2020-08-20

Understanding Canadian Cordillera crustal and upper-mantle structure with joint Bayesian inversion of receiver functions and surface-wave dispersion curves

Smale, Jacquelyn

UNIVERSITY OF CALGARY

Understanding Canadian Cordillera crustal and upper-mantle structure with joint Bayesian inversion of receiver functions and surface-wave dispersion curves

by

Jacquelyn Smale

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA

AUGUST, 2020

Abstract

The accretion of exotic terranes onto the stable Northern American craton during the formation of the Northern Canadian Cordillera (NCC) generated complex structure on a crustal and lithospheric scale. Without making significant model-parametrization assumptions, inverting for structure is challenging. The Bayesian approach quantifies uncertainty due to non-uniqueness of the solution and can avoid subjective choice in model parameterization. Jointly inverting receiver-functions and surface-wave dispersion curves constrains shear-wave velocity structure. With this inversion method, I study the spatial variability of major lithospheric discontinuities, specifically of the Mohorovičić discontinuity (Moho) and lithosphere-asthenosphere boundary (LAB), along a transect from Whitehorse to Toad River. Results show a flat Moho across the NCC at $\sim$ 34-km depth and a heterogeneous LAB with depths between 70- and 100-km depth. Beneath the LAB, I observe a significant low-velocity zone with shear-wave velocities as low as $\sim$ 3.0 km/s. Only the presence of partial melting of 1.5–2.5% can explain these V_s reductions.

Acknowledgements

I acknowledge support from NSERC Discovery Grants to Jan Dettmer and Hersh Gilbert. I would like to thank the Faculty of Science for awarding me with the James Robert Cuthbertson Graduate Bursary in Geology. I would like to thank Pascal Audet and Andrew Schaeffer for supplying regional surface-wave data and Joseph Byrnes for sharing harmonic-decomposition scripts. I acknowledge IRIS Data Management Center and Obspy for making data available and accessible. I acknowledge the Canadian Network (CN), USArray Transportable Array (TA), Yukon Observatory (YO), and Yukon-Northwest Seismic Network (NY) for data used in this thesis. I would also like to thank Geneviève Savard for her help editing and my lab-mates for their support.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols and Abbreviations

Abbreviation	Definition
BAZ	Back-azimuth
CCB	Cordillera-craton boundary
CDF	Cordilleran deformation front
eSTF	Effective source-time function
FMQ	Fayalite-magnetite-quartz
ITD	Iterative time-domain
LAB	Lithosphere asthenosphere boundary
LVZ	Low velocity zone
MCMC	Markov chain Monte Carlo
MAP	Maximum a priori
MH	Metropolis-Hastings
MHG	Metropolis-Hastings-Green
NCC	Northern Canadian Cordillera
NCVP	Northern Cordillera Volcanic Province
PPD	Posterior Probability Density
RF	Receiver function
SNA	Shielded North America
SWD	Surface-wave Dispersion
SCC	Southern Canadian Cordillera
TNA	Teconically active North America
TF	Tintina Fault

Symbol	Definition
a	Gaussian filter pulse width
c	Water-level constant
$\mathbf{C}_\mathrm{d}$	Covariance matrix for data set $\mathbf{d}$
$\mathbf{C}_\mathrm{e}$	Empirically-calculated covariance matrix
$\mathbf{C}^\mathbf{m}$	Model covariance matrix
$\mathbf{d}$	Observed data
$\mathbf{d}(\mathbf{m})$	Predicted data
f	Function applied in marginalization
$G(\omega)$	Gaussian Filter
$H(\omega)$	Fourier transform of RF
I_{in}	Impulse-response function
$\lvert\mathrm{J}\rvert$	Determinant of Jacobian matrix of transformation $\mathbf{m} \rightarrow \mathbf{m}'$
k	Number of interfaces
k_b	Boltzmann's constant
$L(\mathbf{m})$	Likelihood function
$\mathbf{m}$	Model parameter vector
$\hat{\mathbf{m}}$	MAP model

$\bar{\mathbf{m}}$	Mean model
$\mathbf{n}$	Standardized $\mathbf{r}$ vector in non-toeplitz matrix
P	Primary
p	Slowness
$\mathbf{p}\perp$	Horizontal slowness
$P(\mathbf{d})$	Bayeisan evidence
$P(\mathbf{d}\|\mathbf{m})$	Data uncertainty distribution
$P(\mathbf{m})$	Prior distribtuion
$P(\mathbf{m}\|\mathbf{d})$	Posterior probability density
$\tilde{Q}$	Number of models in chosen PPD ensemble
$Q(\mathbf{m})$	Proposal probability distribution
$\hat{Q}(\omega)$	Quality factor
R	Radial component of seismogram
$\mathbf{r}$	Residual errors
$\hat{R}(\omega)$	Radial Frequency Spectra
S	Secondary
SH	Horizontally-polarized S-wave
$S(t)$	eSTF
SV	Vertically-polarized S-wave
$\tilde{S}$	Data sets
T	Transverse component of seismogram
u_i	Observed displacement
$\mathbf{v}$	Vector of velocities
V_p	P-wave velocity
$V_\mathrm{p}/V_\mathrm{s}$	Ratio of V_p and V_s
V_s	S-wave velocity
$\mathbf{w}$	Vector of noise parameters
Z	Vertical component of seismogram
$\mathbf{z}$	Vector of depth of interfaces
$\hat{Z}(\omega)$	Vertical frequency spectra
$\hat{Z}^*(\omega)$	Complex conjugate of $\hat{Z}(\omega)$
α	Acceptance probability
$\hat{\alpha}$	Attenuation frequency dependence factor
β	Tempering parameter
δ	Dirac-delta function
ζ	Scaling factor for C_e
θ	BAZ from North
κ	$V_\mathrm{p}/V_\mathrm{s}$interfaces
ξ	Gaussian normalization factor
ρ	Density
σ^{ML}	Maximum likelihood estimate for σ
σ_RF	Standard deviation of RF noise
σ_SWD	Standard deviation of SWD noise
τ	Temperature of parallel tempering system

$\phi(\omega)$	Water-level factor
φ	Partial melt fraction (%)
ω	Angular frequency
$*$	Convolution

1. Introduction

The northern Canadian Cordillera (NCC) has a long, diverse history along the former margin of Laurentia, including multiple episodes of terrane accretion and rifting. It is a former back-arc that was influenced by the asthenosphere upwelling following the opening of a slab-window at $\sim$42 Ma (Hoffman, 1988; Monger and Price, 2002; Shi et al., 1998; Madsen et al., 2006; Thorkelson et al., 2011). The ancient convergent system thrust exotic terranes upon the Laurentian craton, a transition that is observed today as an abrupt transition between Phanerozoic and Precambrian provinces (Hoffman, 1988; Monger and Price, 2002). This sharp Cordillera-craton boundary (CCB) exists on a lithospheric scale, with hot lithosphere beneath the Cordillera abruptly transitioning to colder temperatures beneath the Proterozoic and Archean cratons (Frederiksen et al., 1998; Lewis et al., 2003; Mercier et al., 2009).

Multiple competing theories based on seismic observations and petrologic data suggest significant lithosphere removal from the Canadian Cordillera by means of delamination (e.g., Bao et al., 2014), thermal erosion from small-scale convection (e.g., Hyndman et al., 2005; Hyndman, 2017; Francis et al., 2010), and continental collision (e.g., Johnston, 2008; Chen et al., 2019). Crustal and upper-mantle structures, such as the Mohorovičić discontinuity (Moho) and lithosphere-asthenosphere boundary (LAB), provide the mechanisms driving lithosphere removal.

While reflection studies provide a view of an anomalously flat Moho across the NCC, despite crossing major faults and various terranes, deeper lithosphere structure is still poorly understood (Cook et al., 2004; Cook et al., 2005; Clowes et al., 2005). Historically, station coverage in the NCC has been sparse in both time and space. The deployment of portable seismometers such

as the CAnadian NOrthwest Experiment (CANOE) (Gaherty and Revenaugh, 2003) have helped bridge spatial gaps but lack longevity. Nevertheless, recent broadband permanent seismometer deployments such as the Yukon-Northwest Seismic Network (NY) (University of Ottawa, 2013) have increased coverage and facilitated the construction of improved seismic models.

Regional studies using body-wave (e.g., Frederiksen et al., 1998; Mercier et al., 2009; Zaporozan et al., 2018; Estéve et al., 2020) and surface-wave tomography (e.g., Dalton et al., 2011; Bao et al., 2014; McLellan et al., 2018) have provided insight into broad shear-wave velocity structure (V_s), confirming the mantle beneath the Cordillera possesses extremely low V_s and is, therefore, warm compared to the craton. SKS splitting (e.g., Courtier et al., 2010), surface-wave tomography (e.g., Bao et al., 2016), and receiver-function (RF) studies in the region (e.g., Cassidy, 1995; Taryoun et al., 2017; Audet et al., 2019) have been inferred to show inherited anistoropy within the lithosphere and flow-driven anisotropy in the asthenosphere.

The Tintina Fault (TF), a right-lateral strike-slip fault, runs parallel to the CCB and corresponds to an abrupt change in seismic velocities and anisotropy (Snyder et al., 2005; Zelt et al., 2006; Mercier et al., 2009; Dalton et al., 2011; McLellan et al., 2018). Moreover, chemical and isotopic changes in alkaline basalts across the TF suggests it straddles two distinct lithosphere mantles (Abraham et al., 2001). These observations suggest that the TF is a primary boundary between the young, warm lithosphere to the west and older, colder lithosphere to the east.

Further investigation of how the Moho and LAB act on a lithosphere scale across the Cordillera provides insight into mechanisms driving the thermal and physical state of the Cordillera. The LAB is challenging to infer, because its geophysical signature is subtle and obscured by other reverberations (Eaton et al., 2009). However, the young, hot lithosphere at shallow depths in the NCC makes the LAB likely easier to observe, provided a shallow, sharp LAB (e.g., Kawakatsu et al., 2009; Chen, 2009; Kind et al., 2012). Moreover, mantle-derived xenoliths in the northern Cordilleran volcanic province (NCVP) help constrain temperatures, depths of the Moho and LAB and origin of mantle rock (Edwards and Russell, 2000; Harder and Russell, 2006; Kilgore et al.,

2018). Improving the understanding of the structure and how the crust, LAB and upper mantle act across the Canadian Cordillera can provide insight into the thermal and physical processes that shaped the western margin of North America.

Complex seismic structure in the NCC makes inverting for V_s challenbing without making assumptions about model parametrization. Inverting RF data is an inherently non-linear and highly non-unique problem; earth models comprised of vastly different parameters can produce similar fits to RF data (Ammon et al., 1990). This non-uniqueness stems from RFs being sensitive to discontinuities in V_s but having poor resolution of absolute values of V_s (Ammon et al., 1990; Juliá et al., 2000). However, incorporating additional information from surface-wave data observations, which are sensitive to absolute V_s, can alleviate this limitation (Juliá et al., 2000).

The Bayesian approach is ideally suited to solve this inverse problem because it quantifies the uncertainty due to the non-uniqueness of the solution and reduces subjective choice (Malinverno, 2002; Tarantola, 2005). Non-linear optimization algorithms can find a global minimum and provide an optimal solution, or a single, best solution. However, optimal solutions ignore the intrinsic non-uniqueness of the problem and preclude rigorous uncertainty quantification. By producing ensembles of the model-parameter vectors in a Bayesian framework, the inversion solution is an approximation of the posterior probability distribution (PPD) and quantifies the degrees of belief about earth structure given the data. The PPD is constrained by data information and prior information. A Bayesian formulation is able to avoid subjective parameterization by treating the inversion as trans-dimensional, where the number of layers as unknown.

I apply Bayesian joint inversion of surface-wave dispersion (SWD) and RF data based on the algorithm from Dettmer et al. (2015) to data in the NCC. An effective exploration of the parameter space is done by integrating the Markov chain Monte Carlo (MCMC) method. The algorithm employs parallel tempering (PT, (Geyer, 1991; Dettmer and Dosso, 2012)), an interactive MCMC method, to sample the PPD. The PT algorithm employs a sequence of Markov chains with increasingly relaxed (tempered) posterior bridging distributions with at least one chain sampling a

distribution identical to the PPD. Only the chain sampling the PPD is used to record samples; the other bridging distributions only contribute sampling efficiency.

Tempering the chains allows for freer movement in the parameter space and higher acceptance rates of the proposed model vectors resulting in a broad exploration of the parameter space. The Metropolis-Hastings (MH) criterion is applied to implement the information exchange between chains and improves efficiency.

This thesis begins by introducing the geologic background of the NCC. Then, I discuss RF processing and SWD acquisition. Subsequently, I describe the Bayesian inversion algorithm. I successfully apply this method to RF and SWD data in the NCC to investigate the Moho and LAB across the NCC. I conclude with a comparison of my observations and previous studies and a discussion of the thermal and geodynamic implications of my research results.

2. Geological Background

2.1 Tectonic Setting

Bounded to the east by the undeformed strata of Ancestral North America, the American Cordillera is a north-south trending mountain belt that extends from Alaska to Patagonia (Kay and Ramos, 2007). Within the American Cordillera, the Canadian Cordillera is comprised of the region that extends from Alaska to the southern US-Canadian border (Figure 2.1). The Canadian Cordillera is currently characterized by a sharp boundary between ancient North America and a Phanerozoic orogenic belt. Topography and elevation reflects the lateral extent of this deformation, observed at the surface as the Cordillera deformation front (CDF).

The NCC was founded on the western margin of the Laurentian craton. It has evolved as an active belt from its inception in the Late Precambrian to the present and has been through diverse tectonic settings and configurations (e.g., Hoffman, 1988; Colpron et al., 2007; Nelson and Colpron, 2007). The Cordillera is formed of an amalgamation of terranes, fault-bounded blocks with distinct geological records (Jones et al., 1983), which generally are older in the east and younger in the west.

During the breakup of Rodinia $\sim$ 750–540 Ma, the ancient passive margin of Laurentia was converted to a convergent margin, marking the beginning of the NCC development (Bond and Kominz, 1984; Monger and Price, 2002; Nelson and Colpron, 2007; Johnston, 2008). The North American craton is encased by the Intermontane terrane and was originally bounded by an outer relic-fore-arc ocean-ward margin (Colpron et al., 2007). The Intermontane terrane grouping is

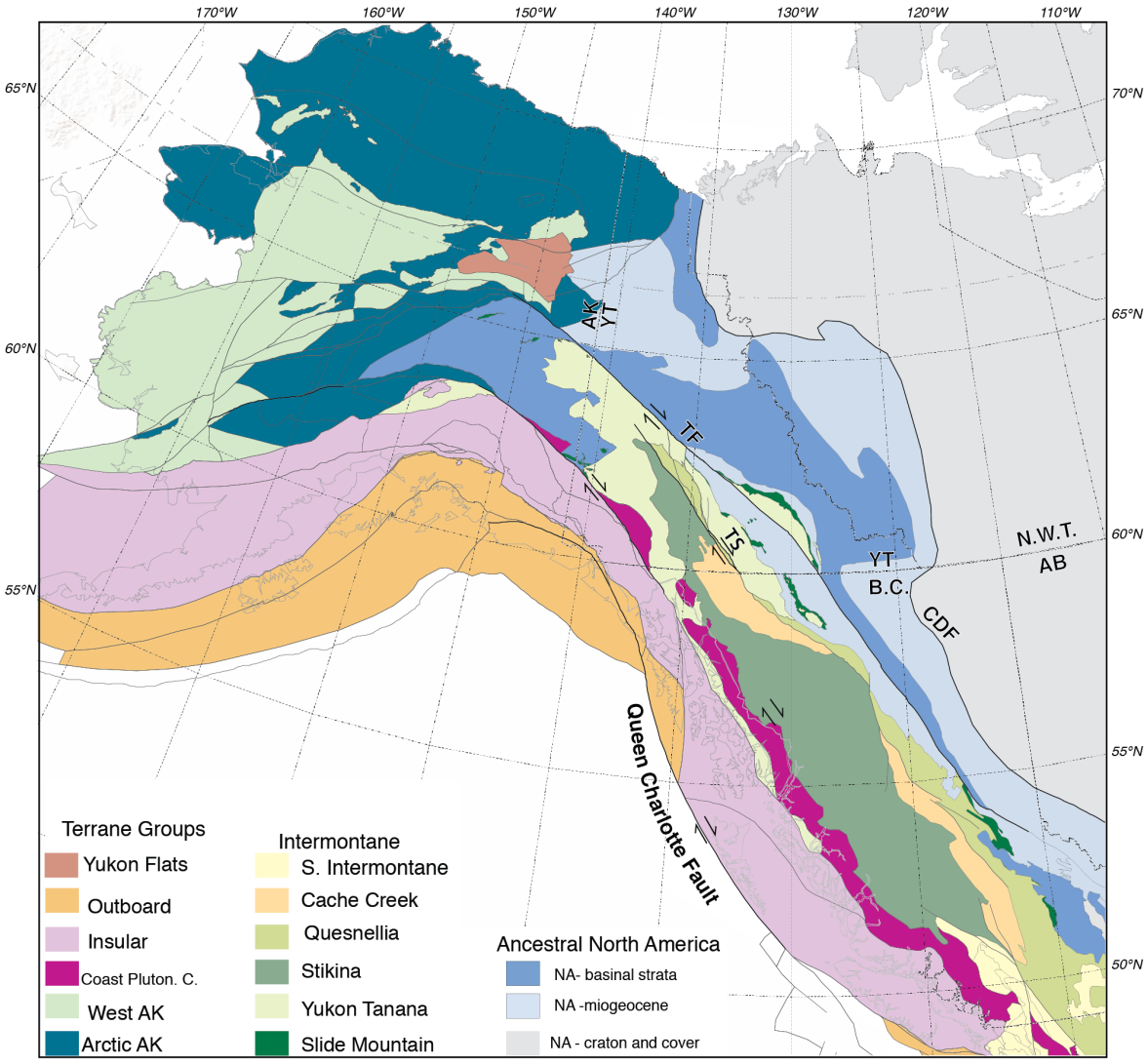

Figure 2.1: Terrane map for the Canadian Cordillera. Terrane groupings and tectonic realms are explained with the legend. Abbreviations represent respective provinces and states: AB- Alberta; AK - Alaska; B.C. - British Columbia; N.W.T. - Northwest Territories; and Y.T.- Yukon Territory. (B.C. and Yukon Geological Survey, 2015a,b).

often referred to as a *superterrane*, as it is an amalgamation of Triassic-Jurassic terranes that joined together prior to accretion (Johnston, 2001; Monger et al., 1982; Johnston, 2008). This belt consists of oceanic assemblages with a similar tectonic history as ancient North American margin, such as remnant slivers of the Slide Mountain oceanic terrane (Monger and Gibson, 2019). Arctic and Insular terranes origins had no relationship to the North American margin prior their accretion on the outer margin of Intermontane terranes by $\sim$ 160 to 150 Ma (van der Heyden, 1992).

With the exception of the Yukatat Terrane in Southeast Alaska, major terrane accretion was completed by the Cenozoic (Nelson and Colpron, 2007). The west coast of the Cordillera was primarily characterized by the eastward subduction of the Kula, Resurrection and Farallon plates by $\sim$ 53 Ma (Madsen et al., 2006). Following the cessation of subduction of the Kula and Resurrection plates, transverse motion from the Queen Charolette-Fairweather fault cut off continuous subduction of the Pacific plate by $\sim$ 42 Ma (Madsen et al., 2006). The shift from a convergent to a transform boundary and the reorganization of stresses to a transtensional environment was accommodated by dextral fault system displacements of crustal rocks $\geq$ 400 km along the Tintina and Denali faults (Hayward, 2015).

Subduction continued and is currently subducting to the north and south of the NCC, where the Pacific Plate is subducting beneath Alaska, and the Explorer and Juan De Fuca plates are subducting beneath the Cascade arc (Madsen et al., 2006; Thorkelson et al., 2011). This break in the subducting slabs, or *slab window*, induced volcanic activity throughout the NCC as upwelling asthenosphere rose to replace the zone once occupied by the slab (Thorkelson et al., 2011). The NCVP encompasses the intraplate volcanism in northwestern B.C., the Yukon Territory, and Alaska is driven by these changes in relative plate motion 10–15 Ma (Edwards and Russell, 2000).

Mafic magmatism in the NCVP involves two xenolith suites corresponding to separate sources: spinel lhzerolite from a shallow, lithsophere source and harzburgite from a deeper astheonsphere source (Francis et al., 2010; Polat et al., 2018). The spinel lherzolite from shallow lithospheric melt is likely representative of the regional lithospheric mantle (e.g., Shi et al., 1998; Francis et

al., 2010; Polat et al., 2018). The second harzburgite suite appears to be fed by melt from a deep, asthenosphere source and resulted from $\sim$ 15% melting from the spinel lhzerolites (Francis et al., 2010). Metasomatized mantle in the Paleogene converted to anhydrous asthenosphere/veined lithospheric mantle with the opening of the slab window (Thorkelson et al., 2011). While the lithosphere is largely anhydrous, water content in the NCC is from the remelting of carbonatite veins rather than the incorporation of water from subducting slab (Kilgore et al., 2018).

2.2 Crustal and Moho structure in the NCC

The Moho appears as a strong seismic signal, as it is a compositional boundary and there are large differences in seismic parameters between the crust and uppermost mantle. The Moho under the Canadian Cordillera is anomalously flat despite underlying an extensively faulted and heterogeneous system (Cook et al., 1992; Cook et al., 2004, Cook et al., 2005, Clowes et al., 2005). Complexities in lithospheric structure are often not associated with any surficial expression, so the CDF does not necessarily mark the CCB at depth (e.g., DiCaprio et al., 2020). Dynamic topography, or the topography driven by density changes from mantle heterogeneity (e.g., velocity anomalies), can account for excess elevation, observed in regions such as the African super-swell (e.g., Lithgow-Bertelloni and Silver, 1998). The influence of dynamic topography is also present in the NCC, and where detachment zones may have been driven by flow in the lower crust and mantle small-scale convection due to high Cordillera-wide temperatures, discussed in Section 2.3 (e.g., Currie and Hyndman, 2006; Eaton, 2005; Hyndman, 2017). Density reduction due to thermal buoyancy also explains the high topography of the Cordillera despite relatively thin crust (Hyndman, 2017). Alternatively, the apparent disconnect between a flat Moho and heterogeneous upper-crustal structures has been interpreted to be evidence for the thrusting of thin slices of accreted crustal material upon the North American crust and mantle (Snyder et al., 2005; Cook et al., 2004, 2005; Dalton et al., 2011). Ancestral North America province has been inferred to extend to the TF from observations of abrupt changes in seismic anisotropy and velocity

at the TF (e.g., Courtier et al., 2010; Dalton et al., 2011) and to the CDF via body wave and surface wave tomography studies (e.g., Mercier et al., 2009; Zaporozan et al., 2018; Estéve et al., 2020). In the southern Canadian Cordillera (SCC), the TF is observed as the Rocky Mountain Trench, and it is inferred that the two features are related, particularly in their role in the formation of the Cordillera. Determining the lateral westward extend of of ancient North America is important when discussing the formation of the CCB and the mechanisms that formed the Cordillera.

2.3 Lithospheric and Upper-mantle Structure in the NCC

The LAB is often more difficult to detect with seismic methods than the Moho, as it is not a compositional boundary and is generally not associated with a sharp contrast. It is defined as a boundary between the rheologically distinct rigid lithosphere and plastic asthenosphere (e.g., Eaton et al., 2009; Fischer et al., 2010). The LAB can also be defined as the transition from where heat is primarily transferred through conduction in the lithosphere, to where heat is transferred through convection in the asthenosphere (Sleep, 2005). Because of this, the LAB can be difficult to image in cratonic provinces but easier to image in Cordilleran regions, because the high heat flow and the presence of partial melt can make it easier to infer (e.g., Chen, 2009; Abt et al., 2010; Kind et al., 2012).

Therefore, an interdisciplinary approach is an appropriate way to accommodate multiple definitions and interpretations of the LAB. While seismic methods are able to observe composition or partial melt variations at depth, petrologic data can be used to infer the depth of the LAB based on partial melt calculations. Magnetotelluric (MT) studies can also detect low-resistivity structures and interpret the presence of partial melt associated with the LAB.

Previously, LAB depths for the NCC have been estimated at 50-km depth using seismic methods (Audet et al., 2019) and between 52–70-km depth in xenolith studies (e.g., Harder and Russell, 2006; Francis et al., 2010; Batir and Blackwell, 2019). Compared to the craton LAB depth of $\sim$ 150–250 km, this range of estimates for the Cordilleran LAB is shallow. Xenolith studies

suggest that the shallow LAB relates to a hot lithosphere with temperatures ranging from 1100 to 1350° C (e.g., Edwards and Russell, 2000; Harder and Russell, 2006; Batir and Blackwell, 2019). Temperatures at the Moho and LAB provide indications of the LAB depth and geothermal gradient. Modelling of the geothermal gradient assuming a long-lived lithospheric structure, LAB depth, and Moho temperature based on Langmuir et al. (2006) suggests a LAB-depth of 75 km corresponds to a Moho at 800° C, while a LAB of 50-km depth corresponds to a Moho at 1000° C (Audet et al., 2019). This geothermal gradient can also be inferred with xenoliths, seismic, and MT studies suggesting Moho temperatures ranging from 800–1000° C (e.g., Harder and Russell, 2006; Francis et al., 2010; Clowes et al., 2005; Ledo and Jones, 2005; Audet et al., 2019). Overall, LAB temperatures between 1100°–1300° C (e.g., Harder and Russell, 2006; Edwards and Russell, 2000) indicate a hot, near-adiabatic mantle at 150 km (Chen et al., 2019). These high geotherms indicate that the lithospheric mantle may contain partial melt from decompressional melting (Thorkelson et al., 2011; Audet et al., 2019).

Heat-flow data in the Cordillera (Figure 2.2) is 100 mW/m^2, which is high compared to heat flow in the craton measured at $\sim$ 42 mW/m^2 (Lewis et al., 2003; Grasby et al., 2012). Even when accounting for radiogenic heat production in the crust, the heat flow across the Cordillera is higher than the craton (Lewis et al., 2003). Such extreme variations between the Cordillera and the craton lead to a significant decrease in effective elastic thickness across the CCB, further indicating the presence of a hot, weak lithosphere and a shallow brittle-ductile transition zone within the NCC (Lewis et al., 2003; Flück et al., 2003; Audet et al., 2007; Hyndman, 2010).

The mechanism of lithosphere removal that lead to the shallow LAB in the NCC remains debated, with possible mechanisms including a large-scale lithosphere delamination event (e.g., Bao et al., 2014), viscous thermal erosion via small-scale thermal erosion induced by upwelling asthensophere (e.g., Hyndman et al., 2005; Hyndman, 2010, 2017; Francis et al., 2010), and terminal suturing during accretion (e.g., Chen et al., 2019).

A terminal suture zone model proposed for the SCC suggests a ribbon continent crust and

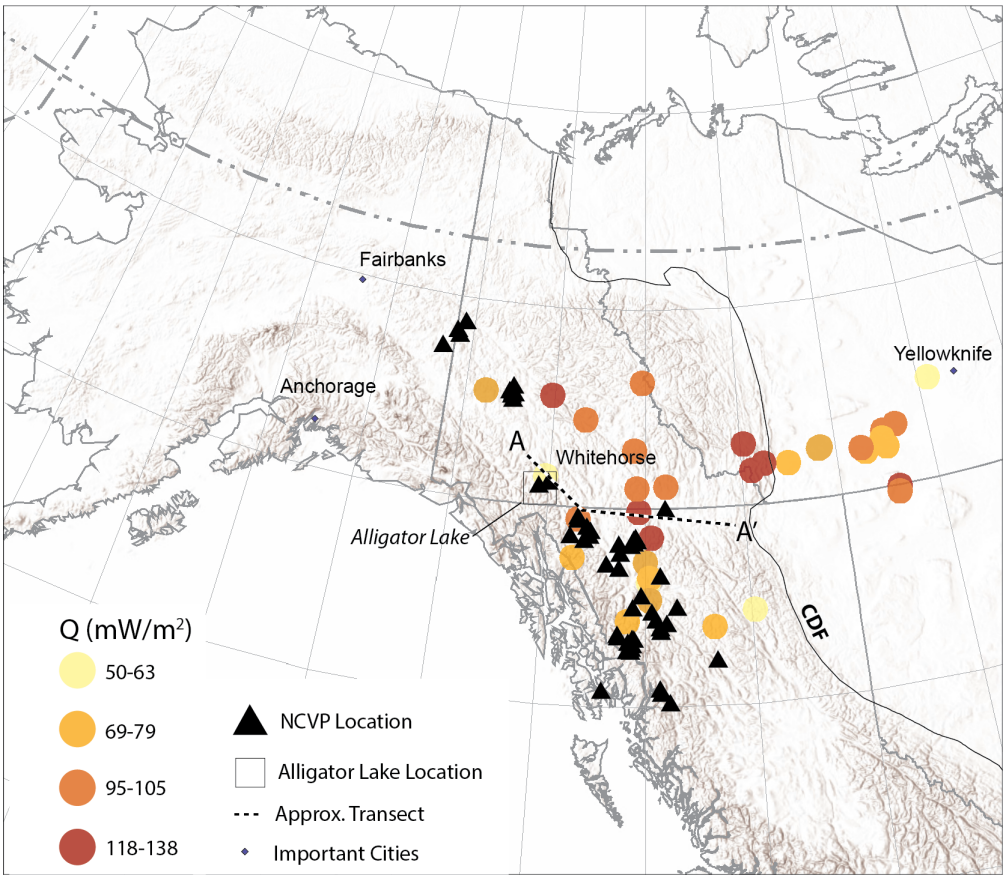

Figure 2.2: Heat flow data across the NCC from Lewis et al. (2003) and NCVP locations from Edwards and Russell (2000). The Alligator Lake NCVP location is noted, as multiple studies referenced in this thesis determine temperatures and lithospheric composition from Alligator Lake xenoliths. An approximate transect of the stations discussed later in this thesis is noted by a black, dashed line, marked from A to A'.

lithosphere was accreted onto the existing craton (Chen et al., 2019). In this case, the LAB is shallow because it is from an exotic source. This model explains the steep, well-preserved west-dipping geometry of the collisional front, sharp velocity contrast between the Cordillera and the craton. However, an exotic lithosphere model fails to address the extreme heat flow patterns observed in the NCC.

Alternatively, the opening of the slab window may have triggered a large-scale delamination event between 5–25 Ma (Bao et al., 2014; Audet et al., 2019) and allowed the displacement and foundering of the lithospheric root (Bao et al., 2014; Thorkelson et al., 2011). However, post-Eocene faults are younger than the suggested delamination event, and it is difficult to reconcile the removal of the lithosphere with present-day motion of North America (Chen et al., 2019).

A more plausible mechanism for removal of the lithosphere is ongoing small-scale convection induced by the high heat flow in the NCC. In this model, the hot mantle creates shallow convection cells (e.g., Currie and Hyndman, 2006; Hyndman et al., 2005; Hyndman, 2017), and the crust is thin and decoupled with the underlying weak upper-mantle (Oldow et al., 1990). Observations of elevated upper-mantle temperatures (e.g., Lewis et al., 2003; Francis et al., 2010), low effective elastic thickness (e.g., Flück et al., 2003; Audet et al., 2008) and a flat, shallow Moho (e.g., Clowes et al., 2005) agree with this model.

All models can explain the shallow LAB previously inferred in the NCC. However, further observations of the lithospheric mantle structure will constrain which mechanisms are driving lithospheric thinning. Understanding the lithosphere removal process will better constrain on which time scales these mechanisms operate and what role the slab window and heat flow plays in this process.

3. Theory

3.1 Receiver-Function Estimation and Surface-Wave Data

I investigate the structure of the crust and upper mantle in the NCC by exploiting the principles of conversion and reflection of seismic body waves at interfaces via RFs and frequency dispersion of surface waves. In this chapter, I discuss two signal processing methods that are commonly applied in RF estimation: water-level deconvolution and iterative-time domain (ITD) deconvolution. While there are other methods to isolate converted waves, these methods are ideal for their simplicity and ability to be combined to address absolute velocity and seismic boundary depths. In addition, I introduce SWD curves that are used to provide additional constraints on earth structure.

3.1.1 Waves and Conversions

Waves travel through the interior of earth as *body waves* and along its free surface as *surface waves*. Body waves propagate through the earth with particle motions in two directions: along the direction of propagation (P waves) and perpendicular to propagation (S waves). The P waves, or primary waves, are the fastest body wave, composed of compression-dilatation motion and some shearing. The S waves, or secondary waves, travel through the earth with pure shear motion.

The S waves exhibit two polarizations of particle motion: within the vertical plane (SV) and the horizontal plane (SH). In media that behave identically in every direction (*isotropic media*), P waves are coupled with SV waves and SH waves propagate separate from the P-SV system (Langston, 1977). When media exhibit directional dependence (e.g., the presence of aligned olivine crystals or a dipping boundary), this assumption is violated, and SH waves are introduced

into the P-SV system (Langston, 1977). Anisotropic effects are minor for P waves and more aligned with the direction of propagation than anisotropic effects on S waves (Crampin et al., 1982). I focus my discussion on P and SV waves for the remainder of this thesis, but discuss complications arising from anisotropy and lateral heterogeneity throughout.

When P and SV waves encounter seismic boundaries (e.g., changes in density and/or velocity) in the earth, their energy is partitioned into reflections, transmissions and conversions to other wave types. In particular, P waves can convert to SV-waves and vice versa. Seismic stations record three orthogonal components of motion and can be rotated to the radial (R), transverse (T), and vertical (Z) components. Rotation into this (R, T, Z) reference system approximately separates the directions of particle motions of P, SV and SH waves.

Isolation of converted wave energy from source and path effects will produce a waveform influenced only by structure underlying the receiver. By deconvolving an estimate of the P-wave source from the radial component of motion, source and path effects are removed. The direct P-wave arrival acts as a proxy for the source-time function and includes path effects and the instrument response (Langston, 1977). Knowledge about an effective source-time function (eSTF) is generally not available independently. This issue can be overcome by making the assumption that the three components of a seismogram separate P and SV particle motions reasonably well. In the case of (R, T, Z) components, the direct P-wave signal is predominantly recorded on the vertical (Z) component, when the teleseismic P-wave incidence angle is nearly vertical below the station. Note that no conversions are produced when the incidence angle is purely vertical with respect to a horizontal seismic discontinuity.

Seismic recordings of earthquakes with epicentral distances between 30–100° from the receiver (*teleseismic events*) can be used to observe P-wave to SV-wave (Ps) conversions in the earth, as schematically described in Figure 3.1. Events closer than 30° may have significant SV energy on the vertical component of the recordings and arrive at too shallow incidence angle, so that crustal effects will complicate the signal and the isolation of the P-wave energy. Events further than 100°

will diffract at the core-mantle boundary, adding significant complexity to the waveform. Such complexities to the signal make the removal of the eSTF more difficult, so only teleseismic events are employed in RF estimation.

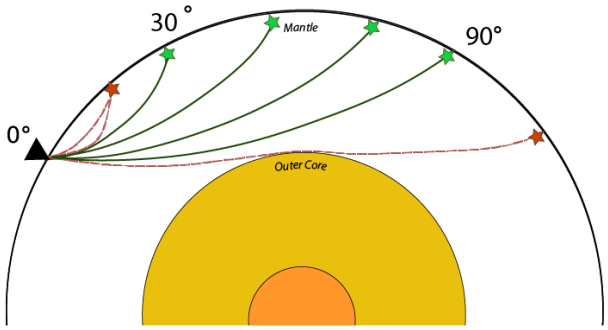

Figure 3.1: Schematic representation of earthquake distances desired in RF calculations after Schneider et al. (2014). Earthquakes between 30° and 90° epicentral distance are ideal in the RF estimation process, as rays have near-vertical incidence beneath the station. Earthquakes originating at $< 30°$ distance will have too much SV-energy on the vertical component for RF analysis. Earthquakes originating from 90° distance can diffract at the core-mantle boundary, complicating the P waveform and RF analysis.

As schematically represented in Figure 3.2, the direct P wave is the first arrival in P-RFs, followed by both primary conversions and multiples. Note that Figure 3.2 does not show other P-wave phases, such as PmP, because these are generally removed by RF processing described below. The observed particle motion at the surface is described as

$$u_i(\mathbf{x},t) = S(t) * I_{in}(\mathbf{x},t;\mathbf{p}\perp), \tag{3.1}$$

where u_i is the observed displacement in spatial direction i at location $\mathbf{x}$ and time t, $S(t)$ is the eSTF, and $*$ represents convolution (Bostock, 2007) . Importantly, I_{in} is a plane wave impulse response function for the receiver-side structure, where the plane wave is assumed to be incident from below (upgoing). The dependence of I_{in} on horizontal slowness $\mathbf{p}\perp$ is required to account for various incidence angles due to considering earthquakes at various epicentral distances and depths (Bostock, 2007).

