

Non-stationary spherical random media and their effect on long-period mantle waves

Matthias Meschede¹ and Barbara Romanowicz^{1,2,3}

¹*Institut de Physique du Globe de Paris, Paris F-75005, France. E-mail: matthias.meschede@googlemail.com*

²*Collège de France, Paris F-75005, France*

³*University of California, Berkeley, CA, USA*

Accepted 2015 August 21. Received 2015 June 28; in original form 2015 February 24

SUMMARY

We present a method to construct non-stationary and anisotropic second-order random model realizations that can be used for numerical wave propagation simulations in various geometries. Models are generated directly from a given covariance matrix using its eigenvector decomposition (principal component or Karhunen-Loëve method). Because this decomposition is very expensive computationally in 3-D, we use model symmetries to reduce the size of the covariance matrix to its non-stationary components. Stationary components can then be described through their power spectrum, such that models with axisymmetric or spherically symmetric statistics can be generated from a 1-D covariance matrix. We focus in particular on models with spherically symmetric statistics that are important to study wave propagation in the Earth. We use this method to show the influence of hypothetical small-scale structure in the Earth's mantle on the elastic wavefield. To this end, we extend tomographic models beyond their spatial resolution limit with different distributions of small-scale scatterers that generate a coda and attenuate direct waves (scattering attenuation). We observe that scattering attenuation of fundamental mode Rayleigh waves is small (0.5–2 per cent of the total attenuation), if the elastic mantle structure does not become significantly stronger at smaller scales. At the examined heterogeneity strengths, scattering attenuation scales linearly with the model variance. The long-period fundamental mode Rayleigh wave coda is difficult to measure because it is weak and overlaps with other signals. However, it can be shown that its intensity also scales linearly with model power, and that it depends strongly on the spherical geometry of the Earth. It can therefore be used to distinguish between models with different small-scale power. We show qualitatively that the coda generated by the type of random models we consider can explain observed scattered energy at long periods (100 s).

Key words: Fourier analysis; Surface waves and free oscillations; Coda waves; Seismic attenuation; Statistical seismology; Wave scattering and diffraction.

1 INTRODUCTION

The seismic wavefield is strongly sensitive to structures with sizes down to at least half its wavelength. Heterogeneities of such sizes are often not resolved in global tomographic models of Earth's elastic structure. They affect the wavefield, but cannot be localized due to the sparse source-receiver distribution on the Earth. Nevertheless, they can be modelled in a statistical way, constraining for example their size and strength but not their exact location. Most notably, incoherently scattered waves (the coda), and attenuation of coherent direct waves (scattering attenuation) can be modelled using statistically distributed scatterers.

Based on analytical, semi-analytical or numerical models, statistical properties of small-scale heterogeneities in the Earth have been inferred from observations of inner core phases (Cormier

2002, 2007; Calvet & Margerin 2008), phases sensitive to core-mantle boundary and D'' structure (Cormier 1999; Margerin & Nolet 2003; Shearer & Earle 2008; Garcia *et al.* 2009; Mancinelli & Shearer 2013), and to upper mantle and crustal structure (Frankel & Clayton 1986; Sato 2002; Maeda *et al.* 2003; Calvet & Margerin 2013).

Scattering can be very strong in other planetary bodies, and cannot be neglected in seismic modelling. In the moon, for example, a strong scattering layer close to the surface is the primary cause of a very long seismic coda (Nakamura *et al.* 1975; Dainty & Toksöz 1981; Blanchette-Guertin *et al.* 2012).

Expressions that describe macroscopic statistical effects on the wavefield due to small-scale structure, like the coda or scattering attenuation, have been found for media with stationary (translation invariant) and normally distributed (second-order) statistics

(e.g. Weaver 1990; Ryzhik *et al.* 1996; Margerin 2006; Calvet & Margerin 2012; Sato *et al.* 2012). Such media can be characterized equivalently by a distance dependent covariance (autocorrelation) function or the Fourier power spectrum of its elastic properties (Wiener-Khinchine theorem).

Importantly, the spectral representation of a stationary covariance function can be used to generate random model realizations with well-defined strength and scale lengths by filtering white noise to the desired power (Frankel & Clayton 1986; Klimeš 2002). Due to its simplicity, this is the most commonly used method to analyse scattering effects numerically (e.g. Frankel 1989; Jahnke *et al.* 2008). Other approaches to construct random model realizations in complex geometries and with complex statistics have been studied primarily in the Geostatistics community (e.g. Stein 1999; Lindgren *et al.* 2011).

Although a stationary and isotropic (rotation invariant) medium can be a good approximation for a small region, it is not a good representation of the Earth as a whole, because its elastic properties vary strongly with depth (Dziewoński & Anderson 1981). Perturbations from the reference structure are often elongated in the horizontal direction (layered structures) or in the radial direction (e.g. plumes/slabs). The required, depth dependent, and radially anisotropic structures cannot be represented by the Fourier power spectrum because it is tied to stationary Cartesian models.

In the first part of this paper, we present a simple and fast method that overcomes these limitations. It can be used to generate realizations of 3-D random media in the full sphere, accounting for non-stationary and radially anisotropic media. As a special case, stationary and isotropic model realizations can be generated in spherical coordinates.

We start by introducing second-order random models and their covariance matrix that defines heterogeneity size and strength. Following that, we present three methods that can be used to construct such random models: (1) the fast and common spectral Fourier method that is suited for stationary Cartesian models (2) the slow, not so common Karhunen-Loève method that can be used to construct non-stationary random models. (3) A combination of both methods that uses the slow Karhunen-Loève approach only for non-stationary model components. Only the combination of both methods allows us to construct non-stationary models on 3-D grids of a useful size.

To use method (3) in a spherical geometry, we need to introduce an appropriate spectral description. To this end, we use a Fourier-Bessel expansion, and show how non-stationary or stationary spherical models can be generated in a similar manner as Cartesian models.

In the second part of the paper, we combine recent tomographic models with random model realizations that represent hypothetical unresolved structure to examine its effect on the propagation of seismic waves. In particular, we study scattering attenuation and the generation of a coda in fundamental mode Rayleigh waves that are commonly used to constrain attenuation models of the Earth. We focus on long-period waves, because they are only sensitive to structure that is at the resolution limit of recent tomographic models and not as much to structure that is significantly smaller.

Although anelastic mechanisms are generally accepted as the explanation for the time decay of seismic signals, the portion of scattering in attenuation models of the Earth is unclear. Estimates range from a few per cent (Sato 2002; Dalton *et al.* 2013) up to a dominant fraction (Yang *et al.* 2007; Ricard *et al.* 2014) of total attenuation. To further constrain scattering attenuation, we measure fundamental mode decay rates in statistical models

that span a large range of possible heterogeneity strengths and shapes.

Additionally, we measure the strength of the long-period coda and examine how it depends on the small-scale structure. Due to the long periods, we need to consider long lapse times where the spherical geometry of the Earth becomes important. Despite overlapping signals, a complex large-scale structure and a complex crustal model, we show that the long-period coda intensity can be used to distinguish between synthetic models with different small-scale structures.

We finally demonstrate that the extended model can generate scattered energy, which is not modelled in current tomographic models, but qualitatively similar to that seen in long-period seismograms. A rigorous comparison with data is beyond the scope of this paper.

2 SECOND-ORDER RANDOM MEDIA

2.1 Covariance matrix

The model values, $f(\mathbf{r})$, of a second-order (called ‘Gaussian’) random medium realization at different locations are by definition drawn from a multivariate normal distribution (for a review see e.g. Abrahamsen 1997). Such media can be characterized completely by their mean $\mu(\mathbf{r}) = \langle f(\mathbf{r}) \rangle$ and their symmetric and positive definite covariance function:

$$C(\mathbf{r}, \mathbf{r}') = \langle f(\mathbf{r})f(\mathbf{r}') \rangle \quad (1)$$

Note that $\langle \rangle$ signifies averaging over multiple model realizations. However, in terms of wave propagation, we are typically interested in spatially averaged properties of a single realization. Fortunately, averages over different realizations of a Gaussian random model can often be interpreted as spatial average (ergodicity): If the covariance function decays rapidly but does not change strongly with location, a local spatial average samples independent model regions with similar statistics. The averaged covariance function of these independent regions then correspond to the average over different realizations.

The mean can then be understood as invariant background structure, and the covariance function describes spatially varying fluctuations around it. The variance, or power, of these fluctuations at a specific location is specified by the diagonal entries $C(\mathbf{r}, \mathbf{r})$. The off-diagonal covariance of two different points, $C(\mathbf{r}, \mathbf{r}')$, describes their shape. We focus in the following on zero mean media, and add independent background structure later.

Alternatively, C can be written in terms of mid-point $\bar{\mathbf{r}} = \frac{1}{2}(\mathbf{r} + \mathbf{r}')$ and distance $\mathbf{d} = \mathbf{r} - \mathbf{r}'$ coordinates:

$$C(\mathbf{r}, \mathbf{r}') = C\left(\bar{\mathbf{r}} + \frac{\mathbf{d}}{2}, \bar{\mathbf{r}} - \frac{\mathbf{d}}{2}\right), \quad (2)$$

where $C(\mathbf{r}, \mathbf{r}')$ is called ‘stationary’ if it is invariant under translation, and it is called ‘isotropic’ if it is invariant under rotation. In the stationary case, it can be written as $C(\mathbf{d})$, and in the isotropic case as $C(|\mathbf{d}|)$.

Both stationarity and isotropy are bound to a coordinate system and an associated measure of distance, that define rotation and translation: Instead of translation along the Cartesian coordinate axes, we could for example understand translation as a shift along the angular and radial axes of a spherical coordinate system. Also angular (arc-distance) and radial distances (difference of radial coordinates) could be considered instead of the usual (Euclidean) distance between two points. A model that is stationary with respect

to such spherical translations is naturally non-stationary with respect to the corresponding Cartesian operation, because the relation between Cartesian and angular distances depends on (radial) location. Therefore, the notion of translation, rotation and distance controls in which coordinate system a model is stationary or isotropic. To avoid confusion, we write for instance ‘radially anisotropic’ or ‘radially stationary’. ‘Isotropic’ and ‘stationary’ refer to a model that is invariant under rotation and translation in the Cartesian sense.

Usually, we construct random model realizations on discrete grids in a finite spatial domain. To this end, we sample the continuous coordinates \mathbf{r} at points \mathbf{r}_i , assuming that the medium is smooth enough to do so. The covariance function (see eq. 1) becomes the covariance matrix:

$$\mathbf{C} = \langle \mathbf{f}\mathbf{f}^T \rangle. \quad (3)$$

Its elements $C_{ij} = \langle f_i f_j \rangle$ are computed from the outer product of the model values $f_i = f(\mathbf{r}_i)$ (T indicates the matrix or vector transpose). Even though we always generate models from a discrete covariance matrix, we use the continuous viewpoint to simplify calculations, and discretize in the end. Furthermore, we do not consider the finite domain size in the continuous equations (integrating over \mathbb{R} instead of the finite domain). This discretization imposes small-scale limits on the models that we describe, and we need to be aware of aliasing and boundary effects.

2.2 Limitations and applicability

Gaussian random models can only describe the second-order correlations of model features. Higher order terms are neglected (e.g. $\langle f(\mathbf{r})f(\mathbf{r}')f(\mathbf{r}'') \rangle$). This is not a fundamental requirement and, for example, more general statistical descriptions take into account the correlation of whole neighbourhoods (e.g. Meerschman *et al.* 2013). However, the simplicity and mathematical accessibility of Gaussian models make them the backbone of many studies. They allow us to focus on the most basic information of heterogeneity shape and strength.

A stronger argument for the important role of these models comes from their ability to predict the covariance of the incoherently scattered wavefield signals: In a medium with weak perturbations, the amplitude of individual scattered waves is linearly related to structural perturbations. The scattered wavefield can be seen as the superposition of many independently scattered waves. In this case, the covariance of the scattered wavefield depends only, and linearly, on the covariance of the medium (e.g. Weaver 1990; Ryzhik *et al.* 1996). Additionally, independently scattered signals become multivariate normally distributed when they overlap and mix, as described by the central limit theorem. Higher order correlations become difficult to observe and are eventually lost. Gaussian models therefore capture the most significant statistical information about a medium with weak perturbations that can be obtained from overlapping, incoherently scattered waves.

We expect the following relations between structure and wavefield in the linear regime that we are going to assess in the later parts of the paper: (1) The (co)variances—not amplitudes—of structure and scattered wavefield are linearly related. (2) Sparse or dense distributions of scatterers have a similar averaged effect on the wavefield, as long as their averaged (co)variance is similar. (3) Extrinsic attenuation, as a measure of dissipated energy, is proportional to the (co)variance (energy) of the scattered wavefield and therefore of the medium. These three linear relations are well known from the analysis of stationary media (e.g. Sato *et al.* 2012).

3 RANDOM MODELS IN CARTESIAN COORDINATES

In this section, we present two different methods to generate random models and a third one which combines the previous two: first we show how to generate random media from a stationary covariance matrix using a spectral description (Fourier method). Second, we present the Karhunen-Loëve method that can be used to generate random models from a general, non-stationary covariance matrix. We then combine both methods such that we can use the fast spectral description for stationary components and the slow Karhunen-Loëve method only for non-stationary components.

3.1 The Fourier method

In Cartesian coordinates, and in particular on a discrete evenly sampled Cartesian grid, a model $f(\mathbf{r})$ can easily be decomposed into its Fourier spectrum using the transform pair:

$$F(\mathbf{k}) = \int_{\mathbb{R}^3} f(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r} \quad (4)$$

$$f(\mathbf{r}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} F(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k}. \quad (5)$$

Here, $F(\mathbf{k})$ is the Fourier spectrum of $f(\mathbf{r})$ and \mathbf{k} is a wavenumber vector in Fourier space. The covariance of the spectrum is related to the covariance of the model (* signifies the complex conjugate):

$$\langle F^*(\mathbf{k})F(\mathbf{k}') \rangle = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \langle f(\mathbf{r})f(\mathbf{r}') \rangle e^{i\mathbf{k}\mathbf{r}-i\mathbf{k}'\mathbf{r}'} d\mathbf{r} d\mathbf{r}' \quad (6)$$

$$= \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} C(\mathbf{r}, \mathbf{r}') e^{i\mathbf{k}\mathbf{r}-i\mathbf{k}'\mathbf{r}'} d\mathbf{r} d\mathbf{r}'. \quad (7)$$

By changing from \mathbf{r}, \mathbf{r}' to midpoint and distance coordinates $\bar{\mathbf{r}}, \mathbf{d}$ (see eq. 2), eq. (7) can be interpreted as a spatial averaging operation:

$$\langle F^*(\mathbf{k})F(\mathbf{k}') \rangle = \int_{\mathbb{R}^3} e^{i((\mathbf{k}-\mathbf{k}')\bar{\mathbf{r}}} d\bar{\mathbf{r}} \int_{\mathbb{R}^3} C(\bar{\mathbf{r}}, \mathbf{d}) e^{i(\mathbf{k}+\mathbf{k}')\frac{d}{2}} d\mathbf{d} \quad (8)$$

Eq. (8) shows that the model’s power spectrum (i.e. $\langle |F(\mathbf{k})|^2 \rangle$) is the Fourier transform of the spatially averaged covariance function ($\int d\mathbf{d}$). If the covariance function decays sufficiently fast, the spectrum $F(\mathbf{k})$ is an average of C in independent model regions which is equivalent to averaging over multiple realizations. In this sense, we can use random models to construct a single realization with the desired spatial statistics.

If C is stationary and does not depend on $\bar{\mathbf{r}}$, eq. (8) simplifies further to:

$$\langle F^*(\mathbf{k})F(\mathbf{k}') \rangle = \delta(\mathbf{k} - \mathbf{k}')(2\pi)^3 \int_{\mathbb{R}^3} C(\mathbf{d}) e^{i\mathbf{k}\mathbf{d}} d\mathbf{d}. \quad (9)$$

Eq. (9) implies that the Fourier coefficients of a stationary model are uncorrelated. A stationary model realization can therefore be generated from independent Fourier coefficients (white noise) that are filtered to the required power spectrum. Due to its simplicity and numerical efficiency, this procedure is ubiquitous in Geophysics.

