Chapter 3

P and S tomography using normal mode and surface wave data with a Neighbourhood Algorithm

Abstract

Traditionally P and S wave tomography is based on the inversion of data which are sensitive to the desired Earth structure, and model covariance is estimated from imperfect resolution and data error propagation. This analysis ignores the usually large null-spaces, and hence significant non-uniqueness of the solution, encountered in seismic tomography problems. Here, we performed a model space search for P- and S-velocity structure to find acceptable fits to recent normal mode splitting and fundamental mode phase velocity data. The survey of the model space employed the Neighbourhood Algorithm of Sambridge which preferentially samples the good data-fitting regions. A Bayesian approach was used subsequently to extract robust information from the ensemble of models. We particularly focussed on posterior marginal probability density functions and covariances for the various model parameters. The covariance matrix obtained is very useful in providing insights on the trade-offs between the different variables and the uncertainties associated with them. We stay in the framework of perturbation theory, meaning that our emphasis is on the null-space of the linear inverse problem rather than the neglected non-linearity. The whole model space (including the null-space) was sampled within reasonable parameter bounds, and hence the error bars were determined by all fitting models rather than

---

The content of this is chapter was published in Beghein,C., Resovsky,J. & Trampert,J., Geophysical Journal International, 149, 646-658 (2002)
subjective prior information. We estimated P and S models for spherical harmonic degree
two only. The uncertainties are quite large and corresponding relative errors can exceed
100% in the mid-mantle for \( V_p \). We found a good correlation between our most likely
S model and previous models, with some small changes in amplitude. Our most likely
P model differs quite strongly from the recent P model SB10L18 and the correlation be-
tween our most likely P and S models is small. However, among all the good data-fitting
models, many have a significant \( V_p - V_s \) correlation. We computed \( d\ln V_s / d\ln V_p \) from
those models that correlate significantly. We found an increase with depth in the top 1500
kilometres. Deeper in the mantle, normal mode data prefer modest values compared to
travel time data.

3.1 Introduction

Several tomographic models of the Earth’s mantle were produced over the past 15 years,
using different kinds of data, parameterizations and inversion techniques. Although there
are some robust patterns, the models present large discrepancies (Resovsky and Ritz-
woller, 1999b). This is partly due to the propagation of data errors through the inversion
operator and imperfect resolution. Tomographic inverse problems are generally ill-posed
(uneven and inadequate sampling of the Earth and inadequate model parameterization)
and ill-conditioned (small errors in the data can lead to large variations in the model es-
timation owing to very small eigenvalues), resulting in large data and model null-spaces.
These null-spaces are usually dealt with by employing some kind of regularisation, i.e.
choosing one particular model out of many which are compatible with the data. Different
authors use different regularisation schemes and hence produce discrepancies in the ob-
tained models. We propose to sample the model space to have an overview of all models
compatible with the data rather than choosing one by some subjective regularisation.

An example of how a model null-space component can change the resulting model
was shown by Deal et al. (1999). They were able to incorporate additional, independent
information in a tomographic image without affecting the misfit. They added a thermal
model of a subducting plate to a high-resolution, three-dimensional tomographic study of
the Tonga-Fiji region. Since the thermal model did not fit the seismic data, they added
only the component of the theoretical slab insensitive to the seismic data using their null-
space shuttle. Vectors of the null-space having no effect on the data prediction, the new
tomographic model was in equally good agreement with the data. Unless the employed
regularisation is derived from true physical information, it can add artefacts to the tomo-
graphic model.

A second advantage of mapping the model space concerns error analysis. Most lin-
earized inversions give a posterior model covariance smaller or equal to the prior covari-
ance by construction (Tarantola, 1987). If the cost function to be minimized has a large
valley, i.e. there is a large model null-space, the posterior covariance can be seriously
underestimated, depending on the prior covariance (Trampert, 1998). We argue that the
width of the valley in the cost function is a realistic representation of the error bars in the
We propose, here, to use a forward modelling approach to explore the model space, including the null-space. We suppose that perturbation theory is valid for our forward modelling. This means that the estimated error bars take the null-space of the linear inverse problem into account, but cannot account for the neglected non-linearity in the forward problem. Since we do not make an inversion, our results are not biased by the introduction of damping or any other non-physical a priori information, and since we stay within the linear theory the starting model is irrelevant. We applied the Neighbourhood Algorithm (hereafter referred to as NA) developed by Sambridge (1999a,b) to survey the parameter space and to find an ensemble of mostly “good” data-fitting models. This method is a novel direct search technique, conceptually very simple and able to exhibit a self-adaptive behaviour by sampling preferentially the regions of lower misfit. Once the survey of the parameter space is achieved, robust information on the ensemble can be extracted using a Bayesian approach, giving valuable indications on the errors and correlation of the model parameters. We applied this method to recent normal mode splitting measurements and fundamental mode surface wave phase velocity maps. The normal modes provide constraints on the long-wavelength structure of the Earth for compressional and shear wave anomalies in the mantle. Fundamental mode Rayleigh waves were included to constrain the upper mantle. The main purpose of this work is to establish the feasibility of our approach to mantle tomography. As a consequence we concentrate only on degree two structure.

An ensemble of “good” data-fitting joint shear and compressional wave velocity models were produced with the present set of normal mode and phase velocity data, and probabilistic information was retrieved. Error bars were then assigned to tomographic models. Having obtained most likely $d\ln V_s$, $d\ln V_p$ models and their respective error bars, we examined their correlation and their ratio, a quantity widely discussed in the mineral physics community and which is of particular interest for geodynamists.

