Bayesian Inference in Seismic Tomography

Xin Zhang

THE UNIVERSITY
of EDINBURGH

Thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy
to the
University of Edinburgh — 2019

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Xin Zhang
October 2019

Abstract

In a variety of scientific applications we require methods to construct three dimensional maps of properties of the interior of solid media, and in the geosciences the medium is usually the Earth's subsurface. For each such map we need the corresponding map of uncertainties in those properties in order to assess their reliability.

Seismic tomography is such a method which has been used widely to study properties of the subsurface of the Earth, for example, using surface wave dispersion data. Surface wave tomography is usually conducted using a two-step method by first estimating two-dimensional (2D) surface wave phase or group velocity maps at a series of frequencies and then inverting those for the 3D spatial velocity structure through a set of 1D inversions for structure with depth beneath each geographical location. Since surface wave tomography is a highly non-linear problem, it is usually solved using Monte Carlo (MC) sampling methods. However, since the 1D inversions in the second step are usually performed independently, lateral spatial correlations of the Earth can be lost. We therefore introduce a one-step MC method which inverts for a 3D velocity structure directly from frequency-dependent surface wave travel time measurements by using a fully 3D parametrization. The method was first applied to a synthetic test and compared with two-step linearised and two-step MC methods. The results show

that by including lateral spatial correlations in the inversion the new method estimates velocity models and associated uncertainty significantly better in the sense that it produces more intuitively reasonable and interpretable results, and the computation cost is also comparable to the two-step MC method.

We apply the 3D MC surface wave tomography method to a real dataset recorded using a dense passive seismic array installed on the North Sea seabed. The ambient noise data of each receiver pair are cross correlated to extract Scholte waves, in which two Scholte wave modes are observed. We separated the two modes using a dispersion compensation method. For each separated mode phase velocity maps are determined using Eikonal tomography. Those phase velocity maps are then used to estimate 3D shear velocities of the subsurface. To further understand the limitation of the approach, we conducted three different inversions: the usual 1D depth inversions, a 2D inversion along a 2D cross section and a fully 3D inversion. With each inversion the shear velocity structure is extracted along the same cross section and compared. The results confirm that 1D inversions can produce errors due to independence of those inversions, whereas 2D and 3D methods improve the results by including lateral spatial correlations in the inversion. The 3D results better match an existing shear velocity model obtained from active source seismic reflection tomography. This is probably because the 3D method uses frequency-dependent measurements directly, which naturally avoids errors introduced in the first 2D Eikonal tomography step. The results show a clear low velocity river channel, and exhibit another low velocity anomaly both in the phase velocity maps at short periods (< 1.6 s) of the fundamental mode and in the shear-velocity model in the near surface (< 250 m). The latter anomaly is correlated with the distribution of seabed pockmarks, indicating that the anomaly might be related to the circulation of near surface fluids.

Apart from surface waves, seismological body wave travel times have also been

used to study the Earth's interior and to characterize earthquakes. Body waves are generally sensitive to structure around the sub-volume in which earthquakes occur and produce limited sensitivity in the near surface, whereas surface waves are more sensitive to the shallower structure. Thus body waves and surface waves can be used jointly to better constrain the subsurface structure. Since the tomographic problem is usually highly non-linear, we apply MC sampling methods to invert for source parameters and velocity models simultaneously using earthquake body wave travel times and ambient noise surface wave dispersion data. The method is applied to a mining site in the U.K. where induced seismicity is recorded using a small local network and ambient noise data are available from the same stations. The results show that by using both types of data, earthquake source parameters and velocity models can be better constrained than in independent inversions. Synthetic tests show that the independent inversion using only body wave travel times can cause biases in the results due to trade-offs between source parameters and velocity models, while this issue can be largely resolved using joint inversion, indicating that the ambient noise data can provide additional information.

Although MC sampling methods have been used widely to solve seismic tomographic problems, they are computationally expensive and remain intractable for large dataset problems. We therefore introduce variational inference methods to solve seismic tomographic problems. Variational inference solves the Bayesian inference problem using optimization, yet still provide probabilistic results. In this thesis we introduce two variational methods: automatic differential variational inference (ADVI) and Stein variational gradient descent (SVGD), and apply them to 2D seismic tomographic problems using both synthetic and real data. We compare the results with those obtained using two different MC sampling methods, and demonstrate that variational inference methods can provide accurate approximations to the results of MC sampling methods at significantly

lower computational cost, provided that the gradient of model parameters with respect to data can be computed efficiently.

Lay Summary

Geoscientists often need to build models of the Earth's subsurface in order to understand the properties and processes in the Earth's interior. Seismic tomography is a method which builds subsurface seismic velocity maps using observed seismic data (e.g. travel times of seismic waves). Three-dimensional (3D) velocity maps (called *models*) built from seismic data are often non-unique since many models fit the measured data. Uncertainties of these models therefore need to be quantified.

It is common to use surface waves to infer velocity models. Surface waves are waves that propagate along the surface of the Earth while oscillating over depth ranges that depend on the period of oscillation. This makes surface wave dispersive: waves of different period travel at different speeds. The dispersion property can therefore be measured and interpreted to study the structure of subsurface. This is usually achieved by using Markov chain Monte Carlo (McMC) inference methods. These generate an ensemble of models that are consistent with the data, which in turn can be used to quantify uncertainties. However, in conventional McMC surface wave tomography, velocity at each location is assumed to be independent of its laterally-neighbouring velocity, which is in general not true in the real Earth. This in turn can produce errors in the results, which are usually smoothed out of final subsurface images in an ad hoc manner.

I introduce a fully 3D McMC inference method which includes lateral spatial correlations in the results by using a 3D model parametrization. The method is first applied to a synthetic example and compared with conventional methods. The results show that the new method can produce more accurate velocities and more intuitively reasonable uncertainty estimates. I then apply the method to a real dataset recorded on a dense passive seismic array installed on the North Sea seabed and compared the results with those obtained using conventional methods. The results confirm that the new 3D method can produce more accurate estimates of velocity and uncertainty.

Apart from surface waves, seismological body wave travel times can also be used to study the subsurface of the Earth and to characterize earthquakes. Body waves generated by earthquakes travel through the Earth's interior. Measurements of arriving body waves therefore provides information about the sub-volume in which earthquakes occur and through which those waves propagate. In comparison to body waves, surface waves are more sensitive to shallow structure. Thus body waves and surface waves can be used jointly to produce a more accurate velocity model.

This joint inversion problem is traditionally solved using linearised methods, which can cause biases in the results and cannot provide accurate uncertainty estimate. I therefore propose a novel inference method which uses 3D McMC methods to solve the tomographic problem using body wave arrival times and surface wave dispersion data jointly. I apply the method to data from a mining site in the U.K., and demonstrate that by using both types of data, both earthquake sources and velocity models can be better constrained than when estimated using independent body and surface wave inversions.

Although McMC methods have been used widely to solve seismic tomographic problems they remain computationally intractable for problems with large

datasets and high dimensionality. To extend uncertainty analysis to larger systems, I introduced two variational inference methods to solve seismic tomographic problems: automatic differential variational inference (ADVI) and Stein variational gradient descent (SVGD). I apply the methods to solve two tomographic problems, one with synthetic and one with real data, and compared the results with those obtained using McMC methods. The results show that variational inference methods can produce accurate approximations to solutions obtained using McMC methods, but at significantly reduced cost. I expect that the methods can be applied to many other type of geophysical inverse problems.

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisor, Prof. Andrew Curtis, for his continuous support through the course, for guidance on the writing of this thesis and for the freedom that he gave me on my research. Thanks to my second supervisors, Dr. Erica Galetti and Dr. Sjoerd de Ridder, for their constructive discussions and suggestions. Thanks to my advisor Dr. Christopher McDermott for his kind support and help. Thanks to my colleagues for their company, Stephanie, Dominic, Angus, Atif, Carlos, Claire, Eva, Hugo and Louise. It has been a great experience to work within such a kind and bright group of individuals.

I would like to thank Dr. Fredrik Hansteen for advising me during my internship in Equinor and for support on the following research after the internship. Thanks to Dr. Matteo Ravasi, Dr. Alexander Kritski, Dr. Long Wu and other colleagues in Equinor for their fruitful discussions and suggestions.

I would like to thank my flatmates, Shibo Shao and Panxia Zhou, for their company and for their amazing cooking. Thanks to my friends, Hao Yu, Yikuo Liu, Yong Zhao, Youbing Zhang, Zhaoliang Hou, Zhaoyu Jin. I reserve special thanks to Ying Liu for her company and help in my life, to Jiadong Xu for his advice and mentorship, to Xing Wu for her company and for the fun she brings to my life.

Thanks to my parents and my sister for their support and encouragement through the study of my Ph.D.

Finally, I would like to thank Edinburgh interferometry sponsors, Equinor, Schlumberger, and Total, for providing me financial support to perform this study. I would like to thank staff members at EPCC for their kind help, and for generously providing me sufficient computational resources.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Bayesian inference

To understand the process and properties of a physical system scientists often define a simplified model to represent the system and a physical relation which predicts measurements or observations from the model. This process is often called the *simulation problem* or the *forward problem.* In practice we often try to characterize models using measurements or observations which is referred as an *inverse or inference problem.* Models are generally parameterized in some way, therefore the inverse problem is also a parameter estimation problem (Tarantola, 2005; Aster et al., 2018).

While the forward problem generally has a unique solution, the inverse problem is often under-determined. Therefore some *prior* information (information that is known independent of data) is often applied to the parameters of the model, and uncertainties of the estimated parameters need to be quantified (Tarantola, 2005).

The inverse problem can be solved in an optimization by seeking an optimal solution that minimizes the misfits between model predicted data and observed data. Since the problem is usually under-determined, some properties are enforced on the model (e.g. smoothness) which is often called *regularization*. This procedure is well established for linear problems and can also be applied to non-linear problems by iteratively approximating the non-linear physics using a linearised relation around the current model and solving the linear problem to update that parameter estimate (Tarantola & Valette, 1982). While the method can produce reasonable uncertainty estimate for linear problems, it is difficult to quantify uncertainties for non-linear problems. Though the regularization can be chosen using some standard methods (e.g., L-curve – Hansen, 1992), valuable information can be concealed by the regularization (Zhdanov, 2002).

Bayesian inference provides a different way to solve the inverse problem by representing the prior information using a probability density function (pdf) and updating the prior probability distribution with new information contained in the data to obtain a probability distribution of parameters post inversion, called a *posterior* probability density function (Box & Tiao, 2011). According to Bayes' theorem the posterior pdf $p(\boldsymbol{\theta}|\mathbf{d}_{obs})$ can be calculated as:

$$p(\boldsymbol{\theta}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{d}_{obs})} \tag{1.1}$$

where $\boldsymbol{\theta}$ is the parameter vector and $p(\boldsymbol{\theta})$ is the prior probability distribution of the parameters; $p(\mathbf{d}_{obs}|\boldsymbol{\theta})$ is called the *likelihood* which is the probability of observing the measured data given a certain parameter vector, and $p(\mathbf{d}_{obs})$ is a normalization factor called the *evidence*.

While the linearised inversion produces one optimal solution, Bayesian inference produces a probability distribution over the parameters. The linearised inversion can be treated as maximizing the likelihood since likelihood is often defined with misfits, and the regularization can be treated as prior information. The linearised inversion can also be implemented to maximize the posterior pdf which is often referred as a maximum a posterior (MAP) solution. In this way one can include broader prior information in the inversion. More connections between linearised inversion and Bayesian inference can be found in Tarantola (2005) and Theodoridis (2015).

The Bayesian inference problem can be solved in at least two different ways: Markov chain Monte Carlo (McMC) and variational inference. A *Markov chain* is a sequence of random variables, $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$... whose conditional distributions obey the rule

$$p(\mathbf{x}_n|\mathbf{x}_{n-1}, \{\mathbf{x}_t : t \in \mathcal{I}\}) = p(\mathbf{x}_n|\mathbf{x}_{n-1}) \tag{1.2}$$

where $\mathcal{I} = \{0, 1, ..., n-2\}$. A Markov chain is called an ergodic Markov chain if it is possible to go from every state to every other state with positive probability (Brooks et al., 2011). In McMC, we construct an ergodic Markov chain whose equilibrium distribution is the target probability distribution (in this case the posterior pdf) and collect samples from this chain. Since the samples are distributed according to the target probability distribution, they can be used to calculate any statistics, e.g., mean and standard deviation (Gilks et al., 1995; Brooks et al., 2011).

The common way to construct an ergodic Markov chain with desired probability distribution is to construct Markov chains that satisfy a detailed balance condition, which ensures that the stationary distribution is the desired distribution. The Metropolis-Hastings algorithm (Metropolis & Ulam, 1949; Hastings, 1970) is one of such methods which has been used widely across various scientific fields (Robert & Casella, 2013). The method has been extended to trans-dimensional inversion, called reversible jump Markov chain Monte Carlo (rj-McMC) which means that the number of parameters (the dimensionality of parameter space) can vary in the inversion and consequently the parameterization itself can be adapted to the data and to any prior information (Green, 1995; Green & Hastie, 2009). In this thesis we will focus mainly on using this method to perform seismic inversions (see details in Chapter 2, 3 and 4).

While McMC methods have been used widely to solve inverse problems, the method cannot be easily applied to problems that have large data set or high-dimensional parameter spaces since the problems become computationally intractable. The methods also appear to be inefficient for complex multimodal probability distributions. In these cases variational inference provides an efficient alternative to McMC methods.

Variational inference solves the Bayesian inference problem using optimization techniques by seeking an optimal approximating probability distribution to the target distribution within a predefined family of probability distributions (Bishop, 2006; Blei et al., 2017). This is achieved by minimizing the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951), which defines the difference between the approximating distribution and the target distribution. The complexity of the approximating family defines the complexity of the optimisation problem and also the ability to approximate arbitrary target probability distributions. Generally the approximating family should be rich enough to approximate many probability distributions and meanwhile the optimizing problem should still be reasonably easy to solve.

Traditionally variational inference is conducted using simple *mean-field* families which mean the parameters are assumed to be mutually independent (Bishop, 2006; Blei et al., 2017) or simply structured families in which dependencies between variables are added (Saul & Jordan, 1996; Hoffman & Blei, 2015). The optimization problem with these families can be solved using a coordinate ascent algorithm (Biondi, 1992; Blei et al., 2017) or the Expectation-maximization algorithm (Nawaz & Curtis, 2018, 2019).

However, these algorithms require tedious derivation and bespoke implementation for each type of problem, and require the analytic solution for the expectation with respect to approximating probability distributions (Bishop, 2006; Blei et al., 2017). This restricts the methods to a relatively narrow range of inverse problems for which those derivations can be performed. To make variational inference easier to use and applicable to a wider class of problems, "black box" variational inference methods have been proposed recently (Kingma & Welling, 2013; Ranganath et al., 2014; Rezende & Mohamed, 2015; Tran et al., 2015; Liu & Wang, 2016; Marzouk et al., 2016; Ranganath et al., 2016; Kucukelbir et al., 2017). In this thesis we will apply two of these methods – automatic differential variational inference (Kucukelbir et al., 2017) and Stein variational gradient descent (Liu & Wang, 2016) – to seismic tomographic problems (see details in Chapter 5).

1.2 Bayesian inference in seismic tomography

Seismic tomography is one type of inverse problem which uses seismic data (e.g., travel times) to study the subsurface of the Earth (Aki & Lee, 1976). The properties (e.g., velocity, density etc.) of the subsurface are first parameterized in some way, and those parameters are then estimated from the observed data. In order to better understand and interpret the results, the uncertainties of these parameters are often required.

Seismic tomographic problems are traditionally solved using linearised methods which has been used widely to study the subsurface structure of the Earth using body wave travel time data (Aki & Lee, 1976; Romanowicz, 1979; Thurber, 1983; Dziewonski & Woodhouse, 1987; Iyer & Hirahara, 1993; Van der Hilst et al., 1997; Curtis & Snieder, 2002; Zhang & Thurber, 2003; Tarantola, 2005) and surface wave dispersion data (Woodhouse & Dahlen, 1978; Nakanishi & Anderson, 1983; Trampert & Woodhouse, 1995; Curtis et al., 1998; Ritzwoller et al., 2002; Simons et al., 2002; Ekström, 2011) that were generated by earthquakes. Since earthquakes are generally distributed inhomogeneously, the resolvability of models in regions of sparse coverage is inevitably limited. The introduction of ambient noise interferometry has greatly increased the size and coverage of surface wave dataset by turning receivers into virtual sources. It has been shown that Green's functions between different receiver pairs can be approximately estimated from cross correlations of ambient noise data recorded at the receivers (Campillo & Paul, 2003; Wapenaar, 2004; van Manen et al., 2005, 2006; Wapenaar & Fokkema, 2006; Curtis et al., 2006). The technique has been used to perform seismic tomography using surface wave dispersion data extracted from cross correlations, which is often referred as *ambient noise* tomography (Shapiro & Campillo, 2004; Shapiro et al., 2005; Yao et al., 2006; Lin et al., 2007; Yang et al., 2007; Lin et al., 2009; Bensen et al., 2009; Behr et al., 2010; de Ridder & Dellinger, 2011; Nicolson et al., 2012, 2014; de Ridder & Biondi, 2013; Mordret et al., 2013a,b; de Ridder et al., 2014; Allmark et al., 2018).

The above studies are generally conducted using linearised methods, which can bias the results and cannot produce accurate uncertainty estimates. Therefore, Monte Carlo methods have been introduced to geophysics to solve geophysical inverse problems (Mosegaard & Tarantola, 1995; Sambridge et al., 1995; Malinverno et al., 2000; Malinverno & Briggs, 2004), and have been extended to trans-dimensional inversions using the rj-McMC method (Malinverno, 2002; Bodin & Sambridge, 2009; Bodin et al., 2012; Galetti et al., 2015, 2017; Galetti & Curtis, 2018). Although many applications have been conducted using McMC methods

(Shen et al., 2012, 2013; Young et al., 2013; Zulfakriza et al., 2014; Zheng et al., 2017; Crowder et al., 2019), they mainly address 1D or 2D tomography problems due to the high computational cost of Monte Carlo methods. Some studies used McMC methods to solve 3D tomographic problems using body wave travel time data (Hawkins & Sambridge, 2015; Piana Agostinetti et al., 2015; Burdick & Lekić, 2017).

However, the fully 3D scheme has not been used in any surface wave tomographic problems and its properties have not been studied. In this thesis, we will apply the fully 3D Monte Carlo scheme to surface wave tomographic problems (see details in Chapter 2 and 3) and extend the method to a joint inversion using both body wave travel time data and surface wave dispersion data (see details in Chapter 4).

Monte Carlo methods are generally computational expensive and cannot be easily applied to large datasets. Therefore, we introduced two variational inference methods: automatic differential variational inference (ADVI – Kucukelbir et al., 2017) and Stein variational gradient descent (SVGD – Liu & Wang, 2016), to solve seismic tomographic problems (see details in Chapter 5). We apply the two methods to solve 2D synthetic and real-data seismic tomographic problems and compare the results to those obtained by Monte Carlo sampling methods to demonstrate that variational inference methods can provide efficient alternatives to Monte Carlo sampling methods.

1.3 Structure of the thesis

Chapter 2 introduces a fully 3D Monte Carlo surface wave tomography method which uses rj-McMC algorithm with a fully 3D model parametrization. In the chapter I first describe the implementation of the method and then compare the method with the standard linearised method and the conventional 2-step McMC method using a synthetic test. The results show that the 3D method estimates the velocity model and uncertainty significantly better than the other two methods, and also does so at comparative cost to the 2-step McMC method.

Chapter 3 describes an application of the 3D method introduced in Chapter 2 to an ambient noise dataset recorded on a dense passive seismic array installed on the North Sea seabed. The chapter first describes cross correlations of the ambient noise data and estimates phase velocity maps using those cross correlations. I then apply the 2-step McMC method and the new 3D method to the dataset to estimate 3D shear velocity models of the subsurface. A 2D McMC method is also applied to a 2D cross section to better understand the limitation of the 1D method. The comparison of the results obtained using the three methods confirms that the 3D method produces the most accurate results and achieves this with a cost that is comparable to the standard 2-step McMC method.

Chapter 4 extends the 3D Monte Carlo method to a joint inversion using earthquake body wave travel times and ambient noise surface wave dispersion data. The method is applied to a mining site in the U.K. where induced seismicity is recorded on a temporary seismic network, and where ambient noise recordings are also available on the same station. The results show that by using both types of data source parameters and velocity models can be better constrained than in independent inversions. Trade-offs between source parameters and velocity models can also be significantly reduced using the joint inversion method.

Chapter 5 introduces two variational inference methods: automatic differential variational inference (ADVI) and Stein variational gradient descent (SVGD), to solve seismic tomographic problems. The two methods are first applied to a 2D synthetic test and compared with two Monte Carlo methods: Metropolis-Hastings

McMC and rj-McMC methods. The results show that variational inference methods can provide accurate approximations to those obtained using Monte Carlo methods at significantly lower computational cost. Finally the two methods are applied to real data described in Chapter 2 to estimate a phase velocity map at 0.9 s.

Chapter 6 first summarizes the work accomplished in this thesis and its contributions to the field. The chapter then discusses limitations of the methods used in this thesis, namely restrictions of the Voronoi parametrization, efficiency problem of the rj-McMC method and limitations of variational inference methods. It also proposes future work to mitigate these issues.

Appendix A describes the derivation of the gradients used in variational inference methods in Chapter 5. The first two parts calculate the gradients in ADVI and the last part calculates the gradients in SVGD. Both sets of gradients are used to minimize the KL-divergence to obtain an optimal approximation to the true posterior probability distribution.

1.4 Publications

The following papers were published in the pursuit of this research, each of which is closely related to a chapter in this thesis.

Chapter 2 published as: **Zhang, X.**, Curtis, A., Galetti, E., & de Ridder, S., 2018. 3-D Monte Carlo surface wave tomography, *Geophysical Journal International*, **215**(3), 1644–1658.

Chapter 3 published as: **Zhang, X.**, Hansteen, F., Curtis, A., & de Ridder, S., 2019. 1D, 2D and 3D Monte Carlo ambient noise tomography using a dense passive seismic array installed on the North Sea seabed, *Journal of Geophysical Research: Solid Earth*, accepted.

Chapter 4 published as: **Zhang, X.**, Roy, C., Curtis, A., Nowacki, A., & Baptie, B., 2019. 3D Tomographic Monte Carlo joint inversion of earthquake body wave travel times and ambient noise surface wave dispersion data, *submitted.*

Chapter 5 published as: **Zhang, X.** & Curtis, A., 2019. Seismic tomography using variational inference methods, *submitted to Journal of Geophysical Research: Solid Earth*, Under Review, Minor Revisions.

And the following co-authored papers were also published:

Roy, C., **Zhang, X.**, Curtis, A., Nowacki, A., & Baptie, B., 2019.Probabilistic local earthquake magnitudes and their uncertainties – implications for traffic light systems for induced seismicity. *submitted.*

Earp, S., Curtis, A., **Zhang, X.**, & Hansteen, F., 2019. Probabilistic Neural Network Tomography across Grane field (North Sea) from Surface Wave Dispersion Data. *submitted to Geophysical Journal International*, Under Review.

The following extended abstracts were also published in the course of this research:

Zhang, X., Hansteen, F., & Curtis, A., 2019. Fully 3D Monte Carlo Ambient Noise Tomography over Grane Field. *In 81st EAGE Conference and Exhibition 2019.*

Curtis, A., Cao, R., Earp, S., **Zhang, X.**, De Ridder, S., & Galetti, E., 2019. Near-Real Time 3D Seismic Velocity and Uncertainty Models from

Ambient Noise, Gradiometry and Neural Network Inversion. *In 81st EAGE Conference and Exhibition 2019 Workshop Programme.*

Chapter 2

3D Monte Carlo seismic surface wave tomography

2.1 Summary

Seismic surface wave tomography is a tried and tested method to reveal the subsurface structure of the Earth. However, the conventional 2-step scheme of inverting first for two-dimensional (2D) maps of surface wave phase or group velocity and then inverting for the 3D spatial velocity structure preserves little information about lateral spatial correlations, and introduces additional uncertainties and errors into the 3D result. We introduce a 1-step 3D non-linear surface wave tomography method that removes these effects by inverting for 3D spatial structure directly from frequency-dependent travel-time measurements. We achieve this using the reversible jump Markov chain Monte Carlo (McMC) algorithm with a fully 3D model parameterization. Synthetic tests show that the method estimates the velocity model and associated uncertainties significantly

better than the conventional 2-step McMC method, and that the computational cost seems to be comparable with 2-step McMC methods. The resulting uncertainties are more intuitively reasonable than those from the 2-step method, and provide directly interpretable uncertainty on volumetrics of structures of interest.

2.2 Introduction

Seismic surface waves travel along the surface of the Earth while oscillating over depth ranges that depend on the period of oscillation. Measured speeds of travel are sensitive to Earth properties within those depth ranges. Consequently, surface waves have been used to study the subsurface structure of the Earth on global scales (Trampert & Woodhouse, 1995; Shapiro & Ritzwoller, 2002; Meier et al., 2007b,a), regional scales (Zielhuis & Nolet, 1994; Curtis et al., 1998; Simons et al., 2002) and reservoir scales (de Ridder & Dellinger, 2011; Mordret et al., 2013a, 2014a; Allmark et al., 2018). In those studies, subsurface information is deduced from the dispersion properties of surface wave phase or group velocities, with different frequency components constraining structures over different depth ranges in the subsurface.

Seismic surface wave tomography is often conducted using a two-step inversion scheme (Nakanishi & Anderson, 1983; Trampert & Woodhouse, 1995; Ritzwoller et al., 2002; Snoke & Sambridge, 2002; Bodin & Sambridge, 2009; Bodin et al., 2012; Galetti et al., 2017). First, a series of 2D phase or group velocity maps for different periods are estimated tomographically at each geographical point of interest using the arrival times of each period as data; the 1D dispersion curve at each geographical location is then inverted to estimate a 1D shear velocity structure beneath that location. Those 1D shear velocity structures placed side-by-side are interpolated to construct a 3D model.

The surface wave inversion problem is usually solved using a linearised procedure which involves approximating the true, nonlinear relation between data and parameters by a linearised relation; that approximate relationship is then used to seek an approximate solution by minimizing the data misfit while applying some regularization (Trampert & Woodhouse, 1995; Ritzwoller et al., 2002). However, the regularization is often chosen by ad hoc means (often trial and error), and valuable information in the data can be concealed by the regularization (Zhdanov, 2002). In addition, due to the irregular distribution of seismic sources and receivers, the subsurface is usually unevenly sampled, which limits the resolution of those region with poor data coverage (Curtis & Snieder, 2002). The introduction of ambient noise interferometry (Campillo & Paul, 2003; Wapenaar, 2004; van Manen et al., 2005, 2006; Wapenaar & Fokkema, 2006; Curtis et al., 2006) and ambient noise tomography (Shapiro et al., 2005; Lin et al., 2007; Yang et al., 2007; Bensen et al., 2009; Behr et al., 2010) has partly resolved this issue because earthquakes are no longer necessary in the region of interest as stations play the role of both receivers and (virtual) sources. However, the seismic stations on the Earth's surface are themselves far from uniformly distributed in many areas, and as a result it is difficult to quantify the uncertainties in information derived from linearised ambient noise tomography (Shapiro & Ritzwoller, 2002; Bensen et al., 2009; Yao & Van Der Hilst, 2009; Weaver et al., 2011; Nicolson et al., 2012, 2014). This limits the ability to determine the uncertainty of inferred subsurface shear-wave velocity structures.

To resolve these issues, Bodin & Sambridge (2009) proposed a method using the Markov chain Monte Carlo (McMC) algorithm to sample models from a posterior probability distribution (pdf), based on a Bayesian framework. McMC methods were introduced to Geophysics by Mosegaard & Tarantola (1995). Thereafter, Malinverno et al. (2000) introduced an extension of the standard method called *reversible jump McMC* (Green, 1995; Green & Hastie, 2009), which allows the number of model parameters (the dimensionality of parameter space) to change during the inversion, and applied it on an inversion of zero-offset vertical seismic profiles and an electrical resistivity sounding inversion (Malinverno, 2002). Bodin & Sambridge (2009) further applied this method to a seismic tomography problem, in which Voronoi tessellations were used to allow spatially irregular model cells to parametrise seismic wave-speed variations, and trans-dimensional inversion allowed the dimensionality of parameter space (the number of cells in the model) to vary. Thus, the model parameterization is dynamically adapted to the irregular data coverage (Curtis & Snieder, 2002), avoiding arbitrary regularization. The method was used successfully to estimate the Rayleigh-wave velocity models across Australia (Bodin & Sambridge, 2009), the upper crustal structure of central Java in Indonesia (Zulfakriza et al., 2014) and SE Tibet in China (Zheng et al., 2017). Young et al. (2013) extended this method to include the second depth-inversion step of surface wave tomography to obtain a 3D shear-wave velocity structure of the Tasmanian crust and its uncertainty. Galetti et al. (2015, 2017) further generalized the method by making it fully nonlinear, and observed the emergence of loop-like topologies of the uncertainty structure around velocity anomalies which define the spatial resolution of those structures.

However, when the two steps of surface wave tomography are conducted separately and sequentially, the solution to the 1-D depth inversion cannot interact directly with the 2D phase and group velocity tomography step. In the second step, usually only the mean and standard deviations of the phase and group velocity maps are used as input (Young et al., 2013; Galetti et al., 2017). This may cause the valuable information contained in the pdf of phase and group velocity maps to be lost, and bias the estimate of the pdf of the 3D shear wave velocity structures.

As a result, the lateral neighbours in the final 3D model typically preserve little of the 2D lateral spatial correlation information in the phase and group velocity maps.

To overcome both problems, we introduce a fully 3D model parameterization to implement 3D non-linearised surface wave tomography in one step, directly from period-dependent phase or group travel-time measurements, using the reversible jump Markov chain Monte Carlo (rj-McMC) method. This preserves the spatial correlation information in 3D, and consequently also in 2D dispersion maps. It also naturally avoids the loss of pdf information from the 2D phase or group velocity maps in the 3D models. Several previous efforts have been made to extend the trans-dimensional tomography method to 3D. For example, Hawkins & Sambridge (2015) used tree structures and wavelets with rj-McMC to implement 3D trans-dimensional tomography and applied it on a synthetic teleseismic body wave tomography problem. Piana Agostinetti et al. (2015) implemented a 3D trans-dimensional local body wave tomography method using 3D Voronoi tessellation. However, to-date, the fully 3D trans-dimensional scheme has not been used in any surface wave tomography problem, and its properties have not been studied.

In the following we first describe our method, as well as the standard linearised method and the 2-step nonlinear McMC method for surface wave tomography. We then compare our new method with the other methods using a synthetic test. We use the results to conclude that from a seismological point of view there seems little reason to use the 2-step nonlinear methods in future as they seem to be just as computationally demanding as the 3D method, and produce intuitively unreasonable solutions.

2.3 Methodology

Following Bodin & Sambridge (2009), we use the reversible jump Markov chain Monte Carlo (rj-McMC) algorithm (Green, 1995) to generate samples from the posterior probability density function (pdf) in our seismic velocity model space. In this section, we first describe and compare three possible methods which can be used for seismic surface wave tomography and which are compared below. Then we provide an overview of the rj-McMC algorithm.

2.3.1 2-step Linearised method

Conventionally surface wave tomographic problems are solved using linearised inversion methods (Iyer & Hirahara, 1993). For example, in a first step, phase or group velocity maps can be estimated by minimizing an objective function:

$$\phi = ||\mathbf{Gm} - \mathbf{d}|| + \lambda||\mathbf{m}|| + \mu||\mathbf{Lm}|| \tag{2.1}$$

where $\mathbf{G} = [\frac{\partial d_i}{\partial m_j}]$ is the first order derivative matrix which defines the assumed linearised physics between parameters $\mathbf{m}$ which describe the tomographic map of localized surface wave speeds, and $\mathbf{d}$ which is the inter-receiver or source-receiver measured dispersion data, $\mathbf{L}$ is a finite-difference derivative matrix and λ and μ are parameters that define the strength of damping and smoothing of the map, respectively. The resulting maps estimated at a number of frequencies can be used to estimate the shear velocity structure beneath each geographical point in a second step that uses a similar linearised expression. Since such tomographic problems are usually under-determined, some forms of regularization is necessary in order to construct a map. Though the regularization parameters λ and μ can be estimated by some accepted methods (e.g. the L-curve method - Hansen, 1992), their values are still ad hoc and valuable information in the data can be destroyed by any applied regularization. In reality, the subsurface is often unevenly sampled, which may lead to poor resolution in those areas with poor data coverage, rendering structure in those areas more susceptible to regularisation than in better-constrained areas.

2.3.2 2-step McMC method

Sampling based methods like McMC are often used to allow non-linearised 2-step inversion in surface wave tomography (Bodin & Sambridge, 2009; Bodin et al., 2012; Khan et al., 2013; Young et al., 2013; Rawlinson et al., 2014; Zulfakriza et al., 2014; Saygin et al., 2015; Galetti et al., 2015, 2017; Zheng et al., 2017). In the first step the 2D plane is usually partitioned into convex polygons using a Voronoi tessellation (Sambridge et al., 1995). Each polygon contains one point (called a site) and is defined by the region of space consisting of all of the points nearer to that site than to any other (Figure 2.1b). Similarly, the model in the second step (depth inversion) can be parameterised using 1D Voronoi tessellation as in Fig 1a (Young et al., 2013; Galetti et al., 2017). In both steps, the McMC method simulates many samples of model space that are consistent with both the data and any available prior information, and varies the tessellation geometry such that it is consistent with the data in a probabilistic sense.

Generally the 1D depth inversions in the second step are run independently at each geographical location without interaction, as this allows perfect parallelisation of what is a computationally demanding task. As a result, the lateral neighbours in the final 3D velocity model may preserve little of the 2D lateral spatial correlation information, and uncertainty estimates based on the set of samples may therefore be incorrect. One option to correct the above errors would be to

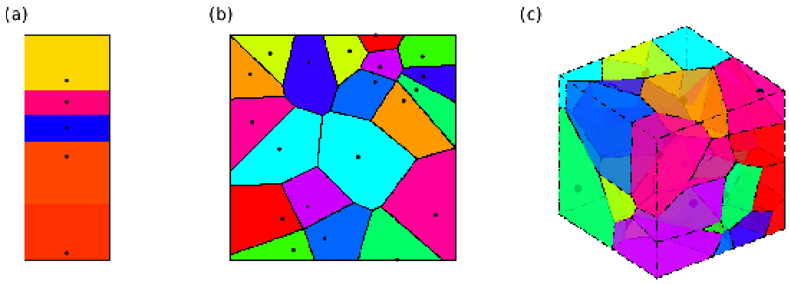


Figure 2.1: Examples of (a) a 1D Voronoi tessellation, (b) a 2D Voronoi tessellation and (c) a 3D Voronoi tessellation of velocity models. Colours represent seismic velocities in each cell. Black dots are the sites that generated each cell.

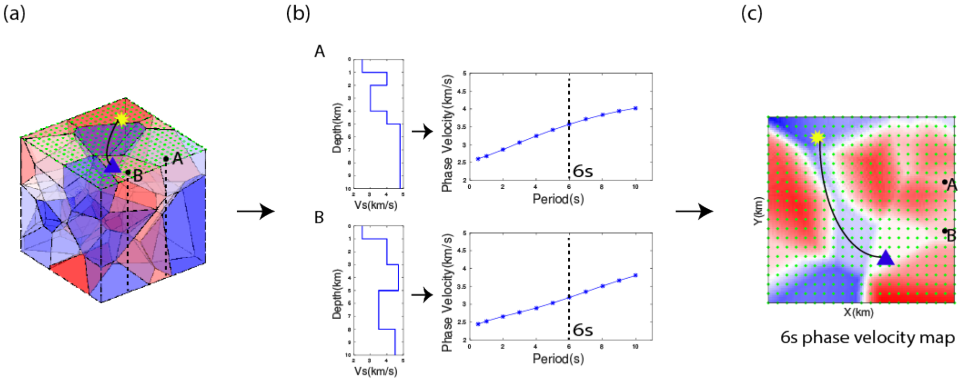
make each step-2 Markov chain interact with the chains of the lateral neighbours in such a way as to preserve lateral correlations observed in the step-1 inversion. However, this would remove or diminish the principal advantage of parallelisation as different chains would need to pass information to each other during the inversion. Alternatively, we can use a fully 3D Voronoi tessellation to parameterise our seismic velocity model (Figure 2.1c). 3D Voronoi tessellations have been used in seismic tomography problems to overcome the uneven distribution of data (Zhang & Thurber, 2005). They have also been used within rj-McMC schemes to implement non-linearised 3D body wave tomography (Piana Agostinetti et al., 2015; Burdick & Lekić, 2017). Here, we introduce the same for 3D single-step, non-linearised surface wave tomography.

2.3.3 Fully 3D Voronoi tessellation

Our 3D seismic velocity field is discretized by a set of Voronoi polyhedral, each of which is determined by its 3D site location $\mathbf{c}_i$ and shear-wave velocity v_i (P-wave velocity and density are linked to the shear velocity - see below) since surface wave phase and group velocities are primarily sensitive to subsurface shear-velocity variations. Here, we use a constant velocity within each polyhedron but other interpolations between sites could be used if desired (Sambridge et al., 1995).

In order to perform 3-D inversion we need a forward modelling method to calculate the surface wave dispersion that would be measured along any source-receiver path in the case that any particular 3D model were true. Ideally a fully 3D wavefield simulation method could be used but these are generally computationally too expensive. We therefore use an approximate 2-step forward modelling method (Reiter & Rodi, 2008). The first step is to compute phase or group velocity maps at each measurement period for our 3D earth model (Figure 2.2a and b); these can be determined by extracting the 1D shear velocity profile beneath each geographical point, and calculating the phase and group velocities for that 1D structure using a modal approximation (Saito, 1988; Herrmann, 2013). To calculate source-receiver phase travel times, we then use the fast marching method (Rawlinson & Sambridge, 2004) to compute travel times through the phase velocity maps for each period (Ritzwoller & Levshin, 1998; Stevens et al., 2001). For group travel times we integrate the group velocities along the ray path traced through the phase velocity map to determine the group travel times (Cerveny, 2005; Reiter & Rodi, 2008).

This forward simulation method is thus based on very similar approximations and assumptions to those made in 2-step inversion methods - that the dispersion properties of surface waves at each geographical point depend only on the velocity structure beneath that point. This has the advantage that it allows the 3D inversion method proposed herein to be compared fairly with the other methods (linearised and non-linearised 2-step inversion). Given this forward simulation



(a) A 3D Voronoi model (b) 1D S-wave velocity profiles and simulated dispersion curves (c) 2D phase velocity map at 6s

Figure 2.2: A schematic figure of the forward modelling method. (a) A 3D Voronoi model discretized by Voronoi cells, with a dense, regular grid of points overlain on the surface. Colours represent different velocities. Points A and B show two example locations at the surface. (b) The 1D S-wave velocity profiles at points A and B extracted from the 3D Voronoi model (centre-left) and the simulated phase velocity dispersion curves using these 1D profiles (centre-right). (c) An example phase velocity map at 6s constructed by juxtaposing the phase velocity at 6s period calculated for all geographical points such as A and B. The source-receiver phase travel time can then be calculated, for example by tracing rays through the model or by using a fast marching method. The yellow star shows the location of one source and the blue triangle shows the location of one receiver.

scheme and measured dispersion data, the rj-McMC method can be used to perform 3D tomography.

2.3.4 Reversible jump Markov chain Monte Carlo

McMC is a class of algorithms to generate a set (or chain) of samples from a target probability density (Sivia, 1996). The Metropolis-Hastings algorithm (Metropolis & Ulam, 1949; Hastings, 1970) is one such algorithms which is used in many fields.

It was introduced to Geophysics over two decades ago (Mosegaard & Tarantola, 1995; Malinverno et al., 2000; Malinverno, 2002; Malinverno & Briggs, 2004) and was first applied to a seismic tomography problem by (Bodin & Sambridge, 2009). In their method, a generalised version of the Metropolis-Hastings algorithm called reversible jump Markov chain Monte Carlo (rj-McMC) was used to allow a trans-dimensional inversion (Green, 1995; Green & Hastie, 2009) which means that the number of model parameters can change along the chain. This allows the parameterization of the seismic velocity model itself to be determined by data and any prior information, avoiding fixing the parameterisation before inversion (Bodin & Sambridge, 2009). Following their lead, we apply the rj-McMC algorithm to 3D seismic surface wave tomography.

In our problem, the target probability density is the posterior probability density function (pdf) of our velocity model $\mathbf{m}$ given the observed data $\mathbf{d}_{obs}$, written $p(\mathbf{m}|\mathbf{d}_{obs})$. According to Bayes' theorem,

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})} \tag{2.2}$$

where $p(\mathbf{d}_{obs}|\mathbf{m})$ is called the *likelihood* which is the probability of observing the measured data conditional on a certain model $\mathbf{m}$ being true; $p(\mathbf{m})$ describes the prior information about model $\mathbf{m}$, and $p(\mathbf{d})$ is a normalization factor called the *evidence*. We choose a Gaussian noise distribution for our likelihood with the data variance as an additional parameter that is also estimated during the inversion in a hierarchical way (for more information see Malinverno & Briggs, 2004; Bodin et al., 2012; Galetti et al., 2017). For the prior pdf, we use an uninformative prior - a Uniform distribution with wide bounds on each parameters.

In the rj-McMC algorithm, a new model $\mathbf{m}^{'}$ in the chain is drawn from a proposal distribution $q(\mathbf{m}^{'}|\mathbf{m})$ that depends on the current model $\mathbf{m}$, and is accepted or rejected with a probability $\alpha(\mathbf{m}^{'}|\mathbf{m})$ given by (Green, 1995),

$$\alpha(\mathbf{m}^{'}|\mathbf{m}) = min[1, \frac{p(\mathbf{m}^{'})}{p(\mathbf{m})} \times \frac{q(\mathbf{m}|\mathbf{m}^{'})}{q(\mathbf{m}^{'}|\mathbf{m})} \times \frac{p(\mathbf{d}_{obs}|\mathbf{m}^{'})}{p(\mathbf{d}_{obs}|\mathbf{m})} \times |\mathbf{J}|] \tag{2.3}$$

where $\mathbf{J}$ is the Jacobian matrix of the transformation from $\mathbf{m}$ to $\mathbf{m}^{'}$ and is used to account for the volume changes of parameter space during jumps between dimensionalities. In our case, it can be shown that the Jacobian is an identity matrix (Bodin & Sambridge, 2009). Once a new model is generated via the proposal distribution, it is accepted or rejected by generating a random number γ from the uniform distribution on [0, 1] and comparing it with the value of the acceptance ratio α. If $\gamma < \alpha$, the new model is accepted; otherwise, the new model is rejected and the current model is repeated as a new sample in the chain. The acceptance ratio α ensures that the density of samples in the Markov chain converges to the posterior probability distribution as the number of samples tends to infinity (Green, 1995).

In seismic tomography problems we have five types of perturbations: adding a cell, removing a cell, moving a cell, changing the velocities and changing the data noise hyperparameters. Thus, our algorithm can be described as:

1. Draw an initial model randomly from the prior pdf.

2. Generate a new model $\mathbf{m}^{'}$ by randomly choosing one of the five possible perturbation types listed above, and then perturbing the current model according to the proposal probability.

3. Calculate the acceptance ratio α and accept or reject the model according to α.

4. Repeat from (ii).

For the proposal probability we choose a Gaussian distribution for the fixed-dimensional perturbation (moving a cell, changing velocities and changing data noise hyperparameters) as also chosen by Bodin & Sambridge (2009). For trans-dimensional perturbations (adding or deleting a cell) we choose to use the prior pdf as the proposal probability since that leads to a higher acceptance ratio compared to using a Gaussian distribution (Dosso et al., 2014). It is a property of McMC methods that in principle the choice of proposal distribution does not affect the fact that the final distribution of samples tends to the posterior pdf as the number of samples tends to infinity.

Note that successive models in a Markov chain are not independent, which for any finite set of samples might cause bias in the estimated posterior probability distribution (Chan & Geyer, 1994). Thus, some thinning of the chain, retaining only every 100th sample of the chain, is applied to obtain a final ensemble of samples. Thereafter, statistical properties of the inverse problem solution (e.g. mean, standard deviation) can be calculated from the remaining ensemble of samples.

Monitoring of McMC convergence is important to ensure that the estimated posterior probability density becomes stationary. However, this appears to be a difficult problem, especially for trans-dimensional chains (Green & Hastie, 2009). In this study, we chose several scalar statistics, such as the misfit and the number of cells, to diagnose apparent convergence. When the misfit and the number of cells become stationary, we assume that convergence is attained and begin retaining every hundredth sample from that point on in the chain, which seems to be sufficient in our synthetic test. However, we note that in other more complicated applications, using such scalar statistics may be insufficient such that a more sophisticated approach may be necessary (Green & Hastie, 2009).

2.4 Results

To validate our method, we conducted a synthetic test using Rayleigh wave phase velocity dispersion data. Our true model is composed of 3 layers with S-wave velocities of 2.5km/s, 4km/s and 5.0km/s respectively and a spherical low velocity anomaly of radius 1km within the second layer (Figure 2.3a, b), of which the velocity is 3.0km/s. A sphere was chosen deliberately because it definitely lies outside of the range of finite-dimensional, straight-sided Voronoi cell parameterisations that can be explored by the Markov chain. Sources and receivers occur at idealised locations to simulate a typical ambient noise experiment where receivers are also used as virtual sources. To calculate the corresponding Rayleigh wave phase velocity dispersion data we use the two-step simulation method described above. First, the phase velocity map is calculated for each period using a 200 × 200 regular grid on the surface (Figure 2.3c). The P-wave velocity is given by a fixed v_p/v_s ratio with a typical crustal value of 1.73, and density is assumed to be dependent on v_p through

$$\rho = 2.35 + 0.036 \times (v_p - 3.0)^2 \tag{2.4}$$

where v_p is given in km/s and ρ is given in g/cm^3 (Kurita, 1973). Then, the phase velocity travel times of surface waves are calculated at each period between each station. We then added 0.01 seconds (∼1%) Gaussian noise to the data. This gives 28 travel times in total for each period, and we use eleven periods between 0.5s to 10s (red dots in Figure2.3d).

To demonstrate our 1-step 3D method, we compared it with the linearised 2-step method and the 2-step McMC method using the synthetic dataset. For the fully 3D McMC method, the velocity prior pdf is set to be a Uniform distribution between 2km/s and 6km/s, which encompasses the true model. The prior pdf

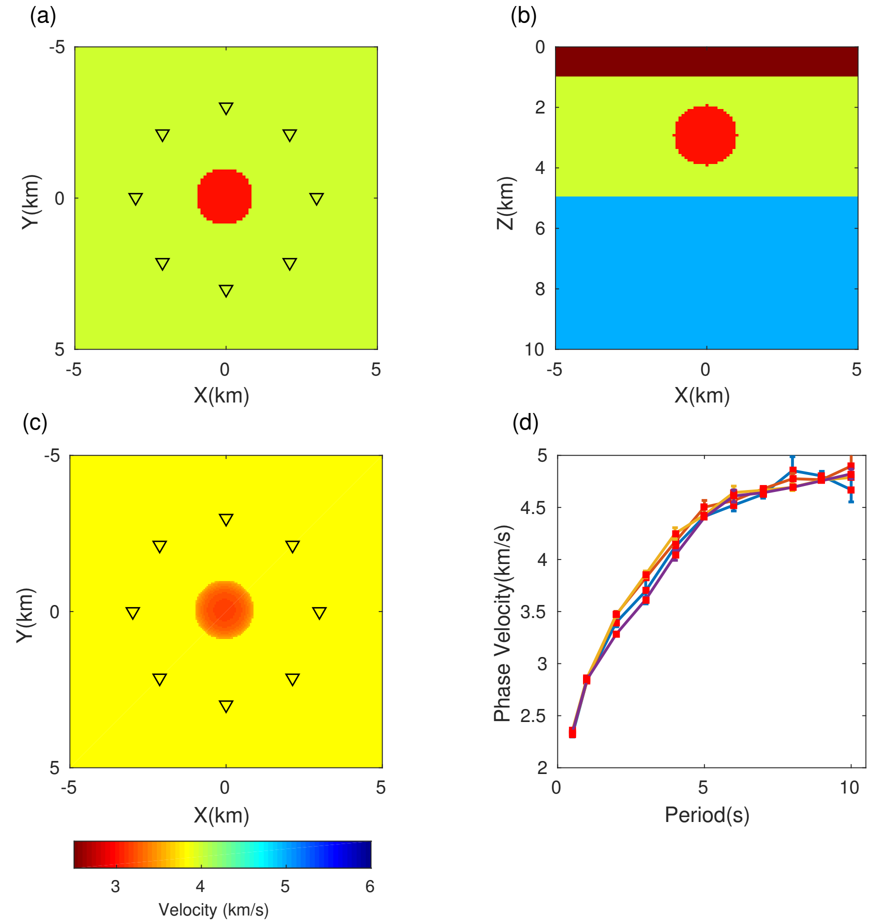


Figure 2.3: True model and data used for the synthetic test. Open black triangles show the locations of sources and receivers, which are colocated to simulate a typical ambient noise experiment. (a) Horizontal section of the true model at depth of 3km; (b) vertical section at Y=0km; (c) An example phase velocity map at 3s period; (d) Examples of modelled inter-receiver phase velocity dispersion curves.

on the number of Voronoi cells is selected to be a discrete Uniform distribution between 4 and 200. The prior pdf of the two noise parameters are set to be a Uniform distribution between 0.00001 and 0.01, and a Uniform distribution between 0 and 0.03, respectively. As described above, we use the *prior* distribution for the trans-dimensional perturbation (cell birth and death) and a Gaussian perturbation for the fixed-dimensional steps (changing velocity, position and noise hyperparameters). The width of each Gaussian perturbation is tuned to produce an acceptance rate between 20% to 50%. In this test, we used 16 independent Markov chains, each generating 4 million samples. After a burn-in period of 1 million samples, we thinned each chain by retaining only every 100^{th} sample.

For the 2-step McMC method, the first step is to determine the phase velocity for each period. Here we use the same method as Galetti et al. (2015) which is a fully nonlinear McMC 2D tomographic method using 2D Voronoi tessellation. The velocity prior for each period is set to be a Uniform distribution with a 2km/s width centred at an averaged velocity of all of the ray paths. The prior for the number of Voronoi cells is selected to be a discrete Uniform distribution between 3 and 100 cells (considering that the true phase velocity maps are relatively simple - e.g., Figure 2.3c). Similarly to above, the width of each Gaussian perturbation for fixed-dimensional steps (changing velocity, position and noise hyperparameters) is tuned to produce an acceptance rate between 20% to 50%. The width of the Gaussian perturbation for trans-dimensional steps (cell birth or death) is selected to give the maximum possible acceptance ratio. For each period, we used 16 independent chains to generate samples from the posterior probability density each with 3 million iterations. After a burn-in of 0.5 million samples, we thinned these chains by retaining only every 100^{th} model to create the final ensemble. Then the phase velocity mean and its standard deviation of the ensemble are calculated at each period on a 100×100 regular grid, and these are taken as the data for the second inversion step. In the second step, we use the McMC based method from Galetti et al. (2017) to determine a 1D shear-wave velocity model beneath each geographic point. For each inversion, we use the same shear-wave velocity prior as in the 3D McMC method – a Uniform distribution between 2km/s and 6km/s. The prior for the number of layers in each 1D depth profile is set to be a discrete uniform distribution between 2 to 20. The proposal distribution for velocity is chosen to be the same Gaussian perturbation as used in the 3D McMC inversion. Other proposal Gaussian distributions are tuned to provide an acceptance rate between 20% to 50%. Similarly, the Gaussian perturbation for the trans-dimensional step (layer birth and death) is selected to give a maximum acceptance ratio. For each 1D inversion beneath each geographical point, we use 6 independent chains, each generating 3 million samples. After a burn-in period of 1 million samples, each chain is thinned by retaining only every 100^{th} model.

For the conventional 2-step linearised inversion, first we need to determine the phase velocity map for each period at which we have data. Here we use fast-marching surface wave tomography (Rawlinson & Sambridge, 2004) to determine the phase velocity. For each period, the initial model is chosen to be a homogeneous model with an average velocity of all of the ray paths. The regularization damping and smoothing factors are selected using the standard L-curve method (Hansen, 1992). The model is parameterized using a 20×20 regular square grid, which was chosen from a series of regular grids (e.g. 5×5, 10×10, 20×20, 30×30) to give the least model parameters that also produced a small misfit tolerance. For the second step, we use the Occam inversion method (Lai & Rix, 1998) to determine the shear-wave velocity structure beneath each grid point. For each 1D inversion beneath each geographical point, the model is parameterized by an evenly spaced layered model with 1km thickness for each layer. The initial velocity of each layer is set to be 1.1 times the phase velocity at the most sensitive period for that particular depth (since for a homogeneous medium the Rayleigh phase velocity is 0.92 times the shear velocity of the medium). The misfit tolerance

Table 2.1: Parametrization, priors and proposal distribution for all 3 inversions

Method	Parametrization	Priors	Proposal distributions
3D McMC	3D Voronoi cells	cells number: Uniform 4-200	Fixed-D: Gaussian
		velocities: Uniform 2 - 6km/s	Trans-D: prior
2-step McMC	2D Voronoi cells	cells number: Uniform 3-100	Fixed-D: Gaussian
		velocities: 2km/s width Uniform	Trans-D: Gaussian
	1D Voronoi layers	layers number: Uniform 2-20	Fixed-D: Gaussian
		velocities: Uniform 2-6km/s	Trans-D: Gaussian
Linearised inversion	20 × 20 grid	homogeneous initial model	NA
	1km layered model	estimated from phase velocities	NA

For the linearised inversion, we regard initial models as priors.

for the Occam inversion is tuned to give a balance between model complexity and data fitting. To conclude, we summarised the parameterization, priors and proposal distributions used in all three inversions in Table 1.

