Interrogating Subsurface Structures using Probabilistic Tomography: an example assessing the volume of Irish Sea basins

Xuebin Zhao[1], Andrew Curtis[1], and Xin Zhang[2]

[1]University of Edinburgh
[2]School of Geosciences

November 24, 2022

Abstract

The ultimate goal of a scientific investigation is usually to find answers to specific, often low-dimensional questions: what is the size of a subsurface body? Does a hypothesised subsurface feature exist? Existing information is reviewed, an experiment is designed and performed to acquire new data, and the most likely answer is estimated. Typically the answer is interpreted from geological and geophysical data or models, but is biased because only one particular forward function is considered, one inversion method is applied, and because human interpretation is a biased process. Interrogation theory provides a systematic way to answer specific questions by combining forward, design, inverse and decision theories. The optimal answer is made more robust since it balances multiple possible forward models, inverse algorithms and model parametrizations, probabilistically. In a synthetic test, we evaluate the area of a low-velocity anomaly by interrogating Bayesian tomographic results. By combining the effect of four inversion algorithms, the optimal answer is very close to the true answer, even on a coarsely gridded parametrisation. In a field data test, we evaluate the volume of the East Irish Sea basins using 3D shear wave speed depth inversion results. This example shows that interrogation theory provides a useful way to answer realistic questions about the Earth. A key revelation is that while the majority of computation may be spent solving inverse problem, much of the skill and effort involved in answering questions may be spent defining and calculating those target function values in a clear and unbiased manner.

Interrogating Subsurface Structures using Probabilistic Tomography: an example assessing the volume of Irish Sea basins

Xuebin Zhao[1], **Andrew Curtis**[1] **and Xin Zhang**[1]

[1]School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom

Key Points:

- We use *interrogation theory* to answer specific questions about the subsurface using probabilistic tomography results.
- This method is shown to give accurate answers about high resolution structures, even given only low resolution tomographic images.
- We apply the method to a real data set and evaluate the volume of the East Irish Sea sedimentary basins using 3D depth inversion results.

Corresponding author: Xuebin Zhao, `xuebin.zhao@ed.ac.uk`

Abstract

The ultimate goal of a scientific investigation is usually to find answers to specific, often low-dimensional questions: what is the size of a subsurface body? Does a hypothesised subsurface feature exist? Existing information is reviewed, an experiment is designed and performed to acquire new data, and the most likely answer is estimated. Typically the answer is interpreted from geological and geophysical data or models, but is biased because only one particular forward function is considered, one inversion method is applied, and because human interpretation is a biased process. Interrogation theory provides a systematic way to answer specific questions by combining forward, design, inverse and decision theories. The optimal answer is made more robust since it balances multiple possible forward models, inverse algorithms and model parametrizations, probabilistically. In a synthetic test, we evaluate the area of a low-velocity anomaly by interrogating Bayesian tomographic results. By combining the effect of four inversion algorithms, the optimal answer is very close to the true answer, even on a coarsely gridded parametrisation. In a field data test, we evaluate the volume of the East Irish Sea basins using 3D shear wave speed depth inversion results. This example shows that interrogation theory provides a useful way to answer realistic questions about the Earth. A key revelation is that while the majority of computation may be spent solving inverse problem, much of the skill and effort involved in answering questions may be spent defining and calculating those target function values in a clear and unbiased manner.

Plain Language Summary

This paper shows how to answer specific questions about the subsurface using probabilistic tomography. Usually tomographic methods are used to estimate images of the subsurface; the 'best' images are then interpreted to answer questions of interest. This work shows that by setting up a formal target function that allows any image to be interpreted automatically, many samples of possible subsurface models can be translated into probabilistic answers to the questions, from which a least-biased answer can be constructed. In the real-data examples presented here the subsurface shape of a sedimentary basin is determined automatically, and a least-biased estimate of its volume is constructed. This method is shown to give accurate answers about high resolution structures even given only low resolution tomographic images; this suggests that the probabilistic results compensate for the lack of resolution.

1 Keywords

Bayesian Inference, Seismic Tomography, Imaging, Probability distribution, Uncertainty Analysis

2 Introduction

Scientific investigations are usually initiated to answer high-level questions posed by investigators. Answers to these questions often lie within low-dimensional spaces: what is the depth of the Moho beneath a particular location? What is the best location to place a new sensor given locations of pre-existing sensors? Can this subsurface aquifer be used for carbon storage? The answers to each of these questions are binary (yes/no) or low-dimensional (Moho depth or sensor location), yet they may depend on high-dimensional parameter spaces, describing the structure of Earth's subsurface for example. We usually seek answers using information that we know already – so-called *prior* information, and to better constrain the answer we collect new data. This involves designing an experiment, acquiring new data by experimentation, and interpreting the data to produce new and useful information. Finally the question is answered by taking both the prior information and the information from new data into account.

More formally, the new data is used to solve a Bayesian inverse problem in which we update the prior information with new information from the data, and seek to describe the resultant state of information by a probability distribution (Tarantola, 2005). Generally, inversion methods can be divided into two categories: linearised and non-linear methods. The former iteratively approximates the possibly complex and non-linear model-data relationship (the *forward* function) by a linear relationship, after which the inverse problem can be solved by minimizing a predefined objective function that measures the misfit between the observed data and synthetic data simulated from a given Earth model (Jackson, 1972). This kind of method requires a good initial model to avoid converging to local minima. In addition, it is not known how to estimate uncertainty or probability robustly from linearised inversion results, which means that we fail to find the solution to the Bayesian inverse problem. This in turn introduces bias when we use the results to answer questions of interest.

In contrast to linearised methods, fully non-linear inversion methods solve the inverse problems under a probabilistic framework. They estimate or characterise the full probabilistic inversion results that describe all information about model parameters given the data – the so called *posterior* probability distribution or density function (pdf). Such problems are often solved using Markov chain Monte Carlo (McMC) which generates an ensemble of samples of the posterior distribution that fit the observed data to within measured data uncertainties. Many different kinds of McMC methods have been introduced for geophysical inversion, e.g.: Metropolis Hastings McMC (MH-McMC) (Mosegaard & Tarantola, 1995), Reversible Jump McMC (RJ-McMC) (Bodin & Sambridge, 2009; Bodin et al., 2012; Galetti et al., 2015, 2017; X. Zhang et al., 2018), Hamiltonian Monte Carlo (HMC) (Fichtner & Simutė, 2018; Fichtner et al., 2019; Gebraad et al., 2020), informed proposal Monte Carlo (Khoshkholgh et al., 2021) and so on. All of these methods become very expensive when dealing with high-dimensional inference problems due to the curse of dimensionality (Curtis & Lomax, 2001). In an attempt to improve the computational efficiency, approaches have been proposed to solve non-linear Bayesian inverse problems using an optimization framework. These include neural network (NN) inversion (Devilee et al., 1999; Meier et al., 2007; Käufl et al., 2014, 2016; Earp & Curtis, 2020; Siahkoohi, Rizzuti, & Herrmann, 2021; Singh et al., 2021) and variational inference (Nawaz & Curtis, 2018, 2019; Nawaz et al., 2020; X. Zhang & Curtis, 2020a; Zhao et al., 2021; X. Zhang et al., 2021; Siahkoohi, Orozco, et al., 2021; Siahkoohi, Rizzuti, Louboutin, et al., 2021). However, the relative efficiency of all of the above methods depends on the problem at hand (Wolpert & Macready, 1997).

The probabilistic results of the inverse problem can be used to answer questions. For non-linear inversion, a common way to achieve this is to interpret the mean model. For example, if we wish to estimate the size of a subsurface structure or feature using Bayesian tomographic inversion results, an intuitive way to proceed is to estimate its size using the mean seismic velocity map. However, answering questions using the mean model alone can be inaccurate since the mean model is only a single statistic of the posterior distribution and may not even represent a model that fits the observed data. In addition, human interpretation is a biased process, which sometimes leads incorrect answers as we show in an example below. In addition, since uncertainty in the result of the inverse problem is not considered, we cannot estimate uncertainty in the answers. Indeed, most of the information within the posterior distribution is summarily discarded when answering questions in this manner, which is extremely wasteful considering the computational cost of Bayesian inversion in non-linear problems.

To address the above deficiencies, we suggest to answer questions using *interrogation theory*, a structured framework to design scientific investigations (Arnold & Curtis, 2018). It combines inverse theory, decision theory and the theory of experimental design to optimise scientific investigations so as to find information that best answers scientific questions of interest. In this paper, we test interrogation theory on real data by using Bayesian non-linear inversion results to answer a specific type of question: *what is the size of a near-surface geological body?* In our test the result is compared to the answer estimated from surface geological mapping.

The rest of this paper is organized as follows. In the next section, we summarise the key components of interrogation theory and how we augment that theory in this paper, and show how optimal answers may be derived using Bayesian inversion results. In section 3, we establish a detailed interrogation procedure using a synthetic example which estimates the area of a subsurface low velocity body based on probabilistic tomographic results. By using a coarse grid parametrisation, we show that human interpretation can be significantly in error, yet in the same case the answer provided by interrogation theory remains accurate. In section 4, we use interrogation theory to answer two real-world questions about the East Irish Sea sedimentary basins. Finally, we provide a brief discussion about this work and draw conclusions.

3 Theory

3.1 Bayesian Inverse Theory

Inverse theory is used to estimate the vector model parameter $\mathbf{m}$ given some observed data $\mathbf{d}$, as shown in Figure 1a. This usually includes solving a forward problem that generates synthetic data corresponding to any parameter $\mathbf{m}$ using a predefined *forward* function $\boldsymbol{f}(\mathbf{m})$. The parameter space is then explored to find values that match the observed data to within their uncertainties. In a Bayesian framework, the inverse problem is solved in a probabilistic way by evaluating the so-called *posterior* probability density function (pdf) $p(\mathbf{m}|\mathbf{d})$ – the probability of model parameter $\mathbf{m}$ given observed data $\mathbf{d}$ – using Bayes' theorem:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})} \tag{1}$$

Here $p(\mathbf{m})$ is the *prior* pdf of model parameter $\mathbf{m}$, that is, the information we know about $\mathbf{m}$ prior to the inversion. The conditional probability $p(\mathbf{d}|\mathbf{m})$ is the *likelihood* of observing data $\mathbf{d}$ given a particular set of values for parameter vector $\mathbf{m}$, and is used to measure how consistent are the sample and the data. In the denominator, $p(\mathbf{d})$ is a normalization constant called the *evidence*.

Markov chain Monte Carlo (McMC) is often used to solve Bayesian inference problems by sampling from the posterior distribution directly, yet it is often highly, if not impossibly, expensive to sample it with representative density due to the curse of dimensionality (Curtis & Lomax, 2001). As an alternative, variational inference solves Bayesian inversion using an optimization framework by seeking the best approximation to the posterior distribution. This can be accomplished by minimizing the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951) between the approximated (so-called variational) distribution and the posterior distribution (Bishop, 2006; Blei et al., 2017; C. Zhang et al., 2018; Nawaz & Curtis, 2018, 2019; Nawaz et al., 2020; X. Zhang & Curtis, 2020a; Zhao et al., 2021; X. Zhang et al., 2021; Siahkoohi, Rizzuti, Louboutin, et al., 2021; Siahkoohi, Rizzuti, & Herrmann, 2021). In this work we combine results from both Monte Carlo and variational algorithms C.

3.2 Interrogation Theory

3.2.1 Fundamentals

Figure 1b outlines the key components of an interrogation problem, and a more detailed algorithmic flow chart is illustrated in Figure 2. Rather than focusing on the model parameter $\mathbf{m}$ in an inverse problem, interrogation theory orientates all theory around a scientific question Q and corresponding optimal answer a^*, which usually lies in a low-dimensional space $\mathbb{A}$. For example, geoscientists may be interested in the volume of a particular subsurface reservoir; the answer to this question would be a (1-dimensional) positive number. For other cases we may pose a binary question such as: is there a geothermal plume beneath this area? The answer would be yes or no. Since low-dimensional answers often lie within high-dimensional model parameters, which are constrained by high-dimensional data, it is hard to interpret data and answer questions directly. Interrogation theory provides a systematic way to investigate optimal answers to those questions.