Representative covariance functions, such as Scale-Free, Exponential, Gaussian, von-Karman or the more general Matern, with their respective power spectral representations are often chosen to categorize different types of media. An overview of these can be found in Klimeš (2002) or Stein (1999). Fig. 1 shows an example of a 3-D random model realization with an exponential

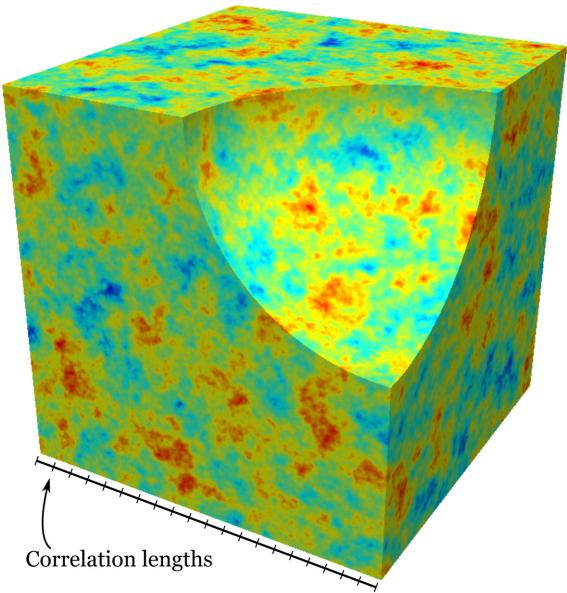


Figure 1. Second-order random model in Cartesian coordinates with a stationary exponential covariance function (tick marks indicate the correlation length a). This model was computed by filtering independent normally distributed Fourier coefficients (i.e. white noise) to the power spectrum that is given by the Fourier transform of the covariance function.

covariance function that was generated using this technique ($C(|\mathbf{d}|) = \sigma^2 \exp(-|\mathbf{d}|/a)$, $\langle |F(\mathbf{k})|^2 \rangle = 8\pi\sigma^2 a^3 / (1 + (|\mathbf{k}|^2 a^2)^2)$ with variance $\sigma^2 = 1$, correlation length $a = 1/20$, cube edge length 1).

Apart from its focus on second-order correlations that we have mentioned before, two principal limitations of this method are:

(1) It relies on the Fourier transform which is bound to Cartesian coordinates. In particular, only anisotropy with respect to the Cartesian axes can be described. This limitation can be overcome by introducing spectral transforms that are adapted to the required coordinate system, like Fourier–Bessel transforms in polar and spherical coordinates.

(2) It is limited to stationary media. We overcome this limitation and construct non-stationary models by directly generating correlated normally distributed random variables from a given covariance matrix.

3.2 The Cholesky and Karhunen-Loëve methods

To generate a model $\mathbf{f} = \{f(\mathbf{r}_i)\}$ with covariance matrix $C(\mathbf{r}_i, \mathbf{r}_j)$, we first generate a set of independent, normally distributed variables \mathbf{y} with unit variance using a pseudo-random generator, and multiply them with a matrix \mathbf{M} to correlate them. The model \mathbf{f} can then be written as:

$$\mathbf{f} = \mathbf{My}. \quad (10)$$

The associated model covariance matrix \mathbf{C} is the ‘square’ of matrix \mathbf{M} :

$$\mathbf{C} = \langle \mathbf{My}^T \mathbf{M}^T \rangle = \mathbf{MM}^T. \quad (11)$$

For a given covariance function, we can therefore find the matrix \mathbf{M} by decomposing \mathbf{C} into a matrix multiplied by its transpose. Different possibilities exist for such a decomposition:

(1) The symmetric and positive definite covariance matrix can be written as $\mathbf{C} = \mathbf{LL}^*$ using the ‘Cholesky’ decomposition \mathbf{L} and its

conjugate transpose \mathbf{L}^* . \mathbf{f} can then be generated from the independent random variables \mathbf{y} using:

$$\mathbf{f} = \mathbf{L} \cdot \mathbf{y}. \quad (12)$$

This decomposition exists for symmetric, positive definite matrices.

(2) Alternatively, \mathbf{C} can be decomposed into the diagonal eigenvalue matrix $\mathbf{\Lambda} = \{\lambda_i \delta_{ij}\}$ and its eigenbasis $\mathbf{E} = \{E_{ij}\}$:

$$\mathbf{C} = \mathbf{E}^T \mathbf{\Lambda} \mathbf{E}. \quad (13)$$

Because \mathbf{C} is positive definite and symmetric, the eigenvalues λ_i are positive and real and the eigenbasis is orthogonal. With this decomposition, a random model can be generated from:

$$\mathbf{f} = \mathbf{E}^T \sqrt{\mathbf{\Lambda}} \mathbf{y} \quad (14)$$

with the diagonal matrix $\sqrt{\mathbf{\Lambda}}_{ij} = \sqrt{\lambda_i} \delta_{ij}$. The orthogonal eigenbasis \mathbf{E} is often called a ‘Karhunen-Loëve’ basis that transforms uncorrelated, independent variables into correlated ones. To our knowledge, this method is not widely used in seismology to construct random model realizations but it has at least been proposed and tested (e.g. Thorne *et al.* 2006).

As the sum of independent normally distributed variables, the random variables \mathbf{f} are also normally distributed. Of the two decompositions (1) and (2), the Cholesky decomposition is often simpler to compute than the eigendecomposition of the covariance matrix. On the other hand, knowledge of the eigenvalues and eigenvectors of the covariance matrix provides additional flexibility: For example, we can truncate the eigenbasis and consider only its strongest components. Moreover, the eigendecomposition can be applied if the matrix \mathbf{C} is not positive definite. Although a covariance matrix should always be positive definite, eigenvalues that are close to zero can become negative due to numerical errors. Such negative small eigenvalues can then be removed, together with their associated eigenvectors, without changing the covariance matrix \mathbf{C} much.

An example of this procedure is shown in Fig. 2. In this 1-D example, we provide a non-stationary covariance matrix that relates model values at points r, r' (Fig. 2a), and decompose it into its eigenbasis (Fig. 2b). With help of this eigenbasis, we can correlate independent random variables and generate a model realization that reflects the heterogeneity scales of the covariance matrix (Fig. 2c).

In 3-D, a particularly simple case arises when the covariance function is separable. In Cartesian coordinates, this means that it can be factorized:

$$C(x, x', y, y', z, z') = C(x, x')C(y, y')C(z, z'). \quad (15)$$

In discrete language, the covariance matrix can then be written as an outer product of smaller submatrices:

$$\mathbf{C} = \mathbf{C}_x \otimes \mathbf{C}_y \otimes \mathbf{C}_z. \quad (16)$$

The complete eigenbasis of \mathbf{C} is then also simply the outer product of the eigenbases of all submatrices.

In case of a 1-D stationary medium that is discretized with equidistant points, the covariance matrix is of Toeplitz structure (constant along descending diagonals). Multiplication with such a matrix corresponds to a convolution, and it is well known that the Fourier basis diagonalizes such a matrix. Therefore the Fourier method that we described in Section 3.1 can be seen as a particularly efficient way of diagonalizing the covariance matrix.

3.3 A combined approach

The principal advantage of the Fourier method is its efficiency. Diagonalization of the full covariance matrix becomes quickly intractable

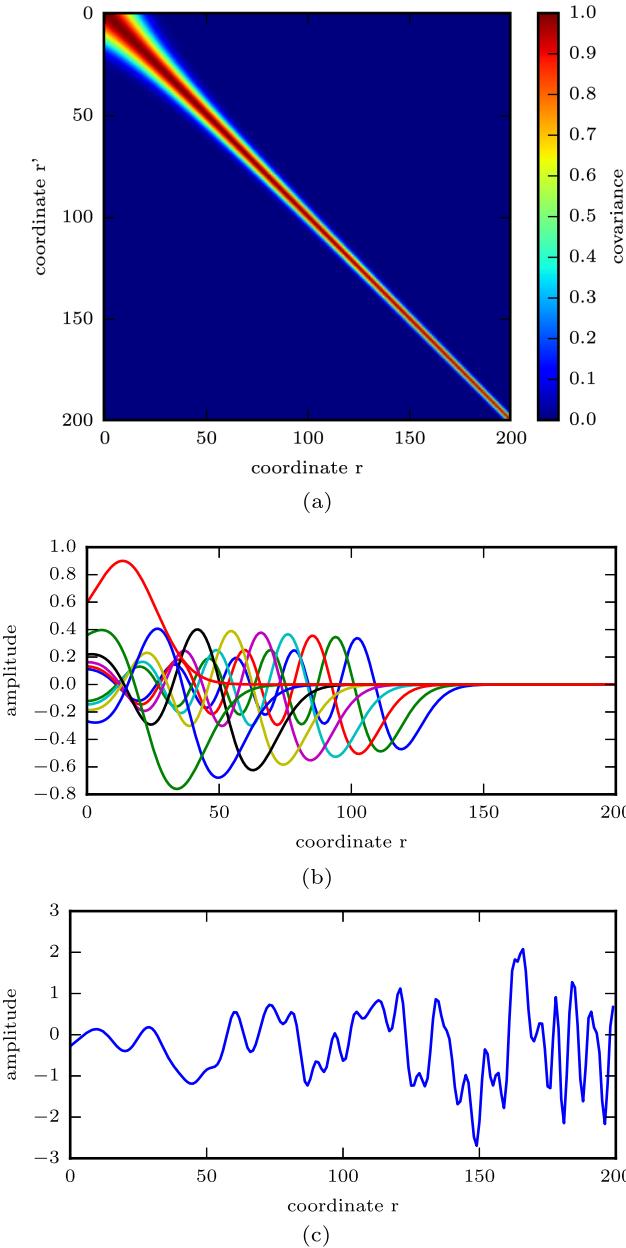


Figure 2. (a) Non-stationary covariance matrix. The scale length of the model becomes smaller with increasing r . (b) Eigenbasis (incomplete) of the covariance matrix from (a). The orthogonal eigenbasis reflects the scale lengths of the covariance matrix. (c) Random model realization that was generated by multiplying independent random variables with the eigenbasis (b). Scale length of the model realization becomes smaller with increasing r .

in realistic model scenarios. In 3-D, for example, the number of model grid points scales with the three dimensions, its covariance matrix with six. A tiny double precision grid of size $32 \times 32 \times 32$ already requires $32^6 * 8 \text{ bytes} \approx 8 \text{ Gb}$ of memory just to store the covariance matrix.

Fortunately, in many situations, a non-stationary covariance matrix has stationary components. These can be treated using the spectral Fourier method to reduce the dimensionality of the diagonalization of the covariance matrix to its non-stationary components. The ‘reduced’ covariance matrix no longer describes the covariance of point pairs but of spectral coefficients that represent stationary

surfaces or lines. In this section, we outline how the two previously presented methods can be combined.

In Cartesian coordinates, for example, the forward and inverse Fourier transforms can be used only for the subset of all spatial dimensions that are considered stationary. For example, with $\mathbf{r}_{xy} = (x, y)$ and $\mathbf{k}_{xy} = (k_x, k_y)$, a model $f(\mathbf{r})$ can be written as:

$$f(\mathbf{k}_{xy}, z) = \int_{\mathbb{R}^2} f(\mathbf{r}) e^{-i\mathbf{k}_{xy} \cdot \mathbf{r}_{xy}} d\mathbf{r}_{xy} \quad (17)$$

$$f(\mathbf{r}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} f(\mathbf{k}_{xy}, z) e^{i\mathbf{k}_{xy} \cdot \mathbf{r}_{xy}} d\mathbf{k}_{xy}, \quad (18)$$

where $f(\mathbf{k}_{xy}, z)$ is the Fourier spectrum of the xy -plane as a function of the z coordinate. It can also be computed from the 3-D Fourier transform $F(\mathbf{k})$ of the model:

$$f(\mathbf{k}_{xy}, z) = \frac{1}{2\pi} \int_R F(\mathbf{k}) e^{ik_z z} dk_z. \quad (19)$$

If $f(\mathbf{r})$ is a random model with a covariance matrix that is stationary in the xy -plane, its \mathbf{k}_{xy} Fourier coefficients are independent (see eq. 8). On the other hand, Fourier coefficients that correspond to the same horizontal basis \mathbf{k}_{xy} on different planes can be correlated. Their covariance can be computed from the 3-D covariance matrix (equivalent to eq. 8):

$$C(\mathbf{k}_{xy}, z, z') = \int_{\mathbb{R}^2} d\bar{\mathbf{r}}_{xy} \int_{\mathbb{R}^2} dd_{xy} C(\mathbf{d}_{xy}, z, z') e^{i\mathbf{k}_{xy} \cdot \mathbf{d}_{xy}} \quad (20)$$

or from a given 3-D Fourier spectrum:

$$C(\mathbf{k}_{xy}, z, z') = \langle f(\mathbf{k}_{xy}, z) f(\mathbf{k}_{xy}, z') \rangle \quad (21)$$

$$= \frac{1}{(2\pi)^2} \int_R dk_z \langle |F(\mathbf{k})|^2 \rangle e^{ik_z(z-z')}. \quad (22)$$

This allows us to define the expected 2-D horizontal power spectra for classical choices such as Gaussian, exponential or von-Karman: The covariance function of coefficient \mathbf{k}_{xy} is the Fourier transform of the 3-D power spectrum along the z -axis. An isotropic power spectrum at position k_x, k_y is, for example, transformed along a hyperbola: $|F(\sqrt{k_x^2 + k_y^2 + k_z^2})|^2$. Therefore, the covariance function of the horizontal Fourier basis function \mathbf{k}_{xy} is a distorted version of the original covariance function that reflects scales with $|\mathbf{k}|^2 \geq k_x^2 + k_y^2$.

The partial covariance matrix (eq. 21) breaks the full covariance matrix into independent sub-matrices for each horizontal basis function $e^{i\mathbf{k}_{xy} \cdot \mathbf{r}_{xy}}$. Each $C(\mathbf{k}_{xy}, z, z')$ provides an independent eigenbasis for each horizontal Fourier coefficient. With this eigenbasis and the corresponding eigenvalues, correlated coefficients can be generated (eq. 14) for all \mathbf{k}_{xy} . These coefficients are finally transformed into a model realization (eq. 18). The size of the 1-D sub-matrices is much smaller than that of the original 3-D covariance matrix. The complexity of the original problem is therefore drastically reduced by this procedure.

If we additionally constrain the model to be isotropic in the xy -plane, the xy spectrum depends only on $\rho_{xy} = \sqrt{k_x^2 + k_y^2}$ and the size of the problem reduces further. Under these assumptions, ρ_{xy} is the only remaining dimension and a polar variant of the Fourier transform (Hankel transform) can be used for random model generation (e.g. Baddour 2009).

Finally, after all simplifications, we only need to diagonalize less than 32 matrices of size 32×32 to construct a model on a regular grid with dimensions $32 \times 32 \times 32$. This is a small problem compared to what we had at the beginning of this section. Much larger

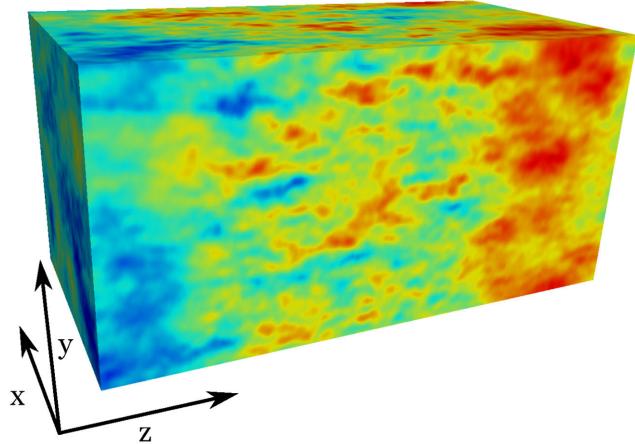


Figure 3. Second-order random model in Cartesian coordinates. This model has an exponential covariance function that is stationary in the xy -plane and non-stationary and anisotropic with respect to the z component. The statistics of the xy components are described by their 2-D Fourier power spectrum and the statistics of the z component by the covariance of the xy -Fourier coefficients.

model sizes can be handled as well because the problem scales much better. If we define the covariance matrix of horizontal planes according to eq. (22), we can recreate well-known 3-D random media. However, in contrast to the full 3-D Fourier method (Section 3.1) we now have full control over the horizontal statistics and can easily change the covariance function as a function of depth.

As an example, we have constructed the 3-D model realization, shown in Fig. 3 from a non-stationary covariance matrix that is locally exponential and anisotropic. Only a single component is non-stationary and changes the model statistics.

4 RANDOM MODELS IN SPHERICAL COORDINATES

4.1 Spherical covariance matrices

In a spherical coordinate system, the covariance matrix of a Gaussian random medium can be written as $C(r, r', \phi, \phi', \theta, \theta')$. This is useful if statistical properties of the medium depend on the radial and angular locations and distances. Furthermore, in case of a medium with spherically symmetric statistics, the angular component of C is stationary and it can be written as $C(r, r', \Delta)$, where Δ is the arc distance between two points.

