Advances in Trans-dimensional Geophysical Inference

Rhys Hawkins

A thesis submitted for the degree of
Doctor of Philosohpy
The Australian National University

August, 2017

Declaration

This thesis is an account of research undertaken between February 2014 and August 2017 at Research School of Earth Sciences, The Australian National University, Canberra, Australia.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Rhys Hawkins

August, 2017

Acknowledgements

No work occurs in isolation. Malcolm Sambridge is thanked for convincing me to give up a well paying career to become a poor student again. I've enjoyed the freedom you've given me to pursue research during the course of my PhD and appreciated all that you've done to get me to this stage.

Jan Dettmer is one of my thesis advisors who deserves special mention. I learnt a great deal from our interactions while we shared an office and enjoyed the collaborative work on Tsunami sea surface deformation [Dettmer et al., 2016], even the fractally wrong initial review we received.

My other thesis advisors, Hrvoje Tkalčić and Stephen Roberts are thanked for their inputs and advice at various stages of this thesis.

Brian Kennett, who whether he wanted it or not became my adopted thesis advisor, is thanked for the many corridor conversations and discussions which contributed to aspects of Chapters 2 and 7.

Erdinc Saygin for discussions on group velocity extraction and ambient noise data processing.

Syeongryong Kim for introducing the paper by Ekström on the Aki phase velocity extraction method.

Josip Stepčivić for his group velocity estimates obtained from Iceland.

I would like to thank Thomas Bodin for the original version of the Voronoi tomography code, Nicholas Rawlinson for making his 3D model and rays available for our 3D teleseismic tests, both of which feature in Chapter 3.

Portions of Chapter 3 were published in Geophyiscal Journal International in 2015 [Hawkins and Sambridge, 2015]. I would like to thank Niklas Linde and Alberto Malinverno for their constructive comments which improved the submitted manuscript and are incorporated into this chapter.

Portions of Chapter 4 were published in Exploration Geophysics in 2017 [Hawkins et al., 2017]. I would like to thank Ross Brodie for this collaborative project. Murrary Richardson and Yusen Ley-Cooper are thanked for reviewing an earlier draft. Niklas Linde and an anonymous reviewer are thanked for constructive comments that improved the submitted manuscript and this chapter.

The data used in Chapter 4 is courtesy of Geoscience Australia from the Broken Hill Managed Aquifer Recharge Project.

I would like to thank the ACT Branch of the Australian Society of Exploration Geophysicists for their scholarship awarded in 2016 which motivated the study in Chapter 4.

The kernel of the idea for the parameterisation used in Chapter 6 came indirectly from Alberto Malinverno in his review of the Trans-dimensional Tree paper [Hawkins and Sambridge, 2015]. Stephanie Durand and Eric Debalye are thanked for making their data available. Eric Debayle is thanked for sharing his code for the inversion of the shear wave velocity model from 1D multi-mode dispersion curves.

Julian Byrne has provided invaluable assistance with advice for the use of the Terrawulf cluster, a computational facility supported through the AuScope Australian Geophysical Observing System (AGOS) and the National Collaborative Research Infrastructure Strategy (NCRIS), both Australian Federal Governments programmes.

This thesis was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

Aspects of this research were supported under Australian Research Council Discovery grant scheme, project DP110102098.

This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

Peter and Yvonne, my parents, for their support and encouragement.

Kathryn, my partner, for her love and support.

Abstract

This research presents a series of novel Bayesian trans-dimensional methods for geophysical inversion. A first example illustrates how Bayesian prior information obtained from theory and numerical experiments can be used to better inform a difficult multimodal inversion of dispersion information from empirical Greens functions obtained from ambient noise cross-correlation. This approach is an extension of existing partition modeling schemes.

An entirely new class of trans-dimensional algorithm, called the trans-dimensional tree method is introduced. This new method is shown to be more efficient at coupling to a forward model, more efficient at convergence, and more adaptable to different dimensions and geometries than existing approaches. The efficiency and flexibility of the trans-dimensional tree method is demonstrated in two different examples: (1) airborne electromagnetic tomography (AEM) in a 2D transect inversion, and (2) a fully non-linear inversion of ambient noise tomography. In this latter example the resolution at depth has been significantly improved by inverting a contiguous band of frequencies jointly rather than as independent phase velocity maps, allowing new insights into crustal architecture beneath Iceland.

In a first test case for even larger scale problems, an application of the trans-dimensional tree approach to large global data set is presented. A global database of nearly 5 million multi-model path average Rayleigh wave phase velocity observations has been used to construct global phase velocity maps. Results are comparable to existing published phase velocity maps, however, as the trans-dimensional approach adapts the resolution appropriate to the data, rather than imposing damping or smoothing constraints to stabilize the inversion, the recovered anomaly magnitudes are generally higher with low uncertainties. While further investigation is needed, this early test case shows that trans-dimensional sampling can be applied to global scale seismology problems and that previous analyses may, in some locales, under estimate the heterogeneity of the Earth.

Finally, in a further advancement of partition modelling with variable order polynomials, a new method has been developed called trans-dimensional spectral elements.

Previous applications involving variable order polynomials have used polynomials that are both difficult to work with in a Bayesian framework and unstable at higher orders. By using the orthogonal polynomials typically used in modern full-waveform solvers, the useful properties of this type of polynomial and its application in trans-dimensional inversion are demonstrated. Additionally, these polynomials can be directly used in complex differential solvers and an example of this for 1D inversion of surface wave dispersion curves is given.

Contents

Chapter 1

Introduction

1.1 Introduction

A large part of our understanding of the formation, composition and geodynamism of the Earth, is based upon geophysical imaging at various scales, from near surface techniques for mineral exploration, site hazard identification, and aquifer studies to near antipodal long period seismic waves that sense the structure of the Earth's core.

In many geophysical inverse problems, the physical nature of the relationship between candidate models of the Earth and limited observations, result in under-determined problems that causes ambiguities. An example is the unavoidable circularity in seismic travel time tomography where understanding wave speed variations in the Earth depends on time and location information of earthquakes measured with seismometers. Accurate determination of earthquake hypo-centres and rupture times depends on a wave speed model of the Earth to compute radial distances from multiple seismometers using triangulation. This circular issue is often described as the source/structure trade-off.

As methods and data improve for studying the deeper Earth, inferences of mantle sensitive observations depend on models of the crust. Similarly, inferences from inner core observations depend on the model of the mantle and crust. If models at various scales, for example, crustal, lithospherical, or mantle, are constructed with robust uncertainty estimates, then inferences of composition, geodynamism, or other physical processes can be performed and tested in statistically meaningful ways.

The focus of this research is on new methods for the inversion of geophysical data using ensemble approaches for uncertainty analysis and model comparison. Although seismic tomography is the focus, many of the concepts presented can be equally applied to more general geophysical imaging problems.

1.2 Overview of geophysical imaging problems

In geophysical imaging problems, the goal is to determine spatially varying parameters of a planetary object from proxy observations that can be related back to properties of interest via mathematical modelling. A non-exhaustive list of examples of these class of problems are inverting for the density variation within planetary bodies from gravity potential field anomalies [Swenson and Wahr, 2002, Pavlis et al., 2012, Zuber et al., 2013], determining conductivity or impedance from passive or actively induced eddy currents [Tikhonov, 1950, Cagniard, 1953, Simpson and Bahr, 2005, Chave and Jones, 2012], and lastly estimating variations in seismic wave speed through the use of recorded seismograms [Aki, 1977, Aki and Richards, 2002, Rawlinson and Sambridge, 2003, Rawlinson et al., 2014]. All these problems fall under the general umbrella of geophysical inverse problems.

In the general formulation of a geophysical inverse problem [Menke, 1989, Tarantola, 2005], there are some observations, **d**, where the bold face represents a vector of values, that is $\mathbf{d} = [d_1 \ldots d_n]^T$. The goal is to find some model, **m**, that best explains the observations. Through the application of a forward model, G, predicted observations can be generated from the model and these can be directly compared to the observations. The solution of this problem generally minimises some norm of the difference between the predictions of the model and the observations, that is, in the simplest case

$$arg\,min_{\mathbf{m}} ||G(\mathbf{m}) - \mathbf{d}||_p. \tag{1.1}$$

In general, methods for solving the inverse problem fall into one of two categories: those that produce a single optimal model given some data fit criteria and model penalties, for example Thurber [1983], and those that produce an ensemble of models, for example Mosegaard and Tarantola [1995].

The focus of this work is on the later class, motivated by the following reasons: (1) uncertainty analysis is important. Ensemble based methods provide greater statistical information for both uncertainty and validation of the formulation of the problem.

In optimisation based methods, uncertainty analysis is typically a linear Gaussian approximation using the Hessian of the model covariance matrix that can produce overly optimistic estimates of uncertainty and fails to account for multi-modal and long-tailed uncertainties. Ensembles on the other hand allow population based statistical analysis for hypothesis testing, statistical significance and model comparisons. (2) The majority of optimisation approaches require stabilisation in terms of either damping to a reference model, a model smoothness constraint, or combination thereof, and these have a tendency to reduce the resolution of features in an inversion and attenuate the recovered magnitude of model parameters of interest. (3) An ensemble provides far greater information about the potential solution than a single optimised model. It has been said that “all models are wrong, but some are useful” [Box and Draper, 1987], but in an ensemble approach it may be more appropriate to say that “all these models are wrong, but together they are useful”.

Common to both optimisation and ensemble based approaches is that in many geophysical problems, the observed measurements are spatially distributed in a sparse and irregular manner or have non-uniform sensitivities over the spatial domain. This can be due to the feasible location of measurement devices, for example, requiring a sensor to be on land when much of the Earth is covered by ocean, or simply that the location of events is unevenly distributed, for example, large magnitude earthquakes are most commonly located along plate tectonic boundaries.

Similarly, in many geophysical problems where sub-surface properties are explored with surface measurements, the ability to resolve features diminishes with depth and can be further compromised where strong near surface features are present.

This causes significant difficulties in parameterisation of the inverse problem due to a trade off between spatial resolvability and the conditioning of the problem. In a simple example, consider a linearised 2D tomography problem in which there are a set of known rays with observed travel times and that is to be inverted for the slowness field. If the model is parameterised as a regular 2D Cartesian grid with constant values in each grid cell, the inverse problem can be solved with linear least squares solution.

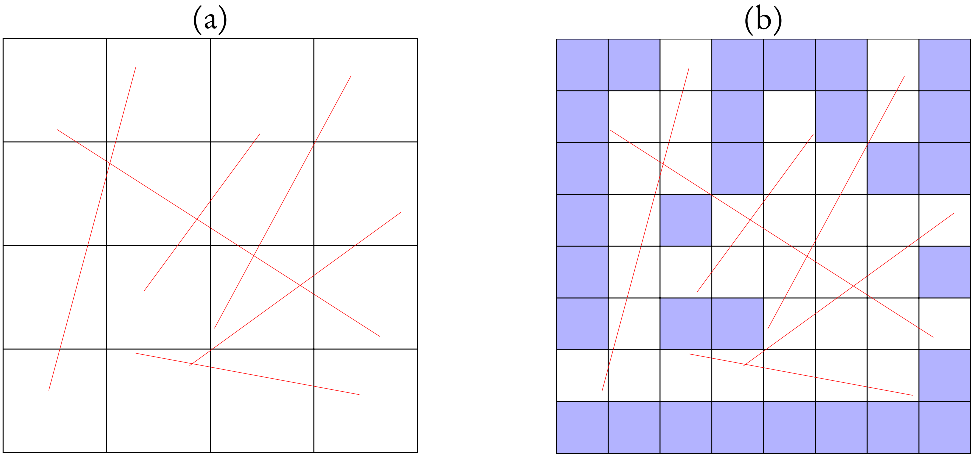


Figure 1.1: A simple example of how the choice of grid cell size can cause instabilities in the inversion. In (a) is a coarse grid where every cell is covered by a ray. In (b) with a finer grid, some cells, indicated with blue shading, have no ray coverage and are therefore unconstrained by the observations.

However, the number of grid cells or resolution of the grid needs to be chosen. If the grid is chosen to be too coarse, then the full information from the observations where coverage is dense is recovered only approximately. Conversely, if the grid is set too fine then the problem becomes ill-conditioned. An example is shown in Figure 1.1 where in (a) a coarse grid is well conditioned as there are rays through every cell, whereas in (b) many grid cells have no ray coverage highlighted in light blue.

Resorting to coarse grids to stabilise the inversion produces models without the expected degree of smoothness of lateral variations. This has prompted some solutions to overcome the problem of low resolution grids that do not present expected laterally smooth variations. An example, in the original seismic tomography paper introducing what is now known as teleseismic tomography, Aki [1977] used two solutions at the same resolution but with a half grid diagonal offset between the two grids. This then allows smooth contour plots, for example Figure 9(a) therein, using 4 point averaging.

This heterogeneity in resolvability has prompted others to use irregular parameterisations that adapt to the data in some fashion. Sambridge and Faletič [2003] use a criterion based on the maximum spatial gradients in the seismic velocity perturbation to iteratively subdivide a tetrahedral grid during the inversion of a 3D mantle model. Plattner et al. [2012] adaptively refine a multi-scale wavelet parameterisation through optimisation in an electrical resistivity tomography inversion. Similarly, Hung et al. [2011] use an adaptive wavelet parameterisation to invert for P and S wave speed using finite frequency kernels [Dahlen et al., 2000]. In linearised tomographic problems, the "null space shuttle" [Deal and Nolet, 1996, de Wit et al., 2012] is an alternate approach to spatially adapting the inversion to adequately resolvable features.

1.3 Geophysical inversion choices

In any geophysical inverse problem, there are a number of choices to be made in the formulation of the problem. The first of these is the parameterisation in which to represent the model. This choice of parameterisation is often relatively arbitrary. Associated with any parameterisation choice is some further choice in terms of model complexity. In grid based parameterisations, this complexity or resolution is represented in the number and size of the grid cells. In basis function parameterisations such as Fourier series or spherical harmonics, there is a maximum degree to determine.

Parameterisation choices can have impacts on geophysical inversions [Trampert and Sneider, 1996, Valentine and Trampert, 2012, 2016, Lever et al., 2016], with all sharing the same inherent trade off between the resolving power of the inversion versus constraint of the model parameters. As the complexity or resolution of the parameterisation is increased, that is, the degrees of freedom are increased, the ability to constrain these model parameters decreases. This trade-off was recognised in a series of papers [Backus and Gilbert, 1968, Backus, 1970a,b], but the application of Backus-Gilbert inversion has found relatively little traction. Recent developments have seen advances from the helio-seismology community showing promise for large scale seismology inversions [Zaroli, 2016].

Coupled with the parameterisation choice, is some form of prior information used to constrain or stabilise the inversion. Geophysical inversions are almost universally under-determined problems and without the inclusion of prior information could produce non-physical model parameters (e.g. negative seismic velocities). Two common approaches for prior constraint in geophysical inversions are damping to a reference model, where again a reference model must be chosen, or imposing penalties on the spatial gradients of the model to impose smoothness.

A last choice is in the forward modelling used. There is unlikely to be an exact forward model for many inverse problems and decision involves selecting a reasonable approximation to the physical processes involved. There is often a spectrum of forward models, for example in seismic travel time tomography there is spectrum of forward models from fixed rays in a reference model through to full seismic waveform solvers.

Some optimisation techniques include regional adaptation and therefore effectively perform a model parameterisation choice. Regional adaptation schemes ask of the observations whether one parameterisation is “better” than another using some heuristic model choice criterion. A key question in an inversion is how to determine which of these choices are appropriate, or more likely representative of the truth, given the observations.

1.4 Bayesian inversion

The focus in this thesis is on a Bayesian approach [Bayes, 1763] to geophysical inverse problems and use Markov chain Monte Carlo (McMC) techniques [Gamerman and Lopes, 2006, Brooks et al., 2011] to generate an ensemble of candidate models. The Bayesian approach has a number of attractive features including a probabilistic ensemble from which can be extracted point estimates, such as the Bayesian maximum *a posteriori* (MAP), mean, mode and median models. Secondly, Bayesian credible intervals can be numerically estimated from the ensemble giving a robust estimate of uncertainties [Hyndman, 1996]. Third, in the Bayesian approach Bayes factors [Kass and Raftery, 1995] or an appropriate criterion such as the Deviance Information Criterion [Spiegelhalter et al., 2002] can be used to perform model comparisons between inversions with differing parameterisations or physics models [Steininger et al., 2014] and obtain objective measures on the model that is best supported by the observations.

In a Bayesian formulation of an inverse problem, rather than inverting for an optimal model, the posterior probability distribution is sought [Tarantola and Valette, 1982a,b], where the probability density represents knowledge of model parameters. This posterior probability is computed using Bayes theorem [Bayes, 1763],

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{m})p(\mathbf{d}|\mathbf{m})}{p(\mathbf{d})}, \tag{1.2}$$

where $\mathbf{m}$ is the model and $\mathbf{d}$ is the observations. The probability distribution $p(\mathbf{m})$ represents a prior probability distribution on model parameters. In a geophysical inverse problem, this is where information for constraining model parameters of under determined problems appears. The term $p(\mathbf{d}|\mathbf{m})$ is the likelihood which in a Bayesian formulation encapsulates the misfit between the model predictions computed by a forward model and the inherent noise processes. Lastly the term $p(\mathbf{d})$ is the "evidence", sometimes called the marginal likelihood, and is a normalising constant for the posterior probability distribution $p(\mathbf{m}|\mathbf{d})$.

The evidence can be obtained by directly integrated using

$$p(\mathbf{d}) = \int_M p(\mathbf{m})p(\mathbf{d}|\mathbf{m})d\mathbf{m}, \tag{1.3}$$

where the domain M is over the entire model parameter space. For simple problems with analytic solutions and few parameters, the evidence and posterior can be computed relatively easily. As the number of parameters increase and with increasing complexity or non-linearity in the forward model, this quickly becomes a difficult numerical computational problem.

In higher dimension and complex non-linear problems, sampling approximations are used to compute the posterior with methods such as Markov chain Monte Carlo (McMC) sampling [Liu, 2001, Brooks et al., 2011]. Often the posterior is then computed without the evidence, as relative inference is sufficient. Such sampling approaches have a long history in geophysical inverse problems [Wiggins, 1969, Mosegaard and

Tarantola, 1995, Sambridge and Mosegaard, 2002].

As stated before, any geophysical inversion that is under determined requires some form of additional information to stabilise the inversion. In a Bayesian approach, this is generally achieved in the prior and is a source for criticism of Bayesian approaches. A prior is another subjective choice in the formulation of the inversion, and often difficult to define well [Scales and Snieder, 1997].

In McMC inversion, from an arbitrary initial model, for example, $\mathbf{m}_i$, a probabilistically perturbation is added to the model to create a proposed model $\mathbf{m}'_i$. This perturbed model is accepted by assigning $\mathbf{m}_{i+1} = \mathbf{m}'_i$, or rejected leaving $\mathbf{m}_{i+1} = \mathbf{m}_i$ according to a criterion. The sequence of models forms a Markov chain which converges to the target posterior density. In this thesis the Metropolis-Hastings acceptance criteria [Metropolis et al., 1953, Hastings, 1970] is exclusively used where the new model is accepted with probability

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = \min \left\{ 1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m} \rightarrow \mathbf{m}')}{Q(\mathbf{m}' \rightarrow \mathbf{m})} \right\}, \tag{1.4}$$

where $p(\mathbf{m})$ is the prior, $p(\mathbf{d}|\mathbf{m})$ is the likelihood and $Q(\mathbf{m} \rightarrow \mathbf{m}')$ is the proposal probability. The Metropolis-Hastings criteria satisfies the mathematical condition known as "detailed balance" [Gamerman and Lopes, 2006] which allows the Markov chain to converge and correctly sample the target posterior distribution. Markov chains are run until they are deemed to have converged using some convergence criteria [Gelman and Rubin, 1992b, Brooks and Gelman, 1998]. Early portions of the Markov chain prior to convergence are typically discarded as "burnin" samples.

Given a Markov chain of models, this forms the result of the inversion from which representative models can be extracted, such as the mean or median of the ensemble. The uncertainty, is also derived from this Markov chain in terms of variances, credible intervals, and marginal histograms of model parameters or variables derived from selected model parameters. The draw back with McMC methods is that the length of the Markov chain required for sufficient statistical inference could be of the order of millions which can become a computational burden for large number of data or complex forward modelling.

A common approach to obtaining more samples is to run multiple independent Markov chains but this has potential to overestimate uncertainty if used incorrectly. A common pitfall is to use multiple short chains run in parallel to save time, however these chains are unlikely to be converged and although the ensemble means of these model will likely be acceptable (as the mean converges quickly), the variances will be large due to lack of convergence.

1.5 Bayesian Model Selection

The Bayesian approach to model selection is Bayes Factors [Kass and Raftery, 1995] which require the calculation of the evidence. The difficulty of computing the evidence is a function of both the model and the forward model complexity. As the dimension of the model increases, so does the dimension of the integral required for computing the evidence. As the forward model complexity increases, analytical solutions to integrals may not be available or approximate numerical integration or sampling may take more computational effort. For all but the simplest problems, a numerical approximation is required and while methods are available [Skilling, 2006], evidence calculation remains a difficult problem. Reliably computing evidence in higher dimensional problems is an active area of research and newer methods show promise in geophysical inverse problems [Brunetti et al., 2017]

Once evidence has been computed for competing model solutions of a given problem, the Bayes factors are simply the evidence ratios of the two candidate models. Put simply, the model with higher evidence has more support and there are criteria for the level of support based on the magnitude of evidence ratios.

1.6 Approximate Bayesian Model Selection

Due to the difficulties and computation cost in the reliable calculation of the evidence, various approximations have been developed [Akaike, 1974, Spiegelhalter et al., 2002, Ando, 2007, 2010]. At their core, these approximations assume a simple shape of the posterior and result in a criterion score as a function of the maximum likelihood solution and the model complexity. In this way, better fitting models are rewarded and overly complex models are penalised in an attempt to prefer models that are simple enough to explain the observations, but no simpler. The draw back of these approximating criteria are that they may not be valid for many non-linear geophysical inverse problems due to their simplifying approximations.

1.7 Relative model selection via Reversible Jump

An extension to McMC samplers is the Birth/Death scheme of Geyer and Møller [1994], generalised to the Reversible Jump by the seminal work of Green [1995]. Reversible jump McMC has become commonly known as trans-dimensional sampling in the geophysics community, through works of Malinverno [2002] and Sambridge et al. [2006].

In trans-dimensional samplers, a proposal distribution is allowed to change the parameterisation of the model and dimension, that is the size of the vector $\mathbf{m}$ of model parameters. A key benefit of allowing the sampling to jump between dimensions is that the data dictates the model complexity resulting in a parsimonious result [Malinverno, 2002]. A trans-dimensional result was shown to be equivalent to the product space of fixed dimension solutions weighted by their evidences, that is, the trans-dimensional result is effectively Bayesian relative model selection [Sambridge et al., 2006].

Hence the trans-dimensional approach can be used for Bayesian model selection without the need to compute evidence directly. Unlike approximating criteria, there are no assumptions on the shape of posterior distribution and non-linearity in the forward is no longer a problem. When compared to computing the evidence directly, a draw back of trans-dimensional sampling is that, since only relative evidence is computed, all competing models must be considered in a single inversion. If another competing model is to be considered at a later date, then this must be incorporated into a new trans-dimensional frame work and the entire inversion rerun. In contrast, the evidence is a fixed quantity for a model parameterisation and can be compared without recalculation at a later date to any other competing model.

The generalisation of the Metropolis-Hastings acceptance criteria to support trans-dimensional steps is

$$\alpha(\mathbf{m} \to \mathbf{m}') = \min\left\{1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m} \to \mathbf{m}')}{Q(\mathbf{m}' \to \mathbf{m})} \, |\mathscr{J}| \right\}, \tag{1.5}$$

where the additional term from (1.4), $|\mathscr{J}|$, is determinant of the Jacobian that maintains detailed balance through variable transformations resulting from trans-dimensional steps. Expression (1.5) may also be used if the dimension is unchanged, but the proposal involves a step from one class of parameterisation to another.

In trans-dimensional McMC inversion, in a similar fashion in which McMC converges to the posterior distribution of interest, trans-dimensional McMC includes proposals that jump between different models. The McMC chain then samples in each model space and converges to give the relative support of the candidate models.

In the original work by Green [1995] with useful reviews by Denison et al. [2002] and Sisson [2005], a common usage of trans-dimensional algorithms is partition modelling. In partition modelling, a domain is partitioned into a set of k contiguous cells used to represent some function. Figure 1.2 shows examples of this in 1D where in (a), a step-wise function defined by a single value in each partition. This form of partition modelling has been applied in geophysical inverse problems for representing 1D velocity models in receiver function studies [Piana Agostinetti and Malinverno, 2010], surface wave dispersion studies [Dettmer et al., 2012], and both [Bodin et al., 2012b]. In (b) the same partitions or cells can be used to construct a piece-wise linear curves by

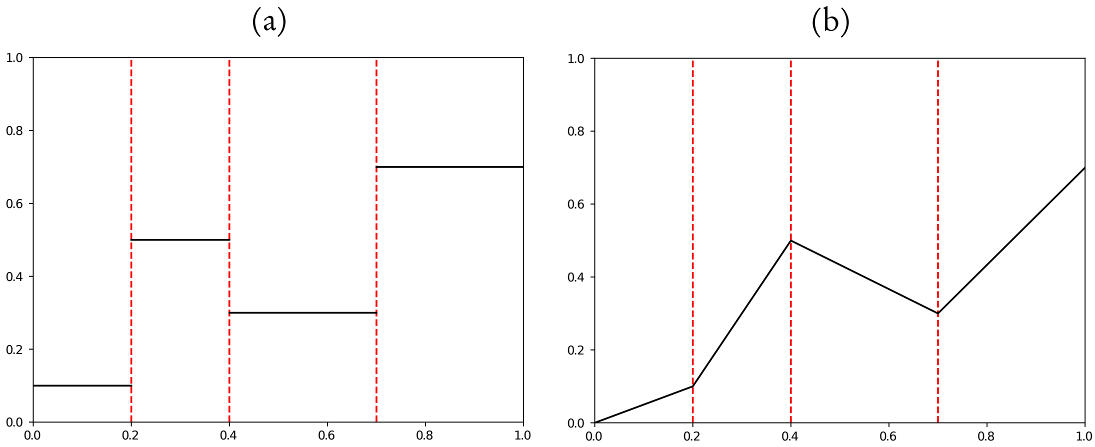


Figure 1.2: Example parameterisations for partition modelling where in (a) a partition with constant values in each partition is shown that the results in a step function. In (b) the natural partitioning scheme is shown with linear segments between partition boundaries which can be used to represent continuous functions.

specifying values at the partition boundaries and this has been in various paleo-climate reconstructions [Hopcroft et al., 2007, Lambeck et al., 2014, Sambridge, 2016].

In partition modelling, a key aspect of the problem is deciding on the number of partitions because this is a measure of resolution. In trans-dimensional partition modelling, the method adds and removes partitions automatically to sample about a number of partitions that is well constrained by the data for a parsimonious result.

Extension of partition modelling to two and higher dimensions have typically involved the use of Voronoi cells [Okabe et al., 1992, Samet, 2006]. Examples in geophysical inverse problems include 2D tomography [Bodin and Sambridge, 2009, Bodin et al., 2009, 2012a, Galetti et al., 2015, Saygin et al., 2016], finite fault inversion [Dettmer et al., 2014], core-mantle boundary tomography [Young et al., 2013] and 3D local earthquake tomography [Piana Agostinetti et al., 2015]. While use of Voronoi cells in trans-dimensional geophysical inversion is now well established, what should be recognised is that the two classes of partition modelling in 1D shown in Figure 1.2 extend to two and higher dimensions as Voronoi cells and Delaunay tessellations, examples of each are shown in Figure 1.3.

The complexity of the models generated from trans-dimensional samplers is sensitive

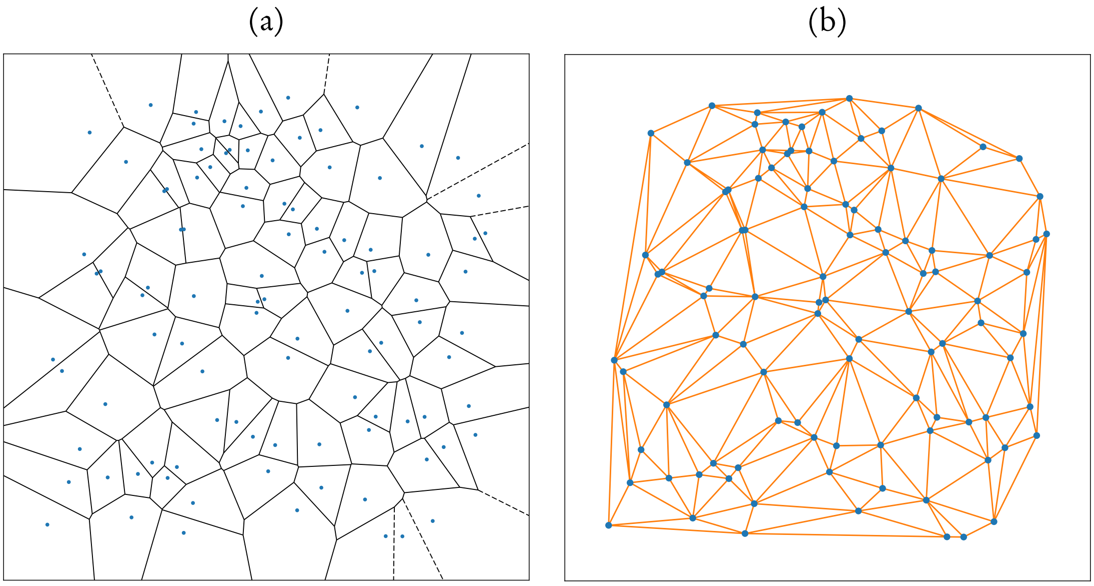


Figure 1.3: Example parameterisations for partition modelling in 2D where in (a) a partition with constant values in each partition or a Voronoi cell approach is shown that can be used for functions with discontinuities. In (b) is shown triangle patches between vertices called a Delaunay triangulation that can be used to represent C^0 continuous functions.

to the estimated noise on the data. If this noise level, encoded within the likelihood function, is under estimated then a trans-dimensional inversion will return overly complex models rather than a parsimonious appraisal. This is due to under estimated noise levels requiring a trans-dimensional solution to fit noise. Conversely, if noise levels are over estimated, then the resulting complexity will be too simple. For this reason, in the case where the noise estimates on observations are poor, or approximations in the formulation of problem introduce theory noise, it is advantageous to use a hierarchical Bayesian inversion [Malinverno and Briggs, 2004] coupled with the standard trans-dimensional sampling. Hierarchical noise estimation allows additional noise parameters to be inverted for as part of the sampling the model parameters of interest [Bodin et al., 2012a].

In trans-dimensional inversion, the development of convergence criteria is not as mature as in the case of more traditional fixed dimension McMC sampling [Sisson and Fan, 2007]. The philosophy taken in this thesis is a pragmatic choice where chains are initialised with simple models, for example, in tomographic examples McMC chains are initialised with homogeneous models. From here the trans-dimensional inversion must iteratively add and remove complexity until convergence is reached. This is deemed preferable to randomly initialising models of varying complexity as the time taken for complex models to converge to simpler can be prohibitive. While starting from simple models and increasing complexity is generally quicker to converge. An additional observation is that some unknown failing in the inversion process will produce results that err on the side of simplicity.

The convergence is judged on two main criteria, firstly and most importantly is the hierarchical noise level which is a proxy for the likelihood as it represents a scale factor or factors of the estimated errors in the data. In the absence of hierarchical noise levels, the likelihood itself would suffice. Secondly, a measure of complexity is monitored, for example the number of partitions. The Gelman-Rubin convergence tests are also considered on these parameters although these tend to converge quite quickly and the qualitative convergence metrics mentioned here are conservative by comparison [Hawkins et al., 2017].

1.8 Thesis Outline

This thesis presents a number of advances in trans-dimensional Bayesian methods for geophysical inversion problems of varying physics and character.

In Chapter 2, an approach to extracting phase velocity information from seismic ambient noise observations is presented using a trans-dimensional partition modelling approach. This method is able to extract Love and Rayleigh wave dispersion jointly from three component data as continuous dispersion curves with uncertainties that can be carried forward in subsequent phase velocity map inversions.

In Chapter 3, the development of a new type of trans-dimensional algorithm called trans-dimensional trees is presented. This scheme is targeted towards higher dimensional geophysical inversion problems where existing Voronoi cell methods have computational deficiencies. This is an abstract trans-dimensional approach with many potential applications. This chapter demonstrates the flexibility of trans-dimensional trees by coupling this with a wavelet parameterisation and compares the results to existing Voronoi cell approaches in simple linearised tomographic problems.

In Chapter 4, the trans-dimensional tree approach is applied to the inversion of Airborne Electromagnetic data along a section of flight line for a 2D profile of resistivity. This approach uses the trans-dimensional tree with a wavelet parameterisation coupled with a 1D forward model to invert a spatially coherent model, that is, with no lateral discontinuities due to the independent 1D inversions. Results show more detail is resolved than an existing inversion using a damped and smoothed least squares optimisation approach. This chapter also introduces hierarchical priors and the generation of estimated covariant noise models to further improve and stabilise the inversion.

In Chapter 5, the results of the dispersion curves generated in Chapter 2 are inverted for Love and Rayleigh phase velocity maps of Iceland. The Fast Marching Method is used "tightly coupled" to the trans-dimensional tree approach with a wavelet parameterisation to invert in a fully non-linear fashion. The difference between non-linear and linear (with fixed ray paths) inversions is demonstrated in synthetic examples. This comparison highlights the generally poorer recovery of features and underestimation of anomaly magnitudes of the linear inversions compared to the fully non-linear approach developed in this chapter. Taking this class of inversion further, all periods of interest are inverted jointly in a 3D fully non-linear inversion to take advantage of the correlation of spatial features between neighbouring frequencies.

In Chapter 6, the trans-dimensional tree is applied to a large problem in the spherical domain, through the inversion of global surface wave data. In this problem, the total number of observations is of the order of 5 million ray paths. This preliminary study test the feasibility of inverting these observations in a reasonable length of time. A key benefit of the trans-dimensional tree approach is that the inversion is stabilised through relative Bayesian model choice rather than smoothing or damping constraints, both of which cause the magnitude of fast and slow anomalies to be underestimated and examples are provide of this in the African rift region.

In Chapter 7, a simple general trans-dimensional partition modelling scheme for 1D problems is introduced that is able to adapt to data best explained with discontinuous features or smoothly varying features, and combinations thereof. The parameterisation used in this approach is polynomials expressed as Gauss-Legendre-Lobatto polynomials where the curve(s) are defined by nodal interpolation points. This ensures that the prior is more intuitive compared to polynomial coefficient priors. It is shown that allowing an inversion to consider more complex combinations of polynomials in a partition modelling scheme produces better results in synthetic regression problems.

In Chapter 8, a common and difficult problem in seismology is that many problems have non-unique solutions and a classic one is whether observations are best explained by a 1D Earth model with a series of homogeneous layers, or smoothly varying structure. To attempt to address this, some synthetic surface wave dispersion problems are considered. A novel spectral element method is derived for computing surface wave dispersion predictions from arbitrary models expressed as a series of elements with arbitrary order. A novelty in this approach is the inclusion of a Laguerre element for representing a half-space that dramatically improves accuracy at longer periods. This method is first validated against known analytic results and existing approaches. Finally in a series of synthetic tests coupling the trans-dimensional approach of Chapter 7 with the spectral element surface wave dispersion forward model, it is examined whether decisive posterior information can be obtained in the inversion of simple structures with slowly varying and homogeneous layers in Love wave, Rayleigh wave and joint inversions.

Chapter 2

Phase velocity determination

2.1 Introduction

This chapter provides a simpler introduction to some of the key concepts relating to Bayesian inversion, Markov chain Monte Carlo (McMC) and trans-dimensional sampling that will be used throughout this thesis. Here the problem is one dimensional and involves inverting for dispersion curves from empirical Greens functions, a key component in ambient noise tomography.

Ambient noise tomography is a relatively new technique for surface wave tomography that uses the ambient seismic wave field, excited by ocean swells, storms and wind, to image relatively near surface structure (up to approximately 100km). The reason for the relatively shallow limit is that ambient noise excitation frequencies are generally limited to between 1 and 30 second period, with a dominant spectrum around 7 to 16 seconds [Bensen et al., 2007, Figure 7(a)]. This limit of useful frequencies translates to a similar limit on resolvable depths.

Early work by Aki [1957] on micro-tremors established much of the more recent theoretical work showing the plausibility of recovering the elastic Green's function between two recording stations through the cross-correlation of the ambient seismic noise or coda [Lobkis and Weaver, 2001, Derode et al., 2003, Snieder, 2004, Wapenaar, 2004, Larose et al., 2005] (see also Larose et al. [2006] for a review article). The Greens function between stations A and B represents the signal observed at station B of an impulse at A and vice-versa. These virtual seismic events between station pairs enables seismic tomography techniques to be applied in the absence of Earthquakes. Early applications of this approach using observed seismic coda were reported by Campillo and Paul [2003], Paul et al. [2005], and similarly for ambient noise Shapiro and Campillo [2004], Sabra et al. [2005].

There are three preliminary stages for ambient noise tomography, as outlined by Bensen et al. [2007], one of the first papers to make key recommendations for these steps. These steps are, (i) pre-processing the continuously recorded seismograms, (ii) cross correlation of seismograms from two stations to obtain inter-station empirical

Greens functions (EGFs), and (iii) extracting path integrated surface wave dispersion information for two stations using these EGFs.

A key recommendation is that continuously recorded seismograms be pre-processed prior to cross correlation. The purpose of this pre-processing is three fold, firstly to remove instrument response from the seismograms that may amplify/suppress measurement of ambient noise, secondly time domain normalisation to remove the effects of local or global seismicity from corrupting latter cross correlation and lastly spectral normalisation or whitening to raise the signal to noise ratio of all frequencies of interest.

For focus of this chapter is the estimation of dispersion based upon the Green's functions recovered from cross-correlations of noise, a crucial component of ambient noise tomography. An overview of the pre-processing steps required is also given as they have important consequences for subsequent processing.

2.2 Pre-processing of Ambient Noise

An early effort by Bensen et al. [2007] sought to establish a standard for the processing of ambient noise data and performing cross correlations. The summary of the steps, in phase one of Bensen et al. [2007] are

1. remove instrument response,
2. remove mean,
3. remove trend,
4. band pass filter,
5. apply time domain normalisation or 1 bit normalisation, and
6. apply spectral whitening.

The first 4 pre-processing steps are fairly standard processing of seismic signals. In fact they can together be considered a single band pass filter as the trend and mean can be considered low frequency components of the signal.

Much of the rationale for the subsequent processing is the removal for spurious signals (from local and global seismicity) and an attempt to amplify signal for the subsequent stacking process. In a Bayesian formulation, it can be argued that this is unnecessary and to some degree counter-productive as the time domain normalisation step distorts phase information in the signal and spectral whitening amplifies spurious signal that can impact uncertainties.

To demonstrate the distortion of phase information, a simple experiment is to generate a random signal with a known phase angle as a linear function of frequency, and then apply the recommended processing to discern the effects on the phase. The raw signal and raw phase are show in Figure 2.1 (a) and (b) respectively. The remaining plots show the effect of various pre-processing steps on the signal and phase information. Low pass filtering, shown in (c) and (d), preserves the phase information up to the cut-off frequency (1Hz) of the filter. It should be noted that the low pass filtering is implemented with a linear filter applied twice, once forward and once backward. This combined filter results in linear phase and as can be seen.

In the work of Bensen et al. [2007], three processing techniques are recommended and these are shown in (e) and (f) for one bit normalisation, (g) and (h) for time domain weighted mean normalisation, and (i) and (j) for spectral whitening. As can be seen in these cases, the phase angle information is changed, and in some cases, for example in the one bit normalisation, quite dramatically.

The purpose of the time domain normalisation processes, that is, the one bit normalisation or the time domain weighted mean, is to mask out the effect of seismicity when cross-correlating the signals. The effect that seismicity has on ambient noise cross-correlations is to create a peak at zero time in the empirical Greens functions. This can be problematic for close proximity stations, but for reasonably spaced arrays it is less of an issue.

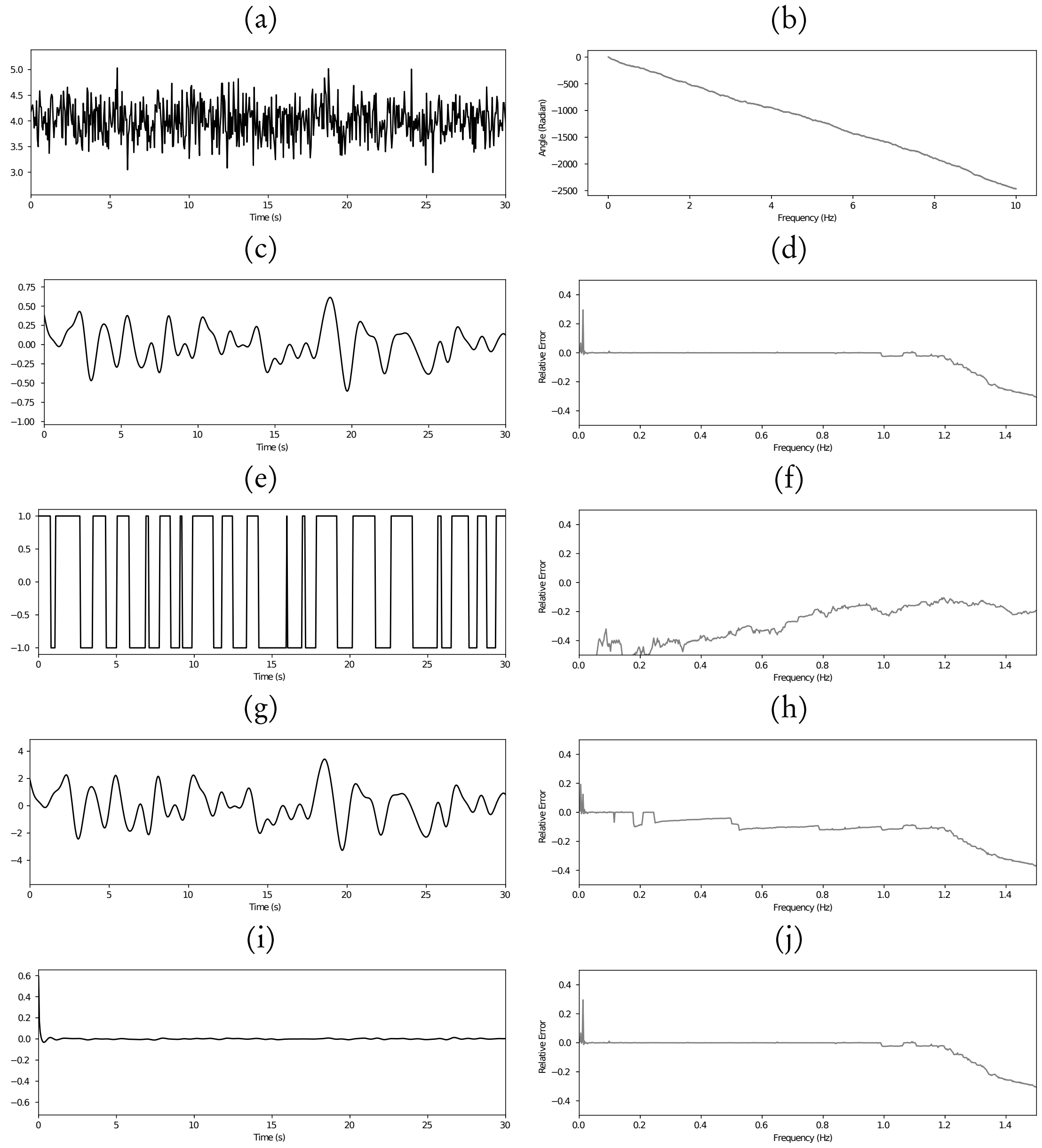


Figure 2.1: The pre-processing can affect the phase information contained in a noise signal. The left column shows the signal and the right column shows the relative difference between the phase angle and the original signal with the exception that in the first row (b) shows the original phase angle. In (c), the de-trended, demeaned and rescaled signal still faithfully preserves phase up to the frequency limit (1 Hz) of the filter in (d). One bit normalisation of the signal in (e) produces large differences in the phase (f). Similarly, time domain averaging, (g) and (h), affects the phase angle. Spectral whitening (i) and (j) though preserves the phase angle over the frequency band of interest.

In a Bayesian approach, where noise is an integral part of the inversion, minimal processing is preferred for realistic uncertainty estimates. Particular to this problem is the potential for additional processing, for example, one bit normalisation, to distort phase information for subsequent cross-correlation. Using a combination of the processing recommendations of Ekström [2014] and Seats et al. [2012], the steps performed in this study are

1. organise seismograms into day lengths with 20Hz sampling,
2. remove instrument response and band pass filter between 10mHz and 1Hz using linear phase filters,
3. down sample seismograms to 2Hz and normalise so they are zero mean and have a standard deviation of one,
4. cross-correlate inter-station seismograms using overlapping windows [Seats et al., 2012] of sufficient length assuming a minimum velocity of 1km/s and the maximum distance between between two stations in the array to set the time duration of the window, and
5. stack individual windows create the inter-station ambient noise correlation function.

The benefit of this minimalist processing is that it preserves phase information through the careful use of phase preserving filtering and a simple normalisation of seismograms across days. Both of these processing steps have minimal impact on the phase within the signal.

The processing steps outlined above were applied to three component data from stations in Iceland using data from the HOTSPOT [Allen et al., 1999] deployment and one IUGG station (BORG). These set of stations is the same as those used in the study by Gudmundsson et al. [2007]. An example of the cross correlations obtained from this sequence of steps is shown in Figure 2.2 where in (a) the time domain empirical Greens function is shown and in (b) the real and imaginary parts of the spectrum are shown. There is clearly asymmetry in the time domain signal, an indication that the noise sources are not spatially isotropic. Another indication of this is that in the frequency domain, the imaginary component of the spectrum is not zero [Aki, 1957, Cox, 1973].

Unlike Bensen et al. [2007], spectral whitening is not performed because doing so biases the uncertainty as a function of frequency. In a Bayesian approach, observational noise is a key component of the inversion and to arbitrarily normalise frequencies that may be poorly excited by the ambient noise field is seen as undesirable. Rather, it is preferable for the uncertainty in the ambient noise observations to propagate throughout the inversion, both of surface wave dispersion and of subsequent use of the dispersion information in following tomographic inversions.

2.3 Obtaining dispersion information

In Bensen et al. [2007], some time is spent discussing the extraction of phase velocity, where they suggest that there was no known suitable method. For this reason, the authors suggest extraction of the group velocity using frequency-time analysis techniques [Dziewonski and Hales, 1972, Herrmann, 2013]. This in turn became the method of choice for many ambient noise studies because there are established codes and it is relatively straight forward albeit labour intensive.

The approach taken here estimates phase velocity dispersion and this is motivated by the fact that phase velocity is more useful than group velocity. Firstly, the group velocity can be uniquely determined from phase velocity dispersion, since phase velocity is given by

$$c_n(\omega) = \frac{k_n(\omega)}{\omega}, \tag{2.1}$$

and group velocity by

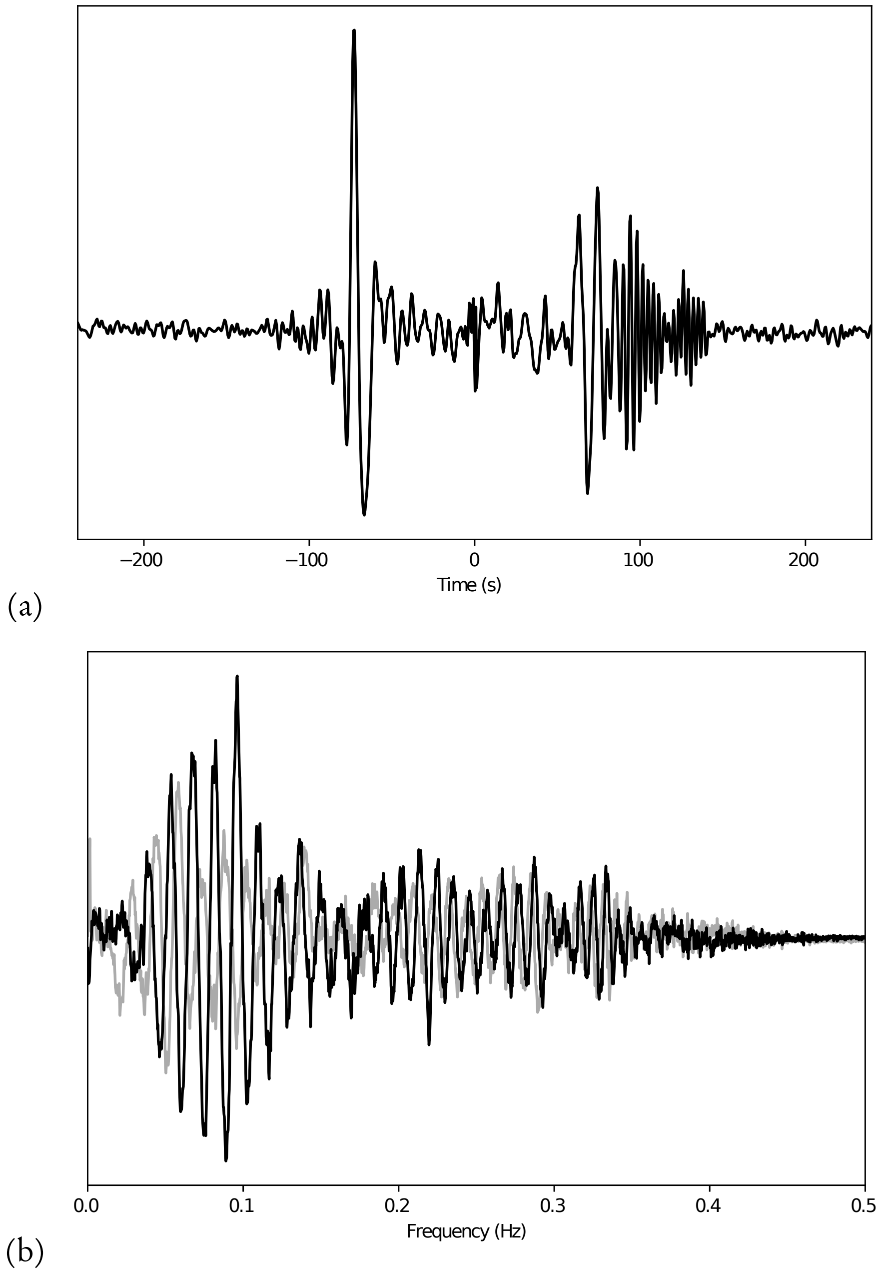


Figure 2.2: An example empirical Greens function is shown resulting from the minimalistic pre-processing approach before cross-correlating signals. In (a) is the empirical Green's functions in the time domain highlighting the causal and acausal parts of the signal. In (b) is the complex spectrum with the real part shown as a dark line to highlight the spectral zero crossings, and the imaginary part shown as a feint line. It is clear from both the time domain and frequency domain that the noise is not spatial isotropic.

$$U_n(\omega) = \frac{\partial \omega}{\partial k_n(\omega)}, \tag{2.2}$$

where c_n is the phase velocity of mode n as a function of frequency ω, k_n is the wave number, and U_n is the group velocity. From these relationships, an expression can be obtained for the group velocity in terms of phase velocity

$$U_n(\omega) = \frac{c_n(\omega)}{1 - \frac{\omega}{c_n(\omega)} \frac{\partial c_n}{\partial \omega}}, \tag{2.3}$$

hence given a differentiable phase velocity curve, the group velocity can be computed. The reverse is not true because expression of c_n as a function of U_n becomes

$$c_n(\omega) = \int_{\omega_0}^{\omega} U_n(x) dx + c_n(\omega_0), \tag{2.4}$$

and the $c_n(\omega_0)$ is not known. Hence any extraction of phase velocity automatically means group velocity is also available assuming differentiability, but the reverse is not true.

Secondly, surface wave ray paths are sensitive to phase velocity, not group velocity [Tanimoto, 1986]. In a non-linear inversion in which ray paths are recomputed, they should be recomputed based upon the phase velocity and not the group velocity. Many previous studies [Saygin et al., 2016, Galetti et al., 2016] have computed ray paths from group velocities which introduces a further approximation. For slowly varying, or nearly linear, dispersion of phase velocity, using the group velocity is a reasonable approximation because the the group velocity dispersion will be a scaled version of the phase velocity curve to first order. Hence the relative change in group velocity and phase velocity coincide and the ray paths generated by each would be similar. Unfortunately, phase velocity dispersion for simple models and exemplar measured dispersion shows that there is a degree of gradient change in the phase velocity in frequency ranges of interest, hence this approximation may cause significant inaccuracies. In Figure 2.3, a reasonable but simple dispersion is shown illustrating the difference between phase

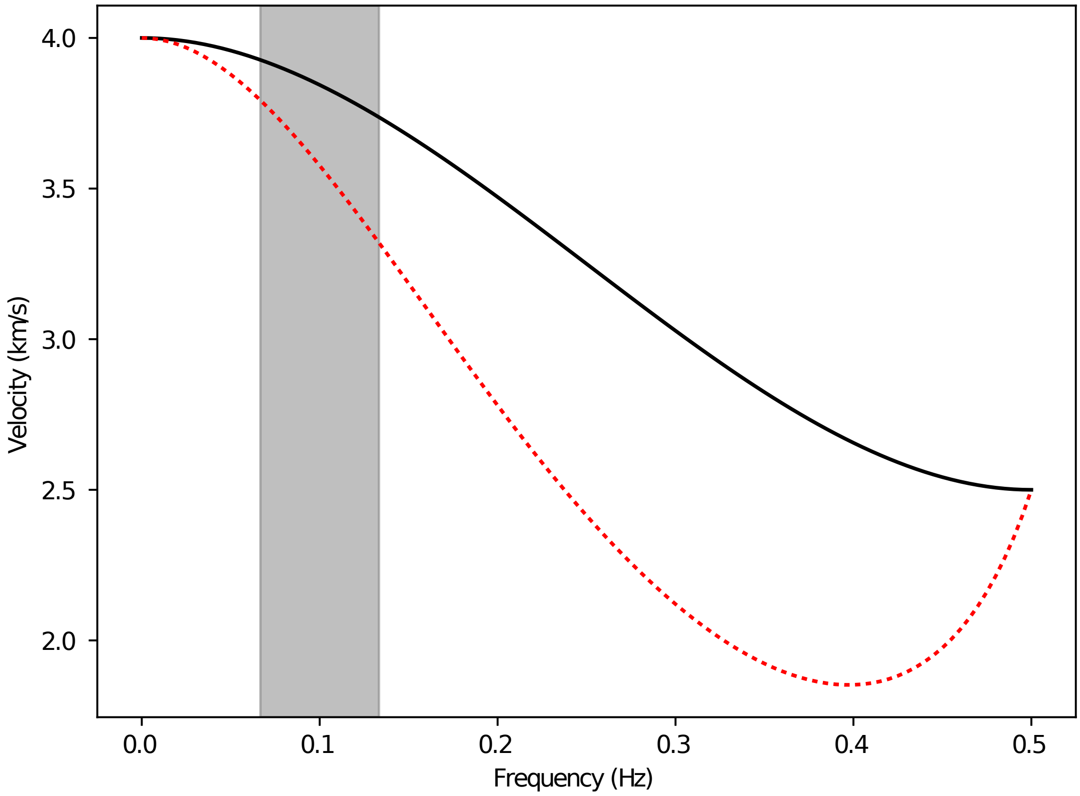


Figure 2.3: An example dispersion curve showing difference between phase (solid black line) and group velocity (red dotted line). Updating rays using group velocity can bias the results because relative change in phase velocity and group velocity can change in the frequency range of interest for ambient noise tomography (grey shaded region).

and group velocity.

Others have similarly expressed the benefits of phase velocity over group, for example, Boschi et al. [2013] provide three motivations for phase velocity over group:

1. Group velocity is less precisely defined than phase velocity,
2. For the fundamental mode, phase velocity information is able to image deeper into the Earth than group velocity,
3. Group velocity measurements are more likely to be contaminated by interfering phases than phase velocity measurements.

Motivated by the benefits of phase velocity over group velocity dispersion information, more recently, different methods have been proposed for the extraction of phase velocity dispersion from empirical Green's functions. Firstly an image based time-domain technique [Yao et al., 2005, 2006], and a frequency domain technique [Ekström et al., 2009, Ekström, 2014] based on an analysis of the statistics of micro-tremor correlations between stations [Aki, 1957]. Both these techniques have been successfully applied to the recovery of phase velocity maps of regions from ambient noise correlations.

To extract phase velocity, a frequency domain method is used, originally proposed by Aki for micro-tremor data [Aki, 1957] and reintroduced for ambient noise by Ekström [Ekström et al., 2009, Ekström, 2014]. Alternatives, such as image based methods exist for extracting phase velocity [Yao et al., 2005], however the frequency domain method does not use the far field approximation and therefore can extract longer period dispersion information from closer stations.

In the original Ekström [Ekström et al., 2009] approach, the zero crossings of the real component of the spectrum of the cross-correlograms were used to construct trial phase velocity curves directly. This uses the result from Aki [1957] that the real component of correlated noise between stations is of the form of a Bessel function of the first kind with order 0. Restating this result here

$$\bar{\rho}(\omega_0, r) = J_0\left(\frac{\omega_0}{c(\omega_0)} r\right), \tag{2.5}$$

where $\bar{\rho}$ is the cross correlation spectrum, ω_0 the angular frequency of the fundamental mode, $c(\omega_0)$ is the frequency dependent phase velocity, and r is inter-station distance. From the observed zero crossings in the empirical Greens function, i.e. a set of zero crossings $z_1 \ldots z_n$ at angular frequencies $\omega_1 \ldots \omega_n$, trial phase velocity curves are constructed using

$$c_m(\omega_n) = \frac{\omega_n r}{z_{n+2m}}, \tag{2.6}$$

with m the integer trial value that is 0, ±1, ±2, etc. From these trial curves, the dispersion curve mostly closely matching a reference dispersion curve for the region is chosen as the observed phase velocity curve. The problem with this approach is that noise inherent in the observations can cause spurious zero crossings as highlighted by Menke and Jin [2015] and this can result in phase velocity dispersion curves that drop precipitously to unfeasible values. These problems were recognised and a subsequent improvement to the method [Ekström, 2014] adds the extra step of fitting a piece wise spline to the real component of the spectrum in an effort to eliminate the spurious zero crossings. A further extension of this general approach was the inclusion of completely fitting the real part of the spectrum by Menke and Jin [2015], which improved the rejection of spurious zero crossings and added the ability to using residuals from the inversion of individual cross-correlated station pairs as quality factors.

Here the aim is to build on these advances by developing a Bayesian approach for extraction of phase velocity information. A key factor in Bayesian approaches is the inclusion of prior information in both the formulation of the problem, the assumptions therein as probability distributions.

2.4 A Bayesian Trans-dimensional Partition modelling approach

In a Bayesian approach, the solution is obtained in a probabilistic sense be inverting for a set of models plausibly could have generated the observations. The plausibility is decided in a probabilistic sense as a combination of a prior probability distribution and a likelihood. This is related to the posterior inference using Bayes theorem [Bayes, 1763]

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{m})p(\mathbf{d}|\mathbf{m}), \tag{2.7}$$

where $\mathbf{m}$ is the model, $\mathbf{d}$ is the observations or data, $p(\mathbf{m}|\mathbf{d})$ the posterior, $p(\mathbf{m})$ the prior and $p(\mathbf{d}|\mathbf{m})$ the likelihood.

The parameterisation to be used in this inversion is an advance on the standard partition modelling approach introduced in Figure 1.2(b). In a general partition modelling approach, the model consists of k, a number of partitions, a k length vector $\mathbf{c}$ of the partition widths and at each partition boundary a phase velocity. The partitions divide up a range of frequencies and from this model. From the partition model, a continuous function of phase velocity versus frequency can be constructed represent the dispersion curve.

Through the action of a forward model, a set of predictions can be computed from the dispersion curve to compute the likelihood. These are the key components of a Bayesian inversion and the following sections describe each, and the reasoning behind choices therein in further detail.

2.4.1 Prior information

In a Bayesian approach, prior information expressed as a valid probability distribution is required to inform the subsequent inversion. This is at times difficult and one of the common criticisms of Bayesian methods, namely, the dependence of the inversion on a prior. This prior information can be obtained from a variety of sources, in order of strength: previous inversions of model parameters of interest, laboratory or experimental measurement, theoretical knowledge or numerical experiments, and lastly assumptions.

Previous inversions of model parameters of interest, even obtained through non-Bayesian means should have uncertainty estimates which can be used as priors for subsequent Bayesian inversion. The benefit here is that it becomes easy to quantify what has been learnt from the inversion of new observations compared to the previous study that provided prior information.

Laboratory measurements of rock samples, including those at temperature and pressure, can be used as guidance for parameters such as density, transmission velocities and attenuation. Theoretical underpinnings can be used both in parameterisation of the problem and in expressing priors. Prior knowledge such as velocity and density must be positive are simple examples. Similarly, numerical simulation can be used to determine a range of plausible values for a model parameters of interest, given prior knowledge of the variables that influence those parameters.

Lastly, when no prior knowledge is available, the prior can be expressed as an assumption in the inversion. While less than ideal, in a Bayesian approach, such assumptions are quantified explicitly in the inversion and can be later challenged and validated with posterior inference. In later chapters, methods of hierarchical sampling to enable more rigour in parameters with ill-defined or problematic priors will be presented.

In the previous sections, an outline of the process to obtain empirical Greens functions between station pairs was given. These Greens functions, either for Love or Rayleigh waves, represent the signal that would be observed at one station given an appropriate impulse excitation at the other. In almost all ambient noise studies dealing with surface waves, only the fundamental mode is of high enough energy to be detected.

To use these empirical Greens functions to invert for local structure using ray theory, the following assumptions are made. Firstly, that lateral variations are gradual so that no interface or caustic effects such as mode conversions or scattering occur. Secondly, that the medium is elastic and therefore attenuation free. Lastly, it is assumed that the phase velocity dispersion is normal and a smooth function of frequency.

Normal dispersion

The assumption that dispersion is normal and a continuous smooth function of frequency allows the use of a linearisation approximation so that the group velocity, U, can be related to the angular frequency ω and wave number k_n for a given mode n by

$$U_n(\omega) = \frac{d\omega}{dk_n(\omega)} = \left[\frac{dk_n(\omega)}{d\omega}\right]^{-1}. \tag{2.8}$$

For the dispersion to be normal, the group velocity must be less than or equal to the phase velocity for all frequencies. The phase velocity is given by

$$c_n(\omega) = \frac{\omega}{k_n(\omega)}. \tag{2.9}$$

Using these two equations, the conditions for normal dispersion are

$$c_n(\omega) \geq U_n(\omega) \tag{2.10}$$

$$\frac{\omega}{k_n(\omega)} \geq \left[\frac{\partial k_n(\omega)}{\partial \omega}\right]^{-1} \tag{2.11}$$

$$k_n(\omega) - \omega \frac{\partial k_n(\omega)}{\partial \omega} \leq 0 \tag{2.12}$$

$$k_n(\omega)^2 \left[\frac{\partial c_n}{\partial \omega}\right] \leq 0, \tag{2.13}$$

and since $k_n(\omega)^2$ is strictly a positive function for $\omega > 0$, giving

$$\frac{\partial c_n(\omega)}{\partial \omega} \leq 0. \tag{2.14}$$

From these relations, the phase velocity dispersion as a function of frequency must be a $\mathscr{C}^1$ continuous, monotonically non-increasing function of frequency, in order for normal dispersion assumption to be maintained.

2.4.2 Partition modelling for dispersion

Spatial priors in 1D

The naive approach to specifying a prior for control points or interpolation points in 1D would be to assume the points are independently generated from a uniform prior across the range. For example, for k points generated between some bounds $x_{\min}$ and $x_{\max}$, a uniform prior would be

$$p(\mathbf{x}|x_{\min}, x_{\max}, k) = k!\,(x_{\max} - x_{\min})^{-k}, \tag{2.15}$$

where $\mathbf{x}$ is the k length vector of x positions [Denison et al., 2002]. The $k!$ factorial term at the front is required since the order of generating the x positions is not important.

Another alternative prior, often labelled the "grid trick" [Denison et al., 2002] is to assume there is a fine grid of N fixed points from which to choose from. In this case the prior becomes

$$p(\mathbf{x}|N,k) = \left[\binom{N}{k}\right]^{-1}. \tag{2.16}$$

It so happens that in the acceptance criteria, the use of this prior coupled with compatible proposals results in the complete cancellation of the binomial combinatorial terms and unspecified value of N, resulting in simple acceptance criteria, hence the popularity of this approach.

A common problem with both the uniform and grid trick priors is that they have no preference for the distribution of spatial points. This can result in two points very close to one another which in regression problems means that the function values at this points are unconstrained, or in forward model problems, may cause instabilities in the forward model. In trans-dimensional sampling, these closely spaced points are generally removed but the fact that they can occur detracts from sampling efficiency. Given prior knowledge of the problem, that is, if prior knowledge suggests the curve is slowly varying, then closely spaced points are unlikely to represent this curve and the prior should reflect this.

In Green [1995], Section 4.1, given k points, the prior is specified in a similar manner to the grid trick where the probability of selecting k points from $2k+1$ points uniformly distributed in the domain of interest using even-numbered order statistics which has the property of probabilistically selecting models with evenly spaced points. This however remains a discrete approximation to setting of a prior for continuous variables and has no way of controlling the degree of "evenness".

A common draw back of these methods is their inability to account for varying the range of the domain. A solution was given by Steininger et al. [2013] where a 1D Earth model was parameterised as a set of layers between the surface and a variable basement. To express this in terms of a probabilistic prior, they used a symmetric Dirichlet distribution. The Dirichlet distribution in 1D partitions the unit interval into k variably spaced partitions. If each of the partition cell widths are x_i and each cell has a weight α_i, then the Dirichlet distribution is

$$p(\mathbf{x}|\alpha_1 \ldots \alpha_k, k) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} x_i^{\alpha_i - 1}, \tag{2.17}$$

with Γ being the gamma function. A symmetric Dirichlet distribution has all α weights equal in which case the distribution simplifies to

$$p(\mathbf{x}|\alpha, k) = \frac{\Gamma(k\alpha)}{\Gamma(\alpha)^k} \prod_{i=1}^{k} x_i^{\alpha - 1}. \tag{2.18}$$

Through simple variable transformation, this prior can be expressed in terms of the partition locations rather than cell sizes and also apply this prior to different scaled regions. For partition locations $\mathbf{y}$ between y_{min} and y_{max} the prior becomes

$$p(\mathbf{y}|y_{\min}, y_{\max}, \alpha, k) = k!\,(y_{\max} - y_{\min})^{-k} \prod_{i=1}^{k} \left[\frac{y_i - y_{i-1}}{y_{\max} - y_{\min}} \right]^{\alpha - 1}, \tag{2.19}$$

and for α set to one, the uniform distribution outlined earlier is recovered, that is

$$p(\mathbf{y}|k) = k!\,(y_{\max} - y_{\min})^{-k}. \tag{2.20}$$

Parameterisation

The aim is to represent dispersion of surface waves using a partition modelling approach introduced in the previous chapter. Then using trans-dimensional McMC sampling to estimate the range of dispersion that predict the observations, in this case the spectrum of an inter-station EGF.

It was shown that, given assumptions, the dispersion curve is continuous and monotonically decreasing with frequency. At first glance it would appear that an obvious choice for representing such a function would be the piece wise linear segment approach shown in Figure 1.2(b). The draw back to this parameterisation is that it doesn't enforce monotonicity and the piece wise linear curve is only C^0 continuous and therefore would give a discontinuous group velocity dispersion curve.

A simple way to enforce monotonicity is to use partition modelling on both frequency and velocity axes which forms a non-uniform grid. By using the vertices of the intersection of the partition boundaries in order, a set of points in frequency and velocity is obtained that have velocity monotonically decreasing as a function of frequency. This is shown in Figure 2.4(a) where a piece-wise linear curve between vertices is plotted.

In order the generate a C^1 continuous function, a spline based interpolation between these points could be used. However, even with monotonically decreasing points, a cubic spline can still produce oscillations that would violate monotonicity. In Figure 2.4(b) a Hermite interpolant constructed by setting the gradient to zero at the end points and computing the gradients at intermediate points using finite differences is shown. At around 0.8 Hz there is a minimum which violates the monotonic requirement.

To remedy this, a variant of Hermite interpolation called monotone piece wise cubic interpolation [Fritsch and Carlson, 1980, Fritsch and Butland, 1984] can be used. This method uses a criterion for specifying the gradients at intermediate points that guarantees monotonicity of the interpolating curves. The results of this interpolation strategy are shown in Figure 2.4(b) and even with sharp deviations, the monotonicity of the points defined by the vertices is preserved by this interpolant.

In the parameterisation proposed here, the frequency range is fixed as part of the inversion, but the velocity domain is left to be variable within some upper and lower limits, effectively setting a uniform prior on the phase velocities. This can be accommodated by first constructing an "outer" prior on the initial and final velocities of the dispersion curve with respect to predefined minimum and maximum values as

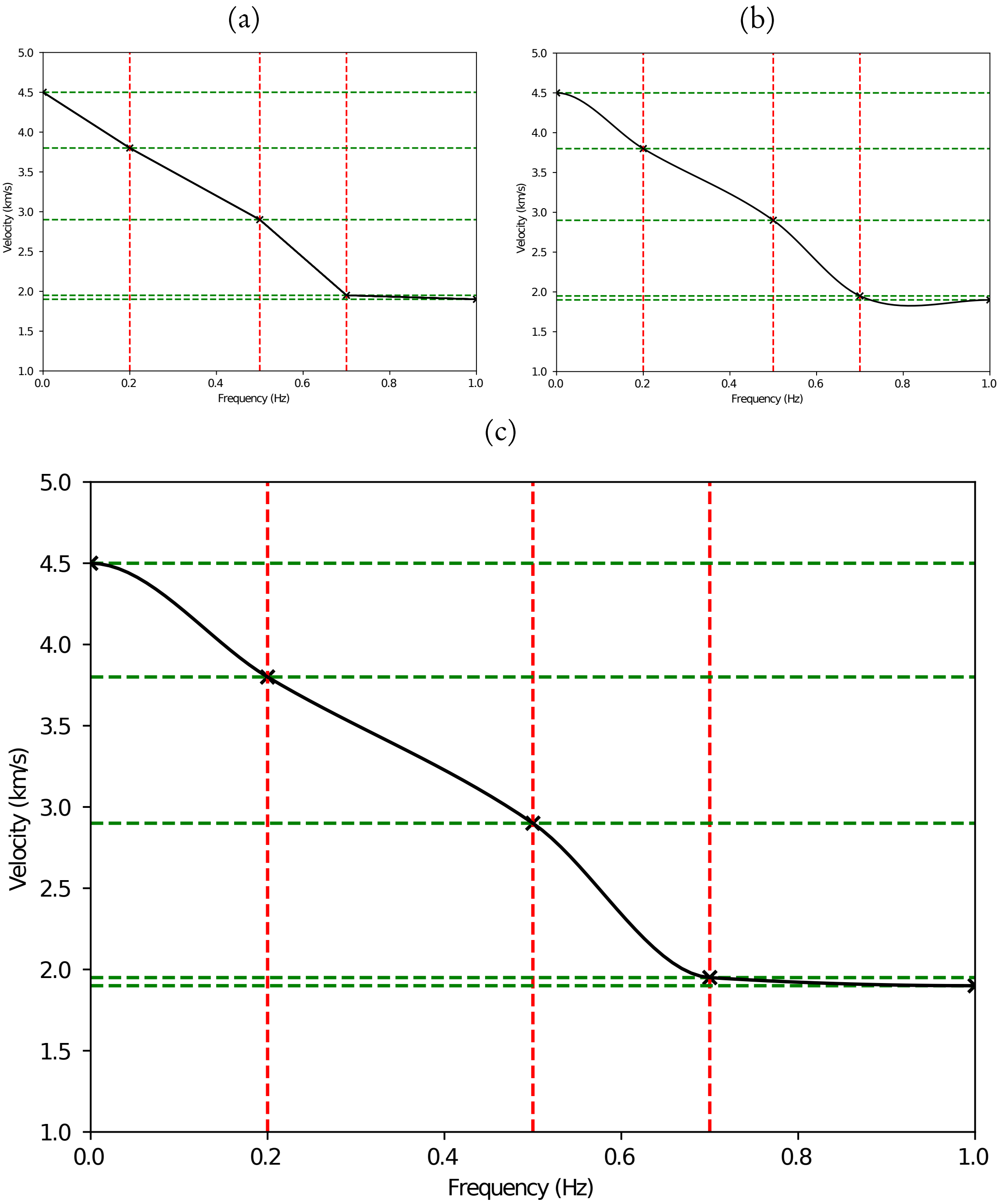


Figure 2.4: Various parameterisation options are shown for a dispersion curve. In (a) a piece wise linear interpolant is used but it is only C^0 continuous. In (b) a Hermite cubic interpolant is used but in some cases this can result in non-monotonic functions. In (c) a monotone cubic interpolant is used giving C^1 continuity while guaranteeing monotonicity.

$$p(v_0, v_k | v_{min}, v_{max}) = 2(v_{max} - v_{min})^{-2}, \tag{2.21}$$

where v_{min} and v_{max} are the prior bounds on the phase velocity, v_0 and v_k are the phase velocities of the lowest and highest frequencies respective. This expression represents a Dirichlet prior with three partitions along the velocity axis. The region defined by f_{min}, f_{max}, v_0 and v_k can be partitioned further into an irregularly spaced grid of k rows by k columns whose vertices define the dispersion curve. The prior expression for this is the combination of two Dirichlet distributions given by

$$\begin{aligned} p(\mathbf{f}, \mathbf{c} | k, v_0, v_k, f_{min}, f_{max}) = \quad & k!(f_{\text{max}} - f_{\text{min}})^{-k} \prod_{i=1}^{k} \left[\frac{f_i - f_{i-1}}{f_{\text{max}} - f_{\text{min}}} \right]^{\alpha_f - 1} \\ & k!(v_k - v_0)^{-k} \prod_{i=1}^{k} \left[\frac{v_i - v_{i-1}}{v_k - v_0} \right]^{\alpha_c - 1} . \end{aligned} \tag{2.22}$$

A schematic demonstrating this construction is shown in Figure 2.4(c) where, for example, v_{min} equals 1 and v_{max} equals 5. The outer Dirichlet prior would give the v_0 of 4.5 and v_k of 1.9.

This resulting prior, a recursive Dirichlet prior, has the properties that it represents the distribution of interpolation points that monotonically decreases between a specified v_{min} and v_{max} across a domain f_{min} and f_{max}. Coupled with the piece wise monotone cubic interpolant gives a prior for C^1 continuous curves between configurable bounds in frequency and phase velocity.

2.4.3 Likelihood and Forward model

In the inversion, the likelihood represents the probability that a model could give rise to the observations and this generally manifests itself as a difference between predicted observations from the model and the observations. The common approach to this is to express this as

$$\mathbf{G}(\mathbf{m}) = \mathbf{d} + \epsilon \tag{2.23}$$

where **G** is the forward model that transforms the vector of model parameters **m** into predicted observations, and ϵ is the randomly distributed error between predicted and observed data. This random error variable encompasses all sources of error including measurement error, errors in various fixed parameters of the inversion, and theory or approximation errors in the forward model. Although other distributions are possible, the assumption is that ϵ is normally distributed which allows a likelihood of the form

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{2\pi|C_d|}} \exp\left\{-\frac{1}{2}(\mathbf{G}(\mathbf{m})-\mathbf{d})^T C_d^{-1}(\mathbf{G}(\mathbf{m})-\mathbf{d})\right\}, \tag{2.24}$$

where C_d is the data error covariance matrix. A simplifying assumption is made here that the data noise is independent resulting in a diagonal error covariance matrix. The noise is unlikely to be truly independent as the observations are the real part of the spectrum of an inter-station EGF, however the assumption is that this is a reasonable approximation. With this simplifying assumption, the likelihood reduces to

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{2\pi\sigma_d^{2N}}} \exp\left\{-\frac{1}{2}\sum_i \frac{(G_i(\mathbf{m})-d_i)^2}{\sigma_d^2}\right\}, \tag{2.25}$$

where σ_d is the independent standard deviation of the noise on each observation.

In the case where the noise level is not precisely known or estimated, hierarchical sampling of noise scaling term can be used [Malinverno and Briggs, 2004]. To implement this, a scaling parameter λ is introduced and the independent noise level becomes

$$\sigma_d = \lambda\sigma_e, \tag{2.26}$$

where σ_e is the noise estimate and λ is a scaling parameter that is inverted for during the inversion. This helps ensure results are not biased by inaccuracy in estimating the noise level. To obtain a noise estimate, the standard deviation of the real part of the

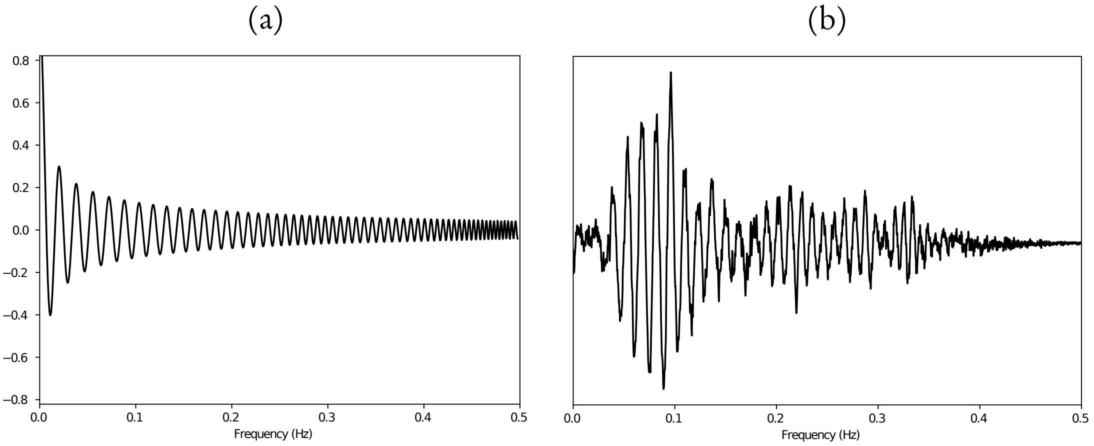


Figure 2.5: Comparison of a Bessel function in (a) of the predicted real part of the spectrum and in (b) an example of the real part of the spectrum obtained from cross-correlation of ambient noise between two stations.

spectrum in a quiescent region is used. This gives an estimate of observational signal noise. Other sources of errors contribute to to the misfit between predictions and the observations such as theoretical modelling approximation errors.

In the forward model of this problem, given a model consisting of a piece wise cubic curve for phase velocity as a function of frequency, the predicted Bessel function can be computed using

$$\bar{\rho}(\omega_0, r) = J_0\left(\frac{\omega_0}{c(\omega_0)} r\right), \tag{2.27}$$

and then over the frequency range of interest compute (2.25) as the difference between the predicted and observed real spectrum. In Figure 2.5, an example Bessel function and an example real spectrum are shown.

In the figure it is evident that the amplitude of the observed real spectrum varies considerably across the frequency domain so fitting the Bessel function is difficult. This is one of the reasons that Ekström et al. [2009] suggested using the zero crossings of the observed part of the real spectrum since this removes issues of fitting Bessel function to a varying amplitude spectrum. As Menke and Jin [2015] showed, the zero crossings method is ill-determined with noise producing spurious zero crossings. Better results can be obtained by considering the entire part of the signal. The spectrum could be whitened, but as stated earlier, such normalisation has implications for uncertainty analysis so is undesirable. A simple approach taken in this study is to assume the observation is a scaled version of the predicted Bessel function and compute an optimal scaling given the predicted Bessel function and the observed spectrum.

If the predicted Bessel function values are **B** for the observed spectrum in **d**, an optimal scaling term γ is required such that

$$|\gamma\mathbf{B}-\mathbf{d}|_2 \tag{2.28}$$

is minimised. This is straight forward and can be computed using

$$\gamma = \left|\frac{\sum_i B_i d_i}{\sum_i B_i^2}\right|, \tag{2.29}$$

where i indexes over each frequency of comparison. The absolute signs are required to prevent negative scaling terms which would cause the Bessel function to flip about the frequency axis resulting in a phase shift (effectively a half cycle skip). Using a maximal likelihood estimator is a common tactic for nuisance parameters, and a similar approach was taken by Dettmer et al. [2015] in a receiver function inversion. An alternative here would be to sample the γ scaling parameter during the inversion, but this parameter is of little consequence in subsequent inference. A more advanced inversion may also jointly solve for the amplitude or envelope of the Bessel function as this may provide extra information that could be used to understand the ambient noise spectrum and perhaps frequency dependent attenuation. At this stage this is not of interest so the simple optimal scaling term suffices, however this may be an area of future research.

2.4.4 Proposals

In a Bayesian McMC inversion, a Markov chain is constructed by starting from some arbitrary model $\mathbf{m}$. This model is then perturbed by sampling some proposal probability density and evaluating the likelihood of this new proposed model. The Metropolis-Hastings-Green acceptance criteria [Green, 1995], that is a new model $\mathbf{m}'$ is accepted with probability

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = \min\left\{1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m}' \rightarrow \mathbf{m})}{Q(\mathbf{m} \rightarrow \mathbf{m}')} |\mathcal{J}|\right\}, \tag{2.30}$$

where prime indicates proposed models, $p(\mathbf{m})$ is the prior, $p(\mathbf{d}|\mathbf{m})$ is the likelihood and $Q(\mathbf{m} \rightarrow \mathbf{m}')$ is the proposal distribution that perturbs the model from the current, $\mathbf{m}$, to the proposed $\mathbf{m}'$ and visa-versa. If the proposed model is accepted, it becomes the current model for the next iteration, otherwise the current model is retained.

Part of the formulation of the problem is determining a set of proposals for the inversion. In this study, standard proposals for partition modelling are used which consist of

Value the velocity of a randomly chosen partition boundary is perturbed

Move the frequency of a randomly chosen partition boundary is perturbed

Birth a randomly chosen partition is split in both frequency and velocity creating a new interpolation node.

Death two randomly chosen neighbouring partitions are merged into one removing an interpolation node.

Hierarchical Noise randomly perturb the hierarchical noise scaling factor, λ, in the likelihood.

At each iteration, a proposal is randomly chosen from one of the above to act on the current model. The individual proposals are described in depth in the following sections.

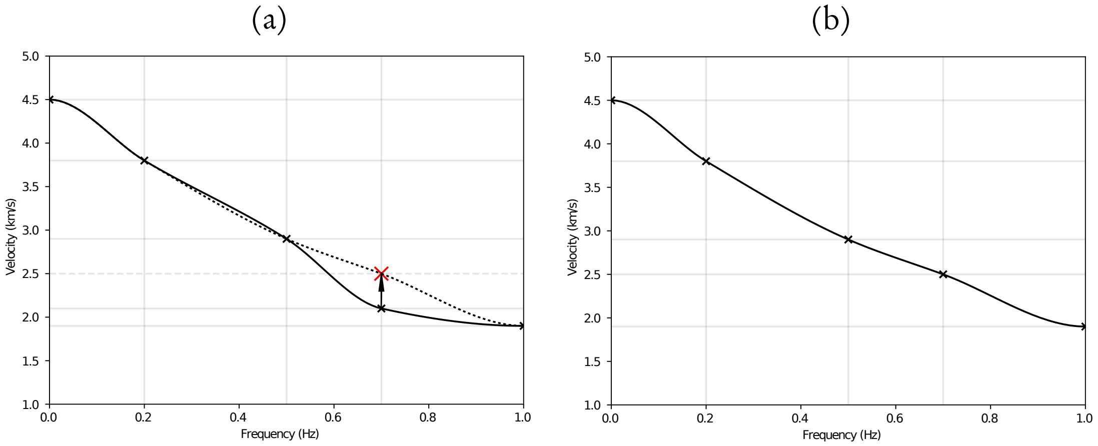


Figure 2.6: This figure gives a schematic of the value proposal. A velocity value of a randomly chosen partition boundary is perturbed to create a new point shown with the arrow and red cross in (a). The new proposed model resulting is shown in (b).

2.4.5 Value

In a value proposal the steps, shown in schematic form in Figure 2.6, are

1. Randomly select a partition boundary including edge boundaries.
2. Perturb the velocity of the model at the boundary by sampling from a symmetric probability distribution (e.g. Gaussian).

The symmetric distribution requirement means that the proposal distribution cancels, for example, if using a Gaussian distribution, the proposal is

$$Q(\theta \rightarrow \theta') = \frac{1}{N_p} N(v_i' - v_i, \sigma_v) \tag{2.31}$$

for some standard deviation σ_v, and the reverse

$$Q(\theta' \rightarrow \theta) = \frac{1}{N_p} N(v_i - v_i', \sigma_v) \tag{2.32}$$

which are exactly the same as the Gaussian distribution is symmetric.

It should be noted that a sufficiently large perturbation could cause the monotonicity

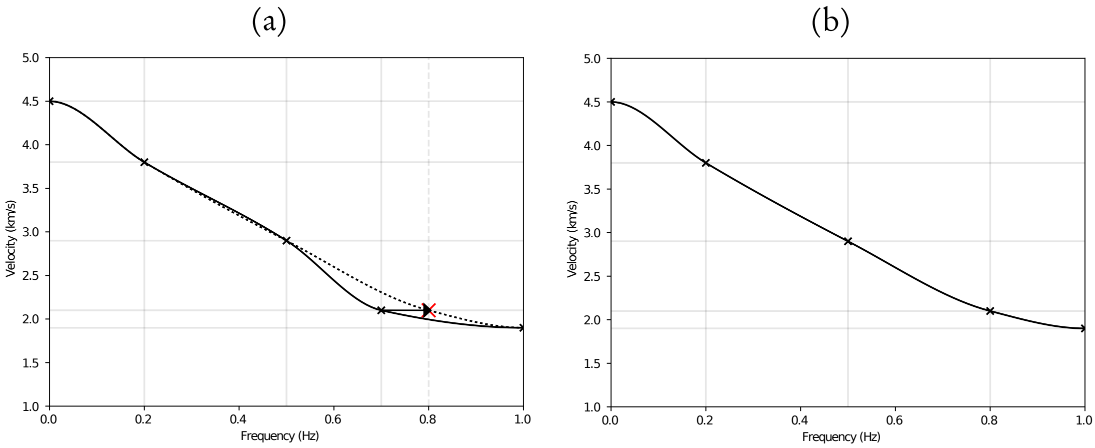


Figure 2.7: This figure gives a schematic of the move proposal. A partition boundary is randomly chosen and perturbed in frequency (horizontally) to create a new point shown with the arrow and red cross in (a). The new proposed model resulting is shown in (b).

or ordering of the velocities. Such proposals are rejected as this would cause negative partition widths in the Dirichlet prior which results in a zero prior ratio.

2.4.6 Move

The move proposal, shown schematically in Figure 2.7, is very similar to the value proposal except the move proposal

1. Randomly selects a partition boundary except the edge boundaries, and
2. Perturbs the frequency of partition boundary by sampling from a symmetric probability distribution.

Similar to value proposals, large changes that would re-order points cause the Dirichlet prior to be zero and subsequent rejection of the proposal and hence monotonicity is preserved.

2.4.7 Birth

In a birth proposal, a new interpolation node is created by splitting an existing cell vertically and horizontally. This is shown in Figure 2.8 where in (a) a cell is split

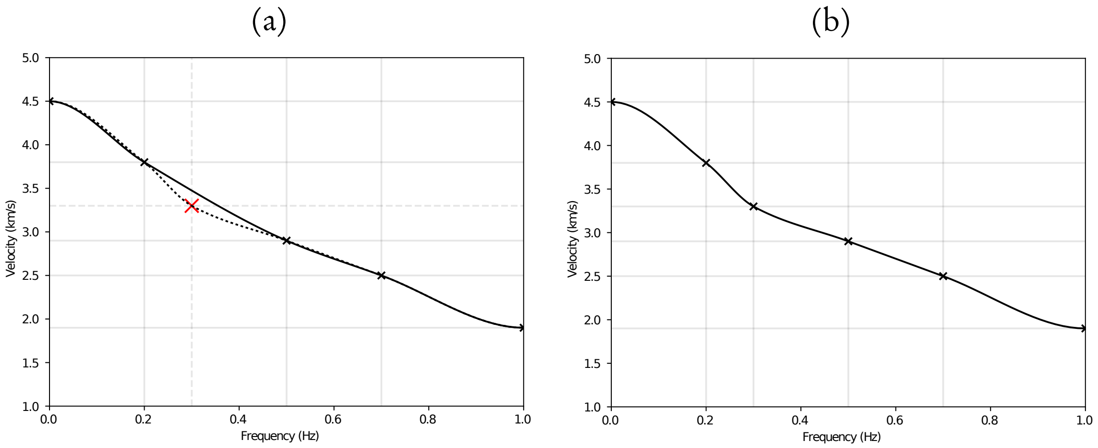


Figure 2.8: This figure gives a schematic of the birth proposal. A cell is chosen at random and in randomly split in frequency (horizontally) and velocity (vertically) to create a new point on the curve shown with a red cross in (a). The new proposed model resulting is shown in (b).

vertically and horizontally as indicated with dotted lines to create a new interpolation node marked with a red cross. The individual steps are

1. Select a random cell
2. Generate random uniform, ξ, between 0 and 1 for a horizontal splitting location
3. Generate random uniform, η, between 0 and 1 for a vertical splitting location
4. Split cell according the ξ and η

The proposal needs to be balanced with the reverse proposal, discussed in more detail in the next section. The forward proposal density can be written

$$Q(\mathbf{m} \rightarrow \mathbf{m}') = \frac{1}{k} p(\xi) p(\eta), \tag{2.33}$$

where k is the number of active cells, ξ is the horizontal splitting random variable and η is the vertical. If these last two random variables are sampled from uniform variates between 0 and 1 then the proposal density reduces to

$$Q(\mathbf{m} \rightarrow \mathbf{m}') = \frac{1}{k}. \tag{2.34}$$

This leaves the Jacobian to compute which can be computed by examining the bijection between variables during the birth process

$$df_i, dv_i, \xi, \eta \longleftrightarrow df_i', dv_i', df_{i+1}', dv_{i+1}' \tag{2.35}$$

where df represents the cell width or change in frequency, dv represents the cell height or negative change in velocity. The proposal is local so that other model parameters remain unchanged during a birth. This means that only four parameters of the two new cell need to be considered for the Jacobian determinant as the full Jacobian would consist of a diagonal of ones and the sub matrix of modified values.

The relation ships between these variables are

$$df_i' = \xi df_i \tag{2.36}$$

$$dv_i' = \eta dv_i \tag{2.37}$$

$$df_{i+1}' = (1-\xi)df_i \tag{2.38}$$

$$dv_{i+1}' = (1-\eta)dv_i \tag{2.39}$$

$$, \tag{2.40}$$

from which a Jacobian can be constructed

$$\mathscr{J} = \begin{bmatrix} \frac{\partial df_i'}{\partial df_i} & \frac{\partial dv_i'}{\partial df_i} & \frac{\partial df_{i+1}'}{\partial df_i} & \frac{\partial dv_{i+1}'}{\partial df_i} \\ \frac{\partial df_i'}{\partial dv_i} & \frac{\partial dv_i'}{\partial dv_i} & \frac{\partial df_{i+1}'}{\partial dv_i} & \frac{\partial dv_{i+1}'}{\partial dv_i} \\ \frac{\partial df_i'}{\partial \xi} & \frac{\partial dv_i'}{\partial \xi} & \frac{\partial df_{i+1}'}{\partial \xi} & \frac{\partial dv_{i+1}'}{\partial \xi} \\ \frac{\partial df_i'}{\partial \eta} & \frac{\partial dv_i'}{\partial \eta} & \frac{\partial df_{i+1}'}{\partial \eta} & \frac{\partial dv_{i+1}'}{\partial \eta} \end{bmatrix} = \begin{bmatrix} \xi & 0 & 1-\xi & 0 \\ 0 & \eta & 0 & 1-\eta \\ \frac{1}{df_i} & 0 & \frac{-1}{df_i} & 0 \\ 0 & \frac{1}{dv_i} & 0 & \frac{-1}{dv_i} \end{bmatrix}. \tag{2.41}$$

With some simple row and column reordering, an analytical expression for the determinant of the Jacobian is

$$|\mathcal{J}| = \frac{1}{df_i \, dv_i}. \tag{2.42}$$

For the reverse proposal, an internal cell is select from all except the last one and the selected cell is merged with the cell to the right. After a birth, there are $k+1$ cells but since the selection is from all but the last cell, the choice is from one of k cells, hence the reverse proposal is simply

$$Q(\mathbf{m}' \to \mathbf{m}) = \frac{1}{k'-1} = \frac{1}{k}, \tag{2.43}$$

and the proposal ratio cancels leaving the acceptance criteria for birth as

$$\alpha(\mathbf{m} \to \mathbf{m}') = \min\left\{1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{1}{df_i \, dv_i}\right\}. \tag{2.44}$$

2.4.8 Death

The death proposal is the reverse of birth from the previous section, that is

1. Select two neighbouring cells or equivalently select cell from `0` to $N_{cells} - 1$
2. Merge two cells into one.

From similar arguments to the birth proposal, the proposal ratio cancels in the acceptance criteria leaving

$$\alpha(\mathbf{m} \to \mathbf{m}') = \min\left\{1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} df_i' dv_i'\right\}, \tag{2.45}$$

where the terms df_i' and dv_i' represent the cell widths and heights respectively of the merged cell created by the proposal and result from the Jacobian in a similar fashion to that shown in the previous section.

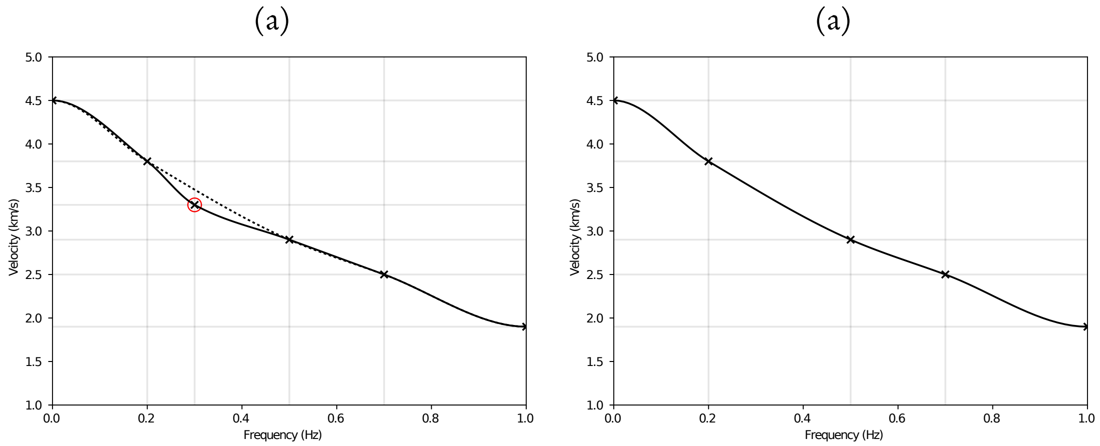


Figure 2.9: This figure gives a schematic of the birth proposal. Two neighbouring cells are chosen at random and merged to remove a point on the curve shown circled in red in (a). The new proposed model resulting is shown in (b).

2.4.9 Hierarchical

For the hierarchical proposal, the error scaling factor λ in (2.26) is perturbed. This proposal is similar to move and value proposals in that the perturbation is sampled from a symmetric proposal distribution. The difference here is that that normalisation term of the likelihood (2.25) must be considered in the acceptance criteria as the σ_d term, that depends on λ affects this.

Validation

Trans-dimensional samplers are inherently difficult to formulate and validation is required to verify the acceptance criteria correctly maintains detailed balance. Detailed balance is critical for trans-dimensional samplers and if the correct balance isn't maintained then the results can be severely biased toward either too simple or too complex models.

A simple test is to run a Markov chain with the likelihood function set to a constant value. This means that the Markov chain receives no additional information and that the posterior should be proportional to the prior within sampling accuracy, that is, the prior on k, the number of partitions, should be recovered in the posterior.

In Figure 2.10 the posteriors retrieved from Markov chains with constant likelihoods are shown. In these tests, a uniform prior for k is set with a maximum of 20, 35, or 50 cells. For the α parameter for the horizontal Dirichlet prior, three separate values are simulated: 1 which is essentially a uniform prior on the location of the points, as well as α values of 2 and 3. For these tests, 10 million iterations are used three times with different random seeds. In the figure, the histograms of the individual runs are shown in light grey with the mean of the three runs shown as the solid black outline histogram. In all cases, the retrieved posterior matches the prior and is uniform to sampling accuracy with little variation between runs with different seeds.

Regression Problem

As another test, a simple regression problem is used to test the trans-dimensional inversion code. In this test a true model is constructed from a simple cubic function

$$v(f) = 4f^3 - 6f^2 + 5. \tag{2.46}$$

To construct regression observations, 50 randomly generated x coordinates in the range 0 … 1 are created and the y value of the function at those points evaluated. To the y coordinates, independent Gaussian noise is added with three different standard deviations, 0.1, 0.2 and 0.3 km/s. The results of the inversion are shown in Figure 2.11 with the left hand column showing the true curve and data points and the right hand column showing the results with the probability density of the curve shown in blue shading (darker blue indicates higher probability) and the mean of the ensemble shown with a dotted yellow line.

This simple test shows good recovery of the underlying true curve with good estimation of the errors, that is, increased data noise leads to increased uncertainty in the results as shown by the thicker blue bands in the results.

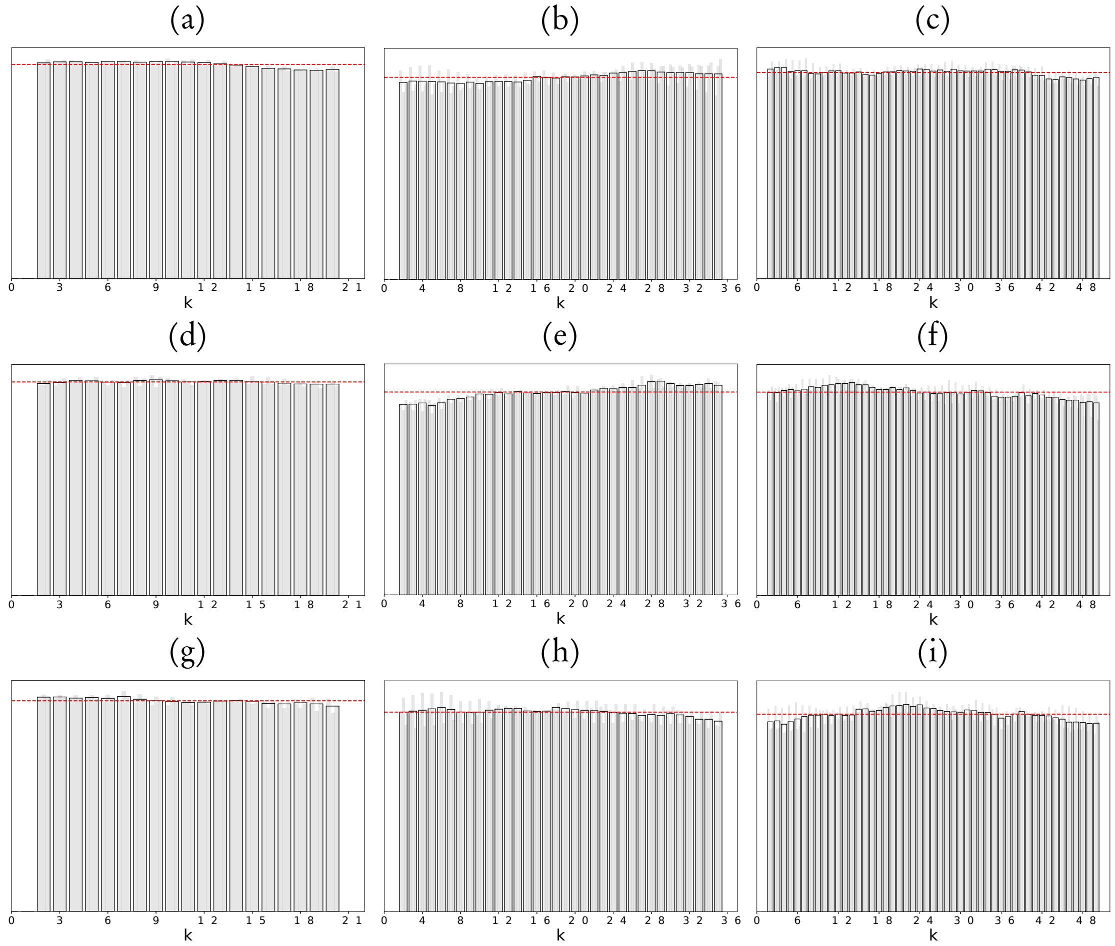


Figure 2.10: Given a uniform prior on k, the number of points defining the dispersion curve, when the likelihood is fixed to a constant a uniform posterior is expected. In these plots the posterior is shown for varying parameters. In (a), (b), (c) with the α parameter set to 1 for 20, 35, and 50 maximum points respectively. Similarly in (d), (e), and (f) the posteriors for α set to 2, and in (g), (h), and (i) for α set to 3. In each test 3 different seed values are used and the plots show the individual histograms faintly with the average of the 3 different runs shown in dark outline. A red dashed horizontal line shows the expected uniform value. In all plots an approximate uniform posterior is retrieved.

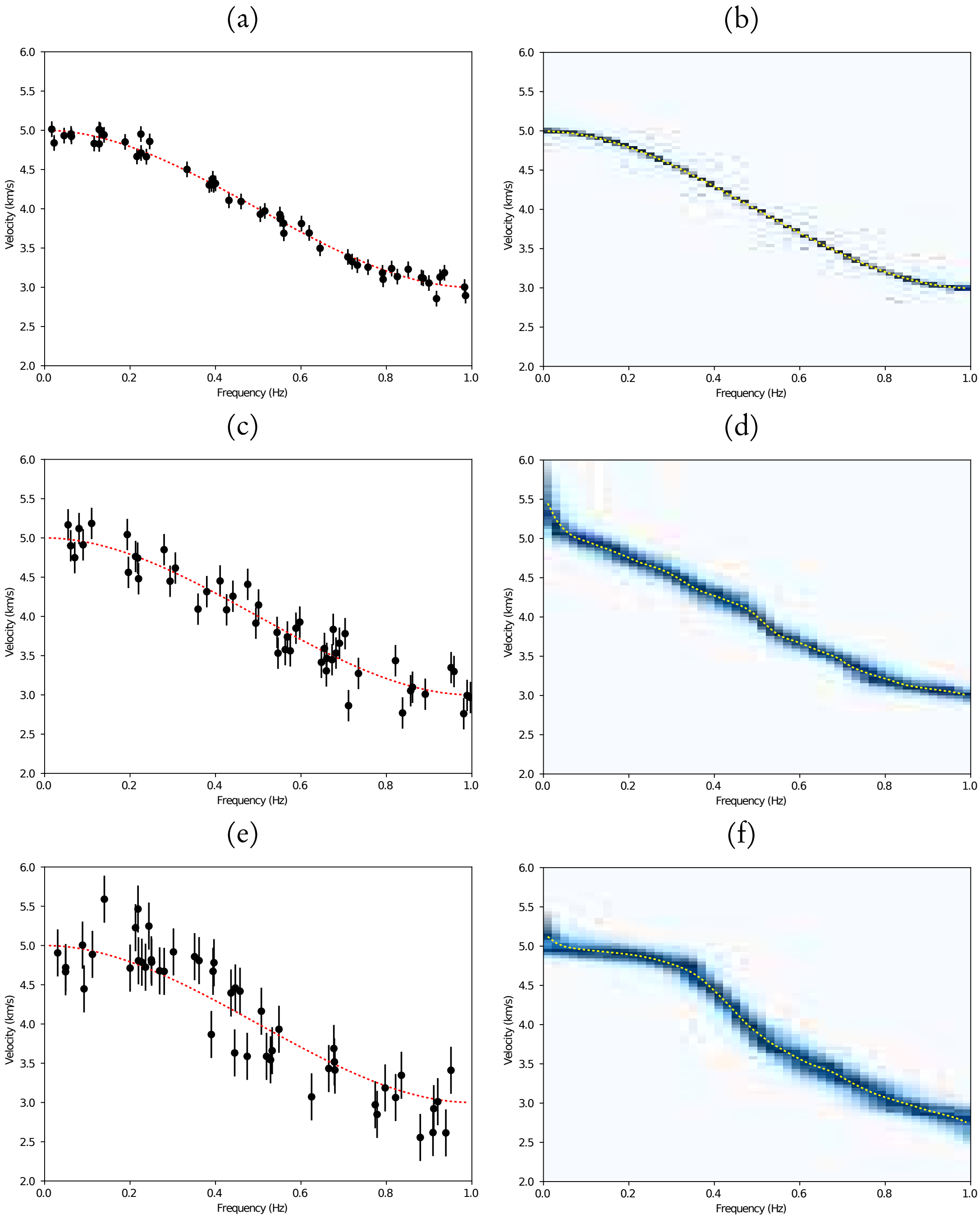


Figure 2.11: The results of inverting for a dispersion curve using a simple regression forward model. In (a), (c), and (e) the true curve is shown with a red dotted line and the data with added noise with points with error bars. In (b), (d) and (f) the posterior ensemble histogram is shown with more intense blues representing more likely regions and the mean of the ensemble shown with a yellow dotted curve.

2.5 Love and Rayleigh Dispersion

Having demonstrated that the inversion methodology is self-consistent and unbiased, this method can now be applied to real EGFs obtained from continuously recording seismometers stationed in Iceland. Detailed examination of two station pairs are discussed, the first (HOT15 - HOT20) has a relatively long (approximately 200km) great circle path that traverses older crust. The second (HOT23 - HOT26) is a short path that crosses an area of active volcanism. These two paths are designed to be indicative of range of possible behaviours within this data set to evaluate the performance of the new method.

For the inversions, 12 independent chains are simulated with a prior range for the phase velocity of 4.5 to 1.5 km/s for Love wave dispersion and 4.0 to 1.0 km/s for Rayleigh wave dispersion. The frequency range for the fit of the Bessel function is between 0 and 0.5 Hz. The McMC chains are simulated for 2,000,000 steps with the first 500,000 iterations removed as burnin. Additionally, parallel tempering [Earl and Deem, 2005, Sambridge, 2014] is used with 4 independent temperatures with a logarithmically spaced temperatures between 1 and 5 with samples taken from only those chains with unit temperature. The reason for this is that the problem is very multimodal and chains could become stuck in local minima. Parallel tempering allows us to properly explore the full range of solutions.

For the first path the results of the inversion are shown in Figure 2.12 where in (a) is the result for the Rayleigh wave dispersion and (b) is for the Love wave dispersion. These results are quite good with low uncertainty, particularly the Love wave dispersion curve. In the Rayleigh wave dispersion curve at around 0.25 Hz, there is some multi-modality and at around 0.4 Hz the curve ensemble becomes incredibly bifurcated.

The reason for both these becomes apparent when looking at the observed spectrum and the ensemble of the predicted Bessel functions as shown in Figure 2.13. In this figure, there is a spectral hole in the real part of the spectrum of the Rayleigh wave

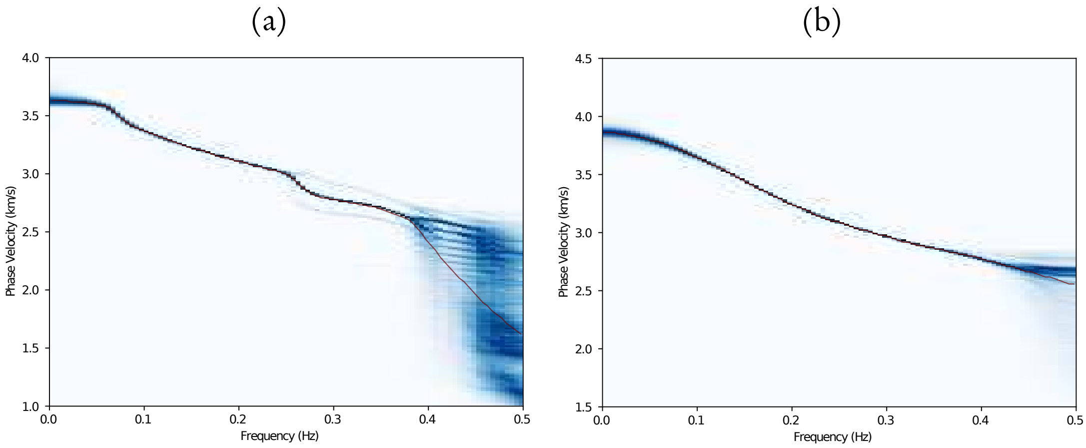


Figure 2.12: The results from inverting the dispersion between stations HOT15 and HOT20 for Rayleigh (a) and Love (b) waves. The blue shading represents the probability of the dispersion curve at a given point with darker blue representing more likely. The red curve is the median dispersion curve of the ensemble.

EGF at around 0.25 Hz and this is the cause for the multi-modality starting around this point. The spectral coverage for the Love wave is relatively consistent throughout the range of interest, so the dispersion curve is unimodal.

For the second example path the results are shown in Figure 2.14 where both Rayleigh and Love waves are multi-modal across a broad range of frequencies. For the Love wave results in (b), the phase velocities are too low and the likely "true" dispersion curve (based on prior knowledge of phase velocities around these frequencies) appears

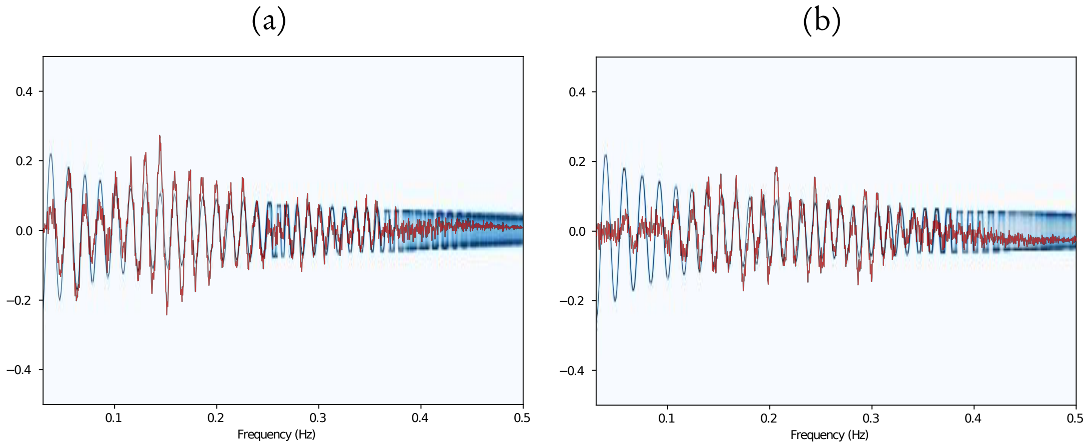


Figure 2.13: The histogram of the Bessel functions in shaded blue (darker blue is more probable regions of the Bessel function) and over plotted the real part of the spectrum of the cross-correlation function between stations HOT15 and HOT20 for Rayleigh (a) and Love (b) waves.

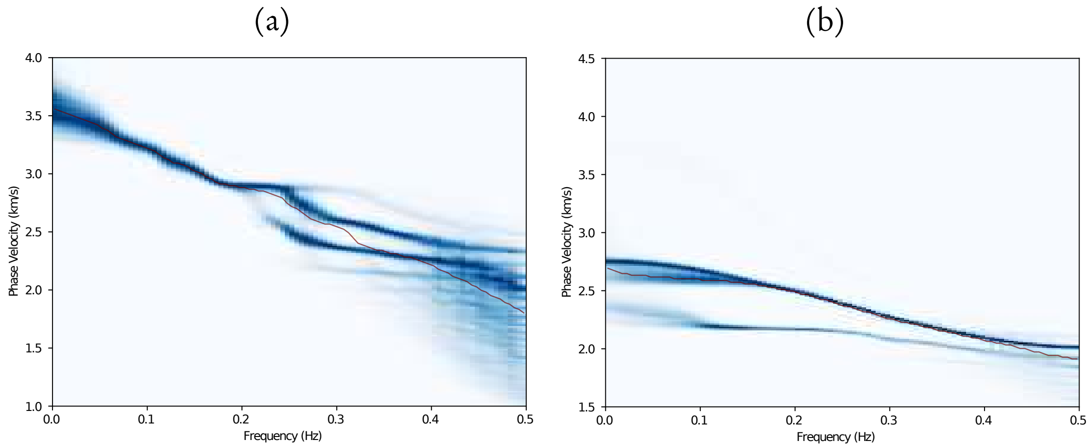


Figure 2.14: The results from inverting the dispersion between stations HOT23 and HOT26 for Rayleigh (a) and Love (b) waves. The blue shading represents the probability of the dispersion curve at a given point with darker blue representing more likely. The red curve is the median dispersion curve of the ensemble. These results demonstrating greater multi-modality that in effect represents cycle skipping of the Bessel function.

in very faint blue starting from just less than 4 km/s at 0 Hz.

Examining the spectrum and Bessel function fits of these two dispersion inversions in Figure 2.15, there is only a very limited amount of coherent signal in the real part of the spectrum. For the Love wave spectrum, there is very little low frequency information which is the cause of the inversion choosing low phase velocities.

These two examples presented here were chosen to highlight the range of behaviour for this inversion method, from well constrained to weakly constrained as in Figure 2.14. If the median from all the results are plotted as shown in Figure 2.16, the dispersion is well grouped as expected since each station pair represents path average dispersion so all station pairs should be correlated with each other. The second example Love wave inversion is a clear outlier and is the single curve at the bottom of Figure 2.16(b)

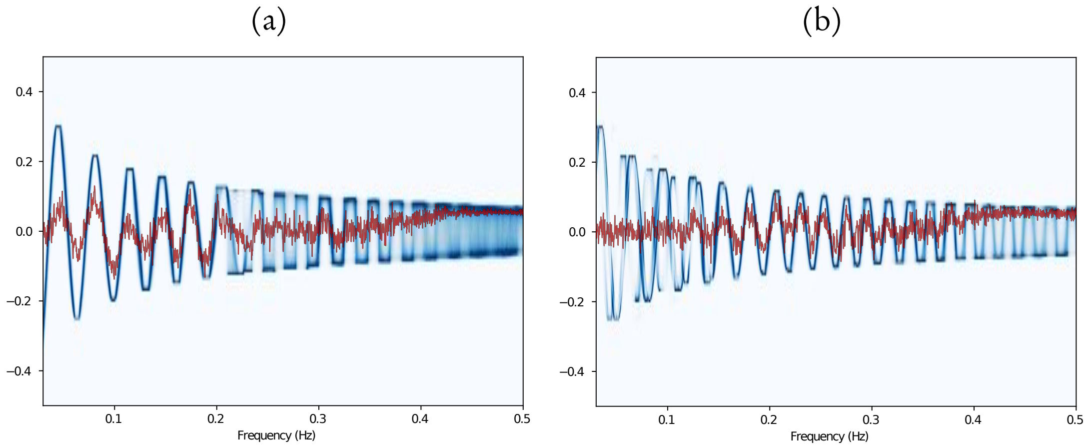


Figure 2.15: The histogram of the Bessel functions in shaded blue (darker blue is more probable regions of the Bessel function) and over plotted the real part of the spectrum of the cross-correlation function between stations HOT23 and HOT26 for Rayleigh (a) and Love (b) waves.

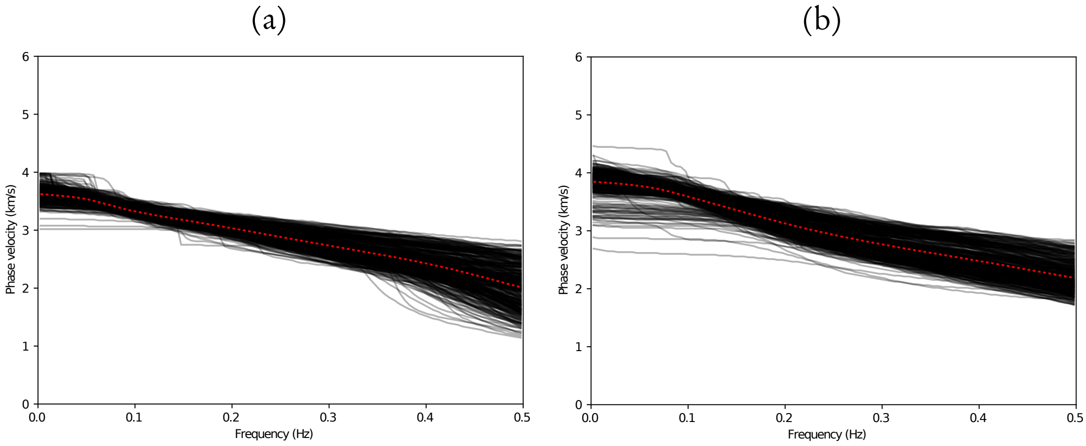


Figure 2.16: The median of the ensemble of all available station pairs for Rayleigh (a) and Love (b) wave EGFs. The inversion is stable and produces a good clustering of dispersion curves with few outliers.

2.6 Joint inversion of Love and Rayleigh wave dispersion

While the results from the independent inversion of Love and Rayleigh wave dispersion are encouraging there remains room for improvement. A clear trait of the real part of the spectrum for Love and Rayleigh dispersion is that they appear to have different excitation frequencies. In Figure 2.13(a) for example, at around 0.25 Hz there is a "hole" in the real part of the Rayleigh wave spectrum whereas in (b) the Love wave spectrum still has relatively strong signal throughout 0.2 to 0.3 Hz. In the worst performing station pair shown in Figure 2.15, the Rayleigh wave signal in (a) has signal from approximately 0 to 0.2 Hz whereas the Love wave signal in (b) has signal from 0.15Hz to 0.3Hz. This suggests the possibility of using joint inversion of both Love and Rayleigh dispersion to smooth over these frequency holes.

Random sampling from crustal models

A generally acknowledged rule of thumb is that at a given frequency, the Rayleigh wave phase velocity is slower than that of a Love wave. Numerical approximations of the probability distribution of this relationship can be obtained by sampling realistic shear wave velocity models and computing dispersion curves for Love and Rayleigh waves using a forward model.

To compute Rayleigh waves, density and P-wave velocity are required. The assumptions used here are that the Earth models are isotropic and that empirical relationship of Brocher [2005] is sufficient where

$$v_p = 0.9409 + 2.0947 v_s - 0.8206 v_s^2 + 0.2683 v_s^3 - 0.0251 v_s^4 \tag{2.47}$$

$$\rho = 1.6612 * v_p - 0.4721 * v_p^2 + 0.0671 * v_p^3 - 0.0043 * v_p^4 + 0.000106 * v_p^5, \tag{2.48}$$

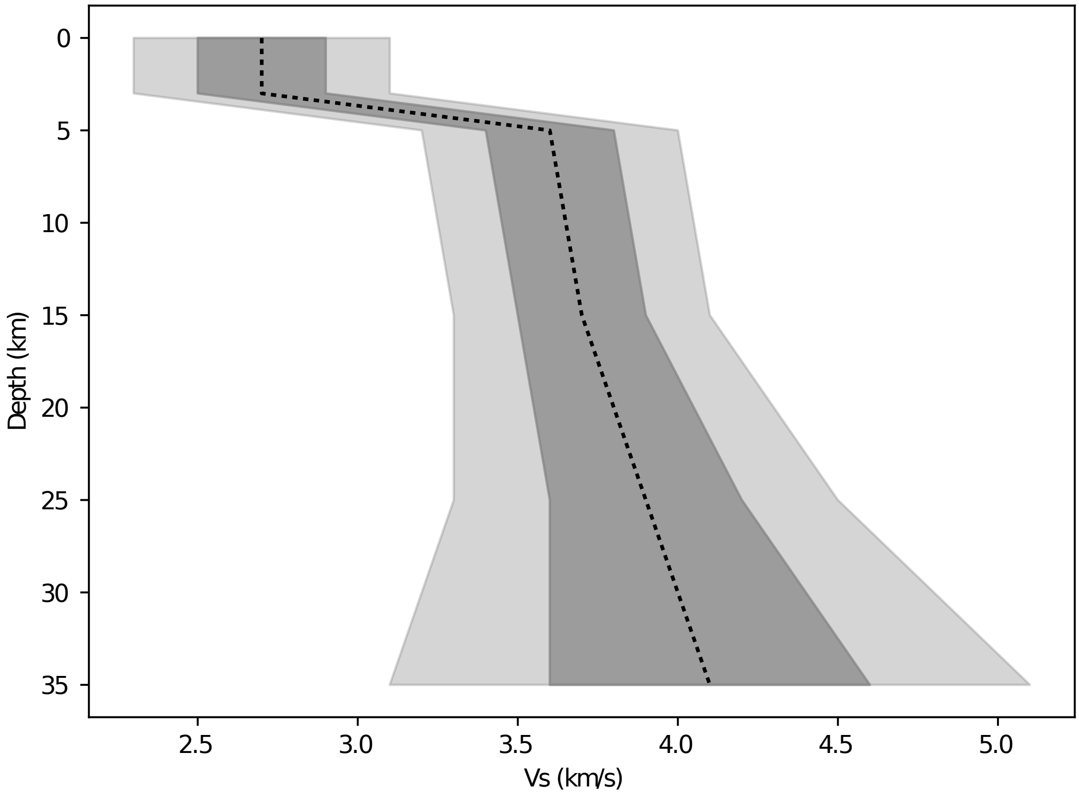


Figure 2.17: This figure shows the prior on shear wave velocity as a function of depth. The prior is Gaussian distributed with a mean show with the solid line. The shaded regions show one and two standard deviation ranges.

leaving only a prior on v_s to define.

