Model parameterization and model selection in geophysical inverse problems

Designing inverse problems that respect a priori geophysical knowledge

Thesis by
Jack Broderick Muir

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended 2021–10–04

Jack Broderick Muir
ORCID: 0000-0003-2617-3420

This thesis was substantively written on the lands of the Tongva, Eora, Yuin, Ngunnawal and Ngambri peoples. I pay my respects to elders, past, and present, for their stewardship of the land. Without their efforts over untold ages, neither the conditions necessary for the development of the science nor the writing of this thesis would have been possible.

ACKNOWLEDGEMENTS

The past six years have been a transformative part of my life. At the Caltech Seismo Lab, I have found a warm and welcoming home, where I have grown in independence and experience — both academically and emotionally. The Seismo Lab has a tremendous energy, and if you immerse yourself in it, it is impossible not to feel like your work is not in some way consequential, which is a great consolation during the long years of a PhD. I am of course particularly indebted to my many supervisors and faculty collaborators. Victor has taught me the value of variety whilst maintaining a firm eye on first principles (and publishing in a timely manner!). Zhongwen's enthusiasm for pushing technical boundaries in instrumentation and methodology has been inspirational. Rob's great expertise in large array seismology and seismic inversion has been indispensable. Jennifer has been a fantastic and thoughtful academic advisor. Zach has been an enthusiastic supporter of some of my quirkier ideas, which should hopefully see publication soon after this thesis is defended! Beyond the faculty that I have worked directly with at the Seismo Lab, I've had the great opportunity to interact informally with other members — especially notable being the famous post-earthquake coffee hours moderated by Mike, Mark, Tom and Hiroo! The student and postdoc cohort at the Seismo Lab has evolved into a wonderful network of peer researchers during my time at Caltech. I am especially thankful for my interactions, both social and academic, over the years with Daniel, Voon and Jorge, and for the strange quirk of fate that allowed me to share an office with Voon on the other side of the world as I rode out 2021 in Australia. Rosemary, Priscilla, Sarah, Donna, and Kim deserve endless appreciation for everything that they do to enable the good work of the Seismo Lab to proceed.

Further afield at Caltech, I've had the great opportunity to work with Andrew Stuart and Ollie Dunbar at CMS. I'm particularly thankful for Andrew's inverse problem course, which proved to be particularly influential for many of the topics explored in this thesis and also lead to my participating in a fantastic inverse problems summer school at Breckenridge under Omar Ghattas, which I will always remember fondly. I've also been supremely fortunate in being able to return to my *alma mater*, the Australian National University, in 2021, where I've been able to have some respite from the COVID pandemic and have resumed my collaborations with Hrvoje Tkalčić, Malcolm Sambridge and Rhodri Davies, all of whom were instrumental in my initial foray into geophysics.

I was lucky to have the opportunity to participate fully in Caltech's extracurricular activities. I'd like to thank Glen for running a fantastic music program, and will always remember the special place that Allen created in Caltech's orchestra. I'd like to thank all of the members of the GSC for their hard work supporting the graduate students of Caltech, and especially Alistair, Daniel, Ana and Nancy for being close collaborators during my time as a GSC officer. Konrad, Xuan, Alex$^{(2)}$ and Reston helped keep me in shape (always important in an office job!) with volleyball and climbing.

Of course, the greatest of thanks must be reserved for my family. My inlaws, Richard and Kris, have provided me with great kindness over the past decade, especially in supplying the refuge from where I write this thesis.

My parents, Kathy and Rod, have supported me in an unfathomable fashion for the past 30 years, have made sure that I have had every opportunity possible to achieve anything that I wished to achieve, and have always been ready to listen.

To my wife Kate and daughter Hadley — thank you for taking this journey with me, sharing the sacrifices that it has entailed and hopefully some of the joys as well. My time with you has always been the best part of every day, and your love has kept me going on many occasions. I look forward to an exciting future with you over new horizons!

ABSTRACT

The vast majority of the Earth system is inaccessible to direct observation. Consequently, the structure and dynamics of the Earth can only be determined indirectly, via geophysical sensing. These methods have the mathematical form of an inverse problem, in which the data and the unknowns are linked by a physical process, such as seismic wave propagation. From the possibly noisy data, we have indirect access to the unknowns. The vast majority of geophysical inverse problems are ill-posed, and require the provision of *a priori* knowledge to stabilize the solution. This thesis investigates methods for designing inverse problems to better take advantage of geophysical or geological constraints, to allow better resolution or more interpretability of the solutions. Four major themes are investigated: In Chapter 2, we study the collection of a novel dataset of Rayleigh wave horizontal-to-vertical ratios to provide stronger constraints on upper-crustal structure in Southern California. In Chapters 3 and 4, we develop a method for wavefield-reconstruction of sparse seismic data, including heterogeneous networks consisting of both displacement and strain instruments. This method amounts to an inversion in data-space, and promises to unlock the potential of wavefield based methods for complex datasets. In Chapters 5 and 6, we investigate a new structural parameterization based on a combination of Gaussian processes and the level-set method, that better models discontinuous geological features such as sedimentary basins. We test our method on a variety of synthetic and real datasets, culminating in a detailed study of the northeastern Los Angeles basin, which we found to be significantly deeper and steeper than in previous models. Finally, we develop a method of model selection for noisy historical datasets, which we investigate using the case study of correcting Oldham's data misinterpretation in the 1906 paper that "discovered" Earth's core.

PUBLISHED CONTENT AND CONTRIBUTIONS

Muir, J. B., R. W. Clayton, V. C. Tsai & Q. Brissaud "Parsimonious velocity inversion applied to the Los Angeles Basin, CA".
In: *Journal of Geophysical Research: Solid Earth*, Submitted
Contributions: JBM, RWC and VCT co-designed the project. JBM wrote the code, analysed the data, designed the figures and wrote the manuscript. QB contributed to the technical analysis, provided an initial figure and provided editorial assistance. RWC and VCT contributed in a supervisory and editorial role.

Muir, J. B. & Z. Zhan "Wavefield-based evaluation of DAS instrument response and array design".
In: *Geophysical Journal International*, Submitted
Contributions: JBM and ZZ co-designed the project. JBM wrote the code, analysed the data, designed the figures and wrote the manuscript. ZZ contributed in a supervisory and editorial role.

Muir, J. B. & Z. Zhan (2021). "Seismic Wavefield Reconstruction using a Preconditioned Wavelet-Curvelet Compressive Sensing Approach".
In: *Geophysical Journal International* 227.1, pp. 303–315.
DOI: 10.1093/gji/ggab222
Contributions: JBM and ZZ co-designed the project. JBM wrote the code, analysed the data, designed the figures and wrote the manuscript. ZZ contributed in a supervisory and editorial role.

Muir, J. B. & V. C. Tsai (2020). "Geometric and level set tomography using ensemble Kalman inversion".
In: *Geophysical Journal International* 220.2, pp. 967–980.
DOI: 10.1093/gji/ggz472
Contributions: JBM and VCT co-designed the project. JBM wrote the code, analysed the data, designed the figures and wrote the manuscript. VCT contributed in a supervisory and editorial role.

Muir, J. B. & V. C. Tsai (2020). "Did Oldham discover the core after all? Handling imprecise historical data with hierarchical Bayesian model selection methods".
In: *Seismological Research Letters* 91.3, pp. 1377–1383.
DOI: 10.1785/0220190266
Contributions: JBM designed the project, wrote the code, analysed the data, designed the figures and wrote the manuscript. VCT provided the initial impetus for the project, and contributed in a supervisory and editorial role.

Muir, J. B. & V. C. Tsai (2017). "Rayleigh-Wave H/V via Noise Cross Correlation in Southern California".
In: *Bulletin of the Seismological Society of America* 107.5, pp. 2021–2027.

DOI: 10.1785/0120170051
Contributions: JBM wrote the code, analysed the data, designed the figures and wrote the manuscript. VCT designed the project and contributed in a supervisory and editorial role.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

LIST OF TABLES

Chapter 1

INTRODUCTION

Elucidating the full 3D heterogeneity Earth's interior processes is one of the grand challenges of 21st century Earth science. From the multi-million-year deep tectonic processes providing Earth's climostat, to the amplification of earthquake shaking in the seconds the energy takes to cross a shallow sedimentary basin, essential questions of Earth structure ranging at all time and spatial scales remain to be resolved. The most detailed information we have about the Earth's interior derives from seismology. The basic 1D wave equation, with displacement u and wavespeed c,

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0, \tag{1.1}$$

is hyperbolic, and has solutions of the form $u(x,t) = f(x - ct) + g(x + ct)$. Qualitatively, this type of controlling equation allows information from deep within the Earth to be maintained and carried to the surface, and as such seismic imaging has proven incredibly fruitful. Despite this useful property, the physics of viscoelastic wave propagation and the incomplete illumination of the Earth's interior by the seismic data recorded to date means that the problem of recovering Earth structure from seismic data is fundamentally ill-posed. This thesis aims to improve both the quality of data and the quality of inversion methods used to solve seismic inverse problems, with a particular focus on upper-crustal structure in the western United States, a region of interesting tectonics and high seismic hazard.

1.1 The Structure of Inverse Problems

At the heart of geophysical inverse problems we usually find an equation of the form

$$\boldsymbol{y} = G(\boldsymbol{m}). \tag{1.2}$$

This equation combines the three key parts of the inverse problem: the data $\boldsymbol{y}$; the unknowns $\boldsymbol{m}$ that we seek to infer from the data; and a mapping G between the two (Dashti and Stuart, 2013). The mapping G may be linear or nonlinear. In practice, data is normally observed with noise, so that in practice we are dealing with the equation

$$\boldsymbol{y} = G(\boldsymbol{m}) + \boldsymbol{e}, \tag{1.3}$$

from which two problems arise; firstly, G may not have a unique inverse, and secondly, that even if G has a unique inverse (so that we could obtain $\boldsymbol{m} = G^{-1}(\boldsymbol{y} - \boldsymbol{e})$), we do not know what $\boldsymbol{e}$ is. For this introduction, we will assume uncorrelated noise with some characteristic level σ, but the noise could in principle have additional structure. Concisely stated, geophysical inverse problems are ill-posed—finite, noisy data will never be sufficient to perfectly recover unknown continuous Earth structure. In order to solve the inverse problem, additional *a priori* information must therefore be added, beyond what is provided by the data. The probabilistic formulation gives the natural setting for adding this *a priori* information. The fundamental object of interest in the probabilistic setting is the joint distribution of data and unknowns $p(\boldsymbol{y}, \boldsymbol{d})$. Using the law of conditional probability and rearranging we have Bayes' theorem:

$$p(\boldsymbol{m}|\boldsymbol{y})p(\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{m})p(\boldsymbol{m}) \tag{1.4}$$

$$p(\boldsymbol{m}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{m})p(\boldsymbol{m})}{p(\boldsymbol{y})}, \tag{1.5}$$

which may intuitively be read as

$$posterior = \frac{likelihood \times prior}{evidence}. \tag{1.6}$$

The posterior is what we are interested in—given the data, what is the probability distribution of unknowns? The likelihood, or the distribution of data given the unknowns, expresses the goodness of fit of our potential solution. The prior encodes

the additional knowledge about the unknowns, and the evidence tells us how likely the data is given the class of model being fit. The evidence is difficult to calculate, and is fixed for a particular class of model, so Bayes' theorem is normally given in the proportional form

$$p(\boldsymbol{m}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{m})p(\boldsymbol{m}). \tag{1.7}$$

The interplay between the class of model (which determines the form of the unknowns $\boldsymbol{m}$, the choice of likelihood function, and the prior, determines the outcome of the inverse problem. The most typical choice of both the likelihood and the prior is the normal distribution, with the *a priori* information being encoded by some operator R, so that the posterior has the form

$$p(\boldsymbol{m}|\boldsymbol{y}) \propto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{||G(\boldsymbol{m})-\boldsymbol{y}||_2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-R(\boldsymbol{m})}, \tag{1.8}$$

from which we can take the negative log to obtain

$$-\log p(\boldsymbol{m}|\boldsymbol{y}) \propto \frac{||G(\boldsymbol{m}) - \boldsymbol{y}||_2}{2\sigma^2} + R(\boldsymbol{m}), \tag{1.9}$$

if we assume a fixed data noise level σ. Finding the minimum of the negative log-posterior is equivalent to finding the *maximum a posteriori* point (MAP point), and shows the close relationship between a typical Bayesian geophysical inverse problem and (non)linear least-squares inversion. The often-raised assertion that probabilistic formulations are inherently subjective and therefore untrustworthy, when compared to regularization based approaches (Parker, 1994), can be seen to be made under false pretenses. The probabalistic formulation and the regularization based formulation have the same structure in most geophysical inverse problems, with the probabilistic approach merely giving a fuller account beyond just the MAP point. The main areas in which subjectivity plays a role are the choice of the regularization operator R, and the form of the model parameterization—how the unknowns $\boldsymbol{m}$ are translated into Earth structure—and regularization based approaches have no special claim

over objectivity in this regard. A particularly common form of regularization or *a priori* knowledge is the setting of R to a linear operator—this is commonly referred to as Tikhonov regularization (Aster et al., 2018), with $R = I$ and $R = \nabla^2$ often seen. If $\boldsymbol{m}$ directly represents the Earth by encoding the values of geophysical parameters at voxels, then these two choices have an obvious interpretation. $R = I$ penalizes deviations from zero, so a model with small amplitude is preferred; this choice is often used if the inverse problem is in a perturbative setting relative to a simpler reference model. $R = \nabla^2$ penalizes the roughness of the model, noting that there is no penalty to constant or linear models with this regularization. It is of course a subjective choice as to what *a priori* knowledge one encodes in an inverse problem—or equivalently, what penalties to include in the regularization!

Equally important is the choice of model parameterization. In geophysics, the two most common approaches are to use voxel based parameterizations, which have the advantage of directly representing Earth structure, and spectral representations, which often have particularly expedient mathematical properties. However, both of these common forms have significant drawbacks. In particular, spectral representations are inherently global and so struggle to represent datasets with spatially dependent resolution optimally. Voxel representations are typically more flexible due to spatial locality, but normally require very strong regularization for inversion stability, which often results in a blurry and uninterpretable output image. Using more opinionated model parameterizations transfers the workload of inversion stability from the regularization onto the parameterization, which is often easier to reason about using *a priori* geological knowledge; this is a concept that we investigate in this thesis in Chapters 5 and 6. Another approach of interest is that we may perform inversion in the data space itself, solving the problem of ill-posedness by synthesising additional data from observations. This is implicitly the approach taken in the recently popular wavefield tomography methods, such as eikonal and

Helmholtz tomography (Lin and Ritzwoller, 2011; Lin et al., 2009), which directly solve for Earth structure products without incorporating a true tomographic inverse problem, *per se*. Instead, they rely on adequately interpolating the observed seismic data such that the constitutive properties can be directly retrieved, i.e. the inverse problem is fundamentally in the data space. Data space inversions are increasingly important given the proliferation of large seismic arrays, and we investigate methods for performing them in Chapters 3 and 4.

1.2 Overview

This thesis presents the full lifecycle of the geophysical inverse problem, from initial collection of data, to preliminary analyses, design of the inverse problem, algorithms for obtaining a solution, and finally the selection of a final model from a principled basis. The work is laid out in roughly this order, although each chapter often cuts across multiple stages of the inverse problem workflow. The predominant theme throughout the work is to go beyond "standard" techniques to explore methods that may allow greater insight into earth processes, whether by obtaining a novel and particularly informative dataset, analysing this data in a powerful way or by bringing in *a priori* geophysical knowledge into the inversion process.

We begin in Chapter 2 with a presentation of the study of the shallow crust of Southern California from the horizontal-to-vertical (H/V) ratios of Rayleigh waves. H/V ratios have a sensitivity depth approximately $1/3$ that of Rayleigh phase velocity measurements at the same frequency, meaning that they are particularly well suited for isolating near-surface structures. By using the ambient-noise cross-correlation method to observe coherent Rayleigh wave propagation between pairs of stations in the Southern California Seismic Network, we observed H/V ratios at substantially higher frequencies than reported by previous authors, allowing for high-resolution imaging of sedimentary basin structure. This chapter broadly illustrates the themes

of the thesis overall, including the selection of an appropriate dataset & the development of an inverse problem to best exploit it.

In Chapters 3 and 4 we develop a comprehensive treatment of the interpolation of sparse seismic wavefields using compressive sensing. Beginning with direct observations of the time derivatives of the displacement wavefield in Chapter 3, we propose performing a two-stage wavelet transform consisting of a discrete wavelet transform in time followed by a Laplacian-preconditioned curvelet transform in space. To handle incomplete spatial data, we perform the curvelet transform in the compressive sensing setting, which poses the transform as an inverse problem with sparsifying regularization. Laplacian preconditioning promotes smoothness of the solution up to the second spatial derivative, promoting interpolations that satisfy the seismic wave equation. Consequently, the interpolated wavefields derived from this method may be used for advanced seismic data processing methods, such as wavefield gradiometry and Helmholtz-Hodge decomposition. In Chapter 4, we extend the methodology to incorporate strain measurements. We test the methodology using data from the Porotomo experiment (Feigl and Parker, 2019), which co-deployed temporary seismometers and fibre-optic distributed-acoustic-sensing (DAS) to a geothermal field at Brady, Nevada, and assessed the performance of both the individual seismic and DAS networks, as well as the heterogeneous mixed network, and found that the wavefield reconstruction paradigm was particularly useful in empirically quantifying the site response of the DAS cable, a current major research question that must be resolved before robsut DAS becomes widespread. Using our reconstruction framework, we also proposed methods for improving the fidelity of DAS records to true ground motions using homogenization theory, and also investigated the optimal design of mixed point-sensor and strain deployments within the compressive sensing setting.

In Chapters 5 and 6, we turn our attention from inverse methods for data collection to

inverse methods for the determination of Earth structure. Given that seismic inverse problems require some kind of *a priori* knowledge for their solution, we developed a methodology for explicitly introducing geologically motivated regularizations via the level-set method. By using level-sets, we can simplify the difficult problem of parameterizing explicit geological features (which may have an unknown topology) by defining them implicitly using the contour levels of a higher-dimensional latent field. We lay out the structure of this method, and its application to several synthetic and real-world travel-time tomography examples, in Chapter 5. In Chapter 6, we applied our method to a large-scale problem of real-world importance—the structure of the northeast Los Angeles (LA) basin. The Mw 6.4 and Mw 7.1 July 2019 Ridgecrest earthquakes generated strongly-excited surface waves within the basin (Filippitzis et al., 2021; Kohler et al., 2020). The pattern of amplification was poorly predicted by existing 3D velocity models, however the recent deployment of the permanent strong-ground-motion Community Seismic Network (CSN) introduced the possibility of updating these models using the Ridgecrest data using a dense network of stations. We used our level-set based method to invert for a basin update using Love-wave phase velocities and amplitudes, and found that the northeast LA basin is most likely significantly deeper, and with steeper sidewalls, than the existing reference model CVM-S4.26, which substantially changes the amplification characteristics of the basin. On the other hand, the basin depth in the most northerly part of the array is shallower than in the seismic models, which brings our model into concordance with the current understanding of the geology of the area. In Chapter 6, we also introduce a novel inversion scheme based on an extension of the Ensemble Kalman Sampler (Garbuno-Inigo et al., 2020) that efficiently calculates a derivative-free Bayesian approximate solution to the inverse problem for hierarchical parameterizations.

Finally, having obtained data and developed models from that data in the previous

chapters, we move on to the assessment of these models in Chapter 7. In the Earth sciences, there are often a range of parameterizations that are *a priori* reasonable, and so before doing the inversion we have no justifiable basis on which to pick one in particular. After performing the inversion, however, we have obtained more information, and can use that to reason about our decision as to which model to use in the future to make further predictions. Picking the best fitting model is rarely the best choice, as we inevitably encounter the problem of overfitting, leading to poor predictive accuracy. Using the famous travel-time curve of R.D. Oldham (Oldham, 1906) as an example, we developed a Bayesian framework for handling model selection for very imprecise data for which inaccuracies were present in both dependent and independent variables (as is often the case in historical datasets). We found that there was robust statistical evidence for Oldham's original conclusion that the Earth had a core using only a small subset of his data; it is now well known that the full dataset was fundamentally misinterpreted in the original paper.

Chapter 2

RAYLEIGH-WAVE H/V VIA NOISE CROSS CORRELATION IN SOUTHERN CALIFORNIA

Muir and Tsai (2017) previously published as

Muir, J. B. & V. C. Tsai (2017). "Rayleigh-Wave H/V via Noise Cross Correlation in Southern California". In: *Bulletin of the Seismological Society of America* 107.5, pp. 2021–2027. DOI: 10.1785/0120170051

2.1 Abstract

We study the crustal structure of Southern California by inverting horizontal-to-vertical amplitudes (H/V) of Rayleigh waves observed in noise cross-correlation signals. This study constitutes a useful addition to traditional phase-velocity based tomographic inversions due to the localized sensitivity of H/V measurements to the near-surface of the measurement station site. The continuous data of 222 permanent broadband stations of the Southern California Seismic Network (SCSN) were used in production of noise cross-correlation waveforms, resulting in a spatially dense set of measurements for the Southern California region in the 2.5–37.5 s period band. The fine inter-station spacing of the SCSN allows retrieval of high signal-to-noise ratio (SNR) Rayleigh waves at periods as low as 2.5 s, significantly improving the vertical resolution of the resulting tomographic image, compared to previous studies with minimum periods of 5–10 s. In addition, horizontal resolution is naturally improved by increased station density. Tectonic sub-regions including the Los Angeles Basin and Salton Trough are clearly visible due to their high short-period H/V ratios, while the Transverse and Peninsular Ranges exhibit low H/V at all periods.

2.2 Introduction

The development of noise cross-correlation techniques constitute one of the major developments of observational seismology in the past 15 years (Shapiro and Campillo, 2004). As a result of the many possible pairs of stations in a seismic network such as the SCSN ($n(n-1)/2$ independent pairs for a network of n stations), and the lack of dependence on earthquake source location, noise cross-correlation has delivered a level of data density that was unprecedented little more than a decade ago.

The fundamental result of the noise cross-correlation technique is that cross-correlating an ambient noise wavefield between two points results in a signal approximately proportional to the causal and anti-causal far-field Green's function between those two points (Boschi and Weemstra, 2015; Snieder, 2004). Practically, waveforms approximating the displacement Green's function between two seismic stations can be constructed by correlating noise traces (of length sufficient to capture the phase of interest) at each station in the time domain, and then stacking them to produce an average (Bensen et al., 2007).

The power spectral density of seismic noise is at a maximum in the 5-20 s period band, as a result of the primary and secondary oceanic microseisms at peak periods of 15 and 7 seconds respectively (Ardhuin et al., 2015). The dominance of oceanic noise also means that the global noise wavefield is primarily generated near the Earth's surface. Consequently, the seismic mode best observed in noise cross-correlation measurements is the short-period fundamental mode Rayleigh wave, as it is well excited by surface noise sources in this period band.

It has been generally thought that Rayleigh waves are more easily observed in noise cross-correlation than Love waves as P-SV motion is excited at the Earth's surface at a greater rate than SH motion. Cross-correlation studies have therefore focused

heavily on Rayleigh-wave techniques, especially the traditional measurements of phase and group velocities for tomographic imaging (Shapiro et al., 2005), although some studies have shown that Love waves may be clearly observed (Lin et al., 2008). As the cross-correlation technique matures, amplitude information derived from noise cross-correlation studies has begun to supplement velocity measurements.

Absolute amplitudes of noise cross-correlations, or even relative inter-station amplitudes, are difficult to interpret theoretically due to the differing effects of real noise distributions (e.g. Tsai (2011)); however, the amplitude ratios between noise cross-correlation components are more robust due to common noise source and wave propagation, particularly if care is taken to jointly, rather than independently, normalize all components (Lin et al., 2014). We note that noise cross-correlation derived H/V match earthquake H/V in their overlapping period range, providing empirical evidence of the robustness of the noise derived H/V (Lin et al., 2014). Similarly to traditional event waveform horizontal-to-vertical (H/V) amplitude ratios of Rayleigh waves, H/V of noise cross-correlations are highly sensitive to upper crustal structure in the immediate vicinity of the receiving station.

H/V ratios derived from single-station noise power spectral densities have the advantage of being easy to derive from even short deployments. However, it is difficult to definitively interpret single station measurements; they have been variously interpreted in terms of Rayleigh waves (Boore and Toksoz, 1969; Fäh et al., 2001), vertically propagating SH waves (Nakamura, 2000) or diffuse-wave theory (Sánchez-Sesma et al., 2011). The noise cross-correlation method presented here does not suffer from this theoretical uncertainty. The use of noise cross-correlation traces instead allows for clear identification of Rayleigh waves by observation of elliptical particle motion and windowing of seismograms around the expected Rayleigh-wave arrival time. Hence, H/V ratios derived from noise cross-correlations may be interpreted using Rayleigh-wave forward modeling. H/V ratios are especially sensitive

to basin structures, in which the large impedance contrast between the basin material and underlying bedrock results in a characteristically peaked H/V spectrum, exhibiting a basin resonance effect.

Figure 2.1 shows that Rayleigh-wave H/V sensitivity is concentrated at the near-surface, relative to the sensitivity of Rayleigh-wave phase velocities for the Hadley-Kanamori 1D velocity model of the Transverse Ranges, which is commonly used as a reference velocity model for Southern California (Hadley and Kanamori, 1977; Hutton et al., 2010). Sensitivity is even greater close to the surface for models containing low-velocity basins, as the eigenfunction amplitudes become sharply peaked in the basin at resonant frequencies. This high surface sensitivity, with suppressed sensitivity at depth relative to phase-velocity measurements, makes Rayleigh-wave H/V ratios an exciting prospect for highly vertically resolved tomographic studies of the near surface.

In this study, we have employed the regionally dense permanent Southern California Seismic Network (SCSN) to observe H/V values in Southern California. Southern California is an ideal test site for new seismic methodologies because it is simultaenously structurally complex, and well studied. Thus, there is a wealth of comparisons that can be made with previous results. Cutting edge tomographic results in especially complex regions, such as the Los Angeles Basin, rely on joint inversion for many seismic observables, up to and including full waveform inversions for which the forward problem is prohibitively expensive for many applications, and a high quality initial 3D model is required. In comparison, ambient noise cross-correlation is relatively inexpensive, with good results available after only a several-months long deployment of a small broadband network. H/V tomography (e.g. Li et al. (2016); Lin et al. (2012, 2014)) has utilized noise cross-correlations and long-period Rayleigh waves in a combined period band of 8–100 s. The 8–100 s period band results in good vertical resolution at a continental scale and combined

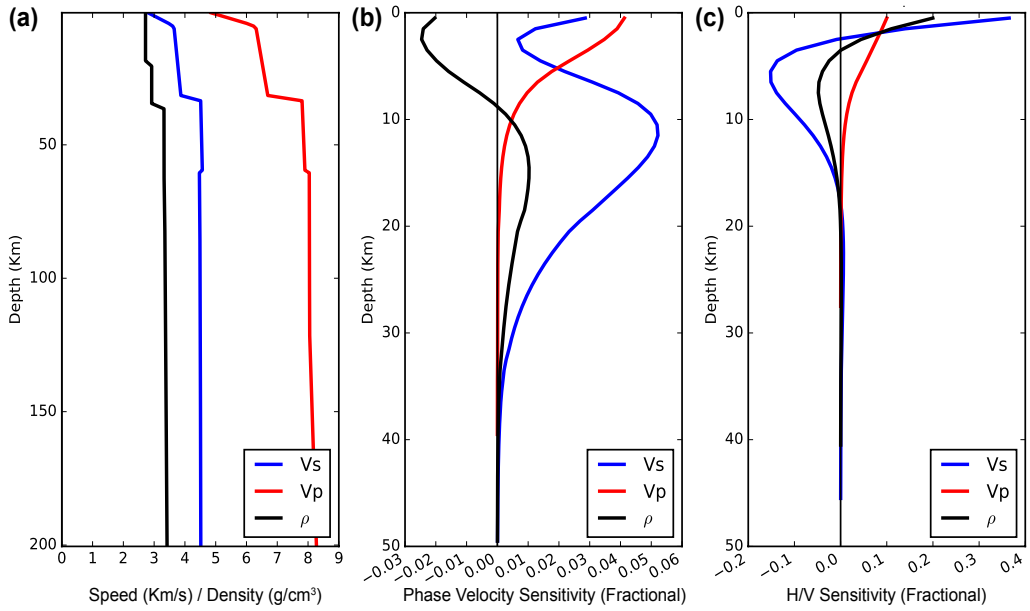


Figure 2.1: (a) Southern California Hadley-Kanamori 1D velocity model for P wave velocity (V_P), S wave velocity (V_S) and density (ρ) to a depth of 200km. (b) Rayleigh-wave phase velocity sensitivity to perturbations in V_P, V_S and ρ at depth at a period of 10s. (c) Rayleigh-wave H/V sensitivity to perturbations in V_P, V_S and ρ at depth at a period of 10s.

with phase-velocity measurements is sensitive to the whole crust; however, it does not contain the short-period data necessary to image the upper 10 km of the crust at kilometric length scales. Expanding the period range of Rayleigh-wave H/V via noise cross-correlation techniques, to a minimum period of 1 s should therefore prove especially useful in maximizing the utility of preliminary regional crustal studies, where events during station deployment are limited and the regional crustal model is of poor resolution.

2.3 Methodology

Three-component broadband (BH channel) data were obtained for the year 2015, for 222 permanent stations of the SCSN. The data were divided into 1 hour segments. The data were then pre-processed using the methods described in Bensen et al. (2007), with the following modifications: In order to facilitate efficient computation, the data were decimated from their natural sampling rate of 40Hz to 5Hz after being low-pass filtered to prevent aliasing. Note that the cutoff for the low-pass filter is 1Hz, well above the 0.4Hz maximum frequency used for computing H/V ratios for this study, and so does not effect reported results. For each station, all three components were normalized in the time and spectral domains using a common normalization signal, to maintain the relative amplitude information necessary for H/V ratio measurement; in contrast to coherency measurements of amplitude this method allows for a meaningful measurement of amplitude ratios. The common signals used to normalize the time and spectral domain records were taken to be the means of the single component signals for each channel, as described in Bensen et al. (2007).

H/V ratio measurements were then performed following Lin et al. (Lin et al., 2008). Nine component cross-correlations between all station pairs were calculated for each hour; the resulting cross-correlations were then stacked for all available hours in

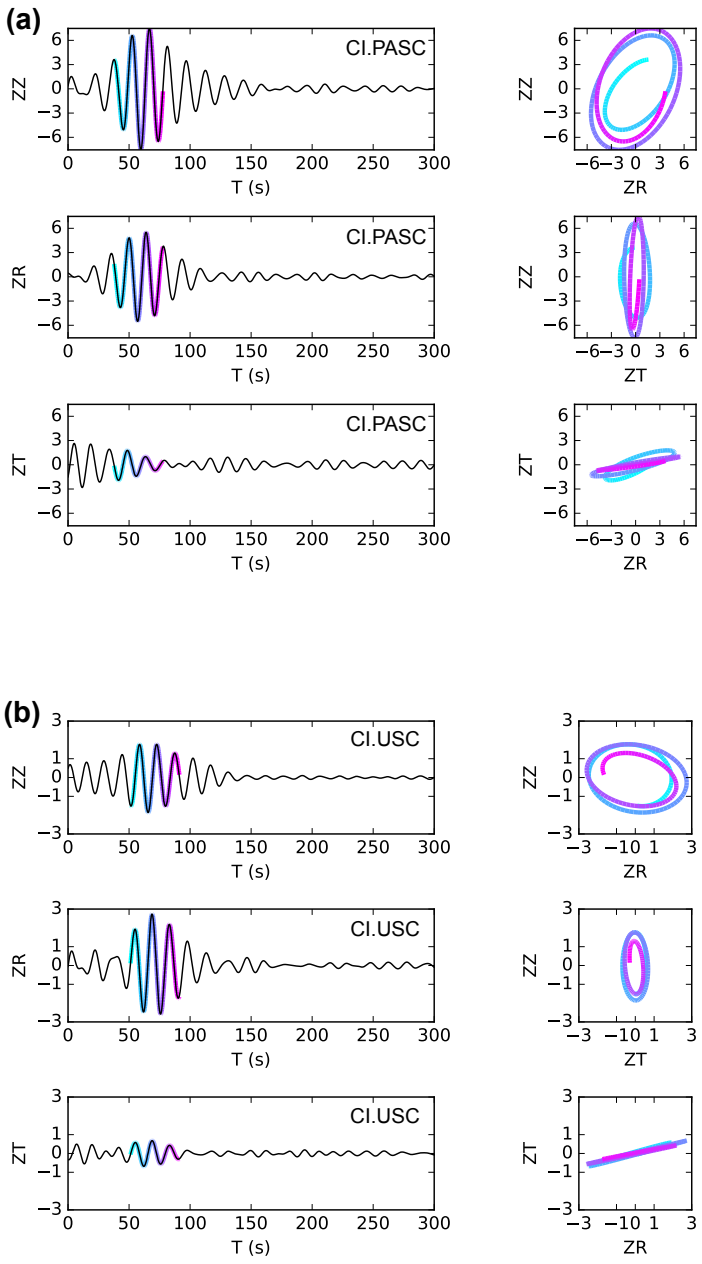


Figure 2.2: Waveforms and particle motions, filtered to a central period of 12.5s, for: (a), a hard rock site (Pasadena Arts Center, SNR of 92); (b) a basin site (University of Southern California, SNR of 73). The colored area shows the time interval used for H/V calculation.

2015 to produce the final hour-long averaged traces. To calculate H/V ratios, the cross-correlations were rotated from the measurement (east-north-vertical or ENZ) frame into the radial-transverse-vertical (RTZ) frame between the two stations, and then filtered to the period of interest using a Butterworth bandpass filter. To simultaneously rotate all 9 components into the correct RTZ frames, the rotation matrix may be simply calculated as $\mathbf{M} = \mathbf{M_1} \bigotimes \mathbf{M_2}$ where $\mathbf{M_{1,2}}$ are the 3 component rotation matrices from the ENZ to the RTZ frame for the individual stations and $\bigotimes$ is the Kronecker product. This calculation extrapolates to the case of a general 3D rotation—see Laub (2005) Chapter 13. We also calculated empirically derived backazimuths from ZR/ZT particle motion ellipses to account for any effect raypath bending away from the great-circle arc may have on the calculated H/V ratios. Using the empirical backazimuths increased the H/V ratios by a maximum of 10% in the 7.5–12.5 s period band; however, as rotation by the empirical backazimuths did not appear to substantially improve the calculated waveforms, we report only the great-circle path rotated results here.

To avoid misidentifying higher-mode Rayleigh waves as the fundamental mode of interest, the theoretical arrival times of the fundamental mode and first overtone were calculated, and the signal before the mean of these times removed. The phase gradient of the identified peak was logged to record the sense of motion of the arrival. As noted by Tanimoto and Rivera (2005), the sense of motion is period dependent and may switch from retrograde to prograde for the fundamental mode in the presence of steep surface velocity gradients, which can complicate the identification of the fundamental mode. As a result, further discrimination between potential first overtone and fundamental mode signals was not performed using the phase gradient. Furthermore, record sections of the calculated noise-cross correlations do not show coherent move-out of the first overtone. The root-mean-squared amplitude of the waveform near the maximum of the envelope of

each component was then used to make amplitude ratio measurements between components. Example waveforms, with the time intervals used for the calculation of H/V intervals highlighted, are shown in Figure 2.2.

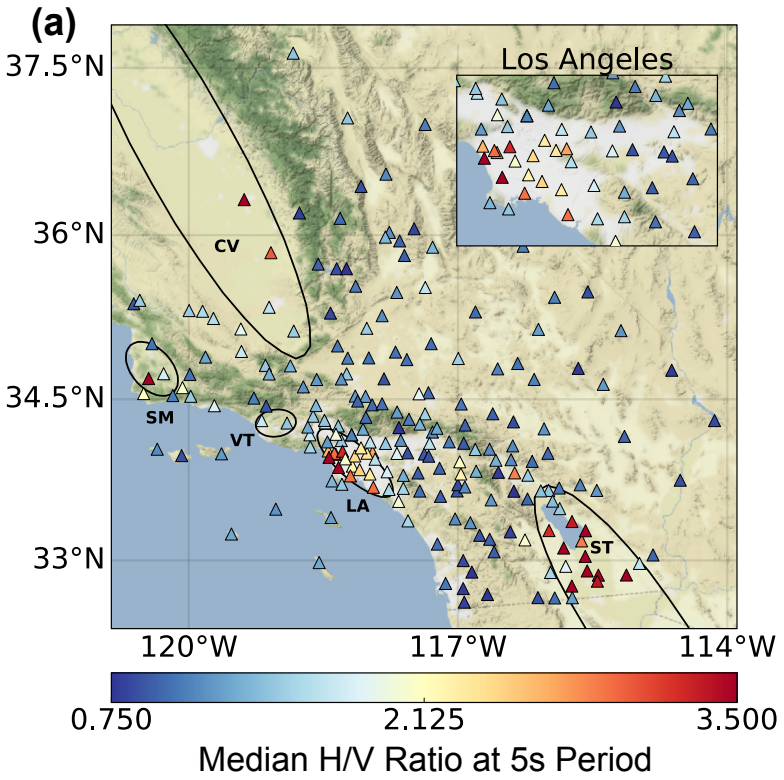
Once the H/V values for each station pair were collected, quality control was further performed by removal of low (<15) signal-to-noise ratio (SNR) measurements, with the last 100 s of the causal cross-correlation taken as the noise reference. The four components of interest in calculating H/V ratios (out of the nine calculated components) are ZZ, ZR, RZ, and RR, where the first letter corresponds to the component of the "source" and the second to the "receiver" for the causal cross-correlation signal (i.e. ZR corresponds to a vertical impulse at the virtual source being recorded on the radial component of the receiver). Both ZR/ZZ (initial vertical impulse) and RR/RZ (initial radial impulse) H/V measurements may be made; however, as the RR component of the causal cross-correlation failed the SNR criterion at a much higher rate than ZZ and ZR, the RR/RZ H/V values are not reported in this study. For each receiver station, each virtual source that satisfied the SNR criterion was used to generate summative H/V spectra. The ZR/ZZ H/V measurements are approximately log-normally distributed; for the purposes of inverting for velocity structure we assume that they follow a log-normal distribution at each period, for each receiver. Inter-station spacing varies widely throughout the SCSN, with the densest spacing in the Los Angeles basin. As H/V is principally dependent only on structure local to the receiver, the station density should not influence the inversion results with the possible exception of fewer short-period data in regions with low station density, as the amplitude of short-period waves decay rapidly with distance.

The resultant log H/V spectra were then used to invert for 1D near-surface structure underneath each of the available SCSN stations using the genetic algorithm global function minimizer provided by MATLAB. A 1D average velocity model

for Southern California provided by SCEC was used as the initial parametrization (Hutton et al., 2010). This smoothed initial parametrization captures some of the efforts of previous tomographic results, without overly conditioning the prior information. Theoretical H/V values were calculated using the finite element method of Lysmer (1970) to solve for Rayleigh-wave fundamental-mode eigenfunctions. The earth structure was parametrized by 5 layers (0-500 m, 500-1000 m, 1000-2000 m, 2000-4000 m, 4000-8000 m), with the V_P/V_S and V_P/ρ ratios set by the empirical relationships of Brocher (2005). This parametrization was chosen since five layers was the minimum required to fully fit the observed H/V peaks; the increasing layer thicknesses with depth helps to avoid overspecification of the forward model; however, the choice of parametrization was ultimately subjective.

2.4 Results and Discussion

Figure 2.3 shows an example of the median of the collected H/V results, in this case for central periods of 5 s and 17.5 s. The median H/V, standard deviation of the H/V logarithms and the counts exceeding SNR at each station at all periods from 2.5–37.5 s s are available in the electronic supplement to this article, figures S1-S15. A table of data containing all recorded amplitudes and SNRs is available as supplementary Table S1. Major sedimentary basin regions of Southern California (the Los Angeles Basin and the Salton Trough) are clearly visible as regions of elevated H/V ratios, particularly in the 5–20 s period band. Basins with smaller land surface expressions (Ventura Basin, Santa Maria Basin) and with fewer deployed seismometers (Central Valley) are also visible on a limited selection of stations. We expect the H/V ratios at each station to be closer to log-normal than normal in distribution (as the amplitudes are a product of many positive multiplicative factors). Maps of the standard deviation of the H/V logarithms show little correlation with known structures, other than a general increase within the Los Angeles Basin (see



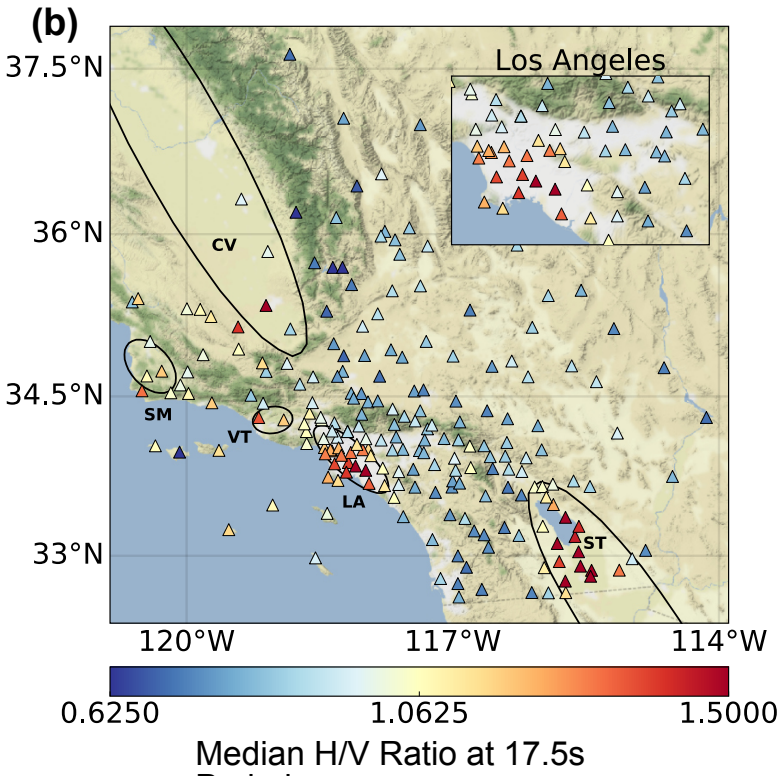


Figure 2.3: H/V aggregate measurements using the ZR & ZZ cross-correlation components at central filtering periods of: (a) 5s; (b) 17.5s. Inset plots show the Los Angeles basin area; note the different color scales for each subplot.

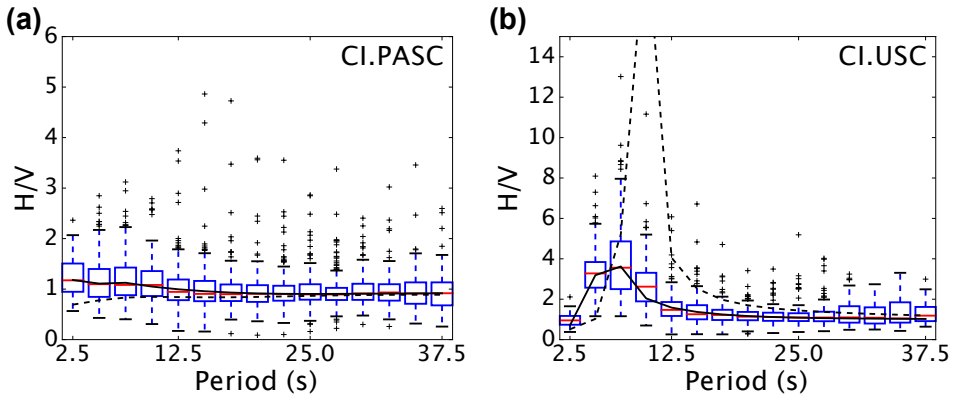


Figure 2.4: Example boxplots of H/V spectra with lines from fitting a layered structure (black) for (a) a hard-rock site (Pasadena Arts Center) and (b) a basin site (University of Southern California). Boxes show the middle 50th quantile of H/V ratio measurements at a particular period, with red bars showing the median. The lower and upper whiskers are at the 25th quantile minus 1.5 times the interquartile range (IQR) and the 75th plus 1.5 times the IQR, respectively, with remaining outliers shown as crosses. The dashed black lines show the predicted absolute H/V values from the Harvard Community Velocity Model (CVM-H).

Figures S31-S45, available in the electronic supplement to this article). It is difficult to speculate on the underlying cause behind the differing variance in H/V ratios, as it does not show consistent trends with either values of the H/V ratio, number of measurements exceeding the SNR criterion, the backazimuth to the source station, distance to the ocean (the predominant noise source at most periods), or station density.

Figure 2.4 shows two examples of H/V spectra as a function of period exhibiting typical hard-rock and basin site features. The H/V spectra is generally well fit by a five layer parameterization for nearly all stations. There is significant tradeoff between basin depth and near-surface velocity within inversions; however, fitting both the width and amplitude of the H/V peak, if present, does allow for basin depth and velocity to be independently resolved. Figure 2.5 shows the inverted V_S maps for the top two layers of the parametrization that result from use of the measured

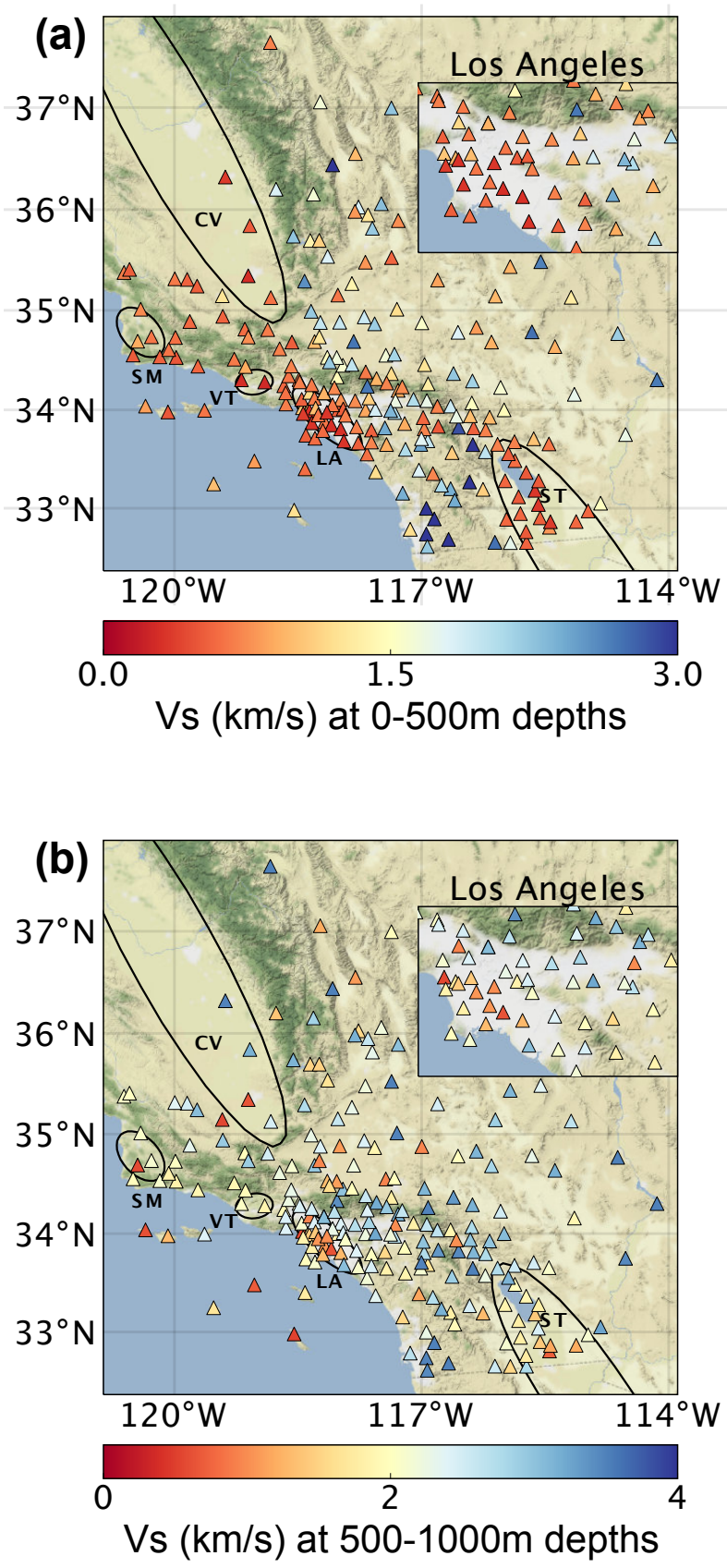


Figure 2.5: V_S models, inverted using ZR/ZZ data, for depth ranges of (a) 0–500 m and (b) 500–1000 m. Inset plots show the Los Angeles basin area; note the different color scales for each subplot.

ZR/ZZ H/V ratios. Basin structures are clearly observable within the inverted V_S model as regions of depressed velocity persisting several kilometers into the crust. We performed an ensemble Markov-Chain Monte-Carlo (MCMC) inversion of the vertical velocity profile underneath CI.USC to assess the uncertainties in the inversion results (see Figure S51, available in the electronic supplement to this article) using the *emcee* Python package, which natively handles the non-linear and highly correlated likelihood surface of the forward problem (Foreman-Mackey et al., 2013). The results show increased confidence in the inversion of the near surface parameters, as expected from the sensitivity behavior of H/V ratio measurements. The extensive computational requirements of the MCMC sampling process preclude us from performing this analysis for all stations.

A notable inference that can be made from Figure 2.5 is that inversions based on H/V ratios suggest a shallower effective basin depth (in the sense of horizontal seismic amplification) than is currently defined within the Southern California community velocity models (CVMs), of which we use the Harvard CVM (CVM-H, Shaw et al. (2015)). Ma and Clayton (2016), using Rayleigh and Love wave dispersion along the Los Angeles Syncline Seismic Interferometry Experiment ("LASSIE"), also found a shallower layer of very low velocities in the Los Angeles Basin than that contained in the CVM-H basin, although their deep velocity structure is more similar to the CVM-H than ours. The theoretical H/V ratios predicted by the CVM-H model are shown in Figure 2.4 as dashed black lines; for the USC station example the CVM-H predicts a larger amplitude and longer period than found in the empirical data. The mismatch between H/V ratios predicted by CVM-H and those measured empirically is frequent within basin areas; as can be seen in the PASC example, in hard rock sites the spectra is quite flat and there is not as strong evidence for potential mismatches within the period range studied here. Matsushima et al. (2014) and others have invoked non-planarity of subsurface interfaces to explain similar

discrepancies between observations and theory in microtremor H/V ratio data, and have obtained similarly reduced amplitudes and shorter periods. However, the effect they observed is significantly weaker than the mismatches between CVM-H predicted H/V and the H/V we have observed. Additionally, the effect of non-planar interfaces (and anisotropic velocity) should be apparent as an azimuthally dependent H/V ratio, for which there is only weak evidence in our dataset. Further exploration of this effect requires a large scale numerical simulation of ambient noise propagation within Southern California and is therefore outside the scope of this paper.

2.5 Conclusions

We have obtained noise cross-correlation derived measurements of Rayleigh-wave H/V ratios for the Southern California Seismic Network in the 2.5–37.5 s s period band. Robust statistics are obtained for the majority of stations, showing clearly elevated H/V ratios within the major basin regions of Southern California within the 5–20 s period band. The H/V ratios have also been inverted for 1-D profiles of V_S beneath the recording stations.

These results confirm the ability of dense seismic arrays to produce useful H/V measurements at shorter periods than in the optimally excited oceanic microseism period band. Recorded patterns of H/V ratios correlate very well with known structure in Southern California throughout the period range studied. These results provide a useful additional constraint on near surface structure that may be folded into large collaborative models such as the CVM-H, even within their high resolution basin zones, and could potentially improve near-surface resolution within the less studied areas of the Southern California region. Within this study, the use of a single theoretical framework (surface wave inversion) results in a self-consistent map of V_S, unlike existing CVMs which are often nonphysically discordant as a result of

resolution changes across the model according to data availability. Furthermore, the good concordance with known structure validates the use of short-period H/V measurements to develop detailed near-surface crustal models underneath other seismic arrays where the underlying crustal structure is less well known.

Inversion of 1D structure underneath a single station using either H/V values, or phase-velocity, is by necessity under-determined. This is a consequence of the significant tradeoffs between the three principle seismic parameters seen in the sensitivity kernels of Figure 2.1, for both H/V ratios and phase velocities. Removal of these tradeoffs would require the calculation of phase-velocity dispersion curves for the stations using this study. Incorporation of this data would also extend the depth to which the inversion is sensitive; however there would also by necessity be some mapping of deeper structure into the shallow upper crust observed by H/V ratios, dependent on the relative weighting of datasets. Use of only H/V ratio measurements in an inversion for velocity structure is an illuminating exercise in that it is by physical necessity only sensitive to the near surface and consequently does not suffer the poorer depth resolution of phase-velocity measurements. Accurate assessment of the relative data uncertainty will be required for a future joint inversion using both H/V ratio and phase velocity datasets. The decrease in measurements exceeding the SNR requirement at shorter periods indicate that this study has approached the limit of the short-period range accessible to noise-cross correlation surface wave measurements using the SCSN. This study therefore presents the best vertical resolution of the near surface crust achievable using regional surface waves in Southern California using current instrumentation.

2.6 Data and Resources

All available waveform data from the SCSN in the year 2015 were downloaded using the Seismogram Transfer Program (Southern California Earthquake Data Center

(SCEDC), 2013; http://scedc.caltech.edu/, last accessed July 2017). Station data were downloaded using the ObsPy International Federation of Digital Seismographic Networks client service (Beyreuther et al., 2010). MATLAB's genetic algorithm global function minimizers were used to invert horizontal-to-vertical (H/V) ratios for velocity structure (www.mathworks.com/products/matlab, last accessed July 2017). Maps were created using Cartopy (http://scitools.org.uk/cartopy/, last accessed July 2017), with background images provided by Stamen Terrain.

2.7 Acknowledgements

The authors would like to acknowledge Daniel Bowden for his thoughts, and advice during initial code development. Jack Muir would also like to acknowledge the General Sir John Monash Foundation and the Origin Energy Foundation for their financial support. We would also like to thank the three anonymous reviewers and the editors of BSSA for their input to the manuscript. This version of the manuscript has been corrected to include a published erratum. The authors would like to thank Ellen Syracuse for identifying the problem and its source within the codes used to process data. The authors would also like to thank Editor-in-Chief Thomas Pratt and Betty Schiefelbein for reviewing this erratum.

2.8 Supplement

The number of supplementary figures included in this paper preclude their reproduction in this thesis, and the author directs readers to the online version of the manuscript for access to them.