As illustrated in Figure 2, at the beginning of an interrogation problem, investigators pose a question Q of interest given some background knowledge B. To answer this

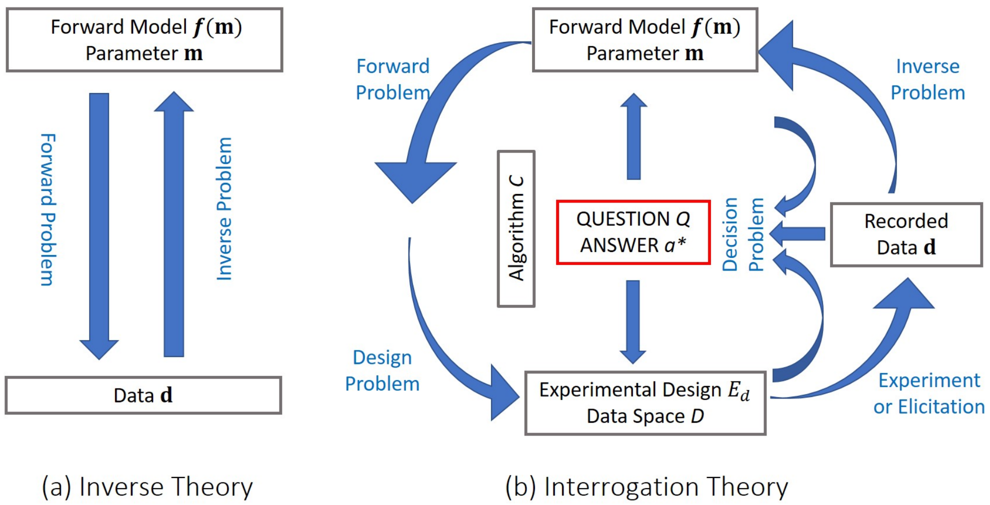

Figure 1. Comparison between inverse theory and interrogation theory. (a) Inverse theory: given observed data $\mathbf{d}$, we estimate model parameter $\mathbf{m}$. This is accomplished by evaluating the data match between the observed data and synthetic data simulated by solving a forward problem $\boldsymbol{f}(\mathbf{m})$. (b) Interrogation theory: given a scientific question or set of questions Q, we wish to find the optimal answer a^*. Forward, design, inverse and decision problems are solved together to maximise information about the answer to question Q, rather than about parameter $\mathbf{m}$. In addition, in this paper the effect of different computational algorithms C for solving these problems is considered to reduce the bias of the final answer.

question, we first define a space of forward models $\mathbb{F}(\mathbb{M})$ in which all of the forward functions are deemed relevant to the question Q. Each element $\boldsymbol{f}(\mathbf{m})$ maps parameter space into corresponding data space, and has a prior density functional $p(\boldsymbol{f}(\mathbf{m}))$ which states the probability that this specific forward function $\boldsymbol{f}(\mathbf{m})$ would accurately represent the parameter-data relationship. The set of forward models satisfies: $\sum_{\boldsymbol{f}\in\mathbb{F}} p(\boldsymbol{f}(\mathbf{m})) = 1$, assuming that the space of forward function is discrete. For each forward model, we define the corresponding model parameter $\mathbf{m}$ and its prior probability distribution $p(\mathbf{m}|\boldsymbol{f}(\mathbf{m}))$ such that $\int_{\mathbf{m}} p(\mathbf{m}|\boldsymbol{f}(\mathbf{m}))d\mathbf{m} = 1$ where the integration is over the entire parameter space. For example, assume we are facing a seismic tomography related project. In this project, we use the following two forward functions to map subsurface velocity structure $\mathbf{m}$ into corresponding first arrival travel time data $\mathbf{d}$ between sources and receivers: ray tracing ($\boldsymbol{f}_1(\mathbf{m})$ – Julian et al., 1977) and the fast marching method ($\boldsymbol{f}_2(\mathbf{m})$ – Rawlinson & Sambridge, 2004). Since the former may fail to find the shortest travel time (the correct ray path) and is not robust for complex velocity structures, whereas the latter is capable of predicting travel times accurately in complex media, we assign prior probability density for these two forward functions as $p(\boldsymbol{f}_1(\mathbf{m})) = 0.2$ and $p(\boldsymbol{f}_2(\mathbf{m})) = 0.8$ respectively. For both forward functions, we use the same Uniform distribution to define our prior information on model parameter $\mathbf{m}$.

To answer question Q, we usually need some additional information, which is obtained by collecting new data. Given a set of forward models $\boldsymbol{f}(\mathbf{m})$, an experimental design problem is solved to select the optimal design E_d to acquire data, selected from the space of designs $\mathbb{E}_d$. The difference between the design problem mentioned here and traditional experimental design problems (e.g., Maurer et al., 2010) is that the former finds a design that is chosen to provide the most relevant information to answer question Q, whereas the latter finds a design that best constrains model parameter $\mathbf{m}$. After implementing the experiment, the recorded data is used to update information about model parameter $\mathbf{m}$ by solving an inverse problem, after which we can answer question Q.

Usually a variety of different computational algorithms can be used to solve forward, design and inverse problems. These may provide significantly different solutions. For example, Zhao et al. (2021) illustrated that different results were obtained when solving the same Bayesian tomographic problem with four different inversion algorithms. Choosing any one of those results is likely to bias any inferred answer to question Q. To reduce bias in the optimal answer, in this paper we account for uncertainties due to the

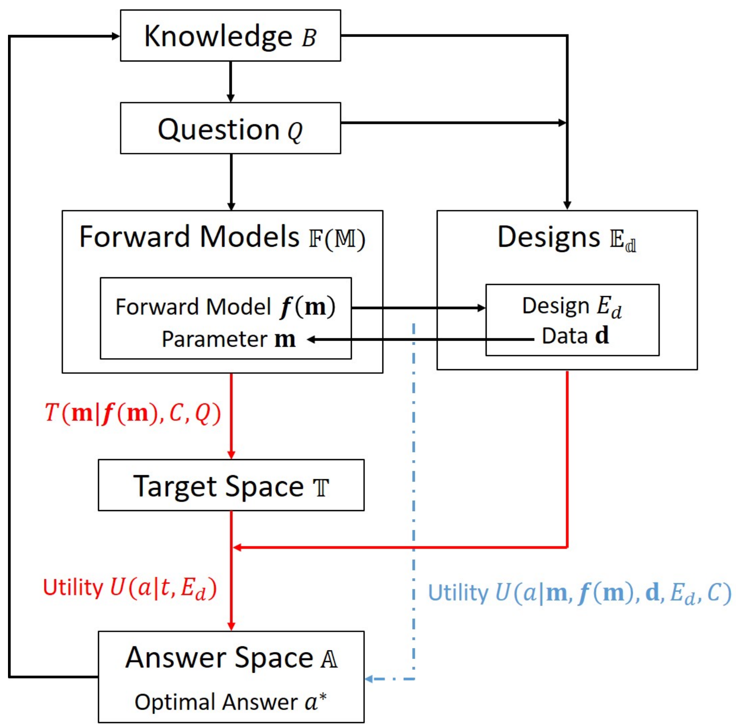

Figure 2. Algorithmic flow chart for interrogation theory. Given background knowledge B and a scientific question Q, we define forward model $\boldsymbol{f}(\mathbf{m})$, the corresponding parameter $\mathbf{m}$, and experimental design E_d to collect new data $\mathbf{d}$. An inverse problem is solved to update the model parameter $\mathbf{m}$ using the acquired data $\mathbf{d}$. A utility function U is constructed and further maximised to obtain the optimal answer a^*. The blue dashed lines show one way to define the utility function by combining all of the above elements directly, which is usually hard to achieve in reality. Instead, we introduce a target space $\mathbb{T}$ and define a target function $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$ to simplify the utility function $U(a|t, E_d)$, as shown by the red lines.

variety of possible computational algorithms C, augmenting the original interrogation framework outlined in Arnold and Curtis (2018).

We define a utility function $U(a)$, which quantifies the net benefits of accepting any particular answer a. The utility is defined such that the optimal answer a^* that maximises the utility function is the one that best satisfies whatever properties we require of our answer (Chaloner & Verdinelli, 1995): $a^* = \arg\max_{a \in \mathbb{A}} U(a)$. Figure 2 shows two approaches to construct the utility function. In the first, we combine all of the information provided in the components of interrogation problems described above, to define a highly structured utility function: $U(a|\mathbf{m}, \boldsymbol{f}(\mathbf{m}), \mathbf{d}, E_d, C)$ as illustrated by dashed blue lines in Figure 2. Note that this utility function is conditioned on the data $\mathbf{d}$ and experimental design E_d to account for the cost of conducting the experiment given a specific design, or to allow the data to provide some components of answer a directly (Arnold & Curtis, 2018). However, the investigator may in general have no means of constructing a utility function of such structure and complexity. Moreover, even when agreeing to a utility function with such a high dimensional set of independent variables, an investigator cannot generally be expected to appreciate all of the consequences of choosing a specific functional form (Curtis & Lomax, 2001). Also, there is no straightforward way to maximise this utility function over the usually discrete choices of forward functions and algorithms under consideration. As an alternative, Arnold and Curtis (2018) introduced a target space $\mathbb{T}$ which is determined by question Q such that Q can be answered directly in $\mathbb{T}$. The target space should be the same for all forward functions $\boldsymbol{f}(\mathbf{m})$ and algorithms C. A target function $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$ is defined to convert the model parameter $\mathbf{m}$ into a target value t. Based on this, a new utility function can be expressed as $U(a|t, E_d)$ which has a much simpler form since it is only conditioned on target value t and design E_d. Usually this is expected to be easier to maximise (shown by red lines in Figure 2).

As an example of a target function, below we will address the question Q, *"What is the volume of a subsurface body?"*. We wish to answer this question using seismic tomographic results. The target function $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$ is defined to transform the model parameter $\mathbf{m}$ – the subsurface velocity structure in this case – into the corresponding volume of the subsurface body of interest. Thus, the target function maps a high-dimensional parameter space into a low-dimensional target space, eliminating nuisance parameters and retaining only information that is essential to represent the answer to the question.

For more details about interrogation theory and these components, we refer readers to Arnold and Curtis (2018).

3.2.2 The Optimal Answer

In this paper, we use the same utility function defined in Arnold and Curtis (2018) – a negative squared error function:

$$U(a|t, E_d) = U(a|t) = -(a-t)^2 \tag{2}$$

in which t is assumed to be the true summarized state of nature in the target space. The utility function in equation 2 is maximized when the estimated answer a is equal to (or is as close as possible to) state t. This results in an analytical solution of the optimal answer a^*: the posterior mean of $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$ averaged over all $\mathbf{m}$, $\boldsymbol{f}(\mathbf{m})$ and C:

$$\begin{aligned} a^* &= \mathbb{E}[T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)|\mathbf{d}, E_d] \\ &= \sum_{\boldsymbol{f}(\mathbf{m}),C} \int_{\mathbf{m}} T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q) p(\mathbf{m}, \boldsymbol{f}(\mathbf{m}), C|\mathbf{d}, E_d)\ d\mathbf{m} \\ &= \sum_{\boldsymbol{f}(\mathbf{m}),C} p\left(C, \boldsymbol{f}(\mathbf{m})\right) \int_{\mathbf{m}} T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q) p(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), \mathbf{d}, E_d, C)\ d\mathbf{m} \\ &= \sum_{\boldsymbol{f}(\mathbf{m}),C} p(\boldsymbol{f}(\mathbf{m})) p\left(C|\boldsymbol{f}(\mathbf{m})\right) \int_{\mathbf{m}} T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q) p(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), \mathbf{d}, E_d, C)\ d\mathbf{m} \end{aligned} \tag{3}$$

where $p(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), \mathbf{d}, E_d, C)$ is the probability of model parameter $\mathbf{m}$ given a specific forward function $\boldsymbol{f}(\mathbf{m})$, observed data $\mathbf{d}$, design E_d and algorithm C, describing the posterior distribution of model parameter $\mathbf{m}$ in Bayesian inversion. Integration in the third line $\int_{\mathbf{m}} T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q) p(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), \mathbf{d}, E_d, C)\ d\mathbf{m}$ calculates the optimal answer given a specific forward model $\boldsymbol{f}(\mathbf{m})$ and computational algorithm C (denoted as $a^*_{\boldsymbol{f}(\mathbf{m}),C}$ below). The third line of equation 3 holds based on the assumption that forward model $\boldsymbol{f}(\mathbf{m})$ and algorithm C are usually independent of design E_d and observed data $\mathbf{d}$. Then, term $p(C, \boldsymbol{f}(\mathbf{m})) = p(\boldsymbol{f}(\mathbf{m}))p\left(C|\boldsymbol{f}(\mathbf{m})\right)$ describes the joint probability density of forward function $\boldsymbol{f}(\mathbf{m})$ and algorithm C, where $p(C|\boldsymbol{f}(\mathbf{m}))$ is the prior probability that a specific algorithm C will find the correct solution given that forward function $\boldsymbol{f}(\mathbf{m})$ does adequately describe the forward physics. Note that C and $\boldsymbol{f}(\mathbf{m})$ are not necessarily independent of each other since some forward functions may preclude the use of different algorithms. For example, we would prefer to use Monte Carlo sampling method if the forward function can be solved cheaply, since the algorithm provides an unbiased approximation of the true solution of a Bayesian inversion problem only as the number of samples becomes large. Therefore we would not consider this algorithm when the forward function is incredibly expensive (for example a full waveform simulator that solves the 3D wave equation). Equation 3 states that the final optimal answer a^* is a weighted sum of $a^*_{\boldsymbol{f}(\mathbf{m}),C}$ over all of the models and algorithms considered. This can be understood intuitively: by considering the effect of different forward models and algorithms, we reduce the bias due to subjective choices and so obtain a more robust interrogation result.

