IDEA League

Master of Science in Applied Geophysics

Research Thesis

Full seismic waveform inversion of Europe and western Asia

with optimally designed adaptive meshes

Carl Josef Schiller

September 29, 2023

Full seismic waveform inversion of Europe and western Asia

with optimally designed adaptive meshes

Master of Science Thesis

for the degree of Master of Science in Applied Geophysics

by

Carl Josef Schiller

September 29, 2023

IDEA LEAGUE
JOINT MASTER'S IN APPLIED GEOPHYSICS

Delft University of Technology, The Netherlands
ETH Zürich, Switzerland
RWTH Aachen, Germany

Dated: *September 29, 2023*

Supervisor(s):

Prof. Dr. Andreas Fichtner

MSc. Sebastian Noe

Committee Members:

Prof. Dr. Andreas Fichtner

MSc. Sebastian Noe

Prof. Dr. Florian Wagner

Abstract

In Geodesy, the development of accurate Digital Elevation Models of the surface helps to provide height data for various applications in other Earth Science disciplines. These models are used as base models for studies. Geophysics can be a support science when providing base models for non-invasive studies inside the Earth. Such models can be multi-dimensional velocity models, where the relation of velocities and depth can be used for other structural parameter studies in e.g. Geodynamics or Geochemistry. With technological advancements in the recent decade, there is incentive to provide new regional continental-scale models. Full-waveform inversion (FWI) is an imaging method that iteratively inverts through seismic waveforms for structural heterogeneities in the Earth to create a high resolution static image. The development of event-specific adaptive SmoothieSEM meshes for regional meshes leads to a substantial decrease in computational costs of the FWI by lowering the amount of mesh elements needed to model a wavefield. This allows to push regional continental-scale tomographies towards period bands that were previously cost-prohibitive. Divided into two tomographic studies, on a bigger continental-scale mesh of Europe and Western Asia and a smaller mesh focused on Europe respectively, the study areas are lowered to a period of 24s. The reference base model is derived from the current generation of the Collaborative Seismic Earth Model (CSEM). Starting at a minimum period of 50s, the mimimum period is pushed in steps towards the target of 24s in a multiscale inversion approach to enable monotonic misfit reduction. The misfit optimization of the synthetic and observed waveforms is based on time-frequency phase misfits while the gradient optimization is based on the trust-region limited memory Broyden-Fletcher-Goldfarb-Shannon (L-BFGS) method and dynamic mini-batches. Waveform fits are quantified by the reduction of the time-frequency phase misfit and normalized amplitude difference misfits between the end model and the CSEM base model. The waveform inversion reveals more heteregeneous structures in Europe and Western Asia compared to previous studies.

Acknowledgements

After 6 years of Studies, I am at the precipice of becoming a Masters degree holder. I am the first person in my family to get a university diploma in Science since my father. May he Rest in Peace, Friedrich Josef Schiller. There have been many bumps on this road. Some that have almost stopped me entirely. If it weren't for the people who have stayed at my side and have inspired me to press on further, I would not be here. It is because of all these people that I could even start and finish my Master's degree at some of my dream schools that I have been looking up to.

I would like to thank my mom, Felisa. Words can not describe how much I appreciate, owe to you and love you. You had to take care of me as a single mother from a young age on. Sacrifice almost all your comfort and familiarity. So that I can have a semblance of a good childhood. When you were standing in front of a difficult decision; Either go back to the Philippines, to a support system of family, or go back to Germany, with no support system, but more opportunities for your son. You chose my education and future above all else. Indeed words can not describe this feeling. That you have given me Life in more ways than once. It has been tough and we don't always agree on everything. But that is just how life is. So long as we have each other and appreciate each other. Thomas, Nicko, Denise and you have been my rock in trying times. Thank you.

I would like to thank my family friends. Karen, Jenny, Torsten, Domingo, Christiana, Luis and Meng. You have been my second family already for more than half of my life till now. Thank you for all the laughs and wisdom. You have made my experience growing up a blast. And I hope you will continue doing so.

Thank you to my family in Regensburg, Oma Justi und Opa Fritz, Tante Andrea und Gabi, Onkel Thomas und meine Cousins und Cousinen, Larissa, Katrin und Matthias. And also to the Berliners, Tante Christine, Onkel Martin and die Spaßvögel, Moritz and Konstantin.

Thank you to my family in the Philippines, Pinay, both Joshuas, Grace, Jehzel, Mira, Trish, Lawrence, Lorraine, R.E. Thank you for being there. I wish I could visit you more often. Stay healthy and see you soon. Rest in Peace Lola and Lolo.

A big thank you to my best friend Fynn. It is funny that by chance we met each other at school, after you came back from America. Your parents could have chosen any other school, but you chose our school. I sometimes still question that decision, but good that it

happened! I have not bonded with someone faster than with you. We had almost identical interests, talking about our niche hobbies and history and politics in the world; and most important video games. It is rare to find someone, where you know, no matter the distance or time passed, when you meet again time seems to not have passed and we have not aged one minute more since we became friends. We continue to evolve as our person, but it seems we evolve together in a similar direction. I do not know where I would be without you. Thank you for being in my life.
Now I need to retribute my old high school. A school that had the impossible task to shoulder the legacy as successor to one of the most prestigious public schools in Bonn, while dealing with the budgeting decisions of the city government. Safe to say, it could not muster up to it. But it did not need to. What I learned from the school is, that not everything can be perfectly done. That is just the reality of having multiple different voices and decisions in one place. But what you can do is accept it and make due with what you have. I appreciate the spirit and ideas the teachers implemented in the classes, so we could still have quality education in all subjects. I have learned to be open and receptive to the ideas and cultures of everyone, so we could manage the limited resources we had. Through the school I met many nice people and forged friendships that have lasted till even today. This school is the reason I even started studying as it fostered my everlasting desire of learning new things. By chance I entered this school in the fifth grade and I could not have imagined the defining experiences it has given me till high school graduation. Thank you to the faculty. And thank you to my friends for never having a boring moment.
I would like to thank some key people from my Bachelor studies, as they are the reason I went on this academic path. Thank you to Mario Valdivia-Manchego, for encouraging me to continue my studies after not doing so well in the first semester and encouraging me to go to Japan. Thank you to Thorsten Wörmann, may he Rest in Peace, for developing my interest in Applied Mathematics. Thank you to Anna Zoporowski, Andreas Kemna and Arne Mansfeld for their guidance and great teaching in Geophysics during my bachelors. I did not continue Geoelectrics, but through you I got on a path to Seismology, Numerical Modelling and Inverse Theory.
These last two years now I have met some incredible people. Thank you to everyone in my year group for these incredible moments we have created throughout these years. I hope to still see you again in the future. I would also like to thank all the professors from the partner universities for their incredible teaching and support. I have learned a lot from you. Not only in the courses. Thank you to everyone at all three universities. You made the experience much better. Your humility as well as willingness to listen to our issues and trying to find solutions for some our predicaments has made a lasting impression. Thank you!
Lastly I would like to thank the Seismology and Wave Physics Group for including me in their Group and their guidance during the Summer Project and my subsequent Master Thesis. You have given me so many ideas and good laughs in the last months, that I still have issues to process everything. I am working on that. Special Thanks to Sebastian Noe who did an absolute amazing job to support me during the Thesis. It would have been harder without him looking into my sometimes mundane mistakes. Special Thanks to DP for also looking into my sometimes mundane mistakes and helping me with a lot of the technical things during the Thesis. Thank you also to Christian, Patrick, Daniel and Sölvi for giving me feedback on certain things and on my Thesis. I enjoy our conversations and I hope we can have many more. And special thanks to Andreas and again Sebastian for supporting me through one of the hardest moments in my life. Thank you for your understanding and willingness to push

me so that I could start and finish the thesis even in this semester. And also special thanks to Andreas for supporting me in general and giving me opportunities I could not have imagined even a year ago.
Life is a series of chances. I can not describe how grateful I am to all the people that have pushed and supported me and given me all these chances in life. While I am proud of myself for this achievement, I am more proud to have met so many amazing people in my life. Thank you all!

ETH Zurich Carl Josef Schiller
September 29, 2023

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Earth, the Planet we call home, has a radius of approximately 6371 km and is drifting through the universe as part of the Solar System. Humans tend to look up and explore undiscovered areas in the universe, but there are many undiscovered areas beneath the surface of the Planet. Oceans or the Inner Earth are inaccessible for humans and other techniques to gain knowledge have to be considered. One of these techniques is Seismic tomography. Similar to medical tomography, where an image of a body part is made with electrodynamic energy sources, an image of the Inner Earth is made with mechanical seismic energy sources from naturally occuring Earthquakes. Seismic tomography is a problem, where we can not directly see physical systems happening. For this, knowing a physical system closely resembling the Earth system, we try to find out more about it through data. Resulting in a model of the Inner Earth and potentially giving clues about the processes inside. We can make a model out of cubes resembling the inside of the Earth and try to find out what the data in those cubes show. But the problem is tricky as data is not available everywhere. The distribution of Earthquakes with sufficient Magnitude and appropriate receivers are locked to major fault lines and relatively quiet land, respectively. Receiver data also has to be of high quality with clear arrivals or the investigation through the model gets in trouble. Therefore data is sparse. Due to the complex nature of waves that propagate through a planetary object like Earth, on an even grid there will be differing amounts of available information and not all grids will be filled similarly with information. Some areas are well covered with plenty of Energy propagating through them. Other areas are shadowed by the reflections and refractions that occur due to the heterogeneous structure of the Earth. One station at one point on the Earth won't be able to receive signals from every corner inside the Earth. This created a need to collaborate on a global level to compare data and to find avenues to gain more information out of previous constrained assumptions e.g. homogeneous layer models or simplified physical assumptions of wave propagation.

Historically, tomographic imaging started in the late 1970s for global- and local/regional-scale studies, where the former explored the deep mantle along a global array (Dziewonski et al. (1977)) and the latter the upper mantle structure (Aki et al. (1977)) along a regional array

using teleseismic P-wave travel-time measurements and ray theory (Romanowicz (2021)). Later with the introduction of Surface wave tomography, radial anisotropy had to be included in models (Nataf et al. (1984)). The first full waveform inversion applications were conducted to gain a shear-wave upper mantle model (Woodhouse and Dziewonski (1984)). Many commonly used 1-D velocity models, depth models that you can compare your results with, have been developed such that it can represent the velocity structure of the Earth with differing purposes. Some examples include AK135 for core phases (Kennett et al. (1995)), IASP91 for later arriving seismic-phases (Kennett and Engdahl (1991)), or for the overall structure the Preliminary Reference Earth Model (PREM) (Dziewonski and Anderson (1981)). The problem can be expanded with more data by including more Earthquake data from around the globe. More data means including information in one problem and computing it. That is costly, which imaging techniques are even beyond Earth Sciences. Implementation has to be balanced with available computing resources. Computing resources are redistributed depending on the hardware limitations of contemporary computers, which are changing in technology and architecture in speed that scientists barely can catch-up trying to properly implement algorithms taking full advantage of these advances in Computing (Bauer et al. (2021)). It is an arduous task to balance out the problem of creating an image between the resources of proper data implementation, computing power and costs, and human resources. This presented many research foci in tomographical problems. Should the research focus on the numerical methods and make them cheaper and more efficient to simulate, should it focus on the acquisition data, should it focus on the inverse problem itself to make it easier to compute and process necessary information kernels or should it simplify the physics to a point were necessary information is constrained while cutting out fluff (Tromp (2020)). Depending on the advancements of solved problems, the focii can shift between each other. Focus is necessary to develop and mature new methods in computational geophysics. For example, the simplification of wave propagation into directed ray paths were the basis of seismic tomography for decades (Liu and Gu (2012)) before it had been expanded by the same principles of simplification with finite-frequency kernels, also called Fréchet derivative kernels (Dahlen et al. (2000)), which can represent the volumetric sensitivity of seismic wave perturbations better than the ray-based principle. Modern waveform inversion can account for the three-dimensional heterogeneity along the finite-frequency ray path.

This study will be based on the Full-waveform inversion with adjoint methods. The development of adjoint methods enabled the exploration of the three-dimensional sensitivity of seismic waveforms to structures in different Earth Science studies (Tromp et al. (2005); Fichtner et al. (2006); Tape et al. (2007); Liu and Tromp (2008); Tape et al. (2009)). Adjoints in Seismology were first discussed in Tarantola (1984) and Tarantola (1988). It makes use of numerical 3D wave propagation e.g. realistic numerical wave propagation with the Spectral-Element-Method (SEM)(Komatitsch and Tromp (1999);Komatitsch and Tromp (2002a); Komatitsch and Tromp (2002b); Fichtner et al. (2009a)) to compute the Fréchet Kernels of the waveform misfits measurements that are sensitive to 3D variations in Earth material properties e.g. through diffraction, scattering, and multipathing. These kernels and their misfits over the adjoint method are iteratively minimized in a nonlinear scheme by improving the observed-to-synthetic data fit. This narrows down the true model according to the observed data starting from a base velocity model. The waveform misfit functional chosen is the time-frequency phase misfit (Fichtner et al. (2008)), which

seperates the observed-to-synthetic data misfit's amplitude and phase information and enables the adjoint sources to be expressed as a weighted phase and envelope difference, meaning that differences arising from interference effects of the Earth's structure that lead to waveform dispersion can be characterized. Thus, the overall description of the problem can account for sensitivity to heterogeneous anomalies important for current Dynamic Earth System studies (Ritsema and Lekić (2020)) and find out what some structures could be.

To infer high resolution models, individual anomalies have to be traced and mapped out within the model space. Depending on the imaging problem, the model will need an appropriate amount of source-receiver waveform data from many Earthquake events that intersect the study area. Data is commonly available from open-access data points around the world and needs to be properly queued for a studies purpose. The data needs to be of high quality for waveforms to add meaningful information as well as not to waste computing costs. Computing a forward and adjoint simulation in a three-dimensional cartesian mesh is costly. During an iteration every event and recordings has their forward and adjoint simulated, while checkpoints and gradients have to be stored for consecutive iterations (and possibly previous iterations if the optimization algorithm has to redo a gradient step). Nowadays, technology is at a point were Parallel Programming methods are used to compute onto several GPU nodes at the same time, utilising Supercomputers that can process and produce even higher amounts of data compared to 10 years ago due to exponential growth of computational power. More computation power means that inversion methods can be developed that can include more information. Nevertheless, data can not be brute forced and has to be elegantly included into the problem, necessatating additional research in proper HPC implementation to reduce computation time and energy consumption. For the FWI workflow, the Large-scale Seismic Inversion Framework, or LASIF (Krischer et al. (2015a); Thrastarson et al. (2021b)) in conjunction with the automated FWI workflow manager Inversionson (Thrastarson et al. (2021a)) is used while Salvus is the Spectral-Element solver driving the forward and adjoint simulations, data stochastics in the form of windowing and station weighting and mesh generation (Afanasiev et al. (2019)). Previous work has produced improvements in numerical wave propagation and HPC implementation of scientific codes that are being used in this study. To compute massive amounts of Earthquake data from hundred of events into one inversion problem, updating models through piecewise iterative updates in dynamic mini-batches has greatly reduced I/O issues aswell as issues in computation time by breaking up one tomographic problem into multiple more manageable smaller problems (van Herwaarden et al. (2020)). This enables the inclusion of a large dataset, as well as adding data evolutionary to one inversion problem (van Herwaarden et al. (2021)), without a penalty in increased computation time and disturbing the global minimum exploration of the model. Developments in wave-adapted- as well as data-adaptive meshing has reduced the computation time of global tomographies using the spectral-element method by multiple factors of difference, depending on the minimum period (van Driel et al. (2020); Thrastarson et al. (2020)). This could be achieved by elongating mesh elements along the azimuthal plane orthonomal to the wave propagation from a source in a simulation. Resulting in less elements compared to a cartesian mesh, while producing almost identical synthetic predicted waveforms. Implemented successfully in global tomography studies (Thrastarson et al. (2022)), more detailed regional studies have also been achieved through dynamic mini-batches. Wavefield-adapted meshes have not been implemented yet on an

regional continental-scale. The study focuses on the developments of wave-adapted and data-adaptive meshing for a regional study on Europe and Western Asia by implementing event-specific adaptive meshes in a dynamic mini-batch approach. Challenges concerned the meshing of the events during an iteration, if the meshes are cut according to the domain and using synthetic tests to investigate potential errors in the FWI workflow e.g. errors through improper placement of absorbing boundaries, mesh interpolation between event-specific- and master mesh that is iterated on, or improper inclusion of an event dataset through not including all source-receiver pairs in the automatic meshing. All of these issues had to be investigated first before implementation in production software.

The event-specific adaptive meshes have orders less elements than the cartesian chunk mesh they are derived from depending on the maximum frequency i.e. minimum period. The study will refer to periods, which is the reciprocal of the Earthquake waveforms frequency. This results in a similar reduction of simulation time for an event simulation. The scaling of compute time by number of events is therefore more favorable with event-specific adaptive meshes. Thus saving compute costs that can be redistributed for simulations at lower periods in the regional tomography than previously possible. The backdrop of this tomography will be previous Europe tomographies of the last decade, that were conducted with differing approaches (Zhu et al. (2012); Fichtner et al. (2013b)). The study also observed the complexity of a regional tomography concerning domain size, by first applying the method onto the bigger domain of Europe and Western Asia as a test and then conditioning the bigger tomography with a smaller domain focused on Europe using the event-specific meshes. Both meshes use real event data.

Preliminary results show that the method not only produces a valid regional continental-scale model, but also saves compute costs by either 6 to 24 factors, depending on the minimum period of investigation. This is based on current results. Further studies into lower periods may show higher orders of magnitude. The imaged structures in the model agree with previous geological as well as previous tomographical studies of the area. Geologically, it can reconstruct older subducted slabs from past models and ongoing tectonic processes. Some anomalies are unrelated to current geological knowledge. This presents a discussion, what the most likely nature of some imaged heterogeneities could be. Lastly, the inclusion of a 1-D attenuation Q-model in the simulations spur further investigations of the modelling accuracy at periods far lower than previously imaged. Attenuation is most sensitive to areas susceptible to temperature processes. Can the attenuation part model thermally active areas or will the method, the data or the mathematical representation fail first. Or even fail at all.

Part I

Development of Event-Specific Adaptive Meshes

September 29, 2023

Chapter 2

Tomography: Or How I Learned to Stop Worrying and Love the Image

This part will look into the framework of Numerical Modelling Inverse Problems in Seismology, specifically how it is applied to large-scale tomographic studies, and how it all fits into each other to achieve a coherent and scientifically sound model of the study area. It will explain the underlying theory, leading into the methods and the necessary cost-saving developments for the second part of the study, the low-period seismic tomography of Europe, to function properly. This includes differentiating what the goal of a model is and what the applied methodologies can and are achieving in a model study. Remember the limitations concerning computing, mathematical representation and data. Every model is simultaneously wrong and profound in what it can describe (Box (1976)). Further information on general Discrete Inverse Theory is provided by Menke (2018), while more specific information on waveform simulation and inversion in the context of different tomographical methodologies is provided by Fichtner (2010). For more information on numerical modelling methodologies and their development history in context of seismological applications refer to Igel (2017). Lastly, information on how to handle waveforms in the context of planetary inversion studies and characterising material and source parameters is provided by Kennett and Fichtner (2020).

2-1 Introduction to the Inverse Problem

Inverse Theory is part of the area of Probabilistic- and Optimal Control Theory used mostly in Applied Mathematics and adjacent scientific subjects. They are some of the most important problems in Natural Sciences as their solution can deliver information about parameters of unknown observables. Given some starting data and a physical principle that governs the data, determine estimates of model parameters. In contrast to Forward Theory which predicts the results of data, given model parameter estimates, the Inverse does the opposite starting

from observations. Drawing a conclusion based on current knowledge from observations, which trivially means inferring knowledge. However, the solution to an Inverse Problem is not trivial. Considering that our observation data is given as d_{obs}, the corresponding model parameters given as m and the Forward operator given as G, which maps model parameters onto itself to determine the state of a physical system and predict its outcome, observables should be produced as:

$$d_{obs} = G(m) \tag{2-1}$$

where data residuals, or misfit χ, that give indication whether the model is realistic, associated with parameters m are given as:

$$\chi = d_{obs} - G(m) \tag{2-2}$$

Discrepancies between the data and model in equation 2-2 can show that the physical system m is not modelled correctly and that an improvement may be necessary. To get the model parameters from observables, the basic Inverse problem can be reformulated from equation 2-1 to:

$$m_{est} = G^{-1} d_{obs} \tag{2-3}$$

If the system made up of these three basic parts is linear with the same dimensions, then it depends on the invertability of the Forward operator and if the solution is unique for a possible solution. Some problems are nonlinear and non-unique, which require a different viewpoint on an approach. Three major viewpoints are considered. A probabilistic approach, a deterministic approach or a wholly continuous approach, where the functions are not numerically discretised but explicitly continuous. Problems can also either be well-posed or ill-posed, where for the former the solution exists, is unique and the solution's behaviour changes continuously with initial conditions. The latter occurs more often, as most inverse solutions are sensitive to changes in the final data. Every problem in sciences uses similar fundamental theorems, but the results from individual problems look superficially different. This also includes seismic tomographic problems.

Seismic tomographic problems can be approached in many ways, as the way to obtain a model is arbitrary depending on the application realm and what data is given. Fundamentally the Earth problem itself can be further explained through first a continuous approach in form of the Backus-Gilbert Inverse Problem (Backus and Gilbert (1967)), although the vast majority of inverse problems are approached in a discrete way to account for heterogeneities. Requiering discretisation of normally continuous model function parameters. This comes with a penalty of inferences and models potentially being flawed by subjective choices made by the scientist e.g. the choice of basis functions for the discrete function expansion. A necessary evil, as this allows the usage of predictable vectors and matrices to represent a continuous system, which is difficult to represent as a model. Still, a continuous approach is advantageous as it can provide a top-down view of an issue. Backus and Gilbert wanted to study the resolving power of natural occuring linear functionals e.g. Earth's gravity field, giving insight on the admissable average data Kernel of multiple finite data collections. Their studies led to natural bounds of an inverse problem. First, the generalized problem is given as:

$$d = \int_V G_i(x) m(x) d^L x \tag{2-4}$$

where $d^L x$ is the volume element in the space of x, and the model function depends on several variables L. To discretize equation 2-4 and estimate the model function $m(x)$ at a specific point x' it has to be reformulated to account local averages m_i^{avg}:

$$m^{avg}(x') = \sum_{i=1}^{N} G_i^{-g}(x') d_i = \int R(x', x) m^{true}(x) d^L x \tag{2-5}$$

$R(x', x)$ is the resolving kernel, the averaging analogy of the model resolution matrix R_{ij}, which determines the goodness of the model parameters. In general, the highest resolution of a model is attained, when the resolution matrices are equal to identity matrices. The overall prediction misfit of the off-diagonal elements can be determined by measuring the L2-norm of the difference between the resolution and identity matrix. The misfit leads to the Dirichlet spread function for the model, which can be applied to data-, model- and unit covariances matrices. The spread of resolution R is given as:

$$spread(R) = ||R - I||_2^2 = \sum_{i=1}^{M} \sum_{j=1}^{M} [R_{ij} - \delta_{ij}]^2 \tag{2-6}$$

$R = I$ suggests, that the $spread(R) = 0$. There is an issue with the Dirichlet spread function, when analysing side lobes i.e. large amplitude regions in the resolution matrix that are far from the main diagonal. The Dirichlet spread function does not do well with the natural ordering of model and data parameters as off-diagonal elements are weighted equally regardless of distance to the diagonal. Preference is then given to a generalized inverse formulation without side lobes. This widens the band of non-zero elements along the diagonal of the resolution matrix. The solution to this resolution matrix can be interpreted as a localized average of adjacent model parameters. For this preference of deltalike resolution matrices to be enacted, a weighting factor to the measure of spread that weights the i-th, j-th element according to the element's distance from the diagonal is added, but where the choice of weighting depends on the dimensionality of ordering. As such, a more convenient choice is to choose the spread such that diagonal elements have no weight, which leads to $w(i, i) = 0$ and $w(i, j)$ being non-negative and symmetric on the elements. The spread function of the Backus-Gilbert generalized inverse becomes:

$$spread(R) = \sum_{i=1}^{M} \sum_{j=1}^{M} [R_{ij} - \delta_{ij}]^2 = \sum_{i=1}^{M} \sum_{j=1}^{M} w(i, j) R_{ij}^2 \tag{2-7}$$

This spread can now be used to to derive new generalized inverse problems.

2-2 Considerations in the tomographic Inverse Problem

Applying the newly derived continuous generalized inverse onto a tomographic problem, shows some cracks in the representation and that further trade-offs are unavoidable for finding a solution. From equation 2-5 it follows, that the optimal resolution in an averaging sense is achieved, when the kernel peaks near the target point x' with a given set of data kernels, $G_i(x)$. To know the goodness of the model resolution, the spread over the volume is quantified

with equation 2-7. The weighting function is zero at point x' and grows monotonically away from the point and behaves normally in a quadratic way. Inserting equation 2-5 into equation 2-7 gives the spread at point x':

$$\begin{aligned} J(x') &= \int w(x',x)R(x',x)R(x',x)d^Lx \\ &= \int w(x',x)\sum_{i=1}^{N} G_i^{-g}(x')G_i(x)\sum_{j=1}^{N} G_j^{-g}(x')G_j(x)d^Lx \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} G_i^{-g}(x')G_j^{-g}(x')\int w(x',x)G_i(x)G_j(x)d^Lx \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} G_i^{-g}(x')G_j^{-g}(x')[S(x')]_{ij} \end{aligned} \tag{2-8}$$

where

$$[S(x')]_{ij} = \int w(x',x)G_i(x)G_j(x)d^Lx \tag{2-9}$$

is the overlap integral that has a large value, when two kernels overlap. This continuous spread function is also analogous to a discrete spread function. To minimize the spread, the continuous form of the generalized inverse of the resolution is given as:

$$G_l^{-g}(x') = \frac{\sum_{i=1}^{N} u_i[S^{-1}(x')]_{il}}{\sum_{i=1}^{N}\sum_{j=1}^{N} u_i[S^{-1}(x')]_{ij}u_j} \tag{2-10}$$

where $u_i = \int G_i(x)d^Lx$. The aforementioned trade-offs involve the nature of data d. Knowing the data is determined up to a certain misfit quantified by the covariance matrix, the localized average $m^{avg}(x')$ is also determined up to a certain correspondent misfit. Minimizing the spread may lead to averages with large bounds and having a slightly less localized average may be more desirable. An arbitrary parameter, which varies between 0 and 1, can be used to quantify a relative weight given to the spread of resolution and the size of variance. The overlap integral , given in equation 2-9, can then be given as:

$$[S'(x')]_{ij} = \alpha\int w(x',x)G_i(x)G_j(x)d^Lx + (1-\alpha)[cov(d)]_{ij} \tag{2-11}$$

where α is the arbitrary parameter. This leads to a trade-off curve between spread of resolution and size of variance depending on the value of α (Figure 2-1).

Before discretising, there are further trade-offs that have to be considered when approximating the continuous problem as a discrete problem. To transfer the continuous representation to a discrete one, the assumption that the model function can be represented by a finite number of coefficients has to be made. The choice of functions has to implicitly provide priori information about the behaviour of the model, which places a certain amount of scrutiny on what choice should be made.

$$m(x) \approx \sum_{j=1}^{M} m_j f_j(x) \quad \text{with} \quad f_j(x) = H(x - j\Delta x) - H(x - (j+1)\Delta x) \tag{2-12}$$

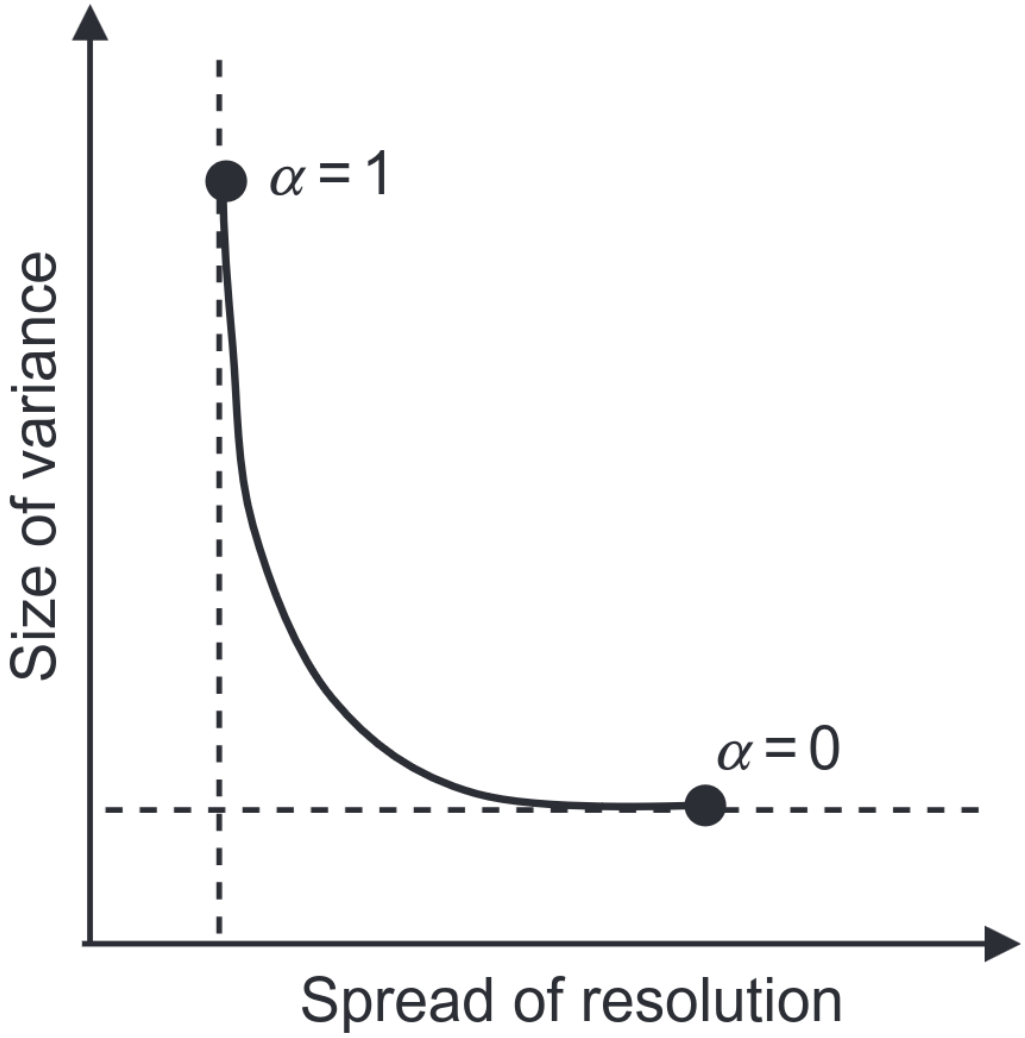

Figure 2-1: Trade-off curve for variation in $\alpha = [0, 1]$. Goal is to find alpha, such that there is good balance between amount of spread and the size in variance (Menke (2018)).

Here H is the Heaviside step function, which is zero for $x < j\Delta x$ and unity for $x > j\Delta x$, accounting for expansion inside the function space. Inserting equation 2-12 into equation 2-4 gives the discrete form of the continuous Backus-Gilbert problem:

$$d_i = \int G_i(x) \sum_{j=1}^{m} m_j f_j(x) d^L x = \sum_{j=1}^{M} \left\{ \int G_i(x) f_j(x) d^L x \right\} m_j = \sum_{j=1}^{M} G_{ij} m_j \tag{2-13}$$

which again gives the discrete form in the form of $d = Gm$. This basic form allows the application of several geophysical inverse problems as long as the physics are well described in the data Kernel, and the basic trade-offs in variance are kept in mind.

2-3 From Rays to Finite-frequency

2-3-1 Ray-theoretical approach

Backus-Gilbert gave a framework to resolve the data of the Earth model depending on the used model function in a continuous and then discretized representation. Early model functions for the tomography problem were continuous curved rays along a ray path. For more information refer to Liu and Gu (2012). In tomography the model function m is a function with two or more variables related to the data d_i as:

$$d_i = \int_{D_i} m[x(l), y(l)] dl \tag{2-14}$$

where the domain D is a curved ray integrated along a path of length l. Expanding as a standard continuous problem, where the data kernel G samples with the Dirac delta function $\delta(x)$, $G_i(x, y) = \delta[x(l) - x_i[y(l)]]dl/dy$. The Dirac delta function is a generalised function that maps every value in its domain to zero except at the origin, while the integral of its domain equals to one. As a linear functional, it maps a continuous function to the value at its origin introducing a weak limit, where the continuous function becomes a sequence of "bumps". Extending to a discretization of a continuous function, the delta function "samples" the value a discretized segment of a continuous function. However, this is not trivial and special cases should be kept in mind. Applying equation 2-4 to equation 2-14, d_i becomes:

$$d_i = \int \int m(x, y) \delta[x(s) - x_i[y(s)]] \frac{dl}{dy} dx dy = \int_{D_i} m[x(l), y(l)] dl \tag{2-15}$$

x varies with y along curve D and y also varies with length l. The goal of the inverse problem becomes the determination of the fractional change of velocity along curve segments and to achieve a good fit for the seismic signals in the forward problem. This forward problem involves computing the traveltime t of the curved ray as:

$$t = \int_{x_s}^{x_r} s \cdot dl \tag{2-16}$$

where x_s, x_r are the source and receiver locations respectively and s is the slowness i.e. the reciprocal velocity for every segment along the ray path. The forward problem holds for when

origin times or locations of the seismic sources are known, such that $\delta t = \int_{perturbered\ ray} \delta s \cdot dl \approx \int_{unperturbed\ ray} \delta s \cdot dl$, as otherwise the formulation needs to also account for perturbations in the earthquake location and origin time accruing (Menke (2018)). The Forward Problem follows Fermat's Principle, which states that that perturbations in the ray path cause only second-order changes in traveltime, an important fact for later derived tomography methods involving adjoints. Discretised the formulation becomes:

$$\delta t_i = \sum_{j=1}^{N_p} \delta s_j l_{ij} \tag{2-17}$$

where δt_i is the difference between observation and reference model prediction for ray i. This formulation is then expanded through linerisation and parameterisation for travel times to be sampled on a global basis. Orthogonal basis functions were adopted to describe the slowness perturbation as a set of linear functions with discrete unknown weights as well as back-projection for parametirisation. Early global traveltime inversions used simple approaches with P wave arrivals to determine isotropic wavespeeds. In gridded blocks j, the slownesses of all determined rays are summed and determined in the inverse problem. This approach helped to image homogeneities sensitive to P-waves (Grand et al. (1997); Van der Hilst et al. (1997)), but some factors limit the application of a ray path assumption from a continuous to discretised representation when it comes to sensitivity to local heterogeneities sensitive to surface waves (Backus (1964)). Scientists tried different heterogeneous mesh partioning as prior regionalisation schemes to address these local heterogeneities as well as the underdetermined data distribution in their study area. Issues arising from the regionalisation schemes were acknowledged with path-average approximations (PAVA) and global spherical harmonic expansion (Ritsema and Van Heijst (2000); Woodhouse and Dziewonski (1984)). In the grand scheme of tomographic inversion problems, these applications were a temporary solution for questions that could not be answered fully with travel-time tomography. Outcomes from well constrained global inverse problems should not be influenced easily by the choice of spherical harmonic functions or B-splines as they were more necessary for localisation issues (Bassin (2000)). When data coverage is sufficient, larger models have a 'same earth' assumption, meaning that more information from wavespeeds has to be gained differently than from only travel times. Given enough data, the output of the models can be controlled through resolution trade-off consideration and more focus is on resolution (Kárason and Van Der Hilst (2000)). Focus is then given on waveforms itself for inversion.

2-3-2 Finite-Frequency

With PAVA came the interpretation of the concept on lateral mantle heterogeneities and sensitivities to perturbations along the propagation path (Woodhouse and Dziewonski (1984)). It states that phase perturbations of surface waves are caused by these lateral heterogeneities when horizontally averaged over the source-receiver trajectory. Cumulatively sum local frequency shifts and perturbations arising along the great-circle path of propagation to a path averaged phase perturbation along the entire curved tube. These shifts and their partial derivatives with respect to model parameters are then used to compute a Fréchet Kernel, representing the volumetric densities, or sensitivities, of constitutive Fréchet derivative of the

model.
Given a generalized inverse in equation 2-4, redefine the model $m(x)$ in terms of aforementioned perturbations $\delta m(x)$ around the reference model $m^{ref}(x)$ as $m(x) = m_{ref}(x) + \delta m(x)$ (Menke (2018)). The data d_i along a path can also be redefined as $d_i = d_i^{ref} + \delta d_i$, where, given also a changing data Kernel G_i along the path, $d_i^{ref} = (G_i, m^{ref})$ is the reference predicted data. This results in an equation that says that a perturbation in the model causes a perturbation in the data, as long as it is sensitive to an entry in the model space:

$$\delta d_i = \int G_i(x)\delta m(x)dx = (G_i, \delta m) \tag{2-18}$$

The discretised form then follows as:

$$\Delta d_i = \sum_{i=1}^{M} G_{ij}^{ref} \Delta m_j \tag{2-19}$$

where $G_{ij} = \frac{\partial d_i}{\partial m_j}|_{m_{ref}}$. With the continuous function $\delta m(x)$ the data Kernel in equation 2-18 is the Fréchet derivative:

$$G_i(x) = \frac{\delta d_i}{\delta m}|_{m^{ref}} \tag{2-20}$$

Through this formulation, synthetic seismograms could be computed by normal mode superposition and with PAVA a linear inverse problem of model perturbations could be formulated (Romanowicz (2008)). The usage of this technique enabled Kernels to be sensitive to heterogeneities. In theory PAVA is sensitive to phase and amplitudes of surface waveforms, but in practice perturbations are dominated by phase differences than by amplitude differences in waveform inversions. Another approach that was suggested to circumvent the aforementioned issues, was inverting in multiple stages in form of Partitioned Waveform Inversion (PWI) (Nolet (1990)). PWI lays the first conceptual foundation of approaching an inverse problem in seismology in multiple-stages and consists of two steps, which are (1) fit waveforms with an average reference model along the path between a source and receiver and (2) solve for 3D-structure via linearized inversions and the PAVA misfits from step 1. Step 1 is the most important step in the multi-stage inversion and requires potentially lots of computational resources not only for the forward simulation, but also for the nonlinear optmization given that, on Earth, data availability is sparse, underdetermined and the model area is multidimensional.
Finite-frequency effects have to be well considered in the overall inverse problem, as ray-theory itself relies on the assumption of weak and relatively large-scale lateral heterogeneities, while the sizes of Fresnel zones are more significant for finite-frequency arrivals from broadband recordings, which need to be considered in tomographic inversions. Fitting Fréchet Kernel theory needs to be applied.

