboratory experiments, *Geophys. J. Int.,* **230**(3), 1818–1823.

Liu, E., Zhu, L., Govinda Raj, A., McClellan, J.H., Al-Shuhail, A., Kaka, S.I. & Iqbal, N., 2017. Microseismic events enhancement and detection in sensor arrays using autocorrelation-based filtering, *Geophys. Prospect.,* **65**(6), 1496–1509.

Liu, H., Ong, Y.-S., Shen, X. & Cai, J., 2018. *When Gaussian Process Meets Big Data: A Review of Scalable GPs,* preprint (arXiv:1807.01065).

Lomax, A., Virieux, J., Volant, P. & Berge-Thierry, C., 2000. *Probabilistic Earthquake Location in 3D and Layered Models,* pp. 101–134, Springer Netherlands.

Lomax, A., Michelini, A. & Curtis, A., 2009. *Earthquake Location, Direct, Global-Search Methods,* pp. 1–33, Springer New York.

Maas, A.L., Hannun, A.Y. & Ng, A.Y., 2013. Rectifier nonlinearities improve neural network acoustic models, in *ICML Workshop on Deep Learning for Audio, Speech and Language Processing (Proc. ICML),* **30, 1,** 3, https://www.bibsonomy.org/bibtex/2479ee3123c43b4a7eb5a329d93e7360f/t_seizinge

Majer, E.L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B. & Asanuma, H., 2007. Induced seismicity associated with enhanced geothermal systems, *Geothermics,* **36**(3), 185–222.

Moseley, B., Markham, A. & Nissen-Meyer, T., 2020. *Solving the wave equation with physics-informed deep learning,* preprint (arXiv:2006.11894).

Moser, T.J., van Eck, T. & Nolet, G., 1992. Hypocenter determination in strongly heterogeneous earth models using the shortest path method, *J. geophys. Res.,* **97**(B5), 6563–6572.

Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. & Beroza, G.C., 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nat. Commun.,* **11**(1), 1–12.

Mukuhira, Y., Asanuma, H., Ito, T. & Häring, M.O., 2016. Physics-based seismic evaluation method: Evaluating possible seismic moment based on microseismic information due to fluid stimulation, *Geophysics,* **81**(6), KS195–KS205.

Noack, M.M. & Clark, S., 2017. Acoustic wave and eikonal equations in a transformed metric space for various types of anisotropy, *Heliyon,* **3**(3), e00260 , doi:10.10.1016/j.heliyon.2017.e00260.

Papoulis, A., Pillai, S. & Pillai, S., 2002. *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill electrical and electronic engineering series, McGraw-Hill.

Pratt, R., 1999. Seismic waveform inversion in the frequency domain; Part 1, theory and verification in a physical scale model, *Geophysics,* **64,** 888–901.

Pratt, R. & Shipp, R., 1999. Seismic waveform inversion in the frequency domain, Part 2: fault delineation in sediments using crosshole data, *Geophysics,* **64,** 902–914.

Pugh, D.J., White, R.S. & Christie, P. A.F., 2016. A Bayesian method for microseismic source inversion, *Geophys. J. Int.,* **206**(2), 1009–1038.

Raissi, M., Perdikaris, P. & Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.,* **378,** 686–707.

Rajaratnam, B. & Sparks, D., 2015. *MCMC-Based Inference in the Era of Big Data: A Fundamental Analysis of the Convergence Complexity of High-Dimensional Chains,* preprint (arXiv:1508.00947).

Rasmussen, C.E. & Williams, C.K.I., 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning),* The MIT Press.

Ross, Z.E., Meier, M.-A. & Hauksson, E., 2018. P wave arrival picking and first-motion polarity determination with deep learning, *J. geophys. Res.,* **123**(6), 5120–5129.

Sethian, J.A., 1996. A fast marching level set method for monotonically advancing fronts, *Proc. Natl. Acad. Sci. USA,* **93**(4), 1591–1595.

Shapiro, S.A., Dinske, C., Langenbruch, C. & Wenzel, F., 2010. Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations, *Leading Edge,* **29**(3), 304–309.

Skilling, J., 2006. Nested sampling for general Bayesian computation, *Bayesian Anal.,* **1**(4), 833–859.

Smith, J., 2019. Geomechanical properties of the Groningen reservoir, *PhD thesis*, University of Cambridge, Department of Earth Sciences - Bullard Laboratories.

Smith, J.D., Azizzadenesheli, K. & Ross, Z.E., 2020. *EikoNet: Solving the Eikonal equation with Deep Neural Networks,* preprint (arXiv:2004.00361).

Song, F. & Toksöz, M.N., 2011. Full-waveform based complete moment tensor inversion and source parameter estimation from downhole microseismic data for hydrofracture monitoring, *Geophysics,* **76**(6), WC103–WC116.

Spurio Mancini, A., Piras, D., Ferreira, A. M.G., Hobson, M.P. & Joachimi, B., 2021. Accelerating Bayesian microseismic event location with deep learning, *Solid Earth Discuss.,* **2021,** 1–36.

Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B. & Hobson, M.P., 2022. CosmoPower: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys, *Mon. Not. R. astr. Soc.,* **511**(2), 1771–1788.

Stähler, S.C. & Sigloch, K., 2014. Fully probabilistic seismic source inversion–part 1: efficient parameterisation, *Solid Earth,* **5**(2), 1055–1069.

