Università degli Studi di Bologna

DIPARTIMENTO DI FISICA

DOTTORATO DI RICERCA IN GEOFISICA

GEO/10

Seismic Rays and Travel Time Tomography in Strongly Heterogeneous Mantle Structures

Dottoranda:

Dott.ssa Paola Serretti

Tutore:

Dott. Andrea Morelli

XIX CICLO

ANNO ACCADEMICO 2006/2007

Ringraziamenti

Un sincero ringraziamento va ad Andrea per avermi dato la possibilita' di intraprendere questo lavoro nelle migliori condizioni e per avermi sempre seguito con enorme pazienza. Ringrazio Peter per l'entusiasmo e il tempo che ha dedicato a questo lavoro, ringrazio Nicola, per avermi salvato nei momenti di "panico informatico" quando tutto sembrava perso e ringrazio tutti coloro che qui nell'istituto che mi hanno offerto il loro aiuto. Un ringraziamento va a Jean Virieux, per tutte le cose che mi ha insegnato nei tre mesi trascorsi a "Le Laboratoire Geoazur".

Ringrazio tutti gli amici dell'istituto con i quali ho condiviso buona parte delle mie giornate...a partire dal secondo risveglio (quello delle 10) alle macchinette del caffe'... grazie Simone, Simona, Isabelle, Renata, Peter, Reneta, Nicola e tutti quanti.

Ringrazio Silvia, Lisa e Chicco per avermi sempre incoraggiato e sostenuto (anche con ottime cenette). Ringrazio tutti gli amici, i piu' vicini, i piu' lontani, per tutti i bei momenti...

Contents

Chapter 1

Introduction

Seismic data represent one of the most valuable sources of information about the internal composition of the Earth. The goal of seismic tomography is the determination of elastic structure, in its three dimensions, at very diverse scales. Different component of the seismic wave record, like travel times, waveform amplitude or spectra, lead to different methods that can be used for determining earth structure [Rawlinson and Sambridge, 2003].

This work follows an approach, known as *seismic travel time tomography*, which is based on a relation between the travel time recorded by a seismogram and the wave speed along the ray path, so that the wave velocity pattern in a medium can be reconstructed by the knowledge of a large set of source-station travel times. Both temperature and composition may influence wave speeds, but most of the mantle velocity variation can probably be attributed to changes in temperature [Ranalli, 1996] while the effect of mantle composition contributes only less than 1% [Sobolev et. al., 1997], [Goes et. al., 2000].

As a wave propagates faster in a cold material and slower in a warm material, images of the wave speed behaviour in a medium, give us insight about the geodynamic and tectonic process involving the earth. Following seismic ray theory, the ray path depends on the velocity structure itself, so that the travel time does not depend linearly on the wave speed. Because of this non-linearity, the inverse problem can be very difficult to solve.

A linearization assumption, consisting of the use of reference ray paths calculated in a 1-D background velocity medium, is commonly used for global tomographic studies but

it has to be considerate valid only when the real model is not so different from the 1-D model. On the contrary the method result inadequate, when strong velocity variation are present. For this reason, many local and regional crust and lithosphere studies are performed with the *non-linear tomography* approach [Thomson & Gubbins, 1982], [Thurber, 1983]. It consists in an iterative scheme in which inversion step is alternated with the 3-D ray tracing step, done, for each successive iteration, in the updated model. Although, this kind of process is time demanding and computationally expensive when applied to large scale studies, recently, due to the more powerful computer availability, this approach has been used also in regional and global scale tomography [Bijwaard and Spakman, 2000], [Gorbatov et. al., 2001].

In order to perform the 3-D rays tracing, various method can be used. As for the general tomographic problem, it is necessary to make a choice between which is faster and which is more accurate, choice that is constrained by our computing resources available and the accuracy necessary. Usually global 3-D ray tracing refers to methods, like ray bending [Červený, 2001], that are quite fast but that fail when strong heterogeneous structures are present. For this reason, local seismic tomographic studies often use instead numerical methods that first compute the wavefront with an accurate numerical scheme and consequently trace rays. Since the latter is computational more expensive, especially for a larger area, the choice must consider our mantle knowledge or what we expect for it. Although, recent tomographic models, show quite smooth lateral velocity variation, speed changes with radius are more significant, and travel times curves are complicated by triplications due to discontinuities. This induces to think that an accurate ray tracing method, capable of tackling strong non-linearities in the dependence of ray path on the seismic structure, may be relevant also for upper mantle studies.

This thesis presents the tomographic method devolopped and an application to the Euro-Mediterranean area. The method is based on the iterative non linear inversion in which the forward computation is done with numerical schemes. The second chapter is a review of the known basic concepts of the forward and inverse calculation. The third exposes the finite difference method, we use for the wavefront and relatives rays computation. Since this part is implemented in a cartesian grid, chapter four explains first how it can be applied to a spherical geometry and then how to compute all the 'ingredients' we need for the inversion step. Chapter five shows two applications referred to the forward finite difference computation. The last two chapters show two applications

of the iterative non linear inversion. Chapter 6 describes the seismic data set used and the inversion method. This part shows also a recovery test done in order to test the robustness of the method. Finally, we present the application to the Euro-Mediterranean area.

Chapter 2

Seismic Travel Time Tomography

When an earthquake occurs, it releases energy in the form of heat and seismic waves. Because of the elastic properties of Earth materials and the presence of the free surface, four main types of seismic waves can be identified on a seismogram. Body waves radiated by the source include *Compressional (P)* and *Shear (S) waves* that propagate through the Earth's interior. Their interaction with the Earth's surface generates the *surface waves* distinguished in *Rayleigh* and *Love waves*. They propagate primarily at and near the Earth's surface with an amplitude that decreases with depth. Each type of wave travels with a different speed and with different motions. Body waves have the advantage to go through all the Earth, therefore they can be used to study deeper structures. P waves, in particular, since they travel faster than the others, are the first seismic waves recorded on a seismogram and for this reason they are easy to identify. In this work we focus our attention on the P wave travel time tomography. The goal of the method is the reconstruction of the seismic wave velocity field knowing the first arrival travel times recorded by the seismic stations.

The travel time of a seismic ray can be expressed as

$$t = \int_{L(v)} \frac{1}{\mathbf{v}(\mathbf{x})} dl \tag{2.1}$$

where L is the ray path and $\mathbf{v}(\mathbf{x})$ the wave velocity field. The dependence on medium properties can be simplified by using slowness $\mathbf{c} \equiv \frac{\mathbf{1}}{\mathbf{v}}$ instead of $\mathbf{v}$, but, since the integration path depends on the velocity (slowness), this equation remains non-linear and consequently the inverse problem of finding the wave velocity (slowness) field can be difficult to solve. The approaches, commonly used, are:

-linear tomography
-iterative non-linear tomography
-fully non-linear tomography

The third is the only that finds a solution without any linearization assumption, but it is computationally extremely expensive, so that its application is limited to relatively small scale inversions involving a few tens of parameters at most. In the next section, initially the basic concept of linear tomography (involved also in the iterative non-linear tomography) will be described and finally the difference between the first two approaches will be explained.

2.1 Overview of Basic Concepts

The tomographic method is based on the relationship between data $\mathbf{d}$ and model parameters $\mathbf{m}$. If we define $\mathbf{m}$ the vector of seismic velocity parameters describing the medium and $\mathbf{d}$ a set of seismic wave traveltimes, the value of each element of $\mathbf{d}$ can be predicted by integration through the model, and we can write:

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) \tag{2.2}$$

For an observed dataset $\mathbf{d^{Ob}}$ and an initial model $\mathbf{m^0}$ the difference $\mathbf{d^{Ob}} - \mathbf{g}(\mathbf{m^0})$ gives an indication of how well the current model predictions satisfy the data. The tomographic method consists in manipulating $\mathbf{m}$ in order to minimize the difference between observed and predicted data. The steps we need to produce a tomographic image from seismic data can be summerized as model parameterization, forward calculation and inversion.

2.1.1 Model Parameterization

The medium has to be represented by a model vector that describes the seismic wave velocity or its inverse (the slowness) and each element may be referred to a cell or to a node.

2.1.2 Forward Calculation

In order to solve the forward problem, $\mathbf{d} = \mathbf{g}(\mathbf{m})$, we need to compute ray traveltimes between the source and the receiver through a given velocity structure. The wave propagation can be thought in terms of the wavefronts or in terms of the ray paths, that are, by definition, everywhere normal to the wavefronts. In the first case the propagation is described by surfaces of constant time, with the equation $T = T(x)$, while in the second case by a curve x=x(T). The traditional means of determining source-receiver traveltimes is based on the second approach with the ray tracing methods [Červený, 2001], [Rawlinson and Sambridge, 2003]. Recently also wavefront tracking scheme, based on a finite difference solution of the eikonal equation, have been employed [Qin et al., 1992], [Vidale, 1988, 1991]. We are going to briefly describe and compare the two strategies.

Ray Shooting and Ray Bending

Shooting methods are based on the initial projection angle. The problem is solved by shooting rays from the source trough the medium. The take off angle at the source is adjusted until the final ray passes sufficiently close to the receiver. In Ray Bending the problem is solved by adjusting the geometry of an initial ray between source and receiver until it becomes a "true" ray that satisfies Fermat's principle.

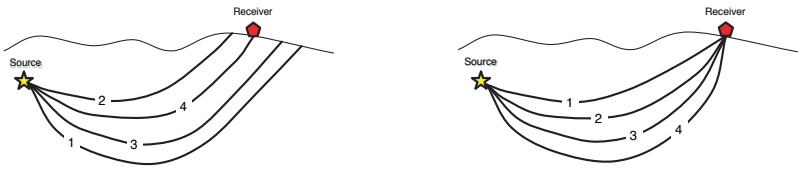


FIGURE 2.1: Schematic illustration of the principle of shooting method (left) and bending method (right)

Wavefront tracking At first it is necessary to find the entire wavefront path in all the medium. Secondly the ray path can be found following the travel time gradient from station to source. Usually wavefront tracking employs finite-difference solutions of the eikonal equation on a regular grid.

How to obtain sufficient accuracy without compromising speed and robustness is often a problem. The trade-off between speed and accuracy is evident in the choice between the

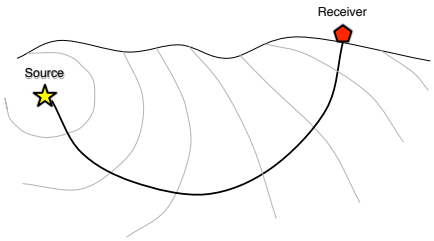


FIGURE 2.2: Schematic illustration of the principle of wavefront tracking method

two strategies described. The numerical eikonal solver, although computationally expensive (especially for a large area), is able to deal with strongly heterogeneous structure. It is often used in local tomography. The first approach, faster, is based on adjustments of an initial guess and fails in the presence of shadow zones or strong velocity variations if the initial guess is not close enough to the actual ray.

2.1.3 Inversion

The inversion step involves the adjustment of the model parameters $\mathbf{m}$ to better satisfy the observed data $\mathbf{d^{Ob}}$. Each element of $\mathbf{m}$ can be written as

$$m_j = m_j^0 + \Delta m_j \tag{2.3}$$

where $\mathbf{m^0}$ is the initial guess of model vector. If the forward problem is weakly nonlinear ($\mathbf{d} = \mathbf{g(m)} = \mathbf{Gm}$) and we know with good approximation the initial model $\mathbf{m^0}$ (model $\mathbf{m}$ is not so different from $\mathbf{m^0}$), the data vector $\mathbf{d}$ can be written, expanding the function $g(\mathbf{m})$ in a Taylor series and keeping only the linear term, as

$$\begin{aligned} \mathbf{d} &\simeq \mathbf{g(m^0)} + \frac{\partial \mathbf{g}}{\partial \mathbf{m}} \mid_{\mathbf{m^0}} \mathbf{\Delta m} \\ &\simeq \mathbf{d^0} + \frac{\partial \mathbf{g}}{\partial \mathbf{m}} \mid_{\mathbf{m^0}} \mathbf{\Delta m} \end{aligned} \tag{2.4}$$

Each element of the data vector $\mathbf{d}$ can be written in terms of the difference between observed and predicted data,

$$\Delta d_i^0 \equiv d_i - d_i^0 \simeq \sum_j \frac{\partial g_i}{\partial m_j} \mid_{\mathbf{m^0}} \Delta m_j \tag{2.5}$$

In travel time tomography, working with slowness $\mathbf{c}$ rather than velocity, equation 2.1 can be written with the current notation:

$$d_i = g_i(m) = \int_{L_i(c)} cdl = \int_{L_i(m)} mdl \tag{2.6}$$

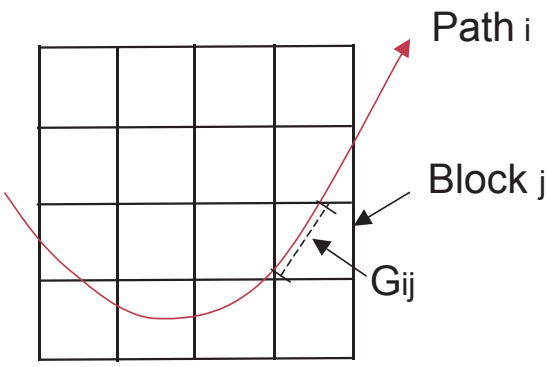


FIGURE 2.3: Schematic illustration of the problem's geometry. The region is divided into blocks j. G_{ij} rapresents the distance the i^{th} ray travels in the j^th block.

If the slowness field is discretized in blocks, or volume elements (*voxels*), $\frac{\partial d_i}{\partial m_j} \equiv G_{ij}$ represents the distance run by the i^{th} ray in the j^{th} block (Figure 2.3). Equation 2.5 becomes

$$\Delta d_i = \sum_j G_{ij} \Delta m_j \tag{2.7}$$

or in matrix notation

$$\mathbf{\Delta d = G \Delta m} \tag{2.8}$$

where G is the partial derivatives matrix.

In travel time seismic tomography the system described by equation 2.8 is, generally, overdeterminated, in fact the number n of data $\mathbf{\Delta d}$ is larger than the number m of unknown model parameters $\mathbf{\Delta m}$. Therefore system 2.8 does not have an exact solution and one common way to solve the inverse problem is to find its *least squares solution*, $\mathbf{\Delta m_{LS}}$, which is the vector that minimizes the misfit between observed and predicted data.

$$\Phi = \| \mathbf{G \Delta m - \Delta d} \|^{\mathbf{2}} \qquad \mathbf{\Phi = min} \Longleftrightarrow \mathbf{\Delta m \equiv \Delta m_{LS}} \tag{2.9}$$

The least squares solution to a linear inverse problem is legitimate when data errors are Gaussian [*Tarantola*, 2005]. It can be shown that $\mathbf{\Delta m_{LS}}$ is the exact solution to,

$$\mathbf{G^T \Delta d = G^T G \Delta m} \tag{2.10}$$

$$\mathbf{\Delta m = (G^T G)^{-1} G^T \Delta d} \tag{2.11}$$

However, due to faulty data coverage, the problem is often also underdeterminated: many (infinite) model can equally well reproduce the data. In such a case, matrix $\mathbf{G^T G}$ is singular and cannot be inverted. Additional conditions, such as jointly minimizing model norm, allow to compute a generalized inverse G_g^{-1}. We can therefore write:

$$\Delta\mathbf{m} = \mathbf{G^{-g}}\Delta\mathbf{d} \tag{2.12}$$

In order to use this method we begin with a starting model $\mathbf{m^0}$ and predict the values expected for the data $\mathbf{d^0} = \mathbf{g}(\mathbf{m^0})$. We then compose the residual vector $\Delta\mathbf{d^0} \equiv \mathbf{d^{Ob}} - \mathbf{d^0}$ and evaluate the matrix of partial derivatives around the starting model and use the (2.12) to find the $\mathbf{\Delta m}$. Thus the new model becomes

$$\mathbf{m^1} = \mathbf{m^0} + \mathbf{\Delta m} \tag{2.13}$$

As written at the beginning of the chapter, ray path depends on the velocity structure, this means that the data do not depend linearly on the model parameters. Depending on this non-linearity, inverse problem can be solved using some linearization assumption.

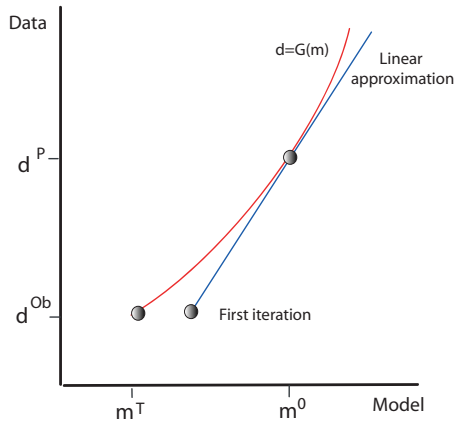


FIGURE 2.4: Schematic illustration of the effect of linearizing about a starting model (m_0). In order to find the new model, we consider difference between the observed data (d_{Ob}) and the predicted (d_P)from the starting model. The worse the linear approssimation is, the more iterations will be needed to reach the true model (m_T).

Linear Tomography

As described in the step 3, the relationship between traveltime residual and velocity perturbation is linearized around a reference model (the matrix of partial derivatives was found by expanding the function that predict the data about the starting model in a Taylor series (equation 2.5), and taking only the linear terms). Ray paths are determined only once and equation 2.13 represents the solution to the problem. This approach assumes that our initial guess $\mathbf{m^0}$ is actually close to the true model $\mathbf{m^T}$ (Figure 2.4) or that, at least, $\mathbf{g(m)}$ is almost linear between $\mathbf{m^0}$ and $\mathbf{m^T}$.

Iterative Non-Linear Tomography

This approach assumes that the ray path is weakly dependent on the velocity and accounts for the non-linearity by iteratively applying corrections and re-tracing rays. Once we have obtained the $\mathbf{m^1}$ model (equation 2.13) the forward computation and the inversion are performed again, therefore we obtain $\mathbf{m^2} = \mathbf{m^1} + \mathbf{\Delta m^1}$. As shown in fig. 2.4 the approximation works well only if $\mathbf{g(m)}$ is sufficiently well-behaved, i.e. the non-linearity is not too strong, and it does not require knowledge of an initial model $\mathbf{m^0}$ very close to the true model $\mathbf{m^T}$ as linear tomography does. The method need to be iterated until convergence-successive iterations produce only small changes in the model, and in the total misfit to the data.

More details about the iterative non-linear inversion scheme that we followed can be found in Section 6.4.

Chapter 3

Finite difference travel time calculation

For an elastic medium, the propagation of seismic wavefronts can be described by the eikonal equation [Aki and Richards, 2002] :

$$(\nabla_{\mathbf{x}} T)^2 = \frac{1}{v(\mathbf{x})^{\mathbf{2}}} \tag{3.1}$$

where $\mathbf{T}(\mathbf{x})$ is the traveltime of the wavefront , $\mathbf{x}$ is the vector of spatial coordinates and $v(x)$ is the velocity in $\mathbf{x}$. This description of wave propagation is valid in the infinite frequency limit. This high frequency assumption assumes that the wavelength of a seismic wave should be much less than the length scale of the velocity variations of the medium through which it passes. If travel time at time t is described by the equation $T_t = T(x)$, the wavefront at time $t + \Delta t$ will be described by $T_{t+\Delta t} = T(x')$ that describes the new geometry and position after a time Δt. Huygens principle states that this wavefront is predictable from the wavefront at time t by considering each point on the known wavefront as a source for a secondary wavefront. The front is assumed to propagate with constant velocity using the velocity at the source point. This is an appropriate approximation if we take a sufficiently small time step. The future wavefront is constructed from the superposition of all possible Huygens' wavefronts.