2-4 Towards Adjoints, banana-doughnuts and FWI

The following section will detail improvements to aforementioned issues. Adjoints were theorised with the advent of appropriate numerical modeling techniques and more computational resources. The development of numerical simulations of seismic wave propagation for

complex and realistic three-dimensional heterogeneous media at multiple scales made accurate synthetic forward modelling possible and the usage of corresponding misfit functionals. Many factors such as attenuation, rotation, gravity forcing, topography and ocean loading could also be included with available resources. Additionally, adjoint methods became more popular in adjacent Earth Science fields like meteorology (Errico (1997)) or geodynamics (Bunge et al. (2003)). This made fitting finite-frequency Fréchet Kernels and the first applications of Full-waveform inversion (FWI) possible. Cornerstone is the computation of Fréchet derivatives of parameters through the interplay between a forward modelled wavefield in a reference model and the adjoint wavefield, which creates gradients of earthquake sources that correspond to the parameters. The advantage of adjoint methods is its efficiency. While in previous methods there were long approximations of the first derivative of the observed data involved, the adjoint method can determine it by solving the forward problem and the corresponding adjoint. Between the forward and adjoint is a misfit term in the gradient, that can be optimised through iteratively fitting the synthetic model with the data of the reference model using an appropriate misfit functional that adjusts and minimises the misfit between iterations. The seismic inverse problem may be treated as an optimisation problem. These gradients are Fréchet derivatives of the misfit function.

2-4-1 Adjoint method

Given a misfit χ with respect to the reference model, where data $d(x)$ is continuous, the misfit is analog to an L2-norm discrete misfit $\chi = (d^{obs} - d)^T(d^{obs} - d)$. The perturbation of the misfit $\delta\chi$ is given as:

$$\delta\chi = \chi - \chi^{ref} = \left(\frac{\delta\chi}{\delta m}|_{m^{ref}}, \delta m\right) \tag{2-21}$$

If the data is related to the model through a linear operator $d = Lm$, then the perturbation of the misfit is given as:

$$\begin{aligned}\delta\chi = \chi - \chi^{ref} &= (d^{obs} - d, d^{obs} - d) - (d^{obs} - d^{ref}, d^{obs} - d^{ref}) \\ &= -2(d^{obs} - d^{ref}, d - d^{ref}) + (d - d^{ref}, d - d^{ref}) \\ &= -2(d^{obs} - d^{ref}, \delta d) + (\delta d, \delta d) \approx -2(d^{obs} - d^{ref}, \delta d) \\ &= -2(d^{obs} - d^{ref}, L\delta m)\end{aligned} \tag{2-22}$$

Now applying a hermitian adjoint operator instead of the linear operator, where the definition of the adjoint is $< Ax, y >=< x, A^*y >$, it follows that equation 2-22 is also:

$$\delta\chi = (-2L^*(d^{obs} - d^{ref}, \delta m)) \tag{2-23}$$

Applying equation 2-23 on equation 2-20 yields the Fréchet derivative of the misfit:

$$\frac{\delta\chi}{\delta m}|_{m^{ref}} = -2L^*(d^{obs} - d^{ref}) \tag{2-24}$$

where the discrete analog is:

$$\frac{\partial\chi}{\partial m_i}|_{m^{ref}} = -2G^T(d^{obs} - d^{pre}) = -2G^T(d^{obs} - Gm) \tag{2-25}$$

and where the misfit χ can be minimized with gradient methods.
Knowing the Fréchet derivatives of the parameters aswell as the Fréchet derivatives of the misfits corresponding to the parameters is the basis for modern waveform tomography enabled through adjoint methods. Dahlen et al. (2000) first described the corresponding Kernels and their banana-doughnut shape, while Tromp et al. (2005) and Fichtner et al. (2006) showed important advantages and applications of the adjoint method in connection to tomographic problems. Assume the gradient in equation 2-25 is set to 0, resulting in the least-square equation for the parameters $G^T G m = G^T d^{obs}$. In the continuous case of equation 2-24 this results in $L^* L m = L^* d^{obs}$. Introducing an identity operator I that satisfies $m = Im$, where $m(x) = \int_{-\infty}^{\infty} \delta(x - x') m(x') dx'$, then the continuous case becomes:

$$m = Im = L^* d^{obs} - (L^* L - I)m \tag{2-26}$$

In the case where $L^* L = I$, equation 2-26 can be seen as a recursion relating an older parameter estimate to a newer one, where i denotes the index of the considered iteration:

$$m^{i+1} = L^* d^{obs} - (L^* L - I)m^i \tag{2-27}$$

so if the recursion starts with the reference model $m^{ref} = 0$, then the parameter estimate becomes $m^{est} = L^* d^{obs}$. Now suppose the linear operator is an indefinite integral such that:

$$d^{obs}(x) = Lm(x) = \int_{-\infty}^{x} m(x') dx' \tag{2-28}$$

and

$$m^{est}(x) = L^* d^{obs}(x) = \int_{x}^{\infty} d^{obs}(x') dx' \tag{2-29}$$

This yields a quite good approximation of the first derivative in a simplified tomography problem from the resulting adjoint by inverting the indefinite integral with another integral. Accuracy of the adjoint is dependent on the residuals in $L^* L - I = 0$. Equation 2-29 states that slowness at a point x is estimated through summing up all the traveltimes of the rays that can sample the slowness and interact with the parameters at x. The summation of the averages introduces only long-wavelength errors in the image.
Continuing, for 3D tomographic problems, representing the Fréchet Derivative of the data through a field, or physical observable, $u(x)$ represented by a Differential Equation can better analyse the inverse problem. Depending on the related parameters, the relationship overall relates a linear operator on the field with the model parameters, which also holds for their respective perturbations, with known boundary conditions:

$$Lu(x) = m(x) \; and \; L \; \delta u(x) = \delta m(x) \tag{2-30}$$

and the data relates to the field in an inner product as follows:

$$d_i = (h_i(x), u(x)) \; and \; \delta d_i = (h_i(x), \delta u(x)) \tag{2-31}$$

where $h_i(x)$ is a known auxiliary function in the inner product. Combining both of these equations yields:

$$\delta d_i = (h_i, \delta u) = (h_i, L_{-1} \delta m) = ((L^{-1})^* h_i, \delta m) = ((L^*)^{-1} h_i, \delta m) \tag{2-32}$$

Comparing equation 2-32 and equation 2-18 gives:

$$G_i(x) = (L^*)^{-1} h_i(x) \; or \; L^* G_i(x) = h_i(x) \tag{2-33}$$

which shows that the field satisfies equation 2-30 and the data kernel equation G_i satisfies the adjunct equation 2-33. The field and corresponding adjoint can be any physical field described by a differential equation, given an appropriate numerical method as a solver and appropriate boundary conditions of the domain depending on the modelled physical phenomena.
There are multiple ways the adjoint can be described and obtained in FWI and the choice is at the discretion of the scientist and which one fits better to the study being conducted. For this study, the modelled Partial Differential Equation is the elastic wave equation:

$$\rho(x)\frac{\partial^2 u(x,t)}{\partial t^2} - \nabla \cdot (C(x) : \epsilon(u)(x,t)) = f(x,t) \tag{2-34}$$

where ρ denotes density, $\epsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$ is the strain tensor, and C is the fourth-order elastic tensor relating strains to stresses. Due to symmetry relations $C_{ijkl} = C_{klij} = C_{jikl}$, the tensor reduces to 21 independent parameters in the three-dimensional case. Another possibility is wavefield seperation into P- and S- wave components respectively by taking the divergence and the curl of the displacement field (Zhou et al. (2022)). The wavefields can further be seperated between horizontal and vertical P- and S-wave parts respectively.

2-4-2 Banana-doughnut kernel

With these ingredients, an waveform tomography problem solving the Fréchet derivatives of the perturbations of velocity parameters $\frac{\delta u}{\delta v_P}, \frac{\delta u}{\delta v_S}, \frac{\delta u}{\delta \rho}$ according to the different scalar wavefields as well as their respective misfits $\frac{\delta \chi}{\delta v_P}, \frac{\delta \chi}{\delta v_S}, \frac{\delta \chi}{\delta \rho}$ can be computed via adjoint methods (Menke (2018)). The perturbations in material constants anywhere in the model will perturb the wavefield at every observer, according to finite-frequency theory. This also underlines the main difference between ray- and finite-frequency traveltime theory, wherein ray theory perturbations are limited to the ray path itself. At intermediate periods, seismic wave fields have arriving phases with a traveltime which are "bump-like" in nature similar to the ones captured by the delta function. Ray-theory does predict that the main effect causing perturbations in slowness also means that a certain point along the ray, the perturbation either advanced or delayed the phase arrival, which is the main aspect carried over to finite-frequency theory. Define a differential traveltime ΔT by comparing time windows of an observed and predicted wave field at an observed phase and determining if they are rather advanced or delayed from each other and cross-correlate between the observed and predicted phase:

$$\Delta T = argmax \; c(\tau) \tag{2-35}$$

where

$$c(\tau) = \int_{t_1}^{t_2} u^{pre}(t-\tau) u^{obs}(t) dt \tag{2-36}$$

The phase arrives in the time window between t_1 and t_2. The point where time lag τ is at its maximum in the cross-correlation is also the travel time difference ΔT. Introducing a

perturbation inside the wavefield, such that $u^{pre}(t) = u_0(t) + \delta u(t)$, the resulting change will also be in the time difference $\Delta T = \Delta T_0 + \delta T$:

$$\delta T = \frac{1}{N} \int_{t1}^{t2} \dot{u_0}(t) \delta u(t) dt \tag{2-37}$$

where $N = \int_{t1}^{t2} \ddot{u_0}(t)\ u_0(t)\ dt$. A window function $W(t)$ enclosing the seismic arrival that is being looked at in a certain time frame, and where everything outside the window is 0 can be such that the resulting perturbation of the time difference is:

$$\delta T = (h(t), \delta u(t)) \tag{2-38}$$

where the auxiliary function $h(t) = \frac{1}{N} W(t)\ \dot{u_0}(t)$ and $N = \int_{-\infty}^{+\infty} W(t)\ \ddot{u_0}(t)\ u_0(t) dt$. Equation 2-38 is of the form of equation 2-31 and the Fréchet Derivative can be calculated with the equations following equation 2-18. What these equations show is that the kernel becomes the banana-doughnut kernel when the period of the seismic waves is decreased, as the high-amplitude part of the derivative becomes more and more concentrated along the ray path giving it its curved shape. Characteristic of the banana-shape is that the largest amplitude of the Fréchet derivative are near the ray path, but is 0 on the ray path itself, as the kernel lacks an intrinsic length scale. Heterogeneities can only perturb the travel time difference to the extent that the travel time from source to heterogeneity to observer is showing differences compared to the difference only between source and receiver, meaning that heterogeneities located inside the ray path have the same difference and the derivative is 0. Wave interference effects also have to be accounted for as they are noticeable through sign reversals in the velocity perturbation near the ray path, which can cause irregularities as to whether there is an advance or delay in traveltime at different locations along the kernel. Finite-frequency theory and Banana-doughnut kernels have certain improvements in sensitivities over ray theory since ray theory is valid only when the wavelength of the waves are smaller then the spatial scale of the velocity heterogeneities. For long-period wavelengths this condition fails, which for studying heterogeneous structures of the Earth proved problematic. Banana-doughnut kernels can circumvent these issues and work well with long perod wavelengths. The conceptual mathematical framework of adjoints is important to start describing the seismic tomographic methodology of Full-waveform inversion as it involves many parts to iteratively get a sharp image of the area of study. A Fréchet Kernel can be computed in four steps:

1. Solve the forward wave equation for the predicted wavefield u^{pre}
2. Compare observed and predicted wavefield using an appropriate misfit functional
3. Solve the corresponding adjoint
4. Multiply the derivative in time of both fields and integrate again in time, resulting in the Kernel.

Between iterations, the misfit, of the parameters between observed wavefield u_i^{obs} and the predicted wavefield u_i^{pre}, has to be minimized. This is a non-trivial issue for multiple reasons. First, the auxiliary function h is determined by the used misfit functional. Secondly, different misfit functionals gain different results as they reduce the misfit according to some function parameter set in the wavefield e.g. phase, amplitude or in the case of the L2 Misfit a metric like

the distance between the observed and predicted wavefield. Lastly, between every iteration, where the four steps are repeated, the model has to be optimized to find the best path for misfit reduction. Since this is a nonlinear optimization problem, appropriate gradient methods have to be chosen. This is also non-trivial as per the No-Free Lunch theorem, which states that there is no shortcuts in the choice of an optimization problem as every problem may have certain conditions and there is no simple "best" method that can be used for every problem.

2-5 Misfit functionals and additional information

Derive the L_2 difference between the observed and predicted wave field:

$$\chi = \sum_{i=1}^{M} \sum_{j=1}^{N} [u_i^{obs}(x_j, t) - u_i^{pre}(x_j, t)]^2 dt \tag{2-39}$$

where M denotes the number of parameters and N denotes the set of locations, like receiver locations. The integral can also be in a given time window/interval. Then the Fréchet derivative of the misfit can be calculated through equation 2-24 or equation 2-25 respectively. The Fréchet derivative of the misfit is important for the numerical computation of the resulting model perturbation in a FWI iteration. As shown starting from equation 2-26 the main advantage of calculating the misfit derivative is to determine the backprojected adjoint source and use the direct adjoint of the wave field, which then leads to the derivative of the misfit perturbation as well as the model perturbation without having to implicitly calculate the wavefield perturbation. The Misfit Fréchet derivative is implicitly the resulting banana-doughnut kernel, which leads to considerations about the choice of Misfit functional and their effect on the sensitivity analysis in waveform tomography. These considerations are (Fichtner (2010)):

1. Complete quantification of the seismic waveform misfit in the frequency range of interest.
2. Seperation of phase and amplitude information.
3. Relaxation of the requirements on waveform similarity needed for the measurement of pure cross-correlation time shifts.
4. Possibility to analyse complete wave trains including body waves, surface waves and interfering phases.

2-5-1 Time-Frequency Analysis

Advantageous is the modularity of the generic form of the Fréchet Kernel; they do not depend on the measurement details of the observed data and the definition of the misfit. But the misfit functionals do control the inversion direction and as to what information is included between each iteration, the possible invertable parameters and the type of nonlinear problem. It is important, depending on the data, that the functional can obtain the best trade-off between exploitable information and computational feasability. Examples are: L_2 waveform misfit, instantaneous phase and envelope misfits, waveform correlation misfits, optimal transport measures, cross-correlation time shifts (Luo and Schuster (1991)) and time-frequency misfits.

In this study the time-frequency phase misfit is used (Fichtner et al. (2008)).
Knowing the wavefield is $u_i^{obs}(x,t)$, where $x = x_r$ at a recorded position x_r, compute the Fourier Transform at the recorded position and omit the dependency on x as well as multiply the wavefield in the frequency domain with a sliding window function $h(t-\tau)$ centered around the maximum cross-correlation time lag τ to track the evolution with time:

$$\tilde{u_i}^{obs}\ \mathfrak{F}[u_i^{obs}](t,\omega) = \frac{1}{\sqrt{2\pi}||W||_2}\int_{-\infty}^{+\infty} u_i^{obs}(\tau)W*(\tau-t)e^{i\omega\tau}d\tau \tag{2-40}$$

where the window function norm $||W||_2 = \sqrt{\int_{-\infty}^{+\infty} h^2(t)dt}$ is assumed to be non-zero. The predicted wavefield is also represented analogous as a Fourier Transform of itself $\tilde{u_i}^{pre}(t,\omega) = \mathfrak{F}[u_i^{pre}](t,\omega)$. These are the time frequency representations of the observed and predicted wavefield. The exponential forms of $\tilde{u_i}^{obs}$ and $\tilde{u_i}^{pre}$ are respectively $\tilde{u_i}^{obs} = |\tilde{u_i}^{obs}|e^{i\Phi_i^{obs}(t,\omega)}$ and $\tilde{u_i}^{pre} = |\tilde{u_i}^{pre}|e^{i\Phi_i(t,\omega)}$ with the phases $\Phi_i^{obs}(t,\omega)$ and $\Phi_i^{pre}(t,\omega)$. The phase misfit is defined as a weighted L_2 norm of the phase difference $\Phi_i^{pre} - \Phi_i^{obs}$, which can be interpreted in terms of a time shift at a certain frequency $\delta\Phi_i = \Phi_i^{pre} - \Phi_i^{obs} = \omega\delta t$:

$$\chi_p = \sum_{i=1}^{M}\sum_{j=1}^{N}\int_{-\infty}^{+\infty} w_p^2(t,\omega)[\Phi_i^{pre}(t,\omega) - \Phi_i^{obs}(t,\omega)]^2 dtd\omega \tag{2-41}$$

where w_p is a positive weighting function. A phase difference includes some unavoidable consequences that occur with circular frequencies e.g. at a given time t, discontinuities in the phases can occur at different frequencies, where their locations jump around the frequency axis. This can lead to differences of $\pm 2\pi$ even at identical data. A solution can be defined under the assumption that the observed and predicted data are approximately in phase. Given a fixed time t the time-frequency representations of the observed and predicted wavefield are the Fourier transforms of the functions $f_t^{obs}(\tau) = u_i^{obs}(\tau)h(\tau-t)$ and $f_t^{pre}(\tau) = u_i^{pre}(\tau)h(\tau-t)$. Define the cross-correlation between the two functions:

$$c(\tau) = \int f_t^{obs}(t')f_t^{pre}(t'+\tau)dt' \tag{2-42}$$

where the Fourier transform is:

$$\begin{aligned}\mathfrak{F}[c](\omega) &= \frac{1}{\sqrt{2\pi}}\int c(\tau)e^{-i\ \omega\tau}d\tau = \sqrt{2\pi}\tilde{u_i}^{pre}(t,\omega)\tilde{u_i}^{obs}(t,\omega)^* \\ &= \sqrt{2\pi}|\tilde{u_i}^{pre}||\tilde{u_i}^{obs}|e^{i(\Phi_i^{pre}-\Phi_i^{obs})} = |\mathfrak{F}[c](\omega)|e^{i(\Phi_i^{pre}-\Phi_i^{obs})}\end{aligned} \tag{2-43}$$

The phase difference can be then expressed as:

$$\delta\Phi_i = -i\ ln\left(\frac{\mathfrak{F}[c]}{|\mathfrak{F}[c]|}\right) \tag{2-44}$$

For a given time in the regions of time-frequency space where u^{obs} and u^{pre} are approximately in phase, only phase differences in range $[-\pi,\pi]$ are admitted. This however still breaks down for when wavefields are highly dissimilar, where the observed and syntehtic waveforms have to either be low-pass filtered or applying a weighting function w_p that can clinch occuring dissimilarities. The choice of weighting function is important to suppress phase differences in time-frequency space occuring in certain regions where meaningful physical representation

breaks down. It helps to exclude time intervals with signals either at 0 or below the noise level, while emphasizing time-windows where higher-mode surface waves can be isolated and improve depth resolution. The weighting function then looks like this for average problems:

$$\widetilde{w_p}(t,\omega) = 1 - e^{-|\tilde{u}^{obs}(t,\omega)|^2/n^2(t,\omega)} \tag{2-45}$$

Time-frequency misfits are considered over cross-correlation time shift- and L_2 amplitude misfits, as the latter do not fare well with several phases interfering while the observed and predicted data are similar. The presence of three-dimensional heterogeneities leads to waveforms dispersing, and single phases being distorted between observed and predicted data. This phase distortion has to be rectified with a suitable misfit, that can capture the amount of distortion for optimisation, which cross-correlation and L_2 do not deliver the physical meaningful information. Therefore, the application of a time-frequency phase misfit, will be able to exploit use and robust information about the Earth's strucutre in waveform tomography. This enables to also potentially include phase related information from different wave types in observed data such as surface waves.

2-5-2 Surface waves

Till now the previous sections were mainly written with the exploitation of body waves in mind. Surface waves, such as Rayleigh- or Love waves, can potentially add more information of the Earth (Menke (2018)). Rayleigh- and Love waves are respectively trapped vertically- and horizontally polarised seismic energy with a velocity gradient increasing by depth. The seismic velocity of surface waves is strongly frequency-dependent, where lower frequencies extend deeper into the earth than the higher frequencies. These are dispersive waveform characteristics caused by impulsive sources, like Earthquakes, and the influence of heterogeneities on surface wave propagation, which leads to the component frequencies propagating at different phase velocities. Phase velocity ν is defined as $\nu = \frac{\omega(k)}{k}$, where ω is angular frequency and $k = \frac{2\pi}{\lambda}; \lambda = \frac{f}{v}$ is the wavenumber. The function $\nu(\omega)$ gives the dispersion curve of the surface wave. It is measured by earthquake wavefields and be predicted by vertically stratified models. For the Inverse Problem, further derivation of parameters and their Fréchet derivatives are then additionally done through the dispersion function. Consider now a Love wave, which can observe resulting perturbations in phase velocity from heterogeneities through perturbations in the shear velocity parameter v_S and density ρ, the perturbation δc over itself is given as:

$$\left(\frac{\delta c}{c}\right)_\omega = \frac{\int_0^\infty [k^2 l_1^2 + (\frac{dI_1}{dz})^2]\delta(v_S^2 * \rho)\ dz - \int_0^\infty [\omega^2 l_1^2]\delta\rho\ dz}{2k^2 \int_0^\infty (v_S^2 * \rho)_0 l_1^2 dz} \tag{2-46}$$

with assumed vertical stratification dz and $u_y = l_1(\omega, k, z)e^{(ikx - i\omega t)}$ is the Love wave displacement. Equation 2-46 is analogous for Rayleigh waves, just that Rayleigh waves consider additionally the compressional wave. This shows that relevant information is also encoded in the frequency-dependent time shift between an observed and predicted surface wavefield and corresponding Fréchet kernels as well as adjoints can be calculated. When adding surface waves to the finite-frequency theory, consider that it would be propagating along a horizontal ray path between a source and receiver. The propagation characteristics of surface waves are

determined by the vertical structure beneath every point along the propagation path (Woodhouse (1974)). The path approximation for each dispersion function along a source-observer path is:

$$\frac{1}{c_i(\omega)} = \frac{1}{D_i} \int_{path\ i} \frac{dl}{c(\omega, x, y)} \tag{2-47}$$

Here, the dispersion function $c(\omega, x, y)$ is extended with positional variables to relate it to line integrals, while D_i is the length of the entire path. The dispersion function can be inverted for estimates of the vertical structure.

2-5-3 Attenuation

Another quality that is considered for heterogeneities is seismic attenuation (Tarantola (1988)), the amplitude loss seismic waves experience due to frictional laws within the Earth. Amplitude loss can be characterized by $\alpha(\omega, x)$, varying with angular frequency ω and position x. When finite-frequency is accurate the fractional loss of Amplitude A is then:

$$\frac{A}{A_0} = e^{-\int \alpha(\omega,x) dl} \tag{2-48}$$

where A_0 is the inital amplitude. The attenuation is assumed to only vary with depth, and can be included through surface wave inversion. The initial amplitude can pose problems, as it is a wildcard on what can be considered the initial amplitude in waves with variational amplitudes like surface waves. When considering it as an unknown parameter, the nonuniqueness problem is similar to having an unknown origin time in classic tomographic problems. It has to be set sensibly. In higher frequencies, the attenuation effects should be more noticeable.

2-6 Iterative Waveform Inversion of Event Data

For this study, the goal is a FWI with real observed Earthquake event data in a regional tomography study. Over the last decade, adjoint method based waveform inversion have matured to gain many more insights into the Earth's mantle. It started with exploiting all the available information data content in the full waveforms, including ambient noise and instrument noise (Fichtner (2010)). Hence the portmanteau Full-waveform inversion. Theoretically, to have the perfect conditions, is to know exactly what the nature of the noise is at the time of recording to neutralise it from the data, which is not going to be possible and for many cases is also not necessary. IRIS reported that by now their servers host 660TB of seismic data. This includes high quality Earthquake data, which is far more abundant compared to twenty years ago, and adding as many independent events as possible into one inversion problem is necessary to gain high resolution tomographic models. The increase in computational resources and the development of robust numerical methodologies made the realisation of realistic three-dimensional seismic-wave propagation feasable (Komatitsch and Tromp (1999); Komatitsch and Tromp (2002a)). Over the years many studies have been carried out to analyse and improve on the selection of appropriate data processing in the FWI approach (Fichtner et al. (2009b); Tape et al. (2009)), treatment of the shallow heterogeneous crust, multi-scale approaches, and resolution analysis (Fichtner and Trampert (2011);

van Herwaarden et al. (2023))
The selection process for event data is mainly governed by the fact, that waveform inversion with the adjoint method is computationally expensive, due to the Forward and Adjoint modelling with the Spectral-Element method on large meshes. As such event data should be of high quality, where noise can be neglected, as well as having as many event-receiver pairs as possible since the costs of adjoint simulation are independent of the amount of receivers.
Due to issues like cycle-skipping (Kennett and Fichtner (2020)), which occurs when the predicted and observed data are more than half a period out of sync and then leads to time offsets, the choice of the initial model has some frequency-dependent requirements. The initial model has to be tuned and close enough to the optimum, so that phase and time alignment have a good benchmark to follow.
Another issue is the crust due to it in some areas around the world being thinner then the shortest seismic wavelengths of interest to image it sufficiently. The dominant wavelength of seismic waves is $\lambda = \frac{v}{f}$, where v is velocity and f is the dominant frequency. The velocity varies increases with depth, while the dominant frequency decreases with depth. The threshold for vertical resolution is given as $\lambda/4$, the quarter-wavelength criterion (Yilmaz (2001)). The crustal thickness of oceanic crust varies between 5 to 10 km and the thickness of continental crust varies between 30 to 50 km. Given a frequency at $f \approx 0.033\ Hz$, which is a period of 30s, and a shear reference velocity of $v = 4\ km/s$, the wavelength criterion is $\frac{\lambda}{4} = \frac{4\ km/s}{4*0.033\ Hz} = 30\ km$. This is not sufficient to image fine-scale heterogeneities in the crust. According to Fichtner et al. (2009b), periods longer than 30s are dominated by surface waves, and the vertical scale for surface waves determined by Tape et al. (2009) is 3 km. The crustal problem results in having to sufficiently discretise the intra-crustal discontinuities like the moho with smaller elements at periods shorter than 30s, leading to an increase of the computational costs. This is also a reason, why three-dimensional studies of certain study areas have not been possible at high frequencies due to their sheer size.
All these issues are subject to efficient information exploitation, for which, in an iterative nonlinear inversion problem, an appropriate optimisation method of the waveform misfit is needed. The aforementioned time-frequency phase misfit functional is used for its benefits regarding emphasis of phase information over amplitude information. For the phase misfit functional to work well with the multitude of receiver information in one event data set, well placed windows have to be set for every source-receiver pair waveform. This is necessary, as noise can contribute in such way that certain amplitudes in a waveform may not be discernable between noise and high-frequency surface wave information as well as low-frequency body wave information to avoid cycle-skipping. Mostly focused on the time window, where the earliest and latest arrival of a seismic phase, given the epicentral distance of a source-receiver pair, is plausible. During inversion a multi-scale approach is chosen. After each iteration, in comparison to the reference model, the smoothed gradients indicate the direction to find the best path to the global minimum while not getting stuck in a local minimum. To accelerate convergence, steps are taken between each period band such that the lowest available frequencies are included first before going into higher frequencies and meanwhile avoiding cycle-skipping.
To avoid cycle-skipping as much as possible, the lowest period band is changed in a logarithmic manner, such that the observed and predicted data still sufficiently overlap. For example, if the current period band is 35s-90s, lower the period band to 28s-90s by about 20%. How big the steps should be chosen is at the discretion of the scientist, given sufficiently small

steps. For the number of required iterations two things have to be considered. First, a certain amount of iterations should lead to misfit reduction to enable a jump in periods without cycle-skipping. Secondly, the number of iterations should not exceed a certain amount to avoid noise being fitted. During misfit reduction, fast convergence, beyond the period band the initial model was tuned for, could be an indicator of aggressive regularisation i.e. smoothing lengths. Smoothing lenghts should be decreased, for the inversion to continue. Through this multi-scale approach more and more fine-scale heteregeneous structures should appear.
The usage of recovery proxies and tests is brought forward to analyse the validity of the final model, with the added disclaimer that its still a model. A model is a realisation of available data and mathematics, while preserving physical laws. Whatever changes can cause some changes in interpretation. To track changes in misfit during the inversion a validation dataset is used to validate and analyse the ongoing waveform fits. Validation events are events that cover the study area in its fundamental directions and are used to track the misfit development of the active fitted data set. Instead of being fitted, they are used in comparison to the current data in an iteration. Confidence can be given, to when the validation events are decreasing in misfit. This does indicate that the inversion is improving on the reference model. There are multiple reasons for stagnating or even increasing validation event misfits. Validation event misfits that increase again may be an indication that the waveforms are being overfitted and noise is being fitted. Here it is important to try to either decrease smoothing lengths and observe if the noise is not being fitted anymore. In the case where the noise is still increasing, then it might indicate the end of a period band. Meaningful measurements of the phase misfit require waveform similarity in the absence of cycle-skipping. The measurement of the phase will be positively biased and large waveform differences will not be included in any time window.
There are also multiple uncertainty tests to analyse the model after the model has been finished and fully inverted, which come with certain benefits as well as certain trade-offs. In the area of recovery tests, approaches that deliberately introduces errors are used to try to recover the resolution of the end model, again through inversion. These include checkerboard tests (Lévěque et al. (1993)) or restitution tests (van Herwaarden et al. (2023)). A checkerboard refers to a model of the same size, which produces synthetic data in a checkerboard pattern. The synthetic data is then used as the observed data and inverted for. Checkerboards are fairly simple to implement and give a good estimate of the possible achievable resolution. Their downside is that the resulting model itself is not tested and disregard noise and insight into the non-uniqueness of the model. In a restitution test, misfit perturbation in form of gaussians are deliberately added to the end model. The perturbed model is again inverted , performing iterations until the misfits are corrected again. The advantage of this test, is that the end model is considered fully and the non-linear nature of FWI is naturally considered. In the area of Hessian-based strategies are point spread functions and L-BFGS approximation. Both of these strategies try to estimate the Hessian matrix, which can interpret which areas inhabit perturbations.

2-7 FWI recipe in a nutshell

Now with all the ingredients given, a recap of Full-waveform inversion helps for modulation and having an overview. The goal is to deliver one Full-waveform inversion iteration with target parameters, including the velocity parameters for horizontal and vertical P- (VPH,

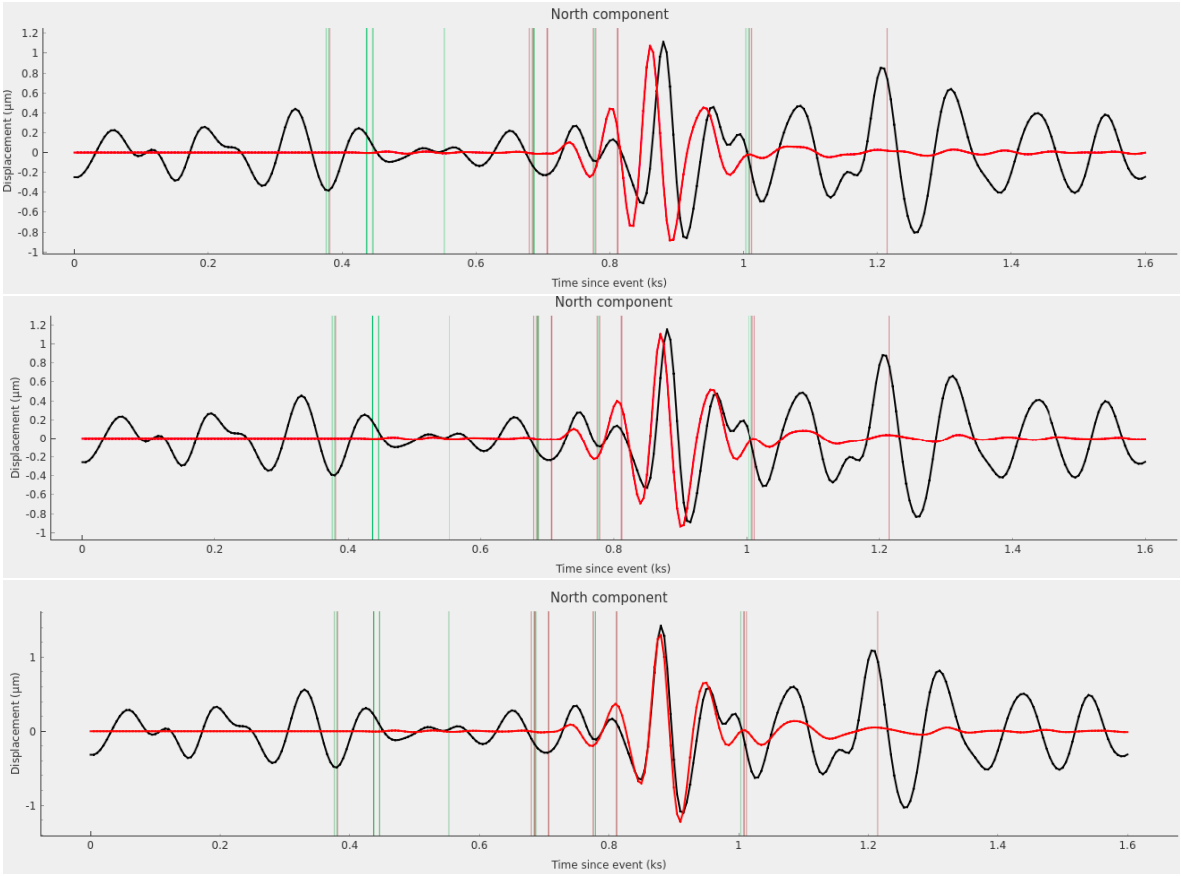

Figure 2-2: Misfit fitting through reduction of the phase difference between observed (in black) and predicted (in red) waveform at a 50s-90s period band of the receiver UP.LUNU.

VPV) and S-waves (VSH, VSV) velocities respectively as well as Rayleigh and Love wave phase velocities, density ρ and η which is a dimensionless parameter. For the inversion, only isotropic P-wave velocity, VP, and transverse isotropic S-wave velocities, VSH and VSV, are inverted for. Due to parameter space, $\eta = 1$ and VP = VPH = VPV. The resulting model is a transverse isotropic velocity model. For this the following workflow has to be followed:

1. Preprocess the observed data by applying a bandpass in the period band of interest e.g. 30s - 120s for broadband data.
2. Forward model the partial differential equation with a suitable numerical method according to a reference model with N-dimensional model space and target parameters.
3. Select window around a proper arrival in the observed data. Over the Inversion process, the phases between the synthetic and observed data are matched and can only be half a period out of sync or else cycle-skipping occurs. The misfit functional is also important in this step as it controls how the observed and synthetic data will be matched.
4. Start regularisation with station weighting (Ruan et al. (2019)). Data availability is not uniform and as such some regions will have a higher station density then other regions. Clustered stations can make the sensitivity of the misfit functional heterogeneous, which is not desirable. As such measure the proximity between stations and apply a normalized weighting coefficient to all stations, where clustered stations will be downweighted.
5. Compute adjoint sources according to the misfit functional and run adjoint simulations.
6. Sum the gradients for each event together in the model and perform a diffusion equation based smoothing over the gradient to regularise out edge values that occur due summing gradients.
7. Update model with non-linear gradient method. In this case L-BFGS will perform a Hessian kernel based trust-region gradient optimization by finding the global minimum. L-BFGS approximates the curvature of the misfit-functional (Wright (2006))
8. End of current iteration, prepare the next iteration and repeat all steps.

Chapter 3

Modelling Methodology

3-1 The Spectral-Element Method

Accurately modelling waves that propagate through a medium obtains information about that medium. The method of modelling has to support the expectations to have an appropriate level of accuracy in the resulting model. For the simulation of broadband period waveforms, the Spectral-Element Method (SEM) is used. Applications where the modelled domain has a complex geometry and boundaries, the SEM is the most effective method for simulating a wavefield. This is due to a specific set of interpolation rules resulting in a globaly diagonal mass matrix, that can be efficiently solved for each time step in the simulation (Igel (2017)). The simulations are handled by the software Salvus (Afanasiev et al. (2019)). Salvus has been conceived as a software that can practically handle the simulation of wavefields with the SEM through unstructured rectangular-, hexahedral- or tetrahedal Meshes. The user should have the flexibility to implement different Mesh forms and to start a simulation on that mesh in two- and three dimensions. It is able to accurately simulate the wavefields used to invert data for fully three-dimensional volumes. The following sections will discuss the mathematical background and formulation of the method, why the formulation does well with simulating wave propagation and the advantages for seismic tomography studies.

3-1-1 Background

Three other frequently used numerical methods in seismic applications are the finite-difference method (FD) (Moczo et al. (2014)), the pseudospectral method (PS), the finite-volume- (FVM) the finite-element method (FEM). FD is a cheap method since it can simulate with a low-order approximation of the partial differential equation in question e.g. the elastic wave equation and can be easily implemented for the one- and two-dimensional case. It is the method of choice for quick implementation, where the domain is relatively simple e.g. equidistant cubes in a rectangular domain. It starts to suffer when it comes to accurately implementing free-surface boundary conditions when met with realistic topography as well as errors slowly accruing at long spatial distances.

PS followed FD with the property of exactly calculating its space derivative, and mitigating the problems of long wave propagation (Komatitsch et al. (1996)). The method can react to discontinuous changes, which made it attractive for anisotropic studies (Igel (2017)), but suffers of poor scaling in parallelisation of codes, and making it difficult to adapt for complex geometries towards three-dimensional problems.
FVM and FEM have the advantage that they can be used also in higher orders, boundary conditions can be implicitly solved in the equations, and they can be implemented for complex geometries with tetrahedal or hexahedral elements. FVM have high computational cost, which makes the FEM more attractive as it is similarly accurate while cheaper. Still, finite elements also suffer since it is a large linear system of equations. The mass matrix, the matrix that has to be inverted for the discrete solution, is also a sparse matrix that is difficult to solve for large systems. Flux terms also have to be computed making it also computationally expensive. SEM can circumvent some of the aforementioned issues.
SEM is FEM with specific choices of basis functions and collocation points. It combines higher-order interpolating polynomials on each element with Gauss-Lobatto-Legendre (GLL) collocation points and quadrature. This results in a globally diagonal mass matrix (Afanasiev et al. (2019)). Elementwise operations can be formulated as matrix-matrix products, which is parallisable. In comparison with FEM, the diagonal matrix of SEM is easier to solve for time evolution. Numerous developments enabled the application of the spectral-element method for stable seismic simulations on heterogeneous meshes. One breaktrough was the successful implementation of seismic wave propagation on a three-dimensional heterogeneous spherical mesh of the Earth (Komatitsch and Tromp (1999)). Since then the spectral-element method is used for seismological studies on two- to three-dimensional meshes with free surface boundary conditions. Making modelling detailed wave propagation models with high simulation times possible.