In the following, we discuss first the spherical harmonics basis that breaks a 3-D covariance matrix with stationary angular component into independent 1-D submatrices $C_l(r, r')$. This is completely analogous to the previously discussed partial Fourier transform. We then discuss how $C_l(r, r')$ looks like if it is stationary in the 3-D Cartesian sense, and how it can be computed from a given Fourier power spectrum. Two examples are shown that illustrate how stationary, and radially non-stationary models can be generated from $C_l(r, r')$ in spherical coordinates.

In this section, we distinguish between the ‘sphere’, that is, the set of all points with fixed distance from the origin, and the ‘ball’, that is, the full volume that is enwrapped by the sphere. Using $C_l(r, r')$, we aim to describe the 3-D ball as a set of correlated spheres (think of the ‘onion’ analogy).

4.2 The angular covariance function

Similar to the Fourier basis in Cartesian coordinates, spherical harmonic functions form a suitable spectral basis on the sphere. They can be written as:

$$Y_{lm}(\theta, \phi) = \sqrt{(2l+1) \frac{(l-m)!}{(l+m)!}} P_{lm}(\cos(\theta)) e^{im\phi} \quad (23)$$

Here P_{lm} is the associated Legendre polynomial of degree l and order m . The $Y_{lm}(\theta, \phi)$ form an orthogonal basis and their normalization is chosen such that their variance is one (4π normalization). At a given radius, a function $f(\mathbf{r}) = f(r, \theta, \phi)$, can then be expanded as:

$$f(\mathbf{r}) = \sum_{lm} f_{lm}(r) Y_{lm}(\theta, \phi) \quad (24)$$

with coefficients $f_{lm}(r)$:

$$f_{lm}(r) = \frac{1}{4\pi} \int_{\Omega} f(r, \theta, \phi) Y_{lm}^*(\theta, \phi) d\Omega \quad (25)$$

‘ $\int_{\Omega} d\Omega$ ’ indicates integration over the spherical surface that is parameterized by θ and ϕ .

The covariance of two spherical layers of $f(\mathbf{r})$ at radii r^i and r^j can be computed from the depth dependent spherical harmonics coefficients $f_{lm}^i = f_{lm}(r^i)$:

$$\frac{1}{4\pi} \int_{\Omega} f(r^i, \theta, \phi) f(r^j, \theta, \phi) d\Omega = \sum_{lm} f_{lm}^{i*} f_{lm}^j = \sum_l C_l^{ij}, \quad (26)$$

where C_l^{ij} is the covariance matrix $C_l(r^i, r^j)$ of the spherical harmonics coefficients of two ‘onion’ layers. The diagonal C_l^{ii} or $C_l(r, r)$ is the ordinary spherical harmonics power spectral density (per degree l) of a horizontal model section.

Importantly, the spherical harmonics coefficients of an isotropic, stationary Gaussian random field on the sphere are independent and Gaussian distributed, just as the Fourier basis is in Cartesian coordinates (e.g. Baldi & Marinucci 2007):

$$\langle f_{lm}^i, f_{l'm'}^j \rangle = \frac{C_l^{ii}}{2l+1} \delta_{ll'} \delta_{mm'}. \quad (27)$$

Stationary and isotropic means in this context that the angular covariance function of the model is invariant under rotation and translation on the spherical surface.

In this case, $C(\mathbf{r}, \mathbf{r}')$ at $\mathbf{r} = r, \theta, \phi$ and $\mathbf{r}' = r', \theta', \phi'$ with similar distance from the centre and arc distance Δ , can be directly calculated from the spherical harmonics power spectrum, and the Legendre polynomial P_l of degree l using the well-known spherical harmonics addition theorem:

$$\langle f(\mathbf{r}) f(\mathbf{r}') \rangle = \sum_{lm'l'm'} \langle f_{lm}^*(r) f_{l'm'}(r') \rangle Y_{lm}^*(\mathbf{r}) Y_{l'm'}(\mathbf{r}') \quad (28)$$

$$= \sum_l C_l(r, r') P_l(\cos(\Delta)). \quad (29)$$

The spherical harmonics power spectrum characterizes $C(\Delta)$ on the spherical surface similar to an isotropic 2-D Fourier power spectrum that characterizes $C(|\mathbf{d}|)$. In case of a non-stationary model, the power spectrum is connected to the ‘angular average’ of the angular covariance function.

In the small-scale and high-degree limit, the spherical surface can be treated as a plane, and the Legendre polynomials P_l become similar to the Bessel functions J_l (Watson 1966; Holschneider 1996):

$$P_l(\cos(\Delta)) \rightarrow J_l(l\Delta) \quad (\text{for } \Delta/l \rightarrow 0). \quad (30)$$

Bessel functions are the radial basis of a polar Fourier transform (Hankel transform) and can be associated with the planar covariance function (e.g. Klimeš 2002). As a consequence, random media with rapidly decaying covariance functions have the same power spectra on the sphere and on the plane (if the normalizations are chosen appropriately).

In this limit, classic choices of covariance function and power spectra are therefore similar on the sphere and on the plane, and the diagonal part of the covariance function $C_l^{ii} = C_l(r^i, r^i)$ that describes the horizontal part of the full covariance function, can be easily understood. The non-diagonal entries that depend on the radial covariance function are the focus of the next section.

4.3 Covariance functions in 3-D spherical coordinates

A model with stationary Cartesian covariance matrix $C(\mathbf{r}, \mathbf{r}')$ has a radially non-stationary covariance function $C_l(r, r')$ because the wavelengths that are associated to the spherical harmonics degrees l depend on radius. In this section, we compute the non-stationary covariance function $C_l(r, r')$ from a given Cartesian Fourier spectrum $F(\mathbf{k})$.

To this end we transform Fourier space, that is the space of Fourier basis functions designated by the wavenumber vector $\mathbf{k} = (\rho, \theta_k, \phi_k)$, into a spherical harmonics basis. This procedure leads to a new set of ‘spherical wave’ basis functions that are superpositions of the original ‘plane wave’ basis. Explicitly, the Fourier basis $e^{i\mathbf{k}\cdot\mathbf{r}}$ can be expressed in a 4π -normalized spherical harmonics basis and spherical Bessel function $j_l(\rho)$ basis as (‘spherical wave expansion’):

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{l=0}^{\infty} \sum_{m=-l}^l (-i)^{-l} j_l(\rho r) Y_{lm}(\theta, \phi) Y_{lm}^*(\theta_k, \phi_k). \quad (31)$$

With this equation and spherical harmonics coefficients $f_{lm}(r)$ and $F_{lm}(\rho)$, we can rewrite the 3-D Fourier transform pair (eqs 4 and 5) as the following Fourier–Bessel transform pair:

$$F(\mathbf{k}) = \sum_{lm} (-i)^{-l} \left[\int dr 4\pi r^2 j_l(\rho r) f_{lm}(r) \right] Y_{lm}(\theta_k, \phi_k) \quad (32)$$

$$f(\mathbf{r}) = \frac{1}{(2\pi)^3} \sum_{lm} (-i)^l \left[\int d\rho 4\pi \rho^2 j_l(\rho r) F_{lm}(\rho) \right] Y_{lm}(\theta, \phi). \quad (33)$$

This transform pair relates Fourier space in spherical coordinates with model space in spherical coordinates. From eq. (33), we can also directly find the horizontal spherical harmonics coefficients $f_{lm}(r)$ that correspond to a model with a given Fourier spectrum:

$$f_{lm}(r) = (-i)^l \frac{4\pi}{(2\pi)^3} \int_0^\infty F_{lm}(\rho) j_l(\rho r) \rho^2 d\rho. \quad (34)$$

In spherical coordinates, the covariance of Cartesian Fourier coefficients of a stationary isotropic model can be written as:

$$\langle F(\mathbf{k}) F(\mathbf{k}') \rangle = |F(\rho)|^2 \delta(\mathbf{k} - \mathbf{k}') \quad (35)$$

$$= \frac{\langle |F(\rho)|^2 \rangle}{\rho^2 \sin \theta_k} \delta(\rho - \rho') \delta(\theta_k - \theta'_k) \delta(\phi_k - \phi'_k). \quad (36)$$

From this expression, we can compute the covariance of the Fourier coefficients $F(\mathbf{k})$ expanded in spherical harmonics using eq. (25) and get:

$$\langle F_{lm}^*(\rho) F_{l'm'}(\rho') \rangle = \frac{\langle |F(\rho)|^2 \rangle}{4\pi \rho^2} \delta_{ll'} \delta_{mm'} \delta(\rho - \rho'). \quad (37)$$

Therefore, an isotropic, stationary random model has uncorrelated Fourier–Bessel coefficients. Their variance corresponds to the Cartesian power $\langle |F(\rho)|^2 \rangle$, but it is now distributed over the spherical surface $4\pi \rho^2$. Eqs (33) and (37) can be used to generate isotropic and stationary random models in spherical coordinates from a given Cartesian Fourier power spectrum.

We can also compute the covariance function of the coefficients $f_{lm}(r)$ using eq. (34):

$$C_{lm}(r, r') = \langle f_{lm}^*(r) f_{lm}(r') \rangle \quad (38)$$

$$= \frac{(4\pi)^2}{(2\pi)^6} \int_0^\infty \langle |F_{lm}(\rho)|^2 \rangle j_l(\rho r) j_l(\rho r') \rho^4 d\rho. \quad (39)$$

This equation can be expressed in terms of the Cartesian power spectrum (eq. 37) as:

$$C_l(r, r') = \frac{4\pi}{(2\pi)^3} \int_0^\infty \langle |F(\rho)|^2 \rangle j_l(\rho r) j_l(\rho r') \rho^2 d\rho. \quad (40)$$

This matrix depends only on the coefficient degree l and is independent of the coefficient order m , which is a direct consequence of the horizontal isotropy of the model. Due to the shape of the spherical Bessel functions j_l , it becomes non-zero only above a certain radius which reduces the effective size of the covariance matrix and the complexity of the eigendecomposition.

Knowledge of $C_l(r, r')$ allows us to construct spherical random models that are stationary and isotropic in the Cartesian sense using its eigendecomposition. In addition, we can modify this matrix, for example, to introduce radial anisotropy or change the spectrum in different regions.

In 2-D polar coordinates (cylindrical symmetry), this procedure is essentially the same (see the Appendix)

4.4 Examples

As a first example, we construct a stationary and isotropic random medium in spherical coordinates. To this end, we choose the power spectrum of a 3-D exponential medium and compute the covariance matrix $C_l(r^i, r^j)$ of a discrete grid by numerical integration of eq. (40). We then decompose \mathbf{C}_l into its eigenvectors \mathbf{E}_l and eigenvalues λ_j (eq. 13). From the eigenbasis, we remove eigenvectors with very small and possibly negative eigenvalues. The total power of the removed eigenvalues should be small compared to the remaining eigenvalues to ensure that we are really reproducing a model with the desired covariance matrix. With the truncated \mathbf{E}_l , we generate correlated coefficients f_{lm}^i from independent coefficients that are normally distributed with unit variance y_{lm}^j .

$$f_{lm}^i = \sum_j \sqrt{\lambda_j} E_l^{ij} y_{lm}^j. \quad (41)$$

The model values at radius r^i are obtained from a spherical harmonics transform of the coefficients f_{lm}^i .

The corresponding model realization is shown in Fig. 4, next to a model that was constructed in Cartesian coordinates. The feature sizes and strengths in both models are visually indistinguishable and no sign of the underlying polar grid remains in the spherical model.

More interesting is an example with a non-stationary covariance function, shown in Figs 5 and 6. Here, we construct a spherical random medium that has three radial regions: The first region, around the origin, has an isotropic and stationary Gaussian covariance function. For the second region, at intermediate radii, we choose another

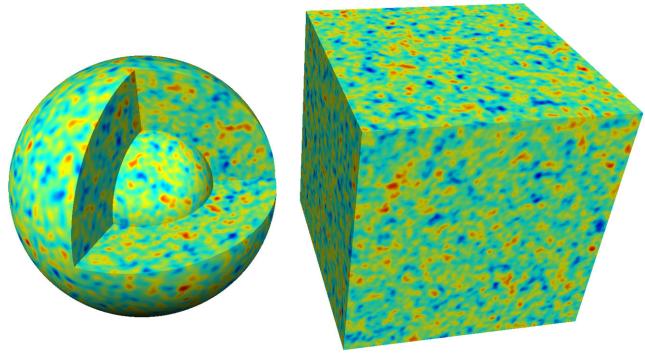


Figure 4. Isotropic second-order random model with exponential autocorrelation function in spherical (left) and in Cartesian (right) coordinates. In contrast to the Cartesian model, the spherical description provides control over the angular and radial statistics of the model.

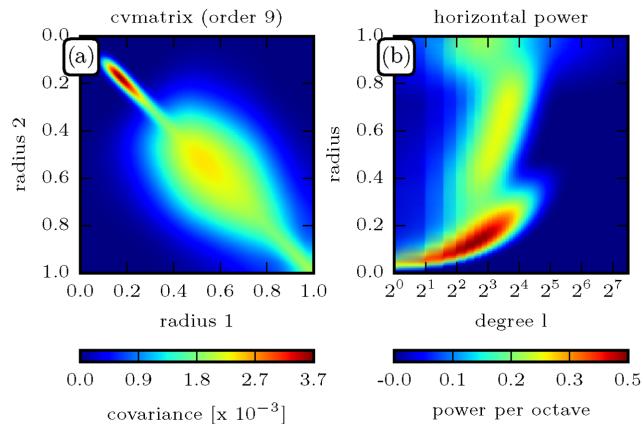


Figure 5. (a) Non-stationary, anisotropic covariance matrix that correlates spherical harmonics with degree 9. Three regions can be identified in the covariance function. In particular, it can be seen that the radial scale length becomes large at intermediate radii. The regions can also be seen in the horizontal power spectra (b). The angular scale lengths become shorter with increasing radius within a stationary region. Furthermore, we can see that horizontal scale lengths are smallest at the centre and largest at the surface. At the centre itself, all model power is in degree 0. The model realization reflects these trends (Fig. 6).

Gaussian covariance function with different scale length and radially elongated (anisotropic) features. In the third region close to the surface, the medium has an exponential covariance function and its features are elongated horizontally.

For each region we generate a covariance matrix C_l^i from eq. (40). Radial anisotropy is introduced by stretching or squeezing the off-diagonal entries away or towards the diagonal. This can be written as a transform $C_{\text{aniso}}(r, r') = C_{\text{iso}}(r_1, r_2)$ with:

$$r_1 = (1 + k) \frac{r}{2} + (1 - k) \frac{r'}{2} \quad (42)$$

$$r_2 = (1 + k) \frac{r'}{2} + (1 - k) \frac{r}{2}. \quad (43)$$

This leaves the horizontal covariance function (diagonal entries) unchanged and increases or decreases the radial scale length. We call k the ‘squeeze factor’. The correlation lengths in a model with squeeze factor k are k times shorter in radial than in horizontal direction. Therefore $k > 1$ corresponds to layered, $k < 1$ to vertical structures in the model.

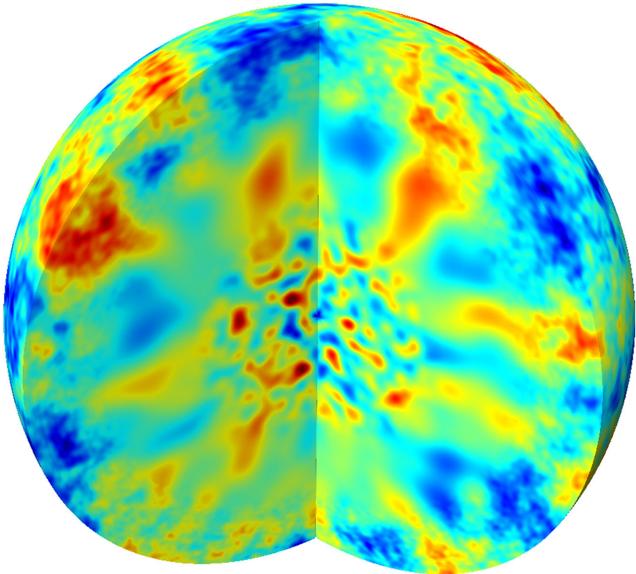


Figure 6. Second-order random model with non-stationary *a priori* covariance function (shown in Fig. 5). It is divided in three distinct regions with isotropic Gaussian covariance in the centre, a larger scale vertically elongated Gaussian covariance at intermediate radii, and an exponential horizontally elongated covariance function close to the surface of the sphere. The different regions are correlated and transfer smoothly into each other.

