UNIVERSITÀ DEGLI STUDI DI BOLOGNA
ALMA MATER STUDIORUM
ANNO ACCADEMICO 2007/2008

TESI DI DOTTORATO IN GEOFISICA
XX CICLO

Multiresolution Spherical Wavelet Analysis in Global Seismic Tomography

PhD Student:

Simona Carannante

Tutor:

Dr. Andrea Morelli

Contents

Introduction

Every seismic event produces seismic waves which travel throughout the Earth.
Seismology is the science of interpreting measurements to derive information about earthquakes and the structure of the Earth. It is usually driven by observations and advantaged by improvements in the instrumentation and progresses in seismological theory.
Global tomography is one of the possible application field of seismic studies. Seismic waves from big enough earthquakes can be observed throughout the world, and as they travel deep through the Earth or along the surface, their arrival times and shapes are impregnated with information on the medium they travelled through. The inverse problem in seismic tomography consists of mapping the Earths two and three-dimensional elastic velocity field from a large quantities of seismic measurements. Over the last two decades numerous tomographic models of the Earth's interior have been derived improving our knowledge of geophysical and geodynamical processes.

Several factors can affect the quality and the reliability of $2D$ or $3D$ global tomographic map derived from seismic tomography. These factors are separately objects of investigation by side of different research groups; the goal of all those studies is to obtain a common knowledge of Earth structure and of the physical processes concerned it.
The nature of data can affect, for example, the characteristics of our model; as it is well known we can use different components of seismic wave record like travel times, waveform, amplitude or spectra but this clearly lead to the application of different methods for determining Earth structure which can produce, sometimes, different or even opposite results.
Theory at the basis of tomographic problem is another important factor of influence. At the moment there are more than one theory used in seismic tomography to implement the tomographic problem; the evaluation of reliability of one theory respect to another one is currently object of discussion in the seismic environment. To the classical and easier *Ray theory* (*Wang & Dahlen*,[38]) which is based on a infinite-frequency approximation is usually set against the *"Scattering theory"* based on Born approximation (*Marquering et al.*, [20]) which, taking in account the finite frequency effect, produce results which are sensitive to the Earth structure off the classical ray-path.
Another interesting factor of influence is the parameterization.
Usually global tomographic images of the Earth can be expressed as linear combination of chosen basis functions. There are several possible choice, known in literature: discrete cells (e.g. *Boschi & Dziewonski*, [9]), splines (e.g., *Boschi & Ekström*, [10]; *Piromallo & Morelli*, [26]), spherical harmonics (e.g., *van Heijst and Woodhouse*,[14] , *Ekström et al.*,[13]) .

With this work we are interested to focus our attention on this last aspect of influence so, using the ray theory and two different types of seismic measurements (those of surface wave phase velocities and those of body wave velocities), we are going to evaluate and testing a new type of parameterization, performed by means of wavelet functions. This parameterization is already used for seismic local study (e.g., *Piromallo et al.*, [25],) but it is rarely applied in global tomography (e.g., *Chiao & Kuo*, [12]).

It is known from the classical Fourier theory, a time series can be expressed as the sum of a, possibly infinite, series of *sines* and *cosines*. This sum is often referred as a *"Fourier expansion"*. The big disadvantage of a Fourier expansion is that it has only frequency resolution and no time resolution. This means that, although we might be able to determine all the frequencies present in a signal, we do not know when they are present. To overcome this problem in the past decades several solutions have been developed which are more or less able to represent a signal in the time and frequency domain at the same time.
The *Wavelet Analysis* (or *Wavelet Transform*) is probably the most recent solution to overcome the shortcomings of Fourier analysis. The fundamental idea behind this innovative analysis is to study signal according to scale. Wavelets, in fact, are mathematical functions that cut up data into different frequency components, and then study each component with resolution matched to its scale, so they are especially useful in the analysis of non stationary process that contains multi-scale features, discontinuities and sharp strike.
Wavelets are essentially used in two ways when they are applied in geophysical process or signals studies:
-as a *basis* for representation or characterization of process;
- as an *integration kernel* for analysis to extract information about the process.
In any analysis or representation the choice of the kernel or basis function determines the nature of information that can be extracted or represented about the process.
With this thesis we attempt to develop a *Multiresolution Wavelet Analysis* in the field of Global Seismic Tomography, with the aim of evaluating and quantifying the power of this approach in terms of earth structure localization and interpretation.

In the first part of work we review the basic properties of wavelets that make them such an attractive and powerful tool for geophysical applications.
We first look at the concept of wavelet function and its properties in the cartesian geometry, more precisely in the time-scale plane (scale being in a way the opposite of frequency, because the term frequency is reserved for the Fourier transform). Then, introducing the concept of *Second Generation Wavelets* (*Sweldens*, [31]; *Schröder & Sweldens*, [29]) , we look at the wavelets and their properties in spherical domain, namely its application on Earth geometry.
The two types of above-mentioned applications of wavelets in geophysical field, are object of study in the second part of this work. At the beginning we use the wavelets as *basis* to represent and resolve the Tomographic Inverse Problem. After a briefly introduction to seismic tomography theory, we assess the power of wavelet analysis in the representation of two different type of *synthetic models*; then we apply it to real data, obtaining surface wave

phase velocity maps and evaluating its abilities by means of comparison with an other type of parametrization (i.e., block parametrization).
For the second type of wavelet application we analyze the ability of Continuous Wavelet Transform in the spectral analysis, starting again with some synthetic tests to evaluate its sensibility and capability and then apply the same analysis to real data (from body wave tomography), to obtain *Local Correlation Maps* between different model at same depth or between different profiles of the same model.

Chapter 1

Wavelets in the Cartesian Spaces

The fundamental idea behind the wavelet approach is to analyze data at different *scales* or *resolutions*. If we look at a signal with a large "window", we would notice gross features. Similarly, if we look at a signal with a small "window", we would notice small features. So the wavelet transform is a method of converting a function (or signal) into another form which either makes certain features of the original signal more amenable to study or enables the original data set to be described more succinctly. The basic idea to perform a wavelet transform is to adopt a wavelet prototype function, called *analyzing wavelet* or *mother wavelet* and then perform the temporal analysis by means of a contracted, high frequency version of this function, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet.

1.1 The Continuous Wavelet Transform

1.1.1 The Wavelet

To perform a wavelet transform we need a set of wavelet functions which, as the name suggests, are localized waveform which must satisfy certain mathematical criteria. In cartesian plane wavelets are generated from a single basic wavelet $\Psi(t)$, the so-called *mother function* (Fig. (1.2)), by scaling and translation (*Valens*, [35]):

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-\tau}{s}\right), \tag{1.1}$$

where s is the scale factor, τ is the translation factor and the factor $s^{-1/2}$ is for energy normalization across the different scales. Equation (1.1) represents a family of wavelet functions: changing the value of s, in fact, has the effect of dilating ($s > 1$) or contracting ($s < 1$) the function $\Psi(t)$ and changing τ has the effect of analyzing function around different points τ (see Fig.(1.1)).

1.1.2 General Proprieties of Wavelets

Wavelets are very powerful tools in the signal processing for their particular proprieties.
Firstly, unlike Fourier basis, sines and cosines, wavelets are functions with compact support

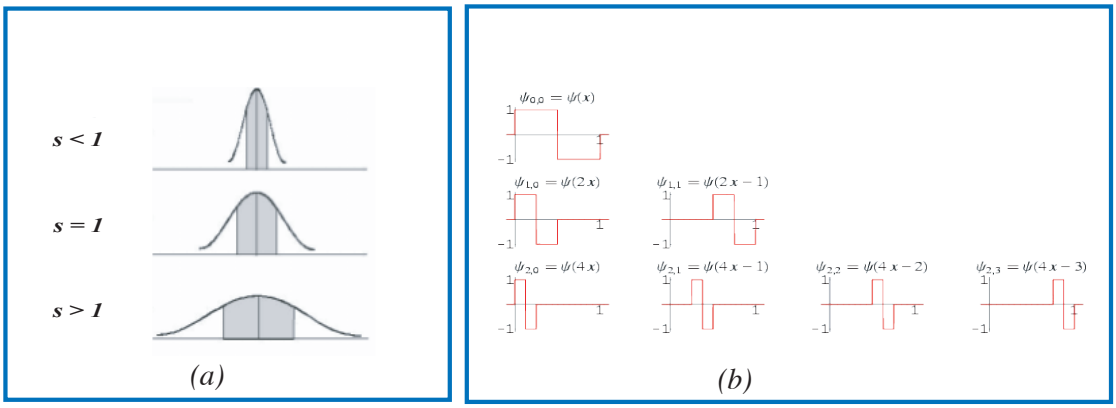


Figure 1.1: *Building of wavelet family - (a) example of dilated ($s > 1$) and contracted ($s < 1$) versions of a mother wavelet; (b) application of building scheme to a particular type of wavelet: the Haar wavelet.*

or sufficiently fast decay to obtain good time - frequency localization. Furthermore, in order to be classified as a wavelet, a function must satisfy certain mathematical criteria (*Addison P.S.*, [1]) .

1. A wavelet must have finite energy :

$$E = \int_{-\infty}^{\infty} |\Psi(t)|^2 dt < +\infty.$$

2. If $\hat{\Psi}(\omega)$ is the *Fourier Transform* of $\Psi(t)$ i.e.:

$$\hat{\Psi}(\omega) = \int_{-\infty}^{\infty} \Psi(t) e^{-i(2\pi\omega)t} dt,$$

then the following condition must hold:

$$C_\Psi = \int_0^{+\infty} \frac{|\hat{\Psi}(\omega)|^2}{|\omega|} d\omega < +\infty \qquad . \tag{1.2}$$

This property is known as *Admissibility Condition* (and gives wavelets their name), while C_Ψ is called *admissibility constant*.

3. Eq.(1.2) implies that the Fourier Transform of $\Psi(t)$ vanishes at the zero frequency:

$$\hat{\Psi}(\omega)\big|_{\omega=0} = 0,$$

this also means that a wavelet is a function with *zero mean* in the time domain:

$$\int_{-\infty}^{\infty} \Psi(t) dt = 0,$$

and therefore it must be *oscillatory*. In other words, $\Psi(t)$ must be a *wave*.

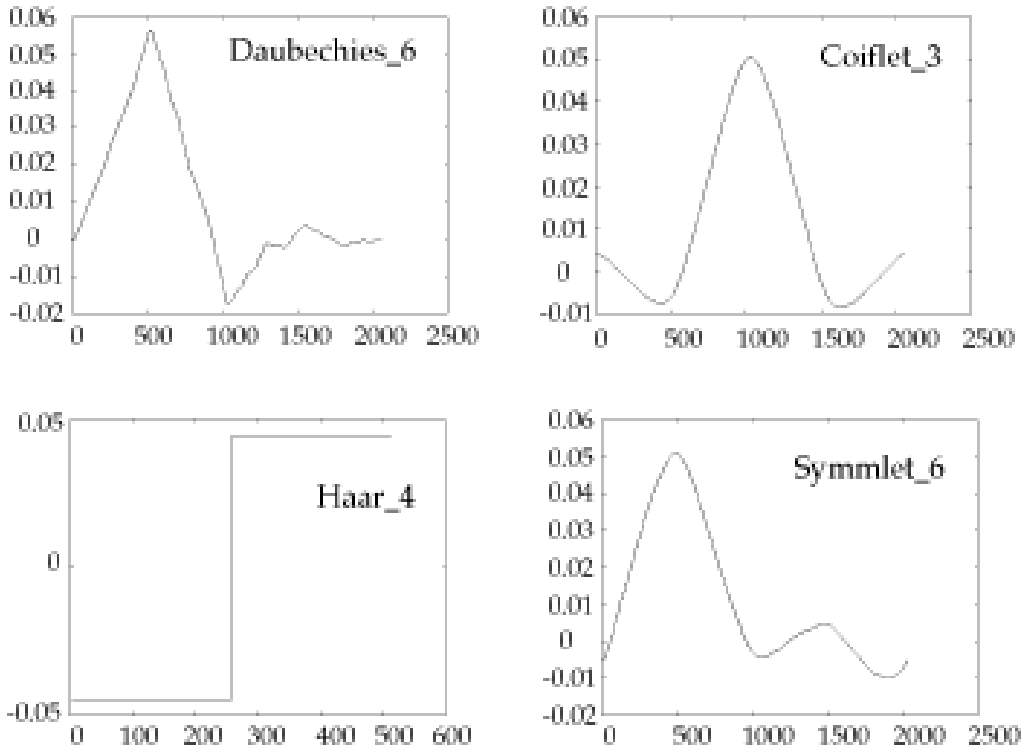


Figure 1.2: *Examples of mother functions (from Amara Graps ,[39]).*

4. Wavelets may also have higher-order moments equal zero [1]:

$$\int_{-\infty}^{\infty} t^k \Psi(t) dt = 0,$$

1.1.3 Continuous Wavelet Transform (*CWT*)

The *continuous wavelet transform* of a signal $f(t) \in \mathcal{L}^2(\Re)$, with respect to a generic set of wavelet functions, is defined as:

$$Wf(s,\tau) = \langle f, \Psi_{s,\tau} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \Psi^* \left(\frac{t-\tau}{s} \right) dt = \int_{-\infty}^{+\infty} f(t) \Psi^*_{s,\tau}(t) dt. \tag{1.3}$$

So the wavelet transform can be rewritten as a convolution product:

$$Wf(s,\tau) = f \star \bar{\Psi}_{s,\tau}(t). \tag{1.4}$$

As its Fourier counterpart, there is an *inverse wavelet transform*, defined as:

$$f(\tau) = \frac{1}{C_\Psi} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} s^{-2} Wf(s,t) \Psi_{t,s}(\tau) dt ds \tag{1.5}$$

Equation (1.5) can be looked at as a way of reconstructing $f(t)$ once its wavelet transform $Wf(s,\tau)$ is known as well as a way to write $f(t)$ as a superposition of wavelets $\Psi_{t,s}(\tau)$. Note that for inverse transform, the original wavelet function is used, rather than its conjugate which is used in the forward transformation. If we limit the integration over a range of s scales rather than all s scales, we perform a basic filtering of the original signal.

1.1.4 The power spectra of a signal

The total energy contained in a signal, $f(t)$, is defined as its integrated squared magnitude:

$$E = \int_{-\infty}^{+\infty} |f(t)|^2 dt = \|f(t)\|^2 \tag{1.6}$$

The relative contribution of the signal energy contained at a specific s scale and τ location is given by the two-dimensional wavelet energy density function:

$$E(s,\tau) = |Wf(s,\tau)|^2 \tag{1.7}$$

A plot of $E(s,\tau)$ is known as *scalogram* (crf. Fig.(1.3)) and is the analogous of a spectrogram, the energy density surface for Fourier transform.
The scalogram can be integrated across s and τ to recover the total energy in the signal using the admissibility constant, C_Ψ:

$$E = \frac{1}{C_\Psi} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} s^{-2} |Wf(s,\tau)|^2 d\tau ds \tag{1.8}$$

[1] Within each family of wavelets there are "wavelet subclasses", distinguished by the number of coefficients. Wavelets are classified within a family most often by the *number of vanishing moments*. This is an extra set of mathematical relationships for coefficients that must be satisfied, and is directly related to the number of coefficients of wavelet expansion.

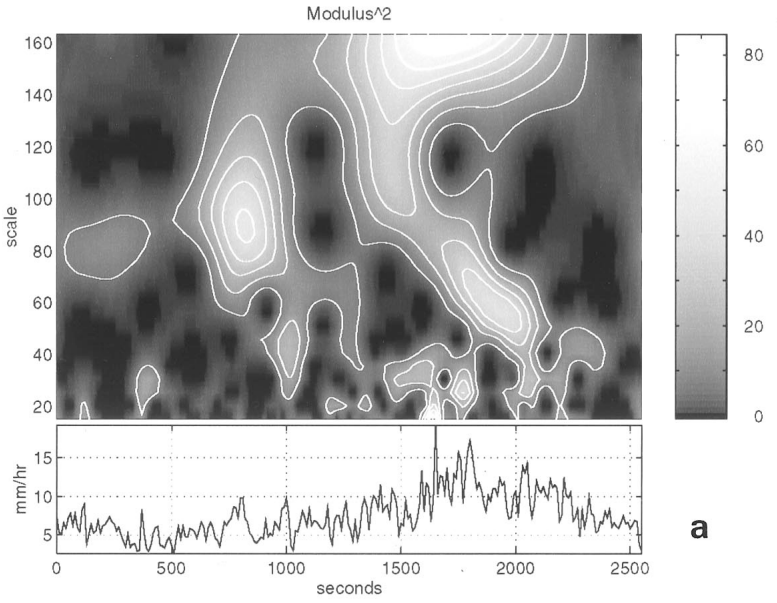


Figure 1.3: *Example of scalogram applied to analysis of temporal rainfall using the Morlet wavelet. The rainfall intensity is shown at the bottom of figure. The scalogram clearly shows the presence of multiscale features and also some embedding of small-scale features within large-scale features [from Kumar, P. & E. Foufoula-Georgiou, [18]]* .

1.2 Wavelet versus Fourier Transforms

• The most important dissimilarity between these two kinds of transforms is that individual wavelet are *localized in space*. Fourier sine and cosine functions are not. This localization feature, along with wavelets localization of frequency, makes many functions and operators, using wavelets, *sparse* when transformed into the wavelet domain. This sparseness results in a number of useful applications such as data compression and removing noise from time series.

• The differences in time-scale resolution between the Fourier analysis and wavelet analysis are also clearly evident if we look at the basis function coverage of time-frequency plane (see Fig.(1.4)). The wavelet transform carries out a decomposition of time-scale plane in cells of different dimensions: in any decomposition the time-scale plane is layered with cells, called *Heisenberg cells*, whose minimum area is determined by the uncertainty principle (which dictates the minimum area but not its shape).
In this way, higher scale/frequencies can be well localized in time, but the uncertainty in frequency localization increases as the frequency increases, which is reflected as taller, thinner cells with increase in frequency.

• It is also remarkable its applicability to analysis of signal contains discontinuities or non-stationary features (like sharp spike) or defined on limited domain, impossible in Fourier domain.

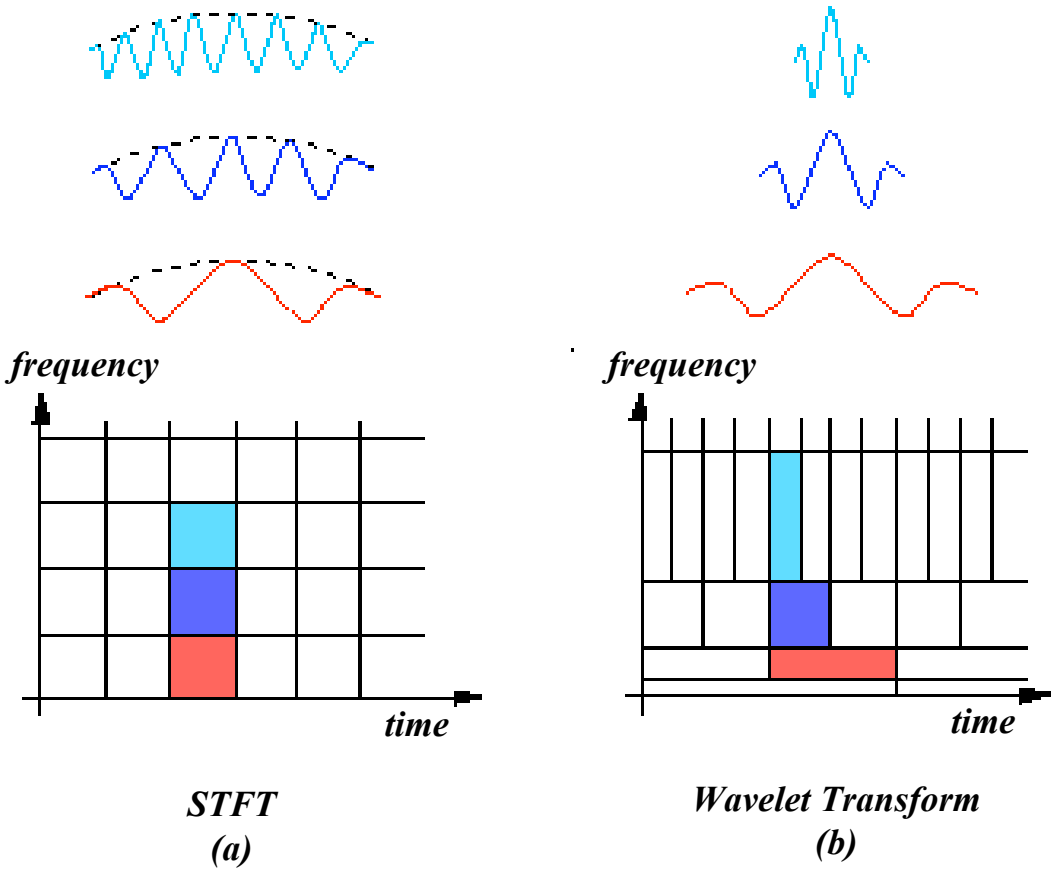


Figure 1.4: *Example of time - frequency plane decomposition using different basis: (a) Fourier basis and (b) Wavelet basis. (from web site [40])*

• Ultimately, it makes possible a *multiresolution analysis* of a signal (see Section 1.4).

1.3 The Discrete Wavelet Transform

The *CWT* maps a one-dimensional signal to a two-dimensional time-scale representation that is highly redundant: the wavelet transform is calculated by continuously shifting a continuously scalable function over a signal and calculating the correlation between the two. It will be clear that these scaled functions will be nowhere near an orthogonal basis. So for practical applications we are some interested to remove this redundancy.
To overcome this problem a *discrete wavelet transform* can be introduced. In fact, we can see that, if certain criteria are met, it is possible to completely reconstruct the original signal using an infinite summations of discrete wavelet coefficients rather than continuous integrals (as required for the CWT). This leads to a fast wavelet transform for the rapid computation of the discrete wavelet transform and its inverse.

1.3.1 Discrete wavelets

The *discrete wavelet transform* (*DWT*) is a particularly useful tool in sampled signals processing when we work with a discrete input signals with finite length.
A natural way to implement the DWT is by means of a discretization of the s scale and τ location parameters. In discretizing the scale and location parameters we can choose $s = s_0^m$, where m is an integer and s_0 is a fixed dilatation step greater than 1, we also choose $\tau = n\tau_0 s_0^m$, where $\tau_0 > 0$ depend upon $\Psi(t)$ and n is an integer.
This kind of discretization of wavelet carry out to the following representation (*Kumar, P. & E. Foufoula-Georgiou*, [18]) :

$$\Psi_{m,n}(t) = \frac{1}{\sqrt{s_0^m}} \Psi\left(\frac{t - n\tau_0 s_0^m}{s_0^m}\right) = s_0^{-m/2} \Psi(s_0^{-m}\tau - n\tau_0). \tag{1.9}$$

The wavelet transform of a continuous signal $f(t)$, using discrete wavelets in the form (1.9), is then:

$$Wf(m,n) = s_0^{-m/2} \int_{-\infty}^{+\infty} f(t)\Psi(s_0^{-m}\tau - n\tau_0)dt, \tag{1.10}$$

which can be expressed as the inner product:

$$Wf(m,n) = \langle f, \Psi_{m,n} \rangle, \tag{1.11}$$

where $Wf(m,n)$ are the discrete wavelet trasform values given on a scale-location grid of index (m,n). They are also known as *wavelet coefficients*.
In the case of a continuous wavelet transform, using (1.3), $Wf(\lambda,t)$ is able to completely characterize $f(t)$. When we use the discrete wavelet $\Psi_{m,n}(t)$ we can also completely characterize $f(t)$ by means of an appropriate choice of s_0 and τ_0. In fact, we can write $f(t)$ as a series expansion if the energy of the resulting wavelet coefficients satisfy a fixed condition.
The *necessary and sufficient condition* for a stable reconstruction is that the above mentioned energy must lie between to positive bounds:

$$A\|f\|^2 \leq \sum_m \sum_n |Wf(m,n)|^2 \leq B\|f\|^2, \tag{1.12}$$

where $\|f\|^2$ denotes the energy of $f(t)$, and $A > 0$, $B < \infty$ are constant independent of function, which can be determined numerically.
When equation (1.12) is satisfied, the family of wavelet functions is a basis and it is referred as a *wavelet frame* with frame bounds A and B.
Given this condition, we can obtain a series expansion for $f(t)$ as:

$$f(t) = \frac{2}{A+B} \sum_m \sum_n Wf(m,n)\Psi_{m,n}(t). \tag{1.13}$$

1.3.2 Orthonormal Wavelet Transform

In general, a frame is not an orthonormal basis (it is a basis only in the limit case $A = B = 1$), so it also provides a redundant representation of the function $f(t)$. Practically, this is equivalent, for example, to representing a vector in the Euclidean plane, using more than two basis

vectors. The ratio A/B is therefore called the *redundancy ratio* or *redundancy factor*. When a frame is redundant the contiguous wavelet coefficients are correlated to each other.
The discrete wavelets can be made orthogonal to their own dilatations and translations by special choices of mother function.
Common choices for discrete wavelet parameters s_0 and τ_0 are 2 and 1, respectively. This particular type of discretization is known as the *dyadic grid arrangement* and it is the simplest and most efficient discretization to obtain an orthogonal wavelet basis (*dilated dyadic wavelets*) (*Addison*, [1])
Substituing $s_0 = 2$ and $\tau_0 = 1$ in equation (1.9), we can write:

$$\Psi_{m,n}(t) = \frac{1}{\sqrt{2^m}}\Psi\left(\frac{t - n2^m}{2^m}\right) = 2^{-m/2}\Psi(2^{-m}\tau - n), \tag{1.14}$$

Discrete dyadic grid wavelets are chosen to be orthonormal. So these wavelets are both orthogonal of each other and normalized to have unit energy:

$$\int_{-\infty}^{+\infty} \Psi_{m,n}(t)\Psi_{m',n'}(t)dt = \begin{cases} 1 & \text{if } m = m' \text{ and } n = n' \\ 0 & \text{otherwise} \end{cases} \tag{1.15}$$

Using the orthonormal dyadic wavelet basis $\Psi_{m,n}(t)$, all square integrable functions $f(t)$ can be approximated, by means of the inverse discrete wavelet transform, by a linear combination as follows:

$$f(t) = \sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} D_{m,n}\Psi_{m,n}(t), \tag{1.16}$$

where the first summation is over scales and at each scale we sum over all translates. The coefficient $D_{m,n}$ measures the contribution of scale s^m at location $n2^m$ to the function.
From (1.12), follows that total energy of function can be obtained as:

$$\int |f(t)|^2 dt = \sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} |D_{m,n}|^2. \tag{1.17}$$

1.3.3 The Haar wavelet

The *Haar wavelet* is the simplest example of an orthonormal wavelet and is given as:

$$\Psi_{m,n}(t) = \begin{cases} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.18}$$

The mother wavelet for the Haar wavelet system, $\Psi(t) = \Psi_{0,0}(t)$ is formed from two dilated unit block pulses sitting next to each other on the time axis, with one of them inverted. From the mother wavelet we can construct the Haar system on a dyadic grid $\Psi_{m,n}$ (see Fig.(1.5)).

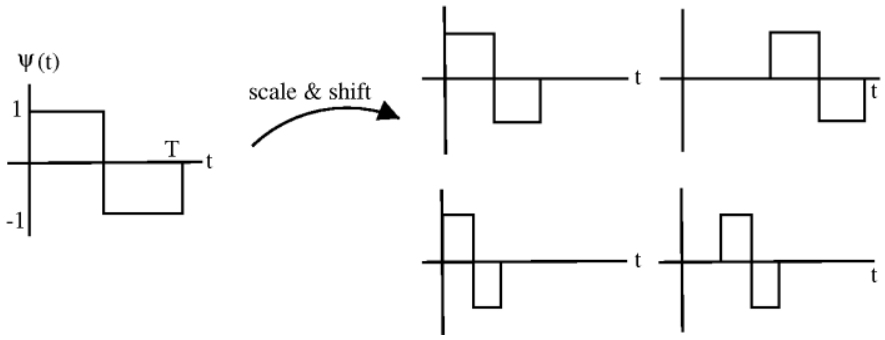


Figure 1.5: *Example of Haar wavelet and its translated and dilated version (from web site [41] .)*

1.3.4 Scaling function

Orthonormal dyadic discrete wavelets are associated with *scaling function* and their dilation equations. The scaling function is associated with the smoothing of the signal and has the same form as the wavelet, given by:

$$\Phi_{m,n}(t) = 2^{-m/2}\Phi(2^{-m}\tau - n). \tag{1.19}$$

They have the property:

$$\int_{-\infty}^{+\infty} \Phi_{0,0}(t) = 1, \tag{1.20}$$

where $\Phi_{0,0}(t) = \Phi(t)$ is sometimes referred as *father wavelet*. The scaling function is orthogonal to translations itself, but not to dilatations of itself. It can be convolved with the signal to produce the so-called *approximation coefficients*:

$$C_{m,n} = \int_{-\infty}^{+\infty} f(t)\Phi_{m,n}(t)dt. \tag{1.21}$$

From the last three equations, we can deduce that the approximation coefficients are simply weighted averages of continuous signal factored by $2^{m/2}$. The approximation coefficients at a specific scale m are, therefore, known also as the *discrete approximation* of the signal at that specific scale.
By means of scaling functions and approximation coefficients we can obtain a *continuous approximation* of the signal at scale m, by summing a sequence of scaling functions at this scale weighting by the approximation coefficients:

$$f_m(t) = \sum_{n=-\infty}^{+\infty} C_{m,n}\Phi_{m,n}(t). \tag{1.22}$$

where $f(t)$ is a smooth, scaling-function-dependent, version of the signal $f(t)$ at the scale index m. From (1.22) we can see that approximation coefficients are used to produce an approximation of the signal which is simply a sequence of scaling functions placed side by side, each factored by their corresponding approximation coefficient.

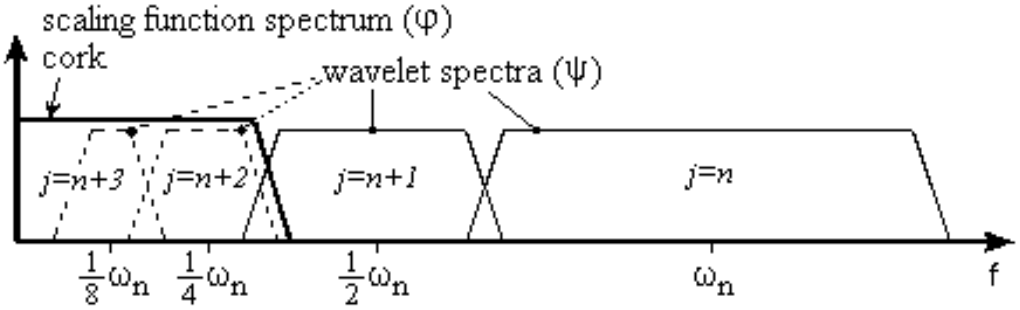


Figure 1.6: *Instead of an infinite set of wavelets it is possible using one scaling function that cover all spectrum for lower frequencies. (from Valens, [35]).*

Using a combined series expansion both the approximation coefficients (1.22) and the wavelet (detail) coefficients (1.16), we can represent a signal $f(t)$ as follows:

$$f(t) = \sum_{n=-\infty}^{+\infty} C_{m_0,n}\Phi_{m_0,n}(t) + \sum_{m=-\infty}^{m_0}\sum_{n=-\infty}^{+\infty} D_{m,n}\Psi_{m,n}(t). \tag{1.23}$$

So we can see from (1.23) that the original continuous signal is expressed as a combination of an approximation of itself, at arbitrary scale index m_0, added to a succession of signal details from scales m_0 down to negative infinity, the *signal detail* at scale m being:

$$d_m(t) = \sum_{n=-\infty}^{+\infty} D_{m,n}\Psi_{m,n}(t). \tag{1.24}$$

The (1.23) also describes the theoretical structure of multiresolution analysis (crf. section (1.4)).