Stähler, S.C. & Sigloch, K., 2016. Fully probabilistic seismic source inversion–part 2: modelling errors and station covariances, *Solid Earth,* **7**(6), 1521–1536.

Sunnåker, M., Busetto, A.G., Numminen, E., Corander, J., Foll, M. & Dessimoz, C., 2013. Approximate Bayesian computation, *PLOS Comput. Biol.,* **9**(1), 1–10.

Tao, Y. & Sen, M.K., 2013. Frequency-domain full waveform inversion with a scattering-integral approach and its sensitivity analysis, *J. geophys. Eng.,* **10,** 065008.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Society for Industrial and Applied Mathematics.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.,* **20**(2), 219–232.

Thornton, M., 2013. Velocity uncertainties in surface and downhole monitoring, *Conference Proceedings, 4th EAGE Passive Seismic Workshop* , doi:10.10.3997/2214-4609.20142344.

Treeby, B.E., Jaros, J., Rohrbach, D. & Cox, B.T., 2014. Modelling elastic wave propagation using the k-Wave MATLAB Toolbox, in *2014 IEEE International Ultrasonics Symposium,* pp. 146–149, doi: 10.1109/ULTSYM.2014.0037.

Usher, P., Angus, D. & Verdon, J., 2013. Influence of a velocity model and source frequency on microseismic waveforms: some implications for microseismic locations, *Geophys. Prospect.,* **61**(s1), 334–345.

Vasco, D.W., Nakagawa, S., Petrov, P. & Newman, G., 2019. Rapid estimation of earthquake locations using waveform traveltimes, *Geophys. J. Int.,* **217**(3), 1727–1741.

Vavryčuk, V., 2001. Inversion for parameters of tensile earthquakes, *J. geophys. Res.,* **106**(B8), 16339–16355.

Vavryčuk, V., 2005. Focal mechanisms in anisotropic media, *Geophys. J. Int.,* **161**(2), 334–346.

Vavryčuk, V., 2015. Moment tensor decompositions revisited, *J. Seismol.,* **19**(1), 231–252.

Virtanen, P. *et al.*, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nat. Methods,* **17,** 261–272.

White, M. C.A., Fang, H., Nakata, N. & Ben-Zion, Y., 2020. PyKonal: A Python Package for Solving the Eikonal Equation in Spherical and Cartesian Coordinates Using the Fast Marching Method, *Seismol. Res. Lett.,* **91**(4), 2378–2389.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J.-W., Das, I. & Droujinine, A., 2019. Full-waveform event location and moment tensor inversion for induced seismicity, *Geophysics,* **84**(2), KS39–KS57.

Wittlinger, G., Herquel, G. & Nakache, T., 1993. Earthquake location in strongly heterogeneous media, *Geophys. J. Int.,* **115**(3), 759–777.

Wuestefeld, A., Greve, S.M., Näsholm, S.P. & Oye, V., 2018. Benchmarking earthquake location algorithms: a synthetic comparison, *Geophysics,* **83**(4), KS35–KS47.

Xu, Y., Li, J. & Chen, X., 2019. *Physics Informed Neural Networks for Velocity Inversion,* pp. 2584–2588, Society of Exploration Geophysicists.

Yao, Y., Rosasco, L. & Caponnetto, A., 2007. On early stopping in gradient descent learning, *Constructive Approx.,* **26,** 289–315.

Zhang, J., Dong, L. & Xu, N., 2020. Noise suppression of microseismic signals via adaptive variational mode decomposition and akaike information criterion, *Appl. Sci.,* **10**(11), 3790 , doi:10.3390/app10113790.

Zheng, J., Lu, J., Peng, S. & Jiang, T., 2018. An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks, *Geophys. J. Int.,* **212**(2), 1389–1397.

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z. & Fang, L., 2019. Deep learning for seismic phase detection and picking in the aftershock zone of 2008 M_w7.9 Wenchuan Earthquake, *Phys. Earth planet. Inter.,* **293,** 106261 , doi:10.1016/j.pepi.2019.05.004.

Zhu, W. & Beroza, G.C., 2018. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.,* **216**(1), 261–273.

Zhu, X., Vondrick, C., Fowlkes, C. & Ramanan, D., 2015. *Do We Need More Training Data?,* preprint (arXiv:1503.01508).

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Typical example of training (solid blue) and validation (dashed coral) loss curves for one receiver in the CLVD case. The vertical dashed black line indicates the epoch with minimum validation loss, which we interpret as corresponding to the best model.
Figure S2. Marginalized 68 per cent credibility contours obtained with our method for a source located at $(x, y, z) =$ (0.20 km, 0.43 km, 0.99 km), indicated by the dashed black lines, comparing different numbers of training data for the emulator (2000, 5000 and 8000). We show these results for three source mechanisms: ISO, DC and CLVD. Note that we are considering 23 receivers and a signal-to-noise ratio of 33 dB. This figure is best viewed in colour.

Figure S3. Marginalized 68 per cent credibility contours obtained with our method for a source located at $(x, y, z) = (0.46\ \text{km}, 0.34\ \text{km}, 1.48\ \text{km})$, indicated by the dashed black lines, comparing different dispositions of the receivers. In particular, 5 receivers refer to the blue squares in Fig. 2 in the main text, and nine receivers refer to the green circles in Fig. 2 in the main text. We show these results for three source mechanisms: ISO, DC and CLVD. Note that we are considering 8000 training data for the emulator and a signal-to-noise ratio of 33 dB. This figure is best viewed in colour.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.