Geophys. J. Int. (2002) **150,** 665–672

Reliable mantle density error bars: an application of the neighbourhood algorithm to normal-mode and surface wave data

Joseph S. Resovsky and Jeannot Trampert

[1] *Utrecht University, Earth Sciences, PO Box* 80.021, 3508 *TA Utrecht, Netherlands. E-mail: resovsky@geo.uu.nl*

Accepted 2002 March 19. Received 2002 March 7; in original form 2001 July 10

SUMMARY

We demonstrate the feasibility and reliability of a new approach to seismic modelling of long-wavelength mantle structure. In doing so, we present new estimates of mantle model uncertainties that successfully explain apparent discrepancies in earlier 3D seismic density models and justify a new generation of model building. The characteristics of good models and of modelling error are described by model space 'maps' that display the data fit of a representative set of potential models. The neighbourhood algorithm (NA), recently developed by Malcolm Sambridge, allows efficient production of such maps for the long-wavelength model parametrizations used in imaging the mantle density using normal-mode data. We observe that when NA is applied to our global modelling problem the results are not only self-consistent and consistent with independent resolution tests, but also explain discrepancies in the results of damped inversions. Synthetic 'mapping' experiments using the sensitivity kernels and measurement errors of several data catalogues allow us to determine the resolution of these data sets. Such resolution tests let us know which model parametrizations and data sets, if any, can yield meaningful results in inversions of real data. The new resolution tests reveal that recent density models from damped inversions of normal-mode data are not robust. However, we also show that the addition of the most recent long-wavelength data will, for the first time, reduce model covariances enough to achieve robust constraints on the depth and size of long-wavelength density heterogeneity throughout the mantle.

Key words: mantle density, model error, normal modes, surface waves.

1 INTRODUCTION

Free-oscillation data have recently played an important role in improving the resolution of long-wavelength mantle models. Models MM2.L12D8 (Resovsky & Ritzwoller 1999a), SPRD6 (Ishii & Tromp 1999, 2001) and S20RTS (Ritsema & van Heijst 2000) all have velocity variations constrained to improve fits to normal-mode splitting function data, which are most extensive at spherical harmonic degrees two and four. MM2.L12D8 and S20RTS both use the normal model catalogue of Resovsky & Ritzwoller (1998) to constrain long-wavelength structure, while SPRD6 combines these data with those of Tromp & Zanzerkia (1995) and He & Tromp (1996). As a result, all three have significantly 'redder' spectra of mid-mantle velocity heterogeneity than do other seismic models, and they are much more consistent with each other at long wavelengths than were the previous generation of models.

Three-dimensional (3D) models of mantle density are more critical to understanding mantle composition and dynamics than are velocity models, but there are significant discrepancies associated with attempts to constrain density models using normal modes. Inversions such as those used in making SPRD6, in which the shear velocity (v_s), compressional velocity (v_p) and density (ρ) components are allowed to vary independently, are relatively rare. Model SPRD6 is accompanied by images of the resolution matrix for the inversions, and by tests with checkerboard and spike input models that examine some particular effects of this resolution. These results imply that the ρ variations and ρ–v_s scaling of this model are robust. However, checkerboard and spike tests have known weaknesses (Leveque *et al.* 1993), and much poorer resolutions were found in tests of similar inversions using different prior constraints (Resovsky & Ritzwoller 1999b; Romanowicz 2001) and in synthetic tests with different forms of input models (Kuo & Romanowicz 2002). Each of these studies concluded that few robust density features could be constrained with existing data. Romanowicz (2001) was able to bound the $d \ln \rho / d \ln v_s$ ratio at most depths, but Resovsky & Ritzwoller (1999b) concluded that ρ–v_s correlation could not be located at any particular depth. Such results imply that resolution matrices from damped inversions can significantly underestimate model error, and also that other resolution estimates depend upon the nature of the resolution test.

The source of such difficulties in model assessment is not mysterious. Rather, it is a familiar consequence of the existence of the model null space, the set of model parameter combinations that have negligible impact on the data fit. The null space is quite significant in the severely underdetermined inverse problems of global tomography. Even when the inverse problem is linear and the data errors are

Gaussian, as is the case for normal-mode inversions, the null space can produce strongly non-Gaussian behaviour in the distribution of likely models. If an inversion is to preserve the usual relationships between the data errors, resolution, and the covariance matrix of the model parameters, the Gaussian behaviour of the model space must be enforced by *a priori* constraints such as damping (Tarantola 1987). In a typical inversion, the parametrization and damping produce a stable result by allowing output models to include only a limited part of the null space, and inversion assessment explores only that part of the null space. Furthermore, the *a priori* constraints assign different weights to different null space model elements according to characteristics such as amplitude or smoothness, and this both distorts the null space and biases model assessments. Even as resolution at long wavelengths is improved by the addition of important new surface wave and normal-mode measurements (e.g. van Heijst & Woodhouse 1999; Masters *et al.* 2000; Widmer-Schnidrig 2002), it is clear that damped inversions cannot be trusted to give reliable estimates of either the resolution or improvement provided by these data. In other words, a more complete description of the model space, including some measure of the range of null space models, is needed before we can know whether it is possible to resolve the mantle density.

This is the primary motivation for implementing techniques that 'map' the mathematical topography (where 'highs' correspond to high likelihoods or good fits to the data and 'lows' to poor fits) of an undistorted multidimensional model space, including the null space. However, because such 'maps' are made by testing ('sampling') a large number of different combinations of model parameters, this approach can be prohibitively computationally intensive. An important new technique for the efficient sampling of a model space is the neighbourhood algorithm (NA) developed by Sambridge (1999a,b), which is much faster than regular or Monte Carlo sampling, and more easily tuned than genetic algorithms. The first successful applications of this algorithm were to inverse problems of up to 24 free parameters. Thus, NA seems well suited to the normal-mode inverse problem, which can be separated into individual inverse problems for each spherical harmonic component of lateral structure variations (Resovsky & Ritzwoller 1999a), and where the radial resolution of the data suggests that there should be approximately ten components of the radial parametrization of the mantle (e.g. Ishii & Tromp 2001; Romanowicz 2001).