Chapter 3

SEISMIC WAVEFIELD RECONSTRUCTION USING A PRECONDITIONED WAVELET-CURVELET COMPRESSIVE SENSING APPROACH

Muir and Zhan (2021) previously published as

Muir, J. B. & Z. Zhan (2021). "Seismic Wavefield Reconstruction using a Preconditioned Wavelet-Curvelet Compressive Sensing Approach". In: *Geophysical Journal International* 227.1, pp. 303–315. DOI: 10.1093/gji/ggab222

3.1 Abstract

The proliferation of large seismic arrays have opened many new avenues of geophysical research; however most techniques still fundamentally treat regional and global scale seismic networks as a collection of individual time series rather than as a single unified data product. Wavefield reconstruction allows us to turn a collection of individual records into a single structured form that treats the seismic wavefield as a coherent 3D or 4D entity. We propose a split processing scheme based on a wavelet transform in time and preconditioned curvelet based compressive sensing in space to create a sparse representation of the continuous seismic wavefield with smooth second order derivatives. Using this representation, we illustrate several applications, including surface wave gradiometry, Helmholtz-Hodge decomposition of the wavefield into irrotational and solenoidal components, and compression and denoising of seismic records.

3.2 Introduction

The character of progress in observational seismology has to a large extent been controlled by the quality and quantity of data. As such, the increasing availability

of seismic arrays with large numbers of instruments (large-N arrays) to the research community has become one of the major instrumentation themes of 21^{st} century seismology. Array seismology brings both opportunities, in the form of spatial analysis techniques such as backprojection (Kiser and Ishii, 2017), eikonal / helmholtz tomography (Lin and Ritzwoller, 2011; Lin et al., 2009) and wave gradiometry (de Ridder and Biondi, 2015; Langston, 2007a,b), as well as challenges associated with processing ever larger data volumes (Kennett and Fichtner, 2020). These technical trends will increase as more large-N array data is recorded, especially with the advent of Distributed Acoustic Sensing (DAS) arrays, from which data sets with many thousands of channels recorded at 100 Hz are now routinely recorded (e.g. Li and Zhan (2018); Lindsey et al. (2017); Williams et al. (2019); Yu et al. (2019)).

One of the most significant opportunities presented by large-N arrays is the transition from collections of individual seismograms to unified analysis of the seismic wavefield as a single entity, for which we can evaluate the ground motion at an arbitrary point in time and space within the array. In this study we use the term wavefield reconstruction to describe the process of synthesising individual seismograms into a coherent product. This product allows for additional analyses, such as robust wavefield gradiometry (de Ridder and Biondi, 2015), that fully utilize the behaviour of the wavefield in both space and time.

Ensuring that seismic instruments are of sufficiently high quality to measure ground motions to high fidelity in time was one of the great instrumentation achievements of the last century of seismology. However in most research settings, spatial sampling of the wavefield is insufficiently dense to capture all details of interest within the desired temporal frequency bandwidth. Because seismic arrays used in research settings typically only take sparse and uneven spatial measurements, the development of methods for optimal reconstruction of the continuous wavefield is an open research question. Industry scale results have typically focused attention to the

problem of optimally filling missing or corrupted channels in an otherwise regularly sampled deployment. Much effort in industry has recently focused on time and scale localized transforms (e.g. Herrmann and Hennenfent (2008); Herrmann et al. (2008)), which can allow for sparser representation of complex wavefields and consequently improved robustness of the reconstruction under noise, but for which the most efficient algorithms require structured data volumes.

At regional scales, methodologies have had to contend with the aforementioned spatial sparsity. Suggested methods have included: perturbations of a reference plane wave by smoothed splines (Sheldrake, 2002), applicability of which is limited to single phases with constant slownesses across the array; radon transforms using a general plane wave basis (Wilson and Guitton, 2007); compressive sensing using a plane wave basis (Zhan et al., 2018); and recently tensor completion methods utilising local rank-reduction to capture curved wavefronts (Chen et al., 2019b). Past regional scale wavefield reconstruction methods have focused primarily on plane-wave based reconstructions due to their attractive theoretical properties and potential for good angular resolution given adequate array design; however the plane wave assumption performs poorly for wavefields close to the source or with local scattering, for which the full spatial frequency spectrum is required to reconstruct the wavefield, leading to difficulties in resolving the required plane wave coefficients.

In this study, we present a wavefield reconstruction method based on wavelet analysis, specifically employing temporal wavelet analysis on individual seismic channels and then employing weighted curvelets for spatial analysis. This algorithm allows for a fully time/space and scale localized transform on unstructured seismic data. The underlying physics of wave propagation in continuous media is controlled by balancing the rate of change of momentum in a volume with the stress gradients (and body forces) applied to the volume through Newton's second law: $\rho\ddot{u}_i = \sigma_{ij,j} + f_i$. If the constitutive relationship of the medium is elastic, the stresses σ_{ij} are expressed

in terms of strains $\sigma_{ij} = c_{ijkl}\epsilon_{kl}$ with $\epsilon_{kl} = \frac{1}{2}(u_{i,j} + u_{j,i})$ (Aki and Richards, 2002). These fundamental relationships of wave physics suggest an important consideration in wavefield reconstruction; namely, that the physical wavefield does not depend on the displacement field itself, but rather its second time derivative, and, approximately, its second spatial derivatives. This implies that any wavefield reconstruction algorithm must attempt to not only fit the data, but also produce physically reasonable second derivatives if wave physics is to be respected. Accurately recording the spatial derivatives of the wavefield places significantly tighter constraints on the quality and density of the recorded wavefield than does merely recording the undifferentiated ground motions. The wavefield reconstruction algorithm presented in this study promotes the reconstruction of wavefields that obey wave physics by introducing a Laplacian based preconditioner into the compressive-sensing optimization problem, and by recognizing that due to the non-stationary power content of earthquake wavefields, a time-frequency representation is required to optimize the regularization of the spatial reconstruction transform. After introducing the algorithm, we present three case studies utilizing real seismic data to emphasise practical applications.

3.3 Wavefield Reconstruction Algorithm

Data Quality Control and Processing

For the purposes of designing a data quality control and processing workflow, the most notable characteristic of the proposed wavefield reconstruction algorithm is that the final, compressive-sensing, step it is performed on the whole data volume of an array simultaenously, rather than on a trace-by-trace basis. Therefore, the algorithm requires that each individual trace has an equivalent instrument response, or, if a heterogeneous collection of instruments is being used, that the instrument response is removed from all traces to obtain true ground motions. Because the

regularized curvelet inversion utilized in the spatial reconstruction step uses a least-squares metric for assessing data fit, it is relatively sensitive to amplitude outliers. Consequently, all examples in this paper using real datasets were manually checked for malfunctioning channels, which were removed. For larger datasets, it is likely that an automated procedure for identifying malfunctioning channels is required, however as the method is likely to change depending on application, we have not presented a general strategy here and instead describe what actions have been taken for each example individually.

Wavelet Transform

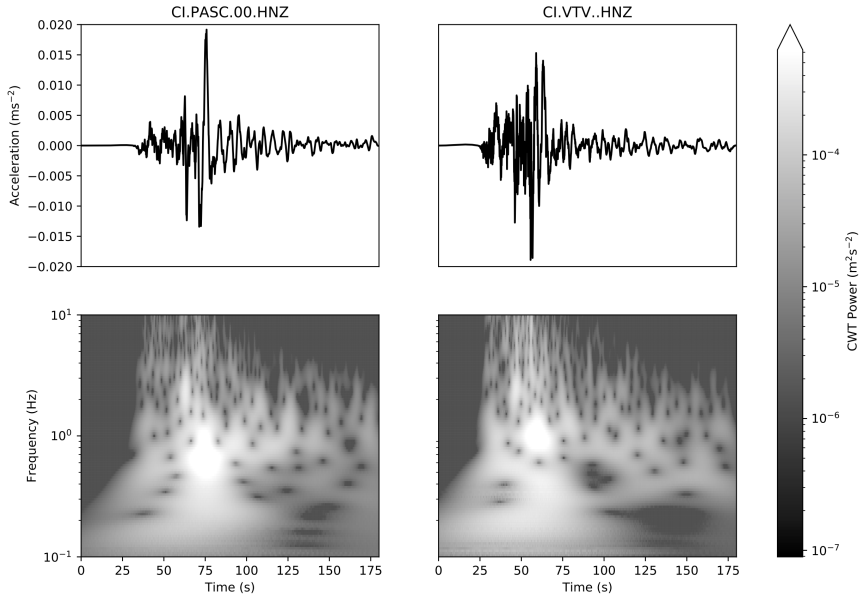


Figure 3.1: Continuous Wavelet Transform (CWT) of two strong ground motion accelerograms of the July 6 2019 Mw 7.1 Ridgecrest Earthquake. The CWT highlights the general observation that the power of signals from seismic events is typically non-stationary in both time and frequency.

After initial processing of the data, each record is then transformed into a time-frequency representation via a wavelet transform. The wavelet transform comes in both continuous and discrete forms; heuristically both act by representing the

signal in terms of a scalings and translations of an underlying mother wavelet. The behaviour with scale encodes the frequency characteristics of the signal, whilst the translations encode behaviour in time. The choice of using a discrete (DWT) or continuous (CWT) wavelet transform is determined by the objective of the study.

For a given mother wavelet $\psi(t)$, the CWT of a given signal $u(t)$ is given by

$$w(j_w, s) = \frac{1}{\sqrt{j_w}} \int_{-\infty}^{+\infty} u(t)\psi^* \left(\frac{t-s}{j_w} \right) dt, \tag{3.1}$$

where ψ^* is the complex conjugate of ψ. As the CWT is overcomplete, the inverse transform is non-unique, however the natural inverse (using the same mother wavelet as the original transform) is given by

$$u(t) = \frac{1}{C} \int_0^{\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{j_w}} w(j_w, s)\psi \left(\frac{t-s}{j_w} \right) \frac{dj_w ds}{j_w^2}, \tag{3.2}$$

with the normalising constant C calculated in the frequency domain by

$$C = \int_0^{\infty} \frac{\tilde{\psi}^*(\omega)\tilde{\psi}(\omega)}{\omega} d\omega. \tag{3.3}$$

In practice, other functions may be simpler to implement—in this paper we use the δ function reconstruction as suggested by Torrence and Compo (1998). This normalising factor calculated in Equation 3.3 implies an admissibility condition, which is that the mean of the mother wavelet (i.e. $\tilde{\psi}(0)$) be 0 so that the integral converges.

The CWT provides high fidelity and for typical seismic wave packets the choice of the Morlet mother wavelet is usually excellent. The (admissibility corrected) Morlet used in this study is given by

$$\psi(t) = \sqrt[4]{\pi} \left(e^{i\omega_0 t} - e^{-\frac{\omega_0^2}{2}} \right) e^{-\frac{t^2}{2}}, \tag{3.4}$$

where ω_0 is a non-dimensional frequency constant, set here to 6 as suggested by Torrence and Compo (1998)—we note that the Morlet is not strictly admissible but the corrected form presented here is sufficiently close for practical purposes.

The CWT has the advantage that it permits explicit processing in the time-frequency domain, for example wavelet based denoising and windowing as suggested by Mousavi and Langston (2016) and Langston and Mousavi (2019). However, the CWT is redundant and so the computational complexity and storage requirements of employing it are high. Figure 3.1 shows two filtered seismic waveforms and the amplitude of their time-frequency representations by employing the CWT using a Morlet mother wavelet. The most salient feature of the time-frequency representations of the CWT is that the signal power depends strongly on time within the signal for each frequency, and strongly on frequency at each time, i.e. the signal is non-stationary within the time-frequency domain. This suggests that the optimal regularization parameter for compressive sensing in the curvelet domain is likely to be different for each wavelet coefficient.

The DWT provides a more computationally efficient transform than the CWT. The implementation details of DWT transforms are substantially more involved than that of the CWT—we refer the reader to Starck et al. (2010) for a review. In contrast to the CWT, the DWT is not redundant, however the choice of mother wavelet for DWT analysis is non-trivial and typically requires some *a priori* knowledge or test data to optimize the DWT representation, as the final reconstruction performance can be significantly impaired by a bad choice of mother wavelet, compared to the generally robust CWT for seismic data using Morelets (Langston and Mousavi, 2019; Mousavi and Langston, 2016). Additionally, the structure of the DWT means that it is not suitable for time-frequency analysis; however, it often forms a suitable representation for data compression and denoising. For reasons of computational efficiency, all of the case studies that follow employ the DWT for time-frequency analysis, however CWT based redundant reconstructions based on the Morelet typically produce the best results and may be necessary if the highest quality is required. We have found that the Daubechies wavelet family typically performs well for the examples shown

in this study.

Compressive Sensing in the Curvelet Domain

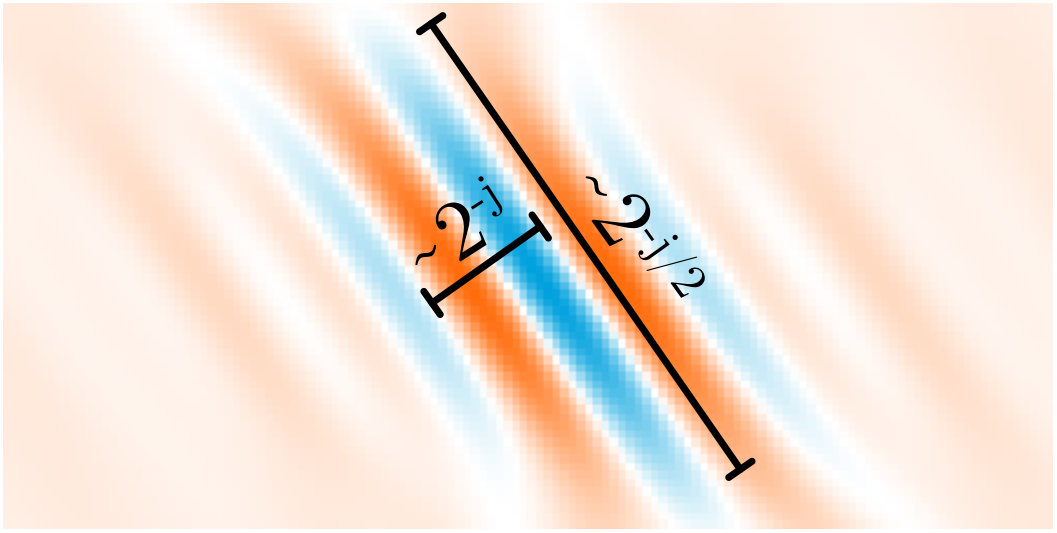


Figure 3.2: Example curvelet in the spatial domain showing angular sensitivity and characteristic parabolic scaling relationships between wavefront-parallel and wavefront-perpendicular directions.

Compressive sensing (CS) provides a mechanism to recover an unknown signal from relatively few measurements, such as is encountered by spatial sampling of seismic data in research settings. CS theory asserts that if the signal d can be sparsely represented by a basis or frame Φ and the signal is sampled incoherently with the basis or frame by a sampling operator Ψ, then solving the Elastic-Net regularized linear inverse problem

$$\hat{\mathbf{m}} = \text{argmin}\left[\frac{1}{2}||\mathbf{\Psi\Phi P m} - \mathbf{d}||^2 + \alpha\lambda||\mathbf{m}||_1 + (1-\alpha)\lambda||\mathbf{m}||_2\right] \tag{3.5}$$

with $\alpha = 1$ provides an accurate and sparse reconstruction of d in the chosen basis or frame (Candes et al., 2008; Davenport et al., 2011; Donoho, 2006) (The $\alpha = 1$ is also known as Lasso regression, while $\alpha = 0$ is known as Ridge regression and corresponds to classical Tikhonov regularization in geophysical inverse problems).

Setting α slightly less than 1 still typically results in a sparse representation of the signal in the chosen basis or frame, but may help to stabilize the inversion in the presence of noise. The requirement that Ψ be incoherent with Φ is achieved by having a sensing matrix that takes point measurements; the coherence of a distributed set of Dirac deltas with a spatially extended frame Φ is low as long as the sensors are not tightly clustered.

A frame that is particularly suited to use in seismic applications is the curvelet frame, which has been designed to promote sparsity for signals with wave-propagation like properties (Candes and Demanet, 2005). The k-space support of an individual curvelet is a dartboard-like wedge segment, which gives curvelets directional sensitivity, as well as scale localisation (Candès et al., 2006). Curvelets are constructed to obey a parabolic scaling relationship with the spatial extent perpendicular to wave propagation scaling like the square root of that parallel to propagation, which accounts for their directional sensitivity (see Figure 3.2). The curvelet frame has consequently been used in industry scale reconstruction and denoising applications using compressive sensing (e.g. Hennenfent et al. (2010); Hennenfent and Herrmann (2006); Herrmann and Hennenfent (2008); Herrmann et al. (2008)), for synthetic regional scale examples applied to surface wave tomography (Zhan et al., 2018) and recently as a filter for scattered waves by Zhang and Langston (2019), who employed the explicit curvelet transform for densely sampled data interpolated to a regular grid rather than the optimization based approach employed here for sparse data.

An optimal interpolation at points sampled by another sampling matrix $\hat{\mathbf{\Psi}}$ is given by $\hat{\mathbf{\Psi}}\mathbf{\Phi}\mathbf{P}\hat{\mathbf{m}}$. This formulation differs slightly from the normal presentation in that we have included an additional weighting matrix $\mathbf{P}$, that rescales the basis coefficients to reflect desired properties. In particular, curvelets are indexed by a scale j_c (as well as rotation and translation indices). To promote smoothness of the second derivative, we employ $\mathbf{P} = Diag(2^{-j_c{}^n})$ as our weighting matrix, which penalises

rapid oscillation of the reconstructed signal. Application of this matrix is in effect somewhat similar to applying Tikhonov regularization to the compressive sensing problem, with $n = 2$ corresponding to smoothing by penalizing the Laplacian of the wavefield, however it removes the need for a second regularization parameter and also puts the problem in a form that can be used by all existing L1 solvers.

To compactify future notation, we employ the traditional notation of linear inverse problems $\mathbf{G} = \mathbf{\Psi\Phi P}$ with i^{th} row $\mathbf{G}_i$ representing the projection of the basis or frame elements on a seismic channel indexed by i. For the proposed wavefield reconstruction algorithm, we solve the inverse problem for each wavelet coefficient

$$\hat{\mathbf{c}}(j_w, s) = \underset{\mathbf{c}(j_w,s)}{\arg\min} \left[\frac{1}{2} \sum_{i=1}^{N_{traces}} (w_i(j_w, s) - \mathbf{G}_i \mathbf{c}(j_w, s))^2 + \alpha\lambda ||\mathbf{c}(j_w, s)||_1 + (1-\alpha)\lambda ||\mathbf{c}(j_w, s)||_2 \right], \tag{3.6}$$

to give a sparse collection of estimated spatial curvelet coefficients $\hat{\mathbf{c}}(j_w, s)$ for each temporal wavelet. As we discussed during the wavelet transform component of the algorithm, broadband seismic signals are typically highly non-stationary in power and frequency content. Consequently, the optimal regularization parameter λ will change for each collection of wavelet coefficients $\{w_i(j_w, s)\}_{i=1}^{N_{traces}}$, sometimes by several orders of magnitude. To some extent this problem can be ameliorated by normalizing the prediction matrix $\mathbf{G}$ and wavelet coefficients $\{w_i(j_w, s)\}_{i=1}^{N_{traces}}$, typically by the standard deviation of $\mathbf{\Psi\Phi}$ and $\{w_i(j_w, s)\}_{i=1}^{N_{traces}}$ respectively. However, this still leaves the question of how to optimally choose λ. We have found good reconstruction results by optimizing the posterior predictive accuracy via 5-fold cross-validation. We choose the value of λ that minimizes the summed squared differences between the left out data and the predictions across the validation data sets. Cross-validation is in general an expensive operation, which motivates the use of a fast cross-validated L1 solver for this step; we employ the Celer solver (Massias

et al., 2018) which has proven itself to greatly outperform other L1 solvers during testing, including those on the GPU. For large scale compressive sensing operations, cross-validation becomes computationally infeasible; in that case, we employ the Corrected Akaike Information Criterion (AICc) using the Lasso.jl software package to perform the inversion, which is a fast Julia language port of the R language glmnet solver (Friedman et al., 2010).

Reconstruction

Reconstruction of the wavefield, either for recorded seismic channels, or at unobserved synthetic channels, is performed by generating the matrix $\hat{\mathbf{G}}$ with rows $\hat{\mathbf{G}}_k$ that describes the sampling-basis transform-weighting product $\mathbf{\Psi\Phi Pm}$ evaluated at the channels indexed by k. The predicted wavelet coefficients for each channel are then given by $\hat{w}_k(j_w, s) = \hat{\mathbf{G}}_k\hat{\mathbf{c}}(j_w, s)$. The time domain signal for the channel is then recovered by performing an inverse wavelet transform $\hat{u}_k(t) = IWT(\hat{w}_k(j_w, s))$. For the CWT, the inverse transform is not unique due to the redundancy of the transform and an appropriately normalized inversion wavelet must be prescribed to take this into account; in contrast the DWT has a unique inverse transform (Daubechies, 1992). It must be noted that this synthesis step encodes the regularizing assumptions inherent in the previous analysis steps that allow us to write the wavefield as a collection of curvelet coefficients in space and wavelet coefficients in time, namely that coherently propagating waves are present in the wavefield and we expect them to have smooth Laplacians. As such, any further usage of the wavefield must be consistent with these assumptions.

Summary

To summarize the methods detailed above, we propose the following generic algorithm for wavelet-curvelet wavefield reconstruction:

1. *Data Preprocessing*: Individual traces $u_i(t)$, indexed by channel i and time t, are quality controlled to ensure timing and amplitude are correct.

2. *Wavelet Transform*: Individual traces $u_i(t)$ are put into a time-frequency representation using a common wavelet transform to give a collection of wavelet coefficients $w_i(j_w, s) = WT(u_i(t))$ indexed by time shift s, channel i and wavelet scale j_w.

3. *Curvelet Transform*: For each wavelet coefficient, an L1 regularized optimization problem of the form given in Equation 3.6 is solved to determine sparse curvelet coefficients $\hat{\mathbf{c}}(j_w, s)$ for shift s and wavelet scale j_w describing the spatial distribution of the wavelet coefficients $w_i(j_w, s)$ across the array.

4. *Reconstruction*: To evaluate the ground motion at a particular location indexed by k, we form the curvelet reconstruction matrix $\hat{\mathbf{G}}_k$, and the reconstructed ground motion is given by $\hat{u}_k(t) = IWT(\hat{\mathbf{G}}_k \hat{\mathbf{c}}(j_w, s))$

3.4 Wavefield Gradiometry for the Southern California Array

Array based seismic wavefield tomography techniques, in particular eikonal tomography (Lin et al., 2009), have, in the last decade, become a major component of the regional-global structural seismic workflow, especially in the context of ambient-noise cross correlation datasets (e.g. Berg et al. (2018, 2020); Bowden et al. (2017); Lin et al. (2014)). These methods use various combinations of derivatives of wavefield observables to obtain information such as surface wave phase velocities. Because of the difficulty of accurately computing derivatives on a sparse collection of data and the resultant amplification of noise, this class of techniques is a good target for wavefield reconstruction as a first step in data processing so as to obtain robust derivatives.

The principle theoretical advantage of array-wavefield techniques, compared to

classical surface-wave tomography, is that they automatically account for wavefront bending without having to iteratively solve for the velocity structure to compute raypaths. However, the commonly used first-order eikonal tomography, and its second-order correction, Helmholtz tomography, rely on accurate identification of a phase delay time gradient $\nabla\tau$, requiring that the wavefield is dominated by a single phase front and that the phase delay time can be measured, with Helmholtz tomography additionally requiring the Laplacian of a fitted amplitude surface. Array methods based on a single wave ansatz include the first order wave gradiometry methods of Langston (2007a,b), which allow time-frequency resolution of a wave-packet containing multiple components including body waves; however the single wave assumption again precludes the ability to handle scattering or multipathing. Array methods that rely on the full wave equation, such as that proposed by de Ridder and Biondi (2015), do not require this assumption and can instead utilize multipathed or interfering wavefields. As both the wave equation based methods and the Helmholtz tomography correction to eikonal tomography utilize second order spatial derivatives of the wavefield, they impose much stricter quality requirements on the spatial resolution of the wavefield compared to eikonal tomography. In its most basic form, wave equation based wavefield gradiometry assumes that the wavefield, filtered at a particular frequency, is dominated by a single-mode surface wave with displacement u, which can be described by the simple acoustic wave equation

$$\frac{\partial^2 u}{\partial t^2} = c_p^2 \nabla^2 u \tag{3.7}$$

for phase velocity c_p. If both the temporal and spatial second derivatives of u can be accurately measured, then the squared phase velocity (with associated error) is given by simple linear regression. In practice, given the difficulties associated with these measurements, the regression is either solved in a hierarchical setting (de Ridder

and Biondi, 2015) or combined with wavefield reconstruction in a partial differential equation (PDE) constrained inverse problem (de Ridder and Maddison, 2018). In all the aforementioned studies, the typical rule-of-thumb given for the station spacing appropriate for accurate computation of the Laplacian of the wavefield is a super-Nyquist $\sim$ 10 stations per wavelength, a requirement that is rarely met in regional–global scale station deployments for wavelengths of interest. Exact performance analysis depends on the geometry of the array, in particular the accuracy of the second-order finite difference operator. An interesting application of the CS based wavefield reconstruction method proposed in this study is to decrease the station density requirements of array based tomography methods, potentially enabling the use of more powerful Laplacian based methods such as Helmholtz tomography and wave equation based gradiometry at the higher frequency/station spacing ratios that are typical for non-industry array deployments. After application of CS in the wavelet domain, the wavefield can additionally be projected onto a dense Cartesian grid, allowing the most accurate computation of numerical second derivatives.

As an example of this, we performed a wave equation gradiometry tomography experiment for the Southern California Seismic Network (SCSN). Wave equation based gradiometry, in particular, places the highest quality requirements on the Laplacian of the wavefield as it utilizes the phase of the wavefield, in addition to its amplitude. Obtaining reasonable values of phase velocity from wave equation gradiometry therefore serves as a practical demonstration that the wavefield reconstruction is accurately recovering the true wavefield, up to the second derivatives in space and time.

The interquartile range of the distances to the four nearest neighbor stations in the SCSN, calculated across the network, is 14–28 km. This corresponds to a station spacing of between 2.5–5 stations per wavelength for a period of 20s, assuming a nominal phase velocity of 3.5 km/s, and up to 7 -14 stations/wavelength for 50s

period waves with, assuming a nominal phase velocity of 4 km/s. As such, large parts of the SCSN fail to match the 10 stations/wavelength requirement for wavefield gradiometry in this period band, which is a typical band for earthquake based array tomography methods.

For the source wavefield, we utilized the Rayleigh wave packet of the November 19, 2017 Mw 7.0 Loyalty Islands earthquake (GCMT code 201711192243A, Dziewonski et al. (1981); Ekström et al. (2012)). This earthquake had a shallow normal faulting mechanism that directed strong Rayleigh waves towards California. We analysed the BHZ channel of the SCSN, integrated to displacement and filtered between 10–100 s. Quality control was performed on a per-trace basis by calculating the maximum normalized cross-correlations between a test trace and the signal at all other stations, and then requiring that at least 50% of these normalized cross-correlation values be better than 0.6 for the test trace to be retained. Additionally, we required that the root-mean-square log amplitude be within three standard deviations of the average across stations. These quality control measures resulted in a final array of 221 stations from an initial set of 234 with data. The wavefield reconstruction was performed using a 64×128 pixel curvelet transform, with the db4 wavelet (the Daubechies family wavelet with 4 vanishing moments) used for the temporal transform. Figure 3.3 shows a cross-section of the reconstruction results for this earthquake. After performing the reconstruction, we apply narrow band Butterworth filters with a width of $1/\sqrt{20}$ the period, and then calculate C_p using Equation 3.7. Examples of this are also shown in Figure 3.3, and show that the scaled Laplacian does typically fit the acceleration records well for this data. The resulting phase velocity curves are physically reasonable, and match the theoretical phase velocity curves from the CVM-H model (Shaw et al., 2015) well, especially within the basin. Figure 3.4 shows a spatial map of the phase velocity recovered for a period of 45 s. The wave equation gradiometry discussed here does not formally

account for finite frequency effects as we do not constrain the wavefield to fit the wave equation; to handle this, and reduce the variance of the results, we apply radial basis function smoothing using a multiquadratic basis function with length scale $45km$ (approximately $1/4$ wavelength), and only report results for points within the convex hull of stations within 90km of that point. The equivalent model without radial basis function smoothing is shown in Supplementary Figure 3.8. These results are comparable to existing surface wave tomography maps in the same period band (Lin and Ritzwoller, 2011), but are derived directly from a single earthquake, giving us good confidence that the reconstruction algorithm is accurately capturing the details of the wavefield.

3.5 Helmholtz-Hodge Decomposition of the Horizontal Wavefield

The horizontal vector wavefield $\mathbf{u}_h$ is by definition tangent to the Earth's surface, and may therefore be represented by two scalar potential functions D and S and a harmonic vector function $\mathbf{r}$ by the Helmholtz-Hodge Decomposition (Bhatia et al., 2013)

$$\mathbf{u}_h = \nabla D + \hat{z} \times \nabla S + \mathbf{r} \tag{3.8}$$

$\mathbf{r}$ being harmonic implies that $\nabla^2 \mathbf{r} = 0$. The irrotational potential function D creates a curl-free displacement field and consequently generates the horizontal projection of the $P-SV$ system in a laterally homogeneous medium, while the solenoidal potential function S creates a divergence-free displacement field and corresponds to the SH system in laterally homogeneous medium. As such, applying the Helmholtz-Hodge decomposition to the horizontal wavefield may allow improved discrimination of wave types in sufficiently smoothly varying media, even in the presence of significant off-great-circle and multipathing wave propagation. In particular, for wavefields comprised solely of surface waves, the Helmholtz-Hodge decomposition allows for discrimination between Rayleigh and Love wave components using potentials

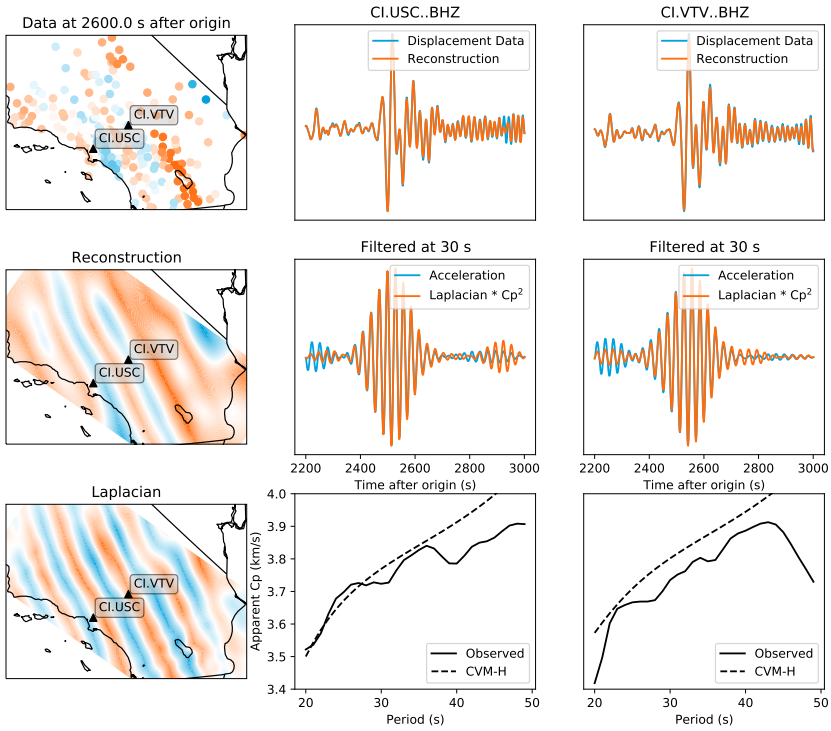


Figure 3.3: Wavefield gradiometry for the Southern California Seismic network applied to the Rayleigh wave packet of the November 19, 2017 Mw 7.0 Loyalty Islands earthquake. We show the BHZ channel integrated to displacement and its reconstruction for one basin station (USC) and one high desert station (VTV), filtered between 10–100s. We also show a comparison between the measured accelerations at 30s and the spatial Laplacian multiplied by a single coefficient, interpreted to be the squared phase velocity at the channel site—showing that the Rayleigh wave packet closely follows the acoustic wave equation at this period. Finally, we show the estimated phase velocity curves and their comparisons to the theoretical phase velocities derived from the CVM-H model; the agreement between the observed data and the theoretical model is quite good, and has been achieved from a single wave packet from a single earthquake.

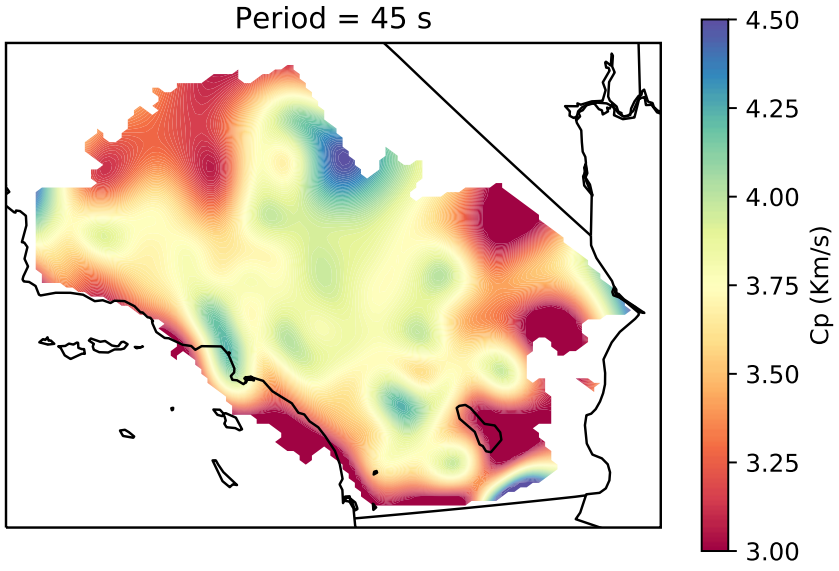


Figure 3.4: Phase velocity map of Southern California at a period of 45 s for points within the convex hull of nearby SCSN stations (<90 km distance), recovered using Laplacian based wavefield gradiometry of a single event, the November 19, 2017 Mw 7.0 Loyalty Islands earthquake. The tomographic model is smoothed using a multiquadratic radial basis function with a length scale of $45km$ to suppress artifacts at lengthscales shorter than $1/4$ of the characteristic wavelength.

that satisfy an acoustic type wave equation for phase velocity. The decomposed wavefield may therefore be used for surface wave tomography directly via wavefield gradiometry (de Ridder and Biondi, 2015), or using a more robust gradient based method such as eikonal or Helmholtz tomography (Lin and Ritzwoller, 2011; Lin et al., 2009).

Using the framework developed in Section 3.3, we can estimate the curvelet coefficients of the potentials D and S from the wavelet transforms of the horizontal components by solving

$$\begin{bmatrix} \frac{\partial \mathbf{G}}{\partial x} & \frac{\partial \mathbf{G}}{\partial y} \\ \frac{\partial \mathbf{G}}{\partial y} & -\frac{\partial \mathbf{G}}{\partial x} \end{bmatrix} \begin{bmatrix} \mathbf{c}_D(j_w, s) \\ \mathbf{c}_S(j_w, s) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_x(j_w, s) \\ \mathbf{w}_y(j_w, s) \end{bmatrix} \tag{3.9}$$

with $L1$ regularization. The harmonic component $\mathbf{r}$ cannot be uniquely determined

without appropriate boundary conditions on a space that are typically not available for finite seismic observations on a particular area. However it may be generated by augmenting either D or S, and is consequently absorbed into the potentials by $L1$ regularization.

As an example of utilizing wavelet-curvelet compressive sensing for performing the Helmholtz-Hodge decomposition, we visualize the wavefield from the Mw 7.1 July 6 2019 Ridgecrest Earthquake, recorded at the Los Angeles United School District subset of the Community Seismic Network (LAUSD-CSN). The LAUSD-CSN is a low-cost, high-density permanent urban seismic deployment utilizing micro-electromechanical system (MEMS) accelerometers, straddling the North-eastern edge of the Los Angeles basin (Clayton et al., 2012; Kohler et al., 2020). The network is optimized for cheap, spatially resolved near real-time strong-ground-motion reporting from earthquakes within the Los Angeles metro area and consequently uses instruments with a high instrument noise floor. However the ground motions from the regional Mw 7.1 Ridgecrest event were sufficiently strong to produce waveforms that are a near match to the acceleration seismograms of co-located permanent strong-ground motion instruments (HN channel, example comparison for station CI.PASC in Supplementary Figure 3.7). The location of the LAUSD-CSN relative to downtown Los Angeles, and the curvelet inversion domain utilized in the Helmholtz-Hodge decomposition are shown in Figure 3.5 A).

The inversion utilized acceleration waveforms bandpass filtered between 2–15 s, a 64×64 pixel spatial curvelet transform, and the db12 wavelet for the time domain transform. The compressive sensing optimization problem was solved with Elastic-Net regression with $\alpha = 0.95$. The LAUSD-CSN team discovered that a number of sensors were misoriented during the Ridgecrest event; these sensor problems have confirmed by site visit and corrected. Additional potentially misoriented sensors have been detected by analysing apparent spatial discontinuities in long-period

particle motion. In that case, rotations in 90° increments have been manually applied to bring the long period motions into alignment; in these cases the misorientation is likely a result of placement of the instrument against the incorrect wall of an interior room. An example time-domain reconstruction is shown in Figure 3.5 B), showing the ability of the Helmholtz-Hodge decomposition to extract the large Love wave component into the solenoidal term. A still frame from the resulting Helmholtz-Hodge decomposition, 75 s after the event origin time, is shown in Figure 3.6 (the full video is available online in the Supplement). The frame shown in 3.6 is during the surface-wave packet. The contributions of the Rayleigh and Love waves can be seen in the irrotational and solenoidal components of the field, respectively. The irrotational component in particular shows the strong bending of Rayleigh waves as they cross into the deep Los Angeles basin from the North-East.

3.6 Wavefield Compression

A final, simple, application of the wavefield reconstruction technique detailed here is as a lossy data compression algorithm. Whilst the wavefield reconstruction algorithm proposed here is not optimized for data compression, the use of a sparsity promoting domain transform naturally induces compression when the wavefield is sufficiently coherent. In general seismic data (especially in research settings) is able to be stored as high fidelity continuous timeseries, however there are at least two significant end-member cases for which compression techniques are potentially applicable. The first is in extremely dense industry-scale surveys, in which the total data volume becomes prohibitively difficult to handle. The tension between periods of high data acquisition and continually advancing computational storage and processing capabilities, has lead to cycles of interest in seismic data compression as a means to reduce the computational burden (e.g. Da Silva et al. (2019); Herrmann et al. (2008); Villasenor et al. (1996)). The majority of these studies have focused

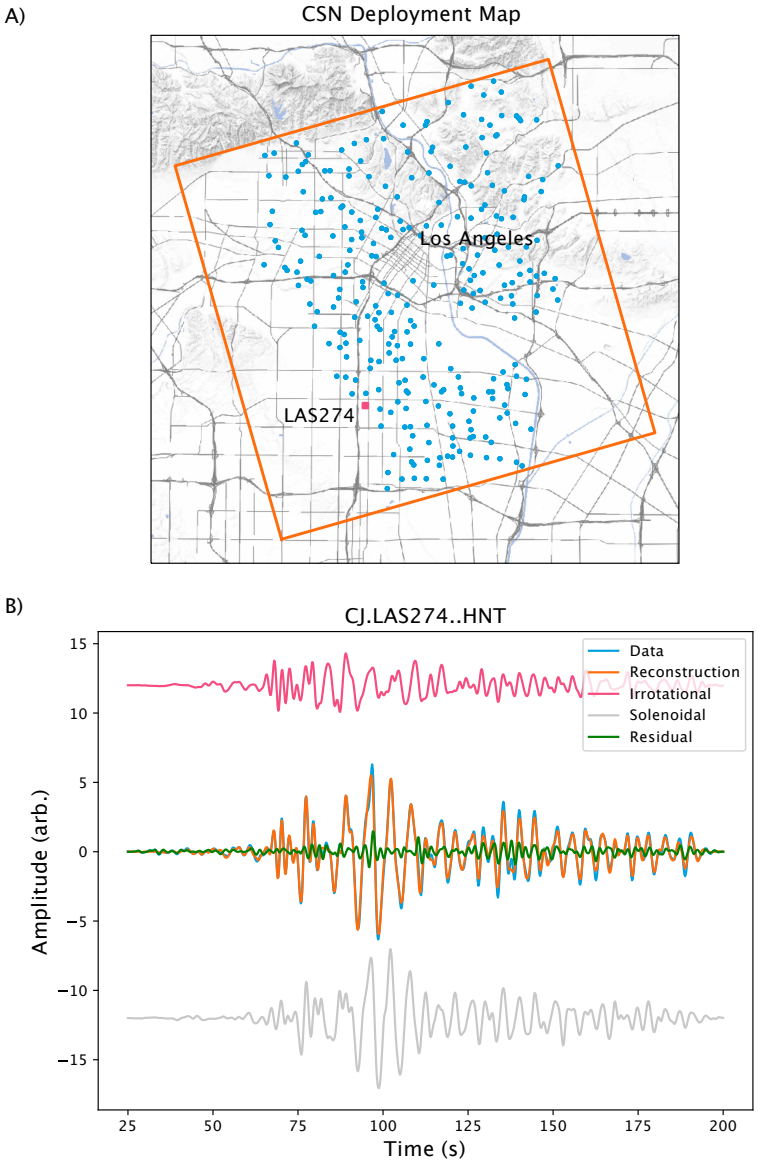


Figure 3.5: A) Los Angeles Unified School District - Community Seismic Network (LAUSD-CSN) accelerometers used in this study are shown in blue, with station LAS274 shown by a pink square. The inversion domain is shown by the orange square. The Los Angeles downtown is the high density area of roads in the center of the figure. B) Data, reconstruction and residual for the tangential component of the Ridgecrest July 5, 2019 Mw 7.1 Earthquake recorded at LAUSD-CSN station CJ.LAS274. The irrotational and solenoidal components are individually shown offset from the main waveform, and show the strong solenoidal Love wave arriving after the SV/SH arrival observed on both components. Waveform is bandpass filtered between 0.5–15 s

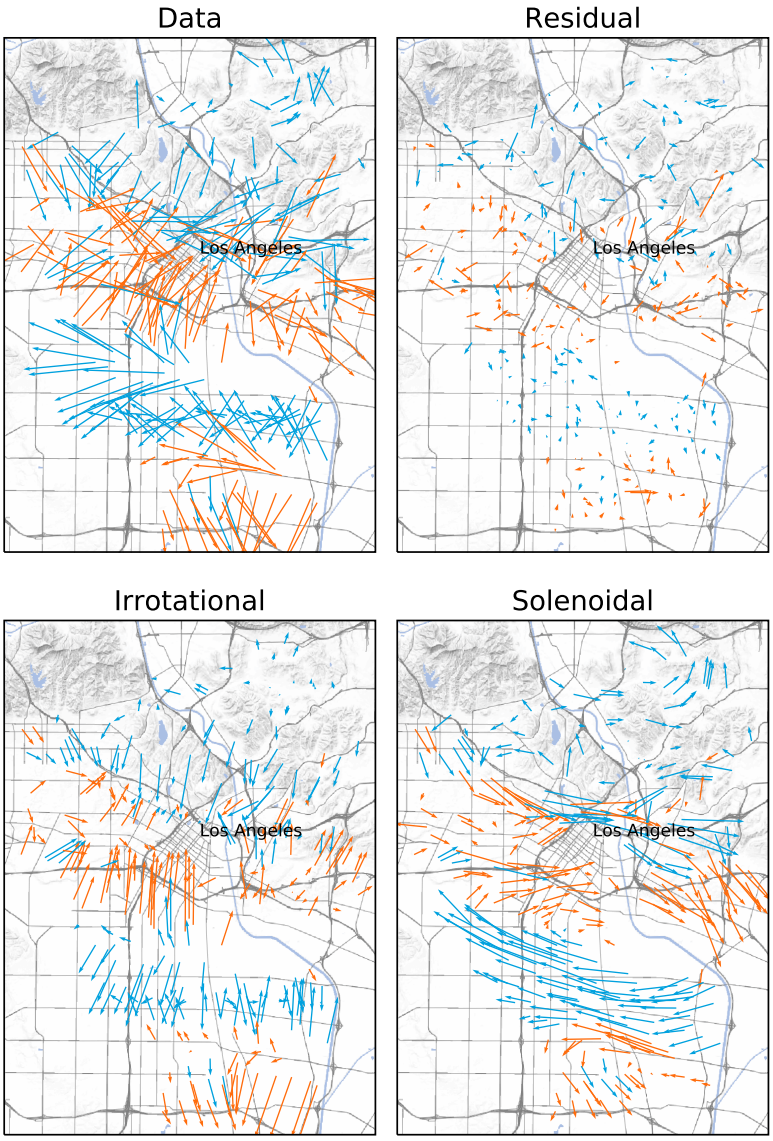


Figure 3.6: Helmholtz-Hodge decomposition of the horizontal wavefield of the Mw 7.1 July 6 2019 Ridgecrest event recorded on the LAUSD-CSN network. The wavefield is plotted 75 s after the event origin time. Arrows show the horizontal particle instantaneous acceleration for both data and reconstruction. Arrows are colored with the sign of the real data vertical component to highlight the oscillatory structure of the wavefield.

on structured seismic data volumes, potentially with some missing elements, and so have achieved high compression ratios whilst maintaining fidelity; a peculiarity of the current study is that our algorithm is targeted at unstructured data volumes, necessitating the mixed-type wavelet-curvelet decomposition described above. The second end-member is that of slow data transmission-rate scenarios, such as those that will be encountered by proposed planetary seismology arrays (e.g. Marsal et al. (2002); Neal et al. (2019); Zhan and Wu (2019)). This scenario motivated initial studies into seismic data compression, such as that of Wood (1974), where the rate of data acquisition outpaced that of early telephone based remote communication protocols. In this scenario, onsite preprocessing before transmission may allow more detailed (either more channels or higher sampling frequency) data to be transmitted than would otherwise be the case. We briefly discuss the ability of the algorithm proposed in this study to compress the wavefield within the context of a local strong-ground motion accelerometer array (CSN).

For CSN dataset resampled at 2 Hz and filtered using a zero-phase bandpass filter between 2–50 s, applying the proposed wavelet-curvelet reconstruction method using a db4 wavelet for DWT achieves a compression ratio of 8.3. The recovery of the original signal can be quantified by the a scaled mean-squared-error fidelity metric $1 - \sqrt{\frac{1}{N_s}\sum_{j=1}^{N_s}\frac{1}{N_d}\sum_{i=1}^{N_d}(\mathbf{d}_{ij} - \hat{\mathbf{d}}_{ij})^2 / Var(\mathbf{d})}$, which gives a value of 0.71 using the for the CSN dataset. The residuals from the above case studies are largely incoherent with the recorded wavefield, suggesting that the reconstruction is acting as a denoising filter (noting that "noise" in this context includes scattered energy that is not able to be sparsely represented using the curvelet frame). We do not attempt to optimize the transform parameters for compression, and it is likely that further investigation would yield improved compression ratios for equivalent fidelity. In particular, for the purposes of compression, it may not be necessary to accurately obtain accurate spatial derivatives, so changing the form of the preconditioning

matrix may promote sparser transforms.

In our scheme, compression occurs entirely during the sparsity promoting Elastic Net regularized curvelet inversion—all wavelet coefficients are inverted for. It is possible that higher compression ratios for equivalent reconstruction accuracy could be obtained by further sparsifying the temporal component in the wavelet domain. Temporal sparsification would, however, have to be designed to be consistent with the spatial patterns of the waveforms, which may be difficult due to the propagation of the wavefield. As such, further studies into compression schemes based on the joint wavelet-curvelet transform are left for future research.

3.7 Discussion and Conclusions

The three real-data examples shown in this work illustrate the required modifications to the framework presented in Zhan et al. (2018) to robustly handle sparse, noisy, realistic wavefield recordings. We have developed the theory in this paper for point sensor recordings, however with minor modifications it can equally be employed for DAS or mixed DAS and point sensor deployments. Because the algorithm presented here relies on time-scale (wavelet) or position-angle-scale (curvelet) transforms, there is significant scope for tailoring the algorithm toward specific datasets, that we have not fully explored. In particular, in the time domain, we have found that the Daubechies family typically works well, but that is not to say that other wavelet families may not be equally or even more performant. In the spatial domain, curvelets are well suited for seismic applications (Herrmann and Hennenfent, 2008; Herrmann et al., 2008), but other transforms such as wave atoms may work well (Leinonen et al., 2013). For large apeture deployments such as the USArray, where the Earth's curvature becomes important, employing natively spherical transforms such as spherical curvelets (Chan et al., 2017) may be necessary. Future work to address, in particular, the best combination of both transforms to promote sparsity

of the spatial transform and hence good compressive sensing performance will improve the accuracy of the reconstruction. Given this flexibility, the key intellectual contribution of this paper is then to propose independent reconstructions in time and space to handle the sparse, unstructured datasets present in research deployments, and to recognize the importance of appropriately reweighting the spatial reconstruction frame to promote continuity of higher order derivatives, if the wave equation is to be satisfied.

Looking forward to future methodology for wavefield reconstructions of unstructured seismic datasets, recent machine learning (ML) applications of physics-informed neural networks (PINNs) (Raissi et al., 2019) promise to allow for grid-free reproduction of seismic wavefields (Karimpouli and Tahmasebi, 2020; Moseley et al., 2020; Song et al., 2020). Current research has focused on training for wavefield solvers based on known synthetic velocity models, however there is potential for the same computational structure to be used in a joint inversion of unknown velocity structure with observed data, under reasonable assumptions that some seismic wave equation is satisfied; given observational constraints the first likely route would be for surface wave reconstruction/phase velocity inversion as is presented in this study. PINN based studies originating from within the geophysics community have utilized computational architectures originally proposed for general discovery of PDE behaviour, i.e they have not been specifically optimized for seismic applications. In particular, these studies have made use of non-periodic activation functions, which fail to exploit the inherent quasi-periodicity of the seismic wavefield. Recent work on neural networks utilizing periodic activation functions (Sitzmann et al., 2020) and/or analysis in the Fourier domain (Li et al., 2020b) may therefore improve the potential of ML based methods for wavefield reconstruction. In particular, the work of Sitzmann et al. (2020), utilizing sinusoidal activation functions in a network they term SIRENS, has proven to be able to very accurately reconstruct the higher order

derivatives of seismic wavefields necessary for computing terms in the seismic wave equation. This is due to the property that the derivative of a SIREN is also a SIREN, which allows for analytical computation of non-vanishing second order derivatives with inherent periodicity. Thus far, the training of PINNs has proven to be relatively computationally expensive and work has been largely limited to synthetic examples, however the field is still in its nascent stage and currently appears to hold great promise for the large class of PDE constrained inverse problems, in which wavefield reconstruction may be included.

We have presented a general strategy for wavefield reconstruction of unstructured seismic data using wavelet/curvelet transforms, as well as specific implementation details for three example applications. This framework allows for the conversion of point seismic time series into a single unified data product, suitable for spatial analyses of the seismic wavefield such as wavefield gradiometry, back-projection, reverse-time migration, etc. Our choice of the curvelet basis for spatial analysis allows for the recovery of complex wavefields including multi-pathing or back-scattering effects. Our framework promotes a physically concordant wavefield by appropriately penalizing short wavelength fluctuations without requiring *a priori* knowledge of the underlying velocity structure or requiring iterative solutions of the wavefield reconstruction problem. It therefore represents a simple and flexible way to compress and represent the seismic wavefield, suitable for irregularly sampled networks at all scales.

3.8 Acknowledgements

The authors would like to thank the editor, Andrew Valentine, and an anonymous reviewer, for their insightful reviews that clarified several points. They would also like to thank Charles Langston for a particularly enthusiastic review. JBM acknowledges the financial support of the Origin Energy Foundation and the General

Sir John Monash Foundation during his PhD studies. ZZ thanks the support from NSF CAREER award EAR 1848166 and the NSF/IUCRC GMG Center. Seismic data were processed using Obspy (Beyreuther et al., 2010), and figures were created using Matplotlib (Hunter, 2007) using Cartopy for mapping (Met Office, 2010).

3.9 Data Availability

Data for the Mw 7.0 Loyalty Islands Earthquake was obtained through the obspy FDSN service using the Southern California Earthquake Data Center as a provider. CSN data for the Mw 7.1 Ridgecrest Earthquake is available through the CSN website at `http://csn.caltech.edu/data/`. Code and data to recreate the Helmholtz-Hodge decomposition is available at `https://github.com/jbmuir/HelmholtzHodgeCSN`.

3.10 Supplement

The supplement also includes a movie of the Helmholtz-Hodge decomposition for the CSN, which may be accessed via the online version of this paper.

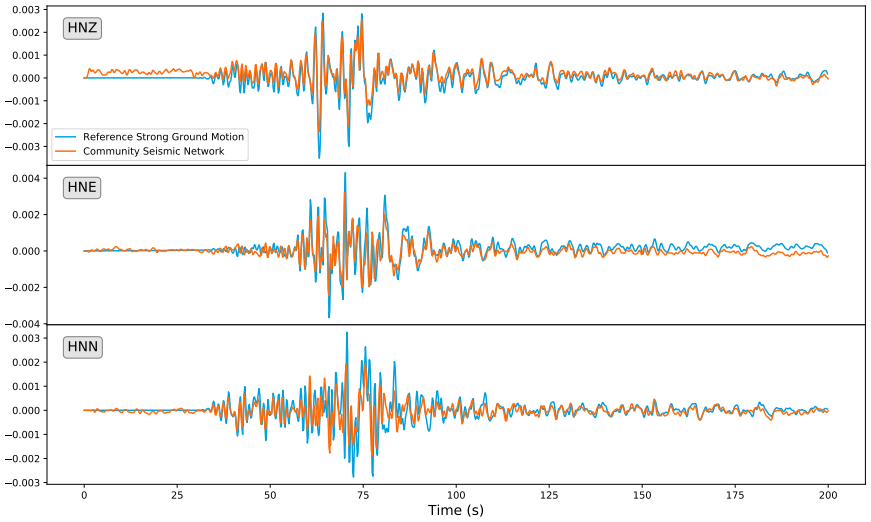


Figure 3.7: Comparison between Southern California Seismic Network strong ground motion station CI.PASC.00.HN* and Community Seismic Network station CJ.T000337..HN* for the Mw7.1 July 5 2019 Ridgecrest earthquake. These stations are approximately co-located. The waveforms have been decimated to 5 Hz, detrended and lowpass filtered at 1 Hz. Amplitudes have been scaled to approximately account for instrument gain.