2.4.1 Model comparison

Figure 2.4 shows the mean, standard deviation and relative error ($|\mathbf{m}_i^{mean} - \mathbf{m}_i^{true}|/\sigma_i$) determined from the ensemble generated by the fully 3D 1-step McMC method. The mean velocity model clearly shows a low velocity anomaly at the center of the second layer. It is not a perfect sphere due to the way we parameterized our model (Voronoi tessellation) as cells have straight edges. We also observe three layers in the mean model, though they are not exactly the same as in the true model. Since surface waves are more sensitive to the shallow structures, the top layers are better determined than bottom layers and have smaller uncertainties. From the relative error map, we can see that the error is within 1 standard deviation across most of the model, except at the bottom boundary. This means we have successfully recovered the true model to within approximately correct uncertainty estimates. The uncertainty map at 3km depth in the upper-middle panel of Figure 2.4 shows low uncertainties in the middle area covered by the data, surrounded by high uncertainties due to lack of data coverage. There is a loop of higher uncertainty around the boundary of the low velocity sphere and at the boundaries between different layers. These loops (in this case actually 3D spherical shells as shown in the cross-section in the lower-middle panel of Figure 2.4) exist due to the multi-modality of the posterior probability density (essentially that the data can not define whether any point in the loop is inside or outside of that sphere, and hence whether it is a point of low or high velocity) and are mainly caused by ray bending - so second or higher order aspects of wave physics (Galetti et al., 2015, 2017). They define uncertainties in the boundary and shape of velocity anomalies, which could be used to help interpret the velocity map. These are the first 3D uncertainty loops that have been observed, though their existence in 3D was conjectured by Galetti et al. (2015).

The noise level generally affects the complexity of the models obtained (Bodin et al., 2012). Here we show some examples of the histograms of two noise parameters in Figure 2.5. The noise level is derived from these two parameters through a linear relationship with the ray length (Bodin et al., 2012; Galetti et al., 2017). However, the results do not converge to the true distribution ($\sigma_0 = 0, \sigma_1 = 0.01$). This is probably due to the fact that we used a step size of 0.001 for the proposal distribution, which makes it difficult for σ_0 to approach zero and consequently leads σ_1 to be smaller. In addition, although we added noise with a standard deviation of 0.01 second to the data, it is also quite reasonable that the post-inversion noise level is smaller or larger than the initial noise level because the posterior distribution also accounts for consistency with

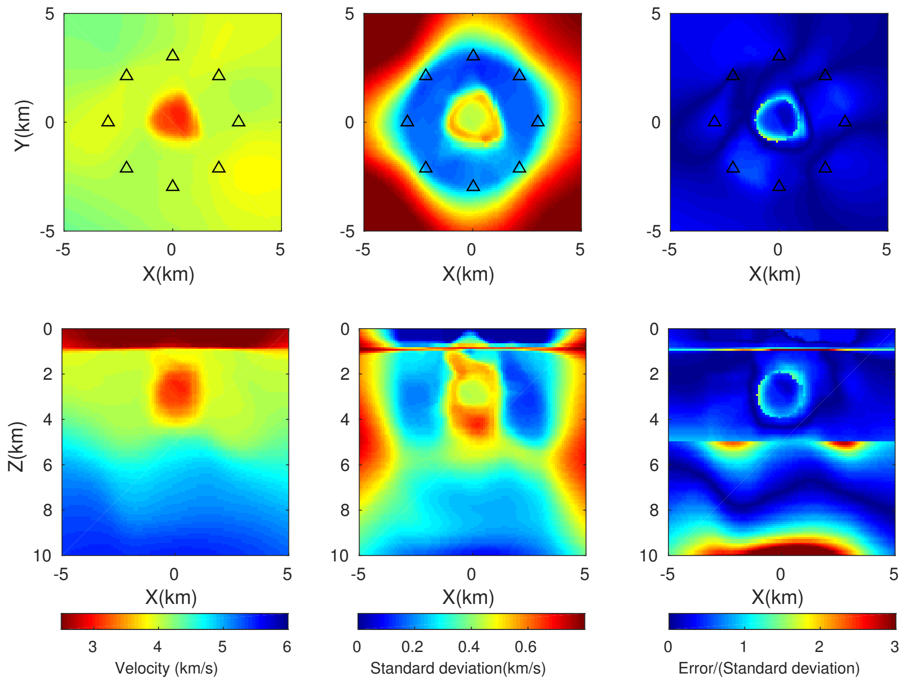


Figure 2.4: The mean velocity model (left), standard deviation (middle) and the number of standard deviations of the error (right) of the horizontal slice of Z=3km (top) and at the vertical slice of Y=0km (bottom) using 3D McMC method.

the prior range of models. Note that here we added independent Gaussian noise for each datum. In reality the data noise might not be independent, especially at neighbouring frequencies.

To compare our new method with the more standard 2-step McMC method, we show the latter results in Figure 2.6. The mean velocity model suggests that the low velocity anomaly is clearly estimated, but similar to the results determined using the fully 3D McMC method, its shape is not perfect recovered. The top layer is clearly recovered due to high sensitivities at shallow depth, except that the layer boundary is deeper at either side which is likely to be a consequence of the prior information due to the lack of data in those areas. Across the model the magnitude of uncertainties are far higher from the 2-step McMC than those determined from

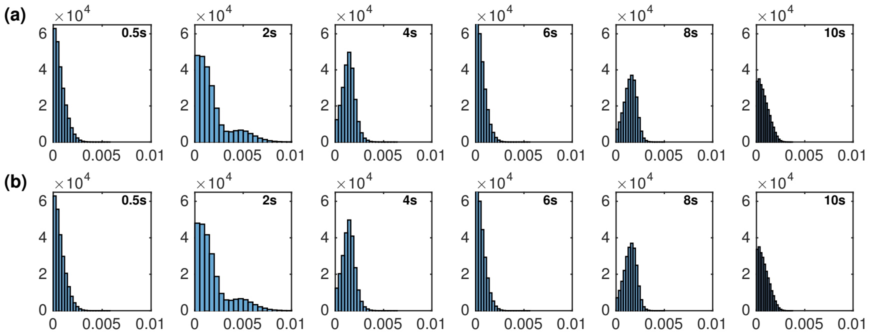


Figure 2.5: Examples of the histograms of two noise parameters at periods of 0.5s, 2s, 4s, 8s and 10s. The top row shows σ_0, the bottom row shows σ_1. The noise level is derived from them by $\sigma = \sigma_0 * raylength + \sigma_1$.

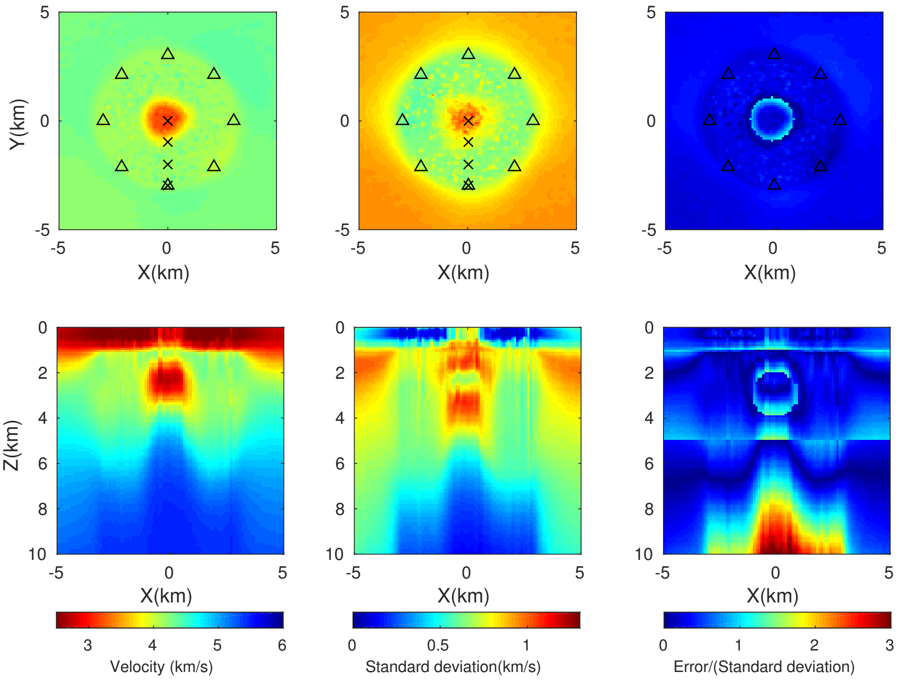


Figure 2.6: The mean velocity model (left), standard deviation (middle) and the number of standard deviations of the error (right) of the horizontal slice of Z=3km (top) and at the vertical slice of Y=0km (bottom) using 2-step McMC method. The crosses show the point location which are referred to in the text.

the 1-step 3D McMC method. Even so, the relative error map suggests that much of the bottom layer still has errors of 2 to 3 standard deviations. Parts of the uncertainty loops are observed in the centre of the model at the boundary of the low velocity anomaly. However these do not join up to create an uncertainty shell in 3D as we could expect intuitively – indeed the uncertainty loop in the horizontal plane is not recovered at all by this method, even though these are clearly observed in the first step of the inversion (Galetti et al. 2015; see Figure 2.7 herein). This shows that the standard 2-step McMC method loses lateral spatial correlations in the second step of the inversion and hence loses the uncertainty loops around the lateral extremities of anomalies, while the fully 3D McMC method provides more intuitively correct results. In other words, the high uncertainty loops at the boundaries of anomalies observed in the results of 3D McMC method spuriously disappear in the results of the 2-step McMC method.

Notice that there is a consistency problem in this comparison: by changing the parametrization of the model, we have also implicitly changed the prior information that is included in the Bayesian solution. This is almost inevitable in such nonlinear problem. It does not diminish the usefulness of comparing solutions, but does mean that we cannot interpret the results as a direct comparison of solutions to exactly the same problem.

To analyse the possible errors introduced by the 2-step McMC method, in Figure 2.8 we show the phase velocity posterior pdf as well as their mean and uncertainties determined in the first step along with the true phase velocities at four points shown in Figure 2.6 (top-left-panel). Those mean phase velocities at points which lie outside of the low velocity anomaly are consistent with the true phase velocities. However, phase velocities within the low velocity anomaly deviate to several standard deviations away from the true velocities at the same location, especially at periods which are sensitive to the low velocity anomaly.

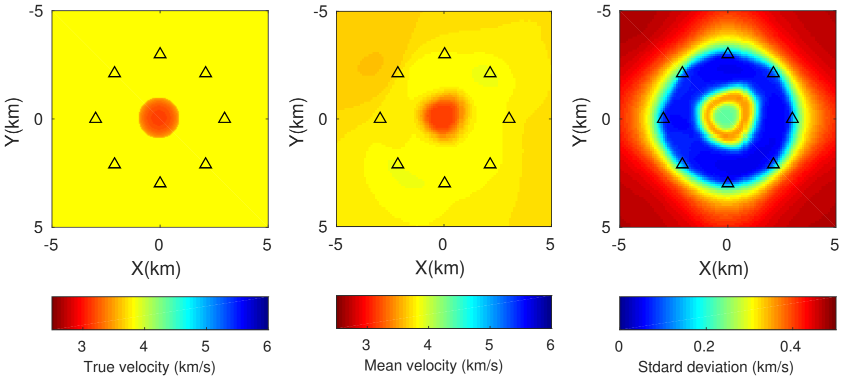


Figure 2.7: An example of the true phase velocity map (left), the mean phase velocity map estimate (middle) and the standard deviation map (right) at 3s period estimated using the first step of the 2-step McMC method.

It is likely that this is because in this first step of the 2-step method we assume that the velocity at each period is completely decoupled from that of every other period, thus correlations across periods that are imposed by the physics of the problem are lost in this method. This leads to bias in the phase velocities, and subsequently to bias in the shear velocities in the second step when we perform a 1D inversion independently at each geographical point. By using a direct 3D method we naturally avoid this drawback because velocities at different periods are naturally coupled and correlated due to the structure of the model with depth: the shear velocity at each depth in the model contributes to the phase velocity over a continuous range of periods, thus imposing phase velocity correlations across periods that are consistent with the assumed physics of the problem. Note that the mean phase velocities at the boundary are far away from true velocities due to the multimodality in the posterior (top-right-panel in Figure 2.8). As a result, when using only the mean and the standard deviation as data in the second step (as is standard practice), the shear velocity structure is biased.

Figure 2.9 shows results from the standard 2-step linearised inversion. The

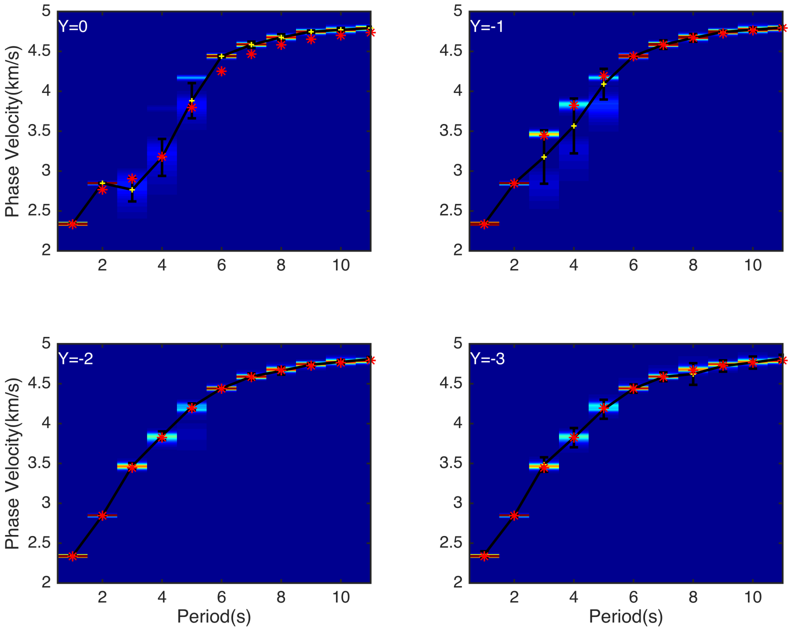


Figure 2.8: The phase velocity posterior probability density distribution (shading - lighter colours have higher probability) and their mean (yellow plus) and standard deviation (error bar) estimated using the first step of the 2-step McMC method at points Y=0, Y=-1, Y=-2 and Y=-3km shown in Figure 2.6 (top-left panel). Red stars represent the true dispersion curves.

velocity model clearly shows the low velocity anomaly and the three layers, though their velocity values are slightly biased which is probably caused by regularization. There are some artefacts around the low velocity anomaly and in the deeper parts of the model, which suggests that the data are overfitted. By comparison, in the McMC based inversions because the parameterization is determined by the data, overfitting is reduced which produces a better estimate of subsurface structure. The magnitude of uncertainties estimated by linearised inversion is significantly larger than those from either of the McMC based methods, which leads to small relative errors in the right panels of Figure 2.9. However, the linearised estimates of uncertainty are generally unreliable because they rely entirely on an estimate of the gradients of data with respect to parameters at only a single point in parameter space. Without prior knowledge of the structure of the problem and its gradients throughout this space, it is really just a chance event whether predicted uncertainties are larger or smaller than they should be in the nonlinear problem. Besides, regularization is often used in linearised problem, making it difficult to quantify the correct uncertainties. Thus, the linearised estimates of uncertainty provide little reliable information to aid interpretation of the velocity model.

2.4.2 Uncertainty analysis

In results from the 1-step 3D and 2-step McMC methods, there are low velocity uncertainties in the very bottom layer, which is counter intuitive since surface waves are less sensitive to deep structures. In retrospect we notice that this was also exhibited in the result from the surface wave dispersion curve inversion of Bodin et al. (2012). To further understand this phenomenon we performed several 1D tests using both trans-dimensional inversion and fixed-dimensional inversion with different period ranges of data (Figure 2.10). The true shear velocity model is set to have 8 layers down to 30km (the blue structure in Figure 2.10). The

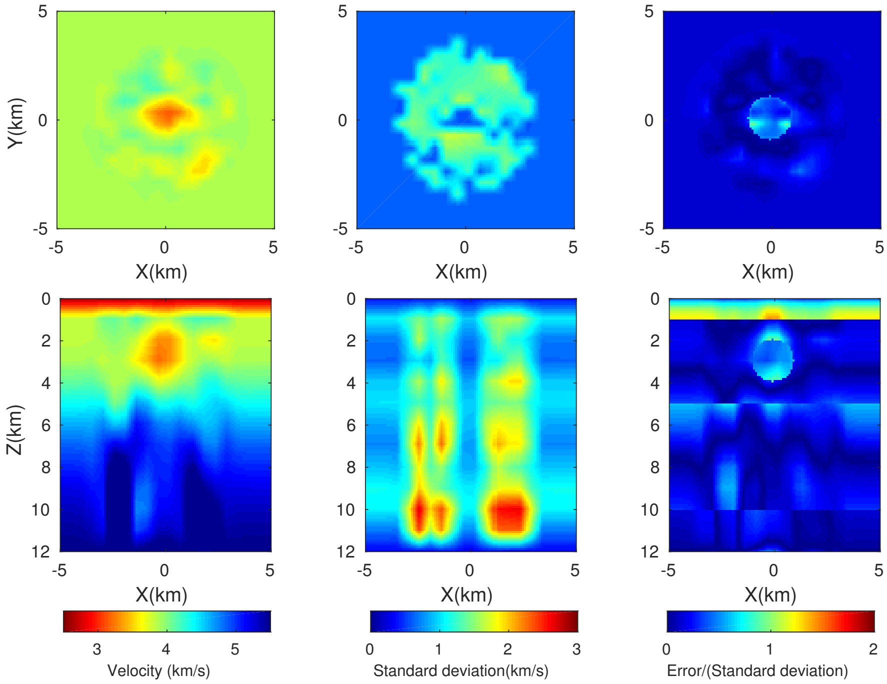
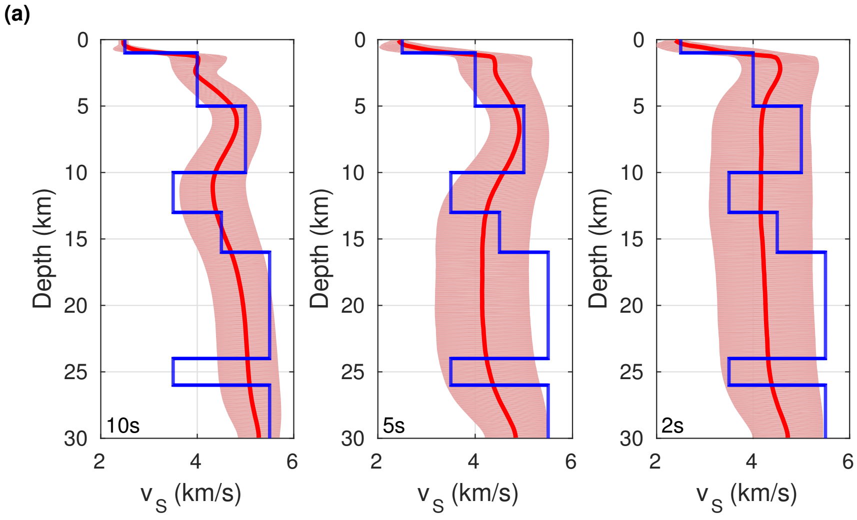
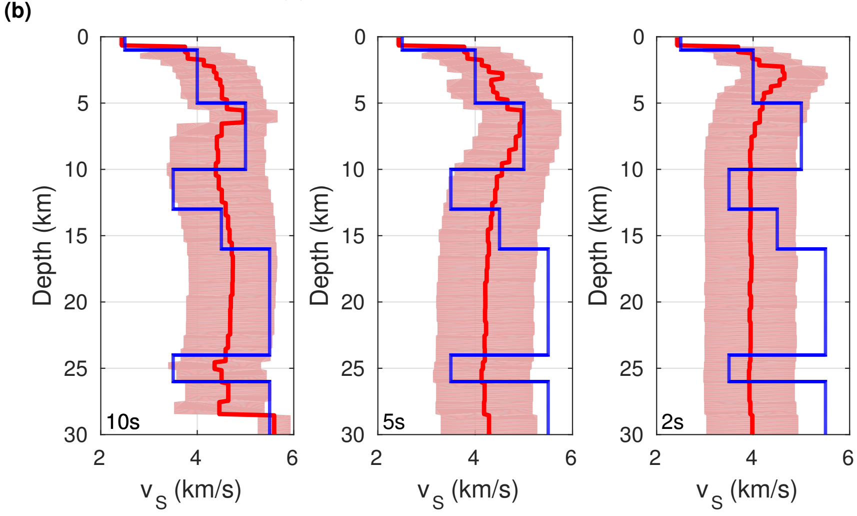


Figure 2.9: The mean velocity model (left), standard deviation (middle) and the number of standard deviations of the error (right) of the horizontal slice of Z=3km (top) and at the vertical slice of Y=0km (bottom) using standard linearised inversion.



(a) Trans-dimensional inversion



(b) Fixed-dimensional inversion

Figure 2.10: 1D dispersion curve inversion using (a) trans-dimensional inversion and (b) fixed-dimensional inversion. The blue line shows the true shear velocity profile, while the red line shows the posterior mean velocity with its uncertainties (pink shading). At the bottom-left of each figure we show the longest period used in the inversion; in each case the lowest period used was 0.5s and then equally spaced periods (spacing 1s) from 1s up to the maximum were included.

period ranges used in these inversions are 0.5-10s, 0.5-5s and 0.5-2s for the left, central and right panels respectively. In the fixed dimensional inversion, the nodes are set to be a regular grid of size 0.5km above 5km depth and size 1km below 5km depth.

Both of the results from 10s inversion show low uncertainties in the deepest layer. The mean velocity shows that we have some resolution at deeper depths (below 10km) because it deviates from the prior mean model of 4km/s towards the true velocity. The relatively more correct mean velocities and smaller uncertainties show that trans-dimensional inversion has slightly higher resolution at deeper depths than the fixed-dimensional inversion. The low uncertainty at the deepest layer is probably due to the accumulated resolutions to deeper depths down to several tens of kilometres at long periods because within the modal approximation used in the forward modellers in our method and that of Bodin et al. (2012), the model is actually assumed to be a half-space below the deepest layers boundary at 30km. When we reduce the longest periods to 5s, there is no low uncertainty at the deepest layer using fixed-dimensional inversion. However, the low uncertainties still exists in trans-dimensional inversion, even if we reduce the longest periods to 2s. The mean velocity model and its uncertainty shows that in reality we do not have any resolution at depths below 15km in this case since the results simply reflect the prior information. Thus, the low uncertainty in the deepest layer is not geophysically interpretable and it always exists in trans-dimensional inversion using Voronoi cells. This is probably due to the fact that natural parsimony of trans-dimensional inversion prefers only one or a few nodes to represent the model where we have little or no resolution, so the lower most Voronoi cell always combines regions where we have no resolution with those in which some resolution exists.

2.4.3 Computational cost

It is clear that the linearised inversion method is substantially less computationally demanding than sampling based methods. However, it usually gives incorrect estimates of uncertainties since it neglects the non-linearity of the system. Therefore, here we compare the computational cost of the two McMC based methods. Generally, McMC methods need large computational power, especially in high dimensional parameter spaces. To overcome this issue, Bodin & Sambridge (2009) fixed the ray geometry during each McMC chain (thus partly linearising the problem), and updated the rays only between successive McMC chains. However, Galetti et al. (2015) showed that this may introduce artefacts and bias in the solution. Thus, in our study we update the ray geometry in every iteration both for the 2D map inversion in the 2-step method, and in our 1-step 3D McMC method. In our method, every new model is a small perturbation of the previous model, which almost always involves only a small number of cells (Jamin et al., 2018). Thus, in the first step of our two-step forward modelling scheme, we only need to update the phase or group velocity dispersion curves affected by these perturbations, which offers a very significant saving in computation. For example, in one million samples in the above examples, the fully 3D McMC method involves ∼3,000 million forward modal simulations of phase or group velocity from 1D shear velocity profiles, while the standard 2-step McMC method involves 10,000 million forward modal simulations. Given that the 'true' model that we use for tests herein is simple, this saving is expected to be substantially greater in a complicated earth structure which needs more cells to represent the model.

However, due to the higher parameter space dimensionality of a 3D model, our new method might require more samples to generate a stationary estimate of posterior probability distributions. For example, in the above synthetic test, the 3D McMC method needs 4 million samples including a 1 million sample burn-in period, while the 2D McMC method needs only 3 million samples including a 0.5 million sample burn-in period. Nevertheless, compared to the time saved in forward modelling described above, this increase in burn-in is not significant in this case. For example, in the above synthetic test the 2-step McMC method involved 30,000 million modal simulations and takes ∼4,000 cpu hours for one chain, while the 3D McMC method involves 12,000 million modal simulations and takes ∼1,600 cpu hours for one chain. To provide an idea of the overall computational time, the above 3D synthetic test costs 160 hours with each chain parallelized with 9 CPU cores.

Note that the computational cost in each case depends strongly on the method used to assess convergence, which is difficult and depends on some subjective choices. In turn, this introduces some subjectivity to the comparison of computational cost between these methods. Despite this it is at least true that the computational cost of the 1-step and 2-step MC method is comparable in our example. We note however that some other more efficient Bayesian inference methods could be used more easily in the 2-step method than in the 1-step method. For example, Meier et al. (2007a) used a Gaussian mixture model to invert for 1D shear-velocity structure from phase velocities, which is more computationally efficient.

2.5 Scaling of Voronoi cells

Voronoi cells are usually defined by a L2-norm distance metric which treats different directions equally, and in particular which has equal lateral and vertical scales. However, seismic velocities often vary more in the vertical direction than laterally; that is, a large aspect ratio model with horizontal major axes may be more likely than an equal aspect ratio model. This will potentially cause models

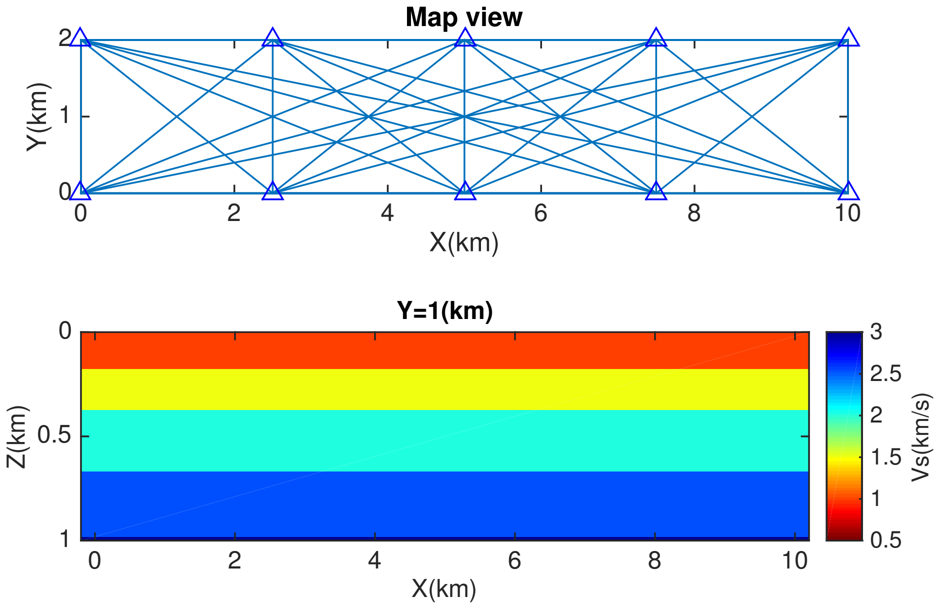


Figure 2.11: The map view of the true model plotted with receivers and data coverage (top panel) and a vertical slice of Y=1km (bottom panel).

to be distorted when using 3D Voronoi cells. In order to address this potential issue, we tested explicitly weighting the vertical scale compared to lateral scales. Here we conducted some synthetic tests to show the potential limitation of 3D Voronoi cells and the effects of different scaling factors added on the vertical scale.

To focus on large aspect ratio models, we created a $10 \times 2 \times 1km$ layered model with a maximum 10:1 lateral and vertical aspect ratio (Figure 2.11). The model is composed of 4 layers with shear wave velocity of 1km/s, 1.5km/s, 2km/s and 2.5km/s, respectively. The P wave velocity is derived from shear wave velocity using a fixed ratio of 1.73. We used 10 receivers, each of which is also used as a virtual source. We used phase velocity data at 9 periods from 0.5s to 2.1s with an equal spacing of 0.2s. To focus on the Voronoi parametrization, we fixed the noise level using a standard deviation of 2 percent of the data.

In order to test the potential effects of different weights added on the vertical scale, we explicitly weighted the vertical scale by factors of 1, 4 and 10. Figure 2.12 shows some results from those three inversions. For each inversion, we used 8 independent Markov chains. In the top panel we show the number of cells versus computational time. After about 300,000 seconds the inversion with scale factor 1 (left panel) has still not converged, while the other two inversions approximately converge. In terms of the number of cells, the inversion with scaling factor 10 converges fastest (converged after 120,000s). For all three inversions we started to collect posterior samples after a burn-in period of 150,000 seconds and thinned each sample chain by a factor of 100. We show the mean velocity models across a vertical slice (Y=1km) in the middle panel of Figure 2.12 and their uncertainties at the bottom panel. The mean velocity of the inversion with scaling factor 1 only contains 2 layers associated with relatively large uncertainties. The boundary between the layers also varies significantly across the model. The other two inversions recovered the model significantly better, though the second layer seems to be smeared out which is probably due to the limited resolution of the data used. In both case, the uncertainty maps show high uncertainties at the boundary of layers. For the boundary location of the bottom layer, the inversion with scaling factor 10 recovers the model slightly better with less variation across the model.

Since qualitatively the same results would be expected to occur in 2D or 3D, we conclude that using either Voronoi cells without scaling could cause distortion and bias in inverted results, but that this could be resolved by explicitly adding a scaling factor on the vertical scale. However, this introduces another parameter to the inversion (the scaling factor). In our simple example, this choice does not affect the results significantly since our model is not particularly high in its aspect ratio, but this issue may need to be considered for more complicated real problem.

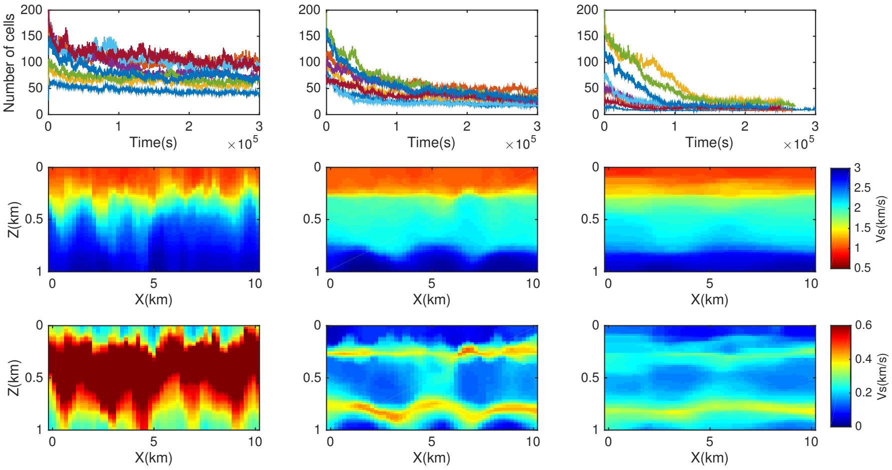


Figure 2.12: The number of cells versus computational time (top panel), the mean velocity model (middle panel) and standard deviation (bottom panel) across the vertical slice of Y=1km using 3 different vertical scaling factors 1 (left panel), 4 (middle panel) and 10 (right panel).

2.6 Discussion

We have shown that using 3D Voronoi tessellation in a McMC surface wave tomography method preserves spatial correlations and better estimates the uncertainties of velocity structures. Because of computational restrictions, we used an approximate forward modelling method in our inversion which still uses the 1D modal approximation to estimate phase velocities at each geographical location across our model. It is certainly possible that this approximation might cause errors in our final model. However, this is the same approximation that is used in the 2-step inversion method, which renders our comparison between the 1-step and 2-step methods fair. Also, there is no impediment to using more precise forward modelling methods if sufficient computational power is available.

In our method, Voronoi cells only need to be updated locally around any change in sites in each iteration. For geometry changes (i.e. move, birth and death) we implement a local change method (Jamin et al., 2018). We keep track of the underlying grid velocities at every iteration so that any velocity change can be updated efficiently. This leads to a fast algorithm. However, if global updating is needed, for example if one were to use a globally updating method like Hamiltonian Monte Carlo method, then updating the Voronoi structure itself would probably become inefficient, especially for large numbers of cells.

Though in our simple example the computational cost is comparable with the standard 2-step McMC method, it still requires large computational resources. In McMC, a key factor affecting efficiency is the proposal mechanism. However, it appears to be difficult to construct efficient trans-dimensional proposals, since the natural ideas of closeness or proximity in fixed-dimensional proposals is no longer intuitive or necessarily true (Green & Hastie, 2009). In our method, we use the prior pdf for the trans-dimensional proposals (Dosso et al., 2014). However, some other efficient design might be used to further reduce the computational cost (Brooks et al., 2003; Ehlers & Brooks, 2008; Green & Hastie, 2009; Karagiannis & Andrieu, 2013). Another possibility is to use some parallel interacting Markov chains technique such as parallel tempering to increase the model mixing (Earl & Deem, 2005; Dettmer & Dosso, 2012; Dosso et al., 2012; Sambridge, 2013).

2.7 Conclusion

For the first time we implemented 3D fully non-linearised surface wave tomography directly from period-dependent travel-time measurements. We used the rj-McMC method and a parameterization based on 3D Voronoi tessellation. This method preserves the 3D horizontal and vertical spatial correlations in Earth properties and in uncertainties which are not preserved using other existing non-linearised methods. A synthetic test shows that the method provides better estimates of the velocity structure and of uncertainties than previous methods, and reproduces uncertainty loops around velocity anomalies in 3D as would be expected intuitively. It also does so at comparable cost to the standard 2-step Monte Carlo tomography method. This shows that our method is a valuable tool to investigate the shear-wave velocity structure of Earth. At least from the points of view of computation and accuracy of final uncertainty estimates, there seems to be little reason to persist in using 2-step method, although of course there can be other reasons to adopt it (ease of implementation and parallelization of the computation, familiarity etc.).

Chapter 3

1D, 2D and 3D Monte Carlo ambient noise tomography using a dense passive seismic array installed on the North Sea seabed

3.1 Summary

In the previous chapter we introduced a 3D fully nonlinear tomography method that inverts for shear velocities directly from frequency-dependent travel time measurements without using intermediate linearisation approximations, and which improves accuracy of the results and better estimate uncertainties. In this chapter we apply that method to real data and compare it to several of those previous methods. We extract Scholte waves from cross-correlations of 6.5 hours of ambient noise data recorded on a permanent monitoring system containing 3458

four-component sensors over Grane field, North sea. Dispersion analysis shows that two main Scholte wave modes are present. We separate the fundamental mode and the first overtone using a dispersion compensation method. Then for each mode, phase velocity dispersion curves are picked automatically and used to determine seabed phase velocity maps using Eikonal tomography. These phase velocity maps are used to study the shear-wave velocity structures of the subsurface. Usually independent one-dimensional depth inversions are performed beneath each geographical location, but this can introduce bias in the resulting 3D shear-velocity models. To further understand the possible limitation of the approach and to determine an accurate shear-velocity model, we applied three different methods to the Grane data: the usual 1D depth inversions, a 2D inversion along a vertical cross-section, and a fully 3D inversion. With each approach we estimate the shear-velocity structure along the same 2D cross-section and compared results. Thus we confirm that the 1D inversion method causes errors in the results due to independence of those 1D inversions, whereas the 2D and 3D inversions improve results by including lateral spatial correlations in the inversion. The 3D inversion bypasses the initial seabed Eikonal tomography step, and therefore avoids the errors it introduces into subsequent 1D and 2D inversions. The results image a clear low shear velocity river channel, and exhibit another low velocity feature both in the phase velocity maps at short periods (< 1.6s) of the fundamental mode, and in the shear-velocity model in the near surface ($<$ 250m). The latter anomaly is correlated with the distribution of seabed pockmarks, indicating that it may be caused by the circulation of near surface fluids. Such near surface low velocities might therefore be useful indicators of fluid leakage from subsurface reservoirs – particularly useful if ambient noise tomography was used to monitor subsurface fluid storage reservoirs, as might be necessary for CO_2 capture and storage scenarios.

3.2 Introduction

Geoscientists often need to image or monitor the subsurface in order to understand the properties and processes of the Earth's interior. Seismic tomography is a technique which has been used widely to produce three-dimensional maps of the properties of the Earth. In order to interpret the imaging results appropriately, and in particular to avoid over-interpretation, it is often desirable to estimate uncertainties in such maps.

Seismic surface waves propagate along interfaces in the Earth across which seismic properties change abruptly (Rayleigh and Love surface waves propagate along the Earth's surface, i.e. the interface between solid and air, whereas Scholte waves propagate along the seabed at the interface between liquid and solid), and oscillate over depth ranges that depend on frequency. This in turn makes surface waves dispersive – different frequencies travel at different speeds, and these speeds are sensitive to different parts of the Earth. By measuring the wave speeds this dispersion property can therefore be used to study the subsurface of the Earth by tomographic imaging on global (Trampert & Woodhouse, 1995; Shapiro & Ritzwoller, 2002; Meier et al., 2007a,b; Ferreira et al., 2010; Ekström, 2011) and regional scales (Zielhuis & Nolet, 1994; Curtis et al., 1998; Simons et al., 2002; Yang et al., 2007; Lin et al., 2008; Zigone et al., 2015).

In the above studies, surface waves were generated by earthquakes, which inevitably limits the resolvability of models in regions of sparse coverage due to the inhomogeneous distribution of seismic sources and stations. The introduction of ambient noise interferometry has greatly increased the size and coverage of our surface wave datasets by turning receivers into virtual (imagined) sources. It has been shown theoretically that Green's functions between different receiver pairs can be retrieved by cross correlations of ambient noise data recorded at the receivers (Campillo & Paul, 2003; Wapenaar, 2004; van Manen et al., 2005, 2006; Wapenaar & Fokkema, 2006; Curtis et al., 2006). Surface waves contained in the Green functions can be extracted and used to study subsurface structure (Shapiro & Campillo, 2004). This method has been used widely to study the regional scale structure of crust and uppermost mantle (Shapiro et al., 2005; Yao et al., 2006; Lin et al., 2007, 2009; Yang et al., 2007; Yao & Van Der Hilst, 2009; Bensen et al., 2009; Behr et al., 2010; Nicolson et al., 2012, 2014) and the near surface uppermost crustal structure (de Ridder & Dellinger, 2011; de Ridder & Biondi, 2013; Mordret et al., 2013a,b, 2014a; de Ridder et al., 2014, 2015; Allmark et al., 2018).

In most of the above studies only the fundamental mode surface wave is used due to the fact that higher modes are often so low in amplitude as to be invisible in the data. However, in surface waves generated by earthquakes it has been shown that in some situations higher modes can be observed (Gabriels et al., 1987; Park et al., 1999a,b) and can be used to further constrain the subsurface structure (Gabriels et al., 1987; Jan van Heijst & Woodhouse, 1999; Xia et al., 2000, 2003). Mordret et al. (2014b) also observed higher modes in the cross correlations of ambient noise data recorded by ocean bottom cable (OBC) sensors over the Valhall oil field, but since this energy was much weaker than the fundamental mode, these higher modes were ignored. However, in other cases the energy of higher modes can be comparable to the fundamental mode and may cause errors in inferred fundamental mode phase or group velocities due to the fact that the modes are mixed together (Xia et al., 2003). Therefore, it is important to correctly identify and separate the different modes.

If multiple source-receiver distances are available, higher modes can often be separated from fundamental modes by frequency-wavenumber (F-K) analysis (Gabriels et al., 1987). However, this process assumes that speeds of each frequency are the same for all source-receiver pairs. In order to perform phase or group velocity tomography, we often need accurate phase/group velocity measurements from each mode for each source-receiver pair independently. This means that individual modes need to be separated at each receiver. This can be done by band-pass filtering if the modes occupy different frequency bands (Crampin & Båth, 1965), but unfortunately this is not always the case. Other methods based on adaptive wavelet transformations (Kritski et al., 2006; Kuttig et al., 2006) or mode-branch stripping (van Heijst & Woodhouse, 1997) have been proposed to quantify the energy of individual modes and to separate them, but those methods risk destroying the phase of individual modes, and thus introducing errors in the phase velocity picks. In this study we therefore used a method based on dispersion compensation (Wilcox, 2003; Xu et al., 2012) or equivalently on time reversal (Fink, 1992; Alleyne et al., 1993; Ing & Fink, 1998). In this method, an individual mode in a dispersive wave is compressed to a short-duration pulse in the time domain by adjusting the phase to undo the dispersion so that each mode can be separated easily from the others e.g., by using a time-windowing function. The method has been used successfully to separate Lamb modes in ultrasonic waves (Xu et al., 2012), but has not previously been applied to seismic surface waves.

Seismic surface wave inversion problems are often solved using a 2-step scheme of first inverting for two-dimensional (2D) geographical maps of surface wave phase or group velocity and then inverting for the 3D spatial velocity structure using 1D inversions for structure over depth beneath each geographical location (Nakanishi & Anderson, 1983; Trampert & Woodhouse, 1995; Ritzwoller et al., 2002; Snoke & Sambridge, 2002; Bodin & Sambridge, 2009; Bodin et al., 2012; Galetti et al., 2017). The 2D tomographic problem in the first step is usually solved by a linearized procedure by minimizing the data misfit while applying some regularization (Trampert & Woodhouse, 1995; Ritzwoller et al., 2002).

However, the regularization is often chosen by ad hoc means (often trial and error), and it can suppress valuable information (Zhdanov, 2002). It has also been shown to be difficult to quantify meaningful uncertainties from linearized ambient noise tomography (Shapiro & Ritzwoller, 2002; Bensen et al., 2009; Yao & Van Der Hilst, 2009; Weaver et al., 2011; Nicolson et al., 2012, 2014). As a result the 1D depth inversions in the second step can be affected by biased 2D velocity and uncertainty estimation (Young et al., 2013).

To resolve these issues nonlinear inversion methods based on the Markov chain Monte Carlo (McMC) sampling algorithm have been introduced to seismic tomography (Mosegaard & Tarantola, 1995). McMC is a class of methods which generate samples from a target probability density (Metropolis & Ulam, 1949; Hastings, 1970; Mosegaard & Tarantola, 1995; Sivia, 1996; Malinverno et al., 2000; Malinverno, 2002; Malinverno & Briggs, 2004). In seismic tomography, a generalised McMC method called the *reversible jump* algorithm (Green, 1995; Green & Hastie, 2009) is often used; this allows a trans-dimensional inversion to be carried out, which means that the dimensionality of the parameter space (the number of model parameters) can vary in the inversion (Bodin & Sambridge, 2009; Hawkins & Sambridge, 2015; Piana Agostinetti et al., 2015; Burdick & Lekić, 2017; Galetti et al., 2017; Galetti & Curtis, 2018). In these methods a class of model parametrization is dynamically adapted to both the prior information and the data. The method has been used to estimate phase and group velocity maps of the crustal structure (Bodin & Sambridge, 2009; Zulfakriza et al., 2014; Galetti et al., 2015; Zheng et al., 2017) and to carry out the second depth-inversion step to obtain 3D shear wave velocity structures of the crust and uppermost mantle (Bodin et al., 2012; Shen et al., 2012, 2013; Young et al., 2013; Galetti et al., 2017).

However, In Chapter 2 we found that due to the independence of the many

1D inversions in the second step and possible phase or group velocity errors introduced in the first step, the 2-step method causes biases in estimated 3D shear-wave velocity models no matter whether linearized or McMC methods are used. We proposed an alternative 3D Monte Carlo method that directly inverts frequency-dependent phase or group traveltime measurements in one step, and showed via synthetic tests that the method improves accuracy of the velocity model estimation and produces more intuitively reasonable uncertainty than the traditional 2-step method. A similar idea has also been used with a linearized inversion method (Fang et al., 2015). In this study we apply the method to study the near surface structure of the Grane field and compare the results to those generated using previous methods.

The Grane oil field is situated in the North sea, about 185 km west of the city of Haugesund, Norway with a water depth of 127m (Figure 3.1) and contains heavy crude oil found in turbidite sandstone from the Tertiary period (Pragt et al., 2012). It was first discovered by Norsk Hydro in 1991 and is currently operated by Equinor ASA. It started production in 2003, which is estimated to last for 25 years. The field is composed of one main reservoir and a few other segments at a depth of 1,700 m. A permanent monitoring system has been deployed in the field, which contains 3458 four-component sensors (Z-vertical, N-north, E-east component and H-hydrophone). This records seismic data from the field continuously (Thompson et al., 2015) and thus provides the possibility to use ambient noise tomography to monitor the reservoir.

In the following we first present the ambient noise data and compute their cross correlations to obtain Scholte waves in section 2. In section 3 we briefly describe the dispersion compensation method and use it to separate the different Scholte wave modes. In section 4 we determine phase velocity maps for both the fundamental mode and the first overtone using Eikonal tomography. In section 5 we first review the standard 2-step method and the new 3D method, then apply them to estimate the shear-wave velocity structure over the Grane field. For the 2-step inversion, the phase velocity maps in section 3 are used as data. To further understand the limitation of independent 1D inversions, we also performed a 2D inversion using a 2D parameterization of a 2D cross-section using the same data as in the 1D depth inversions. We then compare all results across the 2D section. We conclude that the 3D Monte Carlo inversion method produces more realistic results, and achieves this with comparable computation cost compared to the standard 2-step Monte Carlo method.

3.3 Ambient noise interferometry at Grane field

3.3.1 Noise data recorded by the permanent monitoring system at Grane field

Figure 3.1 shows the locations of all 3458 sensors over the Grane field, each of which records samples at 500 Hz. The depth of the sea floor is around 127 m, becoming slightly shallower in the northeast (120 m). The sensors are organized along linear cables and consequently have inline and crossline spacings of approximately 50 m and 300 m respectively. This permanent and continuously recording array provides the potential to use passive seismic interferometry for daily monitoring of the field. In this study, we therefore analysed only 6 hours and 36 minutes of continuous data extracted from recordings on November 2014.

We analysed spectrograms of one-hour intervals of vertical component data extracted from those 6.5 hours of data. Figure 3.2a shows an example of one such dataset from one receiver. Figure 3.2b shows the spectrogram of energy

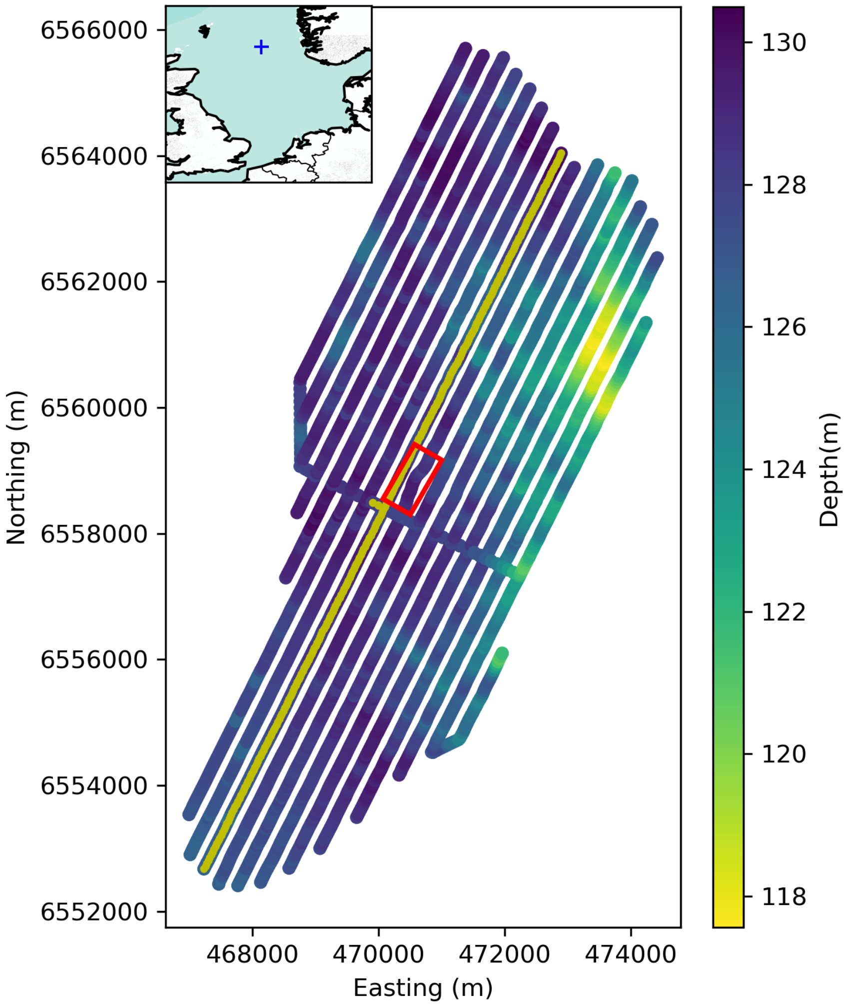


Figure 3.1: The distribution of receivers at the Grane field coloured according to their depths below sea level. The red box shows the location of the platform and the yellow line shows a receiver line used in the text. The blue plus in the inset map indicates the location of Grane field.

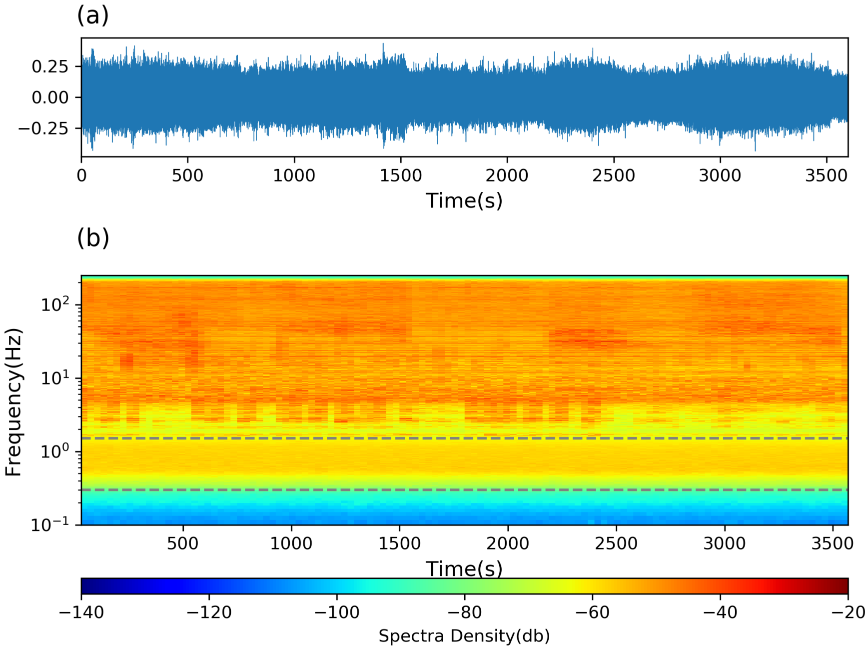


Figure 3.2: **(a)** An example of one-hour of vertical component data recorded by one of the geophones at Grane field; and **(b)** the corresponding spectrogram. The gray dashed lines bound the frequency range used for cross correlations.

across frequency as a function of time computed using 1 minute windows with 50% overlap between windows. The spectrogram is cut off below 0.3 Hz due to the roll-off in receiver sensitivity. The typical frequency response of the receivers is from 3 Hz to 15 KHz. However, it has been shown that lower frequency data can still be used for ambient noise studies (de Ridder & Dellinger, 2011; Mordret et al., 2013a, 2014a) Between 0.3 Hz and 1.5 Hz, the data are dominated by the tail of the secondary microseismic peak (Webb, 1998). This energy can be used to estimate Scholte waves from noise cross correlations. Above 1.5 Hz, field-operation noise sources and active seismic sources dominate the data. While it is also possible to obtain useful information about the subsurface using high frequency data (Mordret et al., 2013a), in this study we focus on the frequency band between 0.3 Hz and 1.5 Hz to obtain Scholte waves and use them to study the near surface structure.

The Scholte waves obtained from noise cross correlations can be biased in the case of an inhomogeneous distribution of noise sources (Wapenaar, 2004; Curtis & Halliday, 2010). Therefore we used the beamforming technique (Cole, 1995; Rost & Thomas, 2002) to characterize the noise distribution of Grane field. The recordings were first band-filtered into a narrow bands (e.g. 0.4-0.6 Hz) to avoid possible blurring of the beamforming results caused by velocity dispersion. The data were then transformed to $\tau - p$ domain by slant stacking (Yilmaz, 2001). Here we carried out beamforming analysis using half-hour segments for all receivers together. Figure 3.3 shows results for two frequency bands: 0.4-0.6 Hz and 0.6-0.8 Hz. Both results show two circles with different phase velocities, which are associated with different Scholte waves modes. The fundamental mode has a phase velocity of ∼580 m/s at the lower band (0.4-0.6 Hz) and a phase velocity of ∼520 m/s at the higher band (0.6-0.8 Hz) while the corresponding phase velocity of the first overtone is ∼910 m/s and ∼830 m/s respectively in each band. Note that the energy of either side of the 0.6-0.8 Hz result is spatial aliasing due to the cross-line

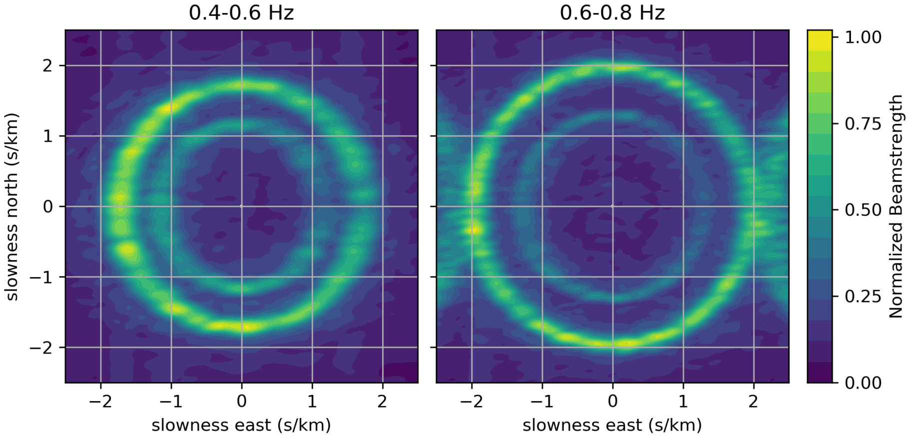


Figure 3.3: The beamforming results of two narrow frequency bands: 0.4-0.6 Hz (left) and 0.6-0.8 Hz (right) calculated using half-hour data segments.

sparsity of receivers (Yilmaz, 2001). Although there is residual inhomogeneity, for example, at the lower frequency band the energy in the west is slightly higher than in the east, in both cases the noise sources are nearly omnidirectional for both modes, which implies that we have sufficiently equidistributed noise sources for noise-based interferometry and tomography.