3.1 Travel Time Field Computation

The wavefront computation is performed using an algorithm proposed by Podvin and Lecompte (1991), P&L. Their approach improved Vidale's finite difference algorithm [Vidale, 1988], [Vidale, 1990]. Although the P&L scheme is inspired from Vidale's approach, it is able to be applied to more contrasted velocity models and it has the advantage to be designed as a massively parallel scheme, although it may also be implemented on a conventional computer.

The method needs a discretization of the velocity model in a cartesian grid with cubic cells. A constant slowness is associated to each cell. The computation relies on a systematic application of Huygens' principle in the finite difference approximation. To each node a time is associated chosen among all the travel times from the adjacent 26 nodes with a minimum time criterium (see figure 3.1). The method is able to take into account the existence of different propagation modes (transmitted diffracted body wave and head waves).

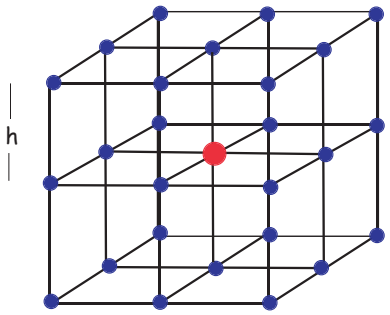


FIGURE 3.1: Data structure used in 3-D. First arrival time is computed at the central grid-point from the surrounding box point (considered as a continuous set of Huygens' source).

The propagation start from the source point, located anywhere in the model. An important feature of seismic travel times is that, if the source and the receiver switch place, the travel times measured in both directions must remain the same. Because it is important that the algorithm obeys reciprocity we use the updated version of the P&L scheme corrected by [Tryggvason & Bergman, 2006]. Due to an error in the point

source initialization, they demonstrated that the travel times computed were different depending if the travel times are computed to the left or to the right along the model discretization.

According to the Huygens principle, travel times precision is dependent on the grid step. Mesh spacing should be at least one order of magnitude smaller than characteristic dimensions of slowness anomalies considered. The choice of the right step size depends on the model dimension together with the computational resources available. The smaller the step is, the smaller will be the error, to the detriment of a large computational effort.

3.2 Rays computation

Once the wavefront is propagated, we know a travel time in each node and rays can be backtracked from station to source following the direction of traveltimes gradient. Travel times partial derivatives $\frac{\partial u(x,y,z)}{\partial x}$, $\frac{\partial u(x,y,z)}{\partial y}$ and $\frac{\partial u(x,y,z)}{\partial z}$ are computed using a finite difference approach and are indicated, respectively, with $u_x|_{rst}$, $u_y|_{rst}$ and $u_z|_{rst}$ (notation and principal formulas are derived in Appendix A). Rays are constructed using a piecewise linear path (see Figure 3.2). Let us assume we arrive at point P, in order to find a new step direction, we look for the angle identified by the three gradient components, $u_x|_P$, $u_y|_P$ and $u_z|_P$. In order to find each component of the gradient, we can use various numerical algorithms which involve only the adjacent nodes or a larger neighbourhood of nodes. We describe three of these methods and their behaviour. For sake of simplicity, figures and descriptions are done for the 2-D case. The 3-D case is the direct consequence.

- First Method: *FD* 1^{st} *order scheme*
  Initially the partial derivative u_x is evaluated in A and B respectively (Figure 3.2), dividing by the step grid h the traveltime difference $T7 - T6$ and $T11 - T10$, then the value in P, $u_x|_P$, is computed using a linear interpolation between the two points. The same approach is followed to find u_y in C and D (dividing by the grid step h the traveltime difference $T10 - T6$ and $T11 - T7$) and then $u_y|_P$ is computed by the interpolation in P.

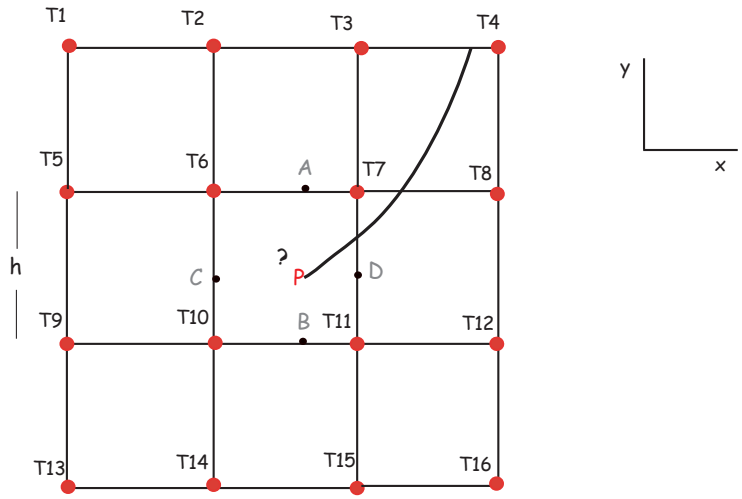


FIGURE 3.2: Shematic illustration of the 2-D geometry used in rays computation.

Referring to figure 3.3 (a), and to equation A.7 and A.8, we have

$$\begin{aligned} u_x|_A &= \frac{u_{r+1,s} - u_{r,s}}{h} \\ u_x|_B &= \frac{u_{r+1,s+1} - u_{r,s+1}}{h} \\ u_y|_C &= \frac{u_{r,s+1} - u_{r,s}}{h} \\ u_y|_D &= \frac{u_{r+1,s+1} - u_{r,s+1}}{h} \end{aligned}$$

- Second Method: *FD* 1^{st} *order forward scheme*
  With this method, partial derivatives $u_x|_P$ and $u_y|_P$ are firstly evaluated in the four adjacent nodes. Partial derivative $u_x|_{rs}$ is computed in $T6$, $T7$, $T10$ and $T11$ dividing by the grid step h, respectively, the travel times differences $T7-T6$, $T8-T7$, $T11-T10$, and $T12-T11$. The same approach is valid for $u_y|_{rs}$, evaluated in the same adiacent nodes, dividing by the step grid the travel times differences $T10-T6$, $T11-T7$, $T14-T10$, and $T15-T11$. Referring to figure 3.3 (b), and

to equation A.7 and A.8

$$\begin{aligned} u_x|_{r,s} &= \frac{u_{r+1,s} - u_{r,s}}{h} \\ u_y|_{r,s} &= \frac{u_{r,s+1} - u_{r,s}}{h} \end{aligned}$$

Values of $u_x|_{rs}$, $u_y|_{rs}$ in P can then be found using a bi-linear interpolation.

- Third Method: *FD 2^{st} order diagonal centered scheme*
  As in the previous method, the partial derivatives are evaluated first in the adjacent four nodes and then interpolated, but the finite difference method considers a larger set of nodes. Partial derivative $u_x|_{rs}$ is computed in $T6$, $T7$, $T10$ and $T11$ dividing by $4h$, respectively, the following sum $(T3 - T1) + (T11 - T9)$, $(T(4) - T2) + (T12 - T10)$, $(T7 - T5) + (T15 - T13)$ and $(T8 - T6) + (T16 - T14)$. The same approach is valid for $u_y|_{rs}$, evaluated in the same adjacent nodes, considering, $(T9 - T1) + (T11 - T3)$, $(T(10) - T2) + (T12 - T4)$, $(T13 - T5) + (T15 - T7)$ and $(T14 - T6) + (T16 - T8)$.
  Referring to figure 3.3 (c), and to equation A.11 and A.12, we have

$$\begin{aligned} u_x|_{r,s} &= \frac{u_{r+1,s+1} - u_{r-1,s+1} + u_{r+1,s-1} - u_{r-1,s-1}}{4h} \\ u_y|_{r,s} &= \frac{u_{r+1,s+1} - u_{r+1,s-1} + u_{r-1,s+1} - u_{r-1,s-1}}{4h} \end{aligned}$$

Once we have calculated the partial derivatives, a new ray step Δs is traced from P to P' considering the components Δx Δy and Δz as,

$$\begin{aligned} \Delta x &= \Delta s \cdot \frac{\frac{\partial u}{\partial x}}{\sqrt{\frac{\partial u}{\partial x}^2 + \frac{\partial u}{\partial y}^2 + \frac{\partial u}{\partial z}^2}} \\ \Delta y &= \Delta s \cdot \frac{\frac{\partial u}{\partial y}}{\sqrt{\frac{\partial u}{\partial x}^2 + \frac{\partial u}{\partial y}^2 + \frac{\partial u}{\partial z}^2}} \\ \Delta z &= \Delta s \cdot \frac{\frac{\partial u}{\partial z}}{\sqrt{\frac{\partial u}{\partial x}^2 + \frac{\partial u}{\partial y}^2 + \frac{\partial u}{\partial z}^2}} \end{aligned}$$

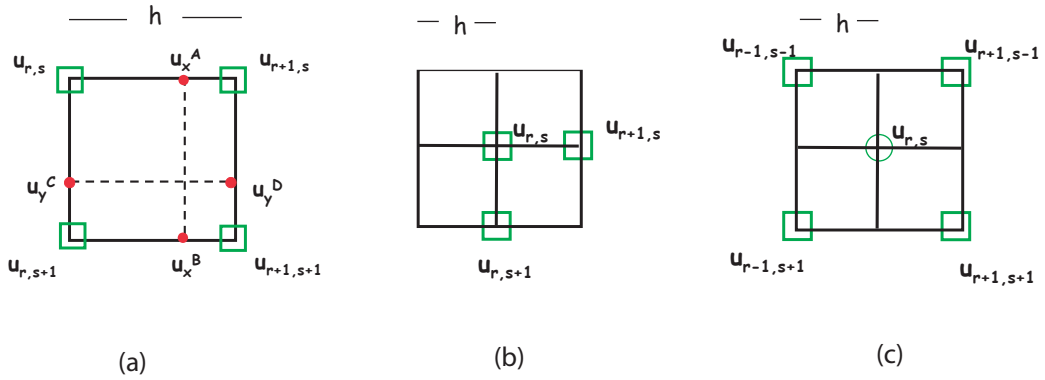


FIGURE 3.3: Shematic illustration of the 2-D grid geometry used in the first (a), second(b) and third (c) method.

Then in P' a new direction has to be found and the computation continues until the ray reaches the source.

Comparing travel times re-computed following the path with travel times resulting from the wavefront computation, for a 1-D model, we noted that rays corresponding to the head wave have a delay, different for each of the three methods. Although they produce apparently the same rays (Figure 3.4) focusing on the interface it is evident that a strong discontinuity, as the Moho, produces a peculiar ray behaviour that influences the travel times.

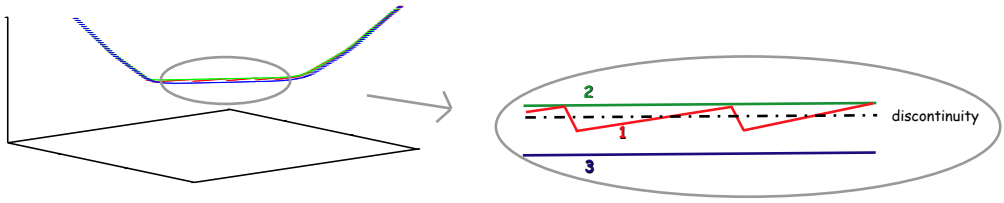


FIGURE 3.4: Schematic illustration of the different methods behaviour around a discontinuity.

The first method produces an instable ray that oscillates around the discontinuity, the ray path is longer with an obvious loss of time. The second method, as it is not

well balanced tends to see a shallower discontinuity, the ray path is shorter but in a slow medium. The third ray is stable and travels below the discontinuity, it tends to be smoother (its calculation employs a larger neighbourhood of nodes) with the disadvantage however of sinking too much in the medium below. Figure 3.5 shows the residuals histogram between the travel times computed in the wavefronts and travel times computed by slowness integration along the rays for the three methods described. Since it has a most stable behaviour, and its path produces the small error, we choose the third method.

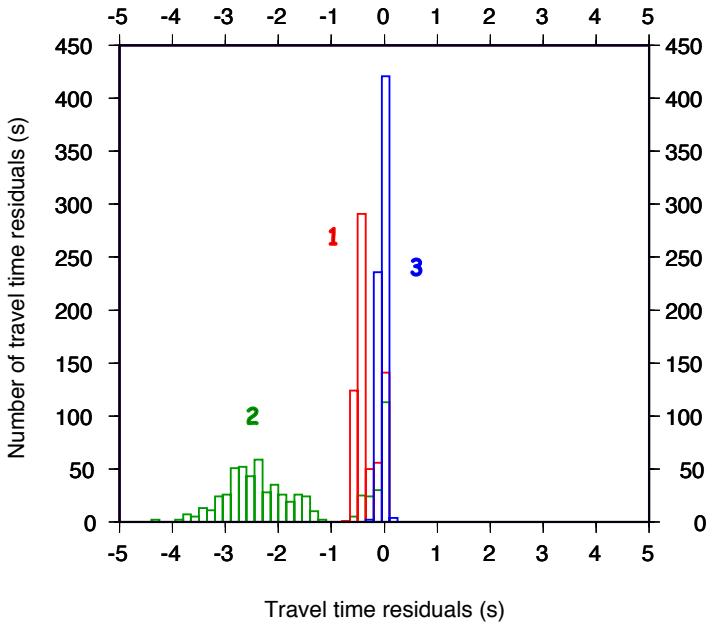


FIGURE 3.5: Histogram of travel time residuals between the P&L travel time values and ray travel times for the three numerical method described. The residuals rms value is respectively, -0.41 s, -2.12 s, -0.05 s for the first, second and third method.

Moreover, in order to limit the inevitable loss of time caused by rays sinking under the discontinuity it is crucial to have a fine discretization at least in the crust to better describe the Mohorovich discontinuity.

3.3 Two-Grid Implementation

As just described, the error introduced by the finite difference approximation depends on the step of the forward computation grid. The smaller the step the smaller the error will be. However, reducing the grid size significantly increase computing time. A compromise must be taken between precision and performance.

The Moho discontinuity between crust and mantle needs particular care. Because of the strong jump in velocity, its location at the correct depth is critical. Furthermore, the Moho originates strong refraction and head wave, it is therefore essential to be able to model such processes with good resolution. On the other hand, seismic wave speed varies much more smoothly in the mantle, so that a rather coarse step size could suffice to model wave propagation.

To maximise precision and, at the same time, keep a reasonable computational cost, we devised and implemented a method with different resolution for the crust and the mantle by managing two separate, partly overlapping, meshes, with a 120 km thick upper grid -with a step of 2 km- and a coarse mantle grid with a step of 6 km. A schematic illustration is shown in Figure 3.6.

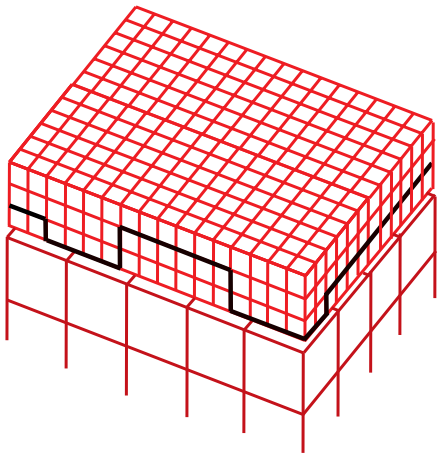


FIGURE 3.6: Schematic illustration of the two grid geometry.

This choice involves a complication. The wavefront computation has to be done in three time steps. The two grids overlap around a reference surface, which we defined to

be at a depth of 102 km. The grid overlap will be necessary for the ray computation. In fact the method chosed, needs each point of the ray, to be surrounded by at least three grid points.

First Step Computation

First travel times are propagated from the source point in the finer superior grid. A travel time value is associated to each node.

Second Step Computation

Then, travel times recorded in the reference surface are used as input for the travel time propagation in the coarse grid.

Third Step Computation

Finally, travel times of the second computation, stored in the reference surface, have to be reported as input for the wavefront superior grid computation (in order to allow rays to emerge on the free surface).

Figure 3.7 shows how values are transmitted to the reference surface. Between the second and third wavefront computation, we use a bi-linear interpolation in order to obtain a value where the coarse grid do not associate it.

Figure 3.8 shows a vertical section of rays and travel time field obtained for a 1-D velocity model. The computation has been done first for a single grid with a step of 2 km, then for a single grid with a step of 6 km and finally for the mixed grid. The area covered is the same and has an horizontal dimension of 3000 km, along the x direction, 90 km along the y direction, and a depth of 900 km. The source is located in the point of coordinates (0,0). The white curves represent the isochrones, the black lines represent the seismic rays, everywhere normal to the wavefronts.

Rays and travel times value obtained for the single grid of 2 km step are used as reference (top left). In the mixed grid section, it is shown respectively, how the second step computation continues directly from the first step (bottom left) and how the third continues directely from the second (bottom right). Figure 3.8 (top right) shows the comparison between the travel times difference of rays computed in the single grid of 6 km step and rays computed in the reference finer grid (red dots), and the travel time difference of rays computed in the mixed grid with rays computed in the reference finer grid (blue dots).

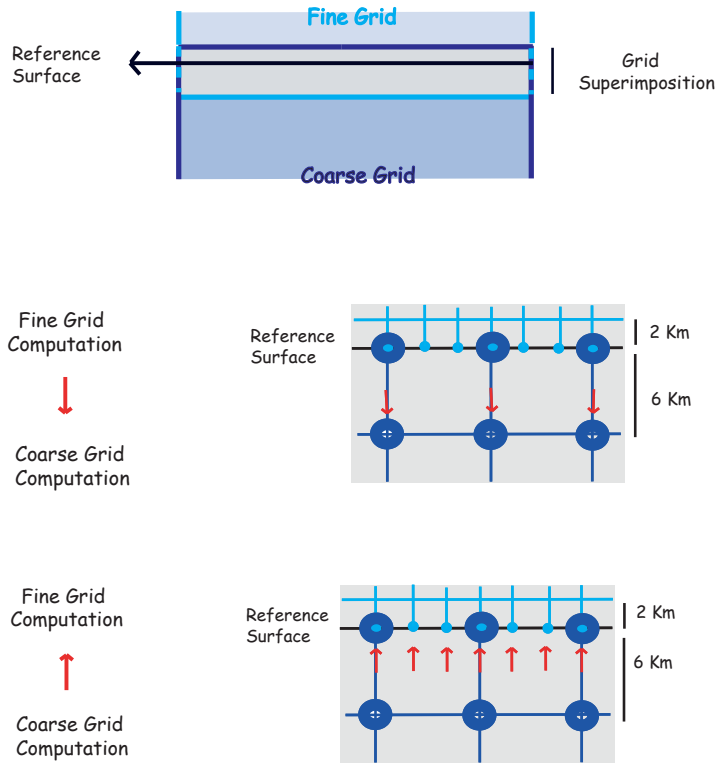


FIGURE 3.7: Schematic illustration of grid superimposition and reference surface location (on the top), and illustration of how travel times values are transmitted on the reference surface before the second wavefront computation step (in the middle) and before the third wavefront computation step (in the bottom).