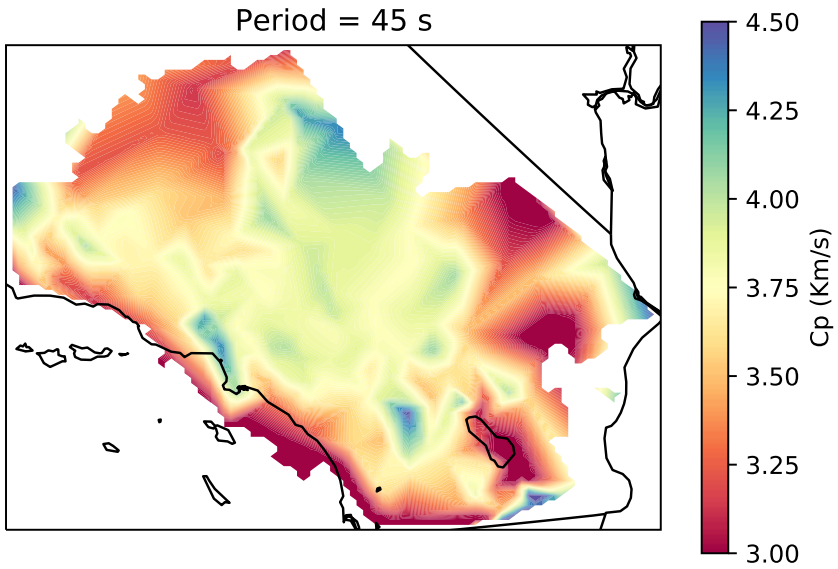


Figure 3.8: Rayleigh fundamental mode phase velocities derived from wavefield gradiometry for Southern California using the November 19, 2017 Mw 7.0 Loyalty Islands earthquake. This version uses simple 2D linear interpolation as opposed to radial basis function smoothing, as shown in Figure 3.4.

Chapter 4

WAVEFIELD-BASED EVALUATION OF DAS INSTRUMENT RESPONSE AND ARRAY DESIGN

To appear as

Muir, J. B. & Z. Zhan "Wavefield-based evaluation of DAS instrument response and array design". In: *Geophysical Journal International*, Submitted

4.1 Abstract

Distributed Acoustic Sensing (DAS) networks promise to revolutionize observational seismology by providing cost-effective, highly dense spatial sampling of the seismic wavefield, especially by utilizing pre-deployed telecomm fiber in urban settings for which dense seismic network deployments are difficult to construct. However, each DAS channel is sensitive only to one projection of the horizontal strain tensor and therefore gives an incomplete picture of the horizontal seismic wavefield, limiting our ability to make a holistic analysis of instrument response. This analysis has therefore been largely restricted to pointwise comparisons where a fortuitious coincidence of reference three-component seismometers and co-located DAS cable allows. We evaluate DAS instrument response by comparing DAS measurements from the PoroTomo experiment with strain-rate wavefield reconstructed from the nodal seismic array deployed in the same experiment, allowing us to treat the entire DAS array in a systematic fashion irrespective of cable geometry relative to the location of nodes. We found that, while the phase differences are in general small, the amplitude differences between predicted and observed DAS strain-rates average a factor of 2 across the array and correlate with near-surface geology, suggesting that careful assessment of DAS deployments is essential for applications

that require reliable assessments of amplitude. We further discuss strategies for empirical gain corrections and optimal placement of point sensor deployments to generate the best combined sensitivity with an already deployed DAS cable, from a wavefield reconstruction perspective.

4.2 Introduction

Distributed Acoustic Sensing (DAS) networks utilize time-of-flight interferometry of Rayleigh backscatter from natural imperfections within fiber optic cables to obtain spatially resolved strain measurements. Current technological standards allow a single DAS interrogator unit to produce meters spaced channels along cables of length ~20-100km; in other words, creating a distributed strain network of many thousands of sensors. While most original application of DAS was in industry, where the spatial resolution and high environmental tolerance of fiber-optic cable makes DAS eminently suitable for borehole deployment, the falling costs of the interrogator units and the increasing utilization of "dark fiber" (already laid inactive telecom fiber) have made surface deployments an attractive proposition for fundamental research purposes. Recent studies have shown that onshore DAS can detect both teleseismic (Lindsey et al., 2017; Yu et al., 2019) and local (Karrenbach et al., 2020; Wang et al., 2018; Zhan, 2020) earthquakes with waveforms that match colocated seismometers. Several studies utilizing offshore cables have also reported success in observations of both earthquakes and ambient environmental seismology (Ide et al., 2021; Lindsey et al., 2019; Matsumoto et al., 2021; Sladen et al., 2019; Spica et al., 2020; Williams et al., 2019). DAS also promises to be one of the key data sources in the nascent field of social seismology, with recent success in monitoring parade traffic in an urban setting (Wang et al., 2020) and real time monitoring of rail-traffic (e.g. Ferguson et al. (2020); Wiesmeyr et al. (2020)). Mixed networks of DAS and point sensors have also proven to be a successful combination for fine

scale monitoring of seismic properties associated with hydrothermal fields (Feigl, 2017; Feigl and Parker, 2019), and hold promise in integration into traditional ambient-noise tomography workflows (Nayak et al., 2021).

While DAS presents great opportunities for very dense deployments, it also comes with several notable challenges. The signal-to-noise ratio of DAS data is typically poorer than that of conventional seismometers, and the instrument response in realistic deployments is also not fully understood, leading to a generally less informative single-channel observations. DAS measures the integrated strain-rate or strain in a particular direction over a finite gauge length, meaning that individual stations or straight lengths of cable do not have access to the full horizontal particle motion as would be obtained from a 3-component point sensor. Finally, the huge volumes of data produced by DAS are highly redundant except at very high frequency, and storing and analysing them poses a great computational challenge. Many of these challenges can be overcome, or better understood, by treating the DAS sensor array as a means of accessing a single underlying wavefield, rather than as a collection of individual channels—that is, seeking a representation of the data as $u(x, y, t)$ rather than $u_i(t)$ where x and y are spatial components and i is a channel number. Such a representation allows us to study in detail the spatial response of the array, including gradient terms, which are essential for strain based measurements like DAS. The wavefield representation offers a simple and coherent mechanism for converting between strains and displacements, which makes it especially well suited for studying the lateral variations of DAS amplitude response. Characterisation of this response will be essential for realizing the promise of DAS as a spatially dense sensing modality for strong ground motions and earthquake early warning (Karrenbach et al., 2020).

In the preceding paper (Muir and Zhan, 2021), we developed a framework for combining an irregular network of spatially distributed sensors into a single unified data

product by using wavefield reconstruction. In this study, we apply our compressive framework to the combined DAS and nodal array deployed during the PoroTomo Experiment at Brady, NV (Feigl, 2016a, 2017; Feigl and Parker, 2019). Nodal arrays are constructed from self-contained cable-free instruments that sense the conventional wavefield (i.e. time derivatives of the particle displacement). Nodal arrays have proven highly successful in temporary dense deployments (i.e. Jia and Clayton (2021); Lin et al. (2013)). Nodes strongly complement DAS arrays in the burgeoning earthquake rapid-response space by providing extra spatial coverage to supplement dark-fibre DAS deployments. We perform a comparison between the strain-rates recorded on the DAS array and those predicted by reconstructing the velocity wavefield recorded on the nodal array. While this analysis has been performed for individual DAS segments using co-located seismometers (e.g. Lindsey et al. (2020); van den Ende and Ampuero (2021); Wang et al. (2018)), our framework allows us to evaluate the DAS array using the entire recorded nodal data volume simultaenously, allowing for non-optimally aligned segments, without favorable reference seismometers, to be studied. The wavefield approach also sidesteps the need to apply fk rescaling (or similar methods that rely on estimating a reference phase velocity which become unstable for small wavenumbers k (Lindsey et al., 2020)) when converting from strain-rate to velocity. This aids in the simplicity of application and the robustness of the results. We further develop methods for simultaenously reconstructing DAS and nodal data, correcting DAS data using the observed nodal field, and optimal design strategies for mixed arrays.

4.3 Inversion of DAS records for particle velocity

DAS measurements are performed by observing the change in back-scattering characteristics from laser pulses sent into a fiber-optic cable. Measurements ideally correspond to the strain or strain-rate averaged along a gauge-length at some point

in the fiber. The strain ϵ_c along a cable is given by the projection of the horizontal strain tensor along the cable azimuth α

$$\epsilon_c = \begin{bmatrix} \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} \\ \epsilon_{xy} & \epsilon_{yy} \end{bmatrix} \begin{bmatrix} \sin\alpha \\ \cos\alpha \end{bmatrix} = \sin^2\alpha\epsilon_{xx} + \sin 2\alpha\epsilon_{xy} + \cos^2\alpha\epsilon_{yy} \quad (4.1)$$

with the strain components given as usual by $\epsilon_{ij} = \frac{u_{i,j}+u_{j,i}}{2}$. The theoretical DAS strain-rate response ϵ_{DAS} is in turn given given by averaging the strain or strain-rate (Bakku, 2015; Wang et al., 2018) over a gauge length L

$$\epsilon_{DAS} = \frac{1}{L}\int_{-L/2}^{L/2} \epsilon_c(l)dl, \quad (4.2)$$

where l indexes the length along the gauge length. Typical gauge lengths are on the order of ∼10m. Work to characterise the phase and amplitude response of DAS cables is rapidly progressing. If a DAS array is deployed such that it has sensitivity to both components of the horizontal wavefield, the general framework of Muir and Zhan (2021) may be employed to cast the recovery of the horizontal wavefield from DAS records as an inverse problem—we develop the necessary inverse problem machinery in Section 4.9. Qualitatively, the framework of Muir and Zhan (2021) creates a sparse representation of the wavefield using a reweighted curvelet basis (Candes and Demanet, 2005), scaled to promote continuity of the wavefield Laplacian. This rescaling, coupled with the use of a curvelet representation which is natively well-suited towards representing wave propagation, promotes interpolations that satisfy the seismic wave equation and result in smooth derivatives that can be used to calculate the predicted DAS strain-rate based on the observed velocity field.

While most extant DAS deployments are typically linear or loop-shaped, the Poro-Tomo experiment deployed at the Brady, Nevada thermal field in 2016 has a suitable space-filling zig-zag deployment geometry for wavefield reconstruction using DAS

(Feigl, 2017). This suitability has also separately lead to analysis of the beamforming capabilities of the PoroTomo DAS array (van den Ende and Ampuero, 2021), where the authors discovered issues with coherence within the array inhibited beamforming performance. The lack of coherence was interpreted to be variously due to the cable geometry, coupling, and the effects of small-scale scattering. By integrating DAS strain records, the beamforming issues highlighted by (van den Ende and Ampuero, 2021) were largely ameliorated, with these results highlighting the potential utility of wavefield reconstruction for improving the spatial continuity of the seismic waveform, thereby making it more suitable for spatial techniques such as beamforming.

The seismic component of the PoroTomo experiment ran for 15 days in March 2016, with the instrumentation consisting of 8700m of fiber optic cable filling a rectangle with approximately 4:1 aspect ratio, and additionally instrumented with 238 three-component seismic nodes. The distribution of nodes and the geometry of the DAS cable are shown in Figure 4.1. An additional 400m of cable was deployed in a borehole, which we do not utilize in our wavefield reconstruction. Wang et al. (2018) showed that by appropriately differencing co-deployed nodal seismic instruments and comparing them to the average strain-rate along cable segments, quantitatively similar strain-rate waveforms could be observed for a ML 4.3 that occurred March 21, 2016 approximately 150km SSE of the PoroTomo experiment. The methodology proposed in this study allows us to perform similar quantitative analysis on the entire DAS array, using a robust wavefield recorded on the entire nodal array, as well as the reverse experiment, to invert for the velocity field perturbations given the strain-rates.

The PoroTomo team has provided data for the Fairfield ZLand 5 Hz three component nodal seismometers, corrected for instrument gain to give units of coil-case velocity in μm s^{-1} (Feigl, 2017). Ringler et al. (2018) found that a collection of similar

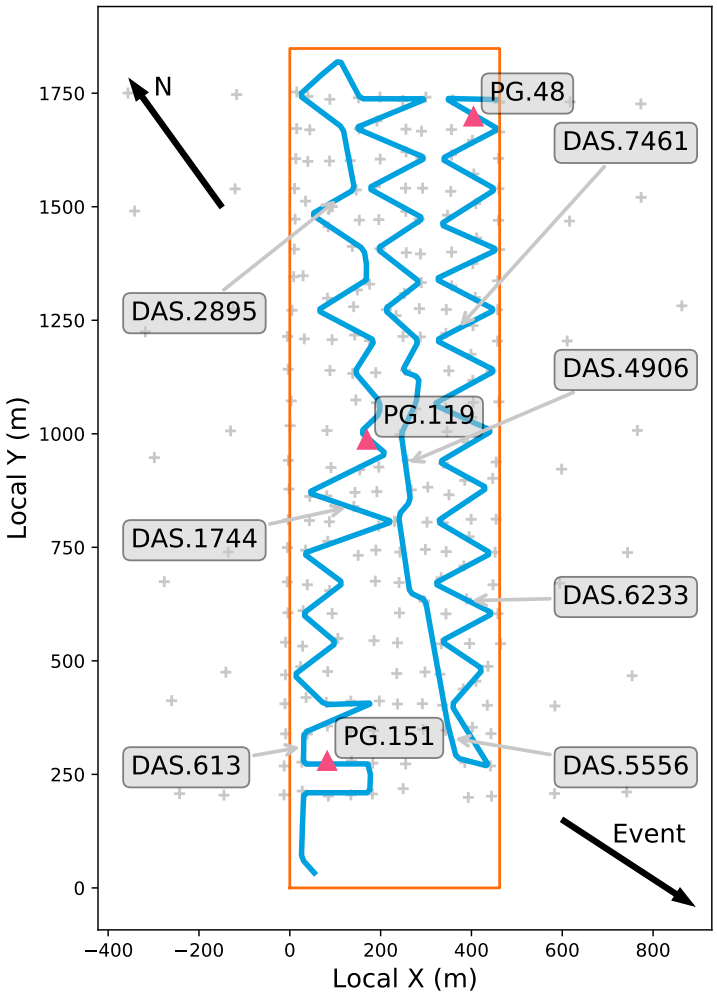


Figure 4.1: Deployment of DAS cable (blue lines) and nodal seismic stations (grey crosses) during the PoroTomo experiment at Brady, Nevada. The orange box shows the reconstruction domain, and pink stations are used for comparison in Figure 4.9

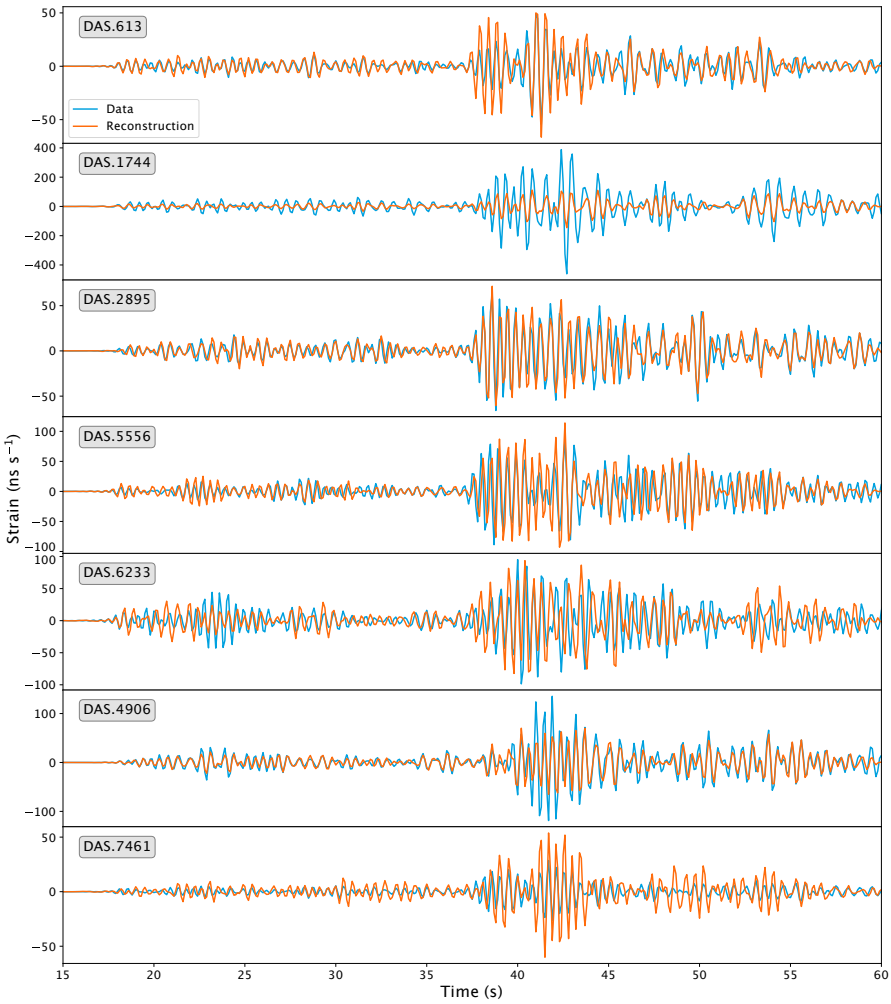


Figure 4.2: Comparisons between DAS channels (blue) and reconstructed strain-rate from nodes (orange) at the seven stations highlighted in Figure 4.1. Waveforms are bandpass filtered at 1–2.5 Hz, and nodal data has been corrected using a nominal Fairfield ZLand 3C instrument response.

instruments can be well characterised by assuming a damped oscillator instrument response, with a mean free oscillation frequency of 4.87 Hz and a damping factor of 0.98 (averaged across all three channels for a collection of three instruments). We use these parameters to correct the nodal data to give ground velocity. DAS records were cleaned by removing channels for which the DAS gauge length included a corner (10 m for the PoroTomo deployment), and by removing traces below the 1^{st} or above the 99^{th} percentile of maximum log amplitude.

To begin with, we first emulate Wang et al. (2018)'s experiment using the compressive sensing framework developed in this study. Utilizing all nodes within the orange study area shown in Figure 4.1, we inverted for a 32×128 pixel curvelet decomposition of the discrete wavelet transform (DWT) of the nodal ground velocity records. We then projected these results onto the DAS cable, and inverse transformed to give the predicted strain-rate. Both nodes and DAS were bandpass decimated to a 10 Hz sampling rate, and a 1–2.5 Hz bandpass filter was applied. 2.5 Hz was chosen as an upper limit based on experimentation, which found it to be the highest frequency able to be reasonably reconstructed using DAS only data as seen in Section 4.10. The Daubechies db12 wavelet with 5 levels was used for the time domain transform (the optimal wavelet was determined by experimentation), and the compressive sensing optimization used pure L1 regularization. The PoroTomo deployment is known to suffer from clock timing mismatches between the nodes and DAS (Wang et al., 2018). We estimated a best-fit uniform clock correction by minimizing the least squares misfit between the reconstructed DAS strain-rate data and the true DAS strain-rate records as a function of lag time. This resulted in a best relative time shift of 17 samples, or 1.7 s, with the DAS clock being faster. We plot the results (including time shift) for seven cable locations (labelled in Figure 4.1) in Figure 4.2. We see that the recovered strain-rate typically match the phase and relative amplitudes of the DAS records well. Absolute amplitudes are normally

well recovered for the P wave. The machinery of Muir and Zhan (2021) can also be used to perform the reverse operation of predicting nodal data using the observed DAS strain-rate without recourse to a plane-wave assumption and the consequent difficulties in obtaining a representative phase velocity, and also to jointly invert DAS and nodal data into a unified data product. The results of these experiments are shown in Appendices 4.10 and 4.11 respectively—notably the framework of wavefield reconstruction allows for a conversion of DAS strain-rates to particle velocity without relying on an inferred horizontal phase velocity that may depend on local constitutive properties and that becomes difficult to stabilize for near-vertical incidence ground motions or where cable segments are insufficiently short for adequate k-space (wavenumber) resolution.

In Section 4.11, we perform a five-fold cross validation experiment to compare the reconstruction performance of a mixed network of DAS and nodes to the node only network. The low prediction error in the nodes-only cross-validation experiment shown in Figure 4.11 illustrates that the nodal sensors at Brady are essentially able to fully capture the major details of ML 4.3 Hawthorn NV wavefield within the 1-2.5 Hz frequency band, an argument that has also been made on the basis of beamforming coherence by van den Ende and Ampuero (2021). The average root-mean-square-error when using 80% of the nodes as a training set to predict the remaining 20% left-out test data is 0.13. This suggests that we should be able to predict the DAS strain-rate data from the reconstructed velocity wavefield derived from the nodal sensors with a high degree of accuracy. While the results of Figure 4.2 indicate that we can capture many features of the DAS data, there still remain substantial discrepancies to be explained. Sources of these discrepancies must relate either to incorrect metadata (imprecisions in the locations of the DAS channels, for example), unexplained errors in the ability of the DAS data to accurately record the true wavefield (i.e. instrument-ground coupling), or theory errors in the

prediction of strain from the observed velocity. Given that strain measurements are acutely sensitive to very small-scale structure, the distinction between true instrument response, coupling between the instrument and the ground, and very local path effects are less distinct for DAS (and other strain-sensing modalities) than they are for point seismometers acting at typical 1 Hz frequencies (e.g., Capdeville et al. (2020); King et al. (1976); Ringler et al. (2019)). Attempting to ascribe the apparent response of the cable to any one of these factors using only the predicted ground motions is difficult, however by studying the characteristics of response across the whole array it may be possible to build a hypothesis as to the predominant factors by searching for a correlation (or lack of correlation) with other datasets, such as tomographic models of the subsurface.

We now look at the characteristics of errors in the predicted wavefield in more detail. While the PoroTomo experiment unfortunately recorded only one earthquake source suitable for wavefield reconstruction, which precludes analysis of the DAS transfer-function in detail as was performed for the FOSSA experiment by Lindsey et al. (2020), the areal coverage characteristics of the PoroTomo experiment offer a unique opportunity to investigate the spatial behavior of the gross-scale metrics of instrument response. We computed zero-lag normalized cross correlations between observed DAS strain-rates and those predicted from the reconstructed nodal field as a metric of waveform shape matching or phase fidelity. We used the default settings for normalized cross-correlation in the ObsPy library (Beyreuther et al., 2010) (i.e. demeaning and normalization of the signals in running windows). Averaged over the entire DAS array, we saw a median normalized zero-lag cross correlation of 0.69, with a histogram of cross-correlation values plotted in Figure 4.3, showing that the majority of the array is clustered at high cross-correlation values, indicating a good overall phase fit, with a long tail of poor cross-correlations, including some with reversed polarity. In order to assess the match in amplitudes, we took the log of the

waveform envelopes, obtained using the Hilbert transform method, and compared their normalized root-mean-square-error (RMSE, normalized by division by the interquartile range of the observed log-envelope to provide a robust assessment of the scale of improvement irrespective of the original scale of amplitudes). The log-envelope metric gives a holistic view of the amplitude fit throughout the waveform, including low and high amplitude sections, with a RMSE $\sim$ 0 under this metric indicating that the amplitude scale is matched throughout the waveform and an RMSE $\sim$ 1 indicating amplitude errors on the scale of the interquartile range of the observations; the median RMSE for this amplitude metric was was 0.61, and the histogram of RMSE values across the array is also plotted in Figure 4.3.

To look at the amplitude fits in greater detail, we performed a time-frequency analysis by calculating the continuous wavelet transform (CWT) of the observed and predicted DAS strain-rate signals using a Morelet wavelet, and then used these to calculate the RMS amplitude of both during the P-wave (16–21 s) and S-wave (36–41 s), for three frequency bands (1.0-1.5 Hz, 1.5–2.0 Hz, 2.0–2.5 Hz, and also the full 1.0–2.5 Hz), with the results plotted in terms of log-amplitude ratios in Figure 4.4. We observe spatially coherent patterns of mispredicted amplitudes across both P and S phases and across frequency bands. The most obvious features are the two red areas of overpredicted amplitude in the lower left and upper right of the array. These areas are coincident with quaternary alluvial surface geology and particularly low near-surface V_P, as reported by (Parker et al., 2018), which together with the large spatial scale of the amplitude response features and the fact that the PoroTomo cable was buried directly in a fresh trench (Feigl, 2017) suggests that the geological conditions encountered by the DAS cable are substantively responsible for the differences between observed and predicted waveforms. In particular there are significant areas in which the observed amplitude of the DAS waveforms is substantially less than what one would expect based on the data recorded on the

nodes. This implies that future studies that utilize DAS for amplitude-critical applications, such as earthquake early-warning, must carefully evaluate the fidelity of DAS amplitudes to expected values corresponding to the true displacement ground motions. Observing that the features in Figure 4.4 are relatively independent of frequency band and wavetype, a naive correction method for amplitude response is to simply take the median amplitude ratio from the 6 rightmost panels and multiply the predicted waveforms; doing this improves the median RMSE log-amplitude fit from 0.61 to 0.54, with the histogram of corrected values also shown in Figure 4.3 showing significant improvement. We note that the two definitions of amplitude fit (RMS ratio for the P and S wave first arrivals vs log-amplitude RMSE for the entire waveform) are quite different so that this naive correction is not just simply curve-fitting. Of course, the naive correction does not improve the normalized cross-correlation values whatsoever, as it does not affect the shape of the waveforms, and as such we are motivated to develop more detailed correction methods that may result in further improvements. Given our previous observation that the spatial distribution of amplitude responses suggest that the strains observed by the DAS for the Porotomo experiment are affected by near-surface geology, we will develop an approach that models the DAS waveforms as a perturbation of the predicted long-wavelength strain-wavefield using homogenization theory.

4.4 Towards correcting DAS amplitude response using two-scale homogenization theory

Homogenization theory (Capdeville et al., 2010a,b) gives a framework for understanding the differences between observed strain-rates and those calculated from the reconstructed velocity field as observed by the nodes. The tutorial study by Capdeville et al. (2020) shows the potential issue starkly: material property distributions which produce smooth displacement waveforms and stress fields may produce

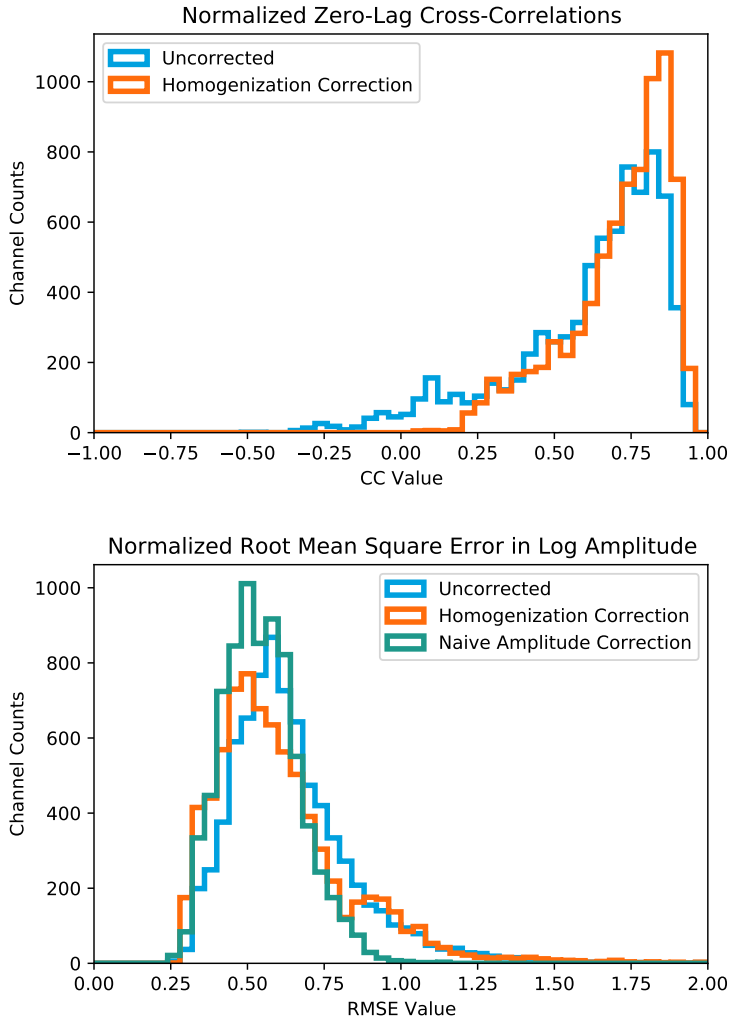


Figure 4.3: Histograms of the normalized zero-lag cross-correlation (after clock-correction) and normalized root-mean-square-error in log amplitudes between the observed DAS strain-rates and strain-rates predicted from the reconstructed nodal-seismic data, both before and after empirical correction based on two-scale homogenization theory.

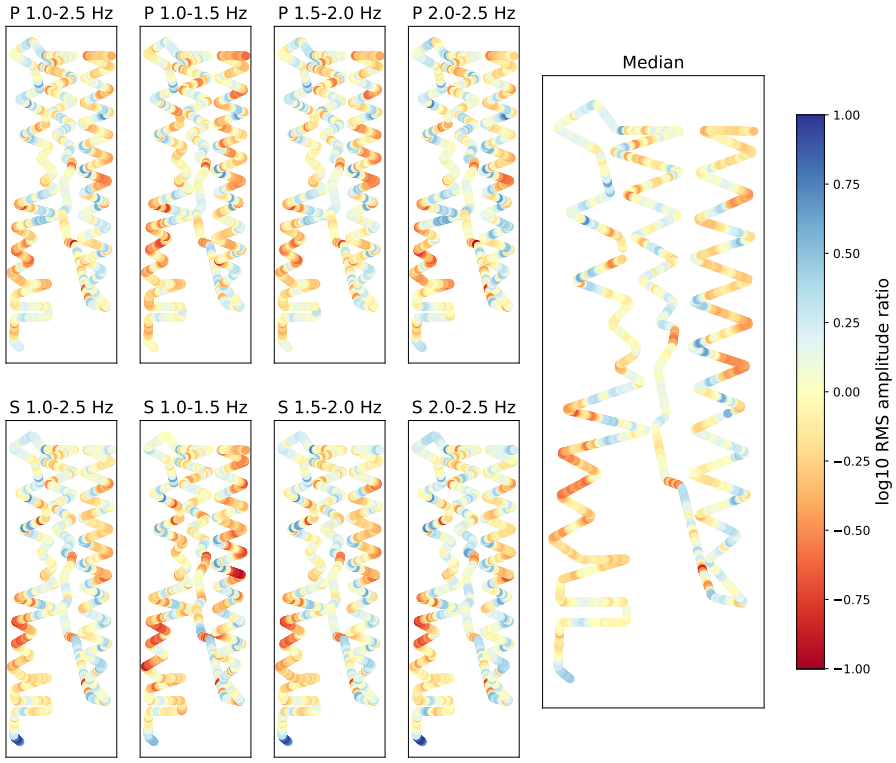


Figure 4.4: RMS amplitude ratios between the observed DAS data and the DAS strain-rates predicted from the nodal data reconstruction, for the P (upper row) and S (lower row) energy packets of the ML 4.3 Hawthorn NV earthquake, both for the full 1–2.5 Hz range of the reconstruction and in 1.0–1.5, 1.5–2.0 and 2.0–2.5 Hz bands. The pattern of ratios shows clear spatial patterns that cut across frequency bands, with the most coherent features being the two overpredicted red patches in the lower left and upper right of the array, which is apparent in the median amplification across the six time-frequency bands.

counterintuitively complicated strain fields, which will complicate the study of the resulting DAS measurements which sense the integrated strain along a gauge length. Singh et al. (2019) studied means to correct for this effect, using the theory of two-scale homogenization, and applied it to rotation measurements recorded using the G-ring laser at Wetzell, Germany. They found substantial improvements could be made to the predicted rotation measurements calculated using PREM by making an additional first order correction, with coefficients trained using observed rotation seismograms. A similar procedure may be used to obtain corrections for DAS seismograms, which we will outline here. For a full description of the theory, we refer the reader to the extensive development presented in Capdeville et al. (2020). We begin with the first-order expansion of the true displacement gradients $\nabla \boldsymbol{u}(x, y, t)$ in terms of reference displacements $\boldsymbol{u}^0(x, t)$ and strains $\boldsymbol{\epsilon}^0(x, t)$, where x is the "large scale" space variable and y is the "microscale" space variable:

$$\nabla \boldsymbol{u}(x, y, t) = \nabla_x u^0(x, t) + (\nabla_y \chi(y)) : \boldsymbol{\epsilon}^0(x, t). \tag{4.3}$$

The reference displacements and strains are calculated using suitably averaged properties, while the term $\chi(y)$ is the first-order correction operator, which is a degree-three tensor. Equivalently, in Einstein-summation notation, and for notational convenience dropping the dependence of the fields on x, y and t we have

$$u_{i,j} = u^0_{i,j_x} + \chi_{inm,j_y} \epsilon^0_{nm}. \tag{4.4}$$

Inserting this relationship into Equation 4.1 and assuming that we have no coupling between the vertical and horizontal components of the wavefield, we have

$$\begin{aligned} \epsilon_c = \sin^2 \alpha (u^0_{1,1_x} + \chi_{1nm,1_y} \epsilon^0_{nm}) + \frac{\sin 2\alpha}{2} (u^0_{1,2_x} + u^0_{2,1_x} + (\chi_{1nm,2_y} + \chi_{2nm,1_y}) \epsilon^0_{nm}) \\ + \cos^2 \alpha (u^0_{2,2_x} + \chi_{2nm,2_y} \epsilon^0_{nm}), \end{aligned} \tag{4.5}$$

where the numerical indices are not summed over and represent the two orthogonal horizontal components and α is the clockwise azimuth relative to component 2. Averaging this equation along the DAS cable over a gauge length L gives us the DAS strain

$$\epsilon_{DAS} = \epsilon^0_{DAS} + \frac{1}{L}\int_{-L/2}^{L/2} \sin^2\alpha\,\chi_{1nm,1_y}(l)\epsilon^0_{nm}(l) + \frac{\sin 2\alpha}{2}(\chi_{1nm,2_y}(l) + \chi_{2nm,1_y}(l))\epsilon^0_{nm} + \cos^2\alpha\,\chi_{2nm,2_y}(l)\epsilon^0_{nm}(l)dl, \tag{4.6}$$

where we have assumed a straight cable about the reference location at $l = 0$. If we further assume that we are in a regime such that the reference strains ϵ^0_{nm} are approximately constant over a gauge length, which we would expect to be true in the regime where displacements are well modelled and which allows the scale separation required by homogenization theory to be valid, then we can write

$$\epsilon_{DAS} \approx \epsilon^0_{DAS} + J_{11}\epsilon^0_{11} + J_{12}\epsilon^0_{12} + J_{22}\epsilon^0_{22}, \tag{4.7}$$

where

$$J_{11} = \frac{1}{L}\int_{-L/2}^{L/2} \sin^2\alpha\,\chi_{111,1_y}(l) + \frac{\sin 2\alpha}{2}(\chi_{211,1_y}(l) + \chi_{111,2_y}(l)) + \cos^2\alpha,\,\chi_{211,2_y}(l)dl, \tag{4.8}$$

etc. The coupling coefficients J are time independent and thus constant both within a particular recorded strain-rate time series, and between records. This analysis depends on assumptions that seem well justified (scale separation allows us to remove the reference strains from the gauge integrand), as well as those that are potentially less so (lack of coupling between vertical and horizontal strains), but does serve as a starting point for the correction of DAS records to local heterogeneity using the theory of two-scale homogenization. The coefficients J can be obtained for each DAS channel by computing

$$\boldsymbol{J} = (\boldsymbol{E}^{0^T}\boldsymbol{E}^0)^{-1}\boldsymbol{E}^{0^T}\boldsymbol{\delta}, \tag{4.9}$$

where for signal sample times $t = t_1, t_2, \ldots, t_n$

$$\boldsymbol{J} = \begin{bmatrix} J_{11} & J_{12} & J_{22} \end{bmatrix}^T, \tag{4.10}$$

$$\boldsymbol{E}^0 = \begin{bmatrix} \epsilon^0_{11}(t_1) & \epsilon^0_{12}(t_1) & \epsilon^0_{22}(t_1) \\ \epsilon^0_{11}(t_2) & \epsilon^0_{12}(t_2) & \epsilon^0_{22}(t_2) \\ \vdots & \vdots & \vdots \\ \epsilon^0_{11}(t_n) & \epsilon^0_{12}(t_n) & \epsilon^0_{22}(t_n) \end{bmatrix}, \tag{4.11}$$

$$\boldsymbol{\delta} = \begin{bmatrix} \epsilon_{DAS}(t_1) - \epsilon^0_{DAS}(t_1) \\ \epsilon_{DAS}(t_2) - \epsilon^0_{DAS}(t_2) \\ \vdots \\ \epsilon_{DAS}(t_n) - \epsilon^0_{DAS}(t_n) \end{bmatrix}. \tag{4.12}$$

When multiple records are used for the inversion, the different t_i need not be from the same record as J is time independent, assuming that the local heterogeneity is not changing. While the PoroTomo experiment only recorded one significant event, which is insufficient to perform a robust analysis similar to Singh et al. (2019), we can solve for J by using the nodal reconstructed wavefield of the Hawthorn NV event to compute the reference strain-rate and see what sort of improvements can be made using the assumptions developed in this section. It is worthwhile to note that the DAS records used here have significantly more high-frequency detail than the long-period rotation measurements in Singh et al. (2019), so the danger of overfitting is lessened. In figure 4.5, we show the corrected strain-rates calculated from the nodal reconstructed wavefield. In five out of seven of the example waveforms, we see a significant improvement in this amplitude metric. Because of the requirement for clock-corrections, these error metrics are calculated using the time range of 15–55 s. Averaged over the entire DAS array, we saw the median normalized zero-lag cross correlation increase from 0.69 to 0.76, and the median normalized RMSE in log amplitude decrease from 0.61 to 0.56—the average gain in waveform shape accuracy is significant, with the tail of negative zero-

lag cross-correlations eliminated completely. However, the improvment in RMSE amplitude fit is not as good as the naive amplitude correction. Histograms showing the distributions of these values across channels are shown in Figure 4.3, with the histogram of homogenized RMSE showing a secondary hump that may be due to the prioritization of the L2 metric of fitting phase over amplitude, which can result in the "best-fit" corrected amplitude (in the L2 sense) having very small amplitude. Further advancements of homogenization-theory based corrections may allow for more accurate matching of recorded DAS strains to predicted waveforms, which may substantially improve the efficacy of DAS arrays for studying source processes. Additionally, there is potentially scope for further inversion of the correction term J for small scale structure, although the requisite theory has yet to be developed.

4.5 Optimal Design of Mixed Networks

In this study, we have so far explored the potential of wavefield reconstruction to characterize a combined DAS and point-seismometer network, with the view that such deployments will become increasingly common in the future. Given that new arrays offer new opportunities for deployment design, a natural extension of our study of wavefield reconstruction is how to best design such mixed networks to maximise their reconstruction performance and ability to correct for DAS gain, and in particular how we might use our mathematical framework to formalize the array design. Within the general framework of Muir and Zhan (2021), the final, spatially resolved step of the wavefield analysis is a linear inverse problem, utilizing a L1 regularization to promote sparse solutions. Procedures for how to make optimal measurements for linear inverse problems have been widely studied for over a century, and have been collectively termed methods for Optimal Experimental Design (OED). While OED is a well established concept, methods for OED within the context of compressive sensing have only been recently developed. In the

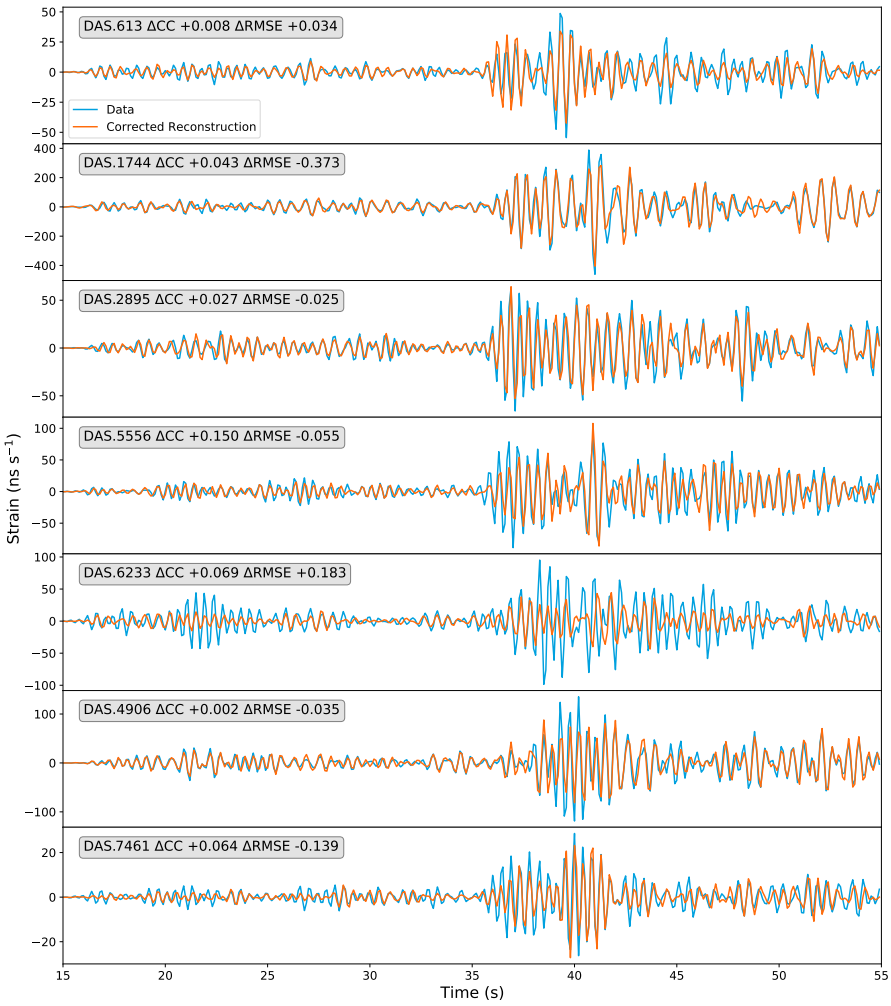


Figure 4.5: Comparisons between DAS channels (blue) and reconstructed corrected strain-rate from nodes (orange) at the seven stations highlighted in Figure 4.1. Corrections are obtained from applying a first order expansion of the observed DAS signal in terms of the predicted DAS signal and predicted strain-rates, as described in the text. Changes to the observed zero-lag normalized cross correlation ΔCC and log-amplitude error ΔRMSE show substantially improved fits compared to Figure 4.2 in almost all cases.

following, we use an algorithm proposed in Ravi et al. (2017) to determine optimal mixed network designs and evaluate their performance for improving wavefield reconstruction relative to unoptimized networks.

Following Ravi et al. (2017), the information matrix $\boldsymbol{M'}$ is defined by the product $\boldsymbol{G}^T\boldsymbol{G}$. For compressive sensing problems, $\boldsymbol{M'}$ is not full rank, so is regularized by $\boldsymbol{M} = \boldsymbol{M'} + \epsilon\boldsymbol{I}$. Using the D-optimal design criterion, which minimizes the confidence ellipsoid of the estimate for the inverted curvelet coefficients, good designs will minimize $\log\det\boldsymbol{M}^{-1}$. We introduce a selection vector $\boldsymbol{s}$, such that each element $\boldsymbol{s}_i \in \{0, 1\}$. We can write the matrix $\boldsymbol{M}^* = \sum_i \boldsymbol{s}_i\boldsymbol{g}_i^T\boldsymbol{g}_i + \epsilon\boldsymbol{I}$ where $\boldsymbol{g}_i$ are the row vectors of $\boldsymbol{G}$—each row vector corresponds to a candidate measurement location. The optimal design problem asks, given a budget of B stations, what the best $\boldsymbol{s}$ with $\text{sum}(\boldsymbol{s}) \leq B$ is, such that the solution of the compressive sensing problem $||Diag(\boldsymbol{s})(\boldsymbol{G}\boldsymbol{m} - \boldsymbol{d})||_2 + ||\boldsymbol{m}||_1$ is similar to that of the unrestricted problem $||\boldsymbol{G}\boldsymbol{m} - \boldsymbol{d}||_2 + ||\boldsymbol{m}||_1$.

Ravi et al. (2017) proposed a design criterion for compressive-sensing problems based on combining D-optimal design with an additional term designed to minimize the coherence of the selected stations. The coherence is defined as the maximum diagonal term of the hat matrix $\boldsymbol{H}^* = Diag(\boldsymbol{s})\boldsymbol{G}\boldsymbol{M}^{*-1}\boldsymbol{G}^T Diag(\boldsymbol{s})$ (this criterion is also known in the literature as G-optimal design). The full objective function to be minimized is therefore given by

$$f(\boldsymbol{s}) = -\log\det(\boldsymbol{M}^*) + \lambda \max\left[Diag(\boldsymbol{H}^*)\right], \tag{4.13}$$

for some regularization parameter $\lambda > 0$, and where the max is the maximum element of the diagonal of the hat matrix $\boldsymbol{H}^*$. Minimizing coherence tends to produce better solutions for compressive sensing, so combining these two criteria promotes a balance of solutions that allows for good sparse signal recovery (from the incoherence condition) while also maintaining overall sensitivity (from the D-

optimal condition). Solving this problem for binary $\boldsymbol{s}$ is an NP-hard nonlinear mixed-integer problem and is thus generally intractable as a polynomial time solution is not known, leading to two solution strategies. The first is to obtain an approximate solution via sequential optimization by finding the minimum of f for an initial station, then fixing that station and subsequently optimizing for the second station, etc. This sequential approach is computationally tractable and will often give a good solution but is not guaranteed to find the optima of $f(\boldsymbol{s})$; this approach has been commonly used for geophysical optimal design problems (e.g. see Bloem et al. (2020) for a recent overview including fully nonlinear design principles for small networks) and has the additional advantage that the same scheme can be also used to study the variant in which the candidate locations are allowed to vary, although we will not consider that possibility in this study. The second approach is to work with a relaxation $0 \leq \boldsymbol{s} \leq 1$, and then minimize $f(\boldsymbol{s})$ simultaneously for all stations; this approach was advocated by Ravi et al. (2017). The resulting relaxed weights give some indication of favorable designs; for instance they could be used as probabilistic weights for random network designs, or further schemes can be used to round the relaxed weights into a binary solution. We implemented the relaxed scheme using the JuMP interface (Dunning et al., 2017) to *ipopt* (Wächter and Biegler, 2006), which utilizes a log-barrier term to relax the constraints into the objective function.

For our mixed network problem, we have $\boldsymbol{s} = [\boldsymbol{s}_{DAS}; \boldsymbol{s}_x; \boldsymbol{s}_y]$. Typically DAS networks will either use existing dark-fiber instrumentation, or if using a greenfield deployment, will require complex constraints on the geometry due the required continuity of the cable, attempts to minimize corners, obstacles, etc.; consequently, we will restrict our discussion to the case that the DAS cable geometry is known and we wish to optimize the locations of deployed point seismometers to supplement the cable. In that case, we have the additional restriction that $\boldsymbol{s}_x = \boldsymbol{s}_y = \boldsymbol{s}_{Seismometers}$, $\boldsymbol{s}_{DAS} = \mathbf{1}$, and the restriction on the number of sensors is

then sum($\boldsymbol{s}_{Seismometers}$) $\leq B_{Seismometers}$. With these changes, the problem statement conforms to that in Ravi et al. (2017), and can be solved as described above.

To test both the sequential and relaxed methods of network design, we utilize a square synthetic array consisting of 384 DAS channels arranged in a cross-hatch formation, with 5m channel spacing and 10m gauge length, along with 64 candidate two-component horizontal seismometer locations evenly spaced over a central square. The geometry of the synthetic setup is shown in Figure 4.6. To determine reconstruction performance, we generated a synthetic Rayleigh-wave propagating across the array by simulating the acoustic wave equation with a point source for a weakly varying checkerboard velocity model (average velocity $\sim 1810 \pm 90$ m/s), and then taking the gradient to determine horizontal motions. We calculate the synthetic traces at each DAS channel and candidate seismometer location, including 10% added Gaussian noise, and also at 256 evenly spaced validation locations within the central square of the DAS array where we expect to be able to achieve adequate reconstruction results. All performance metrics reported are given by the per-trace normalized RMS misfit of the reconstructed wavefield at the validation stations.

For the relaxed analysis, we drew 10 lots of 1000 designs by sampling without replacement using the weights obtained by optimizing Equation 4.13. We set they hyperparameters to $\lambda = 1$ and $\epsilon = 10^{-2}$, with the latter motivated by the cutoff of the eigenvalue spectrum of the design matrix for a DAS array of totally random sensor locations and azimuths, which we would expect to have the best potential reconstruction performance—although that design is certainly not achievable in practice. The choice of λ was then made to give the incoherence term a similar scale to the typical range of perturbations to the D-optimal design term in Equation 4.13. For each of the 10 lots, we then chose the design with the lowest value of Equation 4.13 amongst the 1000 candidates, and performed a wavefield reconstruction on the data. We also performed a wavefield reconstruction on 10 designs chosen by

unweighted random selection without replacement. We tested designs with 4, 8, 16, 32 and 64 stations (with 64 being the complete set of potential locations, so that only 1 design was in fact tested). We also performed wavefield reconstructions for the sequential optima for those numbers of stations. Figure 4.7 shows the relaxed design weights, sequentially optimized design, and example resultant fits to the true synthetic waveform for the tested designs up to 32 included nodes. Overall performance at the example station is quite good, and unsurprisingly gets better as the number of included stations increases. The weights generally prefer stations that are further away from the DAS cable, particularly the outer crossing, which is somewhat intuitive as those areas already have provided information; it is however interesting to note that the center of the array is always weighted at least moderately despite the presence of a DAS crossing, presumably as the strong improvement to the D-optimal design term from occupying the center location outweights the coherence penalty there. Figure 4.8 shows the per-trace-normalized root-mean-square-error (RMSE) for this synthetic experiment as a function of the number of included nodes. In this case, despite there being no guarantee that the sequential design is particularly performant, it consistently has significantly lower RMSE than both the weighted and unweighted random designs, potentially as the regular, dense DAS array design makes choosing the locations of the next element of the sequence simple. For low node density, the weighted design substantially outperforms the unweighted design on average, while it confers no advantage once one quarter of the nodes are included.

4.6 Conclusions

DAS networks promise a paradigm shift within observational seismology by making large-N, highly spatially dense networks financially and logistically feasible for the research community. However, the strain based measurement procedure is

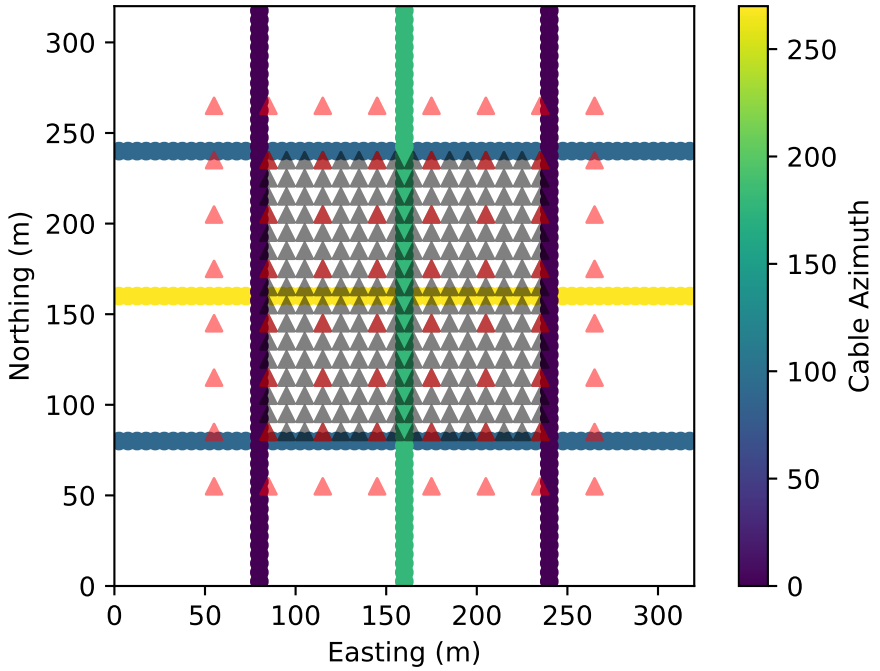


Figure 4.6: Network geometry used for synthetic optimal network design. DAS cable channel locations are shown as circles colored by cable azimuth. Red triangles show the locations of candidate stations, and grey triangles show the locations of the validation sensors placed within the central region where we should expect to be able to achieve good recovery of the wavefield.

highly susceptible to the effects of local heterogeneity and is furthermore sensitive to only one projection of the horizontal strain tensor. Using a wavelet-curvelet compressive sensing based wavefield reconstruction, we have developed a method for simultaenous spatial assessment of DAS array response using a theory that does not require potentially unstable fk rescaling, when a colocated but potentially off-cable point sensor array is available. We have shown that the DAS phase response in the 1-2.5 Hz range is relatively accurate to true ground motion for the ML 4.3 Hawthorne, NV earthquake recorded on the Porotomo DAS array, but that there are however significant amplitude errors that correlate with near-surface geology. These amplitude errors can be well characterized by a single empirical gain

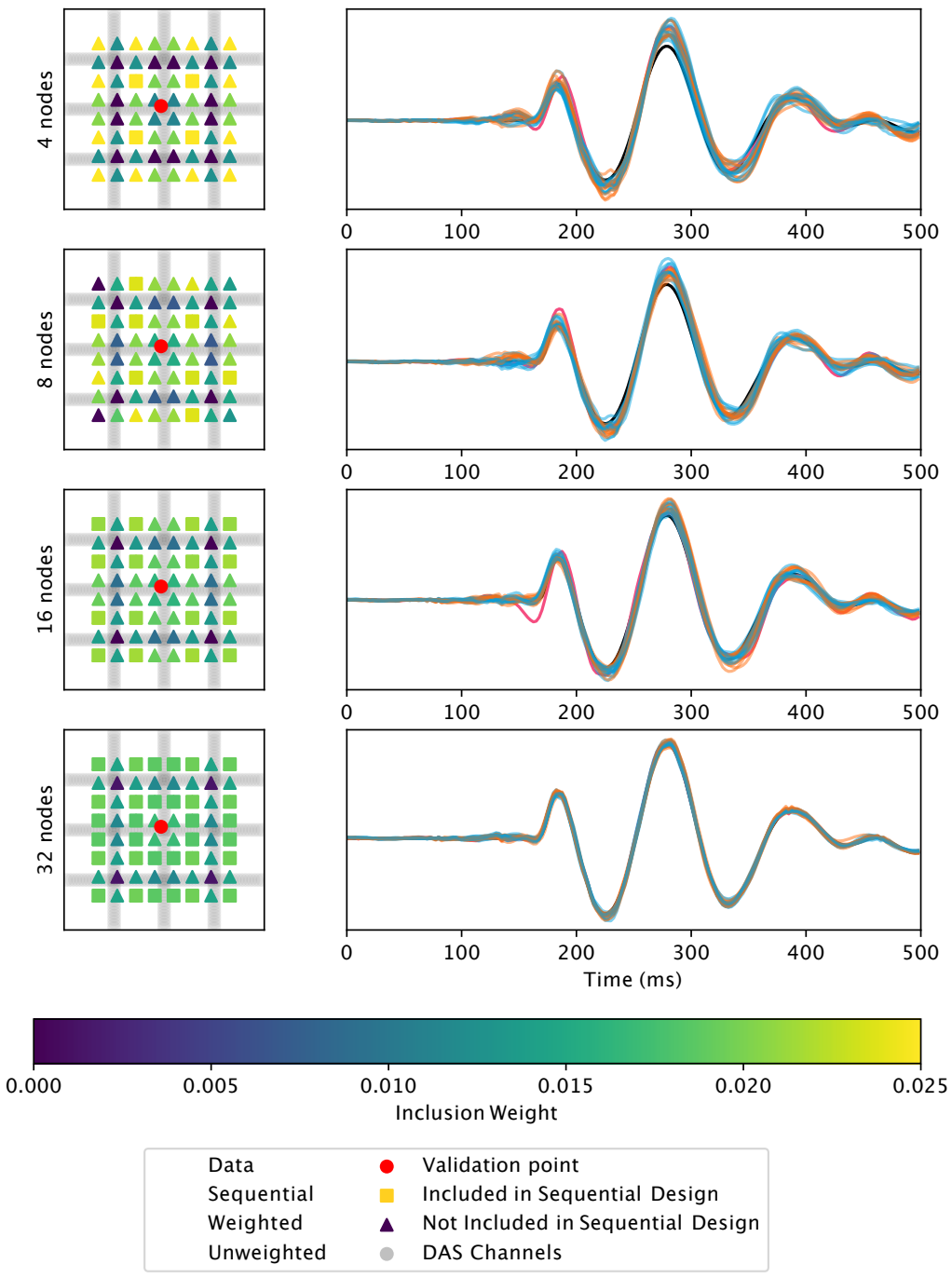


Figure 4.7: Array designs and reconstruction performance for different numbers of included nodes. The colors of the left column shows the inclusion weight for the relaxed OED problem. The shapes show the sequential insertion design, with square symbols showing included stations and triangular symbols showing non-included stations. The right column shows the evaluated reconstruction for the x component of the validation station at location (155,165), near the center of the array, with the black line showing the true data, the pink line the reconstruction for the sequential insertion design, blue lines showing unweighted random designs and orange lines showing weighted random designs.