3.3.2 Cross-correlations of ambient noise

Seismic ambient noise interferometry refers to the construction of Green's functions from virtual sources by noise cross correlations (Campillo & Paul, 2003; Wapenaar, 2004; Curtis et al., 2006). The data are first bandpass-filtered to 0.35-1.50 Hz using a frequency domain taper, and down-sampled to 5 Hz sampling rate. Spectral-whitening is then applied to create data with a uniform amplitude spectrum. This process is carried out using half-hour recording segments with a 50% overlap. Finally the data from every station pair are cross correlated segment by segment and results are stacked over the 6.5 hour interval. We did

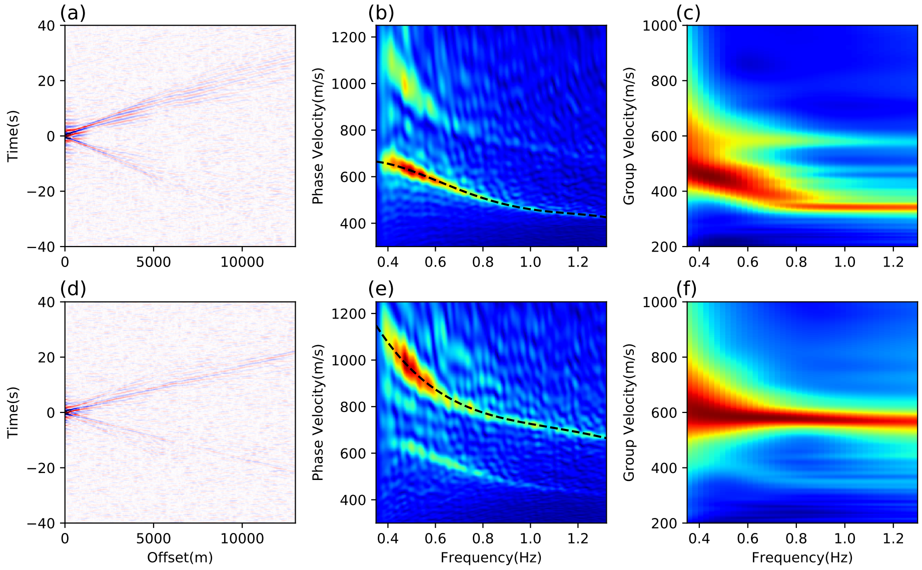


Figure 3.4: Examples of virtual shot gathers constructed using **(a)** vertical components and **(d)** hydrophone components. The receivers used are shown in Figure 3.1. **(b)** and **(e)** show the associated phase velocity dispersion analysis using $f - c$ analysis where c is phase velocity, and **(c)** and **(f)** show the associated group velocity dispersion analysis. The black dashed lines indicate the picked phase velocity for the fundamental mode and first overtone.

not remove instrument responses because they are identical for all receivers. In this study, we only obtained cross correlations using vertical (Z) components and hydrophone components (H) to construct Rayleigh-type waves, but it would be possible to construct Love waves using north (N) and east (E) components in future (Mordret et al., 2013a).

Figure 3.4a and 3.4d show virtual shot gathers along a receiver line indicated in Figure 3.1 (yellow line) constructed using pairs of vertical components and pairs of hydrophone components respectively. At long distances the wave packet spreads out in time due to dispersion. The negative time part has smaller energy than the positive time part, especially at long distances. This may be caused by some residual inhomogeneity in the distribution of noise sources (Figure 3.3), or some deviation of the noise sources from the ideal sources assumed in theory (mutually uncorrelated point sources).

To analyse the Scholte wave dispersion, we carried out frequency-phase velocity $(f - c)$ analysis for the virtual shot gathers in Figure 3.4a and Figure 3.4d. The $f - c$ spectrum $U(c, f)$ of gather $u(x, t)$ is computed using:

$$U(c, f) = \iint\limits_{-\infty}^{+\infty} u(x, t) e^{j2\pi f(t - \frac{x}{c})} \, dx \, dt \tag{3.1}$$

where x is distance along the virtual shot gather, t is time, f is frequency, c is phase velocity and $j = \sqrt{-1}$. For this research we used the symmetric part of the correlation results $u(x, t)$ (the mean of the positive and negative time results) after testing that this at least did not appear to be detrimental to results and appeared to increase stability of results. The results (Figure 3.4b, e) clearly show the two modes that we observed using beamforming analysis (Figure 3.3). Phase velocity varies from 420 m/s to 660 m/s for the fundamental mode and from 660 m/s to 1150 m/s for the first overtone. The fundamental mode dominates the signal in the vertical component cross correlations while in the hydrophone components the first overtone dominates, which has been observed before (Savage et al., 2013; Tomar et al., 2018) and has been shown to be related to a few hundred meters of low velocity sediments below the sea flow (Tomar et al., 2018).

We also analyse group velocity dispersion. Group velocities can be obtained by using the traditional frequency-time analysis (FTAN) method (e.g. Dziewonski et al., 1969; Levshin et al., 1972, 1992; Herrin & Goforth, 1977; Russell et al., 1988; Ritzwoller & Levshin, 1998; Levshin & Ritzwoller, 2001; Yanovskaya et al., 2012). For each seismic trace, a frequency-time domain envelope image can be obtained by applying a Hilbert transform for a set of narrow frequency bands (e.g. by narrow-band Gaussian filters). To estimate the group velocities for possible different modes, we stacked all those envelope images across the receiver line in Figure 3.1 to improve signal to noise ratio (Figure 3.4c, f). Similarly to the phase velocity dispersion analysis, the two modes can be observed in the stacked envelope image. The group velocity of the fundamental mode decreases from 480 m/s to 350 m/s from 0.35 Hz to 0.8 Hz and then shows very little variations after 0.8 Hz. By contrast, the group velocities of the first overtone varies only slightly from 610 m/s to 590 m/s over the frequency range from 0.35 Hz to 1.3 Hz.

3.4 Mode separation

Although it is often the case that only one mode dominates the signal in a wave (the energy of the first overtone is usually lower than the fundamental mode in vertical component recordings - Figure 3.4b, c), energy from other modes will still cause bias in the phase or group velocity of each seismic trace (e.g., Xia et al., 2003). For example, at low frequencies ($<$ 0.5 Hz) the group velocities of the two modes are very close and probably cannot be identified individually easily,

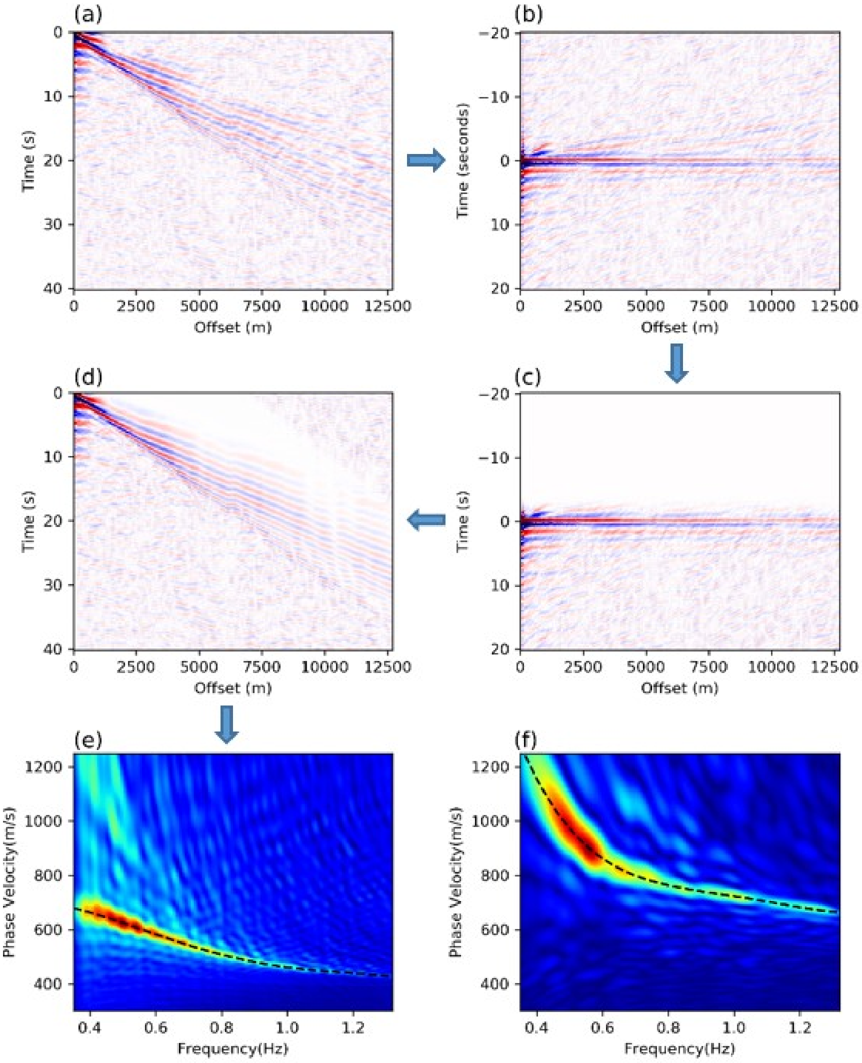


Figure 3.5: An example of the mode separation procedure. **(a)** The virtual shot gather before mode separation (obtained from Figure 3.4a by adding positive and negative times). **(b)** Flattened virtual shot gather obtained by dispersion compensation. **(c)** Filtered flattened virtual shot gather. **(d)** Virtual shot gather after mode separation. **(e)** Phase velocity dispersion analysis after mode separation. **(f)** Phase velocity dispersion analysis after using a similar method to instead isolate the first overtone. Arrows show the order of processing steps. Black dashed lines show the associated phase velocity dispersion curves.

which may cause biases in phase or group velocity picking (see below). This could produce errors in subsequent tomography results. Therefore, in order to obtain accurate phase or group velocity picks for each trace and for each mode, we first need to separate the modes.

3.4.1 Method

We use a dispersion compensation method to separate different modes (Xu et al., 2012). For seismic surface waves we assume that two modes S and A are excited by a broadband source excitation $F(w)$, and that the phase velocity of the modes are constant across the area. In the frequency domain the recorded surface wave at distance x_0 can then be expressed as

$$G_{SA}(w) = [Amp_S H_S(w) + Amp_A H_A(w)]F(w) \tag{3.2}$$

where $H_S(w) = exp(-jk_S(w)x_0)$ and $H_A(w) = exp(-jk_A(w)x_0)$, $k_S(w)$ and $k_A(w)$ are dispersion relations of mode S and mode A respectively, and Amp_S and Amp_A are their amplitudes. If $k_S(w)$ and $k_A(w)$ are known, the process of dispersion can be reversed by back-propagation or dispersion compensation. For example, multiplying equation (3.2) by $H_S^{-1}(w) = exp(jk_S(w)x_0)$ gives

$$\begin{aligned} G^*_{SA}(w) &= H_S^{-1}(w)G_{SA}(w) \\ &= Amp_S F(w) + Amp_A H_S^{-1}(w)H_A(w)F(w) \end{aligned} \tag{3.3}$$

where $G^*_{SA}(w)$ is the surface wave after dispersion compensation. The first term $Amp_S F(w)$ has no dispersion (propagation) term so will be focused to a short-duration pulse $f(t)$ in time domain at zero time, which can be extracted using a window function. The second term is the residual term from mode A which is still dispersive. After separating mode S from the other waves, it can be propagated back to distance x_0 by multiplying the extracted signal by $H_S(w)$. The result then can be used to estimate more accurate phase or group velocities for mode S, and a similar operation can be used to extract mode A.

In practice, the dispersion relations $k_S(w)$ or $k_A(w)$ are usually not constant across the area of study. However, they often vary smoothly across space at least locally. So equation (3.3) can still be used for approximate dispersion compensation. Alternatively, one can integrate along estimated source to receiver ray path to estimate the correct dispersion to each receiver if an approximate estimate of the phase velocity map is known. This dispersion estimate can then be used for $k_S(w)$ or $k_A(w)$ as appropriate. In this study, we simply applied equation (3.3) using estimates of dispersion relations (picked from an initial frequency-wavenumber analysis such as that shown in Figure 3.4).

3.4.2 Application to Grane field

We used the method above to separate different modes across Grane field. The data were first back-propagated to time zero in the frequency domain using equation (3.3) using the dispersion relation of the fundamental mode picked using $f - c$ analysis (black dashed line in Figure 3.4b), and then transformed back to the time domain. Figure 3.5b shows the results after dispersion compensation for the symmetric part of the virtual shot gather in Figure 3.4a. The symmetric gather is shown in Figure 3.5a. After dispersion compensation, the fundamental mode focuses to an impulsive signal at zero time so that energy in the shot gather becomes flat. Those waves that exist at negative time and are not flat are higher modes since higher modes generally travel faster than the fundamental mode. They can be muted using a time-domain window function (Figure 3.5c). Figure 3.5d shows the shot gather after windowing out the higher modes and transforming back to the original propagation time. Figure 3.5e shows the dispersion image obtained from $f - c$ analysis using the virtual shot gather after mode separation (Figure 3.5d). At most frequencies (0.5 - 1.3 Hz) the energy of the first overtone disappears, indicating that the higher modes have been removed successfully. However, at low frequencies (0.35-0.5 Hz) some energy from the first overtone remains. This can be explained by the fact that at these frequencies (0.35 - 0.5 Hz) the group velocities of the two modes are very close (Figure 3.4c) which leads to the modes overlapping even after back-propagation to time zero.

Similarly to the separation of the fundamental mode, we apply the method to obtain waves containing higher modes only. Since the first overtone dominates the signal in cross correlations of hydrophone components (Figure 3.4e, f), we used those cross correlations to retrieve the first overtone. The phase velocity dispersion curve picked from the result of $f - c$ analysis (Figure 3.4e) is used to carry out the phase correction. The fundamental mode is then windowed out using a window function and consequently waves containing only the first overtone are obtained (Figure 3.5f). Finally, after mode separation we obtain Scholte waves that contain only the fundamental mode or the first overtone, which can be used to pick accurate phase or group velocities for each mode.

3.5 Phase velocity tomography

3.5.1 Method

To perform phase or group velocity tomography, we first need to pick phase or group velocities for each cross correlation between each station-pair. There are so many pairs that this process must be automated. Group velocity can be picked using the FTAN method (see description in section 2.2). For phase velocity picking we used an image transformation technique (Yao et al., 2006). First a time-period image ($t - T$) is constructed by applying a set of narrow-band filters to the data. This can be transformed to a velocity-period image ($c - T$) by transforming time to phase velocity, for the moment assuming a straight-ray path between each station pair (Figure 3.6). Finally the phase velocity dispersion curve can easily be identified and automatically picked on the $c - T$ image. The 2π ambiguity of phase velocity measurements can be resolved using our initial estimate of the average phase velocity dispersion curve obtained using $f - c$ analysis (black line in Figure 3.4b; white line in Figure 3.6).

To improve the quality of dispersion data, we applied a series of data selection criteria. A minimum signal to noise ratio (SNR) of 5 is used for the fundamental mode and 2.5 is used for the first overtone. The SNR is calculated using the spectrum of the signals of interest and the spectrum of an interval of noise extracted from the end of the virtual source records. Due to the far-field approximation that is implicit in the ambient noise interferometry method, those station-pairs whose distances are smaller than twice the wavelength at any frequency are discarded (Yao et al., 2006; Lin et al., 2009). Considering the possible biases introduced by mode separation in the frequency range of 0.35 - 0.5 Hz (2 - 2.85 s period) – see Figure 3.5e – in phase velocity tomography we only used phase velocities at frequencies larger than 0.5 Hz ($<$ 2 s period).

We picked phase velocities and group velocities for each station-pair, which can then be used to perform phase or group velocity tomography. Group velocity tomography can be conducted using straight-ray tomography since the data accuracy usually does not merit a more sophisticated approach and since an accurate phase velocity map is not available in order to trace rays to allow group velocity to be calculated along rays (de Ridder & Dellinger, 2011; Mordret et al.,

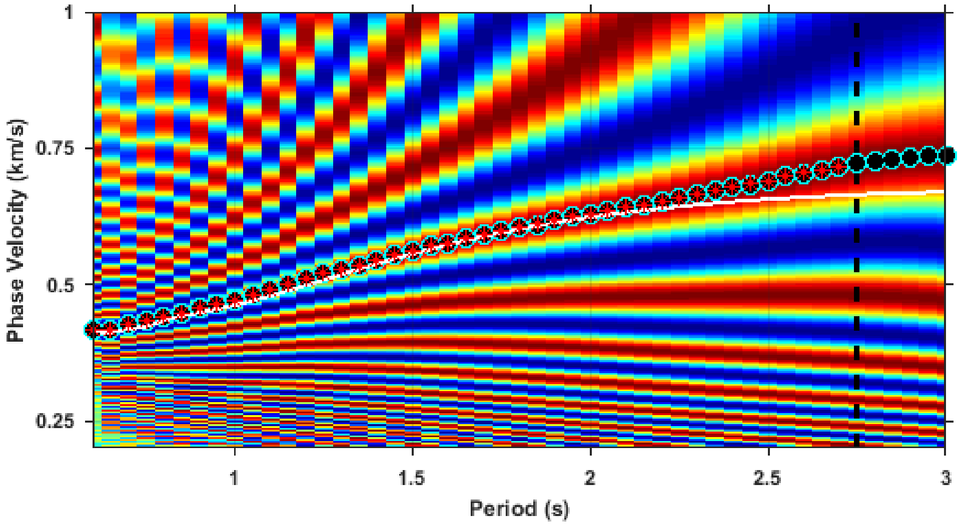


Figure 3.6: An example of the $c-T$ image used to pick phase velocities. The dashed black line shows the maximum period allowed by the far-field approximation (the offset must be larger than twice the wavelength). The black dots denote the picked phase velocity for the whole period range and the red stars show phase velocities that are actually used. The white line shows the phase velocity dispersion curve obtained using $f-c$ analysis.

2013a; Allmark et al., 2018). However, since that the phase velocity measurements are more accurate than group velocities (Yao et al., 2006) and since Eikonal tomography is more efficient and more accurate than straight-ray tomography (Lin et al., 2009), we performed phase velocity tomography using the Eikonal method as we now describe.

In a smoothly heterogeneous medium, the propagation of a single surface wave mode can be expressed using the Eikonal equation (Aki & Richards, 1980; Biondi, 1992; Wielandt, 1993; Shearer, 1999):

$$\frac{1}{c_i(w,\mathbf{r})^2} = |\nabla\tau_i(w,\mathbf{r})|^2 - \frac{\Delta A_i(w,\mathbf{r})}{A_i(w,\mathbf{r})w^2} \tag{3.4}$$

where c is the phase velocity, τ is the travel time, A is the spectral amplitude, w is the angular frequency, $\mathbf{r}$ is the location, and subscript i denotes the i^{th} source. If the second term on the right-hand side can be ignored (see discussion in Lin et al. 2009), this equation becomes:

$$\frac{1}{c_i(w,\mathbf{r})} = |\nabla\tau_i(w,\mathbf{r})| \tag{3.5}$$

In this case the local phase slowness at location $\mathbf{r}$ is simply related to the magnitude of the gradient of the travel time field. Therefore, the local phase velocity can be determined using equation (3.5) by calculating the gradient of the travel time field from each virtual source (Lin et al., 2009).

In order to determine the gradient of the travel time field, we first interpolate the travel time field to a regular grid (Lin et al., 2009). Here we used the biharmonic spline interpolation to regularize the field to a $50m \times 50m$ grid (de Ridder & Dellinger, 2011). To better control the quality of the interpolated field, for each location we only use interpolated travel times that are surrounded by four measurements. Due to small SNR at large offsets we discarded measurements whose offsets are greater than 6km.

Finally, the average phase slowness $s_0(x)$ at location x and its standard deviation $\sigma_{s_0}(x)$ can be computed using all virtual sources by

$$s_0(x) = \frac{1}{n}\sum_{i=1}^{n} s_i(x) \tag{3.6}$$

$$\sigma_{s_0}^2(x) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(s_i(x) - s_0(x))^2 \tag{3.7}$$

where n is the number of sources and i denotes the i^{th} source. Thereafter the phase velocity c_0 and its uncertainty σ_{c_0} can be determined by:

$$c_0(x) = \frac{1}{s_0(x)} \tag{3.8}$$

$$\sigma_{c_0}(x) = \frac{1}{s_0(x)^2}\sigma_{s_0(x)} \tag{3.9}$$

3.5.2 Results

We applied Eikonal tomography for both the fundamental mode and the first overtone in the period range 0.7 s to 2 s. The final mean phase velocity map and its standard deviation are averaged over the 3458 sources for each period. Figure 3.7 shows the results of fundamental mode tomography at periods 0.7 s, 1.0 s, 1.3 s and 1.6 s. At short periods (< 1.6 s), there is a clear low velocity anomaly in the middle of the field (location delineated by the blue solid line box in the 1s period map). At the west edge between Y=8 km and Y=10 km, a clear low velocity anomaly exists at 0.7 s which fades out from 1.0 s to 1.3 s. At 0.7 s a low velocity channel connects this low velocity anomaly to the middle low velocity anomaly.

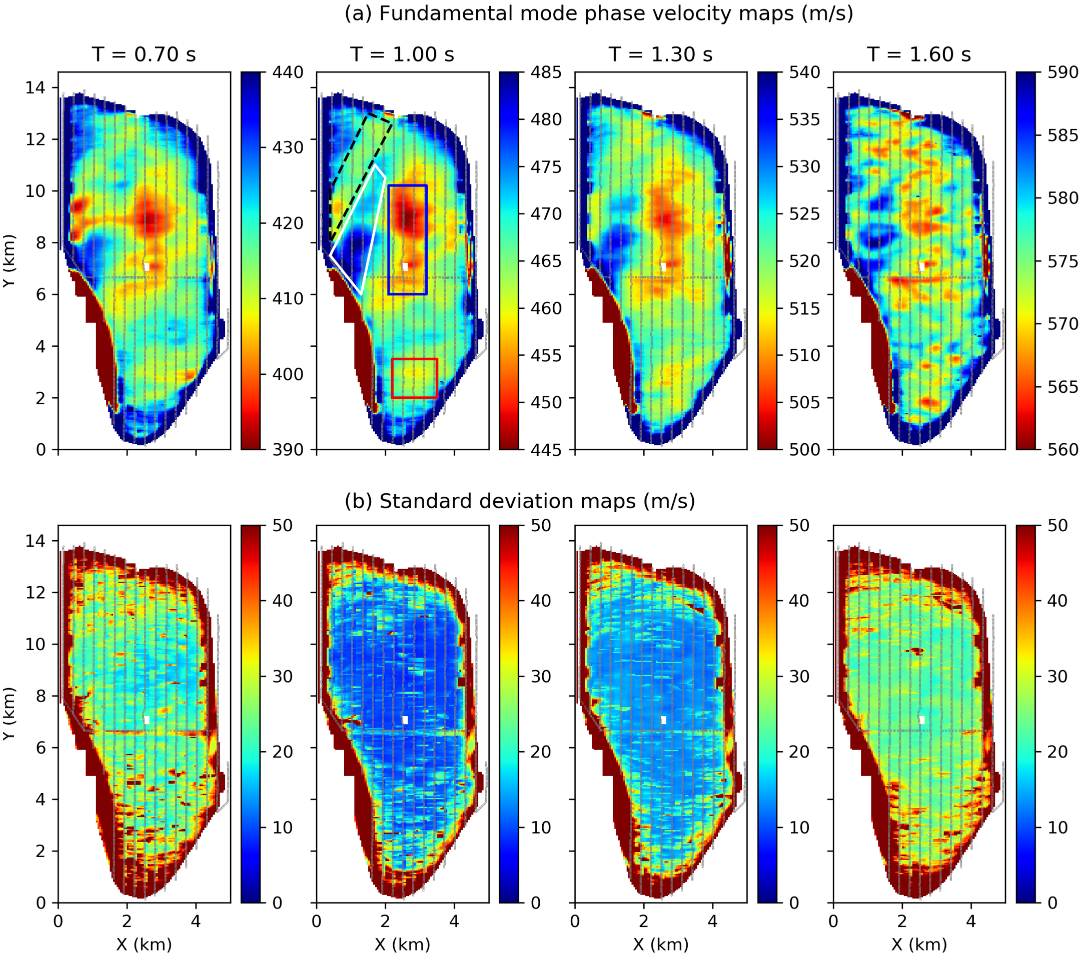


Figure 3.7: **(a)** Phase velocity maps of the fundamental mode and **(b)** their associated standard deviation maps at periods 0.7s, 1.0s, 1.3s and 1.6s. The boxes show locations of features discussed in the text. Gray lines show the distribution of receivers.

However, it disappears at longer periods (> 0.7 s), which indicates that this is probably a very near surface structure. Instead at periods of 1 s and 1.3 s, there is a high velocity channel from middle west to north east (location delineated by white solid line box in the 1s period map) which is interrupted by the low velocity channel at 0.7 s. Next to this high velocity channel there is a parallel low velocity channel (location delineated by the black dashed line box) existing at short periods (< 1.6 s). In the south, a low velocity anomaly also emerges (location delineated by red solid line box). However, at period 1.6 s though there seems to be some degree of similarity in structures with the shorter periods ($<$ 1.6 s), the phase velocity map becomes more complicated. This might due to the poor data quality at longer periods, or may be caused by complex structure at greater depths.

Overall the uncertainties are low at periods of 1.0 s and 1.3 s ($\sim 10m/s$) and are higher at periods of 0.7 s and 1.6 s ($\sim 25m/s$). The high uncertainties at 0.7 s are probably caused by the filtering taper at the lower period side (0.67 s - 0.74 s) while the high uncertainties at 1.6 s probably indicate higher data uncertainties due to lower SNR since phase velocities at longer periods must usually be measured at longer offsets which may also partly explain the complex structure in the mean phase velocity map at 1.6 s. Close to the boundaries all standard deviation maps show very high uncertainties caused by limited data coverage. At the location of the middle low velocity anomaly (blue box), the standard deviation map at 0.7 s shows relatively lower uncertainties; this suggests that the middle low phase velocity anomaly is probably caused by a low velocity structure near to the surface.

To better understand the phase velocity maps, we compared the phase velocity map at 0.7s and 1.0 s with the distribution of pockmarks at the seabed of Grane field (Figure 3.8). Pockmarks are craters in the seabed which have been shown to be related to the seepage of fluids (gas or liquids) (Kvenvolden, 1989). Figure

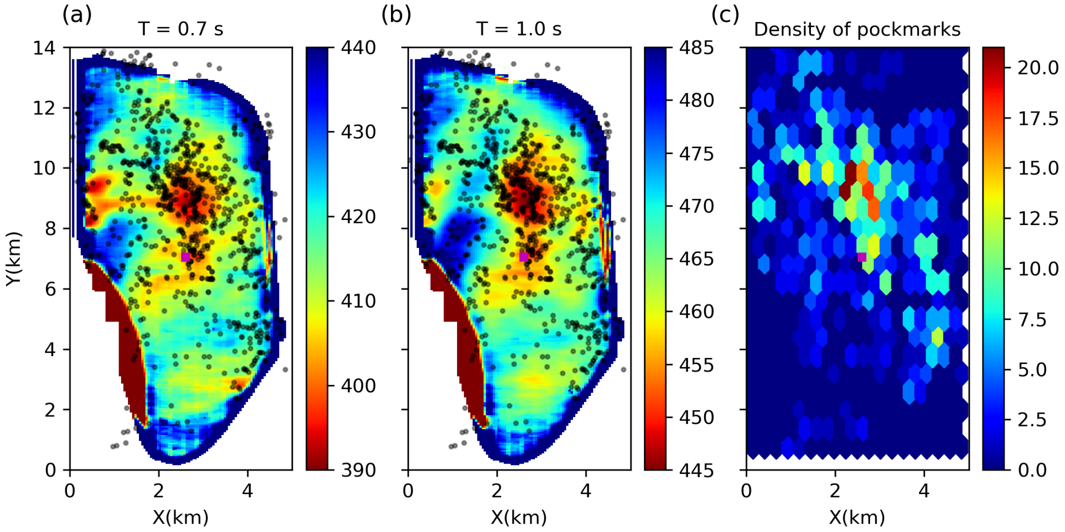


Figure 3.8: Phase velocity maps at (a) 0.7 s and (b) 1.0 s plotted with pockmarks (black dots). (c) The density of pockmark distribution. The magenta rectangle shows the location of the platform.

3.8 shows that there is a dense distribution of pockmarks at the location of the middle low velocity anomaly from the platform to Y=11 km, which suggests that the low velocity anomaly might be caused by near surface fluids. At the west edge, the two low velocity anomalies between Y=8 km and Y=10 km are also consistent with a higher density of pockmark distribution.

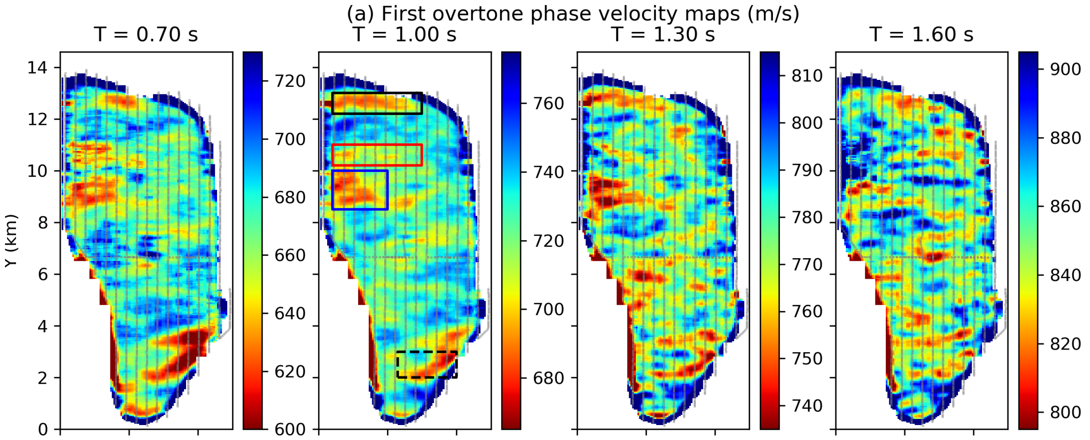
Figure 3.9a shows the mean phase velocity maps of the first overtone at the same periods as for the fundamental mode. We again outline some noticeable features in the phase velocity map at 1 s period. At short periods (< 1.6 s), there is a low velocity anomaly at the west edge (blue line box), with a different shape to that observed in the fundamental mode phase velocity maps. This might suggest that this anomaly is caused by a change in shape with depth since higher mode phase velocities usually have higher sensitivities at greater depths compared to fundamental modes. To the north of this low velocity anomaly there is a low velocity channel at periods of 0.7 s and 1.0 s (red solid line box). At the northern edge, a low velocity channel crosses the field from west to east (black line box). Similarly, this low velocity channel cannot be clearly observed on the fundamental mode phase velocity maps, which indicates that it might be related to deeper structure. In the south of the field there is a similar low velocity anomaly as observed in the fundamental mode phase velocity maps (black dashed line box), which may indicate a consistent low velocity structure from shallow to deeper levels. Overall, at longer periods (i.e. 1.3 s and 1.6 s) the phase velocity maps show very complicated structures as we have seen in the fundamental mode phase velocity map at period of 1.6 s. This may suggest a complicated deeper structure, or may simply be due to the low quality of data at longer periods. Note that the phase velocity maps of the first overtone exhibit much shorter scale structure compared to the fundamental model because of lower data quality of the first overtone. When these phase velocities are used to invert for shear velocities, the shorter scale structure may cause lack of coherence between adjacent 1-D models over depth. To reduce this issue, instead of using Eikonal tomography a regularised inversion might be used to estimate spatially smoother phase velocity maps.

Overall the standard deviation maps of the first overtone show higher uncertainties compared to the fundamental mode (Figure 3.9b) due to the fact that the SNR of the first overtone is lower than that of the fundamental mode. Similarly to the fundamental mode, higher uncertainties are observed at periods of 0.7 s and 1.6 s ($\sim 100m/s$) than at periods of 1 s ($\sim 50m/s$) and 1.3 s ($\sim 80m/s$). The uncertainties are significantly smaller between Y=8 km and Y=10 km at periods of 0.7 s and 1.6 s, showing that this area is well determined, so the low velocity anomaly at this area (blue line box) is well determined. Similarly, there is low uncertainty at the north edge associated with the low velocity channel (black line box) and low uncertainty in the south associated with the low velocity anomaly (black dashed line box). The standard deviation map at 0.7 s shows some higher uncertainty areas ($\sim 160m/s$), e.g., the western edge between Y=10 km and Y=12 km and between Y=6 km and Y= 8 km, which is probably caused by low resolution of those areas. Similarly, there is a high uncertainty area between Y=10 km and Y=12 km at the west edge at period of 1.6 s.

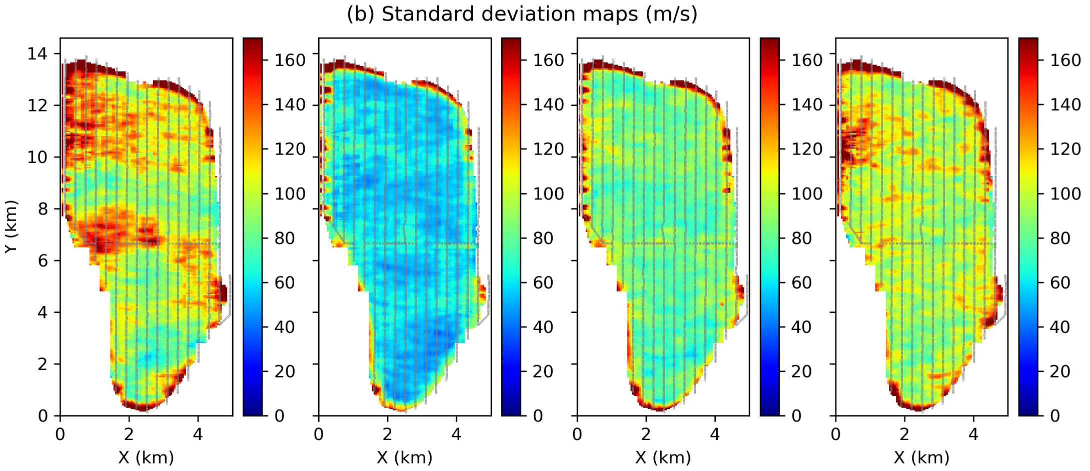
3.6 Shear-wave velocity inversion

Although these phase velocity maps can be interpreted for useful information about the subsurface, such maps cannot provide good indications of the depths of observed structures since Scholte wave phase velocities are a consequence of the velocity structure over a range of depths. In order to better understand the subsurface structure it is necessary to estimate subsurface shear-velocity structures with depth in a separate inversion. Traditionally, a two-step inversion scheme is used to invert for shear-velocity structures where we use the above phase velocity maps as data and perform 1D depth inversions independently beneath each geographical location. However, In Chapter 2 we used synthetic data to show that such a scheme introduces biases in the final 3D shear-velocity structure because each of the depth inversions is conducted independently, whereas in reality they are strongly correlated spatially. We therefore proposed a fully 3D Monte Carlo inversion method using a 3D parameterization which preserves these correlations.

To further understand the limitations of traditional two-step inversion schemes, in this section we compare results from the two-step method and the 3D method on real data. We use Markov chain Monte Carlo (McMC) to perform both the 1D depth inversions and the 3D inversion. To limit the computational cost, we only carried out 1D depth inversions along a 2D cross-section (the yellow line in Figure 3.1). In order to study the effects of independent 1D inversions, we also carried



(a) First overtone phase velocity maps (m/s)



(b) Standard deviation maps (m/s)

Figure 3.9: **(a)** Phase velocity maps of the first overtone and **(b)** their standard deviation maps at periods 0.7s, 1.0s, 1.3s and 1.6s. The boxes in the 1 s period map show locations of features discussed in the text. Gray lines show the distribution of receivers.

out a 2D depth inversion along this cross-section using a 2D parameterization and the phase velocities along the profile on the above maps as data. This inversion is of interest because it uses the phase velocity maps as data similarly to the 1D inversion, but preserves spatial correlations similarly to the 3D method. It would therefore be expected a priori to exhibit intermediate performance compared to the other two methods. In this section, we first describe the two-step methods and the 3D method, and then give an overview of the reversible-jump McMC method and the parallel tempering method (which is used to improve computational efficiency of McMC method). We then apply those methods to the Grane data and compare their results.

3.6.1 Methods

Parameterization

As described in Chapter 2, we use Voronoi tessellations to parameterize the subsurface. A Voronoi cell is defined by a point (called a site) and its volume that consists of all of the points nearer to this site than to any other. Figure 2.1 shows examples of Voronoi tessellations in 1D, 2D and 3D. Each cell contains its location and its properties (e.g., P-wave velocity, shear-wave velocity, density, etc.). Note that in 1D the parametrization with Voronoi cells is inferior to the parametrization with a simple partition model since the same velocity model can be obtained using different configuration of Voronoi cells Green (1995). However, for comparison purpose in this study we still use a 1D Voronoi parametrization. Since seismic surface waves are primarily sensitive to subsurface shear-wave velocity variations, we only invert for shear-wave velocities. P-wave velocity is linked to the shear-wave velocity via an empirical relation (Castagna et al., 1985):

$$V_p = 1.16V_s + 1.36 \tag{3.10}$$

and density is computed from the P-wave velocity empirically (Brocher, 2005):

$$\rho = 1.74V_p^{0.25} \tag{3.11}$$

where V_p and V_s are in km/s, and density rho is in g/cm^3. Similar to Zhang et al. (2018), within each Voronoi cell the velocity is spatially constant.

The specific choice of Voronoi parametrization makes it easy to implement in 1D, 2D and 3D for comparison. At any point in the model, a velocity profile is a layered model whose phase or group velocity dispersion curve can be computed using many available codes (e.g. Herrmann, 2013) without resorting to approximations. However, while the Voronoi parametrization is good at recovery of discontinuities, it can introduce difficulties to recover a smooth model (Hawkins et al., 2019). The scale length of Voronoi cells in higher dimensions can cause models to be distorted and some ad-hoc rescaling is generally required (Zhang et al., 2018). It has also been found that Voronoi parametrization can produce multi-modalities in the posterior which makes interpretation of uncertainties difficult. In these cases one could try some other parametrizations, such as wavelets (Hawkins & Sambridge, 2015; Dettmer et al., 2016), Johnson-Mehl tessellation (Belhadj et al., 2018) and Delaunay and Clouth-Tocher parametrization (Hawkins et al., 2019).

2-step inversion

The shear-wave velocity structure is estimated using a 2-step scheme as follows. In the first step a series of 2-D phase or group velocity maps for different frequencies are estimated tomographically using source-to-receiver or inter-receiver arrival times as data; then at each geographical point, the local dispersion curve is used to invert for a 1-D shear velocity profile beneath that point. For the first step, either linearised (Nakanishi & Anderson, 1983; Trampert & Woodhouse, 1995; Ritzwoller et al., 2002; Snoke & Sambridge, 2002; Nicolson et al., 2012, 2014) or non-linearised methods (Bodin & Sambridge, 2009; Bodin et al., 2012; Khan et al., 2013; Young et al., 2013; Rawlinson et al., 2014; Zulfakriza et al., 2014; Saygin et al., 2015; Galetti et al., 2015, 2017; Zheng et al., 2017) can be used to estimate phase or group velocity maps. In this study, since we have a very dense station network, we used Eikonal tomography to determine phase velocity maps (see above).

For the second step, we use a non-linear McMC method to invert for the 1-D shear velocity profile beneath each point (Bodin et al., 2012; Young et al., 2013; Galetti et al., 2017). Generally those 1D depth inversions are run independently at each geographical location without interaction, as this allows perfect parallelisation of what is a computationally demanding task. We also carried out a 2D depth inversion along the 2D cross-section by using the 2D parameterization described in Figure 2.1b so as to include lateral spatial correlations in the inversion. The data used for the 2D inversion is the same as those used in the 1D inversions (the local phase velocities from Eikonal tomography). For both the 1D and 2D inversions, we used the same forward modelling method, a modal approximation method (Herrmann, 2013), to calculate the phase velocity dispersion curves from the velocity-versus-depth profiles beneath each geographical point.

Fully 3D inversion

In order to determine a 3D shear-velocity model and to be able to compare the three different methods (1D, 2D and 3D inversions), we also performed a 3D inversion using the 3D McMC method described in Chapter 2. The subsurface is discretized by Voronoi cells (Figure 2.1c), each of which is defined by the location of its site and its shear-wave velocity. The forward modelling method is an approximate 2-step method (Ritzwoller & Levshin, 1998; Stevens et al., 2001; Reiter & Rodi, 2008): first a series of phase or group velocity maps at each measurement period are computed by extracting the shear velocity profile with depth beneath each geographical point to what the 1D modal approximation method of Herrmann (2013) is applied to predict group and phase velocities at each period; then for each source-to-receiver or inter-receiver pair, the travel times for each period can be determined by tracing rays through the computed phase velocity map (for which we use the fast marching method – Rawlinson & Sambridge 2004).

However, as shown in Galetti et al. (2017), modal approximation methods that are usually used (Herrmann, 2013; Saito, 1988) produce unrealistic dispersion curves when applied to relatively unusual velocity-depth models. This is due to the fact that these methods solve the period equation for the minimum phase velocity solution; unfortunately when the top layer does not have the lowest shear-wave velocity, the dispersion curve with minimum phase velocity is likely to be one of the trapped modes generated by a low velocity layer at depth (Chen, 1993; Wu & Chen, 2016). These trapped modes generally oscillate within the low velocity layer, meaning that they cannot actually be observed on Earth's surface and hence do not correspond to forward model using the recorded data. Therefore, in order to make the modal approximation modelling package that we used (Herrmann, 2013) valid for our inversion, we added a prior constraint on our models – that

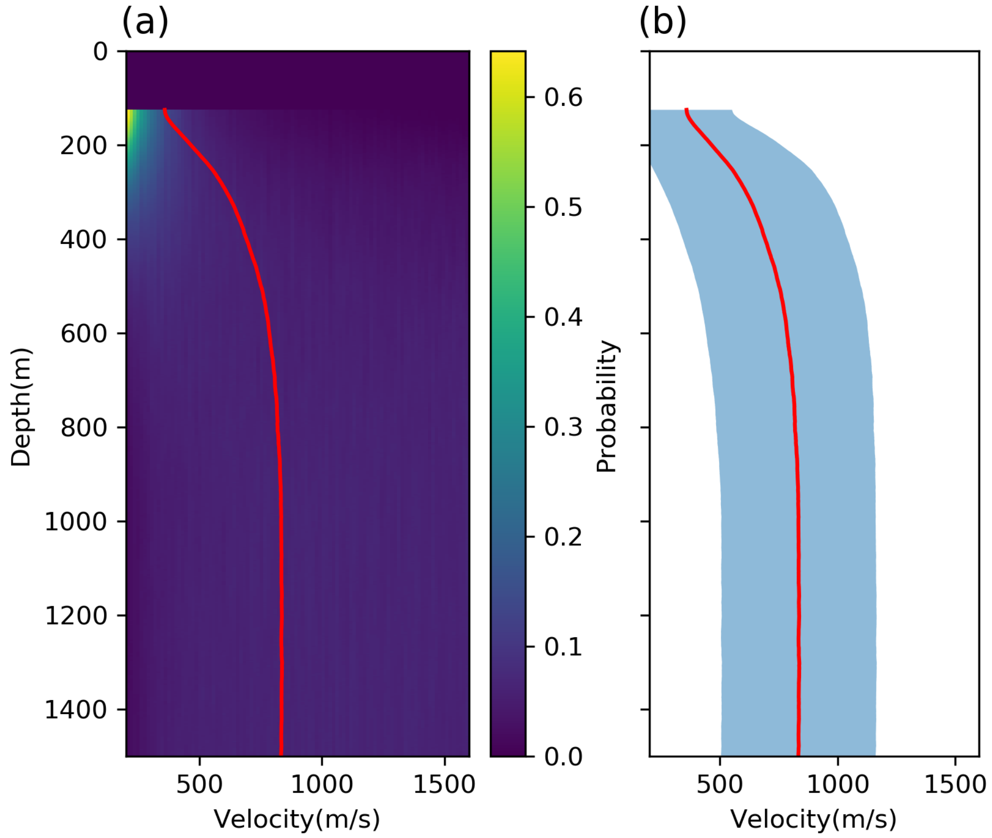


Figure 3.10: **(a)** The marginal prior probability of a 1D example and **(b)** the mean model (red line) and standard deviations (blue area) obtained using reversible jump Markov chain Monte Carlo algorithm. The red line shows the mean model.

the smallest shear-wave velocity must be in the top layer. The prior is achieved by rejecting any proposals of violating models in the Markov chain using a large penalty. Note that this choice of prior causes shear velocities at the near surface to prefer small values (Figure 3.10). Considering that this is generally thought to be true for most of the real Earth, we feel that this is an acceptable and pragmatic solution.

Parallel tempering

To improve efficiency of the 3D Monte Carlo method we apply parallel tempering to Markov chains. Parallel tempering is a technique that mixes information between parallel tempered Markov chains to improve efficiency of McMC methods (Earl & Deem, 2005; Dosso et al., 2012; Sambridge, 2013). First, a set of chains are scaled using different temperatures, such that their target probability can be denoted as:

$$\pi(\mathbf{m}|T_i) = p(\mathbf{m}|\mathbf{d})^{1/T_i}, \tag{3.12}$$

where T_i is the i^{th} temperature, $p(\mathbf{m}|\mathbf{d})$ is the posterior probability density and $\pi(\mathbf{m}|T_i)$ is called the tempered posterior pdf. Those tempered Markov chains are then run in parallel. Models can be swapped between chains randomly based on an acceptance ratio called detailed balance (Earl & Deem, 2005; Sambridge, 2013):

$$\alpha(i,j) = min\{1, \left[\frac{p(\mathbf{m}_j|\mathbf{d})}{p(\mathbf{m}_i|\mathbf{d})}\right]^{1/T_i}\left[\frac{p(\mathbf{m}_i|\mathbf{d})}{p(\mathbf{m}_j|\mathbf{d})}\right]^{1/T_j}\}, \tag{3.13}$$

where $\alpha(i,j)$ is the acceptance ratio of a swap between model $\mathbf{m}_i$ and $\mathbf{m}_j$ at temperature T_i and T_j, respectively. By doing so, one can sample the combined posterior distribution $\pi(\mathbf{m}|T_i), (i = 1, ..., n)$. At higher temperatures the posterior probability density function becomes flatter, which improves the ability of McMC to escape local minima and to explore parameter space more globally. By enabling exchange between different temperatures, the method thus improves the explorative performance of the Markov chain at $T = 1$, which (still) samples our target posterior probability.

The choice of temperature ladder of the parallel chains strongly affects the efficiency of parallel tempering. It has been shown that a power-law temperature schedule is generally more efficient than a uniformly distributed schedule (Calderhead & Girolami, 2009; Sengupta et al., 2015), so here we used a power-law schedule. Given a total of N chains, the temperatures can be distributed as:

$$1/T_i = 1 - \left(\frac{i}{N}\right)^p, \tag{3.14}$$

where T_i is the i^{th} temperature and p is the power coefficient which can be chosen accordingly. To reduce the overhead introduced by synchronization and communication of parallel chains, we only swap models every 50th iteration.

Usually only samples from those chains with $T = 1$ are stored for later inference (Sambridge, 2013; Ray et al., 2017; Galetti & Curtis, 2018). However, for sophisticated problems, parallel tempering often demands a large number of tempered chains to improve the efficiency of McMC methods, which leads to a high computational cost. Those chains with $T > 1$ can also be used for Bayesian inference via an importance resampling scheme (Geyer, 1994; Dosso et al., 2012). Suppose that we have samples from an unnormalized density h and want to calculate an integration with respect to another unnormalized density h_θ. This is important since most of the statistics that we usually wish to calculate (the mean model, variance, etc.) are integrals. Such integrals can be computed by using the importance sampling formula as weighted averages:

$$E_\theta g(X) = \sum_{j=1}^{n} w_\theta(X_j) g(X_j) \tag{3.15}$$

where θ denotes that the expectation is calculated with respect to probability density h_θ, g is the function of which we want to calculate expectation, and

$$w_\theta(x) = \frac{h_\theta(x)/h(x)}{\sum_{j=1}^{n} h_\theta(X_j)/h(X_j)} \tag{3.16}$$

Substituting $h_\theta(x)$ with $\pi(\mathbf{m}|T_0)$, and $h(x)$ with $\pi(\mathbf{m}|T_i)$, the weight $w_{T_0}(\mathbf{m})$ can

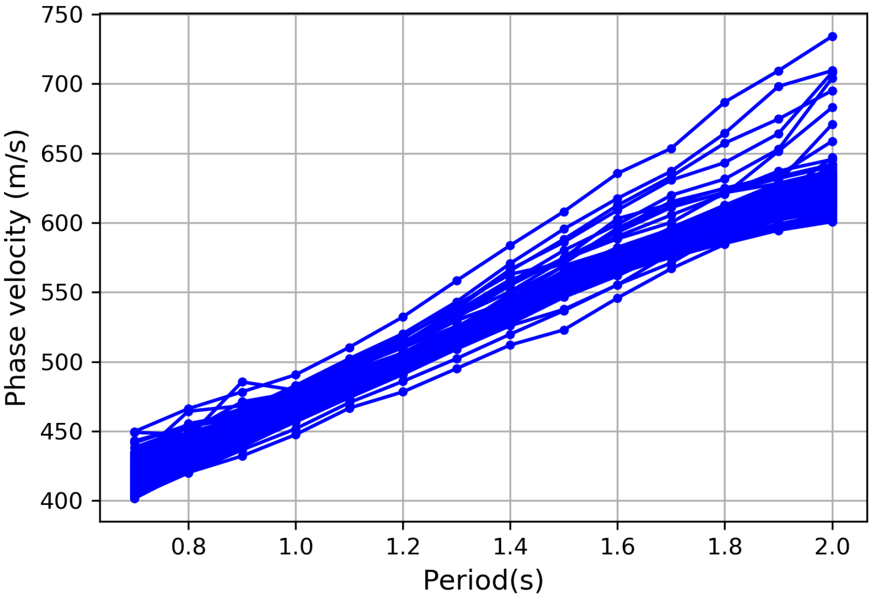


Figure 3.11: Local phase velocity dispersion curves at points along a 2D cross-section (yellow line in Figure 3.1) extracted from phase velocity maps. These were used as data for our two-step inversions (1D and 2D inversions).

be expressed as:

$$w_{T_0}(\mathbf{m}) = \frac{\pi(\mathbf{m}|T_0)/\pi(\mathbf{m}|T_i)}{\sum_{j=1}^{n} \pi(\mathbf{m}_j|T_0)/\pi(\mathbf{m}_j|T_i)} \tag{3.17}$$

where $\mathbf{m}_j$ is the j^{th} sample from the chain with $T = T_i$, $T_0 = 1$, and $\pi(\mathbf{m}|T_0)$ is the density in which we are interested. Combining equation (3.17) with equation (3.12), w_{T_0} will be:

$$w_{T_0}(\mathbf{m}) = \frac{\pi(\mathbf{m})^{1-\frac{1}{T_i}}}{\sum_{j=1}^{n} \pi(\mathbf{m}_j)^{1-\frac{1}{T_i}}} \tag{3.18}$$

where $\pi(\mathbf{m}) = p(\mathbf{m}|\mathbf{d})$ is the target posterior at $T = 1$. Using this equation and equation (3.15), one can calculate any expectations using samples from the tempered chain with $T = T_i$ with respect to the target density $\pi(\mathbf{m})$.

3.6.2 Application to Grane field

We applied the above suite of methods to the Grane field data to estimate shear-wave velocity structures and compared the results. In this section, we used phase velocity dispersion data of fundamental mode Rayleigh-type Scholte waves to invert for the shear-wave velocity structure. For two-step inversions (the 1D and 2D inversions described above) we extracted those local phase velocities that lie along the top of a 2D cross-section (see Figure 3.1) from the phase velocity maps. This produces 257 dispersion curves, one for each geographical location (Figure 3.11). Since the computational cost scales with the minimum of the number of virtual sources and receivers, for the 3D inversion we only used a subset of 36 receivers as virtual sources (Figure 3.12a), each recorded on all 3458 receivers, and picked phase velocities for each virtual source-to-receiver pair. This generates 41842 dispersion curves which constitutes our 3D inversion dataset. Figure 3.12b shows a density map of straight ray paths for those picked phase velocities at 1.0 s. Similarly to the above phase velocity tomography, we only used periods from 0.7 s to 2.0 s with a spacing of 0.1 s (Figure 3.12c).