The impulse response in eq. (3.1) for the receiver-side structure is the desired quantity to be isolated to study structure beneath the station. Therefore, the goal is to remove the eSTF, which

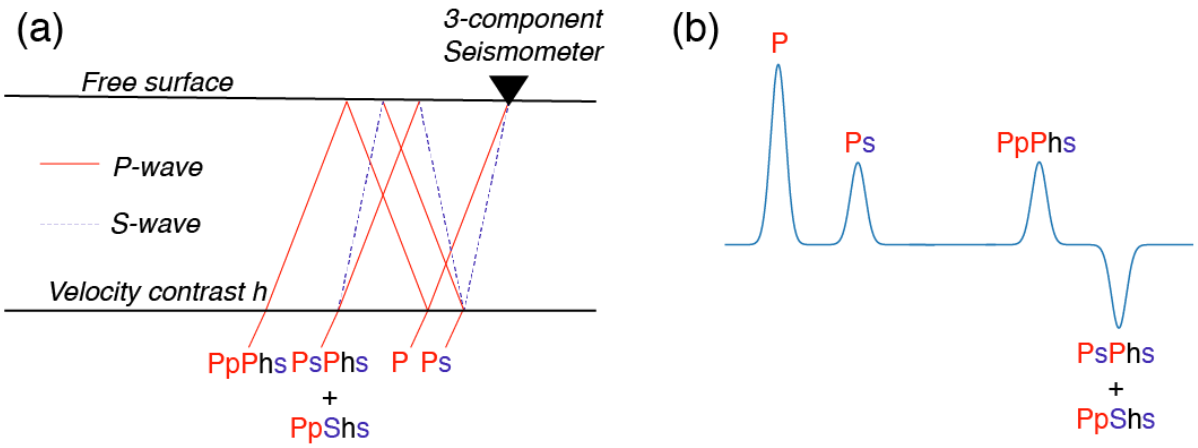

Figure 3.2: (a) Conversions and multiples of P and SV waves from a velocity contrast h. (b) Schematic seismic recording of the conversions arriving at the surface. Modified from Ammon (1991).

combines the source and path effects on the incident wave from the seismogram.

3.1.2 Rotation

Seismogram rotation into the (R, T, Z) coordinate system approximately separates components with P-, SV- and SH- energy. The radial component is aligned with the back-azimuth (BAZ) obtained from earthquake catalogues. With this information, the rotation is applied by matrix multiplication

$$\begin{pmatrix} Z \\ R \\ T \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\theta & sin\theta \\ 0 & -sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} Z \\ N \\ E \end{pmatrix},$$

where Z, N and E denote the vertical, north and east components of the seismogram, and θ is the BAZ angle.

While the vertical component generally contains S-wave motion for obliquely incident waves, using events with at teleseismic distances justifies the assumption that P-wave energy is largely confined to the vertical component. More advanced methods employ rotation to (P, SV, SH) coordinates and account for free surface effects (Kennett and Engdahl, 1991; Bostock, 1998).

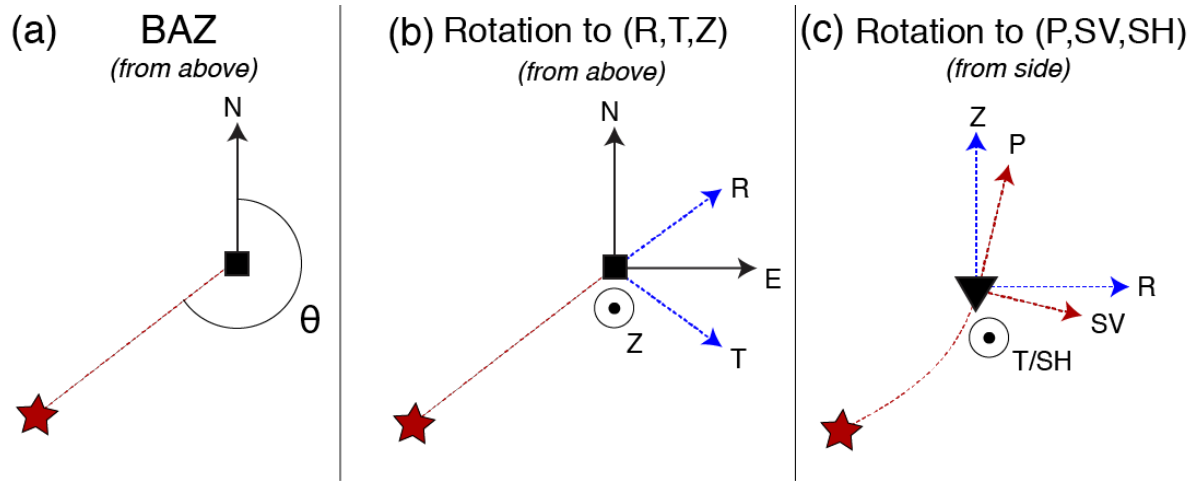

Figure 3.3: Schematic representation of (a) conversions BAZ angle and rotation into (b) (R, T, Z) and (c) (P, SV, SH) coordinate systems. Red star represents the earthquake, the black box represents the seismometer from above, and the black triangle represents the seismometer from the side. N and E coordinates are shown by black arrows, R and T and Z components by blue dashed arrows, and P and SV components with red dashed arrows.

However, this rotation requires knowledge of surface V_p and V_s, density ρ and slowness ($p = 1/V_p$). I elect to remain in (R, T, Z) coordinates in order to avoid the introduction of unnecessary parameters.

In the presence of dipping or anisotropic structure, direct and converted phases exhibit amplitude and arrival time variability with respect to BAZ and slowness on the radial and transverse components (e.g., Langston, 1977; Cassidy, 1992; Frederiksen and Bostock, 2000). A simple model showing these effects of a dipping layer on amplitude and arrival time is shown in Appendix A. In the presence of a dipping layer, P-wave amplitudes vary as a function of BAZ, so that P waves travelling up-dip generate the largest and latest Ps arrivals. Conversely, waves travelling down-dip generate the earliest, smallest arrivals. The direct P wave in the radial receiver function also exhibits azimuthal dependence, which is most pronounced for steeply dipping interfaces (e.g., dipping more than 30°). Therefore, normalizing to unit amplitudes neglects this information. Moreover, dipping layers deflect P and SV waves and introduces a transverse component of ground motion to the R-Z plane (Langston, 1977). Similar azimuthal variation is exhibited in the presence of some forms of anisotropy. Considering this directional dependence, RFs are stacked within a

BAZ window. RFs are stacked from a small corridor of one direction to reduce distortion of RF features from BAZ-dependent conversion arrivals.

In the next subsections, I discuss two methods of eSTF removal: water-level deconvolution and ITD.

3.1.3 Water-level Deconvolution

As discussed, RF estimation assumes the Z component of a seismogram is a good approximation of a P-wave eSTF and the R component is assumed to contain both P-wave energy and converted S-wave energy. Isolating S wave arrivals on the radial component is done by removing P-wave eSTF via deconvolution.

In general, RF deconvolution is achieved by dividing the radial frequency spectra $\hat{R}$ by the vertical spectra $\hat{Z}$

$$H(\omega) = \frac{\hat{R}(\omega)}{\hat{Z}(\omega)}, \tag{3.2}$$

where H is the Fourier transform of the receiver-function estimate, and ω is circular frequency. This process involves division, which means instability issues arise when $\hat{Z}$ becomes small. This is particularly problematic when signals contain noise. Non-random signal generated noise cannot be addressed in water-level deconvolution, but the random fluctuations can be addressed by avoiding small numbers altogether. One way to do so is by including a *water-level* in the denominator (Langston, 1977).

Water-level stabilization is achieved by

$$H(\omega) = \frac{R(\omega)Z^*(\omega)}{\phi(\omega)}G(\omega), \tag{3.3}$$

where $Z^*(\omega)$ denotes the complex conjugate of $Z(\omega)$, G is a lowpass filter, and

$$\phi(\omega) = max\{Z(\omega)Z^*(\omega), c * max\{Z(\omega)Z^*(\omega)\}\}. \tag{3.4}$$

In eq. (3.4), c is the water-level, a constant that determines the minimum amplitude in the denominator of H (Ammon et al., 1990). The water-level removes frequencies for which the vertical

component has small amplitudes, which can lead to instabilities during division. Ideally, the smallest water-level that produces acceptable noise levels in the receiver function is chosen, but must be determined through trial-and-error. Water-level selection is dependent on noise, so it may not be consistent among all stations, thus introducing subjectivity in RF estimation.

The filter G is typically chosen as a low-pass Gaussian filter to reduce high-frequency noise with

$$G(\omega) = \xi \exp(\frac{-\omega^2}{4a^2}), \tag{3.5}$$

where the constant ξ normalizes the filter to unit area and a controls the filter width. The filter gain is unity, and the Fourier transform of a Gaussian is also Gaussian. The mean of G is chosen to be zero and this filter acts as a low-pass filter. As a general rule, the width of the Gaussian pulse in the time domain is $\sim 5/\sqrt{3a}$ (Ammon, 1991). Choosing the filter width is subjective and controls which frequencies are contained in each RF. Layer thickness resolution is another consideration when selecting a Gaussian filter, because higher Gaussian filter values, such as $a = 5$ resolve thin layers of 2–3 km and lower Gaussian filter values, such as $a \sim 1$, successfully resolve thicker layers ≥ 10 km (Cassidy, 1992). For resolution of layers on a lithosphere scale, a Gaussian filter value of $a = 1$ reasonably retains desired information from waveforms in the deconvolution process (Langston, 1979; Cassidy, 1992).

3.1.4 Iterative Time-domain Deconvolution

Water-level deconvolution is applied in the frequency domain, so it can be highly susceptible to instabilities arising from noise. Deconvolution is an inverse problem and it can also be solved in the time domain by applying linear inverse theory (Menke, 2012). While sufficient signal-to-noise ratios and azimuthal coverage lead to similar RFs between different methods, time-domain methods can be advantageous for noisy data (Ligorria and Ammon, 1999). However, with poor signal-to-noise ratios, inadequate azimuthal coverage or short deployments, the chosen RF method becomes more critical (Ligorria and Ammon, 1999).

The ITD deconvolution method was first applied in earthquake source studies (Kickuchi and Kanamori, 1982) and later employed to construct RFs via summation of Gaussian pulses (Ligorria and Ammon, 1999). The ITD method applies the least-squares minimization between the observed horizontal seismogram and predicted signal with a desired misfit tolerance to determine the number of Gaussian pulses utilized to compute a RF. Pulse widths are given by the Gaussian filter defined as for water-level deconvolution. The Gaussian pulses are convolved with an impulse response function for the receiver-side structure. While this process is computationally less efficient than water-level deconvolution, it is more stable for noisy data (Ligorria and Ammon, 1999). Moreover, the quality of the RF is quantified by a misfit value, which can be used to decide which RFs to include in a stack.

The ITD convolution process begins with the cross-correlation of the vertical component with the radial component. This cross-correlation estimates the lag of the first and largest spike in the RF. The spike amplitude is then estimated (Kickuchi and Kanamori, 1982) and convolved with the vertical component seismogram. The resulting waveform is then subtracted from the radial component seismogram. This process is repeated to estimate the other spikes and amplitudes in the seismogram. Misfit between the convolved estimate and the radial component seismogram is calculated at each step, and the steps are repeated to estimate other spike lags and amplitudes. Additional spikes in the RF should reduce the misfit between the vertical and RF convolution of the radial component seismogram. Iteration halts when the reduction in misfit with additional spikes becomes insignificant, which is determined by a subjective misfit tolerance.

Scaling differences between water-level and ITD deconvolution exist due to different normalization conventions. Therefore, it is crucial to scale the RFs before interpreting amplitudes. For a RF with a Gaussian scale parameter of 1, the RF from ITD deconvolution must be normalized by 0.57 to retain a unit-area pulse and be consistent with water-level deconvolution (Ammon, 1991; Ligorria and Ammon, 1999). This normalization is particularly vital in inversion. For computational efficiency, the synthetic seismograms produced by RAYSUM (Frederiksen and Bo-

stock, 2000) are processed via water-level deconvolution to produce RFs. Note that deconvolution instability is not a concern for noise-free predictions.

3.1.5 Surface-Wave Dispersion Data

Discussion of seismic waves thus far has been limited to body waves, but surface waves are also valuable in inferring V_s structure. Surface waves with different periods travel at different velocities, a phenomenon referred to as *surface-wave dispersion* (SWD). SWD data are sensitive to V_s averages, and provide information about absolute V_s values. RF data resolve discontinuities but lack sensitivity to absolute values and SWD data resolve absolute velocities but lack sensitivity to discontinuities. Combining both data in joint inversion provides complementary information to better resolve earth structure (Juliá et al., 2000).

The separation of frequencies results in signal that can be described in terms of individual phases or an overall *wave packet*, where there is destructive interference between the two waves. Phase velocity is defined as the velocity at which the phase of a single frequency component of the wave propagates. Group velocity is the velocity at which the overall shape of the envelope of the wave packet propagates. Generally, phase velocity is greater than group velocity. Longer wavelengths penetrate deeper into the earth and therefore are more sensitive to larger V_s, assuming V_s increases with depth.

Surface-waves recorded travelling along the surface of the earth are either *Rayleigh* waves, which have motion in the vertical and radial direction, or *Love* waves, which are horizontally polarized. For the purposes of this thesis, I limit my discussion to Rayleigh waves, as these waves are caused by the P/SV system. Love waves exist within the transversely-polarized SH system, and are, therefore, sensitive to anisotropic effects, which makes inferring isotropic V_s structure challenging. Both group and phase velocities can also be used to estimate V_s structure, and combining phase and group velocities for different period ranges can successfully be applied in joint inversion. For the purposes of this thesis, inverting dispersion from one velocity is sufficient, because both velocities are sampled across the same range of periods and inverting both will not

add any information. Therefore, I focus my discussion on utilizing phase velocities.

3.2 Joint Bayesian Inference

Estimating V_s structure from RF and SWD data requires solving an inverse problem. Unlike the forward problem, where a unique and stable solution always exists, the inverse problem does not always have a single solution. In fact, there can be an infinite number of solutions (*non-uniqueness*), and small changes in data can lead to large model changes (*unstable*).

Instability in the inverse problem is more severe for ill-posed and highly non-linear problems. Small changes in the data, such as the presence of noise or data errors, can lead to large changes in the model. Data errors arise from measurement and theory errors. Measurement errors are easy to quantify and generally small, but theory errors are more difficult to quantify. Theory errors can come from multiple sources, such as simplified model parameterization or idealized physics of the forward problem.

In this study, I assume zero-mean Gaussian error distributions, so that Gaussian errors are symmetric around zero. Data can be biased, but biases can often be removed during signal processing. According to the central limit theorem, random noise from multiple sources of comparable size can be approximated by Gaussian distributions, regardless of individual source distributions (e.g., Blum et al., 1963; Rényi, 1963).

Classical linear inverse theory is based on *deterministic* statistics, where there is one true model, and the problem is linear or can be linearized, meaning it obeys superposition and scaling. Alternatively, the inverse problem can be solved by a *Bayesian* approach, where model parameters are considered as random variables to be described probabilistically. Classical linear inversion can be used for problems that are at most weakly non-linear, whereas Bayesian inference can be applied to both linear and fully non-linear problems and provides intrinsic uncertainty quantification.

The joint inversion of RF and SWD data is an inherently non-unique, strongly non-linear problem often affected by subjective choices in the starting model when linearization is applied.

Linearization is typically based on over-parametrizing the Earth model with many layers of fixed thickness and then applying regularized or damped methods to infer velocities. Treating layer thicknesses as unknowns causes problems for linearization approaches due to significant nonlinearity.

However, computationally expensive non-linear methods can be applied to solve these issues. To address high computational cost, models are typically under-parametrized, where velocities and thicknesses for a few layers inferred. Under-parametrized models have the limitation of potentially missing important Earth features, under-estimating uncertainty, and relying on subjective choice with respect to model parametrization. Formulating the inverse as trans-dimensional treats the number of unknowns as a parameter to be inverted for (Dettmer et al., 2010). Bayesian inference with RF and SWD data allows observed data to determine model parameterization and reduces the effects of subjective prior choices. Bayesian inference is naturally parsimonious, preferring simpler models with fewer layers over more complex models (Bodin et al., 2012). This is consistent with *Occam's razor*, stating that a hypothesis or model should be as simple as possible.

In this section, I discuss the application of Bayesian inference to the joint inversion of RF and SWD data. First, I discuss how the model is parameterized, the background model, and prior bounds. I then examine Bayes' rule and how to efficiently sample the model space. Then, I discuss the treatment of data noise and uncertainty. Finally, I discuss how a single model is chosen, and uncertainty is quantified.

3.2.1 Bayes' Rule

The PPD of the model parameters $\mathbf{m}$ given the observed data $\mathbf{d}$ is given by Bayes' Rule

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}, \tag{3.6}$$

where $P(\mathbf{d}|\mathbf{m})$ data uncertainty distribution, quantifying the probability that a particular set of parameters gave rise to the data. The prior probability, $P(\mathbf{m})$, represents independent information about the model. The normalizing constant, $P(\mathbf{d})$, is often referred to as Bayesian evidence or

marginal likelihood and provides model selection (Malinverno, 2000; Dettmer et al., 2009).

The Bayesian evidence, the denominator in eq. (3.6), can be interpreted as a likelihood of the model parameterization, a critical part of Bayesian model selection (Sambridge et al., 2006). The evidence normalizes the PPD and can be expressed as an integral of the likelihood times the prior over the model space,

$$P(\mathbf{d}) = \int P(\mathbf{d}|\mathbf{m})P(\mathbf{m})d\mathbf{m}. \tag{3.7}$$

The Bayesian evidence is not directly a function of model parameters, making direct estimation numerically challenging for high-dimensional, non-linear problems such as ours. While important, the un-normalized PPD is sufficient for a given model, due to the computational challenges it poses (Sambridge et al., 2006).

A trans-dimensional formulation incorporates the number of unknowns as an unknown that is integrated over, so that the resulting posterior distribution spans multiples spaces of different dimensions. By treating the problem as trans-dimensional, the posterior includes the uncertainty due to the limited knowledge of the model. The eq. (3.6) can be written as a Bayesian hierarchical model to include a parameter k,

$$P(k, \mathbf{m_k}|\mathbf{d}) \propto P(k)P(\mathbf{d}|k, \mathbf{m_k})P(\mathbf{m_k}|k), \tag{3.8}$$

where k is an index parameter of a single model that spans multi-dimensional spaces (Green, 1995; Dettmer et al., 2010). In eq. (3.8), $P(k)$ is the prior over K models to be considered and $(k, \mathbf{m}_k)$ are the parameters of the inversion for the trans-dimensional case of dimension M_k.

Observed data are fixed, which means $P(\mathbf{d}|\mathbf{m})$ may be interpreted as the likelihood of $\mathbf{m}$, $L(\mathbf{m})$. Formulating the likelihood function requires a distribution assumption about residual-error statistics. We assume Gaussian distributed random errors, so that residuals are the difference of prediction and observation ($\mathbf{r}_i = \mathbf{d}_i - \mathbf{d}_i(\mathbf{m})$, where $\mathbf{d}_i$ is the data vector of the data set and i indexes the number of $\tilde{S}$ data sets) . This leads to a likelihood function described as,

$$L(\mathbf{m}) = \prod_{i=1}^{\tilde{S}} \frac{1}{\sqrt{(2\pi)^{N_i}\mathbf{C}_{\mathbf{d}_i}}} \exp\left(-\frac{1}{2}\mathbf{r}_i^T \mathbf{C}_{\mathbf{d}_i}^{-1}\mathbf{r}_i \right), \tag{3.9}$$

where N_i are the number of data, and $\mathbf{C}_{\mathbf{d}_i}$ is the covariance matrix for the data set. The covariance matrix describes dependency of noise. Diagonal entries on the matrix are noise variances, and off-diagonal elements are co-variances. Error dependence has been shown to be significant for both RF (Bodin et al., 2012) and SWD data (Dettmer et al., 2012).

For the purposes of this study, S includes P-RF and SWD data, but can be used on any data with a positive definite $\mathbf{C}_{\mathbf{d}_i}$ with simplifying assumptions. Data sets are weighted appropriately by errors on data, so weighting does not require subjective specification.

The Metropolis-Hastings-Green (MHG) algorithm (Metropolis et al., 1953; Hastings, 1970; Green, 1995) is a Markov chain Monte Carlo (MCMC) algorithm that draws dependent samples from the posterior by simulating a random walk in the model space, governed by a proposal probability distribution, Q. The proposal depends only on the current model vector of the parameter vector. After proposing a candidate model vector ($\mathbf{m}'$), the algorithm must accept or reject the proposed model vector $\mathbf{m}'$ based on the MH acceptance criterion. If $\mathbf{m}'$ is rejected, the algorithm records the current model vector again and then proceeds to iterate this process until a candidate model vector is accepted, in which it becomes the current model vector.

MHG sampling ensures detailed balance, or reversibility, by drawing samples from the PPD of the Markov Chain, such that

$$P(\mathbf{m}|\mathbf{d})T(\mathbf{m'}|\mathbf{m}) = P(\mathbf{m}'|\mathbf{d})T(\mathbf{m}|\mathbf{m}'), \tag{3.10}$$

where $T(\mathbf{m'}|\mathbf{m})$ represents the transition probability of $\mathbf{m} \rightarrow \mathbf{m}'$, such that

$$T(\mathbf{m}'|\mathbf{m}) = Q(\mathbf{m}'|\mathbf{m})\alpha(\mathbf{m}'|\mathbf{m}), \tag{3.11}$$

where α is the acceptance probability. Note that proposing models from a distribution $Q(\mathbf{m}'|\mathbf{m})$ depends only on the current model. Satisfying this requirement gives a general expression of the MH acceptance criterion as

$$\alpha = \min\left[1, \frac{P(\mathbf{m}'|\mathbf{d})}{P(\mathbf{m}|\mathbf{d})}\frac{Q(\mathbf{m}|\mathbf{m}')}{Q(\mathbf{m}'|\mathbf{m})}\frac{L(\mathbf{m}')}{L(\mathbf{m})}|\mathbf{J}|\right], \tag{3.12}$$

where $|\mathbf{J}|$ is the determinant of the Jacobian matrix of the transformation for $\mathbf{m} \rightarrow \mathbf{m}'$. While this term is important when the dimensions of $\mathbf{m}$ and $\mathbf{m}'$ are different, the Jacobian term is unity ($|\mathbf{J}| = 1$) in this implementation (Dosso et al., 2014).

The number of parameters of an appropriate model of the environment is unknown, so MHG sampling allows jumps between parameter vectors with different numbers of layers. Layers in a stratified earth model are adjusted by adding new layers (*birth steps*), removing layers (*death steps*), or perturbing existing layers (Geyer and Møller, 1994). New layers are birthed by a process that splits an existing layer into two new ones by sampling a new interface position from a uniform proposal. The V_s value for a new layer is sampled from the uniform prior. A death occurs by randomly deleting an existing interface and assigning a V_s to the merged layer by randomly selecting values from the previous two layers.

Given that the proposal distributions for fixed-dimension steps is Gaussian (symmetric; e.g., $Q(\mathbf{m'}|\mathbf{m}) = Q(\mathbf{m}|\mathbf{m'})$) and the proposals for changes in dimension are taken from the prior (Dosso et al., 2014), the acceptance probability for all steps in this implementation simplifies to

$$\alpha = \min\left[1, \frac{L(\mathbf{m'})}{L(\mathbf{m})}\right]. \tag{3.13}$$

The acceptance probability always accepts the proposed model if $L(\mathbf{m'}) \geq L(\mathbf{m})$. Proposed models with $L(\mathbf{m'}) < L(\mathbf{m})$ are accepted with finite probability.

A sub-sample of the accepted models are recorded and saved in the algorithms' ensemble output. The output from the first part of the chain, the *burn-in* period, is discarded. Discarding the burn-in ensures unbiased sampling of the part of the model space supported by observed data. Once the algorithm undergoes sufficient iterations, this sub-sample will provide a converged approximation of the PPD.

3.2.2 Model Parameterization and Data Prediction

Data predictions for RFs and SWD requires forward modelling with the Earth model. I assume the Earth model to consist of homogeneous, isotropic, horizontally stratified layers. As depicted

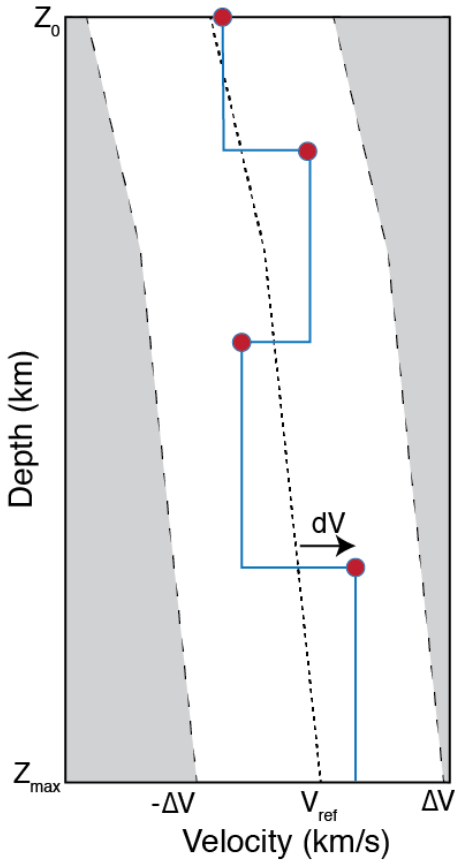

Figure 3.4: Schematic representation of the inversion parameterization, where the blue line represents the model given by the red dots, which represent the V_s perturbations for each layer. Shear-wave velocity is constrained within the white area, determined by prior bounds around a velocity reference model (dashed lines). Depths Z_{min} and Z_{max} are the minimum and maximum depth allowed for layer parameters, with each layer interface situated halfway between adjacent nuclei. No layer node may be placed outside the V_s bounds, the grey-coloured region.

in Figure 3.4, the model described in terms of a variable number, (k), interfaces that partition the Earth into layers to some maximum depth, ($z_{\max}$), of interest. The model parameters can be expressed as a vector $\mathbf{m}$

$$\mathbf{m} = [k, \mathbf{z}, \mathbf{v}, \boldsymbol{\kappa}, \mathbf{w}], \tag{3.14}$$

where k is number interfaces, $\mathbf{z}$ are interface positions (depth), $\mathbf{v}$ are shear-wave velocities, $\boldsymbol{\kappa}$ are $V_{\mathrm{p}}/V_{\mathrm{s}}$ ratios, and $\mathbf{w}$ are noise parameters. Noise parameters are discussed further in Section 3.2.4 and Section 3.2.5, standard deviations for RF and SWD noise levels and the empirically-calculated covariance matrix for RF are incorporated into the noise model parameter, $\mathbf{w}$ (Dettmer et al., 2007).

Depth (km)	V_s (km/s)	V_p/V_s	ρ (g/cm^2
0	3.20	1.8000	2.6
15	4.00	1.8000	2.6
115	4.200	1.8125	2.6
250	4.55	1.8125	2.6

Table 3.1: Background model used for inversion of observed and simulated data.

While RF and SWD data are predominantly sensitive to changes in V_{s}, forward modelling generally requires specification of V_{s}, V_{p} and density for isotropic layers. Many studies assume fixed values for V_{s} and density (e.g., Bodin et al., 2012). However, fixing these quantities can lead to erroneously overestimating layering in V_{s} (Dettmer et al., 2015). Here, I include the $V_{\mathrm{p}}/V_{\mathrm{s}}$ ratio as unknown and constrain density by Birch's law (Birch, 1960). This approach avoids subjectivity due to choosing a specific $V_{\mathrm{p}}/V_{\mathrm{s}}$ or density value and reduces over-parameterization while maintaining physically meaningful parameter combinations (Dettmer et al., 2015).

Both $\mathbf{v}$ and $\boldsymbol{\kappa}$ are treated as perturbations around a reference model (Table 3.1). This is significant, because it allows prior bounds to change with depth as long as these remain at equal width as a function of depth, which is required for the MHG sampling applied. Such constraints help in excluding improbable models from inversion results. The prior for V_{s} perturbations is 3 km/s wide and that for $\boldsymbol{\kappa}$ is 1.0 km wide. Bounds on uniform priors are specified in Table 3.2.

For the complexity model parameter, the number of layers, k, the bounded Poisson distribution

prior is applied

$$p(k) = \exp^{-\lambda} \frac{\lambda^k}{k!}, \tag{3.15}$$

where λ is a scaling parameter (Green, 1995).

Parameter	Lower Bound	Upper Bound
Layers	2	30
Depth	0km	250km
Layer Thickness	1.000	–
dV_s	-1.500	1.500
dV_p/V_s	-0.050	0.050
RF residual error std.	0.015	0.200
SWD residual error std.	0.010	0.200
RF C_e Scaling	1.000	2.000
Poisson prior on k	3.0	-

Table 3.2: Bounds on parameter prior distributions.

Synthetic seismograms are computed using ray-theoretical modeling of teleseismic waves (RAYSUM) (Frederiksen and Bostock, 2000). The predicted RFs are obtained from the synthetic seismograms by applying water-level deconvolution with a Gaussian filter of 1 and water-level of 0.0001. Normal-mode solutions are applied to predict SWD data for Rayleigh waves (Saito, 1988).

3.2.3 Parallel Tempering

Convergence on the target distribution (the PPD) is governed by how efficiently the algorithm samples the high-dimensional regions of the model space. Parallel tempering (PT) accomplishes this by employing multiple Markov chains in parallel at increasingly tempered (*relaxed*) distributions and allowing chains to exchange information (parameter vectors). This process allows the algorithm to converge in a reasonable time (Geyer, 1991; Dettmer and Dosso, 2012). Rooted in the analogy of simulated annealing, PT tempers the likelihood function using various *temperatures*. As likelihood values are relaxed (Figure 3.5), chains sampling at higher temperatures are allowed to make larger jumps between models. Low-temperature chains explore more probable regions of the parameter space, high-temperature chains explore the parameter space more freely, sampling

outside of local high-likelihood nodes. The algorithm accounts for this by updating each chains' acceptance probability by a tempering parameter

$$\beta = \frac{1}{k_b \tau}, \tag{3.16}$$

where β is the tempering parameter and k_b is Boltzmann's constant, which is absorbed by τ, the temperature of the system (Earl and Deem, 2005). The tempering parameter is set between 0 and 1, with at least one chain samples with a β value of 1, which has a distribution identical to the PPD. Information is exchanged between chains with swapping moves that apply the acceptance criterion for chain pairs

$$\alpha_{PT} = \min\left[1, \frac{L(\mathbf{x}_1)}{L(\mathbf{x}_n)}^{\beta_n - \beta_l}\right], \tag{3.17}$$

where $\mathbf{x}_l$ and $\mathbf{x}_n$ are the randomly selected chains l and n, respectively. Therefore, if $\beta \neq 1$, the likelihood is raised to this factor, reducing the amplitudes of highs and lows. Exchange moves apply to trans-dimensional models and allow parameter vectors of different dimensions to be exchanged (Dettmer and Dosso, 2012). This process requires minimal computational effort and communication between chains, so swapping moves are carried out at the end of each model proposal iteration. Note that exchange moves are not limited to neighbouring chains with respect to τ. Rather, performance improves when random pairs may exchange information (Sambridge, 2014)

3.2.4 Inferences from High-dimensional PPDs

Interpretation of the high-dimensional PPD is carried out with marginalization (e.g., Dettmer et al., 2007)

$$I = \int f(\mathbf{m}) p(\mathbf{m}|\mathbf{d}) d\mathbf{m} \approx \frac{1}{\tilde{Q}} \sum_{i=1}^{\tilde{Q}} f(\mathbf{m}_i), \tag{3.18}$$

where $\tilde{Q}$ is the number of models in the ensemble and f is a specific function applied in PPD analysis. These integrals leave well-defined interfaces overlapping while smoothing out poorly-defined interfaces.

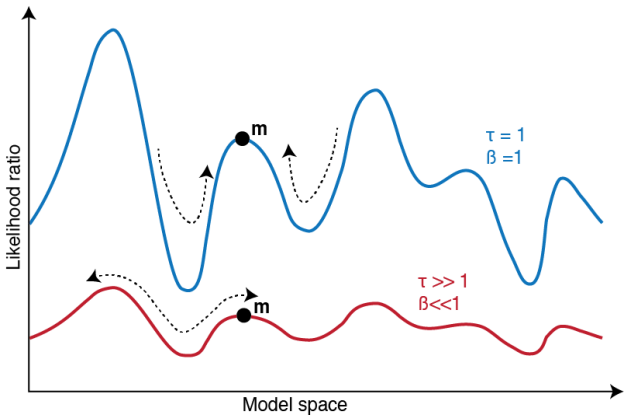

Figure 3.5: Schematic representation of parallel tempering. Here, the blue chain is sampling at a temperature of 1 in blue, and a higher, unspecified temperature chain is sampling in red. The proposed model, **m**, represented by black circles, is trapped in a local maximum in the PPD, so this chain will keep exploring the space around the local maximum. The warmer temperature chain can escape the trap and explore the space more freely, and will be able to find the global maximum. Swapping the chains will allow for the lower-temperature chain to carefully explore the global maximum once the higher-temperature chain discovers it.

Interpretation of the PPD for a single solution is non-trivial due to the multidimensional nature of the PPD. Quantification of parameter estimates and uncertainties is done via computing the maximum *a posteriori* (MAP) model

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max} P(\mathbf{m}|\mathbf{d}), \tag{3.19}$$

mean model

$$\bar{\mathbf{m}} = \int \mathbf{m}' P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \tag{3.20}$$

model covariance matrix

$$\mathbf{C}^{(\mathbf{d})} = \int (\mathbf{m}' - \bar{\mathbf{m}})(\mathbf{m}' - \bar{\mathbf{m}})^T P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \tag{3.21}$$

and marginal probability distributions

$$P(m_i|\mathbf{d}) = \int \delta(m_i' - m_i) P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}', \tag{3.22}$$

where δ denotes the Kronecker-delta function.

With a unimodal, symmetric PPD, the mean and MAP models are identical, but with more complicated models, there may be no good and simple estimate. No further statistical tests are required, as MHG sampling produces solutions with an appropriate level of complexity by adjusting the number of layers of the model (Agostinetti and Malinverno, 2010).

3.2.5 Data Covariance Estimation

Data residuals are interpreted as an approximation of data errors, being some combination of measurement and theory error. While measurement error is relatively easy to determine, theory error is inherently troublesome, as the true magnitude is unknown. Theory error strongly affects the PPD estimate, as uncorrelated noise does occur and neglecting this error can lead to severe uncertainty underestimation (Dettmer et al., 2007). To avoid dependence of results on an initial choice, $\mathbf{C_d}$ is estimated from residual errors based on the optimization result.

No prior information is available for data-error statistics, so the standard deviation, σ_i, becomes a parameter to be included in the inversion. As mentioned earlier, the most commonly applied assumption is that errors are uncorrelated and adhere to a Gaussian distribution. Normal distributions are used for the prior. Likelihood will not be fixed, but have a broad uncertainty. Given $\mathbf{C}^{(d)} = \sigma^2\mathbf{I}$, where σ is considered a parameter in the inversion, the misfit is described as

$$\log(L(\mathbf{m},\sigma)) \propto -\frac{1}{2\sigma^2}\sum_i (d_i - d_i(\mathbf{m})^2) + N\log\sigma. \tag{3.23}$$

Non-stationary covariance matrices account for variation in noise level throughout the data (Dettmer et al., 2007; Dettmer and Dosso, 2012). A non-Toeplitz matrix is required for this scaling. The standard deviation is calculating by scaling by a running-RMS.

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{k=j-N/2}^{j+N/2} r_k^2}. \tag{3.24}$$

Residuals are scaled with these standard deviations, so that

$$n_i = \frac{\sigma_i}{r_j}, \tag{3.25}$$

where n_i are used to replace the residuals in the non-Toeplitz matrix of

$$C_{jl}^{\mathbf{d}} = C_{jl}\sigma_j\sigma_l, \tag{3.26}$$

which allows non-stationary effects.

The inversion process is then repeated, where the inversion is allowed to scale the computed non-Toeplitz covariance matrix

$$\mathbf{C}_{\mathbf{d}_i} = \exp^{2\mathbf{w}_i}\mathbf{C}_{\mathbf{e}_i}, \tag{3.27}$$

where $\mathbf{C}_{\mathbf{e}_i}$ are empirical matrices are scaled by factors $\mathbf{w}$, discussed in Section 3.2.2. This parameter scales non-stationary covariance matrices by from 1 to 2 (Table 3.2) The covariance matrix is incorporated into the likelihood function

$$L(\mathbf{m}) \propto \prod_{i=1}^{\tilde{S}} \exp\big(-N_i w_i - \frac{1}{2\exp^{2w_i}}\mathbf{r}^T\mathbf{C}_{\mathbf{e}_i}^{-1}\mathbf{r}_i\big). \tag{3.28}$$

Parallel tempering is employed to increase efficiency, and because all distributions are considered to be uniform, the acceptance criterion simplifies to the tempered likelihood ratio.

