Available online at www.sciencedirect.com

SCIENCE DIRECT®

EPSL

Earth and Planetary Science Letters 215 (2003) 121–134

www.elsevier.com/locate/epsl

Using probabilistic seismic tomography to test mantle velocity–density relationships

Joseph Resovsky *, Jeannot Trampert

Department of Earth Sciences, Utrecht University, P.O. Box 80021, 3508 TA Utrecht, The Netherlands

Received 9 April 2003; received in revised form 16 July 2003; accepted 18 July 2003

Abstract

We use a neighborhood algorithm to explore the fit to long period seismic data of a wide variety of long wavelength mantle models. This approach to the global tomographic inverse problem yields probability distributions for seismic velocities, density, and related properties as functions of depth. Such distributions can be robust even when individual models are not, and allow us to test several assumptions about the Earth that have long been enforced a priori in inversions. In particular, we are able to test the paradigm of deep mantle heterogeneity that is dominantly thermal in origin, producing velocity and density perturbations that are well correlated and have relative amplitudes given by $\delta\ln\rho/\delta\ln v_s < 0.5$. Our distributions show that such relationships are unlikely, and even though the results are consistent with recent best fitting models from damped seismic inversions, they demonstrate that many specific properties of such models are not robust. The data clearly favor density perturbations that are poorly or negatively correlated with velocity heterogeneity and have amplitudes several times larger (yielding $\delta\ln\rho/\delta\ln v_s > 1.0$) than damped inversions allow. These characteristics are most pronounced in the upper mantle transition zone and the base of the lower mantle, suggesting layered convection. The negative density–velocity correlations favored at these depths imply dominantly chemical heterogeneity, while the likelihood of relatively high amplitude density variations suggests that variable iron content is an important component of this heterogeneity. These results, which we show to be consistent with independent gravity constraints, represent a profound change in the interpretation of seismic constraints. In addition, the distributions show that even though best fitting density models from recent inversions or our sampling are consistent with the data, most specific properties of such models are not robust. This implies that it is more appropriate to use seismic model distributions, rather than individual models, to make geodynamic and geochemical inferences.

Keywords: seismic tomography; mantle density; neighborhood algorithm; chemical heterogeneity

1. Introduction

Seismic images of density variations in the deep Earth can provide crucial constraints on the composition and dynamics of the hidden interior of our planet. Unfortunately, seismic data are dominantly sensitive to seismic velocities and much

* Corresponding author. Tel.: +31-30-253-5138; Fax: +31-30-253-3486.

E-mail addresses: resovsky@geo.uu.nl (J. Resovsky), jeannot@geo.uu.nl (J. Trampert).

0012-821X/03/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0012-821X(03)00436-9

more weakly sensitive to density. This means that in traditional inversions of measurements and sensitivities of Earth models, the density portion of the models must be strongly damped (relative to the velocity components) in order to retrieve stable results. This damping is a form of a priori regularizing information that can almost completely conceal any real information the data contain about the most poorly resolved parameters [1].

In fact, the global seismic inverse problem is mathematically ill-posed. In addition to the null space of data errors and imperfect model parameterization resulting in data that cannot be fit, the data resolution is poor enough that there is also a null space of combinations of model parameters that are effectively invisible to the data. Because of these two null spaces, the functions describing model misfit to the data can have multiple minima, even for a linear inverse problem [2]. This is why inversions (or other searches) for best fitting models must be stabilized by imposing regularizing assumptions.

Stable damped inversions for density generally produce models in which lateral density perturbations relative to the spherical average ($\delta\ln\rho$) are less than half the size of relative shear velocity perturbations ($\delta\ln v_s$). This relative amplitude ratio of $\delta\ln\rho/\delta\ln v_s \approx 0.4$ corresponds closely to the relative resolution of the ρ and v_s parameters [3]. Inversions are often further stabilized by assuming that density and velocity heterogeneity are perfectly correlated and scale with the above ratio [4].

The use of such inversion constraints has proven to be reasonably consistent with the paradigm of mantle models in which both heterogeneity and dynamics are dominantly thermal in nature. Under this set of assumptions, lateral variations of seismic velocity and density are directly caused by lateral variations in temperature maintained by a convective process. This implies, according to high temperature and pressure extrapolations of the behavior of near surface minerals dating back more than 30 years [5], that deep mantle ρ and v_s are perfectly correlated, with $\delta\ln\rho/\delta\ln v_s \approx 0.4$. This ratio matches that imposed by the most natural regularization of seismic inversions, guaranteeing that models from those inversions appear to validate the assumption of thermal heterogeneity, although anelastic effects must be incorporated for full consistency [6]. In addition, geodynamic modeling dominated by thermal convection has proven to be consistent with the same ratio, in the sense that particular viscosity models can be combined with particular examples of correlated velocity and density models to explain surface geodetic and tectonic observables [7,8].

On the other hand, a variety of geophysical studies, old and new, suggest alternatives to the paradigm of dominantly thermal heterogeneity in the deep mantle. Even half a century ago it was proposed that pressure effects serve to decrease the importance of temperature variations in the deep mantle, and that iron oxides could produce strong chemical heterogeneity [9], and these ideas were developed further in recent decades [10]. Still more recent results from mineral physics and geochemistry seem to confirm these inferences. It is now clear that even small variations in iron content can produce large density variations in the deep mantle without significantly perturbing seismic velocities [6]. Geodynamical modeling with deep mantle chemical heterogeneity has shown that such structure can also be consistent with surface observables [11–13], and preliminary thermochemical convection models indicate that large density variations can develop and remain stable over geological times [14–16]. Finally, several recent seismic inversions have produced joint models of shear and bulk sound velocities that are best explained by including significant chemical heterogeneity in the deep mantle [17–19], although inconsistencies among such models make this evidence somewhat ambiguous.

Recent order-of-magnitude improvements in the quantity and quality of seismic data, together with the emergence of new computational tools, now make it possible for seismologists to test the perfect correlations and low $\delta\ln\rho/\delta\ln v_s$ scaling associated with the thermal heterogeneity paradigm by looking at seismic constraints on density without a priori damping which enforces that paradigm. The past several years have seen the publication of extensive normal mode (NM) catalogues measured using spectral fitting [20] and generalized spectral fitting [21]. These data have diverse sensitivities to long wavelength velocity and density heterogeneity throughout the mantle, and have been applied to new attempts to produce density models. The best known of these is model SPRD6 [22], in which the assumption of correlated velocity and density was removed, and density was found to be partly decorrelated from the pattern of velocity heterogeneity, particularly in the deepest mantle. However, this was still a damped inversion, and both earlier [23] and later [24] studies with similar data showed that the results were not robust with respect to choices of damping.

