Geodynamic Tomography

John Keith Magali

N°d'ordre NNT : 2021LYSE1050

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale ED52
(Physique et Astrophysique)

Spécialité de doctorat : Sciences de la Terre
Discipline : Sismologie

Soutenue publiquement le 19/03/2021, par :
John Keith Magali

Tomographie Geodynamique

Devant le jury composé de :

Nom, prénom grade/qualité établissement/entreprise	
Quantin-Nataf Cathy, Professeure des Universités,Université Lyon 1	Présidente du jury
Faccenda Manuele, Professeur Associé, Univeristé de Padoue (Italie)	Rapporteur
Kaus Boris, Professeur, Université de Mayence (Allemagne)	Rapporteur
Garel, Fanny, Maître de Conférences, Université de Montpellier	Examinatrice
Quantin-Nataf Cathy, Professeure des Universités,Université Lyon 1	Examinatrice
Ricard Yanick, Directeur de recherche, CNRS Lyon	Examinateur
Bodin Thomas, Chargé de Recherche, CNRS Lyon	Directeur de thèse

Abstract

Over the last three decades, seismologists have been successful in probing the Earth's internal structure using copious seismic records gathered at its very surface. Widely known as seismic tomography, one of its central goals is to construct three-dimensional elastic models of the Earth using various inversion strategies tailored to the type of seismic data used. A remaining challenge however is the interpretation of the inherent heterogeneities in its elastic structure in terms of several physical properties (*e.g.*, density, chemical/mineralogical composition, temperature) and of the ubiquity of large-scale anisotropy that are crucial to understand plate tectonics and mantle dynamics. For instance, long-wavelength seismic anisotropy observed in tomographic models have often thought to have been caused by crystallographic preferred orientation (CPO) (*i.e.*, the net alignment of intrinsically-anisotropic upper-mantle minerals) due to mantle deformation. Seismic tomography thus provides a great deal of information about the present-day flow in the mantle.

We introduce *Geodynamic Tomography*, a novel approach to the tomographic problem that incorporates geodynamical and petrological constraints to reduce the number of Earth models down to a subset consistent with geodynamic predictions. This approach encompasses several methodologies: mantle flow modeling, texture evolution modeling, thermodynamic modeling, and seismic data inversion into a single self-consistent probabilistic inversion procedure. Ultimately, we aim to retrieve the complete pattern of upper-mantle deformation by inverting seismic data to better understand mantle dynamics from a seismological point of view.

In this work, we mainly focus on the inversion of local anisotropic surface wave phase velocity dispersion measurements since they provide unique constraints to the large-scale anisotropy associated with convective flow in the mantle. Geodynamic tomography addresses one of the most pressing issues of conventional anisotropic surface wave tomography, that is, its inability to resolve the 21-component elastic tensor independently at every location. We formulate the tomographic problem by using geodynamical and petrological constraints to reduce this large number of model parameters. In lieu of inverting for seismic velocities, we parameterize our inverse problem directly in terms of physical quantities governing mantle flow: a temperature field, and a temperature-dependent viscosity field. The full forward problem proceeds as follows: (1) calculation of instantaneous mantle flow using the temperature field, and the viscosity field as inputs, (2) from the flow field and deformation gradient of (1) as inputs, calculation of the induced CPO and seismic anisotropy using a micro-mechanical model for texture evolution, (3) modeling the pressure- and temperature-dependence

of isotropic properties for a given bulk composition using a thermodynamic model, (4) construction of the full elastic tensor given (2) and (3), and finally estimation of azimuthally-varying surface wave dispersion curves.

The non-linearity of the forward problem, the possible existence of trade-offs among the free parameters, and the non-uniqueness of the solution warrants the need to cast the inverse problem in a Bayesian probabilistic framework. The formalism is an ensemble inference approach, where the solution is an ensemble of models representing a posterior probability distribution, accompanied by the uncertainties in each model parameter. In fact, any implicitly-computed variable (*e.g.* deformation and anisotropy) in the forward problem can be recast in terms of a posterior distribution in their respective model space. We efficiently explore the space of candidate Earth models (*e.g.*, temperature field) by employing a Markov chain Monte Carlo (McMC) algorithm.

We demonstrated geodynamic tomography based on two synthetic tests and compared it with an isotropic inversion (*i.e.*, inversion for temperature fields in the absence of deformation-induced anisotropy). The first tests involves the inversion for temperature fields composed of multiple spherical anomalies. We showed how the incorporation of geodynamic constraints outperforms isotropic inversion in every aspect and robustly recovers not only the temperature field but also of the anisotropic structure.

The second test showcases the applicability of the method exhibiting realistic and hence more complex deformation patterns. As such, we considered 3-D instantaneous flows in the upper-mantle induced by subduction. We mimic real, periodically-correlated data by inverting surface wave dispersion measurements at very low noise conditions to retrieve five unknown parameters defining the thermal and the rheological structures of the subduction zone.

The implementation of the method deemed to be a success as both demonstrations exemplified the implicit retrieval of the complete patterns of mantle deformation, and correspondingly, the 21-independent coefficients of the elastic tensor from the inversion of seismic data alone. Geodynamic tomography is therefore a potentially powerful technique to elucidate the structure of the upper mantle, and interpret seismic observations in terms of mantle deformation patterns.

Résumé

Au cours des trois dernières décennies, les sismologues ont réussi à sonder la structure interne de la Terre à l'aide de nombreux enregistrements sismiques recueillis à sa surface. Largement connue sous le nom de tomographie sismique, l'un de ses principaux objectifs est de construire des modèles élastiques tridimensionnels de la Terre en utilisant diverses stratégies d'inversion adaptées au type de données sismiques utilisées. Il reste cependant un défi à relever : l'interprétation des hétérogénéités élastique en termes de propriétés physiques (*e.g.*, densité, composition chimique/minéralogique, température) et de l'omniprésence de l'anisotropie à grande échelle qui sont cruciales pour comprendre la tectonique des plaques et la dynamique du manteau. Par exemple, l'anisotropie sismique à grande longueur d'onde observée dans les modèles tomographiques a souvent été attribuée à l'orientation cristallographique préférentielle (CPO) (*i.e.*, l'alignement net des minéraux intrinsèquement anisotropes du manteau supérieur) due à la déformation du manteau. La tomographie sismique fournit donc une grande quantité d'informations sur le flux actuel dans le manteau.

Nous présentons *Tomographie Geodynamique*, une nouvelle approche du problème tomographique qui intègre des contraintes géodynamiques et pétrologiques afin de réduire le nombre de modèles terrestres à un sous-ensemble conforme aux prévisions géodynamiques. Cette approche englobe plusieurs méthodologies : la modélisation de l'écoulement du manteau, la modélisation de l'évolution de la texture, la modélisation thermodynamique et l'inversion des données sismiques en une seule procédure d'inversion probabiliste autoconsistante. En fin de compte, nous visons à retrouver le schéma complet de la déformation du manteau supérieur en inversant les données sismiques pour mieux comprendre la dynamique du manteau d'un point de vue sismologique.

Dans ce travail, nous nous concentrons principalement sur l'inversion des mesures de dispersion de vitesse de phase des ondes de surface, car elles fournissent des contraintes uniques à l'anisotropie à grande échelle associée à l'écoulement convectif dans le manteau. La tomographie géodynamique aborde l'un des problèmes les plus pressants de la tomographie anisotrope classique des ondes de surface, à savoir son incapacité à résoudre le tenseur élastique à 21 composantes indépendamment à chaque endroit. Nous formulons le problème tomographique en utilisant les contraintes géodynamiques et pétrologiques pour réduire ce grand nombre de paramètres du modèle.

Au lieu d'inverser pour les vitesses sismiques, nous paramétrons notre problème inverse directement en termes de quantités physiques régissant l'écoulement du manteau : un champ de température, et un champ de viscosité dépendant de la température. Le problème d'inversion se déroule comme suit : (1) calcul de l'écoulement mantélique instantané en utilisant le champ de température et le champ de viscosité comme entrées, (2) à partir du champ d'écoulement et du gradient de déformation de (1) comme entrées, calcul de la CPO induite et de l'anisotropie sismique en utilisant un modèle micro-mécanique pour l'évolution de la texture, (3) modélisation de la dépendance en pression et en température des propriétés isotropes pour une composition donnée à l'aide d'un modèle thermodynamique, (4) construction du tenseur élastique complet, et enfin estimation des courbes de dispersion des ondes de surface variant en azimut.

La non-linéarité du problème, l'existence possible de 'trade-offs' entre les paramètres libres et le caractère non unique de la solution justifient la nécessité de placer le problème inverse dans un cadre probabiliste bayésien. Le formalisme est une approche, où la solution est un ensemble de modèles représentant une distribution de probabilité, accompagnée des incertitudes dans chaque paramètre du modèle. En fait, toute variable calculée implicitement (*e.g.* déformation et anisotropie) dans le problème peut être refondue en termes de distribution postérieure. Nous explorons efficacement l'espace des modèles terrestres candidats (*e.g.*, champ de température) en employant un algorithme de Monte Carlo à chaîne de Markov (McMC).

Nous avons fait la démonstration de la tomographie géodynamique à partir de deux tests synthétiques et l'avons comparée à une inversion isotrope (*i.e.*, inversion pour les champs de température en l'absence d'anisotropie induite par la déformation). Le premier test concerne l'inversion pour les champs de température composés de multiples anomalies sphériques. Nous avons montré comment l'incorporation de contraintes géodynamiques surpasse l'inversion isotrope à tous les égards et permet de récupérer de manière robuste non seulement le champ de température mais aussi la structure anisotrope.

Le second test montre l'applicabilité de la méthode en présentant des modèles de déformation réalistes et donc plus complexes. Ainsi, nous avons considéré les flux instantanés en 3D dans le manteau supérieur induits par la subduction. Nous imitons des données réelles, périodiquement corrélées, en inversant les mesures de dispersion des ondes de surface dans des conditions de très faible bruit pour récupérer cinq paramètres inconnus définissant les structures thermiques et rhéologiques de la zone de subduction.

La mise en œuvre de la méthode a été considérée comme un succès car les deux démonstrations ont permis de récupérer implicitement les schémas complets de déformation du manteau et, par conséquent, les 21 coefficients indépendants du tenseur élastique à partir de la seule inversion des données sismiques. La tomographie géodynamique est donc une technique potentiellement puissante pour élucider la structure du manteau supérieur et interpréter les observations sismiques en termes de modèles de déformation du manteau.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Thomas Bodin, for the freedom he gave me to carry out my research, for continuously motivating me to persevere in the most difficult times, for his guidance on the writing of this thesis, and for supporting me throughout my doctoral journey. Thank you for being one of the figures I look up to!

I am grateful for Université de Lyon and CNRS for the opportunity. This research was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 716542. The provision of the computational resources are made possible through the ERC grant - TRANSCALE: Reconciling Scales in Global Seismology project. All computations were carried out through the in-house Transcale cluster situated in Lyon.

I would like to use this opportunity to thank my collaborators, Navid Hedjazian, Henri Samuel, and Suzanne Atkins for the constructive suggestions and invaluable discussions. I am also indebted to Yanick Ricard and Yann Capdeville, that even though our joint efforts to study some aspects of Seismology were momentary, our academic discourse yielded fruitful outcomes.

I also wish to thank my LGL-TPE colleagues, Colin, Florian, Chloe, and Victor for the short coffee breaks, and most importantly the warm hospitality in the group. I also appreciate my Lyon 1 friends, Rania, Sofia, and especially Santasri, who unfortunately, is not with us anymore. We miss you always. Thanks to my Filipino friends in Lyon, Alfie, Aileen, Rick, Romel, Rey, Julio, Justine, Irina, Gerone, and the others. Being with you guys felt like I had a slice of the Philippines. Thanks to my friends from Trieste and the Philippines; a big shoutout to Mostafa, Jeo, Jek, Gian, Macky, Uriel, Edlie, JP, Jiggy, Arjay and James for the fun times we spent together remotely.

This doctoral programme marks the end of my education, and I would like to express my most sincere appreciation to my parents Vilma and Ken, to my siblings Francis, Danielle, Alica, and Stephanie, and to my grandmother Remy for providing me with unconditional love and support. Finally, I would like to thank my dear Anya, for the continuous love, warmth, and affection. In my best moments, and in my worst, you were always there for me. Being away from my family and loved one is not easy. I dedicate this work to all of you..

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Understanding the dynamical evolution of the earth requires one's knowledge of the interplay between different geophysical processes. Of course, providing the full picture may seem far-fetched; with various length and time scales, and wavelengths illustrating different geophysical scenarios, not to mention the coupling of the systems involved. Such highly complex study produces a web of sub-disciplines that describe how the earth evolves through time. Much has been done with scientists teaming up to study different sub-disciplines; with the goal of fitting these puzzle pieces together to illuminate our preconceived knowledge about the planet. And with the advancement of high performance computing and data science applications such as to big data, the dream of reconciling this multidisciplinary study may be within our grasp and already be a part of our reality. However, this comes with a caveat. The evolutionary trajectory of the earth with time cannot be traced with pure accuracy. Although present-day dynamics can be modeled with computers, tracking the earth history requires field data such as indirect observations fleshed out by dating exhumed rocks and other similar geologic methods.

In this work, we will make a joint effort to lay down the foundation of reconciling what initially seems to be two mutually independent sub-disciplines. Seismology, on one hand, and mantle geodynamics on the other, we will see how a given set of simple 3D structures of the Earth can be retrieved using first-order observations measured at the surface.

We show how geodynamic tomography, though being at its infancy, is a potentially powerful technique to constrain the complete patterns of upper mantle deformation, and thus elucidate some aspects of its interior structure and dynamics.

1.1 Dynamics of the Earth's interior

Before understanding the internal dynamics of the Earth, we first have to define, to first-order, the composition and the structure of the Earth. In the early 1900's, there has been a growing consensus about the inherent heterogeneity of the Earth's structure. Until then, it is not new to the general populace that the Earth is comprised of three major layers with associated compositions and mechanical properties: crust, mantle, and the core. It was not until the late 80's when Dziewonski and Anderson (1981) published a comprehensive list of the Earth's elastic structure where it served as one of the foundations that the Earth is indeed composed of several intricate concentric layers. However, whether these divisions are independent or coupled and to what scale these occur is still a subject of extensive research.

1.1.1 A heterogeneous Earth

The crust is the Earth's outermost layer that is exposed to the atmosphere, and where biological life forms thrive. It is characterised to be cold and brittle as it is comprised mainly of exhumed rocks that congeals as they make their way towards the surface. Moreover, the crust is considered to be orders of magnitude thinner relative to the mantle and the core. Because of this, the crust only makes up a tiny fraction (about 1 %) of the Earth's total mass. This pencil-thin layer of mass is then subdivided into two major domains characterised based on their geological features and formation mechanisms: (1) continental crust, which is on average 30 to 50 km thick, and composed mostly of felsic rocks and (2) oceanic crust, which is on average 5 to 10 km thick, and composed mostly of mafic rocks (Daly, 1940).

Probing the Earth deeper, we arrive in the mantle. The boundary between these two layers is known as the "Moho" discontinuity, named after the Croatian seismologist Andrija Mohorovicic who discovered the discontinuity in 1909 (Mohorovičić, 1909). Unlike the crust, the dynamics of the mantle depends on the time scales considered. At relatively short time scale, the upper mantle behaves like a solid as shown by the transmission of a specific type of seismic wave, known as S-wave or 'shear' waves. Unlike compressional waves (also called $P-$ waves), S-waves cannot propagate in low viscosity liquids. Because of this, the uppermost layer of the mantle, and the crust, make up the lithosphere which up to a certain extent, can be described as rigid. Typical thickness of oceanic lithosphere is about $\sim$ 100 km, whereas continental lithosphere can go as far as 250 km (Watts and Burov, 2003; Steinberger and Becker, 2018). Apart from propagation of elastic waves, another compelling evidence about the existence of this brittle layer is through earthquakes. For this reason, the lithosphere may also be

called the seismogenic layer, that is, the layer at which earthquakes mostly originate due to its brittle nature. However, due to the complexity of earthquake mechanisms and the mechanical properties of the region considered, the length scale of the seismogenic layer may not always be in agreement with that of the lithosphere especially in cratonic regions (Watts and Burov, 2003). While earthquakes most certainly occur within the oceanic mantle lithosphere (Wiens and Stein, 1983), the continental mantle lithosphere is relatively stable, and hence, earthquakes rarely occur in this region. Maggi et al. (2000) argued that such observations are attributed to its thermal structure and the presence of water.

The lithosphere is not a homogeneous shell, rather, they are composed of fragmented pieces, known as plates, resulting from unprecedented tectonic activities that took place over millions of years. Coined as plate tectonics, this theory subsumes the continental drift theory proposed by Alfred Wegener in the early 1900's. In the present, the continental plates move past each other at a constant rate of about few millimeters per year as they glide over the mantle. Such motions are sustained by the convecting mantle underneath as first confirmed by Harry Hess in the late 50's (Hess, 1962). However, surface reconstruction experiments also deduced that plate motion acts as one of the driving forces of upper mantle flow (e.g. Tovish, Schubert, and Luyendyk, 1978). This mutual interaction gave birth to a class of geodynamic flow models called 'passive' mantle flow models. It was not until the late 60's when Morgan (1968) laid the groundwork on the basic hypothesis underlying plate tectonics. This was then followed by a group of geophysicists who independently formulated the kinematics of rigid plate motions (McKenzie and Parker, 1967; McKenzie and Morgan, 1969; Le Pichon, 1968). At this point, plate tectonics was then able to explain the apparent clustering of earthquakes along plate boundaries where surface deformation appears to be the largest (Isacks, Oliver, and Sykes, 1968).

The Earth consists of 9 major plates that are separated by distinctive margins known as plate boundaries. Subduction zones are convergent plate margins that result from the density contrast between the oceanic lithosphere and the surrounding upper mantle causing the former to descend due to negative buoyancy. An exquisite example of passive flow is then the constant injection of oceanic lithosphere into the mantle causing large-scale thermal convection (Coltice and Ricard, 1999). On the other end, divergent margins refer to two plates spreading apart such as mid-ocean ridges and continent rift zones. Spreading margins act as conveyors systems as hot mantle materials upwell due to positive buoyancy. Since the rigid plates are locked for extended periods of time, the boundaries accumulate and relieve most of the stresses due to tectonic activity, thus making it a hotbed of earthquake activity. The accumulation phase

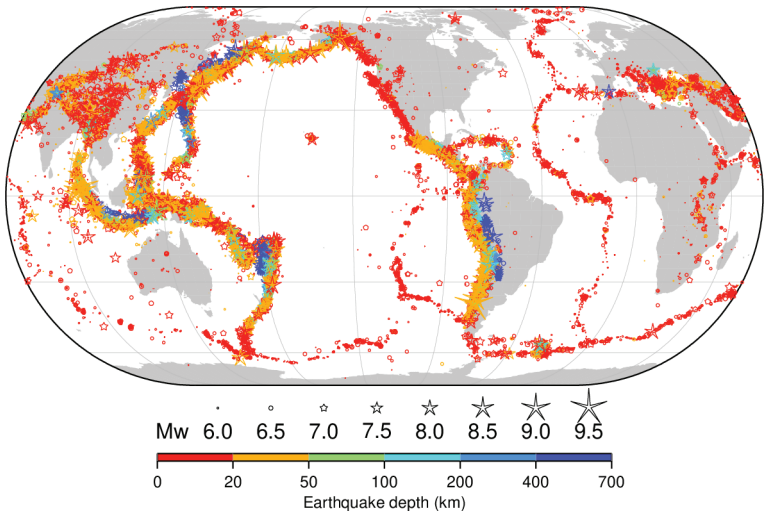

FIGURE 1.1: Global seismicity map as of 9 April 2020. The data are retrieved from the ISC-GEM Catalogue (Storchak et al., 2013; Storchak et al., 2015; Di Giacomo, Engdahl, and Storchak, 2018).

is called interseismic loading. In this period, the stresses stockpile for extended periods of time due to frictional loading between plates as they struggle to move past each other. The locked interfaces will then be ripe for rupture and slip past one another once a critical stress level is attained, hence an earthquake (also called co-seismic slip in the earthquake cycle). This is followed by a transient period, known as post-seismic deformation. In this phase, the fracture heals from the seismic event and most of the delayed deformation caused by stress relaxation will be accounted for in this phase. Fig. 1.1 shows the global distribution of earthquake hypocenters (Storchak et al., 2013; Storchak et al., 2015; Di Giacomo, Engdahl, and Storchak, 2018). As expected, most of the events occur along plate boundaries where deformation is primarily concentrated. The sporadic nature of some of the events highlight the existence of intraplate earthquakes resulting from faulting mechanisms.

Over geologic time scales however, the upper mantle behaves like a liquid. In this state, the mantle is characterised as a highly-viscous material that undergoes deformation creep. Since the temperature difference between the crust and the core is immense, the zone of influence of this temperature difference reaches to the upper mantle; thus undergoing convection over geologic time scales, similar to how warm water rises from the bottom of the pot due to the heating of its base. Seismological studies have then shown the existence of discontinuities between 410 km and 670 km (Anderson, 1967). The presence of the transition zone may seem to be an elegant way to subdivide the mantle into two separate regions comprised of an upper mantle and a lower mantle. To now, solid earth geophysicists debate whether the mantle acts as a single unit or

two separate entities. However, we can safely assume that the mantle is a dynamical unit, with the transition zone being a product of small-scale processes such as thermochemical phase transitions. Just below the lower mantle, there exists a layer that outstretches down to 200 km before the CMB. This is known as the D″ (pronounced as *d double prime*) layer. The D″ layer is characterised as a smooth undulated layer and contains heterogeneities whose structures show uncanny resemblance to continental formation (Czechowski, 1993). Although its mechanism is not yet well-understood, its existence provides plausible implications on mantle convection.

Finally, at the center of the Earth lies the core, an extremely hot and dense material composed mostly of iron-nickel alloy. The transition between the lowermost mantle and the core is called the core-mantle boundary (CMB) located at approximately 2900 km below the Earth's surface, which was discovered by Beno Gutenberg in 1912 (refer to Brush (1980) for a review). Tomographic studies of the deep Earth suggest the existence of a liquid outer core and a solid inner core partitioned by the inner core boundary (ICB), also called the Lehmann discontinuity, named after the Danish seismologist, Inge Lehmann who discovered its existence back in 1936 (Lehmann, 1936). The ICB is located at a depth of approximately 5200 km relative to the Earth's surface, whereas the inner core extends all the way towards the Earth's center at 6370 km. As its names suggests, the outer core behaves in a fluid-like fashion due to the absence of $S-$ wave propagation. This observation, accompanied by the refraction of $P-$ waves by the outer core produces an area of the Earth known as the 'shadow zone' where no elastic body waves are recorded at the surface. The Earth's magnetic field is thought to have been generated by internal convection of magnetic fluids within the outer core, known as the dynamo theory (e.g. Buffett, 2000). Self-consistent theories related to the geodynamo problem is still a hot topic of debate and remain an open problem in the field of geomagnetism.

1.1.2 Convection in the Mantle

In this chapter, we will focus on describing the structure and the dynamics of the mantle. In this way, we can slowly build the foundation towards geodynamic tomography.

The advent of plate tectonic theory expedited the notion of convection in the mantle. However, it was as early as the middle 1800's when Kelvin published his calculation about the age of the Earth when the mantle behaves like a fluid is theorised (Kelvin, 1963). Here, Kelvin argued that the Earth is 100 My which initially sparked controversy among creationists during the Victorian era. In 1895, John Perry subsequently refuted Kelvin's estimate, and argued that heat transfer by convection

could have reconciled the physical and the geological assertions (Perry, 1895). Years later, Holmes (1931) argued that plate tectonics is driven by subsolidus convection that resulted from the generation of heat by radioactivity. Not long after, Hess (1962) confirmed that plate tectonics is indeed driven by mantle convection. Based on his work on the Great Global Rift system, he postulated that the ridge acts as a conveyor that pushes the plates away as the upwelled molten magma oozes out the conduit. The molten magma then congeals, forming a new oceanic crust that branches off the spreading center, a phenomenon known as seafloor spreading. Since then, plate tectonics has been a widely accepted theory and accelerated the advancement of mantle dynamics.

In physics, natural convection is a process at which heat is transferred through the movement of fluid parcels due to density variations. The density structure of the Earth's mantle is heterogeneous and hence, drives thermal convection due to these buoyancy anomalies. To first order, the Earth's temperature decreases with depth which then defines its geotherm (*i.e.*, lines of equal temperature). The geotherm, however, is not consistent radially but is undulated due to several major factors such as the (1) liberation of thermal energy due to exothermic reactions powered by radioactive decay of radiogenic elements, (2) the temperature difference between the superficial layer and the deep lower mantle brought about by the heat transfer from the CMB, and (3) finally, secular cooling of the Earth after its accretion (Schubert, Turcotte, and Olson, 2001). Upon heating, the rocks undergo expansion thus decreasing their density with depth. Since the tendency of this thermodynamic system is to minimise the free energy, the heterogeneous mantle is gravitationally unstable. To achieve stability, the mantle has to upwell deep low density material followed by the downwelling of an equivalent volume of cold high density material. However, as the heavier material flushes to the bottom, its temperature increases. The same can be said for the positively buoyant mantle materials. The result is then a self-sustained convection, a process that occurs over geological time scales (Schubert, Turcotte, and Olson, 2001) (See Fig. 1.2).

Physically speaking, the subsolidus mantle can be sufficiently described in terms of the motion of a highly viscous fluid subjected to a large temperature gradient. In such cases, the fluid parcels undergo convection, however in a creeping fashion due to the negligibility of the advective forces with respect to the viscous forces. For mantle convection to occur, three main conditions have to be fulfilled. The first one involves a a highly viscous fluid bounded by a large temperature difference between the top and the bottom, while subjected to its own gravity. The second is a positive thermal expansivity. In this way, mantle materials tend to increase their volume as soon as they heat up. And finally, thermal conductivity should be relatively low, allowing for heat

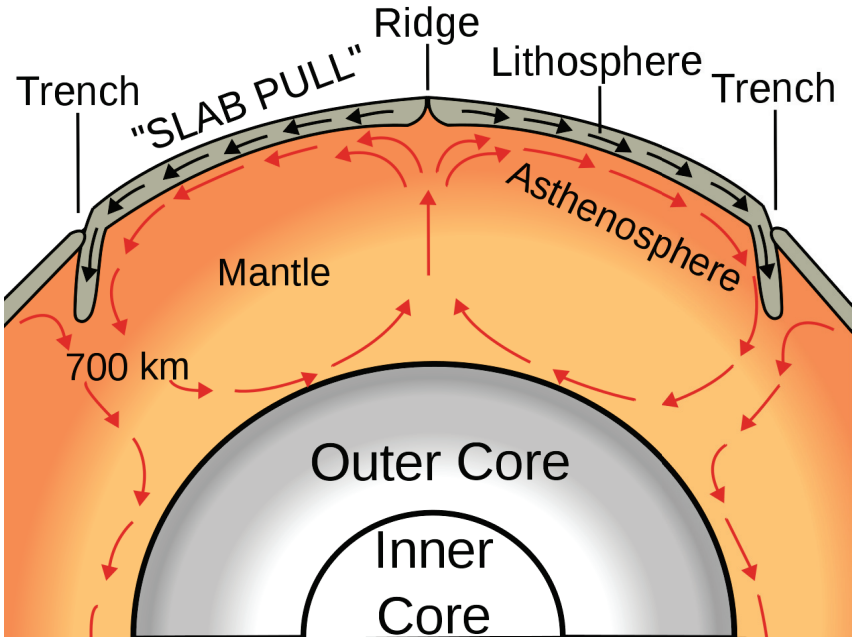

FIGURE 1.2: A cartoon depiction of mantle convection retrieved from the U.S Geological Survey online repository (USGS, 2020).

to be transferred solely by thermal convection. In the mantle, the temperature drop is sustained by the heat transfer at the CMB and the influence of the cold surface that is subjected to atmospheric temperature and pressure. Even though the temperature drop causes the volumetric expansion of the materials and gravitational instability, thermal convection in the mantle is hindered by viscous dissipation and thermal diffusion. To allow for natural convection, the time scale for heat transfer by convection should be much lower than the time scale for heat diffusion. Formally, this is defined by the dimensionless Rayleigh number Ra, named after Lord Rayleigh. Natural convection occurs when Ra exceeds a critical value. Thus, any value below it does not allow for free convection; instead, heat transfers purely by conduction. The characteristic Rayleigh number in the mantle depends on a variety of parameters such as the length scale and the temperature drop across the whole domain. Further details will be elaborated in the succeeding chapter.

It is now clear that several factors come into play that drive convection in the mantle. Of course, it is impossible to dig a hole in the Earth to observe first-hand such processes, not to mention how it is absurd given the time scales of mantle convection. To understand such processes, we have to rely on experiments scaled-down to the size of a laboratory. However, doing such proves to be laborious due to uncontrollable problems associated with the correct scaling relations and given the limited resources

appropriate for these experiments. This calls for the need to use computers to mathematically model mantle convection. Although the physics of convection is already laid upon, it will still be difficult to perform such computations without fine-tuning some input parameters that rely on direct observations. The need for geological data is thus indispensable . Geological constraints such as the mechanical properties of exhumed mantle minerals serve as *A priori* information needed for successfully modeling convection that is representative of the Earth's mantle.

1.1.3 Models of mantle convection

Experimental models of convection

According to George Box (1987), all models are wrong, however, some are useful. Experimental models of mantle convection, although they only constitute some characteristics of the mantle, still reveal key insights about the processes that govern it, and possibly its origin. Laboratory experiments of mantle convection can be categorised in two sections. The first is the replication of the Rayleigh-Benard convection, where a viscous fluid inside a tank forms upwellings and downwellings generated by a large temperature gradient powered by a constant heat source from below and an isothermal bath that cools the fluid from above. Early accounts of such experiments only reach Ra values of about 10^5 (Whitehead Jr and Luther, 1975; White, 1988), which is not well-suited to that of the mantle $10^7 - 10^8$. With the same arrangement, Davaille and Jaupart (1993) result of Ra $10^6 - 10^8$ seem to be more well-grounded as they accounted for large viscosity contrasts using corn syrup. The second class of laboratory setups involve the artificial creation of plumes by injecting a positively buoyant parcel to the surrounding denser fluid, followed by analysing changes in its properties along its trajectory (e.g. Olson and Singer, 1985).

These experiments, however, are unable to address several issues (Lithgow-Bertelloni et al., 2001). The first of which is the inability to account for instabilities along the boundary which supposedly generates plumes. Although this is remedied with the use of the more classical Rayleigh-Benard scheme, these setups are unable to recreate flows associated with plate motions. Moreover, these studies exhibit strong top and bottom boundary layer interactions. As such, the heat transfer through the upper boundary layer strongly affects the kinematics of the upwellings.

Numerical models of convection

At geologic timescales, the mantle behaves like a fluid where rocks creep in a laminar fashion. Such flow is characterised by a very low Reynolds number where advective

inertial forces are negligible compared to viscous forces, also known as Stoke's flow, a very high Rayleigh number where convection is buoyancy-driven, and an infinite Prandtl number where thermal diffusivity is negligible compared to momentum diffusivity. To represent some aspects of mantle dynamics, we then need a model for the physics of convection. These models are formulated in terms of coupled differential equations whose solutions are vector fields that represent flow in the mantle. These equations then need to be set up in such a way that they can be solved numerically. Since numerical models are never perfect, they should converge towards an approximate solution constrained by a tolerance value. The complexity of these equations requires one to make some major assumptions in order to simplify the problem. These may include initial and boundary conditions, and simplified parameterisation schemes when some parameters are poorly known at depth. As an example, mantle flow, to first-order, can be formulated using Stoke's equations. In these equations, one is required to simultaneously solve the conservation of mass, momentum, and energy equations to obtain the velocity that represents the motion of a convecting viscous fluid. Assuming that the mantle is an incompressible fluid, the conservation of mass writes:

$$\nabla \cdot \mathbf{u} = 0 \tag{1.1}$$

In the mantle, the inertial forces are negligible, the conservation of momentum is given by:

$$-\nabla P + \nabla \cdot \boldsymbol{\tau} + f = 0 \tag{1.2}$$

Assuming no radiogenic heat is transferred, the conservation of energy is:

$$\frac{DT}{Dt} = \kappa \nabla^2 T \tag{1.3}$$

where $\mathbf{u}$ is velocity, P is pressure that accounts for the dynamic and the hydrostatic component, η is kinematic viscosity, f is the body force term (gravitational force acting on the density), T is temperature, and κ is thermal diffusivity. This system of differential equations then describes time-dependent flows of a highly viscous and incompressible fluid.

Theory-based mantle convection models have been pioneered in the late 60's with simple assumptions such as uniform material properties, a 2D Cartesian coordinate system implemented, and time-independent flows (e.g. Turcotte and Oxburgh, 1967). Although the models lack complexity and only represent some aspects of mantle flow, it still paved the way to put forward additional insights regarding mantle dynamics

such as the ability of mantle convection to sufficiently provide energy to drive plate motions (e.g. Turcotte and Oxburgh, 1967).

Current developments in the numerical modeling of convective flow have addressed the shortcomings brought initially by classical methods. These include the implementation of (1) lateral viscosity variations based on a temperature-dependent exponential law (e.g. Torrance and Turcotte, 1971; Albers, 2000), (2) self-consistent equations of state (e.g. Ita and King, 1994; Ita and King, 1998), 3D geometry (e.g. Gable, O'connell, and Travis, 1991; Tackley, 1993; Zhong et al., 2000; Zhong et al., 2008), (3) spherical geometry, which permits the extension to global convective flow (e.g. Hager and O'Connell, 1981; Bercovici, Schubert, and Glatzmaier, 1989; Tackley, 1993; Zhong et al., 2000; Zhong et al., 2008), and (4) rigid plate motion, which then induces shear flow unrelated to buoyancy anomalies (also known as toroidal flow) (e.g. Gurnis and Hager, 1988; Gable, O'connell, and Travis, 1991; Tackley, 2008; Piromallo et al., 2006). Modern advancements in this field, as well as the emergence of multi-core architectures, have also included techniques that tremendously increase the computational efficiency of solving the Navier-Stokes problem. One of which is the widely used spatial decomposition method, where the physical modeling domain is subdivided into smaller spatial domains allocated evenly to a set of processors (e.g. Bunge and Baumgardner, 1995; Zhong et al., 2000; Kageyama and Sato, 2004; Aleksandrov and Samuel, 2010).

Following the pioneering work of Turcotte and Oxburgh (1967), one may solve analytically Stoke's equations by imposing plate velocities and a constitutive law that accounts for non-Newtonian flows (e.g. Tovish, Schubert, and Luyendyk, 1978). This family of analytical solutions are called corner flows, and this class of modeling scheme is called passive mantle models. In the latter, mantle flow is assumed to be plate-driven. To the first order, these can be applied in the upper mantle with or without lithospheric coupling. Another class of models are active mantle flow models, where the flow is buoyancy-driven instead of prescribed plate velocities. One of the earliest attempts to produce global mantle flow models based on theory and observation that proved to be a success was the work of Hager and O'Connell (1981). In their work, semi-analytical mantle flow models can be mapped out globally supposing no lateral variations in mantle viscosity. This is carried out by imposing plate velocities as surface boundary conditions, computing crude density anomalies inferred directly from early seismic tomography studies, and evaluating 1D radial viscosity profiles from postglacial rebound, geoid anomalies, and/or dynamic topography (e.g. Hager and O'Connell, 1981; Hager, 1984; Hager and Clayton, 1989; Mitrovica and Forte, 1997; Panasyuk and Hager, 2000, hey). To account for lateral variations in viscosity, the

Stoke's equations should be approached fully numerically, with some even accounting for sharp viscosity contrasts (Christensen and Harder, 1991; Albers, 2000; Tackley, 2008; Ghosh, Becker, and Zhong, 2010; Kronbichler, Heister, and Bangerth, 2012, to name a few).

In our work, we will limit ourselves to instantaneous flow solutions, a subset of active mantle flow models subjected to the steady-state assumption. Here, instantaneous flow solutions are obtained numerically with the following as inputs: 3D density anomalies that depend on temperature using a linear equation of state, 3D viscosity profiles that depend on temperature resembling Arrhenius law for viscous fluids, and a dimensionless Rayleigh number.

1.1.4 Validity of numerical mantle convection models

At this point, we have made our case that mantle geodynamics plays a pivotal role in driving plate tectonics. Numerical models of mantle convection as previously discussed should be testable in order to serve its purpose, that is, to understand mantle dynamics. The key lies on its deformation history which is a permanent signature that something has moved in the geological past. Permanent deformation in the mantle derives from the changes in the volume of a rock as it gets advected along its flow trajectory. Observations of present-day mantle deformation rely on indirect measurements inferred from, but not limited to dynamic topography, plate motions, tectonic stresses, geoid anomalies, and seismic data.

In our work, we will restrict ourselves to seismic data, that is, the observable that maps the elastic response of the medium (*i.e.*, the Earth) due to a disturbance (*e.g.* passive sources such as earthquakes, or anthropogenic sources such as dynamite explosions). The link between mantle deformation and seismic wave propagation lies within a property known as seismic anisotropy. In the upper mantle, the propagation of seismic waves appears to be anisotropic. Thus, seismic anisotropy reveals key insights, not only into the Earth's elastic structure, but also into its present-day dynamics. From this point forward, we will relate this so-called intrinsic anisotropy to the deformation history associated with convection through a specific type of indirect seismic observation - surface waves.

1.2 Probing the deep Earth with long period seismic waves

Over the last decade, the Earth's interior has been successfully imaged thanks to the joint efforts made by seismologists through the use of different state-of-the-art techniques and computational strategies coupled with the advancement of high-performance computing. The idea is to probe the Earth using seismic waves generated by active sources such as man-made explosions and/or passive sources specifically earthquakes. As seismic waves travel within the Earth, its path is strongly affected by the Earth's structure. Knowing the location of the source, the seismic waves recorded at the surface through seismograms thus contain relevant information about a portion of the Earth's structure covered by its path. However, much work remains to be done to interpret these resulting images in terms of various physical properties that govern the dynamics of the mantle. In this section, we will provide an overview of seismic tomography, and then start laying the groundwork for surface wave tomography - an imaging technique which will be of utmost relevance throughout our whole work.

1.2.1 An overview on various available seismic observables

The energy associated with an earthquake travels radially starting at the source with a given speed, and then inevitably arrives at the surface of the Earth where the energy is completely reflected. When recorded at the surface, the ground motion attributed to its propagation is graphically represented as a complex waveform (*i.e.*, a seismogram). The heterogeneity of the Earth's interior structure allows us to decompose this waveform into different components at a given range of frequencies. Since some of the information about the Earth's structure is embedded in the complete waveform, each decomposed signal such as direct, converted, and reflected body waves, surface waves, and ambient noise can be used to extract meaningful information related to different structures, be it that of small-scales of about 1-10 km or large-scale structures of about 10^2 km.

Since each observable exists at specific frequency bands, each of them have different resolving power. Simply put, different data-types sample different length scales. To cite an example, scattering studies use high frequency converted body waves to map seismic discontinuities and small-scale heterogeneities, but fail to sample deep and large scale structures. On the other hand, surface waves are long period observables. As a result, they only recover smooth large scale structures of the mantle, and lack the capability to capture sharp discontinuities and are poor in locating cavities. Another emerging technique is ambient noise tomography (ANT) which is practically useful to image the Earth's superficial layer. As an auxiliary method of surface wave

tomography, current advancements in ANT proved it to be efficient in imaging sharp discontinuities, as well as the Moho, from reconstructed surface waves (e.g. Macquet et al., 2014; Lu et al., 2018).

1.2.2 Seismic tomography

One of the major applications of seismology is the inference of the spatial distribution of seismic velocities, and in fine the elastic structure of the Earth from seismic observations measured at the surface, a technique known as seismic tomography. Seismic velocities are bulk properties, and hence are derived from the mineralogy and the thermochemical state of the Earth's interior. They thus provide basic constraints on chemical composition, petrology, and temperature of the mantle. Seismic tomography is an imaging technique that is loosely based on the concept of an inverse problem. In the context of geophysical studies, the main goal of an inverse problem is to retrieve model parameters that represent a subset of the Earth's physical properties given some data measured at the surface. Mathematically, this translates to:

$$\mathbf{d} = g(\mathbf{m}) \tag{1.4}$$

where $\mathbf{d}$ is the data vector, $\mathbf{m}$ is also a vector containing the Earth model, and g is a function related to the physics of the problem.

In seismic tomography, $\mathbf{d}$ is the seismic observable measured at the surface, $\mathbf{m}$ is our Earth model, and g is a nonlinear mapping related to the path traversed by the propagating wave. For reasons of computational tractability, the fundamental data inverted in tomography is the travel time of seismic waves. Seismograms however, only record the arrival time of these waves. This requires the determination of the origin time and the location of the source if one wants to convert these into travel times. For non-anthropogenic sources such as earthquakes in particular, these should be estimated using other methods, however, this would then introduce uncertainties that will potentially affect the quality of the inverted structure. Knowing the travel times between the source and the receiver, we can simply formulate the forward problem corresponding to that pair as:

$$T(S,R) = \int_S^R \frac{1}{V(\mathbf{r})}\mathbf{dr} \tag{1.5}$$

where the reciprocal of the velocity V, also called the slowness, is integrated along the ray path from the source S to the receiver R. As a simple exercise, let us suppose that

the velocity is constant. In this case, the total travel time is just the sum of the individual times in each segment **dr**, hence providing an integral constraint onto the velocity structure traversed by the rays. This however, introduces non-uniqueness to the problem since other ray paths may share this same characteristic. A single travel time record is therefore insufficient to map the distribution of seismic velocities. This urges the addition of more travel time data to tightly constrain the velocity structure. Ideally, a homogeneous ray path coverage solves the problem, but this is close to impossible given the unresolved problems associated with the distribution of source-receiver pairs.

In the most simple cases, the example above illustrates that if the seismic velocity structure is known, then the travel times can be computed straightforward in a complete forward modeling approach. The inverse problem is much more difficult however, that is, the retrieval of the velocity structure from a set of travel time measurements at the surface. As such, various methods have been developed to tackle the tomographic problem such as the inclusion of other data-types (*e.g.*, surface waves, converted phases, and/or eigenfrequencies of normal modes) in order to better constrain the velocity structure. Still, just like most geophysical inverse problems, seismic tomography is ill-posed (*i.e.*, existence of non-unique solutions) and ill-conditioned (*i.e.*, small changes in the inputs greatly influence the outputs). This arises due to the uneven distribution of sources and receivers in a given area of study causing inadequate ray coverage. Apart from the inhomogeneous sampling, another recurring complication is due to errors in the observed data (*i.e.* instruments are never perfect but are precise to some degree, and that measurements are degraded due to noise).

Formulating the tomographic problem then requires one to choose an appropriate parameterisation scheme. One of the most common practices is to apply regular spatial parameterisations, such as an evenly-spaced grid cells in 2D or blocks in 3D. The standard approach is then to set in advance the velocity structure as basis functions such as splines or spherical harmonics (in the global case). However, this introduces a trade-off between model uncertainty and model resolution. Suppose that one chooses a fine grid. Intuitively speaking, this would tremendously increases model resolution. However, doing so deteriorates the recovered model due to the mapping of the data noise onto the model parameters, thus having poor constraints on the velocity structure. Moreover, without sufficient ray coverage, poorly sampled regions produces small-scale artifacts unrelated to the true velocity structure. On the other hand, coarsening the grid may indeed better constrain the structure, but at the cost of its resolution.

Seismic tomography is indeed a powerful tool; however, it is restricted by the problems associated with inhomogeneous sampling and choice of parameterisation. To

combat this ill-constrained problem, one would then favor a stable and unique solution through regularisation. Various regularisation techniques are available (e.g. Zhdanov, 2002). In the context of tomography, perhaps the two most common techniques are norm damping and smoothing (e.g. Charlety et al., 2013). Damping refers to the coercion of a solution to stay as close as possible to the pre-established model while still being able to explain the data. This pre-established model depends on *a priori* knowledge. As an example, if a velocity structure is already known based from previous studies, this may serve as a reference Earth model where the favored solution depends on a perturbation around it. The second type of regularisation is known as smoothing. Smoothing techniques make use of approximate functions that modify the values of neighboring poorly constrained cells from a given well-constrained cell. The output is therefore a smooth velocity structure that to some extent, eliminates unwanted small-scale artifacts. It is important to note however that regularisation is a user's prerogative, and that such subjective choices are not driven by the data nor the physics of the problem at hand.

At this point, we have talked about the basics of the tomographic problem as well as some conventional techniques used to handle some complications involved in the method. The following sections below briefly discuss two kinds of tomographic strategies: isotropic tomography and anisotropic tomography, and some of the main features observed inside the Earth obtained from them. These two inversion schemes will then be the basis of geodynamic tomography.

1.2.3 Isotropic tomography

Isotropic tomography refers to the inversions for the isotropic shear and compressional wave velocities, V_s and V_p. As the complete elastic structure of the Earth depends on 21-independent coefficients, its isotropic structure only depends on two parameters. In terms of the isotropic velocities, the isotropic Earth can be written in terms of a Voigt-averaged 6×6 stiffness matrix S_{0ij} as:

$$S_{0ij} = \begin{pmatrix} \rho V_p^2 & \rho(V_p^2 - 2V_s^2) & \rho(V_p^2 - 2V_s^2) & 0 & 0 & 0 \\ \rho(V_p^2 - 2V_s^2) & \rho V_p^2 & \rho(V_p^2 - 2V_s^2) & 0 & 0 & 0 \\ \rho(V_p^2 - 2V_s^2) & \rho(V_p^2 - 2V_s^2) & \rho V_p^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\rho V_s^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\rho V_s^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\rho V_s^2 \end{pmatrix} \tag{1.6}$$

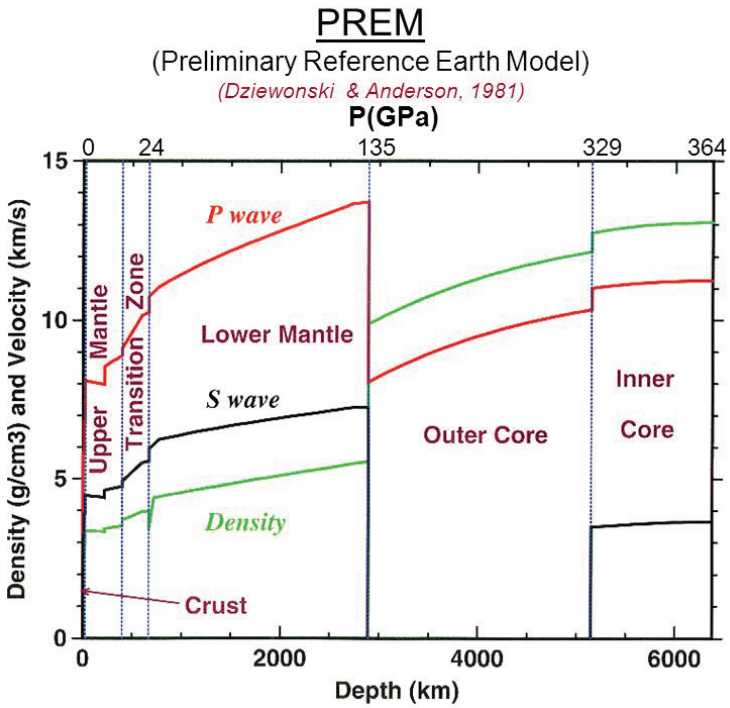

FIGURE 1.3: Preliminary reference Earth model (PREM) (Dziewonski and Anderson, 1981). Image retrieved courtesy of Ada Palmer.

where ρ is the density. In the mid to late 80's, seismologists were able to exploit the growing amount of seismic records, and build 1D reference Earth models reconstructed from direct $P-$ and $S-$ wave travel time residuals and the eigenfrequencies of the normal mode vibrations of the Earth (Dziewonski and Anderson, 1981) (See Fig. 1.3). Other 1D reference models have then been produced such as the IASP91 (Kennett and Engdahl, 1991) and the AK135 (Kennett and Engdahl, 1991). All of these models represent the radial average of the Earth and hence, various physical parameters may be obtained with depth such as attenuation and anisotropy (more on this later on). As such, they served as reference models for 2D and 3D image reconstruction techniques.

Over the last 40 years, isotropic tomography still considers to be an active domain of research in the field of seismology. It has produced countless robust Earth models with a considerable degree of overlap (e.g. See Woodhouse and Dziewonski, 1989; Romanowicz, 2003; Trampert and Fichtner, 2013, for a comprehensive review). To cite some examples, V_p velocity models inferred from the inversion of $P-$ and $S-$wave travel time residuals have successfully captured important features from tectonically active settings such as descending lithospheric slabs (e.g. Hilst, Widiyantoro, and Engdahl, 1997; Romanowicz, 2003). The geometry of the recovered slabs vary with some

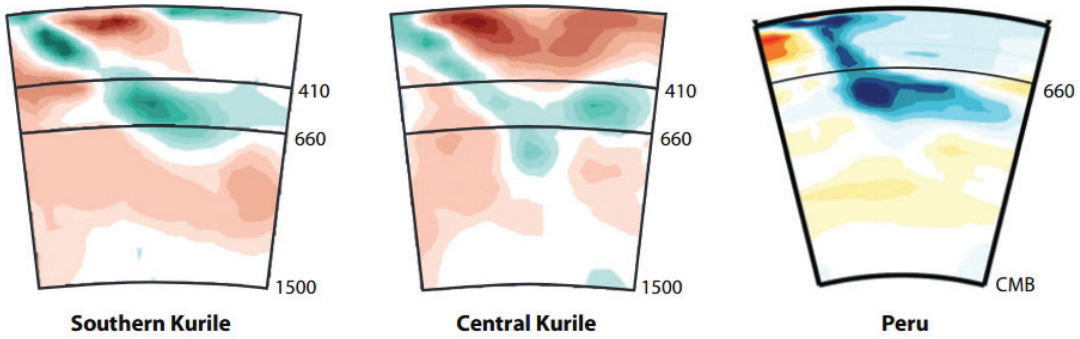

FIGURE 1.4: $P-$wave tomography of stagnant slabs across the transition zone (Gorbatov et al., 2000; Obayashi et al., 2006).

even extending down to the depths of the lower mantle. One of the most notable observations is the stagnation of old subducting slabs at the 660 km discontinuity such as those observed beneath the circum Pacific and the Mediterranean (see Fukao et al., 2009, for a review). Fig. 1.4 shows V_p models of stagnant slabs across the transition zone beneath the North Pacific (Gorbatov et al., 2000) and Peru (Obayashi et al., 2006).

On the other hand, V_s models are inferred using a combination of low frequency travel time residuals, surface wave data, and/or normal-mode splitting functions(e.g. Mégnin and Romanowicz, 2000; Kustowski, Ekström, and Dziewoński, 2008; Ritsema et al., 2011). Low-velocity mid ocean ridges appear to be well-resolved up to a depth of 100-150 km by isotropic inversions for $S-$wave models. Across the oceanic upper mantle, observations from isotropic tomography suggest the existence of radially symmetric seismic velocity gradients which have been attributed to a low velocity zone (LVZ) beneath a high velocity layer (e.g. Zhao, Hasegawa, and Horiuchi, 1992; Matsubara et al., 2005). This high velocity layer coincides with the seismogenic lithosphere. Contrastingly, the LVZ in most parts of the Earth somehow matches with the mechanically weak layer of the upper mantle (*i.e.*, asthenosphere). As mentioned earlier, the mechanical properties of each layer are distinguished based from the fact that the lithosphere skids over the weak upper mantle. Thus, we may in fact deduce that the rough correspondence between the elastic and the mechanical properties of each layer means that seismic observations are indeed sensitive to the mechanical structure of the Earth. Down to the depths of the lowermost mantle, one of the most striking features constrained by V_s models are the existence of the large low-shear-velocity provinces (LLSVPs) in the lowermost mantle beneath Africa and the Pacific as initially documented by Masters et al. (1982), Woodhouse and Dziewonski (1984), and Dziewonski (1984).

In essence, models inferred from travel time tomography are often designated as high resolution models although some have argued that such statements are misleading due to the limitations of a gridded parameterisation (Trampert and Fichtner, 2013). Consequently, V_s models reconstructed from low frequency observations exhibit smooth long-wavelength structures, also partly due to the implementation of lower-order spherical harmonic basis functions. As such, they sample the Earth deeper compared to high-frequency direct $P-$ waves, thus providing good vertical resolution but poor lateral resolution.

1.2.4 Anisotropic tomography

All of the results mentioned above assumes that the elastic properties of the Earth do not vary with the direction of propagation and/or the orientation of the polarised waves. Such assumptions would fail to explain anomalous seismic signatures measured at the surface such as the delay time in core-refracted phases (also known as SKS and SKKS), the azimuthal variations of surface wave phase velocities, and the discrepancies between the phase velocities of Love waves and Rayleigh waves (also known as the Rayleigh-Love discrepancy). Because of this, seismologists resort to anisotropic velocity tomography. As we will see later on, anisotropic tomography allows us to constrain deformation patterns in the mantle, and thus provide insights not only to its structure, but to its dynamics as well.

Elastic anisotropy

In physics, anisotropy is an umbrella term that refers to the changes in the physical properties of a material when measured at different directions. In seismology, seismic anisotropy is a property of an elastic material to assume variations in the velocity of seismic waves passing through it depending on their direction of propagation or polarisation orientation. We will now refer the azimuthal dependence of seismic wave velocities as azimuthal anisotropy, and the variations in seismic wave speeds with polarisation as radial anisotropy. It is worth mentioning that anisotropy differs from heterogeneity in such a way that its physical properties vary with direction whereas the latter depends on the non-uniformity of the composition. The concept of radial anisotropy is a seismological analogue of optical birefringence. Here, seismic anisotropy causes the splitting of an incident $S-$wave into two $S-$waves with horizontal V_{SH} and vertical polarisations V_{SV} and different velocities as shown in Fig. 1.5. Surface waves also suffer from radial anisotropy. As Rayleigh waves are sensitive to V_{SV} whereas

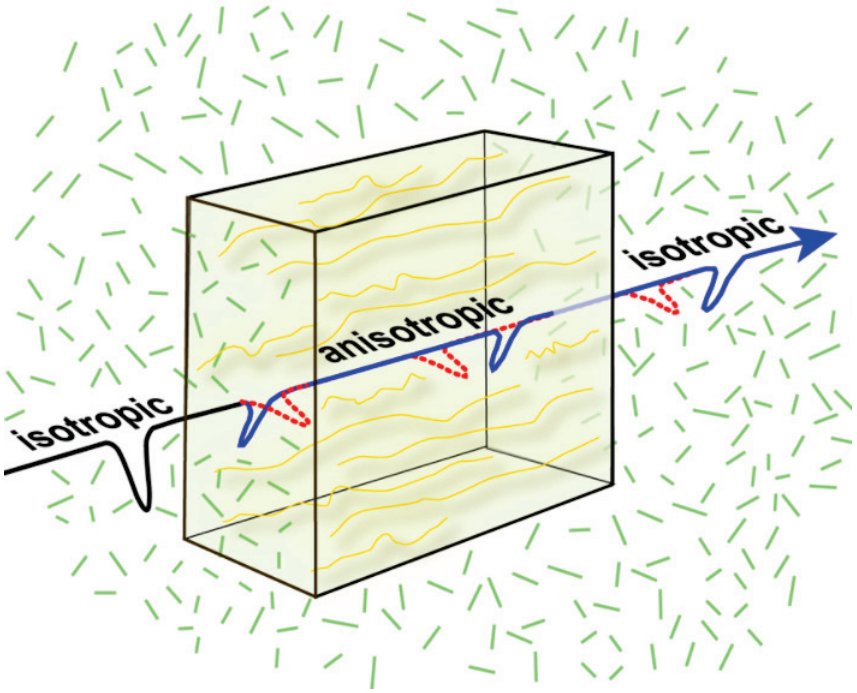

FIGURE 1.5: As soon as an $S-$wave is incident onto an anisotropic media, it splits into two orthogonally polarised $S-$waves with different speeds. The fast wave is indicated in blue and the slow one in red. As it traverses yet into an isotropic media, they have identical speeds again and are separated by a delay time δ t. Image courtesy of Ed Garnero.

Love waves are sensitive to V_{SH}, this causes the well-known Rayleigh-Love discrepancy when these waves propagate into an anisotropic media. As a simple exercise, consider Hooke's law of linear elasticity where the stress σ_{ij} relates to the strain by ϵ_{ij},

$$\sigma_{ij} = c_{ijkl}\epsilon_{ij}, \tag{1.7}$$

Since stress and strain are symmetric tensors where both have 6 independent components, this makes the number of independent elastic parameters to 36. The idea of strain energy further reduces the number of independent elastic parameters to 21. Now we consider plane waves propagating in an anisotropic media. Suppose we have a plane wave solution in a homogeneous and anisotropic media given by,

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{a}\psi(t - \frac{\mathbf{n}\cdot\mathbf{x}}{c}), \tag{1.8}$$

where **a** is the polarisation vector, c is the phase velocity, and the wave is propagating along **n**. The plane wave solution satisfies the anisotropic wave equation,

$$\rho \frac{\partial^2 u_i}{\partial t^2} = C_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l}, \tag{1.9}$$

Substituting equation 1.8 into equation 1.9 we obtain,

$$a_i = m_{il} a_l \frac{1}{c^2}, \tag{1.10}$$

where m_{il} is the Christoffel matrix and is expressed as,

$$m_{il} = \frac{1}{\rho} c_{ijkl} n_j n_k, \tag{1.11}$$

Once we find m_{il} for a given anisotropic medium, we can immediately obtain the phase velocities of the different wave types through its eigenvalues. The corresponding eigenvectors of this matrix give us the polarisation direction of these waves. The speed and the polarisation direction of the wave depend on the propagation direction and the crystallographic structure of the mineral. In the next few discussions we identify one basic type of crystal symmetry arising from a given mineral that exhibit anisotropic properties, namely hexagonal anisotropy.

Seismic waves in a hexagonally symmetric media

In the upper mantle, the propagation of seismic waves appear to be anisotropic, and thus $P-$ and $S-$waves suffer from variations in their wave speeds of about 20% (Kumazawa and Anderson, 1969). Anisotropy in the upper mantle are often attributed to the presence of anisotropic minerals as documented by mineralogical studies (refer to Maupin and Park, 2015, for a review). The complexity of the 21-component elastic tensor and the inability of seismic observables to be sensitive to each component, more often than not, urged seismologists to reduce the number of independent parameters by imposing symmetry constraints. The type of symmetry most often used in seismological studies is hexagonal symmetry with a horizontal axis of symmetry, also called horizontal transverse isotropy (HTI). Imposing HTI reduces the number of independent parameters to only five: A, C, F, L, and N. These are also called the Love parameters. Another class of symmetries is vertical transverse isotropy (VTI) which have often been utilised to resolve the Rayleigh-Love discrepancy problem and shear wave birefringence. Finally, we have tilted transverse isotropy (TTI) where the symmetry axis is tilted and defined by two angles: dip angle θ and the azimuth Ψ. At least

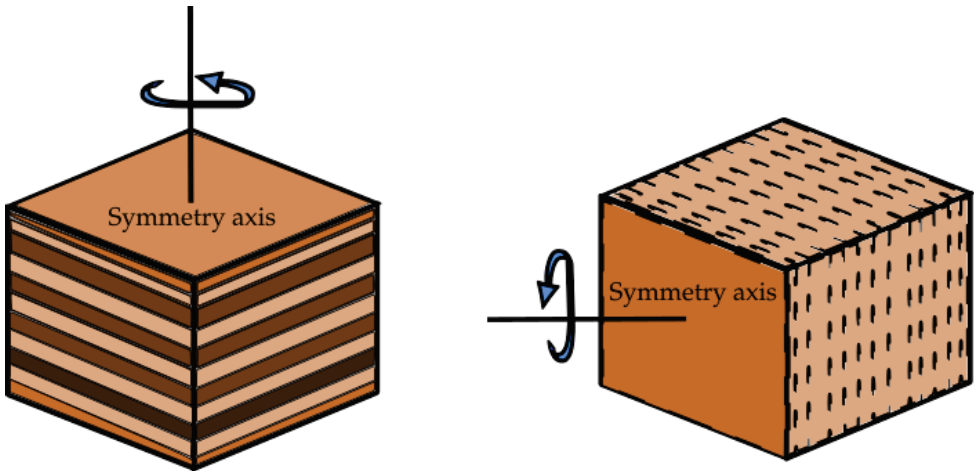

FIGURE 1.6: Cartoon depiction of a vertical transverse isotropic medium (Left) and a horizontal transverse isotropic medium (Right).

in the upper mantle, such assumptions are valid, to first order, as petrological studies suggest that upper mantle minerals exhibit such symmetries (e.g. Montagner and Nataf, 1988).

Going back to our little exercise above, let us consider a VTI solid whose stiffness matrix is given by:

$$\begin{pmatrix} A & A-2N & F & 0 & 0 & 0 \\ A-2N & A & F & 0 & 0 & 0 \\ F & F & C & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 \\ 0 & 0 & 0 & 0 & L & 0 \\ 0 & 0 & 0 & 0 & 0 & N \end{pmatrix}, \tag{1.12}$$

Suppose a wave is traveling in the $+x_1$ direction. The corresponding Christoffel matrix is,

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix},$$

Using equation 1.11 and substituting the components of equation 1.12 into the designated components of M we have,

$$M = \frac{1}{\rho} \begin{pmatrix} A & 0 & 0 \\ 0 & N & 0 \\ 0 & 0 & L \end{pmatrix},$$

By inspection, we easily identify the eigenvalues of M which are essentially the seismic velocities:

$$V_p = \sqrt{\frac{A}{\rho}}, V_{SV} = \sqrt{\frac{L}{\rho}}, V_{SH} = \sqrt{\frac{N}{\rho}}.$$

The polarisation vector can be inferred from the eigenvectors. Thus:

$$\mathbf{a_p} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{a_{sh}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{a_{sv}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The example above shows that 3 waves are produced for a wave incident to a VTI medium traveling along the $+ x_1$ axis. The fastest one being a P-wave polarised parallel to the propagation direction. Two of which are S-waves having two contrasting speeds that depend on the elastic coefficients L and N, and polarisation orthogonal to the propagation direction. This phenomenon where V_{SV} and V_{SH} have different velocities demonstrates shear wave splitting.

Anisotropic surface wave tomography

For the sole purpose of building toward geodynamic tomography, the scope of the anisotropic inversions discussed here will only cover surface wave inversions.

Surface waves are long period observations, and as a result, cannot constrain velocity structures associated with small-scale inhomogeneities. Still, surface wave tomography offers a powerful tool to constrain seismic anisotropy, and thus image the elastic structure of the upper mantle at both regional and global scales. Part of this is due to the fact that the energy associated with surface waves are mostly concentrated across the subsurface. Hence, its fundamental mode is sufficient to evenly sample upper mantle structure along the great circle path traversed from the source to the receiver. With a growing amount of surface wave data, seismologists have produced detailed models of azimuthal anisotropy over the last decades (e.g., Debayle, Kennett, and Priestley, 2005; Deschamps et al., 2008; Adam and Lebedev, 2012; Yuan and Beghein, 2013; Yuan and Beghein, 2014), and radial anisotropy (e.g., Plomerová, Kouba, and Babuška, 2002; Lebedev, Meier, and Hilst, 2006; Nettles and Dziewoński, 2008; Chang et al., 2014; Chang et al., 2015).

Numerous studies have inverted dispersion curves by minimising the difference between observed and synthetic phase and/or group velocities, proving that they can

effectively constrain the depth dependence of anisotropy (e.g., Montagner and Tanimoto, 1990; Ritzwoller et al., 2002). Still, the adequacy of vertical resolution provided by surface waves depends on the frequency range considered, and the type of surface wave used. Moreover, surface waves still contain limited lateral resolution, thus providing only a large-scale image of anisotropy. Recent developments in surface wave studies have indicated the presence of strong anisotropy in oceanic regions at 50 km to 150 km depth, and in old continental regions at about 200 km to 400 km depth with scales consistent to convective flow in the upper mantle (Gung, Panning, and Romanowicz, 2003). Presence of anisotropy in the upper transition zone (410 to 520km) has been observed in global models ()trampert2002global, while some arguing little to no anisotropy (e.g. Kaneshima, 2014), thus still being highly debatable rendering limited understanding of mid-mantle anisotropy.

As mentioned earlier, seismic anisotropy can be described with 21 independent components of the elastic tensor. In practice however, the full tensor cannot be resolved by the seismic data independently at every location, and generally only a restricted number of parameters are inverted for. This is done by assuming hexagonal symmetry, or by using petrological constraints to impose relations between some of the parameters. Surface waves in particular are only sensitive to 13 parameters that are just a linear combination of the elastic constants (Montagner and Nataf, 1986). General practices in surface wave tomography thus investigate: (1) radial anisotropy assuming a VTI medium. Here, radial anisotropy is constrained by comparing the phase velocities of Rayleigh waves with that of Love waves (Babuska and Cara, 1991); or (2) azimuthal anisotropy, which deals with first-order variations of velocities as function of the azimuth of propagation. For example, azimuthal anisotropy can be inferred from the azimuthal terms of the Rayleigh wave phase velocities assuming an HTI medium (Smith and Dahlen, 1973).

Simultaneous interpretations of radial and azimuthal anisotropy have been the subject of extensive research (e.g. Beghein et al., 2014; Burgos et al., 2014). Joint efforts involving the use of *a priori* information have already been conducted to reduce the high dimensionality of anisotropic inversion. Montagner and Anderson (1989) showed that correlations exist between the elastic constants derived from petrological models, thereby reducing the total number of free parameters to be inverted for. This motivated the development of "vectorial tomography" where it involves inverting for 7 parameters instead of 13: two angles defining the strike and dip of the symmetry axis, three coefficients defining the strength of anisotropy, and finally two isotropic coefficients (Montagner and Nataf, 1988; Montagner and Jobert, 1988). Such a medium is also known as tilted transverse isotropy (TTI) and describes the 3D distributions of

anisotropy. This further led to studies revealing that deformation-induced anisotropy can be described by a TTI medium where correlations appear to exist between P and S wave anisotropy (Becker et al., 2006). Such correlations can then be exploited to further simplify anisotropic inversion. Panning and Nolet (2008) then laid the groundwork to derive finite-frequency kernels of surface waves that are explicitly based on a TTI medium. In practice however, constraining the tilt may still be difficult due to sparse azimuthal sampling, alongside other competing factors such as non-uniqueness of the solution and poor data quality. Even so, simultaneous inversions for radial and azimuthal anisotropy using TTI models have already been applied at the regional scale using probabilistic approaches to combat these shortcomings (Xie et al., 2015; Xie et al., 2017).

1.2.5 Probabilistic approaches to surface wave tomography

Surface wave tomography is an ill-posed inverse problem. Just like with any tomographic problem, this arises from the uneven distribution of sources and receivers causing limited ray path coverage, and from noise in the observed seismograms. The type of spatial parameterisation may also lead to ambiguity when interpreting tomographic results. A conventional technique is to separate the problem into two steps. The first step is to construct velocity maps for each considered period, which is an almost linear inverse problem. It is followed by an inversion of each local dispersion curve to build a model of elastic parameters. The inversion is in general performed using a linearised technique, which favors a stable and unique solution through regularisation, for example by adding a spatial smoothness constraint on the model parameters.

More recently, the development of probabilistic approaches using direct sampling of the model space makes it possible to handle the non-uniqueness of the solution and estimate uncertainties on the inferred parameters. These methods require the evaluation of the forward model a large number of times, and hence have a high computational cost. Nevertheless, numerous works have been successful in applying such inversion schemes to seismic data and in particular to the inversion of surface waves dispersion curves (Shapiro and Ritzwoller, 2002; Shen et al., 2012; Bodin et al., 2016; Ravenna and Lebedev, 2017; Xu and Beghein, 2019).

1.3 Bridging the gap between upper mantle structure and dynamics

Seismic anisotropy originates from various processes with depth and can be observed at different spatial scales. Early accounts of upper mantle anisotropy have been explained in terms of the net alignment of intrinsically anisotropic minerals as they get deformed along the trajectory of the flow (Crampin, 1981; Estey and Douglas, 1986; Nicolas and Christensen, 1987). Anisotropy resulting from these is often called intrinsic anisotropy. Some studies however have acknowledged that some of the large-scale anisotropy reconstructed from joint inversion of Love and Rayleigh waveforms cannot be explained by deformation mechanisms in the upper mantle. Instead, the strong anisotropy mapped may be related to fine-layering, or the presence of small-scale heterogeneities. To cite an example, a horizontally-layered isotropic structure when sampled by long period seismic waves may be mapped as smooth radial anisotropic structure instead (Backus, 1962; Capdeville et al., 2013). In the crust for example, apparent anisotropy exists from horizontal layering due to the preferential alignment of cracks which has long been interpreted to be influenced by the compressional components of the regional stresses oriented parallel to the cracks (Nur, 1971). The apparent anisotropy associated with unmapped small-scales is called extrinsic anisotropy. Anisotropic structures retrieved from tomography may therefore be a combination of intrinsic and extrinsic anisotropy. The ambiguity whether a material is intrinsically anisotropic or strongly heterogeneous may thus mislead seismologists in interpreting the structural origin of seismic anisotropy observed in tomographic image.

1.3.1 Origins of upper mantle anisotropy

Extrinsic anisotropy due to shape preferred orientation

Extrinsic anisotropy is produced under two conditions: (i) when the scale of the heterogeneities are much smaller than the minimum wavelength of the observed wavefield, and (2) when the contrast between seismic wave speeds is very large. Fig. 1.7 illustrates this concept.