The approximate full covariance matrix can be obtained from a weighted sum of the submatrices. The weights are chosen such that each covariance matrix dominates a different radii and transfers smoothly into the neighbours. Also the correlation between different radial regions can be estimated by averaging the covariance matrices of the different regions. This can for example be written as:

$$C(r, r') = \sum_{ij} w^i(r) w^j(r') (C_l^i(r, r') + C_l^j(r, r')) / 2. \quad (44)$$

Here, $w^i(r)$ is a window function that windows out region i . The product $w^i(r) w^j(r')$ provides the weights of the different regions. The correlation between different regions is windowed out by the crossterms $w^i(r) w^j(r')$, and is chosen as the simple average of the corresponding covariance functions. Such a weighted sum is only a rough interpolation of the covariance matrices: the combined matrix might, for example, lose its positive definiteness. However, depending on the original matrices, negative eigenvalues are often very small. We have found that this method is sufficient for our purpose and provides reasonably interpolated covariance functions. More elaborate methods and cleaner definitions of non-stationary covariance functions are beyond the scope of this paper (see e.g. Stein 1999; Lindgren *et al.* 2011).

The combined covariance matrix C_5 is shown in Fig. 5(a) in which the different regions are easy to identify. Note that this matrix becomes zero towards the centre of the sphere, because the wavelengths that are associated with degree 5 spherical harmonics becomes much smaller than the scale length of the model. Also, the interpolation method described in eq. (44) provides a covariance function with a smooth transition between different regions. The horizontal power spectra of the sphere are shown in Fig. 5(b) and again the three regions are easy to identify. As expected, horizontal power shifts to lower degrees towards the centre, keeping the Cartesian scale lengths of the model constant.

In the model realization shown in Fig. 6, the three regions can easily be recognized. They are correlated and transfer smoothly into

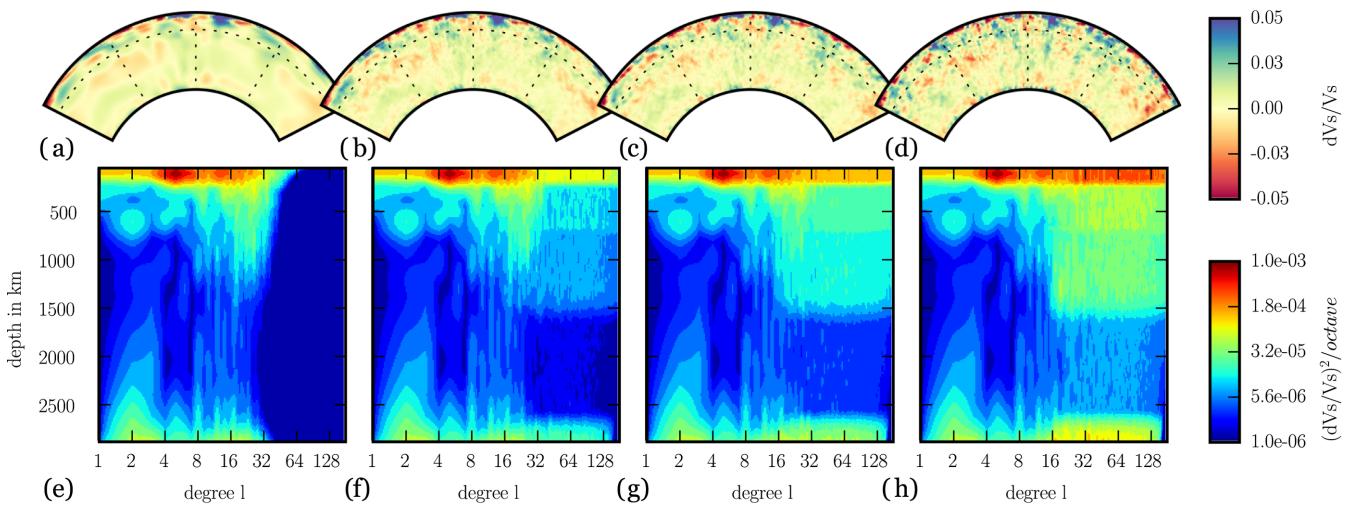


Figure 7. Combined tomographic and statistical models. (a)–(d) show cross-sections of the original tomographic model (a) extended to shorter wavelengths with heterogeneities that are weaker (b), similar (c) or stronger (d) than the original model (colour scale saturates at ± 5 per cent). (e)–(h) show the associated horizontal band power spectra of the models that represent heterogeneity strength at different scales. The low-degree spectrum is fixed to the tomographic model SEMUCB-WM1, high-degree power increases from model (e/a) to (h/d).

each other. This model is a toy example that illustrates the type of models that can be generated with this method.

5 APPLICATION TO SEISMIC WAVE PROPAGATION IN EARTH'S MANTLE

5.1 Model design

In this section, we generate statistical models that represent unknown hypothetical elastic structure in Earth's mantle. We focus in particular on structure with sizes that are smaller than the approximate resolution limit of current generation global tomographic models (1000–500 km, $l \approx 20\text{--}40$) (e.g. Meschede & Romanowicz 2015). As minimum horizontal size, we choose approximately half the wavelength of long-period seismic signals such as a 100 s period Rayleigh wave (100 km, $l \approx 200$).

We use the tomographic model SEMUCB-WM1 (French & Romanowicz 2014) to represent the elastic long wavelength structure (cross-section shown in Fig. 7a). The statistical models add similar perturbations to isotropic V_s , V_p wave velocities as well as density with scaling factors $\frac{dv_p}{vp} = \frac{1}{2} \frac{dvs}{vs}$ and $\frac{dp}{\rho} = \frac{1}{3} \frac{dvs}{vs}$. We do not consider different scaling factors in this study, although they can influence the strength of the coda.

We add statistical heterogeneities to SEMUCB-WM1 that are either weaker (Fig. 7b), similar (Fig. 7c) or stronger (Fig. 7d) than the large-scale structure of the tomographic model. Horizontal power spectra of these models are shown in Figs 7(e)–(h). The spectra show the horizontal logarithmic band power that can be associated with the total variance of localized heterogeneities (e.g. Meschede & Romanowicz 2015). We distribute the statistical heterogeneities evenly over the examined spectral range, using an isotropic scale-free covariance function that has even band power (see Klimeš 2002):

$$\langle |F(\mathbf{k})|^2 \rangle = \log(2) 2\pi^2 \sigma^2 a^3 / (1 + (|\mathbf{k}|^2 a^2)^{3/2}) \quad (45)$$

with variance σ^2 per spherical harmonic octave band, and scale length a which is set below the resolution limit of the tomographic model. We have chosen a minimal resolution of degree $l = 180$ at the surface, which corresponds to a wavelength of about 220 km. This

power spectrum is proportional to $\langle |F(\mathbf{k})|^2 \rangle \sim \frac{1}{|\mathbf{k}|^3}$ for $|\mathbf{k}|^2 a^2 \gg 1$. In 3-D, the total power per $|\mathbf{k}|$ is then proportional to the characteristic $1/|\mathbf{k}|$ decay of a scale-free medium and the power per $\log(|\mathbf{k}|)$ is constant. Different spectra, e.g. with a dominant scale (e.g. exponential or Gaussian), are not examined.

Each model increases the band power, and therefore the variance of the statistical heterogeneities by a factor of two without changing the horizontal and radial correlation functions (and the associated power spectra). Their relative (co)variances are therefore $\frac{1}{2}, 1, 2$. Additionally we are going to use an even stronger model with relative (co)variance 4 (not shown). As can be seen from the horizontal spectra, the variance of the statistical heterogeneities is depth dependent and follows roughly the strength of larger scale heterogeneities in the original model. The medium strength model c) adds, over three to four octaves, perturbations with a total standard deviation of about 2–3 per cent in the heterosphere (upper 200 km), 0.8 per cent above the 660 km discontinuity, 0.6 per cent below 660 km down to 1500 km, and 0.3 per cent down to D'' where they increase again to about 1 per cent. The relative standard deviations of the model perturbations are $\frac{1}{2}, 1, \sqrt{2}, 2$ (including the very strong model whose spectrum is not shown).

All models have a heterogeneous crust based on model Crust2.0 (Bassin *et al.* 2000), or when specified Crust1.0 (Laske *et al.* 2013). Moho and surface topography are explicitly meshed but smoothed. We therefore do not expect to model the full effect of Crustal structure or topography. Oceans are included in an approximate way as well (see Komatitsch & Tromp 2002b).

It is important not to misinterpret the global variance and standard deviations as the strength of individual scatterers. The given numbers are global averages at a certain depth where heterogeneities could be dense or more sparsely distributed. Sparse heterogeneities are stronger than dense ones, if they have the same global standard deviation, because they concentrate the same variance in fewer heterogeneities. Nevertheless, such strong sparse scatterers can still have the same average effect on the wavefield. Therefore standard deviations should be considered as a lower limit on the typical heterogeneity strength. We show an example of sparse and dense scatterers in Section 5.7.

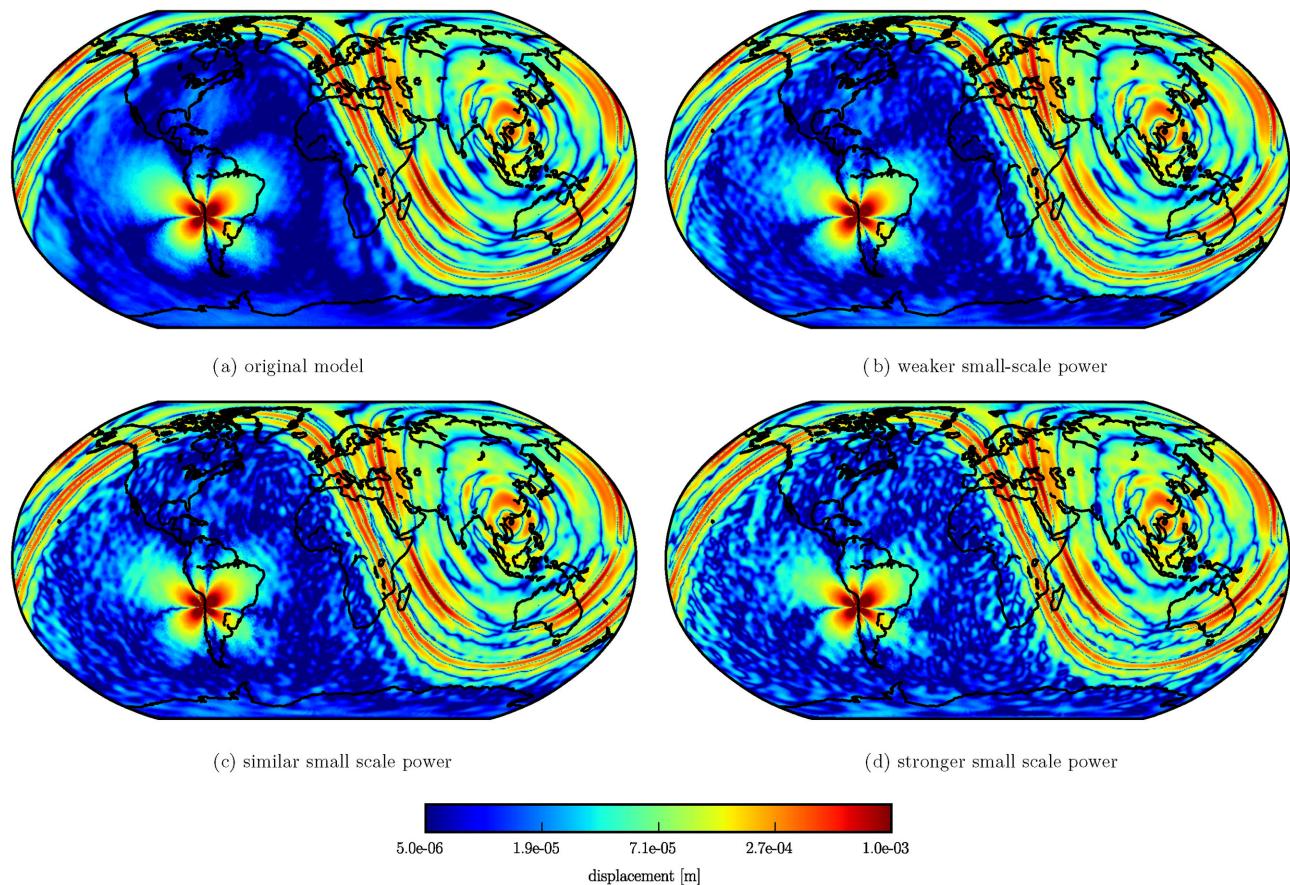


Figure 8. Snapshots of the absolute vertical displacement wavefield at the surface in logarithmic colour scale at lapse time 3134 s. In Chile, the static displacement of the source can be seen. Seismic waves are focusing at the antipode and strong Rayleigh waves are at around 90° arc distance from the source. Already in the original model (a), heterogeneities lead to focusing and defocusing effects that lead to a quite complex wavefield (especially visible at the antipode). In the extended models the wavefield changes only little, even when the additional heterogeneities are strong. Most of the complex interference patterns stay similar to the original models, as they are caused by the larger scale heterogeneities. However, from (b) to (d), a coda develops that is most visible behind the surface wave, where they are largely isolated from other arriving signals.

5.2 Effect on the seismic wavefield

We compute wavefields numerically using the full Earth elastic wave equation solver SPECFEM3D_GLOBE (Komatitsch & Tromp 2002a,b), using a mesh that is accurate to about 35 s period. Note that the statistical models are well sampled by this mesh and smooth compared to its node spacing. Fig. 8 shows snapshots at the surface of the logarithmic absolute vertical displacement at around 50 min lapse time, after a moment-tensor source of a 30 km deep earthquake located in Chile. Figs 8(a)–(d) correspond to the original model (a), and the three extended models (b) to (d), with relative powers $\frac{1}{2}$, 1 and 2, respectively.

A logarithmic colour scale has been used that allows us to identify seismic signals in a large amplitude range. The static displacement and radiation characteristics of the source can be seen in the snapshots. At snapshot time, seismic waves have already reached the antipode (in South-East Asia) where they focus. Strong fundamental mode Rayleigh waves are at about 90° distance from the source.

Already in the unmodified tomographic model (Fig. 8a) the wavefield shows complex focusing and defocusing effects. Especially at the antipode, strong irregular interference can be seen. The additional smaller-scale structure (Figs 8b–d) changes the wavefield only slightly, even when the small-scale structure dominates the total model variance (Fig. 8d).

One reason for the limited influence of the small-scale structure on the direct waves is that travel-times and focusing effects are most sensitive to the integrated effect of the structure along the propagated path. Small perturbations ‘average’ out and therefore do not influence direct waves much. Only if they are sufficiently strong, large-scale effects can become noticeable, even when the perturbations are much smaller than the wavelength of the wave (Capdeville *et al.* 2013). In our examples, perturbations are relatively weak and therefore we do not expect to observe such effects.

Although the direct wavefield is not strongly affected, a coda develops with increasing small-scale heterogeneity strength from Figs 8(a)–(d). It can most easily be seen behind the fundamental mode surface waves where it is strongest, and largely isolated from other signals. This coda is sensitive to the model power as can be seen by comparing the wavefields in Figs 8(a)–(d). It is weak compared to the direct wave amplitudes (1–10 per cent of the maximum Rayleigh wave amplitude), even in model (d) where the small-scale structure dominates the total model power. Therefore the wavefields in all models are, at least at first glance, qualitatively similar to recorded seismograms that have a relatively weak coda.