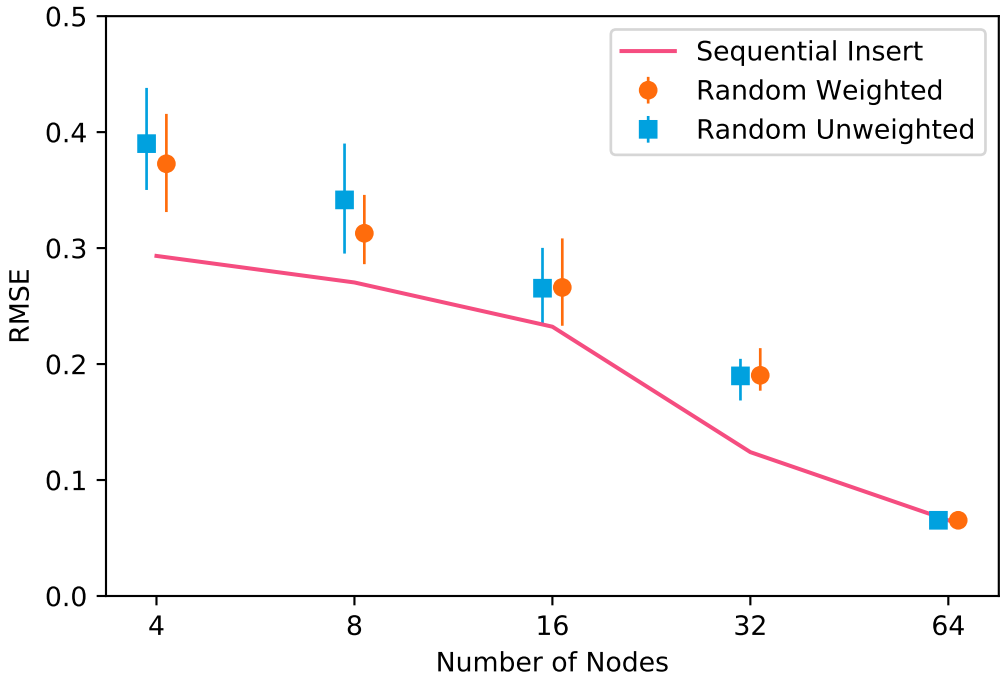


Figure 4.8: Per-trace normalized reconstruction root-mean-square-error (RMSE) calculated using the 256 validation stations shown in 4.6 as a function of number of utilized nodes in the design. The pink line shows the sequentially optimized design, while blue squares and orange circles show results from unweighted and weighted randomized designs respectively. The error bars show the maximum and minimum RMSE across the 10 samples used for each randomized design.

coefficient within the studied frequency band, and further detailed studies using the framework of homogenization theory may yield corrections that improve both phase and amplitude response. Furthermore, we have shown that wavefield reconstruction permits a cohesive framework for combining DAS with point sensors such as 3D nodes, with the combined DAS and nodes network outperforming the nodal network by itself for low node density using the PoroTomo array geometry. With a view to generalizing the applicability of these results beyond the PoroTomo deployment, we have investigated an optimal design strategy for improving the efficiency of mixed DAS and point sensor deployments for wavefield reconstruction. While at this stage we have focused on the relatively simpler task of deciding the best

locations for point sensors given a fixed DAS deployment, as this is computationally more tractable and also corresponds to the common case of utilizing pre-existing dark fiber deployments, the general concept of optimizing network sensitivity while also promoting the incoherence of measurements, will allow for optimal design of general DAS networks in the future. As DAS becomes increasingly prevalent and further integrated with traditional seismic networks, wavefield reconstruction therefore represents a flexible framework for overcoming the weaknesses of DAS as a single component large-N array, instead optimizing its strength as a single unified areal sensor.

4.7 Acknowledgements

The Authors would like to thank the editor, Andrew Valentine, and the reviewers, Eileen Martin and Karen Lythgoe, for their comments, which have greatly improved the manuscript. The Authors would like to thank Ettore Biondi for useful discussions during the preparation of this manuscript. The Authors would like to acknowledge the U. Wisconsin-Madison PoroTomo team for early access and usage instructions for the PoroTomo DAS dataset. JBM acknowledges the financial support of the Origin Energy Foundation and the General Sir John Monash Foundation during his PhD studies. ZZ thanks the support from NSF CAREER award EAR 1848166 and the NSF/IUCRC GMG Center. Seismic data were processed using Obspy (Beyreuther et al., 2010), and figures were created using Matplotlib (Hunter, 2007).

4.8 Data Availability

Data for the Hawthorn NV earthquake can be found at Feigl (2016a,b). The code used to produce this study will be included as a GitHub repository in the final submission.

4.9 Appendix: Wavefield Reconstruction Theory

In Muir and Zhan (2021), we proposed a wavelet decomposition in the time domain, coupled with a compressive-sensing based preconditioned curvelet decomposition in the spatial domain. Mathematically this involves a transform $\boldsymbol{w}(j_w, s) = WT(\boldsymbol{u}(t))$, operating trace-by-trace, from the collection of time-domain signals $\boldsymbol{u}$ to the wavelet domain indexed by scale j_w and time-position variable s, followed by the solution of an $L1$ regularized linear inverse problem $\boldsymbol{G}\hat{\boldsymbol{c}}(j_w, s) = \boldsymbol{w}(j_w, s)$ expressing the spatial behavior of the wavelet coefficients $\boldsymbol{w}(j_w, s)$ in terms of the curvelet coefficients $\hat{\boldsymbol{c}}(j_w, s)$. Individual curvelet coefficients are indexed by a scale j_c, as well as rotation and translation indices. The design matrix $\boldsymbol{G} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{P}$ in our formulation includes both the spatial behaviour of the curvelet basis functions, $\boldsymbol{\Phi}$, the spatial sampling matrix $\boldsymbol{\Psi}$ and a variable preconditioning matrix $\boldsymbol{P} = Diag(2^{-j_c{}^n})$ that allows us to promote smooth wavefields by penalizing the scale factor j_c of the curvelets.

We now develop the theory further to include DAS measurements of the strain field. As indicated in Equation 4.2, DAS strain metrics are calculated by averaging the true strain or strain-rate over a gauge length L. As such, we can invert for the horizontal wavefield curvelet coefficients by solving

$$\left[\frac{1}{L}\int_{-L/2}^{L/2} \boldsymbol{A}_1\frac{\partial \boldsymbol{G}}{\partial x} + \boldsymbol{A}_2\frac{\partial \boldsymbol{G}}{\partial y}dl \quad \frac{1}{L}\int_{-L/2}^{L/2} \boldsymbol{A}_2\frac{\partial \boldsymbol{G}}{\partial x} + \boldsymbol{A}_3\frac{\partial \boldsymbol{G}}{\partial y}dl\right]\begin{bmatrix}\boldsymbol{c}_x(j_w, s)\\ \boldsymbol{c}_y(j_w, s)\end{bmatrix} = \boldsymbol{w}_{\epsilon_{DAS}}(j_w, s) \tag{4.14}$$

$$\begin{bmatrix}\boldsymbol{G}_{DAS,x} & \boldsymbol{G}_{DAS,y}\end{bmatrix}\begin{bmatrix}\boldsymbol{c}_x(j_w, s)\\ \boldsymbol{c}_y(j_w, s)\end{bmatrix} = \boldsymbol{w}_{\epsilon_{DAS}}(j_w, s), \tag{4.15}$$

where $\boldsymbol{A}_1 = Diag(\sin^2\boldsymbol{\alpha})$, $\boldsymbol{A}_2 = Diag(\frac{\sin 2\boldsymbol{\alpha}}{2})$ and $\boldsymbol{A}_3 = Diag(\cos^2\boldsymbol{\alpha})$. The integrals are applied to each row, each of which corresponds to a single measurement for a cable segment with gauge length L and cable azimuth α. This inversion process

is unable to recover static components of the velocity field due to the derivatives involved, but given the normally oscillatory characteristics of solid earth ground motion, this is unlikely to be an issue in practice. Once the curvelet components are found, the ground velocity wavelet coefficients may be recovered by applying the undifferentiated curvelet basis matrix giving $\boldsymbol{w}_{u_x}(j_w, s) = \boldsymbol{G}\boldsymbol{c}_x(j_w, s)$ and $\boldsymbol{w}_{u_y}(j_w, s) = \boldsymbol{G}\boldsymbol{c}_y(j_w, s)$. The ground velocities in time domain can then be recovered by the inverse wavelet transform. In the reverse process, the horizontal components $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$ of 3C seismic deployments may be inverted using an appropriate design matrix $\boldsymbol{G}$, and the above equations used to obtain the predicted DAS signal. Finally, we may simultaneously invert both 3C seismometers and DAS deployments by forming the block matrix equation

$$\begin{bmatrix} \boldsymbol{G}_{DAS,x} & \boldsymbol{G}_{DAS,y} \\ \boldsymbol{G}_{Seismometers} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_{Seismometers} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}_x(j_w, s) \\ \boldsymbol{c}_y(j_w, s) \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_{\epsilon_{DAS}}(j_w, s) \\ \boldsymbol{w}_{u_x}(j_w, s) \\ \boldsymbol{w}_{u_y}(j_w, s) \end{bmatrix}, \tag{4.16}$$

which in the compressive sensing framework is given by L1 regularization

$$\begin{bmatrix} \hat{\boldsymbol{c}}_x(j_w, s) \\ \hat{\boldsymbol{c}}_y(j_w, s) \end{bmatrix} = \arg\min \left[\left\| \begin{bmatrix} \boldsymbol{G}_{DAS,x} & \boldsymbol{G}_{DAS,y} \\ \boldsymbol{G}_{Nodes} & 0 \\ 0 & \boldsymbol{G}_{Nodes} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}_x(j_w, s) \\ \boldsymbol{c}_y(j_w, s) \end{bmatrix} - \begin{bmatrix} \boldsymbol{w}_{\epsilon_{DAS}}(j_w, s) \\ \boldsymbol{w}_{u_x}(j_w, s) \\ \boldsymbol{w}_{u_y}(j_w, s) \end{bmatrix} \right\|_2 + \lambda \left\| \begin{bmatrix} \boldsymbol{c}_x(j_w, s) \\ \boldsymbol{c}_y(j_w, s) \end{bmatrix} \right\|_1 \right], \tag{4.17}$$

solved using the Lasso.jl Julia module (a fast reimplementation of the R code glmnet (Friedman et al., 2010)), with the regularization parameter λ determined automatically using the corrected Akaike information criterion or AICc (Burnham and Anderson, 2004), or via 5-fold cross-validation where computationally feasible.

4.10 Appendix: Reconstruction of nodal data using DAS

Having shown that the dense nodal deployment is able to well-recover the DAS strain-rate, we then performed the opposite procedure: recovery of both horizontal

components of the wavefield using only the DAS cable. We note that this is a significantly more challenging task than the previous nodes-to-DAS experiment, as we can only access the horizontal wavefield through its derivative. In order to stabilize the selection of curvelets when only using DAS, we use mixed L1 and L2 regularization (i.e. Elastic Net Regression) of the form $\alpha||\boldsymbol{c}||_1 + (1-\alpha)||\boldsymbol{c}||_2$ with $\alpha = 0.95$. We have found that this small additional amount of L2 regularization is not necessary when any nodal data is included. Figure 4.9 shows the resulting horizontal components at three node locations, using timing lags optimized for the P wave on each instrument—individually calibrated time corrections were used for this figure as the overall DAS to nodal reconstruction is not of sufficient accuracy to calculate a global correction. In this case, the fit to the phases and amplitudes of the initial P wave is generally good, and performance declines throughout the P wave coda. Waveform fits improve on the East components during the initial S wave, but fits to the North component are generally poor. Sensitivity analysis following Martin et al. (2018) is shown in Figure 4.10 assuming a near surface V_P of 1300 m/s and a Poisson solid relationship for V_S indicates that relative sensitivities for P, SV and SH motions are almost independent of frequency across the frequency band studied here. The cable geometry, near surface velocity structure and earthquake source location combine to give highest sensitivity to P, followed by SH, with little sensitivity to SV. The overall better fit of the P wave data is therefore unsurprising, as the effective wavelength of this low-slowness phase is large across the array, and the relative sensitivity to P motions is large, allowing the CS inversion scheme to better integrate the P wavefield. In the case of the S wave, the higher amplitude of the East vs North nodal channels suggests that the SH component of the S wavepacket is dominant (given that the earthquake is south-southeast of the aray). The higher sensitivity to SH, combined with higher SH component, explains the better fit to the East components of the motion at the nodes. In general, however, it is apparent

that the DAS cable is by itself insufficient to recover the wavefield to a desirable accuracy within the frequency range of this study.

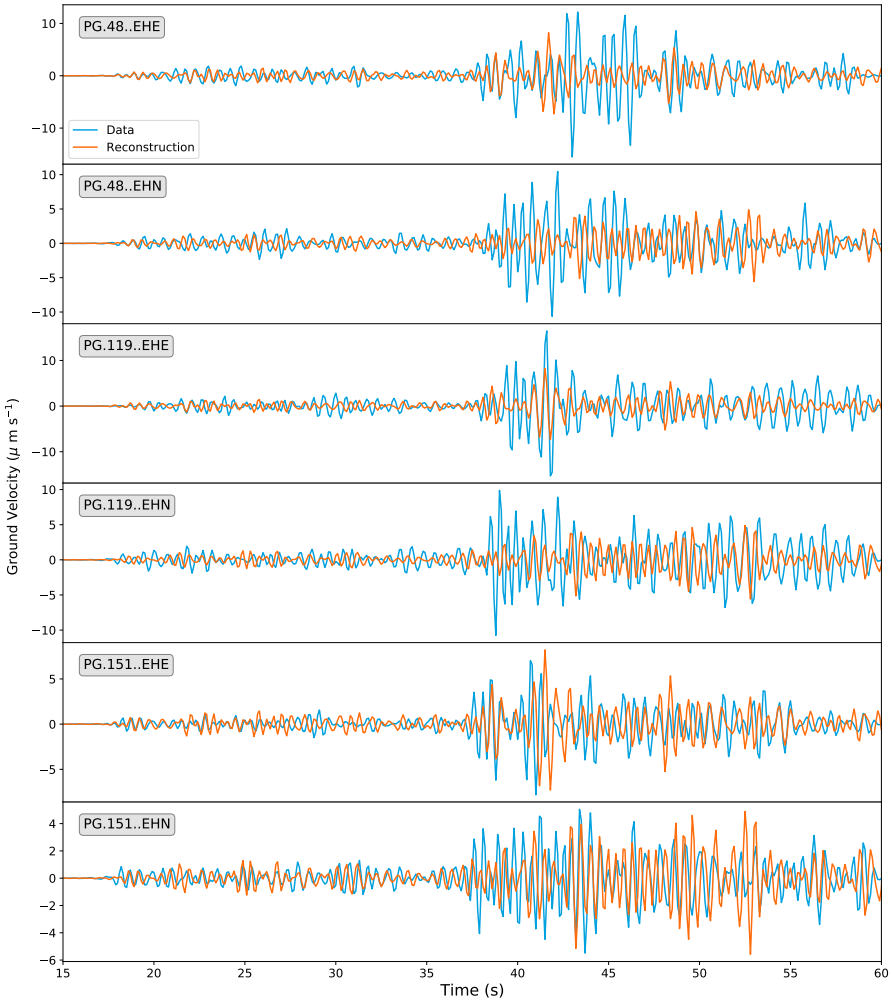


Figure 4.9: Comparisons between nodal seismic channels (blue) and reconstructed ground velocity from DAS (orange) at the three stations highlighted in Figure 4.1. Waveforms are bandpass filtered at 1–2.5 Hz, and nodal data has been corrected using a nominal Fairfield ZLand 3C instrument response.

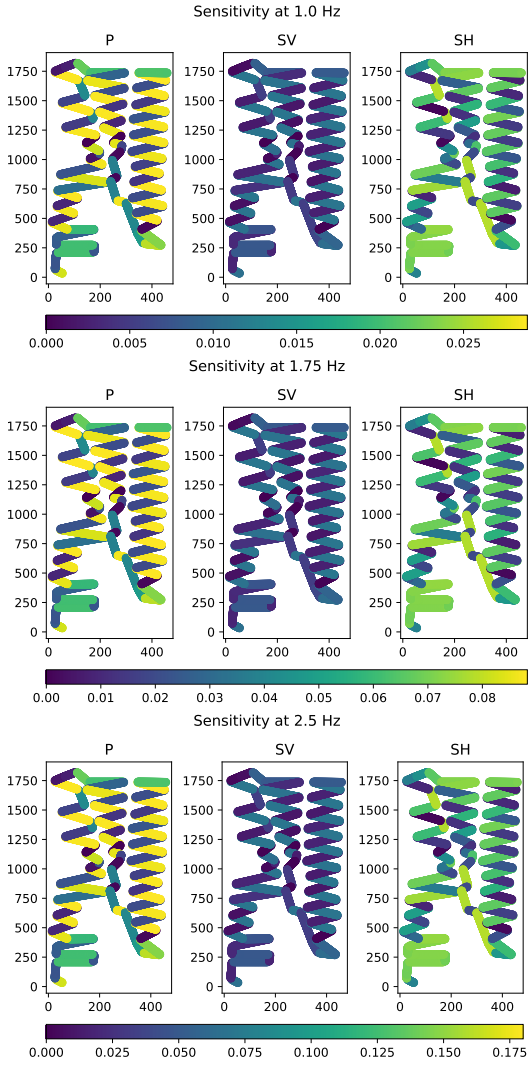


Figure 4.10: DAS cable sensitivity of the PoroTomo experiment at Brady, NV, at a frequencies of 1.0, 1.75 and 2.5 Hz to motions originating from the ML 4.3, March 21, 2016 studied in Wang et al. 2018. Notably, the relative sensitivity of the DAS cable is independent of cable azimuth across this period band

4.11 Appendix: Joint reconstruction of DAS and nodes

Finally, we study the performance of a mixed deployment of horizontally sensitive point sensors (nodes in this case) and DAS cable. We performed a five-fold cross validation experiment to obtain standardized mean square error (MSE) paths for nodal horizontal component reconstruction as a function of nodal instrument density. Specifically, we split the nodal dataset into five folds. With each of these folds acting as a validation dataset, we used between 20% and 100% of the remaining four folds as a training dataset. We then calculated the MSE of the reconstruction both with and without including the DAS data in the reconstruction. We also calculated the MSE of the DAS for each validation fold with no nodal data included. The results are shown in Figure 4.11. With the exception of one poorly performing fold, we see that the inclusion of DAS data substantially improves the MSE performance of reconstruction for low station densities. As the station density increases, the performance of the mixed network saturates whilst the performance of the nodal only network continues to improve. We interpret this saturation effect to be due to the inconsistencies between the nodal and DAS data due to clock errors and uncalibrated DAS-ground coupling; with a mixed network designed from the outset for joint wavefield reconstruction, improved performance at all station densities seems likely. This result highlights the potential of a mixed deployment to act as a unified areal sensor via wavefield reconstruction, even while utilizing a small number of three component seismometers. A network designed and calibrated with wavefield reconstruction in mind from the outset would likely have strongly improved performance when compared to the Porotomo experiment. This style of mixed network would be especially powerful for long-term or permanent deployment, where the costs of installing and maintaining dense three component arrays are substantial.

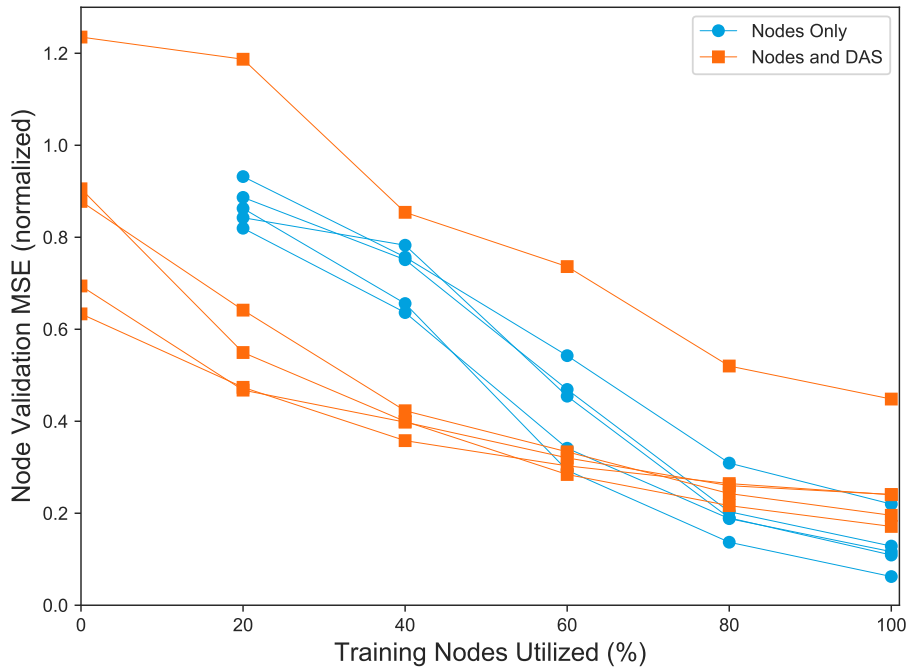


Figure 4.11: Mean Square Error (MSE) paths for a 5-fold cross-validation experiment. Blue lines show paths trained using only the remaining four-fifths of nodal data. Orange lines show paths with the addition of the full processed and quality-controlled DAS dataset.

Chapter 5

GEOMETRIC AND LEVEL SET TOMOGRAPHY USING ENSEMBLE KALMAN INVERSION

Muir and Tsai (2020b) previously published as

Muir, J. B. & V. C. Tsai (2020). "Geometric and level set tomography using ensemble Kalman inversion". In: *Geophysical Journal International* 220.2, pp. 967–980. DOI: 10.1093/gji/ggz472

5.1 Abstract

Tomography is one of the cornerstones of geophysics, enabling detailed spatial descriptions of otherwise invisible processes. However, due to the fundamental ill-posedness of tomography problems, the choice of parametrizations and regularizations for inversion significantly affect the result. Parametrizations for geophysical tomography typically reflect the mathematical structure of the inverse problem. We propose, instead, to parametrize the tomographic inverse problem using a geologically motivated approach. We build a model from explicit geological units that reflect the *a priori* knowledge of the problem. To solve the resulting large-scale nonlinear inverse problem, we employ the efficient Ensemble Kalman Inversion scheme, a highly parallelizable, iteratively regularizing optimizer that uses the ensemble Kalman filter to perform a derivative-free approximation of the general iteratively-regularized Levenberg-Marquardt method. The combination of a model specification framework that explicitly encodes geological structure and a robust, derivative-free optimizer enables the solution of complex inverse problems involving non-differentiable forward solvers and significant *a priori* knowledge. We illustrate the model specification framework using synthetic and real data examples of near-

surface seismic tomography using the factored eikonal fast marching method as a forward solver for first arrival travel times. The geometrical and level set framework allows us to describe geophysical hypotheses in concrete terms, and then optimize and test these hypotheses, helping us to answer targeted geophysical questions.

5.2 Introduction

Geophysical imaging methods, in particular seismic imaging, have offered the strongest constraints on the geometry and material parameters of the internal features of the Earth. Since the origin of geophysical inverse theory in the 1970s (e.g., Aki et al. (1977); Backus and Gilbert (1968); Dziewonski et al. (1977)), imaging methods have rapidly progressed with increasing computational resources, from small-scale linear tomography models to regional and global scale inversions fully utilizing the physics of the governing forward model (e.g., Rawlinson et al. (2010)). Despite these significant advancements, the interpretability of even well-constrained high-resolution seismic imaging results has remained challenging at regional and global scales, resulting in significant disagreements for the implications of seismic images (e.g., Foulger et al. (2013)). The potentially most significant underlying reason is the ill-posed nature of the inverse problem. Since the Earth is a three-dimensional (3D) continuous body, and our data is finitely distributed on or near the surface, there can never be a unique solution to the full continuum inverse problem. This ill-posedness necessitates regularization in imaging, either through explicit Tikhonov type additions to the data misfit function, which are equivalent in the Bayesian formulation to assumptions about the prior distribution of model parameters, or through implicit regularization via basis truncation (Parker, 1994; Rawlinson et al., 2014; Tarantola, 2005). Alternatively, some researchers have sought to use intuition informed by geodynamical considerations to create ad hoc images of the Earth through waveform modelling (Ko et al., 2017; Ni et al., 2002; Song et al., 2009;

Sun et al., 2016). These waveform modelling approaches are particularly important at higher frequencies (∼1Hz) where a combination of computational expense and required model complexity have precluded formal tomographic solutions at regional and global scales. Both the inverse problem approach and the waveform modelling approach have deficiencies. In the latter case, significant expert knowledge is required, and it is likely that only a limited range of candidate models will be tested. The former case does not rely on direct human intervention and is consequently potentially more objective, but the damping and smoothing regularization almost universally used create undesirable tomographic artifacts such as smeared rays and false compensating wave speeds near imaging targets as the misfit function attempts to balance penalties from the data and regularization.

Recent developments in Markov Chain Monte Carlo (MCMC) driven Bayesian tomography have helped to characterize the uncertainty of the results of seismic images (Tarantola, 2005), including the degree of data noise and model complexity in the now popular hierarchical transdimensional formulation (Bodin and Sambridge, 2009). These uncertainty measures can help one to understand poorly constrained parts of the resulting images, allowing more confidence in the predictions drawn from them. Recent results in transdimensional Bayesian tomography have highlighted the important impact of assumptions about the parametrization of internal boundaries on inversion results (Gao and Lekić, 2018; Roy and Romanowicz, 2017). Unlike the waveform modelling approach, which relies on strong *a priori* expectations about what potential structures may look like, seismic tomography in both deterministic and MCMC driven forms has typically only loosely prescribed the forms of acceptable models. We assert that in many cases, strong *a priori* knowledge does in fact exist, and that utilizing it can potentially significantly improve the resulting image in the inverse problem context. In addition, where intuition permits a range of potentially feasible geological structures, explicitly modelling

these options enables us to evaluate them within a hypothesis testing framework, quantitatively ranking potential models and rejecting models that do not fit the data (Claeskens, 2016).

At local scales, objectives of interest include the imaging of anomalous bodies such as tunnels or salt packages, geometric distortions such as faults, and stratigraphic interfaces. At the regional and global scale, there are clear targets of opportunity for which we have strong information from high-frequency waveforms that sharp physical contrasts exist, such as perturbations in important radial discontinuities (the Moho, 410 and 660 discontinuities) and abrupt localized features (slabs, ultra-low velocity zones, sedimentary basins). Specialized methods, such as receiver function analysis, exist to image these structures but they are difficult to use in a traditional tomographic framework. Parameterizing the tomographic inverse problem in such a way that these boundaries are explicitly modeled may help to overcome this limitation. This observation leads to the fundamental idea of this study, which is to pose the geophysical inverse problem as an optimization of explicitly defined geologic structures. Candidate structures for our proposed methodology are shown in Figure 5.1. Defining the inverse problem in this way allows us to better test hypotheses formulated using our *a priori* knowledge, as these hypotheses can be directly modeled. Viewed from another perspective, our inverse framework places waveform-modelling type approaches on a more rigorous footing by allowing greater flexibility in the range of permitted models and supplying the tools necessary for handling larger scale inversions than is possible using an exhaustive full model-space search. An alternative to our explicit modelling viewpoint would be use of the null-space shuttle, which allows a priori information to be added after an optimal solution is obtained (de Wit et al., 2012; Deal and Nolet, 1996; Fichtner and Zunino, 2019).

The purpose of this paper is threefold. The first part will describe a method

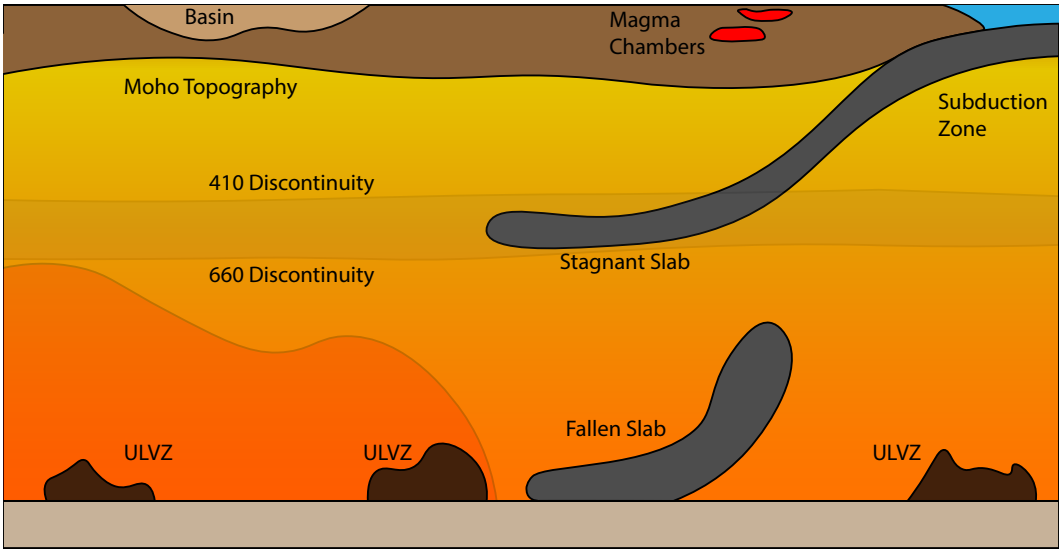


Figure 5.1: Schematic of the types of imaging targets that represent distinct domains with different geophysical properties; these targets are candidates for our proposed methodology.

of defining Earth models that allows for flexible modelling of explicit structures, enabling an improvement in the interpretability of inverse problems. Second, we will introduce from the inverse problem literature a derivative-free optimizer based on the Ensemble Kalman Filter, known as Ensemble Kalman Inversion (EKI), and further describe the details of the algorithm for a geoscience audience. Finally, we will illustrate the use of our model definition scheme and EKI to solve nonlinear travel time tomography problems.

5.3 Model Specification

Parametrization is a fundamental design choice present in all geophysical inverse problems. Parametrizations must seek to accurately represent potential Earth structure, interface with forward solvers, closely predict the data, and lead to solutions of the inverse problem that can be stabilized against the effect of data noise. These potentially conflicting goals have led to a profusion of different parametrization schemes, ranging from simple Cartesian block models, to more exotic basis function sets or spectral domain methods, to irregular multiscale parametrizations designed

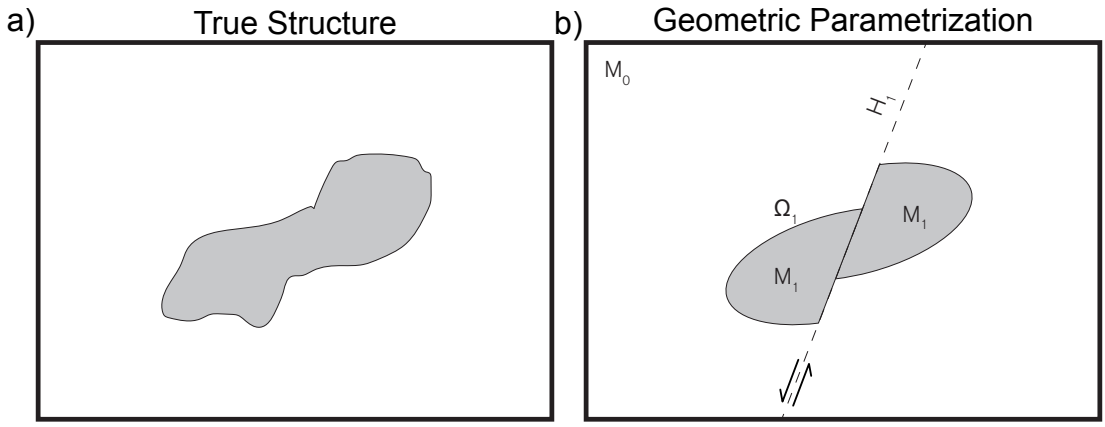


Figure 5.2: Schematic of the geologically motivated parametrization proposed by this study. a) shows some body in the Earth that is the imaging target, for which we have some *a priori* knowledge. b) shows a potential geometric parametrization of the body which we optimize using EKI. M_0 encodes the background model, while Ω_1 and M_1 are the boundary and interior properties of the first model layer, respectively. H_1 is a deformation rule that further alters the model.

to tune model complexity to match the data (Rawlinson et al., 2010). In this study, we aim to introduce a parametrization designed to intuitively describe geological features. Because the model is built up from discrete units that are fixed *a priori*, we use the term model specification rather than parametrization; this highlights that the researcher explicitly introduces their *a priori* knowledge into the inverse problem by determining the number and type of geological features solved for, and also emphasises that the model is independent of the form of discretization used to solve the forward problem.

In the inverse problem context, a model specification for domain X must provide a set of P functions $\{F_p(x)\}_{p=1}^{P}$ that determine the P material properties of interest at an arbitrary point $x \in X$. Our model specification framework describes the inversion domain X via a set of simple layers. The base layer defines a background, or reference, model M_0. The background model has a set of material property functions $F_p(x; M_0)$ that are defined for $x \in X$. So defined, the background model could range from a homogeneous space to a fully 3D model depending on *a priori*

knowledge. On top of the base layer, J objects M_j are defined, each with their own geometries $\Omega_j \subseteq X$ and material parameter functions $F_p(x; M_j)$. We define F_p for a collection of objects as $F_p(x; \{M_j\}_{j=0}^J) = F_p(x; M_{j'})$ where j' is the largest integer with $x \in \Omega_{j'}$—in concrete terms, we select the topmost layer that contains x, reverting to the background if no higher layers are available. Once the objects are assembled, K deformations (such as faults) are included. The deformations are defined by invertible functions $H_k(x), X \to X$. To evaluate the model at a particular point in space, these deformations are reversed, so that $F_p(x; \{M_j\}_{j=0}^J, \{H_k\}_{k=1}^K) = F_p(x'; \{M_j\}_{j=0}^J)$ where $x' = H_1^{-1} \circ H_2^{-1} \circ ... \circ H_K^{-1}(x)$. These operations are shown schematically in Figure 5.2. Figure 5.2a shows an imaging target, while Figure 5.2b shows a geometric parametrization for the body that can be specified using our parametrization framework, and optimized using EKI to fit available geophysical data.

In the applications discussed in this paper, we are typically interested in describing the interface between two or more geologic units (i.e. the boundaries of regions Ω_j). If the interface is expected to be relatively simple—for instance, if we were attempting to image a near-surface tunnel—then an explicit description of the interface is convenient. An explicit description may be based on deformed geometric primitives, or by describing the locations of spline knots or polygon vertices etc. These explicit definitions have the advantage of reducing the number of parameters required to describe interfaces. However, they are relatively inflexible descriptions, especially when data requires that the topology of the interface should be different from that assumed by the explicit definition (for instance, if two bodies should be merged into one or vice versa). These situations may require the use of transdimensional methods in which model parameters are added and removed, which significantly increases the complexity of the inverse problem.

Alternatively, object boundaries may be defined implicitly by means of an auxiliary

function. Implicit definitions handle complex boundaries and changes in topology, while avoiding the need to change the number of parameters during the inversion. In the following sections, we describe the level set method as a way of implicitly defining object boundaries, and Gaussian random fields as a means of controlling the behavior of level set functions.

The level set method

The level set method partitions space into disjoint regions by considering contour lines of a set of n continuous auxiliary functions $\{\phi_i\}_{i=1}^{n}$. The rationale behind the method is that discontinuous fields can be represented in this way by continuous fields of a higher dimension, which often makes the handling of boundaries more mathematically tractable. Associated with the auxiliary fields are regional parameter fields $\{A_j\}_{j=1}^{N}$ that describe the value of the parameter of interest within a region. To construct a parameter field F described by level sets, we may use either a combinatoric or a procedural definition. In this work, we employ the procedural definition as it is simpler to implement and combine with other elements of our model definition; however it does not allow for explicit differentiation of the model. The more commonly used combinatoric definition is given in the appendix for comparison.

Procedural Definition: For N regional parameter fields of interest, set n such that $N = n$. Then $F(x) = A_i(x)$ for the largest i such that $\phi_i(x) > 0$. In this procedural definition, where multiple ϕ are non-zero, we "paint over" with increasing i in a similar fashion to other elements of our model definition framework. Each auxiliary field is individually associated with a spatial region and its associated parameter field, which aids intuition.

Implicit definition of potentially discontinuous boundaries via the level set method has been actively developed since its introduction for the solution of interface evo-

lution problems (Osher and Sethian, 1988). In the level set method, an interface is represented by a particular contour on a continuous auxiliary field—examples are shown in Figure 5.3. Level-set based tomographic methods have recently been intensively studied in the context of Electrical Impedance Tomography (EIT, also known as resistivity tomography, e.g., Chung et al. (2005)), hydrology (Cardiff and Kitanidis, 2009; Iglesias et al., 2013) and in various exploration geophysics contexts, especially crosswell seismic tomography and to a more limited extent gravity and magnetic applications (Isakov et al., 2011; Li et al., 2014, 2017; Li and Qian, 2016; Lu and Qian, 2015; Zheglova et al., 2013, 2018). Existing work has typically assumed piecewise constant fields, often of prescribed value, as this strong a priori knowledge is often available in exploration contexts. Under this framework, authors have found significantly improved reconstruction of interfaces compared to the smoothed images available from traditional Tikhonov regularized tomographic methods. Work within the geophysics community has exclusively employed the level-set evolution equation, which requires the calculation of the Fréchet derivative of the data misfit functional with respect to the level set function. The misfit functional is typically equipped with regularization that penalizes longer interface lengths (i.e. Total Variation, or TV regularization, Osher et al. (2005)). The level set evolution equation allows for efficient inversion but restricts the applicability of the level set formulation to contexts for which the Fréchet derivatives are available. Additionally, existing applications using the level set evolution equation (Li et al., 2017; Zheglova et al., 2018) require significantly more mathematical machinery when multiple level sets are used, limiting their applicability to complex models. When the derivatives are not available, for example when using externally supplied black-box forward models, the level-set evolution equation and also traditional iterative gradient-based tomographic methods break down. An alternative to TV regularization of level sets is specification of a Gaussian random field prior for the

auxiliary field used to generate the level set (e.g., Chada et al. (2018)). Using a Gaussian random field prior allows explicit control of the dominant length scale and roughness of the resultant level set, as shown in Figure 5.3. A possible alternative would be to learn appropriate basis function representations of the level-set auxiliary field from data using a dictionary learning approach (e.g., Bianco and Gerstoft (2018)). Due to its conceptual simplicity, the Gaussian random field based level-set approach is taken in the examples below to specify the boundary of object layers with our model specification framework.

Gaussian random fields

Gaussian random fields (GRFs, also commonly referred to as Gaussian processes, especially in 1D applications) have a long history in geostatistics where they provide the framework for kriging estimators of fields with observed training data (Chiles and Delfiner, 2012). In the inverse problem setting, the quantities of interest are not observed directly. For the linear or weakly nonlinear case, Hansen et al. (2006) has supplied theory for conditioning GRF priors on averaged observations such as travel times in fixed raypath tomography. An intriguing further development in applying GRFs to geophysical inverse problems has recently been provided by Ray and Myer (2019) which utilizes transdimensional MCMC for sampling training points on which the GRFs are conditioned. In this study, we use GRF priors, without conditioning on training data points, for the auxiliary fields used by the level set method. Thus, the material parameter fields are not determined by the GRFs directly, but rather by a nonlinear transform of them that can encode abrupt changes in material properties.

A comprehensive review of GRFs is given by Rasmussen and Williams (2006); here we offer a brief summary of definitions that are important to the model specification scheme outlined in this study. A scalar valued GRF on $\mathcal{R}^n$ is a spatial process

analogous to a Gaussian distribution. It is defined by a mean function $m(x)$ and symmetric covariance function $C(x, y) = C(y, x)$ and has the property that any finite set of points $\{x_k\}_{k=1}^{M}$ on the field are distributed as a multivariate Gaussian distribution $N(\mathbf{m}, \Sigma)$ with

$$\mathbf{m} = \begin{bmatrix} m(x_1) \\ \vdots \\ m(x_M) \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} C(x_1, x_1) & \dots & C(x_1, x_M) \\ \vdots & \ddots & \vdots \\ C(x_M, x_1) & \dots & C(x_M, x_M) \end{bmatrix}. \tag{5.1}$$

The covariance function encodes the GRF's spatial correlation behavior, and may be anisotropic and spatially varying (non-stationary); in many typical applications, a subclass of isotropic, stationary GRFs are employed for which only the distance between x and y matters, i.e. $C(x, y) = C(||x - y||)$. Within this class, the prototypical covariance functions with spatially constant variance σ^2 are the exponential covariance with characteristic lengthscale l

$$C(x, y) = \sigma^2 \exp\left(-\frac{||x - y||}{l}\right), \tag{5.2}$$

and the Matérn covariance function with smoothness parameter ν and length scale parameter ρ

$$C(x, y) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}||x - y||}{\rho}\right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}||x - y||}{\rho}\right), \tag{5.3}$$

where Γ is the gamma function and K_ν is the modified Bessel function of the second kind. The Matérn covariance function includes the exponential covariance function (as well as the squared-exponential variance function) as a special case. Figure 5.3 shows a table of example GRFs defined by Matérn covariances of different length scales and smoothness parameters, as well as level set partitions that can be defined by them. GRFs form a useful class of fields for defining boundaries in models using the level set method as they encode a wide range of potential prior information that can be tailored to a particular geophysical problem.

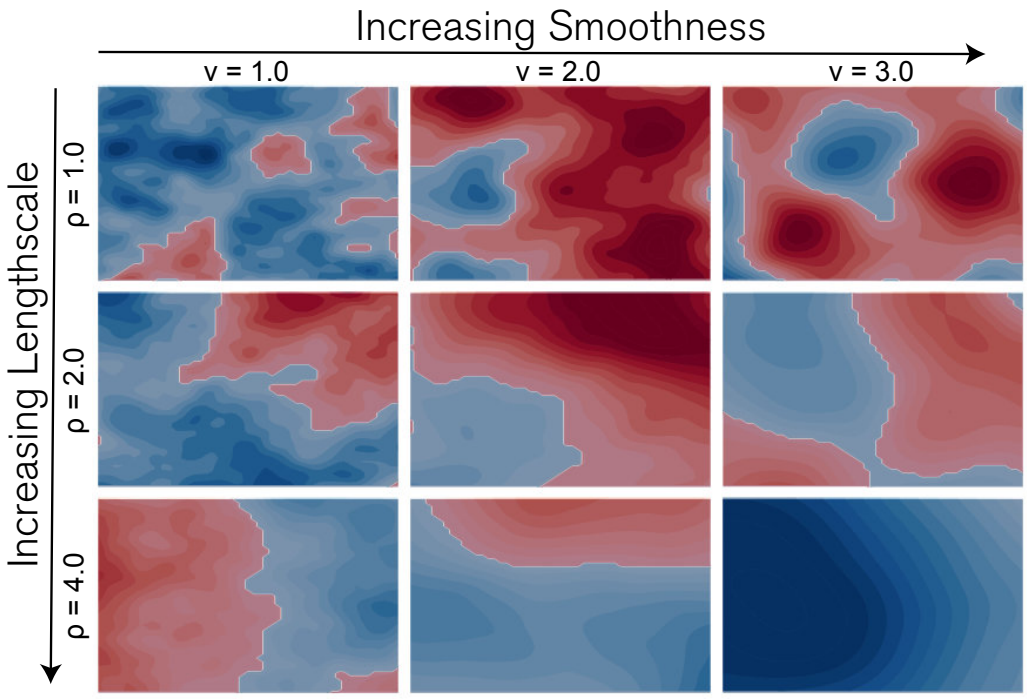


Figure 5.3: A table of sample zero-mean Gaussian random fields (GRFs) shown by continuous contours. These are overlain by a transparent two-color image showing a possible level set partition into two fields, defined by the zero contour level of the GRFs. The underlying continuous GRFs, which are visible underneath the two-color image, give rise to the discontinuous final level set partitioning.

5.4 Ensemble Kalman Inversion

The model specification framework proposed in this paper aims to map the *a priori* information of a researcher into an Earth model in a way that is independent of the computational requirements of the forward solver used by the inverse problem. While defining models in this fashion is advantageous from the perspectives of ease of usage and interpretation, it potentially makes derivatives of the desired geophysical observables with respect to model parameters difficult to calculate using fast analytic or adjoint methods—that is, the model function F may not be easily differentiable. Additionally, derivatives of the physical model may not be available when using closed source or legacy code. Since explicit calculation of derivatives via finite differencing is intractable for models with many parameters, and we

wish to maintain solver independence, in general precluding the use of algorithmic automatic differentiation, we are motivated to employ an efficient derivative-free optimizer for inverse problems defined using our model specification framework. In particular, we have employed the Ensemble Kalman Inversion optimizer (Iglesias et al., 2013), which we define below.

The Ensemble Kalman Inversion (EKI) scheme was introduced by Iglesias et al. (2013) as a derivative-free ensemble-based approximation of the iteratively regularized Levenberg-Marquardt (LM) inversion scheme (Hanke, 1997). Further development has resulted in applications to hydraulic reservoir modelling (Chada et al., 2018; Iglesias, 2014), electrical impedance tomography (Chada et al., 2018), and for optimizing neural-network parameters in machine learning (Kovachki and Stuart, 2018). For comparison, we first describe the regularizing LM scheme from which the EKI scheme is derived. We closely follow the development in Iglesias (2016).

Iteratively Regularized Levenberg-Marquardt scheme: The iteratively-regularized LM scheme considers an inverse problem with model parameters $u \in X$ and data $y \in Y$. X and Y are Hilbert spaces with appropriate norms $||\cdot||_X$ and $||\cdot||_Y$; in a geoscience application X will typically be either a finite dimensional space $\mathbb{R}^P$ or a function space on $\mathbb{R}^P$ and Y will be a finite dimensional observation space $\mathbb{R}^M$. We assume that $y = G(u^\dagger) + \eta$ for some model operator G, "true" set of model parameters $u^\dagger$ and noise η. Using our model framework, we typically have a set of model parameters that are transformed by the model function F to the physical model of interest on an evaluation grid, which are then input into a forward solver H so that $G(u) = H(F(u))$. We assume *a priori* knowledge of the noise level

$$\eta = ||\Gamma^{-1/2}(y - G(u^\dagger))||_Y, \tag{5.4}$$

where Γ is an operator that encodes the measurement precision, so that the absolute

misfit $(y - G(u))$ is weighted to account for the quality of measurements. For finite dimensional observations equipped with the normal Euclidean norm, if we assume that $\eta \sim N(0, \sigma)$ (i.e. Gaussian noise with variance σ^2) and we set $\Gamma = I$ so that observations are equally weighted, then $\eta \approx \sigma\sqrt{M}$ where M is equal to the number of observations.

The objective of any iteratively regularized scheme is to find a model u^η that is a stable approximation of $u^\dagger$ with respect to the noise in the sense that as $\eta \to 0$, then $u^\eta \to u^\star$ for some $u^\star \in X$ with $G(u^\star) = G(u^\dagger)$. In contrast to standard Tikhonov regularization methods, in which the problem is explicitly regularized and then optimized, iteratively regularized schemes fundamentally seek an approximate solution to the unregularized problem but stabilize the parameter updates and terminate at an appropriate level of fitting to avoid being dominated by noise. In the LM scheme, this condition is achieved by solving a succession of Tikhonov regularized updates with regularizing parameter α_n

$$u_{n+1} = u + v^*, \tag{5.5}$$

$$v^* = \arg\min_{v \in X} \left(||\Gamma^{-1/2}(y - G(u_n) - DG(u_n)v)||_Y^2 + \alpha_n ||C^{-1/2}v||_X^2 \right), \tag{5.6}$$

where $DG(u_n)$ is the Frechét derivative of G in respect to u, so that $y - G(u_n) - DG(u_n)v$ is a linear approximation of the misfit about u_n. C is an operator $X \to X$ that encodes regularity or prior information on X, and α_n controls the strength of the regularization at each update step. Note that within the LM scheme, the linear term of the first quadratic form gives rise to a steepest descent update, while the second order term gives the Gauss-Newton approximation of the Hessian (see Appendix 5.9 for the derivation in finite dimensions). The desired stable convergence property of LM was shown by Hanke (1997) to require that α_n at each iteration must satisfy

$$\rho||\Gamma^{-1/2}(y - G(u_n))||_Y \leq \alpha_n ||\Gamma^{-1/2}(y - G(u_n) - DG(u_n)v^*)||_Y \tag{5.7}$$

for a tuning parameter $\rho \in (0, 1)$ that is set *a priori* and fixed for all iterations. The scheme is terminated when

$$||\Gamma^{-1/2}(y - G(u_n))||_Y \leq \tau\eta < ||\Gamma^{-1/2}(y - G(u_{n-1}))||_Y \tag{5.8}$$

for some fixed $\tau > 1/\rho$, where this inequality is required to ensure stable convergence of the scheme (Hanke, 1997). This termination criterion is a form of Morozov's discrepancy principle (Scherzer, 1993), and ensures that the LM scheme does not overfit to the noise; without the termination criterion the scheme is equivalent to a modification of a standard unregularized Levenberg-Marquardt optimizer. Equation 5.6 can be shown to be equivalent to the explicit update

$$u_{n+1} = u_n + (DG^\star(u_n)\Gamma^{-1}DG(u_n) + \alpha_n C^{-1})^{-1}DG^\star(u_n)\Gamma^{-1}(y - G(u_n)), \tag{5.9}$$

with $DG^\star$ the adjoint operator of DG (Iglesias and Dawson, 2013). For finite dimensional X, the m^{th} component of DG is $D_m G(u_n) = J(u_n) \cdot e_m$ where J is the Jacobian of G and e_m is the unit vector for m; Equation 5.9 then simplifies to the standard finite dimensional LM update

$$u_{n+1} = u_n + (J(u_n)^T\Gamma^{-1}J(u_n) + \alpha C^{-1})^{-1}J(u_n)^T\Gamma^{-1}(y - G(u_n)). \tag{5.10}$$

When the Frechét derivative of G is available, the iteratively regularizing LM scheme provides a useful framework for the general solution of nonlinear inverse problems, and has been applied successfully in geophysical applications for groundwater flow (Hanke, 1997; Iglesias and Dawson, 2013).

Ensemble Kalman Inversion scheme: The Ensemble Kalman Inversion (EKI) scheme is an ensemble approximation of the iteratively regularized LM scheme. The general concept is to update an ensemble of particles (where each particle represents a realization of the model) using the ensemble Kalman filter (Evensen, 1994; Iglesias, 2016). Ensemble Kalman filters have recieved recent attention in the seismology

community as a means of uncertainty quantification in large scale full-waveform inverse problems (Eikrem et al., 2019; Thurin et al., 2017), although this usage is still in its preliminary stages. In this study, we utilize the EKI dynamic purely as a high-performance iteratively-regularized optimizer.