4. Field Data

4.1 Station and Event Selection

A map with all of the seismic stations used in this thesis is shown in Figure 4.1. Station selection is based on the length of the available records and overall data quality based on visual inspection of signal-to-noise ratios. I select nine stations on the border of Yukon and British Columbia, spanning the Intermontane terranes and Ancestral North America to the CDF. Station locations and recording start dates are listed in Table 4.1. The multiple networks that contribute data to this investigation are CN, Canadian Network (Geological Survey Of Canada, 1989); NY, Yukon-Northwest Seismic Network (University of Ottawa, 2013); TA, Transportable Array (IRIS Transportable Array, 2003); and YO, Yukon Observatory (Yukon Geological Survey).

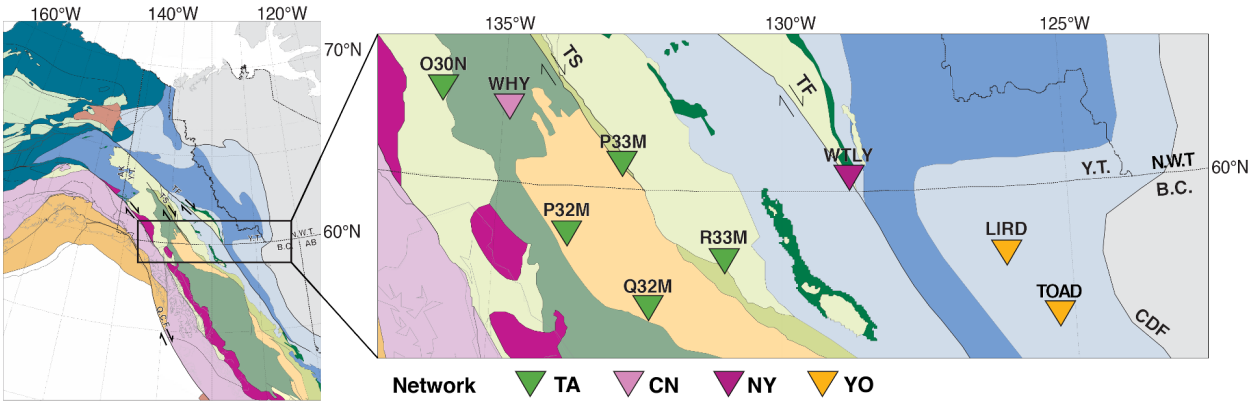

Figure 4.1: Terrane map with locations of seismic stations discussed in this thesis. Abbreviations for networks are TA, Transportable Array (IRIS Transportable Array, 2003); CN, Canadian Network (Geological Survey Of Canada, 1989); NY, Yukon-Northwest Seismic Network (University of Ottawa, 2013); YO, Yukon Observatory (Yukon Geological Survey).

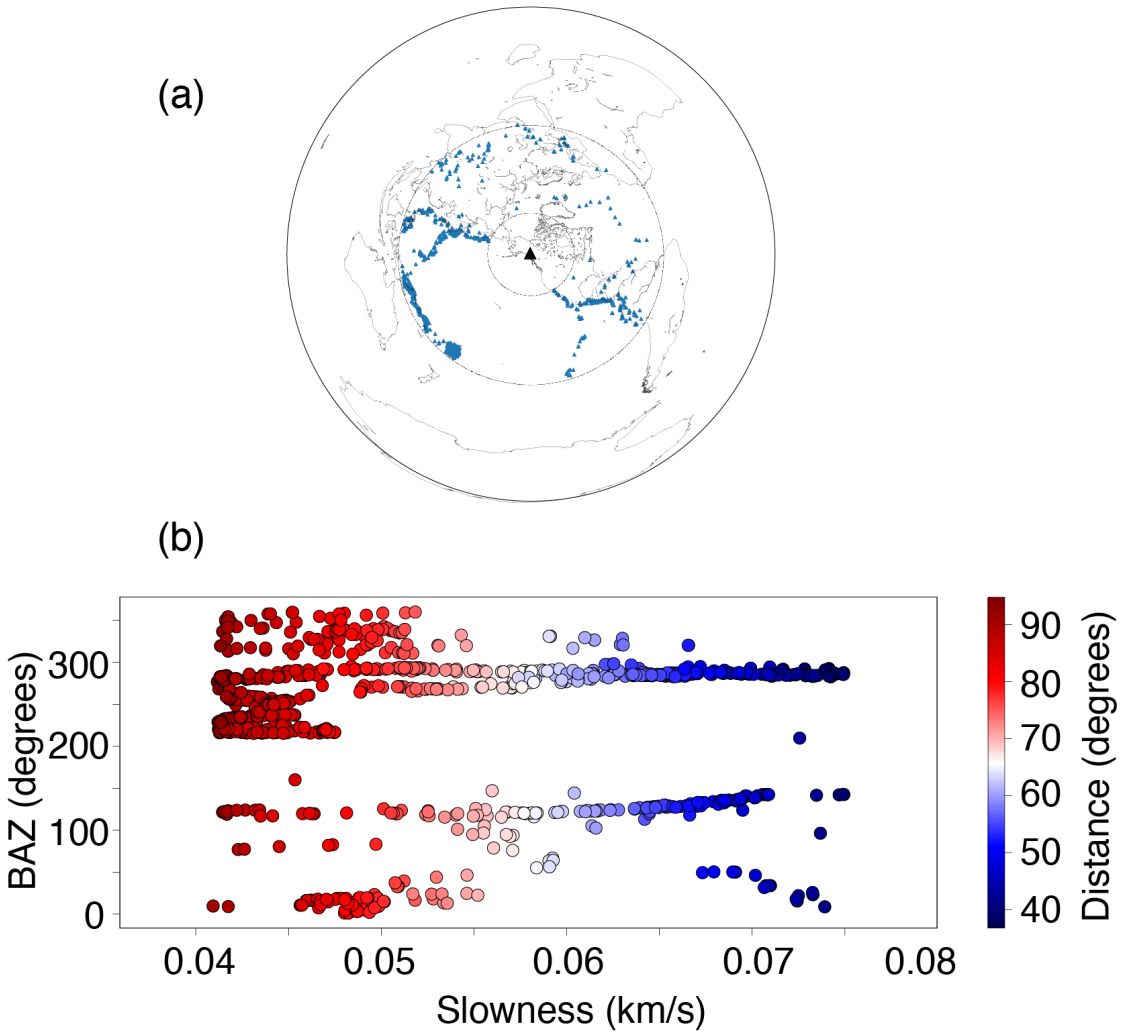

Figure 4.2: Earthquake sampling for station WHY. (a) Map view of the distribution of earthquakes for WHY. (b) Distance, slowness, and BAZ distribution of earthquakes for WHY. Note the abundance of RFs between 280–300°BAZ across all slowness values.

4.2 Data Preparation and RF Estimation

Three-component seismograms are employed for RF processing, and events with magnitudes between 5.5–7.5 and with epicentral distances between 30–90° are selected. A total of 12,152 records from teleseismic events have been selected and processed. Of the total processed events, over half (6,128 events) were processed for the longest-operating station in the region, WHY. In section 4.2(a), locations of events for station WHY are shown. Most events are arriving with a BAZ of 280–300° and originate in subduction zones off the coast of Japan.

RFs are calculated for all available BAZ and slowness values. Seismograms are filtered between 0.05–1.5 Hz, cut to 70-80 s after the theoretical P-wave arrival computed via AK135 (Kennett and Engdahl, 1991) and tapered with a Hanning window. I use *P*-to-*S* wave conversions to calculate RFs using iterative-time domain deconvolution (Ligorria and Ammon, 1999) with a Gaussian filter value of 1.

Processed receiver functions are stacked in BAZ and slowness bins to reduce noise. Sufficiently small BAZ bins are required to satisfy the isotropic assumption in the inversion, so I select RFs in a BAZ of 20°. RF variations as a function of BAZ can be used to study lithosphere anisotropy and dipping layers that produce predictable patterns (e.g., Cassidy, 1992; Frederiksen and Bostock, 2000). RFs vary less as a function of slowness observations compared to BAZ. Because I want to ensure I retain enough RFs to include in each stack, I stack together all RFs with slownesses up to 0.075 s/km for the BAZ bin.

In addition to BAZ and slowness windowing, I discard RFs where the maximum peak is more than ± 1 s from the predicted P-arrival. All events for which the radial or vertical component of the RF has a maximum amplitude greater than one are discarded. These quality controls are applied consistently to RFs from all stations, regardless of deployment duration or signal quality. Finally, RFs with an iterative time-domain deconvolution misfit of $> 70\%$ are stacked to enhance signal-to-noise ratio. For the inversion, the stacks are normalized by a factor of 0.57 (Ammon, 1991), cut from -2 to 26 s after the predicted P-arrival and tapered with a cosine window.

4.3 Surface-Wave Dispersion Data

The GDM52 global surface-wave model (Ekström, 2011) uses 258 stations to determine SWD from $3,330$ earthquakes. These curves are sampled at periods between 25–250 s to determine propagation phase velocity anomalies with respect to the PREM reference model (Dziewonski and Anderson, 1981). Most notably, these data show an increase of V_S in shallow and intermediate depths from west to east across the transect. Note the modest variability in phase velocity in

Station	Lat.	Lon.	Started Recording	Total No. RFs	No. RFs in stack
LIRD	59.40976	-126.09863	2016-06-11	371	21
O30N	60.77040	-136.0906	2016-05-12	884	39
P32M	59.58980	-133.7147	2016-06-02	632	30
P33M	60.21140	-132.8174	2015-10-15	1,091	46
Q32M	58.96010	-132.2691	2016-07-24	491	18
R33M	59.3946	-130.9673	2016-07-21	571	42
TOAD	58.84994	-125.2333	2016-06-13	481	24
WHY	60.65970	-134.8825	1998-01-01	6,128	307
WTLY	60.11326	-128.7961	2013-07-01	1,503	47

Table 4.1: Location, recording start date, total number of RFs computed, and number of RFs included in inversion stacks for stations discussed in this thesis.

section 4.3 along the transect at periods above 150 s. In contrast, variability at lower periods is more significant.

It is important to acknowledge that the GDM52 database is inferred on a global scale with a lateral resolution of $\sim$ 650 km due to the application of significant smoothing/regularization. In contrast, available regional SWD data are often noisy and limited to shorter periods than required to resolve upper-mantle structure (e.g., McLellan et al., 2018). For these reasons, the GDM52 SWD data are better fit for this study as they retain information for the lower crustal and upper mantle depths desired in this study.

However, the noise levels and high smoothing associated with a global model may not be able to accurately resolve low velocity layers present regionally (van Heijst et al., 1994). A likely explanation for the lack of LVZ effects is that the GDM52 data represent a global model that lacks resolution at the scales important to my study. Without other methods to constrain the SWD inversion, observation of absolute velocities associated with a LVZ remains a difficult inversion problem (Killingbeck et al., 2018; Murphy and Egbert, 2019). As a result, observing LVZs with the discussed MHG inversion algorithm may be challenging. Forward-modeled SWD data produced by the DISPER80 modelling code (Galetti et al., 2016) are sensitive to decreases in V_s, so that small velocity perturbations at depth have large data errors. The RFs I observe in this study show clear evidence for LVZs In SWD data, LVZ are expected to cause a reduction in phase velocity

with period. However, such reduction is not observed in the GDM52 model (Ekström, 2011).

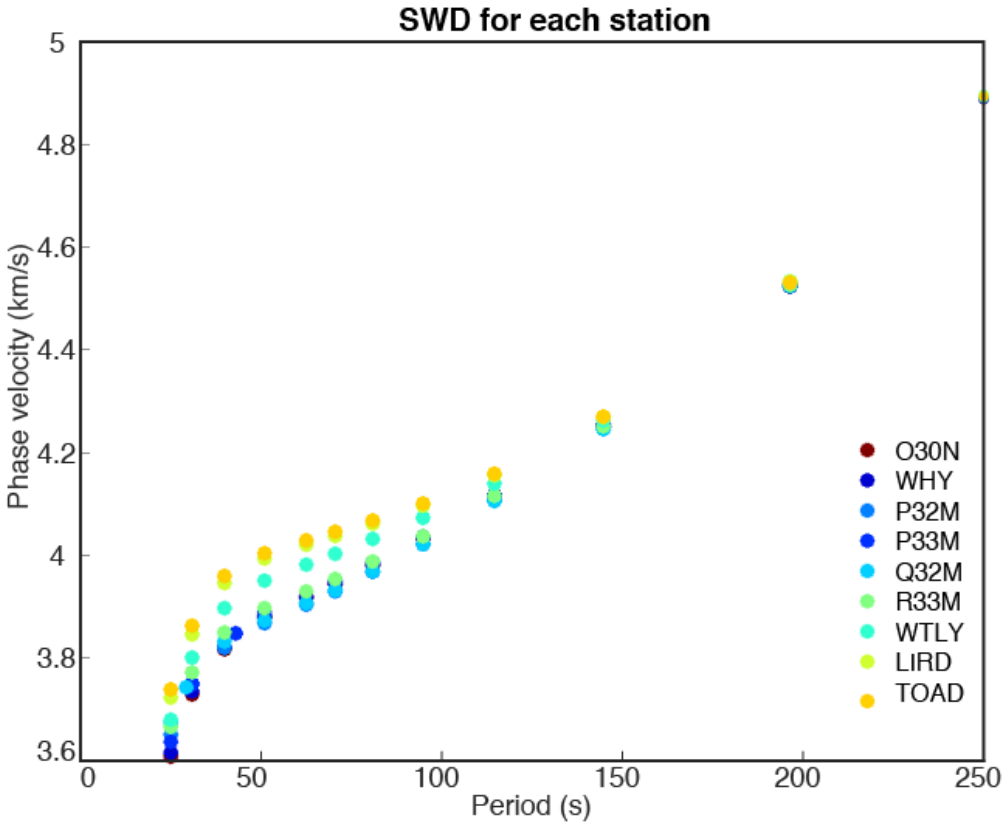

Figure 4.3: SWD as a function of period for selected stations to illustrate the increase of phase velocities from west (cooler colours) to east (warmer colours).

Given these shortcomings of inverting SWD data in the presence of an LVZ, only the RF data will be allowed to constrain V_s at intermediate depths between $\sim$ 50 and 200 km. Figure 4.3 shows sensitivity kernels for Rayleigh waves from a study across the Pacific Basin (Smith et al., 2004). In Figure 4.3, Rayleigh velocities with periods of 25 s, 50 s, 100 s, and 150 s are shown as a function of depth. Periods of 50 s and 100 s have the highest sensitivity between 50-km and 200-km depth. Sensitivity kernels for periods of 25 s and 150 s show highest sensitivity outside the depth range of where LVZ is expected. Therefore, I down-sample the data to 10 uniformly spaced data points with respect to period. Of these, 4 data points are between 20 and 30 s and 6 data points between 150 and 250 s. This resampling aims to reduce the impact of a lack of LVZ information in the SWD data while still retaining enough information on absolute velocities.

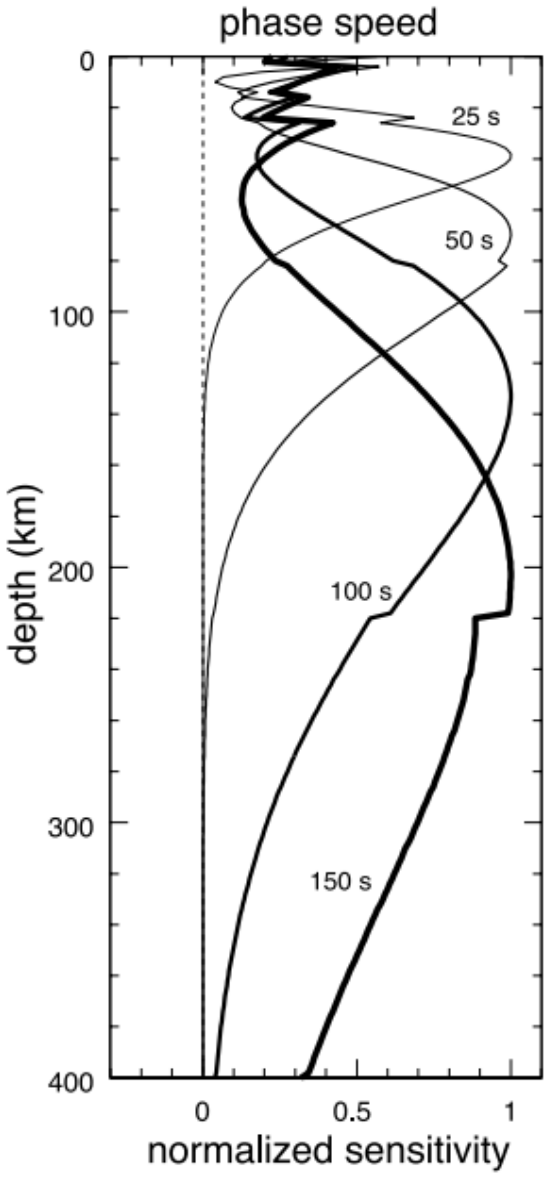

Figure 4.4: Sensitivity kernels for Rayleigh phase velocities at periods of 25 s, 50 s, 100 s, and 150 s across the Pacific Basin from Smith et al. (2004). Note that while specific sensitivity kernels vary across studies, SWD sensitivity overall increases with depth as period increases. Because I aim to reduce the impact of SWD data between 50- and 200-km depth, I retain SWD data with periods less than 30 s and greater than 150 s.

5. Results

5.1 Inversion of Simulated Data

This section applies the Bayesian inversion described in Section 3.2 to simulated P-RF data. In particular, I consider simulations based on station WHY as realistic scenarios to examine LAB resolution by P-RFs in the presence of multiples and the impact of theory error due to anisotropy, which is ignored in the inversion but likely present in the Cordillera.

5.1.1 Impact of Moho Multiples on LAB Resolution

Previous studies have documented that the LAB can be challenging to resolve due to a weak seismic velocity contrast and/or interference with multiples in P-RFs for typical LAB depths. Although the LAB is notoriously difficult to resolve in cratonic provinces (e.g., Eaton et al., 2009), the shallow, sharp LABs of recent orogens can often be resolved (e.g., Chen, 2009; Abt et al., 2010).

I adapt a test from Kind et al. (2012) examining the ability of P-RFs to resolve a LAB-like discontinuity in the presence of Moho multiples. In the original test, it is concluded that to observe a LAB with P-RFs, the Moho needs to be deep enough to separate the LAB from Moho multiples (Kind et al., 2012). Because the Moho is predicted to be at a constant depth of ~ 34 km in the NCC and the inversion algorithm applied here accounts for multiples and non-uniqueness, I carry out a more general test to see how well the LAB-like discontinuity is resolved at depths between 50 and 110 km. Importantly, the simulation study presented here applies uncertainty quantification which accounts for the full non-uniqueness in the parameter

Model	Thickness (km)	Density (kg/m^3)	V_p (km/s)	V_s (km/s)
1	34	2559	5.414	2.962
	16	3189	7.828	4.256
	Halfspace	2893	6.886	3.878
2	34	2559	5.414	2.962
	36	3189	7.828	4.256
	Halfspace	2893	6.886	3.878
3	34	2559	5.414	2.962
	56	3189	7.828	4.256
	Halfspace	2893	6.886	3.878
4	34	2559	5.414	2.962
	76	3189	7.828	4.256
	Halfspace	2893	6.886	3.878

Table 5.1: Models parameters to produce simulated data in Section 5.1.1. All models have the same RAYSUM parameters for each layer, except for the middle layers' thickness.

space. This is an important distinction from other standard inversion approaches that are limited to qualitative conclusions.

Radial P-RFs are computed for four models with two interfaces: a Moho at 34-km depth and LAB between 50- and 110-km depth. The simulated P-RFs for these four models are computed using RAYSUM (Frederiksen and Bostock, 2000) with the parameters listed in Table 5.1. P-RFs are contaminated with Gaussian noise of 0.015 and inverted with a lower prior bound of R-component noise of $\sigma_{RF} = 0.015$ to ensure the P-RF is not over-fitted. (Explanation of inversion marginal profile interpretation is outlined schematically in Appendix B.)

Figure 5.1 shows that the Bayesian inversion algorithm is capable of resolving the LAB in all four models, even in the presence of multiples from the Moho. These results clearly demonstrate the ability of the algorithm to account for Moho multiples in the inversion and recover all LAB interfaces in the models by considering the full non-uniqueness of the problem for considered noise levels. This example also shows an increase of interface depth uncertainty as LAB depth increases due to the inherent trade-off between depth and velocity in P-RF inversion.

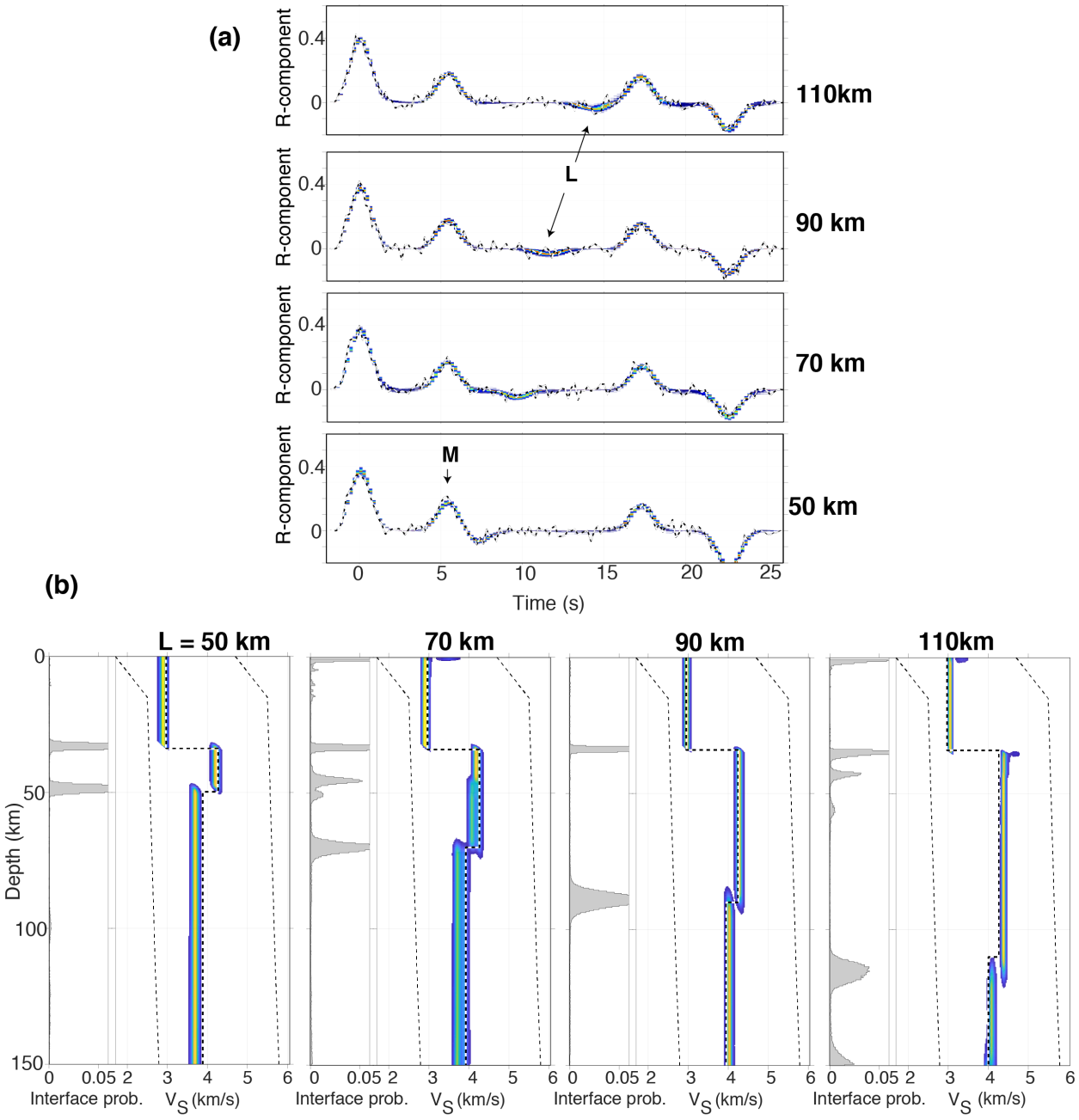

Figure 5.1: Analysis of simulated data following a test adapted after Kind et al. (2012) to evaluate the resolution of the LAB interface at various depths. (a) Noisy P-RFs for models with a Moho at 34 km and LAB at 50, 70, 90, and 110 km with the PPD from the inversion shown in colour. The arrivals from the Moho and LAB discontinuities are fit within the range of P-RF data prediction from the inversion. (b) The 1D inversion results for the four models. The LAB discontinuity is well resolved with in terms of interface probability (left panels). V_s The PPD marginal profiles are shown in the right panels.

Layer	Thickness (km)	Density (kg/m^3)	V_p (km/s)	V_s (km/s)	Anisotropy (%)	Trend	Plunge
1	8	2700	5.50	3.10	-	-	-
2	27	2800	6.30	3.54	-	-	-
3	10	3200	8.00	4.40	12	216°	5°
4	Halfspace	3200	7.80	4.30	-	-	-

Table 5.2: The WHY model from Taryoun et al. (2017) that I employ for simulated data to observe the effects of modest anisotropy.

5.1.2 Impact of Modest Anisotropy on Inversion Results

Due to the computational cost and data requirements for good azimuthal coverage to resolve anisotropic parameters from P-RFs, I consider isotropic parametrizations in the inversions of field data. Resolving anisotropy with trans-dimensional models is still a research frontier (e.g., Bodin et al., 2016; Agostinetti and Bodin, 2018). This section examines the effect of ignoring anisotropy on the inversion results. I apply RAYSUM (Frederiksen and Bostock, 2000) to compute radial P-RFs for an anisotropic earth model. These data are then contaminated with noise and inverted under an isotropic assumption.

The P-RFs from a small BAZ window are stacked, because inverting for a single direction will reduce the possibility of anisotropy or dipping layers complicating the inversion results. I recreate a model by Taryoun et al. (2017) with modest anisotropy for station WHY to determine how the algorithm behaves under the isotropy assumption. This four-layer model is described by RAYSUM (Frederiksen and Bostock, 2000) parameters in Table 5.2. Anisotropy of 12% trending 216° and plunging 5° is introduced in the layer below the Moho and above the half-space.

I obtain a realistic event distribution by randomly sampling 200 events with various BAZ, distance and slowness values at station WHY (Figure 5.2(a)). The P-RFs are computed for this distribution of BAZ, distance and slownesses, and 90 P-RFs with suitable quality between 280–300° BAZ are stacked. Zero-mean, Gaussian-distributed noise with a standard deviation of 0.015 is added to the P-RF stack.

The P-RF stack is inverted by applying uniform priors with bounds of $\pm$1.5 km/s around the

reference model for V_s, ± 0.05 around a ratio of 1.8 for V_p/V_s, and between 0 and 250 km for layer boundaries. For the first phase of the inversion, I allow P-RF noise standard deviations between 0.015 and 0.2 during the hierarchical inversion and the empirically-calculated P-RF matrix to scale from 1 to 2.

Here, 20 processors are used, with each processor simulating a single Markov chain. The β values for parallel tempering are distributed geometrically between 1 and ~ 2000. The algorithm is run for $140,000+$ steps. The first $40,000$ steps from the burn-in period are removed, and the remainder is sub-sampled by a factor of two to retain $55,000$ samples. Residuals are calculated from the MAP model for the remaining samples, which is taken to be the maximum likelihood model with the most probable number of layers (here 5) in the PPD. A covariance matrix is calculated for these residuals (Dettmer et al., 2007; Dettmer et al., 2014).

Profile marginals are shown in Figure 5.2(b). The shallow-most layer is inferred to have a higher V_s than the true model, and the interface probability of the lower bound of the layer is relatively consistent with the true model. There is an extraneous shallow layer of 4-km depth at the top of the model space. An intermediate layer between $\sim$ 4–8-km depth within the shallow-most layer more closely resembles the true model. While the velocities for the anisotropic layer and the half-space are over-estimated, the interfaces are relatively well resolved. The anisotropic layer is resolved as a broader interface probability than the other interfaces.

The range of data predictions produced by the PPD samples for the P-RF is displayed in Figure 5.2(c). After the Moho conversion, there is also relatively higher uncertainty. This region of the P-RF between $\sim$ 5–7 s corresponds to the anisotropic layer, where the algorithm fits an extra layer.

Marginal profiles and the range of predicted P-RF amplitudes are nevertheless well constrained, ensuring significant features of the synthetic waveform are reproduced. The marginal distribution for the covariance matrix scaling is shown in Figure 5.2(d). Note the covariance matrix scaling was constrained to be > 1 by the prior to avoid data over-fitting.

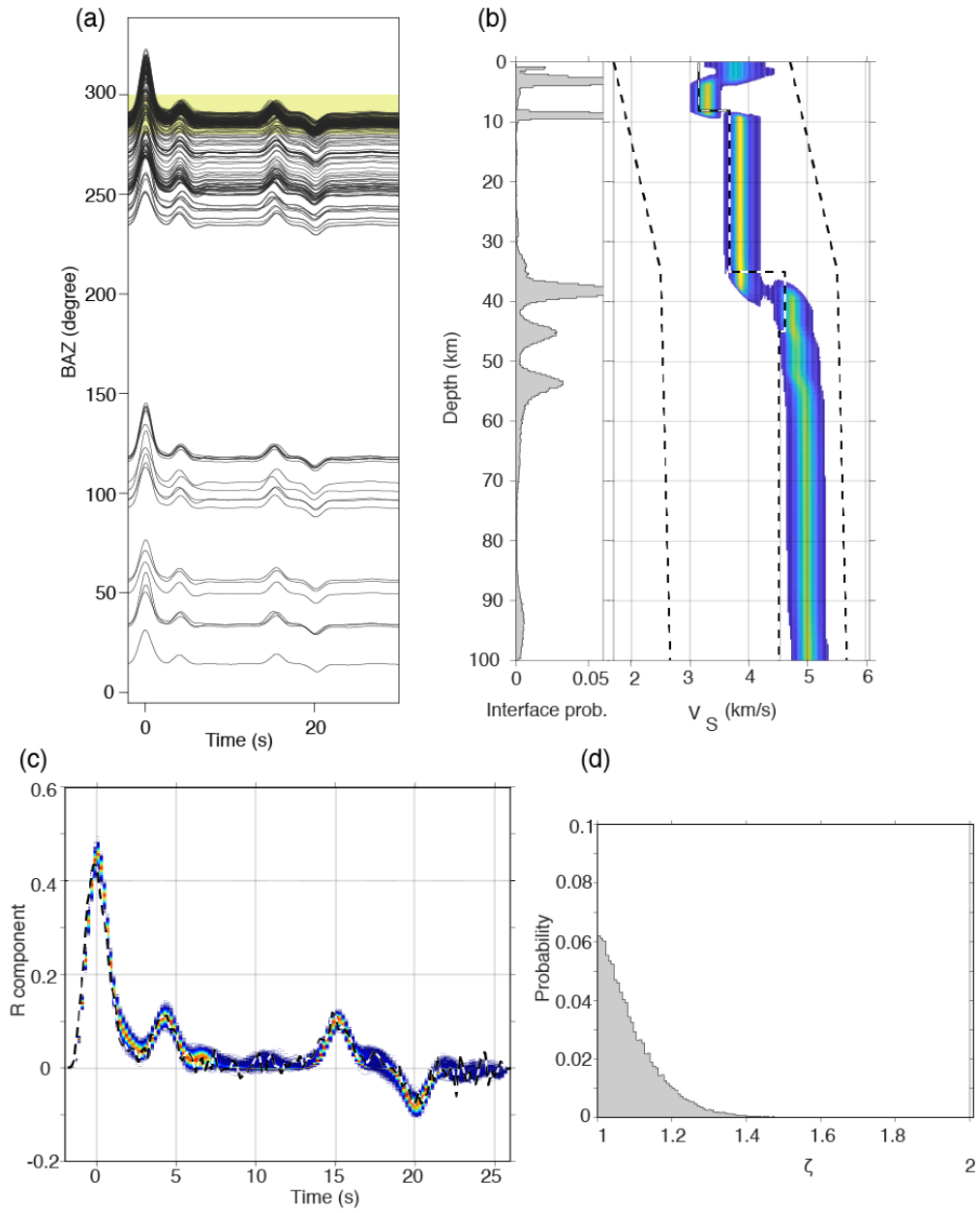

Figure 5.2: (a) RAYSUM output of the model parameters (Table 5.2) with BAZ distribution from WHY employing a random sub-sampling of 200 events. The desired BAZ window is highlighted in yellow between 280–300°. (b) Inversion output for the stacked P-RFs from panel (a). True model parameters shown as a dashed line. Interface probability for model output is shown to the left, and V_s is on the right. Warmer colours are associated with higher probability. The width of the coloured portion reflects the uncertainty of the model. (c) Range of data predictions from the PPD. Observed data are in a dashed black and white line. (d) The scaling factor for the calculated covariance matrix in the inversion.

5.2 P-RFs for Individual Stations

This section discusses observations obtained directly from P-RF estimations. Because station WHY has the longest time record and the most observations, I focus on this station within the following section and present data for remaining stations in Appendix C.

As discussed in Section 3.1.2, P-RFs may vary extensively with respect to BAZ due to dipping layers or anisotropy. I examine the effect of these azimuthal variations on the quality of the stack. From the 6,128 P-RFs computed for station WHY, I select and stack 307 P-RFs associated with a narrow BAZ window of 280–300°and slownesses between 0–0.075 s/km. Figure 5.2 shows that selecting a subset from a small BAZ window still allows for consistent phases to be incorporated into the stack while reducing directional effects.

As described in section 5.2 for station WHY and Figures C.1-C.8 for remaining stations, P-RFs are estimated for all BAZ and stacked for BAZ between 280–300°. The P-RF stacks and V_s profiles are ordered from west-to-east to visualize the variability of lithosphere-scale features across the transect. Stations Q32M and P33M have very similar longitudes, so I switch the order of these two stations, because Q32M is east of the Teslin Fault, whereas P33M is west of the Teslin Fault.

The P-RF BAZ dependence varies by station, with some stations exhibiting relatively minor directional dependence, such as station P32M. Other stations, such as WTLY and TOAD, exhibit strong variations in P-RFs arriving from different directions. Early arrivals exhibit arguably the most directional dependence, with amplitudes and phases varying strongly between the direct P-arrival at 0 s and the Moho arrival at ~ 4 s. Arrivals after the Moho are relatively consistent for stations west of WTLY compared to the stations east of WTLY. Station TOAD appears to have the least consistent P-RF arrivals across BAZ, even within the relatively well-sampled BAZ range of 200–300°. Given the extensive BAZ variations for all stations, stacking by the selected 280–300° corridor proves important to ensure major conversion and multiple phases in each P-RF are incorporated into the stack and not effected by BAZ variations.

