

# Near-real-time design of experiments for seismic monitoring of volcanoes

Dominik Strutz\*<sup>1</sup>, Andrew Curtis<sup>1</sup>

<sup>1</sup>School of GeoSciences, University of Edinburgh, Edinburgh, UK

**Abstract** Monitoring the seismic activity of volcanoes is crucial for hazard assessment and eruption forecasting. The layout of each seismic network determines the information content of recorded data about volcanic earthquakes, and experimental design methods optimise sensor locations to maximise that information. We provide a code package that implements Bayesian experimental design to optimise seismometer networks to locate seismicity at any volcano, and a practical guide to make this easily and rapidly implementable by any volcano seismologist. This work is the first to optimise travel-time, amplitude and array source location methods simultaneously, making it suitable for a wide range of volcano monitoring scenarios. The code-package is designed to be straightforward to use and can be adapted to a wide range of scenarios, and automatically links to existing global databases of topography and properties of volcanoes worldwide to allow rapid deployment. Any user should be able to obtain an initial design within minutes using a combination of generic and volcano-specific information to guide the design process, and to refine the design for their specific scenario within hours, if more specific prior information is available.

## 1 Introduction

Forecasting episodes of volcanic unrest is a significant challenge due to the highly variable characteristics of volcanic systems (Chiodini et al., 2016; Selva et al., 2015; Marzocchi and Bebbington, 2012; Sigmundsson et al., 2010). Monitoring volcanoes serves two main purposes: first, to understand the structure and dynamics of each volcano, as this is crucial for hazard assessment, eruption forecasting, and early warning, and second, for risk mitigation at times of volcanic unrest (Pallister and McNutt, 2015). Monitoring allows accelerating rates of seismic activity to be detected, which often occurs before eruptions (Sigmundsson et al., 2010; Saltogianni et al., 2014; Cannavò et al., 2015), and the location of volcano-tectonic earthquakes to be tracked in near real-time, which may facilitate eruption forecasting (Sigmundsson et al., 2015; Pallister and McNutt, 2015; Falsaperla and Neri, 2015). In addition, long-period earthquakes and micro-earthquakes can be key indicators of imminent eruptions, especially during extended eruptive sequences (McNutt, 2005). In all cases, the location of seismic sources is crucial to allow the underlying processes to be better understood.

Seismicity is usually monitored by a sparse network of seismometers deployed around the volcano. The layout of this network (the experimental design) determines the information content of recorded data, and therefore the quality of derived or interpreted information. The goal of the experimental design process is to optimise the network layout, and sensor types to maximise the information that can be gained from the data (Maurer et al., 2010; Strutz and Curtis, 2023). Despite the

importance of experimental design, and the fact that optimal design methods have existed for decades (e.g., Steinberg et al., 1995), in practice, seismic networks are rarely optimised taking the known physics and uncertainty of the problem into account. This is likely to be due to a lack of awareness of the potential for design methods to enhance survey results, a lack of expertise in understanding the range of design methods and their relative merits (Bloem et al., 2020), or a lack of time in rapid deployment operations.

The goal of this paper is to provide a practical guide to the implementation of Bayesian experimental design for volcano seismic monitoring, and a code package that makes this easily and rapidly implementable by any volcano seismologist. This work is the first to take travel-time, amplitude and array source location methods into account simultaneously, making it suitable for a wide range of volcano monitoring scenarios. The code package is implemented in Python, is available as open source software, and makes use of existing global databases of topography and properties of volcanic systems to provide designs rapidly. The software is designed to be straightforward to use and can easily be adapted to a wide range of scenarios. Any user should be able to obtain a first design within minutes using a combination of generic and volcano specific information to guide the design process, and be able to refine the design for their specific scenario within hours if they have more specific prior information available.

In this paper we first introduce the fundamentals of Bayesian experimental design in the context of monitoring volcano induced seismicity in section 2. We present an example design process in section 3 where we show how to define the prior distribution, the physical model,

\*Corresponding author: dstrutz@ed.ac.uk

how to calculate the expected information gain, and how to optimise the experimental design. We then discuss how to interpret the results, and some advanced topics such as the use of heterogenous velocity models in section 4.

## 2 Methods

This section introduces fundamental theory for Bayesian experimental design for locating and monitoring volcano-induced seismicity. The focus of this paper is to provide a practical guide to the implementation of design methods in real-world scenarios. To keep this section concise and suitable for a broad audience, we do not go into theoretical details, but we refer the reader to relevant literature for a more fundamental introduction.

### 2.1 Bayesian Experimental Design

When designing an experiment we aim to find a set of observation points described by a vector  $\xi$  (the design) that maximises the amount of information that we expect to gain by collecting data at these points. In the context of volcano seismic monitoring, the goal is usually to find a set of station/array locations that maximises the expected information about the location of seismic sources. To account for the full uncertainty and non-linearity of the problem, we use a Bayesian framework for the experimental design process.

#### 2.1.1 Bayesian Inference

Before introducing the experimental design problem, we briefly introduce Bayesian inference. The goal of Bayesian inference is to infer the uncertainty in model parameters  $m$  given some data  $d$  observed using the experimental design  $\xi$ . This uncertainty is characterised by a so-called posterior probability distribution  $p(m | d, \xi)$ . Bayes' theorem states that the posterior distribution can be calculated by

$$p(\mathbf{m} | \mathbf{d}, \xi) = \frac{p(\mathbf{d} | \mathbf{m}, \xi) p(\mathbf{m})}{p(\mathbf{d} | \xi)} \quad (1)$$

where  $p(d | m, \xi)$  is the likelihood of recording data  $d$  given the particular set of parameter values defined by  $m$ ,  $p(m)$  is the prior probability of  $m$ , and  $p(d | \xi)$  is called the evidence of data  $d$ . All terms that involve observed data also depend on design  $\xi$ .

The likelihood is the probability of observing the recorded data given the model parameter values (i.e., assuming that those values are true) and the particular experimental design that was deployed. In our case, the likelihood is the probability of observing a set of travel-times, amplitudes, or array measurements given a source location and the experimental design. The likelihood is assumed to be a multivariate Gaussian, where the mean is the data predicted using the forward model, and the covariance matrix is diagonal with appropriate entries for expected measurement uncertainties given the data types and receiver/array locations in  $\xi$ .

The prior distribution describes where we believe seismicity is most likely to occur in the subsurface before we have observed any data. To make the experimental design process as general as possible, the prior distribution in this study is defined on a discrete grid that spans a subset of the subsurface around each volcano. The distribution is assumed to be uniform within each grid cell. This description makes it straightforward to define the prior distribution, and to sample from it. In section 3.0.2, we show how this grid can be defined and refined using readily available information about any volcano of interest.

#### 2.1.2 Bayesian Experimental Design

The Bayesian framework makes it straightforward to estimate the expected increase in information to be gained by collecting data at a specific set of observation points. This expected information gain (EIG) can be expressed as the average of the information described by the posterior distribution  $p(m | d, \xi)$  compared to the information in the prior distribution  $p(m)$  (Lindley, 1956):

$$\text{EIG} = \mathbb{E}_{p(d|\xi)} \left\{ I[p(\mathbf{m} | \mathbf{d}, \xi)] - I[p(\mathbf{m})] \right\} \quad (2)$$

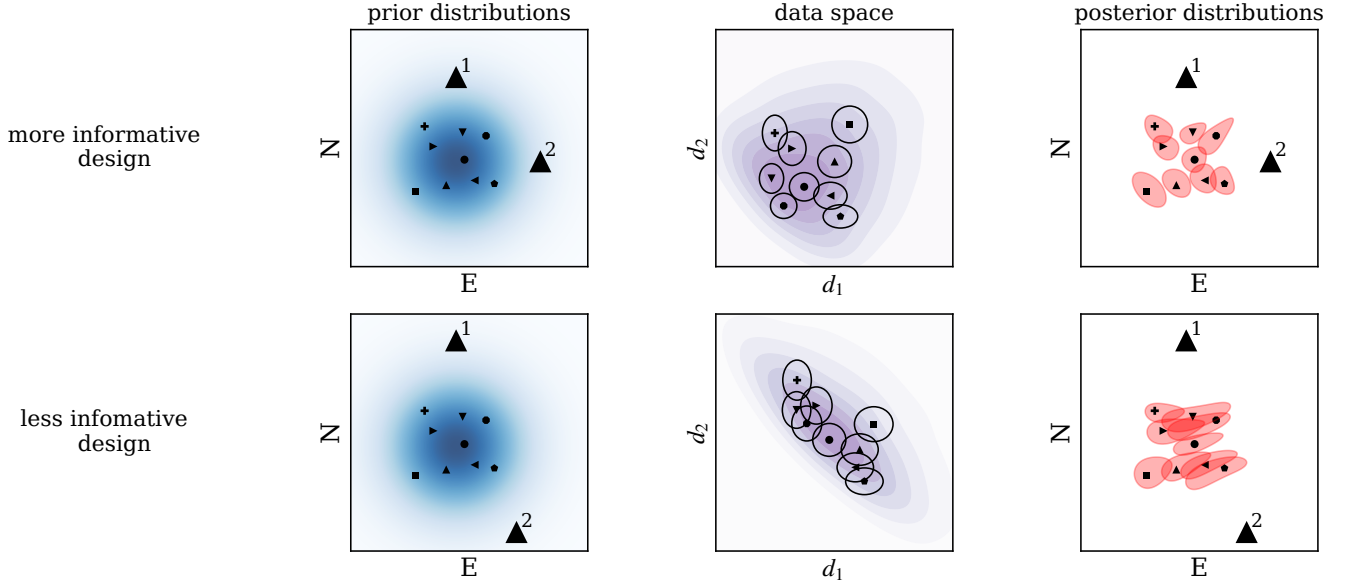
where  $I[p]$  is the Shannon information (Shannon, 1948) of distribution  $p$  (for more information see appendix A), and  $\mathbb{E}_{p(d|\xi)}$  is the expectation over all possible data that is likely to be observed given the available prior information. Intuitively, equation 2 states that the information gain is the reduction in uncertainty in source locations  $m$  given observations  $d$  at stations/arrays  $\xi$ , compared to their uncertainty given only the prior information; since no observed data is usually available for the design process we average over all possible observations (or a representative subset) that are likely to be recorded. We obtain the subset of likely observations by forward modelling data for a set of earthquake locations sampled from the prior distribution, and adding simulated observational noise described by the data likelihood.