For small source-station distances (in this example until about 2300 km) the ray travels enterely in the finer grid and it takes into account the wavefront computed in the first time step. The ray and travel time, in the mixed grid, are exactly the same of those computed entirely in the fine grid of 2 km step.

For larger distances the ray travels in all the three grids. Initially it takes into account the wavefront computed in the third time step, then when it crosses the reference surface the computation continues in the bottom grid, based on the wavefront computed in the second step. Similarly, when it passes the reference surface again, the computation continues in the first wavefront grid computed. The travel time value obtained for the station location decides if computation has to start from the first or the third wavefront computation grid. Rays travelling only in the finer grid will be the ones whose time at

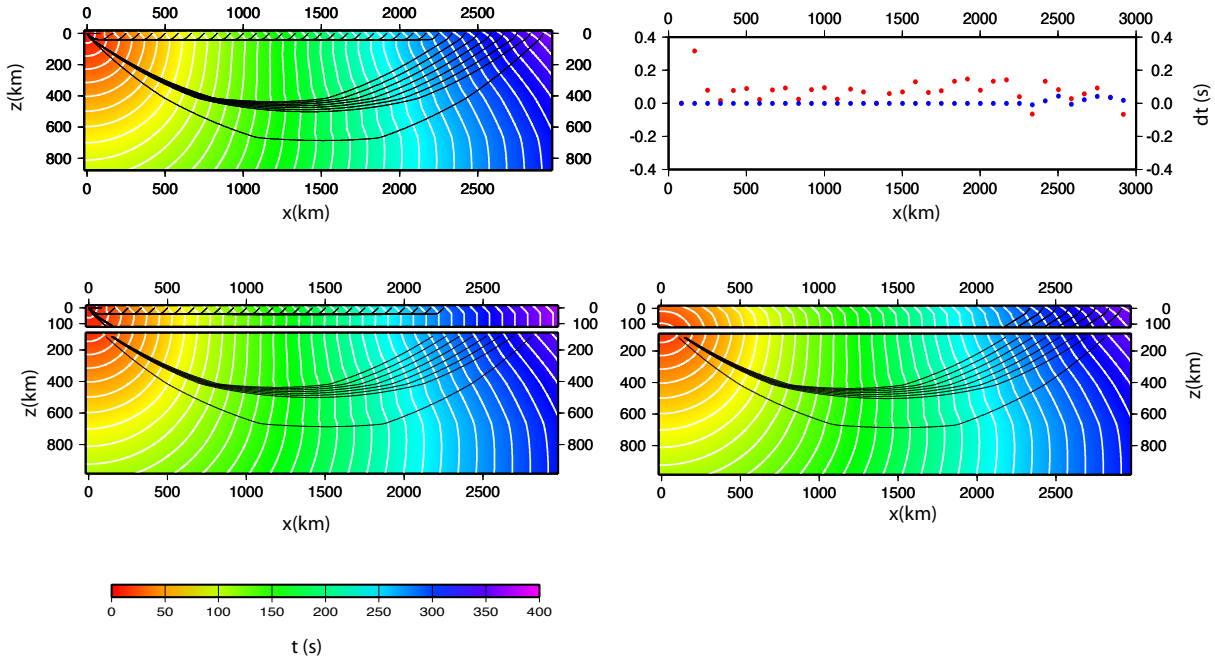


FIGURE 3.8: Comparison of wavefronts and rays using the one-grid (top left) and two-grid strategy (bottom). Travel time differences between rays entirely computed in a coarse grid of 6 km step (red) and rays computed in the mixed grid of 6 km and 2 km step (blue) with respect to a reference solution obtained with a single fine grid of 2 km step (top right).

the station, referred to the first computation is shorter than the time referred to the third computation.

The two grid-strategy allows to have a good compromise between computational effort and accuracy. Rays travelling in the first 100 km have the precision of a single 2 km step grid. Deeper rays have a precision, still higher than that obtained with a single 6 km step grid. The travel times crossings on the reference surface do not involve a significant error. In this example (done in a thin vertical slice, so almost in a contest 2-D) the computational time is about 0.6 and 1000 seconds respectively for the single coarse grid of 6 km step and for the fine grid of 2 km step. The computational time for

the mixed grid is about 60 seconds.

Chapter 4

Implementation of the FD scheme in spherical geometry

Although at continental or global scale a model is described with a geographical grid in spherical geometry and the inversion must be done for the same system, in order to use the finite-difference wavefront computation algorithm [Podvin & Lecompte, 1991] we need to describe the area with a cartesian grid. It can be done in two different manners (see Figure 4.1)

- **Embedding**
  It consists in treating the spherical Earth by embedding it (in part or in whole) within a cartesian box

- **Equidistant azimuthal projection**
  Using this projection the Earth is flattened. The use of the Earth flattening approximation (EFA) introduce in the velocity model a gradient in order to compensated for the sphericity .

Each of these method has advantages and disadvantages with respect to the other. With the former, a flat Moho (and every discontinuity) by being a spherical surface is represented with many steps laterally and it may be possible to have numerical problem in tracing rays relative to the head wave. The latter introduces a geographical projection and, hence, an approximation and, as will be explaned later, it requires a new regridding for each source. We choose the second because, allowing the use of the double grid

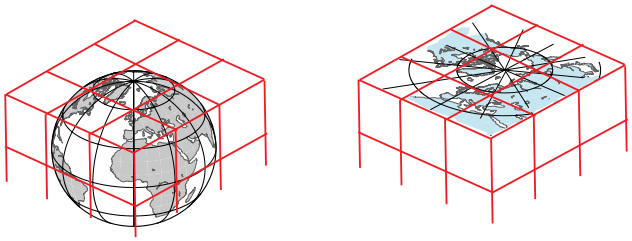


FIGURE 4.1: Schematic illustration of how a cartesian grid can be constructed for the globe, embedding the Earth within a box (left) or flattening the Earth through a projection (right).

explained in Section 3.3, it is less computationally demanding, although at the cost of more complex programming. The comparison between the two strategies will be done in future works.

4.1 Equidistant Azimuthal projection and EFA

Azimuthal projections are projections to a plane placed in a point tangent to the globe. If we consider the north or south pole, meridians are represented as straight lines and parallels as concentric circles (see Figure 4.2). Distances and directions to all places are

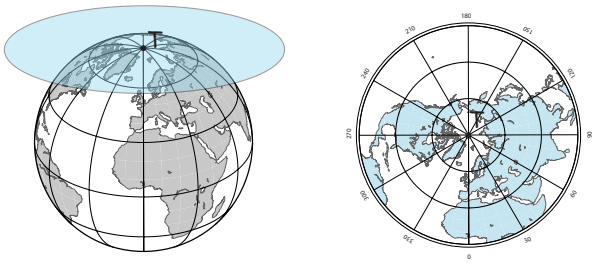


FIGURE 4.2: Schematic illustration of how is performed an equidistant azimuthal projection around a point of tangency T

true only from the central point of projection. Distances are exact between points along the same meridian, while all other distances are approximate. Figure 4.3 shows how the error, introduced by the approximation, in the transversal distance between two points at the same distance from the projection center increases the further away one gets from the central point, or the larger becomes the separation between the points.

As distance between a source and a station has to be mantained, each source will be considered as a projection center. The error, introduced by this approximation, influences the computation only for rays deviating from the great circle path between station and source. As we can see in the Figure 4.3, for a distance TA (corresponding to the source-station distance) of 4000 km and an angle α of 2 degree (corresponding to a ray deviation of about 140 km) , the error will be of about 1 km. We estimate this to be the maximum transversal distance error in which we incur, and it is well acceptable.

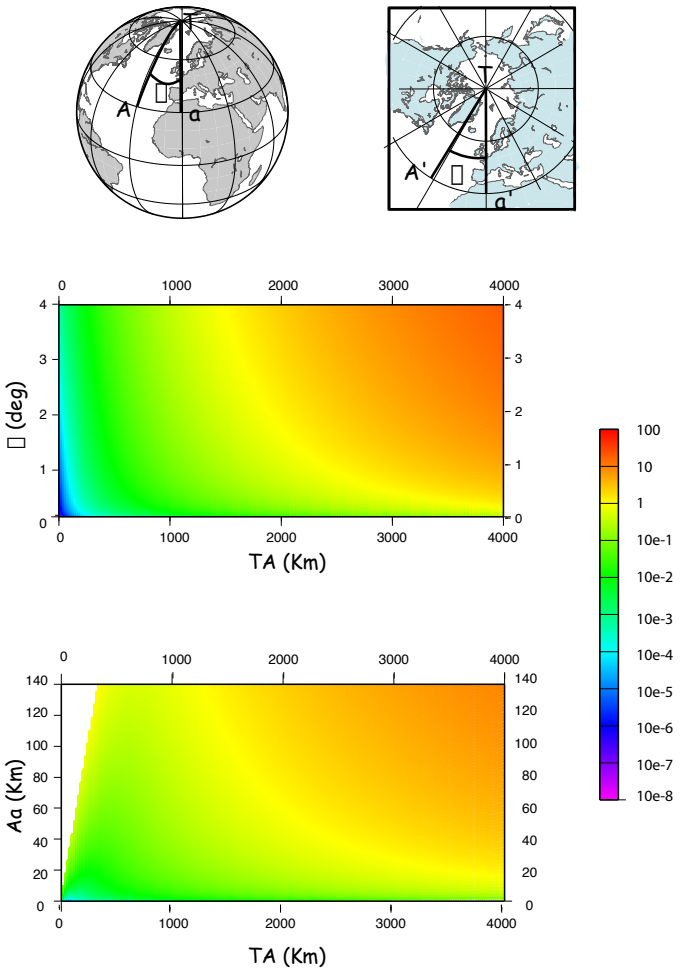


FIGURE 4.3: Colors, in logarithmic scale, represent the difference, in km, between the distance A'a' (on the plane) and Aa (on the sphere) in relation with the distance from the point of tangency, T, and with the opening angle α (middle) and with the distance Aa itself (bottom).

Obviously, rays are not the same if propagated within a sphere or in a cartesian

medium. In order to obtain the same traveltime behaviour, we can use the Earth flattening approximation [Müller, 1971], that introduces a velocity gradient in the flat model.

Defining with z_f and v_f, respectively, the depth and the velocity in the Cartesian geometry, these two simple equations can be found in order to compensate for the sphericity

$$z_f = -r_e \ln\left(\frac{r}{r_e}\right) \tag{4.1}$$

$$v_f(z_f) = \frac{r_e}{r} v_s(r) \tag{4.2}$$

where r_e the Earth's radius, r the distance from the center of the Earth, and v_s is the velocity in the sphere. The EFA approximation preserves both the kinematic and the dynamic properties of body wave propagation [Aki and Richards, 2002].

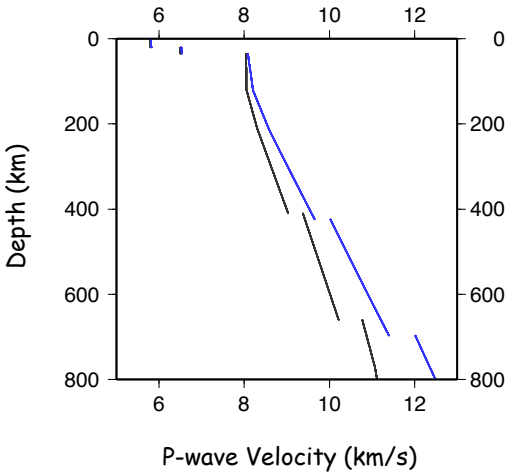


FIGURE 4.4: SP6 velocity profile as function of depth in the sphere (gray) and in the flat (blue)

As it is shown in Figure 4.7, the deeper rays are, the larger becomes the importance of the Earth flattening approximation. As a matter od fact, even if at a depth of 20 km the EFA is not so important (for sp6 model $v_s = 6.50$, $v_f = 6.52$), at a depth of 200 km, it becomes necessary to change the velocity value $v_s = 8.95$ into the flat value $v_f = 9.5$ in order to obtain the same traveltime behaviour .

Figure 4.5 shows how, practically, the forward computation grid is constructed. First, the geographical system is rotated in order to have the new equator (represented by the blue parallel) in the center of the area to grid. It defines the grid cartesian orientation.

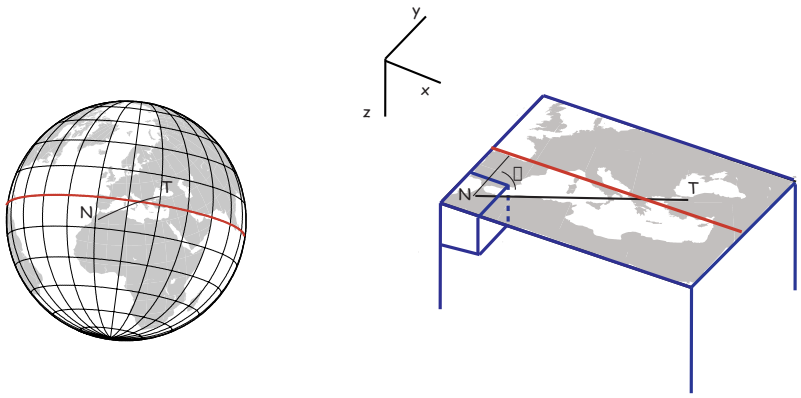


FIGURE 4.5: Schematic illustration of forward computation grid construction

We denote with (Lat_T , Lon_T) and ($\text{Lat}_{T'}$, $\text{Lon}_{T'}$) the spherical coordinate of the source, respectively, in the geographical and rotated system. Computation begins in the cartesian grid, first, given T (0,0,0) and the central point of a cell N (x_N, y_N, z_N), we compute in the cartesian grid the distance TN and angle β. For the nature of the projection the distance TN is preserved in the sphere and β corresponds, with a good precision to the azimuth of the new rotated model. Then, knowing the spherical coordinate ($\text{Lat}_{T'}$, $\text{Lon}_{T'}$), the azimuth and the distance TN it is easy to find the spherical coordinates for N ($\text{Lat}_{N'}$, $\text{Lon}_{N'}$). Finally, with an inverse rotation, the true geographical coordinates (Lat_N, Lon_N) are found and the corresponding P wave velocity associated. As said, especially when we move to deeper cells, the Earth flattening approximation has to be applied, Equation 4.1 can be written as

$$r = re \exp\left(\frac{-z_f}{re}\right) \tag{4.3}$$

The value of velocity that has to be associated to cell N is the one corresponding in the sphere to the point of spherical coordinates (Lat_N, Lon_N,r) after the transformation defined in Equation 4.2.

4.2 Travel time and partial derivatives

In order to solve the tomographic problem, as indicated by equation 2.11, we need to compute the vector of traveltime residuals $\mathbf{\Delta t}$ and the partial derivatives matrix $\mathbf{G}$.
$\mathbf{\Delta d}$ is defined as the difference between the observed travel time, recorded by the station

($\mathbf{d^{Ob}}$) and the predicted travel time in the model where rays are computed ($\mathbf{d^0}$). As described in section 3.2, rays are computed step by step, $\mathbf{d^0}$ can be obtained from the sum of all step traveltimes contributions. If r is the number of step,

$$d^0 = \sum_1^r \Delta s \cdot s_n \tag{4.4}$$

where Δs is the ray step, chosen equal to the ratio $\frac{h}{\sqrt{2}}$ (h is the forward computation grid step), and s_n is the slowness computed in the central point of the step, with a tri-linear interpolation.

G is a matrix composed by n rows and m columns, where n is the number of rays and m the numbers of model parameters. In the linear approximation, we wrote (equation 2.7)

$$\Delta d_i = \sum_j G_{ij} \Delta m_j$$

Given the i^{th} ray, G_{ij} represents the length contribution that multiplied by the slowness anomaly in the j^{th} model parameter gives us the traveltime perturbation in the same cell:

$$G_{ij} = \frac{\partial d_i}{\partial m_j}$$

Figure 4.6 shows the way the computation is done (figure and formulas are referred to the 2-D case). If the i^{th} ray is composed by r step, Gij can be written as,

$$G_{ij} = \frac{\partial d_i}{\partial m_j} = \frac{\partial d_i}{\partial s_1}\frac{\partial s_1}{\partial m_j} + \frac{\partial d_i}{\partial s_2}\frac{\partial s_2}{\partial m_j} + \frac{\partial d_i}{\partial s_3}\frac{\partial s_3}{\partial m_j} + ... + \frac{\partial d_i}{\partial s_r}\frac{\partial s_r}{\partial m_j}$$

where s_1, s_2, ..., s_r are the slowness computed in the central point of each ray's step Δs.

Referring to figure 4.6, we can write

$$\begin{aligned}
G_{i1} &= \frac{\partial d_i}{\partial m_1} = \frac{\partial d_i}{\partial s_2}\frac{\partial s_2}{\partial m_1} + \frac{\partial d_i}{\partial s_3}\frac{\partial s_3}{\partial m_1} + \frac{\partial d_i}{\partial s_4}\frac{\partial s_4}{\partial m_1} + \frac{\partial d_i}{\partial s_5}\frac{\partial s_5}{\partial m_1} \\
G_{i2} &= \frac{\partial d_i}{\partial m_2} = \frac{\partial d_i}{\partial s_1}\frac{\partial s_1}{\partial m_2} + \frac{\partial d_i}{\partial s_2}\frac{\partial s_2}{\partial m_2} + \frac{\partial d_i}{\partial s_3}\frac{\partial s_3}{\partial m_2} + \frac{\partial d_i}{\partial s_4}\frac{\partial s_4}{\partial m_2} \\
G_{i3} &= \frac{\partial d_i}{\partial m_3} = \frac{\partial d_i}{\partial s_1}\frac{\partial s_1}{\partial m_3} + \frac{\partial d_i}{\partial s_2}\frac{\partial s_2}{\partial m_3} + \frac{\partial d_i}{\partial s_3}\frac{\partial s_3}{\partial m_3} \\
G_{i4} &= \frac{\partial d_i}{\partial m_4} = \frac{\partial d_i}{\partial s_2}\frac{\partial s_2}{\partial m_4} + \frac{\partial d_i}{\partial s_3}\frac{\partial s_3}{\partial m_4} + \frac{\partial d_i}{\partial s_4}\frac{\partial s_4}{\partial m_4} + \frac{\partial d_i}{\partial s_5}\frac{\partial s_5}{\partial m_4}
\end{aligned}$$

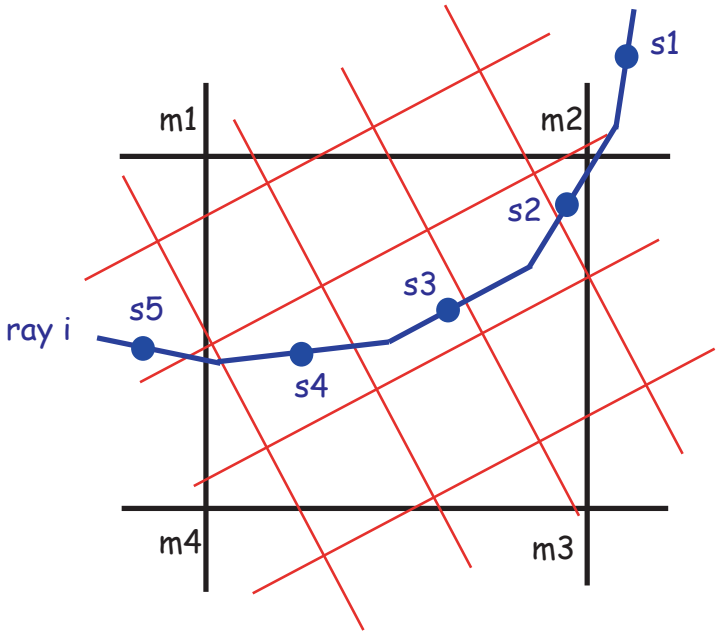


FIGURE 4.6: Schematic illustration about the partial derivatives construction. The black grid represents the model parameterizaton grid described by spherical coordinates. Each element (m1,m2,m3,m4) is the slowness in that point. The red grid represents the forward cartesian computation grid. In blue is represented the i^{th} ray traced step by step.