To conclude, equation 3 answers question Q by interrogating the Bayesian inversion results. It also shows that a design E_d that provides optimal answers to question Q would potentially be very different from one designed to maximise information in the posterior distribution $p(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), \mathbf{d}, E_d, C)$ as has been performed in previous research on Geophysical optimal design (e.g., van Den Berg et al., 2003; Guest & Curtis, 2009; Bloem et al., 2020).

4 Implementation

4.1 Problem Statement

Interrogation theory described above can be used to answer many types of real-world questions. In this paper, we provide a specific application to answer volume-related (3D), area-related (2D), or other shape-related questions about a body or medium of interest using fully non-linear tomographic results. This kind of question appears frequently in both academia and industry where we wish to interpret some geological phenomena from geophysical imaging results, such as to estimate the size of a subsurface body, the volume of a reservoir, or the depth of a particular feature such as the Moho under a specific location.

In this section, we use a 2D synthetic example to establish an interrogation procedure for estimating the area of a 2D subsurface body. Figure 3a shows the true velocity model used in this example: a circular low velocity anomaly of 1 km/s is discretised on a grid size of 0.1 km, and located at the centre of the model, and its surrounding area has a high velocity value of 2 km/s. White triangles display the location of 16 receivers (equivalently 16 virtual sources) to collect traveltime data. Given only seismic travel time data from waves that traverse this velocity model, we pose a scientific question: what is the area of the low velocity anomaly?

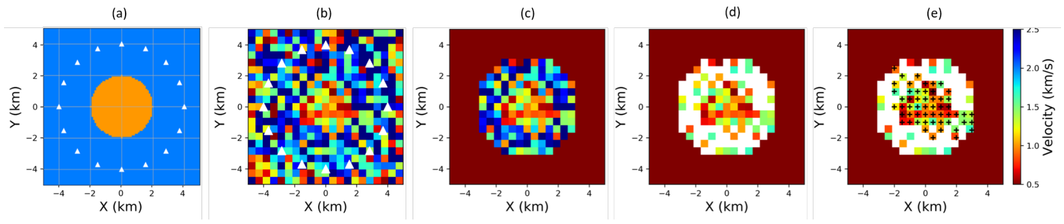

Figure 3. (a) True velocity model used for the 2D synthetic example. (b) A random sample drawn from the posterior distribution of MH-McMC. (c) The same sample in (b) after applying the mask defined in the main text. (d) The retained low velocity pixels after comparing the velocity of every pixel in (c) with the optimal threshold value. (e) Black crosses mark the largest spatially-continuous low velocity body in (d). The defined target function calculates the area of this body.

4.2 Interrogation Procedure

Table 1 summarizes some key elements defined for this interrogation problem. We use the fast marching method (FMM) to represent the model-data relationship. Since this is the only forward model considered in this example, it has a prior probability $p(\boldsymbol{f}(\mathbf{m})) = 1$. The corresponding model parameter $\mathbf{m}$ is the subsurface seismic velocity structure using a regularly-gridded parametrisation, and a Uniform prior distribution is used for the velocity in each cell. To answer the question, we use an experimental design (i.e. source and receiver locations) that contains 16 receivers placed in a circular shape with a radius of 4 km around the low velocity area, as shown by the white triangles in Figure 3a, such that the collected data provides relevant information about the low velocity anomaly. These receivers are also treated as sources, to emulate the use of standard inter-receiver interferometry to provide source to receiver traveltimes (Shapiro et al., 2005; Curtis et al., 2006). Given the collected traveltime data, we solve a Bayesian inference problem to estimate the posterior distribution of the model parameter $\mathbf{m}$. We use four different algorithms to perform non-linear Bayesian tomographic inversion: automatic differential variational inference (ADVI) (Kucukelbir et al., 2017), normalizing flows (Rezende & Mohamed, 2015), Stein variational gradient descent (SVGD) (Liu & Wang, 2016) and Metropolis Hastings McMC (MH-McMC) (Tarantola, 2005); each algorithm is described in Zhao et al. (2021), and the corresponding inversion results are shown in Figures 4a - 4d. The top row of Figure 4 shows the (pixelated) mean velocity maps from the above

Table 1. Key interrogation elements defined for the synthetic test.

Symbol	Meaning	Description
Q	Question	What is the area of the low velocity anomaly?
$\boldsymbol{f}(\mathbf{m})$	Forward model	Fast marching method (FMM)
$\mathbf{m}$	Parameter	Pixelated velocity structure with a Uniform prior pdf
E_d	Design	Source and receiver station locations
$\mathbf{d}$	Data	Source to receiver traveltimes
C	Algorithms	ADVI, Normalizing flows, SVGD and MH-McMC
$T(\mathbf{m})$	Target function	Transform $\mathbf{m}$ into area of low velocity anomaly
$U(a\|t)$	Utility function	$-(a-t)^2$
a^*	Optimal answer	$\mathbb{E}[T(\mathbf{m}\|\boldsymbol{f}(\mathbf{m}), C, Q)\|\mathbf{d}, E_d]$

four methods, while the bottom row shows the corresponding standard deviation maps. In this paper we will not focus on comparing the four inversion results as details about this inversion and a corresponding discussion can be found in Zhao et al. (2021). They concluded that (at least for seismic tomography problems that use FMM as the forward function $\boldsymbol{f}(\mathbf{m})$) ADVI provides an accurate mean velocity model but a biased uncertainty estimation, and the other three methods give similar and accurate mean and uncertainty maps (the same conclusion can be reached by comparing Figure 4a to Figures 4b - 4d). We wish to include the results from ADVI when we determine the optimal answer to the question since this method is relatively efficient and robust (in the sense that the result is highly repeatable), and the mean tends to be accurate in previous tests so it clearly provide information at relatively low computational cost. We downweight the contribution of this algorithm because of the bias expected in its uncertainty estimates by assigning it a relatively low prior probability: $p(C|\boldsymbol{f}(\mathbf{m})) = 0.1$. For the other three algorithms, we assign equal prior values $p(C|\boldsymbol{f}(\mathbf{m})) = 0.3$.

Based on the above elements, we define a target function that maps a posterior sample in high-dimensional parameter space into the area of the central low velocity anomaly in low-dimensional answer space. From the inversion results in Figures 4a - 4d, the low velocity anomaly of interest is located close to the centre of the model. Even though there might be some low velocity anomalies far from the central region, we assume that they

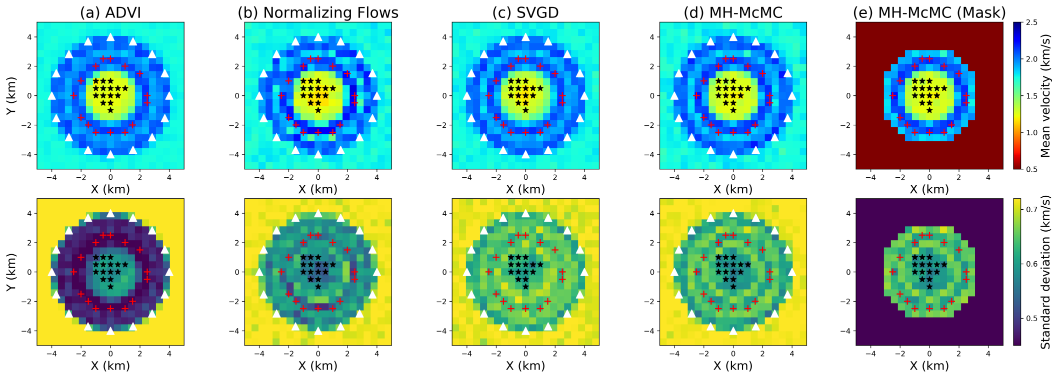

Figure 4. (a) - (d) Pixel-by-Pixel mean (top row) and standard deviation (bottom row) maps of the posterior distributions obtained using ADVI, normalizing flows, SVGD and MH-McMC. (e) The corresponding maps of MH-McMC in (d) after applying the mask introduced in the main text: only the remaining pixels are considered when estimating the area of the low velocity anomaly. White triangles in (a) - (d) illustrate the receiver (and source) locations of the experimental design. Red crosses and black stars in each figure denote the selected pixels used to define the threshold value to discriminate of low and high velocities.

have no relation with the central anomaly in which we are interested since they will be on or outside of the circular array of receivers. To encode this prior assumption, we introduce a mask to confine the region used to calculate the target function. Figure 4e illustrates the effect of the mask, which displays the mean and uncertainty maps of MH-McMC after applying the mask. The area outside of this mask is discarded, and only the remaining velocity pixels are retained to calculate the low velocity area. Thus the target function of each posterior sample $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$ becomes: the area of the low velocity anomaly inside the mask.

Figure 3b shows a posterior sample drawn from the inversion results of MH-McMC, and Figure 3c shows the same sample after applying the defined mask. One way to calculate the target function of this posterior sample is to sum up all of the area of low velocity pixels. This highlights a sub-question that must be answered in order to proceed: "what is the best threshold to discriminate low velocity from high velocity pixels with minimal bias?" If we could estimate such an optimal threshold value, we could classify each pixel as low or high velocity and hence calculate the target function value.

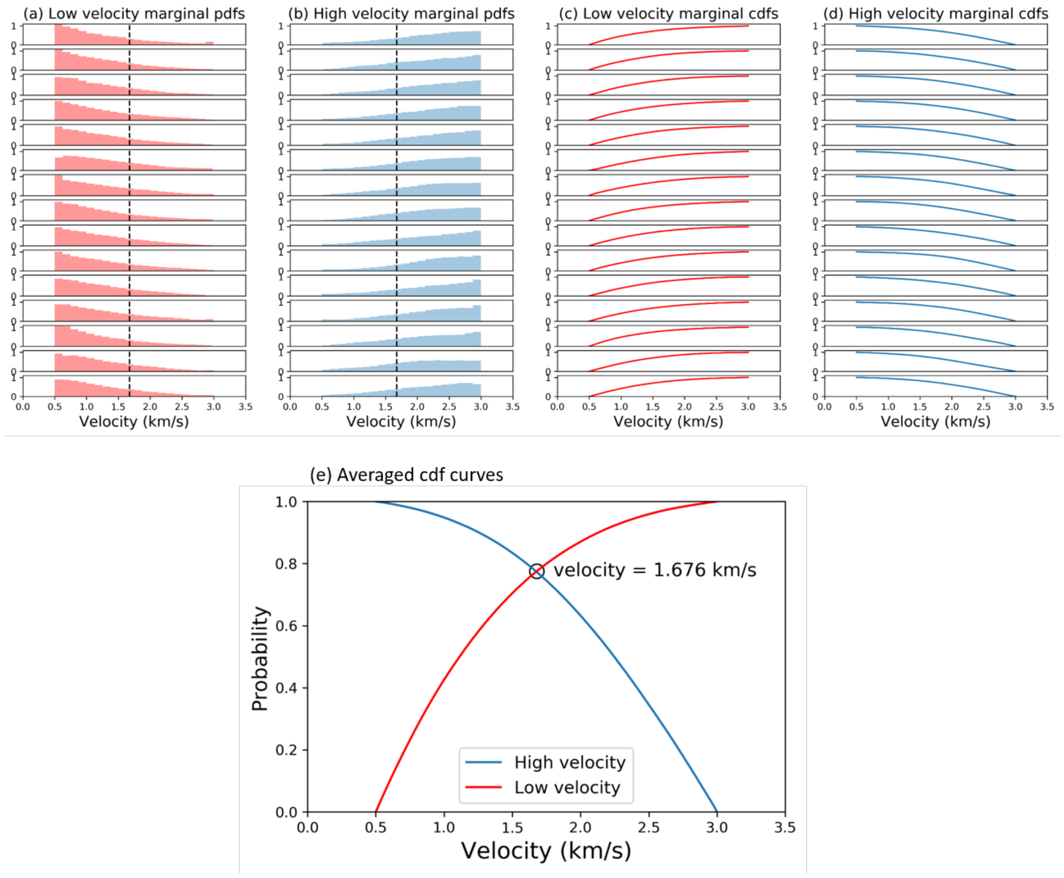

Figure 5. (a) and (b) Marginal pdfs of low and high velocity marked in Figure 4. Dashed black lines denote the crossing point in (e), which is used to classify low and high velocity pixels. (c) and (d) Marginal cdfs obtained by integrating the corresponding pdfs in (a) and (b) in opposite directions. (e) Averaged cdf curves for low (red line) and high (blue line) velocity pixels calculated using (c) and (d). Black dot marks the crossing point of the two curves, and is the threshold value that discriminates low from high velocities with minimal bias.

We define a data-driven way to obtain such a threshold value. Firstly, we pick some pixels that are most likely to be high (and low) velocity cells from the four inversion results. Ideally, these pixels should have higher (lower) mean velocity values relative to the mean, and low uncertainties, as denoted by the red crosses (black stars) in Figure 4. A threshold value estimated from such pixels should represent high and low velocity information better than a value estimated using other, more ambiguous pixels, thus introducing minimal bias.