The primary purpose of this paper is to demonstrate that the NA approach is, indeed, a feasible and reliable approach to the long-period global inverse problem. This must been shown in order to justify future model building and interpretation using the method. To test the method, we employ controlled experiments, with both synthetic and real data, for which certain idealized results are predictable if adequate 'maps' are produced by the NA procedure. Because our synthetic experiments employ the data error and sensitivity kernels of real data, these experiments also allow us to test the resolution of various normal-mode data sets. We hope that by demonstrating both a reliable technique and true resolution, this paper provides justification for a new generation of seismic models of mantle density and velocity.

2 THE SYNTHETIC EXPERIMENTS

2.1 Data and parametrization

In addition to several trials with real data, we examine the reliability of the NA approach with a series of synthetic experiments. These allow us to compare NA 'mapping' estimates of model precision (error bar) with observed modelling inaccuracies, where the latter are the differences between the output models from inversions and the input models used to generate the synthetic data.

The 'input' models used to create synthetic normal-mode data are either the spherically symmetric model PREM (Dziewonski & Anderson 1981) or PREM plus the volumetric part of aspherical model SPRD6. These are each used to construct two data sets: one of 132 splitting functions corresponding to those used by Ishii & Tromp (1999, 2001) (data set 1); and one matching a set of 205 splitting functions with which we hope to construct new long-wavelength mantle models (data set 2). The former data set is a combination of the catalogues of He & Tromp (1996) and Resovsky & Ritzwoller (1998), which includes several dozen pairs of inconsistent (differences greater than combined errors) measurements for the same structure coefficient. The latter data set includes a non-conflicting combination of: the two catalogues of data set 1; the catalogue of Masters *et al.* (2000); several new generalized spectral fitting measurements that improve the resolution of deep-mantle v_p (Pestana 2001); and splitting functions corresponding to 75 different fundamental and overtone surface wave velocity maps from three catalogues (Wong 1989; van Heijst & Woodhouse 1999; Trampert & Woodhouse 2001). The synthetic data are modified by synthetic noise drawn from Gaussian distributions matching the standard deviations reported for real data measurements (or those we have inferred for some surface wave data, as in Beghein *et al.* (2002)).

From a 'starting' model (PREM or PREM + SKS12WM13 of Liu & Dziewonski 1994) we then either perform damped least-squared inversions or use the NA procedure to find perturbations that best fit the synthetic data. The perturbations are parametrized as changes to the mean values of v_s, v_p and ρ (independently) in each of four lower-mantle layers (layer 1 = 2891–2360 km depth; layer 2 = 2360–1800; layer 3 = 1800–1050 km; layer 4 = 1050–670 km) and three upper-mantle layers (layer 5 = 670–400 km; layer 6 = 400–220 km; layer 7 = 220–24 km). The widths of the layers are derived from the sensitivity kernels of Backus & Gilbert (1968) for the data set from Resovsky & Ritzwoller (1998). Our inversions are subjected to an overall damping that produces the most stable combination of the size of the perturbations and the fit to the data (e.g. Backus & Gilbert 1970). The misfit levels that result are within 1 per cent of the best misfits observed with negligible damping and the selected damping level is nearly identical to the damping that minimizes the input–output discrepancies discussed below.

The seven-layer inversion is roughly equivalent to the seven eigenfunction inversions performed by Ishii & Tromp (1999, 2001). Though layered inversions do not produce realistically smooth models, they have approximately the same resolution and can identify the same robust structural features as inversions that enforce smoothness *a priori* through choices of parametrization and damping. This is demonstrated by the similarity in the 3D v_s components of the models of Ishii & Tromp (2001) and Resovsky & Ritzwoller (1999a). We have found the nature of our results to be independent of the combination of input and starting models we use.

2.2 Input–output discrepancies

We observe two forms of discrepancy when we compare the damped least-squares (LS) inversion results to the input models. Our primary test of the NA approach will be its ability to account for these discrepancies.

The inversion produces a set of layer perturbations, which are added to the starting model to give an output model. The