Substantially the wavelet function $\Psi(t)$ (i.e. the mother wavelet) acts as a band-pass filter and scaling it for each level halves its bandwidth. This behavior creates a fundamental problem: in order to cover the entire spectrum, an infinite number of levels would be required. The scaling function $\Phi(t)$ (i.e. the father wavelet) helps us in this purpose because it filters the lowest level of the transform and ensures all the spectrum is covered (crf. Fig. (1.6)).

1.4 Multiresolution Analysis

One of the most important applications of orthonormal wavelet transforms is in *multiresolution analysis*. Multiresolution analysis is concerned with the study of signals or process represented at different resolutions and developing an efficient mechanism for going from one resolution to another. To understand this, imagine that you are looking at a sequence of continuous functions such that the first functions describes only broad features of process (coarse resolution). Each subsequent function progressively adds more detail (higher resolution) such that smaller- and smaller-scale features start appearing as the resolution increases.

1.4.1 Multiresolution representation

Let us assume now we want represent a function $f_m(t)$, approximated by 2^{-m} samples per unit length so that as scale m decreases, the resolution increases. To get the function at the next higher resolution step f_{m-1}, we need to add some detail $f\prime_m(t)$, following the scheme (*Kumar, P. & E. Foufoula-Georgiou*, [18]):

$$f_{m-1} = f_m(t) + f\prime_m(t), \tag{1.25}$$

which is the basic recursive equation of multiresolution theory.
From the previous section we have just seen that this construction require the fundamental use of scaling function; in the wavelet multiresolution framework, $f_m(t)$ is approximated as (crf. (1.22)):

$$f_m(t) = \sum_{n=-\infty}^{\infty} C_{m,n}\Phi_{m,n}(t), \tag{1.26}$$

where $\Phi_{m,n}(t)$ is the above-mentioned scaling function and the coefficients $C_{m,n}$ are:

$$C_{m,n} = \int \Phi_{m,n}(t)f(t)dt. \tag{1.27}$$

The function $\Phi(t)$, substantially, built a discrete approximation of signal at lowest level, behaving as a *sampling function*.
The detail $f\prime_m(t)$ is instead approximated using orthogonal wavelets as (crf. (1.24)):

$$f\prime_m(t) = \sum_{n=-\infty}^{\infty} D_{m,n}\psi_{m,n}(t). \tag{1.28}$$

Fig.(1.7) shows clearly how wavelet multiresolution works: S_λ is the signal at a particular resolution level, $S_{\lambda-1}$ and D_λ are the signal at early resolution level and the detail, respectively, so that $S_{\lambda-1} = S_\lambda + D_\lambda$. Equation (1.25) is called *multiresolution representation*.

1.5 Wavelets in two or more dimensions

The continuous two-dimensional wavelet transform is obtained by treating $\mathbf{t} = (t_1, t_2)$ and $\boldsymbol{\tau} = (\tau_1, \tau_2)$ as vectors (*Addison*, [1]). In this case the 2D wavelet transform is given by :

$$Wf(s,\boldsymbol{\tau}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(\mathbf{t})\frac{1}{s}\Psi^*\left(\frac{\mathbf{t}-\boldsymbol{\tau}}{s}\right)d\mathbf{t} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(\mathbf{t})\Psi^*_{s,\boldsymbol{\tau}}(\mathbf{t})d\mathbf{t}, \tag{1.29}$$

The corresponding inverse wavelet transform is:

$$f(\boldsymbol{\tau}) = \frac{1}{C_\Psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{s=0}^{+\infty} s^{-3}Wf(s,\mathbf{t})\Psi_{\mathbf{t},s}(\boldsymbol{\tau})d\mathbf{t}ds. \tag{1.30}$$

Wavelet transform in higher dimension, D, are also possible simply by extending the length of vectors τ and t to D components. To preserve the energy in the D-dimensional transformation,

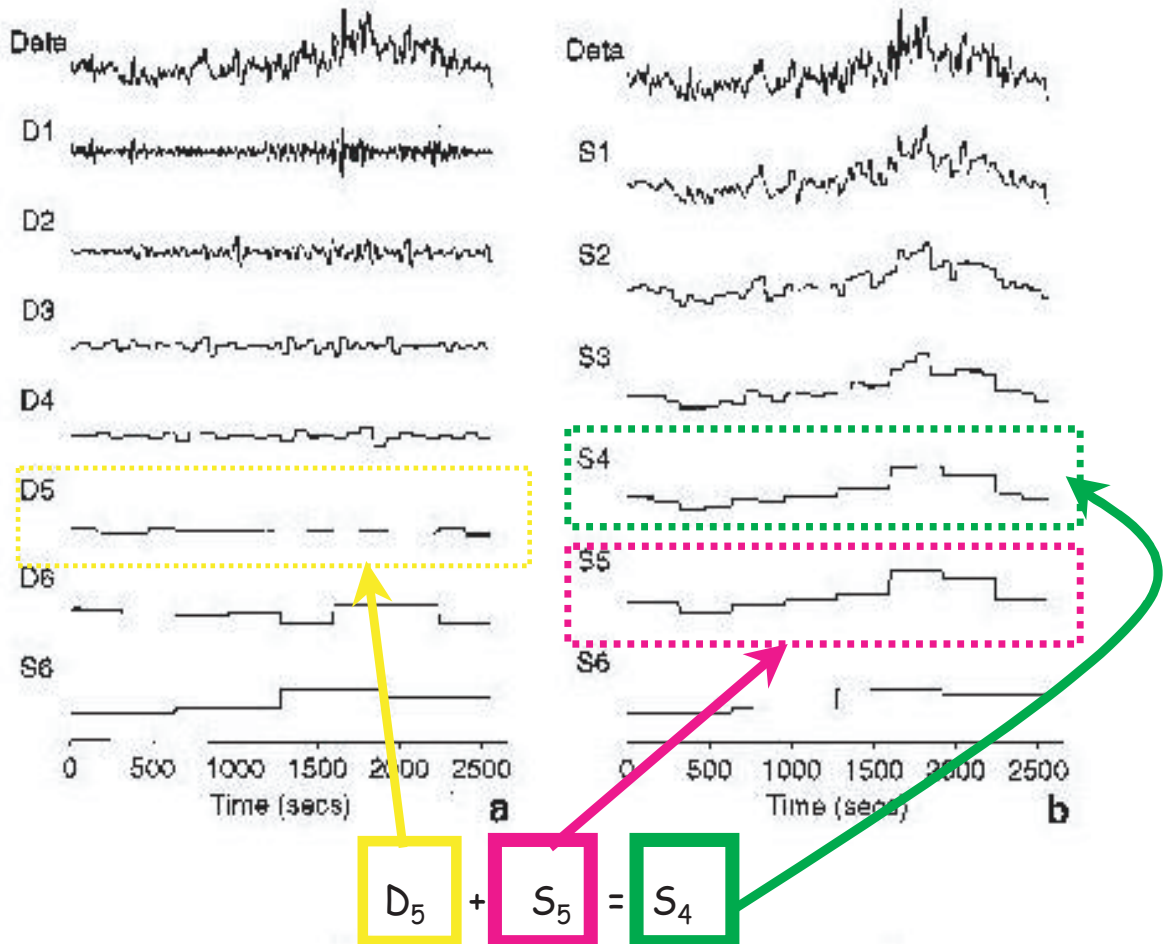


Figure 1.7: Example of Multiresolution Analysis. The general scheme of progression is $S_{\lambda-1} = S_\lambda + D_\lambda$. The example is concerned the use of particular type of wavelet: the Haar wavelet (from Kumar, P. & E. Foufoula-Georgiou, [18]).

the weighting function become $1/s^{D/2}$.
The D-dimensional wavelet transform is then defined as:

$$\Psi_{s,\boldsymbol{\tau}}(t) = \frac{1}{s^{D/2}}\Psi^*\left(\frac{\mathbf{t}-\boldsymbol{\tau}}{s}\right). \tag{1.31}$$

The transform in D dimensions becomes:

$$Wf(s,\boldsymbol{\tau}) = \frac{1}{s}\int_{-\infty}^{+\infty}\Psi^*_{s,\boldsymbol{\tau}}(\mathbf{t})f(\mathbf{t}),d\mathbf{t} \tag{1.32}$$

with inverse:

$$f(\boldsymbol{\tau}) = \frac{1}{C_\Psi}\int_{-\infty}^{+\infty}\int_0^{+\infty} Wf(s,\mathbf{t})\Psi_{\mathbf{t},s}(\boldsymbol{\tau})d\mathbf{t}\frac{ds}{s^{D+1}}. \tag{1.33}$$

As in one-dimension, also in two-dimensional case we can perform a multiresolution representation of a 2D continuous function. Fig.(1.8) shows an example of multiresolution in

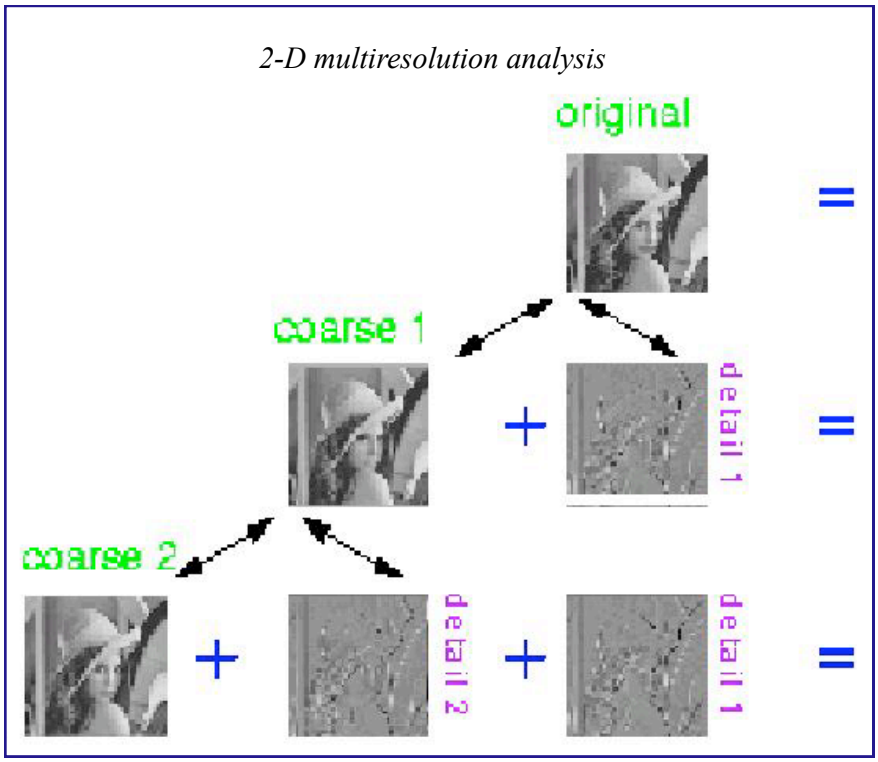


Figure 1.8: Multiresolution analysis applied to imagine processing, that is a 2D signal (from web site [40]).

two-dimensional domain, applied to imaging processing, a successful field of application for wavelet analysis.

Chapter 2

Wavelets on the sphere

Classically wavelet construction has been employed on linear spaces (such as the line $\Re$ and the cartesian plane $\Re^2$) and use the Fourier transform; the underlying reason is that translation and dilation become algebric operation after Fourier transform. This type of construction makes, in fact, impossible the application of wavelet analysis to more general manifolds, where the Fourier transform is not applicable. When we want, for example, to apply multiresolution wavelet analysis to the spherical domain, as we should make in global seismic tomography, an alternative construction is necessary.
The *Lifting Scheme* of *Sweldens* ([31],[29]) does not rely on Fourier transform and can therefore construct wavelet bases over non-translation invariant domains such as bounded region of $\Re^p$ or surfaces.
Following that technique *Bonneau* ([7]) built a new family of *"Nearly-orthogonal spherical Haar wavelets"* that we study in this chapter and we are going to use in the resolution of tomographic inverse problem (see chapter 3).

2.1 The *Second Generation Wavelets (SGW)*

Wavelets are basis functions which represent a given function at multiple level of detail.
Due to their local support in both space and frequency, they are a very useful tool for sparse approximation of functions. Locality in space, as we have already seen, comes from their compact support, while locality in frequency is a direct consequence of their smoothness (decay towards high frequencies) and vanishing moments (decay towards low frequency).
In the classical wavelet setting (i.e., on the real line) wavelets are defined as the dyadic translates and dilates of one particular fixed function.
The basic philosophy behind the new building approach is, therefore, to build wavelets with all desirable properties (localization, fast transform, etc.) adapted to much more general setting.
Firstly we need a construction of wavelets which are adapted to a measure on the surface. While on the real line, the measure is dx, the usual translation invariant Lebesgue measure; on a sphere we denote the usual area by $d\omega$. Adaptive constructions is based on the idea that translation and dilation are not fundamental to obtain the wavelets with the desirable

properties. The fundamental idea that a given function can be expressed as a finite linear combination of basis functions, at finer subdivision level, is retained.
So the main difference between the new family of wavelets that have been built and the old translated and dilated wavelets is that the filter coefficients of the new wavelet basis are not the same throughout, but can change locally to reflect the changing nature of the surface and its measure.

2.1.1 The Lifting Scheme: *Birth* of *SGW*

The *Lifting Scheme* is an innovative method for both designing wavelets and performing the discrete wavelet transform. Actually it is worthwhile to merge these steps and design the wavelet filters while performing the wavelet transform. The technique was introduced by *Sweldens* ([31],[29]) in 1994.
The main difference with classical construction is that it does not rely on the Fourier transform. This way lifting can be used to construct the so-called *"Second generation wavelets"*, wavelets which are not necessarily translates and dilates of one function. The latter we refer to as the *"First generation wavelets"*.
Because of the complexity of the lifting scheme we prefer do not give here the details of the theory pointing to a separated appendix (crf. *Appendix A*). In this section we instead focus our attention to yield a brief overview of the main properties of this technique which make it a so useful tool for wavelet construction.

Advantages in the use of lifting scheme

The use of lifting scheme for wavelet construction has several advantages:

- it allows a faster implementation of the wavelet transform;
- the lifting scheme allows a fully in-place calculation of the wavelet transform. In other words, no auxiliary memory is needed and the original signal can be replaced by its wavelet transform;
- in the classical case, it is not immediately clear that the inverse wavelet transform actually is the inverse operation of the forward transform. With the lifting scheme, the inverse wavelet transform can immediately be found by undoing the operations of the forward transform. In practice, this comes down to simply changing each $+$ into a $-$ and vice versa.

Secondly, the lifting scheme can be used, as we have already suggested, in situations where no Fourier transform is available:

- wavelets on bounded domains (as, for example an interval);
- wavelets on curves and surface (as in our case);
- weighted wavelets;

- wavelets and irregular sampling.

It is obvious that wavelets adapted to these setting can not be formed by translation and dilation. The Fourier transform can thus no longer be used as construction tool. The lifting scheme is, therefore, an alternative approach.

2.2 Spherical Triangular Haar Wavelets

We are interested to a particular family of second generation wavelets: the triangular Haar wavelets.

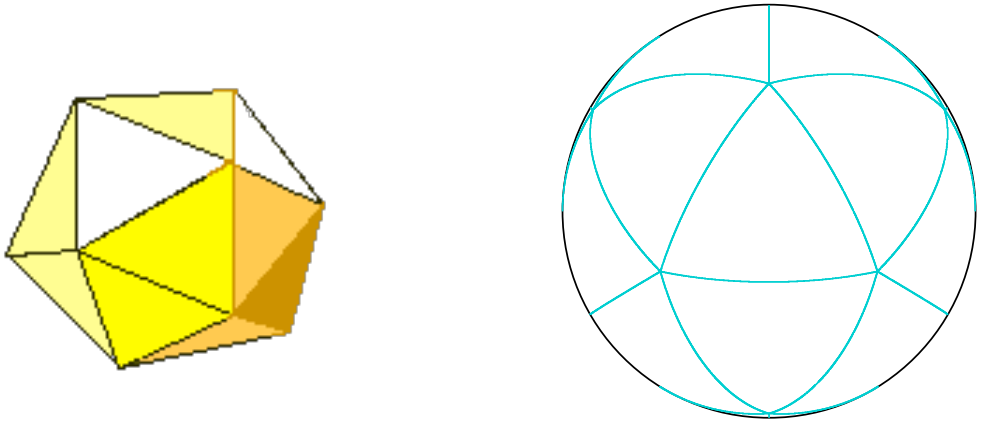


Figure 2.1: *Projecting the faces of an icosahedron (on the left) on spherical surface it is generated a spherical triangular grid of twenty equilateral triangles (on the right)*

The *Spherical Triangular Haar Wavelets* (*STHW*) are defined over a nested triangular grids. These grids are the result of an iterative subdivision of base triangular mesh obtained projecting the faces of an icosahedron[1] (crf. left side of (2.1)) on the sphere.

2.2.1 Building of nested triangular parametric grid

The projection of the faces of icosahedron on spherical surface generates a spherical triangular grid of twenty equilateral spherical triangles. This is the base starting grid indicated as level 0 ($k = 0$) (crf. right side of (2.1)).
Starting from base triangular grid we divide recursively each triangle into four sub-triangles by inserting new vertices at the midpoint of each edge as schematically in Fig. (2.2) shown . Iterative subdivision produce a nested triangular grid in which, the base patches are triangles of “almost equal-area”.
Performing the subdivision process we obtain a sequence of nested grid; Fig.(2.3) shows the tessellation of the spherical surface with the spherical triangular grid up to level 5.

[1]The Icosahedron is one of the five Platonic solids. It is a convex regular polyhedron composed of twenty triangular faces, with five meeting at each of the twelve vertices. It has 30 edges and 12 vertices.

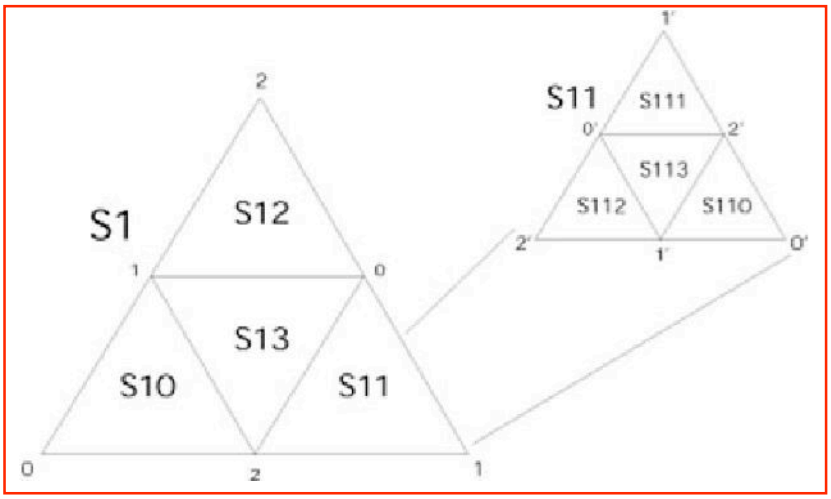


Figure 2.2: *Subdivision scheme.*

2.2.2 Local reconstruction/decomposition with *STHW*

The *STHW* are defined by their values on the four sub-triangles dividing their support.
Let T^k be a triangle at the subdivision depth k (with $k = 0$ associated to the base mesh). Let $T_0^{k+1}, T_1^{k+1}, T_2^{k+1}, T_3^{k+1}$ be the four sub-triangles of T^k. For each triangle at depth k there exist three wavelet functions Ψ_1^k, Ψ_2^k and Ψ_4^k whose support matches exactly with that triangle and a constant function Φ^k whose value is one on T^k. Let r_{ij}^k denote the values of Ψ_j^k on i-th sub-triangle (crf. Fig.(2.4) for exact notations).
The fundamental property of triangular Haar wavelets is that every piecewise constant function on the four sub-triangles can be expressed as a linear combination of the constant function and the three wavelets.
The relation that gives the values x_i^{k+1} of a function at higher resolution level starting by wavelet coefficients y_i^k is known as *local reconstruction*:

$$\begin{pmatrix} x_0^{k+1} \\ x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \end{pmatrix} = \begin{pmatrix} 1 & r_{01}^k & r_{02}^k & r_{03}^k \\ 1 & r_{11}^k & r_{12}^k & r_{13}^k \\ 1 & r_{21}^k & r_{22}^k & r_{23}^k \\ 1 & r_{31}^k & r_{32}^k & r_{33}^k \end{pmatrix} \begin{pmatrix} x^k \\ y_1^k \\ y_2^k \\ y_3^k \end{pmatrix} \tag{2.1}$$

The Fig. (2.5) shows schematically the local reconstruction. The 4×4 matrix R in (2.1) is the so-called *matrix reconstruction*.
The inverse process, referred as *local decomposition*, is obviously:

$$\begin{pmatrix} x^k \\ y_1^k \\ y_2^k \\ y_3^k \end{pmatrix} = \begin{pmatrix} R^{-1} \end{pmatrix} \begin{pmatrix} x_0^{k+1} \\ x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \end{pmatrix} \tag{2.2}$$

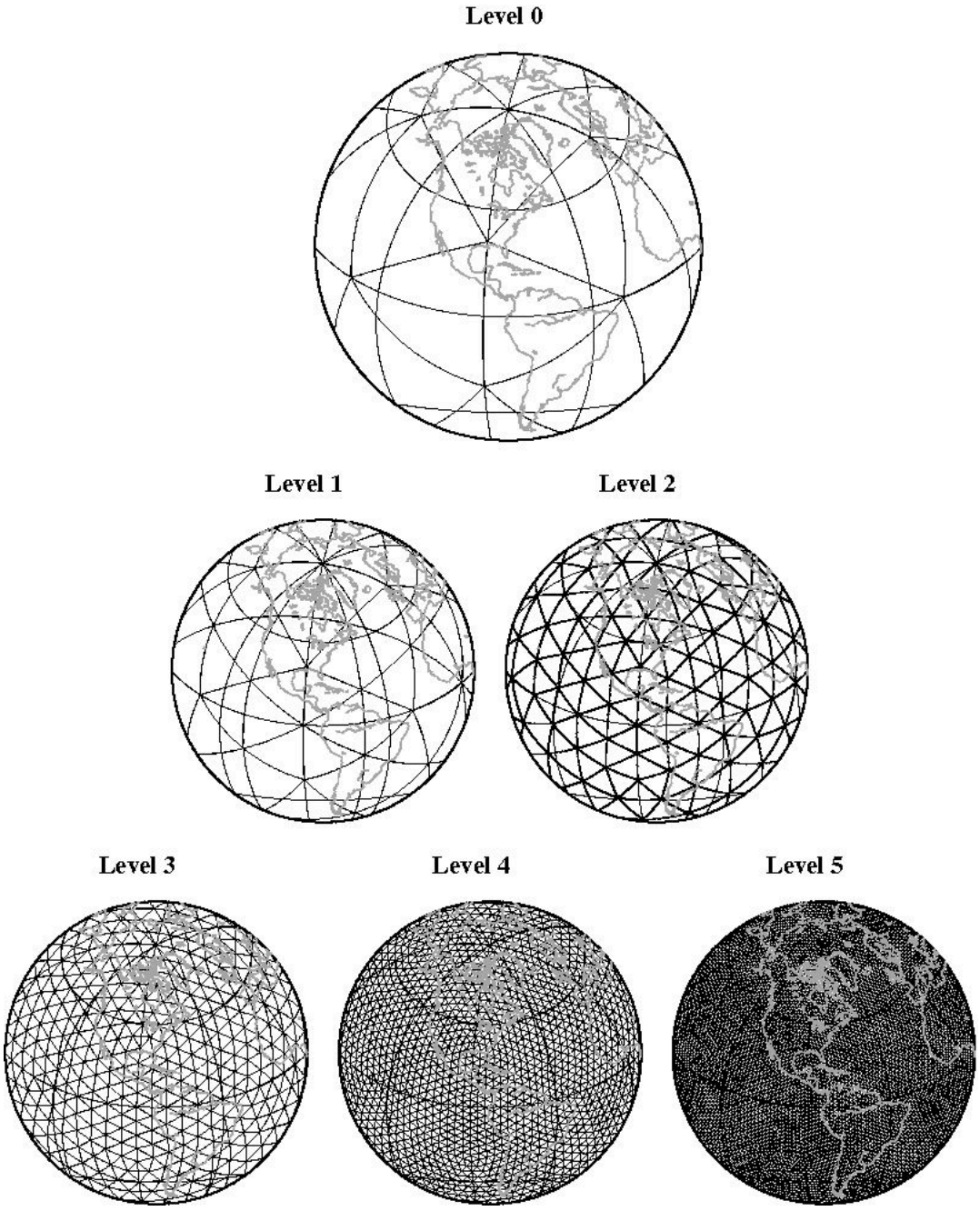


Figure 2.3: *Iterative subdivision of the spherical triangular mesh up to level 5.*

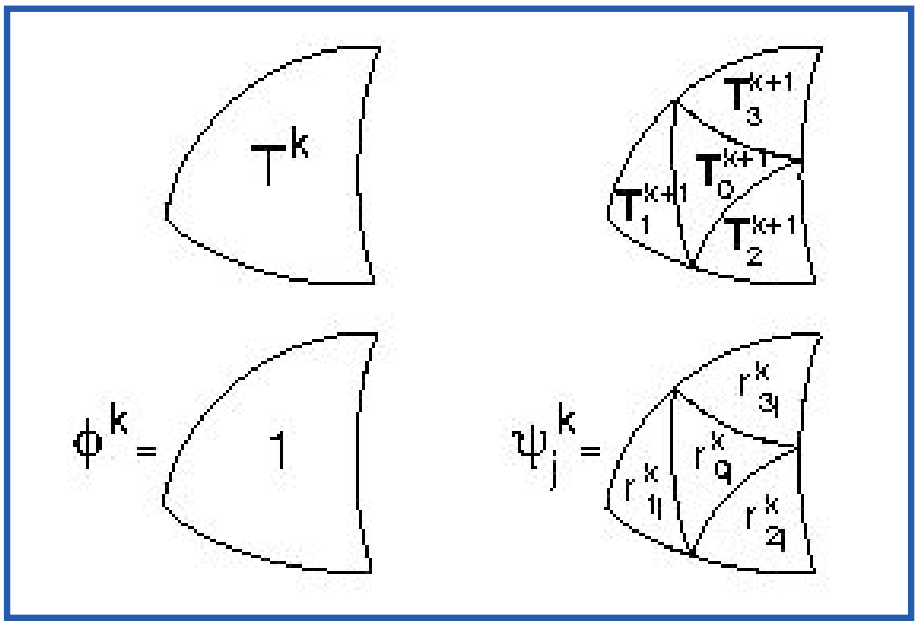


Figure 2.4: *Basis function at level k (from Bonneau G.P, [7]).*

2.2.3 Global reconstruction/decomposition with *STHW*

Given the values (x_i^k) of an input data set at the maximal depth level K, the decomposition in the wavelet basis can be obtained following an iterative pseudo-code:

- for $k = K - 1$ to 0;
- for all triangles T^k at level k;
- perform local decomposition (2.2);
- store $(x^k, y_1^k, y_2^k, y_3^k)$ instead of $(x_0^{k+1}, x_1^{k+1}, x_2^{k+1}, x_3^{k+1})$;

This decomposition process outputs one value on each triangle of the base mesh, and a set of three coefficients for each triangle in the nested grid.
The global reconstruction is the inverse pseudo-code:

- for 0 to $k = K - 1$;
- for all triangles T^k at level k;
- perform local reconstruction (2.1);
- store $(x_0^{k+1}, x_1^{k+1}, x_2^{k+1}, x_3^{k+1})$ instead of $(x^k, y_1^k, y_2^k, y_3^k)$

2.3 Orthogonality on the sphere

It is important to remark that when we work on the sphere we lose the possibility to work with orthonormal basis wavelet functions as we usually can make in cartesian plane.
At the beginning, for completeness, we review some fundamental concepts; for this purpose we will use the notation introduced in (2.2.2) and (2.2.3).

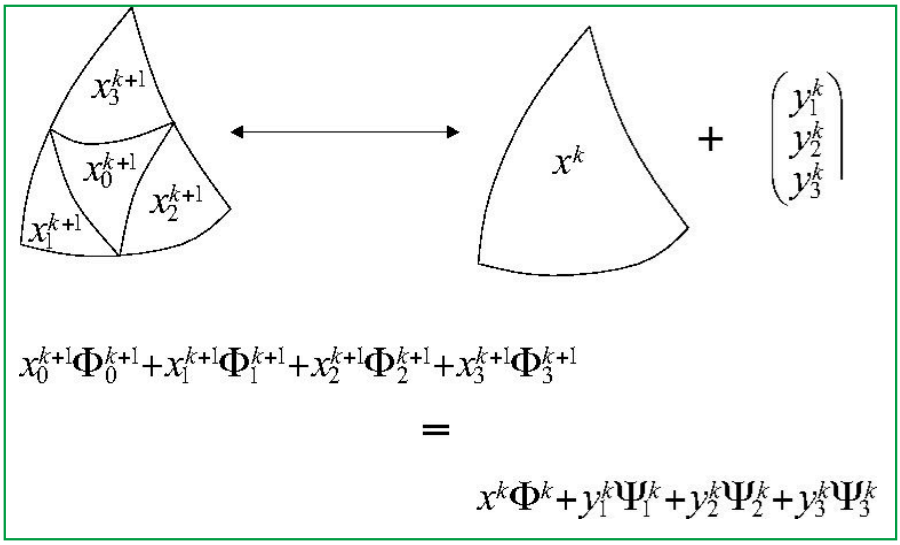


Figure 2.5: *Local reconstruction scheme.*

• A wavelet basis is defined as *semi-orthogonal* if the inner product of the wavelets with the constant function, as defined in (2.2.2), vanishes:

$$\int \Psi_1^k \Phi^k = \int \Psi_2^k \Phi^k = \int \Psi_3^k \Phi^k = 0 \tag{2.3}$$

• A wavelet basis is defined as *orthogonal* if, in addition to the above mentioned property, any two distinct wavelet functions have a vanishing inner product:

$$\int \Psi_1^k \Psi_2^k = \int \Psi_2^k \Psi_3^k = \int \Psi_1^k \Psi_3^k = 0 \tag{2.4}$$

2.3.1 Nearly Orthogonal Spherical Haar Wavelets

In this work, in particular in Chapter 3, we use new spherical Haar wavelets, introduced by *Bonneau* ([7]), that are "nearly-orthogonal".
The concept of nearly orthogonality for the spherical wavelets was introduced by Nielson *et al.* ([21]) and it is based on the observation that the values of spherical wavelets are functions of triangular areas that become uniform at the subdivision depth increases. In the limit case, i.e. when the four sub-triangle areas are equal, the wavelet basis become orthogonal: this is, in practice, the *nearly-orthogonality*.
The starting point to build those wavelets is the observation, from previously introduced reconstruction matrices ([29],[21]) that the wavelet function values are polynomials in the sub-triangles areas $\alpha_0, \alpha_1, \alpha_2, \alpha_3$, with degree 2.

Since we want to find both simple and nearly orthogonal matrices, we start with a 4×3 submatrix containing arbitrary polynomials of degree 1 in $\alpha_0, \alpha_1, \alpha_2, \alpha_3$, then we introduce step by step the required properties.