Figures 5a and 5b show marginal pdfs of the selected low and high velocity pixels, and Figures 5c and 5d display the corresponding marginal cumulative density functions (cdfs). Note that the low velocity marginal cdfs in Figure 5c are obtained by integrating the low velocity marginal pdfs in Figure 5a from low to high velocity (from left to right), whereas the high velocity cdfs in Figure 5d are obtained by integrating the marginals pdfs in Figure 5b in the opposite direction (from high to low velocity). We then average the marginal cdfs in Figures 5c and 5d and plot the averaged cdf curves in Figure 5e. The red line is the averaged cdf for low velocity pixels, and the blue line is that for high velocity pixels, and note that while these curves are close to being mirror images of each other this is not generally the case. The crossing point of the two lines is marked by the black dot with a velocity value of 1.676 km/s. This value is also illustrated by the dashed black line in each pdf curve in Figures 5a and 5b. This point has the property that the probability that the velocities of the selected low velocity pixels (black stars in Figure 4) are lower than this value equals the probability that the velocities of the selected high velocity pixels (red crosses in Figure 4) are higher than this value. This specific threshold value therefore discriminates low from high velocity values with minimal bias.

We compare the velocity value of each pixel in Figure 3c with the optimal threshold, and retain those whose velocity value is smaller than the threshold, as shown in Figure 3d. We interpret these pixels as low velocity bodies in this sample. Question Q demands the area of a single low velocity anomaly, rather than all of the low velocity pixels in Figure 3d. Therefore we add additional prior information that the low velocity anomaly of interest should represent a continuous geological body in space. The target function then becomes the area of the largest continuous low velocity body inside the mask, which is marked by black crosses in Figure 3e (continuity can occur through both laterally and diagonally adjacent pixels). Obviously this target function transforms a high-dimensional velocity vector $\mathbf{m}$ into a (1-dimensional) scalar value, and eliminates nuisance parameters that are less relevant to the question, such that Q can be answered directly in the target space $\mathbb{T}$.

For each of the four inversion results we calculate the target function $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$ for every posterior sample, and plot the corresponding posterior histograms in Figure 6. Given the negative squared error utility function in equation 2, the optimal answer for each algorithm a_C^* can be expressed as the posterior mean of target function $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$

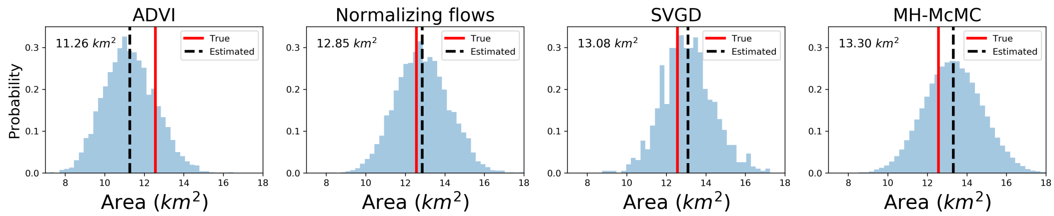

Figure 6. Posterior distributions of the target function for ADVI, normalizing flows, SVGD and MH-McMC, from left to right respectively. The posterior mean value of each target function is displayed at the top-left corner, and is also marked by the dashed black line in each figure. The true answer to this question (12.56 km^2) is denoted by the red lines.

(equation 3), noted at the top-left corner and denoted by the dashed black line in Figure 6. We could further substitute these 4 results, their prior probability values $p(C|\boldsymbol{f}(\mathbf{m}))$, and the prior probability of forward function $p(\boldsymbol{f}(\mathbf{m})) = 1$ into equation 3 to obtain the final optimal answer: 12.89 km^2; this is very close to the true answer (12.56 km^2) which is marked by red lines in Figure 6.

This example illustrates the accuracy of this interrogation procedure. Although the final answer is very close to (even slightly less accurate than) the answer obtained from normalizing flows (12.85 km^2), we usually do not know the true answer to our question for reference, and thus have no means to select the answer from one algorithm over any other. On the other hand, by considering the effect of different algorithms and by defining prior probabilities that each algorithm will provide the correct solution based on their past performance, we would be more confident about the final answer obtained.

Considering the true Earth has infinitely fine structure, whereas in reality we parametrise it with a finite (coarse) grid or number of parameters to reduce the dimensionality of our inverse problem, so it is crucial to consider the effect of different parametrisations when answering questions. In the supporting information associated with this article, we investigate the effect of interrogations carried out using models with different parametrisations. We double the grid size in both directions from 0.5 km to 1 km, which decreases the dimensionality of the tomographic problem from 441 to 121. The results show that both the posterior histograms and their corresponding mean values from the coarser inversion results are quite similar to those obtained from the finer grid parametrisation in Figure 6. The final answer of the coarser grid parametrisation (12.37 km^2) is very close to the true answer (12.56 km^2), as well as that estimated from the finer grid parametrisation (12.89 km^2).

We thus obtain an accurate answer using interrogation theory using either parametrisation. By contrast interpreting the mean map alone provides a severely erroneous answer (9 km^2). This makes interrogation theory more attractive for answering scientific questions since we obtain an accurate answer to the question even under a coarse parametrisation, which usually offers orders of magnitudes of computational cost reduction in real problems.

5 Interrogating the East Irish Sea basins

5.1 Shear Wave Velocity Inversion of the East Irish Sea basins

In the second example, we use interrogation theory to answer questions about the East Irish Sea sedimentary basins. Figure 7a displays 61 seismometer locations (red triangles) around the British Isles used in this test, all of which contain one vertical (Z) and two horizontal (North and East) components to detect ground motion. We consider ambient noise data recorded by these stations during 2001 to 2003, 2006 to 2007 and in 2010. Nicolson et al. (2014) cross-correlated the vertical component of the ambient noise data to estimate inter-receiver traveltimes of Rayleigh waves, and to perform Rayleigh wave tomography of the British Isles. Galetti et al. (2017) used two horizontal components to calculate Love wave group velocity maps at different periods. A more detailed description about the ambient noise data and data processing can be found in Galetti et al. (2017). Since Love waves are dominantly sensitive to the near surface shear velocity structure, we perform shear wave group velocity depth inversion of the East Irish Sea basins using the estimated Love wave traveltime measurements between 4 and 15 s periods, and interrogate the size of those sedimentary basins using the inversion results. Note that the receiver network used in this paper may not be the optimal experimental design to provide the most relevant information about the Irish Sea basins. However, it represents a common situation in seismology where we have fixed legacy designs which are definitely not optimal for every question being posed, and nevertheless wish to find optimal answers to specific questions about the Earth.

We use a two-step scheme for the 3D shear wave group velocity depth inversion. In the first step, we perform Love wave tomography of the British Isles using inter-receiver

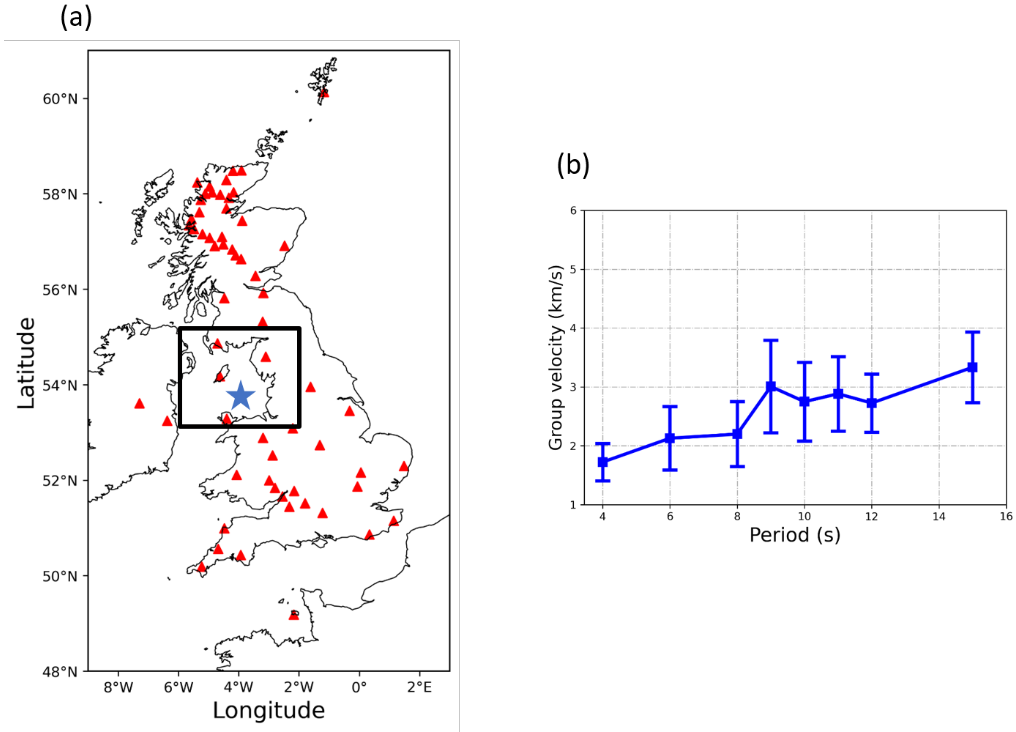

Figure 7. (a) The locations of 61 seismometers (red triangles) around the British Isles used in this paper to record ambient noise data. The recorded data were cross-correlated to provide inter-receiver traveltimes of Love waves at different periods of 4, 6, 8, 9, 10, 11, 12, 15 s (Galetti et al., 2017). We use these data to perform shear wave group velocity depth inversion beneath the East Irish Sea within the black box, via a two-step scheme (see main text for details). (b) One dispersion curve picked from 2D tomographic inversion results of normalizing flows at the geographical point 4°W, 53.5°N, marked by the blue star in (a).

traveltime data at different periods of 4, 6, 8, 9, 10, 11, 12 and 15 s. For each period we perform 2D surface wave tomography, restricting the imaging region to within longitude 9°W – 3°E and latitude 48°N – 61°N, and parametrise the velocity model using a regular grid of 37×40 cells with a spacing of 0.33° in both longitude and latitude directions. The prior distribution is chosen to be a Uniform distribution, and its lower and upper bounds are chosen according to Galetti et al. (2017). The likelihood function is chosen to be a Gaussian distribution, and the traveltime data error of each inter-receiver path is estimated from daily cross-correlations (Galetti et al., 2017). Considering the dimensionality of this fully non-linear inverse problem, we only use three variational methods: ADVI, normalizing flows and SVGD to perform tomography at each period; we do not perform MH-McMC, as the results using that algorithm did not converge acceptably even after drawing 15 million samples in total with 10 chains using 660 hours of elapsed time (Zhao et al., 2021). Previously, Zhao et al. (2021) performed Love wave tomography at 10 s period to compare the performance of different algorithms. In this study we run tomography at all periods, and use these tomographic results to construct dispersion curves at each geographical location. These curves form the dataset that is used to drive the depth inversion (more details on the latter are given below).

Figure 8 shows average velocity maps of the Love wave tomography results using normalizing flows at all of the analysed periods, and Figure 9 shows the corresponding uncertainty results. In order to aid the comparison of velocity structures and uncertainties between the various periods, the same colour scales are used for all of the mean and standard deviation maps in Figures 8 and 9, respectively. Some small structures in Figures 8 and 9 are a bit different compared to those from reversible jump McMC in Galetti et al. (2017) (which uses exactly the same traveltime data for Love wave group velocity tomography). This is due to different parametrisations used in the two studies: Galetti et al. (2017) used a variable parametrisation using Voronoi cells to discretize the velocity model, whereas we use a fixed regularly-gridded parametrisation. Nevertheless, the main features of the mean velocity and uncertainty maps show good consistency with the known geology and previous tomographic studies of the British Isles (Nicolson et al., 2012, 2014; Galetti et al., 2015, 2017). For example, from the tomographic results (especially at smaller periods which usually provide velocity information in the shallow subsurface), we observe a low velocity structure beneath the East Irish Sea within longitude 6°W – 2°W and latitude 53°N – 55°N, marked by the black boxes in Figure 7a and Fig-

