Geophysical Journal International

Geophys. J. Int. (2022) **229,** 750–757
Advance Access publication 2021 December 10
GJI Seismology

https://doi.org/10.1093/gji/ggab496

Interrogating probabilistic inversion results for subsurface structural information

Xin Zhang and Andrew Curtis

School of GeoSciences, University of Edinburgh, Edinburgh EH93FE, UK. E-mail: X.Zhang2@ed.ac.uk

Accepted 2021 December 8. Received 2021 November 30; in original form 2021 October 26

SUMMARY

The goal of a scientific investigation is to find answers to specific questions. In geosciences, this is typically achieved by solving an inference or inverse problem and interpreting the solution. However, the answer obtained is often biased because the solution to an inverse problem is non-unique and human interpretation is a biased process. Interrogation theory provides a systematic way to find optimal answers by considering their full uncertainty estimates and by designing an objective function that defines desirable qualities in the answer. In this study, we demonstrate interrogation theory by quantifying the size of a particular subsurface structure. The results show that interrogation theory provides an accurate estimate of the true answer, which cannot be obtained by direct, subjective interpretation of the solution mean and standard deviation. This demonstrates the value of interrogation theory. It also shows that fully nonlinear uncertainty assessments may be critical in order to address real-world scientific problems; this goes some way towards justifying their computational expense.

Key words: Inverse theory; Probability distributions; Statistical methods; Waveform inversion.

1 INTRODUCTION

Geoscientists often wish to find answers to specific scientific questions: How large is a subsurface body? How deeply does lithosphere subduct? How likely is this volcano to erupt? What method provides the most accurate results? To answer such questions, background research is conducted to reveal existing information, and experiments are designed and performed to acquire new data. The answer to the question is then estimated by interpreting the sum of this information.

Answering a question therefore requires that we obtain useful information relevant to the answer from both existing information (often called the *prior* information) and new data. The process of obtaining useful information usually involves solving an inverse or inference problem (Tarantola 2005). In inversion, one defines a simplified model to represent the physical system, and a relationship that simulates data from the model, called the forward function. The inversion then involves finding a model or a family of models that fit the data to within their uncertainties. For example, to answer questions about the Earth's interior scientists often build subsurface tomographic models from data observed at the surface: seismic velocity structures are obtained from seismic data, or resistivity structures may be constructed from electromagnetic data. This involves solving an inverse problem to estimate a subsurface model or a family of models that are consistent with the data, and an answer to the question may be interpreted from the solution. We address such a case herein.

Due to the nonlinear physical relationship between model parameters and data, insufficient data coverage and noise in the data, the inverse problem almost always has non-unique solutions as many sets of model parameter values fit the data to within their measurement uncertainties. It is therefore important to characterize the uncertainty of a solution such that the final answer can take into account the range of possible models.

The tomographic inverse problem is often solved using either standard optimization or Bayesian inference. In optimization, one seeks a solution that minimizes a misfit function between the observed data and the data predicted from the parameter values of a model (Tarantola 2005; Aster *et al.* 2018). However, since the method only finds one single set of parameter values, it is difficult to characterize the uncertainty and hence the information value of the solution. As a result, the answer inferred from that solution can be biased. Bayesian inference provides a different way to solve the inverse problem. In Bayesian inference, one constructs a probability density function (pdf) that describes the uncertainty of solutions, called the *posterior* pdf, by combining the prior information and the information contained in the data. Statistics or samples of that distribution are estimated in order to characterize the solution. Methods used for Bayesian inference include Monte Carlo sampling (Tarantola & Valette 1982; Mosegaard & Tarantola 1995; Sambridge 1999; Sambridge & Mosegaard 2002; Bodin & Sambridge 2009; Stuart 2010; Brooks *et al.* 2011) and variational inference (Blei *et al.* 2017; Nawaz & Curtis 2018; Zhang & Curtis 2020a, b; Zhang *et al.* 2021).

While many studies have been conducted to solve inverse problems, few have found answers to specific scientific questions based on the solutions. In practice, those questions are typically answered by subjectively interpreting the solution of inverse problems using either the optimal solution or the mean of the posterior pdf together with the standard deviation structure. Since these statistics do not represent the full uncertainty in the solutions, and since human interpretation is a biased process (Polson & Curtis 2010; Bond *et al.* 2012), the answer obtained in such a way is likely to be biased and does not take account of the full uncertainty in the Bayesian solution.

To resolve this issue, Arnold & Curtis (2018) introduced *interrogation theory* that explicitly brings questions and their space of possible answers into a single framework, and combines inversion, inference, design and decision theory to create a systematic way to find optimal answers. In interrogation theory, data are acquired in the way that best answers the question, multiple models are designed and discriminated, parameters of each model can be constrained by multiple algorithms and optimal answers are found using decision theory. One might regard interrogation theory as part of a general inference theory that infers useful information from data. However, inference theory does not usually consider questions and the space of possible answers systematically, whereas interrogation theory clearly states the problem as finding answers to specific questions: it incorporates decision theoretical concepts to define optimal answers within a class, and design theory to optimize data sets that discriminate between answers (Arnold & Curtis 2018).

Arnold & Curtis (2018) is a complex paper that introduces interrogation theory in a general way, but does not show how the theory can be applied in practice. So in this study we demonstrate how to apply interrogation theory in an example that simulates a practical scenario in which one wishes to estimate the size of a subsurface velocity structure observed using seismic full-waveform inversion (FWI). The aim of this study is (i) to translate the previous paper (Arnold & Curtis 2018) into easier language for less mathematical or statistical readers; (ii) to demonstrate a worked example on a realistic question about a high-dimensional parameter space with one of the most strongly nonlinear forward functions in geophysics, to show that the methodology can be used in non-trivial cases; and (iii) to illustrate that while the overall target question of a scientific study may sound simple, commonly it breaks down into less easily answerable subquestions.