Without any conditions on polynomials coefficients, the number of free parameters are 48 (= $4 \times 3 \times 4$).

• The first condition that we impose at our sub-matrix is the *symmetry*. This property states that, when permuting the indices (1,2,3) in the right and left-hand side of the local decomposition/reconstruction relations, the equality should be preserved.
Imposing this property we reduce at 10 the numbers of free parameters.

• The second condition is the *semi-orthogonality*. This condition reduce further the number of free parameters to three. So the reconstruction matrix become:

$$\begin{pmatrix} 1 & a\alpha_1 + b\alpha_2 + b\alpha_3 & b\alpha_1 + a\alpha_2 + b\alpha_3 & b\alpha_1 + b\alpha_2 + a\alpha_3 \\ 1 & -a\alpha_0 + c\alpha_2 + c\alpha_3 & -b\alpha_0 - c\alpha_2 & -b\alpha_0 - c\alpha_3 \\ 1 & -b\alpha_0 - c\alpha_1 & -a\alpha_0 + c\alpha_1 + c\alpha_3 & -b\alpha_0 - c\alpha_3 \\ 1 & -b\alpha_0 - c\alpha_1 & -b\alpha_0 - c\alpha_2 & -a\alpha_0 + c\alpha_1 + c\alpha_2 \end{pmatrix} \tag{2.5}$$

In this case a, b and c are the only free parameters.

• The last condition that we impose is the *nearly-orthogonality*. This implies that in the limit of uniform areas, the wavelets should be orthogonal and so the inner product of function should be zero:

$$\int \Psi_1 \Psi_2 = \int \Psi_2 \Psi_3 = \int \Psi_1 \Psi_3 = 0$$

Imposing the vanishing of the inner product involves:

$$\begin{cases} c = a + b \\ c = -\frac{a+5b}{3} \end{cases}$$

At the end we have two distinct families of simmetris, semi-orthogonal and nearly orthogonal triangular wavelets in which, the only two free parameters are a and b.
If we keep in mind that the wavelet functions must be still normalized, we can reduce the free parameters at one.
In particular, choosing to work with the second families of wavelets, we impose, without lose in generality, $a = 1$; so from the normalization condition, we have $b = -0.067$. The reconstruction matrix is finally:

$$\begin{pmatrix} 1 & a\alpha_1 + b\alpha_2 + b\alpha_3 & b\alpha_1 + a\alpha_2 + b\alpha_3 & b\alpha_1 + b\alpha_2 + a\alpha_3 \\ 1 & -a\alpha_0 - \frac{a+5b}{3}(\alpha_2 + \alpha_3) & -b\alpha_0 + \frac{a+5b}{3}\alpha_2 & -b\alpha_0 + \frac{a+5b}{3}\alpha_3 \\ 1 & -b\alpha_0 + \frac{a+5b}{3}\alpha_1 & -a\alpha_0 - \frac{a+5b}{3}(\alpha_1 + \alpha_3) & -b\alpha_0 + \frac{a+5b}{3}\alpha_3 \\ 1 & -b\alpha_0 + \frac{a+5b}{3}\alpha_1 & -b\alpha_0 + \frac{a+5b}{3}\alpha_2 & -a\alpha_0 - \frac{a+5b}{3}(\alpha_1 + \alpha_2) \end{pmatrix} \tag{2.6}$$

Chapter 3

Wavelet Analysis in 2-D Global Seismic Tomography: The Inverse Problem

Seismic tomography is the most powerful tool for determination of 3D structure of deep Earth's interiors. Tomographic models obtained at the global and regional scales are an underlying tool for determination of geodynamical state of the Earth, showing evident correlation with other geophysical and geological characteristics.
Global tomographic images of the Earth can be written as a linear combinations of basis functions from a specifically chosen set, defining the *"model parameterization"*.
A number of different parameterizations are commonly seen in literature: seismic velocities in the Earth have been expressed, for example, as combinations of spherical harmonics (e.g., *van Heijst and Woodhouse*,[14] , *Ekström et al.*,[13]) or by means of the simpler characteristic functions of discrete cells (e.g., *Boschi and Dziewonski*, [9]). There are several motivations to prefer a parameterization respect to another one: one might be chosen because it simplifies calculations, or because it is efficient in representing lateral variations over a broad spectrum of scale lengths. Different parameterizations however tend to sacrifice resolution at different scale lengths.
For this reason in the last years some multiresolution parameterizations (i.e., parameterizations designed to optimize resolution over broad range of scalelengths) have been also proposed in the global seismological field (e.g., *Sambridge*,[28]).
In fields other than seismology, in the last decades, it became very popular a "*multiscale*" parameterization performed by means of wavelet basis functions, characterized by compact support both in the spatial and frequency domains.

In this chapter we want to apply this multiscale wavelet analysis to solve the surface waves global tomographic inverse problem. In this way we are interested both to obtain better tomographic images of Earth and quantify the potentialities of wavelets in this particular geophysical field.
With this purpose, we firstly introduce some general notions about seismic tomographic the-

ory and informations on the data sets; then we face the tomographic inverse problem resolution using the spherical triangular Haar wavelets, introduced in the previous chapter.
We study and evaluate the wavelet approach in the tomographic field carrying out some synthetic tests. Then we solve the surface wave tomographic problem using real data and make a comparison between wavelet and block parameterization, to quantify the capabilities of first one.

3.1 Surface Wave Tomography

We are interested to study the tomographic problem for seismic surface waves, so we directly introduce the theory concerning this problem.

3.1.1 Phase velocity measurements

Surface waves are generally the strongest arrivals recorded at teleseismic distances and they provide some of the best constraints on the Earth's shallow structure. The seismic surface waves are dispersive: their speed of propagation is a function of their frequency, or, in the other words, individual harmonic components of the surface wave seismogram propagate over the globe at different speed (generally increase with increasing depth).
For any given receiver – source couple, we can isolate, from the seismogram, each harmonic component and measure its average speed (called *phase velocity*). The plot of average phase velocity against frequency is called *dispersion curve*. From a large and uniform set of measured dispersion curves is possible to determinate local phase velocity heterogeneities as function of longitude and latitude (crf. *Stein & Wysession*, [30]). This is substantially the fundamental idea of *Seismic Surface Wave Tomography* .

Given a spherically symmetric reference Earth model, laterally homogeneous and assumed, from the *Ray Theory* (*JWKB theory*) that wavepaths are great circle on the sphere, we can assess a theoretical Rayleigh or Love seismogram for any source-receiver geometry and write it as function of frequency ω, in the form (*Ekström*, [13]):

$$u^0(\omega) = A^0(\omega) \exp[i\Phi^0(\omega)], \tag{3.1}$$

where $A^0(\omega)$ is the amplitude and Φ^0 is the propagation phase for reference surface wave. It is related to phase velocity c^0 in the spherical reference Earth model, by the relation:

$$\Phi^0_P(\omega) = \frac{\omega L}{c^0} \tag{3.2}$$

where L is the propagation path length measured along the great circle.
Using a perturbation approach, we can express the observed surface wave $u(\omega)$ as a perturbation with respect to the reference seismogram (3.1):

$$u(\omega) = [A^0(\omega) + \delta A(\omega)] \exp i[\Phi^0(\omega) + \delta\Phi(\omega)] \tag{3.3}$$

Attributing $\delta\Phi$ to a perturbation of the propagation phase, caused by the real Earth structure, more complex than reference one, we have:

$$\Phi_P = \Phi_P^0 + \delta\Phi = \frac{\omega L}{c^0 + \bar{\delta c}} \tag{3.4}$$

where $\bar{\delta c}$ is the apparent average phase velocity perturbation, calculated for the distance L along the great circle (i.e., our measurement).
The estimation of phase velocity perturbation $\delta c(\omega)$ and amplitude $A(\omega)$ of the fundamental mode of surface wave, is generally performed by means of a complex measurement techniques based on iterative algorithms (crf. *Ekstrom & al.*, [13]) which finally give us a data set of path - average phase velocity as function of period, the above-mentioned *dispersion curve*.

3.1.2 Forward Problem

In surface wave ray theory, the phase delay and the ray path of a surface wave for a certain mode branch (we only deal with the fundamental mode) at given period are completely determined by the phase velocity distribution of the medium. The average phase velocity along a path i between a source and receiver is given by:

$$\frac{\delta c_i(\omega)}{c_0} = \frac{1}{L_i}\int_{path_i} \frac{\delta c(\omega,\theta,\phi)}{c_0} ds, \tag{3.5}$$

where $\delta c_i/c_0$ is the phase velocity perturbation with respect to the phase velocity c_0 in the reference model, L_i is the path length and $\delta c_i(\theta,\phi)$ represents the phase velocity distribution. The integration in (3.5) is along the ray path, which itself is dependent on the phase velocity distribution $\delta c_i(\theta,\phi)$.
In most global surface wave studies the great-circle approximation is used; so the integral between source and riceiver is not carried out over the actual ray path, which is not known a priori, but instead is taken over the corresponding great circle.
Equation (3.5) is the linearized solution to the *Forward Problem*: it allows us to determine perturbation in phase velocity as function of the perturbation to the velocity distribution in the medium.

3.1.3 Inverse Problem

Equations (3.5) is usually used to obtain information on the 2-D and 3-D structure of Earth from surface wave measurements.
Using a collection of dispersion measurements (crf.(3.4)) and the equation (3.5), we can evaluate the unknown function $\delta c[r(s),\theta(s),\phi(s)]$. Chosen a set of n basis functions $(f_1, f2, \cdots, fn)$, we re-write the phase velocity perturbation δc as a linear combination of these function, so that:

$$\delta c(r,\theta,\phi) = \sum_{l=1}^{n} x_l f_l(r,\theta,\phi), \tag{3.6}$$

where x_l are unknown coefficients of linear combination. They do not depend from spherical coordinates so we obtain, substituting in (3.5), we obtain, for the $i-$th observation of phase velocity anomaly:

$$\frac{\delta c_i}{c_0} = \sum_{l=1}^{n} x_l \int_{path_i} f_l[r(s), \theta(s), \phi(s)]ds \quad (i = 1, \cdots, m). \tag{3.7}$$

where m is the total number of available observations and n is the number of parameters.
If we define the $m \times n$ matrix A:

$$A_{il} = \frac{1}{L_i} \int_{path_i} f_l[r(s), \theta(s), \phi(s)]ds, \tag{3.8}$$

we definitively obtain the tomographic relationship in the matrix notation:

$$A \cdot x = d, \tag{3.9}$$

where x is the n-vector of the unknown coefficients x_l, d is the m-vector of the phase velocity anomalies (measurements) while the entries of A are numerically computable by means of (3.8). The linear system (3.9) is the *Tomographic Inverse Problem*.
It is possible to solve separately the linear inverse problem (3.5), for each frequency, obtaining two-dimensional phase velocity maps of Earth at different depths. This is the simplest way to use surface wave phase anomaly data as constraints to the Earth's internal properties.

3.1.4 Solution of the Tomographic Inverse Problem

The relation (3.9) is a linear system, often, overdetermined, meaning that the number m of data are higher than the number n of parameters and are affected by error. So we usually are not able to solve it exactly by a direct inversion. The standard produce is to find its *least square solution*.
The *least squares techniques* arise when all the 'input' probability densities (that of data and that of model) are assumed to be Gaussian. It consists of searching of the model m which minimize the *misfit function*, defined as (*Tarantola*, [34]):

$$\begin{aligned} 2S(\mathbf{m}) &= \|\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}\|_D^2 + \|\mathbf{m} - \mathbf{m}_{pror}\|_M^2 = \\ &= (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})^T C_D^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}) + (\mathbf{m} - \mathbf{m}_{pror})^T C_M^{-1} (\mathbf{m} - \mathbf{m}_{pror}), \end{aligned} \tag{3.10}$$

where $\mathbf{m}_{prior}$ is our a priori information on unknown model whose covariance matrix is C_M and C_D is the covariance matrix for data distribution probability (supposed gaussian).
When the problem is linear, as in our case, we have $\mathbf{g}(\mathbf{m}) = \mathbf{A}\mathbf{m}$ and the least square solution can be written as (*Tarantola*, [34]):

$$m_{LS} = m_{prior} + (A^T C_D^{-1} A + C_M)^{-1} A^T C_D^{-1} (d_{obs} - A m_{prior}). \tag{3.11}$$

We furthermore suppose that $m_{prior} = 0$ (so that $C_M = 0$) and that $C_D = \mathrm{I}$, so, using the notation in (3.9) the final expression of least square solution is:

$$x_{LS} = (A^T A)^{-1} A^T d_{obs}, \tag{3.12}$$

where $(A^T A)^{-1}$ is called *generalized inverse* of A.

3.1.5 Regularization

In order to obtain a stable solution from the ill-conditioned tomographic problem, is often necessary to make use of some a priori informations, generally based on our physic knowledge of the problem. So, the solution of inverse problem is forced to satisfy certain conditions by simply including in the linear system (3.9) a number of additional equations that represent them algebraically.
This procedure is usually called *regularization* or *damping*.
The constraints that we want impose to our model can be describe by a relation:

$$D \cdot x = c. \tag{3.13}$$

Including this new constrain, the system (3.9) becomes:

$$\begin{bmatrix} A \\ \lambda D \end{bmatrix} \cdot x = \begin{bmatrix} d \\ \lambda c \end{bmatrix}. \tag{3.14}$$

Like for (3.9), the (3.14) does not have an exact solution and we again solve it in least square sense. The arbitrary parameter λ has the rule to offset the solution which minimizes $\|Dx - c\|$ and the least squares solution to the original system.
The least square solution for the "damped" system (3.14) is:

$$x_{LS} = [A^T A + \lambda^2 D^T D]^{-1} \cdot [A^T d_{obs} + \lambda^2 D^T c]. \tag{3.15}$$

3.2 Wavelet Tomography

The starting point is the equation (3.5) which relate mean phase velocity $\delta c(\omega)$ along ray-path with unknown local phase velocity $\delta c(\theta, \phi)$.
Now we want to parameterize our images of phase velocity in terms of nearly orthogonal triangular Haar wavelets so we need to expand the unknown function $\delta c(\theta, \phi)$ over a set of spherical triangular Haar wavelets as showed in (3.6) (in our case we have $f_i(\theta, \phi) = \Psi_i^k(\theta, \phi)$, where k is the level of resolution (crf.(2.1)).
Because of the chosen parameterization we can discretize equation (3.5), so we re-write it as a summation:

$$\delta c_i(\omega) = \frac{1}{L_i} \sum_{l=1}^{N} \delta c_l \Delta s_l, \tag{3.16}$$

where N is the number of segments in which the ray path is divided by triangular grid at level $k+1$, Δs_l is the length of one segment and $\delta c_l = \delta c(\theta, \phi)$ is the "local" value of unknown function or, in other words, the value of phase velocity on the single triangle at level $k+1$.
Applying local reconstruction (2.1) to piecewise constant function δc_l we write, for the l-*th* value of data $\delta c(\theta, \phi)$, over the generic triangle T_q^k at level $k+1$:

$$\delta c_l^{(k+1)} = \sum_{q=1}^{4} x_q^{(k)} \Psi_{ql}^{(k)}, \tag{3.17}$$

$\Psi_{ql}^{(k)}$ being the values of qth wavelet basis function on the kth triangle and $x_q^{(k)}$ the wavelet coefficients. Substituting (3.17) in (3.16) we definitely obtain:

$$\delta c_i(\omega) = \frac{1}{L_i} \sum_{l=1}^{N} \sum_{q=1}^{4} x_q^{(k)} \Psi_{ql}^{(k)} \Delta s_l, \tag{3.18}$$

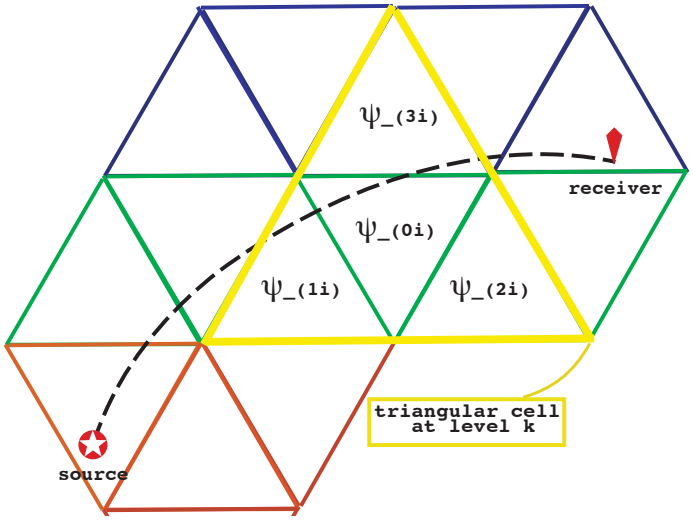


Figure 3.1: *Schematic representation of ray path over spherical triangular grid at level k. The yellow triangle shows the basic support of wavelet function for the fixed level. $\Psi_{(ql)}$ indicate the value of basis function on sub-triangle (q and l being wavelet and sub-triangle index, respectively).*

Clearly, (3.18) is true for a fixed resolution depth k since, changing resolution level we change the area of triangular pixel and consequently the length of ray portions in (3.18) and the values of basis functions $\Psi_{ql}^{(k)}$.

3.2.1 Multiresolution Tomographic Matrix

Let us see how the matrix A in (3.9) is calculated from (3.18) to perform a multiresolution analysis.
By comparison between (3.18) and (3.8) we see that, for any triangular cell at the subdivision level $(k+1)$, the elements of matrix are obtained as the product of value of basis function on its support and the section of ray path which cross the same support (crf. (Fig. (3.1))), normalized with whole length path L_i:

$$\underbrace{\delta c_i(\omega)}_{\text{data}} = \sum_{l=1}^{N} \underbrace{\left(\sum_{q=1}^{4} \frac{1}{L_i} \Psi_{ql}^{(k)} \Delta s_l \right)}_{\text{matrix elements}} \times \underbrace{x_q^{(k)}}_{\text{unknown coefficients}}, \tag{3.19}$$

Starting from the fundamental level we apply the local reconstruction (2.1) (or equally (3.17)) and evaluate the pointed out element in (3.19) for all triangles at level 0.

This first step, substantially, gives us the tools to build (following the forward problem) the function at the higher resolution level (1). If we have the possibility to evaluate model at level 1, i.e. we can assess the value of $x^{(1)}$ in local reconstruction (2.1), in the next step, we do not need to "invert" to obtain the 4 coefficients for a local reconstruction at level 2 but we have only need of the coefficients associated to the three wavelet whose support matches the triangle at level 2 and which value, conveniently weighted by the coefficients, give us the details to be add to value of function at level 1 to obtain its value at level 2.

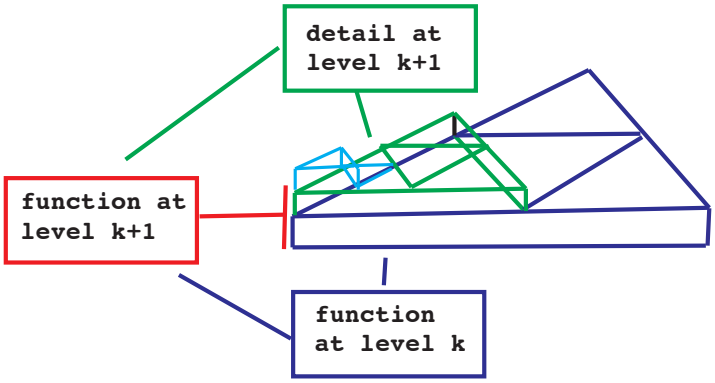


Figure 3.2: *Example of hierarchical construction of function on a single triangle. To obtain the value at $(k+1)$ resolution level (height of red step) we have to add the detail of $(k+1)$ level (green triangle) to basic value at k level (blue triangle) . The azure triangle represent an higher resolution level: we have to add "azure triangle" to blue and green ones to obtain function at level $(k+2)$*

Substantially the multiresolution analysis is performed by means of a hierarchical construction of a function in which, at any level, we add a further detail (crf. Fig.(3.2) for a schematic visualization of hierarchical reconstruction) to obtain a higher resolution level.
So, when we have evaluated the pointed out elements for fundamental level, we store them in a matrix file; then we reply the same procedure for all resolution level but calculating and storing the above-mentioned quantity for only the three wavelets which provide the detail.
At the end of building process we should have a $n \times m$ tomographic matrix A, with n number of data and m number of total parameters ($m = (20 \times 4^{k_{max}})$, with 20 number of triangles for the fundamental grid and k_{max} maximal depth of resolution)

3.2.2 Multiresolution structure of wavelet coefficient vector

When we "invert" the matrix A, or more precisely, when we solve the tomographic inverse problem (3.9) we directly find the m vector $\mathbf{x}$. The entries of $\mathbf{x}$ are all the wavelet coefficients, conveniently ordered to make immediately the multiresolution reconstruction of our model. In fact, $\mathbf{x}$ has a special structure to itself: the 1^{st} element is the solution model for the 1^{st} triangle at level 1; the 2^{nd} to the 4^{th} contain the coefficients on level 1 for the three wavelets, whose support matches the 1^{st} triangle; the 5^{th} element is the solution model for the 2^{nd} triangle at level 1; the 6^{th} to the 8^{th} contain the coefficients on level 1 for the three wavelets,

whose support matches the 2^{nd} triangle and so on until 80^{th} element.
Then from the 81^{th} to the 320^{th}, we have, with the same order, only the coefficients on level 2 for the three wavelets, whose support matches the single triangle at level 2, and so on.
We have currently performed our model up to level 5 so we have a vector $\mathbf{x}$ with $20 \times 4^5 = 20480$ elements.

3.2.3 Inversion Procedure

• Norm Damping

At the moment we have chosen to regularize the inverse problem simply by damping the norm model, i.e., imposing that the size of solution model be minimum.
Using a notation similar to that showed in (3.6), we write the generic unknown solution model $m(\theta, \phi)$ as:

$$m(\theta, \phi) = \sum_{i=1}^{N} x_i f_i(\theta, \phi), \tag{3.20}$$

where f_i are our basis functions (i.e., spherical triangular Haar wavelets) and x_i are the unknown coefficients. We define the norm of $m(\theta, \phi)$:

$$\|m(\theta, \phi)\| = \int_{\Omega} |m(\theta, \phi)|^2 d\Omega, \tag{3.21}$$

where Ω denotes the surface of unit sphere.
The relations (3.20) and (3.21) are quite general and do not depend on the parameterization. If now we substitute (3.20) into (3.21), using our parameterization, i.e. triangular wavelet functions which are piecewise constant functions with a compact support matched with triangle, we obtain:

$$\|m(\theta, \phi)\| = \alpha \sum_{i=1}^{N} x_i^2, \tag{3.22}$$

where α denotes the area of one spherical triangle.
Differentiating with respect to x_i ($i = 1, 2, \cdots, N$), because we need to find the value which minimizes the norm, and converting (3.22) in a matrix notation, we finally determine the norm damping constraint:

$$I \cdot x = 0, \tag{3.23}$$

with I the $N \times N$ identity matrix.
As for (3.14), we can redefine the our damped inverse problem :

$$\begin{bmatrix} A \\ \lambda I \end{bmatrix} \cdot x = \begin{bmatrix} d \\ 0 \end{bmatrix}. \tag{3.24}$$

• LSQR Algorithm

To assess the least squares solution of the linear inverse problem, we make use of an iterative algorithm called *LSQR* which substantially utilize the sparsity of matrices on the left hand

side of (3.9) and (3.14), i.e., their property to have a high numbers of zero elements.
The LSQR algorithm, schematically, determine successive approximations of x_{LS} by means of an iterative procedures, which exploit the sparsity of the matrix A. As many iterative algorithms, the exact convergence of approximate solutions is assured only if a very large number of iterations (typically of the order of number parameters in the model) is performed; but in practice it converge much faster, which makes it extremely convenient.
For our studies we in particular use the LSQR version by *Paige & Saunders*, [23], which is capable of evaluating, at each iteration, whether an acceptable approximation to the actual least squares solution has been achieved, using a "stopping criterion" .

3.3 Synthetic Tests

At the beginning we have tested our multiresolution wavelet analysis with two simplest model:

• a synthetic model composed by longitude bands in which the values of model are, alternatively, $1, 0, -1$.

•the topographic model from CRUST_2.0 (*Bassin et al.*, [4]) which represents the total topography, i.e., the depth with respect to the sea level.

For both test, as for real data, we use the spherical triangular Haar wavelet to lay out the inverse problem as introduced in section (3.2) and (2.2).

3.3.1 Longitude Bands Model

The synthetic model is composed by longitude bands which values are alternatively, $1, 0, -1$. It is reconstructed using paths from a real data set of surface wave phase velocities (those of Ekström *et al.*, [13]). In particular we choose the measurements of Love waves at 45s and use their source– receiver distribution to solve the forward problem (3.5) and assess a synthetic data set on the basis of the longitude bands model. Then we make the inverse procedure, and solve the inverse problem (3.9). The results of inversion are showed in Fig. (3.3).

The purpose of this first trivial test is both to make a simple example of multiresolution approach and, especially, to test the programs and routines developed (during this work) to perform a multiresolution wavelet analysis.
From (3.3) it is evident that our tools correctly work. It is also interesting to observe the hierarchical constructions of our synthetic model. We have performed only one inversion for the highest depth (level=5) but we are able to reproduce our model from level 1 to level 5 by simple using the result of inversion, namely the wavelet coefficients vector, to built the details to be add to go from a level to higher one, with a hierarchical procedure.

3.3.2 Earth Topography

We also show an example of multiresolution analysis applied to a real geophysical model: the topography. In this case, we do not perform a real multiresolution inversion; in fact we work

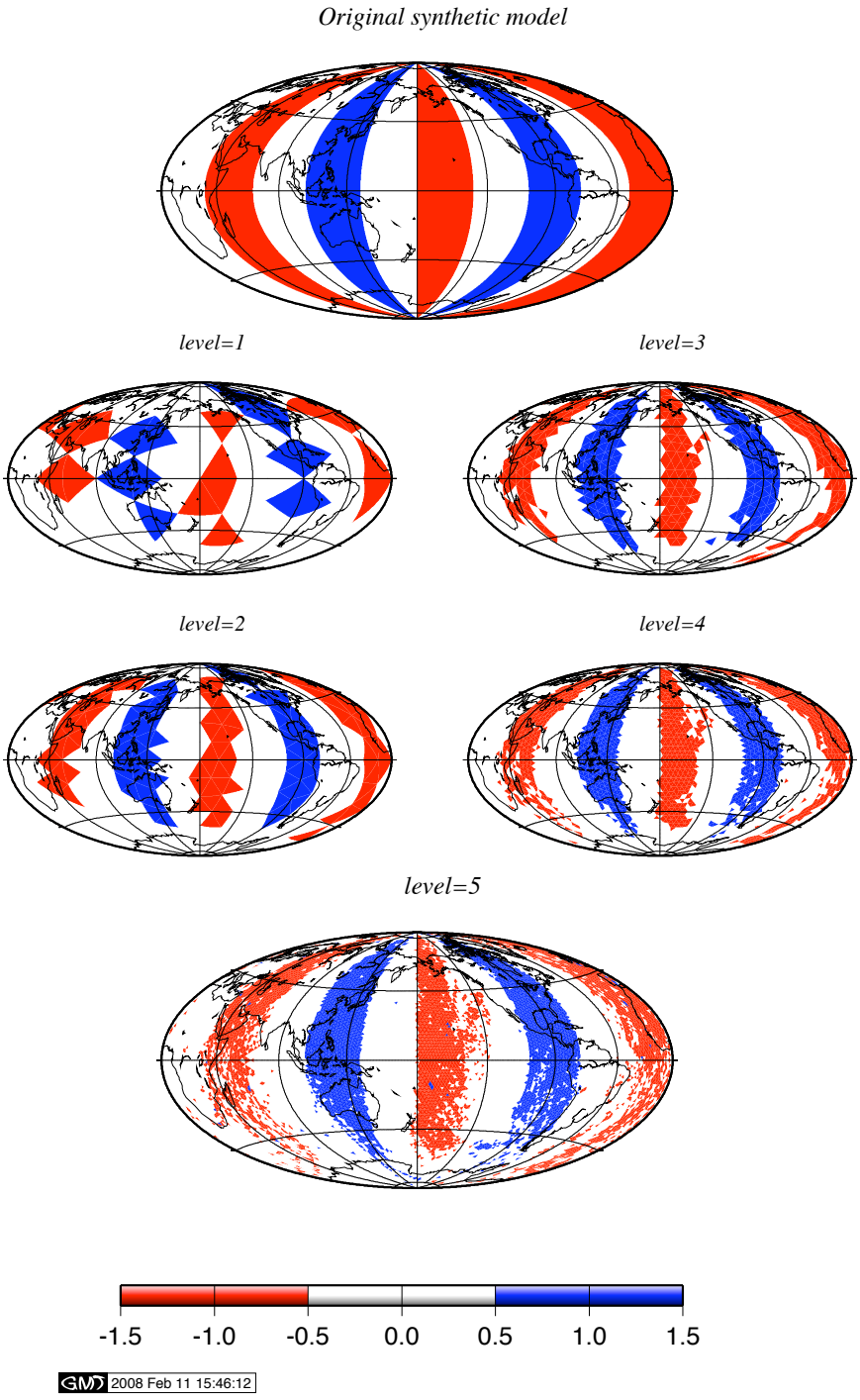


Figure 3.3: *Synthetic test: original longitude bands model (on the top) composed by 10 bands and a multiresolution inversion results at different resolution level. The map on the bottom shows the result of inversion for highest resolution level currently reached.*

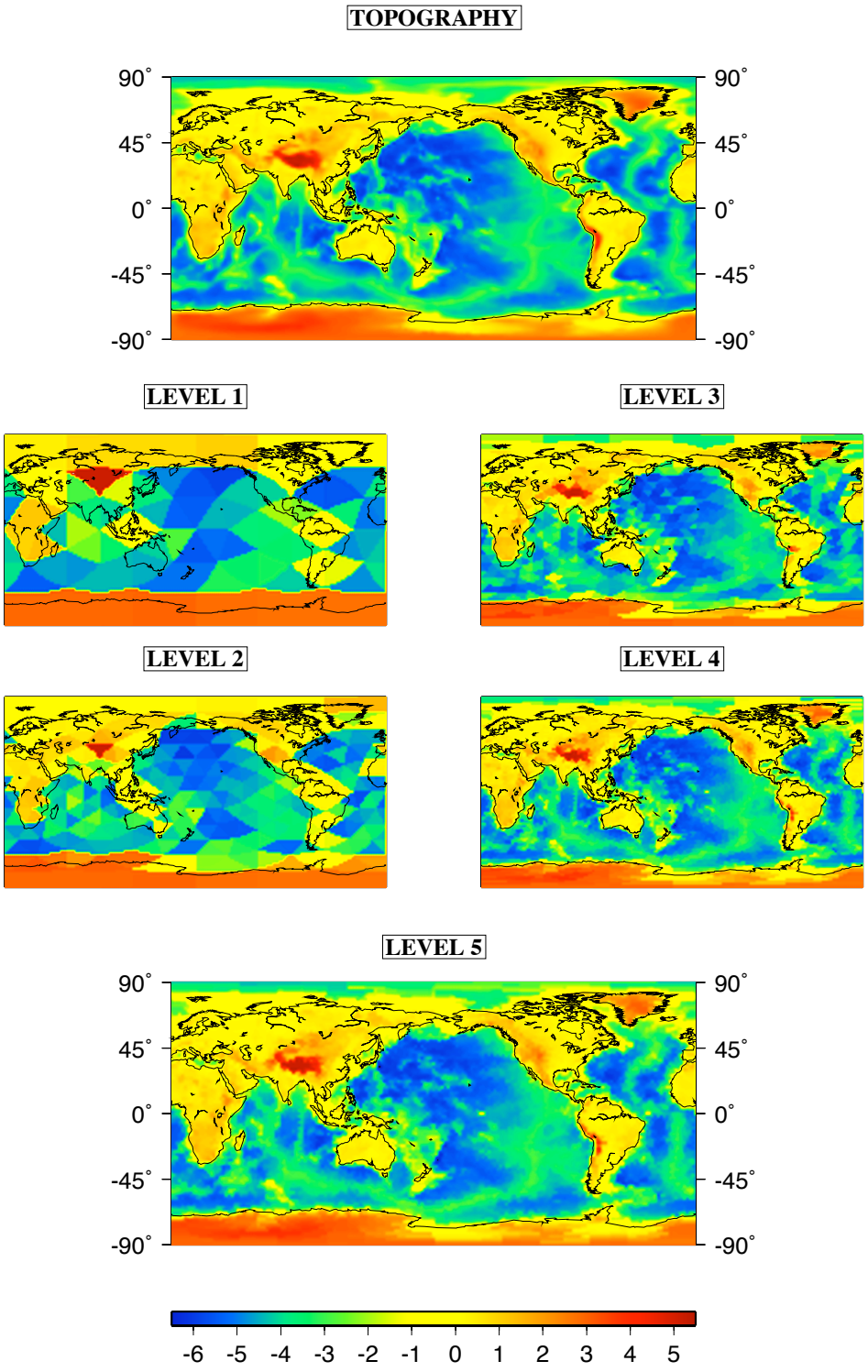


Figure 3.4: *Topographic model reproduced with wavelet analysis at different resolution level.*

with a function characterized by local value, so we do not apply the global reconstruction scheme but only a local reconstruction scheme. We start from forward problem, applied the local decomposition, then, inverting the 4×4 reconstruction matrix R (crf.(2.1)) we assess the 3 coefficients associated to 3 wavelets and that associated to scaling function.
The local reconstruction, then, give us the values of topographic function (crf. Fig.(3.4)).
Although this example is not very useful for testing our algorithms, is however interesting because he shows clearly the hierarchical reconstruction of a well known geophysical model.