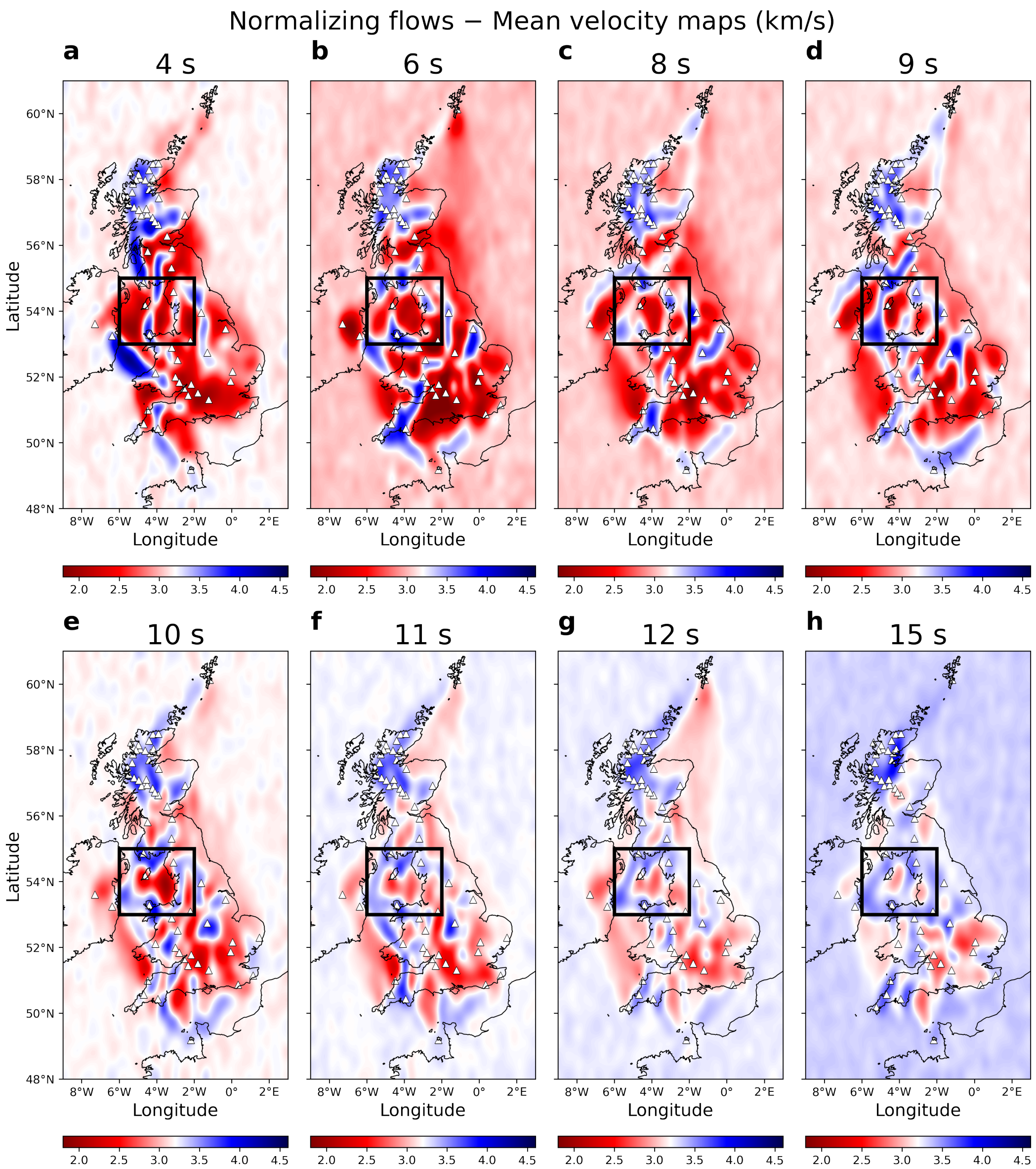

Figure 8. Mean Love wave group velocity maps of the British Isles, interpolated between grid cell locations in the results obtained using normalizing flows at different periods between 4 s and 15 s. All of the mean maps are plotted using the same velocity range for better comparison, and the corresponding period is shown above each map. The black boxes indicate the target region where we pick dispersion curves and perform depth inversion in the second step.

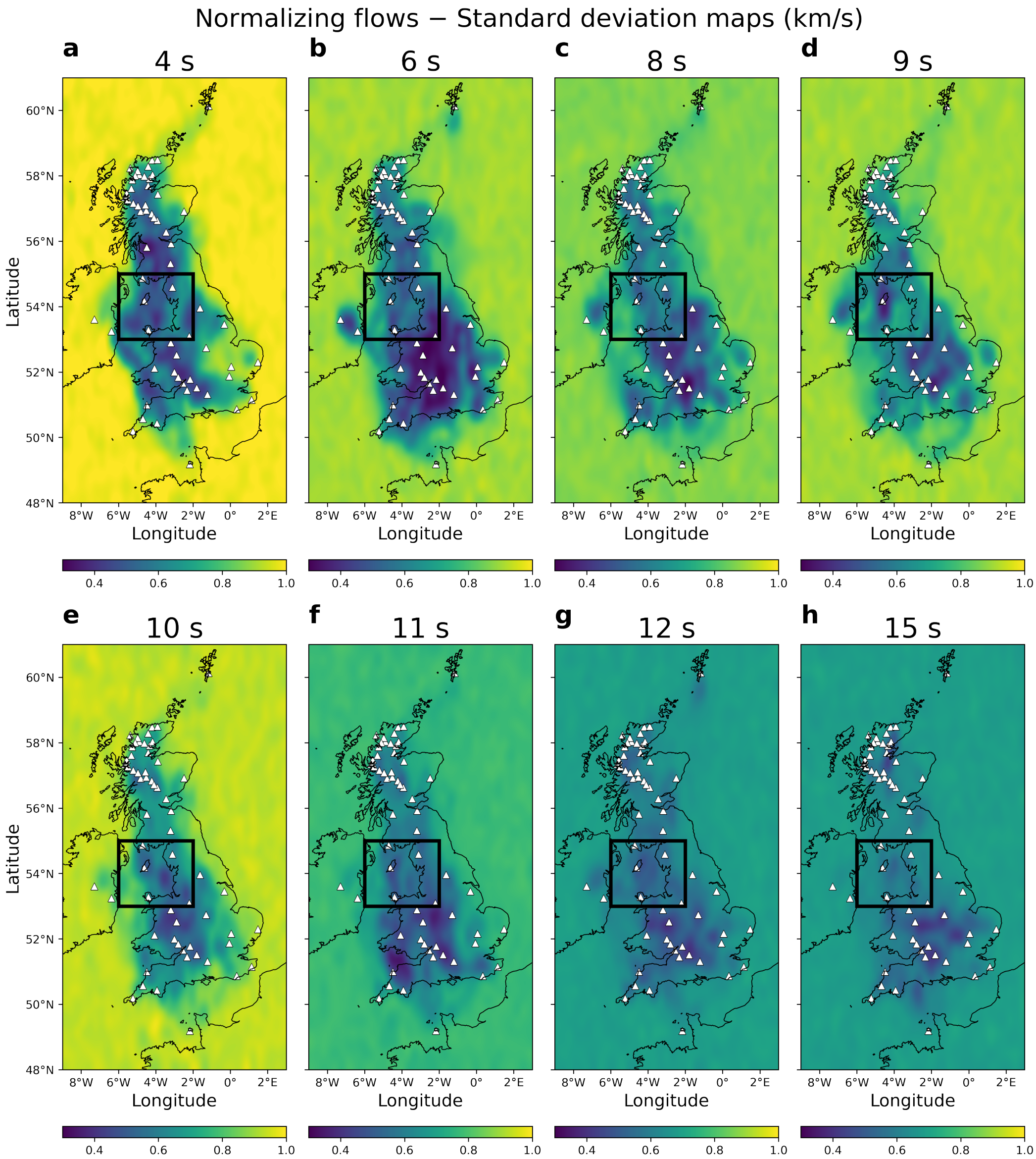

Figure 9. Standard deviation maps of the British Isles, interpolated between grid cell locations in the results obtained using normalizing flows at different periods between 4 s and 15 s, each of which corresponds to one mean velocity map in Figure 8. All the uncertainty maps are plotted using the same range for better comparison, and the corresponding period is shown above each map. The black boxes indicate the target region where we pick dispersion curves and perform depth inversion in the second step.

ures 8 and 9. This low velocity anomaly corresponds to the East Irish sedimentary sedimentary basins (Galetti et al., 2017).

In the second inversion step we focus on the East Irish Sea basins (inside the black box in Figure 7a) and perform dispersion inversion to estimate the 3D shear wave velocity structure at depth using the results from traveltime tomography in the first step. To perform the depth inversion, we construct a dataset of group velocity dispersion curves from the tomographic results. At each geographic point inside the black box in Figure 7a, a dispersion curve can be constructed by taking group velocity values from the 2D mean maps, and uncertainty values from the 2D standard deviation maps at each period. For example, Figure 7b shows one dispersion curve picked from the 2D tomography results in Figures 8 and 9 at 4°W, 53.5°N, the geographical location marked by the blue star in Figure 7a. Given the regular gridded parametrisation scheme we used in the first step, we pick 91 dispersion curves inside the black box around the East Irish Sea.

In order to include lateral spatial correlations in the inversion results, we use the 3D reversible jump Markov chain Monte Carlo (rj-McMC) algorithm of X. Zhang et al. (2018) to perform dispersion inversion in this step. The method parametrises the subsurface velocity model with a 3D Voronoi tessellation, which varies both in shape and number of cells during the inversion. For a given 3D velocity model, the forward problem consists of extracting 1D shear velocity profiles over depth beneath each geographical point, and calculating a group velocity dispersion curve for that 1D structure using a modal approximation (Saito, 1988). Since we obtained different results from the three variational methods in the first step, we obtain three different sets of dispersion curve data for the second step. We therefore perform three independent dispersion inversions to examine the effect of using different algorithms and to reduce the algorithmic bias imposed on our final answer, similar to the approach taken in the synthetic example. For each inversion, the prior distribution is set to be a Uniform distribution on shear velocity in the subsurface between 0.5 and 6 km/s. The prior pdf on the number of Voronoi cells is selected to be a discrete Uniform distribution between 20 and 600 to address the complexity of the shear velocity structure beneath the East Irish Sea. The likelihood function is set to be a Gaussian distribution around the measured data. We perform each inversion by running 16 Markov chains with 3 million iterations, discarding the first 1 million samples from each chain as burn-in, and only retaining every 200th sample there-

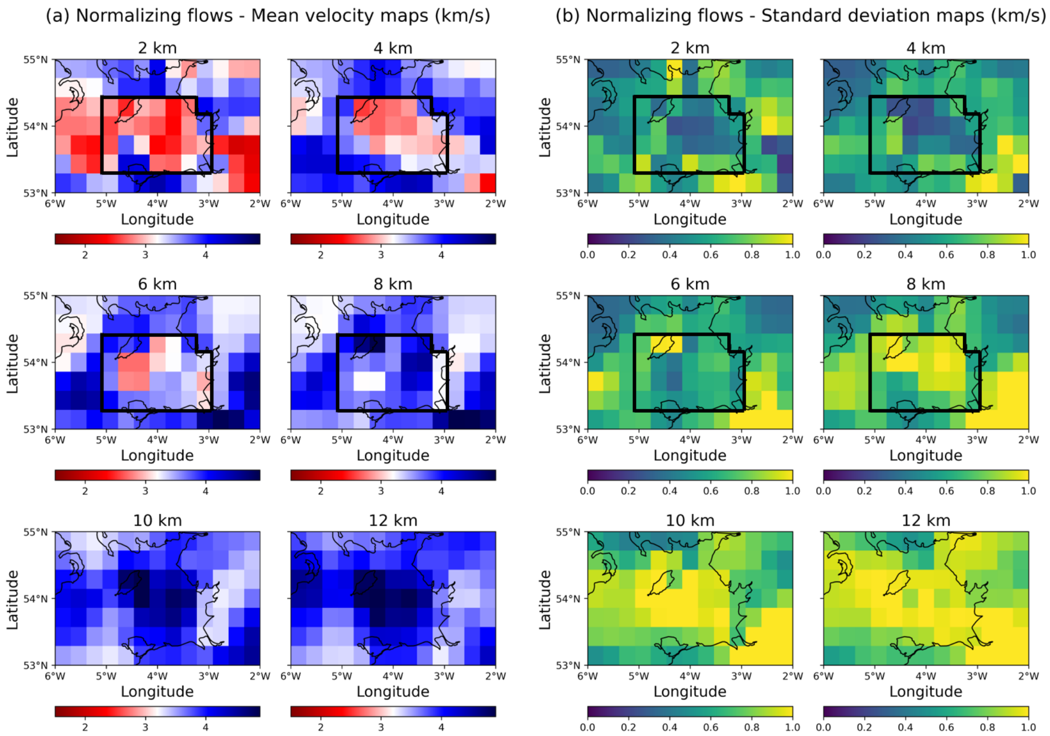

Figure 10. 3D rj-McMC inversion results of shear wave velocity structure constrained by 91 dispersion curves picked from the 2D surface wave tomography results obtained using normalizing flows (within the black boxes in Figures 8 and 9). (a) Mean and (b) standard deviation maps of horizontal slices between 2 km and 12 km depth.

after to calculate statistics of the posterior distribution and to apply interrogation theory.

Considering that each posterior sample is defined using a different 3D Voronoi model parametrisation, we first project all samples onto a regular grid of pixels. In this test, we define a 3D regular grid with a spacing of 0.33° in both latitude and longitude directions and 0.2 km in depth. We then compute the mean group velocity and standard deviation maps across the set of retained samples. Figures 10a and 10b show horizontal slices of the (pixelated) mean and uncertainty maps of the dispersion inversion results between 2 km and 12 km depth, from the inversion result using dispersion data from normalizing flows (Figures 8 and 9). The average shear velocity increases with depth, and the uncertainty also increases since the resolution of Love wave data is lower in the deeper Earth. Again, in Figure 10 we observe similar features compared to those represent by Galetti et al. (2017), which proves the credibility of our results.