The scattered wavefield can be better seen in the difference wavefield between Figs 8(c) and (a), shown in Fig. 9(a). The largest absolute (not necessarily relative) differences are observed around the

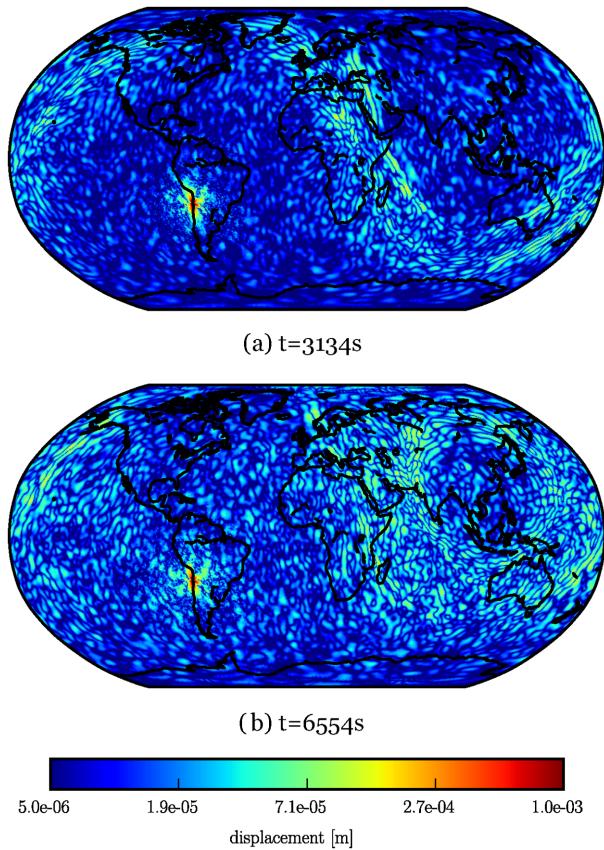


Figure 9. Absolute difference between the vertical displacement of the original tomographic model, and the strong extended model (relative variance 2) at (a) time 3134 s that corresponds to Fig. 8. At this time, waves have not reached the antipode yet. The coda is strongest during and after the fundamental mode Rayleigh wave arrivals. At snapshot time 6554 s (b), these Rayleigh waves have passed the antipode. Due to the spherical geometry, the coda can be stronger before the arriving direct waves.

strong fundamental mode signals. The difference wavefield seems to be distributed incoherently and no clear wave fronts can be identified. It therefore presumably consists of many scattered signals that have interfered. The radiation characteristics of the source are only visible in the direct waves but no longer in the later coda. Due to the dominant amplitude of the reference wavefield, we expect that the coda can be described quite well using a single scattering model at these lapse times. At a later snapshot time, shown in Fig. 9(b), these effects become even clearer. At this time waves have passed the antipode and due to the spherical geometry, the coda can be strong even before the fundamental mode arrival.

5.3 Spectrograms

Spectrograms allow for a more quantitative analysis of the scattered wavefield. We use energy-normalized Morlet wavelets $\Psi(t) = (\pi a^2/\omega^2)^{-1/4} e^{(i\omega t)} e^{(-\omega^2 t^2/(2a^2))}$ to decompose the signal into a time-frequency representation. We choose the scale factor a , that controls the time resolution of the spectrogram, as $a = 8$ if not indicated otherwise, which smooths the seismograms over a few periods. The spectrogram power provides an estimate of the wavefield's variance ('energy') at a fixed location as a function of time and frequency.

We use again a 30 km deep, artificial explosive source and do not take anelastic attenuation into account to simplify the examination of the coda at longer lapse times. Fig. 10, row 1 shows spectrograms

of the vertical displacement field in the four models at a station with 67.5° distance from the source. The periodically arriving minor and major arc fundamental mode Rayleigh waves dominate the overall spectrogram amplitudes. After the first fundamental mode Rayleigh wave arrival (R1), minor and major arc arrivals (R2 and R3) arrive quickly after each other because the station is closer to the source than to the antipode. The overall difference of the spectrograms is small. However, as the amount of small-scale heterogeneities increases, scattered energy appears between the main arrivals.

This scattered energy becomes much clearer in the difference of the extended model's spectrograms with respect to the tomographic model's spectrogram (Fig. 10, row 2). Blue colour indicates reduced energy in the extended model and can be seen mostly at the location of the main arrivals. This shows that energy is scattered out of the direct waves (fundamental modes), generating scattering attenuation. Red colour indicates additional energy in the extended model and, away from the main fundamental mode arrivals, the difference spectrograms are almost exclusively red due to the coda. Coda and scattering attenuation increase with heterogeneity strength and lapse time, as expected.

Fig. 11 shows the average spectrogram power (computed with $a = 16$) in period bands between 400–200 s (a), 200–100 s (b), 100–50 s (c). Clearly a developing coda can be observed, especially at the short periods that have propagated more wavelengths during the same time.

Interestingly, the scattered energy is not evenly distributed: the coda is stronger between the minor and major arc (R2 and R3) arrivals and weaker after the major arc arrival (R3). At first glance, this is surprising because we usually expect the coda to become stronger after a direct wave. An explanation for this is the peculiar spherical geometry: on the sphere, the minor arc is the shortest possible path between two points. Scattered surface wave energy therefore tends to arrive later. On the other hand, the major arc is not the shortest path and scattered waves can take a shortcut to arrive earlier. Therefore, at stations close to the antipode, the scattered energy of the major and minor arcs adds and concentrates in between the arrivals.

This peculiar behaviour of the coda that is caused by the spherical geometry of the Earth provides a method to distinguish the fundamental mode coda from other energy. It has been modelled using isotropic single and multiple scattering models using effective scattering parameters on the 2-D spherical surface (Sato & Nohechi 2001; Sato 2002).

5.4 Scattering attenuation

From the spectrograms, we can measure scattering attenuation due to the elastic statistical heterogeneities in the respective models. Fig. 12 shows measurements of fundamental mode Rayleigh wave attenuation as a function of frequency, relative to the tomographic model without additional scatterers.

To obtain the attenuation rate, we first measured the relative amplitudes of the first six periodically arriving fundamental mode Rayleigh waves at each frequency with respect to the original tomographic model. We then fit a line to the logarithm of the relative amplitudes at each frequency and measure its decay rate. Note that in this measurement procedure, focusing and defocusing effects tend to cancel (e.g. Romanowicz 1990). Because the small-scale heterogeneities do not perturb the direct waves much through traveltime and focusing, this procedure works well and we do not need more elaborate methods (e.g. as in Laske & Masters 1996; Ekström *et al.* 1997) that better isolate the fundamental mode.

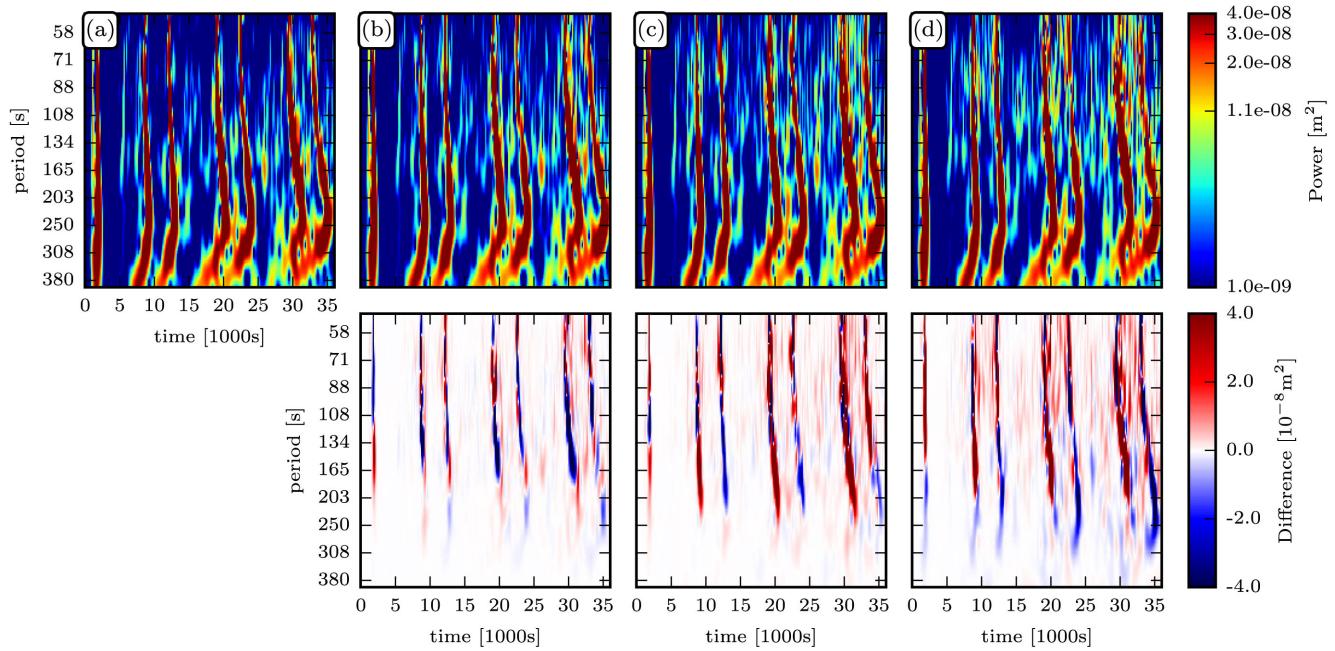


Figure 10. Row 1: spectrograms of the vertical displacement field computed in the original (a) and the extended models with relative variances $\frac{1}{2}$ (b), 1 (c), 2 (d), without attenuation at a station with 67.5° distance from the source. Heterogeneity strength increases from left to right. Most energy is in the strong Rayleigh waves that arrive periodically as they orbit around the Earth. The wavefield amplitude between the direct surface wave arrivals increases with heterogeneity strength as a coda of scattered waves develops. Over longer lapse times, the coda becomes stronger and also the direct wavefield becomes distorted. Row 2: spectrogram differences shows the expected behaviour: energy gets scattered out of direct arrivals (blue colour) into the coda after and in between the direct waves (red colour).

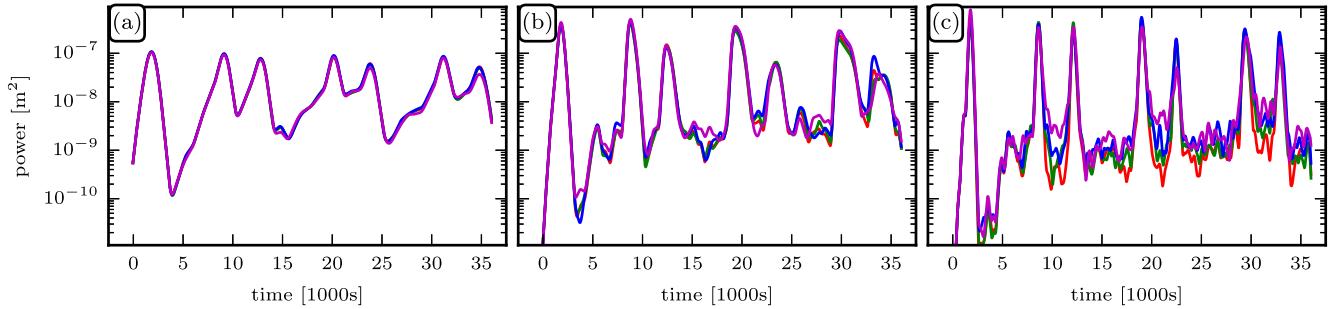


Figure 11. Average spectrogram power of the vertical displacement field in frequency bands (a) 400–200 s, (b) 200–100 s, (c) 100–50 s period, and for the original tomographic model (red) with weak (green), medium (blue), and strong (magenta) additional heterogeneities, corresponding to the spectrogram shown in Fig. 10.

Figs 12(a)–(d) shows the measurements of the attenuation factor $q = 1/Q$ (blue line) and the observed attenuation as predicted by the attenuation model QL6 (Durek & Ekström 1996) (green line) in the models with relative variances $\frac{1}{2}$, 1, 2, 4. At the longest examined periods, 250–400 s, scattering attenuation is undetectable in our measurements. Around 200–250 s, it becomes clearly detectable and increases strongly to about 100 s period. For shorter periods, it remains approximately constant, even with a slight decreasing trend. Compared to the observed attenuation, scattering attenuation is very weak. At 100 s period, it ranges from around $q < 1 \times 10^{-4}$ (a) to $q = 1 \times 10^{-4}$ (b) $q = 2 \times 10^{-4}$ (c) to $q = 4 \times 10^{-4}$ (d). This corresponds to about 0.5 per cent (a) to 1 per cent (b) 2 per cent (c) 4 per cent (d) of the total QL6 attenuation. These values compare well with the scattering coefficient $g = 1 – 4 \times 10^{-6} \text{ km}^{-1}$ that has been estimated in the period range between 200 and 100 s

by matching long-period seismogram envelopes with single and multiple scattering models (Sato 2002): With a group velocity of $v = 3.8 \text{ km s}^{-1}$ this scattering coefficient corresponds to a range of $q = gv/\omega = 0.6 – 4 \times 10^{-4}$.

Interestingly, even in such a complex non-stationary model, the attenuation factor is proportional to the total model variance. This relation allows us to make predictions for scattering attenuation in a large range of models based on a single computation. For example, a similar statistical model with a relative variance of 10 could explain about 10 per cent of the total observed attenuation at 100 s period. Of course, this is only valid if we stay in the linear regime of relatively weak and small heterogeneities but it provides an estimate of the required structure.

The frequency dependence of scattering attenuation depends on the depth dependence of the fundamental mode sensitivity and on

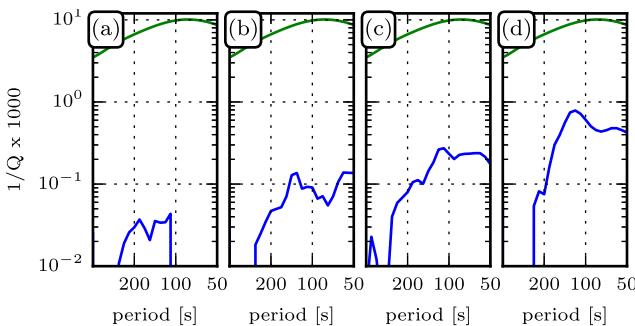


Figure 12. Attenuation measurement of the fundamental mode Rayleigh wave in models with weak (a), medium (b), strong (c) and very strong (d) small-scale heterogeneities. Attenuation is measured in a period range from 400 to 50 s period. From (a) to (d), the power spectra and therefore the model variance is increased three times by a factor of 2. The blue line shows the measured attenuation factor $1/Q (=q)$. The green line shows the observed attenuation as modelled by QL6. Scattering attenuation is in general very weak. At longer periods, scattering attenuation is too weak to be observed from our limited number of measurements.

the non-stationary distribution of the scatterers in the model. It is therefore difficult to explain without a detailed analysis. However, longer periods are sensitive to larger (in horizontal and radial extent) and also deeper heterogeneities, whereas shorter periods are sensitive to smaller and shallower heterogeneities. In our non-stationary models, relative velocity perturbations become much weaker with depth (variance drops to almost 1/10 from 150 to 250 km depth). We therefore expect longer periods to be less attenuated as soon as their dominant sensitivity is in the deeper regions, which can also be observed in Fig. 12. As periods become shorter, dominant fundamental mode sensitivity is in the heterosphere, the scale-free heterogeneities similarly affect a larger frequency band. This is a possible explanation for the approximately constant scattering attenuation at periods < 100 s. The small-scale limit of the statistical models beyond $l = 180$, as well as the unperturbed crust that does not contribute any additional scattering, could explain a slightly decreasing trend. Therefore, a more complete examination of scattering attenuation towards 50 s period, should take into account smaller scale heterogeneities and especially the crust.

5.5 Coda analysis

The decay of a direct wave can be explained similarly by anelastic or elastic effects. Therefore the portion of the latter cannot be distinguished from observations of direct wave amplitudes, although it is important for the construction of intrinsic attenuation models of the Earth (e.g. Gung & Romanowicz 2004; Dalton *et al.* 2008). On the other hand, the strength of the coda can be observed and has been used to study the amount of scattering and to gain insights about elastic structure. Similarly, the absence of a coda can constrain the strength of small-scale features and limit the amount of scattering attenuation. The relation between coda intensity and model power connects observations with the elastic structure and is therefore important.

Again, as for scattering attenuation, we focus on the coda of long-period fundamental mode Rayleigh waves. We examine spectrograms, such as those in Fig. 10, that provide local estimates of the seismogram variance as a function of frequency. The difference in spectrogram power with respect to the unperturbed tomographic model then provides an estimate of the change in wavefield vari-

ance due to the statistical scatterers at a certain position, time and frequency. We measure the ‘coda intensity’ as the mean of the difference spectrogram power between two Rayleigh wave arrivals, as a function of frequency in the different models (10 periods after and before the arrival times as predicted by PREM). Naturally, this definition of the coda intensity mixes scattered waves and perturbations of all direct seismic phases except for the fundamental mode, such that we cannot directly associate this coda with a particular phase. However, for a shallow event most wavefield energy is in the fundamental mode surface waves that furthermore sample the heterogeneous shallow depth regions, such that we expect that they can be associated with much of the scattered variance. Note that with this procedure, we examine a relative coda with respect to a 3-D reference model, in contrast, for example, to Sato (2002).

We measure the coda intensity at stations that are distributed along four great circles that connect source and antipode, in two model realizations, using the same explosive source at 30 km depth without anelastic attenuation as in the previous section. We then average the measurements at stations that are located within a similar distance bin to obtain better results. Fig. 13 shows the measured coda intensity in the four extended models with relative powers $\frac{1}{2}, 1, 2, 4$ from (a) to (d) in nine distance bins as a function of frequency. Our very rough characterization of coda intensity only provides good measurements in the distance bins between 45° and 135° . At these distances, a clear increase in wavefield variance (i.e. coda intensity) can be observed between the Rayleigh wave arrivals due to the additional heterogeneities in the extended models.