In this case, the subquestion concerns how to define, and to discriminate, the boundaries of a low-velocity region with minimal bias. This is itself an important question that has to our knowledge seldom been addressed in a Bayesian manner: tomographers often interpret 'velocity anomalies' in an ad hoc way—without ever actually defining their boundaries. We show that it is necessary to quantify uncertainty in the seismic velocity structure in order to construct an unbiased estimate of an anomaly's boundary. In other words, without using Bayesian, nonlinear methods, this subquestion cannot currently be answered.

In the following section, we first review concepts of interrogation theory. In Section 3, we use the theory to estimate the size of a subsurface velocity structure obtained using FWI. We solve the inverse problem using Stein variational gradient descent (SVGD) to produce the full Bayesian posterior pdf. The size of the velocity structure is then estimated by interrogating the obtained posterior pdf. The results show that the estimated size is very close to the true value, a result that cannot be obtained using only a single best-fitting velocity structure estimate. This demonstrates that the fully nonlinear estimates of uncertainty are critical for decision-making in real-world problems, thus in part justifying their computational expense.

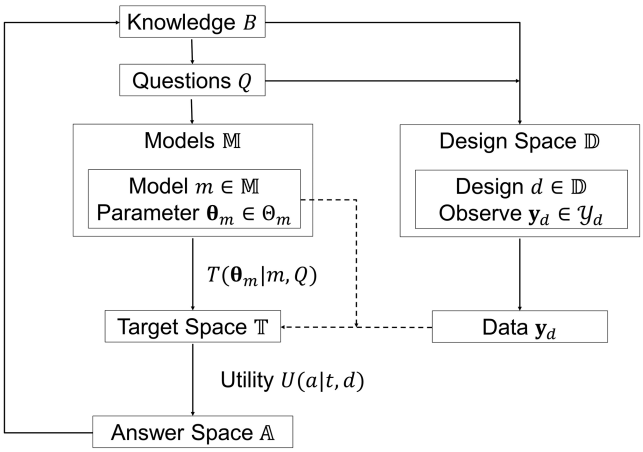

Figure 1. Algorithmic schema of interrogation theory.

2 OVERVIEW OF INTERROGATION THEORY

2.1 Interrogation theory

Interrogation theory provides a systematic way to find optimal answers to specific questions. Fig. 1 shows an algorithmic scheme for the theory. An investigator has prior knowledge B and wishes to answer a question Q. There is a set of answers $\mathbb{A}$ among which a choice needs to be made. For example, in geophysics scientists may ask questions about the depth of the Moho or the size of a reservoir, for which the answer space $\mathbb{A}$ will contain a set of real-numbered values. In other cases, we may want to know whether a specific geological structure exists, and the answer space will contain only two values: yes or no.

In order to answer the question, the investigator defines a set of models $\mathbb{M}$ that are relevant to the investigation. Similarly to Arnold & Curtis (2018), from here on we use the term 'model' in a mathematical sense to mean a relationship between the observed data and the parameters of the model. For example, in FWI we use a model $m \in \mathbb{M}$ consisting of parameters $\boldsymbol{\theta}_m$ in parameter space Θ_m which represent a 3-D seismic velocity structure of the Earth's interior, together with the relationship between this velocity structure and seismic waveforms. The model space $\mathbb{M}$ is therefore related to our prior knowledge B and the question Q. In general, an element or a set of elements in this space are assumed to provide a sufficiently accurate description of the state of nature relevant to answering Q.

To answer Q, an investigator needs to collect new information, which involves designing experiments and collecting data. Here, we use $\mathbb{D}$ to denote the design space that contains all possible experimental designs. For each design $d \in \mathbb{D}$, there is a data space $\mathcal{Y}_d$ that contains all possible observations $\mathbf{y}_d \in \mathcal{Y}_d$. After an experiment a single data set $\mathbf{y}_d^{obs}$ will have been collected.

For some questions Q, there may be many different relevant models in $\mathbb{M}$. For example, to study subsurface Earth structure one can use seismic data to infer the seismic velocity structure, or use resistivity data to infer the resistivity structure. There may also be a variety of rock physics models that relate seismic velocities or subsurface resistivities to parameters of interest such as porosity or rock type. However, no matter which models are used, they must

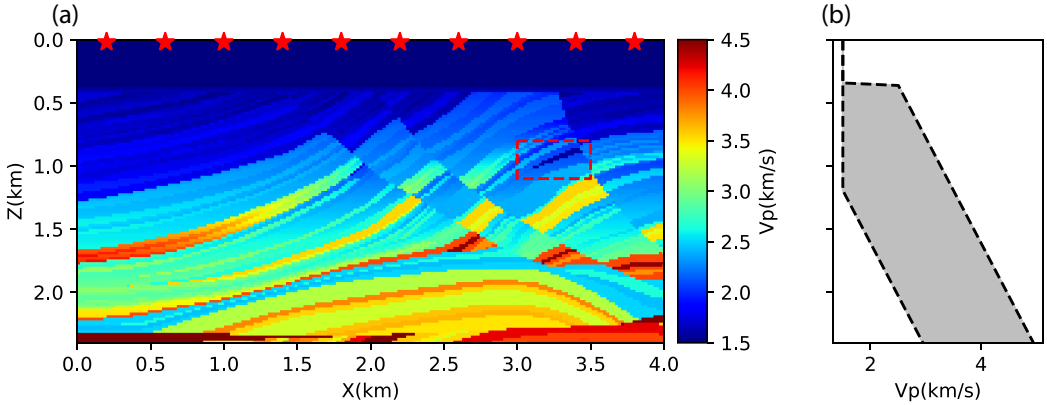

Figure 2. (a) The true velocity model. Red stars denote locations of 10 sources. The 200 receivers are equally spaced along the seabed at 0.36 km depth. (b) The prior distribution of seismic velocity, which is chosen to be a Uniform distribution over an interval of up to 2 km s^{-1} at each depth. A lower velocity bound of 1.5 km s^{-1} is imposed to ensure the velocity is higher than the acoustic velocity in water.

allow us to answer the question Q given the model parameters $\boldsymbol{\theta}_m$. Thus, for each model m and the question Q, there exists a target function $T(\boldsymbol{\theta}_m|m, Q)$ that maps the model parameters $\boldsymbol{\theta}_m$ to a target space $\mathbb{T}$ that is common for all possible models $\mathbb{M}$, and where the function T summarizes information needed to answer question Q.