3.4 Phase Velocity Maps from Surface Wave Measurements

3.4.1 Surface wave data set

Now we are interested to apply the multiresolution analysis to real data to evaluate the capabilities of wavelets in seismic tomography. We have chosen to work with the seismic surface wave phase velocity data set collected by Ekström *et al.* ([13]), 1997. In particular we focus our attention on measurements concerning to Love wave at 35 seconds. That data set is composed by 15780 source-receiver couples (i.e., phase velocity measurements) with a good uniform coverage of Earth (crf. Fig. (3.5)).
We make this choice because that is a data set particularly sensitive to crustal structure and shallowest mantle structure,i.e., earth structure of which we have a good knowledges and models to make a quantitative comparisons.

3.4.2 Multiresolution analysis

We first solve the wavelet tomographic inverse problem (3.9) for the maximum level 5 to show the hierarchical reconstruction of surface wave tomographic model. The Fig.(3.6) shows the results of inversion. It is easy to see how, with a single inversion we are able to reconstruct the model at different wavelengths. At every depth we add details so we pass from a lower to a higher resolution level. The Fig.(3.6) shows the inversion for a particular choice of norm damping (i.e., $\lambda = 1$); from Fig (3.7) we can easily guess that this is a good value to constrains our model as it also appears by the statistics analysis perform in the next section.
This is particularly true for the maximum level where we have mesh elements with the smallest size (about 230 km for side) but we obtain a model smooth so we can deduce that the constrain on norm could be enough strong.

3.4.3 Damping Effect

We want also to test the response of inversion with multiresolution wavelet analysis to norm damping. As we have already explained, we choose to impose only a norm damping. In part, this choice is justified by the smoothing wavelet effect. In fact, we have already seen that, when we decompose a signal using wavelets, we substantilly use filters that act as *averaging* filters and others that produce *details*. Some of the resulting wavelet coefficients correspond to details in the signal. If the details are small, they might be omitted without substantially

affecting the main features of function. By simply application of constrain on the norm that is translated in a condition on wavelet coefficients (crf.(3.23)) we are so able to achieve this partial selection of coefficients, making an implicit smoothing of signal.

Statistics analysis

The effect of norm damping is also evident from some statistic analysis.
We can see that the variance reduction (Fig.(3.8)) decrease with increasing of damping so as the root mean square (Fig.(3.11)) [1] (*RMS*). The decreasing of RMS, in particular, suggest us that the model become more stable when we use a strongest constrain.
This also evident from the peak-to-peak amplitude of interval (Fig.(3.10)) that become narrowest if we mainly constrain the model and the decreasing of number of iteration necessary to LSQR algorithm to reach a stable solution which indicates a fast convergence of the model with increasing of damping value .
On the other hand, the decreasing of variance reduction point out an increase of misfit between our least square solution and real model which fit the data, so we can deduce that, although the model becomes more stable, probably we are "averaging too much" the true model. From (3.7), (3.11) and (3.8) it is quite clear that a good compromise between a stable solution and a low misfit is a norm damping value in the range $(1 \div 2)$.

3.4.4 Block Model parameterization versus Spherical Triangular Haar parameterization

We want make a comparison between a commonly used parameterization, the pixel model and ours. With this goal we have focused our attention on resolution depth 3. In fact, for a consistent comparison we need to achieve both model in similar condition, so we need to perform the inversion without damping, because we do not know how the two different parameterizations reply to damping constrain and we need to obtain a comparable number of parameters. Among all possibilities we have chosen to compare a block parameterization with $6^\circ \times 6^\circ$ size pixel, which gives 1388 parameters with triangular parameterization at level 3, which gives 1280 parameters.
The Fig. (3.12) shows the results of inversion. To have a whole overview we perform the inversion with three different approaches: making an inversion up to level 5, we build the model at level 3, utilizing the multiresolution scheme, the same we make with level 4 to derive level 3, finally we directly solve for level 3. All inversions are performed using the same

[1] We define the *variance reduction* as:

$$\text{variance reduction} = 1 - \frac{\sum_{i=1}^{n}[(A \cdot x_{LS})_i - d_i]^2}{\sum_{i=1}^{n} d_i^2}$$

and the *root mean square* as:

$$\text{root mean square} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{N}}$$

respectively; so the variance reduction is a measure of how well our least squares solution fit the data and it gets closer to 1 as the fit is improved and the root mean square is a statistical measure of the magnitude of a varying quantity and it decreases with increasing of stability of least squares solution.

procedure, the above-mentioned $LSQR$ algorithm.
By comparing the results we can deduce that when we solve for level 3 we obtain a model that is consistent with block model (crf. left column in Fig. (3.12)). In the case of level 4 and 5 we have several differences with block model: this is a consequence of instability of the inverse solution. The use of damping constrains, as we have already explained, have the purpose to make more stable the ill-conditioned tomographic inverse problem; if we do not make use of any constrain we should have an instable solution, particularly if the pixels are small.
Nevertheless, if we focus our attention on, more stable, level 3 we can observe some interesting characteristics of wavelet approach. Without damping and with a comparable number of parameters the two parameterizations exhibit equal stability in the inversion procedure and in the resolution capability. Fig (3.12) shows clearly that an advantage in the use of wavelet approach consists in an intrinsic smoothing effect of our parameterization respect to block one: in the block model the pixel are more noticeable respect to triangular pixel, in fact, generally, the block parameterization require an further constrain to assure a more stable and smoothing solution; this is not true for wavelets where, as we can observe from (3.7), the only constrain on norm is sufficient to obtain stability in the least square solution.
It is also important to make in evidence another advantage of wavelet approach: the possibility of obtaining a model at different resolution level, without the necessity to repeat the inversion. If we face the tomographic problem using a pixel parameterization, we have to split the blocks to obtain a higher resolution (i.e., if we want capture lower wavelengths); but changing the size of elements of grid we need to repeat the inversion procedure.
This is not necessary in the wavelet approach because of its multiresolution nature: it is clear from right column of (3.12) that we can easily perform inverse tomographic problem up to maximum level and then piece together all lower level without the repetition of inversion procedure.

Pixel grid vs. Triangular grid

One of the interesting aspect in the application of spherical triangular Haar wavelet analysis is the use of this particular parametric grid with triangular pixel, in spite of more commonly used block pixel. Utilizing a triangular mesh have some advantages respect to a block parameterization, so the use of triangular wavelets, whose supports match the triangular cells, also hold in itself the goodness of triangular parameterization. In fact, besides to be the natural basis for developing a spherical wavelet analysis,it make possible the implementation of a multiresolution analysis by means of hierarchical subdivision of cells. In the case of square pixels we have a problem relating to spherical block areas: if we want preserve a constant area, the blocks near the poles must be enlarged to be of similar size of blocks near the equator.. This is a safeguard no necessary in the triangular case, so the distorsion effects from the poles are much reduced compared to a mesh built on a lat-long grid.
Moreover blocks show discontinuities from one block to another, causing difficulties in dynamic and cinematic ray tracing.

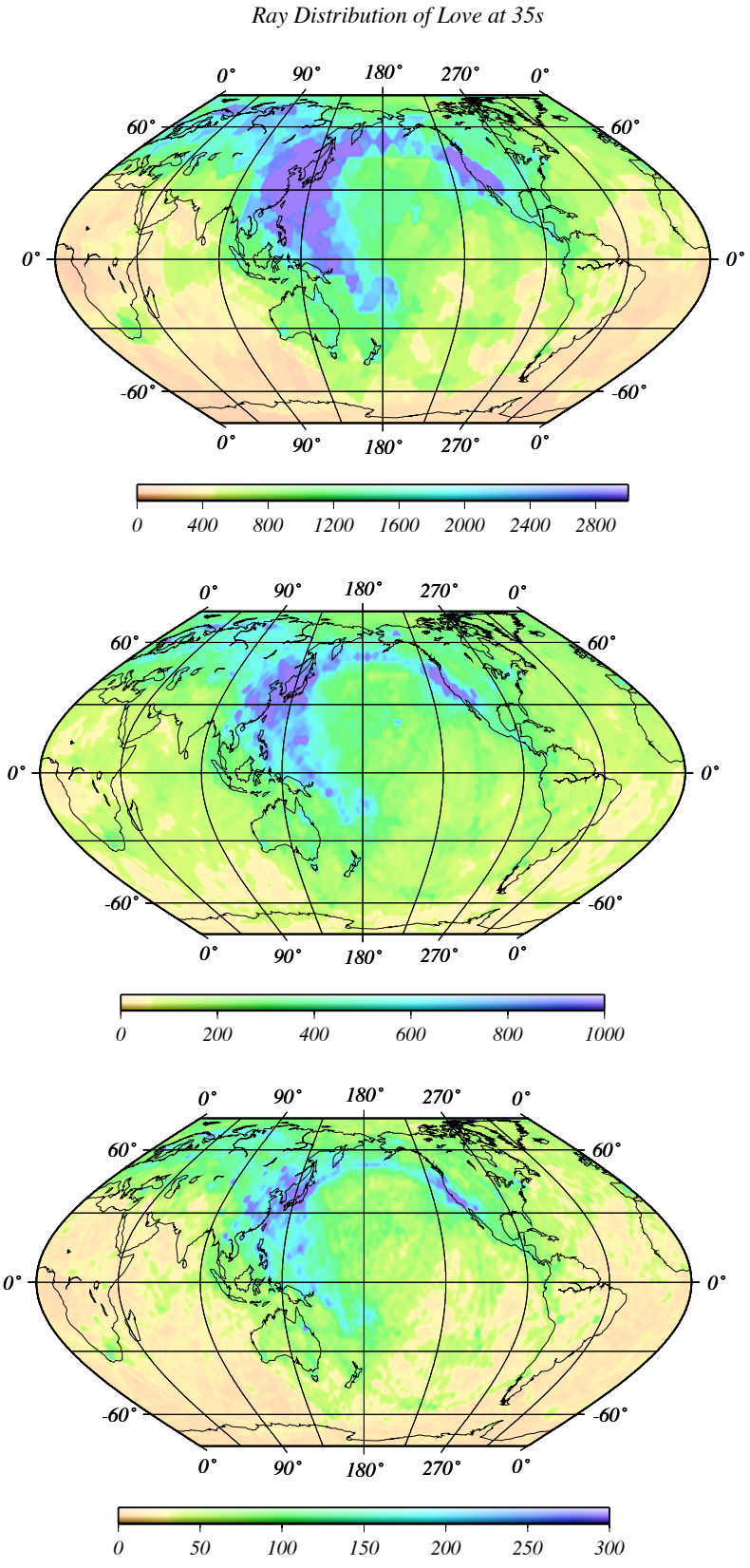


Figure 3.5: *Ray distribution for Love wave at 35s calculated as number of rays which travel along the same triangle of parametric mesh. From top to bottom they are showed the ray distributions for the parametric grid at depth 3, 4 and 5.*

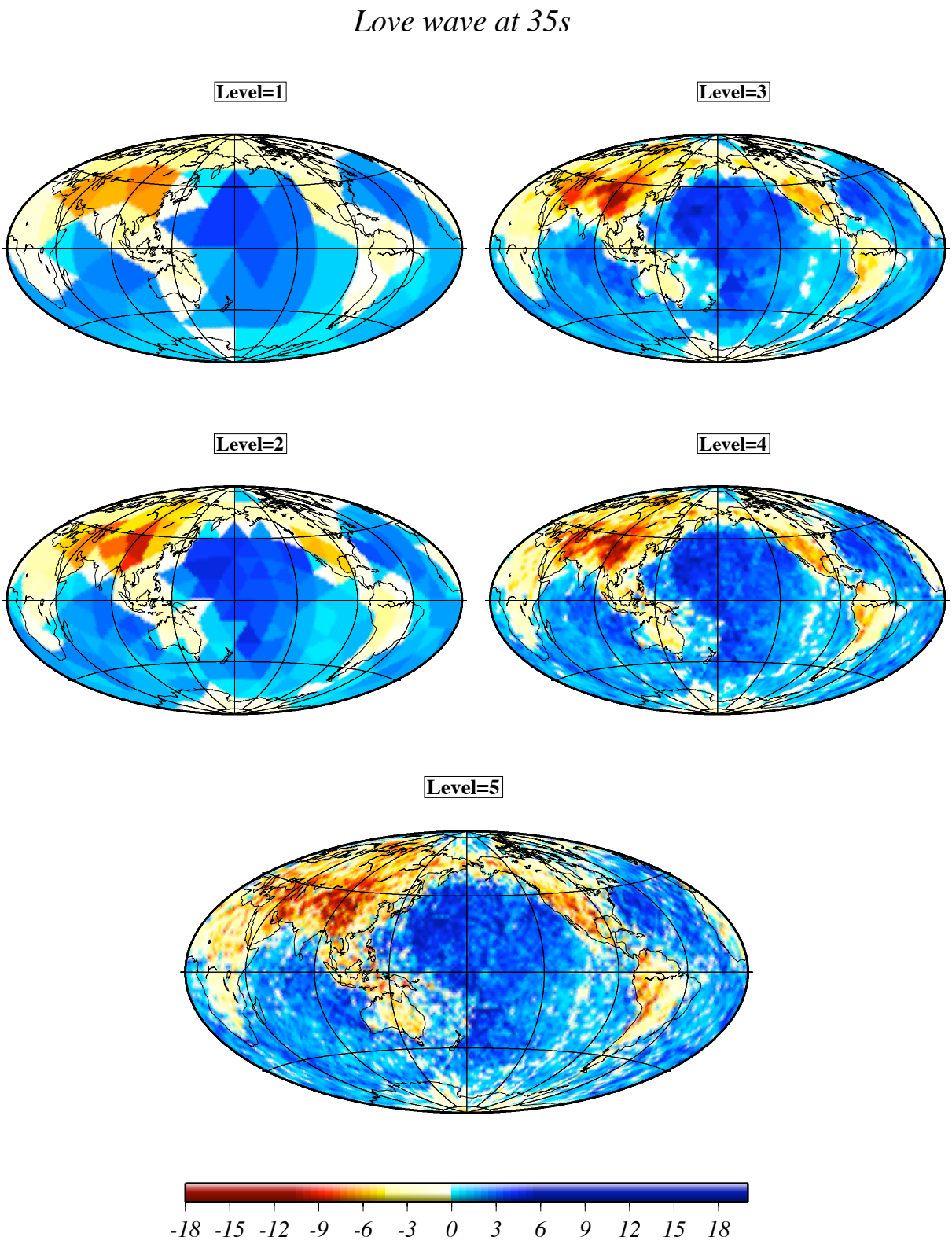


Figure 3.6: *Surface wave velocity tomography for Love wave at 35s, at different level of resolution. The figure clearly shows how the multiresolution is performed. It is the result from a single inversion, achieved with a damping value of 1.*

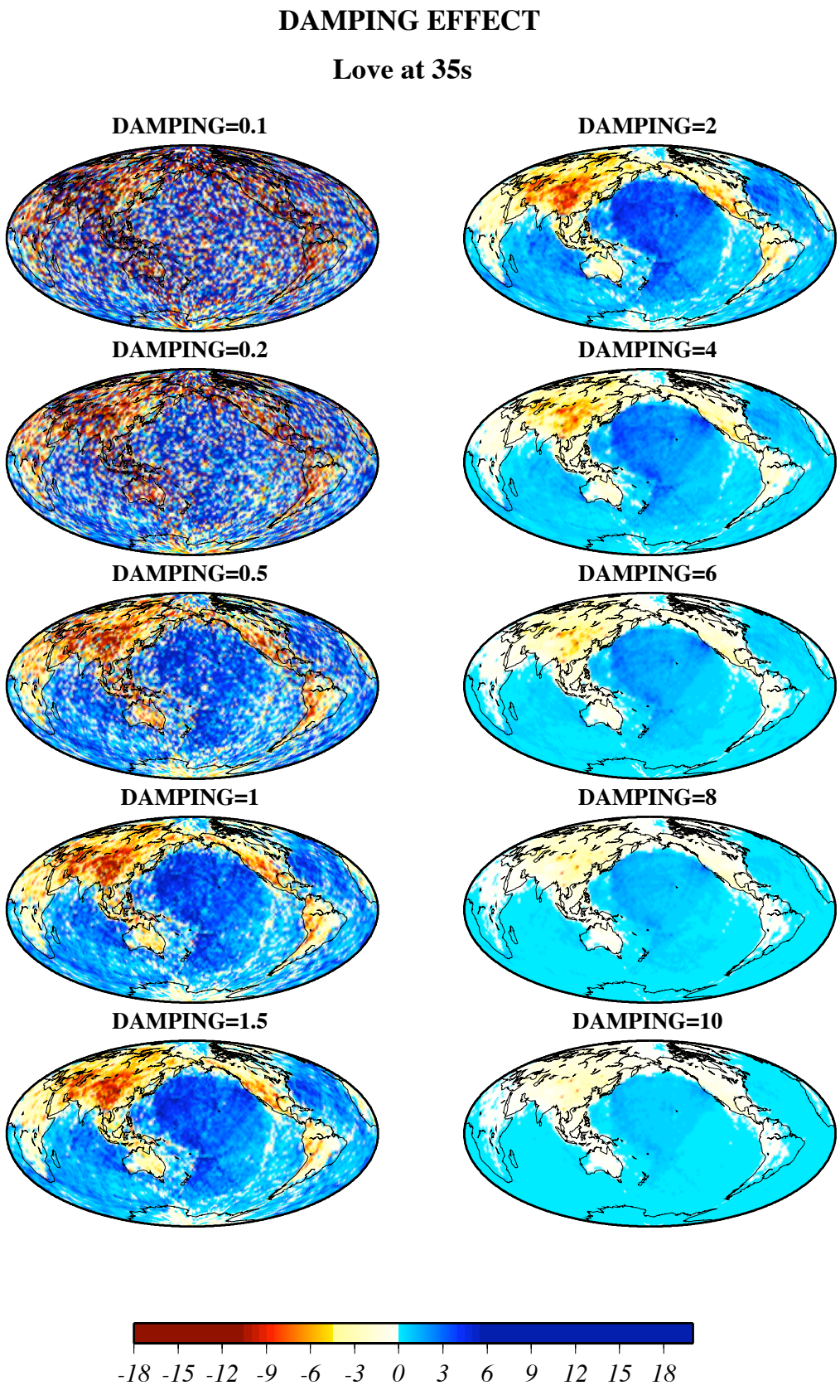


Figure 3.7: *Results from multiresolution inversion for maximum level 5, from different norm damping parameters.*

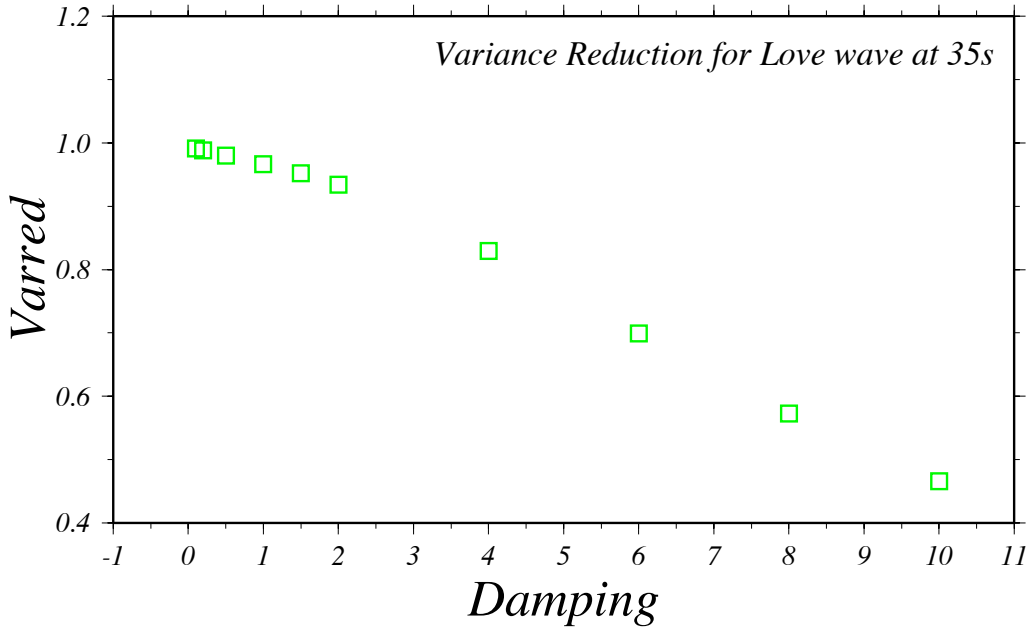


Figure 3.8: *Variance reduction as function of norm damping parameter.*

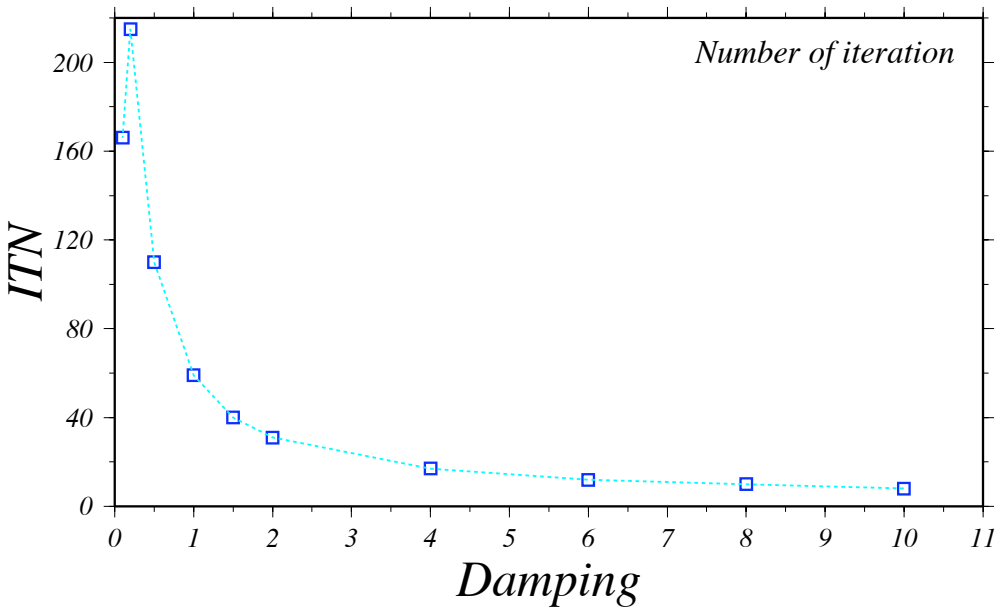


Figure 3.9: *Number of iterations performed by LSQR algorithm as function of norm damping parameter.*

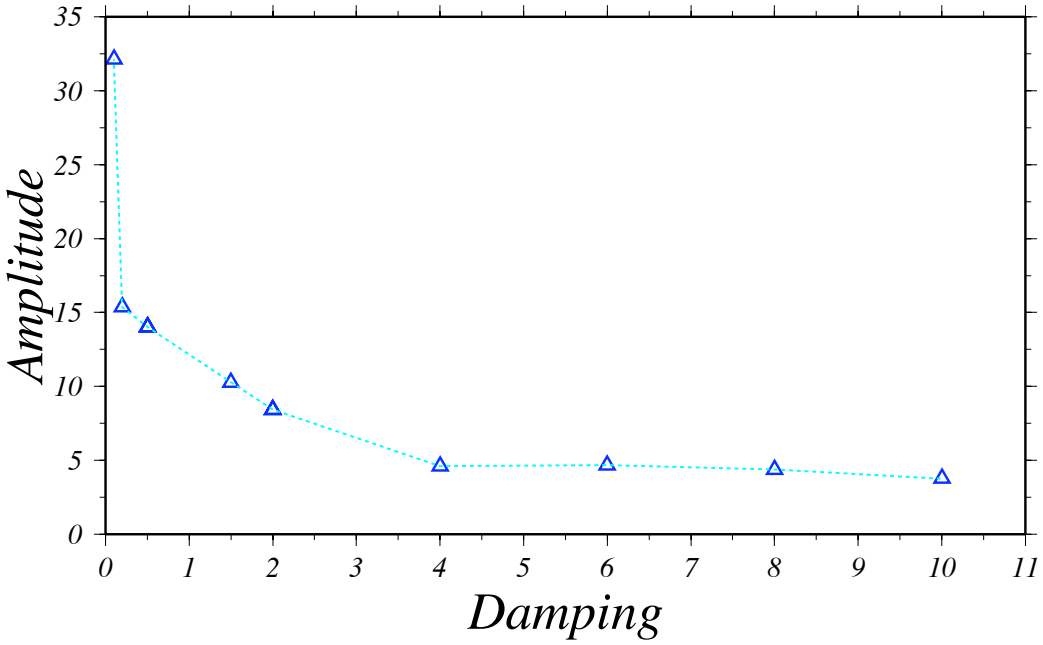


Figure 3.10: *Amplitude of interval between maximum and minimum model value as function of damping.*

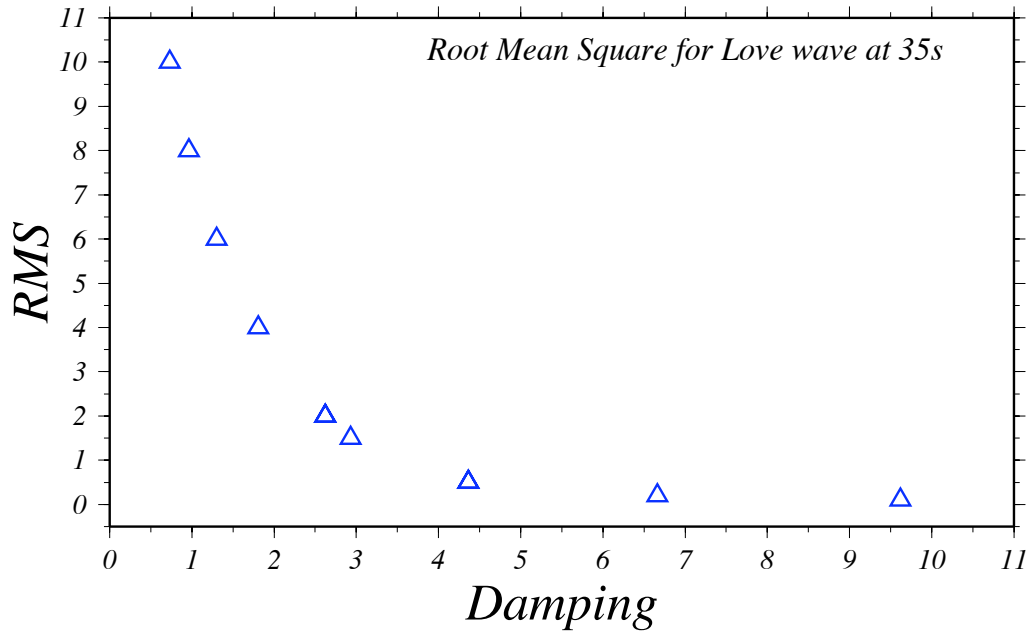


Figure 3.11: *Root mean square as function of norm damping parameter.*

Block vs. Triangular Parameterization

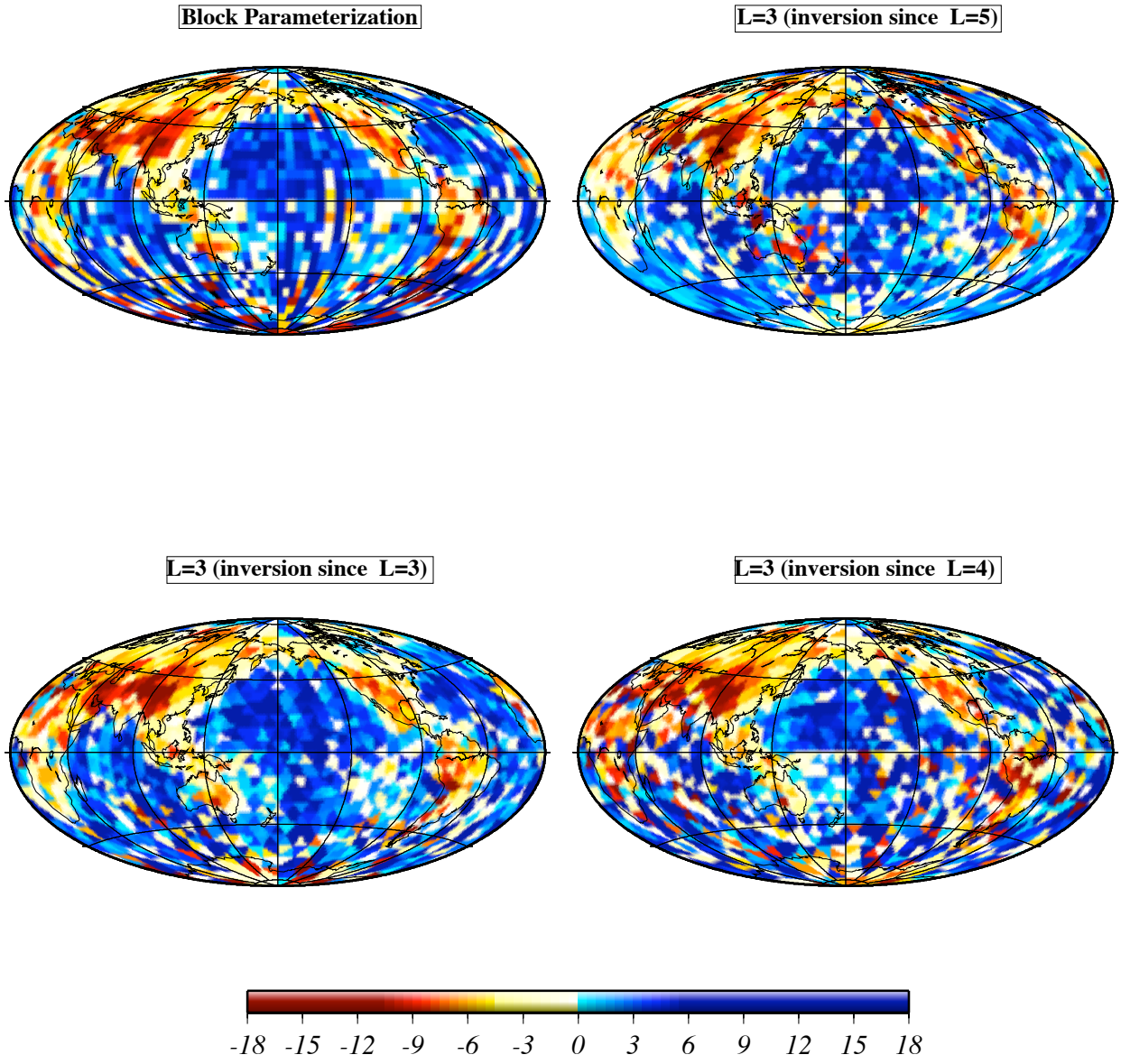


Figure 3.12: *Block model parameterization versus wavelet parameterization. The figure in top lest side of table shows the result of inversion using a parameterization with pixel of* $6^\circ \times 6^\circ$ *size; the others pictures show the results of inversion at depth of resolution 3 obtained by inversion performed at different maximum resolution level: the top left figure show the model at level 3 deduced from inversion at level 5, the two figures on bottom show the model at level 3 obtained from inversion at level 3 (on the left) and at level 4 (on the right).*

Chapter 4

Wavelet Spectral Analysis: Local Correlation Maps

Wavelets are functions with compact support so they have several interesting application for their property of localization both in time and frequency domains.
One interesting field of application for wavelet analysis is in the quantitative comparison between different tomographic models by means of systematic assessment of their spectral content. This is an important field of application for addressing reliability of different reconstructions, and for deriving geodynamic constraints on mantle flow.
Classically, spectral analysis is performed by means of spherical harmonics (e.g., *Becker and Boschi*, [5]), so it only provides an averaged spectral value associated to the global model, for each harmonic degree.
We want now investigate the capabilities of the continuous wavelet transform in the spectral analysis; we use, in particular, a correlation technique based on wavelet analysis (*Kido et al.*, [17]) to retain spatial information together with spectral content. This allow scrutiny of local features, whose representation spans over multiple scales.
One of the mean difficulty in the application of wavelets in the spectral analysis is that most applications, known in literature, are carried out in Cartesian spaces, where 2D and 3D generalization of wavelets is straightforward (e.g.,*Piromallo et al.*, [25]).
While exploration of regional scale tomographic structures can be accomplished by using handy Cartesian formulations, when we look at the globe we are confronted with the problem of scale invariance of the wavelet mother function.
Between different suggested representations for spherical wavelelets, we have chosen to adopt the formulation based on stretching a 2D function on the sphere, that does not constitute a mathematically rigorous wavelet basis, but that can be shown to possess the required scale invariance of the kernel in numerical sense.