## 3.2 Data and parameterization

The data set we used was composed of normal mode splitting functions and fundamental mode phase velocity models, corrected with the crustal model CRUST5.1 (Mooney et al., 1998). The free oscillations of a spherically symmetric, nonrotating and (transversely) isotropic Earth model have specific degenerate frequencies. The addition of asphericities and slight general anisotropy (under the conditions of application of perturbation theory) generates the splitting of multiplets into singlets with eigenfrequencies close to the degenerate eigenfrequency. Let us represent these three-dimensional model perturbations $\delta m(r, \theta, \phi)$ from the reference model in terms of spherical harmonic components:

$$
\delta m(r, \theta, \phi) = \sum_{s=0}^{s_{max}} \sum_{l=-s}^{s} \delta m^s(r) Y^s_l(\theta, \phi)
$$

(3.1)
where $Y^t_s$ are fully normalised and orthogonal spherical harmonics as defined in Edmonds (1960), with harmonic degree $s$ and azimuthal order $t$. The structure coefficients $k^t_s$ of a particular isolated multiplet (denoted by $k$) characterize the way the eigenfrequencies split. If we neglect boundary perturbations, except for the crustal correction, we are left to first order with a single linearized relation between Earth structure and structure coefficients:

$$k^t_s = \int_0^a \delta m^t_s(r) \frac{1}{r^2} \Delta M_s(r) r^2 dr$$

(3.2)

where $a$ is the radius of the Earth and $\frac{1}{r^2} \Delta M_s(r)$ is the volumetric structure kernel for perturbation $\delta m^t_s$ (with respect to PREM (Dziewonski and Anderson, 1981)). For more details about normal modes theory, the reader is referred to chapter 2 of this thesis or to Woodhouse and Dahlen (1978) or Dahlen and Tromp (1998).

Normal mode splitting measurements below 3mHz were recently made by Resovsky and Ritzwoller (1998) for coupled and uncoupled multiplets with good sensitivity to S and P velocity anomalies everywhere in the mantle. In addition to structure coefficients, they also estimated corresponding error bars using Monte Carlo simulations of the effect of theoretical errors and noise. The structure coefficient measurements were used to create synthetic seismograms that were first perturbed with “errors” and noise of the appropriate statistical characteristics, and then inverted for new coefficient estimates. The results of multiple calculations was observed to produce approximately Gaussian coefficient distributions whose widths provided the uncertainties. Also, the coefficients of different angular and azimuthal orders were observed to vary independently. Both measurements and estimated error bars can be found on the internet (http://phys-geophys.colorado.edu/geophysics/nm.dir/). We used their degree two structure coefficients, determined from uncoupled normal mode multiplets, but we excluded modes with sensitivity to the inner core. We kept measurements for 82 uncoupled mode multiplets, in particular 51 spheroidal modes and 31 toroidal modes.

To constrain the uppermost mantle, we added eight fundamental mode Rayleigh wave phase velocity models between periods of 40 and 275 seconds. For periods between 40 and 150 seconds, the models and errors are the average and standard deviation obtained from different studies: Trampert and Woodhouse (1995, 1996, 2001), Ekström et al. (1997), Laske and Masters (1996), Wong (1989) and van Heijst and Woodhouse (1999). For larger periods, we used the models of Wong (1989). Three dimensional models of phase velocity perturbations $\delta c/c$ ($\delta c$ is the phase velocity perturbation relative to a reference phase velocity $c$, PREM in this case) can be expanded into spherical harmonics and their coefficients $\delta c^t_s$ are related to the Earth’s three dimensional structure in a way similar to structure coefficients:

$$l(\delta c^t_s) = \int_0^a \delta m^t_s(r) K(r) r^2 dr$$

(3.3)

Note that kernels for phase velocity perturbations are independent of degree $s$ (they all correspond to kernels with $s = 0$), unlike the kernels relative to structure coefficients,
and neither of them depends on \( t \). \( l \) is an index that discriminates between different frequencies.

Unlike for normal mode splitting data, error bars on phase velocity maps are hard to obtain. Trampert and Woodhouse (2001) showed that the quality of published phase velocity models vary widely with the period. At each selected period, we took all existing models, averaged each spherical harmonic coefficient and estimated its standard deviation. This should account for different measuring techniques of phase velocity, different data coverage and different regularisation-schemes in the construction of the maps. The error bars determined for normal mode structure coefficients by Resovsky and Ritzwoller (1998) have the characteristic of being almost constant at a given degree. There is not much variation between the different orders of spherical harmonic coefficient. By analogy, we decided to assign average uncertainties to \( l \) \((\delta c_t c_s)\) independent of the order \( t \) of spherical harmonic and defined by:

$$
\sigma_t^2 = \frac{1}{2s+1} \sum_{t=1}^{2s+1} l \sigma_s^2
$$

(3.4)

where \( s \) is the degree of the spherical harmonic (degree two in this study) and \( l \sigma_s^2 \) is the variance estimated for one particular spherical harmonic coefficient. Fig. 3.1 shows \( \sigma_t \) as a function of the period of the surface waves considered. It appears that the error decreases almost linearly between 40 and 100 seconds and the curve flattens between 100 and 150 seconds. We decided thus to assign a constant uncertainty to models with periods between 150 and 275 seconds, the value computed at 150 seconds (the model of Wong (1989) being the only one available to us at longer periods). We assumed, for convenience, that the errors have a Gaussian distribution, but there are far too few models to test this hypothesis.

To implement the first part of the NA (sampling the parameter space), we need to define the fit of a model to the data. We chose the \( \chi^2 \) misfit, which is a measure of the average data misfit compared to the size of the error bar. It is defined by:

$$
\chi^2 = \frac{1}{N} \left[ \sum_{k=1}^{N_m} \left( \frac{k \psi_{c_t}^{th} - k \psi_{c_t}^{obs}}{\sigma_k^2} \right)^2 + \sum_{l=1}^{N_s} \left( \frac{l(\frac{\delta c_t}{c})^{th} - l(\frac{\delta c_t}{c})^{obs}}{\sigma_l^2} \right)^2 \right]
$$

(3.5)

where \( N \) is the total number of data, \( N_m \) is the number of normal mode data, \( N_s \) is the number of surface wave data. The upper index “th” stands for theoretical structure coefficients and phase velocity perturbations, predicted by equations 3.2 and 3.3, and “obs” refers to the measurements. \( \sigma_k \) \( (\sigma_l) \) is the estimated error bar corresponding to the \( k \)th \( (l \)th) data.