To find the optimal answer in the answer space $\mathbb{A}$, we define a utility function $U(a|t, d)$ that defines the benefit associated with accepting an answer a given the summarized state of information $t = T(\boldsymbol{\theta}_m|m, Q)$ and the design d. The utility function is conditioned on d so that the benefits can account for the cost of conducting the experiment with design d. An optimal answer is found by maximizing this utility function in the answer space $\mathbb{A}$, and several analytic results that aid this calculation are given by Arnold & Curtis (2018).

2.2 Optimal answers and designs

Given a specific experimental design d and the observed data $\mathbf{y}_d$, the optimal answer to a question Q is found by maximizing the investigator's utility function U. In principle, this problem may be solved by maximizing a utility function $U(a|\boldsymbol{\theta}_m, m, \mathbf{y}_d, d)$ for answer $a \in \mathbb{A}$, parameters $\boldsymbol{\theta}_m \in \Theta_m$ embedded in model m and data $\mathbf{y}_d$ collected under a design d. However, such a utility function $U(a|\boldsymbol{\theta}_m, m, \mathbf{y}_d, d)$ is difficult to specify when considering the spectrum of different parameters $\boldsymbol{\theta}_m$ of different models m, and different data $\mathbf{y}_d$ under different designs d. To resolve this issue, Arnold & Curtis (2018) introduced the above target space $\mathbb{T}$ which is a common space for all possible models. The utility function can then be defined on the level of this common space, that is $U(a|t, d)$, which avoids the need to specify a utility function for every parameter value of every model in model space.

Define $p(\boldsymbol{\theta}_m, m|\mathbf{y}_d, d)$ as the Bayesian posterior probability density function (pdf) for a model m and its parameters $\boldsymbol{\theta}_m$ given the observed data $\mathbf{y}_d$ under the design d. According to Bayes theorem,

$$p(\boldsymbol{\theta}_m, m|\mathbf{y}_d, d) = \frac{p(\mathbf{y}_d|\boldsymbol{\theta}_m, m, d)p(\boldsymbol{\theta}_m|m)p(m)}{p(\mathbf{y}_d|d)}, \quad (1)$$

where $p(\mathbf{y}_d|\boldsymbol{\theta}_m, m, d)$ is the likelihood function of observing data $\mathbf{y}_d$ given parameters $\boldsymbol{\theta}_m$, embodied in a model m and under the design d. $p(\boldsymbol{\theta}_m|m)$ is the prior pdf of $\boldsymbol{\theta}_m$ associated with a model m and $p(m)$ is the prior pdf of model m. $p(\mathbf{y}_d|d)$ is a normalization factor called the 'evidence'. The posterior pdf on the left of eq. (1) can be obtained by Bayesian inference, for example by using Monte Carlo sampling or variational inference methods. The expected posterior utility of answers can then be constructed by integrating or summing over the set of models and over the parameter space of each model:

$$U_p(a|\mathbf{y}_d, d) = \sum_{m\in\mathbb{M}} \int_{\Theta_m} U(a|T(\boldsymbol{\theta}_m|m), d)p(\boldsymbol{\theta}_m, m|\mathbf{y}_d, d)d\boldsymbol{\theta}_m. \quad (2)$$

The optimal answer $a^*(\mathbf{y}_d, d)$ is obtained by maximizing the expected utility function:

$$a^*(\mathbf{y}_d, d) = arg\,max_{a\in\mathbb{A}} U_p(a|\mathbf{y}_d, d). \quad (3)$$

The corresponding maximized utility function is $U^*(\mathbf{y}_d, d) = U_p(a^*|\mathbf{y}_d, d)$.

If the investigator wishes to find the best experimental design, the solution can be obtained similarly by maximizing the utility function in the design space. First, the expected utility for each design d can be obtained by integrating over the data space:

$$U^*(d) = \int_{\mathcal{Y}_d} U^*(\mathbf{y}_d, d)p(\mathbf{y}_d|d)d\mathbf{y}_d, \quad (4)$$

where $p(\mathbf{y}_d|d)$ represents the probability of observing data $\mathbf{y}_d$ under the design d. This distribution corresponds to the evidence in Bayes theorem and is obtained by integrating over the model space and over the parameter space of each model:

$$p(\mathbf{y}_d|d) = \sum_{m\in\mathbb{M}} \int_{\Theta_m} p(\mathbf{y}_d|\boldsymbol{\theta}_m, m, d)p(\boldsymbol{\theta}_m|m)p(m)d\boldsymbol{\theta}_m. \quad (5)$$

The best experimental design is then obtained by maximizing the utility in eq. (4):

$$d^* = arg\,max_{d\in\mathbb{D}} U^*(d). \quad (6)$$

We now define the utility function $U(a|t, d)$. The utility function should be defined based on our knowledge of the question, and to represent desirable qualities in the answer (Arnold & Curtis 2018). In the simplest case, we can assume that the target space and the answer space are identical, that is our question is to estimate the summarized state $t = T(\boldsymbol{\theta}_m|m)$. In this case, we might define the utility function as

$$U(a|t, d) = U(a|t) = -(t - a)^2, \quad (7)$$

where in the first equality we have neglected the cost of conducting the experiment with design d. This utility is maximized when the answer a equals to the true state t. Arnold & Curtis (2018) showed