Using the EIG as our design objective function, the best design can be expressed mathematically as

$$\xi^* = \arg \max_{\xi \in \Xi} \text{EIG}(\xi) \quad (3)$$

where  $\Xi$  is the set of all possible experimental designs, and  $\arg \max$  is mathematical shorthand for finding the design that maximises the EIG with respect to  $\xi$ . In this work, we use a genetic algorithm (Holland, 1975; Sambridge and Gallagher, 1993; Gad, 2023) to optimise the experimental design; this is a population-based optimisation algorithm that operates analogously to processes of natural selection amongst a set of potential designs. It is well suited for experimental design problems since it can optimise designs with mixed data types and constraints. The genetic algorithm is a stochastic (pseudo-random) algorithm, which means that the results can vary between runs. We use a fixed seed for the random number generator to ensure that the results are reproducible.

To evaluate the EIG in the form of Equation (2), we would need to calculate the posterior distribution



**Figure 1** Illustration of the expected information gain (EIG) in the Bayesian experimental design process. The left panel shows a set of models sampled from the prior distribution; the middle panel shows the corresponding observed data and its likelihood (two standard deviation ellipses), and in the background the contours of the evidence calculated using a large number of data samples; and the right panel shows a contour of the posterior distribution for each model. Each model and data sample has a different symbol to allow corresponding pairs of models and data samples to be identified; in the left and middle panels deeper colours represent higher probabilities.

$p(\mathbf{m} | \mathbf{d}, \boldsymbol{\xi})$  for each possible observation  $\mathbf{d}$ , which is computationally very expensive (unless we approximate the posterior distribution (Foster et al., 2019)). To make the problem tractable we rearrange the EIG to explicitly depend on the evidence  $p(\mathbf{d} | \boldsymbol{\xi})$  and the likelihood  $p(\mathbf{d} | \mathbf{m}, \boldsymbol{\xi})$  instead of the posterior distribution  $p(\mathbf{m} | \mathbf{d}, \boldsymbol{\xi})$  and the prior distribution  $p(\mathbf{m})$  (details in appendix B). The EIG can then be expressed as:

$$\text{EIG}(\boldsymbol{\xi}) = \mathbb{E}_{p(\mathbf{m})} \left\{ I[p(\mathbf{d} | \mathbf{m}, \boldsymbol{\xi})] \right\} - I[p(\mathbf{d} | \boldsymbol{\xi})] \quad (4)$$

where the expectation is now over the prior distribution  $p(\mathbf{m})$ .

We can evaluate the EIG in equation 4 with the nested Monte-Carlo (NMC) method (Ryan, 2003). We use the computationally more efficient variation of Huan and Marzouk (2013) which is defined as follows:

$$\begin{aligned} \text{EIG}_{\text{NMC}} &= \frac{1}{N} \sum_{i=1}^N \left\{ \log [p(\mathbf{d}_i | \mathbf{m}_i, \boldsymbol{\xi})] - \log [p(\mathbf{d}_i | \boldsymbol{\xi})] \right\} \\ &= \frac{1}{N} \sum_{i=1}^N \left\{ \log [p(\mathbf{d}_i | \mathbf{m}_i, \boldsymbol{\xi})] - \right. \\ &\quad \left. \log \left[ \frac{1}{N} \sum_{j=1}^N p(\mathbf{d}_i | \mathbf{m}_j, \boldsymbol{\xi}) \right] \right\} \end{aligned} \quad (5)$$

where we sample  $N$  models  $\{\mathbf{m}_i\}_{i=1}^N$  from the prior distribution  $p(\mathbf{m})$ , calculate  $N$  data vectors  $\{\mathbf{d}_i\}_{i=1}^N$ , one vector for each sampled model, and then approximate the EIG of Equation (4) by replacing the expectation with the average  $\frac{1}{N} \sum_{i=1}^N$  over the  $N$  samples.

Calculating the EIG using the NMC method, while unbiased, is computationally expensive. During the de-

sign optimisation process we must evaluate this quantity many times, which can become intractable. We therefore invoke the  $D_N$  method (Coles and Curtis, 2011) in which we approximate  $p(\mathbf{d} | \boldsymbol{\xi})$  by a multivariate Gaussian distribution. This results in the following expression for the EIG:

$$\begin{aligned} \text{EIG}_{\text{DN}} &= \frac{1}{N} \sum_{i=1}^N \log [p(\mathbf{d}_i | \mathbf{m}_i, \boldsymbol{\xi})] - \\ &\quad \frac{3}{2} (1 + \log(2\pi)) + \frac{1}{2} \log |\mathbf{C}_d| \end{aligned} \quad (6)$$

where we again sample  $N$  models  $\{\mathbf{m}_i\}_{i=1}^N$  from the prior distribution  $p(\mathbf{m})$  and calculate corresponding data samples  $\{\mathbf{d}_i\}_{i=1}^N$ . Instead of a nested loop, we now calculate the information in the evidence by making use of the fact that the information in a multivariate Gaussian distribution is given by

$$I[p(\mathbf{d}_i | \boldsymbol{\xi})] = \frac{k}{2} (1 + \log(2\pi)) + \frac{1}{2} \log |\mathbf{C}_d| \quad (7)$$

where  $k$  is the dimension of the data space and  $\mathbf{C}_d$  is the covariance matrix of the data samples  $\{\mathbf{d}_i\}_{i=1}^N$ . The  $D_N$  method is computationally far more efficient than the NMC method, but offers an upper bound for the true EIG and therefore overestimates the EIG. Despite this bias, the method has been shown to perform well in the design of seismic monitoring networks since only relative values of the EIG for different designs are important (Bloem et al., 2020). Thus, unless otherwise stated, in this work we use the  $D_N$  method to calculate the EIG during the design optimisation process, and use the NMC method to validate and interpret the results.

To provide a more intuitive understanding of the EIG, especially of its formulation in data space, we consider

a schematic example, in which we demonstrate how more or less informative designs affect terms in Equation (4). The first step is to sample a set of model samples from the prior distribution  $p(\mathbf{m})$ , shown in the left panel of Figure 1. Then we calculate the data samples (including a realisation of measurement noise described by the likelihood) that would be observed for each model: here, the distance between model and receiver locations, which is equivalent (proportional) to a travel-time from each source to receivers in a homogeneous medium, as shown in the middle panel of Figure 1. The density of the data sample realisations allows the evidence (calculated using a large number of data samples in the background and shown as purple contours in the middle panel of Figure 1) to be visualised. Each data point has an associated likelihood, displayed as the two standard deviation ellipse. This shows graphically how different designs affect the evidence, the likelihood, and, therefore, the EIG.

The term

$$\mathbb{E}_{p(\mathbf{m})} I[p(\mathbf{d} | \mathbf{m}, \boldsymbol{\xi})] = \frac{1}{N} \sum_{i=1}^N \log \{p(\mathbf{d}_i | \mathbf{m}_i, \boldsymbol{\xi})\} \quad (8)$$

computes the expected information content in the data likelihood by averaging the probability of observing the data samples given the (true) source location model samples. This information content will be maximised for small standard deviations and minimised for large standard deviations (small and large ellipses, respectively, in the middle panel of Figure 1). Since we maximise this quantity, the standard deviation of the data samples should be as small as possible.

The evidence term

$$I[p(\mathbf{d} | \boldsymbol{\xi})] = \frac{1}{N} \sum_{i=1}^N \log \left[ \frac{1}{N} \sum_{j=1}^N p(\mathbf{d}_i | \mathbf{m}_j, \boldsymbol{\xi}) \right] \quad (9)$$

or its approximation

$$I[p(\mathbf{d} | \boldsymbol{\xi})] \approx \frac{3}{2} (1 + \log(2\pi)) + \frac{1}{2} \log |C_d| \quad (10)$$

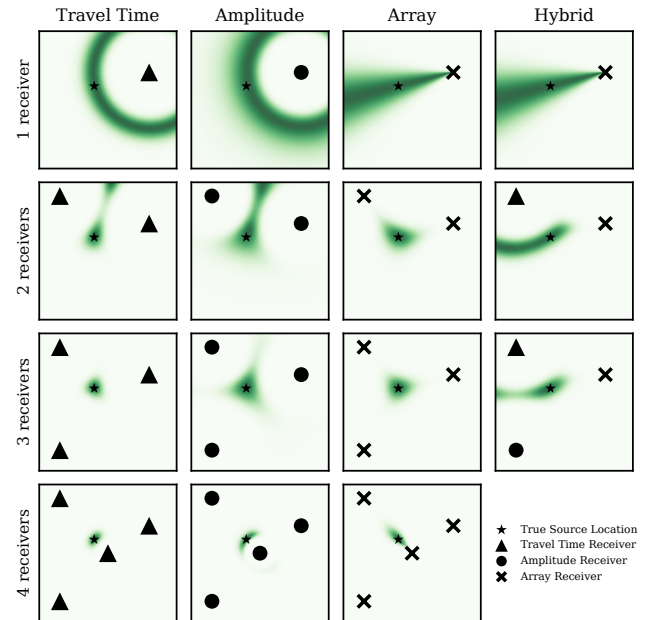
calculates the information content in the evidence, which here is proportional to the inverse of the spread of the data samples (indicated by contours calculated using a large number of samples in the background). A large spread is preferable since it indicates that the data samples are well spread out and can be more easily distinguished from each other (or more formally, that the information in the evidence is low).

The EIG is the difference between the result of evaluating equation 8 and equation 9 or its approximation in equation 10, which is the difference between the average likelihood of the data samples and the measure of the inverse of the spread of the data samples. If data samples are closer together, this means that for fixed data values measured by an experiment, it is hard to distinguish between the source location model that generated those data, and other source location samples - which leads to a low EIG; intuitively this would indicate that the posterior distribution is less spread-out, as shown in the bottom right panel of Figure 1. On the

other hand, if a different experimental design caused the data samples to be more spread out such that their uncertainties do not overlap, the EIG will be high and each source location can be discriminated from the others given a measured data set. This is clearly illustrated in the top right panel of Figure 1.

