High Pressure and High Temperature Research on Lithosphere and Mantle Materials
Proceedings of the International School Earth and Planetary Sciences
Siena 1996, 67-87.

Seismic Tomography and Constraints on the Composition of the Upper Mantle

Guust Nolet - Department of Geosciences, Princeton University, NJ 08544, USA
Kostas Papazachos - Institute of Engineering Seismology and Earthquake Engineering, Thessaloniki, Greece.

INTRODUCTION

A longstanding and fundamental problem in the Earth Sciences has been to determine the composition of the Earth. Whereas the geologist is able to sample rocks at the surface that are representative for the crust, and occasionally the upper part of the mantle, most of the Earth's materials are out of direct reach, and we have to resort to indirect soundings, sophisticated methods of inference, and very demanding laboratory experiments to increase our knowledge of the Earth's interior.

The clues that we have come from fields as diverse as geochemistry, geophysics, and astronomy. Sometimes, these clues are fairly direct. Melts that have found their way to the surface yield geochemical information on the nature of the magma source, and xenoliths provide a direct sampling of the wall rocks within the mantle through which magma migrated. But the deeper we reach, the more blurred the picture becomes. Analyses of isotope ratios or trace element abundances may indicate the existence of different magma reservoirs, but we have difficulty identifying where they are or what their exact nature is. For example, we do not know for certain if hotspots orginate at the core mantle boundary or in the upper mantle. Astronomy gives us the total mass and moment of inertia for the Earth with a very high precision (one part in 10^5), but these constraints are too general for application to a particular region of interest, say, the asthenosphere.

The magnetic, electromagnetic and gravity field of the Earth all contribute to our understanding of the 'big picture' as well, but the geophysical parameters of largest interest to the mineralogist are the seismic velocities. Where a gravity anomaly is found by integrating all density anomalies over a large volume, seismic waves traverse a sharply defined trajectory through the Earth, much like light rays through a lense. The techniques of inverse theory, combined with advances in the manipulation of large matrix systems on fast computers, can be applied to translate the seismic observables into an image of velocity perturbations. The massive growth of digital seismographs that are sensitive to a broad range of frequencies has propelled seismology to the forefront of the battle we fight to discover the Earth's secrets. The fact that seismology is now able to discover heterogeneities at depth (while too long committed to perfecting an idealized Earth model with spherical symmetry) has been

a prime factor in re-establishing the link between this important discipline and the rest of the Earth sciences.

'Seismic tomography' is the name under which a collection of imaging techniques are now collectively known - somewhat inappropriately since this suggest a comparison with medical techniques that is beside the mark, as even a cursory closer inspection shows. It is important to realize that the tomographic images have severe shortcomings despite their indisputable value. As we shall see, these shortcomings arise on the one hand from a lack of data, which prohibits us to average out the often large observation errors, but also from theoretical limitations on what seismic waves can resolve. Whereas improvements are to be expected as the data volume grows with time, some of the theoretical limitations are more fundamental.

To make the seismic measurements useful for the geochemist or mineralogist, the seismic models need to be compared with laboratory data on the elastic properties and density of different rocks and minerals. It is at this point that perhaps the largest uncertainty is introduced. The seismologist measures the seismic velocity *in situ*, under conditions that would prohibit the measurement of seismic velocity in a laboratory, and we must extrapolate these measurements to mantle pressure and temperature - the latter itself being an unknown parameter. And if we can extrapolate, a unique identification is hindered by the existence of many minerals with comparable seismic velocities and the often large influence of temperature or of small quantities of volatiles on elastic properties. Fortunately, the number of candidate minerals decreases strongly as we direct our attention to deeper layers in the Earth.

As this tutorial paper shows, the problem of the determination of the Earth's composition is very much a multi-disciplinary problem. It requires that seismologists understand the needs and wishes of the mineralogists, and that mineralogists learn to evaluate seismological results. The purpose of this paper is to build that bridge from the seismologist's bank of the river.

