Abstract. A Marquardt, nonlinear, iterative inversion is applied to obtain the best-fit parameters for 7 spectral models of heterogeneities under NORSAR using the observed TCF (Transverse Coherence Functions) and ACF (Angular Coherence Functions) at NORSAR. The simplest model is a single-layer model with one power-law spectrum; the most complex model is a two-overlapped-layer model with two different power-law spectra specified by 6 unknown parameters. F-tests are performed for different pairs of models; the two-layer model with two different spectra is a significant improvement over the single layer, one spectrum model. The final results of nonlinear inversion agree with the Flatt-Wu spectral model for heterogeneities under NORSAR.

Keywords: nonlinear inversion, lithospheric heterogeneity, transmission fluctuations, NORSAR

Introduction

The stochastic approach of inverting the transmission fluctuations across a seismic array to derive the statistical properties of heterogeneities under the array has been applied since the early seventies (for a review see Wu and Flatt 1990). Recently Flatt and Wu (1988) introduced the ACF (Angular Coherence Functions) of phase and log amplitude fluctuations for waves with different incident angles. The newly introduced ACF together with the TCF (Transverse Coherence Functions) allow inversion for more general models of nonuniform, multi-scale random media. In this paper we represent models with a small number of parameters, and use a nonlinear method (Marquardt 1963; Press et al. 1986) to find the best parameter values for different models of heterogeneities under the Norwegian Seismic Array (NORSAR). For this study, NORSAR consisted of about 130 seismometers spread over a 110-km-diameter region in Southern Norway. We confine the model of random heterogeneities to be no more complex than a model of two overlapped layers having different power-law spectra (for a comparison of different models see Table 1). This two-layer model has been used by Flatt and Wu (1988) to fit the observed ACF and TCF by a trial and error method. In this paper we will test the Flatt-Wu model using our nonlinear inversion method and compare the goodness of fit to the observations for different models. Finally an F test is conducted to measure the significance of the improvement in fit as the model complexity increases.

Coherence Functions TCF, ACF and the Inversion for Heterogeneity Spectra

The TCF (Transverse Coherence Functions) measure the transverse coherence of the fluctuations as a function of transverse distance x between stations. Let φ be the phase fluctuation of a wave arriving at seismometer i, relative to the phase of arrival of a plane wave fitted to all the seismometers. Let u be the log amplitude fluctuation at seismometer i, relative to the mean u for all seismometers. If we assume small angles with respect to the vertical, and horizontal isotropy for the heterogeneity power spectrum W(Kr,Kz,z), where Kr is the horizontal wavenumber, Kz is the vertical wavenumber, and z is the depth, the TCF can be written as (for details see Flatt and Wu (1988), Wu and Flatt (1990) and Flatt and Moody 1990):

\[
\langle u_1, u_2 \rangle(x) = \frac{4\pi k^2}{R} \int [K_T dK_T J_0(K_T x)] \sin^2 \left( \frac{K_T^2 z}{2k} \right) W(K_T, 0, z) \, dz
\]

\[
\langle \phi_1, \phi_2 \rangle(x) = \frac{4\pi k^2}{R} \int [K_T dK_T J_0(K_T x)] \cos^2 \left( \frac{K_T^2 z}{2k} \right) W(K_T, 0, z) \, dz
\]

where in each case seismometer 1 and 2 are separated by a transverse distance x, and where \( \langle \quad \rangle \) stands for the ensemble average over many observations, k = 2π/λ, λ is the wave length of the propagating wave in the background medium, and R is the thickness of the heterogeneous layer.

The ACF (Angular Coherence Functions) measure the coherence between the fluctuations at a given seismometer of any two incident plane waves with different incident angles:

\[
\langle u_1, u_2 \rangle(\theta) = \frac{4\pi k^2}{R} \int [K_T dK_T J_0(K_T z \theta)] \sin^2 \left( \frac{K_T^2 z}{2k} \right) W(K_T, 0, z) \, dz
\]

\[
\langle \phi_1, \phi_2 \rangle(\theta) = \frac{4\pi k^2}{R} \int [K_T dK_T J_0(K_T z \theta)] \cos^2 \left( \frac{K_T^2 z}{2k} \right) W(K_T, 0, z) \, dz
\]
\[
\langle u_1 \phi_2(\theta) \rangle = 4\pi k^2 \int_0^R \! dz \left[ K_T dK_T J_0(K_T z \theta) \right] \cdot \\
\sin \left( \frac{K_T^2 z}{2k} \right) \cos \left( \frac{K_T^2 z}{2k} \right) W(K_T, 0, z)
\]

where \( \theta \) is the angular separation between the pair of incident waves, labeled by 1 and 2.