## 2.2 Seismic Source Location Methods

Volcanoes are complex systems driven by a wide variety of magmatic and hydrothermal processes, and this leads to a wide range of characters of seismic activity that can be detected at the surface (Zobin, 2017). Locating these seismic sources is a crucial step in understanding the underlying processes. For the observed signals, the characteristic periods span  $10^{-1}$  to  $10^2$  s and corresponding wavelengths span  $10^2$  to  $10^5$  m (Saccorotti and Lokmer, 2021). This wide range of scales leads to a variety of methods for locating seismic sources.



**Figure 2** Example of likelihoods (dark green indicates high likelihood) for travel-time data from a synthetic true event for up to four stations at the same locations for different data types. For details on the different data types see section 2.2. The hybrid column shows how the likelihoods of different data types (represented by symbols) can be combined. In row 1 the hybrid and array results are identical since only a single measurement type is available.

This paper focuses exclusively on data derived from waveforms rather than on waveform data themselves, since the optimal design process and the forward model for full waveforms are computationally costly. Additionally, waveform-based methods are often very sensitive to the subsurface velocity model (e.g. O'Brien et al., 2011); this model is typically only accurately known for already well-monitored volcanoes, for which the presented algorithms may be of more limited use.



### 2.2.1 Travel Time based Methods

Traveltime-based methods are the most common for locating seismic sources and have been widely used in volcano monitoring (Lomax et al., 2001; Saccorotti et al., 2007; Woods et al., 2019). The travel-time is the time taken for a seismic phase (e.g., P-wave, S-wave) to travel from the source to the receiver, and depends both on the locations of the source and receiver, and on the velocity model. If we assume a homogeneous velocity model with a velocity  $v$  which represents the best estimate of the average velocity within the three dimensional region of interest, the observed arrival-time  $t$  of the seismic waves can be calculated as:

$$t = t_{\text{src}} + \frac{1}{v} \left[ (x_{\text{rec}} - x_{\text{src}})^2 + (y_{\text{rec}} - y_{\text{src}})^2 + (z_{\text{rec}} - z_{\text{src}})^2 \right]^{\frac{1}{2}} \quad (11)$$

where  $t_{\text{src}}$  is the origin time,  $(x_{\text{src}}, y_{\text{src}}, z_{\text{src}})$  is the source location, and  $(x_{\text{rec}}, y_{\text{rec}}, z_{\text{rec}})$  is the receiver location. If we assume that there is little prior information on  $t_{\text{src}}$  (the prior distribution is assumed to be uniform within certain spatial boundaries), the origin time can be integrated out since we assume a Gaussian data likelihood (Tarantola et al., 1982; Lomax et al., 2001). The model parameters are then only the source location  $(x_{\text{src}}, y_{\text{src}}, z_{\text{src}})$ , and arrival time observations become equivalent to travel time data. We treat the case of a heterogeneous velocity model separately in section 4, since it requires a more complex forward model.

Uncertainty in the travel-time is expressed as the combination of errors in picking the arrival times of each seismic phase given the seismic waveform, and the contribution from the velocity model uncertainty which arises from the true seismic energy travelling through a heterogeneous medium which leads to a change in travel-time. This forward model error is constant in inverse problems but in experimental design problems it is unknown so we account for it through an uncertainty in the travel-time. A combined standard deviation of  $\sigma_t$  can be calculated as:

$$\sigma_t(t)^2 = \sigma_p^2 + t \cdot \sigma_v^2 \quad (12)$$

where  $\sigma_p$  is the picking uncertainty and  $\sigma_v$  is the relative velocity model uncertainty scaled by the travel-time  $t$ . Scaling the squared velocity model uncertainty by the travel-time is inspired by a Gaussian random walk model in which the mean squared distance (here travel-time noise) from the reference point (here the mean value) is proportional to the time. In a volcano setting the velocity model uncertainty term is typically much larger than the picking uncertainty term. To define it site specific information is necessary, but a value of 0.1 to 0.2 is a reasonable starting point as it corresponds to an average relative velocity model standard deviation of 10% to 20% from the specified homogeneous velocity model. This would result in a standard deviation of 0.1 s to 0.2 s for a travel-time of 1 s. Tests show that while the EIG and derived quantities are sensitive to the value of  $\sigma_v$ , the resulting optimal design is usually not particularly sensitive to it.

Column one of Figure 2 shows the resulting likelihood for travel-time data from a synthetic true event for up to

four stations. For a single station, the uncertainty spans a circle of equal travel-time. For multiple stations, the likelihood is the product of the likelihoods of the individual stations included in the design.

Due to their emergent onset and lack of clear phase arrivals, long-period and tremor signals can not be located using travel-times based on phase arrival picks. However, several other methods, such as coherence based methods (Ohminato et al., 1998; Kawakatsu et al., 2000; Dawson et al., 2004; Almendros, 2003; Saccorotti and Lokmer, 2021) and back-propagation methods (Kao and Shan, 2004; Langet et al., 2014) implicitly use travel-time information to locate sources. Therefore, optimising a network for travel-time observations also provides an optimal network for those methods.

### 2.2.2 Amplitude based Methods

Another common method for locating volcano-induced seismicity is amplitude source location (ASL), in which an attenuation model is assumed for a single wave type (surface or body waves) propagating isotropically in a homogeneous medium (Saccorotti and Lokmer, 2021; Battaglia and Aki, 2003; Taisne et al., 2011; Ogiso and Yomogida, 2012; Kumagai et al., 2013, 2011, 2009; Ogiso et al., 2015; Yamasato, 1997; Jolly et al., 2002; Caudron et al., 2018; Carbone et al., 2008). The source location and strength can then be estimated by comparing observed amplitudes with those predicted from the attenuation model. If we ignore corrections for measurement site amplification, the observed amplitude  $A$  decays as (Kumagai et al., 2013; Morioka et al., 2017):

$$A = \frac{1}{r} \exp(-C \cdot t) \quad (13)$$

where  $r$  is the length of the ray path, and  $C = \pi \frac{f}{Q}$  is the attenuation coefficient which depends on the frequency  $f$ , the quality factor  $Q$ , and the travel-time  $t$  of the seismic phase. The length of the ray path  $r$  as well as the travel-time  $t$  depend on the source location  $(x_{\text{src}}, y_{\text{src}}, z_{\text{src}})$  and receiver location  $(x_{\text{rec}}, y_{\text{rec}}, z_{\text{rec}})$ .

We model the uncertainty in the amplitude through linearised error propagation (Ku, 1966) of the uncertainty terms  $\sigma_r$ ,  $\sigma_t$ , and  $\sigma_Q$  for the ray path, travel-time and quality factor, respectively. The final uncertainty in the amplitude  $\sigma_A$  is typically dominated by uncertainty in  $\sigma_Q$ . Internally, the final amplitude measurements and uncertainties are converted to log-space, which makes the Normal approximation in the  $D_N$  method more accurate.

Column two in Figure 2 shows the resulting likelihood for amplitude data from a synthetic true event for up to four stations. While at first glance, this looks similar to the travel-time likelihood, the uncertainty is asymmetric in the radial direction due to the log-space conversion. For multiple stations the likelihood is again the product of the likelihoods of the individual stations.

### 2.2.3 Array Methods

For the purposes of this paper we define an array as seismic array as a set of seismometers with inter-

station spacings smaller than the wavelengths of interest, which typically have an aperture (largest distance between two stations) significantly smaller than the source-to-array distance (Saccorotti and Lokmer, 2021). This is done in order to discriminate them from a network of seismometers that are processed independently of each other. For this study, we treat an array as an instrument at a single location which can measure a back azimuth  $\theta$  and an incident angle  $i$  (or alternatively, the two components  $p_x$  and  $p_y$  of the horizontal slowness vector). The back azimuth is the angle between the propagation direction of the arriving wavefront and the north direction, while the incident angle is the angle between propagation direction and the vertical direction. These quantities can be estimated straightforwardly using the angles between the source and the array for homogeneous models (or the gradient of the local travel-time grid for heterogeneous velocity models). In practice, the back azimuth and incident angle can be calculated using beamforming (e.g. Rost and Thomas, 2002; Leva et al., 2022), MUSIC (e.g. Inza et al., 2011), or other methods (e.g. Métaixian et al., 2002; Di Lieto et al., 2007).

The uncertainty of the array measurements can either be defined through the standard deviation of the two angular measurements  $\sigma_\theta$  and  $\sigma_i$  or through the standard deviation of the two components of the slowness vector  $\sigma_{p_x}$  and  $\sigma_{p_y}$ . The latter implies that there is an intuitive way to convert an array function to an uncertainty estimate. Within the code package, the data used are  $p_x$  and  $p_y$ , so  $\sigma_\theta$  and  $\sigma_i$  are converted to  $\sigma_{p_x}$  and  $\sigma_{p_y}$  using linearised error propagation.

Column three of Figure 2 shows the resulting likelihood for array data from a synthetic true event for up to four arrays. It is clear that this data type imposes substantially different constraints compared to travel-time and amplitude data: instead of a likelihood focused around a circle of equal travel-time or amplitude, we instead obtain a cone representing the uncertainty in the arrival direction of the seismic energy. And as before, the likelihood from multiple arrays is the product of the likelihoods of the individual arrays. Since the incident angle is very sensitive to the local seismic velocity and can change considerably for a gradient/heterogeneous velocity model, it is also possible to use the back azimuth only. In this case, only  $\sigma_\theta$  is used as the uncertainty in the array measurements.

#### 2.2.4 Hybrid Data Types

The fourth column in Figure 2 shows how likelihoods of the different data types can be combined. As above, these are simply point-by-point multiplications of the likelihoods of each sensor.

