Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Corso di Laurea Magistrale in Fisica del Sistema Terra

Anisotropic Seismic Tomography with the Reversible Jump Markov Chain Monte Carlo

Relatore:
Prof. Andrea Morelli

Correlatori:
Prof. Manuele Faccenda
Dott. Brandon VanderBeek

Presentata da:
Gianmarco Del Piccolo

Sessione I
Anno Accademico 2021/2022

Abstract

The established isotropic tomographic models show the features of subduction zones in terms of seismic velocity anomalies, but they are generally subjected to the generation of artifacts due to the lack of anisotropy in forward modelling. There is evidence for the significant influence of seismic anisotropy in the mid-upper mantle, especially for boundary layers like subducting slabs. As consequence, in isotropic models artifacts may be misinterpreted as compositional or thermal heterogeneities.

In this thesis project the application of a trans-dimensional Metropolis-Hastings method is investigated in the context of anisotropic seismic tomography. This choice arises as a response to the important limitations introduced by traditional inversion methods which use iterative procedures of optimization of a function object of the inversion.

On the basis of a first implementation of the Bayesian sampling algorithm, the code is tested with some cartesian two-dimensional models, and then extended to polar coordinates and dimensions typical of subduction zones, the main focus proposed for this method. Synthetic experiments with increasing complexity are realized to test the performance of the method and the precautions for multiple contexts, taking into account also the possibility to apply seismic ray-tracing iteratively.

The code developed is tested mainly for 2D inversions, future extensions will allow the anisotropic inversion of seismological data to provide more realistic imaging of real subduction zones, less subjected to generation of artifacts.

Sommario

Gli attuali affermati modelli tomografici isotropi mostrano le caratteristiche delle zone di subduzione in termini di anomalie di velocità sismica, ma sono tipicamente soggetti alla generazione di artefatti dovuti alla mancanza di una modellizazione anisotropa, di cui cui c'è stata crescente evidenza nel mantello medio-superiore, in particolare per boundary layers come slab in subduzione. Ne segue che con modelli isotropi gli artefatti potrebbero essere male-interpretati come eterogeneità composizionali o termiche.
In questo progetto di tesi è stata investigata l'applicazione di un metodo trans-dimensionale Metropolis-Hastings per la tomografia sismica anisotropa. Questa scelta nasce come risposta alle importanti limitazioni introdotte dai metodi tradizionali di inversione che utilizzano procedimenti iterativi nell'ottimizzazione di una funzione oggetto dell'inversione.
Sulla base di una prima implementazione dell'algoritmo di campionamento bayesiano, il codice è stato testato su alcuni modelli cartesiani bi-dimensionali, per poi essere esteso a coordinate polari con dimensioni tipiche delle zone di subduzione, che sono il principale focus proposto per questo metodo. Sono stati realizzati esperimenti sintetici con complessità crescente per valutare le capacità del metodo e gli accorgimenti ottimali in diversi contesti, tenendo anche conto della possibilità di tracciare i raggi sismici iterativamente.
Il codice sviluppato è stato testato per inversioni principalmente 2D, future estensioni permetteranno l'inversione anisotropa di dati sismologici per provvedere a una maggiormente accurata ricostruzione delle reali zone di subduzione, meno soggetta a generazione di artefatti.

Contents

Chapter 1

Introduction

Seismic tomography is the branch of Earth sciences which studies the physical properties of planetary interiors on the basis of their influence on the propagation of seismic waves. In seismic tomography with travel times, the seismological data are the intervals of time required by seismic waves to reach seismic stations, after being nucleated from the source event. These data can provide constrains to determine the inner seismic velocity field of the Earth using inversion techniques.
Seismic tomographic problems are generally ill-conditioned due to the limitations in the sampling properties of the seismic rays; the pathological behaviours, which typically arise as non-uniqueness of the solutions, can be treated in multiple ways. Standard methods deal with the ill-condition introducing additional constrains to the model, but incurring in the dependency of the solutions from the way the constrains are integrated.
In this thesis project the main limitations of standard non-linear inversion solvers will be investigated and the application of a stochastic solver for anisotropic seismic tomography will be discussed to overcome them. The main objective is to put the basis for an inversion code which could provide a new insight into the dynamics and the structures of subduction zones using anisotropic forward modelling. Starting from a initial implementation of this method, known as the *Reversible Jump Markov Chain Monte Carlo*, in this thesis extensions of the code and new implementations are proposed.
Forward modelling implemented takes into account the anisotropic contribution in the seismic P-waves velocity fields and its role in subduction zones, where still today interpretation of anisotropy remains elusive. Even if anisotropy in the Earth is important, it has been neglected in most of the tomographic studies since it increases the under-determined nature of the inverse problem. Isotropic velocity models found with the tomographic works of the last forty years are not sufficient to constrain subduction zones, where typically artifacts may be produced in the inversion and misinterpreted as compositional or thermal heterogeneities, due to the assumed isotropic nature of the velocity field. Fully anisotropic inversion could better constrain these regions and remove these artifacts leaving only real heterogeneities.
Main applications of the method here described consist in two-dimensional synthetic experiments, starting from cartesian spaces to polar geometries, where subducting models are tested with an increasing level of complexity.
The last chapter is dedicated to realistic two-dimensional subduction models, where local earthquakes - teleseismic joint tomography is used to better constrain anisotropic velocity fields. Ray-tracing algorithms are described and adopted in the inversion to deal with the intrinsic non-linearity of tomographic problems, where ray-paths depend on the velocity field.

Chapter 2

Theoretical background

2.1 Seismic waves in anisotropic media

On typical time-scales of the seismic-wave fields, deformations of the rocks which constitute the Earth are so small that it is possible to describe elasto-dynamics phenomena in the framework of the *Linear Elastic Theory*.
In linear elasticity the fundamental relationship between strain and stress tensorial fields is represented by:

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl} \tag{2.1}$$

where σ_{ij} and ϵ_{ij} are the stress and strain fields, while C_{ijkl} is a fourth order tensor collecting the elastic coefficients of the medium.
Often the 21 independent coefficients of the elastic tensor are arranged in a 6x6 matrix for arbitrary anisotropic materials.

$$C = \begin{bmatrix} c_{1111} & c_{1122} & c_{1133} & c_{1123} & c_{1113} & c_{1112} \\ c_{2211} & c_{2222} & c_{2233} & c_{2223} & c_{2213} & c_{2212} \\ c_{3311} & c_{3322} & c_{3333} & c_{3323} & c_{3313} & c_{3312} \\ c_{2311} & c_{2322} & c_{2333} & c_{2323} & c_{2313} & c_{2312} \\ c_{1311} & c_{1322} & c_{1333} & c_{1323} & c_{1313} & c_{1312} \\ c_{1211} & c_{1222} & c_{1233} & c_{1223} & c_{1213} & c_{1212} \end{bmatrix} \tag{2.2}$$

Rocks of the Earth can be treated in a first approximation as isotropic media, but there is evidence for many minerals in crust, mantle and core to exhibit anisotropic properties on seismic scale.

In continuum mechanics conservation laws for physical quantities like linear momentum can be formulated for any possible medium.
Using differential form, linear momentum conservation can be formulated as:

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} \right) = \mathbf{f} + \nabla \cdot \boldsymbol{\sigma} \tag{2.3}$$

where ρ is the density of the medium, $\mathbf{u}$ is the displacement vector field, $\mathbf{v}$ is the velocity vector field, $\mathbf{f}$ represents the body forces acting on the medium and $\boldsymbol{\sigma}$ is the stress field. In

linear elasticity, equations 2.3 can be written neglecting the non-linear term of the derivative at the first member. Assuming homogeneous C_{ijkl}, in absence of body forces the equations are:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = C_{ijkl} \frac{\partial}{\partial x_j} \frac{\partial u_l}{\partial x_k} \tag{2.4}$$

These relationships represent the wave equations in an anisotropic linear elastic medium. If a plane wave solution travelling in $\mathbf{n}$ direction is considered:

$$\mathbf{u}(\mathbf{r}, t) = \mathbf{a}\, f(t - \frac{\mathbf{n} \cdot \mathbf{r}}{c}) \tag{2.5}$$

and substituted in equation 2.4 (see [Paullsen, 2014-15] for details), the following relationship holds for $\mathbf{a}$ (polarization versor) and c (wave speed):

$$\frac{1}{\rho} C_{ijkl} n_j n_k a_l = m_{il} a_l = c^2 a_i \tag{2.6}$$

where m_{il} is an element of the Christoffel matrix. If the matrix is diagonalized, three independent wave equations hold, for three possible polarizations. The eigenvalues and the eigenvectors of the Christoffel matrix provide the phase velocities and directions of polarization of independent plane waves in an anisotropic medium.

If the medium is isotropic, the independent coefficients of the elastic tensor are reduced to 2 parameters: λ and μ (rigidity), called Lamè coefficients.

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il}) \tag{2.7}$$

$$C = \begin{bmatrix} \lambda + 2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda + 2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda + 2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{bmatrix} \tag{2.8}$$

The coupling of equations 2.7 and 2.3 (1st order) leads to the well known Cauchy-Navier equations, which describe elastic phenomena in isotropic linear homogeneous elastic media.

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \mathbf{f} + (\lambda + \mu) \nabla (\nabla \cdot \mathbf{u}) + \mu \nabla^2 \mathbf{u} \tag{2.9}$$

Considering an arbitrary propagation versor $(n_1, n_2, n_3)^T$, the Christoffel matrix is written as:

$$M = \frac{1}{\rho} \begin{bmatrix} (\lambda + \mu) n_1^2 + \mu & (\lambda + \mu) n_1 n_2 & (\lambda + \mu) n_1 n_3 \\ (\lambda + \mu) n_1 n_2 & (\lambda + \mu) n_2^2 + \mu & (\lambda + \mu) n_2 n_3 \\ (\lambda + \mu) n_1 n_3 & (\lambda + \mu) n_2 n_3 & (\lambda + \mu) n_3^2 + \mu \end{bmatrix} \tag{2.10}$$

The eigenvalues of M are $\lambda_1 = \lambda_2 = \mu/\rho$ and $\lambda_3 = (\lambda + 2\mu)/\rho$. The eigenvector corresponding to λ_3 is $\mathbf{n}$, while the eigenvectors corresponding to λ_1 and λ_2 are orthogonal to $\mathbf{n}$. Then in a linear isotropic elastic medium longitudinal and transverse waves are admitted (P-waves

and S-waves), with phase velocities (according to equation 2.6): $c_P = \sqrt{(\lambda + 2\mu)/\rho}$ and $c_S = \sqrt{\mu/\rho}$.

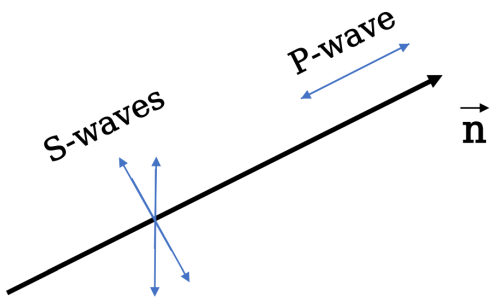

Figure 2.1: Polarization of P- and S- waves in a linear isotropic elastic medium.

Considering an arbitrary point $\mathbf{x}$ in an elastic medium, the travel time $T(\mathbf{x})$ is defined as the interval of time taken by the wave to reach $\mathbf{x}$, starting from the source location. The wave-front is defined as the collection of points in the space characterized by the same travel-time, so it is described by the equation:

$$T(\mathbf{x}) = const \tag{2.11}$$

In an isotropic homogeneous medium the wave-fronts are spherical surfaces for an isotropic point source, while their geometry is more complex in media which deviate from these conditions.

If hexagonal symmetry is assumed for the elastic tensor with symmetry axis along direction $(0, 0, 1)^T$, independent elements are reduced from 21 to 5:

$$C = \begin{bmatrix} A & A-2N & F & 0 & 0 & 0 \\ A-2N & A & F & 0 & 0 & 0 \\ F & F & C & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 \\ 0 & 0 & 0 & 0 & L & 0 \\ 0 & 0 & 0 & 0 & 0 & N \end{bmatrix} \tag{2.12}$$

According to definition of Christoffel matrix in eq. 2.6, and assuming propagation in x-direction $(1, 0, 0)^T$, we have:

$$M = \frac{1}{\rho} \begin{bmatrix} A & 0 & 0 \\ 0 & N & 0 \\ 0 & 0 & L \end{bmatrix} \tag{2.13}$$

With this direction of propagation the Christoffel matrix is already diagonal, with eigenvalues $\lambda_1 = A/\rho$, $\lambda_2 = N/\rho$, $\lambda_3 = L/\rho$, with associated seismic velocities: $c_1 = \sqrt{A/\rho}$

(compressional wave), $c_2 = \sqrt{N/\rho}$ and $c_3 = \sqrt{L/\rho}$ (transverse waves). Wave equation involving Christoffel matrix shows evidence for three different polarization directions, one for the propagation of P-waves, and two for propagation of S-waves. When the wave propagates in x-direction, S-waves propagate with different speeds on the basis of the polarization, leading to the phenomenon known as *shear splitting*.

If the waves propagate in the same direction of the symmetry axis $(0, 0, 1)^T$ of the hexagonal symmetry, the Christoffel matrix is given by:

$$M = \frac{1}{\rho}\begin{bmatrix} L & 0 & 0 \\ 0 & L & 0 \\ 0 & 0 & C \end{bmatrix} \tag{2.14}$$

In this case the compressional wave propagating along $(0, 0, 1)^T$ has speed $c_P = \sqrt{C/\rho}$, while S-waves propagate with the same speed $c_S = \sqrt{L/\rho}$ due to the transverse isotropy, no wave splitting occurs.
Hexagonal symmetry for the elastic tensor has been proved to be a good approximation to represent seismic-scale anisotropic properties of olivine minerals, despite the orthorombic symmetry system of the lattice. This is significative, since olivine constitutes around the 60-70% (mass percent) of the upper mantle rocks. Another fundamental constituent of upper mantle rocks is orthopyroxene, also characterized by orthorombic symmetry structure.
In table 2.1 symmetry systems with the numbers of independent elements of the elastic tensor are summarized with corresponding examples [Paullsen, 2014-15].

Symmetry system	**independent coefficients**	**example**
isotropic solid	2	volcanic glass
cubic	3	garnet
hexagonal	5	ice
trigonal I	7	ilmenite
trigonal II	6	quartz
tetragonal	6	stishovite
orthorombic	9	olivine
monoclinic	13	hornblende
triclinic	21	plagioclase

Table 2.1: Symmetry systems and associated independent elastic coefficients for some media.

2.2 Origin of seismic anisotropy in the Earth

All minerals are intrinsically anisotropic, but since the typical wavelengths for teleseismic waves are larger than 5 km, the origin of seismic anisotropy needs to be investigated in the macroscopic reference frame, which is also the basis reference for the inversion procedure.
According to the accepted models, two main causes of seismic anisotropy occur in the Earth:

- intrinsic anisotropy: this is the result of a common crystallographic alignment among intrinsically anisotropic minerals. The resulting lattice preferred orientation (LPO) extends the anisotropic behaviour from the grain scale (microns to millimeters) to the macroscopic reference frame ($>$ kilometers).

- extrinsic anisotropy: layering, or any other non-random distribution, of (even isotropic) media can result in an anisotropic structure of the seismic velocity field. This anisotropic behaviour can characterize crack-dense regions, where a fluid fills the empty spaces and a common orientation between the cracks exists [Hansen et al., 2021].

Lattice preferred orientation origins from *dislocation glide* (or *slip*). When crystalline media are subjected to the presence of dislocations, these defects can propagate under the influence of a shear stress applied, generating a strain rate.
Any dislocation, interpreted as a linear defect in the crystalline structure, can be represented as the superposition of two basic types [Turcotte & Schubert, 2002]: *screw dislocations* and *edge dislocations*.
For this discussion a simple cubic lattice is considered (Fig. 2.2). An edge dislocation is an extra half-plane of atoms introduced in the lattice which distorts the neighbouring planes of atoms (Fig. 2.3). A screw dislocation is the result of a shear distortion, in Fig. 2.4 the deformation is out of the plane of atoms, solid-points atoms are from the plane below and the slip occurs orthogonal to the figure. The construction of a closed path in the lattice, called *Burgers circuit*, can evidence the deformation introduced by the presence of the dislocation. The relative atomic motion (slip) occurring when a *dislocation line* (frontier of a dislocation surface) passes through a crystalline structure is called the *Burgers vector*. The surface described by the motion of the dislocation line is called the *glide surface* (or *slip plane*).
Two dislocation motions are typically allowed in a lattice: *dislocation climb* and *dislocation slip*. Considering an edge dislocation, in the dislocation climb process the extra half-plane of atoms changes its extension and the dislocation line moves parallel to the dislocation surface adding or removing atoms due to shear stress-induced diffusion. The dislocation slip process occurs when a shear stress is applied to a lattice and an edge dislocation is transferred among parallel atomic planes, the motion of the dislocation line is orthogonal to the extra-plane of atoms.
Dislocation climb is not mass-conserving since it needs diffusion of atoms in order to increase or reduce the half-plane extension, on the other hand the dislocation slip is a mass-conserving motion. This is the reason why dislocation slip is a more rapid process in terms of strain rate generated in the lattice.
Several models have been developed to describe the strain rate generated by the motion of dislocations under the influence of a shear stress. This process is referred to as *dislocation creep*, in opposition to the *diffusion creep*, where the strain rate is generated by the diffusion of point-defects of the lattice under the application of a shear stress. While diffusion creep provides a Newtonian behaviour for the rocks interested by the process, all the models

proposed for the dislocation creep result in a non-Newtonian rheology. Both processes are thermally activated, temperatures typical of the mantle are necessary to induce creep.

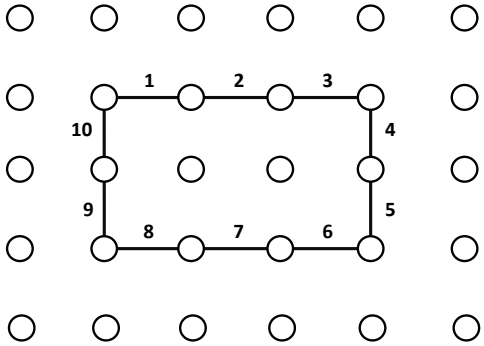

Figure 2.2: Cubic lattice. The Burgers circuit (closed path) represented takes ten steps to complete.

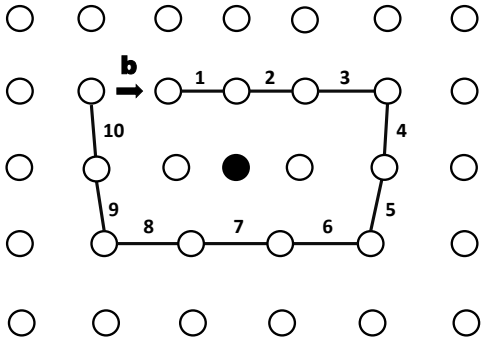

Figure 2.3: Edge dislocation in a cubic lattice. The Burgers circuit takes an additional step to complete, the acquired slip is referred to as the *Burgers vector* **b**. The Burgers vector is parallel to the plane of atoms and orthogonal to the dislocation line, the end of the extra half-plane.

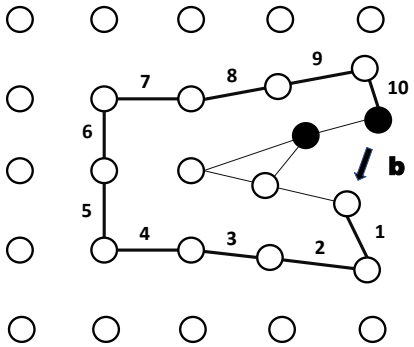

Figure 2.4: Screw dislocation in a cubic lattice. The Burgers vector **b** is orthogonal to the plane of atoms and parallel to the dislocation line.

Deformation induced by dislocation slip is simple shear and therefore it has a rigid body rotational component (Fig. 2.5). A slip system is associated with the dislocation slip deformation, and it is defined is terms of the Burgers vector **b**, which gives the direction and the magnitude of the slip introduced in the lattice, and the slip plane generated by the dislocation glide, identified by its normal versor **n**.

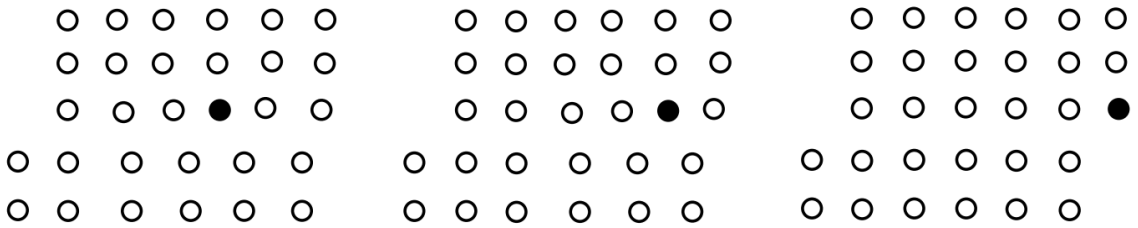

Figure 2.5: Three stages of dislocation slip and the evolving simple shear deformation of the lattice.

The rotational contribution to deformation is described by the second order tensor **Q**:

$$Q_{ij} = \frac{1}{2}(b_i n_j - b_j n_i) \tag{2.15}$$

Then, during the dislocation slip the crystals of the medium rotate around the axis $\mathbf{n} \times \mathbf{b}$. The rotational velocity depends on the orientation of the crystallographic axes with respect to an external fixed reference frame, so the rate of rotation can be written as: $Q_{ij}\dot{\gamma}$, where $\dot{\gamma}$ is the strain rate produced by the motion of dislocations in the reference frame of the slip system. This differential rotational behaviour is the origin of the lattice preferred orientation.
A lattice is generally characterized by multiple possible slip systems, those with a faster

strain rate are called *easy* slip systems, while those with a slower contribution are called *hard* slip systems. There are minerals where the difference in strain rate between easy and hard slip systems is so marked that generally one slip system dominates the deformation process. When a single slip system dominates, slower rotational velocities are associated with grains where the local shear generated by the dislocation motion coincides with the macroscopic shear applied. In these cases the relationship between LPO and applied shear is straightforward: grains tend to align toward the direction where the local shear produced by dislocation glide coincides with the macroscopic shear. Olivine and orthopyroxene belong to this class of minerals [Karato et al., 2008].
Using these results a simple model can be proposed to describe the relationship between LPO and macroscopic flows: if a simple shear deformation is assumed, like the deformation state beneath the lithosphere, olivine and orthopyroxene crystals are aligned with the slip directions and the dislocation glide planes parallel to the macroscopic flows.

The nature of LPO in a lattice is generally determined by the dominant slip system, if it exists. The favoured slip direction is influenced by the direction where the Burgers vector **b** has the smallest length, since the direction is associated with the lowest strain energy. In olivine crystals, Burgers vectors along crystallographic axes [1 0 0] and [0 0 1] have similar lengths: 0.476 nm and 0.598 nm respectively. The third direction [0 1 0] has a length for **b** equal to 1.02 nm. According to the lowest strain energy, directions [1 0 0] and [0 0 1] should be favoured for the slip direction, but since the difference is not strong, we could expect the dominant slip direction to change under the influence of other parameters. In orthopyroxene Burgers vector in direction [0 0 1] has a length equal to 0.518 nm, while directions [1 0 0] and [0 1 0] have lengths for **b** equal to 1.82 nm and 0.881 nm. In this mineral, axis [0 0 1] for slip direction is strongly favoured with respect to the other ones.
Once the dominant Burgers vector has been identified, in order to determine an unique dominant slip system it is necessary to find the slip plane. A criterion to identify the dominant slip plane is based on the Peierls stress, the resistance to shear deformation at T = 0 K, and it consists in finding the plane where the ratio h/b is the largest, with h distance between parallel planes and b the length of the Burgers vector. Despite this criterion, influence of the Peierls stress at high temperature is indirect in determination of dominant slip system and there is evidence for an effect of hydration in Peierls stress reduction.

Development of LPO depends on the physical and chemical state subjecting the media. These contributions can significantly change the dominant slip system occurring in the lattice. If a dominant slip system constitutes a *fabric*, a fabric transition can occur when the strain rate produced by an alternative slip system increases beyond the precedent dominant one.
Strain rate produced by a slip system i can be written as a function of this kind:

$$\dot{\epsilon}_i = f(\, T, P, \sigma; f_{H_2O}, f_{O_2}, a_{MO}\,) \tag{2.16}$$

where T is the temperature, P the pressure, σ the applied deviatoric stress, f_{H_2O} and f_{O_2} are the water and oxygen fugacities and a_{MO} is the activity of oxide. Fugacity is introduced in statistical mechanics in grand canonical ensembles (ensembles that can exchange particles with the environment), its logarithm, multiplied by k_BT is the chemical potential (k_B is the Boltzmann's constant). It represents the effort made when a particle is added to the ensemble.

Generally oxygen fugacity and oxide activity do not change significantly in the upper mantle, so these dependencies can be neglected. Pressure dependence can be made non-direct by normalizing the temperature by the melting temperature $T_m(P)$, which is a function of the pressure, and the deviatoric stress by the shear modulus at temperature T and pressure P, $\mu(T, P)$.

On the basis of the dependencies identified in eq. 2.16, a fabric transition between two slip systems i and j occurs at the fabric boundary identified by the equation:

$$\dot{\epsilon}_i \left(\frac{T}{T_m(P)}, \frac{\sigma}{\mu(T,P)}; f_{H_2O} \right) = \dot{\epsilon}_j \left(\frac{T}{T_m(P)}, \frac{\sigma}{\mu(T,P)}; f_{H_2O} \right) \tag{2.17}$$

Pressure intrinsic contribution may arise if the normalization scheme introduced by $T_m(P)$ and $\mu(T, P)$ is not sufficient to represent the influence of pressure.
Under the approximation made in eq. 2.17, the fabric boundary is described in a three-dimensional space by a relationship:

$$G \left(\frac{T}{T_m(P)}, \frac{\sigma}{\mu(T,P)}; f_{H_2O} \right) = 0 \tag{2.18}$$

Flow laws provide the constitutive relationship to describe the dependencies of strain rate from physical and chemical parameters, using a generic flow law [Karato et al., 2008]:

$$\dot{\epsilon} = A \cdot (1 + B \cdot f_{H_2O}^r) \cdot \sigma^n \cdot exp \left(-\frac{H^*}{RT} \left(1 - \left(\frac{\sigma}{\sigma_p} \right)^q \right)^s \right) \tag{2.19}$$

where H^* is the activation energy, σ_p the Peierls stress, while n, q, r and s are non-dimensional constants and A and B are parameters of the flow law.
If eq. 2.19 is substituted in eq. 2.17, two different fabric boundaries may occur. In the first case, assuming the stress contribution to activation energy small, i.e. $\sigma/\sigma_p << 1$, eq. 2.17 for the boundary is written as:

$$\frac{A_1 (1 + B_1 \cdot f_{H_2O}^{r1}) \cdot \sigma^{n1} \cdot exp \left(-\frac{H_1^*}{RT} \right)}{A_2 (1 + B_2 \cdot f_{H_2O}^{r2}) \cdot \sigma^{n2} \cdot exp \left(-\frac{H_2^*}{RT} \right)} = 1 \tag{2.20}$$

If the exponents of the stress are the same for two different slip systems (this is valid for olivine, see [Bai et al., 1991]), the fabric boundary relationship is:

$$log \frac{A_1}{A_2} \frac{1 + B_1 \cdot f_{H_2O}^{r1}}{1 + B_2 \cdot f_{H_2O}^{r2}} = \frac{H_1^* - H_2^*}{RT} \tag{2.21}$$

Boundary fabric for deformation at low stress is influenced by water fugacity and temperature. This boundary is referred to as *class-I fabric transition.*
The second case considers the influence of stress field on activation energy relevant, while the contribution of water fugacity is not taken into account. Under these assumptions, valid for deformation at high stress and low temperature, eq. 2.17 is written as:

$$log \frac{A_1}{A_2} = \frac{H_1^*}{RT} \left(1 - \left(\frac{\sigma}{\sigma_{p_1}} \right)^{q_1} \right)^{s_1} - \frac{H_2^*}{RT} \left(1 - \left(\frac{\sigma}{\sigma_{p_2}} \right)^{q_2} \right)^{s_2} \tag{2.22}$$

With these assumptions the fabric boundary is influenced by stress field and temperature, defining a *class-II fabric transition*.
Influence of water, temperature and stress has been experimentally addressed to LPO and the fabric transitions described have been observed [Karato et al., 2008]. Water fugacity term can be generally substituted with the water content.
Synthetic water-poor olivine deformed by simple shear geometry was experimentally studied [Zhang & Karato, 1995] at low stress and high temperature, providing results consistent with the dominant slip system [1 0 0] (0 1 0), where crystallographic axis for the slip direction is identified with the square brackets, while the normal to the slip plane with the curved.
By comparing experiments on aggregates of olivine and single crystals [Bai et al., 1991] evidence was found for correspondence between the identification of the weak slip system [0 1 0] (1 0 0). These results allow to describe efficiently the fabric transitions in olivine using single crystal flow laws.
Experiments performed on olivine manifested evidence for the relevant influence of water content in fabric transition. In these studies several fabrics of olivine are identified, each with a different dominant slip system determining the nature of LPO.
Five possible fabrics have been described for olivine, the corresponding slip systems are reported in table 2.2.
In Fig. 2.6 dominant slip system is studied as a function of the water content and stress. Including also dependence by temperature, fabric transition diagrams for olivine can be described using the three parameters identified in equation 2.17.
In Fig. 2.6 it can be noticed that the boundary between fabrics B-type and C-type is not significantly influenced by the stress field, but only by the water content, this recalls the result found in equation 2.22, corresponding to a class-II fabric transition, where only effects of stress and temperature are significant. Similar reasoning can be proposed for the boundary between A- and E- fabrics, where the influence of stress is not relevant. Recalling the result of equation 2.21, this boundary is classified as a class-I fabric transition, where only the effects of water content and temperature are significant.
According to data set in Fig. 2.6, anisotropy distribution can be predicted in some realistic geodynamical contexts. Beneath the lithosphere, where water content is low, with temperature and stress not high, according to the diagram for olivine, A-fabric can be assumed to be dominant. Proceeding into the asthenosphere, where the water content increases with temperature, but stress is kept moderated, fabric transitions between A-type and E-type or C-type may occur.
In Fig. 2.7 an olivine mono-crystal is represented, with compressional and shear-wave velocities for different crystallographic axes. Using this reference it is possible to interpret seismic data of shear wave splitting to infer the olivine-fabrics which describe the observations.

A-fabric	B-fabric	C-fabric	D-fabric	E-fabric
[100](010)	[001](010)	[001](100)	[100]{0*kl*}	[100](001)

Table 2.2: Fabrics identified for olivine. Slip direction is identified using [n l m], while normal to slip planes is identified with (n l m).

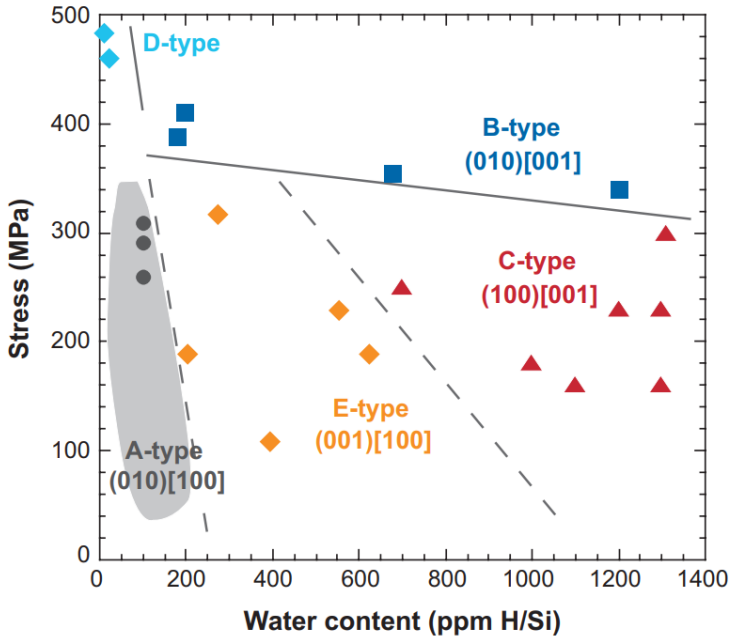

Figure 2.6: Deformation of olivine at temperatures between 1470-1570 K is studied as a function of stress applied and water content. Straight lines predict the fabric boundaries between different dominant slip systems. Data set is from Carter & Avé Lallemant (1970) and Karato et al. (2008), figure from Karato et al. (2008).