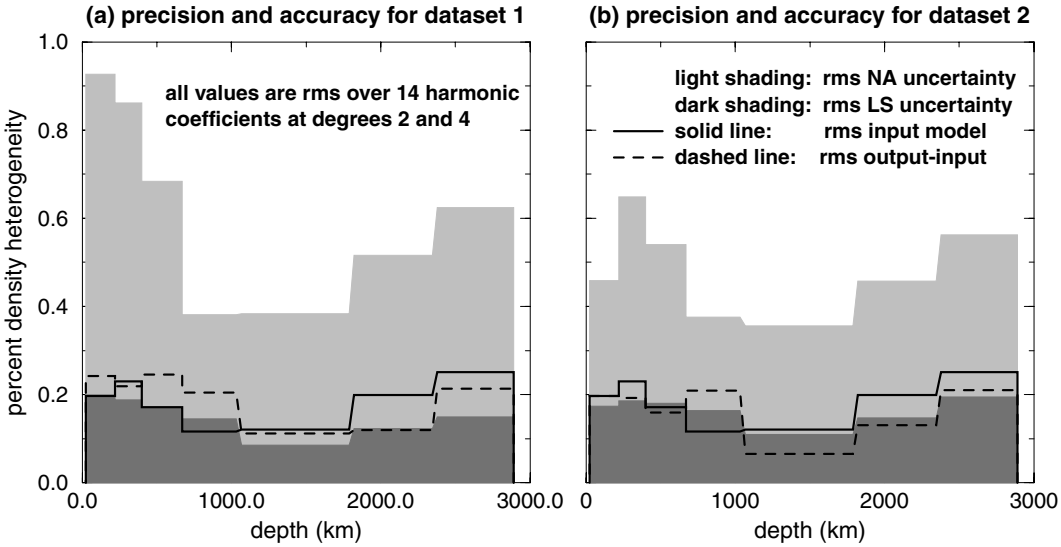

Figure 1. Rms amplitudes of density model inaccuracies and uncertainties, compared with an input model. The input model, the rms of SPRD6 at degrees two and four averaged in each of seven layers, is shown by the solid lines. Dashed lines indicate the rms of inaccuracy (the difference between the input model and output models) for damped least-squared inversions of two synthetic data sets. The dark shading indicates the model uncertainty predicted by the covariances of the damped LS inversions, and light shading shows the model uncertainty estimates from NA 'mapping'. Because the dark shading is usually lower than the level of the dashed lines, there is a discrepancy between the observed and predicted accuracy from the damped inversions. (a) Data set 1 consists of 132 normal-mode splitting functions matching the data used in constructing model SPRD6 (Ishii & Tromp 2001). (b) Data set 2 consists of 205 splitting functions corresponding to the normal-mode and surface waves data that we will be using to 'map' new constraints on mantle density models. The NA uncertainties account for all output–input differences and show the expected improvement from surface wave constraints on the upper mantle.

'input–output' difference, or modelling inaccuracy, is the mean difference between the input model and the output model in each layer. The LS inversion also produces a covariance matrix for the layer perturbation parameters. The covariance matrix is closely related to the resolution matrix, and its diagonal elements give the (square of the) model layer uncertainties or standard deviations (Tarantola 1987).

Fig. 1 provides examples, using data sets 1 and 2, of the input–output difference and LS inversion uncertainty for each layer. This figure shows root-mean-square (rms) amplitudes over 14 degree-two and four spherical harmonic components in each layer. Results are only shown for the density component, but the results for v_s and v_p are similar, except that, relative to the input model, the amplitudes of both input–output differences and uncertainties are smaller. The solid lines give the amplitude of the input model for this experiment (SPRD6), and the dashed lines indicate the amplitude of the input–output difference. The latter is reduced as we increase the number of data points, but remains roughly commensurate with the input amplitudes. Most notably, the input–output differences for data set 1 are often greater than the amplitude of the SPRD6 density variations, despite the fact that this data set is essentially the same as that used to construct that model. These differences are approximately the same as the ranges of good-fitting density models observed by Resovsky & Ritzwoller (1999b) and are a strong indication that the density component of SPRD6 is not robust.

The darkly shaded region indicates the level of uncertainty associated with the LS inversion covariances. These uncertainties are approximately the same for the two data sets, and are usually smaller than the input–output difference. There are two kinds of discrepancy evident in these uncertainties. The first, and most important, is that the uncertainties are smaller than the input–output differences in most layers. The second discrepancy is that data set 2 produces greater uncertainty than data set 1 (except in the top layer), despite having more data with better theoretical resolution (as predicted using the methods of Backus & Gilbert (1968)). This discrepancy is a consequence of the inconsistent measurements in data set 1, which are imitated in our synthetic data. These make it necessary to use a stronger damping that reduces variances while increasing the input–output error relative to data set 2. The first discrepancy is evidence that the most straightforward assessment of damped inversions underestimates modelling error. The second discrepancy shows that even if these error estimates are scaled upward (using 2σ rather than 1σ error bars), they are not reliable estimates of the true resolution of the data or, by extension, of the real uncertainty in the models.

Our synthetic experiments with damped inversions show that the inversion covariances and associated resolution tests are inadequate means of determining model robustness—whether they demonstrate the model to be robust (e.g. Ishii & Tromp 2001) or not (e.g. Kuo & Romanowicz 2002). Variances (and covariances) that underestimate model error, such as those we have observed, are evidence of a significant null space in the inverse problem. This is why we turn to the alternative of making undistorted 'maps' of model spaces that include the null space in describing the fit of models to the data.

3 NEIGHBOURHOOD ALGORITHM EXPLORATIONS

The neighbourhood algorithm, as described in two papers (Sambridge 1999a,b), is a tool for the efficient sampling of a model space. When properly tuned, it simultaneously identifies the model parameters most likely to generate a given data set and finds overall 'maps' of model likelihood in a relatively broad section of model space. The first step in the NA procedure (NA search) explores 'neighbourhoods' of the models that best fit the data, iteratively finding better models and defining new neighbourhoods. NA search is an extension of the 'Gibbs sampler' or 'heat bath' approach to

searching a model space for good models. NA search 'maps' regions of both good and poor fitting models while preferentially exploring the neighbourhoods of good-fitting models.

The second step in the procedure (NA Bayesian) uses a different implementation of Gibbs random walks. Here, the NA search 'maps' are resampled in a way that produces a 'likelihood sampling', where the density of samples near a particular model is inversely proportional to the misfit of that model to the data. The likelihood sampling yields rapidly converging numerical integrals of quantities such as average parameter values, variances and covariances. The resampling algorithm uses relative, rather than absolute, model likelihoods. Thus, integrals are computed without manipulating the very small and numerically unstable probability densities characteristic of highly dimensioned model spaces.

NA search and NA Bayesian can be tuned to a particular inverse problem through adjustments to a few control parameters. The tunability of the NA search algorithm makes it easy to adjust the exploration phase of the algorithm to ensure that searches are neither too narrow nor too broad. Searches that are too narrow can become trapped in relatively unimportant local minima in the misfit function (maxima of the likelihood function) or may accumulate many points near the top of the most-probable peak that slow the NA Bayesian resampling without providing useful information concerning the likelihood distribution. Searches that are too broad fail to identify the likely regions, or do so with so many points that the NA Bayesian converges very slowly.

Searches are broadened by increasing the number (N_r) of best-fitting neighbourhoods in which to test new models, and by decreasing the ratio of N_r to the number of new models (N_s) created in each iteration. For a small number of iterations (N_i) the search is usually too broad, while a very high number of iterations will produce too many points to allow NA Bayesian to run efficiently. When there are significant data misfit minima within the model space being searched, it is not difficult to identify appropriate settings of N_r, N_s and N_i. Such settings produce a sequence of models and their data misfits that progresses from a uniform sampling towards a likelihood sampling. Similarly, the speed of convergence of the Bayesian step is controlled by another two parameters: the number of steps in the random walks (N_w), and the number of different best-fitting cells from which to start the walks (N_c). The NA Bayesian code outputs convergence tests that allow the user to identify appropriate settings of N_w and N_c. The total number of resampled points should be within 50 per cent of the number of points in the original sampling, as long as the NA search produces a reasonable approximation of a likelihood sampling.