SEISMIC VELOCITIES

In this section we briefly recapitulate some basic facts of seismology. In an isotropic Earth, we distinguish compressional or P waves, and transverse or S waves. The velocities of these two wave types relate to the shear modulus μ, the incompressibility κ, and the density ρ as:

$$v_P = \sqrt{\frac{\kappa + \frac{4}{3}\mu}{\rho}} \quad (1)$$

$$v_S = \sqrt{\frac{\mu}{\rho}} \quad (2)$$

At high frequencies, the wave energy travels in narrow bundles or *rays*. When a seismic P wave inflicts on a boundary, it refracts following Snell's law (Figure 1).

It also generates a reflected P wave, and reflected and refracted S waves. Similarly, a verically polarized S wave inflicting on a horizontal boundary such as the Moho may convert part of its energy to P. Such converted waves often provide crucial information on boundaries in the earth, but they are more difficult to observe. A horizontally polarized S wave can only generate P energy if the boundary is not horizontal.

The wavelengths λ of P and S waves scale with the frequency f as $\lambda_P = v_P/f$ and $\lambda_S = v_S/f$. The width of the ray is defined as the region in space occupied by all rays that arrive with a phase delay less than π and is often named 'Fresnel zone' like its analogon in optics. In homogeneous materials, a ray of length L km has a maximum width at its center of $\sqrt{\lambda L}$ km. We commonly use this expression also as a rule of thumb for the real Earth, although heterogeneity does influence the ray width notably in the upper mantle and near the core-mantle boundary. The width of a ray is a fundamental theoretical limitation to the resolving power of the seismic ray, as the Fresnel zone is an approximate indication of the volume in the Earth that influences the wave speed. For example, a P ray with a length of 4000 km, a representative mantle velocity of 11 km/s and a frequency of 2 Hz has $\lambda \approx 11/2 = 5.5$ km and a maximum width of 148 km.

If the Earth is slightly anisotropic, seismic waves keep their compressional or transverse character, but the S wave splits in two waves with different polarization and different velocity. The observation of split S waves is indicative of the existence of a layer where crystals are both anisotropic and aligned, or of a layer with aligned cracks.

SURFACE WAVES

Surface waves are low frequency waves that concentrate their energy close to the free surface of the Earth. In a simplified view, such surface waves can be imagined as S waves trapped between the surface and an interface that reflects them supercritically, such as the Moho for 'fundamental modes' or one of the velocity gradients in the upper mantle for 'higher modes' (Figure 2). Although the individual rays that make up the surface wave travel at an angle with the vertical direction, bouncing up and down, the sum of all these waves creates a wavefront that travels horizontally, with a phase velocity that depends on frequency. This simple view ignores some P wave energy that is usually trapped in the sediment layers or the oceanic water layer, as well as diffracted energy that does not travel in the form of rays, but on the whole it is a useful analogon to understand the properties of surface waves. Horizontally polarized S waves give rise to a type of surface wave known as the Love wave; Rayleigh waves involve vertically polarized S waves and some trapped P energy near the surface. Below the reflector depth, the surface wave energy is not zero but decreases exponentially with depth. Surface waves have longer wavelengths, often of several hundred km, and consequently wider Fresnel zones than P or S rays. Surface waves