In addition, the previously mentioned pattern becomes very clear now: the fundamental mode coda concentrates between close arrivals: for stations that are closer than 90° to the source it is strongest after R2, R4; for stations that are further away than 90° it is strongest after R1, R3, R5. In the 90° distance bin, coda energy is distributed isotropically between the arrivals. Furthermore, the coda intensity depends on the model power and frequency, and increases with lapse time.

Coda intensity is largest for frequencies between 60 and 140 s. As for scattering attenuation, this depends on several factors: (1) the source characteristics. (2) the propagated wavelengths (more for shorter periods). (3) the non-stationary distribution and strength of heterogeneities. A detailed theory or direct comparison with data traces is necessary to analyse it, which is beyond the scope of this article. We can, however, examine the ‘relative’ coda intensity of the different extended models. Because source and non-stationary heterogeneity distribution are similar in all of these measurements, we expect that they generate codas with the same frequency dependence, differing only by a factor, as long as we are in the linear regime.

We measure this frequency dependence by computing the frequency independent scaling factor (inner product) between the coda intensity ‘images’ of Fig. 13. Fig. 14 shows the proportionality factors of the coda between the different models at different distances. A clear trend can be observed at all distances, and in particular in the distance bins 67° – 112° , it is almost linear.

Note that, in this synthetic example, we can distinguish between models with different amounts of small-scale structure from a few seismograms only. This works even though the models have complex larger scale 3-D structures, and a complex 1-D background model that also generate scattered waves. However, in this example, the background structure is known perfectly and provides us with a reliable reference trace. In more realistic examples, an incorrect background structure will certainly also affect the relative coda intensity.

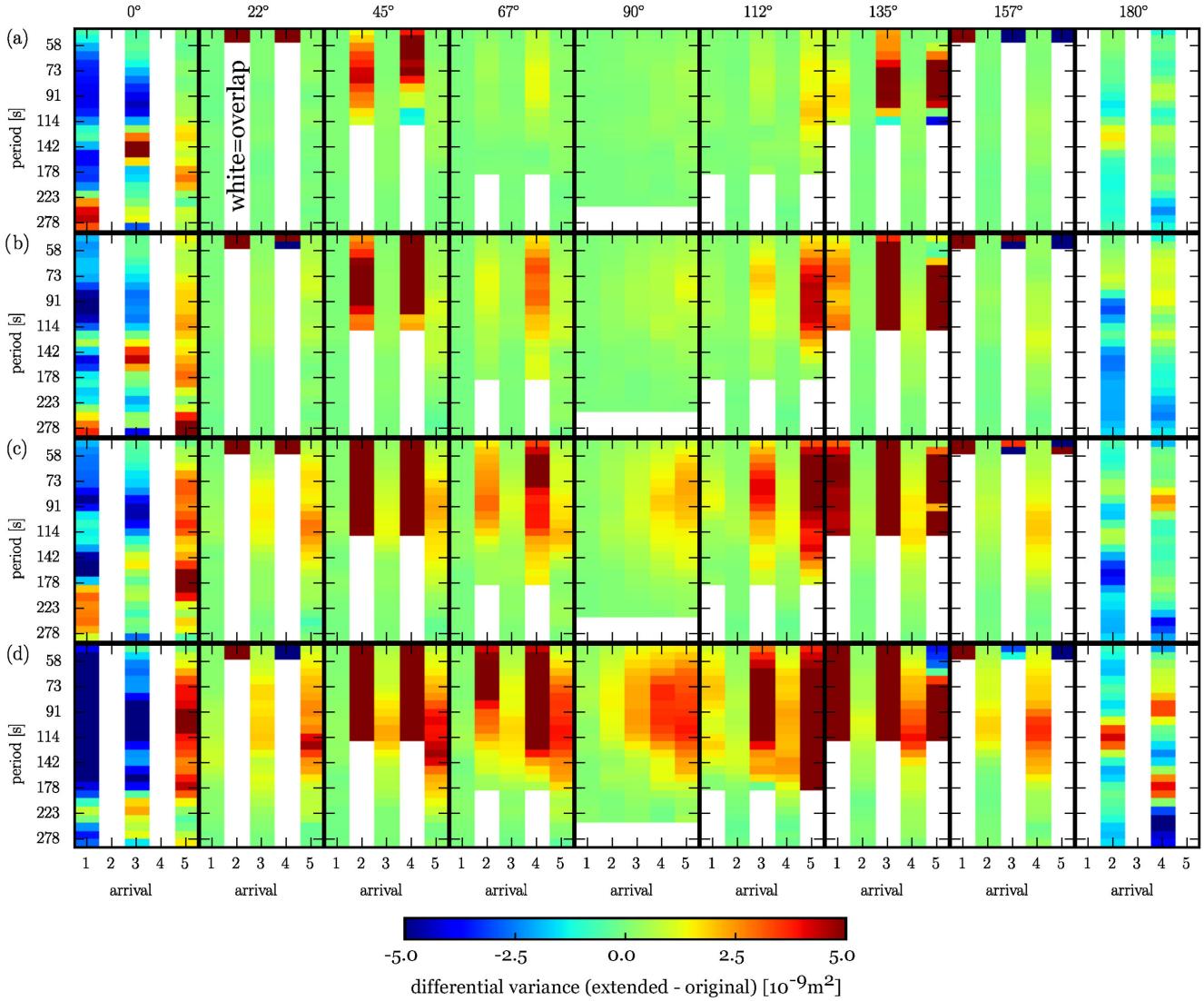


Figure 13. Differential seismogram variance (Spectrogram Power) with respect to the original tomographic model after each successive Rayleigh wave arrival in the four different models (rows a–d). Measurements show the mean spectrogram difference 10 periods after one Rayleigh wave arrival up to 10 periods before the next, averaged over all stations in a specific distance bin (columns). White colour indicates that successive Rayleigh waves arrive within a short time and overlap such that no measurement is possible. The distinct pattern due to the spherical geometry is clearly visible. At 90° distance, the coda is distributed homogeneously. At intermediate distance, we can see a clear dependence of the coda strength on model power.

5.6 Anisotropic structures

We have demonstrated that the relative coda intensity and scattering attenuation can be estimated for models whose covariance matrices differ by a factor. In this case, heterogeneities differ in amplitude but not in shape. On the other hand, heterogeneities in the Earth can have various shapes. In particular, layered structures are often observed, but also vertically extended heterogeneities such as plumes or slabs can be found in Earth's mantle.

In this section, we change the heterogeneity shape using the ‘squeeze factor’ k according to eqs (42) and (43). Values of $k > 1$ correspond to radially squeezed and $k < 1$ to radially elongated heterogeneities with the same horizontal covariance function. Fig. 15 shows two cross-sections of models with heterogeneities that are either radially squeezed by a factor 2 ($k = 2$, Fig. 15a) or radially stretched by a factor 2 ($k = 0.5$, Fig. 15b). Both models have the same large-scale structure and the same horizontal power spectra. Only their vertical correlation length is different.

Structures, which are elongated perpendicular to the direction of wave propagation, have been associated with increased back-scattering (e.g. Hong & Wu 2005; Cormier 2007). We therefore expect that radially elongated structures increase observed scattering attenuation and the coda intensity. Furthermore, surface waves are effectively sensitive to vertically smoothed perturbations. Such perturbations are largest if the radial correlation length is large, which leads to a similar conclusion. However, wave propagation in anisotropic (also called anisomeric) media can be more complex, and a detailed analysis is beyond the scope of this study (see e.g. Margerin 2006).

Fig. 16 shows the relative coda intensity for four models with $k = 0.5, 1, 2, 4$, measured as in the previous section, in the distance bins that are not too close to source or antipode. Indeed the strength of the coda becomes weaker relative to the isotropic model (b) for radially squeezed structures ($k > 1$) and therefore reflects the expected trend. On the other hand, radially elongated structures

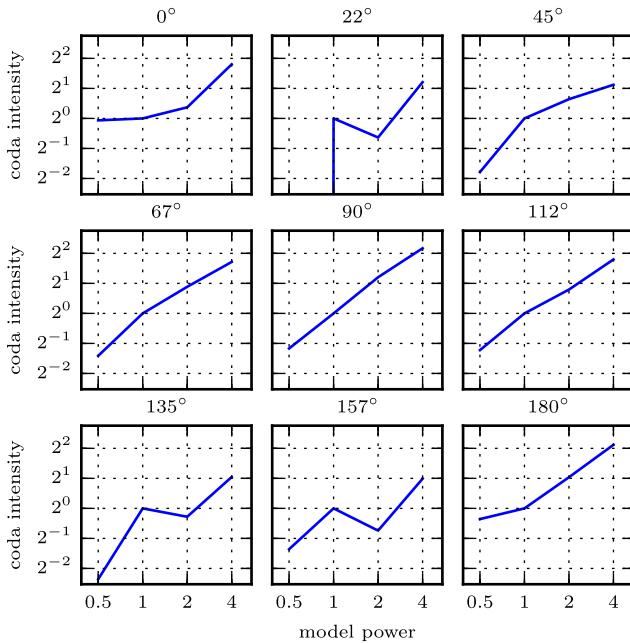


Figure 14. Relative coda intensity, averaged over all frequencies, as a function of relative model variance for stations at different distances. At distances around 90° , the linear trend is clearly visible. Close to the source or to the antipode, we do not get good measurements.

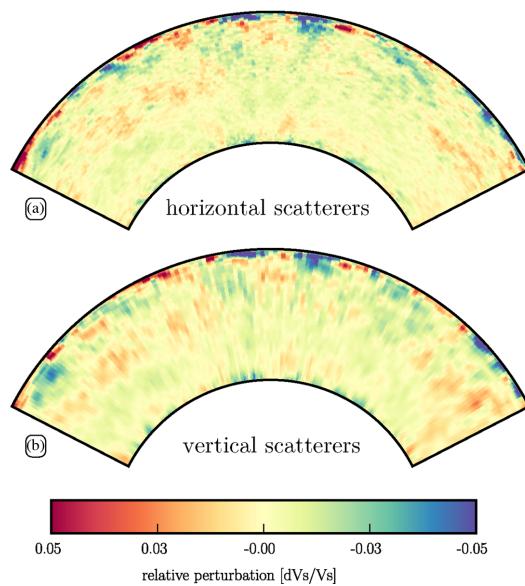


Figure 15. Cross-sections of relative perturbations in V_s (colour scale saturates at ± 5 per cent) in a model with (a) radially squeezed structures (factor $k = 2$) and (b) radially elongated structures (factor $k = \frac{1}{2}$). The horizontal model spectra and the heterogeneity variance are the same in both models.

($k < 1$) do not increase the strength of the coda much. Scattering attenuation, shown in Fig. 17, seems almost independent of weak vertical squeezing ($k = 0.5\text{--}2$) but clearly decreases for strongly squeezed structures ($k = 4$). This shows that the interaction between structure and waves is complex and depends, for example, on their depth sensitivity and the non-stationary heterogeneity distribution of the medium. Important for our purpose, it seems difficult to strongly increase scattering attenuation and coda intensity through radially elongated or squeezed structures.

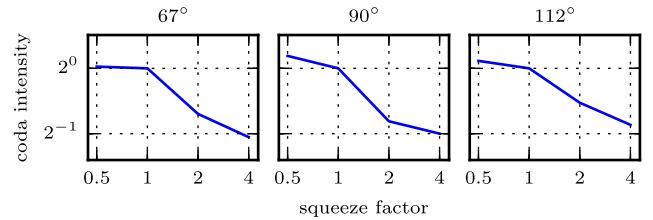


Figure 16. Relative coda intensity (normalized) as a function of radial anisotropy, defined by the squeeze factor k . At these long periods, we expect that stretching heterogeneities in vertical direction strengthens the coda, whereas squeezing in vertical direction weakens the amount of scattering. The exact relationship depends on many factors, most notably the sensitivity of the wavefield at these times and periods, as well as the non-stationary behaviour of the covariance matrix.

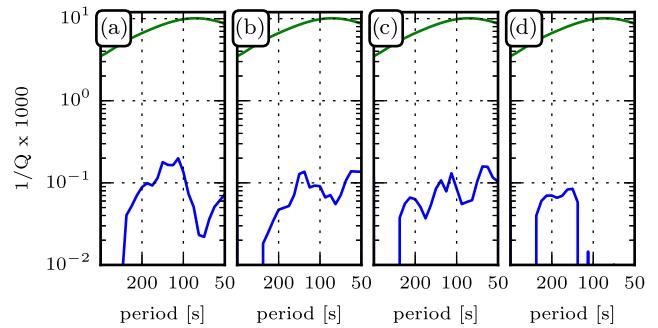


Figure 17. Scattering attenuation of the fundamental mode Rayleigh wave as a function of frequency for squeeze factors $k = 0.5$ (a), 1 (b), 2 (c) and 4 (d). The blue line shows the measured attenuation factor $1/Q (=q)$. The green line shows the observed attenuation as modelled by QL6.

5.7 Sparse heterogeneities

So far we have considered heterogeneities to be densely distributed around the globe (this is a consequence of the normally distributed perturbations in second-order random models). In the Earth, however, we often see situations where heterogeneities are sparsely distributed which means that they concentrate in few locations. In Section 2.1, we have already argued that we do not expect averaged quantities such as the coda or scattering attenuation to depend strongly on the exact distribution of heterogeneities as long as their total variance remains similar. To verify this claim, we examine a model realization with a very different and sparse heterogeneity distribution that has the same spatially averaged covariance function as a model realization with dense heterogeneities.

To this end, we try to concentrate the scatterers at the edges of the large scale structure that is given from the tomographic model. To achieve this, we compute the horizontal gradient in the original model, normalize its variance to one, and multiply the statistical model with this gradient mask at each depth. This concentrates heterogeneities in regions with a strong (horizontal) gradient in the tomographic model. Because the gradient function has larger scale lengths than the statistical heterogeneities, this procedure changes the globally averaged (co)variance of the scatterers only little, and does not influence the horizontal and vertical spectrum much. Heterogeneities in this model are not only concentrated in a few regions, but also correlated to the larger scale structure.

The original tomographic model, and model realizations of the stationary model with densely distributed heterogeneities and the ‘gradient’ model with sparsely distributed heterogeneities are shown in Fig. 18. Although the dense and the sparse model have about the same horizontally averaged covariance functions, they have a

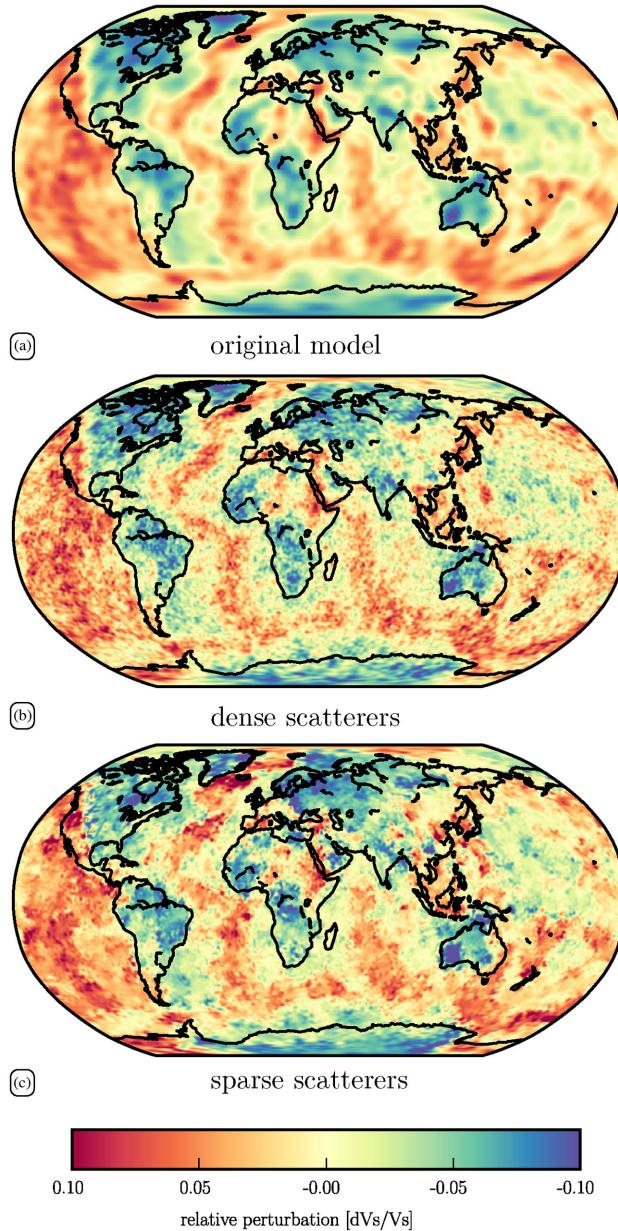


Figure 18. Horizontal maps of relative V_s perturbations (saturated at ± 10 per cent) of the original model (a), a model with densely distributed scatterers (b) and (c) a model with scatterers that are concentrated in regions with a strong gradient in (a). All models have similar horizontal and vertical power spectra, corresponding to the one shown in Fig. 7(g).

very different distribution of heterogeneities. The sparse model has stronger perturbations than the dense, because the same total variance is concentrated in a smaller volume.