#### 2.2.5 Important Assumptions

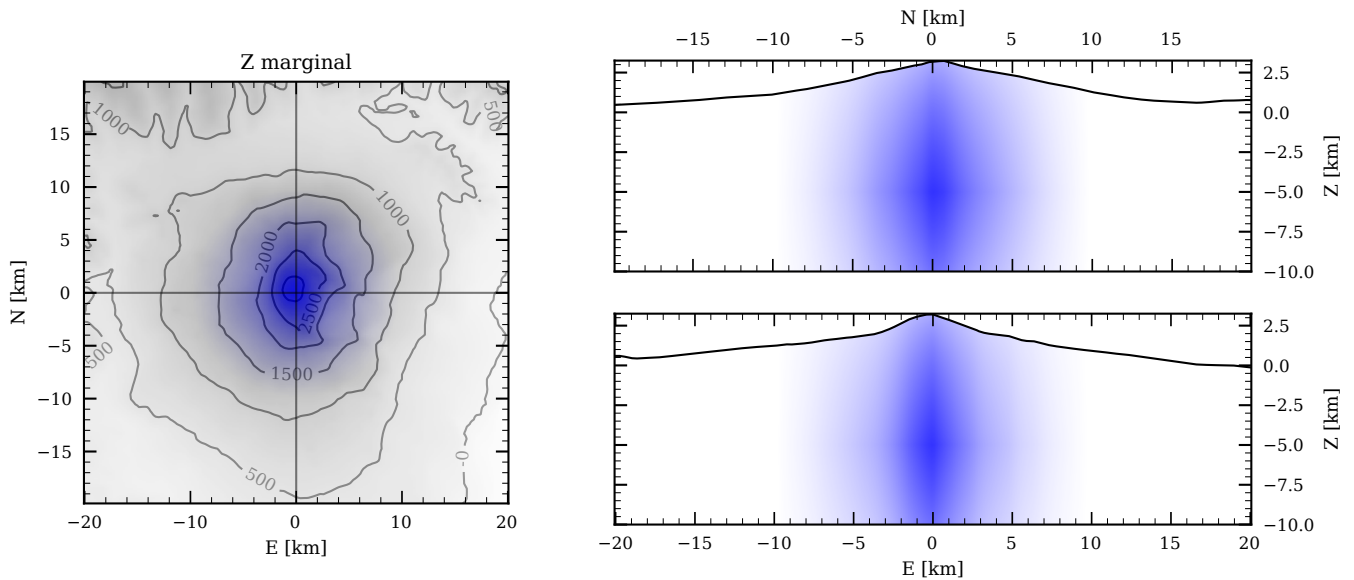
While the above data types capture the usual data used in volcano seismic monitoring, several important assumptions are made to make the design problem tractable, or because information necessary to model the problem more accurately is typically unavailable at the design stage. The most important assumptions are:

- Unless we have a subsurface velocity model available and use a considerably more expensive ray tracing method, we assume a homogeneous velocity model. While this is a severe assumption, it is often the only feasible way to simulate observed data in a volcano monitoring scenario. All of the experimental design methods can be extended to use a heterogeneous velocity model, but this requires a more complex forward function to be available.
- We assume a Gaussian data likelihood. This is a common assumption in Bayesian inference and design studies, but both the choice of variances, and deviations from a Gaussian distribution can impact the quality of results (Lomax et al., 2014). The experimental design methods could readily be extended to use a different likelihood, but this would require much scenario-specific information that is hard to obtain, and which is almost never employed in practice even after data have been collected. Optimised designs are less important in already well-monitored scenarios, where the likelihood might be able to be modelled more accurately since there is usually a diminishing return for a larger number of stations/arrays.
- We assume that the uncertainty in each datum is independent of the others. While this is also a common assumption in Bayesian inference, it can significantly impact the results, especially since different data types at similar locations depend in some way on the same ray paths. The experimental design methods could readily be extended to include correlations between the data types, but estimating the correlations *a priori* is usually difficult and, again, requires a lot of scenario-specific information. In practice, the impact of this assumption may be limited since experimental design algorithms typically spread out the locations of stations over the area of interest.

While these assumptions may seem restrictive, they are chosen to be practically implementable, and to deviate as little as possible from both the natural system and the assumptions typically made in Bayesian inference once data have been collected. In experimental design problems, we average over many samples of the prior distribution, which makes the impact of these assumptions less severe than in a typical inference problem.

### 3 Example Design Process

Finding an optimal experimental design depends on the specific scenario and the goals of the monitoring campaign. Instead of providing a one-size-fits-all solution, we provide a step-by-step guide to design an optimal monitoring network for any specific scenario. Nearly every step in the following example can be adjusted to specific needs. We provide a Jupyter notebook that guides practitioners through the process and serves as a template for each scenario. For more information see the data availability statement (section 7).



**Figure 3** The prior distribution of the source locations for Etna. Dark blue indicates a high likelihood of seismicity, while light blue indicates a low likelihood. The elevation of the volcano is shown in the background. The left figure shows the marginal of the prior distribution along the depth axis, while the right figures show slices of the prior distribution at the Easting and Northing of the synthetic true event.

### 3.0.1 Volcano Data

The Smithsonian Institution’s Global Volcanism Program (GVP) (Venzke, 2024) provides a comprehensive database of volcanic activity, including eruption histories, reports, and data. We use this to obtain basic information about any volcano by providing its name, such as its location, type, and eruption history. This information is used for several of the following steps. The volcano used in this example is Etna, a stratovolcano in Italy. By defining a bounding box around the volcano’s location, we use the OpenTopography API to obtain a digital elevation model (DEM) of the volcano. In this example, we use the SRTM15Plus (Tozer et al., 2019), which has a sufficient resolution for most volcano monitoring scenarios and unlike many other DEM’s it includes bathymetric data. It is also possible to use a custom DEM, but this is often unnecessary.

### 3.0.2 Prior Information

We can define a prior distribution for source locations based on the topographic information. We discretise the subsurface into a grid of cells, and a uniform prior distribution is assumed within each cell. For this example, we assume that the prior distribution is a Gaussian distribution with a standard deviation of 5 km in the Easting and Northing directions, and 8 km in the depth direction, centred the location of the volcano at a depth of 2 km. In addition, we assume that the prior distribution in each vertical column is proportional to the elevation at that point, which gives more weight to areas in which the volcano has a higher relative surface elevation. The result is, of course, a very simplified prior distribution, but it is easy to define, and since it is defined on a grid it is easy for any practitioner to adapt to a more complex distribution that takes a specific geo-

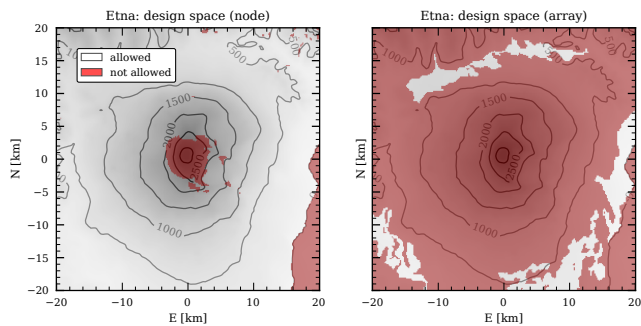
logical setting and other available information into account. Figure 3 shows the prior distribution used in this example.

### 3.0.3 Design Space

When optimising for the best design, we must consider numerous constraints present in each real-world scenario. In our code package, the design space (the set of all locations where a station or array can be placed) is defined on the grid of the (interpolated) DEM, where each cell is assigned a true or false value depending on whether or not a station can be placed there. In this example, we assume that nodal stations can only be placed in areas with an incline of less than 20 degrees, that array stations can be placed in regions with an incline of less than 3 degrees, that no station can be placed below sea level, that there is a safety margin of 3 km around the centre of the volcano if it erupted within the last ten years, and that arrays are placed in areas of at least 10 km<sup>2</sup> of flat (< 3 degrees) area. Figure 4 shows the design spaces for the nodal and array stations, which implement the above constraints.

Having the design space defined by general rules makes it easier to adapt to more complex scenarios with more constraints. Examples might include the use of satellite imagery to avoid recent lava flows, heavily wooded areas, urban areas, areas that are likely to experience high seismic noise, or any other area that is unsuitable for a seismic station. Using a genetic algorithm in the design optimisation process removes the need for constraints to have specific analytical properties, and so also makes it straightforward to include dynamic constraints such as the walking distance between stations, monetary constraints, or any other constraint that can be expressed as a function of the station locations.





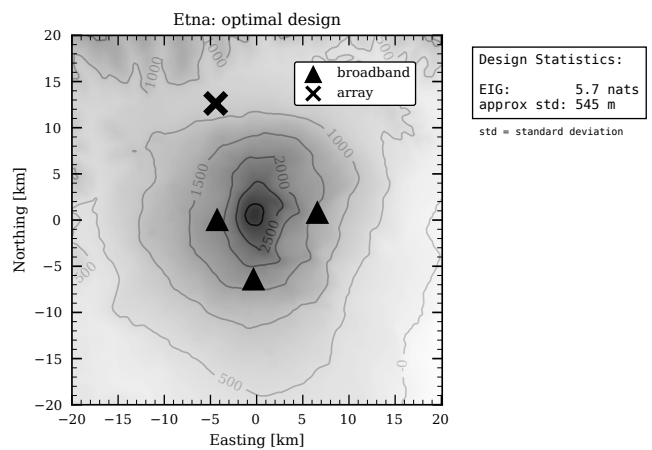
**Figure 4** The design space for the nodal and array stations for the example volcano Etna. Red indicates that a station can not be placed there. The elevation of the volcano is shown by the background grey scale and the grey contour lines.

### 3.0.4 Forward Function and Data Likelihood

To construct the forward function, we need to define several parameters that govern the propagation of seismic waves in the subsurface, with often limited information available. For travel-time data we set the P-wave velocity to 3.5km/s (Patane et al., 2002) and the corresponding uncertainties  $\sigma_t$  to 0.01s and  $\sigma_v$  to 0.1 (corresponds to a characteristic standard deviation of the velocity of 10% along the ray path). For amplitude measurements, the quality factor  $Q$  is set to 50, the frequency of interest  $f$  to 2.0Hz, and the S-wave velocity is set to  $v_s = v_p/\sqrt{3}$  (Patane et al., 2002; Morioka et al., 2017; Kumagai et al., 2009). A literature review, often including other, analogue volcanoes or geologies, is often necessary in order to estimate a suitable quality factor  $Q$  and uncertainty for amplitude data. The low computational cost of the  $D_N$  method allows possible values to be tested rapidly, and in our experience, many different combinations of values typically result in similar experimental designs. The standard deviation of the quality factor is set to 10 in this example, and this, together with the uncertainties of the S-wave travel-time and the distance between the source and receiver allows the uncertainty in amplitude to be calculated. The uncertainty in the azimuth is set to 6 degrees (Inza et al., 2011), and the incident angle is not used in this example. The back azimuth measurements require the least prior information. Within the code, the orientation of the array is corrected for the local topography.