Probabilistic tomography is the necessary extension of these studies, in that it is an approach that not only does away with damping that biases 'best fit' solutions, but also describes important details of the complex model space associated with an ill-posed problem, including the model null space. It does so by creating model space 'maps' that display the fit of a wide range of models to the data [25]. These maps can display multiple minima, and contain much more information about seismic constraints on the mantle than when these constraints are expressed only as a single best fitting inversion model. In particular, the 'maps' show not only the kinds of models that best fit the data, but also which models yield poor fits and which model characteristics are invisible to the data. This information can be expressed by a variety of statistical measures, such as the distributions for correlations and ratios recently presented for $\delta \ln v_s$ and $\delta \ln v_p$ [26], and compared to mineral physics constraints [6,27]. The neighborhood algorithm (NA) [28,29], together with the emergence of new parallelized computing facilities, permits the efficient construction of these maps.

Synthetic experiments have already shown that this approach can be applied reliably to the use of long period data to obtain robust constraints on long wavelength density and velocity heterogeneity in the mantle. These experiments suggested that robust density constraints could be achieved if the NM data are combined with the latest surface wave (SW) data, because the latter help break tradeoffs in NM sensitivity to the upper mantle and lower mantle structures [30]. We now employ real data to do just this, providing a new and more robust test of consistency between seismic data and the paradigm of a mantle dominated by the effects of thermal heterogeneity.

2. Data and methods

We have combined the NM data sets listed in Section 1 with two recent catalogues of SW fundamentals [31] and overtones [32]. In addition, we employ a small but important set of recently measured splitting functions for mantle v_p multiplets [33] recently measured with generalized spectral fitting, a new NM catalogue compiled using an autoregressive form of spectral fitting [34], and a large new collection of NM fundamental and overtone splitting functions retrieved using regionalized stripping [35]. The recent SW catalogues have been augmented, at very long periods, by older measurements [36]. No measurements sensitive to the inner core are included, and all measurements are corrected to remove the signal of the 3D crustal model CRUST5.1 [37], so that this is a purely mantle-sensitive data set. The NM catalogues are accompanied by uncertainty estimates. Errors for the SW data are assigned using comparisons of the results of several different studies [26]. The period range of NM catalogues is 100–2000 s and for the SW data it is 40–300 s.

All of these data are converted into the form of *structure coefficients*. These are the ${}_k c_s^t$ multipliers of the Y_s^t spherical harmonics in expansions of the form:

$$ {}_k\mathrm{F}(\theta,\phi) = \sum_{s,t} {}_k c_s^t \; Y_s^t(\theta,\phi) \tag{1} $$

The functions ${}_k\mathrm{F}$ are global patterns of surface wave velocities and free oscillations that emerge from combinations of many strong earthquakes recorded by widely distributed networks of broadband seismometers. The index k identifies the central frequency and/or eigenfunction of each measurement. Structure coefficient measurements are retrieved from the raw data using linearized regressions, and each is a linear functional of a single harmonic component of Earth structure:

$$_k c_s^t = \int_0^{r_E} \delta\ln\mathbf{m}_s^t(r)\cdot{}_k\mathbf{M}_s(r)r^2\mathrm{d}r + \sum_d \delta\ln h_{sd}^t\ {}_kH_{sd}\ r_d^s \quad (2)$$

Here, perturbations to a spherically symmetric Earth model are given by:

$$\delta\ln\mathbf{m}(r) = \frac{\delta\mathbf{m}(r)}{\mathbf{m}(r)} = [\delta\ln v_\mathrm{p}(r),\ \delta\ln v_\mathrm{s}(r),\ \delta\ln\rho(r)] \quad (3)$$

and by relative perturbations $\delta\ln h_d$ to the locations of radical discontinuities indexed with d, all expanded as in Eq. 1. These are related to the structure coefficients via radial kernels $\mathbf{M}(r)=[P(r),\ S(r),\ R(r)]$ and boundary factors H_d, which are derived from a perturbation theory approximation [38]. The approximate independence of the equations for individual harmonic components is a great advantage of structure coefficient data representation, and is achieved through careful regressions of geographically diverse seismograms [21].

Earth structure is modeled from NM and SW data using the inverse of Eq. 2. This is an approximate relation whose validity depends on $\delta\ln\mathbf{m} \ll 1$ and appropriate accounting for theoretical errors in the structure coefficient regressions. Removing crust and inner core effects helps to meet the perturbation requirement, while most of the recent NM and SW data catalogues have been constructed using methods that greatly reduce theoretical errors such as cross-coupling, source error, and truncation. The reduction of theoretical error also means that data errors become approximately Gaussian. This has been observed directly in the case of the generalized spectral fitting measurements, which are the most important data in our study and are accompanied by the most detailed error analysis [21]. Older, more approximate catalogues [20,36] are assigned greater measurement uncertainties ($_k\sigma_s^t$). These greater uncertainties serve to eliminate the modeling instabilities inherent in having inconsistent measurements in our combined data set.

In all, we have 678 splitting coefficient measurements for each of the five independent degree 2 structural components, and 646 for each of the nine degree 4 components. Our data, error tables, and kernels are available upon request. For the present we have not extended our analysis to higher degrees (shorter wavelengths) or odd degrees, where there are significantly fewer lower mantle-sensitive measurements. Because of the well established dominance of the degree 2 signal in the deep mantle [39], and the frequent discussion of very large-scale features like seismically observed 'superplumes' [17], even probabilistic tomography for only degrees 2 and 4 is a significant test of the thermal paradigm.

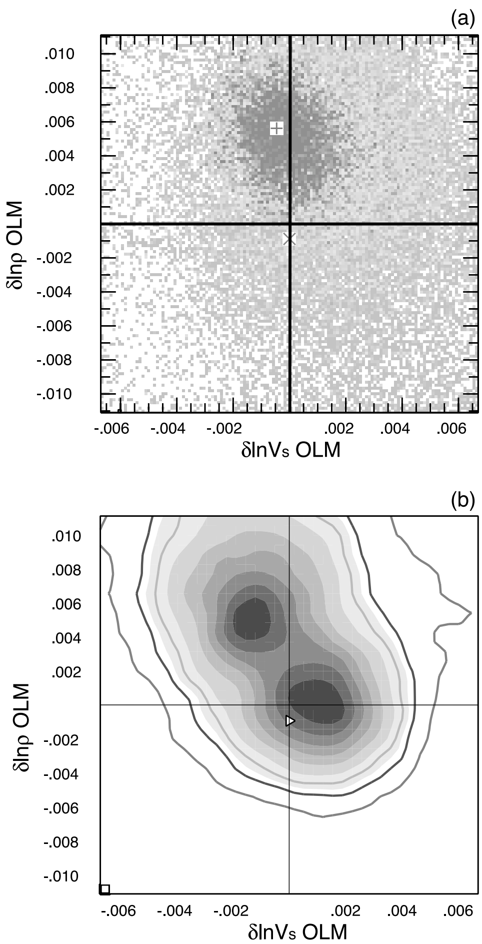
We construct our 3D Earth models using $\delta\ln v_\mathrm{s}$, $\delta\ln v_\mathrm{p}$, and $\delta\ln\rho$ perturbations relative to spherically symmetric preliminary reference Earth model PREM [40] in five layers: uppermost upper mantle (UUM, 24–400 km depth); transition zone (TZ, 400–670 km); outer lower mantle (OLM, 670–1200 km); middle lower mantle (MLM, 1200–2000 km); and lowermost lower mantle (LLM, 2000–2891 km). We also allow perturbations to the 670 km and core–mantle boundary (CMB) discontinuities of the spherical PREM model, for a total of 17 model parameters for each structural component. The effect of the parameter space boundaries is minimized by allowing each parameter to vary in a relatively wide range and adjusting the sampling density within those limits until stable results are achieved.