For a prior on v_s, previous studies such as that of Green et al. [2017] can be used as a guide. Unfortunately the study of Green et al. [2017] only published estimates of shear wave velocity down to 10km, so broad Gaussian priors are added at depth to merge the prior smoothly into to a global reference model [Kennett et al., 1995]. The prior is shown in Figure 2.17.

The procedure for generating random models is as follows:

1. Generate a random number of layers to create between 2 and 10.

2. Generate the interfaces uniformly between 0 and 100km depth.

3. Generate a random shear wave velocity for each layer by sampling from the prior

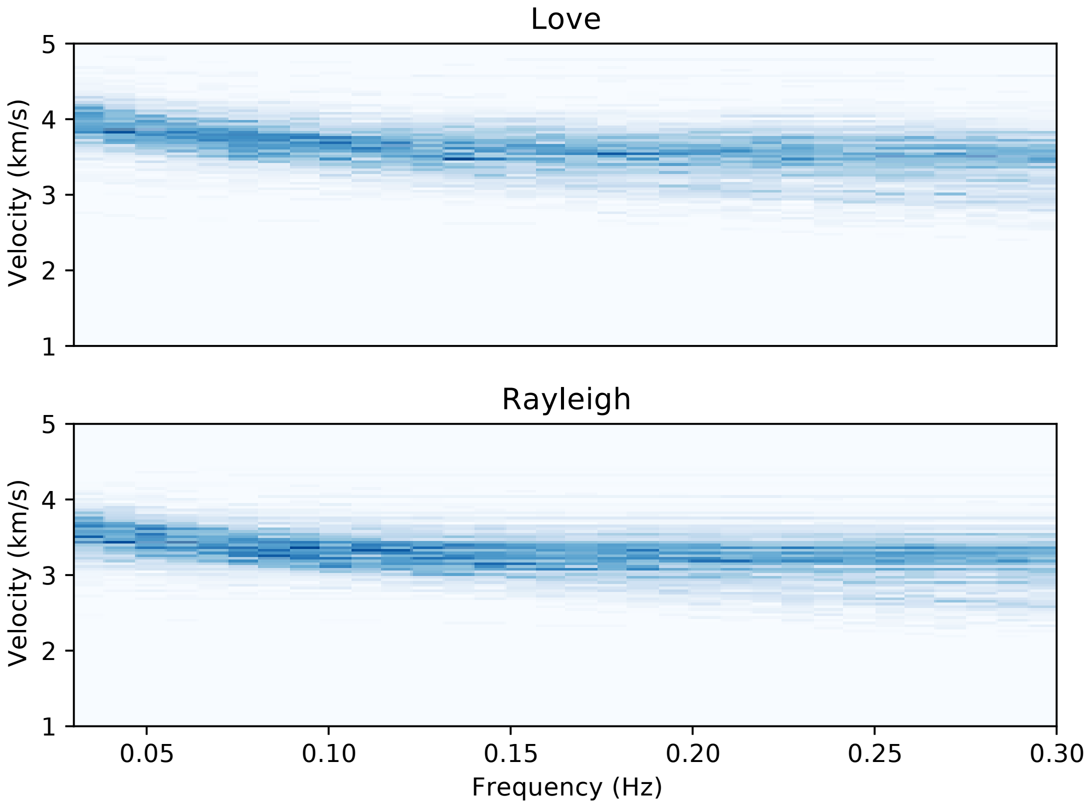


Figure 2.18: The distribution of the dispersion curves obtained for Love and Rayleigh waves by sampling from a broad shear wave velocity model.

at the midpoint of the layer.

4. Compute v_p and ρ

A dispersion curve resulting from this model for both Love and Rayleigh waves using the Thomson-Haskell method [Thomson, 1950, Haskell, 1953]. Occasionally thin layers can cause numerical instabilities in computing the dispersion curve, particularly with Rayleigh wave dispersion, so these curves are discarded from the sample. The final result is a numerical estimate of the probability distribution of dispersion curves given the prior on the shear wave velocity profile. The distribution of dispersion for Love and Rayleigh waves is shown in Figure 2.18.

In the Figures, it is evident that the dispersion curves are monotonic (also verified by a separate check during sampling) and that it appears that the Rayleigh wave phase ve-

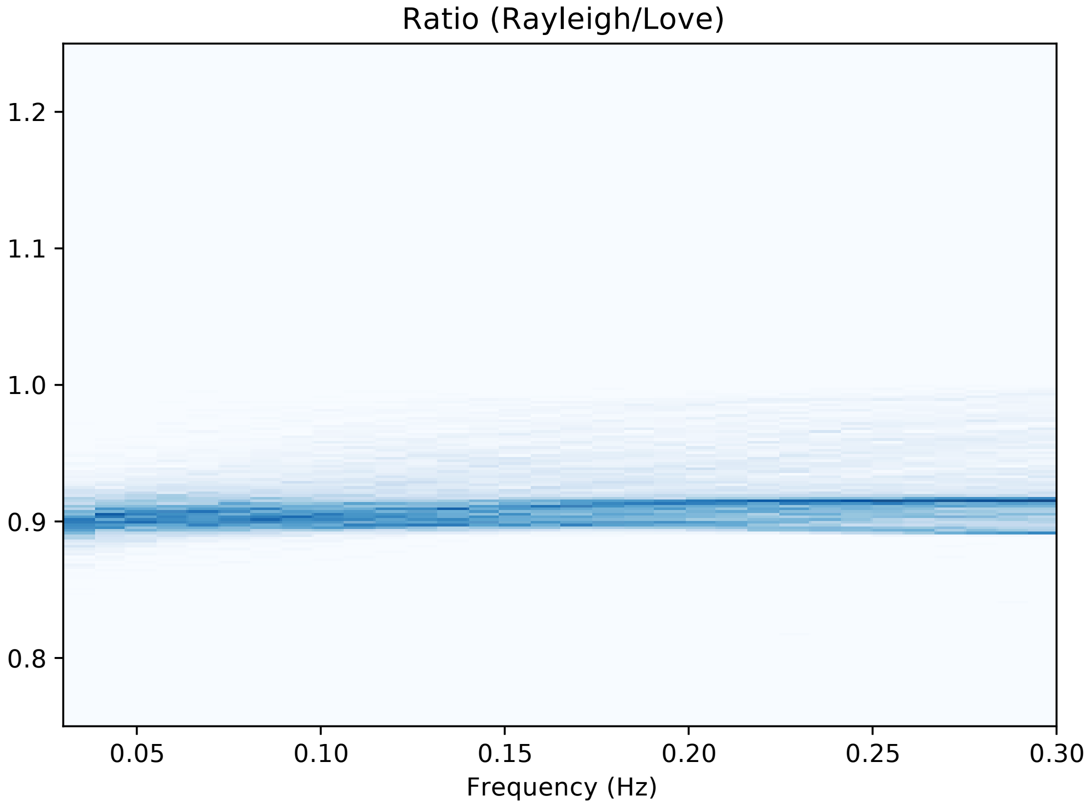


Figure 2.19: The distribution of the ratio of Rayleigh wave phase velocity to Love wave for the range of frequencies of interest in ambient noise tomography.

locity dispersion looks much like a scaled version of the Love wave dispersion. This relationship between Love and Rayleigh wave velocities can be directly shown by computing the ratio of the two and plotting the distribution of this over the frequency range of interest as in Figure 2.19.

From the figure, it can be seen that Rayleigh wave phase velocities appear to be always less than that of Love waves. In examining the distribution across all frequencies as in Figure 2.20, the ratio is strongly confined to approximately 0.9, that is, the Rayleigh phase velocity is always approximately 90% of the Love phase velocity across all frequencies of interest.

These simple numerical simulations can provide valuable insights into a Bayesian formulation of a geophysical inversion problem and this information can be incorporated into a joint inversion of Love and Rayleigh wave ambient noise observations. From

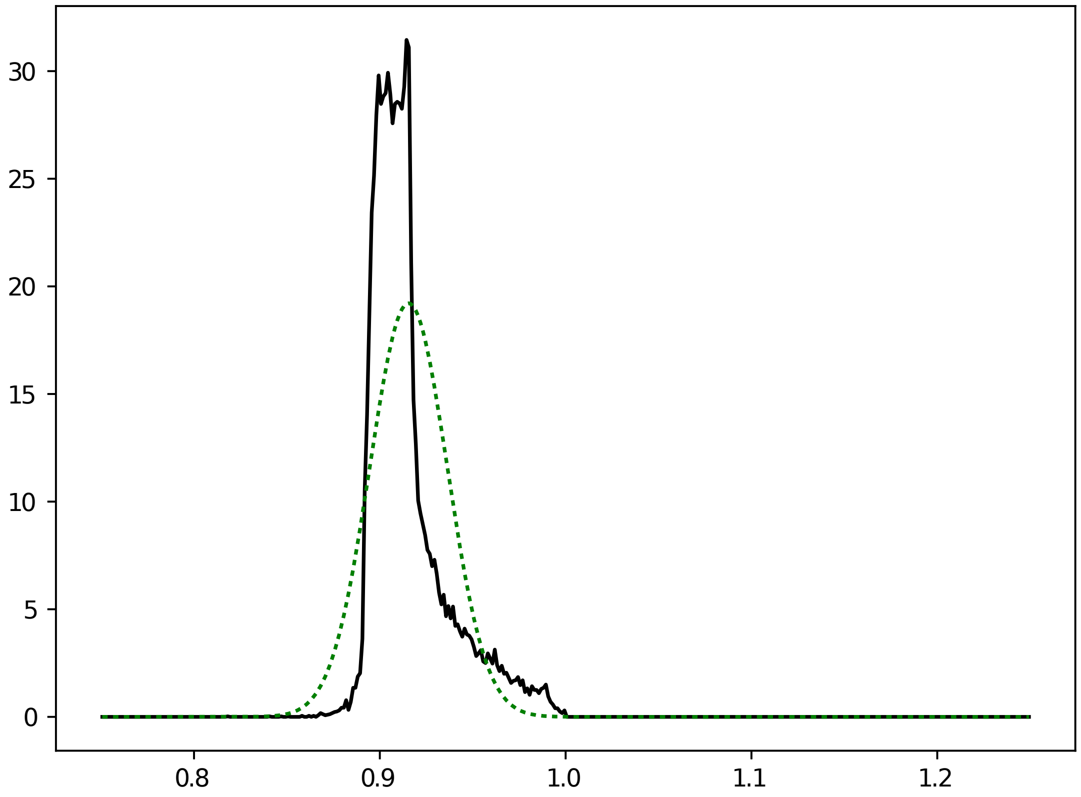


Figure 2.20: The distribution of the ratio of Rayleigh wave phase velocity to Love wave for all frequencies as a single distribution. The black line represents the measured distribution and the green dotted line represents the maximal likelihood estimator of the equivalent Gaussian distribution.

these numerical studies, a prior on the ratio of Rayleigh wave velocity to Love wave velocity can be approximated by a Gaussian with a mean of 0.915 and a standard deviation of 0.021.

2.6.1 A joint inversion approach

The approach taken here is to parameterise the Love dispersion curve as in the single inversion of the previous section, then assign a ratio variable for each control point. From this parameterisation, a Rayleigh dispersion curve can be computed from the Love dispersion curve and associated ratios. For a prior on ratio variable, the information gained from the numerical experiments presented in the previous section is used.

By incorporating this relationship between Love and Rayleigh dispersion curves, it is hoped that the inversion is stabilised further to be able to both resolve the multimodalities and better deal with spectral holes in the ambient noise data. It was previously demonstrated that Love and Rayleigh observations have different frequencies excited, for example in Figure 2.15 the Rayleigh wave observations (red curve) have reasonable signal at the lower frequencies whereas the Love wave observations have only middle frequencies excited. Using the prior information of the relative ratio of the Love and Rayleigh dispersion curves, the aim is to be able to improve the inversion by providing extra constraint, that is, the Rayleigh wave observations provide weak constraint for the Love wave dispersion where Love wave observations are lacking and vice-versa.

2.6.2 Results

All station pair data are inverted using Love and Rayleigh EGFs jointly. The process for the inversion is the same as that of for the independent inversions, that is 12 independent chains are run using parallel tempering with 4 temperatures logarithmically between 1 and 5, for 2,000,000 iterations with 500,000 removed as burnin. The results

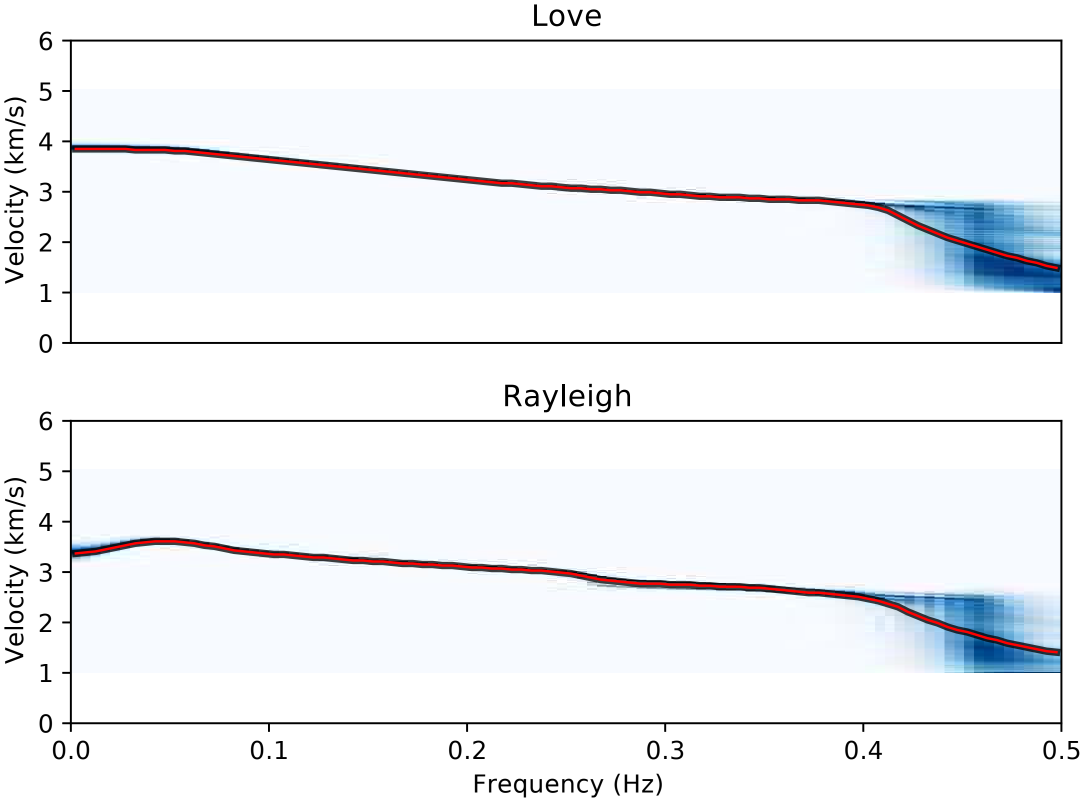


Figure 2.21: The results of the joint inversion of Rayleigh and Love dispersion for station pairs HOT15 and HOT20. The posterior probability is indicated with shaded blue (darker blue equates to more likely) and the mean of the ensemble is indicated with a red line.

for the HOT15—HOT20 observations are shown in Figure 2.21 with the fit shown in Figure 2.22. The previously identified problem at around 0.25Hz for the Rayleigh wave inversion has been remedied. The addition of the prior has meant that the multi-modality has been reduced. In Figure 2.22 the difficulty at around 0.25 Hz is clearly visible.

For the second example path, HOT23—HOT26, the posterior dispersion distribution is shown in Figure 2.23 and the corresponding data and Bessel function fit in Figure 2.24. In comparison to the independent results presented previous, the results have improved in that the multi-modality has been suppressed. The Love wave result has in parts the mode of the distribution lower than that for Rayleigh waves which is unlikely from the numerical experiments. Prior information is in general weak compared to observations and even though there is a tight Gaussian prior on the ratio of

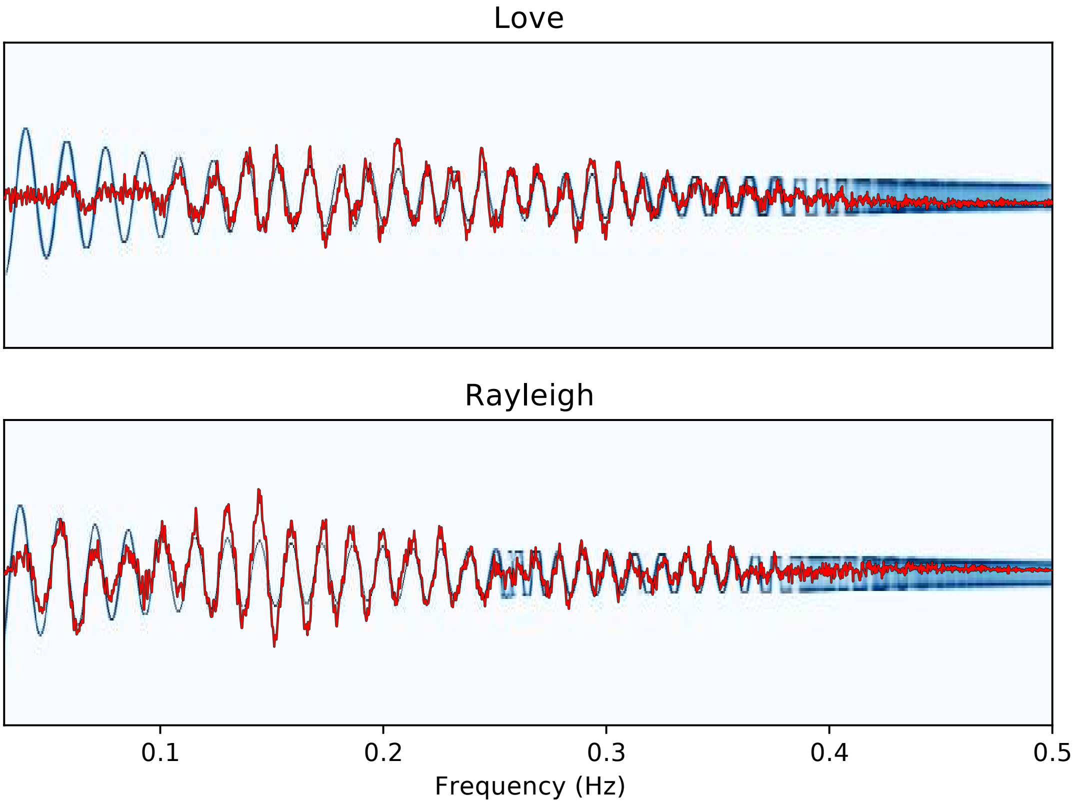


Figure 2.22: The original spectrum for station pair HOT15 and HOT20 is plotted in red with the posterior distribution of the Bessel function shown in shaded blue (darker blue equates to more likely).

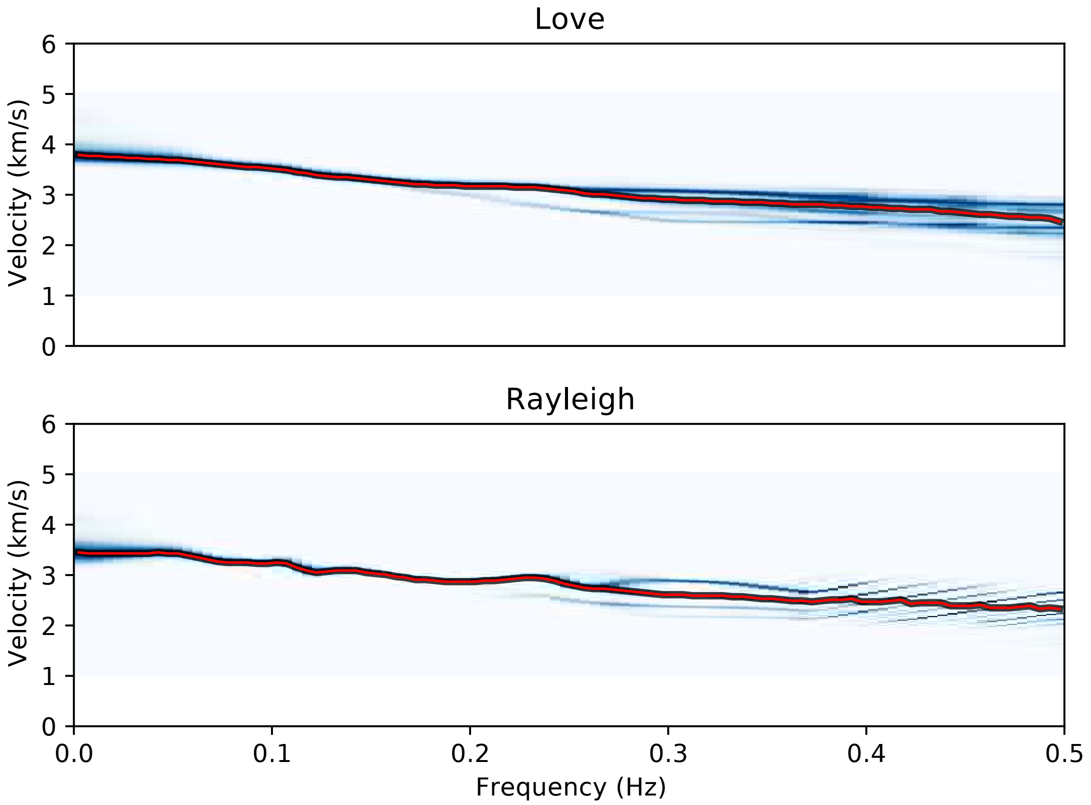


Figure 2.23: The results of the joint inversion of Rayleigh and Love dispersion for station pairs HOT23 and HOT26. The posterior probability is indicated with shaded blue (darker blue equates to more likely) and the mean of the ensemble is indicated with a red line.

Rayleigh to Love phase velocity, the data over powers it in this case. The results for the Rayleigh wave dispersion have similarly improved with reduced multi-modality over the frequency range of interest.

Once again, this second example is a difficult station pair with little signal. All ensemble median dispersion curves in Figure 2.25 and compared to the independent inversions, the clustering of the results shows less variance.

2.6.3 Comparison with manual group velocity

Fortuitously, for the set of seismic stations in this study, Love and Rayleigh wave dispersion curves have also been extracted for group velocity observations manually with the FTAN approach. It is interesting to compare for the manual process, where the

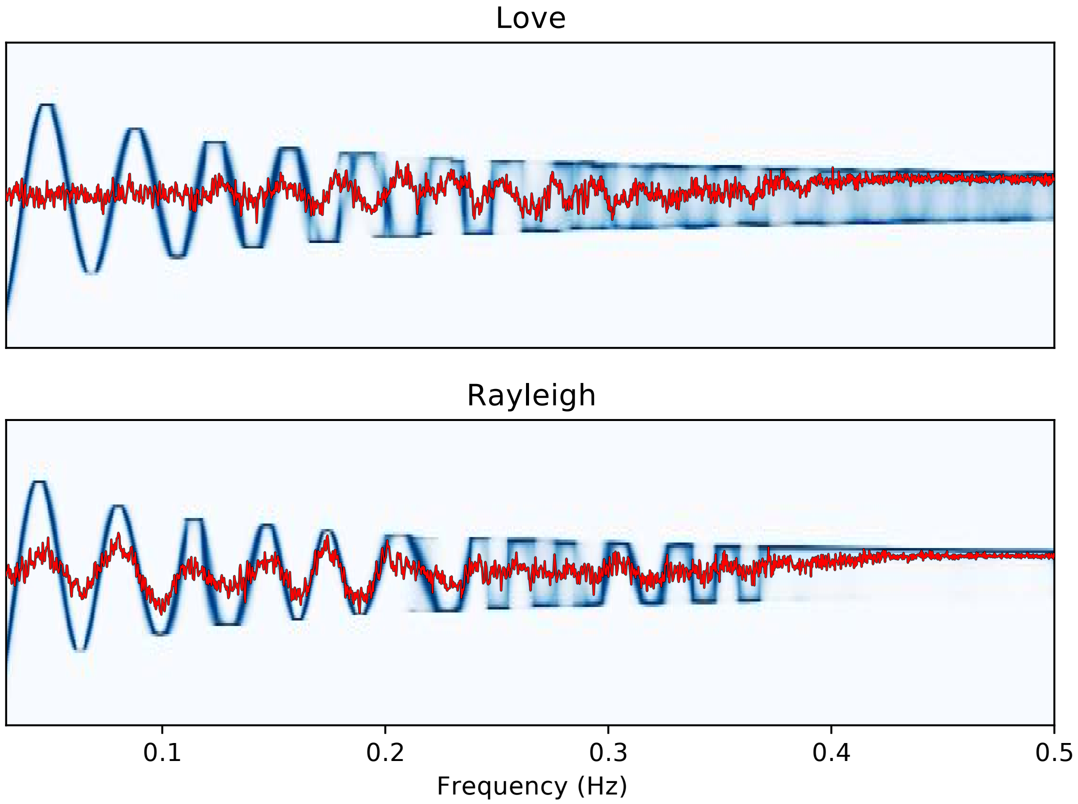


Figure 2.24: The original spectrum for station pair HOT23 and HOT26 is plotted in red with the posterior distribution of the Bessel function shown in shaded blue (darker blue equates to more likely).

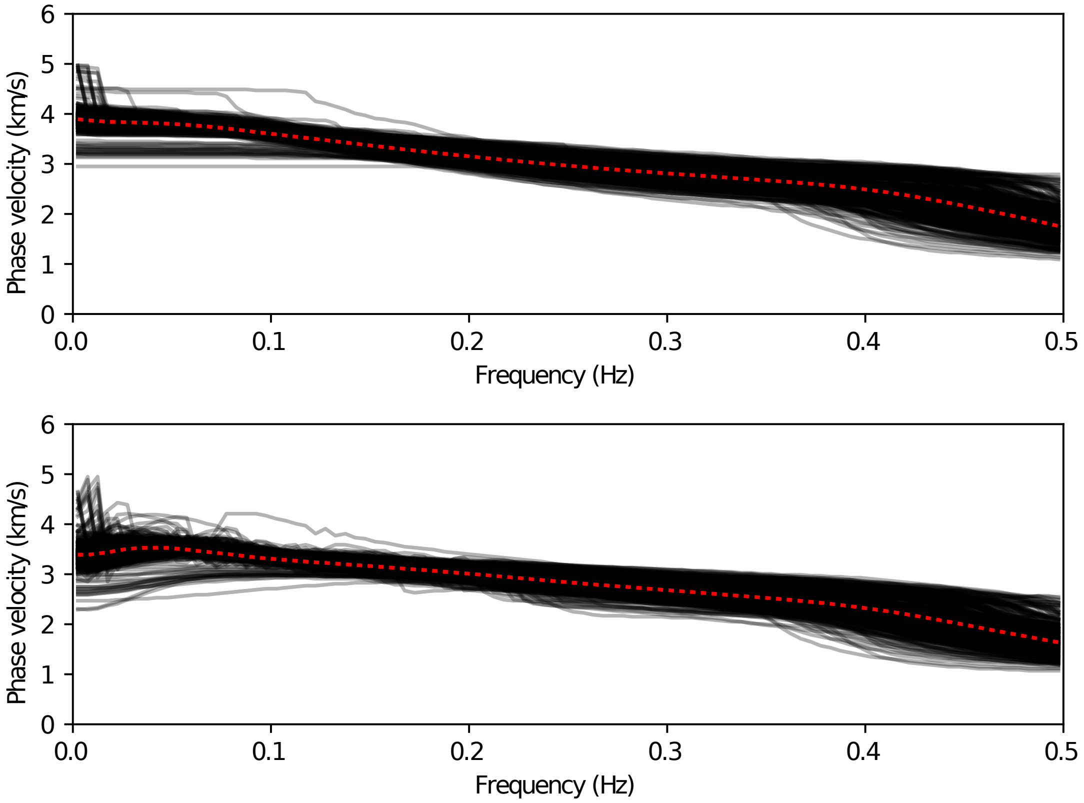


Figure 2.25: The median of all dispersion curves inverted using the joint dispersion approach. The results are better concentrated than those of the independent inversion in the previous section.

FTAN approach is used that has two limitations that reduce usable observations: first, the use of Gaussian taper filters blurs across frequencies rendering weak signals incoherent, particularly at longer periods for ambient noise. Secondly, it uses a far field approximation as FTAN was originally intended for earthquake driven surface wave studies. At longer periods the far field approximation means it is unreliable for station separations on the order of 100 km. These two factors combined mean that when using FTAN method, ray paths observations from stations pairs at longer periods are routinely eliminated.

To give an example of this reduction, in Table 2.1, the count of available station pairs from a manual extraction of group velocity observations for Rayleigh waves where there are 435 available station pairs are listed. The best recovered period is 17s where less that 70 percent of available ray paths are recovered. This number of recovered rays drops precipitously to less than half at longer periods.

In contrast, the method developed in this chapter recovers phase velocity across a frequency range as a continuous curve, with continuous uncertainty estimates. For every frequency, there are 435 observations. In the worst case, where there is poor constraint, consideration could be given to removing some observations using a criteria on the level of uncertainty. There is however no reason these "poorly constrained" dispersion observations couldn't still be used in a subsequent inversion for phase velocity maps.

It is possible to estimate group velocity from the phase velocity curves, although since the phase velocity is a piece-wise cubic curve, the group velocity is a function of a piece-wise quadratic curve. In Figure 2.26, an FTAN image obtained from group velocity processing for the HOT15—HOT20 station pair is shown with darker blue representing higher energy. Contiguous regions of darker blue are where a practitioner determining a group velocity dispersion curve would manually place points by eye.

Plotted in red is the median group velocity estimated from the method present here with 95 percent credible intervals shown with black dotted lines. There is very good agreement with the FTAN image where a practitioner would ideally place there disper-

Period (seconds)	Count	Percent (of possible 435)
1	241	55
2	202	46
3	225	52
4	275	63
5	267	61
6	265	61
7	265	61
8	247	57
9	228	52
10	209	48
11	222	51
12	237	54
13	259	60
14	266	61
15	279	64
16	282	65
17	296	68
18	292	67
19	272	63
20	258	59
22	229	53
24	206	47
26	186	43
28	167	38
30	154	35

Table 2.1: The count of ray paths available for subsequent group velocity map inversion from manual picking of group velocity using the FTAN approach and the percentage of available rays. A large amount of potential information is lost in this process, particularly at higher periods.

sion curve. At the higher frequencies, where the FTAN image becomes incoherent, the uncertainty in the group velocity estimated from the phase velocity inversion method becomes large indicating frequencies above approximately 0.4 Hz cannot be reliably determined here. At the lower frequencies, recovery of well constrained dispersion information is possible with the new method given the low uncertainties evident. In contrast, the FTAN image loses the signal at these lower frequencies because the signal is effectively Gaussian blurred in the frequency domain and with lower energy in the lower frequencies, the image becomes washed out.

In summary the method is consistent with more traditional manual estimation of group velocities, but the recovery is automatic, naturally includes uncertainty estimates, and is better able to resolve dispersion information at lower frequencies that is vital for resolving features at greater depths into the Earth.

2.6.4 Computational Time

The complexity involved in this inversion process comes at some cost. Each of the observations, 2 million iterations are simulated on a cluster computer. The forward model is relatively inexpensive with the largest cost the repeated calculation of the Bessel function. For each station pair, the computational time was between 45 minutes to an hour, which for 435 station pairs corresponds to a total time of 435 hours (taking the upper limit). This 435 hours, with sufficient computational resources available, can be run in parallel and equates to approximately a weekends worth of processing.

This seems a lot, but it is worth considering that previously such extraction of dispersion information was done manually and resulted in a limited set of group velocity observations at certain periods. In contrast, the method presented here automatically extracts phase velocity dispersion curves for both Love and Rayleigh waves as continuous functions of frequency that are reliable over a broader frequency range and naturally come with uncertainty estimates.

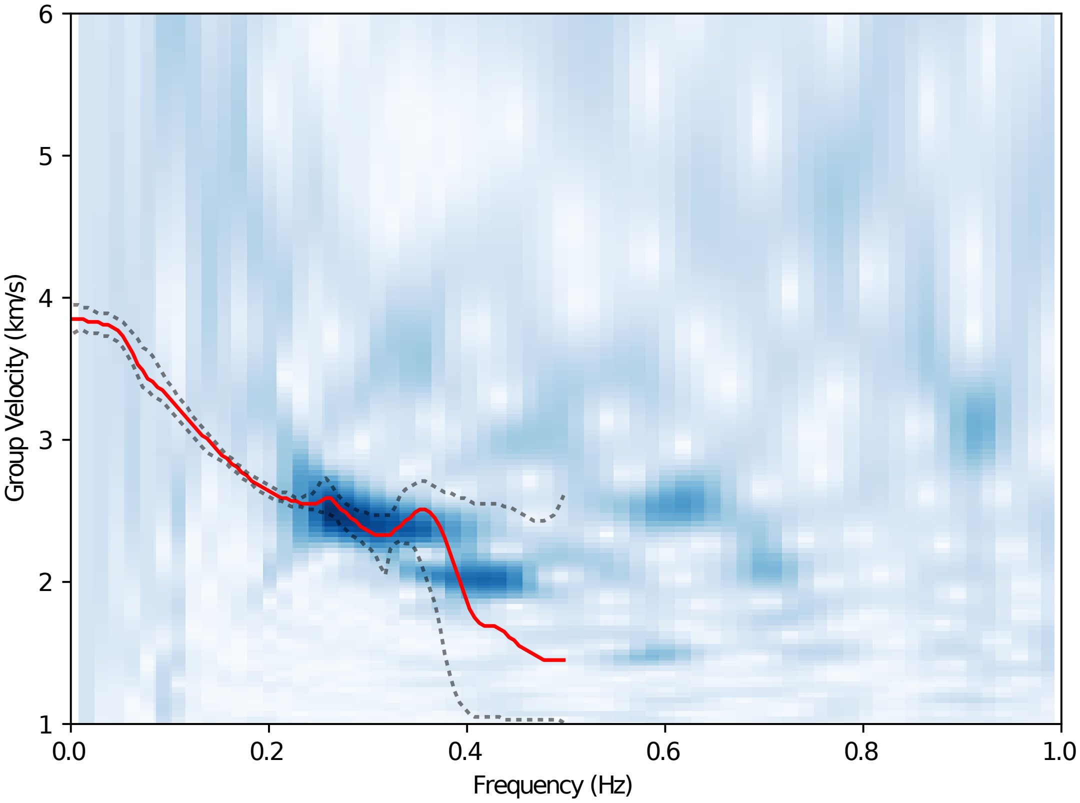


Figure 2.26: Plotted in blue shading is the signal energy computed using the FTAN approach. In red is plotted the median of the ensemble group velocity computed from the inverted phase velocity curves and there is good agreement in regions of higher group velocity energy (darker blue). At higher frequencies, the phase velocity is poorly constrained matching approximately the region where group velocity energy dissipates. At low frequency there is well constrained group velocity estimates where the FTAN method fails to detect a coherent signal.

2.7 Summary

This chapter has presented an approach to robustly extract surface wave dispersion information jointly for Love and Rayleigh ambient noise data. The Bayesian approach allows minimal pre-processing of the data and robust prior information to effect a stable and precise estimate of dispersion with uncertainty estimates that can be used for subsequent tomographic inversion.

In this study the use of focusing parameters in the Dirichlet priors has not been addressed so this could be further explored either with α in the Dirichlet priors with fixed values greater than one or the use of a hierarchical prior on this parameters.

The focus of this study has been on phase velocity alone however the model parameterisation permits easy calculation of the group velocity which could be either used as additional constraint in the inversion by incorporating traditional Frequency-Time Analysis (FTAN) [Dziewonski and Hales, 1972] in the likelihood. Alternatively the group velocities can be estimated from the posterior phase velocity curves as in Figure 2.26. While the group velocity estimate obtained from the phase velocity curves is in good agreement with results obtained with other methods, there is room for improvement of the group velocity results. These could be improved by raising the order of the polynomial further while preserving monotonicity [Dougherty et al., 1989] to enable smoother group velocity estimates.

A simple independent Gaussian noise model with a base level of noise estimated from a quiescent part of the real spectrum. An obvious extension would be to use estimates of covariance errors in the spectrum which could be done per station pair or across the entire array.

Lastly the optimal value of the Bessel function scaling to fit the observed real spectrum could be inverted for as part of the inversion. However, a more interesting approach would be to attempt to recover the amplitude envelope of the spectrum similar to the recovery of the source time function in Dettmer et al. [2015]. This could provide additional information, such as frequency dependent attenuation, that could be used in subsequent tomographic inversion.

Chapter 3

Trans-dimensional Trees

3.1 Introduction

In the previous chapter a trans-dimensional inversion was developed for a one dimensional geophysical problem. Extending trans-dimensional inversion to higher dimensions has typically involved the use of Voronoi cells [Okabe et al., 1992, Samet, 2006] and at first glance this would appear well suited to trans-dimensional geophysical inversions as these cells have a long history in large scale geophysical inversion problems, for example Sambridge and Gudmundsson [1998].

Using Voronoi cells, by specifying the location of the nuclei as well as the value (or values) of Earth properties within each cell, a mobile Voronoi model can be used to represent Earth properties spatially in 2D [Bodin et al., 2012a]. In the first 3D application, Piana Agostinetti et al. [2015] have recently extended the Voronoi cell approach to local earthquake tomography. These Voronoi cell parameterisations are grid free and locally adapt to regions of increased heterogeneity tempered by the resolving power of the data. Although the application of the trans-dimensional Voronoi cell method is now well established for seismic imaging, there are a number of short comings that hinder its application as the number of data and complexity of the Earth model increases.

In ray based seismic tomography, numerical integration along ray paths requires the evaluation of the model at hundreds to thousands of spatial points per observation. For each point along the ray the Voronoi cell parameterisation of the Earth properties is needed, for example, seismic wave speed, and this involves determining in which cell the point resides. A naive algorithm would simply find the nearest Voronoi nuclei by computing the distance to every nuclei of the model and this results in an $\mathcal{O}(n)$ operation, where n is the number of Voronoi cells. In 2D problems, a Delaunay triangulation can be used to speed up the cell look up operation to an $\mathcal{O}(\log n)$ operation [Sambridge and Gudmundsson, 1998]. Even with fast algorithms for incrementally maintaining the Delaunay triangulation [Lawson, 1977], the accounting cost of maintaining the triangulation can be prohibitive as the number of cells increases.

A second feature of the Voronoi cell approach is that they do not lend themselves well to representing a continuous field. In a Voronoi cell parameterisation, the Earth properties within each cell are often represented with constant values, although in principle, any order polynomial is possible. This means that each Earth model consists of an irregular polygonal mesh with discontinuities, both in the function and in its derivatives, at the interfaces between cells. Typically, any single Earth model in the ensemble is rather crude and implausible and it is only by averaging over many such crude representations that it is possible to generate a continuous field. This means that the Voronoi cell approach must utilise multiple independent Markov chains or very large numbers of samples in a single chain in order to produce a continuous field through spatial averaging.

Use of Voronoi cells in 3D imaging has two additional complications. The first is that there is no analogue of fast 2D incremental Delaunay calculation algorithms [Sambridge et al., 1995] and so Voronoi cells must be determined from "scratch" each time the mesh is updated, further adding to the computational burden. The second is that the shape of Voronoi cells in 3D is particularly sensitive to the choice of spatial scaling between lateral and radial directions. For example, Voronoi cells built around nuclei at depth can easily protrude to the surface.

In this chapter a new class of parameterisation for trans-dimensional imaging problems is introduced which overcomes the limitations of Voronoi cells while providing a general efficient framework for dealing with 1D, 2D and 3D problems in Cartesian or spherical geometries. The new framework allows a great deal of flexibility in terms of the choice of basis functions, including multi-scale parameterisations such as wavelets and sub-division surfaces. Due to these new efficiencies and flexibility, with the new algorithm, trans-dimensional inversion of larger scale 3D tomographic problems are more tractable.

3.2 Trans-dimensional trees

Before introducing the trans-dimensional framework for sampling over trees, it is instructive to show how the concept of trees can be used to represent a tomographic Earth model. There are many examples of using hierarchical or multi-resolution analyses of images in 2D, for example the Laplacian pyramid [Burt and Adelson, 1983] and the wavelet transform [Mallat, 1989]. Broadly speaking, within each of these schemes an image is sub-sampled to obtain a coarser but more compact representation. "Corrector" terms are computed representing the difference between the sub-sampled and true image so that with a combination of the sub-sampled predictor image and corrector terms, the original image can be accurately reconstructed. This process can be repeated recursively on each sub-sampled image until the result is a single pixel, representing the mean of the image, and a hierarchical set of corrector terms for each resolution scale. It is a property of continuous tone images that individual pixels are often highly correlated with their neighbours, and as a result, many of the corrector terms are near zero, that is the lower resolution image is a good predictor. For this reason, such multi-resolution image analysis techniques have been used for image compression, for example, the JPEG 2000 image compression standard [Unser and Blu, 2003].

This hierarchy of a single mean value of an image through successive levels of perturbative terms can naturally be represented by a tree structure. Figured 3.1(a) shows how a quaternary tree in which each node has 4 child branches, spans from the single pixel representation of an average value of a field, through successive levels of local perturbation terms. In this example, each node of the tree contains a single parameter value. At the root of the tree, the highest level in Figure 3.1(a), this value represents the mean of the image and all other descendant nodes represent a local deviation from that mean. In this way, each tree level creates an image with a corresponding spatial resolution and each child node adds detail by perturbing the previous level at a finer spatial resolution. From this multi-resolution tree representation, arbitrary 2D images can be constructed and can be used to represent, for example, the seismic wave speed or slowness of a region of the Earth. This same principle, of spanning the subdivision

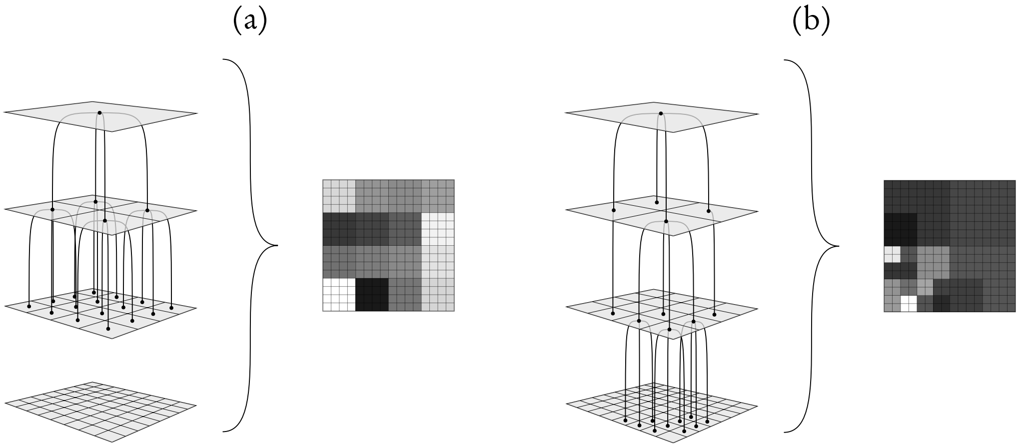


Figure 3.1: The Laplacian Pyramid subdivision showing how a quaternary tree can span from the coarsest resolution to the finest error terms. At the top level there is a single pixel representation of a 2D domain, the root node of the tree, which is subdivided into four sub-pixels at the next level and so on. In (a) is the complete tree structure to the 3rd level. In (b) is an incomplete quaternary tree that can still be used to parameterise a 2D Earth model demonstrating how the tree can locally adapt to regions of localised heterogeneity. In both (a) and (b), the two models have the same number of parameters but represent very different structure.

of grid with a tree, applies equally to 1D, 2D, 3D Cartesian geometries and equally to non-Cartesian geometries such as spherical geometry [Samet, 2006].

It is important to point out here that the tree needn't completely span the underlying 2D grid as shown in the Figure 3.1(a). An incomplete tree is shown in Figure 3.1(b). The 2D image from this tree is constructed in the same way as the full tree with the parameter values of zero at the missing nodes of the tree. This has the potential to compress the model space, or the number of model parameters, by locally adapting to structure or data coverage.

The use of adaptive mesh refinement has been used previously in geophysical inversion, for example Sambridge and Faletič [2003], where a criterion based on the maximum spatial gradients in seismic velocity perturbation was used to iteratively subdivide a tetrahedral grid during the inversion. A similar approach was presented by Plattner et al. [2012] for electrical resistivity tomography where a multi-scale wavelet parameterisation was adaptively refined through optimisation.

Rather than a fixed criterion, the trans-dimensional trees sample over such subdivision refinement choices to obtain posterior information on where the data requires finer scale features. By itself, recasting geophysical inverse problems within a tree structure offers little advantage, but as will be demonstrated, it is highly suited to coupling with a trans-dimensional algorithm within a fully Bayesian framework.

3.3 A General Bayesian Trans-dimensional Framework for Trees

In a Bayesian approach to inference, the solution we obtain is a numerical estimate of the *a posteriori* probability distribution or posterior (see Gelman et al. [2004] for a general overview and Mosegaard and Tarantola [1995], Sambridge and Mosegaard [2002] for an overview of Bayesian inference in a geophysical context). This is the probability density of the model space given the observed data, or written mathematically, $p(\mathbf{m}|\mathbf{d})$, where $\mathbf{m}$ is the vector of model parameters and $\mathbf{d}$ the vector of observations. In all but the simplest of problems, this probability density function is approximated numerically using Markov chain Monte Carlo (McMC) techniques and Bayes theorem [Bayes, 1763], i.e.

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{m})p(\mathbf{d}|\mathbf{m})}{p(\mathbf{d})}. \tag{3.1}$$

This states that the posterior probability density, $p(\mathbf{m}|\mathbf{d})$, is equal to the prior probability distribution, $p(\mathbf{m})$, times the likelihood $p(\mathbf{d}|\mathbf{m})$, which is abbreviated to $p(\mathbf{d}|\mathbf{m})$, normalised by the evidence, $p(\mathbf{d})$. An McMC sampling approach can be applied to the numerator of the right hand side of (3.1) to obtain an estimate of the posterior probability distribution up to the normalising constant of the evidence, which is often difficult to compute explicitly [Sambridge et al., 2006].

An McMC sampler requires the specification of the prior probability distribution, which represents *a priori* information that may be available for the distribution, or plausible range, of the model parameters, and the likelihood which is a probabilistic measure of the fit of the model to the data. An McMC sampler operates by starting from some model at step i of $\mathbf{m}_i$, then creating a new proposed model $\mathbf{m}'_i$ using a proposal in the form of a reversible probability density function $Q(\mathbf{m}_i \to \mathbf{m}'_i)$. The new model is accepted, that is $\mathbf{m}_{i+1} = \mathbf{m}'_i$ or rejected, $\mathbf{m}_{i+1} = \mathbf{m}_i$, based on an acceptance probability, commonly the Metropolis-Hastings acceptance criterion [Metropolis et al., 1953, Hastings, 1970]

$$\alpha(\mathbf{m} \to \mathbf{m}') = \min\left\{1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m}' \to \mathbf{m})}{Q(\mathbf{m} \to \mathbf{m}')}\right\}. \tag{3.2}$$

This can be read as the prior ratio times the likelihood ratio times the proposal ratio. The Metropolis-Hastings criteria satisfies the mathematical condition known as "detailed balance" [Gamerman and Lopes, 2006] which allows the Markov chain to converge and correctly sample the target posterior distribution.

It is common practice to remove some initial number of steps from the final ensemble which are believed to be pre-converged or "burn-in" samples. In most cases, the fact that only the posterior probability distribution up to a normalising constant is obtained is not a problem as relative inferences are generally sufficient.

An extension to McMC samplers is the Birth/Death scheme of Geyer and Møller [1994], generalised to the trans-dimensional framework developed by Green [1995]. In trans-dimensional samplers, a proposal distribution is allowed to change the parameterisation of the model and dimension, that is the size of the vector $\mathbf{m}$ of model parameters. A key benefit of allowing the sampling to jump between dimensions is that the data dictates the model complexity giving in a parsimonious result [Malinverno, 2002, Sambridge et al., 2006]. Additionally, the posterior probability distribution is available on the number of model parameters required by the data given the noise rather than fixing this *a priori*.

The generalisation of the Metropolis-Hastings acceptance criteria to support trans-dimensional steps is

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = \min \left\{ 1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m}' \rightarrow \mathbf{m})}{Q(\mathbf{m} \rightarrow \mathbf{m}')} \left| \mathcal{J} \right| \right\}, \tag{3.3}$$

where the additional term from (3.2), $|\mathcal{J}|$, is determinant of the Jacobian that maintains detailed balance through variable transformations resulting from trans-dimensional steps. Expression (3.3) may also be used if the dimension is unchanged, but the proposal involves a step from one class of parameterisation to another.

The complexity of the models generated from trans-dimensional samplers is dependent on the level of noise applied, that is in general, the lower the noise, the higher the complexity. For this reason, in the case where the noise on the data is unknown or estimated, it is advantageous to use a hierarchical Bayesian step that allows to some extent noise parameters to be inverted for as part of the sampling of model parameters [Malinverno and Briggs, 2004].

A birth/death trans-dimensional sampler will consist of three classes of proposal, a birth proposal where the model vector **m** will increase in size, a death proposal where some model parameters are removed, and a value proposal where the model vector remains the same size, but one or more values will be changed (i.e. the normal class of proposal in fixed dimension McMC samplers).

The aim here is to apply the trans-dimensional framework to the McMC sampling of tree structures that can be used to represent geophysical models of the Earths internal structure. In this framework, a birth proposal will consist of adding one or more new nodes to the tree, a death proposal will consist of removing one or more nodes from the tree, and a value proposal will perturb one or more values located within the existing tree. In the literature, there has been no general treatment of trans-dimensional sampling over tree structures previously presented. Other similar work is that of Denison et al. [1998] which is limited to binary classification trees. This chapter applies the trans-dimensional formalism of Green [1995] to general trees with known structure.

A "general" tree as one in which the maximum number of child nodes, of any node, is fixed. With this restriction a prior can be computed. Rarely in practice would this restriction limit the application of this new framework. In general, the structure of the tree will be restricted by the geometry of the physical application. For example, in the 2D image example earlier, each pixel is subdivided into 4 sub-pixels and this is the upper limit on the number of child nodes. For a 3D volume, each voxel will subdivide into 8 sub-voxels which gives an upper limit on the number of child nodes of 8.

In the following subsections the components of the acceptance criteria will be described and the full general expressions for each type of model perturbation will be derived.

3.3.1 The Model

In the earlier 2D example, the tree structure "template" consists of the complete spanning quaternary tree and two possible tree models conforming to this template are shown in Figure 3.1. A simpler example of such a tree model in a binary tree template appears in Figure 3.2 where the template is shown in outline and an example tree model, consisting of active nodes and value(s) at each node, is shown in solid shading.

Each active node in the tree model has one or more associated values, so given a number of nodes, k, the model space vector would be

$$\mathbf{m} = \langle \mathscr{T}_k, \mathbf{v}_1, \ldots, \mathbf{v}_k \rangle , \tag{3.4}$$

where $\mathscr{T}_k$ represents the arrangement of the k nodes within the template tree structure and $\mathbf{v}$ is the vector of parameters at each node (which may be a single parameter). If there is a unique index for each tree node, $\mathscr{T}_k$ can be represented as a set of indices, that is $\mathscr{T}_k = \langle t_1, \ldots, t_k \rangle$.

3.3.2 The Prior

Given the parameterisation in (3.4), the prior on the model can be written in general terms as a product of conditional probability distribution functions (PDFs),

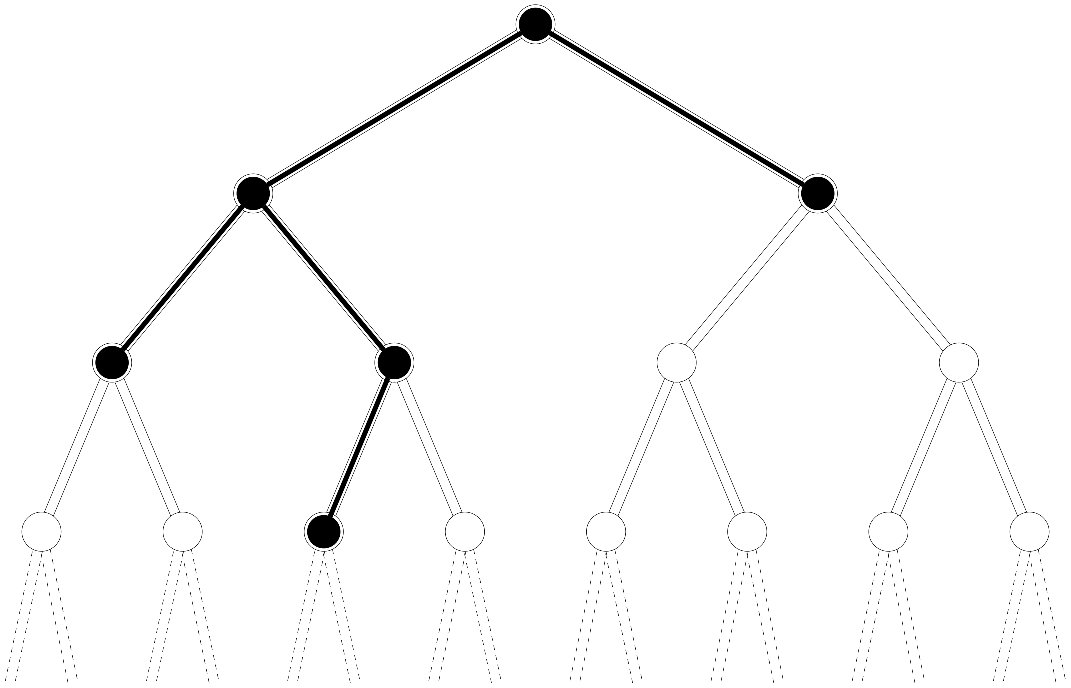


Figure 3.2: The first four levels of a binary tree template shown as outline with an individual tree model highlighted with solid lines.

$$p(\mathbf{m}) = \left[\prod_{i=1}^{k} p(\mathbf{v}_i \mid \mathcal{T}_k, k)\right] p(\mathcal{T}_k|k)p(k). \tag{3.5}$$

Stated simply, the prior is a combination of the probability on the number of nodes in the tree, the probability of the arrangement of the tree within its template and the parameter values at each of the nodes. This prior specification reasonably assumes that each term is independent which results in a separable prior probability distribution function.

Prior on the number of nodes

The prior for the number of nodes is a choice that will be dependent on how the model is mapped from the tree structure. Here the prior is left as a general expression, $p(k)$, but highlight two common choices. Firstly a uniform prior

$$p(k) = \frac{1}{k_{max} - k_{min} + 1}, \tag{3.6}$$

where k_{max} and k_{min} (usually 1) are chosen as the upper and lower bounds on the number of nodes. An alternative is to use a Jeffreys' prior Jeffreys [1939], Jaynes [2003], that is

$$p(k) \propto \begin{cases} \frac{1}{k} & k > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{3.7}$$

This prior is improper because the limit of the integral of $p(k)$ is unbounded as k goes to infinity. Nevertheless, a useful feature is that there is no imposed restriction on the number of nodes unlike in the uniform prior case (see page 238 of Jeffreys [1939]). In experiments to be described here, the posterior PDF of k with either prior is similar, which shows that it is primarily the data which constrains the dimension of the model.

Prior on homogeneous unrestricted trees

The prior on the arrangement of the nodes within the tree template, $p(\mathcal{T}_k|k)$, is the most complicated component of this algorithm and is derived here for the general tree parameterisation. The prior used is a uniform prior on the structure of the tree, this means that given a number of nodes, k, any arrangement of the nodes into a valid tree within its template has equal probability to any other. This is the least informative prior on a tree structure and also the most tractable to compute for the acceptance criteria. The consequences of this prior are that a model that has an even distribution of detail across the region is equally as likely as a model that has localised fine detail. This is illustrated in Figure 3.1 where both models shown have the same number of active tree nodes. In this prior, both of these models are equally likely.

This reduces computing the prior on the structure of the tree into a problem of computing the number of valid tree arrangements possible given a tree structure template and the number of active nodes to form a uniform prior

$$p(\mathcal{T}_k \mid k) = \frac{1}{\mathcal{N}_k}, \tag{3.8}$$

where $\mathcal{N}_k$ is the number of valid trees with k nodes. To evaluate $\mathcal{N}_k$, unrestricted homogeneous trees are first considered, which are defined as those where each node has the same upper limit on the number of child nodes. Binary and quaternary trees fall into this class. Unrestricted means that there are no other constraints on the structure of the tree such as a maximum height and therefore that the tree can grow infinitely large. For this class of trees, there are analytical expressions for computing the number of arrangements, $\mathcal{N}_k$. For binary trees it is known that the number of arrangements follows the sequence of Catalan numbers [Catalan, 1844, Hilton and Pedersen, 1991, Knuth, 2004]

$$\mathcal{N}_k = \frac{1}{k+1}\binom{2k}{k}, \tag{3.9}$$

where $\binom{2k}{k}$ is the standard binomial coefficient. This result has been generalised by Aval [2008] to trees with a maximum number of n children

$$\mathcal{N}_k = \frac{1}{(n-1)k+1}\binom{nk}{k}. \tag{3.10}$$

When $n = 2$ this reduces to (3.9). This expression allows closed form expressions for the prior for homogeneous unrestricted trees. However this only represents a small sub-class of possible trees and this needs to be extended further.

Restricted and heterogeneous trees and their priors

The first restriction on a tree template is an upper limit on height. As seen in the earlier 2D example, the height of the tree corresponds to the level of subdivision of the region. As such, a restriction on the height of the tree imposes a strict upper limit on the minimum resolution scale of the model. In addition, it also constrains the computational complexity of the problem as arbitrarily large trees are no longer possible.

A second variant to be considered is a heterogeneous tree which contains nodes with varying upper limits on the number of child nodes. In later examples, where wavelet parameterisations are used in 2D problems, heterogeneous trees are used where the root of the tree has 3 possible child nodes, and all subsequent nodes have 4 possible branches. Analytic expressions for the number of arrangements of a tree given the number of nodes are only known for trees where each node has the same maximum number of possible child nodes. For both the restricted height and heterogeneous trees, the number of arrangements given k needs to be calculated.

The Catalan sequence for binary trees can be derived from a recurrence relationship using generating functions (see Equation 2.5.10 of Wilf [1990]). The general solution to both these problems is to compute the number of arrangements from a recurrence relationship. Starting from the recurrence relationship for binary trees,

$$\mathcal{N}_k = \begin{cases} 1 & k \leq 0 \\ \sum_{i=0}^{k-1} \mathcal{N}_i \mathcal{N}_{k-i-1} & \text{otherwise} \end{cases}, \tag{3.11}$$

where k is the number of tree nodes, which is a simple integer partitioning problem [Stanley, 1997]. The modification of (3.11) from a binary tree to a ternary tree requires the addition of a third partitioning of the k nodes among the three child branches. To include restrictions on the height of the tree, the relevant terminating conditions need to be added, for example, rewriting the equation

$$\mathcal{N}_{k,h} = \begin{cases} 0 & h = 0 \\ 1 & k \leq 0 \\ \sum_{i=0}^{k-1} \mathcal{N}_{i,h-1} \mathcal{N}_{k-i-1,h-1} & \text{otherwise} \end{cases}. \tag{3.12}$$

Further details of the recurrence relationships and the algorithm developed in this study for computing them in an efficient fashion is outlined in the following section. From here on, $\mathcal{N}_{k,h}$ is assumed to be known from a recurrence relationship like (3.12) and that the prior on the structure of the tree can be calculated as the inverse of the number of arrangements of trees given a number of nodes

$$p(\mathcal{T}_k \mid k, h) = \frac{1}{\mathcal{N}_{k,h}}, \tag{3.13}$$

where h is a maximum height restriction.

Counting arrangements of general trees

In the previous section, the tree structure prior requires the calculation of the number of tree arrangements possible given a restriction the k, the number of active tree nodes, and the height of the tree. Recall that the recurrence relationship for computing the number of arrangements in a binary tree, from (3.11), is

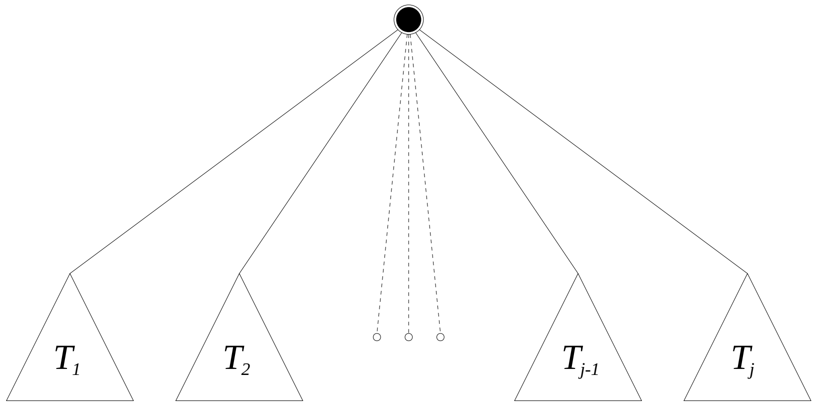


Figure 3.3: An abstract tree node with j sub-trees.

$$\mathcal{N}_k = \begin{cases} 1 & k \leq 0 \\ \sum_{i=0}^{k-1} \mathcal{N}_i \mathcal{N}_{k-i-1} & \text{otherwise} \end{cases} . \tag{3.14}$$

This can be extended to a ternary tree, or a tree in which every node has 3 possible children, as follows:

$$\mathcal{N}_k = \begin{cases} 1 & k \leq 0 \\ \sum_{i=0}^{k-1} \mathcal{N}_i \left[\sum_{j=0}^{k-1-i} \mathcal{N}_j \mathcal{N}_{k-i-j-1} \right] & \text{otherwise} \end{cases} . \tag{3.15}$$

In generalising this further, it should be recognised that this is essentially a restricted integer partitioning problem [Stanley, 1997], or stated simply as how many ways can an integer number of nodes be distributed among some arbitrary number of sub-trees. In Figure 3.3, a general tree node is shown with j possible sub-trees labelled $T_1 \ldots T_j$. It should be noted that each of these sub-trees may have a different structure, that is, a different limit on the number of child nodes at the next level, to each other and to the parent tree. From this generalisation, any tree structure can be constructed.

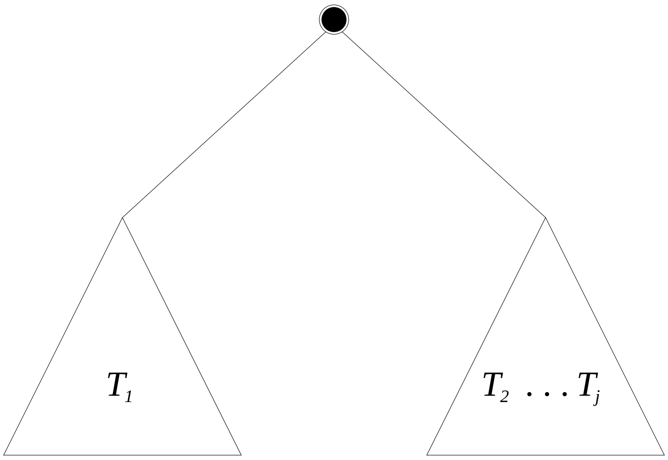


Figure 3.4: Rearrangement of the sub-trees into a binary tree structure by amalgamation $j-1$ *right most sub-trees.*

By grouping the sub-trees appropriately, any number of sub-trees can be reformulated into an expression of the same form as the binary tree case by treating sub-tree T_1 as itself and sub-trees $T_2 \ldots T_k$ as an amalgamated collection of sub-trees. This is shown graphically in Figure 3.4.

Alternatively, when j, the number of sub-trees, is even, the sub-trees can be split evenly into 2 amalgamated collection of sub-trees as shown in Figure 3.5.

In either of these cases where multiple sub-trees are amalgamated into 2 super-sub-trees, if these sub-trees are labelled $\mathscr{A}$ and $\mathscr{B}$, the recurrence relationship can be rewritten

$$\mathscr{N}_k = \begin{cases} 1 & k \leq 0 \\ \sum_{i=0}^{k-1} \mathscr{A}_i \mathscr{B}_{k-i} & \text{otherwise} \end{cases} . \tag{3.16}$$

Note that there is a small difference between this equation and Equation 3.14 in that the number of nodes partitioned to the right branch is $k-i$ rather than $k-i-1$, that is $\mathscr{B}_{k-i}$ instead of $\mathscr{N}_{k-i-1}$. Since the tree is effectively split in two and the left and right

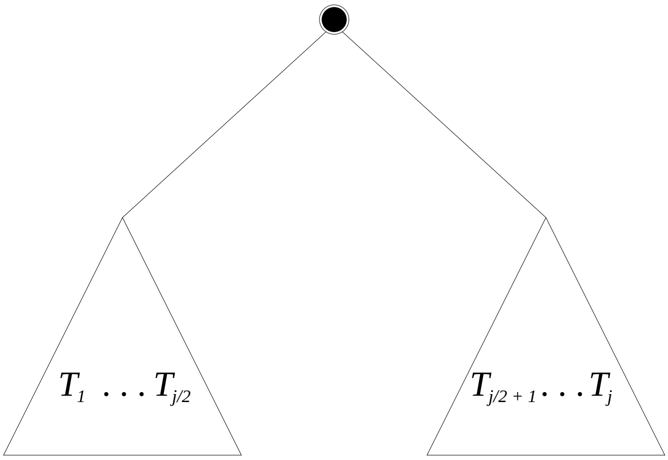


Figure 3.5: Rearrangement of an even number sub-trees into a binary tree structure by an even amalgamation of the the sub-trees.

sides computed, this prevents the root of the tree being counted twice.

For trees or sub-trees with some restriction, for example a restriction on the height, this can be enforced by adding an extra restriction in the recurrence relationship such that

$$\mathcal{N}_k = \begin{cases} 0 & k > k_{max} \\ 1 & k \leq 0 \text{ or } k = k_{max} \\ \sum_{i=0}^{k-1} \mathcal{A}_i \mathcal{B}_{k-i} & \text{otherwise} \end{cases}, \tag{3.17}$$

where k_{max} represents the maximum number of nodes of the current sub-tree. This can be computed recursively using

$$k_{max} = 1 + k_{max}(\mathcal{A}) + k_{max}(\mathcal{B}). \tag{3.18}$$

In all presented thus far, the k_{max} is specified as a height restriction on the tree so that for some sub-trees, that is those with a fixed number of child nodes, an analytical expression can be used to compute the maximum number of nodes. For other trees and sub-trees, these are generally constructed piece wise from generic trees and it is therefore easy and efficient to compute the maximum number of nodes recursively.

A general algorithm for computing the number of arrangements of trees can now be constructed. The first point is that the algorithm incrementally computes the number of arrangements for a given k rather than for all values of k. Secondly, the results of previously computed k in the each sub-tree and the full tree are memoized. The memoize operation is a method of re-using previously computed results, so when some computation is memoized, the first time it is actually computed and every other time it is simply a look-up operation in a stored list of results. For recurrence relationship computations, this is vital to speed up the computation as the same partial results are frequently required. The novel algorithm for this is shown in Algorithm 1.

To give an appreciation of the need to use such an algorithm for computing the number of arrangements, the time for computing the number of arrangements for k equal 1 to 100 for the trees used in the 2D wavelet parameterisation in Section 3.7 was measured. For a naive algorithm, this takes approximately 148 minutes to compute and with the recursive memoization algorithm presented here, the same range of numbers can be computed in approximately 6ms, that is, over a million times faster.

Prior on each parameter value

The prior on the Earth model parameters at each node of the tree will depend the particular basis functions used. Again this prior is often a choice and some alternatives are covered here. The simplest prior is a uniform prior which constrains the parameter values to be within an upper and lower bound. It has been shown that the distribution of wavelet coefficients for a set of representative continuous images follows a generalised Gaussian distribution [Antonini et al., 1990, 1992] suggesting that a generalised Gaussian distribution may be a suitable prior for wavelet based parameterisations. For Bayesian approaches to wavelet based Compressive Sensing, “spike and slab” priors

Algorithm 1 Algorithm for computing the number of tree arrangements

```
$\mathscr{D} = \varnothing$
function MemoizeArrangements(Tree $\mathscr{T}$, Integer $k$)
    if $k = 0$ or $k = k_{max}(\mathscr{T})$ then
        return 1
    end if
    if $k < 0$ or $k > k_{max}(\mathscr{T})$ then
        return 0
    end if
    if $(\mathscr{T}, k) \notin \mathscr{D}$ then
        $j \leftarrow$ NumSubTrees($\mathscr{T}$)
        if $j = 1$ then
            $\mathscr{A} \leftarrow$ Subtrees($\mathscr{T}$, 1, 1)
            $\mathscr{D}(\mathscr{T}, k) \leftarrow$ MemoizeArrangements($\mathscr{A}$, k)
        else if $j$ mod 2 = 1 then
            $\mathscr{A} \leftarrow$ Subtrees($\mathscr{T}$, 1, 1)
            $\mathscr{B} \leftarrow$ Subtrees($\mathscr{T}$, 2, $j$)
            $\mathscr{D}(\mathscr{T}, k) \leftarrow$ ComputeSubtrees($\mathscr{A}$, $\mathscr{B}$, $k$)
        else
            $\mathscr{A} \leftarrow$ Subtrees($\mathscr{T}$, 1, $j/2$)
            $\mathscr{B} \leftarrow$ Subtrees($\mathscr{T}$, $j/2+1$, $j$)
            $\mathscr{D}(\mathscr{T}, k) \leftarrow$ ComputeSubtrees($\mathscr{A}$, $\mathscr{B}$, $k$)
        end if
    end if
    return $\mathscr{D}(\mathscr{T}, k)$
end function
function ComputeSubtrees(Tree $\mathscr{A}$, Tree $\mathscr{B}$, Integer k)
    $sum \leftarrow 0$
    for $i = 0 \ldots k$ do
        $a \leftarrow$ MemoizeArragements($\mathscr{A}$, $i$)
        $b \leftarrow$ MemoizeArragements($\mathscr{B}$, $k - i - 1$)
        $sum \leftarrow sum + a \times b$
    end for
    return sum
end function
```

have been used [Ishwaran and Rao, 2005, He and Carin, 2009]. Any of these choices are possible and the prior on the Earth model parameters at each active tree node is left unspecified and simply write $p(\mathbf{v}_i|\mathcal{T}_k,k)$. In the case of $\mathbf{v}_i$ being of dimension m, this becomes

$$p(\mathbf{v}_i|\mathcal{T}_k,k)=\prod_{j=1}^{m}p(v_{i,j}|\mathcal{T}_k,k), \tag{3.19}$$

where $p(v_{i,j}|\mathcal{T}_k,k)$ is the prior on the jth component of the ith tree node.

Prior Ratios

For each class of proposal, that is birth, death and change value, the prior ratios can be derived. For a simple change in the j-th component of the parameter value in the i-th tree node, the structure of the tree does not alter and the prior ratio is

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})}=\frac{p(v'_{i,j}|\mathcal{T}_k,k)}{p(v_{i,j}|\mathcal{T}_k,k)}. \tag{3.20}$$

For uniform priors, $p(v'_{i,j}|\mathcal{T}_k,k)=p(v_{i,j}|\mathcal{T}_k,k)$ the prior ratio is unity.

For a birth proposal, the structure of the tree changes due to the addition of a new node and the prior of the values cancels except for those of the new node, hence the prior ratio is

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})}=\frac{p(k+1)p(\mathcal{T}_{k+1})p(\mathbf{v}_i|\mathcal{T}_k,k)}{p(k)p(\mathcal{T}_k)}, \tag{3.21}$$

where $p(\mathbf{v}_i|\mathcal{T}_k,k)$ is the prior on the values of the new node. If the prior on k, the number of nodes, is uniform then $\frac{p(k+1)}{p(k)}$ will cancel. Analytical expressions for the the prior ratio on the structure of the tree are generally not available except for some simple unrestricted trees of which some examples are presented in the following sections.

For the death proposal the prior ratio is

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})} = \frac{p(k-1)p(\mathcal{T}_{k-1})}{p(k)p(\mathcal{T}_k)p(\mathbf{v}_i|\mathcal{T}_k, k)}. \tag{3.22}$$

3.3.3 The Likelihood

It is assumed that the model vector can be mapped into the same data space as the vector of observations, **d**, so that a standard misfit can be computed as

$$\Phi(\mathbf{m}) = (\mathbf{G}(\mathbf{m}) - \mathbf{d})^T C_e^{-1} (\mathbf{G}(\mathbf{m}) - \mathbf{d}), \tag{3.23}$$

where **G** is the operator that represents the predictions of data observations from a model, **m** and C_e is the data error covariance matrix which assumes errors follow a Gaussian distribution. The standard normal error distribution can then be used for computing the likelihood

$$p(\mathbf{d} \,|\, \mathbf{m}) = \frac{1}{\sqrt{(2\pi)^n \, |C_e|}} \exp\left\{ -\frac{\Phi(\mathbf{m})}{2} \right\}, \tag{3.24}$$

where n is the number of observations. The operator **G** can take many forms, in Figure 3.1 two examples of how a quaternary tree can be mapped into a 2D image which could be compared to measured data. In later sections, examples will show how trees with the node values representing wavelet coefficients can also be mapped into 2D and 3D images.

3.3.4 The Proposals

For the proposal distribution, $Q(\mathbf{m} \rightarrow \mathbf{m}')$, there are three different classes of proposal: birth, death and change parameter value. Unlike trans-dimensional Voronoi cell approaches, because of the multi-scale parameterisation there is no proposal to move nodes as the tree nodes are spatially fixed. Throughout these explanations the prime superscript to represents proposed quantities, for example $\mathbf{m}'$ is a proposed model generated from the current model, $\mathbf{m}$, via proposal distribution $Q(\mathbf{m} \rightarrow \mathbf{m}')$.

To aid the explanation of the operation of these proposal classes, three sets of nodes within a general tree structure are introduced here. The first set is simply the set nodes in the current tree, or active nodes, S_v. Note that S_v is always non-empty because it will always have at least the root of the tree as an element. The second set, S_d, is the set of all nodes in the tree that have no active child nodes. It is from this set that nodes to remove from the tree are chosen during the death proposal of the algorithm. The third set, S_b, is the set of empty nodes in the tree structure template that are direct children of the nodes in set S_d. This set represents possible locations for adding new tree nodes during the birth proposal of the algorithm. It should be noted that the set S_d is a subset of S_v, whereas the set S_b is disjoint of the other 2 sets.

An example showing each set for a binary tree can be seen in Figure 3.6 with the nodes of each set shaded with a different colour.

Value Proposals

The first and simplest proposal is the change value proposal. This perturbation updates the value of an active node of the tree. If the general case of selecting the jth parameter at the ith node in the tree is considered, then the forward proposal probability density becomes

$$Q(\mathbf{m} \rightarrow \mathbf{m}') = Q(\Delta v_{i,j} \mid i, j) Q(j \mid i) Q(i \mid S_v). \tag{3.25}$$

The last term of the above equation represents the probability of choosing the ith node given S_v. Generally, the choice of which node to perturb for a value proposal will be a uniform one and so

$$Q(i \mid S_v) = \frac{1}{|S_v|}, \tag{3.26}$$

where $|S_v|$ is the number of elements in S_v.

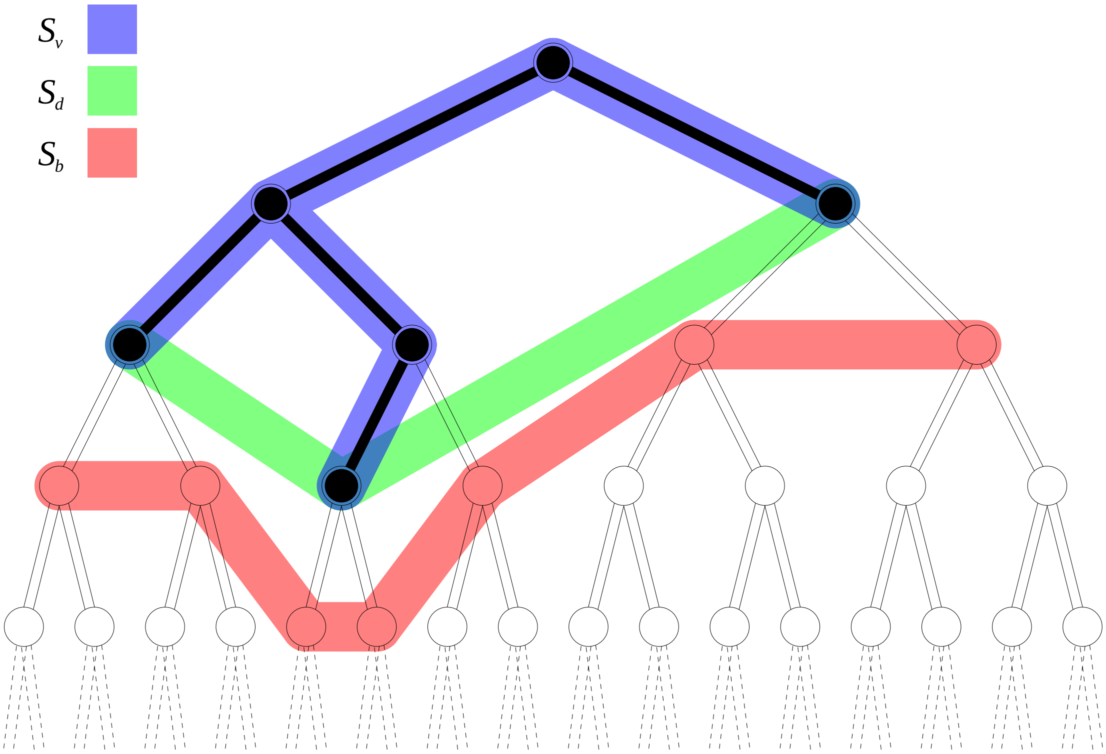


Figure 3.6: The first five levels of a binary tree template are shown in outline with a representative individual tree model drawn with solid lines. The nodes shaded in blue correspond to nodes in the current tree model and are members of the set S_v, or the set of nodes that can be perturbed during a change value proposal. The nodes shaded in green are members of the set S_d and represent nodes that can be removed by the next death proposal. Conversely, the nodes shaded in red are members of the set S_b which contains in-active nodes that could be added to the tree model by the next birth proposal. Although only a binary tree is shown here, these sets can apply equally to any tree structure.

The second term represents the probability of selecting the jth component of the vector of value(s) at the ith tree node. For cases where there is only one value at each node this term disappears. The first term is the actual perturbation of the Earth model parameter value itself. A common approach to perturbing values in McMC samplers is to draw from a symmetric distribution centred about the current value with a pre-defined width tuned to achieve a desired acceptance rate. A common choice is the Normal distribution and in this case the proposal probability will be

$$Q(\Delta v_{i,j} \mid i, j) = \frac{1}{\sqrt{2\pi}\sigma_{i,j}} \exp\left\{-\frac{\Delta v_{i,j}^2}{2\sigma_{i,j}^2}\right\}, \tag{3.27}$$

where $\sigma_{i,j}$ is the standard deviation of the normal distribution for the perturbation of the parameter. Using a proposal of this form, rather than sampling from the prior, can cause proposed values to be outside prior bounds, in which case the prior ratio is zero and the proposal is rejected.

The standard deviation may be the same for all tree nodes or set separately to achieve good acceptance rates. It is also straight forward to use adaptive schemes such as the Single Component Adaptive Monte Carlo (SCAM) approach of Haario et al. [2005] and the adaptive approach of Atchade and Rosenthal [2005].

Regardless of how the standard deviation or width is set, in all cases the new value is generated from a symmetrically distributed random variable. This results in the reverse proposal probability density equal to that of the forward, so the proposal ratio for changing values is unity

$$\frac{Q(\mathbf{m}' \rightarrow \mathbf{m})}{Q(\mathbf{m} \rightarrow \mathbf{m}')} = 1. \tag{3.28}$$

Birth Proposals

The birth proposal probability density may be written

$$Q(\mathbf{m} \rightarrow \mathbf{m}') = Q(\mathbf{v}_i \mid i) Q(i \mid S_b). \tag{3.29}$$

Similarly to the change value proposal, the last term represents the probability of choosing where to place the new node. Unlike in the case of the change value proposal, in some cases there is merit in preferentially choosing to birth nodes closer to the root of the tree. Experiments have been performed using weighted proposal densities of the form

$$Q(i \mid S_b) = \begin{cases} \frac{\mathscr{D}(i)^\alpha}{\sum_{j \in S_b} \mathscr{D}(j)^\alpha} & |S_b| > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.30}$$

where $\mathscr{D}(i)$ is the depth or height of node i and α is the weighting factor. Negative values of α preferentially select lower height nodes and, conversely, positive values preferentially select higher height nodes, whereas a 0 value results in a uniform choice of the birth node. In experiments with a weighted proposal the results were poorer than simply using a uniform proposal to select the position of the new node, so the preference here is a simpler uniform proposal

$$Q(i \mid S_b) = \begin{cases} \frac{1}{|S_b|} & |S_b| > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{3.31}$$

The case for the condition when $|S_b| = 0$ can only occur when there is some restriction on the tree structure template on the total number of nodes in the tree. An example of this would be a tree with a maximum height.

The first term of the proposal probability density in (3.29) reflects how the new parameters are chosen for the new tree node. The simplest method of performing this is to sample the new values from the prior

$$Q(\mathbf{v}_i \mid i) = p(\mathbf{v}_i | \mathscr{T}_k, k). \tag{3.32}$$

Although this is an "unfocused" proposal, birthing from the prior has been shown to result in good mixing by Dosso et al. [2014]. It also simplifies the calculation of the acceptance terms as the prior probability density in the proposal cancels with the prior ratio in the full acceptance expression.

The probability density for the reverse step can be written

$$Q(\mathbf{m}' \to \mathbf{m}) = Q(i \mid S'_d). \tag{3.33}$$

This states that the reverse proposal is simply the probability of selecting the newly added node i from the set S'_d. S'_d is the set of nodes that may be deleted after the proposed birth.

With uniform selection from the two sets involved and sampling from the prior for the new values, the general expression for the proposal ratio is

$$\frac{Q(\mathbf{m}' \to \mathbf{m})}{Q(\mathbf{m} \to \mathbf{m}')} = \frac{|S_b|}{|S'_d| p(\mathbf{v}_i | \mathcal{T}_k, k)}. \tag{3.34}$$

Death Proposals

The proposal probability distribution is essentially the reverse of the birth proposal, so again, for a uniform selection of the node to remove, and sampling from the prior on the reverse step, the proposal ratio for a death step is

$$\frac{Q(\mathbf{m}' \to \mathbf{m})}{Q(\mathbf{m} \to \mathbf{m}')} = \frac{|S_d| p(\mathbf{v}_i | \mathcal{T}_k, k)}{|S'_b|}, \tag{3.35}$$

where the set S'_b represents the set of available points to add nodes to the tree after the selected node is removed.

Jacobian

The last component of the acceptance criteria is the Jacobian which must be determined for each type of proposal.

For the change value proposal, the dimension of the model, **m**, is constant. Since only one value is perturbed at a time using a simple function of a random variable, the Jacobian will always be equal to one in this case.

For a birth proposal, the model space vector can be written as

$$\mathbf{m} = \langle (t_1, \mathbf{v}_1), \ldots, (t_k, \mathbf{v}_k) \rangle , \tag{3.36}$$

where unique indices $t_1 \ldots t_k$ are used to define the currently active nodes of the tree and hence the model vector becomes a set of tuples consisting of the node index and the vector of values associated with that node. The transform for a birth step, which must be bijective, can then be written

$$\begin{array}{c} \langle (t_1, \mathbf{v}_1), \ldots, (t_k, \mathbf{v}_k), (u, \mathbf{w}) \rangle \\ \Longleftrightarrow \\ \left\langle (t_1', \mathbf{v}_1)', \ldots, (t_k', \mathbf{v}_k'), (t_{k+1}, \mathbf{v}_{k+1})) \right\rangle \end{array} , \tag{3.37}$$

where u is a random variable used to choose the unique index of the location of the new node in the tree and $\mathbf{w}$ is the vector of random variables used to generate the values for the new node. To build the Jacobian, the matrix of partial derivatives of the functions used to map values from one model space to the other is constructed. For existing nodes in a birth step no change is required and

$$t_i, \mathbf{v}_i = t_i', \mathbf{v}_i' \quad \forall i \in 1 \ldots k. \tag{3.38}$$

Therefore the partial derivatives for these will be one. The proposals as described in previous sections for the choice of the location of the new node will always mean that

$t_{k+1} = u$ and likewise this will result in a partial derivative of 1.

For death proposals, the Jacobian is the inverse of that of the birth proposal as the birth and death proposals are symmetric to each other.

Non-unit Jacobians

In the case the prior is sampled for the values of the new node, $\mathbf{v}_{k+1} = \mathbf{w}$ and this will result in an identity matrix for the Jacobian and therefore unity for the determinant. This is the scenario that is generally generally advocated but here a potential extension is highlighted that results in some modification to the Jacobian.

The application of the tree structure suggests a multi-resolution hierarchy and as such there is expected to be some relationship between either the parent node and the newly added child node or a newly added child node and its siblings. For example, the values at the child will be less than that of the parent so proposals for the child node values may be drawn from distributions scaled by those of the parent. Alternatively, the mean of all the child nodes to form a distribution centred on zero, and so if there are existing child nodes then the distribution from which values for the a new child are drawn from is tempered by the existing sibling nodes. In either case the mapping takes the form

$$\mathbf{v}_{k+1} = f(\mathbf{w}, \mathbf{v}_j), \tag{3.39}$$

where f is some function of both the random variables and one or more of the existing values of other tree nodes (e.g. the parent or other sibling nodes). This will result in off-diagonal values in the Jacobian matrix. Some choices of the function, f, may also result in non-unity values along the diagonal of the Jacobian and care must be taken to correctly compute the Jacobian scaling term.

The General Acceptance Criteria

Having derived the components of the acceptance criteria, the general acceptance criteria can now be written in full for a trans-dimensional sampler on tree structures by combining the expressions. For a value proposal, the acceptance criterion is

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = min\left\{1, \frac{p(v'_{i,j}|\mathcal{T}_k, k)}{p(v_{i,j}|\mathcal{T}_k, k)} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}\right\}. \tag{3.40}$$

When using a uniform prior, $p(v'_{i,j}|\mathcal{T}_k, k) = p(v_{i,j}|\mathcal{T}_k, k)$, and the above expression reduces to the likelihood ratio.

For a birth step, with the values of the new node sampled from the prior, the acceptance criterion is

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = min\left\{1, \frac{p(k+1)p(\mathcal{T}_{k+1})}{p(k)p(\mathcal{T}_k)} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{|S_b|}{|S'_d|}\right\}. \tag{3.41}$$

And likewise for a death step, the general acceptance criterion is

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = min\left\{1, \frac{p(k-1)p(\mathcal{T}_{k-1})}{p(k)p(\mathcal{T}_k)} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{|S_d|}{|S'_b|}\right\}. \tag{3.42}$$

When using a uniform prior on the number of nodes, $p(k) = p(k+1) = p(k-1)$ and these terms cancel from the birth and death acceptance criteria.

These are conceptually simple criteria for sampling over general tree structures, however, a practical difficulty is in efficiently computing the tree structure prior ratios $\frac{p(\mathcal{T}_{k+1})}{p(\mathcal{T}_k)}$ and $\frac{p(\mathcal{T}_{k-1})}{p(\mathcal{T}_k)}$ for which a fast algorithm was detailed in Section 3.3.2.

3.4 Validation

It is generally acknowledged that the construction of acceptance criteria for trans-dimensional samplers is non-trivial. A small error in these criteria can easily result in a sampler that superficially appears to be working but will nonetheless give biased the results.

A key test of the correctness of a set of acceptance criteria for a trans-dimensional sampler is that the criteria do not bias the posterior on k, which in this case represents the number of nodes. The simplest way to demonstrate this is to allow the algorithm to run for a large number of steps with the likelihood kept at a constant value (hence the likelihood ratio is unity). Then the posterior on the number of nodes should match to sampling accuracy the known prior.

In a first test the general algorithm, the acceptance criteria for simple homogeneous trees is used, that is binary, ternary, quaternary trees etc. For these trees the result of Aval [2008] can be used to write down analytical expressions for the number of arrangements of the trees for a given number of nodes. This results in the following closed form solutions for the acceptance criteria of birth and death proposals respectively

$$\alpha(\mathbf{m}',\mathbf{m})_{\text{birth}} = \min\left\{1, \frac{\left[\prod_{j=2}^{n}(k(n-1)+j)\right](k+1)}{\prod_{j=1}^{n}(nk+j)} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{|S_b|}{|S'_d|}\right\}, \tag{3.43}$$

$$\alpha(\mathbf{m}',\mathbf{m})_{\text{death}} = \min\left\{1, \frac{n\prod_{j=1}^{n-1}(nk-j)}{\prod_{j=1}^{n-1}((n-1)k-j+2)} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{|S_d|}{|S'_b|}\right\}. \tag{3.44}$$

Where n represents the number of child nodes for each tree node, that is $n = 2$ corresponds to a binary tree, $n = 3$ corresponds to a ternary tree etc. Simulations 1 million Markov iterations with a uniform prior on the number of nodes between 1 and a variable k_{max} and for three different values of n were performed. The results of this test are shown in Figure 3.7 with expected histogram shown with a red solid line. In all cases the McMC results approximately match with the uniform prior.

The tests were repeated for a case where the prior PDF on k is not uniform, specifically a truncated Poisson prior on the number of nodes of the form

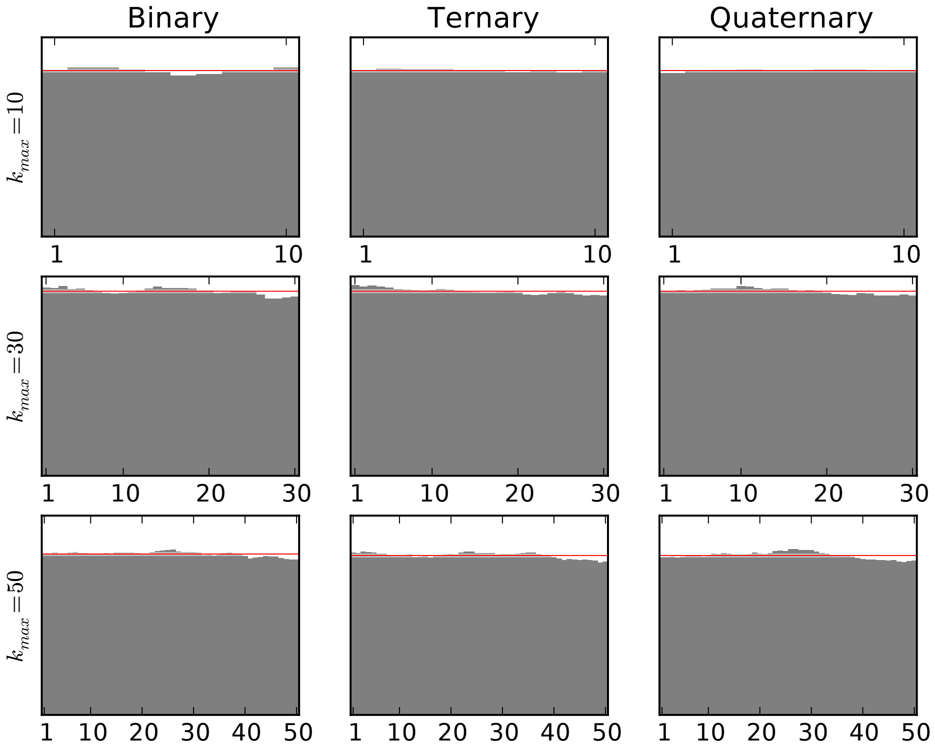


Figure 3.7: The sampled prior of k, the number of active tree nodes, is plotted as a grey histogram for a variety of uniform prior widths with three different classes of trees (binary, ternary and quaternary). In each plot, the solid red line represents the input prior showing there is good agreement between the prior and sampled histogram. This gives confidence that the algorithm maintains detailed balance and therefore will correctly sample the true PPD.

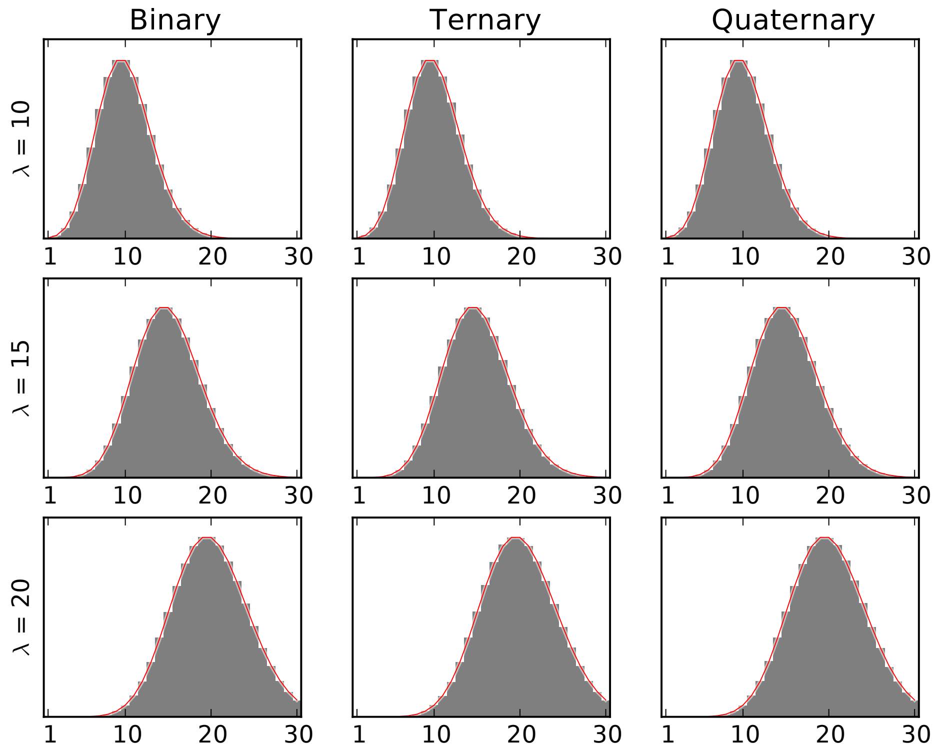


Figure 3.8: The sampled prior obtained when using a truncated Poisson prior is shown with a grey histogram. In each of these tests, the the maximum k is fixed at 30 and the λ parameter of the Poisson prior is varied with different classes of trees (binary, ternary and quaternary). The prior is shown with a solid red line and agrees well with the sampled histogram.

$$p(k) = \frac{\lambda^k}{(e^{\lambda} - 1)k!}, \tag{3.45}$$

where λ represents an approximate expected number of nodes in the tree. The posterior on the number of tree nodes obtained for varying λ and n is shown in Figure 3.8 with the prior over plotted with a solid line. Again the sampled posterior closely matches the analytical prior to within sampling error.

Lastly, this experiment was repeated with a truncated Jeffreys' prior,

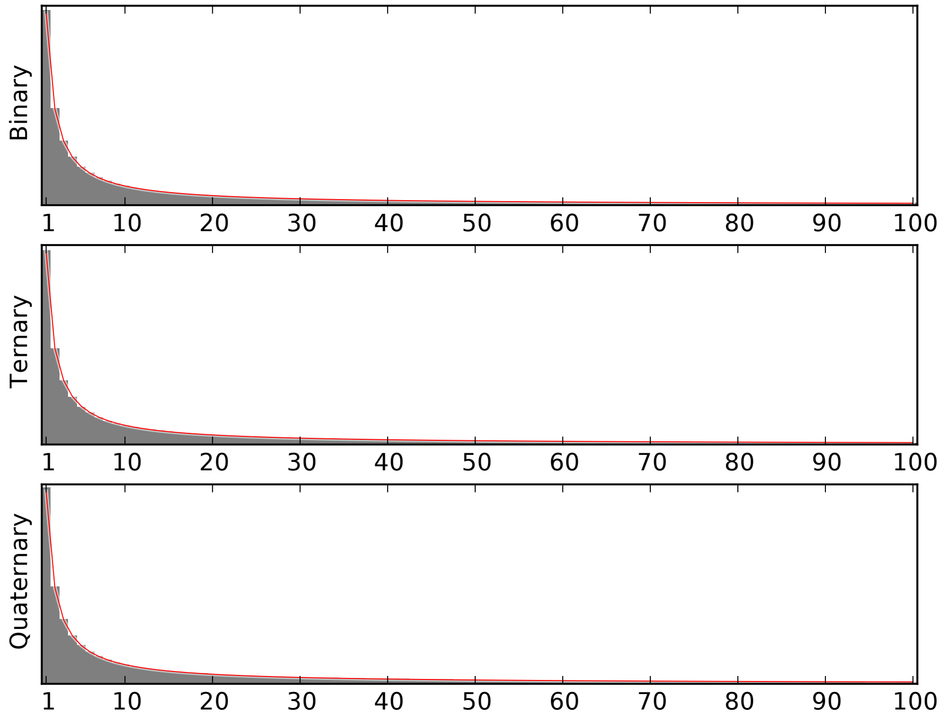


Figure 3.9: The sampled prior obtained when using a truncated Jeffreys' prior is shown with a grey histogram. The maximum k is fixed at 100 and the posterior for three classes of tree, binary, ternary and quaternary are shown. The analytical prior is shown with a solid red line and good agreement is obtained with the sampled histogram.

$$p(k) = \begin{cases} \frac{c}{k} & 1 \leq k \leq k_{max} \\ 0 & \text{otherwise} \end{cases} \tag{3.46}$$

For some normalising constant c and an upper limit on k of k_{max}. The posteriors obtained for different n-ary trees with a k_{max} of 100 are shown in Figure 3.9 along with the true distribution plotted with a solid line. In all cases, the posterior appears to be correctly sampling the prior on the number of tree nodes. This is encouraging and lends support to the proposition that the trans-dimensional formulation presented here is correctly balanced.

3.5 A simple synthetic regression test

To give a simple example of the application of this general framework, the performance is demonstrated in a synthetic 1D regression problem. This uses a binary tree template and a box car basis function of varying width and location at each node of the tree. This is the 1D equivalent of the parameterisation shown in Figure 3.1. Given a boxcar

$$B(x)_{i,j} = \begin{cases} 1 & 2^{-i}j \leq x < 2^{-i}(j+1) \\ 0 & \text{otherwise} \end{cases}, \tag{3.47}$$

where i represents its width and j its offset, a binary tree template can be constructed containing coefficients, $S_{i,j}$, at each node. The 1D regression function to be estimated is then constructed from

$$g(x) = \sum_{i=0}^{i_{max}} \sum_{j=0}^{2^i-1} S_{i,j} B(x)_{i,j}. \tag{3.48}$$

The i coordinate maps to the height in the tree and j runs horizontally starting at 0 for each row. This is shown graphically in Figure 3.10.

To verify that information can be recovered from noisy data, the binary tree template with boxcar basis functions is used to invert data samples from a synthetic step function with added Gaussian noise. The true model is shown in Figure 3.11(a) together with the data samples which are irregularly sampled to create areas of sparse coverage.

A single Markov chain was simulated with 1 million steps with the first 500,000 samples discarded as burnin. The probabilities of the birth, death and change value proposals are set to $p(\text{birth}) = p(\text{death}) = 0.25$ and $p(\text{change value}) = 0.5$. The choice of these probabilities is arbitrary except that they sum to one. In principle these could be tuned for better performance in larger more complex problems, but for this simple problem this is unnecessary. The prior on the coefficients at each node was set to uniform between ± 1, and the change value proposals were normally distributed with standard deviation of 0.1. The initial model was set to have one node (the root of the tree) with

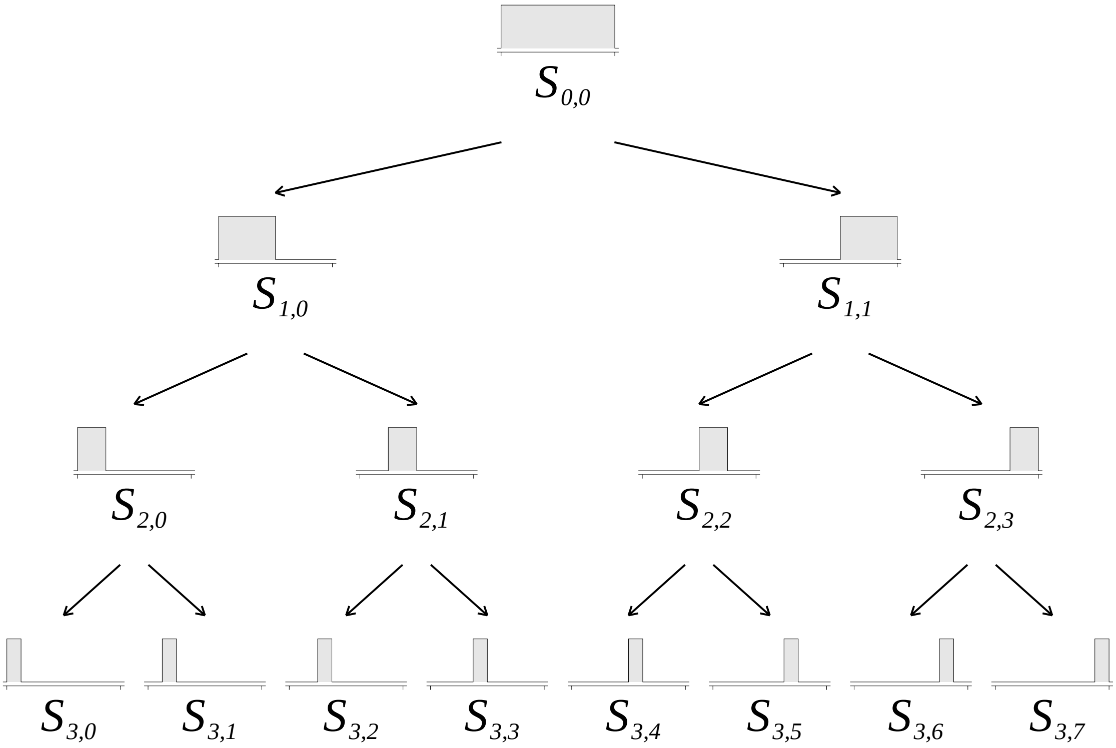


Figure 3.10: In a binary tree template, a boxcar basis function is associated with each tree node. The boxcar basis functions are shown graphically embedded in a binary tree structure. Along each row or at each height of the tree, the basis functions are orthogonal to each other. Conversely, from any parent node, the two child node basis functions are bisecting sub-dividers of the parents basis function. By storing scaling terms at each node of the tree, $S_{i,j}$*, a 1D function can be constructed from the tree expressed as the sum of scaled versions of the basis functions using (3.48).*

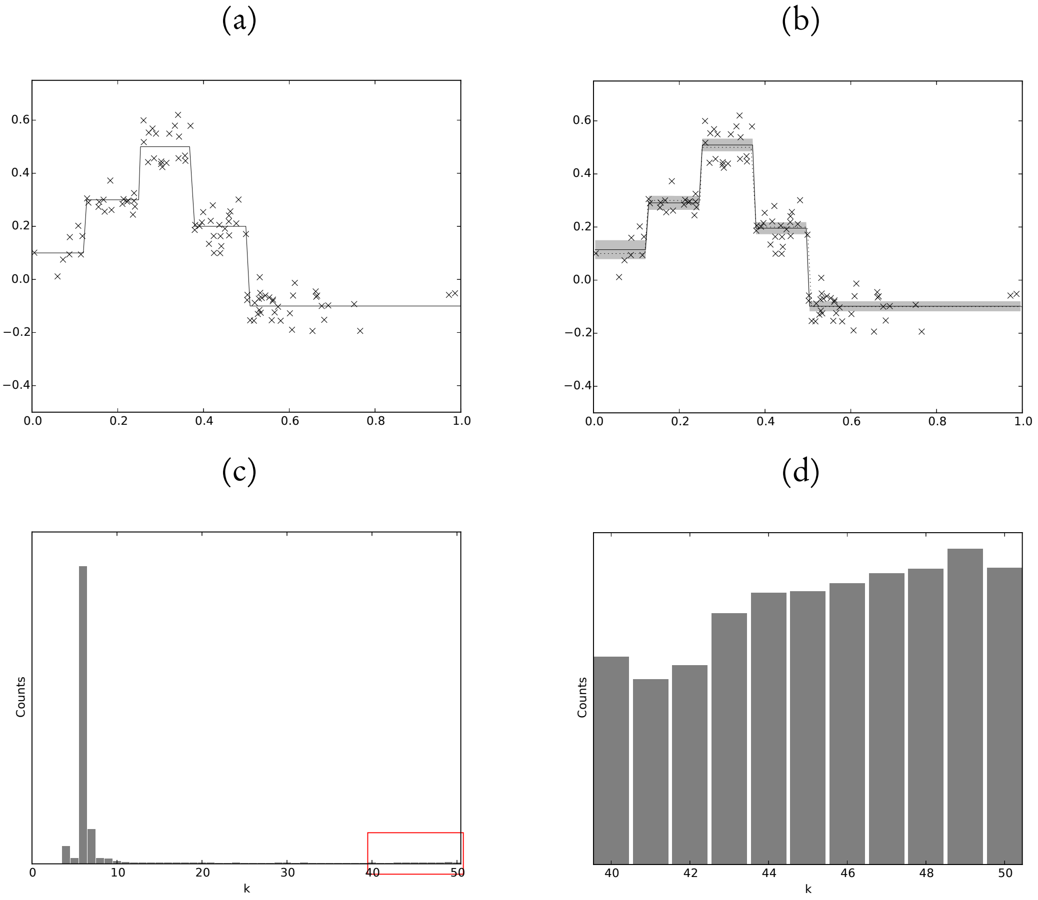


Figure 3.11: The results from the 1D regression experiment using unrestricted binary trees with boxcar basis functions. The synthetic data is shown in (a) which consists of sparsely located data points shown with crosses and the underlying true model is shown with a solid line. The recovered model is shown in (b) with a solid line compared to the true model represented with a dotted line. The shaded region represents ± 3 times the point estimate of the standard deviation from the ensemble models. In (c) is the posterior probability density (PPD) of k, the number of nodes of the tree which has a modal value at 6. In (d) is the PPD of k zoomed in at the higher values of k highlighted with the red box in (c) showing that the posterior has sampled across the entire range of the prior.

its initial value sampled from the prior.

The mean result is shown with a solid line in Figure 3.11(b) compared to the true model (dotted line). The recovery is accurate and additionally the data is not over fit, even in regions of poor data coverage. The variance obtained from the posterior is also low which is expected in this case as the parameterisation can perfectly represent the true model.

Figure 3.11(c) shows the posterior histogram on the number of tree nodes used to represent the data. The modal number of tree nodes is 6 which matches the true model. It is interesting to note that over the course of the sampling, the entire prior range of the number of tree nodes has been sampled, as evidenced by the small number of counts at 50 nodes, shown enlarged in Figure 3.11(d), even though the Markov chain is initiated at $k = 1$ node.

These simple tests lend confidence that the algorithm and acceptance criteria of the general framework derived here are correctly balanced and able to statistically converge.

3.6 Linearised Tomography example

It is useful here to compare the operation of the trans-dimensional tree approach compared to existing techniques in some very simple linearised tomographic problems. This simple set of problems highlight the data adaptability of the trans-dimensional tree approach.

3.6.1 Problem description

In a 2D linearised tomographic problem, commonly used in regional surface wave studies, the observations consist of travel times between two spatial points. In the ray theory approximation the travel time is a path integral along a ray path of the underlying spatially varying velocity. This ray path is dependent on the velocity field and this aspect is ignored in linearised tomography where an approximation is to assume

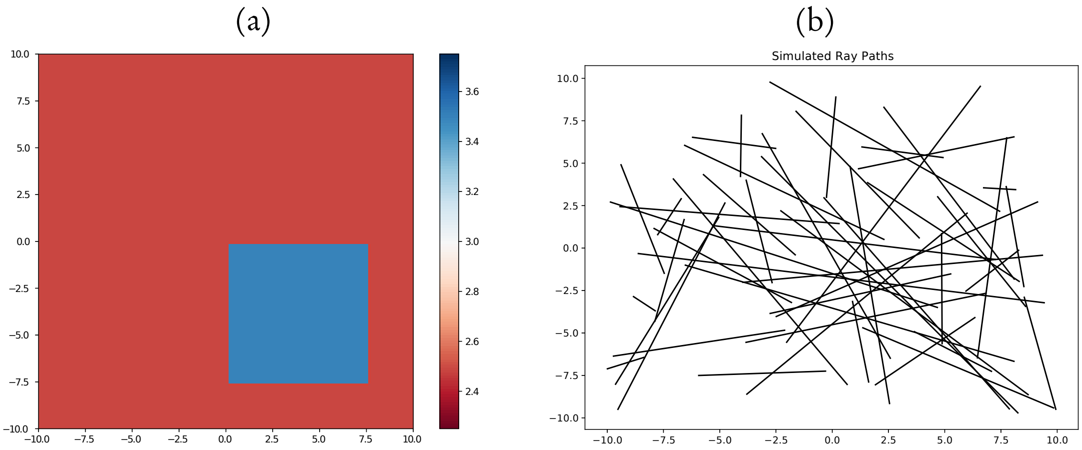


Figure 3.12: The true velocity model for the first test in (a) and the fixed paths through the model for the linearised inversion in (b).

the ray paths are fixed, either to great circle paths or paths computed in some reference model. This reduces the inversion to a linear problem of finding the velocity field that best predicts the travel time observations given the fixed rays.

For these tests, simple true velocity models are used with randomly generated rays. In the first test, shown in Figure 3.12 a simple anomaly is positioned in the south-east with uniformly distributed rays in an 8 by 8 grid. In the second test, shown in Figure 3.13 the anomaly in the south-east now has multi-resolution features and there is a higher density of rays in this area to recreate a common issue in seismology where ray coverage is uneven.

The domain of these tests is a 20 by 20 degree patch of the Earth centred at 0 longitude and latitude. To generate synthetic observations for these comparison tests, the ray path integrals are evaluated along each ray to obtain a true travel time and independent Gaussian noise is added to each.

A common approach to solving this problem is least squares optimisation for the slowness field or the inverse of the velocity which is a linear problem.

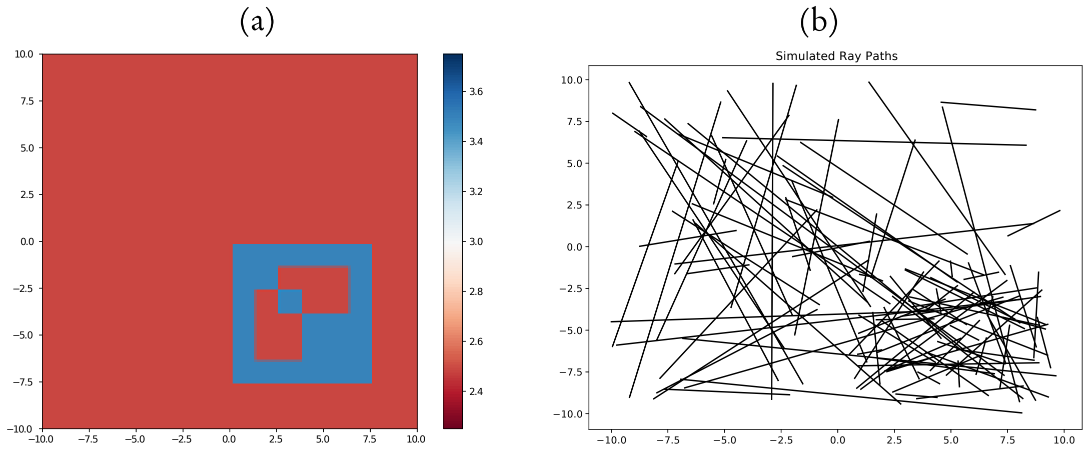


Figure 3.13: The true velocity model for the second test in (a) and the fixed paths through the model for the linearised inversion in (b).

3.6.2 Least squares optimisation

In a least squares optimisation approach to the linearised tomography problem, due to uneven coverage there is generally a requirement of using a penalising norm. These norms either penalise deviation from a reference model, called damping, or penalising large spatial gradients of the model, called smoothing. The choice of the relative weight of the misfit between predictions and the penalising norm is a choice that needs to be made as part of the inversion process. Mathematically, the optimisation problem can be written as the minimisation of

$$|\mathbf{Gm}-\mathbf{d}|_2+\lambda|\mathbf{m}-\mathbf{m}_0|_2 \tag{3.49}$$

for damping regularisation and

$$|\mathbf{Gm}-\mathbf{d}|_2+\lambda|\nabla\mathbf{m}|_2 \tag{3.50}$$

for smoothing.

In the comparison of the least squares optimisation to trans-dimensional tree sampling, the approach used here for choosing the regularisation parameter λ is the discrepancy principle [Menke, 1989, Vogel, 2002]. The rationale for this is that in a synthetic

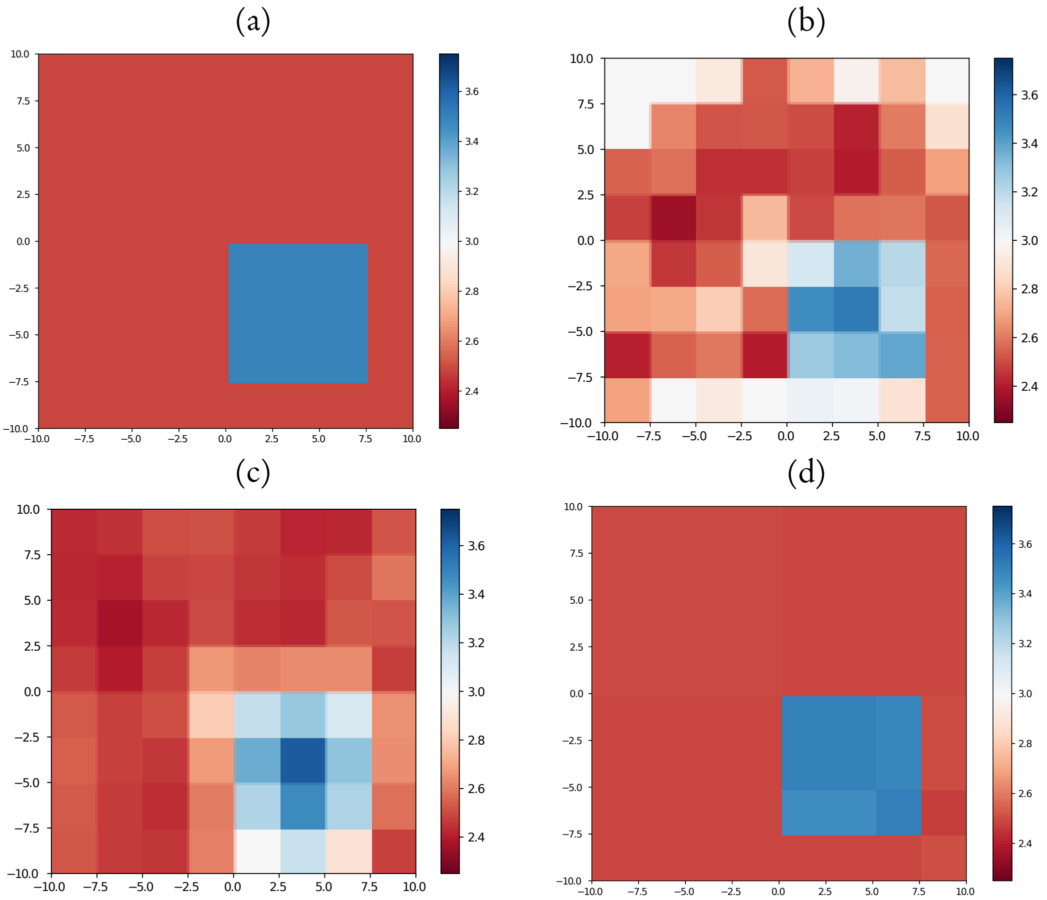


Figure 3.14: In (a) is the true model of the first synthetic test, (b) the least squares optimisation result with damping to a reference model regularisation, (c) the least squares optimisation result with smoothing regularisation, and in (d) the equivalent trans-dimensional tree inversion is shown.

example, where the noise level is precisely known, this gives the residual norm that is approximately equivalent to that of the trans-dimensional tree method. It is not exact as the trans-dimensional tree method will sample about the model space and produce an ensemble of residual norms, but the modal value of the residual norms from the trans-dimensional result should match the optimised residual norm.

In Figure 3.14, a comparison of the results in (b) of a damped least squares inversion, and (c) of a smoothed least squares inversion to the trans-dimensional tree inversion (d) is shown for the first synthetic test case. All the results recover the general character of the true model however the optimisation with damping method has produced spurious artefacts in the inversion compared to the trans-dimensional tree approach. In (c) there is a similar result for the least squares solution with smoothing regularisation. In this

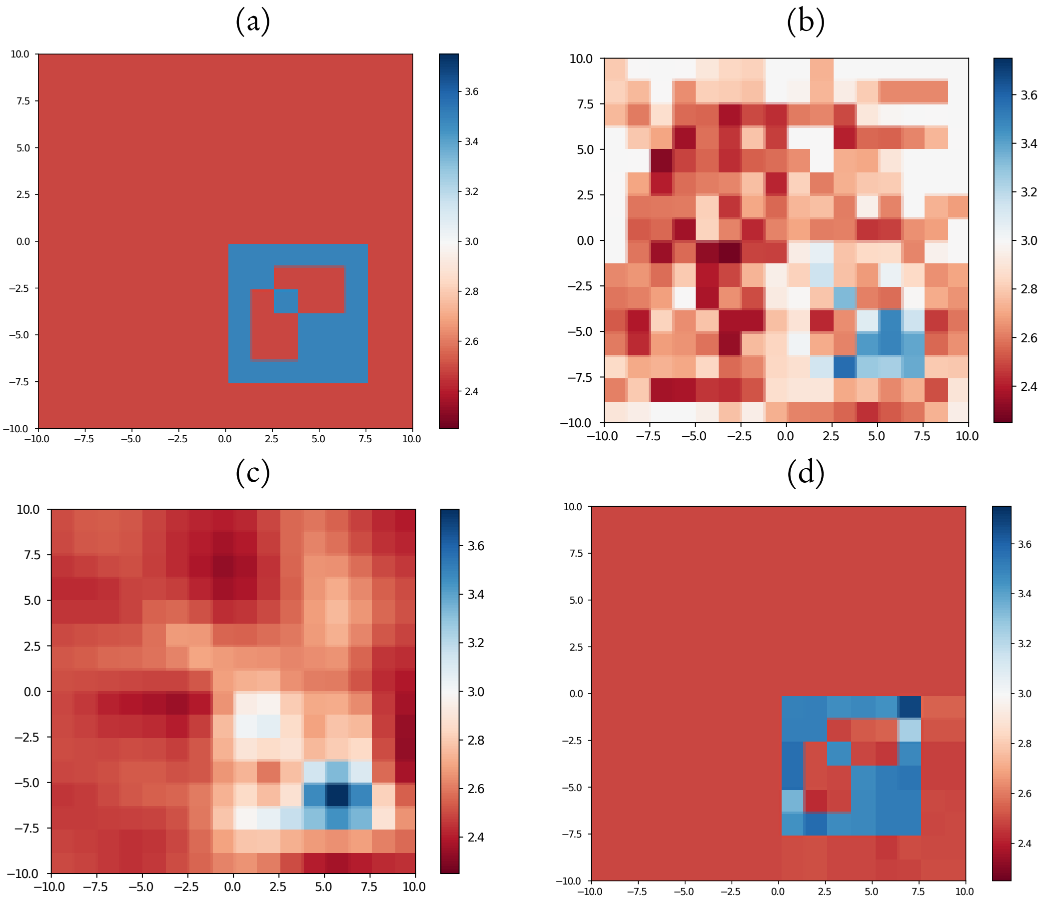


Figure 3.15: In (a) is the true model of the first synthetic test, (b) the least squares optimisation result with damping to a reference model regularisation, (c) the least squares optimisation result with smoothing regularisation, and in (d) the equivalent trans-dimensional tree inversion is shown.

case, the artefacts away from the anomaly are reduced, but the anomaly itself has been smoothed which is to be expected.

In comparing the more complicated synthetic model with multi-resolution features in Figure 3.15 with damping regularisation, in (b) the damping regularisation has performed poorly at recovering some of the finer features. Similarly in (c) the fine features have been lost to the smoothing operator. The trans-dimensional tree approach however has recovered the true model reasonably well by comparison. The primary reasons for this is its adaptability to the resolving power of the data which is broadly related to the ray path density. For the least squares optimisation approach, the controlling parameter for the damping or smoothing is a global parameter across the domain so no such adaptability is available. This means that any least squares inversion formulated in such a way will invert the structure modulated by the lowest resolvable resolution of the data. Attempts have been made to use data adaptive techniques to use the ray path density to guide the inversion however rays may not necessarily give independent information. The trans-dimensional tree approach uses Bayesian model selection to achieve this and more conservatively adapts to the information in ray paths, be they independent or not.