3-1-2 Formulation

For the tomography, formulate the spectral-element method for the three-dimensional elastic wave equation. For brevity it is easier showing the principles of the formulation with a one-dimensional example of the elastic wave equation given in equation 2-34. This is the strong form of the wave equation. To numerically solve the Partial Differential Equation using SEM, the weak- or also variational form of the equation has to be formulated. This is done by multiplying the equation with a test function $\varphi(x)$ and then integrating in a space Ω:

$$\int_\Omega \rho(x)\frac{\partial^2 u(x,t)}{\partial t^2}\varphi(x)dx - \int_\Omega \nabla\cdot(C(x):\epsilon(u)(x,t))\varphi(x)dx = \int_\Omega f(x,t)\varphi(x)dx \tag{3-1}$$

Recall Gauss-Green theorem, where $\iint_\Omega(\nabla\cdot F)dA = \oint_D F\cdot\hat{n}ds$ and D a curved domain. Apply the Gauss-Green theorem to eliminate the second derivative in space:

$$\begin{aligned}&\int_\Omega \nabla\cdot(C(x):\epsilon(u)(x,t))\varphi(x)dx\\ =&-\int_\Omega \nabla(C(x):\epsilon(u)(x,t))\cdot\nabla\varphi(x)dx + \oint_D(\nabla(C(x):\epsilon(u)(x,t))\cdot\vec{n}(x))\varphi(x)dx\end{aligned} \tag{3-2}$$

Inserting equation 3-2 into equation 3-1 gives the weak form of the one-dimensional elastic wave equation:

$$\begin{aligned}&\int_\Omega f(x,t)\varphi(x)dx - \int_\Omega \rho(x)\frac{\partial^2 u(x,t)}{\partial t^2}\varphi(x)dx \\ &= \int_\Omega \nabla(C(x):\epsilon(u)(x,t))\cdot\nabla\varphi(x)dx - \oint_D (\nabla(C(x):\epsilon(u)(x,t))\cdot\vec{n}(x))\varphi(x)dx\end{aligned} \tag{3-3}$$

A solution to the weak form is found if the choice of the test function $\varphi(x)$ can satisfy the wave equation. Following, for the simulation on a numerical grid, a discrete representation of the weak form is needed. A discrete representation can be found with the Galerkin method by approximating the solution for $u(x,t)$ by a superposition of n basis functions $B_i(x)$, where $i = 1, ..., N$ weighted by time-dependent coefficients $u_i(t)$:

$$u(x,t) \approx \bar{u}(x,t) = \sum_{i=1}^{N} u_i(t) * B_i(x) \tag{3-4}$$

where $\bar{u}(x,t)$ is the approximate displacement field. The accuracy of the approximation is dependent on the specific choice of basis functions and the number of functions N. Another important step is the usage of the approximate displacement and the basis functions as the test function. Using equation 3-3, this obtains:

$$\begin{aligned}&\int_\Omega f(x,t)B_i(x)dx - \int_\Omega \rho(x)\frac{\partial^2 \bar{u}(x,t)}{\partial t^2}B_i(x)dx \\ &= \int_\Omega \nabla(C(x):\epsilon(\bar{u})(x,t))\cdot\nabla B_i(x)dx - \oint_D (\nabla(C(x):\epsilon(\bar{u})(x,t))\cdot\vec{n}(x))B_i(x)dx\end{aligned} \tag{3-5}$$

Combining equation 3-4 and equation 3-5 leads to an equation for the time-dependent coefficients:

$$\begin{aligned}&\sum_{i=1}^{N}[\frac{\partial^2 u_i(t)}{\partial t^2}\int_\Omega \rho(x)B_j(x)B_i(x)dx] \\ &+\sum_{i=1}^{N}[u_i(t)\int_\Omega C(x)\nabla B_j(x)\ \nabla B_i(x)dx - \oint_D (C(x)\cdot\vec{n}(x))\nabla B_j(x)\ \nabla B_i(x)dx] \\ &= \int_\Omega B_i f(x,t)dx\end{aligned} \tag{3-6}$$

for all basis functions B_j with $j = 1, ..., n$. This is the Galerkin form for finite-element problems, where the constitutive equations can be rewritten in matrix notation:

$$M\frac{\partial^2 u(t)}{\partial t^2} + Ku(t) = f(t) \tag{3-7}$$

where M_{ij} and K_{ij} are the mass matrix and stiffness matrix respectively. The mass matrix is diagonal given specific basis functions and collocation points. The inversion of the mass matrix is trivial over its determinant, while the stiffness matrix is banded with the bandwidth determined by the number of basis functions in each element. $f(t)$ is a vector containing the source forces propagating the system from a point of rest $t = 0$. This matrix equation has to be solved for the space-independent but time-dependent coefficients $u_i(t)$. This results in a vector of coefficients that will inhabit the displacement values at a global set of points imposed by the given basis functions.

3-1-3 Global assembly

One crucial advantage of the SEM is, that it can react towards and model discontinuities. This is an issue for tomographic models, as the nature of transition zone in the mantle can lead to stark discontinuities in material and therefore in the depth gradient of properties. Spectral-elements do not have this issue due to inherit discretisation of the domain into subdomains on the element level, where one subdomain is an element and has their own basis function. In equation 3-6 this means that every integral of each term is summed up over each element up to the number of subdomains n. The approximation in equation 3-4 its time-dependent coefficient and the basis function are denoted for their respective element with e as a subscript as well as the matrices and vectors in the matrix equation 3-7. This means the matrices become coefficients of the unknown displacement inside an element. The integrals also get a coordinate transform for every element from the global domain to the local domain of an element. This leads to a semi-discrete weak form of 3-6 for one element only, where space is discretized but not time. The next challenge is to fit all the elements into each other, such that it will stay truthful as a discretisation of the entire domain.
As a basis function for every element, a Lagrange polynomial is used. Given a fixed interval between $[-1, 1]$ as x, the polynomial order N determines the number of fixed points between each other as:

$$x_i^N(\xi_i) = \Pi_{j\neq i}^{N+1} \frac{\xi_i - \xi_j}{\xi_i - \xi_j} = 1 \tag{3-8}$$

where $i, j = 1, 2, \ldots, N+1$ and Lagrange polynomials are orthogonal to each other $x_i^N(\xi_j) = \delta_{ij}$ given δ_{ij} is Kroenecker's delta, which is 1 if $i = j$ and 0 otherwise. The aforementioned GLL points help to choose which Lagrange points should be used by trying to minimize the interpolation error in between collocation points through numerical error propagation. This means variable densification of points in the domain. Advantageous for densifying the function interpolation near boundaries to avoid an overshoot. This also can be extended to the modelled geophysical parameters in the weak form notation, which makes it an attractive smoothing feature for seismic wave propagation, but having to adapt and integrate the integrals numerically.
The analytical functions can be replaced through polynomial approximations, which can be analytically integrated. This can be done through a Gauss quadrature scheme. The Gauss quadrature is extended to the GLL quadrature due to the requirement including the boundary points into the polynomial for which the GLL points are used for integration. Here can be observed that the higher the polynomial order and if the integrand is smooth the more accurate the approximation of the numerical integral. What can also be observed is differences in error accumulation. In the spatial domain the only error accumulation is through the numerical integration with the GLL quadrature, while in the time domain its the numerical error propagated by the finite-difference approximation of the derivatives in time, which will be important later. The numerical solution to the spectral-element method for a single element is then:

$$\sum_{i=1}^{N+1} M_{ji}^e \frac{\partial^2 u_i^e(t)}{\partial t^2} + \sum_{i=1}^{N+1} K_{ji}^e u_i^e(t) = f_j^e(t), \; e = 1, \ldots, n_e \tag{3-9}$$

The mass-, stiffness matrix and source vector need to be computed before time extrapolation as a prepatory step for the wavefield. Missing now is a way to compute the derivatives of the Lagrange polynomials at the GLL points. This is done with a pendant called Legendre

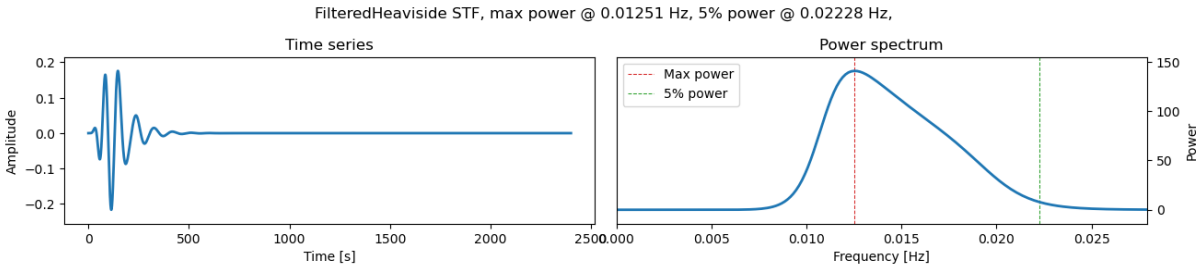

Figure 3-1: Example of the Filtered Heaviside source time function in the 50s to 120s period band. The power is at 5% at the cut-off point before the minimum period of 50s.

polynomials, by evaluating the Lagrange polynomials at the GLL points through them, given as:

$$L_N(\xi) = \frac{1}{2^N N!} \frac{d^N}{d\xi^N} (\xi^2 - 1)^N \tag{3-10}$$

where $L_0(\xi) = 1$, $L_1(\xi) = \xi$, $L_{N\geq 2}(\xi) = \frac{1}{N}[(2N-1)\xi L_{n-1}(\xi) - (n-1)Ln-2(\xi)]$
Now that all parts have been numerically integrated for every sub-element in the global domain, for the full numerical realisation of the wavefield everything has to be assembled into a global system. For this the sub-elements need to be added at the corresponding boundary collocation points. Basically the matrix value of each boundary element are added up together as the boundary collocation points overlap. This results in a global matrix-vector system for the overall wavefield with global mass matrix M_g, global stiffness matrix K_g and global source vector f_g. This leads to a system, where simply the global time-dependent coefficient u_g needs to extrapolated through centred finite-difference scheme:

$$u_g(t + dt) = dt^2[M_g^{-1}(f_g(t) - K_g u_g(t))] + 2u_g(t) - u_g(t - dt) \tag{3-11}$$

Technically feasible is the modification of the elements during time extrapolation, but for the problem in this study the model and mesh will remain constant for the wave propagation in the forward and adjoint problem.

3-1-4 SEM Modelling of events

With Salvus, using one event as a reference event, a forward simulation is started for the cartesian mesh and for the SmoothieSEM mesh with the same simulation settings. For the source injection, a source time function is needed that can imitate the source and sample the outgoing waves in the SEM well. The source time function is a Filtered Heaviside function, which starts sampling only after the resting time at $t = 0$ (Figure 3-1). It initialises energy in a prolonged oscillation with a sampling rate of 10Hz and then decays over time, where the power spectrum should reach 5% at a maximum cut-off frequency of the source. The goal is to have enough energy for sampling and the maximum power and cut-off are around 15 - 20% away from the maximum and minimum frequency respectively, which are done by setting high values for the highpass- and lowpass corners. This changes depending on the frequency range.

For event modelling, the property of SEM to simulate problems with uneven free surfaces helps to accurately model many different propagation problems resulting from moment tensors

e.g. surface waves. At source injection, SEM implements a moment tensor point source (Fichtner (2010)). This point source is given as:

$$f(x,t) = -\nabla \cdot [Z(t)\delta(x - x^s)] \tag{3-12}$$

which is located at the source time funtion injection. The symbols s and $Z(t)$ are the vectorial source time function and moment tensor, respectively. In the SEM, the moment tensor can be implemented either through exact integration or through a polynomial approximation of the delta function. While both are possible, more convenient is to add a polynomial approximation directly to the stress tensor, before computing the injection. This allows an accurate intermediate- to far field simulation of the seismic wavefield, body- and surface waves at moment tensor injection. An issue arises with the near field wavefield, which is inaccurate and needs to either be low-pass filtered or the area around the source in a Gradient set to 0. No errors are caused for the intermediate- and far-field wavefield (Nissen-Meyer et al. (2007)).

3-2 Wavefield adaption

To get tomographies to the target period, computational resources have to be available. The motivation for the developments in this chapter stem from a desire to efficiently save on the overhead of number of mesh elements in our simulation and to keep the computation time low. The required number of elements inside a cartesian mesh scales with the respective simulation frequency to the power of the respective mesh dimension N, such that f^N(Thrastarson et al. (2020)). At low periods i.e. high frequencies, 3-D meshes will therefore scale by a factor of 8 for a cartesian mesh at 10 seconds periods. This results approximately to 14 million elements for the study area. Anisotropic adaptive mesh refinements (aAMR) have been developed from the AxiSEM principle to exploit the approximate axial symmetry of global wavefields and to adapt a conventional cartesian mesh to make use of the shape of a wavefront (Nissen-Meyer et al. (2014a)). Needing less elements overall. For the narrow propagation close to the source to be discretised sufficiently, the source will still have a quasi-rectilinear mesh part around it. The technical aspects of the Spectral-Element-method mentioned before concerning the subdomains enabled the development of this mesh type (Figure 3-2). These wave-adapted meshes will be referred to their production name SmoothieSEM, which have a frequency scaling of dimension N-1. A significant reduction in frequency scaling means that at higher frequencies, the SmoothieSEM meshes can represent these wavefields with only a fraction of mesh elements compared to a standard cartesian mesh.

3-2-1 Mathematical background: Forward Problem

The basic principle builds on the fact, that the complexity of a wavefield propagating through a smooth Earth model is high in the vertical and radial directions, but not in the azimuthal directions from the wavefront (Figure 3-3). Another fact from the Spectral-Element method is that it does well with surface curvature on the elements. Knowing this, try to elongate the elements along the azimuthal curvature while, for the time marching scheme, the grid points per wavelength are conserved along the radial propagation direction. Effectively this turns a cartesian mesh into a irregular hexahedral mesh shaped like the wavefield, which drastically lowers the overall number of elements necessary for a simulation. As a comparison,

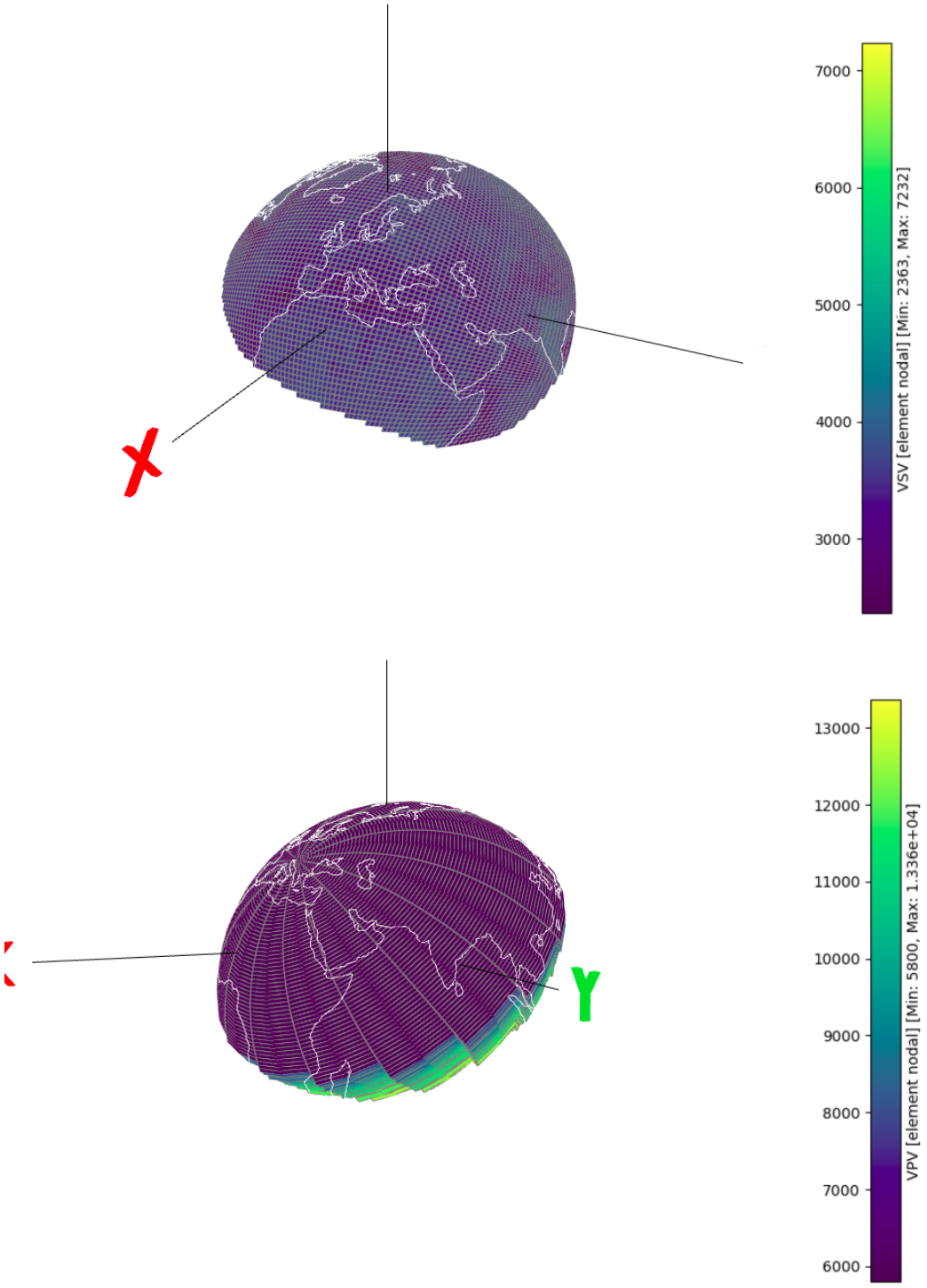

Figure 3-2: Upper figure shows a three-dimensional spherical chunk mesh, while the lower figure shows the equivalent SmoothieSEM mesh

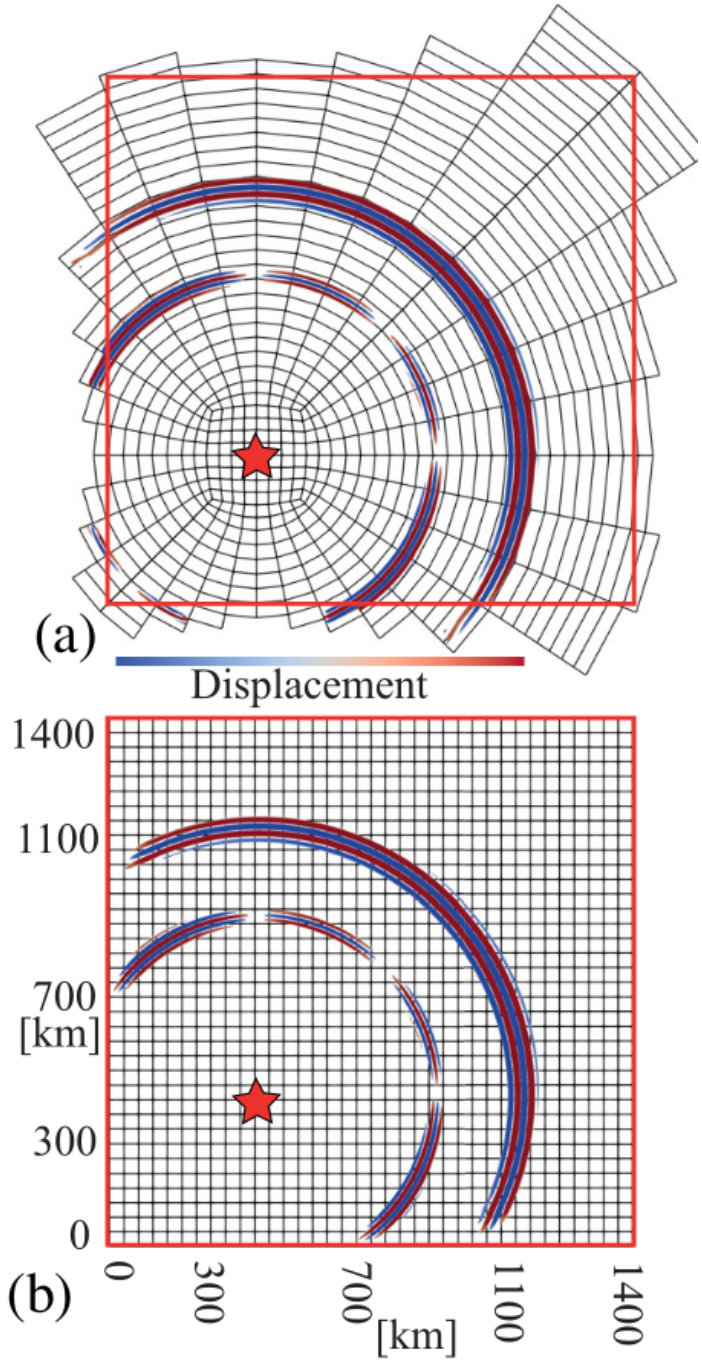

Figure 3-3: (Figure from Thrastarson et al. (2020)) Snapshot of a forward wavefield propagating from a source, denoted by the red star. a) Propagation on a 2-D SmoothieSEM mesh. b) Propagation on a 2-D rectilinear mesh

an equidistant cartesian mesh is distributed in a circular domain such that it can sample the modelled wavefield with 2 elements per minimum wavelength. A SmoothieSEM mesh is made with similar technical properties. What can be directly observed is the difference in number of elements. In an example from Afanasiev et al. (2019), the circular domain with 1.000 km diameter has 253.700 elements in a cartesian mesh while the SmoothieSEM only has 18.500 elements. Immediate comparison of the synthetic wavefield simulated on both meshes show some differences in the phase arrival amplitude. This difference however should be smaller then the differences between observed- and synthetic wavefield. All in all, modelling the wavefield on a SmoothieSEM leads to a reduction in compute cost by a factor depending on the modelled frequency range due to the difference in scaling dimension and the complexity of the modelled structures. At higher frequencies this yields a substantial reduction that enables efficient and cost-effective usage of FWI at higher frequency ranges for a variety of meshing problems e.g. global- and regional-scale FWI In the case of tomographic problems.

3-2-2 Mathematical background: Adjoint Problem

Conceptually when a wavefield is backpropagated from its receivers then the wavefield will originate from the receiver as a point source. A SmoothieSEM is adapted to the forward source, which means that the backpropagated adjoint wavefield wouldn't technically have the wavefied adaption of the mesh. Therefore meaningful capturing of the Fréchet derivative would fall through correct? It turns out the resulting sensitivity kernels of the parameters, compared between cartesian and SmoothieSEM mesh, are nearly identical. In a 2-D study, the resulting gradients and their sums are almost the same (Thrastarson et al. (2020)). The adjoint can be solved, such that it can be simplified and only needs the linear operator L from the forward problem. Meaning that the mesh has negligible influence on the determination of the adjoint parameters. This also shows that the adjoint is not physical and purely a mathematical representation of a physical forward problem determined through physical intuition.

3-2-3 Application in Waveform inversion

The mesh generation of a SmoothieSEM involves as a parameter how many lateral azimuthal elements are used for every quarter of a 360° circle. The choice decides whether the SmoothieSEM will have less or more elements depending on how big the area of propagation is chosen, as well as how much the source will be discretised. For example, if the amount of azimuthal elements is four, one quadrant will have 45° divided by four element rays propagating outwards. Eight elements will have 45° divided by 8 and so on. Depending on that choice, there will be an expected difference between the synthetic wavefield data of a cartesian mesh and a SmoothieSEM mesh depending on the "badness" of the SmoothieSEM mesh. The less azimuthal elements, the more the elements will be stretched out with increasing epicentral distance from the event source. One question in this study is if this ratio of badness also translates to a three-dimensional continental-scale tomographic model and if the error is propagated linearly. In the 2D-tests this has been shown that there is an expected error in the resulting gradient. This error however only deviates under one percent from a test target model reconstruction with 6 azimuthal elements or more (Thrastarson et al. (2020)). Under 6 azimuthal elements there may be a danger that the complexity of the wavefield will

suffer, which is why its advised to not be too excessive with reducing the amount of azimuthal elements. Keeping the amount of elements in SmoothieSEM low while keeping high accuracy, lateral refinements are considered. Given a certain distance from the source, add more azimuthal elements. The source is still sufficiently discretised with a lower amount of azimuthal elements, and in the grander scheme of FWI, the complexity of the wavefield near the source is of less important for large-scale tomographic studies as there is sufficient waveform coverage. Real data applications have been done on a global-scale by Thrastarson et al. (2022). Here a three-dimensional global-scale model was created through a event-Specific meshing approach, where for each event in the model a corresponding SmoothieSEM mesh was created. The parameters from a model update from the spherical global mesh are then interpolated onto the GLL points of the SmoothieSEM meshes and each event has their own Forward simulation. This drastically reduced the costs in the overall FWI problem from the most expensive part, the SEM simulation. This makes the FWI an I/O problem for huge meshes in higher frequency ranges.

3-3 Dynamic mini-batch approach

Batch methods are a type of gradient descent method developed to exploit data redundancy. For waveform inversion having as many high quality events as possible only adds to the potential information that will be included in a tomographic model. Imagine a dataset with over 150 events, and every event has around 1000 waveforms, around 150000 unique ray paths would need to be included in one iteration so the inversion problem is preserved. One realises that the global matrices of this problem would be far too large for even supercomputers. Normally, every event would require painstaking work to process and inspect so that only the best waveforms with clear arrivals are used. The requirements would also change depending on the strength of the event and the available epicentral distance of a waveforms ray path. 150 events would still require a high amount of resources. As a result, developments in the direction of efficient iterative misfit minimisation have been made, that can breakdown the waveform inversion into more manageable problem by dividing the overall inversion problem into a series of smaller more manageable problems. Building on theories in stochastic gradient descent methods, only the misfit of an event batch is considered.

3-3-1 Basic principle

Determine a set amount of events in a mini-batch L. The corresponding misfit is approximated through a normalisation i.e. the sum of misfits of every datum in the batch:

$$\chi \approx \chi(L) = \frac{1}{|L|} \sum_i \chi_i \ |L| < N_d \tag{3-13}$$

This leads to an approximation of the gradient of the dataset for an update of the model. Following, the next iteration picks out the next batch of events. Which events are used is quasi-random with a condition limiting the reselection of the same events in consecutive iterations, which in turn helps the gradient descent to not get locked in a local minima, as the explorable gradient-descent surface changes with each mini-batch iteration and the global minima is easier to distinguish.

3-3-2 Adaptive Mini-Batch Optimization with L-BFGS

? developed an algorithm with adaptive mini-batch optimisation. For waveform inversions of event data, the approximated misfit is the sum of the singular misfits of each Earthquake event. Each iteration chooses a different batch of events and their corresponding misfits. Given a batch L with iteration number subscript k, the first iteration batch, denoted L_1, will select first a random choice of events, where the second event is furthest from the first, and following earthquakes have the maximum distance from all the previous selected earthquakes. This selection algorithm is called Mitchell's best-candidate. This ensures that a batch iteration has good coverage over the study area. To further restrict the selection of the same earthquakes in consecutive iterations, a selection probability is also included, that inversely lowers the chance of reselection by amount of times an event has been chosen.
For the convergence in the gradient method, a center problem is the relationship between batches and the overall convergence of the entire dataset, since in one iteration are different sources then in the previous iterations. For this issue, validation events are chosen that make up a control group. The control group CG_k itself is a subset of the mini-batch χ_k that is also used in the next iteration. The misfit $\chi(CG_k)$ approximates the total misfit E, and a model update will only go through if the model update x_{k+1} of x_k has a smaller misfit $\chi(CG_{k+1}) < \chi(CG_k)$. The selection of events for the control group is done through elimination. Every iteration, the earthquake with least contribution to the gradient approximation is removed until the control group stabilises to a predefined threshold of events. During the selection process, rough gradients are done. After the selection process high accuracy is needed to effectively explore the gradient surface of the models and ensure misfit reduction. This is important for the L-BFGS method, a type of gradient descent method that uses the mini-batch approach to determine a trust value where exactly the control group condition is calculated before moving on in iterations. The control group steers the size of the mini-batch and also the step length in gradient descent adaptively according to the current iteration conditions with the reference model. In conjunction with the L-BFGS method, the method approximates with each iteration also the inverse Hessian matrix as the misfit function changes with each iteration since it also means the analysis of different misfit functions.
L-BFGS is a strong gradient-descent method as it has many checks, such that the inverse Hessian is approximated to a convergence with high certainty. Because of this property it is relatively expensive in computational studies. For waveform inversion this can mean that the convergence takes more iterations then it would be beneficial costwise. As mentioned before, the standard L-BFGS method coupled with the mini-batches creates a strong, but also much cheaper optimisation method, where the expected convergence results are reached faster. This approach was compared with the standard approach to quantify the potential computational gains. What was shown, when using a fixed amount of iterations, that the adaptive mini-batch method converged faster then the standard L-BFGS method. In the end, providing another boost in computational efficiency in the overall FWI-workflow. In this study, the optson software (van Herwaarden and Hapla (2023)) is used. Optson has been developed as a lightweight optimisation tool that can be used with any Inverse problems that use L-BFGS methods. Here, the mini-batch L-BFGS method automatically determines the step length between every iteration and also rejects an iteration with a smaller step length if the trust condition of misfit reduction was not met.

Chapter 4

Experiments and Implementation into FWI workflow

For one iteration in FWI, there are certain steps that have to be taken for convergence. The workfow consists of simple steps: Pre-conditioning $->$ Forward Modelling $->$ Treatment of Forward Modelled data $->$ Calculate Misfit functional $->$ Adjoint Modelling $->$ Misfit optimisation. An extensive overview is given in 4-1. For the computation itself, this study uses the workstation "Dream" of the Seismology and Wave Physics Group as a local host, where the software is started from (Table 4-1). For the remote where the simulations and adjoint processing is done, the Swiss Supercomputer Piz Daint is used on the production budget of the Group, where the GPU can be effectively used for parallelisation (Table 4-2). The production software that handles the FWI of this study is a mix between packages Inversionson (Thrastarson et al. (2021a)), LASIF: the Large-Scale Seismic Inversion Framework (Thrastarson et al. (2021b); Krischer et al. (2015a)), optson (van Herwaarden and Hapla (2023)) and Salvus (Afanasiev et al. (2019)). Every software has a different purpose in the overall workflow and have been adapted for efficient communication between the packages. We also use event-specific SmoothieSEM meshes to lower computational costs of the FWI, as Cartesian meshes become simulation-wise too expensive after a certain period.

System	Dream workstation
OS	Ubuntu 20.04 LTS
CPU	AMD Threadripper PRO 3975WX 2.2GHz à 32 Cores
RAM	258 GB System Memory
Storage	10TB

Table 4-1: Specifications for the Group workstation "Dream" PC

System	Piz Daint CSCS Supercomputer
Project	s1168 allowance of 700.000 node hours
OS	Cray Aries, Dragonfly network topology
CPU	Intel Xeon E5-2690 v3, 2.6 GHz à 12 Cores
GPU	Nvidia Tesla P100 16GB à 600 available Nodes
RAM	256 GB System Memory per Node
Storage Capacity	20TB

Table 4-2: Specifications for the CSCS Supercomputer Cluster on Piz Daint

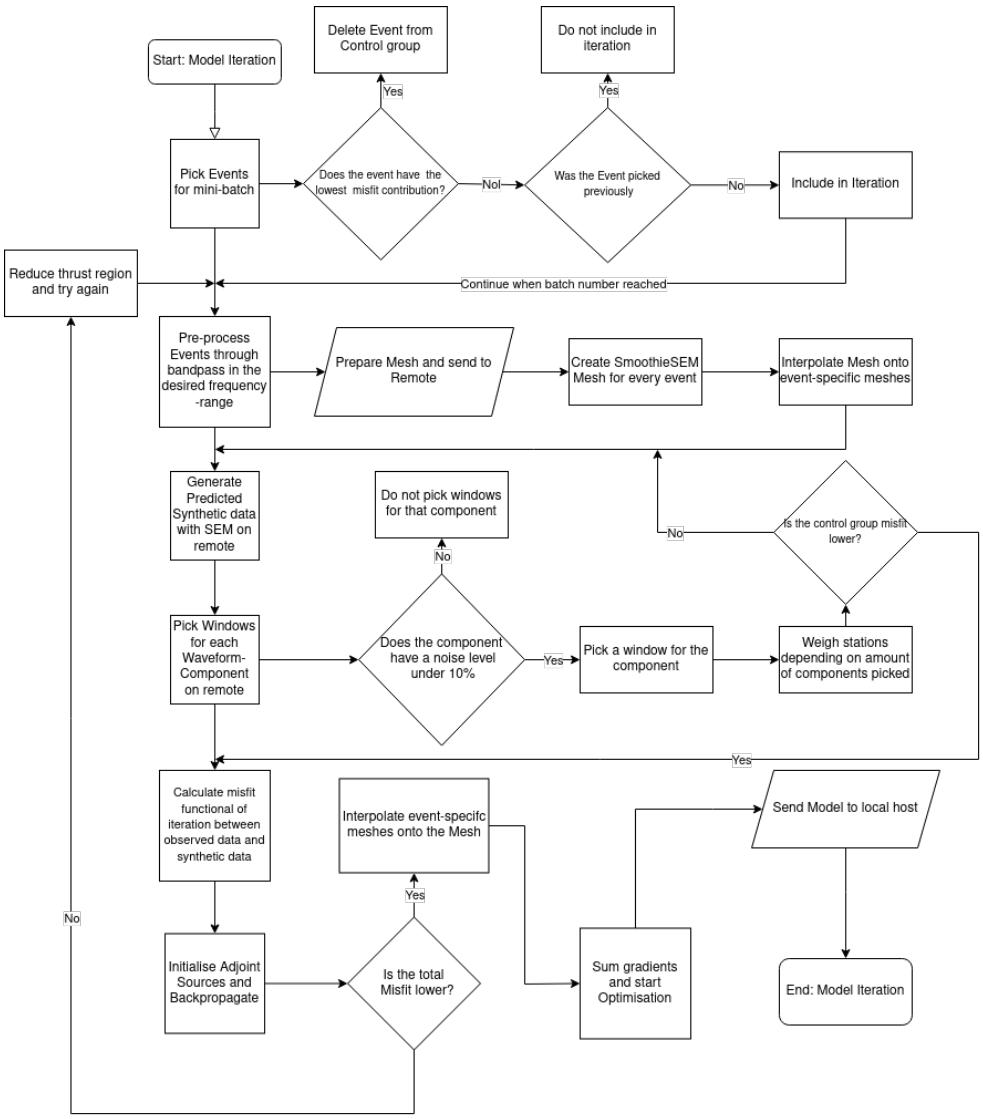

Figure 4-1: Process flowchart of the FWI with Inversionson, MultiMesh and adaptive mini-batches.

4-1 MultiMesh: Using Wavefield adaption

In **?**, the FWI-workflow was extended with the software package MultiMesh. MultiMesh is a seperate environment that handles any sort of meshing and interpolation routines. The goal with the package is to create for every event a specific SmoothieSEM mesh and to use those for the forward- and adjoint simulation to save on SEM costs during the workflow. What has to be specified here is that a Master mesh comprising the domain of the study area will be interpolated from and to. Before the forward simulations are started during an iteration, the SmoothieSEM mesh for an event is created through a general "create mesh" function. Through experiments, that will be mentioned in the next section, we developed a general SmoothieSEM template that ideally can be used for any event in any study area on a continental-scale. After the creation of a SmoothieSEM mesh, the mesh is stored for use on the remote in case the corresponding event is used in another model iteration. The parameters relating to depth from the master mesh are interpolated onto the event-specific meshes. Then the workflow from Forward- to Adjoint Simulation goes on as normal. When the simulations are done, the resulting gradients are interpolated back onto the Master mesh and summed onto it. Following the summation is the L-BFGS optimisation. Exploiting the cost-saving properties of SmoothieSEM, continental-scale studies can be brought to a higher frequency range then it was previously possible, which can possibly reveal new structures that couldn't be constrained before in the upper mantle.

4-2 Experiments and Implementation for Continental-scale FWI

4-2-1 Mesh creation settings

The software Salvus allows the creation of flexible hexahedral meshes with different subgroups of setting paramters (Afanasiev et al. (2019)). These are (1) the basic group, where the base modelling specific parameters are chosen, (2) the source group, where the centre point of the created mesh is located, the advanced group, where detailed modelling parameters like the CFL-number, model parameters or tensor-order can be set, (3) the seismic attenuation group (Nissen-Meyer et al. (2014b)), (4) the spherical group, where the minimum radius and ellipticitity can be given, (5) the refinements group, where the surface refinements of azimuthal elements can be set, and (6) the ocean and topography group, where accurate ocean loading and topography models can be included to the surface datum of the mesh. Importance and scrutiny are advised for all the parameters as the small changes added with each group add up and will affect the model irreperiably. For the event-specific meshes, some parameters will stay the same between frequency-bands and others will change due to the increase in frequency.
In the basic group, the background model will not change and the elements per wavelength and number of lateral elements will be set according to experiments. In the experiments with different SmoothieSEM of varying model parameters, the observations set by Thrastarson et al. (2020) were verified. The minimum required number of lateral elements was set to four, while the elements per wavelength had a minimum of 2 elements according to when the mesh elements have strong effects on the synthetic waveforms. At lower periods, where the effects from scattering at longer wave propagation distances are stronger, mesh refinements

Figure 4-2: Example SmoothieSEM mesh with surface refinements at longer distances. Each refinement doubles the azimuthal elements per quarter. The mesh is more refined at longer propagation distances.

are utilised to discretise the mesh for these effects (Figure 4-2).The background model is the CSEM 2.0 (Noe et al. (2023)), which is tuned for frequencies at 0.02Hz. Starting from that maximum frequency, the frequency band is increased step by step to slowly include more information into the model.
In the source group, the source latitude and longitude have to correspond with the latitude and longitude information of the event the mesh is trying to model. The advanced group will stay untouched as the default parameters are set for isotropic transverse tomographies in mind.