Fig. 19 shows the strength of the coda in these models. Even though they are quite different, coda intensity is very similar, confirming that it is really sensitive to a spatially averaged covariance function of the structure, which is similar in the sparse and dense model. However, we have noticed that individual measurements (in particular of fundamental mode amplitudes) can deviate more strongly from these average properties, as we would expect, if a wave samples an isolated strong heterogeneity. Such strong ‘outliers’ can make it more challenging to measure a reliable average. Also scattering attenuation, shown in Fig. 20, in both models is at a

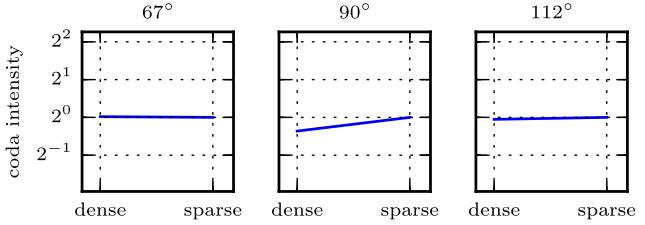


Figure 19. Relative coda intensity (normalized to the sparse model) at different distances in a model with a dense and a sparse distribution of scatterers. The variance of the scattered wavefield almost does not change. This means that this averaged quantity is insensitive to the exact distribution of the scatterers.

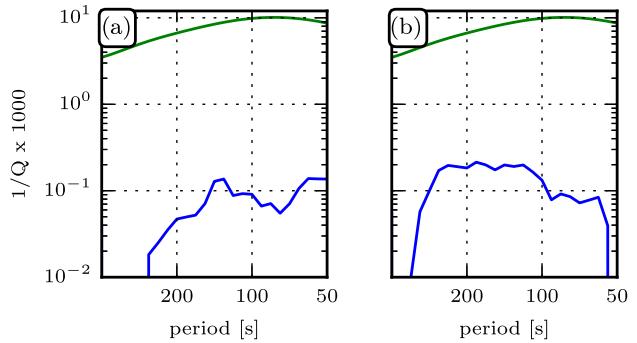


Figure 20. Scattering attenuation in (a) the extended model with medium strength heterogeneities and a uniform distribution of heterogeneities (corresponding to Fig. 12b) and (b) a model with similar spatially averaged covariance function, but sparse distribution of heterogeneities. Although low frequencies seem to be affected a little stronger by the sparse heterogeneities, the overall level of scattering attenuation is comparable.

similar level, although the sparse model seems to affect long periods (>200) more.

5.8 Data example

We only provide one qualitative comparison with an observed spectrogram. Because we are looking at average properties of the wavefield, a single station recording can only be an illustration of the concept. Also we have not examined the influence of the crust so far which can definitely play an important role for scattering. A detailed and robust comparison is beyond the scope of this paper.

Fig. 21 shows vertical displacement spectrogram power of the recorded (a) and computed (b–c) spectrograms after a 70 km deep, M7.2 earthquake near Tonga (event: 200605031527A). This event occurred largely isolated from other strong events. The selected station ‘BBSR’ is closer to the antipode than to the source, and we therefore expect more coda energy to arrive after the minor arc fundamental mode Rayleigh wave (R1) than after the major arc arrival (R2).

The recorded spectrogram (a) is compared with computations in the original tomographic model (b) and in the extended model with medium strength heterogeneities (c) in row 1. In contrast to our previous computations, the observed Rayleigh waves are now strongly attenuated and the source mechanism is more complicated than a simple explosion source. Fortunately, the more complicated source mechanism provides the advantage of nodal paths, at which direct surface wave arrivals are strongly damped. In this case, we expect scattered waves to arrive in these nodal regions because they do not follow a straight great circle path.

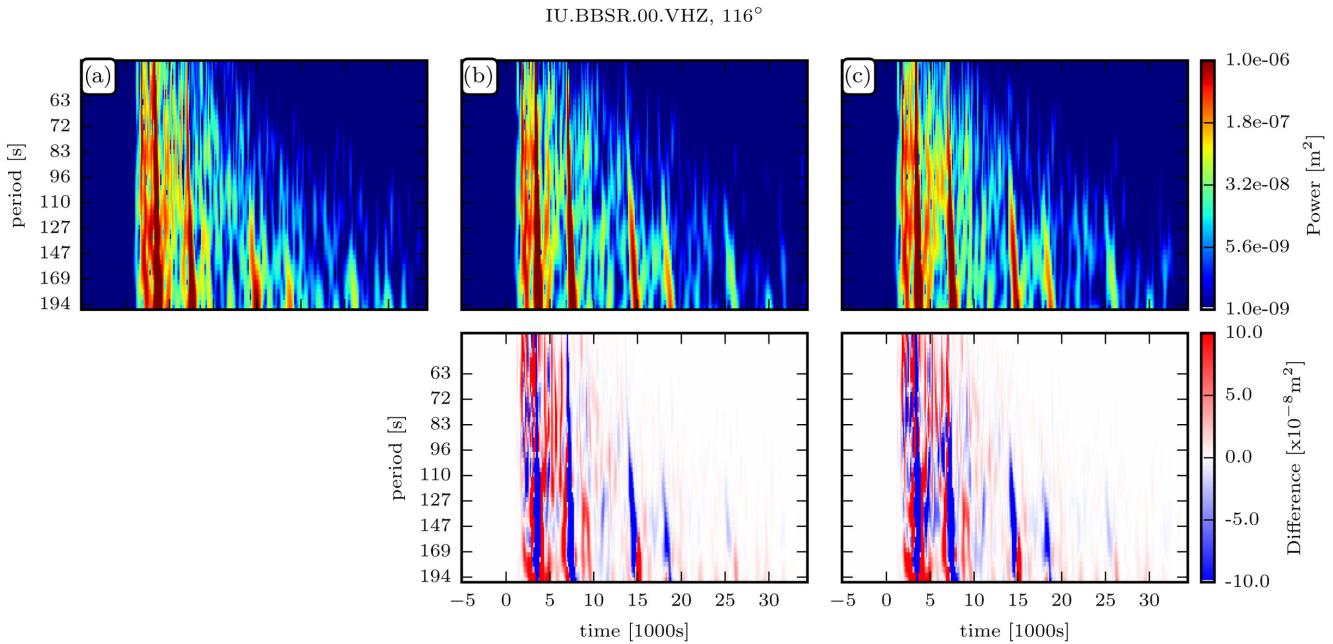


Figure 21. Comparison of a data spectrogram (a) with synthetic spectrograms computed in the original tomographic model (b) and the extended model with medium strength heterogeneities (c).

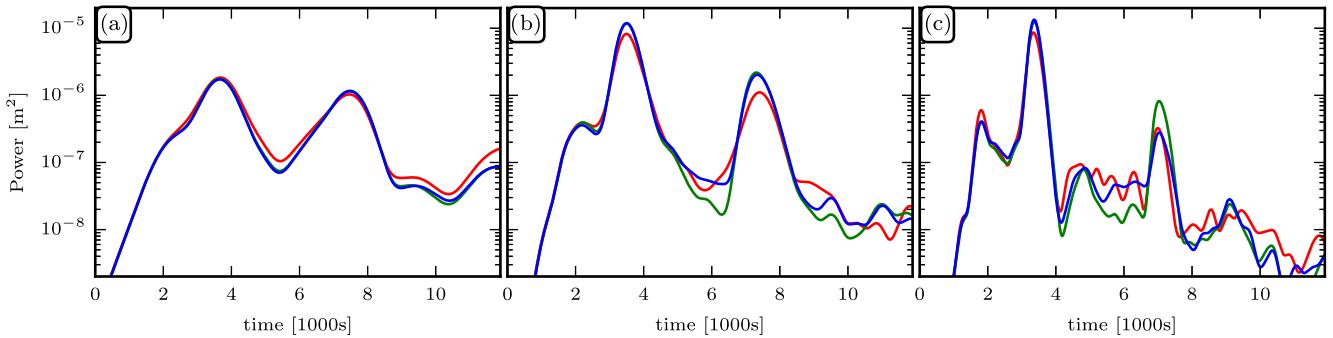


Figure 22. Average spectrogram power of the vertical displacement field in frequency bands (a) 400–200 s, (b) 200–100 s and (c) 100–50 s period, showing data (red), the original tomographic model (green) and the model with medium strength statistical heterogeneities (blue). The corresponding spectrogram is shown in Fig. 21. Note that a higher envelope away from the fundamental mode R1 and R2 can also be due to a different source mechanism and not to waves that are scattered from lateral heterogeneities. Scattered waves can also ‘decrease’ the envelope at a single station, for example, through destructive interference with other direct phases, in contrast to the average envelope of many stations that always increases.

In row 2, we compute difference spectrograms that indicate additional (red) or reduced (blue) energy in the data. Indeed, in this example, we see additional energy in the data between the arrivals R1 and R2 compared to a computation based on model (b). The direct fundamental mode surface waves themselves are weaker than in the computation. The extended model is able to model this effect and produce scattered energy between R1 and R2, as can be seen in the difference spectrogram of model (c) that becomes equally blue and red. Of course, a random model will not be able to generate the scattered energy at exactly the same location, but the average variances of the seismograms become similar.

The average seismogram power (square envelope) in three different period bands from 400–200 s (a), 200–100 s (b) and 100–50 s (c) is shown in Fig. 22 for the data (red), tomographic model (green) and the extended model with medium strength heterogeneities (blue). Here, we have normalized the maximum amplitudes of the seismograms to 1, to ensure a better comparison between energy in and energy out of the direct fundamental mode Rayleigh wave arrivals.

Again, more energy between the fundamental mode arrivals can be seen in the data trace, compared to the synthetic trace based on the original tomographic model, and the extended model is partially able to model this.

As we have mentioned before, this is only a proof of concept. For a meaningful measurement of the variance of the scattered wavefield, we need to examine many events and recordings at many stations. At a single station and recording, additional scatterers can even decrease the coda intensity compared to the direct arrivals, for example due to focusing or defocusing effects. The influence of a more complex or mislocated source will be important: For example, stronger overtone radiation can increase energy between the fundamental mode surface waves and should not be misinterpreted as waves that are scattered due to lateral heterogeneity. In our example, in Fig. 22, it can be seen that even before R1, coda energy is higher in the data than in our simulation. This can indicate that overtones are enhanced in this direction, which could also explain some of the additional energy that appears later.

Also a crustal structure that is more complex than the model Crust2.0 needs to be examined. In a simple test (not shown), we found that exchanging the crustal model Crust2.0 for the newer model Crust1.0 has only a very small effect on coda or scattering attenuation at these periods. We expect that stronger but probably reasonable variations of the elastic structure in the crust could produce a similar coda and similar scattering attenuation as unresolved mantle structure but with somewhat different frequency dependence. Variations definitely have to be stronger in the crust than in the mantle to have an effect on the long-period coda, because it is very thin in large regions, but again this depends on frequency and the exact wavefield sensitivity. Additionally, unmodelled, sharp transitions in moho and surface topography can play an important role.

Due to all of these possible influences, a detailed study beyond this article is necessary to (1) better isolate scattered energy that is due to lateral heterogeneities, (2) associate observed scattered energy with particular phases and (3) examine topography or anisotropy, or other possible perturbations, for example, in the crust that can also generate a coda.

6 CONCLUSIONS

We have presented a method that can be used to construct non-stationary and anisotropic Gaussian random models. In particular, we have proposed to combine a spectral description for components with stationary statistics with the computationally intense Karhunen-Loëve description for the non-stationary components.

We then developed spectral representations for stationary second-order random models in 2-D (see the Appendix) and 3-D polar coordinates. These can be used to generate model realizations in the solid sphere with radially non-stationary and anisotropic covariance functions. Such models can, for example, represent the interior structure of spherical bodies such as the Earth or other planets.

We have applied this method to examine the effect of unresolved small-scale mantle structure on long-period seismic waveforms. A coda and scattering attenuation develops in models with heterogeneities that are similar in power as the tomographic larger scale structures.

Extrinsic attenuation of the long-period waves due to this structure is weak and can explain only around 0.5–2 per cent of the total attenuation if perturbations of the elastic mantle structure do not become significantly stronger at smaller scales. Extrinsic attenuation furthermore scales linearly with model power, even in a complex model, as is expected in the linear scattering regime.

Like scattering attenuation, coda intensity is weak and depends on the variance of the small-scale structure. From measurements of the change in seismogram variance, away from direct surface waves arrivals, we are able to distinguish models with different small-scale power. It is difficult to associate this coda with individual phases because it overlaps with many direct waves at these long periods. However, depending on the station location, the coda becomes stronger after minor and before major arc arrivals and can therefore mostly be associated with fundamental mode surface waves. In our particular measurement method, a linear trend between coda intensity and model variance is visible at stations that are not too close to the source or to the antipode.

At long periods, layered, horizontally elongated heterogeneities have less influence on the coda, as we would expect from the depth sensitivity of the fundamental mode Rayleigh waves. A sparse distribution of scatterers that is correlated to the large-scale structure

generates a similar coda as a dense distribution of uncorrelated heterogeneities as long as they have the same spatially averaged covariance function.

Comparison with data demonstrates that random models are capable of producing observed scattered energy that is unmodelled in current tomographic models but is qualitatively similar to that observed in real data.

A detailed quantitative comparison with data has not been done and is beyond the scope of this paper. Furthermore, models are so far limited to mantle perturbations. The crust is only taken into account through the *a priori* models Crust1.0 and Crust2.0 that have only a minor effect on the results.

ACKNOWLEDGEMENTS

This work was supported by the European Research Council under the EC's 7th Framework Programme (FP7-IDEAS-ERC)/ERC Advanced Grant WAVETOMO. We also thank Nicholas Mancinelli and Marie Calvet for very constructive and helpful comments.

REFERENCES

- Abrahamsen, P., 1997. *A Review of Gaussian Random Fields and Correlation Functions*, Norsk Regnesentral/Norwegian Computing Center.
- Baddour, N., 2009. Operational and convolution properties of two-dimensional Fourier transforms in polar coordinates, *JOSA A*, **26**(8), 1768–1778.
- Baldi, P. & Marinucci, D., 2007. Some characterizations of the spherical harmonics coefficients for isotropic random fields, *Statistics & Probability Letters*, **77**(5), 490–496.
- Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in north America, *EOS, Trans. Am. geophys. Un.*, **81**, F897.
- Blanchette-Guertin, J.-F., Johnson, C. & Lawrence, J., 2012. Investigation of scattering in lunar seismic coda, *J. geophys. Res.*, **117**(E6), doi:10.1029/2011JE004042.
- Calvet, M. & Margerin, L., 2008. Constraints on grain size and stable iron phases in the uppermost inner core from multiple scattering modeling of seismic velocity and attenuation, *Earth planet. Sci. Lett.*, **267**(1–2), 200–212.
- Calvet, M. & Margerin, L., 2012. Velocity and attenuation of scalar and elastic waves in random media: a spectral function approach., *J. acoust. Soc. Am.*, **131**(3), 1843–1862.
- Calvet, M. & Margerin, L., 2013. Lapse-time dependence of coda Q: anisotropic multiple-scattering models and application to the Pyrenees, *Bull. seism. Soc. Am.*, **103**(3), 1993–2010.
- Capdeville, Y., Stutzmann, E., Wang, N. & Montagner, J.-P., 2013. Residual homogenization for seismic forward and inverse problems in layered media, *Geophys. J. Int.*, **194**(1), 470–487.
- Cormier, V.F., 1999. Anisotropy of heterogeneity scale lengths in the lower mantle from pkp precursors, *Geophys. J. Int.*, **136**(2), 373–384.
- Cormier, V.F., 2002. Frequency-dependent seismic attenuation in the inner core 2. A scattering and fabric interpretation, *J. geophys. Res.*, **107**(B12), 2362, doi:10.1029/2002JB001796.
- Cormier, V.F., 2007. Texture of the uppermost inner core from forward- and back-scattered seismic waves, *Earth planet. Sci. Lett.*, **258**(3–4), 442–453.
- Dainty, A.M. & Toksöz, M.N., 1981. Seismic codas on the earth and the moon: a comparison, *Phys. Earth planet. Inter.*, **26**(4), 250–260.
- Dalton, C.A., Ekström, G. & Dziewoński, A.M., 2008. The global attenuation structure of the upper mantle, *J. geophys. Res.*, **113**(B9), doi:10.1029/2007JB005429.
- Dalton, C.A., Hjorleifsdottir, V. & Ekstrom, G., 2013. A comparison of approaches to the prediction of surface wave amplitude, *Geophys. J. Int.*, **196**(1), 386–404.