One of the known configurations at which extrinsic anisotropy is produced is rock-scale shape preferred orientation (SPO). In the Earth's upper mantle, rock-scale SPO are a result of igneous differentiation (Faccenda et al., 2019). These meter-scale heterogeneities are often depicted by stratified layers due to the preservation of the differentiated magma during the congealing phase (Middlemost, 1986). Since magmatically differentiated oceanic lithosphere is composed of a basaltic crustal layer blanketed by

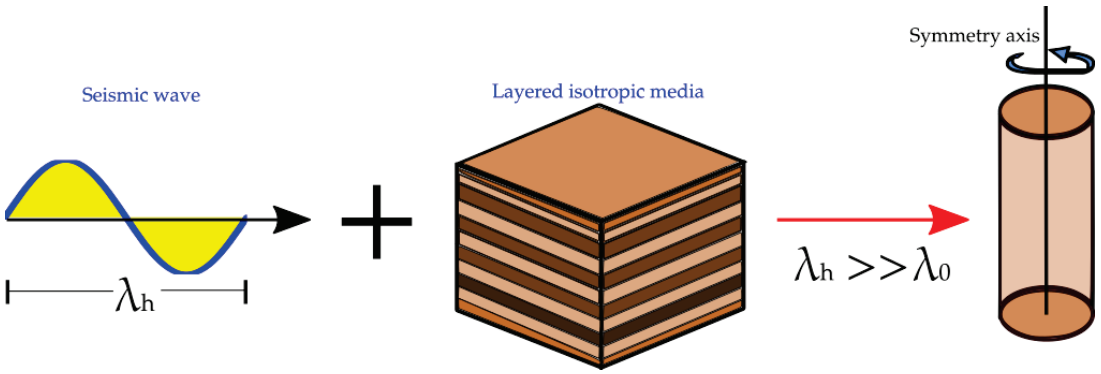

FIGURE 1.7: When seismic waves sample a strongly heterogeneous media whose scales λ_0 are much small than the wavelength of the observed wavefield λ_h, the unmapped small-scales are then removed and instead replaced with effective properties. In the case above, a horizontally stratified isotropic medium is interpreted as a homogeneous radially anisotropic medium with a vertical symmetry axis (VTI).

a depleted harzburgitic mantle (Allègre and Turcotte, 1986), upper mantle structure is often modeled in terms of a mechanical mixture of these two end-member compositions (e.g. Hofmann, 1988; Xu et al., 2008; Ballmer et al., 2015).

Along convergent margins, the density contrast between the oceanic lithosphere and the surrounding upper mantle causes the former to descend due to negative buoyancy. Such processes then also drive large-scale thermal convection in the mantle resulting to the constant injection of oceanic lithosphere into the mantle (Coltice and Ricard, 1999). As it exposes to the mantle, the subducting oceanic lithosphere heats up and as a result, weakens its structural integrity. Upon the reduction of its viscosity, the subducting oceanic lithosphere mechanically stirs with the surrounding mantle and experiences a series of stretching and thinning due to the normal and shear strains associated with mantle convection (Allègre and Turcotte, 1986). The magmatically differentiated lithosphere can then be described in terms of elongated strips that tend to wear-off as deformation dwindles (Olson, Yuen, and Balsiger, 1984). This led Allègre and Turcotte (1986) to develop a geodynamic model of the mantle that would resemble marble cake-like patterns. In their model, the layering may be erased either by dissolution processes or mantle reprocessing at mid-ocean ridges resulting into the cumulative decrease in the length scale of the heterogeneities considered. Assuming that the mixing preserves the physical properties of the two-end members with depth and over geological time scales, such processes may explain rock-scale seismic heterogeneities observed in the mantle (Xu et al., 2008; Stixrude and Jeanloz, 2015).

Intrinsic anisotropy due to crystallographic preferred orientation

Intrinsic anisotropy is interpreted in terms of the development of the crystallographic preferred orientation (CPO) of anisotropic crystals within an aggregate during their plastic deformation (Crampin, 1981; Estey and Douglas, 1986; Nicolas and Christensen, 1987). Due to the physical process at its origin, intrinsic anisotropy can be interpreted in terms of the strain history associated with mantle circulation.

Olivine is the most abundant anisotropic mineral in the upper mantle. Single crystal olivine exhibits orthorhombicity, and thus suffers variations in fast and slow P- and S-wave velocities up to 20 % (Kumazawa and Anderson, 1969). When an ensemble of olivine forms a polycrystalline aggregate, their CPO can be described in terms of a hexagonally symmetric medium with a distinct fast and slow axes (e.g. Montagner and Nataf, 1988). Pyroxene, another naturally occuring mineral in the mantle, is also intrinsically anisotropic. Along with olivine, both constituents contribute to the bulk anisotropy in periodotite (Estey and Douglas, 1986) although numerical simulations of texture evolution suggest that the incorporation of pyroxene may impede the net alignment of the minerals. This results in a net decrease in the overall anisotropy due to the difference in their single crystal properties and their individual responses to a macroscopic deformation (Kaminski, Ribe, and Browaeys, 2004). Although such effects are noticeable, olivine is still twice as abundant as pyroxene and thus has a greater contribution to the large scale anisotropy in the mantle (Babuska and Cara, 1991).

Observations of large-scale anisotropy in tomographic models appear to be ubiquitous at patches associated with strong deformation, and thus have often been interpreted in terms of convective flow (McKenzie, 1979). For instance, tomographic imaging has revealed the presence of positive radial anisotropy (*i.e.*, horizontally propagating $SH-$waves traveling faster than $SV-$waves) of about 4% in the upper $\sim$ 250 km of the mantle and has been interpreted as lateral flow (refer to Long and Becker (2010) for a comprehensive review). Long wavelength seismic anisotropy is also prevalent in the transition zone as documented by some studies (Trampert and Heijst, 2002; Wookey and Kendall, 2004) although its origin is still highly debatable up to this day (Chen and Brudzinski, 2003; Chang and Ferreira, 2019; Sturgeon et al., 2019). Probing deeper depths, the lower mantle appears to be isotropic (e.g. Meade, Silver, and Kaneshima, 1995) barring the D" layer where enough evidence have shown it to be anisotropic (e.g. Kendall and Silver, 1998; McNamara, Keken, and Karato, 2002; Panning and Romanowicz, 2006).

It is worth mentioning however that since CPO anisotropy maps the deformation patterns, CPO may deviate far from the flow direction. This is because the deformation patterns relate not to the velocity field, but to the velocity gradient itself. Moreover,

CPO anisotropy is not instantaneous, but depends on the history of the deformation. As a result, regions with short deformation trajectories such as mid-ocean ridges appear to have underdeveloped CPO anisotropy, and thus would lag behind the direction of shear deformation (Kaminski and Ribe, 2002).

Different proxies have then been utilised to interpret seismic anisotropy directly in terms of mantle flow. First-order seismic observations suggest that the fast axis of azimuthal anisotropy tends to align with horizontal mantle flow (Ribe, 1989; Becker et al., 2003; Becker et al., 2014). However, this behavior may not always be exhibited due to complex local deformation mechanisms associated with CPO evolution. Laboratory experiments have then been performed to mimic such observations. Simple shear experiments suggest that, at low strains, the orientation of the olivine fast axis tends to be aligned with the long axis of the finite strain ellipsoid (FSE) (Zhang and Karato, 1995; Ribe, 1992). The amplitude of anisotropy, on the other hand, can be approximated as a monotonic function relating to the ratio between the long axis and the short axis of the FSE (Ribe, 1992; Hedjazian and Kaminski, 2014). At sufficiently large strains however, CPO evolution deviates from the FSE due to the apparition of dynamic recrystallisation. It tends to align nearly parallel to the direction of shear instead (Zhang and Karato, 1995; Bystricky et al., 2000), although its transient behavior remains complex (Hansen et al., 2014a). Following this observation, a possible proxy is to interpret the orientation of the anisotropy fast axis as the infinite strain axis (ISA), that is, the axis of the FSE in the limit of infinite strains (Kaminski and Ribe, 2002). In practice however, this proxy have had limited success at the global scale (Becker et al., 2014).

1.3.2 The story of a convecting mantle as told by seismic tomography

As previously discussed, large-scale convective flow in the mantle manifest themselves in seismic observables through seismic anisotropy. Initially undeformed mantle minerals develop preferential orientations once they undergo straining as they get advected along the flow trajectory. Since the time scales associated with seismic wave propagation are tremendously lower than the development of CPO anisotropy, and thus the deformation history of the mantle, seismic waves only record the present-day deformation in the mantle which then precludes the evolution history of mantle dynamics. Still, seismology provides a great deal of information about some aspects of mantle dynamics such as the very nature of imbricated convection, and are able to constrain deformation patterns responsible for tectonic plate motions and possibly the existence of hot mantle plumes.

Various upper mantle studies are in concord that the recovered structures from seismic tomography are consistent with the scale of the convective patterns. The surface manifestation of these patterns are in close agreement with different geological features observed at the surface and that they strongly influence the distribution of continents and oceans. As mentioned earlier, isotropic arrival time tomography revealed the penetration of subducting slabs down to the depths of the lower mantle while some appear to stagnate across or below the 660 km transition zone. In the context of mantle dynamics, Hilst and Seno (1993) and Hilst (1995) argued that the conditions whether slabs penetrate deep into the lower mantle or stagnate across the transition zone is an interplay between lateral trench migration (*i.e.* relative plate motions) and slab deformation associated with viscosity variations and/or recession of phase boundaries. To confirm this hypothesis, numerical models of mantle flow have been utilised (e.g. Zhong and Gurnis, 1995) to compare this with the tomographic results. Indeed, adding constraints from geodynamic modeling have implied that the complex flow patterns associated with slab stagnation are a result of various geophysical processes such as plate coupling and viscosity stratification, and that surface plate motions are indeed linked to large-scale convection in the mantle (Kárason and Van Der Hilst, 2000).

In retrospect, none of the studies mentioned above however have integrated computational geodynamic modeling with seismic imaging. Even still, some aspects of mantle dynamics in hindsight remains enigmatic such as the demise of slabs as they penetrate across the transition zone. As suggested by Montagner (1998), the incorporation of seismic anisotropy would allow the complete mapping of mantle deformation beneath trenches which may address the problem mentioned, and perhaps provide us a clear understanding on the relation between the return flow adjacent to the slab and the existence of hot plumes.

Supposing that we only have information provided by isotropic tomography, it is still plausible to directly incorporate seismic velocity distribution as inputs in geodynamic forward models. As mentioned in the first few sections, flow in the mantle is driven by density anomalies. These anomalies stem from temperature variations in the mantle. Intuitively, we know that temperature is inversely proportional to the strength of materials (*i.e.* cold anomalies tend to be more rigid that warm ones). Of course, rudimentary physics is not enough and a more rigorous proof is needed to model such behavior. Thankfully, semi-empirical models that allow us to estimate seismic velocities from temperature for a given bulk composition exist. These models were constrained based on first principles of thermodynamics (*i.e.* free energy minimisation) and from petrological experiments. It is important to keep in mind however that such techniques are valid if we ignore the effect of petrological heterogeneities and that a

more thorough method involves modeling the dependence of seismic wave velocities onto composition. Seismic tomography also has the ability to constrain the Rayleigh number Ra. As we will see later, Ra depends on various parameters and one of them is the scale of the convection L_s. Here, L_s can also pertain to the length scale of the tomographic region considered. Tomographic models can thus be viewed as added geometrical constraints onto convective flow in the mantle.

The manifestation of seismic anisotropy onto seismic observables is guaranteed most of the time especially with surface waves where they sample the anisotropic structure of the upper mantle. Because of this, seismic anisotropy has to be integrated in the geodynamic forward problem. Doing so will also be easier to delineate complex flow patterns that cannot be explained by isotropic tomography alone. In order to explain surface wave anisotropy, particularly in intra-oceanic and young continental regions where the flow appears to be in steady-state, first-order interpretations involve finite strains computed from global circulation models (Becker et al., 2003). In their work, the density field derived from isotropic tomography (Becker and Boschi, 2002) is used to compute instantaneous flow solutions in the upper mantle assuming a homogeneous bulk composition. Finite strain models derived from the flow are subsequently compared with azimuthal anisotropy in surface waves. However, as discussed before, finite strain-derived models may fall short at larger strains due to complex deformation mechanisms such as dynamic recrystallisation (Zhang and Karato, 1995). This urges the use of computational strategies that incorporate texture evolution models to estimate the level of intrinsic anisotropy.

Mechanical models of texture evolution, coupled with geodynamic flow modeling have been developed to replicate the laboratory results and have been extrapolated at scales consistent with upper mantle deformation patterns. Among these is the viscoplastic self-consistent (VPSC) model which is used to explain the mechanical response of polycrystals from plastic deformation (Tommasi et al., 2000). Such tools however are computationally expensive, especially when applied to 3D and non-steady state flows (Lev and Hager, 2008). Another well-received method is a simple kinematic approach that utilises an average field formalism to quantify the net orientation of a polycrystal under an imposed macroscopic deformation (Kaminski, Ribe, and Browaeys, 2004). In their model, the mineral assemblage responds to the deformation by plastic deformation and dynamic recrystallisation. Texture evolution models input the deformation gradient tensor computed from the geodynamic flow field. The output can either be expressed as a function describing the volume fraction and the orientation of the crystals, or a Voigt-averaged elastic tensor with 21-independent coefficients (Mainprice, 1990). The latter can then be used as a reference medium for which seismic

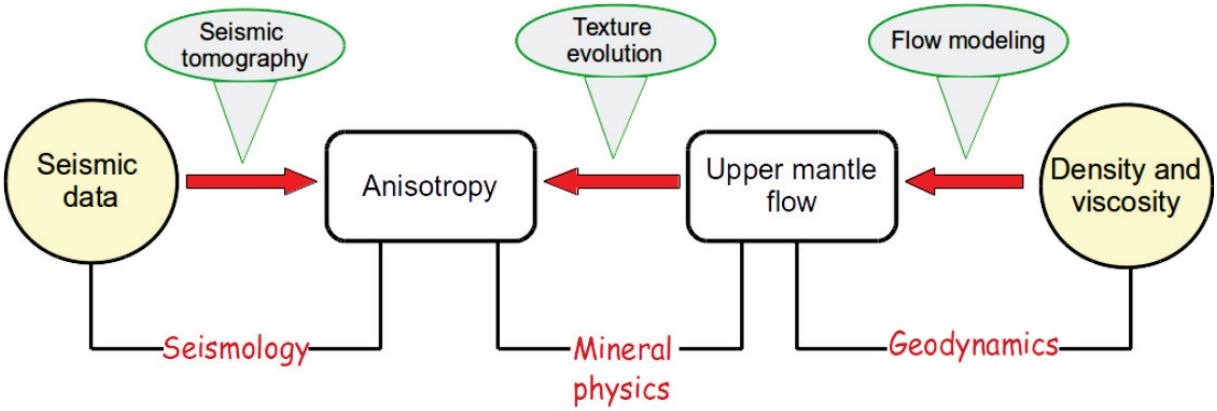

FIGURE 1.8: The story of a convecting mantle as told by seismic tomography. The outline reads as follows: seismic data can be inverted to retrieve the anisotropic structure of the mantle (seismic tomography). Texture evolution modeling on the other hand is utilised to retrieve the same anisotropic structure, but starting from mantle flow models. Lastly, mantle flow models are obtained in a full forward modeling approach with the density field and viscosity fields as inputs. Most tomographic studies and texture evolution modeling schemes cease at the recovery of anisotropy (second box). The latter serves as the point of interest between two disciplines (seismology and upper mantle dynamics) where most geophysicists rely on visual comparison of the anisotropy reconstructed from the two methods (*i.e.* seismic tomography and texture evolution modeling). If we imagine an arrow pointing towards density and viscosity starting from seismic data, we thus have geodynamic tomography.

waves can propagate into, and thus explain anisotropic signatures observed in seismic data recorded at the surface.

Texture evolution models have been extensively applied to predict CPO-induced anisotropy from geodynamic flow models in a forward modeling approach at the regional (Hall et al., 2000; Lassak et al., 2006; Miller and Becker, 2012; Faccenda and Capitanio, 2013) and at the global scale (Becker et al., 2006; Becker, Kustowski, and Ekström, 2008). Forward models such as these assist further in the interpretation of seismic tomography models in terms of mantle circulation patterns. For instance, CPO-induced anisotropy resulting from to 3D numerical simulations of subducting slabs shows consistency with radial anisotropy patterns inferred from global tomographic images (Ferreira et al., 2019; Sturgeon et al., 2019).

In summary, most studies rely on visual comparisons between CPO anisotropy obtained from numerical simulations and tomographic images. To the best of our knowledge, no study yet exists where mantle deformation has been inferred directly from seismic observations using an inverse approach.

1.4 Geodynamic tomography

This motivated us to implement a novel approach to the seismic tomography problem, called *Geodynamic Tomography*, where no symmetry is imposed to the elastic tensor at the outset, and where seismic observations are inverted with constraints from geodynamic modeling and petrological modeling. Although the word 'seismic' from seismic tomography indicates the type of data used in the inversion, we chose 'geodynamic' to stipulate the significance of geodynamic flow modeling in limiting the search for admissible tomographic models. The ultimate goal of the method is to map the complete deformation patterns associated with mantle convection, as opposed to conventional seismic tomography where the process ends at inferring the Earth's elastic structure.

To constrain the patterns of mantle deformation, we jointly invert Love and azimuthally-varying Rayleigh phase velocity dispersion curves to retrieve the present-day 3D thermal structure of the upper mantle. The thermal structure relates to density anomalies through a linear equation of state. Given the density anomalies, one may then be able to compute convective flow in the mantle that matches the seismic predictions observed at the surface through deformation-induced CPO anisotropy.

The complete forward problem proceeds as follows: (1) Given a temperature field, we first numerically solve an instantaneous 3D convection problem with temperature-dependent viscosity (Samuel, 2012a). (2) Using the obtained velocity field and velocity gradient obtained, we track CPO evolution of olivine crystals (Kaminski, Ribe, and Browaeys, 2004) where the steady-state assumption of the flow is implied. This computes for the anisotropic part of the elastic tensor. (3) From the temperature field (and the hydrostatic pressure), we derive the pressure and temperature dependence of the isotropic part using a thermodynamic model for a given bulk composition (Connolly, 2005; Connolly, 2009; Stixrude and Lithgow-Bertelloni, 2011). The result from (2) and (3) is a complete elastic tensor S_{ij} at each point in space. (3) The last step involves computing synthetic surface wave dispersion curves using normal mode summation in a spherical earth (Smith and Dahlen, 1973) and their azimuthal variations from the full S_{ij} (Montagner and Nataf, 1986).

To efficiently look for candidate models that best explain the data, the inversion explores the parameter space using a Markov chain Monte Carlo (McMC) algorithm, and evaluates through Bayesian inference the posterior probability of model parameters. In opposition to conventional tomography where elastic parameters are to be inverted for, our method directly inverts for a single scalar field (*e.g.*, temperature anomalies) and extra information is driven by the physics of mantle convection. The complete solution to our problem is a probability distribution of the 3D present-day thermal

structure of the upper mantle. Since the complete elastic tensor is computed for each sampled model, we can also obtain a posterior distribution of the full elastic tensor. In fact, any variable that is implicitly computed in the forward model can be expressed as a posterior distribution in their respective model space (temperature, flow, deformation, and anisotropy). Thus, geodynamic tomography may be viewed as a technique to reduce model dimension (*i.e.*, the number of inverted parameters) in the inverse problem. Our goal in this study is to lay its proof of concept by applying it to simple synthetic temperature fields. We base our conclusions on the quality of the recovered structures.

1.5 Structure of the manuscript

This thesis is organised in such a way that the reader will chronologically follow all the work I have accomplished during my stint as a PhD candidate. The next chapter covers an introduction to Bayesian methods. Chapter 3 lays the foundation of the full forward problem to geodynamic tomography. Chapters 4 and 5 present synthetic experiments. Chapter 6 covers the concluding remarks and future perspectives of the method. The last chapter reports the project I have done on the side.

Bayesian Inference in Surface Wave Tomography
This chapter presents a detailed description of Bayesian inference and the Markov chain Monte Carlo (McMC) algorithm. As a simple demonstration, we perform conventional anisotropic surface tomography with Bayesian inference to synthetic data generated by a 3-D deforming upper mantle due to a sinking, spherical anomaly. We impose strong *a priori* constraints to the elastic structure by prescribing the correct values associated with the $P-$wave structure. The purpose of this synthetic experiment is that even when placed in a favorable scenario (*e.g.*, some of the elastic properties are well-known), conventional surface wave tomography is stil hampered by several issues such as the dependence on the choice of parameterisation and its inability to resolve sharp gradients in the elastic structure.

Geodynamic Tomography: The Forward Problem
The previous chapter serves as motivation to reformulate the tomographic problem by incorporating geodynamic and petrological constraints to reduce the number of candidate models down to a subset consistent with mantle dynamics. As such, this chapter unravels in detail the full forward problem to geodynamic tomography which consists of several intimately related sub-forward models. We present how one could arrive at seismic data (*i.e.*, surface wave dispersion measurements) starting from a scalar field

(*i.e.* temperature). In the process, the full elastic tensor can be implicitly computed everywhere thus solving one of the prevailing problems associated with conventional surface wave tomography.

Geodynamic Tomography: Joint Inversion of Love and Rayleigh Waves with Azimuthal Variations

This chapter presents the results of geodynamic tomography through synthetic experiments. We applied a Bayesian Monte Carlo approach to jointly invert Love and aniotropic Rayleigh wave dispersion measurements for an ensemble of temperature fields. The synthetic data were generated coming from a true model corresponding to a 3-D deforming upper mantle due to a temperature field made of multiple spherical anomalies. The data is contaminated with random noise to test the ability of the method to recover the temperature field and quantify model uncertainties. We show how such a method implicitly retrieves the complete pattern of mantle deformation, and correspondingly, the full elastic tensor at every location.

Geodynamic Tomography: Application to a 3D Deforming Upper Mantle Beneath a Subduction Zone

We extend the method by applying it with a more physical parameterisation to test its competence in retrieving more complex deformation patterns. In this chapter, we consider synthetic data associated with 3-D instantaneous flows induced by subduction. We replicate realistic (*i.e.*, period-correlated) surface wave dispersion measurements by stipulating very low noise levels to the synthetic data.

Concluding Remarks and Perspectives

The achievements of the methodology is listed briefly in this chapter. We also discuss its limitations, and propose future avenues to delve into in order to address such limitations.

Quantifying Intrinsic and Extrinsic Contributions to Elastic Anisotropy Observed in Tomographic Models

The final chapter digresses from geodynamic tomography, however is still essential as it paves the way towards a new way to interpret tomographic models. Geodynamic tomography only accounts for the intrinsic component of anisotropy due to mantle deformation. In reality however, elastic structures reconstructed from long-period observations may also depict large-scale spurious anisotropy due to small-scale heterogeneities. The coexistence of both thus prompted us to assess their separation contributions in tomographic models for better interpretation. Here, we hypothesise that the effective anisotropy observed in tomographic models is simply the product between

the spurious component and the 'smooth' version of the intrinsic component. To test our hypothesis, we consider 2-D marble cake models of the mantle in the presence of intrinsic anisotropy.

1.6 Publication schedule

One of the chapters in this manuscript Chapter 4, (Magali et al., 2020) is published in *Geophysical Journal International*, Chapter 2 (Magali, 2020) is published as a preprint in `arXiv`, and lastly Chapter 6 is under preparation in pursuit of this thesis. Note that we plan to submit Chapter 6 in a peer-reviewed journal ahead of the dissertation defense schedule. Other than these three papers, I believe that some of the research work I have conducted have the potential to be published. Rest assured that these works will be revisited and that enough material will be produced fit for a publication.

Submitted as preprint:
Chapter 2: Magali, John Keith (2020). "1D Anisotropic Surface Wave Tomography with Bayesian Inference". arXiv:2012.03915 [physics.geo-ph].

Published in a peer-reviewed journal:
Chapter 4: Magali, JK et al. (2020). "Geodynamic Tomography: Constraining Upper Mantle Deformation Patterns from Bayesian Inversion of Surface Waves". In: Geophysical Journal International.

To be submitted to a peer-reviewed journal:
Chapter 6: Magali, JK et al. (2020). "Quantifying Intrinsic and Extrinsic Contributions to Elastic Anisotropy Observed in Tomographic Models". In: Geophysical Journal International.

Chapter 2

Bayesian Inference in Surface Wave Tomography

1D Anisotropic Surface Wave Tomography with Bayesian Inference
Submitted as a preprint in *arXiv*
Can be accessed at `https://arxiv.org/abs/2012.03915`
J.K. Magali[1]
[1] Université de Lyon, UCBL, CNRS, LGL-TPE, 69622 Villeurbanne, France

2.1 Summary

Classically, anisotropic surface wave tomography is treated as an optimisation problem where it proceeds through a linearised two-step approach. It involves the construction of 2D group or phase velocity maps for each considered period, followed by the inversion of local dispersion curves inferred from these maps for 1D depth-functions of the elastic parameters. Here, we cast the second step into a fully Bayesian probability framework. Solutions to the inverse problem are thus an ensemble of model parameters (*i.e.* 1D elastic structures) distributed according to a posterior probability density function and their corresponding uncertainty limits. The method is applied to azimuthally-varying synthetic surface wave dispersion curves generated by a 3-D-deforming upper mantle due to a sinking spherical anomaly. We show that such a procedure captures essential features of the upper mantle structure. The robustness of these features however strongly depend on the wavelength of the wavefield considered and the choice of the model parameterisation. Additional information should therefore be incorporated to regularise the problem such as the imposition of petrological constraints to match the geodynamic predictions.

2.2 Introduction

Conventional surface wave tomography is usually implemented using a two-step approach (e.g Nakanishi and Anderson, 1983; Nataf, Nakanishi, and Anderson, 1984; Trampert and Woodhouse, 1995; Romanowicz, 2002; Ritzwoller et al., 2002). The first step involves the inversion of the arrival times of each period considered in the measured source-receiver dispersion data to infer 2D group or phase velocity maps at a given period. Using the 2D velocity maps, the second step proceeds by inverting a dispersion curve at a given geographical location to estimate the 1D velocity structure beneath this location. One may then build a smooth 3-D velocity model by the juxtaposition of the inferred 1D models followed by interpolating them.

The tomography problem is often solved by applying first-order corrections of the forward function g around a reference model $\mathbf{m}_0$. Mathematically. this translates to:

$$\mathbf{d} = g(\mathbf{m_0}) + \frac{\partial g}{\partial m}\Delta\mathbf{m}. \tag{2.1}$$

Doing so allows it to be treated as a linearised inverse problem. For instance, one may estimate the 2D phase velocity maps by minimizing an objective function containing a data residual term and more than one regularization terms such as:

$$S = ||\mathbf{Gm} - \mathbf{d}|| + R_1||\mathbf{m}|| + R_2||\mathbf{Dm}|| \tag{2.2}$$

where $G = \frac{\partial g}{\partial m}$ is now a mathematical forward operator which now refers to the linearised physics between the model parameters (in this case surface wave velocities), and the data (dispersion data), D is a second-order smoothing operator, and R_1 and R_2 are the damping and smoothing parameters, respectively. As mentioned earlier, the last two parameters will often dictate the trade-off between the data misfit (*i.e.*, how well the model parameters predict the data), the proximity of the estimated model from its reference state, and the degree of smoothing in the inverted model. Moreover, the regularization parameters are chosen ad hoc. Hence, the inverted model may be susceptible to non-data driven constraints and that these constraints shroud essential information provided by the data. Lastly, optimization techniques such as this lack the capability to estimate model uncertainties.

The problems associated with non-uniqueness and quantification of uncertainties, coupled by the ever-growing computational capacity of modern supercomputers led to the advancement of probabilistic approaches to geophysical inverse problems, as first exemplified by Mosegaard and Tarantola (1995). Following this pioneering study,

a volume of studies that involve probabilistic approaches to geophysical inverse problems have been published in seismology (e.g. Lomax et al., 2000; Shapiro and Ritzwoller, 2002; Husen et al., 2003; Bodin and Sambridge, 2009; Debski, 2010; Bodin et al., 2016), and rapidly growing in the field of geodynamics (e.g. Baumann, Kaus, and Popov, 2014; Baumann and Kaus, 2015; Morishige and Kuwatani, 2020; Ortega-Gelabert et al., 2020).

Casting the inverse problem in a probabilistic framework allows one to utilise the original non-linear mapping g between the data and the model within its forward procedure. However due to the use of sampling-based methods, one is then forced to solve the forward model numerous times depending on the number of candidate models to be sampled. Fortunately, various techniques are available to address such complications that are not solely based on heuristics. Here, we restrict ourselves with Bayesian inference, that is, a form of statistical inference that formulates our solution as an *a posteriori* probability initially based on the information we have prior to evaluating the inverse problem. The goal is therefore not to create an ensemble of solutions that follow a certain probability distribution *ex nihilo*, but to update a prior probability based on valuable information provided by the data.

In such schemes, the parameter space has to be explored for best possible model candidates that could match the predictions observed at the surface. Grid-search algorithms however are time consuming and thus uniform sampling may not be performed efficiently. As such, direct-search algorithms have been introduced in geophysical inverse problems that sample candidate models within a subset instead of the entire parameter space. A specific class of ergodic algorithms is called Markov chain Monte Carlo (McMC) methods, which has been initially used to solve problems in physics (Metropolis and Ulam, 1949; Metropolis et al., 1953), but is now widely used in geophysical inverse problems. In this algorithm, the parameter space is randomly sampled based on our current state of knowledge of a given model candidate. The sequence of searching new model candidates depend on an acceptance probability that is often determined by satisfying a detailed balance condition to ensure stationarity of the desired solution, the most common being the Metropolis-Hastings algorithm (Hastings, 1970).

In this chapter, we cast the second step of the surface wave tomography problem in a Bayesian framework. That is, we assume that we have completely inferred 2D phase velocity maps of the regions considered, and invert for 1D velocity structures that explain the dispersion curves built from these maps. The problem is applied to synthetic data for each geographical location considered whose anisotropic signatures are solely influenced by convective flow in the mantle beneath it. The solution is a

marginal posterior distribution of 1D velocity models that best explain the data accompanied by their uncertainty limits. We show that even with strong *a priori* constraints, conventional surface tomography falls short to capture the complete picture related to mantle deformation. Still, some of the notable features are resolved more or less. This exercise builds upon the hypothesis that adding geodynamical and petrological constraints would allows us to reduce the number of acceptable tomographic models that are consistent with the geodynamical predictions.

2.3 Bayesian inference

As opposed to learning based on heuristics, Bayesian framework is a methodical process of logical reasoning that relies on integrating new information with prior beliefs. It emancipated in the late 1700's, thanks to its proponent Thomas Bayes, and still continues today as a mode of learning in decision theory. Bayes' theorem has been successfully applied in various areas of discipline, particularly as a method in statistical inference. In Bayesian inference, valuable information is extracted from a probability distribution *ex post facto* based on evidence.

Bayesian inference offers a practical way to solve the inverse problem by recasting the ensemble of model solutions $\mathbf{m}$ as a posterior probability distribution. The posterior probability is formulated by updating an initial probability distribution that already contains information prior to the inversion process, called the prior distribution or prior, with new information provided by the data d, called the likelihood (Box and Tiao, 2011). Mathematically, Bayes' theorem writes:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})} \tag{2.3}$$

where for any measured quantity A and B, $p(A|B)$ reads as *the probability of A knowing B*. Here, $p(\mathbf{m}|\mathbf{d})$ is the posterior distribution, that is, what we know about the model parameters given the data. $p(\mathbf{d}|\mathbf{m})$ is the likelihood, that is, the probability of observing the data for a given model. $p(\mathbf{m})$ is the prior, or the distribution of allowable models prior to the inversion process. Lastly, $p(\mathbf{d})$ is the marginal likelihood also called the evidence. In most cases, and in our work hereafter, we regard the marginal likelihood as a constant since it does not depend on any given model. As such, we can reformulate eq. (2.3) as:

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \tag{2.4}$$

It is worth noting that these functions are continuous distributions, and that the result of the inverse problem is instead a finite amount of models distributed according to the target distribution $p(\mathbf{m}|\mathbf{d})$. The main challenge is therefore to generate samples that are ergodic in nature to approximate the posterior distribution. This calls for the need to implement direct-searching algorithms that efficiently samples complex probability distributions. One such technique, as previously mentioned, is the Markov chain Monte Carlo (McMC) algorithm, where model updates **m'** solely depends on its current state **m**, and any information prior to the realization of **m** is erased in **m'**.

2.3.1 Likelihood function

The likelihood function $p(\mathbf{d}_{obs} \mid \mathbf{m})$ quantifies how well the model parameters fit the observed data. In the context of our problem, it is loosely based on the L^2-norm cost function in that it measures the level of misfit between the predictions and the observations. Here, it is essential to distinguish the data residuals related from errors in measurement $\boldsymbol{\epsilon_d}$, and from the modeling error due to the use of an incorrect forward model $\boldsymbol{\epsilon_g}$. Assuming that the errors are independent and describe a random process, the forward problem can be written as:

$$\mathbf{d_{obs}} = g(\mathbf{m}) + \boldsymbol{\epsilon_d} + \boldsymbol{\epsilon_g}. \tag{2.5}$$

Thus, $\mathbf{d_{obs}}$ can also be seen as a random process, and the likelihood distribution can be formulated in terms of the pdf of the data errors:

$$p(\mathbf{d_{obs}}|\mathbf{m}) = p(\boldsymbol{\epsilon_d} + \boldsymbol{\epsilon_g}). \tag{2.6}$$

If we then assume that the errors are uncorrelated and follow a univariate Gaussian distribution with zero mean, and variance σ^2 where $\sigma^2 = \sigma_d^2 + \sigma_g^2$, we can write the likelihood function as an exponential giving:

$$p(\mathbf{d_{obs}}|\mathbf{m}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[\frac{-||\mathbf{d_{obs}} - g(\mathbf{m})||^2}{2\sigma^2}\right]. \tag{2.7}$$

where N is the size of the data vector. Since the goal of any optimization problem is to minimise the L^2 cost function, minimising it tantamounts to maximising the probability of the Gaussian likelihood function given by eq. (2.7).

2.3.2 Prior distribution

One of the flexibilities of the Bayesian framework is that it enables one to account for prior information provided that it can be formulated as a probability distribution. The choice of the prior depends on the type of geophysical process we aim to tackle. In the context of seismic tomography, the prior information depends on a range velocity structures that are reasonable for the Earth in general. In practice, the prior information is constrained by existing studies.

At this point forward, we assume minimal prior knowledge and hence, make use of uniform prior distributions with wide bounds. Although we acknowledge that using uniform distributions may be a naive way to setup the prior, working with such a simple distribution would already suffice when demonstrating proofs of concept. Here, we know the exact values of the model parameters, and by imposing wide uniform priors, we are able to assess the efficiency of the method by placing ourselves in the worst case scenario. Indeed, a subject of future work is to consider other forms of the prior distribution, for example non-informative priors or even hierarchical Bayes (Malinverno and Briggs, 2004).

Let us now consider a given model parameter m_i. The prior $p(m_i)$ is prescribed a constant value over a given range of values defined by $[m_{\min}, m_{\max}]$. The prior distribution is thus given by:

$$p(m_i) = \begin{cases} 0 & m_i > m_{\max}, m_i < m_{\min} \\ \frac{1}{\Delta m} & m_{\min} \leq m_i \leq m_{\max}. \end{cases} \tag{2.8}$$

Eq. (2.8) is interpreted as follows. Suppose that we draw a sample for the specific model parameter m_i from the proposal distribution q. If m_i is out of bounds, then the proposal is automatically rejected because the value is not specified by the prior. If m_i is within the prior bounds, then the proposal is accepted with condition based on the the acceptance probability A. Choosing narrow bounds therefore imposes hard constraints to the model parameters giving less emphasis to the information provided by the data.

2.4 Markov chain Monte Carlo algorithm

In the context of a probabilistic inverse problem, McMC is a class of iterative stochastic algorithms used to efficiently sample the parameter space. As opposed to uniform sampling methods, McMC ultimately tends to sample towards a restricted area of high

probability density and operates according to a random walk-type behavior. The algorithm proceeds as follows. An initial model $\mathbf{m}_0$ is randomly drawn from the prior distribution $p(\mathbf{m})$. A perturbation around $\mathbf{m}_0$ according to a proposal probability gives birth to a new model $\mathbf{m}$ within the Markov chain which will then be accepted conditionally based on an acceptance probability. Once a statistically significant amount of models are sampled, known as the burn-in phase, the random walk develops an importance sampling of the parameter space. In this state, the ensemble of models begin to be distributed according to the posterior, and that the random-walk behavior allows for further refinement of the approximation. At this point, the Markov chain is said to be converged. When implemented in parallel with different starting models, the converged chains are also said to be well-mixed.

2.4.1 The proposal distribution

At this point forward, we will make use of Gaussian probability distributions as the proposal probability of choice. Note that the choice of the proposal probability will not affect the nature of the posterior pdf. However, such ad hoc choices may be held accountable to some computational lapses associated with the algorithm's efficiency and convergence. Formally, the proposal probability for a given model parameter m_i can be expressed as a univariate normal distribution with zero mean and standard deviation σ_i assuming uncorrelated model residuals:

$$q(m_i|m_i') \propto \exp\left[-\frac{||m_i - m_i'||^2}{2\sigma_i^2}\right]. \tag{2.9}$$

Conventionally, the proposed model m_i' can be treated as a Gaussian perturbation centered at the current model m_i giving:

$$m_i' = m_i + N(0, \sigma_i) \tag{2.10}$$

where N is a normal distribution. Here, σ_i can be regarded as the degree of perturbation. Take note that using small perturbations avoids drawing samples far from the current model, leading to small changes in the likelihood function. As such, the proposal is most likely to be accepted. Likewise, choosing large perturbations will most likely be rejected as it means larger variations in the likelihood function. Regardless, both situations restrain the movement of the sampler when searching the parameter space, and that the candidate models may strongly depend on their initial state. The game is therefore to choose perturbations that strike balance between the acceptance

and rejection rates. As a rule of thumb, the acceptance rates should be around 30-40% for more efficient sampling (Sambridge and Gallagher, 2011).

2.4.2 An adaptive perturbation scheme

Throughout the course of my PhD thesis that covers McMC inversions, I instead opted for a more dynamic perturbation, that is, the perturbations vary depending on a given situation. Here, we employ the McMC sampler with an adaptive perturbation scheme based on the acceptance rate. This requires keeping track of the acceptance rate for a given model parameter on the fly. Let us now denote N to be the population size, that is, total number of samples in the inversion, and M to be the total number of accepted models within N. The acceptance rate corresponding to the entire population is just:

$$\text{acceptance} = \frac{M}{N} \times 100\% \tag{2.11}$$

Next, we need a sizeable amount of samples n within N to allow for the relaxation of the acceptance rates, that is, the period at which the acceptance rates are in stable conditions. The population N are then separated into different cycles (*i.e.*, sample window) that are multiples of n. For instance, choosing $n = N$ just pertains to the acceptance rates of the entire population, whereas choosing $n = 1$ would constantly reset the counters for the proposal and the accepted models at every iteration. At the end of every cycle, the current values of the acceptance rates are then used to determine whether to increase or decrease the perturbation. If the acceptance is less than 20%, then the perturbations will be reduced by 25% of its current value. Likewise, if the acceptance is more than 50%, then the perturbations will be increased by 25% instead. The counters for the proposal and the acceptance are then reset, and the new acceptance rate is recalculated in the subsequent cycle. As an example, suppose that we have $N = 50000$ samples, and a subset of $n = 5000$, we thus having ten cycles. After the first 5000 iterations l (first cycle), the acceptance rate at $l = 5001$ will determine whether to increase or decrease the perturbations by 25%. Resetting of the counters will then be ensued regardless. The appraisal of the acceptance rates are then marked after every n intervals,that is, at $l = 10000, 15000, 20000$ and so on.

2.4.3 The acceptance probability

The transition to the new state is decided using the Metropolis-Hastings algorithm (Hastings, 1970). The process elaborates as follows. Suppose we have a population of size N with each iteration denoted by l. First, we initialise our sampler by drawing an

initial model $\mathbf{m}_0$ from the prior, and setting $l = 0$. We then proceed with the iterative process. We propose a new model $\mathbf{m'}$ from the proposal distribution $q(\mathbf{m}_l \mid \mathbf{m'})$. Once the new model is drawn, the transition is accepted or rejected with condition depending on an acceptance probability $A(\mathbf{m}_l \mid \mathbf{m'})$. The acceptance probability is computed using the following expression:

$$A(\mathbf{m_l}|\mathbf{m}') = \min\left(1, \frac{p(\mathbf{m}'|\mathbf{d})}{p(\mathbf{m}_l|\mathbf{m}')} \times \frac{q(\mathbf{m_l}|\mathbf{m}')}{q(\mathbf{m'}|\mathbf{m_l})}\right). \tag{2.12}$$

Next, a random number is generated between 0 and 1 from a uniform probability distribution $u \in [0, 1]$. The proposal is accepted, that is $\mathbf{m}_{l+1} = \mathbf{m'}$, if $u \leq A(\mathbf{m}_l \mid \mathbf{m'})$. It is rejected otherwise if $u > A(\mathbf{m}_l \mid \mathbf{m'})$. In this case, the old state is copied in the next state, that is, $\mathbf{m}_{l+1} = \mathbf{m}_l$. Finally, l is updated for the next iteration $l = l + 1$.

Generating samples using McMC is indeed efficient. However, one must be cautious because the models sampled by McMC every l iterations are different from drawing l independent samples from the posterior pdf. This is based from the very nature of the Markov chain itself wherein the samples are correlated. This requires a burn-in period to be implemented, that is, the period at which the sampler moves erratically. After disregarding the burn-in period, we should expect the random-walk to be in a steady-state condition. Only then when the chains can be regarded as quasi-independent. At this point, we could start recording the samples generated by McMC, and hence, infer a posterior distribution out of them. Another path we could take which facilitates drawing independent samples is to parallelise the algorithm.

2.5 Synthetic experiment

In this section, we discuss the full implementation of a 1D anisotropic surface wave tomography in a full Bayesian parameter search approach. We highlight in full detail the (1) model parameterisation, (2) the forward problem, (3) the data, and (4) the inverse approach. The method is applied to synthetic surface wave dispersion curves produced by simple setups of an intrinsically anisotropic upper mantle.

2.5.1 Model parameterisation of a 1D Earth structure

Radial anisotropy component

Surface waves are sensitive to 13 depth parameters which are just a linear combination of the full elastic tensor S_{ij} (Montagner and Nataf, 1986). In particular, Love and

Rayleigh phase velocities are sensitive to five depth depth parameters which make up an azimuthally-averaged vertically transverse isotropic (VTI) medium. These parameters are also known as the Love parameters, and by convention, are designated as A, C, F, L, and N. Note that these functions are constrained by the isotropic phase velocities, and are independent of the azimuth of surface wave anisotropy. As a supplementary, the seismic wave velocities propagating either parallel or perpendicular to the symmetry axis can be written as:

$$V_{PH} = \sqrt{\frac{A}{\rho}} \tag{2.13}$$

$$V_{PV} = \sqrt{\frac{C}{\rho}} \tag{2.14}$$

$$V_{SH} = \sqrt{\frac{N}{\rho}} \tag{2.15}$$

$$V_{SV} = \sqrt{\frac{L}{\rho}}, \tag{2.16}$$

where ρ is the density of the medium. Here, it is important to understand that the vertical symmetry axis need not be the fast axis of anisotropy. Hence, the relative magnitude between N and L, and C and A are interchangeable. Most anisotropic tomography studies interpret $L > N$ as vertical flow, since to first-order, the direction of shear is presumed to be vertical and thus aligned with vertically propagating $S-$waves (e.g. Montagner, 1994). Likewise, $L < N$ is often interpreted in terms of horizontal flow. When the flow has both horizontal and vertical components, then the resulting anisotropy will be ambiguous. This requires resolving the tilt of anisotropy (Montagner and Nataf, 1988; Montagner and Jobert, 1988).

We constrain radial anisotropy by using a more compact form related to the Love parameters. For $P-$waves, the strength of radial anisotropy can be expressed as $\phi = C/A$, whereas for $S-$waves, it is given by $\xi = N/L$. Finally, there exists another anisotropic parameter which relates to the ellipticity $\eta = F/A - 2L$. The phenomenology of η can be understood through the parameter F which controls the velocity along the direction between the fast and slow velocities.

Azimuthal anisotropy component

Assuming a slightly anisotropic medium, azimuthal anisotropy in surface waves can be decomposed into two terms that depend on its azimuth θ. These two terms are usually called the 2θ and 4θ components, and are small perturbations around the isotropic phase velocities. The 2θ and 4θ components are sensitive to eight depth functions 2θ: $G_s, G_c, B_s, B_c, H_s, H_c$, and 4θ: C_s, C_c (Montagner and Nataf, 1986).

We will only work with azimuthal anisotropy in Rayleigh waves. Rayleigh waves are much more sensitive to the 2θ terms than the 4θ terms (Maupin and Park, 2015). Thus, to first-order, we could eliminate the parameters C_s and C_c from the inversions. Surface waves also poorly resolve the parameters H_s and H_c (Bodin et al., 2016). Thus, we could reduce the model dimensionality associated with azimuthal anisotropy down to four parameters G_s, G_c, B_s, and B_c.

Pseudo-regularisation

By using compact notations and first-order approximations, we are able to reduce the possible number of parameters to be inverted for. Four of which: A, L, ξ, ϕ, and η are sensitive to Love and Rayleigh phase velocities, and the remaining four: G_s, G_c, B_s, and B_c are sensitive to the azimuthal variations in Rayleigh phase velocities. Working with synthetic data allows us to access the correct values of the parameters defining our Earth model. As we wish to compare conventional tomography with geodynamic tomography, here, we will be placing ourselves in the best case scenario. We impose strong *a priori* constraints to our solution in the tomographic problem. These constraints can be regarded as regularisation parameters which limit the regions in the parameter space to search through. Here, we assume that we have the correct values relating to the $P-$wave structure A, ϕ, η, B_s, and B_c; and thus, only invert for $S-$wave-related structures L, ξ, G_c, and G_s. The list of parameters to/what not to invert for is summarized in Table 2.1.

The 1D Earth structure spans from the surface down to a depth of 400 km. It is subdivided into 62 layers of equal thicknesses where each layer has a given value of the nine depth functions (see Table 2.1). Thus, the total number of unknowns to be inverted for is $62 \times 4 = 248$. Bayesian inversion is therefore a suitable method to treat such an under-determined inverse problem. To reduce the model dimensions further, the mantle is parameterised using a piecewise cubic Hermite polynomials at fixed control points. In the inversions, we vary the unknown parameters L, ξ, G_c, and G_s at the control points and in-between these points, the parameters are interpolated. Below

TABLE 2.1: Depth functions constrained by surface waves and their azimuthal variations.

Parameter	Fixed	Inverted for
A	✓	
L		✓
ξ		✓
ϕ	✓	
η	✓	
G_c		✓
G_s		✓
B_c	✓	
B_s	✓	

the 400 km, we impose isotropic PREM (Dziewonski and Anderson, 1981). The 1D structure is illustrated in Fig. 2.1

The model vector is thus defined as:

$$\mathbf{m} = [\mathbf{L}, \boldsymbol{\xi}, \mathbf{G_s}, \mathbf{G_c}], \tag{2.17}$$

where bold faces indicate that each parameter is also a vector of size 62.

2.5.2 The forward problem

For each step in the McMC algorithm, the forward problem is evaluated using the proposed model (see Fig. 2.1) as input. The predicted data from this model is then compared with the observed synthetics.

Isotropic Rayleigh $\mathbf{c_R}(T)$ and Love $\mathbf{c_L}(T)$ dispersion curves are computed using normal mode summation in a spherical Earth (Smith and Dahlen, 1973). Here, the computations are carried out in a fully non-linear approach following the method developed by Saito (1967) and Saito (1988). The software package `DISPER80` (Saito, 1988) takes 1D depth profiles of V_p, V_s, ρ, ξ, ϕ, and η as inputs to compute for $\mathbf{c_R}(T)$ and $\mathbf{c_L}(T)$ and their associated sensitivity kernels using a Runge-Kutta matrix integration scheme.

Following the pioneering work of Montagner and Nataf (1986), the azimuthal variations in surface wave phase velocities $\mathbf{c_1}$ and $\mathbf{c_2}$ can be evaluated using the following

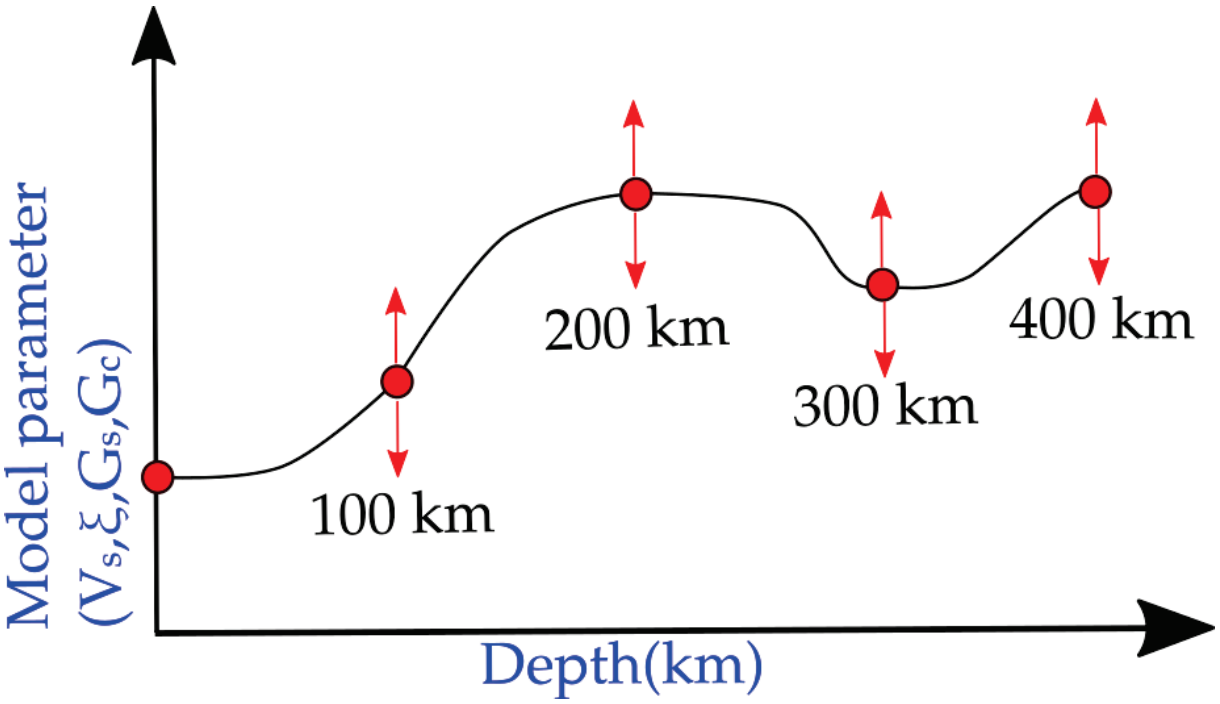

FIGURE 2.1: Schematic diagram of the 1D parameterisation. The entire region is parameterised with a piecewise cubic Hermite polynomials based on a number of control points. Here, the control points are fixed in depth. The model parameters are then varied at these points using an McMC sampling algorithm. The layers in between are interpolated, and anything below the 400 km is isotropic PREM.

expressions:

$$\mathbf{c_1}(T) = \int_{z=0}^{\infty} \left(B_c(z) \frac{\partial \mathbf{c_R}(T)}{\partial A} + G_c(z) \frac{\partial \mathbf{c_R}(T)}{\partial L} \right) dz \tag{2.18}$$

$$\mathbf{c_2}(T) = \int_{z=0}^{\infty} \left(B_s(z) \frac{\partial \mathbf{c_R}(T)}{\partial A} + G_s(z) \frac{\partial \mathbf{c_R}(T)}{\partial L} \right) dz. \tag{2.19}$$

Eqs (2.18) and (2.19) imply that the azimuthal variations in Rayleigh waves are linearized around the reference VTI model after averaging azimuthally. Such approximations are valid assuming the medium is quasi-isotropic (Montagner and Nataf, 1986; Maupin and Park, 2015).

2.5.3 The data

For each geographical location, we can express the local dispersion curve as the sum of the isotropic dispersion curves and their azimuthal variations giving:

$$\mathbf{c}(T, \theta) = \mathbf{c_0}(T) + \mathbf{c_1}(T) \cos(2\theta) + \mathbf{c_2}(T) \sin(2\theta), \tag{2.20}$$

where T is the period, and θ is the azimuth of the propagating surface wave.

For Rayleigh waves, we invert $\mathbf{c_0}$, $\mathbf{c_1}$, and $\mathbf{c_2}$ whereas for Love waves, we only invert $\mathbf{c_0}$. For simplicity, we neglect the higher-order terms associated with the elastic parameter N. such assumptions are valid due to sparse sampling, low sensitivity, or high noise levels.

2.5.4 Quantification of anisotropy

Anisotropic surface wave tomography is capable of constraining azimuthal and radial anisotropy. As previously discussed, the level of radial anisotropy can be quantified through the parameter ξ. Conversely, there are a variety of ways to quantify the strength of azimuthal anisotropy, and its fast azimuth Ψ. Here, we quantify it in terms of the peak-to-peak anisotropy:

$$\text{azi} = \frac{2G}{L}, \tag{2.21}$$

where $G = \sqrt{G_c^2 + G_s^2}$. The azimuth of fast propagation is given by:

$$\Psi = 0.5 \arctan\left(\frac{G_s}{G_c}\right). \tag{2.22}$$

2.5.5 The inverse problem

The inverse problem is cast in a full Bayesian procedure where the solution is an ensemble of models distributed according to the posterior pdf $p(\mathbf{m} \mid \mathbf{d}_{obs})$, accompanied by their uncertainty bounds. In this framework, Bayes' theorem holds:

$$p(\mathbf{m}|\mathbf{d_{obs}}) \propto p(\mathbf{m})p(\mathbf{d_{obs}}|\mathbf{m}). \tag{2.23}$$

The parameter space is searched using a Markov chain Monte Carlo (McMC) algorithm. To produce reasonable acceptance rates, we employed the adaptive perturbation scheme discussed in Section 2.4.2.

The likelihood

Assuming errors are uncorrelated and distributed according to a Gaussian distribution with zero mean and variance σ_c, the likelihood function corresponding to a single dispersion measurement can be written as:

$$p(\mathbf{c_{obs}}|\mathbf{m}) = \frac{1}{(2\pi\sigma_c^2)^{N/2}} \exp\left[\frac{-||\mathbf{c_{obs}} - c||^2}{2\sigma_c^2}\right], \tag{2.24}$$

where $\mathbf{m}$ is the 1D velocity model described in Fig. 2.1, N is the number of discreet periods, and σ_c^2 is the estimated variance of the data noise. The forms of the likelihood functions of the 2θ terms can be cast in the same manner.

The prior

Assuming prior independence, the prior $p(\mathbf{m})$ can be written as a product of 1D priors on each unknown parameter considered giving us:

$$p(\mathbf{m}) = \prod_{i=1}^{\text{Nlyrs}} \left[p(L_i)p(\xi_i)p(Gs_i)p(Gc_i)\right], \tag{2.25}$$

where Nlyrs is the number of layers of the 1D Earth model. Eq. (2.25) implies that the probability of accepting the transition is automatically zero should one of the parameters be outside their respective bounds.

The Sampling algorithm

We use a Markov chain Monte Carlo (McMC) algorithm to search the parameter space for 1D Earth model candidates that could explain the surface wave dispersion measurements. The sampler initiates by randomly drawing a reference model from the prior followed by the evaluation of the likelihood function. Within the Markov chain, the current 1D Earth model is perturbed at their control points (see the red arrows in Fig. 2.1 to transition into a new state. This is performed by randomly selecting one of the following set of moves:

1. Change the Love parameter L values of all control points according to a Gaussian distribution centered at the current value of L.

2. Change ξ values of all control points according to a Gaussian distribution centered at the current value of ξ.

3. Change G_s values of all control points according to a Gaussian distribution centered at the current value of G_s.

4. Change G_c values of all control points according to a Gaussian distribution centered at the current value of G_c.

If the proposed 1D Earth model is within their respective prior bounds, we then solve the forward problem completely. The computed dispersion curves are then compared with the observed synthetics following the evaluation of the likelihood function. The resulting probability is then used to evaluate the acceptance probability via the Metropolis-Hastings algorithm. The outcome of the algorithm determines whether the proposed model is added to the posterior distribution. Should it be rejected, the current model is counted successively in the next iteration.

2.5.6 Application to a 3-D deforming upper mantle induced by a sinking anomaly

We perform 1D anisotropic surface wave tomography at 32 different geographical locations. To efficiently implement the inversions for each location, the algorithm is parallellised in such a way that multiple Markov chains can simultaneously search the parameter space independently from one another. Here, one Markov chain is allocated to one processor. At the end of the inversion procedure, the ensemble of models from each chain are then gathered to construct the posterior probability distribution. Since we consider 20 independent chains for each geographical location, and we have

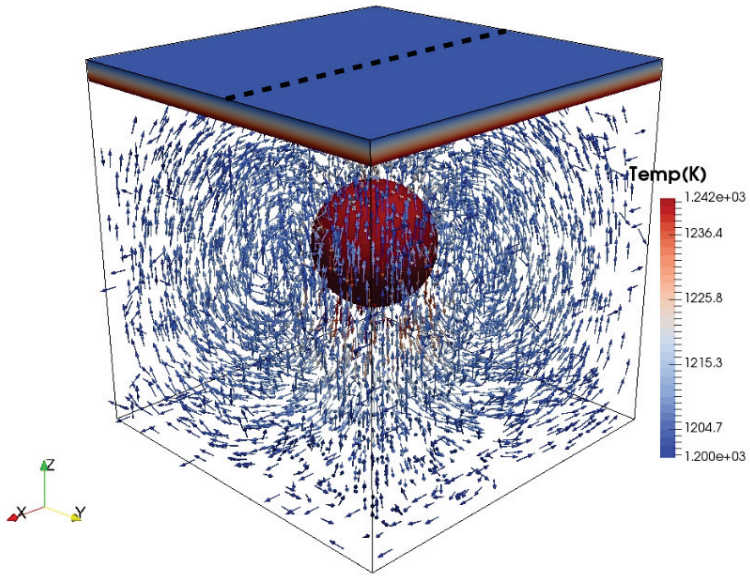

FIGURE 2.2: Snapshot of a 3-D deforming mantle induced by a sinking spherical anomaly. The model domain is of the size 400 km $\times$ 400 km $\times$ 400 km. The isovolumetric gradients correspond to the 3-D temperature profile of the region, and the superimposed vector field is the flow induced by the spherical body. The dashed lines at the surface represent the 32 geographical locations of the local surface wave dispersion curves.

32 locations in total, we need a hefty 32 $\times$ 20 = 640 cores to implement the parallel scheme.

We demonstrate the method in the case of a 3-D deforming upper mantle induced by a negatively buoyant spherical anomaly, akin to a Stoke's sinker. As illustrated in Fig. 2.2, the surrounding material responds to the sinking anomaly by producing a return flow. The return flow, together with the downward motion of the anomaly, generates local convection cells whose scales are consistent with that of the upper mantle (which we set at L_s = 400 km). We setup the geographical locations of the local surface wave dispersion curves in such a way that they provide a good coverage of the anomaly. Here, we use 32 locations that slice the region evenly into two (refer to the dashed lines).

In the succeeding chapter, we will explain in detail the full forward problem that allows us to arrive at azimuthally-varying dispersion curves from the deformation pattern and from the temperature field. For now, we have at hand the observed synthetics generated by a 3-D deforming upper mantle.

To mimic real-Earth observations, we tarnish the observed synthetics with noise. We added random uncorrelated noise with standard deviation $\sigma_{c_{R,L}}$ = 0.001 km/s for

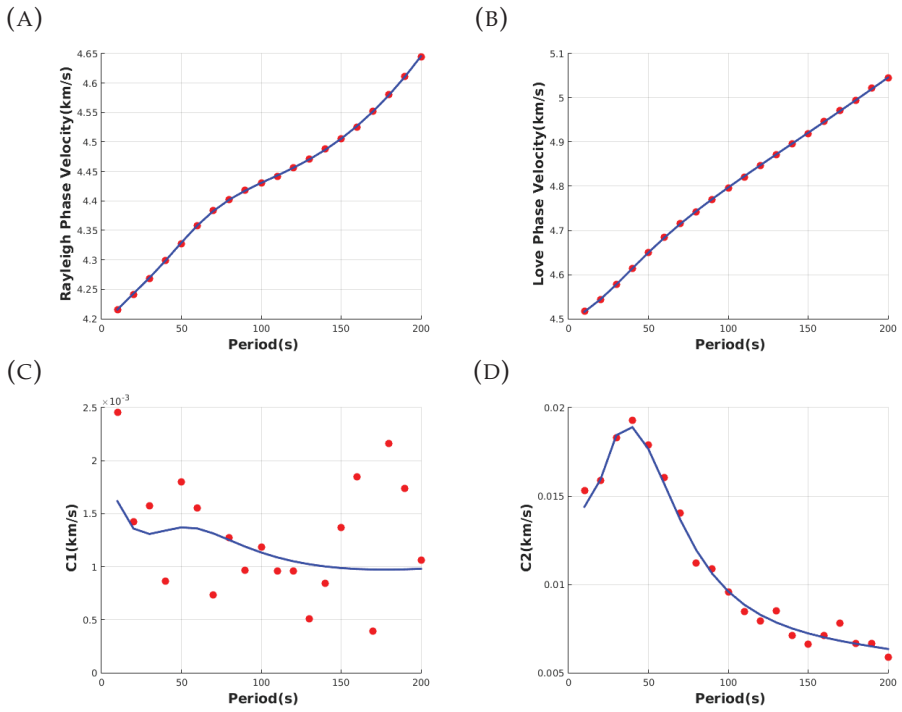

FIGURE 2.3: Surface wave phase velocity dispersion curves and their azimuthal variations at a specific geographical location. Solid blue lines are the correct values and the scatter plots are the ones added with noise and are to be inverted.

TABLE 2.2: Prior ranges of the unknown model parameters.

	L(GPa)	ξ	G_c(GPa)	G_s(GPa)
min	20	0.7	-5	-5
max	150	1.3	5	5

the isotropic dispersion curves $\mathbf{c_R}$ and $\mathbf{c_L}$, and $\sigma_{c_{1,2}} = 0.0005$ km/s for the azimuthal components $\mathbf{c_1}$ and $\mathbf{c_2}$. Fig. 2.3 shows the resulting dispersion curves at one given location. Solid blue lines are the synthetic dispersion curves without noise and the red dots represent the synthetics with added random uncorrelated noise. By stacking the dispersion curves laterally (*i.e.*, placing them side-by-side), we are able to visualize the true structure in the data space. Fig. 2.4 shows the effect of the density anomaly onto the resulting surface wave dispersion maps. The negative subsidence observed in the Rayleigh map corresponds to an increases in its speed as it traverses the cold anomaly. The contaminated data is then used to invert for the elastic structure.

For each geographical location, the inversion consists of 20 independent Markov chains each containing 1.0×10^6 samples initiated at a random 1D Earth model drawn from the prior distribution. The prior bounds of each model parameter is summarized in Table 2.2. The models are then collected after a burn-in period of 900,000 samples

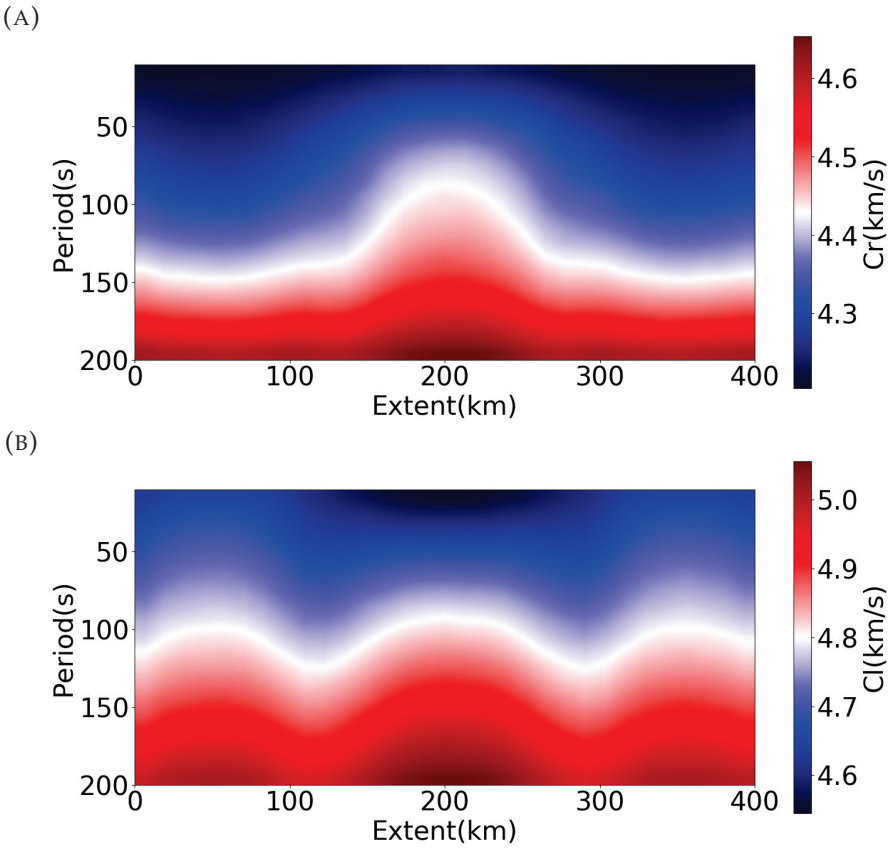

FIGURE 2.4: Constructed 2D surface wave maps with added noise to be used in the inversion. (A) Rayleigh. (B) Love. The negative subsidence in the Rayleigh map corresponds to the cold anomaly mapped in the data space.

to ensure the chains are in steady-state and are properly sampling the posterior distribution. Note that even though the models are randomly initiated, the positions of the control points are fixed, and thus may have strong implications on the recovered structures.

Fig. 2.5 shows the mean velocity structure and the mean radial anisotropy recovered from Bayesian inversion, and the correct structures. The 2D models are constructed by placing the recovered 1D mean structures side-by-side. Surface waves were able to successfully map the most prominent feature, that is, the seismic anomaly associated with the denser sphere. Radial anisotropy was also successfully recovered. However, one of the major drawbacks of surface waves is that they are more sensitive at shallower depths. Hence, the top layers are better resolved than the bottom layers. We thus expect the upper portion to exhibit less model uncertainties that the bottom half. Additionally, some essential features are smeared vertically, which we attribute to the inherent long period nature of surface waves. The structure also appears to be smooth with depth which in part is due to the choice of parameterisation (*i.e.*, cubic splines are smooth functions). The lateral resolution however does not exhibit complete smoothness; instead, appears to be degraded at some regions. This is a result of using randomly uncorrelated data. In this case, the random data noise manifest as small-scale artifacts in the model space.

One of the main advantages of a Bayesian framework is we can express the solution as a marginal posterior pdf where the width of the distribution quantifies the model uncertainties. Fig. 2.6 shows 1D marginal distributions versus depth of L, ξ, peak-to-peak azimuthal anisotropy, and its fast azimuth Ψ at a specific geographical location. These depth profiles were obtained by merging the ensemble of models from the 20 Markov chains. The entire number of models $\mathbf{m}$ used to build the posterior pdf is thus 2.0×10^6. Results show that the profiles successfully capture the true structure although sharp gradients fail to be resolved. The resulting azimuthal anisotropy shows some peculiarities. Since we only make use of five control points which are fixed, the saddle points of the true azimuthal anisotropy fail to be captured. The increase in model uncertainty at depths is again a result of surface waves being concentrated at shallower depths.

2.5.7 Discussion and conclusion

We have demonstrated conventional anisotropic surface wave tomography. Here, the inverse problem is fully Bayesian in that the solution is an ensemble of models distributed according to a posterior probability. Here, the parameter space is searched using a Markov chain Monte Carlo algorithm with importance sampling. This method

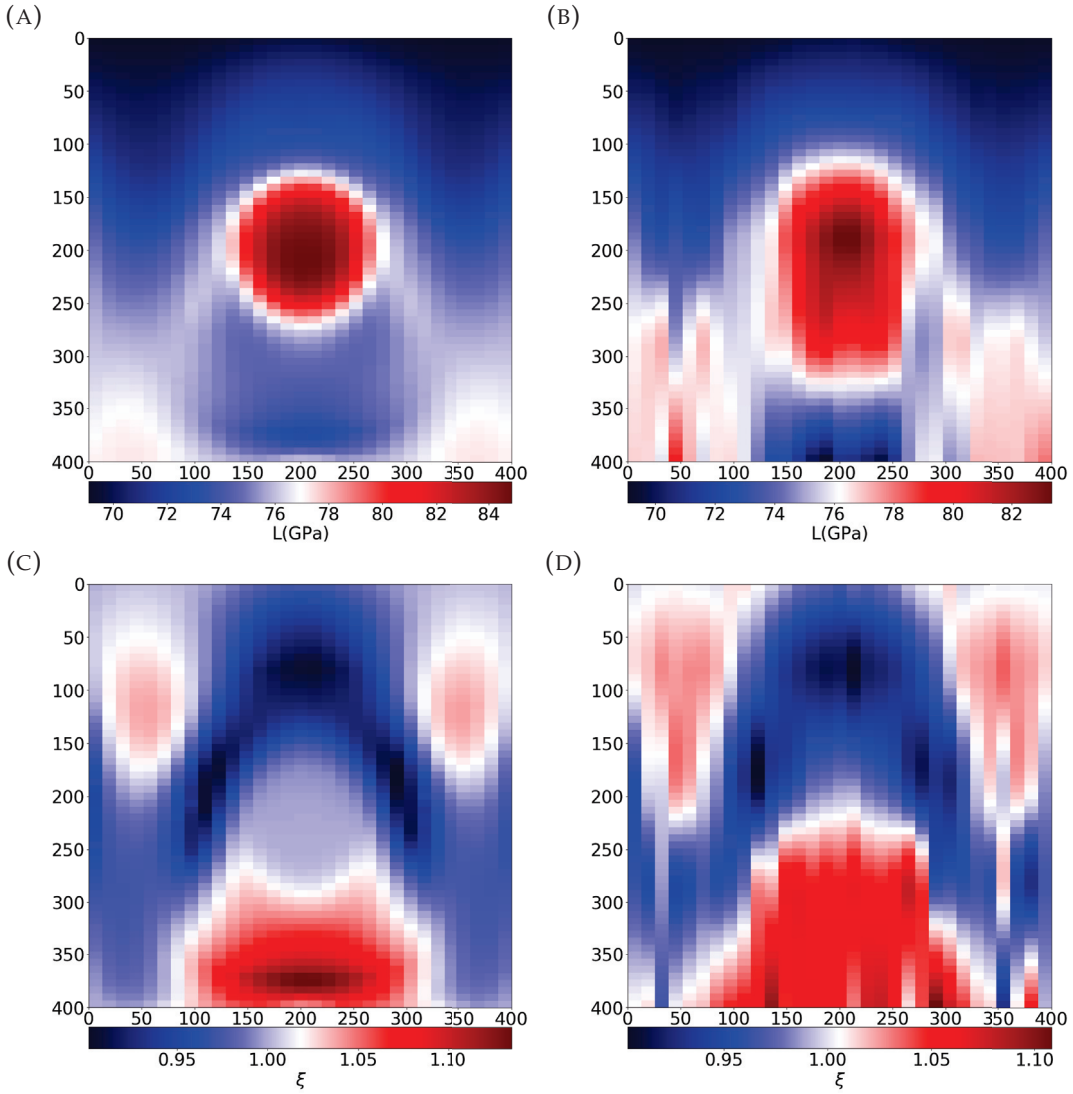

FIGURE 2.5: True models (left), Mean models recovered from the inversion (right). (A) and (B) $L-$ structure, (C) and (D) radial anisotropy.