Each term s_n can be written, using a bi-linear interpolation (or tri-linear for the real geometry 3-D) in terms of the model parameters m_j,

$$s_n = \alpha_1 m_1 + \beta_1 m_2 + \gamma_1 m_3 + \phi_1 m_4$$

In this way, for each central point of the step, the partial derivative $\frac{\partial s_n}{\partial m_j}$ reduces to a factor depending on the position respect to the parameter m_j. Obviously it will be different from zero only when the central step point, where s_n is computed, is in a cell whose boudaries nodes contain the m_j parameter. The partial derivative, $\frac{\partial d_i}{\partial s_n}$ is simply the length of the step Δs.

It is important to note that, as the inversion will be done in a spherical grid also rays traveltimes and the partial derivative matrix have to be referred to the same geometry. Actually both the two partial derivatives, $\frac{\partial s_n}{\partial m_j}$ and $\frac{\partial d_i}{\partial s_n}$, take into account the Earth flattening approximation. In fact s_n is computed on the basis of the true spherical depth

(see Equation 4.2) and consequently on the basis of the true distances between model parameters. Instead the step length Δs depends on the parametrization used [Gorman, 2002], and must be changed. As illustrated in Figure 4.7, in a spherical system the length Δs is not constant as in the cartesian system, but it depends on the depth, or better on the ratio $\frac{r}{re}$.

$$\Delta s = \sqrt{\left(\frac{r_{i+1}l_{i+1} - r_i l_i}{r_e}\right)^2 + (\xi_{i+1} - \xi_i)^2} \tag{4.5}$$

Extension of all these formulas to three dimension is straight forward.

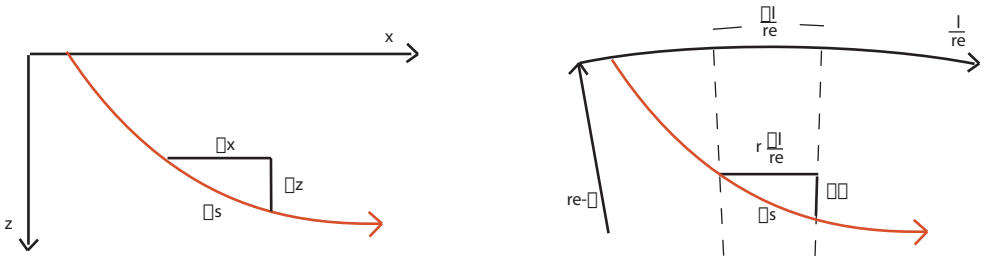


FIGURE 4.7: Schematic illustration about the difference in Δs between a cartesian coordinates system (left) and a spherical one (right). In the spherical system depth z is converted to radial distance r, ξ represents the spherical depth so that $r = r_e - \xi$ is r. l is the the arc length along the surface of the Earth.

Chapter 5

Rays in realistic heterogeneous mantle structure

In this chapter , we show two applications done in order to test the method.
In section 5.1 we describe how also teleseismic data can be included in the analysis. Rays are traced in a model that is a simplified representation of a mantle plume.
In section 5.2 we show an application of the method where rays are traced in a recent tomographic model, and compared with those traced in a lateral homogeneous model.

5.1 Teleseismic Rays and Behaviour in Slow Anomaly

Teleseismic rays are essential for imaging deep mantle structures. Rays bottoming in the model volume (regional distance) are in fact very important to achieve good vertical resolution, but of course only rays rising from deeper mantle can yield the necessary data coverage of the bottom of the upper mantle. Teleseismic rays are commonly used in local and global tomography, from the pioneering work of Aki et al. (1977) and Dziewonski et al. (1977) but application of finite differences would require to extend the calculation grid to the whole Earth, with a dramatic increase of computational requirements. We devise instead a hybrid approach, by which the travel time field from a teleseismic source is initialized at the bottom of the model by means of a tau-p (Buland and Chapman, 1983; Crotwell et al., 1999) calculation, and then it is propagated upwards by the usual finite difference scheme.

We consider a lateral homogeneous model, sp6 [Morelli & Dziewonski, 1993] with a

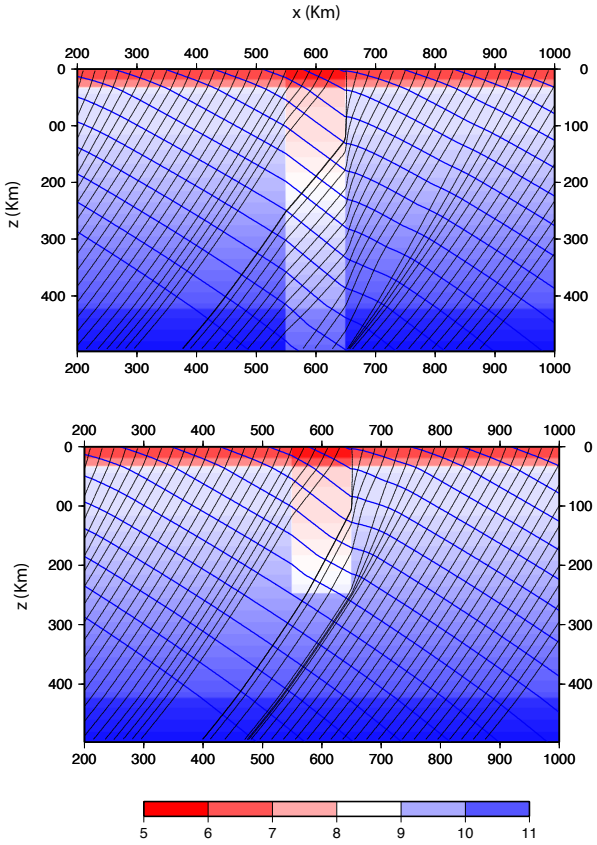


FIGURE 5.1: Illustration of the velocity model used. Colors in the background represents the P-wave velocity (km/s). In blue are traced the wavefronts and in black the rays backtraked for a cylinder 500 km deep (on the top) and 250 deep (on the bottom).

cylindrical slow anomaly superimposed. In order to propagate the wavefront , a slowness value has to be associated to each cell of the forward grid. The area extension is 1200x500 km with a depth of 500 Km which corresponds to a grid composed by 600x250x250 nodes with a step of 2 km. We initialized the wavefield at the base of the model grid with a plane wave with apparent velocity corresponding to the arrival of a teleseismic P wave from a surface focus at the distance of 60 degrees from the center of the grid, calculated by a tau-p interpolation (Crotwell et al., 1999). We compare differences in traveltimes and rays behaviour in two different starting models. In the first, the slow anomaly is represented by a cylinder 500 km deep, while in the second model the anomaly is a

cylinder 250 km deep. Both the anomaly are of 9 % with respect to the 1-D model.

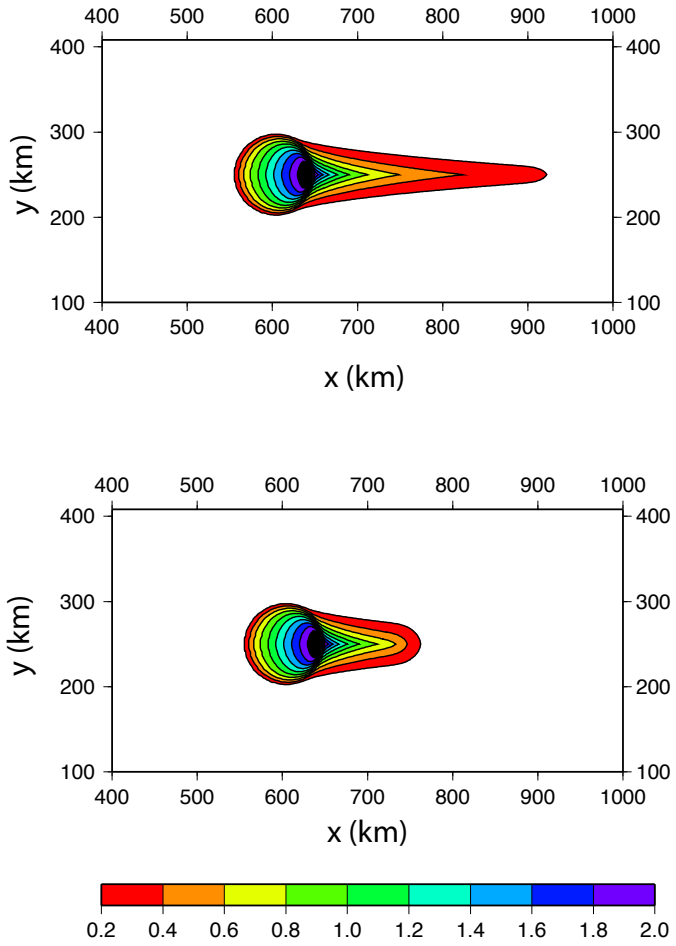


FIGURE 5.2: Traveltime delay respect to the unperturbed model for the model with the cylinder 500 km deep (top) and 250 km deep (bottom)

Figure 5.1 shows the models, wavefronts and rays compared between the two model. Fig 5.2 shows the different behaviour in the traveltime field recorded at the top of the grid, between the two models. Colors represents the delay with respect to the computation done in a lateral homogeneous model. It is evident how the perturbation derived by the deeper anomaly is present for larger distances. The main outcome of this experiment, however, is the demonstration of the ability to include teleseismic rays at an acceptable

computational cost, by resorting to the hybrid scheme.

5.2 Rays in realistic heterogeneous mantle structures

To test our forward calculation scheme we trace P rays in a tomographic model of P wave speed in the upper mantle under the greater European and Mediterranean region [Piromallo and Morelli, 2003], including also crustal structure from CRUST2.0 [Bassin et al., 2000].

5.2.1 P wave Model: PM0.5

The velocity is defined with a regular spacing in a coordinate system rotated with respect to the geographical one, as given by an oblique cylindrical projection, centered at (10 W, 45 N) with pole at (170 E, 45 N). This choice is intended to have the model area lying astride the new equator, which allows to obtain cells of approximately the same size in both direction. The model horizontal dimensions are 6600 km in E-W direction,

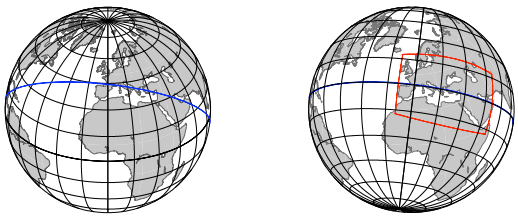


FIGURE 5.3: PM0.5 rotated model

3900 km in N-S direction and 1000 km in depth, and it consist of a 3-D lattice of 180411 nodes (121x71x21), spaced by 0.5° in the horizontal directions and 50 km in the vertical. The velocity is defined in each node.

5.2.2 Crustal Model: Crust2.0

CRUST2.0 is a global crustal model. It is divided in cells of 2x2 degree. To each cell it is associated a profile composed by 7 layers of different depth. Vp and Vs velocity values and density values are given explicitly for each layer that takes in account of the

ice, water, soft and hard sediment. A different elevation topography, bathymetry and Moho depth is also specified for each cell.

The forward computation is done in a single 2 km step grid. A slowness is associated to each cell, from CRUST2.0 for cells whose depth is less than the crustal thickness or, for deeper cells, using a trilinear interpolation from values at the nodes of PM0.5 .

Figure 5.4 shows the reference location of two cross section. For each of these, the

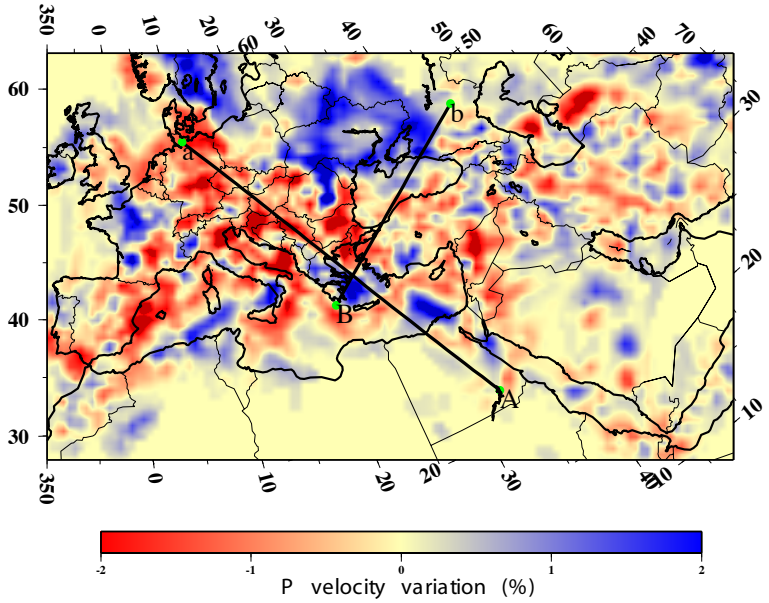


FIGURE 5.4: Reference location of the two sample cross sections (A-a and B-b). P-wave lateral variation in tomographic model PM0.5 at 200 km depth shown in the background.

computation is done separately. The grids are constructed around each section with an horizontal extension of (4000x500) km and a depth of 1000 km. An imaginary source is located, repectively, in A and B and the wavefront computations are performed. Rays are then computed from hypothetical stations, placed along the two profiles, respectively A-a and B-b.

Figure 5.5 shows lateral variation of rays from their great circle path for the profile A-a. Figure 5.6 shows the comparison between rays computed in the model described and rays computed in the one-dimensional model used for reference. Rays are sometimes

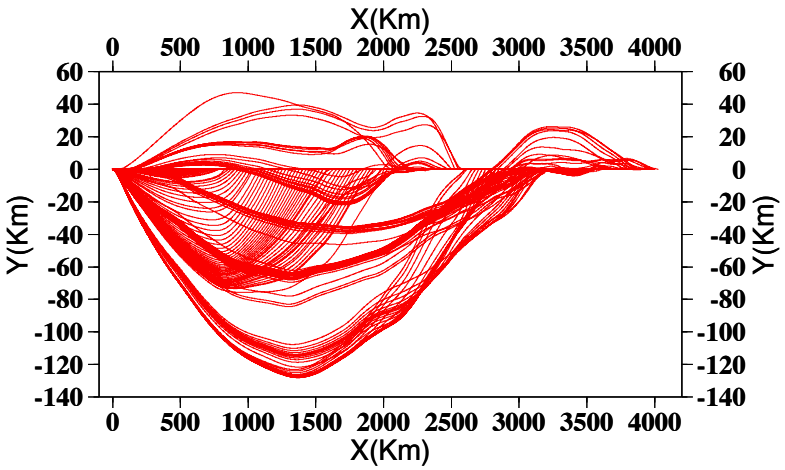


FIGURE 5.5: Horizontal projection of deviation of rays from great circle path along profile A-a

considerably bent. The largest shift, especially with depth, of path do not appear to be possibly obtained by linearly perturbing the reference ray. This analysis proves the effective importance of using numerical wavefront tracking scheme and consequentely numerical ray tracing method, instead of methods such as ray bending, especially when strong heterogeneous structures are present even at upper mantle scales.

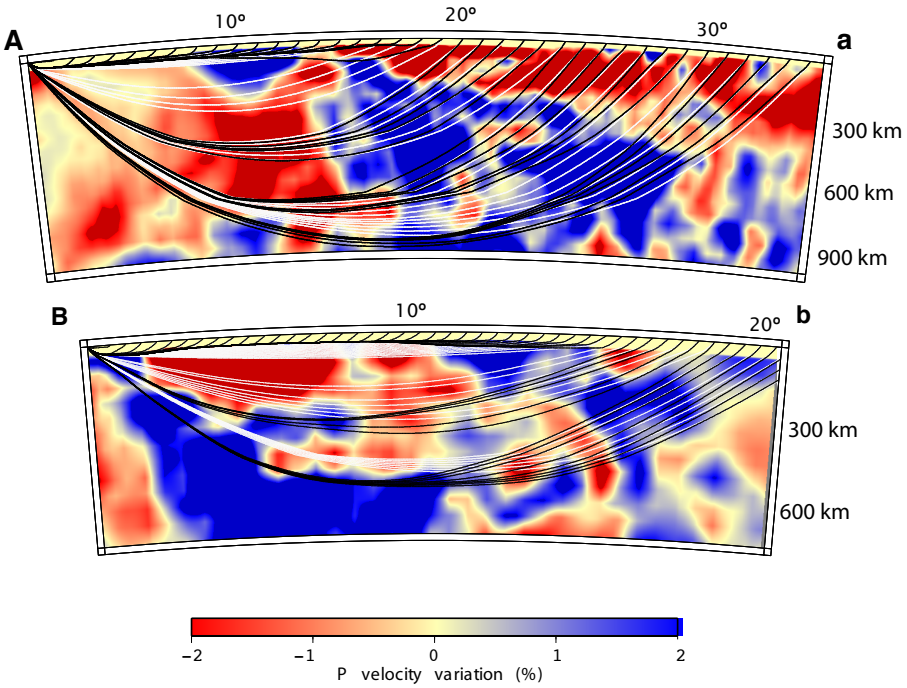


FIGURE 5.6: Rays traced in the 3-D tomographic model (black) compared with those traced in the reference 1-D model (white) along profile A-a (top) and B-b (bottom). Note for instance the dramatic concentration of 3-D rays at the top of the mantle, that leave the uppermost 300 km empty

Chapter 6

Recovery of Mantle Structures

6.1 Synthetic Model

In order to test the robustness of the method it is useful to perform a syntetic test, that attempts to reconstruct a known input model using the same sources and stations position of the real experiment. In fact if a known structure, with similar length scales to the target structure, can be recovered using the real data distribution, then the solution to the real inverse problem should be reliable. The quality criterion is the similarity between the input and the recovered model. Usually this kind of test is performed using a checkerboard model, in which the volume is divided into regularly alternating regions of high and low velocity. This approach is not necessarily as reliable as it may seem [Lévêque et al., 1993]. It is possible for the small scale structure of the checkerboard test to be well retrieved while larger-scale structure is poorly retrieved. If the solution takes into account the non-linearity of the inverse problem, i.e. rays are traced in the heterogeneous model, then the ray path coverage will have a dependence on the velocity distribution. Thus, while a checkerboard reconstruction can account for the non-linearity of the travel time dependence on the checkerboard structure, it cannot account for the non-linearity of the traveltime dependence on the true structure. It may be better to use a different and more realistic synthetic model.