It's this adaptability and robustness of the trans-dimensional tree algorithm that encourages further examination of its application to geophysical inversion problems with sparse or uneven sensitivities.

3.6.3 Direct sampling

A potential criticism of trans-dimensional sampling is one of why bother when it is easier to use a Bayesian sampling approach for a fixed dimension pixel based inversion. In this short comparison, the the same synthetic problems are used to compare a non trans-dimensional Bayesian McMC inversion to the trans-dimensional tree approach.

In Figure 3.16, the ensemble means of the full sampling of the wavelet coefficients and the trans-dimensional tree algorithm are shown. Similarly to the damping result, while the recovery of the anomaly is visually reasonable, there is much unconstrained structure elsewhere. This poor constraint has impacts in the uncertainty estimates and the standard deviation of the ensembles is shown in Figure 3.17. The result of directly sampling increases the uncertainty of the result. This occurs because there are more degrees of freedom in the full McMC sampler which results in higher uncertainty.

This trend continues in the second multi-resolution test with the ensemble means shown in Figure 3.18 and the standard deviations shown in Figure 3.19. The recovery in this case contains far more spurious structure and the uncertainty has increased further. In a Bayesian inversion, full resolution sampling as performed here can be married with a prior with a spatial correlation length to effectively add a smoothing constraint analogous to the smoothing regularisation in the previous section. This

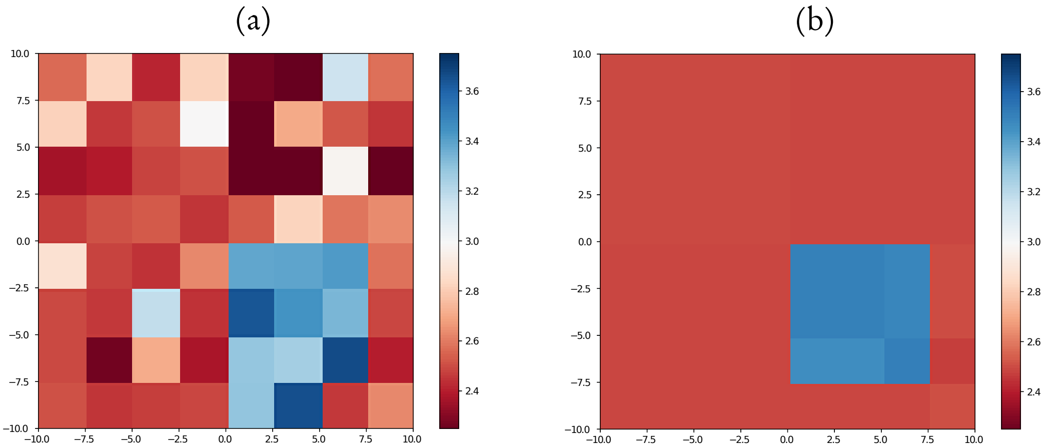


Figure 3.16: In (a) the full Bayesian McMC ensemble mean is shown in (a). In (b) the equivalent trans-dimensional tree inversion is shown.

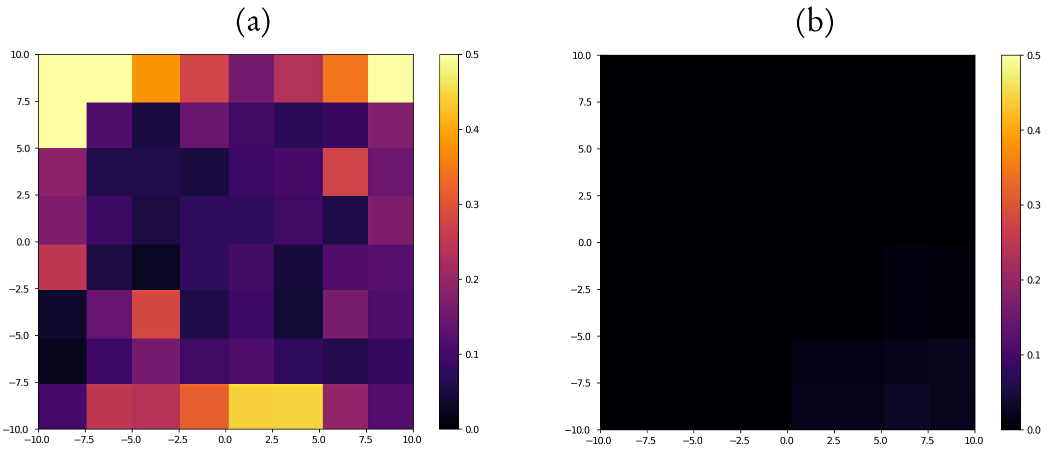


Figure 3.17: In (a) the full Bayesian McMC ensemble standard deviation is shown. In (b) the equivalent trans-dimensional tree inversion ensemble standard deviation is shown.

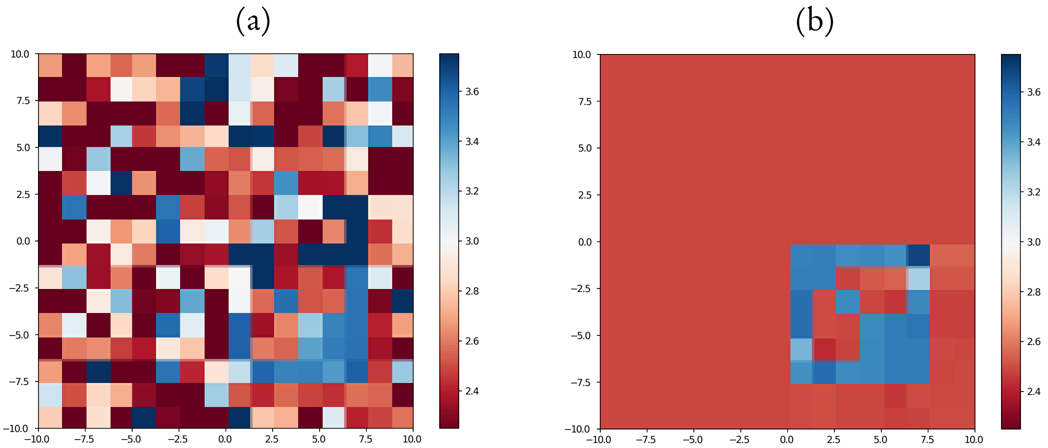


Figure 3.18: In (a) the full Bayesian McMC ensemble mean is shown. In (b) the equivalent trans-dimensional tree inversion is shown.

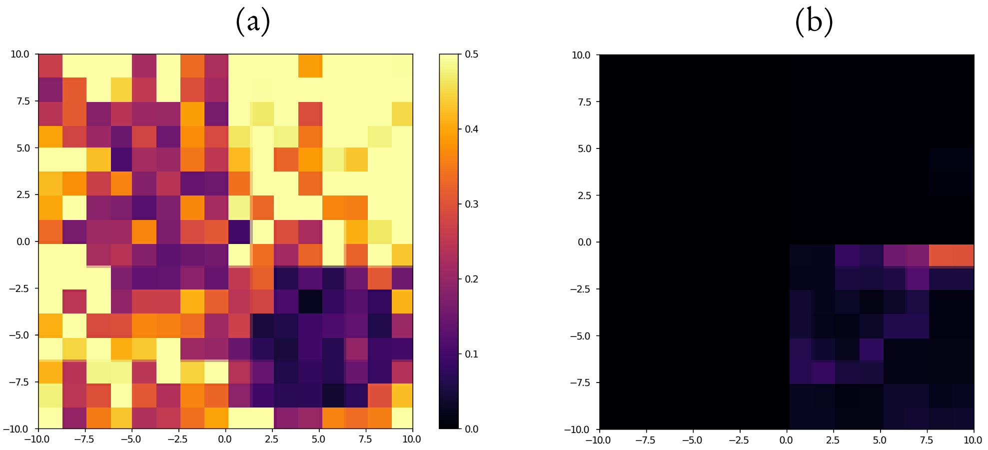


Figure 3.19: In (a) the full Bayesian McMC ensemble standard deviation is shown. In (b) the equivalent trans-dimensional tree inversion ensemble standard deviation is shown.

would improve the inversion to a degree but it would have the similar problems as the smoothing regularisation, that is the choice of the smoothing parameter globally affects the result and cannot adapt to locally resolvable structure.

3.7 Linearised tomography with wavelets

This section expands on the simple 2D linearised tomography by using a variety of wavelet bases and with larger problems more representative of real world applications. The particular target application here is ambient noise tomography which is a technique of obtaining near surface structure information from correlation of noise measurements between spatially distributed receiver stations, introduced to the seismological field by Shapiro and Campillo [2004] (see also review articles by Larose et al. [2006] and Snieder and Larose [2013]). For comparison purposes, the trans-dimensional tree with wavelet parameterisation method is compared to existing trans-dimensional methods using Voronoi cells.

Trans-dimensional travel-time tomography using a Voronoi cell parameterisation was introduced by Bodin and Sambridge [2009] and has been successfully used for inversion of ambient noise measurements for group velocity structure in several regional studies, for example, Young et al. [2011], Pilia et al. [2014], Saygin et al. [2016]. In the following sections, the solution of this problem using the new trans-dimensional tree algorithm with wavelets as basis functions is demonstrated.

3.7.1 A Tree-structured Wavelet Parameterisation

Wavelet analysis may be used to decompose bounded signals in both time and frequency at multiple scales. This is in contrast to Fourier analysis which decomposes signals by frequency only (for an introduction to wavelets see Daubechies [1992] and Mallat [1999]. The Fast Discrete Wavelet transform, following the multi-resolution wavelet transform of Mallat [1989], has been used in a variety of image based problems, notably image compression. Wavelet bases have been previously used in several studies for resolving seismic tomography at various scales, for example Chiao and Kuo [2001], Simons et al. [2011b], Chevrot et al. [2012], Charlety et al. [2013], Fang and Zhang [2014].

The Discrete Wavelet Transform in Cartesian domains has a natural multi-scale hierarchy that can be traversed with a tree structure. In image compression, this has been utilised by Shapiro [1993] and Said and Pearlman [1996]. In Compressive Sensing the same tree structure has been used for 1D signal recovery by La and Do [2005] and 2D image reconstruction by He et al. [2010].

In Figure 3.20(a) the progressive decomposition of a small 4 by 4 pixel image (bottom) by a wavelet transform is shown. As can be seen, at each step the image is reduced by half in each dimension. The wavelet based tree structure of this wavelet decomposition is illustrated in Figure 3.20(b). The progressively shaded regions indicate each level of wavelet decomposition with the darkest top left corner representing the scaling coefficient of the wavelet decomposition at the coarsest level which also corresponds to the root of the tree.

The tree has 3 children from the root node, and 4 children from every other node with the exception of the last nodes representing the finest level of detail which have no children. This is the case for a region in which the width and height are equal.

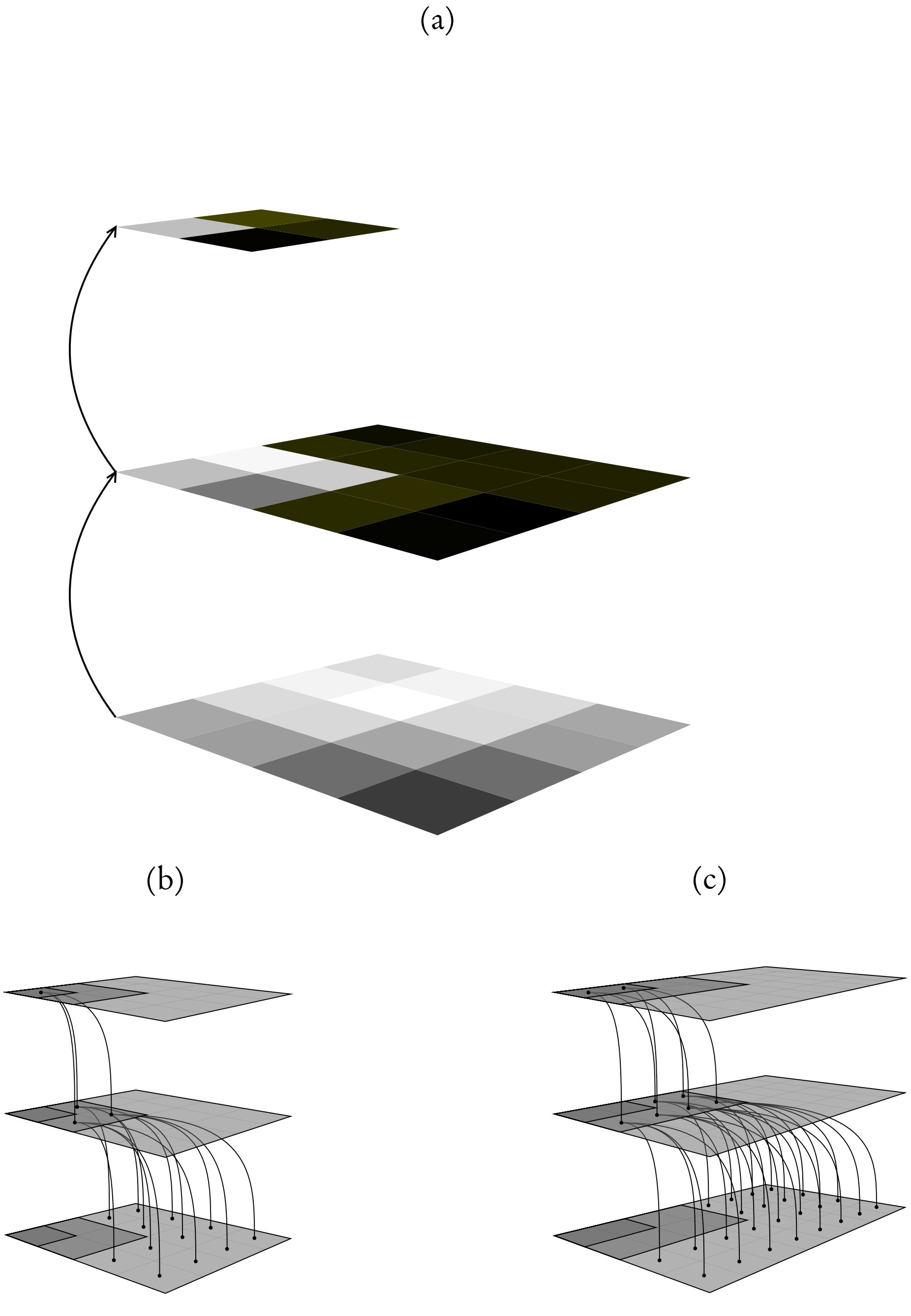


Figure 3.20: In (a) is a simple 4 by 4 image in the lowest panel and two successive wavelet transforms of this image in the panels above. The first forward wavelet transform results in 2 by 2 lower resolution approximation of the input image and a set of wavelet coefficients (shown in darker shade). The next step performs the forward wavelet transform on the 2 by 2 image to obtain a 1 pixel approximation and 3 wavelet coefficients. With this 1 pixel approximation and the 3 plus 12 wavelet coefficients, the original 4 by 4 pixel image can be recovered using the inverse wavelet transform. In (b) is the tree structure that spans the 1 pixel approximation and wavelet coefficients of a 4 by 4 square image. Each level of decomposition is shaded a progressively lighter shade of gray and note how each branching of the tree coincides with the next wavelet decomposition level. In (c) it is demonstrated how a variant of the tree structure can equally apply to rectangular regions by beginning from 2 top level coefficients.

For rectangular regions, a tree can be constructed by treating the initial scaling coefficients of a wavelet decomposition of a rectangular region as a 2D subdivision grid. An example is seen in Figure 3.20(c). In the following examples square images are used for simplicity, however, the only limitation when working with wavelets and this framework is that each image dimension must be a power of two.

At the root of the tree, the parameter value represents the scaling coefficient from a wavelet decomposition of the tomographic image. The parameter values of the remaining tree nodes represent the hierarchy of wavelet coefficients. In contrast to the earlier 2D image example, where values at the tree nodes are directly summed into an output image, an image is reconstructed from these coefficients using the inverse wavelet transform [Mallat, 1999].

3.7.2 The Synthetic Model and Test Procedure

To demonstrate the performance of the new trans-dimensional tree algorithm, it is compared to the Voronoi parameterisation in some synthetic checker board tests with 1,000 fixed ray paths. The ray paths remain fixed during the sampling to allow a direct comparison between the various parameterisations, however, there is no impediment in the new method that prevents either ray path updates at every step for a fully non-linear inversion [Galetti et al., 2015] or periodic updates for an iterative non-linear scheme [Bodin and Sambridge, 2009].

The true model and the ray coverage are shown in Figure 3.21. The region of the test is set to a square bounded at ± 10 degrees longitude and latitude. The model is a smooth (cosine) checker board, results for a discontinuous (boxcar) checker board are also shown in Section 3.8. The observed travel times are computed by integrating along each path and Gaussian noise is added with a standard deviation of 5 seconds which corresponds to approximately a 2.5 percent error on the mean travel time.

For the wavelet parameterisation, the experiment is repeated with three different wavelet bases. These are the Haar wavelet [Haar, 1910], the Daubechies 6-tap wavelet

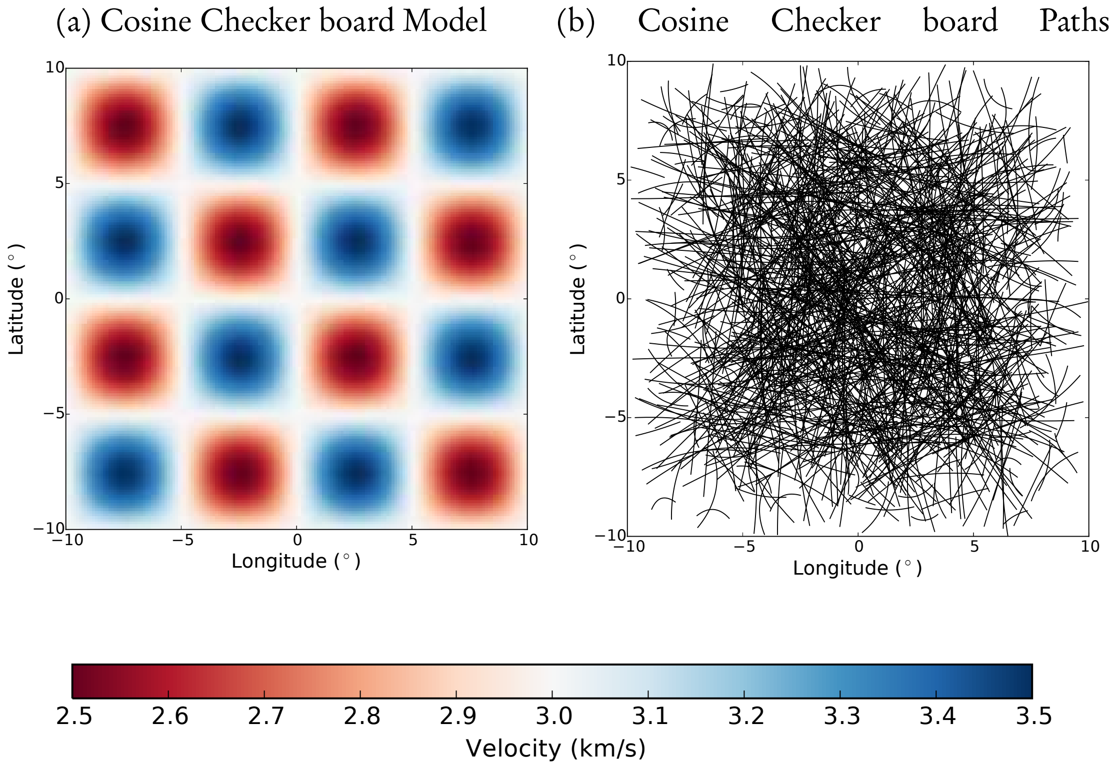

(a) Cosine Checker board Model (b) Cosine Checker board Paths

Figure 3.21: The synthetic models used in the test is a smooth (cosine) checker board with seismic velocities between 2.5 and 3.5 km/s. 1,000 random ray paths are generated through the region and travel times integrated to obtain the synthetic observations to which we add Gaussian noise.

[Daubechies, 1988] and the Cohen-Daubechies-Feauveau 9/7 wavelet [Cohen et al., 1992, see table 6.2] as used in the JPEG-2000 image compression standard [Usevitch, 2001]. The choice of these wavelet bases is designed to give a representative selection of available wavelets with varying degrees of smoothness.

The tests are run under comparable conditions, that is 64 independent Markov chains are used in each case with 10 million steps. At an interval of 1 million steps, each chain is restarted by randomly choosing a new starting model from current population with probability proportional to the mean likelihood of each chain. This approach, detailed by Dettmer et al. [2011], accelerates convergence to sampling the high-probability region of the posterior PDF and prevents individual chains from becoming stuck in local modes.

For the Voronoi case, all chains are started with a single cell corresponding to a tree with a single root node. For birth/death proposals, “birth from the prior” is used for both the Voronoi and Wavelet parameterisation. For change value proposals, fixed Gaussian perturbations are used of the cell values/wavelet coefficients where the proposal width is reasonably tuned to obtain acceptance rates of approximately 20 to 40 percent.

The prior on the number of parameters, $p(k)$, is set to be uniform between 1 and 5,000 parameters (3.6). In the Voronoi parameterisation, a uniform prior on the wave speed is set between 2.0 and 4.0 km/s which encompasses the true range of 2.5 and 3.5 km/s.

For the Wavelet parameterisation, the prior specification is complicated by the fact that the range of values of the coefficients can vary by several orders of magnitude, that is from the coarsest to finest resolution. This means that it is sensible to set a different uniform prior for each level of wavelet decomposition with the prior bounds determined by examining likely velocity variations. This approach suffices for these simulation tests, but a more advanced scheme such as that of Lochbühler et al. [2015], would also be possible.

3.7.3 Ensemble mean and Credible Intervals

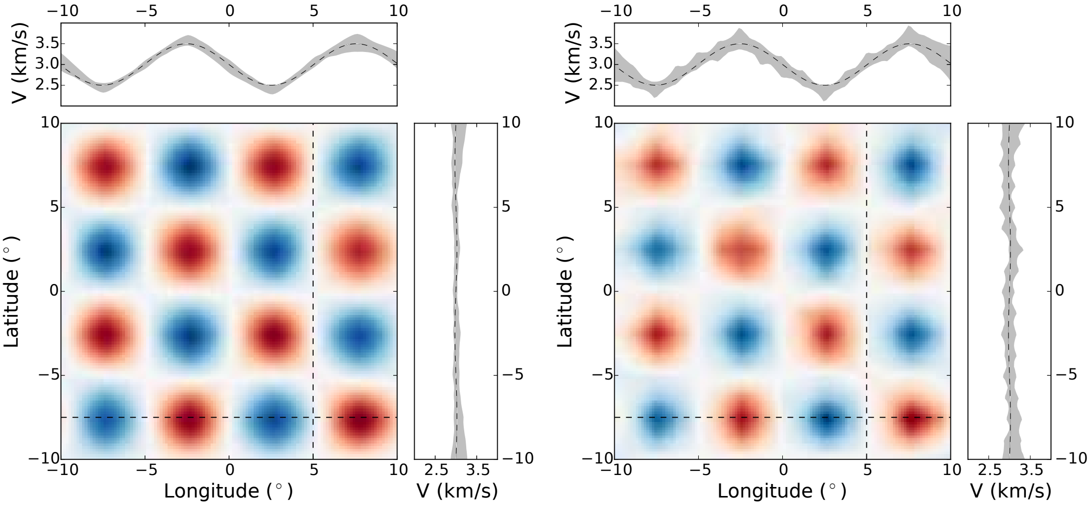
The mean of the ensembles across all chains are shown in Figure 3.22. Subjectively, the CDF 9/7 and Daubechies 6 wavelets have recovered the smooth model better. The Haar wavelet has performed poorly while the Voronoi parameterisation reasonably recovered the broad pattern of the model but introduced polygonal artefacts.

In addition to the mean of the ensemble, point wise 95 percent credible intervals can be extracted. In Figure 3.22, also plotted is the 95 percent credible interval along transects indicated by the dashed line through the ensemble means. The credible interval is shown as a shaded grey range and the true model is shown with a dashed line. The two transects are chosen in this example so that the longitudinal transect samples along peaks and troughs while the latitudinal transect samples along a constant velocity. From Figure 3.22(a) and (b) shows that the CDF 9/7 and Daubechies 6 wavelets have low uncertainties, a characteristic of model parameterisations suited to the underlying data. In contrast, the magnitude of the uncertainties for the Haar wavelet and the Voronoi cell are significantly higher. These results highlight the point that the choice of parameterisation is important both to the recovery and, more importantly, to the uncertainties recovered.

3.7.4 Number of Model Parameters

The number of parameters (coefficients in the trans-dimensional tree based wavelet parameterisation and cells in the Voronoi parameterisation) gives a simple measure of model complexity. Direct comparison between the two parameterisations is not very meaningful because in the Voronoi parameterisation each cell has three parameters, the cell value and its (x, y) coordinates, whereas for the trans-dimensional tree wavelet parameterisation, the most reasonable approach is to assume the model is written as in (3.36) where each parameter has a coefficient value and a unique tree node identifier as variables. This would mean that the number of Voronoi cells should be multiplied by 3 and the number of wavelet coefficients by 2 to obtain a fair comparison. In the

(a) CDF 9/7 (b) Daubechies 6



(c) Haar (d) Voronoi

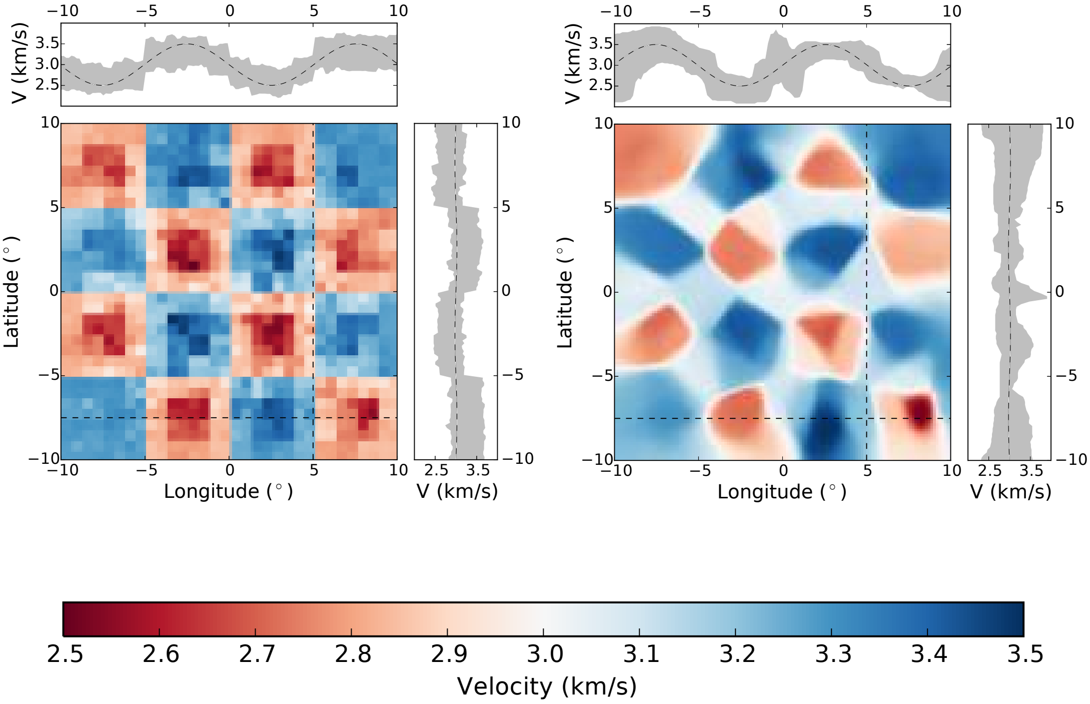


Figure 3.22: The mean of the ensembles obtained for the four different parameterisations. In each plot the uncertainties are shown along longitudinal and latitudinal transects indicated by the dashed lines. These show the 95 percent credible interval as a grey shaded region with the true model over plotted with black and white dashes.

Figure 3.23, the histograms of the raw number of cells/wavelet coefficients are shown.

For the wavelet parameterisations, the number of coefficients is higher than that of the Voronoi cell parameterisation, particularly for the Haar wavelet parameterisation. This may suggest that the wavelet parameterisations is over-parameterised, however, as will be shown in Section 3.7.7 this is not necessarily the case.

3.7.5 Computational Time

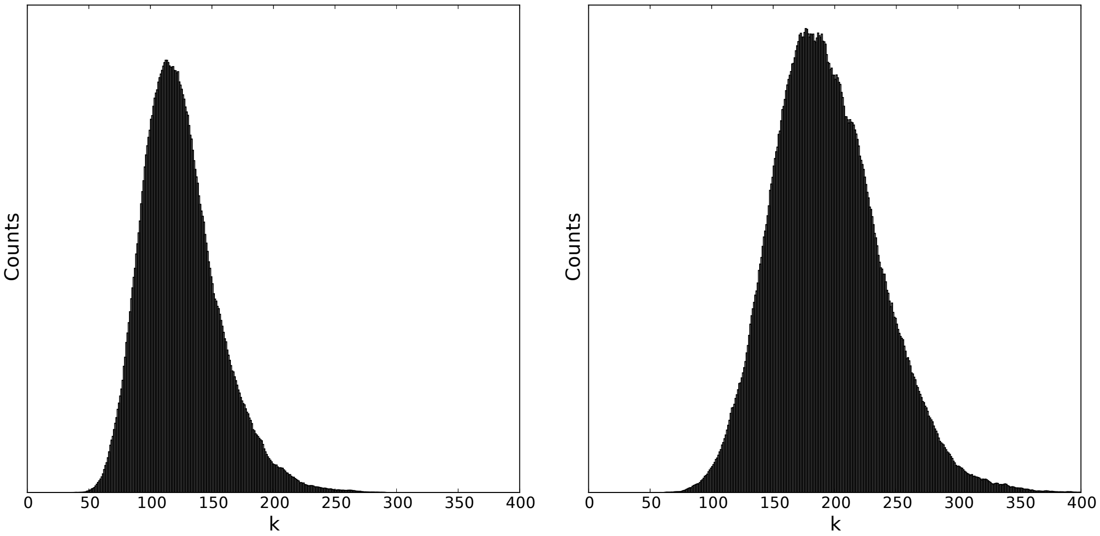
The compute time for the last 1 million steps for each independent chain was recorded and averaged to obtain an estimate of the relative computational cost of each of the parameterisations. The computed times are shown in Table 3.1.

Since the Voronoi parameterisation is grid free, comparing the cost of integrating travel times along ray paths will depend on the sampling rate along the ray paths. To ensure equivalency, as much as possible, of the two methods in terms of forward model accuracy, the ray paths were sampled at approximately the upper limit of grid resolution used by the wavelet parameterisation. As a 128 by 128 grid was used in a 20 by 20 degree region, this sampling spacing was approximately 0.16 degrees.

In the tree based wavelet parameterisation, the forward model cost is dominated by the inverse wavelet transform [Mallat, 1999]. As a general rule, a smoother wavelet will require more computational effort in the transform. For the Daubechies 6 wavelet, the standard Discrete Wavelet Transform (DWT) whereas both the Haar and CDF 9/7 used the Fast Lifted Wavelet transform [Sweldens, 1996, Daubechies and Sweldens, 1998]. This explains the relatively poor performance of the Daubechies 6 parameterisation. It is possible to use a lifted transform version of the Daubechies 6 wavelet in which case the expected time for this transform would lie between that of the Haar and the CDF 9/7 transform. However the number of active coefficients does factor into the computational time as evidenced by the fact that the Haar computational time is greater than that of the CDF 9/7 transform (this is reversed in other examples presented in Section 3.8, where more coefficients are needed by the CDF 9/7 parame-

(a) CDF 9/7

(b) Daubechies 6



(c) Haar

(d) Voronoi

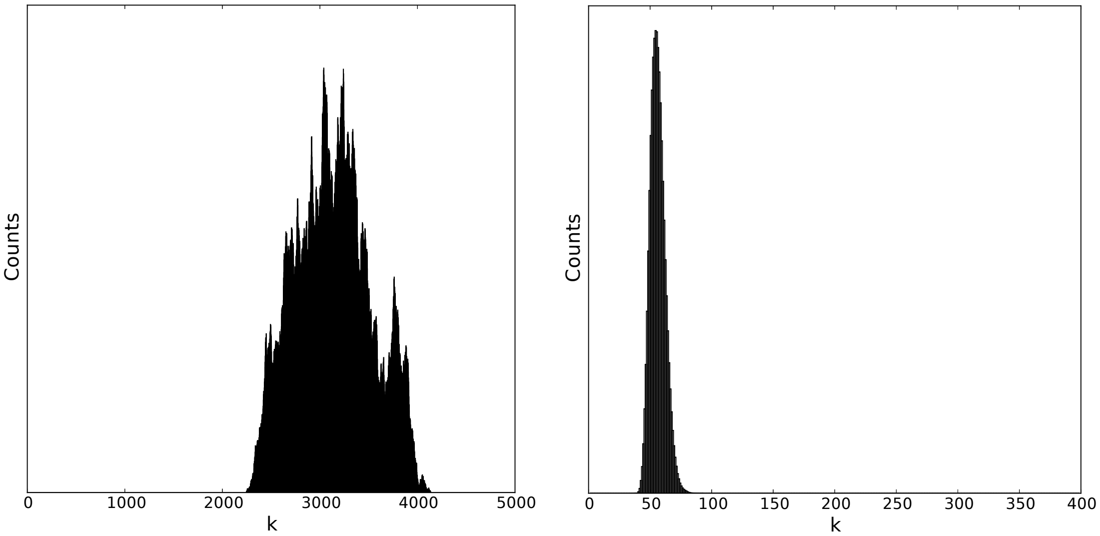


Figure 3.23: The estimated posterior probability distribution on the number of nodes/cells for the different parameterisations from the cosine checker board test.

Parameterisation	Time (s)	Relative Time
Haar	2452.8	1.4
CDF 9/7	1760.7	1.0
Daubechies 6	4735.7	2.7
Voronoi	30684.8	17.4

Table 3.1: Mean computational time per 1 million steps for cosine checker board model.

terisation).

Taking the median of the tree based wavelet parameterisation compute times, for these examples the Voronoi parameterisation is roughly an order of magnitude slower. These synthetic tests have relatively few coefficients. As the complexity of the models increase, the Voronoi parameterisation scales in computational effort as $\mathcal{O}(\log n)$, with n the number of cells, in the best case. In contrast, the dominant cost in the forward model of the trans-dimensional tree wavelet parameterisation, the inverse wavelet transform, is independent of the number of coefficients, suggesting that the wavelet method will scale better to more complex and larger scale tomographic problems.

3.7.6 Convergence

Monitoring convergence is notoriously difficult in McMC. In the trans-dimensional case, measures such as the Gelman-Rubin statistic [Gelman and Rubin, 1992a] are not applicable. In this work it is assumed that the independent Markov chains have converged when the negative log likelihood has reached an equilibrium value consistent with the data and errors. This is sufficient for the simulated problems here but robust convergence metrics for trans-dimensional sampling is an area of further research.

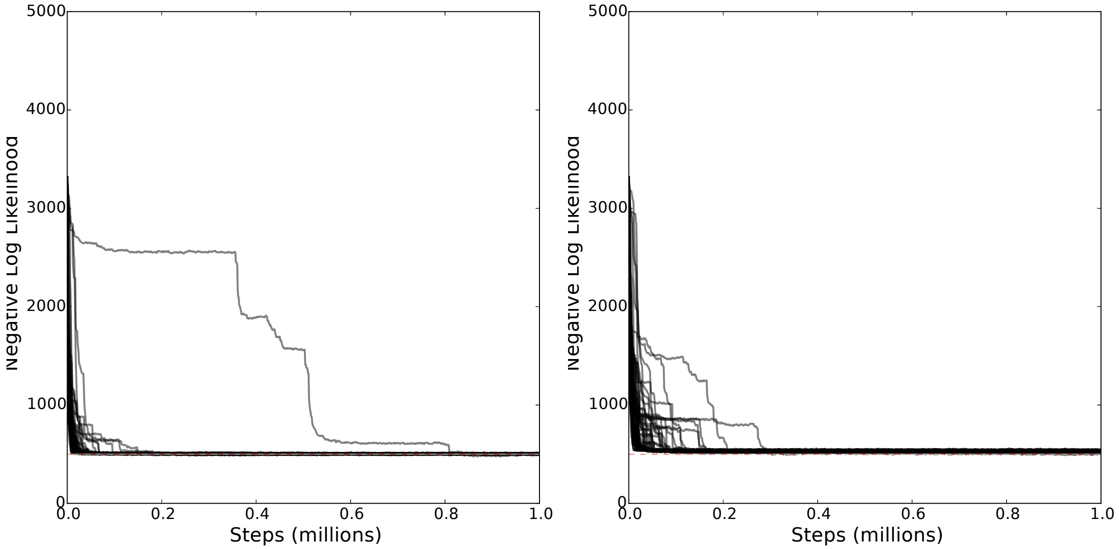
The evolution of the negative log likelihood of each Markov chain is plotted in Figure 3.24 for the first million steps. The trans-dimensional tree wavelet parameterisation has lower variability in the log likelihood across the chains and in some cases convergence has been achieved after a relatively small number of steps.

One reason for this is that, in general, the acceptance rates for a birth or death proposal is typically higher in the trans-dimensional tree wavelet parameterisation than for the Voronoi parameterisation. The acceptance rates were approximately 10 percent for the tree based wavelet method and around 5 percent for the Voronoi method. Hence a birth proposal is approximately twice as likely to be accepted in the tree based wavelet parameterisation than the Voronoi parameterisation.

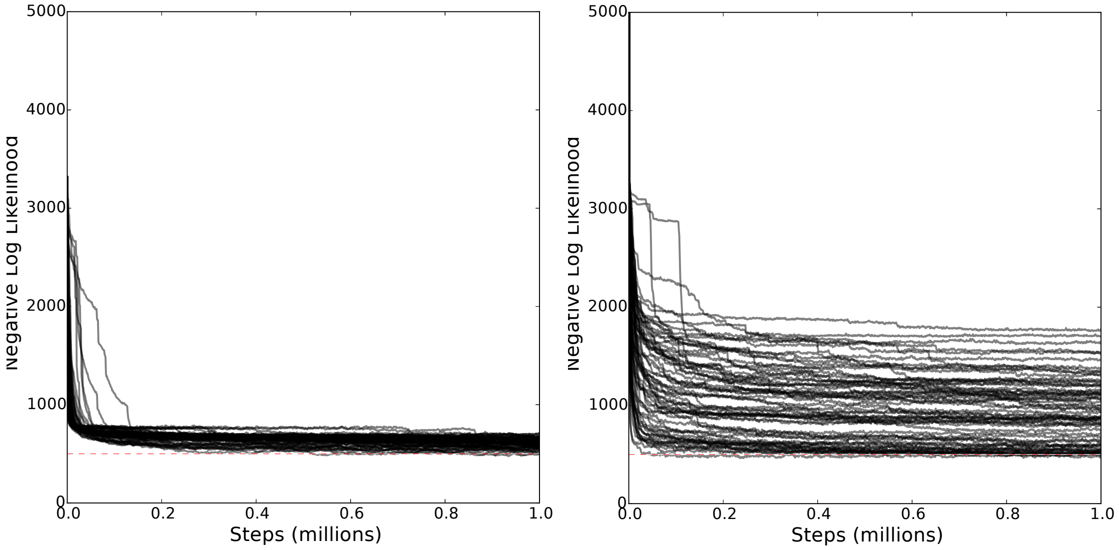
It is a common criticism of trans-dimensional samplers that the acceptance rates for the birth/death proposal are generally quite low and therefore the convergence is hindered due to lack of mixing between model spaces [Denison et al., 2002]. It is this higher acceptance rate for birth/death proposal that results in the faster convergence of the trans-dimensional tree wavelet parameterisation.

This higher acceptance rate is a result of the tree structure coupled with a multi-scale basis. To explain why this is the case, with a Voronoi parameterisation, the order of the births of its cells doesn't matter. In contrast, with a trans-dimensional tree model where the ordering of the birth does matter as a parent node must be birthed before its child nodes. In a multi-scale parameterisation such as wavelets, this means that coefficients that represent broad scale features will be birthed first, and often well constrained, before finer scale features. It also means that from any particular model, any birth will be at a scale length that is appropriate to refining the model rather than wasted on large scale feature changes.

In the case of the two smooth wavelet bases, the rapid convergence and small spread of the negative log likelihood values suggests that these tree based wavelet parameterisations have efficiently explored the parameter space. This implies that large numbers of independent chains, as is needed in the Voronoi based approach, may be less important with the wavelet parameterisation, given an appropriate choice of basis. In Figure 3.25 a comparison of the mean and MAP models of all chains combined is shown compared to a that of an arbitrary single chain. The single chain that was chosen was the chain with the largest minimum likelihood, notionally the worst performing chain. As can be seen in the figure, even the "worst" chain is barely distinguishable from the overall mean.

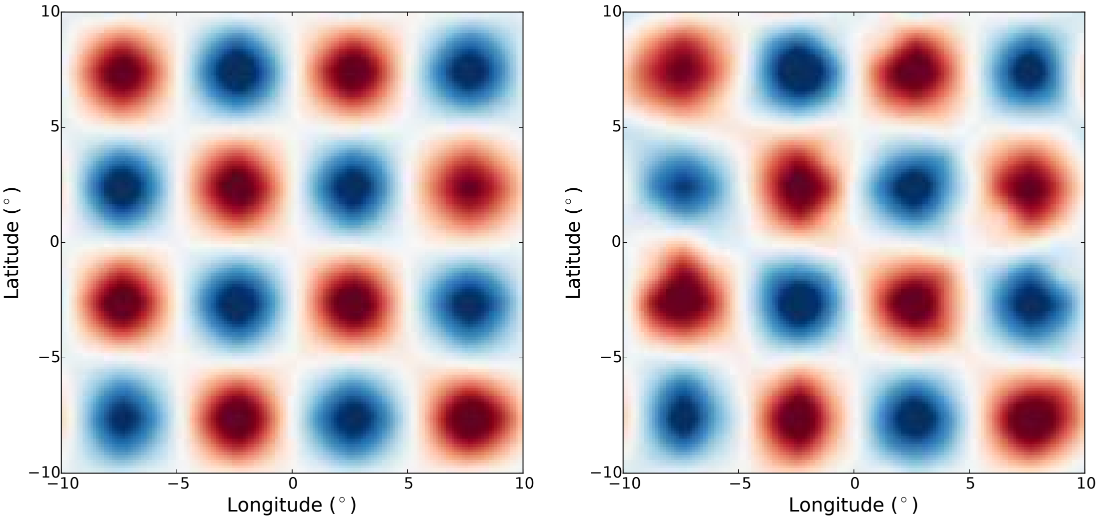

(a) CDF 9/7

(b) Daubechies 6


(c) Haar

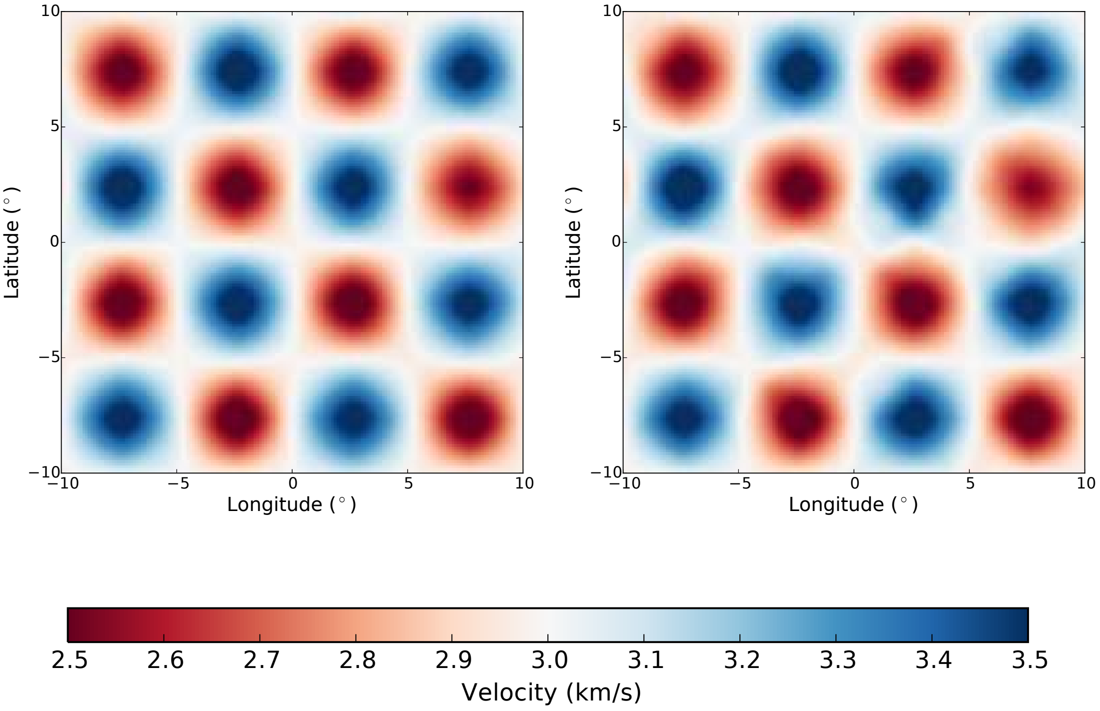
(d) Voronoi

Figure 3.24: For each of the parameterisations compared, the history of the negative log-likelihood is plotted for each of the 64 chains for the first 1 million steps during the recovery tests of the cosine checker board model.



(a) CDF 9/7 mean

(b) CDF 9/7 MAP



(c) CDF 9/7 single chain mean

(d) CDF 9/7 single chain MAP

Figure 3.25: For the CDF 9/7 parameterisation recovering the cosine checker board model, even the "worst" performing Markov chain of the 64 parallel chains obtains results comparable to the overall ensemble solution. In (a) and (c) are shown the mean of the ensemble of the 64 chains and the mean of the single "worst" performing chain respectively. In (b) and (d) are shown the over all best Bayesian maximum a posteriori (MAP) and the MAP model of the "worst" performing chain to show that they contain many similar features.

One of the primary reasons for employing multiple chains in the Voronoi parameterisation is to improve robustness of the chain by averaging. In the Voronoi cell case this is the only way to obtain a more plausible result for ambient noise tomography. The results of these experiments have shown that with a trans-dimensional tree based method, and an appropriate choice of wavelet basis function, multi-chain averaging may be unnecessary. Hence with the new approach it suffices to employ a smaller number of Markov chains, although with more complex and non-linear problems, parallel interacting chain approaches such as Parallel Tempering [Earl and Deem, 2005, Dettmer and Dosso, 2012, Dosso et al., 2012, Sambridge, 2014] may be necessary to adequately overcome local modes and multi-modalities.

3.7.7 Model Comparisons

With the new trans-dimensional tree approach, there is flexibility in the choice of basis function. With this flexibility comes the problem of determining the best basis to use for a given problem. To compare the results of different parameterisations, in synthetic tests an error norm from the “true” model such as the mean squared error (MSE) can be used. One issue with this approach is it does not take into account model complexity and therefore may prefer over fitted models. A second issue is that in real inversions, the “true” model will not be available with which to compare.

Therefore a flexible model comparison criterion is required. A direct comparison between the trans-dimensional tree wavelet approach and the Voronoi method using the Bayesian Information Criteria (BIC) [Schwarz, 1978] is difficult due to the already alluded to issue of fairly estimating the number of parameters in the tree based wavelet parameterisation. Proposed here is the use of the Deviance Information Criteria (DIC) [Spiegelhalter et al., 2002] that has previously been applied in trans-dimensional model comparison by Steininger et al. [2014]. A variation of the DIC proposed in Gelman et al. [2004, Chapter 12] is used in this study, where the DIC is computed as

$$\mathrm{DIC} = \overline{D(\mathbf{m})} + \mathrm{var}(D(\mathbf{m})), \tag{3.51}$$

where the over bar represents the mean, and $D(\mathbf{m})$, the deviance, is given by

$$D(\mathbf{m}) = -2\log p(\mathbf{d}|\mathbf{m}) + 2\log f(\mathbf{d}). \tag{3.52}$$

$f(\mathbf{d})$ is a normalising function of the data which cancels out in model comparison applications and can be ignored when computing the DIC. This formulation is preferred because in trans-dimensional sampling, point estimates can be over parameterised and from experience, using the variance results in a more stable calculation.

The first term in (3.51) rewards a low mean negative log likelihood which penalises too simplistic an ensemble of models. The second term penalises model complexity since more unknowns tend to result in ensembles with larger likelihood variance [Spiegelhalter et al., 2002, Gelman et al., 2004]. A model is said to be a better fit to the data if it has a lower DIC value. The results of the DIC are shown in Table 3.2. The DIC has been computed across all Markov chains (i), with just a single best chain (ii), and across all chains early in the simulation (iii) (for steps 750,000 to 1,000,000).

The DIC results confirm earlier subjective visual comparisons of the mean of the ensemble (Figure 3.22) to the true input models (Figure 3.21) where the CDF 9/7 and Daubechies wavelet parameterisations had recovered the true model more accurately. Note that the mean deviance of the Voronoi parameterisation is less than that of the CDF 9/7 parameterisation implying a better fit to the data. This is an example where using the misfit alone for model comparisons is insufficient. Previously it was shown in Figure 3.23 that the number of parameters in the wavelet parameterisation was higher, suggesting over-fitting. However the DIC shows low variance of the deviance in the wavelet case suggesting a smaller number of effective parameters.

The results of computing the DIC across all chains and a single chain are similar. The DIC was also computed during the last quarter of the first 1 million steps representing

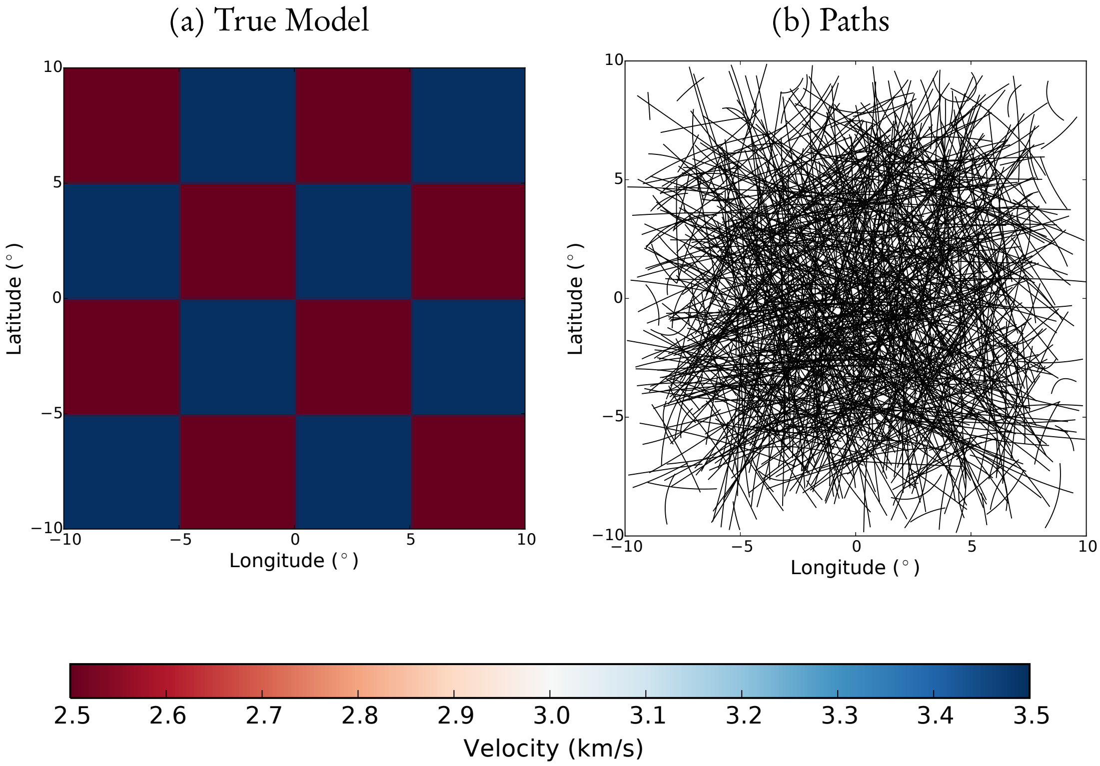
Parameterisation	$\overline{D(\theta)}$	var($D(\theta)$)	DIC
(i) *All chains*			
CDF 9/7	9280.2	82.7	9321.6
Daubechies 6	9294.7	256.4	9422.9
Haar	9232.2	481.4	9472.9
Voronoi	9207.9	540.5	9478.2
(ii) *Best chain*			
CDF 9/7	9274.4	149.8	9349.3
Daubechies 6	9282.9	194.2	9380.0
Haar	9212.1	528.4	9476.3
Voronoi	9191.2	566.1	9474.2
(iii) *Steps 750,000 to 1,000,000*			
CDF 9/7	9283.0	252.7	9409.3
Daubechies 6	9336.2	571.3	9621.8
Haar	9509.7	10817.3	14918.4
Voronoi	10193.4	467600.2	243993.5

Table 3.2: The DIC of the various parameterisations from the cosine checker board recovery test.

the tail end of the burn-in period. In these results there is a great deal more variance, particularly for the Voronoi parameterisation, and these results clearly show the more rapid convergence of the trans-dimensional tree approach in this problem.

With the new trans-dimensional tree wavelet method there is now the ability to choose from a variety of bases. Although prior knowledge of the expected heterogeneity of the tomography can be used to guide the choice of basis, this choice will necessarily be based on incomplete knowledge. A potential solution is to run a sweep of inversions with different basis functions and then compute the DIC (or similar criteria) of the obtained ensembles.

An alternative, which is beyond the scope of this work, is to select the wavelet bases in a hierarchical fashion itself using a trans-dimensional sampler. In this way the choice of basis could be driven by the data.


(a) True Model

(b) Paths

Figure 3.26: The boxcar synthetic models used in the test with seismic velocities between 2.5 and 3.5 km/s. We generate 1,000 random ray paths through the region from which we integrate travel-times to obtain the synthetic observations to which we add Gaussian noise of approximately 5 percent relative to mean travel times.

3.8 Boxcar Checker board Results

In Section 3.7.2, the results of a simulated smooth checker board 2D tomography test were presented. Here the same set of tests are repeated with a discontinuous boxcar checker board with the true model shown in Figure 3.26.

3.8.1 Ensemble mean solutions

The ensemble mean solutions for the boxcar checker board input model are shown in Figure 3.30. The Haar wavelet basis has recovered the input model almost exactly and the Voronoi parameterisation has also performed well. Both of the smooth wavelet parameterisations have recovered the underlying model to a lesser degree and have

Parameterisation	Time (s)	Relative Time
Haar	1742.1	1.0
CDF 9/7	2211.4	1.3
Daubechies 6	5222.6	3.0
Voronoi	19140.1	11.0

Table 3.3: Mean computational time per 1 million steps for the boxcar checker board model.

ringing artefacts. This is due to a property of wavelets where the number of coefficients required to represent discontinuities increases as a basis becomes smoother.

3.8.2 Number of Model Parameters

The histogram on the number of parameters for the boxcar checker board tests are shown in Figure 3.31. For the CDF 9/7 and Daubechies wavelet inversion of the boxcar checker board, the number of coefficients required to get a poorer representation of the model is substantially larger than the other two methods. This is to be expected as the representation of hard edges with smooth wavelets requires many coefficients. Also this then becomes a more challenging search problem to find these larger number of important coefficients and to sample them sufficiently, resulting in a lengthier convergence time.

3.8.3 Computational Time

The computational time for the boxcar checker board tests are shown in Table 3.3. The ordering is the same as for cosine checker board simulation with the Haar and CDF 9/7 parameterisations reversed due to the small but not insignificant computational burden resulting from a large number of coefficients.

Figure 3.27: Daubechies 6

(a) CDF 9/7

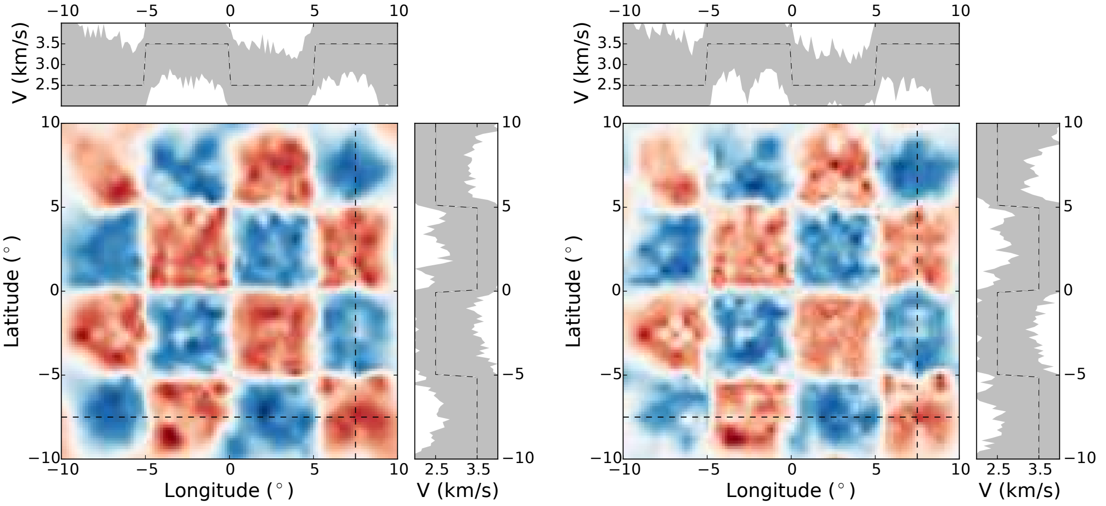


Figure 3.28: Haar

Figure 3.29: Voronoi

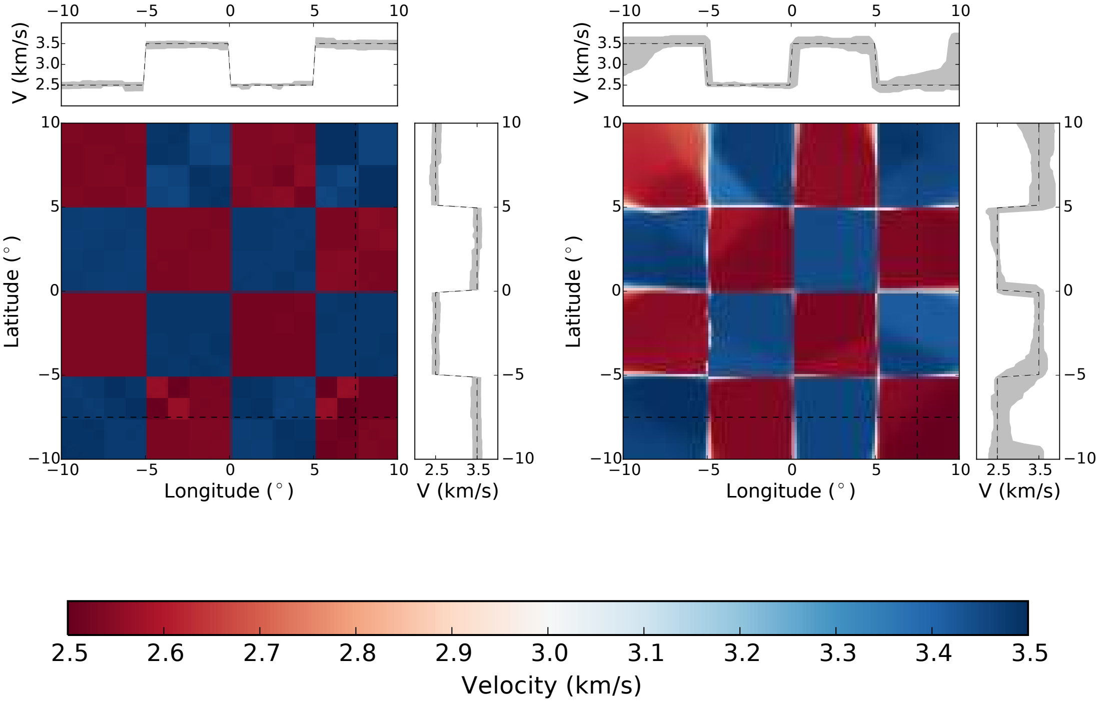
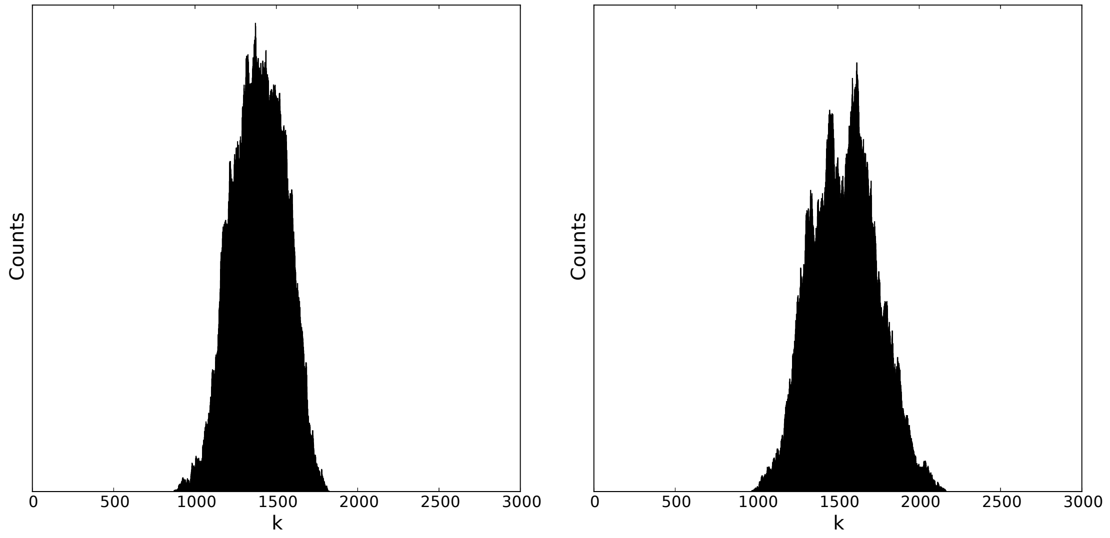
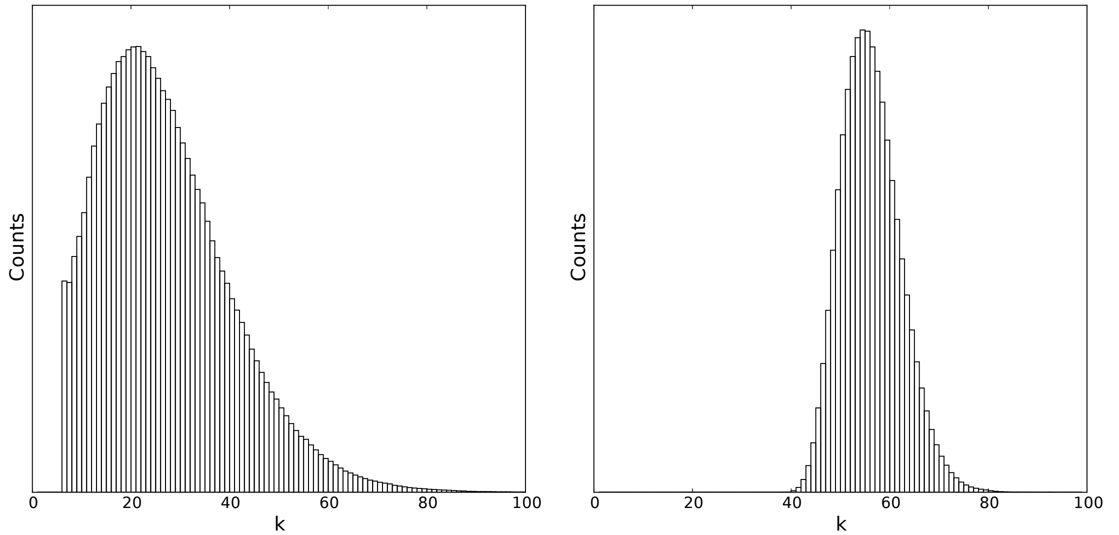


Figure 3.30: The mean of the ensembles obtained for the four different parameterisations used for the boxcar checker board input model.


(a) CDF 9/7

(b) Daubechies 6


(c) Haar

(d) Voronoi

Figure 3.31: The estimated posterior probability distribution on the number of nodes/cells for the different parameterisations from the boxcar checker board test.

3.8.4 Convergence

The evolution of the negative log likelihood of each Markov chain for the boxcar checker board test is plotted in Figure 3.32. The spread in likelihoods of the Voronoi parameterisation is noticeably larger than that of the trans-dimensional tree approach, even for wavelet bases that are not a good match for this input model.

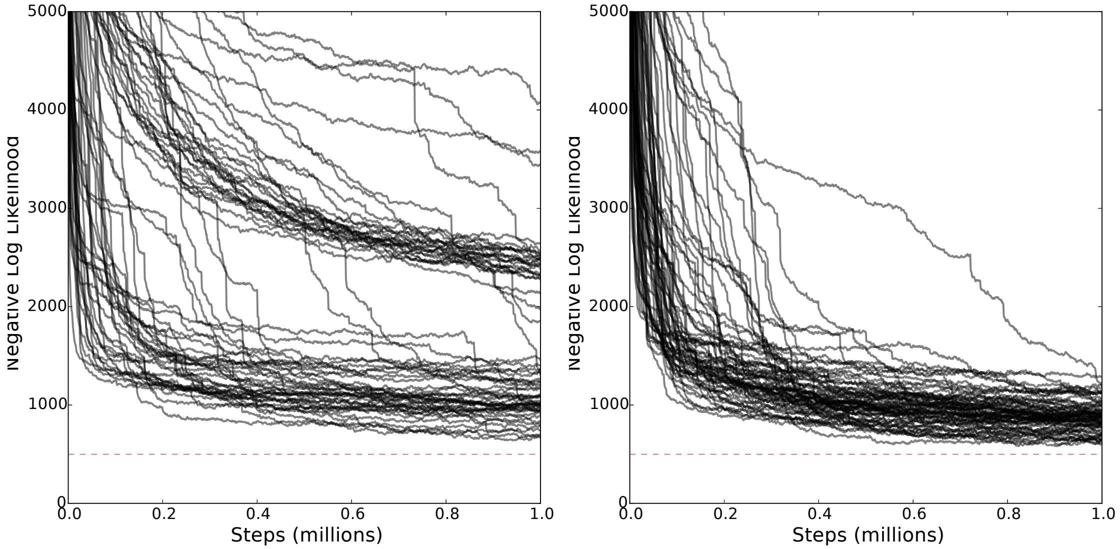
3.8.5 Model Comparisons

The DICs for the various parameterisations for the boxcar checker board tests are shown in Table 3.4. The DIC clearly favours the Haar wavelet representation in this case. It is also interesting to note that the DIC for the Haar parameterisation is almost exactly the same across all chains after 10 million steps as it is during the steps 750,000 to 1,000,000, implying convergence has been reached very quickly. Again the Voronoi parameterisation has the lowest deviance and therefore best overall fit, but is penalised by the variance of the deviance. The other two smooth wavelet parameterisation perform more poorly as expected.

3.8.6 Conclusions

In Sections 3.7.2 and 3.8 the wavelet parameterisation was observed to obtain better results, depending on the choice of wavelet basis, across a series of metrics. For the target application envisaged for this approach, namely ambient noise tomography, the smooth wavelets such as CDF 9/7 seem a logical choice and should provide better performance, both in terms of computational cost and recovery of the underlying model. It is worth bearing in mind though, that physical discontinuities may be better recovered using a Voronoi cell approach, but it is unclear whether these can truly be detected in ambient noise studies.

(a) CDF 9/7 (b) Daubechies 6



(c) Haar (d) Voronoi

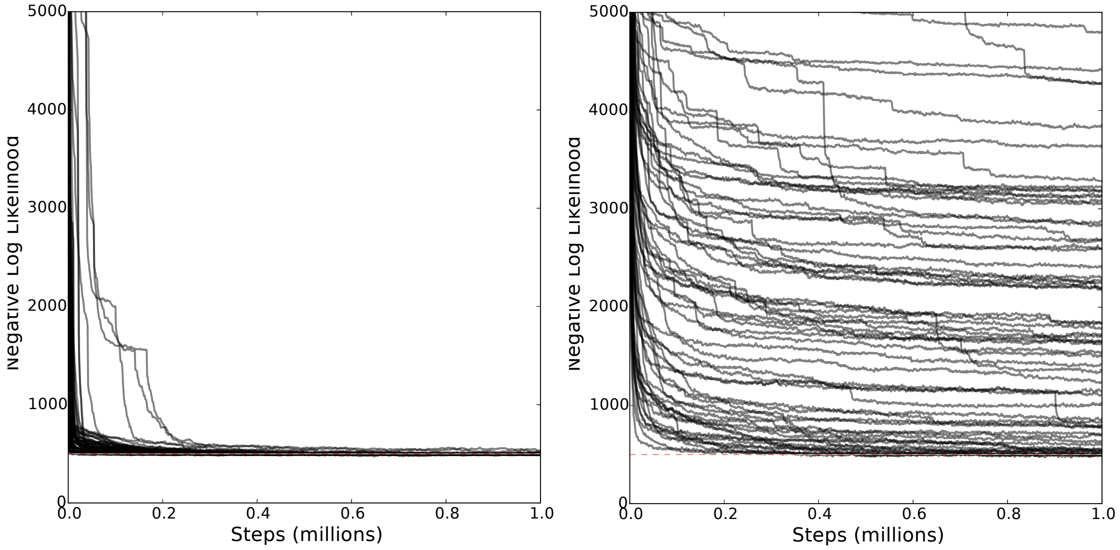


Figure 3.32: For each of the parameterisations compared the history of the negative log-likelihood is plotted for each of the 64 chains for the first 1 million steps during the recovery tests of the boxcar checker board model.

Parameterisation	$\overline{D(\theta)}$	var($D(\theta)$)	DIC
(i) *All chains*			
CDF 9/7	9273.9	1262.4	9905.1
Daubechies 6	9292.0	1583.5	10083.8
Haar	9269.9	64.2	9302.0
Voronoi	9269.2	284.5	9411.4
(ii) *Best chain*			
CDF 9/7	9273.5	1050.2	9798.6
Daubechies 6	9253.0	1367.1	9936.6
Haar	9268.5	76.5	9306.7
Voronoi	9257.0	219.8	9366.9
(iii) *Steps 750,000 to 1,000,000*			
CDF 9/7	11727.4	3090233.3	1556844.0
Daubechies 6	10106.9	145477.7	82845.7
Haar	9272.3	332.8	9438.7
Voronoi	12056.0	5641905.0	2833008.5

Table 3.4: The DIC of the various parameterisations from the boxcar checker board recovery test.

3.9 3D Teleseismic Tomography

For a more substantive test of the new trans-dimensional tree framework, it is applied to the teleseismic inversion of body waves to recover 3D lithospheric structure. The inversion result and ray paths published in Rawlinson et al. [2011] are used of a large scale regional area centred around Victoria, Australia.

To construct simulated data for the inversion, a Gaussian filter is applied on the model obtained by Rawlinson et al. [2011] to remove streak artefacts and use this as the “true” model. The model is then embedded as a deviation from the AK135 Earth reference model [Kennett et al., 1995] in the region of interest, shown in Figure 3.33, and 19897 of 19922 of the original paths are re-integrated to obtain true travel-times (some paths were removed as they were outside the region of interest). Gaussian noise is then added with a standard deviation of 0.5 seconds which corresponds to an approximately 1 percent error on the average travel time through the region. As in the earlier 2D

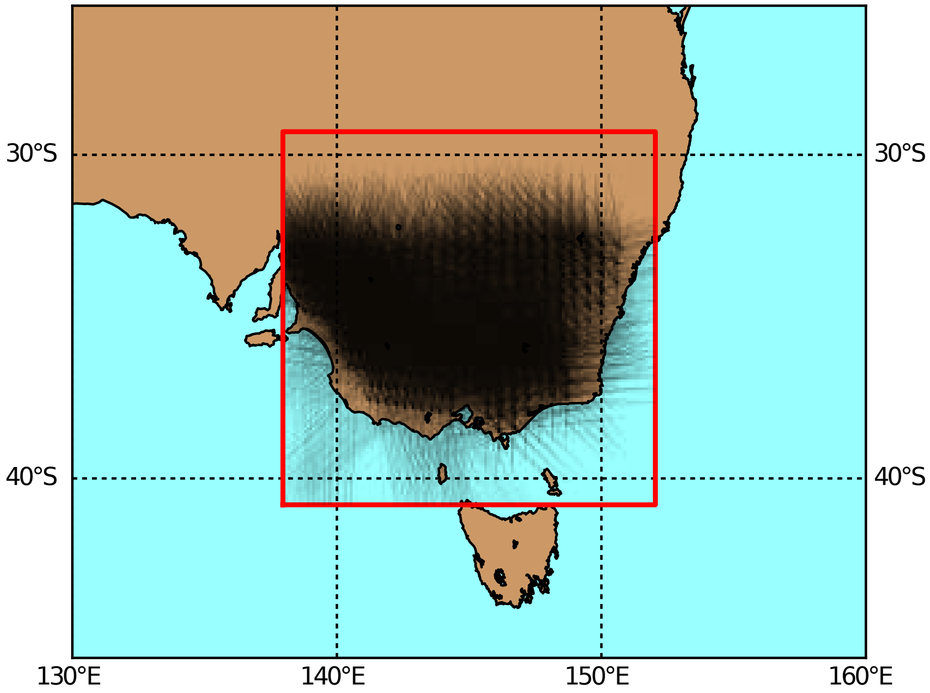


Figure 3.33: The teleseismic paths clipped to the 3D region (red rectangle) used in the inversion are sourced from the published study of Rawlinson et al. [2011]. There are 19897 body wave ray paths in the region of interest located in South Eastern Australia.

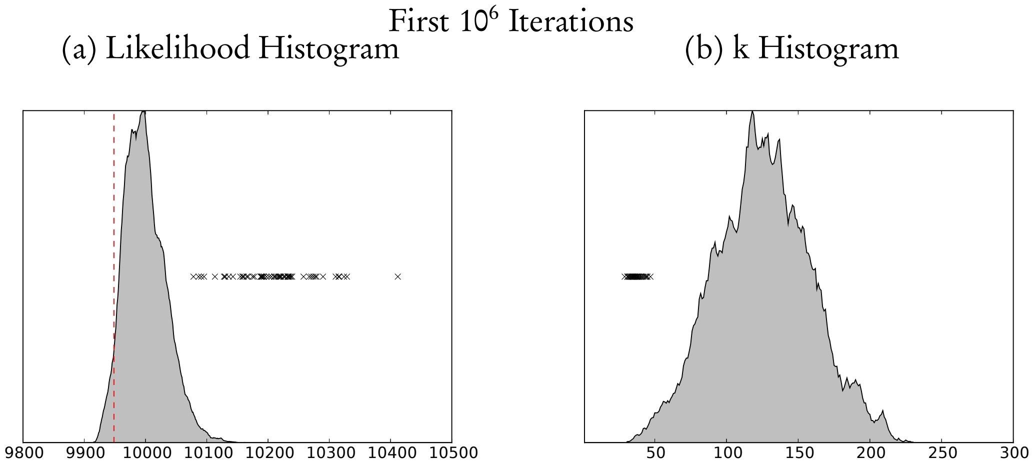
experiments, the problem is effectively linearised by using fixed ray paths.

The parameterisation used for the inversion of this region mostly follows that of the ambient noise tomography example shown earlier. A grid is set that is 128 longitude cells by 128 latitude cells by 32 radial cells to represent the region, this equates to nearly cubical voxels of approximately 10 km size. In this problem the 3D rectangular region requires a tree starting with a 4 by 4 subdivision grid laterally, that is 16 children from the root of the tree, and then progresses to the standard 3D wavelet tree consisting of 7 children from these subdivision nodes and 8 children thereafter (recall that in the 2D case this was 3 children from the root node and 4 thereafter).

To start from a single node of a tree as in the 2D case would likely take a long time to burn in. To accelerate this process, a simple stochastic optimisation scheme is used to generate an initial starting model. The outline of this simple method is as follows: the initial model is set to homogeneous with only the mean component or root of the tree active and the BIC [Schwarz, 1978] is computed. From this model, a large number of candidate models are generated using birth proposals to perturb the current model by adding randomly generated tree nodes. For each of the candidate models, their BICs are computed and the proposed model with the smallest BIC is selected as the next current model. The process is iterated until none of the BICs of the candidate models is less than that of the current model, that is, no improvement was made. The BIC in this case is used to prevent over optimisation of the initial model as the desire here is for simple but good starting models.

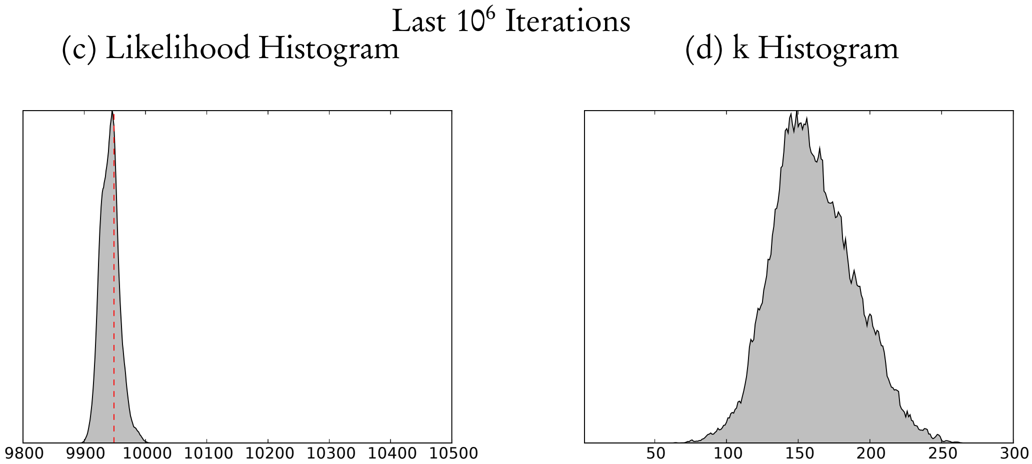
In the tree based parameterisation, it is also possible to restrict the height that the tree is allowed to sample. Given the grid is approximately 10 km on edge, we can equate each depth of the tree of 7 levels with an approximate length scale: that is level 0 represents the overall mean of the model velocity variations, level 1 represents scale lengths of 320 km, level 2 scale lengths of 160 km and so on down to level 7 which corresponds approximately to a 10 km scale length. As an additional restriction, an initial height restriction at level 5 is set (levels 6 and 7 unavailable) so the optimisation scheme only generated models with scale length features down to approximately 40 km. The height restriction is an optional feature of the trans-dimensional tree method that may be used to improve convergence in higher dimension and problems of greater complexity.

60 independent models are generated using the optimisation scheme and from these starting models, 60 independent Markov chains are simulated for 2 million steps. For the first 1 million steps, the height restriction remained in place but was removed for the last million. In Figure 3.34 the spread of the negative log likelihood and number of coefficients generated from the optimisation, the histograms of these during the first million steps with the height restriction, and the last million steps once this restriction is removed are shown. In the negative log likelihood plots, shown with a red dashed line is the theoretical χ^2 limit given the number of data and the level of independent Gaussian noise.


First 10^6 Iterations

(a) Likelihood Histogram

(b) k Histogram


Last 10^6 Iterations

(c) Likelihood Histogram

(d) k Histogram

Figure 3.34: The histogram of likelihood and k, the number of wavelet coefficients, of all Markov chains for the first 10^6 *iterations with the height restriction in place are shown in (a) and (b) respectively. Over plotted with crosses are the spread of likelihoods and k generated via the optimisation scheme for the initial models. The histograms for the last* 10^6 *iterations are shown in (c) and (d). In the likelihood plots, the vertical red dashed line represents the theoretical* χ^2 *limit of the data. These plots illustrate the convergence of the likelihood and the number of coefficients through the three phases we used during the inversion.*

The benefit of the height restriction is that it allows broader scale features to converge before sampling of fine scale features commences. Figure 3.34 shows how the negative log likelihood decreases from the initial models, which are clearly too simple. In the first million steps the algorithm resolves only medium scale features due to the height restriction. In the last million steps, the chains converge to the target posterior and the likelihood distribution is tightly focused on the theoretical χ^2 limit of the data. Similarly for the posterior on k, the number of coefficients, starts from a relatively small number in the optimisation phase and converges to a higher number in the final 1 million steps.

The modal number of wavelet coefficients is approximately 150 whereas the entire 3D image consists of 524,288 coefficients. The reason for this is a combination of two factors: firstly the model is smooth and the CDF 9/7 wavelet basis is used and is good at compressing smooth images, that is the CDF 9/7 wavelet basis is able to represent smooth images with few non-zero coefficients or sparsely. Secondly, the trans-dimensional tree method samples about a range of coefficients that adequately describes the information in the observations, resulting in a parsimonious number of coefficients.

The time taken for this simulation is approximately 15 hours in total with 1 hour required for the optimisation phase and 7 hours for each of the 1 million steps (Intel Xeon CPU E5-2620 at 2.10 GHz). This equates to approximately 25 ms per iteration. A comparable inversion with the Voronoi parameterisation has not been performed. Other studies that have performed 3D trans-dimensional tomographic inversion such as the work of Piana Agostinetti et al. [2015] have reported running times of approximately one month, however their inversion included hypo centre re-locations and ray-path updates which adds significant computational complexity so a direct comparison is not meaningful. It is to be expected that, as with the 2D case, there would be approximately an order of magnitude decrease in computational time for a single chain in the trans-dimensional tree approach when compared to the Voronoi parameterisation for the same scale of problem.

In Figure 3.35 the ensemble mean results of the volume with lateral slices at varying depth are shown. Similar to the 2D results earlier, for each depth the uncertainty along lateral transects indicated by the dashed line is shown. In the uncertainty plots, the shaded grey region shows the 95% credible range, the solid line the true input model, and the dotted line the ensemble mean along the transect. In general, the inversion has achieved a good recovery of the true input model, but there are some cases where the true model does not entirely reside in uncertainty bounds.

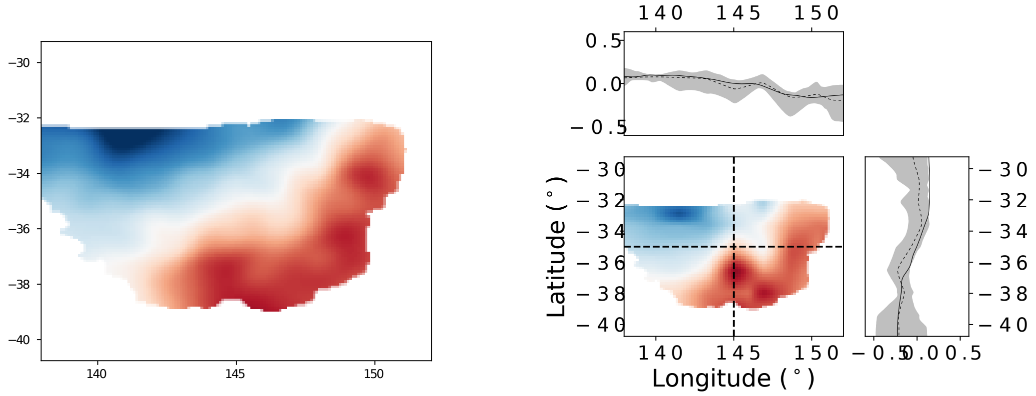
Similarly, in Figure 3.36, two slices of the ensemble mean volume longitudinally and latitudinally are shown to demonstrate the recovery as a function of depth. Again the recovery is quite good and the algorithm has not introduced any noticeable streaking artefacts due to the highly anisotropic ray distribution. Of minor concern is that a feature of both of the plots is a subtle underestimation of the velocity perturbation at the deepest part of the model. This is most likely a result of the poor resolvability of features at depth inherent in teleseismic data sets.

These results show that the tree based wavelet parameterisation can be used for large scale 3D geophysical tomography problems. Further work on the parallelisation or domain decomposition of the wavelet transform, coupled with parallel evaluation of the likelihood (i.e. the integration along the ray paths to obtain the model predicted travel times), would likely improve performance further.

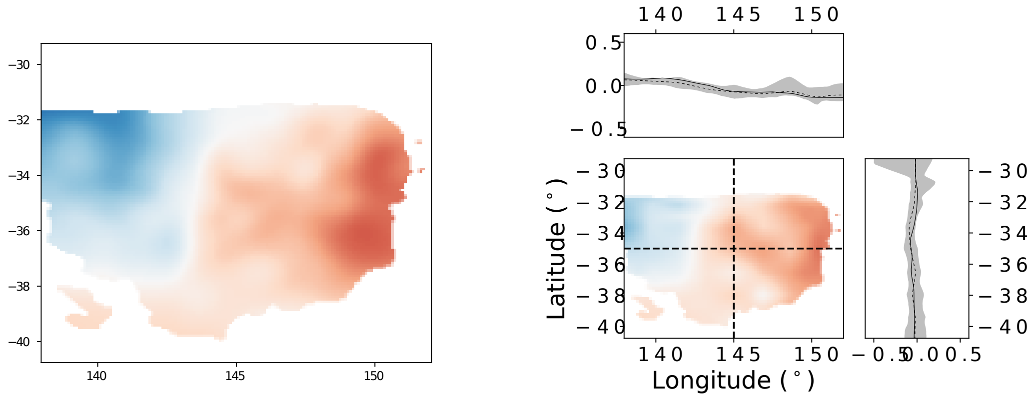
3.10 Summary

In this chapter, a new trans-dimensional framework for solving general image based geophysical inverse problems has been presented that is both efficient and flexible. This new approach is efficient because of three main factors: the first is that models are transformed back to regular grids which enables efficient forward model processing. Secondly, the approach can take advantage of existing fast algorithms such as the Fast Lifted Wavelet Transform for building the Earth models from the trans-dimensional tree representation. Lastly, the tree based approach is inherently multi-scale and there-

(a) True 100 km (b) Mean 100 km



(a) True 200 km (d) Mean 200 km



(a) True 300 km (f) 300 km

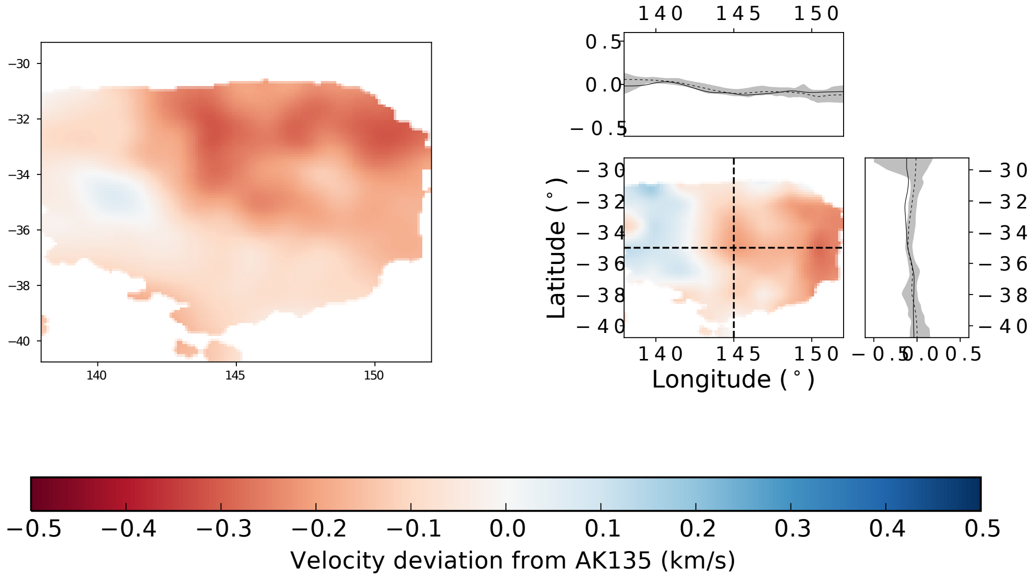
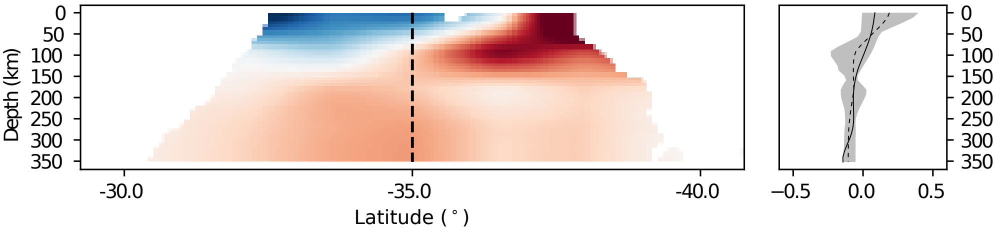


Figure 3.35: At three different depth slices we show the true model in the left and the ensemble mean model is shown on the right. Regions of no ray coverage masked out. In each of the ensemble mean plots we also show uncertainties along transects indicated with the dash lines. In the uncertainty plots, the grey region represents the 95% credible interval, the ensemble mean along the transect is shown with a dotted line, and the true model with a solid line. Generally, the true model falls close to the ensemble mean and is within the uncertainty bounds indicating good recovery in this simulation.

(a) Latitudinal Slice at 145^o



(b) Longitudinal Slice at 35^{o}S

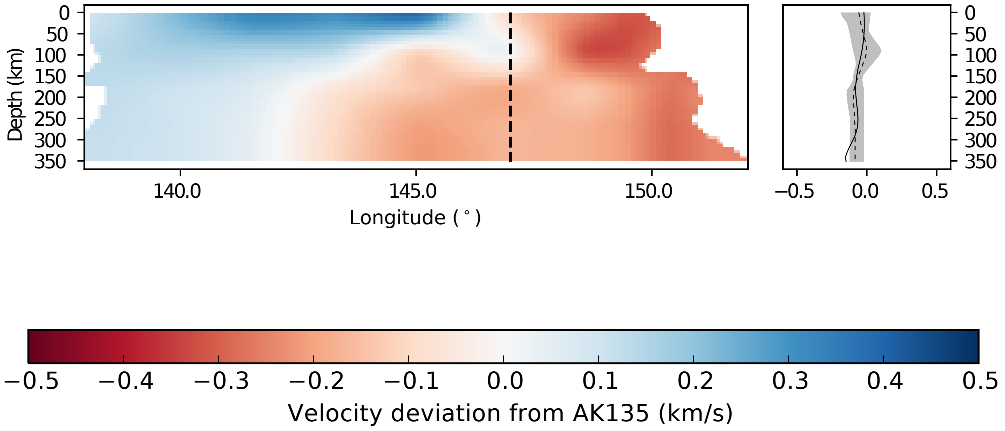


Figure 3.36: The ensemble mean model is shown with a slices along lines of constant longitude in (a) and latitude in (b). In each plot uncertainties are shown along transects indicated with the dash lines. In the uncertainty plots, the grey region represents the 95% credible interval, the ensemble mean along the transect is shown with a dotted line, and the true model with a solid line.

fore constructs models in a top down, coarse to fine scale, fashion, which is particularly well suited to many geophysical imaging problems.

The trans-dimensional framework is flexible because it allows a wide variety of basis functions to be used for representing Earth models, while performing all sampling operations on a common tree structure. We have shown examples of simple boxcar basis functions and wavelet bases, however, more advanced bases can be used such as higher order orthogonal polynomials, curvelets [Candes and Donoho, 1999], and wavelets on the sphere [Schröder and Sweldens, 1995, Leistedt et al., 2013a] or other orthogonal bases.

From the results presented in this chapter, the trans-dimensional tree approach appears to show promise in the probabilistic inversion of large scale geophysical inverse problems including robust uncertainty estimates.

Chapter 4

2D Airborne Electromagnetic Inversion

4.1 Introduction

In the previous chapter, the trans-dimensional tree algorithm was introduced and its abilities to provide robust results and uncertainties in geophysical inverse problems demonstrated in synthetic experiments. In this chapter, the trans-dimensional tree algorithm is used to invert for 2D conductivity profiles for airborne electromagnetic observations in a real world application.

Airborne electromagnetic methods (AEM) [Palacky, 1993] represent a class of non-invasive methods that allow the inference of the distribution of the electromagnetic properties of the Earth's subsurface. AEM uses a transmitting loop and receiver coils towed or carried by an aircraft along flight lines. Either short pulses (time domain AEM) or a combination of different frequency sinusoids (frequency domain AEM) are transmitted from an altitude of between 30 to 150m. These primary electromagnetic fields, generally in a frequency range of 25 Hz to 130 kHz, induce eddy currents in the conductive materials of the Earth according to Maxwell's equations [Maxwell, 1881]. Receiver coils, located near or collocated with the transmitter, are then able to detect secondary electromagnetic fields emitted from conductive regions in the subsurface resulting from the induced eddy currents.

While other electromagnetic properties can be inferred, the most commonly resolved material property is that of conductivity. This has implications for mineral exploration as many ore bodies of economic metals typically have higher conductivity relative to their host rocks. In AEM data, these ore bodies show up as "bumps" in the amplitudes of induced secondary fields, thus motivating the development of AEM techniques as a rapid mineral exploration technique in the 1950s.

More recently, AEM techniques have been adapted to more general geological mapping, hydro-geological and various other subsurface detection applications [Ackman, 2003]. Example applications include ground water mapping [Sattel and Kgotlhang, 2004], salt water intrusion [Fitterman and Deszcz-Pan, 1998], and dry land salinity [Street et al., 1998, Lawrie et al., 2000].

Existing approaches for inferring the subsurface distribution of conductivity generally parameterise the subsurface as a set of 1D conductivity profiles (e.g. stitched 1D inversion), a 2D conductivity-depth section mesh (e.g., laterally constrained 1D inversion schemes and 2.5D inversions) or a 3D mesh (e.g., spatially constrained 1D and 3D schemes). The conductivity at each layer or cell is then solved for by minimising the misfit between observed responses and predictions generated by a forward model. To stabilise the inversion, one or more penalising model norms that impose restrictions on deviation from a reference model (damping) or spatial gradients (smoothing), are required [Tikhonov, 1943, Constable et al., 1987, Farquharson and Oldenburg, 1993, Brodie and Sambridge, 2006] resulting in a minimisation of misfit and model norms with different relative weightings. The relative weighting of the misfit between predictions and observations and the penalising norms need to be controlled with tuning parameters which introduce a degree of subjectivity. While various criteria and automatic methods of setting the hyper parameters are available [Hansen, 1992, Farquharson and Oldenburg, 2004] they are not without their problems [Hanke, 1996, Vogel, 1996].

One drawback of this regularised deterministic approach for inferring conductivity is that only a single estimate of the conductivity structure is recovered. This gives little appreciation of the potential non-uniqueness of the optimised solution or spread of plausible conductivity models that could give rise to the same observed responses resulting from the non-linear physical model and noise on the observations. While uncertainty can be estimated from model covariance matrices obtained from optimisation strategies, these give linearised estimates of uncertainty that can provide biased uncertainty estimates and are unable to properly quantify non-uniqueness of solutions [Menke, 1989, Tarantola, 2005].

In recent times, computing power has advanced sufficiently that ensemble techniques are becoming feasible for geophysical inversion problems. These commonly use Markov chain Monte Carlo (McMC) techniques [Brooks et al., 2011] to sample an *a posteriori* probability distribution based on Bayes' theorem [Bayes, 1763]. In this approach, an ensemble of plausible solutions to the inverse problem are obtained based on a likelihood function and an *a priori* probability distribution for the unknowns. From the posterior ensemble, likely candidate models can be extracted analogous to the optimal models obtained from traditional inversion techniques [Sambridge and Mosegaard, 2002]. In addition, the spread or variation of the ensemble can be statistically interrogated to obtain rigorous uncertainty estimates that can identify model parameterisation trade offs and non-uniqueness.

In general, the inversion results and corresponding uncertainty estimates can be adversely affected by poor choices in the formulation of the inverse problem, for example, in the choice of the number and thickness of layers. As a general rule, if the problem is under parameterised or too simple then the misfits between observations and predictions will be large and somewhat paradoxically, the uncertainties will be underestimated. This is commonly observed when the grid resolution is set too coarsely in tomographic problems. Conversely, if the problem is over parameterised, the model fit will improve yet the uncertainties will be overestimated due to increased degrees of freedom or trade-offs present between model parameters. This trade off between resolution and uncertainty is well known in geophysical inversion [Backus and Gilbert, 1968]. More generally in Bayesian statistics, this trade off is one of the motivations for the field of Bayesian model selection [Ando, 2010].

To rigorously estimate uncertainty in a geophysical inverse problem, the impact of model selection should be considered. Traditional approaches for when the number of candidate models is relatively small include Bayes Factors [Kass and Raftery, 1995] and their various approximate criteria [Akaike, 1974, Spiegelhalter et al., 2002, Ando, 2007].

An advance in McMC sampling techniques that allows dimensional changes, i.e. changes to model parameterisation, is Reversible Jump McMC [Geyer and Møller, 1994, Green, 1995], often called trans-dimensional sampling in the geophysics community. Since the introduction of the trans-dimensional approach to geophysics [Malinverno, 2002], it has been used in a wide variety of inverse problems, including Paleo-temperature reconstructions [Hopcroft et al., 2007] Ambient Noise Tomography [Bodin and Sambridge, 2009], Receiver Functions [Piana Agostinetti and Malinverno, 2010, Bodin et al., 2012b], Surface wave dispersion [Dettmer et al., 2012], Paleoplate motion reconstruction [Iaffaldano et al., 2014], Body wave tomography [Young et al., 2013, Piana Agostinetti et al., 2015], Earth's mantle viscosity [Rudolph et al., 2015], the relative rotation of the Earth's inner core [Tkalčić et al., 2013], Tsunami sea surface deformation [Dettmer et al., 2016], 1D AEM inversion [Minsley, 2011, Brodie and Sambridge, 2012, Minsley and Ley-Cooper, 2015], and lastly 2D AEM inversion [Ray and Key, 2012, Ray et al., 2014]. In trans-dimensional inversion, the model parameterisation is allowed to vary as part of the inversion process. Specifically, the complexity of the spatial features of the model are controlled by the information in the data implicitly through relative Bayesian evidence [Sambridge et al., 2006]. This prevents the ensemble from being dominated by too simple or overly complex model parameterisations and leads to a natural "parsimony" in model complexity, which in turn offers more comprehensive estimates of uncertainty as shown by Dettmer et al. [2016].

In this chapter, the application of trans-dimensional sampling to 2D image based inversion of time domain AEM data to demonstrate its benefits for both inversion, and estimates of uncertainty.

4.2 Method

4.2.1 Overview

The problem domain for the trans-dimensional inversion is a 2D region along a flight line recorded with time domain AEM observations. A standard approach is used for the inversion of the 2D region of the subsurface with one axis in the depth direction, and the other along the flight line or laterally. The region of interest is parameterised as a 2D image with one column of pixels per AEM sounding in the lateral direction and logarithmically spaced pixels in depth. The bottom most row of pixels represents

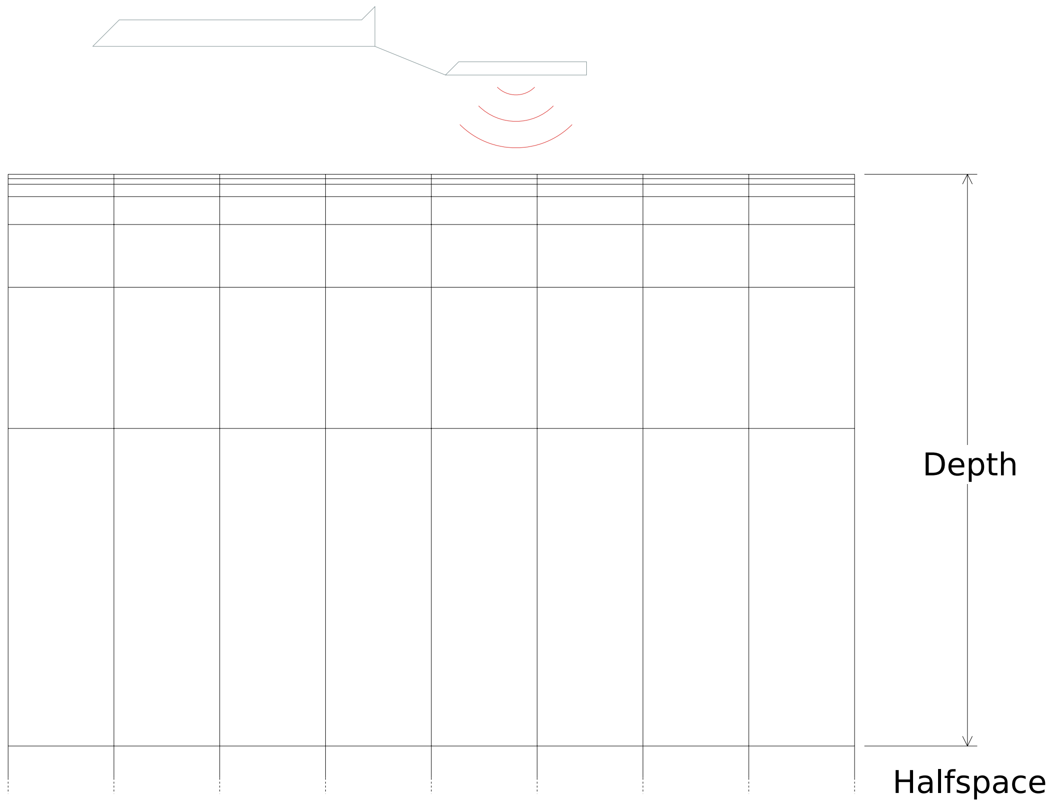


Figure 4.1: A schematic figure illustrating the parameterisation of the inverse problem beneath the flight line in an AEM survey. The conductivity of the subsurface is parameterised as a set of fixed conductivity grid cells with regular spacing in the horizontal direction and logarithmically spaced cells in the depth direction. The deepest row of conductivity cells represent the conductivity of the bottom half space.

the underlying half-space layer of the problem domain. An illustrative schematic of the parameterisation used in the inverse problem is shown in Figure 4.1.

Here the aim is to demonstrate the applicability of trans-dimensional sampling to time domain airborne electromagnetic inversion. In a sampling approach, the forward model that computes the predicted response from a candidate model needs to be evaluated on the order of 1 million times. From a practical point of view, this prohibits the use of 2D or 3D forward modelling with current computing resources. For this reason, this treatment is restricted to using the 1D forward model approach detailed in Brodie and Sambridge [2006] and Brodie [2010], available as open source software [Brodie, 2016].

Since a 1D forward modelling approximation is used to invert the 2D structure, for each AEM sounding an independent 1D forward model is run on the conductivity structure represented by the grid cells directly below the AEM sounding. This assumes that lateral heterogeneities along the flight line are smooth and do not introduce significant modelling errors that would require 2D or 3D forward modelling [Yang and Oldenburg, 2012].

4.2.2 The Bayesian Trans-dimensional Approach

The approach in this chapter uses Bayesian inference to assess probability density functions (PDFs) on model parameters representing conductivities of the subsurface. From these empirical PDFs, inferences on models of likely structure can be obtained using expected values, medians or modes. An additional benefit of this approach though is in being able to estimate uncertainties and non-uniqueness by examining the spread of the ensemble at each point of the model. The Bayesian approach [Brooks et al., 2011, Gelman et al., 2004] uses Bayes theorem,

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

where $\mathbf{m}$ is the vector of M model parameters, $\mathbf{d}$ is the N observed data, $p(\mathbf{m})$ is independent prior information on the model parameters (e.g. physical constraints on the range of plausible conductivities), $p(\mathbf{d}|\mathbf{m})$ is the likelihood function and $p(\mathbf{d})$ is a normalising term commonly referred to as the evidence. Since the time domain AEM problem involves a non-linear forward model, accurately estimating the evidence normalisation term is not feasible analytically, although numerical approximations are available [Skilling, 2006]. Fortunately, relative inferences without computing this term are sufficient for both model inference and uncertainty estimates.

In time domain AEM, the observations at each point consist of a response curve(s) representing the observed response of the secondary field from conducting bodies beneath the surface, an example of which is shown later. Under the assumption that a

Gaussian noise model accurately approximates the noise resulting from measurement and theory error, the likelihood function can be written as

$$p(\mathbf{d}_\mathrm{i}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^j |C_r|}} \exp\left\{-\frac{1}{2}(G(\mathbf{m})_i - \mathbf{d}_\mathrm{i})^T C_r^{-1} (G(\mathbf{m})_i - \mathbf{d}_\mathrm{i})\right\}, \tag{4.2}$$

where $\mathbf{d}_\mathrm{i}$ is the ith AEM sounding along the flight path, j is the number of time windows in the sounding, $G(\mathbf{m})_i$ is the predicted response as a function of the model parameters $\mathbf{m}$, and C_r is the covariance matrix representing the potentially correlated noise on the data.

Markov chain Monte Carlo (McMC) techniques are used to generate samples that converge to the target distribution, in this case, the posterior probability density (PPD), given by $p(\mathbf{m}|\mathbf{d})$ in (4.1). This is an iterative approach that perturbs the current model by sampling a proposal density function $Q(\mathbf{m} \to \mathbf{m}')$ to generate a new candidate model $\mathbf{m}'$. The new model is accepted, that is becomes the current model in the chain, or rejected meaning the previous model is retained, according to the Metropolis-Hastings [Metropolis et al., 1953, Hastings, 1970] probability rule

$$\alpha(\mathbf{m} \to \mathbf{m}') = \min\left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m}' \to \mathbf{m})}{Q(\mathbf{m} \to \mathbf{m}')}\right]. \tag{4.3}$$

The acceptance probability terms ensure correct convergence to sampling the posterior by maintaining "detailed balance" of the Markov chain(s) [Brooks et al., 2011]. The more general Metropolis-Hastings-Green [Green, 1995] acceptance criteria that includes model dimension changes is

$$\alpha(\mathbf{m} \to \mathbf{m}') = \min\left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m}' \to \mathbf{m})}{Q(\mathbf{m} \to \mathbf{m}')} |\mathcal{J}|\right], \tag{4.4}$$

where now $\mathbf{m}'$ may contain a different number of unknowns than $\mathbf{m}$, and the additional term $|\mathcal{J}|$ is the determinant of the Jacobian that represents the variable transformations that may occur when model dimension or parameterisation changes.

Following Hawkins and Sambridge [2015] and Chapter 3, the trans-dimensional tree approach is used with a wavelet parameterisation to represent the image based model. In this approach, the model $\mathbf{m}$ consists of a hierarchy of wavelet coefficients, from coarse scale to fine, that are trans-dimensionally sampled to reconstruct the subsurface distribution of conductivity. The benefits of this approach compared to simply sampling over all pixels is that the parameterisation can adapt to different scale length features. This in turn results in better constraint on the parameters of the inversion and more robust estimates of the uncertainty. Earlier trans-dimensional approaches exist that parameterise 2D regions of interest in terms of Voronoi cells, however the trans-dimensional tree approach with wavelet basis has been demonstrated to be more efficient for geophysical imaging problems, both in terms of computational time and convergence rates. The choice of wavelet basis also leverages the innate ability of wavelets to decorrelate and compress images meaning complex subsurface features can be represented with fewer parameters.

The operation of the trans-dimensional tree is briefly recapitulated here with full details in Hawkins and Sambridge [2015] and Chapter 3. In Figure 4.2, an example abstract tree of wavelet coefficients is shown on the left. In this schematic of the tree, active coefficients are shown as solid dots. Inactive nodes equate to having the corresponding wavelet coefficient set to zero. From top to bottom, each level corresponds to progressively finer structure. Through the application of the inverse wavelet transform using a chosen wavelet basis, this hierarchy of wavelet coefficients can be mapped into a conductivity image shown on the right. For the simulation studies presented, the bi-orthogonal wavelet basis commonly referred to as CDF 9/7 [Cohen et al., 1992] is used which provides good compression of information, as evidenced by its use in the JPEG 2000 image compression standard [Unser and Blu, 2003]. The image constructed from the model of wavelet coefficients is then used by the forward model to generate synthetic response curve predictions. These predictions are then compared with the observations in the likelihood function (4.2).

The trans-dimensional tree algorithm adds and removes nodes from the abstract tree.

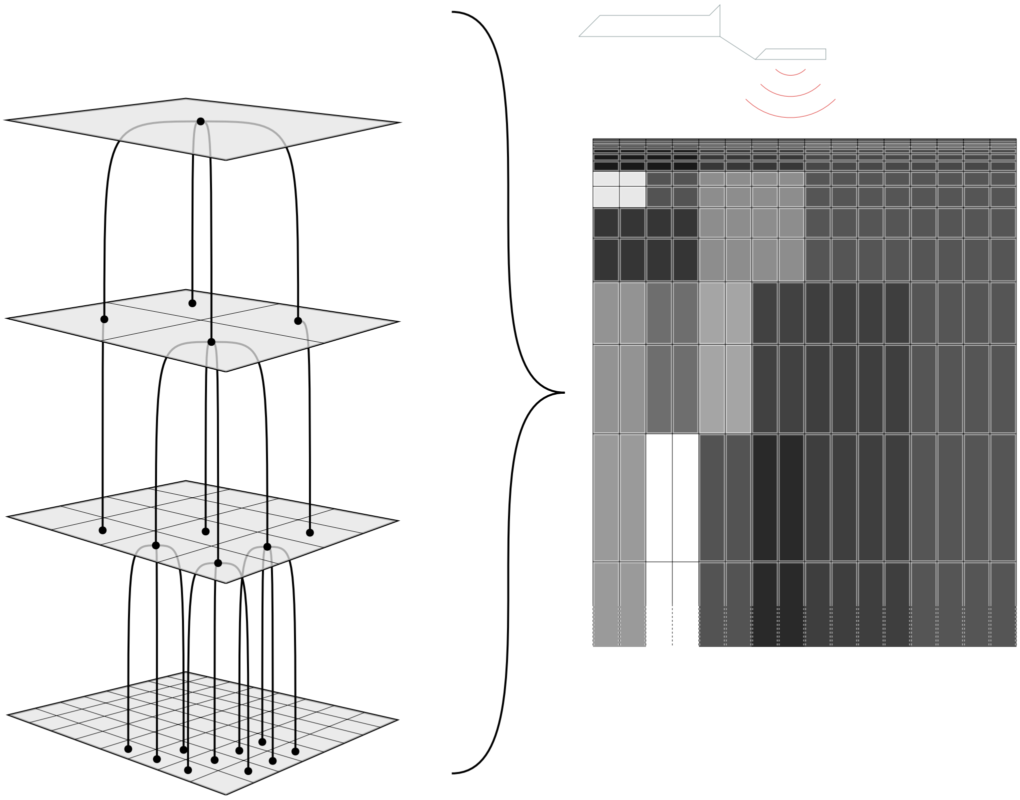


Figure 4.2: A cartoon illustration of the Trans-dimensional Tree method with wavelet parameterisation of the sub-surface conductivity. On the left the sub-surface conductivity is represented abstractly as a hierarchy of wavelet coefficients with different scale lengths. On the right is shown a corresponding sub-surface conductivity image illustrating how the trans-dimensional tree approach can adapt local to varying scale lengths of heterogeneity.

At each step of the Markov chain, the randomly selected perturbations of the current model are: add a new wavelet coefficient to the tree, remove a wavelet coefficient from the tree, or change the value of an existing wavelet coefficient. The probability of adding a new wavelet coefficient is set to the same as that of removing a wavelet coefficient to maintain detailed balance. In the general case, the starting model and the chain of models during convergence are often poor fits to data and are discarded as part of the "burnin" process. The remaining "chain" of candidate models then forms the ensemble from inferences can be made.

Common problems in sampling algorithms are poor convergence due to poor tuning of proposal distributions, sampling local minima due to non-linear effects and the related problem of the difficulty of sampling multi-modal posterior distributions. To overcome these issues, Parallel Tempering [Earl and Deem, 2005, Dosso et al., 2012, Sambridge, 2014] is used to more effectively explore the posterior space during inversion. In this approach, multiple Markov chains are run at different temperatures which reduce the influence of the likelihood in the modified acceptance criteria

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = \min \left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \left[\frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \right]^{\frac{1}{T}} \frac{Q(\mathbf{m}' \rightarrow \mathbf{m})}{Q(\mathbf{m} \rightarrow \mathbf{m}')} |\mathscr{J}| \right], \tag{4.5}$$

where T is the temperature. A set of logarithmically spaced temperatures with multiple chains at each temperature are run with statistical information collected from the set of Markov chains at a temperature of one. At higher temperatures, the influence of the likelihood ratio is reduced and this allows the high temperature chains to more actively explore the prior space. Periodically, model exchanges are attempted between chains at different temperatures which allows sharing of information about posterior regions of interest between chains. This results in better sampling of non-linear problems and more robust and effective sampling of the entire prior space to give greater confidence in the final results, that is a local minimum or a single modality in a multimodal posterior does not bias the results. Similar probabilistic Bayesian approaches have previously been reported [Rosas-Carbajal et al., 2014, Hauser et al., 2015], however in trans-dimensional sampling the observations are used to adapt the resolution as required instead of *a priori* fixed a global correlation length. Additionally, parallel independent Markov chains with parallel tempering are used in this study to more thoroughly explore the range of possible solutions.

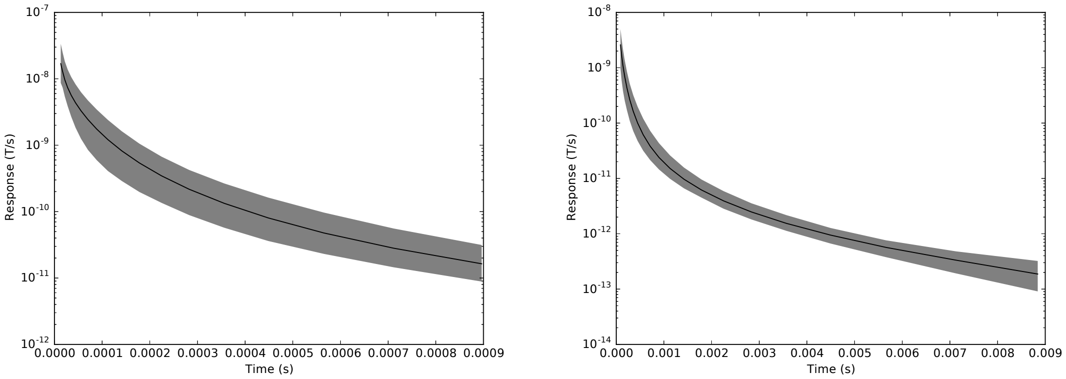
4.3 Application to Broken Hill Managed Aquifer Recharge (BHMAR) Project

4.3.1 Case study overview

This case study uses helicopter borne time domain AEM data acquired as part of the Broken Hill Managed Aquifer Recharge (BHMAR) project, a Geoscience Australia groundwater study. The project's aims were to investigate groundwater sources to help future proof the township of Broken Hill's water supply during times of drought and better manage water resources [Lawrie, 2016].

The AEM data were acquired in 2009 using a SkyTEM system [Sorensen and Auken, 2004] in which the transmitter loop and receiver coils were carried on a frame towed below a helicopter. Interleaved low moment (LM) and high moment (HM) data were acquired at 222 Hz and 25 Hz base frequency respectively. Off time data were recorded at delay window times shown in Tables 4.1 and 4.1. Mean response curves over a representative flight path for both moments are shown in Figure 4.3.

In time domain AEM, each response curve is measured in a series of time windows. As can be seen from Figure 4.3, and as evidenced by the necessity of a log scale for clarity, the response magnitude decays with time suggesting a non-stationary noise model is required. For the studies herein, an empirically derived non-stationary noise model is used where the noise is assumed to be independent Gaussian distributed and the variance is a function of the amplitude of the observed response and a constant background noise level



(a)

(a)

Figure 4.3: The mean of the response curves from a representative section of the BHMAR survey for the Low Moment and High Moment are shown with a solid black line in (a) and (b) respectively. The vertical scale is in log of the magnitude for clarity. The grey shaded regions represent the range of possible response curves over the studied flight line.

$$\sigma_i^2 = (r|d_i|)^2 + a_i^2, \tag{4.6}$$

where σ_i is the standard deviation of the noise on the ith time window, r is the relative level of noise as a scalar multiple of the observed magnitude $|d_i|$, and a_i is the standard deviation of the additive background noise considered constant for each window. The time windows and level of additive noise is shown for both the low and high moment signals in Tables 4.1 and 4.2. For the relative noise, a reference level of 3.6% was used and hierarchical sampling [Bodin et al., 2012a] is used to estimate a scaling term of this reference, i.e. $r = 0.036\lambda$ where λ is the hierarchical scaling term.

4.3.2 Simulation studies

To test the inversion of time domain AEM data with the trans-dimensional tree approach using a wavelet parameterisation, an initial synthetic experiment is conducted using a known true model. A random flight path was generated that included random walk behaviour to simulate changes in altitude and orientation of the transmitter/receiver assembly. Using this flight path, synthetic response curves were generated using the 1D forward model. Random noise was then added to the observations ac-

Window Start (μs)	Window End (μs)	Constant Noise ($\times 10^{-12}$)
11.39	15.00	57.76100
15.39	19.00	7.71540
19.39	24.00	5.78490
24.39	31.00	3.91640
31.39	39.00	3.15020
39.39	49.00	2.51050
49.39	62.00	2.29120
62.39	78.00	1.92100
78.39	99.00	1.73300
99.39	125.00	1.52900
125.39	157.00	1.22580
157.39	199.00	0.96876
199.39	250.00	0.90323
250.39	315.00	0.82181
315.39	397.00	0.74835
397.39	500.00	0.62648
500.39	630.00	0.62901
630.39	793.00	0.57157
793.39	999.00	0.51475

Table 4.1: The parameters for the time windows and additive noise used in the noise model for the SkyTEM system in this study for the low moment signal.

Window Start (μs)	Window End (μs)	Constant Noise ($\times 10^{-12}$)
78.39	99.00	0.255450
99.39	125.00	0.208150
125.39	157.00	0.191440
157.39	199.00	0.159200
199.39	250.00	0.145980
250.39	315.00	0.134020
315.39	397.00	0.127120
397.39	500.00	0.108440
500.39	630.00	0.102140
630.39	793.00	0.097184
793.39	999.00	0.090881
999.39	1258.00	0.084579
1258.39	1584.00	0.077776
1584.39	1994.00	0.069864
1994.39	2511.00	0.066747
2511.39	3161.00	0.059365
3161.39	3980.00	0.053300
3980.39	5011.00	0.048430
5011.39	6309.00	0.042199
6309.39	7942.00	0.037096
7942.39	9742.00	0.035710

Table 4.2: The parameters for the time windows and additive noise used in the noise model for the SkyTEM system in this study for the high moment signal.

cording to the noise model described in the previous section to create synthetic observations.

The inversion was started with a constant conductivity value of 0.25 S/m in all cells of the model, including the half space. The models could be initialised by sampling from the prior, for example, random models, or through some pre-conditioned or optimised model, for example, the results obtained using traditional regularisation techniques could be used as a starting point. However since the trans-dimensional tree is able to represent the conductivity model across multiple scales, the performance of the method is highlighted by initiating from a simple (homogeneous) structure.

The domain was parameterised as a region 200 metres deep to the half space layer with 16 pixels in depth (i.e. 15 layers plus the half-space layer) and 16 pixels laterally. For the priors, a uniform prior on the number of wavelet coefficients is set between 1 and 256. The prior on all wavelet coefficients was set to be uniform between -0.5 and 0.5.

Setting the prior on a wavelet coefficient is non-intuitive. A method of understanding the range of values an image can take given such a prior on the corresponding wavelet coefficients is to sample from the prior and examine the statistics of the resulting conductivity images. The result of this is shown in Figure 4.4 where this experiment is repeated with increasingly larger prior widths. As can be seen from the figure, a uniform range of between -0.5 and 0.5 gives a reasonable distribution of conductivities. Simple coefficient priors have been specified here but more advanced techniques using past inversions or training images would likely improve performance [Lochbühler et al., 2015].

The prior on the wavelet coefficients can have subtle effects on the final outcome, so a second inversion was run with double width priors to test the prior sensitivity, similar to Dettmer et al. [2016], with negligible observed differences. The prior on the hierarchical noise scaling parameter was set to a Jeffreys prior which expresses the lack of knowledge of the relative error scale [Jeffreys, 1939, Jaynes, 2003].