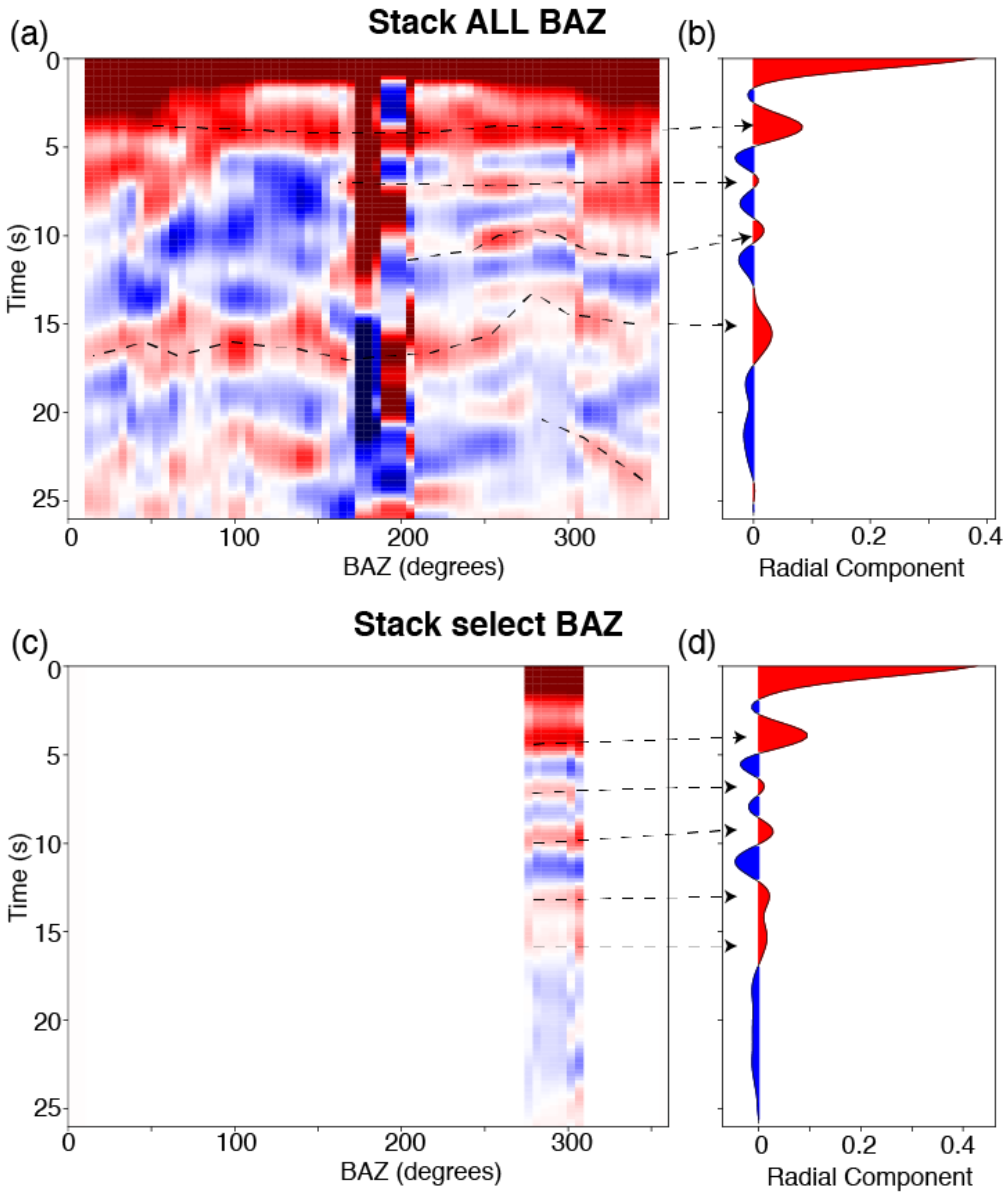

Figure 5.3: (a) BAZ variation plots for all P-RFs computed for station WHY and (b) stacked. (c) BAZ variation and (d) stacked P-RF for window between 280–300°BAZ and 0.0–0.55 s/km slowness.

Stacked P-RFs across the NCC are shown in Figure 5.2. A positive polarity pulse at $\sim$ 4 s is consistently observed across all stations, presumably from the direct Ps Moho arrival. This signal is followed by a negative polarity pulse of varying amplitudes at nearly every station across the Cordillera. Stations WHY and TOAD exhibit smaller P-RF amplitudes after the 4-s arrival compared to neighbouring stations. Conversions are less coherent across stations after $\sim$ 10 s and are more challenging to interpret due to interference with Moho reflection multiples. This emphasizes the importance of a rigorous approach, such as an Bayesian inversion, for the interpretation of P-RF data in order to distinguish conversions from multiples in later P-RF arrivals.

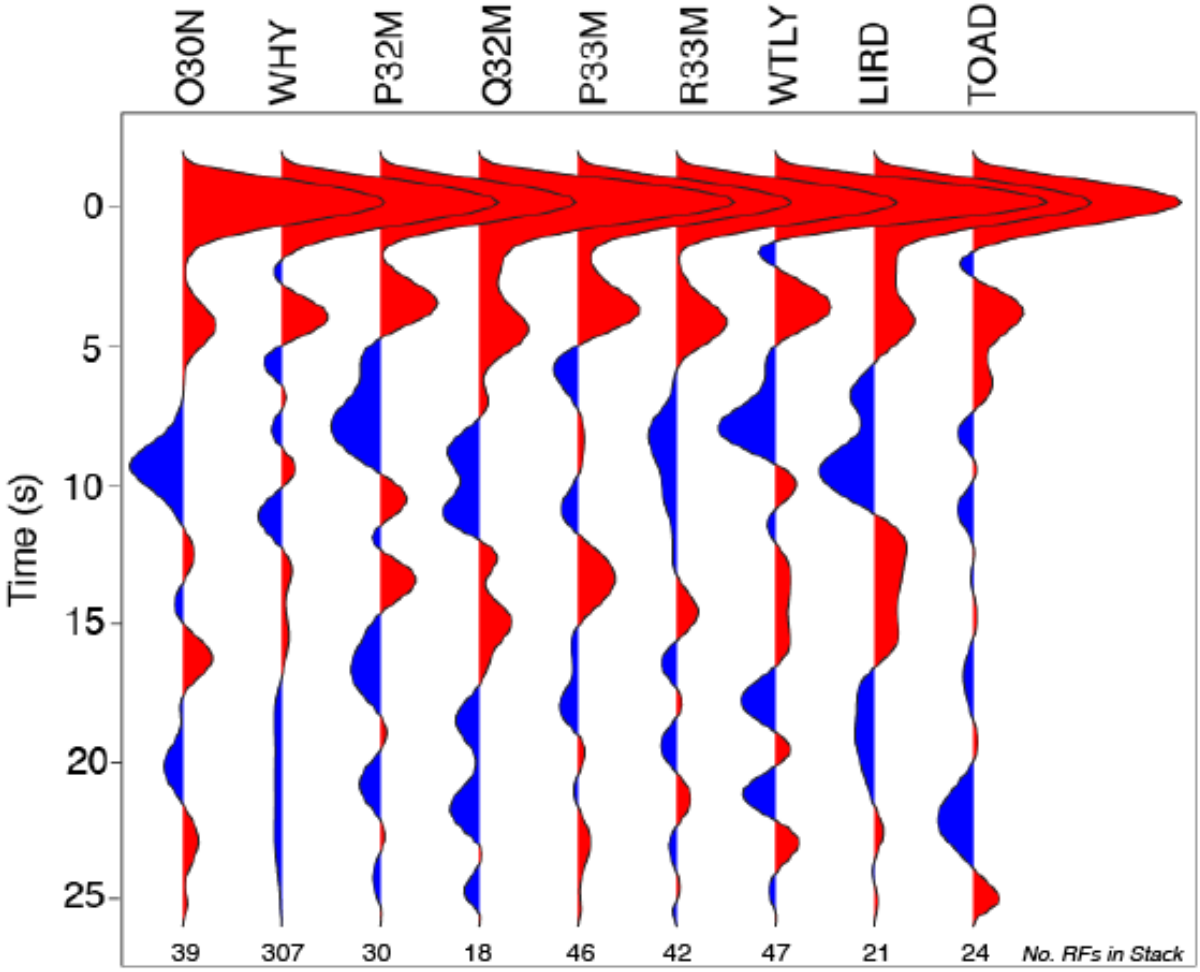

Figure 5.4: Estimations of P-RFs considered in this thesis across the NCC ordered west-to-east. The number of P-RFs in each stack are shown below under the stacked P-RFs and listed in Table 4.1. The locations of stations are listed in Table 4.1 and shown in Figure 4.1.

5.3 1D Inversion Results

Using the trans-dimensional inversion algorithm, I obtain 1D marginal V_s profiles for all nine stations across the NCC. The profiles display several prominent crustal and upper mantle features.

As discussed in subsection 3.1.5, SWD data constrains absolute V_s while RF data inversions constrains velocity discontinuities, with poorer constraints on absolute V_s. By combining the information from both datasets in the joint inversion, the trade-off between absolute velocities and layer structure is balanced to produce a reasonable fit to both P-RF and SWD data. The results for SWD, P-RF, and joint inversion of both datasets for station WHY are shown in Figure 5.5. For the SWD inversion in Figure 5.5(a), the absolute velocity is inferred for the depth profile, but the interface probability across depths is low. Interfaces are better resolved in Figure 5.5(b) for the P-RF inversion, and (c), the joint inversion. While the difference between the P-RF and joint V_s profiles are more subtle than for the SWD inversion, the uncertainty of V_s for the joint inversion is reduced and the interface probabilities are sharper. These improvements in uncertainty in V_s and interface probability for the remaining stations in the joint inversion are show in Appendix D, Figures D.1 to D.8.

Joint inversion results for the V_s marginal profile, data fits, and parameters are shown for station WHY in Figure 5.6. Note the resolution of each P-RF feature in the marginal and PPD in panels (a) and (b). There is relative uncertainty in the depth of the LAB layer, as portrayed by the PPD of the forward-modelled P-RF in Figure 5.6(b) around 80-km depth or 8 s. While SWD data, shown in Figure 5.6(c), are not resolved for intermediate depths, the shortest and longest period features of the SWD data are most accurately modelled, ensuring the incorporation of these velocities into each the final model. Distribution of the number of layer nodes is described in Figure 5.6(d), where the median number of nodes is 9. The PPDs for the scaling factor and the SWD standard deviation are shown in Figure 5.6(e). In Figure 5.6(e), it is important to note that the scaling factor strongly favours the lower prior bound. The SWD data, on the other hand, favour a low standard deviation but the peak probability is not at the lower prior bound. This suggests the algorithm is

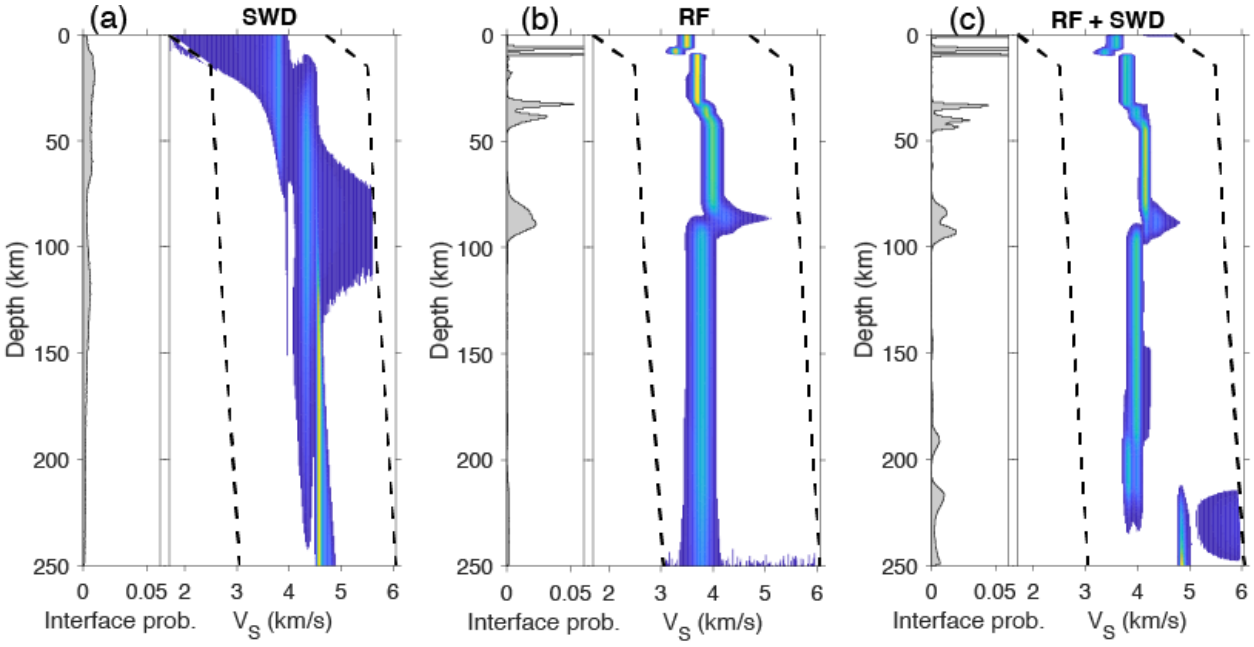

Figure 5.5: Inversion of (a) only SWD data, (b) only P-RF data, and (c) joint inversion of P-RF and SWD data for station WHY. Note how the SWD inversion resolves absolute velocities, compared the P-RF inversion, which infers interfaces. Joint inversions of both datasets constrain both parameters relating to interfaces and velocity well compared to either dataset alone.

preferentially fitting P-RF features and allowing higher noise parameters to SWD data.

Marginal 1D profiles, data fits and supporting statistics for all other stations included in Appendix D in Figures D.9 to D.15. High variability of the number of layer nodes in the PPD between 6–15 across stations confirms that *ad-hoc* choices on the number of layers nodes would be undesirable.

Major P-RF features for all stations are reasonably well fit by the range of data predictions. For stations where there is a minimal decrease in velocity at intermediate depths in the marginal profile (e.g., TOAD, Figure D.16), observed SWD curves are well fit compared to stations with large V_s decreases (e.g., R33M, Figure D.13). While stations with major V_s reduction exhibit poorer fit of short periods of SWD data, all stations are able to fit longer period features reasonably well. By fitting both small and large SWD periods, the algorithm is able to constrain V_s to some degree for the minimum and maximum bounds of the model space. Most stations exhibit reasonable uncertainty, indicated by the width of the profile marginal, except for station P32M, which has a

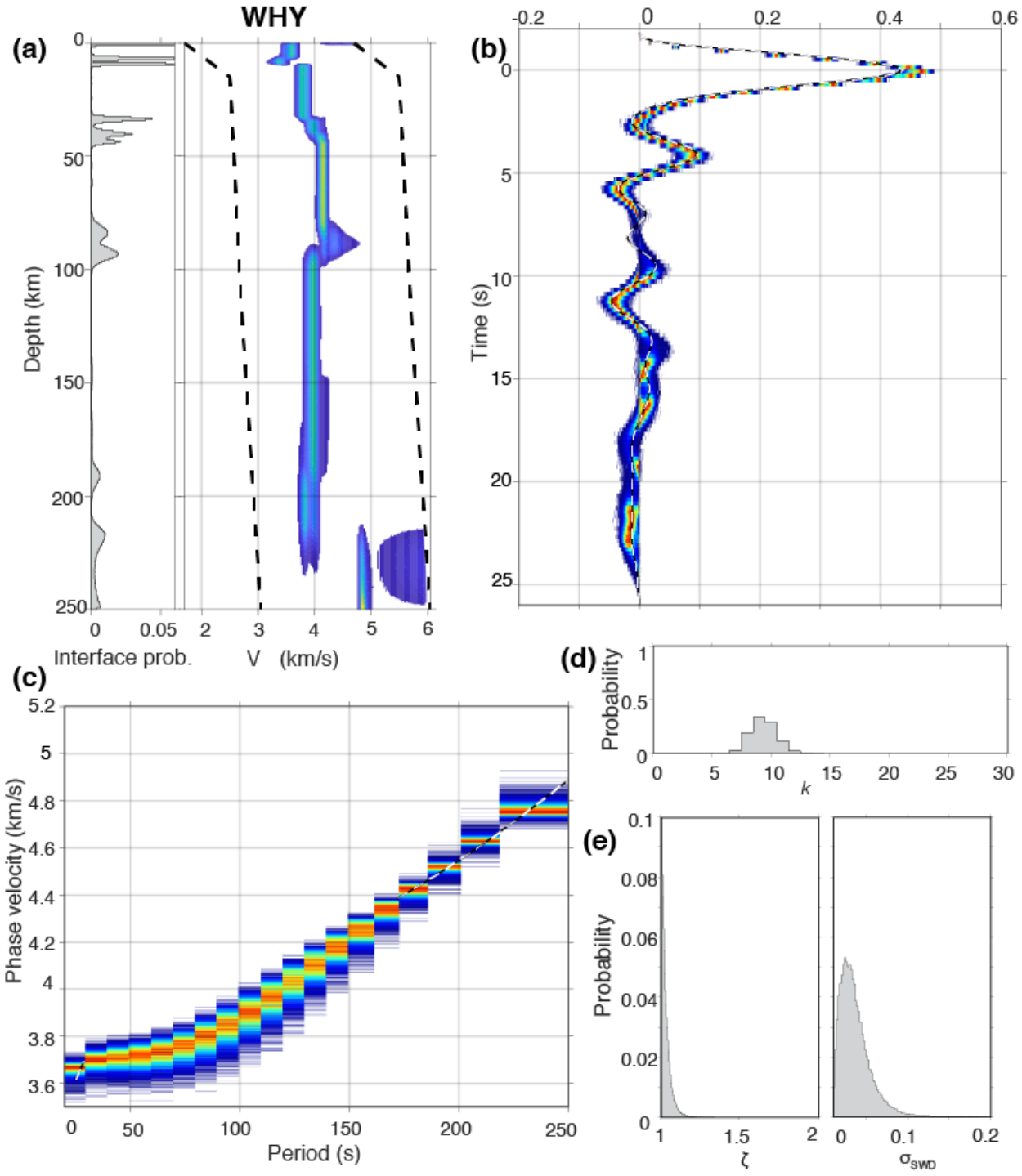

Figure 5.6: (a) Profile marginal plot for WHY; (b) Predicted and observed P-RF data; (c) Predicted and observed SWD data; (d) Marginal of number of layer nodes; (e) Marginals for the scaling factor for R-component and standard deviation for SWD data.

notably thin width along the profile marginal. Such little uncertainty may result from incompatible SWD and RF data for the station or a RF inversion with multiple reasonable solutions (Figure D.2). Both SWD and RF data are reasonably fit as hyperparameters, so it is possible that one dataset is driving the inversion. Overall, note that uncertainties estimated do not require results to be consistent between stations, meaning that similarities arise from the data of each individual station.

Crustal velocities across the NCC generally fall within a range of 3–4 km/s, where most stations infer layer of $\sim$ 3 km/s in the upper crust and a layer of $\sim$ 4 km/s directly above the Moho. Stations WTLY and Q32M both estimate higher velocities than the other stations, where station WTLY has a shallow layer of $\sim$ 4.3 km/s and station Q32M has a lower crustal velocity layer of $\sim$ 4.2km/s. The surficial geology at station WTLY cannot explain the high velocities, as $\sim$ 4.3 km/s is exceptionally fast for V_s at shallow depths and the station sits on a sedimentary rock type. Both stations WTLY and Q32M have relatively poor data range prediction fit to observed SWD data at low periods, so shallow V_s may likely be over-estimated.

Values of V_s for the mantle are observed to vary between $\sim$ 3.0 and $\sim$ 4.7 km/s above 200-km depth. At depths of 250 km, which is the lower bound of the profile, marginal profiles are generally wider with V_s around $\sim$ 5 km/s. Nearly all stations decrease in V_s above 125-km depth, except for station TOAD, which exhibits a V_s decrease at $\sim$ 200-km depth.

6. Discussion

This chapter compares my results to previous seismic, petrologic, and heat flow studies. My probabilistic 1D profiles are represented as a 2D cross-section across the NCC to observe variability in V_s structure Figure 6.1. I interpret the locations for the Moho and LAB and observe a pervasive LVZ spanning the NCC. My results present strong evidence of the existence of pervasive partial melt in the LVZ throughout the NCC. Partial melt fractions have an upper limit of 3.5% but are more likely $< 2\%$. Finally, I discuss the implications of the LVZ and heterogeneous LAB on existing tectonic models, and conclude that the heterogeneity of the lithosphere is a crucial feature to be included in future tectonic models.

6.1 Observations of the Moho, LAB and a LVZ

The inversion results for the west-to-east profile (Figure 6.1) shows the Moho discontinuity depth estimates across the NCC at $\sim$ 31-km depth. Station TOAD exhibits a unique Moho boundary at $\sim$ 31 km that does not necessarily correspond to a jump in velocity. Nevertheless, this feature at 31-km depth is interpreted as the Moho, as it is consistent with the other Moho estimations in the profile. Station locations generally exhibit a diffuse Moho, which is interpreted as an increase in V_s between 25- to 45-km depth. This estimate of a flat Moho at 31-km depth is consistent with previous LITHOPROBE seismic refraction and reflection studies (e.g., Cook et al., 2004; Clowes et al., 2005) and RF studies (e.g., Taryoun et al., 2017).

The LAB depths are also a significant feature in the profile-marginal results in Figure 6.1. For the purposes of this study, I define the LAB as the contrast between the high-velocity lid and a

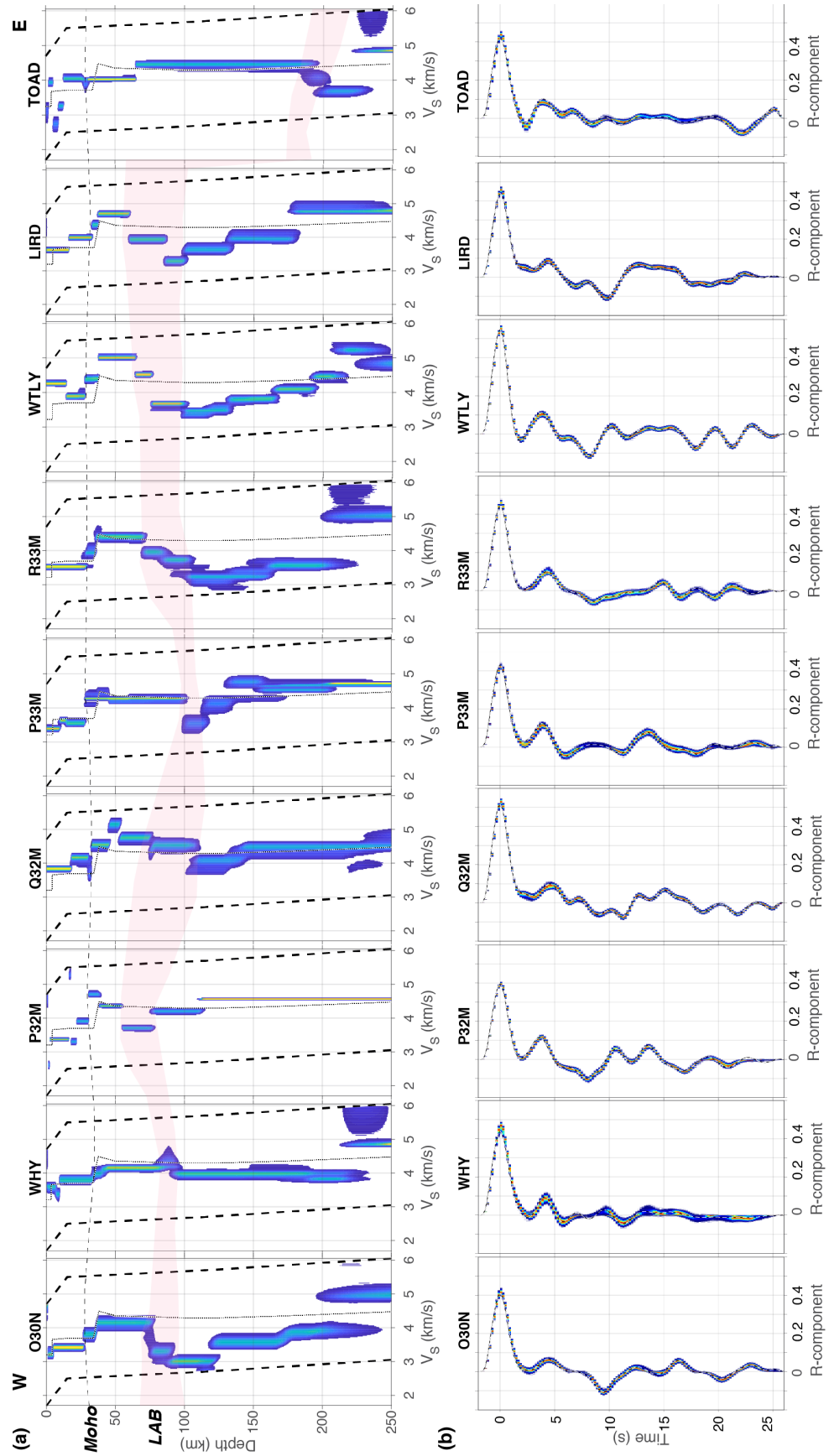

Figure 6.1: (a) Profile marginals for selected stations across the NCC. Vertical dashed lines encasing the marginal are the prior bounds, and the centre, dotted black line is the TNA model (Grand and Helmberger, 1984). The horizontal dashed, line is my estimate of the Moho, at approximately 34-km depth across the cordillera. The light pink shaded area is my interpretation of the LAB, with the vertical width of the shaded area representing the depth uncertainty. The LAB shows high variability across the profile. (b) Range of predicted data (colour scale) and observed data (dashed) for stations.

decrease in velocity directly beneath. As discussed in Section 2.3, this decrease in V_s under a high-velocity lid is the seismic observation likely due to the presence of partial melt underneath the LAB (Eaton et al., 2009; Schmerr, 2012,e.g.,).

I observe a LAB between depths of $\sim$ 80 km for O30N, WHY, and P32M, $\sim$ 100 km for Q32M, P33M and R33M, $\sim$ 70 km for WTLY and LIRD, and $\sim$175 km for TOAD. Compared to the flat, homogeneous Moho, the LAB is a laterally heterogeneous feature across the NCC. LAB depths inferred in this study range from 75–100 km and are deeper than previous RF studies (e.g., Audet et al., 2019) and petrologic studies (e.g., Harder and Russell, 2006; Francis et al., 2010; Batir and Blackwell, 2019) suggest by 10-50 km. Geothermal modelling of the NCC shows a LAB of 75- to 100-km depth corresponds to Moho temperatures of 800° C, resulting in a deeper geotherm than previous seismic scattering and RF data suggest (Langmuir et al., 2006; Audet et al., 2019).

Beneath the LAB, there is a pervasive LVZ with $V_s < 4$ km/s in the upper mantle for most stations. The lowest absolute V_s is observed at station O30N with V_s of $\sim$ 3.0 km/s at 100-km depth. However, these estimates are associated with significant uncertainty with 95% credibility interval widths of 400 m/s. In addition, velocities may be under-estimated due to the lack of reliable SWD data with sensitivity to these depths. The thickness of the LVZ is laterally variable, ranging from $\sim$ 10 km at station WHY to $\sim$ 120 km at station WTLY. The diffusivity of the lower bound of the LVZ is also variable across stations, but most exhibit a gradual increase to $\sim$ 5 km/s. This gradual return from LVZ velocities to typical upper-mantle velocities may be because of the lack of distinct seismic discontinuities underneath the LVZ.

When discussing velocity reductions, it is common practice to consider perturbations from a 1D reference model. Earth models change with respect to the tectonic regime and the choice of reference model can impact inferred V_s as a percentage reduction. For example, in North America at 100-km depth, a reference model in the tectonically active area (TNA) has a V_s of 4.29 km/s, but in the tectonically stable region (SNA), V_s is 4.775 km/s (Grand and Helmberger, 1984). When calculating V_s reduction percentages, I use the TNA model (Grand and Helmberger, 1984) for a

1D reference model (Appendix E).

My observations of the LVZ mostly agree with previous works that showed similar magnitudes in negative velocity perturbations (e.g., Mercier et al., 2009; Schaeffer and Lebedev, 2014; Zaporozan et al., 2018). Similarly extreme LVZs with V_s to ~ 3.6 km/s are observed in the Hawaiian plume (Rychert et al., 2013) and the African Rift (Armitage et al., 2015). While less extreme than the observations in this study, absolute V_s of ~ 4.1 km/s are inferred in regions of asthenosphere upwelling, such as southern California (e.g., Yang and Forsyth, 2006), and the Hangai Dome in Mongolia (e.g., Chen et al., 2015). Overall, the magnitude of V_s reductions observed in this study are high and may be affected by a lack of SWD information sensitive to the LVZ depths. However, my observation of a LVZ agrees with previous studies and I present strong evidence for the pervasive nature of this feature and its depth. In particular, the lateral extent of the LVZ in the NCC has not previously been resolved (e.g., Frederiksen et al., 1998; Schaeffer and Bostock, 2010). Notably, the LVZ in this study extends east of the Tintina Fault, which has previously been inferred as a boundary between two distinct lithospheres and the eastern edge of the NCVP (e.g., Edwards and Russell, 2000; Abraham et al., 2001; Dalton et al., 2011; McLellan et al., 2018).

6.1.1 Reductions of Shear-wave Velocity

Lateral variations to observed V_s arise from changes in characteristics of the upper mantle, major-element composition, water content, temperature or the presence of melt. A single V_s observation cannot determine a single temperature and composition, because these thermochemical parameters have multiple trade-offs (e.g., Takei, 2017; Cobden et al., 2018). Therefore, I consider V_s reductions to arise from the synergistic combination of multiple V_s-reducing mechanisms.

The composition of the mantle can influence V_s, primarily with variations of the magnesium number (Mg# = Mg/(Mg+Fe)) in mantle peridotite. This number corresponds to the ratio of a sample between Mg and Fe end-members of mantle minerals, which is primarily olivine and pyroxene (e.g., Green and Ringwood, 1963; Francis, 1987). Mantle *enriched* in Fe-content has lower seismic velocities compared to Mg-rich, Fe-*depleted* mantle. Depleted mantle can become

refertilized by silicate melts from the asthenosphere. In the case of refertilization, the content of the element (Fe) increases, and makes the lithosphere more *fertile*, thus decreasing V_s. Xenolith peridotites above the detected LVZ in the NCVP are bimodal, composed primarily of a fertile spinel lherzolite and depleted harzburgite (e.g., Shi et al., 1998; Francis et al., 2010). Fertile lherzolite xenoliths with similar compositions are prevalent across the NCC, unlike the depleted harzburgites, which are present in volcanic flows above the slab window. Bulk xenolith analyses for non-anomalous xenoliths in the NCC indicate Mg# that range between 0.893–0.913, which is lower than the standard range for the shallow mantle $\sim 91 <$Mg#$<\sim 93$ (Francis, 1987; Boyd, 1989; Bernstein et al., 2007). Even considering a large Mg# range of 86 and 94, wave velocities only vary by $\sim 2.5\%$ at the depths I am interested in (Lee, 2003). Velocity estimation for the most fertile lherzolite and most depleted harzburgite cannot account for the range in teleseismically detected velocities (Shi et al., 1998; Frederiksen et al., 1998).

The presence of water can effect V_s through several mechanisms, such as directly through the change in bond strength or indirectly through the enhancement of anelastic relaxation (Karato, 1995). The loss of energy of seismic waves (anelasticity), described by the seismic quality-factor, $\hat{Q}$, is often ignored when making simplifying assumptions, but becomes important in the asthenosphere, where $\hat{Q}$ is low (e.g., Karato, 1993, 1995). Reduction of seismic velocities (Karato, 1993)

$$\delta V/V = -(1/2)\cot(\pi\hat{\alpha}/2)\hat{Q}^{-1}(\omega), \tag{6.1}$$

where δV is the velocity reduction due to anelasticity and $\hat{\alpha}$ is a parameter to describe the frequency dependence of (Karato, 1995).

$$\hat{Q}(\omega) \propto (\omega)^{\hat{\alpha}}. \tag{6.2}$$

Considering a range of $0.1 < \hat{\alpha} < 0.3$ and $50 < \hat{Q} < 100$, $\delta V/V \sim -(1\text{–}6\%)$, which suggests that variations in $\hat{Q}$ or $\hat{\alpha}$ will result in significant changes in V_s (Karato, 1995). Note that the precise value of $\hat{\alpha}$ is very uncertain, with most studies placing it in the range of 0.1–0.4 (e.g., Jackson et al., 2002; Fau, 2005; Romanowicz and Mitchell, 2007; Cobden et al., 2018), but is likely in itself

frequency dependent, and may even become negative at long periods (Lekić et al., 2009). Also, note that when the algorithm uses RAYSYM (Frederiksen and Bostock, 2000), it ignores anelastic effects, and, therefore, results may be biased.

Significantly modified seismic properties in the presence of water have been observed at low frequencies in exploratory studies (Aizawa et al., 2008), but recent laboratory studies indicate that reduced velocities arise from prevailing redox conditions and the presence of melt, rather than the presence of fluids or hydrated defects (Cline II et al., 2018). Therefore, oxygen fugacity, or the measure of oxygen available to freely react with elements (Fe or C), may significantly reduce V_s (Cline II et al., 2018). However, Alligator Lake lherzolites, a NCVP volcanic complex located $\sim$ 30 km southwest of Whitehorse, indicate relatively low oxygen fugacity (-1.8 to -0.5 δ fayalite-magnetite-quartz, FMQ), which is similar to oxygen fugacities in the southern Canadian cordillera (-1.5 to -0.5 δ FMQ), and the two lithospheric regions are assumed to be genetically related (Canil et al., 1990; Kilgore et al., 2018). Such values indicate that the redox conditions of the NCC would not significantly reduce V_s (Kilgore et al., 2018; Cline II et al., 2018).

Attenuation and V_s in the NCC are more likely influenced by the presence of partial melt. When considering a dry ($C_{OH} = 50H/10^6Si$), damp ($C_{OH} = 1{,}000H/10^6Si$), and wet mantle ($C_{OH} = 3{,}000H/10^6Si$) water content as a parameter in the anelasticity correction with $\hat{\alpha} = 0.27$, $V_s < 4.0$ km/s at 132-km depth can only be modelled via melting for mantle compositions with melt (Fau, 2005; Cobden et al., 2018). Dry, damp and wet mantle require crustal compositions to reach $V_s < 4.0$ km/s, which would be unreasonable to expect at 132-km depth. Therefore, I consider partial melt the primary V_s reducing mechanism, but note that high temperatures also contribute in inducing melt and reducing V_s.

There are multiple competing models to relate V_s reductions to the amount of melt present. Models considering V_s reductions address the frequency-dependence of V_s, and account for the time scale on which forces are accommodated in microstructures. This time-dependence can be characterized as an instantaneous response (unrelaxed state) or the long-term equilibrium (relaxed

state). In a modelling study from Hammond and Humphreys (2000b), the unrelaxed and relaxed state are considered as cases in which fluid flow between pores has not yet diminished and when pore pressure is in equilibrium, respectively. Anelastic parameters are considered for the relaxed state in this model, because Hammond and Humphreys (2000a) deem it more reasonable for seismic frequencies. Models arising from ultrasonic experimental studies determine a relationship between V_s and melt fractions, and determine V_s relationships for seismic frequencies by adjusting $\hat{\alpha}$ (Chantel et al., 2016).

I consider V_s reductions based on finite difference modelling of melt geometries, including tubules, ellipsoids and random cuspates for melt fractions $< 1\%$ and includes the presence of films for melt fractions $> 1\%$ (Hammond and Humphreys, 2000b). In partially molten upper mantle containing 1% melt, V_s decreases at least $\sim 8\%$ for each percent melt in realistically shaped melt inclusions (Hammond and Humphreys, 2000b). With this linear relationship, I calculate that V_s reductions observed in this study indicate melt fractions of up to 3.5% in the NCC. When estimating partial melt fractions from V_s reduction underneath the LAB, I select up to three probable LAB interfaces for each station. Selecting interfaces at stations P32M and P33M is straightforward, because there is a sharp, localized decrease in V_s. Selecting LAB boundaries for broad, diffuse V_s reductions at stations, such as R33M and R33M, is more complicated. Therefore, selecting multiple interfaces at these stations allows for a more comprehensive range of partial-melt fraction estimations. Compared with heat flow data across the NCC at latitudes $> 59^\circ$ N, melt fractions are generally higher when heat flow is higher.

As shown in Figure 6.2, estimates for partial melt fractions at LAB boundaries range from 0% up to $\sim 3.5\%$. Partial melt estimates range from 0.5% at stations WHY and Q32M to 2.75–3.5% at station O30N. Western stations, from O30N to P32M, have clear LVZs, resulting in a small range of partial melt fraction estimations. Stations P33M, R33M, WTLY and LIRD have significantly wider ranges of possible partial melt fractions compared to the remaining stations. The presence of partial melt may extend to station TOAD, inferred to be near the CCB, because station TOAD is

where partial melt fractions decrease in depth. Overall, ~ 0.5 to 3.5% partial melt beneath the LAB provides a mechanism for the observed V_s reductions across the NCC. However, this is a purely mechanical model that does not address other parameters that can reduce V_s, such as the frequency dependence of anelasticity (e.g., Karato, 1993) or the presence of melt bands (Kawakatsu et al., 2009). Notably, this model assumes that anelastic relaxation occurs sufficiently quickly, so that it is the relaxed modulus that is excited by observed seismic waves (Hammond and Humphreys, 2000a,b).

Experimental results that observe reductions in V_s are now available for representative mantle pressures (Chantel et al., 2016), and results show larger V_s reductions for the same partial-melt fractions for seismic-wave frequencies than modelled by (Hammond and Humphreys, 2000b). Partially molten samples with mixtures of olivine and different basalt composition percentages are observed for in situ ultrasonic velocity frequencies (Chantel et al., 2016). Employing a range of $\hat{\alpha}$ values, as shown in Figure 6.3, accounts for anelasticity and reduces the amount of melt to fit my observations decreases to $\sim 2\%$ melt. I discuss the process in determining the melt relationships for various $\hat{\alpha}$ in Appendix F. The black dots in Figure 6.3 represent the observed V_s reductions for melt fractions at ultrasonic frequencies, and these V_s reductions are comparable to the melt fractions calculated using the linear model in Figure 6.2. Decreasing $\hat{\alpha}$ decreases the amount of partial melt for observed velocities, as indicated by the coloured dots, which represent realistic $\hat{\alpha}$ values of 0.4, 0.26 and 0.2 (Chantel et al., 2016). For stations O30N and LIRD, adjusting $\hat{\alpha}$ reduces partial melt fractions from $\sim 3.5\%$ to $\sim 1.5\%$. Nevertheless, the $\hat{\alpha}$ value in the NCC might be as low as 0.1, which is the value inferred for Southern California (Yang and Forsyth, 2008). Southern California has a similar LVZ to what is observed in this study and observed $\hat{\alpha}$ values in California are likely similar to what would be expected in the NCC. I could not estimate partial melting for $\hat{\alpha} = 0.1$, as it was not included in Chantel et al. (2016), but note that partial melt percentages calculated from V_s reductions using $\hat{\alpha} = 0.1$ would be even smaller than shown in Figure 6.3.