For 1D depth inversions, the prior pdf of the number of layers is chosen to be a discrete uniform distribution between 2 and 20 layers and the prior of shear velocity is set to be a uniform distribution between 200 m/s and 1400 m/s. The noise level of the likelihood is parameterized using a hyperparameter γ which serves as a scaling factor of a prior Gaussian uncertainty that has standard deviation which comes from the results of Eikonal tomography above (Galetti et al., 2017). The prior of this hyperparameter is simply chosen to be a uniform distribution between 0.001 and 1.0. Since the sea floor depth varies smoothly across Grane field, we assumed an averaged water depth of 127 m everywhere. For the proposal distribution we use a Gaussian distribution: the width of the Gaussian for fixed-dimensional steps (velocity change, moving a cell site and

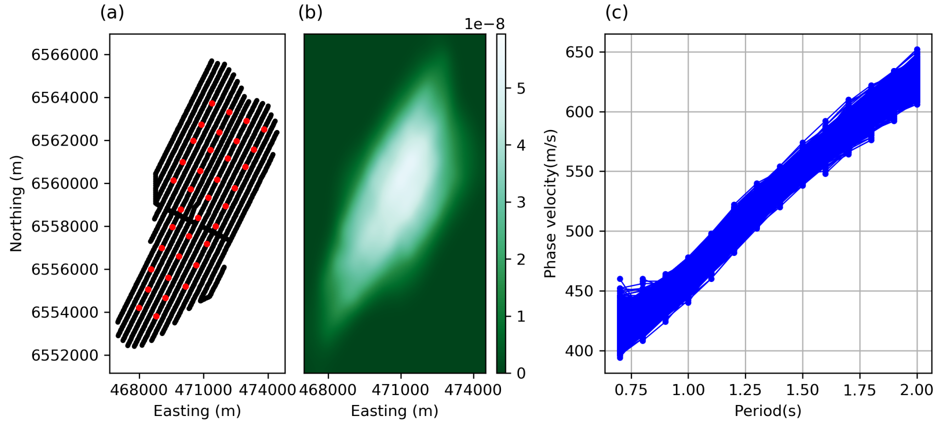


Figure 3.12: **(a)** Receivers used as virtual sources (red dots) for the 3D inversion. Phase velocity dispersion curves are picked from those virtual sources to all 3458 receivers. **(b)** Density of ray paths of phase velocity at 1.0 s and **(c)** phase velocity (travel time) dispersion curves.

hyperparameter change) is chosen by trial and error to ensure the acceptance ratio is between 20 and 50 percent (Hawkins & Sambridge, 2015; Zhang et al., 2018); the width for the trans-dimensional step (birth and death) is selected to produce the maximum possible acceptance ratio. For each inversion at each geographical location we used eight chains, and each chain is run for 3,000,000 iterations with a burn-in period of 1,000,000 during which all samples are ignored for subsequent inference of the posterior pdf. Each chain is thinned by retaining every 100th sample after burn-in, and those samples are used to estimate the posterior pdf's mean and standard deviation.

For the 2D inversion we used a discrete uniform distribution between 100 and 400 as the prior on the number of cells, and the same prior distribution for shear velocity as in 1D inversions. For the noise parameterization we used only one hyperparameter for each period across the section as a scaling factor for uncertainties from Eikonal tomography. Thus we maintain the relative uncertainty structures from Eikonal tomography across the 2D section. The prior for this hyperparamter is chosen to be a uniform distribution between 0.01 and 1. Similarly to above, the proposal distribution for fixed-dimensional steps (velocity change, moving a cell, hyperparameter change) is selected to give an acceptance ratio between 20 and 50 percent. For the trans-dimensional step (birth and death) we used the prior pdf as the proposal distribution (Dosso et al., 2014; Zhang et al., 2018). As shown in Chapter 2, the high lateral-to-vertical spatial aspect ratio of the cross-section can affect the efficiency of McMC sampling when using Voronoi cells. Therefore, we applied a scaling factor of 8 for the vertical dimension to reduce the aspect ratio. We used a total of 16 chains for the 2D inversion and collected 4,000,000 samples from each chain with a burn-in period of 2,000,000. Each chain is thinned by a factor of 100.

For the 3D inversion, the prior of the number of cells is set to be a discrete uniform distribution between 400 and 1500 since the Grane field has a relatively complex structure as indicated by the complex phase velocity maps. We used the same prior for the shear velocity as in 1D and 2D inversions. The noise level is derived from two parameters σ_0 and σ_1 using equation $\sigma = \sigma_0 * predicted\_time + \sigma_1$. The prior density of the two noise hyperparameters are set to be a uniform distribution between 0.0001 and 0.02 and a uniform distribution between 0.0 and 0.1, respectively. The proposal distribution for fixed-dimensional steps are chosen in a similar way to those in the 1D and 2D inversions. For trans-dimensional steps the prior is used as the proposal distribution. As in the 2D inversion, the vertical aspect ratio is scaled by 8 to compensate for the high lateral-to-vertical difference in scaling. To improve the efficiency of 3D McMC we applied parallel tempering. A total of 24 chains are used with 16 chains at temperature 1. The temperatures of the other chains are chosen using equation (3.14) with $p = 3$ and $N = 8$. For each chain we generated 2,700,000 samples with a burn-in period of 1,000,000 and maintained only every 100th sample after burn-in.

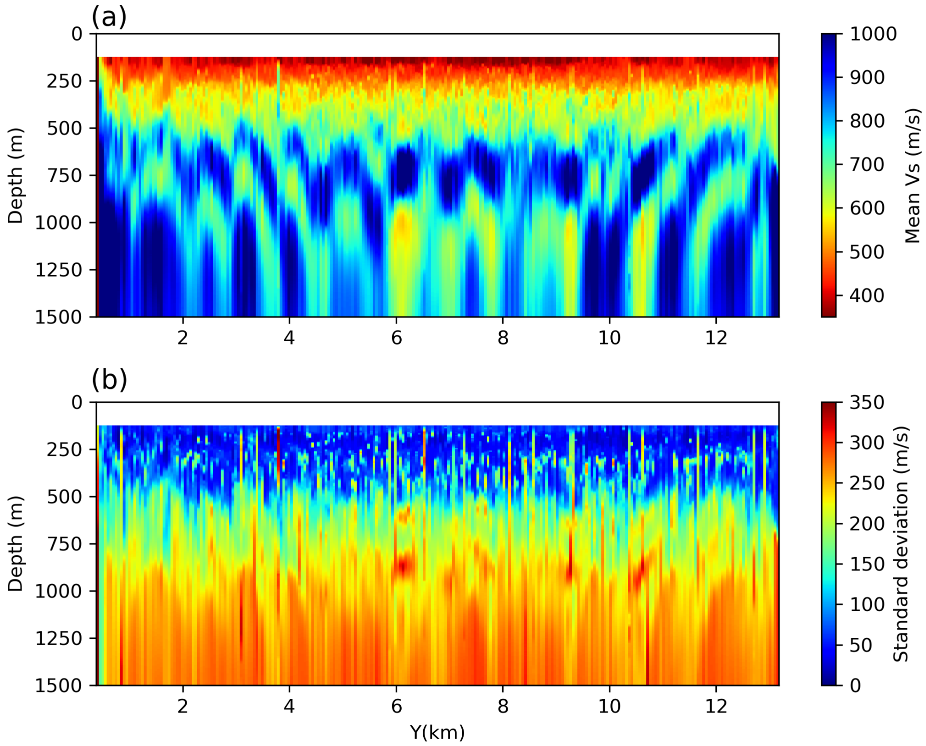


Figure 3.13: **(a)** Mean and **(b)** standard deviation of shear velocity Vs along the 2D cross-section (see Figure 1) from independent 1D Monte Carlo depth inversions. The white top layer represents the water layer which has zero shear velocity.

3.6.3 Results of shear velocity tomography

Figure 3.13 shows the shear-velocity mean and standard deviation from 1D inversions. In the near surface (<250 m), the model has a relatively low velocity layer ($\sim$ 400 m/s, see Figure 3.16a). Between 250 m and 600 m the velocity is slightly higher ($\sim$ 650 m/s) and shows complicated structures which are likely caused by the independence of each 1D inversion. This latter effect is also reflected by some laterally sharp discontinuities across the section. Below 600 m the model shows a high velocity layer ($\sim$ 900 m/s) between Y=6 km and Y=9.5 km down to 800 m. At each side (Y<6 km and Y> 9.5 km), parallel, dipping, alternating high and low velocity anomalies are observed across the section. At the bottom (> 800 m) there exists alternating vertical high and low velocity anomalies. These high and low velocity anomalies are possibly related to similar structures observed in the phase velocity maps at longer periods (Figures 3.7 and 3.9). However, due to high uncertainties (> 200 m/s) at greater depths (>600 m) , those complicated structures are probably not interpretable. The standard deviation map (Figure 3.13b) shows that the near surface structure (< 500 m) is apparently well constrained since surface waves are more sensitive to shallower depths (Figure 3.16a). Note however that there are also lateral discontinuities in the uncertainty map caused by the independence of each 1D inversion; these generally show that the corresponding phase velocity discontinuities are not well resolved (they have very high uncertainty). Note that in this study the phase velocity maps are obtained using Eikonal tomography in which no explicit regularization is imposed. As a result the phase velocity maps show some short-scale structures which may cause some roughness in the shear velocity model estimated by independent 1D inversions. To reduce this issue the conventional regularized tomography may be used to produce smoother phase velocity maps, and consequently to produce a smoother shear velocity model.

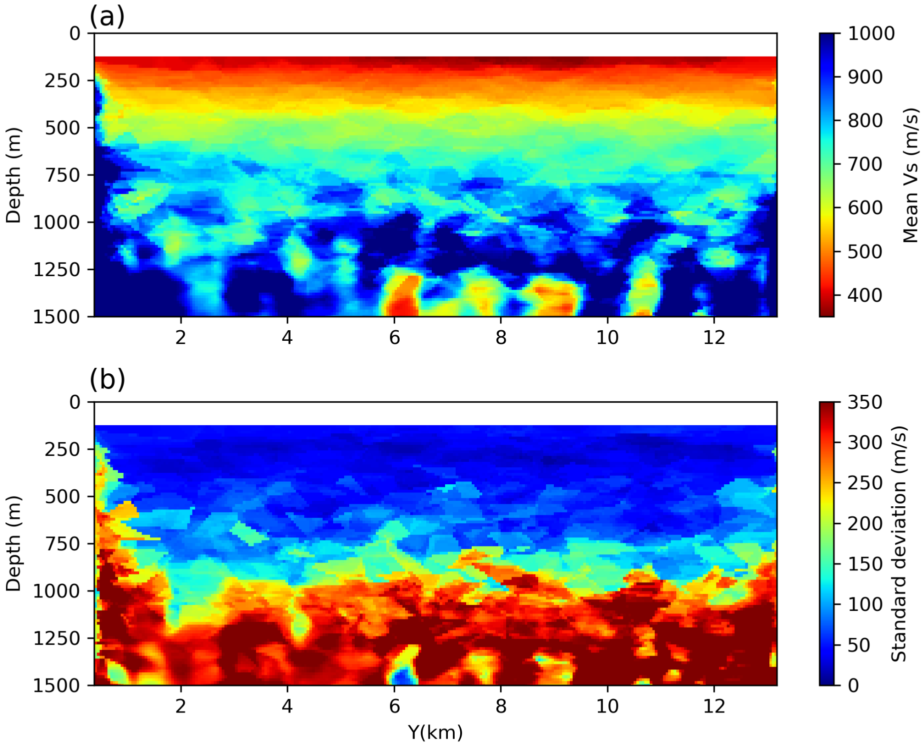


Figure 3.14: (a) Mean and **(b)** standard deviation of shear velocity along the 2D cross-section (see Figure 1) from the 2D Monte Carlo inversions. The white top layer represents the water layer which has zero shear velocity.

For comparison, Figure 3.14 shows the results from the 2D inversion. Overall the mean velocity model is smoother because of lateral interactions that are included in the 2D parameterization. The near surface structure is generally similar to that from 1D inversions: a low velocity layer at depths < 250 m. However, between 250 m and 600 m, the 2D result shows far smoother structures compared to the discontinuous structures in the 1D result. Below 600 m the two results show very different structures. Instead of the clear high velocity anomalies and dipping structures that appeared between 600 m and 800 m in the 1D result, the 2D result exhibits smoother structures. Below 800 m the structure becomes more complicated and at greater depths (> 1000 m) we also observe some vertical high and low velocity anomalies similar to the 1D results. The standard deviation map shows that uncertainties are relatively small (< 100 m/s) from the surface down to 800 m (Figure 3.16b), which is significantly deeper than for the 1D inversions (500 m). This is probably because by including lateral spatial correlations, and because the near surface structure is better determined which further improves the resolution at greater depths. Below 800 m we have very high standard deviations (~ 350 m/s) as expected.

We show the results from the 3D inversion in Figure 3.15. In the near surface (< 250 m) the structure is very similar to the 2D result, showing a clear low velocity layer (Figure 3.16c). Between 250 m and 800 m the structure is smoother compared to both the 2D and 1D results, and does not show the high velocity anomalies and dipping structures that exist in the 1D result. Around 800 m there is a possible high velocity layer from Y=8 km to Y=12 km and beneath it there are some isolated low velocity anomalies at around 1000 m depth. At the west (Y<8 km), the model shows some high and low velocity anomalies between 800 m and 1000 m. Though the structure beneath 1000 m is relatively smooth and shows high velocities (> 900 m/s), there are still some vertical structures similar to those in the 2D and 1D results. This suggests that these structures are probably

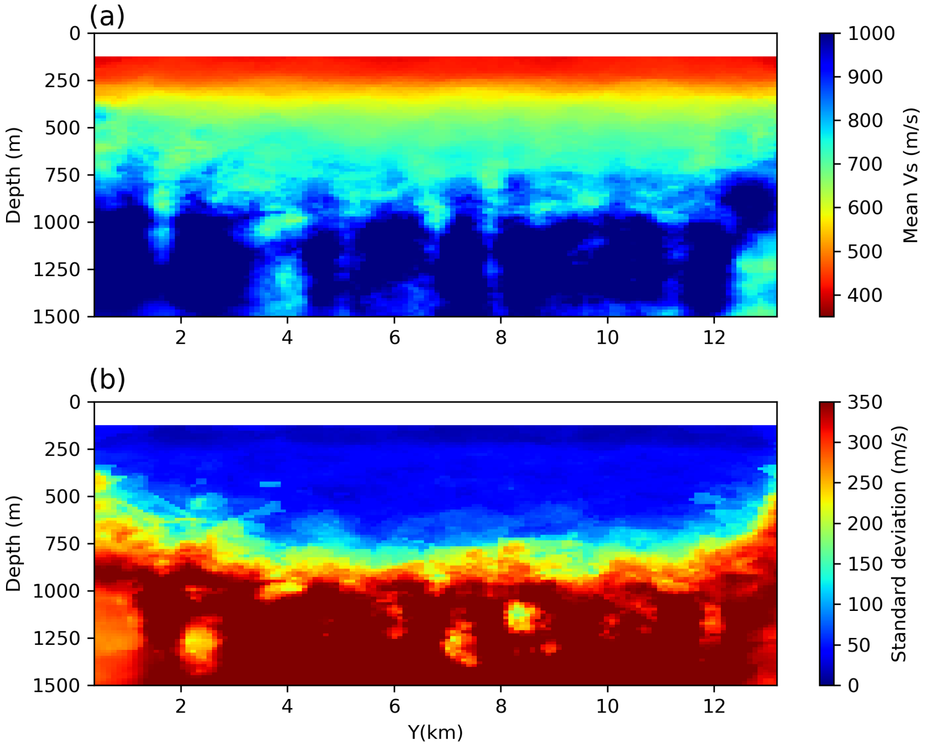


Figure 3.15: **(a)** Mean and **(b)** standard deviation of shear velocity along the 2D cross-section (see Figure 1) from the 3D Monte Carlo inversion. The white top layer represents the water layer which has zero shear velocity.

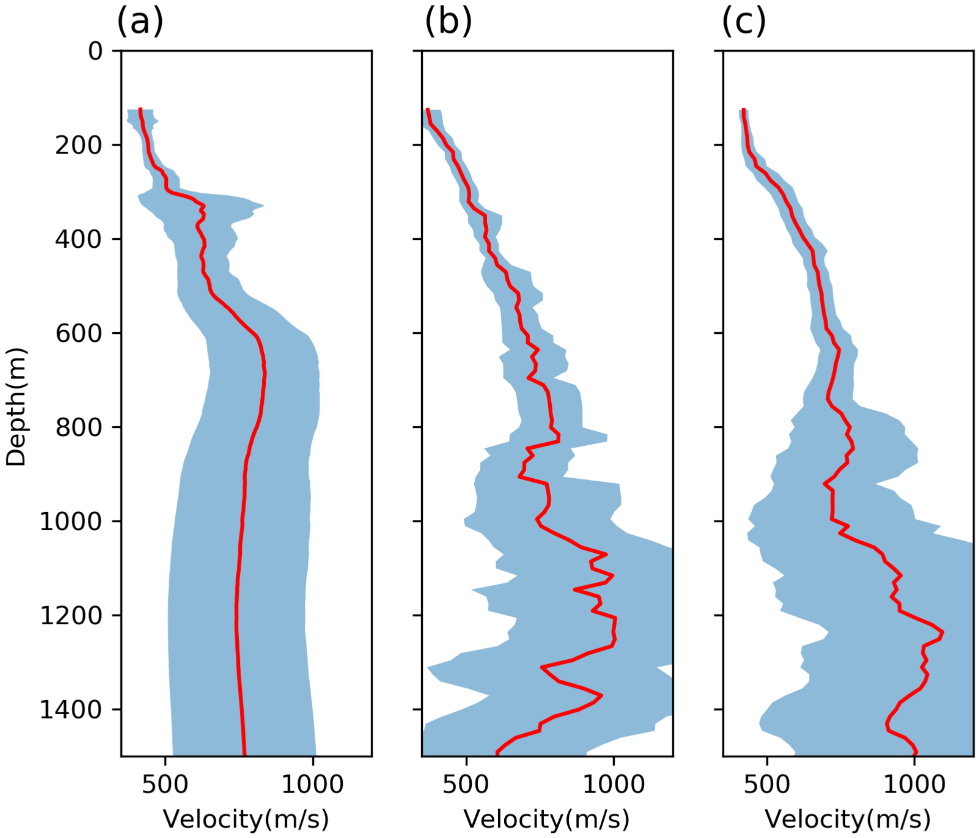


Figure 3.16: 1D marginal mean (red line) and standard deviation (blue area) at the middle (Y=6.7 km) of the 2D cross-section (see Figure 3.1) obtained using **(a)** the 1D Monte Carlo inversion, **(b)** the 2D Monte Carlo inversion and **(c)** the 3D Monte Carlo inversion.

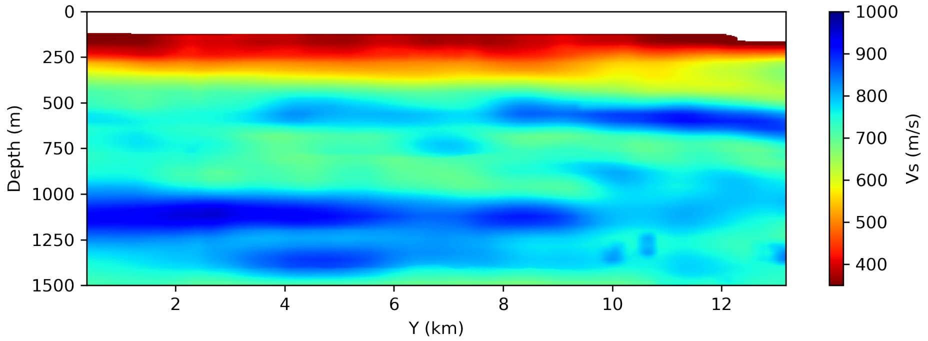


Figure 3.17: Shear-wave velocity Vs model from reflection tomography obtained using active source seismic data. The white top layer represents the water layer which has zero shear velocity.

related to the similar structures in the phase velocity maps and might indicate useful information about the subsurface, or may be caused by poor quality data at longer periods. Similarly to the 2D results, we have low uncertainties from the surface down to 800 m (Figure 3.16c). Therefore, by including lateral spatial correlations in the 2D and 3D inversions, we can have greater confidence at larger depths than in 1D inversions. Compared to the 2D uncertainty result, the 3D result shows lower uncertainties between 400 m and 800 m (Figure 3.16b and c), and they are also spatially smoother. This difference might be caused by errors introduced in the phase velocity maps in the initial 2D Eikonal tomography step (see Chapter 2) since this produces the data used in the 2D McMC inversion. Due to the fact that surface waves are mainly sensitive to the near surface structure, small errors in the phase velocity maps will affect the deeper structure more than the shallow structure – it might be that large velocity variations at greater depths are needed to fit biased data. This may also be the reason why the results of 2D and 3D inversions show different results at depth: since the 3D inversion uses the phase velocity travel time picks directly, it naturally avoids any errors in the phase velocity maps (see Chapter 2). Note that the two sides of the cross-section have higher uncertainties which is caused by lower ray path coverage at the two sides (Figure 3.15b).

To further validate our results and to better understand the three methods, we compare the results with the shear-wave velocity model in Figure 3.17 which was derived from PP-PS simultaneous joint tomography using active source seismic data (Bullock et al., 2015). Overall, the PP-PS tomography model is smoother compared to those from ambient noise dispersion inversions, which might be caused by regularization in the PP-PS tomography (which is not added explicitly in the McMC inversion). In the near surface ($<$ 400 m), the PP-PS tomography model is very similar to the results from 2D and 3D inversions which again suggests that including lateral spatial correlations in the inversion solution improves results. Between 400 and 800 m the PP-PS tomography model shows a high velocity layer around 500 m, which cannot be observed in any of the three models from surface wave dispersion inversion. This is probably because that the frequency content of the two inversion are very different and small scale anomalies present in the PP-PS tomography are unlikely to be resolvable by Scholte wave data. Below 800 m, the PP-PS tomography model is much smoother and does not show the relatively complicated structures of the 2D and 3D results. However even though we have limited resolution below 800 m, there still seems to be some similarities in the PP-PS tomography model and the 3D results, e.g. higher velocities below 1000 m and relatively smoother structure compared to the 1D and 2D results. In conclusion, the 3D inversion seems to produce a shear-velocity model that is more consistent with PP-PS tomography than do 1D and 2D inversions. Note also that the frequency range used in this study is limited to 0.7-2.0 Hz, and we only used fundamental mode surface waves. If we use a larger frequency range and include higher mode data, the results may be improved further.

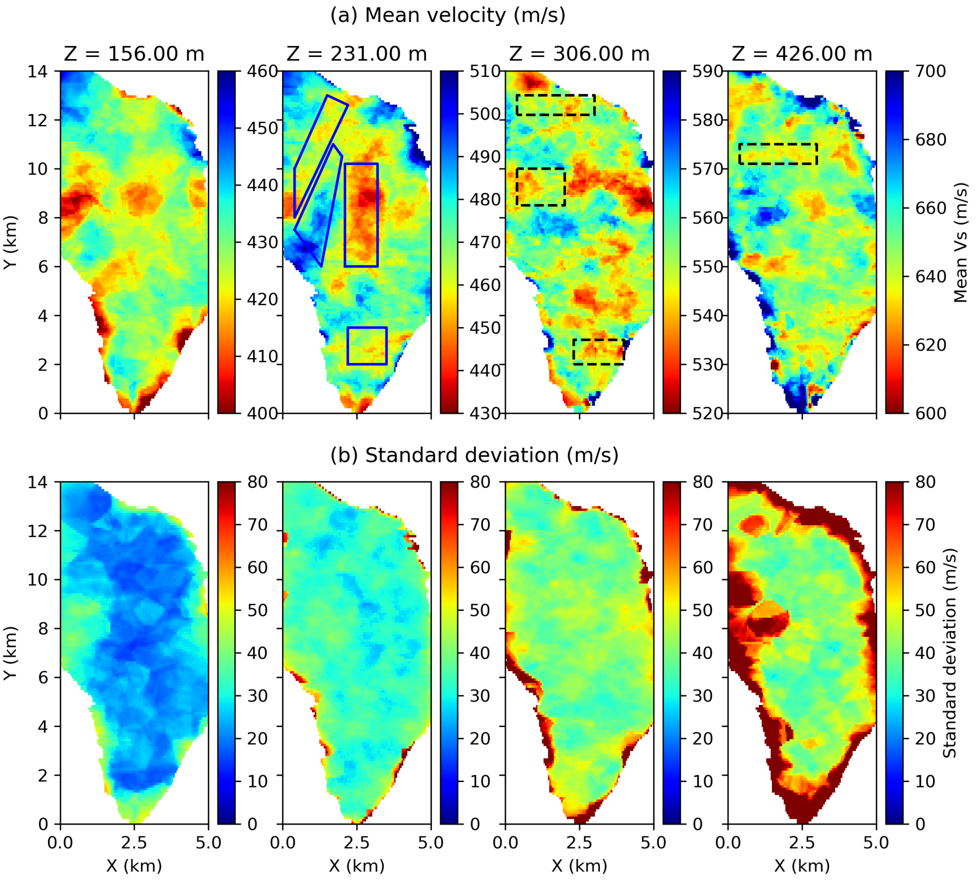


Figure 3.18: Horizontal slices of **(a)** the mean and **(b)** standard deviation of the shear velocity model from 3D Monte Carlo inversion at depths of 156 m, 231 m, 306 m and 426 m from left to right across the figure.

Figure 3.18 shows horizontal slices of the pointwise mean and standard deviation of the 3D shear velocity model estimated using the 3D Monte Carlo method at depths of 156 m, 231 m, 306 m and 426 m. The blue line boxes denote the locations of features in the fundamental mode phase velocity maps discussed above, and the black dashed-line boxes show the locations of features discussed about the phase velocity maps of the first overtone. In the near surface (<250 m) the shear velocity model has similar structures to those in the fundamental mode phase velocity maps at short periods, for example the central low velocity anomaly which might be caused by near surface fluids and the low velocity anomaly in the south. This suggests that the phase velocities at short periods are mostly determined by the near surface structure. Due to possible interpolation errors in the Eikonal tomography step, the edges of phase velocity maps are not as well determined as in the shear velocity model, which is probably the reason why the low velocity anomaly and the high velocity anomaly in the northwest are extended in the shear velocity model compared to the phase velocity maps. In the deeper part (> 250 m), as in the phase velocity maps at longer periods, the velocity structures are more complicated than in the shallow part. However, although the phase velocity maps of the fundamental mode and the first overtone show completely different structures, the shear-velocity model inverted using only the fundamental mode indicates some similar features as observed in the first overtone phase velocity maps (black dashed-line boxes in Figure 3.18). For example, at the depth of 306 m there is a low velocity anomaly at the western edge between Y=8 km and Y= 10 km and a low velocity anomaly at the south around Y=2 km which also appear in the first overtone phase velocity maps. The low velocity channel existing in the first overtone phase velocity maps (red line box in Figure 3.9) can also be clearly observed at a depth of 426 m in the shear-velocity model. This further suggests that the complicated features in the phase velocity maps at longer periods could provide useful information about the subsurface.

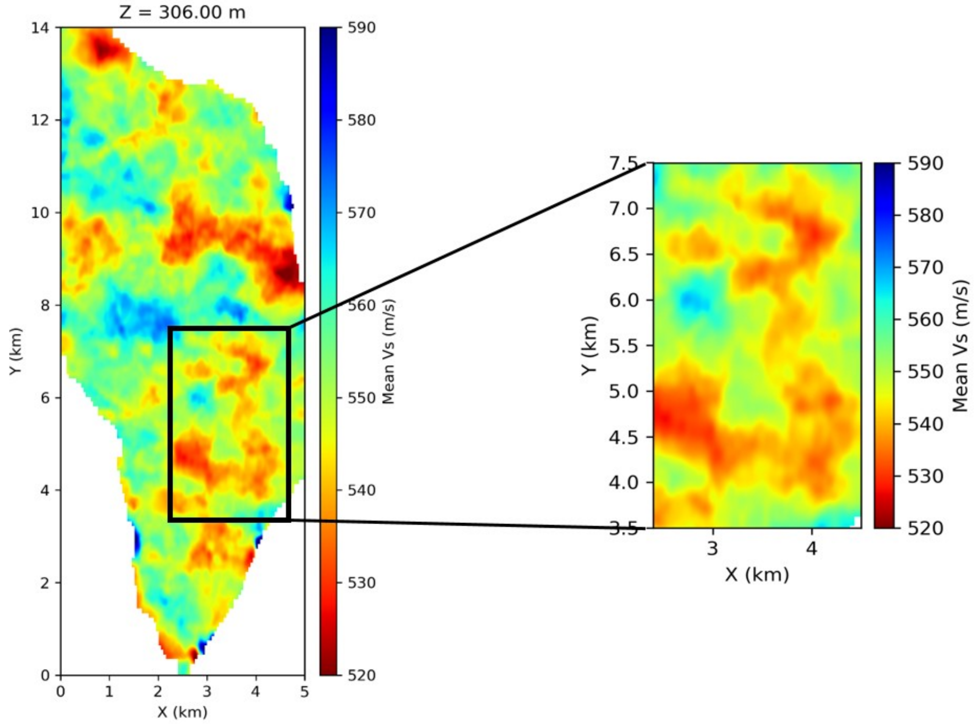


Figure 3.19: The Horizontal slice at depth of 306 m with a detailed structure highlighted in the magnified plot.

At the depth of 306 m there is a low velocity channel feature between Y=3.5 km and Y=7.5km (black box in Figure 3.19). This channel-like feature indicates the presence of a possible palaeoriver channel at the seabed. Note that this feature cannot be observed in either fundamental mode or first overtone phase velocity maps. This suggests that the feature might be averaged out in phase velocity maps since phase velocities are a consequence of structures over a range of depths, a process that our inversion procedures are designed to undo.

Overall the standard deviation maps suggest relatively low uncertainties (< 50 m/s) at all depths, and uncertainties generally increase with depth due to the fact that surface waves have lower sensitivities at greater depth. The standard deviation map at the depth of 231 m shows relatively lower uncertainties at the locations of the velocity anomalies outlined by blue boxes, suggesting that these features are well determined. However, due to the insufficient data coverage at the edges, there is a relatively higher uncertainty area at the western edge around Y=8 km at the depth of 156 m associated with a low velocity anomaly. Similarly, at the western edge between Y=8 km and Y= 10 km the standard deviation map at a depth of 426 m shows high uncertainties.

3.6.4 Analysis of noise level

In this study the data noise level is estimated within the McMC method. Tomographic results are generally sensitive to the noise level as it directly affects the complexity of the model that is needed to fit the data adequately (Bodin et al., 2012). Figure 3.20 shows examples of the data noise level at periods of 0.7 s, 1.0 s, 1.3 s and 1.6 s estimated using the three inversion methods. Figure 3.20a, b show the noise distribution of all of the local phase velocities estimated using the mean scaling factor for the 1D and 2D inversion respectively, and Figure 3.20c

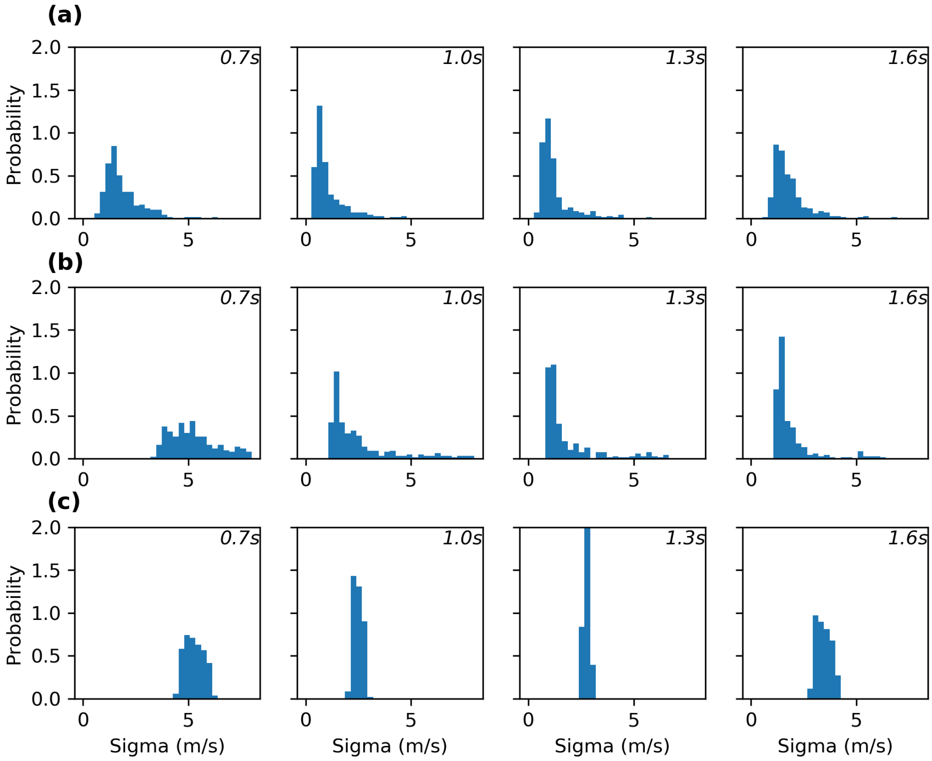


Figure 3.20: Noise distribution of phase velocities from the **(a)** 1D, **(b)** 2D and **(c)** 3D inversions. For each case the distribution is shown at periods of 0.7 s, 1.0 s, 1.3 s and 1.6 s from left to right.

shows the noise distribution of all of the used phase velocities in the 3D inversion estimated using the mean hyperparameters. Note that the noise estimated in the 3D inversion is different from those in the 1D and 2D inversion – the 3D inversion estimates the noise level of travel times between each source-receiver pair; this is transformed to noise on velocities using straight-ray source-to-receiver distances. Overall the noise level from the three methods are relatively consistent with each other – all of them are around 2 to 4 m/s. The noise levels from the 3D inversion are slightly higher than those from the 1D and 2D inversion since they are essentially different quantities. The noise estimated from the 1D and 2D inversions are highly consistent, except at the shortest period of 0.7 s. However, from the Eikonal tomography we observed that the phase velocity map at 0.7 s itself has higher uncertainties than the others. Therefore it is possible that the 1D inversion overestimated the noise level since the method might also account for consistency with the prior range of models (Zhang et al., 2018).

To better understand the effects of the noise level on results, we performed the 1D inversions by fixing the noise level at that estimated using Eikonal tomography (Figure 3.21). Though there are still some discontinuities in the results due to the independence of each 1D inversion, this cross-section has a smoother structure than that in Figure3.13 and does not show the complicated and dipping structures observed previously. However, by doing this we inevitably sacrifice the resolution at greater depths. For example, below 550 m the velocity model is simply equal to the prior. Therefore, the hierarchical Bayesian inversion runs the risk of overfitting data for 1D inversions, while a fixed improper noise estimate might lose resolution (Bodin et al., 2012). In 1D inversions, the model is generally determined by data at tens of discrete frequency points, which might lead the inversion problem to be under-determined when complex models are used. As a result, the McMC result is very sensitive to the noise level since that determines the complexity of the model.

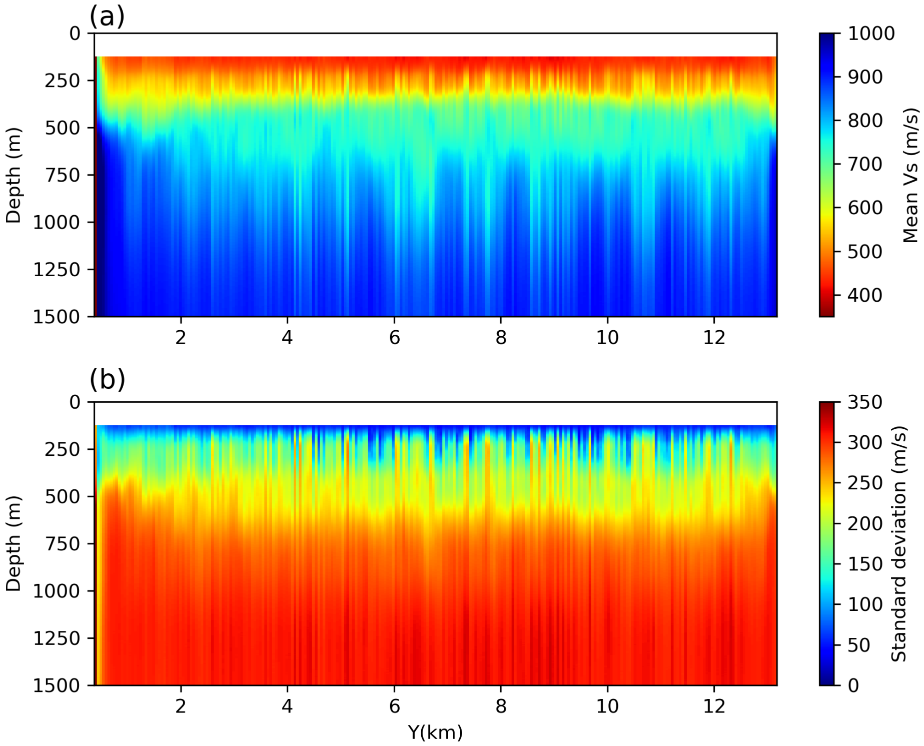


Figure 3.21: **(a)** Mean and **(b)** standard deviation of shear velocity Vs from independent 1D Monte Carlo inversions using only fundamental mode Rayleigh-type Scholte wave phase velocities with noise levels fixed to be uncertainties estimated from Eikonal tomography. The white top layer represents the water layer which has zero shear velocity.

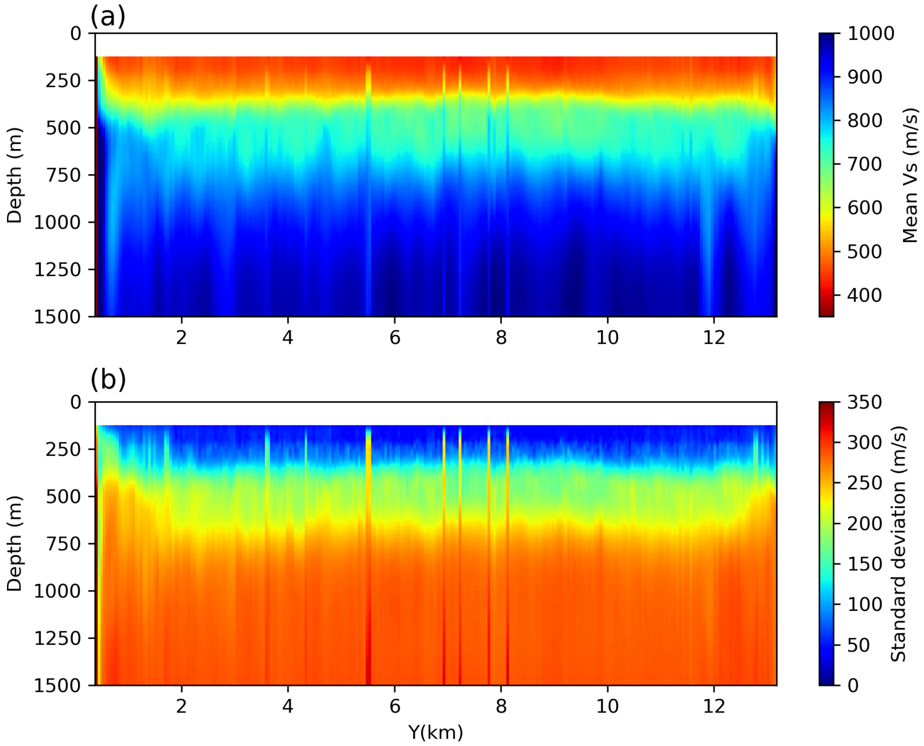


Figure 3.22: **(a)** Mean and **(b)** standard deviation from independent 1D Monte Carlo inversions using both the fundamental mode and the first overtone Rayleigh-type Scholte wave phase velocities, with noise levels fixed to be those uncertainties estimated from Eikonal tomography. The white top layer represents the water layer which has zero shear velocity.

To better constrain the model, we carried out another inversion including the first overtone dispersion data (Figure 3.22). The noise level is fixed at the uncertainties estimated using Eikonal tomography. The mean velocity model (Figure 3.22a) has fewer discontinuities than the result from inversion using only the fundamental mode, and the layer boundaries are also better constrained in depth. The standard deviation model (Figure 3.22b) shows that the near surface layer has smaller uncertainties compared to the previous result. At deeper levels (500 - 700 m), the model is also better determined since it has smaller uncertainties ($\sim$ 180 m/s). However, compared to the results from 2D or 3D inversion, at greater depths (700 - 1000 m) the detailed structure that appeared in the 2D and 3D results cannot be observed and the uncertainty is also higher ($\sim$ 300 m/s), therefore we still lose some resolution.

To conclude, for 1D inversions it is possible that hierarchical Bayesian inversion can overfit the data and produce biased results. However, the noise is generally not easy to estimate and an improper noise level might also cause model resolution to be lost. By including more data (e.g. higher modes dispersion data), this issue can be partly compensated. Alternatively, we have shown that this issue can also be resolved by including lateral spatial correlations in the inversion using 2D and 3D parameterization.

3.7 Discussion

We observed two modes in the cross correlations of ambient noise data and used a dispersion compensation method to separate those modes. However we have shown that even after mode separation there is still some higher mode energy left in the fundamental mode estimate at low frequencies, which limits the frequency range used for tomography. Further research needs to be done in order to find methods to use the full frequency range of the data and thus to obtain more information from ambient noise cross correlations. What is more, there is a concern that the higher mode causes errors in the correlation at the fundamental mode (and vice versa) due to cross-talk in the correlation performed in our initial seismic interferometry (Halliday & Curtis, 2008). Unfortunately there is little we can do about this as mode separation in the original ambient noise remains a largely unsolved problem.

Our method of dispersion analysis needs an estimate of the phase velocity dispersion to resolve the 2π ambiguity in signal phase. This estimate can be obtained using $f - c$ analysis in our case, however it is not always possible to obtain an estimate of the phase velocity dispersion (e.g. if only a sparse array is available). In such cases some other mode separation methods based on single station measurements might be used (e.g. Trampert & Woodhouse, 1995; van Heijst & Woodhouse, 1997; Kritski et al., 2006).

We used Eikonal tomography to determine phase velocity maps by ignoring the amplitude term in equation (3.4). This is justified when the phase velocity map is sufficient smooth so that the spatial variation of amplitude is small (Lin et al., 2009). However, Mordret et al. (2013b) have shown that this could cause some bias in phase velocity maps, especially at long periods (roughly > 1.0 s). According to their study the bias introduced by ignoring the amplitude term is, on average, about 1 m/s which is far smaller than our uncertainties. Our results therefore remain valid.

In this study, we only used the fundamental mode data when we compared the three different methods to estimate a 3D shear-velocity model. However, we have also shown that by including the first overtone dispersion data in 1D inversions, the subsurface structure can be better constrained, as has been observed previously (Gabriels et al., 1987; Xia et al., 2000, 2003). Therefore, future work will be to include those first overtone dispersion data in the 3D inversion to better constrain the subsurface structure and to further improve the resolution of greater depths.

We observed a low velocity anomaly at the center of the field both on the fundamental mode phase velocity maps at short periods (< 1.6 s) and on the shear velocity model at shallow depths (< 250 m), which is correlated with a high density region of pockmark distribution, suggesting that it might be caused by near surface fluids. Therefore, such near surface low velocity anomalies might be used as indicators of fluid leakage from the subsurface reservoirs. This suggests that ambient noise tomography might be used to monitor subsurface fluid storage reservoirs, for example in CO_2 capture and storage scenarios.

Note that in this study we did not take into account any strong anisotropy which might exist in the shallow subsurface (Barkved et al., 2005; Barkved & Kristiansen, 2005; Hatchell et al., 2009). This may introduce some bias in our final results and may explain the complicated phase velocity structures that we observed at long periods (> 1.3 s) and the complicated shear velocity structures at depth (> 250 m). However, our results should at least remain qualitatively valid, and in future it is possible to include anisotropy in 3D Monte Carlo inversions to further improve the results.

The McMC methods are generally very computationally expensive. We now compare the computational cost for the three methods used in this study. For one chain, the 1D depth inversions along the 2D section takes $\sim$186.1 cpu hours while the 2D inversion needs $\sim$206.8 cpu hours; 3D inversion costs $\sim$4824.3 cpu hours for one chain (but of course this produces a complete 3D velocity model across the entire area). Thus the 1D and 2D inversions require almost the same cpu hours which makes sense since they are essentially the same apart from the different parameterization. Note that the 1D inversion and the 2D inversion are performed only along one vertical section. If we assume that we conduct the 1D inversion and the 2D inversion along all vertical sections in the 3D grid, (i.e. across all 100 cross-sections included in the grid used in Eikonal tomography), the cpu hours they consumed would be ~18,610 cpu hours and 20,680 cpu hours respectively – significantly more than for the 3D inversion. However, the computational cost in each case strongly depends on the methods used to assess convergence, which in turn depend on subjective choices. This therefore introduces some subjectivity to the comparison. However, in our experience it is at least true that the cost of the 3D inversion is comparable to that of the 1D or 2D inversions, which has also been found in Chapter 2. To provide an overall idea of the computational cost needed for 3D inversion, the 3D inversion herein takes approximately 22 days with each chain running on 9 cpu cores, so for all 24 chains it requires 216 CPU cores.

Note that in the 3D inversion we used an approximation forward modelling method which involves a 1D modal approximation and a 2D fast marching method. Although the method improves the accuracy of the results, the use of 1D forward modelling and 3D parametrization could fail to accurately capture the underlying properties and structures. For example, Yang & Oldenburg (2012) showed that 1D inversion can cause artefacts in the final results compared to a 3D inversion when using airborne time-domain electromagnetic data. Thus, in the future we hope to use a more accurate forward modelling method.

3.8 Conclusion

We cross correlated about 6.5 hours of ambient noise data from the Grane field, North sea and observed two modes in the constructed seabed Scholte waves. The fundamental mode dominates the signal in the cross correlations of vertical component displacement data while the first overtone dominates in the cross correlations of hydrophone components. We used a dispersion compensation method to separate the fundamental mode and the first overtone. For each mode, we determined phase velocity maps at different periods using the Eikonal tomography method. The fundamental mode phase velocity maps show a low velocity anomaly at the center of the area at short periods (<1.6 s), which might be caused by near surface fluids. At longer periods both modes show complicated phase velocity structures, suggesting that the Grane field might have a complicated geological subsurface.

We then applied three different methods, 1D, 2D and 3D Monte Carlo inversions, to obtain shear wave velocity models of the subsurface using dispersion data of the fundamental mode as data and compared the results. The 1D results show complicated structures at deeper depths (> 250 m) which are probably caused by the independence of individual 1D inversions since the discontinuous structures do not appear in the 2D and 3D results. By including the lateral spatial correlations in the 2D and 3D inversions, we may estimate a more realistic model. The 2D inversion and the 3D inversion show lower uncertainties at greater depths (500 - 750 m), which suggests that the 2D and 3D inversion allow greater confidence at larger depths that the 1D inversion. The 3D inversion results better match a model obtained from reflection tomography than do the results from 2D or 1D inversions. This is probably due to the fact that the 3D inversion uses the measured source-to-receiver travel times directly, and therefore naturally avoids possible errors introduced in the initial (Eikonal) phase velocity tomography step required by the other methods. Though the 3D velocity model is determined using only the fundamental mode dispersion data, it shows some similar features to those which appear in the phase velocity maps of the first overtone. This provides a validation of our 3D results and may suggest that the complex phase velocity maps at longer periods could provide some useful information of the subsurface structure. Overall, the 3D McMC method provides an accurate way to study the subsurface structure using surface wave dispersion data and it is also roughly as computationally efficient as similar 1D and 2D two-step McMC inversions.

Chapter 4

3D Tomographic Monte Carlo joint inversion of earthquake body wave travel times and ambient noise surface wave dispersion data

4.1 Summary

In Chapter 2 and 3 we used Markov chain Monte Carlo (McMC) methods to solve seismic surface wave tomographic problems. Apart from surface wave tomography, seismological body wave travel time tomography have also been used widely to characterise earthquakes and to study the subsurface structure of the Earth. Since these types of problem are often significantly non-linear and have non-unique solutions, McMC methods have been used to find probabilistic solutions. Body and surface wave data are usually inverted separately to produce independent

velocity models. However, body wave tomography is generally sensitive to structure around the sub-volume in which earthquakes occur and produces limited resolution in the shallower Earth, whereas surface wave tomography is often sensitive to shallower structure. To better estimate subsurface properties, we therefore jointly invert for the seismic velocity structure and earthquake locations using body and surface wave data simultaneously. We apply the new joint inversion method to a mining site in the U.K. at which induced seismicity occurred and was recorded on a small local network of stations, and where ambient noise recordings are available from the same stations. The ambient noise is processed to obtain inter-receiver surface wave dispersion measurements which are inverted jointly with body wave arrival times from local earthquakes. The results show that by using both types of data, the earthquake source parameters and the velocity structure can be better constrained than in independent inversions. Trade-offs between source parameters and velocities appear to bias results if only body wave data are used, but we show that this issue is largely resolved by using the joint inversion method.

4.2 Introduction

Seismic tomography is a method to estimate the spatial distribution of properties of the subsurface, and is often used in order to understand heterogeneity and processes in the Earth's interior. In seismic tomography one usually parameterizes subsurface properties in some way to form a subsurface *model*, then solves the parameter estimation problem given observed data and a relationship between the data and the parametrized physical properties.

Seismic tomography problems are traditionally solved using linearised methods to estimate the model parameter values which minimize the misfit between observed and synthetically predicted data. These methods first approximate the non-linear physical relation by a linear relation, and iteratively update both the model and the linear approximation to reduce the misfit predicted by that linearised relation (Aki & Lee, 1976; Dziewonski & Woodhouse, 1987; Iyer & Hirahara, 1993; Tarantola, 2005). Since the problem is often under-determined and ill-posed, regularization is added to the process to enforce particular properties on the model (e.g., smoothness or minimal deviation from a reference model). However, the form of regularization is often arbitrary and the strength of regularization is chosen by trial and error, usually by invoking ad hoc criteria. Valuable information can therefore be concealed by regularization (Zhdanov, 2002). Moreover, it is difficult if not impossible to estimate accurate uncertainties in solutions of non-linear problem when using linearised methods since the family of model parameter values that fit the data is defined by the true nonlinear physics, and not by the linearised relations.

Markov chain Monte Carlo (McMC) methods have been introduced to geophysics to resolve some of these issues (Mosegaard & Tarantola, 1995; Malinverno et al., 2000; Malinverno, 2002; Malinverno & Briggs, 2004; Bodin & Sambridge, 2009; Galetti et al., 2015, 2017; Zhang et al., 2018). These methods solve the problem in a Bayesian sense by generating a set (or chain) of samples whose density approximates a *posterior* probability density function (pdf): this describes the probability of the model given both the observed data and any available *prior* information. The method has been extended to trans-dimensional inversions by using reversible jump McMC (rj-McMC – Green, 1995; Bodin & Sambridge, 2009) such that the dimensionality of the parameters space (the number of parameters and indeed their meaning) can vary in the inversion. This has the advantage that the parameterization can be adapted and simplified so as to best represent information in the data and prior information without over-parameterizing the model, which significantly improves performance in otherwise high dimensional problems (Malinverno & Briggs, 2004; Bodin & Sambridge, 2009; Bodin et al., 2012; Galetti et al., 2015; Zhang et al., 2018). The rj-McMC method has been used to estimate 2D phase or group velocity maps of the crust (Bodin & Sambridge, 2009; Zulfakriza et al., 2014; Galetti et al., 2015; Zheng et al., 2017; Crowder et al., 2019) and to estimate seismic velocity profiles with respect to depth in the Earth (Bodin et al., 2012; Shen et al., 2012, 2013; Young et al., 2013; Galetti et al., 2017; Zhang et al., 2019). The method was recently extended to estimate 3D velocity models using body wave travel time data (Piana Agostinetti et al., 2015; Hawkins & Sambridge, 2015; Burdick & Lekić, 2017). In Chapter 2 and 3 we applied the method to solve surface wave tomographic problems.

In the above studies, body waves and surface waves are used separately to construct velocity models. Seismic body waves are generally sensitive to deeper structure where earthquake sources occur, and produce limited resolution closer to the surface. This is because we usually have a relatively sparse station array compared to the density of sources, which results in relatively sparse body wave ray coverage in the shallower Earth. In comparison, fundamental mode surface waves are generally more sensitive to shallower rather than to deeper structure. Body and surface wave data can therefore usefully be combined to better constrain the subsurface velocity structure.

Such joint inversions have already been used widely to study the crust and upper mantle structure (West et al., 2004; Reiter & Rodi, 2008; Obrebski et al., 2011, 2012; Rawlinson & Fishwick, 2012; Zhang et al., 2014; Syracuse et al., 2015; Fang et al., 2016; Liu & Zhao, 2016; Roecker et al., 2017). However, these studies were performed using linearised inversion methods which renders associated uncertainty estimates questionable at best. In this study we apply the rj-McMC algorithm to fully nonlinear joint inversion using both body wave arrival times and surface wave dispersion data. We show that results are significantly improved over independent body or surface wave inversions, both in terms of velocity structure and earthquake source location uncertainties. In the next section we summarise the rj-McMC algorithm and describe how it is applied to the joint inversion problem. In section 3 we apply the new McMC joint inversion method to data from an ex-mining site located to the north of New Ollerton, U.K, and compare the results with those from individual inversions in section 4. Finally we discuss the implication of this work in section 5 before concluding.

4.3 Methodology

4.3.1 Parametrization

As in Chapter 2 and 3, in order to perform trans-dimensional inversion in three spatial dimensions we use Voronoi cells to parameterize our seismic velocity models (Figure 2.1). A Voronoi cell is defined by a generating point (called a site) and its volume which consists all points that are closer to that site than to any other. Each cell has associated seismic properties, e.g., P-wave velocity and shear wave velocity. In this study, we use constant velocities within each cell. Our velocity model can therefore be parameterized as $(\mathbf{c}, \mathbf{v}_p, \mathbf{v}_s)$, where $\mathbf{c}$ is the vector of positions of Voronoi sites, and $\mathbf{v}_p$ and $\mathbf{v}_s$ are vectors of the associated P-wave velocity and shear wave velocity in each cell. Each earthquake source (number i) is parameterized as $s_i = (x_i, y_i, z_i, t_i)$, where x_i, y_i, z_i is the source location and t_i is the origin time. Our model $\mathbf{m}$ therefore can be represented as $(\mathbf{c}, \mathbf{v}_p, \mathbf{v}_s, \mathbf{s})$.

4.3.2 Reversible jump Markov chain Monte Carlo (rj-McMC)

We use rj-McMC to perform 3D tomographic inversion following the approach described in Chapter 2. Rj-McMC is a generalized Metropolis-Hastings algorithm which generates a chain of samples distributed according to a target probability density. The algorithm allows the number of parameters to change along the chain (Green, 1995), which makes the parameterization adaptable to the data and avoids the need to specify it exactly prior to the inversion (Bodin & Sambridge, 2009). In seismic tomography we are interested in the posterior pdf of model $\mathbf{m}$ given the observed data $\mathbf{d}_{obs}$,

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})} \tag{4.1}$$

where $p(\mathbf{d}_{obs}|\mathbf{m})$ is the *likelihood* which describes the probability of data given a specified model $\mathbf{m}$; $p(\mathbf{m})$ is the prior pdf which describes information that is independent of data and $p(\mathbf{d}_{obs})$ is a normalization factor called the *evidence*. We use a Gaussian distribution for the likelihood, for which the data variance is estimated in a hierarchical way in the inversion (Malinverno & Briggs, 2004; Bodin et al., 2012; Zhang et al., 2018, 2019). The prior $p(\mathbf{m})$ is chosen to be a Uniform distribution.