Four parallel chains with four temperatures logarithmically spaced with temperatures between 1 and 10 were used. Convergence was monitored by comparing the likeli-

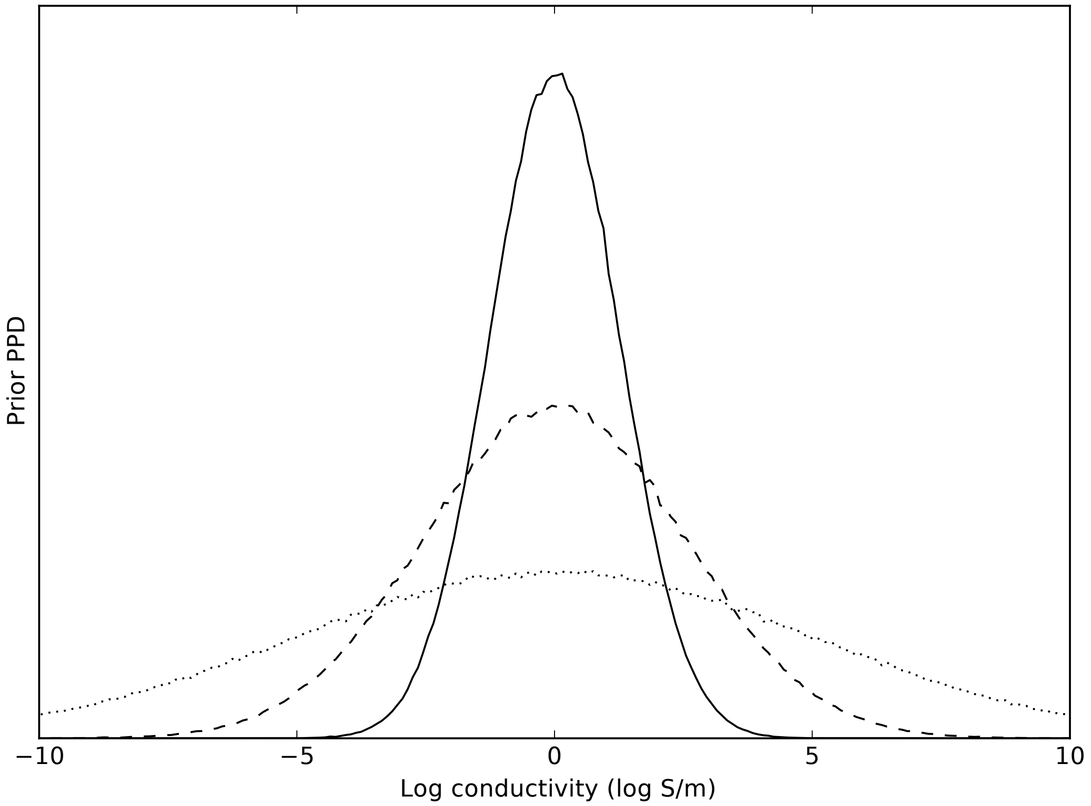


Figure 4.4: The histogram of a central point of the images generated from sampling from uniform wavelet coefficients with varying widths. The solid line corresponds to a uniform prior on the wavelet coefficients of -0.5 … 0.5 as used in this study, the dashed line -1.0 … 1.0 and the dotted line -2.0 … 2.0. The distributions of the conductivities given a prior on the wavelet coefficients is well behaved and follows a generalised Gaussian distribution centred on zero.

hood, the hierarchical noise scaling parameter, and the number of wavelet coefficients required by the parallel chains to ensure they are sampling about the same region. Qualitative indications of convergence are shown in Figure 4.5 where in (a) the un-normalised negative log likelihood (NLL), analogous to the misfit in optimisation regimes, is plotted against Markov chain step, while in (b) the hierarchical error scaling parameter and (c) the number of model parameters or wavelet coefficients is plotted against chain step.

In Figure 4.5(a) the negative log likelihood of each independent chain is correctly sampling about the same value. In this synthetic experiment there are 640 observations (16 lateral columns with 40 time windows for each column). When the noise model is a good match for the true noise, the un-normalised NLL will sample about a value around half the number of observations, that is 320 in this case. This value, defined as the χ^2 limit, is shown as a red dashed line in the Figure. Similarly, in (b) each of the independent chains the hierarchical error scaling parameter has converged to a value of approximately one indicating that noise model is correctly recovered. Finally, in (c) the number of wavelet coefficients has converged to approximately 20 and is much fewer than the maximum of 256 wavelet coefficients from a 16 x 16 grid. This illustrates the power of trans-dimensional sampling in that it automatically parsimoniously samples the number of model parameters to those required by the data which can improve both sampling efficiency as well as the inversion result itself.

In Figure 4.6 the true synthetic model is shown in (a) while the mean, median and mode of the ensemble are shown in (b), (c) and (d) respectively. It can be seen in the figure that these model estimates have generally recovered the structure quite well with the exception of the parts of the model below 100m where the models overestimate the conductivity. In (e) and (f) the top most 50m of the true and ensemble median model are displayed to highlight the good recovery of the true model nearer the surface.

The erroneous features at depth are to be expected, since resolving power of AEM systems decays with distance. The key point of this approach is that this decay in resolvability can be resolved in the posterior uncertainties. To illustrate that this is

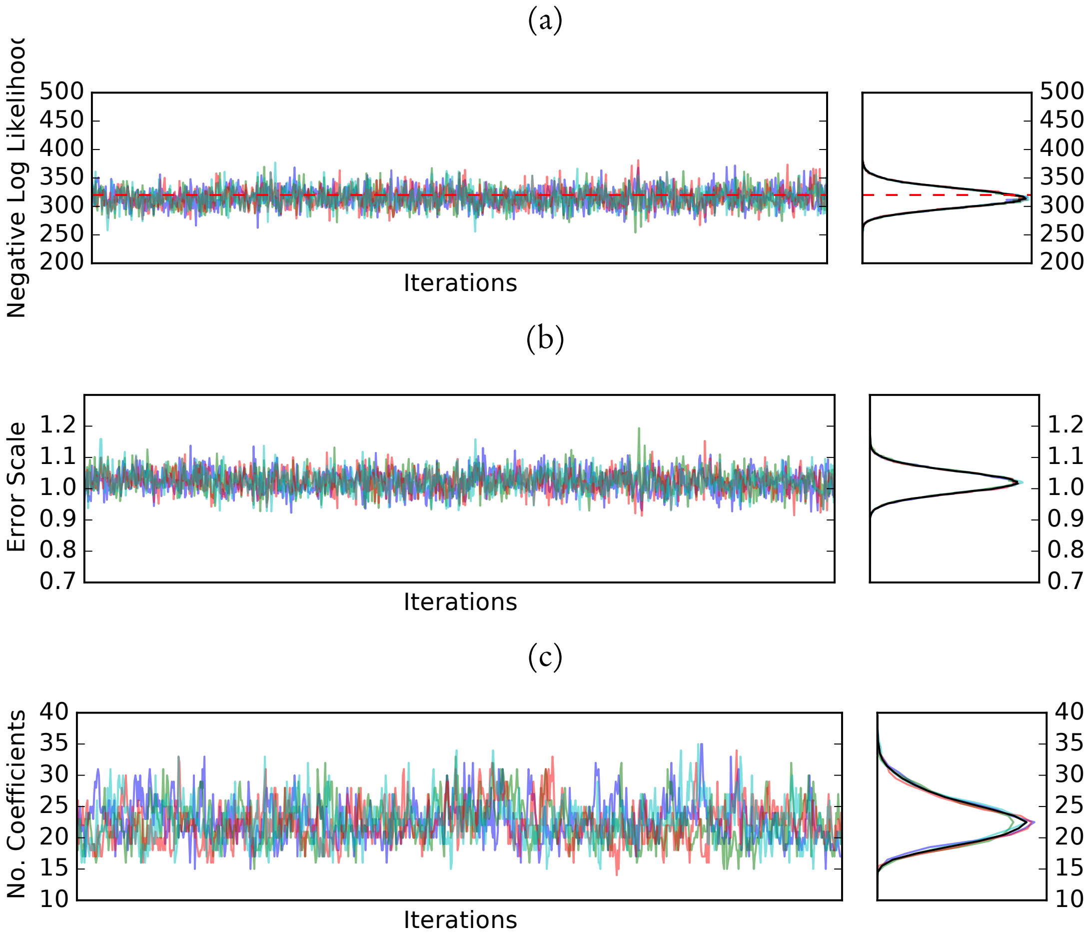


Figure 4.5: The history of 3 parameters that can be used as qualitative indications of convergence. These plots are monitored until each of the chains appears stationary about a similar level. Posterior samples are collected once this qualitative criteria is met. In (a) is the negative log likelihood, in (b) the hierarchical error scaling parameter, and in (c) the number of model parameters or wavelet coefficients required. In each of the plots, the history of the independent chains in different colours on the left and on the right is the posterior histograms with the black histogram representing the histogram of all chains combined. In (a), the red dashed line shows the χ^2 theoretical limit based on half the number of observations (640).

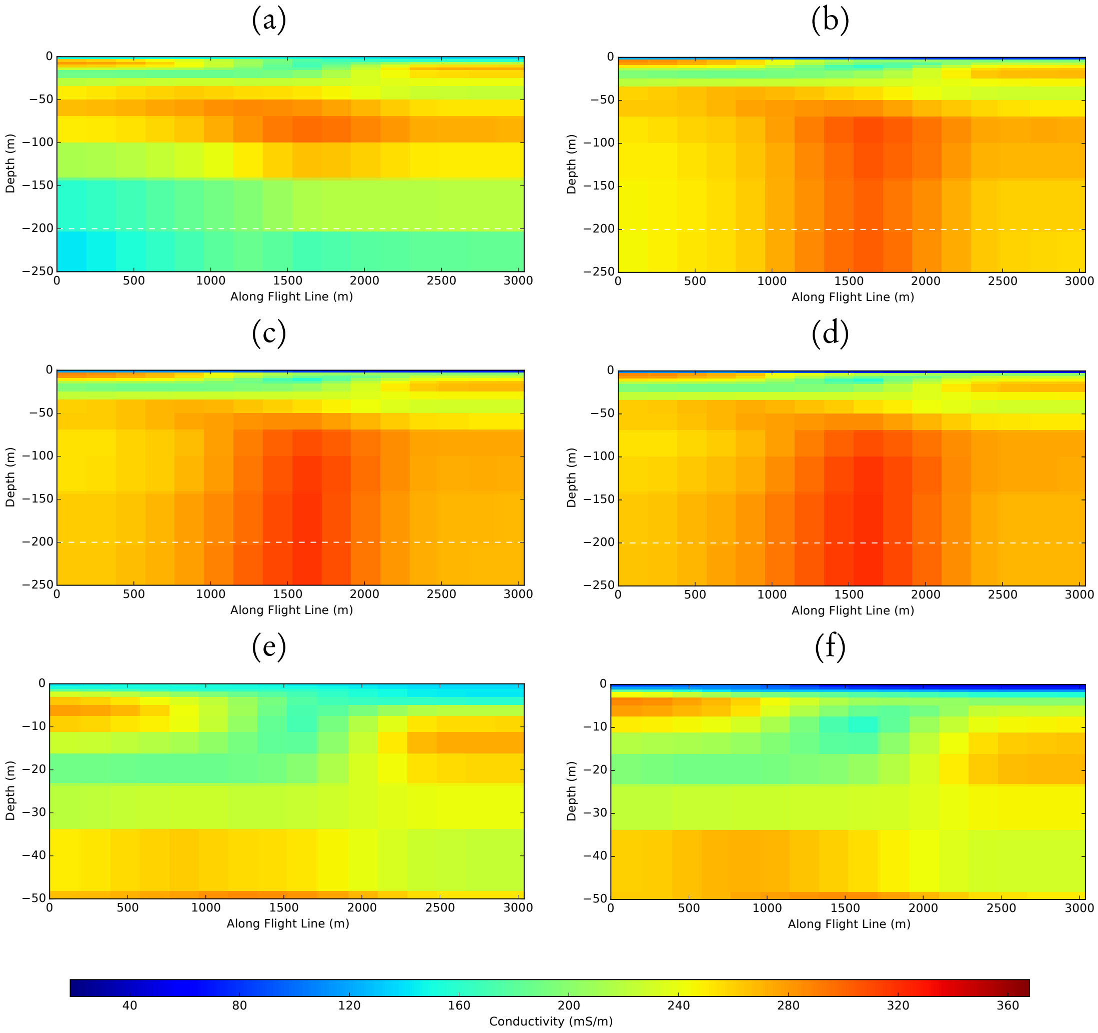


Figure 4.6: Conductivity shown as log S/m. (a) shows the simulated true model while (b) is the mean of the ensemble obtained from the trans-dimensional inversion. Similarly, (c) is the median and (d) is the modal model. The deep low conductivities in the true model are poorly constrained and not recovered by the inversion. Panels (e) and (f) show the top most 50m of the domain where the true model (e) and the median of the ensemble (f) are very similar. In the inversion, no explicit spatial smoothing is imposed. The lateral scales of variation recovered in the inversion results are due to the adaptive trans-dimensional parameterisation.

reflected in uncertainty results from the inversion, in Figure 4.7 shows both the magnitude of half the 95% highest posterior density (HPD) width [Hyndman, 1996] and two times the standard deviation for each conductivity cell in the region. The highest posterior density (HPD) width represents the minimum width (highest density) in log of conductivity that contains 95% of the ensemble models. At any point in the model, as an approximation the uncertainties are $\pm\frac{1}{2}$HPD about the mean, median or mode. For comparison, twice the standard deviation at each point is shown which also approximately corresponds to the range of 95% of the models in the ensemble. In general, the HPD interval is comparable to the standard deviation in magnitude, however this is not always the case. The standard deviation assumes the posterior is Gaussian in shape and this assumption is often violated, particularly in non-linear problems. Regardless, the trend in both measures of uncertainty show consistent increase of uncertainty with depth.

With this simulation study, it has been demonstrated that laterally varying conductivity models can be resolved with the combination of the empirical noise model, and the trans-dimensional tree sampling approach using a wavelet basis. The amplitude of the data noise is also estimated via inclusion of a hierarchical scaling term in the noise model.

4.3.3 Real data

In this study, part of a flight-line of time domain AEM data recorded as part of the BHMAR project is inverted. A small section of a flight line of approximately 3km was selected, which consisted of 32 LM and HM soundings. The inversion was initialised with all Markov chains so to a model with homogeneous conductivity of 0.25 S/m. The same empirical noise model as described earlier is used with hierarchical sampling to estimate an additional amplitude term.

The inversion was run on a cluster with 8 parallel tempered chains of the 8 temperatures each giving a total of 64 independent chains. The 8 temperatures were spaced

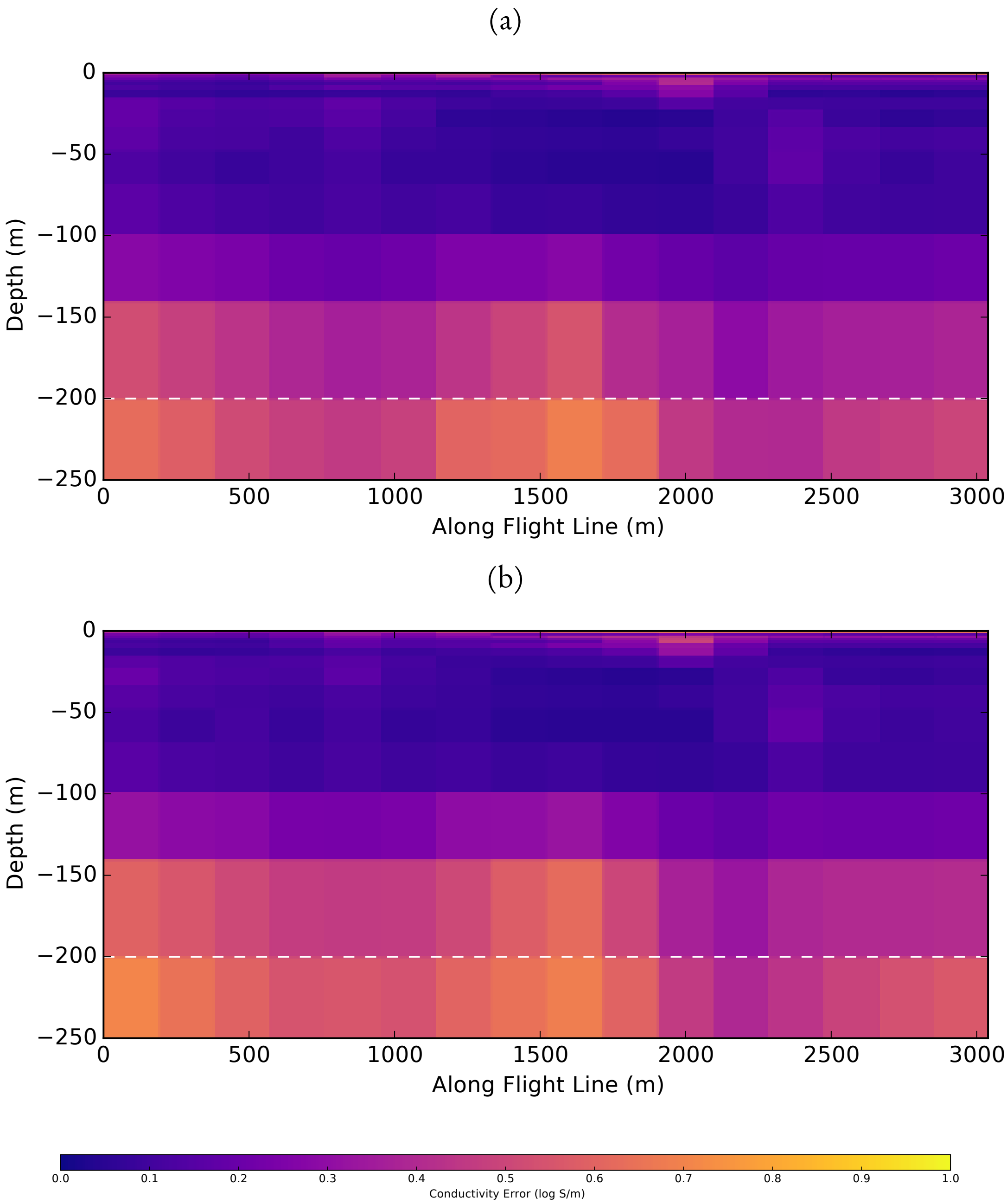


Figure 4.7: In this figure simple estimates of uncertainty, in (a) is half the magnitude of the 95% highest posterior density width in each grid cell. In (b) is twice the standard deviation at each point in the ensemble which is an approximation of the credible interval that assumes the posterior at each point is Gaussian. From these two estimates of uncertainty, features become increasingly uncertain with depth.

logarithmically between a temperature of 1 and 10. The inversion was performed incrementally in lots of 200,000 iterations until the chains appeared converged. After convergence, the posterior ensemble was collected by running for a further 1 million iterations. The inversion process took approximately 27 hours (using 512 cores on a cluster with Intel Xeon CPUs).

Figure 4.8 shows the convergence of the negative log likelihood in (a), the hierarchical error scaling parameter in (b), and the number of wavelet coefficients in (c). These plots show well converged chains both sampling around the same region of the posterior. The hierarchical data noise scaling parameter has converged to a scaling of approximately 0.6 suggesting a relative error closer to 2.2% is perhaps more appropriate for this data. The modal number of coefficients required to represent the solution is approximately 120 from a uniform prior of between 1 and 1000 coefficients (of a possible 1024 for this 32 sounding by 32 layer system). While not the intended purpose of trans-dimensional sampling, this can also be seen as reducing computational effort as instead of sampling 1024 individual pixels, the problem is reduced to sample over a far smaller number of coefficients. More importantly, it has been observed elsewhere that trans-dimensional sampling results in a better constrained inversion with better estimates of uncertainties than fully sampling the entire domain [Dettmer et al., 2016].

In Figure 4.8(a) the negative log likelihood has converged to below the χ^2 limit. There can be many reasons why this can occur but the most likely is that there is correlation in the noise in the observations that is currently not taken into account by the empirical noise model. An approach to verify and/or remedy this is to compute the covariance of the residuals obtained from the inversion as in Dettmer et al. [2016] and re-invert the data using this empirically derived covariance matrix in (4.2). Understanding the covariant attributes of the noise inherent in this system would be a potential area of future research.

In Figure 4.9(a) is shown the median of the ensemble as a representative estimate of the likely distribution of conductivity that generated the measured observations. The median compares well with previously obtained results using the a conventional laterally

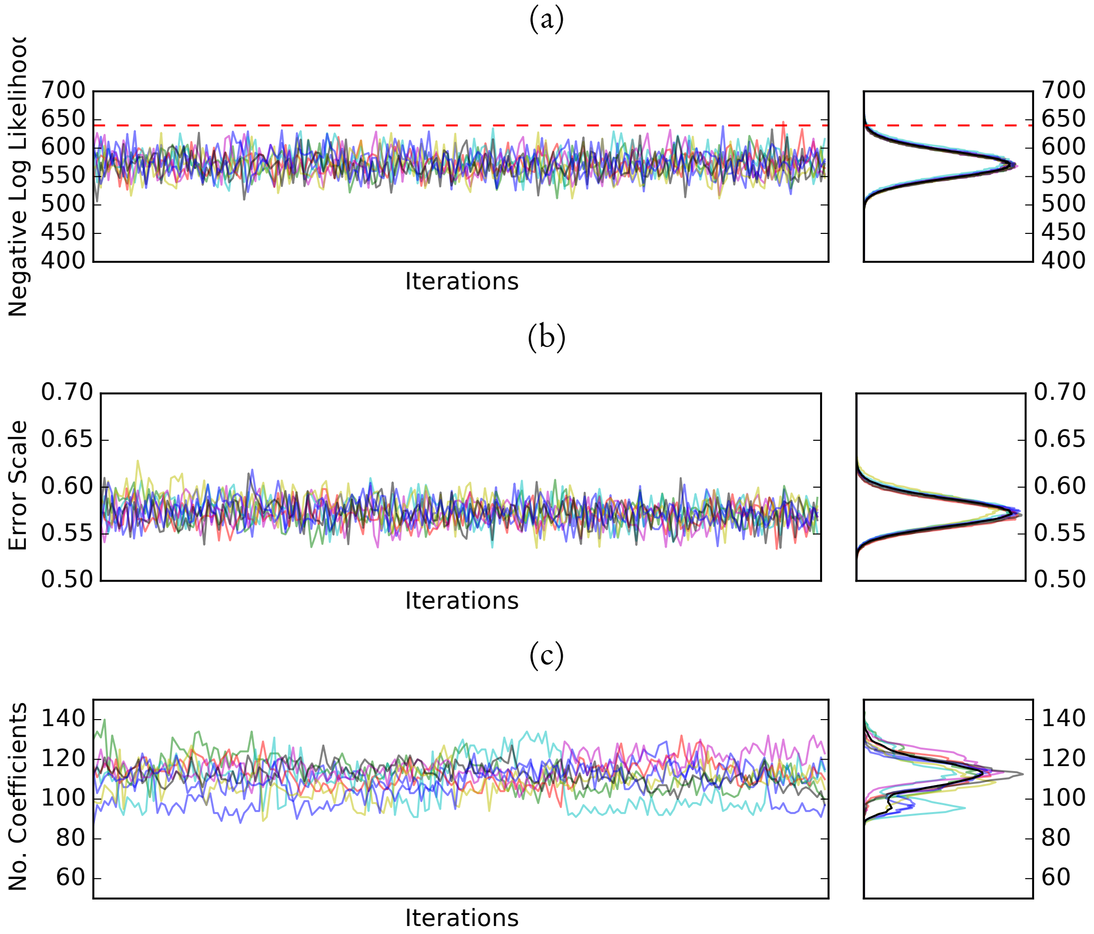


Figure 4.8: The qualitative indications of convergence of the simulation result for the section of data from the BHMAR project. In (a) are individual chains negative log likelihood as a function of iteration as different colours and to the right, histograms of the likelihood with a combined histogram shown in black. Similarly in (b) is the history and posterior histograms of the individual hierarchical error scaling parameter, and in (c), the number of wavelet coefficients required by the inversion.

constrained deterministic least squares approach with damping and smoothing regularisation shown in Figure 4.9(b) [Lawrie, 2016]. The broad features of the inversions are comparable, however there are two striking differences. The trans-dimensional approach appears to have recovered a deep high conductivity region beginning at approximately 75m deep at the left of the region, and a low conductivity anomaly at approximately 40m deep to the right of the region.

In Figure 4.10 the comparison is shown again of the median of the ensemble to the earlier result, zoomed in on the top 50m to examine detail near the surface. Here the trans-dimensional approach achieves similar structure but with greater detail and higher magnitude conductivity anomalies than the traditional least squares approach with damping and smoothing regularisation. The overall amplitude of variations is higher in the trans-dimensional solution suggesting that the least squares solution may have been over damped or smoothed. Additionally, with the trans-dimensional tree approach, the trade off between noise level and complexity is explored through the use of the hierarchical noise scaling parameter.

4.3.4 Ensemble appraisal

A useful feature of a trans-dimensional ensemble based approach is that it is a method that can robustly estimate uncertainties in highly non-linear inverse problems. This uncertainty may become critical in decision making processes relying on AEM studies, for example, determining optimum location of sites for hydro-geological investigations.

Common estimates of uncertainty, such as computing the Hessian in a least squares optimisation approach, give linearised estimates which assume a Gaussian distribution for the posterior. This can lead to poor estimates of uncertainty, both under and over estimates. In a Bayesian sampling approach, point wise standard deviations and HPD intervals can be computed which give similar single value representations of model errors. Once again though, these assume a Gaussian distributed posterior and a single modality posterior respectively, and similarly result in poor estimates of uncertainties

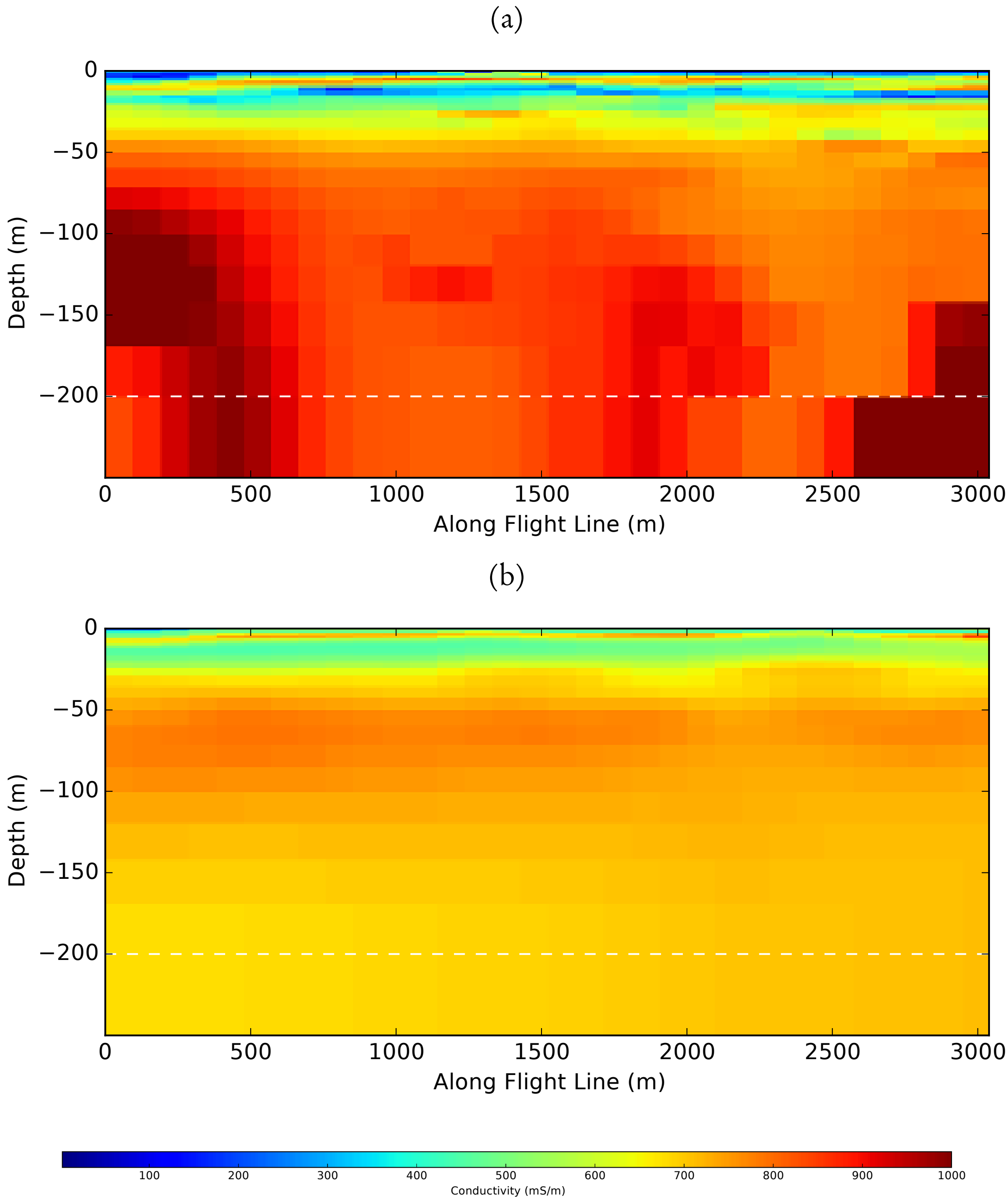


Figure 4.9: The median of the ensemble from the inversion of data from the Broken Hill Managed Aquifer Recharge (BHMAR) project using the Trans-dimensional tree approach is shown in (a). For comparison, the inversion result obtained from traditional least squares inversion with damping and smoothing regularisation is shown for the same area in (b).

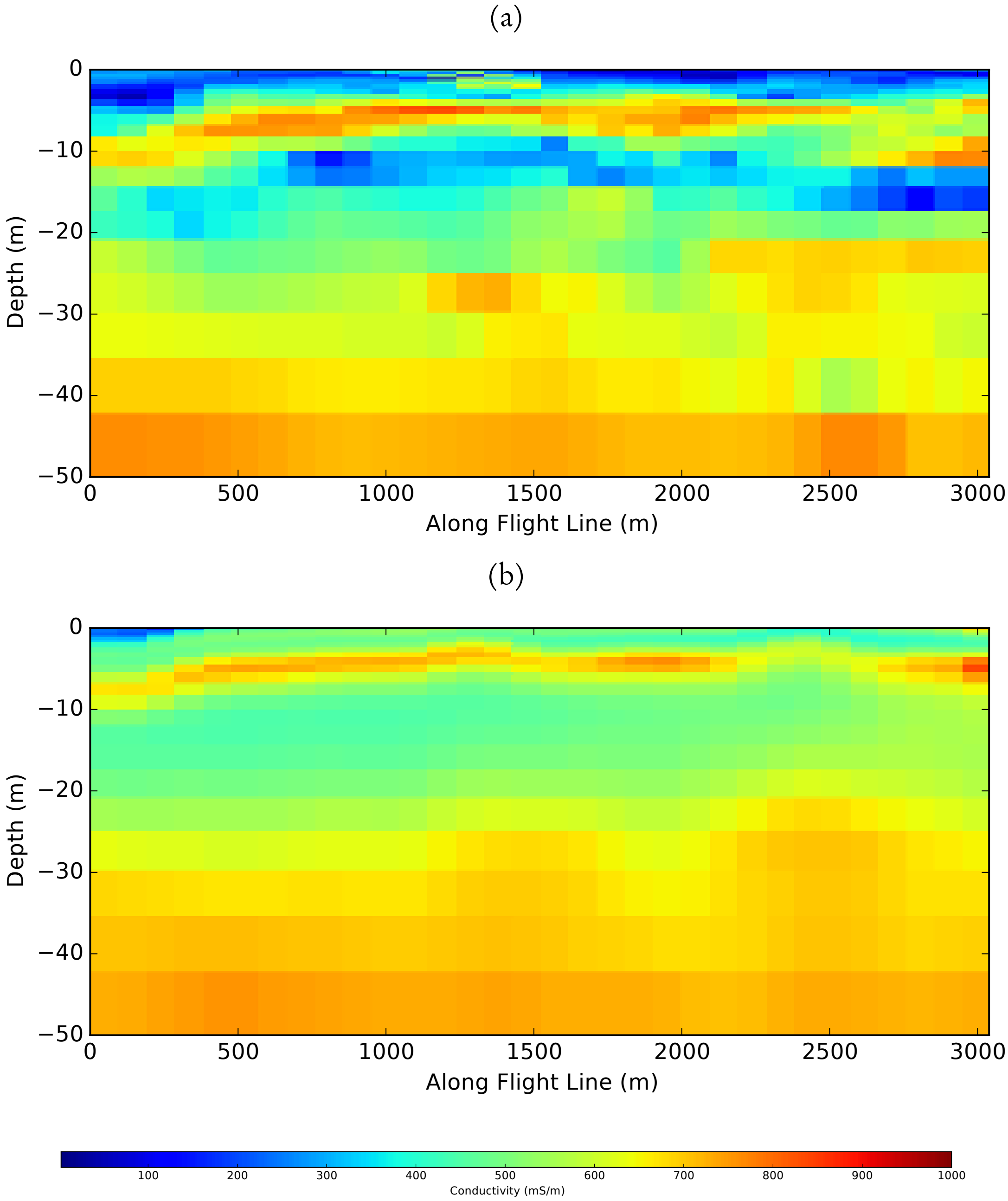


Figure 4.10: As in Figure 4.9 however here only the top 50m is shown of the median of the ensemble in (a) to better examine the detail near the surface. (b) shows the least squares result for the same region.

if these assumptions are violated.

A more effective method of examining the potential uncertainties is to look at 1D marginal posterior probabilities. To give an example, 1D marginal posterior densities along depth profiles at points of interest are shown in Figure 4.11 where (d) shows the model domain overview and location of transects along depth, and individual marginal probabilities for the respective profiles are shown in (a), (b) and (c). In these three plots, relative probability of the conductivity is shown in shaded colour (darker meaning higher probability). Also shown is the median with a red dotted line, and the highest posterior density (HPD) minimum and maximum (assuming a unimodal distribution) with green dashed lines.

A common feature of these marginal plots is their high degree of multi-modality. Even though the likelihoods, hierarchical error scaling and number of model parameters have converged, the posterior shows strong variations indicative of multiple conductivity profiles supported by the observations. This is characteristic of a forward model that is diffusive in nature and can not be quantified in optimisation approaches with simplistic linearised estimates of uncertainty.

Transect A was chosen to investigate the high conductivity anomaly that was recovered by the trans-dimensional approach but not seen in the traditional least squares solution. One can clearly see in Figure4.11(a) that, though the spread of possible conductivities is large from this depth, the support for conductivity around 1 S/m or greater from approximately 75m deep is reasonably strong down to 140m, but deeper than this inference becomes questionable. In contrast, transects B and C show strongest support for conductivities less than 1 S/m in the deeper parts of the model. This strongly suggests that the high conductivity anomaly is not an artefact and in fact constrained by the observations, albeit weakly given the spread of the values. The least squares approach fails to see this anomaly possibly due to over damping to a reference model.

In Figure 4.12, the marginal probabilities are shown again for the top 50m of the domain in order to highlight the near surface structure. In transect B, in the top 5m there is a feature that shows strong layering in the median model. In examining the

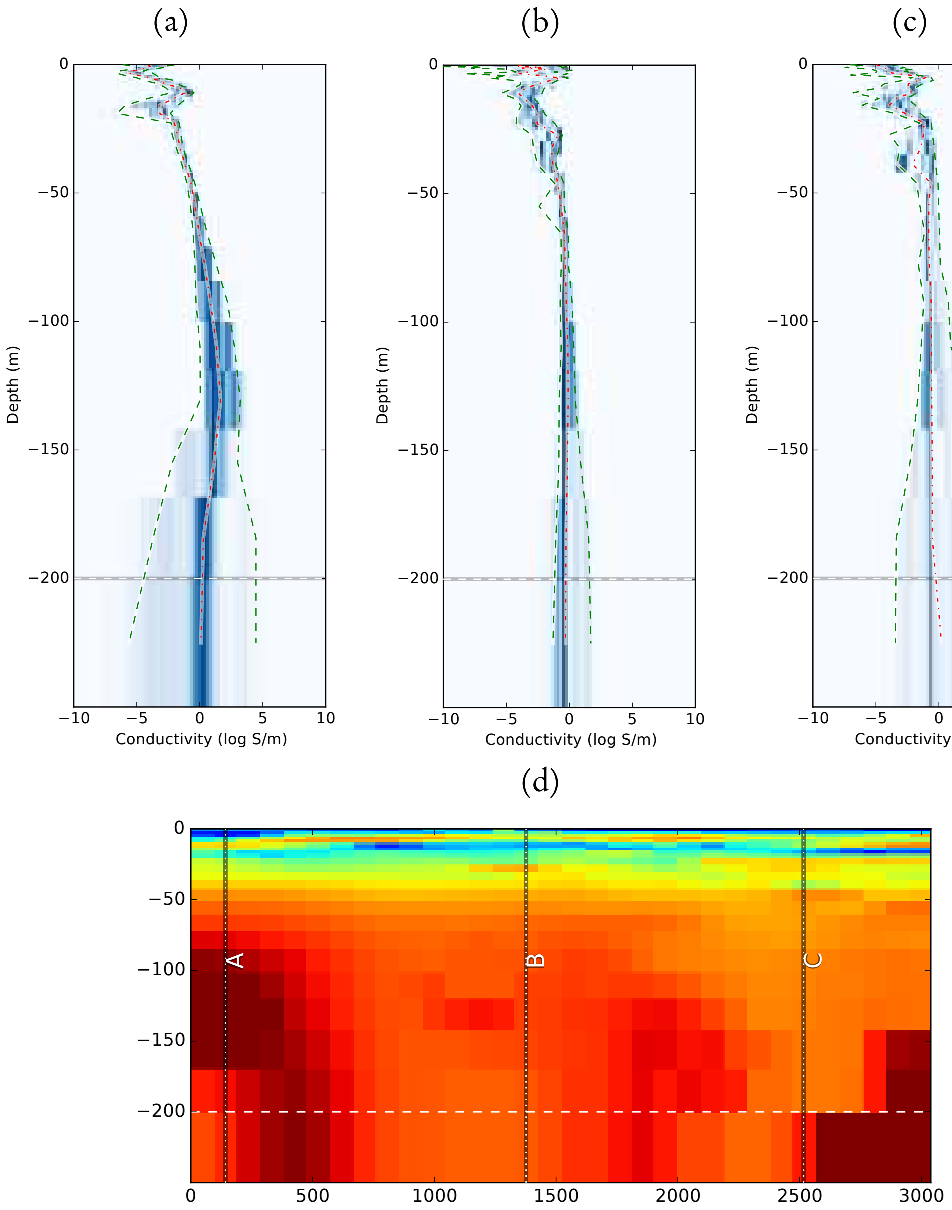


Figure 4.11: The marginal probabilities along vertical profiles with depth at three locations in the modelled domain. The location of the profiles is shown in the median model image in (d). The marginal probability densities for profiles A, B, C are shown in (a), (b), (c) respectively. The blue shading represents the relative probability of the conductivity at a given depth (darker equates to more likely). The median is shown over plotted with a red dotted line, and green dashed lines are used to show the width of the highest posterior density (HPD) region.

marginals for this transect, in comparison to the other transects, there is a larger uncertainty and multi-modality along this transect in the top 5m. This may be caused by a conductive surface feature causing interference of the observations, or another source of observational noise.

Finally, referring to Figure 4.12, in transect C there is a low conductivity anomaly at approximately 40m. In examining the 1D marginal in (c), at this point there is a strong bi-modality between two low conductivity values and the median has been perturbed toward the lowest of these. The modal value at this depth is the higher of the two bi-modal peaks suggesting that this low conductivity anomaly is less likely.

It is clear from the marginal profiles shown at three locations, and in particular this last marginal profile, that presenting the results of a conductivity subsurface in terms of a single model estimate, either from an optimised solution or the mean or median of an ensemble solution, lacks a great deal of the information on the non-uniqueness of the solution. It is only through careful and rigorous interrogation of the marginal distribution of conductivities that constraint of the inversion at all points of the model domain can be evaluated. Ensemble based approaches require considerably more attention by the practitioner, but likewise provide a greater degree of richness in interpretation of the model constraint.

4.4 Hierarchical Laplacian Prior

One of the draw backs of the trans-dimensional tree with a wavelet parameterisation is the difficulty in the setting of the prior. Using numerical studies such as those presented in Figure 4.4 are useful to gauge the range of potential variation in the model, but do little to examine the sensitivity of inversion to the prior.

An approach to remedying this is to use a hierarchical prior, sometimes called a hyper prior [Malinverno and Briggs, 2004]. Much like the hierarchical error estimate where a scaling of the noise level is inverted for as part of the inversion process, in a hierarchical prior the scaling “width” of the prior can be inverted for during the inversions. A

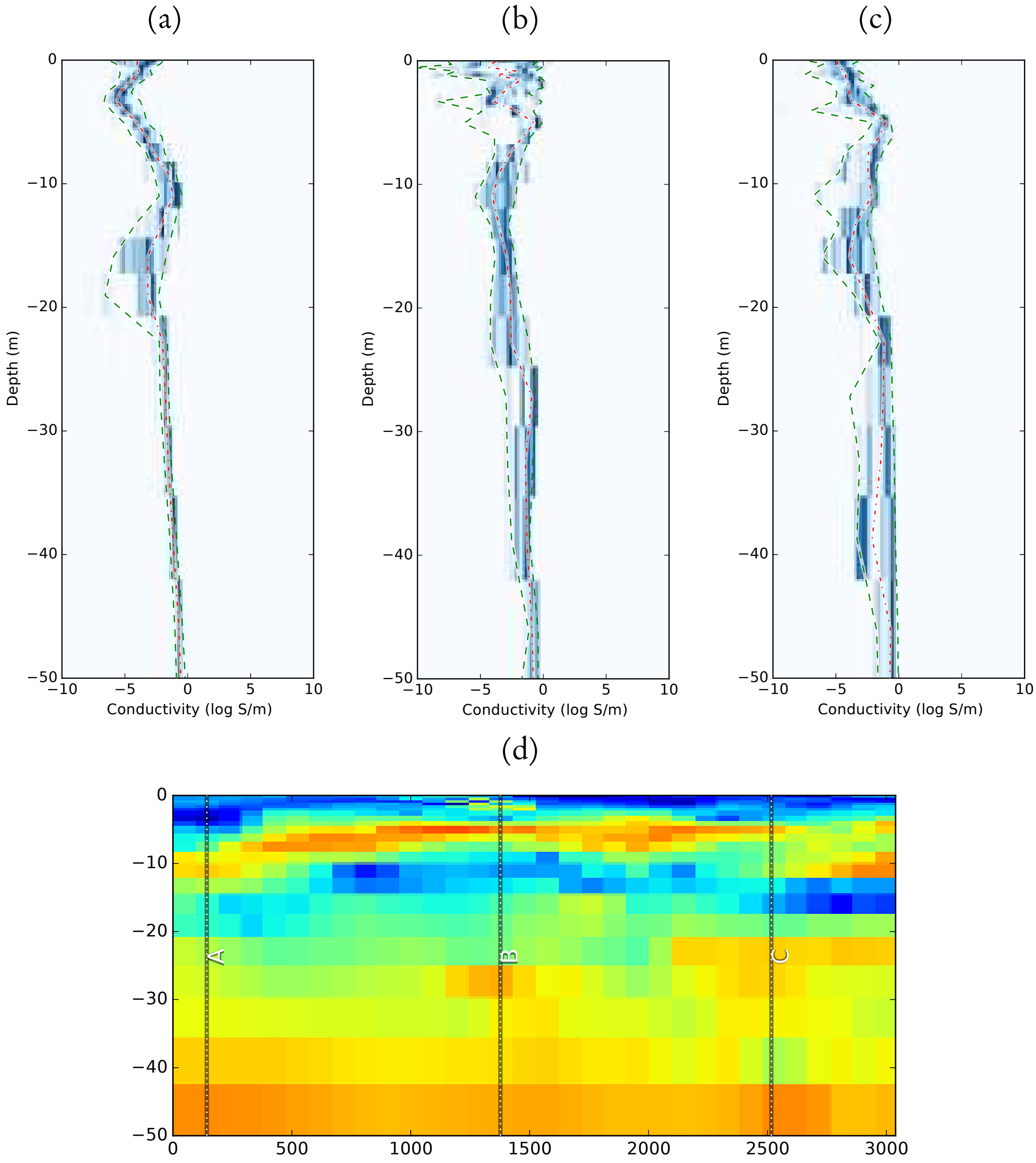


Figure 4.12: As in Figure 4.11, the marginal posterior probabilities for the three indicated transects, this time zoomed in on the top 50m of the domain.

hierarchical prior proposal will randomly perturb this width scaling parameter and while the likelihood will not change, the prior ratio will and this proposal can be accepted or rejected with the standard Metropolis-Hastings acceptance criteria. For uniform priors, this can be problematic as they do not have infinite support and as such a proposal that narrows the width of the uniform prior may cause one or more model parameters to be out of the uniform prior range and therefore invalid.

A way to overcome this is to use a wavelet coefficient prior with infinite support. In early work on image compression using wavelet transforms [Antonini et al., 1990, 1992], numerical studies of the wavelet coefficients of large numbers of images showed that the distribution of the coefficients was best fit by a generalised Gaussian distribution with a shape parameter of approximately 1.7. For reference a Laplace distribution is a generalised Gaussian with a shape parameter of 1, a Gaussian distribution has a shape parameter of 2 and a uniform distribution has a shape parameter of ∞.

In seismic tomography, in recent years many studies have used a wavelet based parameterisation with a l_1 norm regularisation [Chiao and Kuo, 2001, Simons et al., 2011b, Chevrot et al., 2012, Charlety et al., 2013, Fang and Zhang, 2014]. The l_1 norm is in a Bayesian sense analogous to a Laplacian prior. A Laplacian prior also more closely resembles approaches using “Spike and Slab” priors for Bayesian compressive sensing approaches [Ishwaran and Rao, 2005, He and Carin, 2009, Vera et al., 2009, He et al., 2010]. For these reasons, a Laplacian prior for infinite support seems to be an appropriate choice.

Rather than setting a fixed width for the Laplacian prior for the inversion, this width is inverted for as part of the inversion, in much the same way as the hierarchical error scaling term. To achieve this, a prior needs to be set on the width of the Laplacian prior. The common approach to this is to use the conjugate distribution to that of the Laplacian, the inverse Gamma distribution [Park and Casella, 2008]. Assigning the inverse Gamma distribution as the prior on the Laplacian width parameter removes the need to configure a fixed width parameter, however this difficulty has been passed up the chain to the parameters of the inverse Gamma distribution. The probability density function of the inverse Gamma distribution is

$$p(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left\{-\frac{\beta}{x}\right\}, \tag{4.7}$$

where α is a shape parameter and β is the rate parameter. The choice used in this study is to set α to one. This results in a relatively flat prior on the width of the Laplacian distribution. For the rate parameter, this is configured by setting the desired mode of the inverse Gamma distribution using the relationship

$$Mo = \frac{\beta}{\alpha + 1}, \tag{4.8}$$

with Mo the mode of the distribution. The reasoning behind this is that there is some guidance as to a good starting point for the width of the Laplacian distribution from the ensemble of plausible models from the previous inversion. In Figure 4.13, the distribution of wavelet coefficients from the ensemble of models is shown and the Laplacian distribution with a width parameter of 0.5 is plotted with a red dashed line. The two distributions are quite close, further justifying the choice of a Laplacian prior.

To formulate a proposal to perturb the prior width, recall that the prior on the model is constructed as

$$p(\mathbf{m}) = p(\mathbf{c}|\mathscr{T}, k) p(\mathscr{T}|k) p(k), \tag{4.9}$$

where k is the number of wavelet coefficients, $\mathscr{T}$ is the tree structure prior, and $\mathbf{c}$ is the vector of wavelet coefficients. It is this last probability that is affected by a hierarchical prior perturbation, whilst the rest is unchanged. The hierarchical prior can be incorporated by adding dependencies writing

$$p(\mathbf{c}|\mathscr{T}, k) = p(\mathbf{c}|\mathscr{T}, k, b) p(b|\alpha, \beta), \tag{4.10}$$

where b is the width of the Laplacian distribution, and α and β are the parameters for

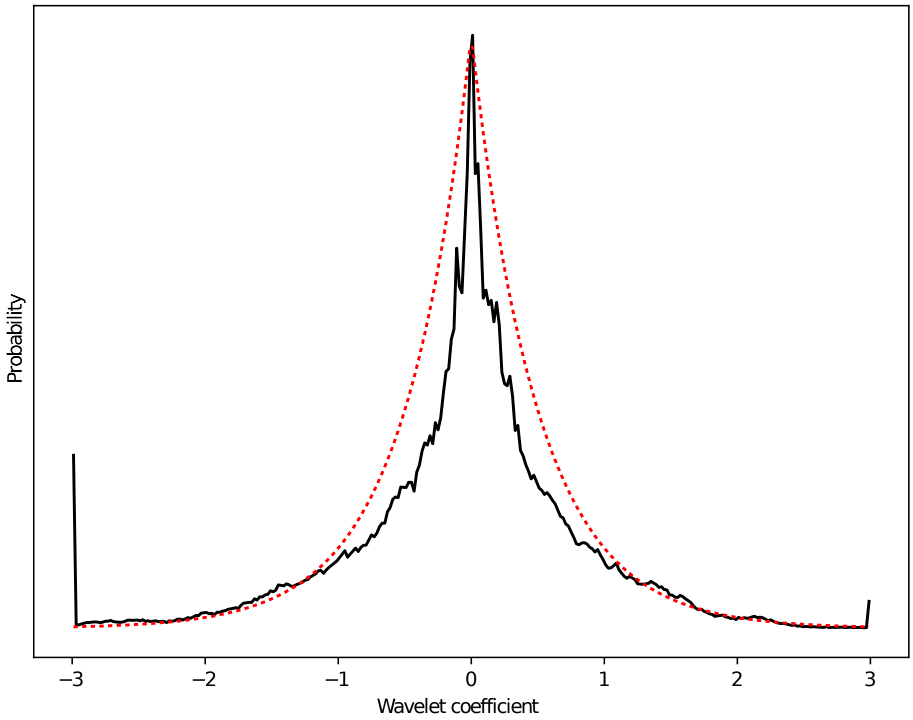


Figure 4.13: From the ensemble of wavelet coefficients obtained from the inversion, a probability distribution can be constructed, shown with the black curve. Over plotted with a red dotted line is a Laplacian distribution with the width parameter of 0.5.

the inverse Gamma distribution. A single prior is used for all wavelet coefficients and so the first part of the expression above is

$$p(\mathbf{c}|\mathscr{T},k,b) = \prod_{i=1}^{k} \frac{1}{2b} \exp\left\{ \frac{|c_i|}{b} \right\}, \tag{4.11}$$

For a proposal to change the hierarchical prior that uses a symmetrical proposal distribution, the likelihood and proposal ratios cancel, reducing the acceptance criteria to the component of the prior ratio dependent on b in (4.10) or

$$\alpha = \min\left\{ 1, \frac{p(\mathbf{c}|\mathscr{T},k,b')p(b'|\alpha,\beta)}{p(\mathbf{c}|\mathscr{T},k,b)p(b|\alpha,\beta)} \right\}, \tag{4.12}$$

where b' is the proposed Laplacian distribution width. This additional proposal was incorporated into the inversion method of the previous section and re-run. The qualitative criteria used to judge convergence are once again shown in Figure 4.14, this time with the addition of the hierarchical prior convergence. As can be seen in the figure, the convergence appears similar to the previous inversion in Figure 4.8.

The median of the ensemble is shown in Figure 4.15 and Figure 4.16 for the near surface features (top 50m). From both figures, it can be concluded that the resulting structure is very similar suggesting that the initial choice of a uniform prior was reasonable and the addition of the hierarchical prior is unnecessary. The issue that the hierarchical prior resolves however is the difficulty of choosing the prior width of a uniform prior. If a uniform prior is configured too narrow, wavelet coefficients can be truncated resulting in artefacts in the inversion. Similarly, if the uniform prior is too wide, acceptance rates in the birth/death proposals can be adversely affected resulting in poor mixing. In contrast, with the Laplacian prior there is infinite support and the width of the hierarchical Laplacian prior adapts to the observations, increasing the birth/death acceptance rates resulting in better mixing.

These results demonstrate that a hierarchical prior can be used to invert for an unknown prior width for wavelet coefficients. Statistics gained from a first inversion

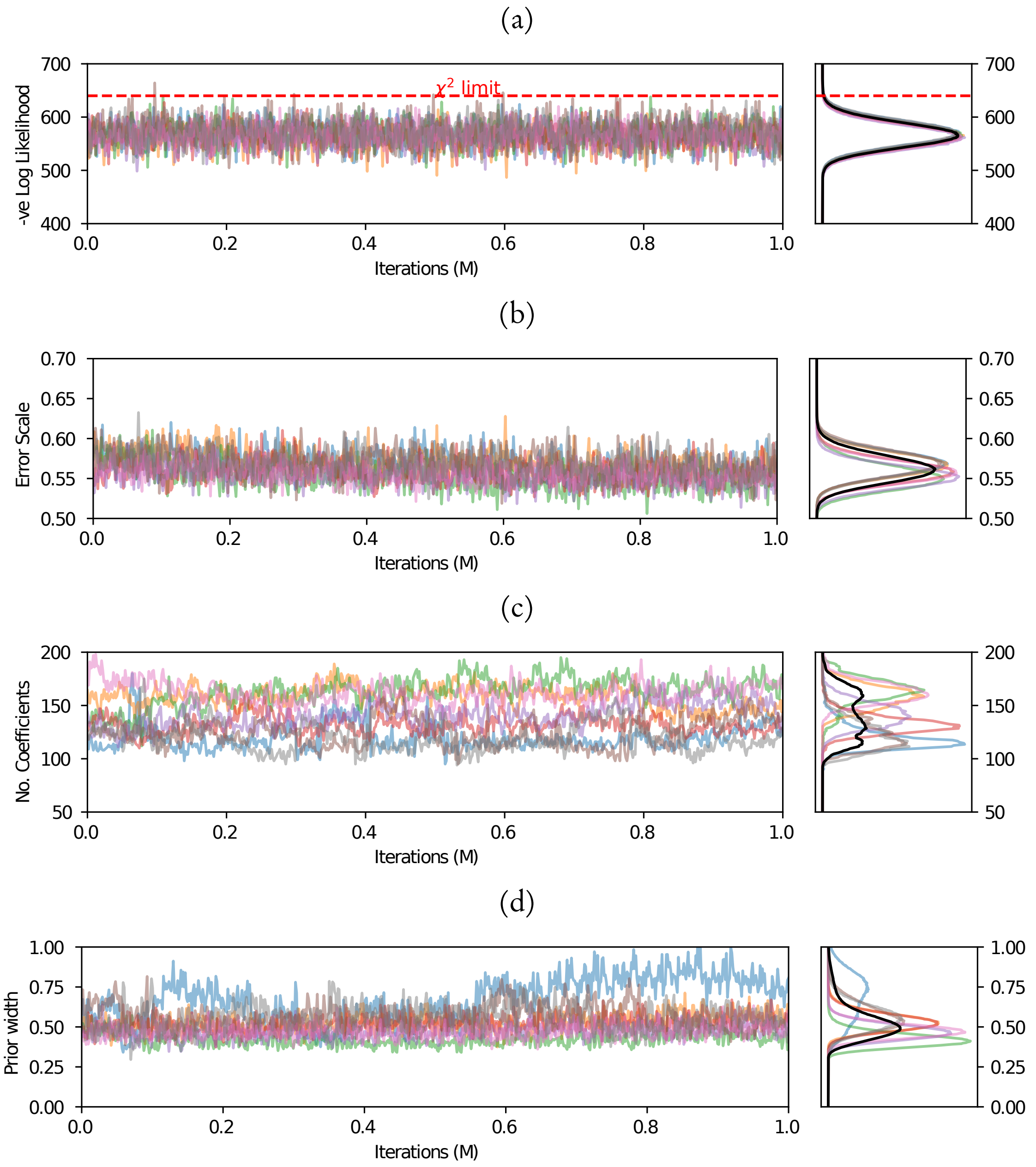


Figure 4.14: This figure gives qualitative indications of convergence of the simulation result for the section of data from the BHMAR project. In (a) are the individual chains negative log likelihood as a function of iteration as different colours and to the right, histograms of the likelihood with a combined histogram shown in black. Similarly in (b) is the history and posterior histograms of the individual hierarchical error scaling parameter, and in (c), the number of wavelet coefficients required by the inversion.

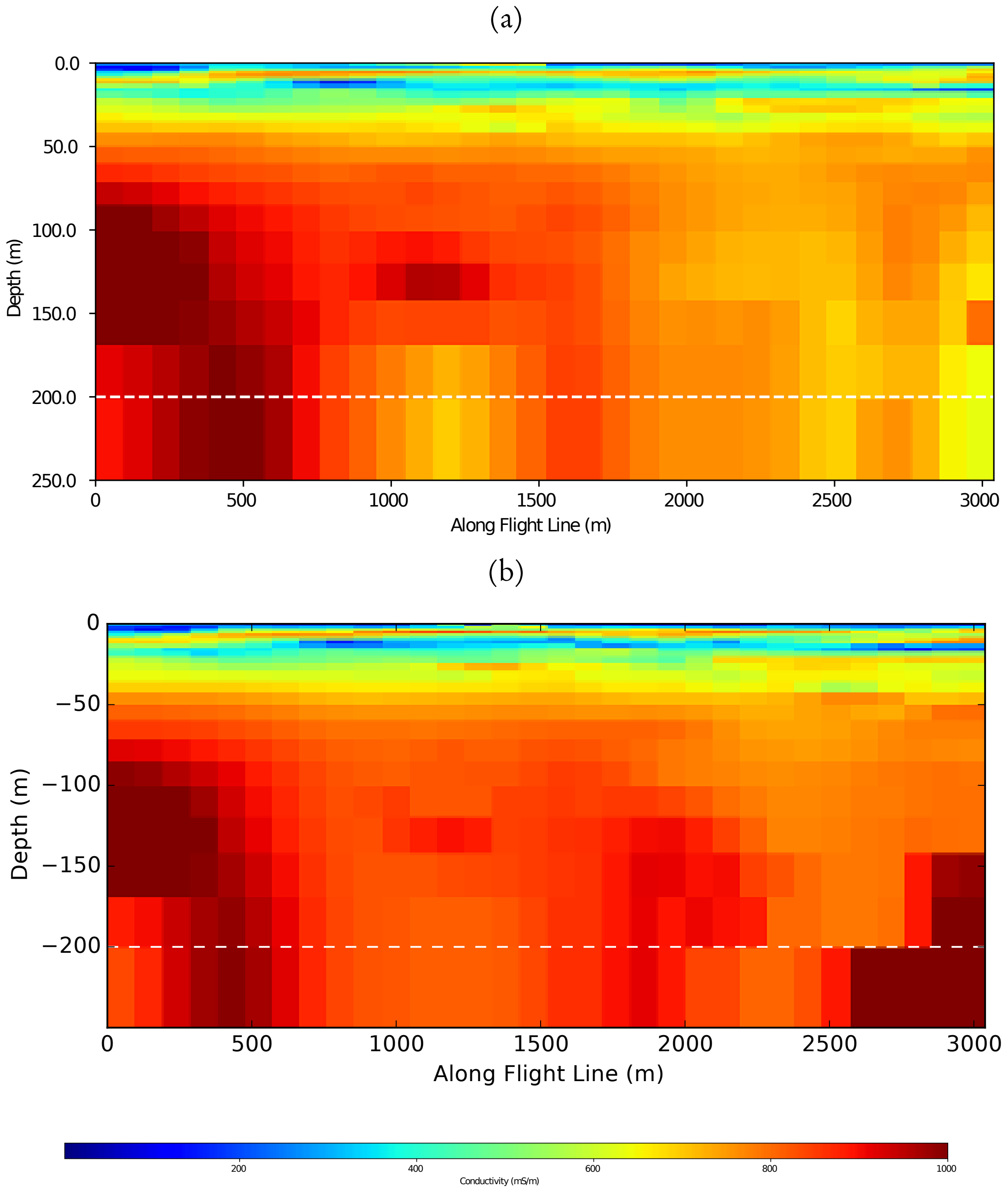


Figure 4.15: The median of the ensemble from the inversion of data from the Broken Hill Managed Aquifer Recharge (BHMAR) project using the Trans-dimensional tree approach with a hierarchical Laplacian prior is shown in (a). For comparison, the inversion result obtained in the previous section is shown for the same area in (b).

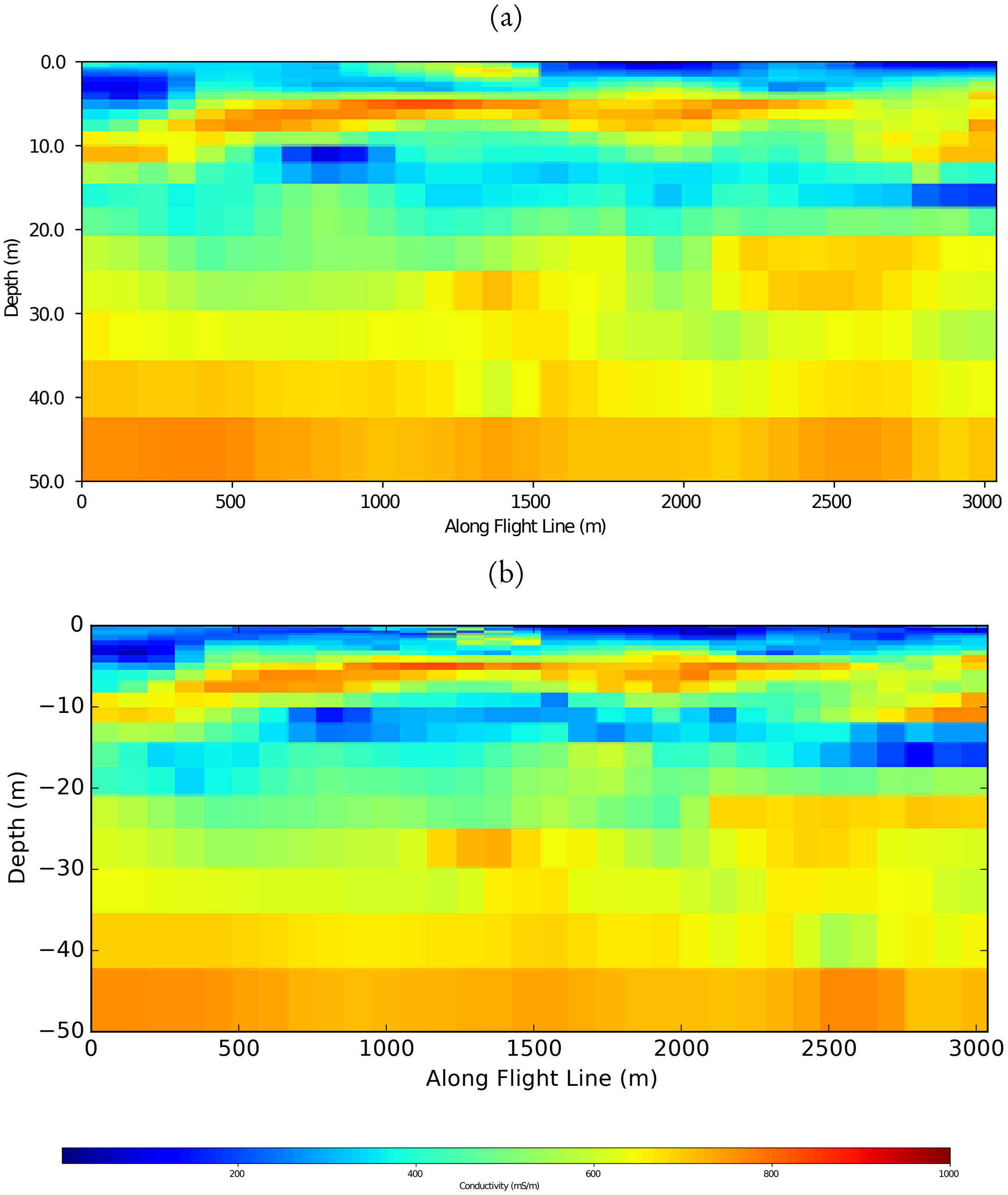


Figure 4.16: As in Figure 4.15 however here only the top 50m of the median of the ensemble is shown.

were useful to guide the setting the parameters of the hierarchical prior, that is, the parameters of the inverse Gamma distribution. From the posterior distribution of the Laplacian distribution width, this proved to be a good estimate. Through the addition of the hierarchical prior, the sensitivity of the final result specification of the wavelet coefficient prior has effectively been removed to give a more thorough investigation of the plausible model/prior space.

The benefit of the hierarchical Laplacian prior is the decreased sensitivity to poor specification of the prior in that a likely value for the modal Laplacian width parameter is set, but the inversion adapts this width to suit the data. A uniform prior on the wavelet coefficients, if poorly specified, can result in truncation of wavelet coefficients or poor birth and death acceptance rates which hinder convergence. For trans-dimensional tree inversions with wavelet parameterisations, the hierarchical Laplacian prior is preferable to fixed uniform coefficient priors.

4.5 Covariant Noise Model

In the previous section, much like the first inversion, the likelihood converges below the theoretical χ^2 limit. A potential cause of this is dependence or correlation between observations. The basic underlying physics of this problem, that is, different frequencies having different sensitivities to depth, suggest that this would be the case, yet in the previous two inversion an empirically derived noise model has been used, but with independent errors.

In previous studies where multiple passes over a single flight line have been performed, correlated errors can be estimated [Green and Lane, 2003]. In the case here, and in most cases, there will be a single pass over each flight line.

A general benefit of a sampling approach is the posterior statistical inference that can be performed on the ensemble, both in terms of model parameters, but also for residual analysis. To examine the effect of correlated noise on the result, the covariance matrices for the residuals of both the high and low moment signals can be computed.

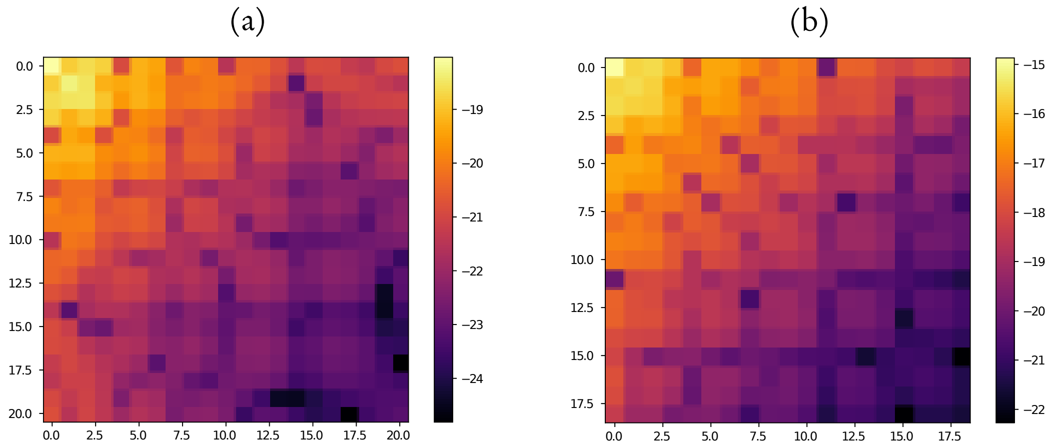


Figure 4.17: These figures shows the $\log_{10}$ *of the magnitude of the covariance for the high (a) and low (b) moment signals obtained from the previous run using the ensemble residuals.*

The covariance matrices are shown in Figure 4.17 and due to their spanning of several orders of magnitude and including negative values, the $\log_{10}$ of the magnitudes are plotted. Their properties along the diagonal show good agreement with the empirical noise model used previously but there is strong off diagonal structure.

A final inversion was run using the hierarchical Laplacian prior and the covariant noise model estimated from the previous run. A hierarchical scaling term was included as previously used, that is, the covariance matrix used in the likelihood function is

$$\mathbf{C}_e = \lambda \mathbf{C}'_e \tag{4.13}$$

where λ is the scaling term, $\mathbf{C}'_e$ is the covariance matrix obtained from the residuals and $\mathbf{C}_e$ is the covariance matrix used in the likelihood function. The convergence metrics used before are once again shown for this inversion that incorporates covariance in Figure 4.18.

As can be seen, the hierarchical scaling parameter has converged to a much lower value than in the two previous inversions, indicating that the estimated covariance is too large in magnitude. Now however, the likelihood has converged to the correct χ^2 limit indicating that the noise model in the inversion is closer to the truth. The process of estimating covariance from the residuals can be iterated to try and improve the noise

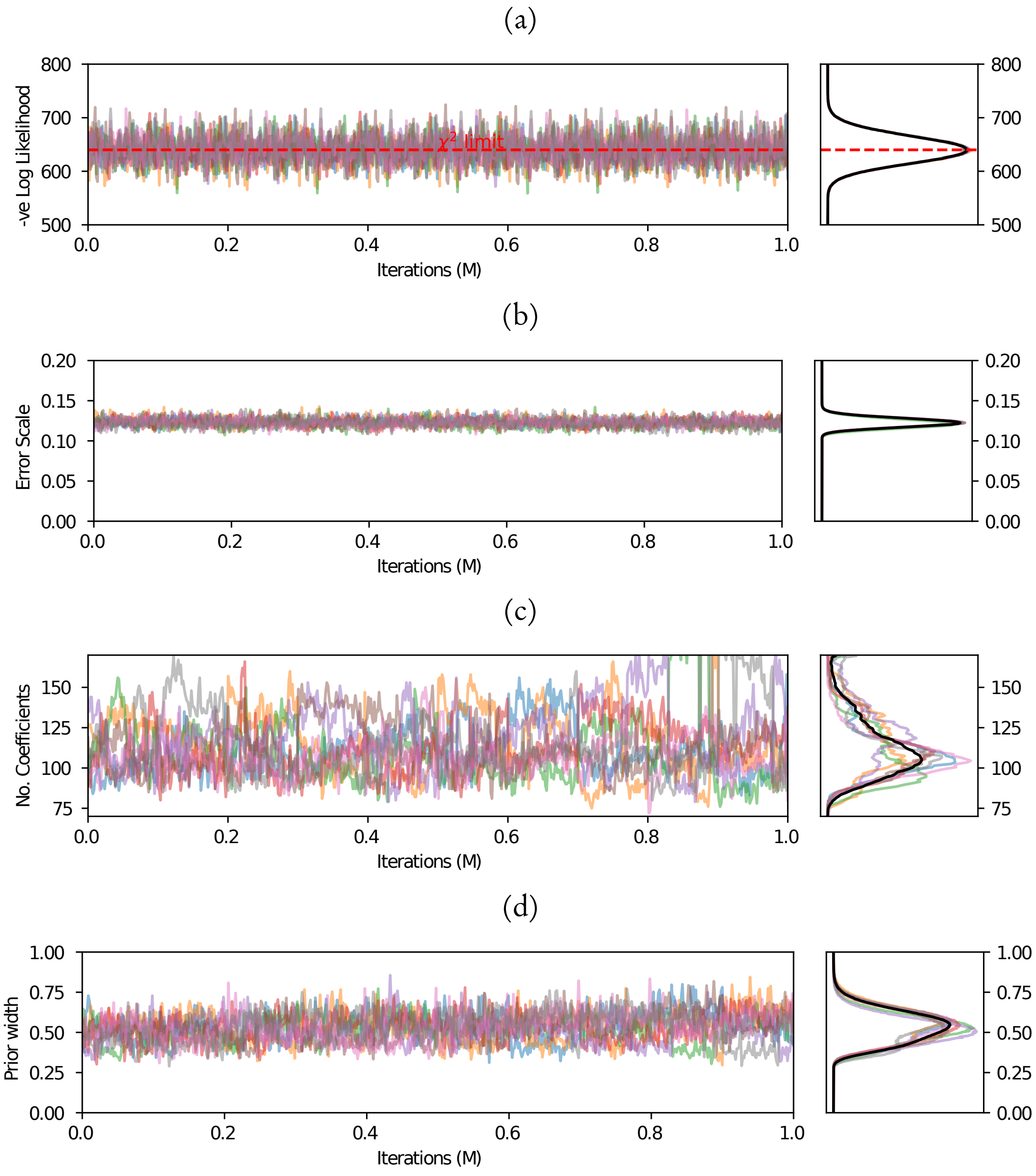


Figure 4.18: This figure gives qualitative indications of convergence of the simulation result for the section of data from the BHMAR project. In (a) are the individual chains negative log likelihood as a function of iteration as different colours and to the right, histograms of the likelihood with a combined histogram shown in black. Similarly in (b) is the history and posterior histograms of the individual hierarchical error scaling parameter, and in (c), the number of wavelet coefficients required by the inversion.

model further, however the robustness of the median of the ensemble compared to the previous inversion as shown in Figure 4.19 and Figure 4.20 suggest that there only small incremental improvements could be made.

As a final comparison, comparison transects of the three inversion at the three separate locations are shown in Figures 4.21, 4.22 and 4.23. The three inversions are in broad agreement with each other, but a noticeable difference, particularly in Figures 4.21 and 4.23 is that the inversion using covariance, plot (c) in the figures, has a narrower distribution of conductivities at the deeper depths. In plot 4.22 this result is not so clear. This suggests that the addition of a covariant noise model has improved the sensitivities to depth. Note however the high conductivity anomaly that was present in both inversions using the empirical noise model, Figure 4.21 (a) and (b), has disappeared with the inclusion of a covariant noise model.

4.6 Summary

The application of a novel trans-dimensional sampling technique using a wavelet parameterisation to the inversion of time domain AEM data has been demonstrated. The new approach, while more computationally costly, provides quantitative insights on the robustness of residual features in the model as well as insight into the potential non-uniqueness of the inversion and more rigorous estimates of uncertainties. It achieves this through full solution of the physical model at each step, albeit with a 1D layered Earth approximation. This means it is not necessary to linearise the problem for the estimation of the conductivity field and the uncertainties.

In the trans-dimensional approach, there is no need for damping and smooth regularisation required by traditional least squares optimisation approaches. The damping penalising norms constrain conductivities toward an *a priori* reference model and this can result in underestimates of conductivity magnitudes for weakly informative data and consequently result in failure to resolve features and underestimates of uncertainties.

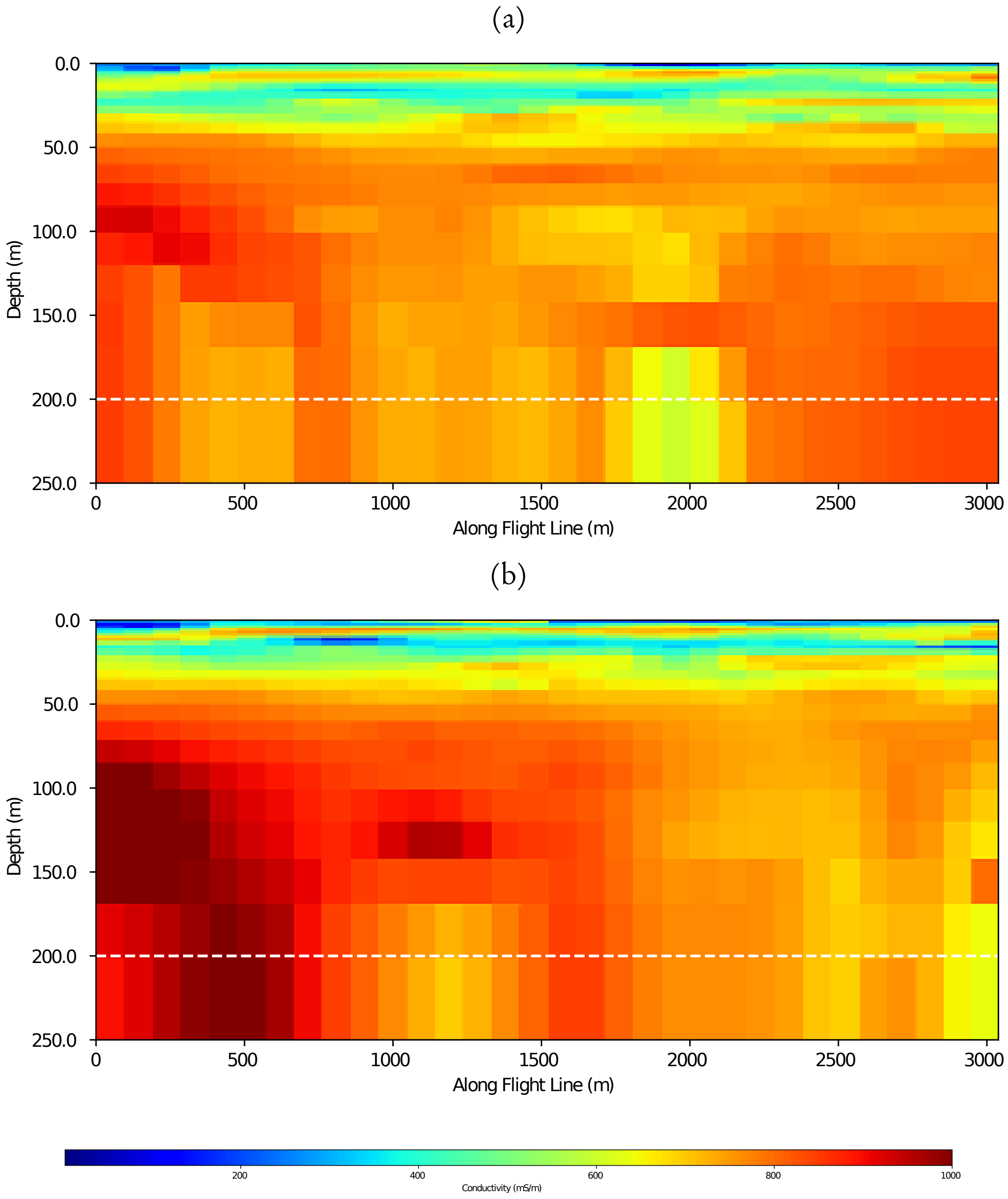


Figure 4.19: The median of the ensemble from the inversion of data from the Broken Hill Managed Aquifer Recharge (BHMAR) project using the Trans-dimensional tree approach with hierarchical prior and covariant noise model is shown in (a). For comparison, the inversion result obtained in the previous section is shown in in (b).

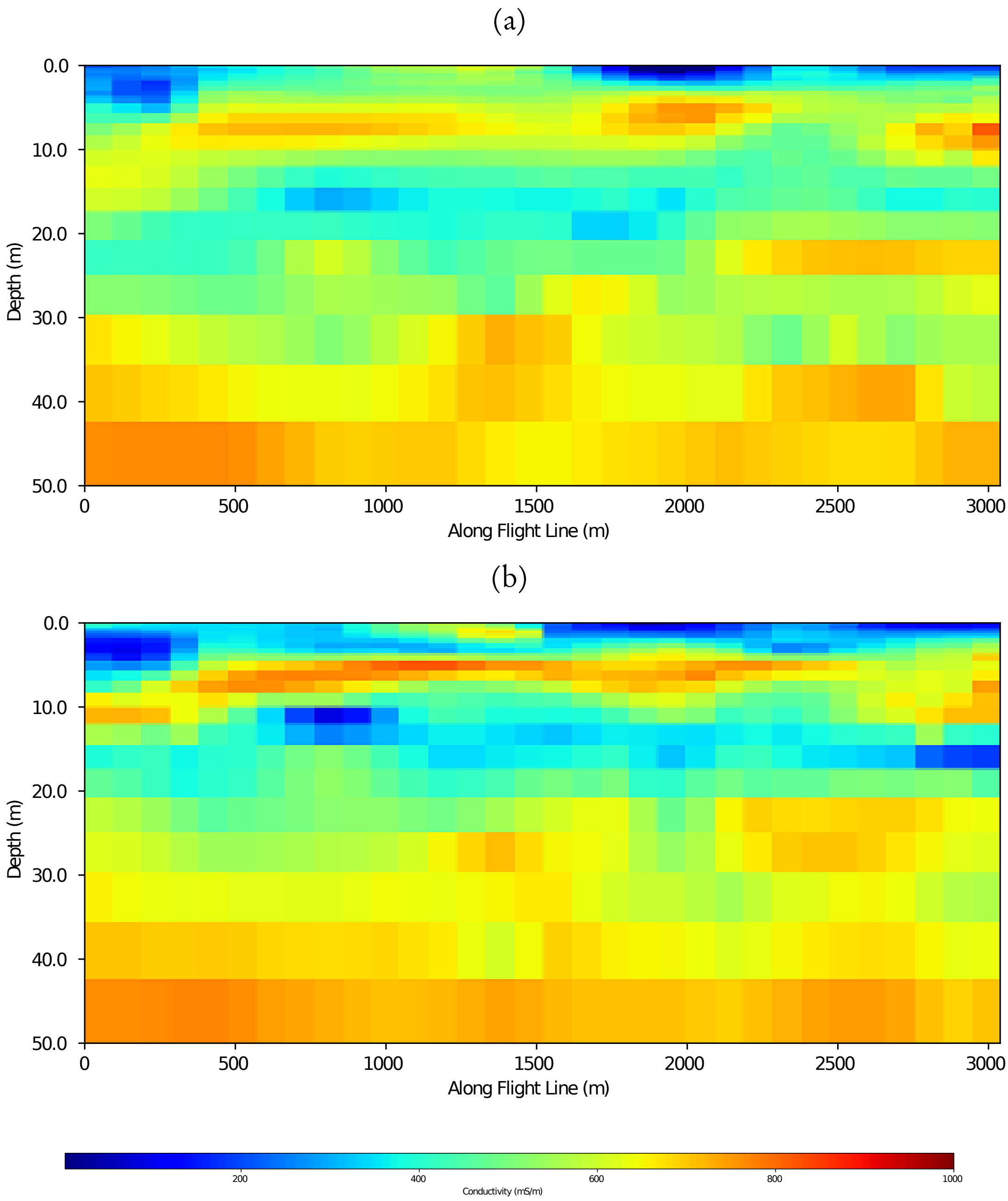


Figure 4.20: As in Figure 4.19 however this time only the top 50m of the median of the ensemble is shown.

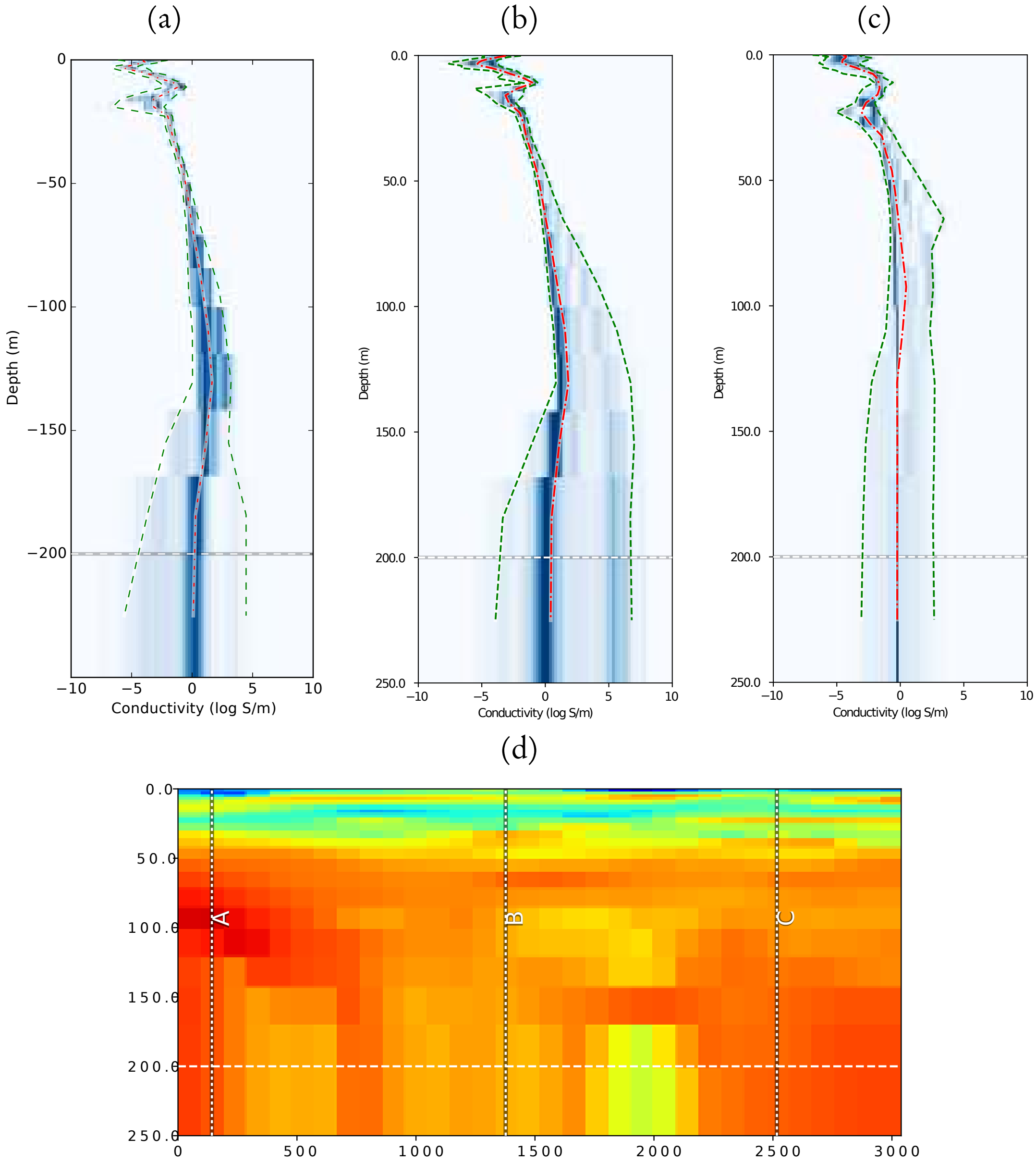


Figure 4.21: The marginal probabilities of three different inversion at transect A. (a) is the original inversion, (b) with hierarchical prior and (c) with hierarchical prior and covariant noise model. The blue shading represents the relative probability of the conductivity at a given depth (darker equates to more likely). The median is shown over plotted with a red dotted line, and green dashed lines are used to show the width of the highest posterior density (HPD) region.

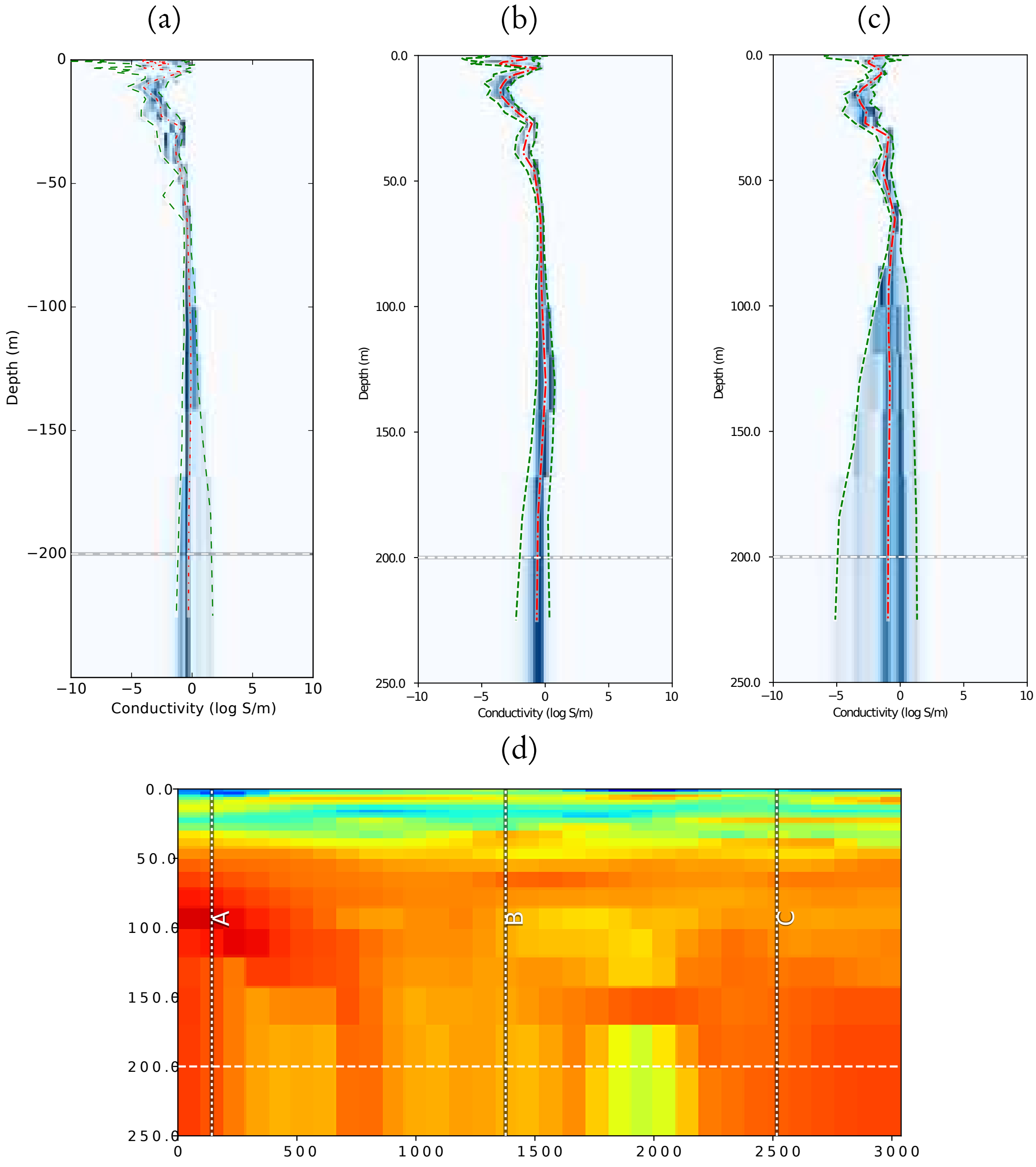


Figure 4.22: The marginal probabilities of three different inversion at transect B. (a) is the original inversion, (b) with hierarchical prior and (c) with hierarchical prior and covariant noise model. The blue shading represents the relative probability of the conductivity at a given depth (darker equates to more likely). The median is shown over plotted with a red dotted line, and green dashed lines are used to show the width of the highest posterior density (HPD) region.

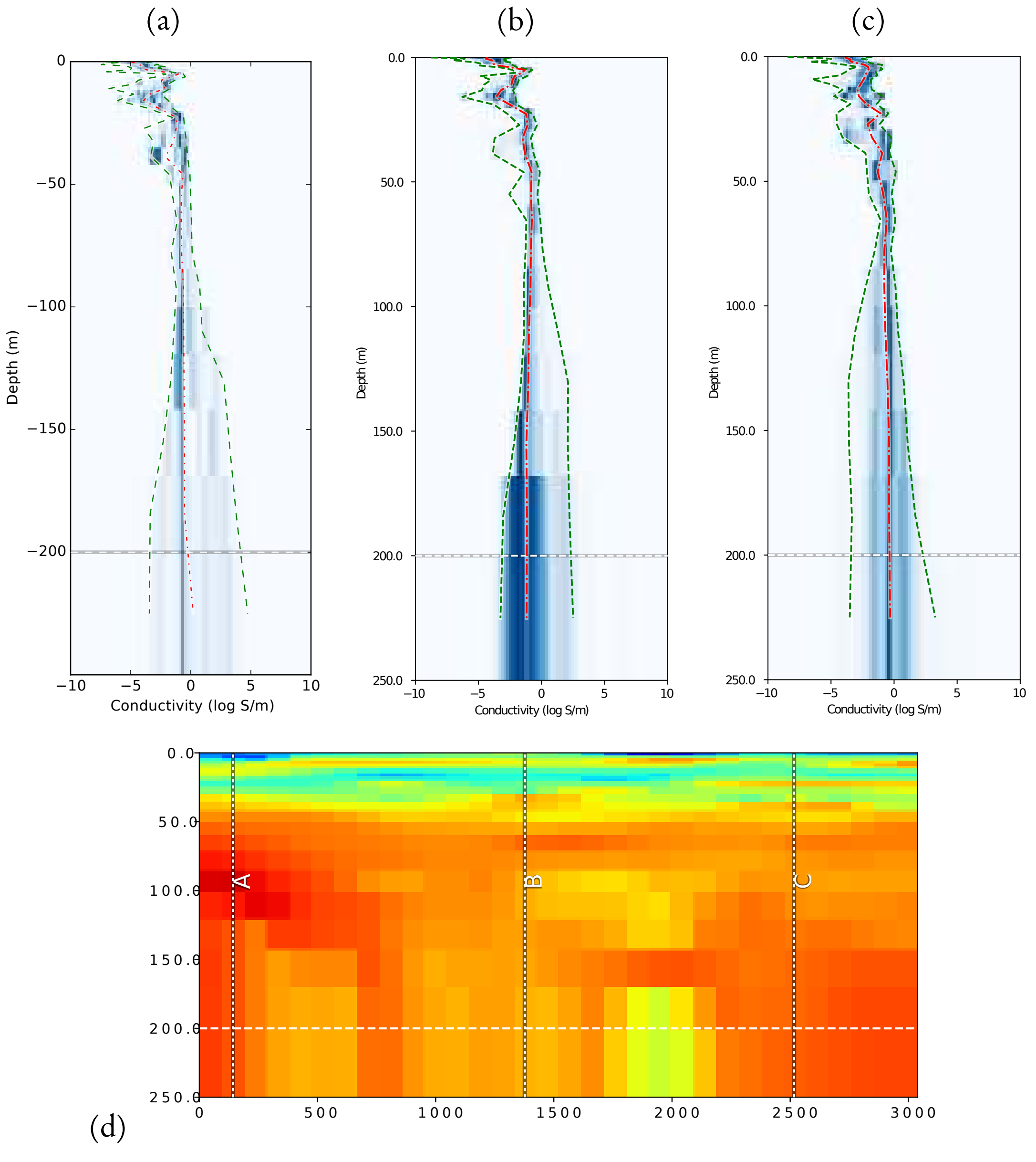


Figure 4.23: The marginal probabilities of three different inversion at transect C. (a) is the original inversion, (b) with hierarchical prior and (c) with hierarchical prior and covariant noise model. The blue shading represents the relative probability of the conductivity at a given depth (darker equates to more likely). The median is shown over plotted with a red dotted line, and green dashed lines are used to show the width of the highest posterior density (HPD) region.

The trans-dimensional approach is able to explore a larger range of conductivities and therefore show a larger range of conductivity profiles that plausibly fit the observations given the forward modelling and noise estimation. Similarly, smoothing penalising norms by their nature reduce resolvability of conductivity contrasts. In the trans-dimensional approach, the inversion adapts to the resolvable features required to support the observations rather than *a priori* imposing, via a smoothing weight, an upper limit on the resolution of the model.

Additionally, through hierarchical Bayesian sampling, the sensitivity of the inversion to a noise model can be explored through hierarchical scaling terms, and likewise, the prior through hierarchically sampling the prior width. This in a sense integrates out the sensitivity to these parameters in the inversion giving greater confidence that the results aren't driven by poor or biased choices.

As stated in the discussion, the inversion took approximately 27 hours to complete on a large computer cluster. This represents a serious impost in terms of computational time over and above traditional least squares optimisation techniques. The reason for this is a least squares optimisation approach may require hundreds of forward model evaluations, whereas in a complex model, a sampling approach may require millions of forward model evaluations and this large number of forward model evaluations is by far the dominant computational effort of the inversion. As access to super-computing facilities becomes widespread it is envisaged that this negative will be outweighed by the benefits of the approach in terms of the more detailed information obtained from the inversion that can be used to rigorously test hypotheses about the structure of the subsurface.

It is worth noting that the trans-dimensional approach does not preclude the use of traditional least squares optimisation, but is complementary to it. Regions previously inverted using traditional techniques where the results prove interesting or paradoxical could be re-examined with the trans-dimensional approach to gain deeper insights into the range of possible subsurface conductivities. Similarly, inversions obtained through traditional least squares approaches could be used as starting models for subsequent trans-dimensional inversion to save computational time. Alternatively, the trans-dimensional approach can be used to study the noise processes and to estimate empirical data noise covariances as shown in the third inversion and these noise models can then be used in large scale deterministic inversions.

By its very nature, time domain AEM is an imaging problem with non-unique solutions. Tackling this through ensemble based methods, such as trans-dimensional sampling is viable, useful, and will become increasingly practical in the future.

Chapter 5

Ambient noise tomography

5.1 Introduction

Chapter 2 showed how to obtain inter-station path average phase velocity dispersion information for both Love and Rayleigh waves, the observations required for ambient noise tomography. The focus of this chapter is on the tomographic inversion of these observations to construct a local 3D model of the structure of the Earth.

There are many approaches to this particular tomographic inversion problem ranging from linearised optimisation approaches [Tarantola and Valette, 1982b], to iteratively non-linear trans-dimensional inversion [Bodin et al., 2012a] to fully non-linear trans-dimensional inversion [Galetti et al., 2016].

The commonality between approaches is a two step approach where regionalised group or phase velocity maps at selected periods are inverted from group or phase velocity information, and in a second step, 1D radial or depth profiles of shear wave velocity are inverted at selected points, to generate desired 2D transects through regions of interest or full 3D models. A recent alternative to this is to directly invert a shear wave model [Fang et al., 2015]. However, this approach linearises the problem about a reference shear wave model.

The two step process will be followed here, with novel use of the trans-dimensional tree method coupled with a fully non-linear forward model for studying the deep crustal structure of Iceland.

5.2 Iceland

Iceland is located in the mid North Atlantic ocean atop a spreading ridge between the Eurasian and North American plates. The high degree of geothermal, volcanic and seismic activity has led to a wealth of studies on the region and much debate about the geodynamics that drive these processes.

Volcanism in Iceland, particularly more recent activity, is concentrated along the spreading ridge which approximately transects Iceland in a North-East direction [Thor-

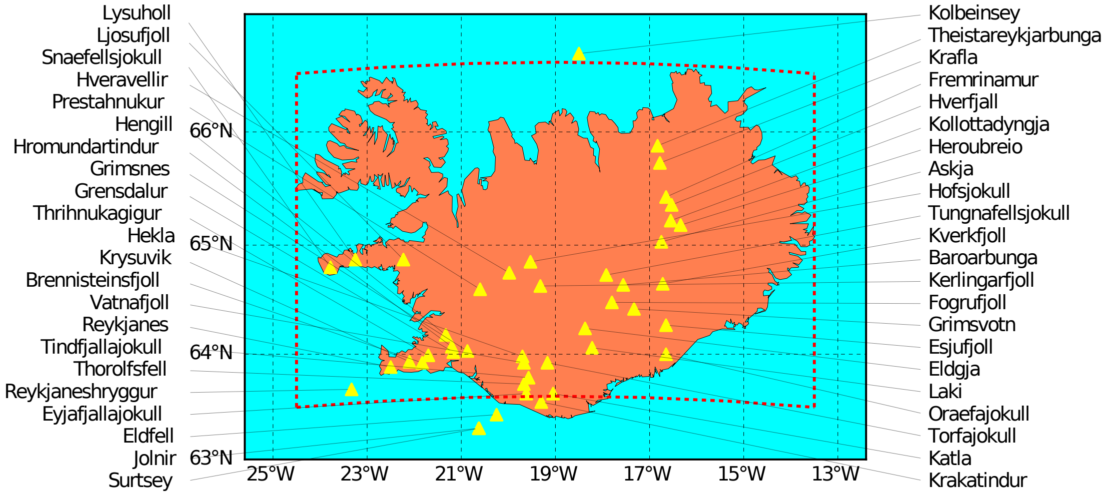


Figure 5.1: The spatial location of recent and historic volcanism in Iceland indicated with yellow triangles from a catalogue [Thordarson and Larsen, 2007]. The red dashed box represents the boundary for this ambient noise study.

darson and Larsen, 2007, Gudmundsson and Högnadóttir, 2007, Tkalčić et al., 2009]. The location of known volcanism in Iceland is shown in Figure 5.1.

Iceland is seismically active [Jakobsdóttir, 2008], again predominantly about the spreading ridge. Refraction studies of the crust [Darbyshire et al., 1998, 2000] suggest considerable crustal thinning in the volcanic regions with a thickness of around 3km, whereas surrounding areas have a crustal thickness between 25 and 40km.

Previously, the structure of Iceland was explored using ambient noise tomography by Gudmundsson et al. [2007] using the stations from the HOTSPOT deployment [Allen et al., 1999]. The same stations are used in this study and are shown in Figure 5.2. A recent study [Green et al., 2017] using a larger number of stations from more recent seismometer deployments has revealed more detail of the crustal structure particularly

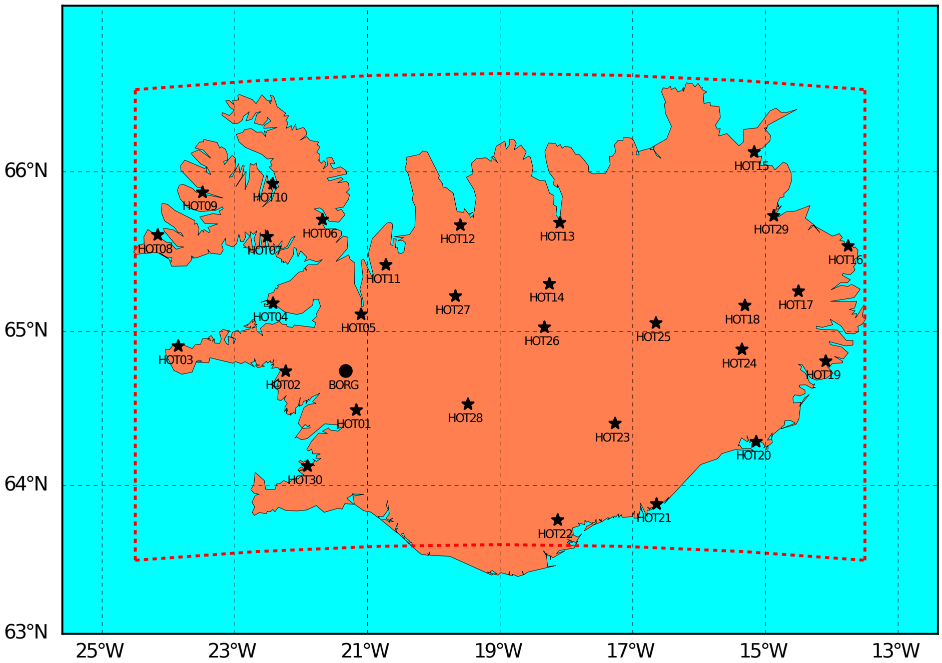


Figure 5.2: The seismic stations used in this ambient noise tomography case study

above the spreading ridge.

The location, continuity and even existence of a mantle plume beneath Iceland has been controversial [Allen et al., 1999]. Earthquake driven surface wave studies such as that of Li and Detrick [2006] and Pilidou et al. [2005] resolve up to 400 km, yet fail to resolve a plume at depth. Early tele-seismic tomography results failed to see deeper than around 600km [Foulger et al., 2001] to 1,000 km [Hung et al., 2004] due to the aperture of stations on Iceland. Receiver function studies have been similarly limited [Shen et al., 1998, 2002, Jenkins et al., 2016]. It has only been more recent regional and global studies that consistently see slow velocity anomalies in the deep mantle that may correspond to a mantle plume beneath Iceland [Ritsema et al., 1999, Rickers et al., 2013].

This study focuses on the application of new techniques to inverting the ambient noise data using the same source stations as those used by Gudmundsson et al. [2007]. The aim is to examine periods up to 30s to resolve features between 10 to 100km though higher uncertainty is expected at greater depths.

5.3 Problem description

Seismic tomography is a well established technique with 40 years of history stemming from early works of Aki [1977]. In traditional seismic tomography, seismic waves propagate from a point source, typically an Earthquake, to a receiving station that observes a waveform some distance away. There are three general classes of seismic tomography that vary in the modelling approach used for the propagation of the seismic signal through the Earth.

The first is based on ray theory [Červený, 2001], which uses a high frequency approximation for the propagation of the seismic waves. The high frequency approximation reduces the sensitivity of travel time observations to a single infinitesimally thin ray path between source and receiver. Hence in ray theory tomography, travel times are picked from received seismograms and these are compared to travel times computed by performing ray path integrals in candidate models.

In the second method, Dahlen et al. [2000] introduced an extension to ray theory that accounted for multiple scattering called finite frequency kernels. This extends the ray path integral to incorporate the frequency dependence of multiple scattering ray paths. However, most implementations of finite frequency kernels have used a single scattering paraxial approximation whereby parameters of scattered rays are approximated using values computed along the direct ray. In this sense, finite frequency kernels using the paraxial approximation are similar to Fresnel volume techniques [Červený and Soares, 1992]. This has the capacity to improve resolving power by both increasing the coverage across the domain due to the thickness of the kernels, and by effectively multiplying the available observations using multiple frequency bands for the same event-station pair. A common criticism of this approach is that it linearises the forward problem about a reference model and sufficient errors in the reference model adversely affect the final result. There has been only one study where kernels were iteratively refined in the literature [Gautier et al., 2008].

Lastly, in recent years with advances in computational power, full waveform solvers based on the spectral element method have become popular for tomography problems at both regional and global scales. The common approach in using full waveform techniques is ad-joint tomography where for each event-station pair, a forward simulation and a reverse simulation are run using a candidate model to generate a sensitivity kernel based on a specified misfit function. The collection of these sensitivity kernels, one for each observation, are then used to perturb the model using a gradient descent approach often modulated with smoothing constraints or constraints based on sensitivity coverage. This process is repeated a number of times, often with relatively few data and model parameters due to the huge computation cost, until the model is judged to have converged.

What these techniques represent is a spectrum through the sophistication of the physical modelling. At the low sophistication end there is linearised tomography based on fixed rays computed using a reference model progressing through non-linear ray tracing, finite frequency kernels, to ad-joint full-waveform tomography. There is always a question of how much improvement is afforded in model construction by adopting more sophisticated physical forward modelling.

In this ambient noise study, the source observations are path average dispersion curves between two stations. Additionally, a sampled Bayesian approach requires a large number of forward model calculations which prohibits the use of computationally expensive physical simulations. The desire here is also to go beyond existing linearised approximations and iteratively non-linear approaches and for these reasons, full non-linear ray theory will be the focus of this study.

5.4 The forward model

In a ray theoretical approximation, at any given frequency, the propagation of the surface wave between two points, either Love or Rayleigh, occurs along an infinitesimally thin ray. The path this ray takes is directly related to the velocity field between two points and dictated by Fermat's principle, that is, the minimal travel time between two points. In a uniform velocity field, the ray path between two points is a straight line. If a slow velocity anomaly were located between the two points, the ray path would appear to deviate away from it. Similarly, fast velocity anomalies tend to bend or focus ray paths towards them.

A key component of the forward problem is accurately predicting the time of first arrival of seismic surface waves between two points in a heterogeneous velocity field. There are many methods for computing seismic rays through a media and they fall into three general classes, shooting, bending and network algorithms [Červený, 2001].

In the shooting algorithms [Julian and Gubins, 1977, Sambridge and Kennett, 1990], the differential equations of the path of the ray are solved numerically from a starting point and shooting direction. The path of the ray will progress through the domain of interest until it intersects with the target destination or it is deemed that it will fail to do so. This process is iterated until the ray intersects sufficiently close to the target point.

Ray bending approaches [Um and Thurber, 1987, Grechka and McMechan, 1996] generally start from a straight path or reference path between two points and the ray is iteratively bent to more closely satisfy Fermat's principle up to some stopping criteria.

Finally, the network algorithms use a combination of a fine mesh or network of control points and a Djikstra [Djikstra, 1959] like algorithm to compute a travel time field from a source location. Ray paths between two points can then be computed using the gradient of the travel time field. This approach has in recent years been improved with a stable finite difference updating scheme and has become known as the fast marching method (FMM) [Sethian and Popovici, 1999, Rawlinson and Sambridge, 2005, Rawlinson et al., 2007].

This last scheme seems a best fit for the problem at hand as the ray path itself is unimportant, all that is needed is the travel time. From the travel time and knowledge of the inter-station distance, the path average phase velocity can be computed. Additionally, observations are between known stations so although there may be a large number of ray paths, the number of stations is relatively small. Due to reciprocity, with FMM the travel times of all ray paths can be computed from travel time fields computed for each station bar one in the worst case. For example, with the ambient noise observations collected for Iceland there are 435 observations from 31 stations (some station pairs do not overlap in time). For this problem 28 travel time field calculations are required to predict travel times for the 435 station pairs. With other ray calculating methods, it would be necessary to compute each of the 435 rays independently introducing a much larger computational burden.

5.4.1 Uniform tests

For the purposes of inversion it is important to understand the limitations of the forward model to accurately predict travel times or path average velocities between points. To explore this, a simple experiment is to use a uniform velocity model of 3 km/s and the station locations from the Iceland case study with fixed ray paths shown in Figure 5.3. The travel time can then be computed between each station pair from which the predicted path average velocity is obtained by dividing by the great circle distance between stations. It should be expected that all inter-station path average velocities will be 3 km/s but due to numerical imprecision and the finite resolution of the grid there will be some degree of error.

The tests that follow were performed using a custom FMM implementation with a second order upwind scheme and local mesh refinement around the source point. Both of these improvements to the basic FMM algorithm have been shown to significantly improve the accuracy of the method [Chopp, 2001, Rawlinson and Sambridge, 2005].

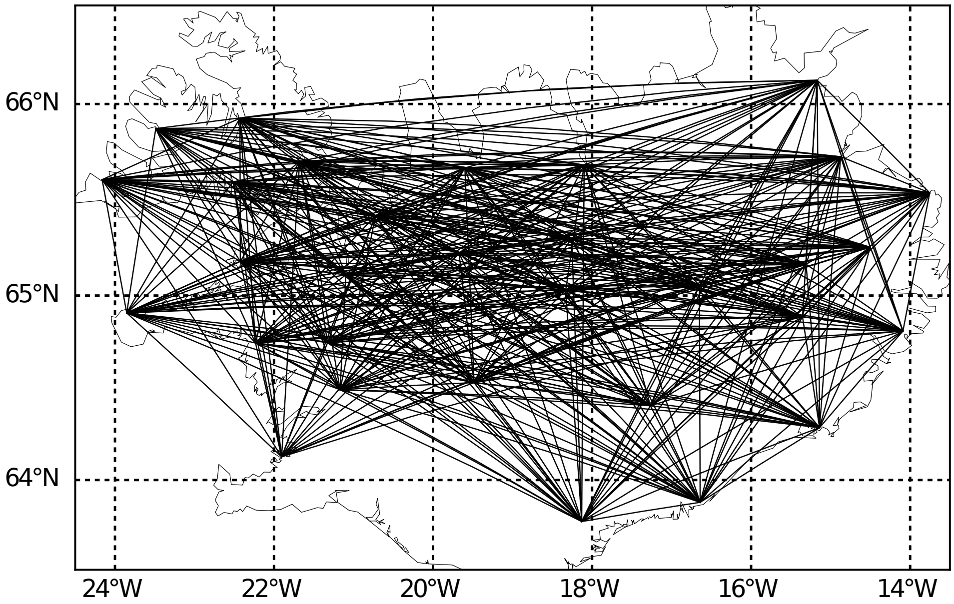


Figure 5.3: The great circle paths between stations used in this study to give an appreciation of the coverage available to synthetic and real data inversions.

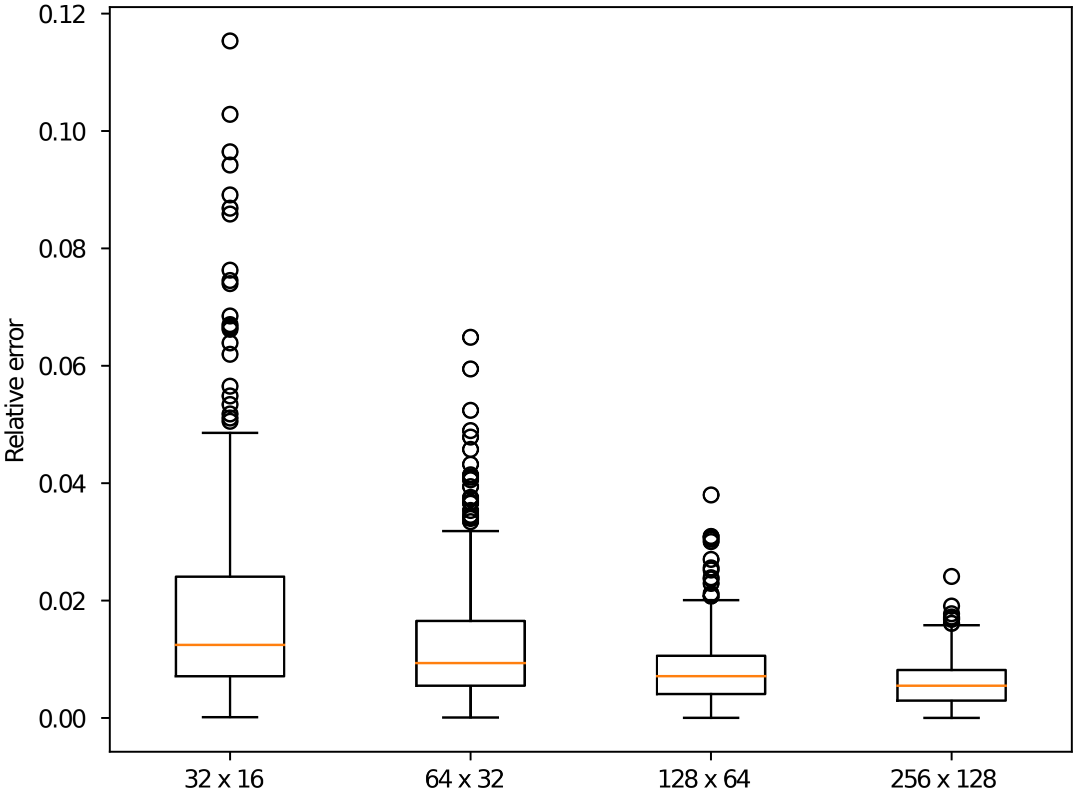


Figure 5.4: The distribution of errors between the known true path average velocity computed with the fast marching method for increasing resolutions in a uniform velocity model. The relative error is computed from the path average velocity obtained with the fast marching algorithm with respect to the uniform velocity. In the box plot, the orange line is the median, the box represents the range of the first and third quartiles, the whiskers the 9th percentile and the 91st percentile, and circles outliers.

For a series of increasing resolution grids, FMM is used to compute the the relative error of the path average velocity from 3 km/s and the results are summarised in Figure 5.4. In the plot, for each resolution the distribution of the relative error for all paths is shown with a box plot where the horizontal orange line is the median error, the box represents the range from the first to the third quartile, the whiskers are at 1.5 times the inter-quartile range, and the circles represent outliers. As the resolution of the underlying grid increases, the distribution of the errors descends to near zero. For example, at a resolution of 128 by 64 cells, all paths have a relative error less than 5 percent and the relative errors are predominantly restricted to less than 2 percent.

The computational cost of the FMM is a function of the number of grid cells. This

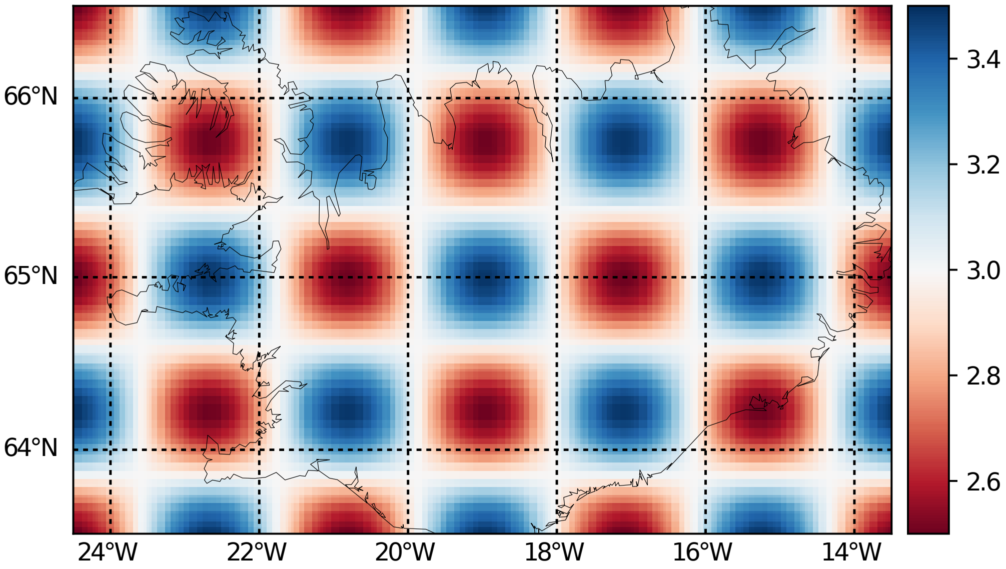


Figure 5.5: The synthetic checker board model used to test the accuracy of the fast marching method in the Iceland region.

means that a double of the number of cells horizontally and vertically results in an approximate four times increase in the computational cost. Hence there is a trade-off between the numerical accuracy of the method and the computational expense.

5.4.2 Checker board comparison tests

To extend this test beyond a simple uniform velocity field, a similar experiment can be performed with a smooth checker board model, shown in Figure 5.5. In this case a known solution is not available and the errors reported are relative to a higher resolution calculation of travel times using a resolution of 256 by 128 cells. The results are shown in Figure 5.6 where the same pattern of decreasing errors with increasing resolution is observed.

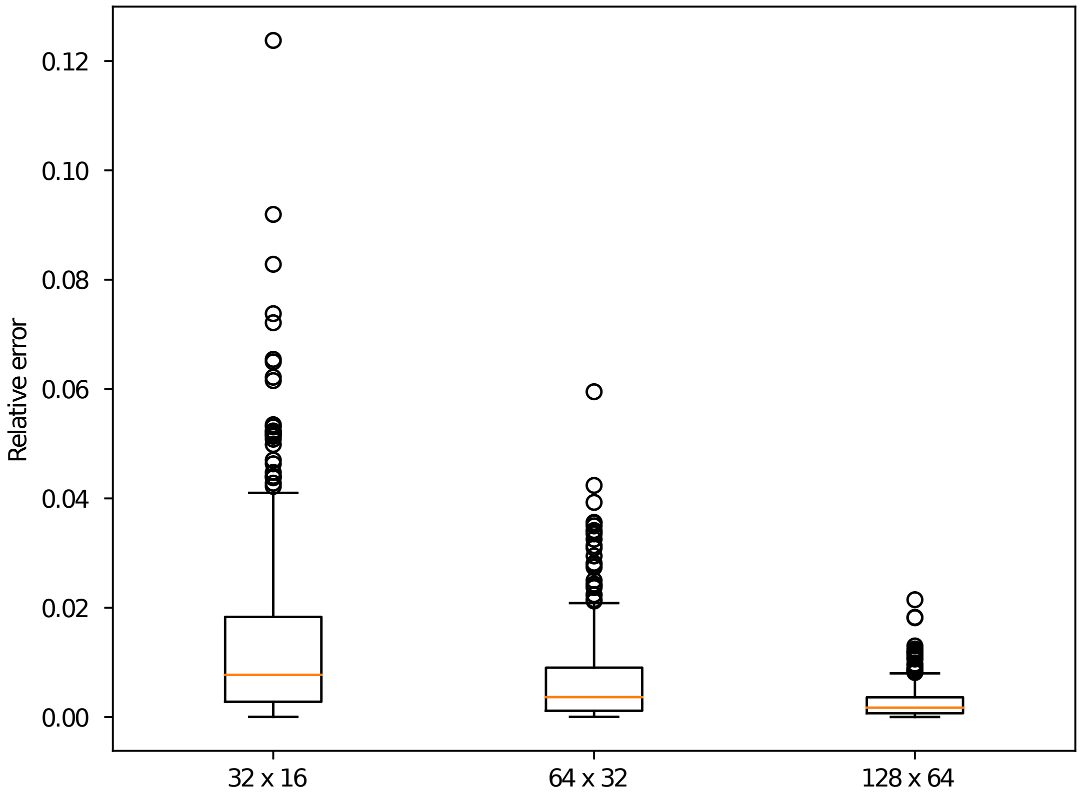


Figure 5.6: The distribution of errors between the path average velocity computed with the fast marching method for varying resolutions compared to a 256 by 128 cell grid resolution solution.

5.4.3 Summary

In two different tests, using the realistic ray geometry taken from the stations used in this study, it was demonstrated that travel times can be computed with a grid resolution of 128 by 64 cells. For this resolution, the median relative error is significantly less that one percent and the maximum relative travel time error is less than four percent. Increasing the number of cells further is possible, but at significant additional computational costs.

5.5 Wavelet super resolution

In a wavelet parameterisation, the model is a vector of wavelet coefficients and an image is constructed from these coefficients using the inverse wavelet transform [Mallat, 1999]. A property of wavelets is that a set of wavelet coefficients uniquely defines an image at any grid resolution that is a multiple of two of its base size. For example, a set of wavelet coefficients representing an image of 32 by 16 pixels, can equally construct an image of 64 by 32, 128 by 64, 256 by 128 and so on. This ability, often called wavelet super resolution, follows from the multi-resolution property of wavelets.

In an geophysical inverse problem, the resolution of the model parameterisation chosen from knowledge of resolvable length scales may not give adequate forward model accuracy. A solution is to use wavelet super resolution to up-scale the model parameterisation to a sufficient resolution to ensure accurate physical modelling.