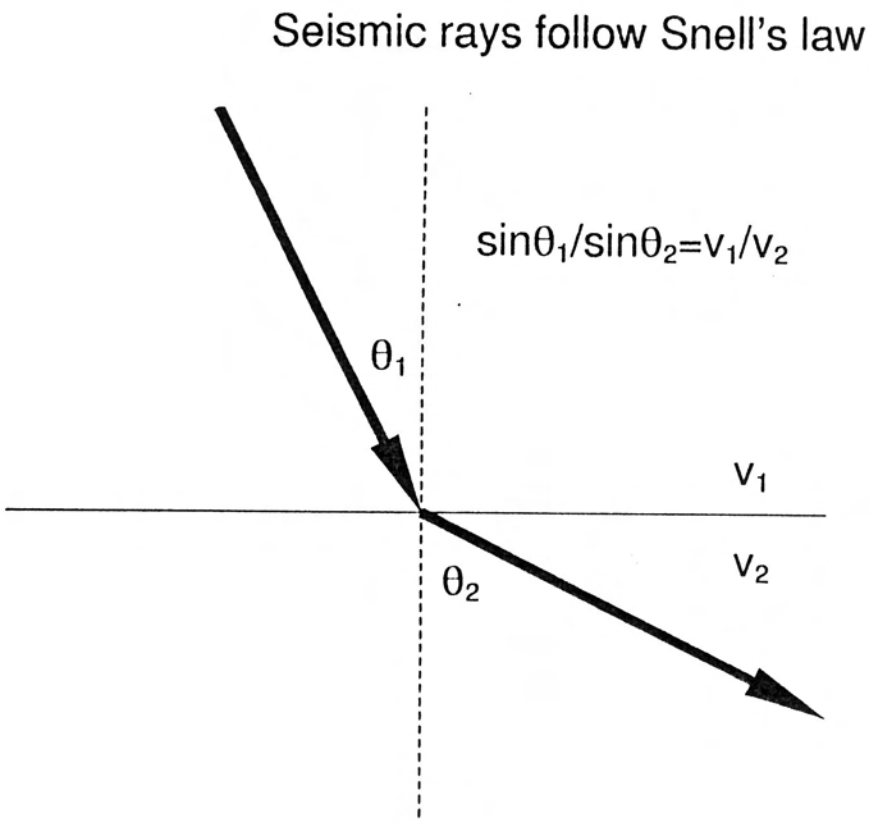

Figure 1: Seismic rays in horizontally media refract according to Snell's law, which says that $\sin\theta/v$ is constant, where θ is the angle with the vertical.

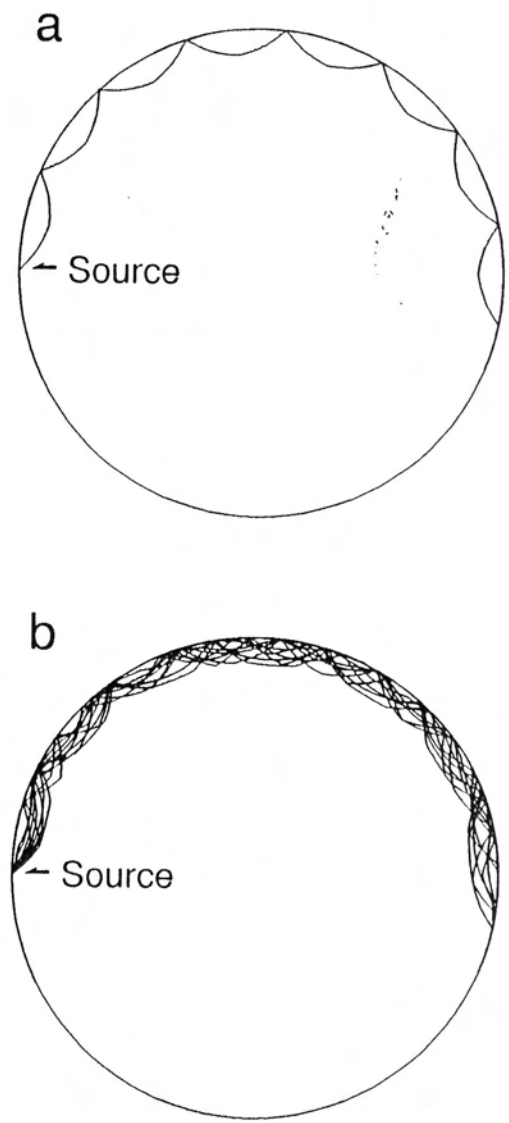

Figure 2: (a) An S wave leaving the source at a shallow angle returns to surface, only to be bounced back again many times. For strong earthquakes, such multiple S waves can circle the entire Earth, (b) there are many such multiple S waves, each starting with a different angle. An observer at the surface will see many different waves. Constructive interference between the multiples gives rise to 'surface waves', travelling with a horizontal phase velocity that depends on frequency