The dynamics driving the EKI ensemble are designed to drive the mean of the particles towards the solution of the inverse problem of interest (Iglesias et al., 2013). At each step, the ensemble of particles solves an approximate Tikhonov regularized update with iteration dependent regularization α_n, the strength of which is controlled by a global regularization parameter ρ (Iglesias, 2014). We let the ensemble at iteration n be $\{u_n^{(j)}\}_{j=1}^J$ where J is the number of ensemble members. Means of collections are denoted by over-bars (i.e. $\bar{u}_n$ is the mean over the collection of $\{u_n^{(j)}\}_{j=1}^J$). Approximating $G(u_n^{(j)})$ to first order about the mean of the ensemble

$$G(u_n^{(j)}) \approx G(\bar{u}_n) + DG(\bar{u}_n)(u_n^{(j)} - \bar{u}_n). \tag{5.11}$$

Iglesias (2016) shows that using this approximation, explicit calculation of the Frechét derivative $DG(u)$ may be eliminated, leading to an approximation of the iteratively regularized LM scheme by the following algorithm:

1. *Initialization*: Draw $\{u^{(j)}\}_{j=1}^J$ ensemble members from prior distribution. Set $\rho \in (0, 1)$ and $\tau > 1/\rho$. Then for $n = 0, 1\ldots$

2. *Prediction*: Evaluate $w_n^{(j)} = G(u_n^{(j)})$; calculate $\bar{w}_n$

3. *Termination*: If $||\Gamma^{-1/2}(y - \bar{w}_n)|| \leq \tau\eta$, terminate and output $\bar{u}_n$ as the solution

4. *Analysis*: At each iteration, an ensemble of perturbed data $\{y_n^{(j)}\}_{j=1}^J$ is generated with additional noise $y_n^{(j)} = y + \eta$. Addition of extra noise helps the ensemble to better explore parameter space by preventing the ensemble from converging to a single point from which ensemble gradients cannot be computed. Let $\langle \cdot, \cdot \rangle_Y$ being the inner product on Y and define covariance operators

C_n^{uw} and C_n^{ww} by

$$C_n^{ww}(\cdot) = \frac{1}{J-1}\sum_{j=1}^{J}(G(u_n^{(j)}) - \bar{w}_n)\langle G(u_n^{(j)}) - \bar{w}_n), \cdot\rangle_Y, \tag{5.12}$$

$$C_n^{uw}(\cdot) = \frac{1}{J-1}\sum_{j=1}^{J}(u_n^{(j)} - \bar{u}_n)\langle G(u_n^{(j)}) - \bar{w}_n), \cdot\rangle_Y, \tag{5.13}$$

then update the model ensemble $\{u_n^{(j)}\}$ with the ensemble of perturbed data $\{y_n^{(j)}\}_{j=1}^J$ by

$$u_{n+1}^{(j)} = u_n^{(j)} + C_n^{uw}(C_n^{ww} + \alpha_n\Gamma)^{-1}(y_n^{(j)} - w_n^{(j)}), \tag{5.14}$$

where α_n is heuristically chosen to be $\alpha_n = 2^i\alpha_0$, with α_0 an initial guess, such that $i \geq 0$ is the first integer with

$$\alpha_n||\Gamma^{1/2}(C_n^{ww} + \alpha_n\Gamma)^{-1}(y - \bar{w}_n)|| \geq \rho||\Gamma^{-1/2}(y - \bar{w}_n)||. \tag{5.15}$$

α_0, ρ and τ are tuning parameters of the scheme; typically $\alpha_0 = 2$ so that $\alpha_n \geq 1$—the choice of $\alpha_n = 2^i\alpha_0$ is a heuristic that tries to balance choosing as small as possible α_n without computing Equation 5.15 many times for each analysis step; theoretically any α_n that satisfies Equation 5.15 is acceptable, but this heuristic provides a good balance of computational effort without over-regularizing. Higher values of ρ provide greater regularization by forcing larger α_n; this typically also results in more iterations until the termination criterion is reached (Iglesias, 2016). If P is the dimension of the model space (potentially after discretization in the case where G operates on fields) and M is the number of observations, then C_n^{uw} is a $P \times M$ matrix and C_n^{ww} is a $M \times M$ matrix. For large data and model spaces, constructing and especially inverting these matricies can be very expensive—$O(M^3)$ for the construction of $(C_n^{ww} + \alpha_n\Gamma)^{-1}$. However, due to their construction, both covariance matrices are of rank at most $\min(J-1, M)$. Consequently, for $J \ll M$, it is more efficient to implement them within the algorithm as operators defined

by Equations 5.12 and 5.13. For constructing the inverse, we employ a low rank approximation of C_n^{ww} to compute the approximate Hermitian eigendecomposition of the operator, as it is symmetric positive semidefinite by construction (Halko et al., 2011). The low rank approximation is exact if an approximation of rank $J-1$ is sought. This decomposition allows us to write

$$C_n^{ww} = Q\Lambda Q^T, \tag{5.16}$$

where Λ is a square diagonal matrix of dimension at most $(J-1)\times(J-1)$ containing the largest eigenvalues of C^{ww} and Q, which is a M by at most $J-1$ matrix, has columns equal to the eigenvectors of C^{ww} corresponding to the elements of Λ. We may then use the Woodbury matrix identity to compute

$$(C_n^{ww}+\alpha_n\Gamma)^{-1} = (Q\Lambda Q^T+\alpha_n\Gamma)^{-1} = \frac{\Gamma^{-1}}{\alpha_n}-\frac{\Gamma^{-1}}{\alpha_n}Q(\Lambda^{-1}+Q^T\Lambda^{-1}Q)^{-1}Q^T\frac{\Gamma^{-1}}{\alpha_n}. \tag{5.17}$$

The matrix $(\Lambda^{-1}+Q^T\Lambda^{-1}Q)$ is of dimension at most $(J-1)\times(J-1)$, and all other inverses are of diagonal matricies. Constructing the Hermitian eigendecomposition requires only matrix-vector products (Halko et al., 2011); due to the structure of the covariance matrix C_n^{ww}, only J vector-vector products are actually required if we use C_n^{ww} in its operator form. This means that C_n^{ww} never needs to be explicitly constructed, which can result in significant memory savings for large data sets. The cost of constructing the eigendecomposition is amortized across the need to update J ensemble members. Consequently, using a low rank approximation and applying the Woodbury matrix identity can dramatically reduce the cost of updating the ensemble in both number of operations and memory. Figure 5.4 shows a schematic of the EKI algorithm applied to a two parameter linear inverse problem. Far from the optimum, ensemble members take scaled gradient descent steps as the regularization provided by $\alpha_n\Gamma$ dominates the dynamics. Closer to the optima, the ensemble becomes more aware of the curvature of the objective as the C^{ww} term dominates.

In the basic EKI algorithm, the final model u^{η} lies in the span of the initial ensemble. Appropriate choice of the initial ensemble therefore acts to encode prior information into the inverse problem. Within our model specification framework, the actual geological model $F(u)$ used to predict data by the forward solver is a nonlinear transform of the model parameter vector u defining the underlying Gaussian random fields and geometric parameters. This allows significant flexibility even when the underlying space of potential models is constrained to lie within a low-dimension subspace of the full space of models. Specifically, though the final parameter vector $\bar{u}_n$ describing the model specification is in the span of the initial ensemble $\{u_0^{(j)}\}_{j=1}^{J}$, the corresponding physical model $F(\bar{u})$ is not necessarily in the span of $\{F(u_0^{(j)})\}_{j=1}^{J}$.

The EKI algorithm offers several compelling benefits for the derivative-free solution of PDE constrained inverse problems. From a theoretical standpoint, the stable convergence to an approximate solution depending on the noise level is appealing. Furthermore, the scheme is practical, easy to implement and handles large parameter spaces. In particular, the calculation of the forward models $G(u_n^{(j)})$ and the updates of models $u_{n+1}^{(j)}$ have no interaction between ensemble members. Consequently, these parts of the algorithm are embarrassingly parallelizable and scale trivially to meet available computational resources (Herlihy and Shavit, 2011). Since the forward model calculations are typically the most expensive part of the algorithm, this is a particularly useful property. Finally, an important consideration for practical employment of the algorithm is that it allows black-box forward models, such as legacy or proprietary closed-source codes for which derivatives of the misfit function with respect to model parameters are not available, to be used without expensive explicit finite differencing. Compared to obtaining the gradient from the adjoint method, the EKI method uses J forward solves for every step, compared to J_a for an adjoint method, so that the ratio of computational effort is J/J_a if the solution of model updates is negligible in cost. J_a depends on the forward model but is

typically 2-3, while the optimum J depends on the problem but is typically larger. Despite this, as the J ensemble members are independent, the ensemble method is particularly amenable to distributed computing even if communication between processes has high latency. Additionally, for non-self-adjoint forward solvers, the EKI algorithm does not utilize a backwards pass and so does not require complex checkpointing schemes for managing storage requirements (e.g. Komatitsch et al. (2016)), which may be a useful property for some problems.

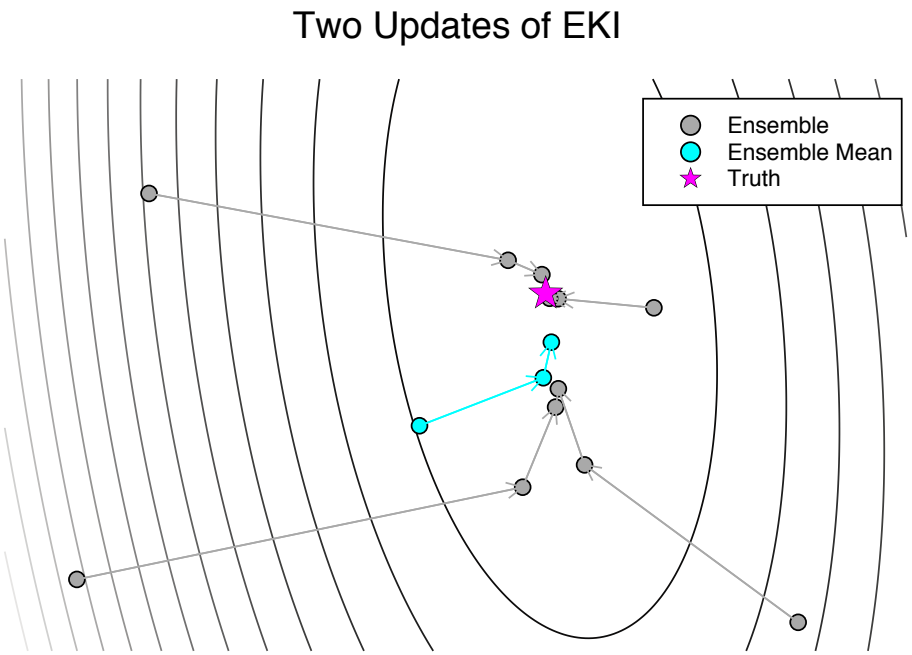


Figure 5.4: Two updates of the EKI alogrithm with four ensemble members for a toy linear objective with two parameters. Elliptical lines show the contours of the objective function.

Inversion Framework Summary

The inversion framework presented in this study consists of a geologically motivated parametrization of the Earth, coupled to an efficient, highly parallelizable and derivative-free solver. Framing geophysical inverse problems as a question of optimizing geological models allows for direct interpretation of the resulting im-

ages, and allows practitioners to compare structurally different models against each other. The parametrization scheme described above has the flexibility to describe models ranging from simple 1D descriptions to fully 3D, heterogeneous models with structural discontinuities in a consistent format. We have shown that by using the implicit level-set method to define geological domains allows the topology of a model to change to fit the data without changing the parametrization, in contrast to explicit definitions of domains in which the parameters must be added or removed to describe changes in topology, significantly complexifying the inverse problem. Our parametrization framework motivates using a derivative-free optimizer because the resulting models are not necessarily efficiently differentiable, and because a goal of this study is to modularize the inverse problem so that the structure of the model is not tied directly to the forward solver. We employ EKI as the optimizer, as it scales well with computational resources, treats the forward solver as a black box, and incorporates iterative regularization to avoid overfitting the data. Utilizing the inherent low-rank structure of the covariance matricies used by EKI allows even large data sets to be handled efficiently. As an iteratively regularized algorithm, EKI does not include explicit Tikhonov damping and will fit the data to within an assumed noise level, without the biases introduced by these terms. The tuning parameters in the EKI scheme instead control the stability of the convergence and the convergence rate.

5.5 Examples

To illustrate the combination of our model parametrization framework and the EKI inversion scheme, we show two synthetic seismic tomography examples and one example using real active source seismic data collected at Carrizo Plains, CA. In all cases, the data are first arrivals of P-waves from known source locations, as is typical in an active source seismic experiment. We have chosen this relatively

simple forward model to concentrate on the details of the model specification and the inversion method. We note however, that both the model specification framework and the EKI solver are independent of the choice of foward model and are not limited to seismic travel-time tomography; for instance joint inversions incorporating potential methods such as gravity could be used, or full seismic waveforms could be used—noting that for full waveform methods the model specification must be very close to the truth or it is likely that the inverse problem will converge to an unrealistic local minimum.

To calculate the arrival times through the model, we solve the eikonal equation using the fast marching method (Osher and Sethian, 1988; Rawlinson and Sambridge, 2004). We employ the factored form of the eikonal equation, accounting for the singularity at the source analytically, resulting in significantly improved travel time calculations along grid diagonals relative to the basic eikonal method (Treister and Haber, 2016). The first example illustrates the advantages of the GRF level-set definition for describing geological domains; the second example shows how our model specification can compose geological objects and deformations; the third example shows that our method is robust for real data and highlights the useful iteratively-regularizing properties of the EKI scheme. In the examples that follow, we take GRFs with fixed length scales for simplicity; solving for GRF length scales may be achieved during the inversion by hierarchical EKI (Chada et al., 2018).

Shape Recovery in First Arrival Crosswell Tomography

Our first example is an application of the level set method with GRF priors to invert first arrival data in a crosswell geometry using the EKI algorithm—as such, this example uses only a subset of the model description framework described in Section 5.3. The purpose of this example is to illustrate the advantages of implicitly defining boundaries via the level set method. This type of shape optimization

problem may be alternatively solved using the level set evolution equation (Li et al., 2014), however, the GRF based formulation used in this study imposes additional *a priori* constraints on the inverse problem—this example shows that shape recovery is still possible under these constraints. We synthesize data from 9 sources in a vertical well with 16m spacing. We record data in a vertical well 96m away, with sensors spaced at 4m, and assuming a nominal data picking error of 0.25 ms. We hypothesize a background model of 1000m/s velocity, with fast inclusions of 1500m/s. The geometry of the true input model is shown in Figure 5.5a.

We assume that we have no knowledge of the number or geometry fast inclusions, while the velocities are known. This makes explicit parametrization of their locations and shapes difficult, as some heuristic must be used to determine the appropriate number and topologies of boundaries. To overcome this issue, we generate an initial ensemble of 200 candidate models using GRF defined level sets, containing a wide range of inclusion topologies—four examples of the starting ensemble are shown in Figure 5.5b, from which we can confirm that the initial ensemble is not strongly tuned to reflect the true input model. The ensemble was generated using a zero-mean Matérn GRF with $\rho = 50$m and $\nu = 1.5$; the choice of a Matérn GRF with $\nu < 2$ is motivated by a desire to have solutions with somewhat rough boundaries. Therefore, for this example, the parameter vector u consists of the values of the latent field, initially drawn from the Matérn distribution, and the model function F is the level set operator assigning values of the latent field that are greater than 0 to 1500m/s, and those below 0 to 1000m/s. The forward operator H is the solution of the factored eikonal equation from the sources to the receivers, and as usual the full forward map may be written as $G(u) = H(F(u))$.

We evolve the initial ensemble using the EKI algorithm using $\rho = 0.75$ and $\tau = 1.6$ until the discrepancy principle termination criterion is satisfied after 40 iterations. The output model and fits to the data are shown in Figure 5.5cd. We see that the

location and approximate geometries of the three inclusions are recovered, and that the data is well fit by the predicted model.

In this example, regularization is provided by the underlying structure of the GRF used to generate the ensemble. In particular, the wavelength parameter ρ of the Matérn covariance was chosen to be comparable to the size of the inversion domain, which suppresses short wavelength structure. The EKI algorithm, as presented in this paper, produces model parameter outputs in the linear span of the initial ensemble, which has the effect of maintaining the GRF structure throughout the iterations of the inversion. The level-set function acts as a nonlinear activation function, allowing the GRF function to produce the shorter lengthscale features required by the data, even though the GRF lengthscale is significantly longer. Setting the GRF lengthscale to be large avoids the introduction of small anomalous features in the final result. Additionally, the discrepancy principle used to terminate EKI serves to avoid overfitting the data; the inversion starts with smooth members of the starting model and evolves greater complexity, stopping immediately once a fit to the data is achieved.

Determining Surface Fill Depth with a Fault

In many geological settings, there may be stronger *a priori* knowledge of potential structures that can be employed in an inversion. A typical example of this would be the inferred presence of faults derived from observed seismicity, surficial rupture or other geological constraints. In this example, we simulate first-break seismic refraction data for a smoothly varying interface between two layers bisected by a vertical fault with some offset—for example, this could be a profile perpendicular to a strike-slip fault with unconsolidated alluvial surface cover. We assume sources spaced every 30m and receivers spaced every 5m along a profile 240m in length. Data were perturbed with 1ms Gaussian noise to simulate picking error estimated

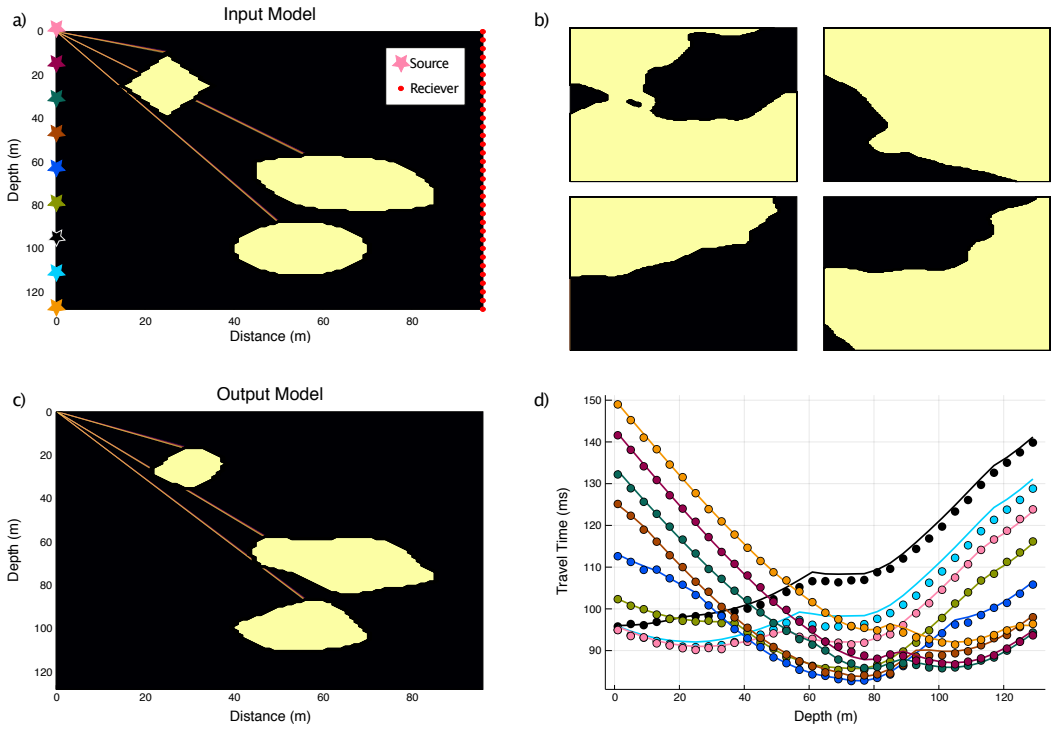


Figure 5.5: Use of Gaussian random field (GRF) level sets for a crosswell tomography boundary identification problem. a) shows the true input model, with source and reciever geometry. Yellow regions are 1500m/s, black 1000m/s. b) shows 4 examples of the initial ensemble of models used for EKI. c) shows the output model, and d) shows the data and fit, with colors corresponding to the source colors in a).

from a real data experiment with equivalent geometry (Example 5.5). The true model is shown in Figure 5.6a. A simple stationary GRF-based level set approach cannot easily represent this kind of model since the smooth covariance structure will suppress the fault, acting similarly to a Tikhonov smoothing regularization. Instead, we explicitly add in the presence of a potential fault in our model description. This has the additional advantage that the parameters related to the fault (e.g., position, dip angle, offset) are immediately physically interpretable. This type of inversion therefore represents a combination of level set inversion and minimum-parameter modelling in the style of Zelt and Smith (1992). The objective of the inverse problem is then to calculate the optimal parameter vector u, which is made up of the GRF latent field describing the interface, the explicit geometrical parametrization

of the fault, which consists of the horizontal location of the fault plane and the amount of vertical offset, and the velocities of the two layers. The forward map can again be written $G(u) = H(F(u))$ where F transforms u into the physical model of interest evalutated on a Cartesian solver grid, and H solves the factored eikonal equation. We use EKI with 256 ensemble members to solve the resulting inverse problem, which converged in 28 iterations. Once again, we show 4 examples of the initial ensemble to illustrate the range of potentially allowable geometries in Figure 5.6b. The final inverted model is shown in Figure 5.6c, together with a comparison in Figure 5.6d to a traditional ray-tracing based tomography performed using the commercial DWTomo Software, which explicitly considers topography and creates a smoothed regularized solution (Geogiga Technology Corporation, 2016).

Without *a priori* knowledge of the expected structures, the traditional refraction tomography smooths the vertical interface and has approximate vertical and horizontal resolution of $\sim$ 10m, controlled by the regularization and data quality, as can be seen in Figure 5.6d. Additionally, the L2 regularization used in the tradition tomography promotes a smooth transition from low to high velocity. Assuming we have appropriate knowledge, our level set / geometric parametrization can much better recover the true model. In this case, appropriate knowledge could be prior mapping of a surface rupture of the fault. The question of whether an explicitly layered model such as this is more appropriate than a smooth model requires assessment of the data, as well as any appropriate geologic knowledge at hand.

Near-Surface Refraction Tomography of the San Andreas Fault at Carrizo Plains

For a final example, we apply our inversion scheme to real seismic refraction data collected on March 20, 2017 at Carrizo Plains, California, USA. Reconstruction of paleoseismicity of the San Andreas Fault (SAF) at Carrizo Plains suggests regular

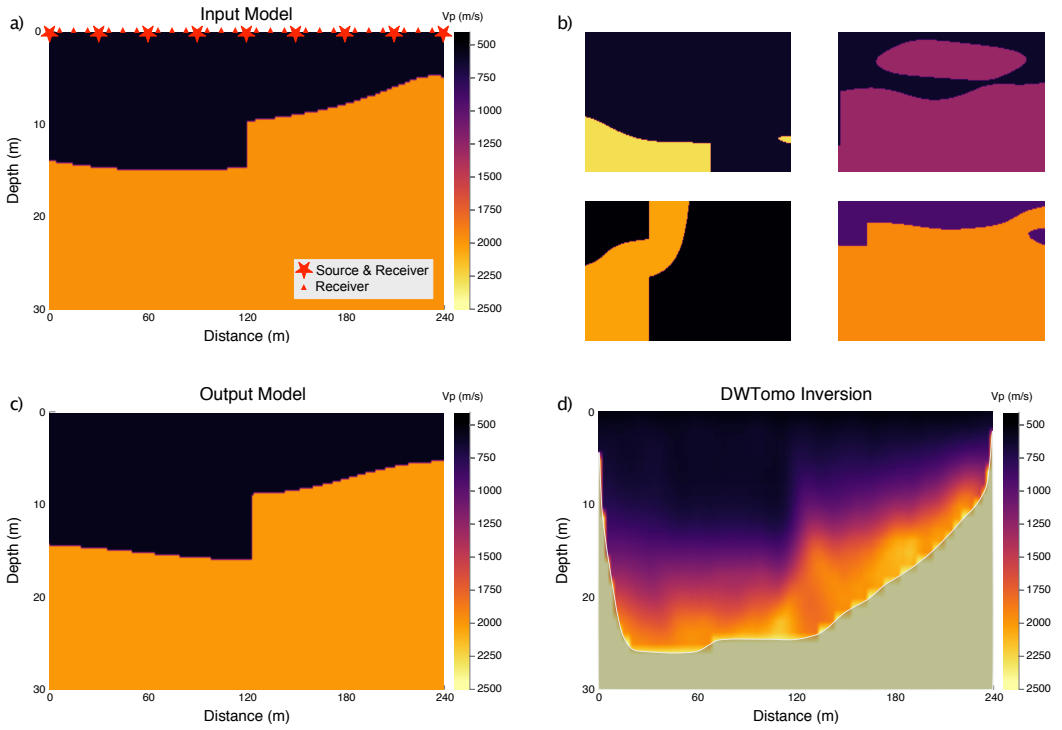


Figure 5.6: Illustration of using a combination of level sets and explicit geometric parametrizations to recover a subsurface interface offset by a vertical fault. a) shows the true input model with source / receiver geometry. b) shows 4 examples of the initial ensemble of models used for EKI. c) shows the output of the inversion. d) shows the traditional inversion using DWTomo; the opaque grey mask shows the boundary of the rays calculated by DWTomo.

slip of up to ~ 5m (Ludwig et al., 2010; Zielke et al., 2010), with trenching implying a potentially > 10m wide band of multiple near-surface fault strands that are likely to be seen as a low velocity damage zone in tomographic images (Akciz et al., 2009). Data were collected along a profile of length 240m, oriented SW to NE, with significant topography, using a 48-channel geophone array. The profile is roughly bisected by the SAF, which can be prominently seen in Figure 5.7a, especially noting the significant stream channel offset near the center of the image. Remington Industrial 8-gauge charges buried approximately 0.25 m deep generated the active sources at 0, 60, 120, 180 and 235m along the profile. First arrival times were then handpicked. We consider the data noise to include picking and triggering errors, imprecision in the source and receiver locations, and errors in the recorded

surface topography. The true noise distribution is consequently unknown; for this application we will assume data is independent, identically Gaussian distributed with equal variance.

The purpose of this example is primarily to show that the combination of the level set formulation and EKI is practical and stable when applied to real data and to compare it against a traditional tomographic image. Inspection of the data suggests a three layered model. Consequently, we choose to invert for a model vector u that consists of two 1D GRFs describing layer interfaces and the constant velocities of the three layers. The model function F computes level sets from the GRFs and assigns velocities to the resulting regions in physical space. The explicit topography derived from the known locations of the receivers is included in F by linear spline interpolation. In this case, we chose to employ 1D GRFs to initialize the ensemble for a 2D model inversion, to avoid overlapping folds in the boundaries between regions. Similarly, in a 3D setting, 2D GRFs may be used to introduce a layered structure with no folds. This type of problem could potentially be solved using a multiple level set evolution equation method such as that in Li et al. (2017), however as shown in section 5.9 the combinatorial complexity of these methods greatly increases with the number of layers, and our GRF formulation provides intrinsic regularity to the solutions which motivates the use of our model specification framework. As previous trenching evidence suggested that the fault was likely to be observed as a distributed damage zone at the lengthscale of this study, we did not employ any deformation layers in our model description. We chose Matérn GRFs with $\rho = 100$m, $\nu = 1.5$ and $\sigma = 5$m. The *a priori* mean depth of the first layer used to generate the EKI ensemble was set to be uniform across the depth range of the model, with the mean depth of the second layer set to be uniformly generated between 0-20m below the first layer. To test the stable convergence properties of EKI, we inverted the data assuming nominal noise standard deviations σ of 6, 4 and 3 ms, with the resulting

models shown in Figure 5.7d,f,h respectively. Note that we estimated a picking error of approximately 1ms from the data, but expect to see significant modelling error from source/receiver geometry errors and modelling errors. Solution of the inverse problem employed 128 ensemble members, and required 7 iterations to reach the 3 ms noise level. Data and fits are shown in Figure 5.7c,e,g. Together, these show that as the assumed noise level is lowered, the data is progressively better fit and the model becomes progressively more featured, without developing obvious artifacts related to lack of sufficient regularization. A traditional tomographic reconstruction (again using DWTomo) is shown in Figure 5.7b and exhibits similar qualitative behavior to the 3ms level set / EKI result, with a slow surface layer with similar undulations and a steep step up across the SAF of a faster third layer.

These results show that our parametrization and optimization scheme is sufficiently robust to apply to real inverse problems. At the 3-ms noise level, all significant features of the data are captured even by the relatively simple 3 layered model proposed here. Lowering the assumed noise level does not significantly qualitatively change the models, but instead sharpens features, especially the primary feature of the step in the fastest velocity across the fault. The initial ensemble of models for all noise levels have on average flat interfaces across the layer boundaries—the progression in Figure 5.7d,f,h illustrates a key property of the iteratively EKI algorithm, in that it evolves the ensemble away from the typically smooth "prior" towards a more featured final model. At higher noise levels, this progression is terminated earlier, and so the ensemble will look more like the smooth prior; hence Figure 5.7d has smoother and flatter interfaces than Figures 5.7f and h, in which the evolution of the ensemble progresses further away from the prior. As in any iterative tomographic method, the starting model, or in this case starting ensemble, has an important impact on the final result when the data is noisy, but becomes progressively less important as the inversion is constrained to closely fit the data; a substantial

difference to traditional tomographic methods is that the final model produced by EKI lies in the linear span of the starting ensemble, so that for implicit GRF parametrizations the covariance structure is maintained throughout the inversion. This may or may not be a desired property of the inversion; if the initial ensemble encodes a model appropriate for the data then the linear span property ensures that the final model reflects the initial ensemble. Alternatively, if greater flexibility is required due to less strong *a priori* constraints on the model, then a hierarchical generalization of EKI may be employed in which hyperparameters are optimized for the fundamental properties of the parametrization, such as the length scales ρ used for GRFs (Chada et al., 2018). As our focus in this study is setting up a general modelling framework, we have chosen not to investigate these generalizations in this paper, however they offer an intriguing extension for situations in which *a priori* information is relatively lacking.

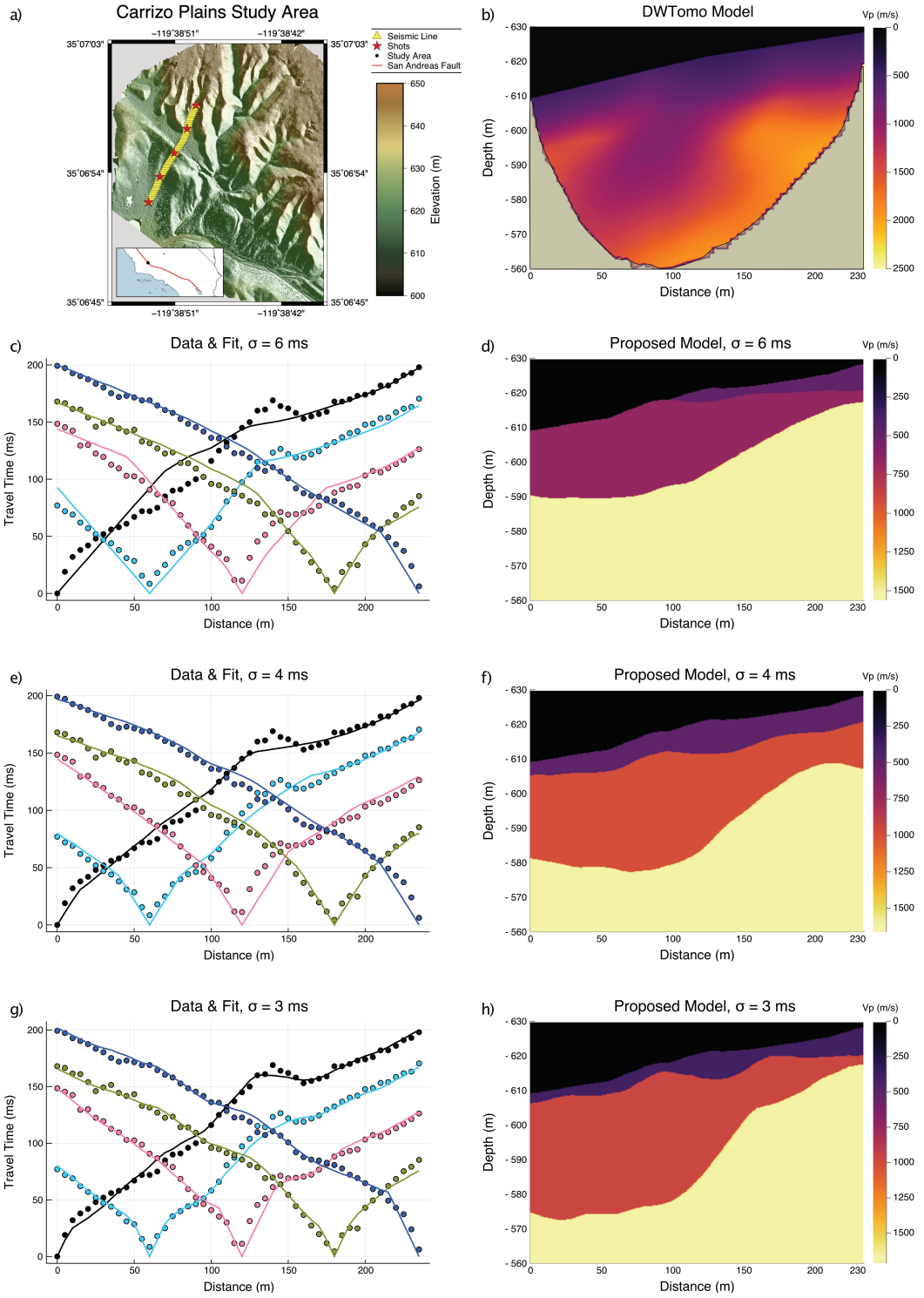


Figure 5.7: Three layer inversion of near surface velocity adjacent to the San Andreas Fault at Carrizo Plains, illustrating the consistent convergence properties of the iteratively-regularized EKI scheme. Black regions of the tomographic images are not inverted, and correspond to air. a) shows the study area and source / receiver geometry. b) shows the traditional inversion using DWTomo; the opaque grey mask shows the boundary of the rays calculated by DWTomo. c), e) and g) show the data and fits for assumed data noise $\sigma = 6, 4, 3$ ms respectively, and d), f) and h) show the corresponding 3-layer inversion models using our scheme.

5.6 Discussion and Conclusions

The objective of this study has been to develop a framework for encoding geological information into geophysical inverse problems in an intuitive way. Using the EKI algorithm, the computational difficulties of taking derivatives of our models are avoided, enabling our definitions to be used to solve large-scale inverse problems defined by nonlinear, possibly black-box forward models. Using our inversion framework we solved three example inverse problems using the P-wave first arrival travel-time problem as a test case. In these examples, the level-set model specification enabled complex boundaries to be inverted using only the *a priori* knowledge of the expected number of domains. Furthermore, we showed how we can incorporate useful *a priori* information, such as the presence of faults, to deliver a yet more parsimonious model that has significantly better resolution than traditional tomographic approaches. We have illustrated how using our inversion framework appropriately may result in tomographic images that are easier to interpret than traditional images produced by standard methods; the practitioner should be empowered to formulate descriptive models that enable targeted exploration of the data. For concenceptual clarity, this study has used examples for which heterogeneity within model regions is sufficiently weak that it is reasonable to assume constant velocity models. However, the model framework permits arbitrary structure within each model layer, which could be modeled with a GRF with no level set function applied. For regions in which there is substantial in-layer heterogeneity, explicit modelling of material interfaces may still be useful when there is a mixture of sharp transitions and smooth variations in mechanical properties which are both relevant to the inverted data - for example, when jointly inverting high-frequency receiver functions with surface wave dispersion for sedimentary basin geometry and internal velocity structure.

One important outcome of the framework not presented in this study is the possibility of formal model selection performed on geologically parameterized models. In the

context of model selection, a practitioner would propose several distinct geological models and then use some criterion to rank the models in a preferred order by balancing their complexity against their ability to predict the data. For the deterministic inverse problems solved in this study, which produce a single optimum model that best fits the model given observed data and prior constraints, various information criterion (IC) such as the Akaike or Bayesian IC may be used (Claeskens, 2016). If computational resources permit, cross-validation techniques are possible and act as a guard against outlier data (Claeskens, 2016). If a Bayesian approach is taken, more robust approaches include predictive performance checks using draws from the posterior predictive distribution of the data—methods such as PSIS-LOO using these draws can emulate leave-one-out cross validation without explicitly resampling the posterior conditioned on subsets of the data (Vehtari et al., 2017). Finally, explicit Bayes factor estimation may be tractable for lower dimensionality models where the practitioner is confident in the priors assigned to the model (Weinberg, 2012). As our experimental evidence shows, even relatively simple models of the Earth can match complex data to within a realistic noise level. It is therefore up to the domain expertise of the inversion practitioner to design candidate models such that any model selection is meaningful. Once appropriate geophysical models are identified from *a priori* knowledge, our study provides a framework by which the models can be defined and optimized to fit the data. The inversion philosophy promoted by this work is more investigative than exploratory when compared to traditional geophysical inversion procedures that typically emphasize removing *a priori* information as much as possible from the inverse problem. As our understanding of the Earth and its structures grows, we believe that methodologies, such as the one presented here, that are driven by our *a priori* knowledge will become increasingly important to ameliorate the fundamental issue of non-uniqueness in geophysical inverse problems.

5.7 Acknowledgements & Funding Sources

The authors would like to thank Prof. Nick Rawlinson and an anonymous reviewer for providing useful commentary that has significantly improved the quality of the manuscript. We would also like to thank Editor Prof. Michael Ritzwoller and the anonymous assistant editor for managing the review process. JBM would like to thank Andrew Stuart (Caltech Computational and Mathematical Sciences) and the 2018 Gene Golub SIAM summer school for useful discussions regarding this study. Data from Carrizo Plains was collected during the 2017 Caltech Applied Geophysics Field Course, for which JBM was a Teaching Assistant. JBM would like to thank the instructors Rob Clayton and Mark Simons, and co-TA Voon Hui Lai, as well as the students, for the course. JBM would also like to thank the General Sir John Monash Foundation and the Origin Energy Foundation for financial support. This study was supported by NSF grant EAR-1453263.

5.8 Resources

All calculations were computed using the Julia language (Bezanson et al., 2017). Code for our model specification language can be found at `https://github.com/jbmuir/EarthModels.jl`. Code for the EKI optimizer can be found at `https://github.com/jbmuir/EnsembleKalmanInversion.jl`. Code for a Julia 1.0+ compliant factored Eikonal fast marching method forward solver can be found at `https://github.com/jbmuir/FEFMM.jl`.

5.9 Appendix

Combinatorial Definition of Multiple Level Sets

For N regional parameter fields of interest, set n such that $N = 2^n$. If N is not a power of 2, we can arbitrarily split regions until we can meet this condition; this will generate a “boundary" without a discontinuity across it. We define the Heaviside

step operator to be

$$H(\phi)(x) = \begin{cases} 1, & \phi(x) > 0 \\ 0, & \phi(x) \leq 0 \end{cases} \tag{5.18}$$

Then let $\iota(i, j)$ be the i^{th} digit of the binary representation of $j - 1$. Then

$$F(x) = \sum_{j=1}^{N} \prod_{i=1}^{n} A_j(x)(\iota(i, j)(1 - H(\phi_i)(x)) + (1 - \iota(i, j))H(\phi_i)(x)) \tag{5.19}$$

This definition is differentiable and potentially requires fewer auxiliary fields than the procedural definition. However, due to the combinatoric nature of the formula differentiation becomes difficult in practice for $n > 2$, and regularization of the inverse problem may result in cross-talk between different regions which share some of the same auxiliary fields—auxiliary fields are not individually associated with regional parameter fields.

Derivation of Explicit Levenberg-Marquardt Update in Finite Dimensions

In finite dimensions, Γ and C are symmetric positive-definite matrices. For compactness, let the prediction error at u_n be $y - G(u_n) = \delta y_n$ We start with the LM update rule in finite dimensions

$$u_{n+1} = u + v^*, \tag{5.20}$$

$$v^* = \arg\min_{v \in \mathbb{R}^M} \left(||\Gamma^{-1/2}(\delta y_n - J(u_n)v)||^2_{\mathbb{R}^M} + \alpha_n ||C^{-1/2}v||^2_{\mathbb{R}^N} \right), \tag{5.21}$$

$$v^* = \arg\min_{v \in \mathbb{R}^M} \left((\delta y_n - J(u_n)v)^T \Gamma^{-1}(\delta y_n - J(u_n)v) + \alpha_n v^T C^{-1} v \right). \tag{5.22}$$

The condition for v^* is that the derivative of the right hand side equals 0, which gives

$$\frac{\partial \left((\delta y_n - J(u_n)v)^T \Gamma^{-1}(\delta y_n - J(u_n)v) + \alpha_n v^T C^{-1} v\right)}{\partial v} \tag{5.23}$$

$$= \frac{\partial(\delta y_n - J(u_n)v)^T \Gamma^{-1}(\delta y_n - J(u_n)v)}{\partial(\delta y_n - J(u_n)v)} \frac{\partial(\delta y_n - J(u_n)v)}{\partial v} + \frac{\partial \alpha_n v^T C^{-1} v}{\partial v} \tag{5.24}$$

$$= -(\delta y_n - J(u_n)v)^T \Gamma^{-1} J(u_n) + \alpha_n v^T C^{-1} \tag{5.25}$$

$$= v^T (J(u_n)^T \Gamma^{-1} J(u_n) + \alpha_n C^{-1}) - \delta y_n^T \Gamma^{-1} J(u_n) \tag{5.26}$$

$$= 0, \tag{5.27}$$

or on taking transposes while noting Γ^{-1} and C^{-1} are both symmetric

$$(J(u_n)^T \Gamma^{-1} J(u_n) + \alpha_n C^{-1})v = J(u_n)^T \Gamma^{-1} \delta y_n, \tag{5.28}$$

which gives the usual explicit LM update

$$v^* = (J(u_n)^T \Gamma^{-1} J(u_n) + \alpha_n C^{-1})^{-1} J(u_n)^T \Gamma^{-1}(y - G(u_n)). \tag{5.29}$$

Chapter 6

PARSIMONIOUS VELOCITY INVERSION APPLIED TO THE LOS ANGELES BASIN, CA

In review as

Muir, J. B., R. W. Clayton, V. C. Tsai & Q. Brissaud "Parsimonious velocity inversion applied to the Los Angeles Basin, CA". In *Journal of Geophysical Research: Solid Earth*, Submitted

6.1 Abstract

The increasing proliferation of dense arrays promises to greatly improve our ability to image geological structures at the scales necessary for accurate assessment of seismic hazard. However, effectively combining the resulting local high-resolution tomographic results with existing regional models presents an ongoing challenge. We have recently developed a framework based on the Tikhonov-regularized level-set method that provides a simple means to infer where local data provides meaningful constraints on seismic observables beyond those found in larger scale regional models—for example the Community Velocity Models (CVMs) of southern California. This technique defines a volume within which updates are made to a reference CVM, with the boundary of the volume being part of the inversion rather than explicitly defined. By appropriately penalizing the complexity of the boundary, a minimal update that sufficiently explains the data is achieved.

To test this framework, we utilize the high-resolution data from the Community Seismic Network, a large 400-station permanent urban deployment. We inverted Love wave dispersion, derived from eikonal tomography of two-station cross-correlation travel-time delays, and relative amplification data, from the Mw 6.4 July 3 and Mw

7.1 July 5 2019 Ridgecrest earthquakes. We invert for an update to CVM-S4.26 using the Tikhonov Ensemble Sampling scheme, a highly efficient derivative-free approximate Bayesian method. We find that the Ridgecrest Earthquake data is best explained by a deepening of the Los Angeles Basin (compared to the CVM-S4.26 reference model) with its deepest part just south of downtown Los Angeles, along with a substantially steeper northeastern wall of the basin. This result offers new progress towards the parsimonious incorporation of detailed local basin models within regional reference models utilizing an objective inverse-problem framework, and highlights the importance of accurate basin geometry models when accounting for the potentially significant amplification of surface waves from regional earthquakes in the high-rise building response band.

Plain Language Summary

Los Angeles is a major city of the United States that is at high risk of damage due to earthquakes, due to the large number of nearby active faults and its location on a deep bowl of weak rock, which tends to amplify earthquake damage. We use a large number of instruments located in Los Angeles district schools to make measurements of earthquakes that occurred near Ridgecrest, California in July 2019. These earthquakes generated a type of energy that is particularly useful for studying the structures responsible for amplification of earthquakes. Using this data, we applied a new imaging technique to create a local model of the northeast Los Angeles basin at higher resolution than had been previously available. Our imaging technique appropriately balances information from previous, lower resolution inversions with the new data obtained in this study.

6.2 Introduction

The Los Angeles (LA) Basin is a deep sedimentary structure whose evolution can be roughly characterized by an initial subsidence and extensional phase during the

establishment of the North America - Pacific plate boundary associated with the opening of the Gulf of California and the rotation of the Transverse Ranges in the Miocene. This was followed by a period of transpression (Ingersoll and Rumelhart, 1999), and the generation of a substantial network of thrust faults within the basin (Wright, 1991). In its current state, the basin contains both active strike-slip faults (e.g. the Newport-Inglewood fault, Whittier-Elsinore fault) and an imbricated stack of blind thrust faults (e.g. the Elysian Park faults, Puente hills thrust), all of which accommodate the transpressional motion of the basin. These faults contribute to local seismic hazard both by providing source surfaces for earthquakes and by controlling local path effects by shaping the basin geometry (Plesch et al., 2007). The evolutionary history of the LA basin, with ample opportunity to produce and bury organic material during extension followed by the estabilishment of stratigraphic traps during compression, allowed LA to be a leading producer of oil in the United States (US), helping to fuel a large rise in population during the mid-20^{th} century. Development took place predominantly on the soft sediments of the main LA, San Fernando, San Gabriel and San Bernadino basins. As a consequence, LA is both one of the largest and most economically important cities in the US, while also being one of the most exposed to significant earthquake hazard due to the complex fabric of active faults and ground-motion amplifying sedimentary structures associated with the geology that has allowed its preeminence.

Seismic hazard within the basin is controlled by the locations and potential for slip on the multiple local and regional faults of southern California, combined with the significant amplifying effect of the basin on ground motions. The importance of path effects, such as wavefield focusing, multipathing, and basin amplification, on LA basin ground motions has led to extensive development of seismic velocity models. The ultimate goal of these models is to produce accurate synthetic waveforms at frequency ranges relevant to infrastructure and building codes within the

basin. Early efforts focused on creating rule-based models of southern California (Magistrale et al., 2000, 1996) using empirically derived velocity laws (Faust, 1951) in combination with inferred geological structure obtained by correlating surface outcrops, borehole profiles and potential methods (Wright, 1991). Since these inital efforts, regional scale models of southern California have assimilated ever greater quantities of seismic data, including seismic reflection profiles, receiver functions, and earthquake source locations and mechanisms, in an effort to better demarcate boundaries, including faults (Magistrale et al., 2000; Plesch et al., 2007), and allowed for more lateral variation of within basin velocity structures by using geostatistical methods to tie together disparate seismic data (Shaw et al., 2015; Süss and Shaw, 2003). Continued development of seismic velocity models of southern California has resulted in two widely used reference Community Velocity Models (CVMs), CVM-S4.26.M01 (Lee et al. (2014), CVM-S hereafter) and CVM-H 15.1.0 (Shaw et al. (2015), CVM-H hereafter), that have incorporated waveform based seismic tomography to further refine the models. CVM-S and CVM-H broadly agree in the positions, average velocity profile, and geometry of the major basins of southern California, however in detail they are quite different, with CVM-H containing more explicit geological information. Figure 6.1 shows a characteristic cross-section of the LA basin for both models, running from Catalina Island, across the Inner Borderland to Palos Verdes, then through the main LA basin, San Gabriel basin and though the transverse ranges to the high desert. This profile makes evident the considerably higher detail present in the CVM-H model due to its construction including explicit geological features (notably including an Inner Borderland basin not present in CVM-S), as well as its significant artefacts associated with changing lateral resolution, as evident in profile marks R1 and R2. In contrast, CVM-S is significantly smoother than CVM-H due to its reliance on wavefield-tomography during the final stages of construction, although several sharp resolution based artefacts are also

evident. While many features of the seismic wavefield within the LA basin, such as phase arrival times and P-to-S amplitude ratios, are captured for local events at frequencies of up to 0.2 Hz (Lai et al., 2020; Taborda et al., 2016), excitations of the basin from the recent large regional Ridgecrest earthquake sequence in July 2019 have illustrated that ground motion amplification predictions from finite-difference wave propagation through the SCEC CVM-H and CVM-S models do not accurately model the observations even at the relatively low frequency 0.1-1Hz range that is relevant for tall buildings within downtown LA (Filippitzis et al., 2021), warranting continued close study of the LA basin velocity model.

Seismic tomography offers the best opportunity for full spatial coverage of the basin at high resolution, especially when dense seismic arrays are utilized. In the southern and central parts of the basin, the deployment of high-density temporary seismic arrays using 10Hz corner-frequency geophone nodes by the petroleum industry has enabled considerable exploration of the shallow structure of the basin using ambient-noise derived observables, such as Rayleigh-wave phase velocities, Rayleigh-wave amplifications, and body-wave travel times (e.g. Bowden et al. (2015); Castellanos et al. (2020); Jia and Clayton (2021); Lin et al. (2013)). However, similarly dense industry deployments have not to date taken place in the northern part of the basin, which encompasses the downtown LA region, with buildings that are highly susceptible to resonant coupling to the basin. The permanent broadband southern California Seismic Network (SCSN), while providing a long time series of excellent quality observations, has already been incorporated into the CVM reference models and does not provide the spatial resolution required for the next generation of basin models. A potential alternative data source is the Community Seismic Network (CSN, Clayton et al. (2012, 2020)), a permanent network of three-component micro-electromechanical system (MEMS) accelerometers, designed to provide real-time strong-ground-motion telemetry in the event of local earthquakes within the

LA basin. The CSN instruments have been designed for inexpensive construction, utilizing off-the-shelf components, and have a maximum observable acceleration of $\pm 2g$, in order to fulfil their primary goal of strong-ground-motion monitoring. As a result, the instrument noise floor is above the amplitude of ground motions produced by smaller regional earthquakes, and is also above the ambient seismic noise level, which precludes the use of ambient-noise cross-correlation methods on CSN data as these methods rely on coherent low-level energy propagation between sensors. However, both the Mw 6.4 and Mw 7.1 2019 Ridgecrest, California earthquakes produced high quality records across the array, allowing for detailed analysis of ground amplification within the basin (Filippitzis et al., 2021; Kohler et al., 2020). The coherent surface-wave energy from these two events, recorded on the CSN, offers a unique opportunity to construct a high-resolution local tomographic model of the northeastern edge of the LA basin. In this study, we use the phase velocity and relative amplitudes of Love waves from both events, along with a 3D surface-wave tomography method based on the level-set method of Muir and Tsai (2020b), to create such a model. The level-set framework extends traditional tomography by allowing for discontinuous interfaces within a velocity model, which are implicitly defined by a contour line of a latent function. For instance, Muir and Tsai (2020b) used the level-set method to image the damage zone of the San Andreas Fault at Carrizo plains using a an implicit three-layer model, while Tso et al. (2021) presented several applications of the level-set method for developing interpretable block models of electrical resistivity. The ability to handle implicitly defined discontinuities significantly extends traditional tomography, which usually require restrictive and unphysical regularization schemes to be well-posed. We use the level-set method to define a basin volume within which we update a local model—this method allows us to only alter the reference CVM model where we have sufficient data constraints to warrant an update. Integration of local models within the SCEC CVM ecosys-

tem will become an important part of hazard modelling within Southern California as high-density arrays allow access to the fine scale detail of path effects. The framework presented in this study represents a parsimonious way to achieve this integration.

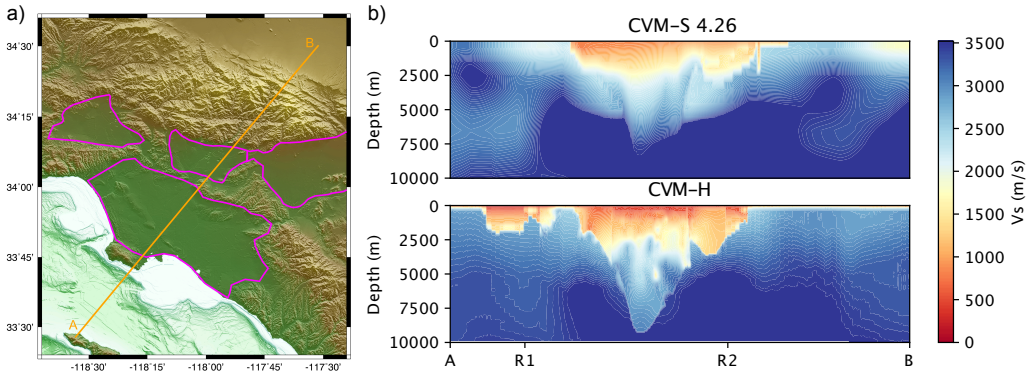


Figure 6.1: a) Shaded elevation model of southern California showing the outline of the major basins (defined by slope-break analysis) in purple and the transect A-B used for profiles shown in orange. b) Characteristic profiles through the Los Angeles basin for the CVM-S and CVM-H models. Abrupt lateral changes in resolution at positions R1 and R2 are seen in the CVM-H model.

6.3 Data Collection

Preprocessing

The data for this study were obtained from the HN accelerometer channels of the Los Angeles Unified School District (LAUSD) subarray of the Community Seismic Network (CSN, Clayton et al. (2012, 2020)), consisting of 200s time series after the Mw 6.4 and Mw 7.1 Ridgecrest earthquakes' origin times and recorded at 50 samples/sec. The network is deployed within school buildings in the City of Los Angeles, and at the time of the Ridgecrest earthquakes consisted of 300 stations spaced approximately 0.5 km apart. We used the components of the CSN located within the northeast LA basin, which is the densest part of the array—the study area, including the locations of the stations, is shown in Figure 6.2. Various display of the Ridgecrest earthquake data are shown in Filippitzis et al. (2021), along with

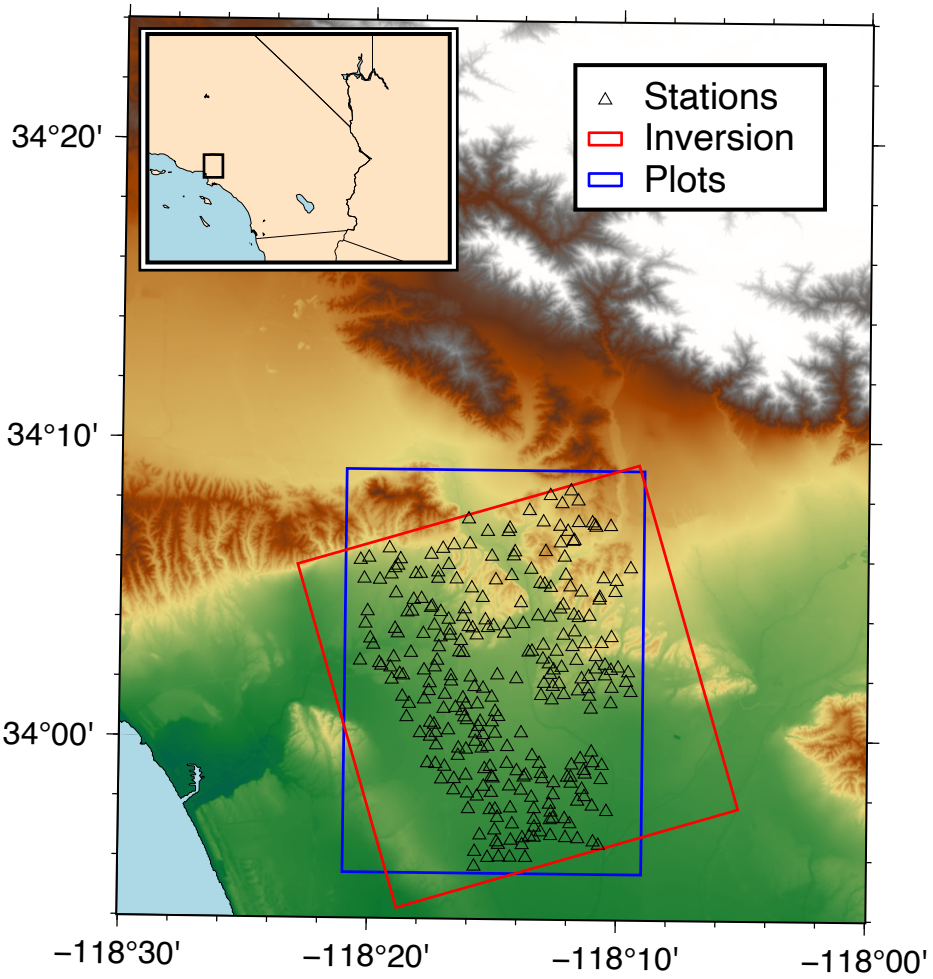


Figure 6.2: Map of the study region, showing the locations of the CSN stations as empty triangles, the boundary of the square inversion region in red, and the boundary of the analysis plots in blue.

a comparison of the data and predicted ground motions by several methods. For our study, data were first detrended, rotated into the *ZRT* frame, decimated to 5 Hz and then detrended once more. Pseudo-spectral accelerations (PSA) were then calculated for both the real data and synthetic 3D finite-difference simulations following the Graves and Pitarka method (Graves and Pitarka, 2010; Pitarka et al., 2019) for both the CVM-H and CVM-S models by convolving the records with a 5% damped harmonic oscillator, with the results for 4–9 s period shown in Figure 6.3. A record section of the high-frequency strong-ground-motion-accelerometer

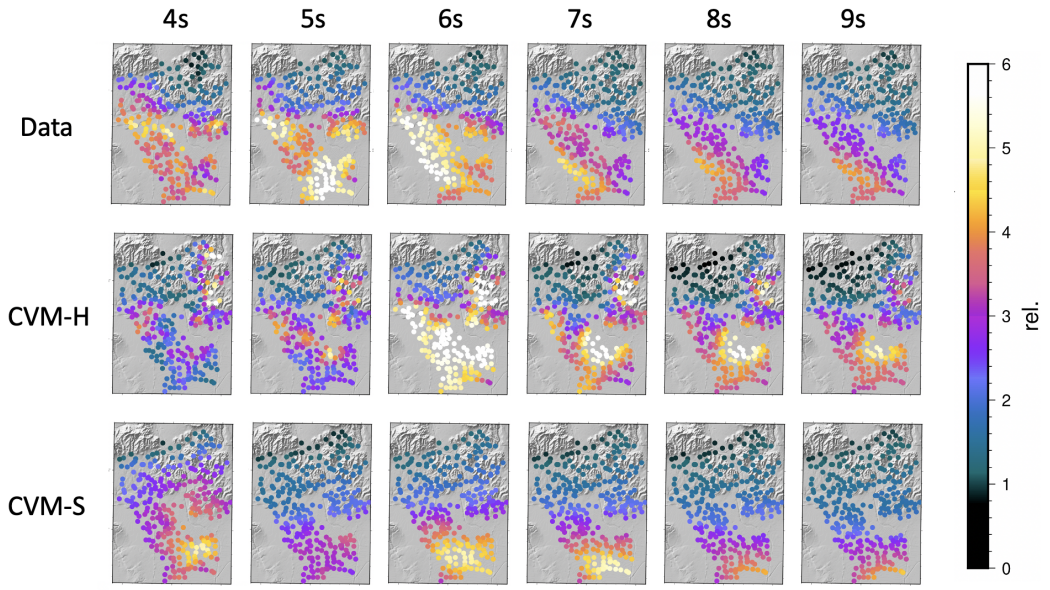


Figure 6.3: Relative amplification of the maximum amplitude of 3 component pseudo-spectral accelerations (PSA) in the range of 4–9 s from the Mw 7.1 July 5 2019 Ridgecrest Earthquake as recorded on the Community Seismic Network (CSN), and as simulated using the Graves and Pitarka rupture generator (Pitarka et al., 2019) and a 3D finite-difference waveform solver for both the CVM-H and CVM-S models.

transverse (HNT) channel showing strong SH polarized phases corresponding to the fundamental Love mode is shown in Figure 6.4.

