Stochastic Tomography: Characterizing Small-scale Heterogeneity in Earth Using Coherence Functions

Yiteng Tian
University of Connecticut - Storrs, yiteng.tian@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation
https://opencommons.uconn.edu/dissertations/1794
Stochastic Tomography: Characterizing Small-scale Heterogeneity in Earth Using Coherence Functions

Yiteng Tian, PhD

University of Connecticut, 2018

This thesis presents a thorough study of the stochastic tomography technique, from theory to numerical validation and application in characterizing small-scale heterogeneity in Earth’s mantle using statistical approach. Fluctuations in amplitude and travel time of teleseismic P waves, measured by amplitude and phase coherences beneath elements of EarthScope seismic array, are used to invert for the heterogeneity spectrum of P velocity in a 1000 km thick region of the upper mantle beneath the array. Best fits to joint transverse coherence functions require a depth-dependent heterogeneity spectrum, with peaks in narrow depth ranges that agree well with the predictions for a temperature derivative of velocity that includes the effects of chemical and phase variations expected for standard models of the silicate mineral assemblage of the upper mantle. The results confirm the existence of significant chemical as well as thermal contributions to observed upper mantle heterogeneity at spatial scales between 50 km to 300 km.
Stochastic Tomography: Characterizing Small-scale Heterogeneity in Earth Using Coherence Functions

By Yiteng Tian

B.S., Shandong University, 2012

M.S., University of Connecticut, 2014

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the University of Connecticut

2018
Copyright by

Yiteng Tian

2018
Stochastic Tomography: Characterizing Small-scale Heterogeneity in Earth Using Coherence Functions

Presented by
Yiteng Tian, M.S. Physics, B.S. Physics

Major Advisor
Vernon F. Cormier

Associate Advisor
Lanbo Liu

Associate Advisor
Richard T. Jones

University of Connecticut
2018
Acknowledgements

First and foremost, I would like to thank my advisor and dear friend, Prof. Vernon Cormier, for his kind patience and enormous help. He has had such an influence on me, both academically and personally. I would like to thank my kindly associate advisor, Prof. Lanbo Liu, for his guidance and great support. The brilliant professors at the physics department, Prof. Richard Jones, Prof. Niloy Dutta and Prof. Winthrop Smith being my advisory committee, Prof. Gerald Dunne being my early academic mentor, this work would not have been done without their help. I also would like to thank the friendly staffs, Heather Osborne, Micki Bellamy and Ann Marie Carroll, for helping and supporting my teaching practice and graduate study.

I would like to thank my peer fellows, Freddy Obrecht, John Mangeri, Shiqin Yin, Brad Clarke, Robert Dabrowski, Zhiwei Zhang, Cheng Tu and Xiang Zhang, for creating such an inspiring community at UConn Physics. I would like to thank my lovely significant other, Fei Huang, and my friends who shared a UConn life with me, Zhitong Fei, Ran Miao, Zhu Luo and Luoyu Zhou. It truly has been a thrilling and emotional adventure.

Lastly, I would like to express the deepest appreciation to my wonderful family, my grandfather having encouraged and guided me to the world of physics, my parents having been standing behind me since always, and all the delightful fruits on my family tree, without whose support I would never become who I am.
To my family
# Table of Contents

**Chapter 1. Introduction**.................................................................1

**Chapter 2. Random Media: a Representation of Small-scale Heterogeneity**.................................................................9

2.1 Random Media.........................................................................................9
2.2 The Generalized Form of Power Spectrum.................................................11
2.3 Numerical Realization of Random Media from Power Spectrum..............14
2.4 Calculation of Power Spectrum from Tomography Model.........................17

**Chapter 3. Inverse Modeling Using Stochastic Tomography:**

**Methodology and Numerical Validation**.................................19

3.1 Theoretical Coherence Functions..........................................................19
3.2 Discretization and Implementation..........................................................23
3.3 Numerical Analysis..................................................................................27
3.4. Predicted Coherence Functions for Different Spectrum.........................33
  3.4.1 Transverse Coherence Functions: Different Types of Spectrum..........36
  3.4.2 Transverse Coherence Functions: Effect of Variant Parameters........39
  3.4.3 Transverse Coherence Functions: Add Depth Dependency...............40
  3.4.4 Angular Coherence Functions............................................................42
3.4.5 Joint Transverse Angular Coherence Functions.................................43

Chapter 4. Measurement and Mining on Seismograms..............47

4.1 Data Source........................................................................................................47
4.2 Event Selection and Preprocess.................................................................49
4.3 Synthetic Reference Seismograms.............................................................52
4.4 Measurements of Fluctuations and Calculation of Coherence...............55
4.5 Cross Coherence of Different Earthquake Groups.................................58

Chapter 5. Stochastic Tomography: Inversion of Coherence Functions.................................................61

5.1 Inversion Scheme with Re-sampling..........................................................61
5.2 Single Layer Assumption and Inversion with L-2 Regularization.............66
5.3 Inversion Regularized by a Generalized Function.....................................69
5.4 Inversion Adding Depth Dependence.........................................................72

Chapter 6. Analysis and Discussion.................................................................75

6.1 Discussion on result with single layer assumption...............................75
6.2 Discussion on Depth dependence.............................................................77

Chapter 7. Conclusion.......................................................................................87

Bibliography........................................................................................................91
Earth models describe the material and mechanical structures of Earth’s interior including the spatial variation of elastic properties, density, and temperatures. In an excellent first order approximation, these properties can be described as functions only of radius. An example of such 1-D global Earth models are PREM (Dziewonski and Anderson, 1981), IASP91 (Kennett, 1991), and AK135 (Kennett et al., 1995). Higher order approximations seek to describe 3-D heterogeneities as perturbations to these 1-D models. Knowledge of these 3-D heterogeneities can help us understand the dynamics and evolution of our planet better. Earth’s interior can be probed by several means, including the study of borehole samples of rocks (e.g. Labo, 1987), the study of chemical compositions from meteorites that are similar to Earth (e.g. Smith, 2003), and the modeling of elastic waves interacting with Earth’s heterogeneities, which will be our main topic in this thesis.

The structure and state of Earth’s interior can be characterized by variations in its seismic velocities. The negative temperature derivative of seismic velocities, for example, has been used to conclude that subducting slabs at convergent tectonic plates are cold due to their high seismic velocities (e.g. Stern, 2002), and the interior of buoyantly rising plumes in the mantle are hot due to their low seismic velocities (e.g. Nataf, 2000).

To retrieve and image Earth’s 3-D velocity structure, seismic tomography has become a
powerful technique by inverting the travel time of body waves ray-traced through Earth (Luo and T., 1991; Nemeth et al., 1997; Pratt and Goulty, 1991; Schuster and A., 1993; Zhu and McMechan, 1989) or by inverting full seismic waveforms (Lailly, 1984; Pica et al., 1990; Tarantola, 1984; Tarantola, 1986, 1988). Tomography, however, fails to resolve small-scale heterogeneities having dimensions less than several dominant wavelengths of band-passed body waves. In practice this scale is less than several hundreds of kilometers. The distribution and shapes of these small-scale heterogeneities carry constraints on plate tectonics and compositional mixing of the mantle by convection. Directionally dependent scattering and focusing and defocusing of seismic waves by these small-scale heterogeneities will introduce a fluctuation in the travel time and amplitude of the seismic waves, affecting estimates of viscoelastic attenuation and anisotropy, which in turn are important for estimating temperature, mineral composition, and mineral phase. An example of these effects is shown in Fig. 1.1. If originating from temperature variations alone, velocity heterogeneities smaller in dimension than several hundred kilometers will tend to be removed by thermal diffusion over several 10s of millions of years. Such small-scale heterogeneities must be explained by chemical heterogeneities (Helffrich, 2006; Kaneshima and Helffrich, 2010), that have persisted since the early history of the Earth. The study of small-scale heterogeneities will thus help in understanding the history and nature of compositional mixing of mantle.
Fig 1.1 Left: example of a 2-D model of mantle heterogeneity from thermochemical convection (Brandenburg et al., 2008). There is a change in the anisotropy of scale lengths between the upper mantle and lower mantle. Middle and Right: the possible heterogeneity of different kinds of scale length anisotropy; its effect on back-scattered body wave coda (Cormier, 2000).