5.5.1 Uniform velocity

Before applying the super resolution property of wavelets, its accuracy needs to be verified in similar tests to the previous section. The configuration for these tests is that travel times are computed in a high resolution grid (256 by 128 cells), and various lower resolution wavelet representations use wavelet super resolution to up-scale to a 256 by 128 cell image for computing travel times. In Figure 5.7, the results of progressively

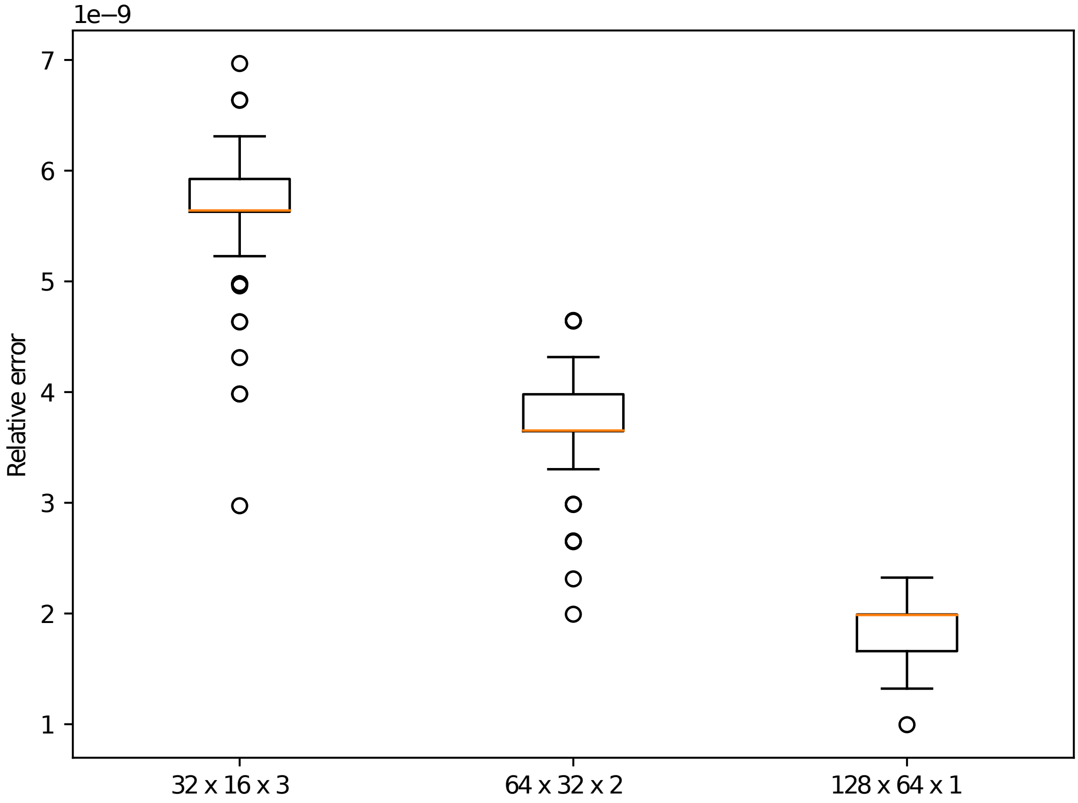


Figure 5.7: The distribution of errors between the path average velocity computed with the fast marching for a series of wavelet super resolution steps relative to a calculation performed on a 256 by 128 grid of a uniform velocity model.

increasing base resolutions to the more accurate result are shown. The notation in the figure labels corresponds to width by height by super resolution steps. For example, 32 by 16 by 3 corresponds to a base resolution of 32 by 16 with 3 super resolution steps each of which doubles the resolution. Hence 32 by 16 by 3 means a wavelet model at 32 by 16 with the travel times computed on a 256 by 128 grid.

As can be seen from the figure, the accuracy is to within machine precision of the wavelet transform which is to be expected for a uniform velocity model.

5.5.2 Checker Board velocity

The same test is again repeated with the true checker board model in Figure 5.5 with respect to a reference solution computed on a 256 by 128 grid. This effectively tests the

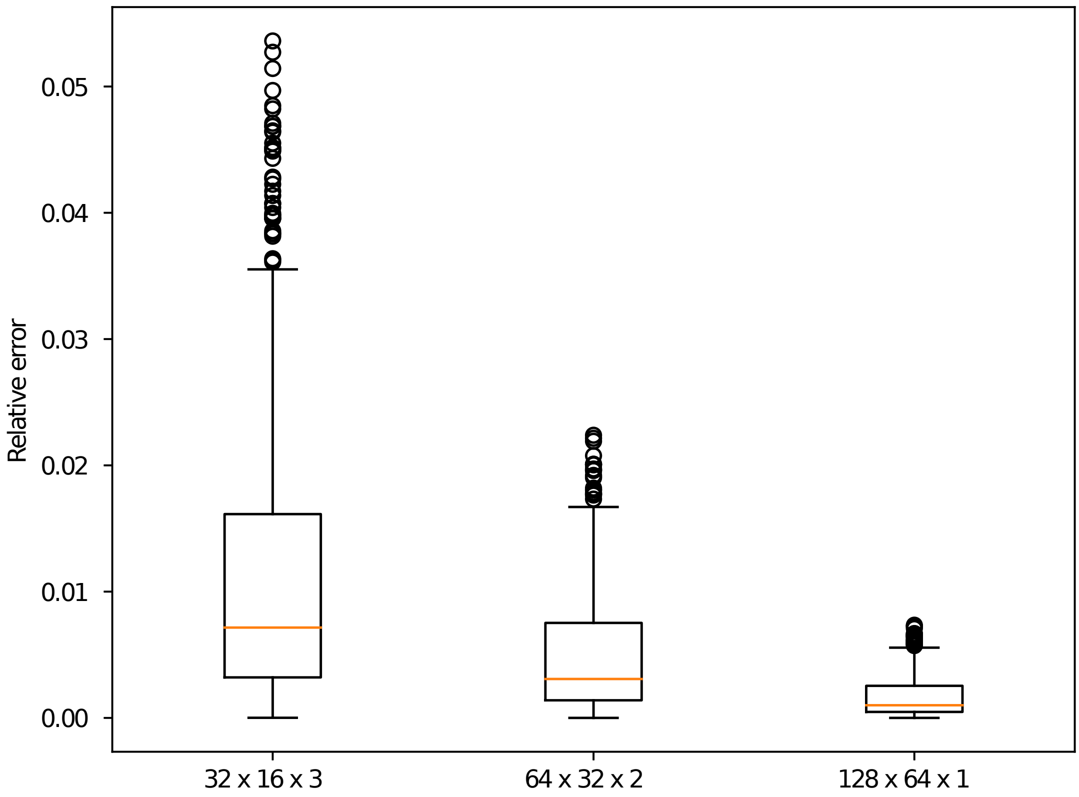


Figure 5.8: The distribution of errors between the path average velocity computed with the fast marching for a series of wavelet super resolution steps to a calculation performed on a 256 by 128 grid of a checker board velocity model.

ability of the wavelet representation to accurately represent the checker board model with a lower resolution or truncated set of wavelet coefficients. The results shown in Figure 5.8 demonstrate that the relative errors introduced by the wavelet super resolution are quite small.

In very complex models where lower resolution wavelet approximations may be poor predictors, errors may be larger. However, for the characteristic scale length of models that are expected to be resolvable with surface wave observations, this is likely sufficient.

5.5.3 Summary

In this section, the use of wavelet super resolution to increase the accuracy of travel time predictions was demonstrated. This allows tight coupling between a model parameterisation specified with wavelet coefficients to be up-scaled into a regular grid image of arbitrary resolution to ensure forward model accuracy. As is common in any geophysical inversion problem, there is a trade off in terms of accuracy versus computational effort, and any choice here must be guided by tests to ensure the accuracy is satisfactory for the purpose at hand. The accuracy tests presented here give confidence to apply this set of methods to a complete tomographic problem.

5.6 Bayesian formulation

This tomographic inversion will apply the trans-dimensional tree approach with a wavelet parameterisation. Stemming from the success of this approach in the Airborne Electromagnetic tomography in the previous chapter, the formulation of the problem shares much in common. Both hierarchical error scaling and the hierarchical Laplacian prior are used. With ambient noise tomography, the observations will in general only be able to resolve smooth features so the CDF 9/7 wavelet basis is used. Parallel tempering is also used to improve the robustness of the inversion. For much of the formulation, the previous chapter provides much of the background and highlight here the differences in this application, namely the quality of the observations, the likelihood, and the forward model.

In the tomographic problem, the observations are path average phase velocities between two points with associated uncertainties obtained using the method in Chapter 2. Previous trans-dimensional tomographic methods [Bodin et al., 2012a] have been based on travel time observations. These were based upon ambient noise group velocity estimated using Frequency-Time analysis (FTAN) [Dziewonski and Hales, 1972]. For error estimation, a hierarchical approach was used to invert for the standard deviation of an identically independently distributed Gaussian error process. The drawback with this method is that the same level of error (in seconds) is assigned with every path regardless of length. This states that the expected level of error on a 10 km path is the same as that on a 1000 km path. In an effort to address this, Galetti et al. [2016] have incorporated a distance factor into their noise model. In the approach outlined in Chapter 2, estimates of phase velocity dispersion curves with uncertainties are obtained and these can be used directly in a likelihood expressed as

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left\{-\sum_{i} \frac{(G(\mathbf{m},f)_i - d_i(f))^2}{2(\lambda\sigma_i(f))^2}\right\}, \tag{5.1}$$

where d_i is the previously measured phase velocity and σ_i its associated posterior uncertainty estimate. Recall that the results in Chapter 2 generated continuous dispersion curves and their uncertainties which form the observations used here, and so f represents the given frequency of interest. $G(\mathbf{m},f)_i$ is the forward model that consists of computing the travel time, coupled with the known inter-station distances, to produce a model predicted velocity. Lastly the λ parameter is the hierarchical scaling parameter. The point here is that good estimates of uncertainty are available in terms of phase velocity and this likelihood is faithful to the observations.

Since good estimates of the uncertainty in the phase velocity measurements are available, it is reasonable to ask why the hierarchical error scaling term is required? The reason is that the assumed noise process is of the form

$$v_{\text{observed}} = v_{\text{true}} + \epsilon_{\text{observational}} + \epsilon_{\text{theory}} + \epsilon_{?}, \tag{5.2}$$

that is, for a given observed phase velocity v_{observed}, the noise is a combination of the observational noise, obtained from the posterior of the inversion in Chapter 2, theoretical modelling errors from approximations in the forward model, and other unknown errors. The likelihood approximates these sources of errors as a single independent Gaussian noise model. So the observational errors should be the dominant source of errors, but the approximations in the physical modelling of the problem introduces additional error that is accounted for by the scaling of the observational errors using a hierarchical error scaling term λ.

5.7 Resolution Tests

A first test of this tomographic method is a synthetic test, and while the use of checker board tests is not without controversy within the seismology community, it remains a standard testing approach. The main criticism of checker board tests stems from their failure to properly give a indication of the resolving power of the inversion [Lévêque et al., 1993, Rawlinson and Spackman, 2016]. Rather than enter into this debate here, checker board resolution tests are used for a synthetic simulation to demonstrate the operation of the algorithm.

In this experiment, a fine 256 by 128 resolution grid is used to trace rays in a series of increasing resolution checker boards between all station pairs available in the Iceland experiment; the same observations as will be used in the full inversion. Fixed great circle ray paths between stations were previously shown in Figure 5.3. To these "true" path average phase velocities, random independent Gaussian noise with a standard deviation of 0.05 km/s (approximately a 2 percent error level) is added to create the observations used in these experiments.

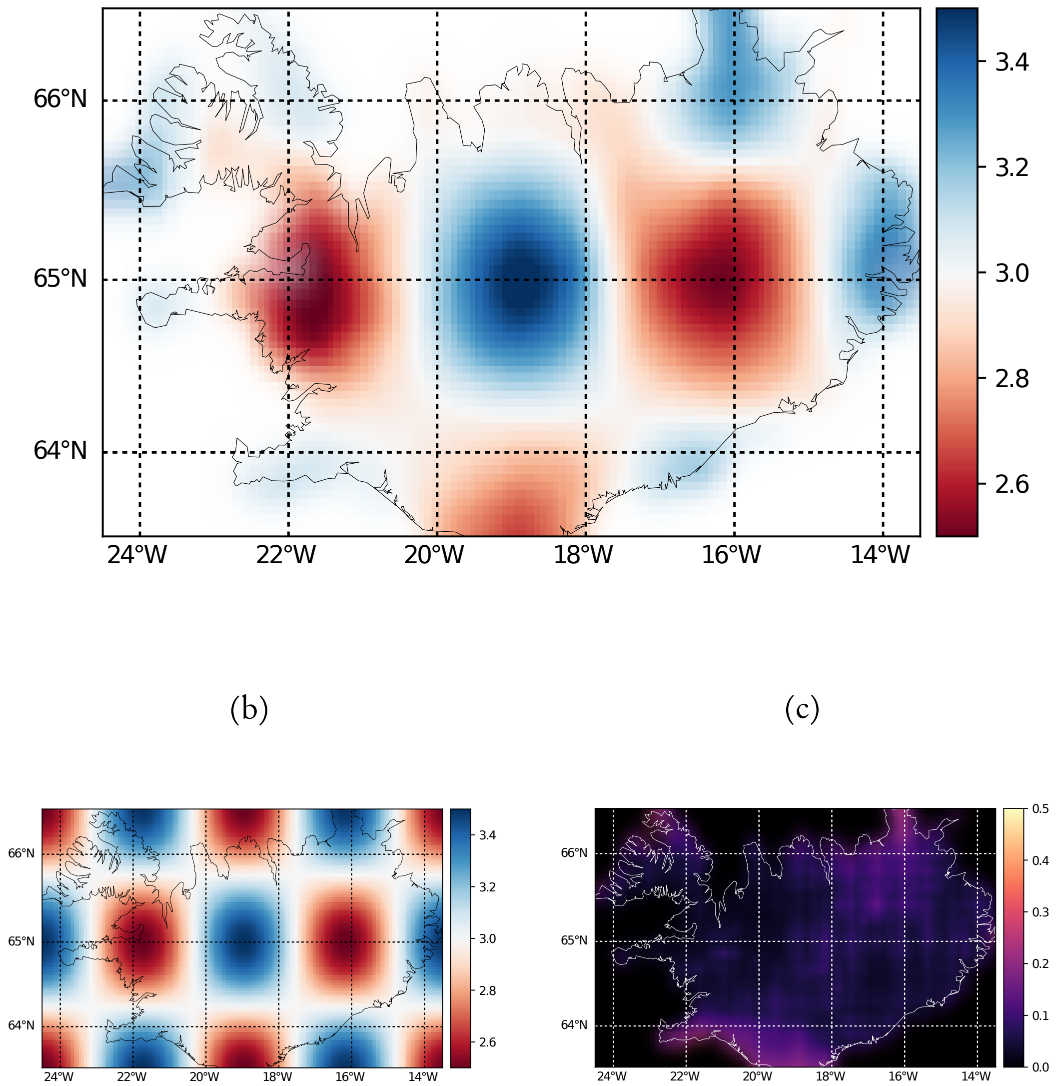
The synthetic observations were then inverted using the trans-dimensional tree with wavelet parameterisation approach in a fully non-linear inversion. A 64 by 32 grid was used for the trans-dimensional tree with CDF 9/7 wavelet parameterisation and one super resolution step. This means that the inversion travel times were calculated on a 128 by 64 resolution grid or half the resolution that the true observations were generated on. A series of inversions with increasingly finer checker boards patterns was run to test the recovery of the synthetic observations with the inter-station paths.

Hierarchical Laplacian priors and an inverse Gamma hyper prior with a mode of 0.25 ($\alpha = 1$ and $\beta = 0.5$) was used. For each inversion, six independent chains were used with parallel tempering using four temperature levels spaced logarithmically between 1 and 5, for a total of 24 chains. Each of the inversions was initialised with a uniform velocity field at the mean path average velocity of the observations for a total of 2 million iterations.

The recovery of the checker board models is shown in Figures 5.9, 5.10, 5.11 and 5.12. In each of these figures, the ensemble median is shown in (a), the true model for comparison in (b), and the ensemble standard deviation in (c). The uncertainty varies depending on the target model, with areas of larger uncertainty tending to concentrate along a band approximately following the presumed location of the spreading ridge and an area of active volcanism. This reflects logistical issues in the siting of seismometers for the HOTSPOT project and unfortunately would appear to limit the ability to resolve this interesting area at high fidelity.

It is also interesting to compare the result with an inversion that uses fixed ray paths for the forward model instead of the fast marching method. To test this, the inversions are re-run with the fast marching method being replaced by fixed ray paths that follow the great circle arcs between stations. The results of these inversions are shown with the corresponding fast marching inversion in Figures 5.13, 5.14, 5.15 and 5.16. From a qualitative point of view, the recovery is similar in both with some slight artefacts apparent in the fixed ray case. More broadly it would seem that as the complexity of the model increases, the fixed ray inversion tends to damp the amplitude of the anomalies and this is particularly noticeable in Figures 5.15 and 5.16. This is due to large scale anomalies causing relatively minor deviations from great circle paths. As such, the results indicate that the fast marching method and linear rays should be comparable for less complex models.

Another qualitative observation in Figures 5.13, 5.14, 5.15, and 5.16 is that the pattern of uncertainty seems to change. While the effect is subtle, there does appear to be higher uncertainty in the inversions with fixed ray paths. This would support the assertions by Galetti et al. [2015] that use of more realistic physics is necessary to capture the true uncertainty in the inversion. Only a comparison between fixed rays and a fully non-linear scheme is presented here. A middle ground approach where rays are updated at some interval as performed by Bodin and Sambridge [2009] has not been

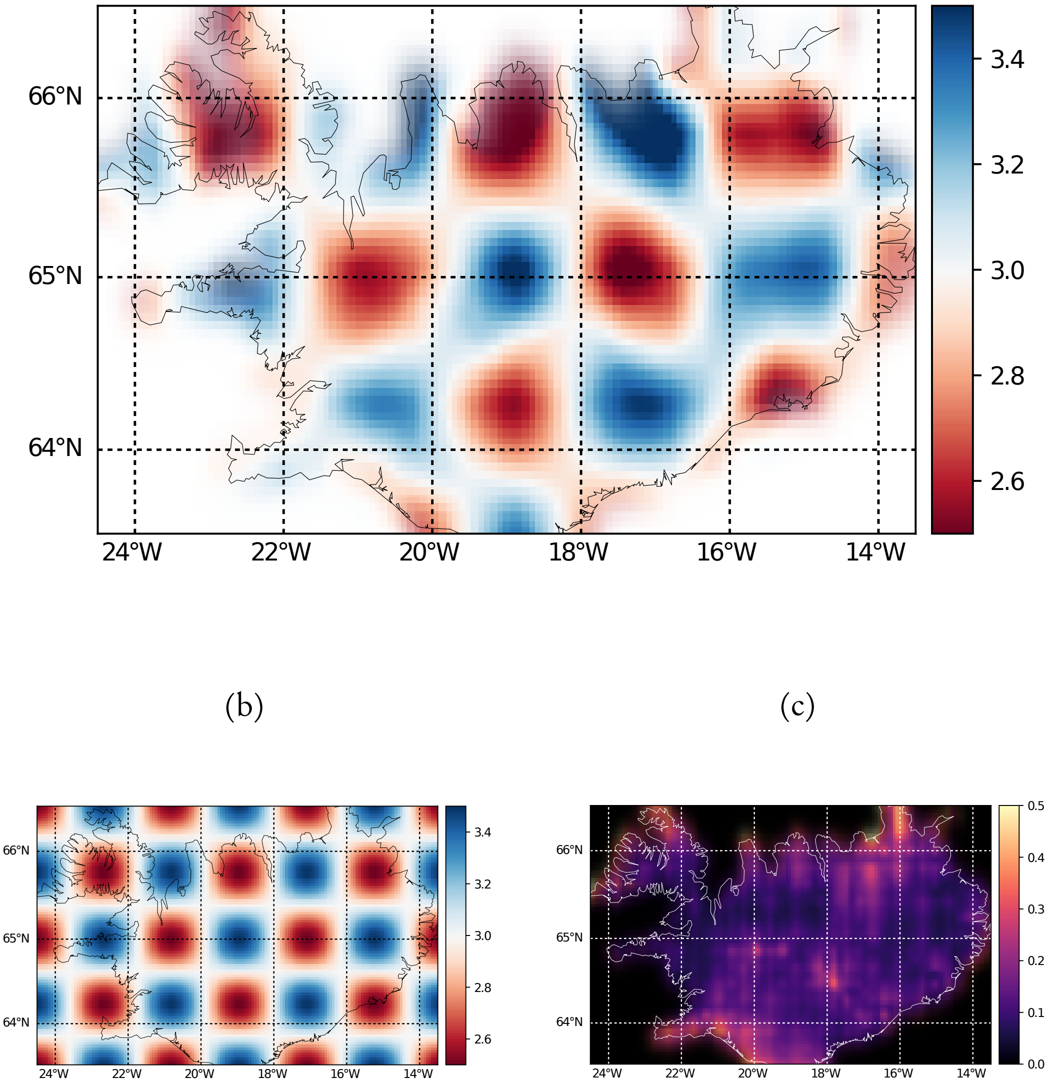


(a)

(b)

(c)

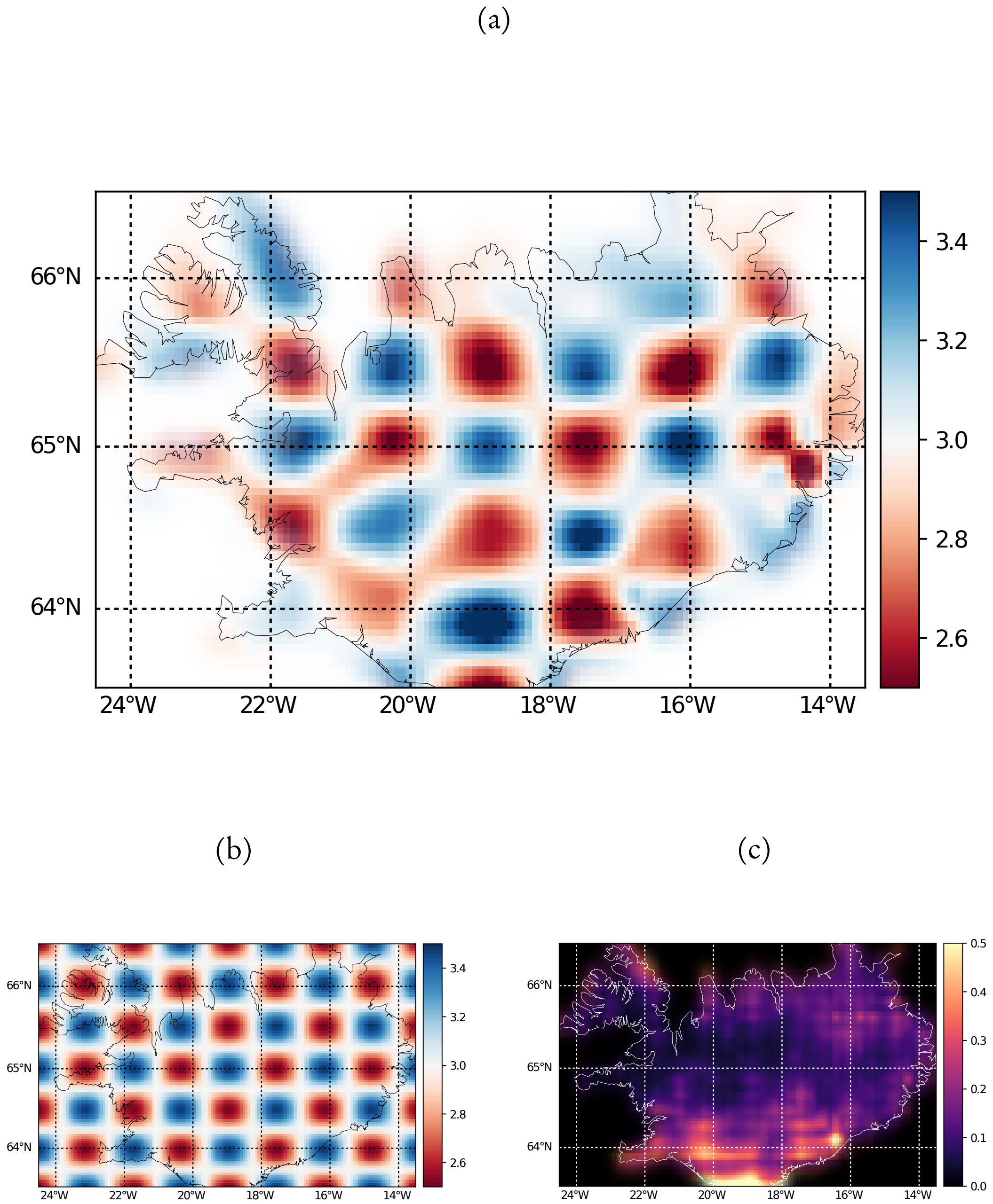
Figure 5.9: Synthetic recovery of the lowest resolution checker board test. The median of the ensemble is shown in (a) which should be compared to the true in (b). The standard deviation of the ensemble is shown in (c) which gives a measure of uncertainty.


(a)

(b)

(c)

Figure 5.10: Synthetic recovery of the medium resolution checker board test. The median of the ensemble is shown in (a) with the true model in (b). The she standard deviation of the ensemble is shown in (c) which gives a measure of uncertainty.


(a)

(b)

(c)

Figure 5.11: Synthetic recovery of the fine resolution checker board test. The median of the ensemble is shown in (a) with the true model in (b). The she standard deviation of the ensemble is shown in (c) which gives a measure of uncertainty.

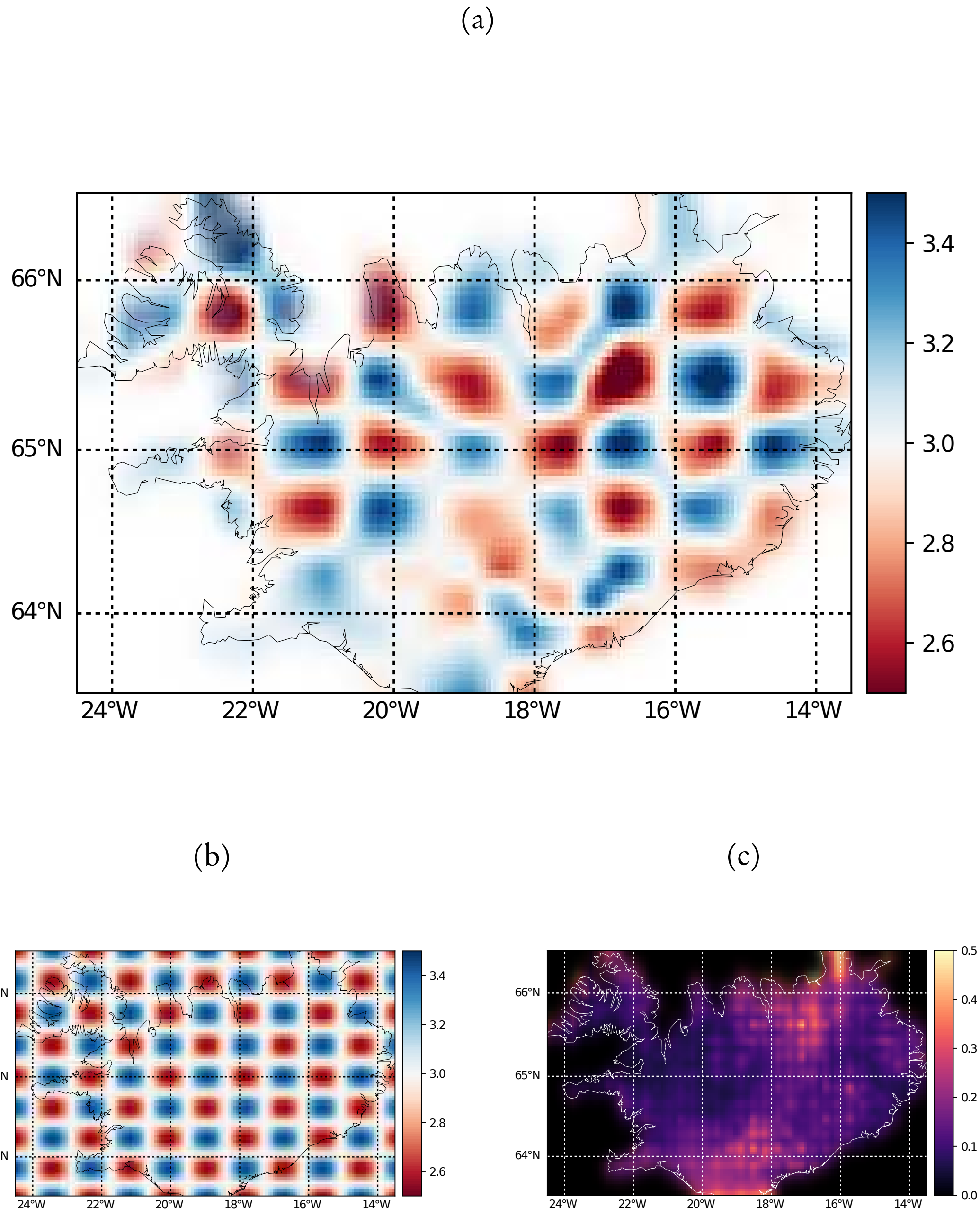


Figure 5.12: Synthetic recovery of the super fine resolution checker board test. The median of the ensemble is shown in (a) with the true model in (b). The she standard deviation of the ensemble is shown in (c) which gives a measure of uncertainty.

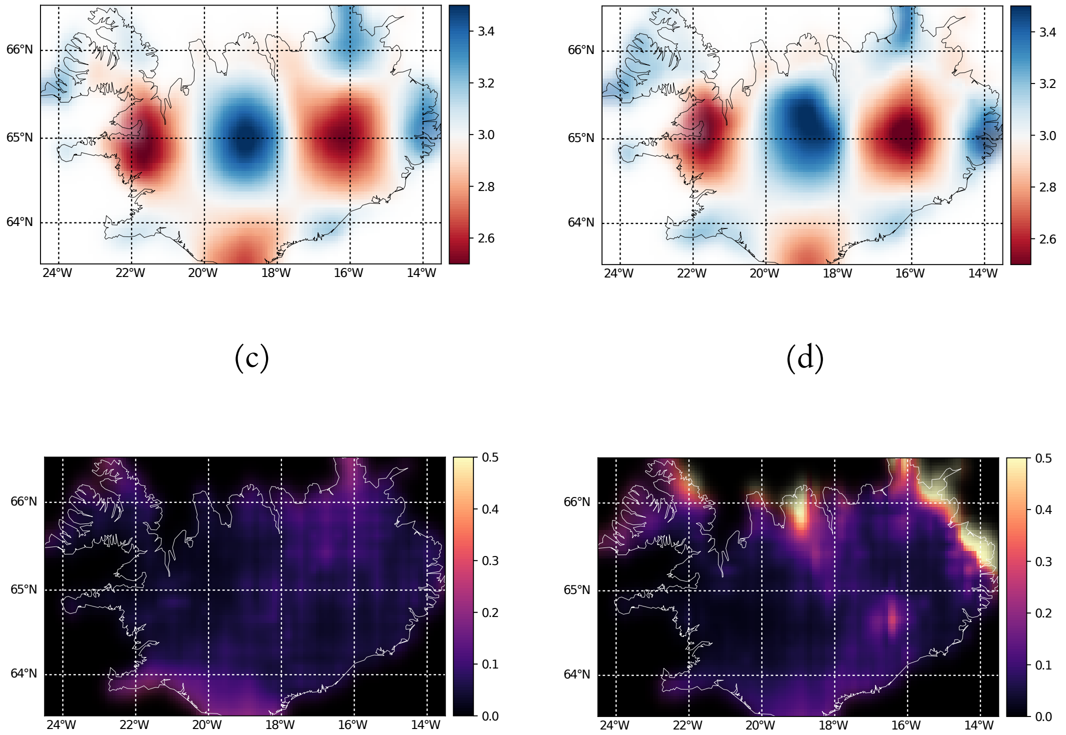

(a)
(b)
(c)
(d)

Figure 5.13: Comparison of synthetic recovery of the fast marching forward model (left column) versus fixed great circle paths (right column) of the low resolution checker board test. The median of the ensemble is shown in (a) and (b) and the standard deviation in (c) and (d).

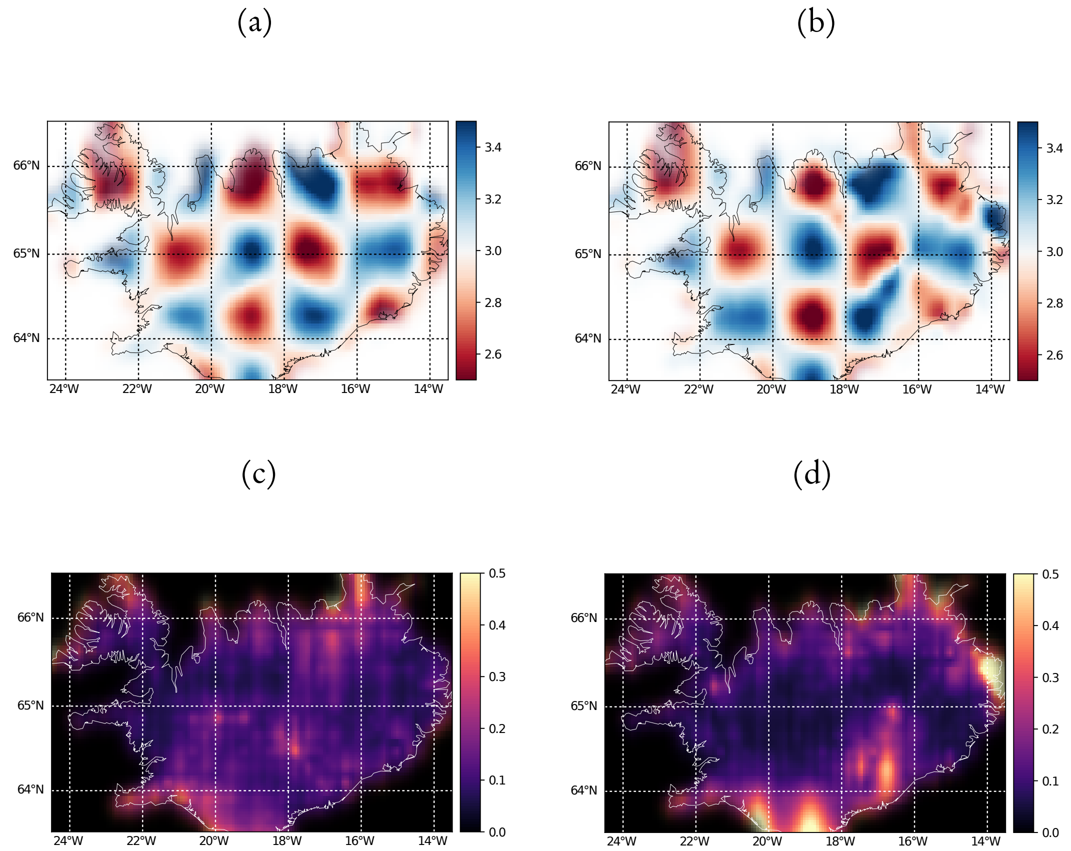


Figure 5.14: Comparison of synthetic recovery of the fast marching forward model (left column) versus fixed great circle paths (right column) of the medium resolution checker board test. The median of the ensemble is shown in (a) and (b) and the standard deviation in (c) and (d).

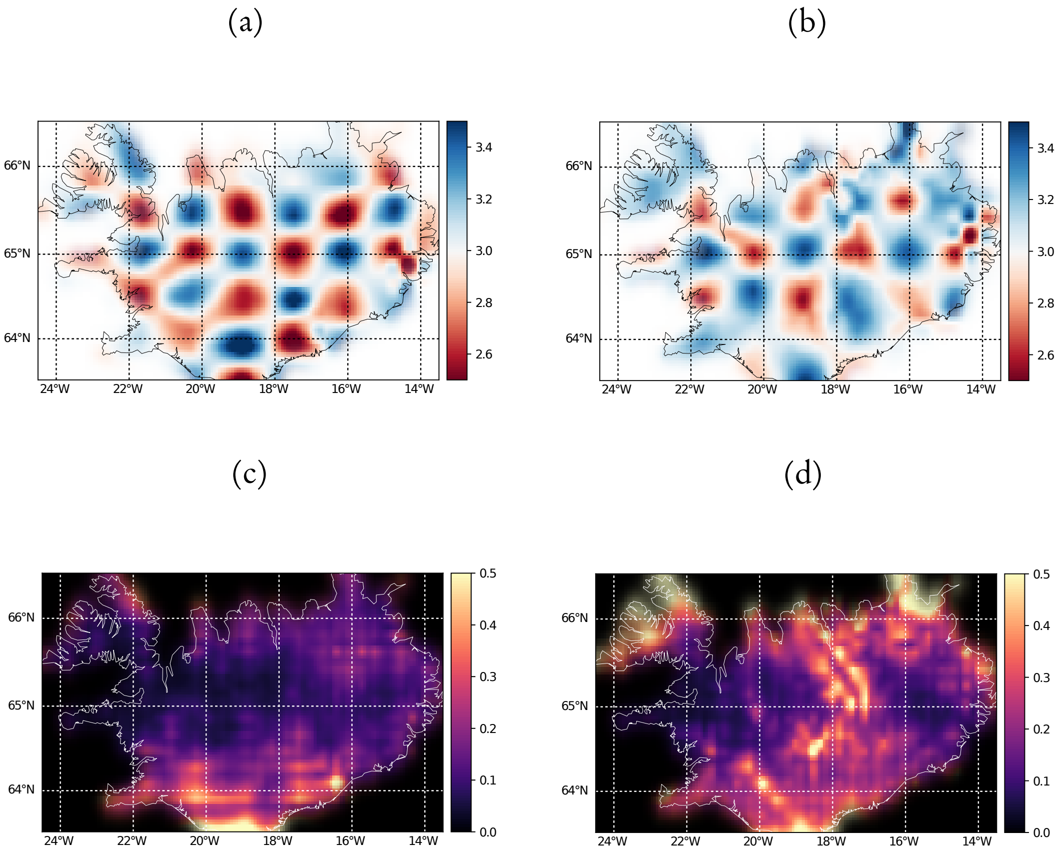


Figure 5.15: Comparison of synthetic recovery of fast marching forward model (left column) versus fixed great circle paths (right column) of the fine resolution checker board test. The median of the ensemble is shown in (a) and (b) and the standard deviation in (c) and (d).

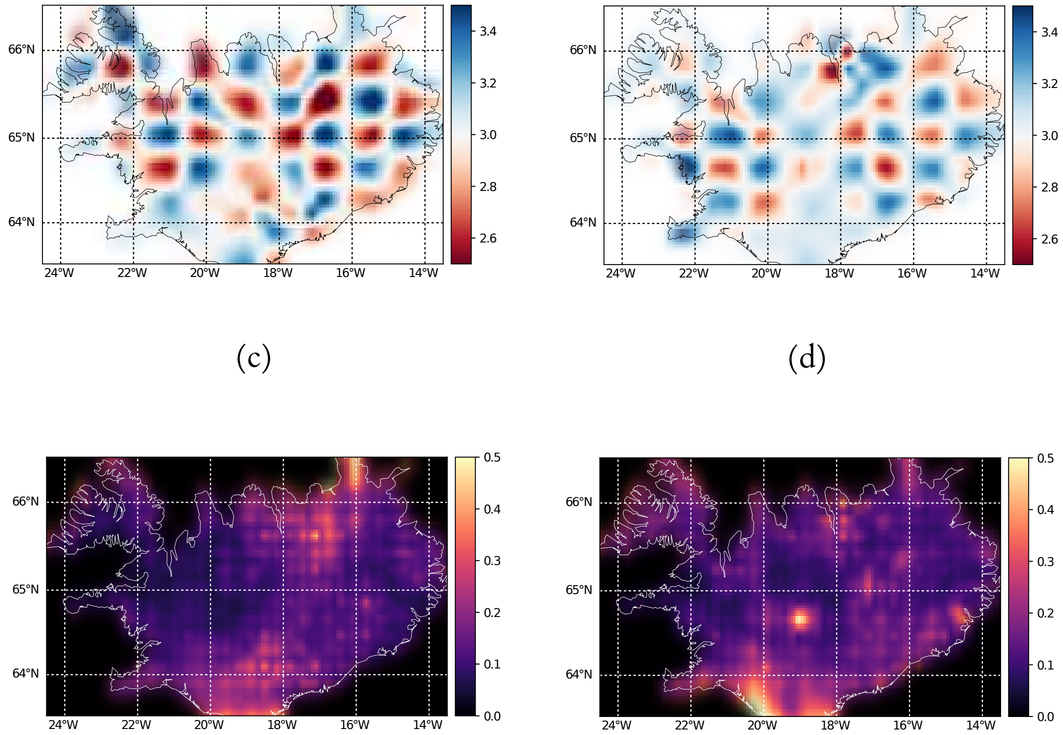

(a) (b) (c) (d)

Figure 5.16: Comparison of synthetic recovery of fast marching forward model (left column) versus fixed great circle paths (right column) of the super fine resolution checker board test. The median of the ensemble is shown in (a) and (b) and the standard deviation in (c) and (d).

evaluated.

One of the assumptions in a Bayesian inversion is that the noise model captures both observational and theory error as well as any other source of errors within the likelihood function. Hierarchical error scaling has been used in both fully non-linear and fixed rays inversions to invert for the level of noise appropriate for the data. Two inversions of the same synthetic models have been run: (1) where the forward model is nearly exact, namely the fast marching method, and (2) an approximation using fixed great circle paths. This is reflected in the hierarchical error scale estimation.

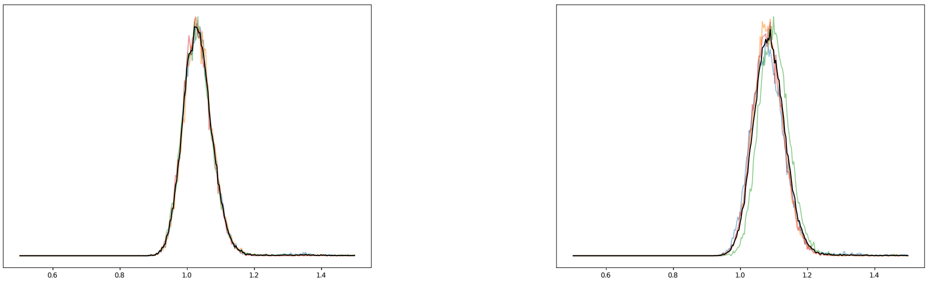
In Figure 5.17 the *a posteriori* histogram of the hierarchical error scale parameter is shown from the non-linear inversion in the left column and for the fixed great circle paths inversion on the right. In these synthetic experiments, since the theory error is zero, this histograms should be centred on the value of one, which appears to be the case for the non-linear inversion. For the linear approximation this also holds for the lower resolution checker board tests (b) and (d) but begins to break down as the resolution of the checker board model increases.

The demonstration of hierarchical error scaling increasing when an approximate forward model is used also goes someway to explain why more simplistic inversion strategies using linearisation assumptions coupled with regularisation strategies produce models of low resolution or that are overly smooth such as that in Green et al. [2017]. The use of physical approximations adds to errors and compensation comes in the form of regularisation parameters that result in over damped or smoothed models. Using fixed ray paths gives a conservative estimate but is inferior to the non-linear case, particularly as the resolution of the checker boards increases.

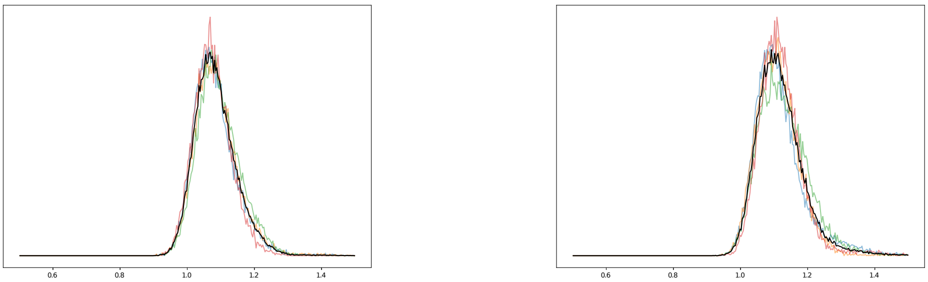
This section has shown several synthetic studies of the recovery of varying resolution checker boards using a new method for the inversion of phase velocity maps using trans-dimensional trees with a wavelet parameterisation and fully non-linear forward model. It has also been demonstrated that using fixed ray paths results in the reduced ability to resolve finer scale structure and the under estimation of velocity anomalies.

For these synthetic tests, the fixed great circle path inversion took approximately 45

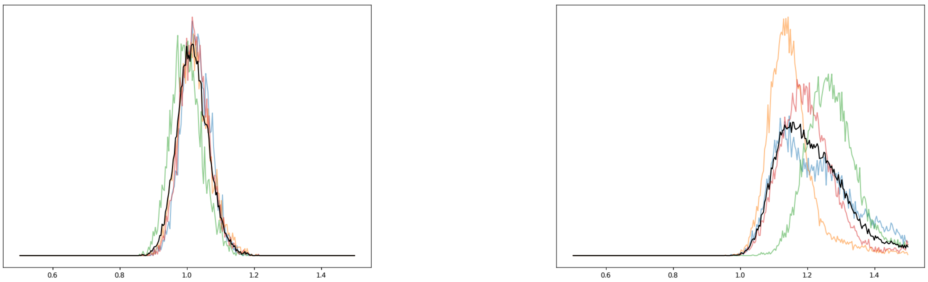
(a) Low non-linear (b) Low fixed



(c) Medium non-linear (d) Medium fixed



(e) Fine non-linear (f) Fine fixed



(g) Super fine non-linear (h) Super fine fixed

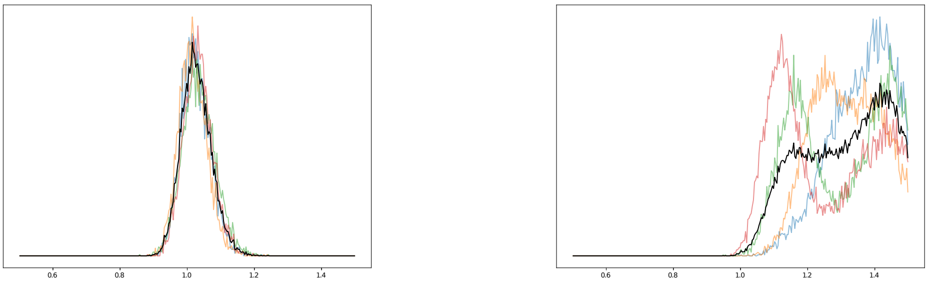


Figure 5.17: The comparison of the hierarchical error scaling estimation between the fully non-linear inversion in the left column to that of the fix ray path (great circle) inversion on the right for the different resolution checker boards from coarse resolution in (a) and (b) to super fine in (g) and (h). The expected result is a histogram centred about one in this synthetic test and as the resolution increase, the performance of using fixed ray paths diminishes with increasing complexity of features.

minutes and the same inversion with the fast marching forward model took 13 hours. This is nearly 20 times slower, although 13 hours of computer time is not prohibitive. While previous applications of trans-dimensional ambient noise tomography have used an iteratively non-linear approach [Bodin and Sambridge, 2009, Young et al., 2013, Saygin et al., 2016], it has been argued by Galetti et al. [2015] that uncertainty should be estimated using a fully non-linear approach and the difference in uncertainties observed in these synthetic tests generally support this.

In the work of Galetti et al. [2016] on ambient noise tomography of the British Isles it is stated that a 2D inversion takes on the order of a month. The trans-dimensional tree approach with wavelet parameterisation is significantly faster owing to the tight coupling of the fast marching method to the wavelet parameterisation allowing for a full 2D inversion to run on a cluster in a day (the inversions of Galetti et al. [2016] ran for 3 million steps, compared to the inversion of 2 million steps so the comparable estimate for 3 million iterations is approximately 20 hours). Given this increase in speed, there is little reason for not using a fully non-linear approach for this size of problem. The results of the synthetic inversions using fixed great circle paths are reasonable approximations and might be used effectively for initial test inversions and as starting points for fast marching inversions to save time on the convergence of independent Markov chains.

5.8 2D Slice Inversion

After successfully demonstrating the operation of the trans-dimensional tree algorithm with a wavelet parameterisation in synthetic experiments, a next step is to apply it to the real observations obtained in Chapter 2. The configuration of the inversion for the real data is the same as that of the synthetic tests in the previous section. The inversion is started from a uniform model with the velocity set to the mean of the path average velocities of the observations. The standard deviations computed for each frequency from the inversion in Chapter 2 are used as the estimated observational errors and this is coupled with a hierarchical error scale estimated as part of the inversion. The assumption here is that the true errors on the observations are accurate to a constant scaling term which seems reasonable. A hierarchical Laplacian prior is used for the wavelet coefficients with a mode of 0.25. Finally, the trans-dimensional tree resolution is set to a 64 by 32 grid and 1 wavelet super resolution step so the travel times are computed with the fast marching method on a 128 by 64 cell grid. This gives a good balance of accuracy and computational cost.

Figure 5.18 shows the median of the ensemble for selected periods. The results generally show the central rifting region of Iceland to have slower propagation velocities than the east and western coastal regions consistent with previous studies [Gudmundsson et al., 2007, Green et al., 2017]. Another general observation is that the Love wave slow speed anomalies are generally larger in magnitude and more widely distribution than that of the Rayleigh wave results. This may either be caused by the greater sensitivity of Love waves to slow speed anomalies or a larger uncertainty in Love wave dispersion information or some combination thereof.

At periods more sensitive to near surface features, for example Figures 5.18(e) and 5.18(f), a striking slow anomaly in the south east is apparent, corresponding to an area of recent volcanism. Indeed, this area erupted during the HOTSPOT deployment.

The uncertainty for these same periods is shown in Figure 5.19. The standard deviation shows similar patterns to that observed in the synthetic tests. That is, areas along the spreading ridge have higher uncertainties. Again this is most likely a limitation of the geometry of the deployment dictated by the logistics of the environment. This could be improved by gaining access to more recently collected data from the study of Green et al. [2017].

In Figure 5.20 the results of this study are compared with two published studies that have produced Rayleigh wave group velocity maps. In (a), the work of Gudmundsson et al. [2007] is shown where the authors used exactly the same source data as this study, (b) is the median of the ensemble of this study, and (c) is the more recent work of Green et al. [2017] which had observations generated from 241 seismometers or roughly six times the number used in this study. A direct comparison is not possible between these studies as they invert a different property. That is, the results of other studies are group velocity maps whereas this study produces phase velocity maps. Nonetheless they should have structural similarities, and indeed comparison suggests they do.

In the inversion of Green et al. [2017], smoothing regularisation and fixed ray paths are used to invert for the group velocity map resulting in the smoothness that is apparent in the figure. Using less data, roughly one sixth of the number of stations, this study has recovered similar structure and in parts it would appear to have more detailed structure. Inferences based upon the models recovered in this study can be supported with uncertainty, where as Green et al. [2017] provide no such justification for their results.

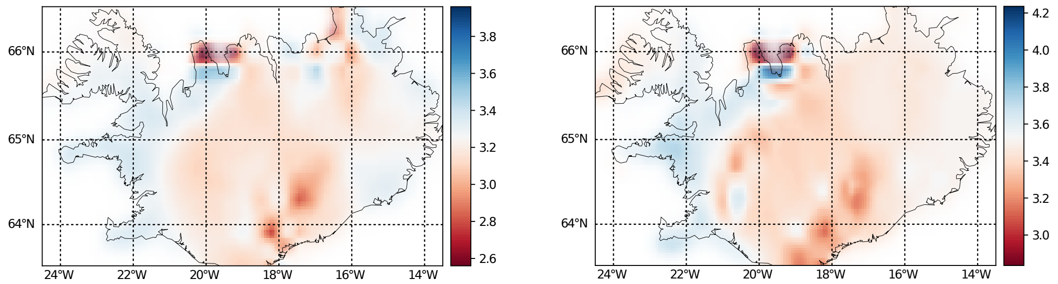
5.9 3D Inversion

By inverting each frequency or period independently, the correlation between neighbouring frequencies is effectively ignored. This expected correlation between neighbouring frequencies stems from the assumptions from Chapter 2 that dispersion is a continuous function.

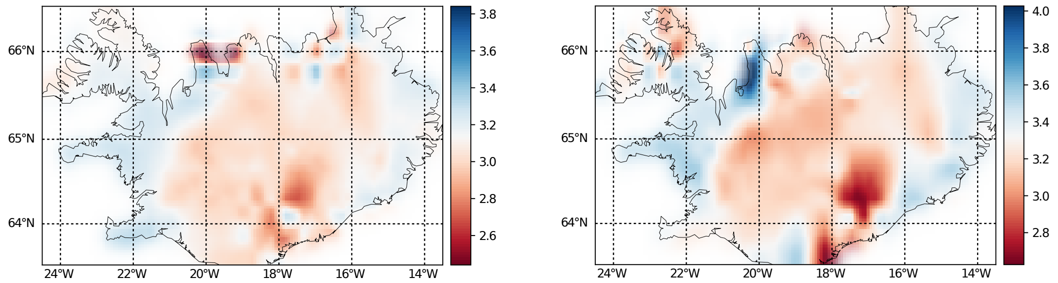
The trans-dimensional tree approach has a great deal of flexibility in both the basis functions that are used and the dimension of the problem. The trans-dimensional tree approach is an abstraction that can be applied to many problems. Here an obvious solution is to invert a set of neighbouring frequencies simultaneously using a 3D wavelet parameterisation.

The details of the inversion change very little. The hierarchical prior, a hierarchical error scale (which applies to all periods) and parallel tempering are all still used. For the inversion, 32 frequencies spaced between 0.034 Hz and 0.296 Hz (or 29.2 s and 3.4 s periods) are inverted jointly, with the starting model set to the mean of the path average phase velocities at each frequency. Recall that the phase velocity observations are continuous functions meaning it is possible to invert any discretised band of fre-

(a) Rayleigh 8.4s (b) Love 8.4s



(c) Rayleigh 6.2s (d) Love 6.2s



(e) Rayleigh 4.1s (f) Love 4.1s

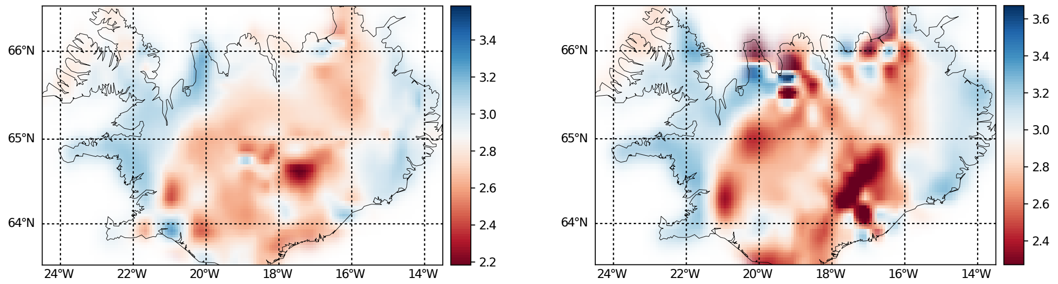
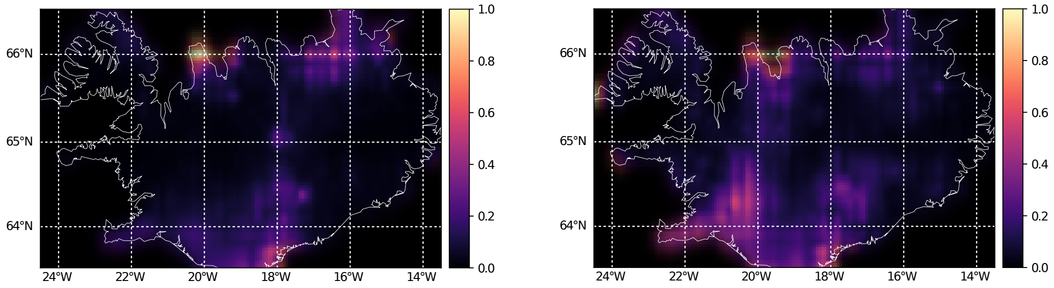
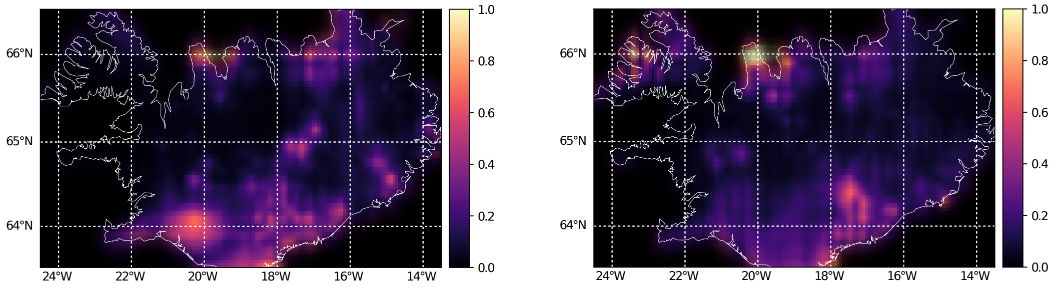


Figure 5.18: Median of the ensemble images for Rayleigh (left column) and Love wave (right column) inversions at selected periods for Iceland.

(a) Rayleigh 8.4s (b) Love 8.4s



(c) Rayleigh 6.2s (d) Love 6.2s



(c) Rayleigh 4.1s (d) Love 4.1s

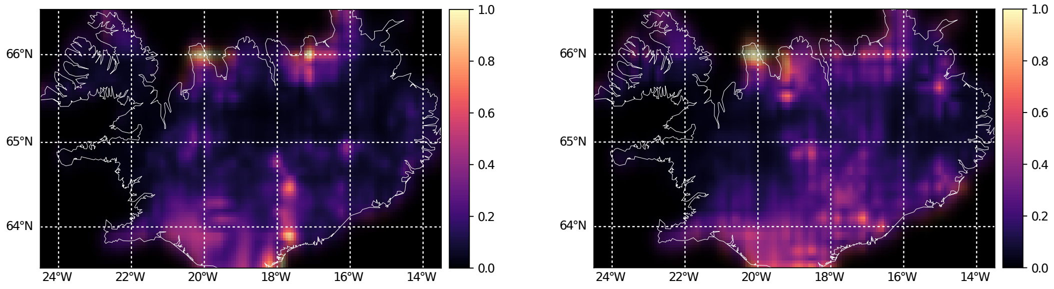
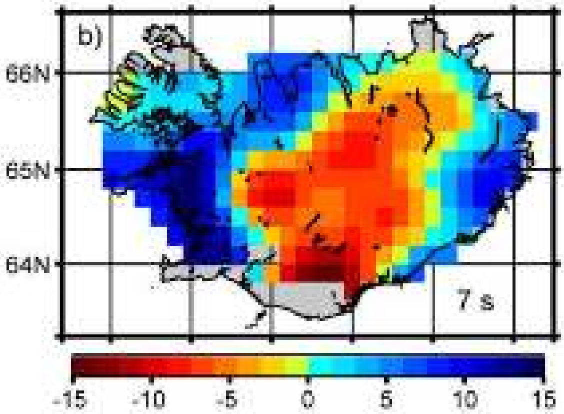
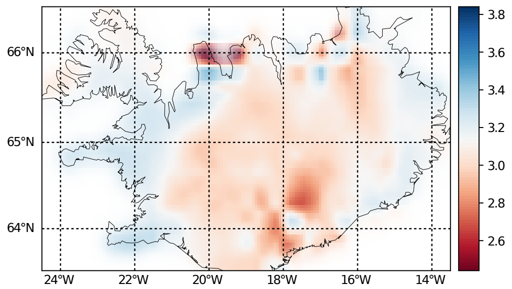


Figure 5.19: The ensemble standard deviation of inversions of ambient noise data of Iceland. The left column is for Rayleigh wave observations and the right for Love wave.

(a) 7 second period Rayleigh wave group velocity



(b) 6 second period Rayleigh wave group velocity



(c) 6.2 second period Rayleigh wave phase velocity

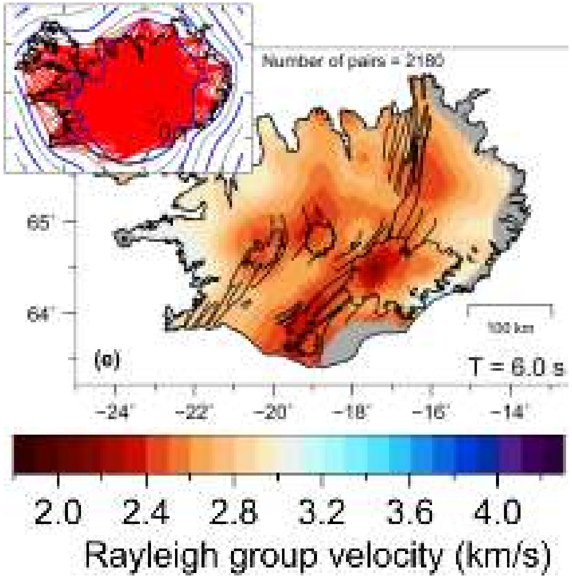


Figure 5.20: A comparison of results from Gudmundsson et al. [2007] in (a), this study in (b), and Green et al. [2017] in (c). The images in (a) and (c) are Rayleigh wave group velocity and (b) is Rayleigh wave phase velocity.

quencies. The choice here is a central band of frequencies that envelop the dominant frequencies excited by ambient noise sources.

With 32 frequencies, the trans-dimensional tree wavelet model becomes a 64 by 32 by 32 volumetric image of longitude by latitude by frequency bin. A single wavelet super resolution step is used for each frequency to improve the accuracy of the fast marching method. Since the variation of phase velocity with frequency is smooth, it is appropriate to use the CDF 9/7 wavelet for both the spatial dimension and that of frequency. However, it is possible to use different wavelet bases in different dimensions where warranted.

The travel times of 435 station pairs across 32 frequencies, or nearly 14,000 ray paths, need to be computed at each iteration. This increases the computational burden substantially and the computational time for 2 million iterations increases from approximately 14 hours for a single frequency inversion to approximately 160 hours (7 days) for a 3D inversion of the same lateral dimension.

The ensemble medians for both the Love and Rayleigh 3D inversions are shown in Figure 5.21. These results are for the same periods as the single frequency inversion described in the previous section, and show similar features to those of the previous 2D inversion where phase velocity maps were inverted independently. In the 3D inversion, some of the anomalous features seen in the 2D inversions are muted. In Figure 5.22, the standard deviation of the ensemble is shown and it has much the same character as before.

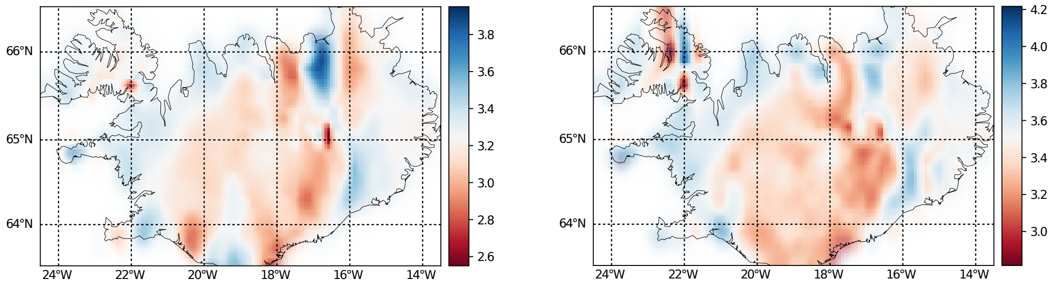
The correlated information between neighbouring frequencies will tend to smooth anomalies across the frequency dimension. This is analogous in a 1D regression problem to the difference between fitting a piece wise linear curve to a set of data points with noise and fitting a single linear function to the data. The single linear function will be smoother but if the underlying true function is linear then a single best fit linear function will be a more faithful representation than a piece wise linear fit. In the case here, the independent inversion of frequencies may over fit the velocity anomalies as addition information from neighbouring frequencies hasn't been made available. This extra information helps stabilise the inversion and allows resolvability of deeper Earth structure.

The median and standard deviations of longer periods are shown in Figures 5.23 and 5.24. For the longest period of nearly 30 seconds, features are resolved in the Rayleigh wave inversion, and while similar features appear in the Love wave inversion, there are some inversion artefacts and high uncertainty suggesting limited resolvability. Nonetheless inversion is completed down to a much longer period than has been completed by either of the previous published ambient noise studies of the region [Gudmundsson et al., 2007, Green et al., 2017].

The convergence metrics for the 3D inversion of both the Rayleigh and Love wave inversions are shown in Figures 5.26 and 5.27 respectively. The hierarchical error scale is well converged with the number of wavelet coefficients showing more disparity. Unlike the airborne electromagnetic inversion in Chapter 4, evidence of larger parallel tempering steps is apparent in the convergence plots (large vertical jumps).

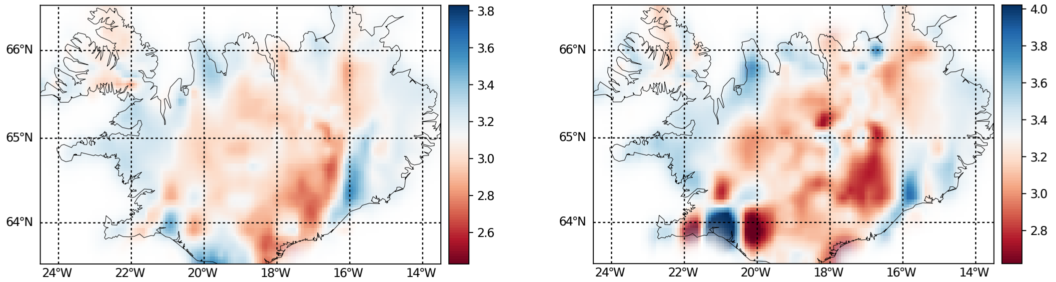
Inverting a band of frequencies at the same time increases the number of parameters and hence the total number of iterations it takes to converge. However, overall the 3D inversion uses significantly fewer parameters for the inversion compared to the 2D when considering the number of frequencies inverted. For example, the 3D inversion uses between 500 and 800 coefficients with the modal value around 600. On average the 2D inversions run in the previous section use of the order of 50 to 200 coefficients with a modal number of coefficients of approximately 100. If this modal number is multiplied by the number of frequencies in the 3D inversion, this would equate to around 3,200 coefficients (or between 1,600 and 6,400) compared to 500 to 800. So although there are inherent difficulties in the joint inversion of a band of frequencies, there are benefits in terms of robustness and ability to resolved features deeper into the Earth.

The results at approximately the 6 second period can again be compared to previous studies and these are shown in Figure 5.25 where, (a) and (b) are the group velocity maps of previous studies, (c) shows the ensemble median of the previous 2D inversion



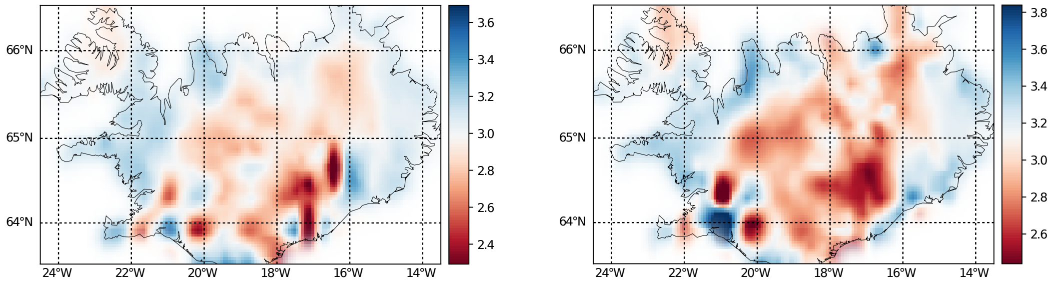
(a) Rayleigh 8.4s

(b) Love 8.4s



(c) Rayleigh 6.2s

(d) Love 6.2s



(e) Rayleigh 4.1s

(f) Love 4.1s

Figure 5.21: The ensemble medians for the 3D inversion of Rayleigh wave (left column) and Love wave (right column) for selected periods.

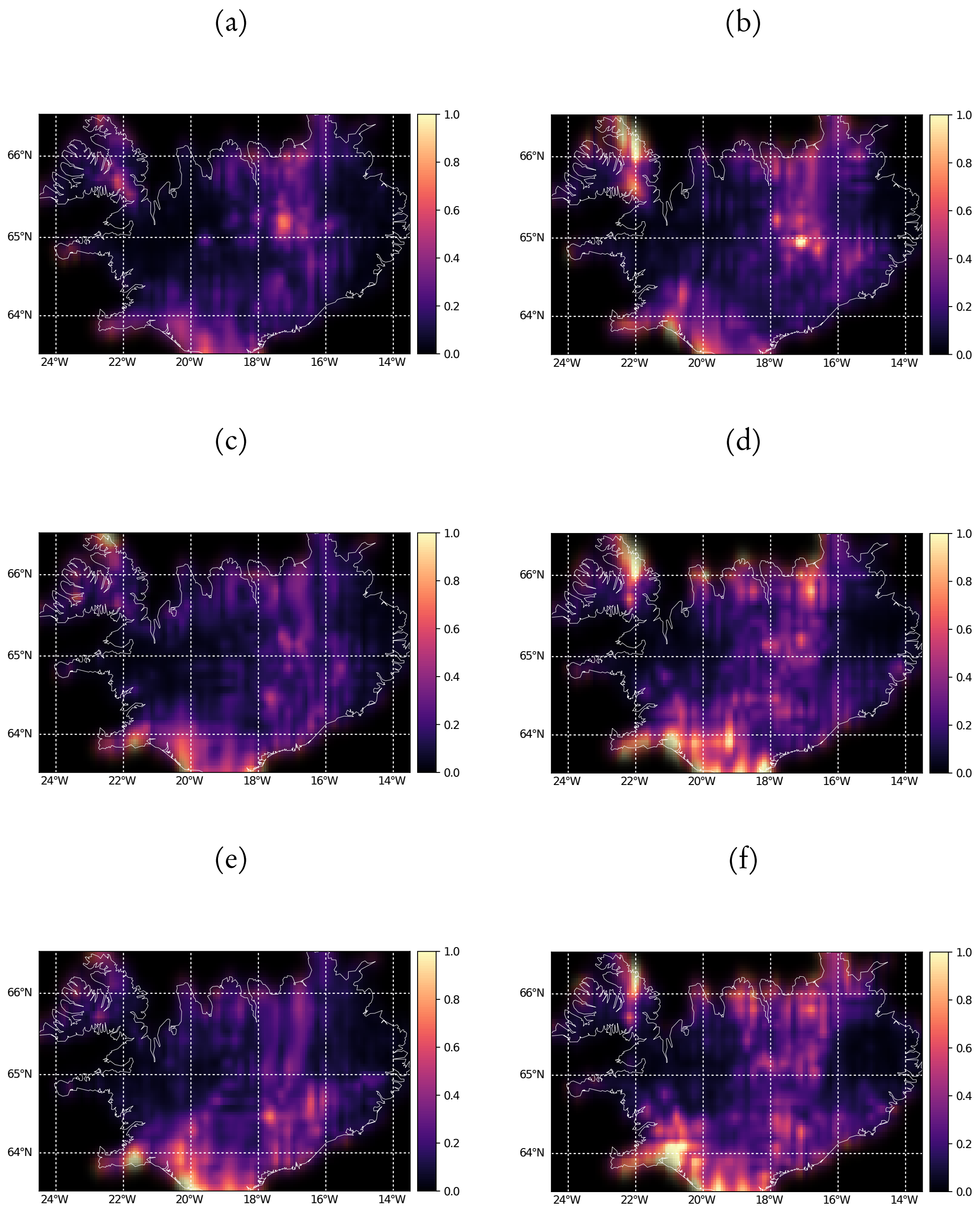
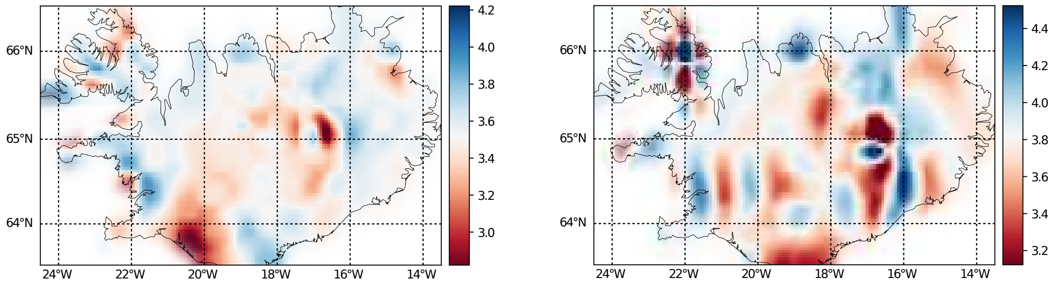
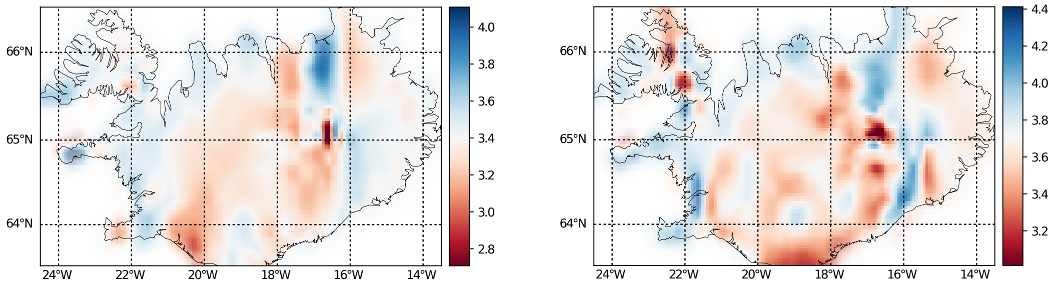


Figure 5.22: The ensemble standard deviations for the 3D inversion of Rayleigh wave (left column) and Love wave (right column) for selected periods.



(a) Rayleigh 29.2s

(b) Love 29.2s

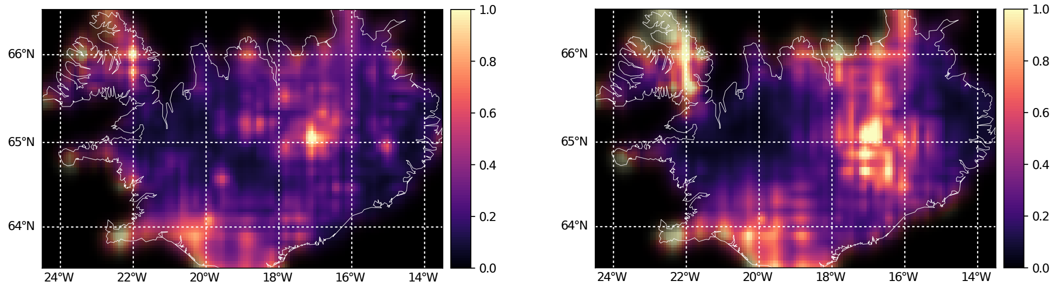


(c) Rayleigh 13.1s

(d) Love 13.1s

Figure 5.23: The ensemble medians for the 3D inversion of Rayleigh wave (left column) and Love wave (right column) for longer periods.

(a) Rayleigh 29.2s (b) Love 29.2s



(c) Rayleigh 13.1s (d) Love 13.1s

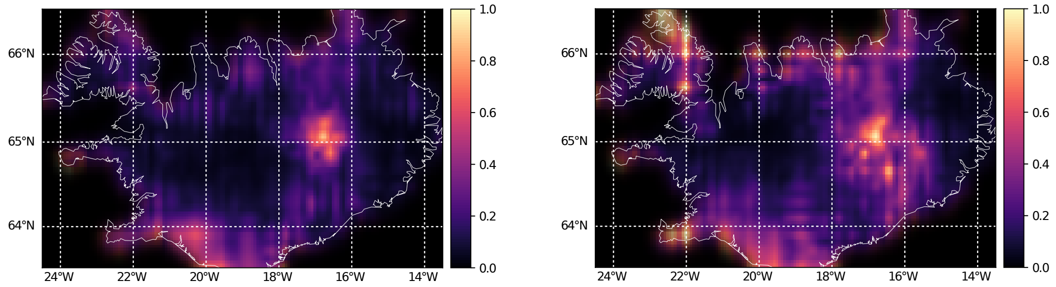
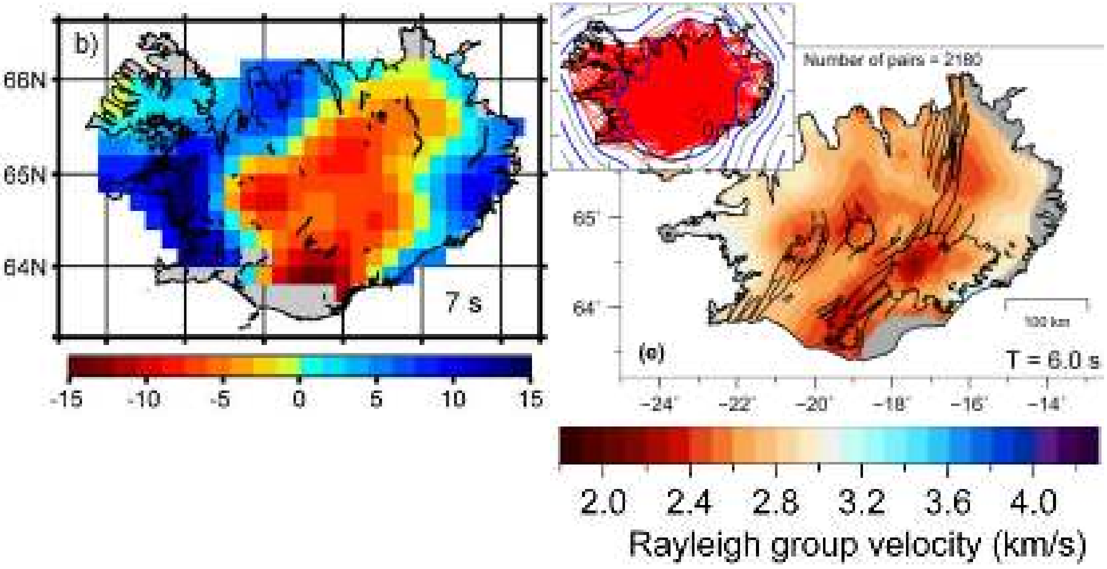


Figure 5.24: The ensemble standard deviations for the 3D inversion of Rayleigh wave (left column) and Love wave (right column) for longer periods.

(a) 7 seconds period group velocity map (b) 6 seconds period group velocity map



(c) 6.2 second period phase velocity map (d) 6.2 second period phase velocity map

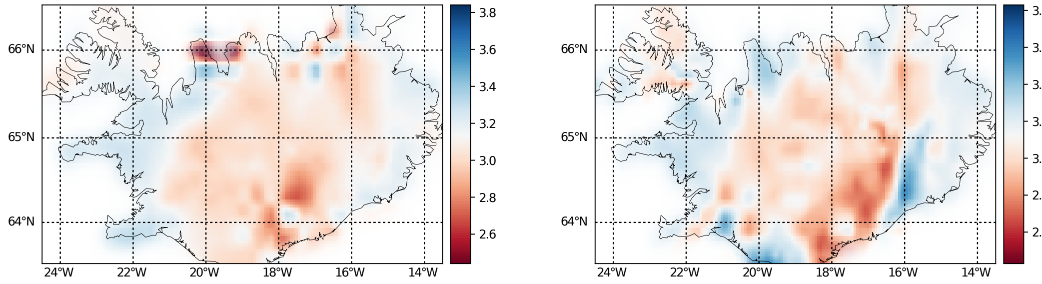


Figure 5.25: A comparison of results from Gudmundsson et al. [2007] in (a), Green et al. [2017] in (b), and from this study, the ensemble median of the 2D inversion in (c) and the ensemble median of the 3D inversion in (d). The images in (a) and (b) are Rayleigh wave group velocity and (c) and (d) are Rayleigh wave phase velocity.

and (d) the ensemble median of the 3D inversion. In first comparing the 2D inversion to the 3D, the structure has be refined considerably. Particular features include the much stronger contrast between the volcanic zone (red) to old crust (blue) in the south-east between 16 and 17 degrees longitude west. Some of the anomalies in the north of the 2D inversion have been replaced with simpler structure in the 3D inversion. In the 3D inversion, there are two distinct bands of slow velocity zones that agree with structure proposed by Green et al. [2017] in their study, which has significantly more data in central Iceland compared with this study. Generally it is found in the 3D inversion that the resolving power of the observations is improved by jointly inverting across a range of frequencies.

In this inversion of phase velocity maps, the Love and Rayleigh wave inversions have been run separately so the relationship between Love and Rayleigh phase velocity ratios from Chapter 2 is not included. Similarly the monotonicity of the dispersion curve is not enforced in the wavelet parameterisation. Incorporating both these factors and jointly inverting for Love and Rayleigh phase velocity maps has the potential to improve results further.

Nonetheless, by looking at 1D profiles of dispersion, these assumptions can be tested, as shown in Figure 5.28. In this figure it is apparent that for Love and Rayleigh wave results, the monotonicity assumptions are violated at low frequencies. Additionally, the numerical studies showing Love wave velocities slightly higher than Rayleigh wave velocities is also violated at higher frequencies. When accounting for errors, shown in shaded region, both violations could be accounted for in range of uncertainties. In future, including joint inversion of Love and Rayleigh wave observations may improve the results and provide greater constraint.

5.10 Shear wave 1D Inversion

In the previous section, phase velocity maps for Love and Rayleigh wave dispersion were inverted from a suite of frequencies. The common next step is to invert dispersion

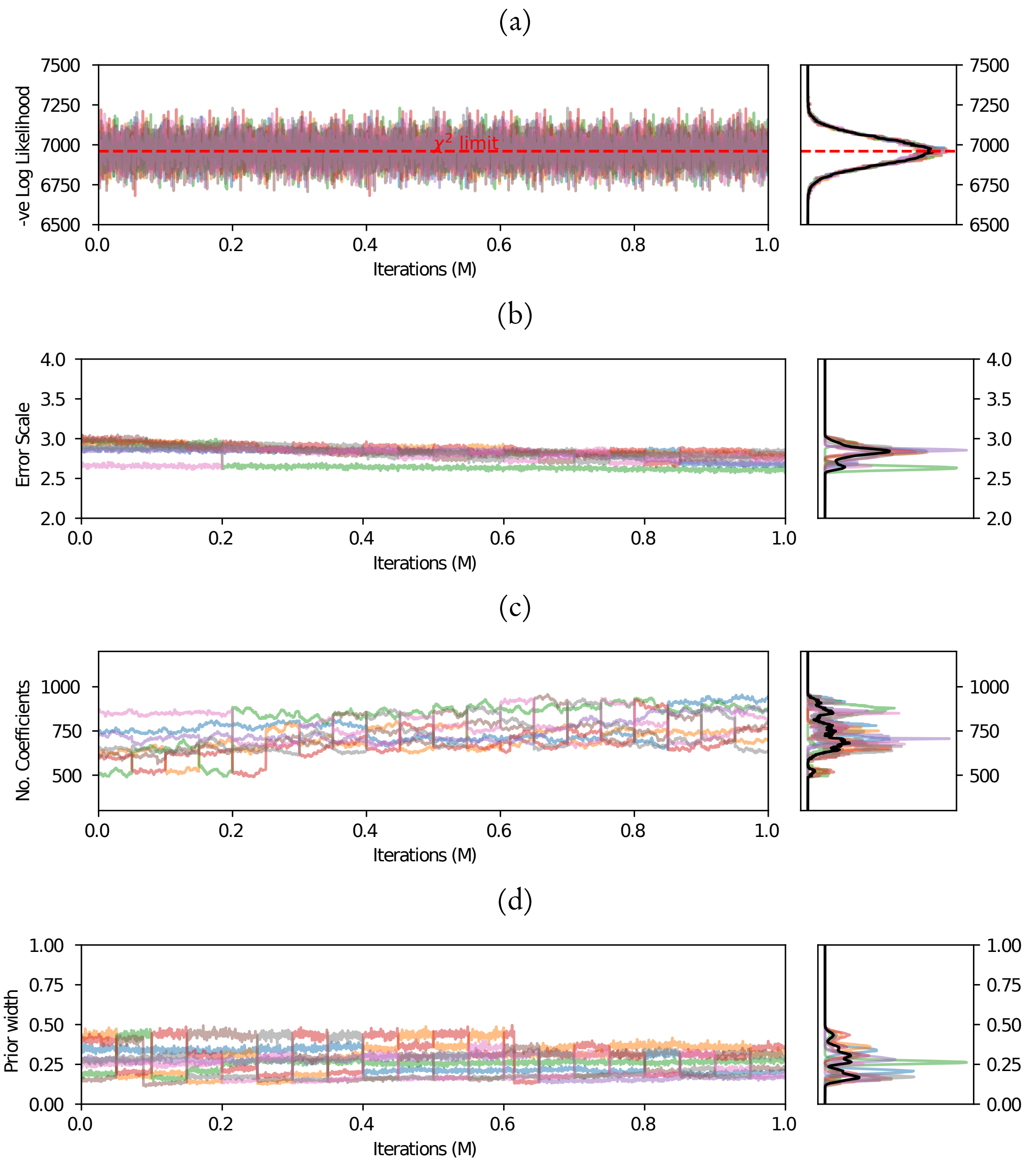


Figure 5.26: Convergence properties of Rayleigh wave 3D Inversion. Plotted here is the negative log likelihood, hierarchical error scaling parameter, number of coefficients and the hierarchical prior width as a function of iteration. Each chain is plotted in a different colour and the black histograms on the right represent the average histogram of all curves.

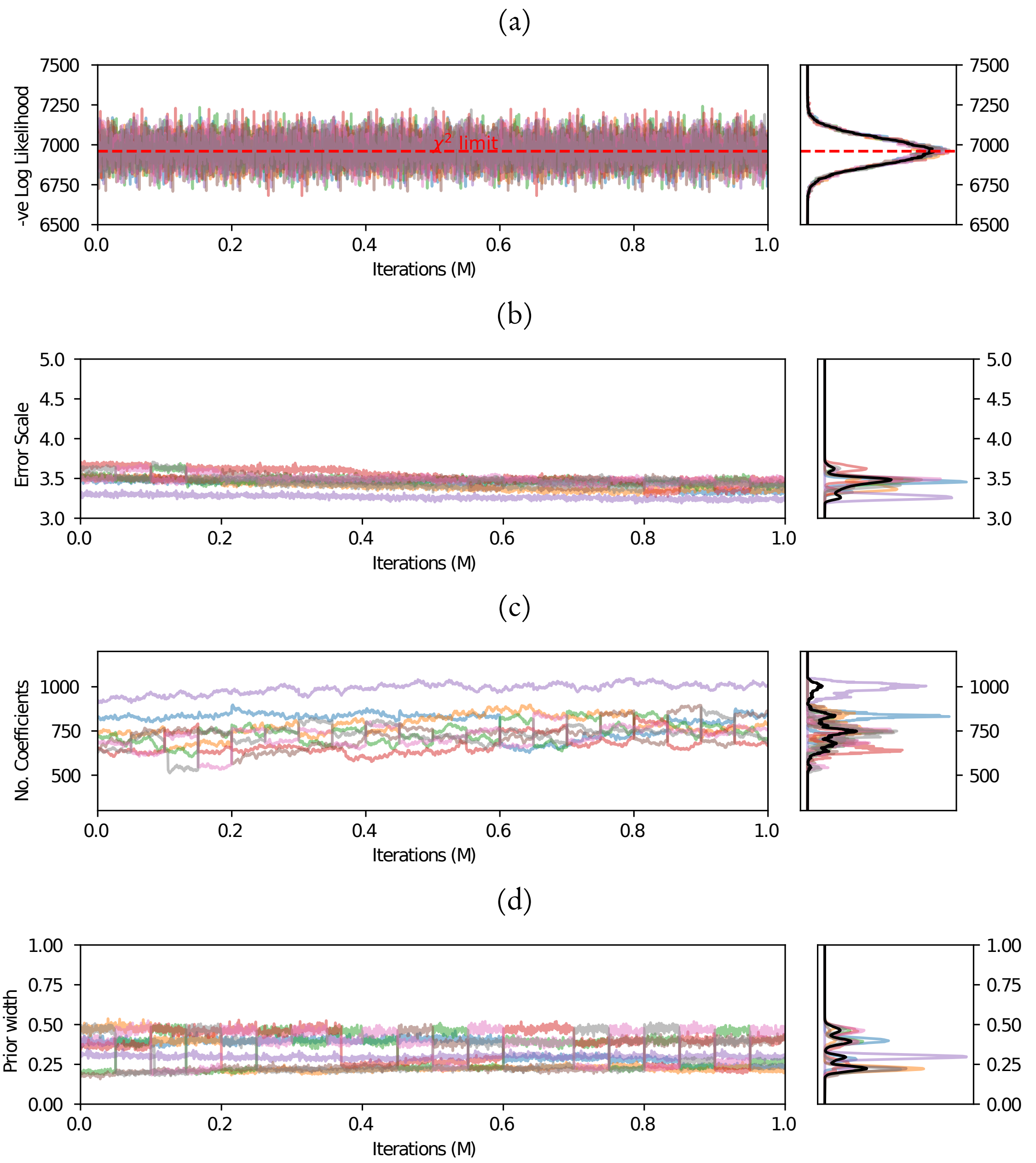


Figure 5.27: Convergence properties of Love wave 3D Inversion. Plotted here is the negative log likelihood, hierarchical error scaling parameter, number of coefficients and the hierarchical prior width as a function of iteration. Each chain is plotted in a different colour and the black histograms on the right represent the average histogram of all curves.

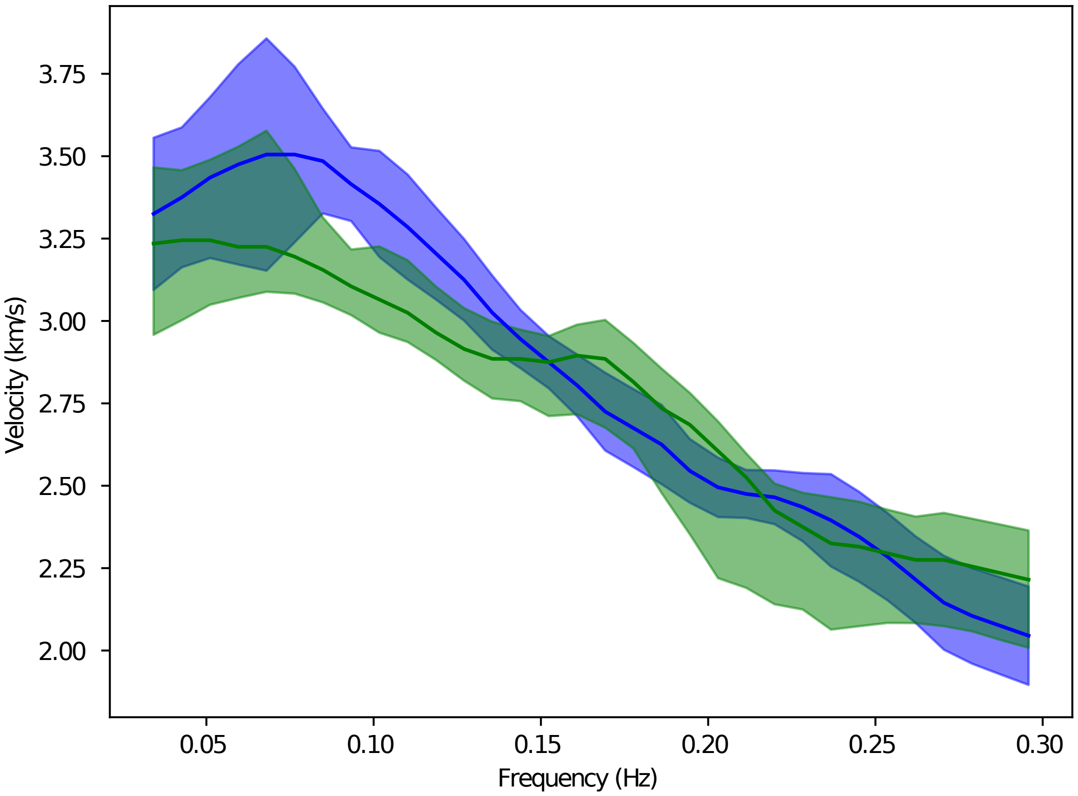


Figure 5.28: Dispersion curves from 3D inversion of phase velocity shown that the relationship between Love (blue) and Rayleigh (green) dispersion is preserved during the inversion to within errors. Shaded regions represent one standard deviation.

curves for local shear wave velocity structure. The approach used in this study is outlined in Bodin et al. [2012b] where a trans-dimensional partition model is used to represent a shear velocity model to a half space. The forward model used is the `DISPERS80` routine of Takeuchi and Saito [1972] and an elastic Earth is assumed to reduce the parameter space. Love waves are sensitive only to the shear wave velocity, however, Rayleigh waves are also sensitive to P-waves and an empirical relationship is used to obtain density [Deschamps et al., 2001] and P-wave velocity [Ritsema and van Heijst, 2002] from a shear wave velocity in each partition of the model.

A shear wave model is jointly inverted using results from the Love and Rayleigh 3D inversion of the previous section and included uncertainty estimates from the posterior. Independent hierarchical scaling parameters are included for the Love and Rayleigh wave observations to allow the data to decide the relative weight rather than enforcing this. This results in a likelihood function of the form

$$\begin{aligned} p(\mathbf{d}|\mathbf{m}) \propto \exp \Bigg\{ & -\sum_i \frac{(G_{\text{Love}}(\mathbf{m}, f_i) - d_{\text{Love}}(f_i))^2}{2(\lambda_{\text{Love}} \sigma_{\text{Love}(f_i)})^2} \\ & -\sum_i \frac{(G_{\text{Rayleigh}}(\mathbf{m}, f_i) - d_{\text{Rayleigh}}(f_i))^2}{2(\lambda_{\text{Rayleigh}} \sigma_{\text{Rayleigh}(f_i)})^2} \Bigg\}, \end{aligned} \tag{5.3}$$

where the forward model $G_{\text{Love}}(\mathbf{m}, f_i)$ computes the phase velocity for the given model $\mathbf{m}$ at the frequency f_i, and $d_{\text{Love}}(f_i)$ is the corresponding observation at the given frequency. $\sigma_{\text{Love}(f_i)}$ is the standard deviation of the observation from the 3D inversion of the previous section and λ_{Love} is the independent hierarchical scaling parameter.

For each spatial grid cell of the 3D inversion domain in the previous section, 1D trans-dimensional joint inversion using four independent chains is performed to obtain a 3D shear wave velocity structure. The prior for the shear wave velocity was set to uniform between 2 and 6 km/s and for the number of layers the prior was set to uniform between 1 and 12 layers. The model domain is fixed to a depth of 100 km to the half space. Several different test inversion were attempted to varying depths to verify

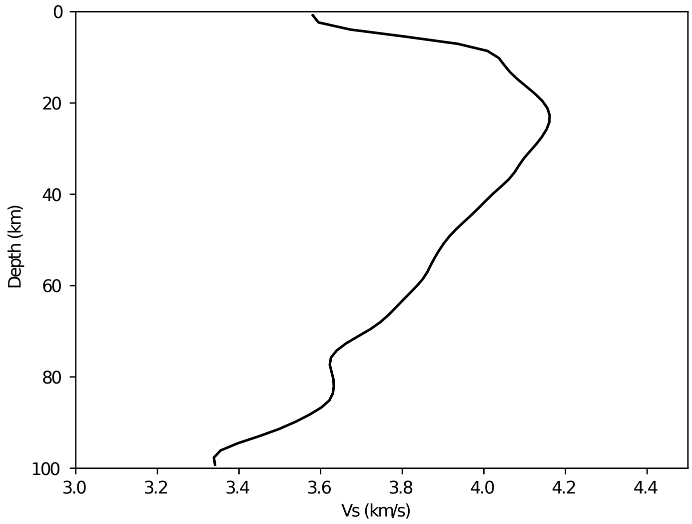


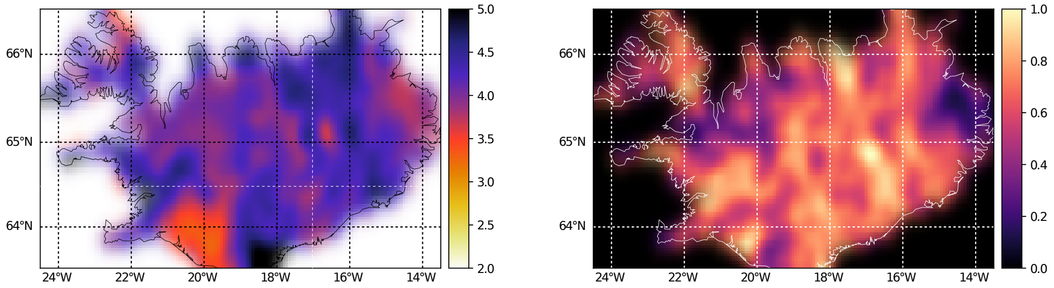
Figure 5.29: The mean shear wave velocity profile of all 1D inversions.

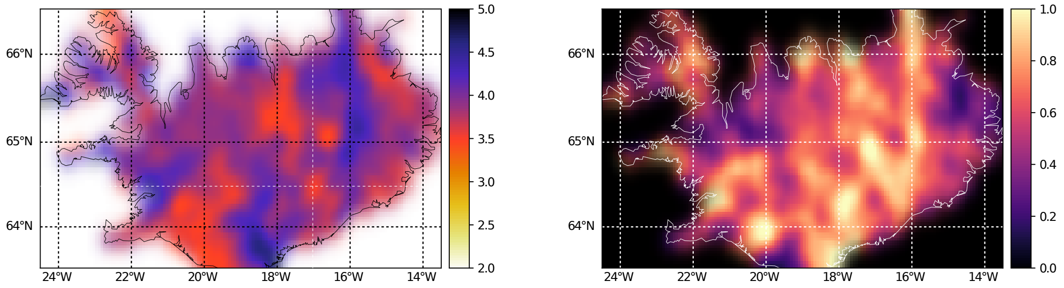
that setting a too shallow domain wasn't artificially constraining the results. Previous similar inversions of the area only used Rayleigh wave group velocity and therefore have lower depth sensitivity than jointly inverting Love and Rayleigh phase velocity.

The mean of all the 1D inversions is plotted in Figure 5.29. In this figure, an interesting feature of this profile is the high velocity layers around 20 to 40 km depth overriding lower velocities. This average profile is consistent with oceanic crust [Knopoff, 1972, see Figure 13] and consistent with previous earthquake driven surface wave studies of Iceland [Li and Detrick, 2006, see Figure 5].

Maps of the inverted shear wave model at various depths are plotted in Figure 5.30. Shear wave velocity inversion from dispersion information is a very ill-determined problem, in that many shear wave models can give rise to the same or similar dispersion curve. Hence, there is a high degree of uncertainty in this inversion as evidenced by the large standard deviation across much of the inversion domain. A consistent feature of the inverted shear wave model is the large slow velocity anomaly slightly to the south-east of central Iceland. This is located near a zone of active volcanism and above the presumed location of the Icelandic Hot spot.

Longitude and latitude transects can show additional structure and these are plotted in


(a) 25 km (b) 25 km


(c) 50 km (d) 50 km

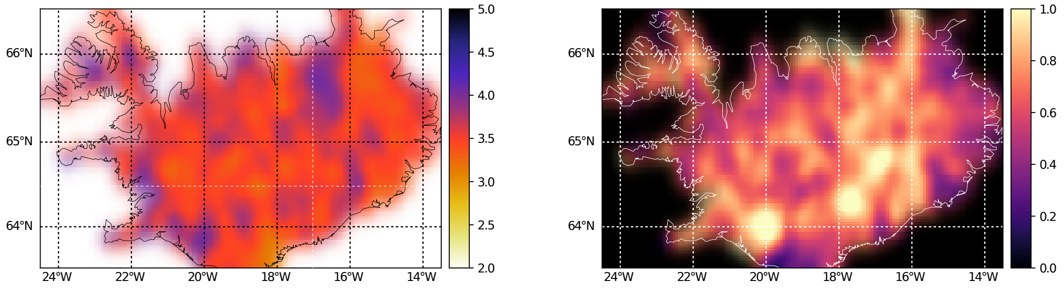

(e) 75 km (f) 75 km

Figure 5.30: Ensemble means (left column) of shear wave velocity and their standard deviations in km/s (right column) of the surface wave dispersion inversion at various depths.

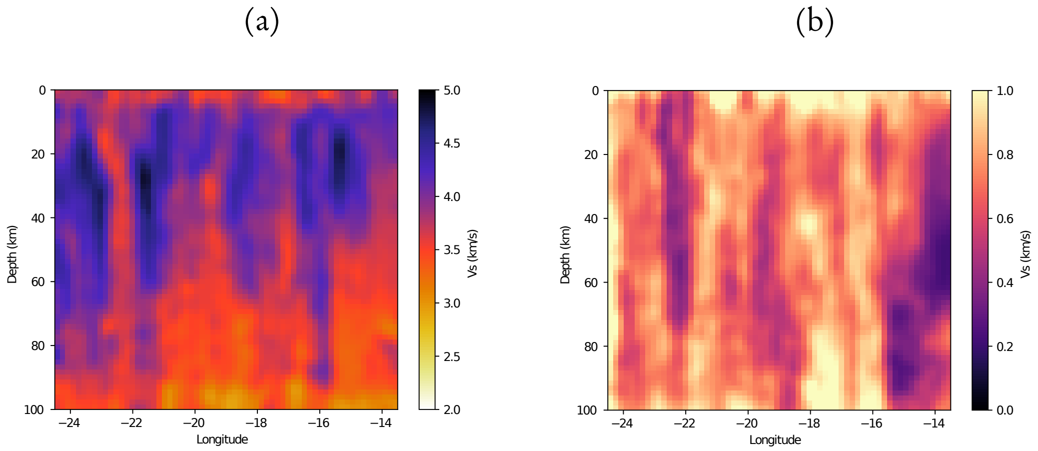

(a) (b)

Figure 5.31: Longitudinal transect through Iceland showing depth profile of shear wave velocity (a) and standard deviation (b) in km/s. The transect is taken along the horizontal dashed line in Figure 5.30

Figures 5.31 and 5.32. The general structures show good agreement with the long period earthquake driven Rayleigh wave tomography inversion of Li and Detrick [2006], namely a crust of around 3.25 km/s increasing to 4 - 4.25 km/s up to 50 km deep, then dropping back to less than 4 km/s below this. These transects are chosen so their intersection point is approximately located in Vatnajökull region of Iceland, an area of active volcanism. In Figure 5.32(a) a slow velocity tube rising from a large slow velocity region at depth is visible directly beneath the area of volcanism. The location of the slow velocity region at depth is consistent with the presumed location of the Icelandic hot spot [Thordarson and Höskuldsson, 2008]. It is tempting to claim that these results show the plumbing between the Iceland hot spot and volcanism at the surface. The strength of this claim is weakened by the large uncertainty in these results and the appearance of many of these structures, some of which correspond to other volcanic regions, for example the other tube feature in Figure 5.32(a) between 65.0 and 65.5 latitude corresponds to a cluster of volcanoes to the north of the Vatnajökull region volcanic provinces. If the transect is plotted as a relative deviation from the mean model as in Figure 5.33, the slow velocity anomaly above the location of the hot spot linked to the active Bárdarbunga and Grímsvötn volcanoes is more apparent. In the plots, this vertical tube structure is indicated with a dashed line and is a consistent slow velocity anomaly continuous from the deepest parts of the model to the surface.

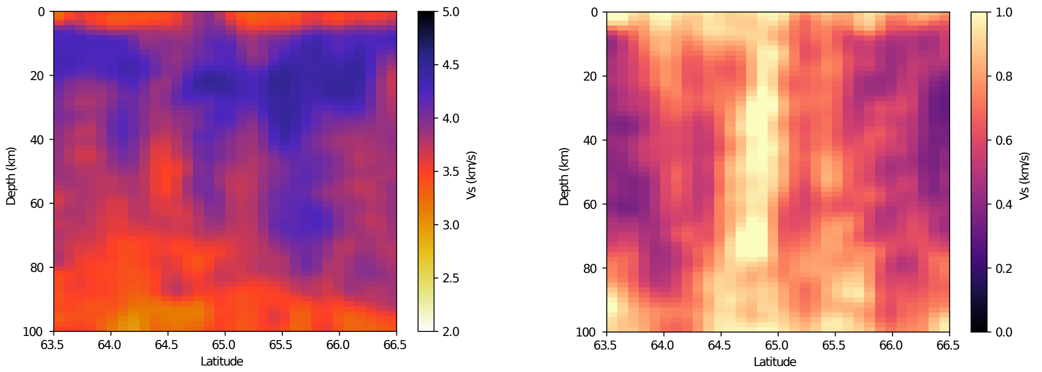

(a) (b)

Figure 5.32: Latitudinal transect through Iceland showing depth profile of shear wave velocity (a) and standard deviation (b) in km/s. The transect is taken along the vertical dashed line in Figure 5.30

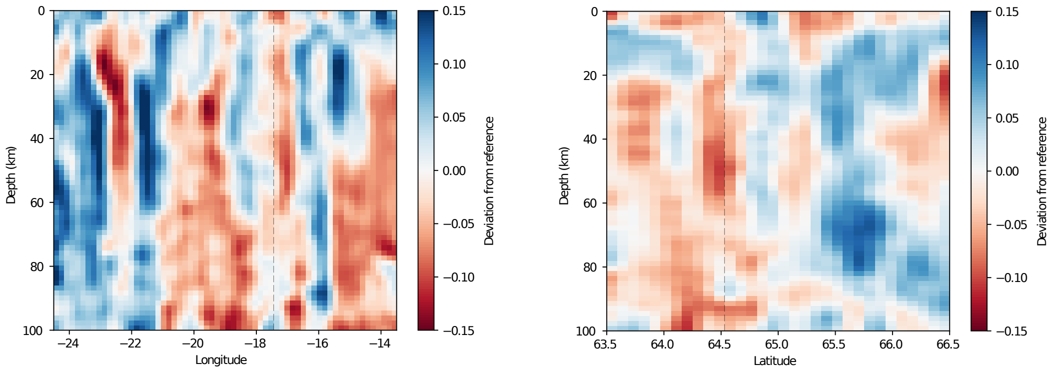

(a) (b)

Figure 5.33: Longitude (a) and Latitude (b) transects plotted as a relative deviation from the ensemble mean velocity, highlighting slow zones. At the dashed line is approximately the location of the active Bárdarbunga and Grímsv otn volcanoes.

The shear wave model inversion and uncertainty could potentially be improved by taking into account spatial correlation between 1D inversions using a technique similar to that employed in the Airborne Electromagnetic inversion from Chapter 4. This would mean that rather than inverting each 1D profile independently, a transect could be inverted as an image using the trans-dimensional tree approach (or the entire volume depending on computational resources). This is an area for potential future research.

5.11 Summary

In this chapter, a novel fully non-linear trans-dimensional inversion of ambient noise data has been developed. Its effectiveness was first demonstrated in the inversion of synthetically generated checker board tests and the results were compared to a linearised inversion where fixed great circle paths were used. This showed that short length scale features tend to be under-estimated in a fixed ray path inversion.

The new method was then applied to produce phase velocity maps for Iceland for both Love and Rayleigh waves. This is the first inversion of phase velocity maps from ambient noise observations in the region and the first ambient noise study to have used a fully Bayesian inversion, where errors from the path average phase velocity estimates in Chapter 2 are propagated through to the phase velocity map inversions.

Rather than inverting individual periods or frequencies in 2D inversions, a method for inverting a band of frequencies jointly to take advantage of the correlation between neighbouring frequencies was demonstrated. This further stabilises the inversion of phase velocity maps, improving the resolving power of the observations and allowing greater resolution deeper into the Earth.

From the phase velocity maps, a joint inversion of Love and Rayleigh wave dispersion was performed to construct a 3D shear wave velocity model down to 100km depth for the Iceland region. This model shows tantalising glimpses of what look like volcanic conduits, however the large uncertainties in this inversion prompt caution. Methods that would allow better constraint of uncertainties during this stage of the inversion would be a fruitful avenue of further research.

Chapter 6

Global surface wave tomography

6.1 Introduction

This chapter seeks to apply the trans-dimensional tree algorithm to a problem at the global scale. Many problems of seismology are naturally at the global scale, either 2D problems on the surface of a sphere, 3D spherical shells (for example, crustal or Mantle models) or full 3D spheres. Previously trans-dimensional approaches have used Voronoi cells for inner core tomography [Young et al., 2013], but there has been limited application of trans-dimensional techniques to such global scale problems. The reason for this paucity of trans-dimensional inversions on the sphere is that global data sets are often large and this coupled with traditional trans-dimensional approaches relying on a computationally inefficient Voronoi cell parameterisation render many problems intractable.

To give an appreciation of recent applications of Voronoi cell parameterisation applied to global scale problems, the work of Young et al. [2013] used approximately one month of compute time to invert 1,871 differential travel time observations. The key reason behind this large computational cost is that Cartesian Voronoi cells were used, that is the surface of the sphere is mapped to the Cartesian plane. This required extra processing to impose continuity across longitudinal boundaries of the Cartesian region.

In re-examining the application of Voronoi cell approaches to global scale tomographic problems, a new spherical Voronoi cell code has been developed as part of this thesis and subsequently applied by others(for example Pejic et al. [2017]). This new spherical Voronoi method treats the problem correctly with proposal distributions that respect spherical geometry. By taking this approach, polar regions are treated without introducing distortions and model continuity is maintained throughout the domain. The spherical Voronoi method has the added benefit of increasing the efficiency of inversions.

One of the yet resolved questions of the inner core is the apparent East-West hemisphericity of travel time anomalies. One way to investigate this is using differential travel times of phases that sense the outer most layer of the inner core. An example of this is the large database of PKIKP-PKiKP differential travel times collected in the study of Waszek and Deuss [2011]. These data consist of P-wave arrivals where one has arrival has reflected off the outer-core/inner-core boundary (PKIKP) and the other has refracted through the upper most layers of the inner-core. It is assumed that the difference in travel times of these two phases has dominant sensitivity to the P-wave velocity of the upper most inner-core as the path of these phases through the crust, mantle and outer-core are similar and therefore cancel out.

In this study a trial inversion of these observations has been performed using trans-dimensional spherical Voronoi cells. The model is parameterised in terms of Voronoi cells with value representing deviations from a reference model, in this case AK135 [Kennett et al., 1995]. The rays are fixed, which is a reasonable approximation as the observations represent short grazing paths of the outer most inner core. Twelve independent chains are run for two million iterations with the first one million removed as burnin. The mean of the ensemble is shown in Figure 6.1 where the globe is shown from both hemispheres to highlight the distinct hemispherical boundary through the Pacific and the less distinct boundary through Europe and Africa.

The data set has 5,477 paths, being slightly more than double the number used in the study of Young et al. [2013]. The inversion using a Voronoi cell approach took approximately a week to compute on a cluster computer. Although this global data set is relatively small and the ray paths for the tomographic problem are relatively short, the inversion is computationally expensive for what is a relatively simple model. Apparent in the mean image as well are Voronoi cell artefacts. These artefacts would likely be remedied by averaging over several more chains. However, this amounts to using even more computational power to mask some of the short comings of Voronoi cell parameterisations.

For this particular problem, where interrogation of the location and strength of inner-core boundaries is undertaken with relatively simple models, the Voronoi cell parameterisation may be well suited. However, for larger scale problems where the number of observations is larger and the scale length of resolvable structure is smaller, that is,

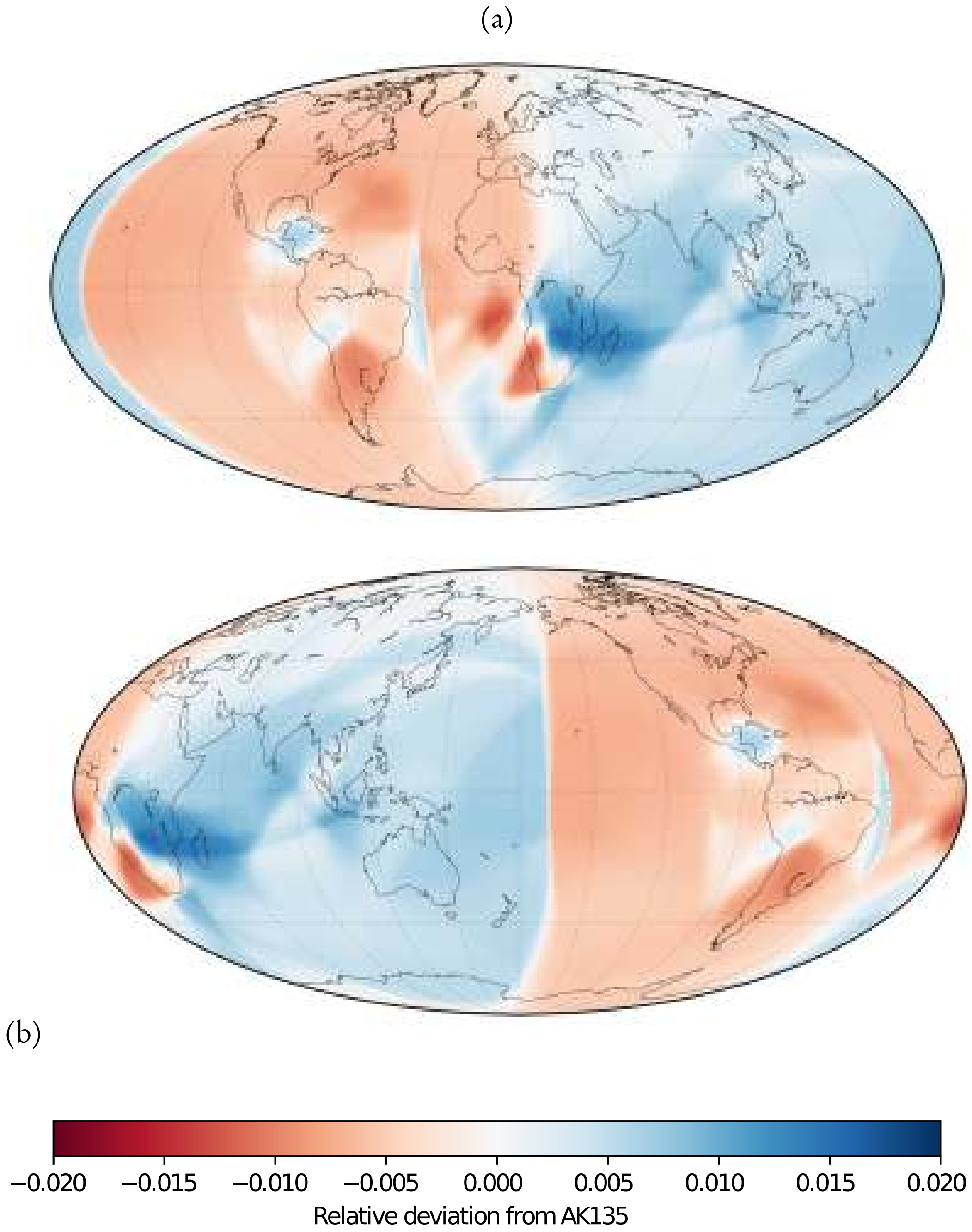


Figure 6.1: The ensemble mean image from inverting PKIKP-PKiKP differential travel times as a deviation from the AK135 reference model. The hemispherical nature of the travel time anomalies is clearly apparent.

greater model complexity, the Voronoi cell parameterisation seems impractical.

6.2 Global surface waves

An example of large global scale data sets are earthquake driven surface wave observations. Global surface wave observations consist of a number of travel times or estimates of phase velocities between earthquake sources and receiving stations or stations approximately aligned in a surface wave propagating direction [Bloch and Hales, 1968] over a broad range of frequencies or modes.

Studies using global surface wave observations have constructed phase velocity maps for from fundamental model Love and Rayleigh waves using earthquake driven sources [Montagner, 1986, Trampert and Woodhouse, 1995, Shapiro and Ritzwoller, 2002, Shaeffer and Lebedev, 2013]. These global surface wave data sets include an order of magnitude greater number of observations and require inversion at a large number of periods rather than a single inversion as required by the differential travel time tomography example in the previous section. Ongoing collection of global surface wave data, including higher order mode phase velocities has only added to the size of this large scale data problem [Debayle and Sambridge, 2004, Debayle and Ricard, 2012].

The common approach to these inversions is to use a least squares inversion using a truncated spherical harmonics parameterisation with a smoothing regularisation term. Spherical harmonics tomographic inversions have a long history in global geophysical inversion, however a common issue with this parameterisation is ringing artefacts resulting from truncation, similar to ringing caused by Fourier series truncation. Remedies have been proposed that reduce the ringing by imposing a tapering of the spherical harmonics order rather than a truncation [Whaler and Gubbins, 1981] but a potential drawback of tapering is a predilection for smooth models and a corresponding reduction in resolution.

The aim of this chapter is to demonstrate in a proof of concept study that a large scale database of multi-mode global surface wave observations can be inverted using a

Period	Mode					
(s)	Fundamental	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
40	84,145	83,597	79,118	72,319	69,061	77,414
45	84,145	83,250	75,976	76,981	68,371	81,487
50	84,150	82,946	71,107	79,417	73,861	82,480
60	84,172	82,457	66,146	76,992		
70	84,189	82,226	69,251	79,747		
80	84,203	82,227	80,496	84,206		
90	84,211	82,350	82,798	84,186		
100	84,193	82,489	83,558			
120	110,462	110,578	110,486			
140	110,336	110,576	43,035			
160	110,458	110,564	43,034			
180	110,335	110,556				
200	110,093	43,040				
220	43,040	43,040				
240	43,040	43,040				
260	43,038					
280	43,029					
300	42,993					
320	42,726					
340	42,338					
360	42,720					
Total	1,568,016	1,232,936	805,005	553,848	211,293	241,381

Table 6.1: The number of observations at each period and mode in the data set of Durand et al. [2015].

sampling trans-dimensional approach. The data in question is published by Durand et al. [2015] and consists of 4,612,479 Rayleigh wave surface wave observations up to the 5th overtone with the distribution of observational periods and modes shown in Table 6.1. This data set is several orders of magnitude larger than has been attempted before with a trans-dimensional sampling approach and therefore provides a good test of the computational viability of trans-dimensional inversion of large data sets.

6.3 Problem description

The problem is similar to the ambient noise study in Chapter 5 in that the observations are path average Rayleigh wave phase velocities between two points on the Earth surface. The distinction between these observations and discussed in Chapter 5 are three fold: firstly, the observations include higher order modes which increases the number of observations. Second, the paths are major arc paths that span large portions of the globe rather than being confined to a small region and means that the path integral in a forward model needs to integrate over a much longer distance for each observations. Finally, the number of observations is several orders of magnitude larger than the ambient noise study. These three factors together mean that the computational effort to evaluate predictions using a forward model for the global surface wave data set is significantly larger than the ambient noise study in Chapter 5.

Characteristic of most seismic studies, the location of the source events of sufficient magnitude to generate surface wave observations at large distances are concentrated in areas of high seismicity such as subduction zones and spreading ridges. Similarly, the seismic recording stations where events are recorded are concentrated on land masses which leads to a uneven distribution of events and stations across the globe as shown in Figure 6.2.

To make this problem more tractable, the complexity of forward model is reduced by assuming that great circle ray paths are sufficient or serve as a suitable approximation. The physical propagation path of surface waves is deflected by slow and fast regions. In the previous chapter it was demonstrated that when using fixed rays as opposed to computing rays, the recovery of models in synthetic tests was affected in both the resolution resolved and in an under-estimation of the magnitude of fast and slow anomalies. The limitations of a linearised approach must be accepted, as the scale of the problem prohibits the use of the Fast Marching method for re-computing travel times at each step for a fully non-linear inversion. This approximation reduces the forward modelling to integrating great circle paths through a proposed model that can be efficiently

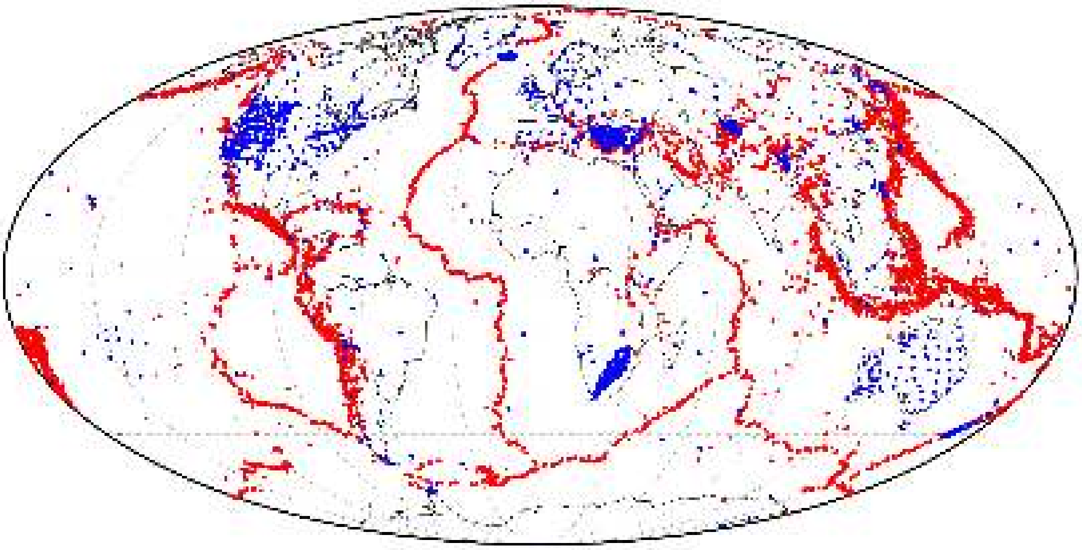


Figure 6.2: The distribution of events (red dots) and recording stations (blue dots) at a period of 100 seconds in the global surface wave data set of Durand et al. [2015]

.

computed with a suitable parameterisation.