There is no guarantee that it will be possible to find control parameters that permit a convergent sampling within a practical time frame and produce results that are insensitive to small changes in these controls. However, in the present examples we are confident that practical and reliable settings have been found. They vary slightly from component to component and data set to data set, but are approximately $N_s = 15$, $N_r = 15$, $N_i = 9000$, $N_\mathrm{w} = 10\,000$ and $N_\mathrm{c} = 7$. For each of 27 spherical harmonic components, approximately 135 000 models in 21-dimensional space (seven layers of perturbations to v_s, v_p and ρ) are created by NA search, and an additional 70 000 by NA Bayesian. The time required for each component was approximately 80 h of time on a single 400 MHz Sun UltraSPARC-II processor. We have been able to implement parallelized versions of the NA codes. With these, 'mapping' of each component requires just over 12 h on eight processors of a Sunfire cluster of 750 MHz UltraSPARC-III processors. We do not anticipate that the computational intensity of this approach will be a problem. This is partly because the long-wavelength nature of these data and models allows us to restrict ourselves to relatively low-dimensioned model spaces, and partly because both NA search and NA Bayesian are extremely well-suited to parallelization and implementation on either supercomputers or one of the increasingly common linked networks of workstations.

There is one additional set of tuning parameters needed to implement the NA technique: the ranges of the model values to be tested. If the ranges are too large, a prohibitively large number of models may need to be sampled before the best-fitting regions are identified. If ranges are too small, important good-fitting models may be excluded *a priori*. In our synthetic experiments, we choose ranges nearly the same as the range of recently published long-wavelength models for v_s; about 150 per cent of the range of recent models for v_p; and about 250 per cent of the range of recent models for ρ. These ranges combine the observed variability among existing models with an approximation of the different effects of damping on v_s, v_p and ρ, given the relative scarcity of constraints on the latter two quantities. We have confirmed that if these ranges are expanded, 'mapping' results with either real or synthetic data remaining consistent as long as equivalent sampling densities are used. Conversely, results for smaller ranges are not consistent for either real or synthetic data.

Examples of NA search and NA Bayesian outputs are given in Fig. 2. It is important to note that the shape of the 2D Bayesian likelihood function (Fig. 2b) does not exactly match the distribution of the sampling (Fig. 2a), because initially lightly and densely sampled regions of equivalent fit are given equal weight by the resampling. The 2D marginals correspond to detailed images of the tradeoffs between parameters, which are more commonly represented by scalar measures such as the elements of resolution or covariance matrices. The 1D marginals (Figs 2c and d) yield Bayesian mean value and standard deviation estimates for model parameters. Our standard deviation estimates are given by the half-range of the x-axis multiplied by the fraction of the x-axis corresponding to likelihoods within a factor of $\mathrm{e}^{-1/2}$ of the maximum likelihood. For a Gaussian distribution, this is identical to the usual σ, but this estimate can also be applied to non-Gaussian distributions such as that of Fig. 2(d). The rms of such error estimates for density are shown by the lightly shaded regions of Fig. 1. These uncertainties from NA are larger than either the uncertainties or the inaccuracies of the damped inversions. It should be noted that, in order to show fine details of the distributions, Fig. 2 displays only the inner 30 per cent of the parameter ranges explored by NA. The full ranges are wide enough to include all significant likelihood regions for most of our parameters.

4 TESTS OF NA-MAPPING RESULTS

We have tested the reliability of our model space maps in three ways: (1) consistency with well-resolved results from damped inversions; (2) consistency with independent resolution tests; and (3) consistency with discrepancies in synthetic inversions.

4.1 Consistency with damped inversions

If NA 'mapping' correctly identifies well-resolved model characteristics, then these characteristics should also be identified by damped inversions. This is because, by definition, the null space has little impact upon the apparent likelihood of well-resolved characteristics of the models. Such characteristics include both individual model