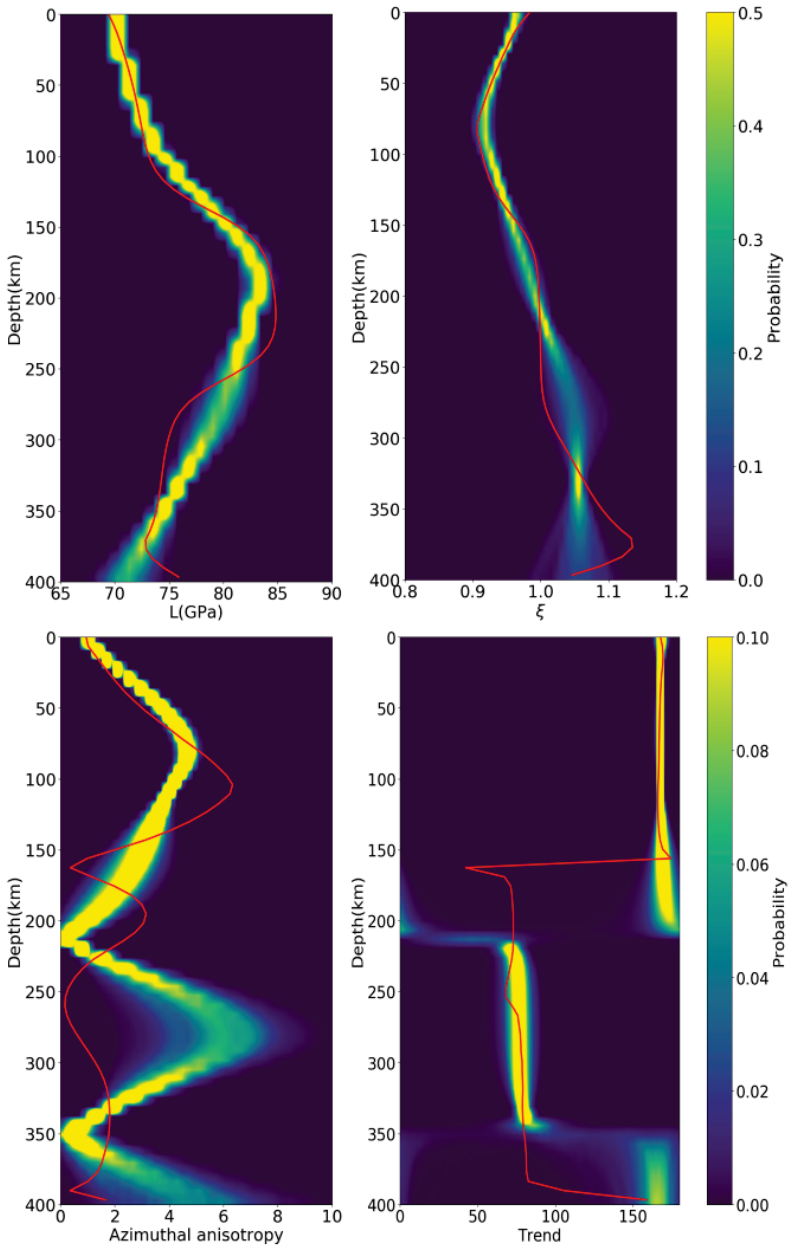

FIGURE 2.6: 1D marginal posterior distributions of L, ξ, peak-to-peak azimuthal anisotropy, and its fast azimuth Ψ at a specific geographical location, inferred from the Bayesian inversion of surface wave dispersion curves. The true structures are plotted in solid red.

was applied to azimuthally-varying surface wave dispersion curves computed from a 3-D deforming upper mantle.

Conventional surface wave tomography was able to recover robust features. However, the resolving power of surface waves limits the vertical resolution of the recovered structures. Surface are long period observations and hence cannot resolve small-scale features. These features are instead spatially-averaged and are smooth as a result. It is also important to emphasize that the choice of parameterisation regularises the inverse problem. As initially stated, the positions of the control points where the values are perturbed are fixed. As a consequence, the results from all the chains inflict strong dependency of the final result on the parameterisation. Such is apparent in the recovered azimuthal anisotropy structures. Finally, since the energy associated with surface waves tend to be more concentrated near the surface, they thus tend to decrease resolution with depth. This is evident in the inversions where below the 250 km mark, the width of the posteriors begin to increase.

To handle such complications, possible future avenues include the incorporation of higher-modes. Throughout the inversions, we only considered fundamental-mode surface waves. Using higher modes would increase the sensitivity of surface waves to deeper structures thus providing adequate resolution with depth (e.g. Simons, Zielhuis, and Van Der Hilst, 1999). Another alternative route we could take is to consider trans-dimensional approaches where the number of model parameters to be inverted for are treated as an unknown. In such cases, the model adapts to the data itself, thus providing a state of balance between model complexity and resolution (Bodin et al., 2012).

Across the horizontal, the recovered structures appear to be less resolved since unlike teleseismic body waves, surface waves exhibit poor lateral resolution. Still, we should expect the horizontal structures to be smooth. This is not the case however as this was a result of using randomly uncorrelated data noise. This random noise maps as small-scale artifacts in the model, which clearly explains the lateral discontinuities. Since real data noise are inherently spatially- and periodically-correlated, it is therefore necessary to build the full data covariance matrix and account for them in the likelihood function.

Finally, even though we placed ourselves in the best case scenario, that is, imposing hard *a priori* constraints in the inversions by setting the correct values for A, ϕ, η, B_c, and B_s, we are still inhibited by the problems associated with conventional surface wave tomography. In practice, these values are treated as unknowns in the inversion or are determined based on empirical relations. Implementing the former complicates

the inversion procedure in a way that it increases the model complexity, and that using only one type of data may not resolve these parameters at every location. One is thus forced to go with the latter, where they utilise simple relations such as the V_pV_s ratio to constrain V_p-related parameters. Velocity models inferred from such simplistic formulations however may not be representative of the would-be recovered structure especially in situations where complex underlying mechanisms dictate the lithological integrity of the region. The tomographic problem should therefore be approached from a different perspective, where no symmetry of the elastic tensor is imposed at the outset, and are instead driven by key geophysical processes.

Chapter 3

Geodynamic Tomography: The Forward Problem

3.1 Summary

As demonstrated in the previous chapter, conventional surface wave tomography is indeed a powerful technique to constrain the seismic structure of the upper mantle. Its interpretation however is equivocal and is usually up to the tomographer's prerogative. Apart from this even though the problem is cast in a Bayesian framework, constraining the 13-depth functions at every geographical location is still proven to be a gruelling task. Even with the imposition of hard constraints on the prior, surface waves still fail to render some important features not only due to its very nature and the choice of parameterisation, but also due to the inherent complexity brought about by a 3-D-deforming upper mantle. This felt the need to redefine the surface wave tomography problem, where geodynamical and petrological constraints are introduced to estimate the full elastic tensor at every geographical location and thus precisely explain the observations measured at the surface. In this way, it also reduces the possible number of Earth models down to a subset that are consistent with geodynamical and mineralogical predictions. Finally, incorporating geodynamical constraints address the high-dimensionality of anisotropic surface wave tomography by only utilising a single scalar field (*i.e.*, temperature). In this chapter, we discuss in detail the complete forward problem which consists of four major steps: (1) geodynamic mantle flow modeling, (2) mineral physics modeling, (3) thermodynamic modeling, and (4) surface wave dispersion calculations.

3.2 Geodynamic modeling of upper mantle flow

3.2.1 3-D instantaneous flow with variable viscosity

We consider buoyancy-driven convection in a highly-viscous, incompressible, creeping element of fluid in a 3-D Cartesian domain. These density heterogeneities pertain to pervasive mantle rocks that are gravitationally unstable. Its density at a given geographical position can be expressed as a function of four major variables:

$$\rho = f(T, P, M, C), \tag{3.1}$$

where T is temperature, P is hydrostatic pressure, M is mineral composition, and C is chemical composition. Assuming the rocks to be independent of C and M, and that the P and T dependence of ρ is separable, one may express the P and T variations of ρ with:

$$\alpha = \frac{-1}{\rho}\frac{\partial \rho}{\partial T}, \tag{3.2}$$

and

$$\beta = \frac{1}{\rho}\frac{\partial \rho}{\partial P}. \tag{3.3}$$

The variables α and β are the thermal expansion and compressibility coefficients, respectively. Typically treated as constants in most instantaneous flow models, α and $\beta > 0$ ensures that the density increases with pressure and decreases with temperature at a constant pace. In real Earth scenarios however, α and β are very much dependent on P and T and thus more realistic equations of state should be considered to model density variations. Solving the differential equations by separation of variables yields:

$$\rho = \rho_0 \exp\left(-\alpha(T - T_0) + \beta(P - P_0)\right) \tag{3.4}$$

where ρ_0 is a reference value for density, and T_0 and P_0 are the temperature and pressure at ambient conditions. Throughout our work, we assume the response of the density to variations in pressure to be negligible compared to temperature, and since rocks tend to exhibit strong resistance to changes in its mechanical properties, we can treat $\alpha(T - T_0)$ to be minuscule and thus exploit a first-order Taylor approximation to eq. (3.4):

$$\rho(T) = \rho_0[1 - \alpha(T - T_0)], \tag{3.5}$$

resulting in a linear equation of state.

The dynamics of this system, just like any classical particle, can be described by Newton's second law but applied to an element of fluid. Its full form is formulated

giving the well-known Navier-Stokes equation:

$$-\nabla P + \nabla \cdot \boldsymbol{\tau} + \rho g \hat{e}_g = \rho \frac{D\mathbf{u}}{Dt}, \tag{3.6}$$

where $\boldsymbol{\tau}$ is the deviatoric stress tensor obtained from a constitutive law, g is gravity, and the last term is the Lagrangian derivative applied to the flow field $\mathbf{u}$. One useful assumption applied to most fluid dynamics models is the incompressibility criterion. In geodynamic modeling for example, rocks are tend to be sturdy enough to resist variations in density with time. Such a case simplifies the constitutive law to:

$$\tau_{ij} = 2\eta\varepsilon_{ij}, \tag{3.7}$$

where η is the dynamic viscosity, and εij is the strain rate tensor. Under this condition, the conservation of mass can be reduced to:

$$\nabla \cdot \mathbf{u} = 0. \tag{3.8}$$

In the convective mantle, rocks pervade in a laminar fashion over geologic time scales. Such can be modeled by considering highly viscous flows where the inertial term of eq. (3.6) vanishes due to its negligibility compared to the surface forces and the body forces. Since we work with buoyancy-driven flows, we apply the Boussinesq approximation. Here, the density variations are only accounted for when they appear in the body force term. The momentum equation simplifies to the Stokes equation:

$$-\nabla P + \nabla \cdot \boldsymbol{\tau} + \rho g \hat{e}_g = 0, \tag{3.9}$$

which is also referred to as creeping flow. The pressure can be decomposed into two parts: (1) a hydrostatic term P_H, and (2) a dynamic term P_D. By the distributive property, the gradient of P can be expressed as:

$$\nabla P = \nabla P_H + \nabla P_D. \tag{3.10}$$

Note that the hydrostatic term can be simplified further by knowing that $\nabla P_H = \rho_0 \mathbf{g}$. Applying the simplification and plugging eq. (3.5) to eq. (3.9) yields:

$$-\nabla P_D + \nabla \cdot \boldsymbol{\tau} + \alpha\rho_0 \mathbf{g}(T - T_0) = 0. \tag{3.11}$$

Finally, apart from the conservation of mass and momentum, dynamic fluid parcels

have to abide by the conservation of energy. With this, geodynamic models have to incorporate heat transport phenomenon to predict variations in the temperature field. However, since we will be dealing with present-day temperature fields and instantaneous flow solutions, the convection-diffusion equation for temperature can be ignored.

3.2.2 Computational strategy

In this section, we explain in detail our computational strategy to solve eq. (3.11) for a given class of temperature fields.

Non-dimensionalisation

We adopt scaling relations to introduce non-dimensional counterparts of the relevant variables as a prerequisite in solving the eq. (3.11) numerically. In this way, it reduces the number of free parameters thereby simplifying the analysis of the problem at hand. The scaling relations are shown below. Here, the primed variables pertain to the non-dimensional form of the free parameters.

$$\begin{aligned} \eta' &= \frac{\eta}{\eta_0} & \mathbf{u}' &= \mathbf{u}\frac{L_s}{\kappa} \\ P' &= P\frac{L_s^2}{\eta_0 \kappa} & T' &= \frac{T}{T_0}, \end{aligned} \tag{3.12}$$

where κ is the thermal diffusivity coefficient and L_s is the characteristic length scale of the model domain. Note that the scaling factor of the deviatoric stress tensor is the same with pressure. Thus, the momentum equation can be written alternatively as:

$$\frac{\eta_0 \kappa}{L^3}(-\nabla P' + \nabla \cdot \boldsymbol{\tau}) = -\rho_0 \alpha g T' \hat{e}_g. \tag{3.13}$$

Finally, the non-dimensional form of the momentum equation is:

$$-\nabla P + \nabla \cdot \boldsymbol{\tau} + R_a T' \hat{e}_g = 0. \tag{3.14}$$

R_a is the well-known Rayleigh number which regulates the degree of natural convection. In the case of instantaneous flows, it controls the magnitude of the velocity.

$$R_a = \frac{\rho_0 \alpha T_0 L_s^3 g}{\eta_0 \kappa}. \tag{3.15}$$

Throughout this work, we impose a power law rheology for η' analogous to Arrhenius law for viscous fluids. Known as the Frank-Kametnetzkii approximation, the viscosity obeys an exponential law depending only on temperature:

$$\eta' = \exp(E(T' - 1)), \tag{3.16}$$

where E is the activation energy for viscosity. Microscopically, E is interpreted as an energy barrier describing the degree of shear resistance between molecules. This barrier has to be overcome by a thermally-activated process for molecules to get past each other (Glasstone, Laidler, and Eyring, 1941).

Multigrid method with variable viscosity

Solving the Stokes equation using direct methods interdicts one from producing fine resolution models due to limitations related to computational memory and computational speed. Such is critical specifically when transitioning from 2D to 3-D since the number of linear equations to be solved and the number of mathematical operations required to solve such equations scale with the dimensionality. Thus to preserve the resolution in 3-D, the amount of linear equations increases by at least two orders of magnitude, and as a consequence the computational operations has to increase by at least three orders of magnitude. To address this problem, iterative methods have been introduced to combat memory limitations. However, most of the well-known iterative techniques, Gauss-Sidel for instance, faces the problem of resolution. Here, the number of iterations needed to come up with a solution model grows with increasing resolution.

To overcome these limitations, we implement the multigrid algorithm which was first formulated by Fedorenko (1964) and has been actively developed ever since. Just like any classical iterative methods, multigrid allocates computational memory efficiently. Its main advantage however lies on the fact that the number of iterations is independent of the grid size thus speeding up the rate of convergence. The basic idea is that the system of equations are solved simultaneously on several grids with varying resolution where information exchange is permitted between these grids.

We adopt the approach of Albers (2000) where we solve the Stokes problem with variable viscosity using multigrid methods on a staggered grid parameterisation. The latter simply means that in a unit cell, scalar fields are assigned/solved at the geometric center whereas vector field components are parameterised at the center of the cell faces with directions parallel to the surface normal. We follow this by employing a

control-volume sequence to discretise the mass conservation and momentum conservation equations:

Mass conservation

$$\frac{u_{\mathbf{x}ijk} - u_{\mathbf{x}i-1jk}}{\Delta x} + \frac{u_{\mathbf{y}ijk} - u_{\mathbf{y}ij-1k}}{\Delta y} + \frac{u_{\mathbf{z}ijk} - u_{\mathbf{z}ijk-1}}{\Delta z} = 0\,, \tag{3.17}$$

Momentum conservation

$$\frac{\tau^{xx}_{i+1jk} - \tau^{xx}_{ijk}}{\Delta x} + \frac{\tau^{xy}_{ijk} - \tau^{xy}_{ij-1k}}{\Delta y} + \frac{\tau^{xz}_{ijk} - \tau^{xz}_{ijk-1}}{\Delta z} = -\frac{p_{i+1jk} - p_{ijk}}{\Delta x}\,, \tag{3.18}$$

$$\frac{\tau^{xy}_{ijk} - \tau^{xy}_{i-1jk}}{\Delta x} + \frac{\tau^{yy}_{ij+1k} - \tau^{yy}_{ijk}}{\Delta y} + \frac{\tau^{yz}_{ijk} - \tau^{yz}_{ijk-1}}{\Delta z} = -\frac{p_{ij+1k} - p_{ijk}}{\Delta y}\,, \tag{3.19}$$

$$\frac{\tau^{xz}_{ijk} - \tau^{xz}_{i-1jk}}{\Delta x} + \frac{\tau^{yz}_{ijk} - \tau^{yz}_{ij-1k}}{\Delta y} + \frac{\tau^{zz}_{ijk+1} - \tau^{zz}_{ijk}}{\Delta z} = -(Ra\frac{T_{ijk+1} + T_{ijk}}{2} + \frac{p_{ijk+1} - p_{ijk}}{\Delta z})\,, \tag{3.20}$$

where

$$\begin{aligned}
\tau^{xx}_{ijk} &= 2\eta_{ijk}(\frac{u_{\mathbf{x}ijk} - u_{\mathbf{x}i-1jk}}{\Delta x}),\\
\tau^{yy}_{ijk} &= 2\eta_{ijk}(\frac{u_{\mathbf{y}ijk} - u_{\mathbf{y}ij-1k}}{\Delta y}),\\
\tau^{zz}_{ijk} &= 2\eta_{ijk}(\frac{u_{\mathbf{z}ijk} - u_{\mathbf{z}ijk-1}}{\Delta z}),\\
\tau^{xy}_{ijk} &= 2\eta^{xy}_{ijk}(\frac{u_{\mathbf{x}ij+1k} - u_{\mathbf{x}ijk}}{\Delta y} + \frac{u_{\mathbf{y}i+1jk} - u_{\mathbf{y}ijk}}{\Delta x}),\\
\tau^{xz}_{ijk} &= 2\eta^{xz}_{ijk}(\frac{u_{\mathbf{x}ijk+1} - u_{\mathbf{x}ijk}}{\Delta z} + \frac{u_{\mathbf{z}i+1jk} - u_{\mathbf{z}ijk}}{\Delta x}),\\
\tau^{yz}_{ijk} &= 2\eta^{yz}_{ijk}(\frac{u_{\mathbf{y}ijk+1} - u_{\mathbf{y}ijk}}{\Delta z} + \frac{u_{\mathbf{z}ij+1k} - u_{\mathbf{z}ijk}}{\Delta y}).
\end{aligned}$$

Eqs (3.17), (3.18), (3.19), and (3.20) form a discrete set of equations that can be expressed in block form as:

$$\begin{pmatrix} \mathbf{B} & \mathbf{A^T} \\ \mathbf{A} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ P \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ 0 \end{pmatrix} \tag{3.21}$$

The 2×2 matrix is acted upon the solution vector $(\mathbf{u}, P)^T$ to produce the source term $(\mathbf{f}, 0)^T$ containing the body force (i.e. gravity). The Jacobian operator contains the constitutive equation $\mathbf{B}$ related to the rheology of the medium. Operators $\mathbf{A^T}$ and

A are the gradient, and the divergence, respectively. As mentioned earlier, classical iterative methods such as Gauss-Seidel, Jacobi, and successive over-relaxation (SOR) are usually used to solve eq. (3.21). The success of these methods however decline with model complexity. We therefore employ the multigrid method to solve eq. (3.21). Here, the algorithm reduces the modeling errors on a coarse grid, and uses the coarse-grid solutions to improve convergence on a refined grid. We briefly elicit the algorithm given a set of linear equations of the form $Ax = b$:

1. A smooth solution is initially obtained iteratively followed by the damping of high-frequency errors. This technique is called *smoothing*. Here, we use the `SIMPLER` algorithm (Patankar, 2018) since classical iterative procedures are invalid when solving block matrices with zeros on the main diagonal.

2. This is followed by *restriction* where information (*e.g.*, residuals) have to be transferred from a fine grid to a coarse grid. This means that low frequency errors ϵ from (1) are mapped onto the coarse grid domain. To obtain coarse grid errors, we solve the equation $A_c\epsilon_c = R\epsilon$ where A_c and ϵ_c are the coarse grid counterparts of the Jacobian and the residual, respectively, and R is the restriction operator.

3. Once restriction is cycled several times, ϵ_c is delineated onto a finer grid. This step is called *prolongation*. The prolongation operator is then applied onto ϵ_c, the resulting prolongation error is then summed with the fine grid solution to increase the rate of convergence. Prolongation is performed until the initial grid size is attained.

The code is written in `Fortran90` and parallelised with `OpenMP` (*i.e.*, shared memory processing among multiple threads). The inputs are the scalar fields temperature and viscosity which are adimensionalised before solving. The algorithm terminates upon reaching a tolerance value that is chosen ad hoc. It determines the accuracy of the method but also evinces trade-offs with the computation time. For instance, choosing small values would increase the model accuracy however at the cost of suppressing the speed of convergence. Once the flow is calculated, we compute the local velocity gradient by means of finite difference. Central differencing we get:

$$L^{xx}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{x}i+1jk} - u_{\mathbf{x}i-1jk}}{\Delta x}, \quad L^{yx}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{y}i+1jk} - u_{\mathbf{y}i-1jk}}{\Delta x}, \quad L^{zx}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{z}i+1jk} - u_{\mathbf{z}i-1jk}}{\Delta x},$$
$$L^{xy}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{x}ij+1k} - u_{\mathbf{x}ij-1k}}{\Delta y}, \quad L^{yy}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{y}ij+1k} - u_{\mathbf{y}ij-1k}}{\Delta y}, \quad L^{zy}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{z}ij+1k} - u_{\mathbf{z}ij-1k}}{\Delta y}, \tag{3.22}$$
$$L^{xz}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{x}ijk+1} - u_{\mathbf{x}ijk-1}}{\Delta z}, \quad L^{yz}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{y}ijk+1} - u_{\mathbf{y}ijk-1}}{\Delta z}, \quad L^{zz}_{ijk} = \frac{1}{2}\frac{u_{\mathbf{z}ijk+1} - u_{\mathbf{z}ijk-1}}{\Delta z}.$$

We test the algorithm both in series and in parallel given an initial grid resolution of $64 \times 64 \times 64$ to weigh up the computation times. Compared to serial computation which ran in about 6.6s, the parallel computation took six times faster within an 8-core processor which we found to be the optimal number of cores. It is worth noting that the speed of convergence does not necessarily scale with the number of cores. As for the physical parameters governing the flow, we chose L_s = 400 km, $\alpha = 2.0 \times 10^{-5}\ K^{-1}$, g = 9.81 m/s^2, $\eta_0 = 10^21$ Pa, T_0 = 1900 K, $\kappa = 10^-6\ m^2$/s, and $\rho_0 = 3.8 \times 10^3$ kg/m^3. From these values, one may then compute the Rayleigh number using eq. (3.15). Fig. 3.1 illustrates the flow field with free-slip boundary conditions across the borders of the box and the second invariant of the velocity gradient tensor for a given class of temperature fields. Since the viscosity obeys a power law rheology with temperature, it is essential for the temperature fields to be smooth. In this way, the viscosity as well avoids sharp contrasts which might deteriorate the convergence towards a stable solution.

3.3 Texture evolution modeling of upper mantle minerals

In this section, we discuss the role of mineral-scale fabrics in interpreting seismic anisotropy measurements and how the latter associates with upper mantle flow. Let us consider a polycrystalline aggregate composed of N_{crystal} crystals. Each crystal has its own crystallographic axis oriented at a given position with respect to a laboratory frame, and a volume fraction V_{frac} (*i.e.*, the ratio between the volume of a single crystal versus all constituents). Each crystallographic axis is measured using a set of three Euler angles given by Φ, Θ, and Ψ. By convention, we utilise the Bunge ZXZ notation (*i.e.*, a rotation by an amount Ψ about the Z_1 axis, followed by Θ about the X_2 axis, lastly Φ about the Z_2 axis). Supposing that the crystals are randomly-oriented, the aggregate contains no distinguishable fabric and is thus considered to be seismically isotropic. Upon the

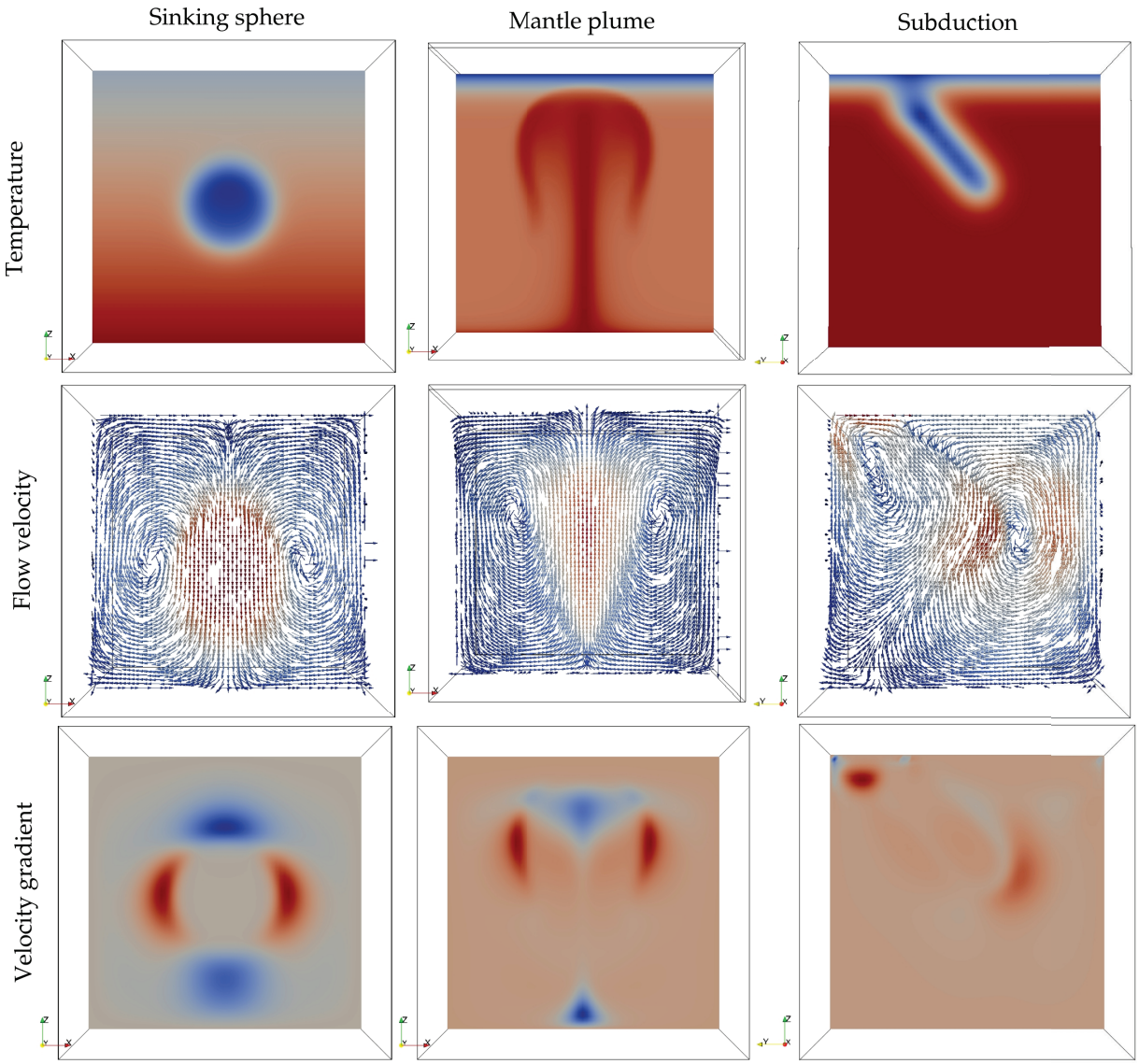

FIGURE 3.1: 2D representation of a 3-D instantaneous flow field and its corresponding deformation gradient (second invariant) for a given family of temperature fields in a $64 \times 64 \times 64$ grid. We use the multigrid method to approximate the solution to eq. (3.21). (Left panels) Sinking spherical anomaly. (Middle panels) Ascending plume. (Right panels) Subduction. Color gradient from blue to red denotes intensity. Flow velocity's magnitude are the strongest where the temperature anomaly is located whereas its gradients specify the zones where shear deformation appears to be the largest.

deformation of the aggregate, each crystallographic axis on average tends to rotate towards the direction of shear. Thus, the vector sum of these individual orientations take a 'preferred' orientation. This phenomenon is called crystallographic preferred orientation (CPO), and is one of the primary sources of long wavelength seismic anisotropy observed in tomographic images.

3.3.1 Finite strain theory

Before we examine the well-established theory governing CPO evolution, we briefly discuss as a preliminary a first-order theory that may explain how seismic anisotropy relates with mantle flow in the absence of constraints from mineral physics. McKenzie (1979) detailed how finite strain theory may be a substitute to constrain patterns of mantle deformation. In his work, he analysed the connection between seismic anisotropy and finite strain under the proposition that anisotropy is influenced only by the properties of the finite strain ellipse (fse). We commence by determining the distance between two particles in an element of fluid with positions x^a and x^b both taken at time t with respect to their initial positions x_0^a and x_0^b at $t = 0$:

$$x^a(t) - x^b(t) = \mathbf{F}(t)[x_o^a - x_o^b], \tag{3.23}$$

where $\mathbf{F}$ is the finite strain tensor responsible for the displacement of both particles resulting to a deformed element. At time $t = 0$, $\mathbf{F}$ is just the identity matrix, hence, no deformation is imposed. Supposing $r(t) = x^a(t) - x^b(t)$, the displacement of these particles at a later time is:

$$r(t + \Delta t) = x^a(t + \Delta t) - x^b(t + \Delta t). \tag{3.24}$$

Subtracting the two equations above and taking the limit as $\Delta t \to 0$ we get the difference between the instantaneous velocities of these particles as:

$$\frac{Dr}{Dt} = u^a(t) - u^b(t), \tag{3.25}$$

where D/Dt is the material derivative. Assuming that the particles are much smaller than the length scale of the convection cells, one may write the right-hand side of eq. (3.25) as a small perturbation around the reference velocity. Equation eq. (3.25) becomes:

$$\frac{Dr}{Dt} = \mathbf{L}r, \tag{3.26}$$

where **L** is the local velocity gradient tensor. Substituting r with the right-hand side of equation eq. (3.23) yields:

$$\frac{D\mathbf{F}}{Dt} = \mathbf{LF}. \tag{3.27}$$

Here, one may think of particles a and b as points along a flow line. Thus, for each element along a flow line, there is a corresponding local velocity gradient **L** evaluated at x, and a time spent t by a tracer particle. If we assume that the flow is time-independent, the advection terms in the material derivative vanish. This means that it easens the computation of **F** since it is evaluated along a streamline. For steady-state solutions we finally get:

$$\frac{\partial\mathbf{F}}{\partial t} = \mathbf{LF}. \tag{3.28}$$

Before we relate **F** with anisotropy, we first perform a polar decomposition of the former by taking the product of two second rank tensors: **F** = **RE**, where **R** and **E** are the rotation and the left-stretch matrices, respectively. We can express **E** as:

$$\mathbf{E} = \mathbf{F}\mathbf{F}^T. \tag{3.29}$$

Essentially, **E** is interpreted as an operator that transforms a sphere into an ellipsoid. Thus starting from an undeformed configuration (*i.e.*, a sphere), a material undergoing progressive deformation along a flow line accumulates strain and forms an ellipsoid. In such a scenario, the major axis of the fse lengthens continuously as strain accumulation progresses. The eigenvalues of **E** given by $\lambda_1 > \lambda_2 >$ and λ_3 pertain to the lengths of the fse axes, and the eigenvectors are their corresponding orientations. It is convenient to introduce a new parameter that characterises the strength of the finite deformation based from the eigenvalues of the fse. Known as natural strains, we measure the amplitude of finite deformation as:

$$\lambda = \frac{1}{2}\ln(\frac{\lambda_1}{\lambda_3}), \tag{3.30}$$

where λ is the natural strain, λ_1 is the length of the major axis, and λ_3 is that of the minor axis. Backed by numerous studies (e.g. Ribe, 1989; Ribe, 1992; Zhang and Karato, 1995; Becker et al., 2003), finite strain theory proved to be a fast and cheap proxy for seismic anisotropy. In principle, the fast axis of mantle olivine at sufficiently low temperatures ($\approx$ 1400 K), and at natural strains up to $\lambda = 1$ was observed to be in line with the long axis of fse. This means that body waves tend to move faster along the direction of maximum extension of strain, and thus mimics the spatial distribution of CPO-driven seismic anisotropy. At high temperatures, and moderate to high strains (λ

≈ 0.75), the fast axis of olivine tends to be parallel to the shear direction. It has been theorised that for simple shear flows, dynamic recrystallisation predominates over plastic deformation at high temperatures causing misalignment between the fast axis of olivine and the long axis of the fse.

3.3.2 Continuum mechanics approach to CPO evolution

In this section, we elaborate the dynamics of CPO evolution using a theory for continuous bodies. As such, we treat the spatial distribution of the individual crystallographic orientations also called the orientation distribution function (odf) as a continuous entity.

Suppose that a single crystal has an orientation represented by the three Euler angles, Φ, Θ, and Ψ, the odf, which we denote as h, is a function that describes the spatial distribution of the individual orientations at time t. If we slightly perturb the orientation α, where $\alpha = [\Phi, \Theta, \Psi]$, by an amount d$\alpha$, h($\alpha$,t)d$\alpha$ is the volume fraction of crystals with orientations between α and α+dα at a given time t. With this, let us consider an ensemble of crystals within an aggregate. If we let the aggregate yield under a locally imposed deformation, individual crystals respond by rotating at a rate $\frac{d\alpha}{dt}$ where we assume no interaction taking place between each crystal. Since the rotation rate is independent of the orientation of a neighboring crystal, the evolution with time of the crystal volume fractions between two different states of orientation is:

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial \alpha}(\dot{\alpha} h) = 0. \tag{3.31}$$

In terms of the Euler angles, eq. (3.31) can be finally expanded into:

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial \Phi}(\dot{\Phi} h) + \frac{1}{\sin\Theta}\frac{\partial}{\partial \Theta}(\dot{\Theta} h \sin\Theta) + \frac{\partial}{\partial \Psi}(\dot{\Psi} h) = 0. \tag{3.32}$$

Eq. (3.32) therefore represents CPO evolution as the time rate of change of crystal volume fractions in an interval [α, α+dα] while being compensated by a net outflow of crystallographic orientations in the same interval at a specified time t.

3.3.3 Deformation mechanisms for texture evolution

The evolution of h as shown in eq. (3.32) depends on ones knowledge of $\dot{\alpha}$. Thus, eq. (3.32) is particularly useful if we specify the rotation rate. Although assigning a function for $\dot{\alpha}$ would already suffice in the case of CPO evolution of a uniaxially compressed olivine, it is still rather simplistic and hence may not accommodate complex

flow patterns. With this, linking seismic anisotropy with realistic flow patterns is ill-advised. To fill this gap, we discuss a kinematic formalism where an analytical expression for $\dot{\alpha}$ is available. The aggregate then responds to an imposed macroscopic deformation through two main deformation mechanisms (e.g. Ribe and Yu, 1991; Kaminski and Ribe, 2001): (1) plastic deformation which drives the changes in crystallographic orientation, and (2) dynamic recrystallisation which administers the development of volume fractions of crystals.

Plastic deformation

Since plastic deformation depends on mineral composition, we only consider models of CPO evolution of a single mineral species, in this case, olivine whose number of slip systems S (*i.e.*, for materials undergoing plastic deformation, a slip system is a family of slip planes where atoms are densely packed, and slip vectors that represent the dislocation direction tangent to the corresponding slip plane.) is three assuring the validity of the theory which we are about to discuss. Let us recall the polycrystalline aggregate which we described at the beginning of the section composed of N_{crystal} crystals with V_{frac} volume fractions. In this formalism, each grain has its own orientation defined by the three Euler angles. Denoting the subscript p as the pth crystal, the aggregate has an ensemble of orientations $\alpha_{p=1,N_{\text{crystal}}} = [\Phi_{p=1,N_{\text{crystal}}}, \Theta_{p=1,N_{\text{crystal}}}, \Psi_{p=1,N_{\text{crystal}}}]$. By considering an unrotated crystal with respect to the lab frame, one can perform a sequence of orthogonal transformations resulting into the rotated configuration. Applying the Bunge convention, the transformation matrix composed of a matrix of direction cosines, for the pth crystal is given by:

$$b_{ij}^{p} = \begin{pmatrix} \cos(\Psi_p)\cos(\Phi_p) - \sin(\Psi_p)\sin(\Phi_p)\cos(\Theta_p) & \sin(\Psi_p)\cos(\Phi_p) + \cos(\Phi_p)\sin(\Psi_p)\cos(\Theta_p) & \sin(\Psi_p)\sin(\Theta_p) \\ -\sin(\Psi_p)\cos(\Phi_p) - \cos(\Psi_p)\sin(\Phi_p)\cos(\Theta_p) & -\sin(\Psi_p)\sin(\Phi_p) + \cos(\Phi_p)\cos(\Psi_p)\cos(\Theta_p) & \cos(\Psi_p)\sin(\Theta_p) \\ \sin(\Phi_p)\sin(\Theta_p) & \cos(\Phi_p)\sin(\Theta_p) & \cos(\Theta_p) \end{pmatrix} \tag{3.33}$$

Individual crystals may respond differently when the crystal aggregate is subjected to a macroscopic deformation. In the case of olivine in the upper mantle, the aggregates are subjected to progressive simple shear along the direction of flow and a rigid body rotation. Locally, the total velocity gradient within an olivine grain with $S = 3$ is just equal to the total amount of simple shear deformation across the three slip systems S_1, S_2, and S_3 accompanied by a rigid body rotation around the crystallographic axis. We assign $\mathbf{l}$ as the local velocity gradient tensor, and $\mathbf{L}$ as the macroscopic velocity gradient

tensor. **l** is expressed as:

$$l_{ij} = \sum_{S=1}^{3} 2\dot{\varepsilon_S}\mathbf{r_i}\mathbf{n_j} - \epsilon_{ijk}\dot{\alpha_k}, \tag{3.34}$$

whereas L is:

$$L_{ij} = \dot{E_{ij}} - \epsilon_{ijk}\Omega_k. \tag{3.35}$$

where ϵ_{ijk} is the levi-civita tensor. The variables **r** and **n** in eq. (3.34) are the unit vectors corresponding to the slip direction and the direction normal to the slip plane, respectively. The first expression in eq. (3.34) just takes the sum of each simple shear contribution at a rate $\dot{\varepsilon}$ for each slip system. E_{ij} is the macroscopic strain rate which is imposed by the flow. The anti-symmetric portions of both equations eq. (3.34) and eq. (3.35) correspond to the single-crystal rotation rate, and the rotation rate of the aggregate, respectively.

In this model of plastic deformation, $\dot{\varepsilon}$ is not implicitly defined. Instead, we apply an empirical relation coming from a product of different studies, that relates the strain rate to the resolved shear stress (RSS) τ_S at a given slip system S. Here, the resolved shear stress is the component of stress that resolves skewed slip planes. In order to relate the constitutive properties of a macroscopic system (*i.e.*, the deformation of the aggregate) with that of its constituents (*i.e.*, deformation of each grain), it is assumed that the stress tensor of both systems are proportional to each other. To quantify the degree of activity occuring in a given slip system with respect to an imposed macroscopic deformation, we utilise the weakest slip system ($S = 1$) resulting to:

$$\frac{\dot{\varepsilon_s}}{\dot{\varepsilon}} = \gamma_S = \frac{I_S\tau_1}{I_1\tau_S}\left|\frac{I_S\tau_1}{I_1\tau_S}\right|^{n-1}, \tag{3.36}$$

where n is the stress component. The variable I is a scalar that links the resolved shear stress with the macroscopic strain rate tensor E_{ij}, $I = r_i n_j E_{ij}$. In this formulation, each slip system is a function of their respective reference RSS values and their matching orientations. Knowing eq. (3.36), one may rewrite eq. (3.34) to match the macroscopic velocity gradient. Fleshed out by Ribe and Yu (1991), the local velocity gradient tensor for each grain is given by:

$$l_{ij} = G_{ij}\dot{\varepsilon} - \epsilon_{ijk}\dot{\alpha_k}, \tag{3.37}$$

where G_{ij} is a dimensionless tensor that controls the activities of all slip systems foisted by a macroscopic deformation:

$$G_{ij} = 2\sum_{s=1}^{S}\gamma_s\mathbf{r_{ij}}\mathbf{n_j}. \tag{3.38}$$

For any given mineral, von Mises criterion dictates that only crystals whose number of independent slip systems greater than 5 could accommodate deformation past the yield point (*i.e.*, the onset of plastic behavior under an imposed deformation). Thus, it is not enough for minerals such as olivine where $S = 3$ to acclimate macroscopic deformation. As a result, one cannot equate the microscopic velocity gradients in individual grains with the macroscopic velocity gradients even if one arbitrarily assigns $\dot{\alpha}$ and $\dot{\varepsilon}$ to coerce l_{ij} into matching with L_{ij}. The discrepancy between the two variables may be alleviated by introducing new means of deformation mechanisms. It may be assumed that this gap may be filled with secondary mechanisms such as dislocation climb and grain boundary sliding. Instead of choosing $\dot{\alpha}$ and $\dot{\varepsilon}$, a more realistic approach is to minimise the volume-averaged differential strain rate (*i.e.*, the difference between the macroscopic and local velocity gradients) to obtain $\dot{\alpha}$ and $\dot{\varepsilon}$. The average strain rate in the aggregate is:

$$\langle D \rangle = \sum_{p=1}^{N_{\text{crystal}}} P_p (L_{ij} - l_{ij})(L_{ik} - l_{ik}). \tag{3.39}$$

P_p is a tensor of the pth crystal that describes the shape currently taken by a crystal where its principal axes is proportional to the length of the symmetry axis of an ellipsoid. In order to proceed with the minimisation, one must assume that the volume average of the local velocity gradients assigned to each crystal be equal to the macroscopic velocity gradient constraining the aggregate:

$$\frac{1}{N} \sum_{p=1}^{N} (G_{ij}^p \dot{\varepsilon}^p - \epsilon_{ijk} \dot{\alpha}_o^p) = L_{ij}. \tag{3.40}$$

Once $\langle D \rangle$ is minimised, a system of $4N$ linear equations appear that describe $\dot{\alpha}$ and $\dot{\varepsilon}$. The solutions to these equations for a pth grain are given by:

$$\dot{\varepsilon_p} = \frac{(L_{ii+1} - L_{i+1i})(G_{ii+1}^p - G_{i+1i}^p) - 2G_{ij}^p L_{ij}}{(G_{kl+1}^p - G_{k+1l}^p)(G_{kl+1}^p - G_{k+1l}^p) - 2G_{kl}^p G_{kl}^p}, \tag{3.41}$$

and

$$\dot{\alpha_p} = \frac{\dot{\varepsilon_p}(G_{i+1i+2}^p - G_{i+2i+1}^p) + (L_{i+2i+1} - L_{i+1i+2})}{2}. \tag{3.42}$$

Rewriting eq. (3.32) in terms of the direction cosine matrix, the development of CPO for a pth crystal due to plastic deformation is finally given by:

$$\frac{Db_{ij}^p}{Dt} = \epsilon_{jkl} b_{il}^p \dot{\alpha}_k^p. \tag{3.43}$$

Eq. (3.43) describes the evolution of CPO by the alignment of the crystallographic axes due to macroscopic deformation imposed onto the aggregate.

Dynamic recrystallisation

When minerals are subjected to stress, they experience certain processes responsible for changes in their physical properties (i.e. CPO evolution) aside from plastic deformation. In metallurgy, this is known as recrystallisation. As defined by Guillope and Poirier (1979), recrystallisation is a process concerning the evolution of crystallographic volume fractions and the reorientation of grains in the absence of chemical change under an imposed macroscopic deformation. In our models, we restrict ourselves in the case where recrystallisation occurs during material deformation, called dynamic recrystallisation. In other cases, recrystallisation can also be induced following deformation (post-tectonic); this is referred to as static recrystallisation.

Albeit complex, dynamic recrystallisation is an essential phenomenon as it unfolds the thermo-mechanical history of rocks beneath the lithosphere given its link with texture evolution. Grains undergoing intracrystalline slip have been observed to increase the global strain energy, or the potential energy due to deformation. This is due to the piling-up of stored energy during dislocation creeping across all slip systems in individual crystals. Studies have shown that dynamic recrystallisation relaxes strain energy buildup (e.g. Guillope and Poirier, 1979; Karato, 1988; Kaminski and Ribe, 2001) in favor of grain growth. During this process, grain boundaries tend to migrate from regions of low-energy crystals to regions of high-energy crystals thereby reducing the global strain energy. This is further reduced by nucleation processes activated by regions of high dislocation density forming new grains. As a result, grain growth is activated by migration of grain boundaries.

Deformation mechanisms respond differently depending on grain size. With dynamic recrystallisation taking over, it is plausible for grain-boundary migration to be strong enough to generate large discrepancies between grain sizes. This disparity, in turn, would activate grain boundary sliding or diffusion creep among small grains (Karato, 1988). Larger grains, on the other hand, tend to deform by dislocation creep. Finally, there have been compelling arguments suggesting that at larger strains, dynamic recrystallisation dominates over intracrystalline slip. Further verified by simple shear experiments, olivine fast axes have been observed to reorient themselves along the direction of shear during dynamic recrystallisation (Zhang and Karato, 1995; Bystricky et al., 2000). Thus, such a mechanism is of great importance when modeling CPO in the asthenosphere where strains are large.

Since dynamic recrystallisation tends to decrease the global strain energy, we first define the strain energy of a grain as the energy stored due to the presence of dislocations. The strain energy is therefore just a function of the dislocation density ρ giving (Kaminski and Ribe, 2001):

$$E = \rho A \mu b^2, \tag{3.44}$$

where A is a non-dimensional constant, b is the magnitude of the Burgers' vector(*i.e.*, a vector quantity representing the magnitude and orientation of a line defect forming the dislocation) and μ is the shear modulus. Note that eq. (3.44) represents the stored energy in a grain in the absence of grain nucleation. Suppose that a grain has been immensely deformed to allow dynamic recrystallisation. The excessive strain energy brought about by the increase in dislocation density triggers nucleation of new strain-free subgrains. The presence of strain-free subgrains reduces the bulk strain energy of the crystal. The resulting strain is:

$$E = \alpha \rho A \mu b^2, \tag{3.45}$$

where α is the volume fraction of crystals that have not undergone recrystallisation. It is convenient to define a nucleation parameter λ that would quantify the efficiency of the nucleation process. The volume fraction of non-recrystallised grains is:

$$\alpha = e^{-\lambda \rho^2}, \tag{3.46}$$

In eq. (3.46), one may infer that the parent grain is one-hundred percent recrystallised ($\alpha = 0$) as lambda approaches infinity. On the other hand, should λ approach zero, the efficiency of nucleation becomes minuscule ($\alpha = 1$). Knowing eq. (3.45), we may construct the evolution equation for the volume fraction of crystals undergoing dynamic recrystallisation based from the fact that the volume fraction V_{frac} depends on the difference between the strain energy of a pth grain versus the average strain energy of the aggregate. In equation form, this gives us:

$$\frac{dV_{frac}}{dt} = -MV_{frac}(E - \overline{E}), \tag{3.47}$$

where $\overline{E}$ is the average strain energy of the aggregate, and M is called the grain-boundary mobility; a parameter quantifying the efficiency of grain boundary migration. Based on eq. (3.47), grain size increases if the average energy of the aggregate is larger than the grain strain energy. Otherwise, the grain shrinks.

To formulate the kinematic model of dynamic recrystallisation, one must first link

the relationship between dislocation density and the associated deformation mechanism. From Kaminski, Ribe, and Browaeys (2004), the non-dimensional strain energy associated to a pth crystal with phase m is:

$$E_m^{p\prime} = \sum_s I_p^s \exp\left[-\lambda_m (I_v^s)^2\right], \tag{3.48}$$

where I_p^s is a scalar known as the Schmidt factor for a given slip system s of a pth particle. Basically, the Schmidt factor measures the amount of shear stress resolved in a given slip system s for a deformed material. It is usually scaled with the non-dimensional RSS for a given phase m , and the reference shear rate $\dot{\varepsilon_o}$ giving:

$$I_p^s = \left(\frac{\tau_o}{\tau_m^s}\right)^{n-n_2} \left|\frac{\dot{\varepsilon}_p^s}{\dot{\varepsilon}_o}\right|^{\frac{n_2}{n}}, \tag{3.49}$$

where $\dot{\varepsilon}_p^s$ is the local shear rate, and n_2 is another stress exponent that relates the dislocation density with stress (Durham, Goetze, and Blake, 1977; Poirier, 1985; Bai and Kohlstedt, 1992).

As a matter of fact, the local shear rate is already obtained from eq. (3.41) whereas the reference shear rate is calculated given the macroscopic velocity gradient. The nucleation rate λ and the reference resolved shear stress for each phase on a slip system S on the other hand are obtained empirically following the work of numerous authors (e.g. Raleigh et al., 1971; Kohlstedt and Goetze, 1974; Bai, Mackwell, and Kohlstedt, 1991; Hanson and Spetzler, 1994; Jin, Bai, and Kohlstedt, 1994; Jung and Karato, 2001). Once the strain energy of each grain is quantified, eq. (3.47) is used to allow for the evolution of the crystallographic volume fractions via grain boundary migration. It is worth noting that the grain boundary mobility M is also estimated from experimental results. For pure olivine, it is shown that $M_{\text{olivine}} = 125 \pm 75$ e.g. Kaminski, Ribe, and Browaeys, 2004.

In summary, mineral aggregates subjected to an imposed macroscopic deformation develop net lattice preferred orientation by two deformation mechanisms: (1) plastic deformation which induces reorientation of each crystallographic axis (alignment of grains) dictated by eq. (3.43), and (2) dynamic recrystallisation which allows for the evolution of crystallographic volume fractions by grain boundary migration dictated by equation equation eq. (3.47).

3.3.4 Computational strategy

Backward tracing of flow streamlines

CPO evolution results from finite deformation accumulated over time. This requires the knowledge of the flow velocity **u**, and the deformation gradient **L**. So far, we have discussed how we can compute **u** and **L** given a single scalar field (*i.e.*, temperature) as an input. The next step now is to assign undeformed mantle minerals at each point in 3-D space subjected to the current configuration of the flow to allow for CPO evolution. Here, instead of allowing the aggregates to be advected forward in time, our strategy is first to perform a backward tracing of the flow along its path down to an initial time t_0. From here then we can assign an undeformed aggregate of a given size N_{crystal} at a point $[x(t = t_0), y(t = t_0), z(t = t_0)]$ followed by integrating the deformation experienced by the aggregate along this path. The result would therefore be a well-developed fabric corresponding to its present-day configuration.

Under the steady-state assumption, the path traversed by an advected particle is simply the flow streamline. We trace the streamline backwards starting at the present-day configuration of the mantle using a fourth-order Runge-Kutta scheme with adaptive time stepping. At each segment along the path, it is imperative to determine the velocity values at arbitrary locations. Fortunately, it is rather straightforward to compute for the velocity, the local velocity gradient tensor, and the time spent by the particle using linear interpolation schemes. Since strain accumulation and thus CPO evolution only depends on its flow trajectory, there is no interaction between neighboring aggregates under this numerical paradigm.

Forward integration along flow streamlines

For first-order observations with finite strain analysis, we apply eq. (3.28) at each segment along the streamline by performing a matrix multiplication of the local velocity gradient tensor with the fse. The total strain accumulated is then just the sum over the whole segment of the discrete contributions in each step along the path. As there are various ways to perform this procedure such as first-order central differencing, here we opt for a Runge-Kutta integration approach instead. At the end of the procedure, we expect to produce the current configuration of a deformed body in terms of the finite strain tensor **F** corresponding to the present-day structure and the present-day dynamics of the mantle.

Fig. 3.2 illustrates some properties of the fse at each grid point due to a negatively buoyant spherical temperature anomaly in a convecting upper mantle. Note that the

model is in 3-D having a grid resolution of $64 \times 64 \times 64$ elements. The image is a vertical cross-section that slices the anomaly evenly through its center. The natural strain is computed using eq. (3.30) where the maximum and minimum eigenvalues of $\mathbf{F}$ (*i.e.*, long and short axis of the fse) are approximated using the Jacobi algorithm. It refers to the amplitude of the accumulated finite deformation. The solid black lines describe the long axis of the fse. As observed, the fse representation maps out the anomaly quite well. This is because the amplitude of the deformation depends on the rheology of the material considered; and since we parameterised viscosity in terms of temperature, the cold, and therefore highly viscous anomaly tends to withstand deformation. Strong deformation on the other hand are ubiquitous in regions that surround the anomaly which correspond to shear zones. Finally, we observe that the orientation of the long axis aligns nearly parallel to the direction of flow. In rare occasions however, the orientation of the fse may be well away from the direction of flow since fse relates directly to the velocity gradient and not to the flow field itself. Thus to first-order, fse can be used as a proxy to infer convection in the mantle.

Micro-mechanical model for texture evolution

The fse framework is a quick yet primitive way to describe simple convection patterns. Its reliability however declines when we consider complex flow patterns or in cases where deformation is large. To accommodate such complexities, we resort to more sophisticated formalism such as that of micro-mechanical models for texture evolution. Here, we still employ fourth-order Runge-Kutta to integrate the local velocity gradients along the flow path. However, in lieu of an initially homogeneous sphere sphere defined by the identity matrix of $\mathbf{F}$, we assign an aggregate of size N_{crystal} whose constituents are initially randomly oriented. Most texture evolution models such as those included in the homogenisation (e.g. Tommasi et al., 2000) or in the kinematic categories (e.g. Kaminski and Ribe, 2001) however adhere to an 'averaged field' formalism. Here unlike 'full-field' models where the aggregate is explicitly deemed as a spatially extended body, it is not necessary to keep track of the interaction among crystals. Instead, the crystals within the aggregate are finite and are treated as a collective entity in a homogeneously isotropic medium whose properties are the weighted mean of the properties of each crystal. Thus in the averaged field formalism, the aggregate can also be initially regarded as a sphere which later transitions to an ellipsoid as it gets progressively sheared along the streamline. In principle, the evolution equations eqs (3.43) and (3.47) are evaluated for each step along the flow path with the accumulated local velocity gradients driving the deformation of the aggregate.

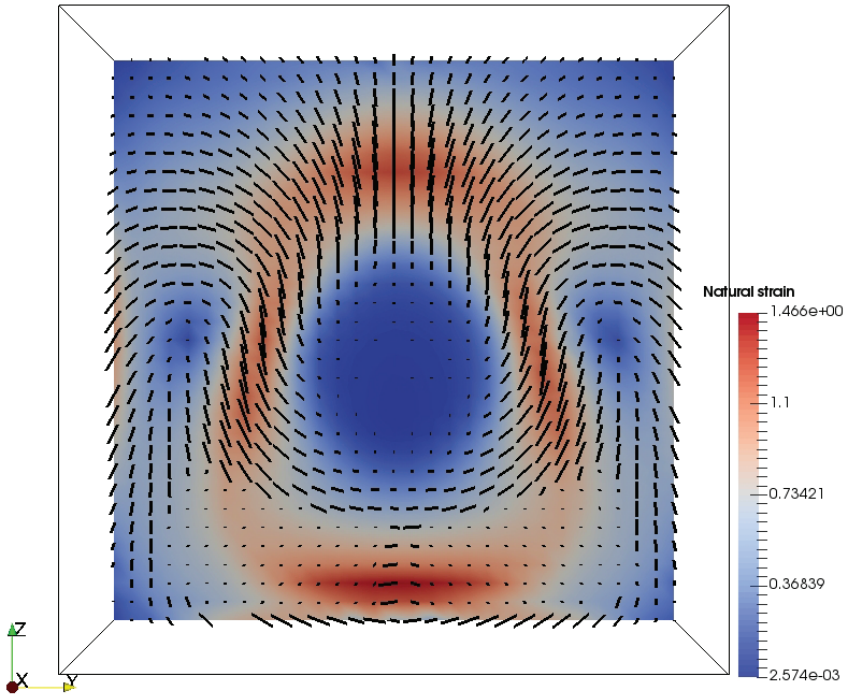

FIGURE 3.2: Finite strain representation of a 3-D deforming upper mantle due to a sinking, cold, spherical anomaly. The surface plot refers to the natural strain (*i.e.* logarithm of the ratio between the long and short axis of the fse) and the superimposed solid black lines are the orientations of the long axis. Since the amplitude of deformation depends on the mechanical properties of the body considered, we expect the cold, and highly viscous anomaly to resist deformation. To first-order, fse can be used to map the flow direction.

Throughout our work, we implement a particular texture evolution model whenever we are obligated to compute for flow-derived seismic anisotropy. We employ `D-Rex` (Kaminski, Ribe, and Browaeys, 2004), a kinematic model that calculates strain-induced crystallographic preferred orientation (CPO) of olivine and enstatite aggregates in arbitrary proportions via plastic deformation, dynamic recrystallisation, and grain boundary sliding. `D-Rex` is written in `Fortran90` and is highly-scalable. Here, we parallelised the code with `MPI` although in the inversions, as we will see later on, we have substituted it with a quick and cheap method because a hybrid `OpenMP` (flow computation) and `MPI`(texture evolution modeling) algorithm has not been considered.

As mentioned earlier, models of plastic deformation and dynamic recrystallisation depend on physical parameters (*e.g.*, slip system activities S for olivine and enstatite) that control each deformation mechanism. Thankfully, the control parameters in `D-Rex` have already been constrained from laboratory experiments (e.g. Zhang and Karato, 1995; Bystricky et al., 2000) and are extrapolated at scales consistent with upper mantle convection. The user-defined inputs are then the (1) initial orientation distribution function (always set to random thus corresponding to an isotropic material), (2) the macroscopic velocity gradient tensor which can be derived directly from the flow, (3) and the percentage of olivine and enstatite making up the aggregate. The raw output of `D-Rex` is the final configuration of the orientation distribution function (*i.e.*, volume fraction and orientation) after the aggregate has been progressively sheared along the streamline. In the way we implement `D-Rex`, it equates to the present-day structure of the mantle.

Knowing its single crystal elastic properties, it is plausible to construct a 21-component average elastic tensor $\mathbf{S}$ for an olivine polycrystal (*i.e.*, the aggregate) using various averaging methods. One notable method is Voigt averaging, that is, taking the dot product over n_s species between the crystallographic volume fractions and the average of the single crystal elastic tensor over N_{crystal} crystals (Mainprice, 1990). For a single mineral species, mathematically this translates to:

$$S_{ijkl}^{\text{voigt}} = \sum_{p=1}^{N_{\text{crystal}}} S_{ijkl} / N_{\text{crystal}}, \tag{3.50}$$

where S_{ijkl} is the elastic tensor associated to a specified grain p with an orientation α_p. S_{ijkl} can be obtained using a series of rotational transformations given by:

$$S_{ijkl} = R_{iq} R_{jr} R_{ks} R_{lt} S_{qrst}, \tag{3.51}$$

where S_{qrst} relates to the single crystal elastic tensor, and $\mathbf{R}$ is a rotational matrix derived from α_p. For any n-rank symmetric tensor, it is convenient to reduce its order for easier transcription. In the case of a fourth-rank S^{voigt}_{ijkl} tensor, we can represent such in terms of a 6×6 matrix with 21 independent coefficients using various notations. The elastic matrix S_{ij} in the Voigt notation is given by:

$$S_{ij} = \begin{pmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} & S_{16} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} & S_{26} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} & S_{36} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} & S_{56} \\ S_{61} & S_{62} & S_{63} & S_{64} & S_{65} & S_{66} \end{pmatrix} \tag{3.52}$$

3.4 Thermodynamic modeling of isotropic properties

Although temperature and pressure variations do not significantly affect anisotropic properties (e.g. Da Silva, Stixrude, and Wentzcovitch, 1997), the same cannot be said for the isotropic properties of rocks. Because of these thermodynamic constraints on various bulk properties including seismic wave velocities, most geophysicists are urged to implement free-energy minimisation algorithms to obtain mineral assemblages and compositions as a function of temperature, pressure, and bulk compositions. Recently developed semi-empirical models have then been utilised to estimate seismic wave velocities given the equilibrium mineral modes and compositions of several abundant ultramafic lithologies (*e.g.* periodotite and eclogite).

3.4.1 Computational strategy

`D-Rex`, the micro-mechanical model to be used for CPO evolution does not account for pressure and temperature dependence of the single crystal elastic parameters. In this work we employ `Perple_X` (Connolly, 2005; Connolly, 2009), a numerical software that solves the Gibbs free energy minimisation problem, to obtain the mineral assemblage of an olivine-dominated composition. We then use the thermodynamic formalism from Stixrude and Lithgow-Bertelloni (2011) to model the pressure and temperature dependence of the isotropic seismic wave velocities (V_P and V_S), and the local density ρ corresponding to the equilibrium mineral assemblage. From V_P,V_S, and ρ, it is rather straightforward to build the isotropic part of the elastic tensor.

Meanwhile, the elastic tensor given by `D-Rex` corresponds at ambient temperature and pressure conditions. One of the properties of the elastic tensor $\mathbf{S}$ is that it can be

broken down into several symmetry classes by performing a cascade of orthogonal projections. Because of this, we can decompose $\mathbf{S}$ into an isotropic part $\mathbf{S_0}$, which remains unaffected regardless of transformations, and an intrinsically anisotropic part $\delta\mathbf{S}$:

$$\mathbf{S}(T_0, P_0) = \mathbf{S_0}(T_0, P_0) + \delta\mathbf{S}(T_0, P_0) \,. \tag{3.53}$$

We replace the isotropic part of the tensor with the one computed from `Perple_X`. To account for the pressure and temperature dependence of the anisotropic part, it is scaled by the ratio between the shear modulus $\mu(T, P)$ at the given pressure and temperature, and the shear modulus at the reference temperature-pressure $\mu(T_0, P_0)$ (Gallego, Ito, and Dunn, 2013). Other methods are available, such as the use of first-order corrections around the elastic tensor at ambient T and P conditions (Estey and Douglas, 1986; Becker et al., 2006). Thus, the full elastic tensor, whose isotropic part depends on pressure and temperature is:

$$\mathbf{S}(T, P) = \mathbf{S_0}(T, P) + \frac{\mu(T, P)}{\mu(T_0, P_0)}\, \delta\mathbf{S}(T_0, P_0) \,. \tag{3.54}$$

3.5 Surface wave dispersion calculation

Up to this point, it is clear to us now how one could build a local elastic tensor without the imposition of any symmetry relations at any arbitrary location. More importantly, the general picture involving the reduction of dimensionality from a 21-component elastic tensor down to a single scalar field (*e.g.*, temperature) by introducing geodynamic and petrological constraints and with the help of a series of intermediate steps (*i.e.*, geodynamic flow modeling and texture evolution modeling) is more or less indisputable.

The next and final step of the geodynamic tomography problem is now the computation of local surface wave dispersion curves for a given depth profile of the elastic tensor $\mathbf{S}$. Once obtained, it is now rather straightforward to compare this result from real Earth observations, and thus infer the structure of the upper mantle that is exclusively explainable by its present-day dynamics using only this set of observations.

As discussed in Chapter 2, we can write the azimuthal dependence of surface wave phase velocities in terms of a Fourier series under the assumption that such a wavefield is propagating through a quasi-isotropic medium. Neglecting higher order terms (*i.e.*, 4θ coefficients of surface wave azimuthal anisotropy), the anisotropic surface wave

phase velocity is expressed as:

$$\mathbf{c}(T,\theta) = \mathbf{c_0}(T) + \mathbf{c_1}(T)\cos(2\theta) + \mathbf{c_2}(\mathbf{T})\sin(2\theta), \tag{3.55}$$

where T is the period and θ is the fast azimuth of surface wave anisotropy. Here, we can interpret $\mathbf{c}$ as a result of small anisotropic perturbation $\delta\mathbf{c}$ related to the 2θ terms around a reference, isotropic surface wave phase velocity $\mathbf{c_0}$.

In practice, $\mathbf{c_0}$ is usually treated in a fully-non-linear fashion whereas the model parameters (*e.g.* the elastic tensor $\mathbf{S}$) linking to the anisotropic perturbations $\mathbf{c_1}$ and $\mathbf{c_2}$ are more often that not linearised around an azimuthally-averaged version of $\mathbf{S}$ thanks to the quasi-isotropic assumption.

3.5.1 Computational strategy

Calculation of isotropic phase velocity dispersion curves

To a certain degree, isotropic surface wave phase velocities $\mathbf{c_0}$ are sensitive to an azimuthally-averaged version of the full elastic tensor $\mathbf{S}$. Surface waves thus 'see' the elastic medium as an equivalent vertically transverse isotropic (VTI) medium with five independent coefficients (*i.e.*, Love parameters). The VTI medium with Love parameters A, C, F, L, and N can be written in a coherent fashion as:

$$\mathbf{S}_{\mathrm{VTI}} = \begin{pmatrix} A & A-2N & F & 0 & 0 & 0 \\ A-2N & A & F & 0 & 0 & 0 \\ F & F & C & 0 & 0 & 0 \\ 0 & 0 & 0 & 2L & 0 & 0 \\ 0 & 0 & 0 & 0 & 2L & 0 \\ 0 & 0 & 0 & 0 & 0 & 2N \end{pmatrix}, \tag{3.56}$$

According to Montagner and Nataf (1986), the Love parameters are just a linear combination of the coefficients of the full elastic tensor $\mathbf{S}$. In terms of S_{ij}, the five depth

functions are given by:

$$\begin{aligned} A &= \frac{3}{8}(S_{11} + S_{22}) + \frac{1}{4}S_{12} + \frac{1}{2}S_{66}, \\ C &= S_{33}, \\ F &= \frac{1}{2}(S_{13} + S_{23}), \\ L &= \frac{1}{2}(S_{44} + S_{55}), \\ N &= \frac{1}{8}(S_{11} + S_{22}) - \frac{1}{4}S_{12} + \frac{1}{2}S_{66}. \end{aligned} \tag{3.57}$$

The isotropic phase velocities for both Love and Rayleigh waves can then be computed from an elastic model with A, C, F, L, and N through normal-mode summation in a spherical Earth (Saito, 1967).

To perform such calculations, we employ `DISPER80` (Saito, 1988), a software package that computes for isotropic phase velocity dispersion following a Runge-Kutta matrix integration scheme for any given depth profile of azimuthally-averaged elastic models. Here, the program inputs 1D depth profiles of: (1) the number of layers and their associated thicknesses, (2) the density ρ of each layer which in our case retrieved from PREM, azimuthally-averaged horizontally propagating (3) $SV-$wave obtained from $V_{SV} = \sqrt{L/\rho}$ and (4) $PH-$wave obtained from $V_{PH} = \sqrt{A/\rho}$, and three parameters describing the strength of anisotropy: (5) $S-$wave radial anisotropy $\xi = N/L$, (6) $P-$wave radial anisotropy $\phi = C/A$, and lastly (6) an ellipticity parameter $\eta = F/A - 2L$. The last parameter η is related to the velocity along the direction intermediate to the fast and slow seismic velocities. The program outputs the local isotropic surface wave dispersion curve at a number of discrete periods T and its associated sensitivity kernels to the elastic parameters (*i.e.*, a sensitivity kernel describes how the behavior of seismic waves is influenced by the changes in the Earth's structure with depth) for any T.

Calculation of the 2θ terms

The 2θ terms are sensitive to the depth functions G_c and G_s, B_c and B_s which relate to the azimuthal variations of the horizontally propagating $S-$ and $P-$wave, respectively. Surface wave azimuthal anisotropy thus 'sees' the elastic structure of the Earth as an effective horizontally transverse isotropic (HTI) medium given by these four

depth functions. Similar to the VTI parameterisation, the elastic constants of the effective HTI medium can be regarded as a linear combination of $\mathbf{S}$:

$$\begin{aligned} G_c &= \frac{1}{2}(S_{55} - S_{44}), \\ G_s &= S_{54}, \\ B_c &= \frac{1}{2}(S_{11} - S_{22}), \\ B_s &= \frac{1}{2}(S_{16} + S_{26}). \end{aligned} \tag{3.58}$$

Another convenient way to describe azimuthal anisotropy of a horizontally propagating $S-$ and $P-$wave in a weakly anisotropic medium is in terms of its strength and orientation:

$$G = \sqrt{G_c^2 + G_s^2}, \tag{3.59}$$

$$B = \sqrt{B_c^2 + B_s^2}, \tag{3.59}$$

$$\Psi_{\text{fast}} = 0.5 \arctan\left(\frac{G_s}{G_c}\right), \tag{3.60}$$

where G and B are the levels of azimuthal anisotropy for $S-$ and $P-$waves, respectively and Ψ_{fast} is their azimuth of fast propagation.