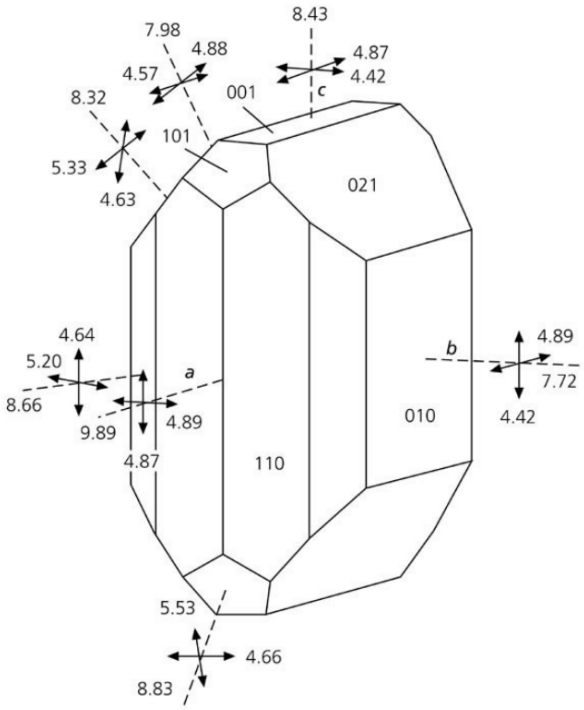

Figure 2.7: Mono-crystal of olivine. Crystallographic axes [100], [010] and [001] are identified by directions a, b and c. Compressional and shear-wave velocities are provided for multiple propagation directions. Original data are from Kumazawa & Anderson, 1969.

There is evidence of anisotropic behaviour of the velocity field across subduction zones from shear-wave splitting measurements at the surface. Data collected in different subduction zones make evident that adopting the old paradigm to interpret the relationship between LPO and macroscopic flow would lead to flows not matching geodynamical simulations in these regions. In Fig. 2.8 the fast shear-wave polarization axes are represented at the trench and away after the volcanic arc. At the trench the polarization axes are trench-parallel, while they become trench-normal on the other side, far from the trench. If A-type fabric were assumed to describe the origin of the anisotropy (old paradigm), according to Fig. 2.7, table 2.2 and vertical direction for the propagation of waves, the associated macroscopic flow should be trench-parallel at the trench and normal away from it. Since flow is expected to be almost vertical at the trench, shear-wave splitting measurements can be described using a B-type fabric, while, assuming a normal flow away from the trench, fabrics C-type and E-type can be proposed. Despite these interpretations, seismic anisotropy understanding remains still today elusive in subduction zones.

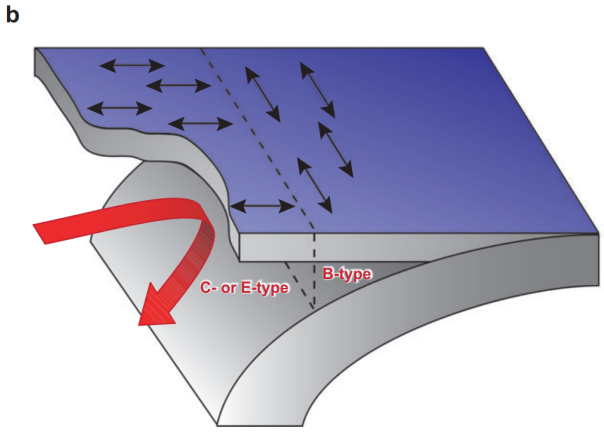

Figure 2.8: Anisotropy evidence at subduction zone (example of Tohoku, Japan). At the trench fast shear-wave velocity axes are trench-parallel, while away from the trench they are normally oriented. These observations cannot be described using A-type fabric and assuming flow direction described by the red arrow. Figure from [Karato et al., 2008]

Extrinsic anisotropy can origin from layering or non-random spatial arrangement of media with different elastic properties. If the wavelengths of seismic waves are much larger than the local scale of layering, a macroscopic anisotropic behaviour can arise, even if single media are intrinsically isotropic. Some sources of extrinsic anisotropy are listed and described in [Hansen et al., 2021]. What arises from works of [Hansen et al., 2021], [Faccenda et al., 2019] and [Tommasi & Vauchez, 2015], is that anisotropy is generally dominated by intrinsic anisotropy of media involved.

At this point it is evident the role of anisotropic seismic tomography in this context, inversion results of seismological data provide the indirect observations of seismic anisotropy. Using these inversions, more accurate characterization of subduction zones can be performed and used to infer the flows in the mantle in the light of the geodynamical models which predict the fabric types and distributions.

Chapter 3

Inversion methods in seismic tomography

3.1 Setup of the tomographic inverse problem

The scientific study of a physical system can be divided in three steps:

1. Parametrization of the system: evaluation of a minimal set of control parameters. The values completely characterize the modelling of the system.

2. Forward modelling: evaluation of the physical laws which allow to make predictions for a given parametrization.

3. Inverse modelling: use of the observations to infer the parameters of the model.

In seismic tomography with travel times the parametrization is the representation of the model (velocity field) in the region investigated. The velocity field is a continuous field, but in the computational framework it is necessary to approximate it using a set of discrete values, which is the minimal set of control parameters. The continuous field can be recovered using two possible approaches: local supports, where the functions used for the parametrization are defined locally, and global supports, the functions are defined over the entire domain.
Considering for simplicity a single-parameter model, but the extension to multi-parametric case is straightforward, the model in represented given a set of basis functions as:

$$m(\mathbf{r}) = \sum_{i=1}^{N} m_i \, f_i(\mathbf{r}) \tag{3.1}$$

where $f_i(\mathbf{r})$ is a function of the basis set and m_i is the *i-th* coefficient of the linear combination.

When the functions of the set are non-zero in all the space, or a large portion, the parametrization is called *global*. A typical example of a global parametrization is the representation of a field defined in a spherical domain using spherical harmonics:

$$m(r, \theta, \phi) = \sum_{n=0} \sum_{l=0} \sum_{m=-l}^{l} m_{n,l,m} \, f_n(r) \, Y_l^m(\theta, \phi) \tag{3.2}$$

where $f_n(r)$ are the radial eigenfunctions and $Y_l^m(\theta, \phi)$ are the surface spherical harmonics, which constitute an orthogonal set of eigenfunctions over the region spanned by $0 \leq \theta \leq \pi$

(co-latitude) and $0 \leq \phi \leq 2\pi$ (longitude):

$$Y_l^m(\theta,\phi) = (-1)^m \left[\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}\right]^{1/2} P_l^m(cos\theta)\, e^{im\phi} \tag{3.3}$$

where $P_l^m(cos\theta)$ are the associated Legendre polynomials.
Spherical harmonics parametrization of the velocity field was successfully used by Dziewonski and others [Dziwsonski et al., 1977] in some of the first works in seismic tomography. An interesting advantage of this procedure, in particular for the time, is that low degree spherical harmonics parametrization applies naturally a low-pass filter to seismological data and it can be easily compared with other maps described with the same formalism, like the geoid, the heat-flux distribution, etc.
The main disadvantage of this parametrization is the large maximum harmonic degree l needed to have high resolution; for a horizontal resolution of 100 km, l_{max} needed is around 200, with a corresponding total number of harmonic functions used around 80000. If this number is multiplied by the number of depth-basis functions we end up with 10^6 needed basis functions for the parametrization. This is the reason why for local tomographic problems the implementation of local basis functions is generally a less computational expensive approach.

An example of local parametrization is the use of pixels or voxels (cells in general, homogeneous or not) to represent the model. The values for the parameters are specified inside the cells and each cell can be seen as function defined locally, so that the basis function $f_n(\mathbf{r})$ is [Nolet, 2008]:

$$f_n(\mathbf{r}) = \begin{cases} \Delta V_n^{-1/2}, & \text{if } \mathbf{r} \text{ is in cell n} \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

where ΔV_n is the volume of the n-th cell.
Local parametrization with cells has been used successfully in the first studies of seismic tomography based on local earthquakes, like [Aki et al., 1977].
In order to have a continuous representation for the fields, interpolation algorithms are generally implemented using a set of nodes placed in correspondence to the cells.
Pixels/voxels discretization is an example of nearest neighbour interpolation, where the value of the field in a generic point in the space is the value associated with the closest node of the parametrization. In alternative, values associated with discrete nodes can be considered and interpolated is different ways. Linear interpolation algorithm performs the interpolation using a linear function whose parameters are determined using a specified set of nodes; for dimensions higher than one the algorithm is iterated along each axis. Lagrange interpolation rule allows to describe the field interpolated over a set of grid nodes as:

$$m(x,y,z) = \sum_{ijk=1}^{2} m_{ijk} m(x_i, y_j, z_k) \tag{3.5}$$

where the linear combination coefficients m_{ijk} are given by:

$$m_{ijk} = \left(1 - \frac{|x-x_i|}{x_2 - x_1}\right)\left(1 - \frac{|y-y_j|}{y_2 - y_1}\right)\left(1 - \frac{|z-z_k|}{z_2 - z_1}\right) \tag{3.6}$$

and $m(x_i, y_j, z_k)$ is the field evaluated in the node (x_i, y_j, z_k).
Cubic interpolation algorithm uses a polynomial function of order 3 to perform the interpolation, fixing the parameters locally for each set of needed nodes.

Interpolation examples for sparse data are represented in Fig. 3.1 using Python interpolation library.
An intuitive property of interpolation algorithms is that the fidelity of the model interpolated depends on the number of nodes generated, regions with high complexity may not be resolved if nodes are too sparse, while regions of lower complexity need a lower number of nodes to be resolved. An example is shown in Fig. 3.2, where lower-scale structures of the true model are resolved only with a sufficiently dense distribution of nodes. This means that in real problems the distribution of nodes is strongly related to the resolution we want to achieve; the distribution can be chosen a priori, on the basis of a prior information or personal choices, or it can be parametrized itself (Monte Carlo methods, see section 3.3).

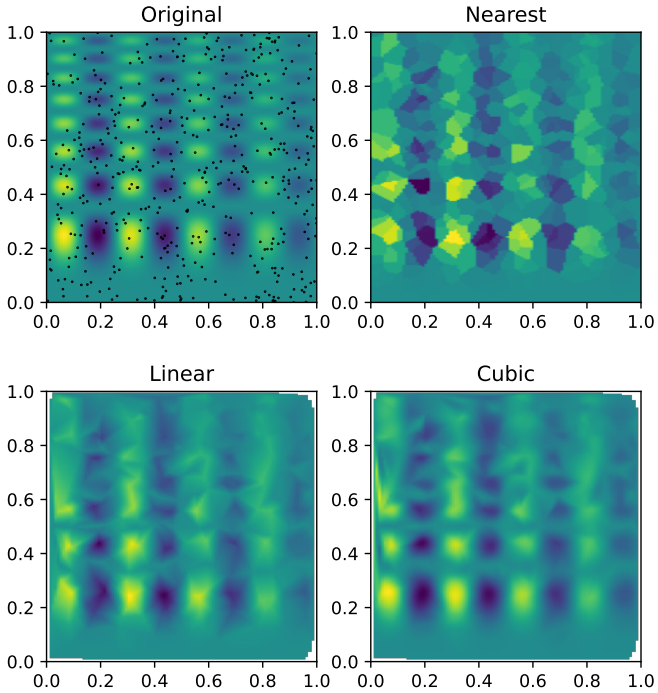

Figure 3.1: Interpolation of sparse data using nearest-neighbour, linear and cubic interpolation.

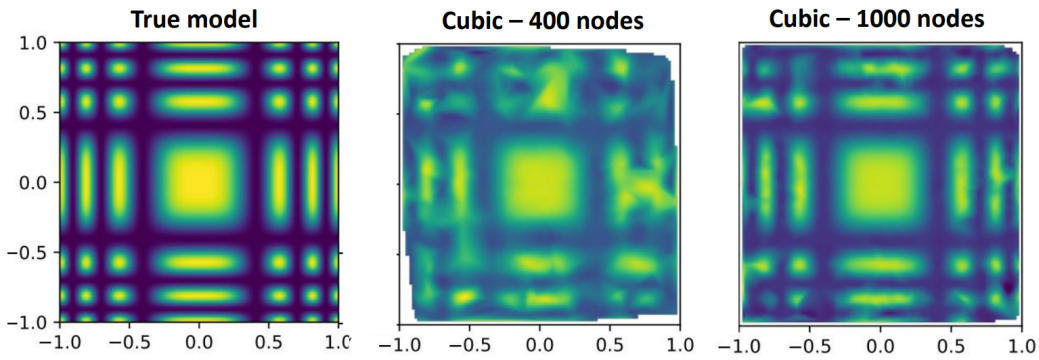

Figure 3.2: Left: true model, center: cubic interpolation with 400 randomly distributed nodes, right: cubic interpolation with 1000 randomly distributed nodes.

Given the representation of the model in terms of the basis functions (equation 3.1), a linear forward modelling allows to describe for a continuous model the predictions according to the relationship:

$$d_i^{pred.} = \int_{Space} G_i(\mathbf{r})\, m(\mathbf{r})\, d\mathbf{r} \tag{3.7}$$

where $G_i(\mathbf{r})$ is the kernel function for the forward modelling.
Expanding the model in terms of the basis functions:

$$d_i^{pred.} = \sum_{j=1}^{N} m_j \int_{Space} f_j(\mathbf{r}) G_i(\mathbf{r})\, d\mathbf{r} \tag{3.8}$$

If the matrix $\mathbf{G}$ is introduced, where the element G_{ij} is defined as:

$$G_{ij} = \int_{Space} f_j(\mathbf{r}) G_i(\mathbf{r}) \tag{3.9}$$

The following linear relationship is found:

$$\mathbf{d}^{pred.} = \mathbf{Gm} \tag{3.10}$$

If forward modelling is non-linear, the general relationship between predictions and model is:

$$\mathbf{d}^{pred.} = \mathbf{g}(\mathbf{m}(\mathbf{r})) \tag{3.11}$$

In isotropic seismic tomography with travel times, forward modelling is performed by computing the integral of the reciprocal of the velocity field along the ray path given. If the slowness (reciprocal of the velocity) field is used and the ray-path is fixed (otherwise non-linearity occurs), the problem is linear and the following relationships hold:

$$t_i^{pred.} = \int_{\Gamma_i} \frac{1}{v(\mathbf{r})}\, dl = \int_{\Gamma_i} c(\mathbf{r})\, dl \tag{3.12}$$

where $v(\mathbf{r})$ is the velocity field, $c(\mathbf{r})$ is the slowness and Γ_i is the ray-path.
If a local parametrization with voxels is assumed (Fig. 3.3) and if the slowness is uniform inside the homogeneous cells, the equation 3.12 is written:

$$t_i^{pred.} = \sum_{j=1}^{K} dl_i^k\, c_{j(ik)} \tag{3.13}$$

where dl_i^k is the length of segment k along the i-th path and $c_{j(ik)}$ is the slowness in the block j(ik), the k-th block crossed by the path Γ_i.
If all the voxels are considered, also the ones not crossed by any ray ($dl_i^k = 0$), the problem can be rearranged as a relationship of the kind of equation 3.10:

$$t_i = \sum_{j=1}^{M} l_{ij} c_j = (\mathbf{Gm})_i \tag{3.14}$$

Inverse problems theory deals mainly with ill-conditioned problems, which are generally over- or under-determined. Over-determined problems have a larger number of observations with respect to the parameters of the model, while under-determined problems are affected by the opposite condition. Tomographic problems are generally both over- and under-determined. In a typical tomographic problem if we adopt a voxels discretization for the velocity field, there are blocks crossed by multiple rays and blocks non crossed by any ray, so some parameters will be over-constrained, generating the over-determined condition of the problem, while other parameters are not enlightened by data, and this makes the solution not unique. Solution may be not unique also when a set of parameters of the model is enlightened by data in a similar way, an example in shown in Fig. 3.4. Here the values of velocity in the two blocks cannot be discriminated because of the non-uniqueness of the solution. The travel times of the rays crossing the blocks are not precise enough to determine the single contributions and multiple combinations of travel times between the two blocks are available for the solution. We will see that generally to avoid this non-uniqueness, the solution with the lowest "amplitude" is chosen.
Under-determined condition for tomographic problem naturally arises when a limitation is introduced in the possible directions of the seismic rays.
Even if, given a fixed ray-path, forward modelling is linear, in seismic tomography ray-trajectories depend on the velocity field, and this makes all tomographic problems intrinsically non-linear. Typically this non-linearity is dealt using an iterative approach.

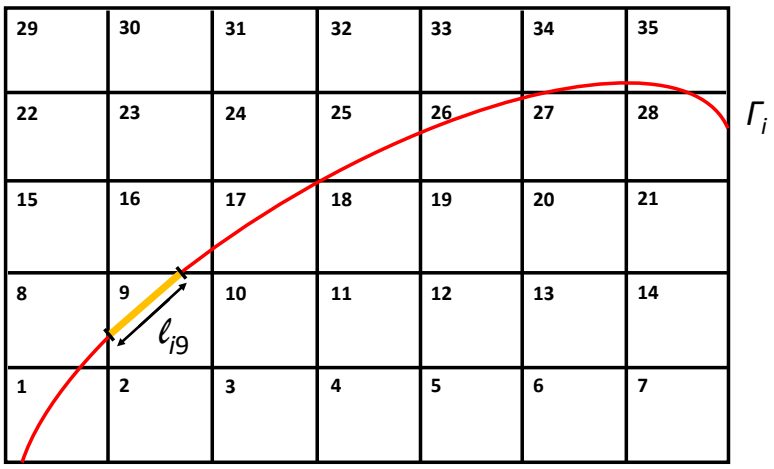

Figure 3.3: Velocity field parametrized using voxels as local functions. The length of the ray path Γ_i in the voxel 9 is l_{i9}.

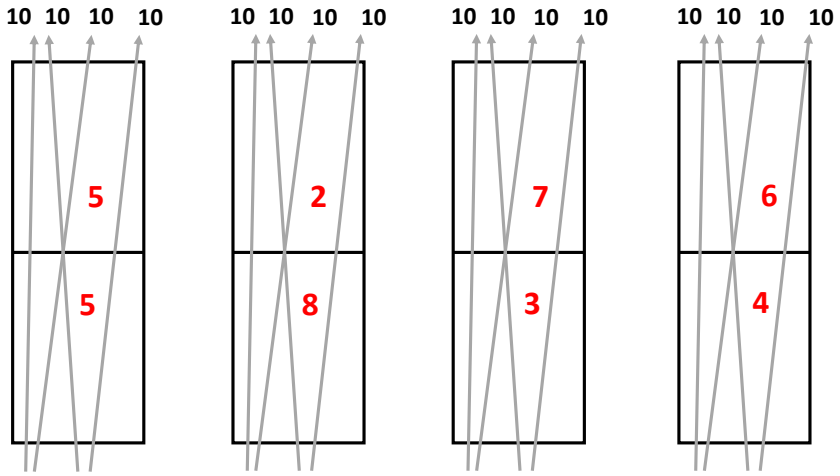

Figure 3.4: When seismic rays constrain a group of parameters without separating in an effective way the single contributions, the solutions may be not unique. Each ray measures a travel time of 10 seconds, but it is not possible to determine the travel times spent in the two blocks.

3.2 Linear and non-linear inversion methods

3.2.1 Linear least squares method

When forward modelling is linear, so $\mathbf{d} = \mathbf{Gm}$ holds, and the inverse problem is over-determined, the matrix $\mathbf{G}$ cannot be inverted (not a square matrix) and formally an exact solution does not exist. If we take into account an error term associated with the observations, we can look for a solution which minimizes the discrepancy between predictions and observations. The error generally takes into account the uncertainty in the measurements and the uncertainty in model.
We assume uncorrelated data with the same uncertainty (covariance matrix $\mathbf{C}_D$) and Gaussian distributed error $\boldsymbol{\epsilon}$. The problem is written as:

$$\mathbf{Gm} = \mathbf{d} + \boldsymbol{\epsilon} \tag{3.15}$$

We look for a solution such that:

$$\mathbf{m} = argmin\{||\mathbf{Gm} - \mathbf{d}||^2\} = argmin\{||\boldsymbol{\epsilon}(\mathbf{m})||^2\} \tag{3.16}$$

The model $\mathbf{m}$ minimizes the function $\Phi(\mathbf{m}) = ||\mathbf{Gm} - \mathbf{d}||^2 = (\mathbf{d} - \mathbf{Gm})^T \mathbf{C}_D^{-1} (\mathbf{d} - \mathbf{Gm})$:

$$\nabla\Phi(\mathbf{m}) = 2\mathbf{G}^T\mathbf{Gm} - 2\mathbf{G}^T\mathbf{d} = 0 \tag{3.17}$$

and then

$$\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} \tag{3.18}$$

which is the classic linear least squares solution, valid for linear over-determined problems with the conditions described above.

3.2.2 Regularization of inverse problems

In most of the tomographic problems the matrix $\mathbf{G}^T\mathbf{G}$ cannot be inverted due to the presence of zero-eigenvalues. This is a consequence of the mixed-determined (over- and under- determined) nature of this class of inverse problems. Reference for this section is [Nolet, 2008]
The matrix $\mathbf{G}^T\mathbf{G}$ is a real symmetric matrix $M \times M$ and an orthonormal basis of eigenvectors exists (spectral theorem) $\{\boldsymbol{v}_i\}_{i=1,\dots,M}$ for the model space with dimension M with associated non-negative eigenvalues λ_i^2.
If the eigenvectors $\boldsymbol{v}_i$ are organized in a orthogonal matrix $\mathbf{V}$, so that $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}$, and the model $\mathbf{m}$ is represented as a linear combination of the eigenvectors $\boldsymbol{v}_i$ with a set of coefficients $\mathbf{b}$, the following relationship holds:

$$\mathbf{G}^T\,\mathbf{G}\,\mathbf{m} = \mathbf{G}^T\,\mathbf{G}\,\mathbf{V}\,\mathbf{b} = \mathbf{V}\,\boldsymbol{\Lambda}^2\,\mathbf{b} = \mathbf{G}^T\,\mathbf{d} \tag{3.19}$$

where $\boldsymbol{\Lambda}^2$ is a diagonal matrix with the eigenvalues λ_i^2.
If both sides of last equation 3.19 are multiplied by $\mathbf{V}^T$ on the left, the following relationship holds:

$$\boldsymbol{\Lambda}^2\,\mathbf{b} = \mathbf{V}^T\,\mathbf{G}^T\,\mathbf{d} \tag{3.20}$$

Equation 3.20 shows that when a component of $\mathbf{b}$ is multiplied by a zero eigenvalue, it has no influence of the discrepancy between predictions of the model and observations, for each possible value b_i. This case is easily represented by a tomographic model where a block of the parametrization is not crossed by any ray, whatever is the value of the associated velocity it has no influence on the data fit and infinite possible models are allowed with the same predictions-observations discrepancy. A unique model can be proposed as solution by minimizing the norm of the model (also intended as a perturbative model with respect to a prior one), which is equivalent to constrain to zero the components b_i which are multiplied by zero eigenvalues.
If the eigenvalues are sorted is a descending order: $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_M^2$, the sub-matrix $\boldsymbol{\Lambda}_K^2$ is selected as the diagonal matrix $K \times K$ (for $K \leq M$) with only non-zero eigenvalues. The corresponding solution model, characterized by the minimum norm, is given by:

$$\mathbf{m}_K = \mathbf{V}_K\boldsymbol{\Lambda}_K^{-2}\mathbf{V}_K^T\mathbf{G}^T\mathbf{d} \tag{3.21}$$

Now the inverse matrix $\boldsymbol{\Lambda}_K^{-2}$ exists since all the eigenvalues along the diagonal are non-zero.

In equation 3.21 model $\mathbf{m}_k$ is a linear transformation of data $\mathbf{d}$, it can be proved that covariance matrix of a linear transformation $\mathbf{Ad}$ is $\mathbf{A}\mathbf{C}_d\mathbf{A}^T$, where $\mathbf{C}_d$ if the covariance matrix associated with $\mathbf{d}$. Assuming diagonal covariance matrix for data $\mathbf{C}_d$ with all equal eigenvalues σ_d, the covariance matrix of the solution model is found [Nolet, 2008]:

$$\mathbf{C}_{\mathbf{m}_k} = \sigma_d^2\,\mathbf{V}_k\boldsymbol{\Lambda}_k^{-2}\mathbf{V}_k^T \tag{3.22}$$

Analysis of the covariance matrix shows evidence supporting that to eliminate non-zero eigenvalues is not generally sufficient, because if the matrix $\mathbf{G}^T\mathbf{G}$ has small eigenvalues they will contribute by making the errors on the model extremely large (they contribute as λ_i^{-2}).

Tikhonov regularization is the most classical method to deal with mixed-determined inverse problems, where the small eigenvalues contributions to the solution create a pathological

behaviour in the errors associated with the model.
Geophysical formulation of Tikhonov regularization finds the solution model $\mathbf{m}$ as the minimum of a function Φ defined as:

$$\mathbf{m} = argmin\{\Phi\} = argmin\{||\mathbf{Gm} - \mathbf{d}||^2 + \epsilon^2||\mathbf{L}(\mathbf{m} - \mathbf{m}_{pr})||^2\} \tag{3.23}$$

where ϵ is the *regularization factor*, necessary to make the solution unique and $\mathbf{m}_{pr}$ is the prior model provided. $\mathbf{L}$ can be the identity matrix $\mathbf{I}$, if the minimization takes into account the norm of the deviation of the model with respect to the prior one, or a finite difference matrix $\mathbf{D}$, if the minimization takes into account the roughness of the model. Also other possibilities are allowed.
Minimum misfit model is obtained inverting the following relationship:

$$(\mathbf{G}^T\mathbf{G} + \epsilon^2\,\mathbf{L}^T\,\mathbf{L})\,\mathbf{m} = \mathbf{G}^T\,\mathbf{d} + \epsilon^2\,\mathbf{L}^T\,\mathbf{L}\,\mathbf{m}_{pr} \tag{3.24}$$

that can be organized as a system of equations:

$$\begin{bmatrix} \mathbf{G} \\ \epsilon^2\mathbf{L} \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{d} \\ \epsilon^2\mathbf{L}\mathbf{m}_{pr} \end{bmatrix} \tag{3.25}$$

So at the end regularization process is equivalent to adding constrains (equations) to model to make the solution unique. Singular values analysis can provide an interesting insight about the role of regularization when small eigenvalues are involved for the matrix $\mathbf{G}^T\mathbf{G}$.
Starting from equation 3.24, taking $\mathbf{L} = \mathbf{I}$ and an eigenvector of $\mathbf{G}^T\mathbf{G}$, $\mathbf{v}_k$, we have:

$$(\,\mathbf{G}^T\,\mathbf{G} + \epsilon^2\,\mathbf{I}\,)\,\mathbf{v}_k = (\,\lambda_k^2 + \epsilon^2\,)\,\mathbf{v}_k \tag{3.26}$$

When regularization is added, the system has the same eigenvectors $\mathbf{v}_k$, but eigenvalues are replaced by $\lambda_k^2 + \epsilon^2$. So the solution model $\mathbf{m}$ in equation 3.21 is given by:

$$\mathbf{m} = \mathbf{V}_K\,(\,\mathbf{\Lambda}_K^2 + \epsilon^2\,\mathbf{I}\,)^{-1}\,\mathbf{V}_K^T\,\mathbf{G}^T\,\mathbf{d} \tag{3.27}$$

The covariance matrix for the model, with the same hypothesis stated in the formulation of equation 3.22, is written as:

$$\mathbf{C}_{\mathbf{m}_k} = \sigma_d^2\,\mathbf{V}_K\,(\,\mathbf{\Lambda}_K^2 + \epsilon^2\,\mathbf{I}\,)^{-2}\,\mathbf{V}_K^T \tag{3.28}$$

It can be noticed that when constrains on the model are added to the equations as regularization, all the eigenvalues below ϵ are filtered out from the solution and from the error estimates. The choice of the regularization parameter ϵ has the effect to move the filter threshold for the eigenvalues.

3.2.3 Synthetic experiment with a linear inversion solver

To make an example of traditional linear solvers at work the following synthetic experiment is considered.
The true model, target of the inversion, is an isotropic velocity field with two heterogeneities (fast and slow) over a smooth-varying background. Stations and events are randomly placed through the 2D cartesian space. The velocity field is parametrized in cells using a regular grid of 20x20 nodes. Straight lines geometry is assumed for the rays, no tracing algorithm is performed.
True model and rays from a single event to all the stations are shown in Fig. 3.5. A Gaussian error with 50 ms as standard deviation is added to synthetic data generated using the true model.

A linear solver was written in Julia programming language for this experiment using the procedure described in section 3.2.2. The solution model is recovered as:

$$\mathbf{m} = argmin\{\Phi\} = argmin\{||\mathbf{G}\,\mathbf{m} - \mathbf{d}||^2 + \epsilon^2\,||\mathbf{m} - \mathbf{m}_{pr}||^2 + \eta^2\,||\mathbf{D}\,\mathbf{m}||^2\} \tag{3.29}$$

where ϵ and η are the *regularization* and *smoothing* parameters, $\mathbf{D}$ is a matrix which implements the finite difference scheme to the model. To add the smoothing term has the effect to take into account in the minimization also the roughness of the model.

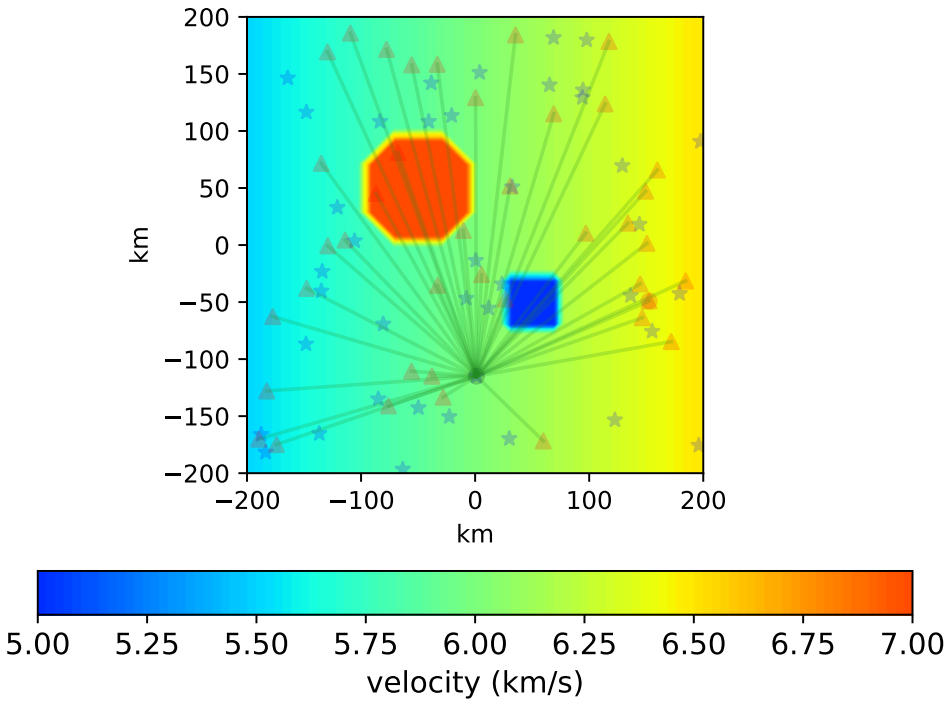

Figure 3.5: True isotropic velocity model, target of the inversion. Stations are represented as red triangles, while seismic sources are blue stars. Rays from a single event are plot.