In this Chapter we want to test the capabilities of this particular representation by means of some preliminary tests with the aim to evaluate the sensibility of kernel as function of assigned parameters as the wavelength or the distance between two different anomalies whose correlation we are interesting to evaluate.
In the second part of the Chapter we apply the wavelet spectral analysis to a real data-set of

body waves to achieve *"Local correlation maps "* between different tomographic models, to make in evidence discrepancies and similarities among them and between different profiles of the same model with the goal to identify some interesting geophysical structures in the crust and upper mantle.

4.1 The Continuous Wavelet Transform in spectral analysis

The wavelet kernel that we use for our analysis is based on the zeroth- order Bessel function and is derived by the azimuthal averaging of the well-known anisotropic Gabor transform. This type of transform is called as "*wavelet-like*" , since the kernel has been adjusted with a spherical correction, which consequently breaks the geometric similarity of the various scales in the kernel, as we going to show n the following section.

4.1.1 Wavelet-like Transform on the sphere

We want now, briefly, describe the construction of the analytical wavelet-like transform, which can be applied over the spherical surface.
Firstly it is important to keep in mind that the minimum mathematical requirements to be satisfied by a mother wavelet are:

- the mean of the wavelet function over the entire interval is zero;

- the geometric similarity must be kept over the different spatial scales.

According to the definition of the wavelet transform as a convolution integral of a kernel with the original field over the surface, we can achieve the continuous wavelet transform of a field $T(\theta, \phi)$ in the point (θ_0, ϕ_0) as:

$$\tilde{T}(\theta_0, \phi_0) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} T(\theta, \phi) K(\theta, \phi) \times \sin\theta d\theta d\phi, \tag{4.1}$$

where $K(\theta, \phi)$ is the integration kernel chosen so that the (4.1) is a wavelet transform.
Following the representation of *Kido et al.*, ([17]), we can start from a preliminary kernel of the 1D Gabor transform $F1_{k,\sigma}(x)$ which is one of the most popular mother wavelets, used for time-series analysis. It is usually composed by three parts:

- $exp[-(kx/2\sigma)^2]$ is the Gaussian window;

- $\cos(kx)$ is the basis function;

- constant factor $exp(-\sigma^2)$ that, subtracted to former value, ensures that the mean value of the kernel be zero.

The expression of $F1_{k,\sigma}(x)$ is therefore :

$$F1_{k,\sigma}(x) = \exp\left(-\left(\frac{kx}{2\sigma}\right)^2\right)[\cos(kx) - exp(-\sigma^2)], \tag{4.2}$$

where k is a wavenumber and σ/k is the width of the Gaussian window along the x-direction. We can then define the kernel of the 2D anisotropic Gabor transform in Cartesian geometry

$Fa_{k,\sigma}(x,y)$, simply by extending $\cos(kx)$ in (4.2) in the y-direction as a constant with the zero-mean adjustment term and the 2D Gaussian window.
If we re-write the 2D kernel in cylindrical coordinate system and impose it to be isotropic we have:

$$Fi_{k,\sigma}(r) = \frac{1}{2\pi}\exp\left(-\left(\frac{kr}{2\sigma}\right)^2\right)[J_0(kr) - \exp(-\sigma^2)], \tag{4.3}$$

where J_0 is the Bessel function of the zeroth-order.

4.1.2 The Gabor-type Transform Kernel

In order to map the 2D Cartesian kernel onto a spherical surface, it is necessary to apply a spherical correction $(\xi/\sin\xi)$ to the basis function, where ξ is an angular distance instead of r in (4.3). Furthermore Kido *et al.* use a normalizing factor l_w^2, where l_w substitutes the wavenumber, and corresponding to the angular degree of the spherical harmonics.
Finally the *Gabor-type transform kernel* for a sphere can be expressed as:

$$F_{l_w,\sigma}(\xi) = l_w^2\exp\left(-\left(\frac{l_w\xi}{2\sigma}\right)^2\right)\times\left[J_0(l_w\xi)\frac{\xi}{\sin\xi} - \exp(-\sigma^2)\right], \tag{4.4}$$

where ξ is the angular distance along a great circle and it can be expressed in the geographical system using ϕ_0, θ_0, ϕ and θ, where (ϕ_0, θ_0) are the longitude and co-latitude at a location to be considered for the transform, namely the central location of kernel and (ϕ, θ) is any geographical location. So we can write for distance ξ:

$$\cos\xi = \cos\theta_0\cos\theta + \sin\theta_0\sin\theta\cos(\phi_0 - \phi). \tag{4.5}$$

The parameter l_w/σ is proportional to window size, while σ is a key parameter which controls the balance of location between space and wavelength (σ sets the width of the Gaussian filter in the spectral domain).
Fig.(4.1) shows the vertical cross section of the Bessel function J_0 (blue line) and the Gabor kernel $F_{l_w,\sigma}$ (red line), for the case of $l_w = 8$ and $\sigma = 2$,as a function of angular distance ξ.
Fig.(4.3) shows instead the trend of kernel on sphere for the same value of l_w and σ: it is clear from the figure the property of localization of the kernel which tend to decrease when the angular distance from the central location increase.
It is important to make in evidence that, using the expression (4.4) we have to choose a value of $l_w \geq 4$. In fact, if we plot $F_{l_w,\sigma}$ as a function of scaled angular distance $(\xi l_w/4)$ we note that, because of the spherical correction, the shape of function depends on l_w, so it violates one of the mathematical requirement for a mother wavelet, namely the geometric similarity among the different scales. This is evident from the Fig(4.2) that shows the plots of Gabor kernel for several values of parameter l_w (i.e., $l_w = 2, 3, 4, 6, 8$ and 12), as function of the scaled angular distance. For $lw \leq 3$ the function has a singular behavior: a prominent feature is a singularity originated in the spherical correction for $l_w = 2$ and 3 and corresponding to the antipodes of the center of the kernel. The problem is partially solved if we set $lw \geq 4$ (blue and azure plot) and is definitely passed if we chose $lw \geq 8$.

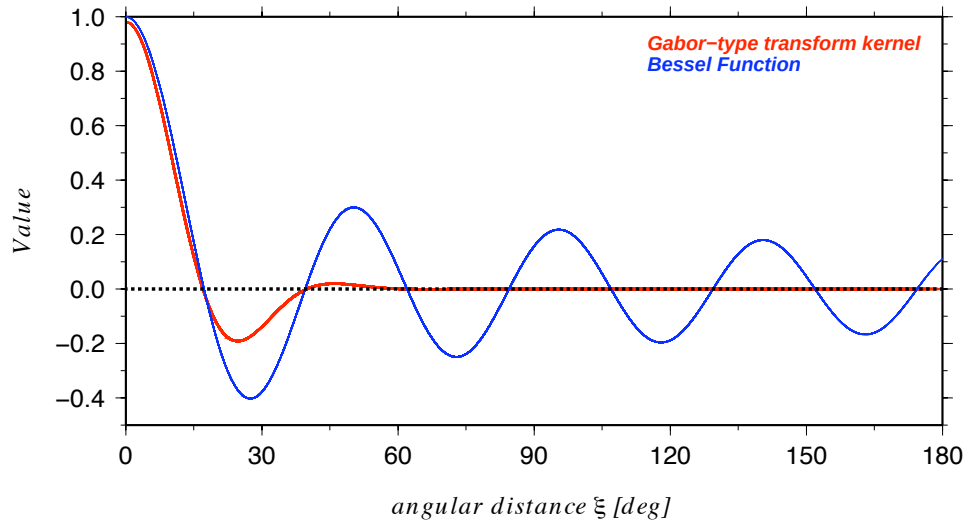


Figure 4.1: *Bessel function $J_0(l_w\xi)$ and kernel of the wavelet -like transform plotted as a function of angular distance ξ*

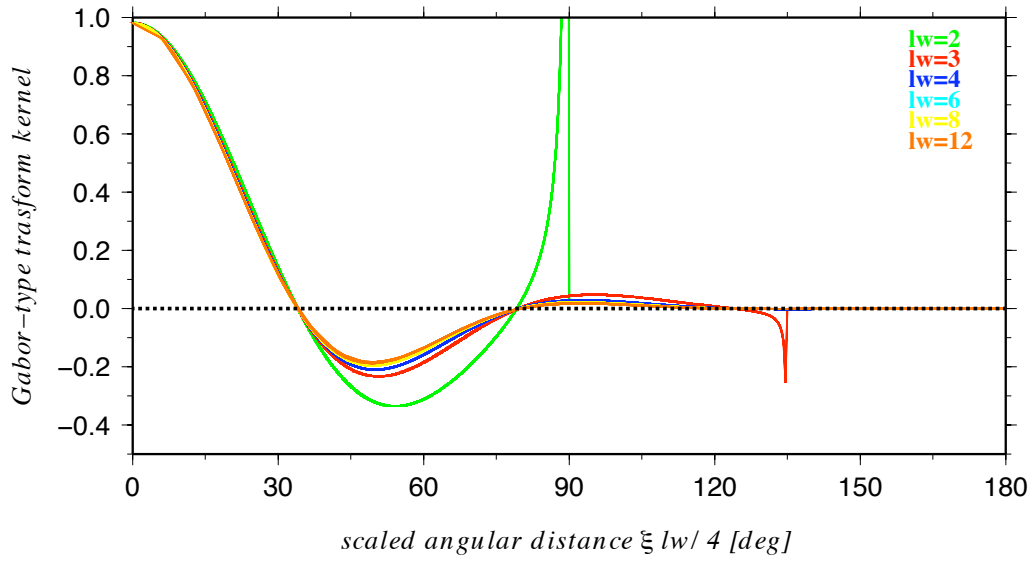


Figure 4.2: *Plots of kernel $F_{l_w,\sigma}$ (without normalization) for each $l_w = 2-4, 6, 8, 12$. The difference in the function near $\xi l_w/4 = 60°$ violates the wavelet requirement of similarity of trend for each value of l_w.*

This is substantially the reason because we use the term *wavelet-like* for our transform.
For our analysis we have therefore decided to test the special value $l_w = 4$ but generally chose

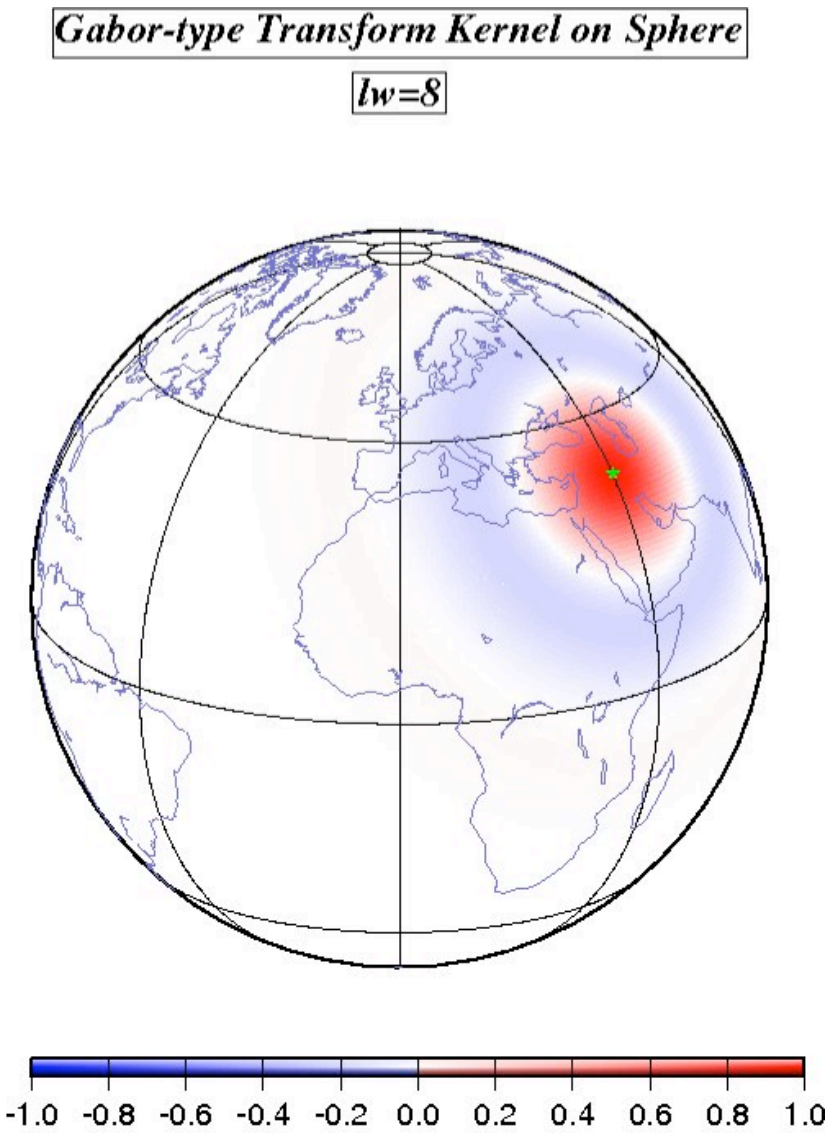


Figure 4.3: *Gabor-type transform kernel on the sphere ($l_w = 8$ and $\sigma = 2$). The green star indicate the center of the kernel.*

$l_w \geq 8$. Moreover, we have also performed some quantitative tests to evaluate the best value for the parameter σ; for brevity we do not show the results of those tests but we directly use the value $\sigma = 2$ (as also suggested by *Kido et al.*, ([17])) because it yields a good filtering capability to our a kernel .

Using the kernel (4.4), we define a wavelet transform of a field $T(\theta, \phi)$ at the geographical location (θ_0, ϕ_0) as:

$$\tilde{T}(\theta_0, \phi_0) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} T(\theta, \phi) F_{l_w, \sigma, \phi_0, \theta_0}(\theta, \phi) \times \sin\theta d\theta d\phi. \tag{4.6}$$

Property of localization of the "wavelet filter" : why use the wavelet transform?

As we have just remarked, one of the most important property of wavelet analysis consist in its property of localization. In the field of spectral analysis this means that we are able not only to achieve the energy contents of a signal but also to "localize" this quantity. This is a great advantage of wavelet transform respect to classical spherical harmonic analysis which only gives us the mean value of the spectral contents for a fixed harmonic degree (i.e., for a fixed wavelength); in other words, given a field , with the wavelet transform we can achieve the energy contents as a function of latitude and longitude, plotting a spherical map of the spectrum while with Fourier analysis we obtain a single value, for each fixed harmonic degree. Fig.(4.4) shows an example of spectral analysis performed with classical and "new-type" analysis, that make the concept of "localized spectrum" more evident and clear.
The original model (obtained from the same data-set used in the $Chapter\ 3$), is showed at the top of (4.4): the shallowest Earth structures and the contrast between continental and oceanic crust are the main observable features. When we apply a Fourier analysis we obtain the value of energy as the summation of square modulus of harmonics coefficients. Clearly, the result of summation is a single value for each harmonics degree. The green plot in the bottom of Fig. (4.4), shows the trend of spherical harmonics spectrum for a maximum harmonic degree equal to 20.
On the other hand, if we apply the continuous wavelet transform to the same signal (crf. eq.(4.6)) we achieve a localized power spectrum which can be plotted on the Earth surface as a function of latitude and longitude, as it is shown in the blue panel. The global map in the blue panel of Fig.(4.4) shows the equivalent result of harmonics analysis for a single value of L, pointed out with a blue star. It is also evident the "filter action" of wavelet transform: the filtered field shows a strongest signal where the tomographic model point out Earth structure associated to remarkable velocity variations, i.e., structure with an higher energy contains.

4.1.3 The Concept of Local Correlation

Starting from the definition of continuous wavelet transform we can compute the spatial correlation $C_{l_w,\sigma}$ between two scalar fields S and T defined on the sphere (that for us can either represent two tomographic models at the same depth, or two sections at different depths of the same 3D model) as:

$$C_{l_w,\sigma}(\theta_0,\phi_0) = \frac{\int_\Omega W_{l_w,\sigma}(\xi)\tilde{S}(\theta,\phi)\tilde{T}(\theta,\phi)d\Omega}{\sqrt{\int_\Omega W_{l_w,\sigma}(\xi)\tilde{S}^2(\theta,\phi)d\Omega \int_\Omega W_{l_w,\sigma}(\xi)\tilde{T}^2(\theta,\phi)d\Omega}}, \tag{4.7}$$

where we make use of a weighting function $W_{l_w,\sigma}(\xi)$ centered on the point of evaluation:

$$W_{l_w,\sigma}(\xi) = \exp\left(-\left(\frac{l_w\xi}{2\sigma}\right)^2\right). \tag{4.8}$$

If we plot C_{l_w} on the sphere as a function of latitude and longitude we obtain a *Local correlation map*, i.e., a map which give us information about how well the two field are correlated and "where" they are correlated.

4.2 Synthetic Test

Before applying the continuous wavelet transform to real data, we have performed two synthetic tests to evaluate the force and sensibility of our approach in the spectral analysis,i.e., its capability in catching the energy content of a signal.

4.2.1 Wavelet Analysis vs. Spherical Harmonics Analysis

We perform a first synthetic test to get a better understanding of relation between wavelet spectral analysis and spherical harmonics analysis (i.e., the classical approach generally used to perform a spectral analysis or to evaluate correlation between different fields).
We are in particular interested to understand the relation between the parameter l_w in (4.6) and the harmonic degree l and also evaluate the "filtering" action of continuous wavelet transform. With this goal we perform a test similar to the classical checkerboard test but using the tesseral harmonics [1].
For brevity we only show the results, obtained with this tests, concerning the lower ($l = 8$) and higher ($l = 20$) harmonic degree tested. From Fig.(4.5) and Fig.(4.6) it is, however, evident the filtering action of the wavelet transform and the relation between the parameter l_w and l: they are sensitive to the same wavelength.

4.2.2 Sensitivity of Correlation to the variations of characteristic parameters

We also carry out a second synthetic test to understand the dependence of local correlation from the dimension of a feature or from the distance between two distinct features that we want correlate. In this case we have built a synthetic model composed by two Gaussian anomalies $\exp(-d^2/2\sigma^2)$, with d distance from the center of the Gaussian and then we have correlated. them. We chose the value $\sigma = 25^\circ$, estimating 2σ as the amplitude of Gaussian anomaly.
Preserving the position of one of the gaussian anomaly, we shift the second one with a step proportional to its amplitude, to evaluate the gradual response of local correction to this change. At the beginning, the second field is shifted with a step that is a fraction of the anomaly; so we start with the two matching anomalies and then we look at partially overlapping anomalies. At the end the distance is a multiple of the dimension so the anomalies are completely separated

[1] Assigned a *real* function $f(\theta, \phi)$ we can expand it in terms of the *real surface spherical harmonics* as follows:

$$f(\theta,\phi) = \sum_{l=0}^{\infty} \left[a_{l0}\mathcal{X}_{l0} + \sqrt{2}\sum_{m=1}^{l} \mathcal{X}_{lm}(a_{lm}\cos m\phi + b_{lm}\sin m\phi) \right]$$

where a_{lm} and b_{lm} are the harmonic coefficients and $\mathcal{X}_{lm}$ are the real spherical harmonics:

$$\mathcal{X}_{lm} = (-1)^m \left(\frac{2l+1}{4\pi}\right)^{1/2} \left[\frac{(l-m)!}{(l+m)!}\right]^{1/2} \mathcal{P}_{lm}(\cos\theta)$$

$\mathcal{P}_{lm}$ being the *associated Legendre functions* of degree l.
When we chose $m \neq 0$ and $m \neq l$ in the expression of $f(\theta, \phi)$ we obtain the *tesseral harmonics*.

Fig.(4.7) shows the original data field, from top to bottom is evident the gradual shift of the anomaly while the first one, not shown here, remains in the first position.
The panel in the right side shows the filtered field for different values of l_w.
Fig.(4.8) shows the local correlation maps for the same values of l_w on the line and for the different relative distances between the two anomalies, on the column.
When the filter is too lower ($l_w = 4$) the dimension of anomalies and their relative distance are so that we are not able to distinguish informations about the anomalies. When the value of l_w increases we begin to observe something:
-with a factor between 0 and 0.33, i.e., with a relative distance so that $0 \leq d \leq (0.33 \cdot 2\sigma)$ we have a perfect correlation (~ 1) because the anomalies are still fully overlapped ;
-when the distance increases with a factor between 0.43 and 0.66 we begin to see the distinct centers of two anomalies characterized by two zone of negative correlation (violet zones localized near the centers of anomalies);
-the two anomalies are perfectly visible with a factor 100, i.e., the anomalies are at a distance equal to their dimension so that they are neighboring.
The same results are visible with $l_w = 12$ and partially with $l_w = 16$. At $l_w = 24$ the value is probably too much higher, so it filters at wavelengths lower than those we are studying.

4.3 Correlation between different Tomographic Models

As we have just underlined with the synthetic tests, one of the most powerful application of continuous wavelet transform consists in its capability of localizing both in frequency and in space domain, the spectral contents of a signal. This property make possible to perform local comparison between different tomographic models by means of the local correlation.Unlike of the classical correlation, obtained by spherical harmonics analysis, correlation calculated starting from wavelet transform give us a global map which shows, spatially localized, the value of correlation and so, where two models are more or less in agree.
We apply the (4.7) to two couples of body wave velocity models.
Firstly we examine two different S-wave tomographic models: the *s20rts* by *Ritsema et al.*,[27] and the *saw642*, the most recent model developed by *Panning & Romanowicz,* [24] (2006). Figures (4.9) and (4.10) show the original models, while figures (4.11) and (4.12) are the filtered versions for several value of parameter l_w ($l_w = 4, 8, 12, 16$ and 24).
Fig. (4.13) and (4.14) finally show the most important results. First of all, we can note that, when the parameter l_w is too much low ($l_w = 4, 8$), the transform is only able to capture structure with highest wavelengths so the models result perfectly correlated. When we go to study the correlation concerning the higher value of filtering parameter we see that the models are generally in agreement , best on the continents respect to ocean zones.
We can also see the presences of some anti-correlated zones (blue-violet anomalies); for example:

- at 200 km it is visible a “negative anomaly” in the correlation, localized in the middle Pacific Ocean which remains up on great filtering value and which can be related with the opposite values of models (positive variation in saw642 and negative in s20rts) visible in (4.9) and

(4.10).
- between 600 and 800 km for value of $l_w = 8, 12, 16$ it also evident a negative correlation zone near middle of Africa, probably correlated to differences (in sign) in this zone between the models; the same is true for the anti-correlated zone in the middle right side of Pacific Ocean .
Clearly the local correlation maps become more patched when we increase the value of l_w: to explain these results is more difficult if we start from the original models, because the wavelengths are enough low, but it becomes clearest if we study the filtered fields.

We repeated the same analysis for two distinct P-wave tomographic models: the *bdp00* by *Antolik et al.*, [2] and the *kh00p* by *Kárason & van der Hilst*, [16]. Also in this case we can deduce the same conclusions for lower value of filtering parameter. When we go to examine higher value we find again some negative correlation zone; for example:
- at 100 km it is visible an anti-correlation zone in the middle of Pacific Ocean consistent with the little positive anomaly visible in $bdp00$ and absent in $kh00p$;
- a similar anomaly is also visible at 400 and 500 km probably compatible with positive anomaly visible in $kh00p$ near Fiji Island, absent in $bdp00$; to this anomaly is coupled another anti-correlation zone near Indian peninsula: also in this case we can notice a positive anomaly in the $bdp00$ model absent in the other one

We can substantially deduce that, in the comparison between two different tomographic models, the local correlation correctly work, detecting locally the zone of discrepancies and that it is, in particular, more sensible to variation in sign of field rather than in amplitude.

4.4 Correlation between different depths

Starting from the just analyzed S-wave tomographic model *saw642*, which is one of the most recent models available in literature, we want now to show another possible and innovative application for continuous wavelet transform : the local correlation maps obtained between two different profile of the same model.
Because of the enormity of analysis which is possible to make in this particular field we chose to show only a few part of our results concerning this application, that appear us more interesting.
Fig (4.21) shows, for example, the auto-correlation maps [2] of profile at 100 km with profiles in an interval around discontinuity at 200 km. The computation is performed for several value of filtering parameter but we show here only those concerning the more significant value ($l_w = 8, 12, 16, 24$). For $l_w = 8$ we observe a good agreement between different profiles with the presence of a zone with a lower correlation which range the Pacific Ocean (crf. map $(0100/0230)$ which shows the local correlation between 100 km and 230 km). Th anomaly become more important when we increase the value of l_w and we can see, for $l_w = 12$ and 16 the presence of another anomaly on the Indian Peninsula, both consistent with a gradual change of sign of velocities field (from positive to negative).

[2] We call *auto-correlation map* the map obtains correlate a model with itself at different depths.

Fig (4.22) and Fig (4.23) shows instead, the auto-correlation maps of couple of profiles with a distance of 100 and 200 km.In this case we chose to focus our attention on deeper layers near discontinuity at 660 km. Differently by (4.21), when we have a narrow analyzing step of 10 km, here we study a wider interval. As we aspect when the value of filtering parameter is low and we examine a correlation distance of 100 km, we have a good correlation between depths. The results change gradually when we increase l_w, so for $l_w = 16$ and 24 we see the presence of two negative anomalies: one on the Indian Peninsula, which is visible in maps $(0400/0600)$, $(0500/0600)$ and $(0500/0700)$ and another negative anomaly between Australia and Papua New Guinea, visible in maps $(0600/0800)$, $(0700/0800)$ and $(0700/0900)$.