According to Montagner and Nataf (1986), the sensitivity kernels of the 2θ terms $\mathbf{c_1}$ and $\mathbf{c_2}$ are equivalent to that of $\mathbf{c_0}$. Under the quasi-isotropic assumption, the azimuthal variations in surface wave phase velocities in any period T can be evaluated through the following expressions:

$$c_1(T) = \int_{z=0}^{\infty} \left(B_c(z) \frac{\partial c_0(T)}{\partial A} + G_c(z) \frac{\partial c_0(T)}{\partial L} \right) dz, \tag{3.61}$$

$$c_2(T) = \int_{z=0}^{\infty} \left(B_s(z) \frac{\partial c_0(T)}{\partial A} + G_s(z) \frac{\partial c_0(T)}{\partial L} \right) dz. \tag{3.62}$$

Eqs (3.61) and (3.62) is straightforward to implement numerically using various numerical integration recipes available. Finally, the anisotropic surface wave dispersion curve $\mathbf{c}$ can be fully realised following eq. (3.55).

3.5.2 Fast forward calculations of seismic anisotropy

Finite strain framework for radial anisotropy

Texture evolution modeling deemed to be the computational bottleneck of the geodynamic tomography forward problem compared to flow modeling and surface wave dispersion calculation. It was documented that even when processed in parallel with more than 100 cores, the texture evolution computations in a $32 \times 32 \times 32$ grid yielded a computing time of 1.8×10^3 seconds. Thus, implementing a computationally expensive routine is unsuitable for an inversion that involves sampling-based parameter search approaches.

As pointed out by Becker et al. (2003) and Becker et al. (2006), the finite strain ellipsoid (fse) framework is a viable first-order approximation to showcase the fast axis of olivine CPO in most places. For this reason, we attempted to fast track the texture evolution modeling step by approximating it with an fse-based method. The inputs are the usual: (1) flow field $\mathbf{u}$ and (2) macroscopic velocity gradient tensor $\mathbf{L}$. Since the raw output is a finite strain tensor $\mathbf{F}$, we need an empirical formula that relates $\mathbf{F}$ to the elastic tensor $\mathbf{S}$ computed from a texture evolution model. However, since surface waves are not sensitive to every component of $\mathbf{S}$, we only seek to relate $\mathbf{F}$ to the azimuthally-averaged VTI tensor with components A, C, F, L, and N. Under the assumption that the 3-axis of the fse is aligned with the symmetry axis of a hexagonally projected $\mathbf{S}$, we can impose a rotational transformation of $\mathbf{S}$ onto the basis of $\mathbf{F}$ giving:

$$S^{\text{fse}}_{ijkl} = R^{\text{fse}}_{iq} R^{\text{fse}}_{jr} R^{\text{fse}}_{ks} R^{\text{fse}}_{lt} S_{ijkl}, \tag{3.63}$$

where $\mathbf{R}^{\text{fse}}$ is a rotation matrix derived from the eigenvectors of $\mathbf{F}\mathbf{F}^T$. This intermediate step is necessary to impose a robust correspondence between the fse and the elastic constants built from a texture evolution model. The next step is to find a scaling law among some properties of $\mathbf{F}$ with that of $\mathbf{S}^{\text{fse}}$. For convenience, we will use the third eigenvalue λ_3 of $\mathbf{F}$ corresponding to the length of its 3-axis, and compare it with the Love parameters A, C, L, and N obtained from a VTI projection of $\mathbf{S}^{\text{fse}}$. For the elastic parameter F, it is somehow more fitting to relate it with λ_{ave}:

$$\lambda_{\text{ave}} = \frac{1}{2} \ln \left(\frac{2\lambda_3}{\lambda_1 + \lambda_2} \right). \tag{3.64}$$

To test whether there is indeed a correspondence, we compute the finite strain tensor and the full elastic tensor from a texture evolution model in each point along a regular grid of $32 \times 32 \times 32$ elements for a 3-D deforming upper mantle due to a

negatively-buoyant, highly-viscous, spherical temperature anomaly located at the center of the box. After dealing with the appropriate rotational transformations and projections mentioned above, for any grid cell in our 3-D model we map a point with coordinates [λ_3, $A/C/L/N$] and [λ_{ave}, F].

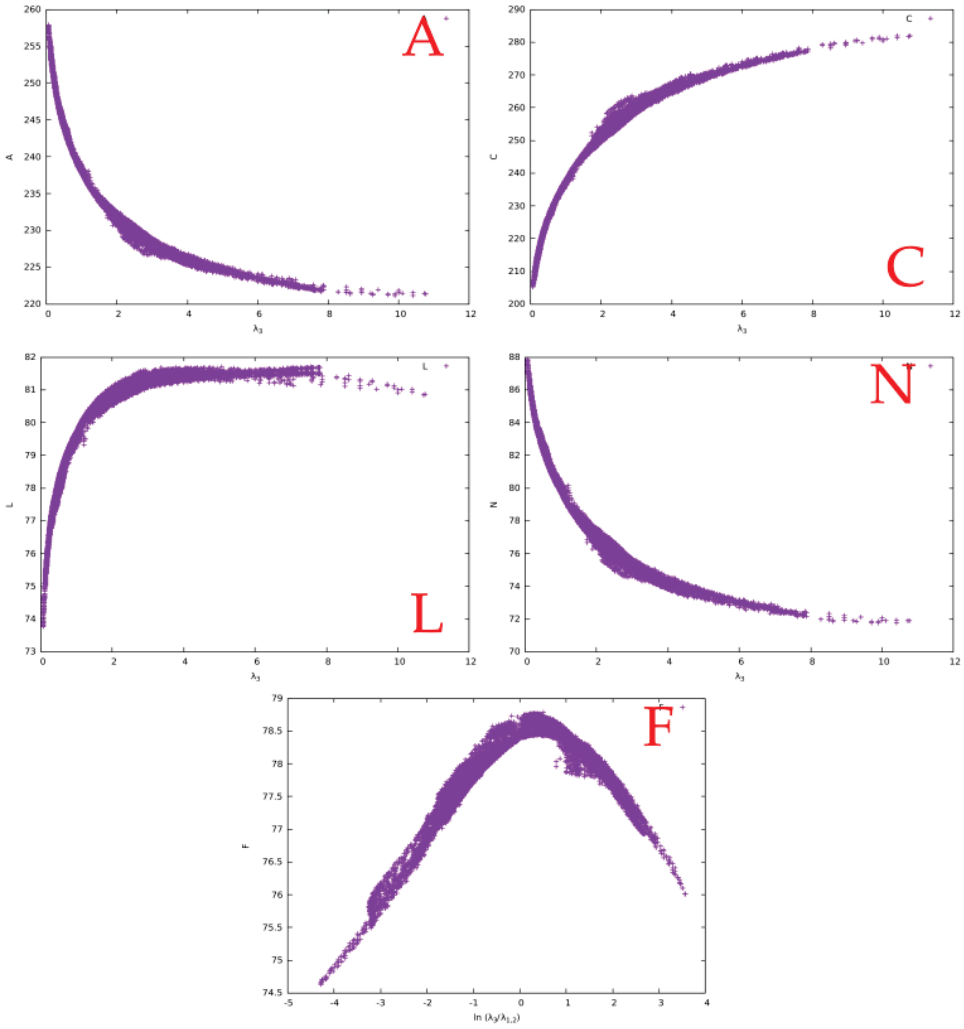

FIGURE 3.3: Relations between some properties of the finite strain ellipse and the elastic tensor computed from a texture evolution model. The plots clearly indicate that an empirical formula can be explicitly inferred from the correspondence between the fse and the elastic tensor. Note however that such is a result by imposing a strong assumption that the third axis of the fse is aligned with the symmetry axis of a hexagonally projected elastic tensor.

Fig. 3.3 illustrates the presence of strong correlation among the Love parameters A, C, L, N of an $\mathbf{S}^{\text{fse}}_{\text{VTI}}$ media with the third eigenvalue λ_3 of the fse, and F with λ_{ave}. From

here it is straightforward to infer an empirical relation between the parameters:

$$A = f_A(\lambda_3), \qquad C = f_C(\lambda_3),$$
$$L = f_L(\lambda_3), \qquad N = f_N(\lambda_3),$$
$$F = f_F(\lambda_{\text{ave}}).$$

Thus, $\mathbf{S}^{\text{fse}}_{\text{VTI}}$ can be viewed as a tilted transverse isotropic (TTI) medium with its symmetry axis aligning the 3-axis of the fse. Such a medium is also called an orthothropic medium defined by the five elastic parameters and two angles defining the symmetry axis: β is the angle of the symmetry axis with respect to the vertical, and Ψ is the azimuth of its projection. Fig. 3.4 shows a schematic representation of this medium. In

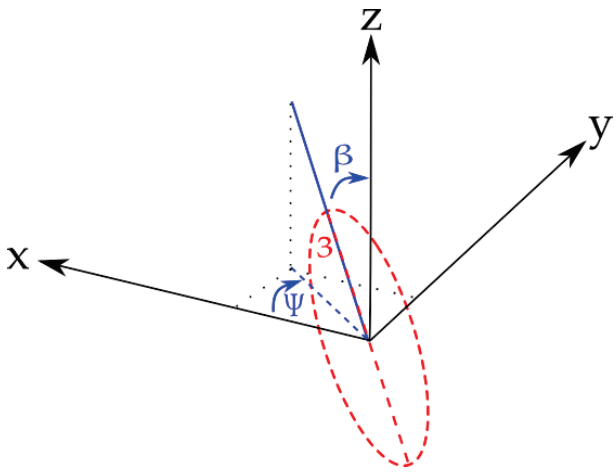

FIGURE 3.4: A tilted transverse isotropic (TTI) medium. Since the TTI is built from the basis of an fse framework (dashed red ellipsoid), its symmetry axis (solid blue lines) is equivalent to the 3-axis of the fse (dashed red lines). The dashed blue lines pertain to the horizontal projection of the symmetry axis. The angles β and Ψ represent the orientation of the symmetry axis. The Cartesian coordinate with axes x, y, and z represent the geographical frame of reference. A word of caution, the 3-axis does not necessarily equate to the long axis of the fse.

order to compute for isotropic surface wave dispersion curves, it is necessary to revert the $\mathbf{S}^{\text{fse}}_{\text{VTI}}$ back from the fse frame to the geographical frame since surface waves probe the elastic medium as an azimuthally-averaged structure with its symmetry axis aligning the geographical vertical axis z. To do this, we must determine the dip angle β from our TTI medium. Fortunately since the symmetry axis aligns with the 3-axis of the fse, we can just extract an angle from its third eigenvector $\mathbf{v}_3$ with respect to the z axis giving:

$$\beta = \arctan\left(\frac{v_{31}}{v_{33}}\right). \tag{3.65}$$

Together with Ψ, we then use these angles to rotate the TTI medium $\mathbf{S}^{\text{fse}}_{\text{VTI}}$ back to the geographical reference frame $\mathbf{S}_{\text{VTI}}$ with components A_0, C_0, F_0, L_0, and N_0. The transformations from $A, C, F, L, N \rightarrow A_0, C_0, F_0, L_0, N_0$ is (Montagner and Nataf, 1988):

$$
\begin{aligned}
A_0 &= \frac{A}{64}(41 + 20\cos 2\beta + 3\cos 4\beta) + \frac{C}{64}(9 - 12\cos 2\beta + 3\cos 4\beta) \\
&+ \frac{F + 2L}{32}(7 - 4\cos 2\beta - 3\cos 4\beta), \\
C_0 &= \frac{A}{8}(3 - 4\cos 2\beta + 4\cos 4\beta) + \frac{C}{8}(3 + 4\cos 2\beta + \cos 4\beta) \\
&+ \frac{F + 2L}{4}(1 - \cos 4\beta), \\
F_0 &= \frac{A}{16}(5 - 4\cos 2\beta - \cos 4\beta) + \frac{C + 4L}{16}(1 - \cos 4\beta) \\
&+ \frac{F}{8}(5 + 2\cos 2\beta + \cos 4\beta) + \frac{N}{2}(1 - \cos 2\theta), \\
L_0 &= \frac{A + C - 2F}{16}(1 - \cos 4\beta) + \frac{L}{4}(2 + \cos 2\beta + \cos 4\beta) \\
&+ \frac{N}{4}(1 - \cos 2\beta), \\
N_0 &= \frac{A + C - 2F}{64}(3 - 4\cos 2\beta + \cos 4\beta) \\
&+ \frac{L}{16}(5 - 4\cos 2\beta - \cos 4\beta) + \frac{N}{2}(1 + \cos 2\beta).
\end{aligned}
\tag{3.66}
$$

A neural network-based approximation for general patterns of elastic anisotropy

This motivated us to endeavor alternative methodologies to model the general patterns of elastic anisotropy. In this section, we use an artificial neural network (ANN) as a surrogate model g_{nn}, to approximate the forward model for texture evolution g_{CPO}. We consider a simple architecture of feedforward neural network called a multi-layer perceptron (MLP) with two hidden layers similar to the work of LeCun, Bengio, and Hinton (2015) defined by:

$$
g_{\text{nn}}(X_l) = \widehat{Y}_l = a_1\left(\sum_{k=1}^{N_{h1}} w^3_{kl} a_2\left(\sum_{j=1}^{N_{h2}} w^2_{jk} a_3\left(\sum_{i=1}^{N_x} w^1_{ij} X_i\right)\right)\right). \tag{3.67}
$$

The output $\widehat{Y}_l$ of the MLP is an estimate of the 21 independent coefficients of the stiffness tensor where l is the index pertaining to one element in the tensor. N_{h1} and N_{h2} are the sizes of the two hidden layers considered, and N_x is the size of the input vector. We design the network such that the input X contains the deformation history along a flow streamline. The streamline is divided into 200 time steps. Each step

contains one L_{ij} matrix and one corresponding dt. Thus, each step has 10 independent components as inputs. The number of inputs in the neural network first layer is $N_x = 2000$ (see eq. (3.67)). The functions a_1, a_2, and a_3 are known as activation functions whose purpose are to introduce non-linearity to the output of one neuron and to constrain its output to a desired range and distribution. Here, we choose them as default rectified linear unit functions to allow faster convergence (Pedregosa et al., 2011). Lastly, the w's refer to the weights which reflect the significance of a given neuron.

To build a suitable surrogate model to D-Rex, the weights w^1, w^2, and w^3 have to be adjusted to the proper value. This is performed by minimising a loss function which is the difference between the training outputs $g_{\mathrm{CPO}}(X)$ and the output of the network itself $g_{\mathrm{nn}}(X)$ using a stochastic gradient descent algorithm (Rumelhart, Hinton, and Williams, 1985). Formally, the loss function is a squared L_2 norm and takes the form:

$$Loss(Y, \widehat{Y}, w) = \frac{1}{2}\|Y - \widehat{Y}\|_2^2 + \frac{\lambda}{2}\|w\|_2^2 \,. \tag{3.68}$$

The second term constrains the weights to avoid data over-fitting, where α is a regularisation parameter that quantifies the degree of penalisation. The weights are updated iteratively by subtracting its current value from the gradient of the loss function with respect to the weights:

$$w_{i+1} = w_i - \epsilon \nabla loss_i \,, \tag{3.69}$$

where ϵ is the learning rate which controls the step-size for updating the weights, and i is the iteration step. The training achieves convergence when the tolerance value tol for the loss function is reached. However, the algorithm may also be stopped once the maximum number of iterations is reached.

The network is trained by considering 30 flow models, each comprising M temperature anomalies to drive thermal convection. Each anomaly has a random position and size, and can either be positively or negatively buoyant. This is to ensure that each flow path we define is unique enough so that the network can learn a variety of input-output combinations. Here, we acknowledge that the choice of flow models is not enough to be able to predict seismic anisotropy in the most general case. However, in this work, we only attempt to predict anisotropy for a small class of flow models (convection due to a collection of temperature anomalies). Since only such classes of models are tested, we can restrict ourselves to this type of model when training the network.

One training input corresponds to one deformation history along a streamline whereas one training output corresponds to one stiffness matrix computed with D-Rex. The training set can be represented as a matrix containing the stiffness coefficients and the

TABLE 3.1: Neural network parameters.

N_{train}	N_x	N_y	N_{h1}	N_{h2}	λ	ϵ	tol	Max iterations
1.2288×10^5	2000	21	100	50	0.1	1.0×10^{-3}	1.0×10^{-4}	1000

input parameters given by $[Y_{l=1,21}, X_{i=1,Nx}]_{n=1,N_{train}}$ where N_{train} is the number of training sets. Thus, the training inputs are of the size $[2000, N_{train}]$ and the training outputs are of the size $[21, N_{train}]$. In this problem, 16^3 input-output combinations for each 3-D flow model are used to train the network. In total, there are $N_{\text{train}} = 1.2288 \times 10^5$ training sets for the network to learn from.

We adopt the Python package `scikit-learn` to train the network (Pedregosa et al., 2011). Table 3.1 below summarises the parameters used to design and build the network.

The network is tested by considering a 3-D deformation due to a sinking anomaly that is not part of the training input. For each pixel in the 3-D map of full seismic anisotropy, a point is plotted containing the predicted elastic coefficients ($y-$axis) versus the elastic coefficients computed with `D-Rex` ($x-$axis) (Fig. 3.5). Each panel corresponds to one element of the anisotropic part of the elastic tensor. The plots are interpreted such that if the slope is one, then the predictions perfectly match the theoretical values. Minor offsets are attributed to several factors such as the imperfections of the network architecture, and the number of training data used.

The elastic tensor computed from g_{nn} is projected into both a VTI medium, thus having elastic parameters A, C, F, L, & N, and radial anisotropy strength ξ, ϕ, & η; and an HTI medium, with parameters G_s, G_c, B_s, & B_c. We compare the results with D-Rex by plotting 1-D marginal distributions of the residuals of each seismic parameter. Each parameter contains a small bias very close to zero which is attributed to the minimisation of the L_2 loss function.

Model comparison of radial anisotropies

We test the robustness of the two surrogate models (*i.e.*, fse framework and neural network) with respect to the true model (*i.e.*, `D-Rex`) by comparing 3-D radial anisotropy ξ profiles obtained from these models. The ξ models correspond to a 3-D deforming upper-mantle due to a cold, sinking, and highly-viscous spherical anomaly placed at the center of a 400 km $\times$ 400 km $\times$ 400 km box. Fig. 3.7 shows the vertical cross-section of the true ξ model. Patches of positive radial anisotropy $\xi > 1$ correlate well with expected lateral flow patterns whereas patches of negative radial anisotropy $\xi < 1$ correlate well with vertical flow patterns. At the location of the spherical anomaly, ξ is

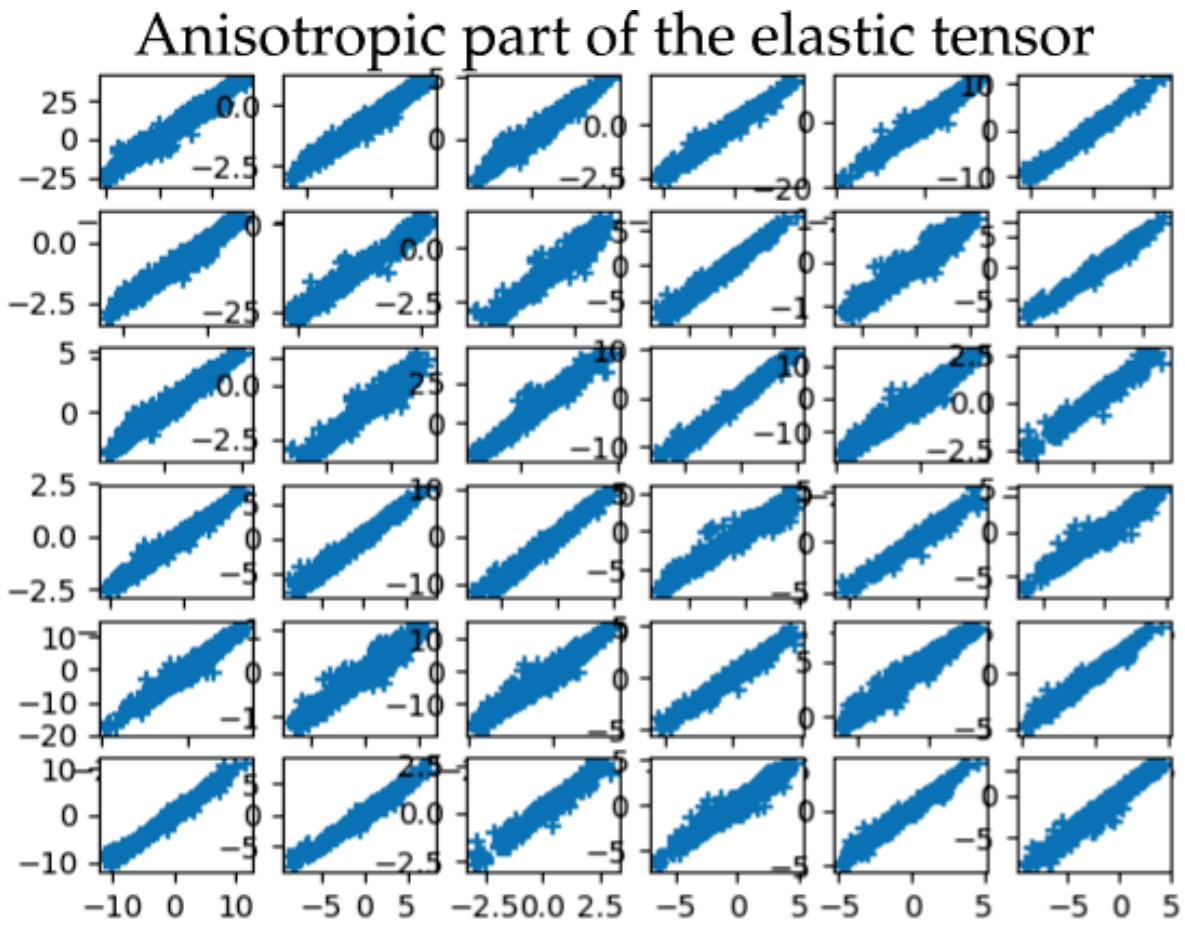

FIGURE 3.5: Predicted elastic coefficients from g_{nn} versus theoretical elastic coefficients from g_{CPO}. Each panel corresponds to one element of the δS_{ij} matrix. A slope of unity simply means that the predictions perfectly match the theoretical values.

approximately unity which implies that the material is more or less isotropic. This result is unsuprising since anisotropy throughout this whole context is induced purely by deformation but because of its rigidity, the anomaly is less likely to deform thus producing dismal levels of anisotropy.

Fig. 3.8 shows vertical cross-sections of ξ models obtained from the fse framework (top left panel) and neural network (top right panel). To compare the results, we calculate for each pixel in the maps the relative error (in %) with respect to the true model for the fse framework (bottom left panel), and the neural network (bottom right panel). We observe the radial anisotropy produced from an fse approximation to be no much different from that of a texture evolution model, with most of the uncertainties of about 2% confined along the edges that delineate large variations in radial anisotropy. On the other hand, the errors in the neural network appear to be more sporadic, with

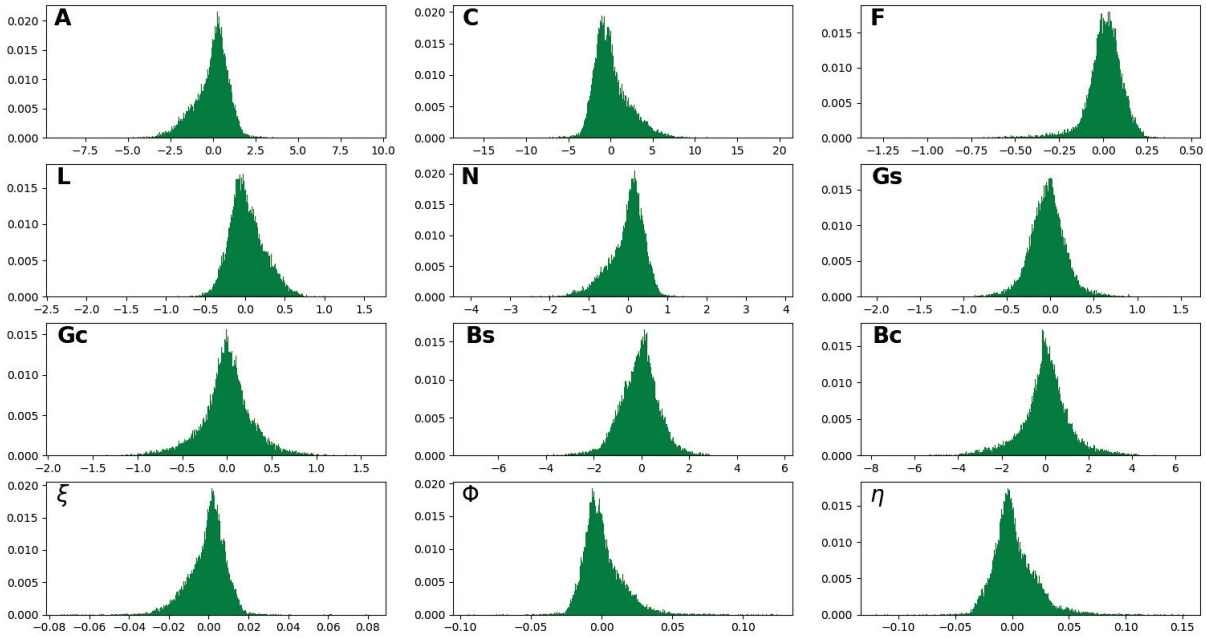

FIGURE 3.6: 1-D marginal distribution of the difference between $g_{\mathrm{CPO}}(X)$ and $g_{\mathrm{nn}}(X)$ in terms of the VTI and HTI-projected elastic tensor.

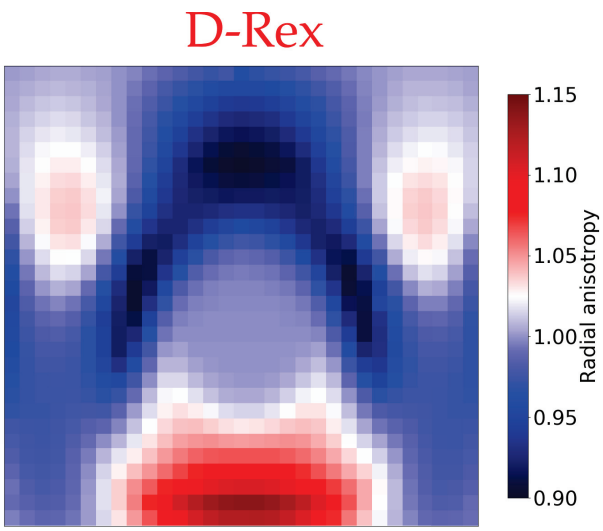

FIGURE 3.7: Radial anisotropy ξ computed from D-Rex. The structure corresponds to a 3-D deforming upper mantle due to a negatively buoyant, highly-viscous, and rigid spherical anomaly. Regions of $\xi > 1$ are interpreted in terms of horizontal flow whereas $\xi < 1$ are vertical flow. At the center where the anomaly is located, $\xi \approx 1$ implies that the anomaly resists deformation and is essentially isotropic.

large uncertainties of about 5-6% flocking across the center top of the box. Such results are expected since the empirical relations inferred from the fse framework were built

TABLE 3.2: Computation times for each subroutine in the forward model.

Routine	D-Rex	FSE	ANN	Flow	Dispersion	Training
Time (s)	73919.83	13.17	21.55	6.6	119.63	603.85

upon an elastic model that is uniquely defined by the sinking spherical temperature anomaly setup. Therefore should this hypothesis be further validated with other types of flow, then the fse framework is a viable substitute to texture evolution models when incorporated in an inversion procedure for radial anisotropy from isotropic surface wave phase velocity dispersion measurements. Apart from the fact that it is easy to implement, its main advantage lies on its cheap computational cost, that is, it is more than three orders of magnitude faster ($\approx$13.17 seconds) than `D-Rex`. As of the moment, this method only thrives in the case of an azimuthally-averaged VTI medium whereas correlations based on an HTI medium and the lateral projection of the fse fails. As such, this method is currently not applicable for the inversion of the azimuthal components of surface waves for azimuthal anisotropy. In contrast, the neural-network architecture is more general in a sense that is tailor-made to handle a multitude of flow patterns induced by a number of temperature anomalies. Similar to the former, the relative speed-up of neural network is over three orders of magnitude compared to performing texture evolution calculations with `D-Rex`.

Table 3.2 shows the computation times for computing anisotropy from `D-Rex`, fse framework, and neural network. For reference, we also give the computation times for network training, flow modeling, as well as for surface wave dispersion curve calculations. Each routine in the forward problem has been executed in a serial fashion for the sake of comparison.

3.6 Conclusion

In this chapter, we formulate the full forward problem to geodynamic tomography. We presented our computational strategy to derive azimuthally-varying surface wave dispersion curves (the data) from a single scalar field - temperature (the model). The four main steps can be summarised as follows:

1. From a temperature field T, calculate the instantaneous solution of upper mantle flow. This outputs the flow field $\mathbf{u}$, and the macroscopic velocity gradient tensor $\mathbf{L}$.

2. From $\mathbf{u}$ and $\mathbf{L}$, calculate the anisotropic part of the elastic tensor $\delta\mathbf{S}$ at any arbitrary position via texture evolution modeling.

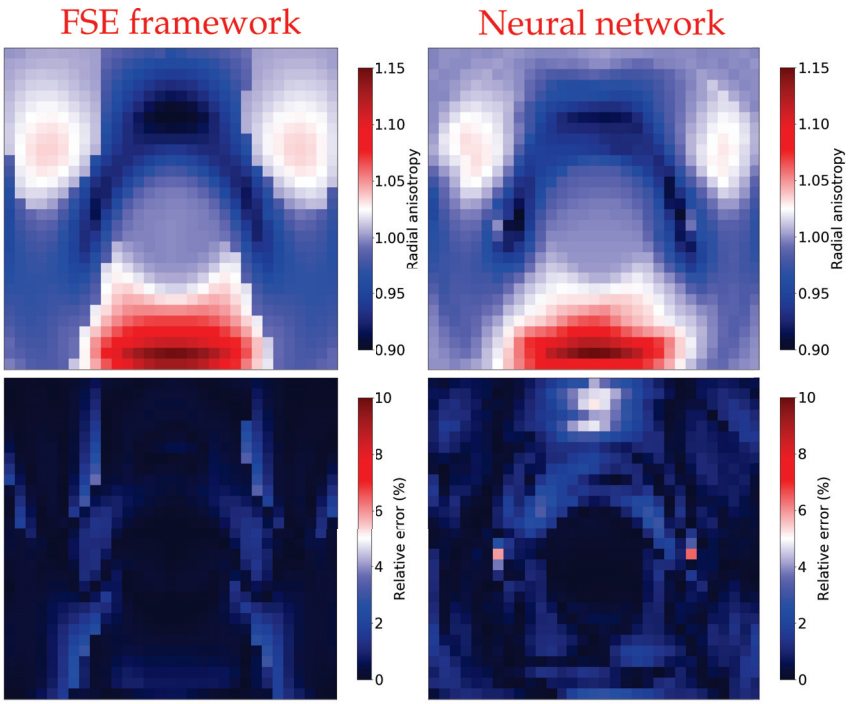

FIGURE 3.8: Top panels: Radial anisotropy ξ computed from: fse framework (left), and neural network (right). Bottom panels: percentage error in ξ with respect to the true model for: fse framework (left), and neural network (right). Results show that the fse framework reproduces better the true model. This is expected since the empirical relations were built upon an elastic tensor solely associated with the sinking anomaly setup whereas the neural-network is trained from a family of flow patterns induced by several temperature anomalies.

3. From T (and hydrostatic pressure P), calculate the pressure and temperature dependence of the isotropic part $\mathbf{S_0}$ using a thermodynamic model for a given bulk composition.

4. From a model of $\mathbf{S}$, calculate the azimuthally-varying surface wave dispersion curve for Love and Rayleigh wave phase velocities.

Fig. 3.9 shows a schematic representation of the full forward problem to geodynamic tomography (in green). For comparison, we also include traditional tomography (in red) which covers only a small portion of the full forward problem.

The estimated dispersion curves can then be compared with observed data in an inverse approach. Because the full forward problem is highly non-linear and comprises several intimately-related sub-procedures, the Bayesian inversion scheme is very much suitable to handle such complexity. In the next two chapters, we will show how such

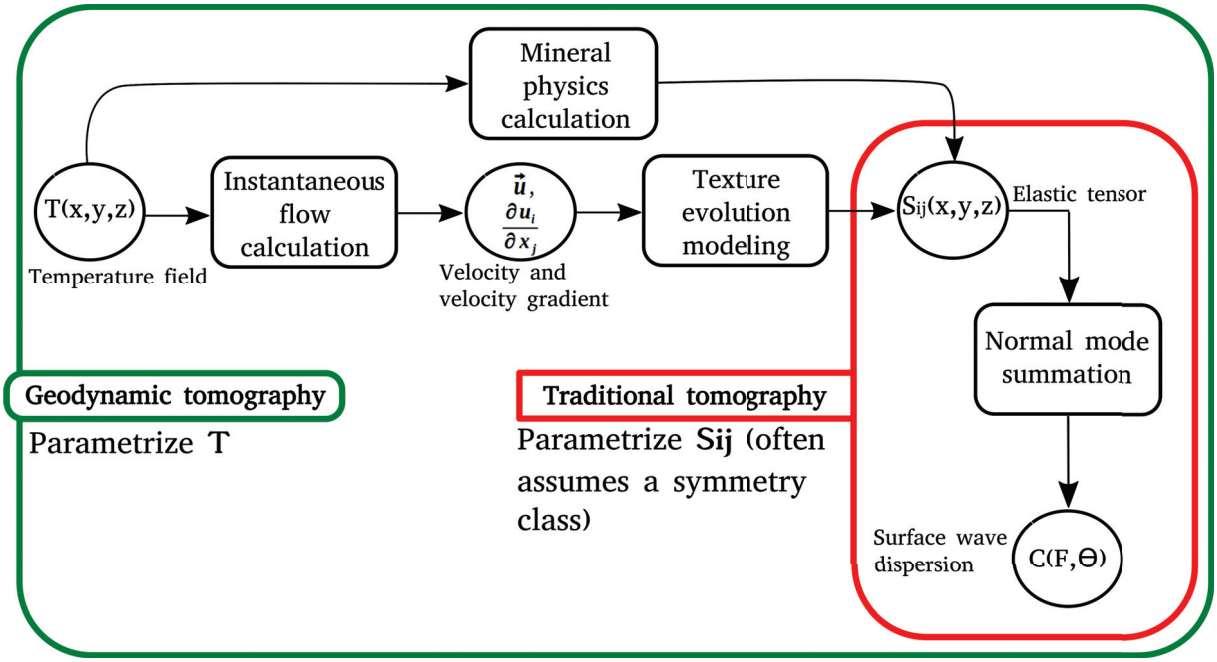

FIGURE 3.9: Geodynamic tomography (green) in comparison with traditional tomographic techniques (red). In geodynamic tomography, the unknown model to be inverted for is the temperature field denoted by T, whereas in traditional tomography, the model is a fourth-order elastic tensor S_{ij} with 21 independent coefficients. Often, seismologists assume a hexagonally symmetric medium onto S_{ij} to reduce the dimensionality of the would be inverted model. The complete forward model (in green) is cast in a Bayesian McMC framework. One of the advantages of geodynamic tomography is the reduction of unknown model parameters due to constraints from geodynamics.

a complex forward problem coupled with a Bayesian inversion approach is a potentially powerful method to constrain the complete pattern of upper mantle deformation and therefore elucidate some aspects associated with the Earth's internal structure and present-day dynamics.

Chapter 4

Geodynamic Tomography: Joint Inversion of Love and Rayleigh Waves with Azimuthal Variations

Geodynamic Tomography: Constraining Upper Mantle Deformation Patterns from Bayesian Inversion of Surface Waves

Published in *Geophysical Journal International*

J.K. Magali[1], T. Bodin[1], N. Hedjazian[1], H. Samuel[2], S. Atkins[3]

[1] Université de Lyon, UCBL, CNRS, LGL-TPE, 69622 Villeurbanne, France.

[2] Université de Paris, Institut de Physique du Globe de Paris, CNRS, F-75005 Paris, France.

[3] Laboratoire de Géologie, Ecole Normale Supérieure, PSL Res. Univ, Paris, France.

4.1 Summary

In the Earth's upper mantle, seismic anisotropy mainly originates from the crystallographic preferred orientation (CPO) of olivine due to mantle deformation. Large-scale observation of anisotropy in surface wave tomography models provides unique constraints on present-day mantle flow. However, surface waves are not sensitive to the 21 coefficients of the elastic tensor, and therefore the complete anisotropic tensor cannot be resolved independently at every location. This large number of parameters may be reduced by imposing spatial smoothness and symmetry constraints to the elastic tensor. In this work, we propose to regularise the tomographic problem by using constraints from geodynamic modeling to reduce the number of model parameters. Instead of inverting for seismic velocities, we parameterise our inverse problem directly in terms of physical quantities governing mantle flow: a temperature field, and a temperature-dependent viscosity. The forward problem consists of three steps:

(1) calculation of mantle flow induced by thermal anomalies, (2) calculation of the induced CPO and elastic properties using a micro-mechanical model, and (3) computation of azimuthally-varying surface wave dispersion curves. We demonstrate how a fully non-linear Bayesian inversion of surface wave dispersion curves can retrieve the temperature and viscosity fields, without having to explicitly parameterise the elastic tensor. Here, we consider simple flow models generated by spherical temperature anomalies. The results show that incorporating geodynamic constraints in surface wave inversion help to retrieve patterns of mantle deformation. The solution to our inversion problem is an ensemble of models (*i.e.*, thermal structures) representing a posterior probability, therefore providing uncertainties for each model parameter.

4.2 Introduction

To our knowledge, no such study yet exists where mantle deformation has been inferred directly from surface wave observations. In summary, two challenges remain as an open problem to interpret surface wave tomography:

1. Surface waves lack the resolving power to constrain anisotropy in space, and seismologists impose non-data-driven constraints to the tomography problem.

2. Surface waves are sensitive to a restricted number of elastic parameters. This is addressed by assuming symmetry relations to the elastic tensor, with some even accounting for transverse isotropic medium with a tilted axis of symmetry. However, this limits its interpretation in terms of deformation patterns.

We propose a complementary approach to estimate the full elastic tensor. This involves the incorporation of geodynamic and mineral physics modeling constraints: the textural evolution of peridotite aggregates during their deformation in the convective mantle. We propose a method to invert directly for the temperature field that produces convective flow and texture evolution. Modeling intrinsic anisotropy in this way removes the issue of low sensitivity from seismic waves since the elastic tensor is not explicitly inverted for, but instead computed directly from texture evolution models. Additionally, the inversion is performed using a Bayesian sampling algorithm, hence provide uncertainties on the obtained temperature field.

In Section 4.3, we explain how geodynamic tomography is implemented, starting with the model parameterisation, followed by the forward problem, the data, and finally the Bayesian inversion scheme. This is followed by Section 4.4, where we apply the method to synthetic data obtained from prescribed temperature fields.

4.3 Methodology

Geodynamic tomography involves two main procedures: (1) evaluate the forward model completely, and (2) implement a fully Bayesian non-linear inversion scheme with an McMC sampling technique. The solution of our inversion scheme is a posterior distribution of thermal structures and their corresponding uncertainty bounds.

4.3.1 Model parameterisation

To parameterise the 3-D thermal structure in a Cartesian domain (x,y,z), we build a basis containing spherical temperature anomalies, on top of an adiabatic temperature gradient. Mathematically, this translates to:

$$T(\mathbf{r}) = T_{\text{background}}(\mathbf{r}) + \sum_{i=1}^{M} T^{i}_{\text{anomaly}}(\mathbf{r}) \,, \tag{4.1}$$

where the background temperature is assumed to be linear and only a function of depth z:

$$T_{\text{background}}(\mathbf{r}) = T_0 + \left(\frac{z}{L_s} - 1\right)(T_0 - 1200\ \text{K}) \,, \tag{4.2}$$

and M is the number of spherical anomalies, $\mathbf{r} = (\mathbf{x}, \mathbf{y}, \mathbf{z})$ defines any point in the 3-D volume, T_0 is the temperature at the bottom (*i.e.*, also the reference value), and L_s is the characteristic length scale. Each anomaly has a distinct size, temperature, and position. We define the basis function for one given spherical anomaly as:

$$T_{\text{anomaly}}(\mathbf{r}) = -\frac{T_c}{2}\left[1 - \tanh\left(\frac{\beta}{L_s}\left(\|\mathbf{r} - \mathbf{r_0}\| - \frac{R}{2}\right)\right)\right] \,, \tag{4.3}$$

where T_c is maximum temperature anomaly reached at the center of the sphere $\mathbf{r_0} = (x_0, y_0, z_0)$, and R controls its size. These five variables are unknown model parameters to be inverted for in our problem. The non-dimensional constant $\beta = 20$ controls the sharpness of the temperature gradient. Additional details can be found in Appendix 4.6.1.

We model the medium rheology by assuming a temperature-dependent viscosity, following the Frank-Kamenetskii approximation to Arrhenius-type viscosity. Here, we only invert for a dimensionless scalar constant E, which plays a similar role to the conventional activation energy (*i.e.*, the sensitivity of viscosity to temperature). The

viscosity field is described by:

$$\eta(\mathbf{r}) = \eta_0 \exp\left[-E\,\frac{(T(\mathbf{r}) - T_0)}{T_0}\right], \tag{4.4}$$

where η_0 is a reference value for viscosity. The total number of parameters defining the model is therefore $5M + 1$, and the corresponding model vector $\mathbf{m}$ is defined as:

$$\mathbf{m} = [\, E, {x_0}^i, {y_0}^i, {z_0}^i, R^i, {T_c}^i, ..., {x_0}^M, {y_0}^M, {z_0}^M, R^M, {T_c}^M \,]\,. \tag{4.5}$$

4.3.2 The forward problem

The forward problem involves three main steps: (1) regional flow modeling in 3-D Cartesian coordinates, (2) modeling texture evolution and computation of the full elastic tensor, and (3) computation of seismic surface wave dispersion curves. We enhance the computational efficiency in Step 2 by using a surrogate model based on an artificial neural network (ANN) to compute the deformation-induced anisotropy.

Flow model

For our instantaneous flow models, we consider the buoyancy-driven convection of a highly-viscous, Newtonian, and incompressible fluid in a 3-D Cartesian coordinate system. The flow is subjected to free-slip boundary conditions. The system of equations describing the flow is given by:

$$\nabla \cdot \mathbf{u} = 0\,, \tag{4.6}$$

and

$$-\nabla P + \nabla \cdot [\eta(\nabla \mathbf{u} + \nabla \mathbf{u}^T)] + \rho\, g\, \hat{e}_g = 0\,, \tag{4.7}$$

where $\mathbf{u}$ is the flow velocity, P is the dynamic pressure, and $\hat{e}_g$ is a unit vector pointing towards the direction of gravity. We assume density ρ to be a function of temperature T using a linear equation of state controlled by a thermal expansion coefficient α, where $\rho(T) = \rho_0 - \rho_0\,\alpha(T - T_0)$. The dimensional values of the governing parameters are listed in Table 4.1. The Stokes equations are discretised using a finite-volume approach (e.g., Patankar, 1980; Albers, 2000), and are solved using the coupled iterative geometric multigrid method using V-cycles (Brandt, 1982; Gerya, 2010), yielding linear convergence with the number of unknowns. The complete code is parallelised with `OpenMP`. The accuracy of the numerical solution has been benchmarked against numerical and analytical solutions (Samuel, 2012a; Samuel, 2018).

TABLE 4.1: Dimensional parameters that define the Rayleigh number.

Symbol	Parameter	Value
η_0	Viscosity	10^{21} Pa s
α	Thermal expansion	$2\ 10^{-5}$ K^{-1}
g	Gravity	9.81 m/s^2
L_s	Layer thickness	400 km
T_0	Temperature scale	1900 K
k	Thermal diffusivity	10^{-6} m^2/s
ρ_0	Density	3800 kg/m^{-3}
Ra	Rayleigh number	$1.05\ 10^6$

Although the code accommodates sharp viscosity contrasts, the latter tend to reduce the speed of convergence. Sharp viscosity contrasts are avoided in this study since smooth thermal structures are considered in our prior distribution. The velocity gradients are obtained by second-order finite differences of the computed velocity field.

Modeling intrinsic anisotropy

Upper mantle minerals develop CPO due to progressive shearing along a flow path. We initially model CPO evolution by employing `D-Rex`, a kinematic model of strain-induced crystal lattice preferred orientation of olivine and enstatite aggregates developed by (Kaminski, Ribe, and Browaeys, 2004). The crystal aggregates respond to an imposed macroscopic deformation by two mechanisms: (1) dislocation creep which induces reorientation of each crystallographic axis, and (2) dynamic recrystallisation, which allows for the evolution of crystallographic volume fractions by grain nucleation and grain boundary migration. In this study, we only consider pure olivine of type-A fabric corresponding to dry upper mantle conditions. The raw output of `D-Rex` is a set of crystallographic orientations and volume fractions for a given aggregate. Finally, its effective elastic properties can be estimated with an averaging scheme such as the Voigt average (Mainprice, 1990). In Voigt notation, the elastic tensor can be represented as a 6×6 matrix with 21 independent elastic coefficients.

We model the temperature and pressure dependence of the isotropic seismic wave speeds of an olivine mantle composition (V_p and V_s) using `Perple_X` (Connolly, 2005; Connolly, 2009) with the thermodynamic model from Stixrude and Lithgow-Bertelloni (2011).

Fast forward approximation for texture evolution calculations

Sampling-based techniques such as Markov chain Monte Carlo schemes can be applied to most geophysical inverse problems provided that the parameter space can be sampled efficiently. In some cases however, the forward model is computationally expensive, and sampling-based techniques may not be efficient at approximating a multi-dimensional probability distribution. Fast approximations of the forward model, such as artificial neural networks (ANN) are sometimes therefore used. Such approximations, however, lead to a theoretical error (also called modeling error). The form of these errors can be estimated and modeled as a Gaussian probability distribution with its resulting variance being accounted for in the likelihood function during the inversion process (Hansen et al., 2014b; Köpke, Irving, and Elsheikh, 2018). In our case, the computational bottleneck is clearly the texture evolution modeling, which we addressed by using an ANN-based surrogate model to approximate seismic anisotropy.

In the field of geophysics, these methods have already been used to approximate the inverse function in a variety of applications in seismology (e.g., Meier, Curtis, and Trampert, 2007; Käufl et al., 2014; Hansen and Cordua, 2017; Hulbert et al., 2019), and in geodynamics (e.g., Shahnas, Yuen, and Pysklywec, 2018). Among these studies, some have already applied surrogate models for fast forward approximations in sampling-based techniques (Hansen and Cordua, 2017; Köpke, Irving, and Elsheikh, 2018; Conway et al., 2019; Moghadas, Behroozmand, and Christiansen, 2020).

These networks are composed of highly non-linear functions that can be trained to approximate a non-linear mapping between an input and an output (Bishop et al., 1995). To approximate such a function, one needs to train this network given a collection of training data consisting of a set of input and output pairs. In this work, we replicate the operator for texture evolution, which we now denote as g_{CPO}. Flow streamlines with assigned local velocity gradients are fed into the network as training inputs. The training output contains the anisotropic part of the elastic tensor $\delta S(T_0, P_0)$ computed from `D-Rex`. The package `scikit-learn` in Python is used to train the network.

Once the network is trained, which we denote as the operator g_{nn}, we perform a simple numerical test of 3-D deformation due to a cold spherical temperature anomaly, and applied both operators to output seismic anisotropy. Fig. 4.1 shows the percentage of total anisotropy found by the two methods. We observe comparable levels of anisotropy. Moreover, the approximation also appears to capture some important features such as the absence of anisotropy at the center, which is ascribed to the larger viscosity of the anomaly in this region. However, the surrogate model tends to underestimate the total anisotropy, which may be attributed to the simplicity of the network

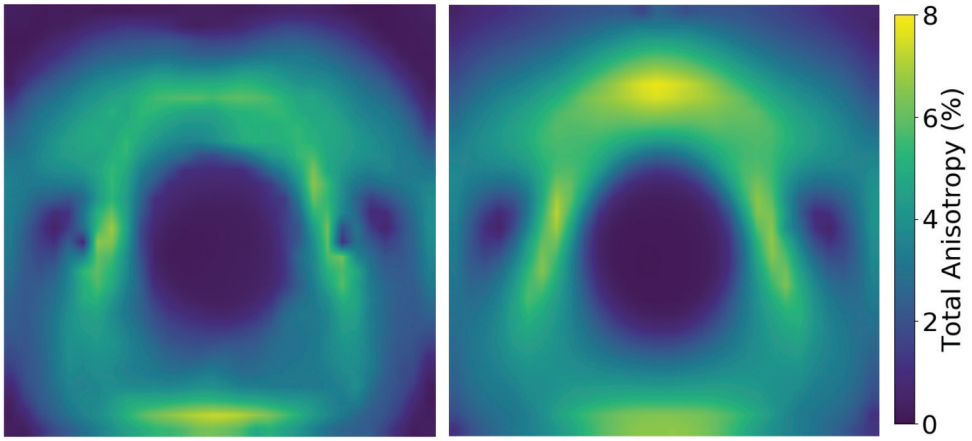

FIGURE 4.1: Vertical cross section of the percentages of total anisotropy obtained from: neural networks (left), and D-Rex (right). The total anisotropy is derived from the norm of the elastic tensor. The slices are oriented along the yz− plane, and taken at the center of the x− axis (*i.e.*, $x = 200$ km)

architecture, and the number of available training data used.

Predicting surface wave data

For any geographical location at the surface, we can extract the 1D velocity profile (*e.g.*, S_{ij} as a function of depth) and compute dispersion curves for Love and Rayleigh waves. The azimuthal dependence of surface wave phase velocity can be treated as the sum of a small anisotropic perturbation around an isotropic phase velocity model (Smith and Dahlen, 1973) giving:

$$\mathbf{c}(T,\theta) = \mathbf{c}_0(T) + \mathbf{c}_1(T)\cos(2\theta) + \mathbf{c}_2(T)\sin(2\theta) + \mathbf{c}_3(T)\cos(4\theta) + \mathbf{c}_4(T)\cos(4\theta)\,, \quad (4.8)$$

where T is the period and θ is the azimuthal angle.

In this work, we only invert $\mathbf{c}_0(T)$, $\mathbf{c}_1(T)$, and $\mathbf{c}_2(T)$ for Rayleigh waves and only $\mathbf{c}_0(T)$ for Love waves. It is not common to to invert other terms, due to low sensitivity or to high levels of noise. For convenience, we denote isotropic Rayleigh wave phase velocity as $\mathbf{c}_R(T)$ and Love wave phase velocity as $\mathbf{c}_L(T)$.

The different terms in eq. (4.8) can be computed from S_{ij} in a fully non-linear fashion by normal mode summation with a Runge–Kutta matrix integration (Takeuchi and Saito, 1972). We refer the reader to Montagner and Nataf (1986) and Bodin et al. (2016) for details. The seismic forward model is computed using a 1D earth assumption beneath each geographical location. We acknowledge that surface waves velocities depend on 3-D heterogeneities, and particularly the fact that surface wave computations exhibit non-linearities due to mode-coupling and finite frequency effects (e.g.,

Sieminski et al., 2007; Ekström, 2011). However, these approximations can be treated as theoretical errors and can be accounted for in the Bayesian inversion procedure.

4.3.3 Bayesian sampling scheme

We formulate the problem in a fully non-linear Bayesian framework (Box and Tiao, 2011; Smith, 1991; Mosegaard and Tarantola, 1995), where the predicted surface wave dispersion curves estimated for a large ensemble of models (3-D temperature fields) are compared to observed data. The solution of the inverse problem is the posterior distribution $p(\mathbf{m}|\mathbf{d})$, the probability model of parameters $\mathbf{m}$ given the data $\mathbf{d}$. According to Bayes' theorem we have:

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{m})\, p(\mathbf{d}|\mathbf{m}) \,. \tag{4.9}$$

The prior distribution $p(\mathbf{m})$ describes our predetermined knowledge on $\mathbf{m}$ (*i.e.*, the position and the amplitude of thermal anomalies, as well as the activation energy). The likelihood function $p(\mathbf{d}|\mathbf{m})$ describes the probability of observing the data given our current knowledge of the model parameters.

Since our forward problem is highly non-linear, the posterior distribution is sampled using a Markov chain Monte Carlo algorithm. It involves direct sampling of the parameter space by random iterative search, where the distribution of the sampled models asymptotically converges towards the posterior distribution.

The likelihood function

The likelihood function $p(\mathbf{d}|\mathbf{m})$ quantifies how well the model parameters explain the observed data (*i.e.* the ensemble of local dispersion curves located at the surface). Supposing that each data-type (*i.e.* $\mathbf{c}_R$ and $\mathbf{c}_L$ for isotropic Rayleigh and Love wave dispersion curves, respectively; $\mathbf{c}_1$ and $\mathbf{c}_2$ for Rayleigh wave anisotropy) is measured independently, the likelihood function gives:

$$p(\mathbf{d}|\mathbf{m}) = p(\mathbf{c_R}|\mathbf{m})\, p(\mathbf{c}_L|\mathbf{m})\, p(\mathbf{c}_1|\mathbf{m})\, p(\mathbf{c}_2|\mathbf{m}) \,. \tag{4.10}$$

For all dispersion curves, we assume that the errors are uncorrelated and follow Gaussian distributions with zero mean, and variances $\sigma^2_{c_R}$, $\sigma^2_{c_L}$, $\sigma^2_{c_1}$, and $\sigma^2_{c_2}$. For isotropic Rayleigh waves and isotropic Love waves $\mathbf{c}_R$ and $\mathbf{c}_L$, respectively, we can express the

likelihood function as a Gaussian distribution:

$$p(\mathbf{c}_{R,L}|\mathbf{m}) = \frac{1}{(2\pi\sigma^2_{c_{R,L}})^{\frac{N}{2}}} \exp\left[\frac{-||\,\mathbf{c}^{\text{obs}}_{R,L} - \mathbf{c}_{R,L}(\mathbf{m})\,||^2}{2\sigma^2_{c_{R,L}}}\right]. \tag{4.11}$$

Here, the likelihood function corresponds to a single dispersion measurement where N is the number of discrete periods. The likelihood functions of the 2θ terms, $\mathbf{c}_1$ and $\mathbf{c}_2$, can be written in the same manner as eq. (4.11).

A maximum likelihood estimate of data errors

In general, it is difficult to estimate $\sigma_{c_{R,L}}$ due to the lack of knowledge on the error distribution. In particular, approximating an elastic tensor with a neural network may introduce errors that are difficult to quantify.

In this work, we use a maximum likelihood estimate (MLE) of the noise parameters $\sigma_{c_{R,L}}$ and $\sigma_{c_{1,2}}$ following the work of Dettmer, Dosso, and Holland (2007). This is performed by maximising the likelihood function over the data standard deviation. The strength of this technique is that it is not necessary to estimate each contribution to the noise parameters individually. Maximising eq. (4.11) over $\sigma_{c_{R,L}}$ yields:

$$\sigma_{c_{R,L}} = \left[\frac{1}{N}\sum_{i=1}^{N}(\mathbf{c}^{\text{obs}}_{R,L} - \mathbf{c}_{R,L}(\mathbf{m}))^2\right]^{1/2}. \tag{4.12}$$

Substituting eq. (4.12) onto eq. (4.11), and taking the log likelihood we obtain:

$$\ln[p(\mathbf{c}_{R,L}|\mathbf{m})] = -\frac{N}{2}\ln\left[\sum_{i=1}^{N}(\mathbf{c}^{\text{obs}}_{R,L} - \mathbf{c}_{R,L}(\mathbf{m}))^2\right]. \tag{4.13}$$

The log likelihood functions of c_1 and c_2 can be defined using the same procedure. This method has two advantages: (1) the absolute value of errors need not be defined, and (2) in the case of joint inversion, we do not have to define the relative weights between each data type. Finally, the full log likelihood function gives:

$$\ln[p(\mathbf{d}|\mathbf{m})] = \ln[p(\mathbf{c}_R|\mathbf{m})] + \ln[p(\mathbf{c}_L|\mathbf{m})] + \ln[p(\mathbf{c}_1|\mathbf{m})] + \ln[p(\mathbf{c}_2|\mathbf{m})]\,. \tag{4.14}$$

The prior distribution

In Bayesian inference, one expresses the *a priori* information in terms of a probability distribution $p(\mathbf{m})$. In geophysical inverse problems, model parameters are typically given a uniform prior distribution with given upper and lower bounds inferred from

prior knowledge (Mosegaard and Sambridge, 2002). Adopting the same formulation, the prior can be written as:

$$p(m_i) = \begin{cases} 0 & m_i > m_{max},\ m_i < m_{min} \\ \frac{1}{\Delta m} & m_{min} \leq m_i \leq m_{max}\ , \end{cases} \tag{4.15}$$

where m_{max}, and m_{min} are the prior bounds for the model. Assuming that the model parameters in our inversion are prior independent, we can express the prior fully as:

$$p(\mathbf{m}) = p(E) \prod_{i=1}^{M} \left[p(x_0^i)\, p(y_0^i)\, p(z_0^i)\, p(R^i)\, p(T_c^i) \right], \tag{4.16}$$

where $p(E)$ is the prior distribution for the activation energy, and M is the total number of spherical temperature anomalies. For an ith temperature anomaly, $p(x_0^i)$, $p(y_0^i)$ and $p(z_0^i)$ are the prior distributions for position; $p(R^i)$ and $p(T_c^i)$ are the prior distributions for the size and temperature, respectively. We choose wide uniform prior distributions. For the prior bounds, we select: (1) the length of the spatial domain (0 km to 400 km) for the positions x_0, y_0, and z_0, (2) 40 km to 240 km for R, (3) 500 K to 1200 K for T_c, and (3) 6 to 12 for E. Choosing wide bounds ensures that the model parameters are loosely constrained from the prior, and more emphasis is given to the information provided by the data.

A random walk to sample the posterior distribution

We use a Markov chain Monte Carlo algorithm to sample the posterior distribution. It begins by randomly selecting an initial temperature model followed by the evaluation of the initial log likelihood. At each iteration, the current model is perturbed to propose a new model. The proposal proceeds sequentially as follows:

1. **Assign local perturbation**: One sphere is randomly picked out of M number of spheres. Once a sphere is picked, we randomly select one of four possible ways to perturb the sphere are as follows:

   - Perturb horizontal position; *i.e.* x_0 and y_0 together;
   - Perturb vertical position z_0;
   - Perturb the size of the sphere R;
   - Perturb the temperature of the sphere T_c.

   Each perturbation is drawn from a univariate normal distribution centered at the current value of the model parameter.

TABLE 4.2: True model parameters defining the synthetic temperature field.

Model parameter	Assigned value
x_0	200 km
y_0	200 km
z_0	200 km
R	120 km
T_c	800 K
E	11.0

2. **Perturb the activation energy**: We then apply eq. (4.1) to define the 3-D temperature field. Alongside, we perturb the activation energy E by using a normal distribution centered at the current value of E, and apply eq. (4.4) to define the 3-D viscosity field. These two scalar fields are used as inputs in the flow calculation.

If the proposed model lies within the prior bounds following eq. (4.16), we evaluate the forward problem completely. The computed dispersion curves from the latter are compared with the observed data using eq. (4.13). The resulting likelihood is then compared to the likelihood of the current model, and the proposed model is either accepted or rejected according to an acceptance probability (Metropolis et al., 1953; Hastings, 1970). If the proposed model is accepted, it becomes the current model for the next iteration. After a sufficient number of iterations, the ensemble of accepted models converges towards the posterior distribution.

4.4 Application with 3-D Synthetic Temperature Fields

4.4.1 Inversion for One Spherical Anomaly

We demonstrate our proof of concept by setting up a simple temperature field consisting of one spherical negative temperature anomaly (*i.e.* negatively buoyant) placed at the middle of a 400 km $\times$ 400 km $\times$ 400 km box. The setup is a very simple toy example inspired by the work of Baumann, Kaus, and Popov (2014) where they applied Bayesian inversion to constrain rheology from gravity anomalies and surface velocities.

Table 4.2 shows the complete list of true model parameters, and Fig. 4.2 displays a cross-sectional view of the temperature field, and its associated instantaneous velocity field.

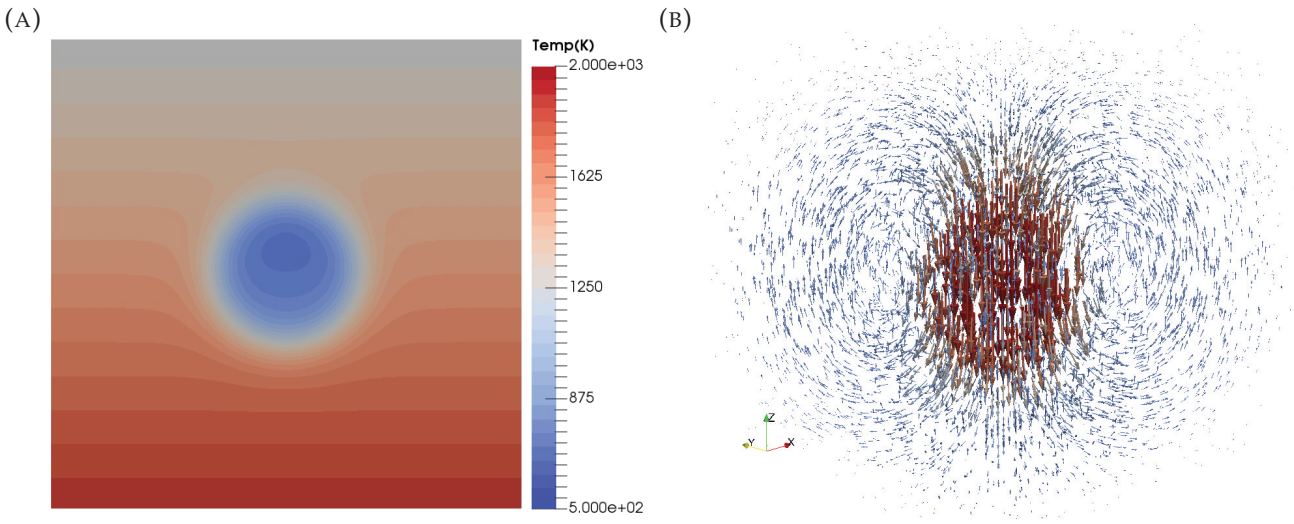

FIGURE 4.2: (a) Cross-sectional view in the yz− plane of the 3-D temperature field. The slice is taken at the center of the x− axis. (b) 3-D flow velocity due to the sinking anomaly. Largest flow magnitudes correspond to the cold anomaly.

We simulate the full forward model given the true model parameters to generate synthetic dispersion curves at periods between 10s and 200s. Fig. 4.3 shows a map of the computed phase velocity and azimuthal anisotropy for Rayleigh waves at 100 seconds. In Fig. 4.3a, the phase velocity is maximum at the middle of the region, due to the presence of the cold anomaly underneath.

Fig. 4.3b shows a map of azimuthal anisotropy in Rayleigh waves. Here, anisotropy is at its minimum at the center, above where the cold more viscous anomaly is located. As a result of this higher rigidity, local velocity gradients are lower, resulting in smaller amounts of deformation and hence lower anisotropy. Another feature is the presence of strong anisotropy at certain locations. These regions are points where shear deformation is at its maximum due to the convergence of flow lines. On top of the level of azimuthal anisotropy is the orientation of its fast axis. Since we expect the flow direction to converge towards the center when observed from the top, the fast axis may be interpreted as the horizontal projection of the flow.

The complete data constitute a regular array of 8×8 locations containing $\mathbf{c}_R$, $\mathbf{c}_L$, $\mathbf{c}_1$, and $\mathbf{c}_2$ spanning the entire surface. We emphasise that the data generated comes from an elastic tensor computed with `D-Rex` whereas during inversion, the estimated data is obtained from an elastic tensor approximated by neural networks.

Finally, we added random uncorrelated noise onto $\mathbf{c}_R$, $\mathbf{c}_L$, $\mathbf{c}_1$, and $\mathbf{c}_2$. Standard deviations for Love and Rayleigh are set at $\sigma_R = 0.05$ km/s and $\sigma_L = 0.05$ km/s, whereas $\sigma_{c_1} = 0.01$ km/s and $\sigma_{c_2} = 0.01$ km/s. Fig. 4.4 illustrates the resulting dispersion curves

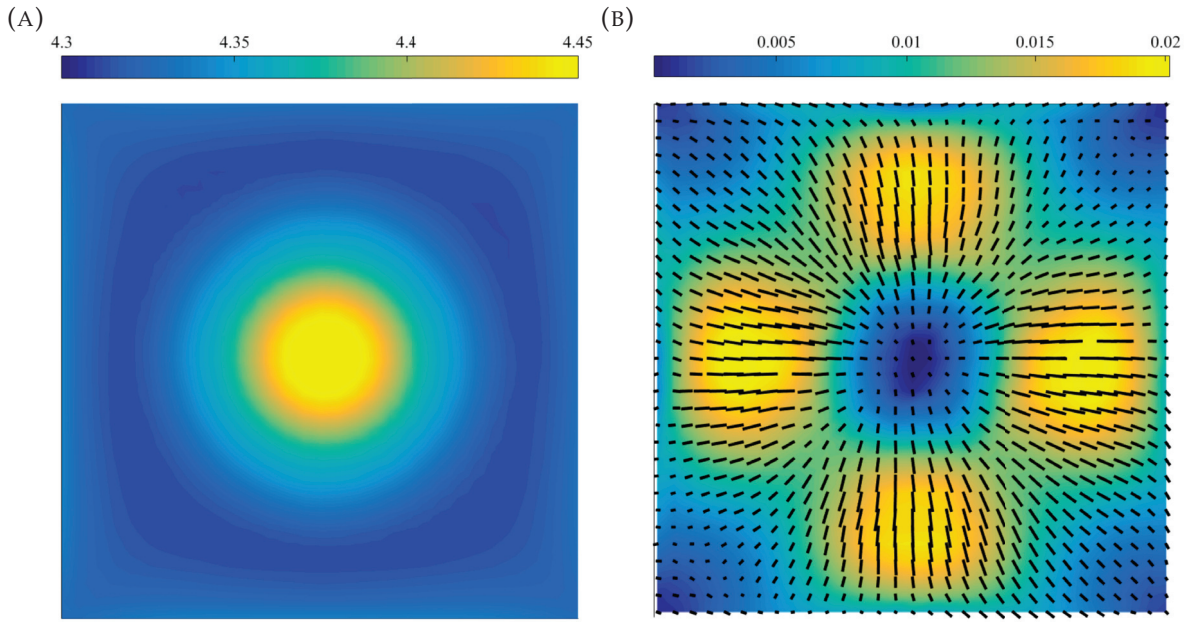

FIGURE 4.3: Phase velocity maps resulting from one sinking anomaly at 100s period. (a) Rayleigh wave phase velocity (km/s). (b) Azimuthal anisotropy in Rayleigh waves (km/s). The solid black lines correspond to the direction of the fast propagation axis. Surface wave maps always lie along the xy− lateral plane.

at one given location with and without noise.

The inversion consists of 20 independent Markov chains each containing 40,000 samples initiated at a random temperature structure. We demonstrate two cases. First is an isotropic inversion, where no anisotropy is involved in the forward model. In this case, it is not necessary to compute instantaneous flow and anisotropy, as isotropic seismic velocities V_p and V_s can be directly scaled with temperature. The inverted data are the isotropic phase velocities $\mathbf{c}_R$ and $\mathbf{c}_L$. Secondly, we present an anisotropic inversion (geodynamic tomography). Both isotropic and anisotropic inversions are given the same wide uniform priors allowing for more mobility when searching the parameter space. We initiate geodynamic tomography by first employing an isotropic inversion. Once the chains have converged in this phase, we then start the actual anisotropic inversion procedure.

The diagonal panels of Figs 4.5 and 4.6 illustrate the ensemble of models recovered from isotropic inversion and anisotropic inversion. The off-diagonal panels depict 2D marginal distributions as 2D histograms to explore possible trade-offs. The black circles indicate the values of the true model parameters. Compared to isotropic inversion, the width of the posterior distribution inferred from geodynamic tomography has been reduced considerably. More information is thus added by introducing geodynamic constraints in the tomographic problem.

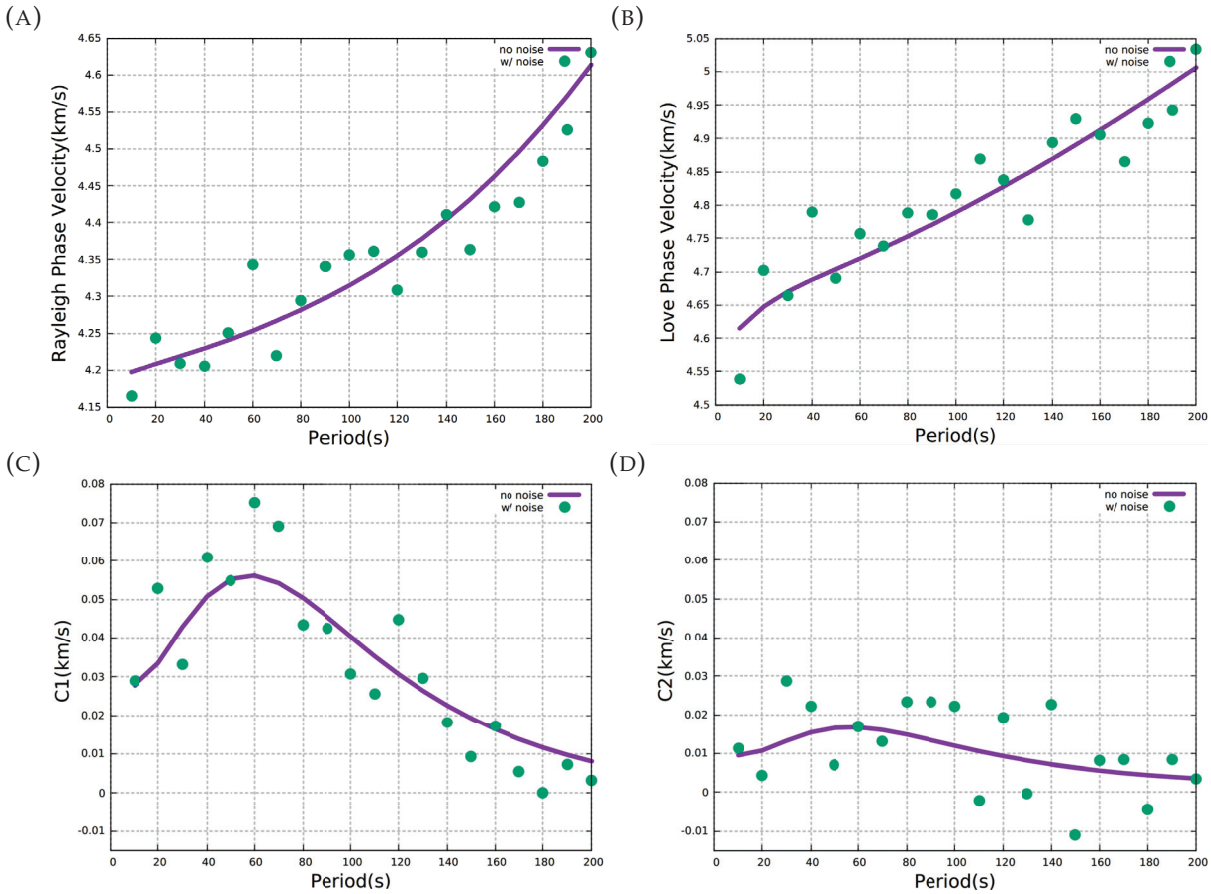

FIGURE 4.4: Synthetic surface wave dispersion curves from 10s to 200s at a given location: (a) Rayleigh wave phase velocity, (b) Love wave phase velocity, (c) Rayleigh anisotropy c_1, (d) Rayleigh anisotropy c_2. Scatter plot: observed dispersion curve with added noise. Line plot: observed dispersion curve without noise.