Within each chain a new model $\mathbf{m}^{'}$ is drawn from a so-called proposal distribution $q(\mathbf{m}^{'}|\mathbf{m})$ that depends on the current model $\mathbf{m}$, and is accepted or rejected with a probability $\alpha(\mathbf{m}^{'}|\mathbf{m})$ given by (Green, 1995)

$$\alpha(\mathbf{m}^{'}|\mathbf{m}) = min[1, \frac{p(\mathbf{m}^{'})}{p(\mathbf{m})} \times \frac{q(\mathbf{m}|\mathbf{m}^{'})}{q(\mathbf{m}^{'}|\mathbf{m})} \times \frac{p(\mathbf{d}_{obs}|\mathbf{m}^{'})}{p(\mathbf{d}_{obs}|\mathbf{m})} \times |\mathbf{J}|] \tag{4.2}$$

where $\mathbf{J}$ is the Jacobian matrix of the transformation from $\mathbf{m}$ to $\mathbf{m}^{'}$ and is used to account for the volume change of parameter space during jumps between different dimensionalities, but where in this case the Jacobian is an identity matrix (Bodin & Sambridge, 2009). The new model $\mathbf{m}^{'}$ is accepted or rejected by generating a random number γ from a Uniform distribution on (0, 1] and comparing it with α. If $\gamma < \alpha$, the new model $\mathbf{m}^{'}$ is accepted; otherwise the new model is discarded and the current model is repeated as a new sample in the chain. The acceptance ratio α guarantees that the density of samples converges to the posterior pdf asymptotically as the number of samples tends to infinity (Green, 1995).

Monitoring the convergence of Markov chains is an important component of McMC methods. In this study, we use the absolute misfits and the number of cells to monitor convergence as used in several previous studies (Bodin & Sambridge, 2009; Bodin et al., 2012; Dosso et al., 2014; Galetti et al., 2015; Hawkins & Sambridge, 2015; Zhang et al., 2018, 2019). For example, when the misfit value and the number of cells become approximately stationary, we assume the chain has reached some sort of dynamic equilibrium. Since consecutive samples are correlated (McMC is a random walk process and only converges to the posterior distribution as the number of samples tends to infinity), the estimated probability pdf from any finite set of samples is often biased (Chan & Geyer, 1994). Therefore, we retain every 50^{th} sample along the chain once equilibrium has been reached, and only those retained samples are used to calculate parameter statistics (mean, standard deviation, etc.).

4.3.3 Joint inversion of body waves and surface waves

In seismic body wave tomography, the earthquake source locations are generally unknown within some volumetric region of uncertainty as are origin times. We therefore include these source parameters in our inversion. This produces a trade-off between source parameters and the seismic velocity model and to reduce this effects Piana Agostinetti et al. (2015) updated the source origin times in an optimization for each velocity model. However that approach may cause errors in the results since sources may converge to incorrect locations and times, and it does not allow correct uncertainty analysis for source parameters. In this study we therefore also include origin times as parameters to be varied in the Markov chain. We start the chains with initial source parameter values obtained using a standard linearised optimization, whereafter they can vary freely within the prior pdf (defined below).

To forward model body wave travel time data we use a 3D fast marching method (Rawlinson & Sambridge, 2004; Valero-Gomez et al., 2013). Due to source-receiver reciprocity, fast marching can be conducted either from sources to receivers or vice versa. Therefore, in practice one chooses the more efficient option based on the minimum number of sources and receivers, and we model from receivers to sources.

For surface wave dispersion data, we use the two step forward modelling method described in Chapter 2 and applied in Chapter 3. First, for each geographical point the local phase velocity at each frequency is computed using the 1D velocity profile beneath that point using a modal approximation (Herrmann, 2013) to create a 2D phase velocity map across the surface. Then, since our dispersion measurements are made between two receivers, for each receiver-to-receiver pair the phase travel time at a specific frequency can be calculated using a 2D ray tracing method (Rawlinson & Sambridge, 2004). Group velocity travel times can be calculated by integrating over the ray path traced through phase velocity maps (Cerveny, 2005; Reiter & Rodi, 2008).

In joint inversion, the relative weights between different data types usually affect the results significantly (Bodin et al., 2012; Shen et al., 2012). In linearised methods, the weight is generally determined by subjective choices which could cause errors in results. In this study we set the data noise level of both data types to be free parameters so that the relative importance of different data types is determined by their own noise level (Bodin et al., 2012; Shen et al., 2012). As in Chapter 2, we hyper-parameterize the noise level using a linear relation with respect to travel times $\sigma = \sigma_0 \times traveltime + \sigma_1$, for each of body and surface wave travel times independently, where σ_0 and σ_1 are free hyperparameters.

In our rj-McMC algorithm there are six types of perturbation: adding a Voronoi cell, removing a cell, moving a cell, changing a cell's velocities, changing the source parameters, and changing the data noise hyperparameters. This results in the following algorithm:

1. Select an initial model from the prior pdf (for seismic velocities) or from a linearised inversion (for source locations and times)
2. Generate a new model $\mathbf{m}^{'}$ by randomly choosing one of the six possible perturbation types listed above, and then perturbing the current model according to the proposal distribution.
3. Calculate the acceptance ratio α in equation (4.2) and accept or reject model $\mathbf{m}^{'}$ with probability α.
4. Repeat from (ii).

For the fixed-dimensional step (moving a cell, changing velocities, changing source parameters and changing the hyperparameters), we use a Gaussian distribution which is centred at the current model as the proposal distribution. The width of the Gaussian is a parameter which needs to be tuned for each inversion (Hawkins & Sambridge, 2015; Zhang et al., 2018). For trans-dimensional steps (adding or deleting a cell), the prior is used as the proposal distribution which usually gives a higher acceptance ratio than using a Gaussian proposal distribution as noted in Chapter 2.

4.4 Application to the New Ollerton mining site

We applied the method to a mining site located to the north of New Ollerton, Nottinghamshire, U.K (Figure 4.1) which operated from 1925 to 2015. A network of seven stations was deployed at the site and recorded 291 microseismic events in 2014. Figure 4.1 shows the location of the stations, event locations from the initial linearised inversion, and a histogram of the depth distribution of event locations. The events mainly occurred around 0.9 km depth with a few occurring significantly shallower or deeper. We used a total of 1725 P-wave arrival times and 923 S-wave arrival times obtained from the British Geological Survey (BGS) (Butcher et al., 2017).

We applied ambient noise interferometry (Campillo & Paul, 2003; Curtis et al., 2006; Bensen et al., 2007) to obtain surface wave dispersion data for each inter-receiver pair. The data are first band-filtered between 0.8 s and 3.0 s to filter out earthquake signals which comprise higher frequencies. Cross-correlations between each receiver pair are then calculated using 24-hour long time segments, which are then stacked over the whole year to improve the signal-to-noise ratio (SNR). The group velocity dispersion of each receiver pair is picked using the frequency-time analysis (FTAN) method (Dziewonski et al., 1969; Levshin et al., 1972, 1992; Herrin & Goforth, 1977; Russell et al., 1988; Ritzwoller & Levshin, 1998; Levshin & Ritzwoller, 2001; Nicolson et al., 2012; Yanovskaya et al., 2012). Figure 4.2 shows an example of the FTAN image used to pick group velocities. We discarded station-pairs for which the SNR is smaller than 5, and those whose inter-receiver distances are smaller than twice the wavelength at any frequency due to far-field surface wave approximation that is implicit within ambient-noise surface wave tomography (Yao et al., 2006; Lin et al., 2009). This results in a total of 12 inter-receiver dispersion curves across the New Ollerton area (Figure 4.1). Since Nicolson et al. (2012) and Galetti et al. (2017) both showed that uncertainties estimated directly from the FTAN images tend to be poor, uncertainties in dispersion curves were estimated hierarchically within the Markov chain.

We performed three different inversions: first using only body wave travel times, second using only surface wave dispersion data, and a third, joint inversion using both types of data. For each inversion the prior pdf of shear velocity is set to be a Uniform distribution between 1.0 km/s and 4.0 km/s at all 3D locations. Similarly, for body wave inversion and joint inversion the prior pdf of P-wave velocity is set to be a Uniform distribution between 1.6 km/s and 6.0 km/s. In the inversion with only surface waves, P-wave velocity is linked to shear velocity using a typical ratio 1.73 and for all three inversions density is computed from the P-wave velocity using a typical crustal relationship $\rho = 2.35 + 0.036 \times (v_p - 3.0)^2$ where v_p is in km and ρ is given in g·cm^{-3} (Kurita, 1973). Since surface waves are much more sensitive to shear velocity than P-wave velocity or density, the approximation should be sufficient in our case. The prior of the number of Voronoi cells is chosen to be a Uniform distribution between 20 and 300. For each event location we use a Uniform distribution across a 2 km box centred at the initial location estimated by BGS using linearised methods (Butcher et al., 2017) as the prior, and for the origin time we used a Uniform distribution with 1 second width centred at the initial origin time. For body wave travel times the prior of the hyperparameters σ_0 and σ_1 are chosen to be Uniform distributions between 0 and 0.1. Similarly for surface wave group travel times the prior of the two hyperparameters are set to be Uniform distributions between 0 and 0.2. Since seismic velocity generally varies more rapidly in the vertical direction than horizontally, we scaled the vertical direction by a factor of 5 larger to ensure vertical and horizontal directions are balanced as suggested by Zhang et al. (2018).

For a fixed-dimensional step (moving a cell, changing velocities, changing source parameters and hyperparameters) we use a Gaussian perturbation centred at the current value as the proposal distribution. The width of the Gaussian distribution is chosen by trial and error to give an acceptance ratio between 20 and 50 percent. For a trans-dimensional step (adding or deleting a cell) the proposal distribution is chosen to be the prior pdf (Dosso et al., 2014; Zhang et al., 2018). For each inversion we used 16 chains; each of which contains 1,600,000 samples including a burn-in of 800,000 to reach apparent equilibrium. To reduce correlations between successive samples we only retain every 50^{th} sample in the chain post burn-in. Those sample are used to calculate parameter means and standard deviations.

4.5 Results

4.5.1 Source parameters

Figure 4.3 shows the mean and standard deviation of event locations obtained from body wave inversion (Figure 4.3a and b) and from joint inversion (Figure 4.3c and d). Both results show that events occur deeper (majority > 1 km) than the initial locations from BGS (majority < 1 km). The results show two clusters: one in the southeast and the other in the north. The southeastern cluster has slightly higher uncertainties than those in the north, which is probably caused by the fact that the stations are distributed to one side of the southeastern cluster. Compared to the standard deviation from body wave inversion (around 0.5 km), the location results from joint inversion show lower uncertainties (around 0.4 km).

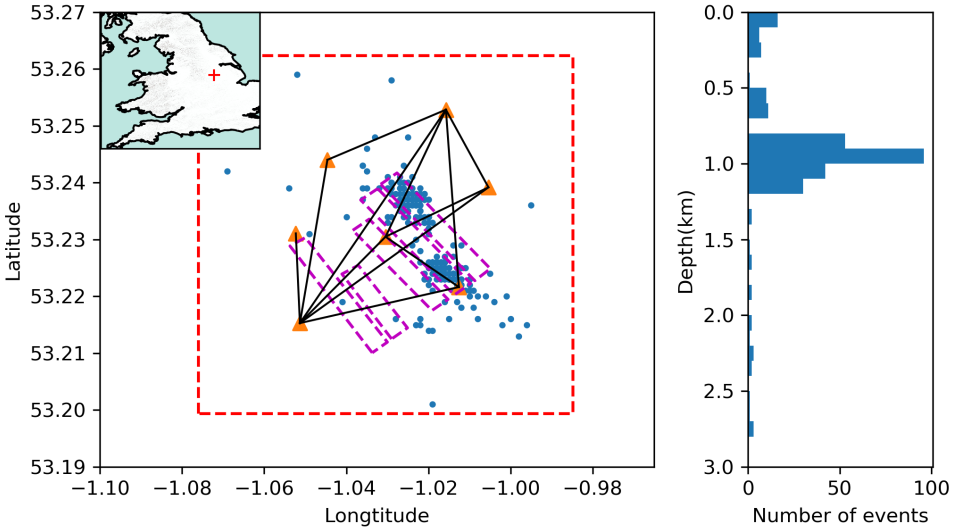


Figure 4.1: Microseismic events (blue dots) recorded with seven seismic stations (orange triangles) at New Ollerton mining site. Black lines between stations show approximate paths along which surface wave dispersion data are available. The pink boxes show the location of coal seams located between 800 and 900 m. The red dashed-line box shows the extent of the other maps herein. The red plus in the inset map denotes the location of the mining site in England. The right panel shows a histogram of event depths.

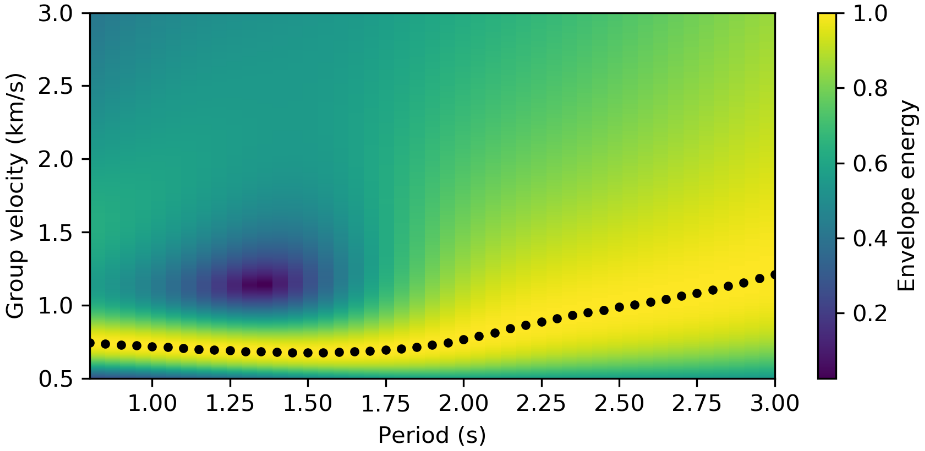


Figure 4.2: An example of a frequency-time analysis (FTAN) envelope image which is used to pick group velocities. The black dots show the group velocities picked in this case.

This suggests that by including even only 12 surface wave dispersion curves the event locations can be better constrained since dispersion data provides additional information about the velocity model between stations. Figure 4.4a and b show histograms of the standard deviations of source origin time obtained using body wave tomography and joint inversion respectively. Most standard deviations from body wave tomography are higher than 0.05 s, while those from joint inversion are centred around 0.05 s. Therefore, by including surface wave dispersion data in the inversion, the source origin time can also be better constrained since this helps to resolve the trade off between origin time and velocity structure.

4.5.2 Velocity models

Figure 4.5 shows horizontal slices through the 3D mean and standard deviation maps of shear wave velocity at depths of 0.2 km, 0.5 km and 1.0 km. The standard deviation map at 0.2 km shows that only a small part of the model is well

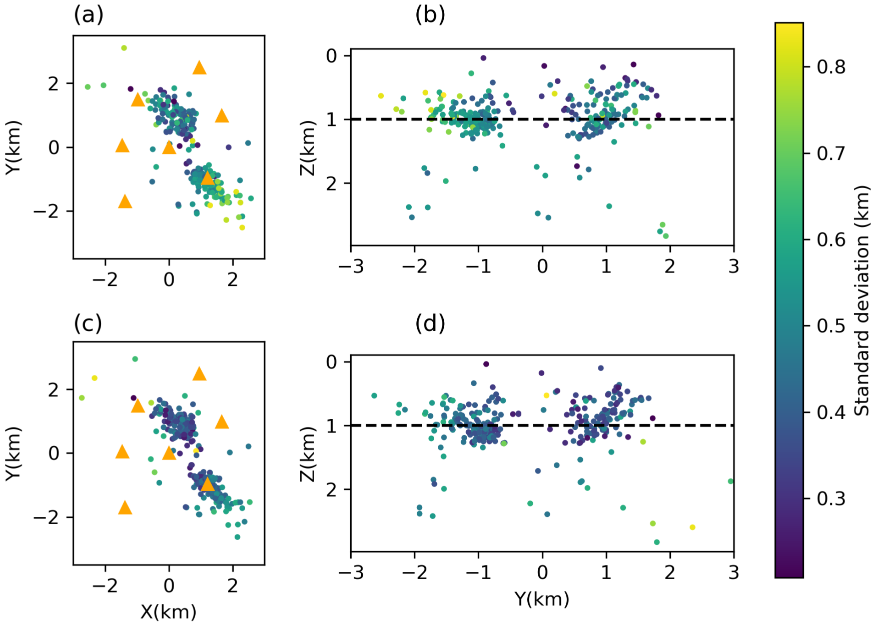


Figure 4.3: Source location results. **(a)** and **(b)** are map view and a cross-section along latitude of source locations obtained using body wave travel time data only. **(c)** and **(d)** are map view and a cross-section of source locations obtained using both body wave travel time data and surface wave dispersion data. The orange triangles show the location of stations. The color of each dot reflects the standard deviations of each source location.

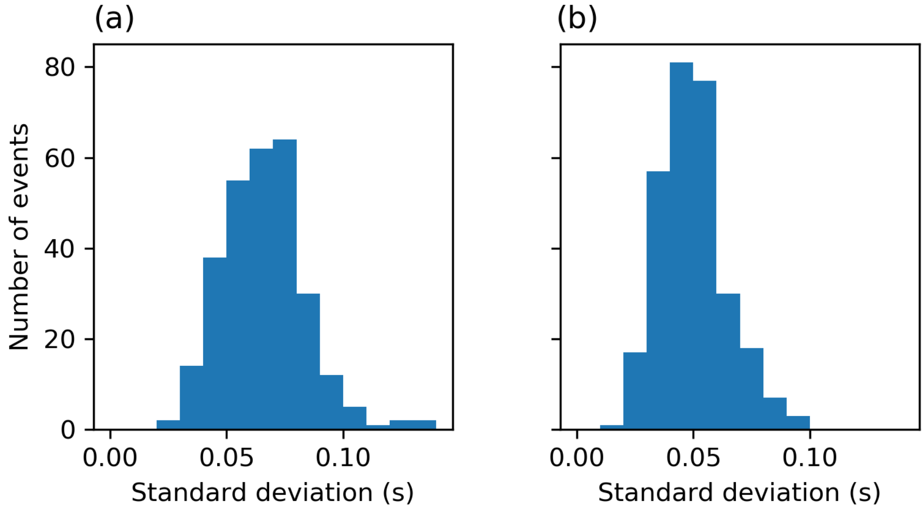


Figure 4.4: Histograms of the standard deviations of source origin time obtained using **(a)** body wave tomography and **(b)** joint body and surface wave inversion.

constrained, which is associated with lower velocities in the mean velocity map. Most of the other maps show the same values as the prior, suggesting that body waves have very limited resolution in the near surface as expected. The mean velocity map at 0.5 km depth shows that the shear velocities in the southwest and northeast are lower than elsewhere. The standard deviation map suggests that most of the structure within the boundary of the array is reasonably well constrained by the data, other than in the southeast which has higher velocities and higher uncertainties, probably caused by the limited data coverage at that area. At 1 km depth the mean velocity map shows that the velocity in the west is lower than the east, the northern earthquake cluster occurs at the boundary of velocity anomalies, and the southeastern earthquake cluster is correlated with a clearly-defined high velocity anomaly. Between the two clusters there are low velocity anomalies. The standard deviation map shows very low uncertainties (< 0.2 km/s) in the southwest associated with the low velocity anomaly, which suggests that the low velocity anomaly is well constrained, whereas slightly higher uncertainties (about 0.4 km/s) are observed elsewhere. There are loops of higher

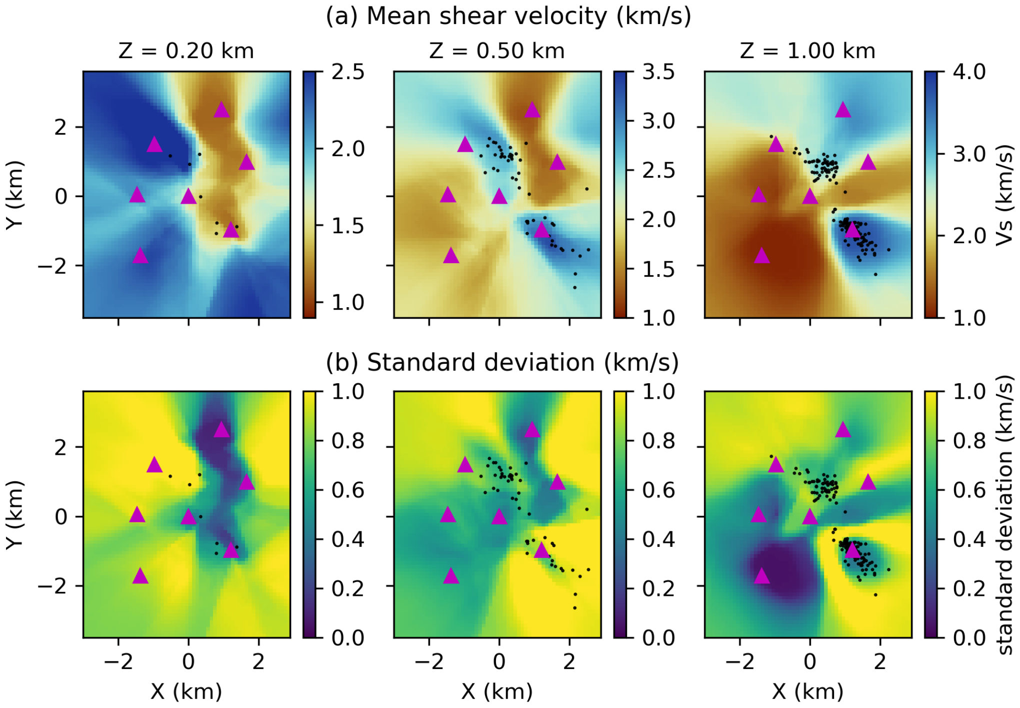


Figure 4.5: Horizontal slices through the 3D shear velocity model at depth of 0.2 km (left), 0.5 km (middle) and 1.0 km (right) obtained using body wave travel time data only. The upper panels **(a)** and the bottom panels **(b)** show the mean velocity maps and standard deviation maps, respectively. At each slice events within 0.2 km of the depth are plotted.

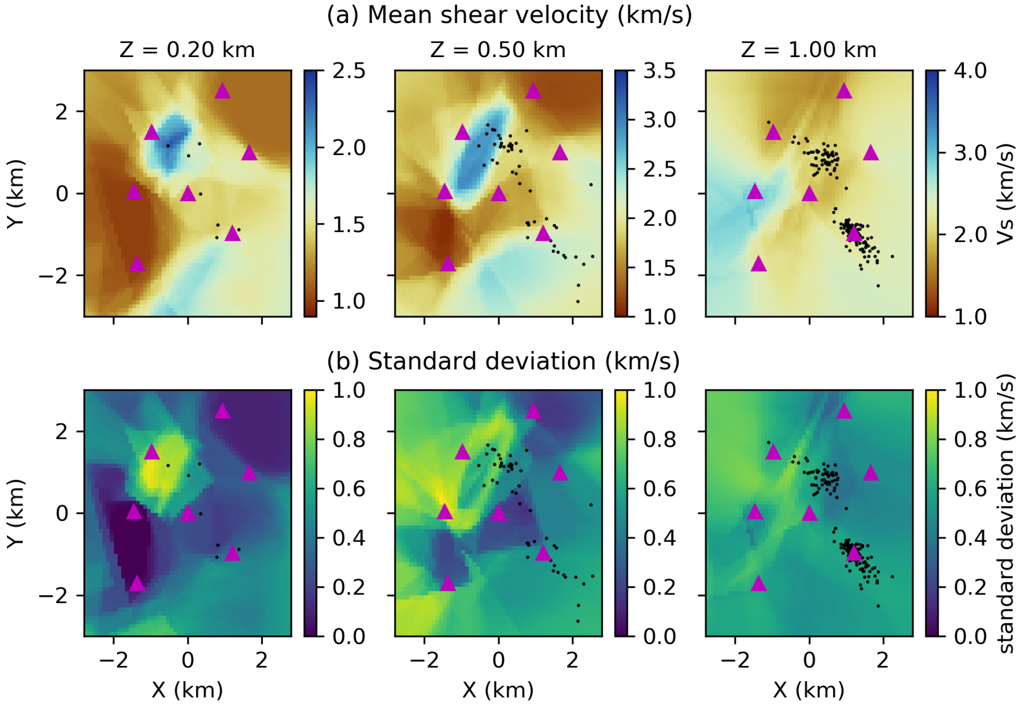


Figure 4.6: Horizontal slices through the 3D shear velocity model obtained using surface wave dispersion data only. Key as in Figure 4.5.

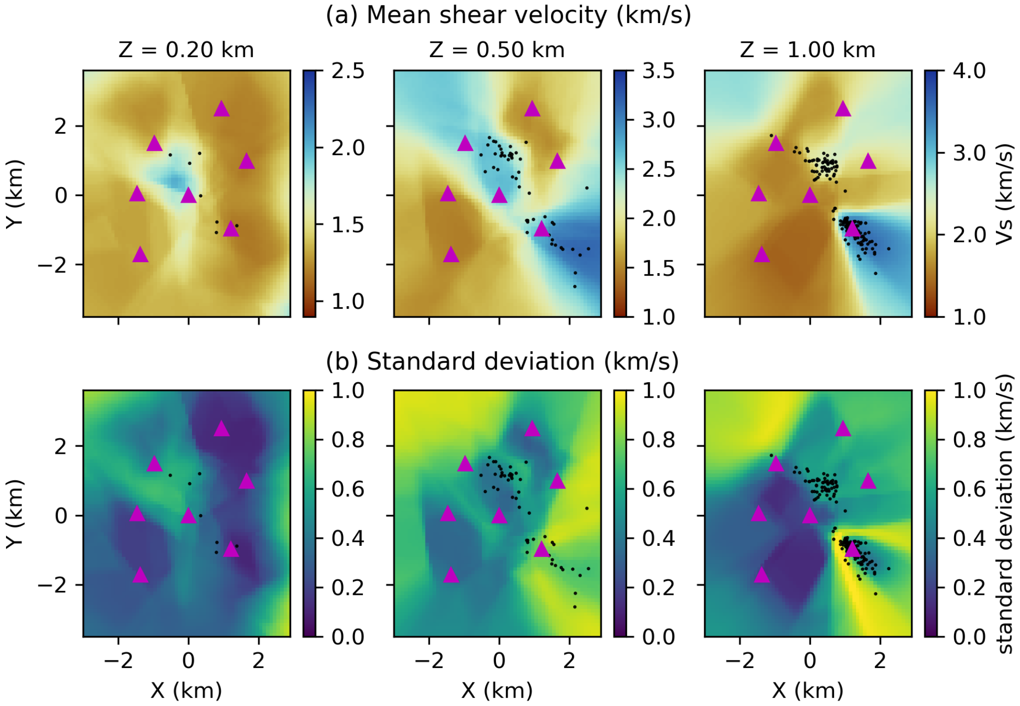


Figure 4.7: Horizontal slices through the 3D shear velocity model obtained from joint body and surface wave inversion. Key as in Figure 4.5.

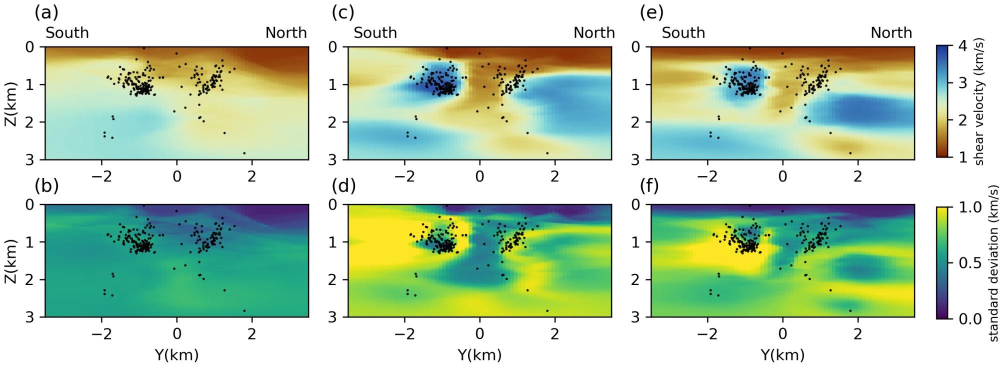


Figure 4.8: Cross sections of the mean (top) and standard deviation (bottom) at X=1 km obtained using surface wave tomography (**a** and **b**), body wave tomography (**c** and **d**) and joint body and surface wave inversion (**e** and **f**). Black dots are events lying within 0.8 km of the cross-section.

uncertainty around the southeastern high velocity anomaly and around the low velocity anomaly between the two clusters. These loops represent uncertainty due to the trade-off between the velocity and the location of velocity anomalies, and hence describe uncertainty in the anomalies' shapes (Galetti et al., 2015; Zhang et al., 2018).

Figure 4.6 shows horizontal slices through the mean and standard deviation obtained from surface wave tomography at the same depths as above. Compared to the results from body wave tomography, the mean shear velocity map at 0.2 km shows lower velocities (about 1.0 km/s) than the results from body wave tomography ($>$ 2.0 km/s), and the standard deviation is also much lower (about 0.2 km/s) than that from body wave tomography (about 1.0 km/s). This is due to the fact that surface waves are more sensitive to near surface structure than body waves. There is a higher velocity anomaly in the northwest which is probably caused by poor data coverage (Figure 4.1). At 0.5 km depth the mean velocity map shows similar patterns of structure to those obtained from body wave tomography: the velocity in the southwest and in the northeast is lower and the velocity in the southeast is higher. The mean velocity map at 1 km depth shows very different results compared to those from body wave tomography and its standard deviation is higher (about 0.6 km/s). This is probably caused by the fact that the frequency range of the surface waves used in the inversion has very low sensitivity at this depth.

Figure 4.7 shows horizontal slices through the mean and standard deviation obtained using joint inversion. Similar to the results of surface wave tomography, the mean velocity map at 0.2 km shows lower velocity values than those from body wave tomography with lower standard deviations: near surface structure can be better constrained by including surface wave dispersion data in the inversion. There is still a higher velocity anomaly between Y=0 km and Y=2 km which is associated with high standard deviations: neither body waves nor surface waves have much resolution in this area so the velocity tends towards the mean of the prior pdf (2.5 km/s). The mean velocity maps at 0.5 km and 1.0 km are very similar to the results from body wave tomography: we have more body wave data than surface wave data that are sensitive to these depths so the body wave data dominate the results. Nevertheless, the velocity magnitudes are slightly different from the results of body wave tomography which is due to the contribution of surface waves, and the standard deviation map shows lower uncertainties within the station array which suggests that surface waves improve the resolution across that entire area. Similarly to the results of body wave tomography, the standard deviation map also shows a higher uncertainty loop around the southeastern high velocity anomaly.

In Figure 4.8 we show vertical cross sections through the mean and standard deviation maps from the three inversions along the X=1 km profile which lies between the two earthquake clusters. The mean velocity model from surface wave tomography (Figure 4.8a) shows that there is a low velocity anomaly between the two clusters. The standard deviation model (Figure 4.8b) shows that the near surface structure ($<$ 0.8 km) is well constrained while the deeper part has very limited resolution. Figure 4.8c and d show the mean and standard deviation cross sections from body wave tomography. The velocity model also shows a low velocity anomaly between the two clusters, however the low velocity anomaly extends to deeper levels and the velocity at either side of the low velocity anomaly is much higher ($>$ 3km/s) than that from surface wave tomography ($\sim$ 2.0 km/s). The standard deviation model shows a low uncertainty area associated with the middle low velocity anomaly suggesting that the anomaly is well determined. There are also higher uncertainty loops around the high velocity anomalies at the two sides of the low velocity anomaly. Figure 4.8e and f show the results from joint inversion. The mean model is similar to that from surface wave tomography at shallow levels, and to that from body wave tomography at depth. However the velocity magnitude of the southern high velocity anomaly is lower than that from body wave tomography, and the velocity of the northern low velocity anomaly around 1 km depth is much lower than that from body wave tomography, both due to the contribution of surface waves. Similarly the standard deviation model shows lower uncertainties at the near surface, and higher uncertainty loops around high velocity anomalies.

Figure 4.9 shows the horizontal slices of the mean and standard deviation of P-wave velocity using body waves only. Similar to the results for shear velocity, at the depth of 0.2 km higher velocities are associated with higher uncertainties since the near surface structure cannot be well constrained by body waves, and at 0.5 km depth the mean Vp velocity model exhibits very similar patterns to those of shear velocity. Similarly to the results for shear velocity, the standard deviation map shows higher uncertainties at the location of the southeastern higher velocity anomaly due to limited data coverage. At the depth of 1.0 km the mean velocity map also shows similar structures to those in the shear velocity results, and the standard deviation map shows higher uncertainty loops around velocity anomalies.

For comparison, we show the results of P-wave velocity from joint inversion in Figure 4.10. The P-wave velocity model at the depth of 0.2 km is better constrained by including surface wave dispersion data: most of the model has lower velocities (< 3 km/s) compared to those from body wave tomography (~ 4 km/s) and lower uncertainties (< 0.5 km/s) than those from body wave tomography (> 1.0 km/s). This is due to the fact that shear velocity is better estimated by including surface waves, so the P-wave velocity can also be better constrained since P and S velocities are correlated through the common earthquake source parameters, and surface waves are also partly sensitive to P-wave velocity at near surface (Zhou et al., 2004; Fang et al., 2016). At greater

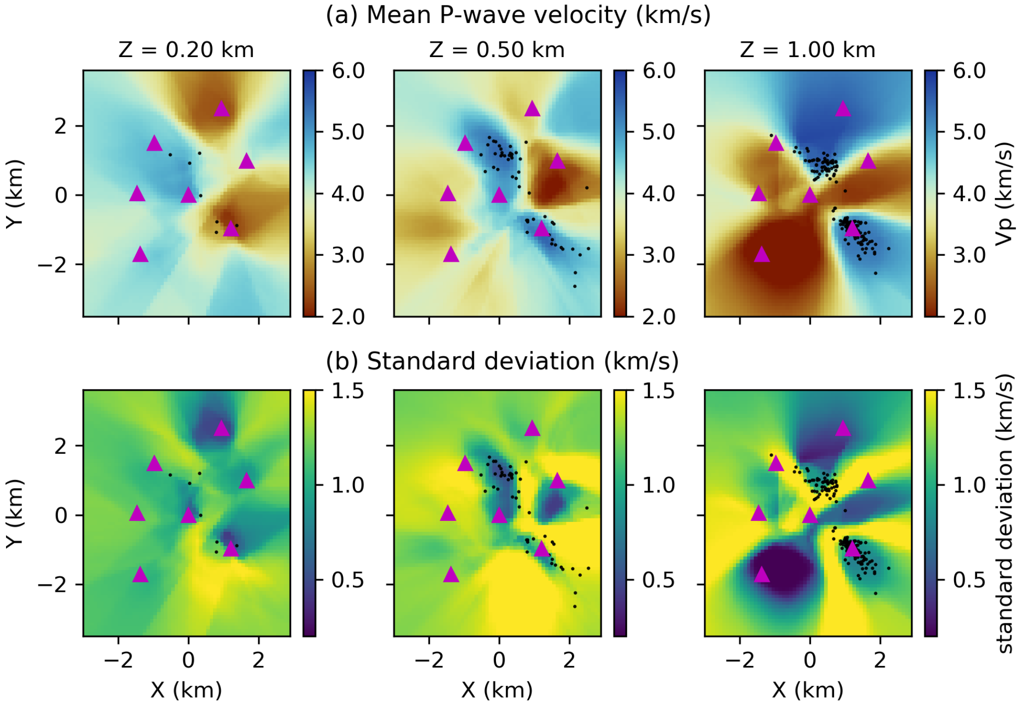


Figure 4.9: Horizontal slices through the 3D P-wave velocity model obtained using body wave travel time only. Key as in Figure 4.5.

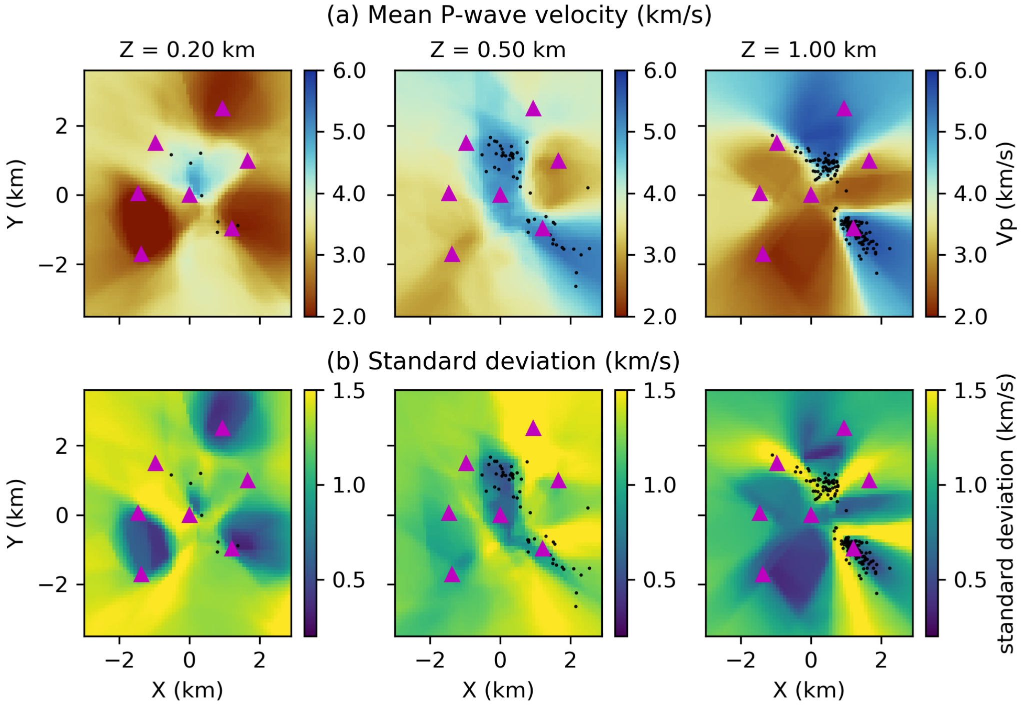


Figure 4.10: Horizontal slices through the 3D P-wave velocity model obtained from joint body and surface wave inversion. Key as in Figure 4.5.

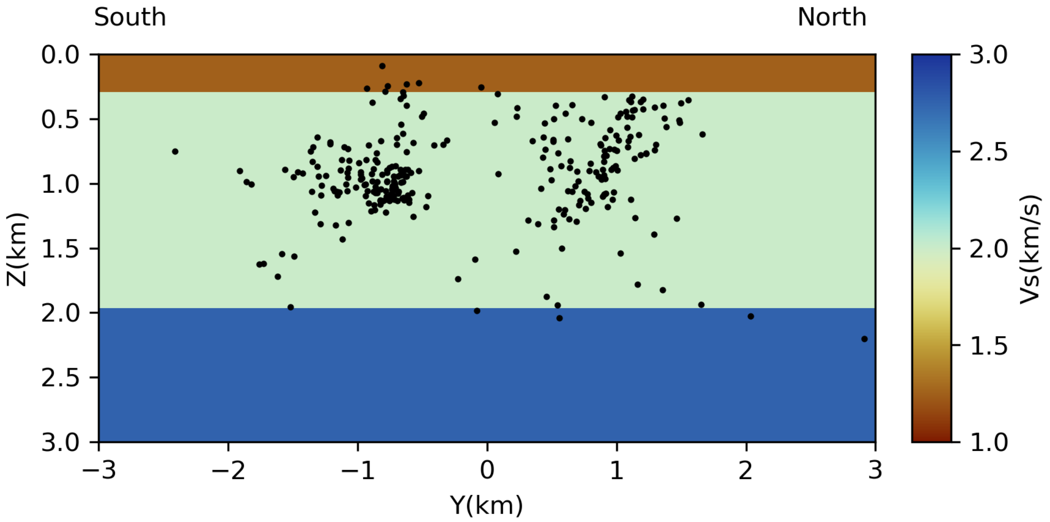


Figure 4.11: Cross section of the synthetic model at X=1 km. Black dots show the event locations which are taken from joint inversion of the real data in Figure 4.3d.

depths (0.5 km and 1.0 km) the mean velocity model is similar to that from body wave tomography since surface waves have very little sensitivity to P-wave velocity structure there.

4.6 Discussion

In the above results there is a high velocity anomaly at the location of the southern earthquake cluster (Figure 4.5, 4.7, 4.8): in the results from joint inversion the magnitude of the velocity anomaly is slightly lower, but is nevertheless clearly identifiable. Similar features have been observed previously and are generally interpreted as earthquake asperities that concentrate stress (Lees, 1990; Eberhart-Phillips & Michael, 1998; Chiarabba & Amato, 2003; Tajima et al., 2009; Li et al., 2013; Zhang et al., 2013). However it is also possible that this correlation is caused by the trade off between source parameters and velocity values.

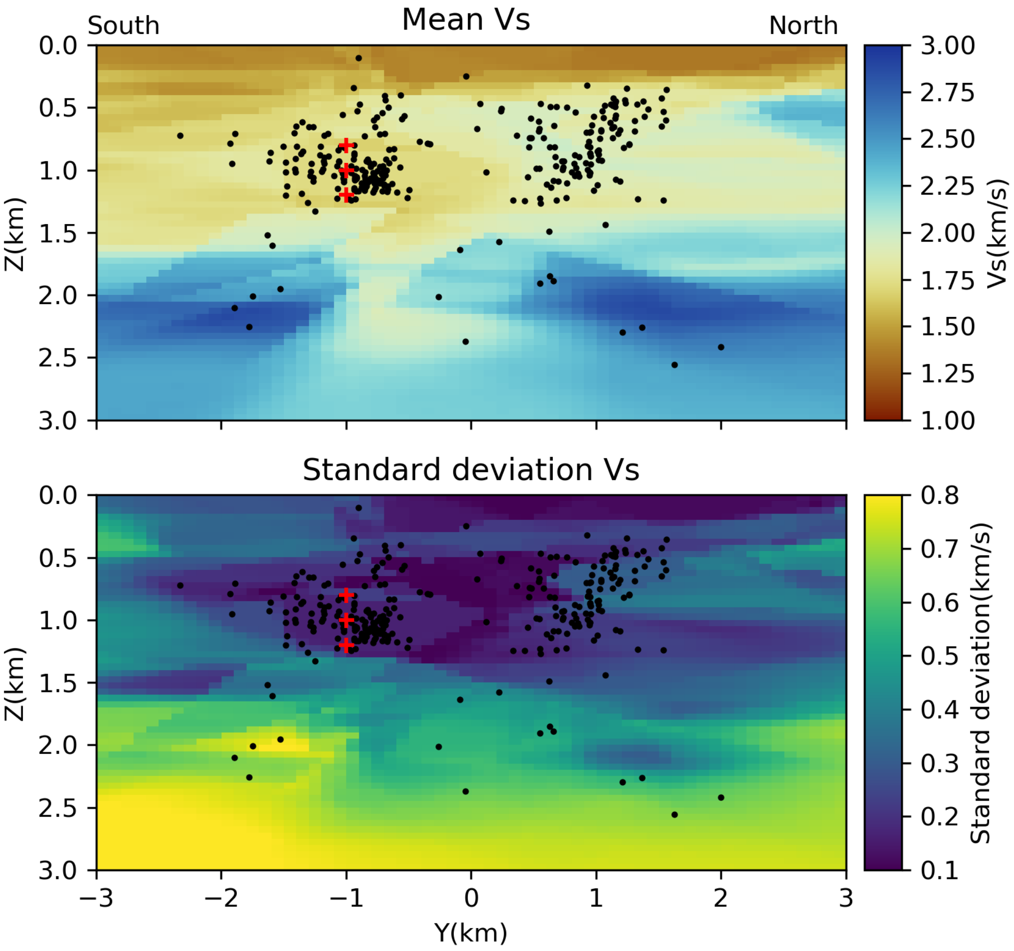


Figure 4.12: Cross sections of the mean and standard deviation at X=1 km obtained by inverting for source parameters and velocity model simultaneously using body wave data only. Black dots show the mean event locations. The red pluses show point locations which are referred to in the text.

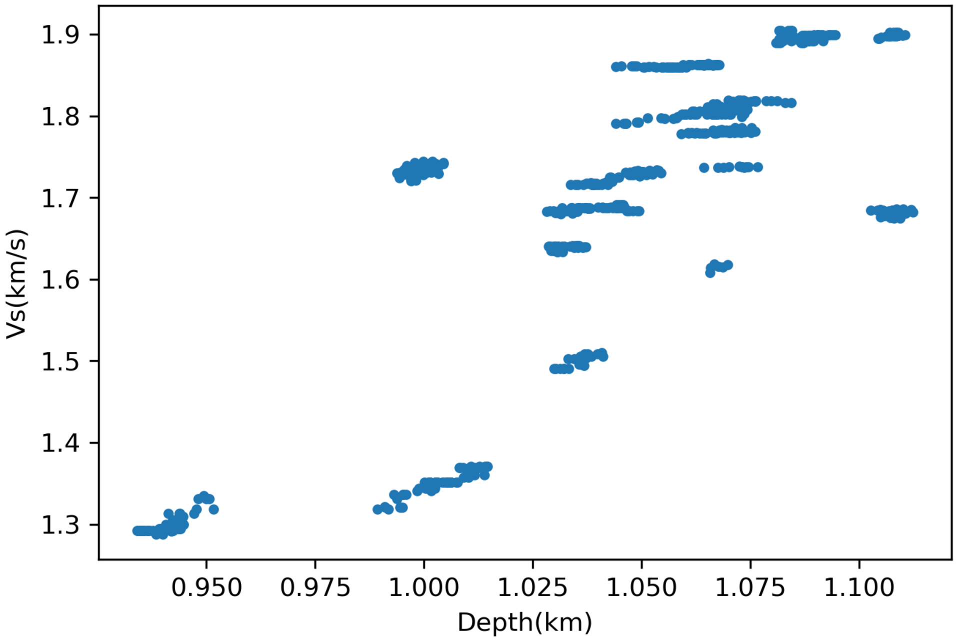


Figure 4.13: Average shear velocity at the location of the southern cluster versus average depth of events of the southern cluster.

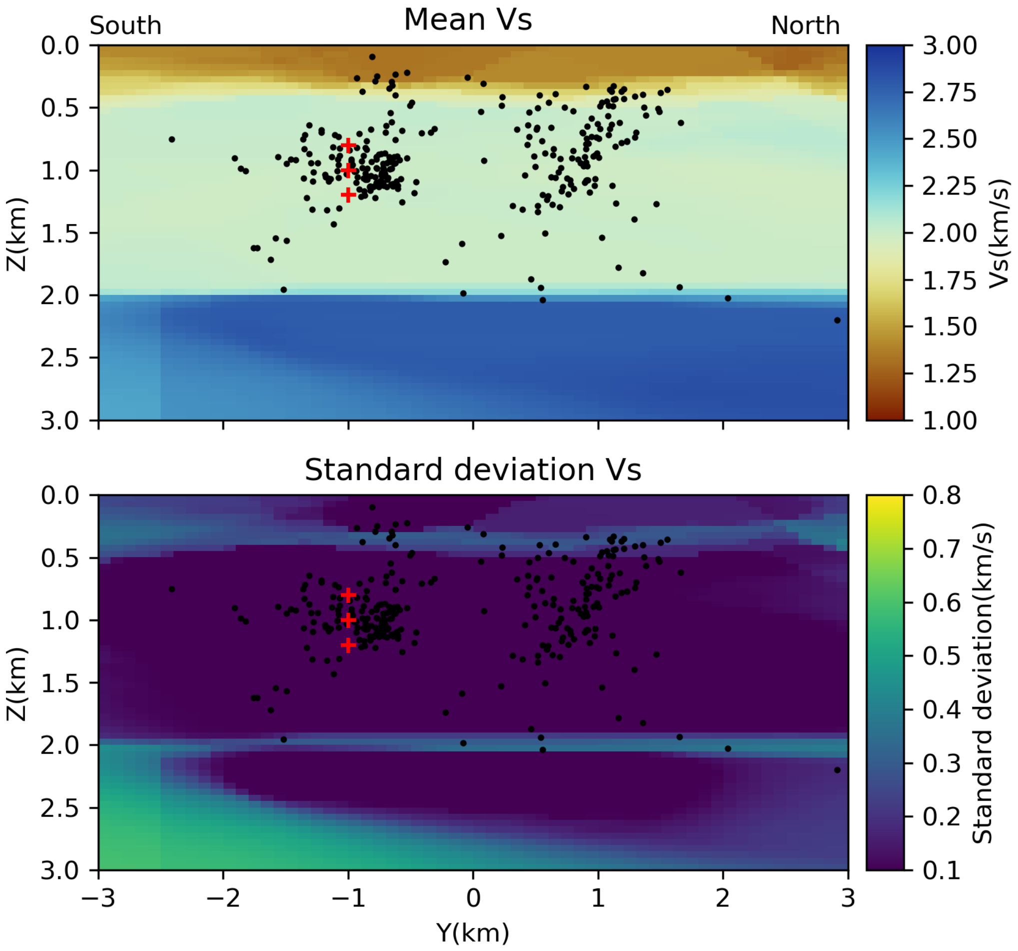


Figure 4.14: Cross sections of the mean and standard deviation at X=1 km obtained by inverting for velocities, and fixing the source parameters at their true values. Black dots show the event locations. Red pluses show point locations which are referred to in the text.

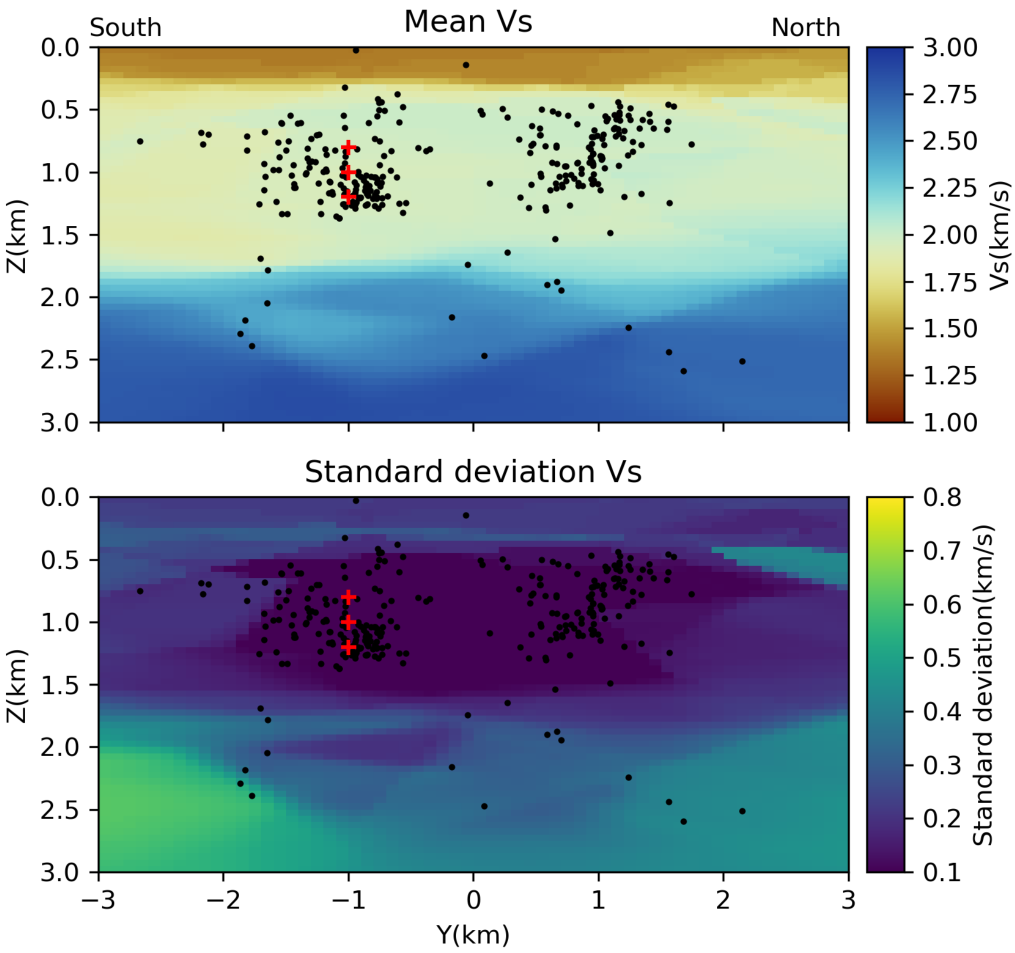


Figure 4.15: Cross sections of the mean and standard deviation at X=1 km obtained by inverting for both velocity and event locations using both body wave and surface wave data. Black dots show the mean event locations. Red pluses show point locations which are referred to in the text.