We parameterized our models with independent isotropic perturbations of the elastic coefficients \( \delta A \) and \( \delta L \) (with \( A = \kappa + \frac{4}{3} \mu = \rho V^2_p \) and \( L = \mu = \rho V^2_s \)) with respect to PREM. Where PREM is transversely anisotropic (at depths between 24 and 220 kilometres), we used the equivalent isotropic PREM. The notation \( A \) and \( L \) was introduced.
by Love (1927) and is usually used to describe radially anisotropic medium. The corresponding anisotropic sensitivity kernels are given in section 2.2.3 of this thesis (see also Tanimoto (1986), Mochizuki (1986) or Dahlen and Tromp (1998)), and were combined to derive the appropriate isotropic kernels. Finally \( \delta \mathbf{m} \) corresponds to \((\delta A, \delta L, \delta \rho)\). However, with the present set of normal mode data, it is not possible to resolve 3-D density perturbations in the mantle. Several authors have confirmed this (Resovsky and Ritzwoller (1999c), Romanowicz (2001), Resovsky and Trampert (2002)). Instead of taking density as one of our model parameters, we decided to scale density anomalies \( \delta \rho \) and shear wave velocity perturbations using \( \frac{d \ln V_s}{d \ln \rho} = 2.5 \) (Anderson et al., 1968). This constraint, together with the size of the model space, introduces some prior information in the problem. Our models were parameterized radially in 7 layers. The bottom and top depths of these layers are, in kilometres, \((2891, 2609), (2609, 2018), (2018, 1526), (1526, 1001), (1001, 670), (670, 220), (220, 24)\). They correspond to radial knots of PREM and are based on the layers defined in Resovsky and Ritzwoller (1998), which in turn are based on a Backus-Gilbert style resolution analysis (Backus and Gilbert, 1968). We gathered some of their layers into one new layer in order to decrease the number of
variables. Equation 3.2 reduces consequently to a sum over these seven layers:

$$k c^i = \sum_{i=1}^{7} \delta \overline{M}_s^i k \overline{M}_s^i$$  \hspace{1cm} (3.6)

with

$$k \overline{M}_s^i = \int_{r_{inf}^i}^{r_{sup}^i} k M_s(r)r^2 dr$$  \hspace{1cm} (3.7)

$r_{inf}^i$ and $r_{sup}^i$ are the lower and upper radii respectively of layer $i$ and $\delta \overline{M}_s^i$ is an average perturbation of parameter $M_s^i$ on layer $i$. A similar relation holds for fundamental mode phase velocity perturbations (equation 3.3). The lateral parameterization is in spherical harmonics:

$$\delta A(r, \theta, \phi) = \sum_{s=0}^{s_{max}} \sum_{t=-s}^{s} \delta A_s^t(r) Y_s^t(\theta, \phi)$$  \hspace{1cm} (3.8)

$$\delta L(r, \theta, \phi) = \sum_{s=0}^{s_{max}} \sum_{t=-s}^{s} \delta L_s^t(r) Y_s^t(\theta, \phi)$$  \hspace{1cm} (3.9)

These expansions allow us to solve the problem spherical harmonic coefficient by spherical harmonic coefficient. Because of the scaling relationship between $d\ln V_s$ and $d\ln \rho$, we were left with only two parameters in each layer, thus 14 model components for each structure coefficient or phase velocity coefficient.

### 3.3 Results

#### 3.3.1 Sampling and appraisal

We applied the NA using a linearized forward problem (equations 3.2 and 3.3). We wanted to survey the model space to find combinations of parameters $\delta A_s^t(r)$ and $\delta L_s^t(r)$ that give an acceptable fit to the data (equation 3.5). Because there is assumed to be no covariance among the data of different angular or azimuthal order, each of the five spherical harmonic coefficients at degree two could be treated independently. In other words, we could explore five parameter spaces separately, with 14 unknowns in each of them.

First, we had to fix the boundaries of the model space. We computed the equivalent degree two parameters $\delta L_s^t(r)$ for six different S models (MM2-L12D8 (Resovsky and Ritzwoller, 1999b), SKS12-WM13 (Su et al., 1994), S20RTS (Ritsema et al., 1999), SAW12D (Li and Romanowicz, 1996), S16B30 (Masters et al., 1996) and SB10L18 (Masters et al., 2000)). In each layer, we chose to search twice the range of the largest absolute amplitude obtained from those six models. Because model SB10L18 is a joint P and S model, we had a P model we could use to fix the range of parameter $\delta A_s^t(r)$. 
We compared the amplitude of the coefficient $\delta A_t(r)$ corresponding to model SB10L18 and those obtained by scaling (factor of two) the other S models. We fixed our range for parameter $\delta A_t$ as twice the absolute amplitude of the largest $\delta A_t$ found. The final result depends on the chosen range (even if the model space is completely sampled, because there is a possibility that a good fitting model exists outside the range), but because of fear of violating perturbation theory, we decided not to increase the range any further. At the same time, we ensured that these six S models were included in the model space we sample, as well as their corresponding P models. Fixing the boundaries is equivalent to taking a boxcar function as prior density probability distribution. This introduces boundary effects and the posterior probability density function is not necessarily Gaussian when the most likely model is close to the edge.