The number of layers is rather small, and was restricted by the computational requirements of searching more highly dimensioned model spaces more than by data resolution. Several tests [22,30] have shown that 7–12 radial layers can be resolved by the available data. Such underparameterization is not as great a problem for us as it is in performing inversions, in which the choice and effects of regularization and damping are usually strong functions of the parameterization. Model space searches without damping will produce the same average characteristics for a wide layer as the mean of the properties of any sublayers into which it is divided. Nonetheless, the use of wide layers will tend to average away interesting structure with shorter radial length scales.

In order to test the thermal paradigm as well as possible with the current data set and a single parameterization, we have chosen layers which we believe give us the best chance of detecting the signature of chemical heterogeneity in the mantle, without necessarily constraining its scale length or ruling out significant chemical heterogeneity of scale lengths much smaller than that of our parameterization. The UUM layer, in particular, is too wide to account for data sensitivity to relatively strong radial gradients of structure likely to exist above 400 km depth. However, the focus of any long period seismic study is on the sensitivity of these data to greater depths, and greater resolution of the UUM could prove misleading because it will tend to absorb any errors in our crustal corrections and because degree 2 and 4 structure has less relative importance at those depths than elsewhere.

The TZ is a major feature of the PREM model and has a width that corresponds roughly with the resolving power of our data, while the OLM and LLM approximate the widest regions proposed for lower mantle boundary layers [14,15,41,42]. If we are able to detect a contrast between the OLM or LLM and the presumably homogeneous MLM, this could be evidence of the presence of boundary layers. Failure to detect contrast will not rule out boundary layers, because there is some chance that our wide layers will average away interesting structure with shorter radial scale lengths. Much more computational power is now available than when this study was begun, and it is already feasible to use finer layering to refine the upper mantle model and constrain the width of the potential lower mantle boundary layers. This will be the subject of future work.

We do not believe the resolution and relevance of our model space 'maps' will be enhanced by changing to a parameterization with radially smooth basis functions. Although the details of layered models are less plausible than smooth models, we are not concerned here with exact models but with depth-averaged characteristics. The two types of parameterizations are different depth averages that yield the same gross characterizations of the mantle, as has been demonstrated by the consistency of layered [39] and smooth [22] long wavelength v_s models.



←

Fig. 1. (a) An example of an NA sampling of a model space. In this case the algorithm has sampled models that fit the subset of long period seismic data sensitive only to the degree 2 zonal spherical harmonic component of 3D mantle structure. The models consist of perturbations to this component of v_s, v_p, and density heterogeneity in five layers and topography on two radial discontinuities, for a total of 17 parameters. 100 000 models have been generated with a sampling density that increases with improving model fit to the data. Here, model fits (darker colors indicate better fits) are projected onto two dimensions of the model space: $\delta\ln v_s$ and $\delta\ln\rho$ in the OLM. The 'x' marks the prediction of model SPRD6, and the '+' marks the best fitting model. The axes give perturbations relative to the spherical PREM model. (b) An example of a 2D probability density marginal. The model space sampling of panel a has been resampled with NA so that sampling density exactly corresponds to the χ^2 misfit in all parts of model space. While the original sampling was concentrated on a single most likely model, the resampling reveals another broad region of good fit near the origin. This region includes the SPRD6 model prediction (white triangle). Note that the prediction has a slight positive correlation for these two parameters, while the best fitting model and the overall distribution each imply a negative correlation (or covariance).

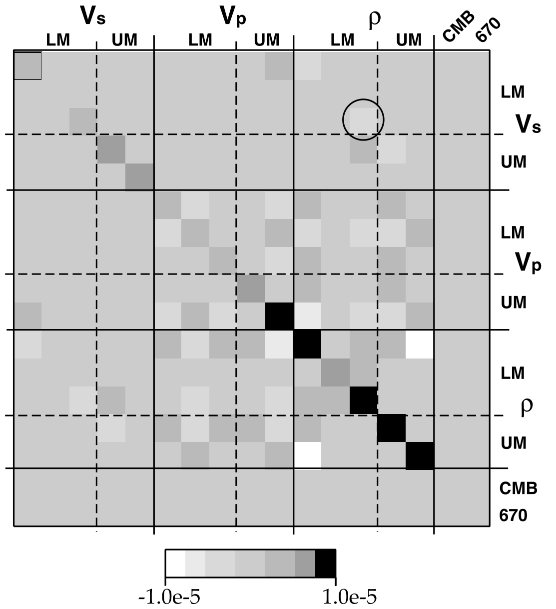


Fig. 2. The covariance matrix for the 17-dimensional model space map projected in Fig. 1b. Each element here is a scalar value describing the shape of a 2D marginal like Fig. 1b. More intense shades indicate greater ellipticity and more significant positive (dark) or negative (light) tilt. The circle indicates the element corresponding to Fig. 1b. The diagonal elements give the variance (squared standard deviation) of each parameter. Density parameters have the greatest uncertainty. The limited model search range allowed for the topography parameters (approximately the amplitude of SPRD6 topography) contributes to their small variances.

For each of the 14 degree 2 and 4 spherical harmonic components, the NA is first used to produce a set of approximately 100 000 models accompanied by a measure of the misfit of each to the corresponding harmonic component of the data. The misfit of a model **(m)** for N data **(d)** is defined by:

$$ {}^{\mathbf{m}}\chi_s^t = \left[\sum_{k=1}^{N} \frac{({}_k^{\mathbf{m}}c_s^t - {}_k^{\mathbf{d}}c_s^t)^2}{({}_k\sigma_s^t)^2} \right]^{1/2} \tag{4} $$

where ${}_k{}^{\mathbf{m}}c_s^t$ is calculated using a discretized version of Eq. 2. Fig. 1a shows an example of such a model sampling, projected from 17-dimensional space onto two dimensions. These sets are then NA resampled to produce smooth 1D probability distributions of each model parameter, selected 2D likelihood marginals (Fig. 1b), covariance matrices (Fig. 2), and collections of most likely parameter ranges (Fig. 3). All are evaluated using a posterior probability density (PPD) defined by PPD(**m**) = exp(${}^{\mathbf{m}}\chi^2/2$) [29]. This definition of the PPD makes it easy to confirm the ill-posed nature of our inverse problem because the approximately Gaussian nature of our data errors would produce Gaussian marginals if our inverse problem were well-posed.