Love Group Arrival Time and Amplitude Picks

To make group arrival picks, raw waveforms were first narrow-band filtered at period P using a zero-phase Butterworth bandpass filter with corners at $1/P \pm 1/(\sqrt{20}P)$ and then cosine tapered over the first 20s of the time series to suppress edge effects. The maximum of the T component envelopes at a central period $P = 12.5$s were set as the first preliminary group arrival pick. The 12.5s filtered waveform envelopes were then again cosine-tapered with a $6P$ taper window with $1P$ edges about this preliminary pick. We then fit a Gaussian function to the waveform envelope, with the center of the Gaussian being used as the finalized group arrival pick at 12.5s and the amplitude of the Gaussian being recorded as the Love wave amplitude. Starting

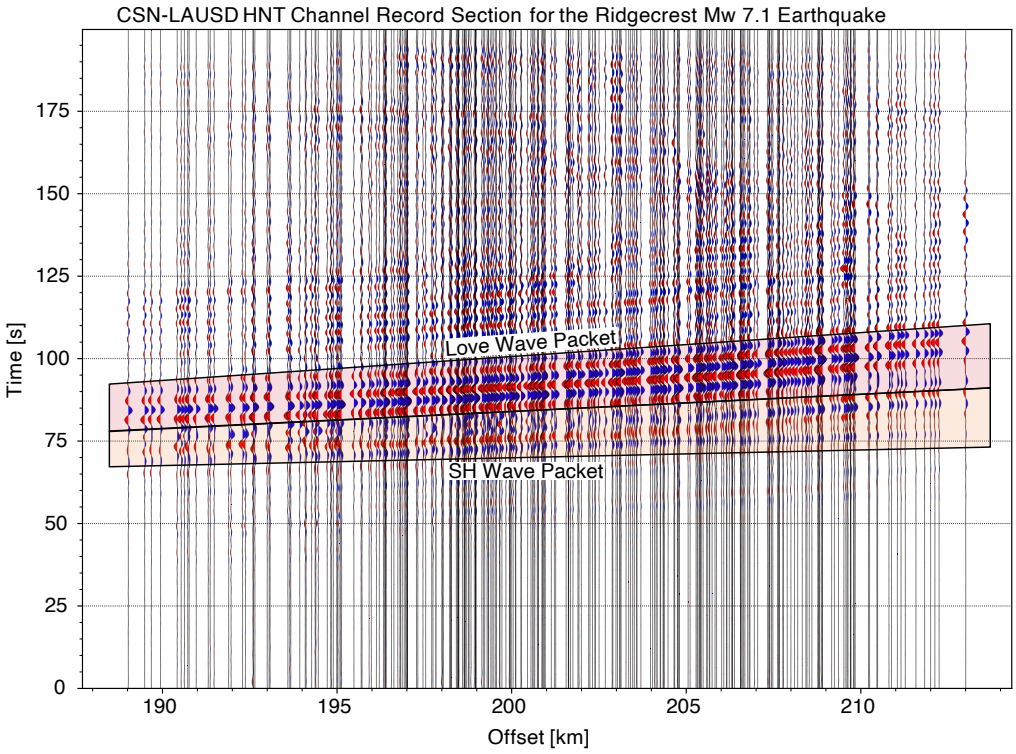


Figure 6.4: Record Section of the Mw 7.1 Ridgecrest earthquake as recorded on the HNT channel of the CSN-LAUSD array, zero-phase bandpass filtered between 4–10s. Two main phases are clearly identifiable, with the first arriving phase exhibiting little delay due to the basin at longer offsets, which we infer to be the primary SH arrival, which is shaded orange. A second, stronger phase, which is delayed by the basin at longer offsets, we infer to be the fundamental Love mode and is shaded red.

with the parameters of the 12.5s Gaussian as initial values, we then proceeded to work down in 0.25s increments on the narrowband filtered waveform envelopes, to a minimum period of 2s. We tapered with the $6P$ width cosine around the Gaussian center of the previous period. We then fit a new Gaussian to the shorter-period waveform, initialized using the previous period's Gaussian fit. This method tracks the Love-wave group arrival from long periods, where it is clearly identifiable as the strongest feature, to shorter periods where other features are present. A characteristic example of the group picks is shown in Figure 6.5.

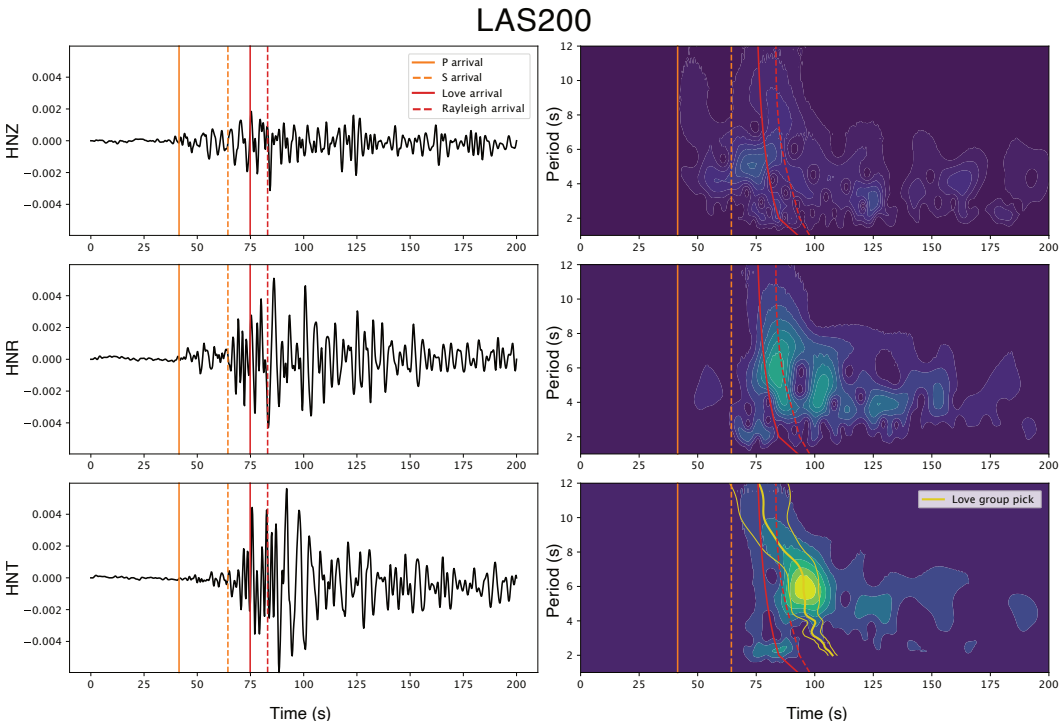


Figure 6.5: HN waveforms and corresponding continuous-wavelet transform spectrograms for the LAUSD CSN station LAS200 from the July 5 2019 Ridgecrest Mw 7.1 earthquake. The solid and dashed orange lines show the theoretical arrival times of the *P* and *S* waves through the laterally averaged CVM-H model from the hypocentral location to LAS200, and the solid and dashed red lines show the theoretical group arrivals for Love and Rayleigh waves, respectively. All theoretical travel times are offset from the event origin time by 10s, which is the approximate peak of the USGS moment rate function. The lemon yellow lines show the center and $\pm 1\sigma$ width of the fitted Gaussian functions to the envelope of the tangential component. The center of these Gaussian functions act as group delay picks for defining the cross-correlation window used for two-station phase delay measurements shown in Figure 6.6.

We took the logarithms of the fitted Gaussian amplitudes and normalized them relative to the mean log at each period to create the amplitude data set. The relatively narrow aperture of the CSN array compared to the distance to the source meant that the geometry was not favorable for traditional tomographic methods. We therefore employed eikonal tomography (Lin et al., 2009, 2014) to calculate surface-wave dispersion curves, which has the additional advantage of naturally handling the curving wavefronts recorded on the CSN, caused by refraction across the basin boundary. While recent studies (Qiu et al., 2019) have attempted to utilize group arrival times for eikonal tomography of group velocity, there is significant noise associated with the group arrival peak. Furthermore, there are strict conditions on the approximations necessary for using eikonal tomography on group delay times which may not be met when the surface-wave arrival experiences refraction across a basin boundary. As such, we did not attempt to utilize group velocity c_g in this study, but rather used the group times as a guide for two-station cross-correlation phase delay times as discussed below.

Eikonal Tomography from Two-Station Cross-Correlation Phase-Delay Times

We employ eikonal tomography (Lin et al., 2009) to obtain phase velocity estimates within the densely spaced CSN array. Eikonal tomography obtains phase velocity c directly from the gradient of the phase delay field: $|\nabla\tau| \approx 1/c$. Eikonal tomography has two principle requirements. Firstly, there must be a clearly identifiable phase delay field τ (i.e. there is no significant multipathing), a requirement which is met for Love waves in the period range of this study. Secondly, eikonal tomography is derived from an approximation of the transport equation $1/c^2 = |\nabla\tau|^2 - \nabla^2 A/A\omega^2$, where ignoring the amplitude correction is typically taken to be valid for velocity models that are sufficiently laterally smooth that the amplitude Laplacian is small. Waves propagating from the Ridgecrest earthquake sequence strike the northeastern

edge of the Los Angeles Basin nearly perpendicularly, so any effect of the basin edge on the Laplacian is limited in extent within the LAUSD-CSN array. It is possible to utilize the full transport equation for determining phase velocity, which is called Helmholtz tomography (Lin and Ritzwoller, 2011), however comparisons between Helmholtz tomography and eikonal tomography show agreement across the basin transition where we would expect the amplitude correction to be strongest, implying that eikonal tomography is sufficient to capture the correct phase velocity in the center of the array. Spurious values of the Helmholtz tomography solutions occur on the edges of the array due to the difficulty of obtaining accurate values of the amplitude Laplacian. Consequently, we limit our data analysis to the phase velocities derived from the eikonal equation as its assumptions appear to be satisfactorily realized and the Helmholtz tomography corrections are not sufficiently robust given our data.

In order to obtain the phase delay field τ at period P (relative to the northernmost station of the array), we first narroband filter wavepackes using central period P and cosine tapered with a flat pass window of width $4P$ and edges of P centered at the group arrival time. We then calculate the cross-correlation time delay $\Delta\tau_{ij}$ between each pair of stations i and j within a circle of radius $r_{ij} < \max(c_g P, c_{min} P)$ with a cutoff velocity $c_{min} = 0.5$km/s. The distance limit reduces the impact of potential cycle skipping on the phase delay observations, while the narrower taper width compared to the group picks also helps to stabilize the cross-correlation calculations. This process is illustrated in Figure 6.6 a) and b). The relative delays $\Delta\tau_{ij}$ form a graph with stations acting as nodes and the delays acting as edge weights. Similarly, the distances between stations Δd_{ij} also form a graph. Appealing to Fermat's principle of least travel time, we extract the minimum spanning tree (MST) of the station distance graph, and then use the geometry of the MST to find the minimum travel time surface. The MST is a unique sub-graph that connects all

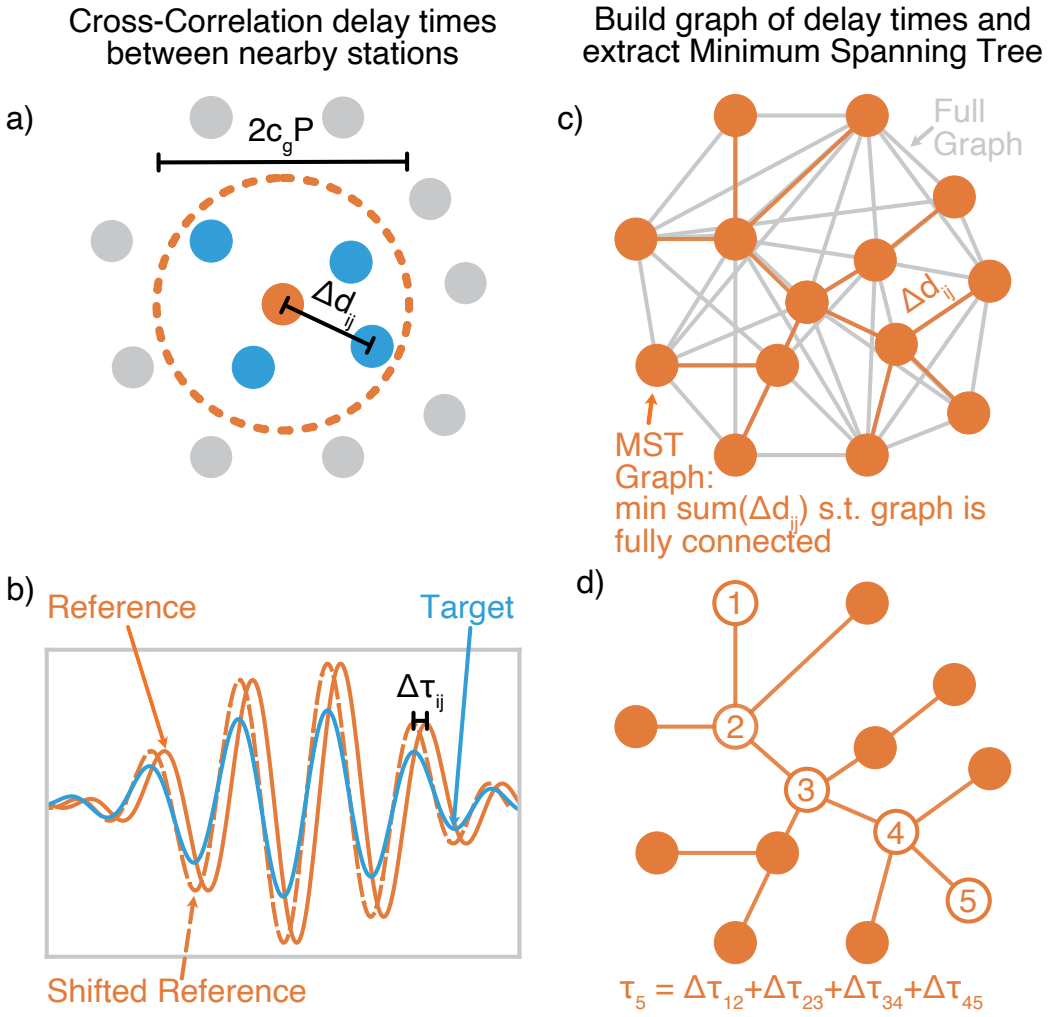


Figure 6.6: Outline of steps used to construct the phase delay field τ from narrow-band filtered records. In the first two steps, the phase delays between all nearby stations are computed. In a), we draw a circle of radius $r_{ij} < \max(c_g P, 0.5P)$ and compute the phase delay for maximum cross-correlation, $\Delta\tau_{ij}$, as shown in b). Only nearby stations are used to suppress cycle skipping. In the second phase, we extract the minimum spanning tree (MST) from the graph of collected phase delay times, as shown in c). The MST is a sub-graph which minimizes the total edge lengths (i.e. Δd_{ij}) such that the graph is still fully connected. Finally, in d) we unroll the MST from the northernmost station, summing $\delta\tau_{ij}$ along the edges to get the τ, a minimum-relative-phase-delay surface concordant with the recorded relative phase delays between individual station pairs.

nodes (stations) with minimum edge weights (distances), with a schematic of this subgraph shown in Figure 6.6 c). Summing phase delays $\Delta\tau_{ij}$ along MST edges from the northernmost station gives a minimum relative travel time surface that is concordant with the observed phase delay data, as shown in Figure 6.6 d). We also tested MSTs extracted from the graph of normalized cross-correlation values, as well as the phase delays themselves, but found that distance weighting gave the best performance. We then smooth the travel-time surface at each period by first fitting a high-tension cubic spline to the data, removing all outlying data points for which the fit residual at that point were greater than one standard deviation of all collected residuals, and then refitting the spline to the remaining data. This smoothed surface τ is then used to calculate phase velocity c at period p using the eikonal equation $|\nabla\tau| = 1/c$.

Estimating Measurement Uncertainty

The only available earthquakes that have produced sufficiently strong ground motions to record at least one octave of frequencies of Love waves are the Mw6.4 and Mw7.1 Ridgecrest events. Two events are insufficient to obtain useful statistical estimates of measurement uncertainty at each individual station. However, given that the surface-wave measurements have a finite area of sensitivity that overlaps substantially between neighbouring stations, we may bin error statistics over subarrays of radius $\lambda/4$ to obtain an estimate of the measurement uncertainty, where λ is the wavelength at the period of measurement. At station i, we calculate the mean of the relative log amplitude $\tilde{a}^i = (a^i_{6.4} + a^i_{7.1})/2$ and phase velocity $\tilde{c}^i = (c^i_{6.4} + c^i_{7.1})/2$ where $a_{6.4}$ and $c_{6.4}$ are the amplitude and phase velocities for the Mw 6.4 earthquake, respectively, and likewise $a_{7.1}$ and $c_{7.1}$ are the amplitude and phase velocity for the Mw 7.1 earthquake. We then estimate the standard error in the mean by averaging

over errors at nearby stations:

$$\sigma_a^i = \sqrt{\sum_{j \in d_{ij} \leq \lambda/4} \left(a_{6.4}^j - \tilde{a}^j\right)^2 + \left(a_{7.1}^j - \tilde{a}^j\right)^2} / \sqrt{2} \tag{6.1}$$

$$\sigma_c^i = \sqrt{\sum_{j \in d_{ij} \leq \lambda/4} \left(c_{6.4}^j - \tilde{c}^j\right)^2 + \left(c_{7.1}^j - \tilde{c}^j\right)^2} / \sqrt{2}, \tag{6.2}$$

where d_{ij} is the distance between stations i and j. The error correlation matrix P_{ij} is estimated using a squared-exponential kernel with characteristic lengthscale equal to one quarter of the average wavelength between the two stations, with the addition of a diagonal term to account for uncorrelated error

$$P_{ij} = \delta_{ij} + \exp(-8d_{ij}^2/(\lambda_i + \lambda_j)^2), \tag{6.3}$$

where δ_{ij} is the Kronecker delta. For each period the error covariance matrices are therefore given by $\Gamma_c = \sigma_c P \sigma_c^T$ and $\Gamma_a = \sigma_a P \sigma_a^T$, where σ_c is the collected vector of individual station phase-velocity error measurements across all periods, and σ_a is likewise the vector of amplitude error measurements. Future work on error modelling could account for a variable scaling between the diagonal and non-diagonal terms in P, and model the correlations between measurements at neighboring periods; however for reasons of computational expediency we do not develop these analyses here.

6.4 Inversion Methodology

Model Parameterization

Having obtained measurements $\tilde{c}$ and $\tilde{a}$ and associated error matrices Γ_c and Γ_a for phase velocity and log-relative amplification within the CSN, we are now in a position to model them and invert for a local basin update. We seek to obtain a parsimonious local update that balances the constraints of new, densely recorded data, with the already well developed models presented in the SCEC CVMs. Ideally, we would perform a fully Bayesian inversion taking a CVM as a prior model; however as robust model uncertainties for the CVMs are not available, this approach would be

highly dependent on subjective estimates for setting the prior, and would furthermore be extremely computationally expensive for the nonlinear forward models required to predict our recorded data. Instead, we recognize that the sensitivity of our data is highly contained within the basin itself, given the characteristic phase velocities c and periods p of our study and the heuristic sensitivity depth of $cp/4$ for Love waves in a power-law basin-style velocity profile, given by Haney and Tsai (2020). Taking advantage of this restricted sensitivity, we utilize the level-set-tomography framework of Muir and Tsai (2020b) to explicitly define a volume within which we perform our model updates as part of the model parameterization, and appropriately regularize the boundary of this volume to achieve the desired parsimony between the *a priori* CVM model and constraints from our newly observed data.

In this study, our model parameterization consists of two parts—a boundary to the inversion domain, and the velocity perturbations within that domain. Both components of the model are given by Gaussian Processes (GP) with a Whittle-Matérn kernel—briefly, this GP model supposes that the outputs are jointly distributed like a multivariate normal distribution with a pairwise covariance between model points with spatial locations x and x' given by

$$C(x, x') = \sigma^2 \frac{2^{1-\beta}}{\Gamma(\beta)} \left(\frac{||x - x'||}{l} \right)^{\nu} K_\beta \left(\frac{||x - x'||}{l} \right), \tag{6.4}$$

where Γ is here the gamma (or extended factorial) function and K_β is the modified Bessel function of the second kind. A comprehensive treatment of classical GP models may be found in (Rasmussen and Williams, 2006). The statistical properties of the GP are controlled by its hyperparameters, which for the Whittle-Matérn kernel are l, the characteristic length scale, σ the characteristic scale of perturbations, and β the regularity parameter. Individual realizations of the GP are $\beta - \frac{1}{2}$ times continuously differentiable. In practice β is very hard to infer in most inverse problems and so it is set to $\beta = 3\frac{1}{2}$ for the remainder of this study, a choice

which generates sufficiently smooth models to ensure that Love-wave eigenvalues are correctly calculated, and which does not introduce artificial roughness into the posterior distribution.

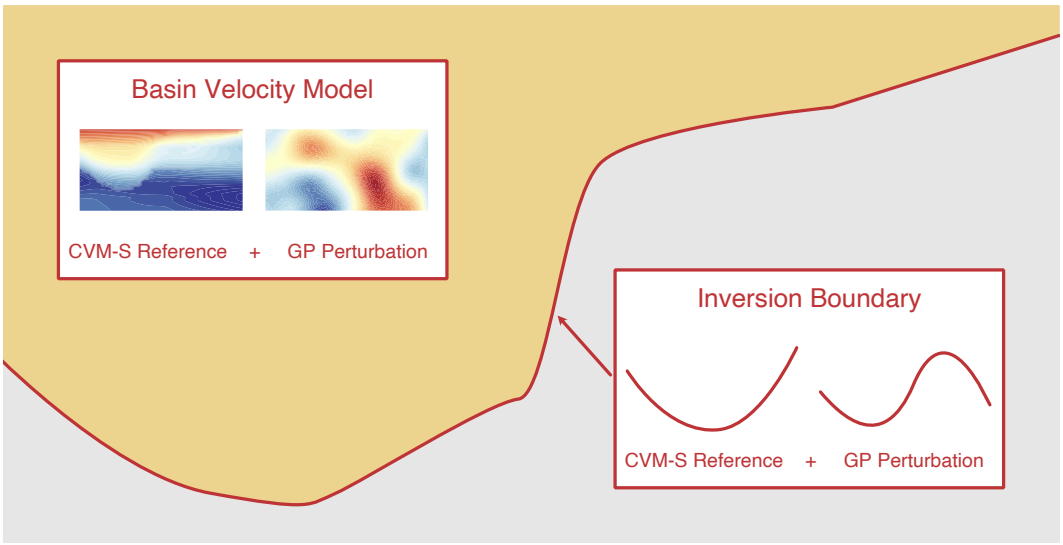


Figure 6.7: Schematic of the model definition, showing the construction of the velocity model update and the boundary of the inversion, both constructed from a CVM-S reference perturbed by a Gaussian Process. The background model, schematically shown in grey, is given by the unaltered CVM-S model.

GP models with variable hyperparameters offer great flexibility, however they are expensive to compute in the spatial domain as they require repeated inversion of the covariance matrix C —an operation of complexity $O(n^3)$ for n model evaluation points. To accelerate the GP computations, we approximate the model by defining it on a regular grid with n_{cell} grid nodes in each dimension, which allows us to specify the model by means of its Fourier coefficients ξ_v and ξ_b for the velocity and inversion boundary components respectively (Chen et al., 2019a; Lindgren et al., 2011). Efficient sampling of the GP can then be performed by an inverse Real Fast Fourier Transform (complexity of order $O(3m^3 \log(m))$ where $m = n_{cells}/2+1 \ll n$), followed by interpolation by cubic splines to the locations required for computing the forward model for phase velocity and amplitude underneath each station. We use the same lengthscale parameter l for both the velocity update and the inversion

boundary; the inversion domain is $22 \times 22 \times 12$ km in size, which must be rescaled to a unit cube for the inverse Fourier transform. The inversion area was determined by finding the smallest square that encompassed the stations, and is shown in in Figure 6.2. We use 16 cells in each dimension, and a rescaled $\tilde{l}$ parameter on the unit cube domain, which induces an effective lengthscale of $l_{xy} \sim 22\tilde{l}$ in the horizontal direction and $l_z \sim 12\tilde{l}$ in the vertical direction—equivalent to assuming vertical heterogeneity approximately twice as sharp as lateral heterogeneity. We denote the evaluation (via inverse FFT) of the velocity GP model given velocity Fourier coefficients ξ_v, lengthscale $\tilde{l}$ and velocity characteristic perturbation amplitude σ_v at a location (x, y, z) by $GPV_{\xi_v,\tilde{l},\sigma_v}(x, y, z)$, and the evaluation of the inversion boundary given boundary Fourier coefficients ξ_b, lengthscale $\tilde{l}$ and boundary characteristic perturbation amplitude σ_b at a location (x, y) by $GPB_{\xi_b,\tilde{l},\sigma_b}(x, y)$. For both GP models, a Whittle-Matérn kernel is assumed, and we use the CVM-S velocity model and basin profile to set mean to ensure initialization near a physical solution. CVM-S was chosen over CVM-H as the mean due to its smoothness, which lends itself to more concordant velocity models across the inversion boundary, and also because it better fits waveforms within the basin (Lai et al., 2019).

The V_s model is therefore given by

$$V_s(x, y, z) = \begin{cases} V_{\text{CVM-S}}(x, y, z) + GPV_{\xi_v,\tilde{l},\sigma_v}(x, y, z) & z < z_{\text{CVM-S}}(x, y) + GPB_{\xi_b,\tilde{l},\sigma_b}(x, y) \\ V_{\text{CVM-S}}(x, y, z) & z \geq z_{\text{CVM-S}}(x, y) + GPB_{\xi_b,\tilde{l},\sigma_b}(x, y), \end{cases} \tag{6.5}$$

where $V_{\text{CVM-S}}$ and $z_{\text{CVM-S}}$ are the S velocity model and basin edge extracted from CVM-S. A graphical schematic of the definition of the discretized model is shown in Figure 6.7. Density and V_p are then calculated from the V_s model using the empirical relationships of Brocher (2005), which are suitable for basins within southern California.

Extracting Reference Basin Depth Profiles from CVM-S

The SCEC CVM-S model is defined by a gridded voxel parametrization of V_P, V_S and ρ, i.e., it does not contain explicit definitions of basin boundaries. To obtain reference boundaries for the CVM-S model, we utilized the following procedure. At each depth slice, we computed the mean and standard deviation of V_S. We then flagged each voxel for which V_S was slower than one standard deviation below the mean of that depth slice as a potential basin candidate. For each 1D depth profile, we then worked from the second (z=500m) depth slice downwards, flagging a voxel to be within a basin only if all voxels above it were also flagged—working from the second depth slice avoids the connection of individual basins due to artificial connectivity in the absence of the geotechnical layer. This process encodes an assumption that basins are strictly convex, which is not true in general but is a useful approximation to begin the inversion process. Using the *scipy* module *ndimage* (SciPy 1.0 Contributors et al., 2020), we then performed image segmentation using the *label* function, which generated 61 individual basins in southern California, of which the most prominent correspond to the Ventura Basin, combined Los Angeles and San Gabriel basins, San Fernando Basin, and the Salton Trough. This workflow is presented in Figure 6.8. The boundaries of the Los Angeles / San Gabriel basin candidate were then utilized as the reference basin bottom surface for the inversion step.

Forward Modelling

In order to predict the data from the final rasterized velocity model given by our model parametrization, we employ the lumped-mass finite element method for surface-wave eigenvalue calculation first proposed by Lysmer (1970), and implemented for Love waves by Haney and Tsai (2020). The rasterized model is interpolated onto a set of finite elements of exponentially increasing thicknesses h given by

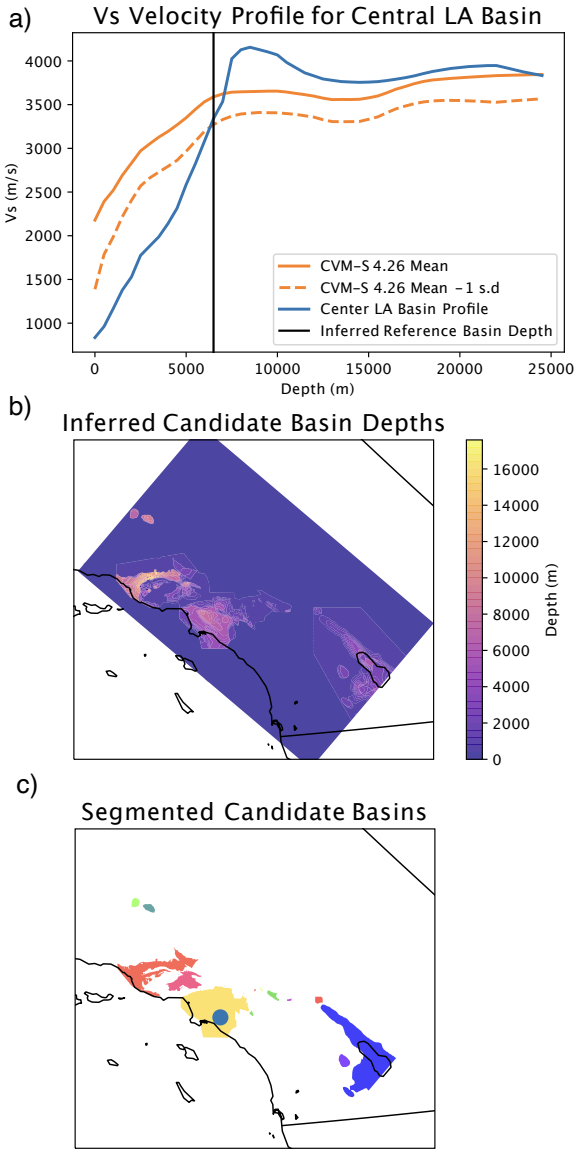


Figure 6.8: Outline of steps used to extract a reference basin surface from the CVM-S. a) for each vertical profile in CVM-S, we determine where (if anywhere) the V_S profile first becomes faster than one standard deviation below the mean CVM-S at that depth. All depths above this level are set to be a potential candidate basin at the location of the profile. In b), we show the extracted candidate basin depths across southern California. In c), we strip off the top 500m (which is highly connected) and then use the SciPy *ndimage label* function to segment the remaining data volume. The three major basin families of southern California are clearly seen in pink (Ventura / San Fernando), yellow (Los Angeles / San Gabriel / San Bernadino) and blue (Salton Trough).

$h_n = \min(c) * \exp(N/(na))/n$ where $N = 50$ is the number of layers in the model, $\min(c)$ is the minimum phase velocity in a reference model, and $a = 0.25$ is a constant used to control the exponential scaling. This exponential scaling heuristically balances the need for finer resolution near the top of the model when calculating shorter period Love waves in a way that is near optimal due to the approximate exponential shape of Love eigenfunctions (Haney and Tsai, 2015, 2017, 2020; Tsai and Atiganyanun, 2014). These layers are stacked on top of 4 layers of thickness $h = 10$ km simulating an infinite half-space to avoid contamination with the locked lower boundary condition. We then set up the finite element stiffness and mass matrices as given by Haney and Tsai (2020), and solve for the maximum slowness eigenfunction u that corresponds to the fundamental Love mode, as well as the phase velocity $c = \sqrt{\nu}\omega$, with ν being the eigenvalue associated with u for angular frequency ω, and group velocity c_g which is a function of c, u and the finite-element mass and stiffness matrices. The relative amplification of Love waves directly observed between two locations can then be calculated by

$$\frac{a_1}{a_2} = \left(\frac{c_{g_1} I_1}{c_{g_2} I_2}\right)^{-1/2}, \tag{6.6}$$

with $I = \int_0^\infty \rho(z) u(z)^2 dz$ (Bowden and Tsai, 2017; Bowden et al., 2017). Transmission coefficients obtained using a 1D mode-conversion theory (Brissaud et al., 2020; Datta, 2018) are plotted in Figure 6.9, and suggest that any potential modelling error from neglecting mode-coupling is small. As we use a derivative-free inversion method, these quantities are sufficient to solve for the optimal model.

Inverse Solver

We use an extension of the Ensemble Kalman Sampler (EKS, Garbuno-Inigo et al. (2020)) to perform the inversion. This method uses an interacting ensemble of particles that follow Langevin diffusion dynamics to infer a Gaussian approximation to the posterior of the inverse problem. The EKS is derivative-free and embarrassingly

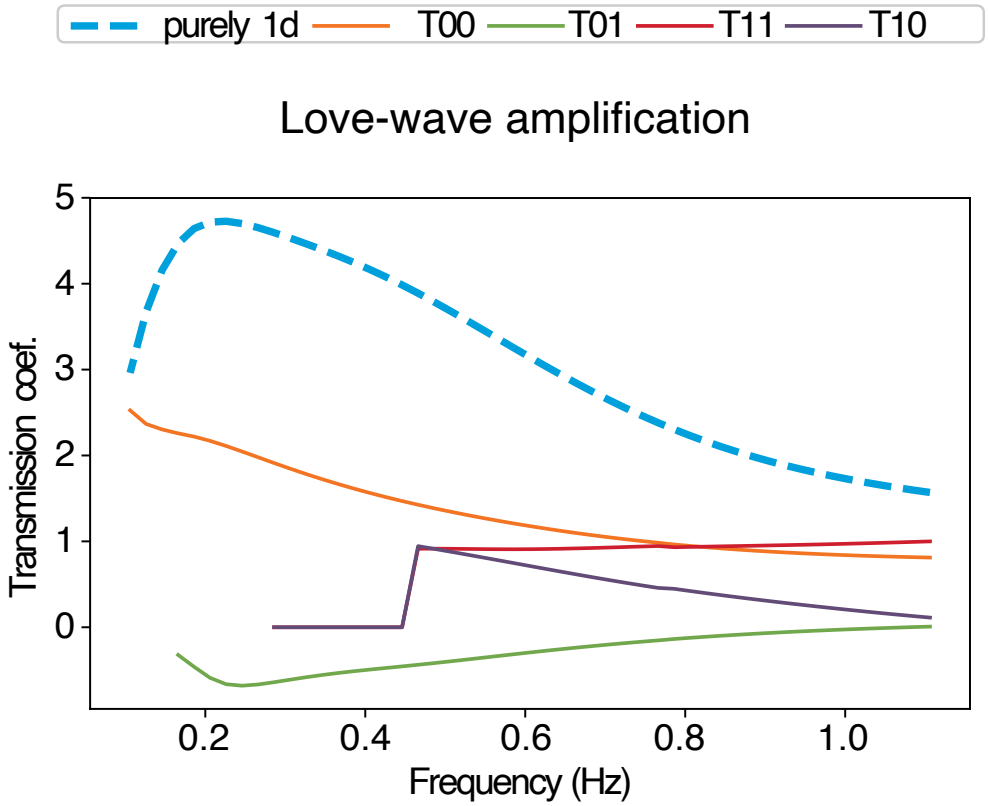


Figure 6.9: Transmission coefficients for a Love wave entering the Los Angeles basin obtained using a 1D mode-coupling theory (Brissaud et al., 2020; Datta, 2018). This represents a worst-case mode-conversion scenario, with the true basin exhibiting a smoother horizontal gradient and hence less conversion. Even in this case, the conversion of energy from the fundamental mode to first overtone T_{01}/T_{00} is relatively small, suggesting that our use of classical Love-amplification theory is appropriate.

parallel in the forward model, which enable rapid user iteration between different datasets and forward modelling methods, as well as easy deployment on heterogenous computing networks. The EKS as outlined in Garbuno-Inigo et al. (2020) assumes that all model parameters have a Gaussian prior. This restricts the model to have fixed hyperparameters (e.g. $\tilde{l}$, σ_v, σ_b, as required to set the statistical behaviour of the model parameterization described in Section 6.4), which introduces a significant potential for practitioner bias as we do not have a good basis for estimating these *a priori*. Consequently, we have further developed the EKS to handle hierarchical

models with variable hyperparameters. The original EKS and our extension to it are discussed in detail in Appendix 6.8. The priors for the velocity hyperparameters are given by $1/\tilde{l} \sim Normal(0, 0.6)$ and $\sigma_v \sim Normal(0, 0.1)$ in scaled inverse km and km/s respectively. Experimentation has shown that the characteristic boundary perturbation amplitude σ_b is not sufficiently identifiable from our data, so we set it to a reasonable value of 0.5 km that is small enough to avoid large, unrealistic changes in the basin geometry whilst allowing a sufficient fit to the data. Using these hyper-priors, we run hierarchical EKS sampling using an initial step length $\Delta t_0 = 50$, and an ensemble size of 32. We double both the step length and the ensemble size every 50 iterations up to iteration 250, and further double the step length only at iteration 300, to finish with 400 iterations. The purpose behind this doubling scheme is to rapidly approach the *maximum a posteriori* (MAP) point using rough gradients from a small number of ensemble members, and then perform more accurate sampling of the posterior using more ensemble members (Garbuno-Inigo et al., 2020). The step length doubling counteracts the tendency of the gradient amplitude to be small near the MAP point. Convergence diagnostics for the inversion run are shown in Figure 6.10. The final inversion reduced the weighted Gaussian misfit function from 8.79 (for the CVM-S model) to 5.33, a variance reduction of 22%, which is a notable improvement from the already highly optimized reference model.

6.5 Results and Implications for the Los Angeles Basin

The results of the inversion are shown in Figures 6.11, 6.12 and 6.13. In Figure 6.11 we plot the mean depth to the inferred basin bottom and the inferred change in the depth of the Los Angeles basin at each station. The change in basin depth is defined by the difference between the reference basin depth extracted from the CVM-S in Section 6.4, and the depth to the same velocity contour in the final model. Figure 6.12 shows the details of the inversion along the profile A–A'. Figure 6.13 shows the

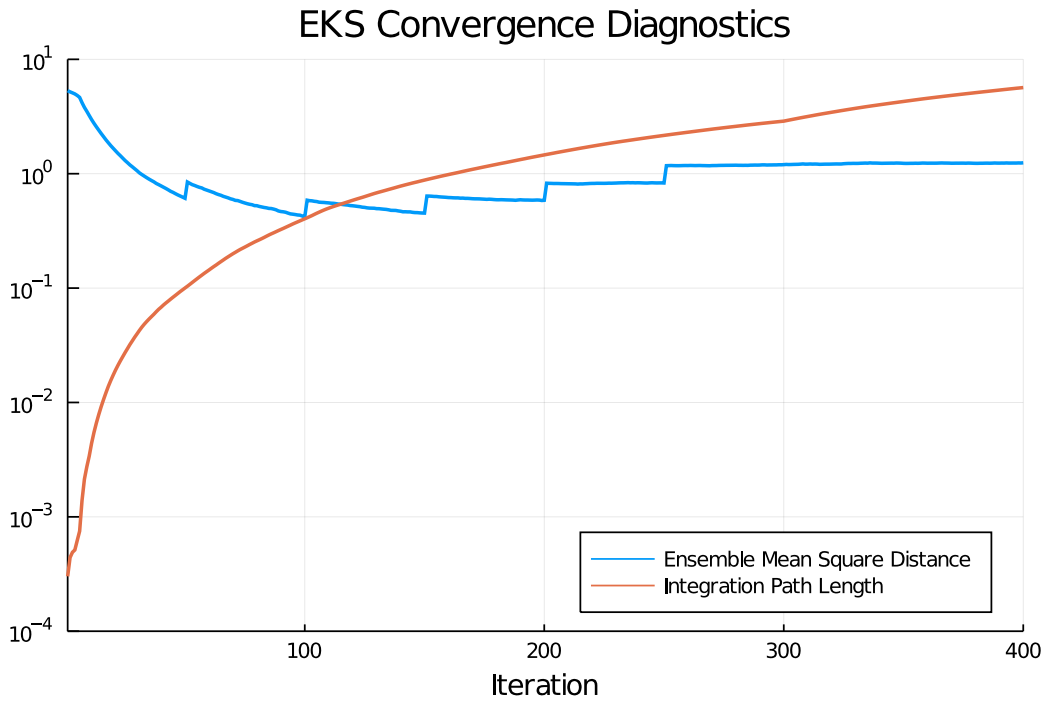


Figure 6.10: Convergence diagnostics of the Ensemble Kalman Sampler (EKS) showing the Ensemble Mean Square Distance converging to a constant approximation of the posterior, and the integration path length increasing steadily (heuristics from Garbuno-Inigo et al. (2020) suggest a path length of 2 is sufficient to approximate the posterior).

approximate posterior distribution of the hyperparameters in the inversion. In Figure 6.12, we also show the reference CVM-S model used to initalize the inversion, the mean of the EKS ensemble, the difference between these two, and the standard deviation of the ensemble. The standard deviation gives a sense of the relative uncertainty of the final inversion; as discussed in Garbuno-Inigo et al. (2020), in the low-particle limit EKS sampling cannot fully capture the range of uncertainty in the true inversion posterior, and so the plotted standard deviations are best assessed in a qualitative fashion. The EKS ensemble indicates that the highest uncertainties are along the boundary of the model. Within the inverted area of the final model, the uncertainties are highest in the deep central basin where the 4-10s Love wave period range offers less sensitivity, and near the northeastern edge of the model where the

phase velocities are high, resulting in small travel time gradients and hence higher uncertainties when employing eikonal tomography.

There are two principle features that are apparent from the results of the inversion. The first and most significant finding is that the data supports a deeper Los Angeles basin along its northeastern edge, with an especially large jump in basin depth in the area immediately abutting the Upper Elysian Park fault as defined in the USGS Quaternary fault map (USGS, 2020). The increase in basin depth reaches its maximum just south of downtown LA, as is seen in the south part of Figure 6.11 b) which shows the change in basin depth. The Upper Elysian Park fault is shown by a thick dashed cyan line in the center-right of the panels of Figure 6.11, and demarcates a steep gradient in the edge of the basin which has been accentuated as a result of the inversion. In Figure 6.12, this tall jump in the depth of the basin edge occurs in the center of the profile A-A', with Figure 6.12 c) showing that the deep parts of the basin to the SSW of the fault are significantly slower in our final model, with the edge of the basin being significantly steeper in our model in a) than the reference model in b). This steepening is spatially coincident with the observations of high amplification further north in the data than in the reference models, seen in Figure 6.3, particularly in 5–7 s band. Extracting the average basin edge gradient from 11.25–13.25 km along profile A–A' in Figure 6.12 gives a dip angle of 72–73°. The SCEC CVMs have evolved from the original models of Magistrale et al. (2000, 1996), which for the Los Angeles basin were based on an empirically determined velocity law for compacted sediments (Faust, 1951), with the spatial distribution of velocities controlled by contacts between two gross scale units (the Repettian and Mohnian), and the inferred basement depth, as reported in Wright (1991). There is a notable gap in the locations of control wells used by Wright (1991), which in turn initialized the SCEC CVMs (either as a starting model for full-waveform inversion used in CVM-S (Lee et al., 2014) or included as a constraint in CVM-H (Shaw

et al., 2015; Tape et al., 2009)), across the steep northeastern boundary of the basin that is now covered by the CSN. Given the position of the basin sidewall is situated between the imbricated blind-thrust faults of the Elysian Park system (Plesch et al., 2007), the high apparent dip angle imaged by surface-wave measurements gives further support to an over-thrusted basin in this region (as is included in the CVM-H model, albiet further to the northeast than is suggested by our results). Further cross-sections through the model are shown in Figure 6.14, and show that this steep basin sidewall continues along the northwest-southeast axis of the northern LA basin wall.

The second notable finding is that the depth of the low velocity zone in the hilly terrain north of the Los Angeles basin is substantially shallower than in the reference model, which can be seen both along the northern edge of Figure 6.11, and in the faster velocities around end A' of the transect in Figure 6.12 c). This shallowing of the basin relative to the CVM-S model is somewhat unsurprising given the high Love wave speeds recorded in the northeast of the array from eikonal tomography, and the relatively lower amplification when compared to the slow, deep sediments in the central basin. Indeed, the northeastern components of the CSN operate within the surface expression of the lower Puente and Topanga units of the LA basin stratigraphic column, which were assembled early within the LA basin sequence and support a shallow sequence of basin rocks towards to the right of profile A–A' (**?**). In the Supplement, we further discuss these two main features in the context of fitting the rule-based CVM1 (Magistrale et al., 2000, 1996) rule-based model to the profile A–A'; by perturbing the locations of the loosely constrained geological contacts that define the CVM1, we can analyse the outcomes of our fully 3D inversion in terms of geological structure, and find that the steep basin sidewall is consistent with recently ($\leq$4 Ma) active deformation, as suggested by our discussion here.

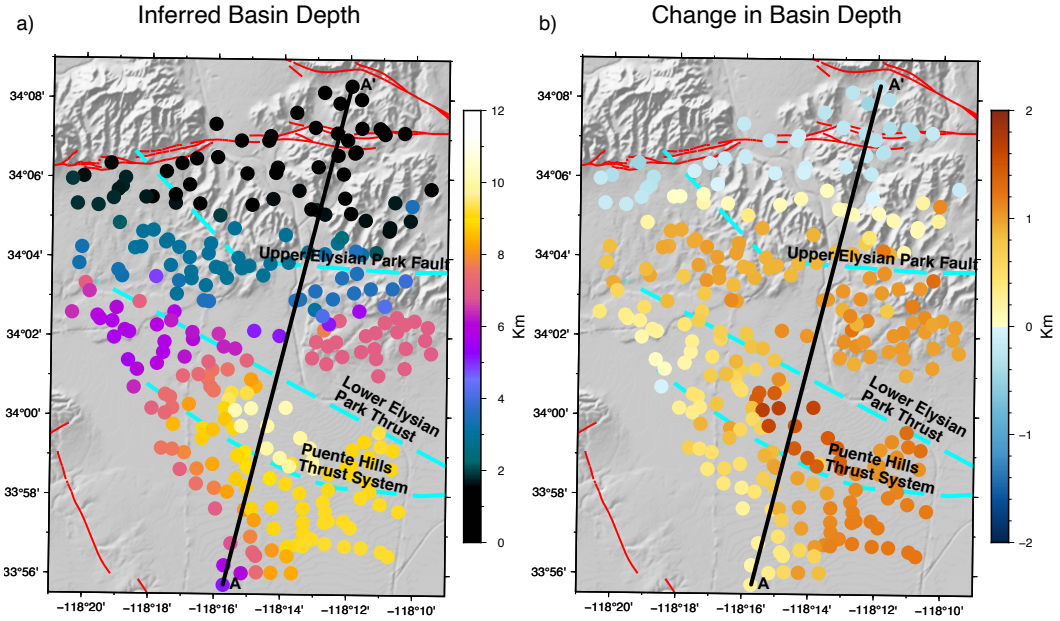


Figure 6.11: a) Mean depth of the inferred basin interface from the final ensemble. b) The inferred change in the depth of the Los Angeles Basin relative to CVM-S, showing deepening of the basin especially south of the Upper Elysian Park fault (top thick dashed cyan line), and shallowing of the model in the hilly terrain to the North of the CSN. In both panels, major late Quaternary faults (<130 Kyr) are shown in red, and other Quaternary faults are shown in thick dashed cyan. The transect A-A' is shown in black.

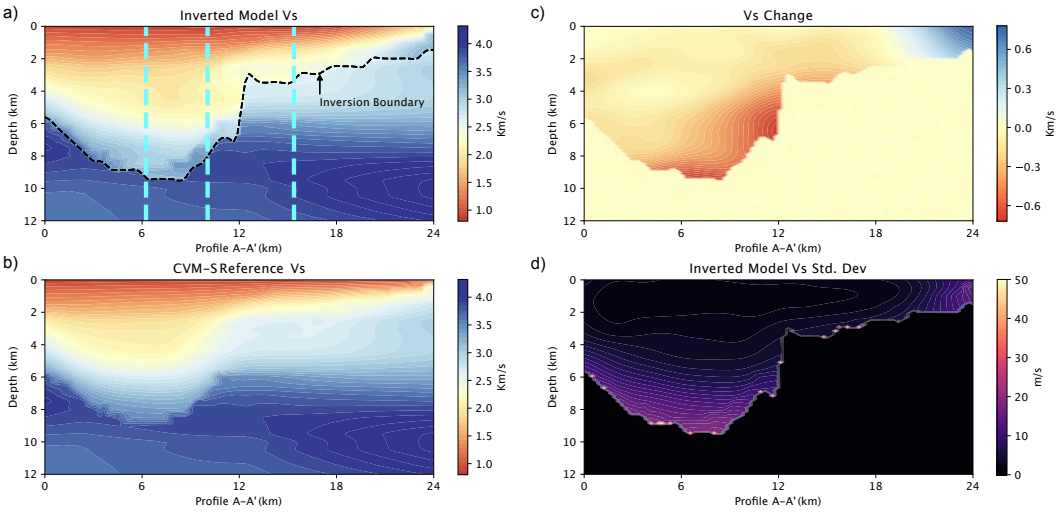


Figure 6.12: a) Mean of the final ensemble V_S model, b) CVM-S reference model V_S, c) difference between final model and reference model, d) standard deviation of the final ensemble V_S model.

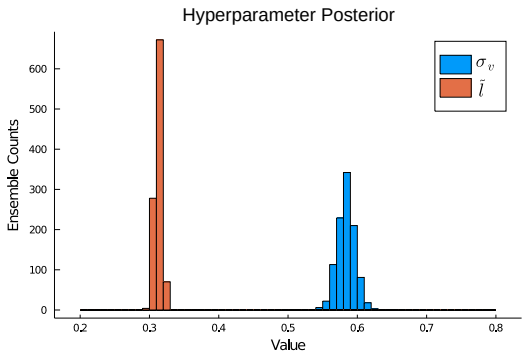


Figure 6.13: Approximate posterior distribution from the final ensemble for the hyperparameters $\tilde{l}$ and σ_v.

6.6 Conclusion

We use Love waves generated by the Mw 6.4 and Mw 7.1 Ridgecrest, CA earthquakes to obtain Love-wave phase velocities and relative amplitudes between 4–10 s period using the Caltech-LAUSD Community Seismic Network, which offers unprecedented high-density coverage of the northeast LA basin. We use the level-set method of Muir and Tsai (2020b) to develop a parsimonious velocity inversion that updates the SCEC CVM-S background model only where empirical estimates of data uncertainty indicate additional complexity is warranted. By employing fully 3D surface-wave inversion, we avoid internal artifacts in the model and make best use of a relatively small dataset. In doing so, we find that the northeast wall of the LA basin is substantially steeper than that of the CVM-S model, allowing for high amplifications of surface waves in the 4–6 s period band travelling within the basin. The constraints provided by this model cover some of the parts of LA with the highest density of population, infrastructure and commercial development, and highlight the continued importance of seismic velocity model evolution in providing the most accurate possible estimates of potential strong ground motions in this important city.