6.4 Parameterisation

Two different applications of the use of wavelets have previously been discussed in Chapter 4 and Chapter 5, where they provided an effective parameterisation for image based geophysical inversions. A key benefits of using wavelets is their ability to represent reasonable structure with relatively small number of parameters. However, their major downside is a non-intuitive prior that has been addressed using a hierarchical prior in Chapter 4. The extension to a 2D spherical surface would appear straightforward, however the representation of fields on the surface of a sphere is a long standing and active area of research, a useful review article is Fasshauer and Schumaker [1998].

An obvious choice of parameterisation for spherical problems would be spherical harmonics. However the difficulty of setting a prior on coefficients with non-physical meaning is the major issue associated with both spherical harmonics and wavelet parameterisations. A second problem with spherical harmonics is that they are a global basis function and this is seemingly incompatible with trans-dimensional sampling. Trans-dimensional sampling, with the exception of simple examples shown in the introductory chapter, tends to locally adapt complexity driven by the information within the observations. This is implicitly the case for Voronoi cell parameterisations and explicitly so for the trans-dimensional tree. However, in a typical inversion using spherical harmonics, the series is truncated up to some degree chosen by the practitioner. Although in principle it would be possible to select this level through trans-dimensional sampling (or Bayesian evidence in linear problems), in practice it would be a very difficult problem as an “increase degree” proposal would change the model by a large amount making such trans-dimensional steps unlikely to succeed.

Regionalised versions of spherical harmonics are possible based on creating maximal energy concentration of spherical harmonics in a region [Simons et al., 2006, Dahlen and Simons, 2008, Simons, 2010, Simons et al., 2011a] with extensions available to the 3D ball [Khalid et al., 2014], were originally used in Fourier analysis by Slepian and Pollak [1960]. Simons et al. [2011a] introduce a multi-resolution version of these Slepian basis functions on the sphere that show promise for trans-dimensional sampling with the trans-dimensional tree algorithm developed in Chapter 3. However, Slepian functions give a basis function set to use in a region and it is not entirely clear how a trans-dimensional approach could be applied, although this may be an avenue for future research. Similarly, spherical splines based on spherical harmonics expansions may be another avenue for future exploration [Michel and Wolf, 2008, Amirbekyan and Michel, 2008, Berkel et al., 2011]

Several approaches have been made effectively using Cartesian wavelets on a sphere [Simons et al., 2011b, Charlety et al., 2013]. Both these methods start with the use of the projection of a cube on to the sphere [Ronchi et al., 1996] to tessellate the body with six square regions and parameterise each of these regions using a 2D Cartesian wavelet basis. However, lack of continuity across each of the six faces leads to artefacts. Simons et al. [2011b] remedied the continuity issue by using slightly overlapping faces with blending, whereas Charlety et al. [2013] wrapped four of the faces around the sphere using the lifting transform [Sweldens, 1996, 1998, Daubechies and Sweldens, 1998]. These solutions treat the problem aesthetically rather than from an inverse problem perspective and still introduce distortions resulting from unequal area of the pixels. The distortion problem can be potentially remedied using a different tessellation scheme that preserves equal area such as HEALPix [Górski et al., 2005], but this approach will always have edge discontinuity problems. Several spherical wavelet formulations have been developed [Antoine and Vandergheynst, 1999, Bonneau, 1999, Bogdanova et al., 2005, Guilloux et al., 2009, Lanusse et al., 2012, Leistedt and McEwen, 2012, Leistedt et al., 2013a] and some have been applied to inverse problems on the sphere [Holschneider, 1996, Holschneider et al., 2003, Chambodut et al., 2005, Michel, 2005, Holschneider and Igelwska-Nowak, 2007, Leistedt et al., 2013a,b]. Many of these approaches are designed for large scale problems and therefore utilise efficient algorithms. Some advances also include directionality [Hayn and Holschneider, 2009] that could be used for tomographic solutions incorporating anisotropy.

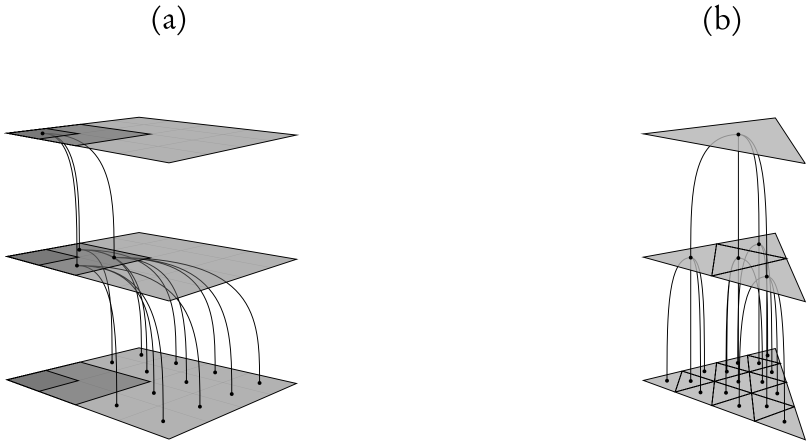


Figure 6.3: Comparison between wavelet pyramid (a) and triangular pyramid scheme (b).

Another class of solution of problems on the sphere are subdivision methods, with the most commonly used icosahedral tessellation [Baumgardner and Frederickson, 1985]. This approach uses a base level subdivision of the sphere that is the projection of an icosahedron onto the spherical surface, resulting in spherical triangles. Each triangle can be subdivided into four equal sub-triangles by dividing each edge of the triangle in half. This subdivision is often known as butterfly subdivision [Dyn et al., 1990]. This icosahedral approach has been used as a grid for a spherical spline scheme [Wang and Dahlen, 1995], and a novel wavelet basis [Schröder and Sweldens, 1995, Schroder and Sweldens, 1995] and for global scale tomographic inversions [Chiao and Kuo, 2001, Sambridge and Faletič, 2003].

An icosahedral parameterisation is well suited to the trans-dimensional tree approach. Figure 6.3 shows side by side the wavelet parameterisation introduced in Chapter 3 in (a) and the triangle subdivision of the icosahedral parameterisation in (b).

In contrast to the use of wavelet coefficients as model parameters, the model parameters can directly specified as the value of the field within each spherical triangle of the subdivision. This means that specifying the prior for the values of the parameters is now more intuitive, for the problem in this chapter it is phase velocity in kilometres per second, something for which there is prior information. The downside of this approach is the lost ability from the wavelet parameterisation to compress or reduce complex fields into a small number of coefficients, particularly for smooth models.

This new method of parameterisation consists of constant values in each triangle and this would, similar to a Voronoi cell approach, result in a model with discontinuities at triangle edges. Recalling the dual of the Voronoi cell tessellation is the Delaunay triangulation, if the node parameters are instead placed at the vertices of the spherical triangles of the icosahedral subdivision, values at points within each triangle could be linearly interpolated. Another option would be to use high order spline interpolants [Wang and Dahlen, 1995]. A complication here is the tree structure is no longer easy to define, although a spanning tree [Wu and Chao, 2004] could be constructed. A simpler approach is use the triangle centred nodes and use spherical barycentric coordinates [Alfeld et al., 1996, Langer et al., 2006] to produce a smooth interpolant between them. In this case the linear interpolant described in Langer et al. [2006] is used and this results in a C^0 continuous field over the sphere for all models with zero gradient at triangle centres.

6.5 Bayesian Formulation

Due to the size of the global surface wave data, all periods are not inverted simultaneously although this may be an area of future work. Instead, the inversion of a single period for a given mode allows the construction of a phase velocity map covering the globe. In the preceding sections the linearised forward model and the trans-dimensional tree parameterisation have been describe. In order to cast this into a Bayesian framework the prior and a likelihood function are required. Recall that a general expression for the prior from Chapter 3 was

$$p(\mathbf{m}) = \prod_{i=1}^{k} p(v_i \mid \mathcal{T}_k, k) p(\mathcal{T}_k|k) p(k), \tag{6.1}$$

of which, only the $p(v_i \mid \mathcal{T}_k, k)$ term remains to be defined. A benefit of parameterise this problem so that the tree nodes represent phase velocity is that the prior can be set in a more meaningful way. An option here is to set uniform priors for the phase velocity based on prior information on their expected maximal range for a given period and mode. An alternative, and the solution used here, is to use an empirical prior [Malinverno and Briggs, 2004] based on the statistics of the observed path average phase velocities. Hence for a given mode and period, the mean, μ_{mode} and standard deviation σ_{mode} of a Gaussian distribution are estimated from the observed path average phase velocities, and the prior is set on individual values for triangle node centres to

$$p(v_i \mid \mathcal{T}_k, k) \propto \exp\left\{-\frac{(v_i - \mu_{\text{mode}})^2}{2\sigma^2_{\text{mode}}}\right\}. \tag{6.2}$$

From the observations, there are estimates of the error for each observation as detailed in Debayle and Ricard [2012]. These can be used to form an independent Gaussian likelihood function

$$p(\mathbf{d} \mid \mathbf{m}) \propto \exp\left\{-\sum_i \frac{(G(\mathbf{m})_i - d_i)^2}{2(\lambda\sigma_i)^2}\right\}, \tag{6.3}$$

where $G(\mathbf{m})_i$ is the predicted path average phase velocity computed using a path integral along the great circle path between the event and receiving station, d_i is the observed path average phase velocity, σ_i the estimated error and λ a hierarchical error scaling term. A hierarchical error scaling term is included since the forward modelling is approximate, that is, isotropic great circle propagation paths are assumed and no ray bending and other physical effects such as anisotropy are considered. The hierarchical scaling term attempts to account for theory error caused by the approximate forward modelling and other unknown formulation errors.

Each period and mode was run with four independent chains for one million iterations starting from a homogeneous model with a phase velocity set to the mean of the path average phase velocity observations. This results in a total of 60 independent inversions. The time taken for each inversion ranges of approximately from 3 to 8 hours.

6.6 Phase velocity maps

The mean of the ensemble of the fundamental mode inversions of 40, 100, 200, and 340 second periods are shown in Figures 6.4, 6.5, 6.6, and 6.7 respectively. For comparison, results obtained by Durand et al. [2015] using a continuous regionalisation scheme [Tarantola and Valette, 1982b, Montagner, 1986, Debayle and Sambridge, 2004] with a smoothing regularisation controlled by a Gaussian covariance are plotted on the same scale as the phase velocity maps.

The inversion using the trans-dimensional approach agrees well with that of the previously published results of Durand et al. [2015]. In fact, the previously published maps have the appearance of smoothed or spatially filtered versions of the trans-dimensional maps. This is a common feature of the continuous regionalisation scheme as the spatial correlation in a smoothing regularised inversion needs to be tuned for the areas of least information to globally stabilise the inversion. In contrast with the trans-dimensional approach, the parameterisation adapts the local resolution of the inversion to the information in the observations, in this case independent crossing rays. This results in both higher resolution where supported by the data, and generally larger magnitude anomalies as they are not blurred or suppressed by some correlation length. Good examples of these are the greater detail in North America and in Northern Africa in Figure 6.4 where ray density is relatively high and are seismically active being subduction zones and areas of continental rifting. In contrast, large areas of the Pacific and Southern oceans where ray coverage is poor, show limited structure.

The benefit of a ensemble approach is that the ensemble can be interrogated to examine whether features in the mean images have strong support. Plotted in Figure 6.8 is the histogram of the 40 second period ensemble along a transect of constant Longitude at 37^o East, which passes through the African rift area whose location is indicated with the red vertical line. From the spread of the histogram in this region it is evident that the phase velocity is highly likely to be well below the reference velocity in this location. In comparison to the published phase velocity maps of Durand et al. [2015]

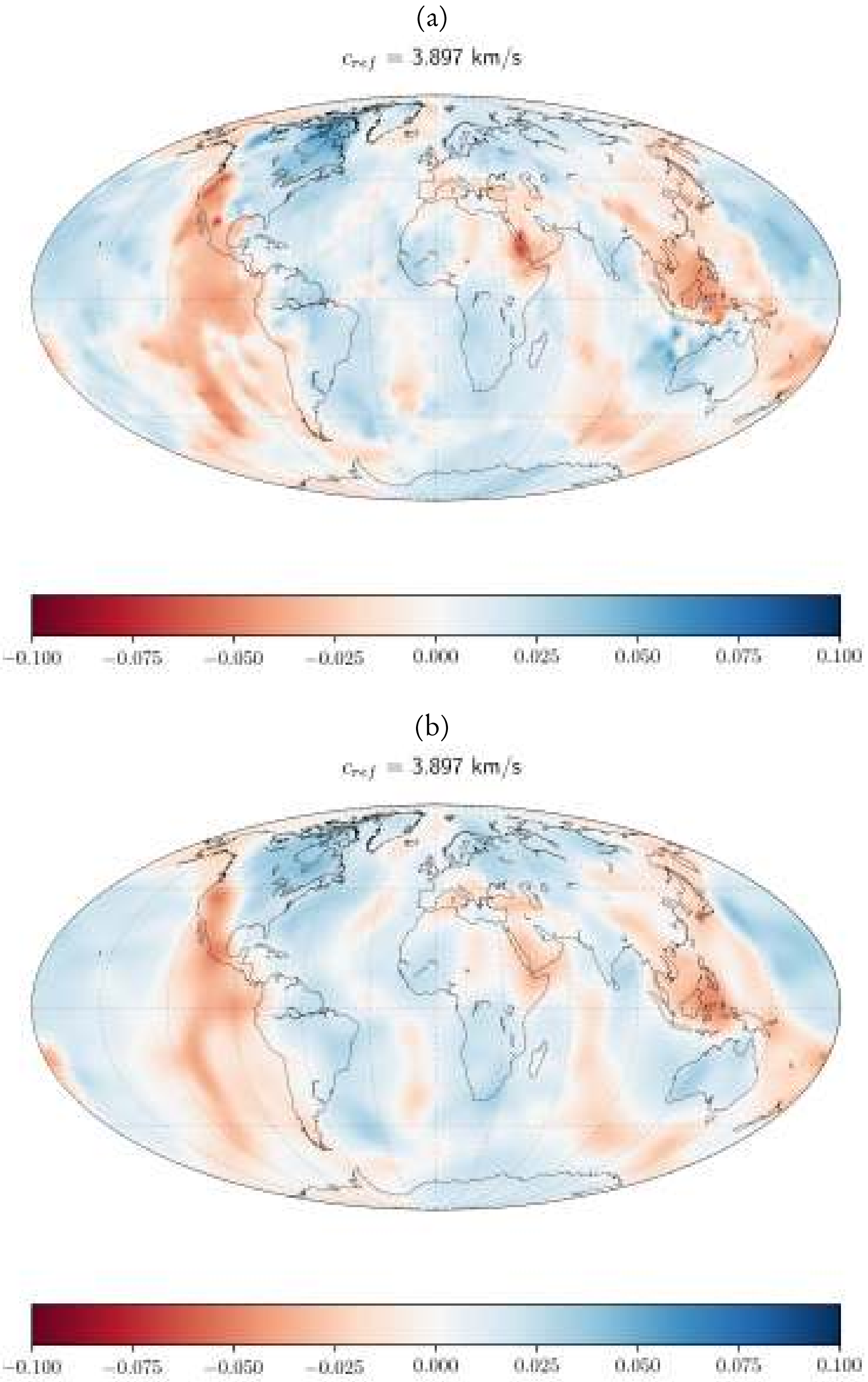


Figure 6.4: Comparison of fundamental mode phase velocity maps for the 40 second period inverted using trans-dimensional tree (a) and published maps of Durand et al. [2015](b)

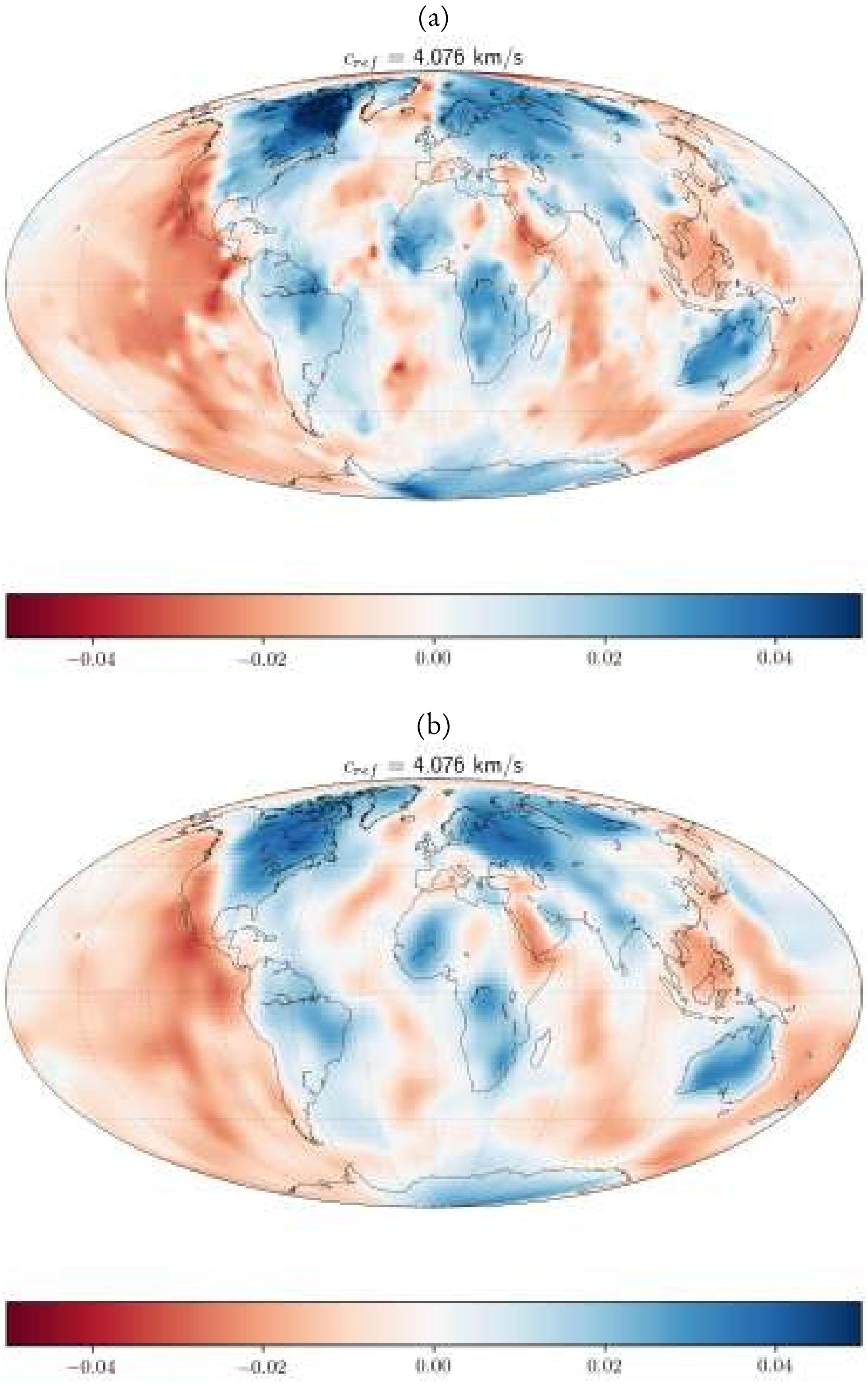


Figure 6.5: Comparison of fundamental mode phase velocity maps for the 100 second period inverted using trans-dimensional tree (a) and published maps of Durand et al. [2015] (b)

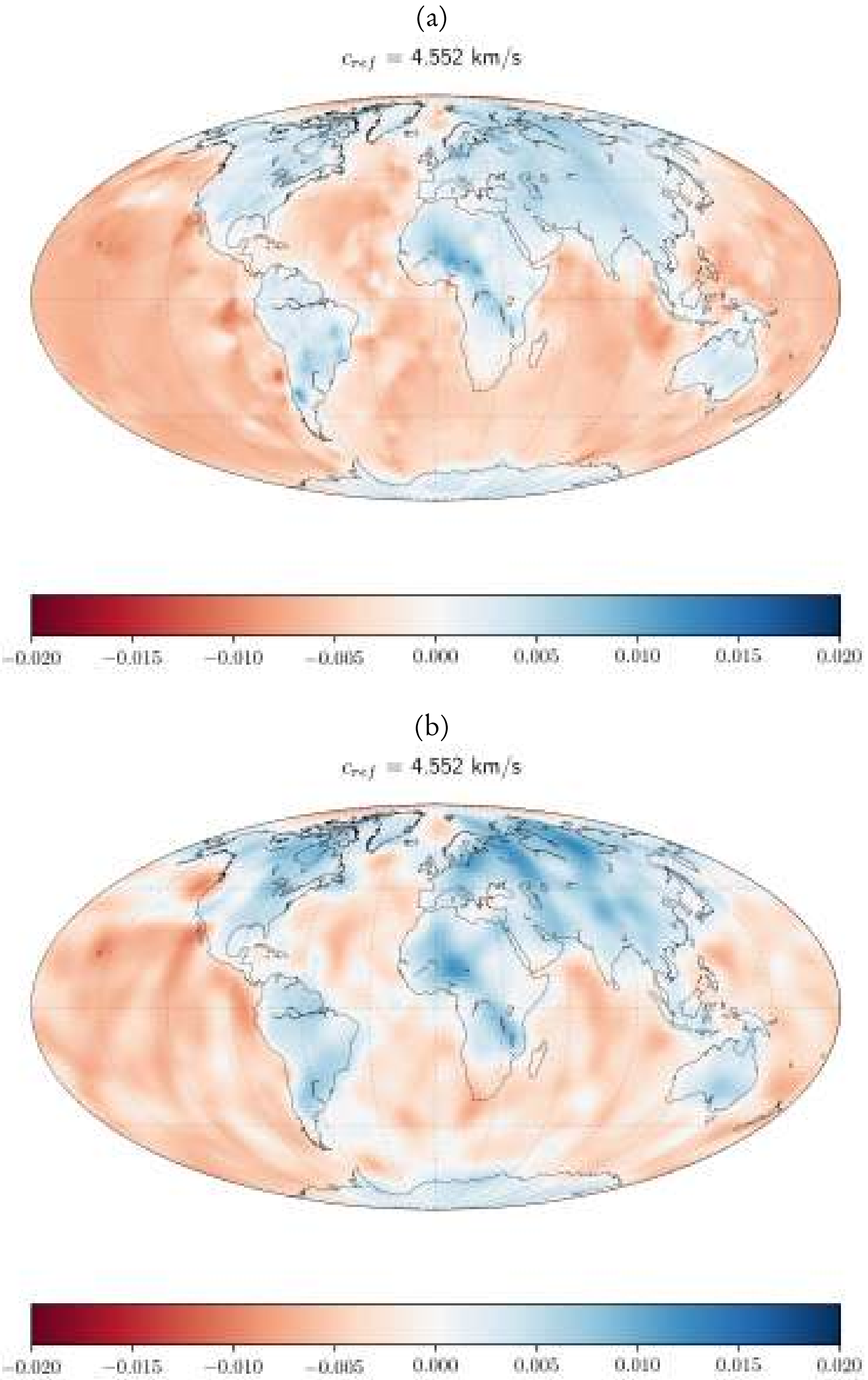


Figure 6.6: Comparison of fundamental mode phase velocity maps for the 200 second period inverted using trans-dimensional tree (a) and published maps of Durand et al. [2015] (b)

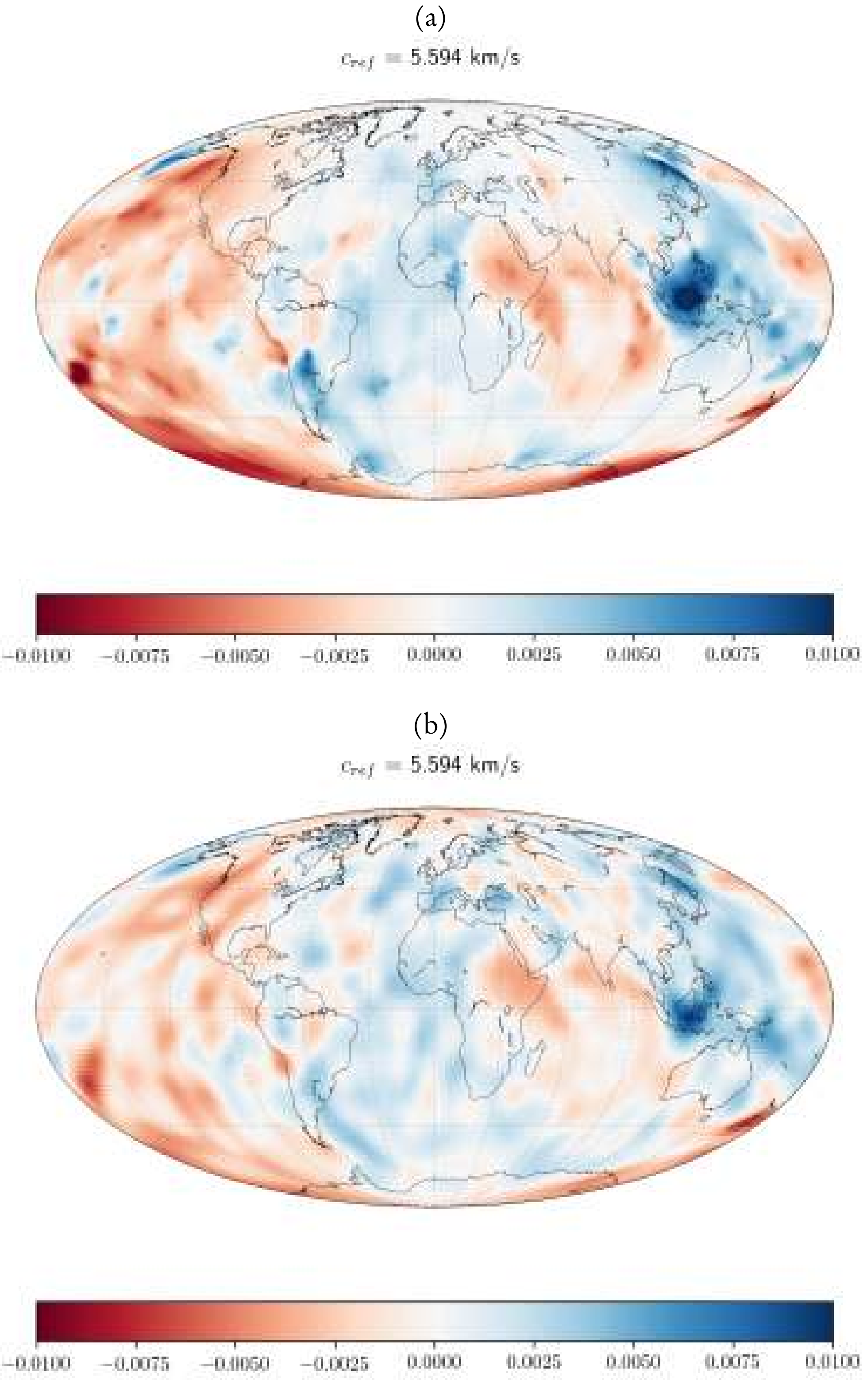


Figure 6.7: Comparison of fundamental mode phase velocity maps for the 340 second period inverted using trans-dimensional tree (a) and published maps of Durand et al. [2015] (b)

shown in this figure with a black dashed line, the smooth continuous regionalisation approach has significantly over estimated the phase velocity in this region. For the remainder of this transect there is generally good agreement.

Another interesting aspect of Figure 6.8 is between approximately -50 to -70 degrees latitude a bi-modality between flat structure and a more complex zig-zag structure is evident and this is an example of a trans-dimensional trade-off where the inversions includes aspects of a simpler model (flat structure) and a more detailed model (zig-zag). However, these figures represent preliminary results and more chains should be run with rigorous convergence criteria prior to more detailed interpretation.

In summary, it has been demonstrated that using the new trans-dimensional tree method, it is possible to trans-dimensionally invert large scale seismic data sets. A key early result is that the trans-dimensional approach produces significantly larger magnitude anomalies in some regions than previous continuous regionalisation approaches. In examining the uncertainties of one of these regions, the African rift, it is clear that this anomaly is well constrained suggesting it is a robust feature of the data. This underestimation of anomaly magnitudes could have consequences in subsequent inversion for a shear wave velocity model.

6.7 Shear wave maps

Once the phase velocity maps have been inverted, multi-mode 1D dispersion curves at each point on a regular grid of the Earth can be constructed using the ensemble mean and standard deviations. Similar to the methods outlined in Chapter 5, inversion is undertaken for a 1D shear wave velocity model beneath each point. Using the same approach as Durand et al. [2015], that is, a linearised inversion based upon depth sensitivity kernels computed for the PREM reference model [Dziewonski and Anderson, 1981], each point on a regular one degree grid is inverted. Two slices of the shear wave velocity model are shown as relative perturbations from PREM in Figures 6.9 and 6.10 for 100 km and 1000 km depth respectively.

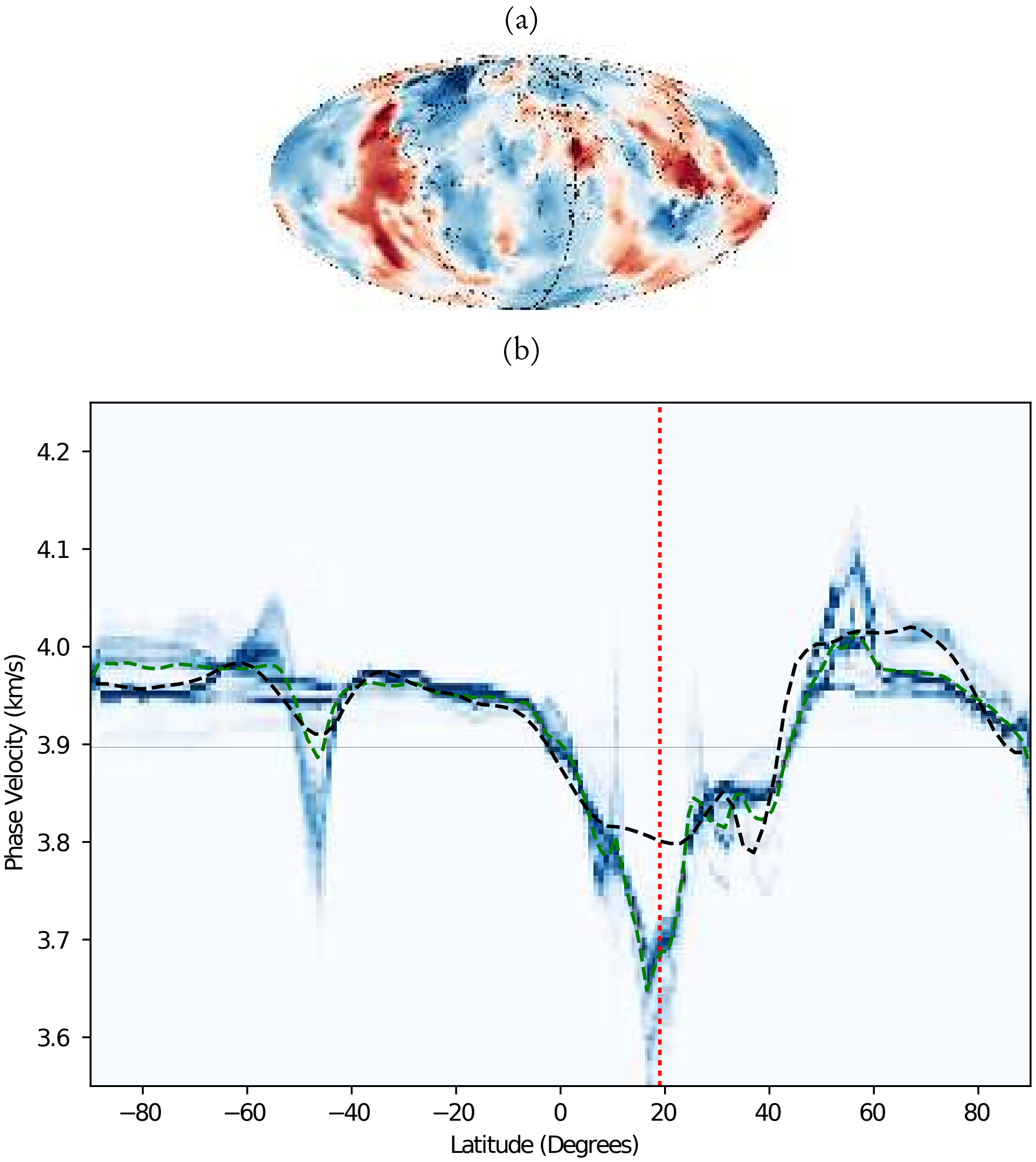


Figure 6.8: In (a) the location of the transect that passes through the African rift area is shown with a dashed line. In (b) is the marginal histogram of the ensemble along longitude 37° East shown with blue shading where darker blue corresponds to higher probability. The ensemble mean is indicated with a green dashed line, and the the phase velocity determined by Durand et al. [2015] is shown with a black dashed line. The location of the African rift is indicated with a vertical red dotted line in the plot. The horizontal line represents the reference velocity used in Figure 6.4.

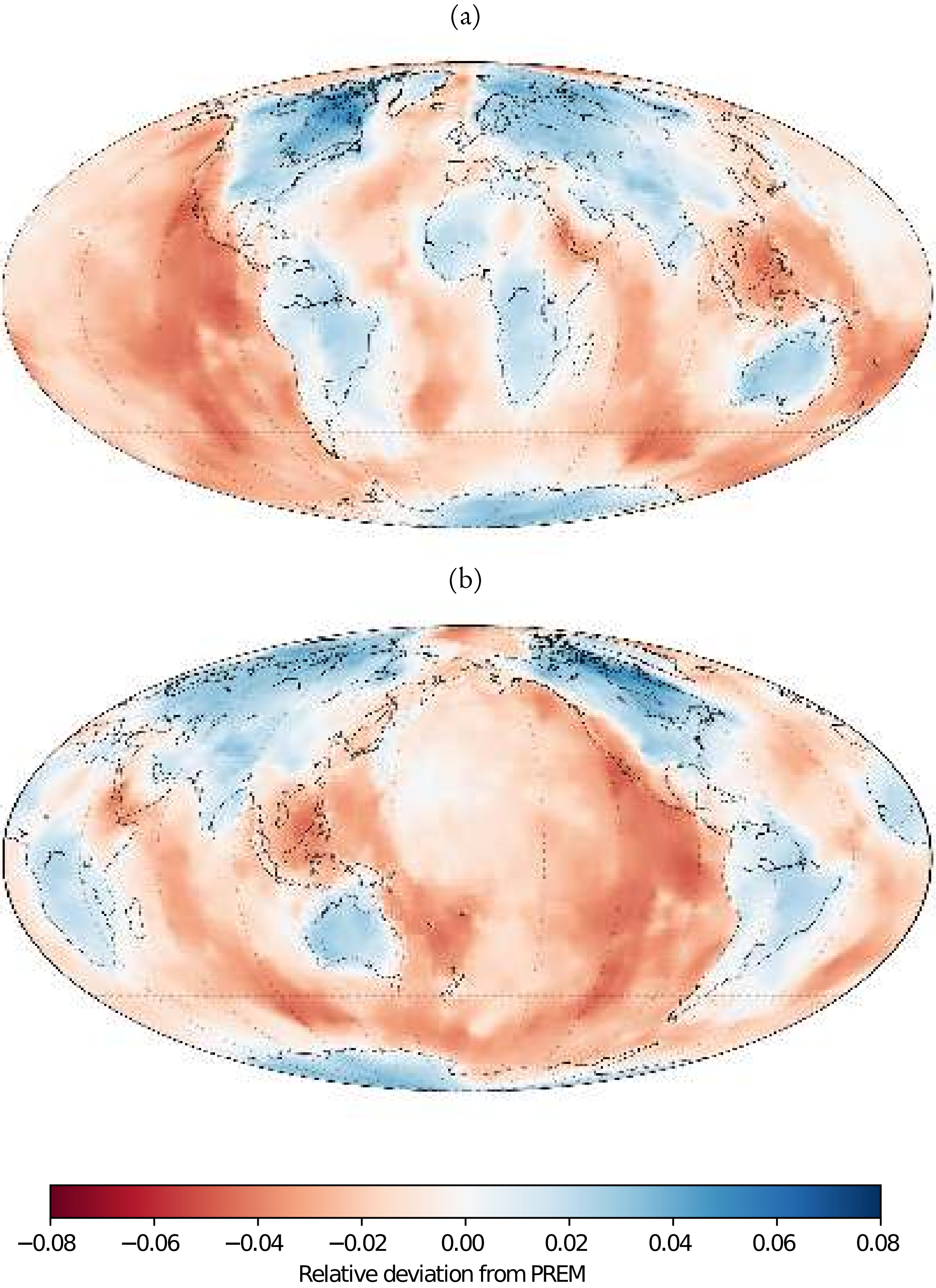


Figure 6.9: Shear wave velocity plotted as relative perturbation from PREM at 100 km depth

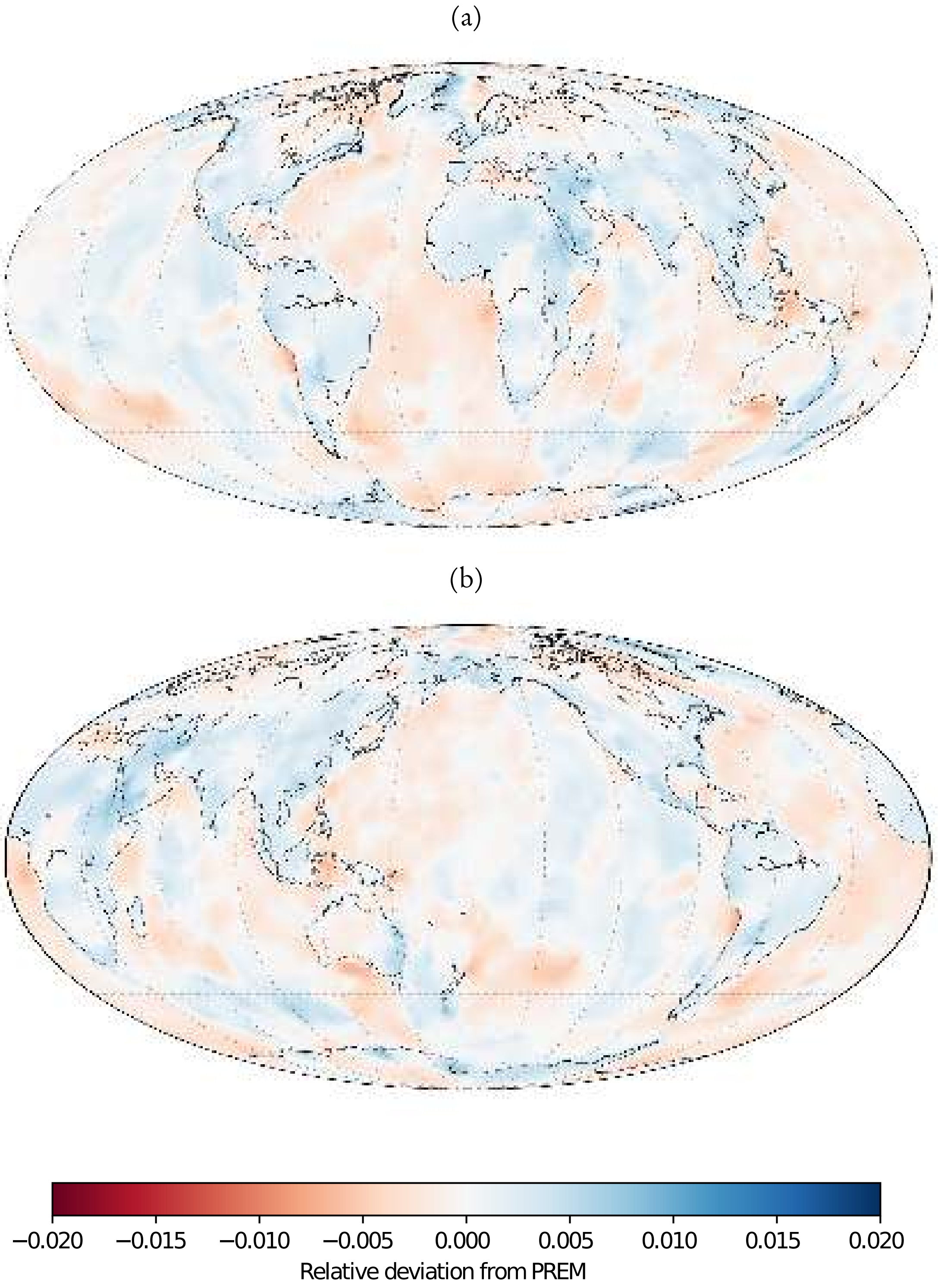


Figure 6.10: Shear wave velocity plotted as relative perturbation from PREM at 1,000 km depth

In the example shear wave velocity slice at 100 km depth, the image contains features of recognisable tectonic significance. Examples include the African rift region, the Atlantic spreading ridge, and the Canadian shield. At 1000 km depth, where the sensitivity of the surface wave observations diminishes, the anomalies are only of the order of one percent from the PREM reference model.

6.8 Summary

This chapter has developed a proof of concept to demonstrate the application of trans-dimensional sampling to a large scale global surface wave data set. By using some approximations such as fixed great circle paths and the assumption of isotropy, a very large number of observations have been inverted in a very modest time frame (ranging from 3 to 8 hours for 60 mode/period observations). This new technique has been successfully demonstrated and shows promise, with a number of different avenues available for future development.

It has been demonstrated in the phase velocity maps that in places where the trans-dimensional result produces significantly larger magnitude anomalies, these are well supported by the data as evidenced by their uncertainties. In contrast, previous continuous regionalisation inversions have perhaps produced overly smooth models in these regions, for example, the African rift. This under-estimation of magnitudes can cause subsequent under-estimation of shear wave velocity models inverted from multiple phase velocity estimates.

At present the inversion methodology uses Monte Carlo sampling to estimate the posterior. However, since a linearised approximation of the forward model and priors with analytical integrals (Gaussian) are used, it would be possible to analytically compute both the posterior of a model, and the evidence. This could be used to save considerable time in an inversion but would have limited such an inversion to a linearised solution. This option could be used to produce approximate starting model for a subsequent non-linear inversion.

The broad reasons for not pursuing this path was that in planned future work, namely the incorporation of extra physics into the inversion. The first of these is to include ray tracing using the fast marching method as was done for the ambient noise study of the Chapter 5. Fast marching is a general technique that can support triangulated domains instead of rectangular grids [Kimmel and Sethian, 1997, Sethian and Vladimirsky, 2000] and can be extended to support anisotropy [Sethian and Vladimirsky, 2003]. Secondly, and a factor that partially motivated the move away from wavelet parameterisations for this work is the incorporation of anisotropy [Visser et al., 2008]. This would entail having at each triangle centre three parameters to represent anisotropy and potentially inverting for the relative support of the data for isotropy versus anisotropy locally.

Finally, the preliminary inversion presented here is rudimentary in nature and could benefit from further parallelisation and include other features presented in earlier chapters such as hierarchically adjusting the prior and parallel tempering. Another potential extension is to take advantage of the coherency between periods and invert all periods for a given mode in a single inversion using ideas from Chapters 2 and 5. Nonetheless, the methods presented here represent significant advances and opportunities for the inversion of global large scale geophysical data sets.

Chapter 7

Trans-dimensional spectral elements

7.1 Introduction

The work of Sambridge et al. [2006] showed the equivalence of trans-dimensional sampling to Bayes factors or evidence based model choice. This was demonstrated in a regression problem where the fit of polynomials of various order formed the model choice problem, that is, which order polynomial best explained the data.

The parameterisation that was used in the study of Sambridge et al. [2006] was the leading coefficient terms of a polynomial. For the model **m** with components m_0, m_1, …, m_k, the polynomial was defined as

$$y(x) = m_0 + m_1 x + m_2 x^2 + \ldots + m_k x^k. \tag{7.1}$$

An issue with this parameterisation is that the prior specification for each of the coefficients is difficult, with perhaps the exception of m_0. It is rare in practice that one would ever have prior constraints on derivatives of a quantity $y(x)$ higher than the first. As the order of the coefficient increases, so does the sensitivity of the polynomial to small perturbations. This also means that proposals to perturb the higher order coefficients in McMC sampling are difficult to tune.

Various other polynomial parameterisations have been used in trans-dimensional sampling. In an early example, Denison et al. [1998] proposed using multiple lower order polynomials in a partition modelling approach for which prior and proposals might be more easily determined. Alternatively, Mallick [1998] proposed the use of Chebyshev orthogonal polynomials with reversible jump change of order proposals. This was integrated into a partition modelling scheme with variable number of partitions and variable order in each partition. In recent work, Bernstein polynomials have been used in a geo-acoustic inversion problem [Quijano et al., 2016].

An alternative scheme using nodal interpolation polynomials is possible, where instead of specifying the coefficients of individual power terms of the polynomial, the values the polynomial must attain at fixed interpolation points is specified. At the outset, this has the advantage that prior specifications are only required on the function values, rather than derivatives. This approach is similar to the Bernstein polynomial work of Quijano et al. [2016], yet subtly different as they specify control points which constrain the polynomial rather than directly specifying the values at known interpolation points.

As an example, for an order p polynomial with a set of $p+1$ fixed points $x_0 \dots x_p$ in a domain and desired values at those points $y_0 \dots y_p$, the unique polynomial coefficients in (7.1) can be obtained using a simple linear system

$$\begin{bmatrix} x_0^0 & x_0^1 & \dots & x_0^p \\ x_1^0 & x_1^1 & \dots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ x_p^0 & x_p^1 & \dots & x_p^p \end{bmatrix} \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_p \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_p \end{bmatrix}, \tag{7.2}$$

where the matrix on the left is commonly known as the Vandermonde matrix.

For higher order polynomials, the stability of this linear matrix problem is determined by the condition number of the Vandermonde matrix. The set of points on an interval $-1 \dots 1$ that minimise the condition number of the Vandermonde matrix are the Fekete points[Fekete, 1923, Saff and Kuijlaars, 1997, Bos et al., 2000]. In one dimension, the Fekete points correspond to the zeros of the Lobatto polynomial

$$L_n(x) = \frac{d}{dx} P_{n+1}(x), \tag{7.3}$$

where P_n is the Legendre polynomial and n is the polynomial order.

At a specified order, a set of orthogonal interpolating polynomials can be constructed using Lagrange cardinal polynomials

$$l_j(x) = \prod_{m=0, m\neq j}^{p} \frac{x - x_m}{x_j - x_m}, \tag{7.4}$$

where x_i is the ith Lobatto node. A continuous function f can then be approximated as a summation over these Lagrange cardinal interpolation polynomials

$$\hat{f}(x) = \sum_{j=0}^{p} f(x_j) l_j(x), \tag{7.5}$$

where $\hat{f}$ is the polynomial approximation of f. As an example, the set of Lagrange cardinal interpolation polynomials for a cubic interpolant is shown in Figure 7.1. A useful property of specifying polynomials in this fashion is the guarantee that for all orders, the maximum absolute value of the Lagrange cardinal interpolation polynomial is one and this occurs at the interpolation node. This results in a parameterisation of the polynomial that is stable for all orders producing minimal overshoot and ringing artefacts. This stability coupled with Gauss quadrature rules, have made these Gauss-Lobatto-Legendre (GLL) polynomials a popular parameterisation for the spectral element method in seismic full waveform simulation[Pozrikidis, 2005, Kopriva, 2009, Fichtner, 2011].

The GLL polynomials can be beneficial for polynomial fitting in a Bayesian framework in a number of ways. Firstly, for a single polynomial of a given order, the model is specified as the values the polynomial takes at the nodal points. These values are directly in the domain of interest and therefore each can share the same simpler prior specification. As an example, if the model represented the function of shear velocity versus depth in a 1D Earth model, the prior on each of the nodal values is a shear wave velocity distribution which is intuitive to specify. In contrast, if instead the polynomial had been parameterised in terms of leading coefficient terms, a different prior is required for each term and the prior for higher order terms would be difficult and non-intuitive to specify.

Another useful property is the minimisation of overshoot and ringing at higher orders that results from the guarantee that the maximum value of Lagrange cardinal polynomials occurs at the interpolation node. This means that if tight priors are set on the nodal values, the prior will be approximately respected over the entire polynomial. Care must be taken here though as there is a degree of overshoot. In Figure 7.2, and

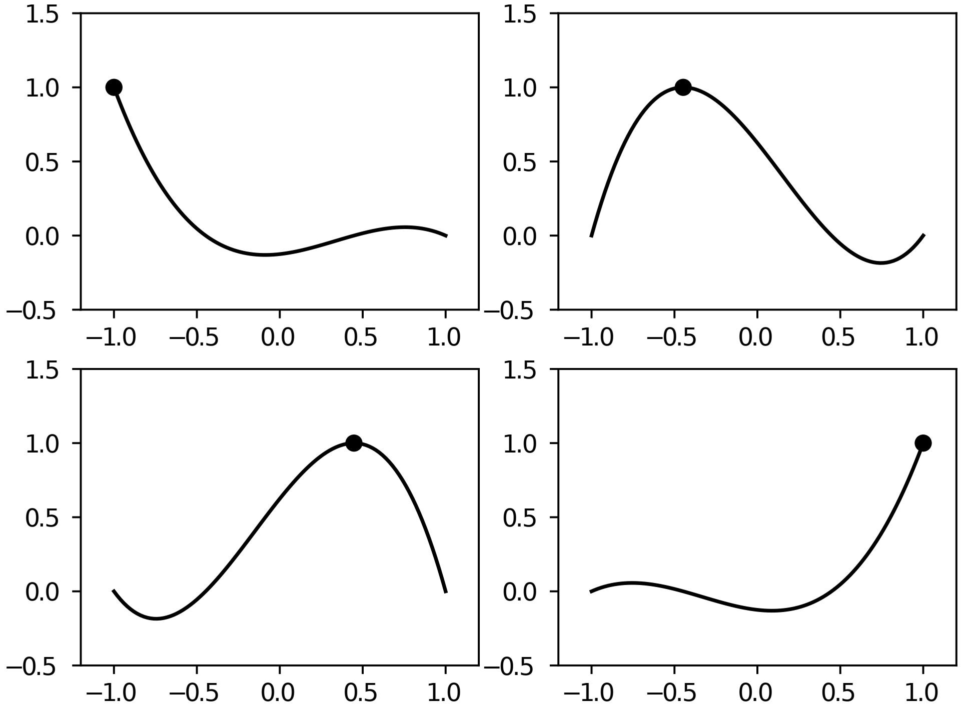


Figure 7.1: The cardinal functions of Gauss-Lobatto-Legendre polynomials for 3rd order polynomials which are polynomial approximations of delta functions at nodal interpolation points.

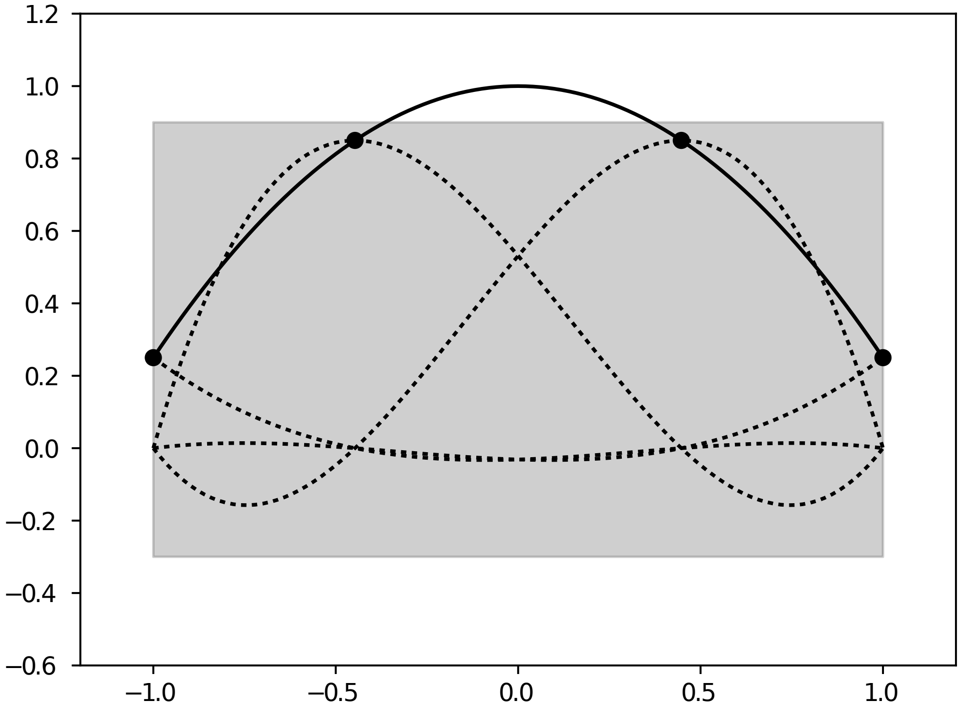


Figure 7.2: An example of the overshoot using GLL polynomials. The prior range is shown as a grey shaded region and although all the nodal interpolation points are within the prior, the resulting polynomial shown as a solid line has sections which are outside the prior bounds.

example of the potential to exceed a prior range is demonstrated. Here all the node values (black dots) are within the prior (grey shading) yet the resultant polynomial (solid line) breaches the range of the prior.

A final useful property is that since Lagrange cardinal polynomials are orthogonal, they can uniquely represent any polynomial with an order less than or equal to themselves. This property can be used for the projection of lower order polynomial models into higher order polynomials for improved forward model accuracy, similar to the wavelet super-resolution approach in Chapter 5. This aspect will be demonstrated in the following chapter.

7.2 Spectral element partition modelling

Having introduced the benefits of the specification of a polynomial in terms of orthogonal Gauss-Lobatto-Legendre (GLL) polynomials, a next step is to incorporate this into a trans-dimensional partition modelling framework. In a spectral element partition model, the model consists of a series of partitions along the x axis. Within each partition, the function is defined by an independent GLL polynomial of an arbitrary order. An example spectral element partition model with a linear, quadratic and cubic polynomial in three partitions is shown in Figure 7.3.

The GLL polynomials offer a number of alternative ways to treat the partition boundaries. First, is to use partition boundaries to represent discontinuities in the model as shown in Figure 7.3(a). A second alternative is to improve continuity of the interpolant across partition boundaries by sharing the boundary interpolation nodes between adjacent partitions which results in a piece wise continuous function with C^0 continuity. A consequence of this parameterisation, shown in Figure 7.3(b), is that the polynomial in each partition must be at least linear. Third, the values at the boundary nodes can be solved for to ensure a piece wise C^1 continuous curve. This is shown in Figure 7.3(c) and requires at least cubic polynomials in each partition.

In the treatment described in this chapter, the discontinuous parameterisation is preferred although all that follows equally applies to the other cases with only minor adjustments. With GLL polynomials, discontinuities of the interpolated function or data can be represented with partition boundaries and smoothly varying structure can be represented with higher order polynomials within partitions. The question of which representation, discontinuous steps or smoothly varying, best explains the data will be determined via trans-dimensional sampling.

The GLL polynomials are defined on the interval $-1\ldots1$ however each partition can be rescaled to arbitrary width with a simple transform

$$\xi = 2\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) - 1, \tag{7.6}$$

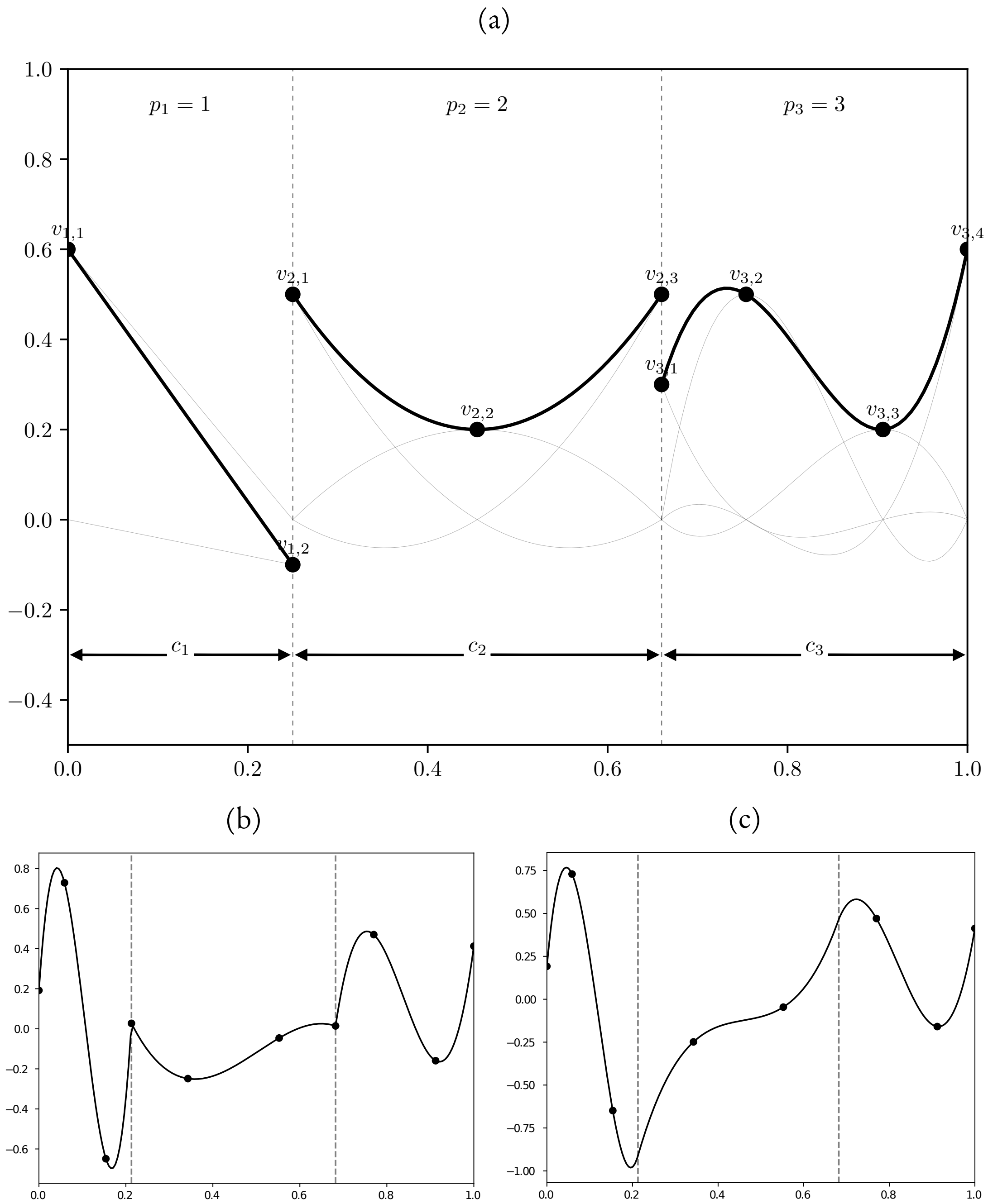


Figure 7.3: An example partition model with spectral element polynomials with discontinuities in (a). The partition polynomial is shown in a solid black line with the nodal interpolation points marked with circles. The individual scaled cardinal interpolation polynomials are shown in faint lines. Vertical dashed lines represent the partition boundaries. In (b) is shown an example where nodes on partition boundaries are shared to create a C^0 *continuous curve. In (c) is an example where nodes on partition boundaries as solved for in a linear system of equations to create a* C^1 *continuous curve.*

where ξ represents the variable x transformed to the $-1 \ldots 1$ domain and x_{max} and x_{min} represent the coordinates of the partition boundaries.

A model in this scheme consists of k, the number of partitions, where in each partition there is a vector $\mathbf{c}$ of the partition widths and a vector $\mathbf{p}$ of the polynomial orders. Lastly, for each partition there is a vector $\mathbf{v}$ of $p+1$ values at the nodal interpolation points. The various model parameters are labelled in Figure 7.3(a). Using this model parameterisation, priors and proposals required by a trans-dimensional framework can be specified.

7.3 Priors

From the description of the model parameters, the prior can be written

$$p(\mathbf{m}) = \prod_{i=1}^{k} p(\mathbf{v}_i|p_i)p(\mathbf{p}|k)p(\mathbf{c}|k)p(k), \tag{7.7}$$

where k is the number of partitions, $\mathbf{c}$ is the vector of partition widths which must sum to width of the entire domain, $\mathbf{p}$ is an integer vector of polynomial orders in each partition and a vector $\mathbf{v}_i$ of nodal interpolation values.

The prior on k is will commonly be uniform between one partition and some maximum number but other choices are possible. For the spatial prior on partition widths, the symmetric Dirichlet distribution is used which is expressed as

$$p(\mathbf{c}|x_{\text{min}}, x_{\text{max}}, \alpha, k) = k!(x_{\text{max}} - x_{\text{min}})^{-k} \prod_{i=1}^{k} \left[\frac{c_i - c_{i-1}}{x_{\text{max}} - x_{\text{min}}} \right]^{\alpha-1}, \tag{7.8}$$

where x_{min} and x_{max} represent the domain boundaries and α is the cell weight which if set to one reduces the symmetric Dirichlet distribution to the uniform distribution.

Similar to k, the prior on the order of polynomial in each partition is a prior on a model choice parameter which will generally be a uniform from zero order up to some maximum.

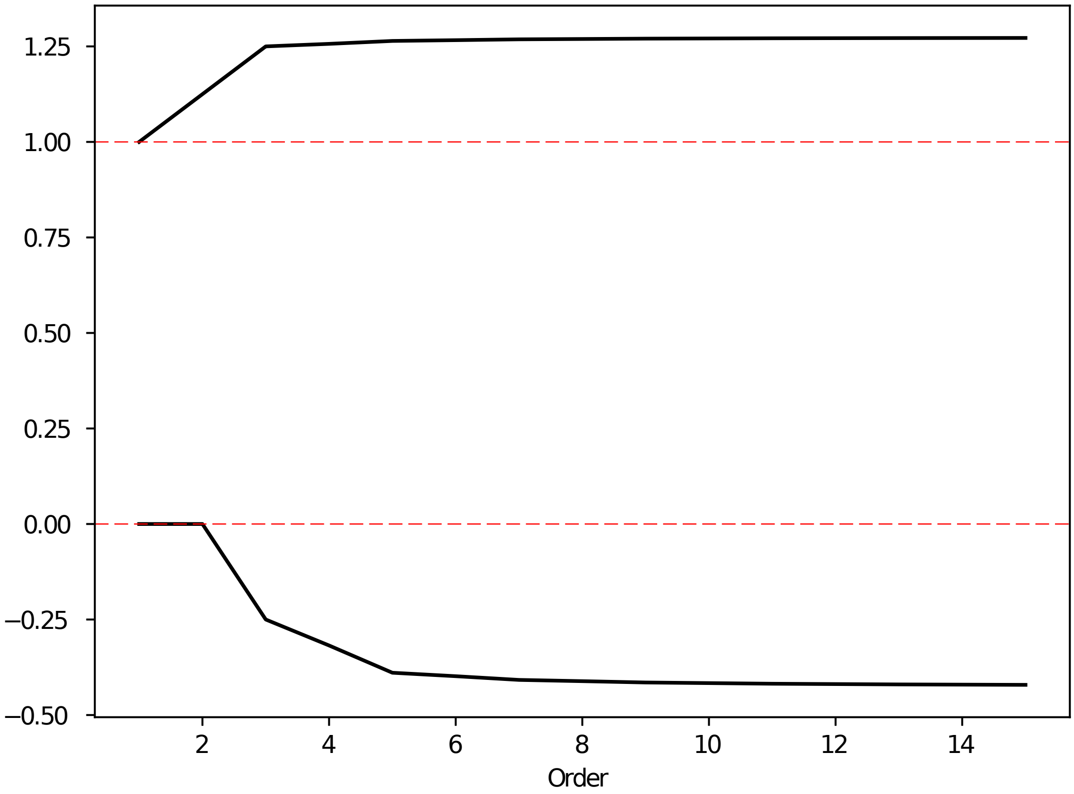


Figure 7.4: For each order the minimum and maximum overshoot is computed for a uniform prior between zero and one. A benefit of GLL polynomials is that the overshoot is bounded for all orders.

As stated earlier, the prior on the values at each node are directly specified in the y axis domain and are therefore easier to specify for most problems. There is the potential for overshoot so care must be taken. As an example, if the y axis represented a parameter such as velocity where negative velocities are invalid the temptation would be to set a prior from zero to some maximum velocity. Unfortunately this prior would allow some parts of the polynomial to be negative which may impact model misfits or forward model validity. Simple calculations can be used for guidance on overshooting problems, such as those presented in Figure 7.4 where the minimum and maximum overshoot of a uniform prior between zero and one are computed. For all orders, the overshoot is bounded, and calculations such as these can be used as a guide for ensuring overshoot doesn't impact forward modelling or model validity.

Since each of the interpolation nodes have a fixed location in the domain, it is also possible to include a prior with a dependence on location.

7.4 Proposals

Recall that the general Metropolis-Hastings-Green acceptance criteria is expressed as

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}') = \min \left\{ 1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{Q(\mathbf{m}' \rightarrow \mathbf{m})}{Q(\mathbf{m} \rightarrow \mathbf{m}')} |\mathcal{J}| \right\} \tag{7.9}$$

which can be expressed in words as the product of the prior ratio, the likelihood ratio, the proposal ratio, and the Jacobian. In this framework there are the following set of proposals:

1. Perturb a nodal value,
2. Perturb the location of a partition boundary,
3. Add a new partition,
4. Remove an existing partition,
5. Increase the polynomial order in a partition by one, and
6. Decrease the polynomial order in a partition by one.

For each of the candidate proposals, the detailed steps taken in the proposal and how the acceptance criteria is formed is described in the following sections.

7.4.1 Value

The simplest proposal is a value perturbation which changes the y value of an individual nodal interpolation point. In Figure 7.5, (a) shows the starting model and (b) the proposed model. The proposal consists of first selecting a partition, and then the nodal value within it to perturb. The selected node is indicated with a red circle in

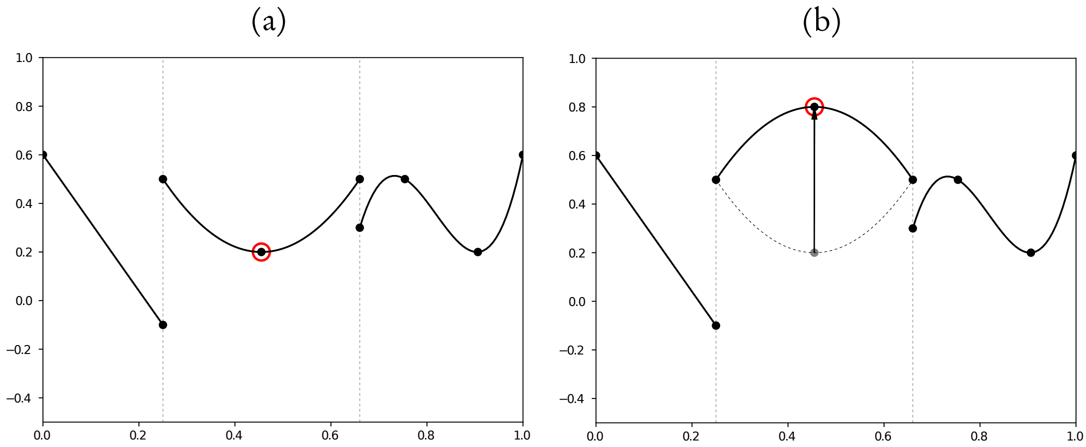


Figure 7.5: The value proposal starts with selecting an individual nodal interpolation point highlighted in (a) with a red circle. In (b) the proposed model is generated by perturbing the y value of the selected node.

Figure 7.5(a), and its perturbed position in Figure 7.5(b). The forward proposal can be written as

$$Q(\mathbf{m} \to \mathbf{m}') = \frac{1}{k}\frac{1}{p_i + 1} Q(v_{i,j} \to v'_{i,j}), \tag{7.10}$$

where p_i is the order of the polynomial in the selected partition and $Q(v_{i,j} \to v'_{i,j})$ is the proposal density for the perturbation of the nodes value, that is, in the y-direction.

It should be clear that if the proposal to perturb the value of a selected node is symmetric, for example, sampled from a Gaussian or Cauchy distribution, the proposal ratio will cancel. For the prior ratio, only one single value is perturbed and the rest of the prior will cancel leaving a term of the form

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})} = \frac{p(v'_{i,j})}{p(v_{i,j})}. \tag{7.11}$$

This ratio will only be unity if the prior on the y-values is uniform. In the case of a prior with finite support where a proposal results in a value outside of the prior range, the prior ratio is zero and the proposal is rejected.

Since there is a one to one relationship between the proposal distribution and the new

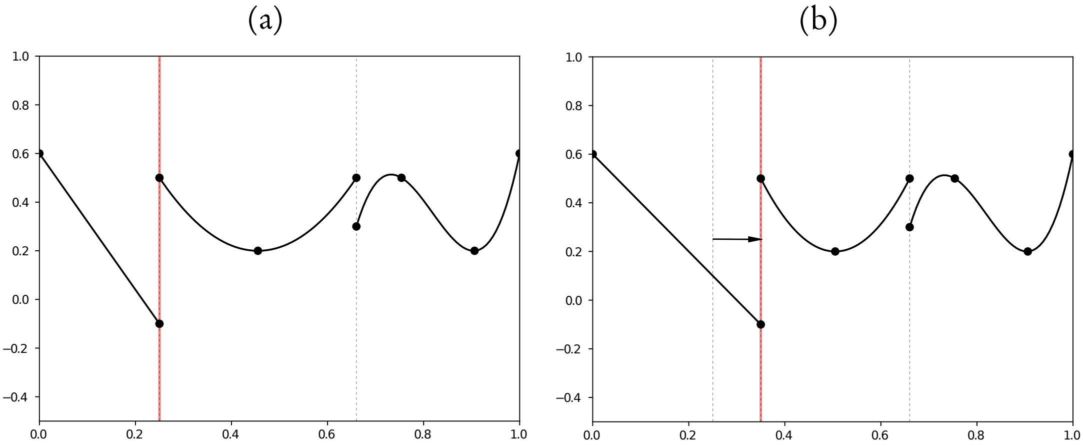


Figure 7.6: The move proposal starts with selecting a partition boundary to move highlighted in red in (a). The location of this boundary is perturbed to generate the proposed model in (b).

model variable, the Jacobian is unity, leaving the acceptance criteria for value proposals as

$$\alpha(\mathbf{m} \to \mathbf{m}')_{\text{value}} = \min\left\{1, \frac{p(v'_{i,j})}{p(v_{i,j})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}\right\} \tag{7.12}$$

7.4.2 Move

For a move proposal, the location of one of the partition boundaries is perturbed so the proposal, shown diagrammatically in Figure 7.6 consists of first selecting a partition boundary highlighted in red in (a) and the adjusting the location of the boundary to a new position in (b).

The forward proposal can be written as

$$Q(\mathbf{m} \to \mathbf{m}') = \frac{1}{k-1} Q(c_i, c_{i+1} \to c'_i, c'_{i+1}), \tag{7.13}$$

where c_i and c_{i+1} are the partition widths on either side of the selected partition boundary. The form of this proposal results from the fact that moving an internal boundary changes the width of two partitions. In some applications perturbing one of the extremal boundaries may also be appropriate[Steininger et al., 2013], in which case only one partition size will be adjusted.

As with the value proposal, if the proposal density of the perturbation is symmetric, the proposal ratio cancels. In the case where only the internal boundaries are perturbed and a symmetric Dirichlet prior is used with the α parameter set to one, the prior ratio will be unity. The exception to this is when a large perturbation of a partition boundary results in the ordering of boundaries changing. In this case the prior ratio is zero due to negative partition widths causing the Dirichlet prior to equate to zero and therefore such a proposal is rejected.

Lastly, there is a one to one correspondence between the proposal distribution and the perturbed model variables so the Jacobian is unity for this proposal leaving only the likelihood ratio for the acceptance criteria,

$$\alpha(\mathbf{m} \to \mathbf{m}')_{\text{move}} = \min\left\{1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}\right\}. \tag{7.14}$$

In the case where extremal boundaries are perturbed, or if the α parameter of the Dirichlet prior is not one, there will be an additional prior ratio term. When applicable, it is straight forward to evaluate (7.8) for both the current and proposed models to obtain the ratio.

7.4.3 Order Birth

The first trans-dimensional proposal is that of *order birth* within a single partition which proposes an increase in the order of a partition by one. This proposal is shown diagrammatically in Figure 7.7 where (a) shows the current model and (b) the proposed model where the central partition has its polynomial increased in order to three from the current model (shown superimposed in a dotted line) of order two.

The proposal consists of selecting the partition in which to increase the order, then proposing new values for each of the new interpolation nodal points. For GLL nodal points, with the exception of the partition end points, none of the nodal points will

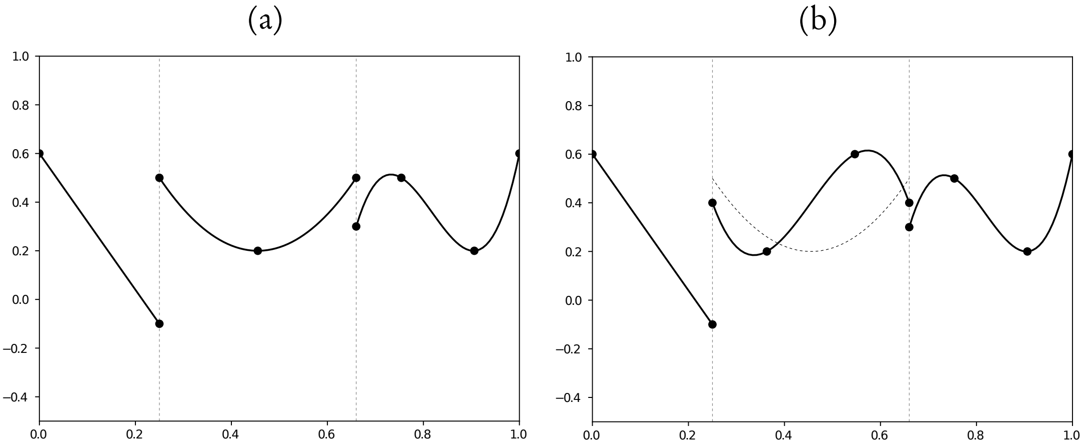


Figure 7.7: The schematic for the order birth proposal where the central partition in (a) with a quadratic polynomial, has the polynomial order increased to three in (b).

coincide between order increments. There are two options which are discussed here, the first and simplest is to resample each point from the prior and ignore the current model values. The second is to project the current model to the higher order interpolation nodes, and then add a perturbation to each value sampled from a proposal distribution. In either case, the general forward proposal can be written

$$Q(\mathbf{m} \to \mathbf{m}') = \frac{1}{k} \prod_{j=0}^{p'_i} Q(v_{i,j} \to v'_{i,j}) \tag{7.15}$$

where p'_i is the order of the proposed polynomial and $Q(v_{i,j} \to v'_{i,j})$ represents the proposal density for each of the new y values. The reverse proposal distribution is of the same form as the forward with the priming of the variables reversed meaning that in the proposal ratio, the leading $\frac{1}{k}$ term will cancel.

Before considering each class of proposal, in either case, the acceptance criteria will reduce to

$$\alpha(\mathbf{m} \to \mathbf{m}') = \min \left\{ 1, \frac{\prod_{j=1}^{p'_i+1} p(v'_{i,j})}{\prod_{j=1}^{p_i+1} p(v_{i,j})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{\prod_{j=1}^{p_i+1} Q(v'_{i,j} \to v_{i,j})}{\prod_{j=1}^{p'_i+1} Q(v_{i,j} \to v'_{i,j})} |\mathscr{J}| \right\}, \tag{7.16}$$

from which can be seen the attraction of "birthing from the prior" where $Q(v'_{i,j} \rightarrow v_{i,j}) = p(v_{i,j})$ and vice-versa, that is the proposal is to sample new values from the prior. Since the prior and proposal ratios cancel, and the Jacobian is unity due to a one to one mapping between random variables and new node values, the acceptance criteria for this type of order birth proposal reduces to the likelihood ratio.

A potential downside of birthing from the prior is that if the prior is relatively wide compared to the posterior, then the acceptance rates of *order birth* steps can drop precipitously decreasing efficiencies of sampling.

Since any polynomial of a given order can be exactly represented with a polynomial of the next highest order, a more focused proposal can take advantage of this by projecting the current polynomial to the next higher order. In a GLL polynomial parameterisation, this entails interpolating the current polynomial to the next highest order nodal points. To the interpolated values at each of the higher order nodal points, more focused perturbations can be added. This results in the acceptance criteria of the form in (7.16), where the prior and proposal ratios need to be fully evaluated.

Unlike in the case of birthing from the prior, the added complexity in this proposal, from the interpolation of the new model parameters from the current, results in a non-identity Jacobian matrix. To evaluate the Jacobian the model bijection relationship is required, which in this case is

$$v_1, \ldots, v_{p_i+1}, u_1, \ldots, u_{p_i+2} \longleftrightarrow v'_1, \ldots, v'_{p_i+2}, u'_1, \ldots, u'_{p_i+1}, \tag{7.17}$$

where v are the nodal values, u are the random variates for the perturbations and p_i is the initial polynomial order. It should be clear that the number of nodal values on the left in (7.17) is one less than on the right and similarly the number of random variables on left is one more than on the right in order for there to be an equal number of variables on both sides.

To compute the Jacobian, the relationships between the variables are also needed. A short hand notation is introduced here where $P(v_1 \ldots v_{p_i+1}, x'_j)$ is the projection, or interpolation function, that projects the polynomial of order p_i to the higher order polynomial at nodal point x'_j. With this short hand notation, the relationship for the proposed values in terms of current values is

$$v'_j = P(v_1 \dots v_{p_i+1}, x'_j) + u_j, \tag{7.18}$$

which represents the projection of the polynomial to the higher order nodal point x'_j plus a random perturbation u'_j. A similar operator can be written for the reverse for the relationships between the random variables u'_j and the current values, that is

$$v_j = P(v'_1 \dots v'_{p_i+2}, x_j) + u'_j. \tag{7.19}$$

The Jacobian for the proposals between each order are independent of the model and represent volume change corrections for the trans-dimensional steps. As such the Jacobian determinant can be computed once and stored. The Jacobians for a suite of *birth order* proposals were numerically evaluated and all found to be unity, so they fall out of the acceptance criteria leaving

$$\alpha(\mathbf{m} \to \mathbf{m}')_{\text{birthorder}} = \min\left\{1, \frac{\prod_{j=1}^{p'_i+1} p(v'_{i,j})}{\prod_{j=1}^{p_i+1} p(v_{i,j})} \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{\prod_{j=1}^{p_i+1} Q(v_{i,j})}{\prod_{j=1}^{p'_i+1} Q(v'_{i,j})}\right\}. \tag{7.20}$$

7.4.4 Order Death

The proposal for *order death* or reducing the order of a polynomial by one, is the reverse of *order birth* and is shown diagrammatically in Figure 7.8. In this class of proposal, a partition is selected to perform the *order death* proposal that decrements the order of the polynomial in the partition. The new values for the lower order polynomial need to be either sampled from the prior or generated from perturbations of the lower order projection of the current model, similarly to the *order birth* proposal.

The acceptance criteria is unchanged from the previous section, and given by (7.20).

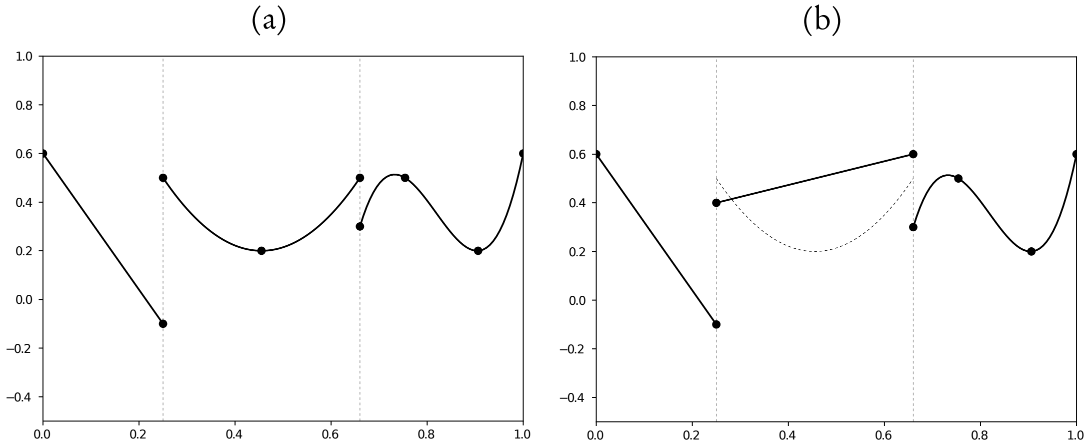


Figure 7.8: The schematic for the order death proposal where the central partition in (a) with a quadratic polynomial, has the polynomial order decreased to one in (b).

7.4.5 Partition Birth

For a partition birth proposal, the steps taken are

1. select a partition,
2. select a partition splitting point to divide this partition, then
3. select new orders for each of the partitions, and
4. choose new values for each of the new nodes.

This is shown diagrammatically in Figure 7.9 where in (a) highlighted is the selected partition to split and (b) shows the new partitions and polynomials with the old polynomial plotted as a dotted line.

This proposal can be expressed mathematically as

$$Q(\mathbf{m} \to \mathbf{m}') = \frac{1}{k}\frac{1}{c_i} Q(p'_i) \prod_{j=0}^{p'_i+1} Q(v'_{i,j})\, Q(p'_{i+1}) \prod_{j=0}^{p'_{i+1}+1} Q(v'_{i+1,j}), \tag{7.21}$$

where $Q(p'_i)$ and $Q(p'_{i+1})$ are the proposal distributions for order of each of the new partitions, and similarly $Q(v'_{i,j})$ and $Q(v'_{i+1,j})$ are the proposal distributions for the

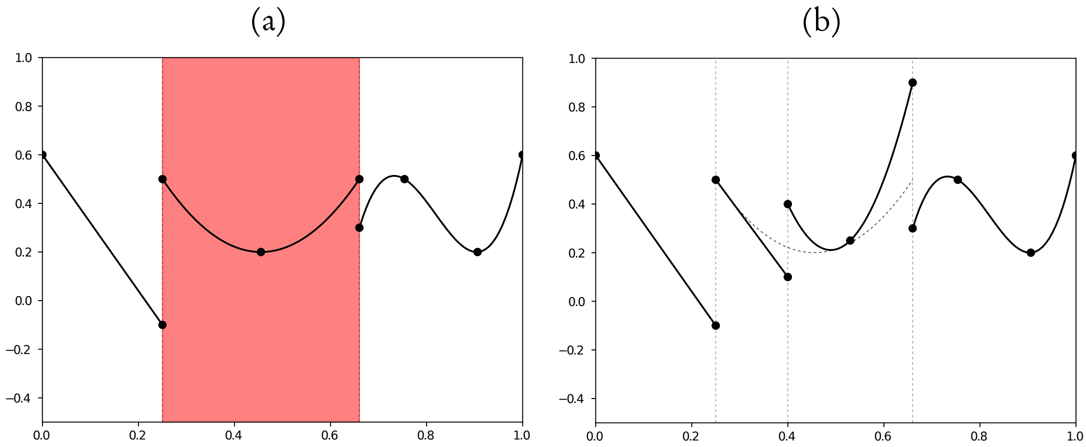


Figure 7.9: For a partition birth proposal, a partition is selected to split, highlighted in red in (a). The partition is split at a randomly chosen location and new polynomials of random order are generated in the two new partitions as shown in (b).

new values in each of the new partitions. Just as in the case for the *order birth* proposal, two alternatives exist for the choosing of the new values, either to sample from the prior, or use projection of the current model onto the two new partition nodal interpolation points and add small perturbations. Due to the added complexity, sampling from the prior is the only partition birth proposal considered, although more focused proposals analogous to the *order birth* proposal are possible.

Proposing using a uniform distribution between $0\ldots1$ for determining the location of the partition splitting point results in a non-unity Jacobian. Similar to the partition approach in Chapter 2, this results in a Jacobian of

$$|\mathcal{J}| = c_i, \tag{7.22}$$

where c_i is the width of the partition being split, and as such the acceptance criteria for this proposal reduces to

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}')_{\text{birth}} = \min\left\{1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} c_i\right\}. \tag{7.23}$$

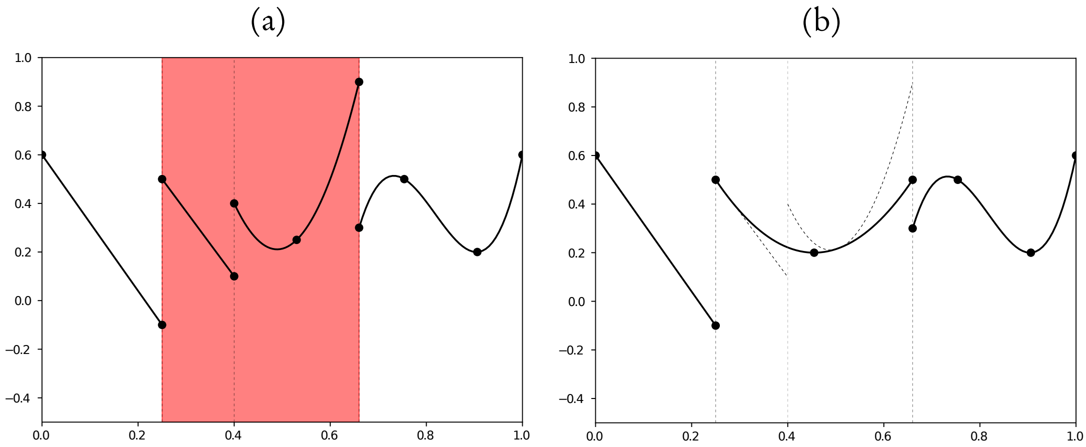


Figure 7.10: For a partition death proposal, two neighbouring partitions are selected to merge highlighted in red in (a). A new partition is formed as the sum of the two neighbouring partitions, and a new polynomial of random order is generated as shown in (b).

7.4.6 Partition Death

For the partition death proposal, shown diagrammatically in Figure 7.10, a partition is selected to be merged with its right neighbour. This means the selection occurs from one of $k-1$ partitions, that is, from all partitions except the last.

The acceptance criteria becomes

$$\alpha(\mathbf{m} \rightarrow \mathbf{m}')_{\text{birth}} = \min\left\{1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \frac{1}{c_i + c_{i+1}}\right\}, \tag{7.24}$$

where c_i and c_{i+1} are the widths of the two partitions being merged into one.

7.5 Synthetic Regression Example

These proposals combined form a general trans-dimensional inversion method for 1D problems where the underlying model may be some combination of a smoothly varying function and discontinuities. To verify this method, a test suite is constructed using observations from three different underlying functions shown in Figure 7.11. These can be described as (a) a simple cubic function, (b) a set of three constant valued step functions, and lastly (c) a set of three partitions with different order polynomials

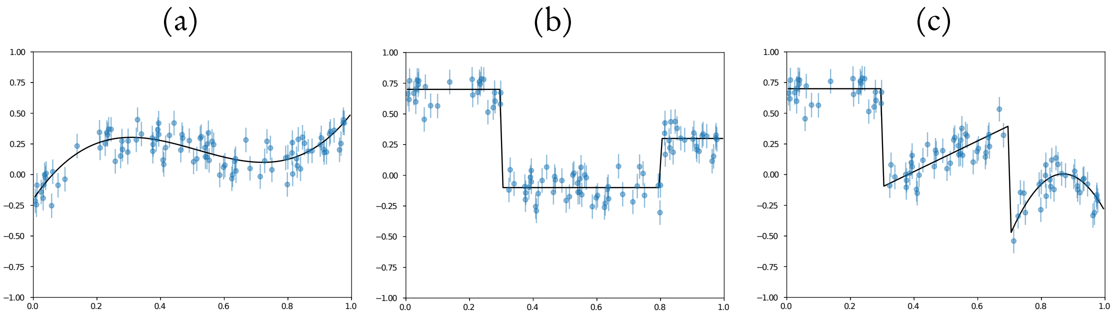


Figure 7.11: The three different synthetic data examples used in this study. (a) is a continuous smooth function, (b) is a a partition model with discontinuities with constant polynomials in each partition, and (c) is a partition model with discontinuities and variable order polynomials in each partition. In each plot, the true data is shown with a black line and the synthetic data observations with points.

in each. For each of three true functions, 100 random sample points are uniformly generated along the x-axis and the true y value of the function is evaluated. To the true value, independent zero mean Gaussian noise is added with a standard deviation of 0.1 or approximately 10% error. In Figure 7.11, the true function is shown with a solid black line, and the noisy data shown with blue points and representative error bars.

Three inversions are performed for each dataset with different trans-dimensional options available, these are

1. a single cell but with trans-dimensional order,
2. trans-dimensional partitions but with fixed zeroth order in each partition, and
3. trans-dimensional partitions with trans-dimensional order in each partition.

For each inversion it is assumed that the data noise is estimated and so a hierarchical error scale is used to estimate the noise level [Malinverno and Briggs, 2004].

Where applicable, a uniform prior is used on the number of partitions of between 1 and 10, and a uniform prior on the polynomial order in each partition of between 0 and 5. With the order birth/death proposals the projection with Gaussian perturbation approach is used. The value, move and hierarchical proposals are configured identically for all of the inversions and tuned to obtain a reasonable acceptance rate across all inversions. Each inversion is run for one million iterations with the first 500,000 iterations removed as burnin.

The results of all nine inversions are shown in Figure 7.12 in a grid layout. The rows correspond to the synthetic data shown in Figure 7.11. The first column shows the results for the single partition inversions with a trans-dimensional order. The second column shows the results of the trans-dimensional partition inversion with fixed zeroth order polynomials in each partition. The third column shows the results with trans-dimensional partitions with trans-dimensional order within each partition. In the results, the blue shaded region shows the posterior histogram of the ensemble of candidate curves and green dotted lines show the mean of the ensemble.

Here the three true models are chosen to best suit one of the three inversion methods used. For example, the first model should be well recovered by the single partition with trans-dimensional order inversion. Similarly, the second model should be well recovered by the trans-dimensional partitions with zeroth order polynomials. It is evident that this is indeed the case in Figure 7.12 (a) and (e), but in both these synthetic models, the trans-dimensional partition with trans-dimensional order polynomials has visually identical results in (c) and (f). The results in Figure 7.12 (c) and (f) are striking examples of the parsimony of trans-dimensional sampling in action. Even though the trans-dimensional partition and order inversion is able to add more partitions in (c) or increase the polynomial order in (f), these extra complexities in the model aren't required to explain the data further, given the inverted for noise level.

In 7.12(i), the method using GLL polynomials with trans-dimensional sampling of partitions and polynomial order, is able to successfully adapt to both discontinuities and variable polynomial order without knowing in advance which is required by the data. It is also clear in (g) and (h) that approximations of the true model can be recovered but the posterior exhibits artifacts related to the limitations within the class of parameterizations available. This has important ramifications of subsequent inferences, for example, a tendency to miss the presence of discontinuities in the data that are detectable, as in (g); or the identification of false partition boundaries that are not

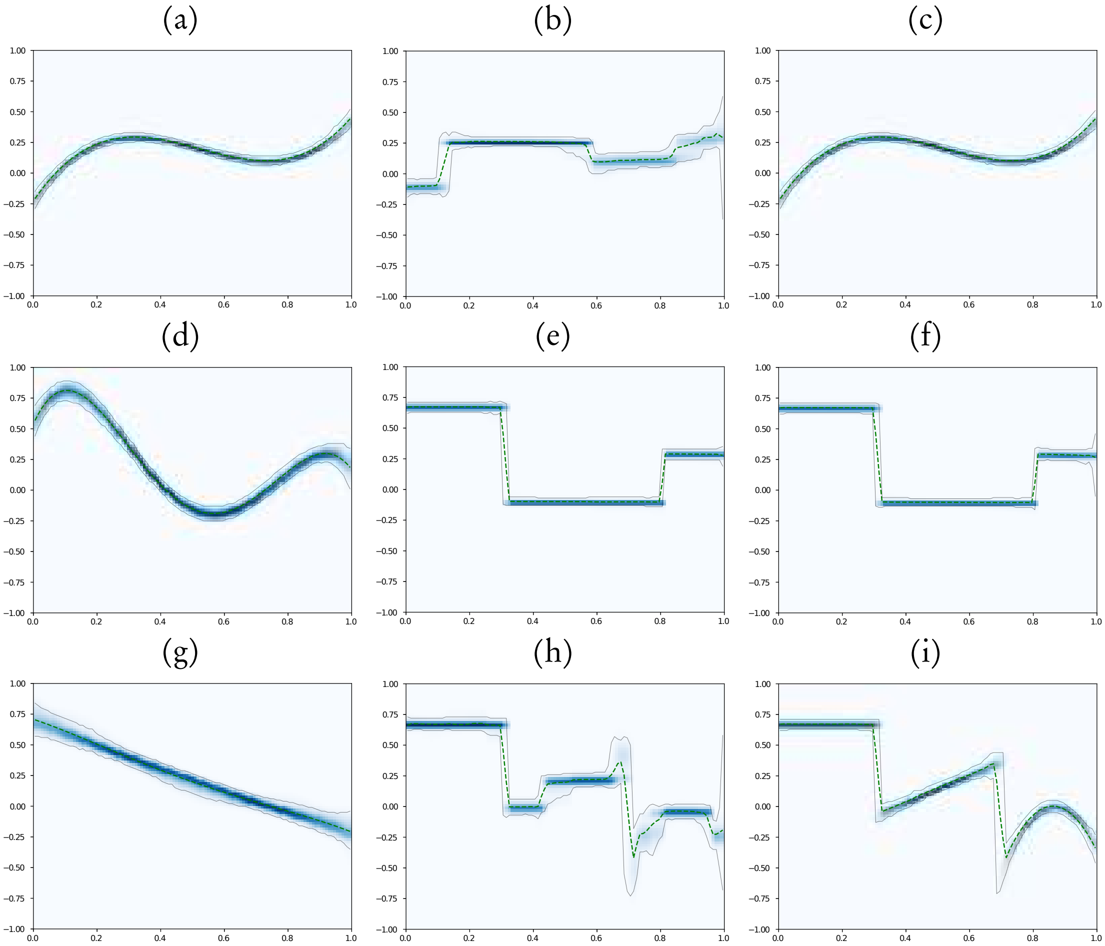


Figure 7.12: The true model and posterior histogram of different trans-dimensional inversion schemes for three synthetic regression experiments. In the first column is the result of a single partition trans-dimensional order inversion. The second column is a trans-dimensional partition model with a fixed zeroth order polynomial in each partition and in the third column is the trans-dimensional partition with trans-dimensional polynomials in each partition. Blue shading shows the ensemble histogram of the curve location, the green dashed line is the ensemble mean, and the faint black lines show the 95% credible intervals.

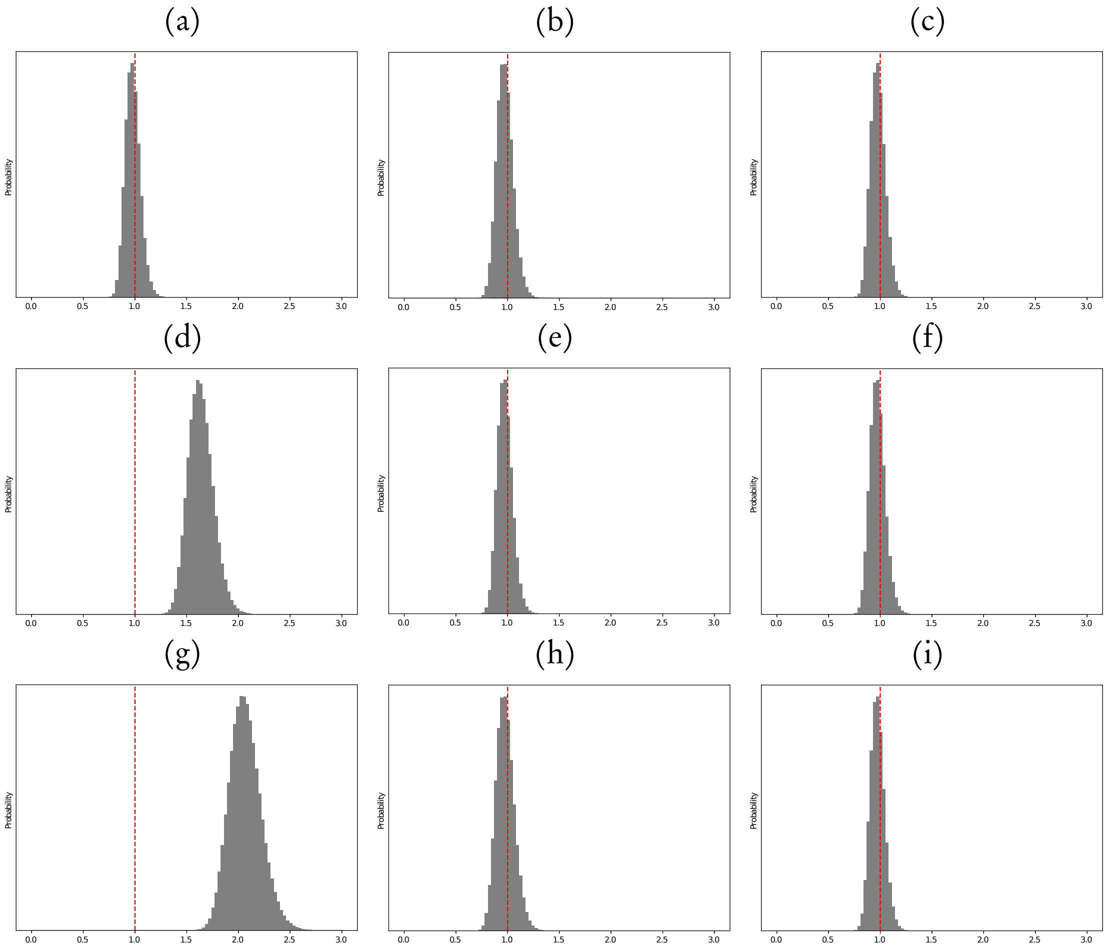


Figure 7.13: This figure shows the hierarchical error scaling histograms for each of the inversions. The first column is the inversions using a single partition with trans-dimensional change of order, the second column is the inversions using trans-dimensional partitions with fixed 0th order polynomials, and the last column using trans-dimensional partitions with trans-dimensional order polynomials.

actually present in the data, for example as in (h). It is only with the more flexible GLL scheme with trans-dimensional partitions and polynomial order where both situations can be resolved correctly.

The histograms for the estimates of the hierarchical scaling factor also show interesting features. In these synthetic regression examples, the true noise is known and the hierarchical scaling parameter inverted for is a multiplier of this true noise level. An indication of a successfully estimated noise level is a histogram centred on unity. In Figure 7.13 the histograms are plotted in the same grid layout as used in Figure 7.12.

From the figure, the histograms are generally well centred about unity with the exception of two of the single partition trans-dimensional order inversion in (d) and (g). It is somewhat surprising that the trans-dimensional partitions with zeroth order polynomials have converged so well to near unity for true models poorly represented by step functions, that is the smooth cubic in Figure 7.13(b) and the three partition model with varying order polynomials in Figure 7.13(h). It is generally expected that if the parameterization is a poor predictor of the observations, then hierarchical error estimates are generally larger but this does not seem to have occured in these cases.

As a final comparison, the posterior histogram of the location of partition boundaries is examined for the two methods with trans-dimensional partitions. This is shown in Figure 7.14 where the first column represents the results of the trans-dimensional partition with zeroth order polynomials (corresponding to (b), (e), and (h) in Figure 7.12). The second column shows the trans-dimensional partitions with trans-dimensional order inversions (corresponding to (c), (f), and (i) in Figure 7.12).

In Figure 7.14(a), in order to represent the smooth cubic function, the trans-dimensional partition model with zeroth order polynomials has created many artificial boundaries in an attempt to fit the smooth function. In contrast, the posterior in (b) correctly shows that no discontinuities are required by the data. In (c) and (d) the results are almost identical with two strong partition locations at the true partition boundary locations indicated with vertical red dashed lines. In the last row, once again due to higher order functions in the true model, the trans-dimensional partition solution with fixed zeroth order polynomials introduces many spurious boundary locations. Conversely, when trans-dimensional polynomial order is enabled in (f), the posterior recovers highly probable partition boundaries representative of the known truth.

A common application of trans-dimensional partition modelling is to detect discontinuities in data series[Green, 1995, Denison et al., 2002]. However here it was demonstrated that spurious discontinuities are possible if only zeroth order polynomials are used within partitions. Partition modelling with zeroth order polynomials is the most common case in existing implementations used in geophysical applications[Piana

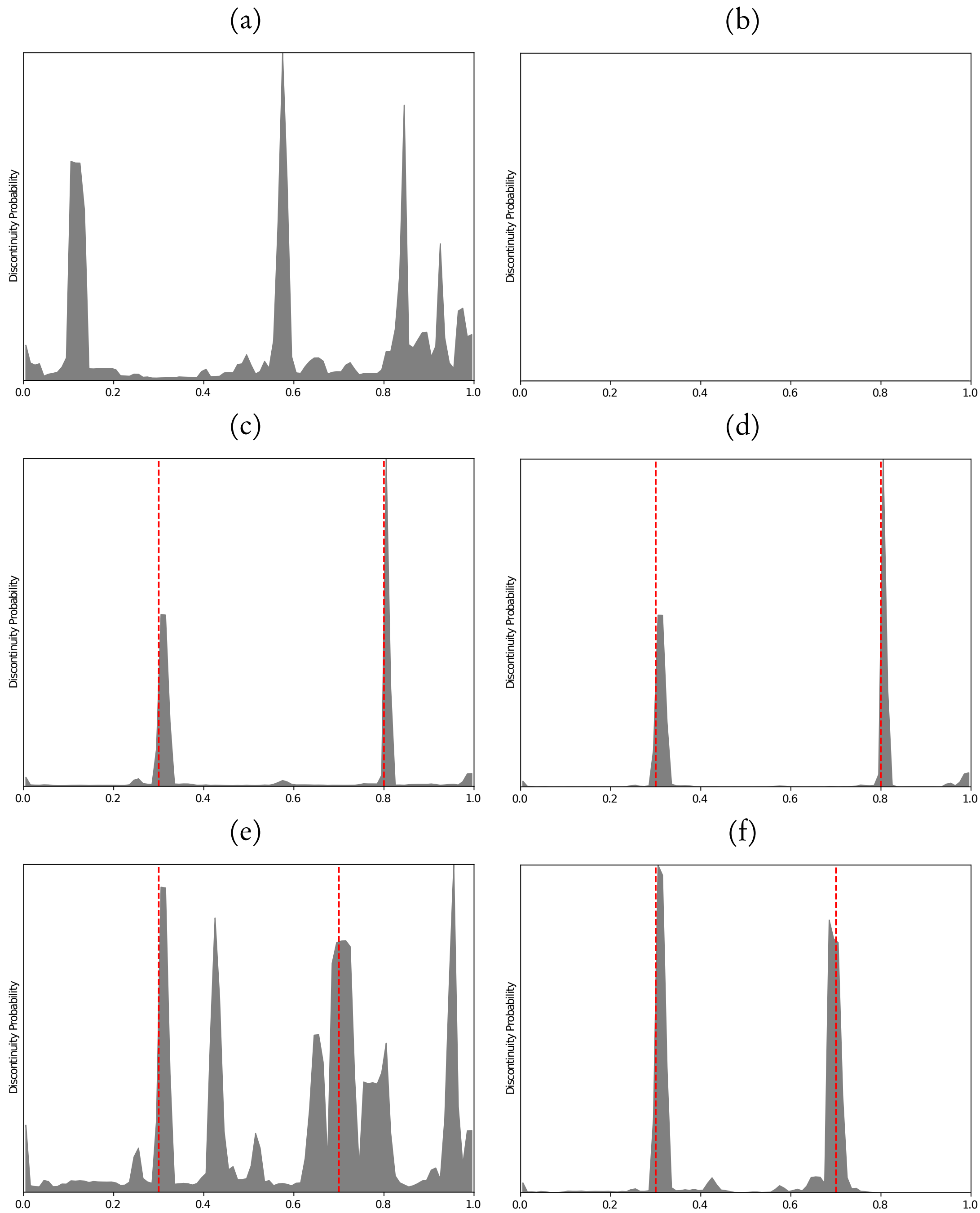


Figure 7.14: This figure shows the histogram of the partition boundaries. In the left column are the results for the trans-dimensional partition inversions with fixed zeroth order polynomials and in the right the results for the trans-dimensional partitions with trans-dimensional order polynomials method.

Agostinetti and Malinverno, 2010, Minsley, 2011, Dettmer and Dosso, 2012, Bodin et al., 2012b, Brodie and Sambridge, 2012]. The trans-dimensional scheme presented here using both partitions and polynomial order using GLL polynomials avoids artifacts of less flexible, fixed order approaches.

7.6 Summary

This chapter introduced a new general method of trans-dimensional partition modelling using Gauss-Lobatto-Legendre polynomials to allow an inversion to adapt to an unknown number of partitions with an unknown order polynomial within each partition.

The combination of partitions and variable order polynomials allows a trans-dimensional inversion to adapt a model to a smoothly varying representation using higher order polynomials, or to a set of discontinuities using partitions. From the results in synthetic tests it was shown that incorporating both trans-dimensional polynomial order and trans-dimensional partitions enables better recovery of the true model, better estimation of noise levels and reduction in spurious detection of discontinuities within the data.

In the following chapter, trans-dimensional spectral elements will be coupled with a non-linear forward model for the solution of a common and difficult geophysical inverse problem.

Chapter 8

Surface wave Inversions

8.1 Introduction

In the previous chapter, a novel trans-dimensional spectral element parameterisation was introduced that allows inversion for both smoothly varying models and models with discontinuities. Its effectiveness in three synthetic regression problems was demonstrated, and due to the parsimonious nature of trans-dimensional sampling, there was not over fitting of the true model.

In this chapter the aim is to apply this trans-dimensional framework to a geophysical problem. The motivation is the common problem in geophysics inversion that observations can often be explained equally well by 1D Earth models that are either smoothly varying, or a series of homogeneous layers with discontinuities. An early example discussing this trade off was described by Bullard et al. [1940] in a seismic refraction experiment (see section 4, particularly point (e)).

There are a number of methods for obtaining localised information of the Earth's structure using 1D models. An example are receiver functions [Langston, 1979, Owens et al., 1984] where the convolution of horizontal and vertical components of seismograms of an event, or series of events, are used to infer where interfaces have caused conversions from P-waves to S-waves. This method is sensitive to the interfaces but is dependent on a velocity model. There is a well established strong trade-off between the velocity model and the location of the interfaces as discussed by Piana Agostinetti and Malinverno [2010].

Another example, which was discussed in the ambient noise study of Chapter 5, is surface wave dispersion [Dettmer et al., 2012]. In this type of problem, an Earth model can predict the dispersion of phase or group velocity as a function of frequency. Surface wave dispersion inversion is highly non-unique as many Earth models can produce the same dispersion curves to within errors. In general, surface wave dispersion observations have low sensitivity to interfaces and is most sensitive to shear wave velocity as a function of depth (Rayleigh waves have some sensitivity to p-wave velocity). For this reason, the complementary sensitivities of receiver functions and surface wave dispersion measurements are often used in joint inversions for local 1D Earth models [Julia et al., 2000, Bodin et al., 2012b].

In surface wave dispersion problems, existing methods for forward modelling are restricted to matrix propagator methods [Thomson, 1950, Haskell, 1953] that can only model a series homogeneous layers. Rayleigh-Ritz methods [Wiggins, 1976] use smooth basis functions and therefore are not able on their own to properly model discontinuous Earth models. The numerical integration scheme of Takeuchi and Saito [1972] can incorporate both smoothly varying structure and discontinuities but is inefficient because it requires multiple numerical integrations to converge. Additionally this method also has numerical accuracy issues [Aki and Richards, 2002]. Another general class of method uses a linearisation from a reference model [Fang et al., 2015]. Typically this approach depends strongly on the choice of reference model and results in both simplistic uncertainties, and most importantly, an inability to properly solve the non-linear model choice problem that is addressed in this chapter.

In order to answer the question of whether the difference between layered structure and smoothly varying velocity can be discerned, an appropriate solution to the forward problem is first needed. This will be one that can use a 1D Earth model parameterised with the spectral element polynomials described in the previous chapter, and accurately compute predictions for dispersion.

8.2 Equations of Motion

This section gives a brief recapitulation of the elastic equations of motion for a body based on the linearisation or small deformation with a flat Earth approximation which is amalgamated from several sources [Love, 1927, Anderson, 1961, Takeuchi and Saito, 1972, Kennett, 1973, Achenbach, 1975, Thomsen, 1988, Chapman, 2004, Fichtner, 2011].

The equations of motion in a vector/tensor form are

$$\rho\ddot{\mathbf{u}} = \nabla \cdot \sigma + \mathbf{f}, \tag{8.1}$$

where ρ is the density, $\ddot{\mathbf{u}}$ is the second derivative of displacement with respect to time, σ is the stress, $\mathbf{f}$ are externally applied forces. For brevity, the dependence of ρ on a spatial coordinate, $\mathbf{x}$, and the dependence of $\mathbf{u}$, σ and $\mathbf{f}$ on the spatial coordinate and time t are omitted.

In a non-dissipative medium, the relationship between the stress and displacements is given by the tensor form of Hooke's law

$$\sigma = \mathbf{C} : \nabla\mathbf{u}, \tag{8.2}$$

where ":" is the tensor contraction operator, ∇ the spatial gradient operator, and $\mathbf{C}$ is the elastic tensor with elements c_{ijkl} which can vary spatially as a function of $\mathbf{x}$. Due to the required symmetry of the elastic tensor, the general case of an anisotropic material in matrix form can be expressed using Voigt notation. Given the 3 dimensional coordinate system where $\mathbf{x} = (x, y, z)$, the stress tensor is

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix}. \tag{8.3}$$

With Voigt notation, the ordering

$$xx \leftarrow 1 \tag{8.4}$$

$$yy \leftarrow 2 \tag{8.5}$$

$$zz \leftarrow 3 \tag{8.6}$$

$$yz \leftarrow 4 \tag{8.7}$$

$$xz \leftarrow 5 \tag{8.8}$$

$$xy \leftarrow 6, \tag{8.9}$$

is used to unwind the stress tensor into a vector

$$\sigma = \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{bmatrix}, \tag{8.10}$$

and with mappings from the terms of the elastic tensor c_{ijkl} to C_{mn} where $m \rightarrow ij$ and $n \rightarrow kl$. The constitutive relationship in matrix form can then be written

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{12} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66} \end{bmatrix} \begin{bmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ 2\epsilon_{yz} \\ 2\epsilon_{xz} \\ 2\epsilon_{xy} \end{bmatrix}, \tag{8.11}$$

where ϵ is the strain, related to the displacement by the template

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{du_i}{dj} + \frac{du_j}{di}\right), \tag{8.12}$$

for example,

$$\epsilon_{xy} = \frac{1}{2}\left(\frac{du_x}{dy} + \frac{du_y}{dx}\right), \tag{8.13}$$

and

$$\epsilon_{xx} = \frac{du_x}{dx}. \tag{8.14}$$

Since the interest here is 1D inversion it makes sense to at most consider transversely isotropic media where the axis of symmetry is defined to be the z, or the radial, direction. This equates to the case where there are different wave speed velocities laterally and azimuthally. In this case, using the terms introduced by Love [1927], the elastic tensor in matrix form becomes

$$\mathbf{C} = \begin{bmatrix} A & A-2N & F & 0 & 0 & 0 \\ A-2N & A & F & 0 & 0 & 0 \\ F & F & C & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 \\ 0 & 0 & 0 & 0 & L & 0 \\ 0 & 0 & 0 & 0 & 0 & N \end{bmatrix}. \tag{8.15}$$

These can be related to seismic wave speed parameters using

$$A = \qquad \rho\alpha_H^2 = \qquad \lambda_\perp + 2\mu_\perp \tag{8.16}$$

$$C = \qquad \rho\alpha_V^2 = \qquad \lambda_{||} + 2\mu_{||} \tag{8.17}$$

$$N = \qquad \rho\beta_H^2 = \qquad \mu_\perp \tag{8.18}$$

$$L = \qquad \rho\beta_V^2 = \qquad \mu_{||} \tag{8.19}$$

where α_H, α_V are the horizontal and vertical P-wave velocity respectively and β_H, β_V are the horizontal and vertical S-wave velocity respectively. Also shown above is the equivalence to the parameters used in Chapman [2004], $\lambda_\perp$, $\mu_\perp$, $\lambda_{||}$, $\mu_{||}$, and not shown, $F = \upsilon$. Finally, in the purely isotropic case these parameters can be reduced to

$$A = C = \lambda + 2\mu \tag{8.20}$$

$$N = L = \mu \tag{8.21}$$

$$F = \lambda, \tag{8.22}$$

where λ and μ are the Lamé elastic parameters.

Finally, restating the equation of motion in terms of vectors

$$\rho \frac{d^2}{dt^2} \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} = \begin{bmatrix} \frac{d}{dx} & \frac{d}{dy} & \frac{d}{dz} \end{bmatrix} \cdot \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}. \tag{8.23}$$

The individual components of the equations of motion can now be obtained in terms of the five transverse isotropic parameters.

$$
\begin{aligned}
\rho\frac{d^2u_x}{dt^2} &= \frac{d}{dx}\left[A\frac{du_x}{dx}+(A-2N)\frac{du_y}{dy}+F\frac{du_z}{dz}\right]+\\
&\quad \frac{d}{dy}\left[N\left(\frac{du_x}{dy}+\frac{du_y}{dx}\right)\right]+\\
&\quad \frac{d}{dz}\left[L\left(\frac{du_x}{dz}+\frac{du_z}{dx}\right)\right]+f_x
\end{aligned}
\tag{8.24}
$$

$$
\begin{aligned}
\rho\frac{d^2u_y}{dt^2} &= \frac{d}{dx}\left[N\left(\frac{du_x}{dy}+\frac{du_y}{dx}\right)\right]+\\
&\quad \frac{d}{dy}\left[(A-2N)\frac{du_x}{dx}+A\frac{du_y}{dy}+F\frac{du_z}{dz}\right]+\\
&\quad \frac{d}{dz}\left[L\left(\frac{du_y}{dz}+\frac{du_z}{dy}\right)\right]+f_y
\end{aligned}
\tag{8.25}
$$

$$
\begin{aligned}
\rho\frac{d^2u_z}{dt^2} &= \frac{d}{dx}\left[L\left(\frac{du_x}{dz}+\frac{du_z}{dx}\right)\right]+\\
&\quad \frac{d}{dy}\left[L\left(\frac{du_y}{dz}+\frac{du_z}{dy}\right)\right]+\\
&\quad \frac{d}{dz}\left[F\frac{du_x}{dx}+F\frac{du_y}{dy}+C\frac{du_z}{dz}\right]+f_z
\end{aligned}
\tag{8.26}
$$

These three coupled equations represent the general equations of motion for a transversely isotropic material. This can be simplified further assuming locally smooth lateral heterogeneities so that the material parameters ρ, A, C, F, L, N depend only on the dimension z, to obtain

$$
\begin{aligned}
\rho\frac{d^2u_x}{dt^2} &= A\frac{d^2u_x}{dx^2}+(A-N)\frac{d^2u_y}{dxdy}+F\frac{d^2u_z}{dxdz}+ \\
&\quad N\frac{d^2u_x}{dy^2}+N\frac{d^2u_x}{dxdy}+ \\
&\quad \frac{d}{dz}\left[L\left(\frac{du_x}{dz}+\frac{du_z}{dx}\right)\right]+f_x \qquad (8.27)\\
\rho\frac{d^2u_y}{dt^2} &= N\frac{d^2u_x}{dxdy}+N\frac{d^2u_y}{dx^2}+ \\
&\quad (A-2N)\frac{d^2u_x}{dxdy}+A\frac{d^2u_y}{dy^2}+F\frac{d^2u_z}{dydz}+ \\
&\quad \frac{d}{dz}\left[L\left(\frac{du_y}{dz}+\frac{du_z}{dy}\right)\right]+f_y \qquad (8.28)\\
\rho\frac{d^2u_z}{dt^2} &= L\frac{d^2u_x}{dxdz}+L\frac{d^2u_z}{dx^2}+ \\
&\quad L\frac{d^2u_y}{dydz}+L\frac{d^2u_z}{dy^2}+ \\
&\quad \frac{d}{dz}\left[F\frac{du_x}{dx}+F\frac{du_y}{dy}+C\frac{du_z}{dz}\right]+f_z. \qquad (8.29)
\end{aligned}
$$

8.2.1 Love Waves

Consider now a 1D Earth model using the previous direction vectors in which the the propagation of a Love wave is in the x direction. Love waves in this configuration oscillate laterally perpendicular to the propagation direction, that is, the y direction. The oscillatory displacement for a given frequency ω and wave number k in each direction can be written

$$
\begin{aligned}
u_x(t) &= 0, \qquad (8.30)\\
u_y(t) &= V(k,\omega,z)\exp[i(kx-\omega t)], \qquad (8.31)\\
u_z(t) &= 0, \qquad (8.32)
\end{aligned}
$$

where V is the depth dependent amplitude of the oscillations. The non-zero stresses associated with this displacement are

$$\tau_{yz} = L\frac{dV(k,\omega,z)}{dz}\exp[i(kx-\omega t)] \tag{8.33}$$

$$\tau_{xy} = ikNV(k,\omega,z)\exp[i(kx-\omega t)]. \tag{8.34}$$

Since u_x, u_z are zero and u_y has no dependence on y, (8.24) and (8.26) are zero. For (8.25), using

$$\frac{d^2u_y}{dt^2} = -\omega^2 V(k,\omega,z)\exp[i(kx-\omega t)] \tag{8.35}$$

$$\frac{d^2u_y}{dx^2} = -k^2 V(k,\omega,z)\exp[i(kx-\omega t)] \tag{8.36}$$

$$\frac{du_y}{dz} = \frac{dV(k,\omega,z)}{dz}\exp[i(kx-\omega t)], \tag{8.37}$$

giving

$$-\omega^2 V\rho = -k^2 NV(k,\omega,z) + \frac{d}{dz}\left[L\frac{dV(k,\omega,z)}{dz}\right], \tag{8.38}$$

after imposing the source free condition, that is $f_y = 0$, and cancelling the oscillating exponential term from both sides.

From (8.38), it is evident that Love waves are only sensitive to the shear wave velocity through the N and L parameters which are related to the horizontal and vertical shear wave velocities respectively.

The solution of these equations require boundary conditions which are given by

$$\lim_{z\to\infty} V = 0 \tag{8.39}$$

$$\tau_{yz}(z_{\text{surface}}) = 0 \tag{8.40}$$

$$. \tag{8.41}$$

These conditions are required by the "source-free" condition where there are no body forces. This condition requires that no source exists at infinity and that the traction vanishes at the surface. Hence τ_{zz}, τ_{xz}, and τ_{yz} must be zero at the surface of which only τ_{yz} is non-zero for Love waves. For a non-zero L parameter, this condition is satisfied by

$$\frac{dV}{dz}(z_{\text{surface}}) = 0. \tag{8.42}$$

Additionally, the displacement V and traction τ_{yz} must be continuous functions of z. In an isotropic homogeneous half space, where ρ, N, L are constant for all z, with L equal to N, the differential equation simplifies to

$$-\omega^2 V\rho = -k^2 LV + L\frac{d^2 l}{dz^2}, \tag{8.43}$$

which has the well known solution

$$V(z) = C\exp\left[-\sqrt{k^2 - \frac{\omega^2\rho}{L}}z\right], \tag{8.44}$$

with C some arbitrary scaling term.