### 3.0.5 Design Optimisation

To find the optimal design, we use a genetic algorithm (Holland, 1975; Sambridge and Gallagher, 1993; Gad, 2023) and the EIG to be optimised is calculated using the  $D_N$  method. In this example, we optimise a network of three nodal stations (travel-time and amplitude information) and one array (travel-time, amplitude, and back azimuth information). Using a population size of 64 designs, around 200 generations (iterations of the algorithm) are needed for the optimisation to converge to a solution, with only slight improvements occurring thereafter up to around 600 generations. Figure 5 shows the optimal design for this scenario.



**Figure 5** The optimal design for a network of three nodal stations and one array on Etna. The elevation of the volcano is shown in the background grey scale and the grey contour lines. For more information about the design statistics given in the box top-right, see section 3.0.6.

On an average laptop, results are available within a few minutes for this example, if the  $D_N$  method is used; the NMC method would take tens of minutes for the same number of samples, or several days if the number of samples is increased by an order of magnitude which would be necessary for an unbiased result. Note that the forward function used here, while not fully optimised, can still be calculated extremely rapidly. The difference between the true EIG and the estimates from either the  $D_N$  or NMC method (for the number of samples used) is mostly constant for the seismic source location example, resulting in similarly well-performing designs Strutz and Curtis (2023). Therefore, both methods can be used in the design optimisation process.

In some scenarios (e.g., prior distribution is far from Gaussian; a multimodal prior distribution; low noise levels), care must be taken when the  $D_N$  method is used so as not to violate the assumptions of a roughly Gaussian evidence too much. While this could be tested by plotting the evidence for a few designs, it is often easier in practice to use the NMC method and run another optimisation (with fewer generations) using the NMC method where the starting population is the optimal design found using the  $D_N$  method (this is implemented in the code package) to see whether the optimal design changes significantly. If it does, the  $D_N$  method may not be suitable for the scenario in question. Special care must be taken in the case of multimodal prior distributions (see Strutz and Curtis (2023) for an example of the  $D_N$  method failing in such a seismic source location scenario). Nevertheless, in general, the  $D_N$  method has proven to be robust in previous design optimisation studies (Strutz and Curtis, 2023; Coles et al., 2013; Bloem et al., 2020).

### 3.0.6 Design Analysis

After an optimal design has been computed it can be deployed. However, in many cases we would first like to analyse its performance in more detail, and possibly re-



fine it either semi-automatically or manually. To understand the expected uncertainties in the inversion results using the final design, we can calculate the posterior distribution for an example event. An example result is shown in Figure 6, where we observe that the (synthetic) true event lies well within the area of high posterior probability. By running several such tests, and by varying the design slightly, we can test how the uncertainty contours depend on the experimental design locations. As expected, the uncertainty in the vertical direction is substantially larger than in the horizontal directions since data is only collected on the surface at substantial offsets from the event. A receiver placed more directly above the event in question would conform more closely to the optimal design for a single event and allow more accurate estimates of the event depth (Bloem et al., 2020; Rabinowitz and Steinberg, 1990; Steinberg et al., 1995), but in this case those locations are precluded (Figure 4).

The main benefit of the EIG calculation is that it provides a measure of the information gain expected over all possible true source locations (according to the prior distribution) instead of just one or a small subset of events. While we could use the EIG directly to compare different designs, it is an unintuitive quantity which is hard to interpret. If we assume that the posterior distribution is a multivariate isotropic Gaussian distribution (same standard deviation  $\sigma_{\text{post}}$  for all parameters, and no inter-parameter correlations - see black circles overlying contours in Figure 6), we can use the expected posterior information  $\bar{I}_{\text{post}} = \mathbb{E}_{p(m)} [I_{\text{post}}] = \text{EIG} - I_{\text{prior}}$ , to estimate the expected standard deviation

$$\bar{\sigma}_{\text{post}}^2 = \exp \left\{ \frac{-\bar{I}_{\text{post}}}{3} - \frac{1}{2} (1 + \log(2\pi)) \right\} \quad (14)$$

by using the analytic expression for the information in a multivariate Gaussian in Equation 7. This expected standard deviation is an intuitive quantity that can be used to compare different designs and to analyse the expected uncertainty in the results. The effect of the isotropic assumption is that  $\bar{\sigma}_{\text{post}}$  will often overestimate horizontal uncertainties and underestimate vertical uncertainties.

We provide interactive plotting tools in the Jupyter notebook that accompanies this paper (see data availability statement in section 7), which allow the effects of different design locations on the posterior distribution and the expected standard deviation to be explored in real time. This makes it easy to refine the design and to include additional design constraints that are hard or impossible to include in the design space or the optimisation process.

Another way to analyse the optimal design process is to observe how the expected information gain and standard deviation evolve as a function of receiver count. Optimal networks of up to 10 receivers which record travel-time and amplitude data, given the same parameters as in the previous example, are shown in Figure 7. We can calculate EIG and  $\bar{\sigma}_{\text{post}}$  for each of these networks and plot them as a function of the number of receivers (Figure 8). This shows how many receivers are

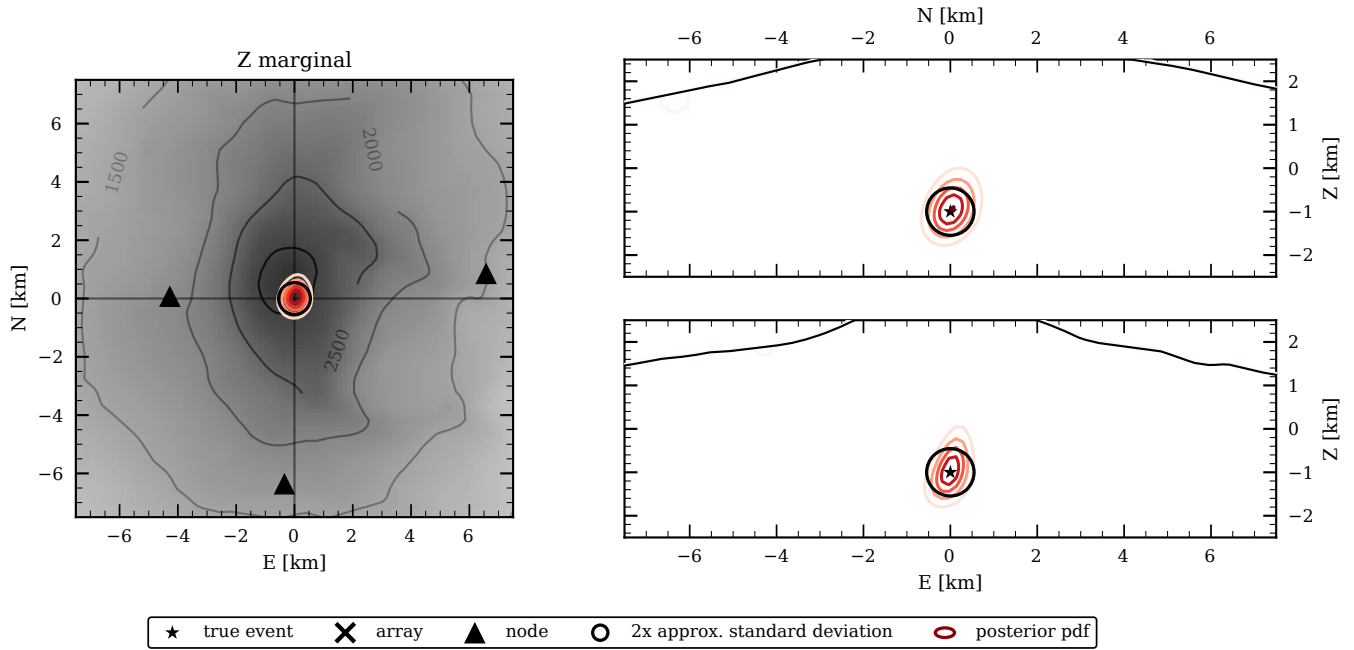
needed to reach any desired level of source location uncertainty.

The  $\bar{\sigma}_{\text{post}}$  results can also be used to compare the optimal designs to randomly placed, or approximately evenly spaced designs. To generate the random designs, we sample the design space according to a uniform distribution within the bounding box. The evenly spaced designs are generated by sampling a Sobol sequence - a quasi-random, space-filling sequence defined on  $(0, 1)^D$  (Sobol', 1967) which is then scaled by a factor sampled uniformly from 0 to 20 km. While the latter procedure is a complicated way to define reference space-filling designs, most of the resulting designs are closer to what we might consider to be a reasonable, uniformly distributed random design than are purely uniformly random designs, which is also reflected in the results. The expected standard deviation of 1000 of the random and Sobol designs are shown in Figure 8 as violin plots. It is clear that the optimal designs are substantially better than the random designs, and that the Sobol designs are better than uniformly random designs, and indeed the latter approach the performance of optimal designs. However, for more than two receivers, an optimal design allows one to achieve the same uncertainty level as the mean of the Sobol designs with one receiver fewer, which is clearly desirable.

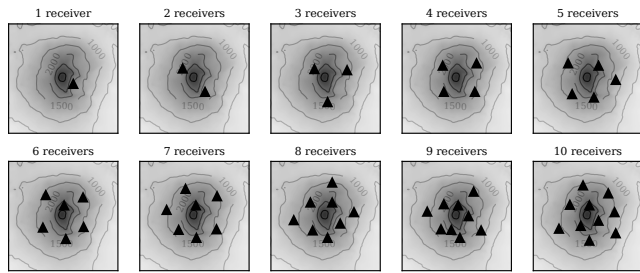
For more than around eight receivers, the reduction in uncertainty for each additional receiver becomes relatively small. This property of diminishing returns is a common feature of optimal design problems (Maurer and Boerner, 1998), but our code package allows this effect to be quantified rapidly. Even if sufficient receivers are used to reach the point of diminishing returns (we refer to such cases as large-N designs, where N is the number of receivers), it is valuable to know how many receivers are required at a minimum to reach consistent performance for any approximately evenly spaced design. While the optimal large-N design process does not add much information in such cases, the confirmation that N is indeed large enough is valuable information in itself.