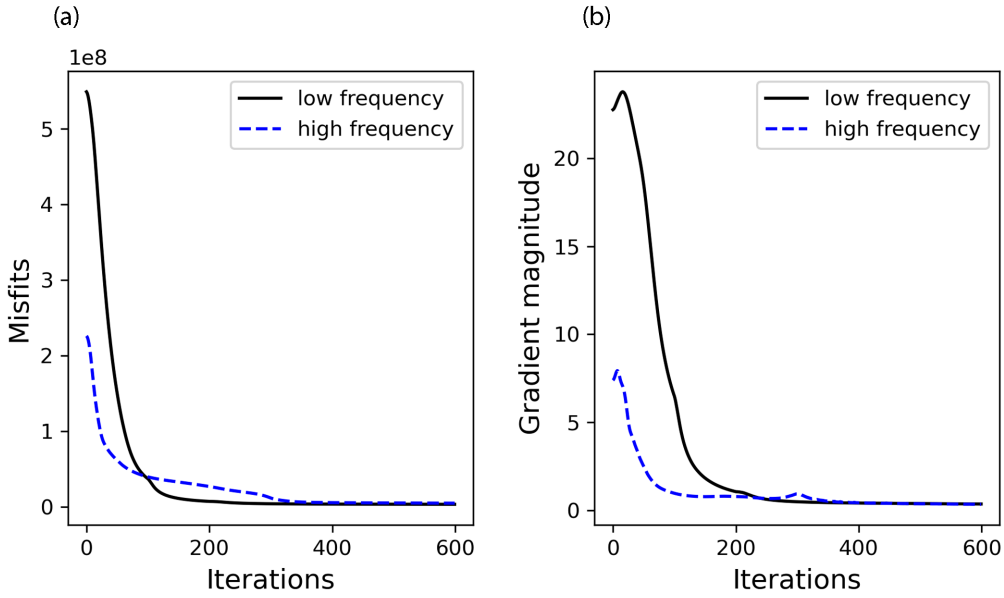

Figure 3. The (a) average misfit and (b) average magnitude of gradients of data with respect to parameters. Both (a) and (b) are averaged across all particles and are plotted as a function of iterations.

that in this case the optimal answer is

$$a^*(\mathbf{y}_d, d) = E[T|\mathbf{y}_d, d] = \sum_{m\in\mathbb{M}} \int_{\Theta_m} T(\boldsymbol{\theta}_m|m)p(\boldsymbol{\theta}_m, m|\mathbf{y}_d, d)d\boldsymbol{\theta}_m, \tag{8}$$

which states that the optimal answer is the posterior mean of the summarized state T. Using different utility functions in eq. (7), or answering different types of question (e.g., categorical questions), results in different forms for the optimal answer in eq. (8).

3 RESULTS

We demonstrate the method by interrogating the size of a subsurface structure. Such questions appear frequently, such as where we wish to estimate the size of a subsurface ore body, or of a reservoir for carbon capture and storage, or to estimate the size of a volcanic magma chamber. In order to answer this question we choose to use seismic FWI to estimate the subsurface seismic velocity structure, and we infer the size of a subsurface structure as defined by velocity anomalies. We use a part of the Marmousi model (Martin *et al.* 2006) as the true velocity structure to demonstrate the method, and simulate 10 sources at 20 m depth in the water layer with 200 equally spaced receivers at a depth of 360 m across the horizontal extent of the seabed (Fig. 2a). This acquisition geometry may not be the optimal design, but it represents geometries that occur in reality where we have fixed seismometers and wish to answer specific scientific questions about the Earth's interior.

The model is discretized in space using a regular 200 × 120 grid of constant-velocity cells; the set of cell velocities constitutes parameters $\boldsymbol{\theta}$. The model's relationship to data is a full-waveform simulation from each source to all receivers through velocity structure $\boldsymbol{\theta}$. The prior pdf of the velocity is assumed to be a uniform distribution at each depth with an interval up to 2 km s^{-1} wide, centred on a 1-D velocity model. To ensure that the rock velocity is higher than the acoustic velocity in water, we also imposed an additional lower bound of 1.5 km s^{-1} (Fig. 2b). We generate two waveform data sets using Ricker wavelets with dominant frequencies of 4 and 10 Hz, respectively, so that the information gained from low-frequency data and high-frequency data can be compared. To obtain the posterior distribution of the velocity, we use SVGD to solve the inverse problem. SVGD is a variational inference method that optimizes a set of samples of parameter space (called particles) to minimize the difference between the pdf represented by their distribution and the true posterior pdf (Liu & Wang 2016; Zhang & Curtis 2020b). For each data set inversion, we used 600 particles that are initially generated from the prior pdf and are then updated for 600 iterations until both the average misfit and the magnitude of gradients of data with respect to parameters averaged across all particles cease to decrease (Fig. 3). More details of the inversion procedure can be found in Zhang & Curtis (2021a). Here, we focus on the dark-blue triangular structure around $X = 3.2$ km and $Z = 1.0$ km (see red box in Fig. 2a).