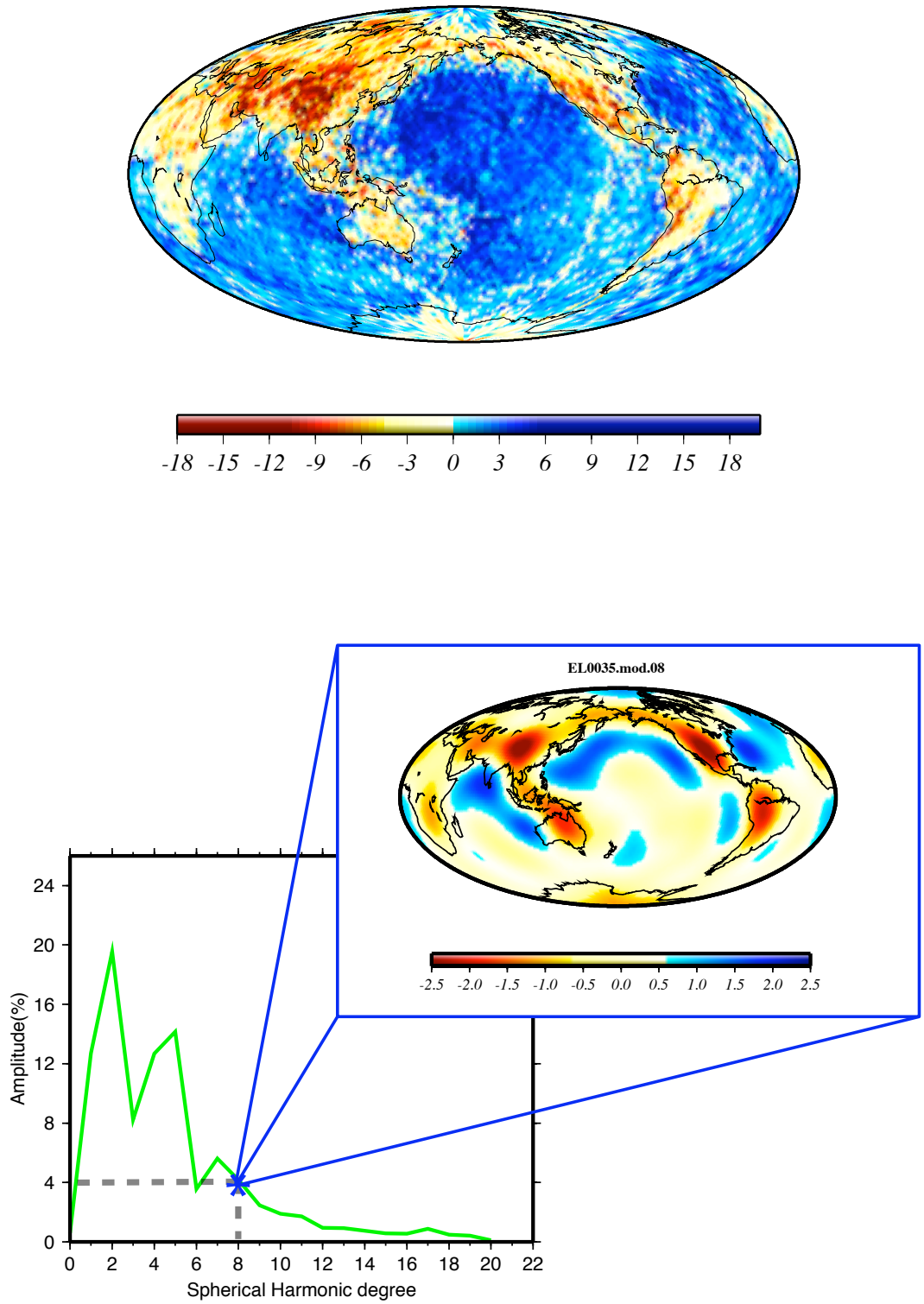


Figure 4.4: *Original model from Love wave at 35 s (on the top). Spectral analysis performed with spherical harmonics (green plot on the left) and with wavelet transform (map in the blue panel on the right)*

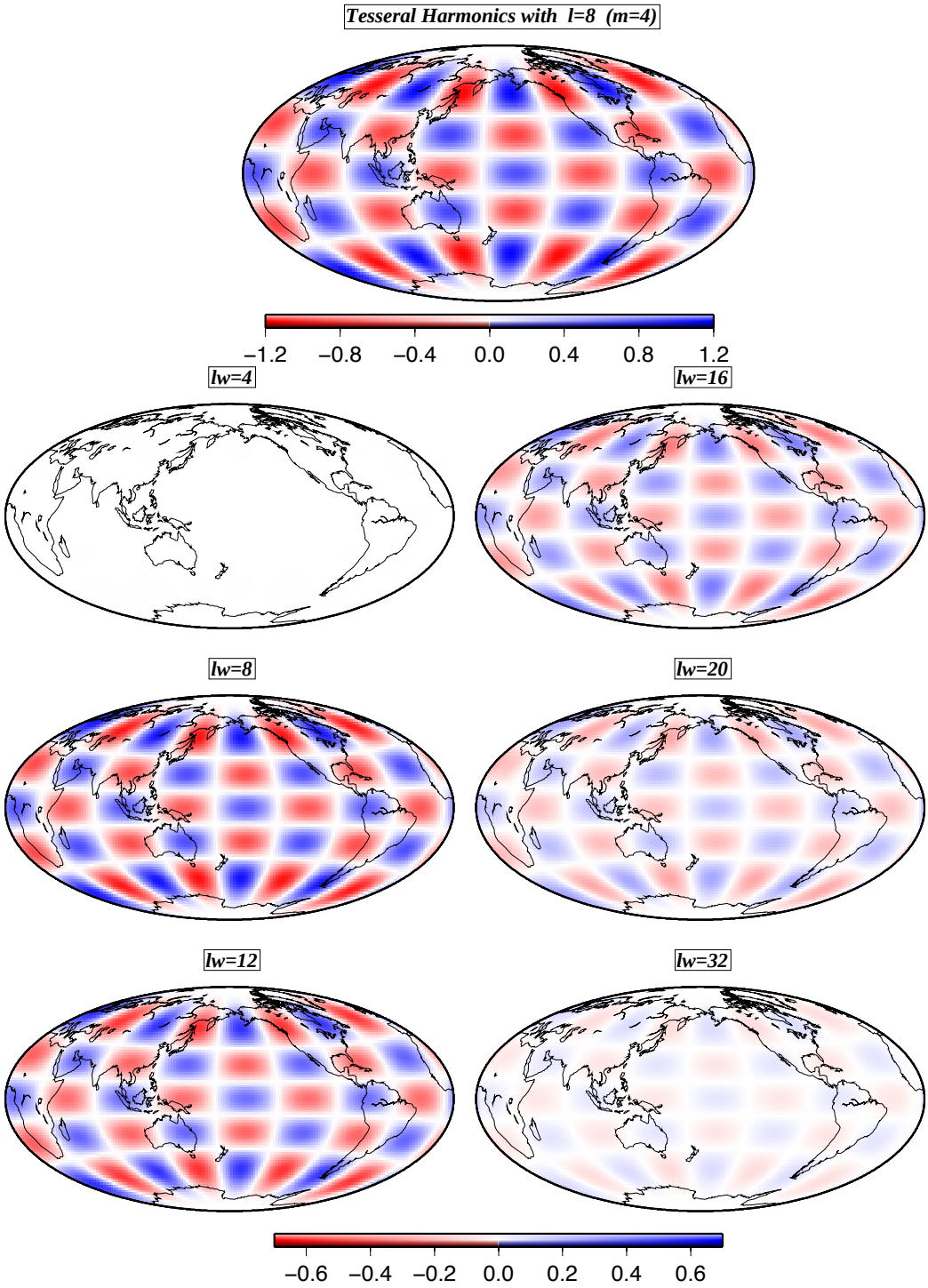


Figure 4.5: *Tesseral harmonics field with* $m = 4$ *and* $L = 8$ *(top of the figure) and its filtered versions obtained applying the "wavelet-like" transform with* ($l_w = 4, 8, 12, 16, 24$).

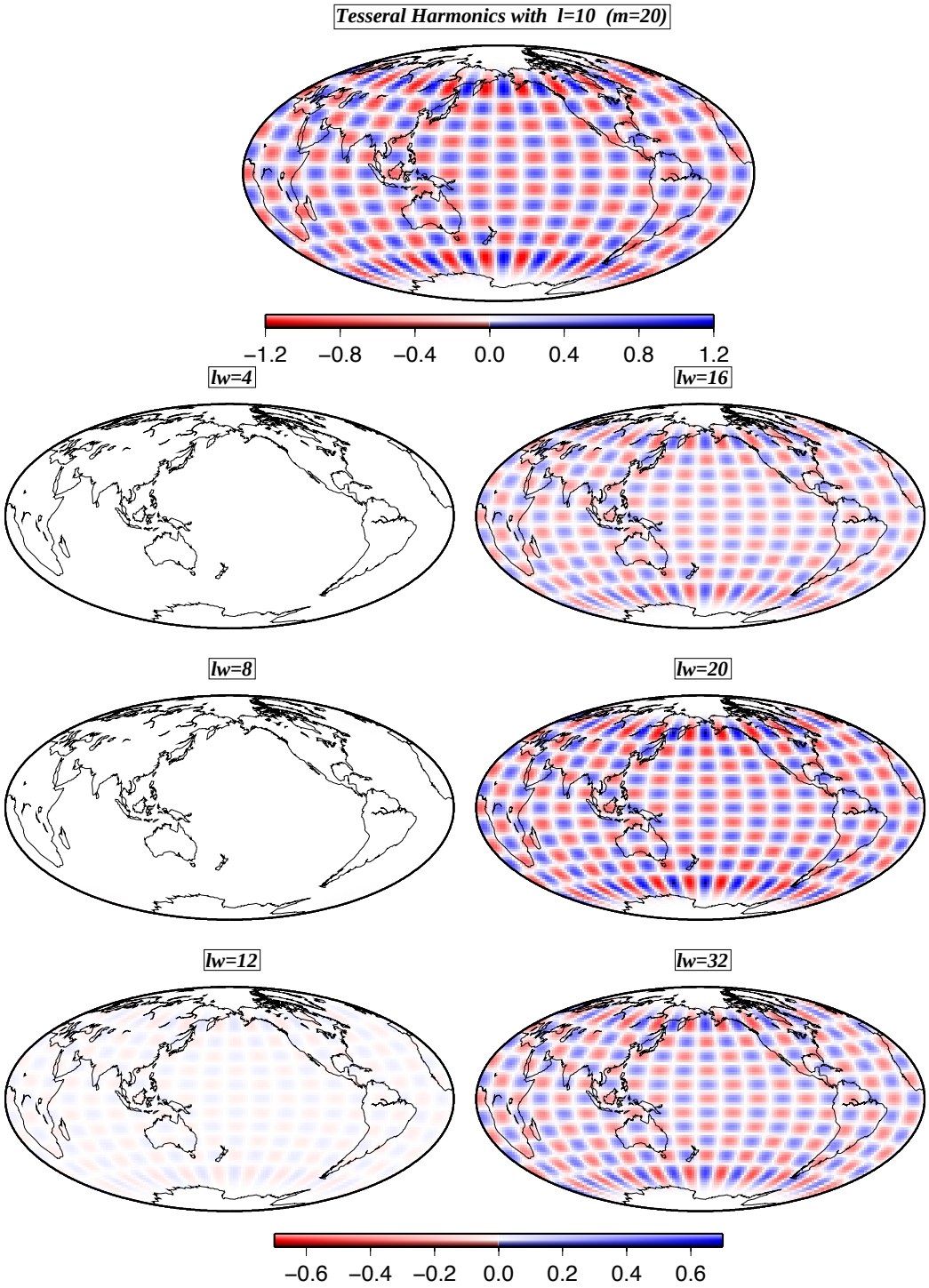


Figure 4.6: *Tesseral harmonics field with* $m = 10$ *and* $L = 20$ *(top of the figure) and its filtered versions obtained applying the "wavelet-like" transform with* ($l_w = 4, 8, 12, 16, 24$).

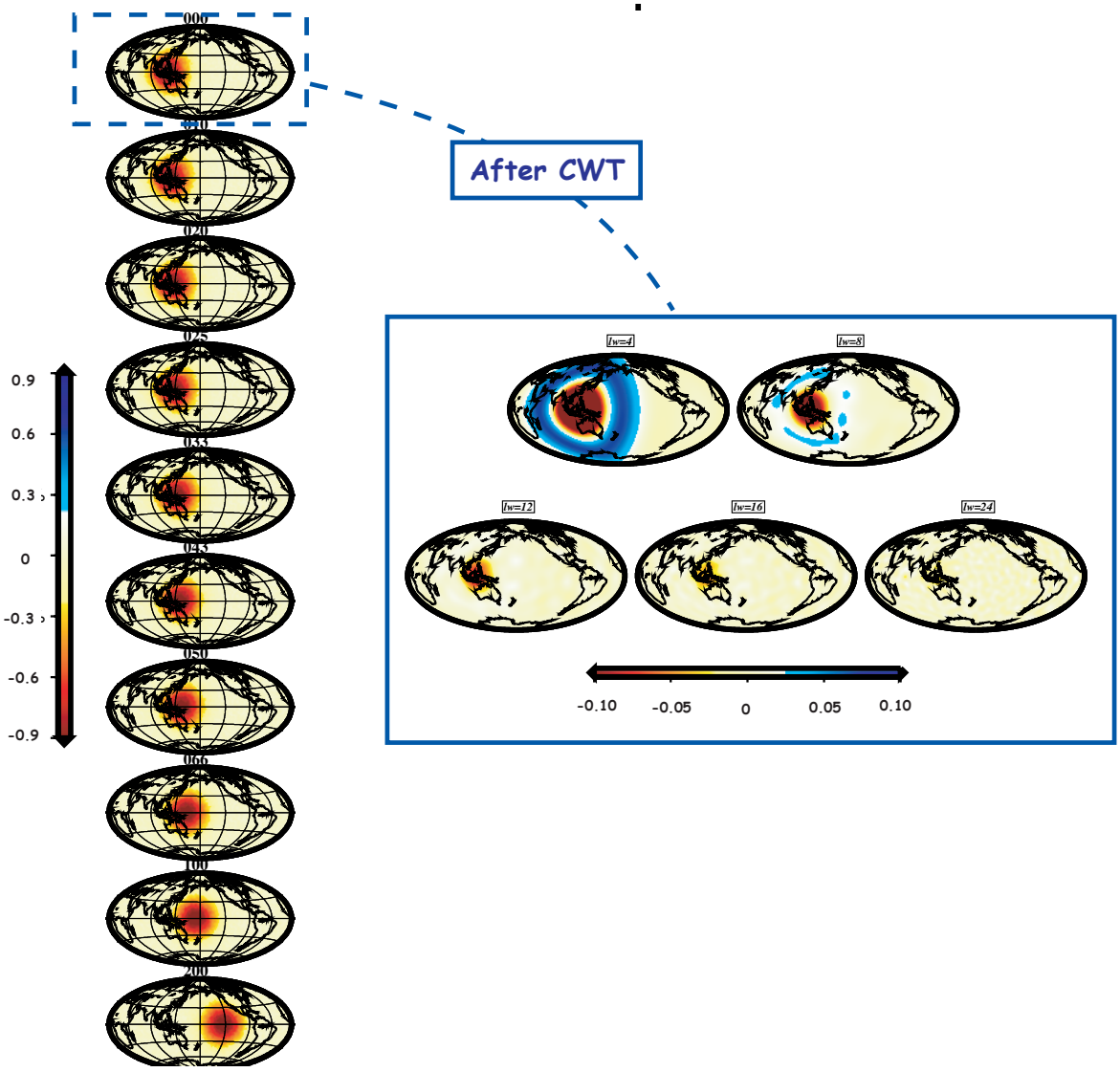


Figure 4.7: *Synthetic test with two shifted gaussian anomalies. In the blue panel is shown the first filtered field for* $(l_w = 4, 8, 12, 18, 24)$*. The other filtered fields are equal but shifted as also shown for the original field.*

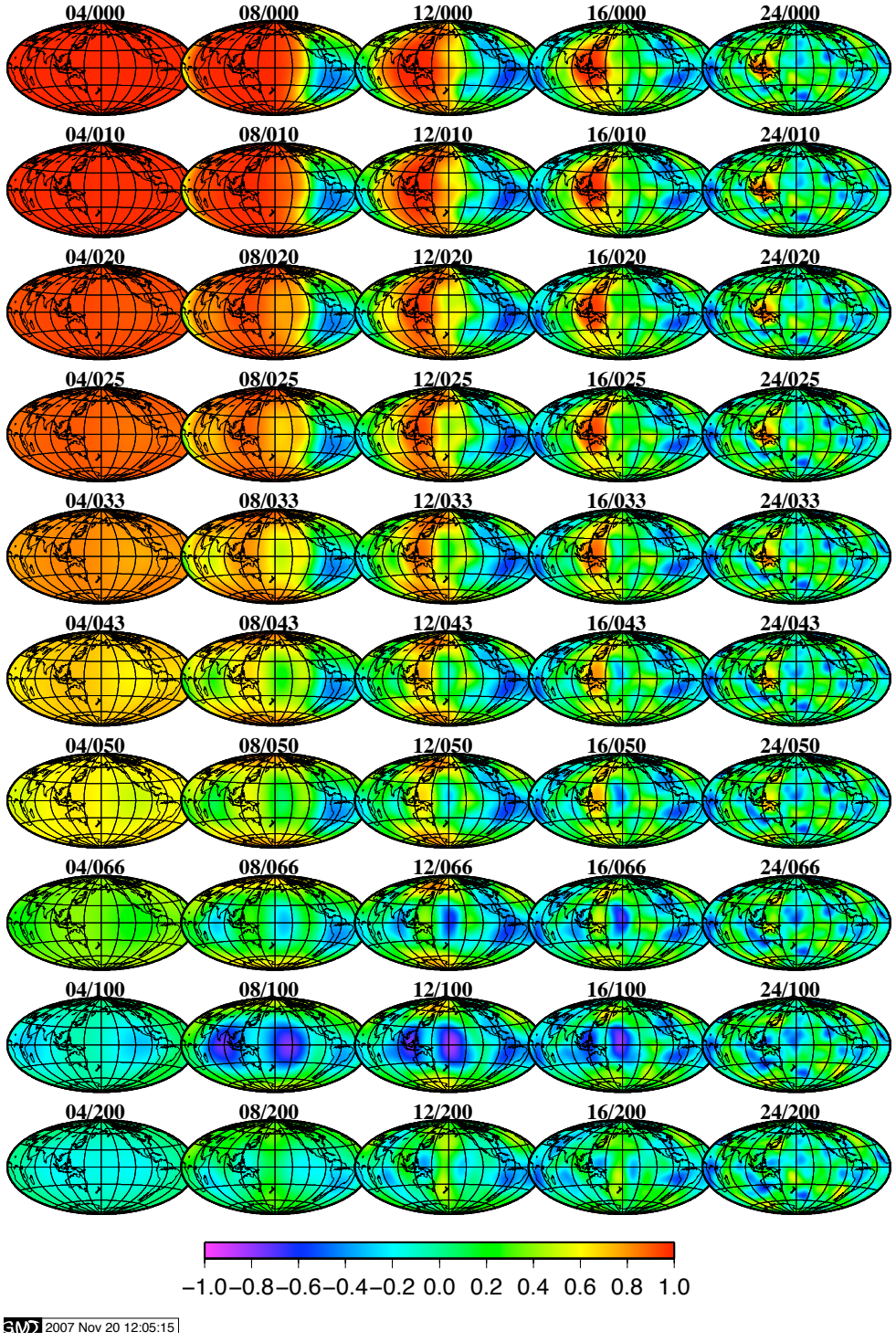


Figure 4.8: *Local correlation maps from gaussian synthetic test: the lines show the result at the same distance, on the columns there are the results for the same values of* l_w.

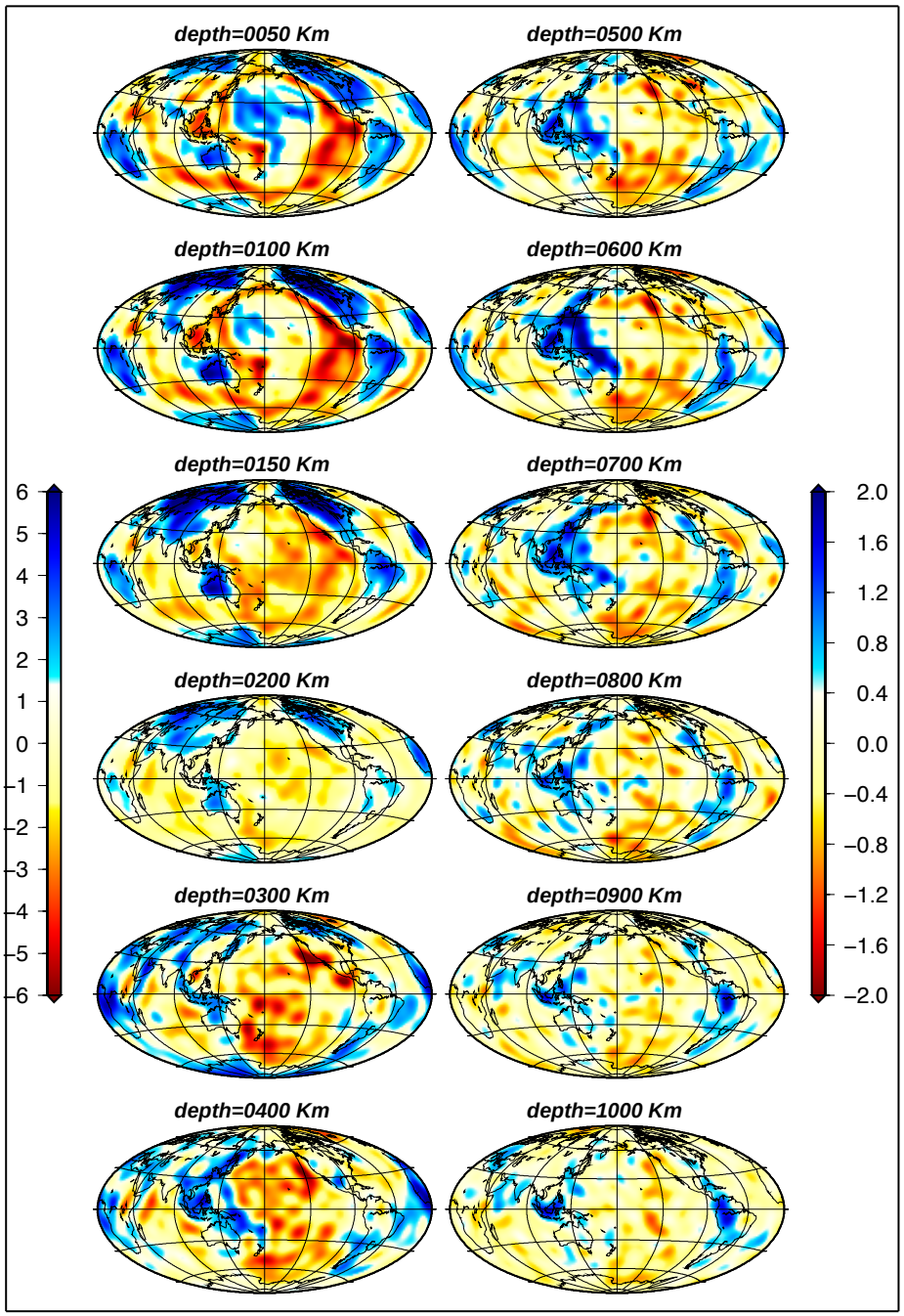


Figure 4.9: *S-wave tomographic model s20rts (Ritsema & van Heijst, [27]).*

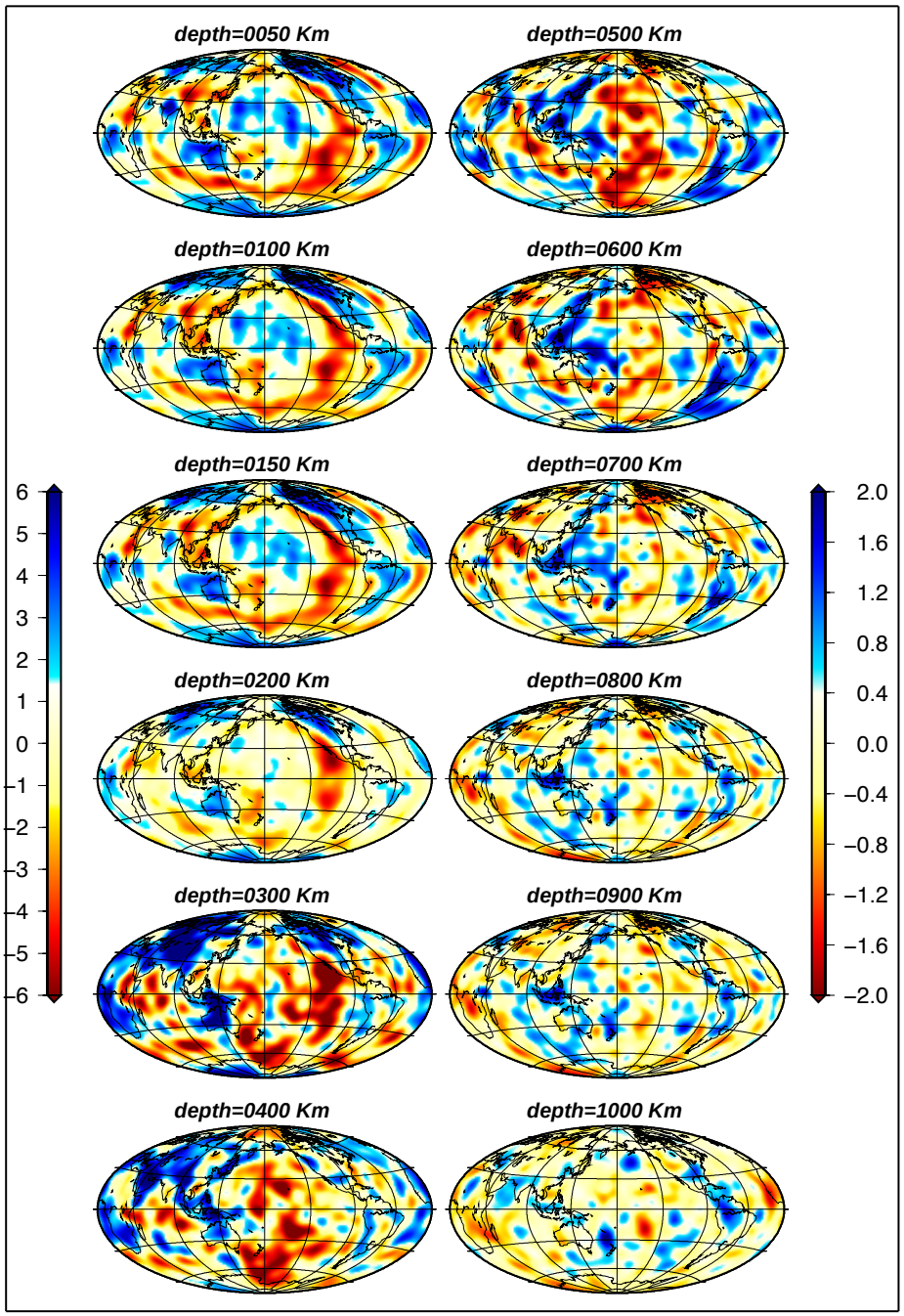


Figure 4.10: *S-wave tomographic model saw642 (Panning & Romanowicz, [24]).*

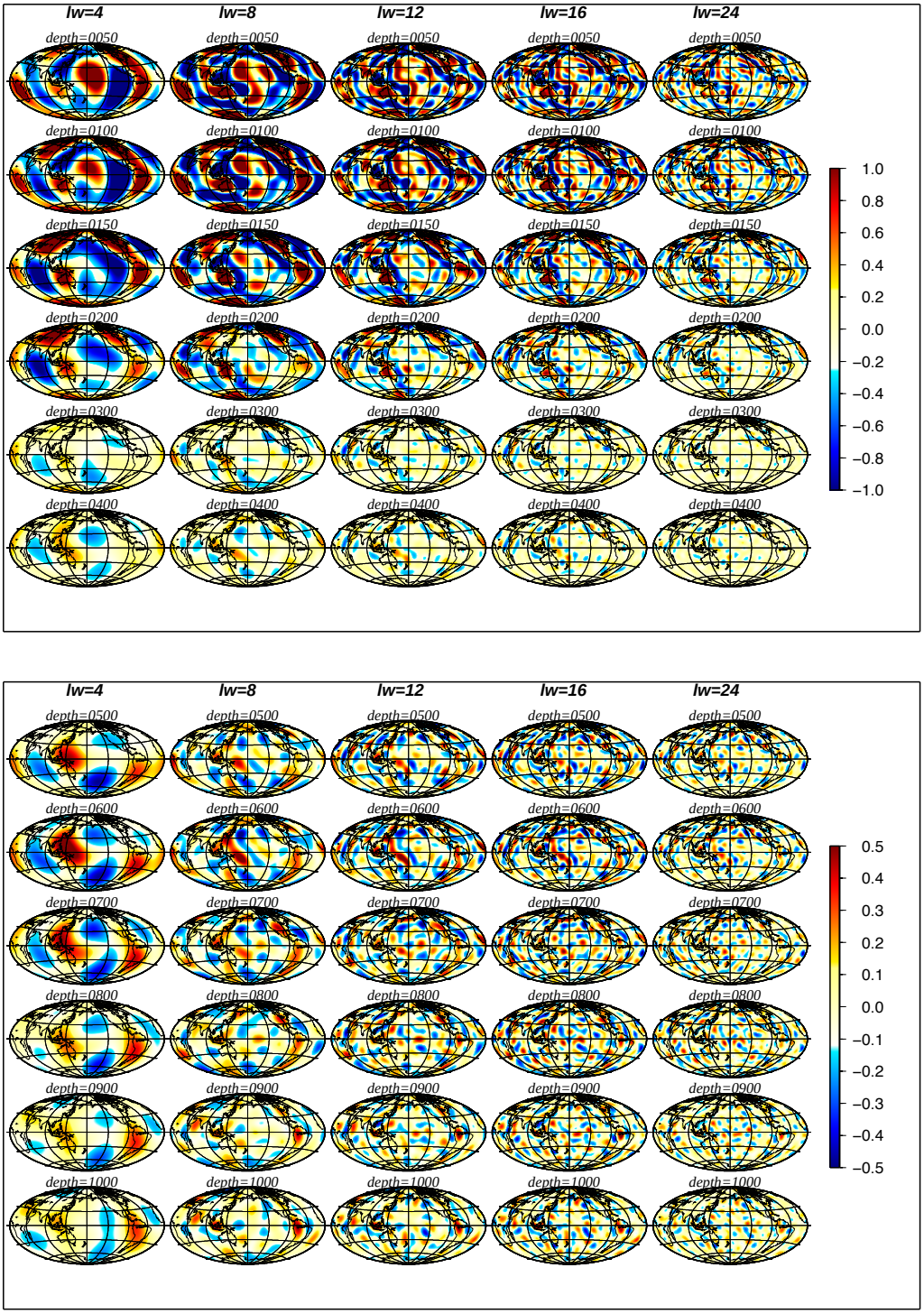


Figure 4.11: *Filtered version of original tomographic model: s20rts .*

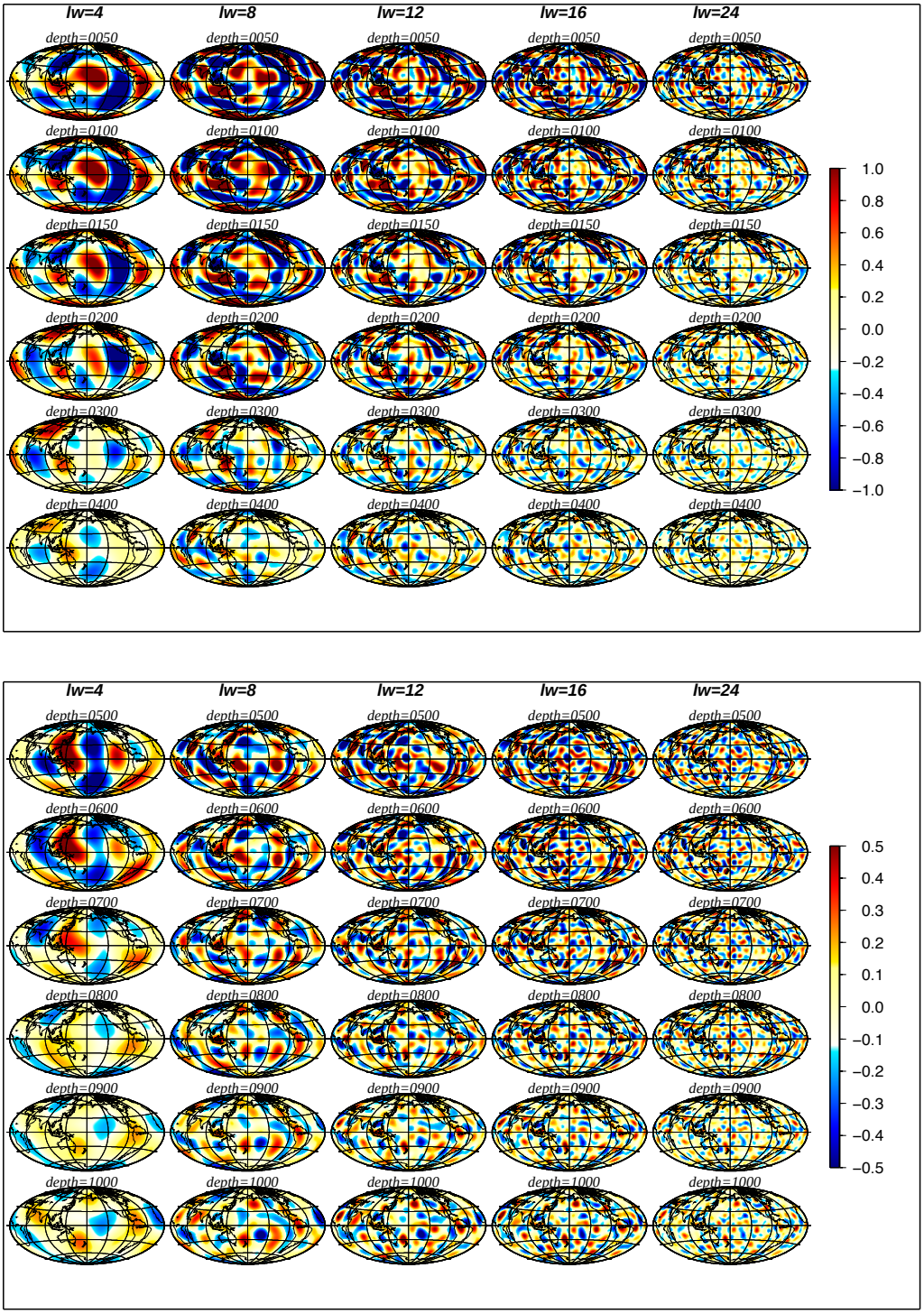


Figure 4.12: *Filtered version of original tomographic model: saw642 .*

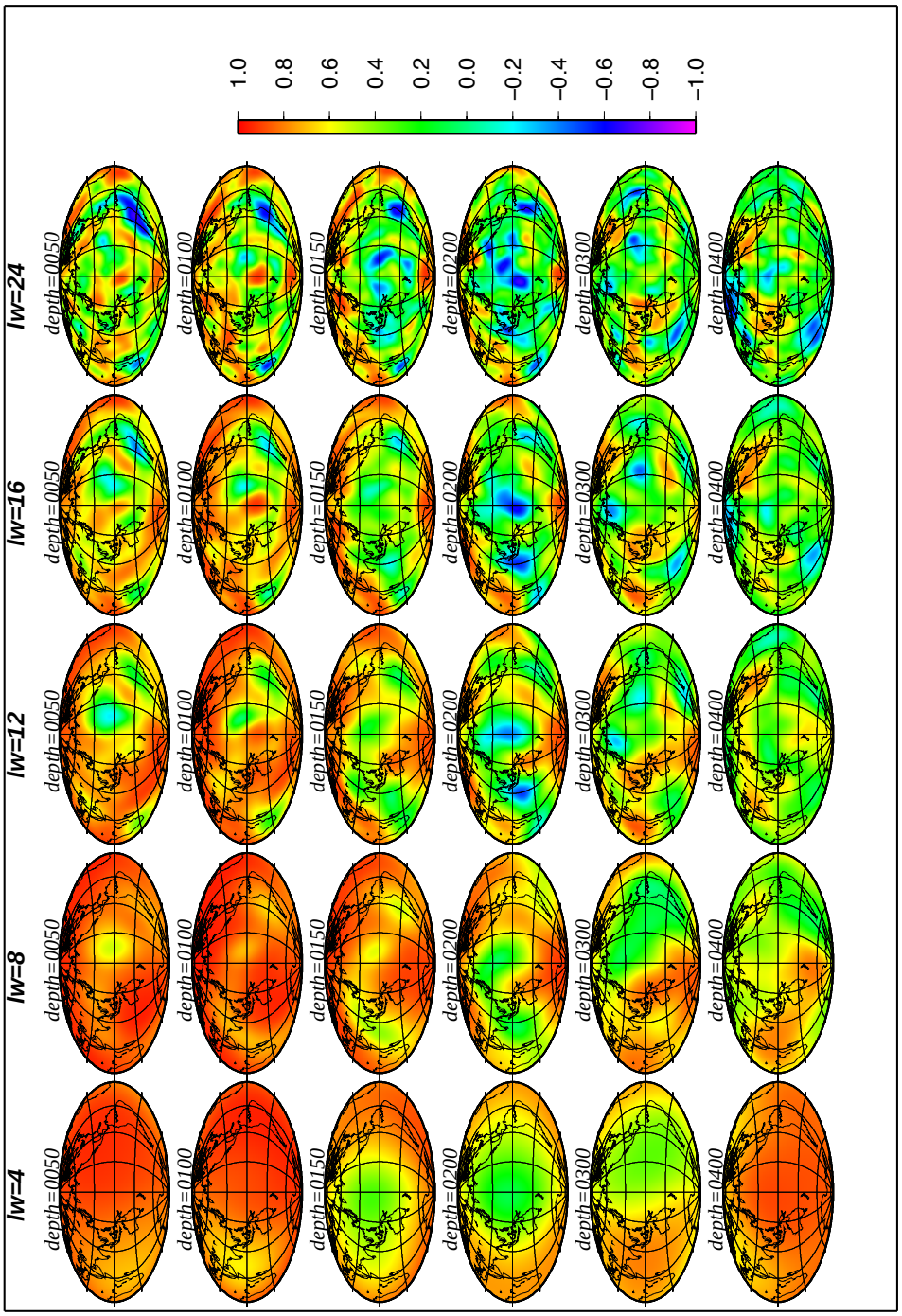


Figure 4.13: *Local correlation maps between S-wave tomographic model: lower depths.*