It is convenient to separate the strength (variances) from the shape of the coherence function by normalizing:

\[
U(x) = \frac{\langle u_1 u_2(x) \rangle}{\langle u_2 \rangle}, \quad \Phi(x) = \frac{\langle \phi_1 \phi_2(x) \rangle}{\langle \phi_2 \rangle}, \quad \text{and} \quad C(x) = \frac{\langle u_1 \phi_2(x) \rangle}{\langle u_1 \rangle \langle \phi_2 \rangle} \frac{\langle u_2 \rangle}{\langle \phi_2 \rangle} \frac{1}{2}
\]

are the normalized TCF for log amplitude, phase, and their cross-correlation. Similar expressions can be written for the normalized ACF: \( U(\theta), \Phi(\theta) \) and \( C(\theta) \). We will use the normalized coherence functions to invert for the shape of heterogeneity spectrum and leave the overall magnitude of the spectrum to be determined by the variances.

The data set (normalized TCF and ACF) was derived by Flatté and Wu (1988) from the NORSAR observations (see Figure 1).

There are 7 models being tested (Table 1). The most complex model, G, has 6 parameters, with two overlapped layers, each layer having a different power-law spectrum \( (W_1 = K^{-p_1}, W_2 = bK^{-p_2}) \). From model G to A, the unknown parameters are gradually reduced. The spectrum can be reconstructed only within a wavenumber band between a low wavenumber limit \( K_1 \) due to the aperture size of NORSAR \( (K_1 = 2\pi/110 \text{ km}^{-1}) \) and a high wavenumber cutoff due to the subarray (diameter 7 km) average \( (K_s = 2\pi/5.5 \text{ km}^{-1}) \). For detailed discussion see Flatté and Wu (1988).

For the inversion, (1) and (2) can be combined into:

\[
d = f(X) + \epsilon
\]

where \( d \) is the data vector containing the observations (coherence function points); \( X \) is the vector of unknowns, and \( \epsilon \) is the noise vector. Because the inversion is highly nonlinear, we adopt Marquardt's method of optimization to search for the best solution. For each step of the iteration, the model modification \( \Delta X \) can be found by solving the following matrix equation:

\[
(A + \Lambda) \Delta X = B \Delta d
\]

where \( \Delta d \) is the data residual, i.e. the difference between the observed data and the model predictions. \( A \) is the Hessian matrix composed of partial derivatives of \( f(X) \) w.r.t. \( X \). After neglecting the second derivatives, the elements of \( A \) and \( B \) can be expressed as

\[
(A)_{ij} = \sum_{i=1}^{N} \frac{w_i \frac{\partial f_j}{\partial x_i}}{\sigma_i^2} \frac{\partial f_j}{\partial x_k}
\]

where \( N \) is the number of the data points and \( \sigma_i \) is the standard deviation of the \( i \)th data point. \( \Lambda \) is a diagonal matrix with elements

\[
(\Lambda)_{ij} = \lambda (A)_{ij}
\]

where \( \lambda \) is a controlling factor inversely proportional to the search step length. When \( \lambda \) is very large, the diagonal elements of \( A \) will dominate the inversion, and equation (3) degenerates to a gradient method, which guarantees that the inversion will march a small step toward a local minimum (with a price of slow convergence). On the other hand, for very small \( \lambda \), equation (3) is equal to a Hessian inversion (Newton's method) which has a fast convergence rate when the solution is close to the minimum. The weighting factor \( w_i \) for the \( i \)th datum is given as:

\[
w_i = 1 - \frac{\theta_i}{\theta_{\text{max}}} \quad \text{or} \quad 1 - \frac{r_i}{r_{\text{max}}}
\]

for ACF or TCF to account for loss of accuracy of the theoretical prediction due to the effect of finite aperture.

From a set of initial parameters we calculate the partial derivatives for \( A \) and \( B \), and the residual \( \Delta d \). With a small \( \lambda \), say 0.001, we invert for the correction \( \Delta X \). If the chi-square of the new model is smaller than the old one, the new parameter set is accepted and the process is to be iterated with a smaller \( \lambda \) (bigger step), otherwise the process is to retreat one step back, and redo it with larger \( \lambda \) (small step). The iteration is stopped when the improvement in chi-square is smaller than a preset threshold or \( \lambda \) is too large. The number of iterations is typically around 20.