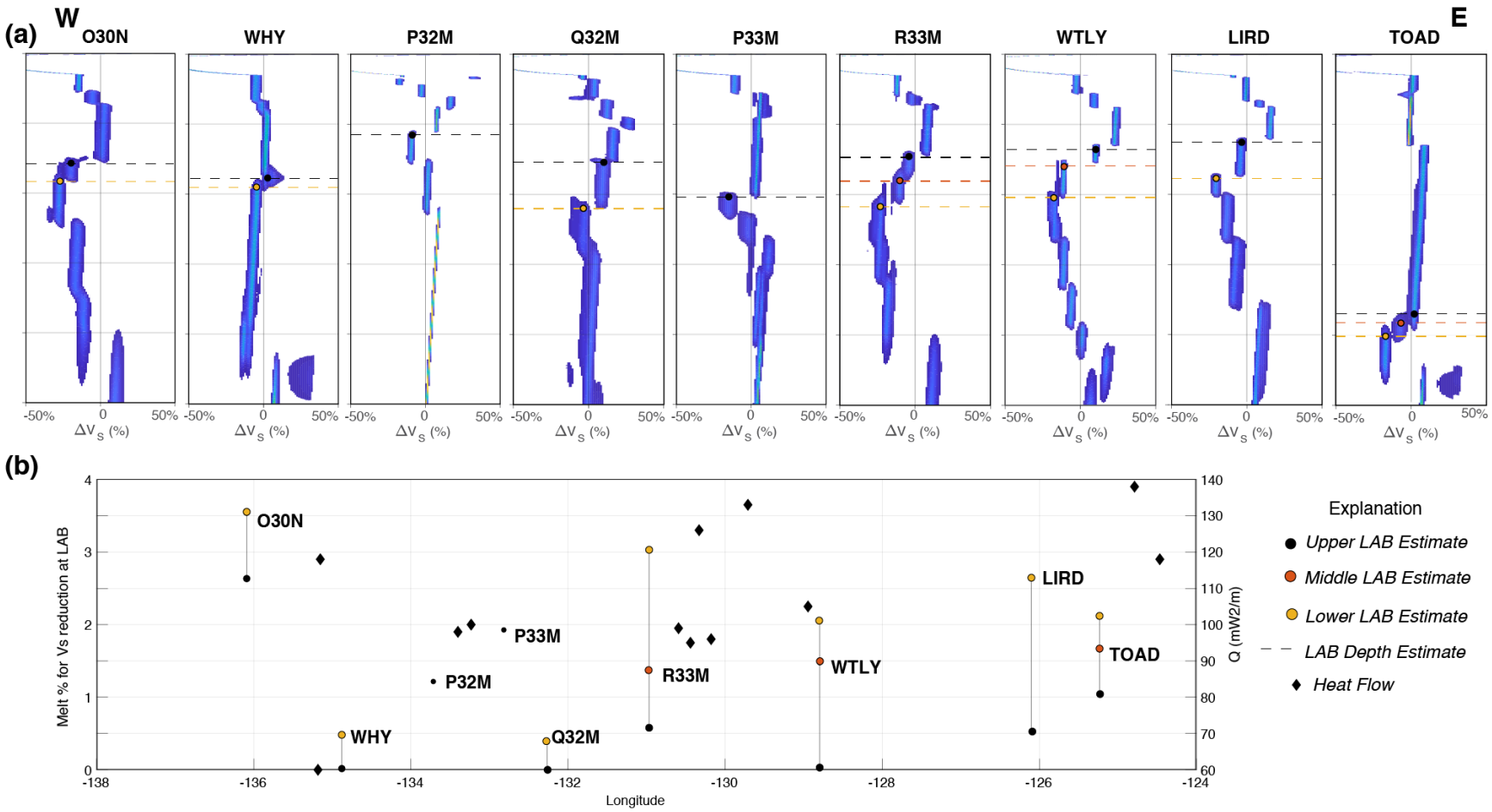

Figure 6.2: (a) Three probable LAB depths are estimated: a shallow upper-bound value, an intermediate value and a deeper, lower-bound value. Shear wave reduction percent of TNA(Grand and Helmberger, 1984) is shown, where the 0 marker represents the TNA velocity at depth. (b) Partial melt fractions as a function of latitude estimated for stations at the LAB with the assumption that 1% partial melt reduces V_S by 7.9% (Hammond and Humphreys, 2000b). Heat flow data from Lewis et al. (2003) for latitudes$>$ 59° North.

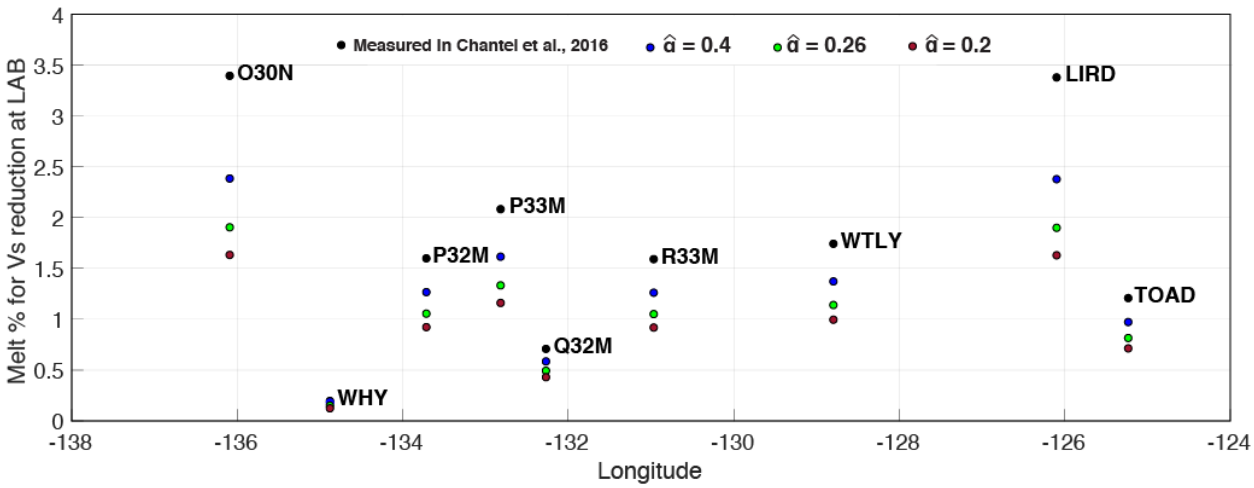

Figure 6.3: Partial melt fractions for all stations with varying attenuation dependence of frequency, $\hat{\alpha}$ described by Chantel et al. (2016). Accounting for the frequency dependence for attenuation in the mantle lowers the melt fraction for each V_s observation.

However, the laboratory-determined model assumes that basalt will have completely melted when assemblage has reached the liquidus of the basalt, so there is a fixed amount of melt despite temperature variations in the cell (Chantel et al., 2016; Weidner et al., 2018). These factors have not all been integrated into a single model. Nevertheless, there are promising models to be further developed. A dynamic melting model (Li and Weidner, 2013) is a promising model that addresses the effects of melt by considering thermodynamic interactions between the solids and liquids (Weidner et al., 2018). Furthermore, when accounting for attenuation effects and mobility of melt at a threshold (Faul, 2001), V_s can be reduced to $\sim$ 3.0 km/s in models of the hot, continental African Rift Zone (Armitage et al., 2015). Overall, further research is required to better constrain the relationship between melting and V_s, but models generally acknowledge assumptions will underestimate the affect of the presence of melt on V_s reductions. In conclusion, my results show significant evidence for the existence of pervasive partial melt in the NCC, with an upper limit of my estimates of 1.5–3% partial melt fractions.

The presence of partial melt in the NCC asthenosphere has been established in previous studies. Geotherms calculated for the NCC indicate to be near-adiabat or greater than the adiabat underneath the LAB, which can explain the presence of partial melt (e.g., Audet et al., 2019; Chen et

al., 2019). The adjacent Southern Canadian Cordillera (SCC) sits above a subducting slab, where low-resistivity structures extend to the TF at ~ 60 km, indicating shallow asthenosphere (Rippe et al., 2013). Profiles in the SCC also exhibit LVZs, which cannot be explained with dry olivine, but rather by the presence of 0.4%–1.5% partial melt (Rippe et al., 2013). Estimated partial melt fractions in the NCC are greater than in the SCC, likely due to the upwelling asthenosphere introducing more melt to the region. Alligator Lake xenoliths suggest the opening of the underlying slab window induced localized partial melting of $\sim$ 20–25% of spinel lherzolite lithosphere (e.g., Francis, 1987; Shi et al., 1998; Francis et al., 2010; Polat et al., 2018). While my inferred melt fraction is less than determined via xenoliths, my observations are across a much larger scale, and may be a more average representation of melt present.

The presence of partial melt could be explained via asthenosphere upwelling, so that upwelling asthenosphere undergoes partial melting via adiabatic decompression to shallow depths (White and Mackenzie, 1989). The LVZ I observe in this study is similar to the ubiquitous LVZ observed under oceanic plates in partial melt fraction, depth and extent (e.g., Schmerr, 2012; Sakamaki et al., 2013). Upwelling in mid-ocean ridges induces LVZs with V_s reductions of 7–8% underneath the LAB (depths of 70–100 km) attributed to partial melting in oceanic lithosphere (Kawakatsu et al., 2009; Hirschmann, 2010; Takei, 2017). Sharp negative velocity contrasts in the ocean can be explained by a LVZ with melt volume of 0.1–3%s (Sato et al., 1989; Schmerr, 2012), and even higher partial melts are possible if volatiles are present (Hirth and Kohlstedt, 1996). Partial melt peak melt mobility at the base of the lithosphere can further explain the depth and extent of the LVZ (Sakamaki et al., 2013). Ratios between melt-solid density in contrast to magma viscosity peaks at pressures corresponding to depths of 120–150 km, approximately the depth at which I observe minimum velocities in the interior of my study region (Sakamaki et al., 2013). Similar LVZs have been observed in continental settings, as well. Partial melt fractions of $\sim$ 3% at 70–100-km depth have been inferred Under the Hangai Dome in Mongolia, where, like the NCC, small-scale upwelling of the asthenosphere induced intraplate volcanism and partial melt (Comeau et al.,

2018). In the Basin and Range province in southwestern US, wet basalts may have begun melting well below the dry solidus, and the depths where melt is ponding correlate to the equilibriation beneath the nominally dry lithosphere (Plank and Forsyth, 2016). Therefore, I conclude that partial melt may be pooling at the base of the lithosphere possibly due to regional asthenosphere upwelling and anomalously high temperatures.

6.2 Tectonic Model with a Heterogeneous LAB and LVZ

Observations of a heterogeneous LAB and underlying LVZ due to the presence of partial melt at the LAB has implications on the tectonic models describing the thermochemical state of the NCC. I infer a drastic increase in LAB depth at station TOAD, nearest to the deformation front. Between TOAD and its neighbouring station to the west, LIRD, the LAB increases in depth by ~ 100 km. This dramatic deepening is interpreted as the CCB, and its location and dramatic nature agrees with previous observations of a sharp CCB (e.g., Dalton et al., 2011; McLellan et al., 2018; Chen et al., 2019). The sharp transition between a shallow LAB across the Cordillera and a deep cratonic LAB is a unique geometry likely formed during the accretion of terranes and can affect the thermochemical state of the NCC, providing insight into its formation (e.g., Chen et al., 2019).

As discussed in Section 2.3, Chen et al. (2019) proposes three mechanisms for forming the CCB: terminal suturing during accretion (Chen et al., 2019); lithosphere delamination (Bao et al., 2014); and viscous thermal erosion (e.g., Currie and Hyndman, 2006; Hyndman et al., 2005; Hyndman, 2010, 2017; Francis et al., 2010). All of these mechanisms have distinct geometries that should be seismically observed on a lithospheric scale, as described in **??**.

In most schematic tectonic models, the LAB is shown as constant and relatively homogeneous, much like the Moho. Like the Moho, the LAB seems to be minimally, if at all, affected by terrane boundaries. However, unlike the Moho, the LAB depth is heterogeneous. Based on xenolith studies observing bimodal suites (e.g., Shi et al., 1998; Edwards and Russell, 2000; Francis et al., 2010) and heterogeneous water content (Kilgore et al., 2018) and on my observations of varying

LVZ and LAB depths, the lithosphere is laterally heterogeneous. Tectonic models proposing exotic lithosphere docked onto the ancient North American craton (e.g., Audet et al., 2019; Chen et al., 2019) suggest the exotic lithosphere is flat and relatively homogeneous in terms of both Moho and LAB depths. Stations west of WTLY in this study are part of this presumed exotic lithosphere, and exhibit various LAB depths and partial melt fractions. While observations in this study do not rule out an exotic origin for western NCCs lithosphere, assuming it is a homogeneous addition to Laurentia ignores large-scale lithosphere variability. Therefore, a model of a terminal suture of ancient North America does not account for lateral variation observations discussed in this study, so I consider other thermochemical models to address these issues.

Alternatively, a large-scale delamination event or small-scale convection could account for observations of partial melt and high heat flow across the NCC. Lithospheric removal may have been removed through a large-scale delamination event or ongoing small-scale convection. The presence of melt could have aided the lithospheric removal process, because the melts and fluids percolating into the lithosphere may have contributed to its weakening (e.g., Bao et al., 2014; Polat et al., 2018; Kilgore et al., 2018). Furthermore, this process may be aided by the edge-driven convection from the sharp CCB (Bao et al., 2014). A large-scale delamination event between 5–25 Ma can reproduce seismic observations (Audet et al., 2019; Chen et al., 2019), so that the opening of the slab window triggered the delamination event and allowed the displacement and foundering of the lithospheric root (e.g., Thorkelson et al., 2011; Audet et al., 2019).

Ongoing lithospheric drip or thermal erosion through small-scale convection cells do provide a mechanism of lithospheric removal while addressing LAB structural variation. The presence of small-scale, thermally-driven convection cells also incorporates the high heat flow observed across the NCC into the tectonic model (e.g., Currie and Hyndman, 2006; Hyndman, 2010). Partial melting is also further addressed in this model, because geodynamic models suggest small-scale drips of weak mantle are prone to melting in a hot, thin lithosphere (Wang and Currie, 2015). Thus, my observations agree best with the thermally-driven convection model presented in **??**.

When considering tectonic models, however, incorporating the presence of heterogeneity in partial melt fraction and LAB depth addresses the main lithosphere features in the study.

7. Conclusion

7.1 Key Contributions

In this thesis, I employ an efficient Bayesian inference algorithm to resolve the lithospheric structure of a transect across NCC. I combine P-RFs for nine stations with Rayleigh phase velocity SWD data from the GDM52 global model (Ekström, 2011) to invert for upper-mantle and crustal V_s structure across the NCC. Approaching the joint inversion with Bayesian methods rigorously quantifies uncertainty associated with the non-uniqueness of RF and SWD inversion. Importantly, the Bayesian formulation avoids most of the subjectivity associated with prior model selection by treating the number of layers as an unknown. Parallel tempering is applied as an efficient optimization algorithm to improve sampling efficiency. Empirical data covariance estimates account for noise level variations throughout the RF during sampling.

While the depth of the Moho is relatively constant across the NCC ($\sim$ 34 km), the LAB discontinuity shows significant lateral heterogeneity, varying from station to station in the range of $\sim$ 70 and $\sim$ 100 km depth. However, like the Moho, lithospheric structure is not effected by terrane boundaries, and seems to has been largely reset by post-accretion processes. The inversion is able to reconcile SWD data that do not show evidence of a LVZ and P-RFs with strong evidence of a LVZ. The magnitude of V_s reductions across the NCC requires the presence of partial melt beneath the LAB. Asthenosphere upwelling may be driving partial melt pooling at the LAB, explaining the pervasive LVZ observed in this study.

The CCB is inferred at the eastern edge of the transect as a dramatic deepening of the LAB.

This sharp, dramatic LAB depth increase agrees with previous seismic tomography studies that inferred a sharp CCB (e.g., Dalton et al., 2011; Zaporozan et al., 2018). This study is the first to resolve a heterogeneous LAB and an extensive LVZ underneath the LAB. V_s in this transect are reduced down to $\sim$ 3.0 km/s, which can be explained with up to 3.5% partial melt, but considering relaxation effects for seismic frequencies, this partial melt value is likely $\sim$ 2%. Future tectonic models need to be able to account for the variations in depth and partial melt beneath the LAB when considering mechanisms forming the CCB and explaining the thin lithosphere in the NCC.

7.2 Limitations and Future work

The data in this study were limited by the resolution of SWD data and the directional dependence in P-RFs. While SWD data from the GDM52 global model were more reliable than other available SWD databases, it had little regional sensitivity to the LVZ observed in P-RF data. Discarding SWD data between 30 and 150 s circumnavigated this issue, but including the full SWD curve with LVZ observations would be advantageous, so that inversion results could better address crustal V_s variations.

While the lack of information for the LVZ in SWD data is the greatest limitation of this study, directional dependence is another limitation I address. Windowing the BAZ for each station reduced complexities from BAZ-dependence in P-RFs, but the effects of anisotropy or dipping layers on the inversion results were not addressed. Future work incorporating harmonic decomposition or another stripping method to remove directionally-dependent features would be advantageous (e.g., Bianchi et al., 2010; Shen et al., 2013). It is important to note that the harmonic decomposition method must be able to account for poor sampling across all BAZ, so a comprehensive procedure accounting for back-azimuth dependencies to obtain a station-representative P-RF would be benefical (e.g., Kim and Rhie, 2019).

Future studies could also further investigate the possibility of addressing anelasticity in forward modelling. An attenuation study across the NCC to establish $\tilde{\alpha}$ values in the NCC would be

helpful when determining partial melt fractions. Incorporating more stations into the transect will also better resolve the variability of the LAB and LVZ across the NCC. Applying this algorithm to a north-south transect could address the increase of heat flow near 60° N latitude. Similarly, repeating this study for a transect in the SCC could address how the slab window affects the lithosphere structure in the NCC and SCC.

A. Effects of Dipping Layers

In the presence of dipping layers, waveforms exhibit directional dependence, the effects of which have been examined in various simulated studies (e.g., Langston, 1979; Cassidy, 1992). A simple test is proposed in Cassidy (1992) to examine the directional effects using absolute RF amplitudes.

This test considers an east-dipping boundary of 15° at 20-km depth and a half-space at 60-km depth, schematically represented in Figure A.1a. Waveforms for BAZ between 0 and 315° are simulated for this model using RAYSUM (Frederiksen and Bostock, 2000) and are shown in in Figure A.1b. The radial and transverse components from this test show the three ways a dipping layer affects RFs:

1. Ps amplitudes vary as a function of BAZ (and slowness, not pictured). P-waves approaching from BAZ = 90°, or travelling *updip*, generate the largest and latest Ps phases. P-wave approaching from BAZ = 270°, or travelling *downdip*, generate the smallest, earliest Ps arrivals.

2. The amplitude of the direct P-wave in the radial RF changes with respect to BAZ direction. In the test from Cassidy (1992), it is noted that this variation is most pronounced when the dipping layer is greater than 30°. Nevertheless, normalizing the RF to unit amplitude neglects this amplitude information, which indicates extreme structure beneath the receiver.

3. The dipping layer deflects waves from the P-SV system, and introduces SH waves recorded by the transverse component (Langston, 1977). The transverse component has zero amplitude from the waves arriving from updip and downdip. Note the change in polarity between the arrivals between BAZ = 180° and BAZ = 0°.

Overall, these three effects of dipping layers on waveforms are important considerations when

determining structure beneath a receiver. Amplitude variation across BAZ indicates structure and incorporating absolute amplitudes helps avoid potential inaccuracies in earth models with dipping structures.

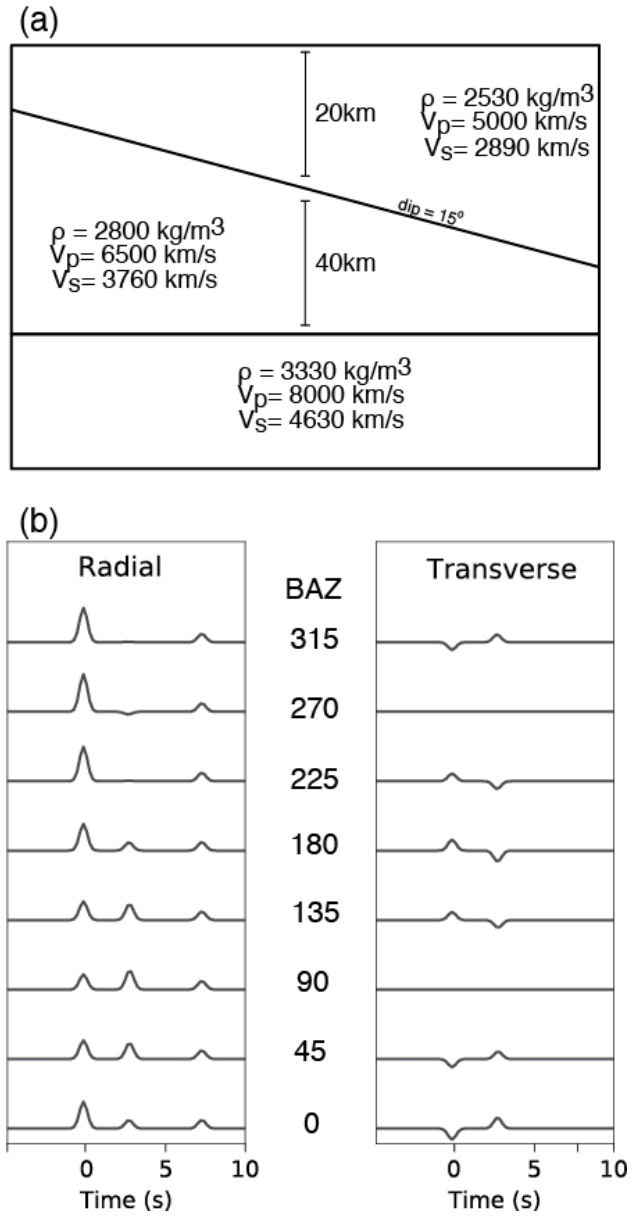

Figure A.1: Simple model from Cassidy (1992) examining the effects of dipping structure on seismograms as a function of BAZ for radial and transverse components (Cassidy, 1992). (a) The model includes a dipping interface of 15°at a range of BAZ. (b) the Radial and transverse components of the seismograms forward modelled using RAYSUM (Frederiksen and Bostock, 2000).

B. Interpretation of Marginal Profiles

Outputs from the algorithm described in section 3.2 are plotted as marginal profiles (Figure B.1.a). Within this plot, the left-most panel describes the interface probability distribution as a histogram showing probability of an interface as a function of depth. The right panel is the profile marginal profile, which describes the ensemble's PPD from z_{min} to z_{max} by colour-coding probabilities of V_s at depth. Warmer colours indicate higher probability of velocity at selected depth, and cooler colours represent lower probabilities. Dashed lines near the bounds of the plot represent the prior bounds of V_s and the middle dotted line is the synthetic input model.

The PPD histograms are shown at two depths:near an interface in Figure B.1.b) and away from an interface in Figure B.1.c). The marginal profile near an interface is wider, and the PPD histogram shows this is because the distribution of solutions is broader. The PPD histogram near the interface shows there are two solutions with nearly the same probability. Away from the interface, the distribution is more narrow in the profile marginal. The corresponding PPD histogram shows that this is because there is one peak in the probability the distribution is more narrow. Higher probability is also expressed as warmer colours in the marginal profile.

Overall, displaying the PPD distribution as the profile marginal plots describes the PPD as a functino of depth and velocity.

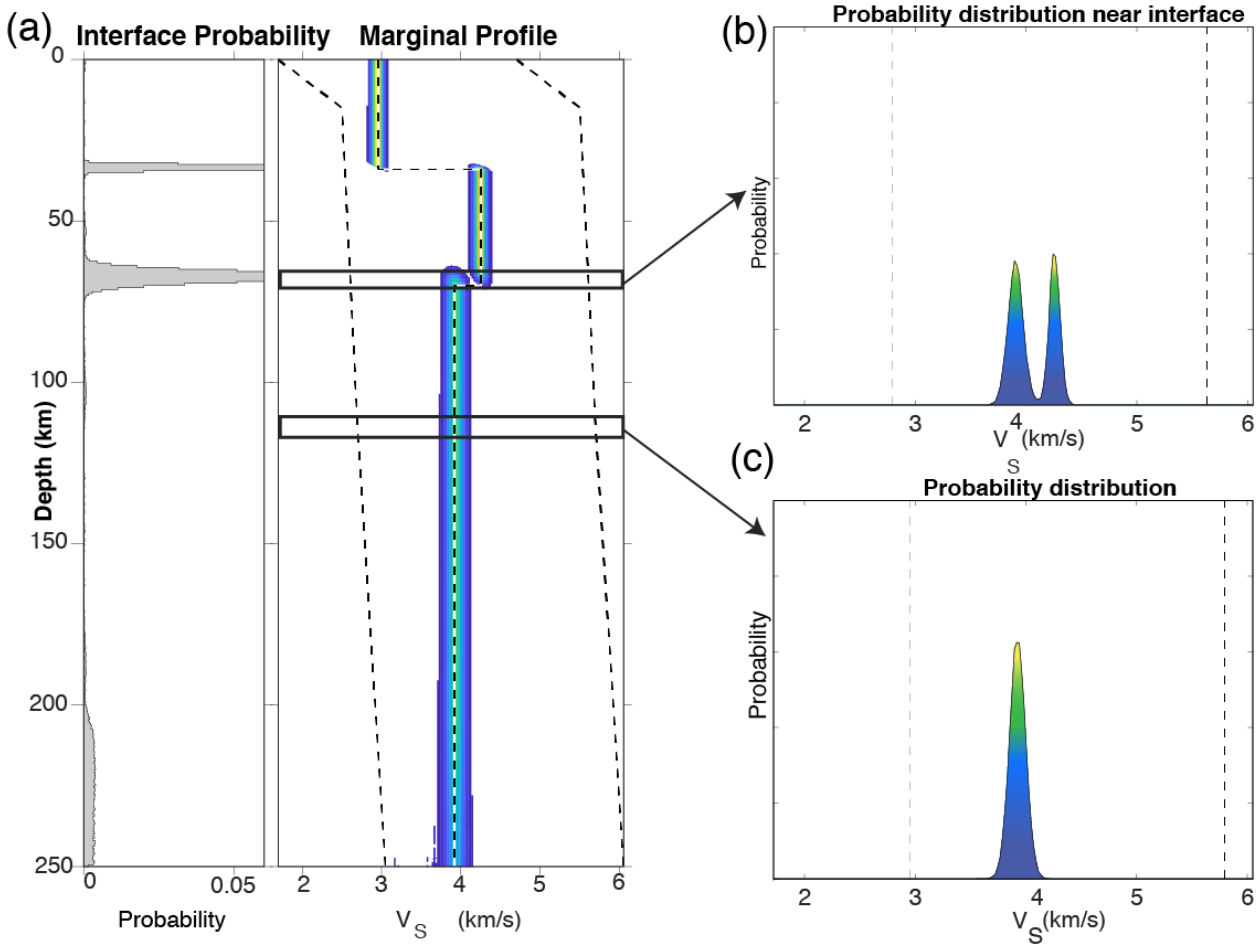

Figure B.1: Schematic representation of PPD for marginal profiles. The interface probability and marginal distribution are shown in Panel a). Thick black boxes represent cross-sections used for Panels b) and c). Panel b) shows the PPD near an interface and panel c) shows the PPD deeper in the profile. In both panels, the vertical, dashed, black lines represents the V_s prior bounds. A single, more probable velocity distribution will have warmer colors and appear thinner in the marginal profile. When uncertainties are higher, the marginal will have cooler colours and appears wider in the marginal profile.

C. P-RFs for Remaining Stations

Additional figures of P-RFs across BAZ along with the resulting stacked P-RF, ordered from west to east, can be seen below. Rather than include the four panels for section 5.2 for all stations, the full BAZ distribution and the stacked P-RF for the BAZ window 280–300° are included. The full BAZ distribution depicts the directional dependence for the station, and the binned, stacked P-RF shows which phases are incorporated into the inversion.

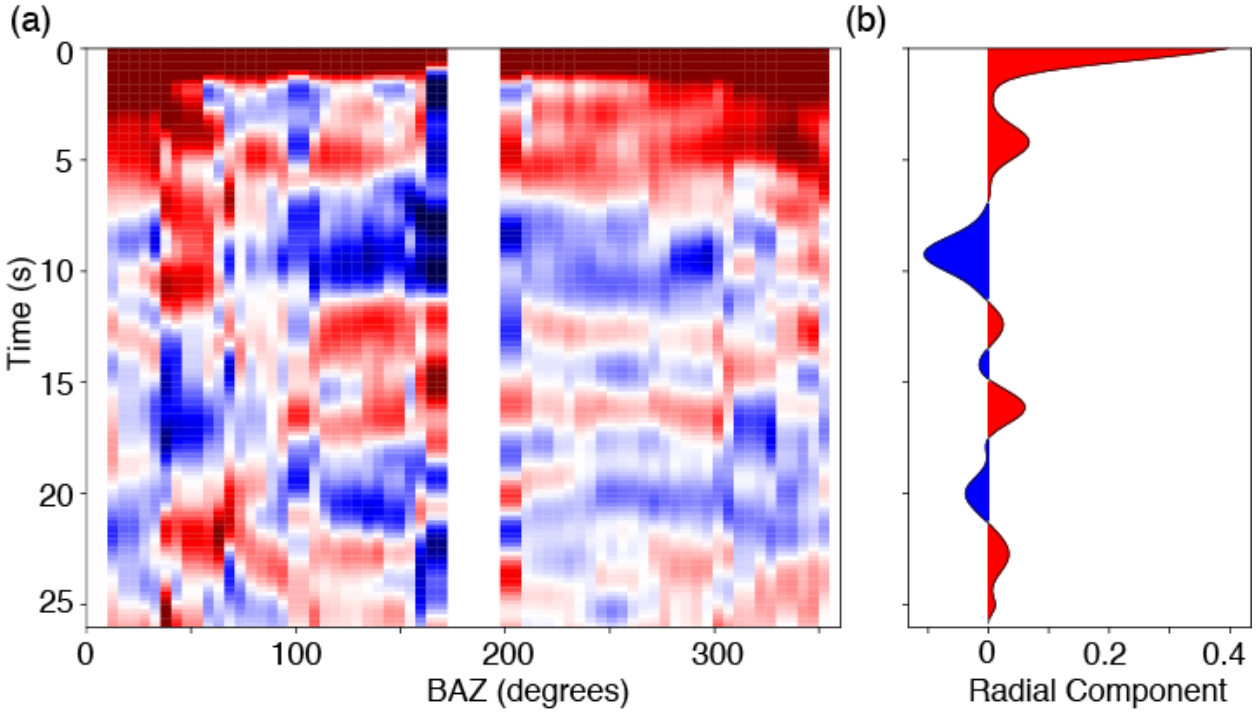

Figure C.1: a) Amplitude of P-RFs by BAZ for O30N. b) Stacked P-RF from 280–300° BAZ.

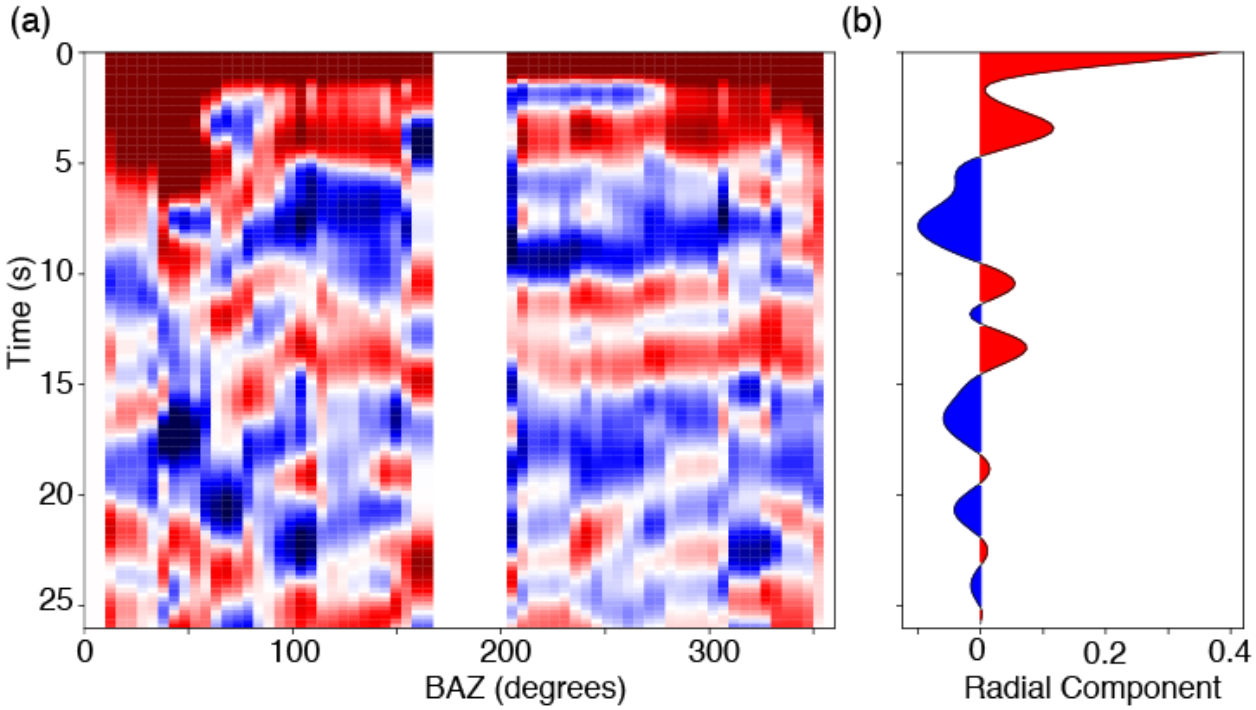

Figure C.2: a) Amplitude of P-RFs by BAZ for P32M. b) Stacked P-RF from 280–300° BAZ.

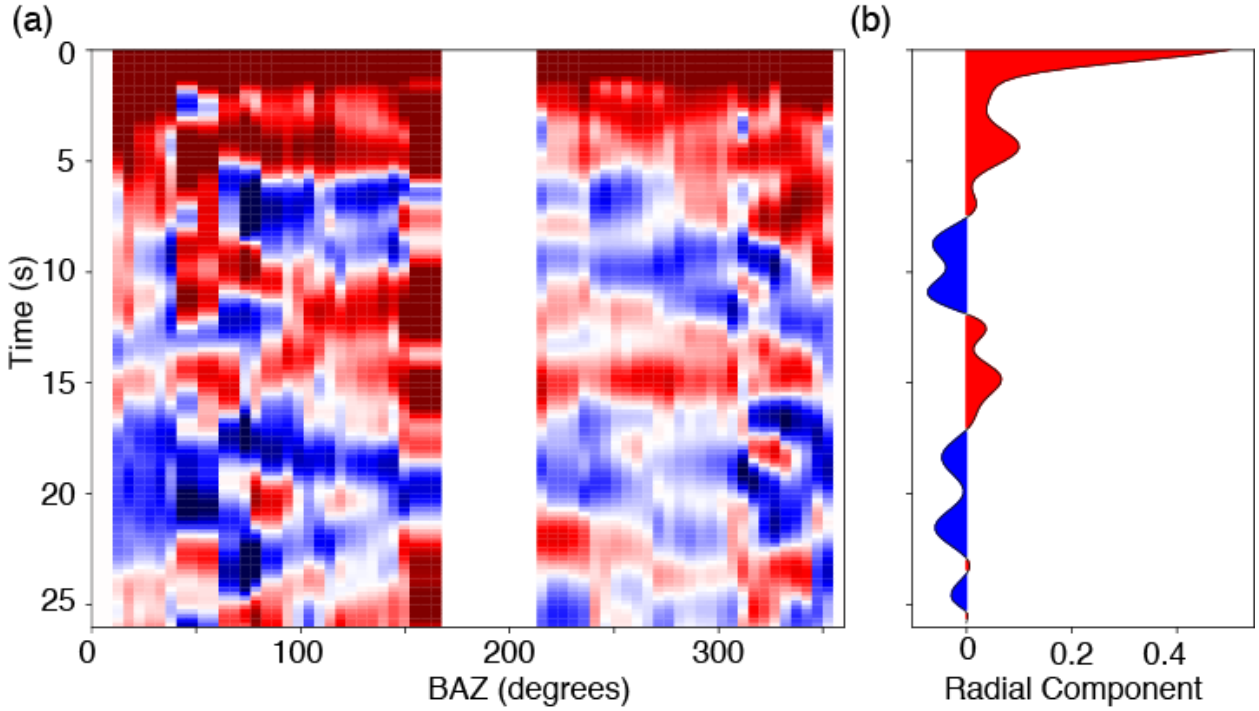

Figure C.3: a) Amplitude of P-RFs by BAZ for Q32M. b) Stacked P-RF from 280–300° BAZ.