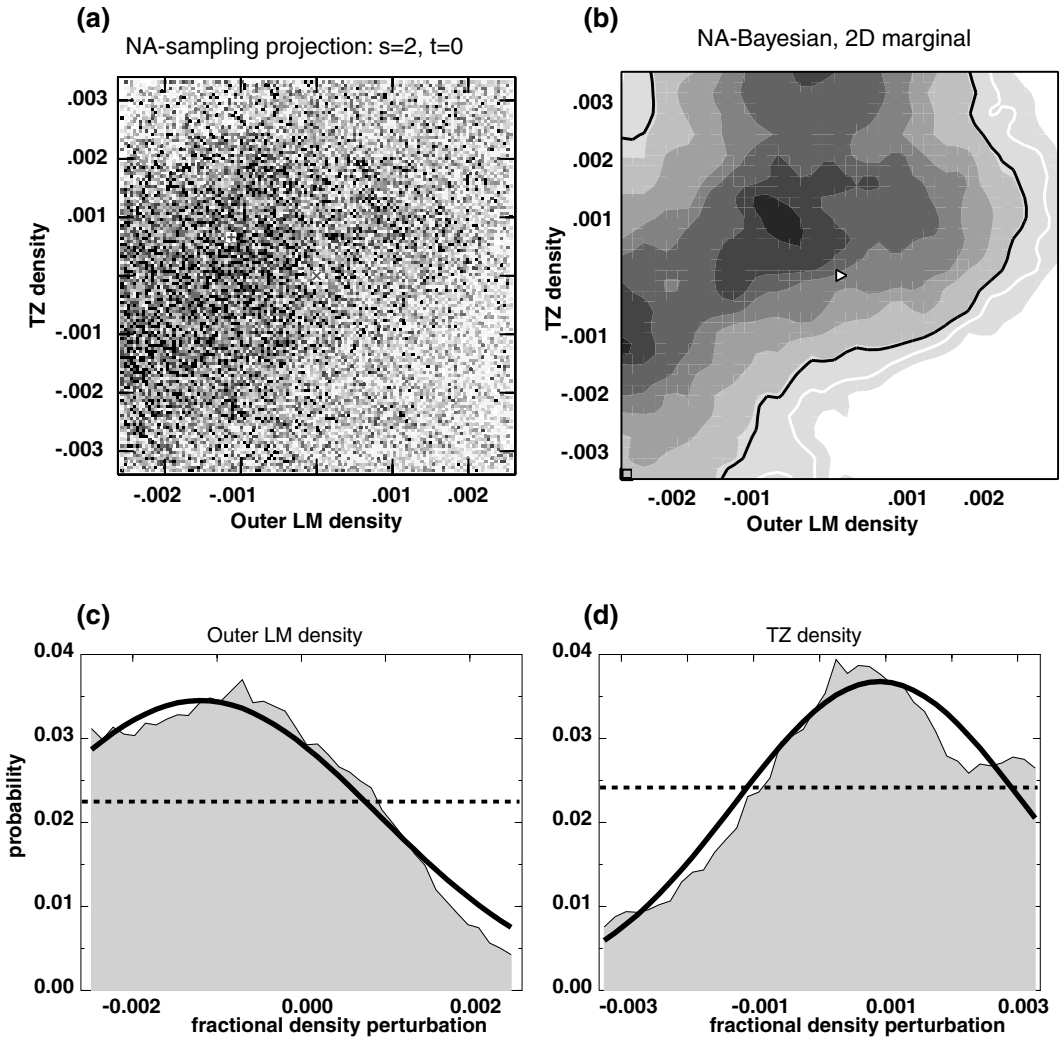

Figure 2. Examples of output for degree-two zonal density heterogeneity in layers 4 and 5 (outermost lower mantle and transition zone), from the NA-sampling and NA Bayesian algorithms applied to data set 2. Only the central 30 per cent of each parameter range is shown, so that it is easier to see patterns within the most-likely region. (a) A projection of a sampling of a 21-dimensional model space on to two dimensions. The density of dots represents the density of sampling, while darker dots indicate models of better fit. The more densely sampled regions are also those of better fit models. (b) the corresponding NA Bayesian 2D marginal, with greater likelihood shown by darker colors and a black line marking the 'confidence' contour that encloses 90 per cent of the total likelihood. (c) and (d) show shaded regions that are the corresponding 1D marginals, together with solid line that indicate Gaussian fits and dashed lines that mark the $e^{-1/2}$ levels. The non-Gaussian shape and wider-than-Gaussian $e^{-1/2}$ level of the TZ marginal are evident.

parameters that are well-resolved and strong trade-offs between pairs of parameters. We have tested NA in this way using the real data of data set 1 for both damped LS inversions and NA mapping. For these experiments, the parameter set was reduced to three layers of velocity perturbations (one for the whole upper mantle and two for the whole lower mantle; $d \ln v_p$ scales with $d \ln v_s$) and four layers (one UM; three LM) for d ln ρ, and SKS12WM13 is used as a starting model. The model parametrization is far simpler than the data require, but this improves our chances of finding the well-resolved parameters and tradeoffs that we need for this test. The range of each parameter is also greatly restricted for our NA 'mapping', so that the model space distortions of damped inversions are not too severe over a damping range sufficient to explore a comparable range of models. At weaker damping levels and greater model ranges, the inversions become too numerically unstable to be meaningful.

When we observe a narrow peak near the centre (starting model) of a 1D marginal from NA 'mapping', as in Fig. 3(a), damped LS inversions of the same inverse problem should yield values for that parameter that are relatively independent of damping and have a value near the marginal peak. This behaviour is apparent in the damped LS results in Fig. 3(a), for nearly an order of magnitude range of damping. When we observe a 2D marginal with a sharp peak in the direction of a corner, which implies a strong tradeoff in the corresponding pair of parameters, the results of LS inversions with decreasing damping should produce a sequence of points moving outward toward the most likely solution. Fig. 3(b) provides an example of this behaviour. The damped LS solutions also progress along a visible 'ridge' in the likelihood contours.

4.2 Consistency with Backus–Gilbert resolution

Backus & Gilbert (1968) introduced resolution filters as a tool for finding optimal inversion damping. Such filters are constructed by summing the data kernels, the functions that describe the sensitivity of each measurement to each model parameter. The filter for each