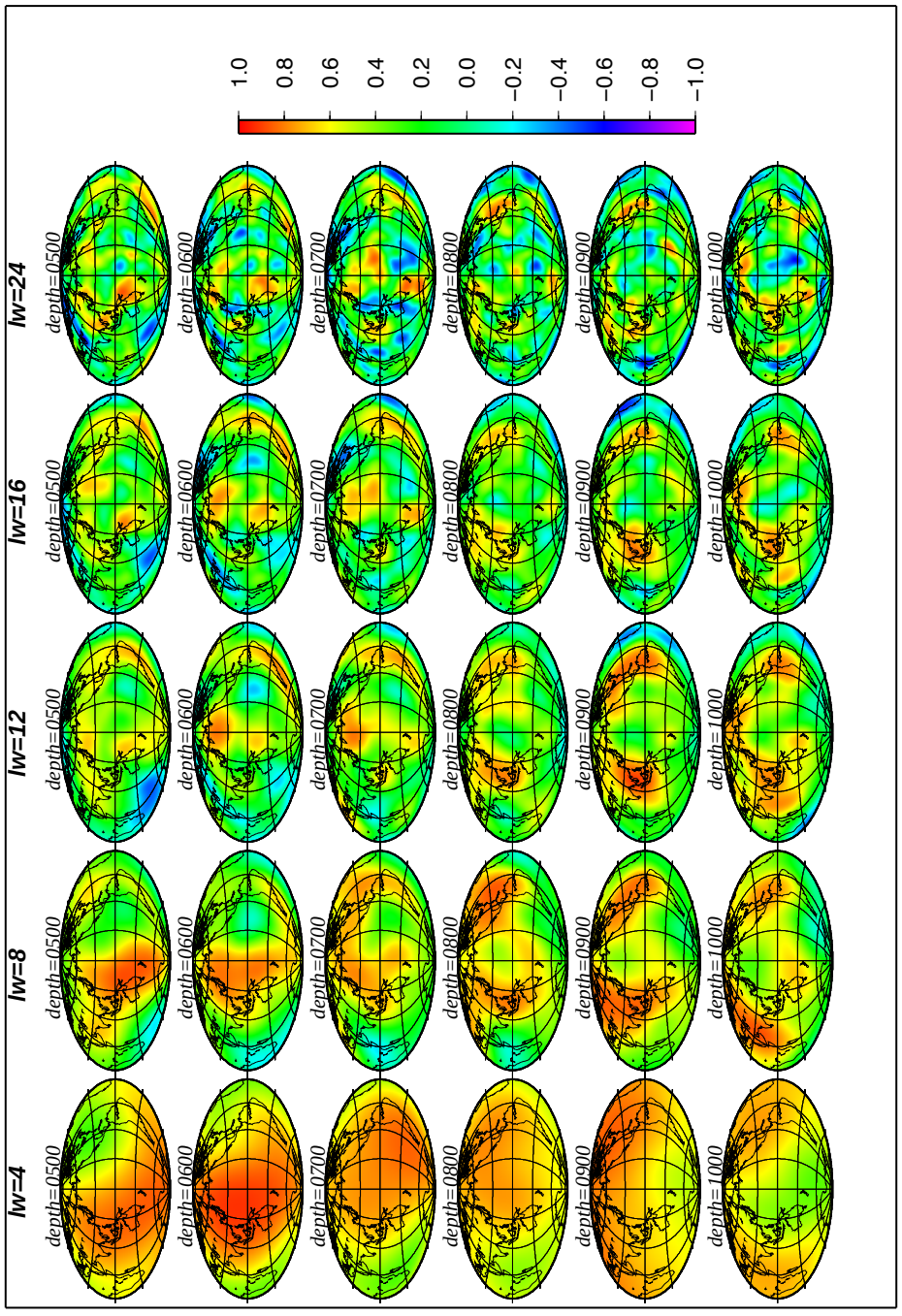


Figure 4.14: *Local correlation maps between S-wave tomographic model: deeper depths.*

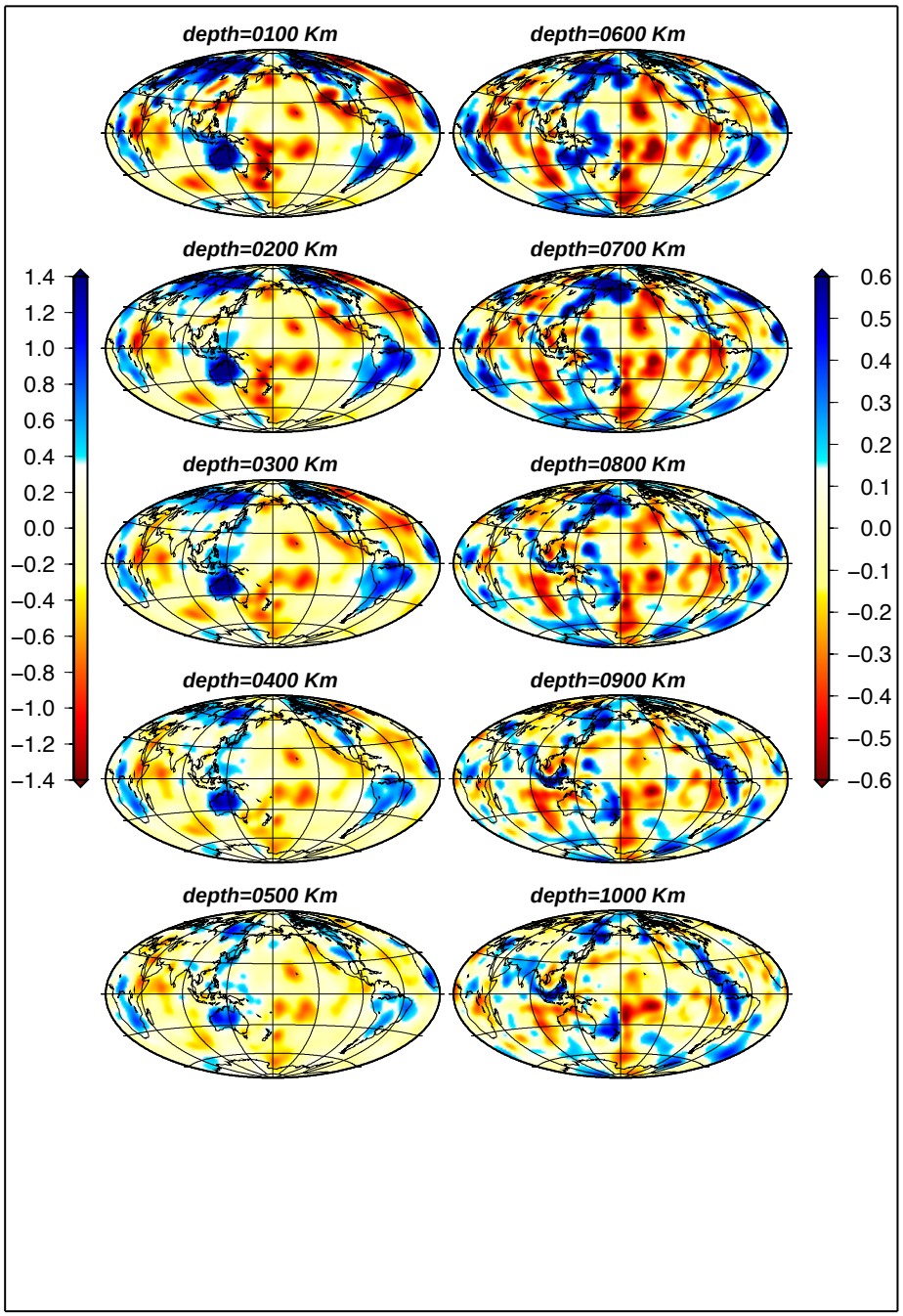


Figure 4.15: *P-wave tomographic model bdp00 (Antolik et al., [2]).*

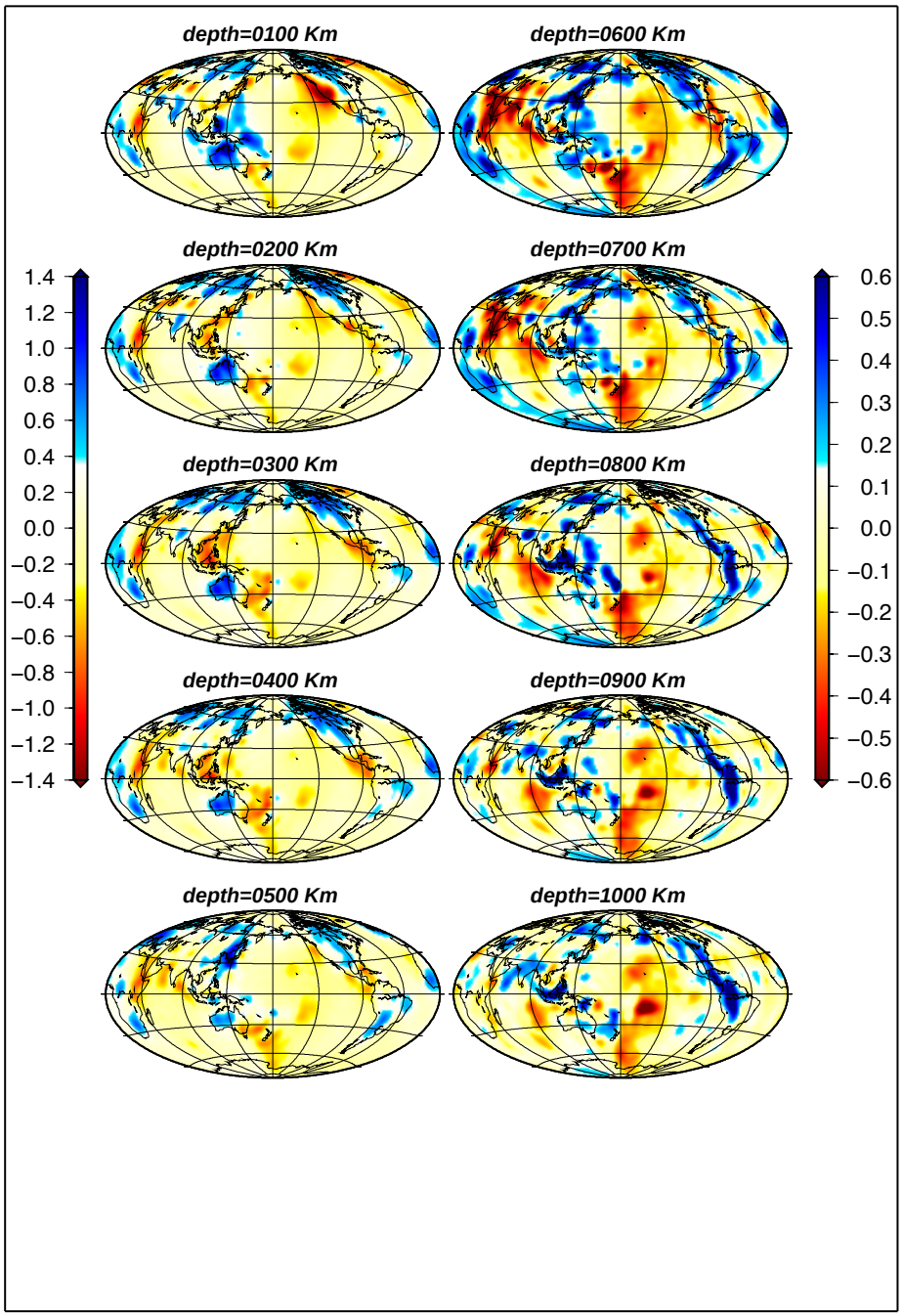


Figure 4.16: *P-wave tomographic model kh00p (Kárason & van der Hilst, [16]).*

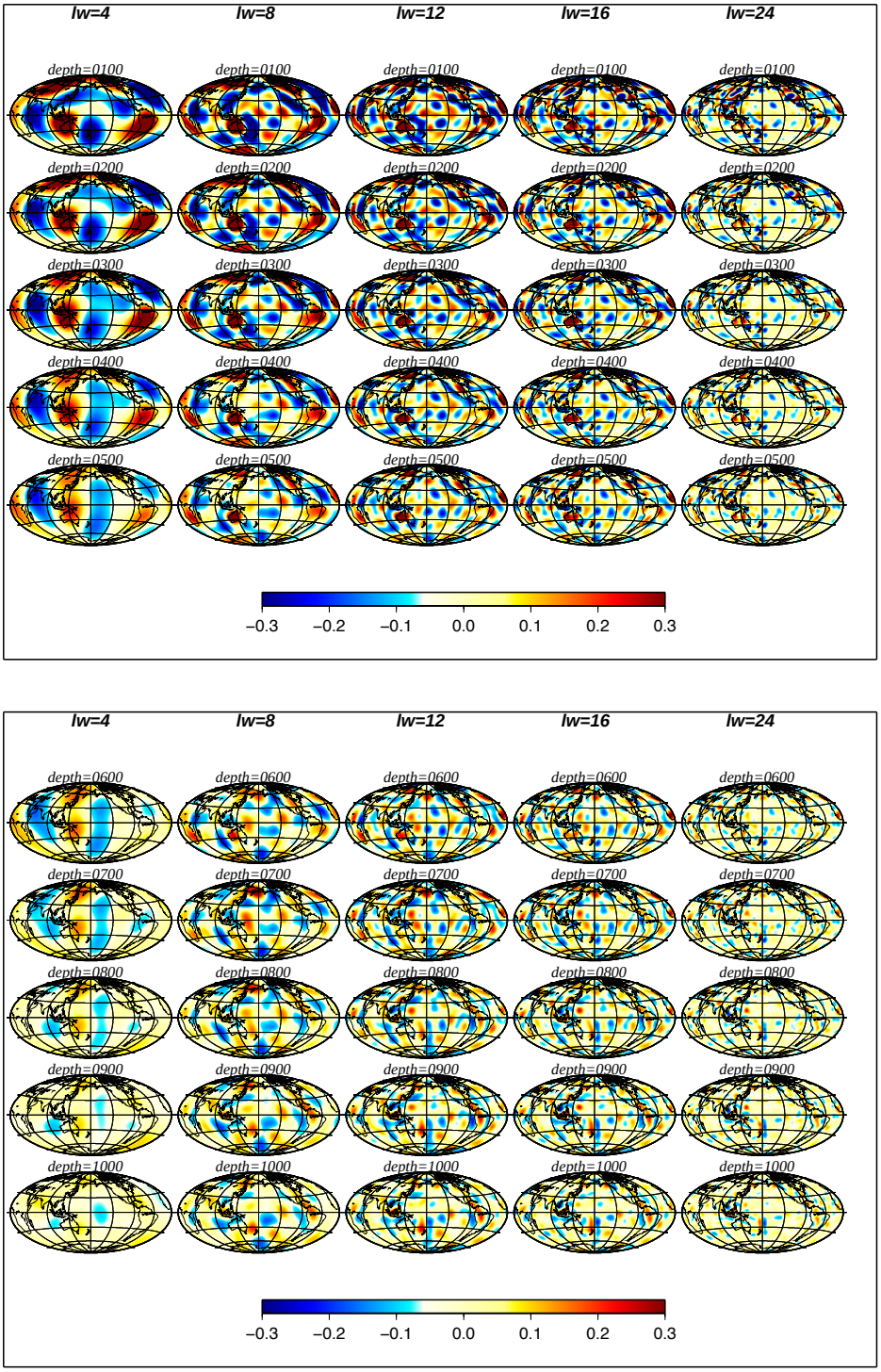


Figure 4.17: *Filtered version of original tomographic model: bdp00 .*

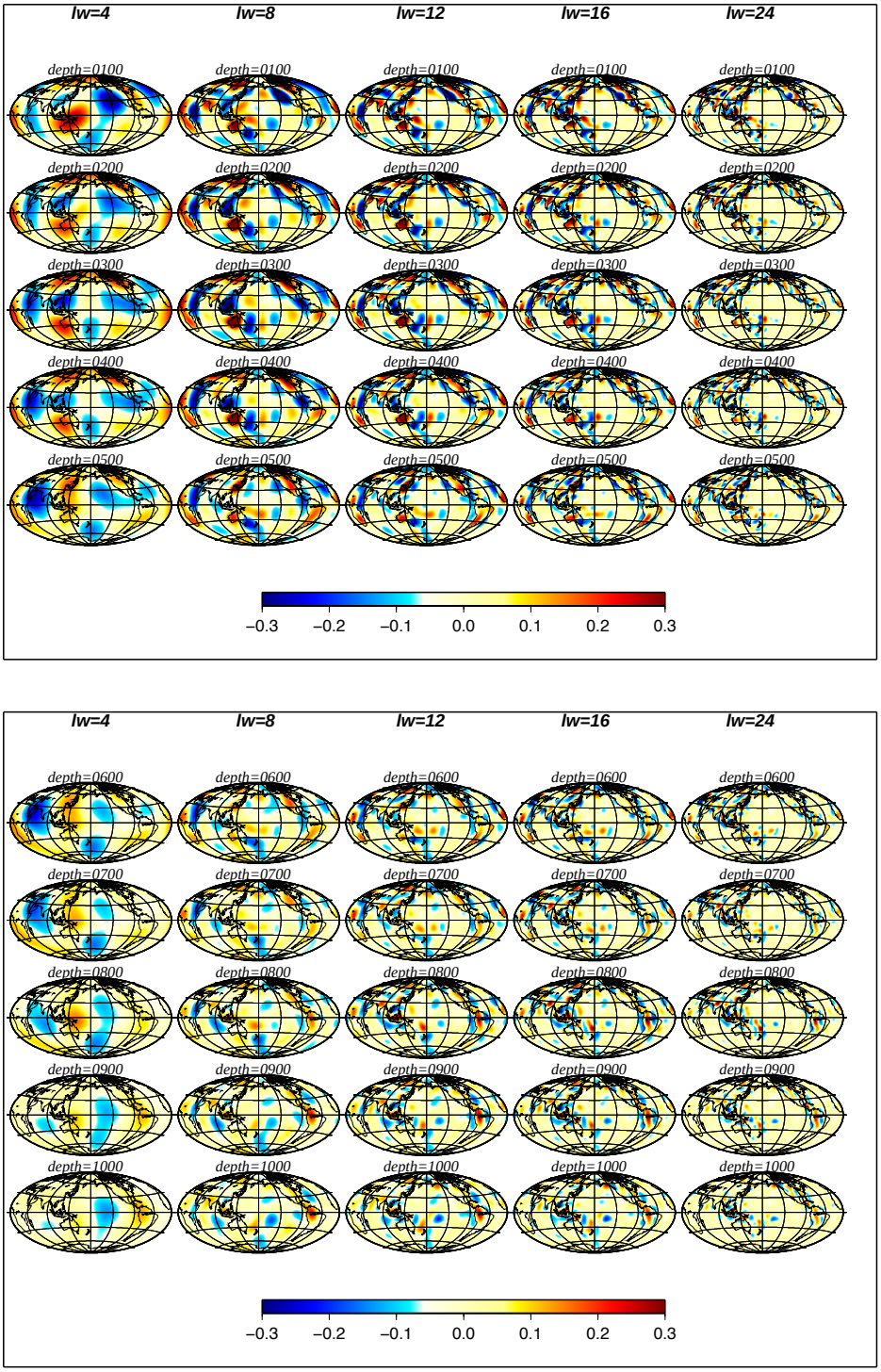


Figure 4.18: *Filtered version of original tomographic model: kh00p .*

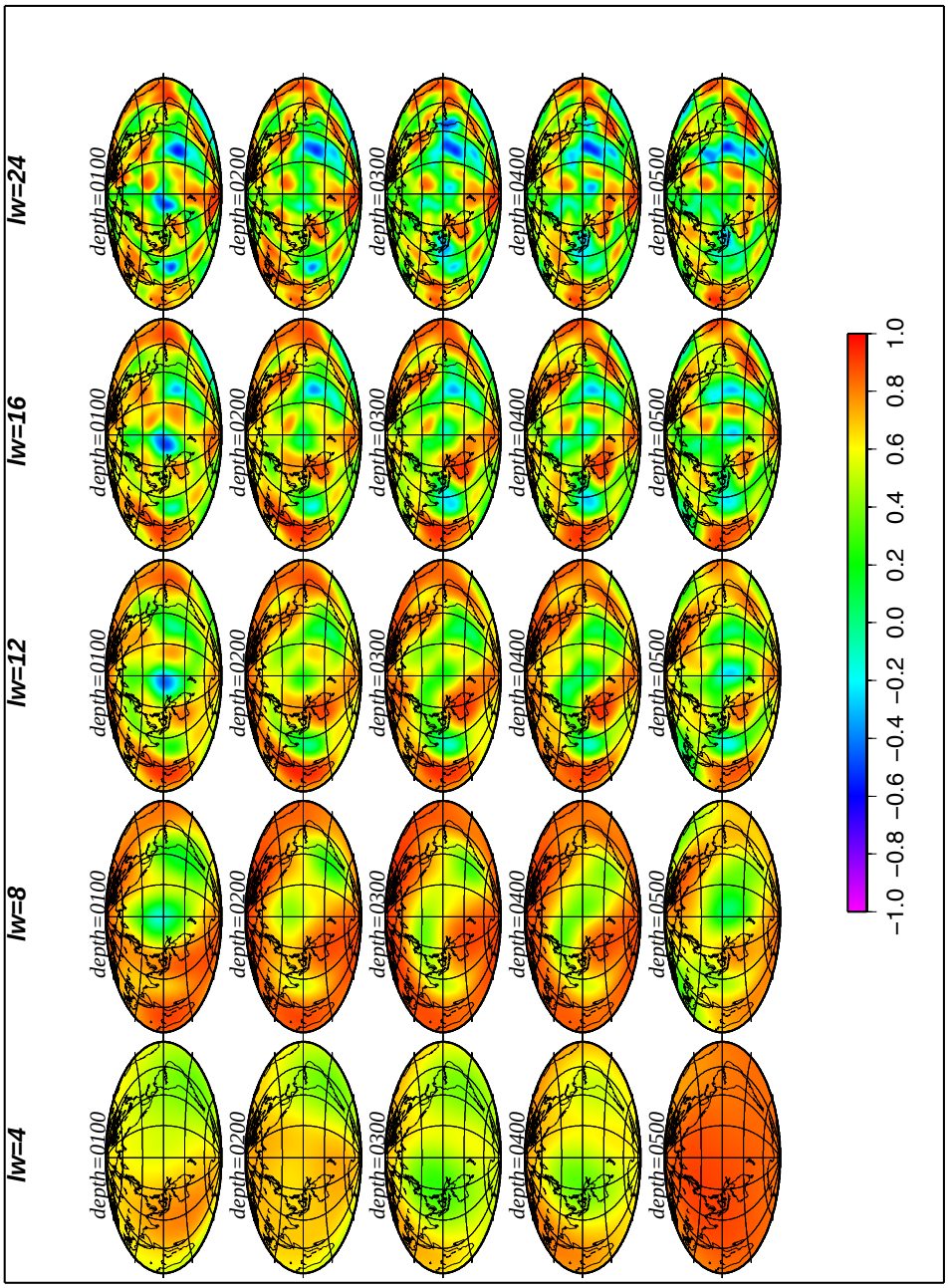


Figure 4.19: *Local correlation maps between P wave tomographic model: lower depths.*

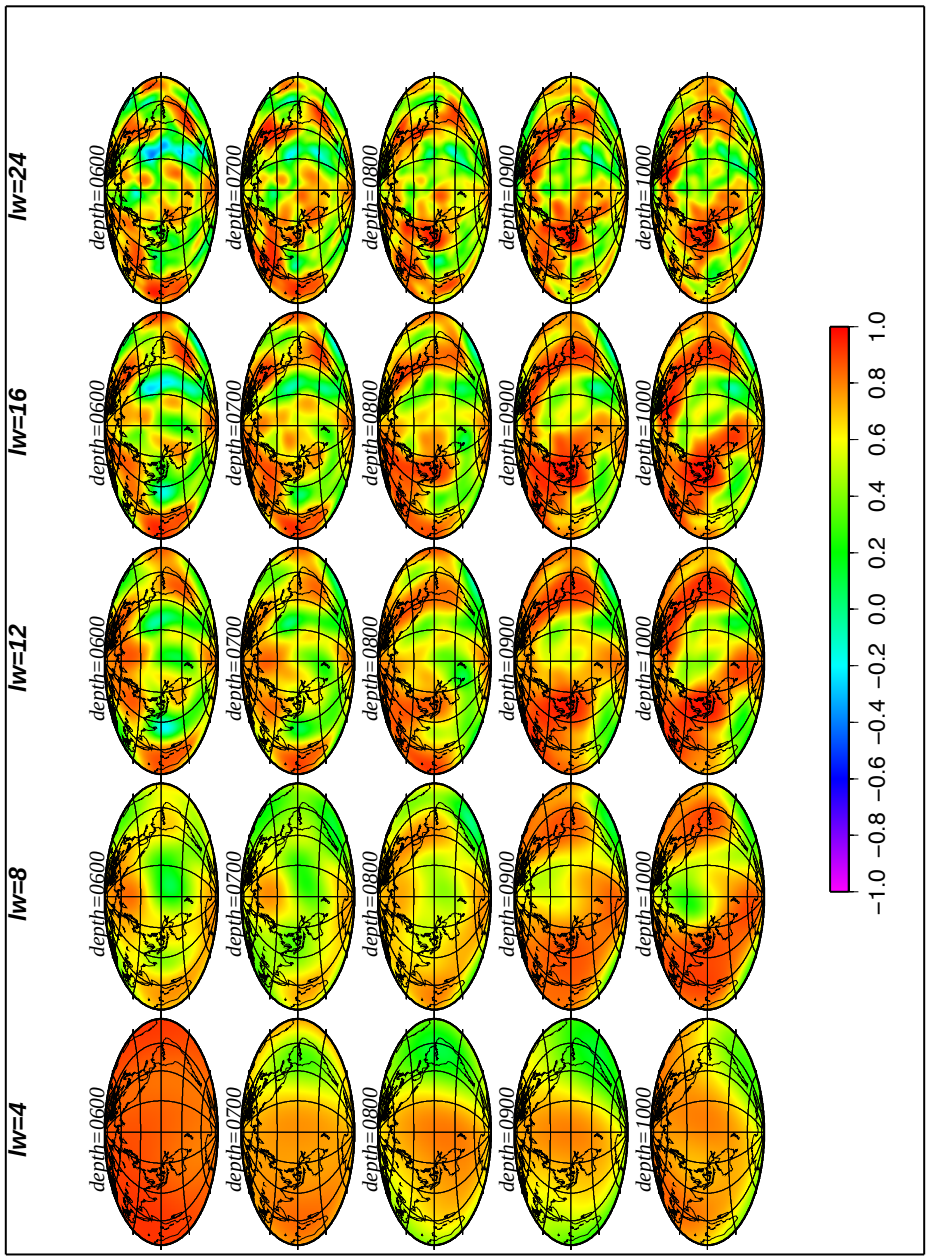


Figure 4.20: *Local correlation maps between P-wave tomographic model: deeper depths.*