Fig. 1. Comparison between the predictions of model F (solid lines) and the observed coherence functions with their error bars (from NORSAR data). On the left are the Transverse Coherence Functions (TCF); on the right are the Angular Coherence Functions (ACF).
Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>description</th>
<th>layer extent</th>
<th>spectra</th>
<th>number of unknowns</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 layer, 1 spectrum</td>
<td>(L_1:(0-R_1))</td>
<td>(W=K^{-p_1})</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>1 layer, 1 spectrum</td>
<td>(L_1:(R_0-R_2))</td>
<td>(W=K^{-p_1})</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>1 layer, 2 spectra</td>
<td>(L_1:(0-R_1))</td>
<td>(W=K^{-p_1}+bK^{-p_2})</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>1 layer, 2 spectra</td>
<td>(L_1:(R_0-R_2))</td>
<td>(W=K^{-p_1}+bK^{-p_2})</td>
<td>5</td>
</tr>
<tr>
<td>E</td>
<td>2 layers, 2 spectra</td>
<td>(L_1:(0-R_1)); (L_2:(0-R_2))</td>
<td>(W_1=K^{-p_1}; W_2=bK^{-p_1})</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>2 layers, 2 spectra</td>
<td>(L_1:(0-R_1)); (L_2:(R_0-R_2))</td>
<td>(W_1=1; W_2=bK^{-p_1})</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>2 layers, 2 spectra</td>
<td>(L_1:(0-R_1)); (L_2:(R_0-R_2))</td>
<td>(W_1=K^{-p_1}; W_2=bK^{-p_1})</td>
<td>6</td>
</tr>
</tbody>
</table>

**Results of Inversion**

We have minimized the typical problem of nonlinear inversion (multiple minima) by not allowing the number of parameters to be very large. We have also started our searches with different initial conditions to provide assurance we have found the global minimum.

Seven models were run against the same data set, to see the significance of contributions from different spectral components and different depths. Results are summarized in Table 2. Model G has the largest number of unknowns and is a model of two overlapped layers. The other models are basically the same, but with one or more parameters fixed to certain values. The parameter \(b\) is the spectral ratio of \(W_2\) (the spectrum of the second layer) versus \(W_1\) (the spectrum of the first layer) at \(K = 1 \text{ km}^{-1}\). To avoid correlations between \(b\) and \(p_2\), we changed the inversion parameter to a dimensionless \(b'= bK^{-p_1}\). We found that fixing \(K_0 = 0.073 \text{ km}^{-1}\) was satisfactory in all cases tested. The number of degrees of freedom is \(v = M - N\), where \(M\) is the number of data points and \(N\) is the number of unknown parameters.

**F-tests** for some pairs of the models are shown in Table 3. The significance of the improvement in fit from one model to another is measured by the chi-square reduction given the difference in the number of degrees of freedom. If the errors are misestimated, or correlated between points, a more robust measure is obtained by doing an F-test. In a sense, \(P(F)\) can be regarded as the probability that the two models fit equally well. The use of the worse rather than the better fit in the denominator of the F-test makes the probabilities conservatively high.

By comparing model A, B, C, and D with E, F and G, we see that the two-layer models are definitely preferred over the one-layer models, and the power index of the first layer is not significantly different from zero. Since the F-test for model G versus model F results in a \(P(F) = 0.16\), we can consider model F and G as equally good in fitting the data. Therefore, we show here only the results of model F. In Figure 1, the solid lines are the model predictions of the coherence functions. We regard the fit to the data as good; this is confirmed by the \(\chi^2\) value of 114 for 127 degrees of freedom. The few outlying points are not unexpected with 130 points, and we have checked that these points are not caused by a systematic effect such as a particular seismometer or event. Model F is in good agreement with the Flatt-Wu model (1988). The inversion has now provided quantitative errors on the parameters of the Flatt-Wu model.

**Geophysical Interpretation**

The physical interpretation of the heterogeneity spectra under NORSAR has been discussed in detail in Flatt and Wu (1988). We summarize that discussion here and add remarks related to the inversion.

The southern section of NORSAR overlays the northern tip of the Oslo graben, whose structure is likely to be confined to shallow depths. The effect of a large-horizontal-scale deterministic structure on our fits is not significant, especially one that is confined to depths small compared with 100 km. Our statistical model naturally has few degrees of freedom at the larger horizontal scales, and our error bars reflect this fact.

The depth of the lithosphere in this geographical region was indicated by a number of studies to be in the range of 150 to 250 km. Those studies were sensitive to lateral scales that are larger than 100 km.