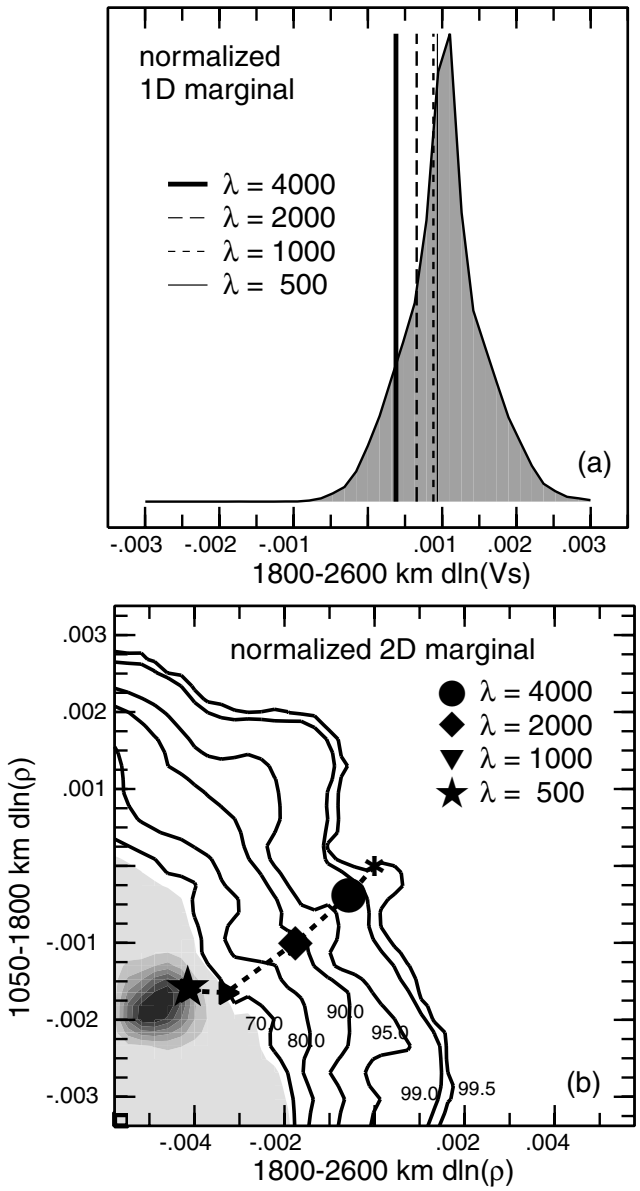

Figure 3. Comparison of results for NA 'mapping' and damped least-squared inversions applied to data set 1 for degree-two zonal structure. This experiment employs real data rather than synthetics, in an inversion for three velocity and four density parameters. (a) The NA Bayesian 1D marginal likelihood for deep-mantle v_s, which corresponds to a diagonal element of the inversion covariance matrix, together with values obtained from least-squared inversions with four damping values. Larger λ implies stronger damping. A wide range of damping values all give results within the high-likelihood region, and weaker damping moves the result closer to the likelihood peak. (b) An NA Bayesian 2D marginal likelihood for two density parameters, which corresponds to an off-diagonal element of the covariance matrix, again shown with results from several damped inversions. Increasing likelihood is shown with darker shading. The lower likelihood contours are shown as confidence levels (confidence increases with the total enclosed likelihood). Weaker damping moves the inversion results from the starting model (the asterisk at the centre) along a 'ridge' toward the likelihood peak.

parameter is the weighted kernel sum that most nearly produces a sensitivity to only that parameter. Backus & Gilbert (1970) showed that model resolution in damped inversions is a function of the difference between a pure filter and the best filter achievable with the given data kernels. This analysis cannot estimate absolute resolution, which is a strong function of the null space and of the form of regularization employed, but it does serve to estimate the relative resolution of two data sets. In other words, we expect that when new data are added, the improvement in the optimal filter for a given model parameter will be matched by the improvement in resolution (or reduction in error) for that parameter.

As noted in Section 2.2, this is not the case for the model error estimates from our damped inversions. In particular, Backus–Gilbert analysis predicts that the introduction of many constraints from surface wave measurements included in data set 2 should substantially reduce the error in upper-mantle density, and even produce significant error reductions for lower-mantle density. No such differences are evident in the uncertainties from damped LS covariances shown in Fig. 1. In contrast, the errors from NA 'mapping' assessments exhibit a pattern of improvement for data set 2 that agrees with the Backus–Gilbert prediction. In general, we have found that for a variety of different data sets the error estimates from NA 'mapping' match both the size and pattern of predicted improvements in resolution.

4.3 Accounting for observed discrepancies

A damped LS inversion distorts the model space to choose as an output model a particular low-misfit solution from the undistorted model space. Therefore, in our synthetic inversions, both the input model and the output model should be within the low-misfit, high-likelihood region of model space. In turn, this implies that if NA-mapping error estimates are good estimates of the size of low-misfit regions, these errors usually will be larger than input–output differences. In other words, the NA-mapping results should account for most of the discrepancies observed in the synthetic experiments.

The lightly shaded regions of Fig. 1 show that this is true of our NA 'mapping' uncertainties for ρ models at degrees two and four. The rms size of the NA Bayesian uncertainties is everywhere larger than the rms size of input–output differences for either data set. The same is true of the v_s and v_p models and for individual spherical harmonic components.

The NA 'mapping' outputs are also self-consistent, in that uncertainties are everywhere larger than the differences between the input model and either the most likely or mean models output by NA Bayesian. These input–output differences from NA 'mapping' are generally 50–100 per cent larger than the input–output differences from the damped LS inversions. This, however, should not be considered a failure of the NA technique, because the 'maps' explore a range of models more than twice as large as the space between the input and starting models of the damped inversions. The latter range is imposed by the damping needed to produce stable inversion results, while the former is large enough to measure the true resolution of the data. We have also performed NA mapping restricted to parameter ranges that match those of the damped inversions. When we do so, the NA 'mapping' uncertainties become less reliable, but both mean and most likely models yield input–output differences smaller than those of the damped inversion results.

Thus, we have now shown that the results of NA mapping are consistent with damped LS results where such consistency is expected, are consistent with an independent test of resolution, are self-consistent, and account for observed discrepancies in synthetic experiments. These observations provide us with confidence that NA mapping is a reliable tool for assessing long-wavelength mantle models.

5 APPLICATION: THE TRUE RESOLUTION OF SEISMIC DENSITY MODELS

The rms uncertainties of Fig. 1 not only test the NA-mapping technique, but also provide estimates of how well seismic data resolve long-wavelength mantle density heterogeneity. Because our synthetic data realistically imitate the error in real data, the error bars for our synthetic experiments approximate the uncertainty in models that use real long-period data to image long-wavelength mantle heterogeneity. These uncertainty levels correspond to the smallest amplitude of structural features for which robust images can be obtained with each data set and parametrization. In addition, model covariances from the 2D marginals of NA Bayesian let us see which model parametrizations, if any, are likely to yield meaningful constraints on mantle density. Strong covariances imply tradeoffs between parameter pairs and may imply that better parametrizations exist. Tradeoffs between adjacent density layers, for instance, would suggest that robust models require wider layers.

The NA 'mapping' error bars of Fig. 1(a) show that density structure with amplitudes of input model SPRD6 are not robust with respect to the resolution of data set 1. Because this data set and our radial parametrization are quite similar to those used to construct SPRD6, we conclude that the model is not robust. This confirms the implications of several experiments with smaller data sets and various parametrizations (Resovsky & Ritzwoller 1999b; Romanowicz 2001; Kuo & Romanowicz 2002).