The inversion results are presented in Fig. 3.6. Different choices for the regularization and smoothing parameters are proposed. The prior model provided to the inversion procedure is a homogeneous velocity field of 6 km/s, to increase the regularization factor ϵ has the effect to constrain stronger the minimization of the second term of the function Φ in equation 3.29, while if the smoothing factor η is increased, smoother models are obtained in the inversion procedure. The larger is the relative weight given to regularization and smoothing through ϵ and η, the lower is the relative weight given to the first term of the function Φ, which takes into account the discrepancy between observations and predictions of the model. It can be noticed that when ϵ and η increase, the RMS, used as an index of the result of the inversion increases from 60 ms (the order of the error added to synthetic data) to 220 ms.
Root mean square error (RMS) is defined as:

$$RMS = \sqrt{\frac{\sum_{i=1}^{Nrays}(t_i^{obs} - t_i^{pred})^2}{Nrays}} \tag{3.30}$$

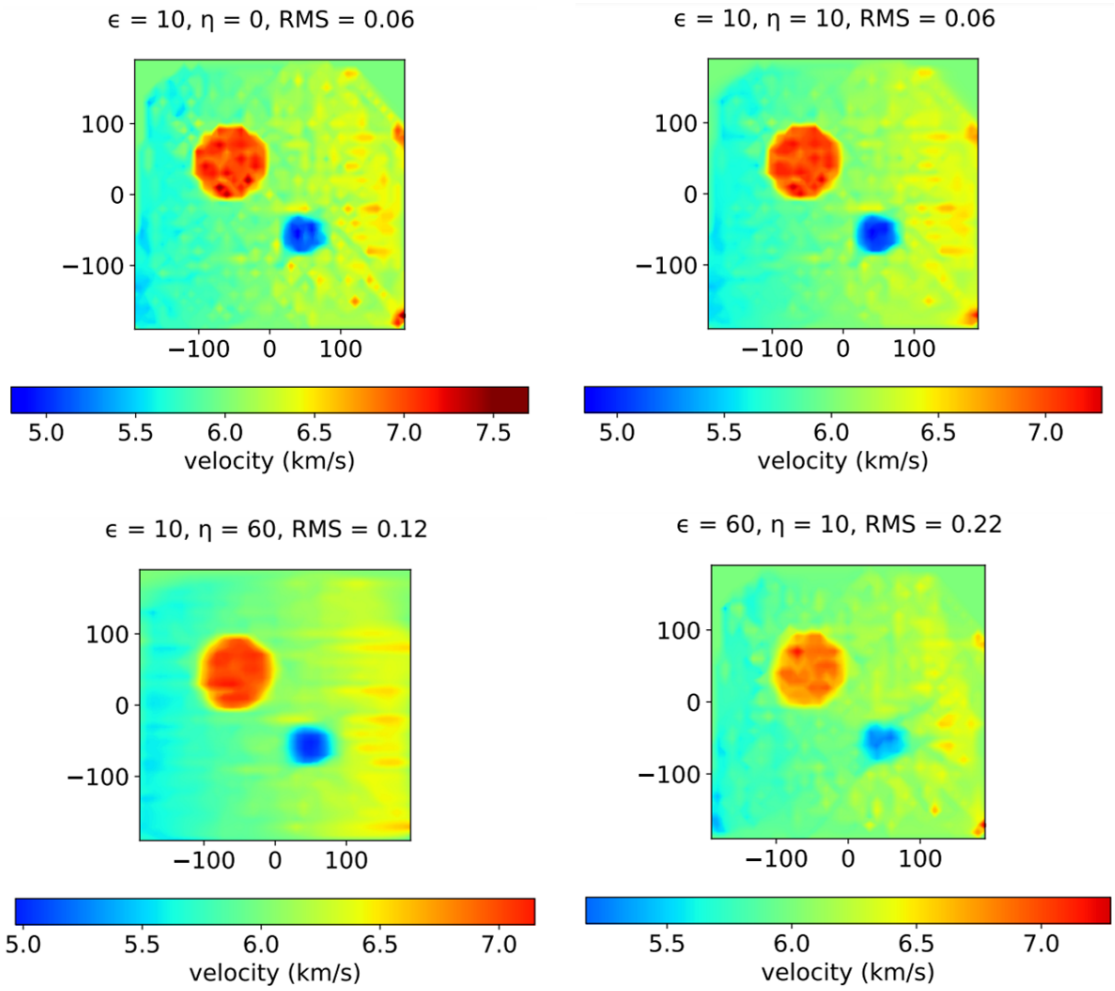

Figure 3.6: Models recovered using different regularization and smoothing parameters.

3.2.4 Non-linear optimization solvers

When forward modelling is not linear, the minimum of the misfit function is not analytically known. If the function $\mathbf{g}(\mathbf{m})$ can be locally expanded as a linear function it is possible to find the minimum proceeding with an iterative perturbative approach.
Solution model is given by:

$$\mathbf{m} = argmin\{\Phi\} = argmin\{||\,\mathbf{d} - \mathbf{g}(\mathbf{m})\,||^2\} \tag{3.31}$$

If the function $\mathbf{g}(\mathbf{m})$ is approximated around a reference model $\mathbf{m}_0$ as:

$$g_i(\mathbf{m}_0 + \boldsymbol{\delta}\mathbf{m}) \approx g_i(\mathbf{m}_0) + \frac{\partial g_i}{\partial m_j}\Big|_{\mathbf{m}_0}(m_j - m_{0j}) = g_i(\mathbf{m}_0) + G_{ij}\,(m_j - m_{0j}) \tag{3.32}$$

The inverse problem has a linear nature around $\mathbf{m}_0$ and the perturbation $\delta\mathbf{m}$ is found by minimizing the local approximation of the misfit function:

$$\Phi = ||\,\mathbf{d} - \mathbf{g}(\mathbf{m}_0) - \mathbf{G}\,\boldsymbol{\delta}\mathbf{m}\,||^2 \tag{3.33}$$

By equating to zero the partial derivatives $\partial\Phi/\partial\mathbf{m}$ we come to:

$$(\mathbf{G}^T\mathbf{G})\,\boldsymbol{\delta}\mathbf{m} = \mathbf{G}^T\,(\,\mathbf{d} - \mathbf{g}(\mathbf{m}_0)\,) \tag{3.34}$$

As seen in previous sections, for mixed-determined inverse problems matrix $\mathbf{G}^T\,\mathbf{G}$ may be singular and adding constrains to the model could be necessary.
If a regularization term is included, eq. 3.34 is written as:

$$(\,\mathbf{G}^T\mathbf{G} + \epsilon^2\,\mathbf{I}\,)\,\boldsymbol{\delta}\mathbf{m} = \mathbf{G}^T\,(\,\mathbf{d} - \mathbf{g}(\mathbf{m}_0)\,) \tag{3.35}$$

Generally multiple iterations are needed to make the algorithm converge. At each iteration $n > 1$ the n-th model is evaluated as:

$$\mathbf{m}_n = \mathbf{m}_{n-1} + \boldsymbol{\delta}\mathbf{m} \tag{3.36}$$

and the regularization factor ϵ can be adjusted.
Levenberg-Marquardt approach estimates the solution model proceeding for multiple iterations where the minimum-misfit model at each iteration is evaluated as:

$$(\,\mathbf{G}^T\mathbf{G} + \epsilon^2\,\mathbf{diag}(\mathbf{G}^T\mathbf{G})\,)\,\boldsymbol{\delta}\mathbf{m} = \mathbf{G}^T\,(\,\mathbf{d} - \mathbf{g}(\mathbf{m}_{n-1})\,) \tag{3.37}$$

This approach scales the weighting of the regularization to $\mathbf{G}^T\mathbf{G}$ to avoid too slow convergence.

Attention should be paid to misfit functions characterized by the presence of local minima. Generally the starting model $\mathbf{m}_0$ is not related to any prior knowledge, but in situations where multiple minima exist the selection of the starting model could be crucial in order to avoid to converge to a local minimum of the misfit function.

3.3 Methods based on Markov Chains Monte Carlo technique

Traditional approaches in seismic tomography based on non-linear optimization solvers have encountered several limitations due to the ill-conditioned nature of tomographic inverse problems. The regularization methods described in section 3.2.2 deal with the under-determined nature of these problems introducing the additional constrains described, but, as pointed previously, a level of arbitrariness is introduced in the selection of the regularization and smoothing factors and with the choice for the density of the grid. Different results are obtained with different choices for these parameters and uncertainty maps could be not accessible in non-linear problems where solution is investigated using multiple iterations for the linearization of the forward modelling until convergence is found. An additional problem is the computation of the derivatives of predictions with respect to the parameters of the model, necessary for the perturbative approach, but not always an easy task.
During the last decades the increase of computational resources and the realization among the geophysical community that linear approximation was not convenient or not even possible for a large class of interesting problems led to a rise in popularity of stochastic methods.
Inverse problems have a natural formalization based on a probabilistic framework, exceptionally described in [Tarantola, 2004] through the Bayes' theorem. Given a prior knowledge for the model $\mathbf{m}$, represented by a prior probability $p(\mathbf{m})$ and a set of observations $\mathbf{d}$ with an associated likelihood function $p(\mathbf{d}|\mathbf{m})$, representing the probability to have the observations $\mathbf{d}$ under assumption that model $\mathbf{m}$ is true, the posterior probability for the model $\mathbf{m}$, on the light of data $\mathbf{d}$, is expressed as:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m}) \times p(\mathbf{m})}{p(\mathbf{d})} \tag{3.38}$$

where $p(\mathbf{d}) = \int_M p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m})\, \mathbf{dm}$ is not a function of the parameters.
Bayes's theorem allows to update the degree of confidence on a model on the basis of the outcome of the observations. This relationship is based on the subjective choice for the prior probability function, where no golden rules exist. Typical choices for prior distributions are uniform probability distributions, to represent the lack of knowledge *a priori*, Gaussian distributions around models which are results of precedent works, Jeffrey's priors when functional invariance is required for the distribution, etc.
If the uncertainty in the forward modelling process is neglected (generally with respect to other sources of uncertainty) and the prior distribution assumed for the data is Gaussian, centered around the predictions, the likelihood function has this form:

$$p(\mathbf{d}|\mathbf{m}) \propto exp\left(-\frac{1}{2}(\,\mathbf{g(m)} - \mathbf{d}\,)^T\, \mathbf{C}_d^{-1}\,(\,\mathbf{g(m)} - \mathbf{d}\,)\right) \tag{3.39}$$

where $\mathbf{g(m)}$ is the set of predictions through the forward modelling and $\mathbf{C}_d$ is the covariance matrix for the data.
If a Gaussian prior centered around a model $\mathbf{m}_{priori}$ is assumed:

$$p(\mathbf{m}) \propto exp\left(-\frac{1}{2}(\,\mathbf{m} - \mathbf{m}_{priori}\,)^T\, \mathbf{C}_m^{-1}\,(\,\mathbf{m} - \mathbf{m}_{priori}\,)\right) \tag{3.40}$$

where $\mathbf{C}_m$ is the covariance matrix for the model, then the posterior probability distribution is expressed as:

$$p(\mathbf{m}|\mathbf{d}) \propto exp\left(-\frac{1}{2}(\,\mathbf{g(m)-d}\,)^T\,\mathbf{C}_d^{-1}\,(\,\mathbf{g(m)-d}\,) - \frac{1}{2}(\,\mathbf{m}-\mathbf{m}_{pr}\,)^T\,\mathbf{C}_m^{-1}\,(\,\mathbf{m}-\mathbf{m}_{pr}\,)\right) \tag{3.41}$$

where

$$2S(\mathbf{m}) = (\,\mathbf{g(m)-d}\,)^T\,\mathbf{C}_d^{-1}\,(\,\mathbf{g(m)-d}\,) + (\,\mathbf{m}-\mathbf{m}_{pr}\,)^T\,\mathbf{C}_m^{-1}\,(\,\mathbf{m}-\mathbf{m}_{pr}\,) \tag{3.42}$$

is the *misfit* function.
The model which maximizes the posterior is the one which minimizes S($\mathbf{m}$).

If forward modelling is linear, $\mathbf{d} = \mathbf{Gm}$, then the posterior probability function is Gaussian, with the form:

$$p(\mathbf{m}|\mathbf{d}) \propto exp\left(-\frac{1}{2}(\,\mathbf{Gm-d}\,)^T\,\mathbf{C}_d^{-1}\,(\,\mathbf{Gm-d}\,) - \frac{1}{2}(\,\mathbf{m}-\mathbf{m}_{pr}\,)^T\,\mathbf{C}_m^{-1}\,(\,\mathbf{m}-\mathbf{m}_{pr}\,)\right) \tag{3.43}$$

In this case properties of the distributions are analytically known; the model with the maximum probability, which is also the expectation value is given by:

$$\begin{aligned} <\mathbf{m}> &= (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1})^{-1}(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d} + \mathbf{C}_m^{-1}\mathbf{m}_{pr}) \\ &= \mathbf{m}_{pr} + (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1})^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{Gm}_{pr}) \\ &= \mathbf{m}_{pr} + \mathbf{C}_m\mathbf{G}^T(\mathbf{GC}_m\mathbf{G}^T + \mathbf{C}_d)^{-1}(\mathbf{d} - \mathbf{Gm}_{pr}) \end{aligned} \tag{3.44}$$

with a covariance matrix:

$$\mathbf{C}_{<m>} = (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1})^{-1} \tag{3.45}$$

We notice that, if the covariance matrix for the model is assumed to be diagonal with all equal eigenvalues σ_m and if no priori information is provided, then this is equivalent to assume that $\sigma_m \to \infty$ and $\mathbf{C}_m^{-1} \to 0$. If the covariance matrix for data is also diagonal with all equal eigenvalues, then equation 3.44(1) is reduced to ordinary linear least squares solution 3.18. This justifies the assumptions made in section 3.2.1.

If covariance matrices for data and model are diagonal (typical assumption in tomography when uncertainties are not well known), the misfit function 3.42 is reduced to:

$$2S(\mathbf{m}) = \frac{1}{\sigma_d^2}\left((\mathbf{Gm}-\mathbf{d})^T(\mathbf{Gm}-\mathbf{d}) + \frac{\sigma_d^2}{\sigma_m^2}(\mathbf{m}-\mathbf{m}_{pr})^T(\mathbf{m}-\mathbf{m}_{pr})\right) \tag{3.46}$$

Calling $\epsilon = \sigma_d/\sigma_m$, minimum is found as:

$$(\,\mathbf{G}^T\mathbf{G} + \epsilon^2\,\mathbf{I}\,)\mathbf{m} = \mathbf{G}^T\mathbf{d} + \epsilon^2\,\mathbf{Im}_{pr} \tag{3.47}$$

or, as a system of equations:

$$\begin{bmatrix} \mathbf{G} \\ \epsilon^2 \mathbf{I} \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{d} \\ \epsilon^2 \mathbf{m}_{pr} \end{bmatrix} \tag{3.48}$$

Equation 3.48 recalls the result of least squares method with regularization and allows to interpret the regularization factor in terms of uncertainties on model and data. As can be expected, a larger level of uncertainty on data with respect to the prior model raises ϵ, giving more weight to the second term of the misfit function, while if data are affected by a lower level of uncertainty it is the deviation between observations and predictions to have a higher weight.
When forward modelling is not linear, $\mathbf{d} = \mathbf{g(m)}$, moments of the posterior distribution are not analytically known and Bayesian estimators take the role for the evaluations. Bayesian estimators compute the moments of the distribution using an ensemble of samples from it and approximating the integral definitions in terms of discrete sums.
Samples from a probability distribution whose analytical expression is known can be generated using methods which implement Markov Chains Monte Carlo techniques. These methods are based on the generation of random numbers and they can be used to explore the model space performing a random sampling of the posterior probability distribution $p(\mathbf{m}|\mathbf{d})$. A systematic exploration becomes rapidly extremely inefficient when the dimension of the model space increases (high-dimensional spaces tend to be *dramatically empty*); two main problems are encountered in large-dimensional space sampling: 1) locating regions of significant probability, 2) sampling the whole regions densely enough. It is possible to implement algorithms which perform a random walk avoiding to leave these regions of significant probability once they have been found.

3.3.1 Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm uses a Markov Chain Monte Carlo method developed by [Metropolis & Ulam, 1949], [Metropolis et al., 1953] and [Hastings, 1970]. It is random and each step of the chain depends only on the previous step.
The algorithm performs a random walk modified according to probabilistic rules. Given a model $\mathbf{m}$, a move to a new model $\mathbf{m'}$ is proposed as sample from the posterior probability $p(\mathbf{m}|\mathbf{d})$ according to a proposal distribution $q(\mathbf{m'}|\mathbf{m})$.
The acceptance ratio α, used to evaluate if a new model is kept, is computed as:

$$\alpha = min \left[1, \frac{p(\mathbf{m'}|\mathbf{d})\, q(\mathbf{m}|\mathbf{m'})}{p(\mathbf{m}|\mathbf{d})\, q(\mathbf{m'}|\mathbf{m})} \right] \tag{3.49}$$

According to the Bayes' theorem this ratio can be replaced in a more convenient way:

$$\alpha = min \left[1, \frac{p(\mathbf{d}|\mathbf{m'})\ \ p(\mathbf{m'})\ \ q(\mathbf{m}|\mathbf{m'})}{p(\mathbf{d}|\mathbf{m})\ \ p(\mathbf{m})\ \ q(\mathbf{m'}|\mathbf{m})} \right] \tag{3.50}$$

The model $\mathbf{m'}$ is added to the chain if a generated random number r, uniform in [0,1] is lower than α. Otherwise the precedent model $\mathbf{m}$ is added again to the chain. The algorithm provides an ensemble of models which converge in distribution to the posterior probability. The convergence speed can be influenced by the choice of the proposal distribution and by

the first random model generated (initial step of the chain). There are no diagnostic methods to test if convergence is achieved, but only if it is not. General procedure is to collect a sufficiently large ensemble of models and discharge an initial sequence of models, called *burn in*, where samples are biased due to the starting random guess. In large-dimension model spaces, since the perturbations have low magnitude, it is inevitable to have correlations between adjacent samples of the chain. In order to have an ensemble of independent models sampled from the posterior, a sub-chain is extracted from the main one, keeping only one model for each interval in the chain of specified length which has to be decided on the basis of the precedent choices (e.g. the proposal).
The selection of the perturbation size of the proposal (example: standard deviation of a Gaussian distribution) has significant consequences on the efficiency of the sampling algorithm, influencing the fraction of models accepted in the sampling procedure and the efficiency of the algorithm in the exploration of the space.
In Fig. 3.7 the fraction of accepted models is represented as a function of the number of iterations for four different choices of the perturbation size. The larger is the perturbation size, the lower is the accepted fraction of models, but with too small perturbations the convergence rate may be affected. Another aspect to take into account is the exploration efficiency of the Metropolis-Hastings algorithm, when the probability distribution to sample is multi-modal, and different modes are highly separated, a small selection for the perturbation size could restrict the sampling procedure only to the "widest" mode, because steps toward the other modes are generally rejected (Fig. 3.8). Sampling with Metropolis-Hastings algorithm should be performed using a perturbation size which makes a good compromise between the convergence rate and the exploration efficiency of the space. In alternative if there is an interest in sampling a particular mode, the selected starting model can be set approximately near to the mode to sample. Another possibility is to start multiple Markov Chains with different starting models and then, once the convergence has been achieved, merge the results.

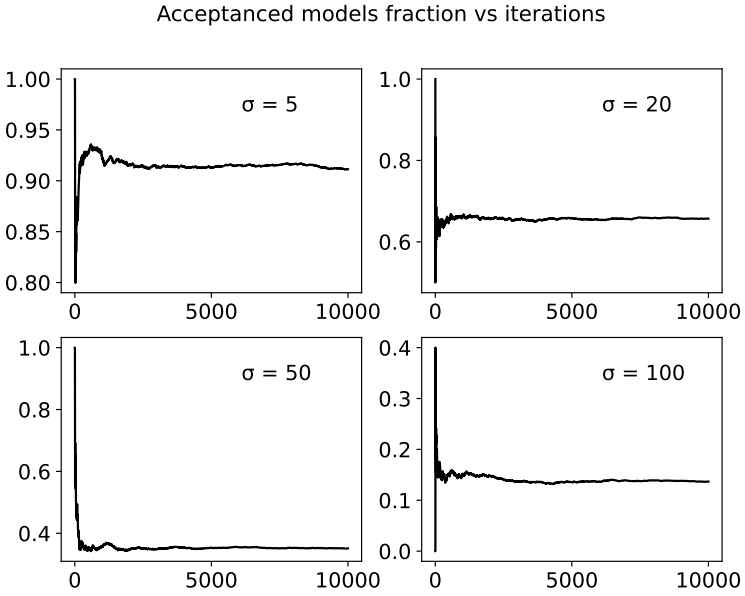

Figure 3.7: Fraction of accepted models as a function of iterations performed by the Markov Chain for different selections of perturbation size of the proposal distribution.

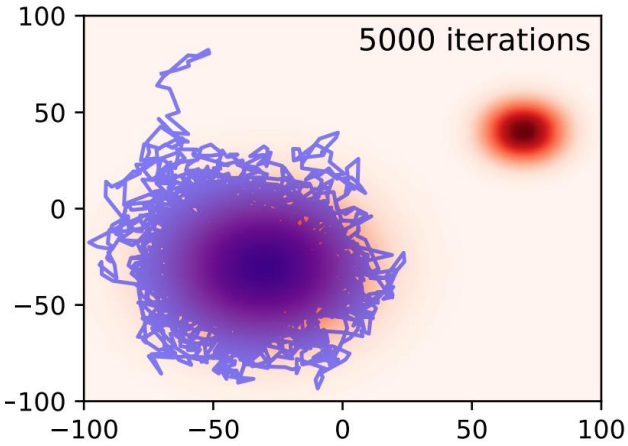

Figure 3.8: The target distribution is multi-modal. A perturbation size = 5 is selected, the narrower mode of the distribution is not sampled with this choice.

3.3.2 Reversible jump Markov Chain Monte Carlo

Metropolis-Hastings algorithm works with a fixed parametrization, which is not always available, for example when the complexity of the fields involved in the modelling is not known a priori. Metropolis-Hastings method can be generalized to consider the number of parameters used to represent the fields as a parameter itself (of the posterior), to investigate in the inversion procedure, we call this generalization *trans-dimensional Metropolis-Hastings*, or *reversible jump Markov Chain Monte Carlo* method. In this section the algorithm is described in its application to 2D cartesian space.
An irregular mesh is considered for this process; nodes are placed through a 2-dimensional space and an interpolation algorithm is implemented to represent the continuous field.
The algorithm performs a random walk starting from a initial model and selecting a move within five possibilities for each step:

1. **change the value of a node**: a random node is selected among those available and the amplitude of the quantity assigned is perturbed according to the selected proposal probability density function (Gaussian). If σ is the standard deviation of the proposal, the new value is given by: $v_n^{'} = v_n + \sigma \times u$, where $v_n^{'}$ is the new generated value, v_n is the old value and u is a random number distributed according to a standard Gaussian.

2. **change the position of a node**: the coordinates of a randomly chosen node are perturbed using two random numbers distributed according to a Gaussian proposal. $x_n^{'} = x_n + \sigma \times u_1$, $y_n^{'} = y_n + \sigma \times u_2$.

3. **create a new node (birth step)**: a node is generated in the space. The coordinates are randomly selected according to an uniform distribution and the value assigned is the result of the interpolation algorithm with the precedent nodes and a Gaussian perturbation.

4. **delete an existing node (death step)**: a randomly selected node is removed from the parametrization.

5. **change the noise parameter**: the error parameter associated with the computation of the likelihood function is perturbed with a Gaussian proposal. This error takes into account data and model uncertainties.

Once the new model has been generated, the algorithm evaluates if it is kept or not. Accounting also the possible change in dimension, acceptance ratio can be computed as:

$$\alpha = min\left[1, \frac{p(\mathbf{d}|\mathbf{m'})\ p(\mathbf{m'})\ q(\mathbf{m}|\mathbf{m'})}{p(\mathbf{d}|\mathbf{m})\ p(\mathbf{m})\ q(\mathbf{m'}|\mathbf{m})}|J|\right] \tag{3.51}$$

where J is the Jacobian of the step-transformation [Green, 1995].

Assuming Gaussian errors on data $\mathbf{d}$ and neglecting model predictions uncertainty ($\delta(\mathbf{d} - g(\mathbf{m}))$), the likelihood function can be expressed as [Tarantola, 2004]:

$$p(\mathbf{d}|\mathbf{m}) \propto exp\left(-\frac{\Phi(\mathbf{m})}{2}\right) \tag{3.52}$$

where $\Phi(m)$, the misfit function is:

$$\Phi(\mathbf{m}) = (\mathbf{g}(\mathbf{m}) - \mathbf{d})^t \, \mathbf{C}_d^{-1} \, (\mathbf{g}(\mathbf{m}) - \mathbf{d}) \tag{3.53}$$

For uncorrelated data with noise σ_d:

$$\Phi(\mathbf{m}) = \frac{\|\mathbf{g}(\mathbf{m}) - \mathbf{d}\|^2}{\sigma_d^2} \tag{3.54}$$

In the Bayesian inversion, the prior knowledge for the model can be parametrized using a probability distribution p($\mathbf{m}$). In this implementation, assuming lack of knowledge a priori, an uniform prior distribution in a fixed range for all the parameters except the number of nodes in considered. The prior distribution for the number of nodes is a log-uniform distribution, which is the relevant Jeffreys prior:

$$p(n) = \left(n \, ln \left(\frac{n_{max}}{n_{min}} \right) \right)^{-1} \tag{3.55}$$

where n_{max} and n_{min} are the maximum and minimum numbers of allowed nodes.
The probability to have a proposed model $\mathbf{m}$' accepted is given by 3.51. For the problem here considered, the determinant of the Jacobian matrix is always 1 ([Green, 1995], [Sambridge et al., 2006]).
Assuming Gaussian perturbative proposals for change and move steps, which are symmetrical, since there is not change in the dimension of the model space, acceptance ratios are computed as eq. 3.56, 3.57.

$$\alpha_{change} = min \left[1, \frac{p(\mathbf{d}|\mathbf{m'})}{p(\mathbf{d}|\mathbf{m})} \right] \tag{3.56}$$

$$\alpha_{move} = \; min \left[1, \frac{p(\mathbf{d}|\mathbf{m'})}{p(\mathbf{d}|\mathbf{m})} \right] \tag{3.57}$$

For the birth and death steps: proposal to generate nodes and associate values are uniform and Gaussian, clearly in this case the forward and inverse steps are not equivalent. Following the derivation of [Bodin and Sambridge, 2009] with our assumptions for the priors, the acceptance ratios for the steps proposed are computed ([Byrnes & Bezada, 2020]; α computation for attenuation tomography problem, here the priors are slightly different) as eq. 3.58, 3.59.

$$\alpha_{birth} = min \left[1, \frac{n}{n+1} \frac{\sigma\sqrt{2\pi}}{\Delta v} exp \left(\frac{1}{2} \left(\frac{\mathbf{v'} - \mathbf{v}_0}{\sigma} \right)^2 \right) \frac{p(\mathbf{d}|\mathbf{m'})}{p(\mathbf{d}|\mathbf{m})} \right] \tag{3.58}$$

$$\alpha_{death} = min \left[1, \frac{n}{n-1} \frac{\Delta v}{\sqrt{2\pi}\sigma} exp \left(-\frac{1}{2} \left(\frac{\mathbf{v} - \mathbf{v}_0}{\sigma} \right)^2 \right) \frac{p(\mathbf{d}|\mathbf{m'})}{p(\mathbf{d}|\mathbf{m})} \right] \tag{3.59}$$

where $\mathbf{v}'$ is the new proposed value associated with the generated node and $\mathbf{v}$ is the same, but for the inverse perturbation in death step, while $\mathbf{v}_0$ is the value resulting from the interpolation with the nodes precedent to the steps. The parameter σ is the standard deviation

of the Gaussian proposal and Δv is the interval of values allowed for the quantity assigned to the node.
In the end, assuming a Gaussian proposal to perturb the value associated with the error parameter, acceptance ratio is computed as eq. 3.60.

$$\alpha_{error} = min\left[1, \frac{p(\mathbf{d}|\mathbf{m'})}{p(\mathbf{d}|\mathbf{m})}\right] \tag{3.60}$$

where the error parameter on the likelihood function is different between numerator and denominator.

Chapter 4

Reversible jump MCMC implementation

4.1 Structure of AnisoTOMO code

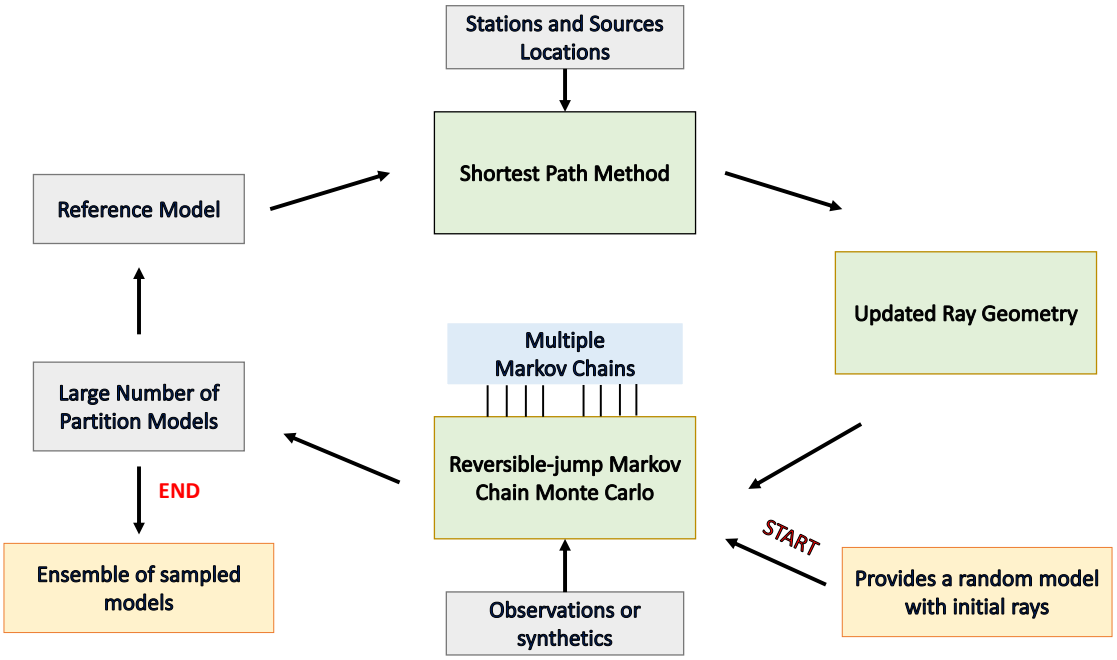

Figure 4.1: Non-linear anisotropic tomography code with rj-MCMC algorithm.

The first implementation of the reversible jump MCMC algorithm used in this project was written by Joseph Byrnes in Julia programming language for two-dimensional cartesian geometry. This version of the code was modified in parametrization and interpolation scheme, then extended to 2-dimensional spherical geometry with the possibility to trace seismic rays using the shortest path method and multiple chains implementation. In the following chapters the main steps in the code development will be discussed with a wide usage of synthetic experiments to show the performance of the inversion algorithm in multiple contexts.
Travel time seismic tomography is intrinsically non-linear, due to the dependence of the ray paths from the velocity field. This non-linearity increases its weight especially when local earthquakes data are used for the inversion procedure. It can be taken into account by mak-

ing repeated updates of the ray paths using reference models for the velocity fields of the seismic waves (Section 5.4).
The inversion procedure developed in this project can be summarized in two nested cycles: initially a random or reference model is provided to the sampling algorithm, which iteratively produces models according to probabilistic rules, then the ensemble of models sampled is used to create a reference model and to compute the shortest paths between sources and receivers. Once the ray paths have been updated, the code executes again the reversible jump loop. This continues for a specified number of nested loops.

4.2 First implementation in 2D cartesian geometry

Original version of the inversion code was characterized by single Markov Chains implementing the rj-MCMC algorithm. Given a synthetic data set of travel times with an added Gaussian noise, an ensemble of models is produced starting from a random one and proceeding according to the probabilistic criteria described in section 3.3.2.

Assuming hexagonal symmetry for the elastic tensor in the upper mantle, which is established to be a good approximation [Becker et al., 2006], the anisotropic velocity field for P-waves is a periodic function of 2α and 4α, where α is the angle between the ray-segment considered along the path and the local hexagonal symmetry axis [Backus, 1965] [Thomsen, 1986]. The 4α-periodic term is generally an order of magnitude lower than the one periodic in 2α, which allows to approximate the dependence of the velocity as:

$$v = \bar{v}\,(1 \pm f cos(2\alpha)) \tag{4.1}$$

where v is the anisotropic magnitude of the velocity along the ray-segment, $\bar{v}$ is the isotropic velocity, f is the anisotropic magnitude fraction, defined as:

$$f = \frac{v_{max} - v_{min}}{v_{max} + v_{min}} \tag{4.2}$$

The sign of f is extrapolated in equation 4.1, a positive value is associated with seismically fast velocity axes (e.g. olivine, A-type fabric, [Karato et al., 2008]), while negative values can be associated with aligned fractures.
If θ and γ are the the ray and symmetry axis elevation angles and ϕ and ψ are the ray and symmetry axis azimuth angles (convention: Fig. 4.2), the equation 4.1 is written as:

$$v = \bar{v}\,[\,1 \pm f\,(\,2\,(cos(\theta)cos(\gamma)cos(\phi - \psi) + sin(\theta)sin(\gamma))^2 - 1)\,] \tag{4.3}$$

When geometrical space is reduced to two dimensions, and only elevation or azimuth are resolved, we can set $\phi = \psi = 0$ for inversions involving only elevation, or $\theta = \gamma = 0$ for inversions involving only azimuth. In both cases the travel time computation formula is reduced to:

$$v = \bar{v}\,[\,1 \pm f\,(\,2\,(cos(\eta - \omega))^2 - 1)\,] \tag{4.4}$$

where η and ω angles are θ and γ (elevation), or ϕ and ψ (azimuth).
Forward modelling implemented in the inversion code works only with positive values for anisotropic magnitude fraction f, discharging the minus sign. This is justified since the upper mantle anisotropic structure is dominated by olivine lattice preferred orientation [VanderBeek & Faccenda, 2021].