We must then combine the 'maps' for individual spherical harmonic components of structure to find the probability distributions of overall model amplitudes and geometries. We do this by drawing random deviates from the relevant marginals. To compare, for instance, likely $\delta\ln v_s$ and

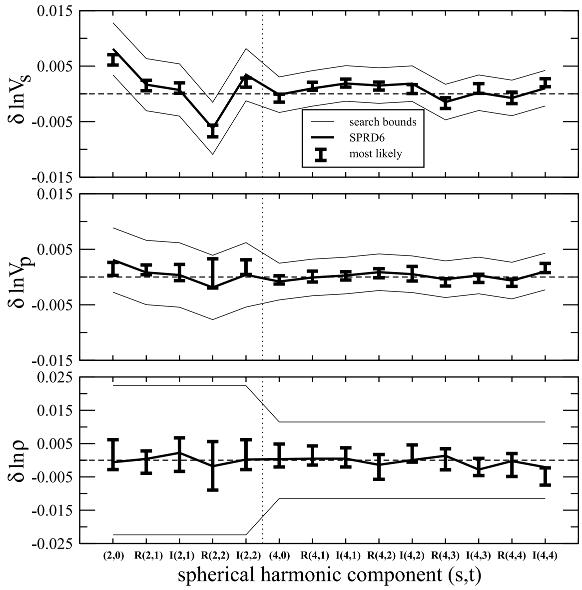
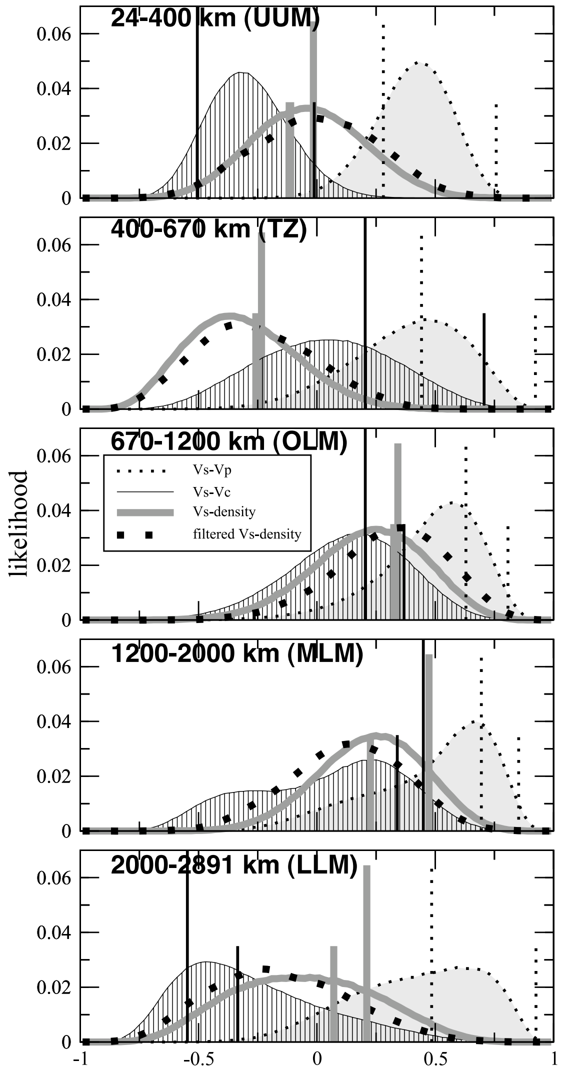


Fig. 3. Most likely $\delta\ln v_s$, $\delta\ln v_p$, $\delta\ln\rho$ density model parameters from NA model space maps in the LLM (below 2000 km depth). The thin solid lines show the parameter space search limits and thick lines give the predictions of model SPRD6. The error bars approximate the range of likely values for model parameters, and, though the distributions are generally asymmetric, are centered on the most likely value for convenience of display. Most likely models and error bars come from 1D marginal projections of 2D marginals like those in Fig. 1b. The error bars give the fraction of each 1D marginal with amplitudes within 1/e of the maximal likelihood. Individual most likely density coefficients are not often robustly different from SPRD6 or PREM. However, unlike the velocity coefficients, they usually have greater amplitudes than the model predictions.

$\delta\ln\rho$ models in the OLM at degree 2, we construct v_s and ρ models from a set of five random deviates, one drawn from each of the five OLM degree 2 $\delta\ln v_s$–$\delta\ln\rho$ 2D marginals (including that of Fig. 1b). We use a sequence of similar 2D random deviates to insure consistency with all 2D marginal projections of the model space. This is an approximation to consistency with the full model space, which, unfortunately, cannot be output by NA without a significant decrease in computational efficiency. The amplitudes and correlations of each set of random deviates are recorded. With enough repetitions of this process (10^5–10^7, depending on the specific application), stable probability distributions of correlations and amplitudes emerge. These are shown in Figs. 4 and 5 as distributions of likelihood, which is a dimensionless renormalization of probability.



We also combine $\delta\ln v_s$ and $\delta\ln v_p$ to produce distributions for bulk sound (v_c) perturbations. In the usual models of thermal heterogeneity, shear modules (μ), bulk modulus (κ), and density vary together at each depth in such a way as to maintain perfect correlations in the lateral variations of ρ, $v_s=\sqrt{\mu/\rho}$, $v_p=\sqrt{(\kappa/\rho)+4(\mu/3\rho)}$, and $v_c=\sqrt{\kappa/\rho}$. Thus, seismic constraints on these correlations are an important test of the extent to which the assumption of dominantly thermal heterogeneity is valid.

3. Results

The 2D marginal of Fig. 1b, the covariances of Fig. 2, and the output parameters of Fig. 3 are typical of our model space maps, and imply that most velocity and density parameters are resolved within our search bounds and independently constrained by the data. The high likelihood regions are well within most of our model space boundaries, and most off-diagonal covariances are significantly weaker than parameter variances. In fact, there is only one off-diagonal covariance,

←

Fig. 4. Likelihood distributions for correlations of lateral velocity and density heterogeneity in five mantle layers, as determined by model fit to long period seismic data. With a discrete probability density, $P(x_i)$ defined so that $\Sigma_i P(x_i)\Delta_x=1$, likelihood is defined by $L(x_i)=P(x_i)/\Delta_x$. In this sense, likelihoods are normalized probabilities. 1 000 000 models were generated by drawing random deviates from 2D $\delta\ln v_s$–$\delta\ln v_p$ and $\delta\ln v_s$–$\delta\ln\rho$ marginals (Fig. 1b) for each spherical harmonic component and layer. Distributions are shown for combined degree 2 and 4 models. Full-scale vertical lines show the correlations given by the coefficients of our most likely model. Half-scale vertical lines give the correlations for these layers predicted by model SPRD6. The distributions for $\delta\ln v_s$–$\delta\ln\rho$ subjected to a gravity filter (see Section 4) are also shown. The great likelihood of low or negative $\delta\ln v_s$–$\delta\ln v_c$ and $\delta\ln v_s$–$\delta\ln\rho$ correlations is the most notable feature of these plots. Anticorrelations for density–velocity are most likely in the TZ and LLM layers.

that between the deepest and shallowest density layers, that is strong for enough individual parameters to be a significant consideration in our layer-average statistics. This tends to increase the uncertainty of LLM by an amount that depends on the amplitude of UUM structure, an effect which is measured by our probabilistic approach. The tradeoff also suggests that crustal errors could contaminate LLM results in a manner not quantified by the NA 'maps', but we have found that for our data such effects are dominated by crustal thickness, errors in which are quite small as degrees 2 and 4. The topography parameters for the 670 and CMB (not shown) are also constrained within the search boundaries, with little topography–density tradeoff.