Figs 4(a) and (c) show the posterior mean obtained using the low-frequency data and high-frequency data, respectively. In both results there is a low-velocity anomaly of which we wish to estimate the size. In order to do this, we first define a 'low velocity' by using a threshold: any cell whose velocity is smaller than the threshold is defined as a low velocity. Although the threshold may be chosen based only on the mean model, this procedure does not account for uncertainty in the velocity structure and hence is likely to introduce bias to the estimate of anomaly size unless properly chosen. For example, Fig. 4 (e) shows the marginal distributions of velocity in cells at $X = 3.29$ km across the vertical extent of the velocity anomaly (pink line in Fig. 4a). While the pixels at depth of 0.90 and 0.92 km are highly likely to be within the low-velocity anomaly, the pixels at depth of 0.86 and 0.98 km clearly do not belong to this anomaly. In comparison, it is difficult to discriminate whether or not the pixels at depths of 0.88, 0.94 and 0.96 km belong to the anomaly. This makes it difficult to choose an appropriate threshold with which the velocity anomaly can be defined. To resolve this issue, we select a set of points that are extremely likely to belong to the anomaly since their velocities are low in almost all particles (red stars in Figs 4a and c) and another set of points that are highly likely to be outside of the anomaly (black crosses in Figs 4a and c), each

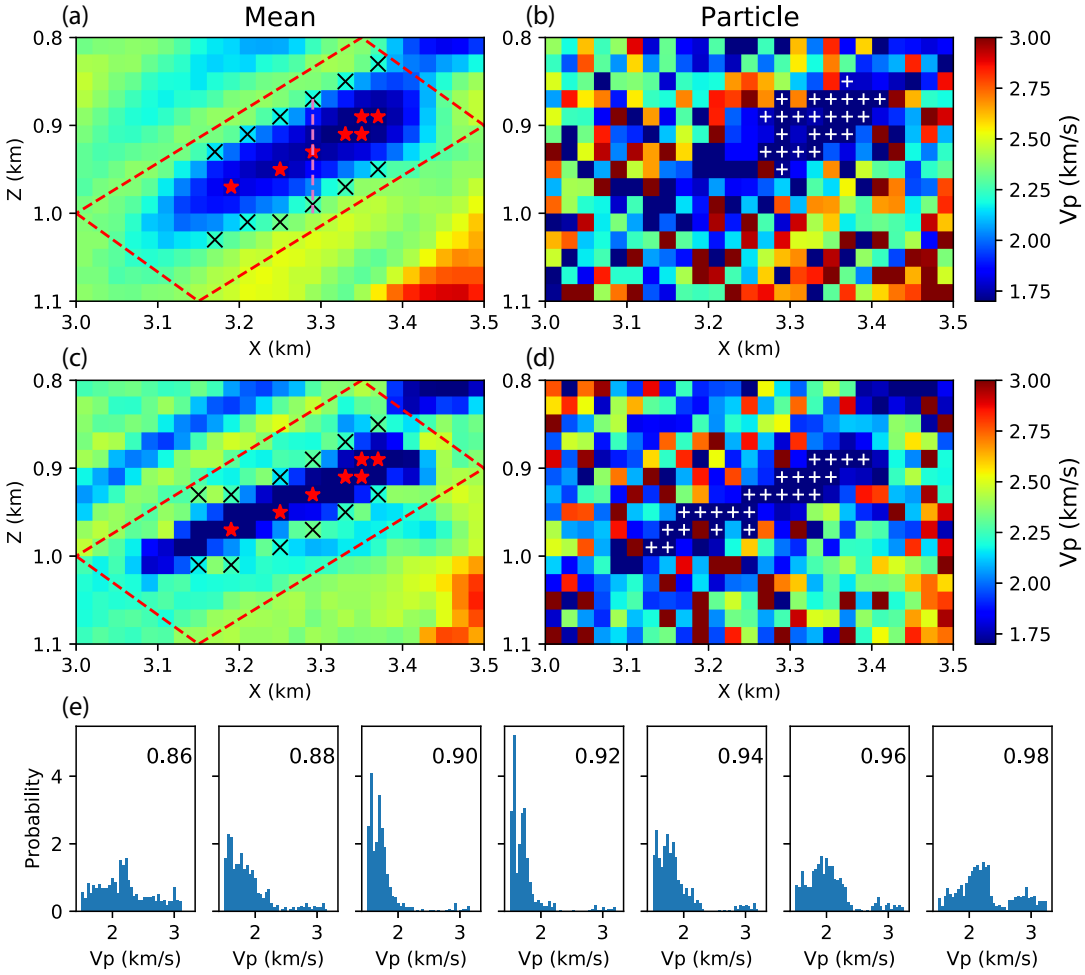

Figure 4. (a) Mean and (b) a random posterior particle (sample) obtained using the low-frequency data. (c) Mean and (d) a random posterior particle obtained using the high-frequency data. Red stars and black crosses denote locations that most likely have low velocity and high velocity, respectively. The red dashed box shows the region where interrogation is performed. White pluses in (b) and (d) show continuous low-velocity anomalies found for each particle. (e) Marginal distributions at $X = 3.29$ km and at regular intervals across the depth range from 0.86 to 0.98 km (pink dashed line in panel a) obtained using the low-frequency data.

chosen using only the clearest results from single-cell marginal distributions. We then calculate the posterior cumulative density function (CDF) of the two sets of points, one accumulated in the positive and the other in the negative Vp direction, and plot them against each other. Figs 5(a) and (b) show the CDF plots obtained using the low-frequency data and high-frequency data, respectively. We define the velocity value where the two CDFs have the same probability as the threshold, because by definition of the CDF the probability that the velocity of those points within the anomaly are lower than this value equals the probability that the velocity of those points outside the anomaly are higher than this value. This velocity threshold therefore discriminates low from high velocities with minimal bias. For the low- and high-frequency examples in this study, the above procedure results in the thresholds 1.778 and 1.745 km s^{-1}, respectively.

We now define a low-velocity anomaly as a continuous area whose velocity is smaller than the above threshold. Because of uncertainty in the posterior velocities (e.g. see Figs 4b and d), there can be many such low-velocity anomalies in a posterior velocity structure. To restrict ourselves to the main low-velocity anomalies observed in the mean velocity structure, for each posterior particle we focus only on the largest continuous low-velocity anomaly within the area of interest (red box in Figs 4a and c). For example, Figs 4(b) and (d) show examples of such low-velocity anomalies that are denoted by white pluses. The size of each anomaly can then be computed, and in this study we simply use the number of interior pixels as the anomaly size. The above procedure of thresholding and counting pixels constitutes our target function $T(\boldsymbol{\theta}_m|m)$.