From the mean velocity maps in Figure 10a, we can observe a low velocity structure beneath the East Irish Sea down to about 8 km depth, which is interpreted to be the East Irish Sea sedimentary basins in previous studies (Mellett et al., 2015; Galetti et al., 2017). Based on the three inversion results, we attempt to answer scientific questions about these sedimentary basins using the interrogation procedure tested above.

5.2 Estimating the area of the East Irish Sea basins in the shallow subsurface

We first estimate the area of the East Irish Sea sedimentary basins in the shallow subsurface using the top cell of the 3D inversion results which extends from surface down to 200m depth. Figure 11 displays the top cell of the three inversion results. From left to right, each column stands for the average velocity (top row) and uncertainty (bottom row) maps of the inversion results using dispersion curve data picked from 2D tomographic results obtained using (a) ADVI, (b) normalizing flows and (c) SVGD (the three variational methods used in the first step only provide different dispersion curves for the second step, and we use the same 3D rj-McMC algorithm for all depth inversions in the second step).

The geological structure beneath the Irish Sea can be divided into a number of bedrock basins, representing depositional zones for the bedrock formations. The largest basins are Triassic in age and comprise the East Irish Sea basins (around 5°W - 3°W and 53.3°N - 55°N: Mellett et al., 2015). Thus we pose a question: *what is the area of the East Irish Sea basins at this depth?* We have a reference answer to this question, which is estimated from a shallow subsurface geological survey (Mellett et al., 2015), and which enables us to validate interrogation theory with real data.

It is known that sedimentary basins often have lower velocities compared to their surrounding regions, thus our question is equivalent to estimating the area of the continuous low velocity body at this depth. We therefore apply exactly the same procedure as we implemented in the synthetic examples above to find the optimal answer. We first define a mask, as marked by the black boxes in Figure 11, meaning that we only consider the seismic velocity information inside the mask. The North, East and South boundaries of the mask are determined by the coastline of mainland Britain, whereas the West boundary is defined based on the bedrock geology beneath the Irish Sea (Mellett et al.,

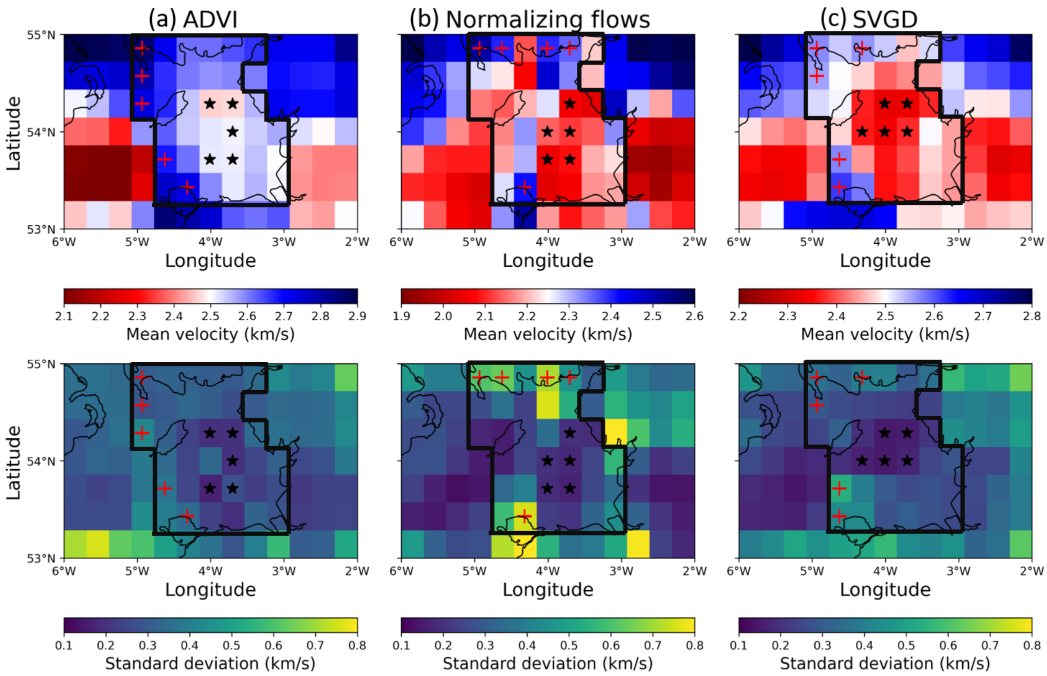

Figure 11. Mean (top row) and uncertainty (bottom row) maps of the top cell (from 0 to 200 m) of 3D shear wave velocity inversion results using dispersion curve data constructed from 2D tomography results obtained using (a) ADVI, (b) normalizing flows and (c) SVGD. In each figure, the black box displays the region where we calculate the area of the sedimentary basins. Black stars and red crosses are used to define the best threshold to discriminate low from high velocities with minimal bias.

2015). We select some points that are likely to belong to the East Irish Sea sedimentary basins (black stars in Figure 11), and another set of points that are highly likely to be outside the basins (red crosses in Figure 11). Given those grid cells, we calculate the best velocity threshold that discriminates low from high velocities with minimal bias using the same data-driven method as used in the synthetic test.

Similarly, we define our target function $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$ as the area of the largest continuous low velocity body inside the mask, and calculate the target function for each posterior sample from each algorithm. Figures 12a – 12c display the posterior distributions of the target function calculated using the inversion results from ADVI, normalizing flows and SVGD. In each figure, the mean value of the posterior target function (the optimal answer considering only each individual algorithm) is denoted by the dashed black line as well as the number below the legend, and the reference answer ($1.12\times10^4 km^2$ estimated from Mellett et al., 2015) is denoted by the red line in each figure.

Given the forward function $\boldsymbol{f}(\mathbf{m})$ used in the second inversion step, we define prior probabilities $p(C|\boldsymbol{f}(\mathbf{m}))$ for different algorithms. We assign $p(C|\boldsymbol{f}(\mathbf{m}))$ as 0.30 for ADVI and 0.35 for normalizing flows and SVGD (where these different algorithms were used for 2D surface wave tomography). The reason we only downweight ADVI slightly is that in this example, the role of these three methods is only to provide different datasets (mean and uncertainty values for dispersion curves) used in the second step depth inversion, in which we use the same algorithm: 3D rj-McMC. Previous studies (X. Zhang & Curtis, 2020a; Zhao et al., 2021) and the synthetic examples above have shown that, ADVI can provide an accurate mean model but a biased uncertainty result; that is the dispersion curves (the observed dataset for the second step) constructed by ADVI would have accurate mean values but inaccurate data uncertainty estimates. We treat these inaccurate data errors as additional unknowns and adjust their values adaptively and hierarchically by a scaling value during 3D rj-McMC inversion (Bodin et al., 2012; Galetti et al., 2017; X. Zhang et al., 2018), so the absolute data uncertainty level of the dispersion curves should have far less effect on inversion results. By using equation 3, we calculate the final optimal answer that considers the effect of different algorithms: $1.22\times 10^4\ km^2$, which provides reasonable accuracy compared to the reference value for this question derived from the geological study ($1.12\times10^4\ km^2$ – Mellett et al., 2015).

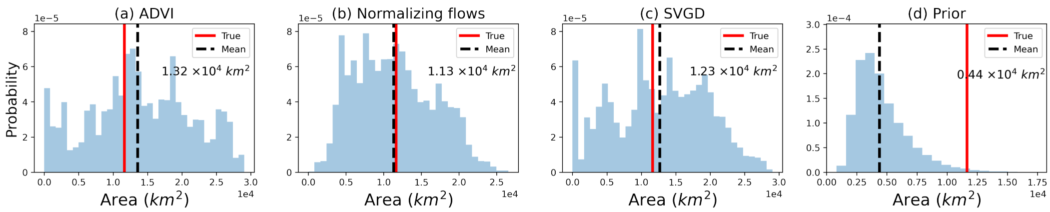

Figure 12. Posterior target functions for the area of the East Irish Sea basins at the shallow subsurface obtained from (a) ADVI, (b) normalizing flows, (c) SVGD and (d) prior distribution, respectively. In each figure, the red line denotes the reference answer to this question (1.12×10^4 km^2) estimated from surface geology (Mellett et al., 2015), and the dashed black line denotes the mean value of each histogram, which is also displayed by the number below the legend.

We note that in Figures 12a – 12c, the three posterior target functions span a very broad range (even the entire answer space from 0 to 3.0×10^4 km^2 that is close to the total area of the defined mask), and the optimal answer we obtained also appears to be close to the mean value of the upper and lower bounds of the answer space (1.5×10^4 km^2). In principle one might argue that this is because the surface wave data used in this example (from 4 s to 15 s period) are relatively insensitive to the near surface at a depth of up to 200 m; hence the posterior samples may not be well constrained by the data, leading to a broadly distributed set of target function values which happen to have the same mean as the true answer. To investigate, we apply the same interrogation procedure using the same velocity threshold as above, to 2 million samples drawn from the Uniform prior distribution, and display the histogram of the calculated target function in Figure 12d. Obviously, the posterior target distributions and the optimal answers obtained from the three inversion results in Figures 12a – 12c are significantly more informative than that estimated from the prior probability distribution which gives an extremely poor answer for the area of sediment. This shows that while it is true that the uncertainty on the final answer is high, the surface wave data are certainly far more informative than the answer that could be obtained from our prior information alone.

Since interrogation theory provides an optimal answer that is close to the answer obtained from an entirely different method based on interpreting surface geology, we have increased confidence in the result. This example as well as the synthetic tests therefore go some way towards validating interrogation theory as a practical method to answer scientific questions. In the next section, we apply the theory to answer a real-world scientific question where we do not know the true answer.

5.3 Estimating the volume of the East Irish Sea basins

We wish to answer a 3D volume-type question about the true Earth: *what is the total volume of the offshore East Irish Sea sedimentary basins?* In this example we need to define a 3D mask inside which we calculate the volume of the basins. As displayed by the black boxes in Figures 10a and 10b, we define such a 3D mask with fixed shape in the depth direction from the surface down to 8 km depth to fully encompass the offshore sediments while excluding most of the land. In the horizontal direction, the boundaries of the mask are defined based on the coastline of mainland Britain as well as on the inversion results in Figure 10.

The target function of this 3D example should account for the volume of the low velocity bodies inside the mask, since sedimentary basins often have relatively lower velocities compared to the surrounding regions. In contrast to 2D cases above where we used a fixed threshold to discriminate low from high velocities, we now need threshold values that vary with depth to allow for the significant velocity changes that occur between different depths due to pressure and temperature increases. We use the following method to obtain such depth-dependent threshold values. Firstly, we calculate 5 independent velocity threshold values at 5 fixed depths of 0 (surface), 2, 4, 6 and 8 km respectively, using exactly the same data-driven method as what we did in the 2D examples, and the obtained optimal threshold values are displayed by the red dots in Figure 13. We further interpolate between these 5 points to obtain the dashed red line in Figure 13. Each velocity value on this line is used as the optimal depth-dependent threshold that discriminates low from high velocities at the corresponding depth. The blue line in Figure 13 shows the average velocity value at different depths from the surface to 8 km. Although these two curves are not exactly the same (and there is no reason why they shoud be), they present similar feature of velocity increasing versus depth, which increases our confidence in the obtained depth-dependent threshold curve.

Given the obtained depth-dependent threshold curve, we classify every pixel inside the 3D mask as a low or high velocity grid cell, retain low velocity pixels and find the continuous low velocity bodies. In contrast to the 2D cases where we treat the largest

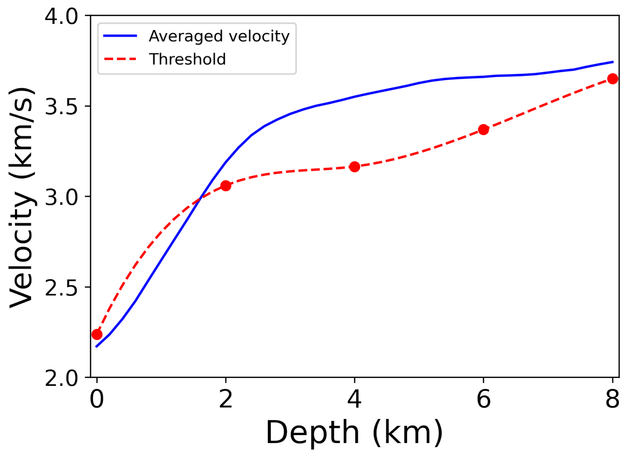

Figure 13. Mean velocity values at different depths from the surface to 8 km (blue line) and the optimal depth-dependent threshold curve to discriminate low from high velocity values with minimal bias (dashed red line).