Figs. 1, 2 and 3 also illustrate the relationship between model space 'maps' and the more conventional seismic inversion outputs. Fig. 1b reveals a velocity–density tradeoff, details of which are concealed by the more common representation as a single element of the correlation matrix in Fig. 2. The high likelihood region of the full 2D marginal is double peaked and extends from the origin in the direction of positive density perturbations and negative v_s perturbations. The multiple local minima are a characteristic signature of the null space of an ill-posed inverse problem [2], and show why both damped inversions and 'downhill' searches for best fits can misrepresent the model space. The predictions from model SPRD6 shown in Figs. 1 and 3 demonstrate that the high probability regions of model space 'maps' are consistent with inversion results. The inversions associated with the SPRD6 study use starting models near the origin of Fig. 1 so it is not surprising that the damping produced results within the likelihood peak nearest that origin and failed to find the high probability models with larger density perturbations.

As demonstrated by Fig. 3, all but a handful of the high likelihood ranges for individual parameters are consistent with SPRD6 values, despite the fact that our likeliest values are often more different from SPRD6 than SPRD6 is from PREM. At the same time, it is evident that while most velocity parameters are robustly non-zero, most of the density perturbation ranges include zero. Thus, specific density models are not very robust, which is consistent with the conclusions of several inversion studies [23,24]. However, Fig. 3 also demonstrates more general characteristics of our density models that prove to be both robust and distinct from inversion results. In contrast with the most likely v_s and v_p perturbation parameters, which are well correlated with one another and have amplitudes near those predicted by SPRD6, it is evident that the most likely density coefficients for the LLM are poorly correlated with the velocities and are of consistently greater amplitudes than the SPRD6 model predictions. Figs. 4 and 5 show that these characteristics are present in other layers and are robust.

Fig. 4 displays likelihood distributions for various lateral correlations of seismic velocities and density. Velocity–velocity and velocity–density correlations for our most likely models generally are consistent with the peaks of the overall distributions and also fairly consistent with the SPRD6 results. The v_s–v_p distributions show that the long period data distinctly favor positive correlations at all depths. Because v_s and v_p have similar dependence upon μ, this is the expected result, and serves as confirmation of the method. However, correlations as high as those of SPRD6 are seen to be unlikely. Indeed, fewer than 10% of our models in any layer have correlations above the 90% confidence level for perfect positive correlation at both degrees 2 and 4 [43], and the peaks of our distributions are near the much lower correlation levels reported by other inversion studies [44]. In contrast, v_c depends only on κ and ρ, and the distributions of Fig. 4 favor $\delta \ln v_c$ poorly or negatively correlated with $\delta \ln v_s$. Together, then, the velocity–velocity correlation distributions imply that our seismic data strongly favor uncorrelated or anticorrelated lateral variations in bulk and shear moduli. This is difficult to explain without the existence of significant chemical heterogeneity [6,13].

We observe greater consistency among the ρ–v_s correlation results (SPRD6, most likely, and likelihood distribution peaks) than among the velocity–velocity correlations. In other tests, we have found that our density models tend to be as consistent with one another and with SPRD6 density

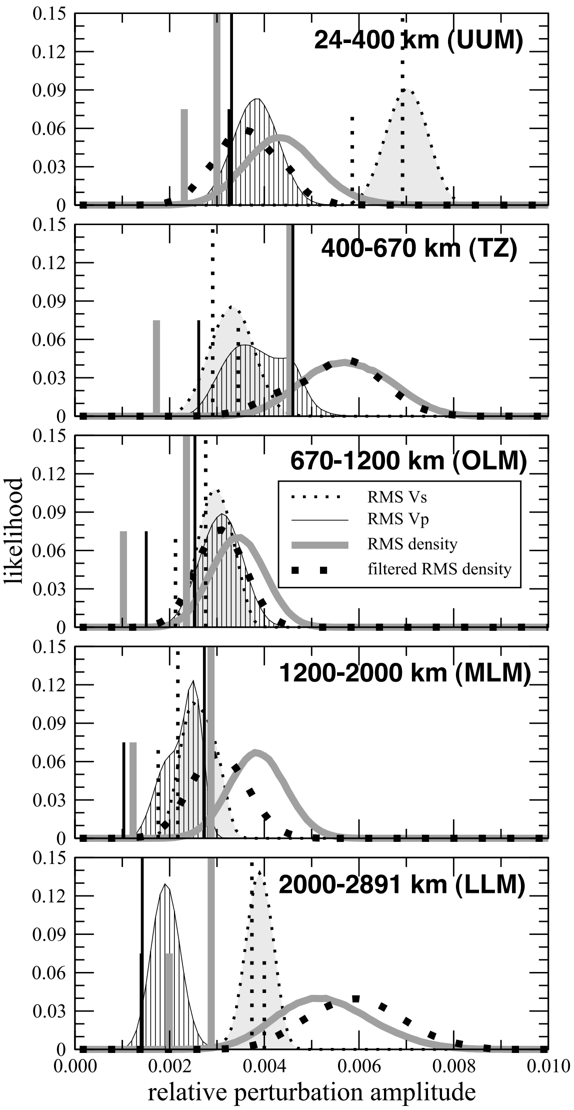


Fig. 5. Probability distributions for RMS amplitudes of global models of $\delta\ln v_s$, $\delta\ln v_p$ and $\delta\ln\rho$, constructed from marginals as in Fig. 4, except that marginal tails are excluded. Distributions are shown only for degrees 2 and 4 together. Full-scale and half-scale vertical lines show, respectively, the amplitudes of our most likely model and those predicted by model SPRD6 for each layer. The SPRD6 model predictions underestimate our observed $\delta\ln\rho$ amplitude distributions, which are likely to exceed $\delta\ln v_s$ everywhere below 400 km depth. Density distributions subjected to a gravity filter (Section 4) are included. These confirm that the elevated densities below 400 km are consistent with static gravity data.

as are various well-established long period velocity models [17,19,44] with one another. The density models also are much better correlated with one another than they are with velocity models. This implies that the data contain a coherent signal from mantle density heterogeneity unexplained by density that scales with velocity. We also observe that our density models in the various layers do not tend to display strong positive or negative correlations with CMB or 670 topography, but it is unclear if resolution of these correlations is good enough to make them meaningful.

While our most likely density–velocity correlations confirm the low correlations of SPRD6 and our most likely velocity–velocity correlations are consistent with several inversions [17,19,44,45], it is the possession of complete distributions that allows us to draw robust conclusions about the unlikeliness of high correlations for v_s–v_c and v_s–ρ. This is relevant because of the existence of apparently conflicting inversion results which yield more strongly positive v_c–v_s correlations [18], and observations of a wide range of acceptable velocity–density correlations [23,24]. The distributions in Fig. 4 have wide ranges of likely values, demonstrating that the observed variety of inversion results is consistent with the data, and can result from the use of different reasonable choices of damping. At the same time, it is clear that high positive correlations are much less likely than small or negative values.