Figs 5(c) and (d) show the distributions of the target function $T(\boldsymbol{\theta}_m|m)$ (the anomaly size) obtained using the low-frequency data and high-frequency data, respectively. According to eq. (8), the optimal answers are the mean of these distributions, which are denoted by dashed red lines in Figs 5(c) and (d). For comparison, the true size is denoted with black lines. The distribution of the anomaly size obtained using low-frequency data is wider than that obtained using high-frequency data, and the optimal answer obtained using only low-frequency data also has a larger error. This demonstrates quantitatively and probabilistically that by using high-frequency data we can obtain more accurate answers to specific scientific questions. In Figs 5(c) and (d), we also show the answers obtained by following the usual procedure of interpreting only the mean (green dotted line) and median (pink dashed–dotted line) velocity

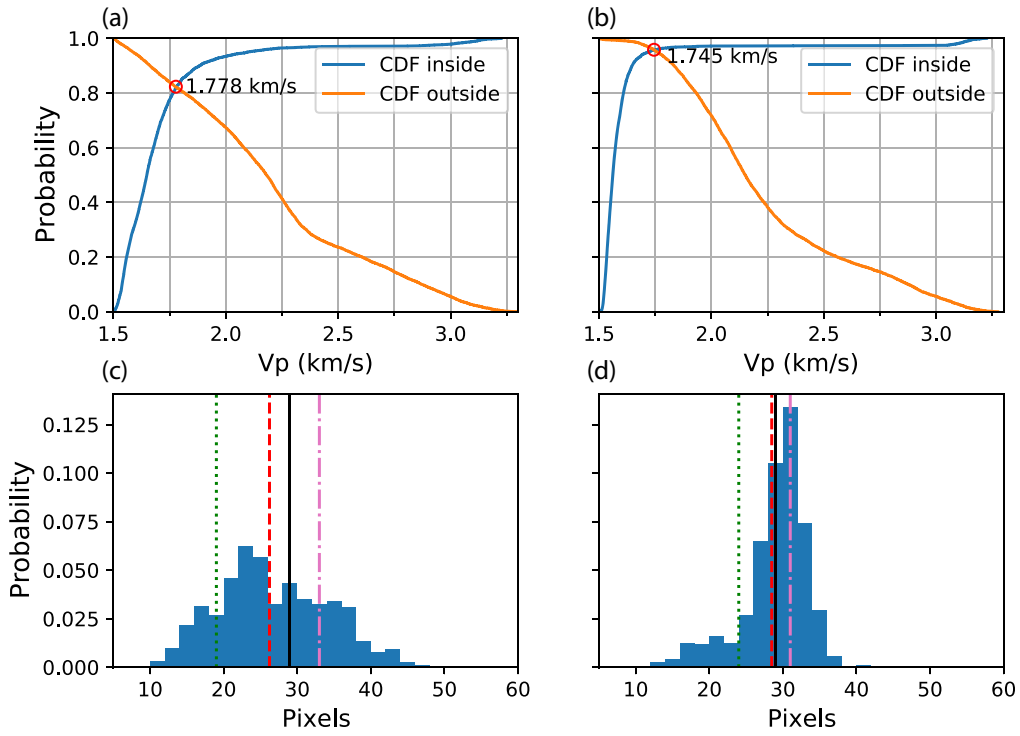

Figure 5. (a) The cumulative density functions (CDF) of points that are most likely within the low-velocity anomaly accumulated in the positive Vp direction (blue line) and outside the low-velocity anomaly accumulated in the negative Vp direction (orange line), both obtained using the low-frequency data. (b) Same as (a) but obtained using the high-frequency data. The red circle denotes the velocity value where the two CDFs have the same value, which is the threshold that discriminates low from high velocities with minimal bias. Panels (c) and (d) show the distributions of the low-velocity anomaly size obtained using the low- and high-frequency data, respectively. Red dashed line denotes the optimal answer obtained using interrogation theory (mean of the distribution—eq. 8). Black line denotes the true size. For comparison, green dotted line and pink dashed-dotted line show the values obtained by directly interpreting the mean and median velocity structures, respectively, in the case that one had already conducted a nonlinear inversion to obtain the pdfs in panels (a) and (b) and hence had access to the minimal-bias velocity thresholds.

structures, using the same threshold value. The size obtained using the mean structure has the largest error, which clearly suggests that this structure provides less information about our question of interest. The size obtained using the median velocity structure has smaller error, probably because in this case the marginal pdfs are multimodal (Fig. 4e) and in such cases the median may represent the true structure better than the mean (which may lie between modes). Nevertheless, in both experiments, neither size is as accurate as those obtained using interrogation theory. And finally notice that neither the mean nor the median results could be obtained using only the single estimate of the velocity structure that best fits the data, as then the minimal bias threshold could not be identified.

4 DISCUSSION

The computational cost of constructing Bayesian solutions to scientific problems can be high. For example, the above example took approximately 6704 CPU hours for each data set, which required 74 hr to run using 90 Intel Xeon E5-2630 CPU cores. Particularly in imaging problems, it is typical to communicate only statistics of the posterior pdf, often the mean or median, and some measure of uncertainty such as point-wise standard deviations. This study highlights the information loss incurred in such communications: not only are the answers derived above from the mean and median models relatively inaccurate, but they could not be calculated at all without additionally having samples of the posterior pdf in order to estimate an unbiased threshold that discriminates low-velocity zones. We therefore advocate that methods to communicate the correlated structure of the posterior pdf are devised and used in future scientific communications. For example, one could publish not only statistics but also all samples of the posterior pdf that are computed as a matter of course by methods such as Monte Carlo and SVGD; alternatively, one could publish the parameters of solutions expressed as normalizing flows (Zhao *et al.* 2021) or invertible neural networks (Zhang & Curtis 2021b), both of which provide posterior samples almost for free. This would have the additional advantage that the same posterior pdf could be interrogated for answers to different questions thereafter. If this approach was widely adopted, the informational value of estimated posterior pdfs would increase, offsetting the cost of their computation.