Figure 6.1 shows the input anomaly model used in this test. It represents a simplified model of high velocity slabs (anomaly of 3%) and slow velocity backarc structures (anomaly of -3%) in the Euro-Mediterranean area. This heterogeneity is superimposed on the 1-D model *sp6*, [Morelli & Dziewonski, 1993], the same used as initial model in

the forward computation of the first iteration.

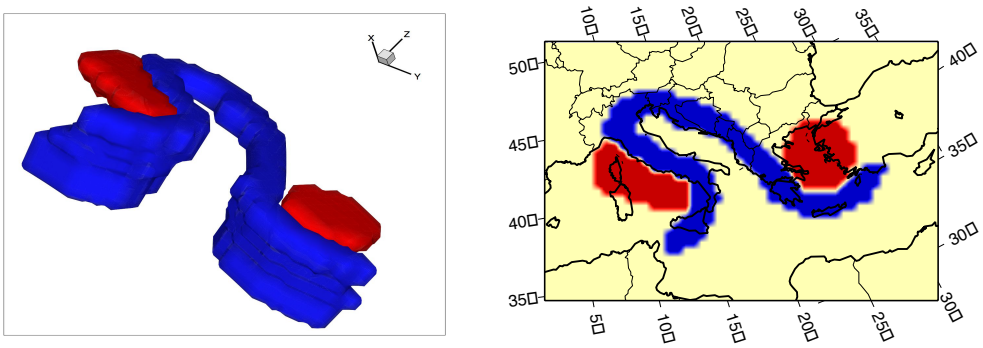


FIGURE 6.1: 3-D view (left) and map view at a depth of 100 km (right) of the synthetic model.

6.2 Data Selection

Data used both in this synthetic test and in the real experiment (described in the next chapter) are P wave travel times reported by the EHB bulletin for the time period 1984-2004. This bulletin, due to Engdahl, van der Hilst and Buland, is an improved version of the ISC bulletin [Engdahl et al., 1998]. They redetermined the identities of over 5 million arrivals, including many previously un-recognized phases. The new phase identifications are based on ak135 travel times, geographic variations in ocean depth and procedures to reduce biases. Partly thanks to these identifications , EHB's hypocentres for the largest events may be better for tomography than those originally computed by the ISC. The results from EHB's reprocessing of data are now available from the ISC web and AutoDRM servers.
Figure 6.2 shows events and stations location in the Euro-Mediterranean area reported by the EHB bulletin for the period 2002-2004.

First arrival travel times are selected from the EHB bulletin on the basis of a few simple criteria, meant to sort out only the well recorded ones. Each event is required to have the following characteristics:

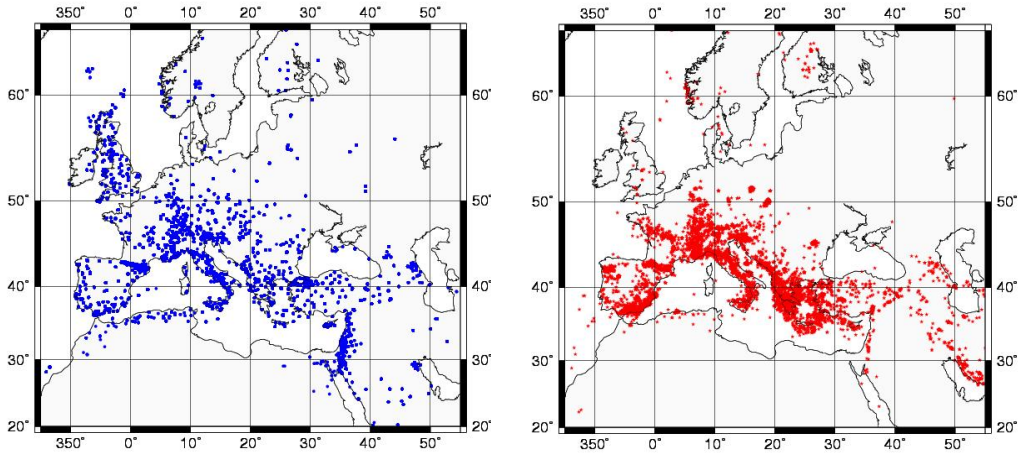


FIGURE 6.2: Stations location (left) and events location (right) for the period 2002-2004

-The reading time precision has not to be larger than $1/10^{th}$ second.
-The secondary azimuth gap, that is the largest azimuth gap filled by a single station, has not to be larger than 180 degree.
Corrections for ellipticity and station elevation are applied.

For the syntetic test, only the sources and stations geometry are taken from the dataset. The travel times are computed in the sample structure, with the same theoretical apparatus that is used for the actual inversion, this means that an error in the theory is not detected by this test.

6.3 Forward Computation

The model parameterization chosen in the test is the same adopted by the PM0.5 velocity model (see section 5.2.1). Velocities are defined in each node of a grid composed by a total of 180411 nodes. The mesh spacing in the orizontal dimension is 0.5 degree while in the vertical is 50 km.

The foward computation grids have a finer discretization (2 km for the crust grid and 6 km for the mantle grid). They are oriented in such a way that they contain as many station and events (and so ray paths) as possible. Figure 6.3 shows the location of seismic stations considerated in the study. The color is referred to the number of events recorded by each station. The grey rectangle includes the area covered by the forward

computation grids. The horizontal dimension is 4080x2100 Km which corresponds to

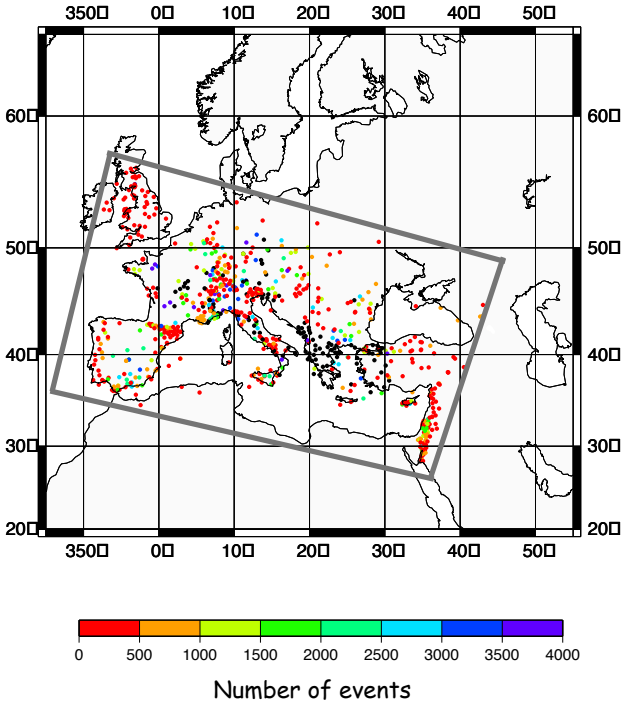


FIGURE 6.3: Number of events recorded by each station for the period 1982-2004.

2041x1051 nodes in the fine grid and to 681x351 nodes in the coarse grid. In the vertical dimension the first 120 km are covered by the fine grid while the depth between 102 km and 756 Km is covered by the coarse grid.

The reciprocity of travel times between sources and stations allow to treat the stations as sources and vice versa. This possibility is very useful when the number of stations is significantly less than the number of events and the computation is particularly expensive. We computed the wavefronts from all the stations for a total of 754 computations. For each station rays are backtracked from the events location. The total number of rays considered in the test is about 3,039,050.

6.4 Iterative nonlinear inversion: method, regularization

As described in section 2.1, system 2.8 is usually underdetermined and the matrix $(\mathbf{G^T G})$ is close to being singular and cannot be inverted. A common approach to deal with ill-posed least squares problems is to impose additional constraints on the problem, a process referred to as regularization. The solution is forced to satisfy other requirements, based on our physical knowledge on the problem. It is comprensible to require that the model solution has to be close to a model, we know to be realistic (a priori model $\mathbf{m^{prior}}$). Following the probabilistic approach [*Tarantola*, 2005] remembering that $\mathbf{\Delta d}$ represents the difference between the observed and predicted data $\mathbf{d^{Ob} - Gm^0}$ and $\mathbf{\Delta m}$ is the perturbation to apply to $\mathbf{m_0}$ Equation 2.11 must now be written, [*Tarantola and Valette*, 1982], as

$$\mathbf{\Delta m} = \left(\mathbf{G^T C_D^{-1} G} + \mathbf{C_M^{-1}}\right)^{-1}\left(\mathbf{G^T C_D^{-1} \Delta d} - \mathbf{C_M^{-1}}(\mathbf{m^0} - \mathbf{m^{prior}})\right) \tag{6.1}$$

where $\mathbf{C_M}$ and $\mathbf{C_D}$ are, respectively, the model and data covariance matrices. This approach assumes that uncertainties both on data and on our a priori model are Gaussian, described by matrices $\mathbf{C_D}$ and $\mathbf{C_M}$. Practically, they act on the solution, weighting the best compromise between the minimization of data misfit and model misfit. If we assume that all data are result of independent measurements and share the same (Gaussian) error; and, similarly, model parameter estimates have indipendent uncertainties, then:

$$\begin{aligned} \mathbf{C_M} &= \sigma_\mathbf{M}^\mathbf{2}\mathbf{I} \\ \mathbf{C_D} &= \sigma_\mathbf{D}^\mathbf{2}\mathbf{I} \end{aligned} \tag{6.2}$$

Equation (6.1) becomes,

$$\mathbf{\Delta m} = \sigma_\mathbf{d}^\mathbf{2}\left(\mathbf{G^T G} + \left(\frac{\sigma_\mathbf{m}}{\sigma_\mathbf{d}}\right)^\mathbf{-2}\mathbf{I}\right)^{-1}\sigma_\mathbf{d}^\mathbf{-2}\left(\mathbf{G^T \Delta d} - \left(\frac{\sigma_\mathbf{m}}{\sigma_\mathbf{d}}\right)^\mathbf{-2}\left(\mathbf{m^0} - \mathbf{m^{prior}}\right)\right) \tag{6.3}$$

Defining the *damping parameter* ϵ as

$$\epsilon = \frac{\sigma_D}{\sigma_M} \tag{6.4}$$

we have,

$$\mathbf{\Delta m} = \left(\mathbf{G^T G} + \epsilon^\mathbf{2}\mathbf{I}\right)^{-1}\left(\mathbf{G^T \Delta d} - \epsilon^\mathbf{2}(\mathbf{m^0} - \mathbf{m^{prior}})\right) \tag{6.5}$$

where $\mathbf{I}$ is the identity matrix. ϵ is a weighting parameter that controls the degree of damping. The functional (introduced by 2.9)minimized by the *damped least squares solution* becomes

$$\Phi = \| \mathbf{G\Delta m - \Delta d} \|^{\mathbf{2}} + \epsilon^{\mathbf{2}} \| \mathbf{\Delta m} \|^{\mathbf{2}} \tag{6.6}$$

By adjusting the parameter ϵ we can control the tradeoff between misfit and model variance. In figure 6.4 we can see how the resulting model depends on the damping parameter choice. The larger ϵ is, the smaller will be the difference from the a priori model and the larger will be the misfit. These constraints add stability to the inversion,

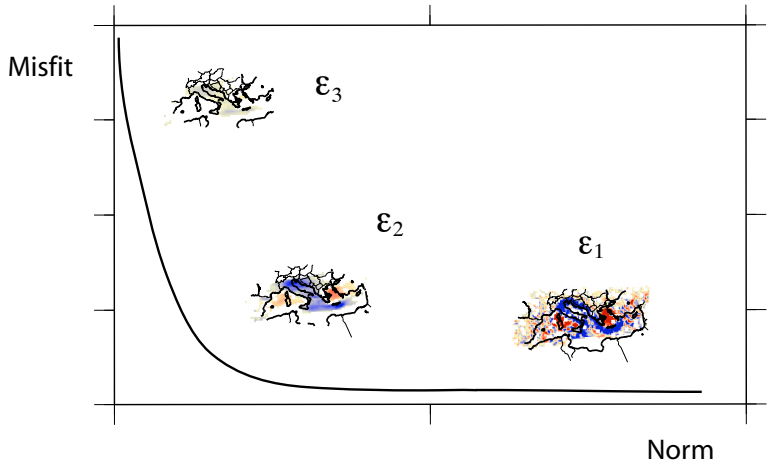


FIGURE 6.4: Schematic curve of tradeoff between misfit and model variance with the variation of damping parameter $\epsilon_1 < \epsilon_2 < \epsilon_3$

perturbations in blocks that are not sampled by rays will go to zero; anomalies will be distribuited equally among blocks that are sampled only with identical ray paths. However, the damped least squares solution will not necessarily lead to a smooth model, since it is the size of the model, not its roughness, that is minimized. Model perturbations in adjacent blocks can be quite different.

A common measure of model roughness for block models is the *Laplacian operator* ∇^2, which can be approximated with a difference operator in both 2-D and 3-D block geometries.

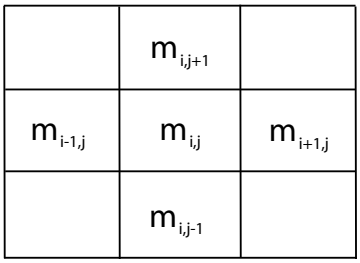
The solution that minimizes the data misfit and the model roughness is

$$\mathbf{\Delta m} = \left(\mathbf{G^T C_D^{-1} G} + \mathbf{L^T C_L^{-1} L}\right)^{-1}\left(\mathbf{G^T C_D^{-1} \Delta d} - \mathbf{C_L^{-1}}(\mathbf{m^0} - \mathbf{m^{prior}})\right) \tag{6.7}$$

where $\mathbf{C_L}$ is the Laplacian covariance matrix, that we can assume to have the simplest form:

$$\mathbf{C_L} = \sigma_\mathbf{l}^\mathbf{2}\mathbf{I} \tag{6.8}$$

L is the finite difference approximation to the Laplacian operator applied over all model blocks. Each row of **L** is given by the difference between the target block and the average of the adiacent cells. For example in a 2-D model the Laplacian, ∇^2_{ij}, of the model cell m_{ij}, becomes (see equation A.16)



$$\nabla^2_{ij} \simeq \frac{1}{4}(m_{i-1,j} + m_{i+1,j} + m_{j+1,i} + m_{j-1,i}) - m_{i,j} \tag{6.9}$$

Equation (6.7) becomes,

$$\mathbf{\Delta m} = \sigma_\mathbf{d}^\mathbf{2}\left(\mathbf{G^T G} + \left(\frac{\sigma_\mathbf{l}}{\sigma_\mathbf{d}}\right)^{-\mathbf{2}}\mathbf{I}\right)^{-1}\sigma_\mathbf{d}^{-\mathbf{2}}\left(\mathbf{G^T \Delta d} - \left(\frac{\sigma_\mathbf{l}}{\sigma_\mathbf{d}}\right)^{\mathbf{2}}(\mathbf{m^0} - \mathbf{m^{prior}})\right) \tag{6.10}$$

Defining the *smoothing parameter* λ as

$$\lambda = \frac{\sigma_D}{\sigma_L} \tag{6.11}$$

we have,

$$\mathbf{\Delta m} = \left(\mathbf{G^T G} + \lambda^\mathbf{2}\mathbf{L^T L}\right)^{-1}\left(\mathbf{G^T \Delta d} - \lambda^\mathbf{2}(\mathbf{m^0} - \mathbf{m^{prior}})\right) \tag{6.12}$$

where λ controls the tradeoff between data misfit and model roughness.

The functional minimized by the *smoothed least squares solution* is

$$\Phi = \| \mathbf{G\Delta m} - \mathbf{\Delta d} \|^2 + \lambda^2 \| \mathbf{Lm} \|^2 \tag{6.13}$$

This type of regularization adds stability to the inversion in a different ways than damped least squares. The resulting models will be smooth, but not necessarily of minimum variance. As in the previous regularization condition, in figure 6.5 we can see how the resulting model depends on the smoothing parameter choice. The larger λ is, the smoother will be resulting model and the larger will be the misfit. Blocks that are not sampled by ray paths will be interpolated between nearby cells,or, more dangerously, extrapolated when they are near the edge of the model.

Both damped least squares and minimum roughness inversions have advantages and

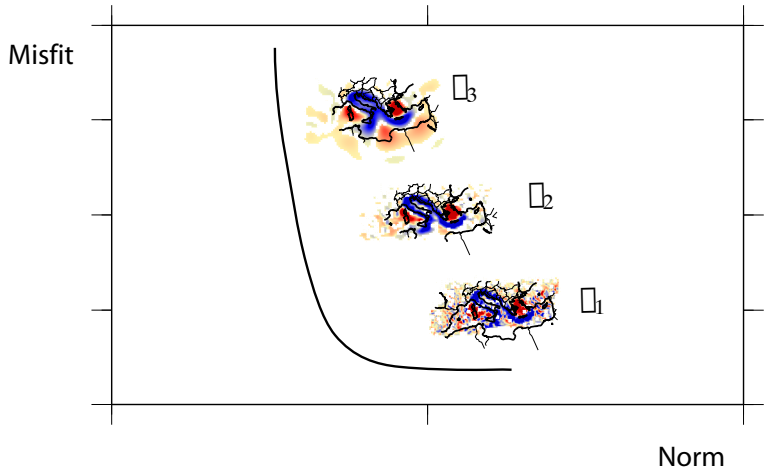


FIGURE 6.5: Schematic curve of tradeoff between misfit and model variance with the variation of laplacian parameter $\lambda_1 < \lambda_2 < \lambda_3$

disadvantages, and the best regularization to use will vary from problem to problem. In this work we use both the two approaches described.