spread their energy over the full channel in which they are trapped, and therefore the velocity of one surface wave frequency gives us only the averaged properties of that layer. However, the depth extent of the channel depends on the wave frequency and this allows us to discriminate between different layers in the Earth. Unfortunately the dependence on frequency is a gradual one and we are unable to obtain information on the sharpness of discontinuities. Because of their limited depth penetration, surface waves are very useful for upper mantle studies.

If the earth is locally faster, the phase velocity is higher, and the phase of the surface wave is advanced. By measuring phase differences of surface waves we obtain information about lateral changes in S-velocity inside the Earth. If we do this for several frequencies, we obtain information about depth variations as well (Figure 3). More sophisticated techniques attempt to fit the complete waveform by changing the velocities in a starting Earth model (Nolet, 1990).

TRAVEL TIME RESIDUALS

It takes a ray ds/v seconds to traverse a distance of ds km (Figure 4). We find the total travel time by integrating along the path:

$$T = \int_{path} \frac{ds}{v} \tag{3}$$

Despite its simplicity, this is actually a complicated equation: we avoided the question of how to specify the path. This is not much of a problem if the Earth is layered, but the calculation of long ray paths in a realistic Earth is difficult and tedious and sometimes impossible with our current computing capabilities. Fortunately, T in (3) is not strongly in error if we make small deviations from the 'true' path. In most computations, seismologists therefore replace the true path by one calculated for a layered Earth. In some cases this can be dangerous. Figure 5, for example, shows that such rays can bend significantly from the true raypaths if the velocity changes are large (10%).

To determine whether the Earth's seismic velocity along a particular path is on average fast or slow, we compare the observed arrival time of the wave with the predicted travel time. Figure 6 shows how the predicted arrival time for the P wave differs 0.3s from the time picked by the observer. The travel time *residual* is found by subtracting from (3) the travel time for a reference Earth with velocity v_0. If we approximate and use the same path for both integrals, and do the Taylor expansion for $v^{-1} = (v_0+\delta v)^{-1} \approx -v_0^{-2}(1-\delta v)$ we find a linear dependence between the residual $\Delta T \equiv T - T_0$ and the velocity anomaly δv:

$$\Delta T = \int_{path} \frac{ds}{v_0 + \delta v} - \int_{path} \frac{ds}{v_0} \approx -\int_{path} \frac{1}{v_0^2} \delta v ds \tag{4}$$

Fermat's Principle states that the seismic ray seeks the path that renders the smallest possible travel time. Rays therefore always bend away from velocity lows and towards