As expected, the posterior distribution on the activation energy E in the isotropic case is flat, as isotropic velocities are only sensitive to temperature and not to viscosity. Anisotropic inversion, on the other hand, constrains E as shown in Fig. 4.6. The distribution, however, appears to be distant from the correct value of E. Such a behavior is also evident in its 2D marginal posterior where the true value is outside the inferred distribution. This clearly exhibits a bias which is deduced from the imperfections of the neural network when computing anisotropy. This effect is eliminated when one uses the correct forward operator for modeling anisotropy. Another distinct feature in these figures is the negative trade-off between T_c and R, which may be attributed to the symmetry of the problem considered. An increase in temperature of the anomaly compensates for an increase in its radius. Such trade-offs may be reduced in the case where the true model exhibits less symmetry.

We also plot the mean temperature models from both inversions (see Fig. 4.7). The figures are obtained by averaging the temperature values at each point. By visual inspection, anisotropic inversion better resolves the 3-D thermal structure. This is further supported by the standard deviation computed around the mean temperature at a given pixel as shown in Figs 4.7c and 4.7d. In both cases, the standard deviations is higher at the center of the box, where the spherical anomaly is located. This is due to the variations in the location and amplitude of the sphere in the ensemble of sampled models. In the anisotropic case, the vertical position of the sphere is less constrained than its horizontal position, as can be seen in the 2D histograms. The ensemble of sampled spheres therefore share the same horizontal position but have a variable vertical position, which explains the shape of the standard deviation map in Fig. 4.7c. The posterior uncertainties are also relatively small compared to the recovered temperature field, implying that sufficient information can be retrieved from the noisy dispersion curves.

Fig. 4.8a shows the 1D depth marginal posterior probability profiles (see captions for further details) for temperature, and Fig. 4.8b for radial anisotropy ξ, peak-to-peak azimuthal anisotropy, and the trend of the fast-axis of azimuthal anisotropy at a given location. Both methods capture the 1D structures for temperature. However, by adding geodynamic constraints (*i.e.*, anisotropic inversion), we observe that the temperature is much better resolved. Additionally, we successfully recover radial anisotropy and azimuthal anisotropy without having to explicitly invert for the elastic tensor (see Fig. 4.8b). Here, due to the positioning of the chosen depth profile for temperature (passing nearly through the center of the anomaly), the azimuthal anisotropy appears to be nonexistent at this location. For that reason, we consider another depth profile ($x = 325$ km, $y = 225$ km) where azimuthal anisotropy is noticeable

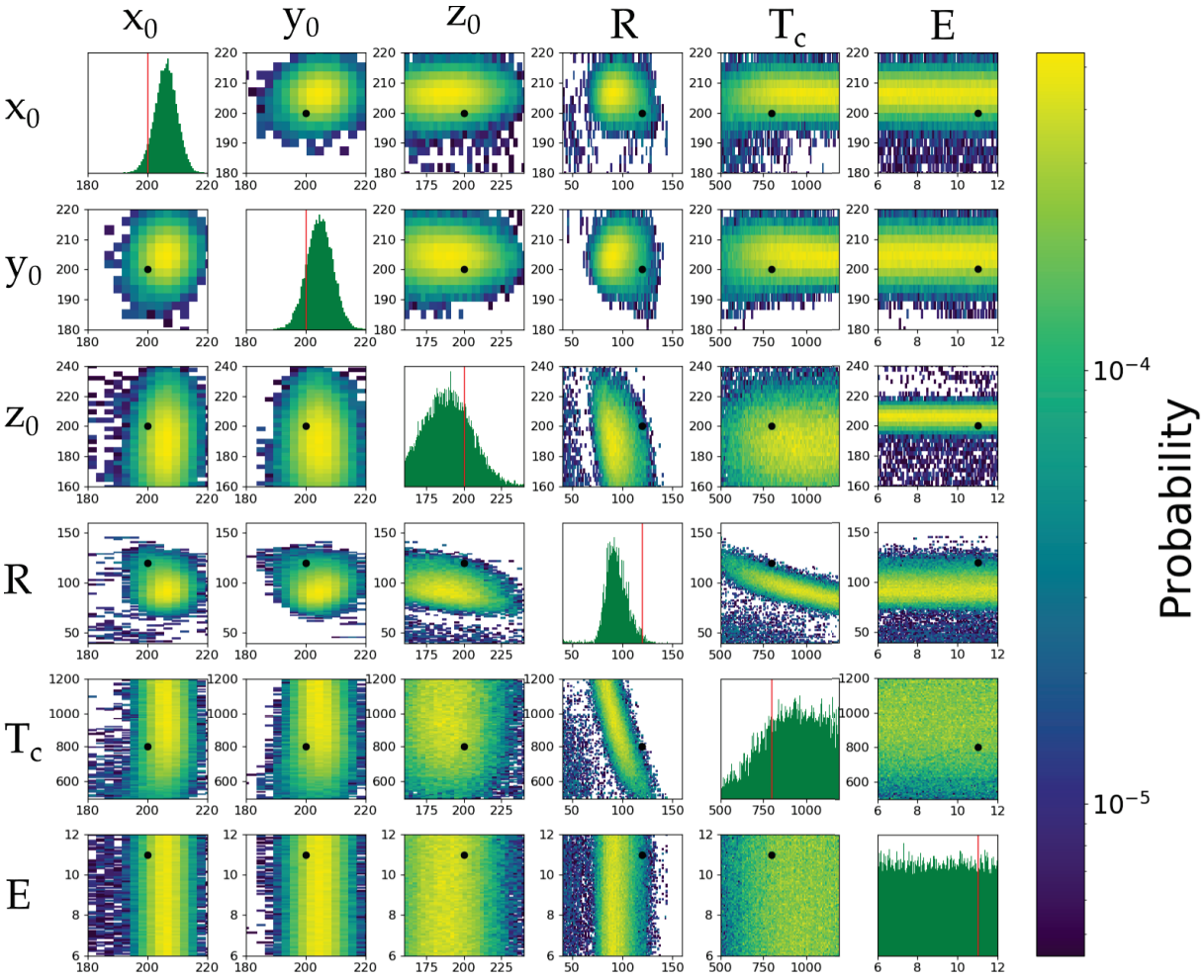

FIGURE 4.5: Posterior probability distribution in the 6-dimensional parameter space inferred from the isotropic inversion $p(\mathbf{m}|\mathbf{c}_R, \mathbf{c}_L)$. Diagonal panels show 1D marginal distributions for each model parameter. Off-diagonal panels show 2D marginal distributions and depict possible trade-offs between pairs of model parameters. The red vertical lines and the black markers indicate the true model values for the diagonal and the off-diagonal panels, respectively. The intensity pertains to the level of posterior probability (*i.e.*, high intensity means high probability, and thus low misfit).

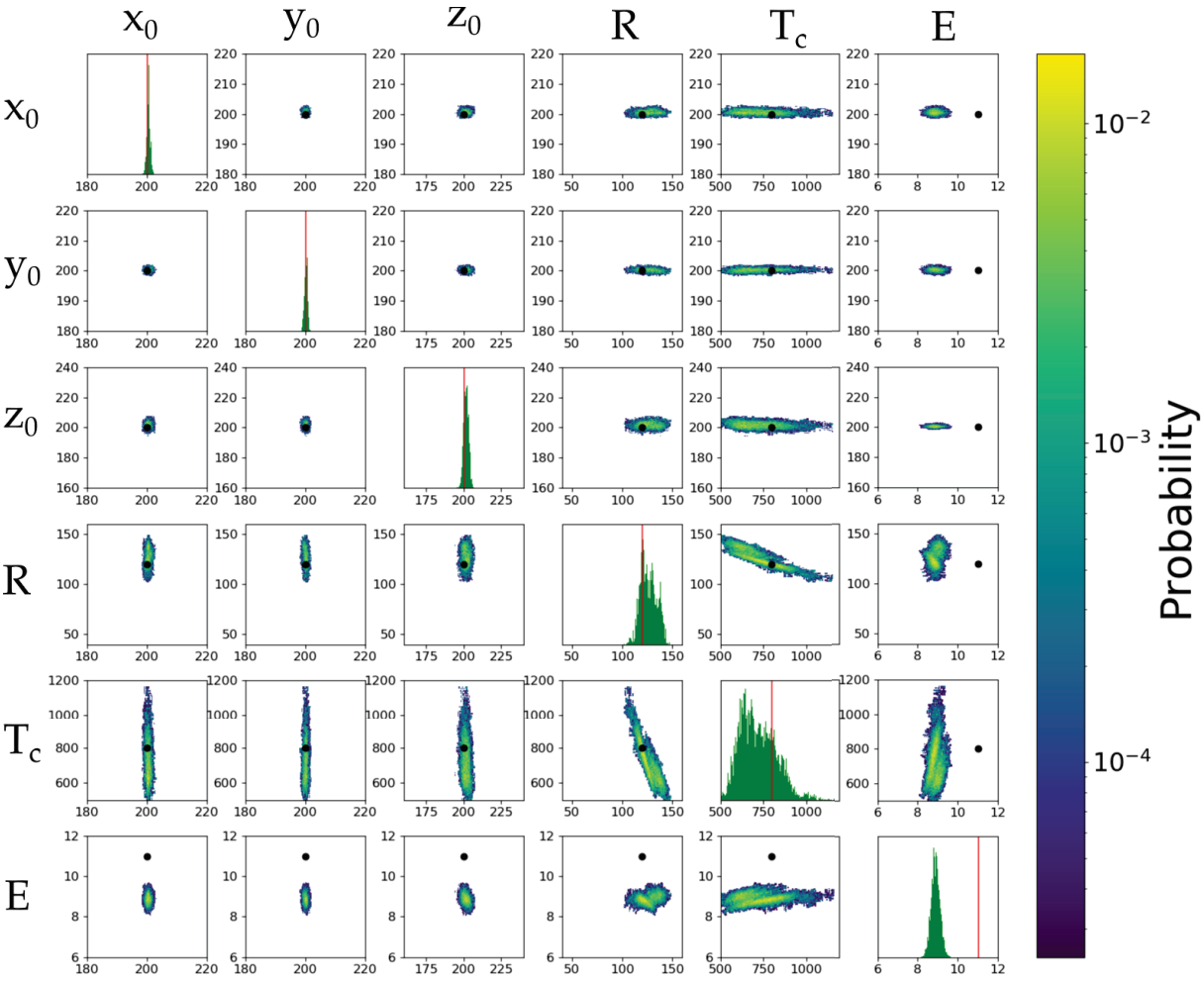

FIGURE 4.6: Posterior probability distribution in the 6-dimensional parameter space inferred from the anisotropic inversion $p(\mathbf{m}|\mathbf{c}_R, \mathbf{c}_L, \mathbf{c}_1, \mathbf{c}_2)$.

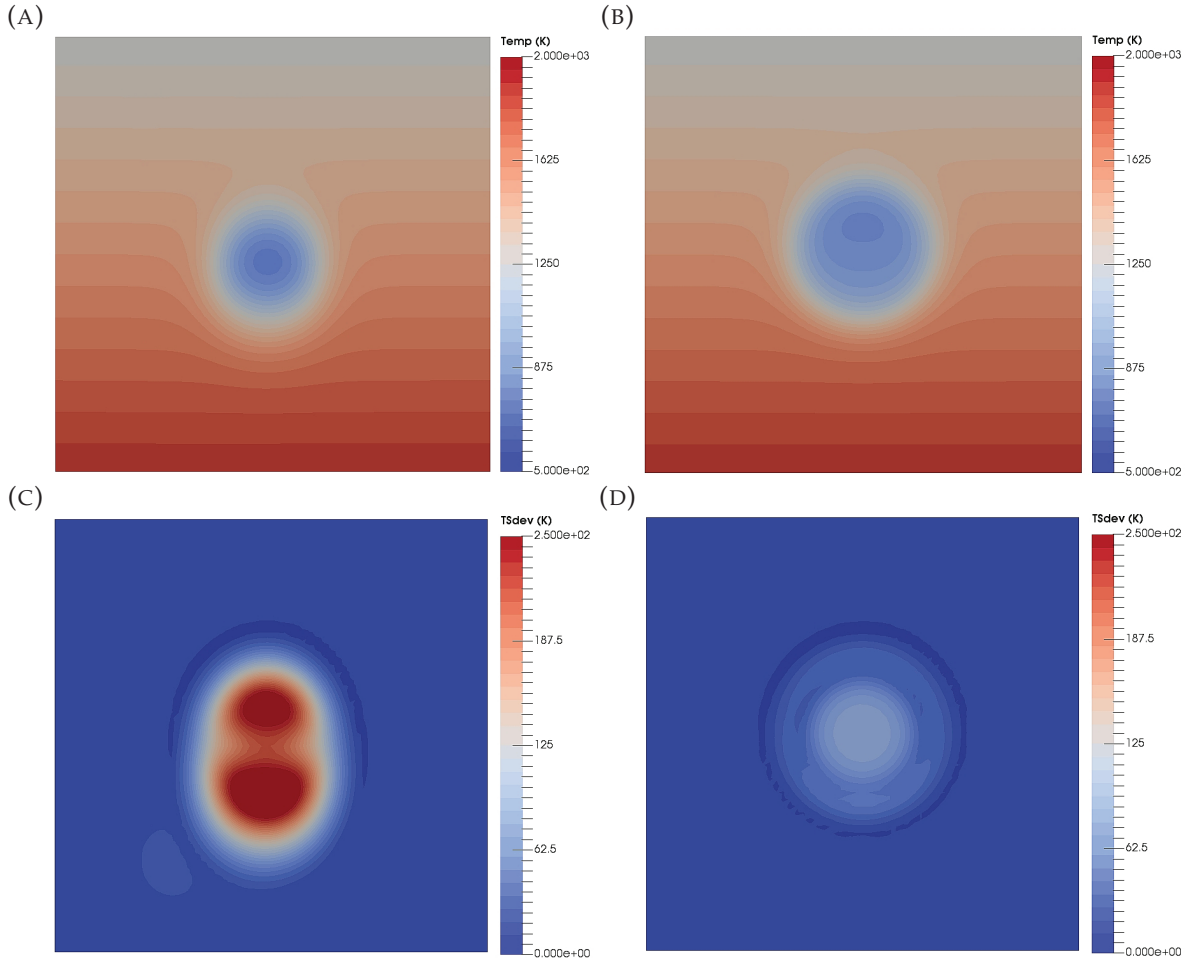

FIGURE 4.7: Upper panel: Cross-sectional view in the yz− plane of the mean temperature field recovered from (a) isotropic inversion, and (b) anisotropic inversion. Lower panel: Standard deviations around the mean temperature fields from (c) isotropic inversion, and (d) anisotropic inversion. These cross-sections are taken at the center of the x− axis.

(Fig. 4.8b−middle).

This method also allows us to resolve 3-D structures of seismic properties. In fact, any implicitly computed variable can be restructured in 3-D. Figs 4.9 and 4.10 show the resulting structures computed from the mean temperature model placed side by side with that of the true model. It appears that the value of anisotropy computed with the neural network is underestimated compared to that of `D-Rex` when using the same input model. This explains why the activation energy E resulting from the inversion is lower compared to the true value: to produce larger anisotropy and replicate the same output as obtained from `D-Rex`, one has to reduce the value of E. Indeed, reducing the viscosity of the material allows for a stronger deformation. The resulting percentage of total anisotropy from both figures are nearly identical. Fig. 4.9 shows the presence of positive radial anisotropy at the bottom, indicating horizontal flow. Due to the imposition of free-slip boundary conditions combined with zero normal velocities imposed on all surfaces, the flow at the bottom of the box is oriented nearly horizontally. The negative radial anisotropy we observe implies vertical flow (see Fig. 4.9 caption for details). This is a result of convection cells forming at the sides of the anomaly as it sinks. At the top of the anomaly, negative radial anisotropy also indicates vertical flow due to downwelling. Finally and as we expect, radial anisotropy at the middle is nearly unity due to the presence of the more viscous anomaly. The difference in the structures may be attributed to the following: (1) imperfections of the forward model used in the inversion; (2) information loss related to data sensitivity, and data noise.

We tested the convergence of the Markov chain by plotting the estimates for data errors with MC steps. For further details, refer to Appendix 4.6.2.

4.4.2 Inversion for Multiple Spherical Anomalies

This section covers the inversion for ten spherical temperature anomalies with different properties (*i.e.*, temperature T_c and radius R), positioned randomly in 3-D space. Such parameterisation scheme may be essential to represent anomalies with complex shapes (*e.g.*, subducting slab) using a collection of several spheres with different characteristics. The synthetic data is generated from a true temperature model consisting of ten spherical anomalies as well. We compare the true temperature model with the mean temperature models obtained from isotropic and anisotropic inversions (Fig. 4.11). Even with this much more complex structure, we are able to recover the main features of the temperature field. Also, as in the test of Section 4.4, anisotropic inversion better recovers the structure than isotropic inversion. Posterior uncertainties are represented in Fig. 4.12 and support this observation. However, some differences with the exact

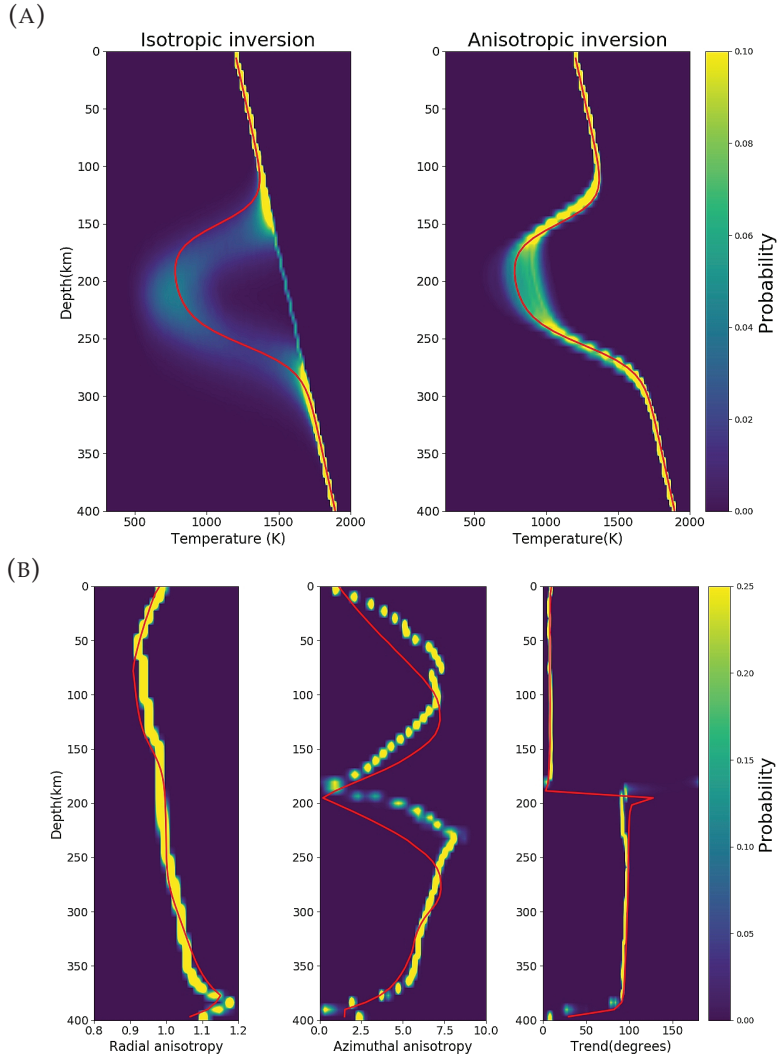

FIGURE 4.8: Upper panel: Probability density plots of temperature with depth. Lower panel: Probability density plots of radial anisotropy, peak-to-peak azimuthal anisotropy, and its fast axis with depth. The depth profiles of temperature and radial anisotropy are taken nearly through the center of the sphere. To show that azimuthal anisotropy is also well-constrained, we took a depth profile at ($x = 325$ km, $y = 225$ km), where azimuthal anisotropy is large. Geodynamic tomography offers the capability to constrain seismic anisotropy. The solid red lines indicate the true structures.

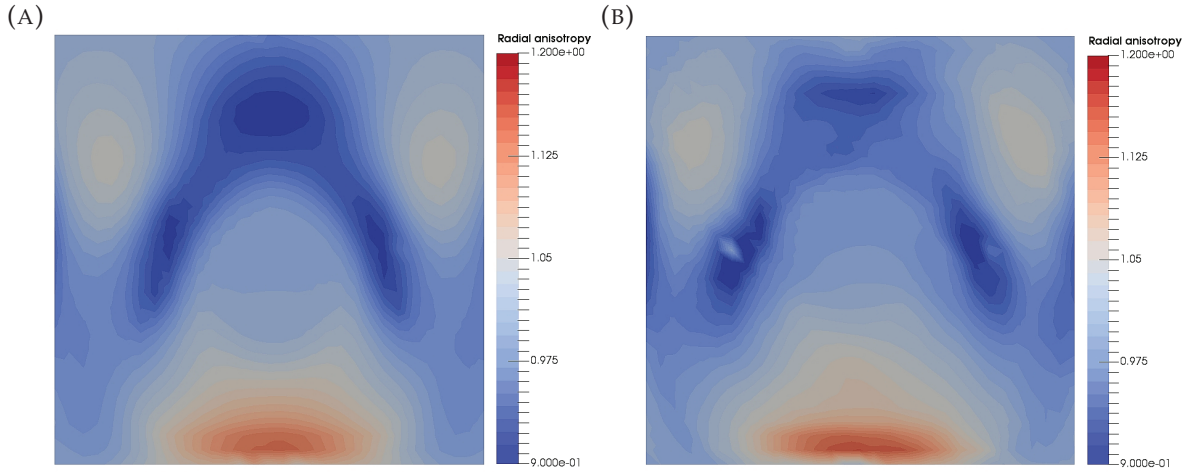

FIGURE 4.9: Cross-sectional view in the yz− plane of the radial anisotropy ξ inferred from (a) true model, (b) mean model. Radial anisotropy is often used as a proxy to infer flow orientation. A $\xi > 1$ (positive radial anisotropy) is often interpreted as horizontal flow. A $\xi < 1$ (negative radial anisotropy) on the other hand, pertains to vertical flow. A $\xi = 1$ indicates the absence of radial anisotropy. The cross sections are taken at the center of the x− axis.

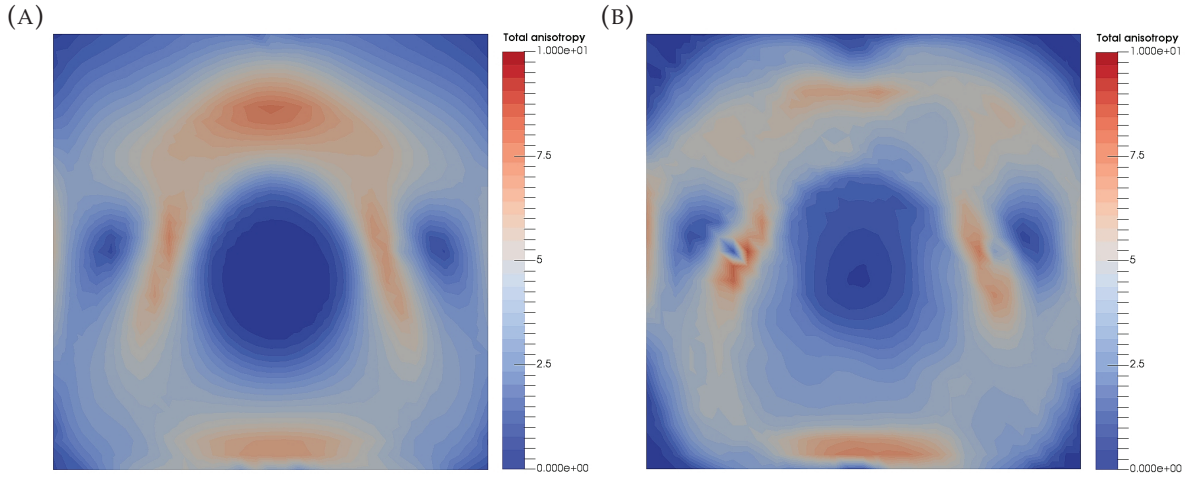

FIGURE 4.10: Cross-sectional view in the yz− plane of the percentage of total anisotropy(*i.e.*, norm of S_{ij}) inferred from (a) true model, (b) mean model. The absence of anisotropy at the center corresponds to a region of minimal deformation for the cold and highly-viscous anomaly. The cross sections are taken at the center of the x− axis.

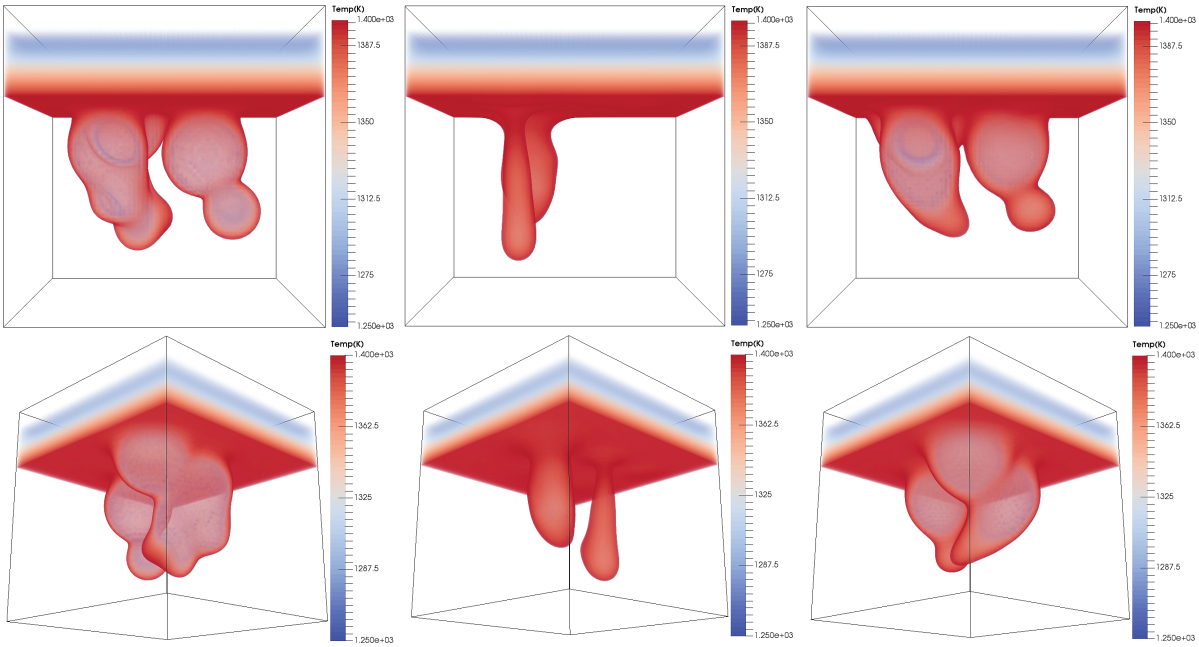

FIGURE 4.11: Isovolumetric view of the temperature fields. Left: True temperature field. Middle: Mean temperature field from isotropic inversion. Right: Mean temperature field from anisotropic inversion.

true structure remain, even using anisotropic inversion. Surface waves are long period observations and hence, small and sharp thermal anomalies may not be resolved. Other contributing factors involve the very nature of the tomographic problem itself as enumerated earlier (*e.g.*, data and modeling errors).

In Fig. 4.13, we choose one depth profile to show the 1D marginal posterior probability densities for temperature, radial anisotropy, and azimuthal anisotropy. The dashed black lines represent the true model. Based on the recovered profiles, anisotropic inversion resolves temperature better than the isotropic case again due to the complementing information brought by geodynamic constraints. Radial and azimuthal anisotropy still appears to be tightly constrained; however with some noticeable deviations from the true model.

4.5 Conclusion

We have laid the groundwork for geodynamic tomography, a novel approach that involves constraints from geodynamic modeling to invert seismic surface waves. Imposing these geodynamic constraints reduces the number of model parameters to a single scalar field (*i.e.*, temperature) and one scalar variable (*i.e.*, activation energy for viscosity). The inverse problem is cast using Bayesian inference where we directly sample the model space using Markov chain Monte Carlo algorithm. Here, instantaneous flow,

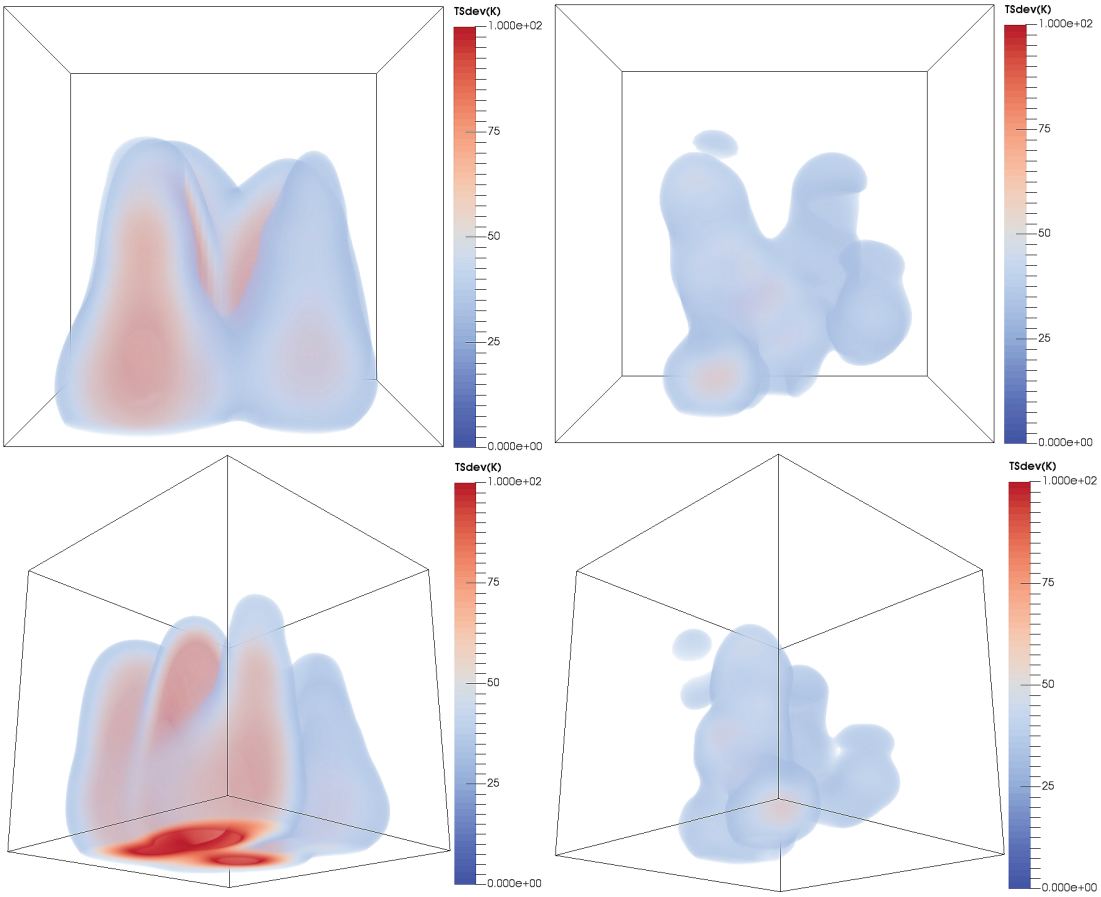

FIGURE 4.12: Isovolumetric view of the standard deviations around the mean temperature models. Left: Standard deviation for the isotropic inversion. Right: Standard deviation for the anisotropic inversion.

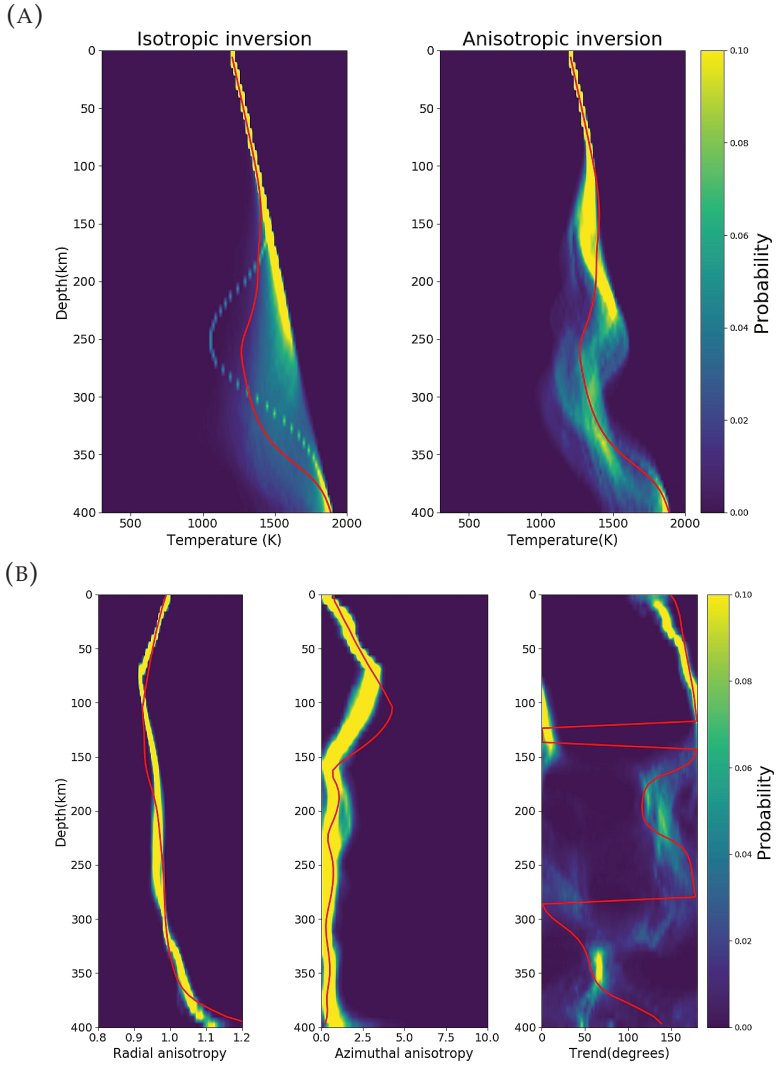

FIGURE 4.13: Upper panel: Comparison between isotropic and anisotropic inversion. Probability density plots of temperature with depth. The profiles are taken nearly through the center of the sphere. Lower panel: Anisotropic inverison: probability density plots of radial anisotropy, peak-to-peak azimuthal anisotropy, and its fast axis with depth. All profiles correspond to the temperature profile above. The solid red lines indicate the true structures.

deformation history, and finally seismic anisotropy are computed in our forward problem. The model space is reduced further by parameterising the temperature field as a sum of spherical temperature anomalies with variable position, size, and temperature.

We tested geodynamic tomography in simple cases, where we successfully recovered synthetic 3-D temperature fields, by jointly inverting fundamental mode anisotropic Rayleigh wave and isotropic Love wave phase velocities. In the process, we are also able to constrain the complete deformation pattern, to provide a quantitative interpretation of seismic anisotropy in the mantle. Given the Bayesian formulation, one may express the ensemble of temperature models, and any implicitly computed variables (such as deformation or anisotropy) as posterior probability distributions, and quantify their associated uncertainties. Geodynamic tomography is therefore a potentially powerful technique to study the structure of the upper mantle, and interpret seismic observations in terms of mantle deformation patterns.

4.6 Appendix

4.6.1 Parameterising temperature with spherical anomalies

For a given anomaly, we define a basis function corresponding to that anomaly using eq. (4.3). The negative sign indicates that the anomaly is colder than the background temperature if T_c is positive (a negatively buoyant anomaly). Should T_c be negative, then the anomaly adds up with the background temperature resulting to a positively buoyant anomaly. The function is designed such that: (1) When $\left\| \mathbf{r} - \frac{\mathbf{r_0}}{L_s} \right\| > \frac{R}{L_s}$ and tanh returns a value of nearly one, then the temperature is just the background temperature. (2) When $\left\| \mathbf{r} - \frac{\mathbf{r_0}}{L_s} \right\| = \frac{R}{L_s}$, then the temperature at just half of the radius of the anomaly is equal to $T_{\text{background}} - \frac{T_c}{2}$. (3) Finally, when $\left\| \mathbf{r} - \frac{\mathbf{r_0}}{L_s} \right\| < \frac{R}{L_s}$ and tanh returns a value of minus one, this corresponds to the temperature at the center of the anomaly $T_{\text{background}} - T_c$.

Here, β controls the sharpness of the temperature gradient and is held at a fixed value. Choosing a very large value for β results in a sharp temperature gradient (see Fig. 4.14). In addition, opting for a smooth function such as hyperbolic tangent avoids very sharp viscosity contrasts when computing for the flow. The advantage of building a basis set is to reduce the number of model parameters. In conventional inversion schemes of scalar fields, we usually invert for a scalar at a given grid point. Hence, the number of model parameters depends on the grid size. In a cube, this would result to N^3 model parameters to constrain, where N^3 is the size of the 3-D block. In our

case, this gives us $5M$ parameters to be inverted, where M is the number of spherical anomalies. Finally, we define the 3-D scalar temperature field as the sum of the background temperature and the spherical anomalies as shown in eq. (4.1).

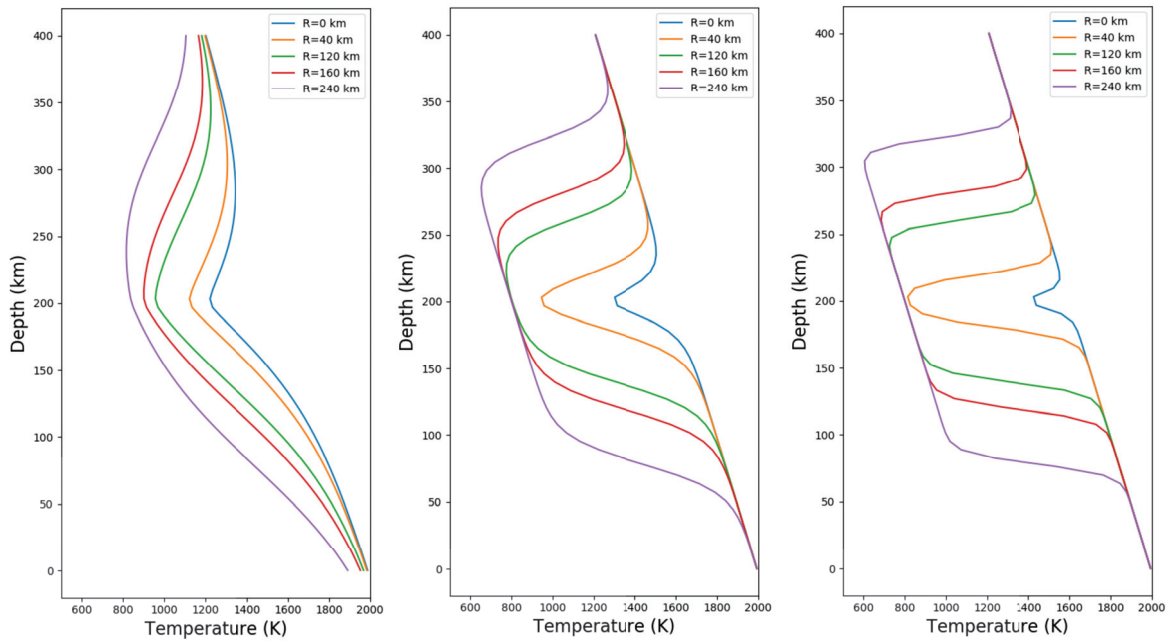

FIGURE 4.14: 1D temperature profiles with depth for different values of R and β. Left: $\beta = 5$. Middle: $\beta = 20$. Right: $\beta = 50$. Here, we consider a spherical anomaly with $T_c = 800$ K located at the center of the 3-D volume. The plots refer to 1D depth profiles of temperature through the middle of the sphere at specified values of R and β. The x and y axes correspond to temperature and depth, respectively. Based on our parameterisation, increasing the value of R at constant β increases the size of the temperature anomaly. At constant R, the anomalies retain their respective sizes but the temperature gradient becomes sharper at increasing β. Thus, choosing an appropriate β is important so as to avoid sharp viscosity contrasts (since η depends on T) when computing flow. In our inversion, we choose to fix $\beta = 20$, and invert for R.

4.6.2 A simple test for convergence

Fig. 4.15 shows the noise estimate plotted against MC step in the one sphere case. The standard deviation of data noise is implicitly computed with MLE (see Section 4.3.3), and is simply given by the level of data fit. The starting point for each plot is the iteration at which anisotropic tomography commences. The trends exhibit well-mixed random walk behaviors indicating that convergence has been achieved. This level of noise estimated by MLE represents the combination of observational errors (white

noise added to the data), and theoretical errors (errors of the surrogate model used for texture evolution).

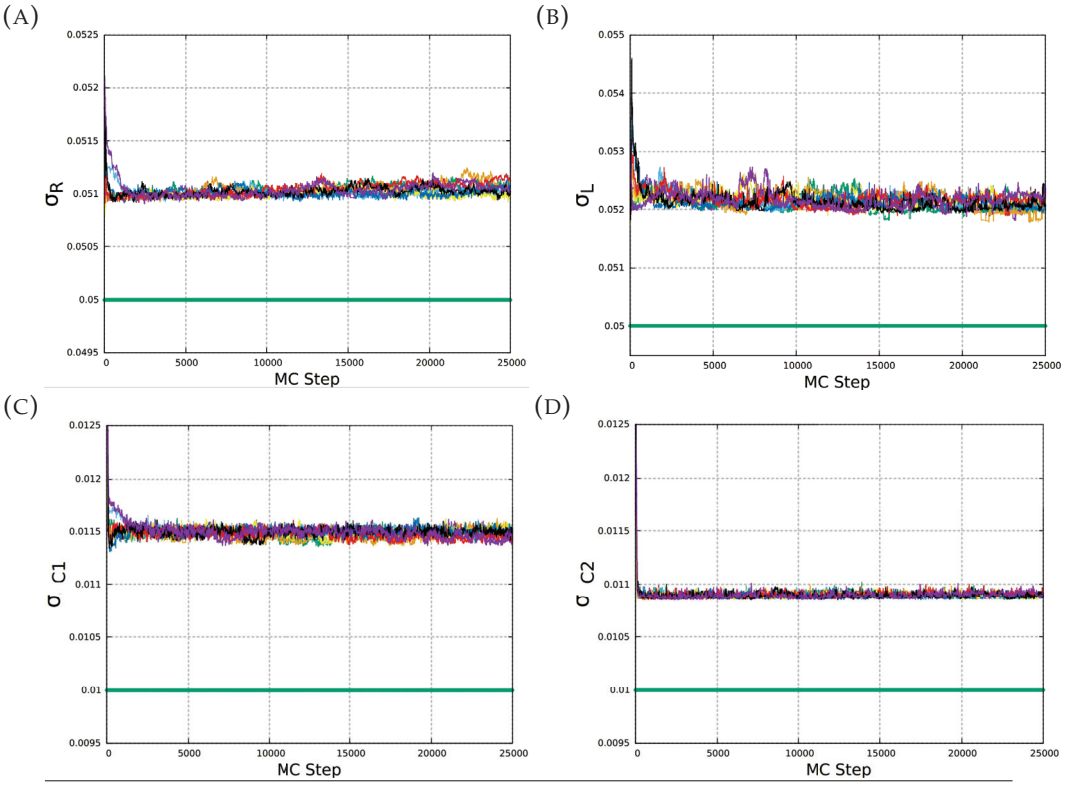

FIGURE 4.15: Noise estimate with MC step for (a) Rayleigh waves, (b) Love waves, (c) c_1, and (d)c_2. Each colored line plot is associated with one independent Markov chain. Solid green line indicates the standard deviation of random errors added to the data.

Chapter 5

Geodynamic Tomography: Application to a 3D Deforming Upper Mantle Beneath a Subduction Zone

5.1 Summary

The previous chapter highlights the success of geodynamic tomography in the recovery of the complete patterns of upper mantle deformation from anisotropic surface wave dispersion measurements in the most simple cases (*i.e.*, classical Stokes sphere-like geometry problem). This chapter explores the ability of the method to capture more complex deformation patterns in the guise of a 3-D deforming upper mantle induced by subduction. Thus this demonstration can be viewed as an intermedial step prior to its full realisation towards real Earth problems.

A myriad of model parameterisations can be implemented. One notable procedure is to represent regional tectonics as geometrical blocks regarded initially as *a priori* constraints (*i.e.*, our preconceived knowledge about the geometrical structure of the region of interest). Constant values of relevant parameters defining its thermal and rheological properties can then be ascribed in each of these blocks. For instance in the case of a subduction geometry, the subducting slab can be assigned with a constant temperature and viscosity that is inherently distinct from the values assigned for the sub-slab mantle and the mantle wedge.

This chapter covers an alternative parameterisation where in lieu of assigning constant values, we apply a continuous parameterisation of the temperature field in terms of hyperbolic tangent basis functions. This implies that the block of subducting slab can be treated as a smooth temperature anomaly defined by this basis function on top of an adiabatic temperature gradient. The basis function depends on four parameters which would be treated as unknowns: (1) slab length L, (2) slab thickness R, (3) dip

angle θ, and (3) slab temperature T_c. For the medium rheology, we still employed temperature-dependent viscosity controlled by a single activation energy E; although it is acknowledged that a realistic parameterisation involves the implementation of viscoplastic behavior with several control parameters other than temperature.

We implement geodynamic tomography to retrieve these five unknown parameters that define the thermal and rheological structure of a synthetic subduction zone. The method is tested to synthetic data prescribed with very low noise levels to mimic periodically-correlated surface wave dispersion measurements. Results show the incorporation of geodynamic and petrological constraints tightly recover these five unknowns, which implies the implicit retrieval of the complete patterns of upper mantle deformation, and correspondingly, the 21-independent coefficients defining elastic anisotropy. The final output is an ensemble of models of L, R, θ, T_c, and E cast in terms of a posterior probability distribution and their uncertainty limits.

5.2 Model parameterisation: thermal structure of a subduction zone

We begin by defining the temperature field T as the sum of a background temperature $T_{\text{background}}$ derived from a half-space cooling model and a thermal anomaly δT. This translates to:

$$T(\mathbf{r}) = T_{\text{background}}(\mathbf{r}) + \delta T(\mathbf{r}), \tag{5.1}$$

where $T_{\text{background}}$ is given by:

$$T_{\text{background}}(\mathbf{r}) = (1900\text{K} - 500\text{K}) \operatorname{erf}\left(\frac{z}{2\sqrt{\kappa t}}\right) + 500\text{K}, \tag{5.2}$$

where $\mathbf{r}$ is any arbitrary position in 3D space defined by the coordinates $\mathbf{r} = [x,y,z]$, κ is the thermal diffusivity, z is depth, and t is the plate age in million years. The anomaly δT is a subducting slab defined by three geometrical parameters: (1) dip angle θ, (2) length of the slab L, (3) thickness R which relates to the rate of heat diffusion.

For simplicity, we assume the subducting slab to be symmetrical about the $y-$ axis. In this way, the trench axis can now be viewed as a point in the $xz-$plane defined by $\mathbf{r}_{\text{trench}} = [x_1, z_1]$. Given θ, L, and $\mathbf{r}_{\text{trench}}$, we are now able to define another point $[x_2, z_2]$, and the line joining these two points delineates the axis of symmetry of the slab about the $xz-$plane. The slope and the z-intercept of the line can then be computed using the two-point formula. For convenience, let's call this line $\overline{AB}$.

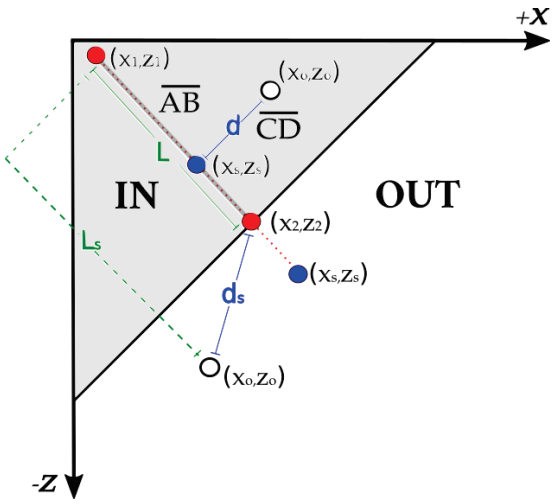

FIGURE 5.1: Simple 2D subduction zone geometry. The gray region encapsulates the grid nodes that are within the bounds of the slab geometry. The line joined by the two red circles designated as $\overline{AB}$ is the axis of symmetry of the slab. The blue circles are the projections of the grid nodes (white circles) along the axis of symmetry of the slab.

The temperature field is constructed in a regular grid of size $N_x \times N_y \times N_z$. To assign a thermal anomaly for each grid node at point $[x_0, z_0]$ that is representative of a slab, one must first determine the distance d from one grid node to $\overline{AB}$. However prior to this step, a grid node is required to be within the bounds of the symmetry axis. To do this, we need to determine another point $[x_s, z_s]$ along $\overline{AB}$ perpendicular to a given grid node. For any line perpendicular to $\overline{AB}$ joined by $[x_s, z_s]$ and $[x_0, z_0]$, its slope is just the negative reciprocal of the slope of $\overline{AB}$. Calling this line $\overline{CD}$, the intersection between $\overline{AB}$ and $\overline{CD}$ is the point $[x_s, z_s]$. The problem can be treated as a system of 2 linear equations, with 2 unknowns (*i.e.*), x_s and z_s. In terms of the slope m and the intercept b, the solution is given by:

$$\begin{aligned} x_s &= \frac{mz_0 + x_0 - mb}{m^2 + 1}. \\ z_s &= \frac{m}{m^2 + 1}(mz_0 + x_0 - mb) + b. \end{aligned} \tag{5.3}$$

The condition at which the grid node is within the said bounds is determined by computing the distance L_s from $[x_1, z_1]$ to $[x_s, z_s]$. If L is greater than L_s, then the grid node of interest is within bounds, otherwise, it is out of bounds. Figure 5.1 shows an illustration of this simple 2D geometry problem.

We model the anomaly in terms of a hyperbolic tangent function, the same way we defined the spherical temperature anomalies in the previous chapter. This coerces the

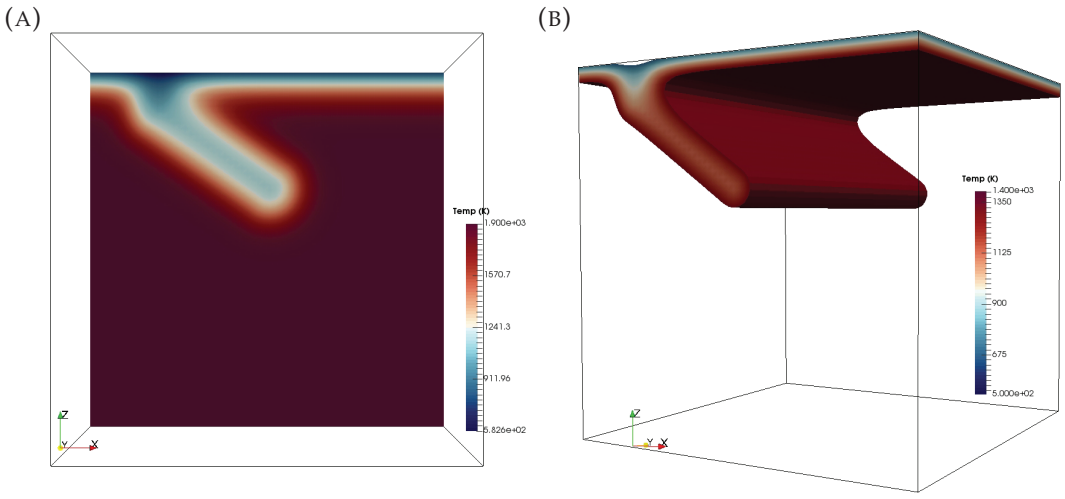

FIGURE 5.2: Thermal structure of a subduction zone parameterised in terms of geometrical points. The structure is rendered using the following input parameters: L = 150 km, $\theta = 35^0$, R = 120 km, T_c = 800 K. The left panel represents the vertical cross-section of the model whereas the right panel corresponds to the isovolumetric contour plot of the temperature field.

temperature to be smaller as it gets closer to the slab's symmetry axis. This also means that the spread of the tanh function relates to the thickness of the slab itself. In here we define it using the input parameter R. Larger values of R tantamount to a thicker slab. When L is greater than L_s, the temperature anomaly at node $\mathbf{r}$=$[x_0, y_0, z_0]$ is:

$$T_{\text{anomaly}}(\mathbf{r}) = -\frac{T_c}{2}\left[1 - \tanh\left(\frac{\beta}{L_{\text{scale}}}\left(d - \frac{R}{2}\right)\right)\right], \tag{5.4}$$

where d is the Euclidean distance between points $[x_0, z_0]$ and $[x_s, z_s]$, T_c is the maximum perturbation around the background temperature located along the axis of symmetry of the slab, and β controls the sharpness of the temperature gradient which is held at a fixed value. When $[x_0, z_0]$ is out of bounds, we can model the curvature of the slab by still using eq. (5.4) but replacing d with d_s (see Fig. 5.1) which is the Euclidean distance between points $[x_0, z_0]$ and $[x_2, z_2]$. Fig. 5.2 shows the thermal structure of the subduction model viewed in 2D (left panel) and in 3D (right panel) using the aforementioned parameterisation. The model parameters possess the following values: L = 150 km, θ = 35^0, R = 120 km, T_c = 800 K.

The medium rheology is modeled following the Frank-Kamenetskii approximation to Arrhenius-type viscosity. This means that the influence of temperature onto viscosity is supervised by the same activation energy E described in the previous chapter.

The viscosity field η is given by:

$$\eta(\mathbf{r}) = \eta_0 \exp\left[-E\frac{T(\mathbf{r}) - T_0}{T_0}\right], \tag{5.5}$$

where η_0 and T_0 are reference values for viscosity and temperature, respectively. In the inversions, the total number of parameters to be inverted for are five: four of which L, θ, T_c, and R characterize the temperature field of the subduction model, and the latter E controls the temperature-dependence of viscosity. The model vector $\mathbf{m}$ is thus:

$$\mathbf{m} = [L, \theta, T_c, R, E]. \tag{5.6}$$

5.3 Modifications in the Bayesian sampling scheme

Since the formulation of the inversion procedure is identical to the previous example, this section highlights instead the minor modifications implemented to the Bayesian sampling scheme tailored to the new parameterisation of the temperature field.

5.3.1 The prior distribution

We assume the model parameters to be independent. In this way, the prior distribution of each model parameter are separable and can be expressed as a product of each distribution:

$$p(\mathbf{m}) = \prod p(L)p(\theta)p(R)p(T_c)p(E). \tag{5.7}$$

Each prior on the model parameters follows a uniform distribution with wide bounds to avoid imposing hard constraints from the prior. Such a setup mimics a scenario where prior knowledge about the regional setting is scant and thus the solution to our inverse problem is more likely driven by the information provided by the data. The prior bounds are as follows: (1) 100 km - 200 km for L, (2) 80 km - 150 km for R, (3) 20^0 - 45^0 for θ, (4) 500 K to 1000 K for T_c, and (5) 5 to 12 for E.

5.3.2 Generation of new models along the Markov chain

At each iteration in the McMC algorithm, a new model $\mathbf{m}'$ is proposed by uniformly randomly selecting one of the possible set of moves:

1. Vary the length of the slab L. The slab length is perturbed according to a univariate Gaussian distribution centered at the current value of L.

TABLE 5.1: True model parameters defining the thermal structure of the subduction model.

Model parameter	Assigned value
L	150 km
R	120 km
θ	35^0
T_c	800 K
E	11.0

2. Vary the dip angle θ. The dip angle is perturbed according to a univariate Gaussian distribution centered at the current value of θ.

3. Vary the thickness of the slab R. The slab thickness is perturbed according to a univariate Gaussian distribution centered at the current value of R.

4. Vary the temperature of the slab T_c. The slab temperature is perturbed according to a univariate Gaussian distribution centered at the current value of T_c

After choosing one of the four possibilities, the proposal is always accompanied by the perturbation of the activation energy E. The activation energy is perturbed using a univariate Gaussian distribution centered at the current value of E.

5.4 Full forward procedure to predict surface wave measurements from the subduction model

Table 5.1 summarises the true model parameters used to generate the data (*i.e.,* local anisotropic surface wave dispersion curves). The data are produced coming from an anisotropic tensor computed with `D-Rex`. Note that the other scalar variables uninvolved in the inversion procedure such as the dimensionless parameters defining the Rayleigh number and the control parameters for CPO evolution modeling are preserved.

To generate the flow model, we employed multigrid methods with variable viscosity on a staggered regular grid parameterisation. This is followed by the interpolation of the velocity components to the center cells of a unit cube making up the grid. The flow is computed in a 400 km $\times$ 400 km $\times$ 400 km box whose resolution is of the size 6.25 km $\times$ 6.25 km $\times$ 6.25 km across all directions. Although most of the computational qualities involved in flow modeling from the previous demonstration are preserved, here we modified the top boundary conditions by imposing plate velocities to replicate real Earth subduction dynamics. The bottom and lateral boundary conditions are

always free-slip. Thus apart from slab descent due to gravitational instability, the slab itself moves laterally due to the influence of the imposed plate velocities.

Fig. 5.3a shows the vertical cross-section of the 3D instantaneous flow field induced by subduction. We observe one of the conspicuous features of subduction-induced flow, that is the existence of a local convection cell beneath the slab tip attributed to retrogade slab migration. This is accommodated by the existence of back-arc motion towards the trench made responsible mainly by trench suction and in part by the induced leftward motion due to the imposed plate velocity across the overriding plate. This also ensues flow ascension in front of the slab. The vigorous mixing observed across the sub-slab mantle resembling roll-back motion is mainly influenced by horizontal boundary effects. Such effects can be reduced by increasing the size of the model domain. Nevertheless, most features observed across the vertical cross-section are mainly predisposed by poloidal flow (*i.e.*, buoyancy-related motion). In essence as with any divergence-free vector field, our velocity field can be decomposed into a poloidal component, and a toroidal component which relates to horizontal flow due to the presence of lateral viscosity contrasts (Gable, O'connell, and Travis, 1991; Bercovici, 1995b). Since we imposed temperature-dependent viscosity, we are compelled to deal with toroidal motion due to lateral variations in viscosity as shown in Fig. 5.3b where we observe local vorticities around the slab edges.

Fig. 5.4 shows the vertical cross-section of the finite strain representation of the subduction model. Solid black lines pertaining to the orientation of the long axis of the finite strain ellipsoid (fse) are superimposed on top of the natural strains (*i.e.*, amplitude of finite deformation in terms of the relative lengths between the long and short axes of the fse). Finite strain orientations to first-order tend to be parallel to the direction of flow, however, may lag behind in some instances where deformation rapidly varies along the flow trajectory. The absence of deformation correlate well with the presence of the fortified and highly-viscous slab.

From the velocity field, we gain access to the macroscopic velocity gradients by second-order finite differencing. The 3D map of the local velocity gradient in conjunction with the temperature field are utilised to construct an elastic model of the synthetic subduction zone using a micro-mechanical model for CPO evolution for the anisotropic part, and a thermodynamic model for the isotropic part, respectively. At any arbitrary location in 3D space, the elastic model contains the elastic tensor $\mathbf{S}$ with 21-independent coefficients. Since it is arduous to interpret a fourth-rank tensor, it is often convenient to decompose $\mathbf{S}$ into a specific symmetry class to better analyse its properties. Fig. 5.5 illustrates the vertical cross-section of the elastic constants A_0 related to $PV-$waves and L_0 related to $SV-$waves associated with the subduction model.

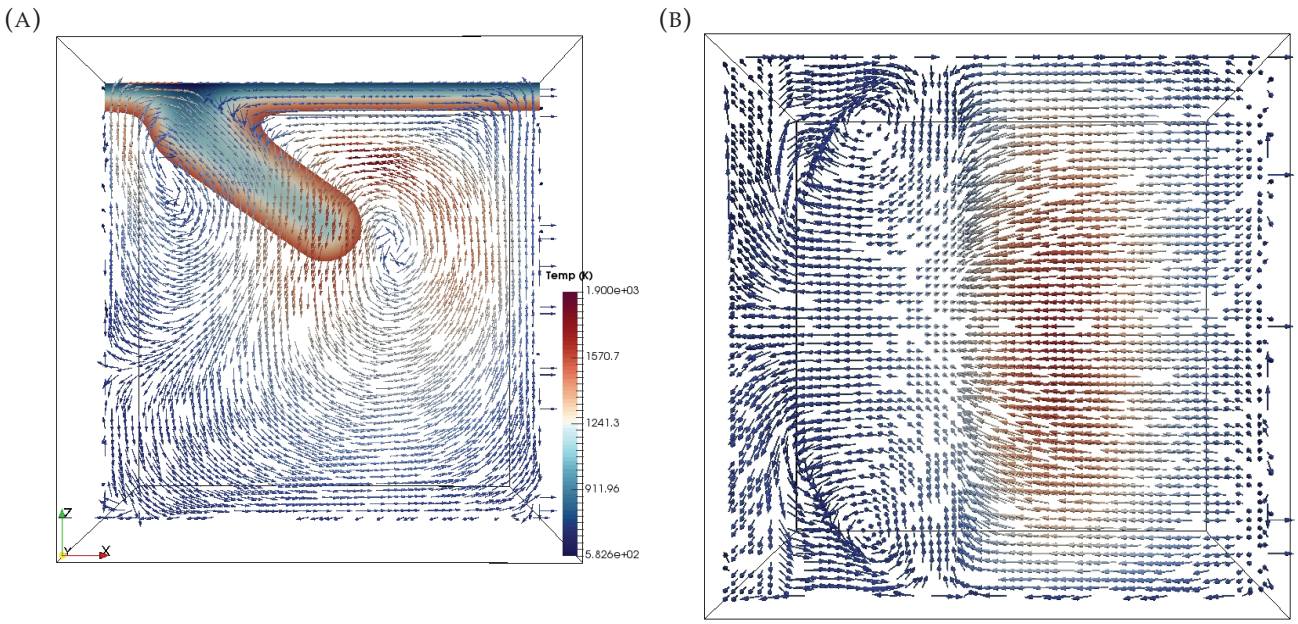

FIGURE 5.3: Instantaneous velocity field induced by subduction. Model domain is of the size $64 \times 64 \times 64$ elements, free-slip boundary conditions are imposed at the lateral and bottom sides. Opposing plate velocities are prescribed at the top to drive horizontal motion. (a) Vertical cross-section of the velocity field with the temperature field superimposed. (b) Overhead view of the velocity field. Fig. (b) illustrates the significance of lateral viscosity variations to produce toroidal fluid flow.

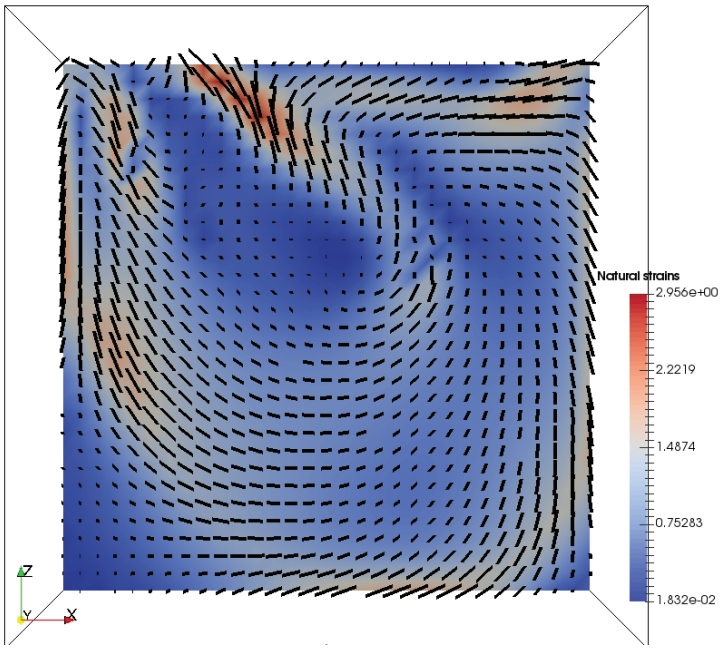

FIGURE 5.4: Cross-sectional view in the xz− plane of the natural strains (*i.e.*, amplitude of the fse). Solid black lines are attributed to the orientaiton of the long axis of the fse. Finite strain framework is usually used as a proxy to infer convective flow in the mantle.

The elastic constants are computed from the elastic projection of **S** to an azimuthally-averaged VTI medium (Montagner and Nataf, 1986). Both panels robustly map the cold subducting slab with L_0 exhibiting sensitivity to temperature variations more fervently than A_0. Since the constants A_0 and L_0 are not the isotropic averages of the fast and slow velocities but a linear combination of the elastic tensor, they possess small anisotropic pertubations thus explaining the presence of smearing in some areas of the maps.

On the other hand, Fig. 5.6 represents the vertical cross-section of $S-$wave radial anisotropy ξ (left panel) and of the total anisotropy (*i.e.*, norm fraction of the elastic tensor with respect to the isotropic component) (right panel). As observed, regions of positive radial anisotropy $\xi > 1$ correlate well with horizontal flow and of negative radial anisotropy $\xi < 1$ with vertical flow. In terms of the total anisotropy, the entrained mantle wedge adjacent to the plunging slab, and beneath the back-arc produced the most CPO due to shear deformation initiated by slab pull and reinforced by trench suction. Strong anisotropy produced across the shallower depths of the sub-slab can be attributed to roll-back motion augmented by boundary effects. Across the slab itself, we expect $\xi \approx 1$ since the material is designed to mimic rigid plates that widthstand deformation. This is also observed at the right panel where total anisotropy across the slab appears to be nearly non-existent. Thus in this case, the subducting slab is

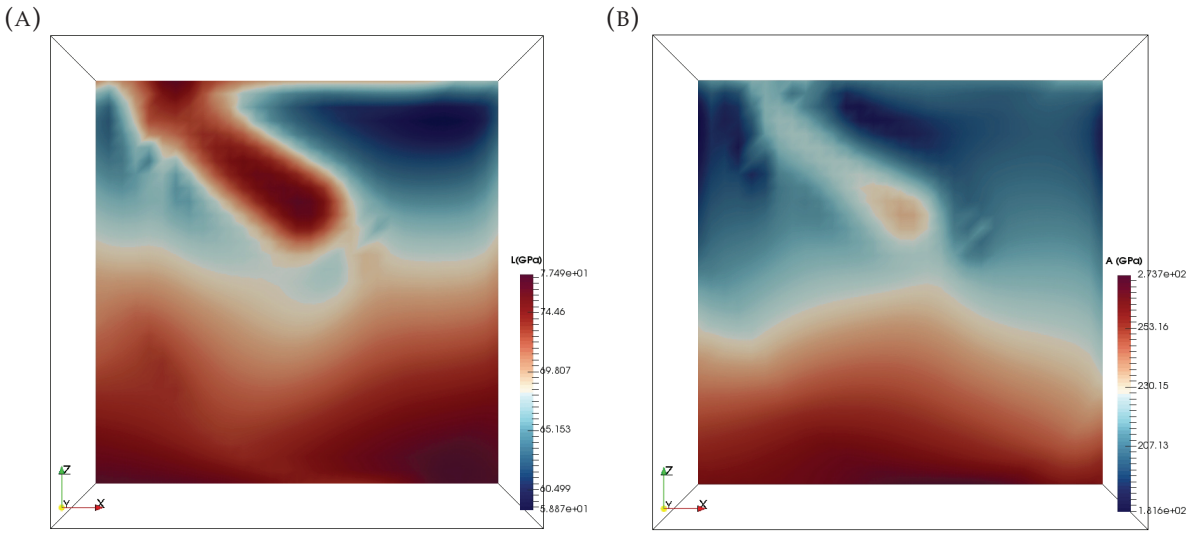

FIGURE 5.5: Cross-sectional view in the xz− plane of the elastic constants L_0 (left panel) and A_0 (right panel). Since elasticity strongly depends on temperature, we can easily map the cold subducting slab in the seismic models. The cross sections are taken at the center of the y− axis.

isotropic since no CPO is generated due to its resilient rheological integrity. The presence of small-scale artifacts in the anisotropic structures may be attributed to numerical errors associated with the forward calculations.

From an elastic model built from the spatial distribution of **S**, it is now permissible to compute 2D phase velocity maps and their azimuthal variations. For instance, Fig. 5.7 shows a map of the computed phase velocity and azimuthal anisotropy for Rayleigh waves at 100 s. The increase in velocity on the left portion of the map shown in Fig. 5.7a indicates the influence of the cold subducting slab. In Fig. 5.7b, the characteristic blue margin in between the yellow regions corresponds to the slab itself. This is also delineated by the shortening of the fast axis of azimuthal anisotropy within its vicinity (solid black lines). Although the orientation of the fast axis is a sufficient proxy to infer the horizontal projection of flow, it may still fail to render some important characteristics such as the presence of a toroidal component in the flow. This is because azimuthal anisotropy in surface waves is an integrated effect of the elastic anisotropy with depth. Furthermore, the latter depends on the deformation trajectory. Hence, absolute flow velocities may be well away from the orientation of its fast propagation. Finally, the deformation induced by subduction seemingly produces about 2 % azimuthal anisotropy in surface waves which spreads out almost evenly throughought the map and only restricted by the existence of the isotropic slab.