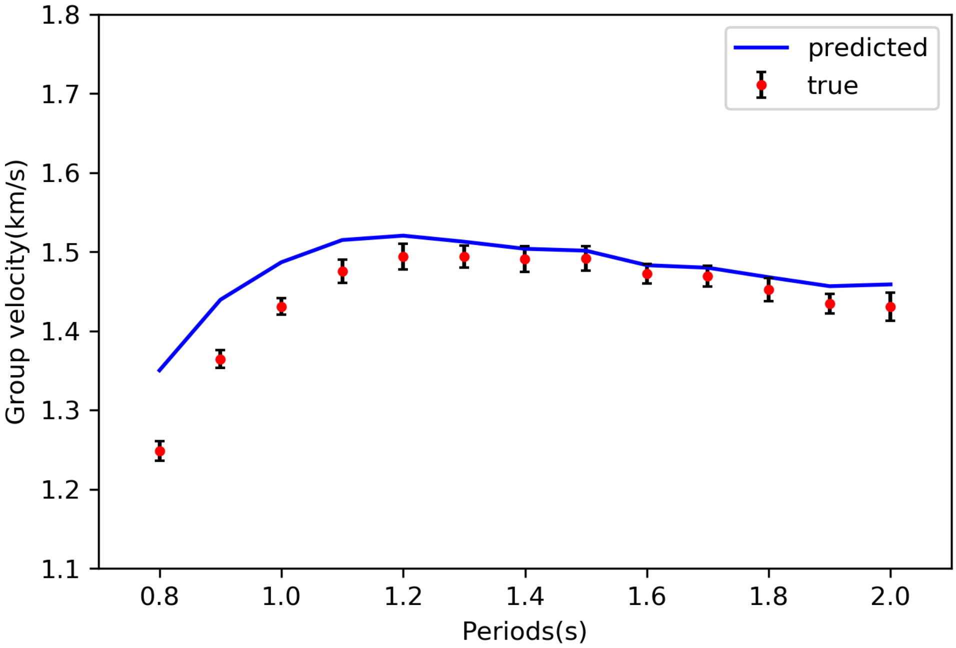


Figure 4.16: Group velocities used in the joint inversion (red dot) plotted with error bars and the average dispersion curve calculated from the mean velocity model (blue line). Since the true model is a 1D model, dispersion curves between different receiver pairs are almost the same except for random noise. Error bars show the standard deviation of group velocities of different dispersion curves.

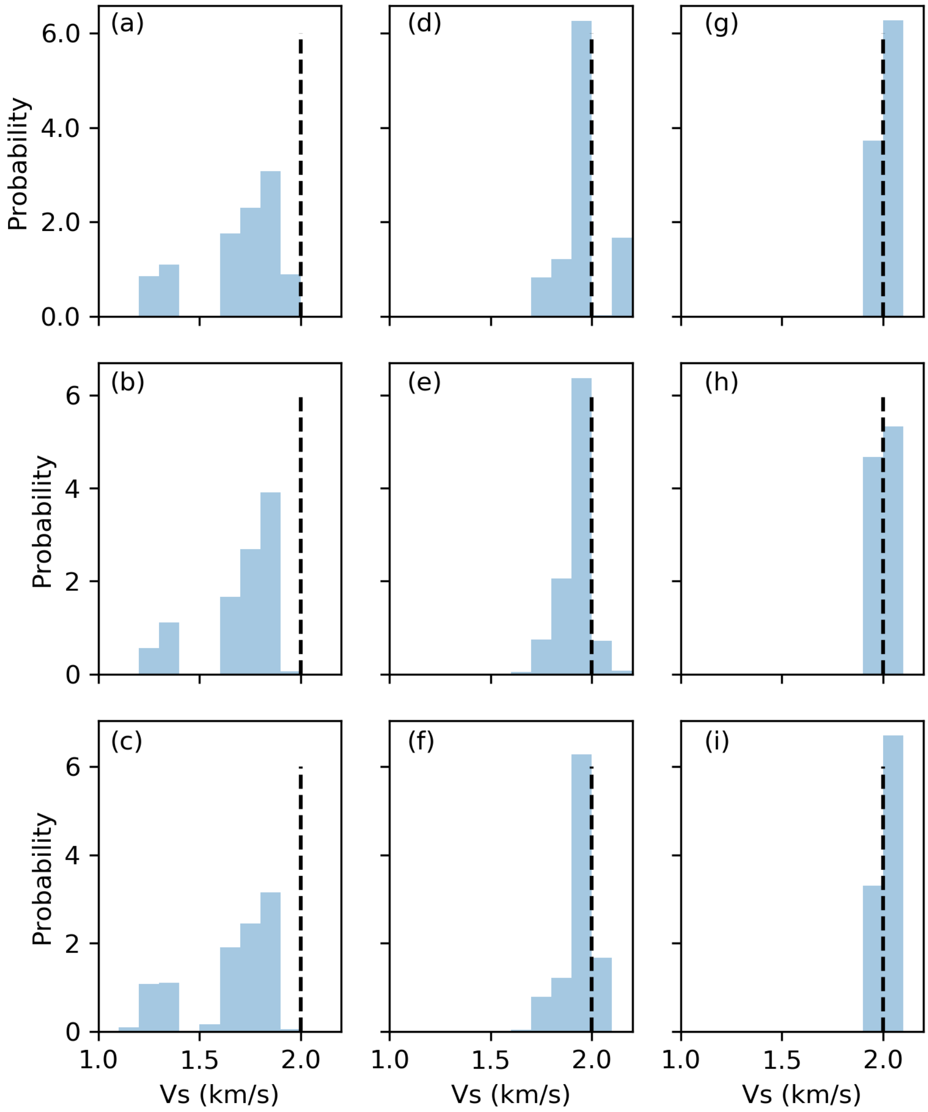


Figure 4.17: The marginal posterior pdfs of shear velocity at three points (pluses in Figure 4.12, 4.14 and 4.15). **(a)**, **(b)** and **(c)** show the marginal posterior pdfs at three points at depth 0.8 km, 1.0 km and 1.2 km obtained by inverting source parameters and velocity model simultaneously using body wave data. **(d)**, **(e)** and **(f)** show the marginal posterior pdfs at three points obtained by joint inversion using both body wave data and surface wave data. **(g)**, **(h)** and **(i)** show the marginal posterior pdfs at three points obtained by fixing the source parameters. The dashed black line denotes the true shear velocity value.

To better understand the correlation of the high velocity anomaly and the earthquake cluster we performed a simple synthetic test in which the "true" model contains three layers and event locations are taken from the joint inversion above (Figure 4.11). We computed synthetic versions of the same body wave travel time data and surface wave dispersion data as used in the above inversion, and added 1 percent noise to the data. We then conducted three inversions: simultaneous inversion for source parameters and velocity model using body wave data, inversion for velocity only using body wave data with sources fixed at their true locations, and joint inversion for sources and velocities using both types of data. The initial source parameters (event locations and origin times) are the same as were used in the real data inversions above. The prior pdfs are chosen to be the same as in the real data example except that the prior for the number of cells is chosen to be a Uniform distribution between 5 and 100 since the model is relatively simple. The proposal pdfs are also tuned in the same way as above. The burn-in and total samples for each chain and the number of chains are also set to be the same as in the real data inversions.

Figure 4.12 shows cross sections of the mean and standard deviation at X=1 km obtained using body wave tomography by simultaneously inverting for source parameters and the velocity model. Though the mean velocity model shows three layers which are to some extent similar to the true model, the velocity value around the southern cluster (around 1.75 km/s) is lower than the true value (2.0 km/s). This suggests that body wave tomography may produce biases in the results around the location of event clusters, caused by the trade off between event locations and velocity values (see Figure 4.13): shallower event depths are generally associated with a lower velocity value to fit the data. The standard deviation model shows low uncertainties from the surface down to around 1.5 km including in the low velocity areas around the southern cluster. This low velocity anomaly is due to the fact that the initial source locations are shallower than the true locations, so in order to fit the data the model decreases the velocity value at the location of event clusters (see Figure 4.13); this trade off creates complex multimodality in the posterior pdf (see Figure 4.17a, b and c below), and since random walk McMC algorithms are generally inefficient for multimodal distributions the chains likely get stuck in modes that have lower velocities. By contrast, in Figure 4.14 we show the results obtained from an inversion with source parameters fixed at their true values. The mean velocity model shows almost the same structure as the true model which again suggests that the non-uniquenesses in the posterior velocity pdfs in the previous results are caused by non-linear trade off between source parameters and velocity values. The standard deviation shows very low uncertainties (< 0.2 km/s) across the whole section except in a small area in the left corner where there are no events. It also shows slightly higher uncertainties at the boundaries between layers which reflect uncertainty in layer boundary locations similar to the uncertainty loops observed above (Zhang et al., 2018). To give an idea of fit to the data, the simultaneous inversion of source parameters and velocity model produces an average residual of 0.81 s while the fixed-source inversion produces a residual of 0.80 s. Thus the two inversions produce almost the same average fit to the data even though they give different estimates of the velocity model; therefore one cannot discriminate the two models based on data fit.

Figure 4.15 shows cross sections of the mean and standard deviation obtained using joint inversion of both data types for both velocities and source parameters. Though the mean velocity model shows slightly different velocity values in the second and third layer compared to the previous two models, it is significantly closer to the true model than that obtained from body wave tomography by simultaneously inverting for source parameters and the velocity model. The standard deviation model shows similar structures to those from the fixed-source inversion, including higher uncertainties at the boundary of layers. Figure 4.16

shows the true dispersion curve and the average dispersion curve calculated using the mean velocity model. At longer periods (> 1.2 s) the average group velocities fit the true values, whereas at shorter periods the average group velocities are higher than true values which is probably caused by the bias produced by body wave data (see Figure 4.12), or because the mean model is not a good representation of the subsurface structure (Zhang & Curtis, 2019).

To better understand the results, in Figure 4.17 we show marginal posterior pdfs obtained using the three methods at three different points (1, -1, 0.8), (1, -1, 1.0) and (1, -1, 1.2) which cross the southern earthquake cluster in the above cross sections. The marginal distributions obtained from body wave tomography show complex multimodal distributions (Figure 4.17a, b and c) and are distributed away from the true value (2 km/s). By contrast, most of the marginal distributions obtained from joint inversion show a unimodal distribution concentrated around the true value (Figure 4.17 e and f) other than a subtle multimodality in Figure 4.17d. The marginal distributions obtained from fixed-source inversion focus to a unimodal distribution around the true value (Figure 4.17g, h and i). Thus, the simultaneous inversion for source parameters and velocity model using only body wave data can bias the results due to the trade off between source parameters and velocity values. By including surface wave dispersion data in the inversion, this problem can be resolved since surface wave data improve the velocity estimate.

Apart from joint inversion using both types of data, the results obtained using surface wave tomography are also used as prior information for body wave tomography to produce a more realistic velocity model. For example, velocity models from surface wave tomography were used as a starting model for body wave tomography (Rawlinson & Fishwick, 2012; Nunn et al., 2013). However in those studies since surface wave data cannot constrain velocity models, trade-offs between source parameters and velocity models could still bias the results and the resulting model can be inconsistent with surface wave data. Thus it is better to jointly invert for a unified model using both types of data.

In the real data results, the high velocity anomaly at the location of the southern cluster therefore may reflect the true structure of the subsurface, e.g., earthquake asperities following previous interpretations (Lees, 1990; Eberhart-Phillips & Michael, 1998; Chiarabba & Amato, 2003; Tajima et al., 2009; Li et al., 2013; Zhang et al., 2013). However, since we still observe subtle multimodilities in the joint inversion results, and the real Earth may have a more complex structure, there is still the possibility that the details of the recovered model are obscured by the trade off between source parameters and velocity models. The synthetic test suggests that the trade off mainly affects the velocity structure at the location of the southern earthquake cluster, so our results at least remain valid for most of the subsurface.

The rj-McMC algorithm is generally not efficient for exploring complex multimodal distributions (Green & Hastie, 2009). In our body wave tomography synthetic test, by simultaneously inverting for source parameters and velocity models the chains may have got stuck at local modes and failed to find the true solution. To reduce this issue, one could use a good initial model (as required by linearised inversion) to ensure that the chains explore around the globally optimal solution. In the real data inversions we used initial source parameters obtained using a 1D initial model, so to further improve the results one could use a good 3D model to estimate better initial locations. If no better model exists (as is often the case) then that is not an option, so methods that are efficient for multimodal distributions can be used, for example grid search (Sen & Stoffa, 2013), non-Markovian importance sampling (Lomax & Curtis, 2001; Lomax et al., 2009) and variational inference methods (see Chapter 5).

4.7 Conclusion

We implemented a Monte Carlo method to perform joint inversion using both body wave arrival time data and surface wave dispersion data, and applied it at a mining site located to the north of New Ollerton, Nottinghamshire, U.K, at which induced seismicity occurs. The results show that by including surface wave dispersion data the shallow structure can be better constrained because surface waves are generally sensitive to the shallow structure, and this further improves estimate of source parameters. We also observed a high velocity anomaly which is correlated with one of the event clusters in the results. To further understand this correlation, we performed inversions using synthetic data generated using the same source and receiver distribution as in the real data experiment. The results show that due to the trade off between source parameters and velocity values, the inversion using only body wave data can produce biases; by including surface wave dispersion data in the inversion the problem can be resolved. We conclude that it is better to include surface wave data in seismic travel time tomographic inversions.

Chapter 5

Seismic tomography using variational inference methods

5.1 Summary

In previous chapters, we used Monte Carlo sampling methods to solve seismic tomographic problems, but they are generally computationally intractable for large datasets and high-dimensional parameter spaces. In this chapter to extend uncertainty analysis to larger systems we use variational inference methods to conduct seismic tomography. In contrast to Monte Carlo sampling, variational methods solve the Bayesian inference problem as an optimization problem, yet still provide probabilistic results. In this study, we applied two variational methods, automatic differential variational inference (ADVI) and Stein variational gradient descent (SVGD), to 2D seismic tomography problems using both synthetic and real data and we compare the results to those from two different Monte Carlo sampling methods. The results show that variational inference methods can

produce accurate approximations to the results of Monte Carlo sampling methods at significantly lower computational cost, provided that gradients of parameters with respect to data can be calculated efficiently. We expect that the methods can be applied fruitfully to many other types of geophysical inverse problems.

5.2 Introduction

In a variety of geoscientific applications, scientists need to obtain maps of subsurface properties in order to understand heterogeneity and processes taking place within the Earth. Seismic tomography is a method that is widely used to generate those maps. The maps of interest are usually parameterised in some way, and data are recorded that can be used to constrain the parameters. Tomography is therefore a parameter estimation problem, given the data and a physical relationship between data and parameters; since the physical relationships usually predict data given parameter values but not the reverse, seismic tomography involves solving an inverse problem (Curtis & Snieder, 2002).

Tomographic problems can be solved using either the full, known physical relationships, or by first creating approximate, linearised physics. In the linearised case, one usually seeks an optimal solution by minimizing the misfits between predicted data and observed data by iteratively linearising the physics around the current best model and solving the linear system to update that model estimate. Since most tomography problems are under-determined, some form of regularization must be introduced to solve the system (Aki & Lee, 1976; Dziewonski & Woodhouse, 1987; Iyer & Hirahara, 1993; Tarantola, 2005). However, regularization is usually chosen using ad hoc criteria which introduces poorly understood biases in the results; thus, valuable information can be concealed by regularization (Zhdanov, 2002). Moreover, in nonlinear problems it is almost always impossible to estimate accurate uncertainties in results using linearised methods. Therefore, partially or fully nonlinear tomographic methods have been introduced to geophysics which require no linearisation and which provide accurate estimates of uncertainty using a Bayesian probabilistic formulation of the parameter estimation problem. These include Monte Carlo methods (Mosegaard & Tarantola, 1995; Sambridge, 1999; Malinverno et al., 2000; Malinverno, 2002; Malinverno & Briggs, 2004; Bodin & Sambridge, 2009; Galetti et al., 2015, 2017; Zhang et al., 2018) and methods based on neural networks (Röth & Tarantola, 1994; Devilee et al., 1999; Meier et al., 2007a,b; Shahraeeni & Curtis, 2011; Shahraeeni et al., 2012; Käufl et al., 2013, 2015; Earp & Curtis, 2019).

Bayesian methods use Bayes' theorem to update a *prior* probability distribution function (*pdf* – either a conditional density function or a discrete set of probabilities) with new information from data. The prior pdf describes information available about the parameters of interest prior to the inversion. Bayes' theorem combines the prior pdf with information derived from the data to produce the total state of information about the parameters post inversion, described by a so-called *posterior* pdf – this process is referred to as Bayesian inference. Thus, in our case Bayesian inference is used to solve the tomographic inverse problem.

Monte Carlo methods generate a set (or chain) of samples from the posterior pdf describing the probability distribution of the model given the observed data; thereafter these samples can be used to estimate useful information about that pdf (mean, standard deviation, etc.). The methods are quite general from a theoretical point of view so that in principle they can be applied to any tomographic problems. They have been extended to trans-dimensional inversion using the reversible jump Markov chain Monte Carlo (rj-McMC) algorithm (Green, 1995), in which the number of parameters (hence the dimensionality of parameter space) can vary in the inversion. Consequently the parameterization itself can be simplified by adapting to the data which improves results on otherwise high-dimensional problems (Malinverno et al., 2000; Bodin & Sambridge, 2009; Bodin et al., 2012; Ray et al., 2013; Young et al., 2013; Galetti et al., 2015, 2017; Hawkins & Sambridge, 2015; Piana Agostinetti et al., 2015; Burdick & Lekić, 2017; Galetti & Curtis, 2018; Zhang et al., 2018, 2019). Although many applications have been conducted using McMC sampling methods (previous references; Shen et al., 2012, 2013; Zulfakriza et al., 2014; Zheng et al., 2017; Crowder et al., 2019), they mainly address 1D or 2D tomography problems due to the high computational expense of Monte Carlo methods. Some studies used McMC methods for fully 3D tomography using body wave travel time data (Hawkins & Sambridge, 2015; Piana Agostinetti et al., 2015; Burdick & Lekić, 2017) and surface wave dispersion (Zhang et al., 2018, 2019), but the methods demand enormous computational resources. Even in the 1D or 2D case, McMC methods cannot easily be applied to large datasets which are generally expensive to forward model given a set of parameter values. Moreover, McMC methods tend to be inefficient at exploring complex, multi-modal probability distributions (Sivia, 1996; Karlin, 2014), which appear to be common in seismic tomography problems.

Neural network based methods offer an efficient alternative for certain classes of tomography problems that will be solved many times with new data of the same type. An initial set of Monte Carlo samples is taken from the prior probability distribution over parameter space, and data are computationally forward modelled for each parameter vector. Neural networks are flexible mappings that can be regressed (trained) to emulate the mapping from data to parameter space by fitting the set of examples of that mapping generated using Monte Carlo (Bishop, 2006). The trained network then interpolates the inverse mapping between the examples, and can be applied efficiently to any new, measured data to estimate corresponding parameter values. The first geophysical application of neural network tomography was Röth & Tarantola (1994), but that application did not estimate uncertainties. Forms of networks that estimate tomographic uncertainties were introduced by Devilee et al. (1999) and Meier et al. (2007a,b) and have been applied to surface and body wave tomography in 1D and 2D problems (Meier et al., 2007a,b; Earp & Curtis, 2019). Nevertheless, neural networks still suffer from the computational cost of generating the initial set of training examples. That set may have to include many more samples than are required for standard Bayesian MC, because the training set must span the prior pdf whereas standard applications of MC tomography sample the posterior pdf which is usually more tightly constrained. Neural networks have the advantage that the training samples need only be calculated once for any number of data sets whereas MC inversion must perform sampling for every new data set. However, in high dimensional problems the cost of sampling may be prohibitive for both MC and NN based methods due to the curse of dimensionality (the exponential increase in the hypervolume of parameter space as the number of parameters increases – Curtis & Lomax, 2001).

Variational inference provides a different way to solve a Bayesian inference problem: within a predefined family of probability distributions, one seeks an optimal approximation to a target distribution which in this case is the Bayesian posterior pdf. This is achieved by minimizing the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951) – a measure of the difference between the approximate and target pdfs (Bishop, 2006; Blei et al., 2017). Since the method casts the inference problem into an optimization problem, it can be computationally more efficient than either MC sampling or neural network methods, and provides better scaling to higher dimensional problems. Moreover, it can be used to take advantage of methods such as stochastic optimization (Robbins & Monro, 1951; Kubrusly & Gravier, 1973) and distributed optimization by dividing large datasets into random minibatches – methods which are difficult to apply for McMC methods since they may break the reversibility property of Markov chains which is required by most McMC methods.

In variational inference, the complexity of the approximating family of pdfs determines the complexity of the optimization. A complex variational family is generally more difficult to optimize than a simple family. Therefore, many applications are performed using simple mean-field approximation families (Bishop, 2006; Blei et al., 2017) and structured families (Saul & Jordan, 1996; Hoffman & Blei, 2015). For example, in Geophysics the method has been used to invert for the spatial distribution of geological facies given seismic data using a mean-field approximation (Nawaz & Curtis, 2018, 2019).

Even using those simple families, applications of variational inference methods usually involve tedious derivations and bespoke implementations for each type of problem which restricts their applicability (Bishop, 2006; Blei et al., 2017; Nawaz & Curtis, 2018, 2019). The simplicity of those families also affects the quality of the approximation to complex distributions. To make variational methods easier to use, "black box" variational inference methods have been proposed (Kingma & Welling, 2013; Ranganath et al., 2014, 2016). Based on these ideas, Kucukelbir et al. (2017) proposed an automatic variational inference method which can easily be applied to many Bayesian inference problems. Another set of methods has been proposed based on probability transformations (Rezende & Mohamed, 2015; Tran et al., 2015; Liu & Wang, 2016; Marzouk et al., 2016); these methods optimise a series of invertible transforms to approximate the target probability and in this case it is possible to approximate arbitrary probability distributions.

We apply automatic differential variational inference (ADVI – Kucukelbir et al., 2017) and Stein variational gradient descent (SVGD – Liu & Wang, 2016) to a 2D seismic tomography problem. In the following we first describe the basic idea of variational inference, and then the ADVI and SVGD methods. In section 3 we apply the two methods to a simple 2D synthetic seismic tomography example and compare their results with both fixed-dimensional McMC and rj-McMC. In section 4 we apply the two methods to real data from Grane field, North Sea, to study the phase velocity map at 0.9 s and compare the results to those found using rj-McMC. We thus demonstrate that variation inference methods can provide efficient alternatives to McMC methods while still producing reasonably accurate approximations to Bayesian posterior pdfs. Our aim is to introduce variational inference methods to the geoscientific community and to encourage more research on this topic.

5.3 Methods

5.3.1 Variational inference

Bayesian inference involves calculating or characterising a posterior probability density function $p(\mathbf{m}|\mathbf{d}_{obs})$ of model parameters $\mathbf{m}$ given the observed data $\mathbf{d}_{obs}$. According to Bayes' theorem,

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})} \tag{5.1}$$

where $p(\mathbf{d}_{obs}|\mathbf{m})$ is called the *likelihood* which is the probability of observing data $\mathbf{d}_{obs}$ if model $\mathbf{m}$ was true, $p(\mathbf{m})$ is the *prior* which describes information that is independent of the data, and $p(\mathbf{d}_{obs})$ is a normalization factor called the *evidence* which is constant for a fixed model parameterization. The likelihood is usually assumed to follow a Gaussian probability density function around the data predicted synthetically from model $\mathbf{m}$ (using the known physical relationships), as this is assumed to be a reasonable approximation to the pdf of uncertainties or errors in the measured data, and because noise reduction is performed by stacking, which through the central limiting theorem implies the use of a Gaussian distribution.

Variational inference approximates the above pdf $p(\mathbf{m}|\mathbf{d}_{obs})$ using optimization. First a family (set) of known distributions $\mathcal{Q} = \{q(\mathbf{m})\}$ is defined. The method then seeks the best approximation to $p(\mathbf{m}|\mathbf{d}_{obs})$ within that family by minimizing the KL-divergence:

$$\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] = \mathrm{E}_q[\mathrm{log}q(\mathbf{m})] - \mathrm{E}_q[\mathrm{log}p(\mathbf{m}|\mathbf{d}_{obs})] \tag{5.2}$$

where the expectation is taken with respect to distribution $q(\mathbf{m})$. It can be shown that $\mathrm{KL}[q||p] \geq 0$ and has zero value if and only if $q(\mathbf{m})$ equals $p(\mathbf{m}|\mathbf{d}_{obs})$ (Kullback & Leibler, 1951). Distribution $q^*(\mathbf{m})$ that minimizes the KL-divergence is therefore the best approximation to $p(\mathbf{m}|\mathbf{d}_{obs})$ within the family $\mathcal{Q}$.

Combining equations (5.1) and (5.2), the KL-divergence becomes:

$$\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] = \mathrm{E}_q[\mathrm{log}q(\mathbf{m})] - \mathrm{E}_q[\mathrm{log}p(\mathbf{m},\mathbf{d}_{obs})] + \mathrm{log}p(\mathbf{d}_{obs}) \tag{5.3}$$

The evidence term $\mathrm{log}p(\mathbf{d}_{obs})$ generally cannot be calculated since it involves the evaluation of a high dimensional integral which takes exponential time. Instead we calculate the evidence lower bound (ELBO) which is equivalent to the KL-divergence up to an unknown constant, and is obtained by rearranging equation (5.3) and using the fact that $\mathrm{KL}[q||p] \geq 0$:

$$\begin{aligned} \mathrm{ELBO}[q] &= \mathrm{E}_q[\mathrm{log}p(\mathbf{m},\mathbf{d}_{obs})] - \mathrm{E}_q[\mathrm{log}q(\mathbf{m})] \\ &= \mathrm{log}p(\mathbf{d}_{obs}) - \mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] \end{aligned} \tag{5.4}$$

Thus minimizing the KL-divergence is equivalent to maximizing the ELBO.

In variational inference, the choice of the variational family is important because the flexibility of the variational family determines the power of the approximation. However, it is usually more difficult to optimize equation (5.4) over a complex family than a simple family. Therefore, many applications are performed using the *mean-field* variational family, which means that the parameters $\mathbf{m}$ are treated as being mutually independent (Bishop, 2006; Blei et al., 2017). However, even under that simplifying assumption, traditional variational methods require tedious model-specific derivations and implementations, which restricts their applicability to those problems for which derivations have been performed (e.g., Nawaz & Curtis, 2018, 2019). We therefore introduce two more general variational methods: the automatic differential variational inference (ADVI) and the Stein variational gradient descent (SVGD), which can both be applied to general inverse problems.

5.3.2 Automatic differential variational inference (ADVI)

Kucukelbir et al. (2017) proposed a general variational method called automatic differential variational inference (ADVI) based on a Gaussian variational family. In ADVI, a model with constrained parameters is first transformed to a model with unconstrained real-valued variables. For example, the velocity model $\mathbf{m}$ that usually has hard bound constraints (such as velocity being greater than zero) can be transformed to an unconstrained model $\boldsymbol{\theta} = T(\mathbf{m})$, where T is an invertible and differentiable function (Figure 5.1a and b). The joint probability $p(\mathbf{m}, \mathbf{d}_{obs})$ then becomes:

$$p(\boldsymbol{\theta}, \mathbf{d}_{obs}) = p(\mathbf{m}, \mathbf{d}_{obs})|det\mathbf{J}_{T^{-1}}(\boldsymbol{\theta})| \tag{5.5}$$

where $\mathbf{J}_{T^{-1}}(\boldsymbol{\theta})$ is the Jacobian matrix of the inverse of T which accounts for the volume change of the transform. This transform makes the choice of variational approximations independent of the original model since transformed variables lie in the common unconstrained space of real numbers.

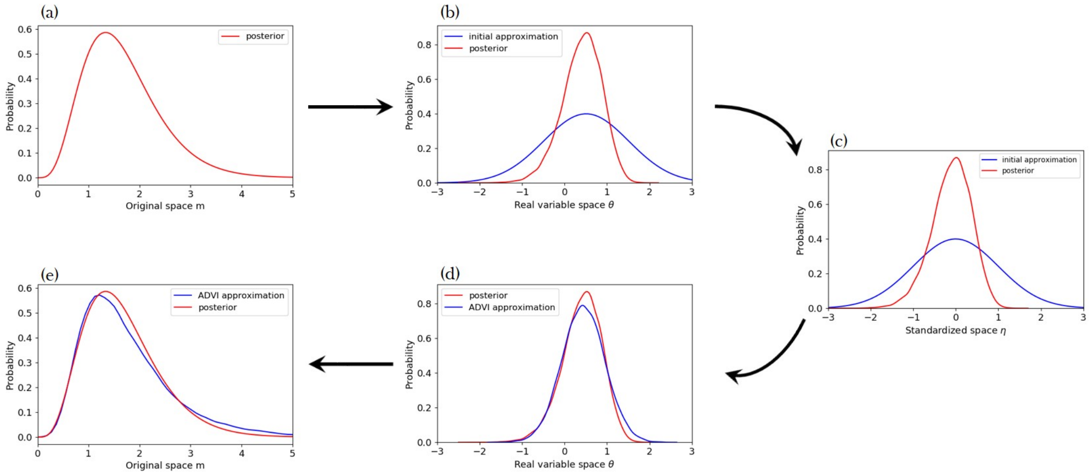


Figure 5.1: An illustration of the workflow of ADVI. **(a)** An example of a posterior pdf in the original positive half space of parameters $\mathbf{m}$. **(b)** The posterior pdf in the transformed real variable space $\boldsymbol{\theta}$ (red) and an initial Gaussian approximation (blue). **(c)** The posterior pdf (red) and the standard Gaussian distribution (blue) in standardized variable $\boldsymbol{\eta}$; gradients with respect to variational parameters are calculated in this space. **(d)** and **(e)** show the posterior pdf (red) and the approximation obtained using ADVI (blue) in the unconstrained real variable space and the original space, respectively.

In ADVI, we choose a Gaussian variational family (e.g., blue line in Figure 5.1b),

$$q(\boldsymbol{\theta};\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu},\mathbf{L}\mathbf{L}^T) \tag{5.6}$$

where $\boldsymbol{\phi}$ represents variational parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. As in Kucukelbir et al. (2017) for computational purpose we use a Cholesky factorization $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$ where $\mathbf{L}$ is a lower-triangular matrix, to re-parameterize the covariance matrix to ensure that it is positive semidefinite. If $\boldsymbol{\Sigma}$ is a diagonal matrix, q reduces to a mean-field approximation in which the variables are mutually independent; in order to include spatial correlations in the velocity model we use a full-rank covariance matrix, noting that this incurs a computational cost since it increases the number of variational parameters.

In the transformed space, the variational problem is solved by maximizing the ELBO, written as $\mathcal{L}$, with respect to variational parameters $\boldsymbol{\phi}$:

$$\begin{aligned} \boldsymbol{\phi}^* &= \arg\max_{\boldsymbol{\phi}} \mathcal{L}[q(\boldsymbol{\theta};\boldsymbol{\phi})] \\ &= \arg\max_{\boldsymbol{\phi}} \mathrm{E}_q\left[\log p(T^{-1}(\boldsymbol{\theta}),\mathbf{d}_{obs}) + \log|det\mathbf{J}_{T^{-1}}(\boldsymbol{\theta})|\right] - \mathrm{E}_q\left[\log q(\boldsymbol{\theta})\right] \end{aligned} \tag{5.7}$$

This is an optimization problem in an unconstrained space and can be solved using gradient ascent methods without worrying about any constrains on the original variables.

However, the gradients of variational parameters are not easy to calculate since the ELBO involves expectations in a high dimensional space. We therefore transform the Gaussian distribution $q(\boldsymbol{\theta};\boldsymbol{\phi})$ into a standard Gaussian $\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})$ (Figure

5.1c), by $\boldsymbol{\eta} = R_{\boldsymbol{\phi}}(\boldsymbol{\theta}) = \mathbf{L}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})$, thereafter the variational problem becomes:

$$\begin{aligned}
\boldsymbol{\phi}^* &= \arg\max_{\boldsymbol{\phi}} \mathcal{L}[q(\boldsymbol{\theta};\boldsymbol{\phi})] \\
&= \arg\max_{\boldsymbol{\phi}} \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})} \left[\log p\Big(T^{-1}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta}) \right), \mathbf{d}_{obs} \Big) + \log|det \mathbf{J}_{T^{-1}}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta}) \right)| \right] \\
&\qquad - \mathrm{E}_q\left[\log q(\boldsymbol{\theta}) \right]
\end{aligned} \tag{5.8}$$

where the first expectation is taken with respect to a standard Gaussian distribution $\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})$. There is no Jacobian term related to this transform since the determinant of the Jacobian is equal to one (Kucukelbir et al., 2017). The second expectation $-\mathrm{E}_q[\log q(\boldsymbol{\theta})]$ is not transformed since it has a simple analytic form as does its gradient (Kucukelbir et al., 2017) – see Appendix A.

Since the distribution with respect to which the expectation is taken now does not depend on variational parameters, the gradient with respect to variational parameters can be calculated by exchanging the expectation and derivative according to the dominated convergence theorem (Çınlar, 2011) and by applying the chain rule – see Appendix B:

$$\nabla_{\boldsymbol{\mu}} \mathcal{L} = \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})} \left[\nabla_{\mathbf{m}} \log p(\mathbf{m}, \mathbf{d}_{obs}) \nabla_{\boldsymbol{\theta}} T^{-1}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log|det \mathbf{J}_{T^{-1}}(\boldsymbol{\theta})| \right] \tag{5.9}$$

The gradient with respect to $\mathbf{L}$ can be obtained similarly,

$$\nabla_{\mathbf{L}} \mathcal{L} = \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})} \left[\left(\nabla_{\mathbf{m}} \log p(\mathbf{m}, \mathbf{d}_{obs}) \nabla_{\boldsymbol{\theta}} T^{-1}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log|det \mathbf{J}_{T^{-1}}(\boldsymbol{\theta})| \right) \boldsymbol{\eta}^T \right] + (\mathbf{L}^{-1})^T \tag{5.10}$$

where the expectation is computed with respect to a standard Gaussian distribution, which can be estimated by Monte Carlo (MC) integration. MC integration provides a noisy, unbiased estimation of the expectation and its accuracy increases with the number of samples. Nevertheless, it has been shown that in practice a low number or even a single sample can be sufficient at each iteration since the mean is taken with respect to the standard Gaussian distribution (see discussions and experiments in Kucukelbir et al., 2017). For distributions $p(\mathbf{m}, \mathbf{d}_{obs})$ for which the gradients have analytic forms, the whole process of computing gradients can be automated (Kucukelbir et al., 2017), hence the name "automatic differential". We can then use a gradient ascent method to update the variational parameters and obtain an approximation to the pdf $p(\mathbf{m}|\mathbf{d}_{obs})$ (e.g. Figure 5.1d).

Note that although the method is based on Gaussian variational approximations, the actual shape of the approximation to the posterior $p(\mathbf{m}|\mathbf{d}_{obs})$ over the original parameters $\mathbf{m}$ is determined by the transform T (Figure 5.1e). It is difficult to determine an optimal transform since that is related to the properties of the unknown posterior (Kucukelbir et al., 2017). In this study we use a commonly-used invertible logarithmic transform (Team et al., 2016),

$$\begin{aligned} \theta_i &= T(m_i) = \log(m_i - a_i) - \log(b_i - m_i) \\ m_i &= T^{-1}(\theta_i) = a_i + \frac{(b_i - a_i)}{1 + exp(-\theta_i)} \end{aligned} \tag{5.11}$$

where m_i represents each original constrained parameter, θ_i is the transformed unconstrained variable, a_i is the original lower bound and b_i the upper bound on m_i. Therefore the quality of the ADVI approximation is limited by the Gaussian approximation in the unconstrained space and by the specific transform T in equation (5.11).

To illustrate the effects of the transform in equation (5.11), we show an example in Figure 5.2. The original variable lies in a constrained space between 0.5 and 3.0 (a typical phase velocity range of seismic surface waves). The space is transformed to an unconstrained space using equation (5.11). If, as in ADVI we assume a standard Gaussian distribution in the transformed space (blue area in Figure 5.2), the associated probability distribution in the original space is shown in orange in Figure 5.2. The actual shape of the distribution in the original space is not

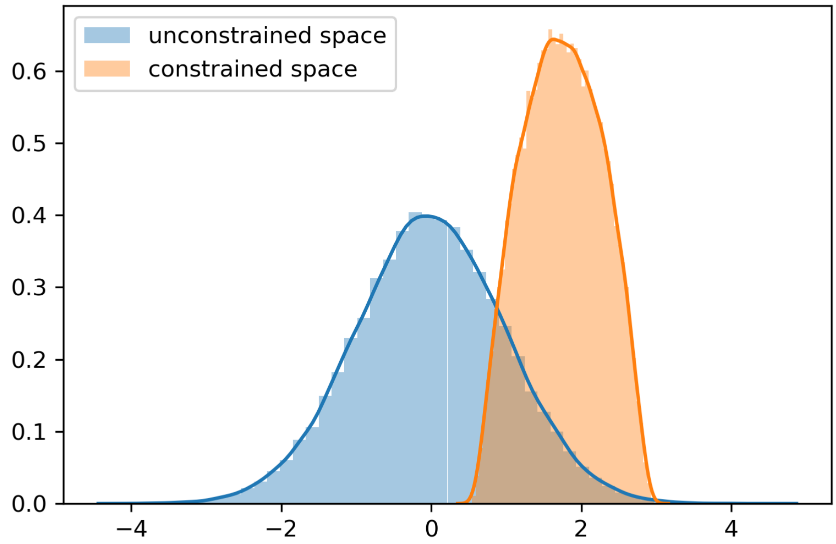


Figure 5.2: An illustration of the transform in equation (5.11). The original variable is in a constrained space between 0.5 and 3.0. The blue area shows a standard Gaussian distribution in the transformed unconstrained space and the orange area shows the associated probability distribution in the original space. The probability distributions are estimated using Monte Carlo samples.

Gaussian but is determined by the transform T in equation (5.11). However, under this choice of T it is likely that the probability distribution in the original space is still unimodal. We thus see that ADVI provides a unimodal approximation of the target posterior pdf around a local optimal parameter estimate. This suggests that the method will not be effective for multimodal distributions, and the estimated probability distribution depends on the initial value of μ and Σ (Kucukelbir et al., 2017). However, since the maximum a posteriori probability (MAP) estimate has been shown to be effective for parameter estimation in practice, the ADVI method could still be used to provide a good approximation of the distribution around a MAP estimate.

5.3.3 Stein variational gradient descent (SVGD)

In practice most applications of variational inference use simple families of posterior approximations such as a Gaussian approximation (Kucukelbir et al., 2017), mean-field approximations (Blei et al., 2017; Nawaz & Curtis, 2018, 2019) or other simple structured families (Saul & Jordan, 1996; Hoffman & Blei, 2015). These simple choices significantly restrict the quality of derived posterior approximations. In order to employ a broader family of variational approximations, variational methods based on invertible transforms have been proposed (Rezende & Mohamed, 2015; Tran et al., 2015; Marzouk et al., 2016). In these methods instead of choosing specific forms for variational approximations, a series of invertible transforms are applied to an initial distribution, and these transforms are optimized by minimizing the KL-divergence. This provides a way to approximate arbitrary posterior distributions since a pdf can be transformed to any other pdf as long as the probability measures are absolutely continuous.

Stein variational gradient descent (SVGD) is one such algorithm based on an incremental transform (Liu & Wang, 2016). In SVGD, a smooth transform $T(\mathbf{m}) = \mathbf{m} + \epsilon\boldsymbol{\phi}(\mathbf{m})$ is used, where $\mathbf{m} = [m_1, ..., m_d]$ and m_i is the i^{th} parameter, and $\boldsymbol{\phi}(\mathbf{m}) = [\phi_1, ..., \phi_d]$ is a smooth vector function that describes the perturbation direction and where ϵ is the magnitude of the perturbation. It can be shown that when ϵ is sufficiently small, the transform is invertible since the Jacobian of the transform is close to an identity matrix (Liu & Wang, 2016). Say $q_T(\mathbf{m})$ is the transformed probability distribution of the initial distribution $q(\mathbf{m})$. Then the gradient of KL-divergence with respect to ϵ can be computed as (see Appendix C):

$$\nabla_\epsilon \mathrm{KL}[q_T||p]\,|_{\epsilon=0} = -\mathrm{E}_q\left[trace\left(\mathcal{A}_p\boldsymbol{\phi}(\mathbf{m})\right)\right] \tag{5.12}$$

where $\mathcal{A}_p$ is the Stein operator such that $\mathcal{A}_p\boldsymbol{\phi}(\mathbf{m}) = \nabla_{\mathbf{m}}\mathrm{log}p(\mathbf{m})\boldsymbol{\phi}(\mathbf{m})^T + \nabla_{\mathbf{m}}\boldsymbol{\phi}(\mathbf{m})$. This suggests that maximizing the right-hand expectation with respect to $q(\mathbf{m})$ gives the steepest descent of the KL-divergence, and consequently the KL-divergence can be minimized iteratively.

It can be shown that the negative gradient of the KL-divergence in equation (5.12) can be maximized by using the kernelized Stein discrepancy (Liu et al., 2016). For two continuous probability densities p and q, the *Stein discrepancy* for a function $\boldsymbol{\phi}$ in a function set $\mathcal{F}$ is defined as:

$$S[q,p] = \underset{\boldsymbol{\phi}\in\mathcal{F}}{\arg\max}\{[\mathrm{E}_q trace\,(\mathcal{A}_p\boldsymbol{\phi}(\mathbf{m}))]^2\} \tag{5.13}$$

The Stein discrepancy provides another way to quantify the difference between two distribution densities (Stein et al., 1972; Gorham & Mackey, 2015). However the Stein discrepancy is not easy to compute for general $\mathcal{F}$. Therefore, Liu et al. (2016) proposed a kernelized Stein discrepancy by maximizing equation (5.13) in the unit ball of a reproducing kernel Hilbert space (RKHS) as follows.

A Hilbert space is a space $\mathcal{H}$ on which an inner product $<,>_{\mathcal{H}}$ is defined. A function is called a *kernel* if there exists a real Hilbert space and a function φ such that $k(x,y) =< \varphi(x), \varphi(y) >_{\mathcal{H}}$ (Gretton, 2013). A kernel is said to be positive-definite if the matrix defined by $K_{ij} = k(x_i, x_j)$ is positive definite. Assuming a positive definite kernel $k(\mathbf{m},\mathbf{m}')$ on $\mathcal{M}\times\mathcal{M}$, its reproducing kernel Hilbert space $\mathcal{H}$ is defined by the closure of the linear span $\{f : f(\mathbf{m}) = \sum_{i=1}^{n} a_i k(\mathbf{m},\mathbf{m}^i), a_i \in \mathcal{R}, n \in \mathcal{N}, \mathbf{m}^i \in \mathcal{M}\}$ with inner products $\langle f,g\rangle_{\mathcal{H}} = \sum_{ij} a_i b_j k(\mathbf{m}^i,\mathbf{m}^j)$ for $g(\mathbf{m}) = \sum_i b_i k(\mathbf{m},\mathbf{m}^i)$. The RKHS has an important reproducing property, that is, $f(x) =< f(x'), k(x',x) >_{\mathcal{H}}$, such that the evaluation of a function f at x can be represented as an inner product in the Hilbert space. In a RKHS, the kernelized Stein discrepancy can be defined as (Liu et al., 2016)

$$S[q,p] = \underset{\boldsymbol{\phi}\in\mathcal{H}^d}{\arg\max}\{\mathrm{E}_q\,[trace\,(\mathcal{A}_p\boldsymbol{\phi}(\mathbf{m}))]^2\,, \quad s.t. \quad ||\boldsymbol{\phi}||_{\mathcal{H}^d} \leq 1\} \tag{5.14}$$

where $\mathcal{H}^d$ is the RKHS of d-dimensional vector functions. The right side of equation (5.14) is found to be equal to,

$$\boldsymbol{\phi}^* = \boldsymbol{\phi}^*_{q,p}(\mathbf{m})/||\boldsymbol{\phi}^*_{q,p}(\mathbf{m})||_{\mathcal{H}^d} \tag{5.15}$$

where

$$\boldsymbol{\phi}^*_{q,p}(\mathbf{m}) = \mathrm{E}_{\{\mathbf{m}'\sim q\}}[\mathcal{A}_p k(\mathbf{m}', \mathbf{m})] \tag{5.16}$$

and for which we have $S[q,p] = ||\boldsymbol{\phi}^*_{q,p}(\mathbf{m})||^2_{\mathcal{H}^d}$. Thus the optimal $\boldsymbol{\phi}$ in equation (5.12) is $\boldsymbol{\phi}^*$ and $\nabla_\epsilon \mathrm{KL}[q_T||p]\,|_{\epsilon=0} = -\sqrt{S[q,p]}$.

Given the above solution, the SVGD works as follows: we start from an initial distribution q_0, then apply the transform $T^*_0(\mathbf{m}) = \mathbf{m}+\epsilon\boldsymbol{\phi}^*_{q_0,p}(\mathbf{m})$ where we absorb the normalization term in equation (5.15) into ϵ; this updates q_0 to $q_{[T_0]}$ with a decrease in the KL-divergence of $\epsilon * \sqrt{S[q,p]}$. This process is iterated to obtain an approximation of the posterior p:

$$q_{l+1} = q_{l[T^*_l]}, \;\; \textit{where} \;\; T^*_l(\mathbf{m}) = \mathbf{m} + \epsilon_l \boldsymbol{\phi}^*_{q_l,p}(\mathbf{m}) \tag{5.17}$$

and for sufficiently small $\{\epsilon_l\}$ the process eventually converges to the posterior pdf p. Note that a large stepsize may lead the Jacobian matrix of transform T to be singular, which in turn makes the approximation probability fail to converge to the true posterior (Liu, 2017).

To calculate the expectation in equation (5.16) we start from a set of particles (models) generated using q_0, and at each step the $\boldsymbol{\phi}^*_{q,p}(\mathbf{m})$ can be estimated by computing the mean in equation (5.16) using those particles. Each particle is then updated using the transform in equation (5.17), and those particles will form better approximations to the posterior as the iteration proceeds. This suggests the following algorithm which is schematically represented in Figure 5.3:

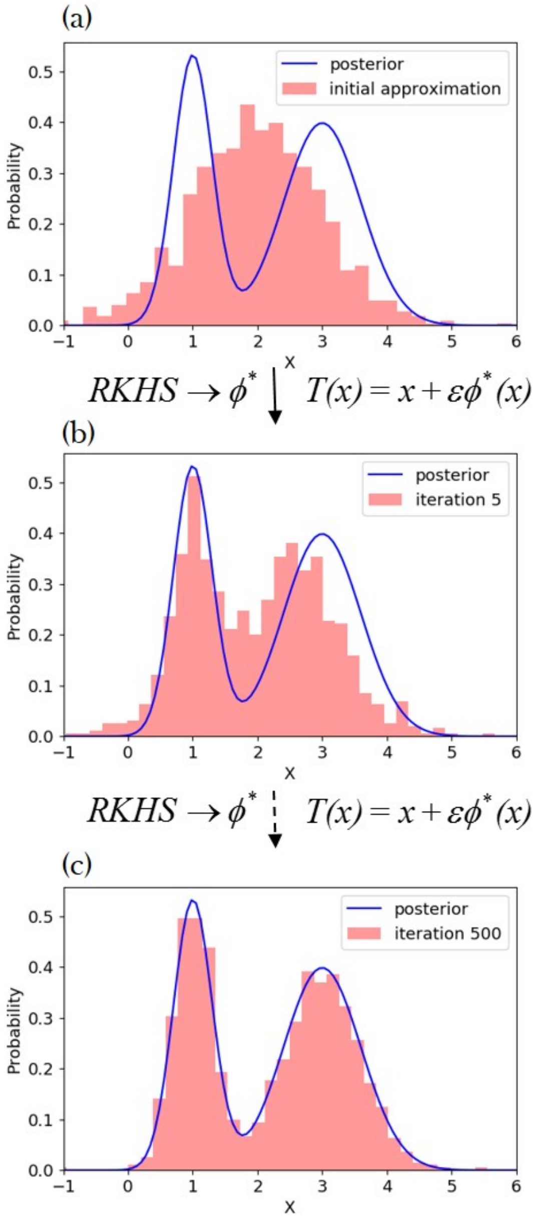


Figure 5.3: An illustration of the SVGD algorithm. The initial pdf is represented by the density of a set of particles (red histogram) in the top plot. The particles are then updated using a smooth transform $T(x) = x + \epsilon\phi^*(x)$, where ϕ^* is found in a reproducing kernel Hilbert space (RKHS). **(a)** An example of a posterior pdf (blue line) and an initial distribution (red histogram). **(b)** The approximating probability distribution after 5 iterations. **(c)** The approximating probability distribution after 500 iterations.

1. Draw a set of particles $\{\mathbf{m}_i^0\}_{i=1}^n$ from an initial pdf estimate (e.g., the prior).
2. At iteration l, update each particle using:

   $$\mathbf{m}_i^{l+1} = \mathbf{m}_i^l + \epsilon_l \boldsymbol{\phi}^*_{q_l,p}(\mathbf{m}_i^l) \tag{5.18}$$

   where

   $$\boldsymbol{\phi}^*_{q_l,p}(\mathbf{m}) = \frac{1}{n}\sum_{j=1}^{n}\Big[k(\mathbf{m}_j^l,\mathbf{m})\nabla_{\mathbf{m}_j^l}\mathrm{log}p(\mathbf{m}_j^l) + \nabla_{\mathbf{m}_j^l}k(\mathbf{m}_j^l,\mathbf{m})\Big] \tag{5.19}$$

   and ϵ_l is the step size at iteration l.
3. Calculate the density of the final set of particles $\{\mathbf{m}_i^*\}_{i=1}^n$ which approximates the posterior probability density function.

For kernel $k(\mathbf{m},\mathbf{m}')$ we use the radial basis function $k(\mathbf{m},\mathbf{m}') = \exp(-\frac{1}{h}||\mathbf{m} - \mathbf{m}'||^2$, where h is taken to be $\tilde{d}^2/\log n$ where $\tilde{d}$ is the median of pairwise distances between all particles. This choice of h is based on the intuition that $\sum_j k(\mathbf{m}_i,\mathbf{m}_j) \approx n\exp(-\frac{1}{h}\tilde{d}^2) = 1$, so that for particle $\mathbf{m}_i$ the contribution from its own gradient and the influence from the other particles in equation (5.19) are balanced (Liu & Wang, 2016). For the radial basis function kernel the second term in equation (5.19) becomes $\sum_j \frac{2}{h}(\mathbf{m} - \mathbf{m}_j)k(\mathbf{m}_j,\mathbf{m})$, which drives the particle $\mathbf{m}$ away from neighbouring particles for which the kernel takes large values. Therefore the second term in equation (5.19) acts as a *repulsive force* preventing particles from collapsing to a single mode, while the first term moves particles towards local high probability areas using the kernel-weighted gradient. If in the kernel $h \to 0$, the algorithm falls into independent gradient ascent that maximizes $\mathrm{log}p$ for each particle.

Note that since SVGD uses kernelized Stein discrepancy, the choice of kernels may affect the efficiency of the algorithm. In this study we used a commonly used kernel: a radial basis function. However, in some cases other kernels may provide a more efficient algorithm, for example, an inverse multiquadric kernel (Gorham & Mackey, 2017), a Hessian kernel (Detommaso et al., 2018) and kernels on Riemannian fold (Liu & Zhu, 2018).

In SVGD, the accuracy of the approximation increases with the number of particles. It has been shown that compared to other particle-based methods, e.g., sequential Monte Carlo methods (Smith, 2013), SVGD requires fewer samples to achieve the same accuracy which makes it a more efficient method (Liu & Wang, 2016). In contrast to sequential Monte Carlo which is a stochastic process, SVGD acts as a deterministic sampling method. If only one particle is used, the second term in equation (5.19) becomes zero and the method reduces to a typical gradient ascent towards the model with the maximum a posterior (MAP) pdf value. This suggests that even for a small number of particles the method could still produce a good parameter estimate since MAP estimation can be an effective method in practice.

In seismic tomography velocities are usually constrained to lie within a given velocity range. In order to ensure that velocities always lie within the constrains, we first apply the same transform used in ADVI (equation 5.11) so that the parameters are in an unconstrained space. We can then simply use equation (5.18) to update particles without explicitly considering the constrains on seismic velocities. The final seismic velocities can be obtained by transforming particles back to the constrained space.

5.4 Synthetic tests

We first apply the above methods to a simple 2D synthetic example similar to that in Galetti et al. (2015). The true model is a homogeneous background with velocity 2 km/s containing a circular low velocity anomaly with a radius of 2 km with velocity 1 km/s. The 16 receivers are evenly distributed around the anomaly approximating a circular acquisition geometry with radius 4 km (Figure 5.4). Each receiver is also treated as a source to simulate a typical ambient noise interferometry experiment (Campillo & Paul, 2003; Curtis et al., 2006; Galetti et al., 2015). This produces a total of 120 inter-receiver travel time data, each of which is computed using a fast marching method of solving the Eikonal equation over a 100×100 gridded discretisation in space (Rawlinson & Sambridge, 2004).

For variational inversions we use a fixed 21×21 grid of cells to parameterize the velocity model $\mathbf{m}$. The noise level is fixed to be 0.05 s ($<$ 5 percent of travel times) for all inversions. The prior pdf of the velocity in each cell is set to be a Uniform distribution between 0.5 km/s and 3.0 km/s to encompass the true model. Travel times are calculated using the same fast marching method as above over a 100×100 grid, but using the lower spatial resolution of model properties parameterized in $\mathbf{m}$. The gradients for velocity models are calculated by tracing rays backwards from receiver to (virtual) source using the gradient of the travel time field for each receiver pair (Rawlinson & Sambridge, 2004). For ADVI, the initial mean of the Gaussian distribution in the transformed space is chosen to be the value which is the transform of the mean value of the prior in the original space, and the initial covariance matrix is simply set to be an identity matrix. We then used 10,000 iterations to update the variational parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$). In order to visualize the results, we generated 5,000 models from the final approximate posterior probability density in the original space and computed their mean and standard deviation. For SVGD, we used 800 particles generated from the prior pdf

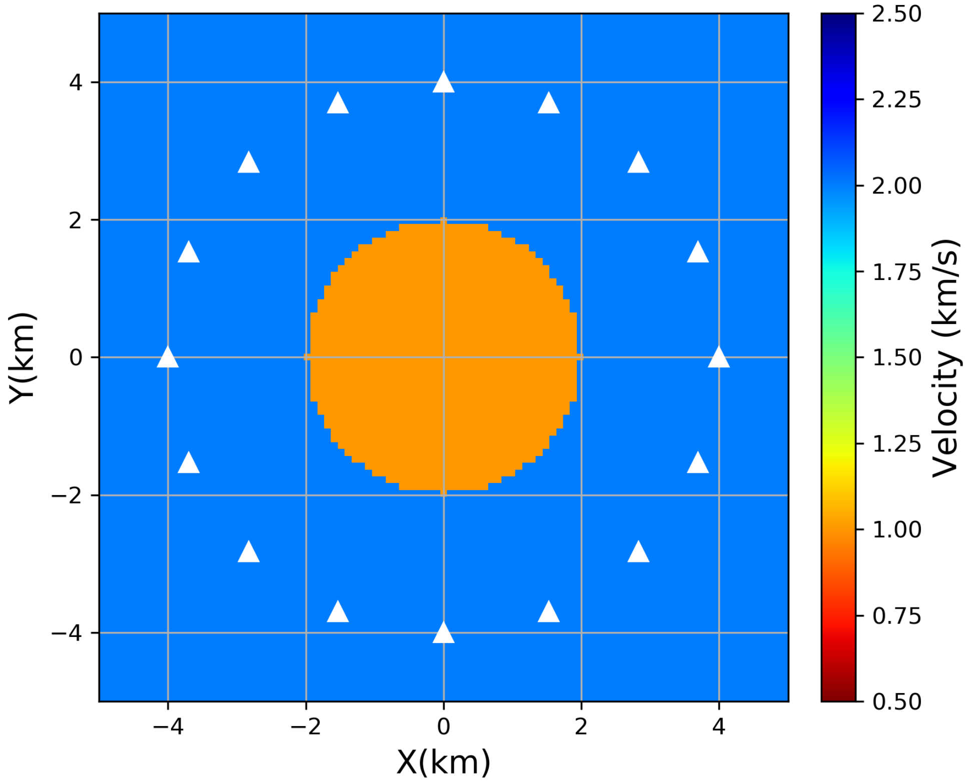


Figure 5.4: The true velocity model and receivers (white triangle) used in the synthetic test. Sources are at the same locations as receivers to simulate a typical ambient noise experiment.

and transformed to an unconstrained space using equation (5.11). Each particle is then updated using equation (5.17) for 500 iterations, then transformed back to seismic velocity. The mean and standard deviation are then calculated using the values of those particles.