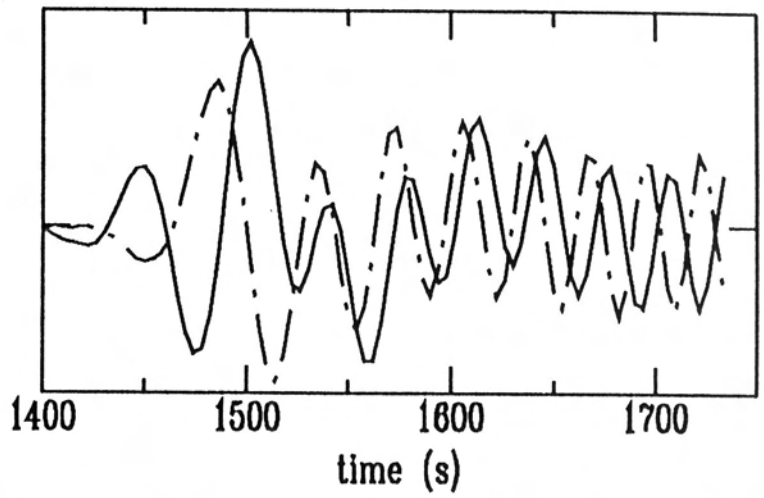

Figure 3: Observed (solid) and predicted phase of a surface wave. For the low frequencies at the start of the waveform, the phase delay is almost 180°, as it is at the end of the record. For intermediate frequencies, the difference is somewhat smaller. Since the low frequency reaches deepest, this tells us the real Earth's deep structure slows the wave down, as does the shallow structure, while the wave catches up at intermediate depth. A more quantitative analysis is able to identify the slow structures with the asthenosphere and the crust.

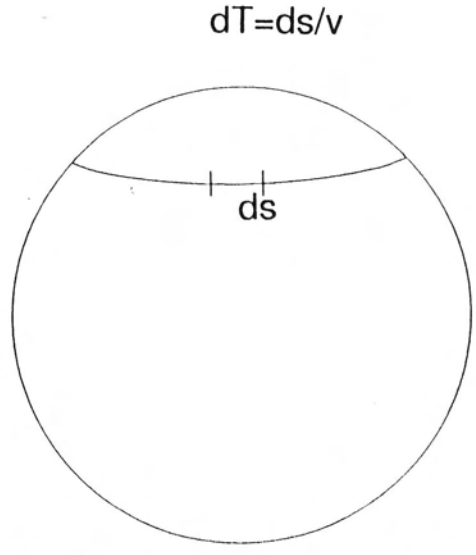

Figure 4: A ray inside the Earth can be split in little segments ds, over which the travel time is given by ds/v.

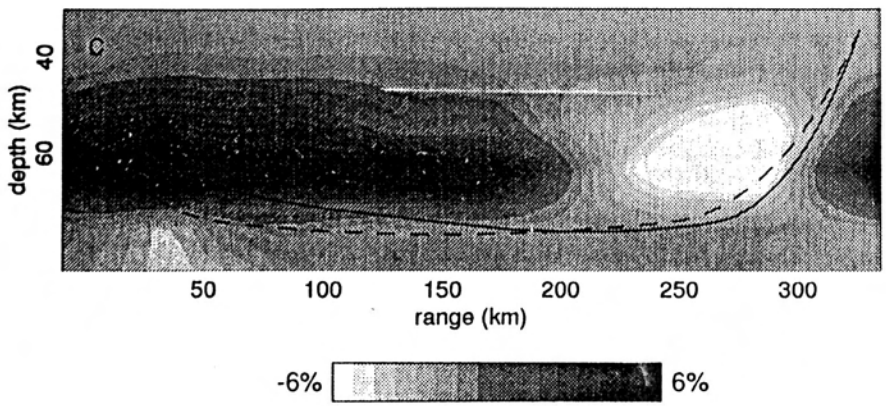

Figure 5: Velocity perturbations of ±6% bend the ray (solid line) away from its trajectory computed for an undisturbed Earth model (broken line). Even at short distances, such deviations can be of the order of 10 km vertically and much larger in the horizontal direction (after Papazachos and Nolet, 1995a).