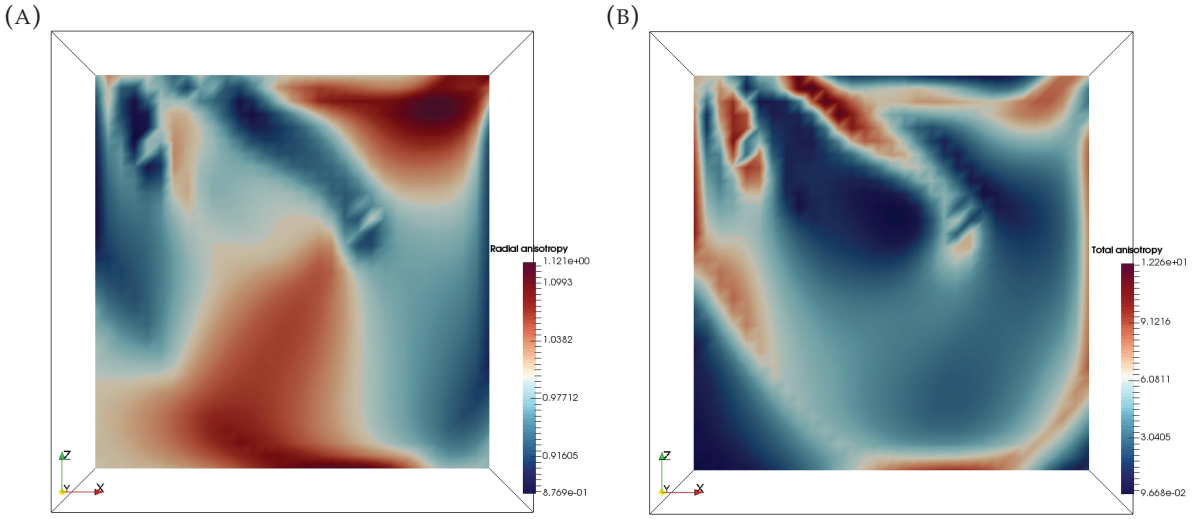

FIGURE 5.6: Cross-sectional view in the xz− plane of the shear wave radial anisotropy ξ (left panel) and the total anisotropy expressed in terms of the tensor norm fraction of $\mathbf{S}$ with respect to its isotropic component (right panel). The cross sections are taken at the center of the y− axis.

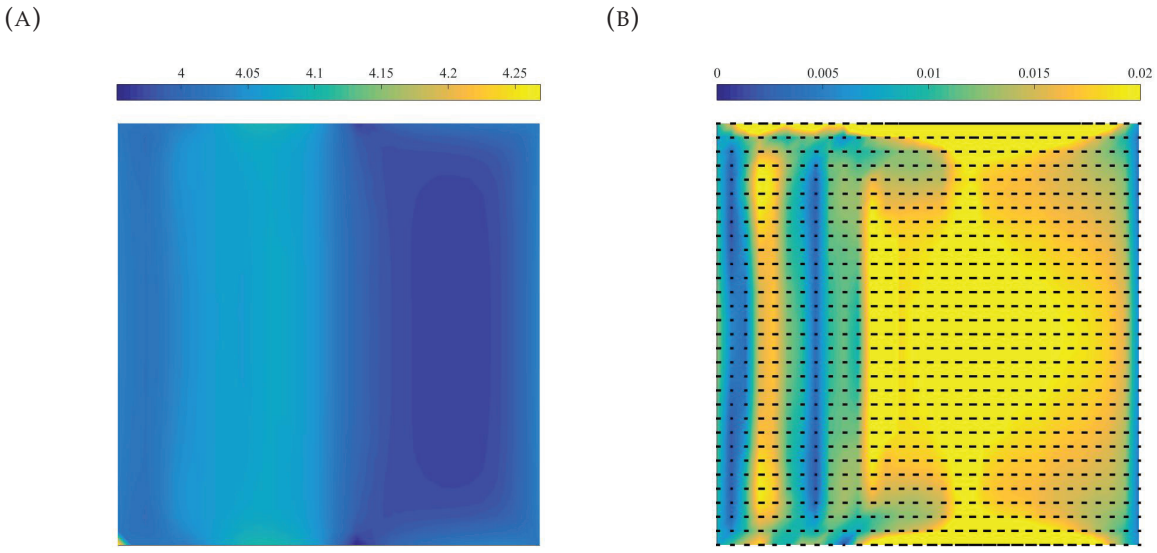

FIGURE 5.7: Phase velocity maps derived from a 3D deforming upper mantle beneath a subduction zone at 100s period. (a) Rayleigh wave phase velocity (km/s). (b) Azimuthal anisotropy in Rayleigh waves (km/s). The solid black lines correspond to the direction of the fast propagation axis. Surface wave maps always lie along the xy− lateral plane.

5.5 Inversion for the five unknown model parameters

Using the true values of the model parameters summarised in Table 5.1, we generated synthetic surface wave dispersion curves and their azimuthal variations at periods between 10 and 200 s with 10-s interval. The complete data consist of a regular array of 8×8 locations containing $\mathbf{c}_R$, $\mathbf{c}_L$, $\mathbf{c}_1$, and $\mathbf{c}_2$ spanning the entire surface. The synthetic data were computed based on elastic tensors calculated with `D-Rex`.

We added Gaussian uncorrelated noise onto $\mathbf{c}_R$, $\mathbf{c}_L$, $\mathbf{c}_1$, and $\mathbf{c}_2$. Since we are pushing the limits of geodynamic tomography by applying it with more physical parameterisations, it is also more appropriate to test the method with realistic synthetic data. As such, to replicate surface wave dispersion measurements that are correlated in period, we prescribed very low noise levels for $\mathbf{c}_R$ and $\mathbf{c}_L$ with $\sigma_{R,L} = 0.001$ km s^{-}1. Conversely, the azimuthal variations were assigned with $\sigma_{1,2} = 0.005$ km s^{-}1. Fig. 5.8 shows a synthetic surface wave dispersion curve with and without added noise at one specific geographical location.

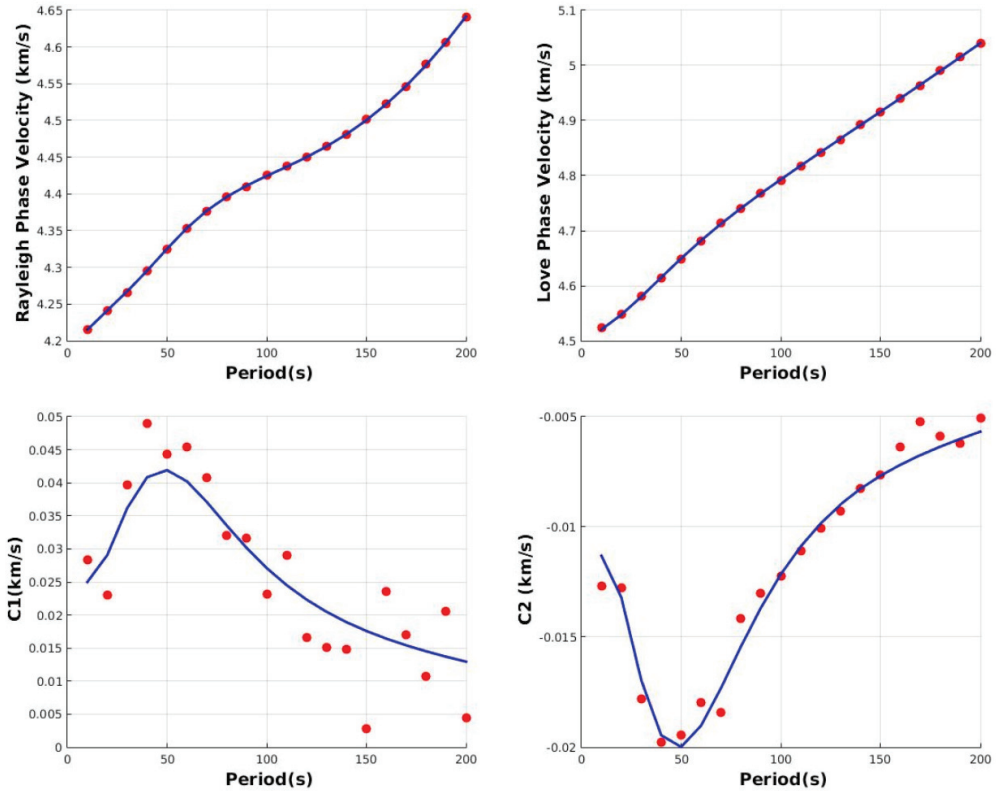

FIGURE 5.8: Synthetic surface wave dispersion curves from 10 to 200 s at a given geographical location (blue lines). The data used in the inversions have been added with Gaussian uncorrelated noise (red circles).

The inversion consists of 20 independent Markov chains containing 40 000 samples each initiated at a random model (*i.e.,* values for L, θ, R, T_c, and E are randomised for all chains) to ensure loose compliance to the initial model. Similar to the previous chapter, we demonstrate two cases: (1) an isotropic inversion and an (2) anisotropic inversion (*i.e.*, geodynamic tomography). Both cases are imposed with wide uniform priors as discussed in Section 5.3.1 allowing for more mobility when searching the parameter space. For efficient sampling, we commenced geodynamic tomography by first employing an isotropic inversion. Once the independent chains have converged in this phase, we then proceeded with the actual anisotropic inversion procedure. It is important to emphasise that we still implemented an artifical neural network (ANN) algorithm to approximate `D-Rex`. Since the current architecture of the ANN is problem-specific, it is designed based on training data generated by flow models produced by a family of thermal subduction models.

Figs 5.9 and 5.10 shows the 1D marginal posterior probability distribution on each model parameter (diagonal panels) and the joint marginal posterior probability distribution between a pair of model parameters (off-diagonal panels) to explore possible trade-offs for isotropic inversion and anisotropic inversion, respectively. The red lines and the black circles indicate their correct values. Both cases have exhibited a single misfit minima for the model parameters that define the thermal structure of the subduction model. However by incorporating geodynamic and petrological constraints, we observe that the entirety (this includes the activation energy E) are much more tightly constrained than the isotropic case, as evidenced by the considerable decrease in the spread of the distributions. The narrow widths of the posterior distributions are also a manifestation of the low noise levels accounted for in the inversions. Between these two effects, it can be implied that the imposition of geodynamic constraints contributes far more toward the robustness of the solutions than the usage of low-noise data.

It is expected that isotropic inversion hardly constrains E since isotropic velocities do not depend on deformation history but are directly derived from temperature and pressure for a given chemical and mineralogical composition. Contrastingly, anisotropic inversion effectively constrains E although the result is clearly biased as it is shown to be two values lower than the correct number. In actuality, bias exists in all the parameters at least except for the length of the slab L. Thus the existence of a misfit minima that are not in agreement with the true model parameters can only be explained by the use of an incorrect surrogate model. Indeed, implementing the correct forward model to compute anisotropy would eliminate this effect. However since it entails to be computationally expensive when employed with direct-search

algorithms, it is imperative to utilise fast-forward approximations such as neural networks. This necessetitates the inclusion of additional training data and/or possibly the partial or complete overhauling of the network architecture. Finally, the existence of a linear trade-off between a pair of parameters appear to be widespread. Such behavior is most apparent between the temperature of the slab T_c and the slab geometry particularly L and R. This is likely due to the accommodation of the increase in the slab temperature by an increase in its size.

Fig. 5.11 illustrates the reconstructed mean temperature field from both inversions (top panels) and their corresponding uncertainties in terms of the standard deviation (bottom panels). By visual inspection, we notice that the mean temperature field from the isotropic inversion to be not much different from the anisotropic case. Due to the low levels of noise in the data, anisotropy does not bring much in the recovery of the temperature field. However in the case of larger noise as demonstrated in the previous chapter, the inclusion of anisotropy in the inversions would be more beneficial. The standard deviation conveys a different story however as observed by its smaller amplitude in the case of geodynamic tomography (Fig. 5.11d). In both cases, the uncertainties are seemingly clustered across subducting slab with two discernible plunging stripes. This indicates a state of relaxation, or more prefereably, convergence of the Markov chains towards a stable solution. The plunging stripes therefore are a result of a random-walk behavior of the subducting slab about its axis of symmetry. The axis of symmetry is delineated by the area of low uncertainty partitioning the two plunging stripes of high uncertainties.

Fig. 5.12 shows the 1-D depth marginal posterior probability profiles at a given location for temperature, radial anisotropy, peak-to-peak azimuthal anisotropy, and the fast azimuth inferred from geodynamic tomography. We successfully jointly recovered azimuthal and radial anisotropy without having to explicitly invert for the elastic tensor. One of the key advantages of geodynamic tomography is its capacity to capture intricate and highly complex features as exemplified by the recovered amplitude of azimuthal anisotropy and its fast azimuth. Furthermore, one of the long standing problems of conventional surface wave tomography is the depletion of its resolving power with depth since its energy is mostly concentrated across the surface. Here we have demonstrated the ability of geodynamic tomography in the apparent eradication of this effect as evidenced by the preservation of the width of the posteriors in depth.

Lastly, geodynamic tomography offers the capability to resolve 3-D structures of any implicitly computed varible. As a demonstration, Fig. 5.13 illustrates the radial anisotropy (left panel) and the total anisotropy (right panel) obtained from the mean temperature model. A recurring issue is the underestimation of seismic anisotropy in

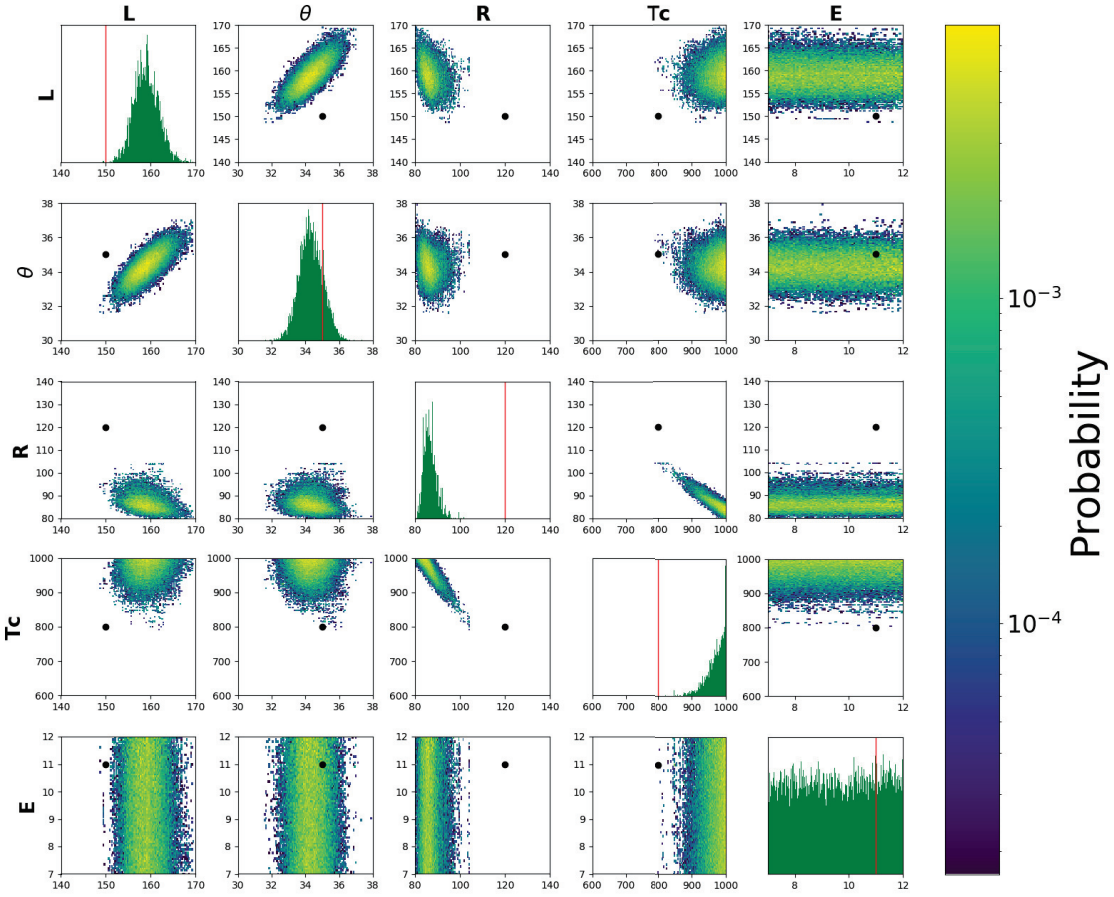

FIGURE 5.9: Posterior probability distribution in the 6-dimensional parameter space inferred from the isotropic inversion $p(\mathbf{m}|\mathbf{c}_R, \mathbf{c}_L)$. Diagonal panels show 1D marginal distributions for each model parameter. Off-diagonal panels show 2D marginal distributions and depict possible trade-offs between pairs of model parameters. The red vertical lines and the black markers indicate the true model values for the diagonal and the off-diagonal panels, respectively. The intensity pertains to the level of posterior probability (*i.e.*, high intensity means high probability, and thus low misfit).

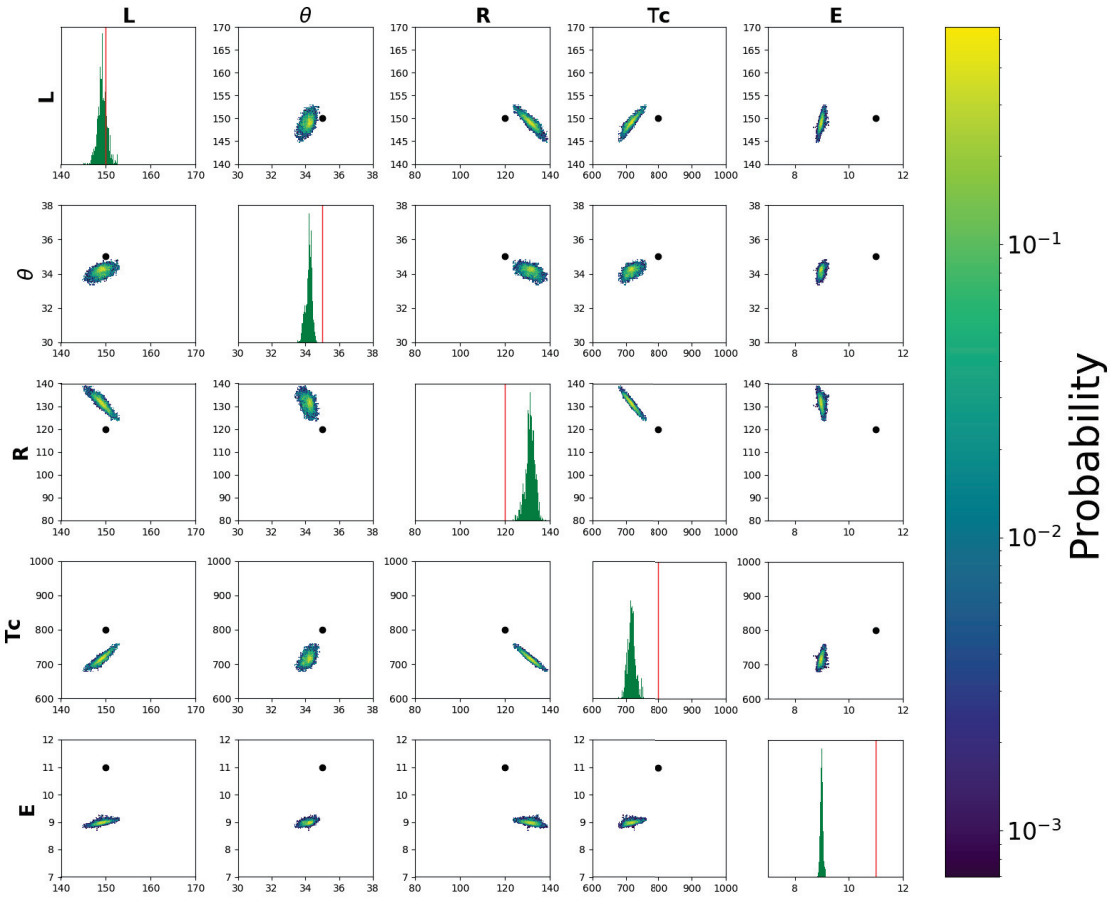

FIGURE 5.10: Posterior probability distribution in the 6-dimensional parameter space inferred from the anisotropic inversion $p(\mathbf{m}|\mathbf{c}_R, \mathbf{c}_L, \mathbf{c}_1, \mathbf{c}_2)$.

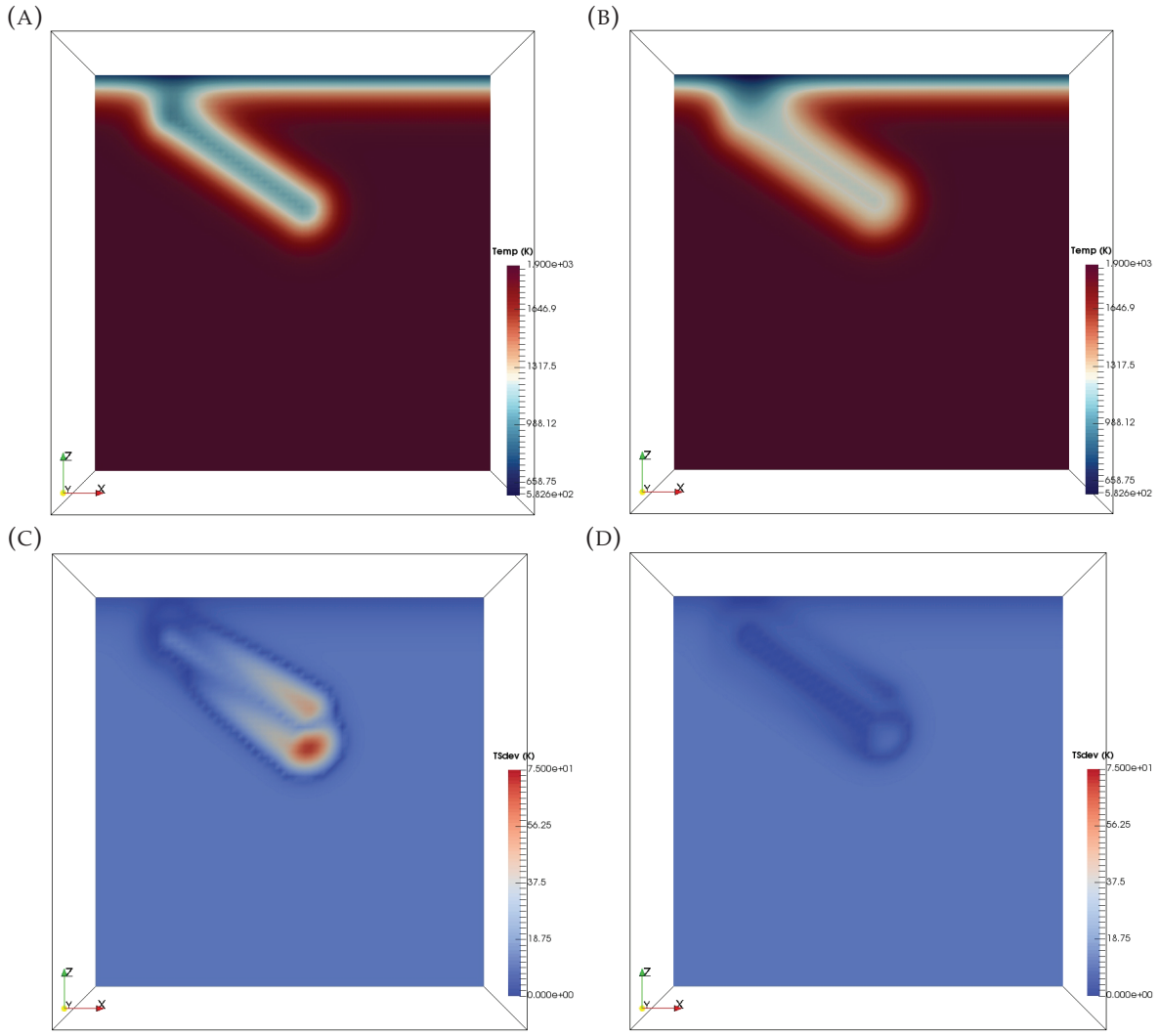

FIGURE 5.11: Upper panel: Cross-sectional view in the xz− plane of the mean temperature field recovered from (a) isotropic inversion, and (b) anisotropic inversion. Lower panel: Standard deviations around the mean temperature fields from (c) isotropic inversion, and (d) anisotropic inversion. These cross-sections are taken at the center of the y− axis.

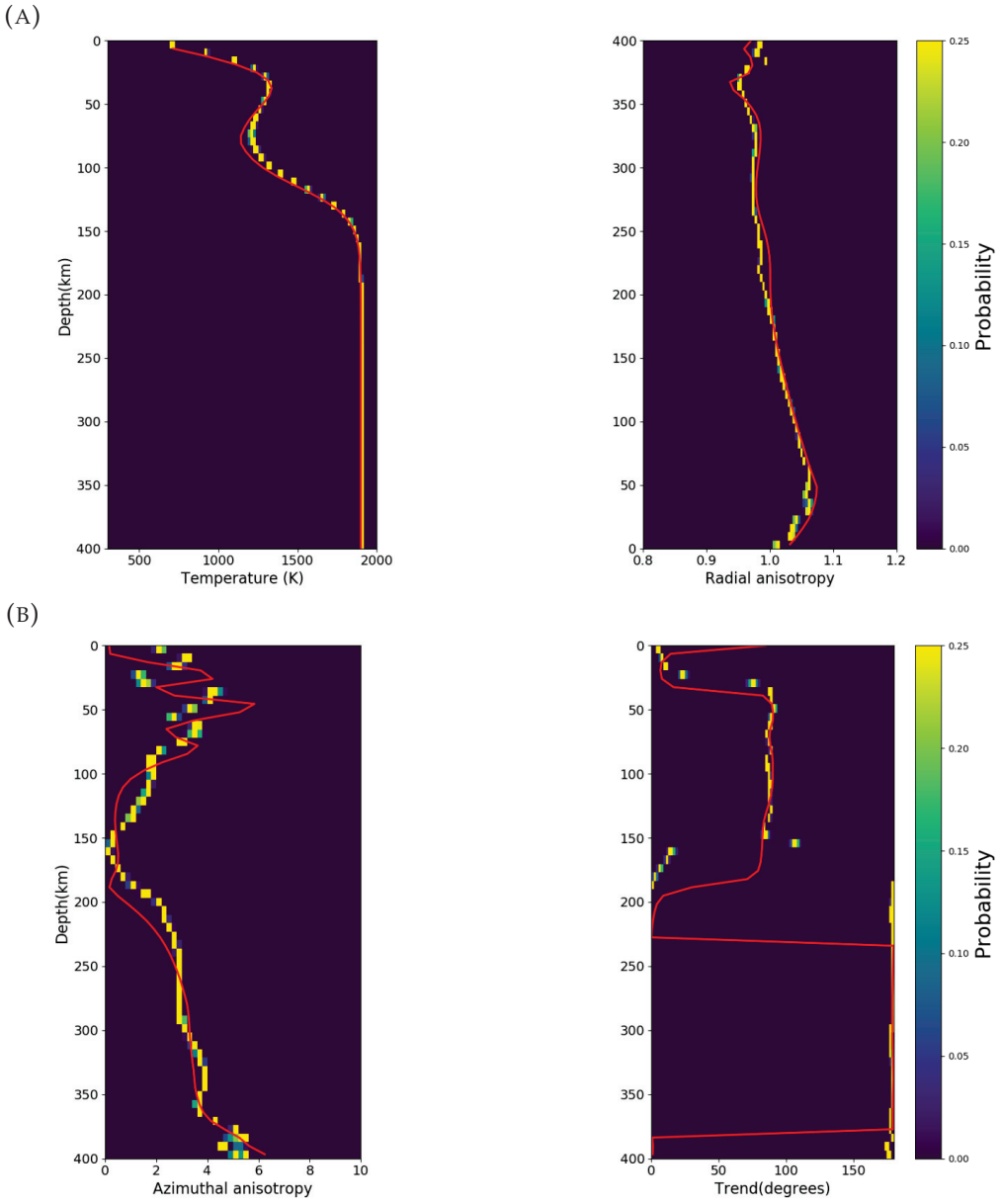

FIGURE 5.12: 1-D marginal posterior probability profiles with depth of several variables inferred from geodynamic tomography. Upper panel: Probability density plots of temperature and radial anisotropy. Lower panel: Probability density plots of peak-to-peak azimuthal anisotropy, and its fast axis with depth. The depth profiles of temperature and radial anisotropy are taken at ($x = 125$ km, $y = 225$ km). To show that azimuthal anisotropy is also well-constrained, we took a depth profile at ($x = 175$ km, $y = 225$ km), where the patterns of azimuthal anisotropy is highly complex. Geodynamic tomography offers the capability to constrain seismic anisotropy. The solid red lines indicate the true structures.

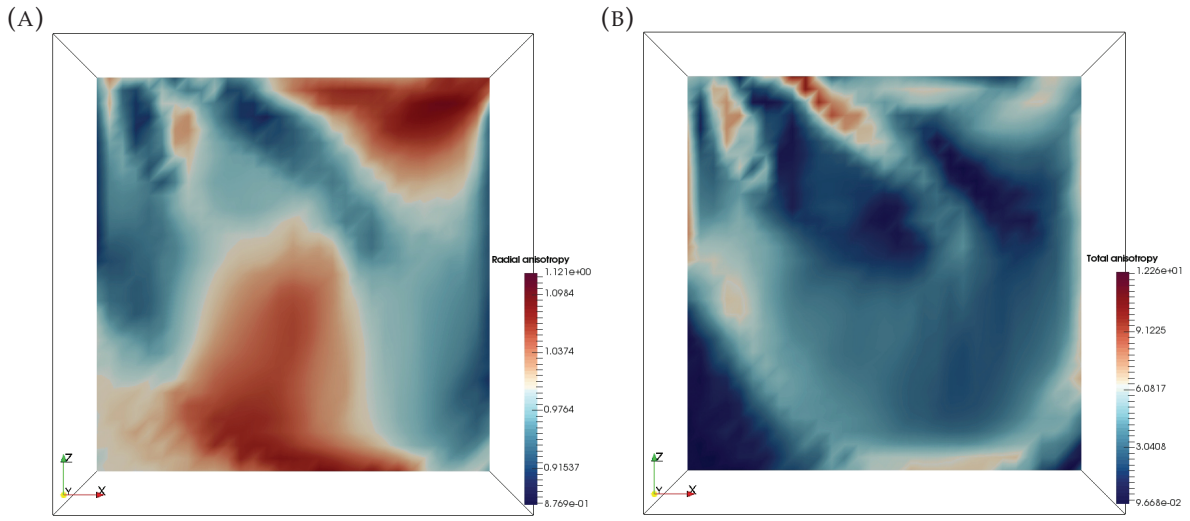

FIGURE 5.13: Cross-sectional view in the xz− plane of the radial anisotropy ξ (left panel) and the total anisotropy (right panel) obtained from the mean temperature model. The cross sections are taken at the center of the y− axis.

comparison with the true model (Fig. 5.6) due to the use of an approximate forward operator to model CPO evolution. This is compensated by the weakening of the slab rheology, by the reduction of the activation energy E (Fig. 5.10 bottom right panel), in order to produce larger levels of anisotropy.

5.6 Conclusion

Geodynamic tomography is an imaging technique that incorporates constraints from geodynamics and mineral physics to restrict the potential number of candidate seismic models down to a subset consistent with geodynamic predictions. Imposition of these constraints eliminate the need to explicitly invert for the elastic tensor everywhere. Instead, the elastic tensors are implicitly computed based on known micro-mechanical models for texture evolution. Such methods tremendously reduce the number of Earth parameters from a spatial distribution of elastic tensors with 21 independent coefficients down to a single scalar field (*e.g.* temperature).

We have tested the applicability of geodynamic tomography to a 3D deforming upper mantle induced by subduction. Isotropic Love and Rayleigh wave phase velocity measurements and their azimuthal variations at a given location were jointly inverted to recover the 3-D thermal structure of a synthetic subduction zone. The method is cast in a single self-consistent Bayesian inversion procedure where the solution is an ensemble of unknown model parameters defining the subduction thermal structure

and rheology. The assembly of these unknown parameters are distributed according to a posterior probability density function.

In the process, not only do we successfully recover the desired thermal structures, we have also constrained the complete pattern of upper mantle deformation induced by subduction, and to provide a quantitative interpretation of intrinsic seismic anisotropy. The Bayesian framework propounds the capability to render marginal posterior probability distributions not only of the target unknown Earth models, but also of any implicitly computed varible such as deformation and anisotropy, and quantify their associated uncertainty limits.

Chapter 6

Concluding remarks and perspectives

For the first time, we have reconciled seismology and mantle geodynamics, two seemingly independent subdisciplines, into a single self-consistent geophysical inversion approach called *Geodynamic Tomography*. Being at its infancy, much work remains to be done towards the recognition of its full potential when applied to real-Earth problems. As such, this chapter elucidates geodynamic tomography from a broader perspective. The first section summarises the achievements of the thesis. This is followed by the limitations of the method and possible future avenues to take in order to address such limitations. The last section explores possible strategies to consider towards the implementation of geodynamic tomography to real-data.

6.1 Geodynamic tomography: a redefinition of the seismic imaging problem

The long standing issue associated with conventional anisotropic seismic tomography, surface wave studies in particular, is its inability to resolve the 21-component elastic tensor independently at every location. Because of this, in practice only a restricted number of parameters are inverted for by assuming tensor symmetry relations. Anisotropic surface wave tomography however, even after the assumption of symmetry relations, remains troublesome due to the high-dimensionality of the free parameters surface waves are sensitive to. This motivated the development of tomographic techniques that exploit any intrinsic correlation among the elastic constants derived from petrological models. The result is the further reduction of the free parameters to be inverted for. In reality however, the full elastic tensor contains every ounce of information regarding the thermo-chemical structure and current state of deformation of the Earth's interior. Conventional anisotropic surface wave tomography may therefore fail to yield some key aspects associated with the Earth's present-day structure and dynamics.

This urged the need to redefine the surface wave tomography problem by combining geodynamic flow modeling, thermodynamic modeling, texture evolution modeling, and finally seismic forward modeling in order to formulate a general geophysical inversion strategy to tackle the problem stated above. The main advantage of such a technique is the confinement of the number of possible Earth models down to a subset that matches with geodynamic predictions, and the total elimination of parameterising the elastic tensor. Constructing the latter does not require the imposition of tensor symmetries at the outset but is purely driven instead by the physics of mantle convection. The underlying idea is that although surface waves are only sensitive to a limited number of elastic constants, the incorporation of geodynamic and petrological constraints serves as leverage to allow for the full elastic tensor to be resolved everywhere. As a favorable consequence of including such constraints, it is plausible to reduce the total amount of free parameters to be inverted for in the tomographic problem to a single scalar field (*e.g.*, density, temperature). The caveat is that the method is computationally expensive due to the use of a multitude of numerical recipes to model various geophysical processes.

Geodynamic tomography consists of two main ingredients. First is the computation of synthetic data (*e.g.*, surface wave dispersion measurements) given an Earth model (*e.g.*, temeprature). The forward model that describes the non-linear mapping between the model and the data involves the following series of steps: (1) geodynamic flow modeling using a scalar field (*i.e.*, temperature) as an input, (2) thermodynamic modeling to compute for the isotropic part of the elastic tensor from temperature and pressure at a given mineralogical and chemical compositions, (3) texture evolution modeling to compute for the intrinsically anisotropic part of the elastic tensor using the flow model derived from (1), (4) constructing the full elastic tensor from (2) and (3), and lastly (5) computing synthetic seismic data given the elastic model built from (4). The second procedure is the actual inversion process which involves the appraisal based on Bayesian inference of a randomly-selected Earth model using direct-search algorithms. The solution to the inverse problem is therefore a posterior distribution in each model parameter and their associated uncertainty bounds.

We first demonstrated this technique to a synthetic upper mantle whose thermal structure is constructed from spherical temperature anomalies. Here, we inverted surface wave phase velocity dispersion measurements and their azimuthal variations to retrieve the temperature and the viscosity fields. In the process, we have successfully recovered the complete pattern of upper mantle deformation. One of the main benefits of the Bayesian framework is the ability to cast any implicitly-computed variable such

as deformation and seismic anisotropy in terms of posterior probability distributions and the quantification of their associated uncertainties.

The second example deviates from the initially simple parameterisation in terms of spheres, towards its application to thermal structures of a synthetic subduction zone. The goal was to test the ability of geodynamic tomography to recover structures harbouring complex deformation patterns. Anisotropic surface wave dispersion measurements were inverted to retrieve five unknown parameters that define the thermal and rheological structure of the subduction zone. Results show that the unknown parameters are tightly constrained with the apparent existence of a single misfit minima in each model pararameter. Each model realisation however fails to swarm around its true value. Such results are attributed to the inability of the surrogate model to accurately replicate the correct forward model for computing anisotropy due to the inherent complexity of the deformation patterns considered.

6.2 Additional comments on geodynamic tomography

6.2.1 Model parameterisation strategies

The goal of Chapter 4 was to test the method in the most simple cases, and we acknowledge that our parameterisation of the temperature field in terms of a sum of spherical anomalies is simplistic. However, such parameterisation can be applied to invert for more complex geometries such as a detached slab, a homogeneous plume, or upper mantle structures beneath cratons. A step further will be to test more realistic approaches. One possible alternative parameterisation is the use of initial temperature models inferred from isotropic tomography, and an iterative update of the structure based on the anisotropy signature at the surface (*i.e.*, anisotropic surface wave dispersion curves). This, however, may only be feasible at the global scale due to boundary effects. It should still be possible to apply this technique at the regional scale, but the structure of interest should be far from the borders of the region considered in order to avoid these boundary effects. Another simple yet effective parameterisation would be to invert for constant parameters (*e.g.*, density, viscosity) within geometrical blocks defined from *a priori* information regarding the tectonics of the region (Baumann, Kaus, and Popov, 2014). This was demonstrated in Chapter 5, however in place of constant values, continuous basis functions were used instead to administer smoother parameterisations. In general, the quality of the results will depend on the choice of the model parameters, and the prior information available for the region of interest.

6.2.2 Neural network-based approach to texture evolution

The computational demands of direct sampling techniques such as McMC is high, as it requires evaluating the forward model a large number of times. Among all routines involved in the forward model, calculating CPO anisotropy proved to be the most costly. We therefore devised a surrogate model that computes texture evolution via a neural network, thus reducing the computation time by three orders of magnitude compared to D-Rex.

However, the surrogate model introduces theoretical errors, which can be reduced by using a network architecture or a training procedure more adapted to the problem at hand. More accurate predictions could be obtained by using a larger training data set, but this has a higher initial computational cost. We observed that the surrogate model does not generalise well. It was trained for specific types of flow (*e.g.*, convective flows due to spherical temperature anomalies/induced by subduction), and thus provided correct predictions only for flow models of the same nature. However, only these specific flow types were tested in the McMC scheme, and it is therefore not necessary here to have a general neural network that applies to any type of flows.

The success of our synthetic tests is in some ways a proof of the quality of the neural network. The inverted anisotropic seismic data sets were calculated using the exact D-Rex model. Therefore, any errors introduced by the network would manifest themselves by producing a poor fit to the observed data. These theoretical errors have been quantified and accounted for in the Bayesian inversion. If we want to treat another problem, such as a sinking slab with complex geometry, one needs to re-train the surrogate model for the specific parameterisation and prior distribution used. A possible future avenue of geodynamic tomography that is independent of this specific step would be to directly parameterise mantle flow, and build a family of expected convection patterns (together with their predicted anisotropy) to investigate flow patterns underneath mid-ocean ridges and subduction zones. Such parameterisation can be easily extended to the global scale by treating these patterns in terms of source and sink models derived from prescribed plate velocities (Bercovici, 1995a).

The Bayesian formulation is a practical tool to quantify and account for the theoretical errors introduced by the parameterisation choice and the surrogate model. Statistics of these errors can be studied by comparing responses obtained with the true forward model and the surrogate model. If the distribution of residuals is approximated as a normal distribution, theoretical errors can be accounted for in the likelihood function (Hansen et al., 2014b). However, the size of the residual vector may not be large enough to properly represent the statistics of errors. Here instead, we used a maximum likelihood estimation (MLE) to implicitly account for these theoretical errors (Dettmer,

Dosso, and Holland, 2007).

6.2.3 Adapting to other data-types

In this work, we assume that the measurement errors in the data are uncorrelated. In reality however, surface wave dispersion measurements are inherently smooth, and correlated both in space and in frequency. A simple improvement when modeling noise can be made by introducing a function that varies with period while still maintaining the assumption of uncorrelated errors, as in the work of Ravenna and Lebedev (2017). One may proceed a step further by constructing a covariance matrix of data noise, more importantly when working on highly spatially-correlated data sets.

It is also worth mentioning that the method is not limited to the use of a single data-type (*i.e.*, surface wave measurements) to effectively constrain the patterns of upper mantle deformation. This calls for the inclusion of other data-types such as gravity anomalies, surface topography, and/or surface velocities in a joint or separate approach. Such strategies have already been successfully implemented to invert for the 3D density structure of the mantle (e.g. Ricard and Wuming, 1991; Baumann, Kaus, and Popov, 2014).

6.3 Physical assumptions imposed

The trade-off between physical complexity and computational cost is evident in every geophysical problem considered. In this work, we chose to decrease the computational cost to massively explore the parameter space (using an inverse problem formulation) but at the price of using simplified physical assumptions.

6.3.1 Nature of the flow model

We assumed that the flow is in steady-state in order to trace the flow streamlines, which is a prerequisite to compute CPO anisotropy. However, this may not be the case in regions where flow appears to be time-dependent such as migrating trenches and mid ocean ridges (Heuret and Lallemand, 2005; Masalu, 2007). A time-dependent flow could be implemented by accounting for the evolution of the surface tectonics (Ricard et al., 1993) and the retrodiction of internal heterogeneities (Bunge, Hagelberg, and Travis, 2003; Steinberger, Sutherland, and O'connell, 2004). Nevertheless, steady-state assumption is still valid in some places such as intra-oceanic regions where flow has been observed to be in steady-state over the last 40 Myr (Becker et al., 2003; Becker et al., 2006).

Another limiting factor is the imposition of arbitrary boundary conditions on the sides of the model domain which strongly impact the nature of the flow. Note that the boundary conditions could be treated as an unknown parameter to be inverted for. An obvious way to address this issue is also to work at the global scale. In this case, a fast and reliable method to compute geodynamic flow in a spherical Earth is indispensable. To cite an example, semi-analytical circulation models such as that of Hager and O'Connell (1981) can be computed from simple density distributions assuming no lateral variations in viscosity. However, the latter may not render a reasonable assumption within the context of geodynamic tomography since lateral viscosity variations affect the flow significantly, and thus may also strongly influence the resulting anisotropy.

The rheological structure of the Earth's mantle is innately heterogeneous. Considering only temperature-dependent viscosity, one can roughly assume that a temperature variation of 100 K yields a viscosity variation by one order of magnitude. Therefore, neglecting lateral viscosity contrasts would only make sense should lateral temperature variations are small. This is probably untrue however in the global scale. For instance, mantle plumes are roughly 100 - 200 K warmer than the surroinding mantle. For subducting slabs, the temperature contrasts are even larger ($>$ 500 K). Thus, even if one is distant from the upper and the lower thermal boundary layers (*i.e.* the lithosphere and the core-mantle boundary, respectively), one should expect lateral viscosity variations to possess several (> 4-5) orders of magnitude. One might be tempted to justify neglecting these specific objects (*e.g.*, plates and plumes) by arguing that they are relatively small compared to the volume of the mantle. However, this would not work because these objects are those that drive the flow since gravitational instabilities originate from temperature differences.

Therefore, one cannot decouple viscosity variations from density variations. In the context of inverse modeling, the inclusion of lateral viscosity variations is indeed computationally more challenging. However, it remains attainable by performing these calculations in a coarser grid to obtain the general pattern of the flow. This step can be followed by interpolating the coarse grid solution on a finer grid prior to the computation of CPO. Using iterative approaches to flow calculations, another practical approach is to degrade the accuracy of the solution should convergence be an impediment. When cast in a Bayesian formulation, the modeling error due the approximation of the flow can be accounted for in the inversion process, similar to how the errors due to the ANN were dealt with. Consequently, texture evolution modeling at the global scale could reasonably be achieved from flows of this nature. The availability of global

surface wave maps on the other end should thus make geodynamic tomography feasible at the global scale.

6.3.2 Composition of the mantle

We assumed that the composition of the mantle to be olivine, with an A-type crystal fabric, corresponding to dry upper mantle conditions. In the real Earth, seismic wave velocities not only depend on temperature and pressure variations, but also on the compositional structure of the minerals. Recently, self-consistent thermodynamic models have already been incorporated in seismic inversion schemes to interpret tomographic images in terms of mantle composition (Ricard, Mattern, and Matas, 2005; Cammarano et al., 2009). While the bulk properties (*i.e.*, seismic wave speeds) obtained from Gibbs minimisation are isotropic, to our knowledge, deformation-induced anisotropy has not yet been formulated cohesively with thermodynamic models, let alone casting it in an inverse problem.

In general, intrinsic anisotropy in the upper mantle results from complex deformation processes, which depend on a plethora of physical parameters that may be linked to one another. Unlike conventional tomographic techniques, the elastic structure recovered in our scheme directly depends on the assumptions made on these upper-mantle processes. As an example, one would expect that the inclusion of enstatite in our models would dilute the overall amplitude of anisotropy in surface waves. In addition, inversion results depend on control parameters for CPO modeling such as the choice of the slip systems of olivine. For the moment, the value of these parameters have been chosen ad hoc, using current available knowledge mostly originating from laboratory experiments, and thus can be viewed as prior (regularisation). Ultimately, the flexibility of Bayesian inference would allows us to treat these parameters as unknown parameters to be inverted for in geodynamic tomography.

6.4 Potential application to a real Earth problem

On the real-data application of geodynamic tomography, we anticipate that the inversion strategy should consist of three stages. In the first stage, we assume that surface wave dispersion maps are freely available, and that we have a general overview of the 3D geometry of the problem. This is arguably the case in most places where surface wave dispersion measurements are widely available thanks to an ever growing amount of seismic records. We then invert an array spanning the entire geographical surface of local isotropic Rayleigh wave phase velocity dispersion curves inferred

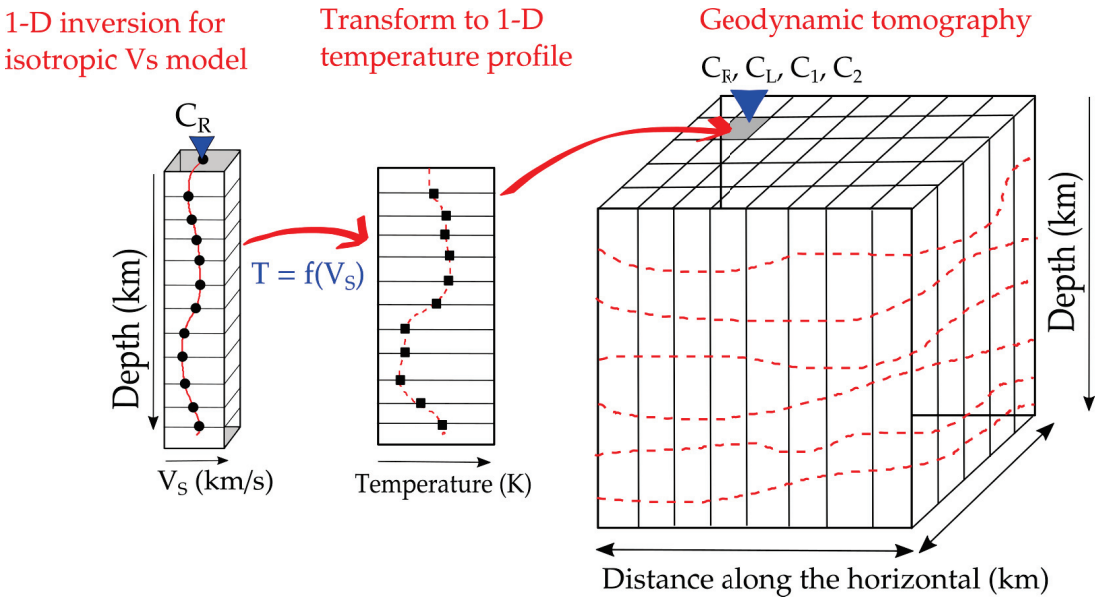

FIGURE 6.1: Three-step approach to geodynamic tomography. The first step involves 1D isotropic surface wave tomography to infer the shear wave structure (solid red lines) from a geographical array of isotropic Rayleigh wave dispersion measurements (blue triangle on top of the 1-D column). From the 1-D shear wave structures, the second step is the estimation of 1D temperature profiles (dashed red lines across the 1-D column). From the set of 1D temperature profiles, one may then build a smooth 3-D model of temperature through interpolation. The 3-D model can thus be viewed as a collection of 1-D columns containing depth profiles of temperature. Geodynamic tomography commences by using the 3-D temperature field (dashed red lines in the 3-D model) as a starting model followed by its iterative update through the inversion of anisotropic surface wave dispersion measurements (blue triangles on top of the 3-D model).

from these maps for 1-D depth isotropic V_S models. From the shear wave structure, the second stage involves the estimation of 1-D depth profiles of temperature T using first-order scaling relations between V_S and T. A more elaborate and precise yet more computationally demanding approach is the inverse implementation of self-consistent thermodyamic models to infer T from V_S for any given bulk composition. The 1-D depth profiles of temperature can then be placed side-by-side followed by refinement through various interpolation schemes to build a smooth 3-D temperature field. The last stage is geodynamic tomography itself, that is, using the 3-D temperature field inferred from the previous stages as the starting model to iteratively update its structure by inverting anisotropic surface wave dispersion curves. Fig. 6.1 is a schematic representation of this three-step inversion strategy.

Chapter 7

Quantifying Intrinsic and Extrinsic Contributions to Elastic Anisotropy Observed in Tomographic Models

Submitted to *Journal of Geophysical Research: Solid Earth*

7.1 Motivation

In the previous chapters, we have assumed that the anisotropy is purely intrinsic, that is, it results from the crystallographic preferred orientation (CPO) of mantle minerals due to finite deformation over time. It therefore provides unique constraints to convective processes in the mantle. Such has been demonstrated by our method, geodynamic tomography, where we auspiciously retrieved the complete deformation patterns in the mantle.

However, small-scale isotropic heterogeneities may generate seismic wavefield patterns that are identical to those produced by anisotropy. Because of this, seismically-unresolved small-scales can be mapped into an apparent/extrinsic anisotropy in tomographic models. This calls for the need to distinguish intrinsic anisotropy from extrinsic anisotropy to avoid misinterpretations about the probable cause of anisotropy observed in tomographic models.

In this chapter, we investigate the separate contributions of the intrinsic and the extrinsic components of anisotropy. The problem is applied to tomographic models of the mantle exhibiting spatial inhomogeneities in both its isotropic and intrinsically-anisotropic components.

7.2 Summary

Large-scale seismic anisotropy inferred from seismic observations has been loosely interpreted either in terms of intrinsic anisotropy due to Crystallographic Preferred Orientation (CPO) development of mantle minerals or extrinsic anisotropy due to rock-scale Shape Preferred Orientation (SPO). The coexistence of both contributions misconstrues the origins of seismic anisotropy observed in seismic tomography models. It is thus essential to discriminate CPO from SPO in the effective anisotropy of an upscaled/homogenized medium, that is, the best possible elastic model recovered using finite-frequency seismic data assuming perfect data coverage. In this work, we investigate the effects of upscaling an intrinsically anisotropic and highly heterogeneous Earth's mantle. The problem is applied to a 2-D marble cake model of the mantle with a binary composition in the presence of CPO predicted from a micro-mechanical model. We compute the long-wavelength effective equivalent of this mantle model using the 3-D non-periodic elastic homogenization technique. Our numerical findings predict that overall, upscaling purely intrinsically anisotropic medium amounts to the convection-scale averaging of CPO. As a result, it always underestimates the anisotropy, and may only be overestimated due to the additive extrinsic anisotropy from SPO. Finally, we show analytically (in 1-D) and numerically (in 2-D) that the full effective radial anisotropy ξ^* is approximately the product of the effective intrinsic radial anisotropy ξ^*_{CPO} and the extrinsic radial anisotropy ξ_{SPO}:

$$\xi^* \approx \xi^*_{\mathrm{CPO}} \times \xi_{\mathrm{SPO}}.$$

Based on the above relation, it is imperative to homogenize a CPO evolution model first before drawing comparisons with tomographic models. Such a composite law can therefore be used as a constraint to better estimate the separate contributions of CPO and SPO from the effective anisotropy. As a demonstration, we use the composite law to identify an SPO model from an existing tomographic model of the upper-mantle underneath a mid-ocean ridge together with a homogenized CPO model.

7.3 Introduction

Seismic anisotropy in the Earth's mantle originates from various processes and can be observed at different spatial scales. At the mineral scale, crystallographic preferred orientation (CPO) of mantle minerals due to progressive shearing over time (Nicolas and Christensen, 1987; Maupin and Park, 2015) produces large-scale intrinsic anisotropy.

On the other hand, rock-scale shape preferred orientation (SPO) such as a stack of layered cracks, seismic discontinuities, and/or preferentially oriented conduits containing fluid intrusions unresolved by long period seismic waves are mapped as large-scale extrinsic anisotropy (Backus, 1962; Crampin and Booth, 1985).

Although these two mechanisms are completely different, a medium may either be intrinsically anisotropic or strongly heterogeneous depending on the minimum wavelength of the observed wavefield used (Maupin et al., 2007; Bodin et al., 2015). Backus (1962) showed that a horizontally-layered isotropic medium is equivalent to a homogeneous radially anisotropic medium with a vertical axis of symmetry when sampled by seismic waves whose wavelength is much longer than the length of layers. This urged seismologists to interpret tomographic models separately depending on the type of data used (*i.e.*, different data-types sample different length scales). Scattering studies use high frequency body waves and interpret small-scale isotropic heterogeneities in terms of phase changes (e.g. Tauzin and Ricard, 2014) or chemical stratification (e.g. Tauzin et al., 2016). Likewise, long period surface waves with typical wavelengths of the order 10^2 km retrieve a smooth anisotropic mantle with scales consistent with convective flow (e.g. Debayle and Ricard, 2013; Bodin et al., 2015; Maupin and Park, 2015). Surface waves however lack the resolving power to recover sharp seismic discontinuities and instead, map these as long wavelength radial anisotropy (Backus, 1962; Capdeville et al., 2013). Anisotropic structures retrieved from tomography may therefore be a combination of apparent extrinsic anisotropy due to SPO and deformation-induced intrinsic anisotropy. The ambiguity whether a material is intrinsically anisotropic or strongly heterogeneous may mislead seismologists in interpreting the structural origin of seismic anisotropy observed in tomographic images.

In this chapter, we extend the work of Alder et al. (2017) by obtaining the long-wavelength effective equivalent of the marble cake model of the mantle hypothesized by Allègre and Turcotte (1986), but in the presence of intrinsic anisotropy. Our aim is to quantify the level of effective anisotropy resulting from elastic homogenization, that is, the relegated version of the true Earth as seen by long-wavelength seismic tomography. Section 7.4 provides an overview of the elastic homogenization theory and highlights a composite law that separates intrinsic and extrinsic anisotropy for a layered and anisotropic media. Here, we analytically prove in 1-D that the effective anisotropy varies with the square of isotropic heterogeneity and with that of intrinsic anisotropy plus a cross term related to their coupling. In section 7.5, we build a 2-D media analogous to the marble cake model where we consider a mechanical mixture of two end-member compositions. We follow this by introducing intrinsic anisotropy due to mantle deformation associated with convection patterns consistent with the

marble cake model. We compute the long-wavelength effective equivalent of the 2-D models using the Fast-Fourier Homogenization algorithm (Capdeville, Zhao, and Cupillard, 2015). Section 7.6 presents the results of the previous section. One of the major findings is that in the absence of isotropic heterogeneities, intrinsic anisotropy is always underestimated upon homogenization due to the spatial averaging of the preferred orientation of the anisotropic minerals. We also verify numerically that the composite law derived in section 7.4 can be extended to 2-D media. Finally in section 7.7, we apply the composite law to infer the extrinsic component of anisotropy from a tomographic model of the upper-mantle beneath a mid-ocean ridge with the help of a homogenized CPO model.

7.4 Elastic homogenization

Seismic tomography outputs only a smooth representation of the real Earth due to limited frequency band. The resulting model however is not just a simple spatial average but is produced from highly non-linear upscaling relations. In the context of wave propagation, such upscaling relations, also known as elastic homogenization, remove seismic heterogeneities whose scales are much smaller than the minimum wavelength of the observed wavefield and instead replace them with effective properties.

This section provides an overview of the elastic homogenisation theory. As a preliminary, we introduce Backus homogenisation for any 1-D elastic media. We then apply this technique to derive analytical expressions for seismic anisotropy in the general case of an intrinsically anisotropic and heterogeneous VTI medium, followed by its application to a more specific problem. The third part covers a brief description of other homogenisation techniques tailored to 2-D and 3-D elastic media. Finally we introduce a jargon tool-box to guide the reader regarding the proper use of some important terminologies.

7.4.1 Definition of terms

Hereafter, what we refer to as the *reference medium* $\mathbf{S}(\mathbf{r})$ is an elastic model of the real Earth varying in space $\mathbf{r}$ that accounts for both intrinsic anisotropy due to CPO and small-scale isotropic heterogeneities that resemble marble cake-like patterns. This reference medium can be treated as a sum of several decompositions resulting from a cascade of orthogonal projections (Browaeys and Chevrot, 2004). One can then express $\mathbf{S}(\mathbf{r})$ in terms of an isotropic tensor $\mathbf{S_I}(\mathbf{r})$ plus an intrinsically anisotropic component

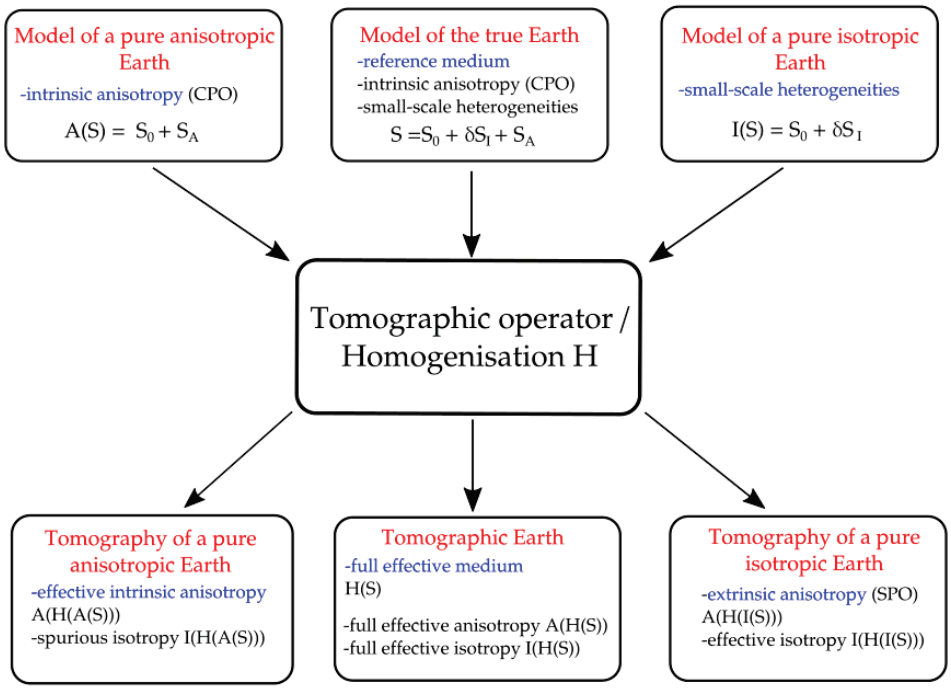

FIGURE 7.1: Homogenization of different Earth models and their respective outputs. The true Earth mantle (top middle box) is described by an average isotropic model $\mathbf{S_0}$, isotropic heterogeneities, $\delta\mathbf{S_I}$ and intrinsic anisotropy $\mathbf{S_A}$, the sum of which being the elastic model $\mathbf{S}$ that tomography tries to recover. However, tomographic methods have only access to a homogenized model $\mathcal{H}(\mathbf{S})$ (or full effective medium). This model has both isotropic components symbolized by $\mathcal{I}(\mathcal{H}(\mathbf{S}))$ and anisotropic components, $\mathcal{A}(\mathcal{H}(\mathbf{S}))$. The goal of this chapter is to quantify the differences between $\mathcal{A}(\mathcal{H}(\mathbf{S}))$ and $\mathcal{A}(\mathbf{S})$, $\mathcal{I}(\mathcal{H}(\mathbf{S}))$ and $\mathcal{I}(\mathbf{S})$. Numerically we can also discuss how an anisotropic model without isotropic anomalies (boxes on the left) can be recovered and if the tomographic inversion can lead to spurious isotropic anomalies. Reciprocally (boxes on the right), one can quantify how much a pure isotropic model is recovered by the tomographic inversion and what is the level of spurious anisotropy (SPO) that can be estimated.

$\mathbf{S_A}(\mathbf{r})$ related to CPO:

$$\mathbf{S}(\mathbf{r}) = \mathbf{S_I}(\mathbf{r}) + \mathbf{S_A}(\mathbf{r}), \tag{7.1}$$

where $\mathbf{S_I}(\mathbf{r})$ can be decomposed further into:

$$\mathbf{S_I}(\mathbf{r}) = \mathbf{S_0} + \delta\mathbf{S_I}(\mathbf{r}). \tag{7.2}$$

Here, $\mathbf{S_0}$ is an isotropic tensor uniform in space, and $\delta\mathbf{S_I}(\mathbf{r})$ is a deviation from $\mathbf{S_0}$ in space related to the small-scale isotropic heterogeneities. The reference medium becomes:

$$\mathbf{S}(\mathbf{r}) = \mathbf{S_0} + \delta\mathbf{S_I}(\mathbf{r}) + \mathbf{S_A}(\mathbf{r}). \tag{7.3}$$

For convenience, let us introduce an operator $\mathcal{I}$ that extracts an elastic tensor from $\mathbf{S}$ solely related to the isotropic heterogeneities, and an operator $\mathcal{A}$ that extracts the anisotropic component from $\mathbf{S}$. Using such notations, $\mathcal{I}(\mathbf{S})$ is simply equation (7.2), whereas $\mathcal{A}(\mathbf{S})$ is given by:

$$\mathcal{A}(\mathbf{S}(\mathbf{r})) = \mathbf{S_0} + \mathbf{S_A}(\mathbf{r}). \tag{7.4}$$

These notations will be used heavily in the rest of the text to denote the isotropic and anisotropic components of the reference medium. The complexity of $\mathcal{A}(\mathbf{S})$ urged geophysicists to use various notations to quantify the level of seismic anisotropy. One of which quantifies the relative strength of horizontally-polarized $S-$waves (V_{SH}) with vertically-polarized $S-$waves (V_{SV}), called radial anisotropy. The *intrinsic radial anisotropy* associated with $\mathcal{A}(\mathbf{S})$ will be denoted by ξ_{CPO}.

In the event where seismic observations sample this reference medium, seismic tomography only outputs a smooth representation of $\mathbf{S}$ due to limited frequency band, a process known as *upscaling* or *homogenization*. Note that these two terminologies are interchangeable but to avoid further confusion, we will go with the latter. Seismic tomography can be viewed as a mathematical operator $\mathcal{H}$ that homogenizes $\mathbf{S}$. From this moment forth, any variable applied by $\mathcal{H}$ will refer to the long-wavelength effective equivalent of it. For $\mathbf{S}$, it is $\mathcal{H}(\mathbf{S}) = \mathcal{H}(\mathbf{S_0} + \delta\mathbf{S_I} + \mathbf{S_A})$ which we now refer to as the *full effective medium*. The anisotropic component of the full effective medium given by $\mathcal{A}(\mathcal{H}(\mathbf{S}))$ will be referred hereafter as the *full effective anisotropy* and its isotropic component $\mathcal{I}(\mathcal{H}(\mathbf{S}))$ is the *full effective isotropy*. We will symbolize the *full effective radial anisotropy* corresponding to $\mathcal{A}(\mathcal{H}(\mathbf{S}))$ with ξ^*.

On the other hand, the tomographic counterpart of a pure anisotropic Earth (*i.e.*, a model where only the anisotropic component varies spatially) is $\mathcal{H}(\mathcal{A}(\mathbf{S})) = \mathcal{H}(\mathbf{S_0} + \mathbf{S_A})$ where $\mathcal{A}(\mathcal{H}(\mathcal{A}(\mathbf{S})))$ is the *effective intrinsic anisotropy*. Due to the non-linearity of $\mathcal{H}$ however, it creates spurious isotropic heterogeneities in the elastic tensor $\mathcal{I}(\mathcal{H}(\mathcal{A}(\mathbf{S})))$

as a byproduct albeit trivial. The *effective intrinsic radial anisotropy* corresponding to $\mathcal{A}(\mathcal{H}(\mathcal{A}(\mathbf{S})))$ will then be designated as ξ^*_{CPO}. Finally, the tomographic counterpart of a pure isotropic Earth (*i.e.*, a model where the isotropic component varies spatially, and the anisotropic component is zero) is $\mathcal{H}(\mathcal{I}(\mathbf{S})) = \mathcal{H}(\mathbf{S_0} + \delta\mathbf{S_I})$ where the non-negligible spurious anisotropic component due to SPO $\mathcal{A}(\mathcal{H}(\mathcal{I}(\mathbf{S})))$ is called *extrinsic anisotropy*. Here, *extrinsic radial anisotropy* will be denoted by ξ_{SPO} (Refer to Figure 7.1 for a comprehensive summary).

7.4.2 Backus homogenization

A vertically transverse isotropic (VTI) medium can be described by five elastic parameters A, C, F, L, and N, also known as the Love parameters (Love, 1906). Supposing that axis 3 is the symmetry axis, the local $\mathbf{S}$ for a VTI solid can be expressed in Voigt notation as:

$$\mathbf{S} = \begin{pmatrix} A & A-2N & F & 0 & 0 & 0 \\ A-2N & A & F & 0 & 0 & 0 \\ F & F & C & 0 & 0 & 0 \\ 0 & 0 & 0 & 2L & 0 & 0 \\ 0 & 0 & 0 & 0 & 2L & 0 \\ 0 & 0 & 0 & 0 & 0 & 2N \end{pmatrix}, \tag{7.5}$$

and the level of shear wave radial anisotropy can be compactly written as:

$$\xi = \frac{N}{L}. \tag{7.6}$$

Following the groundwork laid by Thomson (1950), Postma (1955), and Anderson (1961), Backus (1962) explicitly showed analytical upscaling relations for seismic waves propagating in a 1-D stratified medium. These upscaling relations refer to the filtering processes applied to the non-linear combinations of the fine-scale properties of the 1-D medium provided that they are smaller than the minimum wavelength of the observed wavefield. Thus for a 1-D fine-scale isotropic medium, its long-wavelength effective equivalent primarily assumes the behavior of a homogeneous, transversely isotropic medium.

To perform such calculations, we first define a minimum wavelength λ_{h} of the observed wavefield (*i.e.*, homogenization wavelength). In general, this is required for non-periodic medium with no scale separation. Since scales smaller than λ_{h} are homogenized, we define the threshold wavelength λ_0 for which heterogeneities are considered small scales as (Capdeville, Zhao, and Cupillard, 2015):

$$\lambda_0 = \epsilon_0 \lambda_h, \tag{7.7}$$

where ϵ_0 is a scale-separation parameter that depends upon the choice of the user. In this chapter, we set $\epsilon_0 = 0.5$ equivalent to the sensitivity of seismic waves to heterogeneities whose scales are smaller than half of λ_h in an effective manner e.g. Capdeville, Zhao, and Cupillard, 2015; Alder et al., 2017.

According to Backus (1962) for any 1-D fine-scale media regardless of its elastic nature (*i.e.*, purely heterogeneous, anisotropic or a combination of both), the effective equivalent of the elastic constants, for instance, N and L concerning the shear wave velocities can be expressed in terms of an arithmetic mean and a harmonic mean, respectively:

$$N^* = \langle N \rangle , \tag{7.8}$$

$$L^* = \langle 1/L \rangle^{-1} , \tag{7.9}$$

where $\langle . \rangle$ refers to the spatial average over λ_0, and * denotes the long wavelength effective equivalent. As a supplementary, the effective density ρ^* is simply the arithmetic mean of the local density ρ:

$$\rho^* = \langle \rho \rangle . \tag{7.10}$$

The effective shear wave radial anisotropy ξ^* is essentially the ratio between the effective equivalents of N and L:

$$\xi^* = \frac{N^*}{L^*} = \langle N \rangle \langle 1/L \rangle . \tag{7.11}$$

7.4.3 An analytical expression to quantify CPO and SPO in a 1-D layered media

Let us consider an intrinsically anisotropic (CPO component) and finely-layered (SPO component) VTI medium. Similar as to how we defined our reference medium in section 7.4.1, we regard the elastic parameters N and L as the sum of an isotropic component defined by the shear moduli μ, and local anisotropic perturbations N_A and L_A,

respectively:

$$N(z) = \mu(z) + N_A(z), \tag{7.12}$$

$$L(z) = \mu(z) + L_A(z). \tag{7.13}$$

For any transversely isotropic medium, the isotropic equivalent of μ can be computed following the Voigt averaging method (Montagner, 2007; Maupin et al., 2007):

$$\mu = \frac{1}{15}(C + A - 2F + 6L + 5N), \tag{7.14}$$

where A and C are elastic parameters concerning $P-$waves, and F relates to the so-called 'ellipticity'(*i.e.*, the velocity along the direction interposing fast and slow velocities). Assuming no $P-$wave anisotropy and setting F to unity, one can simplify equation (7.14) to:

$$\mu = \frac{1}{3}(2L + N). \tag{7.15}$$

Knowing equations (7.6) and (7.15), one can re-write N and L in terms of μ and ξ_{CPO} giving:

$$N = \xi_{\mathrm{CPO}}\frac{3\mu}{2 + \xi_{\mathrm{CPO}}}, \tag{7.16}$$

$$L = \frac{3\mu}{2 + \xi_{\mathrm{CPO}}}. \tag{7.17}$$

It is now straightforward to determine the anisotropic components N_A and L_A by equating equations (7.12) and (7.16). For the sole purpose of segregating the isotropic and anisotropic components, the forms of N and L consistent with equation (7.1) is therefore:

$$N(z) = \mu(z) + \left(\xi_{\mathrm{CPO}}(z)\frac{3\mu(z)}{2 + \xi_{\mathrm{CPO}}(z)} - \mu(z)\right), \tag{7.18}$$

$$L(z) = \mu(z) + \left(\frac{3\mu(z)}{2 + \xi_{\mathrm{CPO}}(z)} - \mu(z)\right), \tag{7.19}$$

where any variable as a function of z implies variations in space.