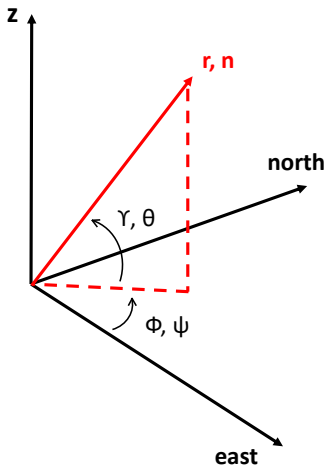

Figure 4.2: Convention for orientation angles. Ray segment or symmetry axis is represented by the directional vector **r** or **n**. Azimuth angles are measured counter-clockwise-positive from east, while elevation angles are measured counter-clockwise-positive from north-east plane.

Originally, the interpolation scheme adopted in the inversion code was the inverse distance weighted mean interpolation, where the value of a continuous field in an arbitrary point of the space, given a set of value-specified nodes $\{\mathbf{u}_i\}_{i=1,..,N}$ at positions $\{\mathbf{x}_i\}_{i=1,..,N}$ is given by:

$$u(\mathbf{x}) = \begin{cases} \frac{\sum_{i=1}^{N} w_i(\mathbf{x})\, u_i}{\sum_{i=1}^{N} w_i}, & if\ d(\mathbf{x}, \mathbf{x}_i) \neq 0\ for\ all\ i \\ u_i & if\ d(\mathbf{x}, \mathbf{x}_i) = 0\ for\ some\ i \end{cases} \tag{4.5}$$

where

$$w_i = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^p} \tag{4.6}$$

where $d(\mathbf{x}, \mathbf{x}_i)^p$ is the distance function with p set to 2.

Inverse distance weighted mean interpolation was substituted with the Voronoi cells interpolation scheme, a traditional choice in seismic tomography, especially when discontinuous velocity fields are resolved.
In Voronoi cells parametrization, the value of the continuous resolved field in an arbitrary point is the value of the closest node. This means that the field is divided in non-overlapping subsets with homogeneous values inside each region and discontinuities at the edges, where

the set of these regions constitutes a Voronoi diagram. In Fig. 4.3 a comparison between these two interpolation schemes is shown.
At this point the parametrization of the anisotropic velocity fields consists in a set of nodes for isotropic velocity, anisotropic magnitude fraction and anisotropic orientation angle, where also the positions of the nodes, the number of nodes and noise are part of the parametrization in the inversion method.

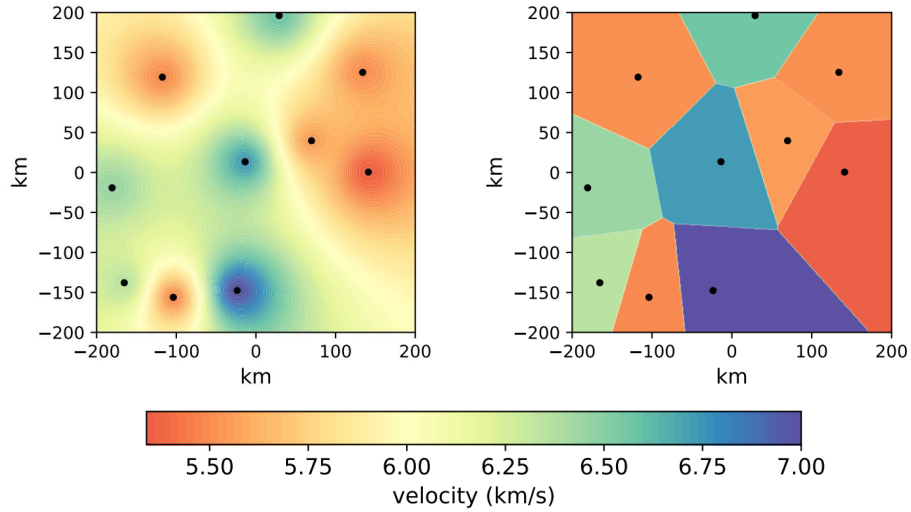

Figure 4.3: left: inverse distance weighted mean interpolation scheme, right: Voronoi cells diagram interpolation. Positions and values of the generated nodes are kept identical in the comparison.

If an ensemble of models has been generated by Markov Chains and the monitoring of some diagnostic quantity, like the RMS (synthetic data and inversion results, see eq. 3.30), has shown evidence to assume that the chains have converged, the burn-in stage can be discharged. All the sampled models are characterized by a collection of nodes for the fields used in the modelling and a reference model can be synthesized using a space-average of the ensemble.
The expectation value of the posterior with distribution p($\mathbf{m}|\mathbf{d}$) is by definition:

$$< \mathbf{m} >= \int_M \mathbf{m}\, p(\mathbf{m}|d)\, d\mathbf{m} \tag{4.7}$$

In the large limit for the ensemble size, the Central Limit Theorem provides the theoretical sustain to the choice of using an estimator to evaluate this moment.

Theorem: Given a sequence of independent random variables X_i, each from a distribution with mean μ_i and variance σ_i^2, then the variable $S = \sum_{i=1}^{n} X_i$ in the limit for $n \to \infty$ converges in distribution to a Normal distribution, with expectation value $E[S] = \sum_{i=1}^{n} E[X_i]$ and variance $Var[S] = \sum_{i=1}^{n} Var[X_i]$

In this context $X_i = \mathbf{m}_i$, and all the variables are from the same distribution, which is the posterior p($\mathbf{m}|\mathbf{d}$); in the large sample limit, the sample mean $\overline{\mathbf{m}} = \sum_{i=1}^{n} \mathbf{m}_i/n$ converges

to the expectation value of $\mathbf{m}$ and it converges in distribution to a Gaussian. According to the Law of Large Numbers the estimator $\overline{\mathbf{m}}$ is also a consistent estimator.
Also standard deviation of the distribution can be computed using an estimator on the ensemble.

Attention should be paid when average operator is applied to angle-like fields, as the azimuth and the elevation. Since the anisotropic velocity is a periodic function in twice the angle between the seismic ray and the symmetry axis of the hexagonal symmetry, two orientations ϕ and $\phi + \pi$ for ray or axis are completely equivalent, but if the standard mean is computed the resulting angle is $\phi + \pi/2$, clearly a misleading conclusion since periodicity is not included. The Julia package *DirectionalStatistics* provides the statistical tools to work with periodic quantities.

4.3 Cartesian synthetic experiments

In order to test the inversion algorithm, some synthetic experiments will be considered. The first group of tests submitted to the method use straight line geometries for rays.

4.3.1 Isotropic data inversion

The first synthetic experiment considered is the inversion of travel times created using an isotropic velocity model (Fig. 4.4). This model is characterized by a smooth-varying background and two discontinuous heterogeneities. Gaussian noise with standard deviation of 50 ms is added to synthetic data. Twenty-four stations (red triangles) and events are placed through the 2D space with a regular distribution. The rays for a single station are represented using green lines. No ray-tracing is applied and only isotropic velocity is parametrized. This experiment is inspired by [Ryberg, 2018]
As setup of the inversion procedure, 500000 iterations are considered and twenty-eight Markov Chains are run in parallel to increase the exploration of the model space. The burn-in phase is set to 300000 (selected using the trends of the RMS for each chain, Fig. 4.5). Only a model every 1000 sampled is kept, in order to avoid the correlations in sequential samplings introduced by the algorithm. Each sampled model is a Voronoi diagram (Fig. 4.6).
Using Voronoi cells interpolation, the recovery of the true model is poor if single samples are considered. In order to have a smoother model, which better recovers the target of the inversion, the space-average of the ensemble of models sampled by the twenty-eight Markov Chains is computed (similar approach with a regularized solver was used in [Fang et al., 2019]). The reference model, resulting from the inversion procedure is represented in Fig. 4.7. Both the smooth and the discontinuous variations are well recovered in magnitude and geometry.
Using statistical estimators on the ensemble it is also possible to compute the standard deviation (Fig. 4.8).
This experiment shows how the reversible jump is able to invert synthetic data in the context of isotropic forward modelling, using this approach no prior knowledge is provided to the inversion procedure, and even if it were, there is not any constrain on the deviation of the actual model from the prior one during the sampling. Some features of the inversion results can be easily recognized as artifacts of the method evaluating uncertainty maps, like the residual Voronoi-cell shapes in Fig. 4.7, which are high-uncertainty regions in Fig. 4.8.

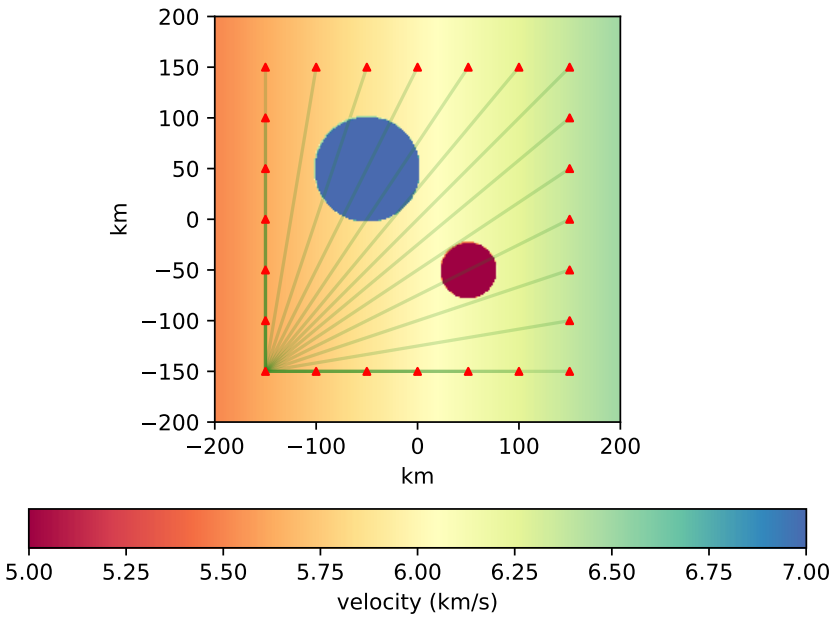

Figure 4.4: True model, used to compute synthetic data. This is the target of the inversion.

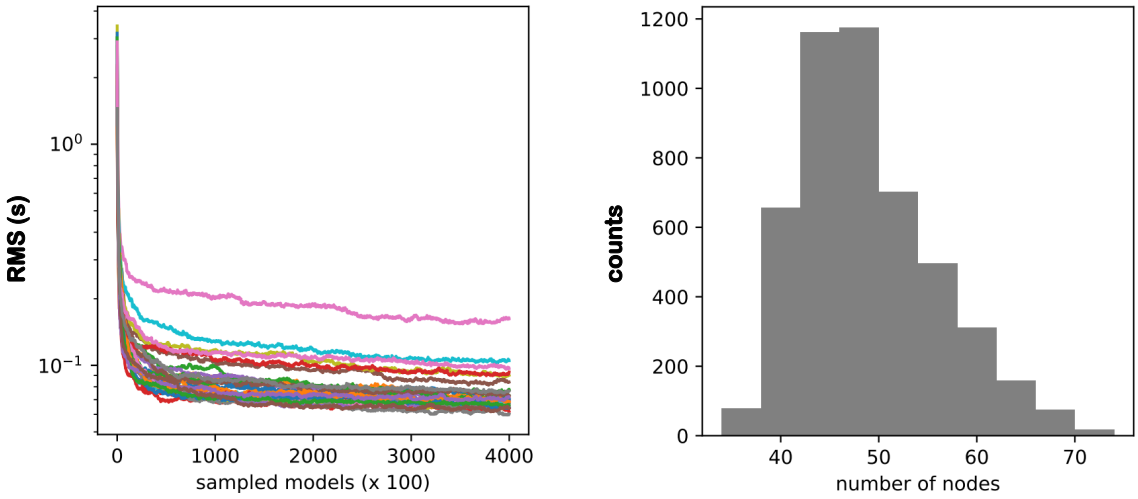

Figure 4.5: Left: trend of the RMS [s] for each chain run in parallel. Right: distribution of nodes from the ensemble.

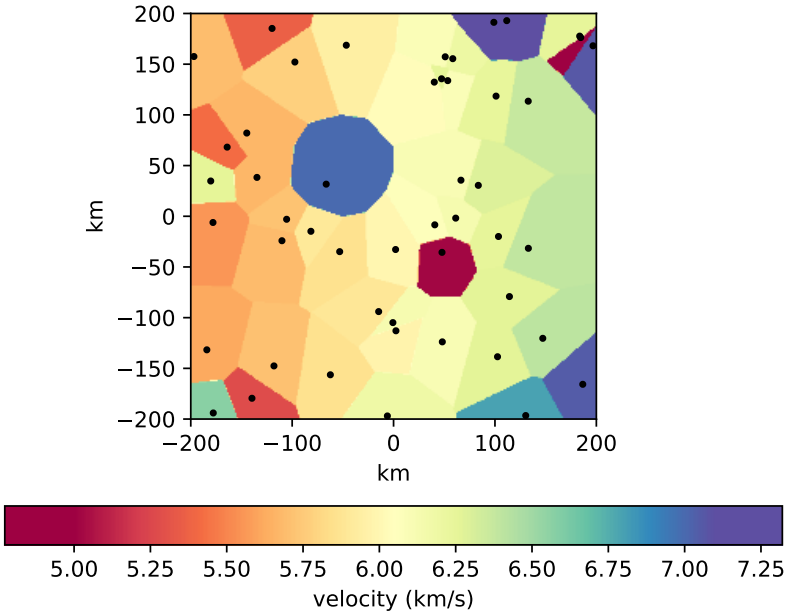

Figure 4.6: Interpolated model sampled from the posterior probability density.

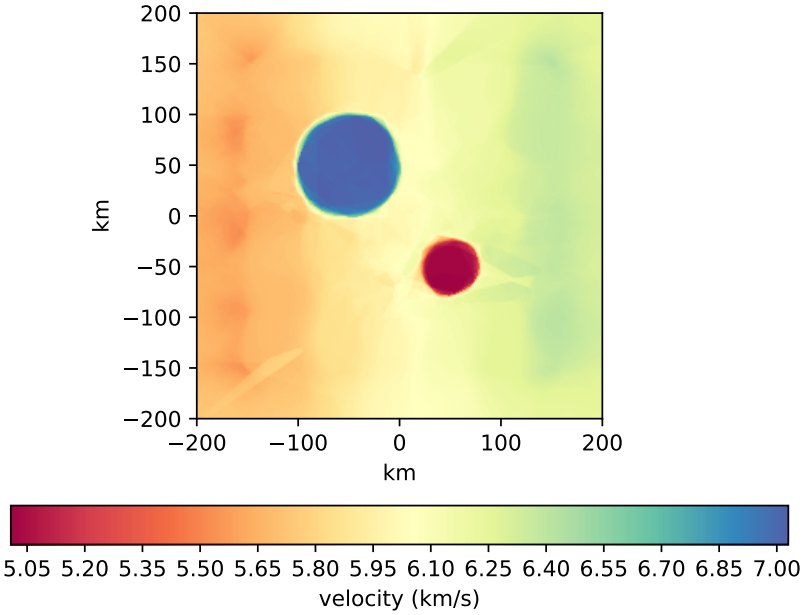

Figure 4.7: Model recovered using the inversion procedure.

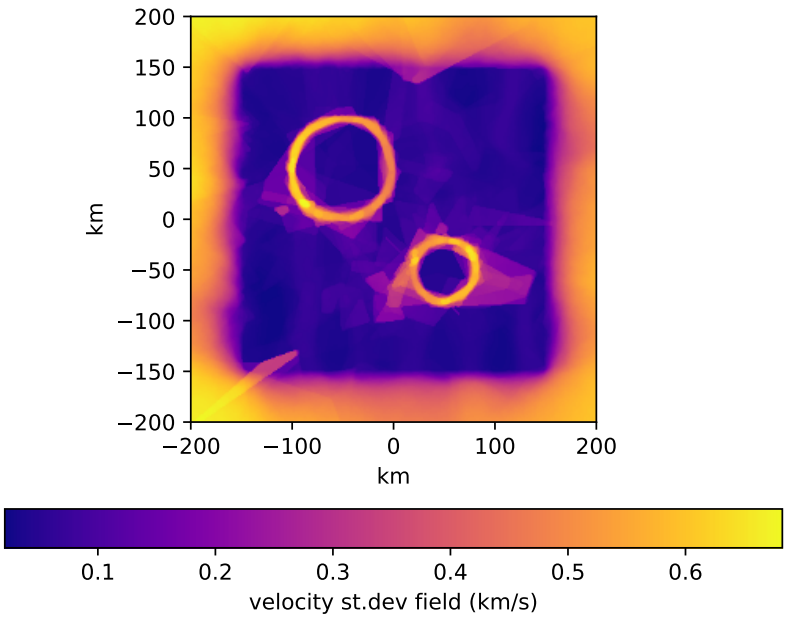

Figure 4.8: Standard deviation of the velocity field recovered.

4.3.2 Results interpretation

In this section some words are spent on the probabilistic interpretation of the results of a stochastic inversion. First point to notice: the reference field constructed as an average of the ensemble is not a real model. Models are the samples of the posterior distribution and they are subjected to all the oscillations and random fluctuations. These features are washed away typically by the average operator, resulting in a field representing mixed information from all the models sampled from the posterior. In some circumstances we know from section 3.3 that the expectation value of the posterior is also a model, but this cannot be sustained in general, especially when the inverse problem is not linear. This leaves space to an interesting way to formulate answers on the basis of the ensemble; if we wanted to know the velocity value of the model at a specific location, we can consider the entire ensemble of solutions and construct an histogram, which represents in the large sample limit, after normalization, the marginal distribution for the specific parameter being investigated.
In Fig. 4.9 four histograms are constructed for four different locations in the model of section 4.3.1. Each histogram represents a distribution of values for the parameter sampled from the posterior and it collects the entire complexity of the posterior distribution. Since this approach is based on Bayesian interpretation of probability, credibility intervals can be generated to represent the degree of belief to have the true value of the investigated parameter in a specified range.

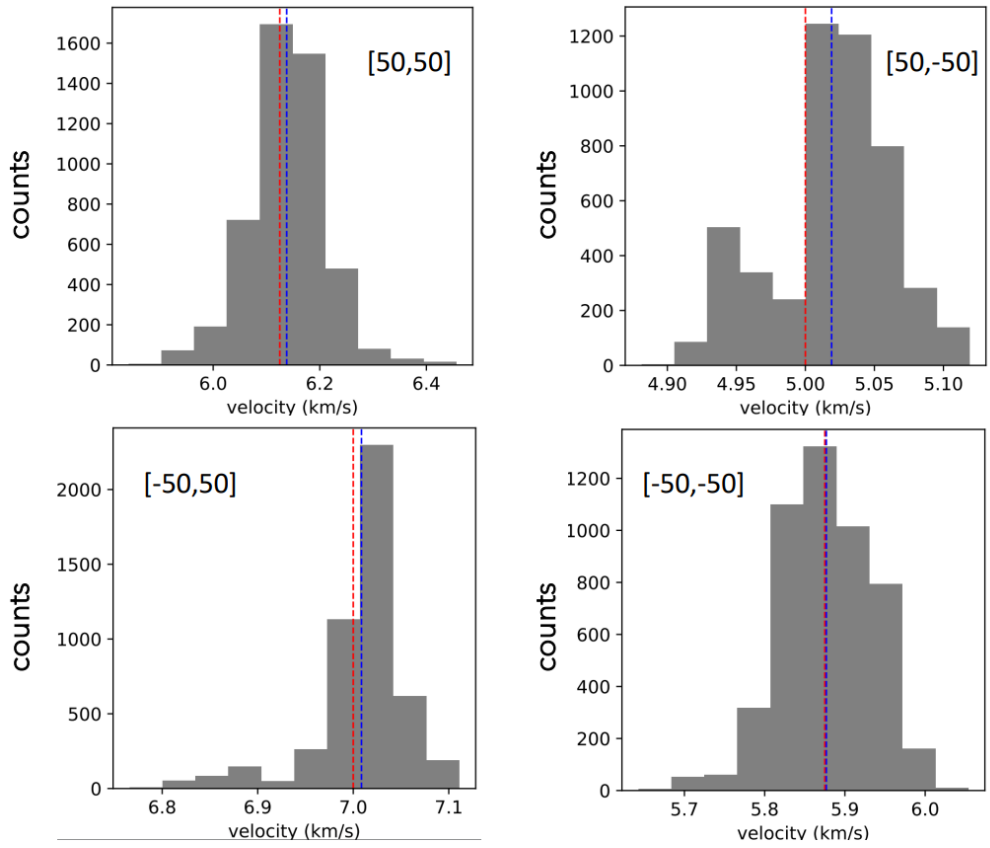

Figure 4.9: Four points are considered in the isotropic velocity model of section 4.3.1, for each point the distribution of velocity values is constructed using the ensemble of models produced by the sampling procedure. Red dashed line corresponds to the true value of the velocity in the location, blue dashed line is the average of the ensemble in the point.

In Fig. 4.10 two-dimensional posteriors are examined for two sections of the model. The limits described by the standard deviation estimate (grey dashed-lines) contain the true values of the velocity. These intervals become significantly large at the margins of the model, where there are not sampling rays, this is consistent with the uncertainty map shown in fig. 4.8. Statistical analysis can be conducted over the solutions provided by the reversible jump MCMC algorithm and they can be extremely useful in the evaluation of the features of the velocity field.

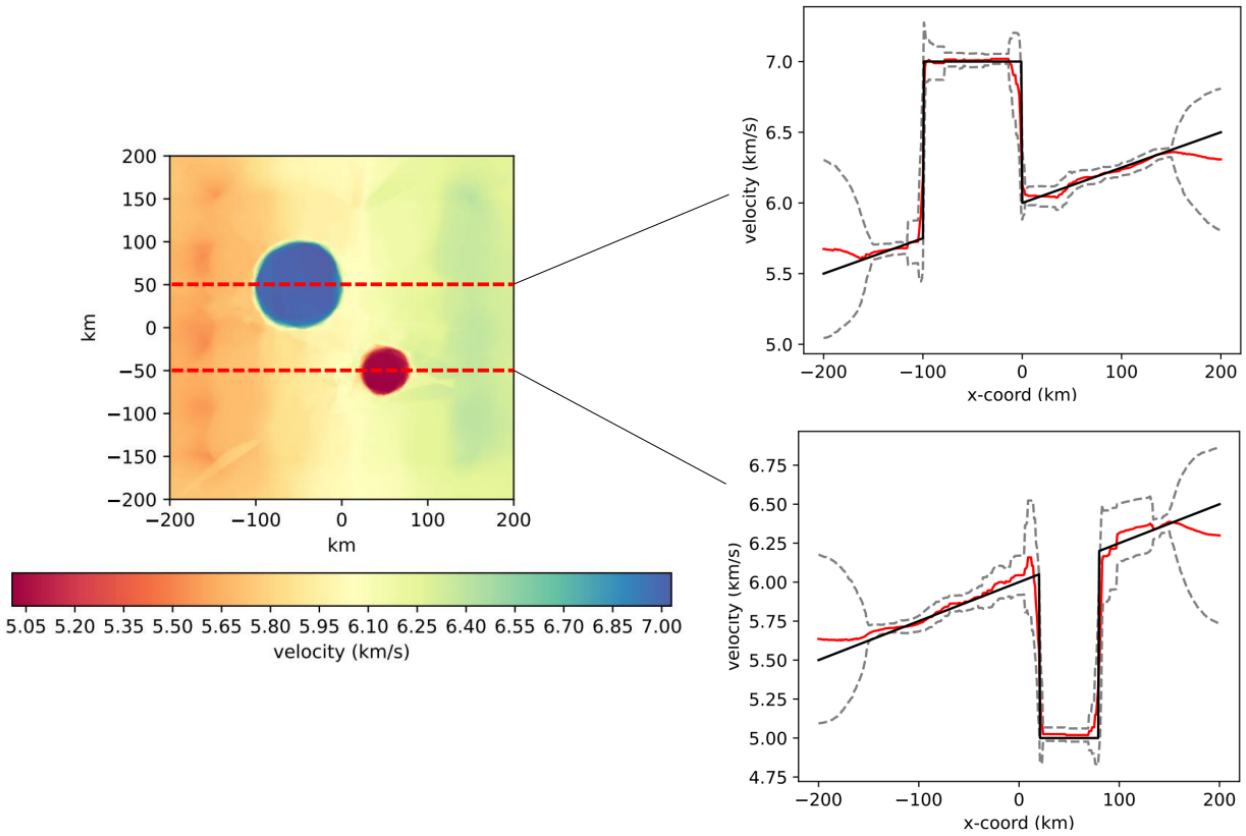

Figure 4.10: Two-dimensional marginal posteriors for velocity at two different sections of the space. Solid black and red lines are true model and average of the ensemble, the grey dashed-line delimits the interval average ± standard deviation.

4.3.3 Anisotropic data inversion with azimuth variability

The second experiment performed is the inversion of synthtetic data generated using an anisotropic velocity field. The field is the superposition of: (i) a circular isotropic fast velocity anomaly and, assuming a hexagonal symmetry for the elastic tensor, (ii) a distribution of fast velocity symmetry axes with circular symmetry (Fig. 4.11). The anisotropic magnitude fraction is homogeneous inside the circular structure and equal to 6%.
This experiment can evidence the performance of the algorithm with a rather complex anisotropic pattern where the orientation changes rapidly in the space. Due to forward modelling, the inverse problem is non-linear. Stations are white triangles in Fig. 4.11 and events are uniformly distributed along the margins of the space.
Gaussian noise with standard deviation of 50 ms is added to the synthetic data. Prior for number of nodes is log-uniform varying between 1 and 100 for each quantity.
For the sampling procedure one million iterations are performed and, considering the trends of the RMS (Fig. 4.12) for twenty-eight Markov Chains, the burn-in phase is set to the first 600000 iterations. The convergence is influenced by the random starting model, as it is evident when multiple chains are run in parallel.
The recovered model is shown in Fig. 4.13 and 4.14; both isotropic and anisotropic anomalies are recovered, the true values are smaller than the recovered ones in some regions, but they are anyway within the uncertainties provided. It is evident the presence of artifacts which are not characteristics of the true model. Since we can produce uncertainty maps with this stochastic method, it is possible to recognize these regions as high-uncertainty features of the velocity field, originated by the sampling procedure (Fig. 4.15, 4.16, 4.17). It is assumed that by increasing the number of chains, the average of the recovered fields should become less dependent on the interpolation algorithm.

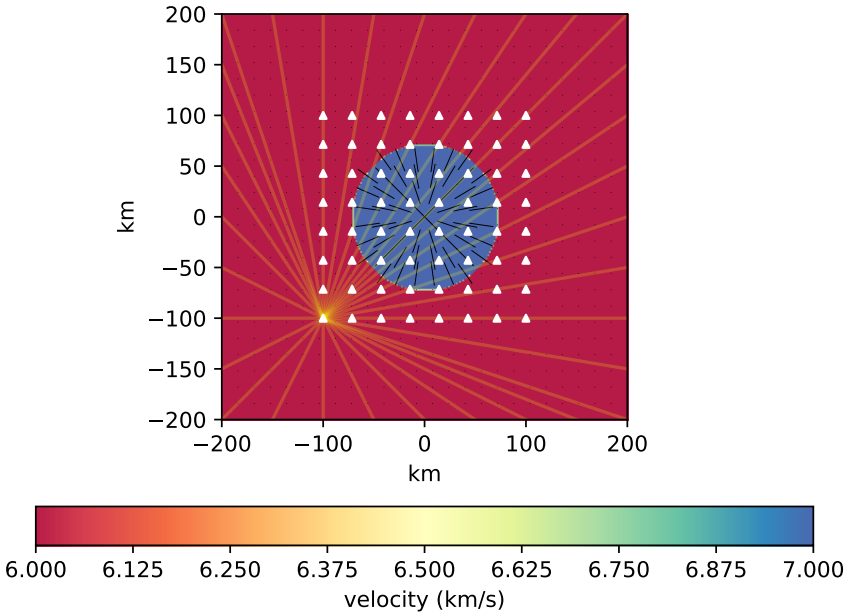

Figure 4.11: True model, used to compute synthetic data.

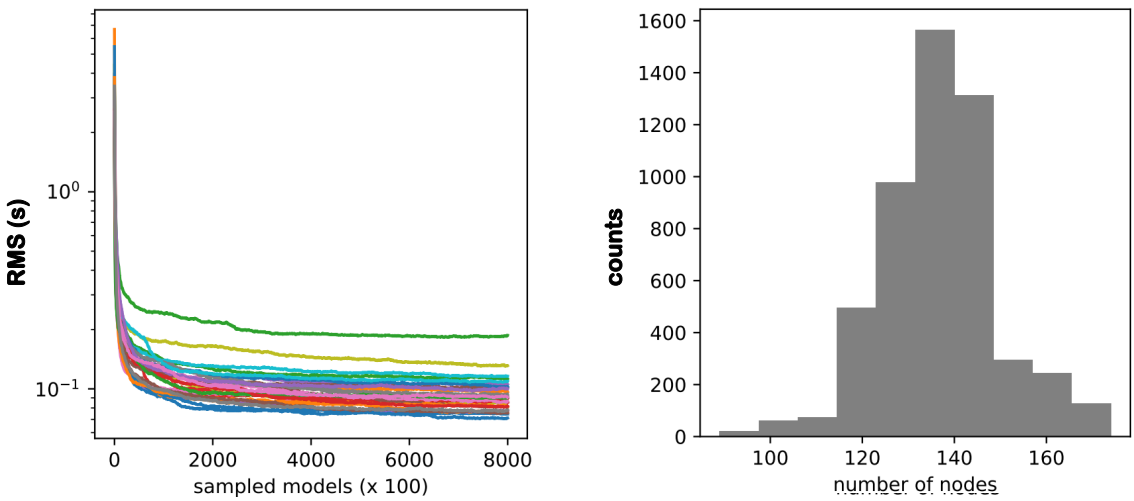

Figure 4.12: Left: trend of the RMS [s] for each chain run in parallel. Right: distribution of nodes from the ensemble.

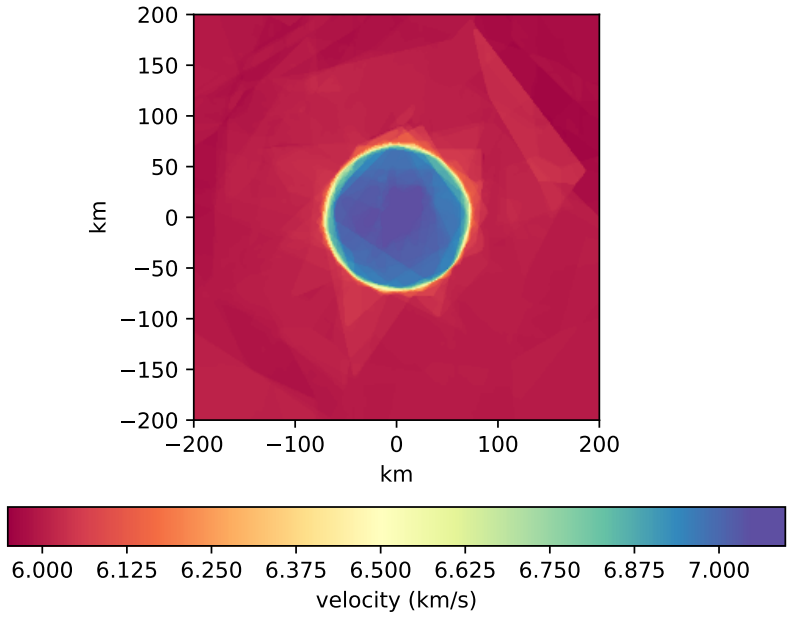

Figure 4.13: Isotropic velocity recovered using anisotropic inversion of anisotropic synthetic data.