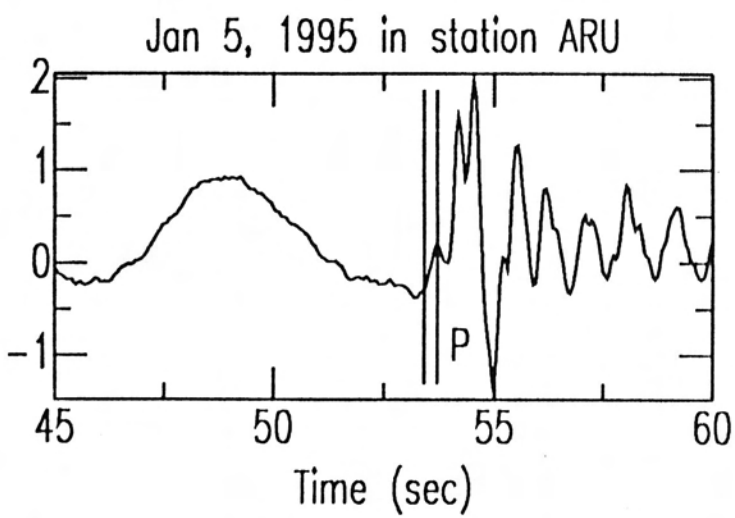

Figure 6: A P wave arrival. The leftmost vertical bar denotes the arrival time picked by the observer, the rightmost is the P arrival time as predicted for a reference earth model. The wave in front of the arrival is typical for microseismic noise.

velocity highs. The result is that we err towards high velocities when we use the path for the reference Earth instead of the true path. This bias has important practical consequences: a small magma body or a narrow plume will not show up as a negative travel time anomaly, but a cold (therefore fast) subduction zone does.

The errors in observed travel times are often of the same order of magnitude as the delays incurred along the path. This is especially true for older delay times, read from photographic recordings and reported to international agencies such as NEIC or ISC. Morelli and Dziewonski (1987) estimate that random errors have a standard error of 1.4s in teleseismic P residuals. Gudmundsson et al. (1990) find errors for P waves between 0.4 and 1.4s, and estimate the signal/noise ratio for ISC residuals to be about 2. The precision of the S wave arrival times is even worse: Vasco et al. (1994) estimate a standard deviation as high as 4.2s for ISC reported S waves. For digital recordings however, the reading error can be much reduced. Nolet et al. (1996) analyse the various error sources in travel times for digitally processed data and find that differences in observed travel time residuals are accurate to 0.1s, but that the absolute travel time variations are uncertain to 0.26s for P and 0.45s for S due to errors in the hypocentral location used to evaluate (3).

The classical way to interpret seismic travel times is to split the integral for (3) or (4) in time terms by splitting the path into segments near source and station and a (large) segment in between (Willmore and Bancroft, 1960):

$$\Delta T = \Delta T_{station} + \Delta T_{mantle} + \Delta T_{source} \tag{5}$$

The rationale behind the time term method is that delays incurred near the source or station will be present in every element of a set of delays for a particular source or station, respectively, and that the other delays will average out if we sum the delays over a whole set. In this way, for example, Dziewonski and Anderson (1981) used a time term analysis to determine station delays. Often we ignore the mantle delays with the assumption that they are much smaller than the source or station delays, but since the discovery of weak but extensive velocity anomalies in the lower mantle (Su et al., 1994), this practice could be disputed.

Figure 7 shows an application of the time-term method on S waves. The histograms present the S wave residual for rays from deep earthquakes after correction for ΔT_{source}, after dividing the stations in two classes: those on Precambrian, stable shields and those on Phanerozoic platforms and orogenic zones. On average, the travel time difference is 2.4s, but there is obviously a scatter well in excess of this signal. We shall try to analyze the cause of the high velocity under the shields in a later section of this paper. First we shall try to deal with the scatter by taking a less simplified approach.

SEISMIC TOMOGRAPHY

Equation (5) is a modest attempt to make inferences about the Earth's lateral

heterogeneity, by assigning a 'delay' to the lithosphere beneath every seismic station (the source term is more difficult to interpret, as it represents the combination of a real delay and the effects of earthquake mislocations). We can carry this further, and split the Earth into N regions or 'cells'. If ds_j is the path length of the ray within cell j, (4) reduces to a sum:

$$\Delta T = -\sum_{j=1}^{N} \frac{1}{v_0^2} \delta v_j ds_j \tag{6}$$

where δv_j is the velocity anomalie in cell j. If we have many observations ΔT_i, $i = 1, ..., M$, each with their own raypath, the path lengths in cell j will differ for every ray, and we should write ds_{ij}. Scaled with the factor $-1/v_0^2$ we can assemble the raylengths in a matrix, to obtain the linear system:

$$\mathbf{A}\delta\mathbf{v} = \Delta\mathbf{T} \tag{7}$$

where $\mathbf{A}$ is a $N \times M$ matrix with $A_{ij} = -ds_{ij}/v_0^2$ and $\Delta\mathbf{T}$ contains the travel time residuals. There is one row in $\mathbf{A}$ for every ray, and one column for every cell.