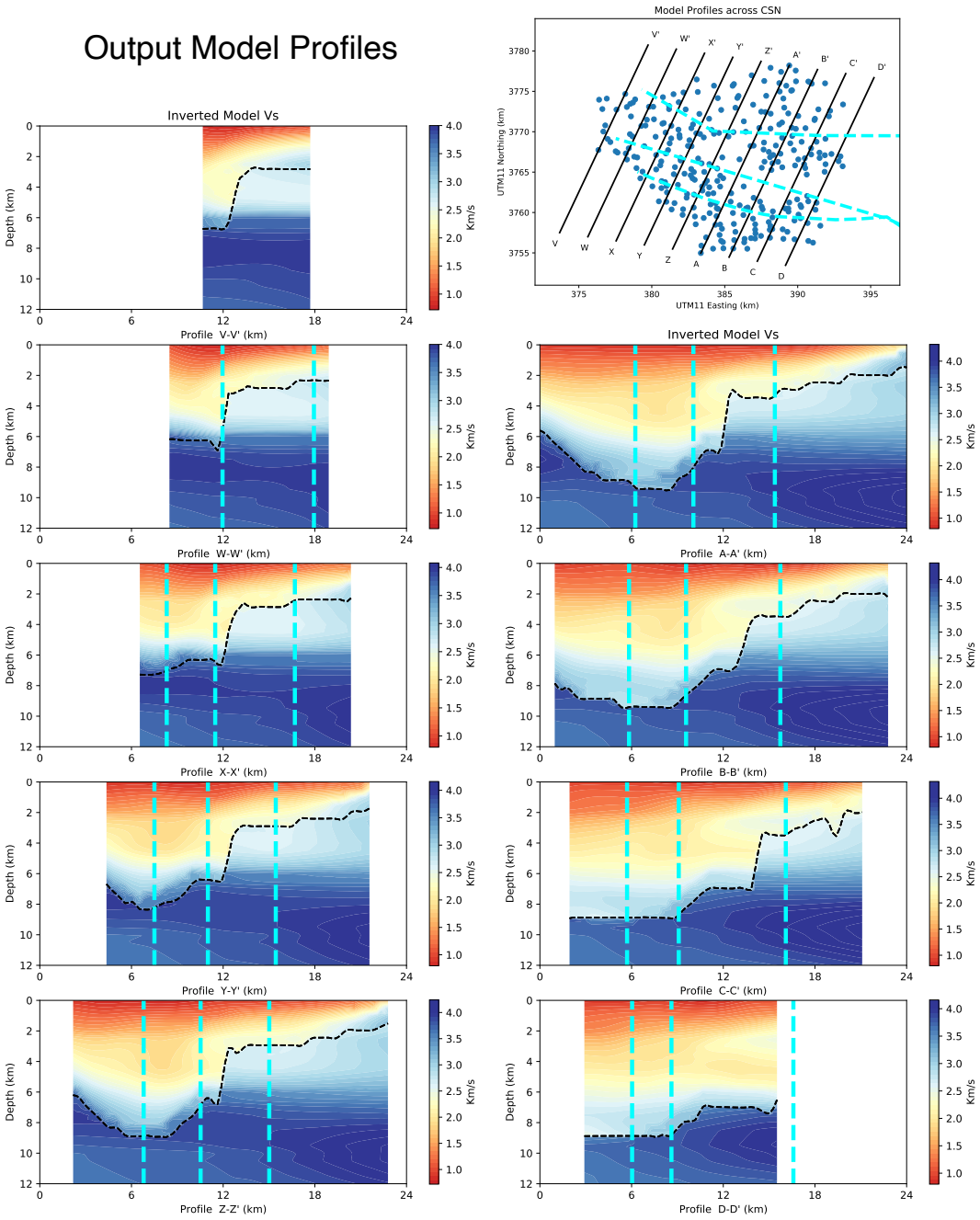


Figure 6.14: Profiles of the mean output V_s across the Los Angeles Basin, with inferred Quaternary faults in dashed cyan and the inferred edge of the inversion shown in dashed black.

6.7 Acknowledgements

This study was supported by the United States National Science Foundation awards EAR-1520081, EAR-2105358 and EAR-2011079, and the Southern California Earthquake Center award 20024. JBM acknowledges the support of the General Sir John Monash Foundation and the Origin Energy Foundation for support during his graduate studies. The CSN data used in this paper are freely available from `http://csn.caltech.edu/data`.

6.8 Appendix

Hierarchical Ensemble Kalman Sampler

The Ensemble Kalman Inversion (EKI) scheme was introduced by Iglesias et al. (2013) by deriving an state-variable augmented Ensemble Kalman Filter (Evensen, 1994, 2003) with dynamics that approximated the Levenberg-Marquardt method. EKI acts as an efficient black-box optimizer for large scale PDE constrained problems for which it is intractable or infeasible to obtain gradients, and has been used successfully in practical geophysical applications (e.g. Muir and Tsai (2020b); Tso et al. (2021)). Subsequent to its initial formulation, much analysis on the EKI scheme has been performed by studying it as a continuous time gradient flow (Kovachki and Stuart, 2018), rather than in its original formulation as a discrete time dynamical system. This has lead to the development of the Ensemble Kalman Sampler (EKS, Garbuno-Inigo et al. (2020)), an algorithm for approximate sampling of the posterior distributions of large-scale Bayesian PDE constrained inverse problems. We utilize a hierarchical variant of the EKS scheme in this study to sample the posterior distribution of our local model update—we will briefly reintroduce the EKS scheme as described in Garbuno-Inigo et al. (2020) and then outline our variant hierarchical formulation. In general, the objective of these schemes is to approximate a posterior

distribution whose negative log-posterior is of the form

$$\Phi(u, d) = ||d - G(u)||_\Gamma + R(u), \tag{6.7}$$

where Γ is the data noise covariance matrix, and where the regularization term $R(u)$ introduces prior information; for instance, a typical choice would be a Tikhonov style regularization term $R(u) = ||u||_{C_0}$ for some prior covariance matrix C_0. The norms here are defined by $||u||_A = \langle u, u\rangle_A = u^T A^{-1} u$.

The EKS scheme is an ensemble-based approximation of a preconditioned over-damped Langevin equation, which is a stochastic differential equation (SDE) of the form

$$\dot{u} = -C(u)\nabla_u \Phi(u) + \sqrt{2C(u)}\dot{W} \tag{6.8}$$

with $C(u)$ a preconditioning operator that depends on u and $\dot{W}$ a Brownian motion term. It can be shown that the long-term behavior of this SDE gives rise to a trajectory that has a distribution given by $p(u|d) \propto \exp(-\Phi(u, d))$, i.e. the desired target posterior (Gelman et al., 1997). In the EKS scheme, an ensemble of particles $U = \{u^{(j)}\}_{j=1}^J$ are used to approximate the gradient of the likelihood, and $C(u)$ to be is chosen to be the empirical covariance $C(U) = \frac{1}{J}\sum_{j=1}^J (u^{(j)} - \bar{u})(u^{(j)} - \bar{u})^T$, where overbars denote means across the particle ensemble. Preconditioning by the empirical covariance acts to approximate the local curvature of the posterior by the ensemble, giving accelerated convergence compared to the unconditioned equation in a similar manner to the difference between Newton's method and gradient descent. The dynamics of this system of particles are given by the following SDE (without the gradient approximation and for Tikhonov-style Gaussian priors)

$$\dot{u}^{(j)} = \frac{1}{J}\sum_{k=1}^{J} \langle (\nabla_u G(u^{(j)})(u^{(k)} - \bar{u}), G(u^{(j)} - d)\rangle_\Gamma u^{(k)} - C(U)C_0^{-1}u^{(j)} + \sqrt{2C(U)}\dot{W}^{(j)}. \tag{6.9}$$

Making the ensemble approximation for the gradient of the forward operator G allows us to rewrite this in a form without an explicit derivative:

$$\dot{u}^{(j)} = \frac{1}{J}\sum_{k=1}^{J}\langle (G(u^{(k)}) - \bar{G}, G(u^{(j)}) - d)\rangle_\Gamma u^{(k)} - C(U)C_0^{-1}u^{(j)} + \sqrt{2C(U)}\dot{W}^{(j)}, \quad (6.10)$$

which is the equation solved by the EKS as described by Garbuno-Inigo et al. (2020). We will define $D(U) = \frac{1}{J}\sum_{k=1}^{J}\langle (G(u^{(k)}) - \bar{G}, G(u^{(j)}) - d)\rangle_\Gamma$ for future convenience, so that the dynamics for the whole ensemble are given by

$$\dot{U} = UD(U)^T - C(U)C_0^{-1}U + \sqrt{2C(U)}\dot{W}. \quad (6.11)$$

We note that at the equilibrium of the ensemble, these dynamics suggest a balance between a Newton-style update of the ensemble (using an empirical covariance matrix to approximate the inverse Hessian) converging to the maximum *a posteriori* point, and the generation of correlated noise scaled to the original ensemble. The final state therefore results in a local Guassian approximation of the posterior.

Often, in geophysical problems, the scale of appropriate regularization (i.e. the choice of operator C_0 for Tikhonov regularized problems) is unknown. As such, much recent effort has been devoted to the development of hierarchical methods for solving inverse problems, in which the prior itself is to some degree unknown and is controlled by some number of hyperparameters (see, e.g.,Malinverno and Briggs (2004)). Additionally, for large-scale problems with Gaussian priors, it may be beneficial for efficient sampling to perform a coordinate transform into diagonalized non-centered coordinates, which remove the correlations in the prior between hyperparameters and the main parameters used in the inverse problem. This class of parametrizations are known as whitened, non-centered hierarchical parametrizations (Chada, 2018; Chada et al., 2018; Chen et al., 2019a). The set of parameters is given by a collection of "regular" parameters ξ and hyperparameters θ. For zero-mean Gaussian priors, the coordinate transform is given by $u = L(\theta)\xi$ for a Cholesky factor $C_0(\theta) = L(\theta)L(\theta)^T$. With this transform, the prior for the parameters ξ is simply

a Gaussian with identity covariance matrix. For reasons of computational efficiency, if the prior covariance C_0 is associated with spatial structure (say if the values of u represent material quantities at particular points in space) an approximate transform based on the solution to a stochastic partial differential equation (SPDE) is used (Lindgren et al., 2011), with the choice of SPDE determined by the particular form of the Gaussian prior to be approximated. For certain choices of prior covariance, and by defining known boundary conditions on a rectangular volume encompassing the model parameters, there are known analytic solutions for the appropriate eigenfunctions $\phi_i(\theta)$ and eigenvalues $\nu_i(\theta)$ with which to solve the SPDE such that truncation of the series of eigenfunctions has the smallest total mean squared error; these eigenfunction-eigenvalue pairs form the Karhunen-Loève (KL) expansion (Dashti and Stuart, 2013). Using the KL expansion, $L(\theta)\xi \sim \sqrt{\nu_i(\theta)}\phi_i(\theta)\xi_i$. By using these known analytic eigenfunctions and appropriately truncating the KL expansion to a reasonable number of eigenfunctions can drastically increase the speed of performing the coordinate transform; for the commonly used Whittle-Matérn family of covariance functions in a rectangular domain, the transform (assuming Neumann boundary conditions) can be calculated using the inverse discrete cosine transform for even greater efficiency.

The hyperparameters θ may have arbitrary priors ρ, which are typically non-Gaussian but do not depend on ξ; consequently the dynamics of the system follow (for ensembles $\Xi = \{\xi^{(j)}\}_{j=1}^{J}$, $\Theta = \{\theta^{(j)}\}_{j=1}^{J}$)

$$\dot{\Xi} = \Xi D(L(\Theta)\Xi)^T - C(\Xi)\Xi + \sqrt{2C(\Xi)}\dot{W} \tag{6.12}$$

$$\dot{\Theta} = \Theta D(L(\Theta)\Xi)^T + C(\Theta)\nabla_\theta \log(\rho(\Theta)) + \sqrt{2C(\Theta)}\dot{W}. \tag{6.13}$$

These dynamics derive from the original EKS by considering an augmented state vector $u = [\xi, \theta]^T$ and allowing arbitrary priors. We have furthermore neglected the cross-covariance terms $\mathrm{Cov}(\Xi, \Theta)$ and assumed a block-diagonal form for the

preconditioning matrix, allowing us to decouple the dynamics as above. In order to solve these equations, we use the same split-step implicit scheme as Garbuno-Inigo et al. (2020), which is given by

$$\Xi^*_{k+1} = \Xi_k - \Delta t_k \Xi_k D(L(\Theta_k)\Xi_k)^T - \Delta t_k C(\Xi_k)\Xi^*_{k+1} \tag{6.14}$$

$$\Theta^*_{k+1} = \Theta_k - \Delta t_k \Theta_k D(L(\Theta_k)\Xi_k)^T + \Delta t_k C(\Theta_k)\nabla_\theta \log(\rho(\Theta^*_{k+1})) \tag{6.15}$$

$$\Xi_{k+1} = \Xi^*_{k+1} + \sqrt{2\Delta t_k C(\Xi_k)}W(\Xi)_k \tag{6.16}$$

$$\Theta_{k+1} = \Theta^*_{k+1} + \sqrt{2\Delta t_k C(\Theta_k)}W(\Theta)_k, \tag{6.17}$$

where $W(\Xi)_k$ and $W(\Theta)_k$ are matrices of standard random normals of the same shape as Ξ and Θ respectively. The timestep Δt_k is calculated adaptively following Kovachki and Stuart (2018). Given a reference timestep Δt_0 we have $\Delta t_k = \Delta t_0/(||D(L(\Theta_k)\Xi_k)|| + \delta)$ where the norm on D is the Frobenius norm and δ is an arbitrary positive constant. Unlike in Garbuno-Inigo et al. (2020), the inclusion of arbitrary non-Gaussian priors for the hyperparameters θ mean that the implicit update is no longer linear, but as the dimension of θ is usually small, the cost of performing this update using an iterative nonlinear solver is normally not overly burdensome; in practice we use forward-mode automatic differentiation for arbitrary priors ρ and the L-BFGS method (Liu and Nocedal, 1989) for solving the implicit update for Θ.

6.9 Supplement

The original versions of the SCEC CVM were based on empirical rule based velocity models to interpolate between the inferred boundaries of large geologic units (Magistrale et al., 2000, 1996). While rule-based models are necessarily simplified compared to the potential complexity of the real Earth in almost all cases, they are often useful from an interpretational standpoint, as rules correspond to real geological features, and additionally serve as a basis for combining disparate datasets within a common framework, such as was done in the initial construction of the

CVM models. Indeed, the level-set tomographic framework Muir and Tsai (2020b) used in our study is an extension of rule-based models to include more flexibility, and combine their benefits with those of standard tomographic models defined via basis function representations.

Within the LA basin, Magistrale et al. (1996) used the sedimentary compaction law of Faust (1951), which has the form $V_P = k(da)^{\frac{1}{6}}$, where d is the depth of maximum burial (corrected for any subsequent positive elevation), a is the age, and k is an calibration factor unique for each basin. Magistrale et al. (1996) used three boundaries - the basin bottom, pegged at an age of 20 Ma, the base of the Mohnian, at 14.5 Ma, and the base of the Repettian, at 4 Ma, with ages linearly interpolated in between these boundaries. The location of these boundaries, as well as the age of the surface, are derived from digitization of older geological studies, principally Yerkes et al. (1965) and Wright (1991). The uplift associated with the Pasadenan deformation is assumed to happen instantaenously at the present (i.e. the entire column is uplifted by an equivalent amount, rather than accounting for any potential deposition during uplift). Magistrale's model is relatively simple; however, such simplicity also results in greater interpretability. Given EKS sampling is a black-box approximate Bayesian method, it is feasible to "post-process" our inversions to interpret them in terms of the rule-based CVM definitions. We apply this to the major A-A' profile of Figure 12. We fix the lower basin boundary at the basin extracted from our inversion, and initialize the surface age, Repettian and Mohnian boundaries at their values in the early CVMs. We then perturb these using 1D Gaussian processes using a Matérn-3/2 kernel with unknown lengthscale and $\sigma = 1$ applied to the log surface-age and boundary depths, clamping the minimum V_P at 1.5 Km / s and using Brocher (2005) to convert to V_S and density. Due to the density of information (in this case fitting to an image, rather than the physical observables in the main inversion), we can use the relatively rough $3/2$ kernel to capture the details

without being concerned about artifacts. The output velocity model, including the locations of the reference and inverted boundaries, is shown in Supplementary Figure 6.15. The boundaries of both the Mohnian and Repettian units agree well with the well-constrained locations (from borehole measurements) in the southern part of the profile. In the northern part of the profile, the deep Mohnian profile agrees with results in Wright (1991), however the CVM1 velocity model specification requires a deep Repettian as well, which is not concordant with the outcropping of Puente and Topanga units at the surface in this area. It is unsurprising that the CVM1 rules provide an outcome inconsistent with the geology here, as they are developed primarily for the deeper basin, whereas our data suggests that the basin units are very shallow in the northern part of the profile. In the central part of the profile, where the reference interface locations are poorly controlled in Wright (1991), the observed velocity model in Figure 12 is best represented by a steep Repettian interface and a deep Mohnian, which conforms with our interpretation of the sharp gradient of the northern basin boundary being controlled by the influence of Quaternary faults, as even the youngest interface in the CVM1 model is highly deformed.

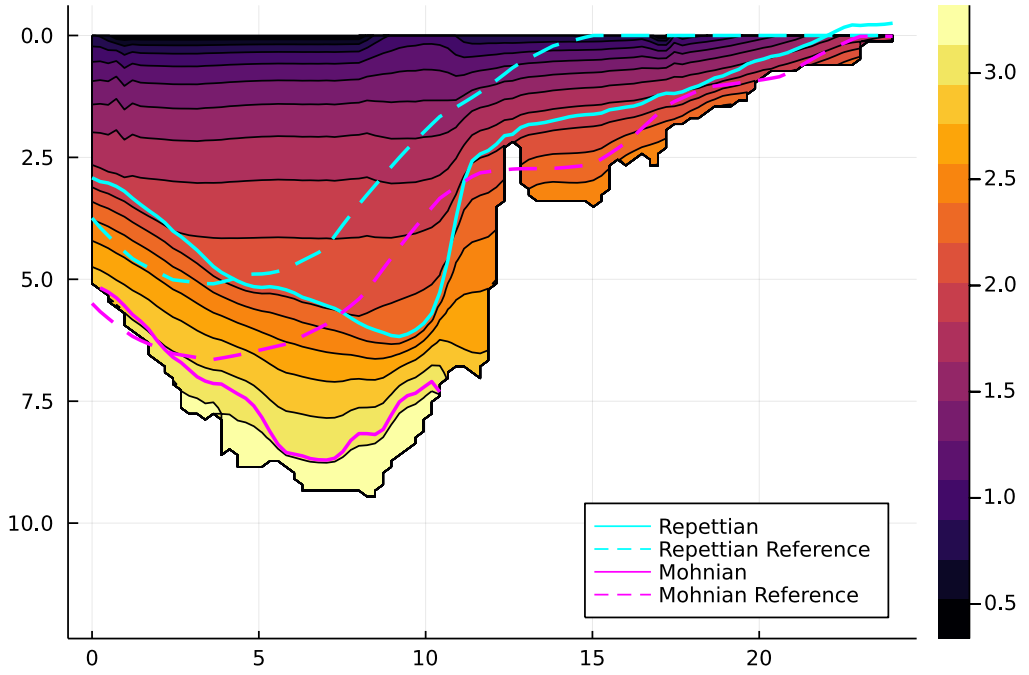


Figure 6.15: Fit of a modified SCEC CVM2 model to the A-A' profile results of Figure 12, with dashed lines showing the CVM2 reference surfaces (the bottom of the Repettian and Mohnian units) and the solid lines showing the inverted interfaces.

Chapter 7

DID OLDHAM DISCOVER THE CORE AFTER ALL? HANDLING IMPRECISE HISTORICAL DATA WITH HIERARCHICAL BAYESIAN MODEL SELECTION METHODS

Muir and Tsai (2020a) previously published as

Muir, J. B. & V. C. Tsai (2020). "Did Oldham discover the core after all? Handling imprecise historical data with hierarchical Bayesian model selection methods". In: *Seismological Research Letters* 91.3, pp. 1377–1383. DOI: 10.1785/0220190266

7.1 Abstract

Historical seismic data is essential to fill in the gaps in geophysical knowledge caused by the low rate of significant seismic events. Handling historical data in the context of geophysical inverse problems requires special care, due to the large errors in the data collection process. Using Oldham's data for the discovery of Earth's core as a case study, we illustrate how a hierarchical Bayesian model selection methodology using leave-one-out cross-validation can robustly and efficiently answer quantitative questions using even poor quality geophysical data. We find that there is statistically significant evidence for the existence of the core using only the P-wave data that Oldham effectively discarded in his discussion.

7.2 Introduction

Seismologists are highly motivated to study historical data due to the long timescales of geophysical processes compared to the human lifespan, and especially compared to the proliferation of modern digital instrumentation. A crucial consideration in seismology is that both the quantity and quality of seismic data are ever increasing,

and that as the density of instrumentation increases, so too does our ability to accurately locate events in space and time and the number of useful recorded events greatly increases. When dealing with historical data, we must therefore unfortunately contend with the reversal of these trends, so that we are left with fewer data of poorer quality. Overcoming these deficiencies requires careful treatment of noise in the data. The required tools are provided by Bayesian analysis, which allow us to rigorously derive posterior probability distributions for models given observed data and explicitly quantified priors (Tarantola, 2005). The ability of Bayesian analysis to quantitatively encode *a priori* information is especially important for historical data, where the information provided by the data is relatively uninformative.

In this study, we focus on a particular type of data: pairs of station-receiver distance and travel times from earthquakes. This data has long been central to geophysical imaging, especially before the advent of computationally feasible waveform inversions. Due to the computational expense of simulating waveforms and the requirement that waveform methods have an accurate starting model, seismic tomography from travel time data still holds a central position in the hierarchy of geophysical methods. When dealing with travel time data, the tomographer's hope is that errors in the spacetime location of an earthquake do not significantly contribute to the observed residuals used for inversion, or that they may at least be minimized by some relocation method. For historical data, the errors are often so large as to make this impossible, so any analysis requires that we explicitly handle errors in both distance and time. Classical regression methods such as Orthogonal Distance Regression can handle this case when the model to be fitted is smooth and the ratio between the errors for distance and time is known (Boggs and Rogers, 1990). However, as the error ratio is generally not known, a full analysis requires marginalizing over all possible reasonable combinations of errors. An analytical solution to this problem for linear models and specific non-informative priors is given in a manuscript by

Jaynes (Jaynes, 1999), left, like much of his work, unfinished by his death. For nonlinear models and arbitrary priors, a numerical approach is required. For this study, we present a Bayesian analysis for nonlinear models of imprecise data using Markov-Chain Monte-Carlo (MCMC) sampling, and show how to incorporate it into a model selection framework. We apply the model selection framework to some of the most historically important data ever presented in seismology—the famous travel time curve of R.D. Oldham (Oldham, 1906), demonstrating how model selection can make a concrete case for the seismic observation of the Earth's core using only a subset of Oldham's data.

It is well known in the seismological community that Oldham provided the first strongly accepted seismic evidence for the Earth's core in his seminal paper *The Constitution of the Interior of the Earth, as revealed by Earthquakes* (Oldham, 1906), for which he is generally credited with the seismic discovery of the core (Brush, 1980). Various geophysical arguments throughout the 1800s had suggested a core, most notably the arguments of Wiechert which determined the parameters of a core model from geodetic observations combined with the calculated moments of inertia of the Earth (Wiechert, 1897). However, direct observation of the core was unavailable until the development of quantitative seismology. Oldham provided a travel time curve for "Primary" and "Secondary" phases derived from teleseismic earthquake records, and correctly postulated their mechanical behavior as being those of P and S waves, respectively. The curvature of travel time strongly suggested to him that the waves travelled deeply in the Earth and were therefore capable of informing us about properties far into the interior. In the travel time curve, there is an apparent break in the behavior of the curves at around 120° epicentral distance, from which Oldham inferred the existence of the core. The change in character is much more apparent for the secondary arrivals than for the primary arrivals; indeed Oldham states that it would have *"probably remained undetected were it not for*

the very conspicuous alteration in the case of the second-phase waves". As such, Oldham predicated most of his argument on the secondary arrivals. Unfortunately, immediately after the publication of the original paper it became apparent that the change in secondary arrival behavior was in fact caused by the difference between S and SS phases, and the apparent large jump in travel time was not due to transmission through the core—consequently Oldham has to a certain extent been lauded for his discovery of the Earth's core under false pretenses (Brush, 1980)! Figure 7.1 shows Oldham's data, taken from Oldham (1906) Table 1 for averaged points and digitized from Oldham (1906) Figure 1 for non-averaged points, and overlaid with modern travel time curves from the ak135 model (Kennett et al., 1995). It is apparent that the later primary phase data are likely core interacting P phases of the PKP family. However, the scatter is extreme for the primary arrivals and it is difficult by eye to confidently claim that there is any meaningful change in the travel time curve. It is therefore a point of historical interest whether it is in fact possible to deduce the seismic existence of the core using only the primary (P) data presented in Oldham's paper. If we can show that there is a statistically significant change in behavior of the P travel time curve, then Oldham's deduction stands up even without the secondary (S, SS) data. Due to the highly imprecise nature of Oldham's data, this question provides an excellent case study for the handling of historical data using hierarchical Bayesian methods.

7.3 Data and Methodology

Oldham's P phase travel time data consist of distances $\mathbf{d}$ and times $\mathbf{t}$. To normalize the data to the interval $(-1, 1)$ for curve fitting, we subtract the mean and divide by the range of both $\mathbf{d}$ and $\mathbf{t}$.

We treat the question of detection of the core as one of quantitative Bayesian model selection. In particular, we propose several models for the data, some containing

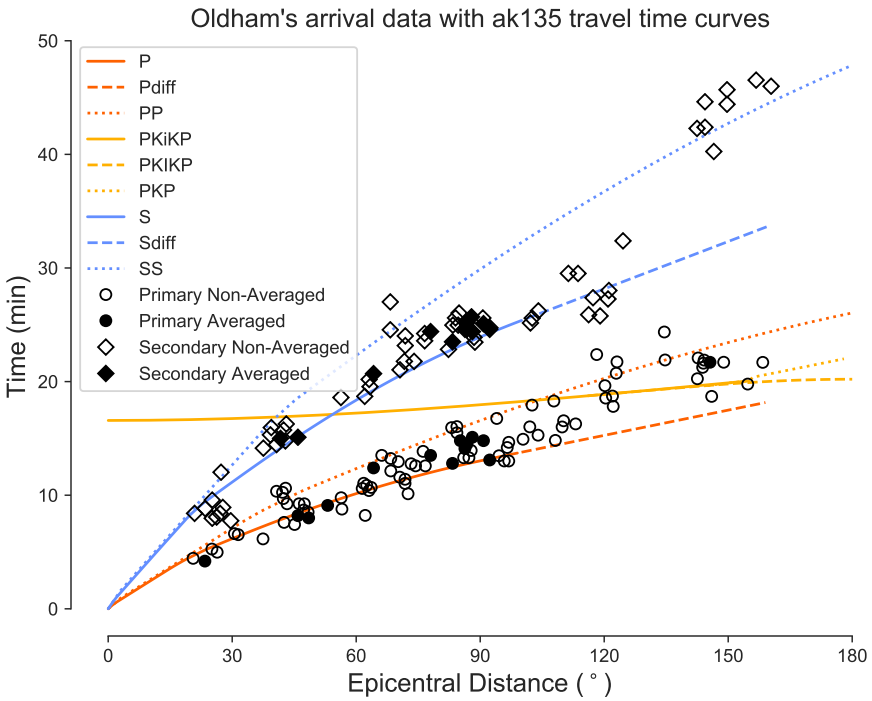


Figure 7.1: Data from Oldham (1906), with modern travel time curves from ak135 overlaid (Kennett et al., 1995). Assignments to Primary/Secondary arrivals are from Oldham.

only one predicted travel time curve, and some containing two. Following Oldham, we infer the presence of the core if a candidate model with two disjoint travel time curves is significantly better at predicting the data than models with only one curve. This decision criterion (i.e. whether to choose the model that has the best predictive performance) is a philosophical choice—for instance, if the true data generating process was one of the candidate models, it is not guaranteed that we would recover it (Shao, 1993). For geophysical data, however, the true data generating process is almost always unavailable and not included as a candidate; selecting a model that best predicts the observations is often the most sensible choice from a practical standpoint. The predictive criterion is data driven and does not explicitly penalize model complexity, but instead relies on the tendency of unnecessarily complex

models to overfit the data and therefore perform poorly at prediction for unseen data.

As mentioned before, Oldham's data is extremely imprecise both in time and epicentral distance compared to modern standards. As such, treating the data properly requires that we handle unknown errors on both axes. As earlier mentioned, frequentist methods such as orthogonal distance regression can fit curves to data with errors in both independent and dependent variables, but require *a priori* knowledge of the relative error. In contrast, hierarchical Bayesian methods allow us to set up the unknown error standard deviations σ_d and σ_t in both distance and time respectively as parameters that are inverted for along with the parameters describing the travel time model—as these parameters describe the form of the model likelihood and prior distributions they are referred to as hyperparameters in the Bayesian geophysical literature (Malinverno and Briggs, 2004). The model parameters are therefore $\mathbf{m}$, describing the form of the fitted travel time curves, and σ_d and σ_t. The functional form of the travel time curve, including any jumps, is given by $f(d, \mathbf{m})$. We assume that the model parameters $\mathbf{m}$ are *a priori* independent from σ_d and σ_t since making observations should not impact the type of models we propose for the travel time structure, and we also assume that σ_d and σ_t are independent since the scales of the errors in distance and travel time are not correlated.

To compare the predictive performance of different models and determine if there is enough evidence in Oldham's P arrival to indicate the existence of the core, we must first derive the posterior distribution for the parameters of the different models given the observed data. From Bayes' rule, the posterior distribution is given by

$$p(\mathbf{m}, \sigma_d, \sigma_t|\mathbf{d}, \mathbf{t}) \propto p(\mathbf{d}, \mathbf{t}|\mathbf{m}, \sigma_d, \sigma_t)p(\mathbf{m}, \sigma_d, \sigma_t) = p(\mathbf{d}, \mathbf{t}|\mathbf{m}, \sigma_d, \sigma_t)p(\mathbf{m})p(\sigma_d)p(\sigma_t), \tag{7.1}$$

assuming independent priors for $\mathbf{m}$, σ_d, and σ_t. Both $\mathbf{t}$ and $\mathbf{d}$ have significant noise,

so we represent the relationship between them as

$$\mathbf{t} = f(\mathbf{d} + \mathbf{e}_d, \mathbf{m}) + \mathbf{e}_t, \tag{7.2}$$

where $\mathbf{e}_d$ and $\mathbf{e}_t$ are the unknown measurement errors in distance and time, respectively. This formulation implicitly assumes errors in distance and time are independent, which given the majority of earthquake origins in Oldham (1906) are from local reports, rather than by triangulation from travel time, is not unreasonable. As we are ranking different functional forms f, we do not include model uncertainty in this analysis.

To write out the posterior, we introduce dummy variables $\mathbf{D} = \mathbf{d} + \mathbf{e}_d$ and $\mathbf{T} = f(\mathbf{D}, \mathbf{m})$; from a Bayesian standpoint $\mathbf{D}$ represents the unknown "true" distances in Oldham's data and $\mathbf{T}$ the corresponding "true" times predicted by the travel time curve. Note that $\mathbf{T}$ is a deterministic function of $\mathbf{D}$ and $\mathbf{m}$ so $p(\cdot|\mathbf{D}, \mathbf{m}) = p(\cdot|\mathbf{T})$. Assuming that $\mathbf{D}$ is independent of $\mathbf{m}$ and given our earlier assumption that uncertainties in $\mathbf{d}$ and $\mathbf{t}$ are independent, we can write the likelihood as a marginalized distribution over $\mathbf{D}$

$$p(\mathbf{d}, \mathbf{t}|\mathbf{m}, \sigma_d, \sigma_t) = \int p(\mathbf{d}, \mathbf{t}|\mathbf{D}, \mathbf{m}, \sigma_d, \sigma_t) p(\mathbf{D}) d\mathbf{D} \tag{7.3a}$$

$$= \int p(\mathbf{d}|\mathbf{D}, \mathbf{m}, \sigma_d) p(\mathbf{t}|\mathbf{D}, \mathbf{m}, \sigma_t) p(\mathbf{D}) d\mathbf{D} \tag{7.3b}$$

$$= \int p(\mathbf{d}|\mathbf{D}, \sigma_d) p(\mathbf{t}|\mathbf{T}, \sigma_t) p(\mathbf{D}) d\mathbf{D}, \tag{7.3c}$$

allowing us to write a fully decoupled marginal posterior

$$p(\mathbf{m}, \sigma_d, \sigma_t|\mathbf{d}, \mathbf{t}) \propto \int p(\mathbf{d}|\mathbf{D}, \sigma_d) p(\mathbf{t}|\mathbf{T}, \sigma_t) p(\mathbf{m}) p(\sigma_d) p(\sigma_t) p(\mathbf{D}) d\mathbf{D}. \tag{7.4}$$

The full posterior, including the dummy variables $\mathbf{D}$, can be written by de-marginalizing Equation 7.4 and applying Bayes' theorem to obtain

$$p(\mathbf{m}, \sigma_d, \sigma_t, \mathbf{D}|\mathbf{d}, \mathbf{t}) \propto p(\mathbf{D}|\mathbf{d}, \sigma_d) p(\mathbf{t}|\mathbf{T}, \sigma_t) p(\mathbf{m}) p(\sigma_d) p(\sigma_t). \tag{7.5}$$

The hierarchical parameterization used in this study may cause difficulties in efficient MCMC sampling due to the structure of the posterior—we discuss a method to avoid this issue in the Supplement. The inclusion of noise in the independent variable $\mathbf{d}$ means that the final inverse problem has a free parameter corresponding to every data pair $(\mathbf{d}_i, \mathbf{t}_i)$, plus those used to specify the error scales σ_d and σ_t and the model variables $\mathbf{m}$, meaning that the problem is fundamentally underdetermined and requires careful selection of priors. Additionally, as the posterior is relatively high-dimensional, explicit integration over it is intractable. We use Markov-Chain Monte-Carlo (MCMC) to calculate integrals with respect to the posterior, specifically using Hamiltonian Monte Carlo (HMC) to sample the high dimensional posterior efficiently (Neal, 2011).

Once the posterior distributions for the candidate models are determined, a metric for comparing them for model selection must be defined. To determine the relative performance of the candidate models, we use leave-one-out cross-validation (LOO-CV). LOO-CV estimates the predictive performance of a model by removing datums from the observations one at a time, fitting the model, and then testing the left out datum against the model predictions. The posterior predictive distribution for the i^{th} left out datum is $p(\mathbf{d}_i, \mathbf{t}_i|\mathbf{d}_{j\neq i}, \mathbf{t}_{j\neq i})$. The LOO-CV estimate for N_{data} data points is given by

$$\text{LOO-CV} = \sum_{i}^{N_{data}} \log p(\mathbf{d}_i, \mathbf{t}_i|\mathbf{d}_{j\neq i}, \mathbf{t}_{j\neq i}). \tag{7.6}$$

For each left out datum, we use MCMC sampling to draw N_{MCMC} samples from the marginal posterior $p(\mathbf{m}, \sigma_d, \sigma_t|\mathbf{d}_{j\neq i}, \mathbf{t}_{j\neq i})$—by using MCMC sampling we avoid explicitly integrating over the nuisance parameters $\mathbf{D}_{j\neq i}$. By writing $p(\mathbf{d}_i, \mathbf{t}_i|\mathbf{d}_{j\neq i}, \mathbf{t}_{j\neq i})$ as a marginalization of the posterior predictive $p(\mathbf{d}_i, \mathbf{t}_i|\mathbf{m}, \sigma_d, \sigma_t)$ with respect to the

held out data, we can estimate $p(\mathbf{d}_i, \mathbf{t}_i | \mathbf{d}_{j\neq i}, \mathbf{t}_{j\neq i})$ using the MCMC draws

$$p(\mathbf{d}_i, \mathbf{t}_i | \mathbf{d}_{j\neq i}, \mathbf{t}_{j\neq i}) = \int p(\mathbf{d}_i, \mathbf{t}_i | \mathbf{m}, \sigma_d, \sigma_t) p(\mathbf{m}, \sigma_d, \sigma_t | \mathbf{d}_{j\neq i}, \mathbf{t}_{j\neq i}) d\mathbf{m} d\sigma_d d\sigma_t \quad (7.7)$$

$$\approx \frac{1}{N_{MCMC}} \sum_{n=1}^{N_{MCMC}} p(\mathbf{d}_i, \mathbf{t}_i | \mathbf{m}_n, \sigma_{d,n}, \sigma_{t,n}), \quad (7.8)$$

where $\mathbf{m}_n$, $\sigma_{d,n}$, $\sigma_{t,n}$ notate the n^{th} MCMC sample of the posterior for the model parameters. For each MCMC sample, we can then calculate

$$p(\mathbf{d}_i, \mathbf{t}_i | \mathbf{m}_n, \sigma_{d,n}, \sigma_{t,n}) = \int p(\mathbf{d}_i, \mathbf{t}_i | \mathbf{m}_n, \sigma_{d,n}, \sigma_{t,n}, \mathbf{D}_i) p(\mathbf{D}_i) d\mathbf{D}_i \quad (7.9)$$

by explicit numerical integration. Higher values of the LOO-CV score indicate better predictive performance. We therefore infer the presence of the core from Oldham's data if a model with a jump in the P travel time has a LOO-CV score at least one standard error higher than all models without a jump. We chose LOO-CV as its estimates of predictive performance are robust and unbiased (Vehtari and Ojanen, 2012). LOO-CV is quite computationally intensive, as it requires an independent MCMC run for each datum, which could motivate the use of less expensive methods such as k-fold cross-validation for large datasets. However, Oldham's data consists of only 90 points, so explicit LOO-CV is feasible—we further discuss these convergence performance considerations in the Supplement.

We use low-degree Chebyshev polynomials of the first kind T_i to define the model travel time curves, following Oldham's expectation that individual travel time curves for a single phase should be smooth. For models that contain a jump in the travel time curve, we use two polynomials to represent the curve before and after the changepoint. We notate the models as $(a, -)$ for single travel time curves of degree a with $-$ signifying no second travel time curve for P arrivals, and (a, b) for double travel time curves of degrees a and b with a jump in travel time. For a single travel time curve of degree a

$$f(x, \mathbf{m}) = \sum_{i=0}^{a} \mathbf{m}_i T_i(x). \quad (7.10)$$

For double travel time curves, the model parameter vector m contains two sets of polynomial coefficients and the location of the changepoint $\mathbf{D}_j$. The HMC method requires that the posterior be continuously differentiable, so a model containing two travel time curves with a true discontinuity between them cannot be used. To model the jump in travel time between curves, we instead use a hyperbolic tangent to transition between two polynomials f_1 and f_2 so that

$$f(\mathbf{D}, \mathbf{m}) = \frac{f_1(\mathbf{D}, \mathbf{m})(1 - \tanh(k(\mathbf{D} - \mathbf{D}_j)) + f_2(\mathbf{D}, \mathbf{m})(1 + \tanh(k(\mathbf{D} - \mathbf{D}_j))}{2}, \quad (7.11)$$

where the factor of $k = 1000$ ensures that the jump is very sharp, but still continuous.

We use products of univariate normal distributions for $p(\mathbf{d}|\mathbf{D}, \sigma_d)$ and $p(\mathbf{t}|\mathbf{T}, \sigma_t)$, which is appropriate given we expect the data to be independent, and the distribution of residuals is approximately normal. For the prior $p(\mathbf{m})$, we also use a product of normals with large standard deviation ($\sigma = 10$); the purpose of this choice is to constrain $\mathbf{m}$ to reasonable parameter ranges for Chebyshev polynomials on the interval $(-1, 1)$, which is important for the two travel time case where the second polynomial may rely on very little data. This choice of prior contains all physically reasonable travel time curves, and so is only as informative as is required to make the posterior sensible. For the dummy variables $\mathbf{D}$, we use an uninformative uniform prior on the whole real line as the distribution of measurment distances is *a priori* unknown. Based on visual inspection of the data, we use a uniform prior on the range $(90°, 130°)$ for the changepoint $\mathbf{D}_j$. Finally, setting the priors for σ_d and σ_t requires special attention. Because the data is highly scattered, the error parameters trade off very strongly with one another, which can lead to parts of the posterior distribution being so highly curved that MCMC sampling is not feasible despite the rescaling mentioned above. These situations have unreasonable choices of the parameter space, with σ_d being almost zero while σ_t is extremely large, or vice versa. Solving this issue requires putting an informative prior in the error parameters that

stops either σ from being very small or very large, which is justified as it is *a priori* clear that the errors in both distance and time are significant but finite. We use inverse-gamma priors tuned so that $P(\sigma_d \leq 1.38°) = P(\sigma_d \geq 13.8°) = 0.01$ and $P(\sigma_t \leq 0.202 \text{ min}) = P(\sigma_t \geq 2.02 \text{ min}) = 0.01$, which are ranges we feel are reasonable for the error standard deviations given Oldham's description of the data.

7.4 Results

We used the STAN HMC sampler (Carpenter et al., 2017) to calculate the LOO-CV score for seven models, detailed in Table 7.1. 5000 samples were generated for 6 chains, with the first 2500 discarded, and the chains were compared to ensure convergence. The best performing model was (1, 1), which fits a line to the first part of the data and a line to the data after the jump in travel times. The difference in LOO-CV scores for all models relative to (1, 1) is given in Table 7.2. Both models with two travel time curves separated by a jump are favored over all models without a jump by at least 7 times the standard error, indicating that within the context of the models chosen, there is a very significant change in the behavior of the travel time curve despite the large scatter in the data. As such, the P phase arrival data alone are sufficient to support Oldham's arguments as to the existence of the core in Oldham (1906). However, the data are not sufficient to distinguish between a quadratic or a line for the first part of the travel time curve, as the difference between models (1, 1) and (2, 1) is not significant. We show model (3, −), which is the best performing model with only one travel time curve, in Figure 7.2 a) and model (1, 1), the best performing model with two travel time curves, in Figure 7.2 b). Note that the mean travel time curve shown for Figure 7.2 b) is smoother than any individual sample of model (1, 1), which has a sharp jump between the two travel time branches. The mean model predictions for both (3, −) and (1, 1) fall between the P/Pdiff and PP phases of the ak135 model for distances less than 120°, suggesting

that Oldham's data is potentially a mix of these phases; for greater distances the better performing $(1, 1)$ model sits between ak135 PKiKP and PP at around 120° before moving towards what are likely to be P core phases at higher distances. The presence of some PP data may explain why the posterior mean of model $(1, 1)$ has negative slowness after the jump in travel time, although the majority of post jump data is closer to the modern core phase times.

From examining the spread of the 90% credible interval (i.e. the area between the 5% and 95% quantiles of the posterior model distribution at each epicentral distance), we can see that $(3, -)$ is more tightly constrained than $(1, 1)$, at the expense of fitting the data significantly worse. For both models, the average correction increases as a function of distance, as is seen in the length of the red connecting lines in Figure 7.2, indicating that measurements were generally worse fit by a single travel time curve at longer epicentral distances. The LOO-CV score balances these concerns and strongly favors models models with two travel time curves. The LOO-CV score degrades substantially from model $(3, -)$ to model $(4, -)$ and model $(5, -)$, which indicates that further higher degrees would perform yet worse in predicting held out data since the single travel time curve models are strictly nested (i.e. $(5, -)$ contains all of $(4, -)$, which contains all of $(3, -)$ as a special case). Overfitting becomes significant even for simple polynomials due to the high scatter in the data.

Model	Two travel time curves?	Total free parameters
$(1, -)$	No	94
$(2, -)$	No	95
$(3, -)$	No	96
$(4, -)$	No	97
$(5, -)$	No	98
$(1, 1)$	Yes	97
$(2, 1)$	Yes	98

Table 7.1: Catalog of models used to test Oldham's data

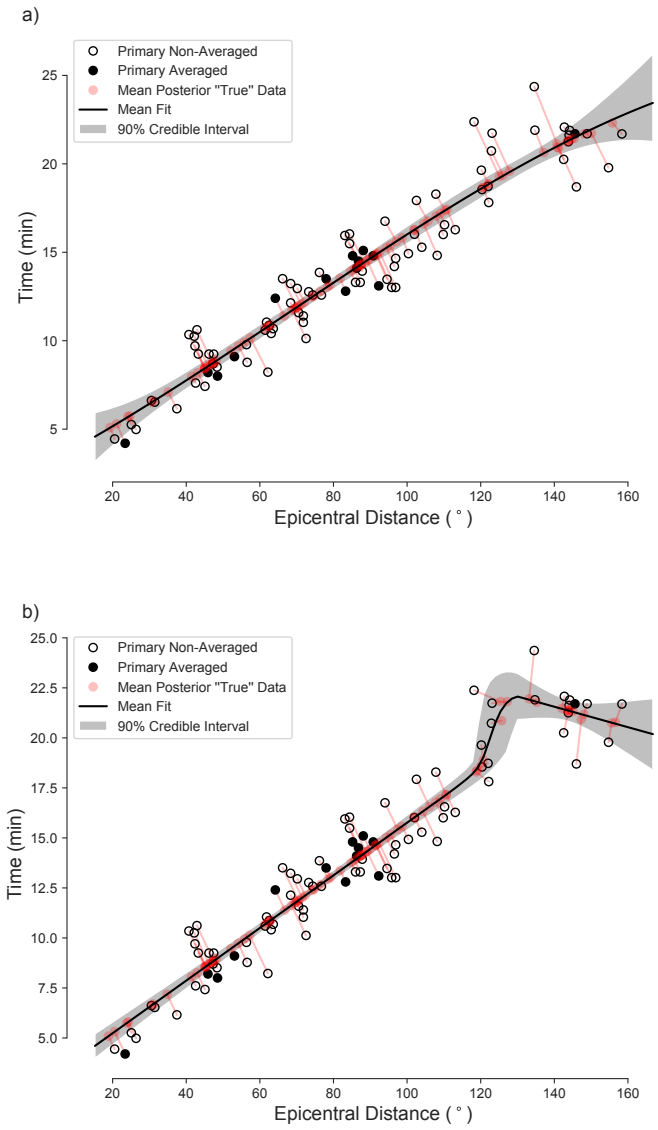


Figure 7.2: a) Posterior distribution for model $(3, -)$, the best performing single travel time curve model. b) Posterior distribution for model $(1, 1)$, the best performing model. Note that the mean shown for b) is smoother than an individual sample of model $(1, 1)$.

Model	LOO-CV score	Difference in LOO-CV	Standard error in difference
(1, –)	23.99	-3.91	0.44
(2, –)	24.03	-3.87	0.40
(3, –)	23.12	-3.78	0.38
(4, –)	23.14	-4.76	0.69
(5, –)	22.55	-5.35	0.95
(1, 1)	27.90	—	—
(2, 1)	27.85	-0.06	0.14

Table 7.2: Difference in LOO-CV score relative to (1, 1), the best performing tested model.

7.5 Discussion and Conclusions

Model selection, provided by the LOO-CV score, strongly supports there being enough evidence solely in Oldham's P-wave data to support two apparently distinct travel time curves, which leads to Oldham's argument for the core. While the scientific question presented in this study has not been in question for over 100 years, the robust statistical tools required to analyze the problem fully have only recently become available. Model selection, in particular, remains at the forefront of statistical research and has great implications for both traditional inverse theory and newer techniques such as machine learning (e.g. Claeskens (2016); Rasmussen and Williams (2006); Wit et al. (2012)). The problem of how to perform model selection is unfortunately less resolved than that of how to sample from the Bayesian posterior, for which MCMC sampling, and in particular HMC, has emerged as the clearly preferred technique (Betancourt, 2017; Fichtner et al., 2019; Neal, 2011). Model selection, in contrast, has a plethora of related techniques, ranging in complexity from penalized fits to the *maximum a posteriori* point such as the Akaike and Bayesian information criteria (Claeskens, 2016), cross-validation methods such as that presented here (Vehtari and Ojanen, 2012), and extending to full calculation of the Bayesian evidence. The Bayesian evidence or Bayes factor calculation, in particular, has received attention in geophysics and astronomy since it can be cleanly derived from Bayes' theorem as explicitly comparing the probability of two models

given the data. Unfortunately, estimating the Bayesian evidence is highly nontrivial and is typically only shown in the literature for low dimensional models due to convergence difficulties, which limits its utility for realistic geophysical problems (Friel and Wyse, 2012; Vehtari and Ojanen, 2012). The LOO-CV method used here proved to be tractable for models with $\sim$ 100 parameters, however, it does suffer from computational difficulties as the number of data becomes large. Vehtari et al. (2017) give a method for using importance sampling on the posterior MCMC samples using *all* data to approximate the results from held out data, which is promising for large geophysical data sets.

In this study, we have tested only simple functional forms for travel time curves with and without jumps, with the further restriction that after the jump point all data are assigned to the second travel time curve. This restriction means that there are no overlapping travel time curves at any epicentral distance, which restricts our analysis to datasets for which there are not multiple groups of phases observed at a particular distance. Visual inspection of Oldham's data suggests that this simple model is the highest level of complexity warranted by the data. With modern seismic data, however, it is likely that observations at a particular distance will contain multiple phases that need to be classified into different classes. In this case, more advanced modeling strategies that allow the expression of uncertainty as to what phase is being observed, such as Gaussian Mixture Modeling, may be useful (e.g. Grana et al. (2017)).

Our study shows how to set up a model that marginalizes over multiple potential sources of error and can be efficiently sampled using HMC. We have shown how careful specification of the prior is especially important in the historical context, where the scale of the data errors are unknown and multiple sources of error may trade off. Together, hierarchical Bayesian modeling and model selection give us a powerful toolbox to explore poor quality historical data and derive robust conclusions

about geophysical processes. In the context of Oldham's travel time data, it allows us to marginalize out the large errors associated with both the distances and travel times to conclude that there is sufficient evidence contained in the P arrivals alone to indicate the existence of the core.

7.6 Acknowledgements

The authors would like to thank two anonymous reviewers for their feedback, and the SRL editor-in-chief Allison Bent, for managing the review process. The authors would also like to thank Luis Rivera for providing an internal review of this article. JBM would like to thank the General Sir John Monash Foundation and the Origin Energy Foundation for financial support during his graduate studies.

7.7 Data and Resources

Historical data were taken from Oldham (1906), either from the reported tables of averaged events or by digitizing the presented travel time curves. All calculations were performed using the PyStan wrapper of the Stan statistical software package (Carpenter et al., 2017). Inversion results for the five models not presented in the paper are found in Figures S1-S5. Additional discussion regarding hierarchical MCMC sampling and LOO-CV vs k-fold CV are also present in the supplement.

7.8 Supplement

Avoiding the hierarchical "funnel" effect

The distribution $p(\boldsymbol{d}|\boldsymbol{D}, \sigma_d)$ is highly curved as a function of σ_d, an effect popularly known as the hierarchical "funnel" (Betancourt and Girolami, 2013). Intuitively, as the scale σ_d decreases, the $\boldsymbol{d}$ and $\boldsymbol{D}$ are forced to be very close. This high curvature greatly impedes the performance of sampling algorithms. To avoid this problem, we rescale $\boldsymbol{D}$ by σ_d, writing $\boldsymbol{D} = \boldsymbol{d} + \sigma_d \bar{\boldsymbol{D}}$ which serves to decouple the scale of $\bar{\boldsymbol{D}}$ from σ_d, and to center it around zero. If the likelihood $p(\boldsymbol{d}|\boldsymbol{D}, \sigma_d)$ is a product of normal

distributions $\boldsymbol{d}_i \sim N(\boldsymbol{D}_i, \sigma_d)$, then this rescaling is equivalent to the likelihood for $\bar{\boldsymbol{D}}_i \sim N(0, 1)$. Equation 7.9 in the main text is then written

$$\int p(\boldsymbol{d}_i, \boldsymbol{t}_i | \boldsymbol{m}_n, \sigma_{d,n}, \sigma_{t,n}, \boldsymbol{D}_i) p(\boldsymbol{D}_i) d\sigma \boldsymbol{D}_i = \int \sigma_{d,n} p(\boldsymbol{d}_i, \boldsymbol{t}_i | \boldsymbol{m}_n, \sigma_{d,n}, \sigma_{t,n}, \bar{\boldsymbol{D}}_i) p(\bar{\boldsymbol{D}}_i) d\bar{\boldsymbol{D}}_i. \tag{7.12}$$

We give uniform priors fo $\bar{\boldsymbol{D}}_i$ as the distribution of measurement distances is a priori unknown.

Performance considerations of LOO-CV $(k = 1)$ **vs k-fold CV** $(k > 1)$

The integral in main text Equation 9 must be calculated explicitly, as it is with respect to the uninformative prior density $p(\boldsymbol{D}_i)$, meaning convergence of an MCMC estimate will be slow. For this reason, cross-validation methods with multiple held out data points such as k-fold cross-validation quickly become computationally infeasible, as both the multidimensional explicit integrals over multiple held out data points, and the MCMC estimates of these integrals, will converge very slowly.

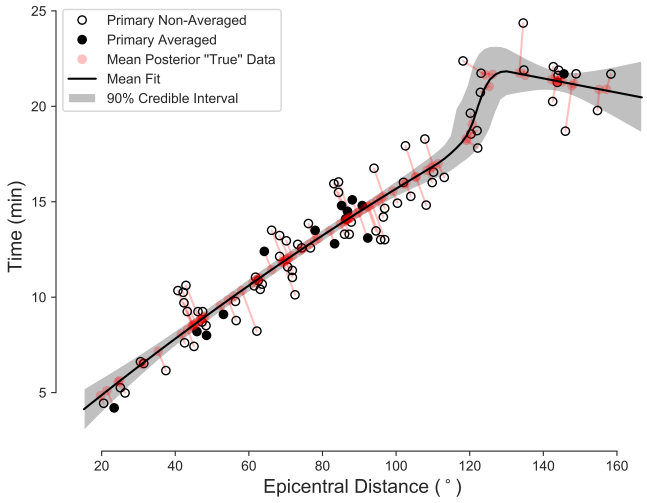


Figure 7.3: Posterior distribution for model (2,1)

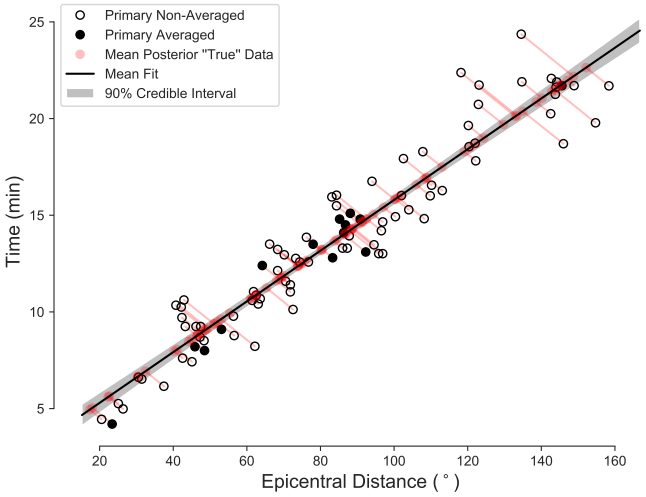


Figure 7.4: Posterior distribution for model (1,-)

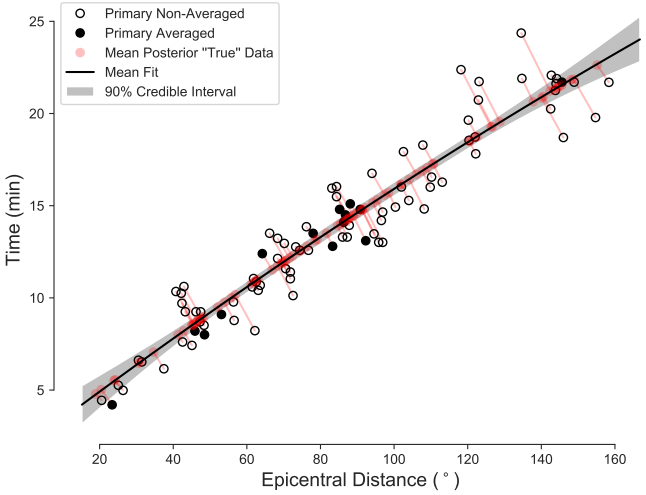


Figure 7.5: Posterior distribution for model (2,-)

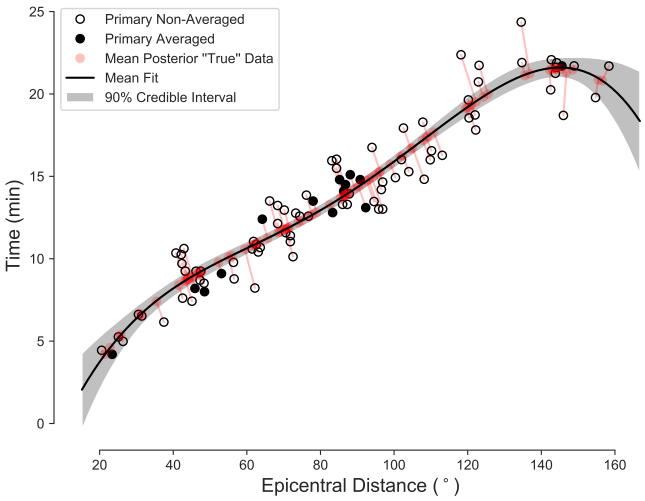


Figure 7.6: Posterior distribution for model (3,-)

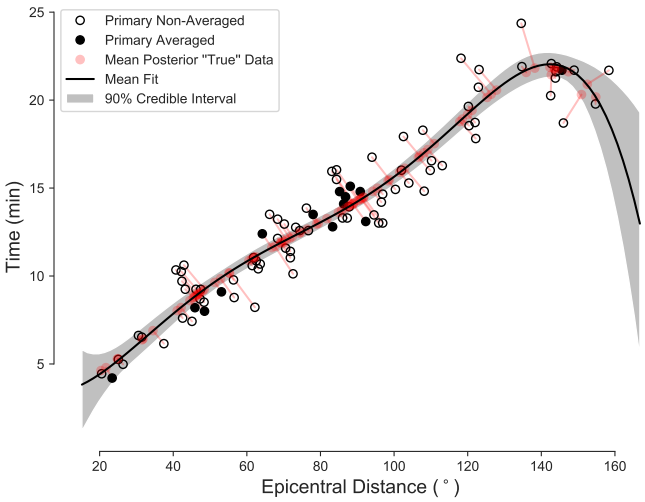


Figure 7.7: Posterior distribution for model (4,-)

Chapter 8

CONCLUSION

Seismic inversion remains a key component of the development of our understanding of the Earth. The overarching theme of this thesis is that inversions, whether in the data space or the model space, may be improved by designing them to better exploit pre-existing geological or geophysical knowledge. In particular, we have looked at how we can design an appropriate combination of model parameterization and regularization to achieve our goals, whether they be to generate optimal seismic data interpolations that respect wave physics, or allow the principled introduction of explicit geological features in structural inversion via the level set method. We have also explored how best to select between different potential classes of model in a geophysical setting, through both the use of hierarchical methods, and post-hoc selection criteria such as the data-driven cross-validation methods or through statistically motivated information criteria. The methodologies presented in this thesis have been extensively tested against both synthetic and real-world datasets, and are suitable for future use by the research community.