An alternative approach is to retrieve a statistical representation of structure from observations of the fluctuation of the amplitude and travel time of teleseismic body waves recorded by array sensors. These fluctuations are created by small-scale heterogeneities beneath the arrays that scatter, focus, and defocus steeply incident body waves, which can be treated as plane waves incident on the upper mantle beneath the receivers from teleseismic sources. While deterministic seismic tomography gives both location and variation of the velocity heterogeneities, the stochastic approach provides an overall description about the assemblage of heterogeneities such as the velocity
perturbations and scale lengths.

With the development of High Performance Computing (HPC) and denser seismic station arrays installed all over the world, studying the fine structure of the Earth has become easier. Beginning in the early 1970’s (Aki, 1973; Berteussen, 1975; Capon, 1974), these small-scale heterogeneities have been described by statistical models that complement the deterministic models retrieved from global travel-time tomography. The observed statistic of the fluctuations wave field amplitudes and travel times can be used to estimate the spatial heterogeneity spectrum of Earth.

To invert for this heterogeneity spectrum, Aki (1973) first proposed using the coherence of the fluctuation in travel time and log amplitude of steeply incident seismic body waves observed at dense arrays of surface seismometers. Flatte and Wu (1988) extended Aki’s methods to include angular coherence, using seismic waves arriving from different incoming directions. Later, several studies (Chen and Aki, 1991; Wu and Flatté, 1990) further extended this method to include observations of Joint Transverse Angular Coherence Functions (JTACF). The inversion using these JTACFs were numerically tested by later work (Wu and Xie, 1991), who named this method “stochastic tomography”. Subsequent work of Zheng and Wu (Zheng and Wu, 2008a) extended it to allow a depth-dependent reference model of seismic velocity. Fig 1.2 demonstrates the different types of coherence functions and wavefronts interacting with heterogeneities beneath the receivers.
Fig. 1.2. Wavefronts of different coherences and their interactions with heterogeneities beneath the receivers, where $\theta$ denotes the incident angle and $\rho$ denotes the lag distance between receivers. Top left: transverse coherences (Aki, 1973); top right: angular coherences (Flatté and Wu, 1988); bottom: joint transverse and angular coherences (Chen and Aki, 1991; Wu and Flatté, 1990).

As a supplement to deterministic tomography, stochastic tomography can be used to study the spatial power spectrum (the Earth model in wave number space) and the spatial correlation of seismic velocities. Instead of retrieving the costly deterministic model profile, the spatial spectrum carries a lot of important physical information such as the
length scale and statistical shape of heterogeneities, and the variance of seismic velocity fluctuations at different depths. The majority of published work on inverting seismic waveform coherence functions assumes a constant background model to retrieve a depth-independent heterogeneity spectrum of a single layer (Flatté et al., 1991a; Flatté and Xie, 1992; Flatté et al., 1991b; Zheng, 2013). Although the inversion theory has been developed for depth-dependent heterogeneity, its application has thus far been limited.

In this thesis, I will focus on characterization and depth dependence of small-scale heterogeneities in the deep Earth using different approaches: a.) comparison of seismic waveform data with the results of forward modeling using numerical methods to simulate wave propagation through a heterogeneous media; and b.) stochastic tomography as an inversion technique, including both numerical validation and application to real seismic events, to derive a depth-dependent heterogeneity spectrum for the upper mantle beneath the USArray from observations of amplitude and phase coherences of teleseismic $P$ waves.

By including the inverted power spectrum in forward modeling experiments we can fill the gap in exploring the effects of the heterogeneity spectrum determined from transmitted body waves using whole-earth tomography and that inferred from the forward scattered coda of body waves (e.g. Cormier et al., 2011). Examination of the effects of these heterogeneous models on the waveforms of seismic body waves may assist in separating the scattering and focusing/defocusing effects of small-scale heterogeneity
from the effects of viscoelasticity and anisotropy, providing new constraints on the composition, phase, and temperature of the mantle.

Structurally, Chapter 2 discusses random media as a representation of heterogeneity; Chapter 3 describes the method of stochastic tomography and some numerical tests and validation; Chapter 4 describes our data and measurement of coherences; Chapter 5 gives the detailed approach we used for inverting for the heterogeneity spectrum from observed coherences; Chapter 6 summarizes the results of inversions for the heterogeneity spectrum for the upper 1000 km of mantle beneath the USArray; Chapter 7 discusses the significance of the results for chemical and phase heterogeneity of the upper mantle; and Chapter 8 gives the conclusion and a few outlooks of this study.
Chapter 2. Random Media: a Representation of Small-scale Heterogeneity

2.1 Random Media

Geological structure can be expressed by a spatial function $U = U(\vec{x})$, where $U$ is a material parameter that may represent either the $P$ or $S$ wave velocity, density, or temperature, etc. Since deterministic Earth models usually have limited spatial resolution, we often use a random media with known statistical characteristics to represent small-scale heterogeneities of these parameters.

Reference Earth models of seismic velocities and mass densities are smooth models representing average, usually radial and large-scale, Earth properties. The spatial variation of these properties can be written as $U_0(\vec{x})$. Multi-scale whole Earth structure can then be written as a perturbation of these average large-scale properties:

\begin{equation}
U(\vec{x}) = U_0(\vec{x}) + \delta U(\vec{x})
\end{equation}
The fluctuation $\delta U(\bar{x})$ is the small-scale heterogeneity added to the reference model $U_0(\bar{x})$. It has the following properties:

$$\langle U(\bar{x}) \rangle = U_0(\bar{x})$$  \hspace{1cm} (2)

$$\langle \delta U(\bar{x}) \rangle = 0$$ \hspace{1cm} (3)

Here the $\langle U(\bar{x}) \rangle$ notation denotes the statistical average of $U(\bar{x})$. We may form a spatial correlation function:

$$F(\bar{x}_1, \bar{x}_2) = \langle \delta U(\bar{x}_1) \cdot \delta U(\bar{x}_2) \rangle$$ \hspace{1cm} (4)

Since the random media is stationary, this function only depends on $\bar{x}$ (Tarantola, 1987), where $\bar{x} = \bar{x}_1 - \bar{x}_2$. In this case, this function becomes an autocorrelation function.

$$F(\bar{x}) = \langle \delta U(\bar{x}_1) \cdot \delta U(\bar{x}_1 + \bar{x}) \rangle_{\text{for all } \bar{x}_1}$$ \hspace{1cm} (5)

If we chose $\bar{x} = 0$, this function becomes
\[ F(0) = \langle \delta U(\bar{x})^2 \rangle_{\text{for all } \bar{x}} = \sigma^2, \]  

where \( \sigma^2 \) is the variance of the random media.

Accordingly, this autocorrelation function \( F(\bar{x}) \) is a characterization of the spatial scale of the random media and a measurement of its magnitude of irregularity. The Fourier transform of the autocorrelation function \( \hat{F}(\bar{k}) \) is also frequently discussed, because the synthetic realization of a random media is based on \( \hat{F}(\bar{k}) \) (see next section). For the 3-D case, the Fourier transform \( \hat{F}(\bar{k}) \) can be interpreted as the Power Spectrum Density Function (PSDF), which will be our primary inversion target in Chapter 5.

### 2.2 The Generalized Form of Power Spectrum

There are various autocorrelation functions studied to represent different types of random media, such as exponential, Gaussian, and self-affine. Klimes (Klimes, 2002) demonstrated that most of the commonly studied random media are special cases of the following function:
\[ \hat{F}(k) = \sigma^2 [a_v^{-2} + k^2]^{\frac{d+2N}{2}} \exp\left(-\frac{a_g^2 k^2}{4}\right), \quad (7) \]

where \(d\) is the spatial dimension, \(k\) is the wave number, \(\sigma^2\) is the variance, and \(N\) is defined as the Hurst parameter. \(a_g\) and \(a_v\) are the correlation lengths, which are the Gaussian correlation length and von Karman correlation length respectively. Note that Eq. (7) is actually a generalized form of the random media's power spectrum. This function can work as a regularization term when inverting seismograms for the power spectrum spectral density function (PSDF). If the complete PSDF is challenging to retrieve by inversion, for example, too many variables and too few constraints, we can still obtain these four parameters \(a_v, a_g, N, \sigma\), to reconstruct the generalized PSDF.