Correlations are useful for detecting chemical heterogeneity, but more information is needed to determine which combinations of thermal and chemical heterogeneity best explain the seismic data. Complementary constraints are provided by heterogeneity amplitudes. Fig. 5 displays likelihood distributions for the root-mean-squared (RMS) amplitudes of $\delta\ln v_s$, $\delta\ln v_p$, and $\delta\ln\rho$ models. These are constructed in the same manner as the correlation distributions, except that we use only the portion of the 2D marginals with likelihoods greater than $1/e$ of the peak likelihood. This adjustment is necessary because amplitude distributions are strongly non-linear functions of the coefficient marginals and are unstable with respect to the shape of the tails of the coefficient distributions. The primary characteristic of the amplitude distributions is that, relative to SPRD6 values, the ranges for v_p and ρ distributions are more strongly elevated than the v_s distributions. The amplitudes of CMB and 670 topography, which are not shown because they are poorly resolved, are similarly elevated relative to SPRD6 amplitudes, but within the range established by other topography studies [46–48]. It should also be noted that density distributions are everywhere wider than the velocity distributions, which reflects the fact that the resolving power of the data is poorer for density than for velocity.

These amplitude distributions contradict the usual conclusion that long period seismology supports $\delta \ln v_p/\delta \ln v_s$ and $\delta \ln \rho/\delta \ln v_s$ ratios less than unity throughout the mantle. The most noticeably elevated amplitudes are for $\delta \ln \rho$ in the TZ and LLM layers, where most of the density distributions are to the right (higher amplitudes) of the v_s distributions. In addition, the favored ranges for $\delta \ln \rho$ and $\delta \ln v_p$ amplitudes in the OLM and MLM, as well as that for $\delta \ln v_p$ in the TZ, mostly overlap the corresponding v_s distributions. Thus $\delta \ln v_p/\delta \ln v_s < 1.0$ is favored only in our top and bottom layers, and $\delta \ln \rho/\delta \ln v_s < 1.0$ is unlikely everywhere below 400 km depth.

The implications of this observation are profound enough that we have performed several additional tests to confirm that it is robust. First, we have confirmed that the v_s–v_p–ρ interdependences implied by our 2D marginals (Fig. 1b) do not result in different ratios than those implied by the separate amplitude distributions of Fig. 5. This is done by using the random deviates to accumulate distributions for the amplitude ratios. Distributions were accumulated both for whole-model RMS amplitudes and point-by-point model comparisons. We also constructed distributions in which only well correlated velocity and density models were included. In all cases the distributions were peaked near the ratios implied by the peaks of the distributions of Fig. 5.

Second, we have inspected our full suite of 1D and 2D marginals to investigate the sources of the difference, visible in Fig. 5, between the velocity and density amplitudes implied by our single most likely model and those implied by the peaks of the likelihood distributions. The amplitudes of the most likely model are uniformly to the left (lower amplitudes) than the amplitude distribution peaks. This can result from either: (1) individual model parameters exhibiting null space effects such as asymmetric and/or multipeaked probability distributions like those of Fig. 1b, or (2) relatively wide distributions centered on small values like most of those for LLM density in Fig. 3.

With the first type of amplitude effect, the null space of our ill-posed problem results in low amplitude most likely models that are poor representatives of a full model distribution for which mean or median models have higher amplitudes. We have found this to be the dominant effect in the TZ density amplitude distribution of Fig. 5. In such cases the amplitude of the most likely model is much less relevant for geophysical inferences than is the location and width of the amplitude distribution peak. With the second type of amplitude effect the likelihood of high amplitudes is primarily attributable to poor resolution. In this case, the amplitude of the most likely model is important because it serves as an approximate lower bound on model amplitudes. It can do so because it contains information about well resolved features like the I(4,4) component of the LLM density in Fig. 3, which are unlikely to disappear if and when resolution is improved.

The amplitude effects for the other peaks in Fig. 5 are a mixture of the two effects that dominate the LLM and TZ density peaks. With these taken into account, our observation of elevated $\delta \ln v_p/\delta \ln v_s$ and $\delta \ln \rho/\delta \ln v_s$ ratios appears to be robust. Even in the least resolved case of $\delta \ln \rho$ in the lowermost mantle, the lower limit for the $\delta \ln \rho/\delta \ln v_s$ is still ~ 0.6, considerably higher than that suggested by recent inversion studies [22,49].

4. Application: consistency with gravity data

The distributions of Figs. 4 and 5 are robust representations of the constraints on mantle velocity and density provided by the long period seismic data. However, distributions for density are relatively wide. This serves to remind us that seismic resolution is poor, and other forms of data could reveal the density of the 'true' Earth to differ significantly from that implied by the peaks of the seismic distributions. Gravity measurements are an important example of such data, and were employed in the SPRD6 inversions [22]. In that study, the gravity data could only be incorporated through joint inversions that introduced further a priori damping and certain geodynamic assumptions. With the seismic constraints represented as distributions, we can now incorporate additional data constraints a posteriori. This avoids the problem of finding damping that appropriately weights different kinds of data. Combining data in this way to make inferences about the interior of the Earth is where our new seismic probability distributions may find their most useful application.

Here, we reconsider the example of combining seismic and gravity constraints. In so doing, we can avoid geodynamic assumptions, because our distributions reflect all tradeoffs between density at various depths and topography on the CMB and 670, and it is trivial to calculate the static gravitational potential for each model of density and topography. The gravity contribution of the CRUST5.1 model is also calculated and subtracted from measurements of the gravitational potential [50] to yield mantle gravity data set. The error in this data set is dominated by uncertainty in the crustal correction, which we estimate by comparing the CRUST5.1 correction to that of an earlier version of the same model. We then remove from our random deviate distributions all density models inconsistent with the crust-corrected gravity data and errors.

This creates new, 'filtered' likelihood distributions that are shifted relative to the original distributions, as seen in Figs. 4 and 5. Only shifts of the v_s–ρ correlations and $\delta\ln\rho$ peaks are shown because the other peak shifts are negligible. The gravity data give enough new information about density to produce visible peak shifts. The most notable shifts are movements of amplitude peaks to lower values. This is not surprising, because of the tendency of amplitude distributions to be elevated by the very distribution tails that are reduced by the improved resolution gravity provides. However, all pairs of filtered and unfiltered peaks are separated from one another by less than a half-width, suggesting that the two data sets are reasonably consistent with one another. Thus, Figs. 4 and 5 clearly demonstrate that the poor v_s–ρ correlations and high $\delta\ln\rho$ amplitudes favored by the seismic data are consistent with existing gravity data.

In an earlier test, we performed gravity filtering on a distribution of density models that was inconsistent with the model space because the random deviates were not drawn from a complete set of 2D marginals (tradeoffs between the density parameters of different layers had been ignored). In that case, consistency with the gravity data was much poorer, with shifts of the correlation peaks as large as the current shifts in the amplitude peaks and clear separation of some filtered and unfiltered amplitude distributions. This enabled us to diagnose the problem in our random deviates.