- Durek, J.J. & Ekström, G., 1996. A radial model of anelasticity consistent with long-period surface-wave attenuation, *Bull. seism. Soc. Am.*, **86**(1A), 144–158.
- Dziewoński, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**(4), 297–356.
- Ekström, G., Tromp, J. & Larson, E.W.F., 1997. Measurements and global models of surface wave propagation, *J. geophys. Res.*, **102**, 8137–8157.
- Frankel, A., 1989. A review of numerical experiments on seismic wave scattering, *Pure appl. Geophys.*, **131**(4), 639–685.
- Frankel, A. & Clayton, R., 1986. Finite difference simulations of seismic scattering: implications for the propagation of short period seismic waves in the crust and models of crustal heterogeneity, *J. geophys. Res.*, **91**, 6465–6489.
- French, S. & Romanowicz, B., 2014. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography, *Geophys. J. Int.*, **199**(3), 1303–1327.
- Garcia, R.F., Chevrot, S. & Calvet, M., 2009. Statistical study of seismic heterogeneities at the base of the mantle from PKP differential traveltimes, *Geophys. J. Int.*, **179**(3), 1607–1616.
- Gung, Y. & Romanowicz, B., 2004. Q tomography of the upper mantle using three-component long-period waveforms, *Geophys. J. Int.*, **157**(2), 813–830.
- Holschneider, M., 1996. Continuous wavelet transforms on the sphere, *J. Math. Phys.*, **37**(8), 4156–4165.
- Hong, T.-K. & Wu, R.-S., 2005. Scattering of elastic waves in geometrically anisotropic random media and its implication to sounding of heterogeneity in the Earth's deep interior, *Geophys. J. Int.*, **163**(1), 324–338.
- Jahnke, G., Thorne, M.S., Cochard, A. & Igel, H., 2008. Global SH-wave propagation using a parallel axisymmetric spherical finite-difference scheme: application to whole mantle scattering, *Geophys. J. Int.*, **173**(3), 815–826.
- Klimeš, L., 2002. Correlation Functions of Random Media, *Pure appl. Geophys.*, **159**(7–8), 1811–1831.
- Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation—I. Validation, *Phys. Earth planet. Inter.*, **149**, 390–412.
- Komatitsch, D. & Tromp, J., 2002b. Spectral-element simulations of global seismic wave propagation—II. Three-dimensional models, oceans, rotation and self-gravitation, *Earth*, **150**, 303–318.
- Laske, G. & Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data, *J. geophys. Res.*, **101**, 59–75.
- Laske, G., Masters, G., Ma, Z. & Pasyanos, M., 2013. Update on crust1.0-a 1-degree global model of earth's crust, *Geophys. Res. Abstracts*, **15**, p. 2658.
- Lindgren, F., Rue, H. & Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(4), 423–498.
- Maeda, T., Sato, H. & Ohtake, M., 2003. Synthesis of Rayleigh-wave envelope on the spherical Earth: analytic solution of the single isotropic-scattering model for a circular source radiation, *Geophys. Res. Lett.*, **30**(6), 1286, doi:10.1029/2002GL016629.
- Mancinelli, N.J. & Shearer, P.M., 2013. Reconciling discrepancies among estimates of small-scale mantle heterogeneity from pkp precursors, *Geophys. J. Int.*, **195**(3), 1721–1729.
- Margerin, L., 2006. Attenuation, transport and diffusion of scalar waves in textured random media, *Tectonophysics*, **416**(1–4), 229–244.
- Margerin, L. & Nolet, G., 2003. Multiple scattering of high-frequency seismic waves in the deep earth: PKP precursor analysis and inversion for mantle granularity, *J. geophys. Res.*, **108**(B11), doi:10.1029/2003JB002455.
- Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M. & Renard, P., 2013. A practical guide to performing multiple-point statistical simulations with the Direct Sampling algorithm, *Comput. Geosci.*, **52**, 307–324.
- Meschede, M. & Romanowicz, B., 2015. Lateral heterogeneity scales in regional and global upper mantle shear velocity models, *Geophys. J. Int.*, **200**(2), 1078–1095.
- Nakamura, Y., Dorman, J., Duennebier, F., Lammlein, D. & Latham, G., 1975. Shallow lunar structure determined from the passive seismic experiment, *The Moon*, **13**(1–3), 57–66.
- Ricard, Y., Durand, S., Montagner, J.-P. & Chambat, F., 2014. Is there seismic attenuation in the mantle?, *Earth planet. Sci. Lett.*, **388**, 257–264.
- Romanowicz, B.a., 1990. The upper mantle degree 2: constraints and inferences from global mantle wave attenuation measurements, *J. geophys. Res.*, **95**, doi:10.1029/JB095iB07p11051.
- Ryzhik, L., Papanicolaou, G. & Keller, J.B., 1996. Transport equations for elastic and other waves in random media, *Wave Motion*, **24**(4), 327–370.
- Sato, H., 2002. Multiple isotropic-scattering model on the spherical Earth for the synthesis of Rayleigh-wave envelopes, *J. geophys. Res.*, **107**(B12), ESE 7–1–ESE 7–9.
- Sato, H. & Nohechi, M., 2001. Envelope formation of long-period Rayleigh waves in vertical component seismograms: single isotropic scattering model, *J. geophys. Res.*, **106**, 6589–6594.
- Sato, H., Fehler, M.C. & Maeda, T., 2012. *Seismic Wave Propagation and Scattering in the Heterogeneous Earth*, Springer.
- Shearer, P.M. & Earle, P.S., 2008. Observing and modeling elastic scattering in the deep earth, *Advances in Geophysics*, **50**, 167–193.
- Stein, M.L., 1999. *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Science & Business Media.
- Thorne, M., Myers, S., Harris, D. & Rodgers, A., 2006. Finite difference simulation of seismic scattering in random media generated with the karhunen-loève transform, in *AGU Fall Meeting Abstracts*, vol. 1, p. 05.
- Watson, G.N., 1966. *A Treatise on the Theory of Bessel Functions*, 2nd edn, Cambridge Univ. Press.
- Weaver, R.L., 1990. Diffusivity of ultrasound in polycrystals, *J. Mech. Phys. Solids*, **38**(1), 55–86.
- Yang, Y., Forsyth, D.W. & Weeraratne, D.S., 2007. Seismic attenuation near the East Pacific Rise and the origin of the low-velocity zone, *Earth planet. Sci. Lett.*, **258**(1–2), 260–268.

APPENDIX: RANDOM MODELS IN POLAR COORDINATES

A1 The polar covariance matrix

In 2-D polar coordinates, the covariance matrix of a Gaussian random medium can be written as $C(r, r', \phi, \phi')$. It is useful if statistical properties of the medium depend on the radial and angular location and distances. In case of a medium with point (cylindrical) symmetric statistics, the angular component of the polar covariance function is stationary, and can therefore be represented by its discrete 1-D Fourier power spectrum (discrete because it is 2π periodic). As a consequence, its angular Fourier coefficients c_k are independent with covariance function $C_k(r, r') = \langle c^*(r)_k c(r')_k \rangle$, or, on a discrete radial grid r^i , with covariance matrix $C_k^{ij} = \langle c_k^{i*} c_k^j \rangle$. Here we used the upper index for the radial non-stationary component and the lower index for the stationary one.

A polar random model can be constructed from a given covariance matrix C_k^{ij} as described in the main article: with the eigenbasis of C we generate correlated normally distributed coefficients c_k^i whose Fourier transform corresponds to the model values at radius r^i . C itself can, for example, be constructed from the angular and radial statistics of the model: The diagonal $C_{ii}^k = |c_k^i|^2$ is defined by the horizontal Fourier power at each radius r^i of the medium that we wish to realize. It fixes the angular covariance function of the medium. The off-diagonal elements of C_k^{ij} are defined by the vertical correlation between the Fourier basis functions at different radii. Usually, this correlation depends on k if the covariance matrix is not separable in the radial and angular direction. Once the covariance matrix is defined, the model can be constructed.

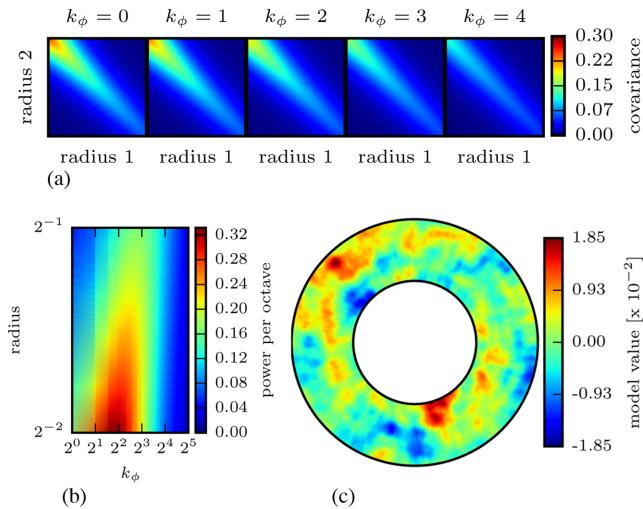


Figure A1. Illustration of a non-stationary polar model covariance matrix and corresponding model realization. The stationary angular model statistics are described by their Fourier power spectrum and the radial statistics by a covariance matrix. (a) Imposed covariance matrices of the first five angular Fourier coefficients. These matrices are non-stationary, because coefficient power (covariance) decreases with increasing radius. (b) Horizontal power spectra as a function of radius. With increasing radius, power shifts to higher degree coefficients. (c) Model realization in polar coordinates. The non-stationary polar covariance matrices generate model which seems roughly stationary in Cartesian space. This is a consequence of increasing angular scales with radius that the particular covariance matrix in this example tries to account for.

An example of *a priori* covariance matrices with $k_\phi < 5$ is shown in Fig. A1(a) where we have imposed an angular and radial covariance of our choice: at each radius, we have selected the horizontal power spectrum of a 2-D exponential medium with an angular scale length that is proportional to the radius. We thus keep the Cartesian scale length roughly constant. As vertical covariance function we have also chosen an exponential medium. All matrices are non-stationary as their entries depend on radius. The horizontal power spectrum as a function of radius can be seen in Fig. A1(b): away from the centre, power shifts towards higher degrees and therefore smaller angular scales, as is necessary to keep the Cartesian model proportions approximately constant. Fig. A1(c) shows a model realization with these statistical properties. It looks approximately stationary and isotropic and we have full control over the radial and angular covariance function of this model.

This approach is useful if a model has well-defined and separable angular and radial statistics, but in many situations this is not the case. In particular, it is essential how the covariance matrix looks like for a model with given stationary Cartesian statistics. In the previous example, we used an improper ‘trick’ and rescaled the angular model scales with radius to achieve model statistics that are stationary and isotropic in the Cartesian sense. On purpose, we have not shown a model that contains the centre point, in Fig. A1(c), where spherical distortions become very strong where this approximation does not work anymore.

A2 The relation of the Cartesian and polar covariance

A model with stationary Cartesian $C(\mathbf{r}, \mathbf{r}')$ or power spectrum $|F(\mathbf{k})|^2$ has naturally a non-stationary covariance function $C(r, r')$,

$k_\phi)$ because the relation between angular and Cartesian distances depends on radius. The connection between both matrices can be found by expanding the model $f(r, \phi)$ in terms of its angular Fourier coefficients:

$$f(r, \phi) = \sum_n f_n(r) e^{in\phi} \quad (A1)$$

$$f_n(r) = \frac{1}{2\pi} \int_{\mathbb{R}} d\phi f(r, \phi) e^{in\phi}. \quad (A2)$$

These coefficients can be synthesized from the Cartesian Fourier spectrum $F(\mathbf{k})$ using a ‘cylindrical wave expansion’:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_n J_n(\rho r) i^n e^{in\phi} e^{-in\theta}. \quad (A3)$$

Here, (ρ, θ) are the radial and angular components of the wavenumber vector \mathbf{k} . With this expansion, the forward and inverse Fourier transforms (eqs 4 and 5) can be written in terms of polar coordinates:

$$F(\rho, \theta) = 2\pi \sum_n i^n e^{in\theta} \int_{\mathbb{R}} f_n(r) J_n(\rho r) r dr \quad (A4)$$

$$f(r, \phi) = \frac{1}{2\pi} \sum_n i^n e^{in\phi} \int_{\mathbb{R}} F_n(\rho) J_n(\rho r) \rho d\rho. \quad (A5)$$

$F_n(\rho)$ are the angular Fourier coefficients of the Cartesian Fourier spectrum $F(\mathbf{k})$. This transform pair (eqs A4 and A5) (‘Hankel’ transform) relates polar Fourier space with polar model space and can therefore be seen as an equivalent of the Fourier transform in polar coordinates. The angular model coefficients $f_n(r)$ that are required to generate a realization can be computed from eqs (A5)–(A2),

$$f_n(r) = \frac{i^{-n}}{2\pi} \int_{\mathbb{R}} F(\rho, \theta) J_n(\rho r) e^{in\theta} \rho d\rho d\theta \quad (A6)$$

$$f_n(r) = i^{-n} \int_{\mathbb{R}} F_n(\rho) J_n(\rho r) \rho d\rho. \quad (A7)$$

To compute the angular model coefficients f_n , we first need to calculate the angular coefficients F_n of the Fourier spectrum.

If we consider a stationary Gaussian random model with Cartesian spectrum $\langle F^*(\mathbf{k})F(\mathbf{k}') \rangle = |F(\mathbf{k})|^2 \delta(\mathbf{k} - \mathbf{k}')$, we obtain for the covariance of the coefficients F_n :

$$\langle F_n^*(\rho) F_{n'}(\rho') \rangle = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \langle F^*(\rho, \theta) F(\rho', \theta') \rangle e^{in\theta - in'\theta'} d\theta d\theta' \quad (A8)$$

$$= \frac{|F(\rho)|^2}{2\pi\rho} \delta_{nn'} \delta(\rho - \rho'). \quad (A9)$$

The polar Fourier coefficients F_n are therefore also independent. Their power $\langle |F_n(\rho)|^2 \rangle$ is similar to the Cartesian power $|F(\rho)|^2$ that is evenly distributed over the angular space $2\pi\rho$.

With the coefficients F_n , a stationary Gaussian model can be directly synthesized in polar coordinates using eq. (A5). Furthermore, from eq. (A7) we can compute the polar covariance

matrix $C_{nn'}(r, r')$:

$$C_{nn'}(r, r') = \langle f_n^*(r) f_{n'}(r') \rangle \quad (\text{A10})$$

$$= \delta_{nn'} \int_{\mathbb{R}} \frac{\langle |F(\rho)|^2 \rangle}{2\pi\rho} J_n(\rho r) J_{n'}(\rho r') \rho^2 d\rho. \quad (\text{A11})$$

From this covariance matrix we can also generate isotropic stationary models in polar coordinates but importantly, we can use it as a basis to construct non-stationary models as well. Different, or smoothly varying Cartesian spectra could for example be chose in different regions.

Fig. A2 shows a comparison between two models with similar Fourier spectrum, generated in Cartesian and polar coordinates, respectively. Both models are similar but in the polar model we have control over the model's radial and angular statistics.

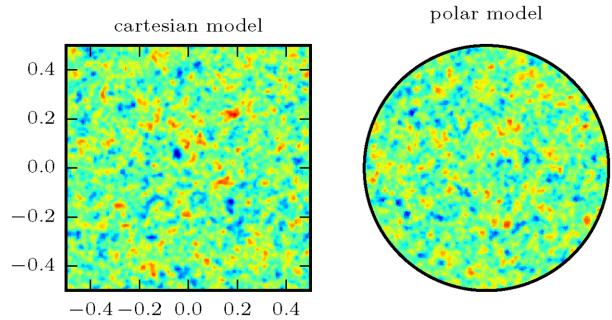


Figure A2. Isotropic Gaussian random field with exponential autocorrelation function in Cartesian (left) and in polar (right) coordinates. The Fourier-Bessel basis allows to generate models that are stationary and isotropic in the Cartesian sense in polar coordinates.