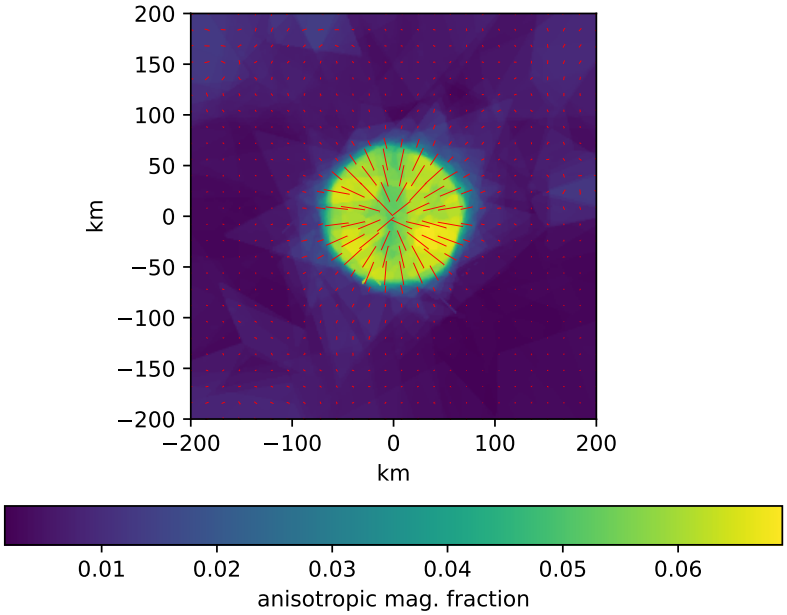

Figure 4.14: Anisotropic magnitude fraction recovered.

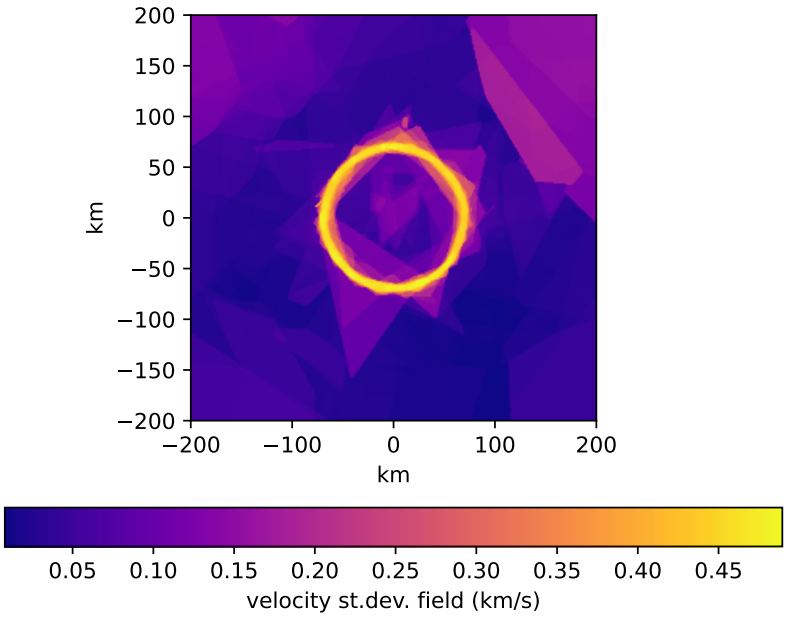

Figure 4.15: Standard deviation of the isotropic velocity field recovered.

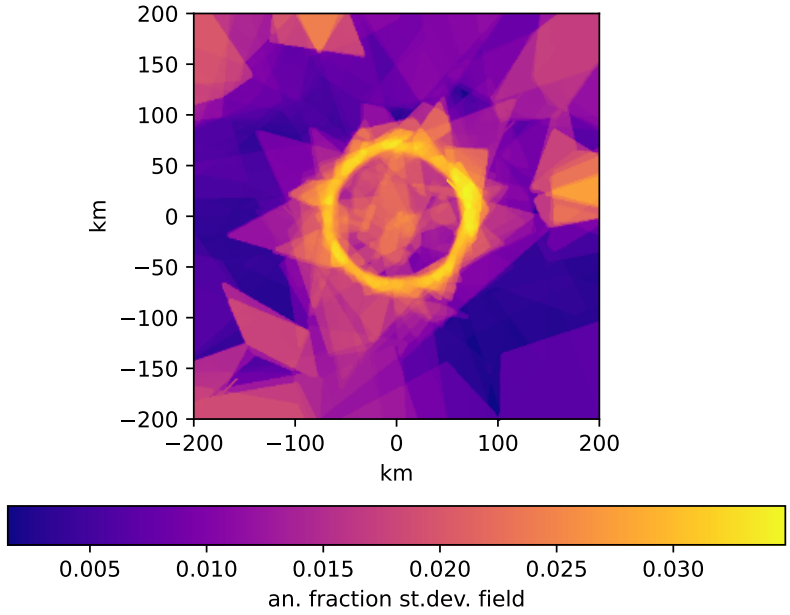

Figure 4.16: Standard deviation of the anisotropic fraction field recovered.

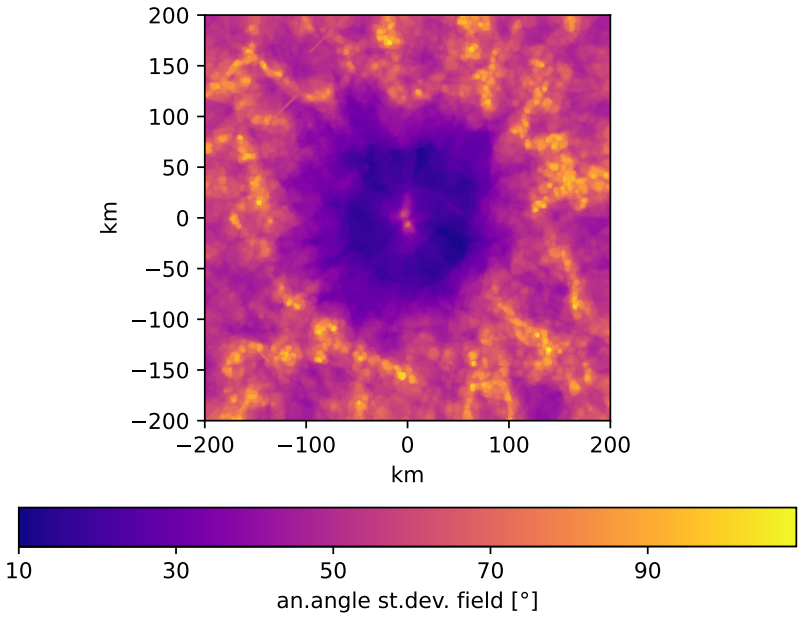

Figure 4.17: Standard deviation of the anisotropic azimuth field recovered.

4.3.4 Anisotropic inversion using stations at the surface

Since the common experimental setup in seismic tomography is characterized by stations at the surface of the Earth, a significant synthetic experiment to test is the situation where all the stations are placed at one side of the 2D space (surface) and a collection of events is randomly distributed below (Fig. 4.18). The target model is made by: (i) an isotropic velocity anomaly and (ii) an anisotropic component of the velocity field with homogeneous fraction (6%) and elevation angle (-45°). Gaussian error with $\sigma = 50\,ms$ is added to data.
One million iterations are performed and, considering Fig. 4.19 the first 600000 are discharged as burn-in phase. In this synthetic experiment twenty-eight Markov Chains are run in parallel.
The recovered isotropic field is represented in Fig. 4.20; the isotropic anomaly is well recovered in geometry, but it has a lower magnitude with respect to the target one, this could be explained by the slight stretching of the anomaly along direction of the rays or by the values assumed in some regions by anisotropic magnitude fraction (Fig. 4.21), the deviation is within the uncertainty associated with this region (Fig. 4.22). The standard deviation map for the anisotropic fraction field (Fig. 4.23) is not smooth and the anomaly is associated with high uncertainty, even if different Markov Chains are averaged. It is assumed that by increasing the number of chains of the ensemble (factor 10) this field should become smoother. In this experiment the geometry of the anomalies generates a large number of artifacts, evidenced also by the uncertainty maps, especially for anisotropic magnitude fraction.

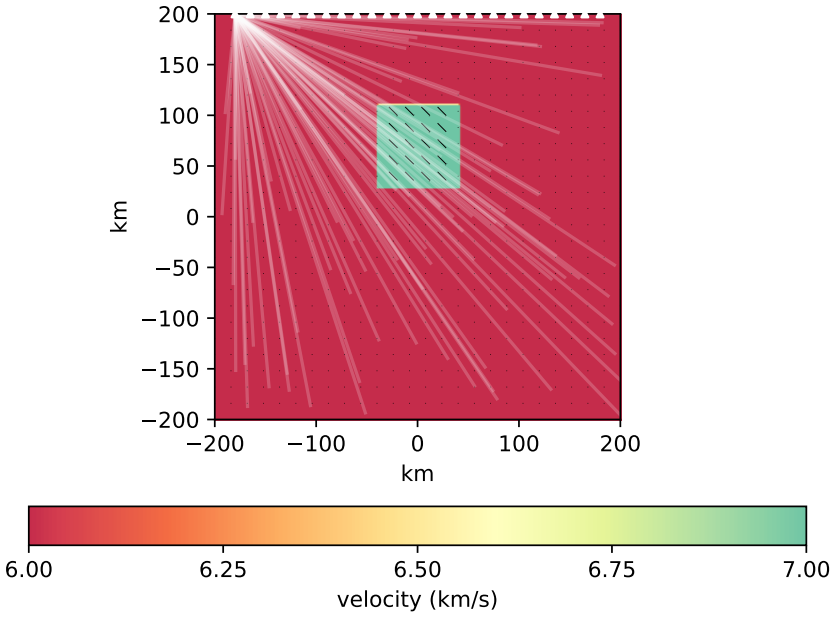

Figure 4.18: True model, used to compute synthetic data. This is the target of the inversion.

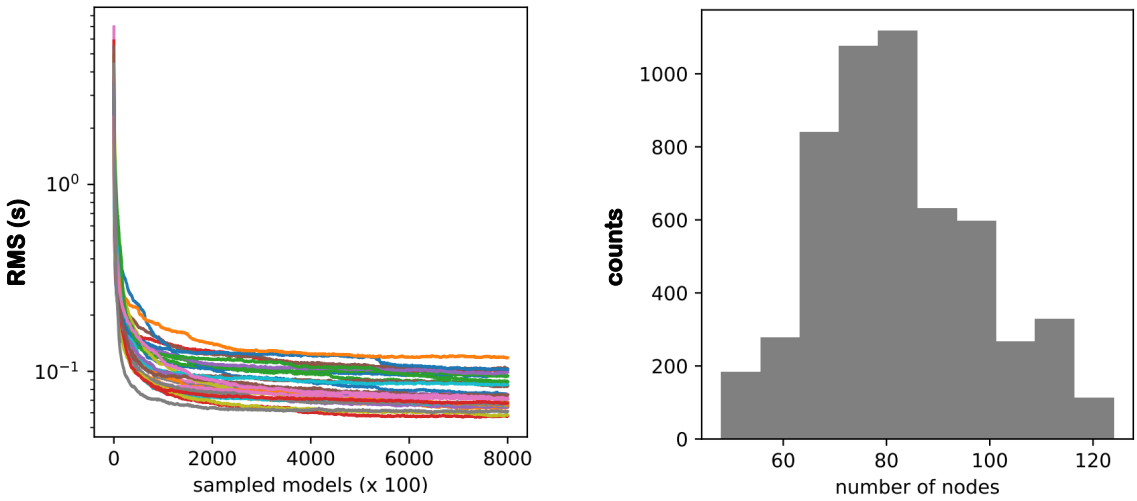

Figure 4.19: Left: trend of the RMS [s] for each chain run in parallel. Right: distribution of nodes from the ensemble.

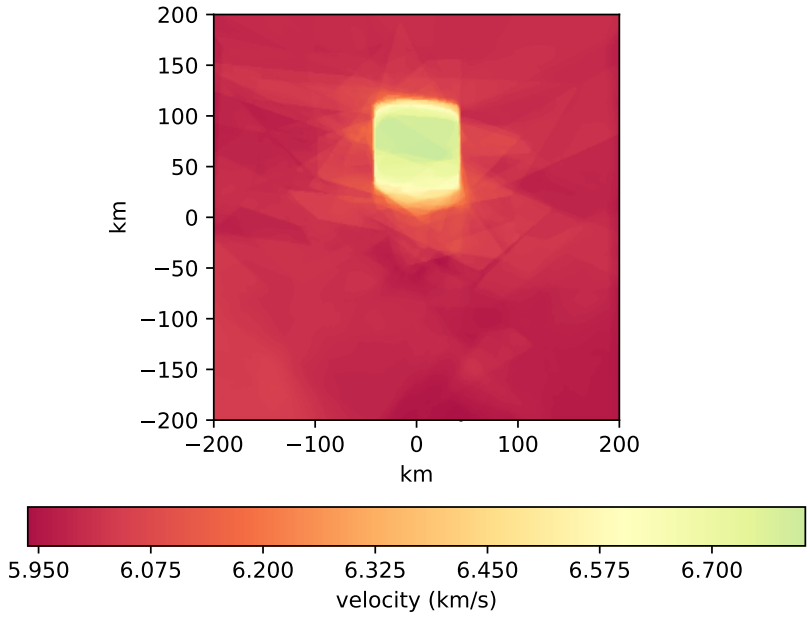

Figure 4.20: Isotropic velocity recovered using anisotropic inversion of anisotropic synthetic data.

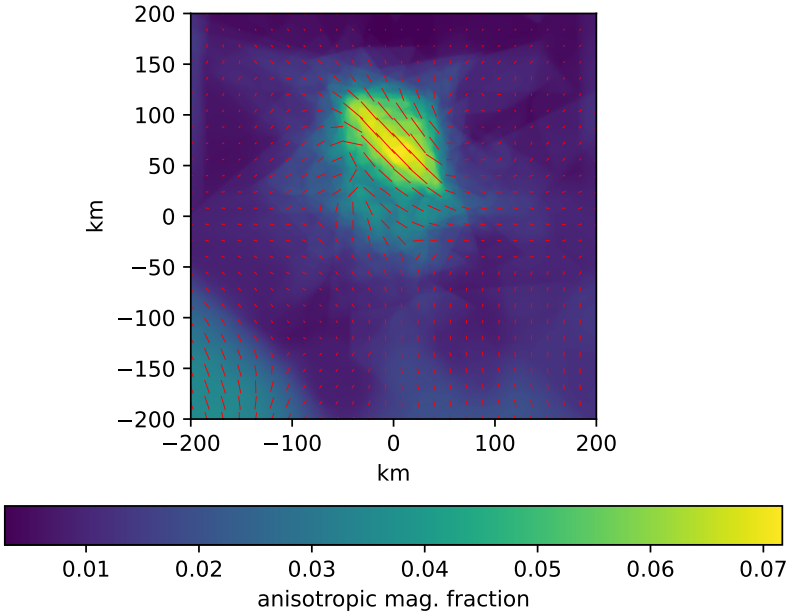

Figure 4.21: Anisotropic magnitude fraction recovered.

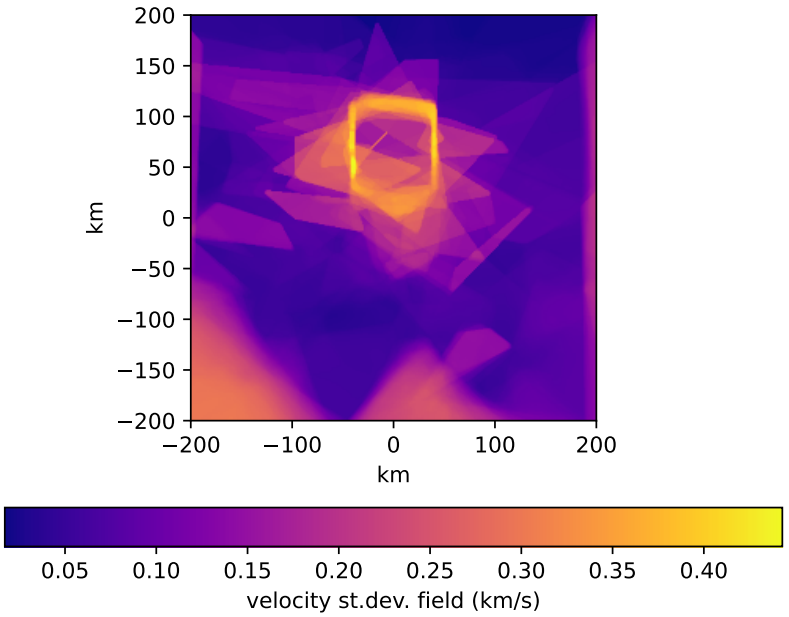

Figure 4.22: Standard deviation of the isotropic velocity field recovered.

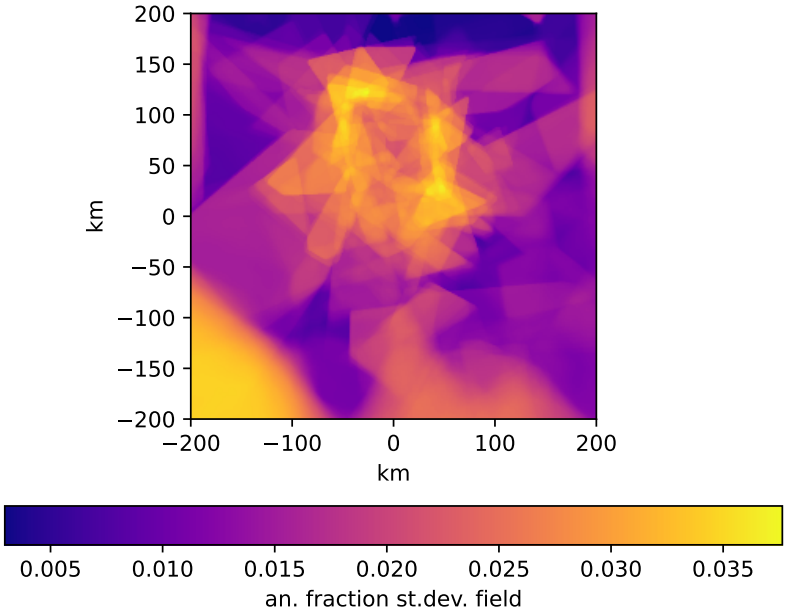

Figure 4.23: Standard deviation of the anisotropic fraction field recovered.

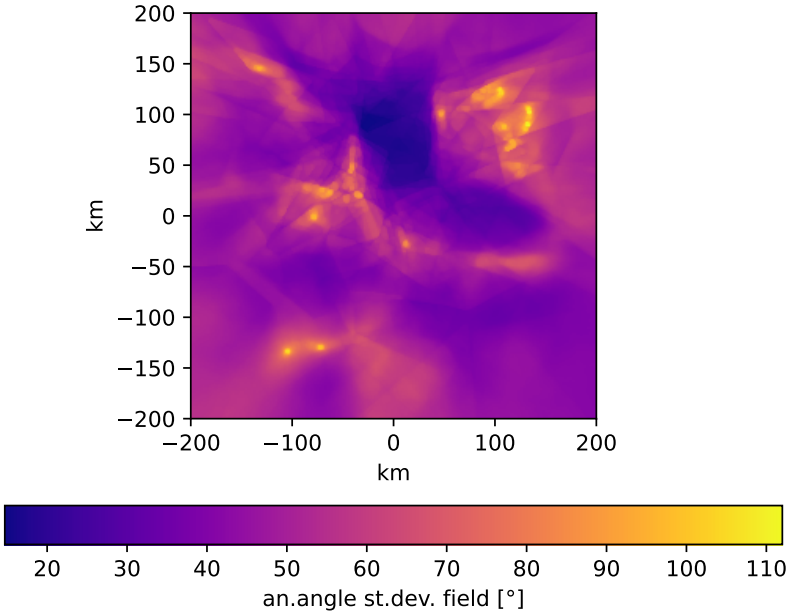

Figure 4.24: Standard deviation of the anisotropic elevation field recovered.

4.4 2D Spherical synthetic experiment

In order to make the code able to invert real observed data, with particular focus in subduction zones, the inversion domain is extended to 2D spherical geometry. Since Voronoi cells nodes are placed and moved with continuity inside the geometrical domain, when a spherical geometry is assumed for the region, this only influences the limits of the domain and the coordinate system used to identify the positions, but no actual grid is used for positioning of the nodes.
To test the ability of the reversible jump MCMC method to recover anisotropic velocity fields in realistic 2D situations, here the procedure is applied to a synthetic data set obtained using a true model characterized by a subducting slab of 100 km width with an associated distribution of fast velocity axes. This model is an extremely simplified representation of a subduction process, more complex and realistic models will be considered in Chapter 5.
From this point forward, Gaussian noise with 50 ms as standard deviation is added to all synthetic data.

4.4.1 Subducting slab with teleseismic waves

In this experiment an extremely simplified model representing a subducting slab is created. The slab is characterized by an isotropic velocity anomaly and a distribution of fast velocity axes which follows the direction of the subduction process. The anisotropic magnitude fraction is assumed homogeneous inside the slab, equal to 6%.
Stations (red points in Fig. 4.25) are distributed at the surface with a regular spacing. The domain ranges for 660 km in depth.
Tomography with teleseismic waves (waves produced at epicentral distances larger than 30°) uses events localized outside the region of the inversion. Rays can be traced using simpler velocity models outside the region and when long period teleseismic waves are used it could be not necessary to apply ray-tracing algorithm iteratively, as a simplifying approximation.
In this synthetic experiment events are placed at the bottom of the space and each station receives only rays with inclination between -60° and 60° (with respect to the tangent), simulating teleseismic waves.
One million iterations are performed. The first 600000 are discharged as burn-in phase. Twenty-eight Markov Chains are run in parallel.
Results of the inversion (Fig. 4.26, 4.27) show that the depth-changing component of the slab is recovered in isotropic velocity and anisotropic distribution of fast velocity axes, while the horizontal components are stretched in the directions of the teleseismic rays, this effect is the consequence of the limitations introduced in the inclination of the rays for this problem.
Standard deviation maps (Fig. 4.28, 4.29, 4.30) help to recognize the regions interested by large uncertainty; the internal portion of the subducting slab is characterized by the smallest standard deviation, proving that this region is well recovered in morphology and amplitude, while regions outside the slab, where the maps in Fig. 4.26, 4.27 deviate from the true one in Fig. 4.25 are the ones with larger uncertainty, as expected. This provides a confirm of the robust properties of uncertainty estimation of this MCMC method.

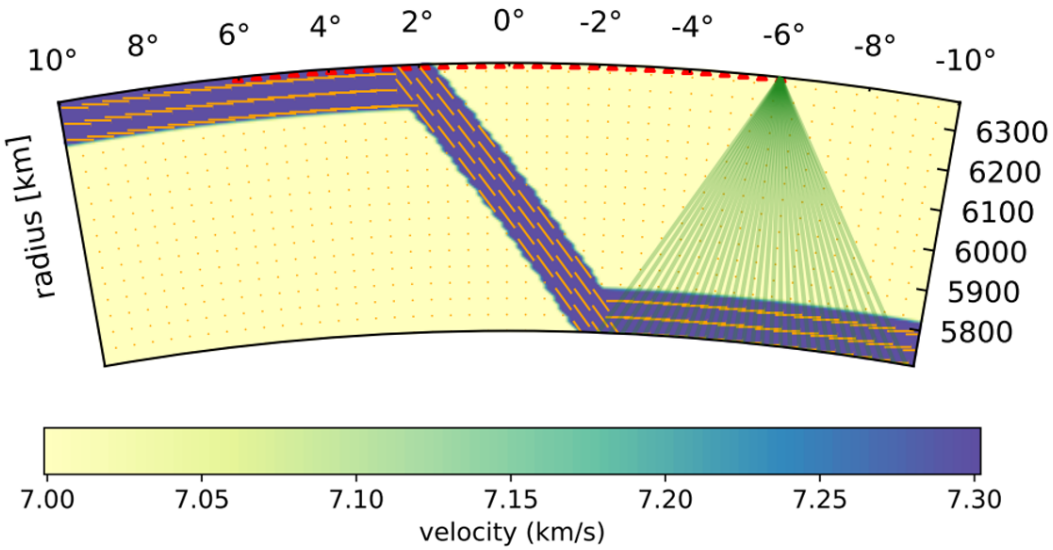

Figure 4.25: True model, used to compute synthetic data. This is the target of the inversion.

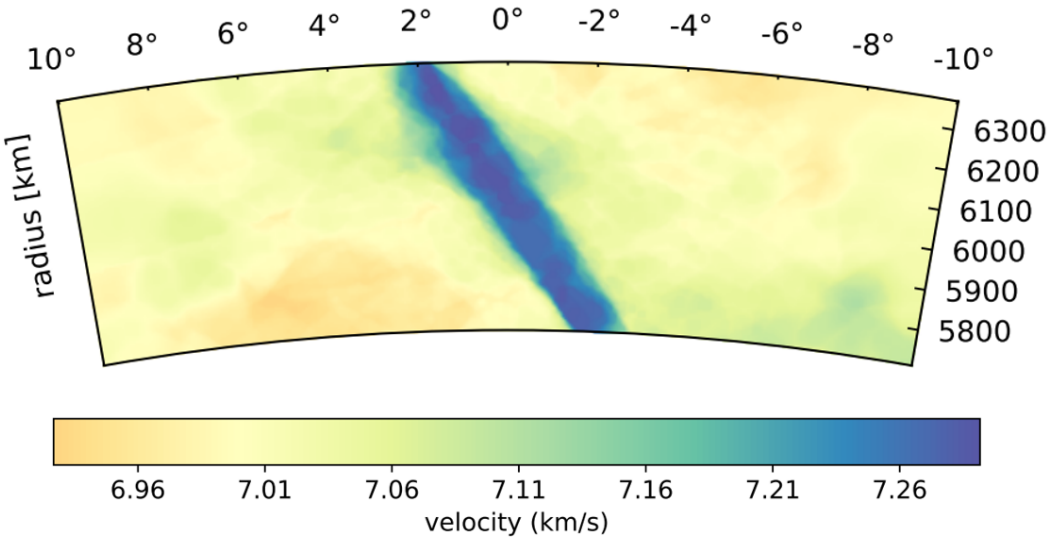

Figure 4.26: Velocity recovered using the inversion procedure.

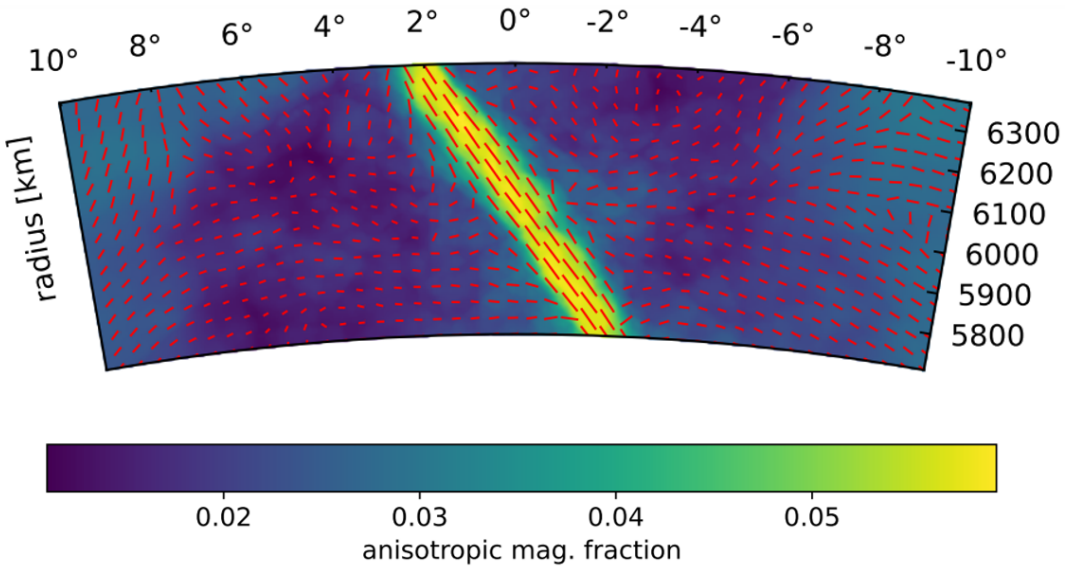

Figure 4.27: Anisotropy recovered using the inversion procedure.

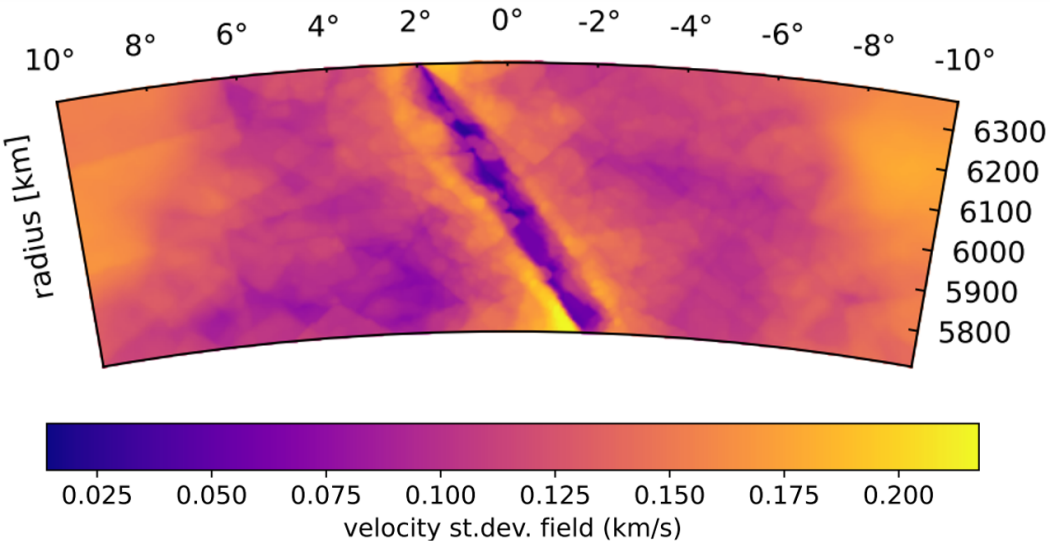

Figure 4.28: Standard deviation field for velocity.

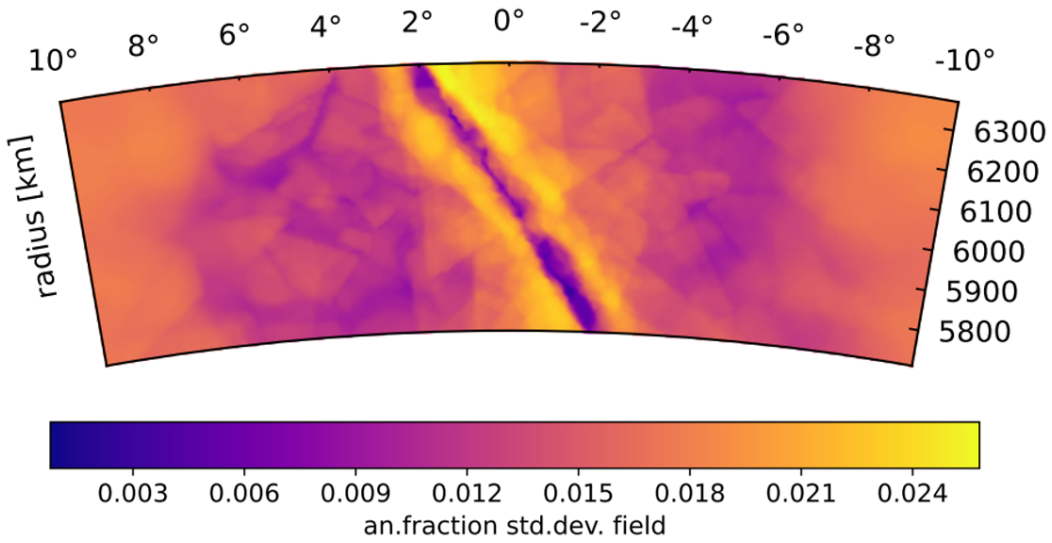

Figure 4.29: Standard deviation field for anisotropic fraction.