Considering a particular case, let \(N = -\frac{d}{2}N\). This function then becomes:

\[ \hat{F}_g(k) = \sigma^2 \exp\left(-\frac{a_g^2 k^2}{4}\right), \quad (8) \]

Its inverse Fourier transform is simply the Gaussian autocorrelation function:
\[
F_{g}(x) = \frac{\sigma^2}{\pi^{d/2} a_g^d} \exp \left( -\frac{k^2}{a_g^2} \right)
\]  

(9)

If we let \( a_g = 0 \), this function becomes:

\[
\hat{F}_v(k) = \sigma^2 \left[ a_v^{-2} + k^2 \right]^{-\frac{d+2N}{2}},
\]

(10)

which is the representation of the von Karman spectrum. For the special case \( N = 1/2 \), we have an exponential random media:

\[
\hat{F}_e(k) = \sigma^2 \left[ a_v^{-2} + k^2 \right]^{-\frac{d+1}{2}}
\]

(11)

\[
F_e(x) = \frac{a_v \sigma^2}{2^d \pi^{d/2}} \exp \left( \frac{x}{a_v} \right)
\]

(12)

And for special case \( a_v \to \infty \), we have a self-affine random media:

\[
\hat{F}_s(k) = \sigma^2 k^{-d-2N} \exp \left( -\frac{a_g^2 k^2}{4} \right)
\]

(13)
\[ F(x) = \frac{\sigma^2 a_g^{2N}}{2^{d+2N} \pi^{d/2}} f\left(-N; d; -\frac{x^2}{a_g^2}\right), \]  

(14)

where \( f\left(-N; d; -\frac{x^2}{a_g^2}\right) \) is Kummer’s hypergeometric function. As we can note from Eq. (13), this is a spectral function representing a random media of the von Karman type multiplied by a Gaussian low-pass filter.

### 2.3 Numerical Realization of Random Media from Power Spectrum

To realize a synthetic random media using a known auto-correlation function or power spectrum, we can take the following steps:

1. Generate random white noise at all spatial grid points \( W(x) \).

2. For an auto-correlation function, calculate its Fourier transform \( \hat{F}(k) \) and \( \hat{f}(k) \) to satisfy this equation: \( \hat{F}(k) = \hat{f}(k)^* \hat{f}(k) \), where \( \hat{f}(k) \) can be treated as a spectral filter.
3. Transform $W(x)$ into wavenumber space $\hat{W}(k)$ and apply the spectral filter $\hat{f}(k)$ to get $\delta \hat{U}(k) = \hat{W}(k) \cdot \hat{f}(k)$

4. Invert transform delta $\delta \hat{U}(k)$ to real space $\delta U(x)$, to get the heterogeneity term $\delta \hat{U}(k)$ in Eq. (1).

Using this scheme, we can generate the example three typical types of random media, as shown in Fig. 2.1.
Fig. 2.1 Random media represented by different types of autocorrelation functions: Gaussian (top left), exponential (top right), self-affine von Karman (bottom)
2.4 Calculation of Power Spectrum from Tomography Model

For determining the power spectrum from a known random media model, we can apply a Monte Carlo approach by using the following steps:

1. For isotropic media, randomly select two points $x_1$ and $x_2$ in the model for a fixed distance $x$, and calculate $U(x_1) \cdot U(x_2)$.

2. For every fixed distance $x$, calculate the correlation function

$$F(x) = \langle U(x_1) \cdot U(x_2) \rangle,$$

where $\langle \cdot \rangle$ stands for the averaging of different realizations. Repeat until the calculation of $F(x)$ converges.

3. $F(x)$ is the auto correlation of the model and its Fourier transform $\hat{F}(k)$ is the power spectrum function.

4. For anisotropic model, replace $x$ with $\bar{x}$ and averaging all $(x_1, x_2)$ pairs with the same directions and distances.

As an example we consider the model NWUS11-P (James et al., 2011). This is a shear-wave tomography model for the northwestern US. The calculated power spectra for different depths are shown in Fig. 2.2.
Fig. 2.2  Tomography model NWUS-11P and corresponding power spectrum at different depth: 150 km, 400 km, 650 km and 900 km.
Chapter 3. Inverse Modeling using Stochastic Tomography: Methodology and Numerical Validation

3.1 Theoretical Coherence Functions

For a plane wave 1, we can measure the log amplitude $u_1(r_1)$ and phase $\phi_1(r_1)$ at a station $r_1$. Similarly, for a second plane wave 2, we can measure the log amplitude $u_2(r_2)$ and $\phi_2(r_2)$ at $r_1$. Using these measured quantities, we can form the coherence functions of log amplitude fluctuation $u$ and phase fluctuation $\phi$ (Zheng and Wu, 2008b):

$$\langle u_1, u_2 \rangle = (2\pi)^{-1} \int_0^\pi d\xi a_1(\xi) a_2(\xi) \int_0^\infty J_0(\kappa R(\xi)) \cos[\omega \theta_1(\xi)] \cos[\omega \theta_2(\xi)] P(\xi, \kappa) \kappa d\kappa$$

$$\langle \phi_1, \phi_2 \rangle = (2\pi)^{-1} \int_0^\pi d\xi a_1(\xi) a_2(\xi) \int_0^\infty J_0(\kappa R(\xi)) \sin[\omega \theta_1(\xi)] \sin[\omega \theta_2(\xi)] P(\xi, \kappa) \kappa d\kappa$$

Where the meaning of each variable is listed below.
Integral limit $H$ : the thickness of the heterogeneity layer

Variable $\xi$ : depth

Variable $\kappa$ : magnitude of transverse wavenumber

Function $a$ :

$$a(p, \xi) = \frac{k^2(\xi)}{k_z(p, \xi)}$$  \hspace{1cm} (16)

Where $k(\xi) = \omega / c(\xi)$ and $k_z(p, \xi)$ is its z component. And $p$ is scalar ray parameter corresponding to specific plane wave and $c(\xi)$ is the reference velocity at depth $\xi$.

The function $\vartheta$ is:

$$\vartheta(p, \xi, \kappa) = \frac{1}{2} \frac{d^2 \tau}{dp^2} \frac{\kappa^2}{\omega^2},$$  \hspace{1cm} (17)

where
\[
\tau(p, \xi) = \int_0^\xi dz \sqrt{\frac{1}{c^2(z)} - p^2}
\]  

(18)

where \( p \) is the ray parameter defined by \( p = r \frac{\sin i}{V_r} \), with incident angle \( i \) and velocity \( V_r \) at radius \( r \) in a spherical geometry. By Snell’s law, ray parameter is constant for single ray propagating through any media, i.e. given a specific event location and receiver location, the ray parameter is definite.

Function \( R \) in Eq. (16) is:

\[
R(r_1, r_2, p_1, p_2, \xi) = \left| \tilde{R}_1(p_1, \xi) - \tilde{R}_2(p_2, \xi) \right|
\]  

(19)

This \( R \) is the horizontal distance of two rays at depth \( \xi \), where \( \tilde{R}_1(p_1, \xi) \) is the ray trajectory connecting station \( r_1 \) for plane wave \( p_1 \), and \( \tilde{R}_2(p_2, \xi) \) is the ray trajectory connecting station \( r_2 \) for another plane wave \( p_2 \). This geometry is as showed in Fig 3.1.

Note that in this figure, it shows a 2D case, where the azimuths of two incoming waves are in the same direction with the station lag vector \( \overline{r_1 r_2} \).

Function \( P(\xi, \kappa) \): the Fourier transform of the spatial correlation function of the heterogeneities. The absolute value of \( P \) is the power spectrum.
Fig. 3.1  Geometry of two rays propagating to different stations.
3.2 Discretization and Implementation

In order to retrieve the power spectrum of the media as a function of the wavenumber, which is the P function of Eq. (15), we discretize the coherence functions into the following format:

\[
\begin{align*}
    \langle u_1 \cdot u_2 \rangle \\
    \langle \phi_1 \cdot \phi_2 \rangle \\
    \langle \phi_1 \cdot u_2 \rangle
\end{align*}
\]

\[= \text{IntegralMatrix} \cdot P(h,k) \quad (20)\]

The left-hand side of the equation is the correlation function considered as “data”, measured from seismograms. It is a function of the stations and source positions. The right-hand side of the equation is an integral, and this can be treated as a linear matrix multiplied by the depth-dependent power spectrum which is the “unknown” we want.