The use of two overlapping layers in our models was motivated by the possibility that either the Moho or the bottom of the lithosphere could be sites of changes in heterogeneity structure. Our fits were free to have the two layers

Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>(p_1)</th>
<th>(p_2)</th>
<th>(b')</th>
<th>(R_0)</th>
<th>(R_1)</th>
<th>(R_2)</th>
<th>(\chi^2)</th>
<th>(v)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2.63±0.01</td>
<td>-</td>
<td>-</td>
<td>4.7±3.1</td>
<td>263±23</td>
<td>-</td>
<td>178.9</td>
<td>130</td>
</tr>
<tr>
<td>B</td>
<td>-</td>
<td>2.62±0.08</td>
<td>-</td>
<td>4.7±3.1</td>
<td>268±13</td>
<td>-</td>
<td>174.9</td>
<td>129</td>
</tr>
<tr>
<td>C</td>
<td>1.59±0.45</td>
<td>5.47±1.09</td>
<td>51.7±8.6</td>
<td>-</td>
<td>268±13</td>
<td>-</td>
<td>131.5</td>
<td>128</td>
</tr>
<tr>
<td>D</td>
<td>1.48±0.50</td>
<td>5.05±1.05</td>
<td>45.7±9.6</td>
<td>9.0±3.1</td>
<td>239±16</td>
<td>-</td>
<td>129.7</td>
<td>127</td>
</tr>
<tr>
<td>E</td>
<td>-1.86±2.69</td>
<td>4.69±0.01</td>
<td>11.9±3.8</td>
<td>0</td>
<td>153±20</td>
<td>265±13</td>
<td>122.9</td>
<td>127</td>
</tr>
<tr>
<td>F</td>
<td>0*</td>
<td>4.99±0.44</td>
<td>19.1±3.4</td>
<td>16.2±4.8</td>
<td>172±14</td>
<td>255±14</td>
<td>113.6</td>
<td>127</td>
</tr>
<tr>
<td>G</td>
<td>-2.76±2.63</td>
<td>4.80±0.39</td>
<td>11.8±3.0</td>
<td>17.4±4.6</td>
<td>146±16</td>
<td>243±13</td>
<td>111.8</td>
<td>126</td>
</tr>
</tbody>
</table>
not overlap, but the results were an overlap in depth of 160 km out of 260 km. The layer extending to 260 km has a spectrum dominated by large-scale heterogeneities, while the layer whose upper surface is at the earth's surface is dominated by small-scale heterogeneities. The data demand that the small-scale heterogeneities extend downward to 175 km and the large-scale heterogeneities extend upward to near the surface. Within our parameterization, this overlapping of the small- and large-scale heterogeneities in depth is accomplished by the overlapping of the two layers, each with a different power-law spectrum.

We believe that a reasonable, though speculative, interpretation of our results is that the layer down to 175 km with small-scale heterogeneities represents the lithosphere, with its rigid nature capable of supporting small-scale temperature, compositional, or mechanical variations. The layer from 175 km to 255 km represents the more fluid upper asthenosphere, dynamically behaving like a thermal boundary layer at the top of the convecting mantle.

The fact that the larger-scale heterogeneities extend throughout the lithosphere would indicate that the processes occurring in the upper mantle are somehow "frozen in" during the formation of the lithosphere. However, since this is the continental lithosphere, it may be that a completely different process created the shallower (~100 km deep) large-scale structure, and our data are not sufficient to distinguish that large-scale process from the equally large-scale upper mantle fluid-dynamical process.

Conclusion

A Marquardt, nonlinear, iterative inversion was applied to obtained the best-fit parameters for 7 spectral models of heterogeneities under NORSAR (Table 1 and 2). F-tests for different pairs of models (Table 3) showed that the two-layer model with two different spectra provides significant improvement over the single-layer, one-spectrum model. The final result of nonlinear inversion is in good agreement with the Flatté-Wu spectral model for heterogeneities under NORSAR, and the inversion has provided statistical errors for the parameters of the model. The physical interpretation of the heterogeneity spectra under NORSAR has been discussed in detail in Flatté and Wu (1988). Briefly, the observations require multiscale heterogeneities down to 260 km depth; the best-fit models require a lithosphere-like layer with both large- and small-scale heterogeneities down to 175 km depth, and a deeper layer from 175 km to 260 km that is dominated by large-scale heterogeneities, which can be interpreted as an upper-mantle convective boundary layer. The specific parameters of the heterogeneity spectra have implications for dynamical models of the crust and upper mantle.

Acknowledgments. This work was supported by Defense Advanced Research Projects Agency grant F19628-89-K-0027 and is contribution #104 of the Institute of Tectonics, University of California, Santa Cruz. We are grateful for a grant from the W. M. Keck Foundation.

References

Flatté, S. M., and T. N. Moody, Weak-fluctuation theory via path integrals, Contribution #106 of the Institute of Tectonics of the University of California, Santa Cruz, 1990.


S. M. Flatté, and R. S. Wu, Institute of Tectonics, University of California, Santa Cruz, CA 95064

Z. K. Shen, Department of Earth and Space Sciences, University of California, Los Angeles, CA 90024

(Received: August 20, 1990 revised: March 15, 1991 accepted: April 1, 1991)