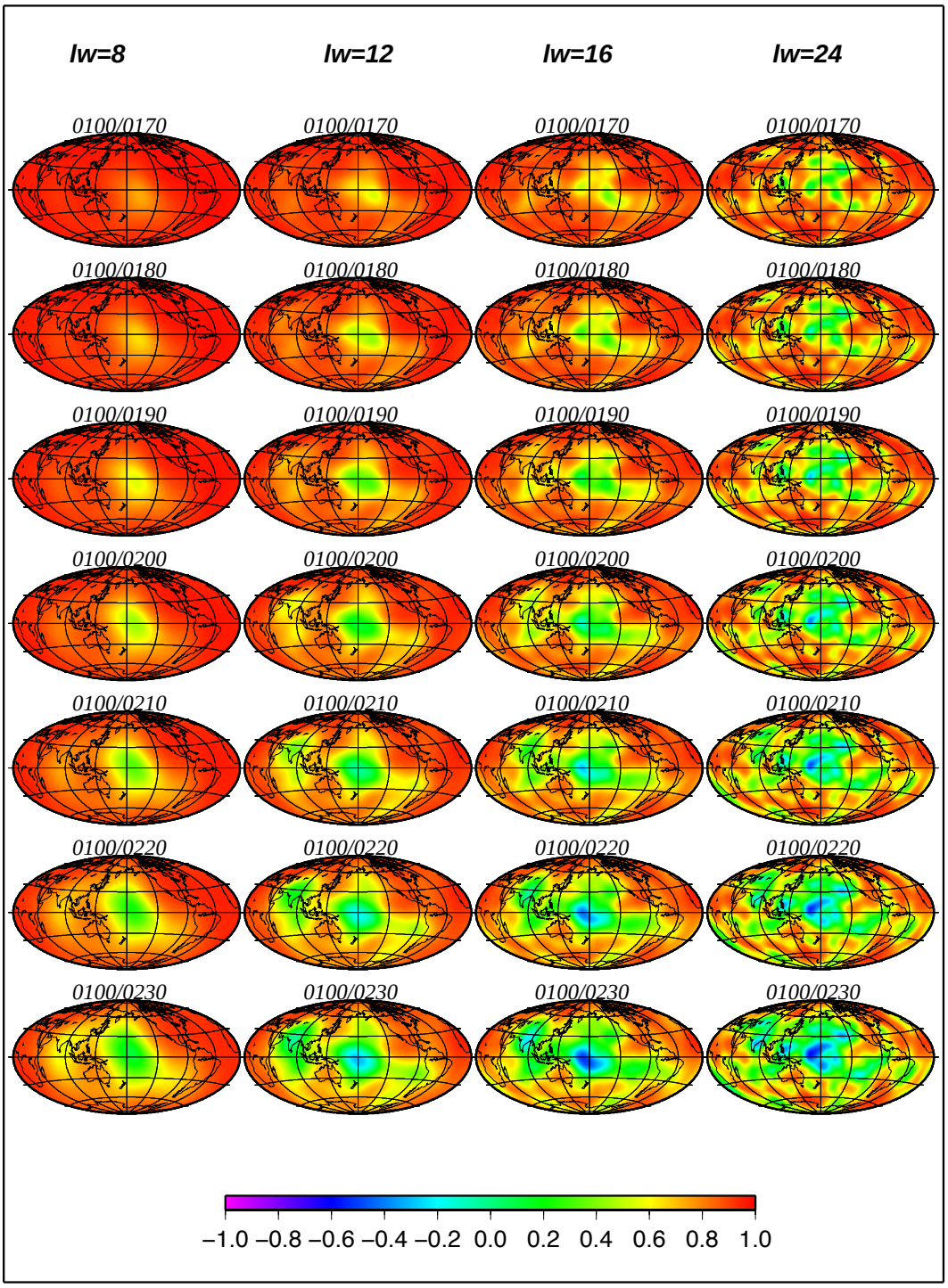


Figure 4.21: *Correlation between model at 100 km and the 7 lower layers.*

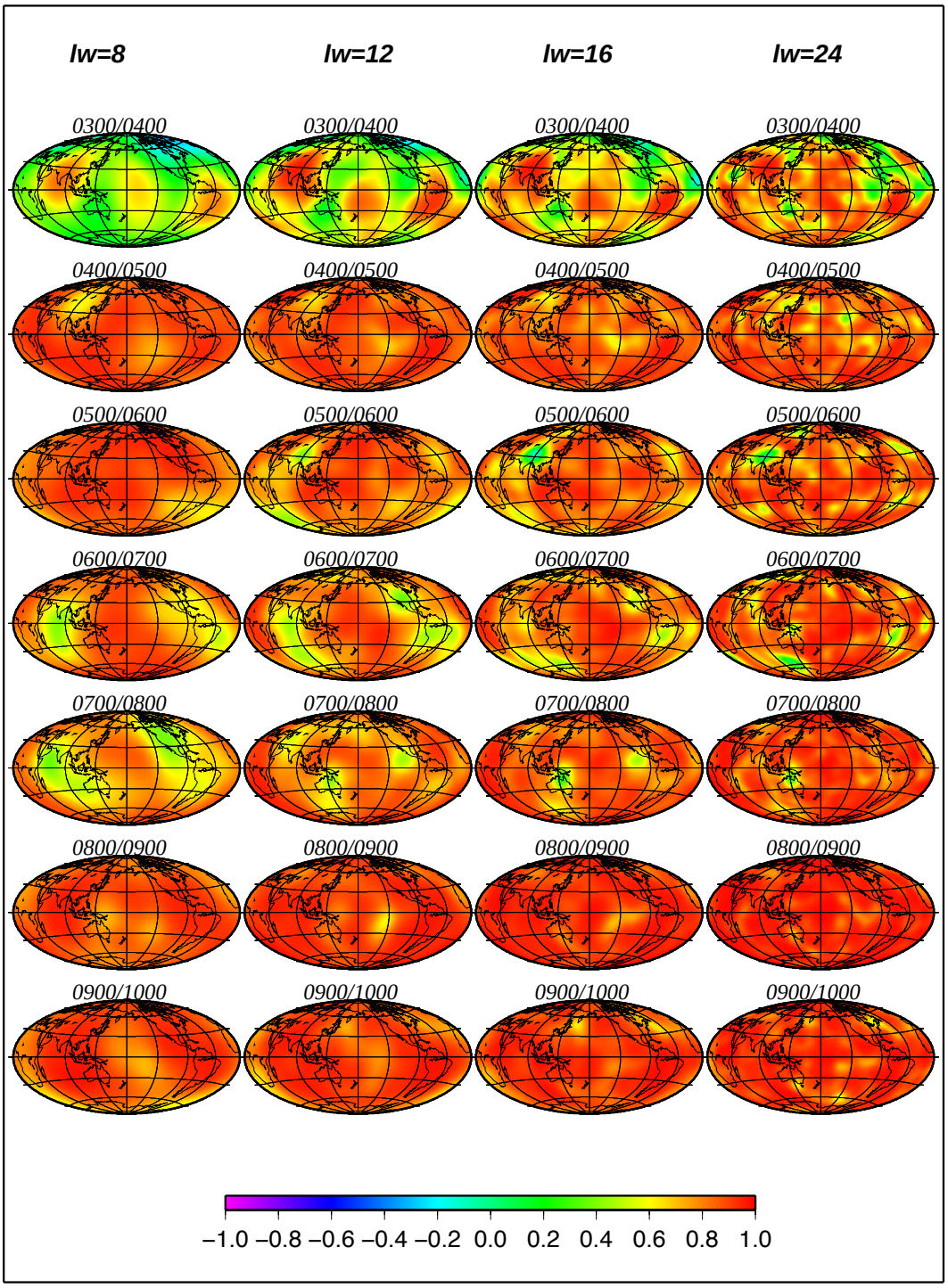


Figure 4.22: *Correlation between depths of model with a distance lower than 100 km .*

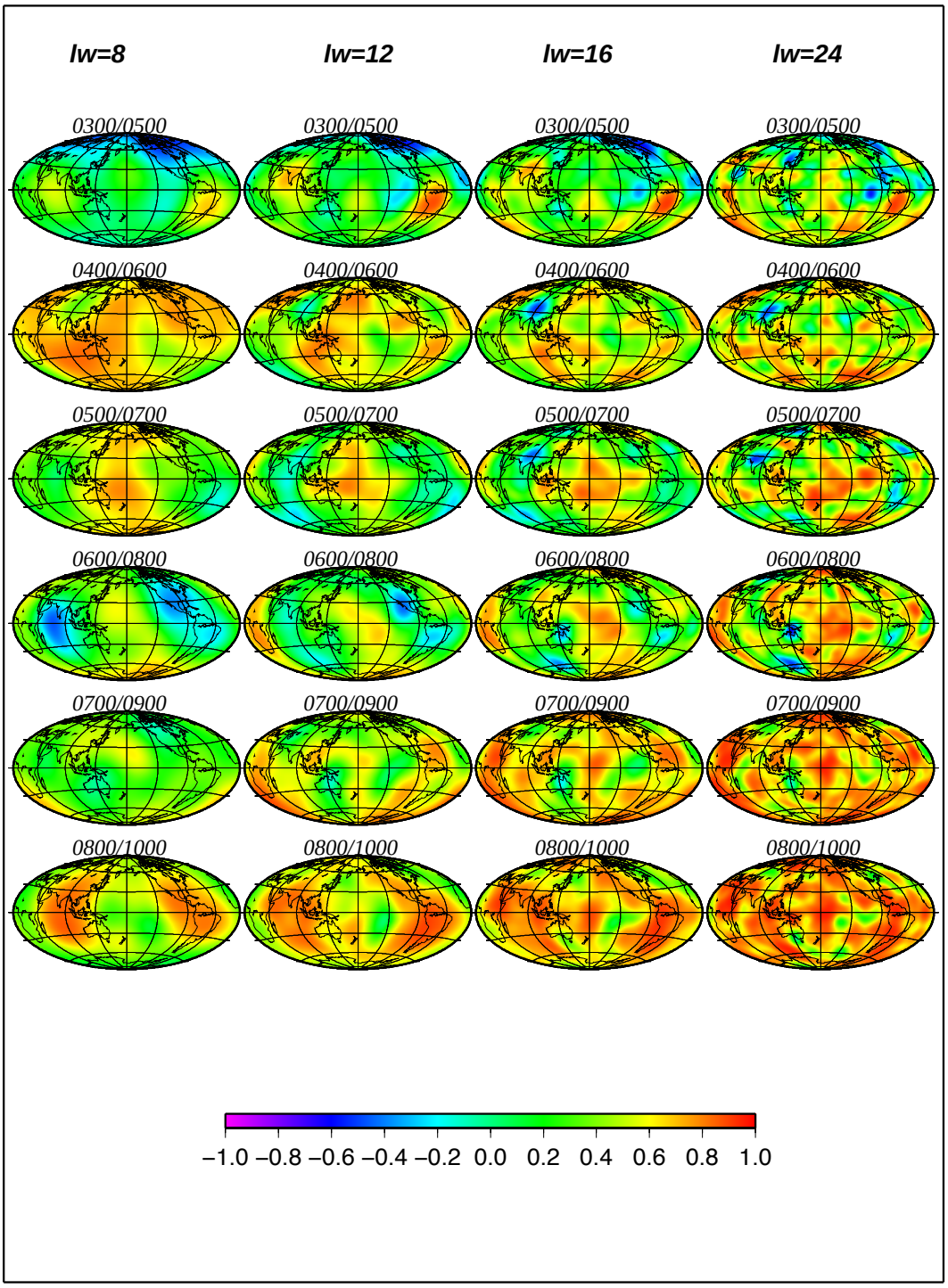


Figure 4.23: *Correlation between depths of model with a distance lower than 200 km .*

Conclusions

The wavelet analysis is an investigation tool widely applied in the field of seismic signal processing but not yet much used in the global tomography.
In this work we developed a multiresolution analysis, performed by mean of the wavelet transform with the aim of applying it to global seismic tomography. We faced this goal following to separate branches, to test the capabilities of this analysis in seismic tomography.
Firstly we solved the tomographic problem using a particular set of basis function, the spherical triangular Haar wavelets. The use of optimal spherical wavelets, which have compact support both in the spatial and wavenumber domains, provides a natural multi-scale representation of a target model and is free from *a priori* prejudices intrinsic to other base functions.
We tested the approach with some preliminary analysis (synthetic tests) and then applied it to real data of surface wave phase anomalies.
The inversion performed with wavelet analysis appeared as stable as that made with block paremeterization if we do not use "a priori" constraints, but its stability increase imposing a norm minimization, In the same condition, the pixel inversion usually require a further constrain on the roughness of the model. This difference in the damping is ascribed to a intrinsic multiresolution nature of wavelet analysis. In fact, when we decompose a signal using set, we, substantially, split it in two main components, one is the average and one is the details; details represent the difference of values of the model in a point and in its neighbourings, but this in fact the value that we minimize when we impose a roughness minimization. So when we solve the tomographic problem with wavelet approach we intrinsically "impose" a smoothing to our model.
Moreover the wavelet analysis assure us, as showed in *Chapter 3*, the possibility to reproduce the model at different resolution levels without the necessity to repeat the inversion, clearly reducing the running time.
In a second part we tested the potentialities of wavelet transform in the field of spectral analysis. In this field the wavelet analysis exhibit a remarkable power and usefulness respect to classical tools. Classically the spectral analysis of a signal, performed by means of the spherical harmonics, is able to provide one value both for energy content and for correlation. These values are substantially an averaged estimate which give us an idea of what is the energy associated to the model or how well two models are correlated. In *Chapter 4* we instead showed the innovative approach by means of wavelet transform which provide us not only the "how" but also the "where" of spectral contains. The tests, performed with synthetic and real data

demonstrated the great innovation of wavelet spectral analysis.

On the basis of results with this thesis we think that a future work could take advantage of multiresolution analysis capability of wavelet functions, developing a criterion to refine the mesh where is necessary, creating a “self-adaptive” parameterization which models itself to take in account the uneven distribution of source and recording stations across the globe.
Morover we are the project , for the future, to continue our comparison between tomographic models using wavelet spectral analysis.

Appendix A

The Lifting Scheme

A lifting is an elementary modification of perfect reconstruction filters, which is used to improve the wavelet properties, decomposing wavelet transforms into a set of stages.
In 1994, *Sweldens* ([32],[31]) has developed the **Lifting Scheme** for the construction on wavelet filters. The main feature of lifting scheme is that all constructions are derived in the *spatial domain*. This is in contrast to the traditional approach which relies heavily on the *frequency domain*. Staying in the spatial domain leads two major advantages.
First, it does not require the machinery of Fourier analysis as a prerequisite. Secondly, lifting leads to algorithms that can easily be generalized to complex geometric situations which typically occur in computer graphic.
The idea behind the lifting scheme is almost complex, so, to make the treatment as accessible as possible, we chose to begin with a simple example of a wavelet transform to introduce the basic ideas and then, eventually, introduce lifting in a more general mathematical background.

A.1 A trivial example: The Haar Wavelet

To make clear the lifting scheme we want in particular, make in evidence, the differences between the classical approach and this new building scheme.
Consider two number a and b and think of them as two neighboring samples of sequence. We propose a well-known, simple linear transform which replaces a and b by their *average* s and *difference* d:

$$\begin{aligned} s &= \frac{a+b}{2} \\ d &= b-a, \end{aligned} \tag{A.1}$$

With this choice we have not lost information because, inverting the 2×2 matrix, we can always recover a and b from given s and d by means of relations:

$$\begin{aligned} a &= a - d/2 \\ b &= s + d/2, \end{aligned} \tag{A.2}$$

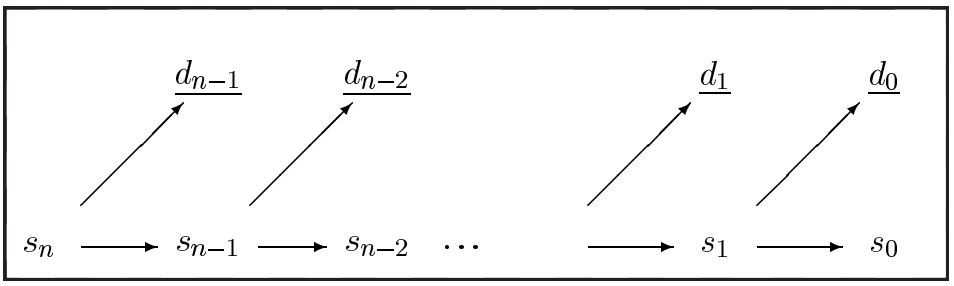


Figure A.1: *Structure of wavelet transform: recursively split into averages and differences.*

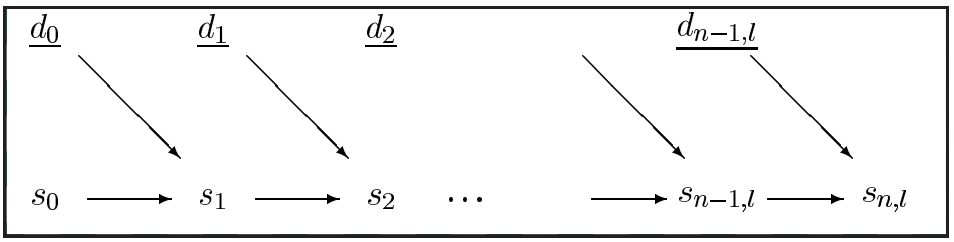


Figure A.2: *Structure of inverse wavelet transform: recursively merge averages and differences.*

This simple observation is the key behind the so-called *Haar wavelet transform*. Consider a signal s_n of 2^n sample values $s_{n,l}$:

$$s_n = \{s_{n,l} | 0 \leq l < 2^n\},$$

Apply the average and difference transform onto each pair $a = s_{2l}$ and $b = s_{2l+1}$. There are 2^{n-1} such pairs ($l = 0, \cdots, 2^{n-1}$); denote the results by $s_{n-1,l}$ and $d_{n-1,l}$: *average s* and *difference d*:

$$\begin{aligned} s_{n-1,l} &= \frac{s_{n,2l} + s_{n,2l+1}}{2} \\ d_{n-1,l} &= s_{n,2l+1} - s_{n,2l}, \end{aligned} \tag{A.3}$$

The input signal s_n, which has 2^n samples, is split in two signals: s_{n-1} with 2^{n-1} averages s_{n-1l} and d_{n-1} with 2^{n-1} differences $d_{n-1,l}$. Given the averages s_{n-1} and differences d_{n-1} one can recover the original signal. We can think of the averages s_{n-1} as a coarser resolution representation of the signal s_n and the differences d_{n-1} as the information needed to go from the coarser representation back to original signal.
We can apply the same transform to the coarser signal s_{n-1} itself. So we split the signal in a (yet) coarser signal s_{n-2} and another difference signal d_{n-2} where each of them contain 2^{n-2} samples. We can do this n times before we run out of samples (crf. Fig. (A.1) and (A.2)).
This is the *Haar transform*. We end up with n detail signals d_j and one signal s_0 on the very coarsest scale. By using the inverse transform we start from s_0 and d_j and obtain s_n again (crf. Fig.(A.2)).

The total number of coefficients after transform is 1 for s_0 plus 2^j for each d_j. This add up to:

$$1 + \sum_{j=0}^{n-1} 2^j = 2^n,$$

which is exactly the number of samples of the original signal. So the whole Haar transform is the result of application of a $N \times N$ matrix to the signal s_n. The cost of computing the transform is only proportional to N. It is the hierarchical structure of wavelet transform which allows switching to and from the wavelet representation in $O(N)$ time.

A.2 Haar and a first approach to Lifting

Now we want to look at a different approach at the Haar transform seen in the previous section. The novelty consists in the way used to compute the difference and average of two number a and b. A first important assumption is that we want work without using auxiliary memory location, by overwritting the location that hold a and b with the value s and d. This can not immediately done with the formulas (A.1), which give us a wrong result. We therefore suggest an implementation in two steps.

First we only compute the difference:

$$d = b - a,$$

and store it in the location for b. As we now lost the value of b we next use a and the newly computed difference d to find the average as :

$$s = a + d/2.$$

This substantially give us the same result because $a + d/2 = a + (b - a)/2 = (a + b)/2$. The advantage of the splitting into steps is that can overwrite b with d and a with s, requiring no auxiliary memory. A *C-like implementation* is given by:

$$b- = a; \quad a+ = b/2;$$

after which b contains the differences and a the average. We can immediately find the inverse without formally solving 2×2 system (A.3): simply run the above code backwards (i.e., change order and flip the signs).

This particular scheme of writing a transform is a first, simple instance of the *lifting scheme*.

A.3 General theory on the Lifting Scheme

Let see now the more general structure of the lifting scheme, describing it in more detail.

Consider a signal s_j with 2^j samples which we want to transform into a coarser signal s_{j-1} and detail signal d_{j-1}. A typical case of a wavelet transform built through lifting consists of three steps: split, predict and update. Let us discuss each stage in mode detail.

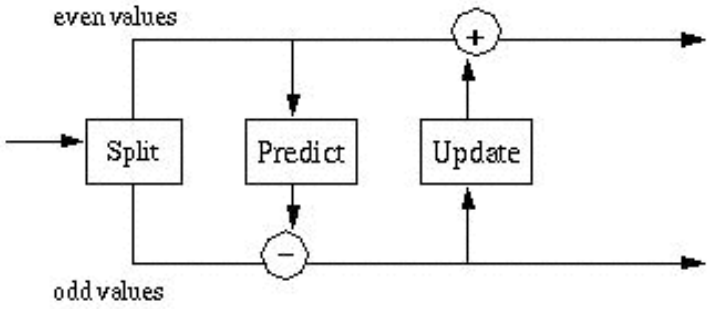


Figure A.3: *Lifting scheme forward wavelet transform: first compute the detail as the failure of a prediction rule, then use that detail in an update rule to compute the coarse signal(from web site [42]).*

- *SPLIT STEP* : This step *splits* the signal into two disjoint sets of samples. In our case one group consists of the even indexed samples s_{2l} an the other group consists of the odd indexed samples s_{2l+1}. Each group contains half as many samples as the original signal. The splitting into even and odd is called *Lazy wavelet transform*. we can built an operator so that:

$$(even_{j-1}, odd_{j-1}) := Split(s_j).$$

Remember that is the previous example a was an even sample while b was an odd sample. - *PREDICT STEP* : If the signal has a local correlation structure, the even and odd subsets will be highly correlated. In other words given one of the two sets, it should be possible predict the other one with reasonable accuracy. We generally use the even set to predict the odd one. An odd sample $s_{j,2l+1}$, in the Haar case, use its left neighboring even sample $s_{j,2l}$ as its predictor. The detail $d_{j-1,l}$ will be the difference between the odd sample and its prediction :

$$d_{j-1,l} = s_{j,2l+1} - s_{j,2l},$$

which defines an operator P such that:

$$d_{j-1,l} = odd_{j-1} - P(even_{j-1}.$$

- *UPDATE STEP* : One of the key properties of the coarser signals is that they have the same average value as the original signal,i.e., the quantity:

$$S = 2^{-j} \sum_{l=0}^{2^j-1} s_{j,l},$$

is independent of j. The update stage ensure this by letting

$$s_{j-1,l} = s_{j,2l} + d_{j-1,l}/2.$$

So we can define an other operator U of the form:

$$s_{j-1} = even_{j-1} + U(d_{j-1}).$$

All this can be computed in-place: the even locations can be overwritten with the averages and the odd ones with the details.
An abstract implementation is given by:
•$(even_{j-1}, odd_{j-1}) := Split(s_j)$
•$odd_{j-1,l} -= P(even_{j-1}$
•$even_{j-1} += U(d_{j-1})$
These three steps are showed schematically in Fig.(A.3) and Fig.(A.4) .

Haar Transform

Lifting version

- split into even and odd

$$(even_{j-1}, odd_{j-1}) := Split(s_j)$$

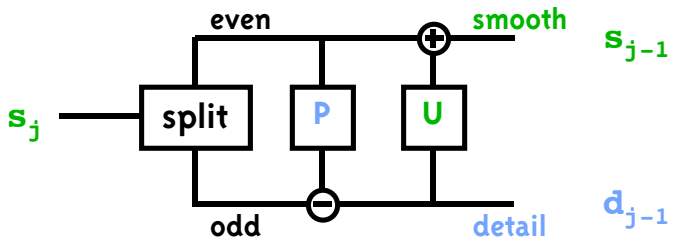
- predict and store difference: detail coefficient

$$d_{j-1} = odd_{j-1} - even_{j-1}$$

- update even with detail: smooth coefficient

$$s_{j-1} = even_{j-1} + d_{j-1}/2$$

Haar Transform



$$d_{j-1} = odd_{j-1} - P(even_{j-1})$$

$$s_{j-1} = even_{j-1} + U(d_{j-1})$$

Figure A.4: *Lifting scheme: a schematic illustration of lifting scheme Haar transform from Sweldens's note [32]*

Bibliography

[1] Addison,P.S. , *The Illustrated Wavelet Transform Handbook: Applications in Science, Engineering, Medicine and Finance* ,IOP Publishing Ltd, 2002.

[2] Antolik, M., G. Ekström, A. M. Dziewonski, Y. J. Gu, J.-F. Pan & L. Boschi, *A new joint p and s velocity model of the mantle parameterized in cubic B-splines*, paper presented at *22nd Annual DoD/DOE Seismic Research Symposium: Planning for Verification of and Compliance with the Comprehensive Nuclear-test-Ban Treaty (CTBT)*, New Orleans, La., 2001.

[3] Audet, P. & J. C Mareschal, *Wavelet analysis of the coherence between Bouguer gravity and topography:application to the elastic thikness anisotropy in the Canadian Shield*, Geophys. J. Int. **168**, 287-298, 2007.

[4] Bassin, C., G. Laske and G. Masters, *The current limits of resolution for surface wave tomography in North America*, EOS Trans. AGU, **81**, p.F897, 2000.

[5] Becker, T. & L. Boschi, *A Comparison of Tomographic and Geodynamic Mantle Models*, Geochemestry, Geophysics, Geosystems, **3**, 2002.

[6] Bergeron, S.Y., D.A. Yuen & A.P. Vincent, *Looking at the insiade of the Earth with 3-D wavelets: apair of new glasses for Geoscientists*, Vis Geosci, textbf5:3,1-14, 2000.

[7] Bonneau, G. P., *Optimal Triangular Haar Bases for Spherical Data*, IEEE Visualization'99, IEEE, October, 1999.

[8] Boschi, L., *Applications of linear inverse theory in modern global seismology*, Ph.D.thesis, Harvard University,Cambridge, 2001.

[9] Boschi, L. & A. M. Dziewonski, *High- and Low- Resolution Images of the Earth's Mantle - Implications of different approaches to tomographic modeling*, J. Geophys. Res., **104**, 25,567 - 25,594, 1999.

[10] Boschi, L., and G. Ekström, *New images of the Earths upper mantle from measurements of surface-wave phase velocity anomalies*, J. Geophys. Res., **107**, 10.1029/2000JB000059, 2002. **104**, 25,567 - 25,594, 1999.

[11] Carannante, S. & L. Boschi, *Databases of Surface Wave Dispersion*, Annals of Geophysics, **48**, 945-955, December 2005.

[12] Chiao, L & B. Kuo, *Multiscale Seismic Tomography*, Geophys. J. Int. , **145**, 517-527, 2001.

[13] Ekström, G. , J. Tromp & E.W.F. Larson, *Measurements and global models of surface wave propagation*, J. Geophys. Res., **102**, 8137-8157, 1997.

[14] van Heijst, H. J. & J. Woodhouse, *Global high-resolution phase distributions of overtone and fundamental-mode surface waves determined by mode branch stripping*, Geophys. J. Int., **137**, 601-620, 1999.

[15] van der Hilst, R.D., Widiyantoro,S., & Engdahl, E. R., *Evidence of Deep Mantle Circulation from Global Tomography*, Nature, **386**, 578-584, 1997.

[16] Kárason, H. & R.D. van der Hilst, *Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP,Pdiff)*, J. Geophys. Res., **106**, 6569-6588, 2001.

[17] Kido, M., D. A. Yuen & A. P. Vincent, *Continuous wavelet - like filter for a spherical surface and its application to localized admittance function on Mars*, PEPI , **135**, 1-16, 2003.

[18] Kumar, P. & E. Foufoula-Georgiou, *Wavelet Analysis for Geophysical Applications*, Reviews of Geophysics, **35**, 4, 385 - 412, 1997.

[19] Mallat, S., *A wavelet Tour of Signal Processing*, Secon edition, Academic Press, **1999**

[20] Marquering, H., F. A. Dahlen & G. Nolet, *Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox*, Geophys. J. Int., **137**, 805-815, 1999.

[21] Nielson, G. M., Il-Hong Jung & Junwon Sung, *Haar wavelets over triangular domains with applications to multiresolution models for flow over a sphere*. In IEEE Visualization 97, 143150. IEEE, November 1997

[22] Nolet, G., *Seismic Tomography with applications in Global Seismology and Exploration Geophysics*, **1987**

[23] Paige, C.C & M.A. Saunders, *LSQR: an algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Soft., **8**, 43-71, 1982.

[24] Panning, M. & B. Romanowicz, *A three-dimensional radially anisotropic model of shear velocity in the whole mantle*, Geophys. J. Int., **167**, 361-379, 2006.

[25] Piromallo, C. , A. P. Vincent, D.a. Yuen & A. Morelli, *Dynamics of the transition zone under Europe inferred from wavelet cross-spectra of seismic tomography*, PEPI, **125**, 125-139, 2001.

[26] Piromallo, C. & A. Morelli, *P-wave tomography of the mantle under the Alpine - Mediterranean area*, J. Geophys. Res., **108**, doi:10.1029/2002JB001725, 2003.

[27] Ritsema, J. & H.J. van Heijst, *Seismic imaging of structural heterogeneity in Earth's mantle: Evidence for large-scale mantle flow*, Sci. Progr., **83**, 243- 259, 2000.

[28] Sambridge, M. & R. Faletič, *Adaptive whole Earth tomography*, Geochem., Geophys., Geosyst., **4**, No. 3, doi:10.1029/2001GC000213, 5 March 2003.

[29] Schröder, P. & W. Sweldens, *Spherical Wavelets: Efficiently Representing Functions on the Sphere*, Computer Graphics Proceedings (SIGGRAPH '95), 161-172, 1995.

[30] Stein, S. & M. Wysession, *An introduction to seismology, earthquakes, and earth structure*, Blackwell Publishing, 2003.

[31] Sweldens, W. ,*The Lifting Scheme, a construction of second generation wavelets*, Siam J. Math. Anal, **29**, 2, 511-546, 1997.

[32] Sweldens, W. & P. Schröder, *Building Your own Wavelets at Home* from *"Wavelets in Computer Graphics"*, ACM SIGGRAPH Course Notes, 1996.

[33] Swedens. W., *Wavelets and the lifting scheme: A 5 minute tour*, Zeitschrift für Angewandte Mathematik und Mechanik, **76** (Suppl. 2), 41-44, 1996.

[34] Tarantola, A., *Inverse Problem Theory and Model Parameter Estimation*, SIAM, 2005.

[35] Valens,C. , *A Really Friendly Guide to Wavelets*

[36] Vecsey L., C.A. Hier Majumder & D. A. Yuen, *Multiresolution tectonic features over the Earth inferred from wavelet transformed geoid*, Vis. Geosci. **8**, 26-44, 2003.

[37] Walnut, D. F., *An introduction to Wavelet Analysis*, Birkhäuser, Boston-Basel-Berlin, 2002.

[38] Wang, Z. & F. A. Dahlen, *Validity of surface - wave ray theory on a laterally heterogeneous Earth*, Geophys. J. Int., **123**, 757-773, 1995.

Web site

[39] $http://www.amara.com/current/wavelet.html$

[40] $https://www-ceb.bo.infn.it/dantone/slides\_pub.htm$

[41] $http://cnx.org/content$

[42] $http://www.bearcave.com/misl/misl\_tech/wavelets$

[43] $http://en.wikipedia.org/wiki/$