The discretization of the integral matrix is as follows:
I.M. = \left\{ \begin{array}{l}
\frac{1}{2\pi} \sum_{i=1}^{H} \left( \sum_{j=1}^{K} d_h \cdot a_{i,j} \cdot a_{2,j} \right) \sum_{j=1}^{K} dk \cdot (j \cdot dk) \cdot \sin(w \cdot \theta_{1,j}) \cdot \sin(w \cdot \theta_{2,j}) \cdot J_0(j \cdot dk \cdot \theta_i) \\
\frac{1}{2\pi} \sum_{i=1}^{H} \left( \sum_{j=1}^{K} d_h \cdot a_{i,j} \cdot a_{2,j} \right) \sum_{j=1}^{K} dk \cdot (j \cdot dk) \cdot \cos(w \cdot \theta_{1,j}) \cdot \cos(w \cdot \theta_{2,j}) \cdot J_0(j \cdot dk \cdot \theta_i) \\
\frac{1}{2\pi} \sum_{i=1}^{H} \left( \sum_{j=1}^{K} d_h \cdot a_{i,j} \cdot a_{2,j} \right) \sum_{j=1}^{K} dk \cdot (j \cdot dk) \cdot \sin(w \cdot \theta_{1,j}) \cdot \cos(w \cdot \theta_{2,j}) \cdot J_0(j \cdot dk \cdot \theta_i)
\end{array} \right. 
\right) (21)

Here \( d_h \) and \( dk \) are numerical integral increments of depth and wave number. Subscript \( i \) denotes the iteration over depth and \( j \) denotes iteration over wavenumber. These functions \( a \) and \( R \) are only functions of depth, while \( \theta \) and Bessel term \( J_0(j \cdot dk \cdot \theta_i) \) are functions of both wave number and depth.

The functions \( a_{1i} \) and \( a_{2i} \) are:

\[
a_{1i} = \left( \frac{w}{c_i} \right)^2 \left( w \sqrt{\frac{1}{c_i} - p_1^2} \right)
\]

\[
a_{2i} = \left( \frac{w}{c_i} \right)^2 \left( w \sqrt{\frac{1}{c_i} - p_2^2} \right)
\]

Then discretize the \( \tau \) function (pre-calculated for all ray parameter \( p \), with increment \( dp \), as explained in next section) and its first derivative given by
\[ \tau_i(p) = \sum_{j=1}^{i} d \sqrt{1 - \frac{1}{c_j^2} - p^2} \]  

(23)

\[ \tau_i'(p) = \frac{[\tau_i(p + dp) - \tau_i(p)]}{dp} \]  

(24)

Functions \( \theta_1 \) and \( \theta_2 \) are:

\[
\theta_{1,i,j} = \frac{1}{2} \left[ \frac{\tau_i'(p + dp) - \tau_i'(p)}{dp} \right]_{p=p_i} (j \cdot dk)^2 \frac{\omega^2}{\alpha^2} 
\]

\[
\theta_{2,i,j} = \frac{1}{2} \left[ \frac{\tau_i'(p + dp) - \tau_i'(p)}{dp} \right]_{p=p_j} (j \cdot dk)^2 \frac{\omega^2}{\alpha^2} 
\]

(25)

Function \( R \) is the horizontal distance (great circle distance) of the two rays at a specific depth. As discussed in Section 3.1, is a function of \( r_1, r_2, p_1, p_2 \) and depth \( \xi \). If we do not consider the effect of azimuth (assuming zero azimuth), and with two given incident plane waves (fixed ray parameters), \( R \) only depends on the depth and the surface lag distance between receivers. With the first derivative of \( \tau \) function (pre-calculated, explained in Sec. 3.3), we can easily compute \( R \) using following formula:

\[ R_i = \text{abs} \left[ \text{lag} - \tau_i'(p_1) + \tau_i'(p_2) \right] \]  

(26)
However, if we take the azimuth into consideration, the calculation of $R$ will differ. For the 3D case, as shown in Fig. 3.2, $p_1$ and $p_2$ are in different planes. We may rotate them to $p_1'$ and $p_2'$ so that the rays are in the same plane.

Fig. 3.2  Geometry of two rays propagating to different stations.
In this case, since \( ZA' \) is a rotation of \( ZA \), these two length are both \( \tau'_1 \), as well as \( ZB \) and \( ZB' \). If we take \( \Delta \theta \) as the rotation angle, then apply the cosine law to triangle \( \Delta ZBB' \) and \( \Delta ZAA' \), we then have following equation for \( R \).

\[
R = \sqrt{R'^2 + 4 \cdot \tau'^2 \cdot \sin^2 \left( \frac{\Delta \theta}{2} \right) + 4 \cdot \tau'_2 \cdot R' \cdot \sin^2 \left( \frac{\Delta \theta}{2} \right)}, \quad (27)
\]

where \( R' \) is from Eq. 26,

\[
R' = \text{abs} \left[ \text{lag} + \tau'_1 - \tau'_2 \right] \quad (28)
\]

### 3.3 Numerical Analysis

We performed a numerical test to implement the discretization and study practical numerical limits of integration required to treat rapidly oscillating terms.
First we set the integral limit over depth to 500km, which assumes a heterogeneous layer of 500km thickness, and used the IASP91 model (Kennett, 1991) is used as our background reference Earth model.

For all ray parameters corresponding to teleseismic events of distance 45 to 90 great circle degrees, we pre-calculate the $\tau$ function and its derivatives, with even sampling over ray parameter $p$ (Sun et al., 2017). Note: generally, the ray paths are determined by event’s location and receiver’s location. If we assume a perfect spherical Earth, the ray paths are only determined by the distance between event and receiver plus a horizontal shift due to symmetry. Applying Snell’s law in this symmetry, the incident angles, seismic event distances, and ray parameters are one-to-one mapped. Here 45 and 90 degrees of great circle distance are about 5000 km and 10000 km, similar to the distance from center of the EarthScope array to Central American and to Northeast Oceania, respectively.

We choose two incoming plane waves of incident angle 15 degree and 20 degrees, which approximates a direct P wave observed at two teleseismic of distances of about 65 and 85 great circle degrees respectively. Then we calculate the different terms to build the integral matrix: $a_i \cdot a_z \cdot \cos(w \cdot \theta_{b,i,j}) \cdot \cos(w \cdot \theta_{2i,j})$, $R$, and the Bessel function term.
To determine the integral limits of $\kappa$, we choose the probe frequency to be $0.5\text{Hz}$ and we compute the $\cos(\omega \theta_{i,j}) \cdot \cos(\omega \theta_{2i,j})$ term for different limits of $\kappa$ : 0.2, 0.3 and 0.5. The result is shown in Fig.2. We can see that there is an oscillation with increasing $\kappa$. 
That means though theoretically we are supposed to integrate $\kappa$ from 0 to $\infty$, we need to set an upper limit to avoid under-sampling the integrand in regions of where it rapidly oscillates. The period is changing with $\kappa$, which is $\frac{4\pi\omega}{k\tau_{pp}}$. When $\kappa$ is large, the period $T$ is small (i.e., oscillating more). We can choose the integration interval to be smaller than $T/10$. That means:

$$dk \leq \frac{2\pi\omega}{5k\tau_{pp}}$$ (29)

In this study here we choose the integration interval to be 0.001 ($2\pi$/km) and the upper limit to be 0.3 ($2\pi$/km) to keep the integration stable and accurate. With this integral limit, the $a_i \cdot a_j$ function, R function and Bessel function plot is shown in Fig. 3.4, Fig. 3.5, and Fig. 3.6. Note we choose the station lag distance between -300 and 300 km with spacing 1km for R.
Fig. 3.4 \( a_1 \cdot a_2 \) function term for calculating integral matrix
Fig. 3.5  R function term for calculating integral matrix
To represent a heterogeneous random medium, we use white noise filtered by a spectral filter $\hat{F}(k)$. If we take the inverse Fourier transform of the product $\hat{F}^*(k)\hat{F}(k)$, we can
get the Fourier transform of the correlation function that describes the statistical properties of this random medium. Therefore the product \( \hat{F}^*(k)\hat{F}(k) \) is actually the power spectrum for which we are looking.