The results for the NA model covariances from 2D marginals, demonstrated in Fig. 4, are as significant as the preceding results from the 1D marginals. There are large off-diagonal elements spread throughout the density columns of the data set 1 covariance matrix (Fig. 4a). Because there are tradeoffs among density parameters at all depths, even parametrizations with broader layers or long-wavelength smooth functions of radius cannot robustly locate even high-amplitude density heterogeneity with data set 1. This is similar to what Resovsky & Ritzwoller (1999b) concluded, using a smaller data set. There are also strong density–v_s and density–v_p tradeoffs that indicate that density heterogeneity is mostly indistinguishable from velocity heterogeneity, as observed by Kuo & Romanowicz (2002). Since robust observations of v_s heterogeneity can be made, it is possible to use the amplitude of the density error bars to infer upper limits to $d \ln v_s / d \ln \rho$ (e.g. Romanowicz 2001). The NA 'mapping' covariances show, however, that subsets of data set 1 cannot give much meaning to these limits because these data cannot place correlated velocity and density variations at any particular depth.

Fig. 1(b) shows that even data set 2 would fail to resolve density variations of SPRD6 amplitudes, and we have found that such amplitudes remain below uncertainty levels when we approximate the addition of the resolving power of the recent Widmer-Schnidrig (2002) catalogue. However, our NA 'mapping' with data set 2 also provides several reasons for optimism concerning the resolution of density heterogeneity with long-period data. First, the covariance matrices for data set 2 (Fig. 4b) show that the additional data greatly reduce tradeoffs among our v_s, v_p and ρ parameters. It is particularly important that the addition of purely upper-mantle constraints from the long-period surface wave data appears to reduce tradeoffs between upper-mantle heterogeneity and the lower-mantle density variations that are of greatest interest. The greatly reduced covariances in the density columns imply that any density heterogeneity large enough to be observed will be distinguishable from velocity variations and can be located at specific depths, with a radial resolution at least as good as that implied by our layer widths. In turn, having density heterogeneity (or just upper limits, where it is too small to be observed) placed at specific depths implies that data set 2 could be used to assess the likelihood and strength of velocity–density correlation as a function of depth.

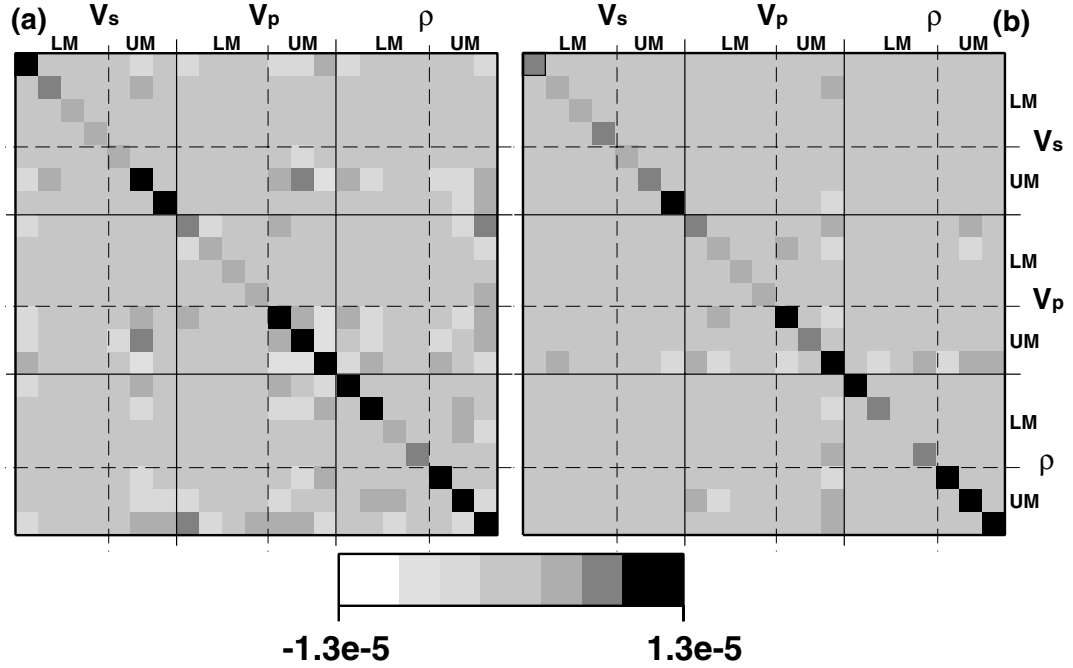

Figure 4. Covariance matrices from NA 'mapping' the degree-two zonal components of data sets 1 (a) and 2 (b). Covariances are derived as in Sambridge (1999b). The scale maximum is half the amplitude of the largest diagonal matrix element for data set 1. Most of the important tradeoffs for data set 1, implied by the visible off-diagonal elements of (a), are reduced by a factor of 2 or more in (b).

A second source of optimism comes from the fact that the poor density resolution evident in Figs 1(a) and 4(a) implies that least-squares inversions with data set 1 require overdamping of the density component to produce stable solutions. Thus the density component of SPRD6 is likely to have artificially suppressed amplitudes. Now that NA has proven to be an effective approach, a far less restricted sampling of the density models can be explored, and there is reason to believe that 'good' models with much greater amplitudes will be found. Poor correlations between v_s and ρ at degrees two and four in SPRD6 in much of the lower mantle could be evidence for strong chemical heterogeneity. Such heterogeneity might produce density perturbations that are too large to be explored by damped inversions. Inspection of SPRD6 also reveals that degree-two density heterogeneity amplitudes above the error levels in Fig. 1(b), if accompanied by lower degree-four amplitudes, could even be consistent with the predominance of thermal heterogeneity, and typical values of $d \ln v_s / d \ln \rho$. Finally, we note that both the long-period data set and the inversion parametrization can be improved. The recent catalogue of Widmer-Schnidrig (2002) includes several hundred new normal-mode splitting functions. These should reduce the uncertainty bounds of Fig. 1(b) by another 10–20 per cent, and produce more diagonal covariance matrices than those of Fig. 4. The number of well-resolved parameters can be improved by finding the optimal parametrization for a given data set. These can be constructed from a simple initial parametrization (such as layers) by examining tradeoffs using a technique such as NA (e.g. Douma *et al.* 1996).