In the above example, we only used one model m, that is a seismic velocity structure and a forward wavefield simulator, and the distribution of parameter values was estimated using only one method of FWI. Eq. (8) allows us to use multiple models. For example, different parametrizations can be included as different models in eq. (8) to account for the uncertainty caused by specific parametrizations. To assess effects caused by different parametrizations, in Appendix A we show the distribution of the estimated size of the low-velocity anomaly obtained using a lower dimensional parametrization (a 100 $\times$ 60 grid) using the low-frequency data. The results show that the distribution is wider than that obtained using the denser grid (200 $\times$ 120) and the optimal answer is larger than the true answer, which is probably because the coarser grid imposes additional smoothness to the structure and consequently the size is overestimated.

The sizes estimated using only the mean and median structures are very different from each other, which again demonstrates that the answer obtained using one single model is not sufficient to answer questions accurately.

Apart from different parametrizations, one could combine different properties of the Earth, for example seismic velocity structures obtained using seismic methods with resistivity structures obtained using electromagnetic methods to answer the same questions. In addition, estimates of posterior pdfs obtained using different inference algorithms may also be combined to answer questions, such that the uncertainty caused by different algorithms can be taken into account in the procedure.

In this study, we assumed a fixed experimental design that may not be an optimal design. To better quantify the answer to a scientific question, an optimal experimental design focused on that question may also be found within the framework of interrogation theory (Arnold & Curtis 2018). However, this involves calculating the integration in eq. (5) multiple times in order to find the optimal value in eq. (6), which might be computationally intractable in practice. In such cases, deterministic methods may be more useful to find approximately optimal designs (Curtis *et al.* 2004; Guest & Curtis 2009; Bloem *et al.* 2020). The utility function in eq. (2) can also be used as a criterion for selecting one among a small number of candidate experimental designs. We also note that there are already geophysical studies that either vary the design of geophysical data sets in near-real time (e.g. Wilkinson *et al.* 2015) or consider many different model parametrizations (Bodin & Sambridge 2009). Interrogation theory would allow such designs and models to be adaptively parametrized to the question and space of answers of interest, rather than adapting only to the parameters of the model.

In this study, we demonstrated interrogation theory by using a 2-D FWI example. The same idea might be too computationally expensive to be applied to 3-D FWI. However, the theory can also be applied to other studies that have been solved using Bayesian inference. For example, nonlinear traveltime tomography problems have been solved using the Monte Carlo method in both 2-D (Bodin & Sambridge 2009) and 3-D (Zhang *et al.* 2018), so the posterior distributions obtained in such cases can be interrogated to answer specific questions.

In reality after a round of interrogation, the investigator may find that the answer to the question is not sufficiently constrained by the data. In such cases, another interrogation can be conducted using an experimental design that is optimized based on the knowledge obtained in the first interrogation. In addition, the information obtained in the first interrogation can be used as prior information for subsequent inversions that can significantly reduce the computational cost (Zhang & Curtis 2021a). This process can be repeated until a satisfactory answer is found.

5 CONCLUSIONS

We used interrogation theory to quantify the size of a subsurface structure by interrogating the probabilistic results obtained from Bayesian seismic FWI. The results demonstrated that the size obtained using interrogation theory provides an accurate estimate of the true size, which cannot be obtained using only one single estimate of the velocity structure. This shows that fully nonlinear uncertainty estimates are important for answering scientific questions, partly justifying their additional computational cost. We expect that the theory can be used to find answers for a range of real-world scientific questions, in particular for quantitative interpretation of geophysical inversion results to better understand the Earth.

ACKNOWLEDGEMENTS

The authors thank the Edinburgh Imaging Project sponsors (BP, Schlumberger and Total) for supporting this research. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/).

DATA AVAILABILITY STATEMENT

There are no real data used in this study. The posterior samples used in this study are available in the paper and in its online supplementary material.

REFERENCES

Arnold, R. & Curtis, A., 2018. Interrogation theory, *Geophys. J. Int.,* **214**(3), 1830–1846.

Aster, R.C., Borchers, B. & Thurber, C.H., 2018. *Parameter Estimation and Inverse Problems,* Elsevier.

Blei, D.M., Kucukelbir, A. & McAuliffe, J.D., 2017. Variational inference: a review for statisticians, *J. Am. Stat. Assoc.,* **112**(518), 859–877.

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

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178**(3), 1411–1436.

Bond, C., Lunn, R., Shipton, Z. & Lunn, A., 2012. What makes an expert effective at interpreting seismic images?, *Geology,* **40**(1), 75–78.

Brooks, S., Gelman, A., Jones, G. & Meng, X.-L., 2011. *Handbook of Markov Chain Monte Carlo,* CRC Press.

Curtis, A., Michelini, A., Leslie, D. & Lomax, A., 2004. A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys, *Geophys. J. Int.,* **157**(2), 595–606.

Guest, T. & Curtis, A., 2009. Iteratively constructive sequential design of experiments and surveys with nonlinear parameter–data relationships, *J. geophys. Res.,* **114**(B4), doi:10.1029/2008JB005948.

Liu, Q. & Wang, D., 2016. Stein variational gradient descent: a general purpose Byesian inference algorithm, in *Advances in Neural Information Processing Systems,* pp. 2378–2386.Curran Associates, Inc.