As described in Chapter 2 and demonstrated by Klimes (2002), commonly studied random media power spectra are all special cases of this equation:

\[
\hat{F}(k) = \kappa \left[ a_v^{-2} + k^2 \right]^{\frac{1+\eta}{2}} \exp \left( -\frac{a_g^{-2} k^2}{8} \right)
\]  \hspace{1cm} (30)

Where \( \eta \) is Hurst parameter, \( \kappa \) is related with RMS velocity of the random media, \( a_v \) is Von Karman correlation length and \( a_g \) is Gaussian correlation length. All typical random media can be fit by this equation by choosing different parameters.

Here we use three typical filters to check how different types of power spectrum will affect the calculated amplitude/phase coherences: Gaussian, Exponential, and Zero Von Karman, as shown in Figure 3.7.
Fig. 3.7  Testing Spectrum: blue is Gaussian spectrum with $N = -1$, $ag = 40$ km and 5% rms; red is Exponential spectrum with $N = 0.5$, $ag = 0$, $av = 30$ km and 5% rms; green is Zero Von Karman spectrum with $N = 0$, $ag = 0$, $av = 30$ km, and 5% rms.
3.4.1 Transverse Coherence Functions: Different Types of Spectrum

With Transverse Coherence Functions (TCF) the incoming seismic wave is approximated as one plain wave, and we measure the amplitude/phase fluctuations at receivers and calculate the coherences over different receivers. It is a function of lag distance between receivers.

In our first test, we calculate the TCF using different types of heterogeneity spectrum. Here we assume the heterogeneity is one layer, with the power spectrum constant over 1000 km depth. The calculated TCFs determined from different model spectra are shown in Fig. 3.8 (a) – (c).
Fig. 3.8 (a) Calculated TCF on Gaussian spectrum. The multiple lines are different incident waves corresponding to distant P wave from 60 to 90 degrees.

Fig. 3.8 (b) Calculated TCF on Exponential spectrum. The multiple lines are different incident waves corresponding to distant P wave from 60 to 90 degrees.
Calculated TCF on Zero Von Karman spectrum. The multiple lines are different incident waves corresponding to distant P wave from 60 to 90 degrees.

Note that the exponential and Gaussian power spectrum show a rather smooth coherence, and there is also a large negative coherence at around 60-70 km. This does not agree with our synthetic results very well. Therefore we choose a Von Karman spectrum for further tests.

Also we can see that these TCFs are symmetric over positive and negative lag distance. This proves that the TCF is only a function of lag distance between stations and is irrelevant to its direction. For different propagation distances, the result also shows an energy loss due to geometric spreading.
3.4.2 Transverse Coherence Functions: Effects of Variant Parameters of Von Karman Model.

For a zero Von Karman model (we set the Hurst parameter as zero for easy computing), the effect on the TCF due to the perturbation parameter and correlation length is shown in Fig 3.9 (a) and (b).

Fig. 3.9 (a) Calculated TCF on Von Karman spectrum. With fixed perturbation (rms) 5%, the multiple lines are from different correlation lengths from 10 km to 100 km.
Fig. 3.9 (b)  Calculated TCF on Von Karman spectrum. With fixed correlation length 30 km, the multiple lines are from different perturbations from 1% to 10%.

3.4.3  Transverse Coherence Functions: Add Depth Dependence

To test the effects of depth dependence of the heterogeneity spectrum, we build two models as shown in Fig 3.10. The two models are a Von Karman spectrum for $N = 0$, correlation length $= 30$ km, with dependence of perturbation parameters ascending from
1% to 5% and descending from 5% to 1%. The predicted coherence functions are shown in Fig 3.11.

Fig. 3.10  Depth-dependent Von Karman spectrum, with N=0 and fixed correlation length 30 km, the perturbation parameters linearly ascending from 1% to 5% (left) and linearly descending from 5% to 1% (right).
Fig. 3.11 Calculated TCFs on different Von Karman spectrum. The red is from ascending model and blue is from descending model in Fig. 9

3.4.4 Angular Coherence Functions

With the Angular Coherence Function (ACF), we measure the fluctuations for different seismic waves received at the same receiver, and calculate the coherences over different incoming waves. It is a function of azimuth angles and incident angles (corresponding to specific P wave propagating distances). Here we assume the same azimuth angle, and Fig 3.12 shows the calculated ACF from the Von Karman spectrum in Fig 3.7.
Fig. 3.12  Calculated ACFs on Zero Von Karman spectrum 30 km, 5% rms. Coherences of different events corresponding to P wave from 60 degree to 90 degree, assuming same azimuth angle.

3.4.5  Joint Transverse-Angular Coherence Functions

For the most general case, Joint Transverse-Angular Coherence Functions (JTACF) can be measured for either different events or different stations.
Since the JTACF is a function of both angles and lag distances between stations, here we present the example to show the coherence functions over different stations for two given waves, observed at 60 and 80 degree great circle distances and at the same azimuthal angle. The test spectra are shown in Fig 3.13 and the calculated coherence functions are shown in Fig. 3.14. Our result agrees with Zheng and Wu’s result (Zheng and Wu, 2008a) for the coherence functions measured from seismograms synthesized in heterogeneous media.

Fig. 3.13  Testing Spectrum: the right one is a simple Gaussian spectrum; the left one is more dispersed spectrum with noise.
Fig. 3.14 Calculated coherence functions. The upper two are amp-amp and phase-phase coherences based on the simple Gaussian spectrum in Fig. 12. The lower two are amp-amp and phase-phase coherences based on the noised spectrum in Fig. 12.
Chapter 4 Measurements and Data Mining on Seismograms

4.1 Data Source

To apply stochastic tomography to observed seismic wave fields, we select seismic events from different earthquake groups to measure the actual coherence functions observed across EarthScope USAArray. The EarthScope USAArray is the seismic portion of the EarthScope project, which includes different “observatory” networks: the Transportable Array, the Flexible Array, and the Reference Network. The typical station spacing is from 50 to 300 km.
Fig. 4.1 Stations in EarthScope USArray.

We invert these measurements to retrieve the power spectrum of the sub-surface heterogeneity. The derived Joint Transverse Angular Coherence Functions (JTACFs) on the left-hand side of Eq. (1.1) are functions with 10 degrees of freedom: two seismic rays of different event depths, receiver longitude/latitude and source longitude/latitude,

\[
\langle u_1(\hat{s}, \hat{r}) u_2(\bar{s}, \bar{r}) \rangle = \text{IntegralMatrix} \ast P(z, \bar{k})
\]  

(31)
We can apply a horizontal “quasi-homogenous” approximation (Flatté and Wu, 1988) to reduce the degrees of freedom. This approximation assumes, for each different depth in the model, that the 2-D horizontal heterogeneity is isotropic. This is equivalent to assuming the power spectrum is only 2-D $P(z,k)$, where $z$ is depth and $k$ is wavenumber. Under this approximation, the degrees of freedom of JTACFs $\langle u_1u_2 \rangle$ are reduced and become 6-D: two depths, two incident angles, lag distance of the receivers, and azimuth difference.

In order to use earthquakes as sources, however, 6-D is still too high for practical inversion. Due to the irregular distribution of event locations, it is also almost impossible to form a generalized mapping from the spectrum to the coherence functions. Therefore we make more approximations to study particular observations described in the next section.

### 4.2 Event Selection and Pre-process

First, we start with the Transverse Coherence Functions (TCF), which are transverse coherences from different receivers of a single incoming wave. The TCFs are just functions of the lag distance between the two receivers.
Since the seismic waves from the same locations will result in the same TCFs, we can sum up all seismic waves from nearby areas as an “earthquake group”, assuming that they share the same event depth and location (longitude/latitude). Also, by averaging the seismic waveforms, we can reduce noise, similar to filtering data via stacking.

For data, we used waveforms recorded by all available stations in the IRIS database (i.e. Incorporated Research Institutions for Seismology, the data manage center for USArray seismometer networks) in the western US between latitudes 30° N to 50° N, and longitudes 100° W to 125° W. These stations are shown in Fig 4.1.

For seismic events, we select P waveform data from three different earthquake groups, with the time range from 2000-01-01 to 2017-10-01, falling into these criteria:

1. Earthquakes from the Japan area: latitude from 25° N to 35° N, longitudes from 135° E to 145° E, depth from 400 km to 500 km, magnitude from 5.8 mw to 6.2 mw.