In the attenuation group, the default values are set to follow a one-dimensional memory variable approximation of the attenuation. In the spherical group the minimum radius is set to be sufficiently large, such that the mesh can be cut to the domain later, and the ellipticity set to the standard WGS84 value. The ocean and topography models have to be on the model at all times. The global topography dataset is a combination of the EGM2008 Geoid and surface data from Earth2014 in relation to the WGS84 ellipsoid standard. The ocean loading

global bathymetry dataset represents the depth of the oceans and is the difference between the entire Earth2014 data and the surface data.

4-2-2 Data adaptivitiy of SmoothieSEM

In comparison to a global domain, a continental-scale study will have clearly defined boundaries beyond just the spherical surface. There are some prerequisites that have to be met. First, does the event-specific mesh represent the event wavefield in the study area well. Secondly, what is the maximum epicentral distance of the entire study area and which refinements are necessary. Lastly, can the event-specific mesh be cut according to the study area. The latter implies that it is a prerequisite for the first point.
The first set of tests concerned the mesh cutting. Which parameters in the mesh creation where generally necessary to create a mesh that will always represent the event and receivers well? Given a clearly defined domain, mask out all the elements that are outside of this domain. The cutting procedure is as follows:

1. Roughly mask out the elements outside the Master mesh range
2. Adjust the depth between the Master mesh and the SmoothieSEM mesh
3. Surface Mask according to the locations of the event source and receivers
4. laterally refine the azimuthal elements to enable sufficient discretisation at large epicentral distances
5. Apply proper absorbing boundaries

The first two points will roughly cut the SmoothieSEM mesh to the Master mesh dimensions. Surface masking is described in van Herwaarden et al. (2020) and cuts out elements that are not located in a radius around the source and receivers on the mesh. A radius should be selected that allows for the placement of absorbing boundaries without sources and receivers being located inside of them. The data-adaptivity of surface asking can account for all source- and receiver points and their absorbing boundaries (Figure 4-3).

4-2-3 Absorbing Boundaries

Since the domain is limited, the simulation has to support the ongoing propagation of the wavefield in the specified simulation time. This would lead to the wavefield bouncing of the domain boundaries, which would artificially reflect waves back to the receivers and pollute the synthetic data with noise that can't be differentiated between reflections from heterogeneities in the domain as they have similar magnitudes. A solution is needed for waves to be absorbed at the domain boundaries. Absorbing boundary conditions honor the time of the propagation and therefore the physics. Mathematically, absorbing boundaries are some of the hardest problems with many ongoing research questions e.g. Perfectly Matched Layers for Galerkin spectral-elements. This is due to efficiency. Current absorbing boundary schemes have to be computed at the same time, which, depending on the boundary conditions, can also be costly. Normally for propagating waves in a cartesian mesh, sponge layers are used as they account

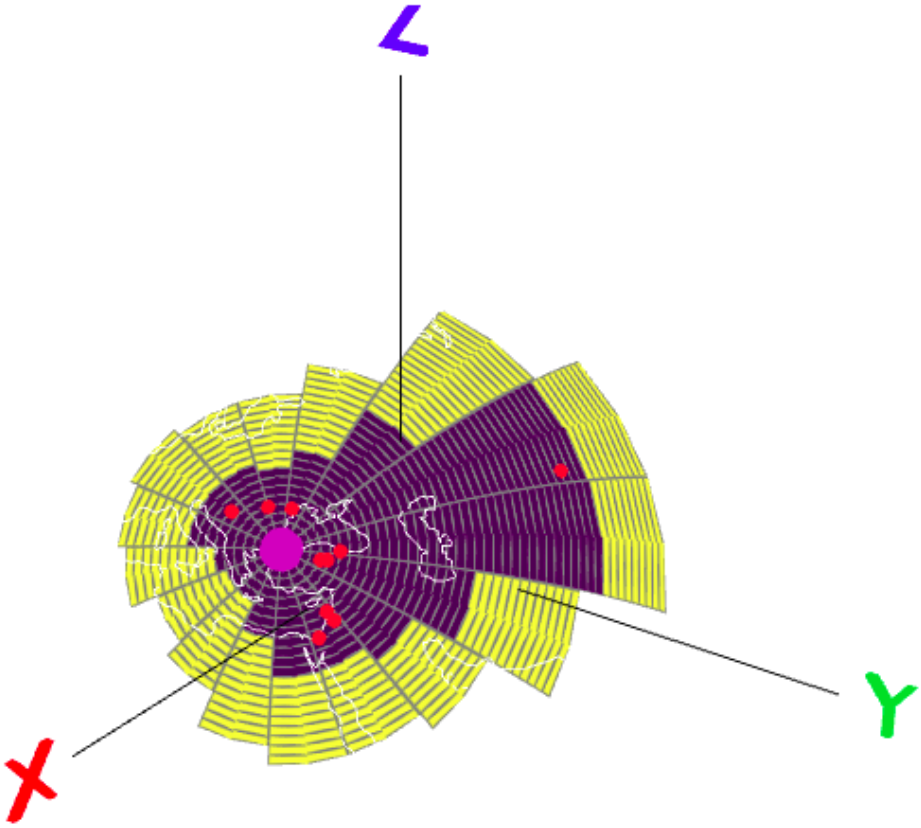

Figure 4-3: SmoothieSEM meshing test. Far away source or receiver points are considered well in the surface mask. The pink point denotes a source, the red points denote receivers. Yellow marked elements are absorbing boundary elements

for different angles of incidence and absorb energy coming into the boundary. Given a certain width d, the sponge layer can act as a taper to the incoming wavefield (Kosloff and Kosloff (1986)) with the equation:

$$2\gamma\partial_t u + \gamma^2 u \tag{4-1}$$

where γ is defined as function of width from the domain boundary:

$$\gamma(d) = \begin{cases} \alpha(1 - sin^2(\frac{\pi d}{2l}) & 0 \leq d \leq l, \\ 0 & \text{else.} \end{cases} \tag{4-2}$$

where α denotes the amplitude. These boundary conditions suffice for most wave propagation problems in bounded domains. In the SmoothieSEM the wavefield angle of incidence is the normal incident angle. For waves with normal incident angle a cheaper boundary condition for elastic wave propagation can be used called Clayton-Engquist condition (Clayton and Engquist (1977)).

$$C : \epsilon(u) \cdot \vec{n} = v_p \rho \partial_t u_{\vec{n}} + \sum_i v_s \rho \partial_t u_{\vec{r_i}} \tag{4-3}$$

where $\partial_t u_{\vec{n}}$ denotes the normal component of the velocity field and $\partial_t u_{\vec{r_i}}$ the tangential component and relates the normal traction on the boundary to the components. In the weak form representation of Equation 3-3 the left-hand side vanishes. This condition works well to absorb a majority of the incoming wavefield. But some reflections still come through, which, for the application in this study, still pollutes the data too much. Therefore, sponge layers are implemented for the simulation meshes.

4-2-4 Simulation differences between Cartesian- and SmoothieSEM mesh

For the simulation differences between both mesh types, a set of tests have been constructed to determine the accuracy of waveforms from a cartesian- and SmoothieSEM mesh by frequency range, differences between the observed and predicted data and relative misfit between synthetic data and observed data.

The synthetic data of the simulation tests has to qualitatively show two things. First, that the phases between the two meshes match. Secondly, that the absorbing boundaries work, indicated by a lack of wiggles after phase arrivals. Depending on the frequency this can change, as a taper has difficulties to absorb high frequencies. In Figure 4-4 is a common occurence of differences between a SmoothieSEM mesh and a cartesian mesh. The SmoothieSEM shows higher amplitude peaks on the phase arrivals than the cartesian mesh. This suggests that the SmoothieSEM elements indicate higher energies then the corresponding cartesian mesh. Intuitively, given the azimuthal stretching of the element meshes with higher epicentral distance, cartesian meshes at similar azimuthal distance weight the amplitude value a bit lower since the amplitude value is divided up by more elements than in the SmoothieSEM mesh. This would also suggest another question, whether in the SmoothieSEM meshes there is a linear increase in amplitudes relative to the amplitudes in the cartesian mesh and if this makes a difference in the overall FWI. Another property of SmoothieSEM was the preservation of grid points per wavelength in the radial direction, which is important for the phase arrivals of the predicted synthetic data. Both simulations match in phase. Looking back at the exploitation of heterogeneities in waveform data, the accurate prediction of phase arrivals is more sensitive to transverse heterogeneities then amplitude values. That the synthetic phases match,

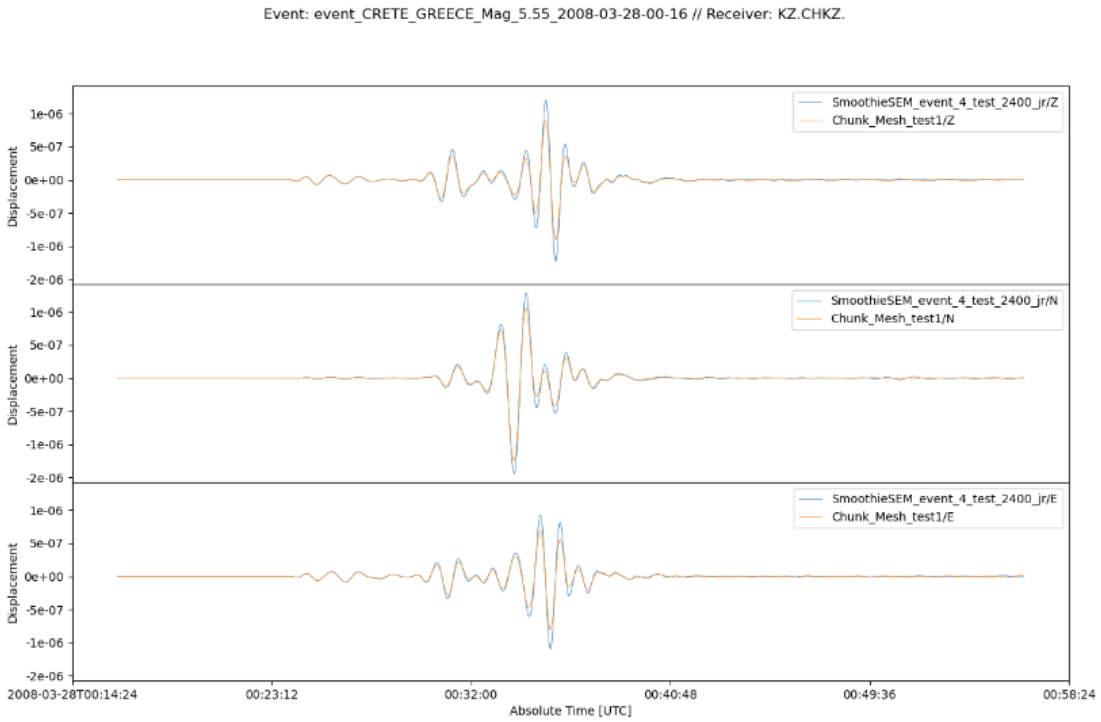

Figure 4-4: Example test from a Crete, Greece event between the cartesian and SmoothieSEM mesh at a period band from 50s to 120s. The example shows a frequently occuring pattern in most waveform tests between the cartesian and SmoothieSEM mesh, where the amplitude of synthetic waveforms is higher in SmoothieSEM meshes.

makes a one to one comparison of a model made with a cartesian mesh and a model made with SmoothieSEM meshes possible as well as the effective usage of SmoothieSEM meshes for seismic tomographic problems. Lastly, one can observe that after the phase arrivals there are minimal to no wiggles, showing that the absorbing boundaries work.

Given our initial starting model, we can compare initial synthetics to the observations. In Figure 4-4 the synthetic data follows the waveform of the observed data and amplitude wise they are similar. Phase wise there is a minimal phase shift between the observed and synthetic data. In the inversion this kind of setup is optimal for the misfit functional. Time-frequency phase misfit functionals will identify this phase shift and reduce the misfit between observed and synthetic predicted waveform after each model iteration, like in Figure 2-2. Ideally, in any frequency-band, this is the optimal setup, where the phase difference is not large enough to cause cycle-skipping issues.

The last question about the SmoothieSEM that needs to be answered is the differences between the observed and predicted data and relative misfit between synthetic data and observed data or rather will the differences in the synthetic data between the two mesh types be noticeable in the waveform inversion. Quantitatively this is hard to answer. Qualitatively, an approach is to compare with the observed data, which is the data the inversion is based on. The cartesian mesh is the reference waveform. If the difference between the observed data

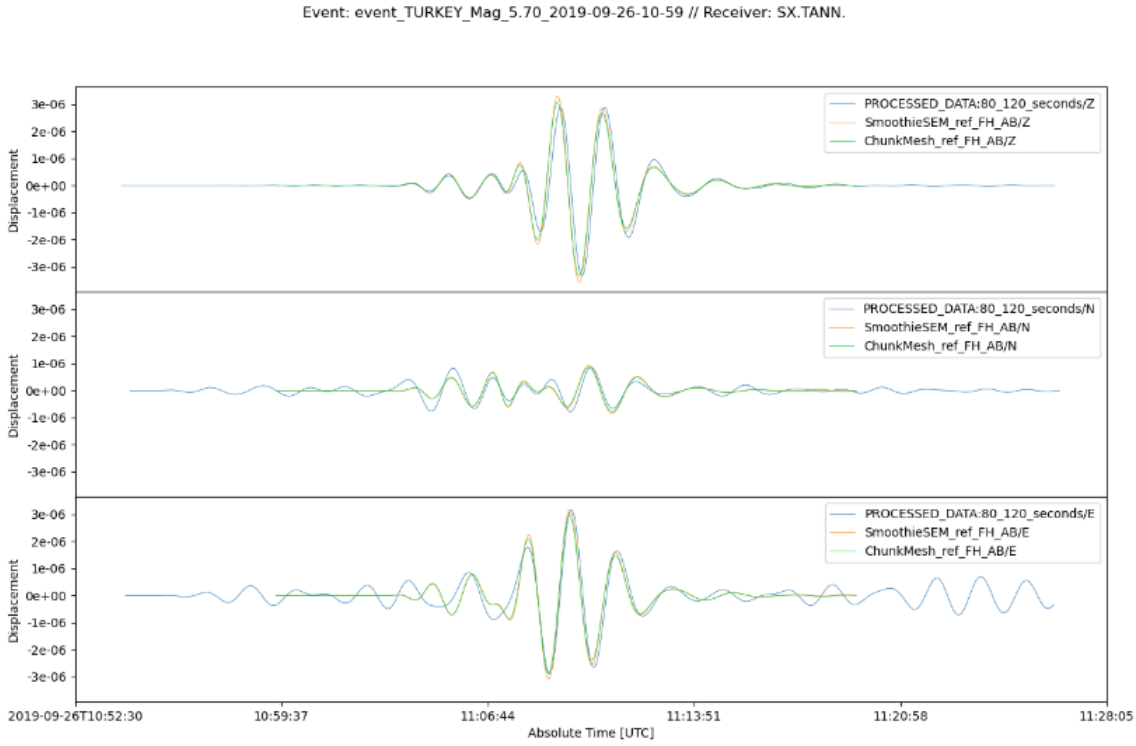

Figure 4-5: Comparison between the observed waveform of the event and the synthetic waveforms simulated by each mesh type. There is a slight phase delay between synthetic and observed waveform.

and the cartesian mesh is smaller then the difference between the cartesian mesh and the SmoothieSEM, then the difference between the two mesh types becomes relevant. Ideally, the opposite should be the case. In 4-6 it can be observed that the difference between observed and predicted synthetic data overlap the difference between the synthetic data of the two mesh types with an exception at epicentral distances higher than 50°. For the application in waveform inversion, the resulting event kernels will overlap in the relevant study area, such that these differences at higher epicentral distances will not matter too much in the optimisation step.

4-2-5 Gradients

When directly comparing the Fréchet Kernel Gradients from one event between the two mesh types, there are immediate observable differences (Figure 4-7). The upper figure is the event gradient on a cartesian mesh, while the lower figure is the event gradient on a SmoothieSEM mesh. The size difference between the two meshes is a result from the rough mesh cutting. To exaclty fit the SmoothieSEM inside the master domain, euclidean distances between the point clouds of the meshes have to be calculated. For large meshes, the calculation is prohibitive, such that GPU-boosted euclidean distance algorithms are considered. The most sensitivity to structures is located in the polarised coloured regions that alternate between the negative and positive vertical shear wave velocity. These regions are typically located on the perimeters of the Kernel as can be seen in the upper figure. In the lower figure however the SmoothieSEM shows a much different behaviour, where the most sensitive regions from the adjoint seem to be located behind the event Kernel. The direction and orientation of the conal structure coming off from the Kernel, the doughnut, seems to be the same. The conal structure is where the ray path of all source and receiver pairs of the event average to. This orientation corresponds to azimuthal orientation with the most windowed receivers during window selection. This apparent difference in the sensitivity kernel is not an issue for waveform inversion and sensitivities of an entire iteration. When all event gradients are summed up, then the actual area of depth sensitivity can be seen, which is then smoothed through a diffusion equation to smooth out edges. Comparing these directly shows that the main area of sensitivity is constrained by both mesh types. A small restitution test was done for the minimum period of 50s. This period corresponds to the tuned period of the CSEM base model, making convergence towards an acceptable misfit quick and a good base line for expected differences in shear wave velocity between both mesh types. The difference in vertical shear wave velocity, considering the specifications of the used SmoothieSEM meshes, should be the expected difference described in Thrastarson et al. (2020). Any further errors may come up at lower periods, where attenuation and heterogeneities appear stronger.

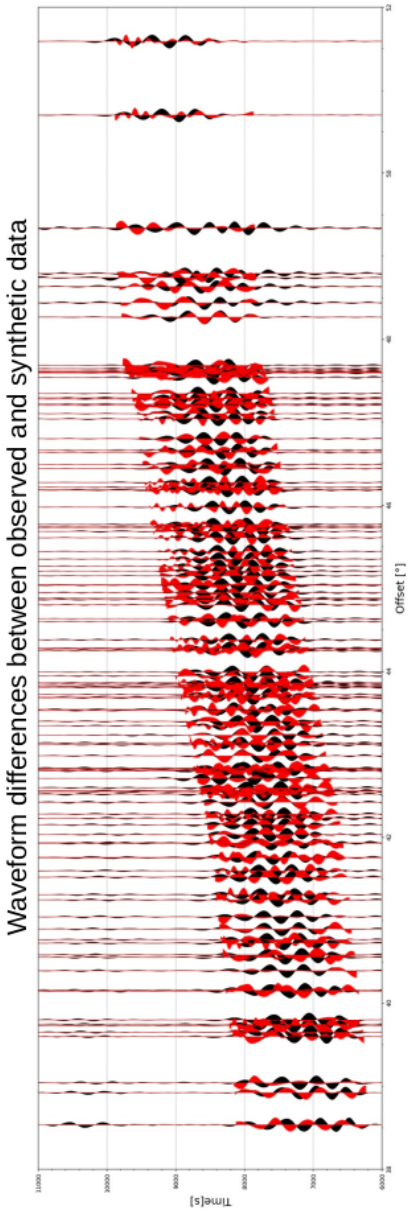

Figure 4-6: Red waveforms denote the waveform difference between the synthetic predicted waveforms of the cartesian and SmoothieSEM meshes. Black waveforms denote the waveform difference between the synthetic predicted waveform of the cartesian mesh and the observed waveform data.

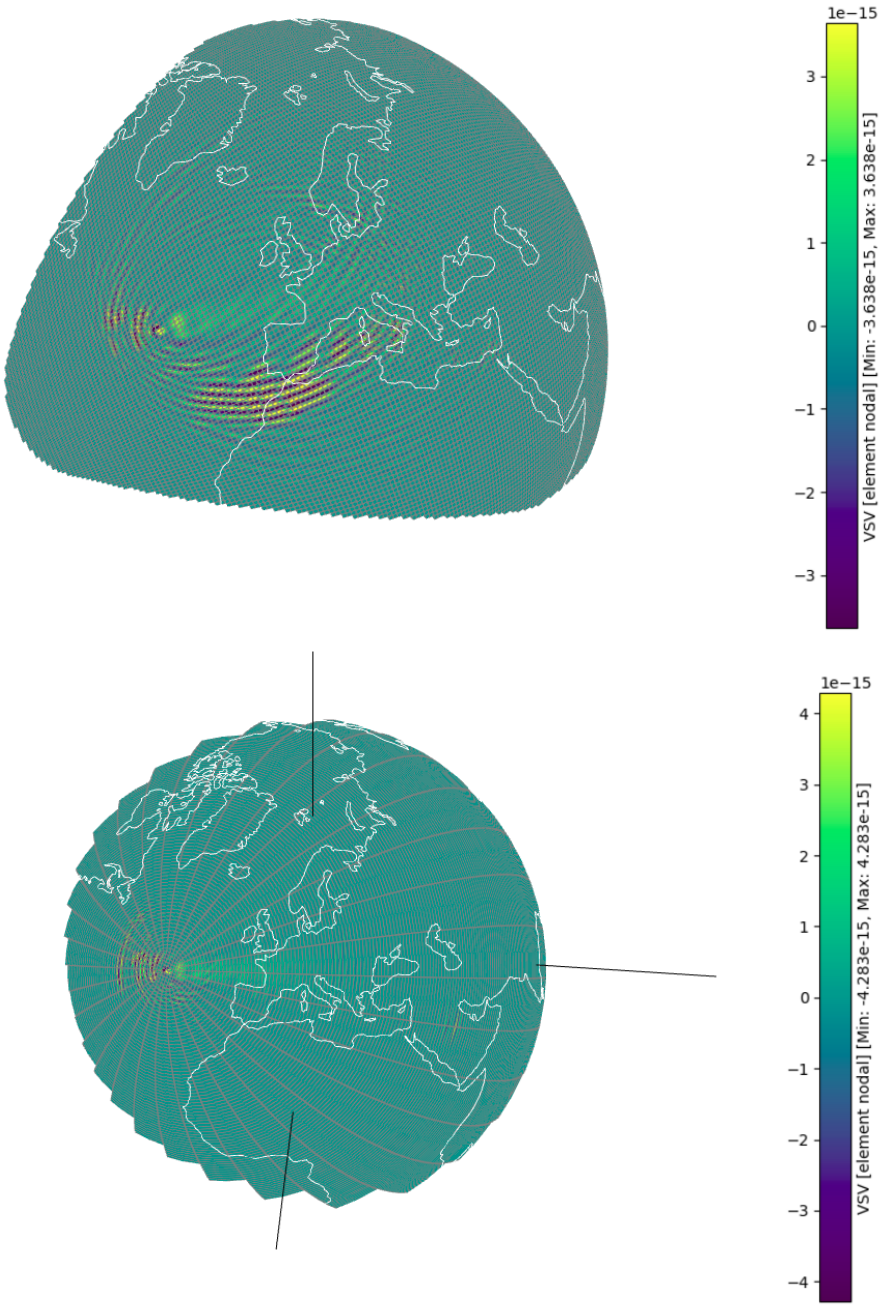

Figure 4-7: Differences in Gradients between the two mesh types. The upper figure shows the Gradient for a Cartesian type mesh, whilte the lower figure shows the Gradient for a SmoothieSEM type mesh. Strongest polar values show the areas where the Gradient is most sensitive to in the mesh.

Part II

Regional Full-waveform inversion Seismic Tomography of Europe and Western Asia

September 29, 2023

Chapter 5

Geological Background

A way to verify a model is through interpreting imaged anomalies with geological models. If an image correlates with known structural geology of today on the surface and from the past in hot mantle, then it only strengthens the confidence in the model parameters for further studies. The area covered spans from the Mid-Oceanic-Ridge (MOR) and Azores, Portugal to Sichuan Province, China. The geological information that can be covered could fill a book. For simplification, the study will focus on structures sensitive to velocity perturbations that can be confidently resolved in the upper mantle. This includes sutur zones, plate boundary processes, graben and fault structures, thermally hot and active areas, and cratons. The areas most sensitive to the FWI are encompassed by the Western half of the Eurasian Plate and includes all processes that are happening at the Eurasian-African-Arabian and Eurasian-Indian plate boundaries as well as processes along the MOR. Inner plate processes will also be covered. Going deeper in the models reveals processes and relics from past geological activity. Being able to link the past with the present is important to also base studies into the future on the model. To interpret what and why something is imaged, the nature of apparent heterogeneities in a model have to be explained. In a wavespeed model, differences between the horizontal- and vertical shear wavespeed need to be examined as one component is more sensitive to certain structures and processes then the other.

5-1 Europe

With "Europe" is meant the European peninsula until the Ural mountains, and the Near East (Figure 5-1). It is bordered by the MOR to the west and north-west till Greenland and Iceland, by the African plate to the south and by the Anatolian plate to the east. Inside Europe are multiple structures of orogenic nature, volcanically active regions, prominent sutur and graben and cratons. They all indicate large regions of currently active change. The peninsula has been extensively researched by geologists since the establishment of the scientific field. There are many fine-scale structures mentioned in literature or are being discussed in contemporary geological research.

Conceptually, the Eurasian and African plate started to collide and formed multiple mountain ranges in a E-W orientation along the mediterranean Sea (Dilek (2006)). Prominent orogenies are the Alps, the Dinaric Alps and Balkan Mountains, the Carpathians, the Apennines, the Pyrenees and the Ural Mountains. The Alps span from the southeast of France over Switzerland, the south of Germany, Austria, the north of Italy till Slovenia. The Alps are a strongly folded mountain range with a complex geological history of successive folding and thrust faults that uplifted marine sedimentary rocks like flysch (Schmid et al. (2004)). It consists of layers of rock from the European-, African plate as well as oceanic layers from a successive transgression and regression collisions involving the Tethys. A series of breaking microplates that formed seas and then slammed back into each other. This gives the alps a sedimentary record of metamorphic rocks of felsic or marine origin that follow each other up (Schmid et al. (2004)). Currently, the orogeny is active sitting on a stable crystalline rock basement (De Graciansky et al. (2011)). Starting from the southeast of the Alps are the Dinaric Alps, which span along the Balkan coast from Slovenia till Albania and North Macedonia. Part of the Alps orogeny, it is a still ongoing collision between the Adria microplate with the Serbo-Macedonian and Rhodope Massifs (Kilibarda and Schassburger (2018)). It is a Karst region mostly built out of Mesozoic limestone and dolomite. This combination results in canyons and gorges forming among the hard, stable dolomite rock through the erosion of limestone features (Schmid et al. (2008)). South of the Alps along the Italian peninsula lie the Appennines. This is a young mountain, with the orogeny starting in the early Neogene. It is distinct from the alpine orogeny through seperation through the Po Valley in Northern Italy, which is a filled trench that was a subduction zone between the Alps and the Appennines (Ollier and Pain (2004)). The Carpathians are a series of mountains that form a crescent starting from Bohemia, around Slovakia and Hungary and ending in Romania. The Carpathians formed during the alpine orogeny with a series of microplate collisions and subducting ocean crust, that introduced marine sediments that were then metamorphically altered into e.g. Flysch and back-arc volcanism (Schmid et al. (2008)). In front of the mountain a back-arc basin formed the Pannonian Basin. This basin underwent rare crustal thinning through extension. Lastly, the Ural Mountains in Russia will be treated as the geological boundary between Europe and Western Asia. It is one of the oldest extant mountain ranges and has an interesting orientation which is directly a N-S orientation (Brown et al. (2008)).
Inside the Eurasian plate are processes that aren't the result from plate boundaries, but from active tectonic or thermal activity. This includes, the Massif Central, Eifel mountains, which are large volcanic regions (Goes et al. (1999)). Graben and Valley structures experiencing rifting or extension like the Rhône Valley, the Upper Rhine Plain or Kannergat in the Oresund Strait (Dèzes et al. (2004)). Suture zones like the Teisseyre- and Sorgenfrei-Tornquist Zone are areas where two different tectonic cratons/units have joined together (Mazur et al. (2018)). Fault zones like the Northern- and Eastern Anatolian Fault (Şengör et al. (2005); Duman and Emre (2013)), which are prone to earthquakes. Lastly, there are certain parts of Europe that are relatively stable and the effects they are experiencing are due to buoyancy and isostasy, like the Fennoscandian Shield or the East European Craton (Artemieva (2003)) or belong to old orogenies like the Variscides in Central Europe or the Caledonides in Northern Europe and the British Isles (McCann (2008)).

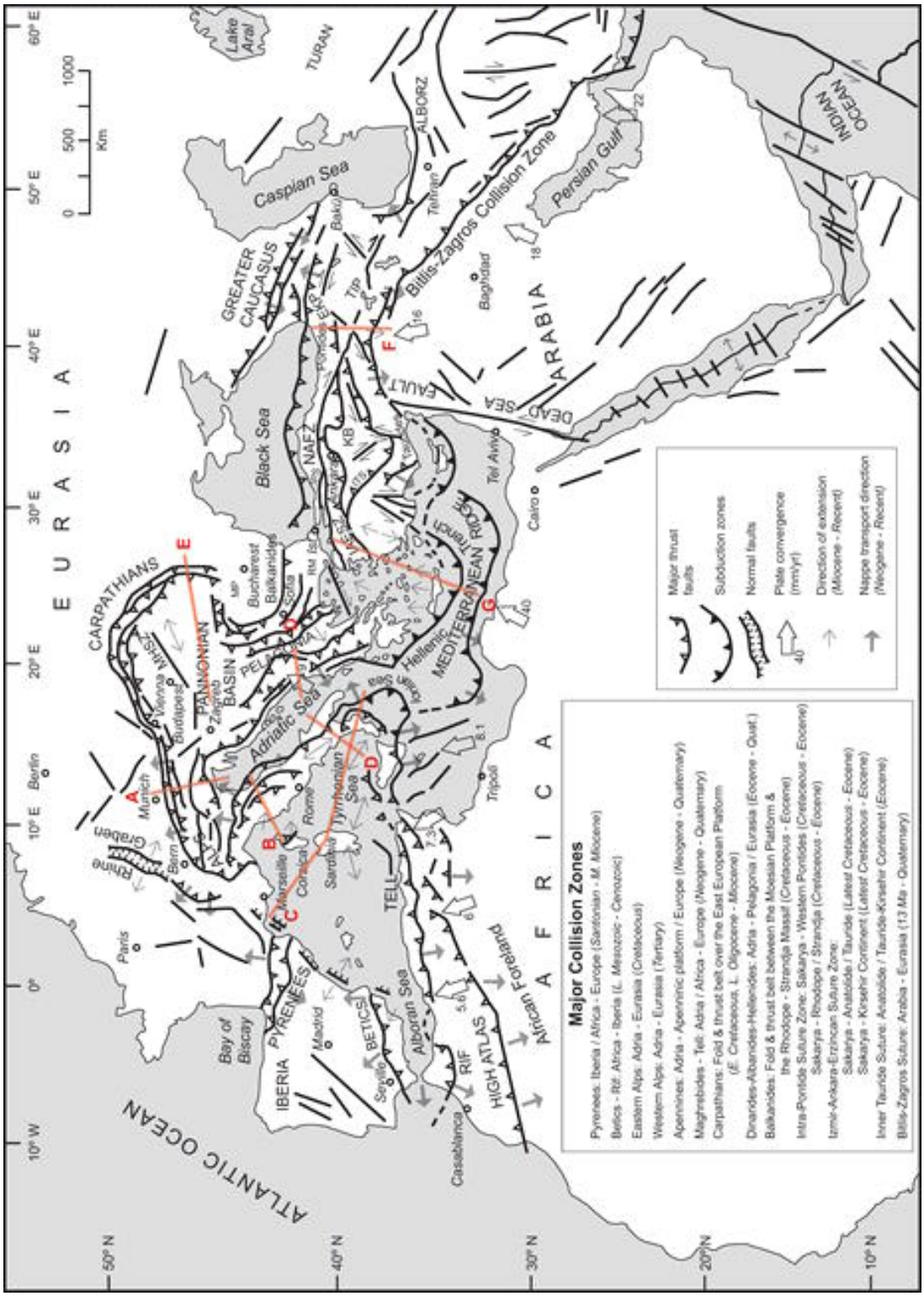

Figure 5-1: Tectonic map of the Mediterranean region showing plate boundaries and collision zones. Red lines are profile lines of additional figures in Dilek (2006).

5-2 Western Asia

Here, Western Asia spans from the Ural Mountains and the Middle-East over Central Asia to Mongolia, the Himalayas and Xinjiang Province, China (Figure 5-2). This part includes many different geological structures, old orogenies, failed rifts and strong influences of the Indian Plate movement, resulting in the Himalaya Mountains and the Tibetan Plateau (Burbank et al. (1996)). All plate boundaries in this area are involved with the Eurasian plate, which dominates the orogenic processes through interaction with several other plates (Gatinsky et al. (2011)). East from the alpine orogeny, the Persia-Tibet-Burma Orogeny starts (Liu and Bird (2008)). Plates affected in the study area are the microplates of the Aegean Sea- and Anatolian Plate, the Arabian Plate, the Indian Plate, the Amur Plate and the Burma Plate. The Aegean Sea- and Anatolian Plate are moving away from the Eurasian Plate, but are slammed back by the African Plate at the same time (Dilek (2006)). This results in complicated fault systems experiencing a high amount of loading in a shear transitional strike-slip movement with additional influence from the Arabian Plate from the east slamming into the Anatolian Plate moving it towards the African Plate (Chang et al. (2010)). The Arabian Plate is moving towards the Eurasian Plate and collides into it in a convergent Plate boundary. It is currently experiencing a divergent plate boundary with the African Plate due to ongoing rifting along the East-African Rift (EAR) and the Triple Junction along the Red Sea. The uplift from the lift results in the Arabian Plate tilting, leading to the Arabian Plate to slip under the Eurasian Plate and causing the Persian orogeny along Iran, and Afghanistan (Stern and Johnson (2010)). Mountains that are a direct result from this process are the Zagros Mountains and the Central Iranian Range.
The Indian Plate moves approximately 5cm per year upwards in a NE-SW orientation (Copley et al. (2010)). It is converging with a force that leads to the Plate forcing itself underneath the Eurasian Plate resulting in the formation of geologic surface structures in Northern India including the Himalayan Mountains and the Tibetan Plateau. During the collision event, multiple rifts and microcontinents contributed to orogeny in Central Asia resulting in the mountain ranges of Tian Shan, Karakoram, Kunlun, Hindu Kush. Lake Baikal is a direct result of the failed Baikal Rift Zone (ten Brink and Taylor (2002)), contributing to its depth. Along Pakistan and Myanmar, the convergent plate boundaries between the Indian Plate and the Eurasian Plate forced slabs to get pushed in an almost 90° horizontal orientation in the mantle direction due to the torque forced by the Indian's Plate fast movement (van Hinsbergen et al. (2011)).

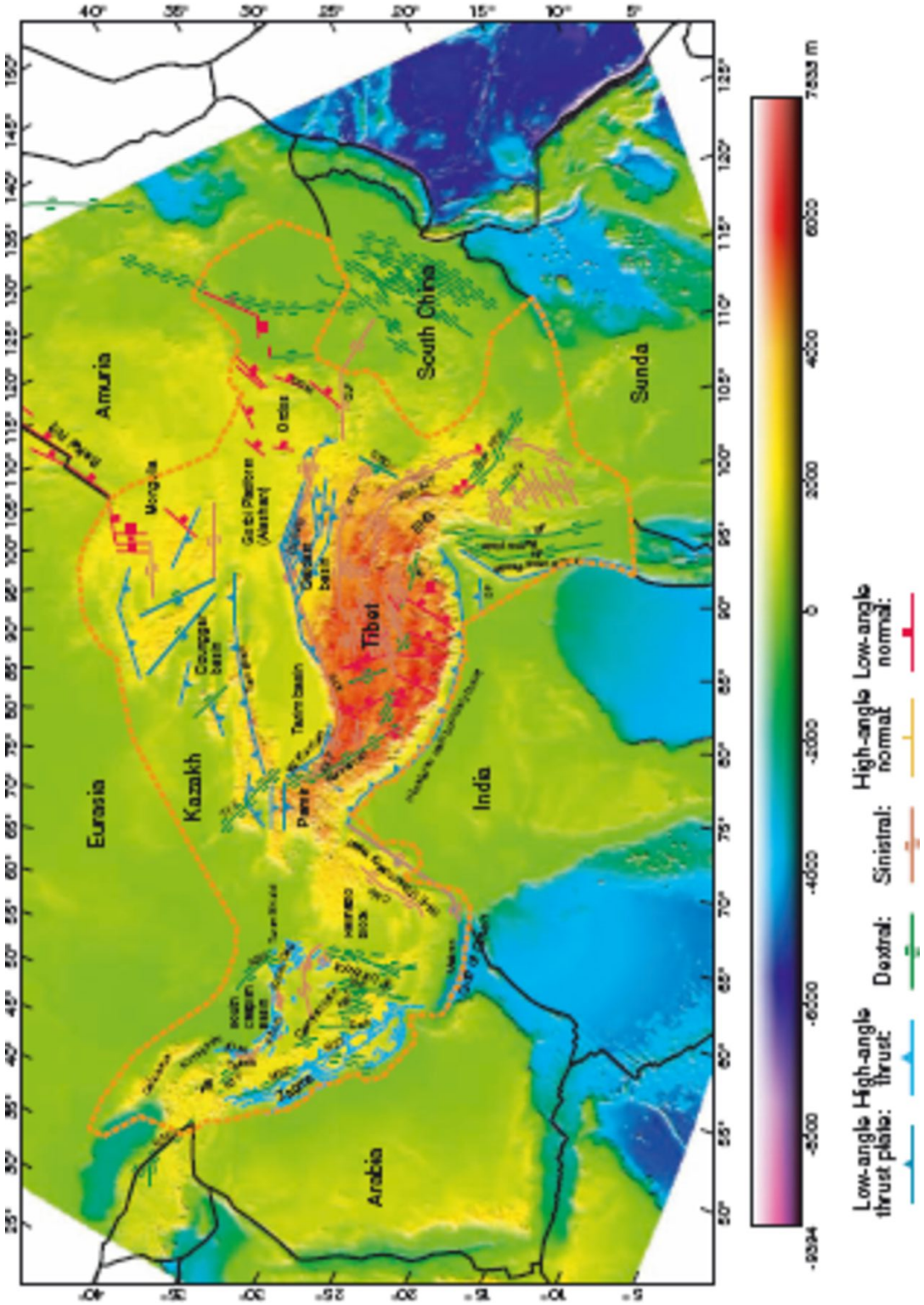

Figure 5-2: Tectonic Map of major orogenies in Asia. Active faults are represented by thin lines. Symbols on the fault traces and colours represent fault type: open triangle and light blue trace = high-angle thrust fault; solid triangle and dark blue trace = low-angle thrust fault; green = dextral fault; brown = sinistral fault; thin line and yellow trace = low-angle normal fault; solid square and red trace = high-angle normal fault. (Liu and Bird (2008))

Chapter 6

Data

Event data from Earthquakes forms the datasets of the waveform inversion. Each event is downloaded from public servers on a HDF5 file format through the LASIF package. The Federation of Digital Seismograph Networks (FDSN) is a global organisation comprised of multiple member institutions that provide three-component broadband seismic data with a standardised data format from their networks. Depending on the network access, waveform data slices from receivers can be accessed with a metadata query that specifies for which timeframe the data should be downloaded from the server. LASIF automates the process of data downloading by collecting the metadata of earthquakes and mass downloading waveforms from receivers within a specified domain area on the globe. Event metadata is provided by the Global Centroid-Moment-Tensor Project's (GCMT) Catalog (Ekström et al. (2012)). (1) Create HDF5 files for every event inside the domain that corresponds to the search parameters and store the origin time, location- and moment tensor information. (2) Given a reference velocity, find out the arrival time at a receiver of an event. (3) Download the broadband waveform data from a FDSN receiver in a timeframe starting at the arrival time and ending at the arrival time plus the specified timeframe of interest e.g. 3600s. (4) Incorporate waveforms in the waveform group of the HDF5 file and store the metadata for each receiver name and location. (5) Remove inconsistent receiver data i.e. chopped up, non-consistent data in the specified timeframe. The event data is stored in a LASIF Project, where it can be processed, forward modelled etc. LASIF in conjunction with obspy (Beyreuther et al. (2010); Megies et al. (2011); Krischer et al. (2015b))allows to plot on global map projections, where you can double check event and receiver locations as well as the raydensity of the entire event dataset. If hands-on quality control is needed, LASIF has a built in GUI, where individual waveforms and their components, synthetic waveforms from progressed iterations in an inversion process and picked windows can be plotted and inspected. The study areas comprise to datasets: One for EUWA (Figure 6-1) and one for EU (Figure 6-2)
Additional closed-source data has been provided by Zaher Hossein Shomali from Uppsala University, Sweden (Lund et al. (2021)), Prof. Martin Mai's Group from King Abdullah University of Science and Technology, Saudi-Arabia and Pálmi Erlendsson from the Icelandic Meteorological Office.