Our conclusions, therefore, are as follows: (1) NA is a practical and reliable tool for exploring and assessing long-wavelength mantle models using long-period seismic data; (2) NA provides the first meaningful error bars for seismic models of density heterogeneity, which are necessary if seismic density models are to become more useful to the mineral physics and geodynamics communities; (3) the error bars and covariances associated with typical splitting function data sets are much poorer than suggested by damped inversions, and previous seismic models of long-wavelength 3D density variations in the mantle cannot be robust; (4) by using the most recent long-period surface wave and normal-mode data to reduce covariances and using the NA approach to avoid the limitations of damped inversions, it is now possible, for the first time, to provide robust seismic constraints on density heterogeneity and the likelihood of correlation between density and velocity variations. Our own applications of NA to the degree-two zonal component of real splitting function data have already retrieved robust lower-mantle density variations. We expect a more complete model to emerge when the parametrization is optimized and the latest data are added.

ACKNOWLEDGMENTS

Most of the computations for this research were performed on the computers of the Edinburgh Parallel Computer Centre, access to which was provided through the TRACS visitor programme of the European Commission's Access to Research Infrastructures initiative. We are very grateful to Malcolm Sambridge of the Australian National University for making his algorithms and visualization tools readily available. We are similarly indebted to Miaki Ishii for making the details of SPRD6 available via the internet, to Guy Masters and Gabi Laske for their compilation of normal-mode data catalogues, and to Hendrik van Heijst for his surface wave data. This research has been supported by the Netherlands NWO, Earth sciences grant number 809_31.003.

REFERENCES

Backus, G. & Gilbert, F., 1968. The resolving power of gross Earth data, *Geophys. J. R. astr. Soc.,* **16,** 169–205.

Backus, G. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data, *Phil. Trans. R. astr. Soc.,* A, **266,** 123–192.

Beghein, C., Resovsky, J.S. & Trampert, J., 2002. *P* and *S* tomography with a neighbourhood algorithm, *Geophys. J. Int.,* **149,** 646–558.

Douma, H., Snieder, R. & Lomax, A., 1996. Ensemble inference in tems of empirical orthogonal functions, *Geophys. J. Int.,* **127,** 363–378.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary Reference Earth Model, *Phys. Earth planet. Inter.,* **25,** 297–356.

He, X. & Tromp, J., 1996. Normal-mode constraints on the structure of the Earth, *J. geophys. Res.,* **101,** 20 053–20 082.

Ishii, M. & Tromp, J., 1999. Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle, *Science,* **285,** 1231–1236.

Ishii, M. & Tromp, J., 2001. Even-degree lateral variations in the Earth's mantle constrained by free oscillations and the free-air gravity anomaly, *Geophys. J. Int.,* **145,** 77–96.

Kuo, C. & Romanowicz, B., 2002. On the resolution of density anomalies in the Earth's mantle using spectral fitting of normal mode data, *Geophys. J. Int.,* **150,** 162–179.

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

Liu, X.-F. & Dziewonski, A.M., 1994. Lowermost mantle shear wave velocity structure, *EOS, Trans. Am. geophys. Un.,* **77,** F663.

Masters, G., Laske, G. & Gilbert, F., 2000. Matrix autoregressive analysis of free-oscillation coupling and splitting, *Geophys. J. Int.,* **143,** 478–489.

Pestana, R., 2001. New normal modes observations. Presented at *The Deep Earth: Theory, Experiment and Observation; Mantle Processes.* Espinho, Portugal, 15–20 Sept., 2001, EURESCO.

Resovsky, J.S. & Pestana, R., 2002. Improved lower mantle V_p constraints from spectral fitting of normal mode data, submitted, *Geophys. Res. Lett.*

Resovsky, J.S. & Ritzwoller, M.H., 1998. New and refined constraints on three-dimensional Earth structure from normal modes below 3 mHz, *J. geophys. Res.,* **103,** 783–810.

Resovsky, J.S. & Ritzwoller, M.H., 1999a. A degree 8 mantle shear velocity model from normal mode observations below 3 mHz, *J. geophys. Res.,* **104,** 100–110.

Resovsky, J.S. & Ritzwoller, M.H., 1999b. Regularization uncertainty in density models estimated from normal mode data, *Geophys. Res. Lett.,* **26,** 2319–2322.

Ritsema, J. & van Heijst, H., 2000. Seismic imaging of structural heterogeneity in Earth's mantle: evidence for large-scale mantle flow, *Science Progr.,* **83,** 243–259.

Romanowicz, B., 2001. Can we resolve 3D density heterogeneity in the lower mantle, *Geophys. Res. Lett.,* **28,** 1107–1110.

Sambridge, M., 1999a. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.,* **138,** 479–494.

Sambridge, M., 1999b. Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble, *Geophys. J. Int.,* **138,** 727–746.

Tarantola, A., 1987. Inverse problem theory, methods for data fitting and model parameter estimation, Amsterdam, Elsevier.

Trampert, J. & Woodhouse, J., 2001. Assessment of global phase velocity models, *Geophys. J. Int.,* **144,** 165–174.

Tromp, J. & Zanzerkia, E., 1995. Toroidal splitting observations from the great 1994 Bolivia and Kuril Islands earthquakes, *Geophys. Res. Lett.,* **22,** 2297–3000.

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

Widmer-Schnidrig, R., 2002. Application of regionalized multiplet stripping to retrieval of aspherical structure constraints, *Geophys. J. Int.,* **148,** 201–213.

Wong, Y.K., 1989. Upper mantle heterogeneity from phase and amplitude data of mantle waves, *PhD thesis,* Harvard University, Cambridge.