2. Earthquakes from the Tonga area: latitude from 15° S to 30° S, longitudes from 175° E to 175° W, depth from 500 km to 700 km, magnitude from 5.8 mw to 6.2 mw.

3. Earthquakes from the Chile area: latitude from 15° S to 30° S, longitudes from 100° W to 125° W, depth from 500 km to 700 km, magnitude from 5.8 mw to 6.2 mw.
These events and stations locations are as shown in Fig 4.2.

Fig. 4.2  Dataset Selection: three groups of teleseismic events.
We use the FDSN service (i.e. International Federation of Digital Seismograph Networks) to retrieve seismograms recorded at EarthScope USArray receivers. In this study, we only use direct P wave arrivals.

To preprocess the data, first we remove the instrument response and filter the seismogram with a band pass window 0.3 Hz to 1.5 Hz, then select the high Signal to Noise Ratio (SNR) waveforms. The SNR is defined by the Euclidean norm of P-wave signal over 10 second versus pre-event signal over the same time window. Of all seismograms (21205 records), about 40% are usable under this criterion.

4.3 Synthetic Reference Seismograms

Since \( u_1 \) and \( u_2 \) (or \( \phi_1 \phi_2 \)) here are fluctuations of the amplitude or phase, we need the reference waveforms from a base model. Therefore the synthetics computed in a reference model are required.

For robust synthetics algorithm, we use the AxiSem (Krischer et al., 2017; Nissen-Meyer et al., 2014) and the IRIS Syngine service (van Driel et al., 2015) to compute synthetic seismograms. AxiSem is a spectral-element method that requires axisymmetric
background models and runs within a 2D computational domain, thereby reaching all
desired highest observable frequencies (up to 2Hz) in global seismology. This
axisymmetry agrees with our use of the horizontal quasi-homogeneous approximation.
The IRIS Syngine service stores a pre-calculated database of Green Functions for
reference models using the AxiSem Method. (In this study we use IASP91 (Kennett,
1991) as the reference model.)

To use the IRIS Syngine service, we need to input momentum tensors and source time
functions as earthquake source parameters. We use archived data from the Lamont-
Doherty Earth Observatory for Global Centroid-Moment-Tensor (GCMT) solutions
(Dziewonski et al., 1981; Ekström et al., 2012) to retrieve event metadata to represent the
P wave radiation pattern of the individual event. The source time functions are the far-
field time behaviors of the earthquake source, representing the time history of slip on a
fault plane. This slip history varies from earthquake to earthquake and can be complex.
It is easy, however, to retrieve this function empirically in some distance ranges where
the P wave is simple, consisting of a simple direct body wave that has not suffered multi-
paths or overlapping reflections from upper mantle discontinuities. For the source time
functions, we apply the following method:

1. Shift the waveforms to align them at the predicted P arrival using the TauP
   program’s travel time calculation for standard reference Earth models (Crotwell et
   al., 1999).
2. Cross-correlate and shift all waveforms to fully align the P arrival.

3. Stack all the shifted waveforms to retrieve the empirical source time function.

   This process is illustrated in Fig 4.3.

Fig. 4.3  Procedure of aligning and stacking P waveforms used to retrieve the empirical source-time function.
4.4 Measurement of Fluctuations and Calculation of Coherences

From measurements made on seismograms and synthetic waveforms, we can calculate the amplitude and phase fluctuations.

For amplitude fluctuations, the measurement is straightforward. Due to the single-scattering approximation, the waveform should be in the time window that covers a single wavelet. After Fourier transforming of the time-domain waveform, using a multitaper method (Percival and Walden, 1993; Zhang et al., 2016), the log amplitude is measured at the desired frequency (0.7 Hz for this case). For the phase fluctuations, we use a waveform cross-correlation method (VanDecar and Crosson, 1990). We also use the empirical source time function (calculated in the previous step) to cross correlate all the waveforms twice for better precision,

After applying this method to both seismograms and reference synthetics, subtract observed from predicted references to get the fluctuations of log amplitude and phase.

The next step is to calculate the coherence functions using these fluctuations. For the TCFs, we care about the coherence over the lag distance of the receivers. Therefore for every two fluctuations (either log amplitude or phase), calculate the respective lag
distance and round it to 10 km increment intervals, multiply the fluctuations as coherence, and take the statistical mean of all products at each 10km interval. Since seismic waves from the similar epicenter locations will result in the same TCFs, we can sum up the results associated with all seismic waves from nearby areas as an “earthquake group”, assuming that they share the same event depth and location (longitude/latitude). By averaging the seismic waveforms, we reduce noise, similar to filtering data via stacking. The summation of earthquake waveforms from groups of events in broad region will also tend to remove and smooth over any unhealed amplitude and phase fluctuations induced by heterogeneities in the upper mantle beneath the source region.

After all these steps, the coherence functions are calculated for these three different earthquake groups. The result is shown in Fig. 4.4.
Fig. 4.4  TCFs from Japan (upper) Chile (middle) Tonga (lower) earthquake groups, depths 400 km to 700 km, magnitude 5.8 mw to 6.2 mw
4.5 Cross Coherence of Different Earthquake Groups

The joint transverse angular coherence functions (JTACF) describes the cross coherence of different events recorded at different receivers. As we’ve already discussed, it is a 10-D function depending on two receiver locations and two source locations (latitudes, longitudes; or equivalently incident angle, azimuth and lag distance). We applied a horizontal “quasi-homogenous” approximation (Flatté and Wu, 1988) to reduce the degrees of freedom. This approximation assumes, for each different depth in the model, that the 2-D horizontal heterogeneity is isotropic. This is equivalent to assuming the power spectrum $P(z,k)$ is a 2-D function of depth and wavenumber. Under this approximation, the degrees of freedom of JTACFs are reduced and become 6-D: two depths, two incident angles, lag distance of the receivers, and the azimuth difference of the two incident plane waves. In order to use earthquakes as sources, however, 6-D is still too high for practical inversion. Due to the irregular distribution of event locations, it is also almost impossible to form a generalized mapping from the spectrum to the coherence functions without further approximations. We assumed the events from the same earthquake group share the same incident angle. The cross coherence of different earthquake groups are as shown in Fig. 5. Due to the irregular distribution of events and receivers, we used a k-nearest neighbors algorithm (KNN) (Altman, 1992) to interpolate the data.
Fig. 4.5  Measured joint angular-transverse coherence functions (JTACFs)
Cross Coherence of different earthquakes: Japan group (JP), Chile
group (CL), Tonga group (TG).
Chapter 5. Stochastic Tomography: Invert Coherence Functions

5.1 Inversion Scheme with Re-sampling

Equation (15) can be written as

$$C(\tilde{x}) = M(\tilde{x}, \tilde{k}, z) \cdot P(\tilde{k}, z),$$

(32)

where $C(\tilde{x})$ are the measured coherence functions, $M(\tilde{x}, \tilde{k}, z)$ is an integral operator in Eq. (15); $P(\tilde{k}, z)$ is the 2.5-d power spectrum density function with respect to depth and wavenumber. With an axisymmetric approximation, this can be considered as the layered anisotropic power spectrum at each depth.

Using the measured coherence functions, we can solve for the power spectrum of the heterogeneity by minimizing

$$\left\| M(\tilde{x}, \tilde{k}, z) \cdot P(\tilde{k}, z) - C(\tilde{x}) \right\|$$

(33)
As we have discussed in Chap. 3, the integral operator $M$ can be discretized as an integral matrix. Then the equation becomes

$$C_{ij} = M_{ijmn} P_{mn}, \quad (34)$$

where $M_{ijmn}$ is the integral matrix in Eq. (16), with a dimension of $(ij \times mn)$, where $i$ and $j$ are the iterations over different discrete receivers, $m$ and $n$ are discrete intervals of depth and wave number respectively. $i$ and $j$ can be reduced to one variable or expanded to more variables depending on the calculation of either the Transverse Coherence Functions (TCF), or Angular Coherence Functions (ACF) or both.