We refer the reader to Sambridge (1999a,b) for details about the NA. The first stage of the algorithm, the sampling, makes use of a geometrical construct, the Voronoi cells, to approximate the misfit function and to drive the search towards the best data-fitting regions while continuing to sample a relatively wide variety of different models. It is relatively easy to tune since only two parameters have to be set: $n_s$, the number of models generated at each iteration, and $n_r$ the number of “best” data-fitting Voronoi cells in which random walks are performed at each iteration. We were careful with the choice of these values, since it is very important to have a good initial sampling in order to make a meaningful Bayesian interpretation (which is the second stage of the NA). We had to avoid directing the search towards a local minimum and we had to sample the posterior probability density (PPD) with the highest possible accuracy, to sample all the good fitting regions of the model space. A way to do so is by increasing the values of the tuning parameters. As both tuning parameters increase together, the algorithm is more explorative as a sampler but also less efficient at mapping details of the most important (the best fitting) parts of the model space. It is not possible to draw general conclusions on the tuning parameters since every problem is different and requires specific parameters. We had to find their most appropriate values by trial and error. Sambridge explained that the minimum sample size required is very sensitive to the dimension of the problem (Sambridge, 1998). His experiments showed that, to get a good enough approximation of the posterior probability density, the required sample sizes have to be increased when the dimension of the problem becomes larger. We decided to have $n_s$ equal to $n_r$, which is the minimum value $n_s$ can take, in order to broaden the survey. We started to sample the model space with some relatively low values of $n_s$ and $n_r$ (10 or 20), and we increased the tuning parameters successively and compared the different results. For some variables, we got different results if we used such small values. This indicates that for these tuning parameters, the results are not independent of the tuning. The chosen values of $n_r$ are not large enough to identify all the models compatible with the data. After some trials with $n_r$ ranging up to 200, we decided to use $n_r = 50$. It was the smallest value above which the results appeared to be independent of the tuning and that explored all the good fitting regions. The algorithm is very effective in finding the regions of lowest misfit. To choose the number of iterations, we looked at the evolution of the misfit with time. $\chi$
3.3 Results

decreased very rapidly, the sampling being directed towards the cells having the best fit. We decided to stop the survey when the misfit has “flattened”, and the model distribution was observed to approximate a likelihood sampling. The latter condition indicates that the sampling is adequate for analysis using the NA Bayesian resampling (see below). When $n_r$ is increased, more models must be generated and hence the survey requires more time. The sampling was completed in about two hours on a SUN Ultrasparc machine (400 MHz) for $n_r = 200$ and 300 iterations and it only took half an hour for $n_r = 50$ and 600 iterations.

Models generated by the Neighbourhood Algorithm

![Diagram showing models generated by the Neighbourhood Algorithm](image)

Figure 3.2: Results of the sampling using the real part of $c^1_2 (\delta c^1_2 / c)$. The scale represents the $\chi$ misfit as defined in equation 3.5. Parameters on the horizontal axis correspond to perturbations in elastic coefficient $A (= \delta (\rho V_p^2))$ and parameters on the vertical axis correspond to perturbations in $L (= \delta \mu = \delta (\rho V_s^2))$, with respect to PREM. The full range of the models space is shown.
As an example, we show in Fig. 3.2 the result using the real part of $c_1^2$ in four of our seven layers. It is a way to depict the shape of the model space with regions of higher misfit (light grey) and regions of better fit (dark), where the sampling density is larger. We see that there is a clear global minimum in the upper mantle. For some other variables, the best fitting region is more elongated. This shows that, within the boundaries of our model space, a large range of values for these variables are compatible with the data. It is the case of perturbations in $\delta A = \delta (\rho V_p^2)$ in the mid-mantle and deeper in the mantle. Several variables have their global minimum situated at the edges of the model space. Increasing the range of the survey has the effect of reducing the minimum misfit somewhat but, because of the trade-offs among the model parameters (visible in the correlation matrix, Fig. 3.3), moving the global minimum of one of them implies that other variables move as well and may, in turn, be directed towards the edge. Therefore we believe that, as long as there are trade-offs between model parameters, the model space cannot be surveyed guaranteeing that no solution is on the edge of the model space. This is not a major problem though, because we can quantify the trade-offs as explained below. Further, increasing the range of search substantially will violate perturbation theory and require a complete new set-up of the problem.

We then needed to extract quantitative information on the models previously generated. Sambridge (1999b) provides an approach to the appraisal problem, based on a Bayesian point of view. This is the second stage of the algorithm. This appraisal of the ensemble is the most time-consuming part of the algorithm. In most model space search techniques, inferences are drawn from the good fitting part of the ensemble only, and sometimes even from a single member. What is new in this algorithm is that the entire ensemble is used, the “bad” data-fitting models as well as the “good” ones, and an efficient summary of the sampled models is provided. In a Bayesian approach, the information contained in the models is represented by a posterior probability density function (PPD). In the absence of restrictive prior information on the models, the model that maximizes the PPD is the model with the best data fit. The PPD can be used to compute quantities such as the posterior mean model, the posterior model covariance matrix and marginal posterior probability density functions. An integration over the parameter space is performed using a “likelihood sampling” whose density corresponds to the PPD. To do this, the algorithm uses the misfit of the sampling to create a likelihood sampling. This requires a new ensemble of points to be generated (the “resampled” ensemble) whose distribution follows the approximate PPD. This is one of the main factors influencing the computation time. Once the resampling is done, computing the Bayesian integrals requires only simple averages over the resampled ensemble.