8.1 Future Outlook for Incorporating *a priori* Knowledge in Geophysical Inversion

The ultimate goal of geophysical imaging is to enhance geological or geodynamical understanding, however bridging the divide between inversion outcome and interpretation is difficult. Existing methods to regularize geophysical imaging blur the structures within the Earth, hindering geodynamical interpretability. As such, many of the fundamental questions about Earth structure remain unanswered, e.g., does the entire mantle recycle or are there long-lived reservoirs of primordial material?

How does nonuniform heat flux across the core-mantle boundary affect the magnetic field of the Earth and its reversals? How efficiently does the subducting crust transport water into the interior and what effect does that have on the viscosity governing convection? Answering these questions will require true multiscale imaging of the deep Earth. Recent developments in machine learning have allowed alternative forms of regularization based on deep networks, that assimilate the features of target structures and deliver resulting images that reflect these features. These advanced regularization schemes provide a principled basis on which to input complex and detailed *a priori* geophysical constraints into tomographic problems — for instance, promoting the recovery of features obtained from high resolution computational geodynamical models within the context of global seismic imaging.

Deep-learning based inverse problem solvers have to date typically employed a scheme where observed data is mapped directly into the desired physical model space through the application of a multi-layer network. This is distinctive compared to traditional inverse-problem solutions in that there is an absence of the forward problem - inverse problem loop, as the forward model is never evaluated for the observed data. To train these networks, a large number of data-model pairs must be constructed; this is potentially infeasible for large-scale seismic tomography where the solution of the forward problem is a very computationally intensive effort. Additionally, in the seismic tomography problem there exist no "ground truth" models except in very limited lab settings, so any such mapping would necessarily be based on theoretical earth models. Instead, it is possible to leverage much of the existing machinery of seismic tomography, treating the imaging problem as a non-linear inverse problem to be solved by traditional optimization or sampling methods rather than via push-forward through a deep neural network operator. In order to weight our a priori knowledge, two methods may be employed. The first is to explicitly parametrize the model in terms of geological features, such as

we developed in Chapters 5 and 6. To create complex models that incorporate a priori knowledge without being overly proscriptive, a deep-neural network can be employed in a generative fashion, taking inputs in an abstract space and mapping them into a physical model, for which the predicted data is calculated and the inverse problem solved as normal. While these methods are powerful, they render the mere creation of the model highly nonlinear, and also restrict the model to lie within the range of the data used to train the generative process. The second method is to utilize regularization, in conjunction with a simple voxel-based parameterization. As has been often noted in this thesis, all realistic inverse problems in geophysics are ill-posed and require regularization to enable a unique solution to be found. Traditional regularization methods have utilized simple principles such as the penalizing the amplitude or roughness of model. Using deep-learning methods, more informative types of regularization may be encoded by learning to extract features from synthetic structural models that reflect our understanding of Earth's evolution—this type of approach has been termed NETT (network Tikhonov) by Li et al. (2020a). The advantage of a regularization-based approach to encoding a priori knowledge is that it can fully utilize the existing structure of large-scale seismic inversion, as inversion is still performed in the physical model space. This represents a considerable advantage over other methods as the infrastructure required to solve global seismic imaging problems is non-trivial (Krischer et al., 2016). As an example of a potential application of this method, by testing a suite of regularisers conditioned on a variety of modes of mantle convection, and assessing their performance, it would be possible to improve the interpretability and quantification of uncertainty of seismic images and allow beyond state-of-the-art hypothesis testing of different models of mantle convection. Future work in data-space inversion will also lead to an improvement in our ability to utilize "tomography free" methods, that may be able to operate using short-time deployments of large arrays; we discuss some future prospects for

advancing data-space inversions in Chapter 3.

8.2 Summary

In this thesis, we presented four studies representing the full spectrum of the geophysical inversion process. In Chapter 2, we describe a structural inversion of the upper crust of Southern California using ambient-noise derived Rayleigh wave horizontal-to-vertical (H/V) ratios, finding that H/V ratios provided a strong independent constraint for the slow basin velocities in the Los Angeles area and the Salton Trough. In Chapters 3 and 4, we developed a wavefield-reconstruction method based on a temporal wavelet transform combined with spatial compressive sensing using a preconditioned curvelet basis. This set of algorithms greatly improves the condition of interpolated wavefields on sparse networks, and provides a quantitative framework for assimilating heterogeneous strain and displacement datasets into a single data product. In Chapters 5 and 6, introduced a new parameterization method for seismic studies, the combined Gaussian process and level set parameterization, that allows researchers to include discontinuous geological features into seismic inverse problems in a particularly simple way. We applied our method to several synthetic and real world datasets, culminating in a study of the northeastern Los Angeles basin using Love wave energy originating from the Mw 6.4 and Mw 7.1 July 2019 Ridgecrest earthquakes as recorded on the Community Seismic Network. We found that the observed data required a steeper and deeper northeastern basin edge, changing the outlook for ground motion amplification in downtown Los Angeles. Finally, we investigated robust Bayesian model selection methods for noisy historical datasets, correcting the interpretation of R. D. Oldham's famous discovery of the Earth's core along the way. In all cases, by making full use of the suite of inversion techniques available to us, we are able to advance our knowledge of Earth processes beyond what we could do using traditional methods — and by maintaining a strong focus on

methodological development into the future, important discoveries about our world are sure to follow.

BIBLIOGRAPHY

Akciz, S. O., L. Grant Ludwig, and J. R. Arrowsmith
2009. Revised dates of large earthquakes along the Carrizo section of the San Andreas Fault, California, since A.D. 1310 ± 30. *Journal of Geophysical Research: Solid Earth*, 114(B1).

Aki, K., A. Christoffersson, and E. S. Husebye
1977. Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, 82(2):277–296.

Aki, K. and P. G. Richards
2002. *Quantitative Seismology*. University Science Books.

Ardhuin, F., L. Gualtieri, and E. Stutzmann
2015. How ocean waves rock the Earth: Two mechanisms explain microseisms with periods 3 to 300 s. *Geophysical Research Letters*, 42(3):765–772.

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

Backus, G. and F. Gilbert
1968. The resolving power of gross earth data. *Geophysical Journal International*, 16(2):169–205.

Bakku, S. K.
2015. *Fracture Characterization from Seismic Measurements in a Borehole*. PhD thesis, Massachusetts Institute of Technology.

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

Berg, E. M., F.-C. Lin, A. Allam, H. Qiu, W. Shen, and Y. Ben-Zion
2018. Tomography of Southern California Via Bayesian Joint Inversion of Rayleigh Wave Ellipticity and Phase Velocity From Ambient Noise Cross-Correlations. *Journal of Geophysical Research: Solid Earth*, 123(11):9933–9949.

Berg, E. M., F.-C. Lin, A. Allam, V. Schulte-Pelkum, K. M. Ward, and W. Shen
2020. Shear Velocity Model of Alaska Via Joint Inversion of Rayleigh Wave Ellipticity, Phase Velocities, and Receiver Functions Across the Alaska Transportable Array. *Journal of Geophysical Research*, 125:22.

Betancourt, M.
2017. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [stat]*.

Betancourt, M. J. and M. Girolami
2013. Hamiltonian Monte Carlo for Hierarchical Models. *arXiv:1312.0906 [stat]*.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann
2010. ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, 81(3):530–533.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah
2017. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98.

Bhatia, H., G. Norgard, V. Pascucci, and P.-T. Bremer
2013. The Helmholtz-Hodge Decomposition—A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(8):1386–1404.

Bianco, M. J. and P. Gerstoft
2018. Travel Time Tomography With Adaptive Dictionaries. *IEEE Transactions on Computational Imaging*, 4(4):499–511.

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

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

Boggs, P. T. and J. E. Rogers
1990. Orthogonal distance regression. *Contemporary Mathematics*, 112:183–194.

Boore, D. M. and M. N. Toksoz
1969. Rayleigh Wave Particle Motion and Crustal Structure. *Bulletin of the Seismological Society of America*, 59(1):331–&.

Boschi, L. and C. Weemstra
2015. Stationary-phase integrals in the cross correlation of ambient noise. *Reviews of Geophysics*, 53(2):411–451.

Bowden, D. C. and V. C. Tsai
2017. Earthquake ground motion amplification for surface waves: Ground Motions for Surface Waves. *Geophysical Research Letters*, 44(1):121–127.

Bowden, D. C., V. C. Tsai, and F. C. Lin
2015. Site amplification, attenuation, and scattering from noise correlation amplitudes across a dense array in Long Beach, CA. *Geophysical Research Letters*, 42(5):1360–1367.

Bowden, D. C., V. C. Tsai, and F.-C. Lin
2017. Amplification and Attenuation Across USArray Using Ambient Noise Wavefront Tracking: USArray Noise Amplitudes. *Journal of Geophysical Research: Solid Earth*, 122(12):10,086–10,101.

Brissaud, Q., D. C. Bowden, and V. C. Tsai
2020. Extension of the Basin Rayleigh-Wave Amplification Theory to Include Basin-Edge Effects. *Bulletin of the Seismological Society of America*, 110(3):1305–1322.

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

Brush, S. G.
1980. Discovery of the Earth's core. *American Journal of Physics*, 48(9):705–724.

Burnham, K. P. and D. R. Anderson
2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2):261–304.

Candès, E., L. Demanet, D. Donoho, and L. Ying
2006. Fast Discrete Curvelet Transforms. *Multiscale Modeling & Simulation*, 5(3):861–899.

Candes, E. J. and L. Demanet
2005. The curvelet representation of wave propagators is optimally sparse. *Communications on Pure and Applied Mathematics*, 58(11):1472–1528.

Candes, E. J., M. B. Wakin, and S. P. Boyd
2008. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905.

Capdeville, Y., P. Cupillard, and S. Singh
2020. An introduction to the two-scale homogenization method for seismology. In *Advances in Geophysics*, volume 61, Pp. 217–306. Elsevier.

Capdeville, Y., L. Guillot, and J.-J. Marigo
2010a. 1-D non-periodic homogenization for the seismic wave equation. *Geophysical Journal International*.

Capdeville, Y., L. Guillot, and J.-J. Marigo
2010b. 2-D non-periodic homogenization to upscale elastic media for P-SV waves: 2-D non-periodic homogenization, PSV case. *Geophysical Journal International*, 182(2):903–922.

Cardiff, M. and P. K. Kitanidis
2009. Bayesian inversion for facies detection: An extensible level set framework. *Water Resources Research*, 45(10).

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell
2017. *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1).

Castellanos, J. C., R. W. Clayton, and A. Juarez
2020. Using a Time-Based Subarray Method to Extract and Invert Noise-Derived Body Waves at Long Beach, California. *Journal of Geophysical Research: Solid Earth*, 125(5).

Chada, N. K.
2018. Analysis of Hierarchical Ensemble Kalman Inversion. *arXiv:1801.00847 [math]*.

Chada, N. K., M. A. Iglesias, L. Roininen, and A. M. Stuart
2018. Parameterizations for ensemble Kalman inversion. *Inverse Problems*, 34(5):055009.

Chan, J. Y. H., B. Leistedt, T. D. Kitching, and J. D. McEwen
2017. Second-Generation Curvelets on the Sphere. *IEEE Transactions on Signal Processing*, 65(1):5–14.

Chen, V., M. M. Dunlop, O. Papaspiliopoulos, and A. M. Stuart
2019a. Dimension-Robust MCMC in Bayesian Inverse Problems. *arXiv:1803.03344 [stat]*.

Chen, Y., M. Bai, and Y. Chen
2019b. Obtaining free USArray data by multi-dimensional seismic reconstruction. *Nature Communications*, 10(1):4434.

Chiles, J.-P. and P. Delfiner
2012. *Geostatistics: Modeling Spatial Uncertainty*, volume 497, second edition. John Wiley & Sons.

Chung, E. T., T. F. Chan, and X.-C. Tai
2005. Electrical impedance tomography using level set representation and total variational regularization. *Journal of Computational Physics*, 205(1):357–372.

Claeskens, G.
2016. Statistical Model Choice. *Annual Review of Statistics and Its Application*, 3(1):233–256.

Clayton, R. W., T. Heaton, M. Chandy, A. Krause, M. Kohler, J. Bunn, R. Guy, M. Olson, M. Faulkner, M. Cheng, L. Strand, R. Chandy, D. Obenshain, A. Liu, and M. Aivazis
2012. Community Seismic Network. *Annals of Geophysics*, 54(6):738–747.

Clayton, R. W., M. Kohler, R. Guy, J. Bunn, T. Heaton, and M. Chandy
2020. CSN-LAUSD Network: A Dense Accelerometer Network in Los Angeles Schools. *Seismological Research Letters*, 91(2A):622–630.

Da Silva, C., Y. Zhang, R. Kumar, and F. J. Herrmann
2019. Applications of low-rank compressed seismic data to full-waveform inversion and extended image volumes. *GEOPHYSICS*, 84(3):R371–R383.

Dashti, M. and A. M. Stuart
2013. The Bayesian Approach To Inverse Problems. *arXiv:1302.6989 [math]*.

Datta, A.
2018. SWRT: A package for semi-analytical solutions of surface wave propagation, including mode conversion, across transversely aligned vertical discontinuities. *Geoscientific Instrumentation, Methods and Data Systems*, 7(1):101–112.

Daubechies, I.
1992. *Ten Lectures on Wavelets*, volume 61. Siam.

Davenport, M., M. Duarte, Y. Eldar, and G. Kutyniok
2011. Introduction to compressed sensing. In *Compressed Sensing: Theory and Applications*, Pp. 1–68. Cambridge, Uk: Cambridge University Press.

de Ridder, S. A. L. and B. L. Biondi
2015. Near-surface Scholte wave velocities at Ekofisk from short noise recordings by seismic noise gradiometry. *Geophysical Research Letters*, 42(17):7031–7038.

de Ridder, S. A. L. and J. R. Maddison
2018. Full Wave Field Inversion of Ambient Seismic Noise. *Geophysical Journal International*.

de Wit, R. W. L., J. Trampert, and R. D. van der Hilst
2012. Toward quantifying uncertainty in travel time tomography using the null-space shuttle: Robustness in travel time tomography. *Journal of Geophysical Research: Solid Earth*, 117(B3).

Deal, M. M. and G. Nolet
1996. Nullspace shuttles. *Geophysical Journal International*, 124(2):372–380.

Donoho, D. L.
2006. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(6):797–829.

Dunning, I., J. Huchette, and M. Lubin
2017. JuMP: A Modeling Language for Mathematical Optimization. *SIAM Review*, 59(2):295–320.

Dziewonski, A. M., T.-A. Chou, and J. H. Woodhouse
1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research: Solid Earth*, 86(B4):2825–2852.

Dziewonski, A. M., B. H. Hager, and R. J. O'Connell
1977. Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research*, 82(2):239–255.

Eikrem, K. S., G. Nævdal, and M. Jakobsen
2019. Iterated extended Kalman filter method for time-lapse seismic full-waveform inversion. *Geophysical Prospecting*, 67(2):379–394.

Ekström, G., M. Nettles, and A. M. Dziewoński
2012. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200-201:1–9.

Evensen, G.
1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143.

Evensen, G.
2003. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367.

Fäh, D., F. Kind, and D. Giardini
2001. A theoretical investigation of average H/V ratios. *Geophysical Journal International*, 145:535–549.

Faust, L. Y.
1951. Seismic velocity as a function of depth and geologic time. *GEOPHYSICS*, 16(2):192–206.

Feigl, K.
2016a. Brady's Geothermal Field DAS Earthquake Data.

Feigl, K.
2016b. Brady's Geothermal Field Nodal Seismometer Earthquake Data.

Feigl, K. L.
2017. Overview and preliminary results from the PoroTomo project at Brady Hot Springs, Nevada: Poroelastic tomography by adjoint inverse modeling of data from seismology, geodesy, and hydrology. In *42nd Workshop on Geothermal Reservoir Engineering*, Stanford University.

Feigl, K. L. and L. M. Parker
2019. PoroTomo Final Technical Report: Poroelastic Tomography by Adjoint Inverse Modeling of Data from Seismology, Geodesy, and Hydrology. Technical Report 3.1, 1499141, University of Wisconsin-Madison.

Ferguson, R. J., M. A. McDonald, and D. J. Basto
2020. Take the Eh? train: Distributed Acoustic Sensing (DAS) of commuter trains in a Canadian City. *Journal of Applied Geophysics*, 183:104201.

Fichtner, A. and A. Zunino
2019. Hamiltonian Nullspace Shuttles. *Geophysical Research Letters*, 46(2):644–651.

Fichtner, A., A. Zunino, and L. Gebraad
2019. Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, 216(2):1344–1363.

Filippitzis, F., M. D. Kohler, T. H. Heaton, R. W. Graves, R. W. Clayton, R. G. Guy, J. J. Bunn, and K. M. Chandy
2021. Ground motions in urban Los Angeles from the 2019 Ridgecrest earthquake sequence. *Earthquake Spectra*, P. 875529302110039.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman
2013. Emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306.

Foulger, G. R., G. F. Panza, I. M. Artemieva, I. D. Bastow, F. Cammarano, J. R. Evans, W. B. Hamilton, B. R. Julian, M. Lustrino, H. Thybo, and T. B. Yanovskaya
2013. Caveats on tomographic images. *Terra Nova*, 25(4):259–281.

Friedman, J., T. Hastie, and R. Tibshirani
2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1).

Friel, N. and J. Wyse
2012. Estimating the evidence–a review. *Statistica Neerlandica*, 66(3):288–308.

Gao, C. and V. Lekić
2018. Consequences of parameterization choices in surface wave inversion: Insights from transdimensional Bayesian methods. *Geophysical Journal International*, 215(2):1037–1063.

Garbuno-Inigo, A., F. Hoffmann, W. Li, and A. M. Stuart
2020. Interacting Langevin Diffusions: Gradient Structure and Ensemble Kalman Sampler. *SIAM Journal on Applied Dynamical Systems*, 19(1):412–441.

Gelman, A., W. R. Gilks, and G. O. Roberts
1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1).

Geogiga Technology Corporation
2016. DWTomo, http://geogiga.com/en/dwtomo.php.

Grana, D., T. Fjeldstad, and H. Omre
2017. Bayesian Gaussian Mixture Linear Inversion for Geophysical Inverse Problems. *Mathematical Geosciences*, 49(4):493–515.

Graves, R. W. and A. Pitarka
2010. Broadband Ground-Motion Simulation Using a Hybrid Approach. *Bulletin of the Seismological Society of America*, 100(5A):2095–2123.

Hadley, D. and H. Kanamori
1977. Seismic structure of the Transverse Ranges, California. *Geological Society of America Bulletin*, 88:1469–1478.

Halko, N., P. G. Martinsson, and J. A. Tropp
2011. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review*, 53(2):217–288.

Haney, M. M. and V. C. Tsai
2015. Nonperturbational surface-wave inversion: A Dix-type relation for surface waves. *GEOPHYSICS*, 80(6):EN167–EN177.

Haney, M. M. and V. C. Tsai
2017. Perturbational and nonperturbational inversion of Rayleigh-wave velocities. *GEOPHYSICS*, 82(3):F15–F28.

Haney, M. M. and V. C. Tsai
2020. Perturbational and nonperturbational inversion of Love-wave velocities. *GEOPHYSICS*, 85(1):F19–F26.

Hanke, M.
1997. A regularizing Levenberg - Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Problems*, 13(1):79–95.

Hansen, T. M., A. G. Journel, A. Tarantola, and K. Mosegaard
2006. Linear inverse Gaussian theory and geostatistics. *Geophysics*, 71(6):R101–R111.

Hennenfent, G., L. Fenelon, and F. J. Herrmann
2010. Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach. *Geophysics*, 75(6):WB203–WB210.

Hennenfent, G. and F. J. Herrmann
2006. Seismic denoising with nonuniformly sampled curvelets. *Computing in Science & Engineering*, 8(3):16–25.

Herlihy, M. and N. Shavit
2011. *The Art of Multiprocessor Programming*. Morgan Kaufmann.

Herrmann, F. J. and G. Hennenfent
2008. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173(1):233–248.

Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam
2008. Curvelet-based seismic data processing: A multiscale and nonlinear approach. *Geophysics*, 73(1):A1–A5.

Hunter, J. D.
2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95.

Hutton, K., J. Woessner, and E. Hauksson
2010. Earthquake Monitoring in Southern California for Seventy-Seven Years (1932-2008). *Bulletin of the Seismological Society of America*, 100(2):423–446.

Ide, S., E. Araki, and H. Matsumoto
2021. Very broadband strain-rate measurements along a submarine fiber-optic cable off Cape Muroto, Nankai subduction zone, Japan. *Earth, Planets and Space*, 73(1):63.

Iglesias, M. A.
2014. Iterative regularization for ensemble data assimilation in reservoir models. *arXiv:1401.5375 [math, stat]*.

Iglesias, M. A.
2016. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Problems*, 32(2):025002.

Iglesias, M. A. and C. Dawson
2013. The regularizing Levenberg–Marquardt scheme for history matching of petroleum reservoirs. *Computational Geosciences*, 17(6):1033–1053.

Iglesias, M. A., K. J. H. Law, and A. M. Stuart
2013. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4):045001.

Ingersoll, R. V. and P. E. Rumelhart
1999. Three-stage evolution of the Los Angeles basin, southern California. *Geology*, 27(7):593–596.

Isakov, V., S. Leung, and J. Qian
2011. A Fast Local Level Set Method for Inverse Gravimetry. *Communications in Computational Physics*, 10(04):1044–1070.

Jaynes, E. T.
1999. Straight line fitting - A Bayesian solution; available at https://bayes.wustl.edu/etj/articles/leapz.pdf. *Unpublished Manuscript*.

Jia, Z. and R. W. Clayton
2021. Determination of Near Surface Shear-Wave Velocities in the Central Los Angeles Basin With Dense Arrays. *Journal of Geophysical Research: Solid Earth*, 126(5).

Karimpouli, S. and P. Tahmasebi
2020. Physics informed machine learning: Seismic wave equation. *Geoscience Frontiers*, 11(6):1993–2001.

Karrenbach, M., Z. Shen, Z. Li, S. Cole, E. Williams, A. Klesh, X. Wang, Z. Zhan, L. LaFlame, V. Yartsev, and V. Bogdanov
2020. Rapid deployment of distributed acoustic sensing systems to track earthquake activity. In *SEG Technical Program Expanded Abstracts 2020*, SEG Technical Program Expanded Abstracts, Pp. 490–494. Society of Exploration Geophysicists.

Kennett, B. and A. Fichtner
2020. Introduction. In *Exploiting Seismic Waveforms: Correlation, Heterogeneity and Inversion*, Pp. 1–10. Cambridge: Cambridge University Press.

Kennett, B. L. N., E. R. Engdahl, and R. Buland
1995. Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, 122(1):108–124.

King, G., W. Zurn, R. Evans, and D. Emter
1976. Site Correction for Long Period Seismometers, Tiltmeters and Strainmeters. *Geophysical Journal International*, 44(2):405–411.

Kiser, E. and M. Ishii
2017. Back-Projection Imaging of Earthquakes. *Annual Review of Earth and Planetary Sciences*, 45(1):271–299.

Ko, J. Y.-T., D. V. Helmberger, H. Wang, and Z. Zhan
2017. Lower Mantle Substructure Embedded in the Farallon Plate: The Hess Conjugate. *Geophysical Research Letters*, 44(20):10,216–10,225.

Kohler, M. D., F. Filippitzis, T. Heaton, R. W. Clayton, R. Guy, J. Bunn, and K. M. Chandy
2020. 2019 Ridgecrest Earthquake Reveals Areas of Los Angeles That Amplify Shaking of High-Rises. *Seismological Research Letters*, 91(6):3370–3380.

Komatitsch, D., Z. Xie, E. Bozdağ, E. Sales de Andrade, D. Peter, Q. Liu, and J. Tromp
2016. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion. *Geophysical Journal International*, 206(3):1467–1478.

Kovachki, N. B. and A. M. Stuart
2018. Ensemble Kalman Inversion: A Derivative-Free Technique For Machine Learning Tasks. *arXiv:1808.03620 [cs, math, stat]*.

Krischer, L., J. Smith, W. Lei, M. Lefebvre, Y. Ruan, E. S. de Andrade, N. Podhorszki, E. Bozdağ, and J. Tromp
2016. An Adaptable Seismic Data Format. *Geophysical Journal International*, 207(2):1003–1011.

Lai, V. H., R. Graves, Z. Zhan, C. Yu, and D. Helmberger
2019. Analyzing Shallow Basin Effects in Los Angeles Basin using 3D Simulations and Dense Array Analysis. In *SCEC Annual Meeting*, volume Contribution 9804.

Lai, V. H., R. W. Graves, C. Yu, Z. Zhan, and D. V. Helmberger
2020. Shallow Basin Structure and Attenuation Are Key to Predicting Long Shaking Duration in Los Angeles Basin. *Journal of Geophysical Research: Solid Earth*, 125(10).

Langston, C. A.
2007a. Spatial gradient analysis for linear seismic arrays. *Bulletin of the Seismological Society of America*, 97(1):265–280.

Langston, C. A.
2007b. Wave gradiometry in two dimensions. *Bulletin of the Seismological Society of America*, 97(2):401–416.

Langston, C. A. and S. M. Mousavi
2019. Separating Signal from Noise and from Other Signal Using Nonlinear Thresholding and Scale-Time Windowing of Continuous Wavelet Transforms. *Bulletin of the Seismological Society of America*, 109(5):1691–1700.

Laub, A. J.
2005. *Matrix Analysis for Scientists and Engineers*, volume 91. Siam.

Lee, E.-J., P. Chen, T. H. Jordan, P. B. Maechling, M. A. M. Denolle, and G. C. Beroza
2014. Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research: Solid Earth*, 119(8):6421–6451.

Leinonen, M., R. J. Hewett, L. Demanet, X. Zhang, and L. Ying
2013. High-dimensional wave atoms and compression of seismic data sets. In *SEG Technical Program Expanded Abstracts 2013*, Pp. 3591–3596. Society of Exploration Geophysicists.

Li, G., H. Chen, F. Niu, Z. Guo, Y. Yang, and J. Xie
2016. Measurement of Rayleigh wave ellipticity and its application to the joint inversion of high-resolution S wave velocity structure beneath northeast China. *Journal of Geophysical Research: Solid Earth*, 121(2):864–880.

Li, H., J. Schwab, S. Antholzer, and M. Haltmeier
2020a. NETT: Solving inverse problems with deep neural networks. *Inverse Problems*, 36(6):065005.

Li, W., S. Leung, and J. Qian
2014. A level-set adjoint-state method for crosswell transmission-reflection traveltime tomography. *Geophysical Journal International*, 199(1):348–367.

Li, W., W. Lu, J. Qian, and Y. Li
2017. A multiple level-set method for 3D inversion of magnetic data. *Geophysics*, 82(5):J61–J81.

Li, W. and J. Qian
2016. Joint inversion of gravity and traveltime data using a level-set-based structural parameterization. *Geophysics*, 81(6):G107–G119.

Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar
2020b. Fourier Neural Operator for Parametric Partial Differential Equations. *arXiv:2010.08895 [cs, math]*.

Li, Z. and Z. Zhan
2018. Pushing the limit of earthquake detection with distributed acoustic sensing and template matching: A case study at the Brady geothermal field. *Geophysical Journal International*, 215(3):1583–1593.

Lin, F.-C., D. Li, R. W. Clayton, and D. Hollis
2013. High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic array. *GEOPHYSICS*, 78(4):Q45–Q56.

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

Lin, F.-C. and M. H. Ritzwoller
2011. Helmholtz surface wave tomography for isotropic and azimuthally

anisotropic structure: Helmholtz surface wave tomography. *Geophysical Journal International*, 186(3):1104–1120.

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

Lin, F. C., B. Schmandt, and V. C. Tsai
2012. Joint inversion of Rayleigh wave phase velocity and ellipticity using USArray: Constraining velocity and density structure in the upper crust. *Geophysical Research Letters*, 39(12):n/a–n/a.

Lin, F. C., V. C. Tsai, and B. Schmandt
2014. 3-D crustal structure of the western United States: Application of Rayleigh-wave ellipticity extracted from noise cross-correlations. *Geophysical Journal International*, 198(2):656–670.

Lindgren, F., H. Rue, and J. Lindström
2011. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach: Link between Gaussian Fields and Gaussian Markov Random Fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Lindsey, N. J., T. C. Dawe, and J. B. Ajo-Franklin
2019. Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing. *Science*, 366(6469):1103–1107.

Lindsey, N. J., E. R. Martin, D. S. Dreger, B. Freifeld, S. Cole, S. R. James, B. L. Biondi, and J. B. Ajo-Franklin
2017. Fiber-Optic Network Observations of Earthquake Wavefields. *Geophysical Research Letters*, 44(23):11,792–11,799.

Lindsey, N. J., H. Rademacher, and J. B. Ajo-Franklin
2020. On the Broadband Instrument Response of Fiber-Optic DAS Arrays. *Journal of Geophysical Research: Solid Earth*, 125(2):e2019JB018145.

Liu, D. C. and J. Nocedal
1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528.

Lu, W. and J. Qian
2015. A local level-set method for 3D inversion of gravity-gradient data. *Geophysics*, 80(1):G35–G51.

Ludwig, L. G., S. O. Akciz, G. R. Noriega, O. Zielke, and J. R. Arrowsmith
2010. Climate-Modulated Channel Incision and Rupture History of the San Andreas Fault in the Carrizo Plain. *Science*, 327(5969):1117–1119.

Lysmer, J.
1970. Lumped Mass Method for Rayleigh Waves. *Bulletin of the Seismological Society of America*, 60(1):89–&.

Ma, Y. R. and R. W. Clayton
2016. Structure of the Los Angeles Basin from ambient noise and receiver functions. *Geophysical Journal International*, 206(3):1645–1651.

Magistrale, H., Day, Steven, Clayton, Robert W., and Graves, Robert
2000. The SCEC Southern California Reference Three-Dimensional Seismic Velocity Model Version 2. *Bulletin of the Seismological Society of America*, 90(6B):S65–S76.

Magistrale, H., K. McLaughlin, and S. Day
1996. A geology-based 3D velocity model of the Los Angeles basin sediments. *Bulletin of the Seismological Society of America*, 86(4):1161–1166.

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

Marsal, O., M. Venet, J.-L. Counil, F. Ferri, A.-M. Harri, T. Spohn, and J. Block
2002. The NetLander geophysical network on the surface of Mars: General mission description and technical design status. *Acta Astronautica*, 51(1-9):379–386.

Martin, E. R., N. Lindsey, J. Ajo-Franklin, and B. Biondi
2018. Introduction to Interferometry of Fiber Optic Strain Measurements. Preprint, EarthArXiv.

Massias, M., A. Gramfort, and J. Salmon
2018. Celer: A fast solver for the Lasso with dual extrapolation. *arXiv preprint arXiv:1802.07481*.

Matsumoto, H., E. Araki, T. Kimura, G. Fujie, K. Shiraishi, T. Tonegawa, K. Obana, R. Arai, Y. Kaiho, Y. Nakamura, T. Yokobiki, S. Kodaira, N. Takahashi, R. Ellwood, V. Yartsev, and M. Karrenbach
2021. Detection of hydroacoustic signals on a fiber-optic submarine cable. *Scientific Reports*, 11(1):2797.

Matsushima, S., T. Hirokawa, F. De Martin, H. Kawase, and F. J. Sanchez-Sesma
2014. The Effect of Lateral Heterogeneity on Horizontal-to-Vertical Spectral Ratio of Microtremors Inferred from Observation and Synthetics. *Bulletin of the Seismological Society of America*, 104(1):381–393.

Met Office
2010. *Cartopy: A Cartographic Python Library with a Matplotlib Interface*. Exeter, Devon.

Moseley, B., A. Markham, and T. Nissen-Meyer
2020. Solving the wave equation with physics-informed deep learning. *arXiv:2006.11894 [physics]*.

Mousavi, S. M. and C. A. Langston
2016. Hybrid Seismic Denoising Using Higher-Order Statistics and Improved Wavelet Block Thresholding. *Bulletin of the Seismological Society of America*, 106(4):1380–1393.

Muir, J. and V. C. Tsai
2020a. Did Oldham Discover the Core After All? Handling Imprecise Historical Data with Hierarchical Bayesian Model Selection Methods. *Seismological Research Letters*, 91(3):1377–1383.

Muir, J. B. and V. C. Tsai
2017. Rayleigh-Wave H/V via Noise Cross Correlation in Southern California. *Bulletin of the Seismological Society of America*, 107(5):2021–2027.

Muir, J. B. and V. C. Tsai
2020b. Geometric and level set tomography using ensemble Kalman inversion. *Geophysical Journal International*, 220(2):967–980.

Muir, J. B. and Z. Zhan
2021. Seismic wavefield reconstruction using a pre-conditioned wavelet–curvelet compressive sensing approach. *Geophysical Journal International*, 227(1):303–315.

Nakamura, Y.
2000. Clear Identification of Fundamental Idea of Nakamura's Technique and its Applications. In *12WCEE 2000*.

Nayak, A., J. Ajo-Franklin, and The Imperial Valley Dark Fiber Team
2021. Measurement of Surface-Wave Phase-Velocity Dispersion on Mixed Inertial Seismometer – Distributed Acoustic Sensing Seismic Noise Cross-Correlations. *Bulletin of the Seismological Society of America*.

Neal, C., R. Weber, S. Kedar, C. Beghein, P. Chi, N. Currier, S. Dell'Agnello, R. Garcia, I. Garrick-Bethell, R. Grimm, M. Grott, H. Haviland, T. Kawamura, P. Lognonné, S. Nagihara, Y. Nakamura, C. Nunn, L. Ostrach, N. Petro, and M. Panning
2019. The Lunar Geophysical Network Mission. In *Lunar Planety Science Conference*.

Neal, R. M.
2011. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11).

Ni, S., E. Tan, M. Gurnis, and D. Helmberger
2002. Sharp Sides to the African Superplume. *Science*, 296(5574):1850.

Oldham, R. D.
1906. The constitution of the interior of the Earth, as revealed by earthquakes. *Quarterly Journal of the Geological Society*, 62(1-4):456–475.

Osher, S., M. Burger, D. Goldfarb, J. Xu, and W. Yin
2005. An Iterative Regularization Method for Total Variation-Based Image Restoration. *Multiscale Modeling & Simulation*, 4(2):460–489.

Osher, S. and J. A. Sethian
1988. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49.

Parker, L. M., C. H. Thurber, X. Zeng, P. Li, N. E. Lord, D. Fratta, H. F. Wang, M. C. Robertson, A. M. Thomas, M. S. Karplus, A. Nayak, and K. L. Feigl
2018. Active-Source Seismic Tomography at the Brady Geothermal Field, Nevada, with Dense Nodal and Fiber-Optic Seismic Arrays. *Seismological Research Letters*, 89(5):1629–1640.

Parker, R. L.
1994. *Geophysical Inverse Theory*. Princeton University Press.

Pitarka, A., R. W. Graves, and A. J. Rodgers
2019. Strong Ground Motion Simulation for the M7.1, 2019 Ridgecrest California Earthquake. *AGU Fall Meeting Abstracts*, 31.

Plesch, A., J. H. Shaw, C. Benson, W. A. Bryant, S. Carena, M. Cooke, J. Dolan, G. Fuis, E. Gath, L. Grant, E. Hauksson, T. Jordan, M. Kamerling, M. Legg, S. Lindvall, H. Magistrale, C. Nicholson, N. Niemi, M. Oskin, S. Perry, G. Planansky, T. Rockwell, P. Shearer, C. Sorlien, M. P. Suss, J. Suppe, J. Treiman, and R. Yeats
2007. Community Fault Model (CFM) for Southern California. *Bulletin of the Seismological Society of America*, 97(6):1793–1802.

Qiu, H., F.-C. Lin, and Y. Ben-Zion
2019. Eikonal Tomography of the Southern California Plate Boundary Region. *Journal of Geophysical Research: Solid Earth*, 124(9):9755–9779.

Raissi, M., P. Perdikaris, and G. Karniadakis
2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707.

Rasmussen, C. E. and C. K. I. Williams
2006. *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning. Cambridge, Mass: MIT Press.

Ravi, S. N., V. K. Ithapu, S. C. Johnson, and V. Singh
2017. Experimental Design on a Budget for Sparse Linear Models and Applications. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, P. 18, New York.

Rawlinson, N., A. Fichtner, M. Sambridge, and M. K. Young
2014. Seismic Tomography and the Assessment of Uncertainty. In *Advances in Geophysics*, volume 55, Pp. 1–76. Elsevier.

Rawlinson, N., S. Pozgay, and S. Fishwick
2010. Seismic tomography: A window into deep Earth. *Physics of the Earth and Planetary Interiors*, 178(3-4):101–135.

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

Ray, A. and D. Myer
2019. Bayesian geophysical inversion with trans-dimensional Gaussian Process machine learning. *Geophysical Journal International*, 217(3):1706–1725.

Ringler, A. T., R. E. Anthony, M. S. Karplus, A. A. Holland, and D. C. Wilson
2018. Laboratory Tests of Three Z-Land Fairfield Nodal 5-Hz, Three-Component Sensors. *Seismological Research Letters*, 89(5):1601–1608.

Ringler, A. T., D. C. Wilson, W. Zürn, and R. E. Anthony
2019. Rayleigh wave ellipticity measurement uncertainty across the IRIS/USGS and New China Digital Seismograph Networks. *Geophysical Journal International*, 217(1):219–237.

Roy, C. and B. A. Romanowicz
2017. On the Implications of A Priori Constraints in Transdimensional Bayesian Inversion for Continental Lithospheric Layering. *Journal of Geophysical Research: Solid Earth*, 122(12):10,118–10,131.

Sánchez-Sesma, F. J., M. Rodríguez, U. Iturrarán-Viveros, F. Luzón, M. Campillo, L. Margerin, A. García-Jerez, M. Suarez, M. A. Santoyo, and A. Rodríguez-Castellanos
2011. A theory for microtremor H/V spectral ratio: Application for a layered medium. *Geophysical Journal International*, 186(1):221–225.

Scherzer, O.
1993. The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems. *Computing*, 51(1):45–60.

SciPy 1.0 Contributors, P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J.

Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt
2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.

Shao, J.
1993. Linear model selection by cross-validation. *Journal of the American statistical Association*, 88(422):486–494.

Shapiro, N. M. and M. Campillo
2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31(7):n/a–n/a.

Shapiro, N. M., M. Campillo, L. Stehly, and M. H. Ritzwoller
2005. High-resolution surface-wave tomography from ambient seismic noise. *Science*, 307(5715):1615–8.

Shaw, J. H., A. Plesch, C. Tape, M. P. Suess, T. H. Jordan, G. Ely, E. Hauksson, J. Tromp, T. Tanimoto, R. Graves, K. Olsen, C. Nicholson, P. J. Maechling, C. Rivero, P. Lovely, C. M. Brankman, and J. Munster
2015. Unified Structural Representation of the southern California crust and upper mantle. *Earth and Planetary Science Letters*, 415:1–15.

Sheldrake, K. P.
2002. Regional Wavefield reconstruction for teleseismic P-waves and Surface waves. *Geophysical Research Letters*, 29(11):1544.

Singh, S., Y. Capdeville, and H. Igel
2019. Correcting wavefield gradients for the effects of local small-scale heterogeneities. *Geophysical Journal International*, P. ggz479.

Sitzmann, V., J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein
2020. Implicit Neural Representations with Periodic Activation Functions. *arXiv:2006.09661 [cs, eess]*.

Sladen, A., D. Rivet, J. P. Ampuero, L. De Barros, Y. Hello, G. Calbris, and P. Lamare
2019. Distributed sensing of earthquakes and ocean-solid Earth interactions on seafloor telecom cables. *Nature Communications*, 10(1):5777.

Snieder, R.
2004. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Physical Review E*, 69(4):046610.

Song, C., T. Alkhalifah, and U. bin Waheed
2020. Solving the acoustic VTI wave equation using physics-informed neural networks. *arXiv:2008.01865 [physics]*.

Song, T.-R. A., D. V. Helmberger, M. R. Brudzinski, R. W. Clayton, P. Davis, X. Perez-Campos, and S. K. Singh
2009. Subducting Slab Ultra-Slow Velocity Layer Coincident with Silent Earthquakes in Southern Mexico. *Science*, 324(5926):502–506.

Spica, Z. J., K. Nishida, T. Akuhara, F. Pétrélis, M. Shinohara, and T. Yamada
2020. Marine Sediment Characterized by Ocean-Bottom Fiber-Optic Seismology. *Geophysical Research Letters*, 47(16).

Starck, J.-L., F. Murtagh, and J. M. Fadili
2010. *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge university press.

Sun, D. Y., D. Helmberger, M. S. Miller, and J. M. Jackson
2016. Major disruption of D" beneath Alaska. *Journal of Geophysical Research-Solid Earth*, 121(5):3534–3556.

Süss, M. P. and J. H. Shaw
2003. Pwave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California. *Journal of Geophysical Research: Solid Earth*, 108(B3).

Taborda, R., S. Azizzadeh-Roodpish, N. Khoshnevis, and K. Cheng
2016. Evaluation of the southern California seismic velocity models through simulation of recorded events. *Geophysical Journal International*, 205(3):1342–1364.

Tanimoto, T. and L. Rivera
2005. Prograde Rayleigh wave particle motion. *Geophysical Journal International*, 162(2):399–405.

Tape, C., Q. Liu, A. Maggi, and J. Tromp
2009. Adjoint Tomography of the Southern California Crust. *Science*, 325(5943):988–992.

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

Thurin, J., R. Brossier, and L. Métivier
2017. An ensemble-transform Kalman filter: Full-waveform inversion scheme for uncertainty estimation. In *SEG Technical Program Expanded Abstracts 2017*, Pp. 1307–1313, Houston, Texas. Society of Exploration Geophysicists.

Torrence, C. and G. P. Compo
1998. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

Treister, E. and E. Haber
2016. A fast marching algorithm for the factored eikonal equation. *Journal of Computational Physics*, 324:210–225.

Tsai, V. C.
2011. Understanding the amplitudes of noise correlation measurements. *Journal of Geophysical Research-Solid Earth*, 116(B9).

Tsai, V. C. and S. Atiganyanun
2014. Green's Functions for Surface Waves in a Generic Velocity Structure. *Bulletin of the Seismological Society of America*, 104(5):2573–2578.

Tso, C.-H. M., M. Iglesias, P. Wilkinson, O. Kuras, J. Chambers, and A. Binley
2021. Efficient multiscale imaging of subsurface resistivity with uncertainty quantification using ensemble Kalman inversion. *Geophysical Journal International*, 225(2):887–905.

USGS
2020. U.S. Geological Survey and California Geological Survey, Quaternary fault and fold database for the United States, https://www.usgs.gov/natural-hazards/earthquake-hazards/faults, Accessed April 12, 2021.

van den Ende, M. P. A. and J.-P. Ampuero
2021. Evaluating seismic beamforming capabilities of distributed acoustic sensing arrays. *Solid Earth*, 12(4):915–934.

Vehtari, A., A. Gelman, and J. Gabry
2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

Vehtari, A. and J. Ojanen
2012. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6(0):142–228.

Villasenor, J., R. Ergas, and P. Donoho
1996. Seismic data compression using high-dimensional wavelet transforms. In *Proceedings of Data Compression Conference - DCC '96*, Pp. 396–405, Snowbird, UT, USA. IEEE Comput. Soc. Press.

Wächter, A. and L. T. Biegler
2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57.

Wang, H. F., X. Zeng, D. E. Miller, D. Fratta, K. L. Feigl, C. H. Thurber, and R. J. Mellors
2018. Ground motion response to an ML 4.3 earthquake using co-located distributed acoustic sensing and seismometer arrays. *Geophysical Journal International*, 213(3):2020–2036.

Wang, X., E. F. Williams, M. Karrenbach, M. G. Herráez, H. F. Martins, and Z. Zhan
2020. Rose Parade Seismology: Signatures of Floats and Bands on Optical Fiber. *Seismological Research Letters*, 91(4):2395–2398.

Weinberg, M. D.
2012. Computing the Bayes Factor from a Markov Chain Monte Carlo Simulation of the Posterior Distribution. *Bayesian Analysis*, 7(3):737–770.

Wiechert, E.
1897. Ueber die massenverteilung im inneren der erde. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1897(3):221–243.

Wiesmeyr, C., M. Litzenberger, M. Waser, A. Papp, H. Garn, G. Neunteufel, and H. Döller
2020. Real-Time Train Tracking from Distributed Acoustic Sensing Data. *Applied Sciences*, 10(2):448.

Williams, E. F., M. R. Fernández-Ruiz, R. Magalhaes, R. Vanthillo, Z. Zhan, M. González-Herráez, and H. F. Martins
2019. Distributed sensing of microseisms and teleseisms with submarine dark fibers. *Nature Communications*, 10(1):5778.

Wilson, C. and A. Guitton
2007. Teleseismic wavefield interpolation and signal extraction using high-resolution linear radon transforms. *Geophysical Journal International*, 168(1):171–181.

Wit, E., E. van den Heuvel, and J.-W. Romeijn
2012. 'All models are wrong...': An introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236.

Wood, L. C.
1974. Seismic Data Compression Methods. *Geophysics*, 39(4):499–525.

Wright, T. L.
1991. Structural Geology and Tectonic Evolution of the Los Angeles Basin, California. In *Active Margin Basins*, AAPG MEMOIR, Pp. 35–134. American Association of Petroleum Geologists: AAPG Special Volumes.

Yerkes, R. F., R. H. Campbell, R. M. Alvarez, and K. R. Bovard
2005. Preliminary geologic map of the Los Angeles $30' \times 60'$ Quadrangle, southern California. *US Geological Survey Open File Report*, 1019:1559–1573.

Yerkes, R. F., T. H. McCulloh, J. E. Schoellhamer, and J. G. Vedder
1965. Geology of the Los Angeles Basin, California; an introduction. Technical Report 420-A, United States Geological Survey.

Yu, C., Z. Zhan, N. J. Lindsey, J. B. Ajo-Franklin, and M. Robertson
2019. The Potential of DAS in Teleseismic Studies: Insights From the Goldstone Experiment. *Geophysical Research Letters*, 46(3):1320–1328.

Zelt, C. A. and R. B. Smith
1992. Seismic traveltime inversion for 2-D crustal velocity structure. *Geophysical Journal International*, 108(1):16–34.

Zhan, Z.
2020. Distributed Acoustic Sensing Turns Fiber-Optic Cables into Sensitive Seismic Antennas. *Seismological Research Letters*, 91(1):1–15.

Zhan, Z., Q. Li, and J. Huang
2018. Application of wavefield compressive sensing in surface wave tomography. *Geophysical Journal International*, 213(3):1731–1743.

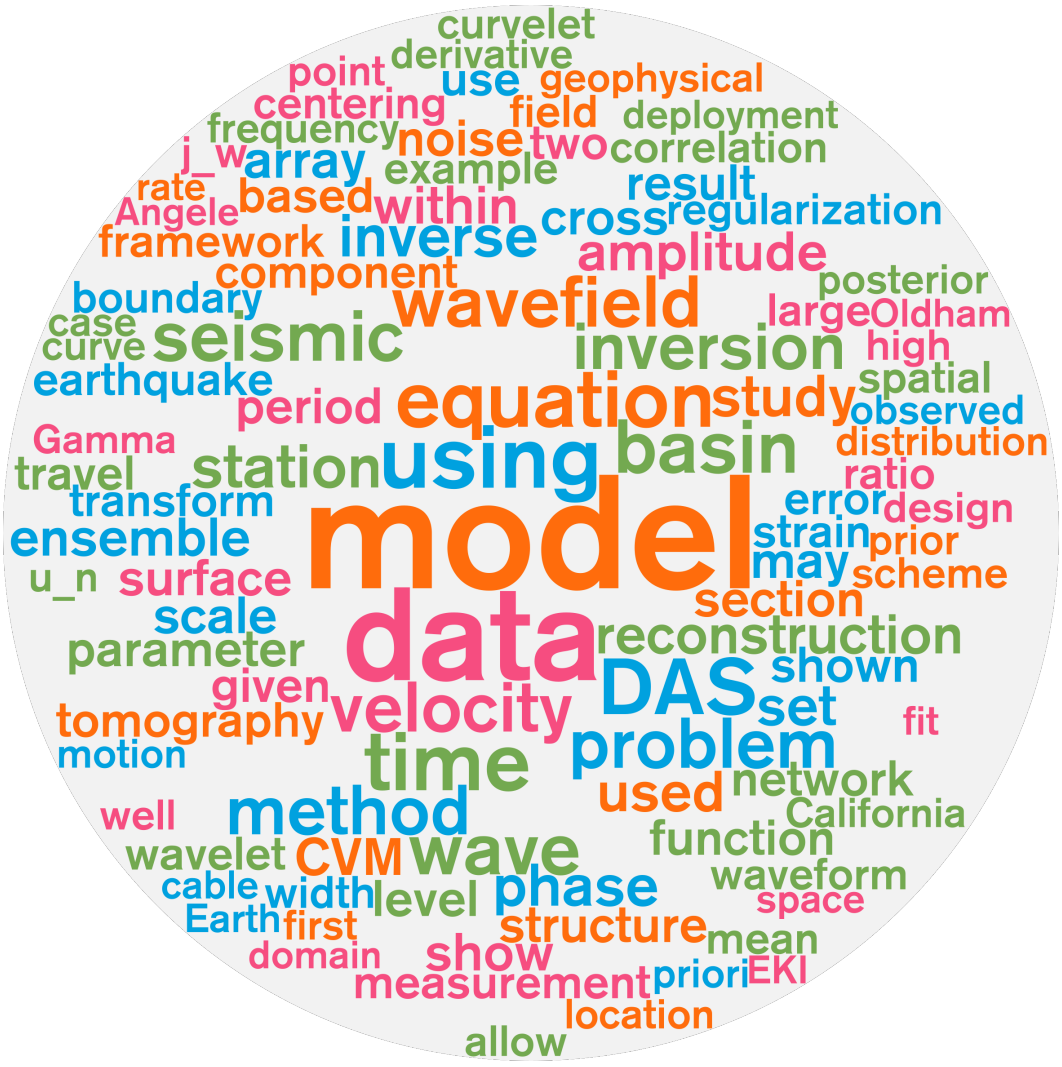
Zhan, Z. and W. Wu
2019. Fiber seismic networks on the Moon. In *AGU Fall Meeting Abstracts*, Pp. P31C–3441.

Zhang, J. and C. A. Langston
2019. Separating the scattered wavefield from teleseismic P using curvelets on the long beach array dataset. *Geophysical Journal International*, P. ggz487.

Zheglova, P., C. G. Farquharson, and C. A. Hurich
2013. 2-D reconstruction of boundaries with level set inversion of traveltimes. *Geophysical Journal International*, 192(2):688–698.

Zheglova, P., P. G. Lelièvre, and C. G. Farquharson
2018. Multiple level-set joint inversion of traveltime and gravity data with application to ore delineation: A synthetic study. *Geophysics*, 83(1):R13–R30.

Zielke, O., J. R. Arrowsmith, L. G. Ludwig, and S. O. Akciz
2010. Slip in the 1857 and Earlier Large Earthquakes Along the Carrizo Plain, San Andreas Fault. *Science*, 327(5969):1119–1122.


curvelet
derivative
point
use
geophysical
centering
field
deployment
frequency
noise
two
correlation
j_w
array
example
result
rate
based
within
cross
regularization
Angele
framework
inverse
amplitude
component
posterior
boundary
wavefield
large
Oldham
case
seismic
inversion
high
curve
earthquake
spatial
period
equation
study
observed
Gamma
basin
distribution
travel
station
using
ratio
transform
error
design
model
ensemble
strain
prior
may
u_n
surface
scheme
section
scale
data
reconstruction
parameter
shown
given
DAS
set
velocity
tomography
fit
problem
motion
time
network
used
well
method
California
function
wavelet
CVM
wave
waveform
cable
width
level
phase
space
Earth
first
structure
mean
domain
show
measurement
priori
EKI
location
allow