We need to consider, however, the integral stability limitation, as discussed in Chapter 3,

$$dk \leq \frac{4\pi\omega}{k\tau_{pp}}, \quad (35)$$

If we let the $dk$ be small enough to satisfy this limitation, $P(\tilde{k},z)$ may become a function with more unknown parameters than $C(\tilde{x})$, making this inversion an underdetermined problem. We may perform a linear transform to deal with this issue. Assuming the operator $M$, calculate the integral over wave number first, then the integral over depth, and the transform will be
\[ P_{m'n'} = T_1 T_2 P_{mn} \] (36)

\[
T_1 \bigg|_{m' \times mn} = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
0 & 0 & \cdots & 1 \\
\vdots & & & \\
0 & 0 & \cdots & 1 \\
\end{bmatrix}
\]

\[
T_2 \bigg|_{m' \times m'n'} = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 \\
\end{bmatrix}
\]

\[ (37) \]

\[ P_{m'n'} \] is the transformed target spectrum function and \( P_{mn} \) is the real target spectrum function with fewer parameters. The matrices of T1 and T2 actually work as an upsampling transform, which is illustrated in Fig 5.1:
Fig. 5.1  Transformation $T_1$ and $T_2$ in Eq. (37) as an up-sampling mapping, which is used to satisfy the limitation Eq. (36).
If we let $M_i = M_{ijmn}T_1T_2$, the object function then becomes

$$\min \left\| M_i P_{mn} - C_{ij} \right\|^2$$  \hspace{1cm} (38)

The most straightforward way to do the inversion is to use a Moore–Penrose inverse

$$P = M_i^T \left( M_i M_i^T \right)^{-1} C$$  \hspace{1cm} (39)

Or a more computationally efficient way, apply gradient descent method, by iterating

$$P_{new} = P_{old} - \gamma \cdot \nabla \left\| M_i P_{old} - C \right\|$$  \hspace{1cm} (40)

Since the direct inversion is hard to regularize (too many unknown parameters for a 2-D spectrum), we may invert the coherence functions step by step: first assuming a single layer spectrum, then add depth dependence.
5.2 Single Layer Assumption and Inversion with L-2 Regularization

First, we may assume the power spectra are the same at different depths. The objective function is

\[
\min \left( \left\| M_i P_{mn} - C_i \right\|^2 + \lambda L P_{mn} \right) \quad (41)
\]

The second term is a Tikhonov regularization (Tikhonov and Arsenin, 1977), where \( L \) is first order derivative operator matrix and \( \lambda \) is a small real number that can be determined by the L-curve technique (Hansen, 1992). This regularization is used to avoid over fitting.

In this computation, we use 100 unknown parameters, and a 0.003 km\(^{-1}\) wavenumber grid spacing (integration limitation and stability discussed in Chapter 3). The result is shown in Fig. 5.2. A comparison of synthetic coherence functions and observations is shown in Fig. 5.3. The coefficient of determination (Zhang et al., 2015) is 0.74 and the reduced chi-squared, which is used to show the goodness of fit (Taylor, 1997), is 1.76.
Fig. 5.2 Inversion Result: spectrum retrieved under a single layer approximation and L-2 regularization.
Fig. 5.3 Comparison: synthetic coherence functions by spectrum of Fig. 5.2 VS observed coherences.
5.3 Inversion Regularized by a Generalized Function

In Chapter 2, we discussed that the most frequently studied random media spectra are special cases of Eq. (7), derived by Klimes (Klimes, 2002). Using this equation, the P term can be written as a function of 4 unknown parameters. The object function then becomes

\[
\min \left( \| M_i P(a_r, a_g, \kappa, N) - C_q \|^2 \right)
\]  

(42)

Using a constraint on the Hurst parameter, for physically reasonable scaling properties of geological structures (Addison, 1997; Turcotte, 1989):

\[-\frac{3}{2} \leq N \leq 1\]  

(43)

Since P is a non-linear function, the gradient descent iteration for this function is in a different form. With \( X \) standing for one parameter of \( (a_r, a_g, \kappa, N) \), we have

\[
X_{\text{new}} = X_{\text{old}} - \gamma \cdot \nabla \left\| M_i P(X_{\text{old}}) - C_q \right\|
\]  

(44)
which is

\[ X_{\text{new}} = X_{\text{old}} - \gamma \left( M_i^T \left( M_i P(X_{\text{old}}) - C_i \right) \right)^T J_{PX} \]

(45)

Where \( J_{PX} \) is the Jacobian of function \( P \) with respect to parameter \( X \). The inversion result is that \( a_g = 28, N = -0.4, \kappa = 0.9 \); \( a_v \) becomes extremely large as the gradient descent iteration continues, so we set \( a_v = 2000 \) to the physical reasonable limit: 2000 km as the scale invariance up-limit length (Mandelbrot, 1977). The comparison of synthetic coherence functions and observations is shown in Fig. 5.4. The coefficient of determination of 0.52 and reduced chi-squared is 1.71. Note although the prediction looks not a good fit, with a significant smaller coefficient of determination, the chi-square per degree of freedom is very similar to the previous inversion with Tikhonov regularization using more parameters. This suggests the Klimes function is still a good representation of the shape of the heterogeneity spectrum in wavenumber because it uses much fewer parameters (only four).
Fig. 5.4  Comparison: synthetic coherence functions by parameter-regularized spectrum VS observed coherences.
5.4 Inversion Adding Depth Dependence

The previous inversion assumes that the heterogeneity power spectrum is the same over depth. We can add depth dependency to make the predicted coherence function fit more with our measurement. We may let:

\[ P(k,z) = \sigma_z^2 P(k) \]  

\((46)\)

\(P(k)\) is the inversion result from the previous section and \(\sigma(z)\) becomes the new inversion target function. The \(\sigma\) is a measure of standard deviation of random media. In this case the gradient descent iteration becomes:

\[ \sigma_{new} = \sigma_{old} - 2\lambda \left( M^T_i \left( M_i \left( \sigma_{old}^2 P_k \right) - C \right) \right) \sigma_{old} P_k \]  

\((47)\)

We discretize the 1000 km model into 40 layers with 25 km spacing. The inversion result is shown in Fig 5.5. The comparison of synthetic coherence functions and observations is shown in Fig. 5.6. Note the coefficient of determination is 0.80 and the chi-squared per degree of freedom is 1.05. This shows a significantly improved goodness of fit. The reduced chi-squared is close to 1, indicating neither over fitting nor under fitting. The number of parameters is very well chosen.
Fig. 5.5 Inversion Result: depth dependency added to P velocity variance.
Fig. 5.6 Comparison: synthetic coherence functions by depth-dependent spectrum VS observed coherences.
Chapter 6 Analysis and Discussion

6.1 Discussion on Inversion Result with Single Layer Approximation

Analysis of Fig. 5.2 leads to the following interpretations:

1. Large power coefficient at low wavenumbers (< 0.018 km\(^{-1}\)) indicates that the dominant heterogeneities are still large-scale (> 350km).

2. A power peak at 0.022 km\(^{-1}\), indicates a characteristic length scale of 280 km. We can also see ripples between 0.058 km\(^{-1}\) and 0.11 km\(^{-1}\), representing smaller heterogeneity peaking 60 km to 100 km scale length.

3. Power decays with increasing waveumber. For wavenumber larger than 0.12km\(^{-1}\) the power coefficient tends to zero. Though multiple studies (Frankel and Clayton, 1986; Sato and Fehler, 1998) show that a pure Gaussian spectrum cannot explain both seismic wave scattering and travel-time variations, the inverted power spectrum is shaped like a low pass filter. A Gaussian low pass filter could be, hence, a good supplement to represent the heterogeneity at least in the resolved wavenumber band of this study. For
example, the Klimes function (Klimes, 2002) that we discussed in Chap. 2, is used to represent this random media by multiplying a spectrum with a Gaussian low pass filter.

The inverted parameters of the regularized Eq. (7), are \( a_g = 28, a_v > 2000, \ N = -0.4, \ \kappa = 0.9 \). The inner and outer cutoff scales of the self-affine random media (Mandelbrot, 1977) are the upper and lower limit for self-affine length scale:

\[
a_g \ll x \ll a_v
\]  

(48)

The result \( a_v > 2000 \) shows that the geological structure may be self-affine to even larger scale (hundreds of kilometers). This agrees with the study on well-logging data that the Earth is fractal (Goff and Holliger, 1999; Jones and Holliger, 1997; Wu et al., 1994), as well as some other fields of study (Zhang et al., 2017). \( a_g \) works as the Gaussian lower pass filter's characteristic length scale. The result \( a_g = 28 \) indicates that smaller heterogeneities (smaller than 28 km) are excluded in this generalized spectrum, which agrees with our inversion result in Fig. 5.2. Though these parameters have reasonable physical meanings, the calculated coherence functions, however, are under-fit in Fig. 5.4. This is probably because the spectrum is over regularized by the Klimes function, meaning the true spectrum should contain much more information than either low-pass or hi-pass filters.
6.2 Discussion on Depth Dependence

The inverted depth-dependent P-wave variance in Fig. 4.5 shows a rather interesting result. Notice there are velocity variance peaks at different depths. One possible guess is that these large velocity variances are the result of topography at the transition zone boundaries. We can prove, however, that this is not this case in Fig 6.1 – Fig 6.3.