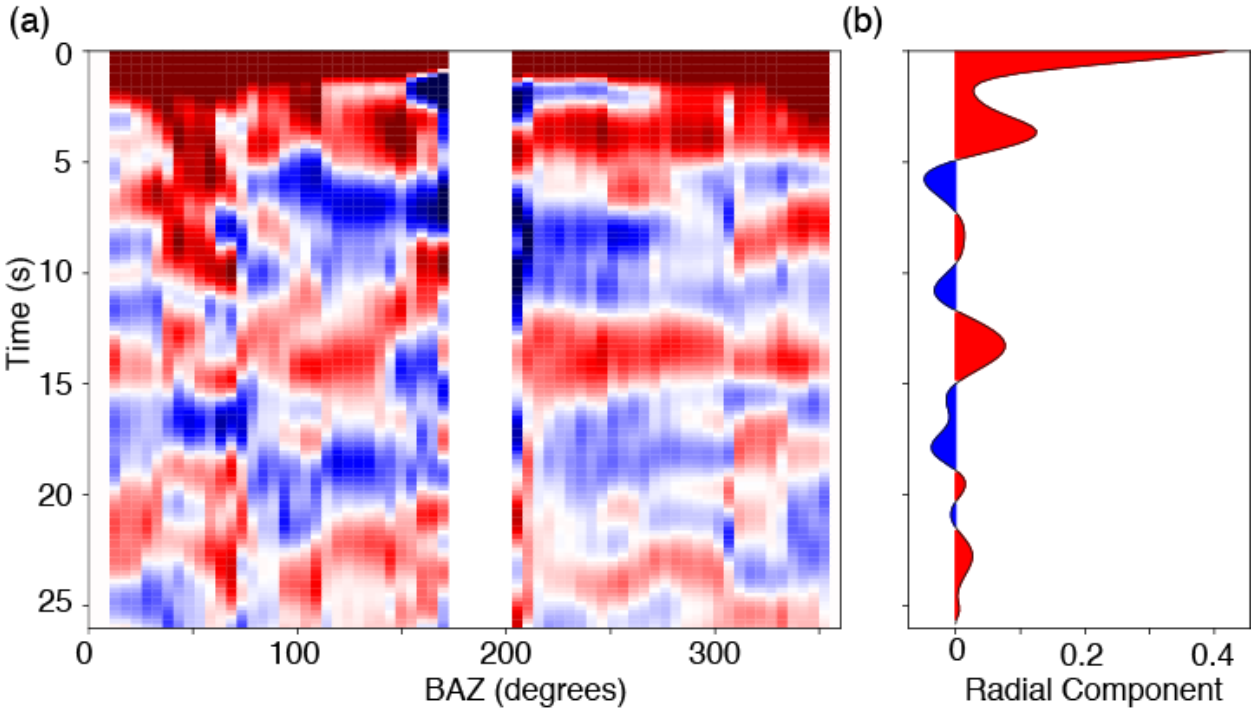

Figure C.4: a) Amplitude of P-RFs by BAZ for P33M. b) Stacked P-RF from 280–300° BAZ.

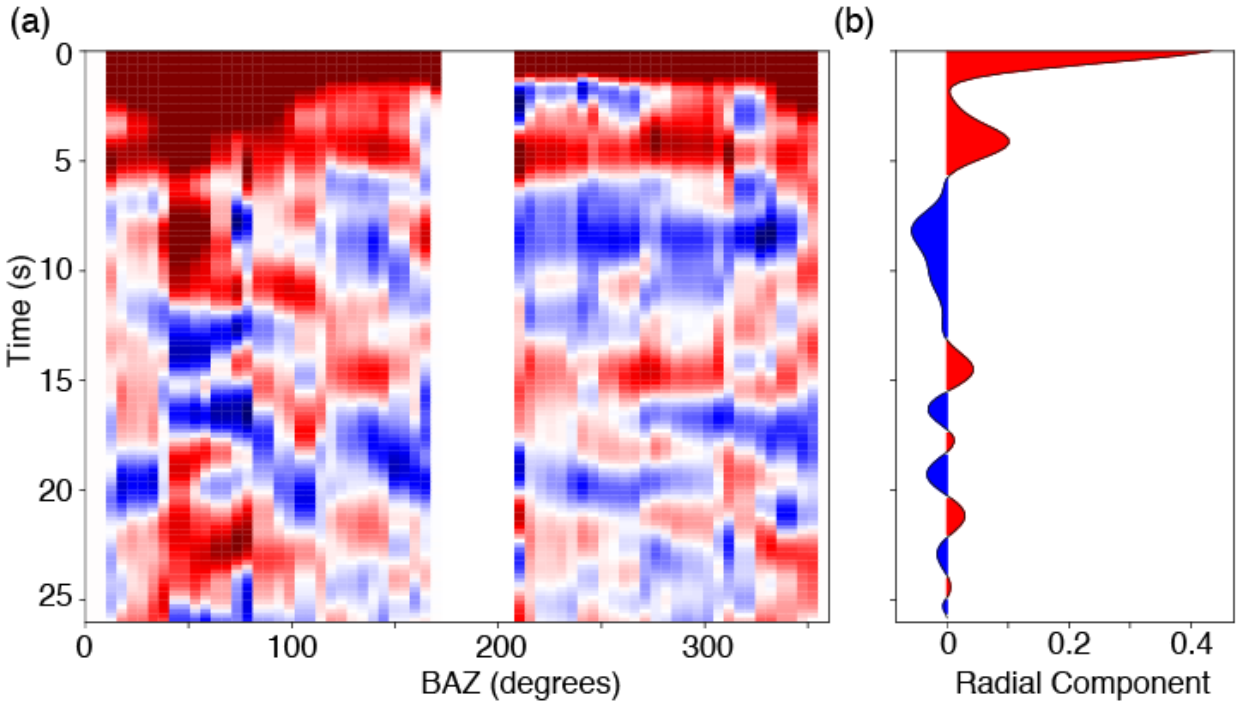

Figure C.5: a) Amplitude of P-RFs by BAZ for R33M. b) Stacked P-RF from 280–300° BAZ.

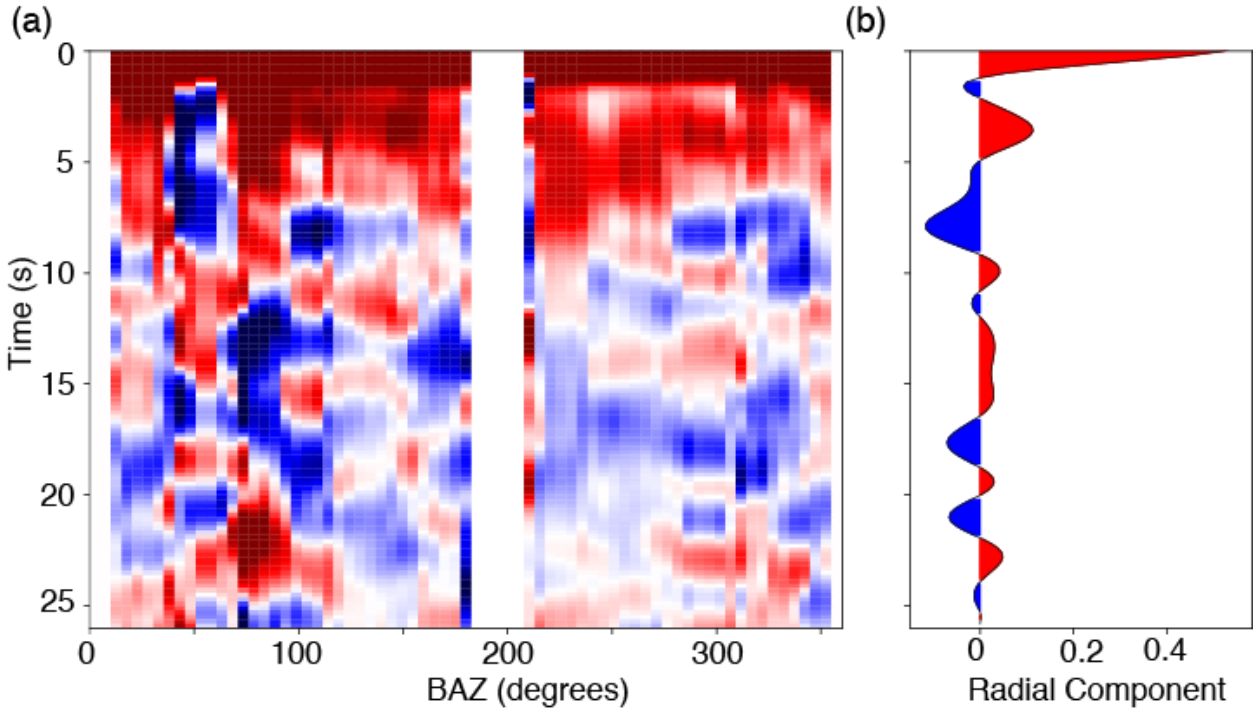

Figure C.6: a) Amplitude of P-RFs by BAZ for WTLY. b) Stacked P-RF from 280–300° BAZ.

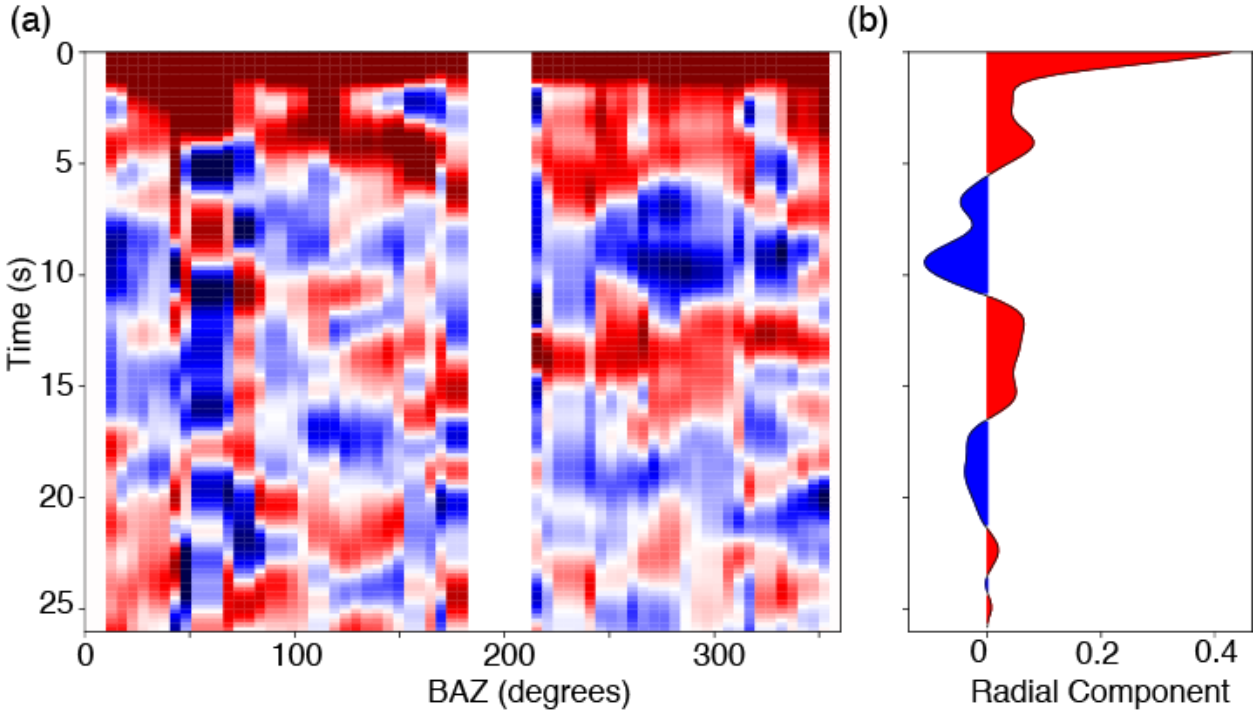

Figure C.7: a) Amplitude of P-RFs by BAZ for LIRD. b) Stacked P-RF from 280–300° BAZ.

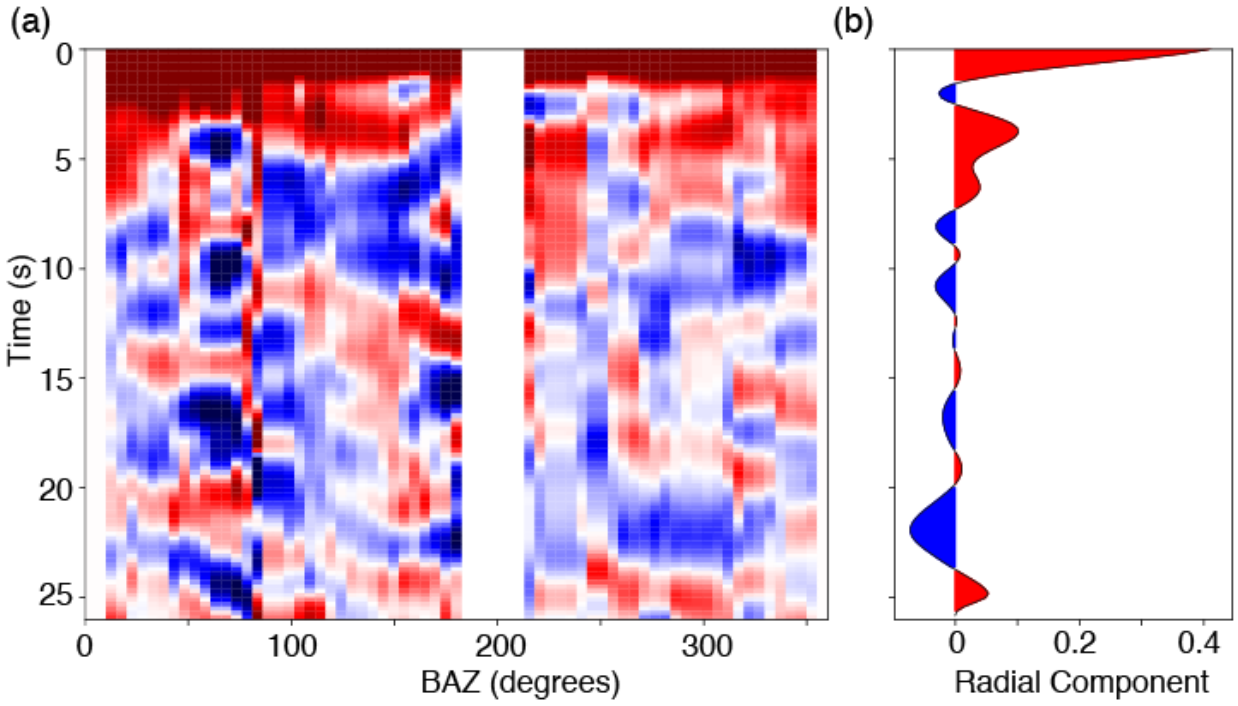

Figure C.8: a) Amplitude of P-RFs by BAZ for TOAD. b) Stacked P-RF from 280–300° BAZ.

D. 1D Inversion Results

In this Appendix, the remaining inversion results depicting SWD, P-RF and joint inversions and full inversion results are shown that were not shown in the main text. Stations are ordered from west-to-east.

D.1 Comparison between SWD, RF and joint inversions

For all stations, the SWD inversion describes an absolute V_s profile, void of any interface probability of notable probability. There is a notable change between SWD and P-RF inversions, because the P-RF inversions exhibit probable interfaces. Joint inversions of both datasets visually appear more similar to P-RF inversions, but are more narrow. While some stations' V_s at 250-km depth are similar between the P-RF and joint inversions, some stations' (e.g., R33M) V_s increase with the joint inversion.

D.2 Joint Inversion Marginals and Statistics

The profile marginals, data ranges for P-RF and SWD, node probability distribution and P-RF scaling factor and SWD noise standard deviations are described for all stations. Note in the P-RF data fit ranges how well major features are resolved for all stations. The SWD data ranges are less consistent, with some stations (e.g., P33M, TOAD) fitting lower period SWD well. Stations exhibiting a large LVZ (e.g., R33M, WTLY) have larger misfits between the low period SWD data. Layer distributions vary across stations, ranging from 8 to 15 layers. All P-RF empirical covariance

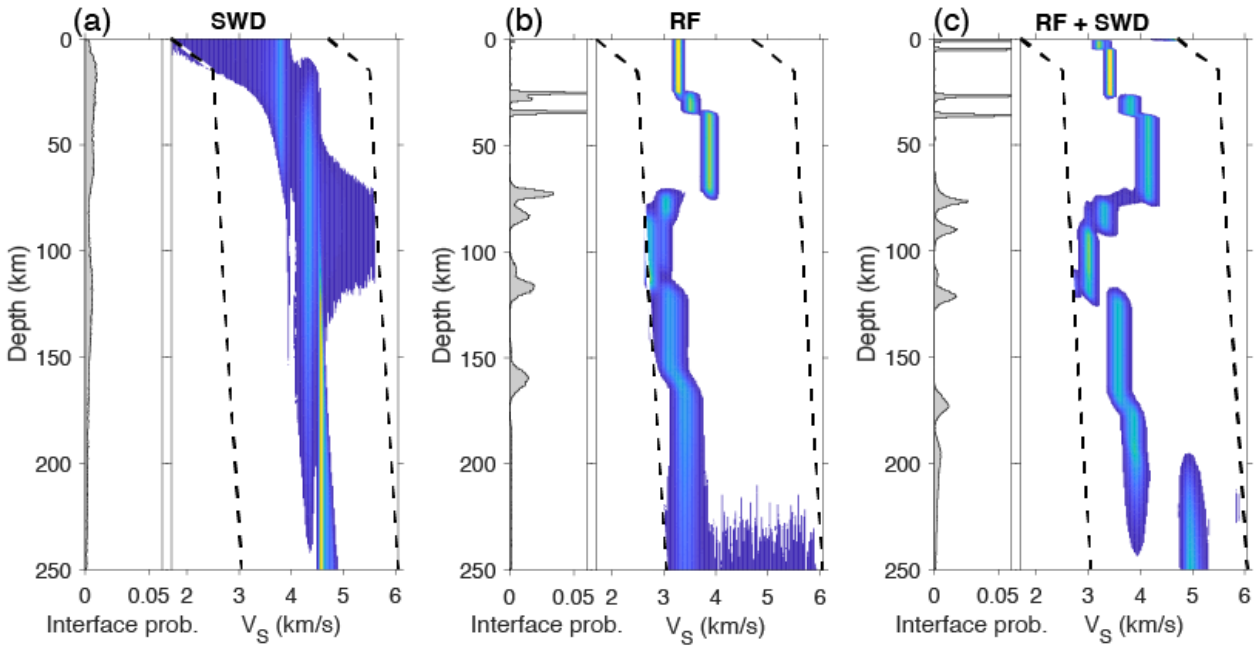

Figure D.1: (a) SWD-only, (b) RF-only and (c) joint inversion for station O30N.

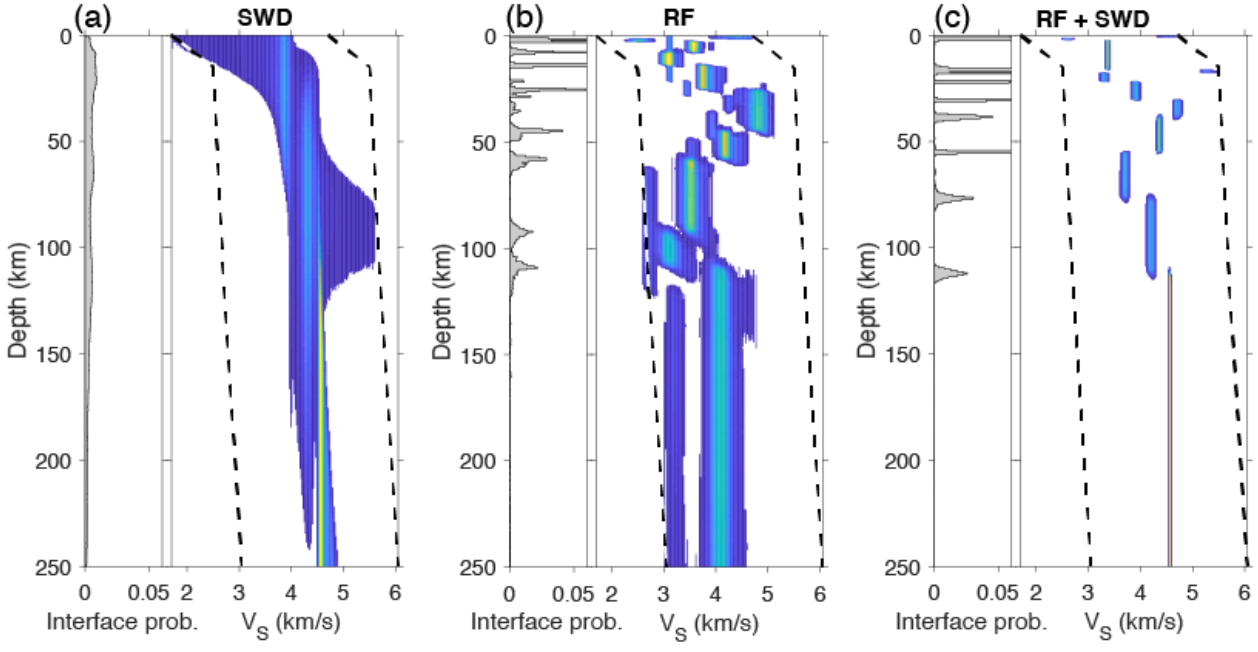

Figure D.2: (a) SWD-only, (b) RF-only and (c) joint inversion for station P32M.

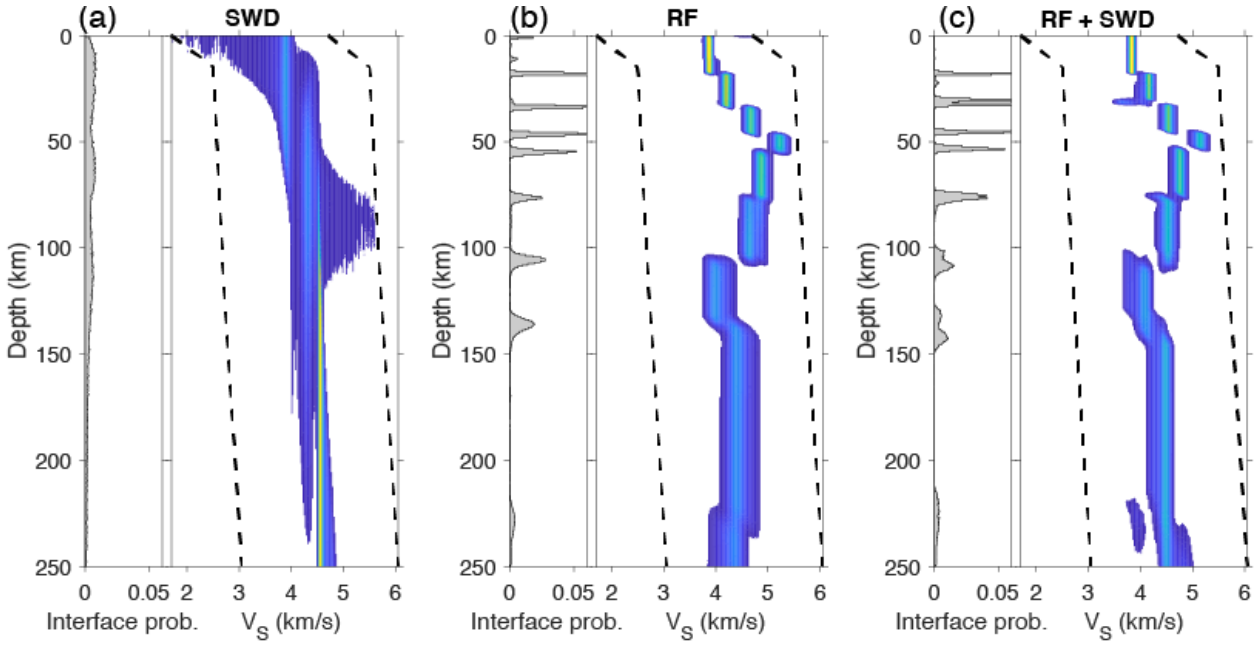

Figure D.3: (a) SWD-only, (b) RF-only and (c) joint inversion for station Q32M.

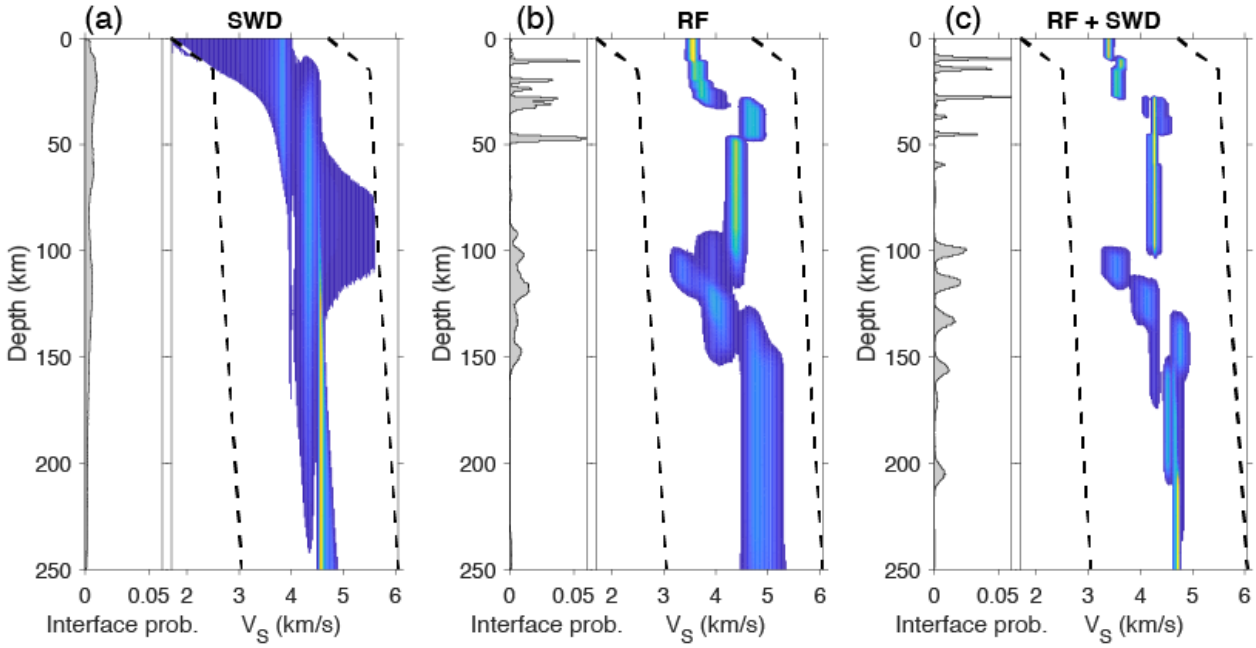

Figure D.4: (a) SWD-only, (b) RF-only and (c) joint inversion for station P33M.

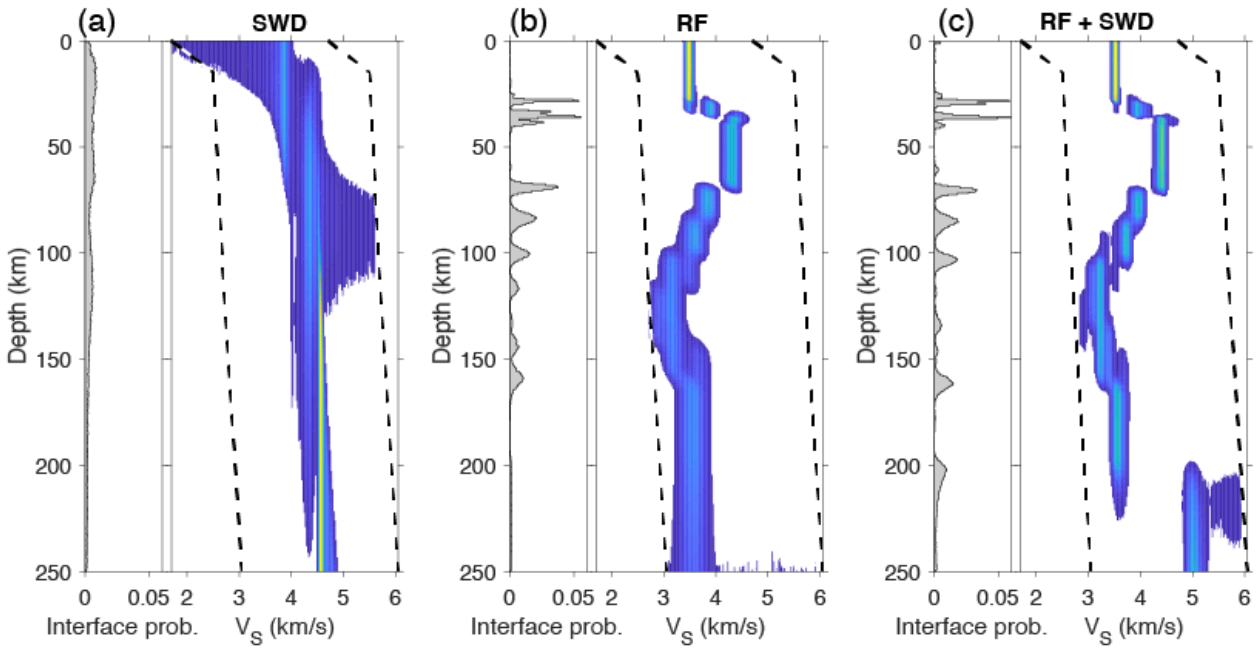

Figure D.5: (a) SWD-only, (b) RF-only and (c) joint inversion for station R33M.

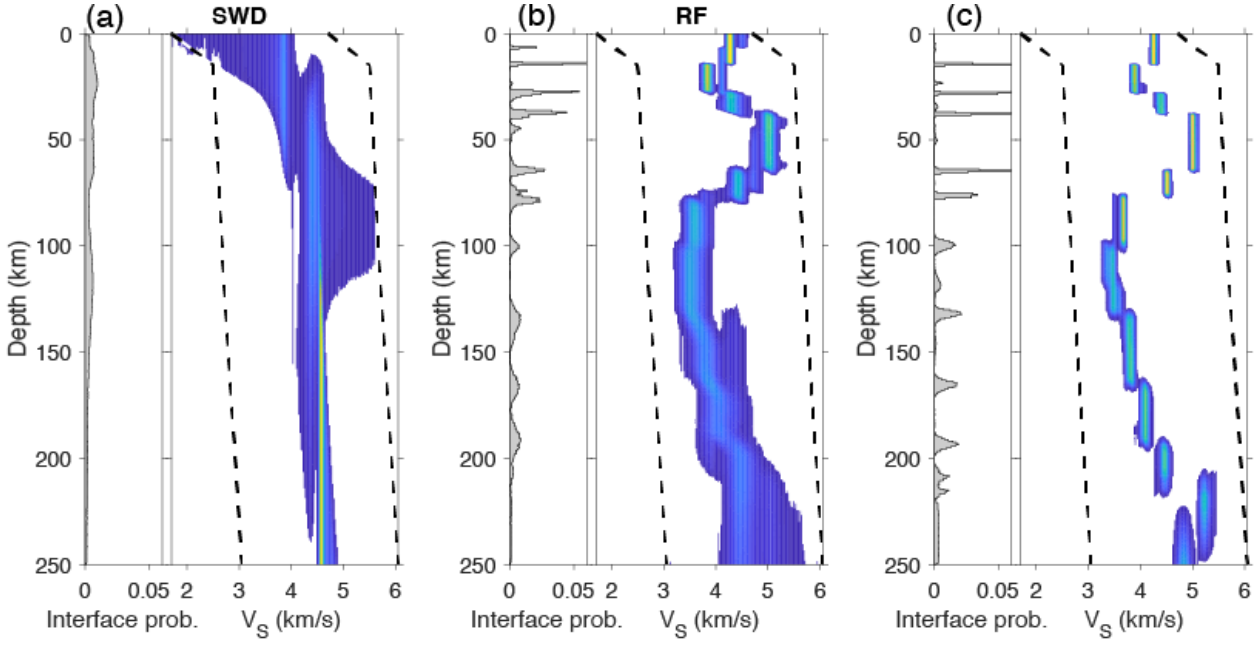

Figure D.6: a) SWD-only, b) RF-only and c) joint inversion for station WTLY.

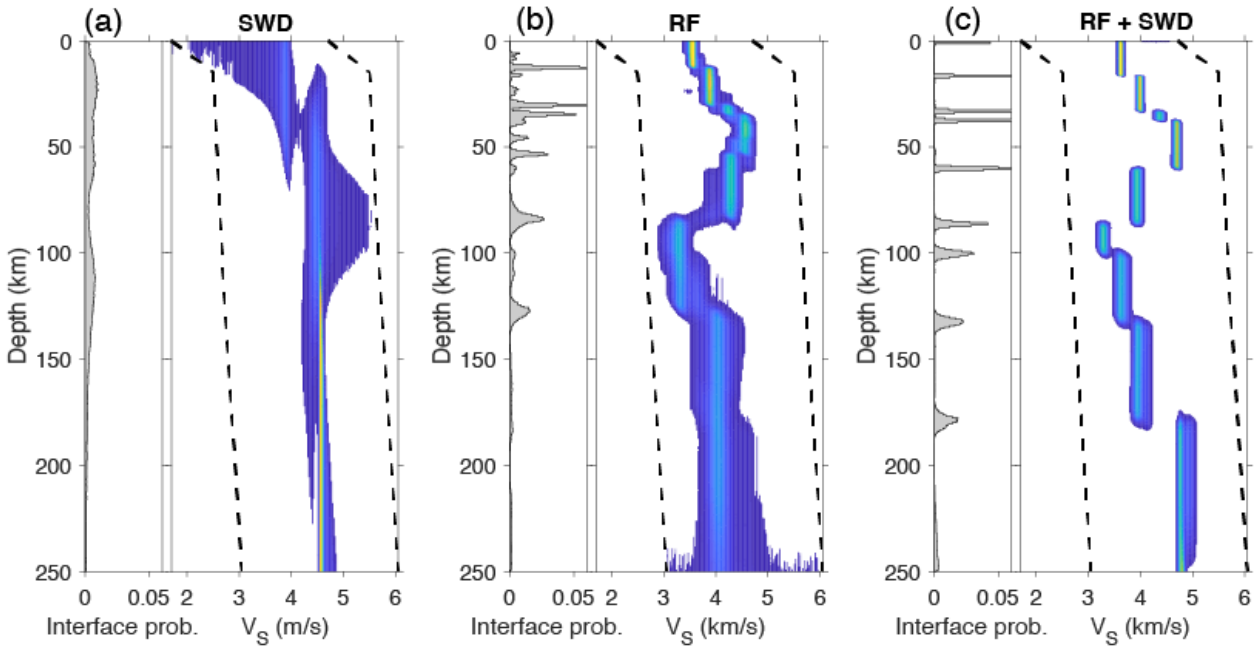

Figure D.7: (a) SWD-only, (b) RF-only and (c) joint inversion for station LIRD.