## 4 Heterogenous Velocity Model

This section briefly explores how a heterogeneous seismic velocity model affects the formulation of the experimental design problem and the resulting optimal designs. The first immediate effect is that the forward function for all three data types is substantially more computationally expensive since an eikonal solver or seismic raytracer is needed to calculate travel-times, ray lengths and arrival angles. We use the open-source software package Pykonal (White et al., 2020) to calculate travel-time fields from which the other data types can be derived. To make the experimental design process more efficient, we precompute the data for every possible receiver/array location for a sufficient number of prior distribution samples such that results became stable under additional samples (typically around  $10^3 - 10^4$  samples) and store the results in a lookup table. We tested a layered model (Villaseñor et al., 1998) and a full 3D heterogeneous velocity model in which we added



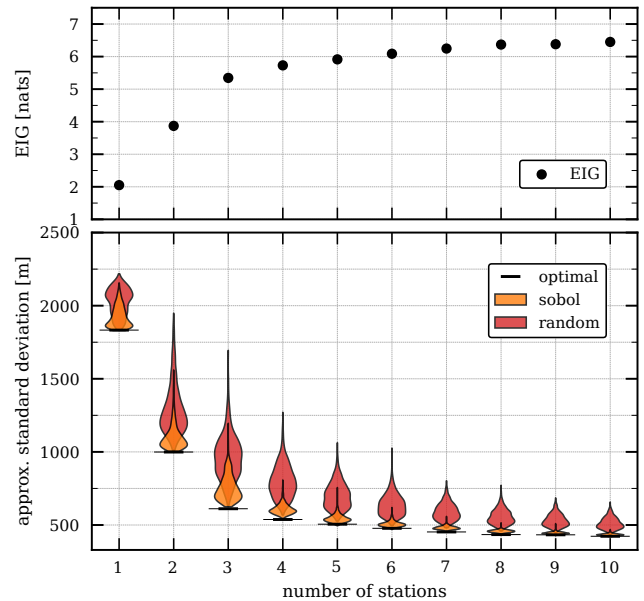
**Figure 6** The posterior distribution for a synthetic true event for Etna, recorded with the optimal design. The synthetic true event is shown as a black star, and the uncertainty contours of the Bayesian posterior distribution are shown in red (high) to orange (low probability density). The expected standard deviation of the posterior distribution derived from the EIG is shown as a black circle with radius  $\bar{\sigma}_{\text{post}}$ . The left figure shows the marginal of the posterior distribution integrated over the depth axis. The right figures show slices of the posterior distribution at the Easting and Northing of the synthetic true event.



**Figure 7** Example of optimal designs for Etna with up to 10 receivers. The coordinate axes span -15 to 15 km for both Easting and Northing.

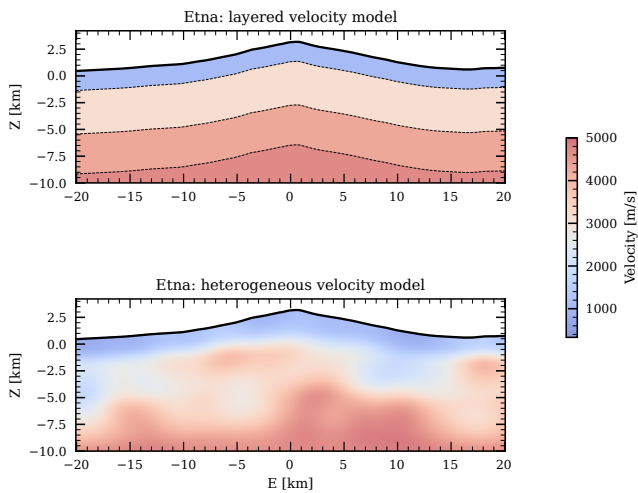
variations with a spatial wavelength of around 3 km and a relative amplitude of up to 50% to a smoothed version of the layered velocity model (none of the velocity models in the literature was openly available e.g., Villaseñor et al. (1998); Patane et al. (2002)). Both models are shown in Figure 9.

Since the computation time for the data table lookups is in the order of the computation time for the NMC method, in this case we use the more accurate NMC method for the optimal design process. For this experiment, we only consider travel-time and array-derived data since calculating the large number of ray paths necessary for the amplitude measurements would be computationally expensive, and previous tests showed that this data had a small effect on the optimal design process. Again, we use 500 generations in the genetic algorithm with a population size of 64, which still enables us to calculate the travel-time table and the optimal design within a few hours on a standard laptop.

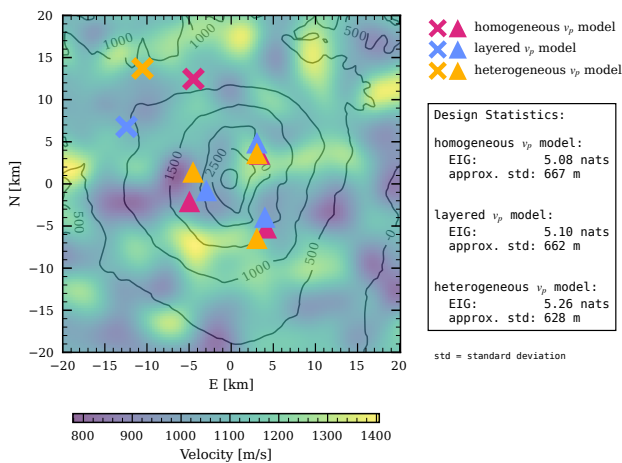


**Figure 8** The expected standard deviation and the EIG of the posterior distribution as a function of the number of receivers deployed in optimal configurations for the example volcano, Etna (Fig. 7). The violin plots for random and sobol designs show the density of the expected standard deviation for 1000 random and sobol designs. The wider the violin plot, the more designs have that expected standard deviation. Optimal designs provide the lowest expected standard deviation in all cases.

The results of the optimal design process for a homogeneous, layered and heterogeneous velocity model



**Figure 9** The layered and full 3D heterogeneous P-wave seismic velocity models used in this study. The layered model is the velocity model from Villaseñor et al. (1998), and the heterogeneous model is a 3D model in which noise with a spatial wavelength of around 3 km and a relative amplitude of up to 50% is added to a smoothed version of the layered model.



**Figure 10** The optimal design for a homogeneous, layered and heterogeneous velocity model for three receivers and a single array. The EIG and approximated expected standard deviation are calculated using the heterogeneous velocity model for all three optimal designs and shown in the box on the right. The seismic velocity model at the surface is shown in the background. Crosses denote arrays and triangles denote nodal stations.

for three receivers and a single array, derived with the same settings, are shown in Figure 10. EIG and approximated expected standard deviation are calculated with the heterogeneous velocity model for all three optimal designs. The performance of the homogeneous and layered models are very similar, but as expected the optimal design for the heterogeneous model performs better. While the relative placement of the nodes follows a similar pattern, the arrays are placed in different locations. Another observation not shown in the summary statistic is that the three designs perform very similarly for events at shallow or medium depth, and the increase in performance of the heterogeneous models is primarily due to improved performance for deep events.

The results show that the optimal design process is relatively robust towards the choice of velocity model, but a performance increase can be achieved by going beyond a 1D layered media. This also suggests that care must be taken for 3D velocity models with high uncertainty, since the optimal design process will be sensitive to the velocity model employed.

## 5 Discussion and Conclusion

We have presented a novel combination of experimental design methods for the optimal design of seismic monitoring networks for volcano monitoring. The methods are based on Bayesian experimental design to take the full uncertainty of the inversion process into account. We have shown that the expected standard deviation of the posterior distribution is a valuable quantity to evaluate the design quality, and present ways to interpret how much value an optimal design can add to post-survey inversion results.

In the pursuit of accessibility, several assumptions were made to allow for a straightforward use, even if limited information is available. Most assumptions can be relaxed if more information is available, as shown in the example of the heterogeneous velocity model. The methods are also flexible enough to include more complex constraints and data types, but since those are typically scenario-specific, we have not included them in this work. Extending the methods and code package to design networks that are optimised to constrain the full source solution including the moment tensor is a challenging problem that is not addressed in this work. The methods presented here can be extended for this purpose, but it is currently not clear how this can be done in such a computationally efficient way.

The main advance of this work is the use of both fast ( $D_N$ ) and accurate (NMC) methods to calculate the expected information gain, and setting up the codes with access to public databases such that they can be deployed rapidly for any volcano in the world. This approach allows practitioners to optimise designs within minutes and have interactive tools available to refine the results. We hope that this work will enable more researchers to use optimal design methods without needing to be experts in the field, and will allow both experts and non-experts alike to use optimal designs even when responding rapidly to evolving eruptive scenarios.



## 6 Acknowledgements

The implementations of the algorithms in this work would not have been possible without extensive use of open-source software. Not all of them have been included in the respective sections to ease readability. All the code was written in Python (Van Rossum and Drake, 2011), and the libraries NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), Matplotlib (Hunter, 2007), PyGAD (Gad, 2023), and xarray (Hoyer and Hamman, 2017) have been used extensively.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 955515—SPIN ITN ([www.spin-itn.eu](http://www.spin-itn.eu)).

This work is based on data services provided by the OpenTopography Facility with support from the National Science Foundation under NSF Award Numbers 1948997, 1948994 & 1948857.

## 7 Data Availability

All code necessary to reproduce the results is available at <https://github.com/dominik-strutz/WoWED-volcano>. The interactive Jupyter notebook that accompanies this paper and guides the user through the design process is available in the same repository. All other data is publicly available and referenced in the respective sections.

## References

Almendros, J. Performance of the radial semblance method for the location of very long period volcanic signals. *Bull. Seismol. Soc. Am.*, 93(5):1890–1903, Oct. 2003. doi: 10.1785/0120020143.

Battaglia, J. and Aki, K. Location of seismic events and eruptive fissures on the Piton de la Fournaise volcano using seismic amplitudes. *J. Geophys. Res.*, 108(B8), Aug. 2003. doi: 10.1029/2002jb002193.

Bloem, H., Curtis, A., and Maurer, H. Experimental design for fully nonlinear source location problems: which method should I choose? *Geophys. J. Int.*, 223(2):944–958, July 2020. doi: 10.1093/gji/ggaa358.