To calculate the long-wavelength effective equivalent of ξ, let us first write the parameters μ and ξ_{CPO} as:

$$\mu(z) = \langle\mu\rangle + \delta\mu(z), \tag{7.20}$$

$$\xi_{\mathrm{CPO}}(z) = \langle\xi_{\mathrm{CPO}}\rangle + \delta\xi_{\mathrm{CPO}}(z), \tag{7.21}$$

where $\langle\mu\rangle$ and $\langle\xi_{\mathrm{CPO}}\rangle$ are the spatially-averaged counterparts, and $\delta\mu$ and $\delta\xi_{\mathrm{CPO}}$ with $\langle\delta\mu\rangle$ and $\langle\delta\xi_{\mathrm{CPO}}\rangle = 0$ are small-scale heterogeneities in the shear modulus and intrinsic radial anisotropy, respectively. The long-wavelength effective equivalents N^* and $1/L^*$ are:

$$N^* = \langle N\rangle = \left\langle \xi_{\mathrm{CPO}} \frac{3\mu}{2+\xi_{\mathrm{CPO}}} \right\rangle = \left\langle (\langle\xi_{\mathrm{CPO}}\rangle + \delta\xi_{\mathrm{CPO}}) \frac{3(\langle\mu\rangle+\delta\mu)}{2+\langle\xi_{\mathrm{CPO}}\rangle+\delta\xi_{\mathrm{CPO}}} \right\rangle, \quad (7.22)$$

$$1/L^* = \langle 1/L\rangle = \left\langle \frac{2+\xi_{\mathrm{CPO}}}{3\mu} \right\rangle = \left\langle \frac{2+\langle\xi_{\mathrm{CPO}}\rangle+\delta\xi_{\mathrm{CPO}}}{3(\langle\mu\rangle+\delta\mu)} \right\rangle. \quad (7.23)$$

We attempt to simplify equations (7.22) and (7.23) by calling forth two simplifying assumptions:

1. Weak contrast among seismic wave velocities (and hence, in the shear moduli) so that $\delta\mu/\langle\mu\rangle \to 0$.

2. Weak contrast in intrinsic radial anisotropy so that $\delta\xi_{\mathrm{CPO}}/2 + \langle\xi_{\mathrm{CPO}}\rangle \to 0$. Although not physically intuitive, such an assumption does not necessarily mean weak anisotropy, but may also imply strong levels of intrinsic anisotropy but nearly homogeneous (*i.e.*, small spatial variations).

Using these two simplifying assumptions, and knowing $\langle\delta\mu\rangle$ and $\langle\delta\xi_{\mathrm{CPO}}\rangle = 0$, we are now in the position to evaluate equations (7.22) and (7.23). Approximating to the second-order we get:

$$N^* \approx \frac{3\langle\mu\rangle}{2+\langle\xi_{\mathrm{CPO}}\rangle}\left(\langle\xi_{\mathrm{CPO}}\rangle - \frac{2}{(2+\langle\xi_{\mathrm{CPO}}\rangle)^2}\langle\delta\xi_{\mathrm{CPO}}^2\rangle + \frac{2}{\langle\mu\rangle(2+\langle\xi_{\mathrm{CPO}}\rangle)}\langle\delta\mu\cdot\delta\xi_{\mathrm{CPO}}\rangle\right), \quad (7.24)$$

$$1/L^* \approx \frac{2+\langle\xi_{\mathrm{CPO}}\rangle}{3\langle\mu\rangle}\left(1 + \frac{1}{\langle\mu\rangle^2}\langle\delta\mu^2\rangle - \frac{1}{\langle\mu\rangle(2+\langle\xi_{\mathrm{CPO}}\rangle)}\langle\delta\mu\cdot\delta\xi_{\mathrm{CPO}}\rangle\right). \quad (7.25)$$

As a supplement, L^* is computed by simply approximating equation (7.25) to the second order:

$$L^* = \frac{3\langle\mu\rangle}{2+\langle\xi_{\mathrm{CPO}}\rangle}\left(1 - \frac{1}{\langle\mu\rangle^2}\langle\delta\mu^2\rangle + \frac{1}{\langle\mu\rangle(2+\langle\xi_{\mathrm{CPO}}\rangle)}\langle\delta\mu\cdot\delta\xi_{\mathrm{CPO}}\rangle\right). \quad (7.26)$$

Multiplying equations (7.24) and (7.25) and neglecting terms higher than order two, the full effective radial anisotropy ξ^* due to fine-layering and intrinsic anisotropy is

therefore:

$$\xi^* \approx \langle \xi_{\mathrm{CPO}} \rangle - \frac{2}{(2+\langle \xi_{\mathrm{CPO}} \rangle)^2} \langle \delta \xi_{\mathrm{CPO}}^2 \rangle + \frac{\langle \xi_{\mathrm{CPO}} \rangle}{\langle \mu \rangle^2} \langle \delta \mu^2 \rangle + \frac{2 - \langle \xi_{\mathrm{CPO}} \rangle}{\langle \mu \rangle (2+\langle \xi_{\mathrm{CPO}} \rangle)} \langle \delta \mu \cdot \delta \xi_{\mathrm{CPO}} \rangle. \tag{7.27}$$

Equation (7.27) explicitly shows the separate effects of the small-scales in the isotropic component and in the intrinsically anisotropic component onto the effective anisotropy as 'seen' by seismic waves. Assuming the medium to be devoid of intrinsic anisotropy (*i.e.*, $\xi_{\mathrm{CPO}} = 1$ and $\delta\xi_{\mathrm{CPO}} = 0$), the full effective radial anisotropy ξ^* directly relates to the level of heterogeneities in the shear moduli. Here, $\left\langle \delta\mu^2 \right\rangle$ refers to the variance of the shear modulus for scales much smaller than λ_0. It can be interpreted as the extrinsic radial anisotropy ξ_{SPO} due to the seismically unresolved small-scale isotropic heterogeneities. It varies as the square of the heterogeneities which is in agreement with the result of Alder et al. (2017). On the other hand, when the isotropic component is uniform (*i.e.*, $\delta\mu = 0$), ξ^* also varies with the square of heterogeneities, but now in intrinsic anisotropy. This can be interpreted as the effective intrinsic radial anisotropy ξ^*_{CPO}, or the radial component of the intrinsic anisotropy that gets smoothed out as a result of upscaling. Interestingly, its overall effect is to underpredict intrinsic anisotropy as indicated by the minus sign in front of the second term. In the absence of small-scale isotropic heterogeneities, we anticipate anisotropy to be always underestimated by tomography. Finally, equation (7.27) apparently suggests the existence of a cross term $\langle \delta\mu \cdot \delta\xi_{\mathrm{CPO}} \rangle$ due to the coupling of intrinsic anisotropy and the shear modulus. Supposing spatial variations in both components are significant such as in interfaces, the correlation term should strongly influence the anisotropy mapped in tomographic models. Nevertheless it is propounded that this correlation term is trivial and hence may need not be accounted for (Bakulin, 2003).

In the Earth's asthenosphere, we do not expect velocity variations (*i.e.* $\Delta V_S/(V_{S1} + V_{S2})$) oftentimes to be larger than 5% e.g. Xu et al., 2008; Stixrude and Jeanloz, 2015. This roughly translates to ∼10% heterogeneities in the shear modulus assuming constant density ρ. To perform a numerical estimate, we consider binary-structured horizontal laminations where CPO only exists in one of the two phases (*e.g.* in the harzburgite phase when considering a mechanical mixture of eclogite and harzburgite) with $\langle \xi_{\mathrm{CPO}} \rangle = 1.025$, where $\xi_1 = 1$ and $\xi_2 = 1.05$ (*i.e.*, ∼5% radial anisotropy analogous to what is observed in the upper-mantle e.g. Long and Becker, 2010), with the other parameters being $\rho = 3500$ kg·m^{-3}, $V_{S1} = 4.4$ km·s^{-1}, $V_{S2} = 4.7$ km·s^{-1}, and $\langle\mu\rangle$=72.5 GPa. Clearly by applying equation (7.27), the intrinsic component (first term) contributes the most to the effective anisotropy with $1 - \langle \xi_{\mathrm{CPO}} \rangle = 0.025$ wherein its spatial variations' overall effect is to tone-down the amplitude of anisotropy by an amount of

$\sim -10^{-4}$ (second term). This is followed by the SPO component (third term) which is responsible for the amplification of anisotropy ($\sim +10^{-3}$). Lastly, the cross term provides the least contribution ($\sim \pm 10^{-4}$) and therefore can reasonably be ignored in most cases. The $\pm$ sign denotes that it may increase or decrease anisotropy depending on the coupling pattern between the shear modulus and intrinsic anisotropy.

As an added bonus of deriving equations (7.24) and (7.26), one could also compute for the effective Voigt-averaged shear modulus μ^* through:

$$\mu^* = \frac{2L^* + N^*}{3}. \tag{7.28}$$

Plugging these expressions to equation (7.28), we get:

$$\mu^* = \langle\mu\rangle - \frac{2}{\langle\mu\rangle(2 + \langle\xi_{\mathrm{CPO}}\rangle)}\langle\delta\mu^2\rangle - \frac{2\langle\mu\rangle}{(2 + \langle\xi_{\mathrm{CPO}}\rangle)^3}\langle\delta\xi^2_{\mathrm{CPO}}\rangle + \frac{4}{(2 + \langle\xi_{\mathrm{CPO}}\rangle)^2}\langle\delta\mu \cdot \delta\xi_{\mathrm{CPO}}\rangle. \tag{7.29}$$

Based on the above equation ignoring intrinsic anisotropy, the homogenized shear modulus μ^* is always smaller than its spatially-averaged version $\langle\mu\rangle$. Such a result is logical in the 1-D case. Here, radial anisotropy induced by fine-layering is always positive thereby having $N^* > L^*$. Since L 'counts' twice and N once in its isotropic average, its long-wavelength effective equivalent μ^* is always slower than $\langle\mu\rangle$. Contrastingly if one neglects isotropic heterogeneities and only consider variations in intrinsic anisotropy, homogenization also results in the underestimation of the shear modulus. One would predict that homogenization at shorter wavelengths leads to the creation of spurious isotropic heterogeneities due to small-scales in CPO. Lastly and as expected, the cross term recurs due to the spatial correlation between the shear modulus and intrinsic anisotropy.

We acknowledge that the homogenized expressions given by equations (7.27) and (7.29) in terms of the isotropic shear modulus μ may not be particularly convenient for seismologists. In practice, seismologists observe spatial distributions in V_S and not in μ. If one assumes that density is uniform, then $\delta\mu/\mu$ can be simply replaced by $2\delta V_S/V_S$. On the other hand, if one assumes that density increases with V_S, one could also establish long-wavelength effective equivalent expressions for V_S in the same manner as μ using simple empirical relations for density such as that of Tkalčić et al. (2006).

As an example, let us examine a stack of planar layers with alternating shear moduli values determined by $\pm\ \Delta\mu/2$ (Figure 7.2a middle panel). This is translated from shear wave velocity variations $\Delta V_S/(V_{S1} + V_{S2})$ of about 10%. The 1-D depth profile

spans from $0-1000$ km with layers of equal thicknesses of 5 km. Positive intrinsic radial anisotropy ($\xi = 1.2$) is prescribed in the even layers, whereas the odd layers are isotropic ($\xi = 1$) (Figure 7.2a right panel). Using equations (7.16) and (7.17), one can then compute for the Love parameters L and N (Figure 7.2a left panel). In practice, Backus homogenization can be implemented by filtering out the spatial variations in N and L much smaller than $\lambda_0 = \varepsilon_0 \lambda_h$. Upon upscaling at an extremely long homogenization wavelength $\lambda_h \sim 10^3$ km, the resulting signals for N and L are flat, and simply given by their arithmetic and harmonic means, respectively (Figure 7.2a left panel). Once the long-wavelength effective equivalents N^* and L^* are acquired, one can then compute for the full effective radial anisotropy ξ^* through equation (7.11) (solid red line in Figure 7.2a right panel), and the effective Voigt-averaged shear modulus μ^* through equation (7.28) (solid red line in Figure 7.2a middle panel). For reference, we also estimate ξ^* and μ^* using equations (7.27) and (7.29) respectively (dashed blue lines in Figure 7.2a middle and right panels), whereby showing robust predictions. Of course, both expressions are only valid assuming small heterogeneity perturbations and hence they are expected to worsen when considering strong contrasts. Finally, Figure 7.2b illustrates a different scenario where ξ only exists in the odd layers (Figure 7.2b right panel). In essence when the shear modulus and intrinsic anisotropy are uncorrelated, the homogenized parameters μ^* and ξ^* should be the same regardless. However, a slight offset in μ^* and ξ^* of Figure 7.2b with respect to Figure 7.2a can be observed which is exclusively attributed to this cross term as hinted by equations (7.27) and (7.29). Strictly speaking, the reduction in the amplitude of the effective properties arises from the switch in signs in the cross term from positive to negative $\langle \delta\mu \cdot \delta\xi \rangle$, implying that in the second scenario, the shear modulus and intrinsic anisotropy are now anti-correlated.

In the following section, we investigate two special cases, that is, a purely isotropic (*i.e.*, finely-layered) and a purely anisotropic (*i.e.*, no spatial variations in isotropic component) 1-D media by finding equivalent expressions for extrinsic radial anisotropy ξ_{SPO} and effective intrinsic radial anisotropy ξ^*_{CPO}. By doing so, our goal is to elicit a simple composite law related to equation (7.27) that can be extrapolated to 2-D and 3-D media.

7.4.4 Composite law for radial anisotropy

In the case of an isotropic medium with spatially-varying shear modulus, the terms of equation (7.27) in $\delta\xi_{\text{CPO}}$ are zero while $\langle \xi_{\text{CPO}} \rangle = 1$. The radial anisotropy is entirely

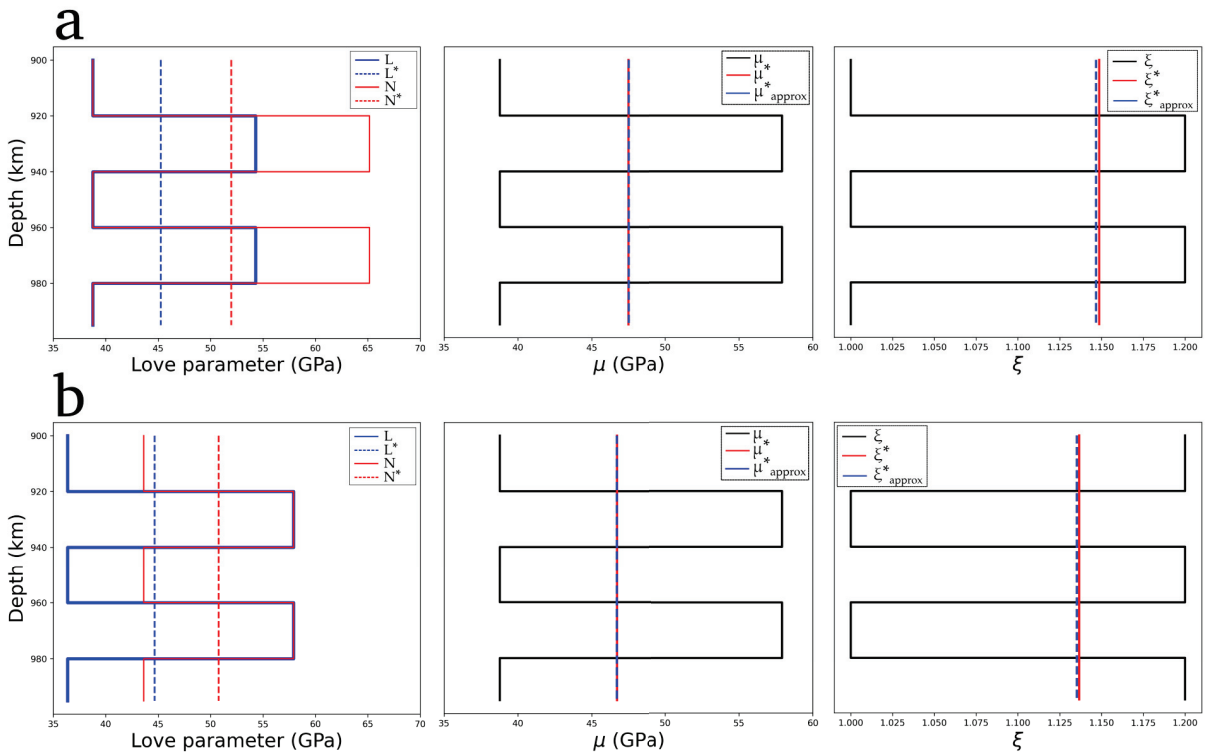

FIGURE 7.2: 1-D binary and periodic media with 10% isotropic heterogeneities in terms of $\Delta V_S/(V_{S1}+V_{S2})$ in the presence of intrinsic anisotropy with 20% heterogeneities in its radial component prescribed across: (a) even layers, and (b) odd layers. Upon homogenization over very long wavelengths (*e.g.*, λ_h = 2000 km), the resulting signals are flat (variables denoted by (*)). The dashed blue lines at the middle (μ^*_{approx}) and left panels (ξ^*_{approx}) correspond to the predicted long-wavelength effective equivalents using equations (7.29) and (7.27), respectively. The difference in the homogenized shear moduli and radial anisotropy between (a) and (b) is attributed to the cross term as implied by equation (7.27). Since the medium is periodic, it is enough to only display a portion of the medium.

due to SPO:

$$\xi_{\mathrm{SPO}} \approx 1 + \frac{1}{\langle\mu\rangle^2}\langle\delta\mu^2\rangle. \tag{7.30}$$

On the other hand, a purely anisotropic medium without spatial variations in the shear modulus leads to an effective anisotropy obtained from equation (7.27) by assuming $\delta\mu = 0$ yielding:

$$\xi^*_{\mathrm{CPO}} \approx \langle\xi_{\mathrm{CPO}}\rangle - \frac{2}{(2 + \langle\xi_{\mathrm{CPO}}\rangle)^2}\langle\delta\xi^2_{\mathrm{CPO}}\rangle. \tag{7.31}$$

Notice by comparing equation (7.27) with equations (7.30) and (7.31), neglecting terms higher than order two, one has simply:

$$\xi^*_{\mathrm{CPO}} \times \xi_{\mathrm{SPO}} \approx \langle\xi_{\mathrm{CPO}}\rangle - \frac{2}{(2 + \langle\xi_{\mathrm{CPO}}\rangle)^2}\langle\delta\xi^2_{\mathrm{CPO}}\rangle + \frac{\langle\xi_{\mathrm{CPO}}\rangle}{\langle\mu\rangle^2}\langle\delta\mu^2\rangle \approx \xi^* \tag{7.32}$$

In practice, ξ^* can be inferred from a seismic tomography model, whereas ξ^*_{CPO} is overwhelmingly difficult to estimate without access to any elastic homogenization tools. For this reason, seismologists often compare ξ^* with the intrinsic radial anisotropy ξ_{CPO} computed from a CPO model rather than to the homogenized model they could at best obtain. Furthermore, equations (7.27) and (7.32) suggest that there is a non-negligible spurious component of anisotropy due to the unresolved small-scale isotropic heterogeneities. Therefore while it is difficult to rigorously establish analytical solutions in the case of a 2-D/3-D complex media, following the logic above, we hypothesize that the mismatch often observed between homogenized CPO models and tomographic models is the extrinsic radial anisotropy ξ_{SPO}. The full effective radial anisotropy can be quantified through the following composite law:

$$\xi^* \approx \xi^*_{\mathrm{CPO}} \times \xi_{\mathrm{SPO}}. \tag{7.33}$$

7.4.5 Homogenization in 2-D and in 3-D media

In 2-D and in 3-D media, one is compelled to solve the so-called periodic cell problem. This however is only applicable to media exhibiting spatial periodicity. Here, it is still feasible to analytically derive upscaling relations but only in the case of a layered media. For complex media, deriving analytical solutions would be impractical. One resorts to numerical methods involving finite element schemes to tackle the periodic cell problem. The true Earth however may display irregularities, exhibit minimal natural scale separation, or pose any kind of spatial statistical invariance. This precludes the use of any homogenization technique not limited to the ones mentioned above

to upscale such a complex media. To alleviate this problem, Capdeville and Marigo (2007), Capdeville, Guillot, and Marigo (2010), Guillot, Capdeville, and Marigo (2010), and Capdeville, Zhao, and Cupillard (2015) developed a non-periodic homogenization method. So far, this method serves as a pre-processing step by enabling one to solve the elastostatic wave equation using a simple mesh, speeding up the computations for wave propagation in complex media (Capdeville and Marigo, 2007; Guillot, Capdeville, and Marigo, 2010; Capdeville, Zhao, and Cupillard, 2015). It also has been extensively utilized for the discovery of a specific class of stable and fine-scale models of the mantle (Alder et al., 2017).

Assuming perfect data coverage, Capdeville and Métivier (2018) numerically verified that for any elastic medium, a tomographic model is equivalent to a homogenized model. Homogenization can be viewed as a first-order tomographic operator, that is, when applied to an elastic medium, the homogenized model is considered to be the best image one could get from tomography. This can be translated to:

$$\mathbf{S}^* = \mathcal{H}(\mathbf{S}) \tag{7.34}$$

where $\mathcal{H}$ is the tomographic operator, $\mathbf{S}$ is the reference medium, and the homogenized model $\mathbf{S}^*$ is the full effective medium (*i.e.*, the best resulting image as seen by a wavefield of a given maximum frequency f_{max} assuming perfect data coverage). In this chapter, we extend the 'tomographic operator' hypothesis to a 2-D composite medium by upscaling the marble cake model in the presence of deformation-induced anisotropy. We compute the full effective medium $\mathbf{S}^*$ using the Fast Fourier homogenization algorithm developed by Capdeville, Zhao, and Cupillard (2015).

7.5 Methodology

7.5.1 Elastic isotropy in a 2-D mechanically-mixed mantle

To define our 2-D incompressible flow model imitating mantle convection, we use a stream function similar to that of Alder et al. (2017):

$$\Psi(x,z,t) = \sin(a\pi z)\Big[\sin(b\pi x) + \alpha(t)\sin((b+1)\pi x) + \beta(t)\sin((b+2)\pi x)\Big] \tag{7.35}$$

where $\alpha(t)$ and $\beta(t)$ are sinusoidal functions of time that introduces chaotic mixing. The variables a and b relate to the number of distinguishable convection cells and are chosen arbitrarily. The velocity field components are computed using the following

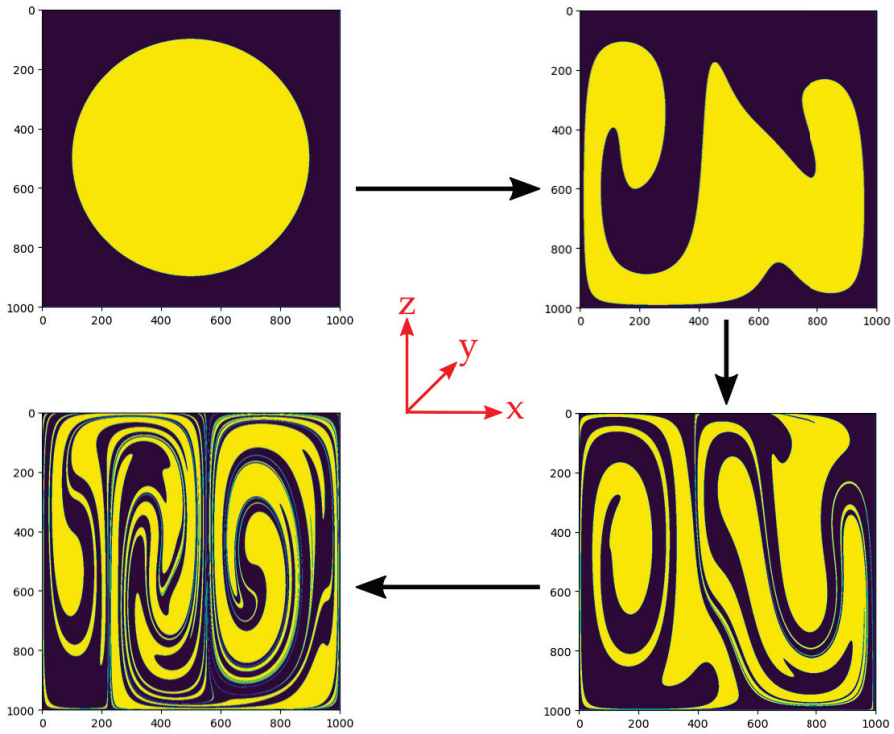

FIGURE 7.3: Initially a circle, the anomaly is deformed progressively until the medium reaches a stage resembling marble cake-like patterns.

expressions:

$$u_x = \frac{\partial \Psi}{\partial z},\ u_z = -\frac{\partial \Psi}{\partial x}\ . \tag{7.36}$$

The form of the function Ψ ensures free-slip boundary conditions. Finally, the velocity field is scaled using a reference value of 1 cm·yr^{-1}.

We replicate the marble cake patterns by deforming a circular anomaly at the center of the box using our prescribed flow field. To do this, control points that define the contour of the anomaly are advected using fourth-order Runge Kutta methods with variable time-stepping (Press et al., 1992). Figure 7.3 illustrates the evolution of the pattern when subjected to the flow field defined in equation (7.35). Setting $a = 1$ and $b = 2$, we have a well-mixed medium with two characteristic convection cells.

Using the last panel of Figure 7.3, the binary system is defined by assigning a reference S-wave velocity value $V_{S_2} = 4.52$ km·s^{-1} to the yellow region, and $V_{S_1} = 3.7$ km·s^{-1} to the blue region so that $100\% \times (V_{S_1} - V_{S_2})/(V_{S_1} + V_{S_2})) = 10\%$. $P-$wave velocities are computed by imposing a constant ratio $V_P/V_S = 1.7$ (Obrebski et al., 2010). Following the work of Tkalčić et al. (2006), we compute the density ρ using the empirical relation $\rho = 2.35 + 0.036(V_P - 3)^2$. To fully describe the elastic medium, one needs to define the local isotropic tensor $\mathbf{S_I}$. This requires the computation of the bulk K and

the shear μ moduli:

$$K = \rho(V_P^2 - V_S^2/3) \tag{7.37}$$

$$\mu = \rho V_S^2 \, . \tag{7.38}$$

The isotropic tensor at a given point is therefore:

$$\mathbf{S_I} = \begin{pmatrix} K+4\mu/3 & K-2\mu/3 & K-2\mu/3 & 0 & 0 & 0 \\ K-2\mu/3 & K+4\mu/3 & K-2\mu/3 & 0 & 0 & 0 \\ K-2\mu/3 & K-2\mu/3 & K+4\mu/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu \end{pmatrix} \tag{7.39}$$

7.5.2 Kinematic modeling of Crystallographic Preferred Orientation

Crystallographic preferred orientation evolves due to strain accumulation over time. This requires the knowledge of the macroscopic velocity gradient tensor $\boldsymbol{\nabla}\mathbf{u}$ driving the net alignment of the crystals within an aggregate. We model CPO evolution using `D-Rex`, a program which calculates strain-induced crystallographic preferred orientation of olivine-enstatite aggregates by plastic deformation, dynamic recrystallization, and grain-boundary sliding. The code inputs the externally-imposed velocity gradient tensor $\boldsymbol{\nabla}\mathbf{u}$ which can be easily derived from the stream function Ψ, the number of crystals N_{crystal}, the amount of olivine and enstatite, and a random initial orientation of the individual grains making up the aggregate. Apart from this, it inputs four other parameters that are already constrained from laboratory experiments (Zhang and Karato, 1995; Bystricky et al., 2000) namely: the activity of the slip systems of olivine and enstatite, a dimensionless grain boundary mobility, a dimensionless nucleation rate, and a dimensionless threshold volume for grain boundary sliding. The main output is a Voigt-averaged elastic tensor with 21 independent coefficients at every location. Users also have the option to output the best-fit hexagonally-symmetric medium calculated from the elastic decomposition method of Browaeys and Chevrot (2004). `D-Rex` has been extensively applied in the regional (Lassak et al., 2006; Miller and Becker, 2012; Faccenda and Capitanio, 2013) and at the global scale (Becker et al., 2006; Becker, Kustowski, and Ekström, 2008).

In our numerical experiments, we compute CPO everywhere (*i.e.*, in both yellow and blue phases of the mantle model). We consider 100% olivine of type-A crystal fabric corresponding to dry upper-mantle conditions. We scale the elastic tensor derived

from D-Rex so that its isotropic component is built from the binary system derived in Section 7.5.1. The reference medium can be constructed from equation (7.1) where $\mathbf{S_I}$ now relates to the small-scale isotropic heterogeneities in the mechanically-mixed mantle, and $\mathbf{S_A}$ is the intrinsically anisotropic component computed with D-Rex.

7.5.3 Fast-Fourier Homogenization

We use the 3-D Fast-Fourier Homogenization algorithm developed by Capdeville, Zhao, and Cupillard (2015). It inputs the local elastic tensors and the local densities of the reference medium and relies on the extensive usage of Fast-Fourier Transforms. The code also requires the scale-separation parameter ε_0 as an input. For now, the value is chosen ad-hoc and has been kept constant throughout the numerical experiments. For the output, it provides the local elastic tensor and the local density at any arbitrary point of the effective medium.

In this chapter, we implement the Fast-Fourier Homogenization algorithm to obtain the long-wavelength effective equivalent of the reference medium $\mathbf{S}$. We also homogenize its constituents, $\mathcal{I}(\mathbf{S})$ and $\mathcal{A}(\mathbf{S})$ separately to analyze their individual contributions to the full effective anisotropy.

7.5.4 Quantifying the level of anisotropy

In this section, we quantify the level of seismic anisotropy for any given elastic tensor $\mathbf{S}$. Since the 3-D elastic tensor contains 21 independent coefficients, quantification of anisotropy necessitates the elastic projection of the tensor in a specific symmetry class. In our case, we use the Montagner and Nataf (1986) projection where we treat the projected tensor as an azimuthally-averaged vertically transverse isotropic (VTI) medium. The Love parameters (Love, 1906) of a hexagonally symmetric medium can be regarded as a linear combination of the 21-component elastic tensor (Montagner and Nataf, 1986). The parameters L and N, which relate to the azimuthally-averaged vertically- and horizontally-polarized $S-$waves, can be computed as follows:

$$L = \frac{1}{2}(S_{44} + S_{55}) \tag{7.40}$$

$$N = \frac{1}{8}(S_{11} + S_{22}) - \frac{1}{4}S_{12} + \frac{1}{2}S_{66}. \tag{7.41}$$

The level of radial anisotropy is then given by equation (7.6). Anisotropic inversion studies generally interpret $\xi > 1$ as horizontal flow, and $\xi < 1$ as vertical flow (refer to Montagner (1994) and Bodin et al. (2015) for a comprehensive review).

Another convenient way to quantify anisotropy is to compute the percentage of total anisotropy by taking the norm fraction of the anisotropic part of the elastic tensor with respect to the isotropic part. This however entails the extraction of the isotropic part of the elastic tensor. To do this, Browaeys and Chevrot (2004) proposed a decomposition method for the isotropic tensor in terms of the dilatational **d** and Voigt stiffness **v** tensors derived from **S** (Cowin and Mehrabadi, 1987):

$$\mathbf{d} = \begin{pmatrix} S_{11}+S_{12}+S_{13} & S_{16}+S_{26}+S_{36} & S_{15}+S_{25}+S_{35} \\ S_{16}+S_{26}+S_{36} & S_{12}+S_{22}+S_{32} & S_{14}+S_{24}+S_{34} \\ S_{15}+S_{25}+S_{35} & S_{14}+S_{24}+S_{34} & S_{13}+S_{23}+S_{33} \end{pmatrix}, \tag{7.42}$$

and

$$\mathbf{v} = \begin{pmatrix} S_{11}+S_{66}+S_{55} & S_{16}+S_{26}+S_{45} & S_{15}+S_{35}+S_{46} \\ S_{16}+S_{26}+S_{45} & S_{66}+S_{22}+S_{44} & S_{24}+S_{34}+S_{56} \\ S_{15}+S_{35}+S_{46} & S_{24}+S_{34}+S_{56} & S_{55}+S_{44}+S_{33} \end{pmatrix}. \tag{7.43}$$

where the bulk K and shear moduli μ can be expressed in terms of the traces of the dilatational and Voigt tensors as (Fedorov, 2013):

$$K = \frac{\mathrm{tr}(\mathbf{d})}{9}, \tag{7.44}$$

and

$$\mu = \frac{3\mathrm{tr}(\mathbf{v}) - \mathrm{tr}(\mathbf{d})}{30}. \tag{7.45}$$

The isotropic tensor $\mathbf{S_I}$ can be computed in terms of μ and K following equation (7.39). The total anisotropy is given by:

$$a_{\mathrm{tot}} = \frac{||\mathbf{S} - \mathbf{S_I}||_2}{||\mathbf{S_I}||_2} \tag{7.46}$$

where a_{tot} is the tensor norm fraction of the total anisotropy with respect to $\mathbf{S_I}$.

7.6 Elastic homogenization of a 2-D mechanically-mixed mantle in the presence of CPO

Figure 7.4 displays some seismic properties of the reference medium **S** before and after homogenization. The left panels are the true structures, whereas the middle and right panels are the structures equating to the full effective medium $\mathcal{H}(\mathbf{S})$ at homogenization wavelengths λ_h of 200 km and 500 km, respectively. The first row depicts the shear

wave velocities. The anisotropy characterized by its radial component, and by its norm fraction are depicted in the second row and in the third row, respectively. Each pixel initially contains an isotropic part derived from the marble cake model with a mixing time for advection $T_{\text{SPO}} \sim 75$ My , and an anisotropic part computed from a CPO model with a time scale for CPO evolution of $T_{\text{CPO}} \sim 40$ My. Hereafter, we define $\lambda_{\text{max}} = 1000$ km as the length scale of our simulations.

As expected, homogenization results in the smoothing of the structures with the level modulated by λ_h. Several glaring features can also be observed such as the presence of positive radial anisotropy ($\xi > 1$) at the top and bottom boundaries where we expect the flow to be sub-horizontal, and likewise negative ($\xi < 1$) at regions where the flow is expected to be sub-vertical. In addition, the resolution of the small-scales is diminished everywhere (top left panels) and the full effective medium is devoid of anisotropy at some patches (bottom left panels) post-homogenization. This implies that homogenization is not just a simple spatial average but a product of highly non-linear upscaling relations.

By decomposing $\mathbf{S}$ into an isotropic tensor $\mathcal{I}(\mathbf{S})$ and an anisotropic tensor $\mathcal{A}(\mathbf{S})$ through equations (7.2) and (7.4), respectively, one could also homogenize each tensor and analyze its effects. Figure 7.5 shows the level of effective radial anisotropy of the two separate components after homogenization. The top panels recreate the results of Alder et al. (2017). Indeed, homogenizing a fine-layered medium produces extrinsic radial anisotropy ξ_{SPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{I}(\mathbf{S}))$). Notice that the effective intrinsic radial anisotropy and the extrinsic radial anisotropy maps are roughly similar as they both induced a positive radial anisotropy $\xi > 1$ in the horizontal layers (*i.e.*, the stretched heterogeneities that induce SPO become elongated along the direction of the maximum principal strain rate that also controls the CPO). As for the homogenization of $\mathcal{A}(\mathbf{S})$, it is noteworthy that analogous as to how a purely isotropic medium produces extrinsic anisotropy, a purely anisotropic medium (*i.e.*, a model where only the anisotropic component varies with space) synthesizes spurious isotropic heterogeneities upon homogenization (See Figure 7.6). It however appears to be small with maximum artificial velocity perturbations of about 0.25 % at λ_h = 200 km and 0.2 % at λ_h = 500 km considering the severity of intrinsic anisotropy produced from our CPO experiments (*i.e.*, 100 % A-type olivine with an evolution time scale of ∼40 My). We also find that the effective intrinsic radial anisotropy ξ^*_{CPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{A}(\mathbf{S}))$ is slightly underestimated compared to ξ_{CPO} of the previous figure. This is better illustrated in Figure 7.7 where we plot the standard deviation of radial anisotropy of the entire 2-D model domain against λ_h.

Some essential information can be extracted from Figure 7.7.

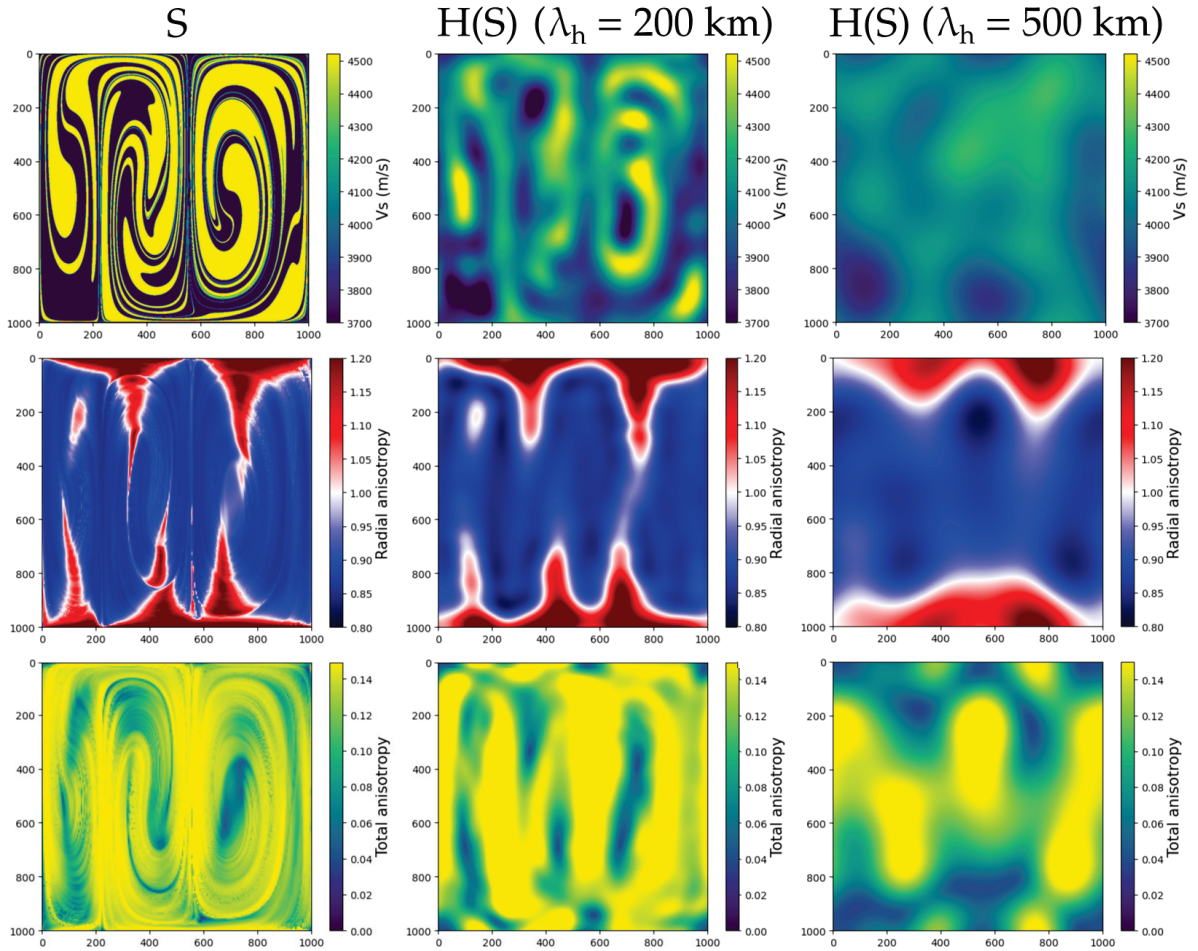

FIGURE 7.4: Some of the seismic properties of the reference medium **S** before and after homogenization. The model dimensions are 1000 km × 1000 km × 1000 km, with cell sizes of 5 km × 5 km × 5 km. Here, each pixel contains an **S** which consists of small-scale isotropic heterogeneities and an intrinsically anisotropic perturbation computed with D-Rex (Kaminski, Ribe, and Browaeys, 2004). The present-day marble cake patterns correspond to a mixing time for advection $T_{\mathrm{SPO}} \sim 75$ My, whereas the time scale for CPO evolution is $T_{\mathrm{CPO}} \sim 40$ My. We homogenized **S** using the Fast-Fourier homogenization algorithm of Capdeville, Zhao, and Cupillard (2015). (From left to right) First row: V_s models derived from **S**, $\mathcal{H}(\mathbf{S})$ at $\lambda_h = 200$ km, and $\mathcal{H}(\mathbf{S})$ at $\lambda_h = 500$ km. Second row: ξ_{CPO}, ξ^* at $\lambda_h = 200$ km, and ξ^* at $\lambda_h = 500$ km. Last row: Total anisotropy in terms of the norm fraction of **S**, $\mathcal{H}(\mathbf{S})$ at $\lambda_h = 200$ km, and $\mathcal{H}(\mathbf{S})$ at $\lambda_h = 500$ km. Elastic homogenization can be viewed as the best possible model resolved by seismic tomography assuming perfect ray-path coverage.

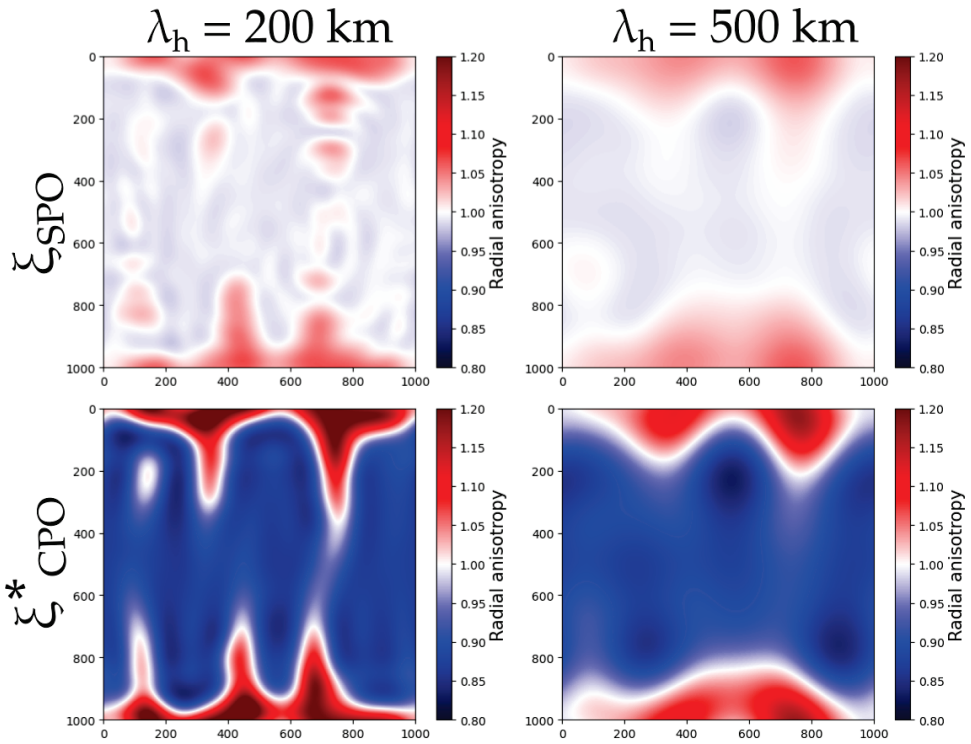

FIGURE 7.5: Extrinsic radial anisotropy ξ_{SPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{I}(\mathbf{S}))$) (top panels) at two different wavelengths of homogenization λ_h. Here, $\xi_{\text{SPO}} > 1$ is now interpreted as vertical layering whereas < 1 as horizontal layering. The bottom panels are the effective intrinsic radial anisotropy ξ^*_{CPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{A}(\mathbf{S}))$).

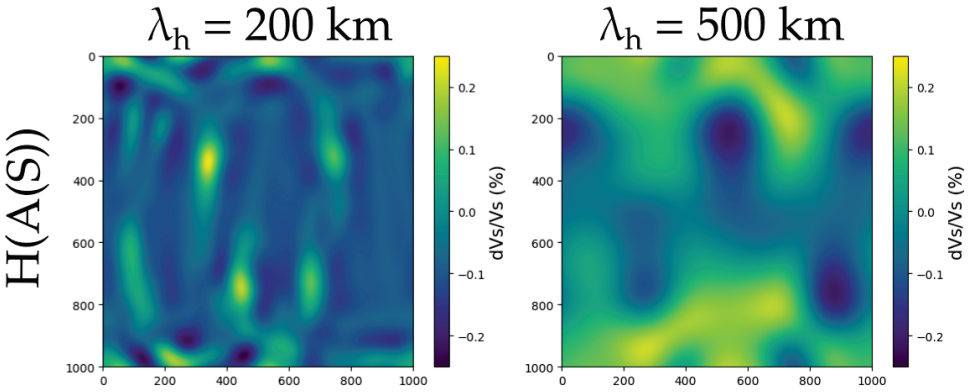

FIGURE 7.6: Spurious isotropic velocity perturbations with respect to a mean velocity V_S at two different wavelengths of homogenization λ_h. $\mathcal{H}(\mathcal{A}(\mathbf{S}))$ pertains to the homogenized model of a purely anisotropic medium. Even when placed in a very favorable scenario for intrinsic anisotropy, homogenizing a purely anisotropic medium produces a meager 0.25% artificial heterogeneities at $\lambda_h = 200$ km and 0.2% at $\lambda_h = 500$ km. When superimposed with the bottom left panels of Figure 7.4, it appears that spurious isotropy fills the void in anisotropy.

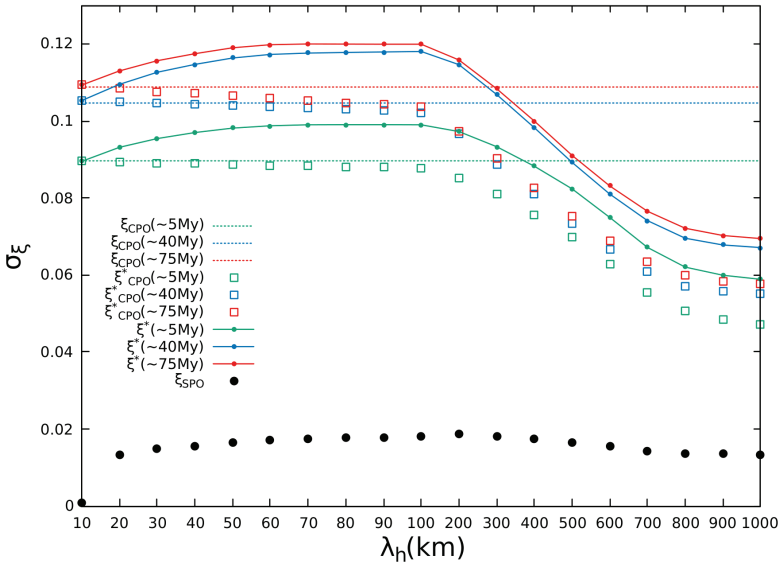

FIGURE 7.7: Level of effective radial anisotropy as a function of the wavelength of homogenization. Here, we quantified the effective radial anisotropy in terms of its standard deviation σ_{ξ} over the entire 2-D image. The time scales indicated in million years pertain to the evolution history of CPO (*i.e.*, larger time scale means well-developed CPO). The dashed lines represent the standard deviation of ξ_{CPO} (*i.e.*, radial anisotropy of model $\mathbf{S}$) and serve as reference values. In this experiment, ξ_{SPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{I}(\mathbf{S}))$) (black circles) deemed to be five times smaller than ξ^*_{CPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{A}(\mathbf{S}))$) (hollow squares), and since SPO is mostly in-phase with CPO, the two anisotropic components add constructively giving the full effective radial anisotropy ξ^* (solid line-dots).

(i) First, we compute the intrinsic anisotropy component of the reference medium **S** using `D-Rex` assuming that the CPO develops over a time T_{CPO} of 5, 40 or 75 Myr. The resulting intrinsic radial anisotropy ξ_{CPO} in terms of its standard deviation over the entire convecting box is reported with dashed horizontal lines. Not surprisingly, the overall anisotropy increases with T_{CPO} although some saturation is observed (*i.e.*, the orientation of crystals depends mostly on their recent deformation, and lose the memory of the deformation they underwent too long ago).

(ii) This anisotropy, if it were not associated with velocity anomalies could be recovered from seismic tomography as an effective intrinsic radial anisotropy ξ^*_{CPO} (squares) which is always lower than the reference value ξ_{CPO}. This is in qualitative agreement with equation (7.31).

(iii) On the contrary, the full effective radial anisotropy ξ^* at short wavelengths of homogenization λ_h are larger than ξ_{CPO}. This again is in agreement with the analytical expression given by equation (7.27). This additional anisotropy is of course due to the existence of SPO (black circles) which reinforces the total level of effective anisotropy.

(iv) Both ξ^*_{CPO} and ξ^* converge toward ξ_{CPO} at infinitely short homogenization wavelengths. Only in this unrealistic case, would seismologists be able to map the intrinsic anisotropy without smoothing or addition of spurious SPO. At λ_h larger than the typical size of the anomalies (of order 100 km for the stripes of Fig 7.3), the recoverable anisotropies rapidly decrease.

(v) Extrinsic radial anisotropy ξ_{SPO} here is much lower than ξ^*_{CPO}. For a well-developed CPO fabric ($\sim$75 My), this amounts to $\approx$ 10.5% for λ_h = 100 km. This is five times bigger than ξ_{SPO}. Such a result however is specific to this numerical experiment, and that CPO is indeed stronger than SPO might not be always true. A longer stirring time (here T_{SPO} $\sim$75 My) would have resulted in thinner and more complex layering that would have increased the SPO. We are unfortunately limited by the number of tracers necessary to describe the phase stirring which is exponentially increasing with time.

7.6.1 Effect of homogenization to the preferred orientation of crystals

We now explore as to why the effective intrinsic anisotropy is always underestimated, or the fact that the full effective total anisotropy appears to be diminished in some areas. Figure 7.8 depicts the best-fit hexagonally symmetric medium to **S** (left panel) and the effective intrinsic total anisotropy (*i.e.*, total amount of anisotropy in $\mathcal{H}(\mathcal{A}(\mathbf{S}))$) (top right panels) and the difference in the angles of the symmetry axis of a hexagonally projected **S** and $\mathcal{H}(\mathcal{A}(\mathbf{S}))$ (bottom left panels) at two different homogenization wavelengths. Regions of high misfit in the symmetry angles (Figure 7.8 bottom right

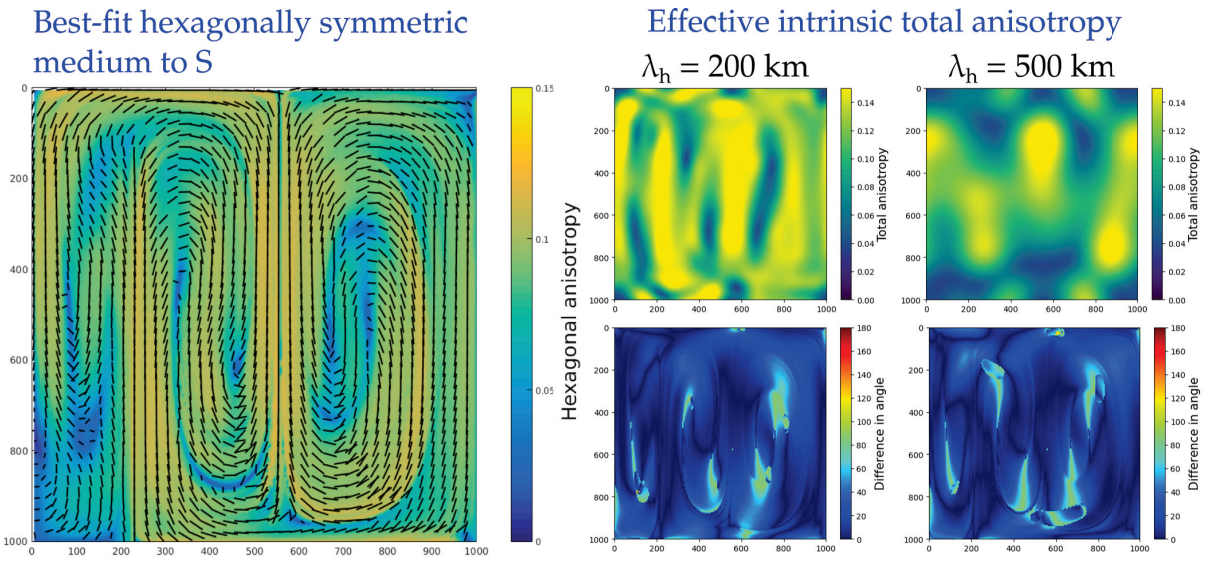

FIGURE 7.8: (Left panel) Amplitude and symmetry axis orientation (solid black lines) of the best-fit hexagonally symmetric medium to **S**. We use the elastic decomposition method of Browaeys and Chevrot (2004) to arrive at such a medium. The symmetry axis maps well regions of high hexagonal anisotropy associated with strong shear deformation. It is often used as a proxy to identify the fast axis of olivine which appears to be also in agreement with the shear direction. (Top right panels) Effective intrinsic total anisotropy (*i.e.*, total amount of anisotropy in $\mathcal{H}(\mathcal{A}(\mathbf{S}))$) at two different wavelengths of homogenization. (Bottom right panels) Difference in the angle of the symmetry axis of a hexagonally-projected **S** and $\mathcal{H}(\mathcal{A}(\mathbf{S}))$. Regions of high-misfit in the angle correlates well with low amplitudes of anisotropy.

panels) correlate well with regions of low effective intrinsic total anisotropy (top right panels). Other features can also be noticed such as the dilution of the effective intrinsic total anisotropy at [600 km, 800 km] (Figure 7.8 top right panels). When juxtaposed with the left panel of Figure 7.8, these are areas where the symmetry axis appear to be misaligned compared to their neighbors. This implies that localized preferential orientations cannot be discerned by a wavefield of λ_h = 500 km. On the other hand, shorter period wavefields as exemplified by λ_h = 200 km preserve the anisotropy at [600 km, 800 km] since these waves are capable of delineating the preferential orientations of the left panel at the same location.

In theory, $\lambda_h \to 0$ km localizes the observations down to the crystal level where we can pinpoint the net crystallographic orientation, thus converging towards the original CPO model where the amplitudes of intrinsic anisotropy are at a maximum. Figure 7.9 reaffirms this hypothesis where we plot ξ^*_{CPO} in terms of σ_ξ versus $\lambda_h/\lambda_{\text{max}}$ for different length scales of convection λ_c (*i.e.*, λ_c = 500 km corresponds to mantle flow with two characteristic convection cells, 333.33 km for three convection cells, and so on). Let

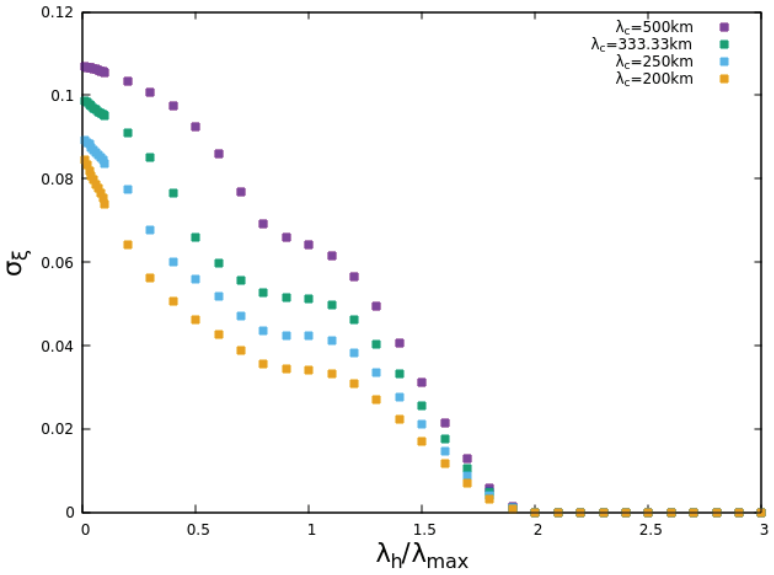

FIGURE 7.9: Level of effective intrinsic radial anisotropy ξ^*_{CPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{A}(\mathbf{S}))$ expressed as σ_ξ of the entire 2-D model of ξ^*_{CPO} versus wavelength of homogenization. Each plot pertains to one family of convection patterns with a given length scale λ_c. Here, λ_c = 500 km is for the 2-cell pattern, 300 km for the 3-cell pattern, and so on. The propensity of the effective intrinsic anisotropy to amplitude loss appears to be well-correlated with increasing complexity in the convection patterns.

us put our attention to the trend of σ_ξ when $\lambda_h/\lambda_{\mathrm{max}} \leq 0.5$. The most salient characteristic is that the slope generally decreases with increasing complexity of the convection patterns. This is a clear indication that the preferential orientations indeed average out. When $0.5 < \lambda_h/\lambda_{\mathrm{max}} < 1.75$, the slope decreases with decreasing complexity. This is because at longer wavelengths of homogenization, the wavefield 'sees' a net vertical structure for the 5-cell pattern whereas orientations from the 2-cell pattern tend to cancel out more since both horizontal and vertical orientations are equally prominent. In a case where the convection cells are stacked vertically, we expect this same behavior. At $\lambda_h/\lambda_{\mathrm{max}} > 1.75$, the wavefields almost fail to render any preferential orientations which is why the effective intrinsic radial anisotropy eventually stagnates at a value of zero.

7.6.2 Verifying the composite law $\xi^* = \xi^*_{\mathrm{CPO}} \times \xi_{\mathrm{SPO}}$ in 2-D

CPO versus homogenized CPO

Earlier we have shown that in the 1-D case, ξ_{SPO} can be directly estimated from ξ^* inferred from a tomographic model and from ξ^*_{CPO} computed from a homogenized CPO

model assuming CPO and SPO are uncorrelated. To verify whether equation (7.33) still holds in 2-D, we plot for each pixel in our 2-D maps a point containing ξ_{SPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{I}(\mathbf{S}))$) $\times$ ξ^*_{CPO} (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathcal{A}(\mathbf{S}))$) against ξ^* (*i.e.*, radial anisotropy of model $\mathcal{H}(\mathbf{S})$) for the $x-$ and $y-$coordinates, respectively. Note that in this demonstration, CPO is computed everywhere and hence, the effect of the cross term is mitigated. Figure 7.10b shows just this for different homogenization wavelengths λ_h. As expected, the relation holds exceptionally well even when λ_h is large.

In practice however, seismologists and geodynamicists alike painstakingly compare tomographic models of ξ^* directly with ξ_{CPO} computed from CPO models. We mimic a conventional scenario where a homogenized CPO model is unavailable by comparing $\xi_{\text{SPO}} \times \xi_{\text{CPO}}$ instead with ξ^* (Figure 7.10a). As it turns out, the relation only holds for small values of λ_h. This is expected since as stated earlier, full effective anisotropy converges towards intrinsic anisotropy when $\lambda_h \to 0$ since not only do the short period wavefields delineate the geographically-localized CPO, but also extrinsic anisotropy eventually vanishes. At larger values of λ_h, the trend appears to be more dispersed as a consequence of the averaging process, losing its viability to some extent. In the absence of a homogenized CPO model, we project that this composite law would remain true in general under the condition that the minimum wavelength used in tomography is sufficiently small.

Effect of the rigidity-anisotropy cross term

To test the impact of the rigidity-anisotropy cross term, we consider another mantle model where CPO is confined in one of the colored regions of the 2-D marble cake illustrated in Figure 7.3. In addition, we increase the percentage of isotropic heterogeneities in V_S to 15%. In this way, we indirectly amplify its effect. Figure 7.10c displays the numerical solution at $\lambda_h = 50$ km and 200 km when CPO is computed in the yellow phases alone. Based on our analytical results, the points are much more spread-out than that of Figure 7.10b. In this scenario, CPO now varies abruptly and appears heterogeneous locally (*i.e.*, $\delta\xi_{\text{CPO}}$ terms are now much larger), and as expected the cross term is much more apparent. Nonetheless, it still exhibits strong linearity implying that the predictions carried out by the composite law are robust.

Let us now explore what happens when we administer CPO in the blue phase and in the yellow phase separately. Neglecting higher-order terms, one could numerically compute for the cross terms by simply taking the residual radial anisotropy ε_ξ from $\varepsilon_\xi = \xi^* - \xi^*_{\text{CPO}} \times \xi_{\text{SPO}}$ at each pixel in the 2-D maps of radial anisotropy. Figure 7.11

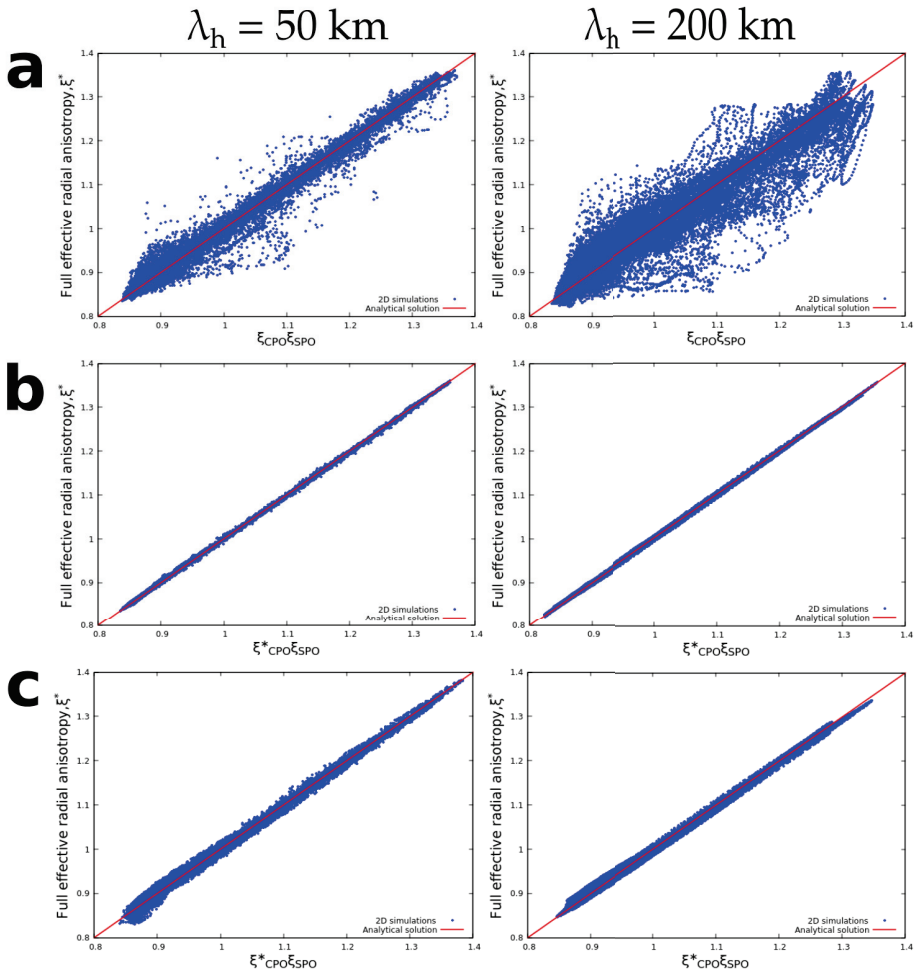

FIGURE 7.10: For each pixel i on our 2-D maps of ξ^*, ξ_{CPO}, ξ^*_{CPO}, and ξ_{SPO} Figure 7.10a: , we plot a cloud of points with coordinates: (a) [$\xi_{CPO} \times \xi_{SPO}(i)$, $\xi^*(i)$], (b) [$\xi^*_{CPO} \times \xi_{SPO}(i)$, $\xi^*(i)$] overlaying the analytical solution equation (7.33) (solid red lines). Our numerical simulations suggest that CPO models should be homogenized first before comparing with tomographic models. Figure 7.10c displays the effect of the cross term by increasing isotropic heterogeneities to 15% and by prescribing CPO only in the yellow phases of the marble cake model in Figure 7.3.

displays 2-D maps of ξ^*, ξ^*_{CPO}, ξ_{SPO}, and ε_ξ homogenized at $\lambda_h = 50$ km for a mantle model where CPO is confined in the yellow regions (top panels) and in the blue regions (bottom panels). Firstly, the cross terms appear to change sign depending on where CPO is prescribed which is in agreement with equation (7.27). One one hand when CPO exists only in the yellow regions, the cross terms are negative and tend to diminish anisotropy (*i.e.*, ε_ξ becomes more negative across horizontal laminations and horizontal flow direction, and more positive across the vertical), implying that the shear modulus and intrinsic anisotropy are anti-correlated. On the other hand when CPO exists only in the blue regions, the opposite effect is observed implying that now, they are positively correlated. Secondly, zones where $\varepsilon_\xi \sim 0$ correlates well with zones where $\xi_{\mathrm{SPO}} \sim 1$. Such a result is expected. Because the filaments in these zones are comparable to the size of λ_h, they appear homogeneously isotropic, and hence do not produce any extrinsic anisotropy nor cross terms since both depend on small-scale isotropic heterogeneities to exist. Finally when the two ε_ξ maps are merged, the result is close to zero. This is because the magnitudes of the cross terms in both configurations (*i.e.*, CPO in blue or CPO in yellow) are approximately the same, and that the distinction between them is determined solely by their signs.

7.6.3 Discussion

We extended the work of Alder et al. (2017) in that we primarily investigated the effects of elastic homogenization to a specific class of stable, and fine-scale models of the mantle in the presence of deformation-induced anisotropy. We homogenized our mantle models using the 3-D Fast-Fourier homogenization algorithm developed by Capdeville, Zhao, and Cupillard (2015). The homogenization procedure can be viewed as a tomographic operator applied to a reference elastic model (Capdeville et al., 2013). The result is the best possible image one might recover from seismic tomography assuming perfect data coverage.

According to our experiments, we showed that the extrinsic radial anisotropy produced by fine-layering could reach up to 2% (see Figure 7.7). This is much lower than those induced by CPO where the effective intrinsic radial anisotropy could peak at nearly 11%. This result however is restricted by some parameters that regulate the level of effective anisotropy. For example, the layered filaments produced by our marble cake models are of the order $10 - 100$ km whereas of those proposed by Allègre and Turcotte (1986) are much thinner and can stretch even further down to the centimeter scale. By increasing the mixing time for advection T_{SPO} in our simulations, we expect the filaments to be more elongated and in turn, induce larger extrinsic anisotropy.

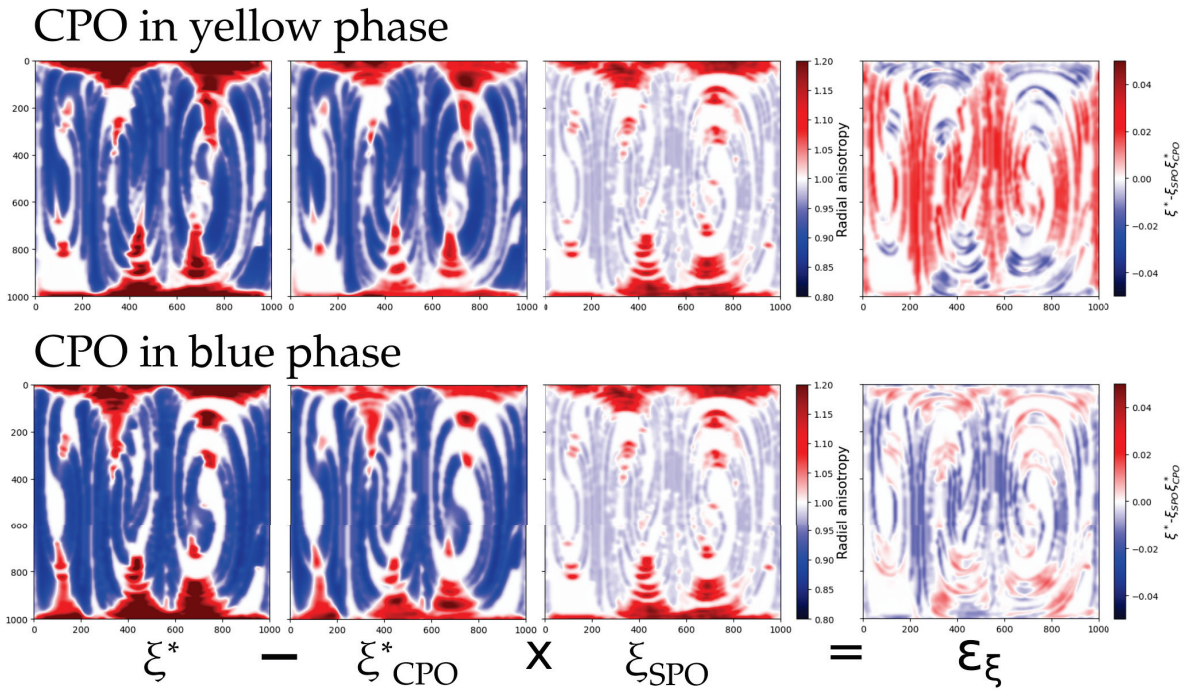

FIGURE 7.11: Quantification of the coupling between the shear modulus and intrinsic anisotropy in terms of the residual radial anisotropy ε_ξ computed from $\varepsilon_\xi = \xi^* - \xi^*_{\text{CPO}}\xi_{\text{SPO}}$ at each pixel in the 2-D maps of radial anisotropy under two scenarios: CPO is only computed in (1) the yellow phases (top panels), and (2) the blue phases (bottom panels). The reference 2-D maps $\mathbf{S}$, $\mathcal{A}(\mathbf{S})$, and $\mathcal{I}(\mathbf{S})$ are homogenized at λ_h= 50 km to obtain ξ^*, ξ^*_{CPO}, and ξ_{SPO}, respectively. Based on the ε_ξ maps, shear modulus and intrinsic anisotropy are anti-correlated in (1), and positively correlated in (2). In both cases, the magnitudes of the cross terms are essentially equivalent, and only differ by their signs.