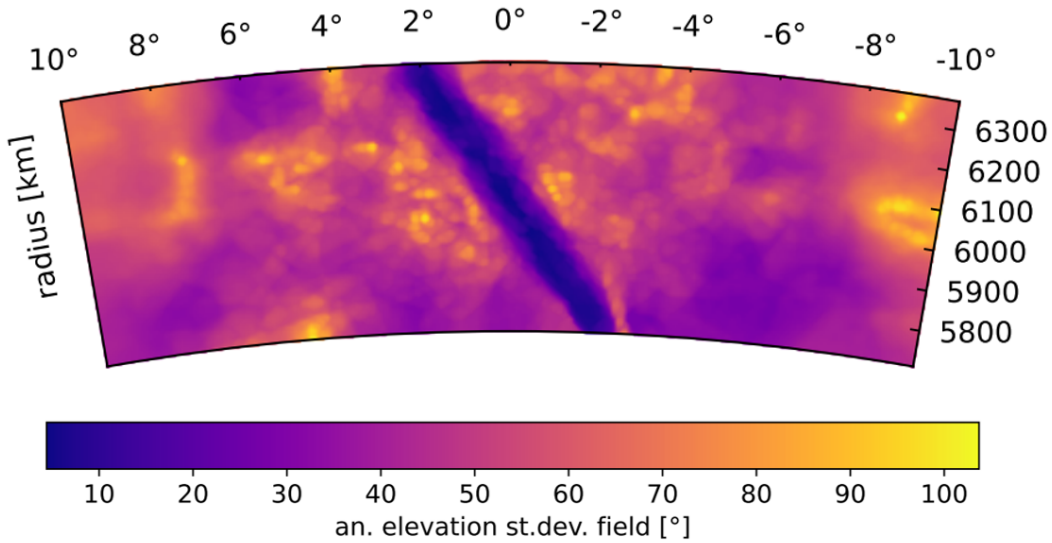

Figure 4.30: Standard deviation field for elevation angle.

Chapter 5

Inversion of realistic 2.5D subduction models with the rj-MCMC algorithm

In this chapter more realistic subduction models are considered for the inversion algorithm. So far synthetic experiments have been performed using parametrization for the velocity field in terms of isotropic velocity, anisotropic magnitude fraction and azimuth/elevation. For these models azimuth and elevation will be both integrated in the parametrization, using the complete formula for the evaluation of the anisotropic velocity field (eq. 4.3). In order to constrain the azimuth of the models it is necessary to generate synthetic data using also rays travelling outside the plane which passes through the model. Since the velocity model is kept two-dimensional, it is extended along the third dimension. This extension is generally referred to as 2.5-dimensional model.

5.1 Subduction of oceanic lithosphere with fossil fabric

The first model tested is generated using a 1D teleseismic reference velocity model, IASP91. The subducting slab is 5% faster than reference and fast axes of anisotropy are distributed along the subduction direction. Anisotropic magnitude fraction is homogeneous and equal to 6% inside the slab. Azimuth angle for anisotropy is set to zero for the entire model (Fig. 5.1). A Gaussian noise with standard deviation equal to 50 ms is added to synthtetic data. For this inversion teleseismic waves are traced from hypocenters to the stations distributed at the surface using TauP toolkit [Crotwell et al., 1999]. This package provides the ray-tracing in 1D reference model (IASP91, [Kennett, 1991]), then the ray-geometries are kept fixed during the inversion procedure. This is an approximation, often made for simplification, in section 5.4, where local events are introduced, rays generated by local earthquakes are traced using an iterative procedure. The approximation for teleseismic waves follows the approach of [VanderBeek & Faccenda, 2021], but in this analysis travel times are generated using ray-theory, which is valid in the high frequency limit, instead of a more consistent approach based on the computation of travel times in the finite-frequency framework. Since in this work synthetic experiments are idealized (synthetic data and inversion results are computed with the same code) we can rely on the ray-theory approximation. With realistic observations it would be necessary to change the forward modelling for teleseismic waves in order to have

more accurate estimates for teleseismic travel times.
Hypocenters are distributed according to the Fig. 5.2 and paths created by TauP toolkit are interpolated using a constant spacing (10 km) from the point where they enter into the domain to the stations.

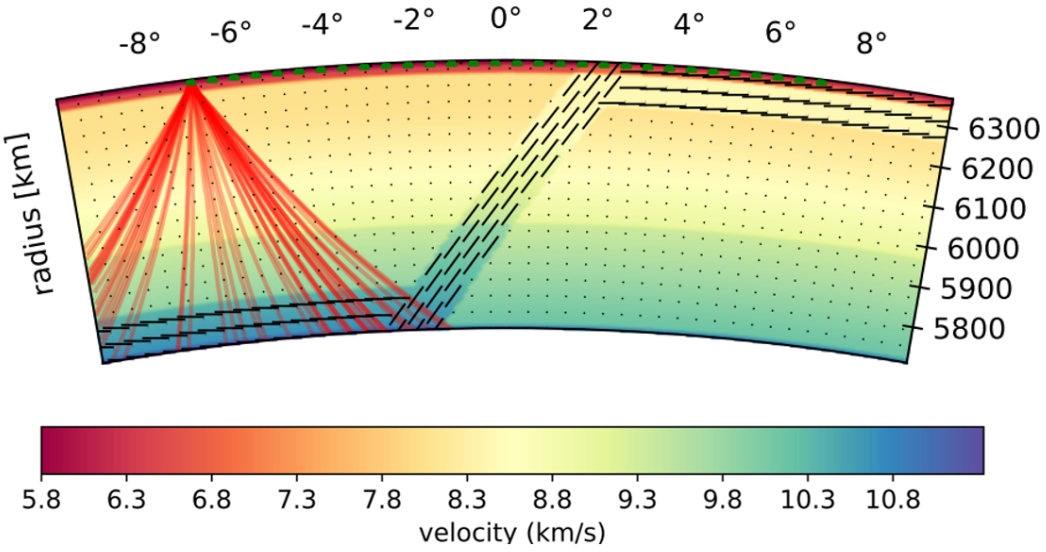

Figure 5.1: True model, used to compute synthetic data. This is the target of the inversion.

In the inversions performed so far, initial models of the Markov Chains were always randomly created. Since this field has a significant higher level of complexity and it is characterized mainly by vertical variation, we can expect limitations in the inversion with teleseismic waves. If an inversion is run with a random initial model, the solution is a velocity field which averages the velocity variation along the direction of the rays, resulting in a poor constrain of the true model. This produces a reasonable data fit, so it can be explained on the basis of the shape of the posterior. The *averaged* velocity field, producing reasonable misfit and RMS, could be a local maximum of the posterior distribution, and since it involves averaged contributions we can expect this local maximum to be wide in the model space. When a random initial model is used, it is likely that the sampling procedure will end being stuck in this local maximum and the maximum corresponding to the region we would like to sample remains isolated.
The solution proposed is this: initial model is not randomly selected, the parametrization of the velocity field is no more absolute, but treated as a perturbation of an isotropic reference velocity field 1D. Initial model is created randomly distributing nodes in the space with isotropic perturbation and anisotropic contributions set to zero, then the sampling chain starts. Using this parametrization it is more likely to start around the absolute maximum of the posterior.
For the first time a prior information is provided to the sampling algorithm, but in this case, in opposition to non-linear optimization solvers, no regularization/smoothing is applied, and

the chain can proceed sampling arbitrary far from the starting model.

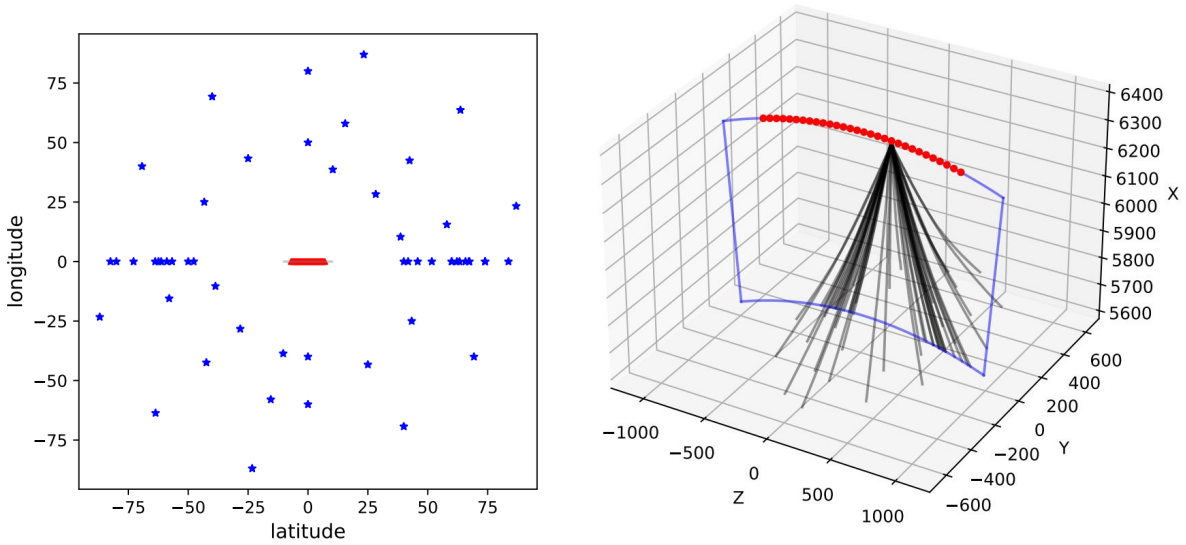

Figure 5.2: Left: distribution of seismic events (blue stars) and stations (red triangles). Right: rays traced with TauP toolkit and interpolated inside the inversion domain, paths for a single station are plot.

For this inversion one million iterations are run with 28 parallel Markov Chains, the first 600000 are discharged as burn-in phase. As reported in Fig. 5.3, 5.4 the subducting portion of the slab is well recovered in morphology and anisotropic magnitude/orientation, the amplitude of the isotropic velocity field is slightly faster than the true model, but this deviation can be bounded by the standard deviation map for this quantity (Fig. 5.5). As expected horizontal portions of the subducting slab are not recovered, even if the stretching of the anomalies is evident along the radial direction, from precedent studies we know that this is an effect of the inversion with teleseismic waves.
Standard deviation maps Fig. 5.5, 5.6, 5.7 and 5.8 evidence that the smallest-uncertainty regions are inside the subducting slab, while largest-uncertainty regions are identified at the frontiers of the slab and in regions interested by artifacts of the inversion procedure.
As shown in these results, acting on the initial model selection and parametrization successfully drives the sampling procedure toward the absolute maximum of the posterior distribution.

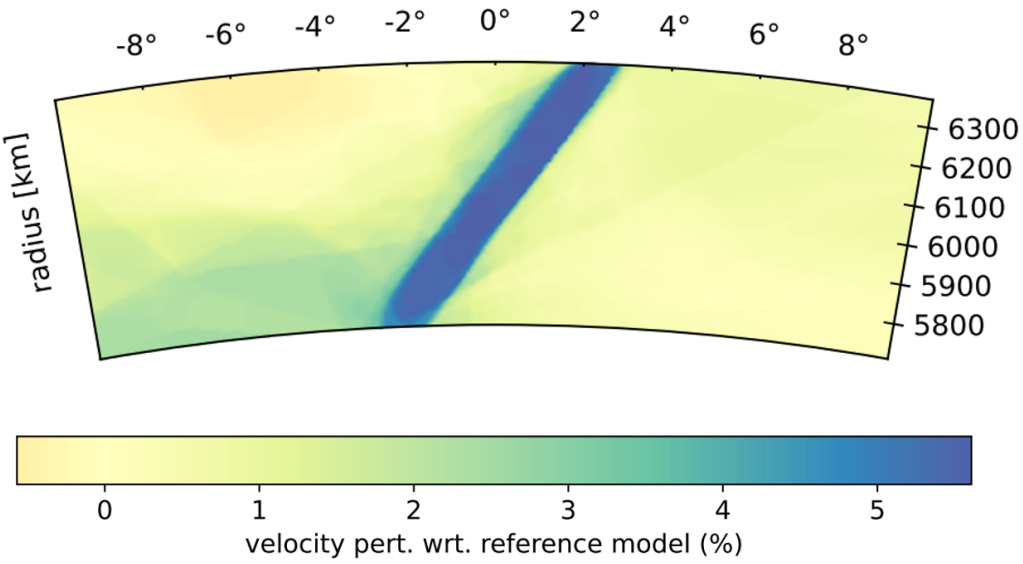

Figure 5.3: Velocity perturbation recovered using the inversion procedure.

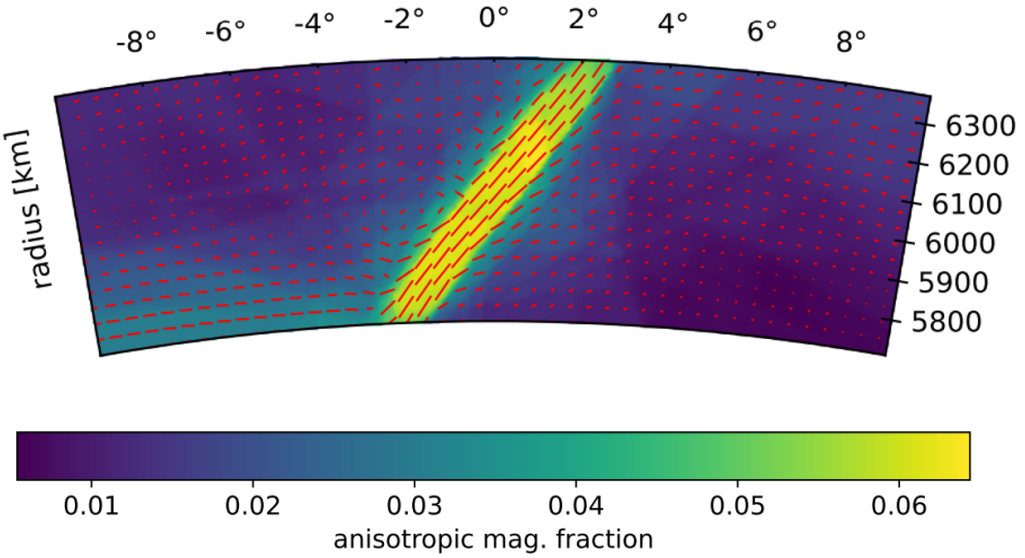

Figure 5.4: Anisotropic magnitude fraction recovered using the inversion procedure.

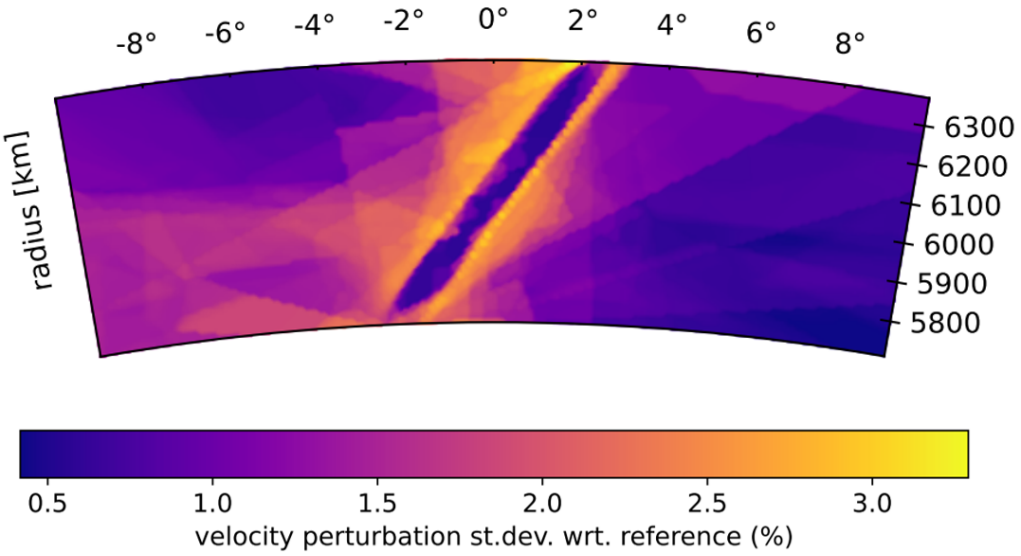

Figure 5.5: Standard deviation of the velocity field recovered.

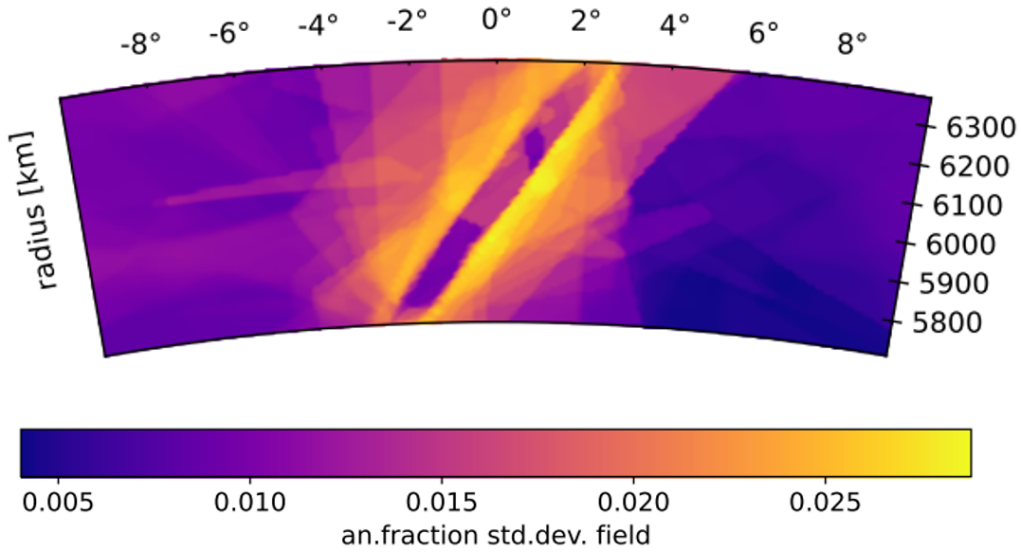

Figure 5.6: Standard deviation of the anisotropic fraction field recovered.

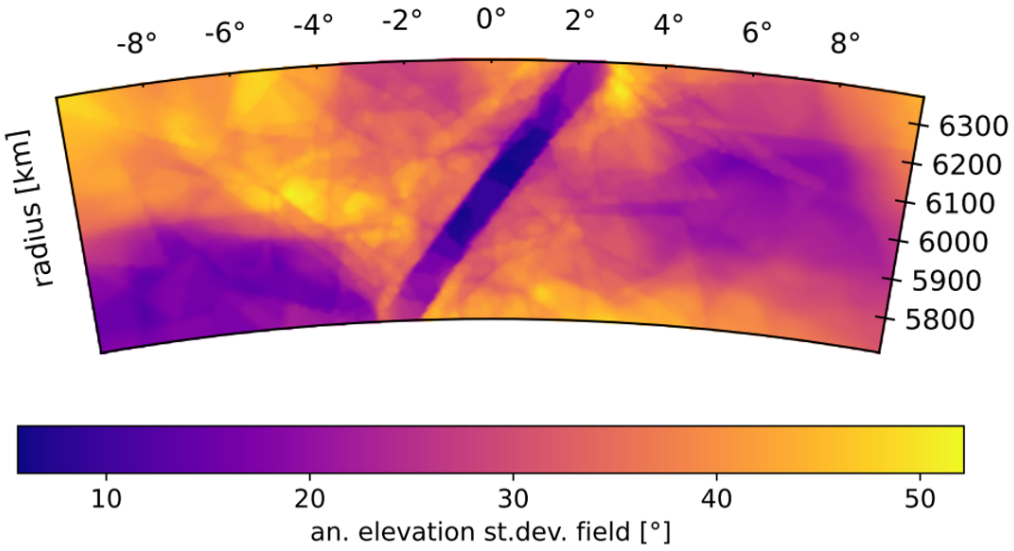

Figure 5.7: Standard deviation of the anisotropic elevation field recovered.

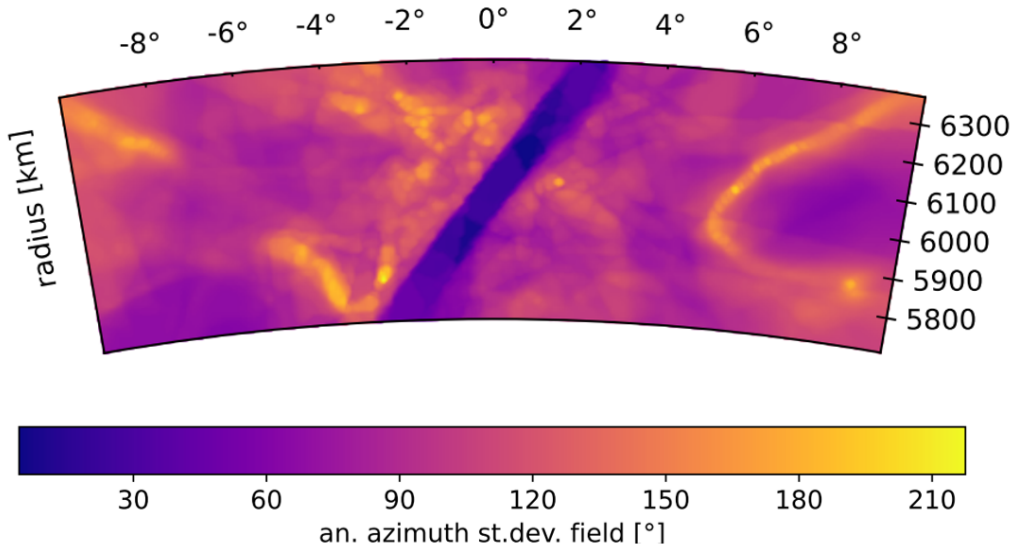

Figure 5.8: Standard deviation of the anisotropic azimuth field recovered.

5.2 Realistic 2.5D subduction with flow induced anisotropy

Following the direction of increasing complexity, now a subduction model resulting from a geodynamical simulation is considered. The petrologic-thermomechanical modelling is described in [Faccenda, 2014].
The model consists in a slab subducting under the action of negative buoyancy force and stagnating at 660 km of depth. Seismic anisotropy is modelled in the LPO framework; the lower transition zone is assumed to be isotropic, while anisotropy in the mid-upper mantle is developed across the subducting slab, under the influences of stress and strain fields. This model has not a crustal layer or topography.
From the geodynamical simulation a 2D section is extracted and tested for the inversion. The isotropic velocity field is represented in Fig. 5.9, teleseismic rays are traced using TauP toolkit and kept fixed in the inversion. Isotropic reference model used as basis for the perturbative parametrization of the anisotropic velocity field is extracted from the actual model in lateral regions. Using ratio between actual isotropic velocity and reference velocity, the isotropic perturbation field is plot in Fig. 5.10. Along the slab the P-wave velocity field is between 4-6% faster than reference, while horizontal anomalies are due to the differences in the seismic discontinuities between the 1D field and the 3D produced by the geodynamical model.
During the evolution of the subduction process, the anisotropic fabric for each polycrystalline aggregate is computed. Two contributions are considered: flow-induced fabrics parallel to the subduction and a frozen fabric characterizing the oceanic lithosphere, assumed to have been generated at the spreading center. Hexagonal symmetry is assumed to approximate the elastic tensors involved in the simulation, this is justified in the light of the dominant role of this symmetry system. In Fig. 5.11 the distribution of fast velocity axes is plot.
Since recovered magnitude for isotropic velocity is lower than true one, the color scale for velocity perturbation is cut between -3% and 3%.

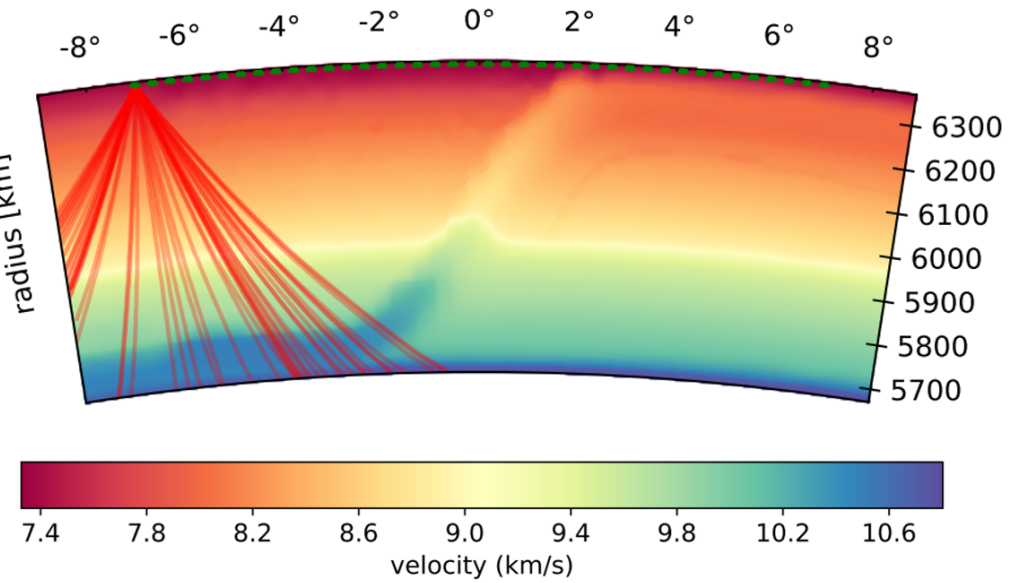

Figure 5.9: Absolute isotropic velocity of the subduction model used as target of the inversion

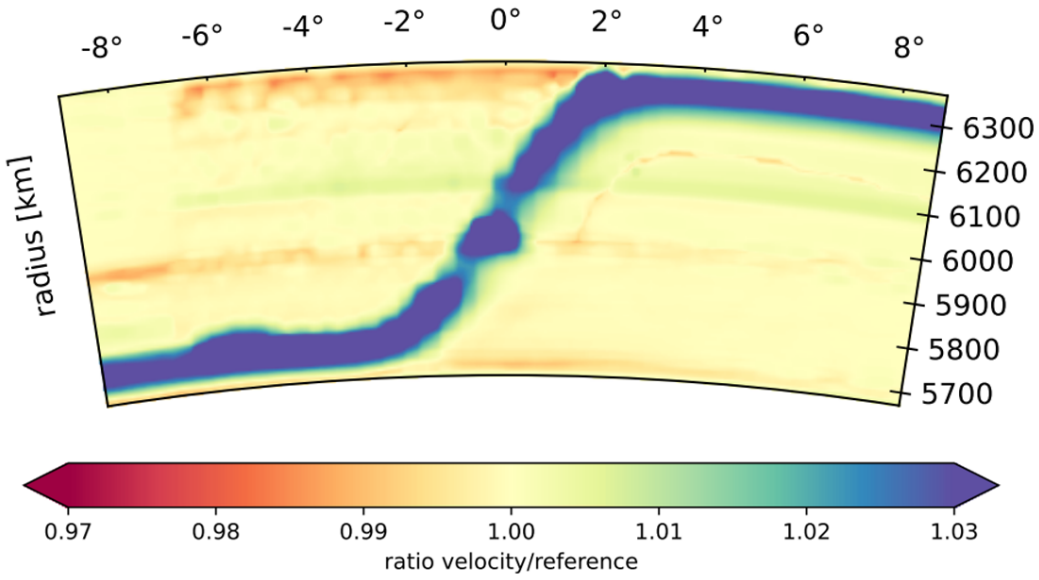

Figure 5.10: Isotropic velocity of the true model as ratio referred to the reference model.

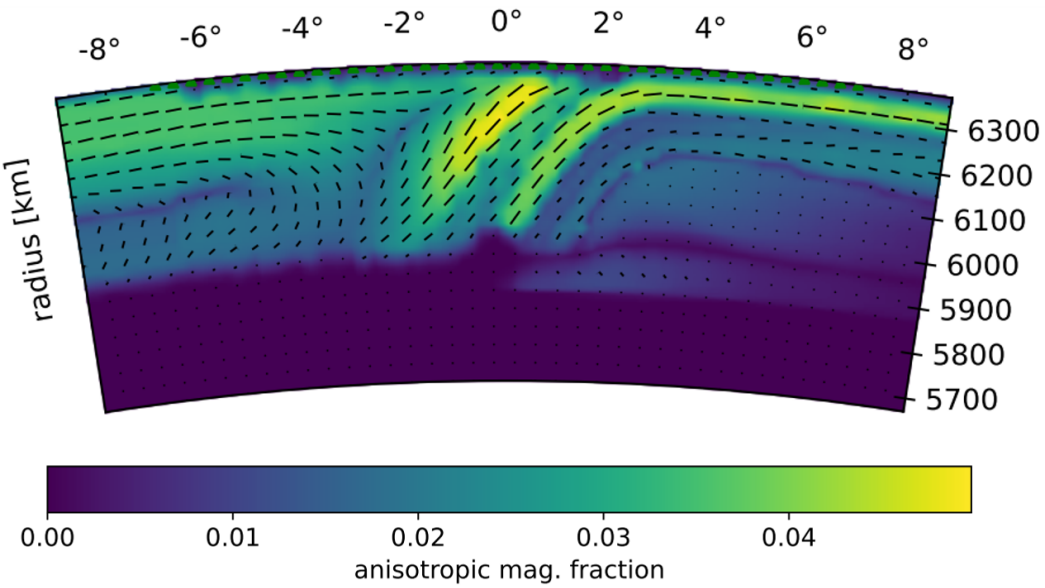

Figure 5.11: Anisotropic structure of the true model.

Inversion results with 28 parallel Markov Chains follow. In this anisotropic velocity field, flow-induced anisotropy is separated from the isotropic velocity anomaly associated with the subducting slab, in Fig. 5.12, where the recovered isotropic velocity field is plot, it is evident that using only teleseismic waves and starting from the reference model described above the sampling algorithm is not able to separate the anisotropic and isotropic contributions as in the target model. The resolved fields create an averaged contribution between anisotropic magnitude fraction and isotropic perturbation to velocity field, probably using as starting model a reference model including an anisotropic distribution, sampling method could proceed toward the absolute maximum of the posterior probability distribution. Despite this, flow-induced and frozen fabrics are resolved in the anisotropic distribution (Fig. 5.13), with limitations in morphology and amplitude recovered.
The reasoning proposed above is sustained also by the error maps (Fig. 5.14, 5.15). Uncertainties for the isotropic perturbation evidence high standard deviation in regions where the isotropic anomalies are stretched, like the zones outside the slab. The order of magnitude of the uncertainty is high also inside the slab, this in compatible with the fact that not the entire magnitude of isotropic anomaly is recovered. Also uncertainty map for anisotropic magnitude fraction can provide useful information to interpret the result of the inversion. Regions where flow-induced and frozen fabrics exit are characterized by uncertainties around 1-1.5%, justified by the fact that anisotropic magnitude in these regions is not fully recovered, due to the mixed effect with isotropic field. Anisotropic magnitude below 6000 km has no correspondence in the true model, this is consistent with the large uncertainty associated.

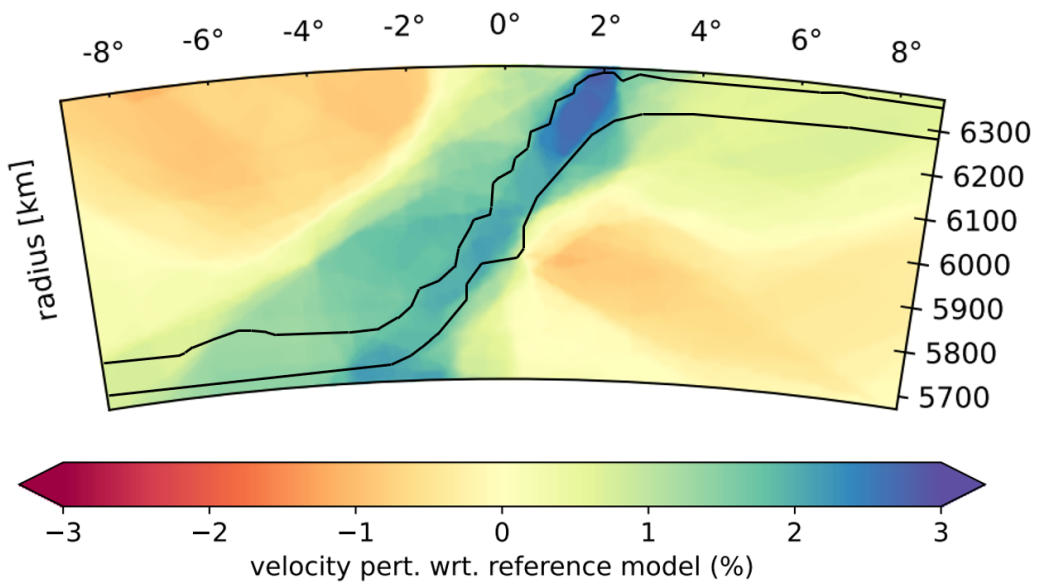

Figure 5.12: Velocity perturbation recovered using the inversion procedure. Profile of the subducting slab is superimposed.