The functional minimized by the *damped smoothed least squares solution* is

$$\Phi = \| \mathbf{G\Delta m} - \mathbf{\Delta d} \|^2 + \epsilon^2 \| \mathbf{m} \|^2 + \lambda^2 \| \mathbf{Lm} \|^2 \tag{6.14}$$

and the solution becomes

$$\mathbf{\Delta m} = \left(\mathbf{G^T C_D^{-1} G} + \mathbf{C_M^{-1}} + \mathbf{L^T C_L^{-1} L}\right)^{-1} \cdot \left(\mathbf{G^T C_D^{-1} \Delta d} - \mathbf{C_M^{-1}}(\mathbf{m} - \mathbf{m^{prior}}) - \mathbf{L^T C_L^{-1} L}(\mathbf{m^0} - \mathbf{m^{prior}})\right) \tag{6.15}$$

or after some algebra,

$$\mathbf{\Delta m} = \left(\mathbf{G^T G} + \epsilon^{\mathbf{2}}\mathbf{I} + \lambda^{\mathbf{2}}\mathbf{L^T L}\right)^{-1}\left(\mathbf{G^T \Delta d} - \epsilon^{\mathbf{2}}(\mathbf{m^0} - \mathbf{m^{prior}}) - \lambda^{\mathbf{2}}(\mathbf{m^0} - \mathbf{m^{prior}})\right) \tag{6.16}$$

Equation 6.14 can also be seen as the solution of the following system:

$$\begin{bmatrix} \mathbf{G} \\ \epsilon\mathbf{I} \\ \lambda\mathbf{L} \end{bmatrix} \Delta\mathbf{m} = \begin{bmatrix} \Delta\mathbf{d} \\ 0 \\ 0 \end{bmatrix} \tag{6.17}$$

It's important to note that the Equation 6.14 and system 6.17, as it is written, solves a linear tomographic problem (or the first iteration of a non linear tomography). In general the damping condition, as we do not want the resulting model to be too different from the starting model, must act on the total difference $\mathbf{m_n} - \mathbf{m_0}$, the smoothing condition acts on the total model $\mathbf{m}$, while the solution $\mathbf{\Delta m}$ refers to $\mathbf{m_n} - \mathbf{m_{n-1}}$, where n is the number of iteration done. Equation 6.14 for the n^{th} iteration (with $n > 1$) becomes:

$$\mathbf{\Delta m} = \mathbf{m_{n+1}} - \mathbf{m_n} = \left(\mathbf{G_n^T G_n} + \epsilon^{\mathbf{2}}\mathbf{I} + \lambda^{\mathbf{2}}\mathbf{L^T L}\right)^{-1} \cdot \left(\mathbf{G_n^T \Delta d} - \epsilon^{\mathbf{2}}(\mathbf{m_n} - \mathbf{m^{prior}}) - \lambda^{\mathbf{2}}\mathbf{L^T L}(\mathbf{m_n} - \mathbf{m^{prior}}))\right) \tag{6.18}$$

where $\mathbf{\Delta d}$ is the difference between the observed data and data predicted by the model obtained after the n^{th} iteration. System 6.17 is now:

$$\begin{bmatrix} \mathbf{G} \\ \epsilon\mathbf{I} \\ \lambda\mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{m_n} - \mathbf{m_{n-1}} \\ \mathbf{m_n} - \mathbf{m_0} \\ \mathbf{m_n} \end{bmatrix} = \begin{bmatrix} \Delta\mathbf{d} \\ 0 \\ 0 \end{bmatrix} \tag{6.19}$$

When the partial derivatives matrix is very large, it is necessary to use some algorithm that calculates the solution without explicitly computing the inverse matrix, in a computationally efficient way. System 6.19 can be seen as

$$\begin{bmatrix} \mathbf{G} \\ \epsilon\mathbf{I} \\ \lambda\mathbf{L} \end{bmatrix} \left[\mathbf{m_n} - \mathbf{m_{n-1}}\right] = \begin{bmatrix} \Delta\mathbf{d} \\ -(\mathbf{m_{n-1}} - \mathbf{m_0}) \\ -(\mathbf{m_{n-1}}) \end{bmatrix} \tag{6.20}$$

LSQR (Paige and Saunders, 1982; Nolet, 1987) is an iterative approach, based on the coniugate gradient algorithm, that approximates the exact lest squares. The order of the approximation depends on the number of iterations done. In this work, a stable solution is obtained after about 60 iterations. As damping parameter and Laplacian parameter we chose, respectevely $\epsilon = 2000$ and $\lambda = 500$

6.5 Recovery test

The forward computation, described by equation 2.8 and performed in the way described in the previous chapters, is done both for the syntetic model $\mathbf{m^S}$ and for the 1-D model $\mathbf{m^0}$. The former gives us the data vector $\mathbf{d^{Ob}}$, the latter gives us the the predicted vector in the 1-D model $\mathbf{d^0}$. Then, the inversion is performed and a new model $\mathbf{m^1} = \mathbf{m^0} + \mathbf{\Delta m}$ is obtained. As explained in chapter two, the approximation leading to the solution formulation, works well only if the actual model is close to the starting model. If this is not the case the method has to be iterated until successive iterations produce only small changes in the model and in the data misfit. In this test, the procedure is followed for three successive iterations.

Figure 6.6 shows points on the curve of trade-off between model norm and data misfit for the following iteration. We refer to the relative data misfit defined as

$$\Phi = \frac{\| \mathbf{G\Delta m^n} - \mathbf{d^{Ob}} \|^\mathbf{2}}{\| \mathbf{G\Delta m^0} - \mathbf{d^{Ob}} \|^\mathbf{2}} \tag{6.21}$$

where $\mathbf{m^n}$ is the model vector obtained after the n^{th} inversion. In fact, since the predicted vector $\mathbf{d^3}$, of the model $\mathbf{m^3}$, is computed only in the fourth iteration the plot is referred to the first two. It shows a misfit reduction of about 90 % after the first iteration and a trend towards a complete reduction with successive iterations.

Figure 6.7 shows another way of seeing the same data misfit reduction. The histogram represents the traveltime residuals after the first (in red) and third (in blue) iteration. It is evident the non-normality and asymmetry of the starting residuals and the successive

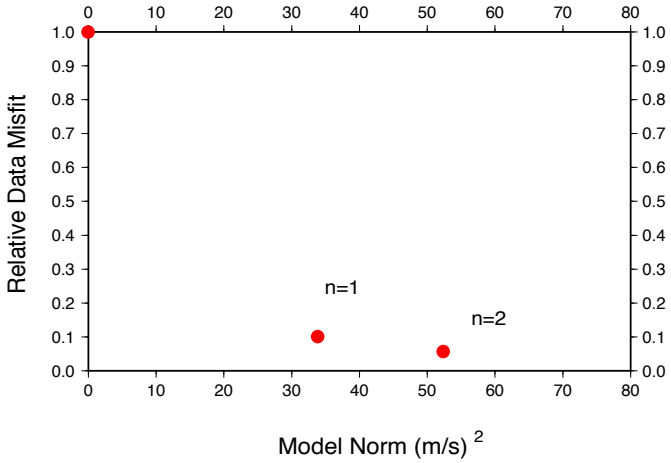


FIGURE 6.6: Curve of tradeoff between misfit and model norm for different iteration.

closing to the zero after the third iteration.

Ray paths provide a rather inhomogeneous illumination of the mantle structure. Various measures of the rays sampling can be adopted, such as the ray density tensor (Kissling 1988), or the cell hit count. The cell hit count is defined as the number of rays sampling a cell. The importance of this quantity is not only that it provides insight into how well cells (and hence mantle areas) are sampled, but also that it has a strong effect on the solution, in the sense that amplitudes of imaged structure correlate to some extent with the cell hit count (Spakman 93). Therefore before inspecting the result of the inversion, it is important to analyze the path distribution over the model volume. We use a similar quantity but referred to our model parametrization. For each node we compute the number of times an adjacent cell has been sampled by a ray.

Figure 6.8 shows rays sampling at different depth, rays are traced in the 1-Dimensional model $\mathbf{m^0}$. The ray sampling varies considerably in the model. The best coverage is located beneath central Europe while the poorest illuminated cells are found below northern Africa and beneath the eastern Europe. In these regions the coverage is limited to small areas, unevenly sampled, due to sparse location of events and stations and to their position in the margin of the grid. The first layer is strongly dependent on earthquakes and station location (refer to Figure 6.2 and 6.3). The 50 km depth layer is dominated by the horizontally travelling P_n rays sampling the lithosphere. Events

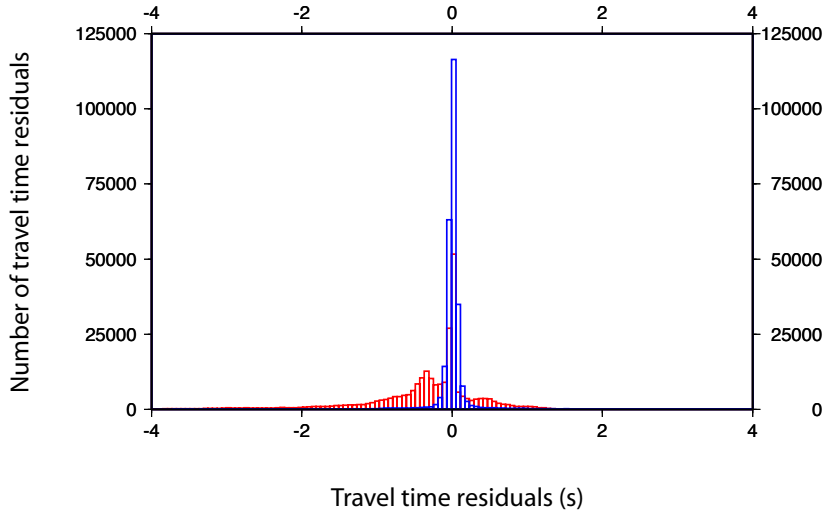


FIGURE 6.7: Histogram of travel time residuals for the first iteration (red) and for the third iteration(blue). Count for a bin interval of 0.06 s.

located in the Hellenic-Aegean region recorded by European stations provide a good coverage of the central part of the model, although most rays have a preferential direction SE-NW. Below 150 km the pattern becomes smoother, restricted to the cental grid area and the maximum density becomes smaller.

We show now the result of the syntetic test for the first 200 km. Figure 6.9 shows the model expressed as percentage velocity perturbation with respect to the reference model sp6.

From a comparison of Figure 6.8 and 6.10, we observe the correlation between ray sampling and the obtained anomaly amplitude. It is evident a very good agreement between the input and the resulting model in the first two layers, the correlation is 0.9 for the first iteration result and increases to 0.93 after the third. With this geometry, ray coverage decreases significantly from 150 km and below, therefore the input moTyrrhenian and Aegean seas, and generally enhances the picture, showing the importance of nonlinearities in ray tracing.

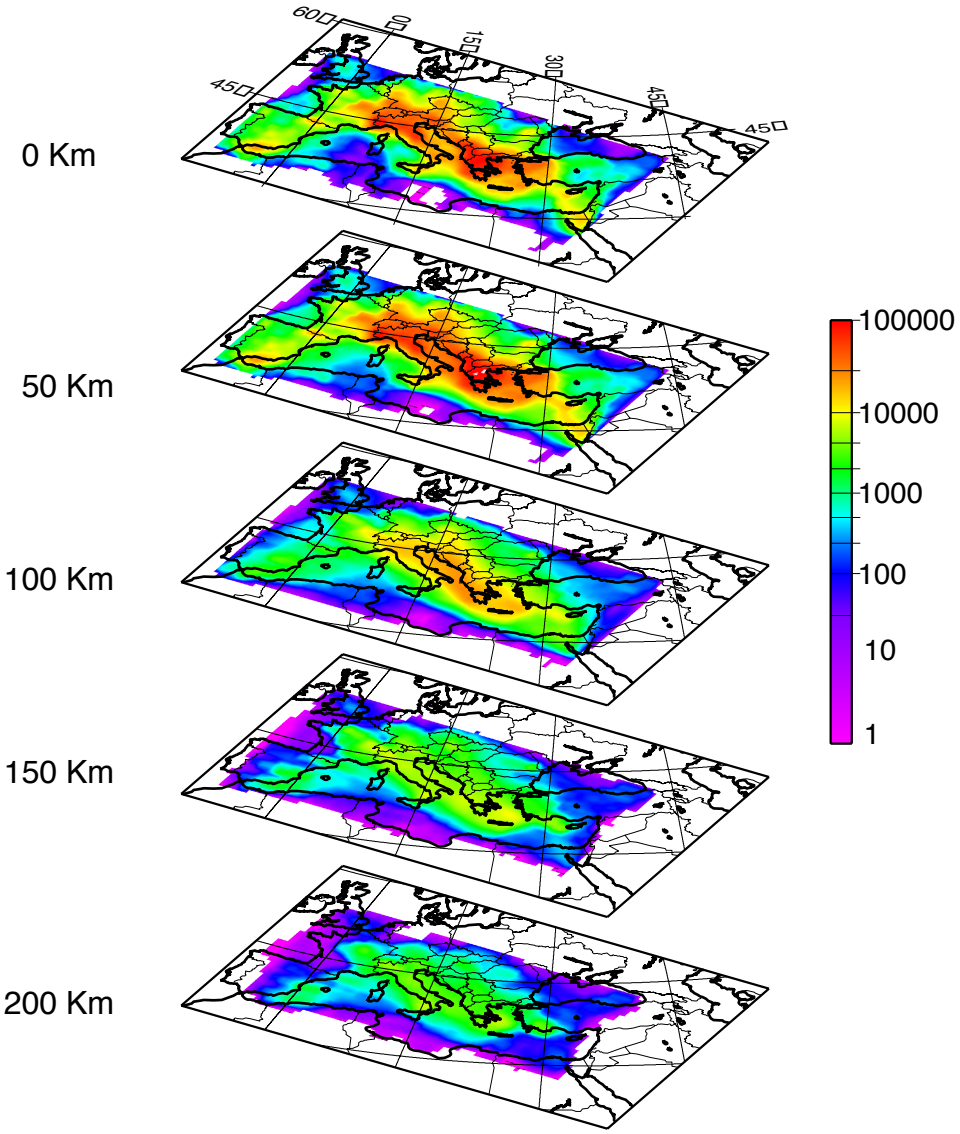


FIGURE 6.8: Ray density express by the number of times a node is adjacent to a cell sampled by a ray.

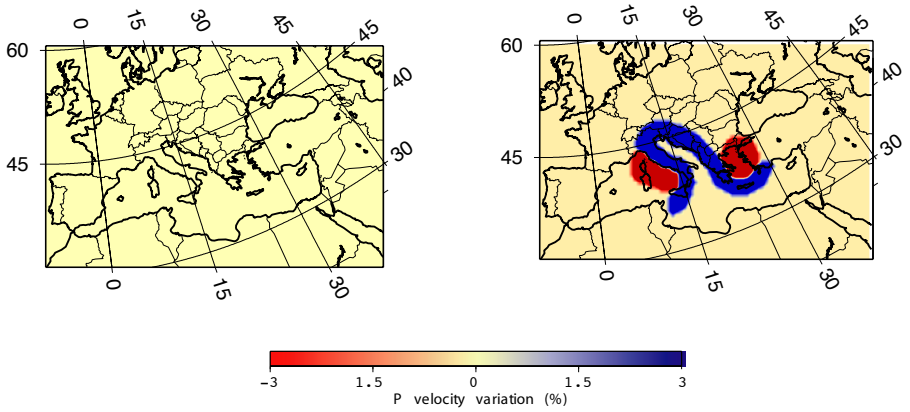


FIGURE 6.9: Input model to be reconstructed. Surface layer (left) is without any anomaly. The right plot shows the anomaly for the layer between 50 and 200 km.

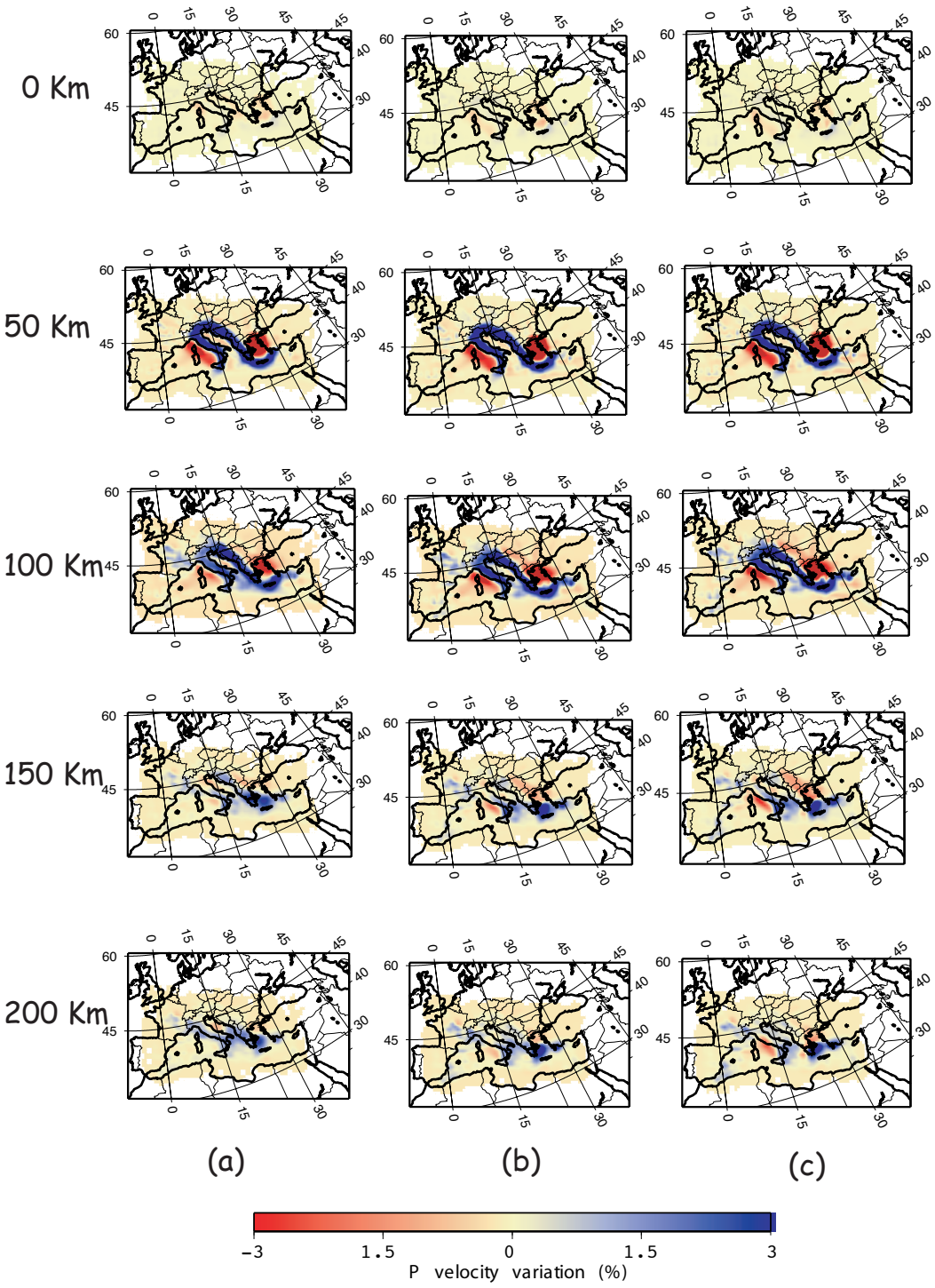


FIGURE 6.10: Result of sintetic test after the first (a), second (b) and third (c) iteration.

Chapter 7

Iterative non linear Inversion of Seismic Traveltime Data

As for the recovery test, we chose the same parameterization adopted by PM0.5. A slowness, from the 1-Dimensional model sp6, is associated to each node of a 3D lattice of 121x71 nodes in the horizontal direction and 21 nodes in the vertical direction for a total of 180411 model parameters. The mesh spacing is of 0.5 degree in the horizontal direction and 50 km in the vertical.

P wave travel times reported by the EHB bulletin for the time period 1982-2004 constitute our data set. Events and stations location are the same used in the syntetic test and can be seen in Figure 6.2 and 6.3. The forward problem is solved, in the way described in Chapter 3 and 4, for the 1-Dimensional model parameterized by $\mathbf{m^0}$. Then, the data vector is computed by the difference between the real observed and predicted data $\mathbf{d^{Ob}} - \mathbf{d^0}$, considering only residuals with a travel time less than 8 s. This produces a selection of about 3,039,000 rays. The solution to the inverse problem is sought by the least squares minimization through the iterative LSQR algorithm. In addition to the minimum norm request, we introduce a minimum roughness condition, to constrain ill-determined model parameters. The roughness condition is applied by minimizing the Laplacian of each cell (description in section 6.3) As described in the previous chapter for the recovery test, we find a solution following the iterative non linear tomography. Inversion steps are alternated with 3-D ray tracing to update ray path and traveltimes. We show the result of the first three iterations (the forward computation of the fourth

iteration has been done in order to find how well $\mathbf{m^{(3)}}$ satisfy the data).