Another factor that needs to be mentioned is the spatial roughness of the model domain. Indeed that models constructed using finer grid would exhibit more extrinsic anisotropy than with models contrived from coarse parameterizations. Last is more of on the intrinsic aspect. Since CPO results from finite strain accumulation over time, the amplitude of intrinsic anisotropy increases with the time scale for CPO evolution T_{CPO}, although recrystallization and damage would limit the CPO that can be eventually accumulated (Ricard and Bercovici, 2009). Such presumptions may only be valid in regions where rock deformation varies over extended periods of time. Therefore, whether CPO accounts for most of the bulk anisotropy observed in tomographic images remains inconclusive and needs further verification.

In light of the simulations conducted, we expect large-scale anisotropy to be only overestimated when CPO coexists with SPO at some point as exemplified in our simulations. In the absence of SPO, homogenization can only decrease the strength of anisotropy. By accounting for both contributions, we showed that $\xi > 1$ is attributed to a combination of flow ascent and horizontal layering, and $\xi < 1$ is a combination of lateral flow and vertical layering. Indeed, the direction of shear not only dictates the net orientation of the anisotropic minerals, but also of the orientation of the folded strips that gives rise to fine-layering.

The repercussion of homogenizing intrinsic anisotropy alone amounts to the convection-scale averaging of the CPO as evidenced by our simulations. By definition, intrinsic anisotropy develops from the net alignment of the minerals in the crystal level. This explains why we can map-out the fast axis of olivine at an arbitrary location. When long period observations sample an intrinsically anisotropic medium, the wavefield unknowingly spatially-averages these geographically localized orientations. As a result, crystallographic orientations that are products of imbricated convection tend to be more disoriented within their vicinity, thereby ostensibly losing effective intrinsic anisotropy in the process. In contrast, spatially-coherent crystallographic orientations that are produced by simpler convection patterns are less susceptible to the dilution of effective intrinsic anisotropy when sampled by long period observations.

The applicability of equation (7.33) in a 2-D complex media as demonstrated may be of interest to geodynamicists and tomographers alike because not only does it permit one to directly quantify the discrepancy between the full effective radial anisotropy inferred from a tomographic model and the effective intrinsic radial anisotropy computed from a homogenized CPO model, it further solidifies the supposition that the mismatch is indeed a result of extrinsic radial anisotropy due to the seismically-unresolved small-scale isotropic heterogeneities. We have conducted several numerical experiments to prove that the composite law still holds exceptionally well even when the

rigidity-intrinsic anisotropy cross term is amplified. Erroneous interpretation of a CPO model to explain tomographic observations may therefore be avoided upon homogenization.

The conclusions reached in this section are based on a number of simplifying assumptions:

(1) Although the elastic homogenization code used is completely 3-D, our mantle models were only 2D. Hence, any 3-D structural effects that might have thwarted some of the results in this study were completely ignored. A possible extension of this work therefore is to consider 3-D geodynamic flow models of the mantle e.g. Zhong et al., 2000; Samuel, 2012b to accommodate any structural effects.

(2) We held the isotropic velocity contrast at a fixed value and assumed it to be representative of the entire mantle. In reality however, V_S variations between basalt and harzburgite generally decrease with depth (Xu et al., 2008; Stixrude and Jeanloz, 2015). This is not to mention the local presence of melt and water that contributes to the variations in wavespeeds, and hence the strength of heterogeneities which completely alters the level of apparent anisotropy.

(3) We completely disregarded the dependency of the elastic constants built from our mantle models on pressure P and temperature T. Future avenues one could take would be to incorporate $P-T$ dependence using empirical relations constrained from laboratory experiments. For instance, one may compute $P-T$ dependence using first-order corrections around a reference elastic tensor at ambient $P-T$ conditions (Estey and Douglas, 1986). The growing availability of self-consistent thermodynamic models based on free-energy minimization schemes (Connolly, 2005; Connolly, 2009) can also be employed in lieu of the simpler relations for more accurate predictions of seismic wavespeeds in any given bulk composition (Stixrude and Lithgow-Bertelloni, 2011).

(4) We considered olivine of type-A crystal fabric as the solitary anisotropic mineral in our mantle models. Because of this, the intrinsic anisotropy produced from finite deformation should be seen as an upper bound. Inclusion of other anisotropic minerals such as pyroxene which make up a fraction in mantle periodotite (Maupin and Park, 2015) would change the net anisotropy. For instance, we anticipate that including a substantial amount of enstatite would dilute the net anisotropy e.g. Kaminski, Ribe, and Browaeys, 2004 bringing intrinsic anisotropy and extrinsic anisotropy to comparable levels.

7.7 Separating SPO from CPO in tomographic models: Application to radial anisotropy beneath oceanic plates

7.7.1 Radial anisotropy beneath oceanic plates

Within the context of seismic tomography, surface waves offer the capability to image upper-mantle structure providing an in-depth view of large-scale anisotropy. They are sensitive to both azimuthal and radial anisotropy. The depth-distribution of radial anisotropy underneath oceanic basins is positive, $V_{SH} > V_{SV}$, characterized by a layer of strong signatures lying in between $\sim$ 80 km $-$ $\sim$ 250 km corresponding to the asthenosphere e.g. Montagner, 1985; Ekström and Dziewonski, 1998; Panning and Romanowicz, 2006; Nettles and Dziewoński, 2008. Meanwhile, geodynamic models of upper-mantle flow beneath ridges reveal that the flow is primarily horizontal. Coupled with micro-mechanical models of CPO evolution, the patterns of radial anisotropy across this depth-range can be explained by the net alignment of anisotropic aggregates with upper-mantle flow (Becker et al., 2006; Becker, Kustowski, and Ekström, 2008; Becker et al., 2014).

Traversing shallower depths, plate-averaged radial anisotropy across the oceanic lithosphere appears frozen-in and displays modest levels of about $1-3\%$ which cannot be explained by CPO alone (Hansen, Qi, and Warren, 2016). It is proposed that the preservation of quasi-laminated structures due to the transportation of crystallized or unextractable melt during lithospheric thickening and magmatic underplating expounds this frozen-in signature of anisotropy e.g. Auer et al., 2015; Hansen, Qi, and Warren, 2016; Debayle et al., 2020. With this, there is a growing consensus that rock-scale SPO may also be a potential mechanism, and that a substantial fraction of the observed lithospheric anisotropy may be a spurious effect due to small-scale isotropic heterogeneities (Wang et al., 2013; Kennett and Furumura, 2015). However, constraints from geological experiments suggest that the spatial distribution of melt conduits across the oceanic lithosphere, especially beneath slow-spreading ridges, is not perfectly-layered but displays sporadic patterns e.g. Kelemen, Braun, and Hirth, 2000 which would then eventually suppress the level of extrinsic anisotropy by an insurmountable degree (Faccenda et al., 2019). In summary, although asthenospheric anisotropy is well-understood, lithospheric anisotropy still poses ambiguities.

The corroboration of the composite law in a 2-D complex media prompted us to assess the discrepancy between a tomographic model and a CPO model of upper-mantle

radial anisotropy underneath a mid-ocean ridge. In our hypothesis, this should commensurate to the extrinsic radial anisotropy due to the unresolved small-scales in seismic velocities.

7.7.2 The tomographic model

In conjunction with the pre-existing global V_{SV} model of the upper-mantle constrained from Rayleigh wave data `DR2012` (Debayle and Ricard, 2012), we adopt the recent global V_{SH} model `CAM2016SH` of Ho, Priestley, and Debayle (2016) to acquire a plate-averaged 2-D profile of radial anisotropy associated with slow-spreading oceanic ridges. The V_S models were reconstructed by independently inverting Love (for V_{SH} models) and Rayleigh (for V_{SV} models) waveforms up to the fifth overtone between the period range $50-250$ s using an extension of the automated waveform inversion approach of Debayle (1999). We refer the reader to Debayle and Ricard (2012) and Ho, Priestley, and Debayle (2016) for a more detailed description of the inversion procedure.

From the V_{SV} and V_{SH} models of the upper-mantle, we compute the tomographic counterpart of radial anisotropy using $\xi^* = (V_{SH}/V_{SV})^2$. Here, we acknowledge that ξ^* is not directly inferred from simultaneous inversions of Love and Rayleigh data but is a rudimentary estimate from the two $S-$wave velocity models that may conceivably have different qualities. We view this demonstration as a proof-of-concept and therefore one must proceed with caution.

The depth distribution of ξ^* spanning from $35-400$ km is shown in Figure 7.12 (top panel). Positive radial anisotropy values ($\xi^* > 1$) are confined in the upper $\sim$ 200 km of the model domain which is in close agreement with previous studies e.g. Montagner, 1985; Ekström and Dziewonski, 1998; Panning and Romanowicz, 2006. The maximum positive vertical gradient of ξ^* at ~ 100 km is usually interpreted as the depth of the lithosphere-asthenosphere boundary (LAB). Since this gradient appears to be constant laterally, radial anisotropy is often used as a seismic proxy to trace an age-independent LAB (Burgos et al., 2014; Beghein, Xing, and Goes, 2019). Although the origin of anisotropy in the asthenosphere is well-understood purely in terms of CPO, lithospheric anisotropy may be an amalgamation of CPO and SPO (Wang et al., 2013). Our task is to invoke the composite law to isolate SPO from CPO in this tomographic model with the help of a homogenized CPO model.

7.7.3 The CPO model

In this section, we re-interpret the results of Hedjazian et al. (2017) where they examined radial anisotropy profiles predicted from CPO models produced by plate-driven

flows underneath a mid-ocean ridge. From their work, we borrowed two CPO models that correspond to a fast-developing CPO and a slow-developing CPO. The rate is dictated by the dimensionless grain boundary mobility parameter M which controls the degree of dynamic recrystallization (Kaminski, Ribe, and Browaeys, 2004). In the first case, a classical value of $M = 125$ constrained from laboratory experiments was imposed. Subsequently, the second case considers a case where $M = 10$ (*i.e.*, slower CPO evolution) which also reproduces experimental results but in the case of a non-random initial (*i.e.*, initially-developed) CPO (Boneh et al., 2015). Hedjazian et al. (2017) compared the CPO anisotropy directly with tomographic models. They concluded that the patterns of radial anisotropy predicted with the slow CPO evolution were in better agreement with tomographic models. We homogenize the two CPO models and obtain their long-wavelength effective equivalent, and again appraise the resulting profiles in comparison with tomographic observations.

The intrinsic CPO mineralogical model

2-D surface-driven flows were acquired by solving the incompressible momentum and energy equations befitting to mantle convection using the code `Fluidity` (Davies, Wilson, and Kramer, 2011). In both CPO models, upper-mantle deformation is governed by a composite dislocation and diffusion creep (*i.e.*, mixed rheology) following the implementation of Garel et al. (2014). `D-Rex` was used to model CPO evolution. Crystal aggregates comprised of 70% type-A olivine and 30% enstatite starting with a random initial CPO at 400 km are subsequently deformed along their flow trajectory. A complete description of the methodology can be found in Hedjazian et al. (2017)

Figure 7.12 displays the intrinsic radial anisotropy profiles ξ_{CPO} belonging to the fast-evolving CPO with reference `D-Rex` values $M = 125$ (model A) and the slow-evolving CPO with $M = 10$ (model B). Model A predicts a layer with strong levels of intrinsic radial anisotropy of about 10% ($\xi_{\mathrm{CPO}} \approx 1.1$) at a depth of ~ 80 km starting at approximately 20 My. At about the same depth, tomography models yield approximately 5% radial anisotropy e.g. Panning and Romanowicz, 2006; Nettles and Dziewoński, 2008; Burgos et al., 2014. Hence, it is argued that model A overpredicts the level of large-scale anisotropy in the upper-mantle (Hedjazian et al., 2017). On the contrary, model B predicts modest levels of intrinsic radial anisotropy, about 5% ($\xi_{\mathrm{CPO}} \approx 1.05$) across the oceanic lithosphere (> 80 km) which is more consistent with tomographic observations.

The homogenized CPO model

Figure 7.12 also shows the effective intrinsic radial anisotropy profiles ξ^*_{CPO} of model A* and model B* . In both cases, the ensuing patterns of radial anisotropy are smoothed out as a result of homogenization. For instance, the apparent two-layered distribution of intrinsic radial anisotropy with depth (down to $\sim$ 250 km) in model A vanishes upon homogenization. The depth profiles of effective intrinsic radial anisotropy as a result contain one layer of radial anisotropy centered at $\sim$ 100 km, making it compatible with expected patterns from tomographic models of the asthenosphere. Furthermore, it was implied that radial anisotropy predicted with laboratory parameters exceeds tomographic observations. Here, we argue that it may be other way around. Due to finite-frequency effects and eventually limitations in resolution power, seismic tomography underestimates the strength of intrinsic anisotropy, at least in the absence of small-scale isotropic heterogeneities. The level of radial anisotropy in models A and B are therefore larger than their homogenized/tomographic counterparts (models A* and B*). As opposed to common practice, the physical parameters used in CPO models of which are initially constrained by experimental data may need not be manually tuned, and perhaps that the action of varying such parameters to conform with tomographic observations deems unnecessary. We therefore conclude that direct visual comparison between a CPO model and a tomographic model could lead to wrong interpretations, and that homogenization is necessary to have correct interpretations of the CPO models.

7.7.4 The SPO model

The SPO models of Figure 7.12 illustrate the extrinsic radial anisotropy ξ_{SPO} profiles computed from ξ^* of the tomographic model, and ξ^*_{CPO} of model A* (model C) and ξ^*_{CPO} of model B* (model D) using eq. (7.33) for each pixel on the 2-D maps. Strong levels of positive extrinsic radial anisotropy clustering near the ridge axis may be due to the inability of surface waves to register vertical flow because of its limited lateral resolution. At the moment, one should look past the 20 My mark. Based on our results, to infer a realistic SPO model from a tomographic model, one should favor a CPO model with a fast-evolving texture over a slow-evolving one. This is expected since the slow model D was tailored to fit the observations from CPO only. On the other hand, the fast model C displays positive extrinsic radial anisotropy above 200 km. This is more consistent with the observed lithospheric anisotropy underneath oceanic plates which has been partly due to the existence of lateral fine-scale structures associated with magmatic underplating and congealing e.g. Auer et al., 2015; Kennett and Furumura, 2015;

Hansen, Qi, and Warren, 2016. Although our SPO models based on first-order principles are far from perfect, this demonstration serves as a proof of concept.

7.8 Conclusion

In retrospect, differentiating the relative contributions of CPO and SPO to the full effective medium is not a simple, straightforward process. The tomographic operator $\mathcal{H}$ is highly non-linear hence a full effective medium cannot be explicitly decomposed into two independent components; not to mention the manifestation of non-negligible spurious artifacts that only exacerbates the problem. It would be seemingly implausible to exactly measure CPO and SPO discretely from tomographic models. One of the most logical courses of action to address such problems is to compare the results with existing micro-mechanical models of CPO evolution. Should there be any discrepancy between the two models however, to the best of our knowledge, there is still no consistent way to independently quantify CPO and SPO from the full effective medium.

Here, we proposed a very simple composite law that directly quantifies the separate contributions of CPO and SPO from the full effective radial anisotropy ξ^* inferred from tomographic models:

$$\xi^* = \xi_{\text{SPO}} \times \xi^*_{\text{CPO}},$$

which we have verified numerically using simple 2-D toy models of an intrinsically anisotropic and a heterogeneous mantle. Although our numerical experiments were mainly a proof of concept, constraining a CPO model from an existing tomographic model is unwarranted and that we highly recommend homogenizing a CPO model as an intermedial step. Doing so may therefore yield essential observations that could benchmark a suite of tomographic models based on this numerical paradigm.

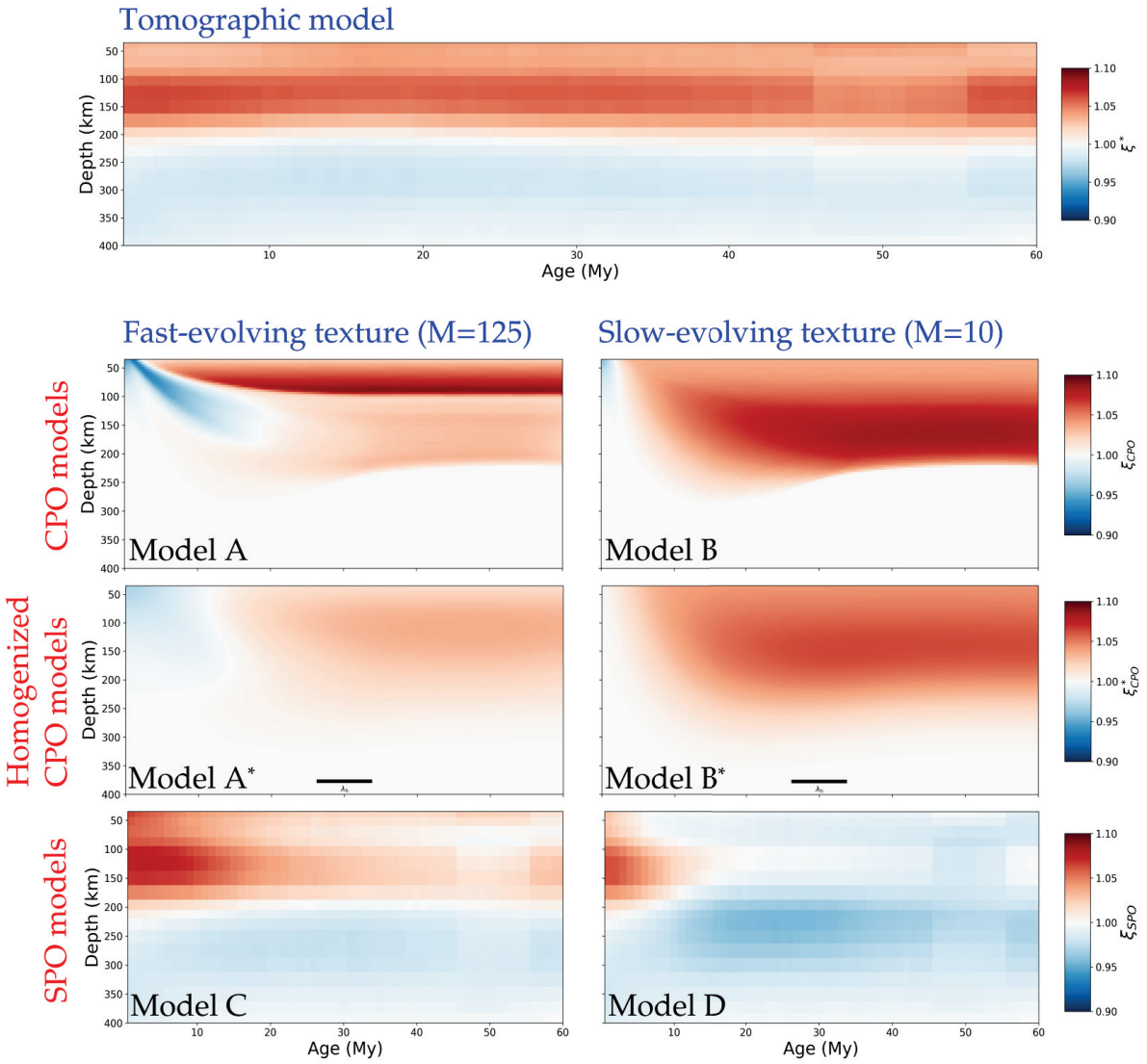

FIGURE 7.12: Radial anisotropy profiles across the upper-mantle underneath a mid-ocean ridge obtained from a tomographic model (top panel), reference CPO models corresponding to fast and slow-evolving textures (models A and B), homogenized versions of model A (model A*) and of model B (model B*). Upon homogenization, the strength of radial anisotropy is curtailed. CPO models constructed from reference parameter values therefore do not overestimate anisotropy, instead, seismic tomography underestimates anisotropy due to limited frequency band. Models C and D, respectively, are the extrinsic radial anisotropy profiles computed from ξ^* of the tomographic model, and ξ^*_{CPO} of model A* , and of model B* using the composite law. Positive lithospheric radial anisotropy in model C implies the existence of horizontally-laminated structures consistent with observations. This is absent in model D which is expected since model B* is designed to fit observations.

Bibliography

Adam, Joanne M-C and Sergei Lebedev (2012). "Azimuthal anisotropy beneath southern Africa from very broad-band surface-wave dispersion measurements". In: *Geophysical Journal International* 191.1, pp. 155–174.

Albers, Michael (2000). "A local mesh refinement multigrid method for 3-D convection problems with strongly variable viscosity". In: *Journal of Computational Physics* 160.1, pp. 126–150.

Alder, C et al. (2017). "Quantifying seismic anisotropy induced by small-scale chemical heterogeneities". In: *Geophysical Journal International* 211.3, pp. 1585–1600.

Aleksandrov, Vladislav and Henri Samuel (2010). "The Schur complement method and solution of large-scale geophysical problems". In.

Allègre, Claude J and Donald L Turcotte (1986). "Implications of a two-component marble-cake mantle". In: *Nature* 323.6084, pp. 123–127.

Anderson, Don L (1961). "Elastic wave propagation in layered anisotropic media". In: *Journal of Geophysical Research* 66.9, pp. 2953–2963.

— (1967). "Phase changes in the upper mantle". In: *Science* 157.3793, pp. 1165–1173.

Auer, Ludwig et al. (2015). "Thermal structure, radial anisotropy, and dynamics of oceanic boundary layers". In: *Geophysical Research Letters* 42.22, pp. 9740–9749.

Babuska, Vladislav and Michel Cara (1991). *Seismic anisotropy in the Earth*. Vol. 10. Springer Science & Business Media.

Backus, George E (1962). "Long-wave elastic anisotropy produced by horizontal layering". In: *Journal of Geophysical Research* 67.11, pp. 4427–4440.

Bai, Quan and David L Kohlstedt (1992). "High-temperature creep of olivine single crystals, 2. Dislocation structures". In: *Tectonophysics* 206.1-2, pp. 1–29.

Bai, Quart, SJ Mackwell, and DL Kohlstedt (1991). "High-temperature creep of olivine single crystals 1. Mechanical results for buffered samples". In: *Journal of Geophysical Research: Solid Earth* 96.B2, pp. 2441–2463.

Bakulin, Andrey (2003). "Intrinsic and layer-induced vertical transverse isotropy". In: *Geophysics* 68.5, pp. 1708–1713.

Ballmer, Maxim D et al. (2015). "Compositional mantle layering revealed by slab stagnation at˜ 1000-km depth". In: *Science Advances* 1.11, e1500815.

Baumann, Tobias S, Boris JP Kaus, and Anton A Popov (2014). "Constraining effective rheology through parallel joint geodynamic inversion". In: *Tectonophysics* 631, pp. 197–211.

Baumann, TS and Boris JP Kaus (2015). "Geodynamic inversion to constrain the nonlinear rheology of the lithosphere". In: *Geophysical Journal International* 202.2, pp. 1289–1316.

Becker, Thorsten W and Lapo Boschi (2002). "A comparison of tomographic and geodynamic mantle models". In: *Geochemistry, Geophysics, Geosystems* 3.1.

Becker, Thorsten W, Bogdan Kustowski, and Göran Ekström (2008). "Radial seismic anisotropy as a constraint for upper mantle rheology". In: *Earth and Planetary Science Letters* 267.1-2, pp. 213–227.

Becker, Thorsten W et al. (2003). "Comparison of azimuthal seismic anisotropy from surface waves and finite strain from global mantle-circulation models". In: *Geophysical Journal International* 155.2, pp. 696–714.

Becker, Thorsten W et al. (2006). "Statistical properties of seismic anisotropy predicted by upper mantle geodynamic models". In: *Journal of Geophysical Research: Solid Earth* 111.B8.

Becker, Thorsten W et al. (2014). "Origin of azimuthal seismic anisotropy in oceanic plates and mantle". In: *Earth and Planetary Science Letters* 401, pp. 236–250.

Beghein, Caroline, Z Xing, and S Goes (2019). "Thermal nature and resolution of the lithosphere–asthenosphere boundary under the Pacific from surface waves". In: *Geophysical Journal International* 216.2, pp. 1441–1465.

Beghein, Caroline et al. (2014). "Changes in seismic anisotropy shed light on the nature of the Gutenberg discontinuity". In: *Science* 343.6176, pp. 1237–1240.

Bercovici, Dave, Gerald Schubert, and Gary A Glatzmaier (1989). "Three-dimensional spherical models of convection in the Earth's mantle". In: *Science* 244.4907, pp. 950–955.

Bercovici, David (1995a). "A source-sink model of the generation of plate tectonics from non-Newtonian mantle flow". In: *Journal of Geophysical Research: Solid Earth* 100.B2, pp. 2013–2030.

— (1995b). "On the purpose of toroidal motion in a convecting mantle". In: *Geophysical research letters* 22.23, pp. 3107–3110.

Bishop, Christopher M et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Bodin, T et al. (2016). "Imaging anisotropic layering with Bayesian inversion of multiple data types". In: *Geophysical Journal International* 206.1, pp. 605–629.

Bodin, Thomas and Malcolm Sambridge (2009). "Seismic tomography with the reversible jump algorithm". In: *Geophysical Journal International* 178.3, pp. 1411–1436.

Bodin, Thomas et al. (2012). "Transdimensional inversion of receiver functions and surface wave dispersion". In: *Journal of Geophysical Research: Solid Earth* 117.B2.

Bodin, Thomas et al. (2015). "Interpreting radial anisotropy in global and regional tomographic models". In: *The Earth's heterogeneous mantle*. Springer, pp. 105–144.

Boneh, Yuval et al. (2015). "Modeling olivine CPO evolution with complex deformation histories: Implications for the interpretation of seismic anisotropy in the mantle". In: *Geochemistry, Geophysics, Geosystems* 16.10, pp. 3436–3455.

Box, George EP and George C Tiao (2011). *Bayesian inference in statistical analysis*. Vol. 40. John Wiley & Sons.

Brandt, A. (1982). "Guide to multigrid development". In: *Lecture Notes in Mathematics* 960, pp. 220–312.

Browaeys, Jules Thomas and Sébastien Chevrot (2004). "Decomposition of the elastic tensor and geophysical applications". In: *Geophysical Journal International* 159.2, pp. 667–678.

Brush, Stephen G (1980). "Discovery of the Earth's core". In: *American journal of physics* 48.9, pp. 705–724.

Buffett, Bruce A (2000). "Earth's core and the geodynamo". In: *Science* 288.5473, pp. 2007–2012.

Bunge, Hans-Peter and John R Baumgardner (1995). "Mantle convection modeling on parallel virtual machines". In: *Computers in physics* 9.2, pp. 207–215.

Bunge, Hans-Peter, CR Hagelberg, and BJ Travis (2003). "Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography". In: *Geophysical Journal International* 152.2, pp. 280–301.

Burgos, Gael et al. (2014). "Oceanic lithosphere-asthenosphere boundary from surface wave dispersion data". In: *Journal of Geophysical Research: Solid Earth* 119.2, pp. 1079–1093.

Bystricky, M et al. (2000). "High shear strain of olivine aggregates: Rheological and seismic consequences". In: *Science* 290.5496, pp. 1564–1567.

Cammarano, Fabio et al. (2009). "Inferring the thermochemical structure of the upper mantle from seismic data". In: *Geophysical Journal International* 179.2, pp. 1169–1185.

Capdeville, Yann, Laurent Guillot, and Jean-Jacques Marigo (2010). "2-D non-periodic homogenization to upscale elastic media for P–SV waves". In: *Geophysical Journal International* 182.2, pp. 903–922.

Capdeville, Yann and J-J Marigo (2007). "Second order homogenization of the elastic wave equation for non-periodic layered media". In: *Geophysical Journal International* 170.2, pp. 823–838.

Capdeville, Yann and Ludovic Métivier (2018). "Elastic full waveform inversion based on the homogenization method: theoretical framework and 2-D numerical illustrations". In: *Geophysical Journal International* 213.2, pp. 1093–1112.

Capdeville, Yann, Ming Zhao, and Paul Cupillard (2015). "Fast Fourier homogenization for elastic wave propagation in complex media". In: *Wave Motion* 54, pp. 170–186.

Capdeville, Yann et al. (2013). "Residual homogenization for seismic forward and inverse problems in layered media". In: *Geophysical Journal International* 194.1, pp. 470–487.

Chang, Sung-Joon and Ana MG Ferreira (2019). "Inference on water content in the mantle transition zone near subducted slabs from anisotropy tomography". In: *Geochemistry, Geophysics, Geosystems* 20.2, pp. 1189–1201.

Chang, Sung-Joon et al. (2014). "Global radially anisotropic mantle structure from multiple datasets: a review, current challenges, and outlook". In: *Tectonophysics* 617, pp. 1–19.

Chang, Sung-Joon et al. (2015). "Joint inversion for global isotropic and radially anisotropic mantle structure including crustal thickness perturbations". In: *Journal of Geophysical Research: Solid Earth* 120.6, pp. 4278–4300.

Charlety, Jean et al. (2013). "Global seismic tomography with sparsity constraints: Comparison with smoothing and damping regularization". In: *Journal of Geophysical Research: Solid Earth* 118.9, pp. 4887–4899.

Chen, Wang-Ping and Michael R Brudzinski (2003). "Seismic anisotropy in the mantle transition zone beneath Fiji-Tonga". In: *Geophysical Research Letters* 30.13.

Christensen, U and H Harder (1991). "3-D convection with variable viscosity". In: *Geophysical Journal International* 104.1, pp. 213–226.

Coltice, Nicolas and Yanick Ricard (1999). "Geochemical observations and one layer mantle convection". In: *Earth and Planetary Science Letters* 174.1-2, pp. 125–137.

Connolly, JAD (2009). "The geodynamic equation of state: what and how". In: *Geochemistry, Geophysics, Geosystems* 10.10.

Connolly, James AD (2005). "Computation of phase equilibria by linear programming: a tool for geodynamic modeling and its application to subduction zone decarbonation". In: *Earth and Planetary Science Letters* 236.1-2, pp. 524–541.

Conway, Dennis et al. (2019). "Inverting magnetotelluric responses in a three-dimensional earth using fast forward approximations based on artificial neural networks". In: *Computers & Geosciences* 127, pp. 44–52.

Cowin, Stephen C and Morteza M Mehrabadi (1987). "On the identification of material symmetry for anisotropic elastic materials". In: *The Quarterly Journal of Mechanics and Applied Mathematics* 40.4, pp. 451–476.

Crampin, Stuart (1981). "A review of wave motion in anisotropic and cracked elastic-media". In: *Wave motion* 3.4, pp. 343–391.

Crampin, Stuart and David C Booth (1985). "Shear-wave polarizations near the North Anatolian Fault–II. Interpretation in terms of crack-induced anisotropy". In: *Geophysical Journal International* 83.1, pp. 75–92.

Czechowski, Leszek (1993). "The Origin of Hotspots and The D" Layer". In: *Geodesy and Physics of the Earth*. Springer, pp. 392–395.

Da Silva, Cesar, Lars Stixrude, and Renata M Wentzcovitch (1997). "Elastic constants and anisotropy of forsterite at high pressure". In: *Geophysical Research Letters* 24.15, pp. 1963–1966.

Daly, Reginald Aldworth (1940). *Strength and Structure of the Earth*. Prentice-Hall New York.

Davaille, Anne and Claude Jaupart (1993). "Transient high-Rayleigh-number thermal convection with large viscosity variations". In: *Journal of Fluid Mechanics* 253, pp. 141–166.

Davies, D Rhodri, Cian R Wilson, and Stephan C Kramer (2011). "Fluidity: A fully unstructured anisotropic adaptive mesh computational modeling framework for geodynamics". In: *Geochemistry, Geophysics, Geosystems* 12.6.

Debayle, E (1999). "SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion". In: *Geophysical Journal International* 137.3, pp. 747–754.

Debayle, Eric, Brian Kennett, and Keith Priestley (2005). "Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia". In: *Nature* 433.7025, p. 509.

Debayle, Eric and Yanick Ricard (2012). "A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements". In: *Journal of Geophysical Research: Solid Earth* 117.B10.

— (2013). "Seismic observations of large-scale deformation at the bottom of fast-moving plates". In: *Earth and Planetary Science Letters* 376, pp. 165–177.

Debayle, Eric et al. (2020). "Seismic evidence for partial melt below tectonic plates". In: *Nature* 586.7830, pp. 555–559.

Debski, Wojciech (2010). "Seismic tomography by Monte Carlo sampling". In: *Pure and applied geophysics* 167.1-2, pp. 131–152.

Deschamps, Frédéric et al. (2008). "Azimuthal anisotropy of Rayleigh-wave phase velocities in the east-central United States". In: *Geophysical Journal International* 173.3, pp. 827–843.

Dettmer, Jan, Stan E Dosso, and Charles W Holland (2007). "Uncertainty estimation in seismo-acoustic reflection travel time inversion". In: *The Journal of the Acoustical Society of America* 122.1, pp. 161–176.

Di Giacomo, Domenico, E Robert Engdahl, and Dmitry A Storchak (2018). "The ISC-GEM Earthquake Catalogue (1904–2014): status after the Extension Project." In: *Earth System Science Data* 10.4.

Durham, WB, C Goetze, and B Blake (1977). "Plastic flow of oriented single crystals of olivine: 2. Observations and interpretations of the dislocation structures". In: *Journal of Geophysical Research* 82.36, pp. 5755–5770.

Dziewonski, Adam M (1984). "Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6". In: *Journal of Geophysical Research: Solid Earth* 89.B7, pp. 5929–5952.

Dziewonski, Adam M and Don L Anderson (1981). "Preliminary reference Earth model". In: *Physics of the earth and planetary interiors* 25.4, pp. 297–356.

Ekström, Göran (2011). "A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25-250 s". In: *Geophysical Journal International* 187.3, pp. 1668–1686.

Ekström, Göran and Adam M Dziewonski (1998). "The unique anisotropy of the Pacific upper mantle". In: *Nature* 394.6689, pp. 168–172.

Estey, Louis H and Bruce J Douglas (1986). "Upper mantle anisotropy: a preliminary model". In: *Journal of Geophysical Research: Solid Earth* 91.B11, pp. 11393–11406.

Faccenda, Manuele and FA Capitanio (2013). "Seismic anisotropy around subduction zones: Insights from three-dimensional modeling of upper mantle deformation and SKS splitting calculations". In: *Geochemistry, Geophysics, Geosystems* 14.1, pp. 243–262.

Faccenda, Manuele et al. (2019). "Extrinsic elastic anisotropy in a compositionally heterogeneous Earth's mantle". In: *Journal of Geophysical Research: Solid Earth* 124.2, pp. 1671–1687.

Fedorenko, Radii Petrovich (1964). "The speed of convergence of one iterative process". In: *USSR Computational Mathematics and Mathematical Physics* 4.3, pp. 227–235.

Fedorov, Fedor I (2013). *Theory of elastic waves in crystals*. Springer Science & Business Media.

Ferreira, Ana MG et al. (2019). "Ubiquitous lower-mantle anisotropy beneath subduction zones". In: *Nature Geoscience* 12.4, pp. 301–306.

Fukao, Yoshio et al. (2009). "Stagnant slab: a review". In: *Annual Review of Earth and Planetary Sciences* 37, pp. 19–46.

Gable, Carl W, Richard J O'connell, and Bryan J Travis (1991). "Convection in three dimensions with surface plates: Generation of toroidal flow". In: *Journal of Geophysical Research: Solid Earth* 96.B5, pp. 8391–8405.

Gallego, A, Garrett Ito, and RA Dunn (2013). "Investigating seismic anisotropy beneath the Reykjanes Ridge using models of mantle flow, crystallographic evolution, and surface wave propagation". In: *Geochemistry, Geophysics, Geosystems* 14.8, pp. 3250–3267.

Garel, Fanny et al. (2014). "Interaction of subducted slabs with the mantle transition-zone: A regime diagram from 2-D thermo-mechanical models with a mobile trench and an overriding plate". In: *Geochemistry, Geophysics, Geosystems* 15.5, pp. 1739–1765.

Gerya, T. V. (2010). *Introduction to numerical geodynamic modeling*. Cambdridge University Press, p. 345.

Ghosh, A, TW Becker, and SJ Zhong (2010). "Effects of lateral viscosity variations on the geoid". In: *Geophysical Research Letters* 37.1.

Glasstone, Samuel, Keith J Laidler, and Henry Eyring (1941). *The theory of rate processes*. McGraw-hill.

Gorbatov, A et al. (2000). "Signature of remnant slabs in the North Pacific from P-wave tomography". In: *Geophysical Journal International* 142.1, pp. 27–36.

Guillope, M and JP Poirier (1979). "Dynamic recrystallization during creep of single-crystalline halite: An experimental study". In: *Journal of Geophysical Research: Solid Earth* 84.B10, pp. 5557–5567.

Guillot, Laurent, Yann Capdeville, and Jean-Jacques Marigo (2010). "2-D non-periodic homogenization of the elastic wave equation: SH case". In: *Geophysical Journal International* 182.3, pp. 1438–1454.

Gung, Yuancheng, Mark Panning, and Barbara Romanowicz (2003). "Global anisotropy and the thickness of continents". In: *Nature* 422.6933, pp. 707–711.

Gurnis, Michael and Bradford H Hager (1988). "Controls of the structure of subducted slabs". In: *Nature* 335.6188, pp. 317–321.

Hager, Bradford H (1984). "Subducted slabs and the geoid: Constraints on mantle rheology and flow". In: *Journal of Geophysical Research: Solid Earth* 89.B7, pp. 6003–6015.

Hager, Bradford H and Robert W Clayton (1989). "Constraints on the structure of mantle convection using seismic observations, flow models, and the geoid". In.

Hager, Bradford H and Richard J O'Connell (1981). "A simple global model of plate dynamics and mantle convection". In: *Journal of Geophysical Research: Solid Earth* 86.B6, pp. 4843–4867.

Hall, Chad E et al. (2000). "The influence of plate motions on three-dimensional back arc mantle flow and shear wave splitting". In: *Journal of Geophysical Research: Solid Earth* 105.B12, pp. 28009–28033.

Hansen, Lars N, Chao Qi, and Jessica M Warren (2016). "Olivine anisotropy suggests Gutenberg discontinuity is not the base of the lithosphere". In: *Proceedings of the National Academy of Sciences* 113.38, pp. 10503–10506.

Hansen, Lars N et al. (2014a). "Protracted fabric evolution in olivine: Implications for the relationship among strain, crystallographic fabric, and seismic anisotropy". In: *Earth and Planetary Science Letters* 387, pp. 157–168.

Hansen, Thomas Mejer and Knud Skou Cordua (2017). "Efficient Monte Carlo sampling of inverse problems using a neural network-based forward—applied to GPR crosshole traveltime inversion". In: *Geophysical Journal International* 211.3, pp. 1524–1533.

Hansen, Thomas Mejer et al. (2014b). "Accounting for imperfect forward modeling in geophysical inverse problems—Exemplified for crosshole tomography". In: *Geophysics* 79.3, H1–H21.

Hanson, David R and Hartmut A Spetzler (1994). "Transient creep in natural and synthetic, iron-bearing olivine single crystals: Mechanical results and dislocation microstructures". In: *Tectonophysics* 235.4, pp. 293–315.

Hastings, W Keith (1970). "Monte Carlo sampling methods using Markov chains and their applications". In.

Hedjazian, N and E Kaminski (2014). "Defining a proxy for the interpretation of seismic anisotropy in non-Newtonian mantle flows". In: *Geophysical Research Letters* 41.20, pp. 7065–7072.

Hedjazian, Navid et al. (2017). "Age-independent seismic anisotropy under oceanic plates explained by strain history in the asthenosphere". In: *Earth and Planetary Science Letters* 460, pp. 135–142.

Hess, Harry H (1962). "History of ocean basins". In: *Petrologic studies* 4, pp. 599–620.

Heuret, Arnauld and Serge Lallemand (2005). "Plate motions, slab dynamics and back-arc deformation". In: *Physics of the Earth and Planetary Interiors* 149.1-2, pp. 31–51.

Hilst, Rob van der and Tetsuzo Seno (1993). "Effects of relative plate motion on the deep structure and penetration depth of slabs below the Izu-Bonin and Mariana island arcs". In: *Earth and Planetary Science Letters* 120.3-4, pp. 395–407.

Hilst, Robert Dirk Van der, Sri Widiyantoro, and ER Engdahl (1997). "Evidence for deep mantle circulation from global tomography". In: *Nature* 386.6625, pp. 578–584.

Hilst, Ron van der (1995). "Complex morphology of subducted lithosphere in the mantle beneath the Tonga trench". In: *Nature* 374.6518, pp. 154–157.

Ho, Tak, Keith Priestley, and Eric Debayle (2016). "A global horizontal shear velocity model of the upper mantle from multimode Love wave measurements". In: *Geophysical journal international* 207.1, pp. 542–561.

Hofmann, Albrecht W (1988). "Chemical differentiation of the Earth: the relationship between mantle, continental crust, and oceanic crust". In: *Earth and Planetary Science Letters* 90.3, pp. 297–314.

Holmes, Arthur (1931). "XVIII. Radioactivity and Earth movements". In: *Transactions of the Geological Society of Glasgow* 18.3, pp. 559–606.

Hulbert, Claudia et al. (2019). "Similarity of fast and slow earthquakes illuminated by machine learning". In: *Nature Geoscience* 12.1, pp. 69–74.

Husen, S et al. (2003). "Probabilistic earthquake location in complex three-dimensional velocity models: Application to Switzerland". In: *Journal of Geophysical Research: Solid Earth* 108.B2.

Isacks, Bryan, Jack Oliver, and Lynn R Sykes (1968). "Seismology and the new global tectonics". In: *Journal of Geophysical Research* 73.18, pp. 5855–5899.

Ita, Joel and Scott D King (1994). "Sensitivity of convection with an endothermic phase change to the form of governing equations, initial conditions, boundary conditions, and equation of state". In: *Journal of Geophysical Research: Solid Earth* 99.B8, pp. 15919–15938.

— (1998). "The influence of thermodynamic formulation on simulations of subduction zone geometry and history". In: *Geophysical research letters* 25.9, pp. 1463–1466.

Jin, ZM, Quan Bai, and David L Kohlstedt (1994). "High-temperature creep of olivine crystals from four localities". In: *Physics of the Earth and Planetary Interiors* 82.1, pp. 55–64.

Jung, Haemyeong and Shun-Ichiro Karato (2001). "Effects of water on dynamically recrystallized grain-size of olivine". In: *Journal of Structural Geology* 23.9, pp. 1337–1344.

Kageyama, Akira and Tetsuya Sato (2004). ""Yin-Yang grid": An overset grid in spherical geometry". In: *Geochemistry, Geophysics, Geosystems* 5.9.

Kaminski, E and NM Ribe (2001). "A kinematic model for recrystallization and texture development in olivine polycrystals". In: *Earth and Planetary Science Letters* 189.3-4, pp. 253–267.

Kaminski, Édouard and Neil M Ribe (2002). "Timescales for the evolution of seismic anisotropy in mantle flow". In: *Geochemistry, Geophysics, Geosystems* 3.8, pp. 1–17.

Kaminski, Edouard, Neil M Ribe, and Jules T Browaeys (2004). "D-Rex, a program for calculation of seismic anisotropy due to crystal lattice preferred orientation in the convective upper mantle". In: *Geophysical Journal International* 158.2, pp. 744–752.

Kaneshima, Satoshi (2014). "Upper bounds of seismic anisotropy in the Tonga slab near deep earthquake foci and in the lower mantle". In: *Geophysical Journal International* 197.1, pp. 351–368.

Kárason, Hrafnkell and Rob D Van Der Hilst (2000). "Constraints on mantle convection from seismic tomography". In: *Geophysical monograph* 121, pp. 277–288.

Karato, Shun-ichiro (1988). "The role of recrystallization in the preferred orientation of olivine". In: *Physics of the Earth and Planetary Interiors* 51.1-3, pp. 107–122.

Käufl, Paul et al. (2014). "A framework for fast probabilistic centroid-moment-tensor determination—inversion of regional static displacement measurements". In: *Geophysical Journal International* 196.3, pp. 1676–1693.

Kelemen, Peter B, Michael Braun, and Greg Hirth (2000). "Spatial distribution of melt conduits in the mantle beneath oceanic spreading ridges: Observations from the Ingalls and Oman ophiolites". In: *Geochemistry, Geophysics, Geosystems* 1.7.

Kendall, JM and PG Silver (1998). "Investigating causes of D" anisotropy". In: *The core-mantle boundary region* 28, pp. 97–118.

Kennett, BLN and ER Engdahl (1991). "Traveltimes for global earthquake location and phase identification". In: *Geophysical Journal International* 105.2, pp. 429–465.

Kennett, BLN and Takahashi Furumura (2015). "Toward the reconciliation of seismological and petrological perspectives on oceanic lithosphere heterogeneity". In: *Geochemistry, Geophysics, Geosystems* 16.9, pp. 3129–3141.

Kohlstedt, DL and C Goetze (1974). "Low-stress high-temperature creep in olivine single crystals". In: *Journal of Geophysical Research* 79.14, pp. 2045–2051.

Köpke, Corinna, James Irving, and Ahmed H Elsheikh (2018). "Accounting for model error in Bayesian solutions to hydrogeophysical inverse problems using a local basis approach". In: *Advances in water resources* 116, pp. 195–207.

Kronbichler, Martin, Timo Heister, and Wolfgang Bangerth (2012). "High accuracy mantle convection simulation through modern numerical methods". In: *Geophysical Journal International* 191.1, pp. 12–29.

Kumazawa, Mineo and Orson L Anderson (1969). "Elastic moduli, pressure derivatives, and temperature derivatives of single-crystal olivine and single-crystal forsterite". In: *Journal of Geophysical Research* 74.25, pp. 5961–5972.

Kustowski, B, G Ekström, and AM Dziewoński (2008). "Anisotropic shear-wave velocity structure of the Earth's mantle: A global model". In: *Journal of Geophysical Research: Solid Earth* 113.B6.

Lassak, Teresa Mae et al. (2006). "Seismic characterization of mantle flow in subduction systems: Can we resolve a hydrated mantle wedge?" In: *Earth and Planetary Science Letters* 243.3-4, pp. 632–649.

Le Pichon, Xavier (1968). "Sea-floor spreading and continental drift". In: *Journal of Geophysical Research* 73.12, pp. 3661–3697.

Lebedev, Sergei, Thomas Meier, and Rob D van der Hilst (2006). "Asthenospheric flow and origin of volcanism in the Baikal Rift area". In: *Earth and Planetary Science Letters* 249.3-4, pp. 415–424.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep learning". In: *nature* 521.7553, p. 436.

Lehmann, Inge (1936). "P', Publ". In: *Bur. Centr. Seism. Internat. Serie A* 14, pp. 87–115.

Lev, Einat and Bradford H Hager (2008). "Prediction of anisotropy from flow models: A comparison of three methods". In: *Geochemistry, Geophysics, Geosystems* 9.7.

Lithgow-Bertelloni, C et al. (2001). "Plume generation in natural thermal convection at high Rayleigh and Prandtl numbers". In: *Journal of Fluid Mechanics* 434, p. 1.

Lomax, Anthony et al. (2000). "Probabilistic earthquake location in 3D and layered models". In: *Advances in seismic event location*. Springer, pp. 101–134.

Long, Maureen D and Thorsten W Becker (2010). "Mantle dynamics and seismic anisotropy". In: *Earth and Planetary Science Letters* 297.3-4, pp. 341–354.

Love, Augustus Edward Hough (1906). *A Treatise on the Mathematical Theory of Elasticity, by AEH Love*. The University Press.

Lu, Yang et al. (2018). "High-resolution surface wave tomography of the European crust and uppermost mantle from ambient seismic noise". In: *Geophysical Journal International* 214.2, pp. 1136–1150.

Macquet, Marie et al. (2014). "Ambient noise tomography of the Pyrenees and the surrounding regions: inversion for a 3-D Vs model in the presence of a very heterogeneous crust". In: *Geophysical Journal International* 199.1, pp. 402–415.

Magali, JK et al. (2020). "Geodynamic Tomography: Constraining Upper Mantle Deformation Patterns from Bayesian Inversion of Surface Waves". In: *Geophysical Journal International*.

Magali, John Keith (2020). *1D Anisotropic Surface Wave Tomography with Bayesian Inference*. arXiv: `2012.03915 [physics.geo-ph]`.

Maggi, Alessia et al. (2000). "Earthquake focal depths, effective elastic thickness, and the strength of the continental lithosphere". In: *Geology* 28.6, pp. 495–498.

Mainprice, David (1990). "A FORTRAN program to calculate seismic anisotropy from the lattice preferred orientation of minerals". In: *Computers & Geosciences* 16.3, pp. 385–393.

Malinverno, Alberto and Victoria A Briggs (2004). "Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes". In: *Geophysics* 69.4, pp. 1005–1016.

Masalu, Desiderius CP (2007). "Mapping absolute migration of global mid-ocean ridges since 80 Ma to Present". In: *Earth, planets and space* 59.9, pp. 1061–1066.

Masters, Guy et al. (1982). "Aspherical Earth structure from fundamental spheroidal-mode data". In: *Nature* 298.5875, pp. 609–613.

Matsubara, Makoto et al. (2005). "Low-velocity oceanic crust at the top of the Philippine Sea and Pacific plates beneath the Kanto region, central Japan, imaged by seismic tomography". In: *Journal of Geophysical Research: Solid Earth* 110.B12.

Maupin, V and J Park (2015). "1.09—Theory and observations—Seismic anisotropy". In: *Treatise on Geophysics*, pp. 277–305.

Maupin, V et al. (2007). "Theory and observations—wave propagation in anisotropic media". In: *Seismology and the Structure of the Earth. Treatise on Geophysics* 1, pp. 289–321.

McKenzie, Dan (1979). "Finite deformation during fluid flow". In: *Geophysical Journal International* 58.3, pp. 689–715.

McKenzie, Dan P and WJ Morgan (1969). "Evolution of triple junctions". In: *Nature* 224.5215, pp. 125–133.

McKenzie, Dan P and Robert L Parker (1967). "The North Pacific: an example of tectonics on a sphere". In: *Nature* 216.5122, pp. 1276–1280.

McNamara, Allen K, Peter E van Keken, and S Karato (2002). "Development of anisotropic structure by solid-state convection in the Earth's lower mantle". In: *Nature* 416.6878, pp. 310–314.

Meade, Charles, Paul G Silver, and Satoshi Kaneshima (1995). "Laboratory and seismological observations of lower mantle isotropy". In: *Geophysical Research Letters* 22.10, pp. 1293–1296.

Mégnin, Charles and Barbara Romanowicz (2000). "The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms". In: *Geophysical Journal International* 143.3, pp. 709–728.

Meier, Ueli, Andrew Curtis, and Jeannot Trampert (2007). "Global crustal thickness from neural network inversion of surface wave data". In: *Geophysical Journal International* 169.2, pp. 706–722.

Metropolis, Nicholas and Stanislaw Ulam (1949). "The monte carlo method". In: *Journal of the American statistical association* 44.247, pp. 335–341.

Metropolis, Nicholas et al. (1953). "Equation of state calculations by fast computing machines". In: *The journal of chemical physics* 21.6, pp. 1087–1092.

Middlemost, Eric AK (1986). "Magmas and magmatic rocks: an introduction to igneous petrology". In.

Miller, Meghan S and Thorsten W Becker (2012). "Mantle flow deflected by interactions between subducted slabs and cratonic keels". In: *Nature Geoscience* 5.10, p. 726.

Mitrovica, Jerry X and Alessandro M Forte (1997). "Radial profile of mantle viscosity: Results from the joint inversion of convection and postglacial rebound observables". In: *Journal of Geophysical Research: Solid Earth* 102.B2, pp. 2751–2769.

Moghadas, Davood, Ahmad A Behroozmand, and Anders Vest Christiansen (2020). "Soil electrical conductivity imaging using a neural network-based forward solver: Applied to large-scale Bayesian electromagnetic inversion". In: *Journal of Applied Geophysics*, p. 104012.

Mohorovičić, Andrija (1909). *Potres od 8*.

Montagner, Jean-Paul (1985). "Seismic anisotropy of the Pacific Ocean inferred from long-period surface waves dispersion". In: *Physics of the earth and planetary interiors* 38.1, pp. 28–50.

— (1994). "Can seismology tell us anything about convection in the mantle?" In: *Reviews of Geophysics* 32.2, pp. 115–137.

— (1998). "Where can seismic anisotropy be detected in the Earth's mantle? In boundary layers..." In: *Pure and Applied Geophysics* 151.2-4, pp. 223–256.

— (2007). "Upper mantle structure: Global isotropic and anisotropic elastic tomography". In: *Treatise on geophysics* 1, pp. 559–589.

Montagner, Jean-Paul and Don L Anderson (1989). "Petrological constraints on seismic anisotropy". In: *Physics of the earth and planetary interiors* 54.1-2, pp. 82–105.

Montagner, Jean-Paul and Nelly Jobert (1988). "Vectorial tomography—ii. Application to the Indian Ocean". In: *Geophysical Journal International* 94.2, pp. 309–344.

Montagner, Jean-Paul and Henri-Claude Nataf (1986). "A simple method for inverting the azimuthal anisotropy of surface waves". In: *Journal of Geophysical Research: Solid Earth* 91.B1, pp. 511–520.

Montagner, Jean-Paul and Toshiro Tanimoto (1990). "Global anisotropy in the upper mantle inferred from the regionalization of phase velocities". In: *Journal of Geophysical Research: Solid Earth* 95.B4, pp. 4797–4819.

Montagner, JP and HC Nataf (1988). "Vectorial tomography. Part I: Theory". In: *Geophysical Journal International* 94, pp. 295–307.

Morgan, W Jason (1968). "Rises, trenches, great faults, and crustal blocks". In: *Journal of Geophysical Research* 73.6, pp. 1959–1982.

Morishige, M and T Kuwatani (2020). "Bayesian inversion of surface heat flow in subduction zones: a framework to refine geodynamic models based on observational constraints". In: *Geophysical Journal International* 222.1, pp. 103–109.

Mosegaard, Klaus and Malcolm Sambridge (2002). "Monte Carlo analysis of inverse problems". In: *Inverse problems* 18.3, R29.

Mosegaard, Klaus and Albert Tarantola (1995). "Monte Carlo sampling of solutions to inverse problems". In: *Journal of Geophysical Research: Solid Earth* 100.B7, pp. 12431–12447.

Nakanishi, Ichiro and Don L Anderson (1983). "Measurement of mantle wave velocities and inversion for lateral heterogeneity and anisotropy: 1. Analysis of Great Circle Phase Velocities". In: *Journal of Geophysical Research: Solid Earth* 88.B12, pp. 10267–10283.

Nataf, H-C, I Nakanishi, and Don L Anderson (1984). "Anisotropy and shear-velocity heterogeneities in the upper mantle". In: *Geophysical Research Letters* 11.2, pp. 109–112.

Nettles, Meredith and Adam M Dziewoński (2008). "Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America". In: *Journal of Geophysical Research: Solid Earth* 113.B2.

Nicolas, Adolphe and Nikolas I Christensen (1987). "Formation of anisotropy in upper mantle peridotites-A review". In: *Composition, structure and dynamics of the lithosphere-asthenosphere system* 16, pp. 111–123.

Nur, Amos (1971). "Effects of stress on velocity anisotropy in rocks with cracks". In: *Journal of Geophysical Research* 76.8, pp. 2022–2034.

Obayashi, Masayuki et al. (2006). "High temperature anomalies oceanward of subducting slabs at the 410-km discontinuity". In: *Earth and Planetary Science Letters* 243.1-2, pp. 149–158.

Obrebski, M et al. (2010). "Anisotropic stratification beneath Africa from joint inversion of SKS and P receiver functions". In: *Journal of Geophysical Research: Solid Earth* 115.B9.

Olson, Peter and Harvey Singer (1985). "Creeping plumes". In: *Journal of Fluid Mechanics* 158, pp. 511–531.

Olson, Peter, David A Yuen, and Derick Balsiger (1984). "Mixing of passive heterogeneities by mantle convection". In: *Journal of Geophysical Research: Solid Earth* 89.B1, pp. 425–436.

Ortega-Gelabert, O et al. (2020). "Fast Stokes flow simulations for geophysical-geodynamic inverse problems and sensitivity analyses based on reduced order modeling". In: *Journal of Geophysical Research: Solid Earth* 125.3, e2019JB018314.

Panasyuk, Svetlana V and Bradford H Hager (2000). "Inversion for mantle viscosity profiles constrained by dynamic topography and the geoid, and their estimated errors". In: *Geophysical Journal International* 143.3, pp. 821–836.

Panning, Mark and Barbara Romanowicz (2006). "A three-dimensional radially anisotropic model of shear velocity in the whole mantle". In: *Geophysical Journal International* 167.1, pp. 361–379.

Panning, Mark P and Guust Nolet (2008). "Surface wave tomography for azimuthal anisotropy in a strongly reduced parameter space". In: *Geophysical Journal International* 174.2, pp. 629–648.

Patankar, S. V. (1980). *Numerical Heat Transfer and Fluid Flow*. New York: Hemisphere Publishing Corporation, p. 197.

Patankar, Suhas (2018). *Numerical heat transfer and fluid flow*. Taylor & Francis.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Piromallo, C et al. (2006). "Three-dimensional instantaneous mantle flow induced by subduction". In: *Geophysical Research Letters* 33.8.

Plomerová, Jaroslava, Daniel Kouba, and Vladislav Babuška (2002). "Mapping the lithosphere–asthenosphere boundary through changes in surface-wave anisotropy". In: *Tectonophysics* 358.1-4, pp. 175–185.

Poirier, Jean-Paul (1985). *Creep of crystals: high-temperature deformation processes in metals, ceramics and minerals*. Cambridge University Press.

Postma, GW (1955). "Wave propagation in a stratified medium". In: *Geophysics* 20.4, pp. 780–806.

Press, William H et al. (1992). *Numerical recipes in Fortran 77: volume 1, volume 1 of Fortran numerical recipes: the art of scientific computing*. Cambridge university press.

Raleigh, CB et al. (1971). "Slip and the clinoenstatite transformation as competing rate processes in enstatite". In: *Journal of Geophysical Research* 76.17, pp. 4011–4022.

Ravenna, Matteo and Sergei Lebedev (2017). "Bayesian inversion of surface-wave data for radial and azimuthal shear-wave anisotropy, with applications to central Mongolia and west-central Italy". In: *Geophysical Journal International* 213.1, pp. 278–300.

Ribe, Neil M (1989). "Seismic anisotropy and mantle flow". In: *Journal of Geophysical Research: Solid Earth* 94.B4, pp. 4213–4223.

— (1992). "On the relation between seismic anisotropy and finite strain". In: *Journal of Geophysical Research: Solid Earth* 97.B6, pp. 8737–8747.

Ribe, Neil M and Yang Yu (1991). "A theory for plastic deformation and textural evolution of olivine polycrystals". In: *Journal of Geophysical Research: Solid Earth* 96.B5, pp. 8325–8335.

Ricard, Y and D Bercovici (2009). "A continuum theory of grain size evolution and damage". In: *Journal of Geophysical Research: Solid Earth* 114.B1.

Ricard, Yanick, Estelle Mattern, and Jan Matas (2005). "Mineral physics in thermochemical mantle models". In: *Hilst, R., Bass, JD, Matas, Trampert, J.,(Eds.), Com-position, Structure and Evolution of the Earth Mantle, AGU monograph* 160, pp. 283–300.

Ricard, Yanick and Bai Wuming (1991). "Inferring the viscosity and the 3-D density structure of the mantle from geoid, topography and plate velocities". In: *Geophysical Journal International* 105.3, pp. 561–571.

Ricard, Yanick et al. (1993). "A geodynamic model of mantle density heterogeneity". In: *Journal of Geophysical Research: Solid Earth* 98.B12, pp. 21895–21909.

Ritsema, J et al. (2011). "S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements". In: *Geophysical Journal International* 184.3, pp. 1223–1236.

Ritzwoller, Michael H et al. (2002). "Global surface wave diffraction tomography". In: *Journal of Geophysical Research: Solid Earth* 107.B12, ESE–4.

Romanowicz, Barbara (2002). "Inversion of surface waves: a review". In: *International Geophysics Series* 81.A, pp. 149–174.

— (2003). "Global mantle tomography: progress status in the past 10 years". In: *Annual Review of Earth and Planetary Sciences* 31.1, pp. 303–328.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1985). *Learning internal representations by error propagation*. Tech. rep. California Univ San Diego La Jolla Inst for Cognitive Science.

Saito, Masanori (1967). "Excitation of free oscillations and surface waves by a point source in a vertically heterogeneous earth". In: *Journal of Geophysical Research* 72.14, pp. 3689–3699.

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

Sambridge, Malcolm and Kerry Gallagher (2011). *Inverse Theory and the Monte Carlo Method*.

Samuel, H. (2012a). "Time-Domain Parallelization for Computational Geodynamics". In: *G-cubed*, doi:10.1029/2011GC003905.

— (2018). "A Deformable Particle-In-Cell Method for Advective Transport in Geodynamic Modelling". In: *gji* 214. doi:10.1093/gji/ggy231, pp. 1744–1773.

Samuel, Henri (2012b). "Time domain parallelization for computational geodynamics". In: *Geochemistry, Geophysics, Geosystems* 13.1.

Schubert, Gerald, Donald Lawson Turcotte, and Peter Olson (2001). *Mantle convection in the Earth and planets*. Cambridge University Press.

Shahnas, MH, DA Yuen, and RN Pysklywec (2018). "Inverse problems in geodynamics using machine learning algorithms". In: *Journal of Geophysical Research: Solid Earth* 123.1, pp. 296–310.

Shapiro, NM and MH Ritzwoller (2002). "Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle". In: *Geophysical Journal International* 151.1, pp. 88–105.

Shen, Weisen et al. (2012). "Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach". In: *Geophysical Journal International* 192.2, pp. 807–836.

Sieminski, Anne et al. (2007). "Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods". In: *Geophysical Journal International* 168.3, pp. 1153–1174.

Simons, Frederik J, Alet Zielhuis, and Rob D Van Der Hilst (1999). "The deep structure of the Australian continent from surface wave tomography". In: *Developments in Geotectonics*. Vol. 24. Elsevier, pp. 17–43.

Smith, Adrian Frederick Melhuish (1991). "Bayesian computational methods". In: *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 337.1647, pp. 369–386.

Smith, Martin L and FA Dahlen (1973). "The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium". In: *Journal of Geophysical Research* 78.17, pp. 3321–3333.

Steinberger, Bernhard and Thorsten W Becker (2018). "A comparison of lithospheric thickness models". In: *Tectonophysics* 746, pp. 325–338.

Steinberger, Bernhard, Rupert Sutherland, and Richard J O'connell (2004). "Prediction of Emperor-Hawaii seamount locations from a revised model of global plate motion and mantle flow". In: *Nature* 430.6996, pp. 167–173.

Stixrude, L and R Jeanloz (2015). "Constraints on seismic models from other disciplines-constraints from mineral physics on seismological models". In.

Stixrude, Lars and Carolina Lithgow-Bertelloni (2011). "Thermodynamics of mantle minerals-II. Phase equilibria". In: *Geophysical Journal International* 184.3, pp. 1180–1213.

Storchak, Dmitry A et al. (2013). "Public release of the ISC–GEM global instrumental earthquake catalogue (1900–2009)". In: *Seismological Research Letters* 84.5, pp. 810–815.

Storchak, Dmitry A et al. (2015). "The ISC-GEM global instrumental earthquake catalogue (1900–2009): introduction". In: *Physics of the Earth and Planetary Interiors* 239, pp. 48–63.

Sturgeon, William et al. (2019). "On the Origin of Radial Anisotropy Near Subducted Slabs in the Midmantle". In: *Geochemistry, Geophysics, Geosystems* 20.11, pp. 5105–5125.

Tackley, Paul J (1993). "Effects of strongly temperature-dependent viscosity on time-dependent, three-dimensional models of mantle convection". In: *Geophysical Research Letters* 20.20, pp. 2187–2190.

— (2008). "Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid". In: *Physics of the Earth and Planetary Interiors* 171.1-4, pp. 7–18.

Takeuchi, H and M Saito (1972). "Seismic surface waves". In: *Methods in computational physics* 11, pp. 217–295.

Tauzin, B and Y Ricard (2014). "Seismically deduced thermodynamics phase diagrams for the mantle transition zone". In: *Earth and Planetary Science Letters* 401, pp. 337–346.

Tauzin, Benoit et al. (2016). "Multi-mode conversion imaging of the subducted Gorda and Juan de Fuca plates below the North American continent". In: *Earth and Planetary Science Letters* 440, pp. 135–146.

Thomson, William T (1950). "Transmission of elastic waves through a stratified solid medium". In: *Journal of applied Physics* 21.2, pp. 89–93.

Tkalčić, H et al. (2006). "A multistep approach for joint modeling of surface wave dispersion and teleseismic receiver functions: Implications for lithospheric structure of the Arabian Peninsula". In: *Journal of Geophysical Research: Solid Earth* 111.B11.

Tommasi, Andréa et al. (2000). "Viscoplastic self-consistent and equilibrium-based modeling of olivine lattice preferred orientations: Implications for the upper mantle seismic anisotropy". In: *Journal of Geophysical Research: Solid Earth* 105.B4, pp. 7893–7908.

Torrance, KE and DL Turcotte (1971). "Thermal convection with large viscosity variations". In: *Journal of Fluid Mechanics* 47.1, pp. 113–125.

Tovish, Aaron, Gerald Schubert, and Bruce P Luyendyk (1978). "Mantle flow pressure and the angle of subduction: Non-Newtonian corner flows". In: *Journal of Geophysical Research: Solid Earth* 83.B12, pp. 5892–5898.

Trampert, Jeannot and Andreas Fichtner (2013). "Global imaging of the Earth's deep interior: seismic constraints on (an) isotropy, density and attenuation". In: *Physics and chemistry of the deep Earth*, pp. 324–350.

Trampert, Jeannot and Hendrik Jan van Heijst (2002). "Global azimuthal anisotropy in the transition zone". In: *Science* 296.5571, pp. 1297–1299.

Trampert, Jeannot and John H Woodhouse (1995). "Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds". In: *Geophysical Journal International* 122.2, pp. 675–690.

Turcotte, DL and ER Oxburgh (1967). "Finite amplitude convective cells and continental drift". In: *Journal of Fluid Mechanics* 28.1, pp. 29–42.

USGS (2020). *Earthquake Glossary*. URL: `https://earthquake.usgs.gov/learn/glossary/?term=shadow+zone`.

Wang, Nian et al. (2013). "Intrinsic versus extrinsic seismic anisotropy: The radial anisotropy in reference Earth models". In: *Geophysical Research Letters* 40.16, pp. 4284–4288.

Watts, AB and EB Burov (2003). "Lithospheric strength and its relationship to the elastic and seismogenic layer thickness". In: *Earth and Planetary Science Letters* 213.1-2, pp. 113–131.

White, David B (1988). "The planforms and onset of convection with a temperature-dependent viscosity". In: *Journal of Fluid Mechanics* 191, pp. 247–286.

Whitehead Jr, John A and Douglas S Luther (1975). "Dynamics of laboratory diapir and plume models". In: *Journal of Geophysical Research* 80.5, pp. 705–717.

Wiens, Douglas A and Seth Stein (1983). "Age dependence of oceanic intraplate seismicity and implications for lithospheric evolution". In: *Journal of Geophysical Research: Solid Earth* 88.B8, pp. 6455–6468.

Woodhouse, John H and Adam M Dziewonski (1984). "Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms". In: *Journal of Geophysical Research: Solid Earth* 89.B7, pp. 5953–5986.

Woodhouse, John Henry and AM Dziewonski (1989). "Seismic modelling of the Earth's large-scale three-dimensional structure". In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 328.1599, pp. 291–308.

Wookey, J and J-M Kendall (2004). "Evidence of midmantle anisotropy from shear wave splitting and the influence of shear-coupled P waves". In: *Journal of Geophysical Research: Solid Earth* 109.B7.

Xie, Jiayi et al. (2015). "Inferring the oriented elastic tensor from surface wave observations: preliminary application across the western United States". In: *Geophysical Journal International* 201.2, pp. 996–1021.

Xie, Jiayi et al. (2017). "Crustal anisotropy across eastern Tibet and surroundings modeled as a depth-dependent tilted hexagonally symmetric medium". In: *Geophysical Journal International* 209.1, pp. 466–491.

Xu, H and C Beghein (2019). "Measuring Higher-Mode Surface Wave Dispersion Using a Transdimensional Bayesian Approach". In: *Geophysical Journal International*.

Xu, Wenbo et al. (2008). "The effect of bulk composition and temperature on mantle seismic structure". In: *Earth and Planetary Science Letters* 275.1-2, pp. 70–79.

Yuan, Kaiqing and Caroline Beghein (2013). "Seismic anisotropy changes across upper mantle phase transitions". In: *Earth and Planetary Science Letters* 374, pp. 132–144.

— (2014). "Three-dimensional variations in Love and Rayleigh wave azimuthal anisotropy for the upper 800 km of the mantle". In: *Journal of Geophysical Research: Solid Earth* 119.4, pp. 3232–3255.

Zhang, Shuqing and Shun-ichiro Karato (1995). "Lattice preferred orientation of olivine aggregates deformed in simple shear". In: *Nature* 375.6534, p. 774.

Zhao, Dapeng, Akira Hasegawa, and Shigeki Horiuchi (1992). "Tomographic imaging of P and S wave velocity structure beneath northeastern Japan". In: *Journal of Geophysical Research: Solid Earth* 97.B13, pp. 19909–19928.

Zhdanov, Michael S (2002). *Geophysical inverse theory and regularization problems*. Vol. 36. Elsevier.

Zhong, Shijie and Michael Gurnis (1995). "Mantle convection with plates and mobile, faulted plate margins". In: *Science* 267.5199, pp. 838–843.

Zhong, Shijie et al. (2000). "Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection". In: *Journal of Geophysical Research: Solid Earth* 105.B5, pp. 11063–11082.

Zhong, Shijie et al. (2008). "A benchmark study on mantle convection in a 3-D spherical shell using CitcomS". In: *Geochemistry, Geophysics, Geosystems* 9.10.