8.2.2 Rayleigh Waves

Similarly to the Love wave case, 1D equations for the oscillations of a propagating Rayleigh wave can be formulated. Rayleigh waves, when propagating in the x direction in the coordinate system, consist of an oscillation in the x direction and an out of phase oscillation in the z direction, resulting in elliptical particle motion in the xz plane. Again the oscillations for a given frequency ω and wave number k in each direction can be expressed independently as

$$\begin{aligned}
u_x(t) &= U(k,\omega,z)\exp[i(kx-\omega t)] && (8.45)\\
u_y(t) &= 0 && (8.46)\\
u_z(t) &= iW(k,\omega,z)\exp[i(kx-\omega t)], && (8.47)
\end{aligned}$$

where U is the depth dependent amplitude of oscillations in the x direction and W in the z. The non-zero stresses associated with this displacement are

$$\begin{aligned}
\tau_{zz} &= i\left(FkU(k,\omega,z)+C\frac{W(k,\omega,z)}{dz}\right)\exp[i(kx-\omega t)] && (8.48)\\
\tau_{xz} &= L\left(\frac{dU(k,\omega,z)}{dz}-kW(k,\omega,z)\right)\exp[i(kx-\omega t)]. && (8.49)
\end{aligned}$$

Since u_y is zero, both sides of (8.25) are zero leaving two equations. Using

$$\begin{aligned}
\frac{d^2u_x}{dt^2} &= -\omega^2U(k,\omega,z)\exp[i(kx-\omega t)] && (8.50)\\
\frac{d^2u_x}{dx^2} &= -k^2U(k,\omega,z)\exp[i(kx-\omega t)] && (8.51)\\
\frac{d^2u_z}{dxdz} &= -k\frac{dW(k,\omega,z)}{dz}\exp[i(kx-\omega t)] && (8.52)\\
\frac{du_x}{dz} &= \frac{U(k,\omega,z)}{dz}\exp[i(kx-\omega t)] && (8.53)\\
\frac{du_z}{dx} &= -kW(k,\omega,z)\exp[i(kx-\omega t)], && (8.54)
\end{aligned}$$

after cancelling oscillation terms, this simplifies (8.24) to

$$-\rho\omega^2 U = -k^2 AU - kF\frac{dW}{dz} + \frac{d}{dz}\left[L\left(\frac{dU}{dz} - kW\right)\right]. \tag{8.55}$$

For (8.26), using

$$\frac{d^2 u_t}{dt^2} = -i\omega^2 W(k,\omega,z)\exp[i(kx-\omega t)] \tag{8.56}$$

$$\frac{d^2 u_x}{dxdz} = ik\frac{dU(k,\omega,z)}{dz}\exp[i(kx-\omega t)] \tag{8.57}$$

$$\frac{d^2 u_z}{dx^2} = -ik^2 W(k,\omega,z)\exp[i(kx-\omega t)] \tag{8.58}$$

$$\frac{du_x}{dx} = ikU(k,\omega,z)\exp[i(kx-\omega t)] \tag{8.59}$$

$$\frac{du_z}{dz} = i\frac{dW(k,\omega,z)}{dz}\exp[i(kx-\omega t)] \tag{8.60}$$

results in

$$-\rho\omega^2 W = Lk\frac{dU}{dz} - Lk^2 W + \frac{d}{dz}\left[FkU + C\frac{dW}{dz}\right]. \tag{8.61}$$

The solution of these equations require boundary conditions which are given by

$$\lim_{z\to\infty} U = 0 \tag{8.62}$$

$$\lim_{z\to\infty} W = 0 \tag{8.63}$$

$$\tau_{zz}(z_{\text{surface}}) = 0 \tag{8.64}$$

$$\tau_{xz}(z_{\text{surface}}) = 0. \tag{8.65}$$

Additionally, for the same reasons as in the Love wave case, U, W, τ_{zz} and τ_{xz} must be continuous functions of z.

8.3 Forward modelling

The most common method for the solution of computing the dispersion curve from a model of the Earth's crust is the propagator matrix method due to Thomson [1950] and Haskell [1953]. In this approach, the Earth is modelled as a stack of thick blocks and within each block the elastic properties and density are constant. The last or deepest block is a half space model. The benefit of this approach is that it is relatively quick to solve for a dispersion curve with computational time a function of the number of layers. The drawback is that a smooth, continuously varying velocity function must be crudely approximated by a many layered Earth model.

Takeuchi and Saito [1972] proposed an alternate iterative scheme that begins from a rigid boundary, or predetermined starting condition at a great depth, and numerically integrating the displacement to the surface using a trial value for the wave number k. For any given value of k, the traction free surface condition may be violated, so an iterative search technique is used to find valid k values such that the surface traction is zero (or close to zero given numerical precision). As a result, this method allows Earth models to have more complex structure, but results in a more computationally expensive forward model and convergence can become difficult at higher frequencies.

A finite element scheme is known to be an effective method for the solution of Love and Rayleigh waves [Lysmer, 1970, Lysmer and Drake, 1972]. In this scheme, a number of cells down to a rigid basement are constructed from which a linear system of equations can be formulated into an Eigen value problem that can be solved with standard approaches. This Finite Element approach to surface waves is analogous to beam vibration problems from structural mechanics and the rigid base is generally only a problem for long period waves which have greater sensitivity at depth. This approach has recently been revisited by Haney and Douma [2011] for Love waves.

A similar approach to finite elements is the Rayleigh-Ritz method which was employed by Wiggins [1976] with smooth basis functions. A draw back of the choice of basis functions is that sharp discontinuities are poorly represented.

In recent years, the spectral element method (SEM) has gained traction in simulating full waveform propagation through 2D and 3D media [Komatitsch and Tromp, 1999, 2002a,b, Fichtner et al., 2009]. In particular, the method easily incorporates free-surface boundary condition and can accurately represent the propagation of surface waves. The spectral element method is arbitrary order, restricted only by computational limits, and can incorporate spatial discontinuities in Earth parameters. To this end, it seems a good fit for the forward modelling of 1D surface wave dispersion for Love and Rayleigh waves.

8.4 Spectral element solution

The development of a continuous Galerkin projection spectral element method for the calculation of Love and Rayleigh wave dispersion has not previously appeared in the literature and this is an original contribution of this thesis. The details of the derivation is given in Appendix A and follows a similar treatment to introductory texts [Pozrikidis, 2005, Kopriva, 2009]. The results of the derivation are summarised here.

8.4.1 Love waves

From the derivation of the spectral element method for Love waves which solve the differential equation in (8.38), a matrix equation of the form

$$\left[\omega^2\mathbf{A} - k^2\mathbf{B} - \mathbf{C}\right]\mathbf{V} = 0, \tag{8.66}$$

is obtained with **A** and **B** diagonal matrices, **C** a block diagonal matrix, and **V** a vector representing the values of the Eigen function at the nodal interpolation values. Here the diagonallity of **B** means its inverse can be trivially computed allowing rearrangement to

$$\left[\mathbf{D} - k^2\mathbf{I}\right]\mathbf{V} = 0, \tag{8.67}$$

where $\mathbf{D} = \mathbf{B}^{-1}\left[\omega^2\mathbf{A} - \mathbf{C}\right]$. For non-trivial solutions of $\mathbf{V}$, it is required that

$$\det\left|\mathbf{D} - k^2\mathbf{I}\right| = 0, \tag{8.68}$$

and this can be efficiently solved using standard Eigen value calculation codes [Moler and Stewart, 1973]. The Eigen values of the system can be complex where the imaginary part is a decay term resulting in a Love wave that does not propagate. For negative real Eigen values, the wave number is purely imaginary and the wave does not propagate. Therefore, the positive Eigen values are the only propagating modes, and these are called static instabilities in the vibrational analysis community [Chang et al., 2010]. Lysmer and Drake [1972] gives a more detailed discussion of the Eigen values and their meaning than the summary presented here.

So for a given Earth model and frequency, it is possible to assemble the Eigen problem in (8.68) and solve for the real positive Eigen values. Each of these represent a surface wave mode whose wave number is the square root of the Eigen value from which the phase velocity can be computed. Additionally the Eigen vectors are the displacement function of the oscillations. In summary, this derivation forms the basis for a novel spectral element approach to computing dispersion information for Love wave propagation from arbitrary 1D Earth models.

8.4.2 Rayleigh waves

From the derivation of the spectral element method for Rayleigh waves as solutions of the coupled equations (8.55) and (8.61), the matrix equation

$$\left\{\omega^2\begin{bmatrix} A_x & 0 \\ 0 & A_z \end{bmatrix} + k^2\begin{bmatrix} B_x & 0 \\ 0 & B_z \end{bmatrix} + k\begin{bmatrix} 0 & C_x \\ C_z & 0 \end{bmatrix} + \begin{bmatrix} D_x & 0 \\ 0 & D_z \end{bmatrix}\right\}\begin{bmatrix} \mathbf{r}_x \\ \mathbf{r}_z \end{bmatrix} = 0, \quad (8.69)$$

can be constructed where the vectors $\mathbf{r}_x$ and $\mathbf{r}_z$ represent the eigen functions of the horizontal and vertical oscillations. Note that the coupling between the two systems is confined to factors of k. For non-trivial solutions, it is required that

$$\det\left|\omega^2\mathbf{A} + k^2\mathbf{B} + k\mathbf{C} + \mathbf{D}\right| = 0, \quad (8.70)$$

which is a quadratic Eigen value problem [Tisseur and Meerbergen, 2001]. The approach to solving this problem is a two step process of scaling the problem for numerical accuracy and restating the equation in one of the many companion forms which reduce the problem to a general Eigen problem. The details are in Hammarling et al. [2013], but the synopsis is that first scaling terms are computed using

$$\gamma = \sqrt{\frac{||\omega^2\mathbf{A} + \mathbf{D}||_2}{||\mathbf{B}||_2}} \quad (8.71)$$

$$\delta = \frac{2}{||\omega^2\mathbf{A} + \mathbf{D}||_2 + \gamma||\mathbf{C}||_2}, \quad (8.72)$$

and then solve the second companion form of (8.70) resulting in the general Eigen system

$$\det\left|\begin{bmatrix} \gamma\delta\mathbf{C} & -\mathbf{I} \\ \delta(\omega^2\mathbf{A} + \mathbf{D}) & \end{bmatrix} - \lambda\begin{bmatrix} -\gamma^2\delta\mathbf{B} & 0 \\ 0 & -\mathbf{I} \end{bmatrix}\right| = 0. \quad (8.73)$$

This solution will give Eigen values that need to be scaled to obtain the wave number, that is, for each Eigen value λ, the wave number k is given by

$$k = \gamma \lambda. \tag{8.74}$$

If the Earth model is parameterised in terms of N spectral element nodes, the Eigen vectors of this solution will be of dimension of $4N$. The first N elements of the Eigen vector give the amplitude function for the lateral displacement and the next N are for the radial displacement. In summary this derivation forms the basis for a novel spectral element approach to computing dispersion information for Rayleigh wave propagation from arbitrary 1D Earth models.

8.4.3 Boundary condition at depth

The discussion of how to treat the boundary of the solution at depth was deferred until now. The boundary condition of the differential equations, both for Love and Rayleigh waves, is that the displacement function(s) decay to zero as depth tends to infinity. In previous methods [Lysmer, 1970, Lysmer and Drake, 1972, Haney and Douma, 2011] for computing phase velocities from a 1D or 2D Earth model, a rigid boundary at a sufficiently large depth was applied as an approximation to the infinite boundary condition. This approach is a reasonable approximation if the depth of the model is sufficient that for all frequencies and modes considered, as the eigen functions are near zero at the basement.

In the spectral element field, there have been various approaches to providing robust solutions to half space problems using either a mapping from a finite to a half space, for example $-1 \ldots 1$ to $0 \ldots \infty$ [Boyd, 1987, 2001]. The solution can then proceed with exactly as described in Appendix A with the exception that the last element has different partial derivative terms due to the modified affine transformation, that is, the same GLL polynomials and quadrature rules are used. This type of mapping technique has very recently been applied to full waveform spectral element codes for representing the gravity potential [Gharti and Tromp, 2017] to enable larger periods to be modelled.

An alternative approach to the "finite to infinite" domain transformation technique is the use of either Laguerre polynomials or Laguerre functions introduced by Mavriplis [1989]. Laguerre polynomials and functions can be seamlessly incorporated into an Gauss-Lobatto-Legendre framework with the coupling of a finite domain to an infinite one [Shen, 2000, Valenciano and Chaplain, 2005]. In a comparison of the mapping approach to Laguerre polynomials, Black [1998] found that the relative merits of one over the other was problem dependent.

Laguerre polynomials are defined recursively as

$$\begin{aligned} L_0(x) &= 1 \\ L_1(x) &= 1 - x \\ L_n(x) &= \frac{2n-1-x}{n} L_{n-1}(x) - \frac{n-1}{n} L_{n-2}(x), \end{aligned} \tag{8.75}$$

and Laguerre functions

$$\hat{L}_n(x) = L(x) \exp\left\{\frac{-x}{2}\right\}. \tag{8.76}$$

The attraction of the Laguerre functions is that their formulation in a spectral element solution is exactly the same as that for Gauss-Lobatto-Legendre polynomials, with only the quadrature weights and differential operators differing. This means an infinite domain element can be readily incorporated using the existing derivation in Appendix A that correctly models a decay to zero of the amplitude functions with depth. In such a formulation, the Earth model will consist of some number of Gauss-Lobatto-Legendre cells from the surface to the half space layer, then one Gauss-Laguerre-Legendre cell to model the half space. In Figure 8.1, in (a) the Gauss-Lobatto-Legendre orthogonal polynomials are shown and in (b) the corresponding Gauss-Laguerre-Legendre polynomials for the same order are shown where the same orthogonal properties are evident, except the the Laguerre polynomials decay to zero as x tends to positive infinity.

For finite spectral elements, a physical part of the domain is mapped to the $-1 \ldots 1$

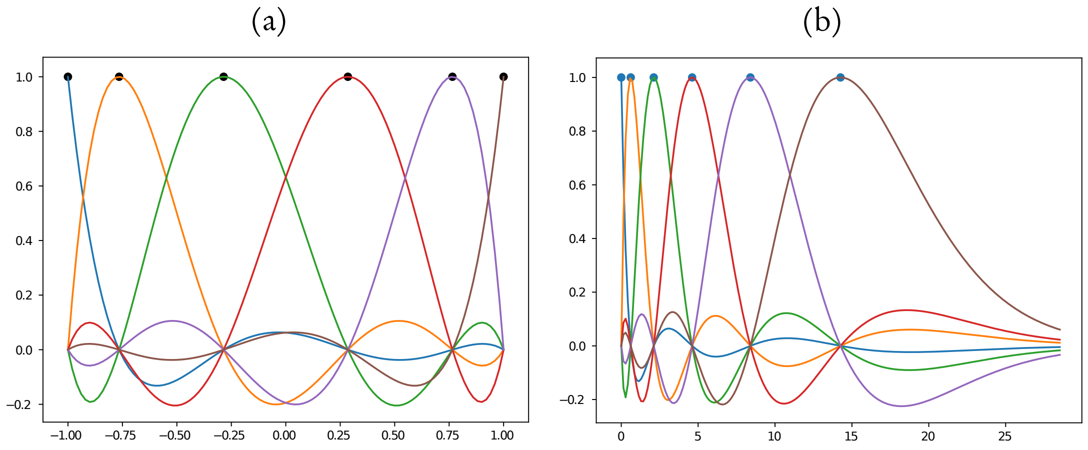


Figure 8.1: The similarity between the Gauss-Lobatto-Legendre (a) and Gauss-Laguerre-Legendre (b) for the 5th order set of cardinal polynomials is shown.

interval. For an infinite spectral element, a half space in the physical domain needs to be mapped from $z..\infty$ to $0..\infty$. The offset here is of little interest, however a scaling allows the Laguerre cell to optimally to fit the problem at hand. For Love wave dispersion, in a constant layer over a half-space, the well known analytical solution for the eigen functions is given by

$$V(z) \propto \exp\left[-\sqrt{k^2-\frac{\omega^2\rho}{L}}z\right]. \tag{8.77}$$

Hence when determining the transform from the physical domain to the spectral element domain, the transform

$$z(\zeta) = \frac{\zeta}{\sqrt{k^2-\frac{\omega^2\rho}{L}}}, \tag{8.78}$$

is optimal, and the Laguerre polynomials can perfectly fit the corresponding Eigen function. Unfortunately there is a circularity here in that in order to solve for k, k needs to be known for each frequency ω. The approach here is to iterate and use the result of a previous computation from which k is known to have changed little. Possibilities are to initially estimate k using a previous k from a neighbouring frequency when computing a dispersion curve.

	ρ (kg/m^3)	V_s (m/s)	V_p (m/s)
Layer	2,800	3,000	5,000
Half space	3,200	5,000	8,000

Table 8.1: The Earth model parameters for the calculation of Love wave phase velocity. The model is a simple homogeneous 10km thick layer over a half space.

8.4.4 Accuracy

In order to evaluate this new method of computing phase velocity from a 1D Earth model, in this section several comparisons are made between analytic results and solutions obtained using existing techniques. Arguably the most common approach used by practitioners for the calculation of phase velocity for a given Earth model and frequency is the Thomson-Haskell method using propagator matrices [Thomson, 1950, Haskell, 1953]. It is a fast, efficient and generally stable method although for Rayleigh waves the prior sampling of Chapter 2 did on occasion numerically fail for some velocity models.

For Love waves, the Earth model used is from Aki and Richards [2002] that consists of an homogeneous 10 km layer over a half-space with material properties shown in Table 8.1.

The phase velocity over a frequency range of near 0 to 1 Hertz was computed for the various methods. The comparison is performed between the results of a Thomson-Haskell calculation, a spectral element solution with a fixed boundary condition, a spectral element solution with a Laguerre boundary condition with fixed scaling, and a spectral element solution with a Laguerre boundary condition with automatic scaling. Each of these methods is compared to the known analytical solution derived in Aki and Richards [2002].

For the spectral element solution with a fixed boundary condition, the domain is extended to 100 km depth with additional cells to approximate the half-space. For the spectral element solution with a Laguerre half space, a fixed scaling of 1.0×10^{-4} was

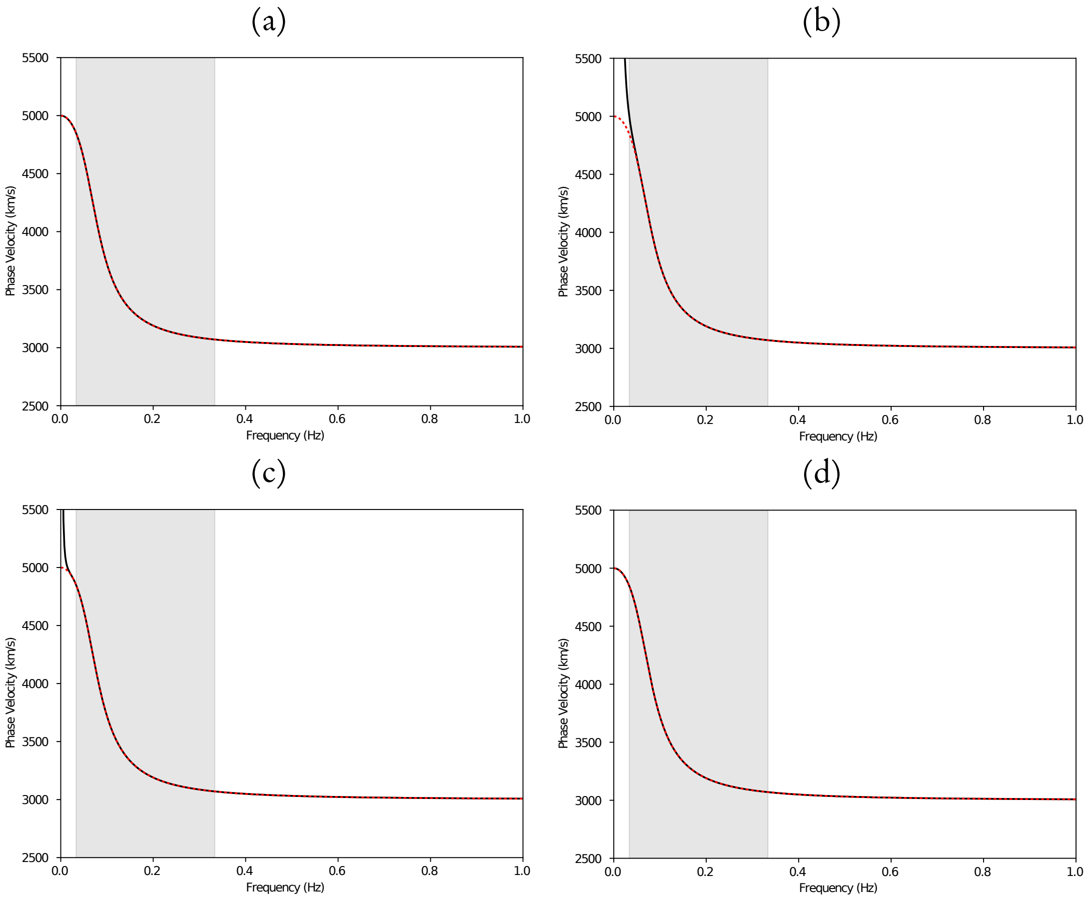


Figure 8.2: These plots show dispersion curves computed with different methods for the model from Aki and Richards [2002]. The methods are (a) the Thomson-Haskell propagator matrix method , (b) the spectral element method with fixed boundary , (c) the spectral element method with a Laguerre boundary with fixed scale, and (d) the spectral element method with automatic scaling. In each plot the computed dispersion curve is shown in solid black and the analytical solution with a red dotted line. The grey shaded region represents the frequency range generally of interest to ambient noise studies. While (b) and (c) have artefacts at low frequencies, (d) is accurate across all frequencies.

used and for the automatic scaling, the phase velocity is computed from high frequency (where incorrect scaling has little impact) to low and the Laguerre scaling term is updated using the wave number of the previous result. For all spectral element solutions, 5th order polynomials are used. The results of the calculations are plotted in Figure 8.2

It can be seen from the plots that in general the match between the analytical result and computed dispersion curve is visually identical with the exception of (b) and (c). In (b), as frequency approaches zero, there is divergence from the analytic result. This is because at low frequencies, the wave number becomes sensitive to the un-physical

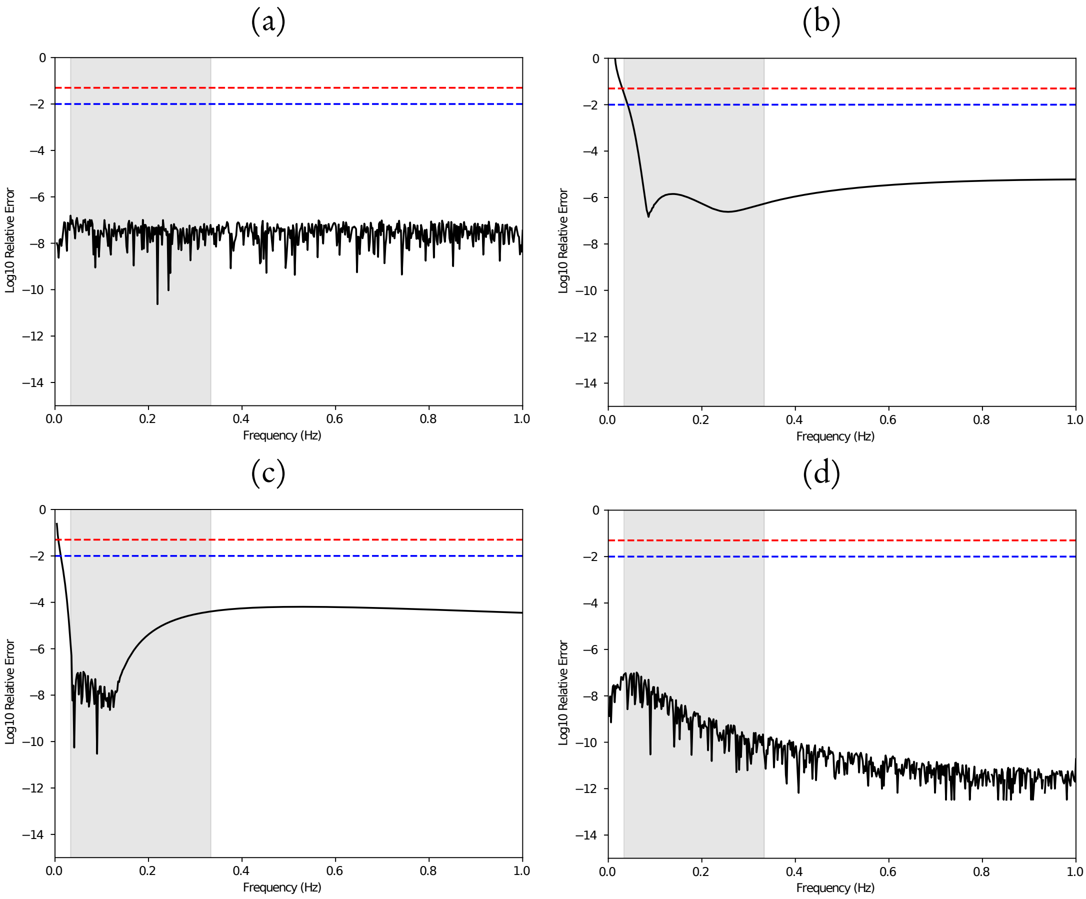


Figure 8.3: These plots show the $\log_{10}$ *relative error of the various methods tested. The methods are the Thomson-Haskell propagator matrix method (a), the spectral element method with fixed boundary (b), the spectral element method with a Laguerre boundary with fixed scale (c), and the spectral element method with automatic scaling (d). In each plot, the horizontal red and blue dashed lines represent 5 and 1 percent relative errors respectively. The grey shaded region represents the frequency range generally of interest to ambient noise studies.*

fixed boundary and causes the phase velocity to become unstable. In (c) there is a similar effect except that in this case it is poor scaling of the Laguerre spectral element representing the half-space. In (c), the Laguerre scale is tuned so that performance is good in the range of frequencies of interest to ambient noise studies. In (d), with the automatic updating of Laguerre scaling, the accuracy of the lower frequencies is far better.

In Figure 8.3, the $\log_{10}$ of the relative error compared to the analytic result for each of the calculations is shown. In each of the plots, the horizontal blue dashed line represents a one percent relative error and a reasonable threshold for accuracy. In

	ρ (kg/m^3)	V_s (m/s)	V_p (m/s)
Layer 1	2,800	3,000	5,100
Layer 2	3,000	4,000	6,800
Half space	3,200	5,000	8,500

Table 8.2: The Earth model parameters for the calculation of Rayleigh wave phase velocity. The Earth model consists of two simple homogeneous 10km thick layers over a half space.

(b) the spectral element solution with a fixed boundary can satisfactorily compute the dispersion across the range of frequencies of interest by sufficiently padding the model (down to 100km depth in this case). Similarly in (c), a fixed Laguerre scale can ensure accuracy over a frequency range of interest. However, the most accurate method in this case is the spectral element method with the automatic updating of the Laguerre scaling term in (d). It additionally out performs the Thomson-Haskell method at higher frequencies.

Similar tests are performed with the Rayleigh wave solvers for a two layer model over a half-space with material properties shown in Table 8.2. In this experiment, there is no analytic solution with which to compare and so here the Thomson-Haskell solution is adopted as the best estimate of the truth.

In Figure 8.4 the results of computing each of the spectral element solutions are shown. In (a) is the spectral element solution with a fixed boundary condition, (b) with a Laguerre boundary condition and fixed scale, and (c) the Laguerre boundary condition with automatic scaling. Across the range of frequencies of interest, the match is visually good with the exception of (a) at the low end of the grey shaded region.

The relative error plot, compared to a Thomson-Haskell solution, is shown in Figure 8.5. Both of the solutions using the Laguerre boundary element are well below the one percent error line across the range of frequencies highlighted. One disconcerting feature is that in this case the error increases as frequency increases. In this problem there are factors that are responsible: firstly the Rayleigh wave dispersion problem is known to be unstable at higher frequencies [Takeuchi and Saito, 1972], and secondly,

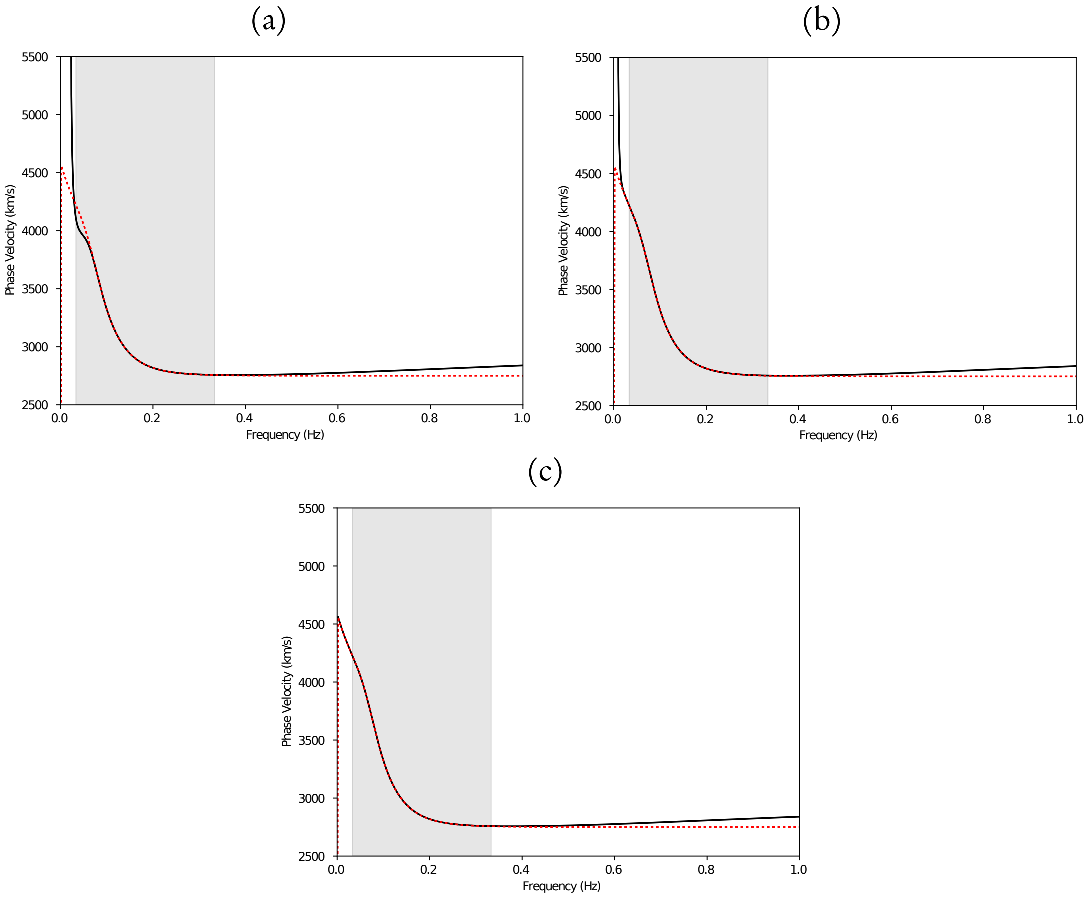


Figure 8.4: These plots show the dispersion curve computed with different methods. The methods are the spectral element method with fixed boundary (a), the spectral element method with a Laguerre boundary with fixed scale (b), and the spectral element method with automatic scaling (c). In each plot the computed dispersion curve is shown in solid black and the Thomson-Haskell solution with a red dotted line. The grey shaded region represents the frequency range generally of interest to ambient noise studies.

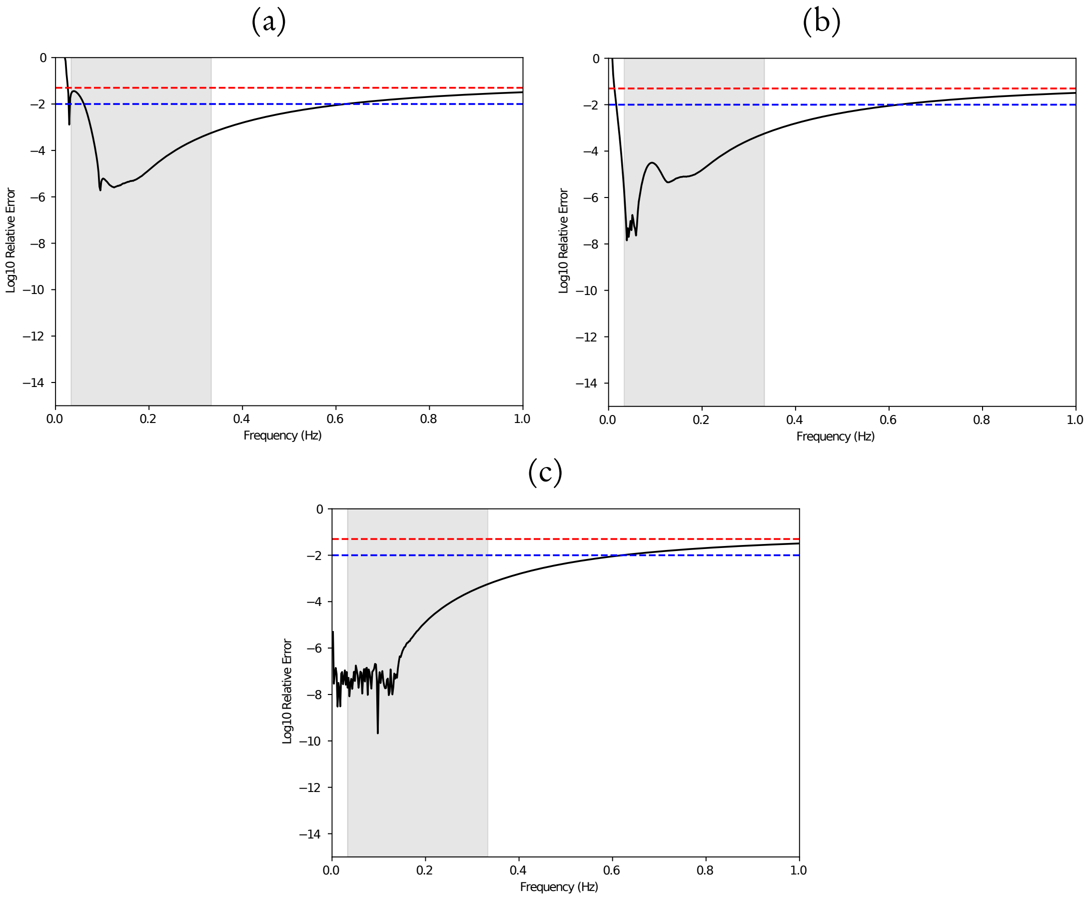


Figure 8.5: These plots show the $\log_{10}$ *of the relative errors with respect to the Thomson-Haskell solution for the different methods. The methods are the spectral element method with fixed boundary (a), the spectral element method with a Laguerre boundary with fixed scale (b), and the spectral element method with automatic scaling (c). The grey shaded region represents the frequency range generally of interest to ambient noise studies.*

the quadratic Eigen problem is numerically difficult.

If higher frequencies are of interest, then the options available are the typical options available to all spectral element solutions. Namely, the grid can be made finer or the order of the polynomial in each cell can be increased. These are known as h-refinement and p-refinement in the broader finite element research community. As an example, if the order of the polynomial is increased from 5 to 10, the results shown in Figure 8.6 are obtained where the Thomson-Haskell solution is matched across the entire frequency range. The downside of increasing the order is that this increases the size of the matrices in the generalised Eigen problem and therefore the computational cost

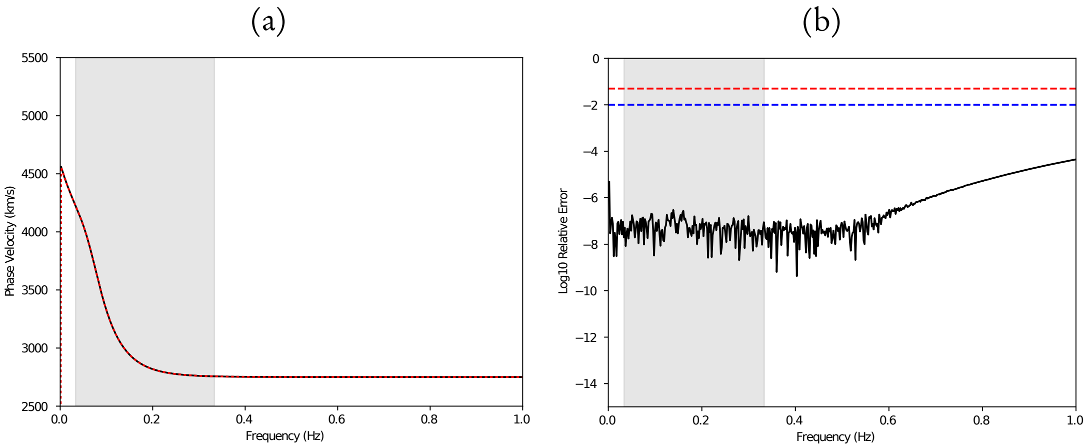


Figure 8.6: If the order of the polynomials is increased in the solution, the accuracy is improved at higher frequency. In (a) is the dispersion curve computed with higher order elements compared to the Thomson-Haskell method, and (b) shows the $\log_{10}$ *of the relative error.*

[Hammarling et al., 2013].

A final verification is to check that the boundary conditions of the differential equations are satisfied. Recall that for Love waves, the displacement and the traction, dictated by the leading term $L\frac{dl}{dz}$, must both be continuous functions of depth. Additionally the traction must be zero at the free surface and decay to zero at depth. In Figure 8.7 the displacement is shown in (a), and the traction shown in (b), for three representative periods normalised so the maximum amplitude is one. In this Figure, both displacement and traction are continuous with depth, and the traction is zero at the surface. In the Figure, the red dashed line indicates the half space boundary and the Laguerre element, which represents these functions below this point, accurately reflects an exponential decay. Careful inspection of the 10 second period curve in (a) reveals an extremely subtle oscillation which is a reflection of Laguerre scale mismatch. This could be remedied through increasing the order of the Laguerre function, or through further iterative refinement of the Laguerre scaling term, but this oscillation is negligible.

For Rayleigh waves, two tractions, τ_{xz} and τ_{zz} must be zero at the surface and decay to zero at depth. Similarly, the displacements must be continuous and decay to zero. In Figure 8.8 the horizontal displacement is shown in (a) and vertical displacement in

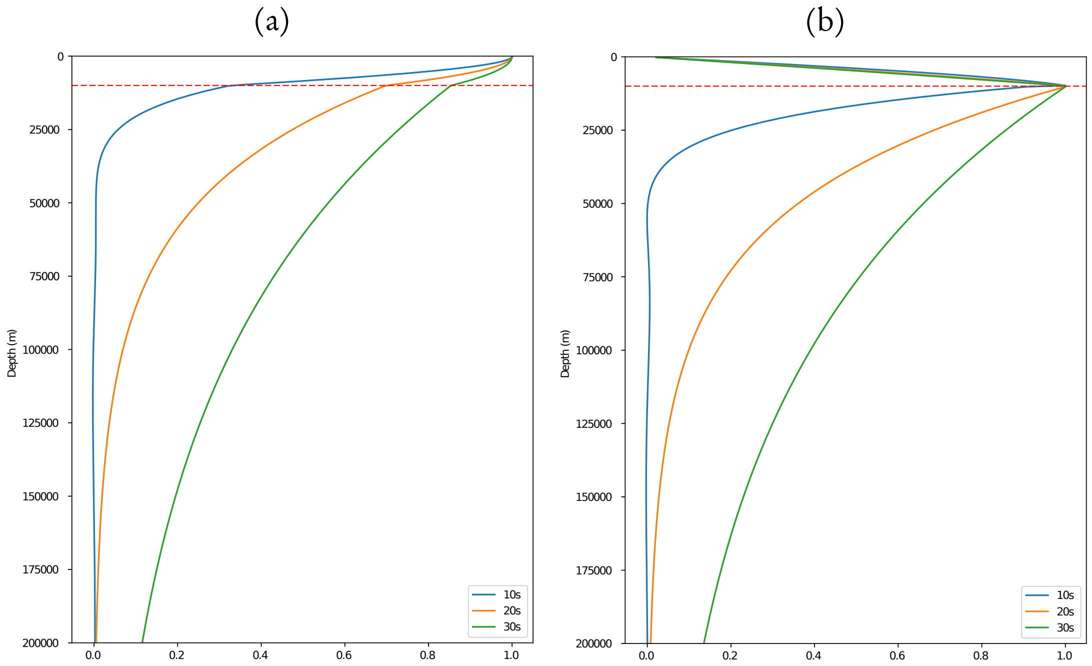


Figure 8.7: The normalised eigen functions for Love waves of selected periods for the synthetic model. (a) is the horizontal displacement, (b) is $L\frac{dl}{dz}$.

(b). The traction τ_{xz} is shown in (c) and τ_{zz} (d). In this model, there is an extra layer compared, and the first layer boundary is represented with a black dashed line and the half space boundary with a red dashed line. Visual inspection reveals that all boundary conditions and continuity requirements are satisfied for the Rayleigh wave tests.

8.4.5 Computational Time

A last comparison between the existing approach using the Thomson-Haskell method and the spectral element method developed here is the computational effort required. In Table 8.3, the relative time taken for the dispersion curve points computed in the accuracy test. In the table, the newly developed spectral element method is competitive with the Thomson-Haskell method. The spectral element method with Laguerre boundaries are marginally quicker than the fixed boundary case as extra elements need to be added to the model to approximate a half space and this padding increases the matrix dimension of the solution. The computational cost is dictated by the size of the matrix which in turn is controlled by the number of layers and the polynomial order

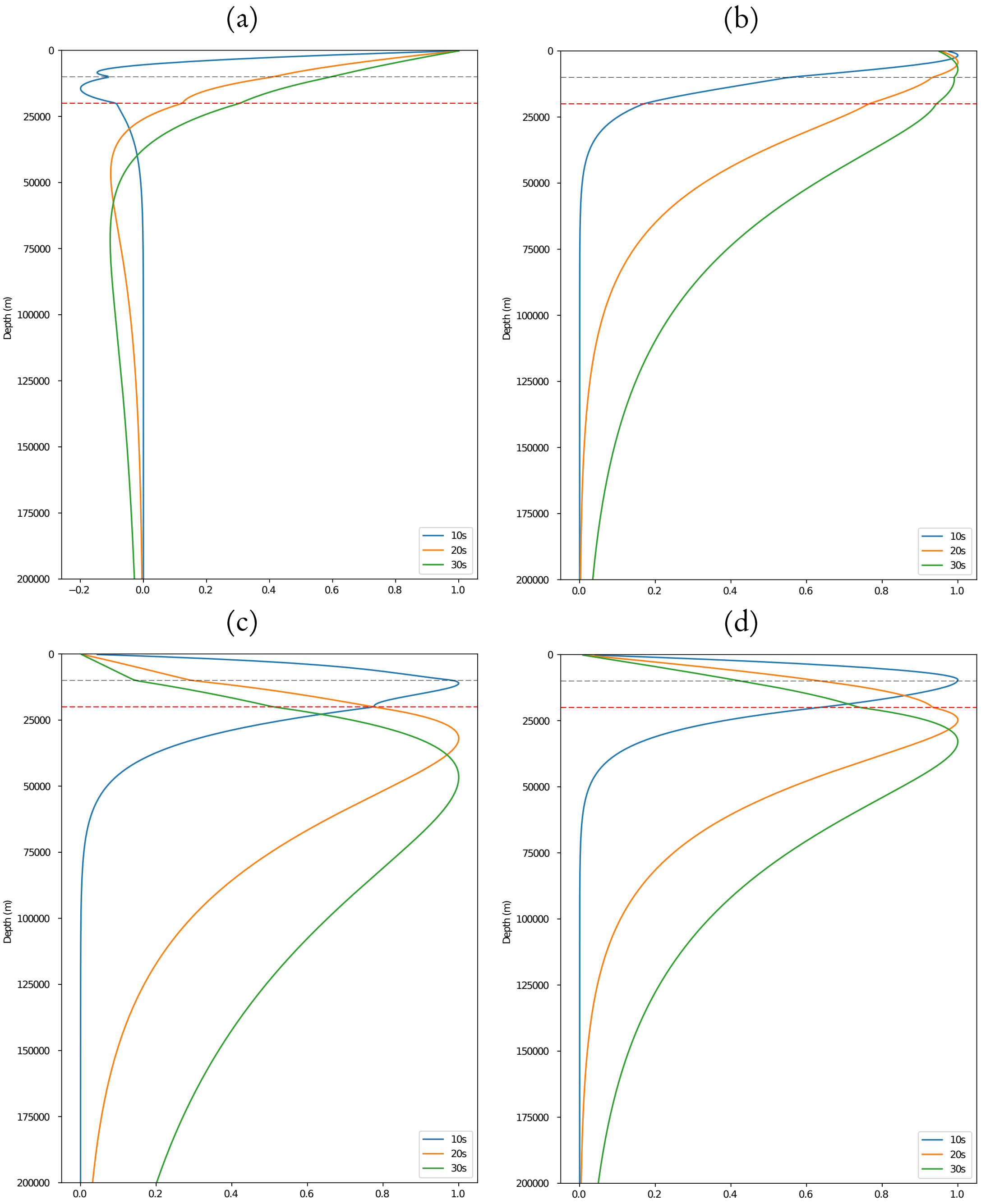


Figure 8.8: The normalised eigen functions for Rayleigh waves of selected periods for the synthetic model. (a) is the horizontal displacement, (b) is the vertical displacement, (c) is $L(\frac{dU}{dz} - kW)$ *and (d) is* $FkU + C\frac{dW}{dz}$*.*

Method	Time (seconds)
Analytic	0.010
Thomson-Haskell	0.150
Spectral Element (Fixed)	0.200
Spectral Element (Laguerre)	0.140
Spectral Element (Auto-Laguerre)	0.160

Table 8.3: A comparison of the computational time for the methods of inverting for Love wave dispersion. The times quoted represent the time to evaluate 512 different frequencies and as can be seen all non-analytic methods are competitive in terms of computational cost.

Method	Time (seconds)
Thomson-Haskell	0.53
Spectral Element (Fixed)	5.07
Spectral Element (Laguerre)	3.01
Spectral Element (Laguerre)	3.09

Table 8.4: A comparison of the computational time for the methods of inverting for Rayleigh wave dispersion. The times quoted represent the time to evaluate 512 different frequencies and as can be seen that for Rayleigh waves, the spectral elements are approximately six times slower.

used in the cells.

For Rayleigh waves a similar comparison between Thomson-Haskell and the spectral element method is shown in in Table 8.4. In this case there is a significant extra cost in the spectral element method, approximately a factor of six times slower. The primary cause of this is that the matrix necessary to solve is a factor of four larger than an equivalent Love wave spectral element solver. In the current implementation, the matrix equation is reconstructed for every calculation, so it is possible that this could be optimised. Nonetheless, for typical dispersion inversions involving Rayleigh waves, a forward model computation would cost less than a second and is feasible for an inversion algorithm.

8.4.6 Summary

In summary, an accurate and effective method of computing surface wave dispersion for both Love and Rayleigh waves using the spectral element method has been demonstrated. A novel aspect of this approach is the introduction of the Gauss-Laguerre-Legendre element for more faithfully representing the boundary condition at depth. Calculations of the new method have been compared to analytic results and existing approaches and showed a high level of accuracy, although particular care must be taken with Rayleigh wave dispersion at higher frequencies.

A key difference between this new method and existing approaches is that it can model smoothly varying Earth models with higher order polynomials and interface discontinuities. Previous matrix propagator techniques [Thomson, 1950, Haskell, 1953] can only model interface changes with homogeneous model structure whereas Rayleigh-Ritz techniques [Wiggins, 1976] can only model smoothly varying structure. Iterative numerical integration schemes are possible [Takeuchi and Saito, 1972, Fichtner and Igel, 2008] however these require multiple numerical integrations of two coupled equations for Love waves and four coupled equations for Rayleigh waves iterating using trial wave numbers. The reason that the number of equations double for the numerical integration scheme is that this is required to enforce continuity of the tractions τ_{xz}, τ_{yz} and τ_{zz} whereas in the spectral element method this requirement is enforced implicitly in the weak form of the equations, hence only a single solution of one or two coupled equations needs to be solved in a generalised Eigen problem for Love and Rayleigh waves respectively.

Many other geophysical forward models use homogeneous layer approximations and the general approach here could be equally applied to those problems. An example is the Airborne Electromagnetic forward model used in Chapter 4 which is similarly a propagator matrix solution to Maxwell's equations [Brodie and Sambridge, 2006, Brodie, 2010].

An important question now presents itself: when inverting for Earth models, can the difference between sharp contrasts and smoothly varying structure be resolved.

8.5 Trans-dimensional inversion

A spectral element method for solving for Love and Rayleigh phase velocity given a 1D Earth model parameterised as a series of elements with arbitrary polynomials in each element has been developed in the previous section. This can be used to represent smoothly varying structure with a higher order polynomial, or structural discontinuities with element boundaries.

Given this general forward model, in an inversion there are many options available. A possibility would be to compute gradients of a misfit function with respect to the model parameters through modelling of perturbations and the adjoint state method [Takeuchi and Saito, 1972, Plessix, 2006] as is commonly used in both Thomson-Haskell based inversions in surface wave dispersion problems and in full waveform tomographic problems using the spectral element method. However, the aim here is to explore the model choice problem between representing a 1D Earth model with a series of homogeneous layers, or a smaller number of layers with smoothly varying structure.

In order to test this, the trans-dimensional spectral elements inversion framework from the previous chapter is coupled to the spectral element forward model developed and verified in the previous sections. An issue here is that in some cases, particularly for Rayleigh wave calculations, the polynomial order must be reasonably high, for example, 5th order and perhaps higher for higher frequencies. In trans-dimensional inversion, the desire is to infer whether the observations are best predicted by some number of constant layers (zeroth order) or smoothly varying structure (higher order), and zeroth order may not provide sufficient forward model accuracy.

Fortunately, the orthogonality of the spectral polynomials can be used to ensure accuracy in the forward model when computing dispersion for lower order models, since any lower order polynomial is exactly represented by a higher order spectral element polynomial. An example is shown in Figure 8.9 where in (a) a simple quadratic function is shown with a solid black line with faint lines showing the cardinal functions. In (b) this quadratic can be exactly represented with a higher order polynomial simply

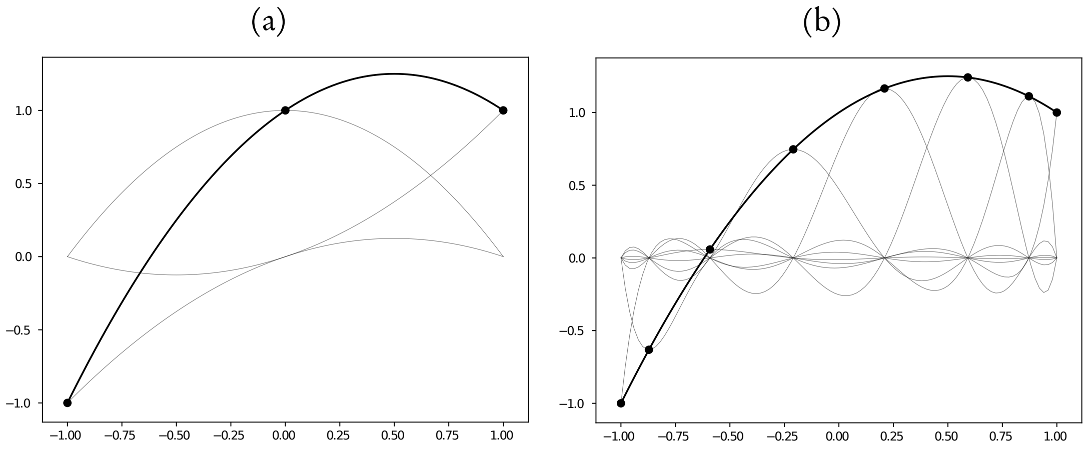


Figure 8.9: With orthogonal polynomials, lower order polynomials can be perfectly represent with higher order polynomials, for example, the quadratic function shown in (a), can be represented with higher order polynomials as shown in (b).

by interpolating the quadratic function at each of the higher order polynomials nodal points. This same property is already used in polynomial change of order proposals of the previous chapter. Here it is used to project low order models to higher order to ensure forward model accuracy, in a similar fashion to Chapter 5 where wavelet super resolution was used to improve the accuracy of the Fast Marching method.

The question that is a focal point of this chapter is under what conditions can surface wave dispersion observations be used to distinguish between homogeneous layers and slowly varying layers in a 1D Earth model? To attempt to answer this, a simple 1D Earth model has been constructed with a 2 km thick homogeneous layer, above an 18 km thick layer with a linear function over a homogeneous half space. The spectral element method developed in previous sections has been used to compute dispersion curves for both Love and Rayleigh wave dispersion across the range of frequencies used in the Iceland ambient noise study in Chapter 5. Independent Gaussian noise is added to each observation to create synthetic observations.

For this test, three inversion are performed with the same observations with different restrictions on the trans-dimensional inversion algorithm used. The three inversions all have trans-dimensional partitions but differ in the maximum order allowed within each partition. The first has fixed `0`th order polynomials and is therefore equivalent to existing techniques, the second can trans-dimensional select between 0th and 1st order polynomials in each partition, and the third allows up to quadratic polynomials.

For each of the inversions in the following sections, eight independent chains were simulated without parallel tempering. Each independent chain was initialised with a single zeroth order cell model with the velocity in the cell sampled from the prior and simulated for 1 million iterations. A hierarchical error scaling estimate of the true noise level within the observations is used, hence the hierarchical scaling term should converge to approximately one. The model is parameterised in terms of shear wave velocity at each nodal interpolation point and the parameters required in the forward model, namely ρ, A, C, F, L, and N, are computed using the assumption of isotropy and the empirical relationships between shear wave velocity, density and P-wave velocity of Brocher [2005].

A symmetric Dirichlet prior is used for layer thicknesses with a variable depth to the lower most half space. This allows the depth of the model to expand or contract to fit the observations [Steininger et al., 2013].

For the prior on the shear wave velocity, three separate priors are used to test the effect of the prior on the inversion. The prior is a depth dependent Beta prior specified by a reference model and a concentration parameter. The Beta distribution is given by

$$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \tag{8.79}$$

where α and β are shape parameters and x is between $0 \ldots 1$. Through a simple change of variable the Beta distribution can be defined for a range of shear wave velocities, in the case here, 1 to 6 km/s (the true model ranges from 2 to 5 km/s). In the priors used in this experiment, the α and β parameters are set as a function of depth using a reference model to defined the mode, then

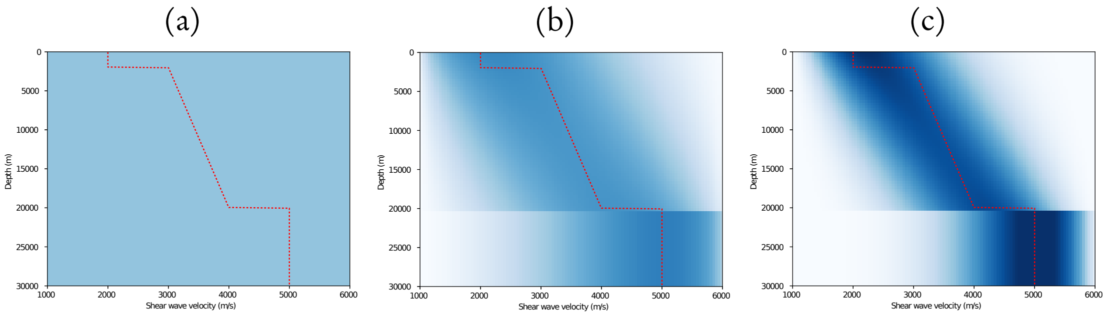


Figure 8.10: The three priors used in the synthetic inversion. In (a) is a uniform prior represent no prior information. In (b) is a Beta prior with the mode set using a reference model near the true model (red curve) and a concentration parameter of 4 to represent a weakly informative prior. Lastly in (c) is again a Beta prior this time with a concentration parameter of 8 for a strongly information prior.

$$\alpha(z) = M_o(z)(\varkappa - 2) + 1 \tag{8.80}$$

$$\beta(z) = (1 - M_o(z))(\varkappa - 2) + 1, \tag{8.81}$$

where $M_o(z)$ is the mode or reference model value at depth z and $\varkappa$ is the concentration parameter. A convenient feature of this formulation of the prior is that setting the concentration parameter to two results in a uniform prior between the range of shear wave velocities specified. For these experiments, three separate values of concentration parameter are used, 2 (giving a uniform prior), 4 and 8. These are shown as probability density plots in Figure 8.10. They aim to represent little prior information in (a), weak prior information in (b) and strong prior information in (c) for the experiments to follow.

For each forward model/observations problem, nine separate inversions are simulated for the tree trans-dimensional inversions and three priors. The results are presented in a three by three grid format with columns corresponding to the fixed zeroth order inversions, trans-dimensional zeroth and linear inversions, and trans-dimensional zeroth, linear and quadratic inversion respectively. The rows correspond to the uniform prior, Beta prior with $\varkappa = 4$ and Beta prior with $\varkappa = 8$ respectively. Only summary results are presented here, with more detailed plots shown in Appendix B.

8.5.1 Love wave results

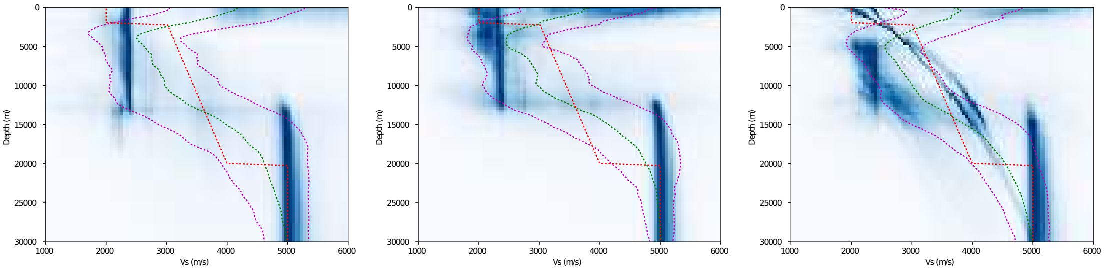
The results for the Love wave inversions are shown in Figure 8.11 as a posterior histogram of the ensemble with blue shading representing relative probability. The sensitivity of Love waves to structure is very broad resulting in poor depth resolution.

For each of the inversions, the results are poor in different ways. In the top row where the prior is uniform, there is a strong preference for fast velocity at the surface which is far from the true model. In a real world inversion, this range of velocities at the surface would be implausible. In the more informative priors in the 2nd and 3rd rows, this issue has been resolved. In (f) and (i), the results are better with the top of the Earth model being represented by a quadratic function that appears to average the low velocity layer of the true model and the top part of the next layer. In (c) and (h), the ensemble supports two linear layers which reasonably approximate the true model.

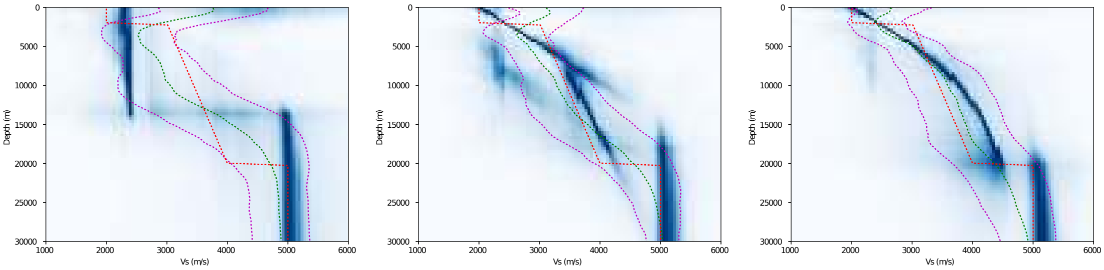
A common motivation for partition modelling in 1D geophysical problems is to estimate the location of interfaces at depth in a probabilistic sense using the histogram of partition boundaries as a proxy for likely interfaces. The low information content of the posterior is also evident in the posterior histograms of partition boundary locations shown in Figure 8.12 where there is no strong signal for either of the two true velocity contrast interfaces. The posterior histograms have preference for partition boundaries near the surface from which no strong inference as to the location of interfaces could be reliably made.

With these generally poor results, the fit to the data could be expected to be poor. These are shown in Appendix B in Figure B.2 where it can be seen that the fit to data in all inversions is to within noise level, and yet the Earth models are generally poor predictors of the true model. Similarly, the hierarchical noise scale is approximately one in Figure B.1 of the appendix indicating good convergence to the true noise level. This result gives an indication of the difficulty posed by this inversion problem, in that, many Earth models can equally well predict the observations to within noise.

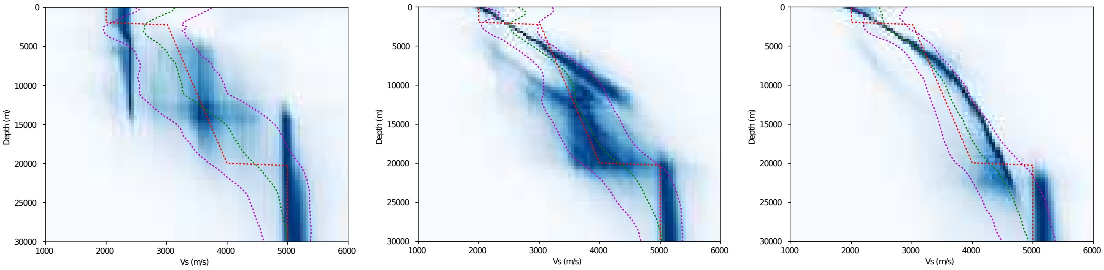
In summarising, with only Love wave dispersion information resolve ambiguities be-



(a) 0th/Uniform (b) 1st/Uniform (c) 2nd/Uniform



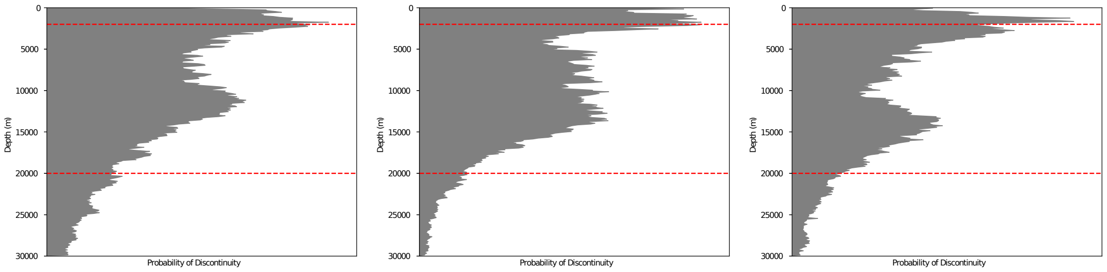
(d) 0th/Beta($\varkappa = 4$) (e) 1st/Beta($\varkappa = 4$) (f) 2nd/Beta($\varkappa = 4$)



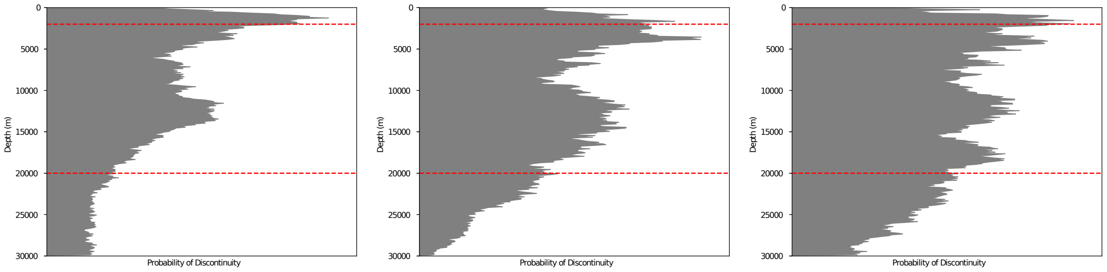
(g) 0th/Beta($\varkappa = 8$) (h) 1st/Beta($\varkappa = 8$) (i) 2nd/Beta($\varkappa = 8$)

Figure 8.11: Love wave inversion histograms with blue shading representing regions of higher probability, the green curve the mean of the ensemble, the red curve the true model and the purple curves showing the spread of the standard deviation from the mean. The first column represents trans-dimensional solutions with 0th order only, the second column trans-dimensionally selects between 0th and 1st order, and the third trans-dimensionally selects between 0th, 1st, and 2nd order polynomials. The first row uses a uniform prior between 1 and 6 km/s, the second row a Beta prior with the mode set from a reference model and a concentration parameter of 4 and the last row is again a Beta prior with a concentration parameter of 8.

(a) 0th/Uniform (b) 1st/Uniform (c) 2nd/Uniform



(d) 0th/Beta($\varkappa = 4$) (e) 1st/Beta($\varkappa = 4$) (f) 2nd/Beta($\varkappa = 4$)



(g) 0th/Beta($\varkappa = 8$) (h) 1st/Beta($\varkappa = 8$) (i) 2nd/Beta($\varkappa = 8$)

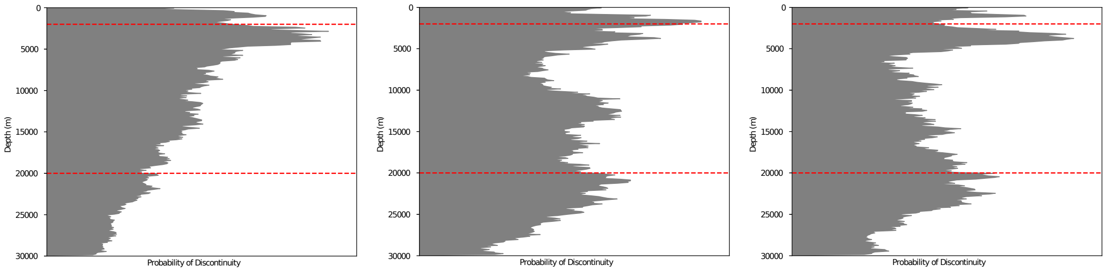


Figure 8.12: Love wave boundaries histogram showing the posterior location of partition boundaries in the ensemble. The first row is from inversions with a uniform prior, the second with a Beta prior with a concentration value of 4, the third with a concentration parameter of 8. The first column is an inversion with 0th order only, the second allows trans-dimensional selection between 0th and 1st order polynomials, and the third allows trans-dimensional selection between 0th, 1st, and 2nd order polynomials.

tween structure with this formulation of the problem. Even with stronger prior information, the mean of the ensemble converges close to the true model but regions of high likelihood in the ensemble posterior do not share similarities. The convergence of the mean for the more informative prior inversions, that is the Beta priors with concentration parameters of four and eight, is likely to be a reflection of the prior rather than the observations.

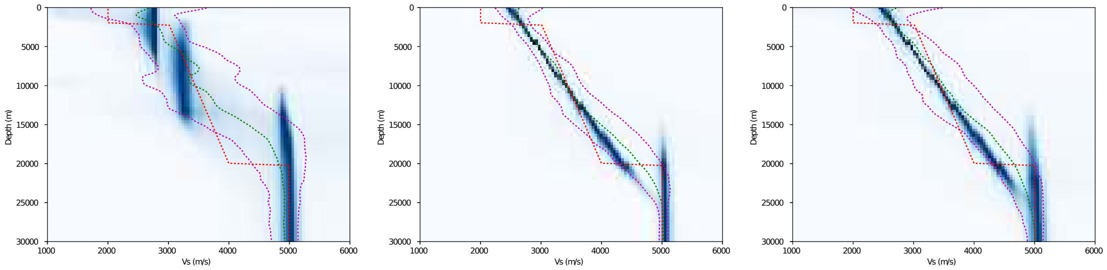
8.5.2 Rayleigh wave results

For the experiments with Rayleigh wave dispersion observations the results of the inversions are shown in Figure 8.13. The results show improvement compared to the Love wave inversions. This is likely due to the different depth sensitivities of Love and Rayleigh wave dispersion observations.

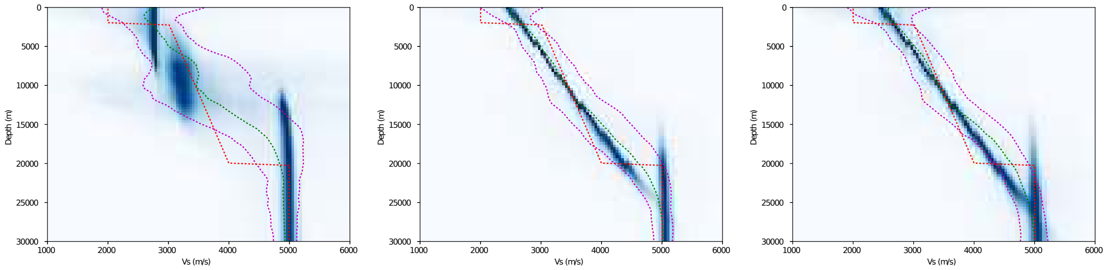
Here, unlike the Love wave inversion, the high probability regions of the histograms generally follows the true model. For the 0th order polynomials in the first column, the true model has been approximated with three homogeneous layers. In contrast, the inversions that allow trans-dimensional selection of the polynomial order, that is the second and third columns, have a strong preference for two layers: a linear gradient over a half-space. In the posterior, the slope of the first layer does not match that of the linear gradient in the true model's second layer. On closer inspection it is evident that the model that these inversions have preferred is averaging the shallow homogeneous first layer and the second layer with the linear gradient. Another interesting aspect in these results is that the three different priors has relatively little impact in the posterior.

The posterior histograms of the location of partition boundaries are shown in Figure 8.14. For the inversions, there does seem to be support for an interface around 2km deep as in the true model. In the zeroth order inversion, there is also less precise support for an interface at approximately 10km deep, however this is a spurious interface and caused by the parameterisations inability to model the linear gradient of the second layer in the true model. For the two inversions that allow trans-dimensional selection

(a) 0th/Uniform (b) 1st/Uniform (c) 2nd/Uniform



(d) 0th/Beta($\varkappa = 4$) (e) 0th/Beta($\varkappa = 4$) (f) 0th/Beta($\varkappa = 4$)



(g) 0th/Beta($\varkappa = 8$) (h) 1st/Beta($\varkappa = 8$) (i) 2nd/Beta($\varkappa = 8$)

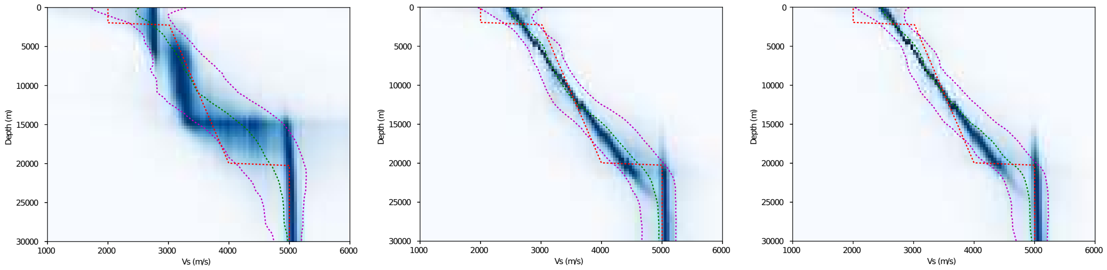
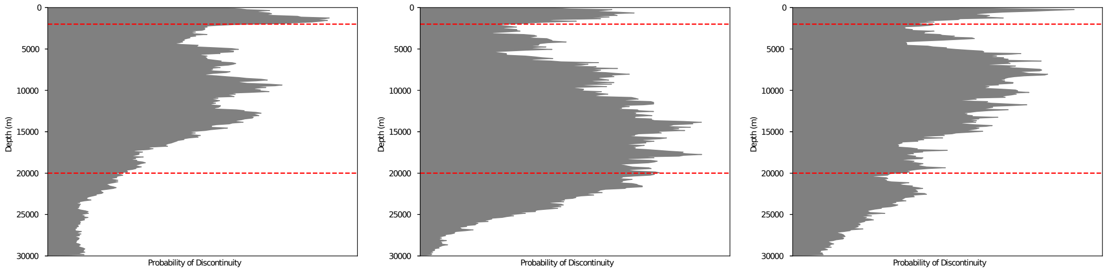
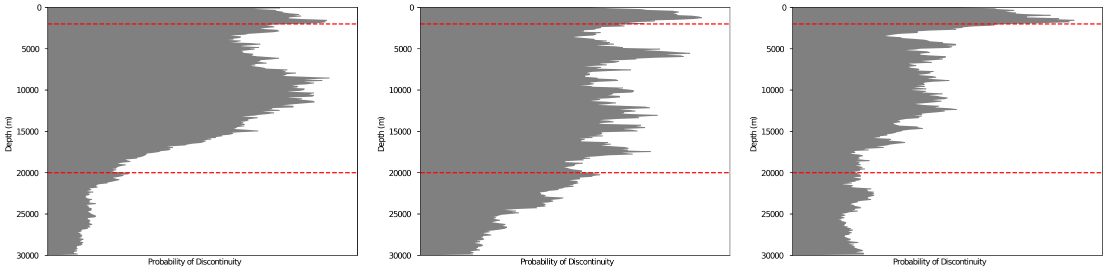


Figure 8.13: Rayleigh wave inversion histogram with blue shading representing regions of higher probability, the green curve the mean of the ensemble, the red curve the true model and the purple curves showing the spread of the standard deviation from the mean. The first column represents trans-dimensional solutions with 0th order only, the second column trans-dimensionally selects between 0th and 1st order, and the third trans-dimensionally selects between 0th, 1st, and 2nd order polynomials. The first row uses a uniform prior between 1 and 6 km/s, the second row a Beta prior with the mode set from a reference model and a concentration parameter of 4 and the last row is again a Beta prior with a concentration parameter of 8.

(a) 0th/Uniform (b) 1st/Uniform (c) 2nd/Uniform



(d) 0th/Beta($\varkappa = 4$) (e) 1st/Beta($\varkappa = 4$) (f) 2nd/Beta($\varkappa = 4$)



(g) 0th/Beta($\varkappa = 8$) (h) 1st/Beta($\varkappa = 8$) (i) 2nd/Beta($\varkappa = 8$)

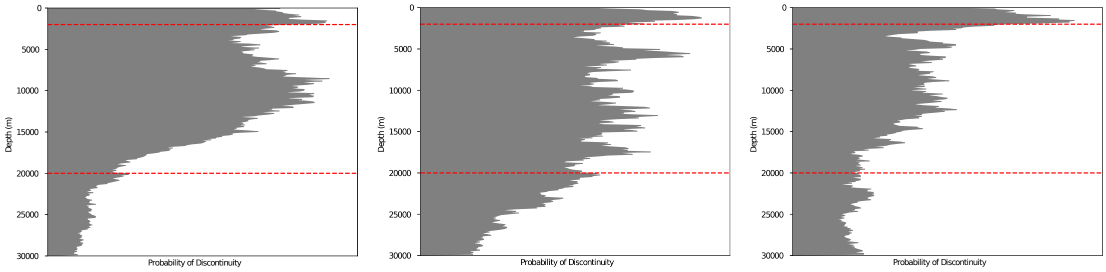


Figure 8.14: Rayleigh wave boundaries histogram showing the posterior location of partition boundaries in the ensemble. The first row is from inversions with a uniform prior, the second with a Beta prior with a concentration value of 4, the third with a concentration parameter of 8. The first column is an inversion with 0th order only, the second allows trans-dimensional selection between 0th and 1st order polynomials, and the third allows trans-dimensional selection between 0th, 1st, and 2nd order polynomials.

of the polynomial order, there is no strong indication of an interface at depth.

In summarising, with only Rayleigh wave dispersion information there is better resolvability of slowly varying structure with this formulation of the problem. Rayleigh waves have better depth sensitivity to depth that Love waves which can be observed in the Eigen function plots in Figures 8.7 and 8.8, and this in part explains the marginally better recovery in these results.

8.5.3 Joint Love and Rayleigh wave

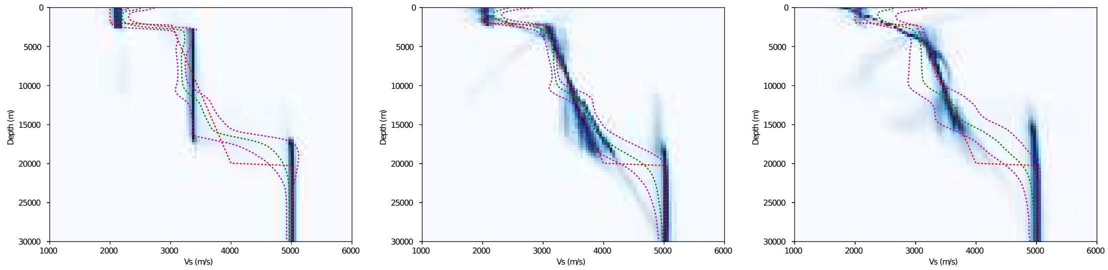
In a final test, Love and Rayleigh dispersion observations are jointly inverted. The details remain the same as summarised previously with a minor distinction being that there are now independent hierarchical error scaling terms for the Love and Rayleigh observations. The results of these inversions are shown in Figure 8.15 where there is much better recovery of the true model. Another general observation is that the combination of Love and Rayleigh observations has reduced the uncertainty in the models.

In the first column where the inversions are only allowed 0th order polynomials the first homogeneous layer of the true model is well recovered. The second layer in the true model with a linear gradient is approximated with a homogeneous layer in the posterior.

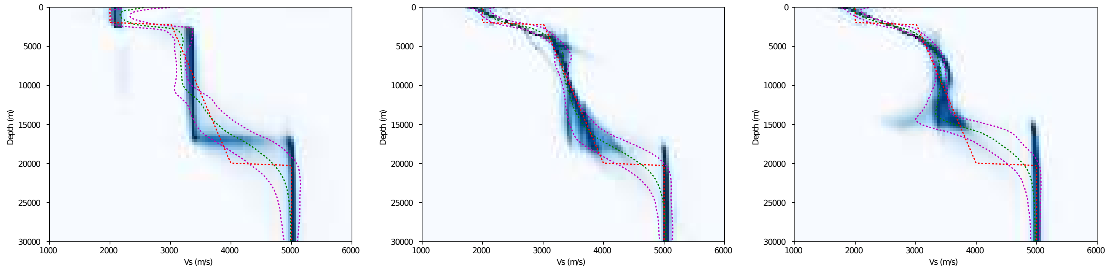
For the inversion that allows up to 1st order polynomials, that is the second column where features from the first column are evident in the posterior, but linear features are also visible. Careful examination of the mean curve (green dashed lines) shows that the three inversions in the second column have generally recovered the true model better than the first, particularly in the mid range of depths.

For the third column, there is evidence of quadratic features, particularly in (c) and (f). The result in (i) is perhaps the best fitting posterior to the true model and visually appears to consist of a linear first layer and quadratic second layer. The histogram on the number of parameters shown in the Appendix in Figure B.14(i) shows that the posterior modal number of parameters is eight compared to the true number of six.

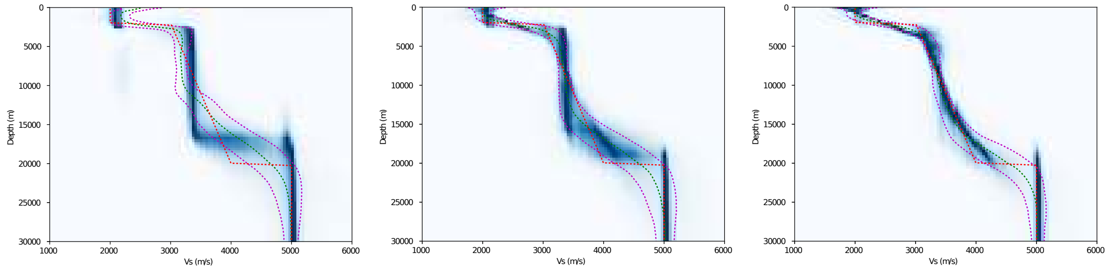
The posterior histogram of discontinuities in each of the inversions is shown in Figure 8.16. The discontinuity at 2km depth is well recovered in all inversions. In the first column, the true second interface at 20km is under-estimated at approximately 15 to 16km. In reviewing the results of the first column in Figure 8.15, the reason for this is that in order to approximate the linear layer with homogeneous layers, a consequence is a shallower second layer causing mis-identification of the true interface.



(a) 0th/Uniform (b) 1st/Uniform (c) 2nd/Uniform



(d) 0th/Beta($\varkappa = 4$) (e) 1st/Beta($\varkappa = 4$) (f) 2nd/Beta($\varkappa = 4$)



(g) 0th/Beta($\varkappa = 8$) (h) 1st/Beta($\varkappa = 8$) (i) 2nd/Beta($\varkappa = 8$)

Figure 8.15: Joint Love and Rayleigh wave inversion histograms with blue shading representing regions of higher probability, the green curve the mean of the ensemble, the red curve the true model and the purple curves showing the spread of the standard deviation from the mean. The first column represents trans-dimensional solutions with 0th order only, the second column trans-dimensionally selects between 0th and 1st order, and the third trans-dimensionally selects between 0th, 1st, and 2nd order polynomials. The first row uses a uniform prior between 1 and 6 km/s, the second row a Beta prior with the mode set from a reference model and a concentration parameter of 4 and the last row is again a Beta prior with a concentration parameter of 8.

In the second column, where the parameterisation can perfectly recreate the true model, the second interface is generally recovered although with larger uncertainty and perhaps a little under-estimated. The uncertainty is a product of reduced resolvability at greater depths. This highlights an important point in that incorrectly representing a linear (or perhaps higher order) layer with a 0th order layer then results in incorrect recovery of discontinuities in the Earth model.

The third column has good recovery of the second interface at 20km deep, with the exception of the result in 8.16(i). The reason for the poor result in (i) is that the inversions have preferred a quadratic function from the surface to represent the observed dispersion and this incorrect parameterisation causes the partition boundaries to not reflect interfaces in the true model. This is again caused by the ill-determined nature of the problem, that is there are many Earth models that give rise to the same dispersion curve.

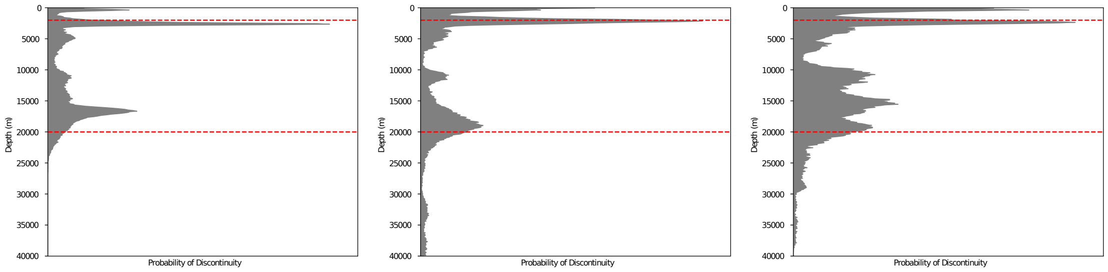
Nonetheless, an important result here is that with joint observations of Love and Rayleigh dispersion information, an Earth model with discontinuities and smoothly varying structure can be feasibly resolved with trans-dimensional partitions and trans-dimensional order polynomials. Additionally, the location of discontinuities is important for inference of phase transitions so the methods presented here may be important for better constraining the depths of these discontinuities.

8.6 Summary

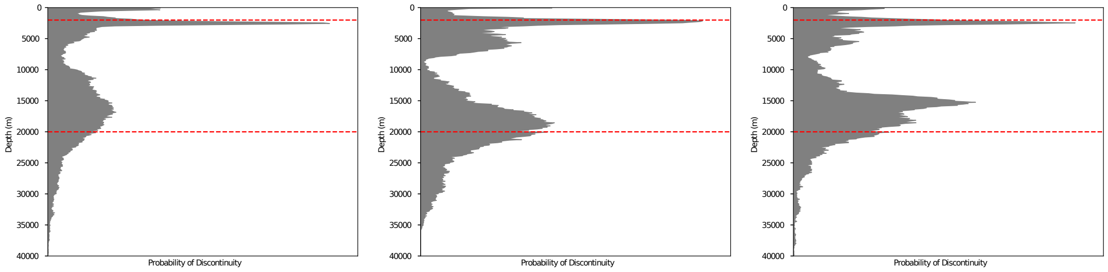
This chapter has derived a stable spectral element method for computing phase velocity from a 1D Earth model. Its accuracy was demonstrated against known analytical solutions and existing codes. A key advantage of this new method is that smoothly varying Earth structure and discontinuities can be modelled together which may more faithfully represent the true Earth.

In a series of synthetic tests, comparisons between inversions using fixed zeroth order partitions and trans-dimensional polynomial partitions for the recovery of known

(a) 0th/Uniform (b) 1st/Uniform (c) 2nd/Uniform



(d) 0th/Beta($\varkappa = 4$) (e) 1st/Beta($\varkappa = 4$) (f) 2nd/Beta($\varkappa = 4$)



(g) 0th/Beta($\varkappa = 8$) (h) 1st/Beta($\varkappa = 8$) (i) 2nd/Beta($\varkappa = 8$)

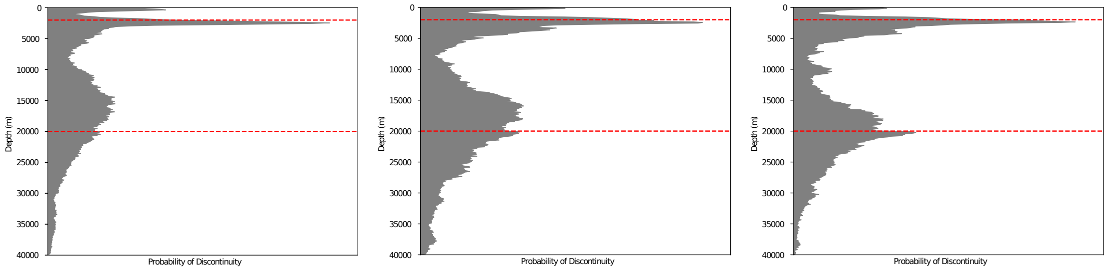


Figure 8.16: Boundaries histogram for the joint Love and Rayleigh wave inversions showing the posterior location of partition boundaries in the ensemble. The first row is from inversions with a uniform prior, the second with a Beta prior with a concentration value of 4, the third with a concentration parameter of 8. The first column is an inversion with 0th order only, the second allows trans-dimensional selection between 0th and 1st order polynomials, and the third allows trans-dimensional selection between 0th, 1st, and 2nd order polynomials.

Earth models were made. It was seen that in general that the recovery by Love waves on their own was not sufficient to disambiguate between slowly varying structure and discontinuities. Rayleigh wave observations appear to be on the cusp of being able to resolving the difference between smooth and discontinuous structure. It is only with the combination of Love and Rayleigh waves where it was demonstrated that inclusion of variable polynomial orders as part of the trans-dimensional inversion made a significant difference. Additionally, the uncertainty is substantially reduced and with the addition of higher order polynomials to represent smoothly varying structure, it was demonstrated that we can get better estimates of the location of discontinuities at depth, which may have important ramifications for inference on phase transitions.

The limitations of the study presented here are: firstly that the independent synthetic noise level may not represent reasonable attainable levels in inversion of real observations. Secondly, that independent noise may be an invalid assumption. Lastly, the independent chains are simulated without parallel tempering and therefore can fixate on single models rather than efficiently traverse the prior. The addition of parallel tempering should be included to ensure that all reasonable models are included in the posterior.

In this synthetic study, only isotropic Earth models have been considered although the derivation of the spectral element forward model is parameterised in terms of standard parameters introduced by Love [1927]. This means that equally the inversion could trans-dimensionally select between azimuthal anisotropy and isotropy similar to Bodin et al. [2016], and determine the support of isotropy versus anisotropy in addition constant layering versus smoothly varying structure. This is likely to result in a more difficult convergence and further ambiguities and may require the addition of other information such as receiver functions or body wave data. It is interesting to speculate here whether azimuthal anisotropy reported in the literature is indeed a well supported inference given the observations or an artefact of inversions where the forward modelling cannot account for smoothly varying structure. This may prove a fruitful area for further research.

The spectral element method presented here is for surface waves but a similar forward modelling approach could be derived for receiver functions. This would enable joint inversion for surface wave dispersion with receiver functions and allow smoothly varying structure in each layer. The sensitivity of receiver functions to interface changes and P-wave velocity could complement the sensitivity of surface wave dispersion and improve the resolvability of local structure.

The spectral element method derived in this chapter can accurately recover the Eigen functions of Rayleigh waves and at higher frequencies, this could be used to invert for Rayleigh wave ellipticity for basin hazard studies. The sedimentary layer often resolved by these studies tends to have a power law shear wave velocity model with respect to depth. This smooth shear wave velocity model is generally approximated with constant layers due to limitations in existing techniques. With the methods presented here, there is the potential to recover these features more robustly and in turn better recover the interfaces representing the sedimentary/bedrock interface, which may have important implications for local seismic hazards.

At the other end of the frequency spectrum, for very long period surface waves, the spherical nature of the Earth needs to be accounted for. Additionally, the change in the gravity potential field needs to be included for Rayleigh waves [Alterman et al., 1959, Woodhouse, 1974, Wiggins, 1976, Tromp and Dahlen, 1992a,b]. It is possible to extend the derivation presented here in order for the spectral element method to be suitable for long period surface waves such as those inverted in Chapter 6. In the formulation presented, each mode of propagation is inverted for at the same time so a single inversion obtains the wave number (and hence phase velocity) for multiple modes. However, there is an issue of accuracy since higher order modes have more oscillatory Eigen functions that require either a finer mesh or higher order polynomials for their accurate representation. This increases computational cost but it is feasible that the methods presented here could be extended to global scale frequencies and multi-modal observations, to answer similar questions of the deeper Earth.

Chapter 9

Conclusions

9.1 Contributions of this thesis

This thesis has considered new approaches to the field of geophysical inversion across a diverse range of applications.

Chapter 2 presents a new approach for extracting phase velocity dispersion information from ambient noise cross-correlations of signals recorded with three component seismic stations. The benefit over existing techniques is that this joint inversion provides both Love and Rayleigh dispersion observations in a single automatic inversion. Jointly inverting Love and Rayleigh observations is shown to reduce the multi-modality of the posterior distribution, a difficulty that is common in other methods. The lack of a stable and consistent method of extracting phase velocities from cross-correlated noise has resulted in many previous authors using group velocity data that while simpler to obtain, provides less precise dispersion information. The new method presented in Chapter 2 provides a stable solution for ambient noise tomography practitioners for the extraction of phase velocity dispersion information with uncertainties.

A major contribution of this thesis is the development of the trans-dimensional tree method for geophysical imaging presented in Chapter 3. This method provides a more flexible and computationally efficient than existing trans-dimensional Voronoi cell approaches. In this chapter, which includes published material [Hawkins and Sambridge, 2015], the algorithm is compared to the existing Voronoi cell approach. The method has now been applied in a number of geophysical problems, some of which feature in this thesis (in Chapters 4, 5, and 6).

An example, not specifically discussed in this thesis, demonstrating the contribution of the development of the trans-dimensional tree method was a collaboration with Jan Dettmer in which the initial sea surface deformation resulting from the great Tohoku earthquake was inverted. This inversion applied the trans-dimensional tree method to a complex joint inversion of disparate observations from ocean bottom pressure (OBP) gauges, global positioning sensors (GPS) and coastal wave gauges (CWG). This work was published in Dettmer et al. [2016] and demonstrated the effectiveness of the trans-dimensional tree method over existing approaches.

In appendices of Dettmer et al. [2016], the benefit of a trans-dimensional inversion compared to a fixed dimensional inversion of the region was shown in its ability in reducing inversion artefacts and uncertainties. Analysis of the effects of uniform prior widths on wavelet parameters were included to demonstrate the sensitivity of the result to different priors. The results show that although the result is sensitive to the prior, the effect is subtle and this is important to understand in trans-dimensional formulations of geophysical inverse problems.

The inverted initial sea surface displacement showed high resolution in most parts of the source region, that is, near the earthquake hypo centre where there was good coverage. This region was well constrained in the inversion with low uncertainties. An interesting feature, not previously observed in other studies, was a peak of the source elongated parallel to the trench showing a well resolved bi-modal finger-like feature in the northern source region that closely follows the trench. Consequently the trans-dimensional tree method developed as part of this thesis has enabled greater understanding of the rupture mechanics of large tsunamigenic events through the inversion of sea surface deformations.

A second application of the trans-dimensional tree approach was to the inversion of Airborne Electromagnetic (AEM) observations in 2D transects. This work was published in Hawkins et al. [2017] and reproduced in Chapter 4 with additional development using hierarchical priors and covariant noise models.

This approach is the first application of a trans-dimensional image based inversion that takes advantage of expected spatial correlation as a means to stabilise the inversion. The study showed improved resolution compared to existing results using a damped and smoothed optimisation approach and also included a quantification of uncertainties of anomalies at greater depths. Further to the material in the published work, the trans-dimensional tree AEM inversion was extended to include a hierarchical Laplacian prior that showed similar results to the tuned uniform prior used in the original study. The success of the hierarchical prior in matching the result of a tuned uniform prior offers a solution to the difficulty in specification of the prior for trans-dimensional tree inversions with wavelet parameterisations, that is, to hierarchically invert for the prior. This provides easier specification of the prior for practitioners and was demonstrated to improve mixing by adapting the prior width to the data.

One of the difficulties in a Bayesian framework is in adequate modelling of the noise. In many inverse problems, noise can be correlated between observations, for example, in the case in Airborne Electromagnetic observations, where different frequencies have different but overlapping sensitivities to depth. In a final AEM inversion, estimates of Gaussian correlated noise were obtained from the posterior of a previous inversion to construct a correlated noise model. This method can be applied even for single flight lines. Chapter 4 demonstrated that inverting with the estimated correlated noise produced simpler models suggesting that properly accounting for correlated noise is important for producing a truly parsimonious result in trans-dimensional inversion. This suggests further synthetic tests to determine the impact of incorrectly assuming an independent noise process when the truth is known to be covariant.

In Chapter 5, a fully non-linear method for the inversion of phase velocity maps from ambient noise dispersion observations was detailed. In some numerical studies, the accuracy of the Fast Marching forward model used for predicting travel times was tested as a function of the grid resolution. The super resolution property of wavelets was used to promote arbitrary models to a resolution that gives acceptable accuracy. In controlled experiments, 2D phase velocity maps were inverted from observations created with synthetic checker board patterns, demonstrating that the trans-dimensional tree approach is an effective scheme for this problem. Additionally the difference between using a fixed ray path forward model and a fully non-linear approach was shown. The conclusion was that the use of fixed ray paths results in failure to resolve higher resolution features and the underestimation of anomaly magnitudes.

Traditionally, group or phase velocity maps are created at specific periods or frequencies independently. Another innovation of this thesis is a new method for the joint inversion of a contiguous band of frequencies. This joint approach acknowledges that the observations, obtained using the methods in Chapter 2, are continuous functions of frequency and therefore naturally correlated. The method in effect is a three dimensional inversion of phase velocity maps with two spatial dimensions and a frequency dimension. This would have been a difficult problem with the previous Voronoi cell approach but is a straight forward extension for the trans-dimensional tree approach developed in Chapter 3 due to its inherent flexibility. By inverting a band of frequencies jointly, the inversion is further stabilised, and comparisons between independently recovered phase velocity maps and those jointly inverted demonstrated improved images at selected frequencies.

Finally, the trans-dimensional tree method was applied on a global scale to invert surface wave observations to build a global tomographic shear wave model. Using a fixed ray approximation, the method is able to invert a large data set of approximately 5 million ray paths, previously intractable with trans-dimensional inversion due to the computational cost of existing Voronoi cell parameterisations.

It was also demonstrated that in previous inversions using a continuous regionalisation scheme with smoothing regularisation, that the magnitude of anomalies was underestimated compared to the trans-dimensional result. Additionally, uncertainties in these regions are low, that is the magnitude of the anomalies are well constrained by the observations. This is caused by two factors, first, the smooth regularisation in the continuous regionalisation scheme will tend to dampen higher magnitudes of smaller scale lengths, and second, the trans-dimensional tree method is able to adapt to local features, thus improving resolution where the data allows it.

Typically, these phase velocity maps are used in subsequent 1D inversions to invert for a shear wave velocity model of the Earth's mantle. Any under or over estimated phase velocity anomalies would consequently be amplified in the shear wave velocities. Although Chapter 6 is essentially a proof of concept, it nonetheless demonstrates that the trans-dimensional tree approach is superior to existing techniques. Future applications of the trans-dimensional tree approach on a global scale include exploration of the Earth's mantle using surface wave observations.

In summary, the trans-dimensional tree approach has been applied in geophysical inverse problems and the success of the method is demonstrated in a variety of applications, including airborne electromagnetic, ambient noise and global surface inversions. Additionally, the approach was published in an initial sea surface deformation inversion [Dettmer et al., 2016] as part of the study of large tsunamigenic earthquakes. Critically, these applications demonstrate the efficiency and flexibility of the trans-dimensional tree method compared to existing approaches making this new method a significant advance for improving geophysical imaging.

In a final contribution of this thesis, the concept of trans-dimensional spectral elements was introduced and its usefulness demonstrated in performing model choice between functions that are smoothly varying and those that have discontinuities in some simple synthetic regression tests. An assumption of trans-dimensional sampling is that a parsimonious result is always obtained, yet this only holds if the "true" (or approximately true) parameterisation is available in the candidate parameterisations in the formulation of the trans-dimensional problem.

This approach was extended by developing a new spectral element method for the solution of Love and Rayleigh wave dispersion for 1D Earth models and running some synthetic tests to determine if the difference between layered or smoothly varying structure can be resolved. It was shown that this is still a very difficult problem, but that in a joint Love and Rayleigh wave dispersion inversion some true models can be recovered, but still with some evidence of ambiguity. This combination of trans-dimensional spectral elements with a spectral element solver is likely to feature in many future inversions with different forward models. The ground work presented here is a major contribution of this thesis.

In short, the concepts and methods developed within this thesis have allowed us to see deeper into the Earth with higher resolution while quantifying uncertainties.

9.2 Future work

In three and a half years, one cannot hope to solve all problems and therefore no thesis is the final word on a subject. There are many avenues that may be pursued and improvements to make.

In Chapter 2, one of the aspects of the inversion that could be improved was the use of optimal scaling for fitting the Bessel function computed in the forward model to the observed real part of the spectrum. This could be extended to fit the observed amplitude as another part of the inversion process to better constrain the inversion in general. It also has the potential to enable attenuation information to be obtained in the inversion from the frequency dependent magnitude of the Bessel function.

In the trans-dimensional tree approach, two parameterisations featured in this thesis: wavelets were used in the trans-dimensional tree chapter, in the inversion of airborne electromagnetic and in the inversion of ambient noise data, whereas in the global surface wave inversion, phase velocity values were directly specified. Each of these parameterisations are choices made in the formulation of the problem. Firstly, there are many other bases available and Discrete Cosine Transforms [Ahmed et al., 1974] or over complete bases such as Curvelets [Candes and Donoho, 1999] might be better suited to certain problems.

Ultimately, following the trans-dimensional philosophy of allowing the data to determine the parameterisation, an extension would be to use the same ideas to select between competing bases. In principle, this could be implemented as jumps between wavelet bases within the same families, a good example are the Daubechies wavelets [Daubechies, 1988] where there is an increasing regularity with order in the sequence of wavelet basis from D2 or the Haar wavelet which is discontinuous, through D4, D6 (used in Chapter 3) and so on. While jumps between families of related bases are plausible, jumps between unrelated bases are likely to be more difficult, for example, jumps between wavelets and curvelets. In these cases, recent advances in the numerical evaluation of the evidence in a Bayesian inversion may be more appropriate.

The ambient noise study in Chapter 5 developed a tightly coupled Fast Marching implementation to the trans-dimensional tree with a wavelet parameterisation. This inversion assumed elastic isotropic propagation of surface waves but many studies have proposed crustal anisotropy. Fast marching can be readily extended to include anisotropy [Sethian and Vladimirsky, 2003], however a potential short-coming of the wavelet parameterisation is the difficulty in encoding anisotropy with wavelets. The common parameterisation of anisotropy for surface waves [Backus, 1965, Smith and Dahlen, 1973] consists of 3 or 5 parameters. There are wavelet transforms for complex numbers [Shukla, 2003, Selesnick et al., 2005] or two parameters but none for 3 or 5 parameters. A naive solution where three separate wavelet images representing the three anisotropy parameters are possible but may not be efficient. A method of efficiently applying the trans-dimensional tree method to anisotropy would be an important area of future research.

In the subsequent surface wave dispersion inversion to recover a shear wave velocity model from the 3D phase velocity maps, recovered uncertainties were rather high. An under-utilised aspect of the 3D inversion is that posterior model covariances can be estimated from the ensemble and may be used in the subsequent surface wave dispersion inversion. It remains to be seen if this can be used to reduce the uncertainties in the shear wave velocity model.

In Chapter 5, using a fixed ray path approximation in the forward model, to a fully non-linear inversion where ray paths are updated from the velocity model, was shown to produce lower magnitude velocity anomalies. For the global surface wave inversions in Chapter 6, fixed ray paths were used and given the relatively efficient inversion there is room to incorporate the Fast Marching method. The Fast Marching method is general enough to be equally applicable to computing travel times on the surface of a sphere [Kimmel and Sethian, 1997] and as stated before can be extended to include anisotropy. It is here where the non-wavelet parameterisation used in the global surface wave inversion may have an advantage as it would be relatively simple to include trans-dimensional jumps within active tree nodes between isotropic phase velocity and anisotropic, that is jumps between one and three parameters for each active tree node.

In Chapter 8 three example problems of surface wave dispersion were presented. Another that could be explored is Rayleigh wave ellipticity. This method is important for basin hazard studies and is used to determine the thickness of the sedimentary layer and first interface which typically represents the sedimentary/bedrock transition. Of interest is that the shear wave velocity of the sedimentary layer typically follows a power law relationship with depth and this would be better represented by a higher order polynomial. Methods developed in this thesis mean that both the trans-dimensional infrastructure and a forward model capable of being applied to this problem are now available to potentially give greater constraints on sedimentary cover. An additional application would be as part of the Australian "Uncover" project that has the major goal of mapping regolith thickness [Blewett, 2017].

The derivation of the spectral element solver in Chapter 8 and Appendix A uses a flat Earth approximation which is suitable to relatively high frequency surface waves. At the other end of the frequency spectrum, this could be re-targeted to a derivation in a spherical Earth and then this method could be applied to longer period Love and Rayleigh waves and may even be applied to the trans-dimensional inversion of the phase velocity maps generated in Chapter 6. In Chapter 8, trans-dimensional jumps between different polynomial orders were used in the examples, but the forward model used supports anisotropy meaning that jumps, or model choice questions, could be proposed between isotropic and anisotropic layering. Ultimately though, the surface wave dispersion problem is highly non-unique and as such may require joint inversion with other data types to more confidently resolve these model choice problems.

As a more general comment, relatively simplistic sampling approaches have been used in all the inversions performed in this thesis, with the exception that Parallel Tempering was heavily relied upon. This was a deliberate decision so as to not conflate the ability of the new methods developed during this thesis with performance improvements afforded by more advanced sampling methods. Relatively small numbers of independent chains have been used in the presented inversions but much larger numbers of independent chains have become popular and allow the population statistics to improve convergence, for example Liu and Chen [1998], Andrieu et al. [2010]. These population based approaches have been used in conjunction with trans-dimensional methods and seem to provide faster convergence to the posterior [Jasra et al., 2007a,b, 2008, Koutsourelakis, 2009, Dettmer et al., 2011].

A simple idea for improving McMC proposals is the concept of using parallel proposals [Liu et al., 2000, Liu, 2001, Calderhead, 2014]. Similarly, where gradients can be computed, there are methods ranging from simple incorporation of direction into proposals [Roberts and Tweedie, 1996, Roberts, 1998] to the more complex Hybrid or Hamiltonian Monte Carlo [Duane et al., 1987, Neal, 2011].

The convergence of trans-dimensional samplers is in general hampered by the low acceptance rates of trans-dimensional steps. Various approaches [Tierney and Mira, 1999, Green and Mira, 2001, Al-Awadhi et al., 2004, Sen and Biswas, 2016] have been proposed. Similarly, in the improving or optimising the acceptance rates of more typical fixed dimensional proposals, there are many methods that could be used to adapt proposals for improved convergence rates [Haario et al., 2005, 2006, Andrieu and Thoms, 2008, ter Braak and Vrugt, 2008, Vrugt et al., 2009, Vrugt and ter Braak, 2011, Laloy and Vrugt, 2012]. Often these schemes are adapted to the fixed dimension case and required some modification to suit trans-dimensional algorithms.

The inversions have also typically been initialised using homogeneous models to demonstrate the ability of the methods presented here to converge to complex models. In real world applications it may be beneficial to initialise McMC chains with some optimised model. An interesting area of future work would be to look at the compatibility between the trans-dimensional tree with a wavelet parameterisation using compressive sensing ideas [Candes et al., 2006] in which l_1 norms are used for regularisation of sparse model parameterisations. If a wavelet parameterisation is indeed a sparse parameterisation for a problem, then comparing the result of l_1 normed inversions to the trans-dimensional tree approach would be interesting. It would be hoped that they are largely compatible so that the l_1 normed optimisation could be used for generating a good starting model, and the trans-dimensional tree approach could be used to explore the posterior about this model to estimate uncertainties.

As a last note, a qualitative metric for convergence for trans-dimensional chains was generally used. Where the qualitative convergence metric was tested against known criteria [Sisson and Fan, 2007], it has been shown to be conservative. However, formal effective metrics for trans-dimensional convergence should be an area of future work.

Appendix A

A spectral element solution for 1D surface wave dispersion

A.1 Introduction

This appendix provides the detailed derivation of the new spectral element solution to 1D surface wave dispersion for Love and Rayleigh waves used in Chapter 8.

A.2 Love waves

This section presents a spectral element solution of the differential equation that describes the propagation of Love waves in a 1D Earth model with a flat Earth approximation. For Love waves, the solution describes the amplitude of oscillations in the y direction as a function of depth or z. These oscillations are perpendicular to the propagation direction x.

Restating (8.38)

$$-\omega^2 V\rho = -k^2 N V(k,\omega,z) + \frac{d}{dz}\left[L\frac{dV(k,\omega,z)}{dz}\right]. \tag{A.1}$$

The domain is divided into E elements from the surface, z_{surface} which is negative in the coordinate system, to the basement, $z=0$ and then forming the weak form of the equation by multiplying by a smooth test function v and integrating to obtain

$$\begin{aligned}
&\omega^2 \sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} \rho V v dz \\
&-k^2 \sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} N V v dz \\
&+\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} \frac{d}{dz}\left[L\frac{dV}{dz}\right] v dz = 0.
\end{aligned} \tag{A.2}$$

The function v is arbitrary except for the fact that it must obey the boundary conditions equally imposed on the solution for the displacement V.

This last term can then be integrated by parts to obtain

$$\begin{aligned}&\omega^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\rho V v dz\\&-k^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} N V v dz\\&-\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} L\frac{dV}{dz}\frac{dv}{dz}dz+\\&\sum_{e=1}^{E}L\frac{dV}{dz}v\bigg|_{z_e-1}^{z_e} = 0.\end{aligned}\tag{A.3}$$

Recall that the displacement V and the traction $L\frac{dV}{dz}$ must be continuous everywhere which requires that the last term perfectly cancels except for the surface and basement values leaving

$$\begin{aligned}&\omega^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\rho V v dz\\&-k^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} N V v dz\\&-\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} L\frac{dV}{dz}\frac{dv}{dz}dz+\\&L(z_b)\frac{dV(z_b)}{dz}v(z_b)-L(z_{\text{surface}})\frac{dV(z_{\text{surface}})}{dz}v(0) = 0.\end{aligned}\tag{A.4}$$

This can be reduced further by using the fact that the traction must vanish at the surface, hence $L(0)\frac{dV(0)}{dz} = 0$, and

$$\begin{aligned}
&\omega^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \rho V v dz \\
&-k^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} N V v dz \\
&-\sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L \frac{dV}{dz} \frac{dv}{dz} dz + \\
&L(z_b) \frac{dV(z_b)}{dz} v(z_b) = 0. \qquad \text{(A.5)}
\end{aligned}$$

An option here is to assume that the distance from the surface to basement is sufficient so that oscillation amplitudes are approximately zero at the basement over the range of frequencies considered. That is $V(0)$ is near zero and therefore $v(0) \approx 0$ to remove the last term, but this is left to be treated later.

A key component of the spectral element method is the use of Lagrange nodal interpolation polynomials with nodal points arranged at the zeros of Lobatto polynomials. The domain for these Gauss-Legendre-Lobatto (GLL) nodes in 1D is [-1, 1] so each of the element integrals requires transformation to this domain. This is achieved with a simple linear transformation so that

$$z_e(\zeta) = z_{e-1} + \frac{\zeta + 1}{2}(z_e - z_{e-1}) = z_{e-1} + \frac{(\zeta + 1)\Delta z_e}{2}, \qquad \text{(A.6)}$$

where ζ is the spectral element coordinate. The substitutions for the required change of variables are

$$dz = \frac{\Delta z_e}{2} d\zeta \qquad \text{(A.7)}$$

$$\frac{d}{dz} = \frac{2}{\Delta z_e} \frac{d}{d\zeta}. \qquad \text{(A.8)}$$

With this linear transform, the system of equations becomes

$$\begin{aligned}
&\omega^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \int_{-1}^{1} \rho V v d\zeta \\
&-k^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \int_{-1}^{1} N V v d\zeta \\
&-\sum_{e=1}^{E} \frac{2}{\Delta z_e} \int_{-1}^{1} L \frac{dV}{d\zeta} \frac{dv}{d\zeta} d\zeta + \\
&L(z_b) \frac{dV(z_b)}{dz} v(z_b) = 0.
\end{aligned} \tag{A.9}$$

If each of the varying functions of ζ are approximated with Gauss-Legendre-Lobatto polynomial approximations, that is

$$f(\zeta) \approx \sum_{i=0}^{P} f_i l_i^{(P)}(\zeta_i^{(P)}), \tag{A.10}$$

where $l_i^{(P)}$ is the Pth order Lagrange interpolating polynomial, ζ_i is the ith Lobatto co-location point for and f_i is short hand for $f(\zeta_i^{(P)})$.

The integrals of functions of ζ can then be approximated

$$\int_{-1}^{1} f(\zeta) d\zeta \approx \sum_{i=0}^{P} f_i w_i^{(P)}, \tag{A.11}$$

where $w_i^{(P)}$ is the quadrature weight.

Now taking the first two terms of (A.9) in turn, the first two are

$$\omega^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \int_{-1}^{1} \rho V v d\zeta \approx \omega^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \sum_{i=0}^{P} v_i V_i \rho_i w_i, \tag{A.12}$$

and

$$-k^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \int_{-1}^{1} N V v d\zeta \approx -k^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \sum_{i=0}^{P} v_i N_i V_i w_i. \tag{A.13}$$

For the third term, the derivatives of the functions add a complication. Given the approximation of the functions of ζ with polynomials, the derivatives of said functions are given by

$$\frac{df(\zeta)}{d\zeta} \approx \sum_{i=0}^{P} f_i \frac{dV_i^{(P)}}{d\zeta}(\zeta_i^{(P)}), \tag{A.14}$$

and the third term can be written

$$-\sum_{e=1}^{E} \frac{2}{\Delta z_e} \int_{-1}^{1} L \frac{dV}{d\zeta} \frac{dv}{d\zeta} d\zeta \approx -\sum_{e=1}^{E} \frac{2}{\Delta z_e} \sum_{j=0}^{P} w_j L_j \left[\sum_{k=0}^{P} V_k \frac{dl_k j^{(P)}}{d\zeta} (\zeta_j^{(P)} \right] \left[\sum_{i=0}^{P} v_i \frac{dl_i^{(P)}}{d\zeta} (\zeta_j^{(P)} \right], \tag{A.15}$$

in which the ordering of summation can be rearranged to

$$\sum_{e=1}^{E} \frac{2}{\Delta z_e} \sum_{i=0}^{P} v_i \left[\sum_{k=0}^{P} V_k \left[\sum_{j=0}^{P} w_j L_j \frac{dl_i^{(P)}}{d\zeta} (\zeta_j^{(P)}) \frac{dl_k^{(P)}}{d\zeta} (\zeta_j^{(P)}) \right] \right] \tag{A.16}$$

Combining these, the full equation becomes

$$\begin{aligned} \sum_{e=1}^{E} \sum_{i=0}^{P} v_i^e \Bigg[& \omega^2 \frac{\Delta z_e}{2} V_i \rho_i w_i \\ & - k^2 \frac{\Delta z_e}{2} N_i V_i w_i \\ & - \frac{2}{\Delta z_e} \left[\sum_{k=0}^{P} V_k \left[\sum_{j=0}^{P} w_j L_j \frac{dl_i^{(P)}}{d\zeta} (\zeta_j^{(P)}) \frac{dl_k^{(P)}}{d\zeta} (\zeta_j^{(P)}) \right] \right] \Bigg] \\ & + L(0) \frac{dl(0)}{dz} v(0) = 0. \end{aligned} \tag{A.17}$$

Now since the interpolation polynomials are orthogonal, the v_i^e's are linearly independent with the exception that at element boundaries the condition $v_P^e = v_0^{e+1}$ must be met for continuity of the test function v. This means that internal collocation points must satisfy

$$\omega^2 \frac{\Delta z_e}{2} V_i \rho_i w_i - k^2 \frac{\Delta z_e}{2} N_i V_i w_i - \frac{2}{\Delta z_e} \left[\sum_{k=0}^{P} V_k \left[\sum_{j=0}^{P} w_j L_j \frac{d l_i^{(P)}}{d\zeta}(\zeta_j^{(P)}) \frac{d l_k^{(P)}}{d\zeta}(\zeta_j^{(P)}) \right] \right] = 0. \tag{A.18}$$

These independent equations can be assembled into a matrix equation

$$\left[\omega^2 \mathbf{A} - k^2 \mathbf{B} - \mathbf{C}\right] \mathbf{V} = 0, \tag{A.19}$$

with **A** and **B** diagonal matrices, **C** a block diagonal matrix, and **V** a vector representing the values of the Eigen function at the nodal interpolation values. Here the diagonallity of **B** means its inverse can be trivially computed and rearrange the equation to

$$\left[\mathbf{D} - k^2 \mathbf{I}\right] \mathbf{V} = 0, \tag{A.20}$$

where $\mathbf{D} = \mathbf{B}^{-1}\left[\omega^2 \mathbf{A} - \mathbf{C}\right]$. Non-trivial solutions for **V** require

$$\det\left|\mathbf{D} - k^2 \mathbf{I}\right| = 0, \tag{A.21}$$

and this can be efficiently solved using standard Eigen value calculation codes [Moler and Stewart, 1973]. The Eigen values of the system can be complex of which the imaginary part in effect is a decay term resulting in a Love wave that does not propagate. For negative real Eigen values, the wave number is purely imaginary and the wave does not propagate. Therefore, the positive Eigen values are the only propagating modes, and these are called static instabilities in the vibrational analysis community [Chang et al., 2010]. Lysmer and Drake [1972] gives a more detailed discussion of the Eigen values and their meaning than the summary presented here.

So for a given Earth model and frequency, the Eigen problem in (A.21) can be assembled and the eigen problem solved for the real positive Eigen values. Each of these represent a surface wave mode whose wave number is the square root of the Eigen value from which the phase velocity can be computed. Additionally the Eigen vectors are the displacement function of the oscillations.

A.3 Rayleigh waves

This section presents the spectral element solution of the coupled equations that describe the propagation of Rayleigh waves in a 1D Earth model with a flat Earth approximation. The solution for Rayleigh waves consists of a system of two equations, a radial equation with oscillations in the x direction along the direction of wave propagation and a vertical equation with oscillations in the z direction, or into the Earth.

A.3.1 Radial equation

Restating (8.55)

$$-\rho\omega^2 U = -k^2 AU - kF\frac{dW}{dz} + \frac{d}{dz}\Big[L(\frac{dU}{dz} - kW)\Big]. \tag{A.22}$$

The domain is divided into E unevenly sized cells from the surface, $z = z_{\text{surface}}$, to the basement, $z = 0$, and then forming the weak version of the equation by multiplying by a smooth test function, v, and integrating. After these steps the equation becomes

$$\begin{aligned}
-\omega^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \rho U v dz = &-k^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} AU v dz \\
&-k \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} F\frac{dW}{dz} v dz \\
&+\sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \frac{d}{dz}\Big[L\Big(\frac{dU}{dz} - kW\Big)\Big] v dz.
\end{aligned} \tag{A.23}$$

Integrating the last term by parts gives

$$
\begin{aligned}
-\omega^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\rho U v\,dz = & -k^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} A U v\,dz \\
& -k\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} F\frac{dW}{dz}v\,dz \\
& -\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} L\left(\frac{dU}{dz}-kW\right)\frac{dv}{dz}dz \\
& +\sum_{e=1}^{E} L\left(\frac{dU}{dz}-kW\right)v\bigg|_{z_{e-1}}^{z_e} . \qquad \text{(A.24)}
\end{aligned}
$$

The term $L\left(\frac{dU}{dz}-kW\right)$ is the amplitude of the stress σ_{zx} which must be continuous across cell boundaries. This means that most of these terms cancel will each other in the summation leaving only the end points at $z=z_{\text{surface}}$ and $z=0$, that is

$$
\begin{aligned}
-\omega^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\rho U v\,dz = & -k^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} A U v\,dz \\
& -k\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} F\frac{dW}{dz}v\,dz \\
& -\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} L\left(\frac{dU}{dz}-kW\right)\frac{dv}{dz}dz \\
& + L\left(\frac{dU}{dz}-kW\right)v\bigg|_{z_{\text{surface}}}^{0} . \qquad \text{(A.25)}
\end{aligned}
$$

At $z=z_{\text{surface}}$ the traction σ_{zx} must vanish and therefore the term $L\left(\frac{dU}{dz}-kW\right)$ must be zero. For now it is assumed the basement will be treated separately, leaving

$$
\begin{aligned}
-\omega^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \rho U v dz = & -k^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} A U v dz \\
& -k \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} F \frac{dW}{dz} v dz \\
& -\sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L \left(\frac{dU}{dz} - kW \right) \frac{dv}{dz} dz.
\end{aligned} \tag{A.26}
$$

Lastly, the right hand side is rearranged to collect polynomial terms of k giving

$$
\begin{aligned}
-\omega^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \rho U v dz = & -k^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} A U v dz \\
& -k \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} F \frac{dW}{dz} v - LW \frac{dv}{dz} dz \\
& -\sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L \frac{dU}{dz} \frac{dv}{dz} dz.
\end{aligned} \tag{A.27}
$$

A change of variable is required to adjust the integrals in each cell to span from $-1 \ldots 1$. The affine transform from the ζ to z coordinate in each cell is given by

$$
z_e(\zeta) = z_{k-1} + \frac{\zeta + 1}{2}(z_e - z_{k-1}) = z_{k-1} + \frac{(\zeta + 1)\Delta z_e}{2}. \tag{A.28}
$$

From these, the substitutions for the change of variables are

$$
dz = \tfrac{\Delta z_e}{2} d\zeta \tag{A.29}
$$

$$
\frac{d}{dz} = \tfrac{2}{\Delta z_e} \tfrac{d}{d\zeta} \tag{A.30}
$$

Substituting these into the equation gives

$$\begin{aligned}-\omega^2\sum_{e=1}^{E}\frac{\Delta z_e}{2}\int_{-1}^{1}\rho^e r_x^e v^e d\zeta = &-k^2\sum_{e=1}^{E}\frac{\Delta z_e}{2}\int_{-1}^{1}A^e r_x^e v^e d\zeta\\ &-k\sum_{e=1}^{E}\int_{-1}^{1}F^e\frac{dW^e}{d\zeta}v^e - L^e W\frac{dv^e}{d\zeta}d\zeta\\ &-\sum_{e=1}^{E}\frac{2}{\Delta z_e}\int_{-1}^{1}L^e\frac{dr^e}{d\zeta}\frac{dv^e}{d\zeta}d\zeta,\end{aligned} \tag{A.31}$$

where the superscript e represents the piece wise function in element e. In a similar manner to the Love wave derivation, integrals can be replaced with their approximate Gauss-Lobatto-Legendre quadrature rule equivalents, that is

$$\Phi(\zeta)\approx\sum_{j=0}^{N}l_j(\zeta)\Phi(\zeta_j), \tag{A.32}$$

where Φ is the interpolated function, l_j is the jth Lagrange cardinal polynomial and ζ_j is the jth Lobatto point. The substitutions of each of the terms of (A.31) follow.

The first two terms are straight forward substitutions

$$\int_{-1}^{1}\rho^e r_x^e v^e d\zeta \approx \sum_{j=0}^{P} w_j \rho_j r_{x_j} v_j, \tag{A.33}$$

and

$$\int_{-1}^{1}A^e r_x^e v^e d\zeta \approx \sum_{j=0}^{P} w_j A_j r_{x_j} v_j. \tag{A.34}$$

The third term has derivative terms that can be computed from the polynomial representation. It is useful to arrange the summations with the outer summation over the smooth test function v, for this reason the integral is initially split the integral where individual summations are computed the rearranged to

$$\begin{aligned}\int_{-1}^{1} F^e \frac{dW^e}{d\zeta} v^e - L^e W \frac{dv^e}{d\zeta} d\zeta &\approx \sum_{j=0}^{P} w_j v_j F_j \sum_{k=0}^{P} r_{z_k} l_k'(\zeta_j) - \sum_{k=0}^{P} w_k L_k r_{z_k} \sum_{j=0}^{P} v_j l_j'(\zeta_k) \\ &\approx \sum_{j=0}^{P} w_j v_j F_j \sum_{k=0}^{P} r_{z_k} l_k'(\zeta_j) - \sum_{j=0}^{P} v_j \sum_{k=0}^{P} w_k L_k r_{z_k} l_j'(\zeta_k) \\ &\approx \sum_{j=0}^{P} v_j \sum_{k=0}^{P} \left[F_j w_j r_{z_k} l_k'(\zeta_j) - L_k w_k r_{z_k} l_j'(\zeta_k) r \right] \\ &\approx \sum_{j=0}^{P} v_j \sum_{k=0}^{P} r_{z_k} \left[F_j w_j l_k'(\zeta_j) - L_k w_k l_j'(\zeta_k) r \right] \qquad \text{(A.35)}\end{aligned}$$

The last term similarly has derivatives the summation is re-ordered so that the outer summation is over the smooth test function v

$$\begin{aligned}\int_{-1}^{1} L^e \frac{dU^e}{d\zeta} \frac{dv^e}{d\zeta} d\zeta &\approx \sum_{l=0}^{P} w_l L_l \left[\sum_{k=0}^{P} r_{x_k} l_k'(\zeta_l) \right] \left[\sum_{j=0}^{P} v_j l_j'(\zeta_l) \right] \\ &\approx \sum_{j=0}^{P} v_j \left[\sum_{k=0}^{P} r_{x_k} \sum_{l=0}^{P} w_l L_l l_k'(\zeta_l) l_j'(\zeta_l) \right] \qquad \text{(A.36)}\end{aligned}$$

Now since the interpolation polynomials are orthogonal, the v_i^e's are linearly independent with the exception that at element boundaries the condition $v_P^e = v_0^{e+1}$ must be met for continuity of the test function v. This means that when combining all substitutions above, the internal collocation points must satisfy

$$\begin{aligned}\sum_{e=0}^{E}\sum_{j=0}^{P} v_j^e \Bigg[&-\omega^2 \frac{\Delta z_e}{2} w_j \rho_j^e r_{x_j}^e \\ &+ k^2 \frac{\Delta z_e}{2} w_j A r_{x_j}^e \\ &+ k \sum_{k}^{P} r_{z_k}^e \Big(w_j F_j^e l_k'(\zeta_j) - w_k L_k L_j'(\zeta_k) \Big) \\ &+ \frac{2}{\Delta z_e} \sum_{k=0}^{P} r_k^e \sum_{l=0}^{P} l_k'(\xi_j) l_l'(\xi_j) w_l L_l^e \Bigg] = 0. \end{aligned} \tag{A.37}$$

Next is the vertical equation solution required before coupling the two results into a single system of equations.