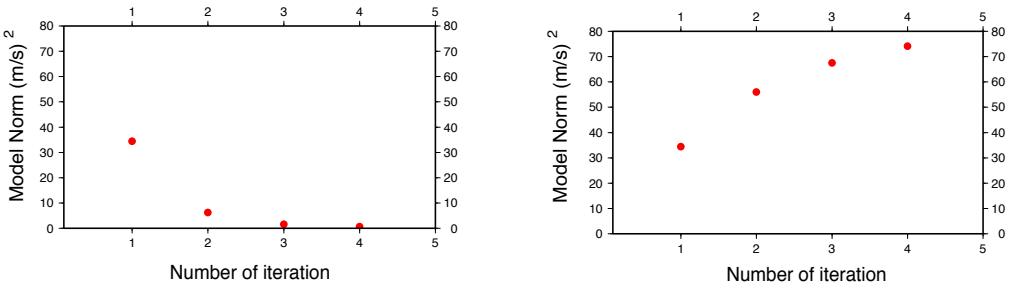


FIGURE 7.1: Model Norm variation respect to previous iteration model (left) and Model Norm variation respect to initial model (right)

Figure 7.1 shows the model variance behaviour as a function of the number of iteration. The model change $\mathbf{m_n} - \mathbf{m_{n-1}}$, where n is the iteration number, is plotted to the left. The total model change from the initial model $\mathbf{m_n} - \mathbf{m_0}$, is shown to the right. It is evident how successive iterations produce smaller and smaller changes. Figure 7.2

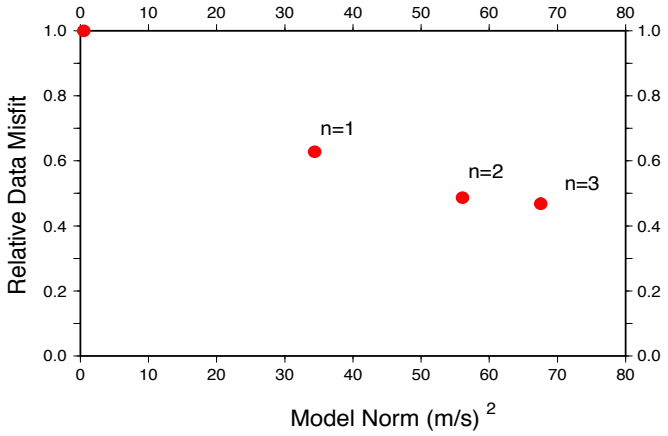


FIGURE 7.2: Tradeoff curve between relative data misfit and model norm for different iteration.

shows the trade-off curve between model norm and data misfit for successive iterations. Although the largest reduction is for the first iteration (about 36%), successive iterations lead to a significant improvement in data fit. After the third iteration the misfit

reduction is about 55%.

Figure 7.3 shows rays sampling for the first and the last iteration. As we use the same initial model and the same stations and events location, considerations done in the previous chapter are still valid. The cell hit count is the same for rays travelling in the 1-Dimensional model. The best coverage is located beneath central Europe while the poorest illuminated cells are found below northern Africa and beneath the eastern Europe. In these regions the coverage is limited to small areas, unevenly sampled, due to sparse location of events and stations and to their position in the margin of the grid. The first layer is strongly dependent on earthquakes and stations location. The 50 km depth layer is dominated by the horizontally travelling P_n rays sampling the lithosphere. Events located in the Hellenic-Aegean region recorded by European stations provide a good coverage of the central part of the model, although most rays have a preferential direction SE-NW. Below 150 km the pattern becomes smoother, restricted to the cental grid area and the maximum density becomes lower.
Figure 7.4 shows the difference of the cell hit count between the first and the last iteration for a 50 km depth layer (top) and 100 km depth layer (bottom).

Although the global geometry of ray paths does not seem to be very different from the linearized inversion (represented by the first iteration), strong differences in path density exist, we note how the hit count for the non-linear inversion after few iterations is generally increased in the subduction zone regions.

Figure 7.5 and 7.6 show, respectively, the model resulting from the linear inversion and the model resulting from the non linear inversion after three iterations.
The model is expressed as percentage velocity perturbation with respect to the 1-D model sp6. The average anomaly over a horizontal layer is subtracted for clarity (for this reason the 50 km map view has a gray background). According to [Bijwaard and Spakman, 2000] a strong change in the anomaly pattern is not expected but systematic changes due to ray tracing in the updated models, occur in region with a sufficient level of velocity heterogeneity. As observed for the recovery test, the amplitude of the model is strictly dependent on rays coverage. Since teleseismic data, essential to image deeper structure, are not used in this experiment, we limit our analysis to the shallower 200 km. The non linear method leads to a general increase in the model amplitude and more

focused images. Misfit is also subject to a reduction that increases with the first two iterations and seems to become stable after the third.

As noted by [Piromallo and Morelli, 2003] the crustal layer at 0 km depth may suffer from the inhomogeneous distribution of stations and events, the effects of unmodeled shallow structure, and of smearing from the close layer at 50 km. Amplitudes are large in well sampled areas and velocity contrasts are quite sharp. This velocity pattern may be affected by a trade-off between variations in crustal thickness and velocity anomalies. We chose not to explicitly invert for crustal thickness and velocity because we do not think that the data set we used is able to break that trade-off.

In the 50 km deep layer, we see how fast anomalies characterize the lithosphere of the basins (Adriatic, Ionian basin and Mediterranean). The fast anomaly in the easthern part of Mediterranean is also visible in deeper layers. Slow anomalies are visible below the Tyrrhenian and Aegean sea. The best resolved structure is the fast velocity anomaly along the Dinarides, Hellenides and the Hellenic chain, that especially for the non-linear inversion assumes a strong continuity at a depth of 100 km. According with PM0.5, the model shows strong difference between the northern and southern Appennines, in fact below the northern part of the chain we find a fast anomaly visible until 100 km depth while in the southern part a slow anomaly is present. At a depth of 150-200 km a fast velocity body is visible along the cost of Tyrrhenian sea. It may represent the subduction of the Adriatic plate under the Italian peninsula. In agreement with other models, slow anomalies are also displayed below the Massif Central and Rhenish Massif.

This experiment was carried on using a x86_64-based NUMA[1] system. The system is equipped with 4 dual-core opteron 8214 CPUs at 2.2 Ghz each and 32 GB of DDR-2 RAM. The computational time required by each iteration, for the grid described, is about 10 hours. Calculations are inherently parallel, therefore the code can easily be parted on a parallel computer cluster, where we expect excellent scalability.

[1] Non Uniform Memory Access

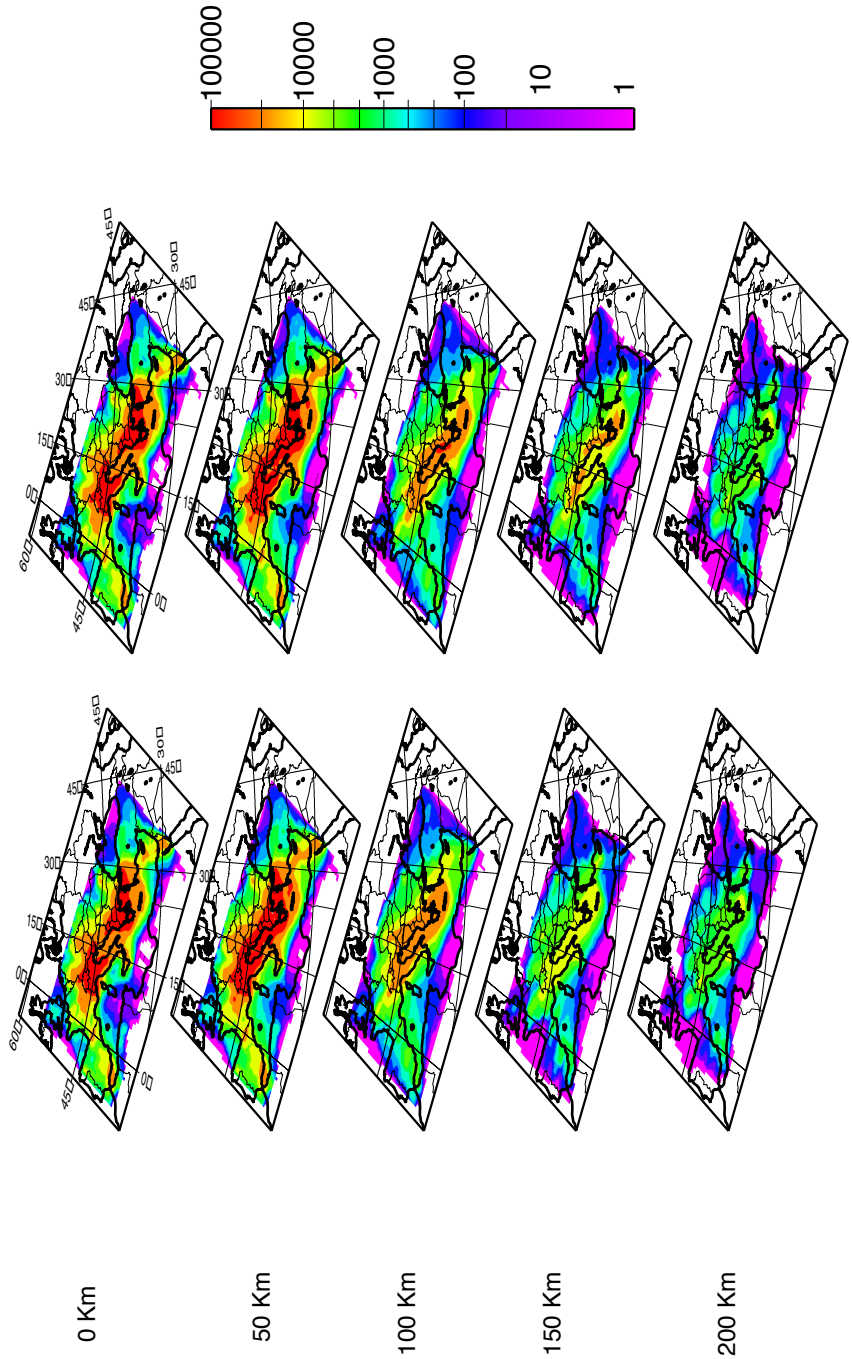


FIGURE 7.3: Ray density express by the number of times a node is adjacent to a cell sampled by a ray.

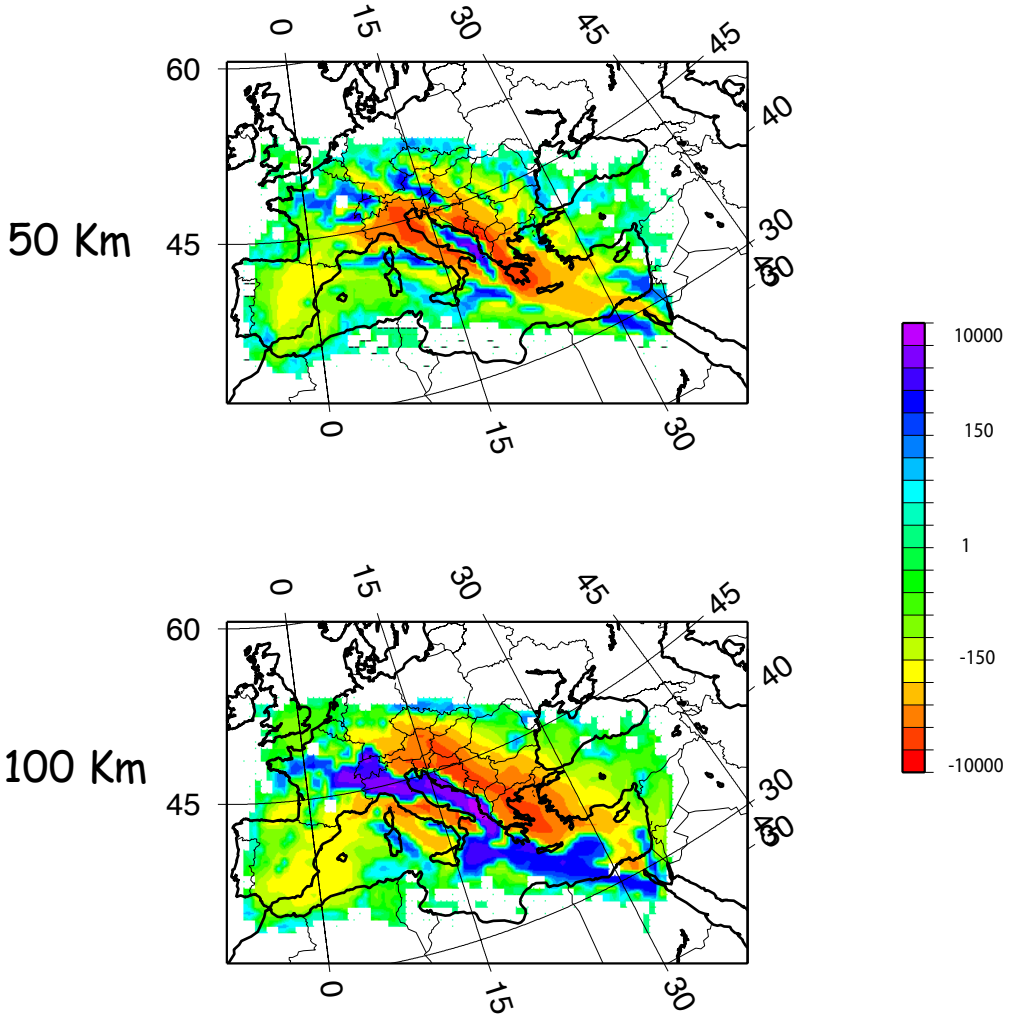


FIGURE 7.4: Logarithmig hit count pattern espressed by the difference between rays traced in the 3^{rd} iteration resulting model and rays traced in the 1-D model. Map view at 50 km (top) and 100 km (bottom)

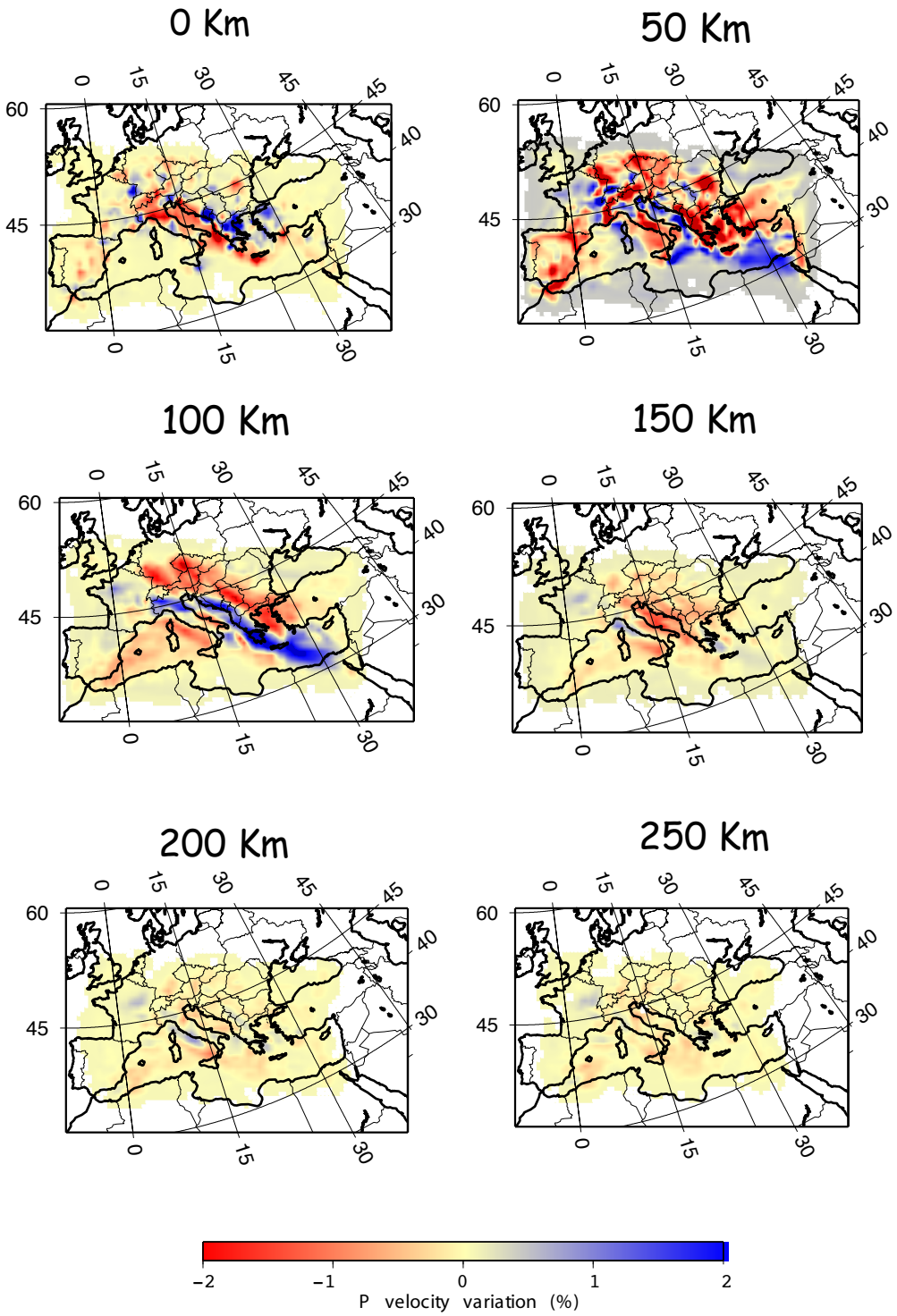


FIGURE 7.5: Map views of the model resulting from the first iteration. P velocity perturbation with respect to reference model sp6 is displayed. Not illuminated areas are blanked. Grey color, at 50 km depth is an artificial effect due to our choice of plot each layer without the mean.