Cannavò, F., Camacho, A. G., González, P. J., Mattia, M., Puglisi, G., and Fernández, J. Real time tracking of magmatic intrusions by means of ground deformation modeling during volcanic crises. *Sci. Rep.*, 5(1):10970, June 2015. doi: 10.1038/srep10970.

Carbone, D., Zuccarello, L., and Saccorotti, G. Geophysical indications of magma uprising at Mt Etna during the December 2005 to January 2006 non-eruptive period. *Geophys. Res. Lett.*, 35(6), Mar. 2008. doi: 10.1029/2008gl033212.

Caudron, C., White, R. S., Green, R. G., Woods, J., Ágústsdóttir, T., Donaldson, C., Greenfield, T., Rivalta, E., and Brandsdóttir, B. Seismic Amplitude Ratio Analysis of the 2014–2015 bárabunga-Holuhraun dike propagation and eruption. *J. Geophys. Res. Solid Earth*, 123(1):264–276, Jan. 2018. doi: 10.1002/2017jb014660.

Chiodini, G., Paonita, A., Aiuppa, A., Costa, A., Caliro, S., De Martino, P., Acocella, V., and Vandemeulebrouck, J. Magmas near the critical degassing pressure drive volcanic unrest towards a critical state. *Nat. Commun.*, 7(1):13712, Dec. 2016. doi: 10.1038/ncomms13712.

Coles, D. and Curtis, A. Efficient nonlinear Bayesian survey design using DNOptimization. *Geophysics*, 76(2):Q1–Q8, Mar. 2011. doi: 10.1190/1.3552645.

Coles, D., Yang, Y., Djikpesse, H., Prange, M., and Osypov, K. Optimal nonlinear design of marine borehole seismic surveys. *Geophysics*, 78(3):WB17–WB29, May 2013. doi: 10.1190/geo2012-0265.1.

Cover, T. M. and Thomas, J. A. *Elements of Information Theory*. John Wiley & Sons, Nashville, TN, 2 edition, July 2006. doi: 10.1002/047174882x.

Dawson, P., Whilldin, D., and Chouet, B. Application of near real-time radial semblance to locate the shallow magmatic conduit at Kilauea Volcano, Hawaii. *Geophys. Res. Lett.*, 31(21), Nov. 2004. doi: 10.1029/2004gl021163.

Di Lieto, B., Saccorotti, G., Zuccarello, L., Rocca, M. L., and Scarpa, R. Continuous tracking of volcanic tremor at Mount Etna, Italy. *Geophys. J. Int.*, 169(2):699–705, May 2007. doi: 10.1111/j.1365-246x.2007.03316.x.

Falsaperla, S. and Neri, M. Seismic footprints of shallow dyke propagation at Etna, Italy. *Sci. Rep.*, 5(1):11908, July 2015. doi: 10.1038/srep11908.

Foster, A., Jankowiak, M., Bingham, E., Horsfall, P., Whye Teh, Y., Rainforth, T., and Goodman, N. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., Mar. 2019.

Gad, A. F. PyGAD: an intuitive genetic algorithm Python library. *Multimed. Tools Appl.*, 83(20):58029–58042, Dec. 2023. doi: 10.1007/s11042-023-17167-y.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2.

Holland, J. H. *Adaptation Natural Artificial Systems*. University Michigan Press, Ann Arbor, 1975.

Hoyer, S. and Hamman, J. xarray: N-D labeled Arrays and Datasets in Python. *J. Open Res. Softw.*, 5(1):10, Apr. 2017. doi: 10.5334/jors.148.

Huan, X. and Marzouk, Y. M. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput. Phys.*, 232(1):288–317, Jan. 2013. doi: 10.1016/j.jcp.2012.08.013.

Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, 9:90–95, May 2007. doi: 10.1109/MCSE.2007.55.

Inza, L. A., Mars, J. I., Métaixian, J. P., O'Brien, G. S., and Macedo, O. Seismo-volcano source localization with triaxial broadband seismic array: 3C seismic antenna on volcano. *Geophys. J. Int.*, 187(1):371–384, Oct. 2011. doi: 10.1111/j.1365-246x.2011.05148.x.

Jolly, A. D., Thompson, G., and Norton, G. E. Locating pyroclastic flows on Soufriere Hills Volcano, Montserrat, West Indies, using amplitude signals from high dynamic range instruments. *J. Volcanol. Geotherm. Res.*, 118(3-4):299–317, Nov. 2002. doi: 10.1016/s0377-0273(02)00299-8.

Kao, H. and Shan, S.-J. The Source-Scanning Algorithm: mapping the distribution of seismic sources in time and space. *Geophys. J. Int.*, 157(2):589–594, Apr. 2004. doi: 10.1111/j.1365-246x.2004.02276.x.

Kawakatsu, H., Kaneshima, S., Matsubayashi, H., Ohminato, T., Sudo, Y., Tsutsui, T., Uehira, K., Yamasato, H., Ito, H., and Legrand, D. Aso94: Aso seismic observation with broadband instru-

- ments. *J. Volcanol. Geotherm. Res.*, 101(1-2):129–154, Aug. 2000. doi: 10.1016/s0377-0273(00)00166-9.
- Ku, H. H. Notes on the use of propagation of error formulas. *J. Res. Natl. Bur. Stand.*, 70C(4):263, Oct. 1966. doi: 10.6028/jres.070c.025.
- Kullback, S. and Leibler, R. A. On Information and Sufficiency. *Ann. Math. Stat.*, 22(1):79–86, 1951.
- Kumagai, H., Palacios, P., Maeda, T., Castillo, D. B., and Nakano, M. Seismic tracking of lahars using tremor signals. *J. Volcanol. Geotherm. Res.*, 183(1-2):112–121, May 2009. doi: 10.1016/j.jvolgeores.2009.03.010.
- Kumagai, H., Placios, P., Ruiz, M., Yepes, H., and Kozono, T. Ascending seismic source during an explosive eruption at Tungurahua volcano, Ecuador. *Geophys. Res. Lett.*, 38(1), Jan. 2011. doi: 10.1029/2010gl045944.
- Kumagai, H., Lacson, Jr, R., Maeda, Y., Figueroa, II, M. S., Yamashina, T., Ruiz, M., Palacios, P., Ortiz, H., and Yepes, H. Source amplitudes of volcano-seismic signals determined by the amplitude source location method as a quantitative measure of event size. *J. Volcanol. Geotherm. Res.*, 257:57–71, May 2013. doi: 10.1016/j.jvolgeores.2013.03.002.
- Langet, N., Maggi, A., Michelini, A., and Brenguier, F. Continuous kurtosis-based migration for seismic event detection and location, with application to piton de la fournaise volcano, la reunion. *Bull. Seismol. Soc. Am.*, 104(1):229–246, Feb. 2014. doi: 10.1785/0120130107.
- Leva, C., Rumpker, G., and Wölbner, I. Multi-array analysis of volcano-seismic signals at Fogo and Brava, Cape Verde. *Solid Earth*, 13(8):1243–1258, Aug. 2022. doi: 10.5194/se-13-1243-2022.
- Lindley, D. V. On a Measure of the Information Provided by an Experiment, 1956. doi: 10.1214/aoms/1177728069.
- Lomax, A., Zollo, A., Capuano, P., and Virieux, J. Precise, absolute earthquake location under Somma-Vesuvius volcano using a new three-dimensional velocity model. *Geophys. J. Int.*, 146(2):313–331, Aug. 2001. doi: 10.1046/j.0956-540x.2001.01444.x.
- Lomax, A., Michelini, A., and Curtis, A. Earthquake location, direct, global-search methods. In *Encyclopedia of Complexity and Systems Science*, pages 1–33. Springer New York, New York, NY, 2014. doi: 10.1007/978-3-642-27737-5\_150-2.
- Marzocchi, W. and Bebbington, M. S. Probabilistic eruption forecasting at short and long time scales. *Bull. Volcanol.*, 74(8):1777–1805, Oct. 2012. doi: 10.1007/s00445-012-0633-x.
- Maurer, H. and Boerner, D. E. Optimized and robust experimental design: a non-linear application to EM sounding. *Geophys. J. Int.*, 132(2):458–468, Feb. 1998. doi: 10.1046/j.1365-246x.1998.00459.x.
- Maurer, H., Curtis, A., and Boerner, D. E. Recent advances in optimized geophysical survey design. *Geophysics*, 75(5):75A177–75A194, Sept. 2010. doi: 10.1190/1.3484194.
- McNutt, S. R. Volcanic seismology. *Annu. Rev. Earth Planet. Sci.*, 33(1):461–491, May 2005. doi: 10.1146/annurev.earth.33.092203.122459.
- Métaxian, J.-P., Lesage, P., and Valette, B. Locating sources of volcanic tremor and emergent events by seismic triangulation: Application to Arenal volcano, Costa Rica. *J. Geophys. Res.*, 107(B10):ECV 10–1–ECV 10–18, Oct. 2002. doi: 10.1029/2001jb000559.
- Morioka, H., Kumagai, H., and Maeda, T. Theoretical basis of the amplitude source location method for volcano-seismic signals. *J. Geophys. Res. [Solid Earth]*, 122(8):6538–6551, Aug. 2017. doi: 10.1002/2017jb013997.
- O’Brien, G. S., Lokmer, I., De Barros, L., Bean, C. J., Saccorotti, G., Metaxian, J.-P., and Patane, D. Time reverse location of seismic long-period events recorded on Mt Etna. *Geophys. J. Int.*, 184(1): 452–462, Jan. 2011. doi: 10.1111/j.1365-246x.2010.04851.x.
- Ogiso, M. and Yomogida, K. Migration of tremor locations before the 2008 eruption of Meakandake Volcano, Hokkaido, Japan. *J. Volcanol. Geotherm. Res.*, 217-218:8–20, Mar. 2012. doi: 10.1016/j.jvolgeores.2011.12.005.
- Ogiso, M., Matsubayashi, H., and Yamamoto, T. Descent of tremor source locations before the 2014 phreatic eruption of Ontake volcano, Japan. *Earth Planets Space*, 67(1), Dec. 2015. doi: 10.1186/s40623-015-0376-y.
- Ohminato, T., Chouet, B. A., Dawson, P., and Kedar, S. Waveform inversion of very long period impulsive signals associated with magmatic injection beneath Kilauea volcano, Hawaii. *J. Geophys. Res.*, 103(B10):23839–23862, Oct. 1998. doi: 10.1029/98jb01122.
- Pallister, J. and McNutt, S. R. Synthesis of volcano monitoring. In *The Encyclopedia of Volcanoes*, pages 1151–1171. Elsevier, 2015. doi: 10.1016/b978-0-12-385938-9.00066-3.
- Patane, D., Chiarabba, C., Cocina, O., De Gori, P., Moretti, M., and Boschi, E. Tomographic images and 3D earthquake locations of the seismic swarm preceding the 2001 Mt. Etna eruption: Evidence for a dyke intrusion. *Geophysical Research Letters*, 29(10): 135–1–135–4, May 2002. doi: 10.1029/2001GL014391.
- Rabinowitz, N. and Steinberg, D. M. Optimal configuration of a seismographic network: A statistical approach. *Bull. Seismol. Soc. Am.*, 80(1):187–196, Feb. 1990. doi: 10.1785/BSSA0800010187.
- Rost, S. and Thomas, C. Array seismology: Methods and applications. *Rev. Geophys.*, 2002. doi: 10.1029/2000RG000100.
- Ryan, K. J. Estimating Expected Information Gains for Experimental Designs With Application to the Random Fatigue-Limit Model. *J. Comput. Graph. Stat.*, 12(3):585–603, Sept. 2003. doi: 10.1198/1061860032012.
- Saccorotti, G. and Lokmer, I. Chapter 2 - A review of seismic methods for monitoring and understanding active volcanoes. In Papale, P., editor, *Forecasting and Planning for Volcanic Hazards, Risks, and Disasters*, volume 2, pages 25–73. Elsevier, Jan. 2021. doi: 10.1016/B978-0-12-818082-2.00002-0.
- Saccorotti, G., Lokmer, I., Bean, C. J., Di Grazia, G., and Patanè, D. Analysis of sustained long-period activity at Etna Volcano, Italy. *J. Volcanol. Geotherm. Res.*, 160(3-4):340–354, Feb. 2007. doi: 10.1016/j.jvolgeores.2006.10.008.
- Saltogian, V., Stiros, S. C., Newman, A. V., Flanagan, K., and Moschas, F. Time-space modeling of the dynamics of Santorini volcano (Greece) during the 2011-2012 unrest. *J. Geophys. Res. Solid Earth*, 119(11):8517–8537, Nov. 2014. doi: 10.1002/2014jb011409.
- Sambridge, M. and Gallagher, K. Earthquake hypocenter location using genetic algorithms. *Bull. Seismol. Soc. Am.*, 83(5):1467–1491, Oct. 1993. doi: 10.1785/bssa0830051467.
- Selva, J., Marzocchi, W., Sandri, L., and Costa, A. Operational short-term volcanic hazard analysis. In *Volcanic Hazards, Risks and Disasters*, pages 233–259. Elsevier, Jan. 2015. doi: 10.1016/b978-0-12-396453-3.00009-5.
- Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.
- Sigmundsson, F., Hreinsdóttir, S., Hooper, A., Arnadóttir, T., Pedersen, R., Roberts, M. J., Oskarsson, N., Auriac, A., Decriem, J., Einarsson, P., Geirsson, H., Hensch, M., Ofeigsson, B. G., Sturkell, E., Sveinbjörnsson, H., and Feigl, K. L. Intrusion triggering of the 2010 Eyjafjallajökull explosive eruption. *Nature*, 468(7322):426–430, Nov. 2010. doi: 10.1038/nature09558.