6-1 Base Model

The base model is the starting reference for a waveform inversion problem given in 2-18. Between iterations, the L-BFGS history tracks changes in relation to the base model. Putting importance on the choice of the reference. If the reference is flawed, the end model will be flawed. For the base model the second generation Collaborative Seismic Earth Model (CSEM) was chosen (Afanasiev et al. (2016); Fichtner et al. (2018); Noe et al. (2023)). The general concept of the CSEM is to provide an evolutionary, multiscale inversion model of geophysical data. Different models, depending on their purpose, scale and dataset, are difficult to include into one all encompassing model, as the formulation of the inversion problem differs between models. For example, regional studies are done with higher data density in comparison to global studies were the data density is more coarse. With technologies improving and more data being recorded, CSEM sets out to save on human resources in research, by allowing successive updates of models with more information, while staying consistent with prior information. This makes the CSEM a general base model with prior knowledge not only from seismic data, but also multiscale gravity, electromagnetic and topographical and bathymetrical Global Positioning System data. Included models in the study area are: Anatolian EU model (Fichtner et al. (2013b)), Central-Eastern and Western Mediterranean (Blom et al. (2020); Fichtner and Villaseñor (2015)), Iceland and Jan-Mayen (Rickers et al. (2013)), the LOWE model (Thrastarson et al. (2022)), the Africa model (van Herwaarden et al. (2023)), and the China model (Ma et al. (2022)). These models will not be thematised further, as the structures imaged from these studies will be observable in the end model.

6-2 EUWA Dataset

EUWA is a shorthand for "Europe and Western Asia and encompasses an area from the MOR to the Himalayas with a maximum epicentral distance of 150°. The longest source-receiver distance in the dataset is 110°. Earthquake events with a Moment Magnitude M_0 between 5.5-6.7M_0 were collected for this dataset, since sufficient energy is needed to have clear body wave arrivals and an acceptable signal-to-noise ratio. The dataset consists of 160 events, with 258.335 unique raypaths and 7.370 unique stations. The events were selected such that there is sufficient coverage of the areas of interest. A bimodal difference in receiver density shows Europe is filled with receivers, while the receiver density in Western Asia is sparse. This resulted in widening the study area towards Western Asia to include events and receivers in Central Asia and the Himalayas. Events in Western Asia then span the study area more evenly with raypaths towards receivers in Europe. Resulting in higher coverage along the Black Sea, Anatolia, Saudi Arabia and Iran. The densest coverage of raypaths is therefore in Continental Europe, Anatolia and the Caucasus.
Beyond focusing on events with sufficient Moment Magnitude, a focus was also given on downloading newer events. Similar studies of EUWA have been done before 2015, when multiple Array missions with upgraded data quality were started (Hetényi et al. (2018); Network (2015); Lund et al. (2021); Institute Earth Sciences "Jaume Almera" CSIC (ICTJA Spain) (2007)). More data is available in quantity as well as in quality since 2015. Revisiting a waveform inversion of the study area is therefore promising and may reveal more heterogeneous structures than previously imaged.

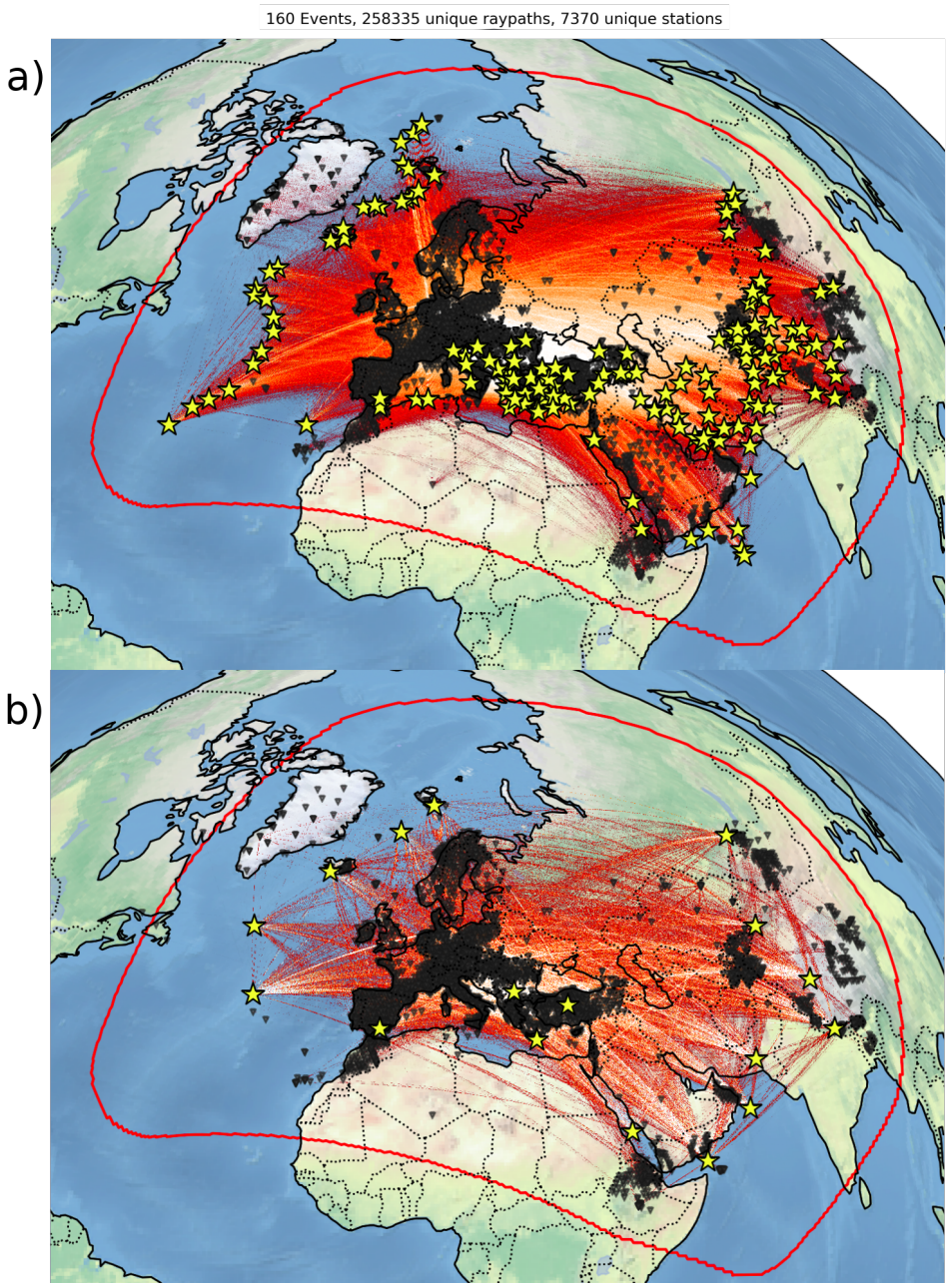

Figure 6-1: a) Raydensity plot of the EUWA dataset. White areas denote the highest density of overlapping raypaths. Yellow stars denote the event sources and the black triangles denote the receivers. b) Raydensity plot of the validation dataset. The vaidation dataset track the overall total misfit of the inversion. Placements of validation events are focused on radial coverage as well as covering areas where events cluster. This follows evident influence of the FWI.

6-3 European Peninsula Dataset

A dataset has been prepared focusing on a smaller area centered around the European Peninsula and exploiting information from more numerous available events with Magnitudes between 4.2-5.7M_0. The dataset has been prepared should the EUWA dataset reach limitations concerning data fitting at sub 20s periods or data sizes of meshes, resulting in speed limitations of I/O between the local and remote supercomputer host. The dataset consists of 300 events, with 492.602 unique raypaths and 5.277 unique stations. It is intended to use the currently full available receiver and waveform density in Europe. However, this does not guarantee there will be more imaged structures compared to the EUWA dataset at similar period bands. The advantage of this dataset would be the circumvention of issues that would arise with the EUWA dataset. These issues concern the sensitivity of far-distanced source-receiver pairs between Europe and Western Asia and if data at lower periods will be able to propagate clear arrivals from Western Asia till Europe. If that is not the case, a domain focused on Europe is considered, as there are more events and receivers available.

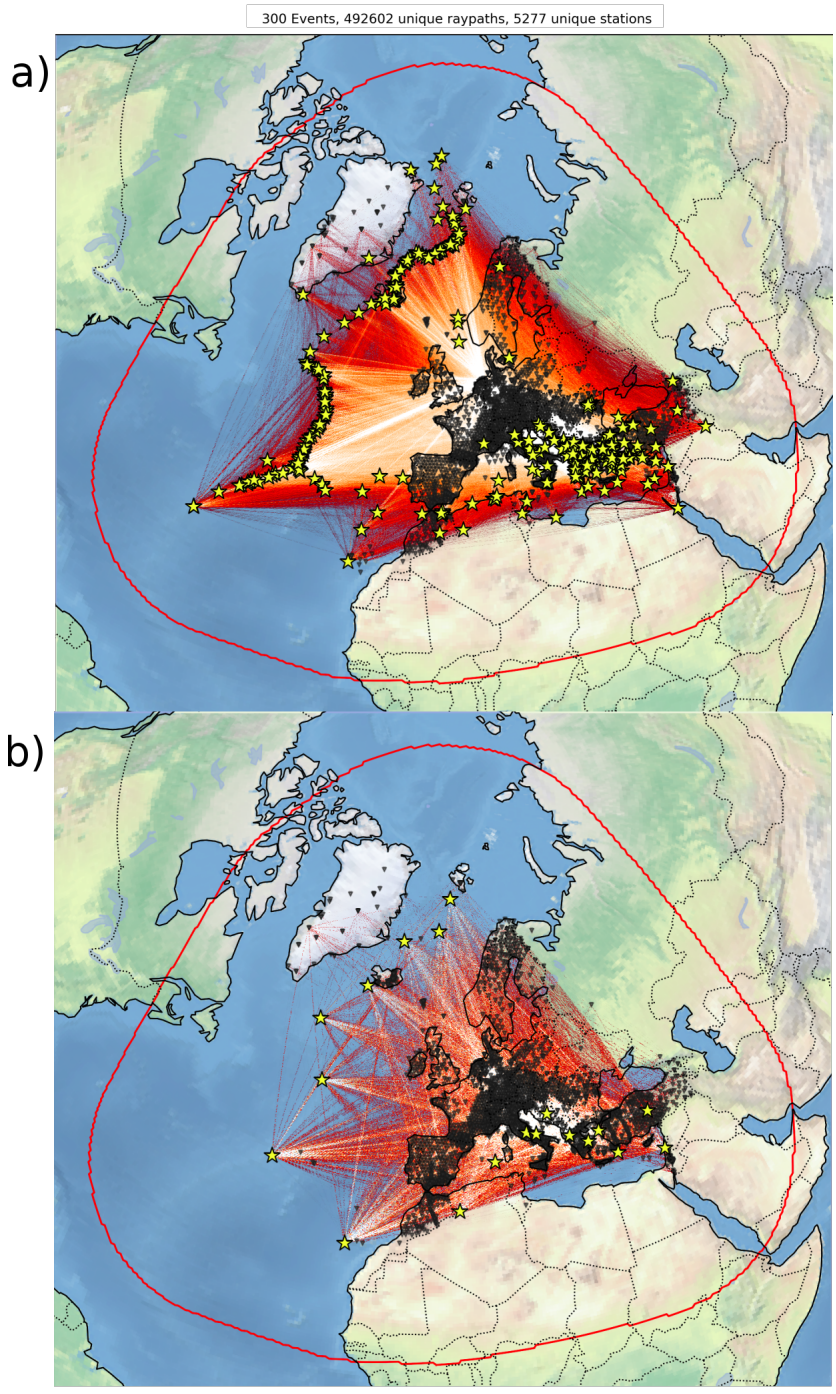

Figure 6-2: a) Raydensity plot of EU dataset. b) Raydensity plot of validation events.

Chapter 7

Results and Discussion

7-1 Computational Costs

SmoothieSEM meshes on a regional continental-scale have shown in the first period bands to achieve a FWI model with negligible differences when compared with a cartesian mesh FWI model. SmoothieSEM meshes save computational costs by multiple factor of difference depending on the minimum period of SEM modelling (Figure 7-1). The period- i.e. frequency scaling of required mesh elements is in agreement with the frequency scaling result shown in (Thrastarson et al. (2020)). For example, comparing the meshes at a minimum period of 35s, there is a noticeable difference in mesh elements. A rule of thumb, is one node per 3.000 elements. The cartesian mesh has over 1 Million elements at a 35s period. To compute the Forward and Adjoint Problem of one iteration on a cartesian mesh, 264 nodes out of 600 nodes had to be used. Given an average compute time of 12 minutes for 40 events per iteration, 5 iterations would lead to a usage of 2.112 node hours each for the Forward and Adjoint Problem. 2.112 node hours of 700.000 node hours with every 5 iterations can increase the resource usage to an unviable cost-benefit ratio. A SmoothieSEM at the same period will vary in number of elements depending on the location of an event. Taking a median the number of elements at 35s is. Only 84 nodes out of 600 nodes had to be used. Furthermore, SmoothieSEM meshes can have a higher time step then cartesian mesh, due to only needing the minimum time step from the radial direction. Average compute time is 6 minutes for 40 events per iteration. 10 iterations leads to a usage of 336 node hours each for the Forward and Adjoint problem with every 10 iterations.

To illustrate this usage on a production level, the cumulative usage graph of the of the Seismology and Wave Physics Group's quarterly compute budget is shown in Figure 7-2. Using Inversionson with every event simulation on a standard cartesian over a span of three days and about 20 Iterations, 8% of the quarterly budget was used i.e. circa 20.000 node hours. This is 2% of a year's compute budget, which makes the method on large cartesian meshes prohibitively expensive. In comparison, the MultiMesh method of Inversionson coupled with SmoothieSEM meshes used in the same amount of iterations and days circa 1.500 node hours. This is approximately a 13 times factor of difference.

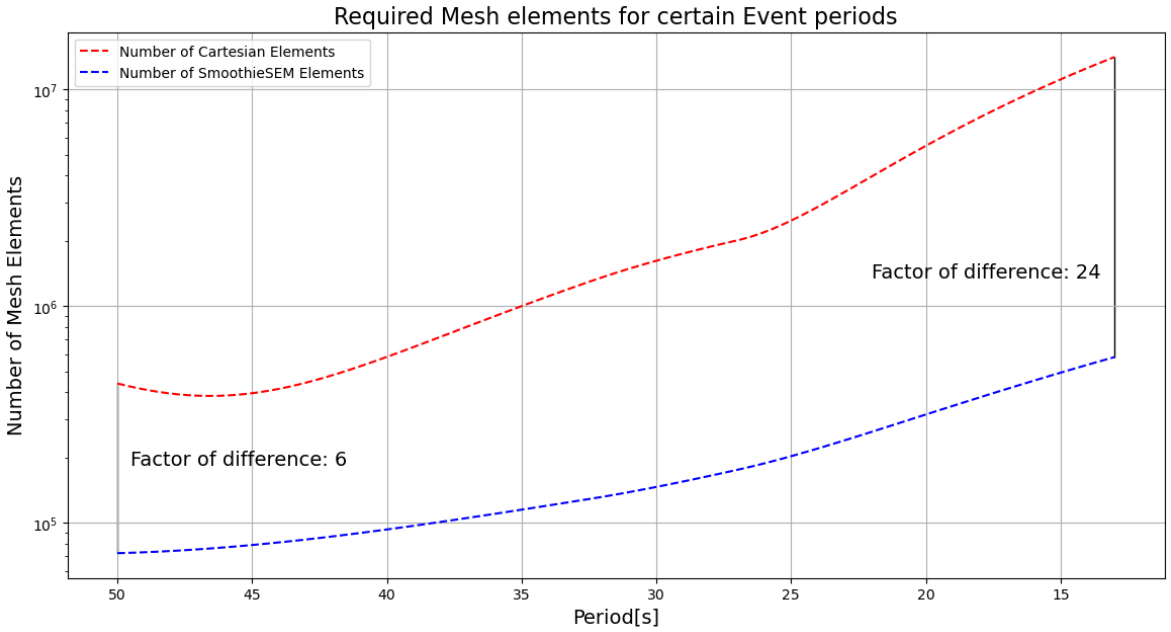

Figure 7-1: Red dotted lines show the increase in mesh elements with lower period, while the blue dottel lines show the increase in mesh elements for a range between 50s and 13s. The factor of difference increases with the increase in period.

These results show that the SmoothieSEM meshes save compute resources for regional continental-scale FWI. A model at low periods can be iterated while using 10 or more times less computing resources, enabling to push FWI tomographies further into lower period bands then previously possible. As long as low period waveforms have clear arrivals, and the information can be exploited in the FWI. One concern remains in relation to the SmoothieSEM meshing and required elements. In Figure 7-3, periods lower than 24s involve SmoothieSEM meshes with lateral refinements appropriate for 24s. What refinements are necessary for which period of data are tested, before implementation in Inversionson. The mesh elements required for periods lower than 24s could be subject to change, but the factor of difference will only differ by a bearable amount between ±1-2 factors. With this development, the constraining costs pivots from the actual simulation, to Data Management and efficient I/O. Currently, the issue is data storage as the individual model updates can reach up to 10-20GB each, resulting into easily filling up 1TB in just 5 iterations. Luckily, the actual memory storage and I/O of meshes has been addressed in a previous study by Hapla et al. (2021) such that only having sufficient hard drive storage or memory becomes a bottleneck.

7-2 Current iteration and Misfit evolution

Before the SmoothieSEM method was fully developed to work with MultiMesh in Inversionson, the Cartesian mesh was used for the period bands 50s-120s, 35s-120s and 35s-90s respectively. Two things were considered. First, what settings concerning smoothing lengths, mini-batch size, simulation time and optimal period bands. Lastly, following results in van

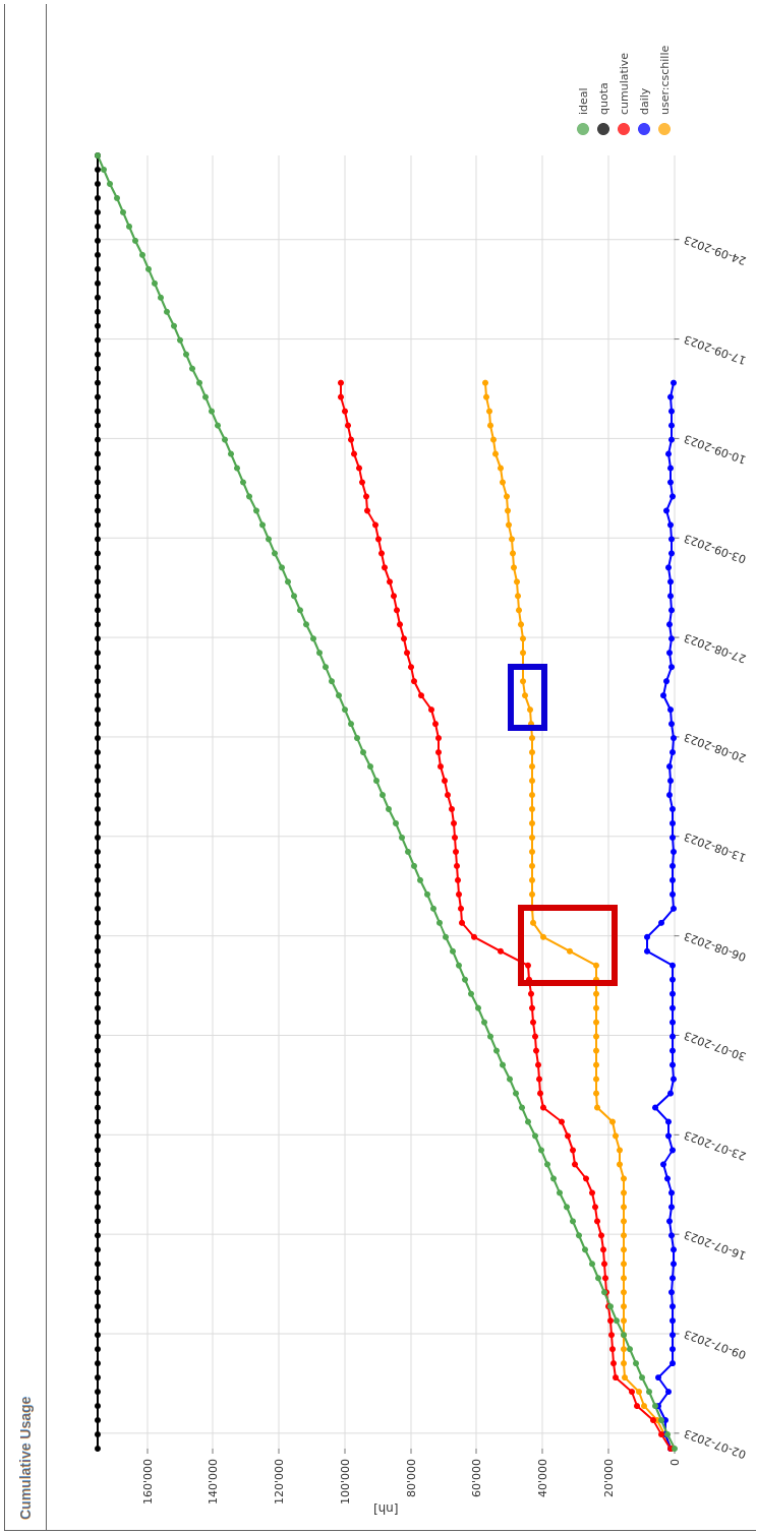

Figure 7-2: Quarterly compute budget of project s1168 for the fourth quarter (01.07. - 30.09.). Red Box: Node hour usage over three days using a cartesian mesh for every event. Blue Box: Node hour usage over three days using a SmoothieSEM mesh for every event.

Herwaarden et al. (2023) about restitution tests, the period bands 50s-90s and 35s-90s were redone using the same base model and the MultiMesh method with SmoothieSEM meshes to compare directly with the results from the Cartesian mesh. During an inversion, the misfit history has to be checked frequently. If the misfit increases or does not decrease in a converging manner, then lowering the smoothing lengths should be considered. The period band of 35s-120s was an outlier, as the inversion history shows changes in the smoothing lengths, but also simulation time and lowering of the maximum period from 120s to 90s.
The first period band at 50s-120s was started using the CSEM 2.0 base model at 50s. The base model is tuned for 50s periods, which results in faster convergence of the information content towards a minimum after 20 iterations. The model iteration with the lowest misfit is taken as the base model of the next period band at 35s-120s (Figure 7-5). This period band has shown some issues with simulation and mini-batch settings and needed three restarts. The starting mini-batch amount was 20 events per iteration, with a 1600s simulation time for the propagating wavefield. First, the period band seems to converge too quickly after 15 iterations. Secondly, events and source-receiver pairs with long epicentral-distances have a much smaller amount of picked windows compared to a 50s period band. Inspecting the data shows that, for the size of the domain, a simulation time of 1600s is too short for lower periods and some body wave arrivals as well as surface waves, could not be captured fully in one window. Altering these settings require a restart and the new information to converge. First the simulation time was increased from 1600s to 2400s. Restarting the inversion with the increased simulation time did not show improvements in window picking. Some further tests showed, that for many waveforms the information in periods 90s-120s introduced a significant amount of noise (Figure 7-3). Enough for the window picking algorithm to not pick windows. Background noise spectra of seismic data show an increase in noise for periods higher than 90s (Peterson (1993)). Exclusion of data between 90s-120s may mean a decrease in senstivity at depths near the Core-Mantle boundary and Lower Mantle. The trade-off was made with the deliberation, that for the goal of pushing to lower periods, investigating the crust and effects of the 1-D Q-attenuation model at shallow depths is of more interest. Additionally, sensitivities in the lower mantle are provided by the LOWE model, that is included in the base model (Thrastarson et al. (2022)). After these changes, the total misfit could decrease and converge again with more picked windows.

With the implementation of SmoothieSEM meshes for regional continental-scale FWI, the method was started from the beginning at 50s-90s using the CSEM base model. Similar to a restitution test, the inversion is done with the SmoothieSEM, to see if the errors introduced through interpolation between Master- and SmoothieSEM can still converge to a similar result. This would also show that in the FWI, as long as the same data is available, the model can be restituted. A difference between this test and the restitution test in the aforementioned study, is that the final model has not been deliberately perturbed and then inverted, but that the inversion is done from the same starting point. A restitution test of the final model from deliberate perturbation will be considered in the Outlook. The inversion with the SmoothieSEM mesh shows a similar convergence behaviour as with the Cartesian meshes (Figure 7-4). The difference between the final iteration of the SmoothieSEM and Cartesian mesh show that there are negligible differences similar to Thrastarson et al. (2020) (Figure 7-6). This also shows, that at a minimum period of 50s, the noise at higher maximum periods seem to not influence the results, most likely to the base model tuned for 50s. The restitu-

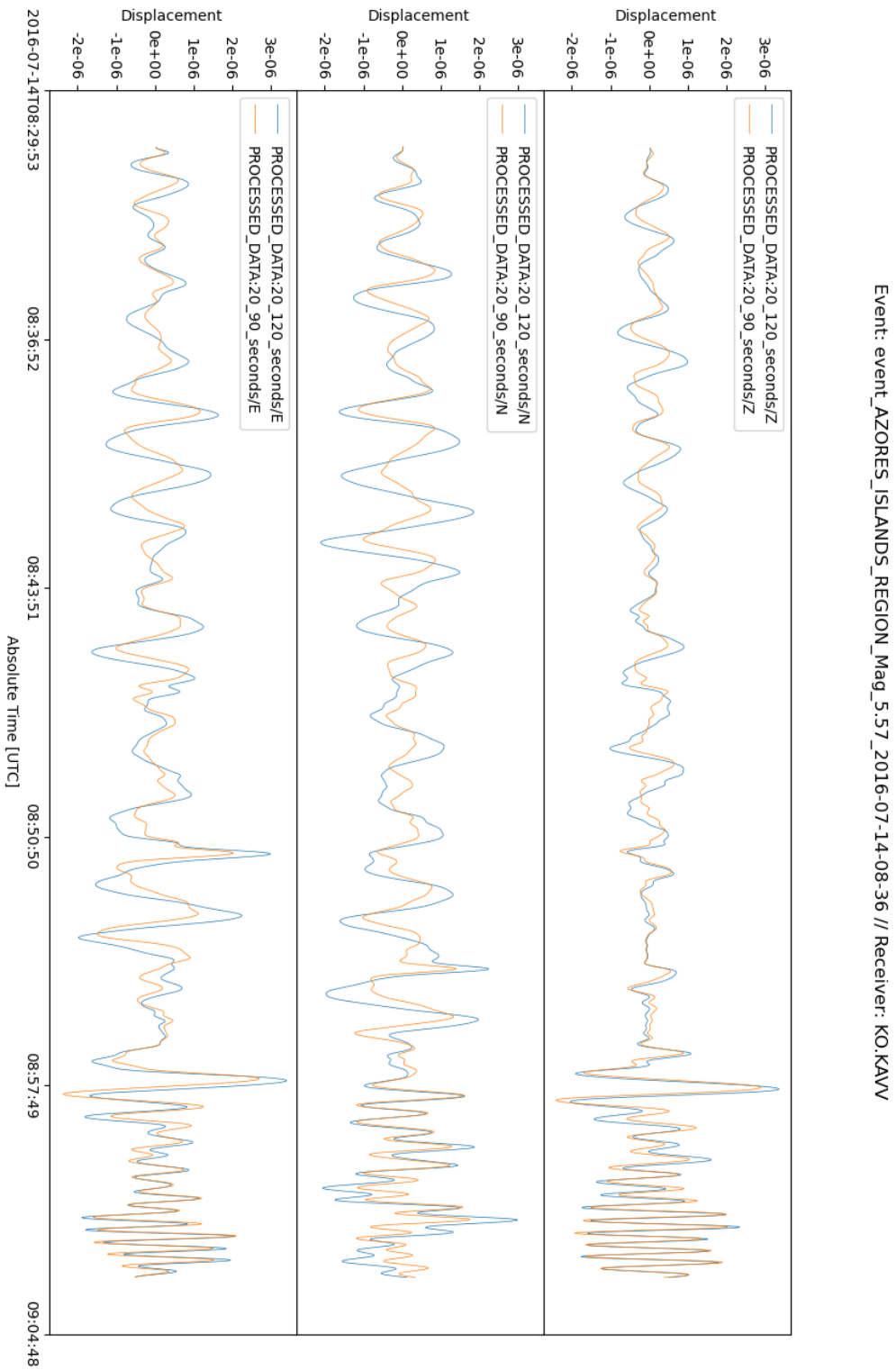

Figure 7-3: Displacements from components N,E,Z of the observed data of Receiver: KO.KAVV from the Azores Island event from 14th of July 2016. The data was processed to include frequencies between the periods 20s - 90s and 20s - 120s respectively. Excluding frequencies between 90s - 120s shows clearer arrivals and less noise. Blue: Data processed for 20s-120s. Orange: Data processed for 20s-90s

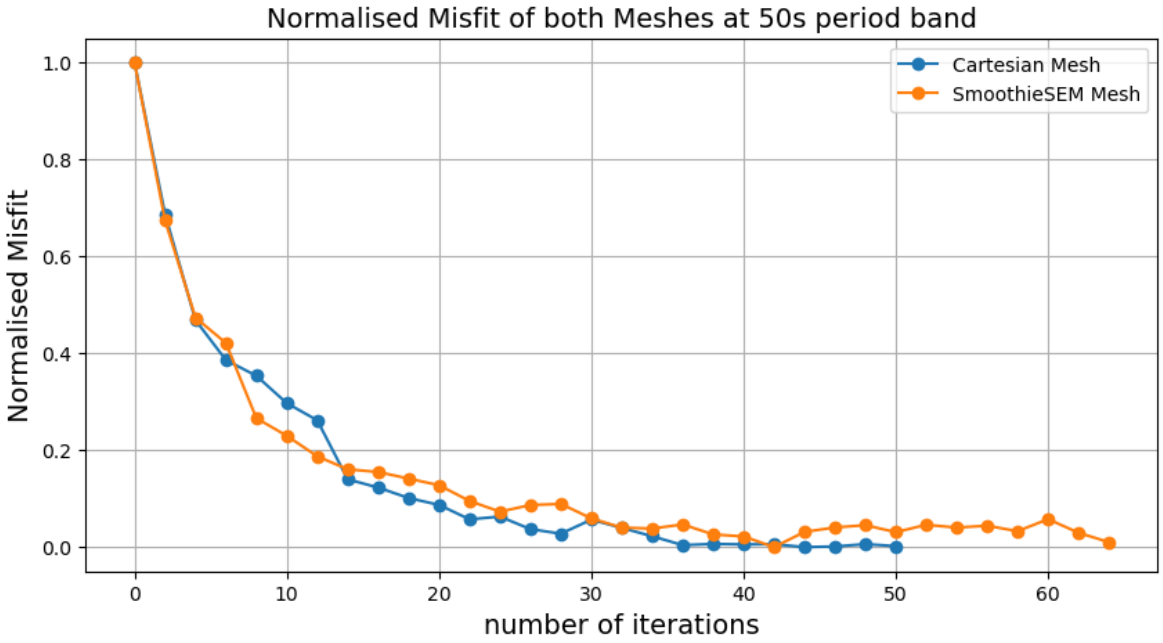

Figure 7-4: Total misfit evolution of the validation data set at 50s-120s/90s. Comparison between the Cartesian mesh and SmoothieSEM mesh show similar behaviour.

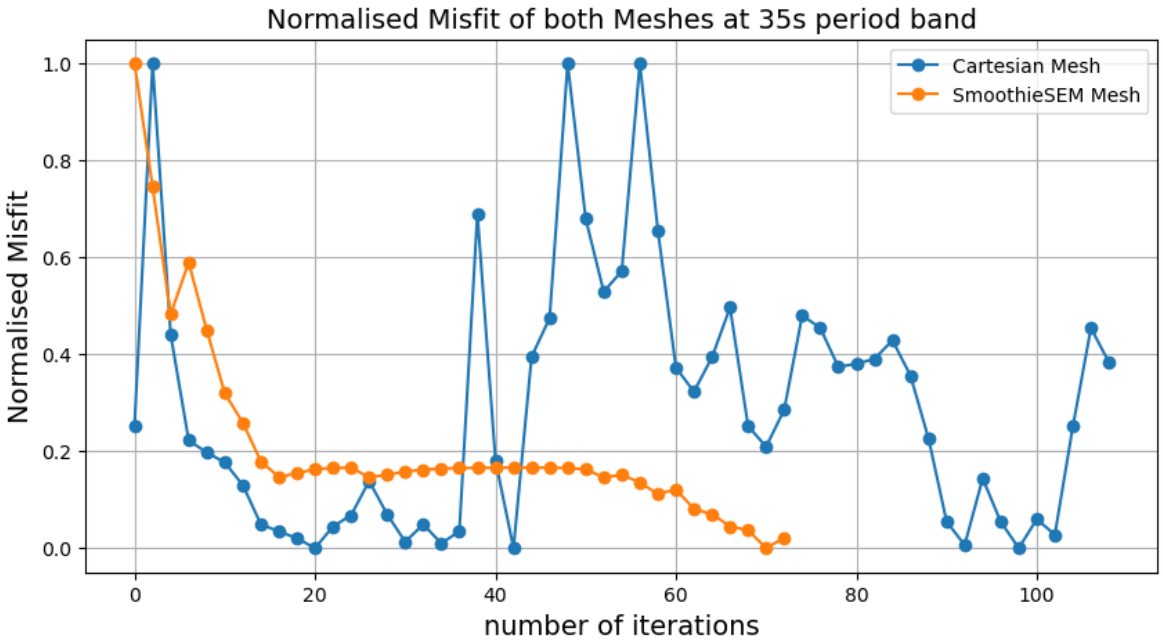

Figure 7-5: Total misfit evolution of the validation data set at 35s-120s/90s. Comparison between the Cartesian mesh and SmoothieSEM mesh. The rocky history of the Cartesian mesh is due to experiments on simulation time, batch size and background noise. The SmoothieSEM mesh shows how the general behaviour looks like with the changed settings.

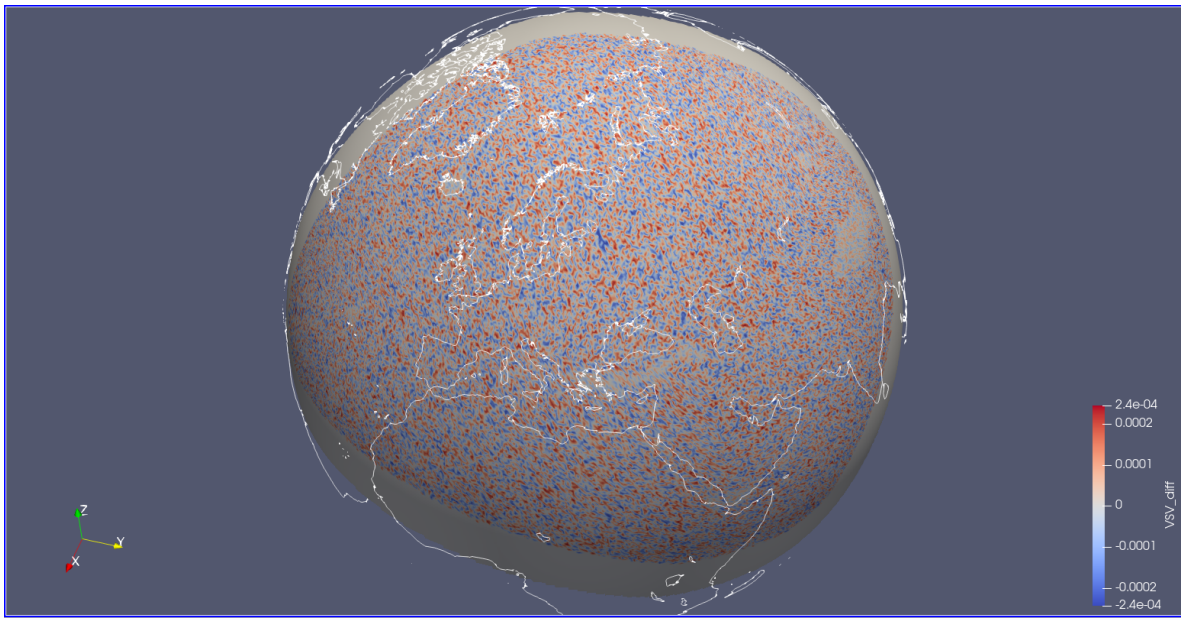

Figure 7-6: Absolute difference between the Cartesian mesh and SmoothieSEM mesh at a period band of 50s-120s at a similar converged misfit. It shows a random pattern of values and no coherent differences. The difference is under 1% similar to the results on differences between 2-D Cartesian and SmoothieSEM meshes (Thrastarson et al. (2020))

tion shows that the SmoothieSEM meshes can reconstruct the model with a similar misfit behaviour, while using 6 times less resources than the Cartesian mesh at a minimum period of 50s. The SmoothieSEM mesh inversion at 35s-90s shows an acceptable misfit behaviour of convergence, when using the changed settings (Figure 7-5). Raising questions on the importance of the maximum period of investigation and how to exploit waveforms for a given study case.