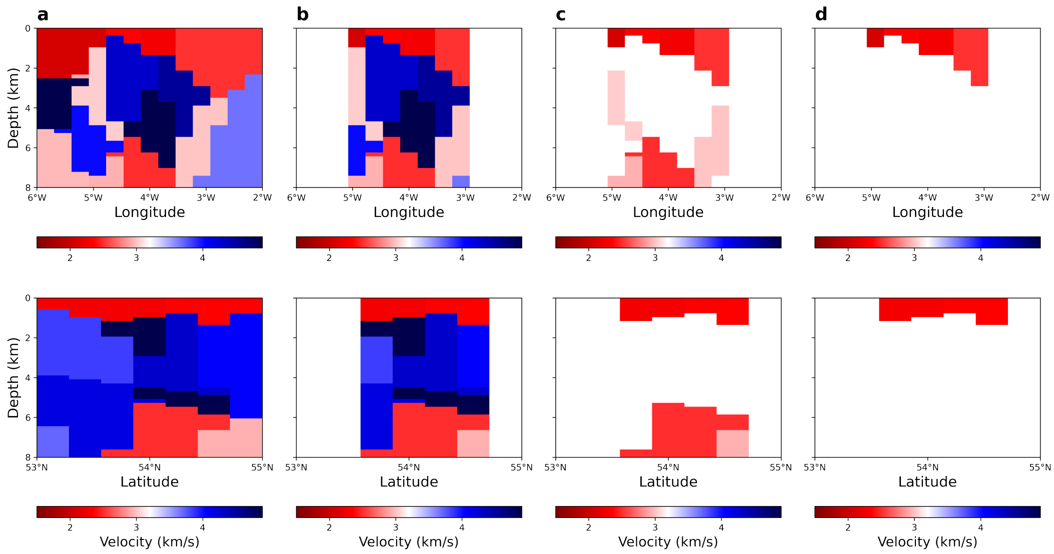

Figure 14. Vertical sections of a posterior sample drawn from the inversion results. The top row shows the vertical section at 53.67°N latitude, and the bottom row shows that at 4.33°W longitude. (a) Two vertical slices of this posterior sample. (b) The same vertical slices as in (a) after applying the 3D mask. (c) Two continuous low velocity bodies classified by the depth-dependent threshold curve. (d) The largest continuous low velocity body that starts from surface, whose volume is treated as the target function of this posterior sample.

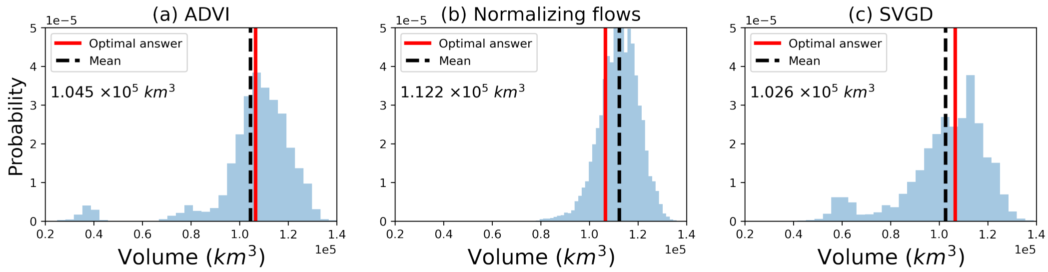

Figure 15. Posterior target functions for the volume of the East Irish Sea basins obtained from (a) ADVI, (b) normalizing flows and (c) SVGD, respectively. In each figure, the dashed black line denotes the posterior mean value of each algorithm, which is also displayed by the number below each legend, and the red line denotes the final estimated answer to this 3D question ($1.065 \times 10^5\ km^3$).

continuous low velocity body as the target function, we need to consider additional geological prior information when defining the target function for this 3D question. To illustrate, Figure 14 presents vertical slices of one posterior sample drawn from the 3D inversion results. The top row shows the depth slice at 53.67°N latitude, and the bottom row shows the vertical section at 4.33°W longitude. The two depth slices of this posterior sample are shown in Figure 14a, and the same slices after applying the 3D mask are displayed in Figure 14b. By comparing each velocity value with the depth-dependent threshold curve, we retain low velocity pixels and obtain two continuous low velocity bodies (Figure 14c). Given that we seek to estimate the volume of sedimentary basins, and considering that the basins in question have been observed to are often assumed to exist at least at the surface rather than only in deeper parts of the crust, we define the target function as the largest continuous low velocity body that starts from the surface. Therefore we interpret the upper low velocity body as the sedimentary basins of interest (shown in Figure 14d) and calculate its volume as the target function of this posterior sample, even though the lower one is larger.

We calculate this target function for each posterior sample obtained from ADVI, normalizing flows and SVGD, and display their posterior target histograms in Figure 15. The mean value of each posterior histogram represents the optimal answer estimated from each corresponding algorithm, which is denoted by the black dashed line and the number below the legend in each figure. We substitute those values and the 3 prior probabilities $p(C|\boldsymbol{f}(\mathbf{m}))$ into equation 3, and obtain the final estimated answer to our question: $1.065 \times 10^5\ km^3$ (the red lines in Figure 15).

6 Discussion

We used interrogation theory to answer real-world, unanswered scientific questions about the Earth based on Bayesian inversion results represented by posterior probability distributions. Previously, similar questions were usually answered by interpreting mean or maximum likelihood models directly. In the synthetic example, we have proved that direct interpretation of the mean model alone provides an inaccurate answer, especially under a coarse model parametrisation. The true Earth has infinitely fine structure, whereas we often use a relatively coarser parametrisation to reduce the dimensionality of the inversion problem. It is therefore likely that the answer obtained in this way is always biased at some level. On the other hand, the examples presented above show that the optimal answer obtained from interrogation theory is very close to the true (reference) answer, despite the relatively coarse model parametrisation (the grid size) employed.

The above result arises because the target function $T(\mathbf{m}|\boldsymbol{f}(\mathbf{m}), C, Q)$, which projects model parameter $\mathbf{m}$ into target space $\mathbb{T}$ where the question can be answered directly, is applied stochastically. In the synthetic example, consider a fixed pixel that spans the boundary of the true velocity anomaly. In some samples it is classified as part of the low velocity anomaly by the defined target function (suppose we label those pixels as 1), while in other samples it is not (we label them as 0). By applying equation 3, we account for the posterior mean of those labels, resulting in a fraction between $[0, 1]$, which denotes the probability that this pixel belongs to the low velocity anomaly. For comparison, if we only interpret the mean model (or any other single model) alone, this same pixel always either belongs or does not belong to the low velocity anomaly, so it always contributes either 1 or 0. As a result, the effective resolution of the answer obtained from interrogation theory can be much higher than might be apparent from the grid cell size alone, since we consider all of the posterior samples together in a statistical manner. Thus the answer is still accurate even when using a coarser parametrisation as observed in the supporting information associated with this paper.

Bayesian non-linear inversion is many times more expensive than linearised inversion, especially for high dimensional problems due to the curse of dimensionality (Curtis

& Lomax, 2001). Typically geophysicists only present, publish and use a small amount of the statistical information obtained from Bayesian inversion results, such as mean and point-wise standard deviations; most of the valuable information within the posterior pdf is discarded, which can introduce errors and biases when answering questions. This paper shows that interrogation theory provides a way to make use of all posterior samples obtained from Bayesian inversion, in a way that gives answers of improved accuracy. This goes some way to justifying the computational expense of solving inverse problems nonlinearly and probabilistically.

We considered the effect of different computational inversion algorithms C, and combined them to calculate optimal answers (equation 3). Thus the uncertainty caused by the use of any single algorithm was taken into account and the bias of the obtained answer was reduced. On the other hand, all of the above examples only used a single forward function $\boldsymbol{f}(\mathbf{m})$, the fast marching method (together with a modal approximation for the 3D example) to map model parameter $\mathbf{m}$ into the corresponding data $\mathbf{d}$. Future improvements in interrogation applications might focus on involving different forward models to answer area-type (or volume-type) questions, for example using full wave simulators as the forward model and using full waveform inversion to solve Bayesian inverse problems (Gebraad et al., 2020; X. Zhang & Curtis, 2020b, 2021), such that we can reduce the uncertainty caused by different model-data relationships.

Prior information is often critical in order to define a reasonable target function. In the synthetic example we defined the largest continuous low velocity body to be the low velocity anomaly of interest rather than simply including all of the low velocity bodies inside the mask. In the field data test, we interpreted sedimentary basins as low velocity bodies considering that basins often have relatively lower velocities compared to their surrounding rocks, and further interpreted the largest continuous low velocity body that starts from the surface to be the 3D basins of interest since these basins have been observed in the near surface geologically. The target function will always be more accurate if we consider more realistic prior information, and thus the optimal answer should be more reliable.

In the field data test, we used interrogation theory to find the best answer a^* conditioned on a particular traveltime dataset $\mathbf{d}$ and a fixed design E_d (receiver stations in Figure 7). In reality, it is common that the (predefined) design used to collect data is not the optimal one for the question posed because when networks are established it is always difficult to define a design that can best answer all questions that may be of interest in future. Interrogation theory provides a methodology to solve design problems to create an experiment that optimises information on one or more questions, but this does not remove the need to define the questions up front.

For real-world applications, it is possible that our ultimate question may not be addressed clearly within one interrogation procedure. To better answer the original question, a set of new questions are usually posed to provide more background knowledge, and a sequential interrogation process is required until a satisfactory answer is obtained. For more details about sequential interrogation, we suggest readers refer to Arnold and Curtis (2018).

Interrogation theory as presented in Arnold and Curtis (2018) appears to be highly structured and formalized. One purpose of this paper is to translate the theory into usable form, and to provide a concrete example of answering a specific type of question. One of the main theoretical advances of Arnold and Curtis (2018) was to introduce the target function in order to allow utilities to be defined in a simpler, more tractable form, even when a variety of parametrizations and forward functions are considered. A key revelation from our examples above is that much of the skill and work involved in answering real-world questions may be spent defining and calculating those target function values in a clear and unbiased manner. We hope to use interrogation theory to answer a wide range of real-world scientific questions in future.

7 Conclusion

In this paper, we use interrogation theory to answer a specific type of question about the Earth: to estimate the shape, area or volume of a subsurface structure by interrogating probabilistic Bayesian tomographic results. We establish an interrogation procedure by using a 2D synthetic example. By considering the effect of different computational algorithms, we reduce the bias of the optimal answer and obtain an accurate estimation of the question. The results using different parametrisations show that the same question can be answered accurately even on a relatively coarse grid, which reduces the computational cost of Bayesian inversion by orders of magnitude. We further apply interrogation theory to answer realistic questions about the East Irish Sea basins. The first application to estimate the horizontal area of the shallow part of the basins validates the theory, as the answer coincides to within 10% of that obtained from surface geological survey mapping. Finally, we use the method to estimate the total volume of the East Irish Sea basins for which no previously published answer exist. The theory established here is quite general, and can be applied to find answers for many other real-world scientific questions.

8 Open Research

Data associated with the field data example are available at British Geological Survey (`http://www.earthquakes.bgs.ac.uk/data/data_archive.html`).

Acknowledgments

We thank the Edinburgh Imaging Project (EIP) sponsors (BP and TotalEnergies) for supporting this research.

References

Arnold, R., & Curtis, A. (2018). Interrogation theory. *Geophysical Journal International*, *214*(3), 1830–1846.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* springer.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), 859–877.

Bloem, H., Curtis, A., & Maurer, H. (2020). Experimental design for fully nonlinear source location problems: which method should i choose? *Geophysical Journal International*, *223*(2), 944–958.

Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, *178*(3), 1411–1436.

Bodin, T., Sambridge, M., Rawlinson, N., & Arroucau, P. (2012). Transdimensional tomography with unknown data noise. *Geophysical Journal International*, *189*(3), 1536–1556.

Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 273–304.

Curtis, A., Gerstoft, P., Sato, H., Snieder, R., & Wapenaar, K. (2006). Seismic interferometry—turning noise into signal. *The Leading Edge*, *25*(9), 1082–1092.

Curtis, A., & Lomax, A. (2001). Prior information, sampling distributions, and the curse of dimensionality. *Geophysics*, *66*(2), 372–378.

Devilee, R., Curtis, A., & Roy-Chowdhury, K. (1999). An efficient, probabilistic neural network approach to solving inverse problems: inverting surface wave velocities for eurasian crustal thickness. *Journal of Geophysical Research: Solid Earth*, *104*(B12), 28841–28857.

Earp, S., & Curtis, A. (2020). Probabilistic neural network-based 2d travel-time tomography. *Neural Computing and Applications*, *32*(22), 17077–17095.

Fichtner, A., & Simutė, S. (2018). Hamiltonian monte carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, *123*(4), 2984–2999.

Fichtner, A., Zunino, A., & Gebraad, L. (2019). Hamiltonian monte carlo solution of tomographic inverse problems. *Geophysical Journal International*, *216*(2), 1344–1363.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D., & Nicolson, H. (2017). Transdimensional love-wave tomography of the british isles and shear-velocity structure of the east irish sea basin from ambient-noise interferometry. *Geophysical Journal International*, *208*(1), 36–58.

Galetti, E., Curtis, A., Meles, G. A., & Baptie, B. (2015). Uncertainty loops in travel-time tomography from nonlinear wave physics. *Physical review letters*, *114*(14), 148501.

Gebraad, L., Boehm, C., & Fichtner, A. (2020). Bayesian elastic full-waveform inversion using hamiltonian monte carlo. *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB018428.