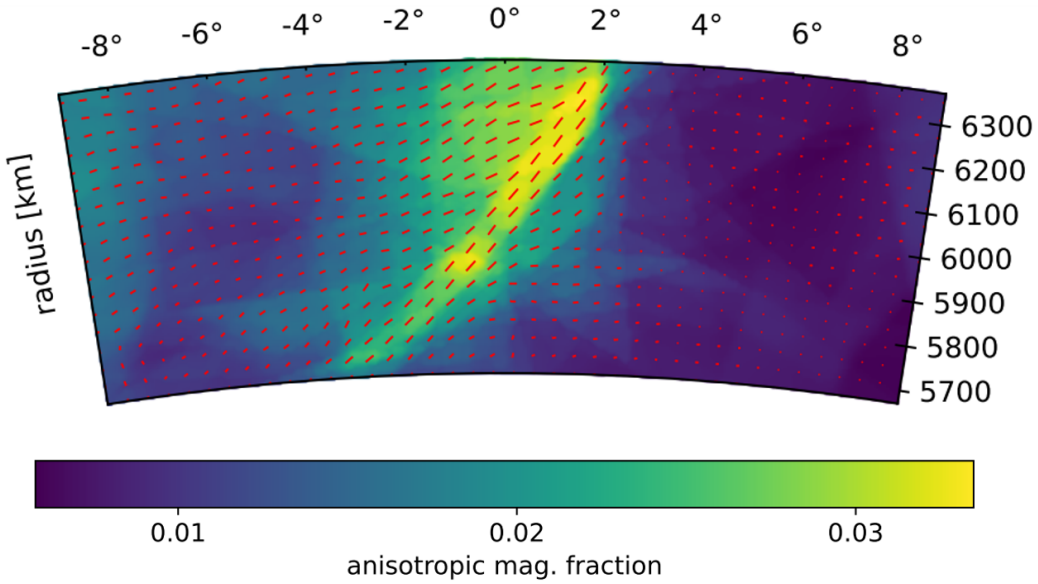

Figure 5.13: Anisotropic magnitude fraction recovered using the inversion procedure.

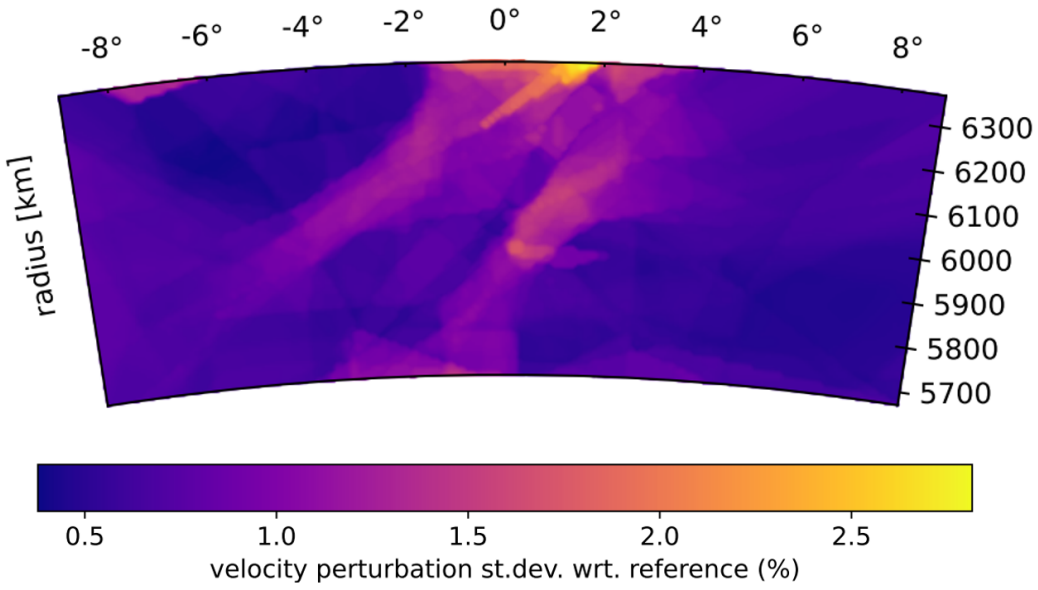

Figure 5.14: Standard deviation of the velocity field recovered.

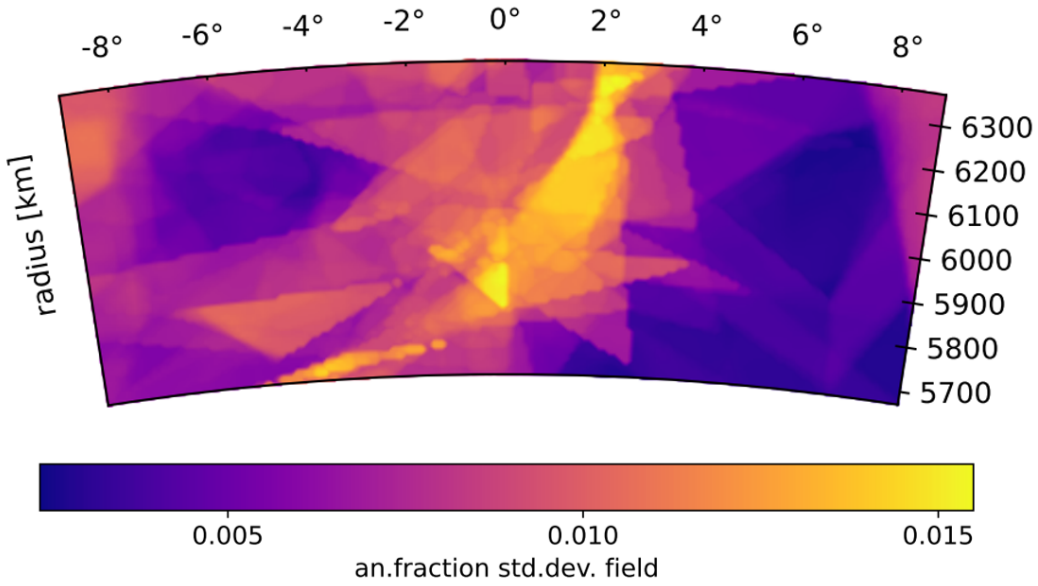

Figure 5.15: Standard deviation of the anisotropic fraction field recovered.

Once the main limitations in this inversion are recognized in the de-coupling between isotropic and anisotropic features, and this can be tested running an isotropic inversion of the isotropic version of the subduction model (Fig. 5.16, subducting slab is well constrained), the question is: how to overcome the limitations introduced in these inversions with teleseismic waves? As said before, a possibility is to change the starting model selecting a reference around the maximum of the posterior to sample, for this purpose models produced by non-linear optimization solvers or other inversions can be tempted. Second possibility is to increase the discrimination among the observations using also local earthquakes. Rays from these events allow to have a larger interval of elevation to investigate the subducting structure, using this approach we can expect to resolve better the isotropic and anisotropic components. But introducing local earthquakes in realistic velocity fields makes necessary to trace the rays using tracing techniques.

In real situations attention should be paid with local earthquakes, since tracing is performed in an iteratively changing velocity field, earthquakes should be re-localized in space and time. This can be accomplished in two ways: a localization algorithm can be run each specified interval of iterations, or, following the philosophy behind Monte Carlo methods, the space-time parameters for each earthquake can be taken into the parametrization allowing perturbations according to the probabilistic criteria defined for the Metropolis-Hastings. Clearly this can increase significantly the number of the free parameters of the inversion.

For the inversion of this subduction model we propose a joint inversion of teleseismic waves and rays coming from local earthquakes (traced in 2D), whose positions are assumed to be fixed in space and time. In the next section the discussion about ray-tracing is approached with the solutions proposed for this context.

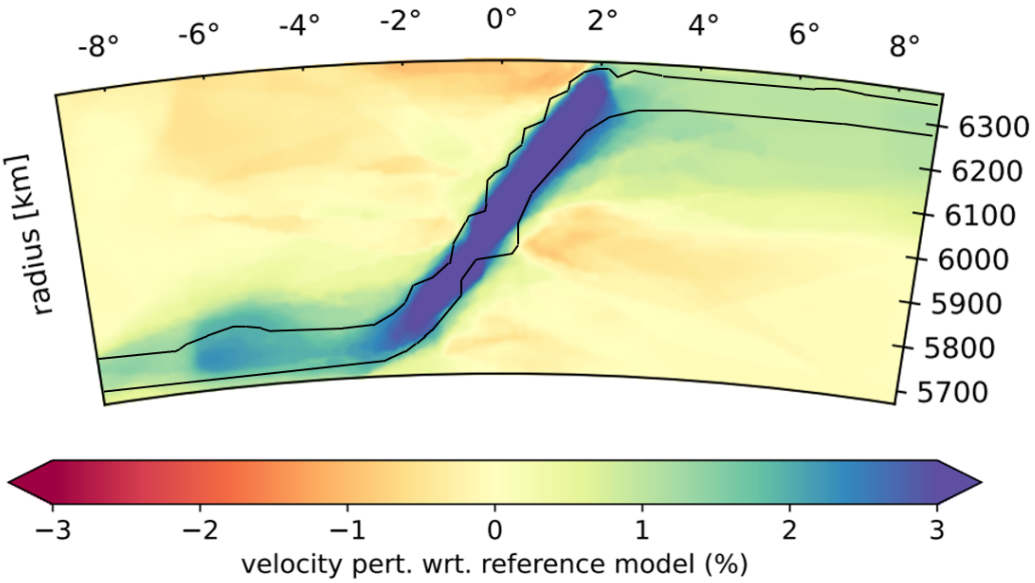

Figure 5.16: Isotropic inversion of synthetic data generated using isotropic version of the subduction model.

5.3 Ray-tracing

In this section some common algorithms used to trace rays in anisotropic velocity fields are described. Recalling the structure of AnisoTOMO code (Fig. 4.1), reference velocity fields used in ray-tracing are constructed as ensemble average of sampled models by multiple chains in the inner loop.

5.3.1 Shortest Path Method in graphs

During each iteration of the outer loop, the reference model constructed is used to compute the shortest paths between sources and receivers using the method developed by Edsger W. Dijkstra. The Shortest Path Method (SPM) [Dijkstra, 1959], was initially developed to find the shortest path between a node in a graph (source) and any other node (receiver).
Here a brief description of the algorithm is provided. In Fig. 5.17 a graph is represented, with the source node labelled with "1". At the beginning all the nodes are marked as "unvisited" and a tentative distance (with respect to the source node) is assigned to each one; the distance of the source is set to zero, while all the other nodes have an infinite distance at this step.
For the current node, distance is evaluated for all the neighbours labelled as "unvisited"; if the distance is smaller than the precedent one assigned, the new value is assigned and the current node is labelled as "visited", it will not be considered again in the algorithm. If every node of the graph (or only a particular receiver) has been labelled as "visited", the algorithm stops, otherwise the node with the smallest distance becomes the current node and the algorithm re-starts from the evaluation of the neighbours.

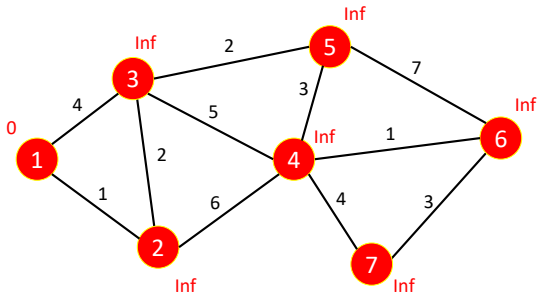

Figure 5.17: Initial state of the Graph. Node labelled with "1" is the source. Tentative distances are set to infinite for all the nodes, except the source (1), where distance is zero.

In Fig. 5.18, 5.19, 5.20 three steps of the algorithm are represented using a simple graph.
In Fig. 5.18 the current node is the one labelled with "1", the source. The neighbours "2" and "3" are unvisited nodes and their distance is evaluated using the edge connecting them with "1". In both cases the new distance is smaller than the precedent one (infinite) and the values are updated. The closest node is "2", which is now the current node, while "1" is set as "visited". In Fig. 5.19 the current node is "2", with neighbours "1", "3" and "4", but since "1" is marked as "visited", only "3" and "4" are considered. The distances evaluated are smaller than the precedent values and they are updated. Now the current node is "3" and "2" is set as "visited". The same procedure is applied in the third step (Fig. 5.20).

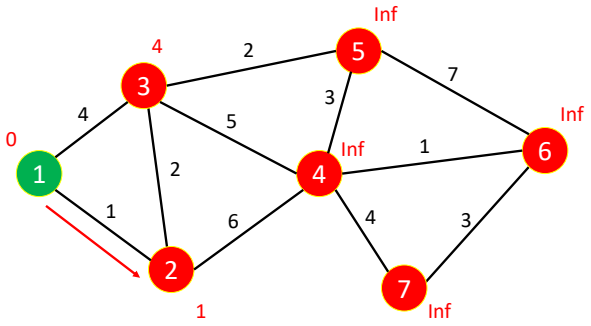

Figure 5.18: Distances of the nodes "2" and "3" are evaluated and the assigned values are updated. The node "2" is now the current node (the one with the minimum distance among the unvisited.)

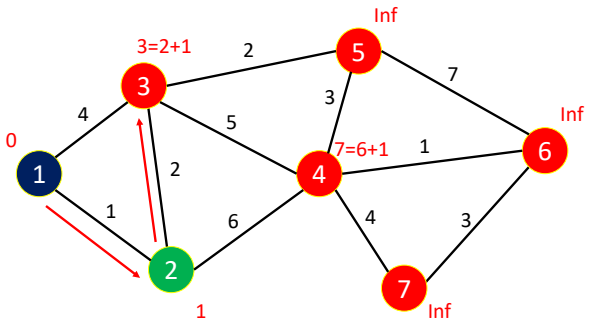

Figure 5.19: Distances of the nodes "3" and "4" are evaluated and the assigned values are updated. The node "3" is now the current node.

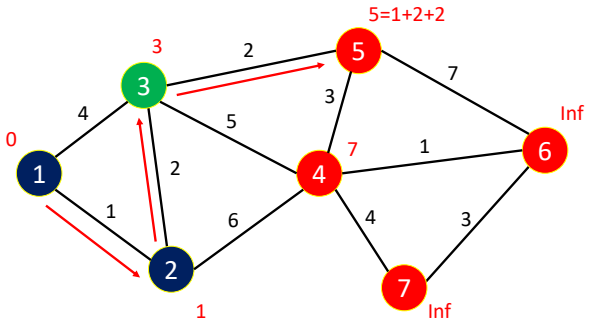

Figure 5.20: Distances of the nodes "5" and "4" are evaluated, only the value associated with node "5" is updated. The node "5" is now the current node.

In order to trace the shortest path among the nodes for a source-target pair, a quantity Q is instantiated. This structure collects the nodal indices associating them with keys. To have $Q[i] = j$ means that the node precedent to i is j, along the shortest path identified so far.
Every time a distance value of a node i is changed, Q is updated putting $Q[i] = Q_i$, where Q_i is the actual node, whose node i is a neighbour.
At the end of the execution of the algorithm it is possible to reconstruct the shortest path between *source* and *target* using the structure Q. Starting from $Q[target]$ and iterating to the *source*, the path is found proceeding backward.
An alternative implementation of the shortest path method using the Bellmann-Ford-Moore algorithm is used in sections 5.3.4 and 5.4. In this implementation at every iteration all the nodes which have changed distance values are *relaxed*. The relaxing process considers for every node counted the neighbours and computes the new distances, proceeding also in the updating of the structure Q described previously. This is iterated until no more changes are proposed for the distance values. This version of the shortest path method can be easily parallelized, especially in the relaxing process, allowing to reduce significantly the computational time by making the tracing algorithm run on multiple cores.

5.3.2 Shortest Path Method with grid model

Shortest path method was initially developed to work in connected graphs, but it was successfully applied for ray tracing in seismic tomography in [Nakanishi & Yamaguchi, 1986].
This method can be easily applied to a structured grid, but it requires the evaluation of the neighbouring nodes for each node of the grid.
When a *grid model* network geometry is used, the grid instantiated is divided in geometrical elements, delimited by groups of nodes. In the case of a 2D cartesian grid the geometrical elements are rectangles, each delimited by four nodes considered as neighbours. Each node has naturally eight neighbours.

The introduction of a space discretization limits the geometries allowed for the rays and the accuracy of the corresponding ray paths (Fig. 5.21). It is possible to expand the nodal-incidence by considering as neighbours of a node also the neighbours of the neighbours (Fig. 5.22). This can be iterated for different orders, but while the effects of the space discretization are mitigated, larger gaps between neighbouring nodes are allowed and smaller features of the velocity field could be ignored in the computation of the shortest path. The connection geometry is referred to as a forward star. The level of a forward star is the number of layers of neighbouring nodes allowed for each node.
Two errors characterize the shortest path method: the error due to the geometry of the grid is known as the *network geometry error*, while the error due to the gaps beyond small features of the velocity field is known as the *heterogeneity error*.

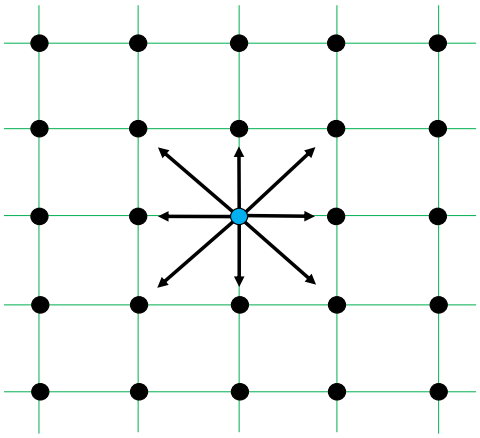

Figure 5.21: First level forward star connections.

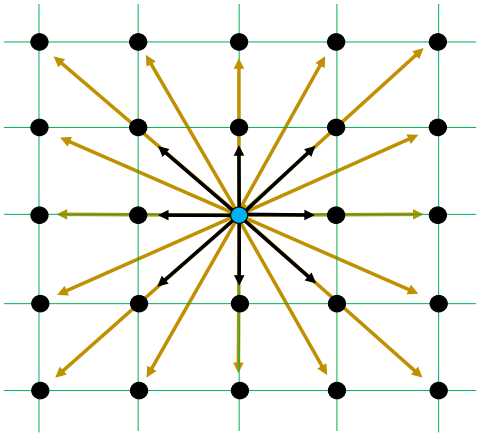

Figure 5.22: Second level forward star connections. A larger flexibility is allowed for rays.

When ray-tracing with the shortest path method is performed in an isotropic velocity field, the velocity values are assigned to the nodes of the grid and the shortest path is interpreted as the shortest travel time for the rays. The travel time of a ray segment between two nodes of the grid is defined as:

$$\Delta T_{pair\,i,j} = \frac{L}{mean(U[i], U[j])} \tag{5.1}$$

where L is the length between nodes i and j and U is the isotropic velocity field.

A point of force of the shortest path method is the possibility to extend it also to anisotropic media. Assuming a parametrization for the anisotropic velocity field in terms of an isotropic velocity U, an anisotropic magnitude fraction F and anisotropic orientation angles Γ (elevation) and Ψ (azimuth), the travel time between two nodes is modified (section 4) in:

$$\Delta T_{pair\,i,j} = \frac{L}{\bar{U} \times (1 + \bar{F} \times (2(cos(\theta)cos(\bar{\Gamma})cos(\phi - \bar{\Psi}) + sin(\theta)sin(\bar{\Gamma}))^2 - 1))} \tag{5.2}$$

where $\bar{U}$, $\bar{F}$, $\bar{\Gamma}$ and $\bar{\Psi}$ are the averages between the values assumed by U, F, Γ and Ψ in the nodes i and j. The angle θ is the elevation of the ray-segment, while ϕ is the azimuth.
In Fig. 5.23 ray-tracing with the shortest path is performed is an isotropic and an anisotropic velocity field.

Some strategies can be proposed to reduce the network geometry error: a possibility is to reduce the discretization size of the grid, which leads to more possible directions for the ray paths, but the computational cost can increase significantly. In alternative the network geometry can be revised (shortest path method with cell model, see section 5.3.4) or ray-bending algorithm can be adopted (section 5.3.5).

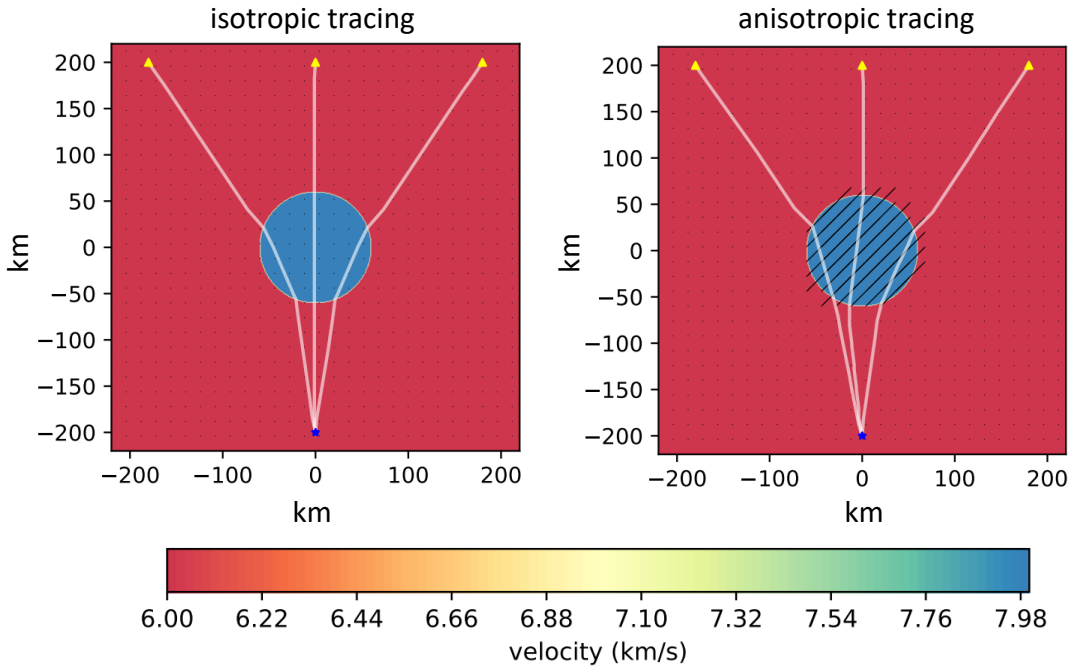

Figure 5.23: Left: Ray-tracing with the shortest path method (third level forward star) in an isotropic velocity field. Right: Ray-tracing with the shortest path method (third level forward star) in an anisotropic velocity field. Anisotropic magnitude fraction is set to 1% with 45° as azimuth. Grid generation script and Dijkstra's algorithm were written in Julia by Albert de Montserrat, ray-tracing was extended to anisotropic media in this thesis.

5.3.3 Synthetic experiments with the Shortest Path Method and grid model

To test the implementation of the shortest path method in the AnisoTOMO code, a simple cartesian synthetic experiment is realized using an isotropic velocity field (Fig. 5.24).
200000 iterations are performed, the first half is discharged as burn-in. Second level forward star is chosen. Reference velocity field is generated every 10000 iterations using twenty-eight Markov Chains. The inversion follows the scheme of Fig. 4.1. Events and stations are distributed with a highly idealized geometry and density.

The velocity field target of the inversion is well recovered (Fig. 5.25), the same is valid for the seismic rays traced. As expected, the fast velocity anomaly (blue) tends to make rays denser in it, while rays tend to bend around the slow velocity anomaly (red).
In Fig. 5.26 the uncertainties are presented; as can be expected, the largest uncertainty is associated to the slow velocity anomaly, resulting from the less dense sampling of the seismic rays bent around.
This experiment shows the successfully implementation of the shortest path method with a grid model in the AnisoTOMO code.

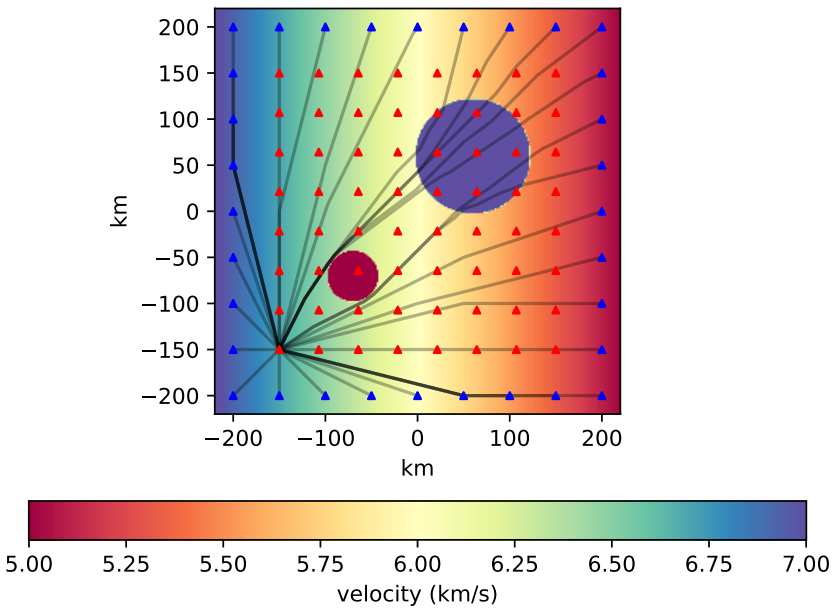

Figure 5.24: True model, used to computer synthetic data. This is the target of the inversion. Blue and red triangles are events and stations. Rays arriving to the first station are plot.

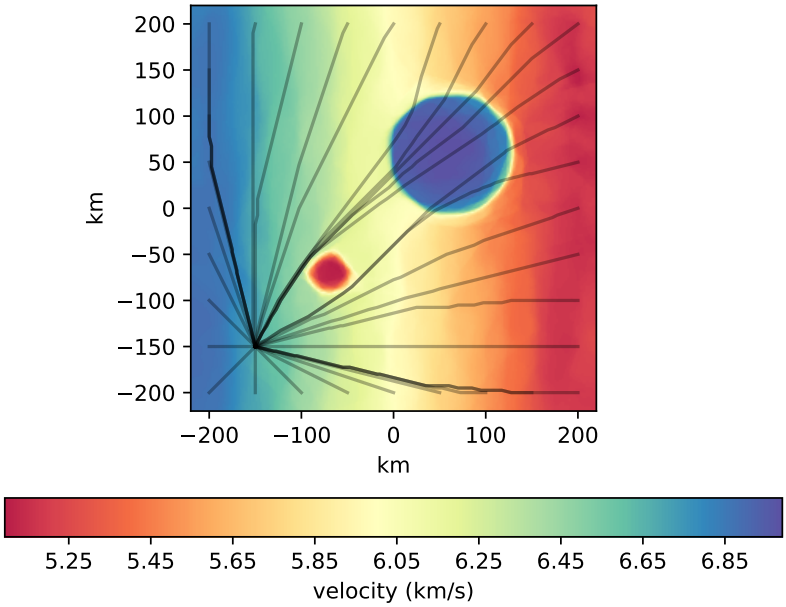

Figure 5.25: Model recovered using the inversion procedure. Initial model is homogeneous velocity (6 km/s) with straight lines for rays. Then trajectories are updated every 10000 iterations.

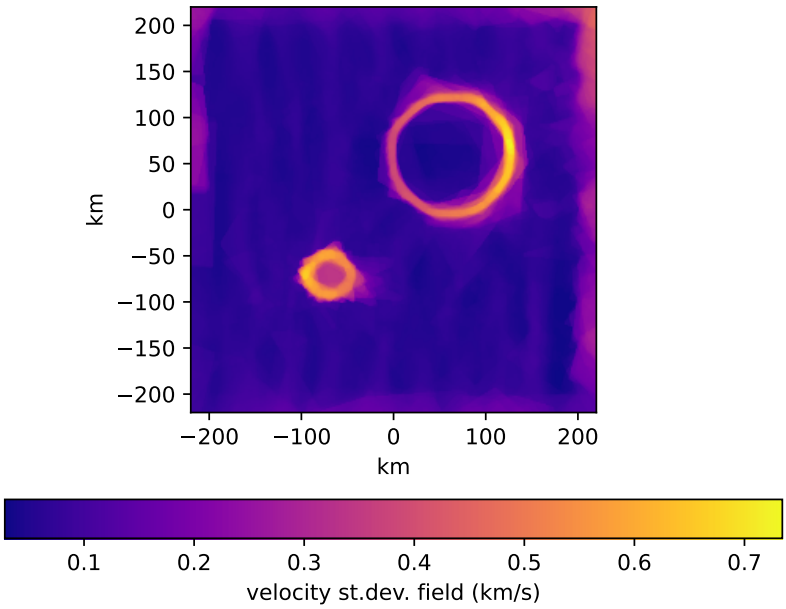

Figure 5.26: Standard deviation of the isotropic velocity field recovered.

5.3.4 Shortest Path Method with cell model

An alternative network geometry for shortest path method implementation is the *cell model*. Values of the fields are specified in regularly placed nodes, called *primary* nodes, which form cells. In two dimensions each cell is delimited by four primary nodes at the vertices, in three dimensions eight nodes are necessary to define a cell. Once the primary nodes have been generated and cells specified, *secondary* nodes are placed at the borders of the cells. With this network geometry, secondary nodes at the borders of the same cell are considered neighbours (multiple implementations are allowed, also using primary nodes as neighbours with secondary ones). As it is done for the grid model, multiple level forward star configurations can be implemented to increase the number of angles allowed for segments of rays.

For cell model, increasing the density of secondary nodes has been a typical solution to improve the accuracy in ray-tracing. In [Bai et al., 2007] cell model and grid model are compared and evidence shows that a lower number of nodes is required by cell model to trace rays with the same accuracy of the grid model.

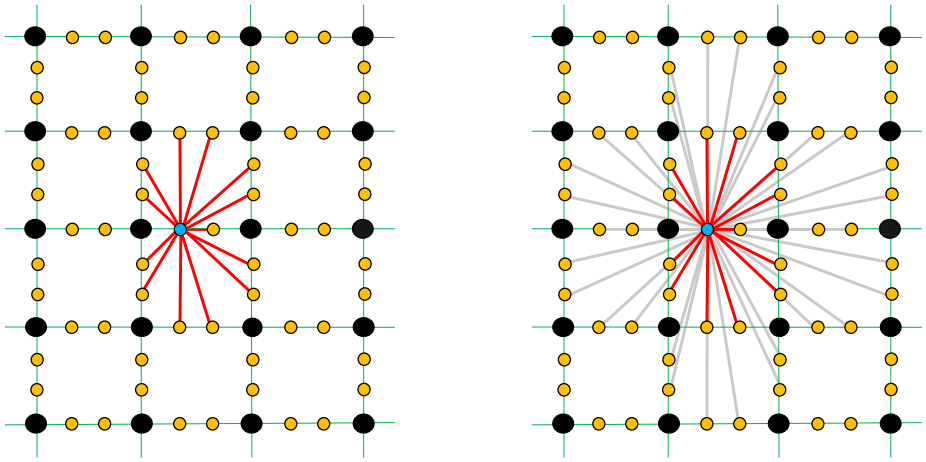

Figure 5.27: Cell model with two secondary nodes per border; left: first level forward star, right: second level forward star.

This geometry network was implemented for an annulus grid by Albert de Montserrat, where the shortest path method algorithm adopted is the Bellmann-Ford-Moore version, with the possibility to distribute the tasks among multiple cores. Annulus grid is too large for the model we are trying to invert, so a portion of this grid is isolated. In order to test the tracing algorithm, comparison with TauP toolkit for tracing in 1D reference field is studied.
Considering IASP91 as velocity field, a grid corresponding to the subduction model with secondary nodes is instantiated setting the number of primary nodes to 200x70. The selected spacing for secondary nodes is set to 1 km.

A bunch of rays is generated for epicentral distances between 2 and 8.5 degrees. In Fig. 5.28 the differences between the travel times predicted by SPM and TauP toolkit are represented. In this accuracy plot points labelled as *TauP - SP1*, *TauP - SP2*, *TauP - SP3* represent the differences in travel times computed using TauP toolkit and shortest path method with first, second and third level forward stars. As expected, to increase the level of the forward star increases also the accuracy of the travel times computed, but while difference is significant between first and second levels, it is not relevant between second and third.
According to these results, SPM method with secondary nodes and second level forward star provides an accurate estimate for travel times, differences computed are generally of the order of -150 ms, but almost systematic. This means that in real situations it is possible to adjust the origin time parameters of the events, which are generally known with an uncertainty of this order of magnitude, and reduce the discrepancy.