Currently the inversion stopped at a minimum period of 24s. The total misfit history shows the convergence behaviour at each period band (Figure 7-7). Bands 50s-90s and 35s-90s have been mentioned in the previous paragraph. Bands 30s-90s and 24s-90s show a different and more interesting behaviour. Every period band started with smoothing lengths at $[0.4, 0.8, 0.8]$. At period band 30s-90s, the smoothing lengths were lowered at the beginning to $[0.35, 0.7, 0.7]$ because of a growing misfit. At iterations where it increased, the misfit was lowered to $[0.3, 0.6, 0.6]$. Leading to the misfit steadily decreasing over the course of the inversion. Normally a convergence behaviour should be observed like in period bands 50s-90s and 35s-90s. As soon as the misfit was sufficiently low, the inversion ended to save computating resources. Period band 24s-90s had another issue. The misfit behaviour dropped off quickly during inversion and individuel event misfits were low. Considering the changes made during period band 35s-120s, this could either be an issue with the data and picked windows or the batch size. The batch size remained 40 events per iteration since band 35s-90s. The batch size was increased to 70 events seen as the spike at iteration 102. The expectation is that with a higher batch count, the dataset can be explored faster and the L-BFGS algorithm will be able to move towards the global minimum with a bigger exploration radius.

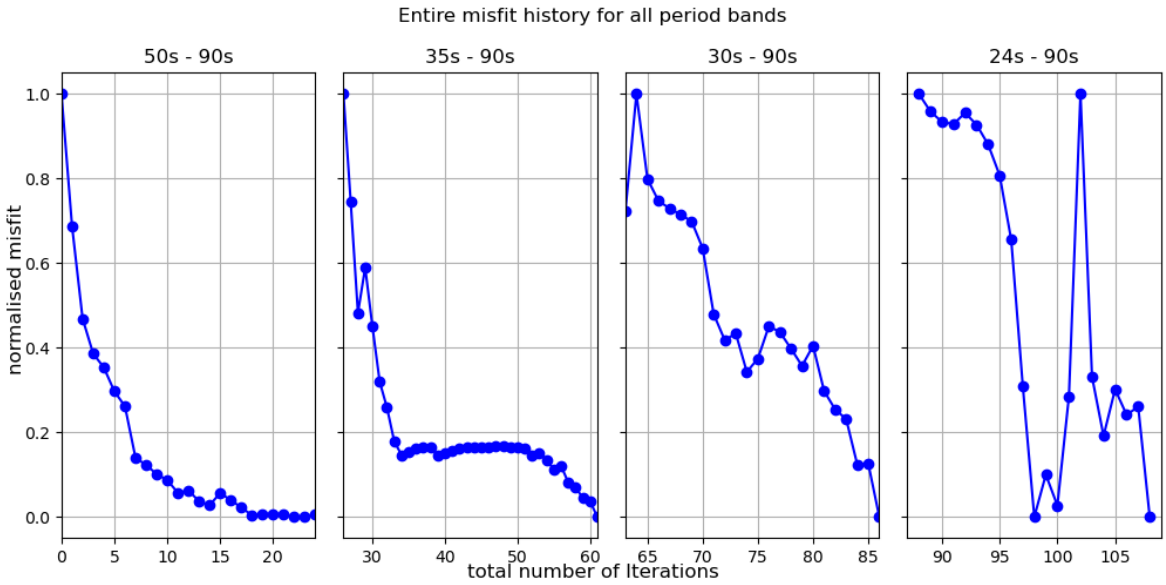

Figure 7-7: Misfit evolution of the validation dataset of all inverted period bands using the SmoothieSEM meshes. Period band 24s-90s is mid-inversion.

7-3 Imaged Structures

S wave velocity will be shown over P wave velocity as it is most sensitive to surface waves (van Herwaarden et al. (2023)). This fact is supported by previous study in S-wave velocity and sensitivities towards intrinsic structures (Panning and Romanowicz (2006); Capdeville et al. (2010)). Only known apparent structures in the upper mantle of the study area will be pointed out as many tomographies beforehand (Fichtner and Villaseñor (2015); Zhu et al. (2012); Fichtner et al. (2013b)) have overlapping results in shear wave anomalies. Any mantle-geodynamic interpretations will not be attempted as seperating quantitative contributions in anisotropy is not possible (Fichtner et al. (2013a)). What should be observed, is that for certain structures VSH and VSV agree, and that any new structures from lower periods, will more likely be added in VSV through surface wave contribution.
The tomography is divided in slices at 40 km, 70 km, 120 km and 240 km of depth for the VSV. At 40 km depth (Figure 7-8), noticeable geologic structures as slow velocity anomalies are most mountain ranges in Europe and Western Asia in high detail. This includes the mountain ranges of the Pyrenees, the Alps, the Appennines, the Carpathians, the Dinaric Alps, and the Pontic Alps in Europe. In Western Asia, the Central Iranian Range, Zagros Mountains, the Hindu Kush mountain range, the Tian Shan mountain range and the Himalayas can be identified. Some areas of subsidence can be identified as fast velocity anomalies, including the Po Valley, the Pannonian Basin, the canyons among the Dinaric dolomites and the Indus Valley in India. At 40 km for VSH (Figure 7-12), the mountain ranges are still identifiable as slow velocity anomalies and more prominently than in the VSV data, Important is the imaged slow-velocity sutur zone, the Teisseyre-Tornquist Zone, which has a NW-SE orientation along Poland, Ukraine, Moldova and Romania, that divides Europe from a fast velocity anomaly, the stable East European Craton. In northern Europe slow velocity anomalies in Iceland are imaged originating most likely from interactions with the MOR. More areas with low-velocity

anomalies are Greenland, Scandinavia and Northern Russia. These velocity anomalies may originate from isostatic loading and rebound.
At 70 km (Figure 7-9), the roots of mountain ranges are still visible as low velocity anomalies. In Africa, the root of the Atlas mountains can be observed as a "tear"-like structure along the border of Libya and Chad. For low velocity rift zones, prominent in East Africa is the EAR and its Triple Junction, with the tilting of the Arabian peninsula. In Europe the Massif Central as well as the Rhône Valley and the Upper Rhine Valley as part of the Mediterrinean-Myosen Zone are imaged. Other low-velocity anomalies in the area could be mountain ranges in Germany, and a geothermal gradient in the Netherlands. VSH data also shows the suture zone between the European Cratons (Figure 7-13). Additionally, it is now showing a prominent low-velocity anomaly, the MOR along the Atlantic Ocean. VSH seems sensitive towards the discontinuities caused by the transform faults of the MOR, as some transform faults are clearly imaged. It also shows a finger-like structure near Iceland and the Jan-Mayen islands. This structure agrees with shear velocity results from a previous study about the area and the MOR (Rickers et al. (2013)), which is the interplay of two hotspots with southeast tilted conduits.
At 120 km (Figure 7-10), there seems to be a slow shear velocity structure along the Western Mediterranean, that corresponds to some known structures like the Central Massif. It is a bigger structure that may indicate an area of ongoing change in the upper mantle. Some low shear velocity structures correspond to known areas of ongoing rifting. High shear velocity areas correspond with stable cratons and former microplates. In the North Sea a high velocity anomaly is appearing. There is not much information for interpretation about this high velocity. This anomaly and other blobs of high velocity anomalies appearing around the British Isles could maybe be remnants of old plates. In India, the 90 degree folding of the slabs subducting against the Eurasian and Burmese plate can be seen. At 240 km (Figure 7-11), a low shear velocity anomaly can be seen going along from the Central Massif, Rhône, Switzerland, over Bohemia, the Carpathians to Romania, the Black Sea and Western Anatolia. At these depths, the imaged lines may correspond with old plate boundaries and sutur zones from past geological processes. The deeper a boundary is imaged, the older the imaged plates could be e.g. Serbian microplate. Other low shear velocity anomalies could be the most bottom parts of a mountain root, Conduits from hotspots originating from the lower mantle, break-off blobs from compacted orogeny. Geological reconstruction as well as chemical evidence from these depths would support interpretations.

7-4 Comparison with CSEM Base Model

The current model should add data and anomalies, while preserving certain structures. For example, known mountains that were in the base model should ideally be observable with higher resolution in the current model. Prior information information where the base model corresponds with geological models, should not be overwritten. In the scenario of overwriting certain structures, two basic questions arise. Either the inversion and data was faulty or the overwritten structure was not there. The latter would suggest, when the structure was corresponding with prior geological and geochemical information, that all interdisciplinary models are wrong. This can be the case with overwhelming evidence, but Occam's Razor suggests that the former question is more likely and more easily answered.
Comparing the CSEM base model with the current model at different depth slices of VSV,

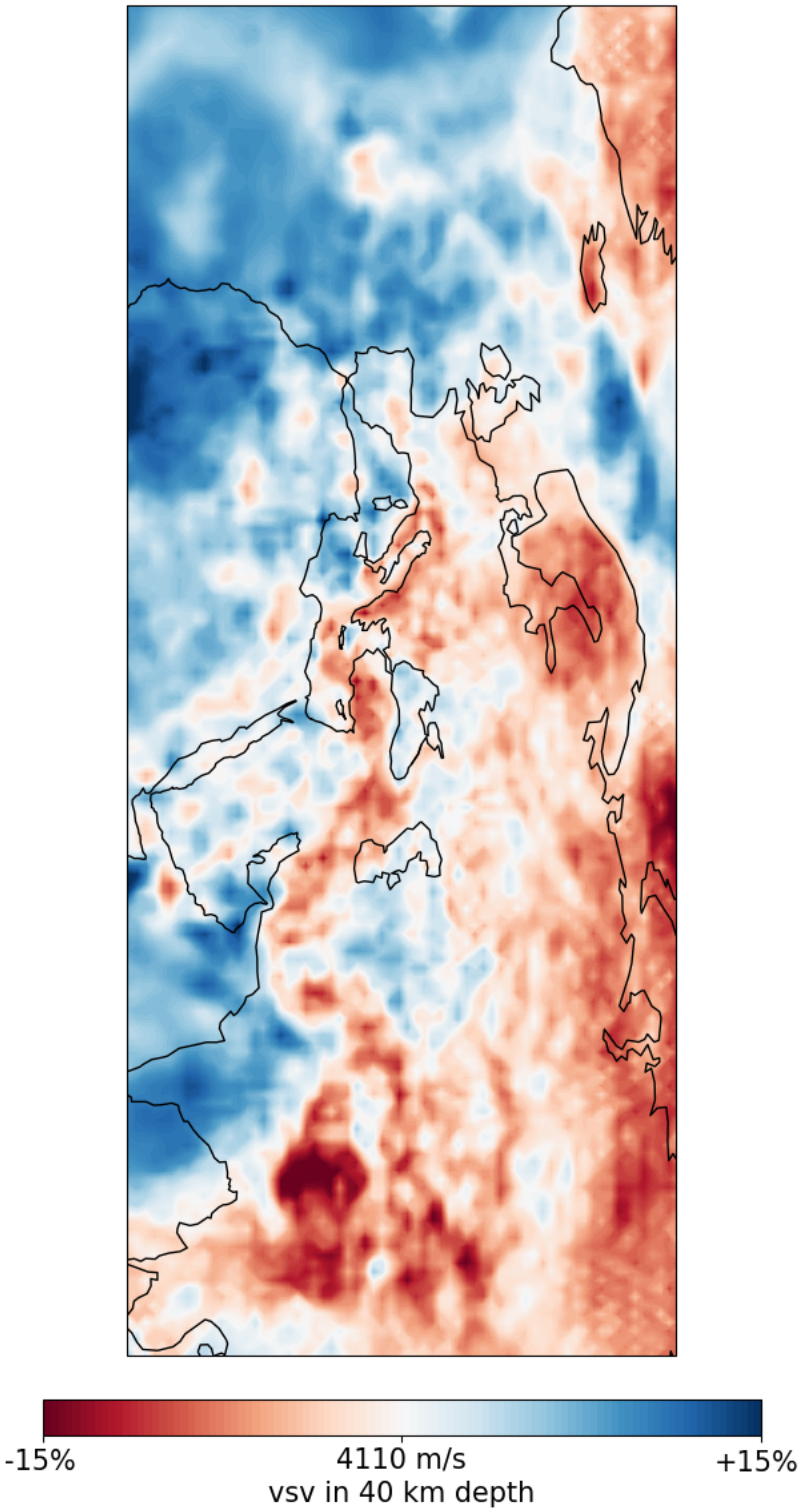

Figure 7-8: Vertical shear velocity model at 40km depth. Perturbation of the model data are plotted relative to the lateral mean. The percentage of relative change, that can show the perturbations clearly, is chosen

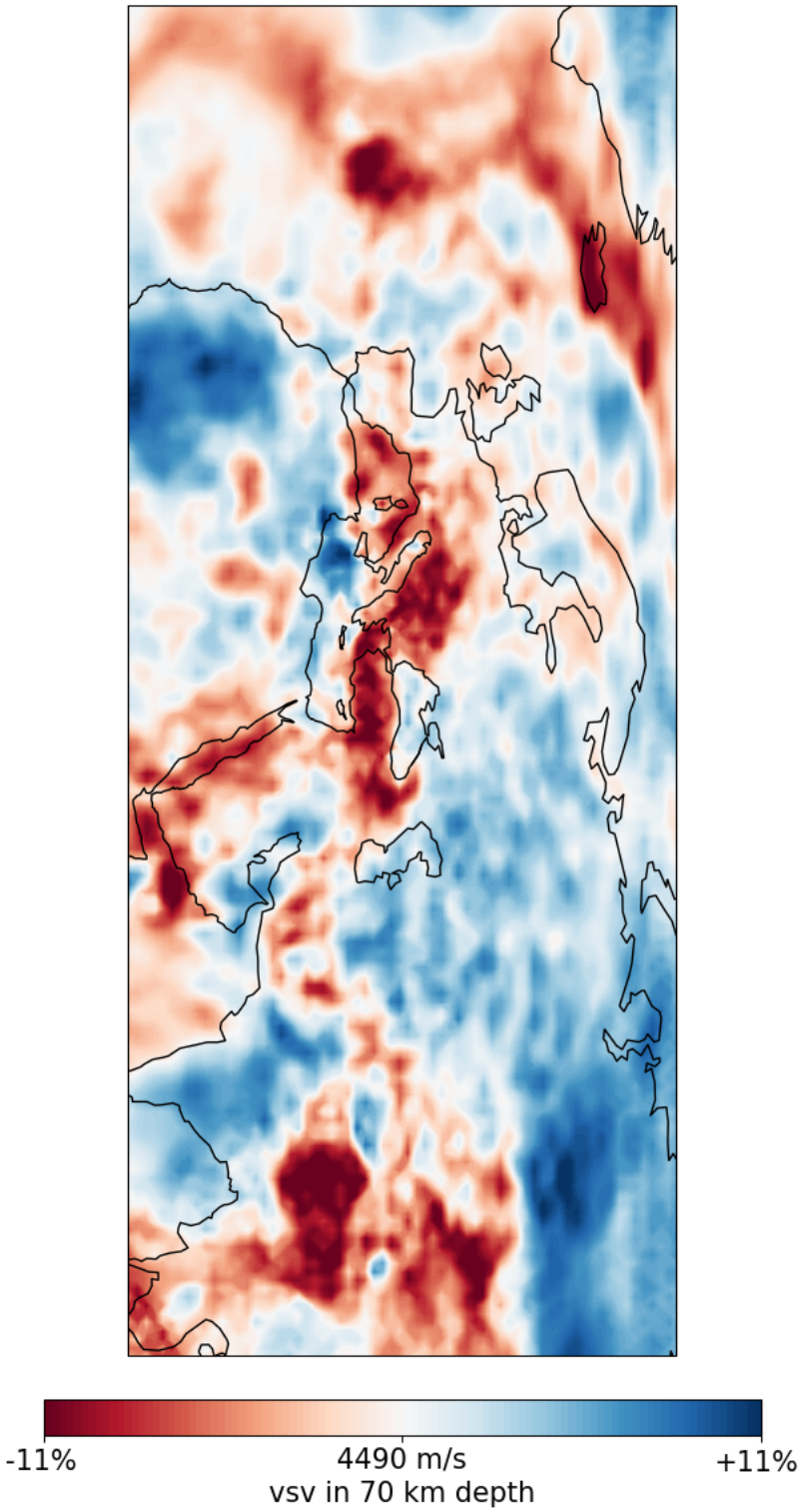

Figure 7-9: Vertical shear velocity model at 70km depth.

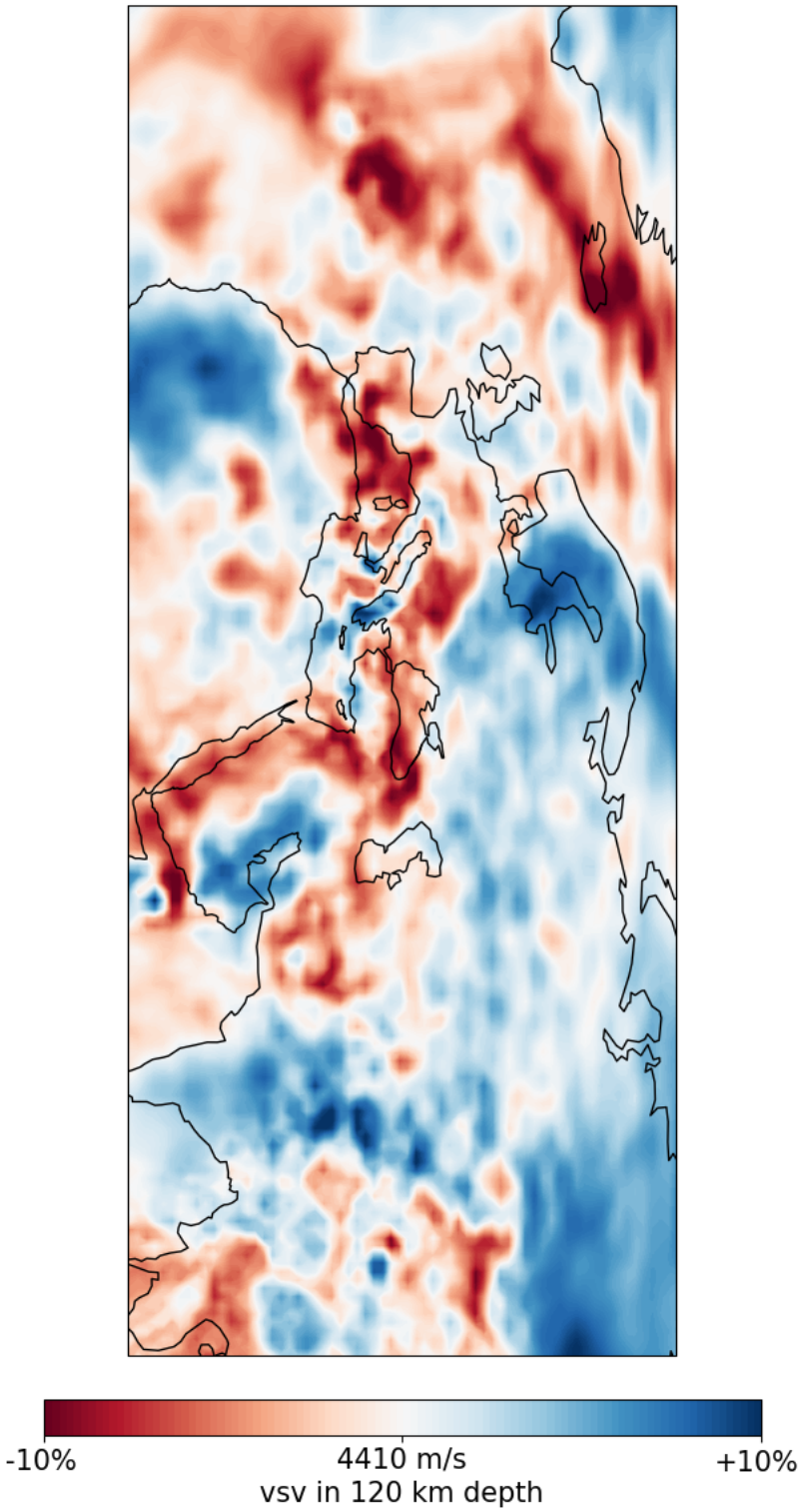

Figure 7-10: Vertical shear velocity model at 120km depth.

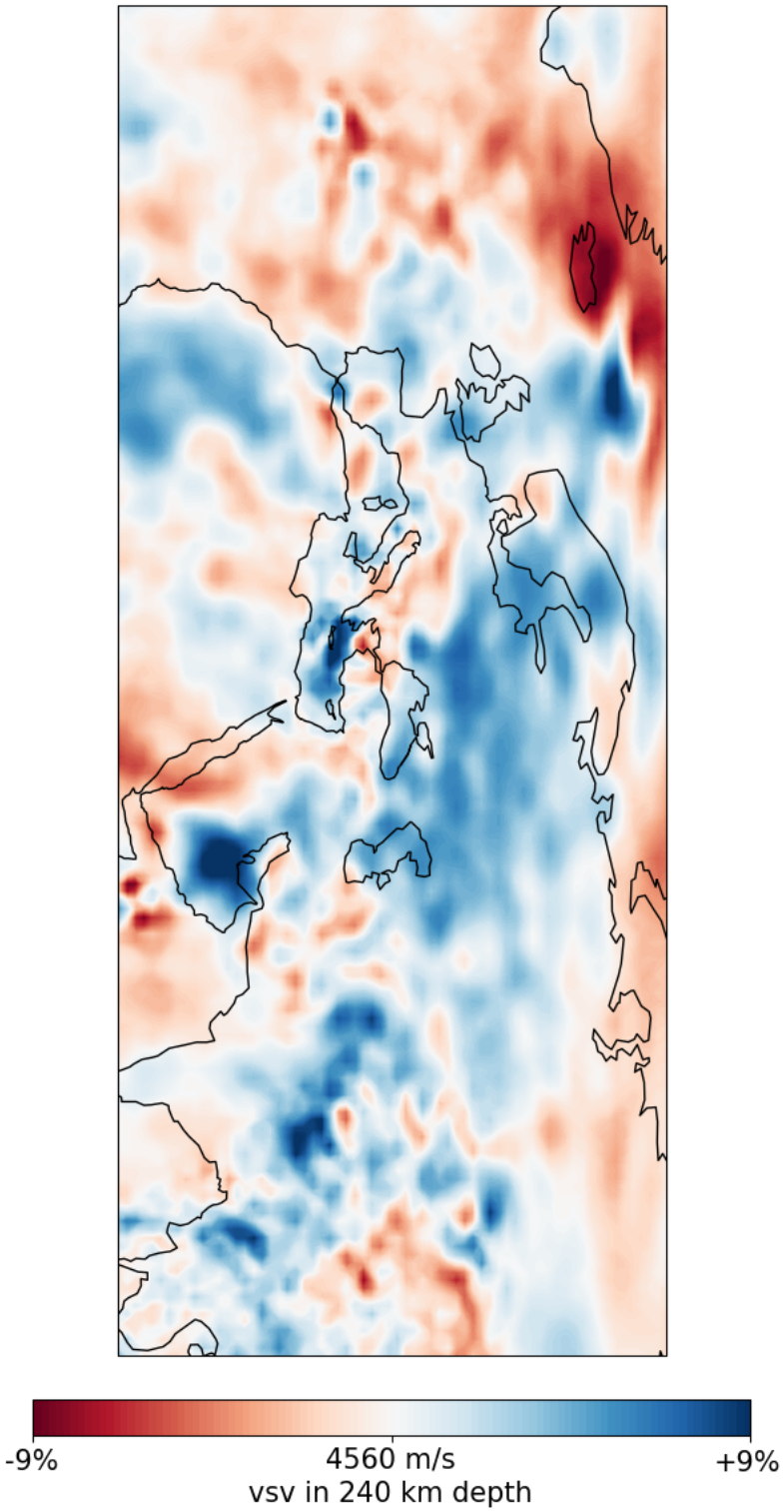

Figure 7-11: Vertical shear velocity model at 240km depth.

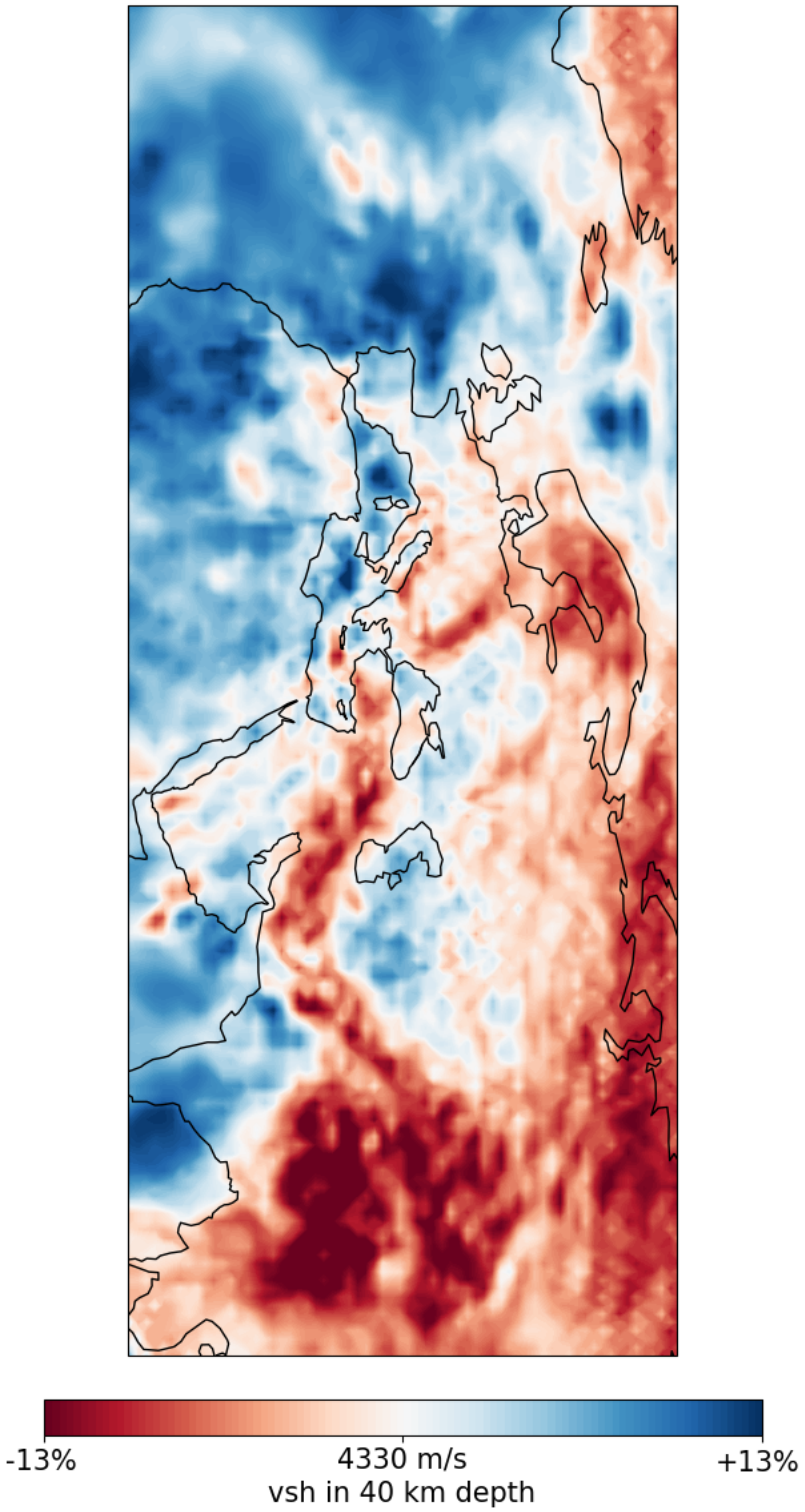

Figure 7-12: Horizontal shear velocity model at 40km depth. Perturbation of the model data are plotted relative to the lateral mean. The percentage of relative change, that can show the perturbations clearly, is chosen

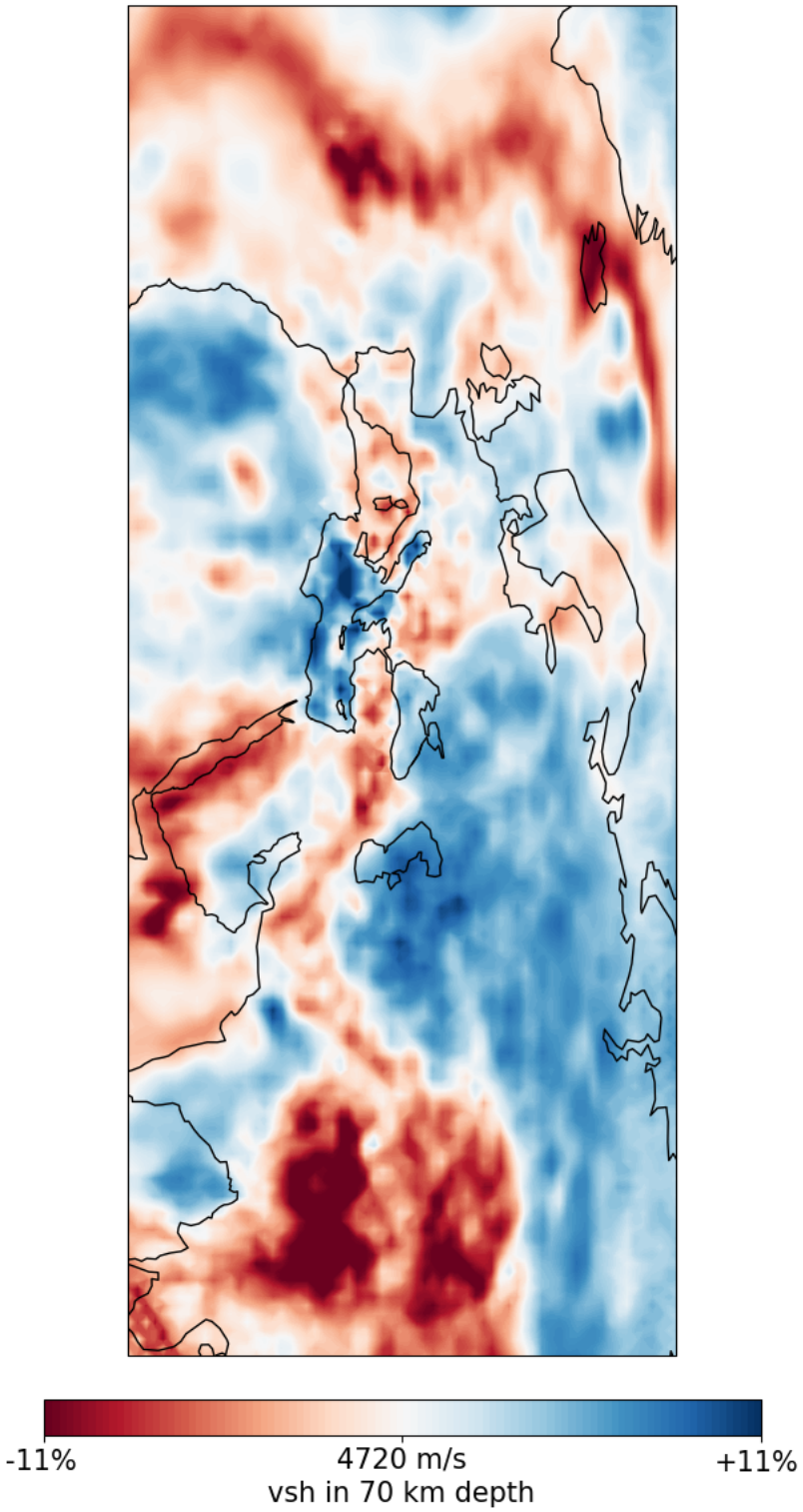

Figure 7-13: Horizontal shear velocity model at 70km depth.

shows that prior information is mostly preserved, while more information has been added by dense raypaths in the covered areas. For example, at depths 40km - 120km an increase in information through more low period data content can be observed. At 40 km, all major mountain ranges in the study area have been preserved and expanded and are more resolved (Figure 7-14). This can additionally be observed through the widening of the velocity data range from 8% to 10%. Northern Europe in particular increased in information content. The thermal processes around Iceland are also more highly resolved. One concern is related to areas outside of apparent raypaths. Some structures have shrunk or vanished. The question arises, if this is true or that data has been averaged away through lack of coverage. At 120 km, the preservation and expansion of information is also visible (Figure 7-15). What is also clearly visible, is the shrinking of structures outside of the area of data coverage e.g. the Himalayas. Another discrepancy seems to be highly resolved point anomalies of opposite polarity close to each other in some areas.
Overall the current model has expanded on the base model and preserved key structures. Supporting that the inversion has added content in the intended mathematical description of the FWI problem. More comparison figures are available in the appendix e.g. of the mantle transition zone (Helffrich and Wood (2001)).

7-5 Discrepencies and Potential Causes

The aforementioned discrepancy concerning point anomalies of opposite polarity are source imprints. Some events at the edge of the domains have a much higher misfit compared to neighbouring events. Normally an area around the source is set to 0 in the gradients. Some events have large moment tensors such that if the area around the event is not sufficient, the influence from the source on the gradient is carried on in the summed gradient of the iteration. After multiple iterations, the source influence stacks, creating point anomalies. In Figure 7-15, the point anomalies can be seen in the Gulf of Aden near the Arabian peninsula, in the Xijiang Province, China and in the Atlatic Ocean. The occurence of point anomalies suggests that point anomalies have been occuring for all events in the dataset. This is not an issue for the overall events and areas inside the area of coverage. The summed gradients of an iteration show the average coverage of Fréchet Kernels in the study area (Figure 7-16). The area of coverage is overlain by the Kernels and the model is most sensitive to these areas. After the gradients are summed, they are smoothed with a diffusion equation with the aforementioned smoothings lengths (Figure 7-17). Smoothing should remove spikes in data likely caused by e.g. source imprints. The smoothing is dependent on the neighbouring values. It would explain why source imprints are clearly visibile at the edges of the dataset and not in the area of coverage. Furthermore, not all sources at the edge have source imprints. This would suggest that not all events have source imprints. The solution to these imprints is also the solution to apparent shrinking of structures outside of the area of raypath coverage. First, remove the events causing source imprints by inspecting individual event misfits and determining anomalous events with relatively high misfits. Focusing on the area of coverage, taper the area around the anomalous events and lowering high polarised values. Lastly, in the eventual inclusion of the final model in the CSEM, the models located around the area of coverage will take precedent.

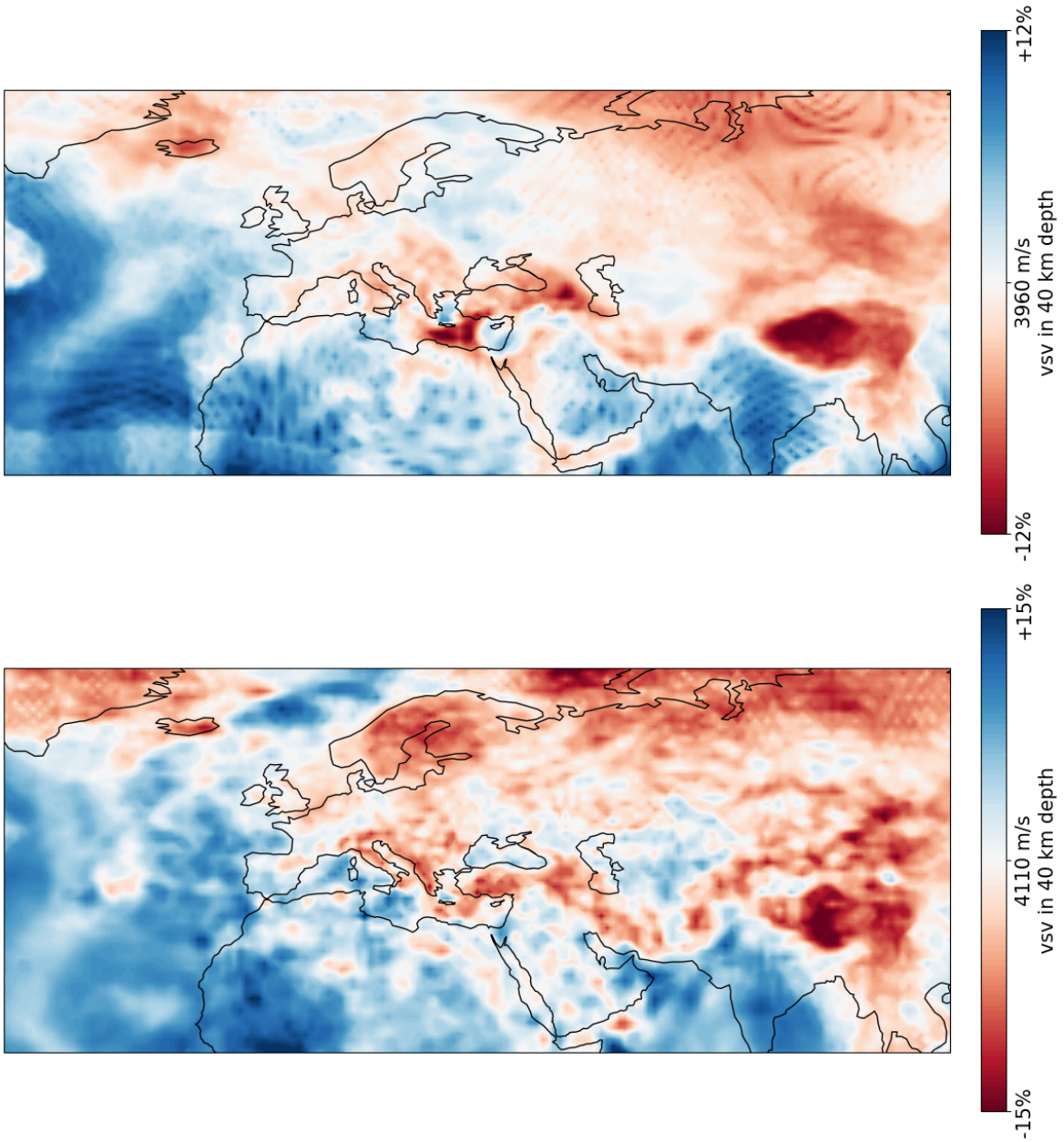

Figure 7-14: Comparison between CSEM base model (upper figure) and the current model (lower figure) in vertical shear velocity at 40 km depth. All apparent structures along the raypaths have been preserved and expanded on.