Guest, T., & Curtis, A. (2009). Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships. *Journal of Geophysical Research: Solid Earth*, *114*(B4).

Jackson, D. D. (1972). Interpretation of inaccurate, insufficient and inconsistent data. *Geophysical Journal International*, *28*(2), 97–109.

Julian, B., Gubbins, D., et al. (1977). Three-dimensional seismic ray tracing. *Journal of Geophysics*, *43*(1), 95–113.

Käufl, P., P. Valentine, A., W. de Wit, R., & Trampert, J. (2016). Solving probabilistic inverse problems rapidly with prior samples. *Geophysical Journal International*, *205* (3), 1710–1728.

Käufl, P., Valentine, A. P., O'Toole, T. B., & Trampert, J. (2014). A framework for fast probabilistic centroid-moment-tensor determination—inversion of regional static displacement measurements. *Geophysical Journal International*, *196* (3), 1676–1693.

Khoshkholgh, S., Zunino, A., & Mosegaard, K. (2021). Informed proposal monte carlo. *Geophysical Journal International*, *226* (2), 1239–1248.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *The Journal of Machine Learning Research*, *18* (1), 430–474.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22* (1), 79–86.

Liu, Q., & Wang, D. (2016). Stein variational gradient descent: A general purpose bayesian inference algorithm. In *Advances in neural information processing systems* (pp. 2378–2386).

Maurer, H., Curtis, A., & Boerner, D. E. (2010). Recent advances in optimized geophysical survey design. *Geophysics*, *75* (5), 75A177–75A194.

Meier, U., Curtis, A., & Trampert, J. (2007). Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophysical Research Letters*, *34* (16).

Mellett, C., Long, D., & Carter, G. (2015). Geology of the seabed and shallow subsurface: The irish sea. *British Geological Survey Commissioned Report CR/15/057*, 1–42.

Mosegaard, K., & Tarantola, A. (1995). Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, *100* (B7), 12431–12447.

Nawaz, A., & Curtis, A. (2018). Variational bayesian inversion (vbi) of quasi-localized seismic attributes for the spatial distribution of geological facies. *Geophysical Journal International*, *214* (2), 845–875.

Nawaz, A., & Curtis, A. (2019). Rapid discriminative variational bayesian inversion of geophysical data for the spatial distribution of geological properties. *Journal of Geophysical Research: Solid Earth*, *124* (6), 5867–5887.

Nawaz, A., Curtis, A., Shahraeeni, M. S., & Gerea, C. (2020). Variational bayesian inversion of seismic attributes jointly for geological facies and petrophysical rock properties. *Geophysics*, *85*(4), 1–78.

Nicolson, H., Curtis, A., & Baptie, B. (2014). Rayleigh wave tomography of the british isles from ambient seismic noise. *Geophysical Journal International*, *198*(2), 637–655.

Nicolson, H., Curtis, A., Baptie, B., & Galetti, E. (2012). Seismic interferometry and ambient noise tomography in the british isles. *Proceedings of the Geologists' Association*, *123*(1), 74–86.

Rawlinson, N., & Sambridge, M. (2004). Wave front evolution in strongly heterogeneous layered media using the fast marching method. *Geophysical Journal International*, *156*(3), 631–647.

Rezende, D. J., & Mohamed, S. (2015). Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*.

Saito, M. (1988). Disper80: A subroutine package for the calculation of seismic normal-mode solutions. *Seismological algorithms*, 293–319.

Shapiro, N. M., Campillo, M., Stehly, L., & Ritzwoller, M. H. (2005). High-resolution surface-wave tomography from ambient seismic noise. *Science*, *307*(5715), 1615–1618.

Siahkoohi, A., Orozco, R., Rizzuti, G., Witte, P. A., Louboutin, M., & Herrmann, F. J. (2021, 09). Fast and reliability-aware seismic imaging with conditional normalizing flows. In *Intelligent illumination of the earth.* ((KAUST, virtual))

Siahkoohi, A., Rizzuti, G., & Herrmann, F. J. (2021). Deep bayesian inference for seismic imaging with tasks. *arXiv preprint arXiv:2110.04825*.

Siahkoohi, A., Rizzuti, G., Louboutin, M., Witte, P. A., & Herrmann, F. J. (2021). Preconditioned training of normalizing flows for variational inference in inverse problems. *arXiv preprint arXiv:2101.03709*.

Singh, J., Tant, K. M., Curtis, A., & Mulholland, A. J. (2021). Real-time super-resolution mapping of locally anisotropic grain orientations for ultrasonic non-destructive evaluation of crystalline material. *arXiv preprint arXiv:2105.09466*.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation* (Vol. 89). siam.

van Den Berg, J., Curtis, A., & Trampert, J. (2003). Optimal nonlinear bayesian experimental design: an application to amplitude versus offset experiments. *Geophysical Journal International*, *155*(2), 411–421.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67–82.

Zhang, C., Bütepage, J., Kjellström, H., & Mandt, S. (2018). Advances in variational inference. *IEEE transactions on pattern analysis and machine intelligence*, *41*(8), 2008–2026.

Zhang, X., & Curtis, A. (2020a). Seismic tomography using variational inference methods. *Journal of Geophysical Research: Solid Earth*, *125*(4), e2019JB018589.

Zhang, X., & Curtis, A. (2020b). Variational full-waveform inversion. *Geophysical Journal International*, *222*(1), 406-411.

Zhang, X., & Curtis, A. (2021). Interrogating probabilistic inversion results for subsurface structural information. *Geophysical Journal International, in press*.

Zhang, X., Curtis, A., Galetti, E., & De Ridder, S. (2018). 3-d monte carlo surface wave tomography. *Geophysical Journal International*, *215*(3), 1644–1658.

Zhang, X., Nawaz, A., Zhao, X., & Curtis, A. (2021). An introduction to variational inference in geophysical inverse problems. *Advances in Geophysics*, *62*, 73-140.

Zhao, X., Curtis, A., & Zhang, X. (2021). Bayesian seismic tomography using normalizing flows. *Geophysical Journal International*, *228*(1), 213–239.

Supporting Information for "Interrogating Subsurface Structures using Probabilistic Tomography: an example assessing the volume of Irish Sea basins"

Xuebin Zhao[1], Andrew Curtis[1]and Xin Zhang[1]

[1]School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom

Interrogation with different parametrisations

In this supporting file, we investigate the effect of interrogations carried out using models with different parametrisations. The true Earth has infinitely fine structure, whereas in reality we parametrise it with a finite (coarse) grid or number of parameters to reduce the dimensionality of our inverse problem, so it is crucial to consider the effect of different parametrisations when answering questions. The previous inversion results in Figure 4 in the main text used a relatively fine regular grid of 0.5 km in both directions to discretize the true velocity model, which leads to an inference problem with 441 unknown parameters. Now we double the grid size in both directions to 1 km, which decreases the dimensionality of the tomographic problem to 121. Obviously the computational cost would be reduced significantly. Using this parametrisation we perform tomography

---

Corresponding author: Xuebin Zhao, School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom. (xuebin.zhao@ed.ac.uk)

again. All of the other inversion settings are the same as those used in the finer grid test and in Zhao, Curtis, and Zhang (2021), and the corresponding inversion results are shown in Figure S1. Figures S1a – S1d show the mean (top row) and uncertainty (bottom row) maps of the posterior distributions from automatic differential variational inference (ADVI) (Kucukelbir et al., 2017), normalizing flows (Rezende & Mohamed, 2015), Stein variational gradient descent (SVGD) (Liu & Wang, 2016) and Metropolis Hastings McMC (MH-McMC) (Tarantola, 2005), respectively. From the four mean velocity maps, we still observe a low velocity anomaly at the centre of the model. However, if we answer the question Q by interpreting the 4 mean velocity maps directly, we would obtain an estimated answer of 9 km^2 from all of the four inversion results, which is strongly biased compared to the true answer (12.56 km^2).

Now using interrogation theory to find the optimal answer, we define a similar mask as for the finer grid test in the main text, shown in Figure S1e. We select some high and low velocity points from the inversion results (black stars and red crosses in Figure S1), and calculate the optimal velocity threshold for this parametrisation: 1.644 km/s. Then we calculate the area of the largest continuous low velocity body inside the mask as the target function value. The corresponding posterior target functions for the four algorithms are displayed in Figure S2. The dashed black line and the value at the top-left corner in each figure mark the mean value of each posterior target histogram, which is interpreted as the optimal answer obtained from each specific algorithm. The red line stands for the true answer to this question. Both the posterior histograms and their corresponding mean values in Figure S2 are quite similar to those obtained from the finer grid parametrisation in Figure 6 in the main text. The final answer of the coarser grid parametrisation (12.37

km^2) is very close to the true answer (12.56 km^2), as well as that estimated from the finer grid parametrisation (12.89 km^2).

We thus obtain an accurate answer using interrogation theory using either parametrisation. By contrast interpreting the mean map alone provides a severely erroneous answer (9 km^2). This makes interrogation theory more attractive for answering scientific questions since we obtain an accurate answer to the question even under a coarse parametrisation, which usually offers orders of magnitudes of computational cost reduction in real problems.

References

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *The Journal of Machine Learning Research*, *18*(1), 430–474.

Liu, Q., & Wang, D. (2016). Stein variational gradient descent: A general purpose bayesian inference algorithm. In *Advances in neural information processing systems* (pp. 2378–2386).

Rezende, D. J., & Mohamed, S. (2015). Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation* (Vol. 89). siam.

Zhao, X., Curtis, A., & Zhang, X. (2021). Bayesian seismic tomography using normalizing flows. *Geophysical Journal International*, *228*(1), 213–239.

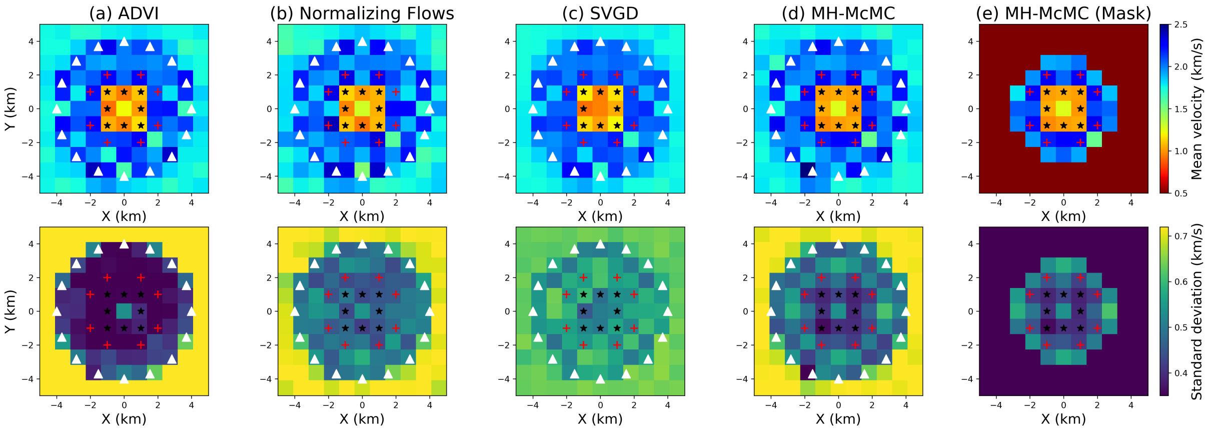

Figure S1. (a) - (d) Pixel-by-Pixel mean (top row) and standard deviation (bottom row) maps of the posterior distributions obtained using ADVI, normalizing flows, SVGD and MH-McMC with a coarser grid parametrisation of 1 km in both directions. (e) The posterior mean and uncertainty maps of MH-McMC after applying the mask introduced in the main text. Red crosses and black stars in each figure denote the selected pixels used to define the threshold value to discriminate low and high velocities.

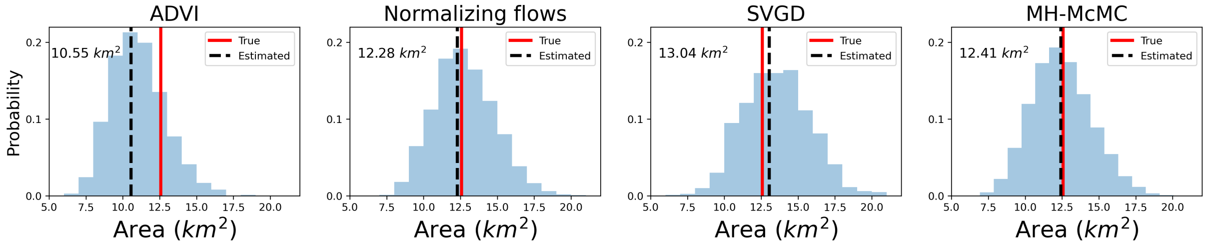

Figure S2. Posterior distributions of the target function for ADVI, normalizing flows, SVGD and MH-McMC with the coarser grid parametrisation in Figure S1. The posterior mean of each target function is displayed at the top-left corner, and is also marked by the dashed black line in each figure. The true answer to this question (12.56 km^2) is denoted by the red line.