To demonstrate the variational methods we compare the results with the fixed-dimensional Metropolis-Hastings McMC (MH-McMC) method (Metropolis & Ulam, 1949; Hastings, 1970; Mosegaard & Tarantola, 1995; Malinverno et al., 2000) and the rj-McMC method (Green, 1995; Bodin & Sambridge, 2009; Galetti

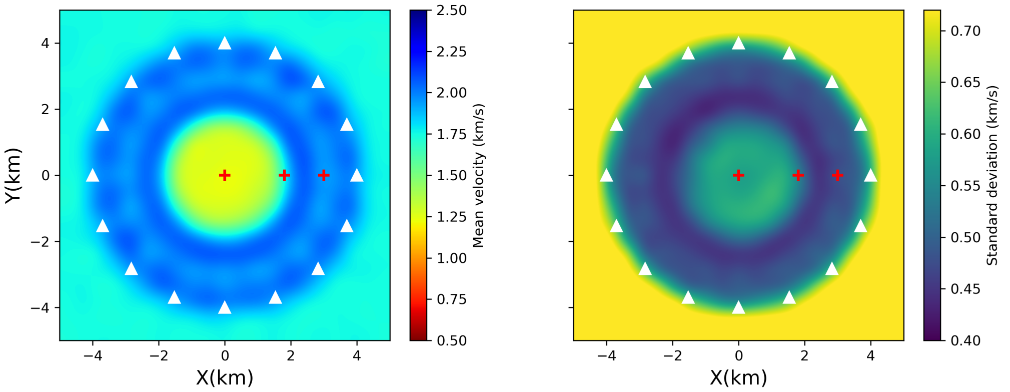


Figure 5.5: The mean (left) and standard deviation (right) found using ADVI. The red pluses show locations which are referred to in the main text.

et al., 2015; Zhang et al., 2018). For MH-McMC inversion we used the same parameterization as for the variational methods (a 21 × 21 grid). A Gaussian perturbation is used as the proposal distribution used to generate potential McMC samples, for which the step length is chosen by trial and error to give an acceptance ratio between 20 and 50 percent. We used a total of 6 chains, each of which used 2,000,000 iterations with a burn-in period of 1,000,000 iterations. To reduce the correlation between samples we only retain every 50^{th} sample in each chain after the burn-in period. The mean and standard deviation are then calculated using those samples. For rj-McMC inversion we use Voronoi cells to parameterize the model (Bodin & Sambridge, 2009), for which the prior pdf of the number of cells is set to be a Uniform distribution between 4 and 100. The proposal distribution for fixed-dimensional steps (changing the velocity of a cell or moving a cell) is chosen in a similar way as in MH-McMC. For trans-dimensional steps (adding or deleting a cell) the proposal distribution is chosen as the prior pdf (Zhang et al., 2018). We used a total of 6 chains, each of which contained 500,000 iterations with a burn-in period of 300,000. Similarly to the fixed-dimensional inversion the chain was thinned by a factor of 50 post burn-in.

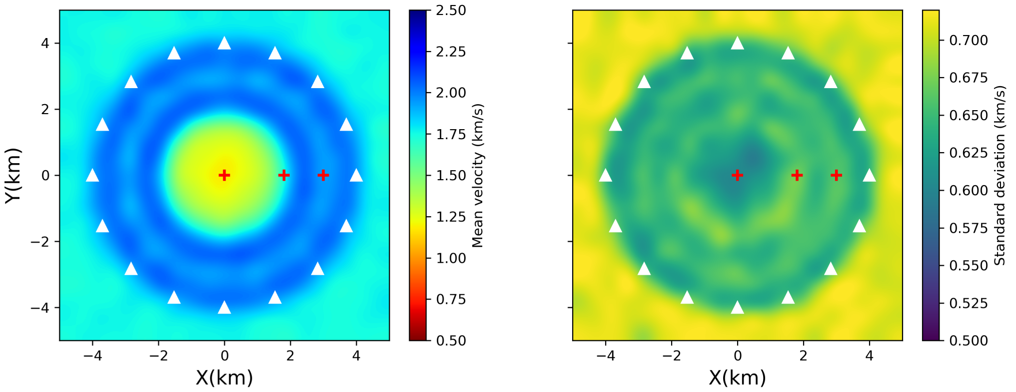


Figure 5.6: The mean (left) and standard deviation (right) found using SVGD. The red pluses show locations which are referred to in the main text.

5.4.1 Results

Figure 5.5 shows the mean and standard deviation calculated using ADVI. The mean model successfully recovers the low velocity anomaly within the receiver array except that the velocity value is slightly higher ($\sim 1.2\, km/s$) than the true value ($1.0\, km/s$). Between the location of the central anomaly and that of the receiver array there is a slightly lower velocity loop. The standard deviation map shows standard deviations similar to that of the prior (0.72 km/s) outside of the array, and clearly higher uncertainties at the location of the central anomaly. The standard deviations around the central anomaly are slightly higher than those at the center. Figure 5.6 shows the results from SVGD. Similarly, the velocity of the low velocity anomaly ($\sim 1.2\, km/s$) is slightly higher than the true value and a slightly lower velocity loop is also observed between the central anomaly and the receiver array. There is a clear higher uncertainty loop around the central anomaly; this has been observed previously and represent uncertainty due to the trade-off between the velocity of the anomaly and its shape (Galetti et al., 2015; Zhang et al., 2018). There is also another higher uncertainty loop associated with the lower velocity loop between the central anomaly and the receiver array. In contrast to this result, the loop cannot be observed in the results of ADVI.

To validate and better understand these results, Figure 5.7 shows the results from MH-McMC. The mean velocity model is very similar to the results from ADVI and SVGD. For example, the velocity value of the low velocity anomaly is higher than the true value, which suggests that the mean value of the posterior under the specified parameterization is genuinely biased towards higher values than the true value. A lower velocity loop is also observed between the circular anomaly and the receiver array. The standard deviation map shows similar results to those from SVGD: there is a higher uncertainty loop around the central anomaly and another one associated with the lower velocity loop between the circular anomaly and the receiver array. The latter loop suggests that this area is not well constrained by the data, and therefore the mean velocity tends towards the mean value of the prior which is lower than the true value. We do not observe the clear higher uncertainty loops in the result of ADVI which may be due to the Gaussian approximation which is used to fit a non-Gaussian posterior. In Figure 5.8 we show the results from rj-McMC. Compared to the results from the fixed-parameterization inversions, the mean velocity is a more accurate estimate of the true model and uncertainty across the model is also lower. For example, the middle low velocity anomaly has almost the same value as the true model and has standard deviation of only $\sim 0.3\,km/s$ compared to values significantly greater than 0.3 km/s for all other methods. Between the middle anomaly and the receivers, the model is determined better than in the fixed-paramterization inversions (with a standard deviation smaller than $0.1\,km/s$). This is because in rj-McMC the model parameterization adapts to the data which usually results in a lower-dimensional parameter space due to the natural parsimony of the method. For example, the average dimensionality of the parameter space in the rj-McMC inversion is around 10; for comparison the fixed-parameterization inversions all have dimensionality fixed to be 441. The standard deviation map from the rj-McMC also shows a clear higher uncertainty loop within the array around the low velocity anomaly, and high uncertainties outside of the array where there is no data coverage.

The results in Figure 5.8 do not show the double-loop uncertainty structure that is observed in the SVGD and MH-McMC results. The rj-McMC method contains an implicit natural parsimony – the method tends to use fewer rather than more cells whenever possible. While this may be useful in order to reduce the dimensionality of parameter space, it is also possible that it causes some detailed features of the velocity or uncertainty structure to be omitted, much like a smoothing regularization condition in other tomographic methods. Since the double-loop structure appears to be a robust feature of the image uncertainty, we assume that the parsimony has indeed regularised some of the image structure out of the rj-McMC results.

Note that the result from rj-McMC is fundamentally different from results obtained using the fixed-parameterization inversions (ADVI, SVGD and MH-McMC) because of its entirely different parameterization. While the other inversion results are parameterized with a regular grid and can themselves be regarded as pixelated images, rj-McMC produces a set of models that are vectors containing positions and velocities of Voronoi cells, which can be translated to an image on a regular grid (Figure 5.9). Note that the Voronoi parametrization imposes prior restricts on the models, for example, the points in the image within each Voronoi cell have the same vleocity. As a result rj-McMC produces very different results as those obtained using the other methods. In fact the choice of parameterizaiton in rj-McMC imposes restricts on models and different parameterizations can produce very different standard deviation

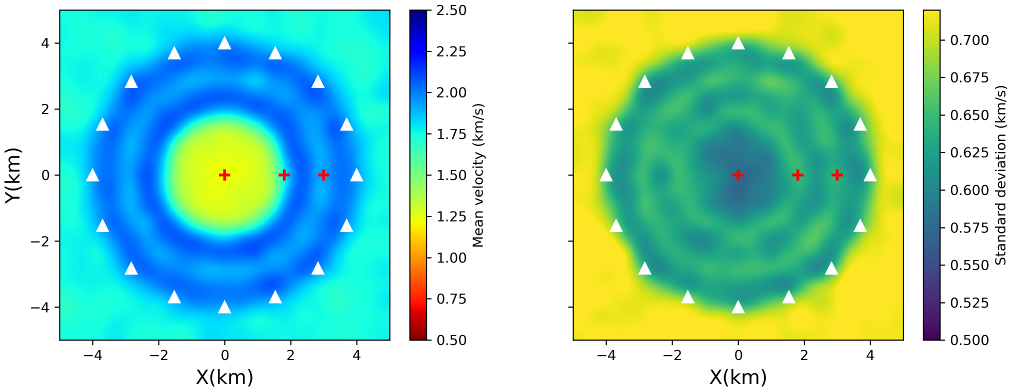


Figure 5.7: The mean (left) and standard deviation (right) found using MH-McMC. The red pluses show the point location which are referred to in the text.

structures (Hawkins et al., 2019). Thus the results of rj-McMC should be interpreted along with the specific parameterization.

To further analyse the results, in Figure 5.10 we show marginal probability distributions from the different inversion methods at three points (plus signs in Figure 5.5, 5.6, 5.7, and 5.8): point (0, 0) at the middle of the model, point (1.8, 0) at the boundary of the low velocity anomaly which has higher uncertainties, and point (3, 0) which also has higher uncertainties in the results from SVGD and MH-McMC. Due to symmetries of the model, marginal distributions at these three points are sufficient to reflect much of the entire set of single-parameter marginal probability distributions. At point (0, 0), the three fixed-parameterization methods produce similar marginal probability distributions. However, the marginal distribution from rj-McMC is narrower and concentrates around the true solution ($1.0\,km/s$). This is likely due to the fact that in rj-McMC we have a much smaller parameter space than in the fixed-parameterization inversions. To assess the convergence we show the marginal distributions obtained by doubling the number of iterations in ADVI and SVGD with an red line in

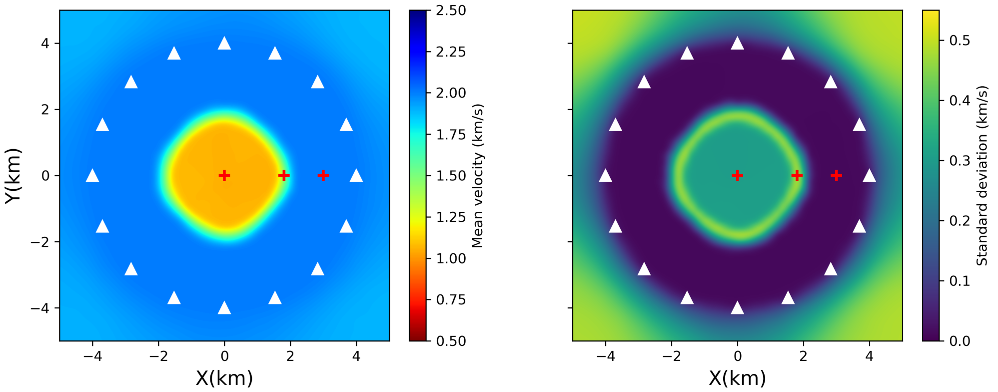


Figure 5.8: The mean (left) and standard deviation (right) found using trans-dimensional rj-McMC. The red pluses show the point location which are referred to in the text.

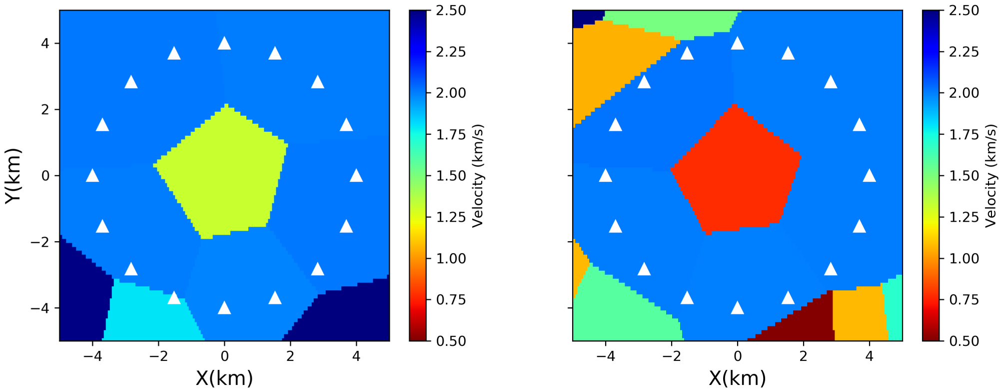


Figure 5.9: Two examples of the models sampled using rj-McMC algorithm.

Figure 5.10a and b. The results show that increasing iterations only slightly improves the marginal distributions, suggesting that they have nearly converged. The black line in Figure 5.10b shows the marginal distribution obtained using more particles (1,600) with the same number of iterations (500). The result is almost the same as the result obtained using the original set of particles which suggests that 800 particles are sufficient in this case. At point (1.8, 0), the marginal distributions from the three fixed-parameterization inversions become broader which explains the higher uncertainty loops observed in the standard deviation maps. The distribution from ADVI is more centrally focussed than the other two, which is again suggestive of the limitations of that method caused by the Gaussian approximation. The distributions from SVGD and MH-McMC are more similar to each other and are close to the prior – a Uniform distribution – which suggests that the area is not well constrained by the data. By contrast, the result from rj-McMC shows a clearly multimodal distribution with one mode centred around the velocity of the anomaly (1 km/s) and the other around the background velocity (2 km/s) as discussed in Galetti et al. (2015). This multimodal distribution reflects the fact that it is not clear whether this point is inside or outside of the anomaly which produces the higher uncertainty loop in the standard deviation map. This suggests that there are different causes of the higher uncertainty loops in the different models. In the fixed-parameterization inversions (ADVI, SVGD and MH-McMC) the higher uncertainty loops are mainly caused by the low resolution of the data at the boundary of the low velocity anomaly which produces broader marginal distributions. In the rj-McMC inversion, the higher uncertainty loops are mainly caused by multimodality in the posterior pdf. At point (3.0, 0) similarly to the point (0, 0), the marginal distributions from the three fixed-parameterization inversions have similar shape and are much broader than the result from rj-McMC. Compared to the results from SVGD and MH-McMC, the result from ADVI again shows a more centrally-focussed distribution reminiscent of the Gaussian limitation implicit in ADVI. In the result of rj-McMC the marginal distribution concentrates to a very narrow distribution around the true value. Overall the marginal distributions from the fixed-parameterization inversions are broader than the result from rj-McMC due to their far larger parameter space. Note that although the marginal distributions from SVGD and MH-McMC have slightly different shape which causes differences in the magnitudes of their standard deviation maps, the maps are essentially similar from these quite different methods which suggests that the results are (approximately) correct.

5.4.2 Computational cost

Table 1 summarises the computational cost of the different methods. ADVI involves 10,000 forward simulations which takes 0.45 CPU hours. However, note that in ADVI we used the full-rank covariance matrix which becomes huge in high dimensional parameter spaces which could makes the method inefficient. SVGD involves 400,000 forward simulations which takes 8.53 CPU hours. This appears to make it less efficient than ADVI, however SVGD can produce a more accurate approximation to the posterior pdf than ADVI which is limited by the Gaussian approximation. Note that SVGD can easily be parallelized by computing the gradients in equation (5.19) in parallel, making the method more time-efficient. For example, the above example takes 0.97 hours when parallelized using 10 cores. In comparison, MH-McMC requires 2,000,000 simulations for one chain which takes about 80.05 CPU hours, so for all 6 chains it requires 480.3 CPU hours in total. The rj-McMC run involved 500,000 simulations for one chain which takes about 17.1 CPU hours, so 102.6 CPU hours in total for 6 chains. The Monte Carlo methods use evaluations of the likelihood and prior distribution at each sample whereas both variational methods also deploy the information in the various gradients in equations 5.9, 5.10 and 5.19. The number of simulations is therefore

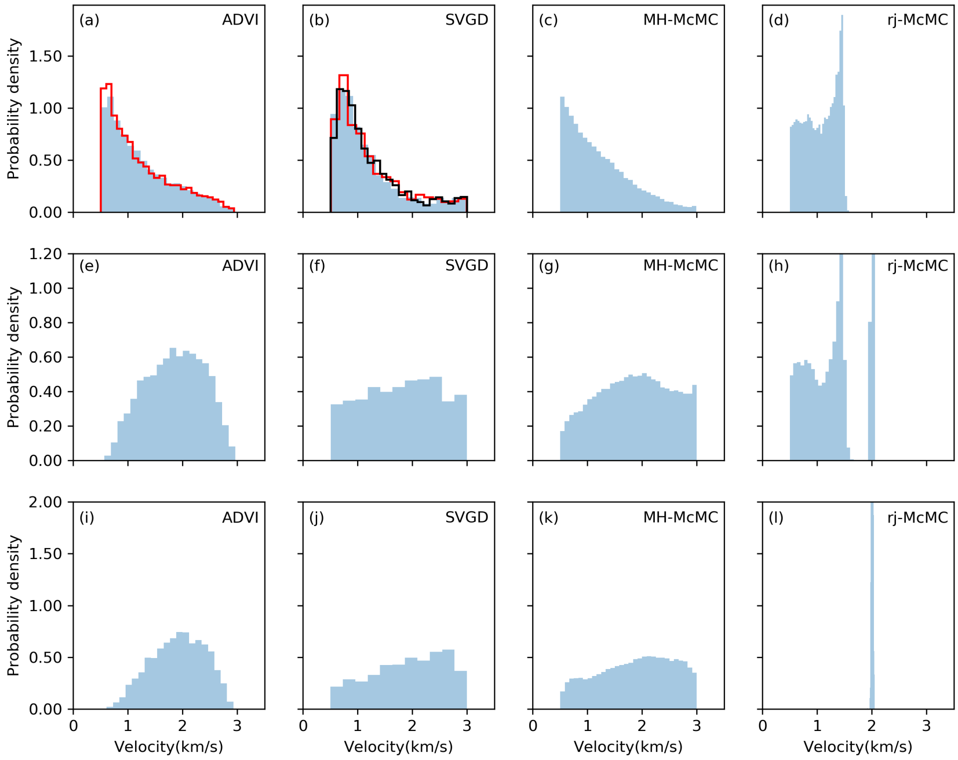


Figure 5.10: The marginal posterior pdfs of velocity at three points (pluses in Figure 3,4,5,6) derived using different methods. **(a)**, **(b)**, **(c)** and **(d)** show the marginal posterior distributions of velocity at the point (0,0) from ADVI, SVGD, MH-McMC and rj-McMC respectively. **(e)**, **(f)**, **(g)** and **(h)** show the marginal distributions at the point (1.8,0) from the four methods respectively, and **(i)**, **(j)**, **(k)** and **(l)** show the marginal distributions at the point (3,0) from the four methods respectively. The red lines in **(a)** and **(b)** are marginal distributions obtained by doubling the number of iterations and the black line in **(b)** shows the marginal distribution obtained using 1,600 particles.

not a good metric to compare the four methods, since the gradients in this case are calculated by ray tracing which require more calculations per simulation in Table 1 compared to MC. CPU hours is a fairer metric for comparison, but of course this depends on the mechanism by which gradients are obtained: in other forward or inverse problems it is even possible that the variational methods take longer than Monte Carlo if estimating gradients requires extensive computation.

In the comparison in Table 1, rj-McMC is more efficient than MH-McMC due to the fact that rj-McMC explores a much smaller parameter space than the fixed parameterization in MH-McMC. However, note that this might not always be true since trans-dimensional steps in rj-McMC usually have a very low probability of being accepted (Bodin & Sambridge, 2009; Zhang et al., 2018) and the method is generally significantly more difficult to tune (Green & Hastie, 2009). Overall, obtaining solutions from variational methods (ADVI, SVGD) is more efficient than Monte Carlo methods since they turn the Bayesian inference problem into an optimization problem. This also makes variational inference methods applicable to larger-datasets, and offers the advantage that very large datasets can be divided into random minibatches and inverted using stochastic optimization (Robbins & Monro, 1951; Kubrusly & Gravier, 1973) together with distributed computation. Monte Carlo methods are very computationally expensive for large datasets. Of course, the above comparison depends on the methods used to assess convergence for each method, which introduces some subjectivity in the comparison so that the absolute time required by each method may not be entirely accurate. Nevertheless, from all tests that we have conducted it is clear that variational methods produce solutions far more efficiently than Monte Carlo methods.

Table 5.1: The comparison of computational cost for all 4 methods

Method	Number of simulations	CPU hours
ADVI	10,000	0.45
SVGD	400,000	8.53
MH-McMC	12,000,000	480.3
rj-McMC	3,000,000	102.6

5.5 Application to Grane field

The Grane field is situated in the North Sea, and contains a permanent monitoring system composed of 3458 four-component sensors measuring 3 orthogonal components of particle velocity and water pressure variations due to passing seismic waves. This allows us to use ambient seismic noise tomography to study the subsurface of the field. To reduce the computational cost, in this study we down-sampled the number of receivers by a factor of 10 which results in 346 receivers, and we only used 35 receivers as virtual sources (Figure 5.11). Cross-correlations are computed between vertical component recordings at pairs consisting of a virtual source and a receiver using half-hour time segments, and the set of correlations for each pair were stacked over 6.5 hours. This process produces approximate virtual-source seismograms of Rayleigh-type Scholte waves (Campillo & Paul, 2003; Shapiro et al., 2005; Curtis et al., 2006). Phase velocity dispersion curves for each (virtual) source-receiver pair are then automatically picked using an image transformation technique: for all processing details see Chapter 3 which presents a complete ambient noise analysis of the field and presents tomographic phase velocity maps at various frequencies as well as estimated shear-velocity structure of the near seabed subsurface. Here we use the recording phase velocity data at 0.9s period.

We apply the variational inference methods ADVI and SVGD, and rj-McMC to the data to obtain phase velocity maps at 0.9 s and compare the results. For variational methods, the field is parametrized using a regular 26×71 grid with a spacing of 0.2 km at both x and y directions giving a velocity model dimensionality of 1846. Due to its computational cost in high dimensional spaces we do not apply MH-McMC. The data noise level is set to be 0.05 s, which is an average value estimated by the hierarchical Bayesian Monte Carlo inversion in Chapter 3. The prior pdf of phase velocity in each model cell is set to be a Uniform distribution between $0.35\, km/s$ and $0.55\, km/s$, which is selected to be wider than the minimum ($0.4\, km/s$) and maximum ($0.5\, km/s$) phase velocity picked from cross-correlations. We then applied 10,000 iterations for ADVI and for SVGD we used 1000 particles and 500 iterations. Similarly to the synthetic test above for rj-McMC we use Voronoi cells to parameterize the model. The prior pdf of the number of cells is set to be a discrete Uniform distribution between 30 and 200, and the data noise level is estimated hierarchically during the inversion (see Chapter 2). Proposal distributions are the same as in the synthetic test above. We used a total of 16 chains, each of which contains 800,000 iterations including a burn-in period of 400,000. To reduce the correlation between samples we only retain every 50^{th} sample post burn-in for our final ensemble.

Figure 5.12 shows the mean and standard deviation maps from ADVI. The mean phase velocity map shows a clear low velocity anomaly around the centre of the field from Y=6 km to Y=10 km and another at the western edge between Y=8 km and Y=10 km. These were also observed in Chapter 3 using Eikonal tomography, where we showed that they are correlated with areas of higher density of pockmarks on the seabed, suggesting that they are caused by near surface fluid flow effects. At the western edge between Y=6 km and Y=8 km and at the northwestern edge there are high velocity anomalies which were also observed in the results in Chapter 3. In the north between Y=11 km and Y=12 km and along the eastern edge between Y=7 km and Y=10 km the model shows some low

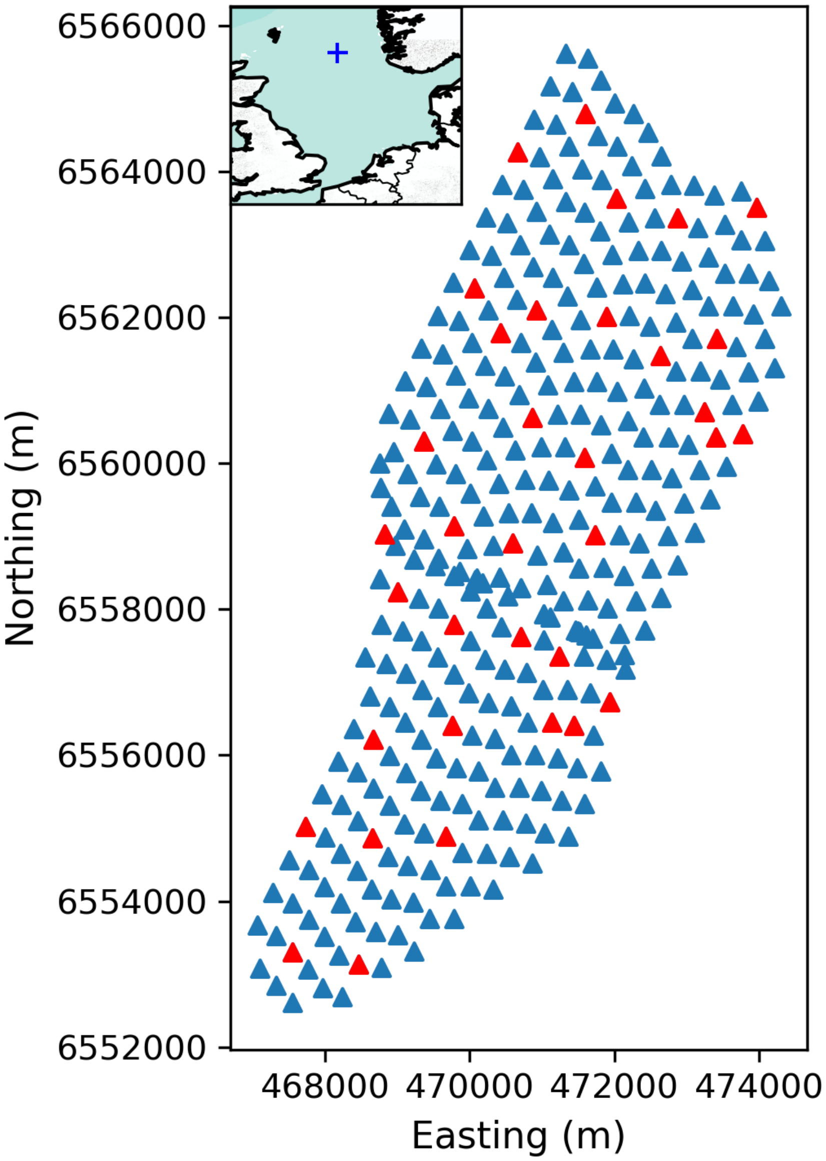


Figure 5.11: The distribution of receiver (blue and red triangles) across the Grane field used in this study. Red triangles show the receivers that were used as virtual sources. The blue plus in the inset map shows the location of Grane field.

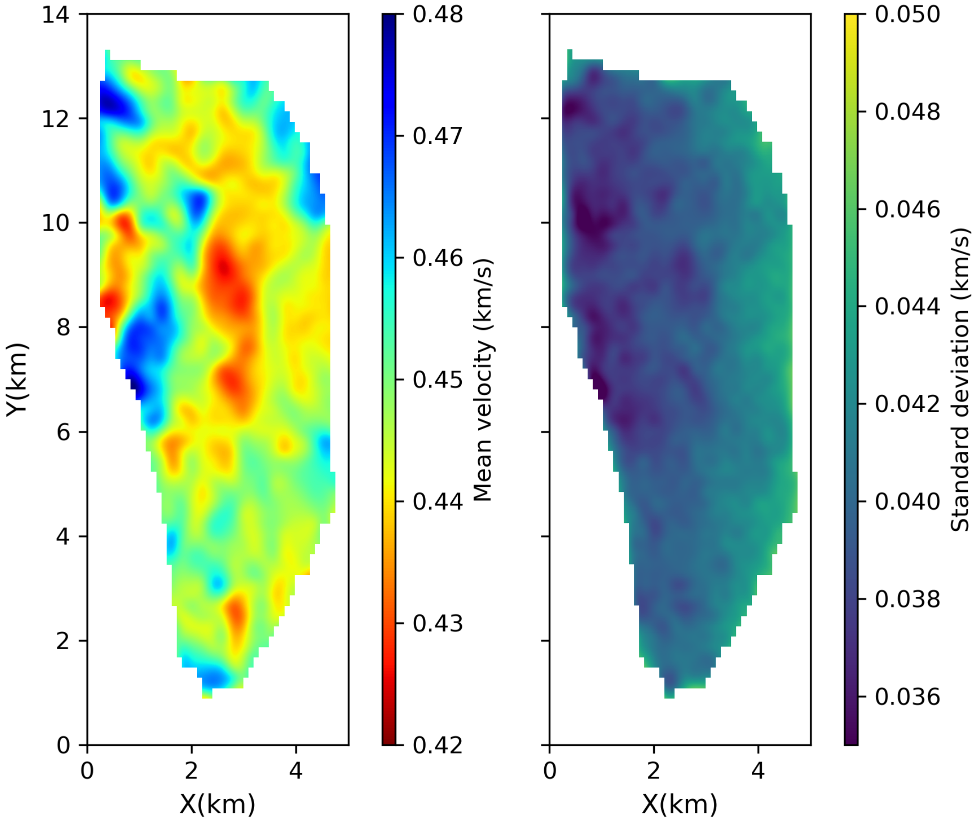


Figure 5.12: The mean (left) and standard deviation map (right) from ADVI.

velocity anomalies. Moreover, there are some small anomalies distributed across the field. For example, to the south of the central low velocity anomaly around Y=6 km there are several other low velocity anomalies. Similarly there is a small low velocity anomaly and a small high velocity anomaly in the south of the field around Y=2.5 km, and a small high velocity anomaly in the north around Y=10.5 km.

Overall the standard deviation map shows that uncertainty in the west is lower than in the east. At the western edge there are some low uncertainty areas which

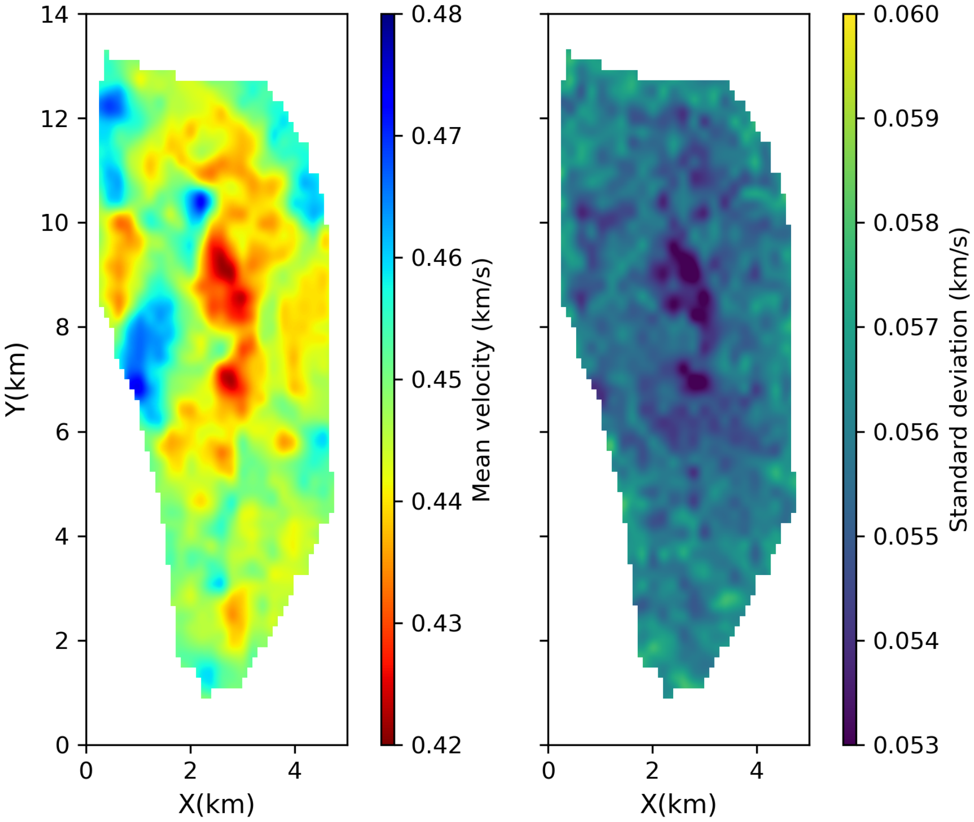


Figure 5.13: The mean (left) and standard deviation map (right) from SVGD.

are associated with velocity anomalies. For example, the low uncertainty area between Y=6 km and Y=8 km is associated with the high velocity anomaly at the same location. Similarly the high velocity anomaly at the northwestern edge around Y=12 km shows a lower uncertainty, and the middle low velocity anomaly also shows slightly lower uncertainties. This might suggest that these velocity structures are well-constrained by the data. However, in the synthetic tests we noticed that the ADVI can produce biased standard deviation maps due to the Gaussian approximation, so these uncertainty properties may not be robust.

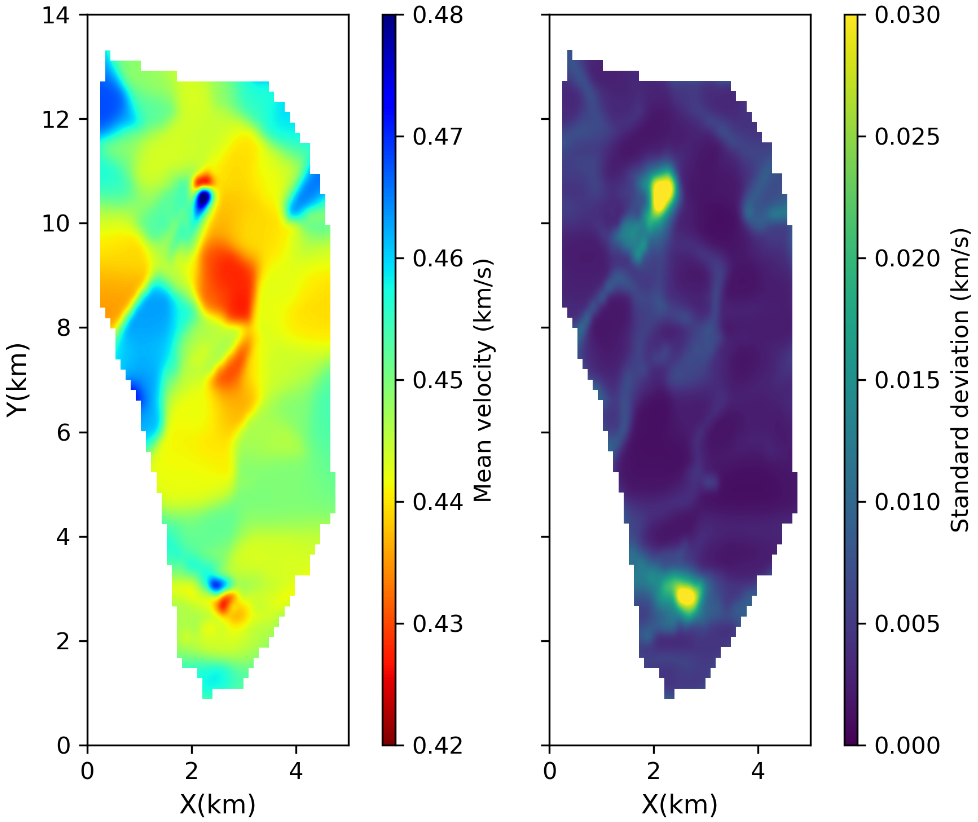


Figure 5.14: The mean (left) and standard deviation map (right) from rj-McMC.

We show the mean and standard deviation maps obtained using SVGD in Figure 5.13. The mean velocity map shows very similar structures to the result from ADVI, except that the velocity magnitudes are slightly different. For example, we observe the central low velocity anomaly and one at the western edge which appeared in the mean velocity map from ADVI and are related to the density distribution of pockmarks. Similarly there are high velocity anomalies at the western edge and a low velocity anomaly at the eastern edge. Even for more detailed structure, e.g., the low velocity anomalies at the north ($Y > 10$ km), the low velocity anomalies around $Y=6$ km and the small velocity anomalies around $Y=2.5$ km, the two results show highly consistent properties between the two methods. This suggests that we have obtained accurate mean phase velocity maps given the fixed, gridded model parameterization and the observed data.

Despite the similarity in the mean results, the standard deviation map from SVGD is quite different from the results from ADVI, which is consistent with similar variations that we observed in the synthetic tests. For example, there is no clear magnitude difference between the west and the east as appeared in the result from ADVI. There is a clear low uncertainty area associated with the central low velocity anomaly, which is slightly lower in magnitude than the result from ADVI. Similarly there is a slightly lower uncertainty area at the western edge associated with the low velocity anomaly at the same location. The south-central low velocity anomaly around $Y=6$ km also exhibits relatively lower uncertainties, which suggests that those small low velocity anomalies in this area may reflect true properties of the subsurface. Similarly there are some low uncertainty structures at the north around $Y= 11$ km which are associated with low velocity anomalies. Note that due to the Gaussian approximation in ADVI, the standard deviation results from SVGD show different magnitudes as we saw in the synthetic tests.

Figure 5.14 shows the mean and standard deviation maps obtained from rj-McMC.

The mean velocity map shows broadly similar structures to the results from ADVI and SVGD. For example, we also observed the middle low velocity anomaly, the low velocity anomalies at the western and eastern edges and the high velocity anomalies at the western edge. However, compared to the previous results these structures are smoother which is probably caused by the natural parsimony that is implicit within the rj-McMC inversion method (Green, 1995; Bodin & Sambridge, 2009) similarly to the synthetic tests above. The small velocity anomalies in the previous results disappear in the result from rj-McMC; this may also be caused by the natural parsimony of rj-McMC, or by overfitting of data in the variational methods due to the fixed parameterization. However, the small high and low velocity anomalies around Y=2.5 km and around Y=10.5 km still exist, which suggests that these detailed velocity structures may represent real properties of the subsurface (or are caused by a consistent bias in the data).

Similarly to the synthetic tests, the standard deviation map from rj-McMC shows significantly smaller uncertainties (< 0.01 km/s) than the results from ADVI (~ 0.04 km/s) and SVGD (~ 0.055 km/s), which is probably caused by a lower dimensionality of parameter space used in rj-McMC (around 60 Voronoi cells were used) than in variational methods (1846), resulting in fewer trade-offs between parameters. However, there are higher uncertainties at the location of the small velocity anomalies at Y=2.5 km and at Y=10.5 km, which is probably due to the fact that not all chains found these small structures.

In the inversion, ADVI involved 10,000 forward simulations which took 5.1 CPU hours and SVGD involved 500,000 forward simulations which required 141.8 CPU hours. By contrast the rj-McMC involved 12,800,000 forward simulations to obtain an acceptable result which required 1,866.1 CPU hours. In real time, SVGD was in fact parallelised using 12 cores which took 12.1 hours to run, while rj-McMC was parallelised using 16 cores which therefore took about 5 days.

We conclude that, although the variational methods produce higher uncertainty estimates, they can produce similar parameter estimates (mean velocity) at hugely reduced computational cost, and indeed our synthetic tests suggest that the variational SVGD image uncertainty results may in fact be more correct.

5.6 Discussion

We have shown that variational methods (ADVI and SVGD) can be applied to seismic tomography problems and provide efficient alternatives to McMC. ADVI produces biased posterior pdfs because of its implicit Gaussian approximation. However, it still generates an accurate estimate of the mean model. Given that it is very efficient (only requiring 10,000 forward simulations) the method could be useful in scenarios where efficiency is important and a Gaussian approximation is sufficient for uncertainty analysis. Alternatively mixture Gaussian approximations might be used to improve the accuracy of the algorithm (Zobay et al., 2014; Arenz et al., 2018). In a very high dimensional case, ADVI could become less efficient because of the increased size of the Gaussian covariance matrix. In that case one could use a mean-field approximation (setting model covariances to zero), or use a sparse covariance matrix to reduce computational cost since seismic velocity in any cell is often most strongly correlated with that in neighbouring cells.

SVGD can produce a good approximation to posterior pdfs. However, since it is based on a number of particles, the method is more computationally costly than ADVI. In this study we parallelized the computation of gradients to improve the efficiency, and for large datasets further improvements can be obtained by using random minibatches to perform the inversion (Liu & Wang, 2016). Such a strategy can be applied to any variational inference method (e.g. also ADVI) since variational methods solve an optimization rather than a stochastic sampling problem. In comparison, this strategy cannot easily be used in McMC based methods since it may break the detailed balance requirement of McMC (Blei et al., 2017). Though it has been shown that SVGD requires fewer particles than particle-based sampling methods (e.g., sequential Monte Carlo) in the sense that it reduces to finding the MAP model if only one particle is used, the optimal choice of the number of particles remains unclear, especially for very high dimensional spaces. In the case of very high dimensionality another possibility is to use normalizing flows – a variational method based on a series of specific invertible transforms (Rezende & Mohamed, 2015).

Monte Carlo and variational inference are different types of methods that solve the same problem. Monte Carlo simulates a set of Markov chains and uses samples of those chains to approximate the posterior pdf, while variational inference solves an optimization problem to find the closest pdf to the posterior within a given family of probability distributions. Monte Carlo methods provide guarantees that samples are asymptotically distributed according to the posterior pdf as the number of samples tends to infinity (Robert & Casella, 2013), while the statistical properties of variational inference algorithms are still unknown (Blei et al., 2017). It is possible to combine the two methods to capitalise on the merits of both. For example, the approximate posterior pdf from an efficient variational method (e.g. ADVI) can be used as a proposal distribution for Metropolis-Hastings (De Freitas et al., 2001) to improve the efficiency of McMC, or McMC steps can be integrated to the variational approximation to improve the accuracy of variational methods (Salimans et al., 2015).

We used a fixed regular grid of cells to parameterize the tomographic model in the variational methods, which might introduce overfitting of the data. For example, the mean velocity models in the synthetic tests show a slightly lower velocity loop between the low velocity anomaly and the receivers, and the uncertainties obtained from fixed-parameterization inversions are significantly higher than the results from rj-McMC. However, it is not easy to determine an optimal grid since this introduces a trade off between resolution of the model and overfitting of the data. Therefore, it might be necessary to use a more flexible parameterization, e.g., Voronoi cells (Bodin & Sambridge, 2009; Zhang et al., 2018) or wavelet parameterization (Fang & Zhang, 2014; Hawkins & Sambridge, 2015; Zhang & Zhang, 2015). It may also be possible to apply a series of different parameterizations and select the best one using model selection theory (Walter & Pronzato, 1997; Curtis & Snieder, 1997; Arnold & Curtis, 2018).

In our experiments the results from rj-McMC are significantly different from the results obtained using variational methods or MH-McMC. This is essentially caused by different parameterizations. In ADVI, SVGD and MH-McMC we invert for a pixelated image, while in rj-McMC we invert for a distribution of parameters that represent locations and shapes of cells and their constant velocities, the pointwise spatial mean of which is visualized as an image. Therefore even though we visualized them in the same way, the results are essentially not directly comparable. Nevertheless, the comparison with rj-McMC is interesting because until now a quite different alternative probabilistic method was never used to estimate the posterior of images from the same realistic tomography problem. The results here demonstrate that the rj-McMC method as applied in most tomography papers gives significantly different solutions than we might previously have thought; specifically, it does not produce the posterior distribution of the pixelated image that is usually shown in scientific papers (e.g., Bodin & Sambridge, 2009; Galetti et al., 2015; Zulfakriza et al., 2014; Crowder et al., 2019). Rather, it samples a probability distribution in a particular irregular and variably parametrized model space and results should be interpreted as such.

In this study we used a fixed data noise level in the variational methods. It has been shown that an improper noise level can introduce biases in tomographic results (Bodin & Sambridge, 2009; Zhang et al., 2019), so in our example we used the noise level estimated by hierarchical McMC. It can also be estimated by a variety of other methods (Bensen et al., 2009; Yao & Van Der Hilst, 2009; Weaver et al., 2011; Nicolson et al., 2012, 2014), and in future it might also be possible to include the noise parameters in variational methods in a hierarchical way.

In this study we applied variational inference methods to simple 2D tomography problems, but it is straightforward to apply the methods to any geophysical inverse problems whose gradients with respect to the model can be computed efficiently. For example, variational methods can be applied to 3D seismic tomography problems to provide efficient approximation, which generally demands enormous computational resources using McMC methods (Hawkins & Sambridge, 2015; Zhang et al., 2018, 2019). The methods also provide possibilities to perform Bayesian inference for full waveform inversion, which is generally very expensive for McMC (Ray et al., 2017) and suffers from notorious multimodality in the likelihoods. SVGD provides a possible way to approximate these complex distributions given that theoretically it can approximate arbitrary distributions.

5.7 Conclusion

We introduced two variational inference methods to geophysical tomography – automatic differential variational inference (ADVI) and Stein variational gradient descent (SVGD), and applied them to 2D seismic tomography problems using both synthetic and real data. Compared to the Markov chain Monte Carlo (McMC) method, ADVI provides an efficient but biased approximation to Bayesian posterior probability density functions. In contrast, SVGD is slightly slower than ADVI but produces a more accurate approximation. The real data example shows that ADVI and SVGD produce very similar mean velocity models, even though their uncertainty estimates are different because of a Gaussian approximation made implicitly within ADVI. The mean velocity models are very similar to those produced by reversible jump McMC (rj-McMC), except that the mean model from rj-McMC is smoother because of the much lower dimensionality of its parameter space. Variational methods thus can provide efficient approximate alternatives to McMC methods, and can be applied to many geophysical inverse problems.

Chapter 6

Discussion and Future work

In this thesis I applied Monte Carlo sampling methods and variational inference methods to solve Geophysical Bayesian inference problems, specifically seismic tomographic problems. Compared to traditional optimization methods, Bayesian inference methods provide probabilistic solutions to a usually under-determined and ill-posed problem and can be used to quantify uncertainties of solutions.

In Chapter 2 I introduced a 3D Monte Carlo surface wave tomography method which inverts for subsurface shear velocity structure directly from frequency-dependent travel time measurements. Compared to previous methods, the method preserves lateral and vertical spatial correlations in subsurface properties and in uncertainties, which is achieved by using reversible jump Markov chain Monte Carlo (McMC) method and a 3D parametrization based on Voronoi tessellations. Synthetic tests show that the method estimates velocity and uncertainty structure significantly better than previous methods and also does this in a comparable cost to the standard two-step McMC method.

In Chapter 3 I applied the 3D method introduced in Chapter 2 to ambient noise data recorded using a dense passive seismic array installed on the North Sea seabed. The ambient noise data are first cross correlated to obtain Scholte waves between each receiver pair and two modes are observed in each of those constructed waves. I therefore used a dispersion compensation method to separate the two modes. Phase velocity dispersion curves between each receiver pair are then automatically picked for each separated mode and are used to estimate phase velocity maps. At shorter periods (< 1.6 s) the fundamental mode phase velocity maps show a low velocity anomaly at the central area which might be related to near surface fluids. I then applied 1D, 2D and 3D Monte Carlo methods to obtain shear velocity models of the subsurface using dispersion data of the fundamental mode and compared their results along a cross section. The results show that the 1D inversion can cause errors in the final shear velocity model because of independence of each 1D depth inversion, whereas the 2D and 3D inversion significantly improved the results by including lateral spatial correlations. The 3D result better matches a model obtained from reflection tomography using active source seismic data than do the results from 2D or 1D inversions. This is probably because the 3D method estimate velocity structure directly from frequency-dependent travel time measurements which naturally avoids possible errors introduced in the first Eikonal tomography step. The shear velocity results show a central low velocity anomaly at the near surface (< 250 m) similar to that which appeared in fundamental phase velocity maps, and a clear low shear velocity river channel which is not observed in the phase velocity maps. I therefore confirm that the 3D Monte Carlo method is the most accurate method to estimate subsurface shear velocity structure from surface wave dispersion data.

In Chapter 4 I extended the 3D Monte Carlo method to a joint inversion using earthquake body wave travel times and ambient noise surface wave dispersion data and applied it to a mining site located to the north of New Ollerton, U.K. at which induced seismicity occurs. The results show that by including surface wave dispersion data in the inversion, both source parameters and velocity models can be better constrained than independent inversions. The results also show a high velocity anomaly which is correlated with one of the event clusters. To better understand the correlation, I conducted three synthetic tests: simultaneous inversion for source parameters and velocity using body wave data, inversion only for velocity using body wave data with source parameters fixed to the true values, and joint inversion for source parameter and velocity using both types of data. The results show that the independent inversion using only body wave data can produce biases in the results when inverting for source parameters and velocity models simultaneously, whereas the problem can be largely resolved by including surface wave data in the inversion. Thus I conclude that it is better to include surface wave dispersion data in seismic travel time tomographic inversions.

In Chapter 5 I introduced two variational inference methods: automatic differential variational inference (ADVI) and Stein variational gradient descent (SVGD) to solve Geophysical Bayesian inference problems. The two methods are first applied to a 2D synthetic tomographic problem and compared with Metropolis-Hastings McMC (MH-McMC) method and reversible jump McMC (rj-McMC) method. The results show that ADVI produces an efficient but biased approximation to the posterior probability density function compared with those obtained using MH-McMC method, whereas SVGD is slightly slower but produces a more accurate approximation. In contrast, the results obtain using rj-McMC are significantly different from those obtain using ADVI, SVGD and MH-McMC. This is essentially caused by different parametrizations: in ADVI, SVGD and MH-McMC we invert for a pixelated image, while in rj-McMC we invert for a distribution in a particular irregular and parametrized model space. Thus every model in the rj-McMC distribution consists of large, constant velocity patches which is not the case for the other three methods. I then applied the two variational methods to a real dataset from Grane field to estimate phase velocity distributions at 0.9

s and compared the results with rj-McMC method. The results from two variational inference methods show very similar mean velocity models but different uncertainties because of a Gaussian approximation made implicitly in ADVI. In contrast the results from rj-McMC show a smoother mean velocity model and significantly smaller uncertainties because of a much lower dimensionality of parameter space and a prior information (assumption) that models consist of large, constant-velocity patches. Overall the variational inference methods can provide efficient approximate alternatives to Monte Carlo sampling methods and can be applied to many types of geophysical problems.

In this thesis I mainly used rj-McMC algorithm with Voronoi tessellations to perform seismic tomography. While Voronoi tessellations are easy to implement in 1D, 2D and 3D for comparison and are good for estimating the location of discontinuities, it is difficult to obtain smooth velocity variations (as a smooth model can only be obtained by averaging many ensemble members, and indeed that smoothed, averaged result is a statistic of the solution, and not in itself a model). This often results in multi-modal posterior distributions. To reduce these issues other parametrizations, such as wavelet parametrization (Hawkins & Sambridge, 2015), Johnson-Mehl tessellation (Belhadj et al., 2018) and Delaunay and Clough-Tocher parameterizations (Hawkins et al., 2019) can be used. Note that these parametrizations can also be used in variational inference methods to provide more flexibility in models.

The efficiency of Metropolis-Hastings algorithm strongly depends on the choice of proposal distributions. However, it is difficult to find an optimal proposal distribution as it depends on the unknown posterior distribution (Neal, 1993; Green & Hastie, 2009; Green et al., 2015). In practice proposals are generally designed to be a small perturbation of the current model, which can be inefficient due to its random-walk behaviour. To resolve this issue some other better choices can be used. For example, a Gaussian proposal distribution can be updated along the chain using the information of previous samples, which is often referred as adaptive McMC algorithm (Haario et al., 2001; Atchade et al., 2009; Dosso et al., 2014). The nature of a problem can also be explored to design an efficient proposal distribution (Mosegaard, 2019). Instead of Metropolis-Hastings algorithms other Monte Carlo sampling methods may also be used to improve efficiency, such as Hamiltonian Monte Carlo (Duane et al., 1987; Neal et al., 2011; Sen & Biswas, 2017; Fichtner et al., 2018) and slice sampling (Neal et al., 2003).