5. Discussion

Figs. 4 and 5 are simple representations of a wealth of information about long period seismic constraints that we have obtained in this study, and we will eventually employ our probability marginals in much more detailed analyses. Nonetheless, these figures provide several immediately useful pieces of information. In particular: (1) high positive $\delta\ln v_s$–$\delta\ln v_c$ and $\delta\ln v_s$–$\delta\ln\rho$ correlations at long wavelengths are unlikely almost anywhere in the mantle; (2) the seismic data favor $\delta\ln\rho$ amplitudes equal to or greater than $\delta\ln v_s$ amplitudes in the TZ and below; and (3) these characteristics are most pronounced in the TZ and LLM, although less robust in the latter. Our model space observations are only as good as the data that inform them, but should be more representative of seismic constraints than are the often very different results from damped inversions of similar data.

As noted in Section 3, observation (1), combined with the relative strength of the degree 2 and 4 signal in deep mantle seismics [21], implies that thermal heterogeneity does not dominate chemical heterogeneity below 400 km depth. Observation (2) offers an important clue to the nature of the chemical component of heterogeneity. Recent studies have confirmed that density variations of greater relative amplitude than velocity variations are most readily achieved with variable iron concentrations [6]. Iron contamination from the core or iron repartitioning among mantle minerals is a potential source of such variation in the deepest mantle [51], and the formation of iron-depleted cratons [12,52] is a potential source in the upper parts of the mantle. Observation (3) suggests that processes of mantle formation or convection have concentrated the chemical heterogeneity above the major seismic discontinuities of the mantle, although, due to resolution limitations, our evidence for chemical heterogeneity in the OLM and MLM is almost as strong as for that in the LLM.

While this kind of heterogeneity could not persist in the most typical models of thermally driven whole-mantle convection [53], it may well be compatible with recent models of ocean crust recycling [54], partial convective layering [14,15,55], or more complete models of thermo-chemical convection [16]. Such hypotheses can be tested by extending this study to shorter wavelengths and greater radial resolution and then extracting more detailed information from our model probability marginals. The most obvious tests to be performed include the accumulation of probability distributions for density and velocity at each point on the globe. The way these distributions vary from place to place can then be compared to maps of subduction history [45,56], or to results of body wave tomography for the CMB region [57]. Geographical correlations of these properties can determine how long wavelength density heterogeneity and other geophysical observables are related. Additionally, the statistical nature of our model space 'maps' makes them well suited to comparisons with statistical descriptions of geodynamic models. These include probability distributions for Bullen's parameter [58], heterogeneity spectra [59], and correlation length [60,61].

As long as no systematic errors are believed to exist in the long period data, the above evidence for geochemical convection and stratification from robust tomography should be accounted for in any future geochemical and geodynamic mantle models. In addition, the contrast between our results and those of damped inversions serves to highlight the misrepresentation of data that can be produced by damping. We hope that this motivates the use of robust approaches to other inverse problems.

Acknowledgements

Most of the computations for this research were performed using the Edinburgh Parallel Computer Centre facilities, access to which was provided through the TRACS program of the EC. Access to Research Infrastructure initiative. The facilities of the Minnesota Supercomputer Institute were also employed, with the cooperation of David Yuen. We are deeply indebted to Malcolm Sambridge for making available his algorithms, and we appreciate the constructive criticism provided by Barbara Romanowicz, Don Anderson and an anonymous reviewer. The research was supported by the Netherlands NWO, Earth Sciences grant 809_31.003.***[SK]***

References

[1] A. Tarantola, Inverse Problem Theory, Methods for Data Fitting and Model Parameter Estimation, Elsevier, Amsterdam, 1987.
[2] A.N. Tikonov, V. Arsenin, Solutions of Ill-Posed Problems, John Wiley, New York, 1977.
[3] T. Tanimoto, Waveform inversion for three-dimensional density and s wave structure, J. Geophys. Res. 96 (1991) 8167–8189.
[4] J. Woodhouse, A. Dziewonski, Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms, J. Geophys. Res. 89 (1984) 5953–5986.
[5] O. Anderson, E. Schreiber, R. Liebermann, N. Soga, Some elastic constant data on minerals relevant to geophysics, J. Geophys. Res. 89 (1984) 5953–5986.
[6] S. Karato, B. Karki, Origin of lateral variation of seismic wave velocities and density in the deep mantle, J. Geophys. Res. 106 (2001) 21771–21784.
[7] M. Richards, B. Hager, Geoid anomalies in a dynamic earth, J. Geophys. Res. 89 (1984) 5987–6002.