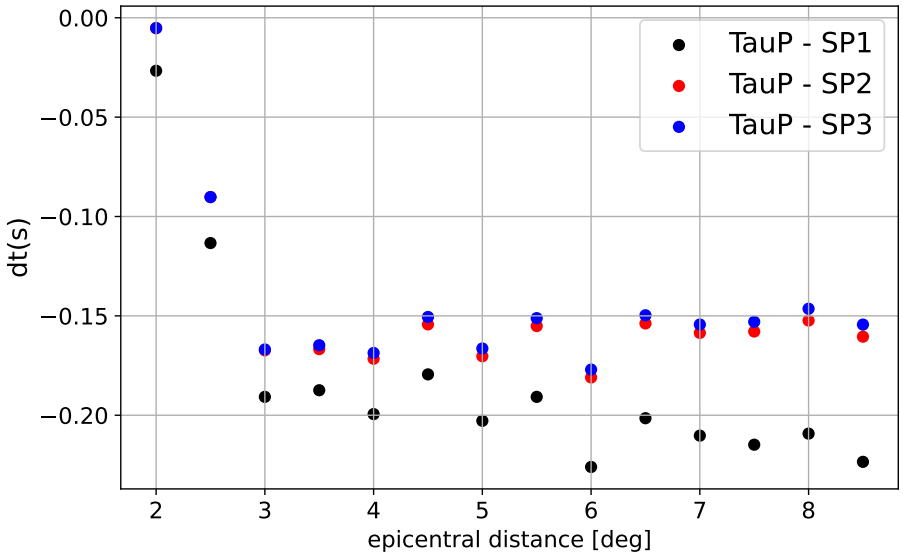

Figure 5.28: Differences between travel times computed using SPM and TauP toolkit. First, second and third level forward stars are implemented.

5.3.5 Ray-bending

Ray-bending is a ray-tracing technique which perturbs an initial guess of the ray trajectory toward the minimization of the travel time, according to the Fermat's principle.
Ray-bending method was introduced in seismology by [Wesson, 1971], [Chander, 1975], and [Julian & Gubbins, 1977]. Ray-tracing equations are written as:

$$\frac{d}{ds}\left(\frac{1}{c(\mathbf{r})}\frac{d\mathbf{r}}{ds}\right) - \nabla\left(\frac{1}{c(\mathbf{r})}\right) = 0 \tag{5.3}$$

where $\mathbf{r}$ is the position vector in the space, s is the arc parameter introduced along the ray direction and c($\mathbf{r}$) is the velocity field.
Two different approaches are proposed: the first proposes to discretize the ray in k+1 points $(\mathbf{r}_0, ..., \mathbf{r}_k)$ and to use finite difference method for the evaluation of the derivatives. Using an initial guess of the ray trajectory and writing 5.3 as a linearized system of equations, it is possible to perturb iteratively the initial guess in order to reduce to zero the left-hand term of the relationship. Limitations of this method arise mainly from the precision of the numerical differentiation. If the system of equations is ill-conditioned, as is could be for complex velocity models, the errors introduced by the numerical differentiation could prohibit the convergence of the algorithm.
The second approach [Um & Thurber, 1987] proposes to minimize the travel time T(γ) associated with the ray by acting on the ray path γ

$$T(\gamma) = \int_\gamma \frac{ds}{c} \rightarrow Min. \tag{5.4}$$

In this method the ray is discretized in k+1 points, connected using straight lines. The integration of the travel time is performed by summing all the contributions along the segments of the ray path.
This method is implemented by defining a travel time function with discretization points of the ray between event and receiver as variable. This function can be minimized using optimization methods provided by the Julia package *Optim*. The gradient of the travel time function can be evaluated using analytical calculations and provided to the optimization procedure, or it is approximated by the minimization algorithm using finite differences.
The bending method suffers a lot the optimization in complex velocity fields. In Fig. 5.29 the algorithm is performed in a continuous velocity field using two different initial guesses: a straight line between event and station and the result of the shortest path method. The results of the minimization are very similar, but the needed number of iterations in the first case is much larger than the second one.
In Fig. 5.30 there is an evidence sustaining the limitation of the ray-bending technique in discontinuous velocity fields. While in a continuous velocity field like the one in Fig. 5.29 the effect of the initial guess affects only the computational time, in this case it affects significantly the result of the optimization. Using a straight line as initial guess the ray path found with the procedure is not the global minimum of the travel time, which is found using a better initial guess, like the result of the shortest path method. This results allow to investigate the application of ray-bending technique in complex velocity fields (with discontinuities) only when it is coupled with the shortest path algorithm (Hybrid Shortest Path) or another algorithm which provides a good initial guess. Examples of coupling between SPM with grid model and ray-bending are shown in Fig. 5.31.

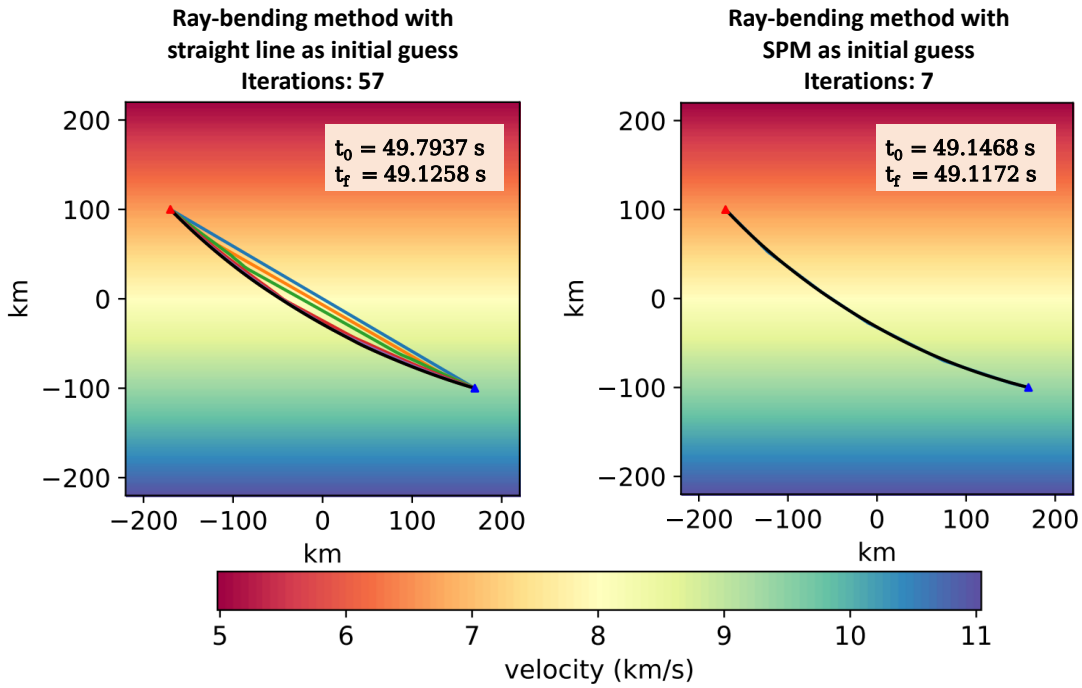

Figure 5.29: Results of bending technique applied to two different initial conditions in a continuous velocity field. Times t_0 and t_f are the initial e final (after ray-bending) travel times. Final travel times and paths are very similar with the two procedures, but with a significant difference in the number of iterations performed.

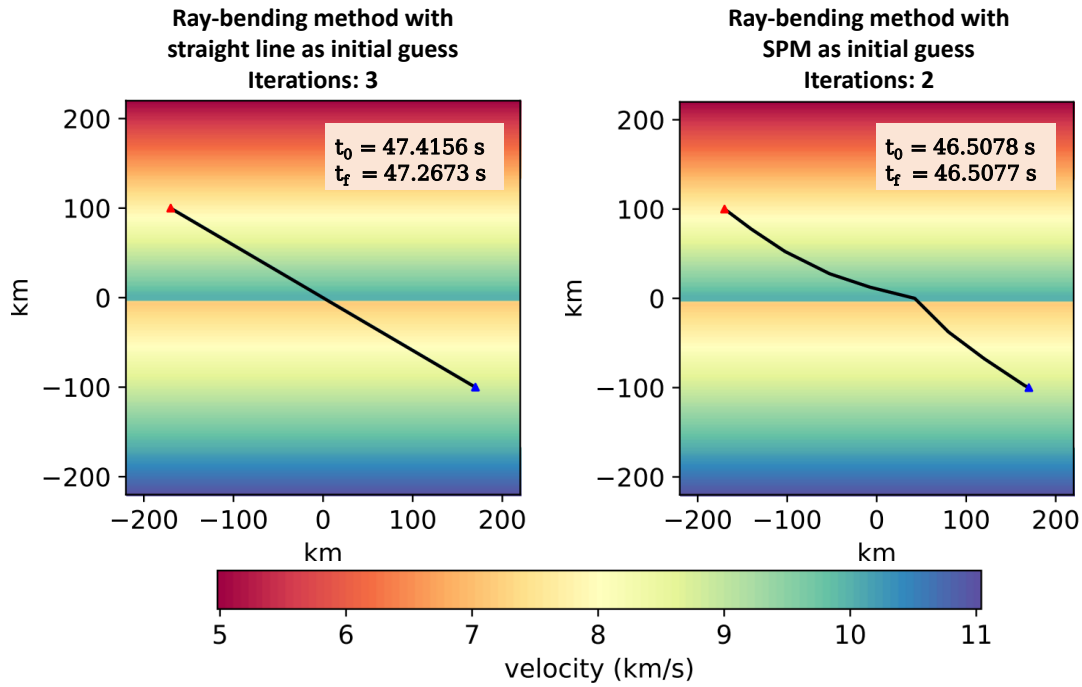

Figure 5.30: Results of bending technique applied to two different initial conditions in a discontinuous velocity field. The choice of the initial guess makes the first optimization to fail to find the absolute minimum of the travel time.

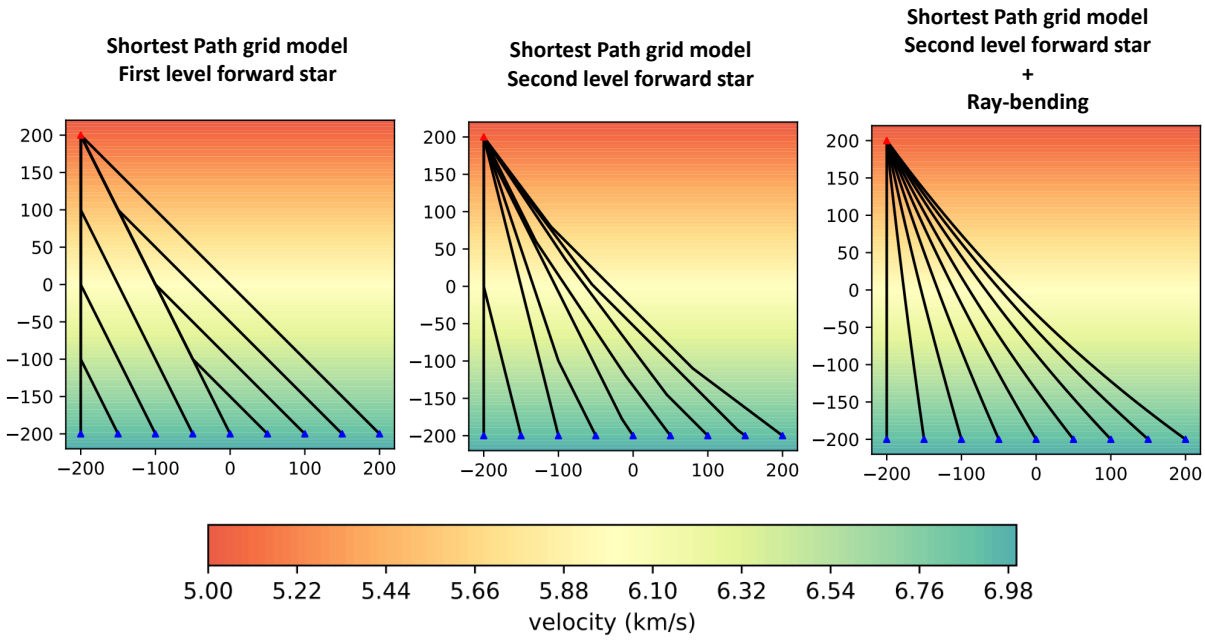

Figure 5.31: Ray-tracing using (left): Shortest Path method (first level) in a grid with spacing of 5 km, (center): Shortest Path method (second level) in a grid with spacing of 2.5 km, (right): Shortest Path method (second level) in a grid with spacing of 2.5 km and ray-bending. Difference in travel times between SPM and hybrid SPM (SPM + ray-bending) is of the order of 10-100 ms.

Coupling between shortest path method and raybending has been investigated in this project, but since accuracy of ray-tracing with cell model has proven to be acceptable, the computational cost required by the raybending technique was not worth the full implementation in the inversion method. Clearly this way is left open to future plans.

5.4 Joint inversion teleseismic waves - local earthquakes

On the basis of the motivations explained in section 5.2 more discrimination among the observations can be achieved using also local earthquakes for the inversion. Twenty local events are distributed approximately along the slab. Local rays have been traced in the true model using anisotropic SPM with cell model and the parameters adopted in the accuracy test (section 5.3.4), then using these paths the corresponding synthetic data have been computed in the ray-theory framework. Teleseismic waves are traced with TauP in IASP91 velocity field. Inversion takes 600000 iterations, every 20000 iterations the reference field is computed and SPM runs to update the trajectories of the local rays, leaving the teleseismic waves unchanged. For this experiment 28 Markov Chains are run in parallel.
In Fig. 5.33 the isotropic velocity field recovered is plot. Using local earthquakes the subducting structure is better constrained, even if we suspect that this result may improve with other selections for the parameters of the AnisoTOMO code or a denser coverage in local events. Lateral regions characterized by artifacts in the isotropic field have correspondence as high-uncertainty regions in the isotropic standard deviation field (Fig. 5.35), so they can be interpreted as products of the inversion algorithm.
Anisotropy distribution recovered is plot in Fig. 5.34, magnitude and orientation are better constrained with respect to the results with only teleseismic waves, including also horizontal structures. There is still a large stretching of anisotropic magnitude along vertical direction in the left portion of the plot, probably due to the limited number of local earthquakes introduced.
The RMS of the chains converges around 100-110 ms, the Gaussian noise added has standard deviation of 50 ms, so the models fit well the synthetic data, and this could lead to the conclusion that the limitations in the inversion could be overcome adding local events.
These results are encouraging to proceed the development of the code in this direction and evaluate how results of teleseismic tomography and joint tomography can be improved.

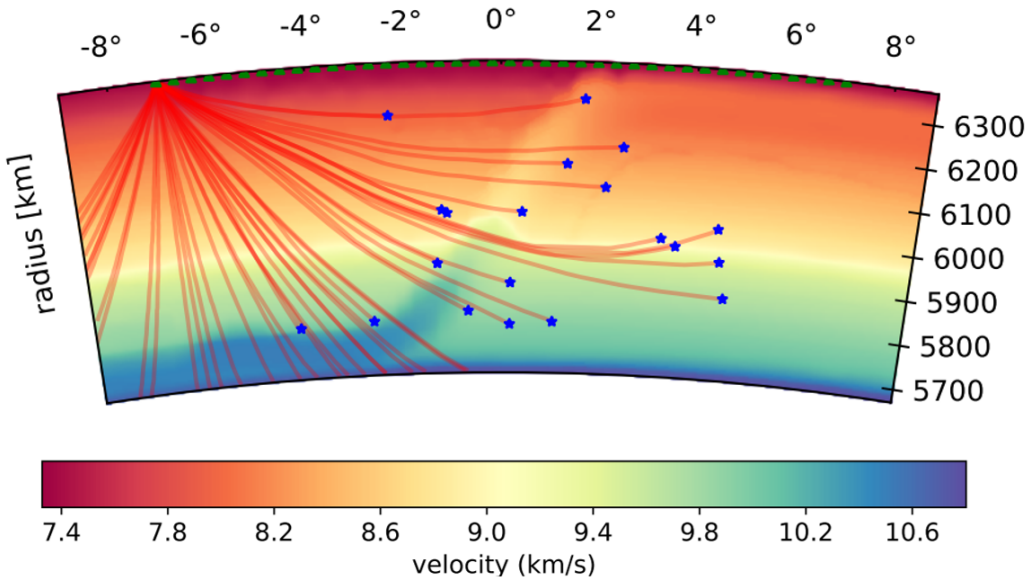

Figure 5.32: Ray paths in the domain. Local events introduced are the blue stars, trajectories are computed using anisotropic shortest path method

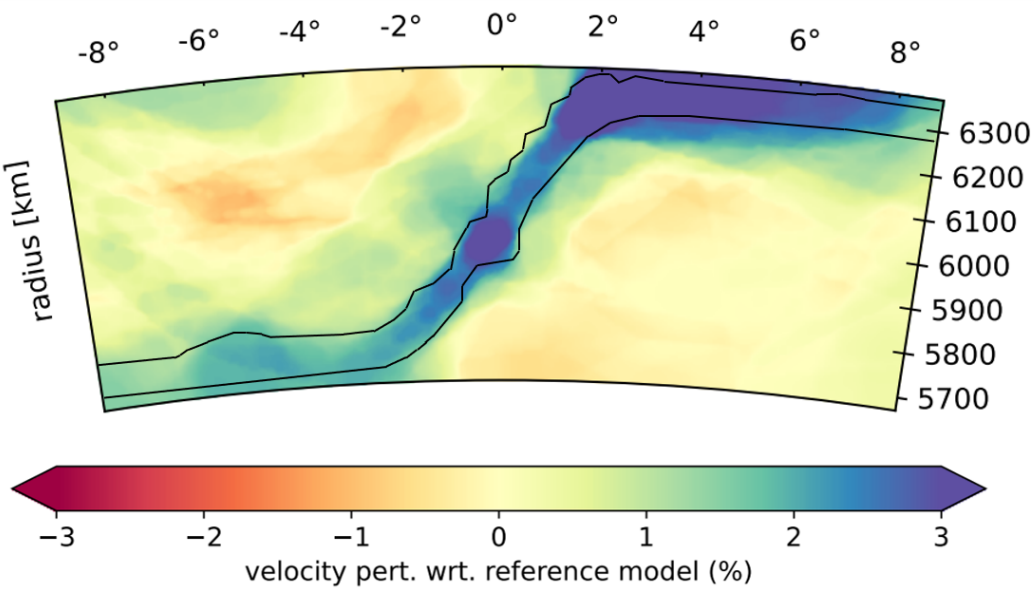

Figure 5.33: Isotropic perturbative velocity field recovered

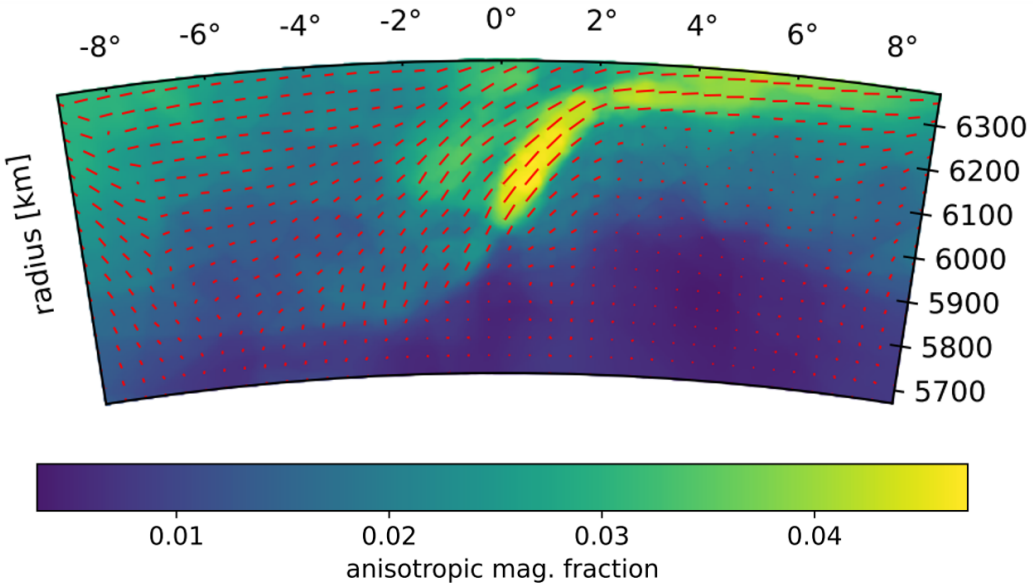

Figure 5.34: Anisotropy distribution recovered.

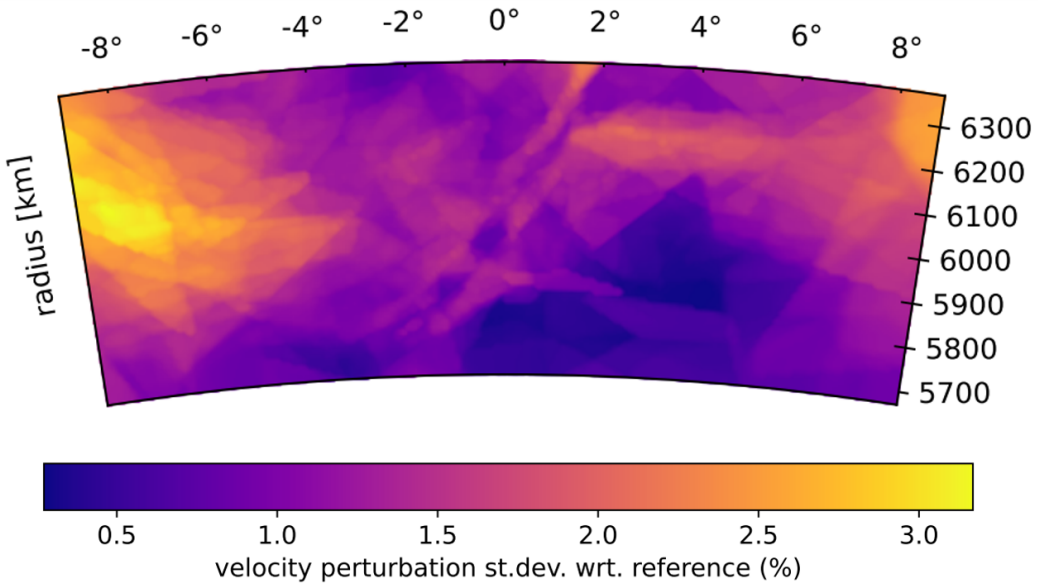

Figure 5.35: Standard deviation of the velocity field recovered.

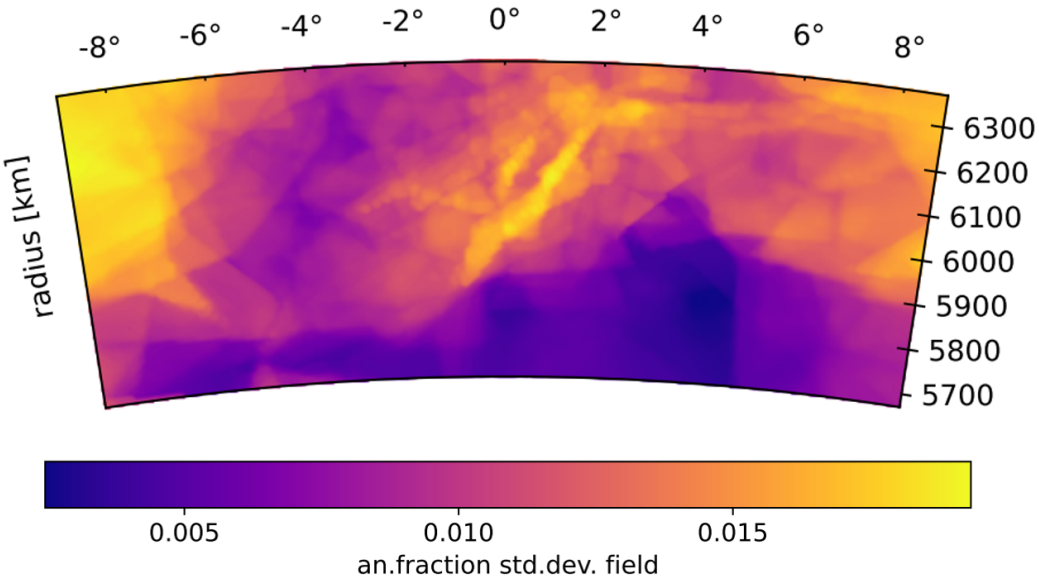

Figure 5.36: Standard deviation of the anisotropic fraction field recovered.

Chapter 6

Conclusions and future plans

The investigation carried out so far has shown the performances of the Bayesian inversion method in realistic two-dimensional geometries and amplitudes for subduction zones with anisotropic forward modelling. Limitations have arisen with the use of teleseismic waves in complex anisotropic patterns but some solutions are proposed to be tested in this direction: changes in the sampling procedure acting on the starting model or on the posterior distribution, use of local earthquakes to better constrain the anisotropic pattern.
This project has shown how this inversion technique can overcome the main limitations introduced by the standard non-linear optimization methods, at the cost of an increase in the computational time. One of the main point of force of the procedure is the possibility to produce uncertainty maps using the ensemble of models sampled. These maps can help to perform a statistical analysis on the resolved fields, identify artifacts of the inversion and evaluate outcomes.

In order to compare results of the inversion with the synthtetic three-dimensional model tested in [VanderBeek & Faccenda, 2021], the next step is the extension of the code to three dimensions. At the actual state inversion can take rays coming from all the three dimensions of the space, but the model is still 2D, this implementation is straightforward using teleseismic waves, while with local earthquakes a three-dimensional adapted grid needs to be implemented to trace rays with the shortest path method and the cell model.
Once the code is able to invert seismological data for a three-dimensional model and it is tested with synthtetic models, real observations from the Mediterranean area can be inverted to resolve anisotropic velocity fields corresponding to the region interested by the subducting slabs and surrounding areas. High quality data from local earthquakes in this region are provided by INGV (Istituto Nazionale di Geofisica e Vulcanologia).

Ringraziamenti

In questo luogo vorrei spendere alcune parole per ringraziare i professori Morelli e Faccenda e il Dott. VanderBeek per aver condiviso tempo e riflessioni con me durante questi mesi.

Ringrazio i miei cari genitori Elisabetta e Gianfranco, che così forte risplendono.

Bibliography

[Aki et al., 1977] Aki, K., Christoffersson, A., and Husebye, E. S. (1977), Determination of the three-dimensional seismic structure of the lithosphere, J. Geophys. Res., 82(2), 277– 296.

[Backus, 1965] Backus G.E., 1965. Possible forms of seismic anisotropy of the uppermost mantle under oceans, J. geophys. Res., 70(14), 3429–3439.

[Bai et al., 1991] Bai Q, Mackwell SJ, Kohlstedt DL. 1991. High temperature creep of olivine single crystals. 1. Mechanical results for buffered samples. J. Geophys. Res. 96:2441–63.

[Bai et al., 2007] Bai, C. Y., Greenhalgh, S. Zhou, B., 2007. 3D ray tracing using a modified shortest-path method, Geophysics, 72(4), T27-T36.

[Becker et al., 2006] Becker T.W., Chevrot S., Schulte-Pelkum V., Blackman D. K., 2006. Statistical properties of seismic anisotropy predicted by upper mantle geodynamic models, J. geophys. Res., 111(B8).

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

[Byrnes & Bezada, 2020] Byrnes, J. S., Bezada, M. (2020). Dynamic upwelling beneath the Salton Trough imaged with teleseismic attenuation tomography. Journal of Geophysical Research: Solid Earth, 125, e2020JB020347.

[Chander, 1975] Chander, R.: On tracing seismic rays with specified end points. J. Geophys. 41, 173-177, 1975.

[Crotwell et al., 1999] Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, Seismological Research Letters 70, 154–160.

[Dijkstra, 1959] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. Numerische Mathematik, 1(1), 269–271.

[Dziwsonski et al., 1977] Dziewonski, A. M., Hager, B. H., and O'Connell, R. J. (1977), Large-scale heterogeneities in the lower mantle, J. Geophys. Res., 82(2), 239– 255.

[Fang et al., 2019] Fang, Hongjian et al. "Parsimonious Seismic Tomography with Poisson Voronoi Projections: Methodology and Validation." Seismological Research Letters (2019).

[Faccenda, 2014] Faccenda M., 2014. Mid mantle seismic anisotropy around subduction zones, Phys. Earth planet. Inter., 227, 1–19.

[Faccenda et al., 2019] Faccenda, M., Ferreira, A. M. G., Tisato, N., Lithgow-Bertelloni, C., Stixrude, L., Pennacchioni, G. (2019). Extrinsic elastic anisotropy in a compositionally heterogeneous Earth's mantle. Journal of Geophysical Research: Solid Earth, 124, 1671– 1687.

[Green, 1995] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82(4), 711–732.

[Hansen et al., 2021] Lars N. Hansen, Manuele Faccenda, Jessica M. Warren, A review of mechanisms generating seismic anisotropy in the upper mantle, Physics of the Earth and Planetary Interiors, Volume 313, 2021, 106662, ISSN 0031-9201.

[Hastings, 1970] Hastings, W.K. (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 57, 97-109.

[Julian & Gubbins, 1977] Julian, B.R. and Gubbins, D. (1977). Three-dimensional seismic ray tracing. J. Geophys. 43, 95–113.

[Karato et al., 2008] Geodynamic Significance of Seismic Anisotropy of the Upper Mantle: New Insights from Laboratory Studies Shun-ichiro Karato, Haemyeong Jung, Ikuo Katayama, Philip Skemer Annual Review of Earth and Planetary Sciences 2008 36:1, 59-95.

[Kennett, 1991] Kennett, B.L.N. (Compiler and Editor). 1991. "IASPEI 1991 Seismological Tables." Bibliotech, Canberra, Australia, 167 pp.

[Metropolis & Ulam, 1949] Metropolis, N. and Ulam, S. (1949) The Monte Carlo Method. Journal of the American Statistical Association, 44, 335-341.

[Metropolis et al., 1953] Metropolis, N., et al. (1953) Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics, 21, 1087-1092.

[Nakanishi & Yamaguchi, 1986] Nakanishi, I. Yamaguchi, K., 1986. A numerical experiment on nonlinear image construction from first-arrival times for two-dimensional island arc structure, J. Phys. Earth, 34, 195–201.

[Nolet, 2008] Nolet, G. (2008). A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun.

[Paullsen, 2014-15] Paullsen, H., Handout seismic anisotropy, Theoretical Seismology course 2014-2015, Utrecht University.

[Ryberg, 2018] Ryberg, T., Bayesian approach in seismic tomography, Seismic Tomography for Non-Seismologists: Methods Potential Applications to Imaging Mountain Belts (February 2018)

[Sambridge et al., 2006] Sambridge, M., Gallagher, K., Jackson, A., Rickwood, P. (2006). Trans-dimensional inverse problems, model comparison and the evidence. Geophysical Journal International, 167(2), 528–542.

[Tarantola, 2004] Tarantola A., Inverse Problem Theory and Methods for Model Parameter Estimation. December 2004.

[Thomsen, 1986] Thomsen L., 1986. Weak elastic anisotropy, Geophysics, 51(10), 1954–1966.

[Tommasi & Vauchez, 2015] Tommasi, Andréa Vauchez, Alain. (2015). Heterogeneity and anisotropy in the lithospheric mantle. Tectonophysics. 661. 10.1016/j.tecto.2015.07.026.

[Turcotte & Schubert, 2002] Turcotte, D.L. and Schubert, G. (2002) Geodynamics. Cambridge University Press, Cambridge.

[VanderBeek & Faccenda, 2021] Brandon P VanderBeek, Manuele Faccenda, Imaging upper mantle anisotropy with teleseismic P-wave delays: insights from tomographic reconstructions of subduction simulations, Geophysical Journal International, Volume 225, Issue 3, June 2021, Pages 2097–2119.

[Um & Thurber, 1987] Um, Ju-Hwan and Clifford H. Thurber. "A fast algorithm for two-point seismic ray tracing." Bulletin of the Seismological Society of America 77 (1987): 972-986.

[Wesson, 1971] Wesson, R.L.: Travel-time inversion for laterally inhomogeneous crustal velocity models. Bull. Seism. Soc. Amer. 61, 729-746, 1971

[Zhang & Karato, 1995] Zhang, S., Karato, Si. Lattice preferred orientation of olivine aggregates deformed in simple shear. Nature 375, 774–777 (1995).