For the resampling of the model space, tuning parameters have to be set: the number of random walks to perform and the number of steps per random walk. Finding the parameters that optimize the coverage of the Gibbs sampler is a trial and error process, which is time-consuming but it is characteristic to every direct search technique. It is preferable to use multiple random walks instead of a single one, and the number of steps per walk must be large enough to insure the convergence. After each trial, convergence
3.3 Results

When the PPD is Gaussian, the trade-off between model parameters can be represented by the correlation matrix, which is derived from the off-diagonals of the posterior

Figure 3.3: Correlation matrix corresponding to the real part of $c_2^1 (\delta c_2^1 / c)$. 

can be checked, numerical errors evaluated and the length and number of random walks can be adjusted accordingly. For ensembles generated with 50 cells and 600 iterations, we needed between 6000 and 10000 steps per walk, depending on the coefficient treated, and four to six walks in order to achieve the convergence of the integrals. It took approximately seven hours on a SUN Ultrasparc (400 MHz) to compute 1-D marginals, 2-D marginals and the correlation matrix. We also computed Bayesian integrals for some of the ensembles generated with 200 cells, to make sure the results were the same as the ones obtained with 50 cells. Since their sampling required more iterations, more points were needed for the computation of the integrals. The appraisal then required three to four days.
model covariance matrix. The correlation matrix corresponding to our problem is different for each spherical harmonic coefficient. The one obtained for the real part of $c_{12}^1$ is represented in Fig. 3.3. We can see that variables of the mid-mantle and lower mantle are highly correlated to one another. They are also correlated to model parameters of the upper mantle, for instance $\delta A$ between 2018 and 2609 km of depth (layer 2) with $\delta L$ in the uppermost layer. There are, to some extent, trade-offs among all the model parameters. In order to improve these results, we would need to add independent data sensitive to one or the other of a given pair of correlated model parameters. The correlation matrix is an excellent tool to see what is actually resolvable and what additional data is most desirable.

The covariance matrix can also be used to get the variances of the model parameters (obtained from its diagonals). However, the use of a covariance matrix only makes sense when the PPDF is a Gaussian distribution. A more general way of looking at variances and trade-offs is given by marginal distributions. A one-dimensional marginal is the probability of each value of a particular model parameter, given all possible variations of the others. The width of those posterior marginals gives a further indication on the constraint we have on each variable, and can be assimilated to error bars. A two-dimensional marginal shows the likelihood of each combination of values for a pair of model parameters. Computing the joint marginal of two model parameters implies an integration over all the other parameters. Fig. 3.4 shows 2-D marginals of some pairs of variables for the real part of $c_{12}^1$. These marginals are a robust way to look at model parameter trade-offs. The diagonal elongated shape observed for some pairs of variables shows the trade-off between them. For instance, the top-left marginal of Fig. 3.4 shows a trade-off between $\delta A$ in the lowermost layer (layer 1) and the layer above (layer 2, between 2018 and 2609 kilometres of depth) and the bottom middle marginal shows a trade-off between P anomalies ($\delta A$) in layer 4 (between 1000 and 1526 kilometres of depth) and S anomalies ($\delta L$) in layer 3 (at depths between 1526 and 2018 km). Even with the additional constraints provided by surface waves, the upper and the lower mantle are not completely independent. The bottom-left marginal shows a trade-off between $\delta L$ in the upper mantle (layer 7) and $\delta A$ between 2018 and 2609 kilometres of depth (layer 2). We also show two pairs of model parameters that almost do not correlate in Fig 3.3 : $\delta A$ and $\delta L$ between 2018 and 2609 kilometres of depth (bottom-right) and $\delta A$ and $\delta L$ at depths between 670 and 1000 km (top-right). Neither shows the elongated diagonal pattern characterizing a trade-off.

In Fig. 3.5, we show the 1-D marginals for a few model parameters. $\delta A$ at depths between 2018 and 2609 kilometres can take a large range of values compatible with the data. The width of the 1-D marginal is large. In this particular case, the sign of the perturbation is barely constrained. On the contrary, in the upper layers, between 220 and 1000 km of depth, both P and S anomalies are better determined. If we compare the 1-D marginal and Fig. 3.2 for $\delta A$ between 1000 and 1526 km of depth, we observe a difference. From Fig. 3.2, one could expect a wide range of possible values for that parameter and thus a 1-D marginal with a shape similar to the one obtained for $\delta A$ between 2018 and 2609
kilometres of depth. Its width is relatively large but it clearly peaks towards the edge, at a positive value. This difference between the sampling and the Bayesian interpretation of this sampling comes from the resampling of the ensemble that has to be made in order to evaluate an approximate PPD. These marginals also show that all the parameters in the model space do not exactly have a Gaussian distribution. When a most likely parameter is located towards the edge of the model range, boundary effects destroy the Gaussian shape. Another strong influence on the Gaussian distribution comes from the null-space. If the model parameters are Gaussian distributed, the PPD is Gaussian. In the presence of a null-space the PPD will show an elongated valley and appear non-Gaussian. Away

2-D Marginals

Figure 3.4: 2-D marginals corresponding to the real part of $c_2^1$ ($\delta c_2^1/c$). The white triangle denotes the values for model SB10L18. and 90 % (blue) confidence levels. Going from inside towards the edges, the solid lines represent the 30 % (white), 50 % (grey) and 90 % (black) confidence levels. Parameters on the horizontal axis correspond to perturbations in elastic coefficient $A \left( = \delta(\rho V_p^2) \right)$ and parameters on the vertical axis correspond to perturbations in $L \left( = \delta \mu = \delta(\rho V_s^2) \right)$, with respect to PREM. The full range of the models space is shown. Layer 1 corresponds to the deepest layer and layer 7 represents the range 24-220 km depth.
1-D Marginals

lower mantle
2018 km < d < 2609 km

mid-mantle
1001 km < d < 1526 km

670 km < d < 1000 km

220 km < d < 670 km

Figure 3.5: 1-D marginals corresponding to the real part of \( c_2^1 (\delta c_2^1 / c) \). The vertical black line is the position of model SB10L18. The full range of the models space is shown.

from the edges of our sampling range, the 1-D marginals are thus a measure of the size of the null-space. The 2-D marginals and the off-diagonal elements of the correlation matrix compare qualitatively well, but the amplitudes of the correlations are affected by large non-Gaussian distributions. Using a Gaussian assumption for the PPD will underestimate the posterior uncertainties on the model parameters in the presence of a large null-space.