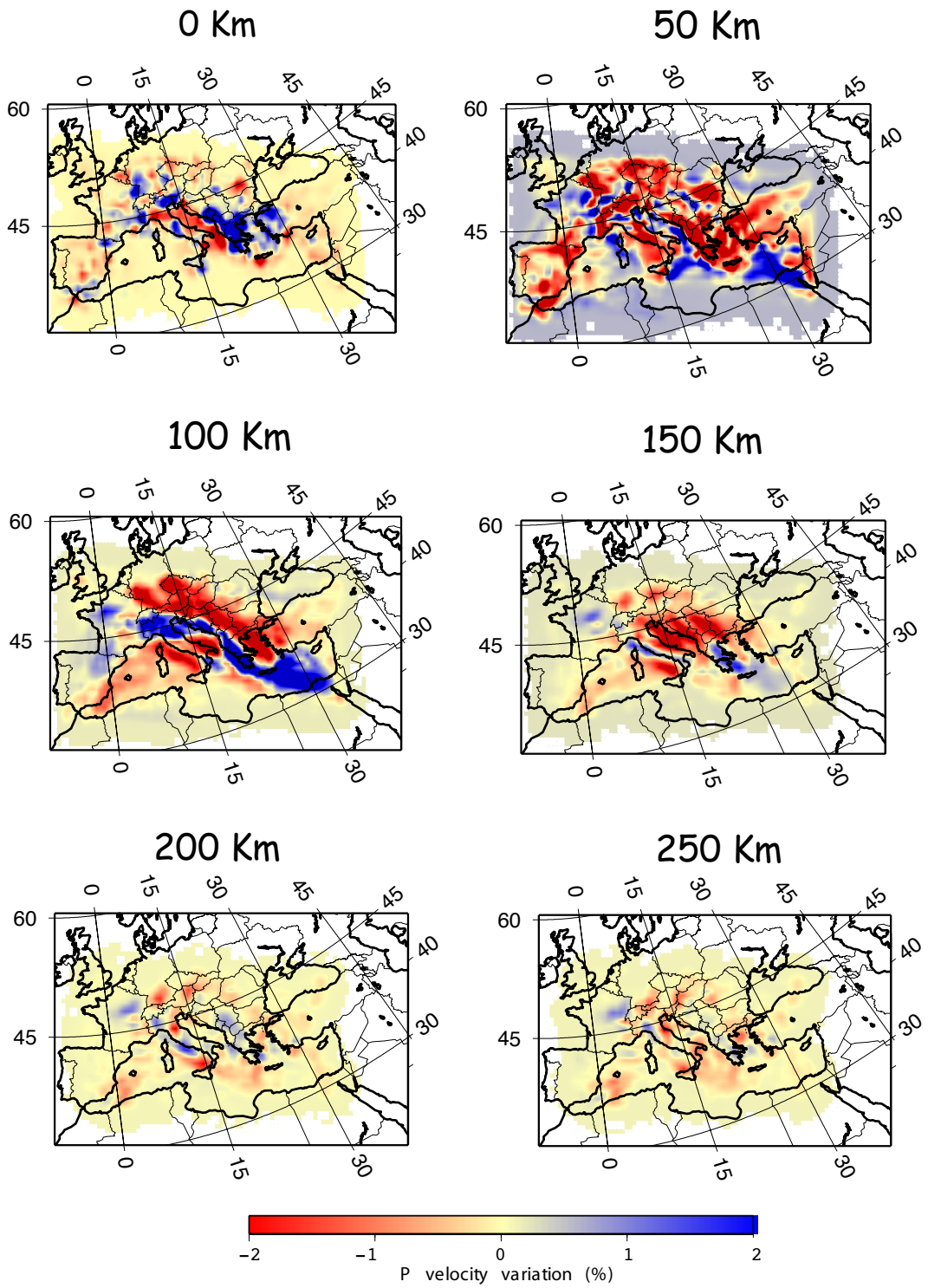


FIGURE 7.6: Map views of the model resulting from the third iteration. P velocity perturbation with respect to reference model sp6 is displayed. Not illuminated areas are blanked. Grey color, at 50 km depth is an artificial effect due to our choice of plot each layer without the mean.

Chapter 8

Conclusion

We developped a method based on iterative non linear tomographic inversion in which the forward computation is done with finite-difference numerical schemes. The method is computationally expensive but, with recent computing improvements, it is appropriate for use for regional and continental scale upper mantle seismic tomography.

We verify that travel time and path calculations, performed with finite difference methods, are very significant and can be particularly important for correctly locating the ray with depth, when heterogeneities are present. The test has been done in a realistic tomographic 3-D model.

The Two-Grid implementation is an innovative (for our knowledge) way that allows to perform global tomography with a good compromise between computational cost and accuracy. In fact it allows to have a finer grid discretization in the crust, where it is necessary to better describe the Moho discontinuity, but it also preserves computational efficiency by use of the the coarser grid describing the mantle.

From the application of the non linear tomography to the Euro-Mediterranean area we see how the method results more effective, as the data misfit of the model, obtained after three iterations, is reduced by about a 20% with respect to the linear approach. The synthetic test performed shows how the model reconstruction is improved by the non-linear method.

As the forward computation is completely independent for each source, it can be easily calculated separately. As a consequence, the method scales well from a single CPU desktop system to a multiprocessor high performance computer and it is suitable for large-scale parallel computing. In fact, although the method is appropriate for use

for continental scale tomography, the computational time required suggests to adopt the implementation on a parallel cluster.

Future developments will address the comparison between the embedding strategy and the azimuthal equidistant projection used in this work. Although the latter introduces an approximation, the former could produce numerical problems due to the step representation of the Moho discontinuity. With recent computational performance improvement, the comparison now can be performed using the same forward grid discretization.

Future work will also include teleseismic data into the tomographic inversion, as they are essential to image deeper mantle structure. As explained in Chapter 5, this is feasible by using a hybrid approach that first initializes the travel time in the bottom of the grid, taking the global lower mantle anomalies into account and then propagating the wave field by the usual finite difference scheme.

Finally, the method easily allows to include advanced non-linear techniques for earthquake location, such as the NonLinLoc [Lomax et. al., 2000 and 2001] in a joint inversion.

Appendix A

Finite Difference Methods Review

Finite difference methods are useful where the domain of interest is represented by a set of points or nodes. Taylor series expansions play a fondamental role in the formulation of finite difference schemes. Given a continuous function $u(x)$ we can discretize the x domain into a set of nodes (Figure A.1) such that

$$u(x_r) \equiv u(rh) \equiv u_r, \qquad r = 0, 1, 2, ...$$

The Taylor expansion for $u(x)$ can be written at the point x_r as

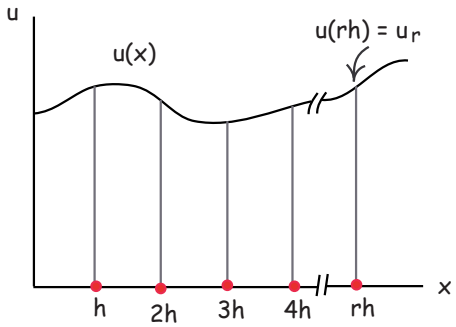


FIGURE A.1: Finite difference discretization of the x domain for a given continuous function $u(x)$

$$u(x_r + h) = u(x_r) + hu_x\Big|_r + \frac{h^2}{2!}u_{xx}\Big|_r + \frac{h^3}{3!}u_{xxx}\Big|_r + ...$$

or

$$u(x_r - h) = u(x_r) - hu_x\Big|_r + \frac{h^2}{2!}u_{xx}\Big|_r - \frac{h^3}{3!}u_{xxx}\Big|_r + ...$$

where $u_x|_r$, $u_{xx}|_r$ and $u_{xxx}|_r$ represent, respectively, the first, the second and the third derivative with respect to x.
The first derivative $u_x|_r$ can be written as

$$u_x\Big|_r = \frac{u(x_r+h)-u(x_r)}{h} - \frac{h}{2!}u_{xx}\Big|_r - \frac{h^2}{3!}u_{xxx}\Big|_r - ... \tag{A.1}$$

$$u_x\Big|_r = \frac{u(x_r)-u(x_r-h)}{h} + \frac{h}{2!}u_{xx}\Big|_r - \frac{h^2}{3!}u_{xxx}\Big|_r + ... \tag{A.2}$$

Two possible approximations are:

$$u_x\Big|_r = \frac{u(x_r+h)-u(x_r)}{h} \equiv \frac{u_{r+1}-u_r}{h} \tag{A.3}$$

$$u_x\Big|_r = \frac{u(x_r)-u(x_r-h)}{h} \equiv \frac{u_r-u_{r-1}}{h} \tag{A.4}$$

Clearly, because the series has been truncated, there is an error, associated with this approssimation. This error, E_r, can be characterized by the first and largest term of the truncated series, which yields

$$E_r = \pm\frac{h}{2}u_{xx}\Big|_\xi = O(h), \qquad \begin{matrix} x_r \le \xi \le x_r + h \\ x_r - h \le \xi \le x_r \end{matrix}$$

For sufficiently small h, this error (of the order of h, $O(h)$) is in absolute value smaller than Ah (A is constant)
If we add (A.1) and (A.2) we obtain

$$u_x\Big| = \frac{u_{r+1}-u_{r-1}}{2h} \tag{A.5}$$

with an error $O(h^2)$ in fact the first truncated term is

$$-\frac{h^2}{6}u_{xxx}\Big|_\xi, \qquad x_{r-1} \le \xi \le xr+1$$

In this way, we can continue to develop more formulas, for example we can solve the second derivative of u at x_r subtracting (A.2) from (A.1)

$$u_{xx}\Big|_r = \frac{u_{r+1}-2u_r+u_{r+1}}{h^2} \tag{A.6}$$

The results, just described may be extended in a straightforward manner to derive finite difference approximation in two or in three dimension. Given a continuous function of

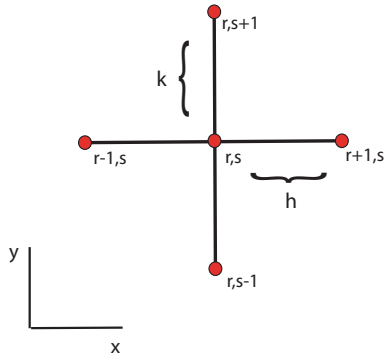


FIGURE A.2: Schematic illustration of the two-dimensional finite-difference grid

two independent variables $u(x, y) \equiv u_{r,s}$, the first derivative with respect to x can be derived from (A.3)

$$\frac{\partial u(x,y)}{\partial x} \equiv u_x\Big|_{r,s} = \frac{u_{r+1,s} - u_{r,s}}{h} + O(h) \tag{A.7}$$

In this approximation $u_x|_{r,s}$ the subscript s is held costant (the partial derivative with respect to x implies that y is held costant). In the same way we obtain the first derivative with respect to y

$$\frac{\partial u(x,y)}{\partial y} \equiv u_y\Big|_{r,s} = \frac{u_{r,s+1} - u_{r,s}}{h} + O(k) \tag{A.8}$$

where the subscript r is held costant. Other formulas with a smaller error can be obtained, considering a larger number of nodes, in analogy with A.5

$$u_x\Big|_{r,s} = \frac{u_{r+1,s} - u_{r-1,s}}{2h} + O(h^2) \tag{A.9}$$

$$u_y\Big|_{r,s} = \frac{u_{r,s+1} - u_{r,s-1}}{2h} + O(k^2) \tag{A.10}$$

or

$$u_x\Big|_{r,s} = \frac{u_{r+1,s+1} - u_{r-1,s+1} + u_{r+1,s-1} - u_{r-1,s-1}}{4h} + O(h^2) \tag{A.11}$$

$$u_y\Big|_{r,s} = \frac{u_{r+1,s+1} - u_{r+1,s-1} + u_{r-1,s+1} - u_{r-1,s-1}}{4h} + O(k^2) \tag{A.12}$$

The second derivative, with respect to x and y can be written, in analogy with (A.6)

$$u_{xx}\Big|_{r,s} = \frac{u_{r+1,s} - 2u_{r,s} + u_{r-1,s}}{h^2} + O(h^2) \tag{A.13}$$

$$u_{yy}\Big|_{r,s} = \frac{u_{r,s+1} - 2u_{r,s} + u_{r,s-1}}{k^2} + O(k^2) \tag{A.14}$$

We could continue to obtain more complex formulas with different degree of approximation, for a more exhaustive description refer to [Lapidus & Pinder, 1982]. To illustrate how we can apply the previous formulas, we consider the Laplace equation

$$u_{xx} + u_{yy} = 0 \tag{A.15}$$

it can be approximated as

$$[u_{xx} + u_{yy}]_{r,s} = \frac{u_{r+1,s} - 2u_r, s + u_{r-1,s}}{h^2} + \frac{u_{r,s+1} - 2u_r, s + u_{r,s+1}}{k^2} + O(h^2 + k^2) = 0$$

when $h = k$ it becomes

$$u_{r+1,s} + u_{r-1,s} + u_{r,s+1} + u_{r,s-1} = 4u_{r,s} + O(h^2) \tag{A.16}$$

This is the finite difference representation for Laplace's equation with error $O(h^2)$.

Bibliography

Aki, K., Christoffersson, A., and H. Husebye, 1977. Determination of the three-dimensional seismic structureof the lithosphere,*J. Geophys. Res.Int.,* **82**, 277-296

Aki, K. and P.G. Richards, 2002. Quantitative seismology, Second Edition. University Science Books, Sausalito, California, USA

Bassin, C., Laske, G. and G. Masters, 2000. The Current Limits of Resolution for Surface Wave Tomography in North America, EOS Trans AGU, 81, F897.

Bijwaard, H., and Spakman, W., 2000. Non-linear global P-wave tomography by iterated linearized inversion.*J.Geophys. Res.,* **141**, 71-82

Buland, R., and C. H. Chapman, 1983. Determination of the three-dimensional seismic structureof the lithosphere, *Bull. Seis. Soc. Am., 73(5), 1271-1302.*

Červený, V., 2001. Seismic ray theory, Cambridge University Press, Cambridge.

H. P. Crotwell, T.J. Owens, and J. Ritsema, 1999. The taup toolkit: Flexible seismic travel-time and ray-path utilities. *Seismological Research Letters* **70**, 154-160

Dziewonski, A. M., and D. L. Anderson, 1981. Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, **25**, 297 - 356.

Dziewonski, A.M, B.H. Hager and R.J. OConnel (1977): Large-scale heterogeneities in the lower mantle, *J.Geophys. Res.,* **82**, 239-255

Ehgdahl, R., R. van der Hilst, and R. Buland, 1998. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seismol. Soc.,* **88**, 722-743, 1998.

Goes, S., R. Govers, and P. Vacher, 2000. Shallow mantle temperatures under Europe frm P and S wave tomography, *J. Geophys. Res.,* **105**, 11,153-11,169

Gorman, A.R., 2002. Ray-theoretical seismic traveltime inversion: modifications for a two-dimensional radially parametrized Earth. *Geophys. J. Int.,* **151**, 511-516.

Gorbatov, A., Fukao, Y., Wdiyantoro, S., & Gordeev., E., 2001. Seismic evidence for a mantle plume oceanwards of the Kamchatka-Aleutian trench junction, *Geophys. J. Int.,* **146**, 282-288. /]ISC1 Internation Seismological centre, *On line Bulletin*, http://www.isc.ac.uk, Internatl. Seis. Cent., Thatcham, United Kingdom, 2001

Lapidus, L. and Pinder, G.F., 1982. Numerical solution of partial differential equations in science and engineering. John Wiley & Sons.

Lévêque, J.J., Rivera, L. and G. Wittlinger, 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int*, **115**, 313 - 318.

Lomax, A., J. Virieux, P. Volant and C. Berge, 2000. Probabilistic earthquake location in 3D and layered models: Intoduction of a Metropolis-Gibbs method and comparison with linear locations, *Advances in Seismic Event Location* Thurber, C.H., and N. Rabinowitz (eds.), Kluwer, Amsterdam, 101-134

Lomax, A., A. Zollo, P. Capuano, and J. Virieux, 2001. Precise, absolute earthquake location under SOmma-Vesuvius volcano using a new 3D velocity model, *Geophys. J. Int.*, **146**, 313-331

Lucente, F.P., Chiarabba, C., Cimini, G.B. and D. Giardini, 1999. Tomographic constraints on the geodynamic evolution of the Italian region, *J. Geophys. Res.* , **104**, 20307 - 20327.

Morelli, A. & Dziewonski A., 1993. Body wave traveltimes and a a spherically symmetric P- and S-wave velocity model, *Geophys. J. Int.,* **112**, 178-194.

Morelli, A., and C. Piromallo, 2000. The late stage of retreating subduction in the Alpine-Mediterranean region: constraints from travel time seismic tomography, in: Problems in geophysics for the new millennium, E. Boschi, G. Ekstrom, and A. Morelli eds., Editrice Compositori, Bologna, pp. 179-216.

Morelli, A., and C. Piromallo, 2000. Imaging the Mediterranean upper mantle by p-wave travel time tomography, *Annals of Geophys.,* **40, N.4.**

Approximate treatment of elastic body waves in media with spherical symmetry, *Geophys.J.Roy.Astron.Soc,* **23**, 435-449

Nolet, G., Seismic wave propagation and seismic tomography, in *Seismic tomography*, edited by G. Nolet,pp.1-23, D.Reidel, Norwell,Mass., 1987

Piromallo, C. and A. Morelli, 2003. P wave tomography of the mantle under the Alpine-Mediterranean area, *J. geophys. Res.*, **108**, NO. B2, 2065, doi:10.1029/2002JB001757

Paige, C. C., and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Software, 8,43-71, 1982*

Ritsema, J. and H. J.van Heijst, 2000. New seismic model of the upper mantle beneath Africa, *Geology*, **28**, 63 - 66.

Podvin, P. & Lecompte, I.,1991. Finite difference computation of traveltimes in very contrasted velocity models. A massively parallel approach and its associates tools, *Geophys. J. Int.,* **105**, 271-284.

Qin, F., Luo, Y., Olsen, K.B., Cai, W., Schuster, G.T.,1992. Finite difference solution of the eikonal equation along expanding wavefronts, *Geophysics*, **57**, 478-487.

Ranalli, G., 1996. Seismic tomography and mineral phisics, in *Seismic Modelling of the earth structure*, edited by E. Boschi et al., pp 443-459, Ed. Compositori, Bologna, Italy

Rawlinson, N. and Sambridge, M., Seismic traveltime tomography of the crust and lithosphere *Advances in Geophysics*, **46, 81-197**

Peter M. Shearer, 1999. Introduction to seismology. Cambridge University Press, Cambridge.

Sobolev, S. V., H. Zeyen, M. Granet, U. Achauer, C. Bauer, F. Werling, R. Altherr, and K. Fuchs. Upper mantel temperatures and lithosphere-asthenosphere system beneath the French Massif Central constrained by seismic, gravity, petrologic and thermal observations. *Tectonophysics,* **275**, 143-164.

Spakman, W., Van der Lee, S. and R.D. Van der Hilst, 1993. Travel-time tomography of the European-Mediterranean mantle down to 1400 km, *Phys. Earth Planet. Inter.*, **79**, 3 - 74

Stein, S., M. Wysession, 2003. An introduction to Seismology, Earthquakes, and Earth Structure. Edited by Blackwell Publishing Ltd.

Tarantola, A., 2005. Inverse Problem Theory and Model Parameter Estimation, SIAM, Philadelphia

Tarantola, A., and B. Valette, 1982. Generalised nonlinear inverse problems solved using the least square criterion, *Rev. Geophys.*, **20**, 219 - 232

Thomson, C.J. & Gubbins, D., 1982. Three-Dimensional lithospheric modelling at NORSAR: linearity of the method and amplitude variations from the anomalies, *Geophys. J. R. astr. Soc.,* , **71**, 1-36

Trampert, J. and R. D. Van der Hilst, 2005. Towards a quantitative interpretation of Global Seismic Tomography, in: Earth's Deep Mantle: Structure, Composition and Evolution. *Geophysical Monograph Series*, **160**, AGU.

Thurber, C.H., 1983. Earthquakes location and three-dimensional crustal structure in the Coyote Lake area, central California, *J. Geophys. Res.,* **88**, 8226-8236.

Tryggvason, A. & Bergman, B., 2006. A traveltime reciprocity discrepancy in the Podvin & Lecompte *time3d* finite difference algorithm, *Geophys. J. Int.,* **165**, 432-435.

Vidale, J., 1988. Finite difference calculation of travel times, *Bull. seism. Soc. Am.,* **78**, 2062-2076

Vidale, J., 1990. Finite difference calculation of travel times in 3D, *Geophysics,* **51**, 889-901

Wessel, P., and W. H. F. Smith, 1995. New version of the Generic Mapping Tools released, Eos Trans. AGU, 76, 329