A.3.2 Vertical equation

Restating (8.61),

$$-\rho\omega^2 W = Lk\frac{dU}{dz} - Lk^2 W + \frac{d}{dz}\left[FkU + C\frac{dW}{dz}\right]. \tag{A.38}$$

The domain is divided into E unevenly sized cells from the surface, $z = z_{\text{surface}}$, to the basement, $z = 0$, and then forming the weak version of the equation by multiplying by a smooth test function, v, and integrating. After these steps the equation becomes

$$\begin{aligned}-\omega^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \rho W v dz = &-k^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L W v dz \\ &+ k \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L \frac{dU}{dz} v dz \\ &+ \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \frac{d}{dz}\left[C\frac{dW}{dz} + FkU\right] v dz. \end{aligned} \tag{A.39}$$

Integrating the last term by parts gives

$$\begin{aligned}
-\omega^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\rho W v dz = & -k^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} LWv dz \\
& +k\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} L\frac{dU}{dz} v dz \\
& -\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\left(C\frac{dW}{dz}+FkU\right)\frac{dv}{dz}dz \\
& +\sum_{e=1}^{E}\left(C\frac{dU}{dz}+FkU\right)v\Big|_{z_{e-1}}^{z_e} . \qquad (A.40)
\end{aligned}$$

The term $C\frac{dU}{dz}+FkU$ is the amplitude of the stress σ_{zz} which must be continuous across cell boundaries. This means that most of these terms cancel will each other in the summation leaving only the end points at $z = z_{\text{surface}}$ and $z = 0$, that is

$$\begin{aligned}
-\omega^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\rho W v dz = & -k^2\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} LWv dz \\
& +k\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e} L\frac{dU}{dz} v dz \\
& -\sum_{e=1}^{E}\int_{z_{e-1}}^{z_e}\left(C\frac{dW}{dz}+FkU\right)\frac{dv}{dz}dz \\
& +\left(C\frac{dW}{dz}+FkU\right)v\Big|_{z_{\text{surface}}}^{0} . \qquad (A.41)
\end{aligned}$$

At $z = z_{\text{surface}}$ the traction σ_{zz} must vanish and therefore the term $C\frac{dU}{dz}+FkU$ must be zero. The basement will be treated separately and it is removed from this derivation, leaving

$$\begin{aligned}
-\omega^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \rho W v dz = & -k^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L W v dz \\
& + k \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L \frac{dU}{dz} v dz \\
& - \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \left(C \frac{dW}{dz} + F k U \right) \frac{dv}{dz} dz.
\end{aligned} \tag{A.42}$$

Lastly the right hand side is rearranged to collect terms of a polynomial in k giving

$$\begin{aligned}
-\omega^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} \rho W v dz = & -k^2 \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} L W v dz \\
& - k \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} F U \frac{dv}{dz} - L \frac{dU}{dz} v dz \\
& - \sum_{e=1}^{E} \int_{z_{e-1}}^{z_e} C \frac{dW}{dz} \frac{dv}{dz} dz.
\end{aligned} \tag{A.43}$$

The affine transform for the mapping of the cells from z to the ζ coordinates is used to obtain

$$\begin{aligned}
-\omega^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \int_{-1}^{1} \rho^e r_z^e v^e d\zeta = & -k^2 \sum_{e=1}^{E} \frac{\Delta z_e}{2} \int_{-1}^{1} L^e r_z^e v^e d\zeta \\
& - k \sum_{e=1}^{E} \int_{-1}^{1} F^e U \frac{dv^e}{d\zeta} - L^e \frac{dU^e}{d\zeta} v^e d\zeta \\
& - \sum_{e=1}^{E} \frac{2}{\Delta z_e} \int_{-1}^{1} C^e \frac{dW^e}{d\zeta} \frac{dv^e}{d\zeta} d\zeta,
\end{aligned} \tag{A.44}$$

where the superscript e represents the piece wise function in element e. As before, each integral term in the above equation is replaced with their corresponding polynomial approximations and evaluate the integrals using GLL quadrature. The first two terms are straight forward substitutions

$$\int_{-1}^{1} \rho^e r_z^e v^e d\zeta \approx \sum_{j=0}^{P} w_j \rho_j r_{z_j} v_j, \tag{A.45}$$

and

$$\int_{-1}^{1} L^e r_z^e v^e d\zeta \approx \sum_{j=0}^{P} w_j L_j r_{z_j} v_j. \tag{A.46}$$

The third term has derivative terms that can be computed from the polynomial representation. It is useful to arrange the summations with the outer summation over the smooth test function v, for this reason the integral is split and individual summations computed before rearranging to give

$$\begin{aligned}
\int_{-1}^{1} FU\frac{dv^e}{d\zeta} - L^e\frac{dU^e}{d\zeta}v^e d\zeta &\approx \sum_{k=0}^{P} w_k r_{x_k} F_k \sum_{j=0}^{P} v_j l_j'(\zeta_k) - \sum_{j=0}^{P} w_j L_j v_j \sum_{k=0}^{P} r_{x_k} l_k'(\zeta_j) \\
&\approx \sum_{j=0}^{P} v_j \sum_{k=0}^{P} r_{z_k} w_k F_k l_j'(\zeta_k) - \sum_{j=0}^{P} w_j L_j v_j \sum_{k=0}^{P} r_{x_k} l_k'(\zeta_j) \\
&\approx \sum_{j=0}^{P} v_j \sum_{k=0}^{P} \left[r_{z_k} F_k w_k l_j'(\zeta_k) - w_j L_j r_{z_k} l_k'(\zeta_j) \right] \\
&\approx \sum_{j=0}^{P} v_j \sum_{k=0}^{P} r_{z_k} \left[F_k w_k l_j'(\zeta_k) - w_j L_j l_k'(\zeta_j) \right]
\end{aligned} \tag{A.47}$$

The last term similarly has derivatives and again the summation is rearranged so that the outer summation is over the smooth test function v

$$\begin{aligned}
\int_{-1}^{1} C^e \frac{dW^e}{d\zeta}\frac{dv^e}{d\zeta} d\zeta &\approx \sum_{l=0}^{P} w_l C_l \left[\sum_{k=0}^{P} r_{z_k} l_k'(\zeta_l) \right] \left[\sum_{j=0}^{P} v_j l_j'(\zeta_l) \right] \\
&\approx \sum_{j=0}^{P} v_j \left[\sum_{k=0}^{P} r_{z_k} \sum_{l=0}^{P} w_l C_l l_k'(\zeta_l) l_j'(\zeta_l) \right]
\end{aligned} \tag{A.48}$$

Now since the interpolation polynomials are orthogonal, the v_i^e's are linearly independent with the exception that at element boundaries the condition $v_P^e = v_0^{e+1}$ must be met for continuity of the test function v. This means that when combining all substitutions above, the internal collocation points must satisfy

$$\begin{aligned}\sum_{e=0}^{E}\sum_{j=0}^{P} v_j^e \Bigg[& -\omega^2 \frac{\Delta z_e}{2} w_j \rho_j^e r_{z_j}^e \\ & + k^2 \frac{\Delta z_e}{2} w_j L_j r_{z_j}^e \\ & + k \sum_{k}^{P} r_{x_k}^e \left(F_k w_k l_j'(\zeta_k) - w_j L_j l_k'(\zeta_j) \right) \\ & + \frac{2}{\Delta z_e} \sum_{k=0}^{P} r_{z_k}^e \sum_{l=0}^{P} l_k'(\xi_j) l_l'(\xi_j) w_l L_l^e \Bigg] = 0. \end{aligned} \tag{A.49}$$

A.3.3 Coupled equation

Using the same approach as for the Love wave spectral element formulation, combining(A.37) and (A.49) into a single matrix equation gives

$$\left\{ \omega^2 \begin{bmatrix} A_x & 0 \\ 0 & A_z \end{bmatrix} + k^2 \begin{bmatrix} B_x & 0 \\ 0 & B_z \end{bmatrix} + k \begin{bmatrix} 0 & C_x \\ C_z & 0 \end{bmatrix} + \begin{bmatrix} D_x & 0 \\ 0 & D_z \end{bmatrix} \right\} \begin{bmatrix} \mathbf{r}_x \\ \mathbf{r}_z \end{bmatrix} = 0, \tag{A.50}$$

where the vectors $\mathbf{r}_x$ and $\mathbf{r}_z$ represent the eigen functions of the horizontal and vertical oscillations. Note that the coupling between the two systems is confined to factors of k. Non-trivial solutions require

$$\det \left| \omega^2 \mathbf{A} + k^2 \mathbf{B} + k\mathbf{C} + \mathbf{D} \right| = 0 \tag{A.51}$$

which is a quadratic Eigen value problem [Tisseur and Meerbergen, 2001]. The approach to solving this problem is a two step process of scaling the problem for numerical accuracy and restating the equation in one of the many companion forms which reduce the problem to a general Eigen problem. The details are in Hammarling et al. [2013], but the synopsis is that scaling terms are computed using

$$\gamma = \sqrt{\frac{||\omega^2\mathbf{A}+\mathbf{D}||_2}{||\mathbf{B}||_2}} \tag{A.52}$$

$$\delta = \frac{2}{||\omega^2\mathbf{A}+\mathbf{D}||_2+\gamma||\mathbf{C}||_2}, \tag{A.53}$$

and then solve the second companion form of (A.51) which results in the general Eigen system

$$\det\left|\begin{bmatrix}\gamma\delta\mathbf{C} & -\mathbf{I}\\ \delta(\omega^2\mathbf{A}+\mathbf{D}) & \end{bmatrix}-\lambda\begin{bmatrix}-\gamma^2\delta\mathbf{B} & 0\\ 0 & -\mathbf{I}\end{bmatrix}\right| = 0. \tag{A.54}$$

This solution will give Eigen values that need to be scaled to obtain the wave number, that is, for each Eigen value λ, the wave number k is given by

$$k = \gamma\lambda. \tag{A.55}$$

If the Earth model is parameterised in terms of N spectral element nodes, the Eigen vectors of this solution will be of dimension of $4N$. The first N elements of the Eigen vector is the amplitude function for the lateral displacement and the next N are for the radial displacement.

Appendix B

Synthetic surface wave inversion results

B.1 Introduction

This appendix contains extra results from the synthetic inversions in Chapter 8.

The figures are grouped into a grid layout with the rows corresponding to the three priors used, a uniform prior, a Beta prior with the mode set to a reference model and a concentration parameter of 4, and a Beta prior with the mode set to a reference model and a concentration parameter of 8. The columns, from left to right are the three different trans-dimensional inversions. Hence, the lettering in the figures to follow is

(a) Fixed `0`th order polynomial inversion with uniform prior,

(b) Trans-dimensional `0`th or 1st order polynomial inversion with uniform prior,

(c) Trans-dimensional `0`th, 1st, or 2nd order polynomial inversion with uniform prior,

(d) Fixed `0`th order polynomial inversion with a Beta prior and concentration parameter of four,

(e) Trans-dimensional `0`th or 1st order polynomial inversion with a Beta prior and concentration parameter of four,

(f) Trans-dimensional `0`th, 1st, or 2nd order polynomial inversion with a Beta prior and concentration parameter of four,

(g) Fixed `0`th order polynomial inversion with a Beta prior and concentration parameter of eight,

(h) Trans-dimensional `0`th or 1st order polynomial inversion with a Beta prior and concentration parameter of eight,

(i) Trans-dimensional `0`th, 1st, or 2nd order polynomial inversion with a Beta prior and concentration parameter of eight.

There are figures are grouped in order by the Love wave, Rayleigh wave and joint Love and Rayleigh wave inversions. The first set of plots are the hierarchical histogram plots. Successful inversion is indicated by a histogram centred about unity.

The second set of plots is the posterior dispersion histograms demonstrating the fit to the observations. In these plots, the true dispersion is shown with a red dotted line, the observations and their errors are shown with black dots and whiskers and the posterior histogram of the dispersion curve(s) are shown with blue shading representing posterior probabilities (dark blue corresponds to more likely).

The third set of plots is the posterior histogram on the number of partitions in the model and the fourth set of plots is the number of parameters. The number of parameters for each model which is computed using

$$N_p = N_c + \sum_{i}^{N_c} (p_i + 1) + 1, \tag{B.1}$$

where N_c is the number of cells, p_i is the polynomial order in cell i. The addition of 1 captures the shear wave velocity parameter in the half space. The true model consists of 6 parameters, that is 2 cells of 0th and 1st order.

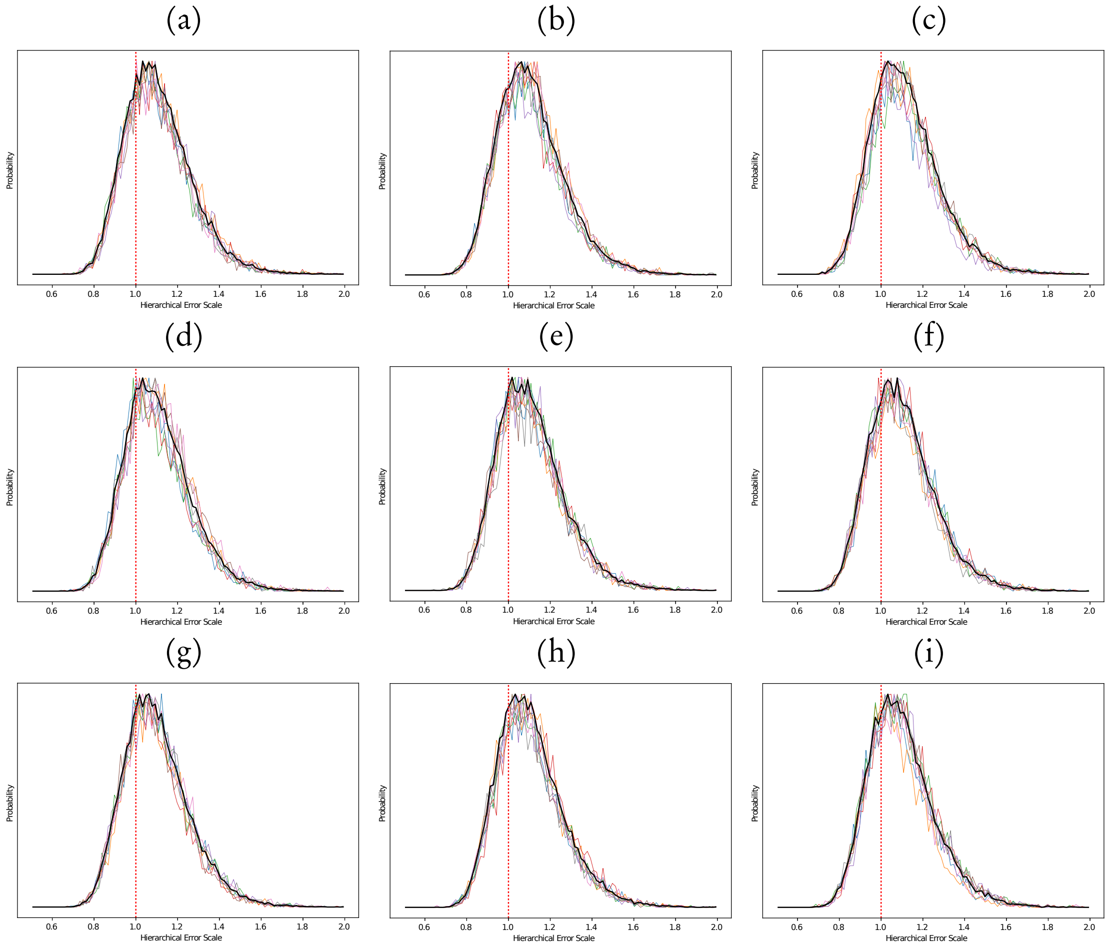
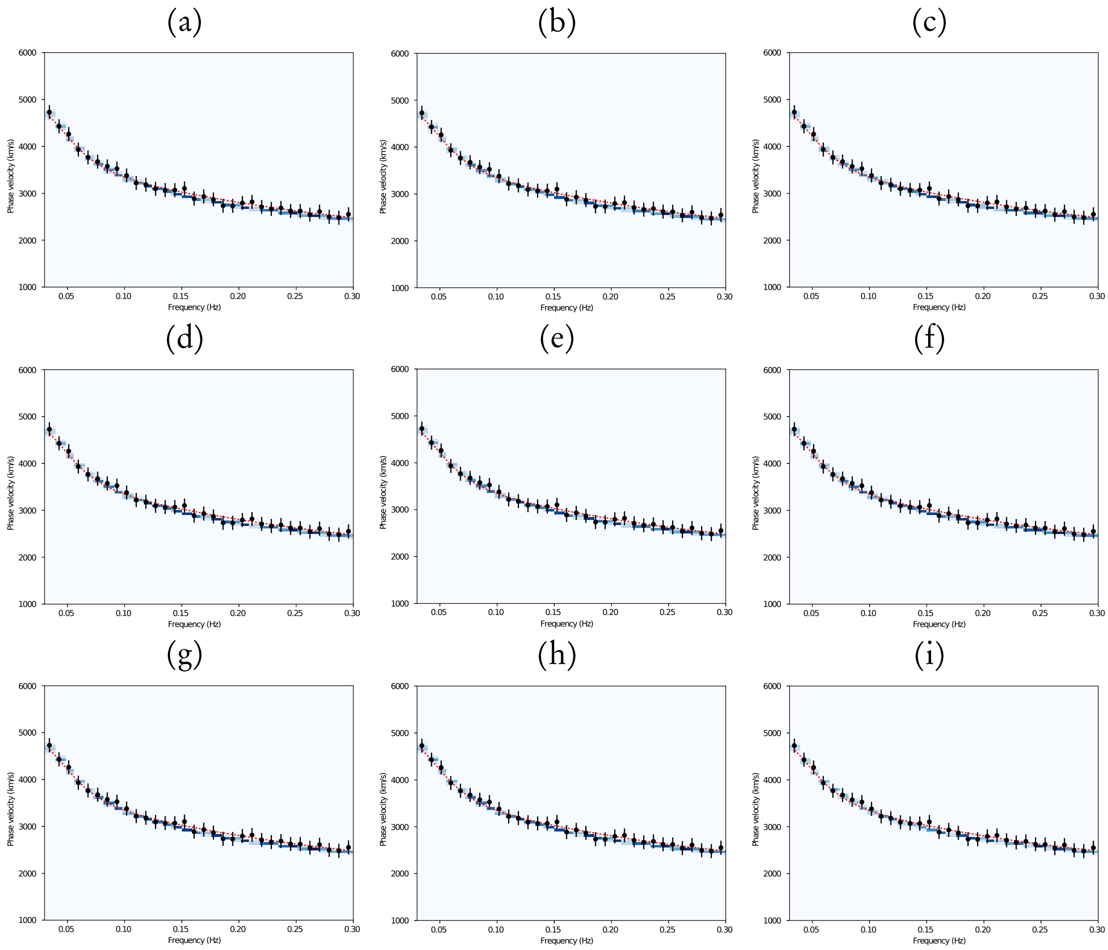


Figure B.1: Hierarchical error scale histogram for each of the Love wave inversions. The coloured lines represent each of the independent chains and the solid black line is the combined histogram across chains.


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure B.2: Love wave dispersion histogram with blue shading representing regions of higher probability, the black points are the observations with error scale marked and the red dashed line is the true model.

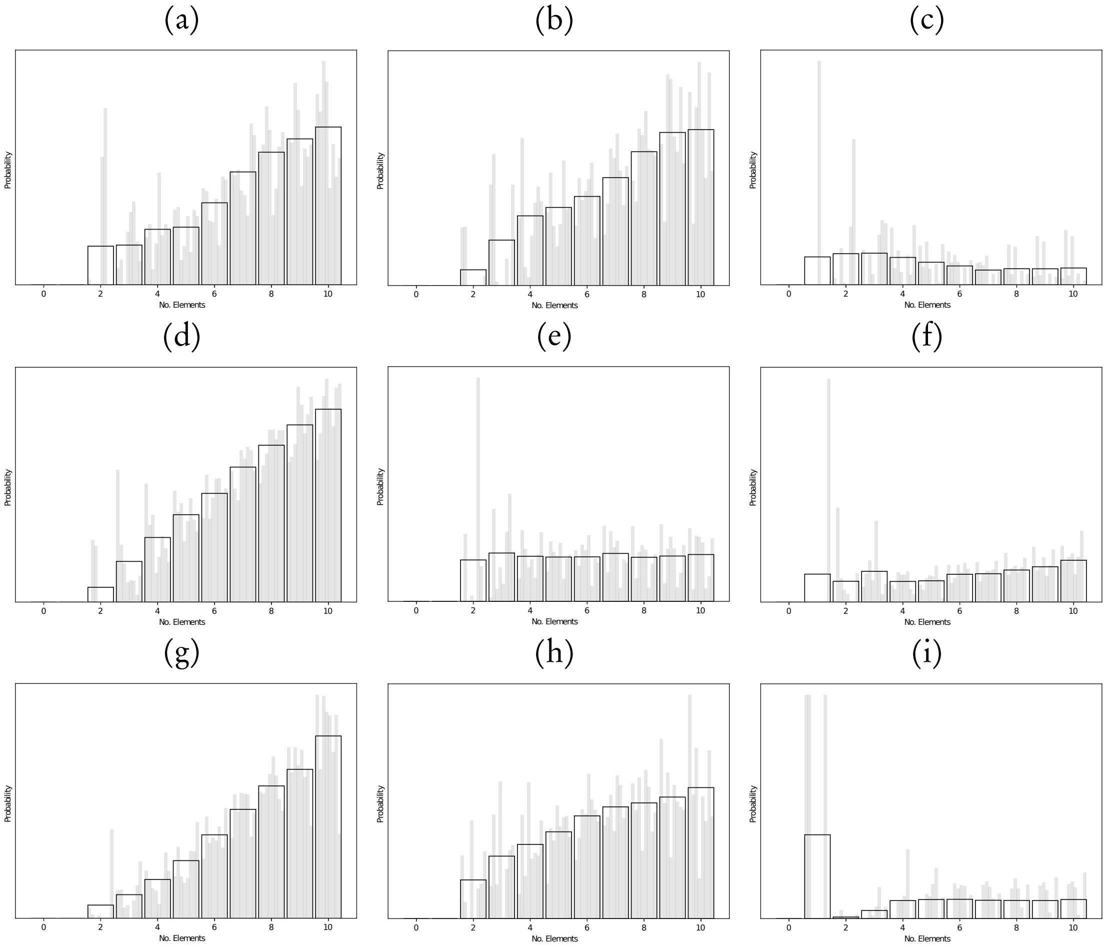


Figure B.3: Love wave histogram of the number of cells.

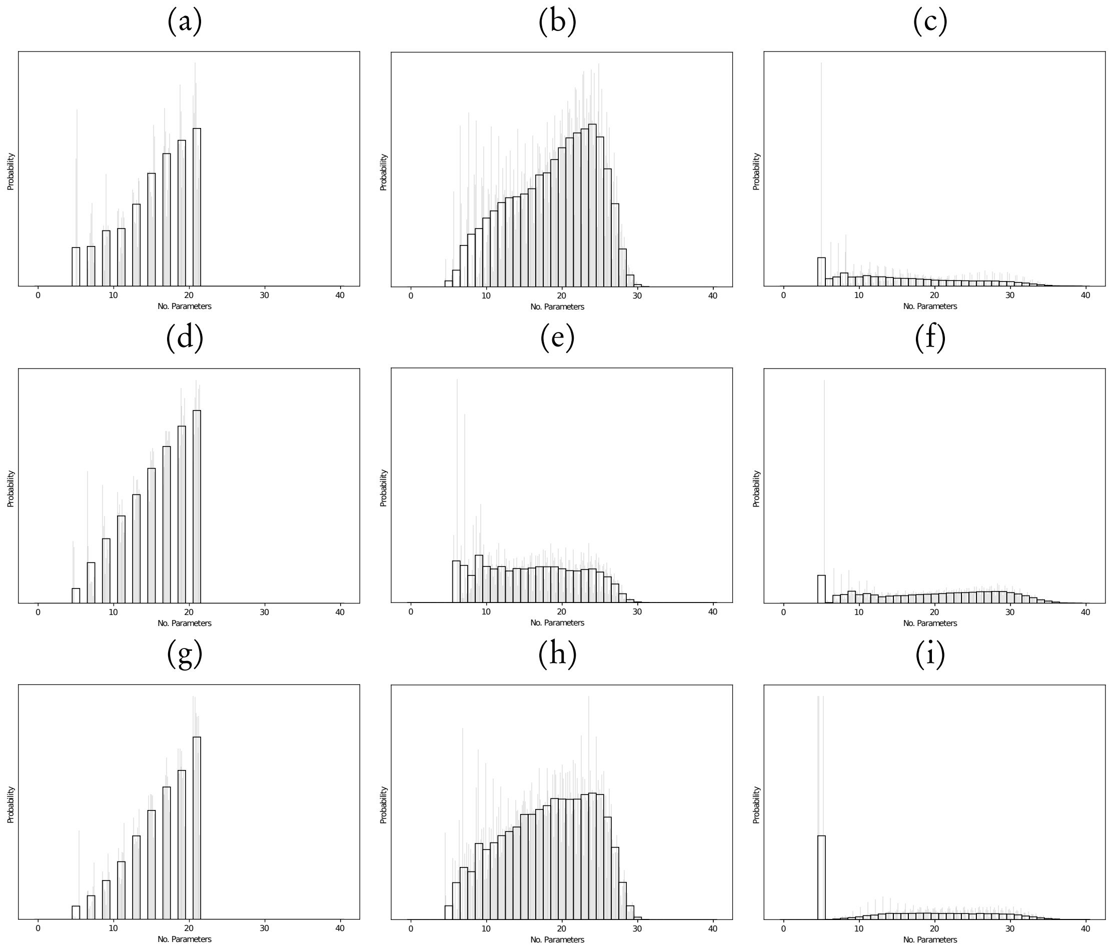


Figure B.4: Love wave histogram of the number of parameters.

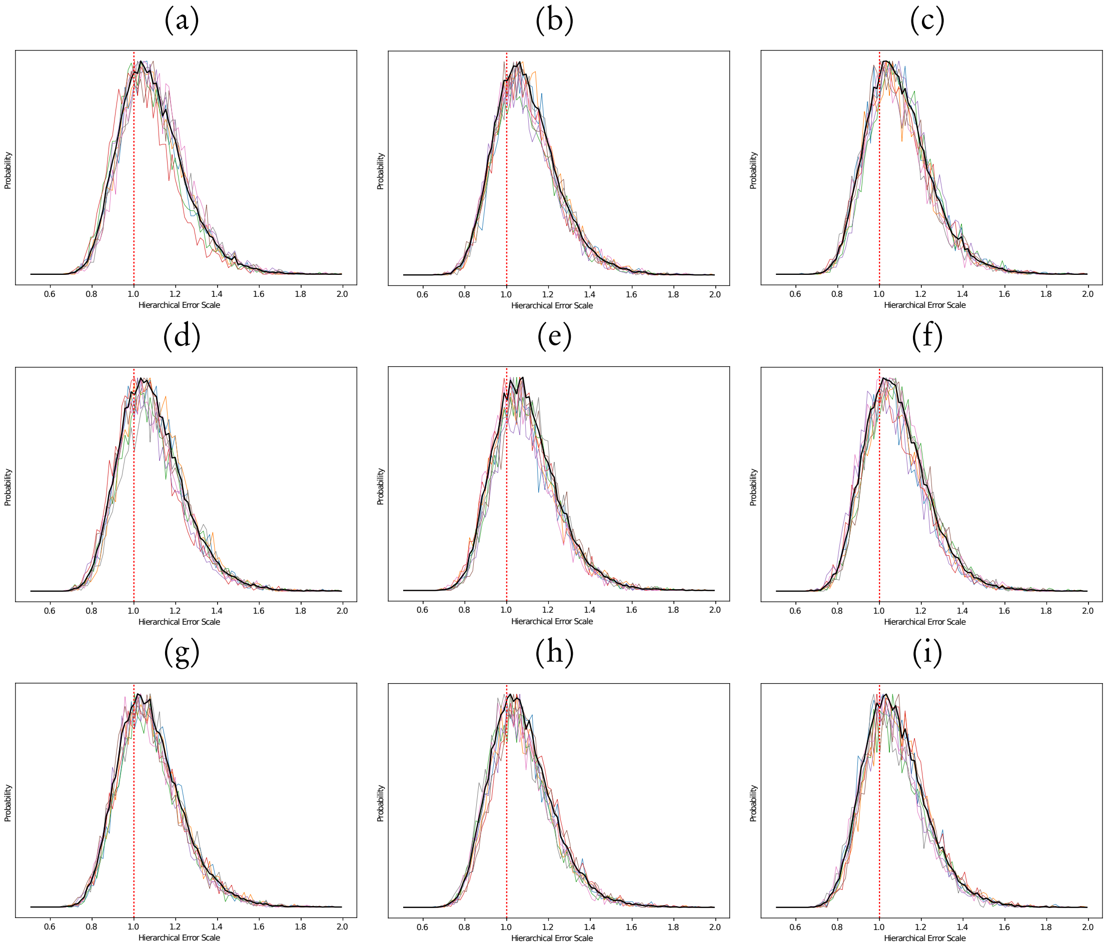


Figure B.5: Hierarchical error scale histogram for each of the Rayleigh wave inversions. The coloured lines represent each of the independent chains and the solid black line is the combined histogram across chains.

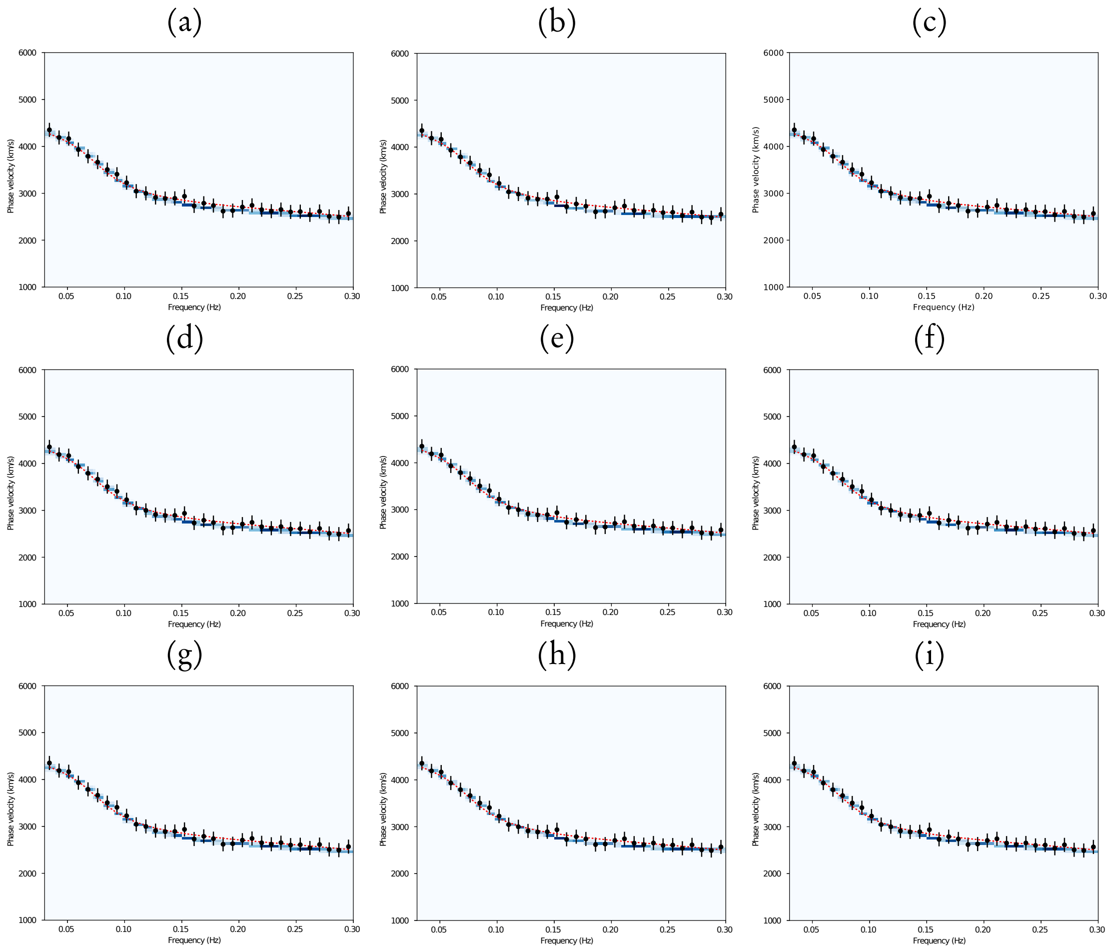


Figure B.6: Rayleigh wave dispersion histogram with blue shading representing regions of higher probability, the black points are the observations with error scale marked and the red dashed line is the true model.

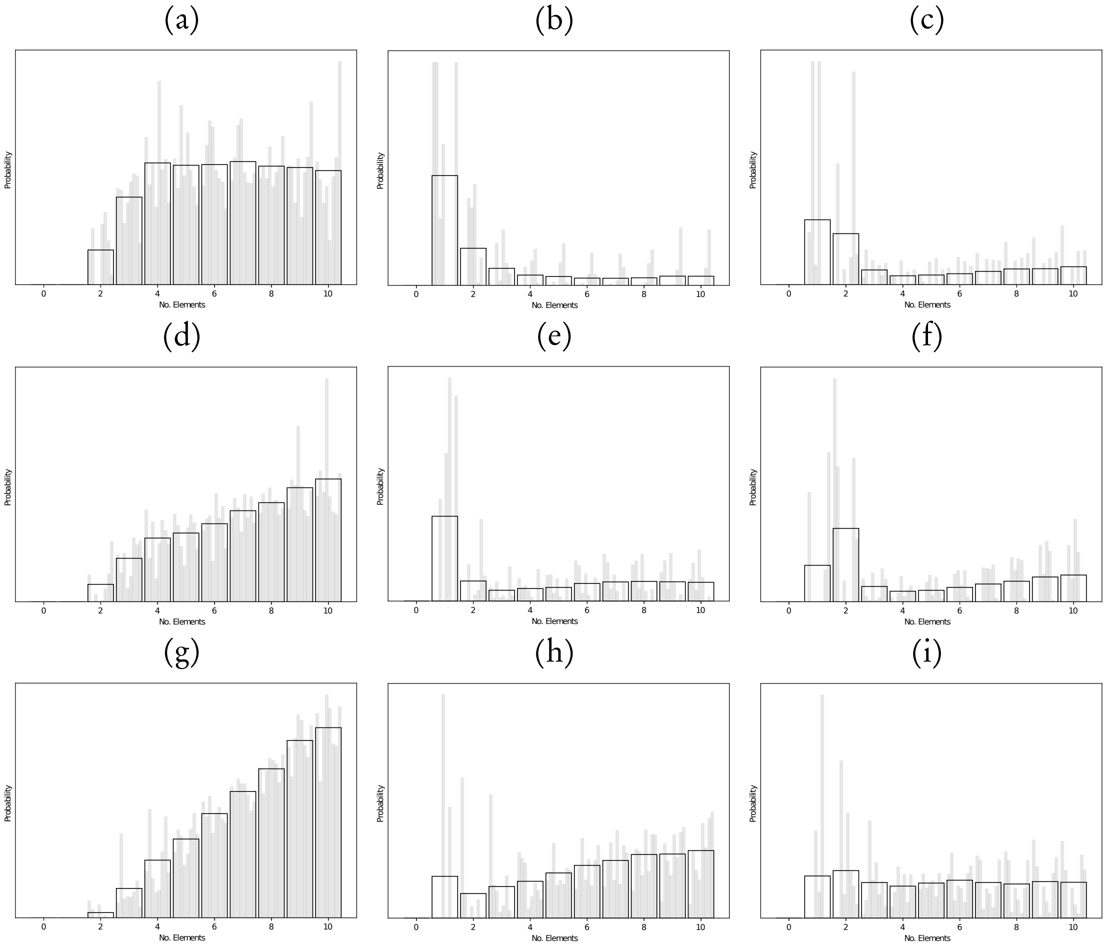


Figure B.7: Rayleigh wave histogram of the number of cells.

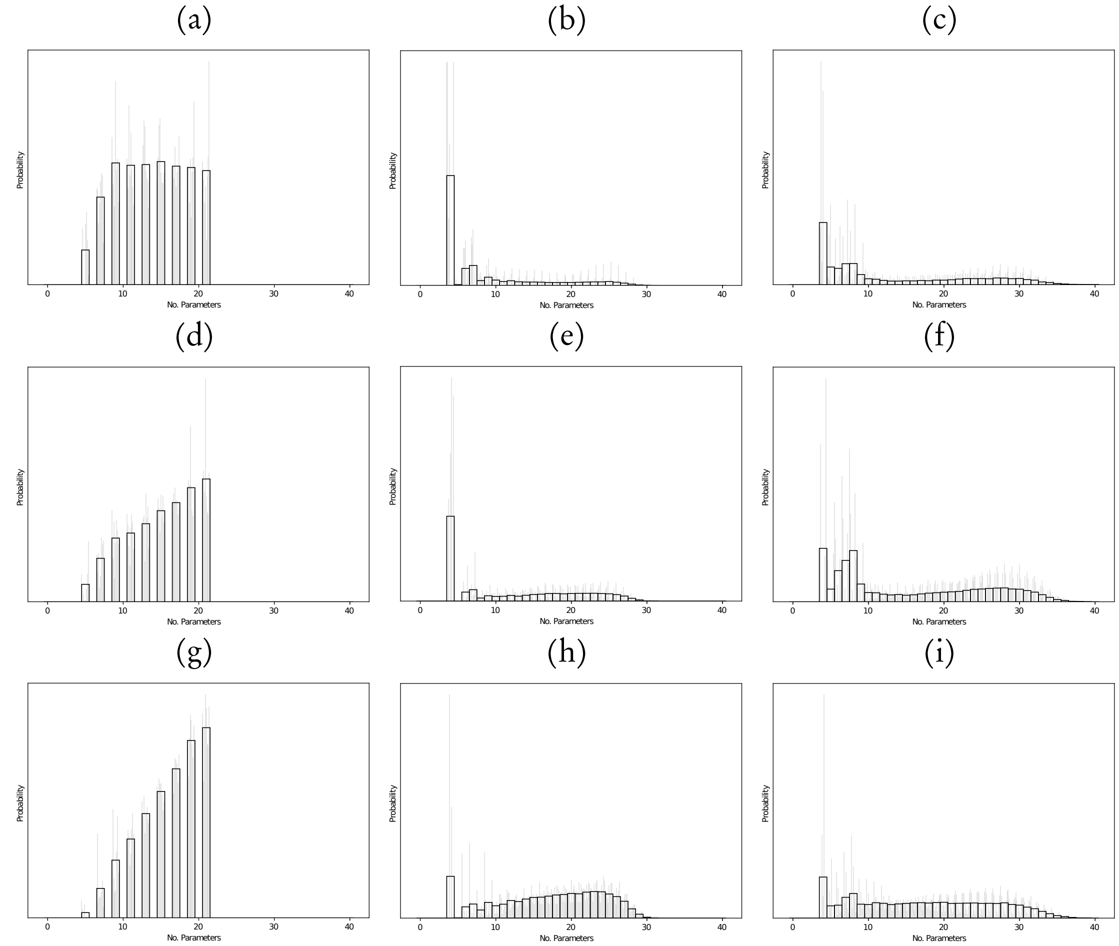


Figure B.8: Rayleigh wave histogram of the number of parameters.

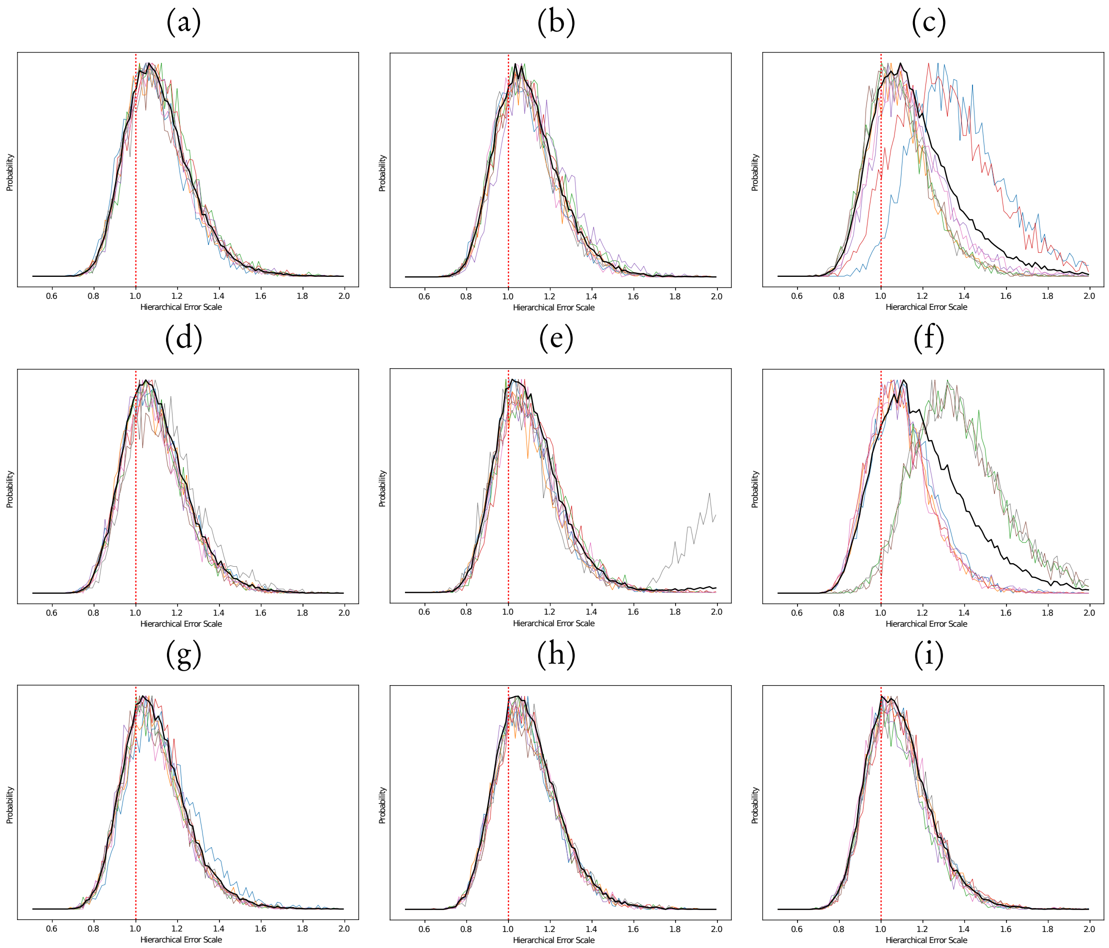


Figure B.9: Histogram of the hierarchical Love wave error scale for each Joint inversions. The coloured lines represent each of the independent chains and the solid black line is the combined histogram across chains.

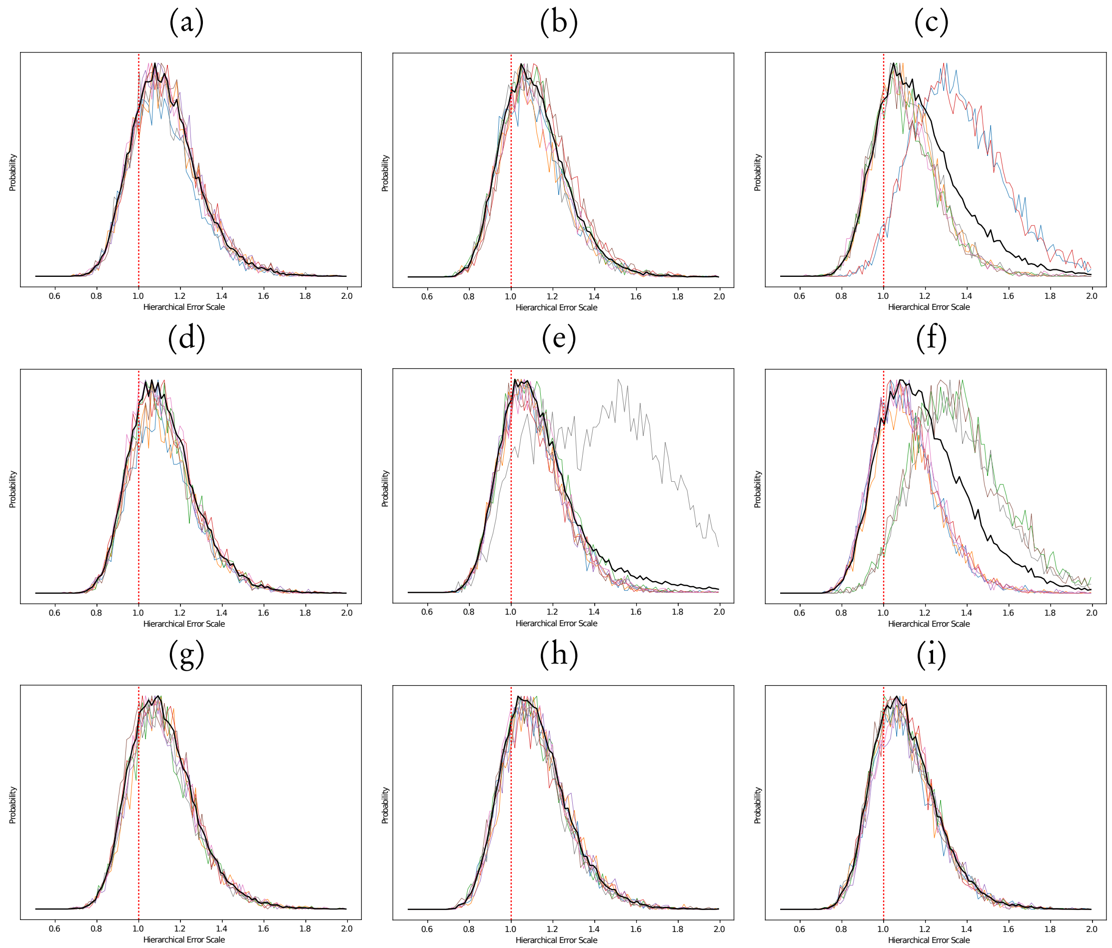


Figure B.10: Histogram of the hierarchical Rayleigh wave error scale for each Joint inversions. The coloured lines represent each of the independent chains and the solid black line is the combined histogram across chains.

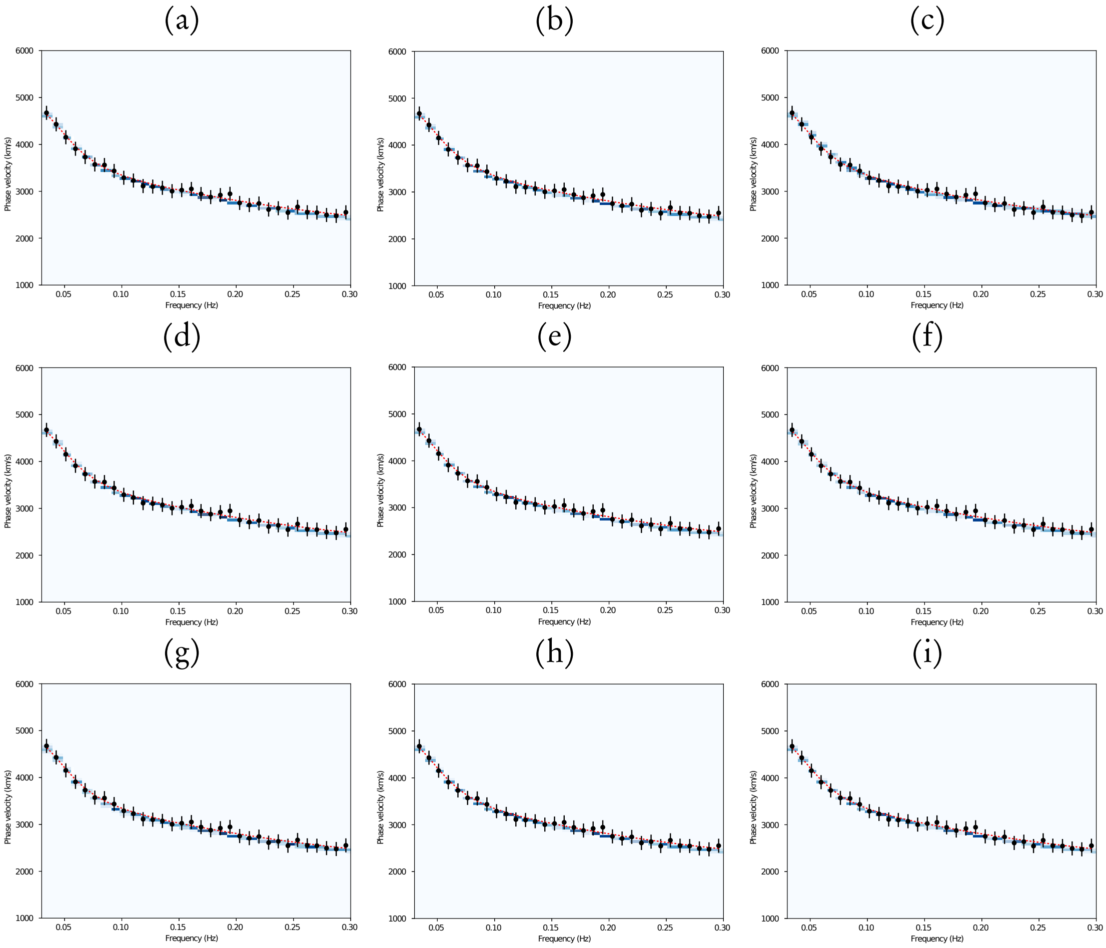


Figure B.11: Love wave dispersion histograms from the joint inversion with blue shading representing regions of higher probability, the black points are the observations with error scale marked and the red dashed line is the true model.

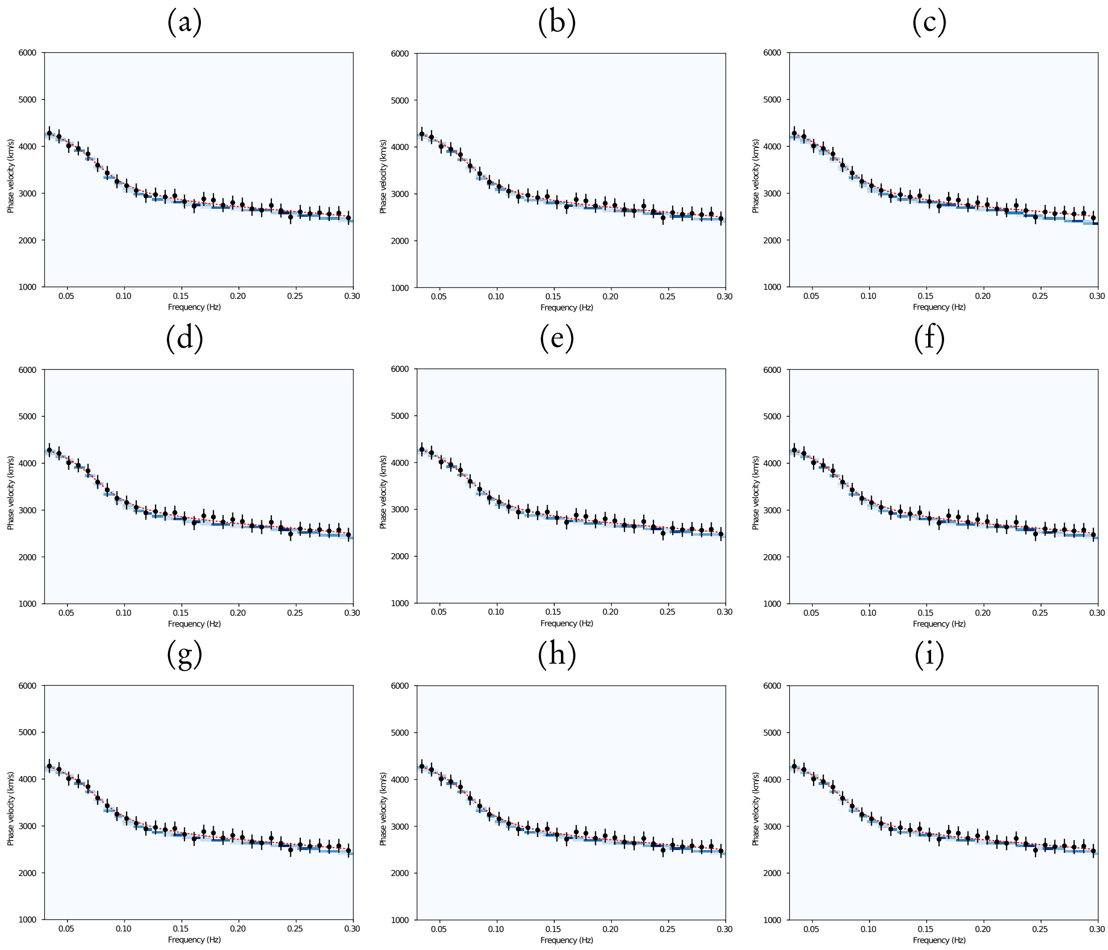


Figure B.12: Rayleigh wave dispersion histograms from the joint inversion with blue shading representing regions of higher probability, the black points are the observations with error scale marked and the red dashed line is the true model.

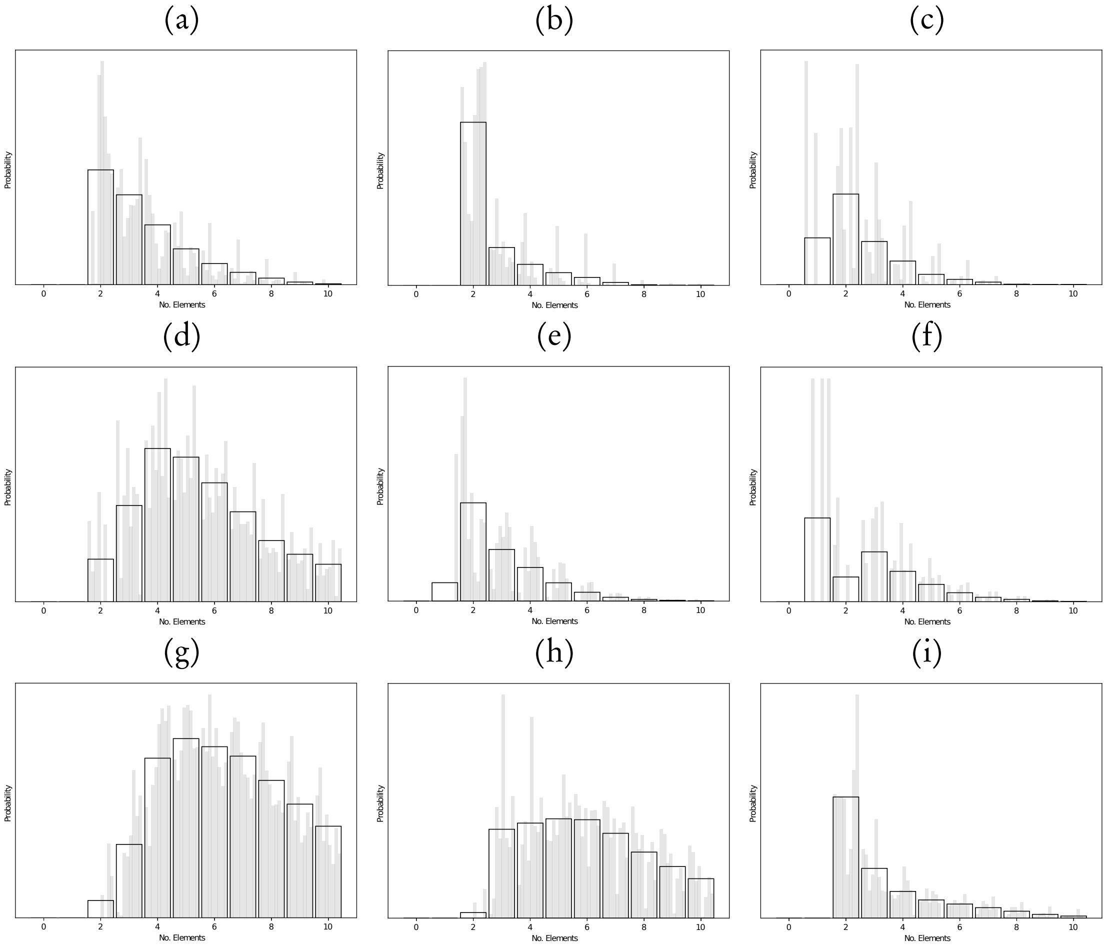


Figure B.13: Joint inversion histogram of the number of cells.

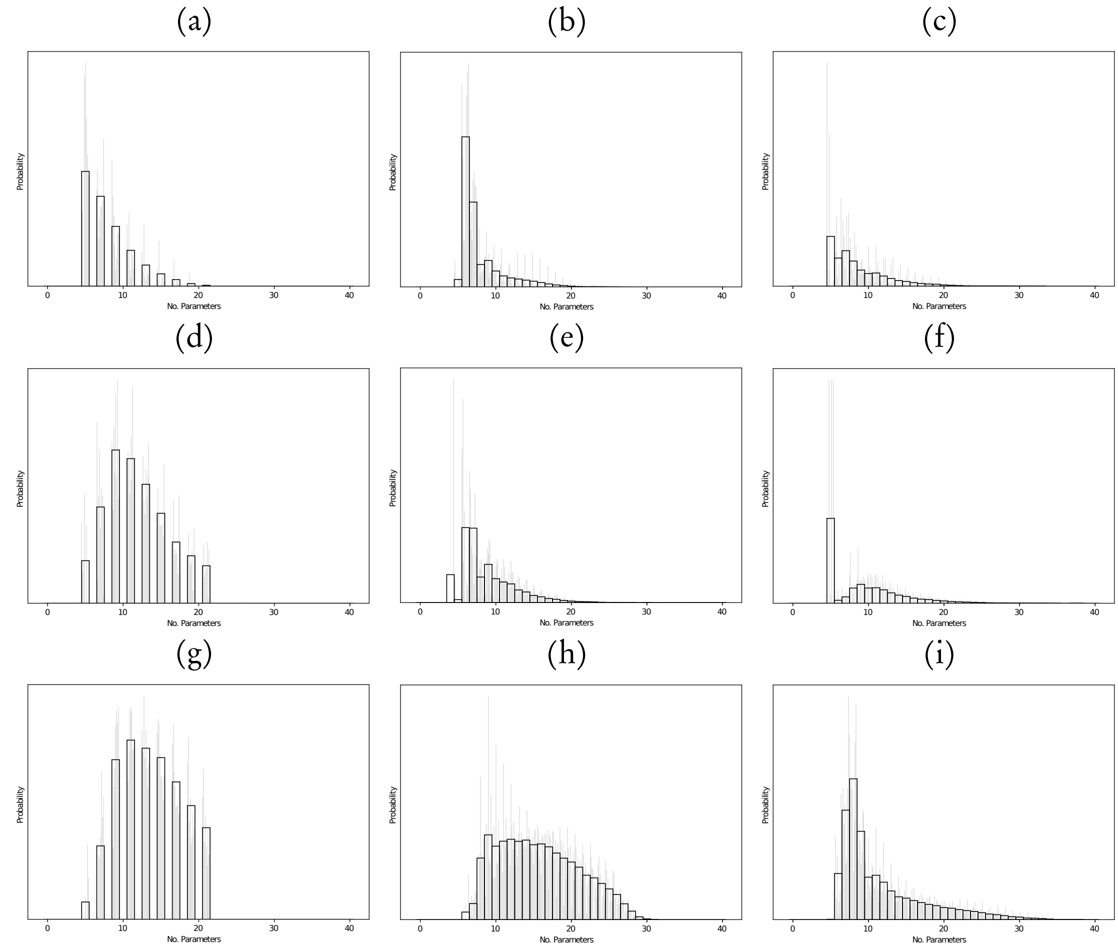


Figure B.14: Joint inversion histogram of the number of parameters.

Appendix C

Research output during PhD

During the course of this study, I was fortunate to be able to contribute to the following Journal artical publications:

Heslop D., Roberts A. P., and Hawkins R., *A statistical simulation of magnetic particle alignment in sediments.*, Geophysical Journal International, 197 (2), 828–837, 2014.

Iaffaldano G., Hawkins R., and Sambridge M. *Bayesian noise-reduction in Arabia/Somalia and Nubia/Arabia finite rotations since 20Ma: Implications for Nubia/Somalia relative plate motions.*, *Geochemistry, Geophysics, Geosystems*, 15 (4): 845 – 854, 2014.

Iaffaldano G., Hawkins R., Bodin T., and Sambridge M., *REDBACK: Open source software for efficient noise-reduction in plate kinematic reconstructions. Geochemistry, Geophysics, Geosystems*, 15 (4): 1663 – 1670, 2014

Ingham L., Heslop D., Roberts A. P., Hawkins R., and Sambridge M., *Is there a link between geomagnetic reversal frequency and paleointensity? a Bayesian approach.*, *Journal of Geophysical Research: Solid Earth*, 119 (7): 5290 – 5304, 2014.

Charles C., Bish A., Boswell R., Dedrick J., Greig A., Hawkins R., and Ho T. S., *A short review of experimental and computation diagnostics for radiofrequency plasma micro-thrusters.*, *Plasma Chemistry and Plasma Processing*, 1–16, 2015

Charles C., Hawkins R., and Boswell R., *Particle in cell simulation of a radio frequency plasma jet expanding in vacuum. Applied Physics Letters*, 106 (9), 2015

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

Saygin E., Cummins P. R., Cipta A., Hawkins R., Pandhu R., Murjaya J., Masturyono, Irsyam M., Widiyantoro S., and Kennett B.L.N., *Imaging architecture of the Jakarta basin, indonesia with transdimensional inversion of seismic noise. Geophysical Journal International*, 204 (2): 918–931, 2016.

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

Hayward K. S., Cox S. F., Fitz Gerald J. D., Slagmolen B. J. J., Shaddock D. A., Forsyth P. W. F., Salmon M. L., and Hawkins R. P. *Mechanical amorphization, flash heating, and frictional melting: Dramatic changes to fault surfaces during the first millisecond of earthquake slip. Geology*, 2016.

Kallenberg B., Tregoning P., Hoffman J. F., Hawkins R., Purcell A., Allgeyer S., and Koulali A. *A new approach to estimate ice dynamic rates using satellite observations.*, *Cryosphere*, 2017.

Hawkins R., Brodie R., and Sambridge M. *Trans-dimensional Bayesian inversion of airborne electromagnetic data for 2D conductivity profiles. Exploration Geophysics*, 2017.

Bibliography

J. D. Achenbach. *Wave propagation in elastic solids*. Elselvier, 1975.

T. E. Ackman. An introduction to the use of airborne technologies for watershed characterization in mined areas. *Mine Water and the Environment*, 22:62 – 68, 2003.

N. Ahmed, T. Nataraja, and K. R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, pages 90–93, 1974.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6):716–723, 1974.

K. Aki. Space and time spectral of stationary stochastic waves with special reference to microtremors. *Bulletin of the Earthquake Research Institute, University of Tokyo*, 35: 415 – 456, 1957.

K. Aki. Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, 82(2):277 – 296, 1977.

K. Aki and P. G. Richards. *Quantitative Seismology*. University Science Books, California, 2nd edition, 2002.

F. Al-Awadhi, M. Hurn, and C. Jennison. Improving the acceptance rate of reversible jump MCMC proposals. *Statistics and Probability Letters*, 69:189 – 198, 2004.

P. Alfeld, M. Neamtu, and L. L. Schumaker. Bernstein-Bézier polynomials on spheres and sphere-like surfaces. *Computer Aided Geometric Design*, 13(4):333 – 349, 1996.

R. M. Allen, G. Nolet, W. J. Morgan, K. Vogfjörd, B. H. Bergsson, P. Erlendsson, G. R. Foulger, Jakobsdóttir, B. R. Julian, M. Pritchard, S. Ragnarsson, and R. Stefánsson. The thin hot plume beneath Iceland. *Geophysical Journal International*, 137:51 – 63, 1999.

Z. Alterman, H. Jarosch, and C. L. Pekeris. Oscillations of the Earth. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 252(1268):80 – 95, 1959.

A. Amirbekyan and V. Michel. Splines on the three-dimensional ball and their application to seismic body wave tomography. *Inverse Problems*, 24:1 – 25, 2008. doi: 10.1088/0266-5611/24/1/015022.

D. L. Anderson. Elastic wave propagation in layered anisotropic media. *Journal of Geophysical Research*, 66(9):2953 – 2963, 1961.

T. Ando. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, 94(2):443–458, 2007.

T. Ando. *Bayesian Model Selection and Statistical Modeling*. CRC Press, 2010.

C. Andrieu and J. Thoms. A tutorial on adaptive mcmc. *Statistical Computing*, 18: 343–373, 2008. doi: 10.1007/s11222-008-9110-y.

C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3): 269–342, 2010.

J.-P. Antoine and P. Vandergheynst. Wavelets on the 2-sphere: A group-theoretical approach. *Applied and Computational Harmonic Analysis*, 7:262–291, 1999.

M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using vector quantization in the wavelet transform domain. In *International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1990.

M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2):205–220, 1992.

Y. F. Atchade and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

J. C. Aval. Multivariate Fuss-Catalan numbers. *Discrete Mathematics*, 308(20):4660–4669, 2008.

G. Backus. Inference from inadequate and inaccurate data I. *Proceedings of the National Academy of Sciences*, 65(1):1–7, 1970a.

G. Backus. Inference from inadequate and inaccurate data II. *Proceedings of the National Academy of Sciences*, 65(2):281–287, 1970b.

G. Backus and F. Gilbert. The resolving power of gross earth data. *Geophys. J. R. astr. Soc.*, 16:169–205, 1968.

G. E. Backus. Possible forms of seismic anisotropy of the uppermost mantle under oceans. *Journal of Geophysical Research*, 70(14):3439, 1965.

J. R. Baumgardner and P. O. Frederickson. Icosahedral discretization of the two–sphere. *SIAM J. Numer. Anal.*, 22(6):1107–1115, 1985.

T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53:370–418, 1763.

G. D. Bensen, M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Yang. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, 169:1239 – 1260, 2007. doi: 10.1111/j.1365-246X.2007.03374.x.

P. Berkel, D. Fischer, and V. Michel. Spline multiresolution and numerical results for joint gravitational and normal-mode inversion with and outlook on sparse regularisation. *Int J Geomath*, 1:167 – 204, 2011. doi: 10.1007/s13137-010-0007-5.

K. Black. Spectral elements on infinite domains. *SIAM J. Sci. Comput.*, 19(5):1667 – 1681, 1998.

R. S. Blewett. Unlocking australia's hidden mineral resource potential. `http://www.ga.gov.au/scientific-topics/minerals/unlocking-resource-potential`, 2017. [Online; accessed 20-September-2017].

S. Bloch and A. L. Hales. New techniques for the determination of surface wave velocities. *Bulletin of the Seismological Society of America*, 58(3):1021 – 1034, 1968.

T. Bodin and M. Sambridge. Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, 178:1411 – 1436, 2009.

T. Bodin, M. Sambridge, and K. Gallagher. A self-parametrizing partition model approach to tomographic inverse problems. *Inverse Problems*, 25:055009, 2009. doi: 10.1088/0266-5611/25/5/055009.

T. Bodin, M. Sambridge, N. Rawlinson, and P. Arroucau. Transdimensional tomography with unknown data noise. *Geophysical Journal International*, 189:1536–1556, 2012a. doi: 10.1111/j.1365-246X.2012.05414.x.

T. Bodin, M. Sambridge, H. Tkalčić, P. Arroucau, L. Gallagher, and N. Rawlinson. Trans-dimensional inversion of receiver functions and surface wave dispersion. *Journal of Geophysical Research*, 117:B02301, 2012b. doi: 10.1029/2011JB008560.

T. Bodin, J. Leiva, B. Romanowicz, V. Maupin, and H. Yuan. Imaging anisotropic layering with Bayesian inversion of multiple data types. *Geophysical Journal International*, 206:605–629, 2016.

I. Bogdanova, P. Vandergheynst, J.-P. Antoine, L. Jacques, and M. Morvidone. Stereographic wavelet frames on the sphere. *Applied and Computational Harmonic Analysis*, 19:223–252, 2005.

G. P. Bonneau. Optimal triangular haar bases for spherical data. In *Visualization*. IEEE, 1999.

L. Bos, M. A. Taylor, and B. A. Wingate. Tensor product Gauss-Lobatto points are Fekete points for the cube. *Mathematics of Computation*, 70(236):1543 – 1547, 2000.

L. Boschi, C. Weemstra, J. Verbeke, G. Ekstr om, A. Zunino, and D. Giardini. On measuring surface wave phase velocity from station-station cross-correlation of ambient signal. *Geophysical Journal International*, 192:346 – 358, 2013. doi: 10.1093/gji/ggs023.

G. E. P. Box and N. R. Draper. *Empirical Model Building and Response Surfaces*. John Wiley and Sons, New York, 1987.

J. P. Boyd. Orthogonal rational functions on a semi-infinite interval. *Journal of Computational Physics*, 70:63 – 88, 1987.

J. P. Boyd. *Chebyshev and Fourier spectral methods*. Dover, 2 edition, 2001.

T. M. Brocher. Empirical relationships between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America*, 95(6):2081 – 2092, 2005. doi: 10.1785/0120050077.

R. Brodie and M. Sambridge. A holistic approach to inversion of frequency-domain airborne EM data. *Geophysics*, 71(6):G301 – G312, 2006. doi: 10.1190/1.2356112.

R. Brodie and M. Sambridge. Transdimensional Monte Carlo inversion of AEM data. In *22nd International Geophysical Conference and Exhibition*. Australian Society of Exploration Geophysics, 2012.

R. C. Brodie. *Holistic inversion of airborne electomagnetic data*. PhD thesis, Australian National University, 2010.

R. C. Brodie. Geoscience Australia AEM source code repository. `https://github.com/GeoscienceAustralia/ga-aem`, 2016. [Online; accessed 20-September-2016].

S. Brooks, A. Gelman, G. L. Jones, and X. Meng, editors. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.

S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434 – 455, 1998.

C. Brunetti, N. Linde, and J. A. Vrugt. Bayesian model selection in hydrogeophysics: Application to conceptual subsurface models of the South Oyster Bacterial Transport Site, Virginia, USA. *Advances in Water Resources*, 102:127 – 141, 2017.

E. C. Bullard, T. F. Gaskell, W. B. Harland, and C. Kerr-Grant. Seismic investigations on the Paleozoic floor of East England. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 239:4 – 94, 1940.

P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communcations*, 31(4):532–540, 1983.

L. Cagniard. Basic theory of the magnetotelluric method of geophysical prospecting. *Geophysics*, 18:605 – 635, 1953.

B. Calderhead. A general construction for parallelizing Metropolis-Hastings algorithms. *Proceedings of the National Academy of Sciences*, 111(49):17408–17413, 2014. doi: 10.1073/pnas/1408184111.

M. Campillo and A. Paul. Long-range correlation in the diffuse seismic coda. *Science*, 299:547 – 549, 2003. doi: 10.1126/science.1078551.

E. J. Candes and D. L. Donoho. Curvelets: a surprisingly effective nonadaptive representations for objects with edges. In *Curves and Surface Fitting : Saint-Malo 1999*, 1999.

E. J. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8): 1207–1223, 2006.

E. Catalan. Note extraite d'une lettre adressée à l'éditeur par Mr. E. Catalan, répétiteur à l'école polytechnique de paris. *Journal für die reine und angewandte Mathematik*, 27:192–192, 1844.

V. C̀ervený. *Seismic Ray Theory*. Cambridge University Press, 2001.

V. C̀ervený and J. E. P. Soares. Fresnel volume ray tracing. *Geophysics*, 57(7):902 – 915, 1992.

A. Chambodut, I. Panet, M. Mandea, M. Diament, M. Holschneider, and O. Jamet. Wavelet frames: an alternative to spherical harmonic representation of potential spheres. *Geophysical Journal International*, 163:875–899, 2005.

J.-R. Chang, W.-J. Lin, C.-J. Huang, and S.-T. Choi. Vibration and stability of an axially moving Rayleigh beam. *Applied Mathematical Modelling*, 34:1482 – 1497, 2010. doi: 10.1016/j.apm.2009.08.022.

C. H. Chapman. *Fundamentals of Seismic wave propagation*. Cambridge University Press, 2004.

A. Charlety, S. Voronin, G. Nolet, I. Loris, F. J. Simons, K. Sigloch, and I. C. Daubechines. Global seismic tomography with sparsity constraints: Comparison with smoothing and damping regularization. *Journal of Geophysical Research: Solid Earth*, 118:4887 – 4899, 2013.

A. D. Chave and A. G. Jones, editors. *The Magnetotelluric method*. Cambridge, 2012.

S. Chevrot, R. Martin, and D. Komatitsch. Optimized discrete wavelet transforms in the cubed sphere with the lifting scheme - implications for global finite-frequency tomography. *Geophysical Journal International*, 191:1391–1402, 2012.

L. Chiao and B. Kuo. Multiscale seismic tomography. *Geophysical Journal International*, 145:517 – 527, 2001.

D. L. Chopp. Some improvements to the fast marching method. *SIAM Journal of Scientific Computation*, 23(1):230–244, 2001.

A. Cohen, I. Daubechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, XLV:485 – 560, 1992.

S. C. Constable, R. L. Parker, and C. G. Constable. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 52(3):289 – 300, 1987.

H. Cox. Spatial correlation in arbitrary noise fields with applications to ambient sea noise. *J. Acoust. Soc. Am.*, 54:1289 – 1301, 1973.

F. A. Dahlen and F. J. Simons. Spectral estimation on the sphere in geophysics and cosmology. *Geophysical Journal International*, 174:774–807, 2008.

F. A. Dahlen, S.-H. Hung, and G. Nolet. Fréchet kernels for finite-frequency travel times - I. theory. *Geophysical Journal International*, pages 157–174, 2000.

F. A. Darbyshire, I. T. Bjarnason, R. S. White, and Ó. G. Flóvenz. Crustal structure above the Iceland mantle plume imaged by the ICEMELT refraction profile. *Geophysical Journal International*, 135:1131 – 1149, 1998.

F. A. Darbyshire, R. S. White, and K. F. Priestly. Structure of the crust and uppermost mantle of Iceland from a combined seismic and gravity study. *Earth and Planetary Science Letters*, 181:409 – 428, 2000.

I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909 – 996, 1988.

I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *Journal of Fourier Analysis and Applications*, 4.3:247 – 269, 1998.

R. W. L. de Wit, J. Trampert, and R. D. van der Hilst. Toward quantifying uncertainty in travel time tomography using the null-space shuttle. *Journal of Geophysical Research*, 117:B03301, 2012. doi: 10.1029/2011JB008754.

M. M. Deal and G. Nolet. Nullspace shuttles. *Geophysical Journal International*, 124: 372 – 380, 1996.

E. Debayle and Y. Ricard. A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements. *Journal of Geophysical Research*, 117:B10308, 2012. doi: 10.1029/2012JB009288.

E. Debayle and M. Sambridge. Inversion of massive surface wave data sets: Model construction and resolution assessment. *Journal of Geophysical Research*, 109:B02316, 2004. doi: 10.1029/2003JB002652.

D. G. T. Denison, B. K. Mallick, and A. F. M. Smith. A Bayesian CART algorithm. *Biometrika*, 85(2):363 – 377, 1998.

D. G. T. Denison, C. C. Holmes, B. K. Mallick, and A. F. M. Smith. *Bayesian Methods for Non-linear Classification and Regression*. John Wiley and Sons, 2002.

A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink. Recovering the Green's function from field-field correlations in an open scattering medium(l). *J. Acoust. Soc. Am.*, 133(6):2973 – 2976, 2003. doi: 10.1121/1.1570436.

F. Deschamps, R. Snieder, and J. Trampert. The relative density-to-shear velocity scaling in the uppermost mantle. *Physics of the Earth and Planetary Interiors*, 124:193 – 211, 2001.

J. Dettmer and S. E. Dosso. Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains. *J. Acoust. Soc. Am.*, 132(4):2239–2250, 2012.

J. Dettmer, S. E. Dosso, and C. W. Holland. Sequential trans-dimensional Monte Carlo for range-dependent geoacoustic inversion. *J. Acoust. Soc. Am.*, 129(4):1794–1806, 2011.

J. Dettmer, S. Molnar, G. Steininger, S. E. Dosso, and J. F. Cassidy. Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models. *Geophysical Journal International*, 188:719 – 734, 2012.

J. Dettmer, R. Benavente, P. R. Cummins, and M. Sambridge. Trans-dimensional finite-fault inversion. *Geophysical Journal International*, 199:735–751, 2014.

J. Dettmer, S. E. Dosso, T. Bodin, J. Stipčević, and P. R. Cummins. Direct-seismogram inversion for receiver-side structure with uncertain source-time functions. *Geophysical Journal International*, 203:1373 – 1387, 2015.

J. Dettmer, R. Hawkins, P. R. Cummins, J. Hossen, M. Sambridge, R. Hino, and D. Inazu. Tsunami source uncertainty estimation: The 2011 Japan tsunami. *Journal of Geophysical Research: Solid Earth*, 121:4483 – 4505, 2016. doi: 10.1002/2015JB012764.

E. W. Djikstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959. doi: 10.1007/BF01386390.

S. E. Dosso, C. W. Holland, and M. Sambridge. Parallel tempering in strongly nonlinear geoacoustic inversion. *Journal of the Acoustic Society of America*, 132(5):3030–3040, 2012.

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

R. L. Dougherty, A. Edelman, and J. M. Hyman. Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation. *Mathematics of Computation*, 52(186):471 – 494, 1989.

S. Duane, A. D. Kennedy, B. J. Pendelton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216 – 222, 1987.

S. Durand, E. Debayle, and Y. Ricard. Rayleigh wave phase velocity and error maps up to the fifth overtone. *Geophys. Res. Lett.*, 42, 2015. doi: 10.1002/2015GL063700.

N. Dyn, D. Levine, and J. A. Gregory. A butterfly subdivision scheme for surface interpolation with tension control. In *ACM transactions on Graphics (TOG)*, volume 9. ACM, 1990.

A. M. Dziewonski and D. L. Anderson. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, 25(4):297 – 356, 1981. doi: 10.1016/0031-9201(81)90046-7.

A. M. Dziewonski and A. L. Hales. *Methods in computational physics*, volume 11, chapter Numerical analysis of dispersed seismic waves, pages 39 – 85. Academic Press, 1972.

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

G. Ekström. Love and Rayleigh phase-velocity maps, 5 - 40s, of the western and central USA from USArray data. *Earth and Planetary Science Letters*, 402:42 – 49, 2014. doi: 10.1016/j.epsl.2013.11.022.

G. Ekström, G. A. Abers, and S. C. Webb. Determination of surface-wave phase velocities across USArray from noise and Aki's spectral formulation. *Geophysical Research Letters*, 36:L18301, 2009. doi: 10.1029/2009GL039131.

H. Fang and H. Zhang. Wavelet-based double-difference seismic tomography with sparsity regularization. *Geophysical Journal International*, 199:944–955, 2014. doi: 10.1093/gji/ggu305.

H. Fang, H. Yao, H. Zhang, Y.-C. Huang, and R. D. van der Hilst. Direct inversion of surface wave dispersion for three-dimensional shallow crustal structure based on ray tracing: methodology and application. *Geophysical Journal International*, 201:1251 – 1263, 2015. doi: 10.1093/gji/ggv080.

C. G. Farquharson and D. W. Oldenburg. Inversion of time-domain electromagnetic data for a horizontally layered Earth. *Geophysical Journal International*, 114:433 – 442, 1993.

C. G. Farquharson and D. W. Oldenburg. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophysical Journal International*, 156:411 – 425, 2004. doi: 10.1111/j.1365-246X.2004.02190.x.

G. E. Fasshauer and L. L. Schumaker. Scattered data fitting on the sphere. *Mathematical Methods for Curves and Surfaces II*, pages 117–166, 1998.

M. Fekete. Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. *Mathematische Zeitschrift*, 17(1):228 – 249, 1923.

A. Fichtner. *Full seismic waveform modelling and inversion*. Springer, 2011.

A. Fichtner and H. Igel. Efficient numerical surface wave propagation through the optimization of discrete crustal models - a technique based on non-linear dispersion curve matching (dcm). *Geophysical Journal International*, 173:519 – 533, 2008. doi: 10.1111/j.1365-246X.2008.03746.x.

A. Fichtner, H. Igel, H. P. Bunge, and B. L. Kennett. Simulation and inversion of seismic wave propagation on continental scales based on a spectral-element method. *Journal of Numerical Analysis, Industrial and Applied Mathematics*, 4(1-2):11–22, 2009.

D. V. Fitterman and M. Deszcz-Pan. Helicopter EM mapping of saltwater intrusion in Everglades National Park, Florida. *Exploration Geophysics*, 29(1):240 – 243, 1998.

G. R. Foulger, M. J. Pritchard, B. R. Julian, J. R. Evans, R. M. Allen, G. Nolet, W. J. Morgan, B. H. Bergsson, P. Erlendsson, S. Jakobsdottir, S. Ragnarsson, R. Stefansson, and Vogfjörd. Seismic tomography shows that upwelling beneath Iceland is confined to the upper mantle. *Geophysical Journal International*, 146:504 – 530, 2001.

F. N. Fritsch and J. Butland. A method for constructing local monotone piecewise cubic interpolants. *SIAM J. Sci. Stat. Comput.*, 5(2):300 – 304, 1984.

F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, 17(2), 1980.

E. Galetti, A. Curtis, G. Angelo Meles, and B. Baptie. Uncertainty loops in travel-time tomography from nonlinear wave physics. *Physical Review Letters*, 114:148501, 2015. doi: 10.1103/PhysRevLett.114.148501.

E. Galetti, A. Curtis, B. Baptie, D. Jenkins, and H. Nicolson. 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*, 2016. doi: 10.1093/gji/ggw286.

D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo*. Chapman and Hall/CRC, 2nd edition, 2006.

S. Gautier, G. Nolet, and J. Virieux. Finite-frequency tomograph in a crustal environment: Application to the western part of the Gulf of Corinth. *Geophysics Prospecting*, 56:493 – 503, 2008. doi: 10.1111/j.1365-2478.2007.00683.x.

A. Gelman and D. B. Rubin. *Bayesian statistics*, chapter A single series from the Gibbs sampler provides a false sense of security, pages 625–631. Oxford University Press, 1992a.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457 – 472, 1992b.

A. Gelman, J. B. Carlin, S. Hal, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, 2nd edition, 2004.

C. J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21:359–373, 1994.

H. N. Gharti and J. Tromp. A spectral-infinite-element solution of PoissonâĂŹs equation: an application to self gravity. *arXiv*, page 1706.00855v1, 2017.

K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622:759 – 771, 2005.

V. Y. Grechka and G. A. McMechan. 3-d two-point ray tracing for heterogenous, weakly transverse isotropic media. *Geophysics*, 61:1883 – 1894, 1996.

A. Green and R. Lane. Estimating noise levels in AEM data. In *16th Geophysical Conference and Exhibition*. ASEG, 2003. Extended Abstracts.

P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 4:711 – 732, 1995.

P. J. Green and A. Mira. Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika*, 88(4):1035–1053, 2001.

R. G. Green, K. F. Priestley, and R. S. White. Ambient noise tomography reveals upper crustal structure of Icelandic rifts. *Earth and Planetary Science Letters*, 466:20 – 31, 2017.

M. T. Gudmundsson and Högnadóttir. Volcanic systems and calderas in the Vatnajökull region, central Iceland: Constraints on crustal structure from gravity data. *Journal of Geodynamics*, 43:153 – 169, 2007.

Ó. Gudmundsson, A. Khan, and P. Voss. Rayleigh-wave group-velocity of the icelandic crust from correlation of ambient seismic noise. *Geophysical Research Letters*, 34 (L14314), 2007. doi: 10.1029/2007GL030215.

F. Guilloux, G. Fay, and J.-F. Cardoso. Practical wavelet design on the sphere. *Applied and Computational Harmonic Analysis*, 26:143–160, 2009.

A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331 – 371, 1910.

H. Haario, E. Saksman, and J. Tamminen. Componentwise adaptation for high dimensional McMC. *Computational Statistics*, 20:265–273, 2005.

H. Haario, M. Laine, A. Mira, and E. Saksman. DRAM: Effecient adaptive MCMC. *Stat Comput*, 16:339–354, 2006.

S. Hammarling, C. J. Munro, and F. Tisseur. An algorithm for the complete solution of quadratic eigenvalue problems. *ACM Trans. Math. Softw.*, 39(3):Article 18, 2013. doi: dx.doi.org/10.1145/2450153.2450156.

M. Haney and H. Douma. Inversion of Love wave phase velocity, group velocity and shear stress ratio using finite elements. In *SEG San Antonio 2011 Annual Meeting*. SEG, 2011.

M. Hanke. Limitations of the L-curve method in ill-posed problems. *BIT Numerical Mathematics*, 36(2):287–301, 1996.

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

N. A. Haskell. The dispersion of surface waves on multilayered media. *Bulletin of the Seismological Society of America*, 43(1):17–34, 1953.

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

J. Hauser, J. Gunning, and D. Annetts. Probabilistic inversion of airborne electromagnetic data under spatial constraints. *Geophysics*, 80(2):E135 – E146, 2015. doi: 10.1190/GEO2014-0389.1.

R. Hawkins and M. Sambridge. Geophysical imaging using trans-dimensional trees. *Geophysical Journal International*, 203(2):972–1000, 2015. doi: 10.1093/gji/ggv326.

R. Hawkins, R. Brodie, and M. Sambridge. Bayesian trans-dimensional inversion of Airborne Electromagnetic 2D conductivity profiles. *Exploration Geophysics*, 2017. doi: 10.1071/EG16139.

M. Hayn and M. Holschneider. Directional spherical multipole wavelets. *Journal of Mathematical Physics*, 50:073512, 2009. doi: 10.1063/1.3177198.

L. He and L. Carin. Exploiting structure in wavelet-based Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 57(9):3488–3497, 2009.

L. He, H. Chen, and L. Carin. Tree-structured compressive sensing with variational Bayesian analysis. *IEEE Signal Processing Letters*, 17(3):233–236, 2010.

R. B. Herrmann. Computer programs in seismology: An evolving tool for instruction and research. *Seismological Research Letters*, 84(6):1081 – 1088, 2013. doi: 10.1785/0220110096.

P. Hilton and J. Pedersen. Catalan numbers, their generalization, and their uses. *The Mathematical Intelligencer*, 13(2):64–75, 1991.

M. Holschneider. Continuous wavelet transforms on the sphere. *Journal of Mathematical Physics*, 37(8):4156 – 4165, 1996.

M. Holschneider and I. Igelwska-Nowak. Poisson wavelets on the sphere. *The Journal of Fourier Analysis and Applications*, 13(4), 2007. doi: 10.1007/s00041-006-6909-9.

M. Holschneider, A. Chambodut, and M. Mandea. From global to regional analysis of the magnetic field on the sphere using wavelet frames. *Physics of the Earth and Planetary Interiors*, 135:107–124, 2003.

P. O. Hopcroft, K. Gallagher, and C. C. Pain. Inference of past climate from borehole temperature data using Bayesian reversible jump Markov chain Monte Carlo. *Geophysical Journal International*, 171(3):1430–1439, 2007. doi: 10.1111/j.1365-246X.2007.03596.x.

S.-H. Hung, Y. Shen, and L.-Y. Chiao. Imaging seismic velocity structure beneath the Iceland hot spot: A finite frequency approach. *Journal of Geophysical Research*, 109: B08305, 2004. doi: 10.1029/2003JB002889.

S.-H. Hung, W.-P. Chen, and L.-Y. Chiao. A data-adaptive, multiscale approach of finite-frequency, traveltime tomography with special reference to P and S wave data from central Tibet. *Journal of Geophysical Research*, 116:B06307, 2011. doi: 10.1029/2010JB008190.

R. J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120 – 126, 1996.

G. Iaffaldano, R. Hawkins, and M. Sambridge. Bayesian noise-reduction in Arabia/Somalia and Nubia/Arabia finite rotations since ∼20 Ma: Implications for Nubia/Somalia relative motion. *Geochemistry, Geophysics, Geosystems*, 15(4), 2014.

H. Ishwaran and J. S. Rao. Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, 33(2):730–773, 2005.

S. S. Jakobsdóttir. Seismicity in iceland: 1994 - 2007. *Jökull*, 58:75 – 100, 2008.

A. Jasra, D. A. Stephens, and C. C. Holmes. Population-based reversible jump markov chain monte carlo. *Biometrika*, 94(4):787–807, 2007a.

A. Jasra, D. A. Stephens, and C. C. Holmes. On population-based simulation for static inference. *Statistical Computing*, 17:263–279, 2007b.

A. Jasra, A. Doucet, D. A. Stephens, and C. C. Holmes. Interacting sequential monte carlo samples for trans-dimensional simulation. *Computational Statistics and Data Analysis*, 52:1765–1791, 2008.

E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

H. Jeffreys. *Theory of Probability*. Clarendon Press, 3rd edition, 1939.

J. Jenkins, S. Cottaar, R. S. White, and A. Deuss. Depressed mantle discontinuities beneath Iceland: Evidence of a garnet controlled 660 km discontinuity? *Earth and Planetary Science Letters*, 433:159 – 168, 2016.

J. Julia, C. J. Ammon, R. B. Herrmann, and A. M. Correig. Joint inversion of receiver function and surface wave dispersion observations. *Geophysical Journal International*, 143:99–112, 2000.

B. R. Julian and D. Gubins. Three-dimensional seismic ray tracing. *Journal of Geophysics*, 43:95 – 113, 1977.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

B. L. N. Kennett. *Scattering and diffraction of seismic waves in multilayered media*. PhD thesis, University of Cambridge, 1973.

B. L. N. Kennett, E. R. Engdahl, and R. Buland. Constraints on seismic velocities in the earth from travel times. *Geophysical Journal International*, 122:108–124, 1995.

Z. Khalid, R. A. Kennedy, and J. D. McEwen. Slepian spatial-spectral concentration on the ball. *arXiv*, 2014.

R. Kimmel and J. A. Sethian. Computing geodesic paths on manifolds. *Proceedings of the National Academy of Sciences*, 95(15):8431–8435, 1997.

L. Knopoff. Observation and inversion of surface-wave dispersion. *Tectonophysics*, 13 (1):497–519, 1972.

D. E. Knuth. *The Art of Computer Programming*, volume 4. Addision-Wesley, 2004.

D. Komatitsch and J. Tromp. Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International*, 139:806 – 822, 1999.

D. Komatitsch and J. Tromp. Spectral-element simulations of global seismic wave propagation - i. Validation. *Geophysical Journal International*, 149:390 – 412, 2002a.

D. Komatitsch and J. Tromp. Spectral-element simulations of global seismic wave propagation - ii. Three-dimensional models, oceans, rotation and self-gravitation. *Geophysical Journal International*, 149:390 – 412, 2002b.

D. A. Kopriva. *Implementing spectral methods for partial differential equations*. Springer, 2009.

P. S. Koutsourelakis. A multi-resolution, non-parametric, bayesian framework for identification of spatially-varying model parameters. *J. Comput. Phys.*, 228(17):6184–6211, 2009.

C. La and M. N. Do. Signal reconstruction using sparse tree representations. *Proceedings of SPIE*, 5914:59140W, 2005.

E. Laloy and J. A. Vrugt. High-dimensional posterior exploration of hydrologic models using multiple-try dream(zs) and high-performance computing. *Water Resour. Res.*, 48:W01526, 2012. doi: 10.1029/2011WR010608.

K. Lambeck, H. Rouby, A. Purcell, Y. Sun, and M. Sambridge. Sea level and global ice volumes from the last glacial maximum to the holocene. *Proceedings of the National Academy of Sciences*, 111(43):15296–15303, 2014. doi: 10.1073/pnas.1411762111.

T. Langer, A. Belyaev, and H.-P. Seidel. Spherical barycentric coordinates. *Eurographics Symposium on Geometry Processing*, pages 81–88, 2006.

C. A. Langston. Structure under mount rainier, washington, inferred from teleseismic body waves. *Journal of Geophysical Research*, 84(B9):4749–4762, 1979.

F. Lanusse, A. Rassat, and J-L. Starck. Spherical 3d isotropic wavelets. *Astronomy and Astrophysics*, 540(A92), 2012. doi: 10.1051/0004-6361/201118568.

E. Larose, A. Derode, D. Clorennec, L. Margerin, and M. Campillo. Passive retrieval of Rayleigh waves in disordered elastic media. *Physical Review E*, 72:046607, 2005. doi: 10.1103/PhysRevE.72.046607.

E. Larose, L. Margerin, A. Derode, B. van Tiggelen, M. Campillo, N. Shapiro, A. Paul, L. Stehly, and M. Tanter. Correlation of random wavefields: An interdisciplinary review. *Geophysics*, 71(4):S111 – S121, 2006.

K. Lawrie. Broken hill managed aquifer recharge. `http://www.ga.gov.au/about/projects/water/broken-hill-managed-aquifer-recharge`, 2016. [Online; accessed 20-September-2016].

K. C. Lawrie, T. J. Munday, D. L. Dent, D. L. Gibson, R. C. Brodie, J. Wilford, N. S. Reilly, R. N. Chan, and P. Baker. A geological systems approach to understanding the processes involved in land and water salinisation; the Gilmore project area, central-west New South Wales. *AGSO Research Newsletter*, 32:13 – 15, 2000.

C. L. Lawson. *Mathematical Software III*, chapter Software for C^1 Surface Interpolation, pages 161 – 194. Academic Press, 1977.

B. Leistedt and J. D. McEwen. Exact wavelets on the ball. *IEEE Transactions on Signal Processing*, 60(12):6257 – 6269, 2012.

B. Leistedt, J. D. McEwen, P. Vandergheynst, and Y. Wiaux. S2let: A code to perform fast wavelet analysis on the sphere. *Astronomy and Astrophysics*, 558(A128), 2013a. doi: 10.1051/0004-6361/201220729.

B. Leistedt, H. V. Peiris, and J. D. McEwen. Flaglets for studying the large-scale structure of the universe. In *SPIE Optical Engineering and Applications*, 2013b.

J.-J. Lévêque, L. Rivera, and G. Wittlinger. On the use of checker-board test to assess the resolution of tomography inversions. *Geophysical Journal International*, 115:313 – 318, 1993.

J. Lever, M. Krzywinski, and N. Altman. Model selection and overfitting. *Nature Methods*, 13(9):703 – 704, 2016.

A. Li and R. S. Detrick. Seismic structure of Iceland from Rayleigh wave inversions and geodynamic implications. *Earth and Planetary Science Letters*, 241:901 – 912, 2006.

J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer, 2001.

J. S. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998. doi: 10.1080/01621459.1998.10473765.

J. S. Liu, G. Liang, and W. H. Wong. The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association*, 95(449):121 – 134, 2000. doi: 10.1080/01621459.2000.10473908.

O. I. Lobkis and R. L. Weaver. On the emergence of the Green's function in the correlations of a diffuse field. *J. Acoust. Soc. Am.*, 110(6):3011 – 3017, 2001. doi: 10.1121/1.1317528.

T. Lochbühler, J. A. Vrugt, M. Sadegh, and N. Linde. Summary statistics from training images as prior information in probabilistic inversion. *Geophysical Journal International*, 201(1):157–171, 2015.

A. Love. *A treatise on the mathematical theory of elasticity*. Cambridge University Press, 4th edition, 1927.

J. Lysmer. Lumped mass method for Rayleigh waves. *Bulletin of the Seismological Society of America*, 60(1):89 – 104, 1970.

J. Lysmer and L. A. Drake. *Methods in Computation Physics*, volume 11 of *Seismology: Surface waves and Earth oscillations*, chapter Seismic surface waves, pages 181 – 216. Academic Press, 1972.

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

A. Malinverno and V. A. Briggs. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, 69(4):1005 – 1016, 2004. doi: 10.1190/1.1778243.

S. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Transactions of the American Mathematical Society*, 315(1):69–87, 1989.

S. Mallat. *A Wavelet Tour of Signal Processing*. CA: Academic, 2nd edition, 1999.

B. K. Mallick. Bayesian curve estimation by polynomials of random order. *Journal of Statistical Planning and Inference*, 70(1):91 – 109, 1998.

C. Mavriplis. Laguerre polynomials for infinite-domain spectral elements. *Journal of Computational Physics*, 80:480 – 488, 1989.

J. C. Maxwell. *A treatise on electricity and magnetism*, volume 1. Clarendon Press, 1881.

W. Menke. *Geophysical data analysis: discrete inverse theory*. Academic Press, 1989.

W. Menke and G. Jin. Waveform fitting of cross spectra to determine phase velocity using Aki's formula. *Bulletin of the Seismological Society of America*, 105(3):1619 – 1627, 2015. doi: 10.1785/0120140245.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1986–1992, 1953.

V. Michel. Regularized wavelet-based multiresolution recovery of harmonic mass density distribution from data of the Earth's gravitational field at satellite height. *Inverse Problems*, 21:997 – 1025, 2005. doi: 10.1088/0266-5611/21/3/013.

V. Michel and K. Wolf. Numerical aspects of a spline-based multiresolution recovery of the harmonic mass density out of gravity functionals. *Geophysical Journal International*, 173:1 – 16, 2008. doi: 10.1111/j.1365-246X.2007.03700.x.

B. J. Minsley. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data. *Geophysical Journal International*, 187:252–272, 2011. doi: 10.1111/j.1365-246X.2011.05165.x.

B. J. Minsley and A. Y. Ley-Cooper. Quantifying model structural uncertainty and facies prediction for locating groundwater supplies in Timor-Leste using AEM data. In *ASEG Extended Abstracts*, volume 1, pages 1 – 4, 2015.

C. B. Moler and G. W. Stewart. An algorithm for generalized matrix eigenvalue problems. *SIAM J. NUMER. ANAL.*, 10(2):241 – 256, 1973.

J. Montagner. Regional three-dimensional structures using long period surface waves. *Annales Geophysicae*, 4:283 – 294, 1986.

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

R. M. Neal. *Handbook of Markov chain Monte Carlo*, chapter McMC using Hamiltonian Dynamics. Chapman and Hall/CRC, 2011.

A. Okabe, B. Boots, and K. Sugihara. *Spatial tesselations: Concepts and applications of Voronoi diagrams*. John Wiley & Sons, 1992.

T. J. Owens, G. Zandt, and S. R. Taylor. Seismic evidence for and ancient rift beneath the cumberland plateau, tenessee: A detailed analysis of broadband teleseismic p waveforms. *Journal of Geophysical Research*, 89(B9):7783–7795, 1984.

G. J. Palacky. Use of airborne electromagnetic methods for resource mapping. *Advances in Space Research*, 13(11):5 – 14, 1993.

T. Park and G. Casella. The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681 – 686, 2008. doi: 10.1198/016214508000000337.

A. Paul, M. Campillo, L. Margerin, E. Larose, and A. Derode. Empirical synthesis of time-assymmetrical Green functions from correlation of coda waves. *Journal of Geophysical Research*, 110:B08302, 2005. doi: 10.1029/2004JB003521.

N. K. Pavlis, S. A. Holmes, S. C. Kenyon, and J. K. Factor. The development and evaluation of the earth gravitational model 2008 (egm2008). *Journal of Geophysical Research*, 117:B04406, 2012.

T. Pejic, R. Hawkins, H. Tkalcic, and M. Sambridge. Trans-dimensional Bayesian attenuation tomography of the upper inner core. *Journal of Geophysical Research*, 2017. In preparation.

N. Piana Agostinetti and A. Malinverno. Receiver function inversion by trans-dimensional Monte Carlo sampling. *Geophysical Journal International*, 181(2):858–872, 2010.

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

S. Pilia, N. Rawlinson, N. G. Direen, A. M. Reading, R. Cayley, L. Pryer, P. Arroucau, and M. Duffett. Linking mainland Australia and Tasmania using ambient seismic noise tomography: Implications for the tectonic evolution of the east Gondwana margin. *Gondwana Research*, 2014.

S. Pilidou, K. Priestly, E. Debayle, and O. Gudmundsson. Rayleigh wave tomography in the North Atlantic: high resolution images of the Iceland, Azores and Eifel mantle plumes. *Lithos*, 79:453 – 474, 2005.

A. Plattner, H. R. Maurer, J. Vorleoper, and M. Blome. 3-d electrical resistivity tomography using adaptive wavelet parameter grids. *Geophysical Journal International*, 189: 317–330, 2012. doi: 10.1111/j.1365-246X.2012.05374.x.

R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167:495 – 503, 2006. doi: 0.1111/j.1365-246X.2006.02978.x.

C. Pozrikidis. *Finite and spectral element methods using Matlab*. Chapman & Hall/CRC, 2005.

J. E. Quijano, S. E. Dosso, J. Dettmer, and C. Holland. Geoacoustic inversion for the seabed transition layer using a Bernstein polynomial model. *The Journal of the Acoustical Society of America*, 140(6):4073 – 4084, 2016.

N. Rawlinson and M. Sambridge. *Advances in Geophysics*, volume 46, chapter Seismic traveltime tomography of the crust and lithosphere. Academic Press, 2003.

N. Rawlinson and M. Sambridge. The fast marching method: an effective tool for tomographic imaging and tracking multiple phases in complex layered media. *Exploration Geophysics*, 36:341–350, 2005.

N. Rawlinson and W. Spackman. On the use of sensitivity tests in seismic tomography. *Geophysical Journal International*, 205:1221 – 1243, 2016. doi: 10.1093/gji/ggw084.

N. Rawlinson, J. Hauser, and M. Sambridge. Seismic ray tracing and wavefront tracking in laterally heterogeneous media. *Advances in Geophysics*, 49:203–267, 2007.

N. Rawlinson, B. L. N. Kennett, E. Vanacore, R. A. Glen, and S. Fishwick. The structure of the upper mantle beneath the Delamerian and Lachlan orogens from simultaneous inversion of multiple teleseismic datasets. *Gondwana Research*, 19:788–799, 2011.

N. Rawlinson, A. Fichtner, M. Sambridge, and M. Young. *Advances in Geophysics*, volume 55, chapter Seismic tomography and the assessment of uncertainty. Academic Press, 2014.

A. Ray and K. Key. Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm. *Geophysical Journal International*, 191:1135 – 1151, 2012. doi: 10.1111/j.1365-246X.2012.05677.x.

A. Ray, K. Key, T. Bodin, D. Meyer, and S. Constable. Bayesian inversion of marine CSEM data from the Scarborough gas field using transdimensional 2-d parameterization. *Geophysical Journal International*, 199:1847 – 1860, 2014. doi: 10.1093/gji/ggu370.

F. Rickers, A. Fichtner, and J. Trampert. The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion. *Earth and Planetary Science Letters*, 367:39 – 51, 2013. doi: 10.1016/j.epsl.2013.02.022.

J. Ritsema and H.-J. van Heijst. Constraints on the correlation of P and S-wave velocity heterogeneity in the mantle from P, PP, PPP and PKPab traveltimes. *Geophysical Journal International*, 149:482 = 489, 2002.

J. Ritsema, H. J. van Heijst, and J. H. Woodhouse. Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science*, 286:1925 – 1928, 1999.

G. O. Roberts. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(1):255 – 268, 1998.

G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341 – 363, 1996.

C. Ronchi, R. Iacono, and P. S. Paolucci. The "cubed sphere": A new method for the solution of partial differential equations in spherical geometry. *Journal of Computational Physics*, 124:93 – 114, 1996.

M. Rosas-Carbajal, N. Linde, Kalscheuer, and J. A. Vrugt. Two-dimensional probabilistic inversion of plane-wave electromagnetic data: methodology, model constraints and joint inversion with electrical resistivity data. *Geophysical Journal International*, 196:1508 – 1524, 2014. doi: 10.1093/gji/ggt482.

M. L. Rudolph, V. Lekic, and C. Lithgow-Bertelloni. Viscosity jump in Earth's mid mantle. *Science*, 350(6266):1349 – 1352, 2015.

K. G. Sabra, P. Gerstoft, P. Roux, and W. A. Kuperman. Extracting time-domains Green's function estimates from ambient seismic noise. *Geophysical Research Letters*, 32:L03310, 2005. doi: 10.1029/2004GL021862.

E. B. Saff and A. B. J. Kuijlaars. Distributing many points on a sphere. *The Mathematical Intelligencer*, 19(1):5–11, 1997.

A. Said and W. A. Pearlman. A new fast and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3):243–250, 1996.

M. Sambridge. A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, 192:357–374, 2014.

M. Sambridge. Reconstructing time series and their uncertainty from observations with universal noise. *Journal of Geophysical Research: Solid Earth*, 121(7):4990 – 5012, 2016. doi: 10.1002/2016JB012901.

M. Sambridge and R. Faletič. Adaptive whole earth tomography. *Geochemistry Geophysics Geosystems*, 4(3), 2003.

M. Sambridge and O. Gudmundsson. Tomographic systems of equations with irregular cells. *Journal of Geophysical Research*, 103(B1):773–781, 1998.

M. Sambridge and K. Mosegaard. Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3):1–29, 2002.

M. Sambridge, J. Braun, and H. McQueen. Geophysical parameterization and interpolation of irregular data using natural neighbours. *Geophysical Journal International*, 122:837–857, 1995.

M. Sambridge, K. Gallagher, A. Jackson, and P. Rickwood. Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, 167:528–542, 2006.

M. S. Sambridge and B. L. N. Kennett. Boundary value ray tracing in a heterogenous medium: A simple and versatile algorithm. *Geophysical Journal International*, 101: 157 – 168, 1990.

H. Samet. *Foundations of Multidimensional and Metric data structures*. Morgan Kaufmann, 2006.

D. Sattel and L. Kgotlhang. Groundwater exploration with AEM in the Boteti area, Botswana. *Exploration Geophysics*, 35:147 – 156, 2004.

E. Saygin, P. Cummins, A. Cipta, R. Hawkins, R. Pandhu, J. Murjaya, M. Masturyono Irsyam, S. Widiyantoro, and B. L. N. Kennett. Imaging architecture of the Jakarta basin, Indonesia with trans-dimensional Bayesian seismic noise tomography. *Geophysical Journal International*, 204(2):918 – 931, 2016. doi: 10.1093/gji/ggv466.

J. A. Scales and R. Snieder. To Bayes or not to Bayes? *Geophysics*, 64(4):1045 – 1046, 1997.

P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 1995.

P. Schroder and W. Sweldens. *Rendering Techniques*, chapter Spherical Wavelets: Texture Processing. Springer-Verlag, 1995.

G. E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

K. J. Seats, J. F. Lawrence, and G. A. Prieto. Improved ambient noise correlation functions using Welch's method. *Geophysical Journal International*, 188:513 – 523, 2012. doi: 10.1111/j.1365-246X.2011.05263.x.

I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, pages 123–151, November 2005.

M. K. Sen and R. Biswas. Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm. *Geophysics*, 82(3):R119 – R134, 2016. doi: 10.1190/GEO2016-0010.1.

J. A. Sethian and A. M. Popovici. 3-d traveltime computation using the fast marching method. *Geophysics*, 64(2):516–523, 1999.

J. A. Sethian and A. Vladimirsky. Fast methods for the Eikonal and related Hamilton-Jacobi equations on unstructured meshes. *Proceedings of the National Academy of Sciences*, 97(11):5699 – 5703, 2000. doi: 10.1073/pnas.090060097.

J. A. Sethian and A. Vladimirsky. Ordered upwind methods for static Hamilton-Jacobi equations: Theory and algorithms. *SIAM J. Numer. Anal.*, 41(3):325 – 363, 2003.

A. J. Shaeffer and S. Lebedev. Global shear speed structure of the upper mantle and transition zone. *Geophysical Journal International*, 194:417 – 449, 2013. doi: 10.1093/gji/ggt095.

J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, 1993.

N. M. Shapiro and M. Campillo. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31(L07614), 2004.

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

J. Shen. Stable and efficient spectral methods in unbounded domains using Laguerre functions. *SIAM J. Numer. Anal.*, 38(4):1113 – 1133, 2000.

Y. Shen, S. C. Solomon, I. T. Bjarnason, and C. J. Wolfe. Seismic evidence for a lower-mantle origin of the Iceland plume. *Nature*, 395:62 – 65, 1998.

Y. Shen, S. C. Solomon, I. Bjarnason, G. Nolet, W. J. Morgan, R. M. Allen, Vogfjörd, Jakobsdóttir, R. Stefánsson, B. R. Julian, and G. R. Foulger. Seismic evidence for a tilted mantle plume and north-south mantle flow beneath Iceland. *Earth and Planetary Science Letters*, 197:261 – 272, 2002.

P. D. Shukla. Complex wavelet transforms and their applications. Master's thesis, University of Strathclyde Glasgow, 2003.

F. J. Simons. *Handbook of Geomathematics*, chapter Slepian functions and their use in signal estimation and spectral analysis, pages 891–923. Springer Berlin Heidelberg, 2010.

F. J. Simons, F. A. Dahlen, and M. A. Wieczorek. Spatiospectral concentration on a sphere. *SIAM Review*, 48(3):504–536, 2006.

F. J. Simons, I. Loris, E. Brevdo, and I. C. Daubechies. Wavelets and wavelet-like transforms on the sphere and their application to geophysical data inversion. In *Wavelets and Sparsity XIV*. SPIE, 2011a.

F. J. Simons, I. Loris, G. Nolet, I. C. Daubechies, S. Voronin, J. S. Judd, P. A. Vetter, J. Charlety, and C. Vonesch. Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. *Geophysical Journal International*, 187(2):969 – 988, 2011b.

F. Simpson and K. Bahr. *Practical magnetotellurics*. Cambridge, 2005.

S. A. Sisson. Transdimensional markov chains: A decade of progress and future perspectives. *Journal of the American Statistical Association*, 100(471):1077–1089, 2005.

S. A. Sisson and Y. Fan. A distance-based diagnostic for trans-dimensional markov chains. *Statistical Computing*, 17:357–367, 2007.

J. Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4): 833 – 860, 2006.

D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty - I. *Bell Syst. Tech. J.*, 40:43 – 63, 1960.

M. L. Smith and F. A. Dahlen. The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium. *Journal of Geophysical Research*, 78 (17):3321 – 3333, 1973.

R. Snieder. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Physical Review E*, 69:046610, 2004. doi: 10.1103/PhysRevE.69.046610.

R. Snieder and E. Larose. Extracting earth's elastic wave response from noise measurements. *Annual Review of Earth and Planetary Sciences*, pages 183–206, 2013.

K. I. Sorensen and E. Auken. SkyTEM - a new high-resolution helicopter transient electromagnetic system. *Exploration Geophysics*, 35:191 – 199, 2004.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64:583–639, 2002.

R. P. Stanley. *Enumerative Combinatorics*, volume 1. Cambridge University Press, 1997.

G. Steininger, J. Dettmer, and S. E. Dosso. Trans-dimensional joint inversion of seabed scattering and reflection data. *J. Acoust. Soc. Am.*, 133(3), 2013. doi: 0.1121/1.4789930.

G. Steininger, S. E. Dosso, C. W. Holland, and J. Dettmer. Estimating seabed scattering mechanisms via Bayesian model selection. *J. Acoust. Soc. Am.*, 136(4):1552–1562, 2014.

G. J. Street, G. Pracilio, P. Nallan-Chakravartula, C. Nash, D. Sattel, M. Owers, D. Triggs, and R. Lane. National dryland salinity program airborne geophysical surveys to assist planning for salinity control; 1. Willaura SALTMAP survey interpretation report. *National Airbourne Geophysics Project, World Geoscience Corporation*, 1998.

W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis*, 3(15):186–200, 1996.

W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, 29(2):511–546, 1998.

S. Swenson and J. Wahr. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *Journal of Geophysical Research*, 107(B9):2192, 2002. doi: 10.1029/2001JB000576.

H. Takeuchi and M. Saito. *Methods in Computation Physics*, volume 11 of *Seismology: Surface waves and Earth oscillations*, chapter Seismic surface waves, pages 217 – 295. Academic Press, 1972.

T. Tanimoto. Surface-wave ray tracing equations and Fermat's principle in an anisotropic earth. *Geophys. J. R . astr. Soc*, 88:231 – 240, 1986.

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

A. Tarantola and B. Valette. Inverse problems = Quest for information. *Journal of Geophysics*, 50:159 – 170, 1982a.

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

C. J. F. ter Braak and J. A. Vrugt. Differential evolution Markov chain with snooker updater and fewer chains. *Stat Comput*, 18:435 – 446, 2008.

L. Thomsen. Weak elastic anisotropy. *Geophysics*, 51(10):1954 – 1966, 1988.

W. T. Thomson. Transmission of elastic waves through a stratified solid medium. *Journal of Applied Physics*, 21:89–93, 1950. doi: 10.1063/1.1699629.

T. Thordarson and Á. Höskuldsson. Postglacian volcanism in Iceland. *Jökull*, 58:197 – 228, 2008.

T. Thordarson and G. Larsen. Volcanism in Iceland in historical time: Volcano types, eruption styles and eruptive history. *Journal of Geodynamics*, 43:118 – 152, 2007.

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

L. Tierney and A. Mira. Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in medicine*, 18(1718):2507–2515, 1999.

A. N. Tikhonov. On the stability of inverse problems. *Doklady Akademii Nauk SSSR*, 39(5):195 – 198, 1943.

A. N. Tikhonov. On determination of electric characteristics of deep layers of the Earth's crust. *Doklady Academii Nauk SSSR*, 151:295–297, 1950.

F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Review*, 43 (2):235–286, 2001. doi: 10.1137/S0036144500381988.

H. Tkalčić, D. S. Dreger, G. R. Foulger, and B. R. Julian. The puzzle of the 1996 bárdarbunga, Iceland, earthquake: No volumetric component in the source mechanism. *Bulletin of the Seismological Society of America*, 99(5):3077 – 3085, 2009.

H. Tkalčić, M. Young, T. Bodin, S. Ngo, and M. Sambridge. The shuffling rotation of the earth's inner core revealed by earthquake doublets. *Nature Geoscience*, 6(6): 497–502, 2013.

J. Trampert and R. Sneider. Model estimations biased by truncated expansions: possible artifacts in seismic tomography. *Science*, 271:1257 – 1260, 1996.

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

J. Tromp and F. A. Dahlen. Variational principles for surface wave propagation on a laterally heterogeneous Earth - i. time-domain JWKB theory. *Geophysical Journal International*, 109:581 – 598, 1992a.

J. Tromp and F. A. Dahlen. Variational principles for surface wave propagation on a laterally heterogeneous Earth - ii. frequency-domain JWKB theory. *Geophysical Journal International*, 109:599 – 619, 1992b.

J. Um and C. Thurber. A fast algorithm for two-point seismic ray tracing. *Bulletin of the Seismological Society of America*, 77:972 – 986, 1987.

M. Unser and T. Blu. Mathematical properties of the JPEG2000 wavelet filters. *IEEE Transactions on Image Processing*, 12(9):1080–1090, 2003.

B. E. Usevitch. A tutorial on modern lossy wavelet image compression: Foundations of JPEG2000. *IEEE Signal Processing Magazine*, pages 22–35, 2001.

J. Valenciano and M. A. J. Chaplain. A Laguerre-Legendre spectral-element method for the solution of partial differential equations on infinite domains: application to the diffusion of tumour angieogenesis factors. *Mathematical and Computer Modelling*, 41:1171 – 1192, 2005. doi: 10.1016/j.mcm.2005.05.010.

A. P. Valentine and J. Trampert. Assessing the uncertainties on seismic source parameters: Towards realistic error estimates for centroid-moment-tensor determinations. *Physics of the Earth and Planetary Interiors*, 210 - 211:36 – 49, 2012. doi: 10.1016/j.pepi.2012.08.003.

A. P. Valentine and J. Trampert. The impact of approximations and arbitrary choices on geophysical images. *Geophysical Journal International*, 204:59 – 73, 2016.

E. Vera, L. Mancera, S. D. Babacan, R. Molina, and A. K. Katsaggelos. Bayesian compressive sensing of wavelet coefficients using multiscale laplacian priors. In *IEEE/SP 15th Workshop on Statistical Signal Processing*. IEEE, 2009.

K. Visser, J. Trampert, and B. L. N. Kennett. Global anisotropic phase velocity maps for higher mode Love and Rayleigh waves. *Geophysical Journal International*, 172: 1016 – 1032, 2008.

C. R. Vogel. Non-convergence of the L-curve regularization parameter selection method. *Inverse Problems*, 12:535–547, 1996.

C. R. Vogel. *Computational methods for inverse problems*, volume 23 of *Frontiers in Applied Mathematics*. SIAM, 2002.

J. A. Vrugt and C. J. F. ter Braak. Dream(d): an adaptive markov chain monte carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrol. Earth Syst. Sci.*, 15:3701 – 3713, 2011.

J. A. Vrugt, C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon. Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3):271 – 288, 2009.

Z. Wang and F. A. Dahlen. Spherical-spline parameterization of three-dimensional Earth models. *Geophysical Research Letters*, 22(22):3099–3102, 1995.

K. Wapenaar. Retrieving the elastodynamic Green's function of an arbitrary inhomogenous medium by cross correlation. *Physical Review Letters*, 94:254301, 2004. doi: 10.1103/PhysRevLett.93.254301.

L. Waszek and A. Deuss. Distinct layering in the hemispherical seismic velocity structure of the Earth's upper inner core. *Journal of Geophysical Research: Solid Earth*, 116, 2011. doi: 10.1029/2011JB008650.

K. A. Whaler and D. Gubbins. Spherical harmonic analysis of the geomagnetic field: an example of a linear inverse problem. *Geophysical Journal International*, 65(3): 645–693, 1981.

R. A. Wiggins. Monte Carlo inversion of body-wave observations. *Journal of Geophysical Research*, 74(12):3171–3181, 1969.

R. A. Wiggins. A fast, new computational algorithm for free oscillations and surface waves. *Geophys. J. R. astr. Soc.*, 47:135–150, 1976.

H. W. Wilf. *Generatingfunctionology*. Academic Press, 1990.

J. H. Woodhouse. Surface waves in a laterally varying layered structure. *Geophys. J. R. astr. Soc.*, 37:461 – 490, 1974.

B. Y. Wu and K.-M. Chao. *Spanning trees and optimization problems.* Taylor and Francis, 2004.

D. Yang and D. W. Oldenburg. Three-dimensional inversion of airborne time-domain electromagnetic data with applications to a porphyry deposit. *Geophysics*, 77(2):B23 – B34, 2012. doi: 10.1190/GEO2011-0194.1.

H. Yao, G. Xu, L. Zhu, and X. Xiao. Mantle structure from inter-station rayleigh wave dispersion and its tectonic implication in western china and neighboring regions. *Physics of the Earth and Planetary Interiors*, 148:39 – 54, 2005. doi: 10.1016/j.pepi.2004.08.006.

H. Yao, R. D. van der Hilst, and M. V. de Hoop. Surface-wave array tomography if SE Tibet from ambient seismic noise and two-station analysis - I. Phase velocity maps. *Geophysical Journal International*, 166:732 – 744, 2006. doi: 10.1111/j.1365-246X.2006.03028.x.

M. K. Young, N. Rawlinson, P. Arroucau, A. M. Reading, and H. Tkalčić. High frequency ambient noise tomography of southeast Australia: New constraints on Tasmania's tectonic past. *Geophysical Research Letters*, 38(13):L13313, 2011.

M. K. Young, H. Tkalčić, T. Bodin, and M. Sambridge. Global p wave tomography of Earth's lowermost mantle from partition modeling. *Journal of Geophysical Research: Solid Earth*, 118:5467–5486, 2013.

C. Zaroli. Global seismic tomography using Backus-Gilbert inversion. *Geophysical Journal International*, 207:876–888, 2016.

M. T. Zuber, D. E. Smith, M. M. Watkins, S. W. Asmar, A. S. Konopliv, F. G. Lemoine, H. J. Melosh, G. A. Neumann, R. J. Phillips, S. C. Solomon, M. A. Wieczorek, J. G. Williams, S. J. Goossens, G. Kruizinga, E. Mazarico, R. S. Park, and D.-N. Yuan. Gravity field of the Moon from the Gravity Recovery and Interior Laboratory (GRAIL) mission. *Science*, 339:668 – 671, 2013. doi: 10.1126/science.1231507.