We did not test different parameterizations. Choosing another parameterization would probably slightly modify the results since, as shown by the correlation matrix, all model parameters are to some extent correlated to one another. The layers we chose were based on those of Resovsky and Ritzwoller (1998), which approximate the optimal depth layers that can be resolved for the given depth kernels. Modifying the layers would thus only degrade the posterior covariance. The important point is that the results should only be considered with the correlation matrix.
3.3 Results

3.3.2 Shear and compressional wave velocity models and error bars

This section deals mainly with the most likely models. These are the models corresponding to the maximum of the 1-D marginals and not the mean model as calculated directly by the NA. The reason for this is that the mean model is only meaningful when the PPD is Gaussian and it is not exactly the case for all the parameters in the model space, as discussed in section 3.3.1. For the same reason, we preferred estimating the width of the 1-D marginals instead of reading them directly in the covariance matrix. We took this width as the largest distance where the amplitude of the most likely models has decreased by a factor 1/e. It usually gives slightly larger variances than those read in the covariance matrix.

![Graphs showing root mean square amplitudes for S and P anomalies](image)

Figure 3.6: Root mean square amplitudes of our most likely models, their robust part and model SB10L18 as a function of depth. Fig. 3.6a (left) corresponds to S anomalies. Fig. 3.6b (right) corresponds to P anomalies.
In Fig. 3.6, we compare the root mean square (rms) amplitude of our most likely velocity models, their robust part (the most likely model from which we subtract the uncertainties) and the rms of model SB10L18, as a function of depth. SB10L18 is a joint P and S model derived from the inversion of body waves, surface waves and normal mode splitting data. The rms of the robust part of our models corresponds to a lower limit for S and P model amplitudes. The size of the anomalies in SB10L18 and in our most likely model is similar. They differ the most between 220 and 670 km depth and in below 1000 km for $V_s$. For $V_p$, the amplitudes are close in the top 670 kilometres and between 1000 and 1526 kilometres of depth. The differences are partly due to the presence of body wave data in SB10L18.

Figure 3.7: Correlation coefficient between our most likely models and other models as a function of depth. Fig. 3.7a (left) represents the correlation between our most likely S model with 5 other S models. Fig. 3.7b (right) represents the correlation between our most likely models with the joint P and S model SB10L18.
We show in Fig. 3.7a the geographical correlation of our $dlnV_s$ model with other S models (degree two): MM2-L12D8 (Resovsky and Ritzwoller, 1999b), SKS12-WM13 (Su et al., 1994), S20RTS (Ritsema et al., 1999), SAW12D (Li and Romanowicz, 1996) and S16B30 (Masters et al., 1996). These models were derived from linearized inversions. For most models, in some layers, the correlation is above the 90% confidence level for degree two maps (corresponding to a correlation coefficient of 0.73 (Eckhardt, 1984)) and there are two layers, in the mid-mantle, where the correlation is very low, with any other model. One could expect it to be better for models MM2-L12D8 and S20RTS that incorporated the same set of normal mode data as we used. This poor correlation is related to the trade-off between $V_p$ and $V_s$ in the mid-mantle (Fig. 3.3). We searched for a joint P and S model, whereas MM2-L12D8 and S20RTS assume a scaling between $dlnV_p$ and $dlnV_s$. These very different constraints on $V_p$ lead to different $V_p$ models and, in the mid-mantle, the trade-offs imply that $V_s$ can also be very different. Fig. 3.7b shows the geographical correlation for degree two of our most likely models with model SB10L18, for both P and S since SB10L18 is a joint P and S model. It is interesting to see that the correlation coefficient between our S model and the S model of SB10L18 is above the 90% confidence level in most layers. For the P models, the correlation is not as good, especially in the lowermost mantle where the correlation coefficient is close to zero. They correlate reasonably well between depths of 220 and 670 kilometres and between 1000 and 1526 km.

In Fig. 3.8, we plotted the maps corresponding to our most likely degree two S and P models in the seven layers we used. A robust feature from all tomographic models to date is that the degree two structure dominates the two lowermost layers ($2609 \text{ km} \leq d \leq 2891 \text{ km}$ and $2018 \text{ km} \leq d \leq 2609 \text{ km}$) and in the transition zone (Resovsky and Ritzwoller, 1999b). These maps are thus good approximations to complete tomographic models only at those depths. As we see from the correlation coefficient in Fig. 3.7b, both S models agree quite well at most depths. On the contrary, the two P models differ much more.

In Fig. 3.9 we plotted the correlation between our most likely P and S models. The correlation is significantly high (above the 90% confidence level) only between depths of 220 and 670 km. The lowest correlations are situated in the lowermost and uppermost layers. This is not representative in the uppermost mantle because degree two structure is not dominant there. On the contrary, the lowermost mantle is believed to be dominated by that degree. We cannot, however, conclude directly that there is no correspondence between P and S anomalies in the lowermost mantle. One of the advantages of using the NA is that it provides error bars on the models. Within these error bars, there may be models compatible with the data that have a higher $dlnV_s - dlnV_p$ correlation. Thus, a better way to look at that correlation is by taking into account the uncertainties on the models.
Figure 3.8: Degree two maps of our most likely S (left) and P (right) models compared to degree two maps of models SB10L18.
3.3 Results