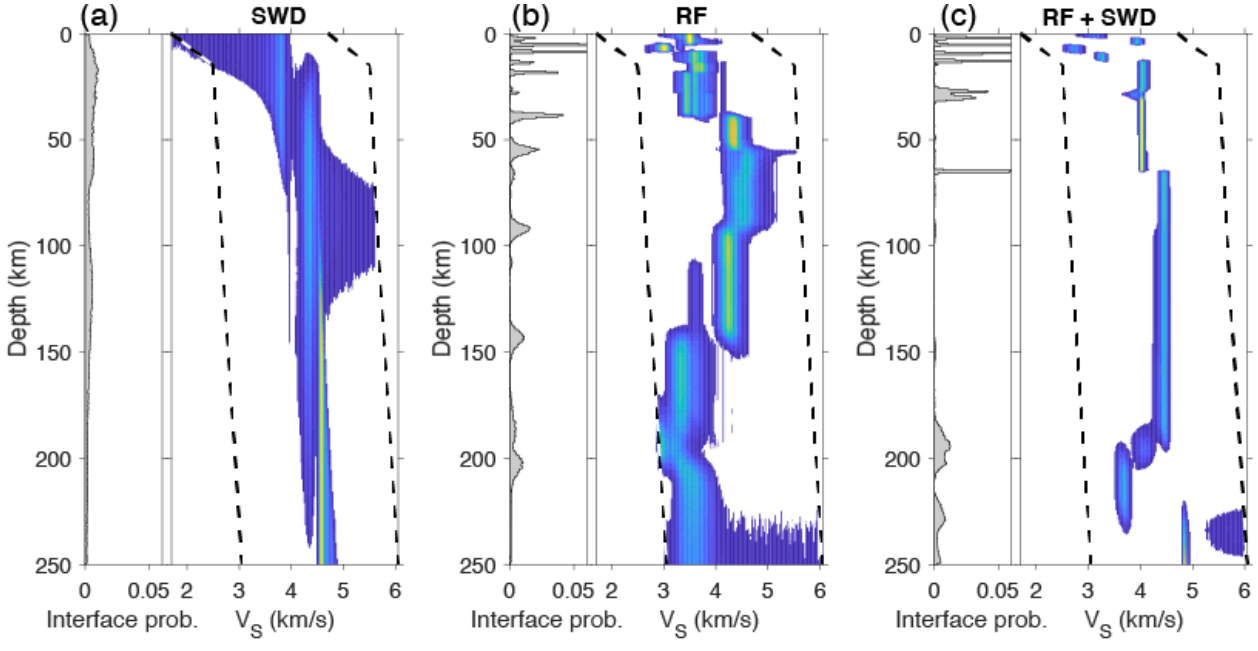

Figure D.8: (a) SWD-only, (b) RF-only and (c) joint inversion for station TOAD.

matrix scaling factors are scaled closer to the minimum bound of 1. Noise parameters for SWD data are more variable. Much like the SWD data fit ranges, the SWD noise parameters are higher for stations with a large LVZ.

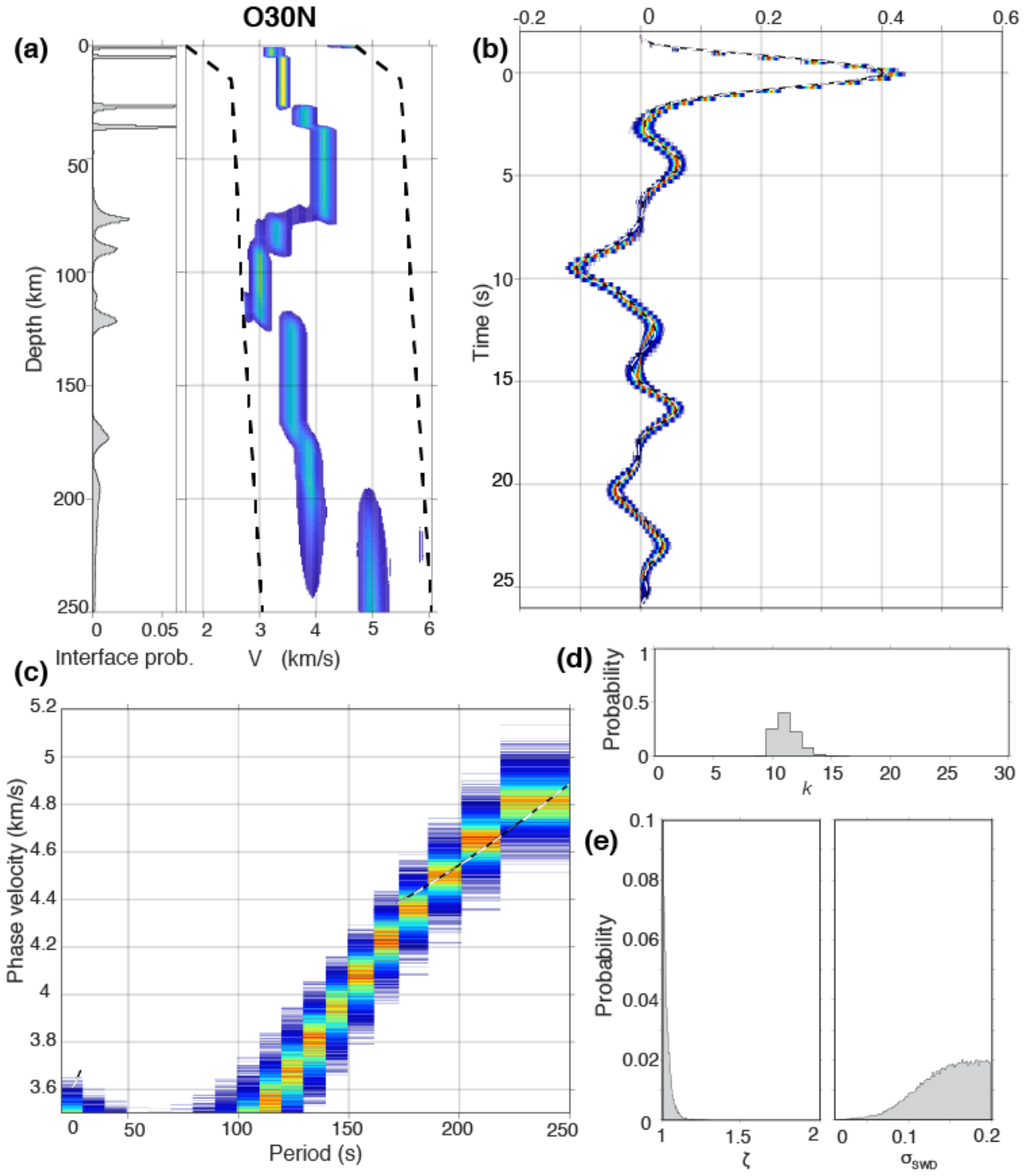

Figure D.9: (a) Marginal density plot for P33M. (b) P-RF predicted data from inversion PPD. (c) SWD predicted data from PPD. (d) Probability of number of layer nodes. (e) Scaling factor for R-component and standard deviation for SWD data.

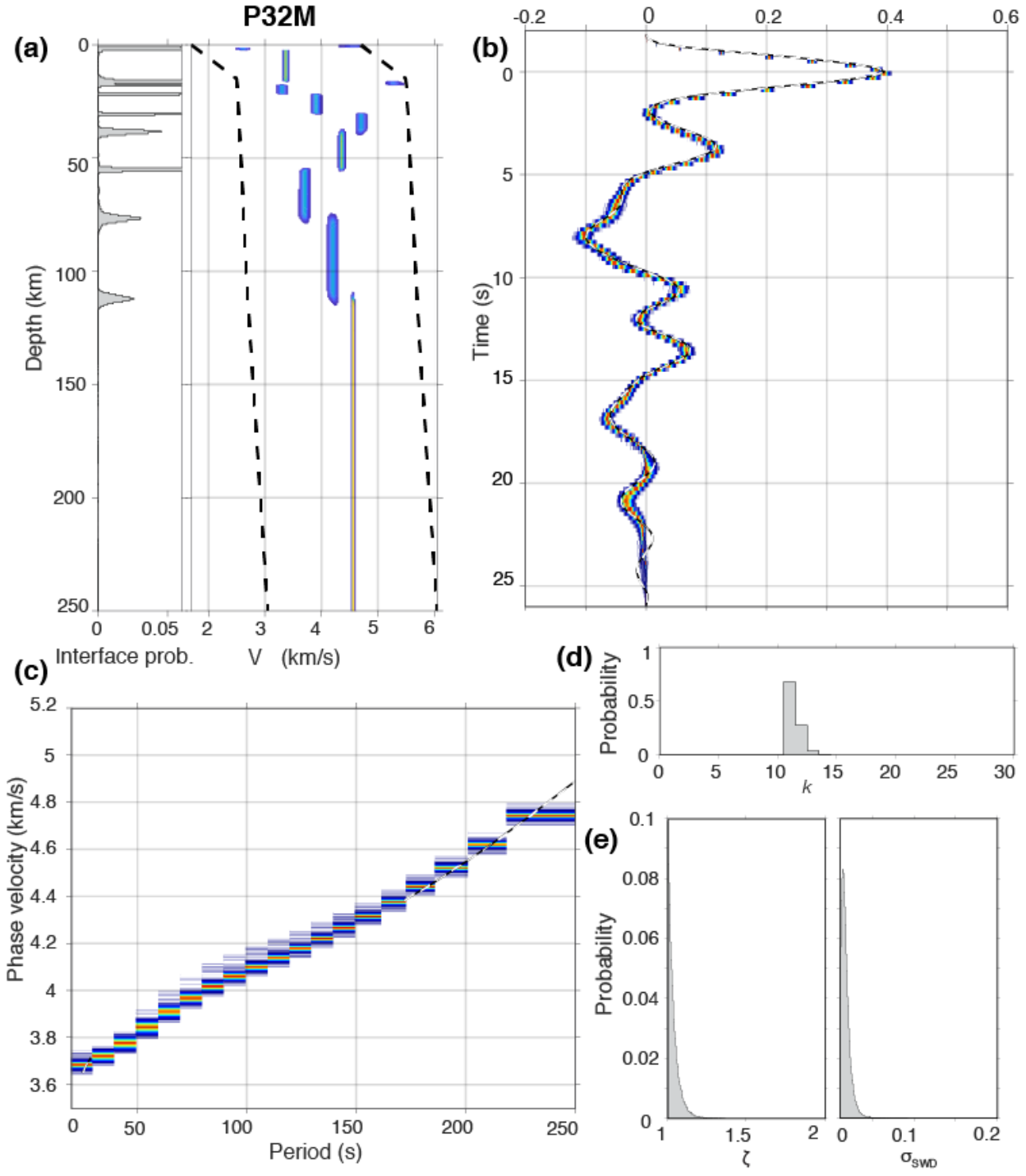

Figure D.10: a) Marginal density plot for P32M; b) RF predicted data from inversion PPD. c) SWD predicted data from PPD. d) Probability of number of layer nodes. e) Scaling factor for R-component and standard deviation for SWD data.

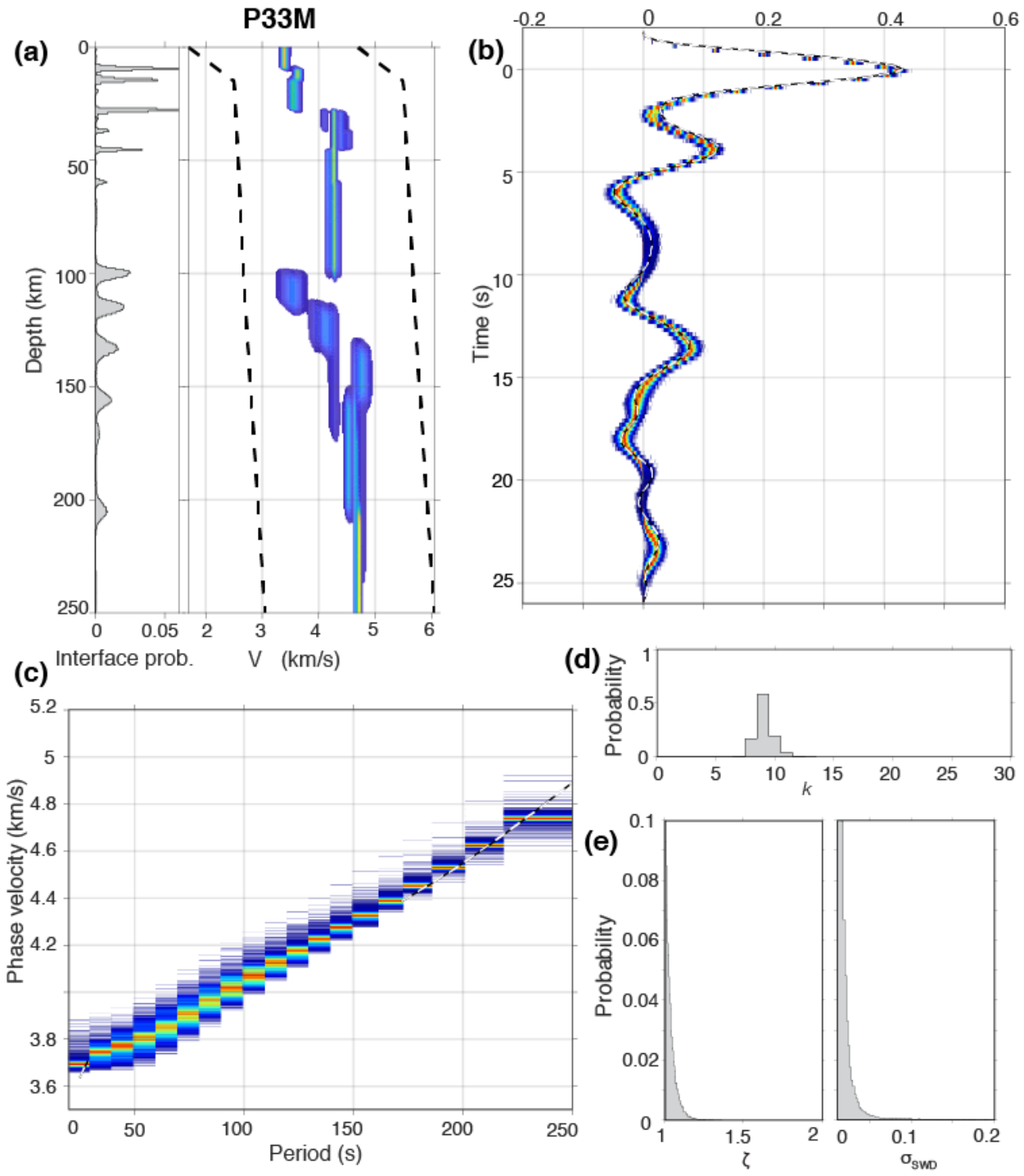

Figure D.11: (a) Marginal density plot for P33M. (b) P-RF predicted data from inversion PPD. (c) SWD predicted data from PPD. (d) Probability of number of layer nodes. (e) Scaling factor for R-component and standard deviation for SWD data.

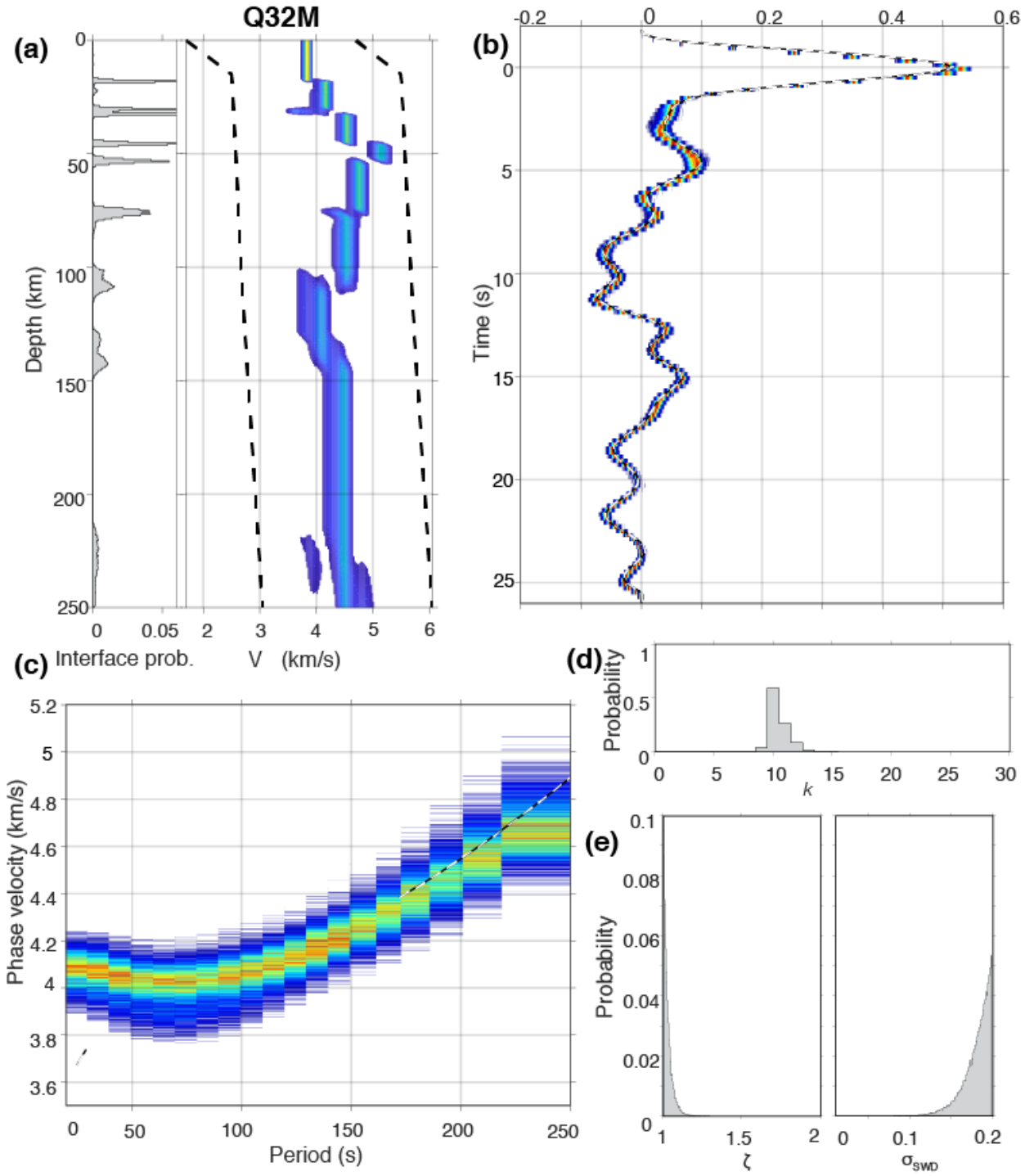

Figure D.12: (a) Marginal density plot for Q32M. (b) P-RF predicted data from inversion PPD. (c) SWD predicted data from PPD. (d) Probability of number of layer nodes. (e) Scaling factor for R-component and standard deviation for SWD data.

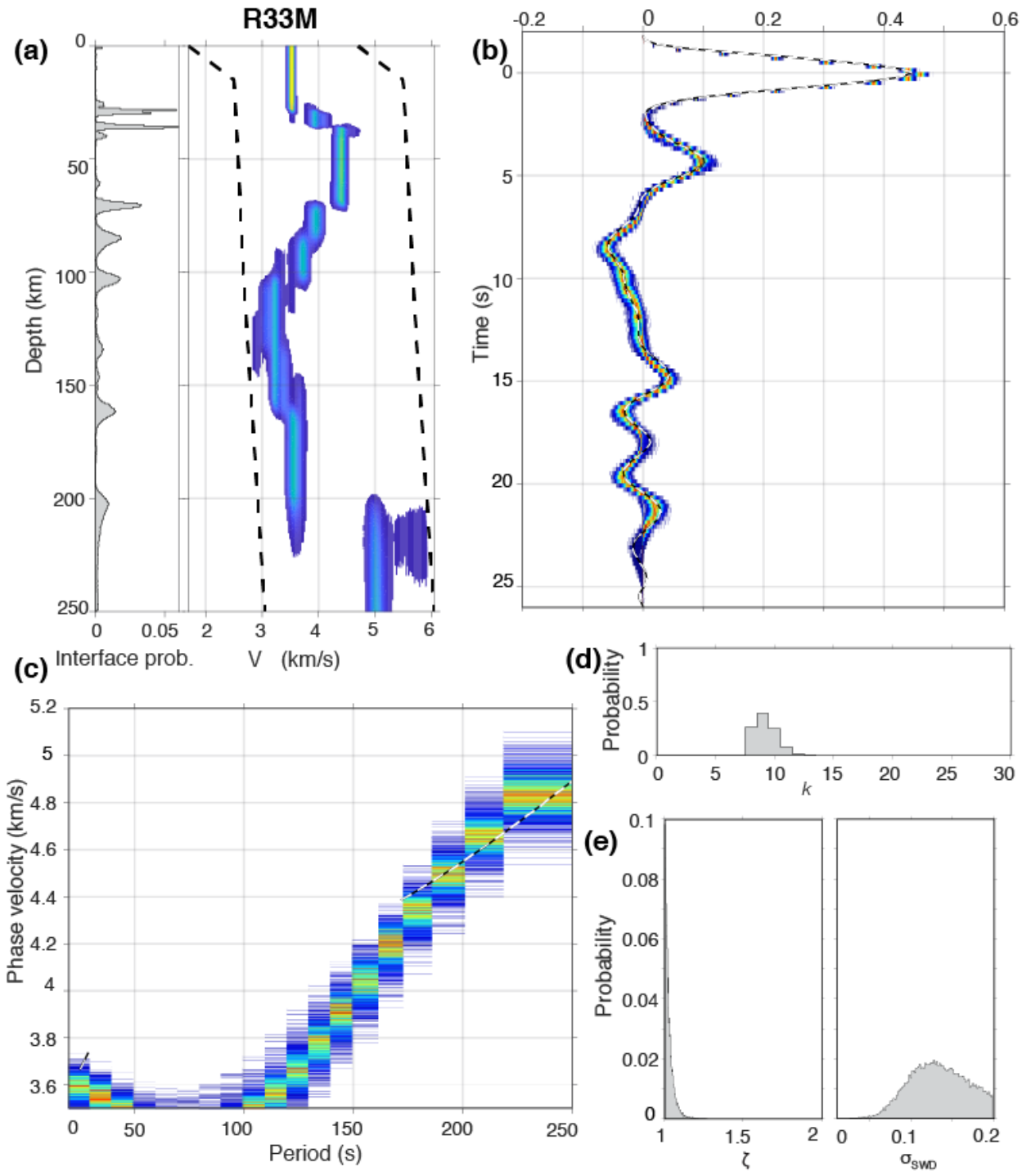

Figure D.13: (a) Marginal density plot for R33M. (b) P-RF predicted data from inversion PPD. (c) SWD predicted data from PPD. (d) Probability of number of layer nodes. (e) Scaling factor for R-component and standard deviation for SWD data.

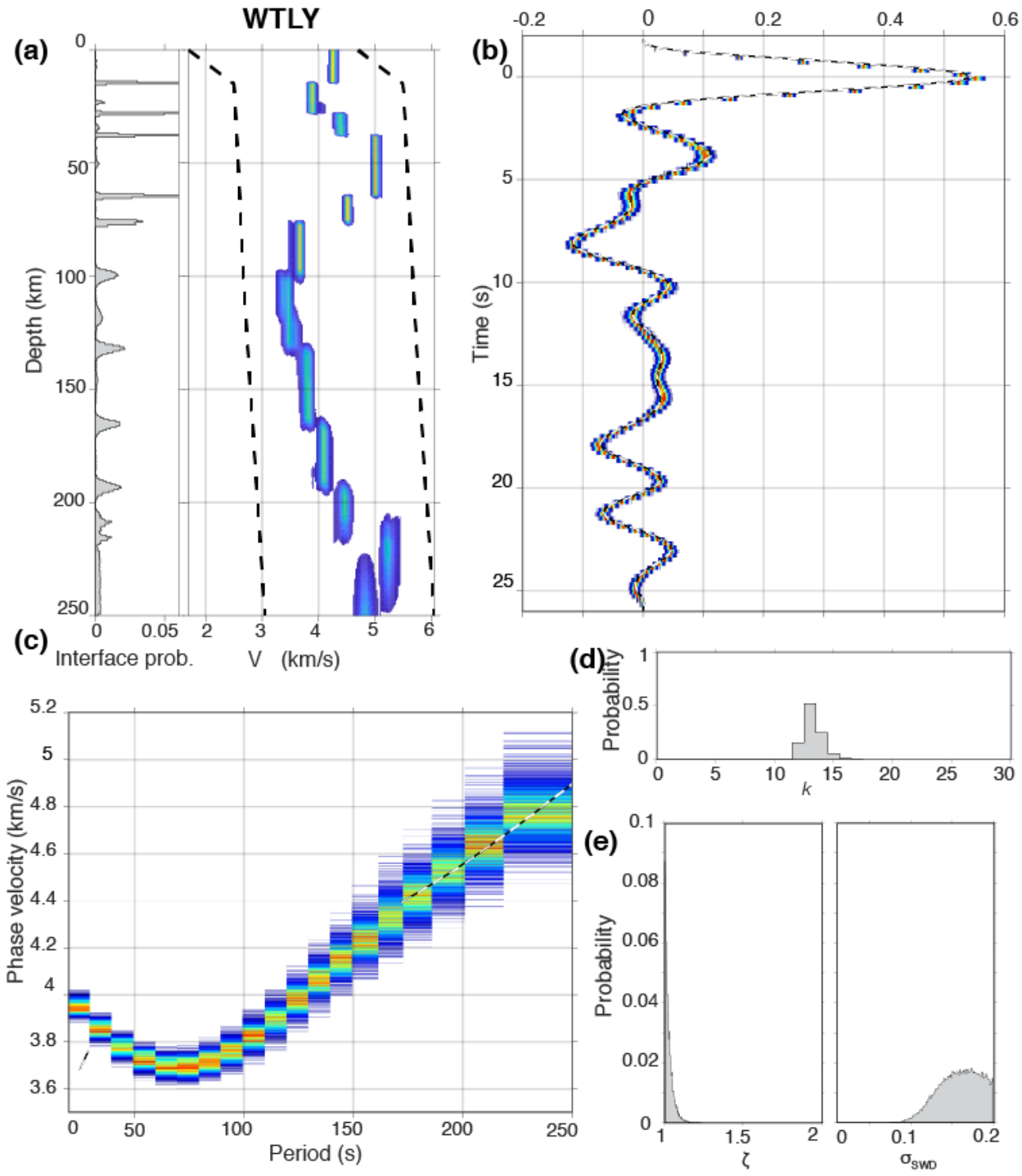

Figure D.14: a) Marginal density plot for WTLY. b) RF predicted data from inversion PPD. c) SWD predicted data from PPD. d) Probability of number of layer nodes. e) Scaling factor for R-component and standard deviation for SWD data.

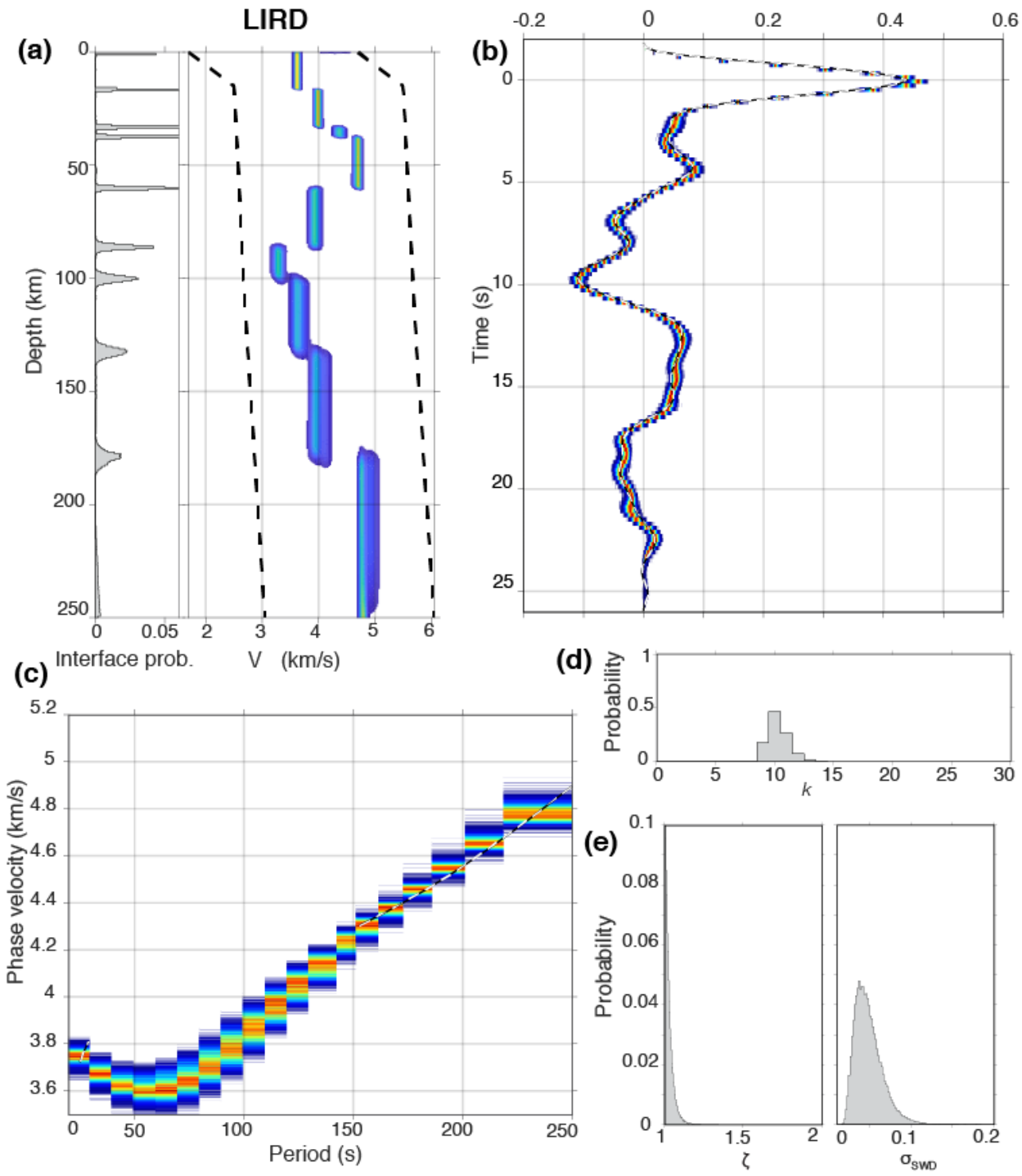

Figure D.15: (a) Marginal density plot for LIRD. (b) P-RF predicted data from inversion PPD. (c) SWD predicted data from PPD. (d) Probability of number of layer nodes. (e) Scaling factor for R-component and standard deviation for SWD data.

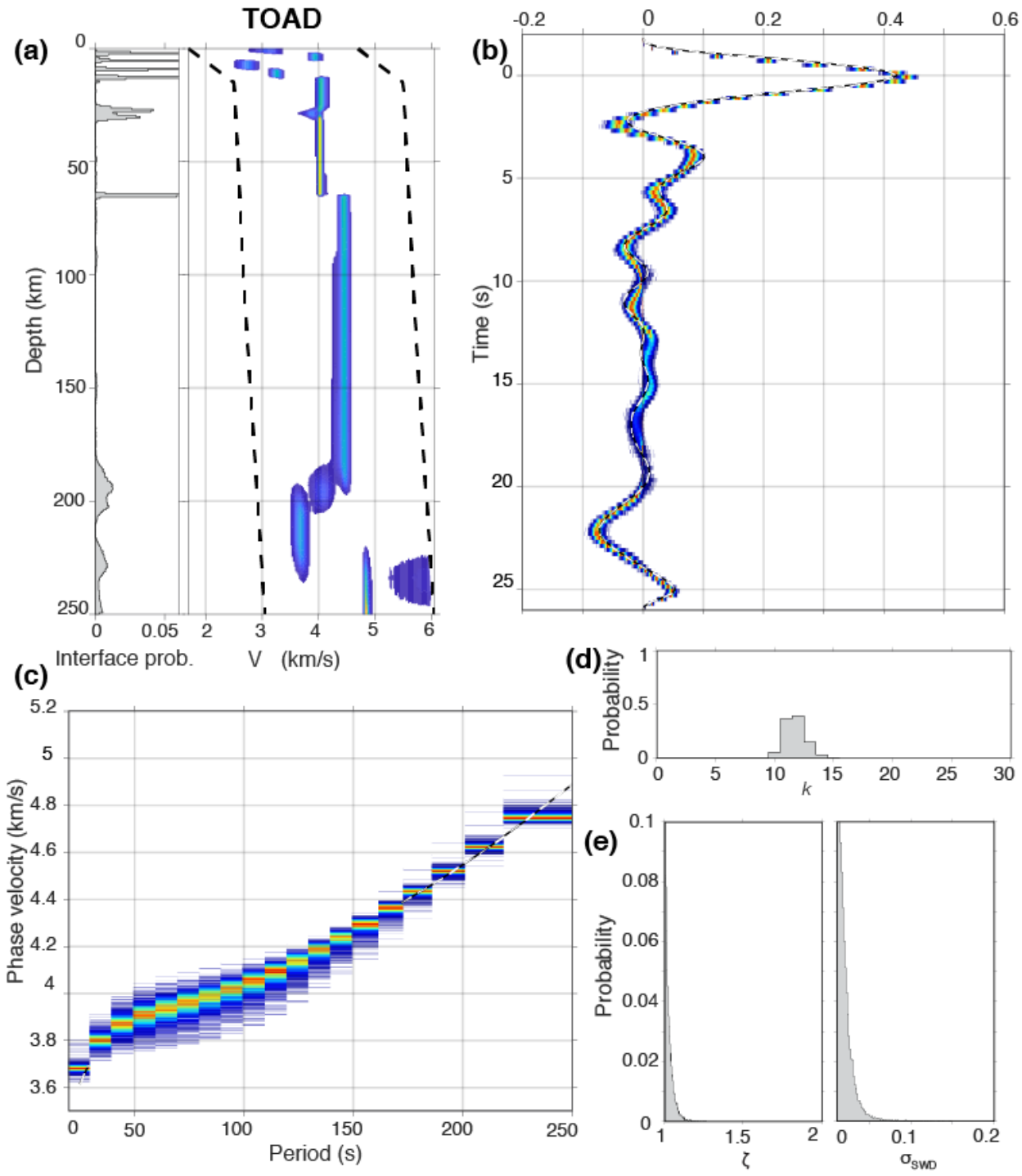

Figure D.16: (a) Marginal density plot for TOAD. (b) RF predicted data from inversion PPD. (c) SWD predicted data from PPD. (d) Probability of number of layer nodes. (e) Scaling factor for R-component and standard deviation for SWD data.

E. The TNA and SNA Models

V_s reductions and partial melt calculations are estimated in this thesis using the tectonically active North American (TNA) model (Grand and Helmberger, 1984). This model accounts for upper mantle heterogeneity in North America for two distinct regions by determining shear wave velocities from S and SS phase waveforms and travel times (Grand and Helmberger, 1984). The V_s models are determined for the tecontically active region and stable shield (SNA) of North America. These models are described in Appendix E and Figure E.1.

V_s in the TNA is about 10% lower than the shield for depths 38 to 170 km. Both models exhibit lids with lower velocities beneath, indicating the LAB. This TNA decrease in V_s is shallower ($\sim$ 75 km) than in the SNA ($\sim$ 150 km). V_s remain lower in the TNA than the SNA until the two models merge around the 410-discontinuity. Overall, these difference between the TNA and SNA models emphasizes the large lateral heterogeneity within North America, and how the selection of the TNA model in partial melt estimates is a more accurate calculation of V_s reduction.