There are alternative methods to obtain a linear system such as (7) that do not depend upon a parametrization of the Earth into cells, but these need not concern us here as we shall wish to study the generic problems associated with solving such large linear systems. The following in this section will require some knowledge of linear algebra from the reader. For those who prefer to dodge the math and skip to section (3.3) we summarize the results first:

1. equation (7) cannot be solved exactly because of data errors, and we have to seek solutions that satisfy (7) 'as well as possible' in some sense.

2. (7) cannot be solved exactly because $\mathbf{A}$ is too large to fit inside a computer memory, and we have to use algorithms that approximate the solution.

3. there are many different models $\delta\mathbf{v}$ that satisfy the data equally well and we have to make a choice for the resulting image. Such a choice can be the model that is closest to the reference model, or the one that is smoothest in some sense. Whatever we do, we have to be aware of artifacts in the model that are a consequence of insufficient sampling of the subsurface.

If $M = N$, and $\det \mathbf{A} \neq 0$, the system (7) allows for an unique solution of the model $\delta\mathbf{v}$. We should never be so foolish to attempt to do this, though! Every good experimenter tries to reduce the influence of errors in her observations by averaging over many measurements. The case of seismic tomography is no exception, and we shall always try to have many more data than unknowns (i.e. $M \gg N$). With luck, we can even obtain rays sampling the same path and get a direct idea of our measurement accuracy by computing averages and standard deviations over this path - a technique in which the averages are known as 'summary rays'. Of course, if we

have more data than unknowns, the model $\delta\mathbf{v}$ can never be chosen such that we satisfy every measurement except in the unlikely case that there are no errors. Rather, we wish to minimize the discrepancy between the left and right hand side of (7). A common yardstick for the discrepancy is the squared length $\mathcal{L}$ of the vector that makes up the difference:

$$\min \mathcal{L} = \min |\mathbf{A}\delta\mathbf{v} - \Delta\mathbf{T}|^2 \tag{8}$$

A well known argument leads from this to Gauss' least squares equations (Figure 8). The range $\mathcal{R}$ of $\mathbf{A}$ is made up of all possible outcomes of the multiplication of $\mathbf{A}$ with an arbitrary N-dimensional vector $\mathbf{x}$. Since x_k weighs the k-th column of $\mathbf{A}$, the range $\mathcal{R}$ is a subspace made up of the (M-dimensional) columns. If all columns are independent, the dimension of $\mathcal{R}$ is N, since we have N columns. It may be less than N, but in any case it is less than M, so that the data vector may stick out of the subspace into M dimensional space (Figure 8). The closest we can get to $\Delta\mathbf{T}$ with a vector in $\mathcal{R}$ is by choosing $\delta\mathbf{v}$ such that the difference vector $\mathbf{A}\delta\mathbf{v} - \Delta\mathbf{T}$ is perpendicular to $\mathcal{R}$. This means that the inner product $(\mathbf{Ax}, \mathbf{A}\delta\mathbf{v} - \Delta\mathbf{T}) = 0$ for all $\mathbf{x}$. Using the identity $(\mathbf{Ax}, \mathbf{y}) = (\mathbf{x}, \mathbf{A}^T\mathbf{y})$, where $\mathbf{A}^T$ is the transpose of $\mathbf{A}$:

$$(\mathbf{x}, \mathbf{A}^T[\mathbf{A}\delta\mathbf{v} - \Delta\mathbf{T}]) = 0 \tag{9}$$

Since this is zero for all vectors $\mathbf{x}$, the right side of the inner product should be identical to 0, or:

$$\mathbf{A}^T\mathbf{A}\delta\mathbf{v} = \mathbf{A}^T\Delta\mathbf{T} \tag{10}$$

which are known as the normal equations. Since the normal equations have M unknowns and M data, we know a solution can be found as long as $\det \mathbf{A}^T\mathbf{A} \neq 0$.