To take into account all the models, we sampled $\delta A^t_s(r)$ and $\delta L^t_s(r)$ within their estimated error bars with a random number generator, we deduced the corresponding $dlnV^t_s(r)$ and $dlnV^t_p(r)$ and we computed the correlation coefficient for all the models generated. Note that we can easily compute $dlnV^t_s(r)$ and $dlnV^t_p(r)$ from $\delta A^t_s(r)$ and $\delta L^t_s(r)$, because we imposed a scaling relationship between $\delta \rho$ and $\delta V_s$. The distributions of correlation values can be plotted as histograms. It is interesting to note that the histograms are largely independent on how we sampled the models (uniformly, Gaussian or corresponding to the actual 1-D marginals). We took the median of the correlation coefficient and we estimated its uncertainty. The uncertainty of the median was obtained by computing its scaled median absolute deviation (SMAD) (Bevington, 1969). We used the SMAD as uncertainty on the overall correlation coefficient. In Table 3.1 are listed the median correlation coefficients and their uncertainty at different depths, and Fig. 3.10 gives the histograms for different depths. In our lowermost layer, where our most likely P and S models do not correlate (the correlation coefficient was -0.417), we see that that the median value is also very low (-0.45) and lots of models are anticorrelated. The uncertainty is, however, very large and there are a few models compatible with the data that correlate well. There are models, in the layers above, that correlate well and some that correlate

![Figure 3.9: S-P correlation as a function of depth.](image)
Figure 3.10: Distribution of $d\ln V_s - d\ln V_p$ correlation coefficient between all the models compatible with the data as a function of depth.

significantly. In the uppermost layer, correlation is poor again, but there, degree two is far from dominant. This is one illustration of the importance of error bars in tomographic models.

Error bars on the spherical harmonic components of $d\ln V_s$ and $d\ln V_p$ can easily be derived from those on $dA_s^j$ and $dL_s^j$ because of the scaling between $d\ln \rho$ and $d\ln V_s$. In Fig. 3.11, we show the relative error bars for the rms amplitude, $d(rms)/rms$, for both shear and compressional wave velocities. This represents the size of the error bars on the rms relative to the size of the most likely model. The error bars are large, especially for $V_p$. These uncertainties are mainly due to the large null-space associated with our problem. It shows what can be determined by the data alone. Prior information can of course reduce the nullspace, but we argue that choosing a model should be based on
3.3 Results

<table>
<thead>
<tr>
<th>Depth (km)</th>
<th>median</th>
<th>SMAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2609 &lt; d &lt; 2891</td>
<td>-0.535</td>
<td>0.425</td>
</tr>
<tr>
<td>2018 &lt; d &lt; 2609</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>1526 &lt; d &lt; 2018</td>
<td>0.205</td>
<td>0.36</td>
</tr>
<tr>
<td>1001 &lt; d &lt; 1526</td>
<td>0.20</td>
<td>0.27</td>
</tr>
<tr>
<td>670 &lt; d &lt; 1001</td>
<td>0.003</td>
<td>0.42</td>
</tr>
<tr>
<td>220 &lt; d &lt; 670</td>
<td>0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>24 &lt; d &lt; 220</td>
<td>-0.29</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 3.1: Correlation between all models generated within their error bars. The second column gives the median of the distribution and the third column gives its robustness.

physical information rather than a subjective damping parameter. Unless such a physical prior information exists, it is preferable to consider all models compatible with the data. Another way to reduce the size of the null-space is of course by adding more data. We expect that including body wave data will significantly reduce the uncertainties.

Perturbations in $V_s$ are believed to be larger than perturbations in $V_p$. The value of their ratio $R = \frac{d\ln V_s}{d\ln V_p}$ is commonly used as a diagnostic whether the heterogeneities in the mantle have a chemical or thermal origin. The value of $R$ in the deep mantle is still a controversy. A low ratio could be explained by a thermal origin of the anomalies and a high value (above 2.5) could indicate a chemical component in the heterogeneities. Looking at various studies, there is a large variety of values for $R$. A good review on the subject can be found in Masters et al. (2000). Because different authors compute $R$ in different ways, Masters et al. (2000) recalculated a spherical averaged $R$ for various models. Generally, studies using body wave travel time data only prefer a high value of $R$ in the lowermost mantle. Robertson and Woodhouse (1996) used ISC data and found a ratio slowly increasing from 1.7 to 2.5 for depth between 600 and 2000 km of depth. Bolton (1996) used long period body wave data and his results also showed an increase of $R$ with depth up to a value slightly below 3.5 at the bottom of the mantle. Both studies were constrained inversions, i.e. a perfect proportionality between $P$ and $S$ anomalies at every depth was imposed but $R$ was allowed to change. Su and Dziewonski (1997) used ISC data, long period body wave data and surface wave data to perform an unconstrained inversion for bulk sound and shear wave speed. They obtained a value of $R$ of almost 3.5 in the lowermost mantle. Saltzer et al. (2001) employed ISC data to produce $P$ and $S$ models of the mantle. They distinguished between regions where there has been subduction in the last 120 million years and where there has not. They found a peak value of $R = 3$ around 2200 km of depth in non-slab regions and a ratio smaller than 2 at all depths in slab regions. There are two models using ISC data that show a low ratio at all depths. These are the models of Vasco and Johnson (1998) and Kennett et al. (1998). Their values of $R$ are smaller than 1.5, and even smaller than 1 for Vasco and Johnson (1998). When normal modes are included in the data set, the tendency observed in joint
Figure 3.11: Relative uncertainties for the root mean square amplitude of our most likely S model (dashed line) and our most likely P model (solid line) as a function of depth.

inversions is to have a more modest ratio in the lowermost mantle (model SB10L18), except for Romanowicz (2001) who found a value between 1 and 2 everywhere but at depths greater than 2000 km where $R$ is larger than 2.5. Model SB10L18 was obtained from the joint inversion for bulk sound and shear wave speed, whereas Romanowicz (2001) performed inversions for $V_p$ and $V_s$ directly. In their paper, Masters et al. (2000) also observed that constrained inversions for bulk sound and shear wave speed usually give P and S models that correlate significantly, while there is less correlation in inversions for $V_p$ and $V_s$.