Martin, G.S., Wiley, R. & Marfurt, K.J., 2006. Marmousi2: an elastic upgrade for Marmousi, *Leading Edge,* **25**(2), 156–166.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Nawaz, M.A. & Curtis, A., 2018. Variational Bayesian inversion (VBI) of quasi-localized seismic attributes for the spatial distribution of geological facies, *Geophys. J. Int.,* **214**(2), 845–875.

Polson, D. & Curtis, A., 2010. Dynamics of uncertainty in geological interpretation, *J. Geol. Soc.,* **167**(1), 5–10.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.,* **138**(2), 479–494.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 3–1.

Stuart, A.M., 2010. Inverse problems: a Bayesian perspective, *Acta Numer.,* **19,** 451–559.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Vol. 89, SIAM.

Tarantola, A. & Valette, B., 1982. Inverse problems= quest for information, *J. Geophys.,* **50**(1), 159–170.

Wilkinson, P.B., Uhlemann, S., Meldrum, P.I., Chambers, J.E., Carriere, S., Oxby, L.S. & Loke, M., 2015. Adaptive time-lapse optimized survey design for electrical resistivity tomography monitoring, *Geophys. J. Int.,* **203**(1), 755–766.

Zhang, X. & Curtis, A., 2020a. Seismic tomography using variational inference methods, *J. geophys. Res.,* **125**(4), e2019JB018589.
Zhang, X. & Curtis, A., 2020b. Variational full-waveform inversion, *Geophys. J. Int.,* **222**(1), 406–411.
Zhang, X. & Curtis, A., 2021a. Bayesian full-waveform inversion with realistic priors, *Geophysics,* **86**(5), 1–20.
Zhang, X. & Curtis, A., 2021b. Bayesian geophysical inversion using invertible neural networks, *J. geophys. Res.,* **126,** e2021JB022320.
Zhang, X., Curtis, A., Galetti, E. & de Ridder, S., 2018. 3-D Monte Carlo surface wave tomography, *Geophys. J. Int.,* **215**(3), 1644–1658.
Zhang, X., Nawaz, M.A., Zhao, X. & Curtis, A., 2021. Chapter two - an introduction to variational inference in geophysical inverse problems, in *Inversion of Geophysical Data, Vol. 62 of Advances in Geophysics,* pp. 73–140, ed. Schmelzbach, C., Elsevier.
Zhao, X., Curtis, A. & Zhang, X., 2021. Bayesian seismic tomography using normalizing flows, *Geophys. J. Int.,* **228**(1), 213–239.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

suppl_data.

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.

APPENDIX A: EFFECTS OF PARAMETRIZATION

To assess effects of different parametrizations on the answer to a question, we conduct another inversion using a lower dimension parametrization (a 100 × 60 grid) using the same low-frequency data as in Section 3. The forward simulation is performed using the same denser grid (200 × 120) which is interpolated from the coarser grid, and the SVGD is conducted in exactly the same way as described in Section 3. The obtained particles are then processed in the same way to estimate the size of the low-velocity anomaly. To be comparable to the results presented in Section 3, we display the size as the number of dense grid pixels.

Fig. A1 shows the obtained mean velocity structure and the associated size estimate. The distribution of the estimated size is wider than that obtained using the denser grid and the optimal answer (red dashed line) is larger than the true answer (black solid line). This is probably because the coarser grid imposes extra smoothness on the velocity structure and consequently the size is overestimated. Similar to the results presented in Section 3, the size obtained by directly interpreting the mean structure (green dashed line) and the median structure (pink dashed line) are very different from each other, which again suggests that the answer obtained using only one single velocity structure is not sufficient.

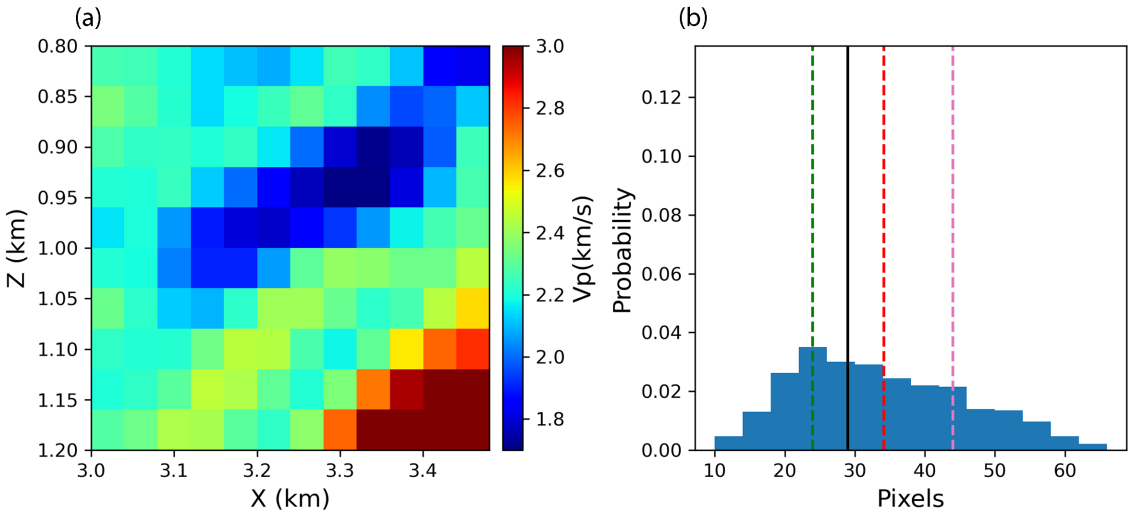

Figure A1. The (a) mean velocity structure and (b) distribution of the low-velocity anomaly size obtained using a lower dimensional parametrization (a 100 × 60 grid). Lines denote the same value as in Fig. 5.