Depth (km)	TNA (km/s)	SNA (km/s)	Depth (m)	Velocity (km/s)
0.0	3.200	3.200	500.0	5.190
4.0	3.200.	3.200	525.0	5.240
4.0	3.650	3.650	550.0	5.290
6.0	3.657	3.657	575.0	5.345
16.0	3.690	3.690	600.0	5.395
30.0	3.690	3.730	625.0	5.445
34.0	3.690	3.750	659.0	5.500
38.0	4.400	4.800	660.0	5.910
50.0	4.350	4.790	675.0	5.980
75.0	4.320	4.775	700.0	6.050
100.0	4.290	4.775	725.0	6.130
125.0	4.290	4.775	750.0	6.200
150.0	4.320	4.775	775.0	6.220
175.0	4.350	4.710	800.0	6.240
200.0	4.390	4.630	825.0	6.260
225.0	4.430	4.640	850.0	6.275
250.0	4.470	4.670	875.0	6.290
275.0	4.510	4.695	900.0	6.305
300.0	4.570	4.720	925.0	6.320
325.0	4.630	4.740	950.0	6.335
350.0	4.680	4.755	975.0	6.350
375.0	4.730	4.765	1000.0	6.365
405.0	4.780	4.780	1025.0	6.385
406.0	5.000	5.000	1050.0	6.405
425.0	5.050	5.050		
450.0	5.090	5.090		
475.0	5.140	5.140		

Table E.1: Reference models of V_S at depths 0- to 1050-km depth for TNA and SNA (Grand and Helmberger, 1984)

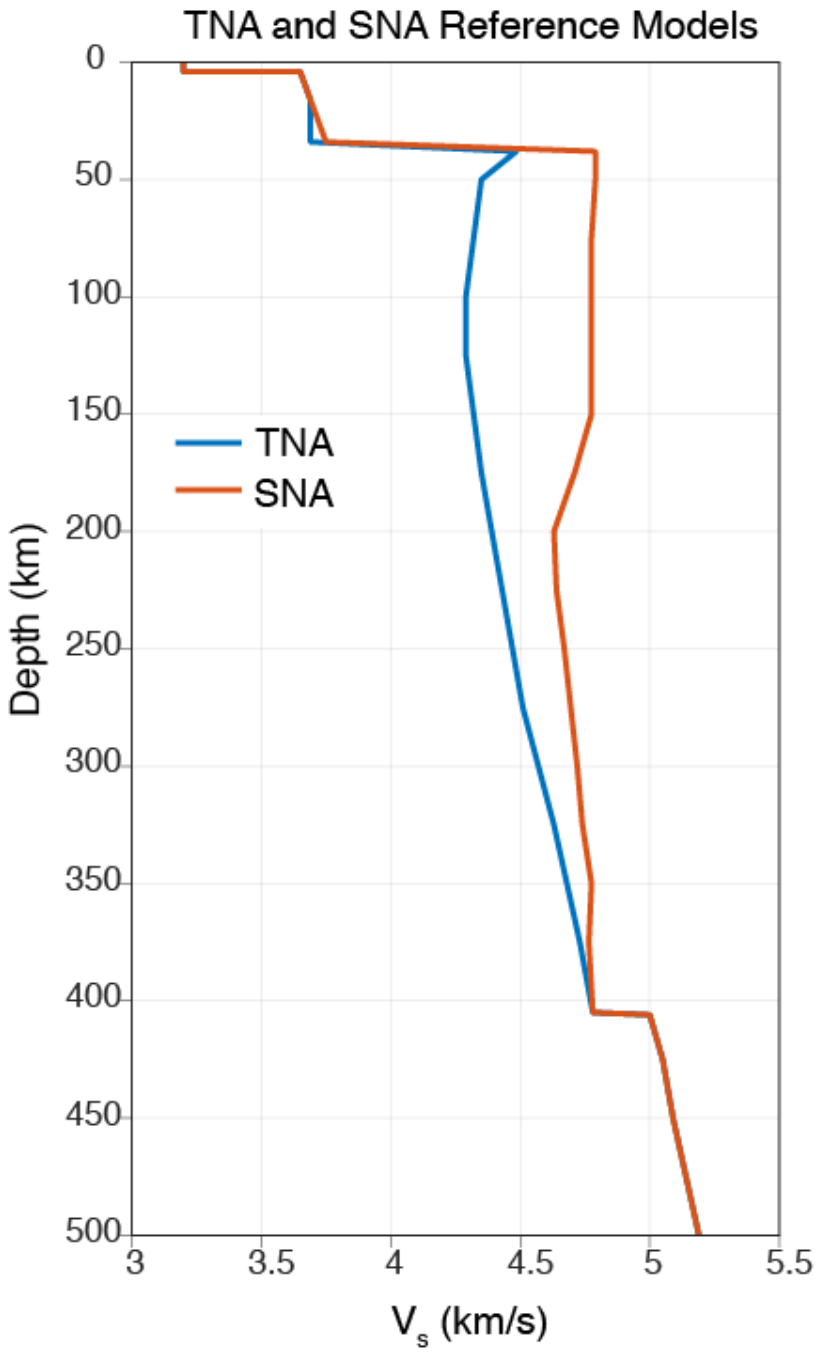

Figure E.1: Comparison of earth reference V_s models TNA and SNA (Grand and Helmberger, 1984).

F. Melt Fraction Estimates

Partial melt estimates for $\hat{\alpha}$ values of 0.2, 0.26, and 0.4 are obtained by digitizing curves representing V_s dependence on melt fraction in Chantel et al. (2016). The Chantel et al. (2016) study fit experimental data with the relationship

$$V_s = 0065\varphi^2 - 0.5565\varphi + 4.4211, \tag{F.1}$$

where φ is the melt fraction. Because the measurements are preformed at megahertz frequencies, effects of anelasticity at seismic frequencies must be considered. Significant anelastic relaxation occurs when the frequency of an elastic wave is shorter or equal the the characteristic frequency of relaxation (Karato, 1993; Jackson et al., 2002). The empirical relationships for all $\hat{\alpha}$ values were not included in Chantel et al. (2016), so I obtain these relationships by digitizing points along the curves for the range of $\hat{\alpha}$. I fit a 2nd order polynomial to fit the points, with coefficients shown in Table F.1. The resulting curves are plotted against the curves determined in Chantel et al. (2016) in Figure F.1. For the ranges of partial melt fractions that I calculate, these fits are sufficient.

$\hat{\alpha}$	a	b	c
N/A	0.0617	-0.5371	4.4300
0.4	0.0881	-0.5371	4.4355
0.26	0.1089	-0.8129	4.4339
0.2	0.1234	-0.9016	4.4240

Table F.1: Coefficients for second-order polynomials for a $\hat{\alpha}$ values of 0.4, 0.26, 0.2 determined by digitizing curves from Chantel et al. (2016).

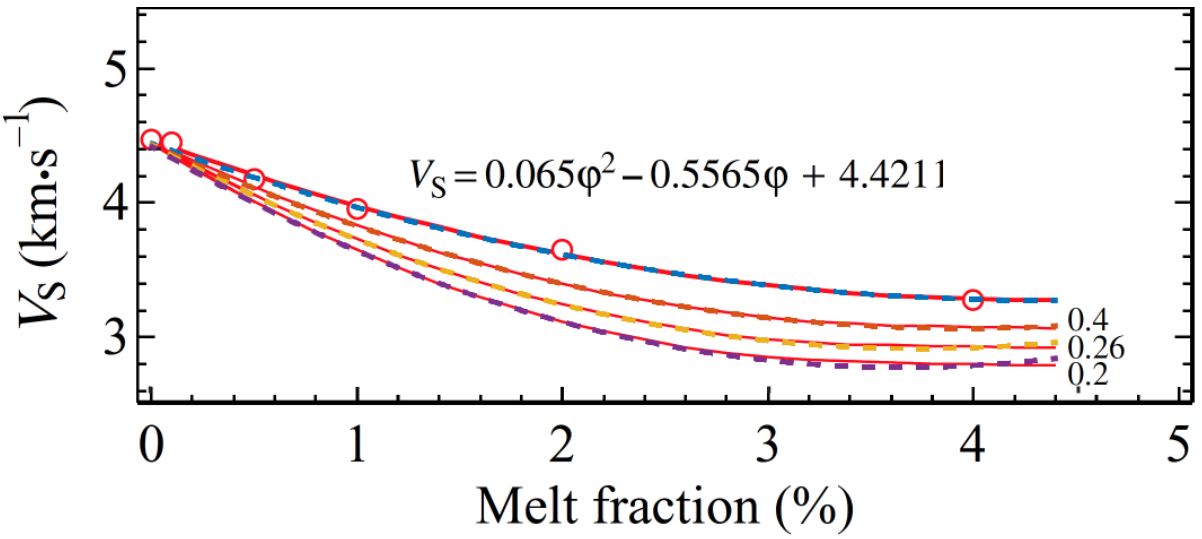

Figure F.1: Dependence of V_s on melt fraction adapted from Chantel et al. (2016). Red dots represent experiment data and the red line connecting the red dots represents the associated fit, represented by the corresponded equation. The dashed, purple line atop this line is my digitized estimate of the experimental results. The lines below these lines represent modelled values that include anelastic effects expected for the seismic waves at high temperatures for $\hat{\alpha}$ values of 0.2, 0.26, and 0.4. Like the first line, the solid red lines are calculations from Chantel et al. (2016), and the dashed lines represent my digitized estimates.

Bibliography

Abraham, A., Francis, D., and Polve, M., 2001, Recent alkaline basalts as probes of lithospheric roots of the Northern Canadian Cordillera: Chemical Geology, **175**, 361–381.

Abt, D., Fischer, K., French, S., Ford, H., Yuan, H., and Romanowicz, B., 2010, North American lithospheric discontinuity structure imaged by Ps and Sp receiver functions: J. Geophys. Res., **115**, no. B09301.

Agostinetti, N., and Bodin, T., 2018, Flexible coupling in joint inversions: a Bayesian structre decoupling alogirthm: J. Geophys. Res. Solid Earth, **123**, 8798–8826.

Agostinetti, P. N., and Malinverno, A., 2010, Receiver function inversion by trans-dimensional Monte Carlo sampling: Geophys. J. Int., **181**, no. 2, 858–872.

Aizawa, Y., Barnhoorn, A., Faul, U., Fitzgerald, J., Jackson, I., and Kovács, I., 2008, Seismic properties of Anita Bay dunite: an exploratory study of the influence of water: J. Petrol., **49**, 841–855.

Ammon, C., Randall, G., and Zandt, G., 1990, On the nonuniqueness of receiver function inversions: J. Geophys. Res., **95**, no. 10, 15303–15318.

Ammon, C., 1991, The isolation of receiver effects from teleseismic P waveforms: Bull. Seis. Soc. Am., **81**, no. 6, 2504–2510.

Armitage, J., Ferguson, D., Goes, S., Hammond, J., Calais, E., Rychert, C., and Harmon, N., 2015, Upper mantle temperature and the onset of extension and break-up in Afar, Africa: Earth Planet. Sci. Lett., **418**, 78–90.

Audet, P., Jellinek, A., and Uno, H., 2007, Mechanical controls on the deformation of continents at convergent margins: Earth Planet. Sci. Lett., **264**, no. 1–2, 151–161.

Audet, P., Bostock, M., Mercier, J., and Cassidy, J., 2008, Morphology of the Explorer–Juan de Fuca slab edge in northern Cascadia: Imaging plate capture at a ridge-trench-transform triple junction: Geology, **36**, no. 11, 895–898.

Audet, P., Currie, C., Schaeffer, A., and Hill, A., 2019, Seismic evidence for lithospheric thinning and heat in the northern Canadian Cordillera: Geophys. Res. Lett., **46**, 4249–4257.

Bao, X., Eaton, D., and Guest, B., 2014, Plateau uplift in western Canada caused by lithospheric delamination along a craton edge: Nat. Geosci., **7**, 830–833.

Bao, X., Eaton, D., and Gu, Y., 2016, Rayleigh wave rayleigh wave azimuthally anisotropic phase velocity maps beneath western Canada: J. Geophys. Res. Solid Earth, **121**, 1821–1834.

Batir, J., and Blackwell, D., 2019, Thermal evolution of the Northern Cordilleran Volcanic Province: implications for heat flow in remnant back-arc regions: Int. Geol. Rev.

B.C. and Yukon Geological Survey. Major Faults:, 2015. ArcMap layer obtained from Yukon government's Corporate Spatial Warehouse.

———. Terranes:, 2015. ArcMap layer obtained from Yukon government's Corporate Spatial Warehouse.

Bernstein, S., Kelemen, P., and Hanghøj, K., 2007, Consistent olivine Mg in cratonic mantle reflects Archean mantle melting to the exhaustion of orthopyroxene: Geology, **35**, no. 5, 459–462.

Bianchi, I., Park, J., Piana Agostinetti, N., and Levin, V., 2010, Mapping seismic anisotropy using harmonic decomposition of receiver functions: An application to Northern Apennines, Italy: J. Geophys. Res., **115**, no. B12317.

Birch, F., 1960, The velocity of compressional waves in rocks to 10 kilobars, part 1: J. Geophys. Res., **65**, no. 4, 1083–1102.

Blum, J., Hanson, D., and Rosenblatt, J., 1963, On the Central Limit Theorem for the Sum of a Random Number of Independent Random Variables: Z. Wahrscheinlichkeitstheorie, **1**, 389–393.

Bodin, T., Sambridge, M., Takli, H., Arroucau, P., Gallagher, K., and Rawlinson, N., 2012, Transdimensional inversion of receiver functions and surface wave dispersion: J. Geophys. Res., **117**, B02301.

Bodin, T., Leiva, J., Romanowicz, B., Maupin, V., and Yuan, H., 2016, Imaging anisotropic layering with Bayesian inversion of multiple data types: Geophys. J. Int., **206**, 605–629.

Bond, G., and Kominz, M. A., 1984, Construction of tectonic subsidence curves for the early Paleozoic miogeocline, southern Canadian Rocky Moutnains: implications for subsidence mechanisms age of breakup, and crustal thinning: Geol. Soc. Am. Bull., **95**, 155–173.

Bostock, M., 1998, Mantle stratigraphy and evolution of the Slave province: J. Geophys. Res., **103**, no. B9, 21183–21200.

Bostock, M., 2007, Theory and observations-teleseismic body-wave scattering and receiver-side structure *in* Romanowicz, B., and Dziewonski, A., Eds., Seismology and Structure of the Earth:: Elesvier, 219–246.

Boyd, F., 1989, Compositional distinction between oceanic and cratonic lithosphere: Earth Planet. Sci. Lett., **96**, 15–26.

Canil, D., Virgo, D., and Scarfe, C., 1990, Oxidation state of mantle xenoliths from British Columbia, Canada: Contrib. to Miner. and Pet., **104**, 453–462.

Cassidy, J., 1992, Numerical experiments in broadband receiver function analysis: Bull. Seis. Soc. Am., **823**, 1453–1474.

Cassidy, J., 1995, A comparison of the receiver structure beneath stations of the Canadian National Seismograph Network: Can. J. Earth Sci., **32**, 938–951.

Chantel, J., Manthilake, G., Andrault, D., Novella, D., Y, T., and Wang, Y., 2016, Experimental evidence supports mantle partial melting in the asthenosphere: Sci. Adv., **2**, no. 5.

Chen, M., Niu, F., Liu, Q., and Tromp, J., 2015, Mantle-driven uplift of Hangai Dome: New seismic constraints from adjoint tomography: Geophys. Res. Lett., **42**, 6967–6974.

Chen, Y., Gu, Y., Currie, C., Johnsston, S., Hung, S., Schaeffer, A., and Audet, P., 2019, Seismic evidence for a mantle suture and implications for the origin of the Canadian Cordillera: Nature Comm., **10**, 2449.

Chen, L., 2009, Lithospheric structure variations between the eastern and central North China Craton from S- and P- receiver function migration: Phys. of Earth and Planet. Int., **173**, 216–227.

Cline II, C., Faul, U., David, E., Berry, A., and Jackson, I., 2018, Redox-influenced seismic properties of uppermantle olivine: Nature, **555**, 355–358.

Clowes, R., Hammer, P., Fernández-Viejo, G., and Welford, J., 2005, Lithospheric structure in northwestern canada from lithoprobe seismic refraction and related studies: a synthesis: Can. J. Earth Sci., **42**, 1277–1293.

Cobden, L., Trampert, J., and Fichtner, A., 2018, Insights on upper mantle melting rheology and anelastic behavior from seismic shear wave tomography: Geochemistry, Geophysics, Geosystems, **19**, 3892–3916.

Colpron, M., Nelson, J., and Murphy, D., 2007, Northern Cordilleran terranes and their interactions through time: GSA Today, **17**, no. 4/5, 4–10.

Comeau, M., Käufl, J., und A. Kuvshinov, M. B., Grayver, A., Kamm, J., Demberel, S., Sukhbaatar, U., and Batmagnai, E., 2018, Evidence for fluid and melt generation in response to an asthenospheric upwelling beneath the Hangai Dome, Mongolia: Earth Planet. Sci. Lett., **487**, 201–209.

Cook, F., Clowes, R., Snyder, D., van der Velden, A., Hall, K., Erdmer, P., and Evenchick, C., 2004, Precambrian crust beneath the Mesozoic northern Canadian Cordillera discovered by Lithoprobe seismic reflection profiling: Tectonics, page 23.

Cook, F., Hall, K., and Lynn, C., 2005, The edge of northwestern North America at ∼1.8 GA: Can. J. Earth Sci., **42**, 983–997.

Courtier, A., Gahetry, J., Revenaugh, J., Bostock, M., and Garnero, E., 2010, Seismic anisotropy associated with continental lithosphere accretion beneath the CANOE array, northwestern Canada: Geology, **38**, no. 10, 887–890.

Crampin, S., Stephen, R., and McGonigle, R., 1982, The polarization of P-waves in anisotropic media: Geophys. J. R. astr. Soc., **68**, 477–485.

Currie, C., and Hyndman, R., 2006, The thermal structure of subduction zone back arcs: J. Geophys. Res., **111**, (B08404).

Dalton, C., Gaherty, J., and Courtier, A., 2011, Crustal VS structure in northwestern Canada: Imaging the Cordillera-craton transition with ambient noise tomography: J. Geophys. Res., **116**, B12315.

Dettmer, J., and Dosso, S., 2012, Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains: J. Acc. Soc. Am, **132**, no. 4, 2339–2250.

Dettmer, J., Dosso, S., and Holland, C., 2007, Uncertainty estimation in seismo-acoustic reflection time inversion: J. Acc. Soc. Am., **122**, 161–176.

Dettmer, J., Dosso, S., and Holland, C., 2009, Model selection and Bayesian inference for high-resolution seabed and reflection inversion: J. Acc. Soc. Am, **125**, 706–716.

Dettmer, J., Dosso, S., and Holland, C., 2010, Trans-dimensional geoacoustic inversion: J. Acoust. Soc. Am., **128**, no. 6, 3393–3405.

Dettmer, J., Molnar, S., Steininger, G., Dosso, S., and Cassidy, J., 2012, Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models: Geophys. J. Int., **188**, 719–734.

Dettmer, J., Dosso, S., Bodin, T., Stipčević, J., and Cummins, P. R., 2015, Direct-seismogram inversion for receiver-side structure with uncertain source-time functions: Geophys. J. Int., **203**, 1373–1387.

DiCaprio, L., Maiti, T., Dettmer, J., and Eaton, D., 2020, Moho Structure Across the Backarc-Craton Transition in the Northern U.S. Cordillera: Tectonics, **39**, no. e2019TC005489.

Dosso, S., Dettmer, J., Steininger, G., and Holland, C., 2014, Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation: Inverse Problems, **30**, no. 11.

Dziewonski, A., and Anderson, D., 1981, Preliminary reference Earth model: Phys. Earth Plan. Int., **25**, 297–356.

Earl, D., and Deem, M., 2005, Parallel tempering: Theory, applications, and new perspectives: Phys. Chem., **7**, 3910–3916.

Eaton, D., Darbyshire, F., Evans, R., Grütter, H., Jones, A. G., and Yuan, X., 2009, The elusive lithosphere–asthenosphere boundary (LAB) beneath cratons: Lithos, **109**, 1–22.

Eaton, D., 2005, Multi-genetic origin of the continental Moho: insights from Lithoprobe: TerraNova, **18**, 34–43.

Edwards, B., and Russell, J., 2000, Distribution, nature, and origin of Neogene–Quaternary magmatism in the northern Cordilleran volcanic province, Canada: GSA Bulletin, **112**, no. 8, 1280–1295.

Ekström, G., 2011, A global model of Love and Rayleigh surface wave dispersion and anistropy, 25-250s: Geophys. J. Int., **187**, 1668–1686.

Estéve, C., Audet, P., Schaeffer, A., Schutt, D., Aster, R., and Cubley, J., 2020, The Upper Mantle Structure of Northwestern Canada from Teleseismic Body Wave Tomography: J. Geophys. Res.: Solid Earth, **125**.

2005, The seismological signature of temperature and grain size variations in the upper mantle: Earth Planet. Sci. Lett., **234**, no. 1-2, 119–134.

Faul, U., 2001, Melt retention and segregation beneath mid-ocean ridges: Nature, **410**, 920–923.

Fischer, K., Ford, H., Abt, D., and Rychert, C., 2010, The lithosphere-asthenosphere boundary: Ann. Rev. Earth Planet. Sci., **38**, 551–575.

Flück, P., Hyndman, R., and Lowe, C., 2003, Effective elastic thickness Te of the lithosphere in western Canada: J. Geophys. Res., **108**.

Francis, D., Minarik, W., Proenza, Y., and Shi, L., 2010, An overview of the Canadian Cordilleran lithospheric mantle: Can. J. Earth Sci, **47**, 353–368.

Francis, D., 1987, Mantle-melt interaction recorded in spinel lherzolite xenoliths from the Alligator Lake Volcanic Complex, Yukon, Canada: J. of Petrology, **28**, no. 3, 569–597.

Frederiksen, A., and Bostock, M., 2000, Modelling teleseismic waves in dipping anisotropic structures: Geophys. J. Int., **141**, 401–412.

Frederiksen, A., Bostock, M., VanDecar, J., and Cassidy, J., 1998, Seismic structure of the upper mantle beneath the northern Canadian Cordillera from teleseismic travel-time inversion: Tectonophysics, **294**, 43–55.

Gaherty, J., and Revenaugh, J. Collaborative Research: Canadian Northwest Seismic Experiment:, 2003.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D., and Nicolson, H., 2016, Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry: Geophys. J. Int., **208**, 36–58.

Geological Survey Of Canada. Canadian National Seismograph Network:, 1989.

Geyer, C., and Møller, J., 1994, Simulation procedures and likelihood inference for spatial point processes: Scand. J. Stat., **21**, no. 4, 359–373.

Geyer, C., 1991, Markov Chain Monte Carlo maximum likelihood: Comp. Sci. and Stat., Proceedings from the 23rd Symposium on the Interface, pages 156–163.

Grand, S., and Helmberger, D., 1984, Upper mantle shear structure of North America: Geophys. J. R. astr. Soc, **76**, 399–438.

Grasby, S., Allen, D., Bell, S., Chen, Z., Ferguson, G., Jessop, A., Kelman, M., Ko, M., Marjorowicz, J., Moore, M., and nd R. Therrien, J. R., 2012, Geothermal Energy Resource Potential of Canada: Geological Survey of Canada Open File 6914, page 301P.

Green, D., and Ringwood, A., 1963, Mineral assemblages in a model mantle composition: J. Geophys. Res., **68**, no. 3, 937–945.

Green, P., 1995, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination: Biometrika, **82**, no. 4, 711–732.

Hammond, W., and Humphreys, E., 2000a, Upper mantle seismic attenuation:Effects of realistic partial melt distribution: J. Geophys. Res., **105**, no. B5, 10,987–10,999.

——— 2000b, Upper mantle seismic wave velocity: Effects of realistic partial melt geometries: J. Geophys. Res., **105**, no. B5, 10975–10986.

Harder, M., and Russell, J., 2006, Thermal state of the upper mantle beneath the Northern Cordilleran Volcanic Province (NCVP), British Columbia, Canada: Lithos, **87**, 1–22.

Hastings, W., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 711–732.

Hayward, N., 2015, Geophysical investigation and reconstruction of lithospheric structure and its control on geology, structure, and mineralization in the Cordillera of northern Canada and eastern Alaska: Tectonics, **34**, 2165–2189.

Hirschmann, M., 2010, Partial melt in the oceanic low velocity zone: Phys. Earth Plan. Int., **179**, 60–71.

Hirth, G., and Kohlstedt, D., 1996, Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere: Earth Planet. Sci. Lett., **144**, 93–108.

Hoffman, P., 1988, United Plates of America, the birth of a craton: Early Proterozoic assembly and growth of Laurentia: Ann. Rev. Earth Planet. Sci., **16**, 543–603.

Hyndman, R., Currie, C., and Mazzotti, S., 2005, Subduction zone backarcs, mobile belts, and orogenic heat: GSA Today, **15**, no. 2, 4–10.

Hyndman, R., 2010, The consequences in elevation and recent tectonics of Canadian Cordillera thermal regime: a review: Can. J. Earth Sci., **47**, 621–632.

Hyndman, R., 2017, Lower-crustal flow and detachment in the North American Cordillera: a consequence of Cordillera-wide high temperatures: Geophys. J. Int., **209**, 1779–1799.

IRIS Transportable Array. USArray Transportable Array:, 2003.

Jackson, I., Gerald, J. F., Faul, U., and Tan, B., 2002, Grain size-sensitive wave attenuation in polycrystalline olivine: J. Geophys. Res., **107**, no. B12:2360.

Johnston, S., 2001, The Great Alaska Terrane Wreck: reconciliation of paleomagnetic and geologic data in the northern Cordillera: Earth Planet. Sci. Lett., **193**, 259–272.

Johnston, S., 2008, The Cordilleran Ribbon Continent of North America: Ann. Rev. Earth Planet. Sci., **36**, 495–530.

Jones, D., Howell, D., Coney, P., and Monger, J., 1983, Recognition, character and analysis of tectono-stratigraphic terranes in western North America *in* Hashimoto, M., and Uyeda, S., Eds., Accretion Tectonics in the Circum-Pacific Region:: Terra, 21–35.

Juliá, J., Ammon, C., Herrmann, R., and Correig, A., 2000, Joint inversion of receiver functions and surface wave dispersion observations: Geophys. J. Int., **143**, 99–112.

Karato, S., 1993, Importance of anelasticity in the interpretation of seismic tomography: Geophys. Res. Lett., **20**, no. 15, 1623–1626.

Karato, S., 1995, Effects of water on seismic wave velocities in the upper mantle: Proc. Japan Acad, **71**, no. B, 61–66.

Kawakatsu, H., Kumar, P., Takei, Y., Shinohara, M., Kanazawa, T., Araki, E., and Suyehiro, K., 2009, Seismic evidence for sharp lithosphere-asthenosphere boundaries of oceanic plates: Science, **324**, no. 5926, 499–502.

Kay, S., and Ramos, V., 2007, Backbone of the Americas: From Patagonia to Alaska—A Super Rock Star Event: GSA Today, **17**.

Kennett, B., and Engdahl, E., 1991, Traveltimes for global earthquake location and phase identification: Geophys. J. Int., **105**, 429–465.

Kickuchi, M., and Kanamori, H., 1982, Inversion of complex body waves: Bull. Seis. Soc. Am., **72**, 491–506.

Kilgore, M., Peslier, A., Brandon, A., and Lamb, W., 2018, Water and oxygen fugacity in the lithospheric mantle wedge beneath the Northern Canadian Cordillera (Alligator Lake): Geochm., Geophys, and Geosys., **19**, 3844–3869.

Killingbeck, S., Livermore, P., Booth, A., and West, L., 2018, Multimodal layered transdimensional inversion of seismic dispersion curves with depth constraints: Geochem, Geophys, Geosys, **19**, 4957–4971.

Kim, S., and Rhie, J., 2019, Calculation of station-representative isotropic receiver functions: Pure and Appl. Geophys., **176**, 2367–2383.

Kind, R., Yuan, X., and Kumar, P., 2012, Seismic receiver functions and the lithosphere-asthenosphere boundary: Tectonophysics, **536-537**, 25–43.

Langmuir, C., Bezos, A., Escrig, S., and Parman, S., 2006, Chemical systematics and hydrous melting of the mantle in back-arc basins in D. M. Christie and C. R. Fisher (Eds.): Back-arc spreading systems: Geological, biological, chemical and physical interactions, Geophysical Monograph, **166**, 87–146.

Langston, C., 1977, The effect of planar dipping structure on source and receiver responses for constant ray parameter: Bull. Seis. Soc. Am., **67**, no. 4, 1029–1050.

Langston, C., 1979, Structure under Mount Rainier, Washington, inferred from teleseismic body waves: J. Geophys. Res., **84**, no. B9, 4749–4762.

Ledo, J., and Jones, A., 2005, Upper mantle temperature determined from combining mineral composition, electrical conductivity laboratory studies and magnetotelluric field observations: Application to the intermontane belt, Northern Canadian Cordillera: Earth Planet. Sci. Lett., **236**, 258–268.

Lee, C., 2003, Compositional variation of density and seismic velocities in natural peridotites at STP conditions: implications for seismic imaging of compositional heterogeneties in the upper mantle: J. Geophys. Res., **108**, no. B9, 2441.

Lekić, V., Matas, J., Panning, M., and Romanowicz, B., 2009, Measurement and implications of frequency dependence of attenuation: Earth Planet. Sci. Lett., **282**, no. 1–4, 285–293.

Lewis, T., Hyndman, R., and Flück, P., 2003, Heat flow, heat generation, and crustal temperatures in the northern Canadian Cordillera: Thermal control of tectonics: J. Geophys. Res., **108**, no. 2316.

Li, L., and Weidner, D., 2013, Effect of dynamic melting on acoustic velocities in a partially molten peridotite: Phys. Earth Plan. Int., **222**, 1–7.

Ligorria, J., and Ammon, C., 1999, Iterative deconvolution and receiver-function estimation: Bull. Seis. Soc. Am., **99**, no. 5, 1395–1400.

Lithgow-Bertelloni, C., and Silver, P., 1998, Dynamic topography, plate driving forces and the african superswell: Nature, **395**, 269—272.

Madsen, J., Thorkelson, D., Friedman, R., and Marshall, D., 2006, Cenozoic to Recent plate configurations in the Pacific Basin: Ridge subduction and slab window magmatism in western North America: Geosphere, **2**, no. 1, 11–34.

Malinverno, A., 2000, A Bayesian criterion for simplicity in inverse problem parameterization: Geophys. J. Int., **140**, 267–285.

Malinverno, A., 2002, Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem: Geophys. J. Int., **151**, no. 3, 675–688.

McLellan, M., Schaeffer, A., and Audet, P., 2018, Structure and fabric of the crust and uppermost mantle in the northern Canadian Cordillera from Rayleigh-wave tomography: Tectonophysics, pages 28–41.

Menke, W., 2012, Geophysical Data Analysis: Discrete Inverse Theory: Elsevier, 3 edition.

Mercier, J., Bostock, M., Cassidy, J., Dueker, K., Gaherty, J., Gareno, E., Revenaugh, J., and Zandt, G., 2009, Body-wave tomography of western Canada: Tectonophysics, **475**, 480–492.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Equation of state calculations by fast computing machines: J. Chem. Phys., pages 1087–1092.

Monger, J., and Gibson, H., 2019, Mesozoic-Cenozoic deformation in the Canadian Cordillera: The record of a“Continental Bulldozer”?: Tectonophysics, **757**, 153–169.

Monger, J., and Price, R., 2002, The Canadian Cordillera: Geology and tectonic evolution: CSEG Recorder, **27**, 17–36.

Monger, J., Price, R., and Tempelman-Kluit, D., 1982, Tectonic accretion and the origin of the two major metamorphic and plutonic welts in the Canadian Cordillera: Geology, **10**, 70–75.

Murphy, B., and Egbert, G., 2019, Synthesizing seemingly contradictory seismic and magnetotelluric observations in the southereastern United States to image physical properties of the lithosphere: Geochemistry, Geophysics, Geosystems, **20**, 2606–2625.

Nelson, J., and Colpron, M., 2007, Tectonics and metallogeny of the British Columbia, Yukon and Alaskan Cordillera, 1.8 GA to the present: Goodfellow, W.D., ed., Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods: Geological Association of Canada, Mineral Deposits Division, Special Publication, , no. 5, 755–791.

Oldow, J., Bally, A., and Lallemant, H., 1990, Transpression, orogenic float, and lithospheric balance: Geology, **18**, 991–994.

Plank, T., and Forsyth, D., 2016, Thermal structure and melting conditions in the mantle beneath the Basin and Range province from seismology and petrology: Geochem. Geophys. Geosyst., **17**, 1312—-1338.

Polat, A., Frei, R., Longstaffe, F., Thorkelson, D., and Friendman, E., 2018, Petrology and geochemistry of the Tasse mantle xenoliths of the Canadian Cordillera: A record of archean to quaternary mantle growth, metasomatism, removal, and melting: Tectonophysics, **737**, 1–26.

Rényi, A., 1963, On the central limit theorem for the sum of a random number of independent random variables: Acta Mathaematica Academiae Scientiarum Hungarica, **11**, no. 1–2, 97–102.

Rippe, D., Unsworth, M., and Currie, C., 2013, Magnetotelluric constraints on the fluid content in the upper mantlebeneath the southern Canadian Cordillera:Implications for rheology: J. Geophys. Res. Solid Earth, **118**, 5601–5624.

Romanowicz, B., and Mitchell, B., 2007, Deep Earth structure Q of the Earth from crust to core *in* Schubert, G., Ed., Treatise on geophysics:: Elesvier, 731–774.

Rychert, C., Laske, G., Harmon, N., and Shearer, P., 2013, Seismic imaging of melt in a displaced Hawaiian plume: Nature Geosci., **6**.

Saito, M., 1988, DISPER80: A subroutine package for the calculation of seismic normal-mode solution: Seismological Algorithms Computational Methods and Computer Programs, pages 239–319.

Sakamaki, T., Suzuki, A., Ohtani, E., Terasaki, H., Urakawa, S., Katayama, Y., Funakoshi, K., Wang, Y., Hernlund, J., and Ballmer, M., 2013, Ponded melt at the boundary between the lithosphere and asthenosphere: Nature Geoscience, **6**, 1041–1044.

Sambridge, M., Gallagher, K., Jackson, A., and Rickwood, P., 2006, Trans-dimensional inverse problems, model comparison and the evidence: Geophys. J. Int., **167**, 528–542.

Sambridge, M., 2014, A parallel tempering algorithm for probabilistic sampling and multimodal optimization: Geophys. J. Int., **196**, no. 1, 357–374.

Sato, H., Sacks, I., and Murase, T., 1989, The Use of Laboratory Velocity Data for Estimating Temperature and Partial Melt Fraction in the Low-Velocity Zone' Comparison With Heat Flow and Electrical Conductivity Studies : J. Geophys. Res., **94**, no. B5, 5689–5704.

Schaeffer, A., and Bostock, M., 2010, A low-velocity zone atop the transition zone in northwestern Canada: J. Geophys. Res., **115**, no. B06302.

Schaeffer, A., and Lebedev, S., 2014, Imaging the North American continent using waveform inversion of global and USArray data: Earth Planet. Sci. Lett., **402**, 26–41.

Schmerr, N., 2012, The Gutenberg Discontinuity: Melt at the Lithosphere-Asthenosphere Boundary: Science, **335**, 1480–1483.

Schneider, F., Kind, R., Tilman, F., Shapiro, S., Sobel, E., and Yuan, X., 2014, Imaging an intra-continental subduction in Central Asia with teleseismic receiver functions: Ph.D. thesis, Freie Universität Berlin.

Shen, W., Ritzwoller, M., and Schulte-Pelkum, V., 2013, A 3-D model of the crust and uppermost mantle beneath the Central and Western US by joint inversion of receiver functions and surface wave dispersion: J. Geophys. Res., **118**, 262–276.

Shi, L., Francis, D., Ludden, J., Frederiksen, A., and Bostock, M., 1998, Xenolith evidence for lithospheric melting above anomalously hot mantle under the northern Canadian Cordillera: Contributions to Mineralogy and Petrology, **131**, 39–53.

Sleep, N., 2005, Evolution of the continental lithosphere: Ann. Rev. Earth Planet. Sci., **33**, 369–393.

Smith, D., Ritzwoller, M., and Shapiro, N., 2004, Stratification of anisotropy in the Pacific upper mantle: J. Geophys. Res., **109**, no. B11309.

Snyder, D., Roberts, B., and Gordey, S., 2005, Contrasting seismic characteristics of three major faults in northwestern Canada: Can. J. Earth Sci., **42**, 1223–1237.

Takei, Y., 2017, Effects of partial melting on seismic velocity and attenuation: A new insight from experiments: Ann. Rev. Earth Planet. Sci., **45**, 447–470.

Tarantola, A., 2005, Inverse Problem Thoery and Methods for Model Parameter Estimation: Society for Industrial and Applied Mathematics.

Taryoun, A., Audet, P., Mazzotti, S., and Ashoori, A., 2017, Architecture of the crust and uppermost mantle in the northern Canadian Cordillera from receiver functions: J. Geophys. Res. Solid Earth, **122**, 5268–5287.

Thorkelson, D., Madsen, J., and Sluggett, C., 2011, Mantle flow through the Northern Cordilleran slab window revealed by volcanic geochemistry: Geology, **39**, no. 3, 267–270.

University of Ottawa. Yukon-Northwest Seismic Network:, 2013.

van der Heyden, P., 1992, A middle Jurassic to early Tertiary Andean-Sierran arc model for the Coast Belt of British Columbia: Tectonics, **11**, no. 1, 82–97.

van Heijst, H., Snieder, R., and Nowack, R., 1994, Resolving a low-velocity zone with surface-wave data: Geophys. J. Int., **118**, 333–343.

Wang, H., and Currie, C., 2015, Magmatic expressions of continental lithosphere removal: J. Geophys. Res.: Solid Earth, **120**, 7239–7260.

Weidner, D., Li, L., Whitaker, M., and Triplett, R., 2018, Ultrasonic Acoustic Velocities During Partial Melting of a Mantle Peridotite KLB-1: J. Geophys. Res. Solid Earth, **123**, 1252–1261.

White, R., and Mackenzie, D., 1989, Magmatism at rift zones: a generation of volcanic, continental margins and flood basalts: J. Geophys. Res., **94**, no. B6, 7685–7729.

Yang, Y., and Forsyth, D., 2006, Rayleigh wave phase velocities, small-scale convection, and azimuthal anisotropy beneath southern California: J. Geophys. Res., **111**, no. B07306.

Yang, Y., and Forsyth, D., 2008, Attenuation in the upper mantle beneath Southern California: Physical state of the lithosphere and asthenosphere: J. Geophys. Res., **113**, no. B03308.

Yukon Geological Survey. Yukon Observatory Seismic Network:.

Zaporozan, T., Ferderiksen, A., Bryksin, A., and Darbyshire, F., 2018, Surface-wave images of western Canada: lithospheric variations across the Cordillera–craton boundary: Can. J. of Earth Sci., **55**, no. 8, 887–896.

Zelt, C., Ellis, R., and Zelt, B., 2006, Three-dimensional structure across the Tintina strike-slip fault, northern Canadian Cordillera, from seismic refraction and reflection tomography: Geophys. J. Int., **167**, no. 3, 1292–1308.