Sigmundsson, F., Hooper, A., Hreinsdóttir, S., Vogfjörð, K. S., Ófeigsson, B. G., Heimisson, E. R., Dumont, S., Parks, M., Spaans, K., Gudmundsson, G. B., Drouin, V., Árnadóttir, T., Jónsdóttir, K., Gudmundsson, M. T., Högnadóttir, T., Fridriksdóttir, H. M., Hensch, M., Einarsson, P., Magnússon, E., Samsonov, S., Brandsdóttir, B., White, R. S., Ágústsdóttir, T., Greenfield, T., Green, R. G., Hjartardóttir, Á. R., Pedersen, R., Bennett, R. A., Geirsson, H., La Femina, P. C., Björnsson, H., Pálsson, F., Sturkell, E., Bean, C. J., Möllhoff, M., Braiden, A. K., and Eibl, E. P. S. Segmented lateral dyke growth in a rifting event at Bárðarbunga volcanic system, Iceland. *Nature*, 517(7533):191–195, Jan. 2015. doi: 10.1038/nature14111.

Sobol', I. M. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802, 1967.

Steinberg, D. M., Rabinowitz, N., Shimshoni, Y., and Mizrachi, D. Configuring a seismographic network for optimal monitoring of fault lines and multiple sources. *Bull. Seismol. Soc. Am.*, 85(6): 1847–1857, Dec. 1995. doi: 10.1785/BSSA0850061847.

Strutz, D. and Curtis, A. Variational Bayesian experimental design for geophysical applications: seismic source location, amplitude versus offset inversion, and estimating CO2 saturations in a subsurface reservoir. *Geophys. J. Int.*, 236(3):1309–1331, Dec. 2023. doi: 10.1093/gji/ggad492.

Taisne, B., Brenguier, F., Shapiro, N. M., and Ferrazzini, V. Imaging the dynamics of magma propagation using radiated seismic intensity. *Geophys. Res. Lett.*, 38(4), Feb. 2011. doi: 10.1029/2010gl046068.

Tarantola, Valette, and Others. Inverse problems= quest for information. *J. Geophys.*, 50(1):159–170, 1982.

Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., and Wessel, P. Global bathymetry and topography at 15 arc sec: SRTM15+. *Earth Space Sci.*, 6(10):1847–1864, Oct. 2019. doi: 10.1029/2019ea000658.

Van Rossum, G. and Drake, F. *The python language reference manual*. Network Theory, Bristol, England, Mar. 2011.

Venzke, E. Global Volcanism Program [Database] Volcanoes of the World (v. 5.2.1; 3 Jul 2024) Distributed by Smithsonian Institution, compiled by Venzke, E. <https://doi.org/10.5479/si.GVP.VOTW5-2024.5.2>. Accessed: 2024-7-23.

Villaseñor, A., Benz, H. M., Filippi, L., De Luca, G., Scarpa, R., Patanè, G., and Vinciguerra, S. Three-dimensional P-wave velocity structure of Mt. Etna, Italy. *Geophysical Research Letters*, 25(11):1975–1978, June 1998. doi: 10.1029/98GL01240.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, 17(3):261–272, Mar. 2020. doi: 10.1038/s41592-019-0686-2.

White, M. C. A., Fang, H., Nakata, N., and Ben-Zion, Y. 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, July 2020. doi: 10.1785/0220190318.

Woods, J., Winder, T., White, R. S., and Brandsdóttir, B. Evolution of a lateral dike intrusion revealed by relatively-relocated dike-induced earthquakes: The 2014–15 Bárðarbunga–Holuhraun rifting event, Iceland. *Earth Planet. Sci. Lett.*, 506:53–63, Jan.

2019. doi: 10.1016/j.epsl.2018.10.032.

Yamasato, H. Quantitative analysis of pyroclastic flows using infrasonic and seismic data at unzen volcano, japan. *J. Phys. Earth*, 45(6):397–416, 1997. doi: 10.4294/jpe1952.45.397.

Zobin, V. M. Fundamentals of volcanic seismology. In *Introduction to Volcanic Seismology*, pages 35–60. Elsevier, 2017. doi: 10.1016/b978-0-444-63631-7.00003-0.

## A Shannon Information

Shannon information (Shannon, 1948) is an intuitive measure of information with several beneficial properties (e. g., linear additivity of information from independent sources). The Shannon information  $I[\cdot]$  of an arbitrary continuous probability density function  $p(x)$  is defined as

$$\begin{aligned} I[p(x)] &= \mathbb{E}_{p(x)} [\log_b(p(x))] \\ &= \int_{\mathcal{X}} p(x) \log_b(p(x)) dx \end{aligned} \quad (15)$$

where  $x \in \mathcal{X}$  is a random variable distributed according to  $p(x)$  and  $\mathbb{E}_{p(x)}$  is the expectation with respect to  $p(x)$ , which is defined by the right-most expression. Depending on the context, information is also often expressed through the entropy  $H[p(x)] = -I[p(x)]$ , where entropy  $H$  is defined to be the negative of either expression on the right of Equation (15). This absolute information measure can be extended to the relative information content of one pdf relative to another, also called the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951)

$$\text{KL}(P||Q) = \int_{\mathcal{X}} p(x) \log \left( \frac{p(x)}{q(x)} \right) dx \quad (16)$$

For further information on the properties of information, the reader is referred to Cover and Thomas (2006).

## B Rearranging the Expected Information Gain

$$\text{EIG}(\xi) = \mathbb{E}_{p(d|\xi)} [I[p(\mathbf{m} | \mathbf{d}, \xi)] - I[p(\mathbf{m})]] \quad (17)$$

$$= \mathbb{E}_{p(\mathbf{d}, \mathbf{m}|\xi)} \left[ \log \frac{p(\mathbf{m} | \mathbf{d}, \xi)}{p(\mathbf{m})} \right] \quad (18)$$

$$= \mathbb{E}_{p(\mathbf{d}, \mathbf{m}|\xi)} \left[ \log \frac{p(\mathbf{m}, \mathbf{d} | \xi)}{p(\mathbf{m})p(\mathbf{d} | \xi)} \right] \quad (19)$$

$$= \mathbb{E}_{p(\mathbf{d}, \mathbf{m}|\xi)} \left[ \log \frac{p(\mathbf{d} | \mathbf{m}, \xi)}{p(\mathbf{d} | \xi)} \right] \quad (20)$$

$$\text{EIG}(\xi) = \mathbb{E}_{p(\mathbf{m})} [I[p(\mathbf{d} | \mathbf{m}, \xi)] - I[p(\mathbf{d} | \xi)]] \quad (21)$$