We computed $R$ on a $10^\circ \times 10^\circ$ grid. We took into account all the models obtained for $V_s$ and $V_p$, in the same way we did to compute the correlation between all our P and S models. We generated a series of values of $\delta A_i(r)$ and $\delta L_i(r)$ within their estimated error bars, recombined the five spherical harmonic coefficients to get $\delta A(\theta, \phi)$ and $\delta L(\theta, \phi)$ on each grid point, and we computed the corresponding $d\ln V_p(\theta, \phi)$, $d\ln V_s(\theta, \phi)$ and $R$ for all the possible combinations. The distributions obtained expressed the variation of $R$ for various $(\theta, \phi)$ and various models generated within the posterior model uncertainties. We determined the median of these distributions, which we took as the spherically averaged estimate of $R$. The SMAD is used to derive the uncertainty on $R$ (as in Masters et al. (2000)). We found that $R$ takes modest values in the lowermost mantle (see Fig. 3.12, dotted line). The question is whether $R$ is low in the lowermost mantle (where degree two

---

Relative error bars for root mean square amplitudes

degree 2

$V_p$, $V_s$

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

$d\text{ln} V_p$ vs. depth (km)

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

$d\text{ln} V_s$ vs. depth (km)

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

d$\text{rms}$/rms

Figure 3.11: Relative uncertainties for the root mean square amplitude of our most likely S model (dashed line) and our most likely P model (solid line) as a function of depth.
3.3 Results

Figure 3.12: Ratio between degree two $d\ln v_s$ and $d\ln v_p$ as a function of depth.

structure is dominant) because the correlation is low or because the normal mode data generally prefer a lower ratio. To answer that, we computed $R$ by only taking the models which correlate significantly (above 0.73) in one particular layer. The values are plotted in Fig. 3.12 (solid line). We observe an increase of $R$ with depth in the top 1500 km of the mantle, taking values between 1 and 1.75, and modest values in the lowermost mantle. This shows that normal mode data favour low values for $R$ in the deeper mantle, in contradiction with Romanowicz (2001), who found $R$ up to 3.5 at depths > 2000 km, using degree two normal mode splitting data and a layered parameterization. The possible reasons for this difference are that Romanowicz does not include surface wave measurements in her data and introduces strong prior information in the inversions (the damping was chosen so that $d\ln v_s/d\ln v_p$ matches the range 1.5 - 2 in the top 1500 km, slightly higher than what we find).
3.4 Conclusion

The aim of this paper was to explore a model space with a direct search method to identify good data-fitting isotropic Earth models. We used the Neighbourhood Algorithm developed by Sambridge (1999a,b), which is a new derivative-free direct search technique that preferentially samples the good data-fitting regions of a model space. A Bayesian approach was used subsequently in order to extract robust information from the ensemble of models generated. We examined the posterior marginal probability density functions, the variances of the various model parameters and the correlations among them. The whole model space, including the null-space, was sampled within reasonable bounds, and the error bars are consequently more realistic than traditional inversion error estimates. This new technique appears to be very efficient in finding the best data-fitting regions in a high dimensional space (provided the size of the model space has been established). It is easy to tune since there are only two tuning parameters but they have to be chosen carefully to avoid the search being trapped in a local minimum and to importance sample the entire parameter space, while not reducing the efficiency of the algorithm. The optimal values have to be found by trial and errors. The subsequent use of the Bayesian algorithm does not require any further solutions of the forward problem. The parameter space is resampled instead, using only information from the initial survey of the model space. The accuracy of the Bayesian integrals will depend on the way the model space was initially sampled.

We applied the Neighbourhood Algorithm to the search of isotropic mantle shear and compressional wave velocity models, using recent normal mode splitting measurements and fundamental mode phase velocity data. We found an ensemble of joint P and S models (spherical harmonic degree two only) for which posterior marginal probability density functions, correlation and covariances were computed. 1-D marginals give information on how well a parameter is constrained, and show whether it is Gaussian distributed or not. They are used to infer error bars on the various model parameters. 2-D marginals show the trade-offs among pairs of model parameters. The posterior covariance matrix obtained under a Gaussian assumption gives fair representation of the correlation, although error bars may be underestimated in case of strong non-Gaussian distribution. The uncertainties and correlations constitute essential information in order to make meaningful analysis of the models obtained. We see, for instance, that there is a correlation between parameters of the lower mantle and parameters of the upper mantle. This implies that we could improve our models by adding data, sensitive exclusively to either the upper or lower mantle in order to decorrelate those parameters. Our most likely S model is highly correlated with other S models, such as MM2-L12D8, SKS12-WM13, S20RTS, SAW12D or S16B30, except where there is a high trade-off between P and S perturbations. The correlation with the recent S model SB10L18 is very high at most depths, but our most likely P model is very different from the P model of SB10L18 in most layers. Degree two error bars on the rms velocity perturbations of our mean models are quite large, especially for P. The correlation between our most likely P and S model is low but among all the
3.4 Conclusion

models generated and compatible with the data, there are some that have a high $V_p - V_s$ correlation. We calculated the ratio $R = \frac{d\ln V_s}{d\ln V_p}$, and we found an increase with depth in the top 1500 kilometres up to a value of 1.75. Deeper in the mantle, there is a decrease and $R$ oscillates between 0.8 and 1.3. Many models using body wave data alone seem to find a high value for $R$ in the deep mantle. This is, however, not always the case, as shown by the ratios found by Saltzer et al. (2001) in slab regions. Our study, for which only surface wave and normal mode data have been employed, tends to favour low values. The model of Masters et al. (2000) obtained from surface waves, normal mode and body wave data gives an intermediate ratio at these depths.

We have a powerful new tool to explore a model space, including the null-space, and to estimate ranges of "good" data-fitting models. In the absence of true physical information, we prefer considering the full range of models consistent with the data. Correlations between model parameters can be computed and give valuable indications on what kind of independent data should be added to constrain the models better. Because our problem can be solved spherical harmonic coefficient by spherical harmonic coefficient, it is trivial to parallelize the procedure resulting in huge timesavings for computing a full tomographic model.