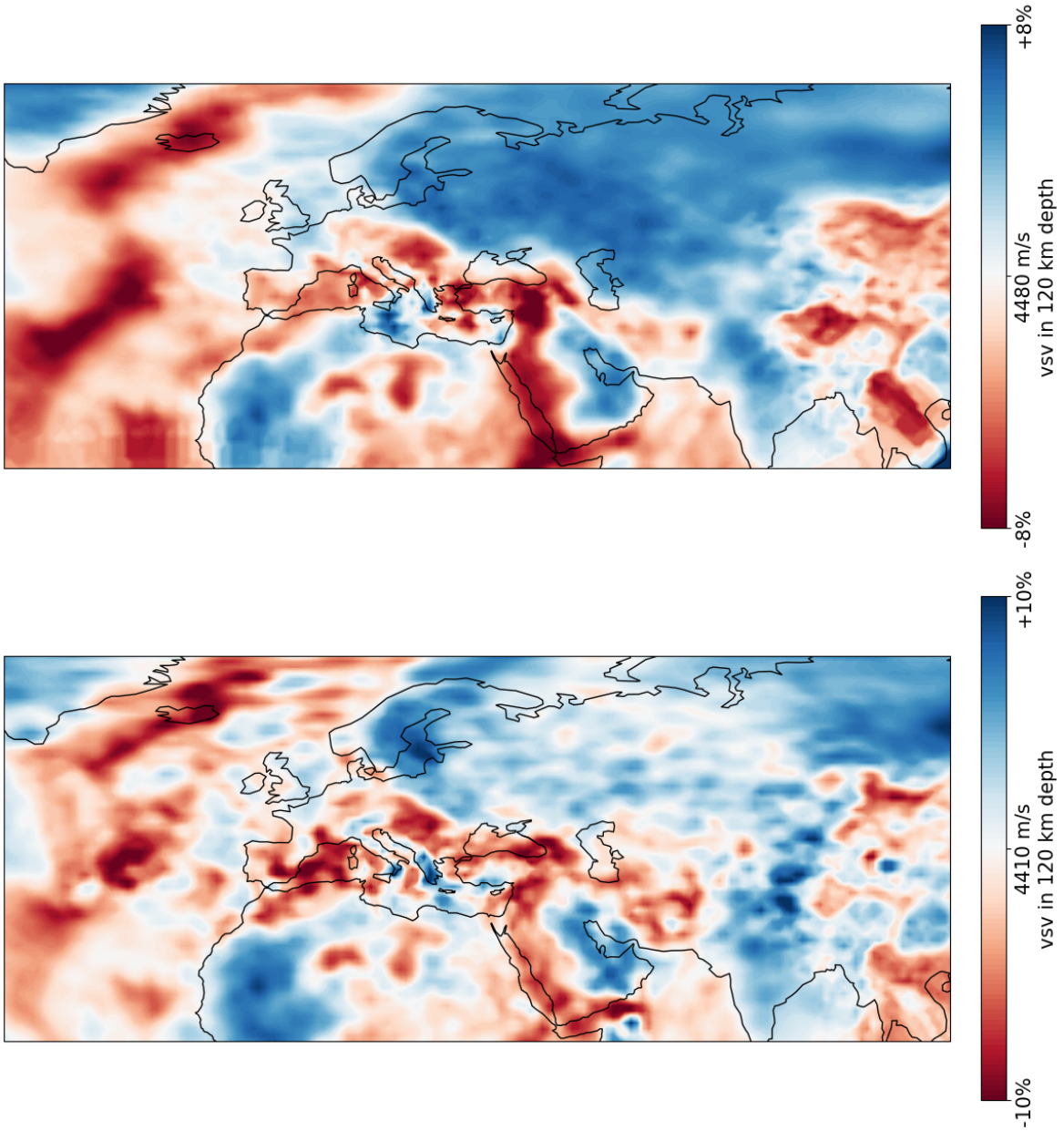

Figure 7-15: Comparison between CSEM base model (upper figure) and the current model (lower figure) in vertical shear velocity at 120 km depth. Main depth of investigation at the current minimum period shows a complete expansion of all known structures from the base model inside data coverage.

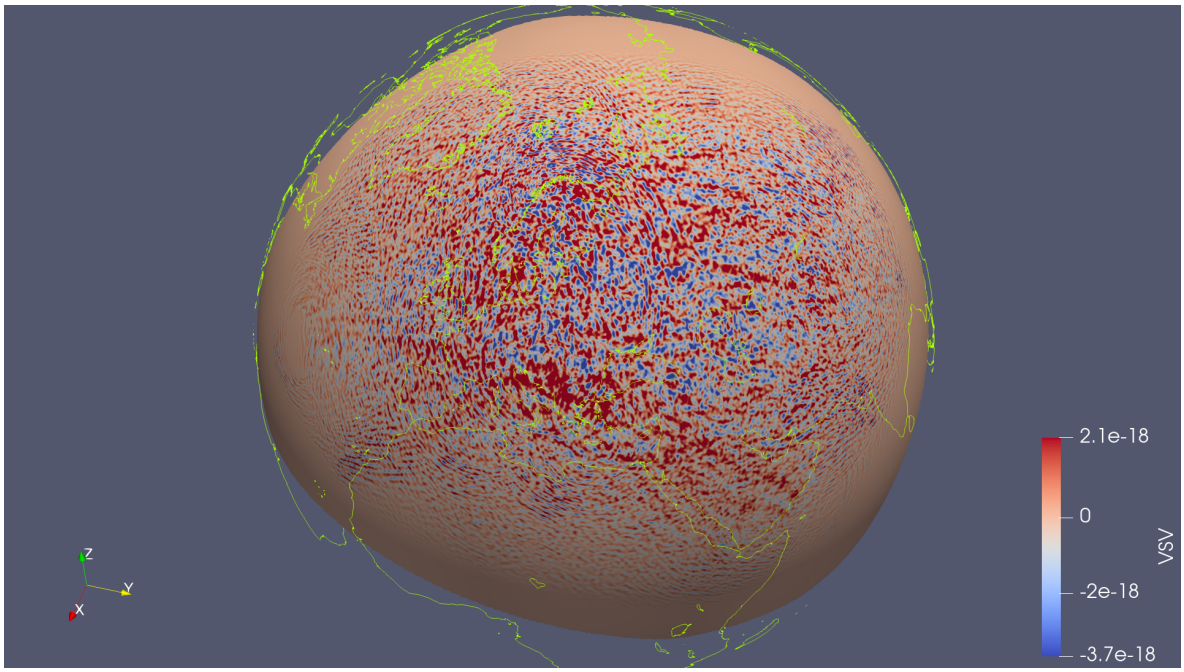

Figure 7-16: Summed Gradient slice at 100 km depth on the EUWA mesh of the current iteration. Gradients of a 70 event batch have been summed up, resulting in an overall gradient sensitive to velocity parameters. Smoothing is applied to solve for more coherent structure.

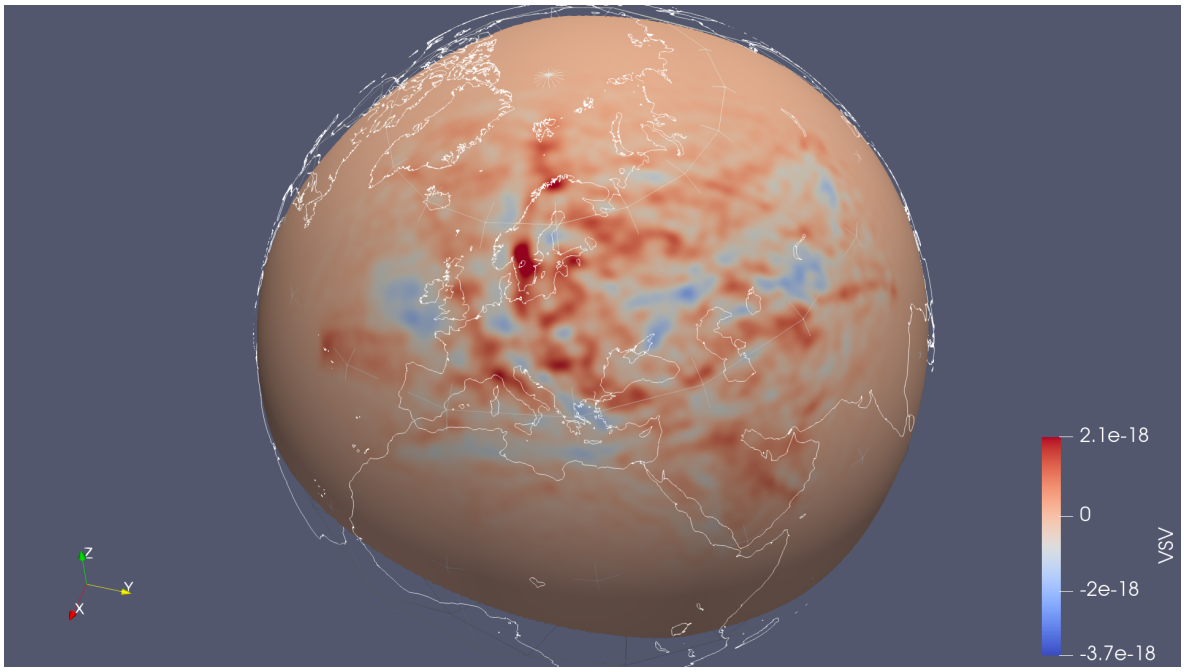

Figure 7-17: Smoothed Gradient showing coherent structures at 100 km depth.

Chapter 8

Conclusion and Outlook

The following goals have been attained, laying the foundation of more cost-effective FWI for regional continental-scale models: (1) Implementing SmoothieSEM meshes that are optimally adapted to Events inside a bounded domain in a FWI-workflow. (2) Showcase the factor of difference in computing costs and achieving cheaper FWI iterations. (3) Resulting models are preserving and adding information starting from a base model. (4) The possibility to include higher frequent data content inside waveform tomographies and (5) using more information given by surface waves and attenuation. Cost-effective FWI also enables studies on data quality for waveform inversion itself or uncertainty analysis. For the outlook, the model needs to be inverted further into lower periods. Close attention should be given to the fitting ability of the available data content at lower periods as well as the added information in the shallow crust. The prototype of the MultiMesh method will be properly implemented into Inversionson on a production level. Making the method available for use in other study areas around the world. Lastly, uncertainty analysis should be done for the eventual final model of the study area. This may include proper restitution testing, checkerboard tests and point spread functions. Further information from geological or geochemical information should also be used for cross-correlation. Proper testing and uncertainty analysis is needed to further support the validity of the final model. Overall, the model in its current state has and will have more information than previous models of the study area.

Bibliography

Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D. A., Knepley, M. G., and Fichtner, A. (2019). Modular and flexible spectral-element waveform modelling in two and three dimensions. *Geophysical Journal International*, 216(3):1675–1692.

Afanasiev, M., Peter, D., Sager, K., Simutė, S., Ermert, L., Krischer, L., and Fichtner, A. (2016). Foundations for a multiscale collaborative earth model. *Geophysical Journal International*, 204(1):39–58.

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

Artemieva, I. M. (2003). Lithospheric structure, composition, and thermal regime of the east european craton: implications for the subsidence of the russian platform. *Earth and Planetary Science Letters*, 213(3-4):431–446.

Backus, G. E. (1964). Geographical interpretation of measurements of average phase velocities of surface waves over great circular and great semi-circular paths. *Bulletin of the Seismological Society of America*, 54(2):571–610.

Backus, G. E. and Gilbert, J. F. (1967). Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal International*, 13(1-3):247–276.

Bassin, C. (2000). The current limits of resolution for surface wave tomography in north america. *Eos Trans. AGU*, 81:F897.

Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., and Wedi, N. P. (2021). The digital revolution of earth-system science. *Nature Computational Science*, 1(2):104–113.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. (2010). Obspy: A python toolbox for seismology. *Seismological Research Letters*, 81(3):530–533.

Blom, N., Gokhberg, A., and Fichtner, A. (2020). Seismic waveform tomography of the central and eastern mediterranean upper mantle. *Solid Earth*, 11(2):669–690.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.

Brown, D., Juhlin, C., Ayala, C., Tryggvason, A., Bea, F., Alvarez-Marrón, J., Carbonell, R., Seward, D., Glasmacher, U., Puchkov, V., et al. (2008). Mountain building processes during continent–continent collision in the uralides. *Earth-Science Reviews*, 89(3-4):177–195.

Bunge, H.-P., Hagelberg, C., and Travis, B. (2003). Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography. *Geophysical Journal International*, 152(2):280–301.

Burbank, D. W., Leland, J., Fielding, E., Anderson, R. S., Brozovic, N., Reid, M. R., and Duncan, C. (1996). Bedrock incision, rock uplift and threshold hillslopes in the northwestern himalayas. *Nature*, 379(6565):505–510.

Capdeville, Y., Guillot, L., and Marigo, J.-J. (2010). 2-d non-periodic homogenization to upscale elastic media for p–sv waves. *Geophysical Journal International*, 182(2):903–922.

Chang, S.-J., Van Der Lee, S., Flanagan, M. P., Bedle, H., Marone, F., Matzel, E. M., Pasyanos, M. E., Rodgers, A. J., Romanowicz, B., and Schmid, C. (2010). Joint inversion for three-dimensional s velocity mantle structure along the tethyan margin. *Journal of Geophysical Research: Solid Earth*, 115(B8).

Clayton, R. and Engquist, B. (1977). Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the seismological society of America*, 67(6):1529–1540.

Copley, A., Avouac, J.-P., and Royer, J.-Y. (2010). India-asia collision and the cenozoic slowdown of the indian plate: Implications for the forces driving plate motions. *Journal of Geophysical Research: Solid Earth*, 115(B3).

Dahlen, F., Hung, S.-H., and Nolet, G. (2000). Fréchet kernels for finite-frequency traveltimes—i. theory. *Geophysical Journal International*, 141(1):157–174.

De Graciansky, P.-C., Roberts, D. G., and Tricart, P. (2011). The birth of the western and central alps: Subduction, obduction, collision. In *Developments in Earth surface processes*, volume 14, pages 289–315. Elsevier.

Dèzes, P., Schmid, S., and Ziegler, P. (2004). Evolution of the european cenozoic rift system: interaction of the alpine and pyrenean orogens with their foreland lithosphere. *Tectonophysics*, 389(1-2):1–33.

Dilek, Y. (2006). Collision tectonics of the mediterranean region: causes and consequences.

Duman, T. Y. and Emre, Ö. (2013). The east anatolian fault: geometry, segmentation and jog characteristics. *Geological Society, London, Special Publications*, 372(1):495–529.

Dziewonski, A. M. and Anderson, D. L. (1981). Preliminary reference earth model. *Physics of the earth and planetary interiors*, 25(4):297–356.

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

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

Errico, R. M. (1997). What is an adjoint model? *Bulletin of the American Meteorological Society*, 78(11):2577–2592.

Fichtner, A. (2010). *Full seismic waveform modelling and inversion*. Springer Science & Business Media.

Fichtner, A., Bunge, H.-P., and Igel, H. (2006). The adjoint method in seismology: I. theory. *Physics of the Earth and Planetary Interiors*, 157(1-2):86–104.

Fichtner, A., Igel, H., Bunge, H.-P., and Kennett, B. L. (2009a). Simulation and inversion of seismic wave propagation on continental scales based on a spectral-element method. *Journal of Numerical Analysis, Industrial and Applied Mathematics*, 4(1-2):11–22.

Fichtner, A., Kennett, B. L., Igel, H., and Bunge, H.-P. (2008). Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain. *Geophysical Journal International*, 175(2):665–685.

Fichtner, A., Kennett, B. L., Igel, H., and Bunge, H.-P. (2009b). Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725.

Fichtner, A., Kennett, B. L., and Trampert, J. (2013a). Separating intrinsic and apparent anisotropy. *Physics of the Earth and Planetary Interiors*, 219:11–20.

Fichtner, A. and Trampert, J. (2011). Resolution analysis in full waveform inversion. *Geophysical Journal International*, 187(3):1604–1624.

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., and Villasenor, A. (2013b). Multiscale full waveform inversion. *Geophysical Journal International*, 194(1):534–556.

Fichtner, A., van Herwaarden, D.-P., Afanasiev, M., Simutė, S., Krischer, L., Çubuk-Sabuncu, Y., Taymaz, T., Colli, L., Saygin, E., Villaseñor, A., et al. (2018). The collaborative seismic earth model: generation 1. *Geophysical research letters*, 45(9):4007–4016.

Fichtner, A. and Villaseñor, A. (2015). Crust and upper mantle of the western mediterranean–constraints from full-waveform inversion. *Earth and Planetary Science Letters*, 428:52–62.

Gatinsky, Y., Rundquist, D., Vladova, G., Prokhorova, T., et al. (2011). Up-to-date geodynamics and seismicity of central asia. *International Journal of Geosciences*, 2(01):1.

Goes, S., Spakman, W., and Bijwaard, H. (1999). A lower mantle source for central european volcanism. *Science*, 286(5446):1928–1931.

Grand, S. P., Van der Hilst, R. D., and Widiyantoro, S. (1997). High resolution global tomography: a snapshot of convection in the earth. *Geological Society of America Today*, 7(4).

Hapla, V., Knepley, M. G., Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., and Fichtner, A. (2021). Fully parallel mesh i/o using petsc dmplex with an application to waveform modeling. *SIAM Journal on Scientific Computing*, 43(2):C127–C153.

Helffrich, G. R. and Wood, B. J. (2001). The earth's mantle. *Nature*, 412(6846):501–507.

Hetényi, G., Molinari, I., Clinton, J., Bokelmann, G., Bondár, I., Crawford, W. C., Dessa, J.-X., Doubre, C., Friederich, W., Fuchs, F., et al. (2018). The alparray seismic network: a large-scale european experiment to image the alpine orogen. *Surveys in geophysics*, 39:1009–1033.

Igel, H. (2017). *Computational seismology: a practical introduction*. Oxford University Press.

Institute Earth Sciences "Jaume Almera" CSIC (ICTJA Spain) (2007). Iberarray.

Kárason, H. and Van Der Hilst, R. D. (2000). Constraints on mantle convection from seismic tomography. *Geophysical monograph*, 121:277–288.

Kennett, B. and Engdahl, E. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105(2):429–465.

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

Kennett, B. L. and Fichtner, A. (2020). *Exploiting seismic waveforms: Correlation, heterogeneity and inversion*. Cambridge University Press.

Kilibarda, Z. and Schassburger, A. (2018). A diverse deep-sea trace fossil assemblage from the adriatic flysch formation (middle eocene–middle miocene), montenegro (central mediterranean). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 506:112–127.

Komatitsch, D., Coutel, F., and Mora, P. (1996). Tensorial formulation of the wave equation for modelling curved interfaces. *Geophysical Journal International*, 127(1):156–168.

Komatitsch, D. and Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical journal international*, 139(3):806–822.

Komatitsch, D. and Tromp, J. (2002a). Spectral-element simulations of global seismic wave propagation—i. validation. *Geophysical Journal International*, 149(2):390–412.

Komatitsch, D. and Tromp, J. (2002b). Spectral-element simulations of global seismic wave propagation—ii. three-dimensional models, oceans, rotation and self-gravitation. *Geophysical Journal International*, 150(1):303–318.

Kosloff, R. and Kosloff, D. (1986). Absorbing boundaries for wave propagation problems. *Journal of Computational Physics*, 63(2):363–376.

Krischer, L., Fichtner, A., Zukauskaite, S., and Igel, H. (2015a). Large-scale seismic inversion framework. *Seismological Research Letters*, 86(4):1198–1207.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., and Wassermann, J. (2015b). Obspy: A bridge for seismology into the scientific python ecosystem. *Computational Science & Discovery*, 8(1):014003.

Lévěque, J.-J., Rivera, L., and Wittlinger, G. (1993). On the use of the checker-board test to assess the resolution of tomographic inversions. *Geophysical Journal International*, 115(1):313–318.

Liu, Q. and Gu, Y. (2012). Seismic imaging: From classical to adjoint tomography. *Tectonophysics*, 566:31–66.

Liu, Q. and Tromp, J. (2008). Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods. *Geophysical Journal International*, 174(1):265–286.

Liu, Z. and Bird, P. (2008). Kinematic modelling of neotectonics in the persia-tibet-burma orgen. *Geophysical Journal International*, 172(2):779–797.

Lund, B., Schmidt, P., Shomali, Z. H., and Roth, M. (2021). The modern swedish national seismic network: two decades of intraplate microseismic observation. *Seismological Research Letters*, 92(3):1747–1758.

Luo, Y. and Schuster, G. T. (1991). Wave-equation traveltime inversion. *Geophysics*, 56(5):645–653.

Ma, J., Bunge, H.-P., Thrastarson, S., Fichtner, A., Herwaarden, D.-P. v., Tian, Y., Chang, S.-J., and Liu, T. (2022). Seismic full-waveform inversion of the crust-mantle structure beneath china and adjacent regions. *Journal of Geophysical Research: Solid Earth*, 127(9).

Mazur, S., Krzywiec, P., Malinowski, M., Lewandowski, M., Aleksandrowski, P., and Mikołajczak, M. (2018). On the nature of the teisseyre-tornquist zone. *Geology, Geophysics and Environment*, 44(1):17–30.

McCann, T. (2008). *The geology of central Europe: Mesozoic and Cenozoic*. Geological Society of London.

Megies, T., Beyreuther, M., Barsch, R., Krischer, L., and Wassermann, J. (2011). Obspy–what can it do for data centers and observatories? *Annals of Geophysics*, 54(1):47–58.

Menke, W. (2018). *Geophysical data analysis: Discrete inverse theory*. Academic press.

Moczo, P., Kristek, J., and Gális, M. (2014). *The finite-difference modelling of earthquake motions: Waves and ruptures*. Cambridge University Press.

Nataf, H.-C., Nakanishi, I., and Anderson, D. L. (1984). Anisotropy and shear-velocity heterogeneities in the upper mantle. *Geophysical Research Letters*, 11(2):109–112.

Network, A. S. (2015). Alparray seismic network (aasn) temporary component. *AlpArray Working Group, Other/Seismic Network, doi*, 10.

Nissen-Meyer, T., Fournier, A., and Dahlen, F. A. (2007). A two-dimensional spectral-element method for computing spherical-earth seismograms–i. moment-tensor source. *Geophysical Journal International*, 168(3):1067–1092.

Nissen-Meyer, T., van Driel, M., Stähler, S. C., Hosseini, K., Hempel, S., Auer, L., Colombi, A., and Fournier, A. (2014a). Axisem: broadband 3-d seismic wavefields in axisymmetric media. *Solid Earth*, 5(1):425–445.

Nissen-Meyer, T., van Driel, M., Stähler, S. C., Hosseini, K., Hempel, S., Auer, L., Colombi, A., and Fournier, A. (2014b). Axisem: broadband 3-d seismic wavefields in axisymmetric media. *Solid Earth*, 5(1):425–445.

Noe, S., van Herwaarden, D.-P., Thrastarson, S., and Fichtner, A. (2023). The collaborative seismic earth model: Generation 2. Technical report, Copernicus Meetings.

Nolet, G. (1990). Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs. *Journal of Geophysical Research: Solid Earth*, 95(B6):8499–8512.

Ollier, C. and Pain, C. (2004). *The origin of mountains*. Routledge.

Panning, M. and Romanowicz, B. (2006). A three-dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophysical Journal International*, 167(1):361–379.

Peterson, J. (1993). Observations and modeling of seismic background noise. *USGS Open File Report*, 93-322:94 pp.

Rickers, F., Fichtner, A., and Trampert, J. (2013). The iceland–jan mayen plume system and its impact on mantle dynamics in the north atlantic region: evidence from full-waveform inversion. *Earth and Planetary Science Letters*, 367:39–51.

Ritsema, J. and Lekić, V. (2020). Heterogeneity of seismic wave velocity in earth's mantle. *Annual Review of Earth and Planetary Sciences*, 48:377–401.

Ritsema, J. and Van Heijst, H.-J. (2000). Seismic imaging of structural heterogeneity in earth's mantle: evidence for large-scale mantle flow. *Science Progress (1933-)*, pages 243–259.

Romanowicz, B. (2008). Using seismic waves to image earth's internal structure. *Nature*, 451(7176):266–268.

Romanowicz, B. A. (2021). Seismic tomography of the earth's mantle. In Alderton, D. and Elias, S. A., editors, *Encyclopedia of Geology (Second Edition)*, pages 587–609. Academic Press, Oxford, second edition edition.

Ruan, Y., Lei, W., Modrak, R., Örsvuran, R., Bozdağ, E., and Tromp, J. (2019). Balancing unevenly distributed data in seismic tomography: a global adjoint tomography example. *Geophysical Journal International*, 219(2):1225–1236.

Schmid, S. M., Bernoulli, D., Fügenschuh, B., Matenco, L., Schefer, S., Schuster, R., Tischler, M., and Ustaszewski, K. (2008). The alpine-carpathian-dinaridic orogenic system: correlation and evolution of tectonic units. *Swiss Journal of Geosciences*, 101:139–183.

Schmid, S. M., Fügenschuh, B., Kissling, E., and Schuster, R. (2004). Tectonic map and overall architecture of the alpine orogen. *Eclogae Geologicae Helvetiae*, 97:93–117.

Şengör, A., Tüysüz, O., Imren, C., Sakınç, M., Eyidoğan, H., Görür, N., Le Pichon, X., and Rangin, C. (2005). The north anatolian fault: A new look. *Annu. Rev. Earth Planet. Sci.*, 33:37–112.

Stern, R. J. and Johnson, P. (2010). Continental lithosphere of the arabian plate: a geologic, petrologic, and geophysical synthesis. *Earth-Science Reviews*, 101(1-2):29–67.

Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2009). Adjoint tomography of the southern california crust. *Science*, 325(5943):988–992.

Tape, C., Liu, Q., and Tromp, J. (2007). Finite-frequency tomography using adjoint methods—methodology and examples using membrane surface waves. *Geophysical Journal International*, 168(3):1105–1129.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. *Scattering and attenuations of seismic waves, part i*, pages 365–399.

ten Brink, U. S. and Taylor, M. H. (2002). Crustal structure of central lake baikal: Insights into intracontinental rifting. *Journal of Geophysical Research: Solid Earth*, 107(B7):ETG–2.

Thrastarson, S., van Driel, M., Krischer, L., Boehm, C., Afanasiev, M., Van Herwaarden, D.-P., and Fichtner, A. (2020). Accelerating numerical wave propagation by wavefield adapted meshes. part ii: full-waveform inversion. *Geophysical Journal International*, 221(3):1591–1604.

Thrastarson, S., van Herwaarden, D.-P., and Fichtner, A. (2021a). Inversionson: fully automated seismic waveform inversions. *EarthArXiv*.

Thrastarson, S., Van Herwaarden, D.-P., Krischer, L., Boehm, C., van Driel, M., Afanasiev, M., and Fichtner, A. (2022). Data-adaptive global full-waveform inversion. *Geophysical Journal International*, 230(2):1374–1393.

Thrastarson, S., van Herwaarden, D.-P., Krischer, L., and Fichtner, A. (2021b). Lasif: Large-scale seismic inversion framework, an updated version. *EarthArXiv*.

Tromp, J. (2020). Seismic wavefield imaging of earth's interior across scales. *Nature Reviews Earth & Environment*, 1(1):40–53.

Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1):195–216.

Van der Hilst, R. D., Widiyantoro, S., and Engdahl, E. (1997). Evidence for deep mantle circulation from global tomography. *Nature*, 386(6625):578–584.

van Driel, M., Boehm, C., Krischer, L., and Afanasiev, M. (2020). Accelerating numerical wave propagation using wavefield adapted meshes. part i: forward and adjoint modelling. *Geophysical Journal International*, 221(3):1580–1590.

van Herwaarden, D. P., Afanasiev, M., Thrastarson, S., and Fichtner, A. (2021). Evolutionary full-waveform inversion. *Geophysical Journal International*, 224(1):306–311.

van Herwaarden, D. P., Boehm, C., Afanasiev, M., Thrastarson, S., Krischer, L., Trampert, J., and Fichtner, A. (2020). Accelerated full-waveform inversion using dynamic mini-batches. *Geophysical Journal International*, 221(2):1427–1438.

van Herwaarden, D.-P. and Hapla, V. (2023). Optson.

van Herwaarden, D.-P., Thrastarson, S., Hapla, V., Afanasiev, M., Trampert, J., and Fichtner, A. (2023). Full-waveform tomography of the african plate using dynamic mini-batches. *Journal of Geophysical Research: Solid Earth*, 128(6):e2022JB026023.

van Hinsbergen, D. J., Steinberger, B., Doubrovine, P. V., and Gassmöller, R. (2011). Acceleration and deceleration of india-asia convergence since the cretaceous: Roles of mantle plumes and continental collision. *Journal of Geophysical Research: Solid Earth*, 116(B6).

Woodhouse, J. (1974). Surface waves in a laterally varying layered structure. *Geophysical Journal International*, 37(3):461–490.

Woodhouse, J. H. and Dziewonski, A. M. (1984). Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms. *Journal of Geophysical Research: Solid Earth*, 89(B7):5953–5986.

Wright, S. J. (2006). Numerical optimization.

Yilmaz, Ö. (2001). *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of exploration geophysicists.

Zhou, X.-Y., Chang, X., Wang, Y.-B., Wen, X.-T., You, J.-C., and Sun, C. (2022). Non-artifact vector p- and s-wave separation for elastic reverse time migration. *Petroleum Science*, 19(6):2695–2710.

Zhu, H., Bozdağ, E., Peter, D., and Tromp, J. (2012). Structure of the european upper mantle revealed by adjoint tomography. *Nature Geoscience*, 5(7):493–498.

Appendix A

The back of the thesis

A-1 Other Model slices

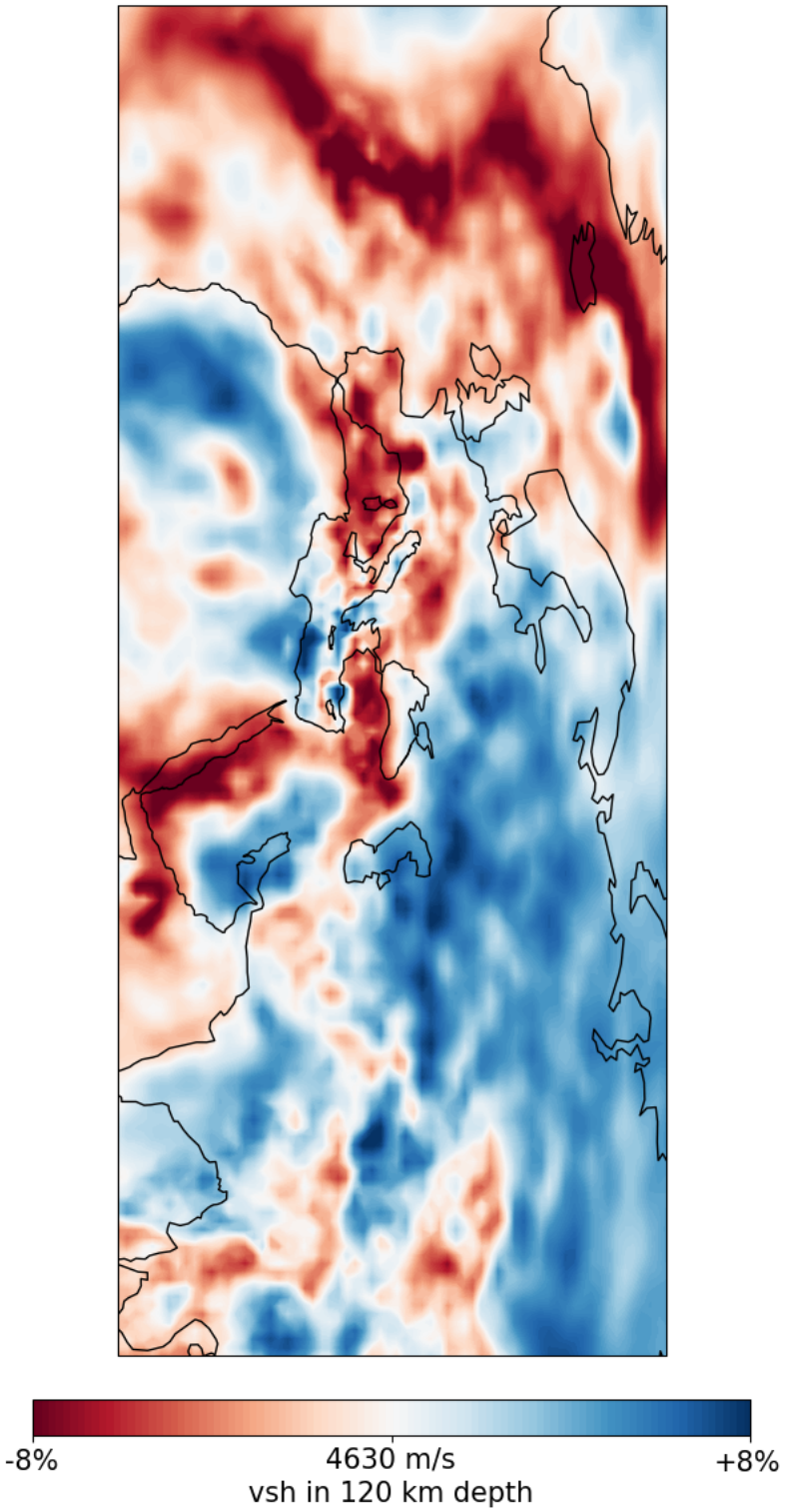

Figure A-1: Horizontal shear velocity model at 120km depth.

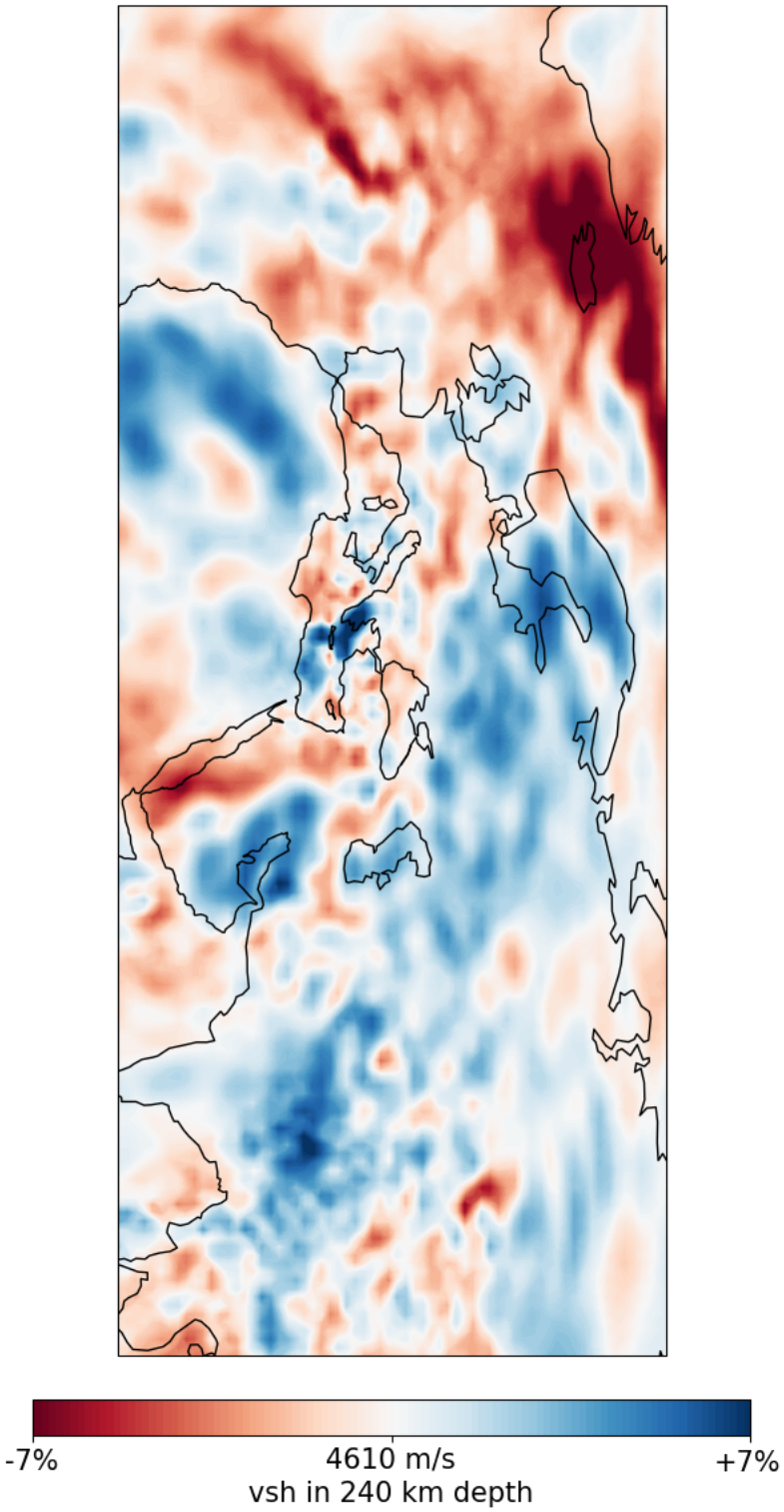

Figure A-2: Horizontal shear velocity model at 240km depth.