[8] A. Forte, R. Woodward, A. Dziewonski, Joint inversion of seismic and geodynamic data for models of three-dimensional mantle heterogeneity, J. Geophys. Res. 99 (1994) 21,857–21,877.
[9] F. Birch, Elasticity and constitution of the earth's interior, J. Geophys. Res. 57 (1952) 227–286.
[10] D. Anderson, Theory of the Earth, Blackwell Scientific, Boston, MA, 1989.
[11] T. Jordan, Composition and development of the continental tectosphere, Nature 274 (1978) 544–548.
[12] A.M. Forte, C.H. Perry, Geodynamic evidence for a chemically depleted continental tectosphere, Science 290 (2000) 1940–1944.
[13] A.M. Forte, J. Mitrovica, Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data, Nature 410 (2001) 1049–1056.
[14] L.H. Kellogg, B.H. Hager, R.D. van der Hilst, Compositional stratification in the deep mantle, Science 283 (1999) 1881–1884.
[15] D. Anderson, The case for irreversible chemical stratification of the mantle, Int. Geol. Rev. 44 (2002) 97–116.
[16] P. Tackley, Strong heterogeneity caused by deep mantle layering, Geochem. Geophys. Geosyst. 3 (2002) available at www.g-cubed.org.
[17] W.-J. Su, A. Dziewonski, Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, Phys. Earth Planet. Inter. 100 (1997) 135–156.
[18] B. Kennett, S. Widiyantoro, R. van der Hilst, W. Warren, G. Ackland, J. Crain, Joint tomography for bulk sound and shear wave speed in the Earth's mantle, J. Geophys. Res. 10 (1998) 12,469–12,493.
[19] G. Masters, G. Laske, H. Bolton, A. Dziewonski, The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: Implications for chemical and thermal structure, in: S. Karato (Ed.), Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale, Seismology and Mineral Physics, Geophys. Monogr. Ser., Vol. 117, AGU, Washington, DC, 2000, pp. 63–87.
[20] X. He, J. Tromp, Normal-mode constraints on the structure of the Earth, J. Geophys. Res. 101 (1996) 20,053–20,082.
[21] J.S. Resovsky, M.H. Ritzwoller, New and refined constraints on three-dimensional Earth structure from normal modes below 3 MHz, J. Geophys. Res. 103 (1998) 783–810.
[22] M. Ishii, J. Tromp, Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle, Science 285 (1999) 1231–1236.
[23] J.S. Resovsky, M.H. Ritzwoller, Regularization uncertainty in density models estimated from normal mode data, Geophys. Res. Lett. 26 (1999) 2319–2322.
[24] C. Kuo, B. Romanowicz, On the resolution of density anomalies in the Earth's mantle using spectral fitting of normal mode data, Geophys. J. Int. 150 (2002) 162–179.
[25] M. Sambridge, Exploring multidimensional landscapes without a map, Inverse Problems 14 (1998) 427–440.
[26] C. Beghein, J.S. Resovsky, J. Trampert, P and s tomography with a neighbourhood algorithm, Geophys. J. Int. 149 (2002) 646–658.
[27] J. Trampert, P. Vacher, N. Vlaar, Sensitivities of seismic velocities to temperature, pressure and composition in the lower mantle, Phys. Earth Planet. Inter. 124 (2001) 255–267.
[28] M. Sambridge, Geophysical inversion with a neighbourhood algorithm – I. Searching a parameter space, Geophys. J. Int. 138 (1999) 479–494.
[29] M. Sambridge, Geophysical inversion with a neighbourhood algorithm – II. Appraising the ensemble, Geophys. J. Int. 138 (1999) 727–746.
[30] J.S. Resovsky, J. Trampert, Reliable mantle density error bars: an application of the neighbourhood algorithm to normal model and surface wave data, Geophys. J. Int. 150 (2002) 665–672.
[31] J. Trampert, J. Woodhouse, Assessment of global phase velocity models, Geophys. J. Int. 144 (2001) 165–174.
[32] H.J. van Heijst, J. Woodhouse, Global high-resolution phase velocity distributions of overtone and fundamental-mode surface waves determined by mode branch stripping, Geophys. J. Int. 137 (1999) 601–620.
[33] J. Resovsky, R. Pestana, Improved lower mantle Vp constraints from spectral fitting of normal mode data, Geophys. Res. Lett. 30 (2003) 1383–1386.
[34] G. Masters, G. Laske, F. Gilbert, Matrix autoregressive analysis of free-oscillation coupling and splitting, Geophys. J. Int. 143 (2000) 478–489.
[35] R. Widmer-Schnidrig, Application of regionalized multiplet stripping to retrieval of aspherical structure constraints, Geophys. J. Int. 148 (2002) 201–213.
[36] Y.K. Wong, Upper Mantle Heterogeneity from Phase and Amplitude Data of Mantle Waves, Ph.D. Thesis, Harvard University, Cambridge, MA, 1989.
[37] W. Mooney, G. Laske, G. Masters, Crust 5.1: a global crustal model at 5 deg×5 deg, J. Geophys. Res. 103 (1998) 727–747.
[38] F. Dahlen, J. Tromp, Theoretical Global Seismology, Princeton University Press, Princeton, NJ, 1998.
[39] J.S. Resovsky, M.H. Ritzwoller, A degree 8 mantle shear velocity model from normal mode observations below 3 MHz, J. Geophys. Res. 104 (2001) 100–110.
[40] A. Dziewonski, D. Anderson, Preliminary reference Earth model, Phys. Earth Planet. Inter. 25 (1981) 25,297–25,356.
[41] R.D. van der Hilst, H. Kárason, Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: toward a hybrid convection model, Science 283 (1999) 1885–1887.
[42] D. Anderson, Top-down tectonics?, Science 293 (2001) 2016–2018.
[43] D. Eckhardt, Correlations between global features of terrestrial fields, Math. Geol. 16 (1984) 155–171.
[44] D. Vasco, L. Johnson, Whole earth structure estimated from seismic arrival times, J. Geophys. Res. 103 (1998) 2,633–2,671.
[45] R. Saltzer, R.V. der Hilst, H. Karason, Comparing p and

s wave heterogeneity in the mantle, Geophys. Res. Lett. 28 (2001) 1335–1338.

[46] M. Flanagan, P. Shearer, Global mapping of topography on transition zone velocity discontinuities by stacking ss precursors, J. Geophys. Res. 103 (1998) 2673–2692.

[47] Y. Gu, A. Dziewondki, C.B. Agee, Global decorrelation of the topography of transition zone discontinuities, Earth Planet. Sci. Lett. 157 (1998) 57–67.

[48] L. Boschi, A. Dziewonski, Whole earth tomography from delay times of p, pcp, pkp phases: lateral heterogeneities in the outer core, or radial anisotropy in the mantle?, J. Geophys. Res. 105 (2000) 25567–25594.

[49] B. Romanowicz, Can we resolve 3d density heterogeneity in the lower mantle, Geophys. Res. Lett. 28 (2001) 1107–1110.

[50] F.G. Lemoine, S.C. Kenyon, J.K. Factor, R.G. Trimmer, N.K. Pavlis, D.S. Chinn, C.M. Cox, S.M. Klosko, S.B. Luthcke, M.H. Torrence, Y.M. Wang, R.G. Williamson, E.C. Pavlis, R.H. Rapp, T.R. Olsen, The development of the joint nasa gsfc and national imagery mapping agency (NIMA) geopotential model EGM96, NASA/TP-1998-206861 NASA, 1998, GSFC, Greenbelt, MD.

[51] J. Badro, G. Fiquet, F. Guyot, J.-P. Rueff, V. Struzhkin, G. Vanko, G. Monaco, Iron partitioning in the earth's mantle: toward a deep lower mantle discontinuity, Science 300 (2003) 789–791.

[52] F. Deschamps, R. Snieder, J. Trampert, Anomalies of temperature and iron in the uppermost mantle inferred from gravity data and tomographic models, Phys. Earth. Planet. Inter. 129 (2002) 245–264.

[53] P. Keken, E. Hauri, C. Ballentine, The generation, preservation, and destruction of chemical heterogeneity, Annu. Rev. Earth. Planet. Sci. 30 (2002) 493–525.

[54] N. Coltice, Y. Ricard, Geochemical observations and one layer mantle convection, Earth Planet. Sci. Lett. 174 (1999) 125–137.

[55] L. Wen, D. Anderson, Layered mantle convection: A model for geoid and topography, Earth Planet. Sci. Lett. 145 (1997) 367–377.

[56] C. Lithgow-Bertelloni, M. Richards, The dynamics of cenozoic and mesozoic plate motions, Rev. Geophys. 36 (1998) 27–78.

[57] E. Garnero, J. Revenaugh, Q. Williams, T. Lay, L. Kellogg, Ultralow velocity zone at the core-mantle boundary, in: The Core-Mantle Boundary, AGU, Washington, DC, 1998, pp. 319–334.

[58] C. Matyska, D.A. Yuen, Bullen's parameter η: a link between seismology and geodynamic modelling, Earth Planet. Sci. Lett. 198 (2002) 471–483.

[59] P. Tackley, D. Stevenson, G. Glatzmaier, G. Schubert, Effects of multiple phase transitions in a three-dimensional spherical model of convection in earth's mantle, J. Geophys. Res. 99 (1994) 15877–15901.

[60] T. Jordan, P. Puster, G. Glatzmaier, P. Tackley, Comparison between seismic earth structures and mantle flow models based on radial correlation functions, Science 261 (1993) 1427–1431.

[61] G.R. Helffrich, B.J. Wood, The Earth's mantle, Nature 412 (2001) 501–507.