Fig 6.1(a) gives an example medium of 3 layers of different heterogeneity spectra, with discontinuities at 400 km and 650 km. After multiple realizations of random media using the procedure as described in Chap. 2, we calculate corresponding velocity variances versus depth. The result is shown in Fig. 6.1(b). Then we add sinusoidal topography at the boundaries, as shown in Fig. 6.2(a). We repeat this calculation and plot the velocity variance versus depth in Fig. 6.2 (b). We can see that the sinusoidal topography will not add ripples to the plot, i.e., making the velocity variance larger at the boundaries. The topography effects simply behave like a transition zone with the velocity variance gradually changing from one layer to another layer.

The complete spectra are shown in Fig. 6.3. The original sub-models are synthetic random media using a Gaussian filter. The sinusoidal topography region is not merely the average of the two random media. The spectra at 400 km and 650 km show a non-zero trailing in the high wave number zone. The spectra of sinusoidal topography of the upper
mantle, mineral phase discontinuities are another proof that a simple Gaussian media cannot be a good approximate of the random heterogeneity of the mantle because it will eliminate all high wavenumber components.

Fig. 6.1  (a) Three-layer random media example; (b) corresponding velocity variance depth dependency.
Fig. 6.2  (a) 3-layer random media with the sinusoidal topography at the boundaries; (b) corresponding velocity variance depth dependency.
Returning to the discussion of the velocity variance peaks at several depths in the inversion, we can note the velocity variance peaks especially at the transition zone boundaries 400 km and 650 km. The transition zone is the layer between two the discontinuities due to polymorphic phase changes. At 400 km, the mineral phase changes from olivine to wadsleyite ($\alpha -$ to $\beta -$ phase of $Mg_2SiO_4$). At 650 km, it is more complicated and generally linked to the mineral phase transition from ringwoodite to
bridgmanite and periclase. Therefore a reasonable guess to explain the variance peaks is chemical heterogeneity or phase heterogeneity between the major phase boundaries.

Stixrude and Lithgow-Bertelloni (Stixrude and Lithgow-Bertelloni, 2007; Stixrude and Lithgow-Bertelloni, 2012) showed that the temperature derivative of the velocity is a functional characterization of phase heterogeneity in the deep earth. The isomorphous and metamorphic part is reproduced in a phase change diagram in Fig. 6.4. The phases included are: orthopyroxene (opx), clinopyroxene (cpx), high-pressure Mg-rich clinopyroxene (hpcpx), garnet (gt), olivine (ol), wadsleyite (wa), ringwoodite (ri), perovskite (pv), CaSiO$_3$ perovskite (capv), and ferropericlase (fp), and stishovite (st).
Fig. 6.4 Calculated phase equilibrium (Cammarano et al., 2003; Ita and Stixrude, 1992; Stixrude and Lithgow-Bertelloni, 2007). Red line assumes an adiabatic temperature profile, which can be considered as the isomorphic contribution. Blue lines are complete phase equilibrium, which can be considered as metamorphic contribution.
The metamorphic contribution to mantle heterogeneity is hard to recognize by traditional deterministic tomography, as shown in Fig 6.5.

Fig. 6.4 Temperature derivative of the shear wave velocity (Stixrude and Lithgow-Bertelloni, 2012) isomorphic (dash-line) and metamorphic (thick green-line) and 200km moving-average filtered to mimic the
tomography resolution (thin green-line), compared with tomography model SAW24B16 (Megnin and Romanowicz, 2000) and S20RTS (Ritsema et al., 2004).

If we compare our inversion to Stixrude and Lithgow-Bertelloni’s study (Stixrude and Lithgow-Bertelloni, 2007), as in Fig 6.5, we can see that the inverted spectra seem to capture the metamorphic contribution very well. This explains that velocity variance peaks in our inversion result are an effect of lithofacies metamorphosis. In this mechanism 3-D differences in temperature can move the heterogeneous collection of silicate minerals that comprise Earth’s mantle into different regions of their phase diagrams, affecting their elastic moduli and densities, and hence their seismic velocities.
Fig. 6.5  Inversion result of velocity variance (upper) compared with temperature derivative of velocity (Stixrude and Lithgow-Bertelloni, 2007) (lower). The red dashed line is isomorphic and blue line is the metamorphic contribution.
Chapter 7  Conclusion

By modeling the waveform of body waves propagating through Earth, traditional seismic tomography has become a powerful technique to retrieve and image Earth’s 3-D velocity structure. However, it fails to resolve small-scale heterogeneities. We use an alternative approach, stochastic tomography, to retrieve a statistical representation of structure from observations of the fluctuation of the amplitude and travel time of teleseismic body waves recorded by array sensors.

Instead of deterministic imaging, stochastic tomography is used to resolve the power spectrum of heterogeneous media. Throughout this thesis we present a thorough study of this technique, from theory to numerical validation and the application in characterizing small-scale heterogeneity in Earth’s mantle by inverting coherence functions.

We choose the seismic stations of the USAArray across the western US continent as our data source and study three earthquake groups in the Tonga, Japan and Chile areas. Fluctuations in amplitude and travel time of teleseismic P waves, measured by amplitude and phase coherences beneath elements of the EarthScope seismic array, are used to compute coherence functions across different stations or sources. These coherence functions are used to invert for the heterogeneity spectrum of P velocity in a 1000 km thick region of the upper mantle beneath the array.
Respectively, we invert the coherence functions under a single-layer assumption and a depth-dependent heterogeneity spectrum. The single-layer inversion result reveals a multi-scale of heterogeneities of characteristic length 280 km, and 60 km to 100 km. Power decays with increasing wave number. Also the geological structure may be self-affine at all probing scales. Best fits to joint transverse coherence functions require a depth-dependent heterogeneity spectrum, with velocity peaks in narrow depth intervals around 250 km, 420 km, 500 km and 600 km.

To understand the velocity variance peaks at different depths, we compared our inversion result with the estimated temperature derivative of seismic velocity, to show that this result is a signal of phase change in upper mantle. The shift in temperature will result in a shift in pressure-temperature boundaries in mineral phase diagram, indicating the lithofacies metamorphosis. Specifically, in this mechanism 3-D differences in temperature can move the heterogeneous collection of silicate minerals that comprise Earth’s mantle into different regions of their phase diagrams, affecting their elastic moduli and densities, and hence their seismic velocities. The 420 km and 600 km peaks are in accordance with the depths of known transition zone boundaries. The 250 km and 500 km signals are weaker but also predicted. They agree well with the predictions for a temperature derivative of velocity that includes the effects of chemical and phase variations expected for standard models of the silicate mineral assemblage of the upper mantle. The results confirm the existence of significant chemical as well as thermal contributions to observed upper mantle heterogeneity at spatial scales between 50 km to
300 km.

Compared with other seismic imaging approaches, the depth-dependent inversion result is the first example of seismic imaging that has well captured additional phase changes between the 400 and 650 km depth interval. This study shows a promising future for stochastic tomography. Besides seismic imaging, this technique could also be useful complement to other waveform inversion or modeling applications.
Bibliography


Sun, Z., Sun, J., Han, F., Wang, X., Tian, Y., and Qi, P., 2017, Multiples ray tracing using multistage group marching upwind hybrid method in regional multiscale


Zhang, Z., Tian, Y., and Fernando, G. W., 2017, From a $\mathbb{Z}_2$ topological insulator to a square ribbon: Tuning quantum spin Hall states and conductivity: Journal of Applied Physics, v. 121.


Zheng, Y., and Wu, R. S., 2008a, Theory of Transmission Fluctuations in Random Media with a Depth-Dependent Background Velocity Structure, v. 50, p. 21-41.