The rj-McMC algorithm performs a trans-dimensional inversion so that a class of model parametrization is dynamically adapted to the data, whereas in fixed-dimensional McMC and variational inference a specific parametrization is applied which might produce biases in the results (e.g., overfitting). However, the proposal distribution for trans-dimensional step in rj-McMC is difficult to tune and usually produces a very low acceptance ratio (Green & Hastie, 2009) which makes the algorithm inefficient. Alternatively one can select a best parametrization using Bayesian model selection theory (Walter & Pronzato, 1997; Curtis & Snieder, 1997; Arnold & Curtis, 2018). For example, algorithms such as nested sampling (Skilling, 2004; Skilling et al., 2006; Feroz & Hobson, 2008) can be used to calculate the evidence term in Bayes theorem and perform model selections. The evidence lower bound (ELBO) calculated implicitly in variational inference methods may also be used as a model selection criterion (Sato, 2001; Bernardo et al., 2003; McGrory & Titterington, 2007).

In this study prior probabilities are chosen as a simple Uniform distribution. In cases that we have more knowledge about the subsurface, a more informative prior probability can be used to improve the results. In variational inference and MH-McMC I used a regular grid parameterization and a Uniform distribution, which produces complex individual models because of high dimensionality. This might not be appropriate since the real Earth may have a smoother structure (de Pasquale & Linde, 2016; Ray & Myer, 2019). In this case, some prior regularization may be used to produce individual smoother models (MacKay & Mac Kay, 2003). Gaussian process may also be used to inject prior information with adaptable complexity into inference scheme (e.g. Ray & Myer, 2019).

In conclusion, in this work I have shown that 3D trans-dimensional McMC outperforms more standard two-step 2D and 1D trans-dimensional McMC methods, I have performed the first 3D Monte Carlo tomography that includes both body and surface waves, I have introduced two new variational tomographic methods to Geophysics, and I have applied all of these methods to real data. Both the introduction of 3D Monte Carlo and variational methods should significantly extend theoretical and practical developments in the field of tomography, and particularly variational methods appear to present real opportunities to enhance tomography in future due to their apparent excellent performance at reduced computational cost. Time will tell.

Acknowledgments

The authors would like to thank the Grane license partners Equinor ASA, Petoro AS, ExxonMobil E&P Norway AS, and ConocoPhillips Skandinavia AS for allowing us to publish this work. The views and opinions expressed in this thesis are those of the authors and are not necessarily shared by the license partners. The authors thank the Edinburgh Interferometry Project sponsors (Schlumberger, Equinor and Total) for supporting this research. This work used the Cirrus UK National Tier-2 HPC Service at EPCC (http://www.cirrus.ac.uk).

Appendix A

Gradients calculation in variational inference methods

A.1 The entropy of a Gaussian distribution

The entropy $\mathrm{H}\left[q(\boldsymbol{\theta};\boldsymbol{\phi})\right]$ of a Gaussian distribution $\mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu},\mathbf{L}\mathbf{L}^T)$ is:

$$\begin{aligned}
\mathrm{H}\left[q(\boldsymbol{\theta};\boldsymbol{\phi})\right] &= -\mathrm{E}_q[\log q(\boldsymbol{\theta})] \\
&= -\int \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu},\mathbf{L}\mathbf{L}^T)\log\mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu},\mathbf{L}\mathbf{L}^T)d\boldsymbol{\theta} \\
&= \frac{k}{2}+\frac{k}{2}\log(2\pi)+\frac{1}{2}\log|det(\mathbf{L}\mathbf{L}^T)|
\end{aligned}$$

where k is the dimension of vector $\boldsymbol{\theta}$. The gradients with respect to $\boldsymbol{\mu}$ and $\mathbf{L}$ can be easily calculated (see Appendix A.2).

A.2 Gradients of the ELBO in ADVI

We first describe the dominated convergence theorem (DCT) (Çınlar, 2011): **Theorem** Assume $X \in \mathcal{X}$ is a random variable and $f : \mathbb{R} \times \mathcal{X} \to \mathbb{R}$ is a function such that $f(t, X)$ is integrable for all t and $\frac{\partial f(t,X)}{\partial t}$ exists for each t. Assume that there is a random variable Z such that $|\frac{\partial f(t,X)}{\partial t}| \leq Z$ for all t and $\mathrm{E}(Z) < \infty$. Then

$$\frac{\partial}{\partial t}\mathrm{E}(f(t, X)) = \mathrm{E}(\frac{\partial}{\partial t}f(t, X))$$

The proof of this theorem is given in Çınlar (2011).

We then calculate the gradients in equation (5.9) and (5.10) based on Kucukelbir et al. (2017). The ELBO $\mathcal{L}$ is:

$$\mathcal{L} = \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\left[\log p\big(T^{-1}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right), \mathbf{d}_{obs}\big) + \log|det\mathbf{J}_{T^{-1}}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right)|\right] + \mathrm{H}\left[q(\boldsymbol{\theta};\boldsymbol{\phi})\right]$$

where $\mathrm{H}\left[q(\boldsymbol{\theta};\boldsymbol{\phi})\right] = \mathrm{E}_q\left[\log q(\boldsymbol{\theta})\right]$ is the entropy of distribution q. Assume $\frac{\partial}{\partial \phi}\log p$ is bounded where $\boldsymbol{\phi}$ represents variational parameters $\boldsymbol{\mu}$ and $\mathbf{L}$, then the gradients can be computed by exchanging the derivative and the expectation using the dominated convergence theorem (DCT) and applying the chain rule:

$$\begin{aligned}\nabla_{\boldsymbol{\mu}}\mathcal{L} &= \nabla_{\boldsymbol{\mu}}\big\{\mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\left[\log p\big(T^{-1}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right), \mathbf{d}_{obs}\big) + \log|det\mathbf{J}_{T^{-1}}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right)|\right] \\ &\quad +\mathrm{H}\left[q(\boldsymbol{\theta};\boldsymbol{\phi})\right]\big\}\end{aligned}$$

Applying the DCT and since H does not depend on $\boldsymbol{\mu}$,

$$\nabla_{\boldsymbol{\mu}}\mathcal{L} = \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\left[\nabla_{\boldsymbol{\mu}}\left\{\log p\big(T^{-1}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right), \mathbf{d}_{obs}\big)\right\} + \nabla_{\boldsymbol{\mu}}\left(\log|det\mathbf{J}_{T^{-1}}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right)|\right)\right]$$

Applying the chain rule,

$$\begin{aligned}\nabla_{\boldsymbol{\mu}}\mathcal{L} &= \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\big[\nabla_{\mathbf{m}}\mathrm{log}p(\mathbf{m},\mathbf{d}_{obs})\nabla_{\boldsymbol{\theta}}T^{-1}(\boldsymbol{\theta})\nabla_{\boldsymbol{\mu}}R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta}) \\ &\quad +\nabla_{\boldsymbol{\theta}}\mathrm{log}|det\mathbf{J}_{T^{-1}}(\boldsymbol{\theta})|\nabla_{\boldsymbol{\mu}}R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\big] \\ &= \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\left[\nabla_{\mathbf{m}}\mathrm{log}p(\mathbf{m},\mathbf{d}_{obs})\nabla_{\boldsymbol{\theta}}T^{-1}(\boldsymbol{\theta})+\nabla_{\boldsymbol{\theta}}\mathrm{log}|det\mathbf{J}_{T^{-1}}(\boldsymbol{\theta})|\right]\end{aligned}$$

The gradient with respect to $\mathbf{L}$ can be obtained similarly,

$$\begin{aligned}\nabla_{\mathbf{L}}\mathcal{L} &= \nabla_{\mathbf{L}}\Big\{\mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\left[\mathrm{log}p\left(T^{-1}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right),\mathbf{d}_{obs}\right)+\mathrm{log}|det\mathbf{J}_{T^{-1}}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right)|\right] \\ &\quad +\frac{k}{2}+\frac{k}{2}\mathrm{log}(2\pi)+\frac{1}{2}\mathrm{log}|det(\mathbf{L}\mathbf{L}^{T})|\Big\}\end{aligned}$$

Applying the DCT

$$\begin{aligned}\nabla_{\mathbf{L}}\mathcal{L} &= \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\Big[\nabla_{\mathbf{L}}\left\{\mathrm{log}p\left(T^{-1}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right),\mathbf{d}_{obs}\right)\right\} \\ &\quad +\nabla_{\mathbf{L}}\left(\mathrm{log}|det\mathbf{J}_{T^{-1}}\left(R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\right)|\right)\Big]+\nabla_{\mathbf{L}}\frac{1}{2}\mathrm{log}|det(\mathbf{L}\mathbf{L}^{T})|\end{aligned}$$

and applying the chain rule we obtain

$$\begin{aligned}\nabla_{\mathbf{L}}\mathcal{L} &= \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\big[\nabla_{\mathbf{m}}\mathrm{log}p(\mathbf{m},\mathbf{d}_{obs})\nabla_{\boldsymbol{\theta}}T^{-1}(\boldsymbol{\theta})\nabla_{\mathbf{L}}R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta}) \\ &\quad +\nabla_{\boldsymbol{\theta}}\mathrm{log}|det\mathbf{J}_{T^{-1}}(\boldsymbol{\theta})|\nabla_{\mathbf{L}}R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\eta})\big]+(\mathbf{L}^{-1})^{T} \\ &= \mathrm{E}_{\mathcal{N}(\boldsymbol{\eta}|\mathbf{0},\mathbf{I})}\left[\left(\nabla_{\mathbf{m}}\mathrm{log}p(\mathbf{m},\mathbf{d}_{obs})\nabla_{\boldsymbol{\theta}}T^{-1}(\boldsymbol{\theta})+\nabla_{\boldsymbol{\theta}}\mathrm{log}|det\mathbf{J}_{T^{-1}}(\boldsymbol{\theta})|\right)\boldsymbol{\eta}^{T}\right]+(\mathbf{L}^{-1})^{T}\end{aligned}$$

A.3 Gradients of KL-divergence in SVGD

We calculate the gradient in equation (5.12) following Liu & Wang (2016). Denote T^{-1} as the inverse transform of T. Then by changing the variable,

$$\mathrm{KL}[q_T||p]=\mathrm{KL}[q||p_{T^{-1}}]$$

and hence

$$\begin{aligned}\nabla_\epsilon \mathrm{KL}[q_T||p]\,|_{\epsilon=0} &= \nabla_\epsilon \mathrm{KL}[q||p_{T^{-1}}]\,|_{\epsilon=0} \\ &= \nabla_\epsilon \left[\mathrm{E}_q \mathrm{log} q(\mathbf{m}) - \mathrm{E}_q \mathrm{log} p_{T^{-1}}(\mathbf{m})\right]\end{aligned}$$

and since $q(\mathbf{m})$ does not depend on ϵ

$$\nabla_\epsilon \mathrm{KL}[q_T||p]\,|_{\epsilon=0} = -\mathrm{E}_q \left[\nabla_\epsilon \mathrm{log} p_{T^{-1}}(\mathbf{m})\right]$$

where $p_{T^{-1}}(\mathbf{m}) = p(T(\mathbf{m})) \cdot |\det\left(\nabla_{\mathbf{m}} T(\mathbf{m})\right)|$. Therefore

$$\nabla_\epsilon \mathrm{log} p_{T^{-1}}(\mathbf{m}) = \left(\nabla_{\mathbf{m}} \mathrm{log}\left(p(\mathbf{m})\right)\right)^{\mathrm{T}} \nabla_\epsilon T(\mathbf{m}) + trace\left(\left(\nabla_{\mathbf{m}} T(\mathbf{m})\right)^{-1} \cdot \nabla_\epsilon \nabla_{\mathbf{m}} T(\mathbf{m})\right)$$

where $T(\mathbf{m}) = \mathbf{m} + \epsilon \boldsymbol{\phi}(\mathbf{m})$, $\nabla_\epsilon T(\mathbf{m} = \boldsymbol{\phi}(\mathbf{m})$ and $\nabla_{\mathbf{m}} T(\mathbf{m})|_{\epsilon=0} = \mathbf{I}$, and so

$$\begin{aligned}\nabla_\epsilon \mathrm{KL}[q_T||p]\,|_{\epsilon=0} &= -\mathrm{E}_q \left[\left(\nabla_{\mathbf{m}} \mathrm{log}\left(p(\mathbf{m})\right)\right)^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{m}) + trace\left(\nabla_{\mathbf{m}} \boldsymbol{\phi}(\mathbf{m})\right)\right] \\ &= -\mathrm{E}_q \left[trace\left(\nabla_{\mathbf{m}} \mathrm{log}\left(p(\mathbf{m})\right) \boldsymbol{\phi}(\mathbf{m})^{\mathbf{T}}\right) + trace\left(\nabla_{\mathbf{m}} \boldsymbol{\phi}(\mathbf{m})\right)\right] \\ &= -\mathrm{E}_q \left[trace\left(\mathcal{A}_p \boldsymbol{\phi}(\mathbf{m})\right)\right]\end{aligned}$$

where $\mathcal{A}_p \boldsymbol{\phi}(\mathbf{m}) = \nabla_{\mathbf{m}} \mathrm{log} p(\mathbf{m}) \boldsymbol{\phi}(\mathbf{m})^T + \nabla_{\mathbf{m}} \boldsymbol{\phi}(\mathbf{m})$ is the Stein operator.

References

Aki, K. & Lee, W., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. a homogeneous initial model, *Journal of Geophysical research*, **81**(23), 4381–4399.

Aki, K. & Richards, P. G., 1980. *Quantitative seismology*.

Alleyne, D., Pialucha, T., & Cawley, P., 1993. A signal regeneration technique for long-range propagation of dispersive lamb waves, *Ultrasonics*, **31**(3), 201–204.

Allmark, C., Curtis, A., Galetti, E., & de Ridder, S., 2018. Seismic attenuation from ambient noise across the North Sea Ekofisk permanent array, *Journal of Geophysical Research: Solid Earth*.

Arenz, O., Zhong, M., & Neumann, G., 2018. Efficient gradient-free variational inference using policy search, in *International Conference on Machine Learning*, pp. 234–243.

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

Aster, R. C., Borchers, B., & Thurber, C. H., 2018. *Parameter estimation and inverse problems*, Elsevier.

Atchade, Y., Fort, G., Moulines, E., & Priouret, P., 2009. Adaptive markov chain monte carlo: theory and methods, *Preprint*.

Barkved, O. I. & Kristiansen, T., 2005. Seismic time-lapse effects and stress changes: Examples from a compacting reservoir, *The Leading Edge*, **24**(12), 1244–1248.

Barkved, O. I., Kristiansen, T., & Fjær, E., 2005. The 4D seismic response of a compacting reservoirâĂŤexamples from the Valhall field, Norway, in *SEG Technical Program Expanded Abstracts 2005*, pp. 2508–2511, Society of Exploration Geophysicists.

Behr, Y., Townend, J., Bannister, S., & Savage, M., 2010. Shear velocity structure of the Northland Peninsula, New Zealand, inferred from ambient noise correlations, *Journal of Geophysical Research: Solid Earth*, **115**(B5).

Belhadj, J., Romary, T., Gesret, A., Noble, M., & Figliuzzi, B., 2018. New parameterizations for bayesian seismic tomography, *Inverse Problems*, **34**(6), 065007.

Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A., Lin, F., Moschetti, M., Shapiro, N., & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International*, **169**(3), 1239–1260.

Bensen, G., Ritzwoller, M., & Yang, Y., 2009. A 3-D shear velocity model of the crust and uppermost mantle beneath the United States from ambient seismic noise, *Geophysical Journal International*, **177**(3), 1177–1196.

Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., West, M., et al., 2003. The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures, *Bayesian statistics*, **7**, 453–464.

Biondi, B., 1992. Solving the frequency-dependent Eikonal equation, in *SEG Technical Program Expanded Abstracts 1992*, pp. 1315–1319, Society of Exploration Geophysicists.

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.

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

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *Journal of Geophysical Research: Solid Earth*, **117**(B2).

Box, G. E. & Tiao, G. C., 2011. *Bayesian inference in statistical analysis*, vol. 40, John Wiley & Sons.

Brocher, T. M., 2005. Empirical relations between elastic wavespeeds and density in the earth's crust, *Bulletin of the seismological Society of America*, **95**(6), 2081–2092.

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L., 2011. *Handbook of markov chain monte carlo*, CRC press.

Brooks, S. P., Giudici, P., & Roberts, G. O., 2003. Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(1), 3–39.

Bullock, A., Bekhtin, Y., Ortin, M., Mathewson, J., Henneberg, K., & Sæbø, A., 2015. Improved imaging with PP-PS simultaneous joint tomography over the Grane field, in *SEG Technical Program Expanded Abstracts 2015*, pp. 2094–2097, Society of Exploration Geophysicists.

Burdick, S. & Lekić, V., 2017. Velocity variations and uncertainty from transdimensional p-wave tomography of North America, *Geophysical Journal International*, **209**(2), 1337–1351.

Butcher, A., Luckett, R., Verdon, J. P., Kendall, J.-M., Baptie, B., & Wookey, J., 2017. Local magnitude discrepancies for near-event receivers: Implications for the UK traffic-light scheme, *Bulletin of the Seismological Society of America*, **107**(2), 532–541.

Calderhead, B. & Girolami, M., 2009. Estimating Bayes factors via thermodynamic integration and population MCMC, *Computational Statistics & Data Analysis*, **53**(12), 4028–4045.

Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science*, **299**(5606), 547–549.

Castagna, J. P., Batzle, M. L., & Eastwood, R. L., 1985. Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks, *Geophysics*, **50**(4), 571–581.

Cerveny, V., 2005. *Seismic ray theory*, Cambridge university press.

Chan, K. S. & Geyer, C. J., 1994. Discussion: Markov chains for exploring posterior distributions, *The Annals of Statistics*, **22**(4), 1747–1758.

Chen, X., 1993. A systematic and efficient method of computing normal modes for multilayered half-space, *Geophysical Journal International*, **115**(2), 391–409.

Chiarabba, C. & Amato, A., 2003. Vp and Vp/Vs images in the Mw 6.0 Colfiorito fault region (central Italy): A contribution to the understanding of seismotectonic and seismogenic processes, *Journal of Geophysical Research: Solid Earth*, **108**(B5).

Çınlar, E., 2011. *Probability and stochastics*, vol. 261, Springer Science & Business Media.

Cole, S. P., 1995. *Passive seismic and drill-bit experiments using 2-D arrays*, no. 86, Stanford University.

Crampin, S. & Båth, M., 1965. Higher modes of seismic surface waves: mode separation, *Geophysical Journal International*, **10**(1), 81–92.

Crowder, E., Rawlinson, N., Pilia, S., Cornwell, D., & Reading, A., 2019. Transdimensional ambient noise tomography of Bass Strait, southeast Australia, reveals the sedimentary basin and deep crustal structure beneath a failed continental rift, *Geophysical Journal International*, **217**(2), 970–987.

Curtis, A. & Halliday, D., 2010. Directional balancing for seismic and general wavefield interferometry, *Geophysics*, **75**(1), SA1–SA14.

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

Curtis, A. & Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics*, **62**(5), 1524–1532.

Curtis, A. & Snieder, R., 2002. Probing the earth's interior with seismic tomography, *International Geophysics Series*, **81**(A), 861–874.

Curtis, A., Trampert, J., Snieder, R., & Dost, B., 1998. Eurasian fundamental mode surface wave phase velocities and their relationship with tectonic structures, *Journal of Geophysical Research: Solid Earth*, **103**(B11), 26919–26947.

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

De Freitas, N., Højen-Sørensen, P., Jordan, M. I., & Russell, S., 2001. Variational MCMC, in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 120–127, Morgan Kaufmann Publishers Inc.

de Pasquale, G. & Linde, N., 2016. On structure-based priors in bayesian geophysical inversion, *Geophysical Journal International*, **208**(3), 1342–1358.

de Ridder, S. & Biondi, B., 2013. Daily reservoir-scale subsurface monitoring using ambient seismic noise, *Geophysical Research Letters*, **40**(12), 2969–2974.

de Ridder, S. & Dellinger, J., 2011. Ambient seismic noise Eikonal tomography for near-surface imaging at Valhall, *The Leading Edge*, **30**(5), 506–512.

de Ridder, S., Biondi, B., & Clapp, R., 2014. Time-lapse seismic noise correlation tomography at Valhall, *Geophysical Research Letters*, **41**(17), 6116–6122.

de Ridder, S., Biondi, B., & Nichols, D., 2015. Elliptical-anisotropic Eikonal phase velocity tomography, *Geophysical Research Letters*, **42**(3), 758–764.

Detommaso, G., Cui, T., Marzouk, Y., Spantini, A., & Scheichl, R., 2018. A stein variational newton method, in *Advances in Neural Information Processing Systems*, pp. 9169–9179.

Dettmer, J. & Dosso, S. E., 2012. Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains, *The Journal of the Acoustical Society of America*, **132**(4), 2239–2250.

Dettmer, J., Hawkins, R., Cummins, P. R., Hossen, J., Sambridge, M., Hino, R., & Inazu, D., 2016. Tsunami source uncertainty estimation: The 2011 japan tsunami, *Journal of Geophysical Research: Solid Earth*, **121**(6), 4483–4505.

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.

Dosso, S. E., Holland, C. W., & Sambridge, M., 2012. Parallel tempering for strongly nonlinear geoacoustic inversion, *The Journal of the Acoustical Society of America*, **132**(5), 3030–3040.

Dosso, S. E., Dettmer, J., Steininger, G., & Holland, C. W., 2014. Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation, *Inverse Problems*, **30**(11), 114018.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D., 1987. Hybrid monte carlo, *Physics letters B*, **195**(2), 216–222.

Dziewonski, A., Bloch, S., & Landisman, M., 1969. A technique for the analysis of transient seismic signals, *Bulletin of the seismological Society of America*, **59**(1), 427–444.

Dziewonski, A. M. & Woodhouse, J. H., 1987. Global images of the Earth's interior, *Science*, **236**(4797), 37–48.

Earl, D. J. & Deem, M. W., 2005. Parallel tempering: Theory, applications, and new perspectives, *Physical Chemistry Chemical Physics*, **7**(23), 3910–3916.

Earp, S. & Curtis, A., 2019. Probabilistic neural-network based 2D travel time tomography, *arXiv preprint arXiv:1907.00541*.

Eberhart-Phillips, D. & Michael, A. J., 1998. Seismotectonics of the Loma Prieta, California, region determined from three-dimensional Vp, Vp/Vs, and seismicity, *Journal of Geophysical Research: Solid Earth*, **103**(B9), 21099–21120.

Ehlers, R. S. & Brooks, S. P., 2008. Adaptive proposal construction for reversible jump MCMC, *Scandinavian Journal of Statistics*, **35**(4), 677–690.

Ekström, G., 2011. A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25-250 s, *Geophysical Journal International*, **187**(3), 1668–1686.

Fang, H. & Zhang, H., 2014. Wavelet-based double-difference seismic tomography with sparsity regularization, *Geophysical Journal International*, **199**(2), 944–955.

Fang, H., Yao, H., Zhang, H., Huang, Y.-C., & van der Hilst, R. D., 2015. Direct inversion of surface wave dispersion for three-dimensional shallow crustal structure based on ray tracing: methodology and application, *Geophysical Journal International*, **201**(3), 1251–1263.

Fang, H., Zhang, H., Yao, H., Allam, A., Zigone, D., Ben-Zion, Y., Thurber, C., & van der Hilst, R. D., 2016. A new algorithm for three-dimensional joint inversion of body wave and surface wave data and its application to the Southern California plate boundary region, *Journal of Geophysical Research: Solid Earth*, **121**(5), 3557–3569.

Feroz, F. & Hobson, M. P., 2008. Multimodal nested sampling: an efficient and robust alternative to markov chain monte carlo methods for astronomical data analyses, *Monthly Notices of the Royal Astronomical Society*, **384**(2), 449–463.

Ferreira, A., Woodhouse, J., Visser, K., & Trampert, J., 2010. On the robustness of global radially anisotropic surface wave tomography, *Journal of Geophysical Research: Solid Earth*, **115**(B4).

Fichtner, A., Zunino, A., & Gebraad, L., 2018. Hamiltonian monte carlo solution of tomographic inverse problems, *Geophysical Journal International*, **216**(2), 1344–1363.

Fink, M., 1992. Time reversal of ultrasonic fields. i. basic principles, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, **39**(5), 555–566.

Gabriels, P., Snieder, R., & Nolet, G., 1987. In situ measurements of shear-wave velocity in sediments with higher-mode Rayleigh waves, *Geophysical prospecting*, **35**(2), 187–196.

Galetti, E. & Curtis, A., 2018. Transdimensional electrical resistivity tomography, *Journal of Geophysical Research: Solid Earth*, **123**(8), 6347–6377.

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.

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.

Geyer, C. J., 1994. Estimating normalizing constants and reweighting mixtures, *Technical Report*.

Gilks, W. R., Richardson, S., & Spiegelhalter, D., 1995. *Markov chain Monte Carlo in practice*, Chapman and Hall/CRC.

Gorham, J. & Mackey, L., 2015. Measuring sample quality with Stein's method, in *Advances in Neural Information Processing Systems*, pp. 226–234.

Gorham, J. & Mackey, L., 2017. Measuring sample quality with kernels, in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1292–1301, JMLR. org.

Green, P. J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, pp. 711–732.

Green, P. J. & Hastie, D. I., 2009. Reversible jump MCMC, *Genetics*, **155**(3), 1391–1403.

Green, P. J., Łatuszyński, K., Pereyra, M., & Robert, C. P., 2015. Bayesian computation: a summary of the current state, and samples backwards and forwards, *Statistics and Computing*, **25**(4), 835–862.

Gretton, A., 2013. Introduction to RKHS, and some simple kernel algorithms.

Haario, H., Saksman, E., Tamminen, J., et al., 2001. An adaptive metropolis algorithm, *Bernoulli*, **7**(2), 223–242.

Halliday, D. & Curtis, A., 2008. Seismic interferometry, surface waves and source distribution, *Geophysical Journal International*, **175**(3), 1067–1087.

Hansen, P. C., 1992. Analysis of discrete ill-posed problems by means of the L-curve, *SIAM review*, **34**(4), 561–580.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**(1), 97–109.

Hatchell, P., Wills, P., & Didraga, C., 2009. Production induced effects on near-surface wave velocities at Valhall, in *71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009*.

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophysical Journal International*, **203**(2), 972–1000.

Hawkins, R., Bodin, T., Sambridge, M., Choblet, G., & Husson, L., 2019. Trans-dimensional surface reconstruction with different classes of parameterization, *Geochemistry, Geophysics, Geosystems*, **20**(1), 505–529.

Herrin, E. & Goforth, T., 1977. Phase-matched filters: application to the study of Rayleigh waves, *Bulletin of the Seismological Society of America*, **67**(5), 1259–1275.

Herrmann, R. B., 2013. Computer programs in seismology: An evolving tool for instruction and research, *Seismological Research Letters*, **84**(6), 1081–1088.

Hoffman, M. D. & Blei, D. M., 2015. Structured stochastic variational inference, in *Artificial Intelligence and Statistics*.

Ing, R. K. & Fink, M., 1998. Time-reversed lamb waves, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, **45**(4), 1032–1043.

Iyer, H. & Hirahara, K., 1993. *Seismic tomography: Theory and practice*, Springer Science & Business Media.

Jamin, C., Pion, S., & Teillaud, M., 2018. 3D triangulations, in *CGAL User and Reference Manual*, CGAL Editorial Board, 4.12 edn.

Jan van Heijst, H. & Woodhouse, J., 1999. Global high-resolution phase velocity distributions of overtone and fundamental-mode surface waves determined by mode branch stripping, *Geophysical Journal International*, **137**(3), 601–620.

Karagiannis, G. & Andrieu, C., 2013. Annealed importance sampling reversible jump MCMC algorithms, *Journal of Computational and Graphical Statistics*, **22**(3), 623–648.

Karlin, S., 2014. *A first course in stochastic processes*, Academic press.

Käufl, P., Valentine, A. P., O'Toole, T. B., & Trampert, J., 2013. A framework for fast probabilistic centroid-moment-tensor determination – inversion of regional static displacement measurements, *Geophysical Journal International*, **196**(3), 1676–1693.

Käufl, P., Valentine, A., de Wit, R., & Trampert, J., 2015. Robust and fast probabilistic source parameter estimation from near-field displacement waveforms using pattern recognition, *Bulletin of the Seismological Society of America*, **105**(4), 2299–2312.

Khan, A., Zunino, A., & Deschamps, F., 2013. Upper mantle compositional variations and discontinuity topography imaged beneath Australia from Bayesian inversion of surface-wave phase velocities and thermochemical modeling, *Journal of Geophysical Research: Solid Earth*, **118**(10), 5285–5306.

Kingma, D. P. & Welling, M., 2013. Auto-encoding variational Byes, *arXiv preprint arXiv:1312.6114*.

Kritski, A., Vincent, A., Yuen, D., & Carlsen, T., 2006. Adaptive wavelets for analyzing dispersive seismic waves, *Geophysics*, **72**(1), V1–V11.

Kubrusly, C. & Gravier, J., 1973. Stochastic approximation algorithms and applications, in *1973 IEEE conference on decision and control including the 12th symposium on adaptive processes*, pp. 763–766, IEEE.

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.

Kurita, T., 1973. Regional variations in the structure of the crust in the central United States from P-wave spectra, *Bulletin of the Seismological Society of America*, **63**(5), 1663–1687.

Kuttig, H., Niethammer, M., Hurlebaus, S., & Jacobs, L. J., 2006. Model-based analysis of dispersion curves using chirplets, *The Journal of the Acoustical Society of America*, **119**(4), 2122–2130.

Kvenvolden, K. A., 1989. Seabed pockmarks and seepages: Impact on geology, biology and the marine environment, *Science*, **244**(4904), 590–592.

Lai, C. G. & Rix, G. J., 1998. Simultaneous inversion of rayleigh phase velocity and attenuation for near-surface site characterization.

Lees, J. M., 1990. Tomographic P-wave velocity images of the Loma Prieta earthquake asperity, *Geophysical Research Letters*, **17**(9), 1433–1436.

Levshin, A. & Ritzwoller, M., 2001. Automated detection, extraction, and measurement of regional surface waves, in *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Surface Waves*, pp. 1531–1545, Springer.

Levshin, A., Ratnikova, L., & Berger, J., 1992. Peculiarities of surface-wave propagation across central Eurasia, *Bulletin of the Seismological Society of America*, **82**(6), 2464–2493.

Levshin, A. L., Pisarenko, V., & Pogrebinsky, G., 1972. On a frequency-time analysis of oscillations, in *Annales de Geophysique*, vol. 28, pp. 211–218, Centre National de la Recherche Scientifique.

Li, Z., Tian, B., Liu, S., & Yang, J., 2013. Asperity of the 2013 lushan earthquake in the eastern margin of Tibetan Plateau from seismic tomography and aftershock relocation, *Geophysical Journal International*, **195**(3), 2016–2022.

Lin, F.-C., Ritzwoller, M. H., Townend, J., Bannister, S., & Savage, M. K., 2007. Ambient noise Rayleigh wave tomography of New Zealand, *Geophysical Journal International*, **170**(2), 649–666.

Lin, F.-C., Moschetti, M. P., & Ritzwoller, M. H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophysical Journal International*, **173**(1), 281–298.

Lin, F.-C., Ritzwoller, M. H., & Snieder, R., 2009. Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array, *Geophysical Journal International*, **177**(3), 1091–1110.

Liu, C. & Zhu, J., 2018. Riemannian stein variational gradient descent for bayesian inference, in *Thirty-Second AAAI Conference on Artificial Intelligence*.

Liu, Q., 2017. Stein variational gradient descent as gradient flow, in *Advances in neural information processing systems*, pp. 3115–3123.

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

Liu, Q., Lee, J., & Jordan, M., 2016. A kernelized Stein discrepancy for goodness-of-fit tests, in *International Conference on Machine Learning*, pp. 276–284.

Liu, X. & Zhao, D., 2016. P and S wave tomography of Japan subduction zone from joint inversions of local and teleseismic travel times and surface-wave data, *Physics of the Earth and Planetary Interiors*, **252**, 1–22.

Lomax, A. & Curtis, A., 2001. Fast, probabilistic earthquake location in 3D models using oct-tree importance sampling, in *Geophys. Res. Abstr*, vol. 3, p. 955.

Lomax, A., Michelini, A., & Curtis, A., 2009. Earthquake location, direct, global-search methods, *Encyclopedia of complexity and systems science*, pp. 2449–2473.

MacKay, D. J. & Mac Kay, D. J., 2003. *Information theory, inference and learning algorithms*, Cambridge university press.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophysical Journal International*, **151**(3), 675–688.

Malinverno, A. & Briggs, V. A., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics*, **69**(4), 1005–1016.

Malinverno, A., Leaney, S., et al., 2000. A Monte Carlo method to quantify uncertainty in the inversion of zero-offset VSP data, in *2000 SEG Annual Meeting*, Society of Exploration Geophysicists.

Marzouk, Y., Moselhy, T., Parno, M., & Spantini, A., 2016. An introduction to sampling via measure transport, *arXiv preprint arXiv:1602.05023*.

McGrory, C. A. & Titterington, D., 2007. Variational approximations in bayesian model selection for finite mixture distributions, *Computational Statistics & Data Analysis*, **51**(11), 5352–5367.

Meier, U., Curtis, A., & Trampert, J., 2007a. Global crustal thickness from neural network inversion of surface wave data, *Geophysical Journal International*, **169**(2), 706–722.

Meier, U., Curtis, A., & Trampert, J., 2007b. A global crustal model constrained by nonlinearised inversion of fundamental mode surface waves, *Geophysical Research Letters*, **34**, L16304.

Metropolis, N. & Ulam, S., 1949. The Monte Carlo method, *Journal of the American statistical association*, **44**(247), 335–341.

Mordret, A., Landès, M., Shapiro, N., Singh, S., Roux, P., & Barkved, O., 2013a. Near-surface study at the Valhall oil field from ambient noise surface wave tomography, *Geophysical Journal International*, **193**(3), 1627–1643.

Mordret, A., Shapiro, N. M., Singh, S. S., Roux, P., & Barkved, O. I., 2013b. Helmholtz tomography of ambient noise surface wave data to estimate Scholte wave phase velocity at Valhall life of the fieldnoise tomography at Valhall, *Geophysics*, **78**(2), WA99–WA109.

Mordret, A., Landès, M., Shapiro, N., Singh, S., & Roux, P., 2014a. Ambient noise surface wave tomography to determine the shallow shear velocity structure at Valhall: depth inversion with a neighbourhood algorithm, *Geophysical Journal International*, **198**(3), 1514–1525.

Mordret, A., Shapiro, N. M., & Singh, S., 2014b. Seismic noise-based time-lapse monitoring of the Valhall overburden, *Geophysical Research Letters*, **41**(14), 4945–4952.

Mosegaard, K., 2019. Efficient monte carlo uncertainty quantification through problem-dependent proposals, in *81st EAGE Conference and Exhibition 2019 Workshop Programme*.

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

Nakanishi, I. & Anderson, D. L., 1983. Measurement of mantle wave velocities and inversion for lateral heterogeneity and anisotropy: 1. analysis of great circle phase velocities, *Journal of Geophysical Research: Solid Earth*, **88**(B12), 10267–10283.

Nawaz, M. & Curtis, A., 2019. Rapid discriminative variational Byesian inversion of geophysical data for the spatial distribution of geological properties, *Journal of Geophysical Research: Solid Earth*.

Nawaz, M. 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.

Neal, R. M., 1993. *Probabilistic inference using Markov chain Monte Carlo methods*.

Neal, R. M. et al., 2003. Slice sampling, *The annals of statistics*, **31**(3), 705–767.

Neal, R. M. et al., 2011. Mcmc using hamiltonian dynamics, *Handbook of markov chain monte carlo*, **2**(11), 2.

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.

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.

Nunn, C., Roecker, S. W., Tilmann, F. J., Priestley, K. F., Heyburn, R., Sandvol, E. A., Ni, J. F., Chen, Y. J., Zhao, W., & Team IV, t. I., 2013. Imaging the lithosphere beneath NE Tibet: teleseismic P and S body wave tomography incorporating surface wave starting models, *Geophysical Journal International*, **196**(3), 1724–1741.

Obrebski, M., Allen, R. M., Pollitz, F., & Hung, S.-H., 2011. Lithosphere–asthenosphere interaction beneath the western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities, *Geophysical Journal International*, **185**(2), 1003–1021.

Obrebski, M., Allen, R. M., Zhang, F., Pan, J., Wu, Q., & Hung, S.-H., 2012. Shear wave tomography of China using joint inversion of body and surface wave constraints, *Journal of Geophysical Research: Solid Earth*, **117**(B1).

Park, C. B., Miller, R. D., & Xia, J., 1999a. Multichannel analysis of surface waves, *Geophysics*, **64**(3), 800–808.

Park, C. B., Miller, R. D., Xia, J., Hunter, J. A., & Harris, J. B., 1999b. Higher mode observation by the MASW method, in *SEG Technical Program Expanded Abstracts 1999*, pp. 524–527, Society of Exploration Geophysicists.

Piana Agostinetti, N., Giacomuzzi, G., & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophysical Journal International*, **201**(3), 1598–1617.

Pragt, J., Herberg, W., Meister, M., Clemmensen, C. C., Grindhaug, G., Hanken, K. J., et al., 2012. Reaming on demand-selective activation of an integrated under reamer at the Grane field in the North Sea, in *SPE Deepwater Drilling and Completions Conference*, Society of Petroleum Engineers.

Ranganath, R., Gerrish, S., & Blei, D., 2014. Black box variational inference, in *Artificial Intelligence and Statistics*, pp. 814–822.

Ranganath, R., Tran, D., & Blei, D., 2016. Hierarchical variational models, in *International Conference on Machine Learning*, pp. 324–333.

Rawlinson, N. & Fishwick, S., 2012. Seismic structure of the southeast Australian lithosphere from surface and body wave tomography, *Tectonophysics*, **572**, 111–122.

Rawlinson, N. & Sambridge, M., 2004. Multiple reflection and transmission phases in complex layered media using a multistage fast marching method, *Geophysics*, **69**(5), 1338–1350.

Rawlinson, N., Fichtner, A., Sambridge, M., & Young, M. K., 2014. Seismic tomography and the assessment of uncertainty, in *Advances in Geophysics*, vol. 55, pp. 1–76, Elsevier.

Ray, A. & Myer, D., 2019. Bayesian geophysical inversion with trans-dimensional gaussian process machine learning, *Geophysical Journal International*, **217**(3), 1706–1726.

Ray, A., Alumbaugh, D. L., Hoversten, G. M., & Key, K., 2013. Robust and accelerated Bayesian inversion of marine controlled-source electromagnetic data using parallel tempering, *Geophysics*, **78**(6), E271–E280.

Ray, A., Kaplan, S., Washbourne, J., & Albertin, U., 2017. Low frequency full waveform seismic inversion within a tree based Bayesian framework, *Geophysical Journal International*, **212**(1), 522–542.

Reiter, D. T. & Rodi, W. L., 2008. A new regional 3-D velocity model for Asia from the joint inversion of p-wave travel times and surface-wave dispersion data, Tech. rep., WESTON GEOPHYSICAL LEXINGTON MA.

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

Ritzwoller, M. H. & Levshin, A. L., 1998. Eurasian surface wave tomography: Group velocities, *Journal of Geophysical Research: Solid Earth*, **103**(B3), 4839–4878.

Ritzwoller, M. H., Shapiro, N. M., Barmin, M. P., & Levshin, A. L., 2002. Global surface wave diffraction tomography, *Journal of Geophysical Research: Solid Earth*, **107**(B12).

Robbins, H. & Monro, S., 1951. A stochastic approximation method, *The annals of mathematical statistics*, pp. 400–407.

Robert, C. & Casella, G., 2013. *Monte Carlo statistical methods*, Springer Science & Business Media.

Roecker, S., Ebinger, C., Tiberi, C., Mulibo, G., Ferdinand-Wambura, R., Mtelela, K., Kianji, G., Muzuka, A., Gautier, S., Albaric, J., et al., 2017. Subsurface images of the Eastern Rift, Africa, from the joint inversion of body waves, surface waves and gravity: investigating the role of fluids in early-stage continental rifting, *Geophysical Journal International*, **210**(2), 931–950.

Romanowicz, B. A., 1979. Seismic structure of the upper mantle beneath the united states by three-dimensional inversion of body wave arrival times, *Geophysical Journal International*, **57**(2), 479–506.

Rost, S. & Thomas, C., 2002. Array seismology: Methods and applications, *Reviews of geophysics*, **40**(3), 2–1.

Röth, G. & Tarantola, A., 1994. Neural networks and inversion of seismic data, *Journal of Geophysical Research: Solid Earth*, **99**(B4), 6753–6768.

Russell, D. R., Herrmann, R. B., & Hwang, H.-J., 1988. Application of frequency variable filters to surface-wave amplitude analysis, *Bulletin of the Seismological Society of America*, **78**(1), 339–354.

Saito, M., 1988. DISPER80: A subroutine package for the calculation of seismic normal-mode solutions, *Seismological algorithms*, pp. 293–319.

Salimans, T., Kingma, D., & Welling, M., 2015. Markov chain Monte Carlo and variational inference: Bridging the gap, in *International Conference on Machine Learning*, pp. 1218–1226.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm – i. searching a parameter space, *Geophysical journal international*, **138**(2), 479–494.

Sambridge, M., 2013. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophysical Journal International*, p. ggt342.

Sambridge, M., Braun, J., & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophysical Journal International*, **122**(3), 837–857.

Sato, M.-A., 2001. Online model selection based on the variational bayes, *Neural computation*, **13**(7), 1649–1681.

Saul, L. K. & Jordan, M. I., 1996. Exploiting tractable substructures in intractable networks, in *Advances in neural information processing systems*, pp. 486–492.

Savage, M. K., Lin, F.-C., & Townend, J., 2013. Ambient noise cross-correlation observations of fundamental and higher-mode rayleigh wave propagation governed by basement resonance, *Geophysical Research Letters*, **40**(14), 3556–3561.

Saygin, E., Cummins, P., Cipta, A., Hawkins, R., Pandhu, R., Murjaya, J., Irsyam, M., Widiyantoro, S., & Kennett, B., 2015. Imaging architecture of the Jakarta Basin, Indonesia with transdimensional inversion of seismic noise, *Geophysical Journal International*, **204**(2), 918–931.

Sen, M. K. & Biswas, R., 2017. Transdimensional seismic inversion using the reversible jump hamiltonian monte carlo algorithm, *Geophysics*, **82**(3), R119–R134.

Sen, M. K. & Stoffa, P. L., 2013. *Global optimization methods in geophysical inversion*, Cambridge University Press.

Sengupta, B., Friston, K. J., & Penny, W. D., 2015. Gradient-free MCMC methods for dynamic causal modelling, *NeuroImage*, **112**, 375–381.

Shahraeeni, M. S. & Curtis, A., 2011. Fast probabilistic nonlinear petrophysical inversion, *Geophysics*, **76**(2), E45–E58.

Shahraeeni, M. S., Curtis, A., & Chao, G., 2012. Fast probabilistic petrophysical mapping of reservoirs from 3D seismic data, *Geophysics*, **77**(3), O1–O19.

Shapiro, N. & Ritzwoller, M., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophysical Journal International*, **151**(1), 88–105.

Shapiro, N. M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophysical Research Letters*, **31**(7).

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.

Shearer, P., 1999. *Introduction to Seismology*, Cambridge University Press.

Shen, W., Ritzwoller, M. H., Schulte-Pelkum, V., & Lin, F.-C., 2012. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophysical Journal International*, **192**(2), 807–836.

Shen, W., Ritzwoller, M. H., & Schulte-Pelkum, V., 2013. A 3-D model of the crust and uppermost mantle beneath the central and western US by joint inversion of receiver functions and surface wave dispersion, *Journal of Geophysical Research: Solid Earth*, **118**(1), 262–276.

Simons, F. J., Van Der Hilst, R. D., Montagner, J.-P., & Zielhuis, A., 2002. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle, *Geophysical Journal International*, **151**(3), 738–754.

Sivia, D., 1996. Data analysis: A Bayesian tutorial (oxford science publications).

Skilling, J., 2004. Nested sampling, in *AIP Conference Proceedings*, vol. 735, pp. 395–405, AIP.

Skilling, J. et al., 2006. Nested sampling for general bayesian computation, *Bayesian analysis*, **1**(4), 833–859.

Smith, A., 2013. *Sequential Monte Carlo methods in practice*, Springer Science & Business Media.

Snoke, J. A. & Sambridge, M., 2002. Constraints on the S wave velocity structure in a continental shield from surface wave data: Comparing linearized least squares inversion and the direct search neighbourhood algorithm, *Journal of Geophysical Research: Solid Earth*, **107**(B5).

Stein, C. et al., 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, The Regents of the University of California.

Stevens, J. L., Adams, D. A., & Baker, G. E., 2001. Improved surface wave detection and measurement using phase-matched filtering with a global one-degree dispersion model, Tech. rep., SCIENCE APPLICATIONS INTERNATIONAL CORP SAN DIEGO CA.

Syracuse, E. M., Maceira, M., Zhang, H., & Thurber, C., 2015. Seismicity and structure of Akutan and Makushin Volcanoes, Alaska, using joint body and surface wave tomography, *Journal of Geophysical Research: Solid Earth*, **120**(2), 1036–1052.

Tajima, R., Tajima, F., & Kato, A., 2009. Seismic structure in and around the source area of the 2004 mid-Niigata, Japan, earthquake: 3-D waveform modelling based on local tomography images, *Geophysical Journal International*, **177**(1), 145–160.

Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*, vol. 89, SIAM.

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

Team, S. D. et al., 2016. Stan modeling language users guide and reference manual, *Technical report*.

Theodoridis, S., 2015. *Machine learning: a Bayesian and optimization perspective*, Academic Press.

Thompson, M., Andersen, M., Elde, R., Roy, S., & Skogland, S., 2015. The startup of permanent reservoir monitoring for snorre and grane, in *77th EAGE Conference and Exhibition 2015*.

Thurber, C. H., 1983. Earthquake locations and three-dimensional crustal structure in the coyote lake area, central california, *Journal of Geophysical Research: Solid Earth*, **88**(B10), 8226–8236.

Tomar, G., Stutzmann, E., Mordret, A., Montagner, J.-P., Singh, S. C., & Shapiro, N. M., 2018. Joint inversion of the first overtone and fundamental mode for deep imaging at the valhall oil field using ambient noise, *Geophysical Journal International*, **214**(1), 122–132.

Trampert, J. & Woodhouse, J. H., 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophysical Journal International*, **122**(2), 675–690.

Tran, D., Ranganath, R., & Blei, D. M., 2015. The variational Gaussian process, *arXiv preprint arXiv:1511.06499*.

Valero-Gomez, A., Gomez, J. V., Garrido, S., & Moreno, L., 2013. The path to efficiency: fast marching method for safer, more efficient mobile robot trajectories, *IEEE Robotics & Automation Magazine*, **20**(4), 111–120.

Van der Hilst, R. D., Widiyantoro, S., & Engdahl, E., 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**(6625), 578.

van Heijst, H. J. & Woodhouse, J., 1997. Measuring surface-wave overtone phase velocities using a mode-branch stripping technique, *Geophysical Journal International*, **131**(2), 209–230.

van Manen, D.-J., Robertsson, J. O., & Curtis, A., 2005. Modeling of wave propagation in inhomogeneous media, *Physical Review Letters*, **94**(16), 164301.

van Manen, D.-J., Curtis, A., & Robertsson, J. O., 2006. Interferometric modeling of wave propagation in inhomogeneous elastic media using time reversal and reciprocity, *Geophysics*, **71**(4), SI47–SI60.

Walter, E. & Pronzato, L., 1997. *Identification of parametric models from experimental data*, Springer Verlag.

Wapenaar, K., 2004. Retrieving the elastodynamic green's function of an arbitrary inhomogeneous medium by cross correlation, *Physical review letters*, **93**(25), 254301.

Wapenaar, K. & Fokkema, J., 2006. GreenâĂŹs function representations for seismic interferometry, *Geophysics*, **71**(4), SI33–SI46.

Weaver, R. L., Hadziioannou, C., Larose, E., & Campillo, M., 2011. On the precision of noise correlation interferometry, *Geophysical Journal International*, **185**(3), 1384–1392.

Webb, S. C., 1998. Broadband seismology and noise under the ocean, *Reviews of Geophysics*, **36**(1), 105–142.

West, M., Gao, W., & Grand, S., 2004. A simple approach to the joint inversion of seismic body and surface waves applied to the southwest US, *Geophysical research letters*, **31**(15).

Wielandt, E., 1993. Propagation and structural interpretation of non-plane waves, *Geophysical Journal International*, **113**(1), 45–53.

Wilcox, P. D., 2003. A rapid signal processing technique to remove the effect of dispersion from guided wave signals, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, **50**(4), 419–427.

Woodhouse, J. & Dahlen, F., 1978. The effect of a general aspherical perturbation on the free oscillations of the earth, *Geophysical Journal International*, **53**(2), 335–354.

Wu, B. & Chen, X., 2016. Stable, accurate and efficient computation of normal modes for horizontal stratified models, *Geophysical Journal International*, **206**(2), 1281–1300.

Xia, J., Miller, R. D., & Park, C. B., 2000. Advantages of calculating shear-wave velocity from surface waves with higher modes, in *SEG Technical Program Expanded Abstracts 2000*, pp. 1295–1298, Society of Exploration Geophysicists.

Xia, J., Miller, R. D., Park, C. B., & Tian, G., 2003. Inversion of high frequency surface waves with fundamental and higher modes, *Journal of Applied Geophysics*, **52**(1), 45–57.

Xu, K., Ta, D., Moilanen, P., & Wang, W., 2012. Mode separation of lamb waves based on dispersion compensation method, *The Journal of the Acoustical Society of America*, **131**(4), 2714–2722.

Yang, D. & Oldenburg, D. W., 2012. Three-dimensional inversion of airborne time-domain electromagnetic data with applications to a porphyry deposit, *Geophysics*, **77**(2), B23–B34.

Yang, Y., Ritzwoller, M. H., Levshin, A. L., & Shapiro, N. M., 2007. Ambient noise Rayleigh wave tomography across Europe, *Geophysical Journal International*, **168**(1), 259–274.

Yanovskaya, T., Levshin, A., Its, E., Lander, A., Bukchin, B., Barmin, M., & Ratnikova, L., 2012. *Seismic surface waves in a laterally inhomogeneous Earth*, vol. 9, Springer Science & Business Media.

Yao, H. & Van Der Hilst, R. D., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet, *Geophysical Journal International*, **179**(2), 1113–1132.

Yao, H., van Der Hilst, R. D., & De Hoop, M. V., 2006. Surface-wave array tomography in SE tibet from ambient seismic noise and two-station analysis–i. phase velocity maps, *Geophysical Journal International*, **166**(2), 732–744.

Yilmaz, Ö., 2001. *Seismic data analysis*, vol. 1, Society of exploration geophysicists.

Young, M. K., Rawlinson, N., & Bodin, T., 2013. Transdimensional inversion of ambient seismic noise for 3D shear velocity structure of the Tasmanian crust, *Geophysics*, **78**(3), WB49–WB62.

Zhang, H. & Thurber, C., 2005. Adaptive mesh seismic tomography based on tetrahedral and voronoi diagrams: application to parkfield, california, *Journal of Geophysical Research: Solid Earth*, **110**(B4).

Zhang, H. & Thurber, C. H., 2003. Double-difference tomography: The method and its application to the hayward fault, california, *Bulletin of the Seismological Society of America*, **93**(5), 1875–1889.

Zhang, H., Maceira, M., Roux, P., & Thurber, C., 2014. Joint inversion of body-wave arrival times and surface-wave dispersion for three-dimensional seismic structure around SAFOD, *Pure and Applied Geophysics*, **171**(11), 3013–3022.

Zhang, X. & Curtis, A., 2019. Seismic tomography using variational inference methods, *submitted*.

Zhang, X. & Zhang, H., 2015. Wavelet-based time-dependent travel time tomography method and its application in imaging the Etna volcano in Italy, *Journal of Geophysical Research: Solid Earth*, **120**(10), 7068–7084.

Zhang, X., Zhang, H., Wang, H., & Pei, S., 2013. Seismic velocity imaging of the aftershock zone of the 2013 Mw 6.6 lushan Earthquake, China, in *AGU Fall Meeting Abstracts*.

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., Hansteen, F., & Curtis, A., 2019. Fully 3D Monte Carlo ambient noise tomography over Grane field, in *81st EAGE Conference and Exhibition 2019*.

Zhdanov, M. S., 2002. *Geophysical inverse theory and regularization problems*, vol. 36, Elsevier.

Zheng, D., Saygin, E., Cummins, P., Ge, Z., Min, Z., Cipta, A., & Yang, R., 2017. Transdimensional Bayesian seismic ambient noise tomography across SE Tibet, *Journal of Asian Earth Sciences*, **134**, 86–93.

Zhou, Y., Dahlen, F., & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophysical Journal International*, **158**(1), 142–168.

Zielhuis, A. & Nolet, G., 1994. Deep seismic expression of an ancient plate boundary in Europe, *Science*, **265**(5168), 79–81.

Zigone, D., Ben-Zion, Y., Campillo, M., & Roux, P., 2015. Seismic tomography of the Southern California plate boundary region from noise-based Rayleigh and Love waves, *Pure and Applied Geophysics*, **172**(5), 1007–1032.

Zobay, O. et al., 2014. Variational bayesian inference with gaussian-mixture approximations, *Electronic Journal of Statistics*, **8**(1), 355–389.

Zulfakriza, Z., Saygin, E., Cummins, P., Widiyantoro, S., Nugraha, A. D., Lühr, B.-G., & Bodin, T., 2014. Upper crustal structure of central Java, Indonesia, from transdimensional seismic ambient noise tomography, *Geophysical Journal International*, **197**(1), 630–635.