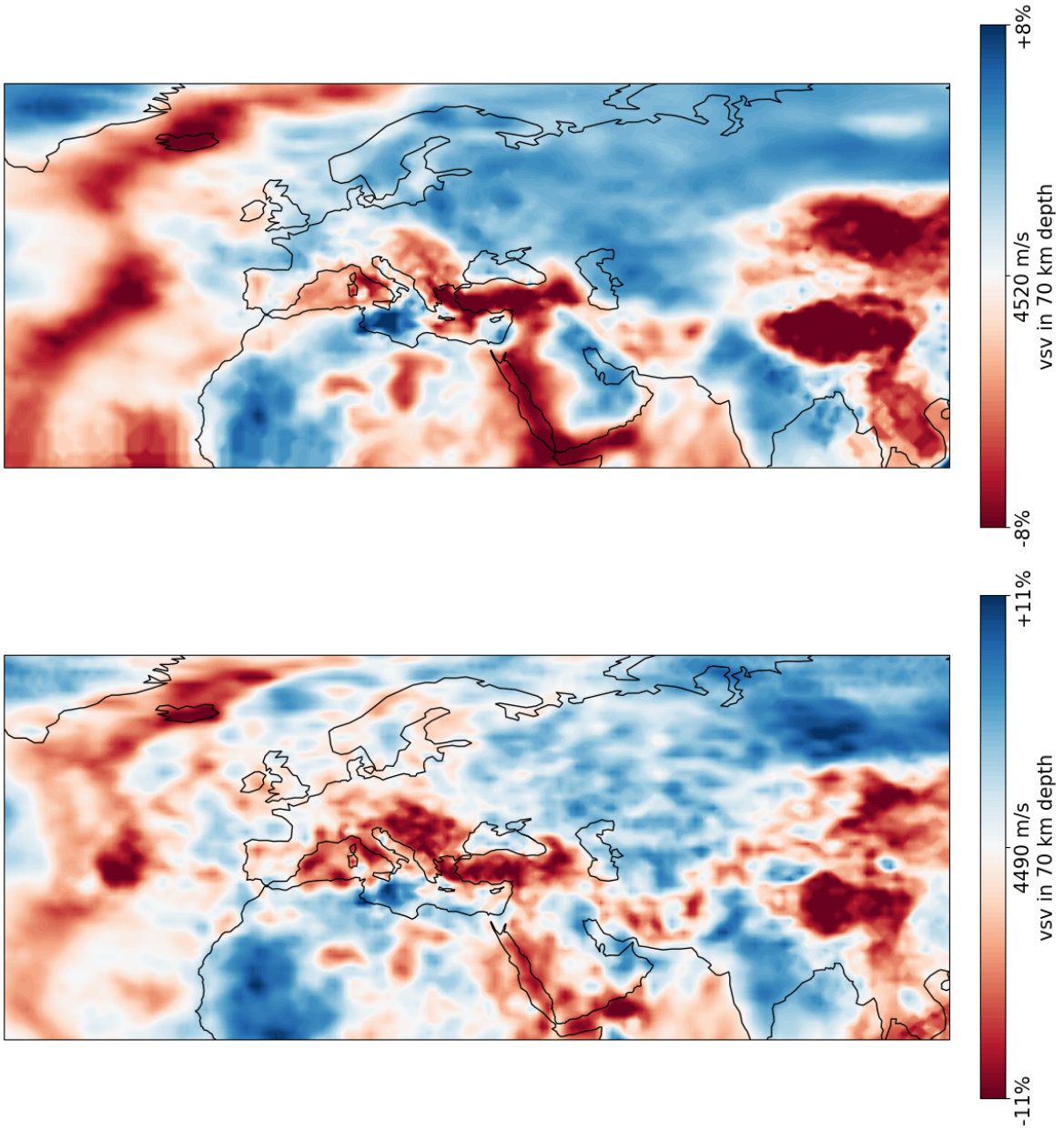

Figure A-3: Comparison between CSEM base model (upper figure) and the current model (lower figure) in vertical shear velocity at 70 km depth.

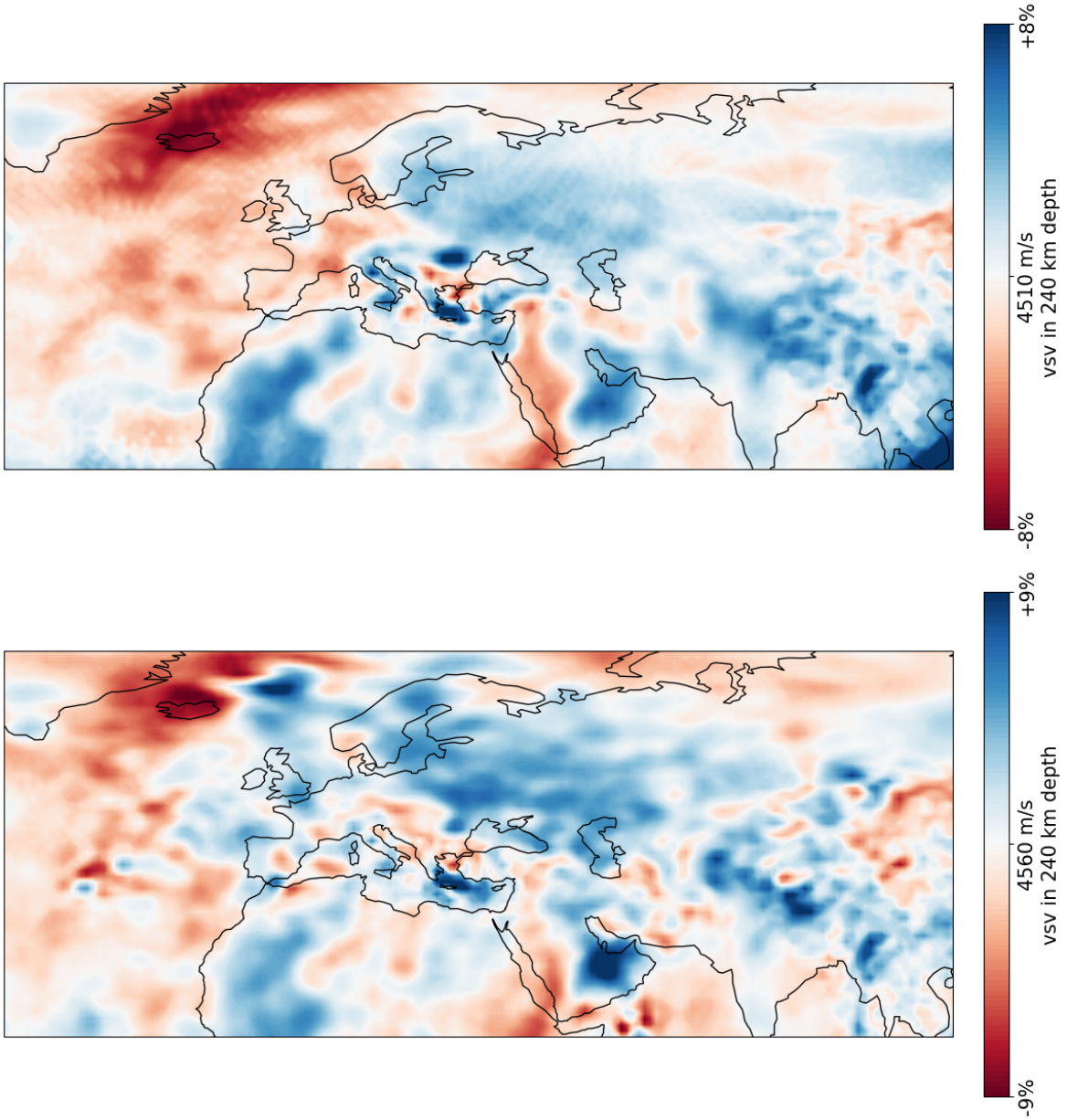

Figure A-4: Comparison between CSEM base model (upper figure) and the current model (lower figure) in vertical shear velocity at 240 km depth.

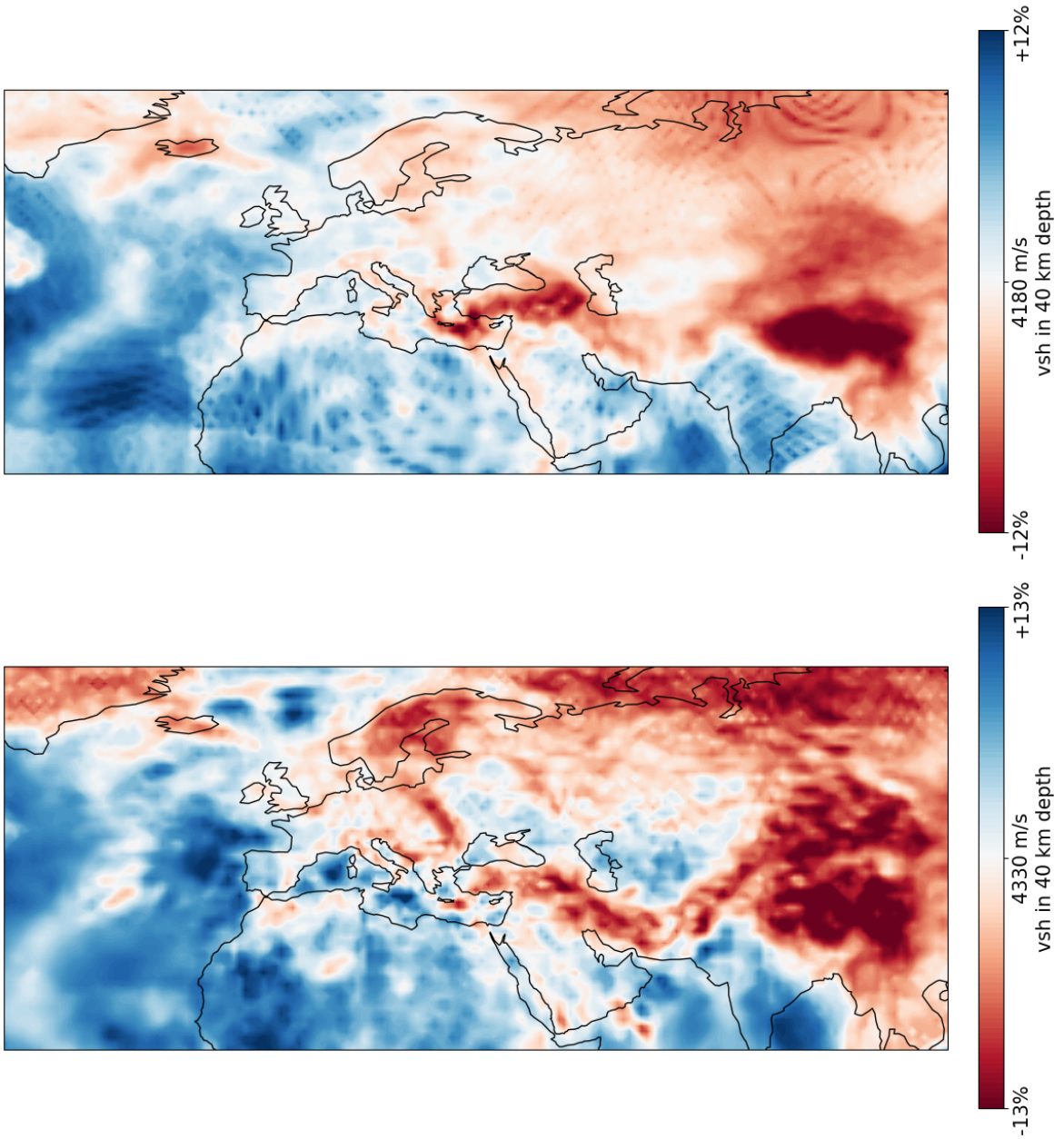

Figure A-5: Comparison between CSEM base model (upper figure) and the current model (lower figure) in horizontal shear velocity at 40 km depth.

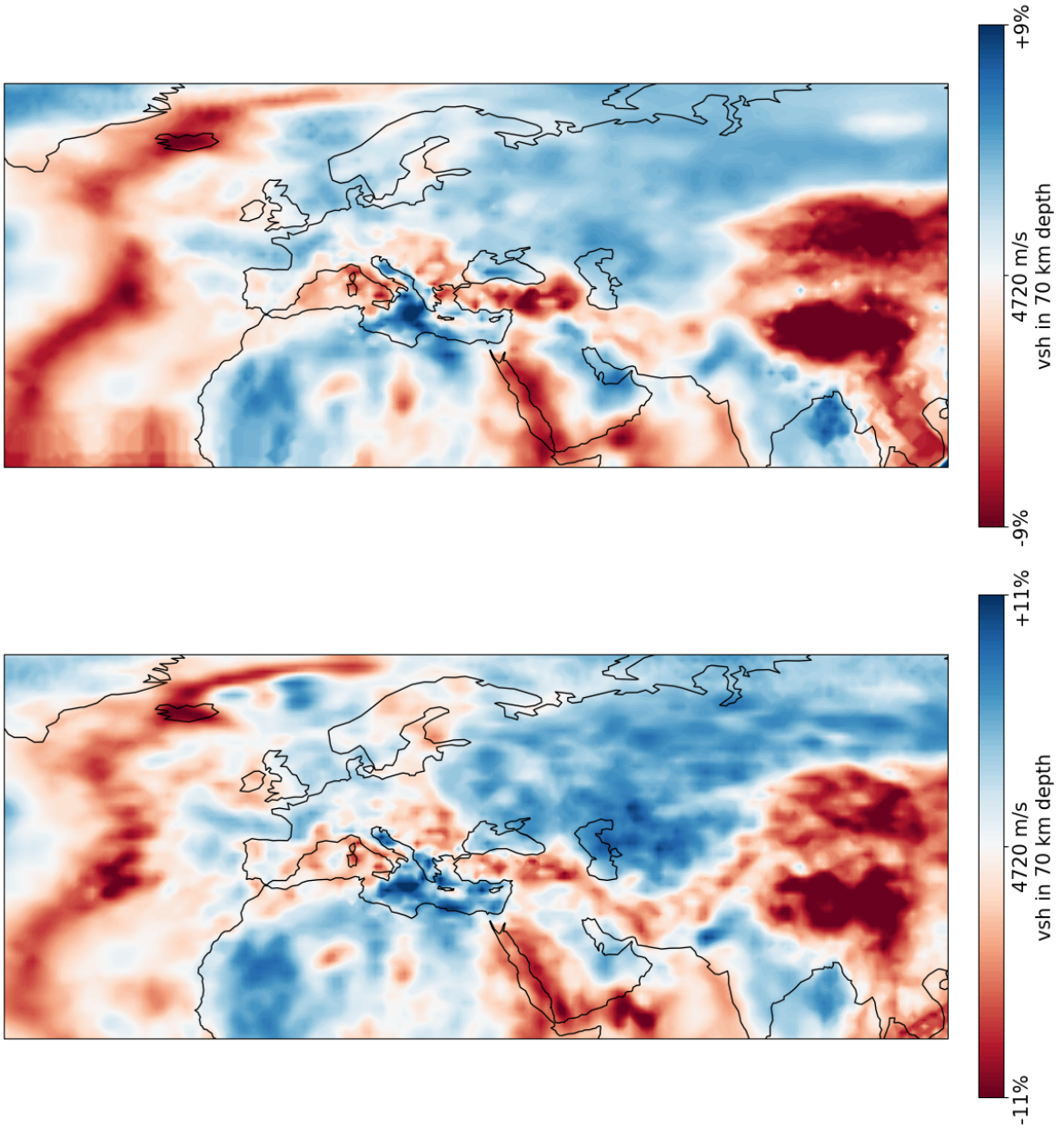

Figure A-6: Comparison between CSEM base model (upper figure) and the current model (lower figure) in horizontal shear velocity at 70 km depth.

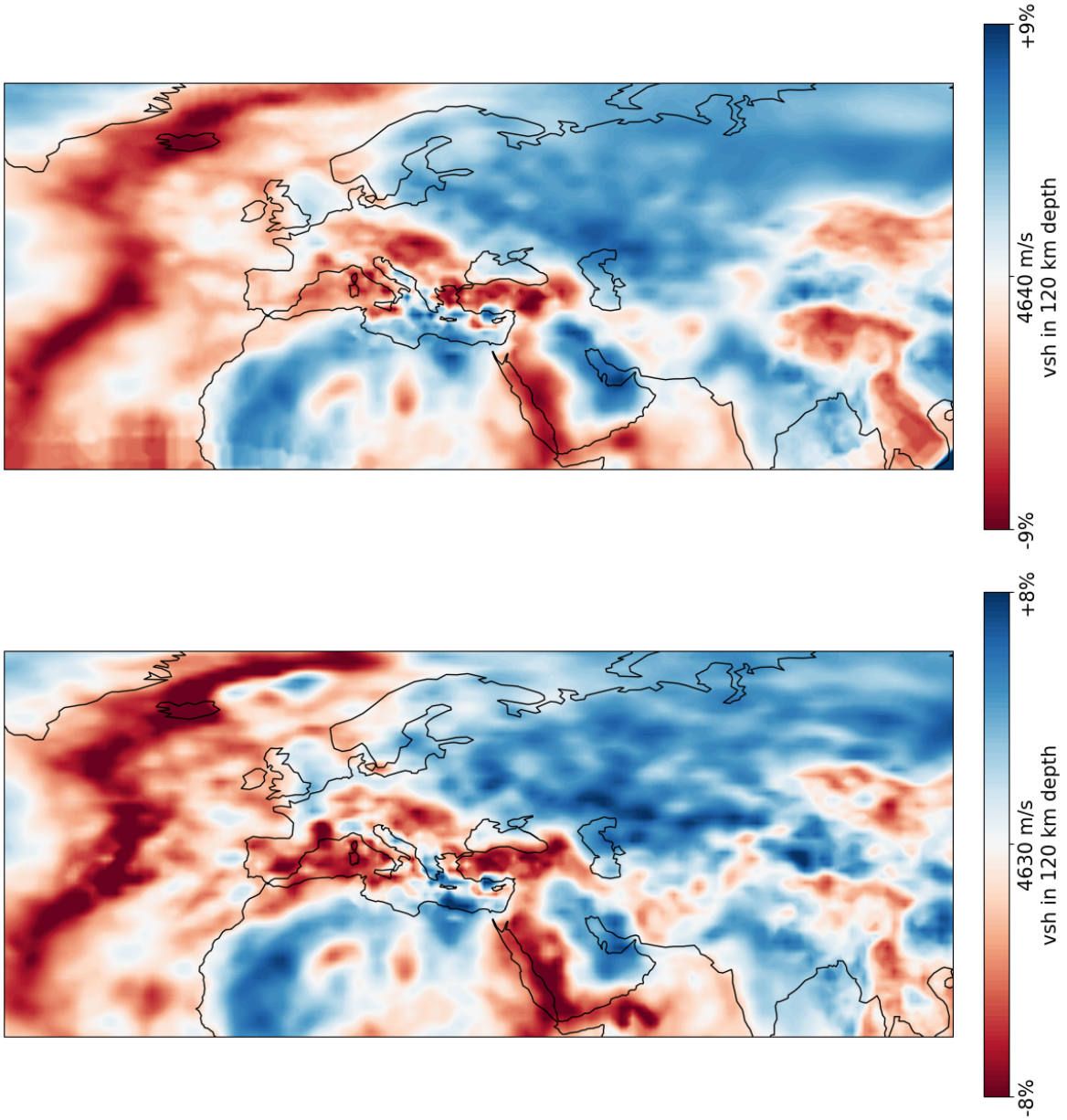

Figure A-7: Comparison between CSEM base model (upper figure) and the current model (lower figure) in horizontal shear velocity at 120 km depth.

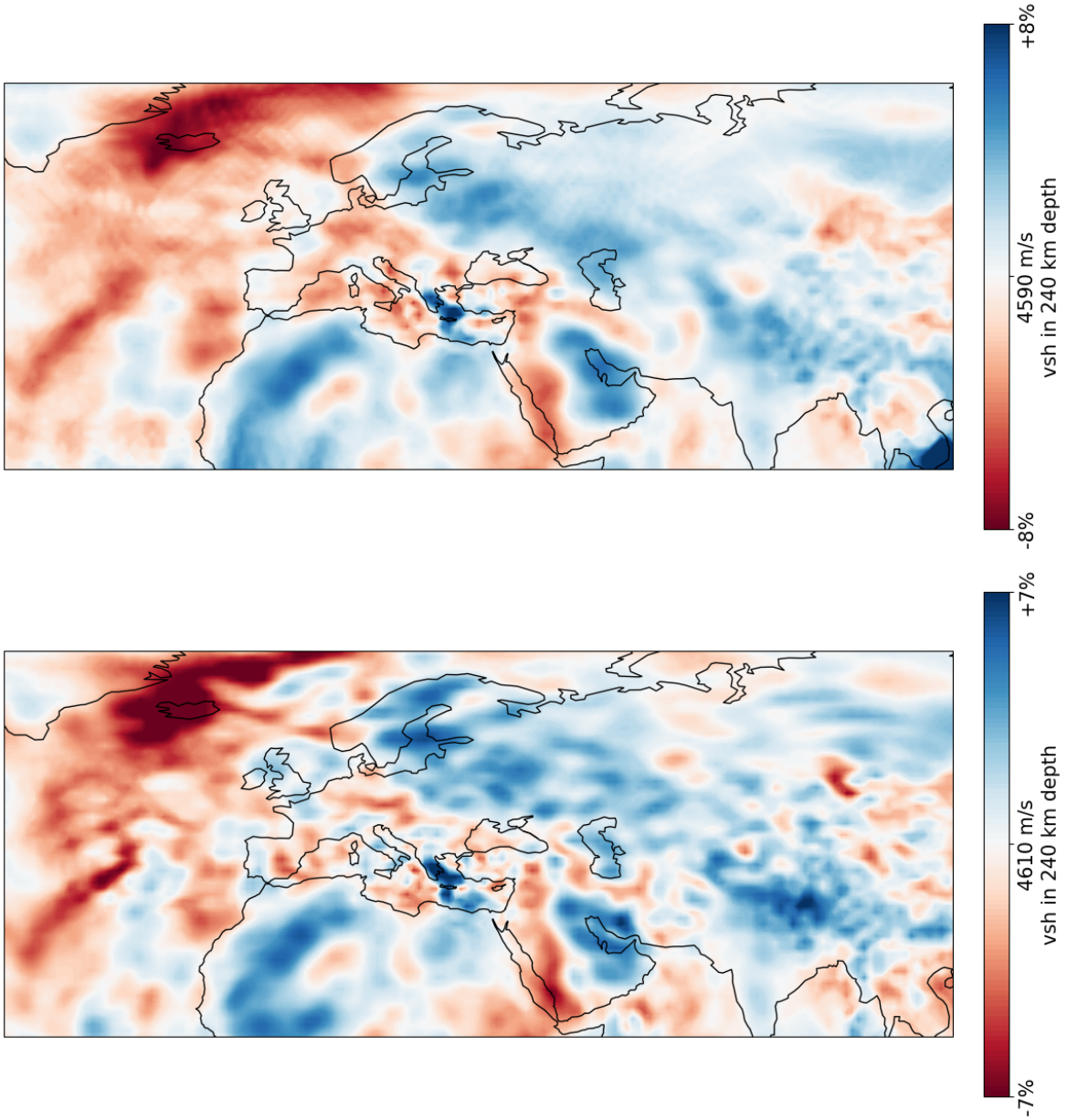

Figure A-8: Comparison between CSEM base model (upper figure) and the current model (lower figure) in horizontal shear velocity at 240 km depth.

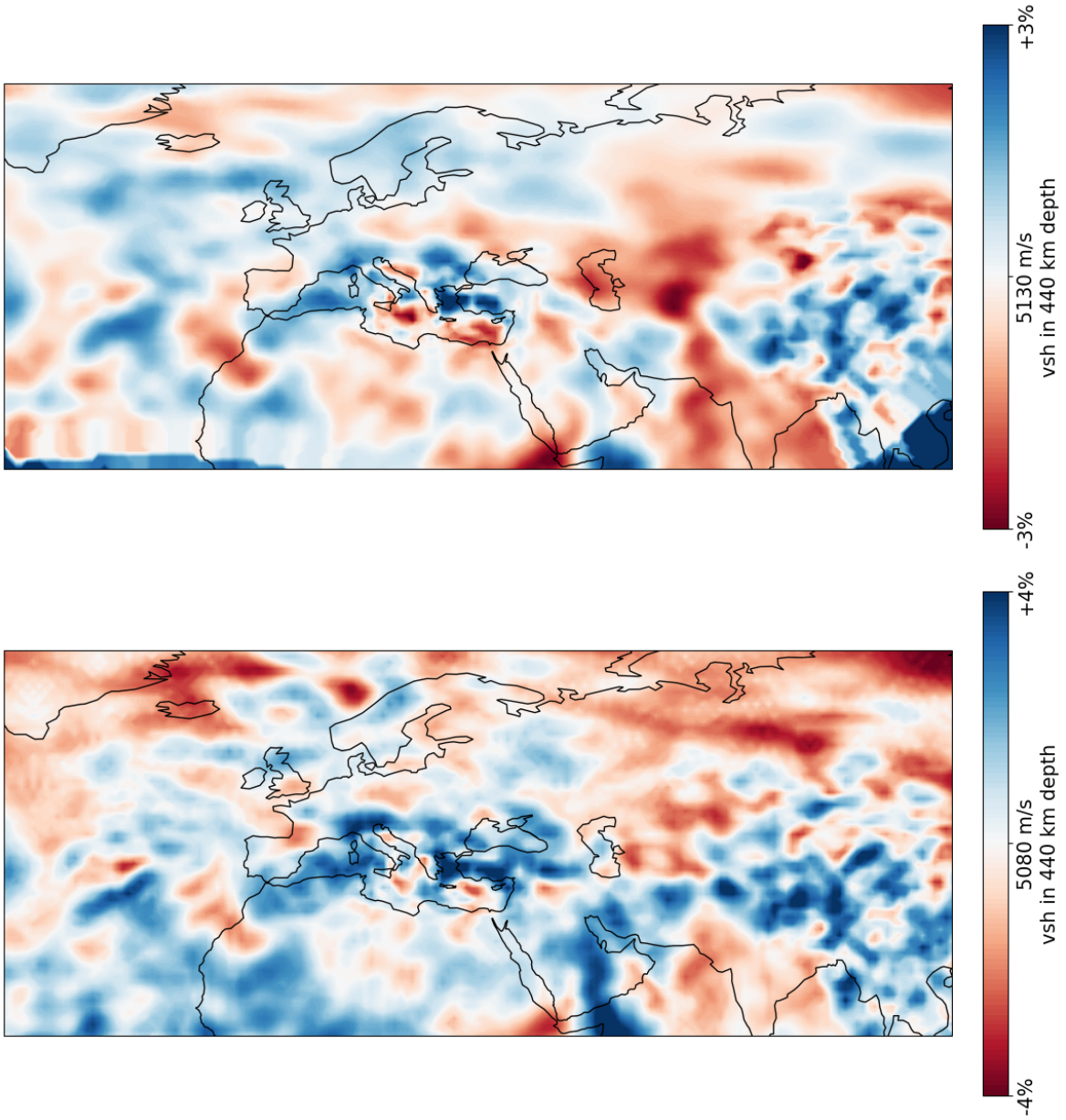

Figure A-9: Comparison between CSEM base model (upper figure) and the current model (lower figure) in horizontal shear velocity at 440 km depth. This depth corresponds with the boundary between the upper mantle and the beginning of the mantle transition zone.

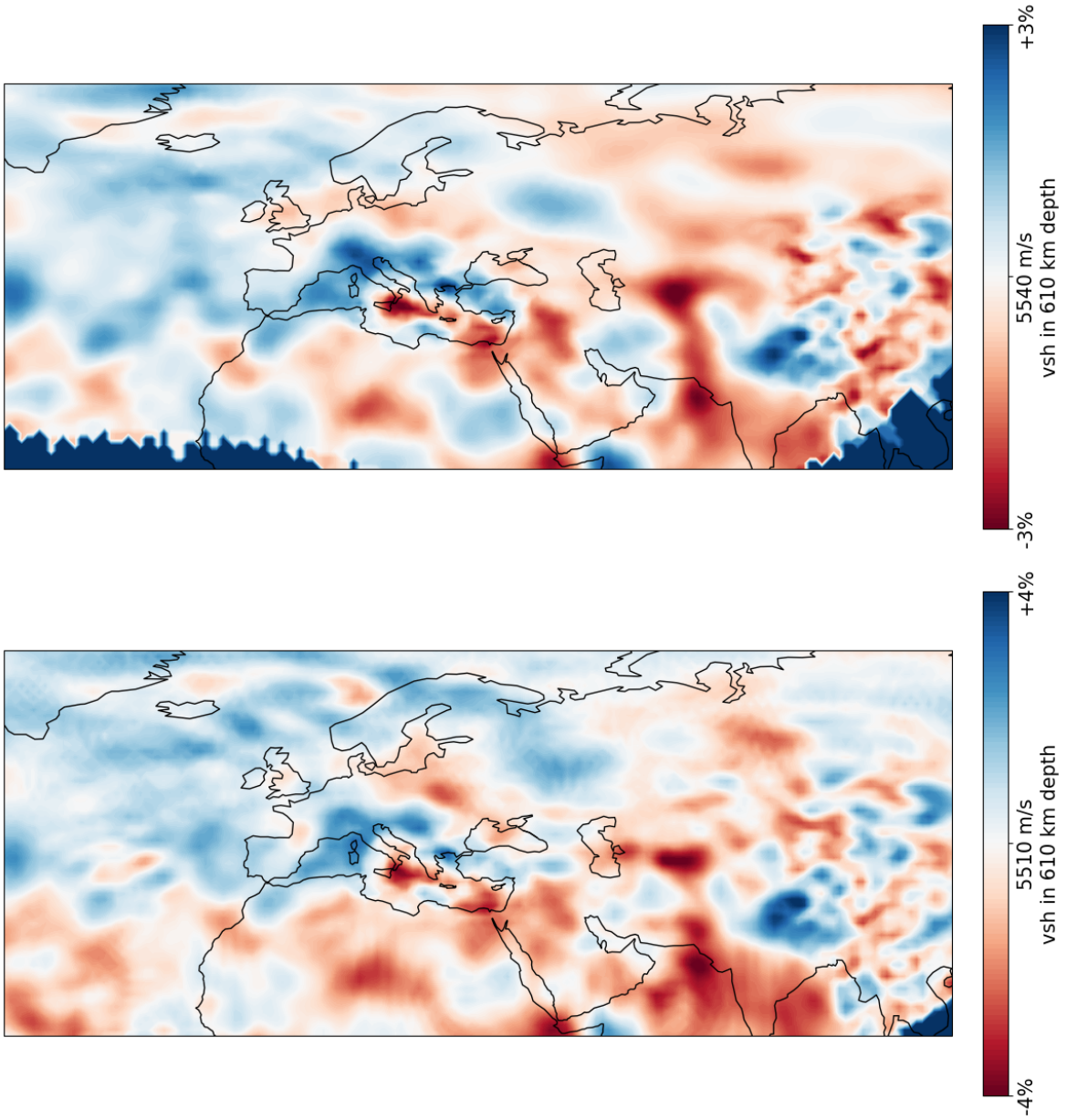

Figure A-10: Comparison between CSEM base model (upper figure) and the current model (lower figure) in horizontal shear velocity at 610 km depth. This depth corresponds with the boundary between the end of the mantle transition zone and the beginning of the lower mantle.

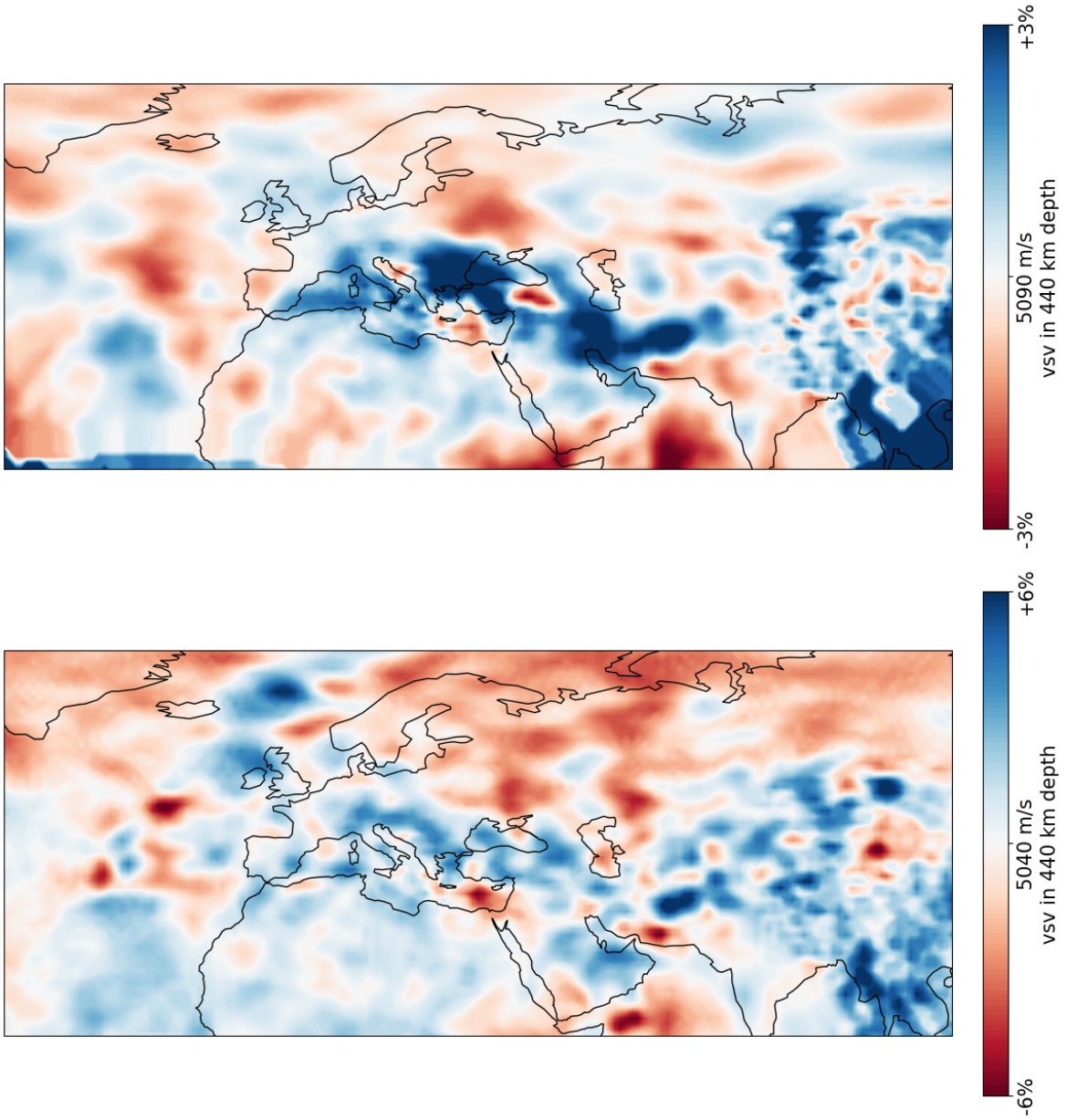

Figure A-11: Comparison between CSEM base model (upper figure) and the current model (lower figure) in vertical shear velocity at 440 km depth. This depth corresponds with the boundary between the upper mantle and the beginning of the mantle transition zone.

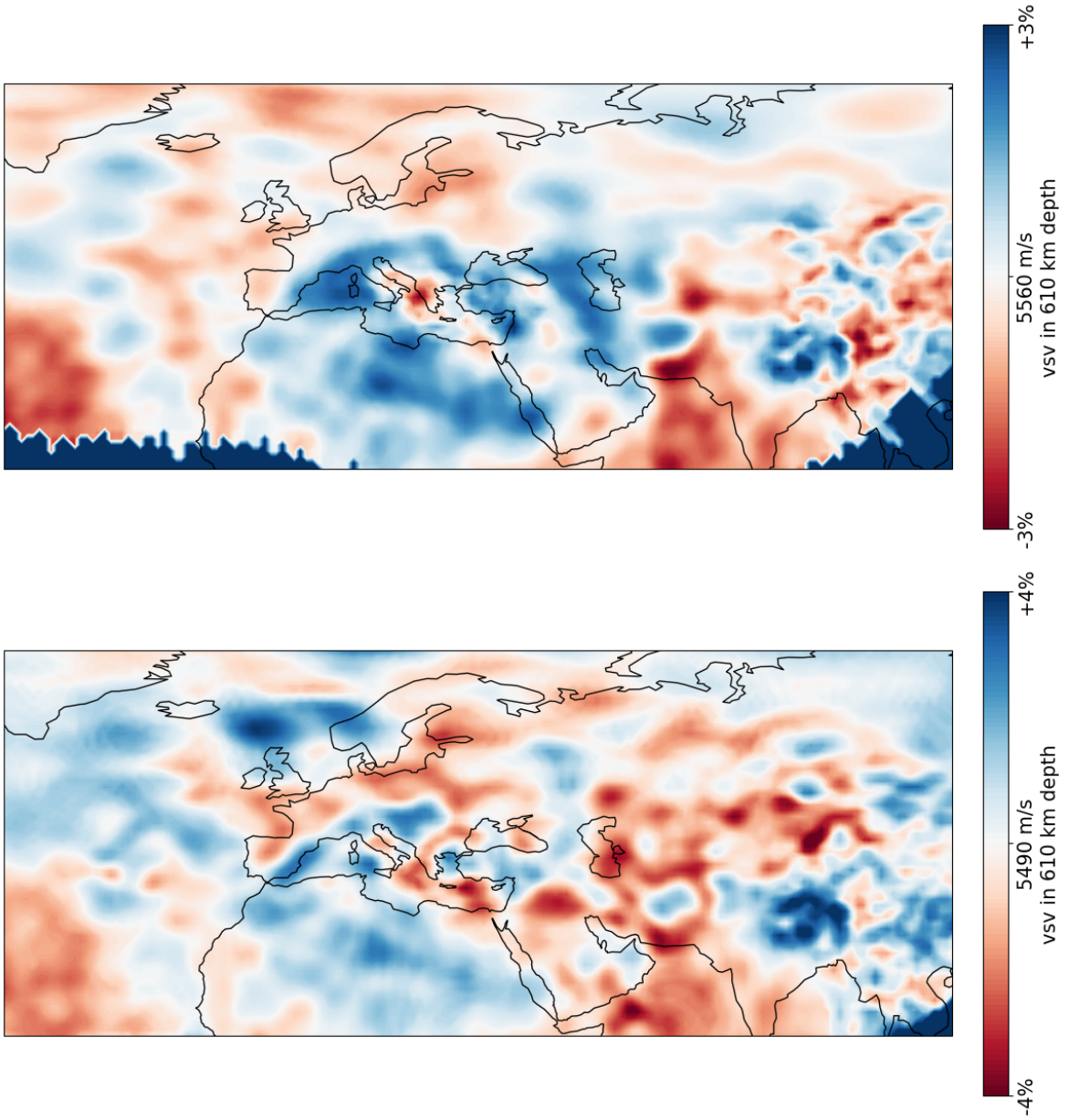

Figure A-12: Comparison between CSEM base model (upper figure) and the current model (lower figure) in vertical shear velocity at 610 km depth. This depth corresponds with the boundary between the end of the mantle transition zone and the beginning of the lower mantle.