DAMPING

Unfortunately, the determinant of $\mathbf{A}^T\mathbf{A}$ in (10) is usually 0 (or very small). That means that there are nonzero vectors $\mathbf{x}$ with $\mathbf{Ax} = 0$ (or negligible). Such vectors span the nullspace $\mathcal{N}$. In that case we have many solutions that give an equally good fit to (7). For, if $\delta\mathbf{v}$ is a solution that minimizes $\mathcal{L}$, $\delta\mathbf{v} + \mathbf{x}$ will fit equally well if $\mathbf{x}$ is in $\mathcal{N}$. This is most evident if there is a cell in our model that is not visited by any ray. We can assign a velocity change in this cell without affecting the data fit. More subtle insufficiencies arise when several cells are visited by only one ray or by a set of almost parallel rays, so that rows of $\mathbf{A}$ are not independent.

If more than one model fits the data, which one do we choose? *This turns out to be the major difficulty in seismic tomography, and there is no 'best' answer.* A common strategy that one will often find is to minimize the norm $|\delta\mathbf{v}|^2$ of the model. This is done by suppressing any components from the nullspace. Usually, we go further and also 'damp' the contribution of those in components $\delta\mathbf{v}$ that have a nonzero but small influence on the data vector. Such 'damping' biases the solution towards the reference model. There is nothing wrong with that as long as the reference model

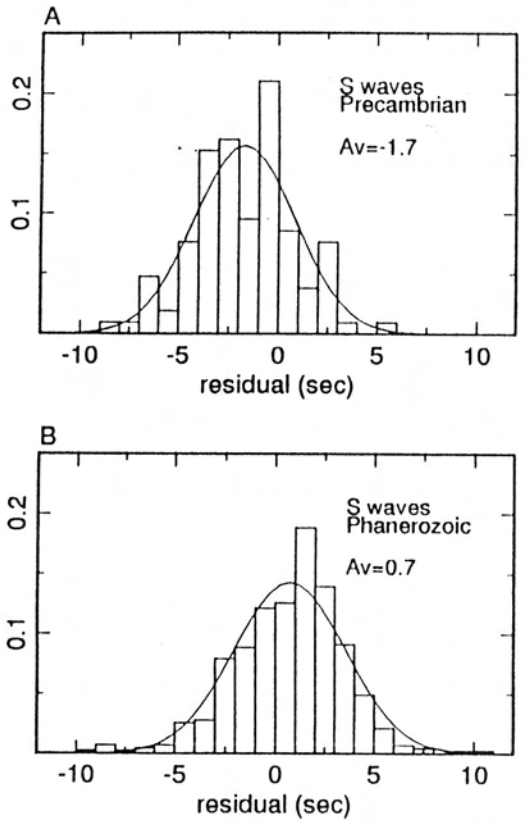

Figure 7: Normalized histograms of observed residuals for S waves, with a best fitting Gaussian probability density function superposed. The systematic difference with the geological age of the station location warrants the use of time terms.

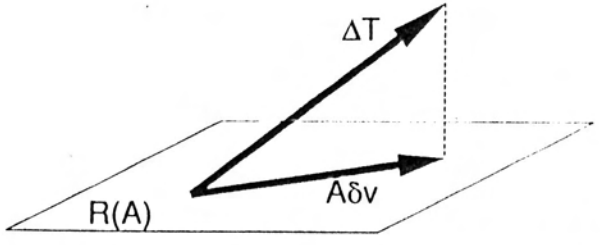

Figure 8: The range $R(A)$ or $\mathcal{R}$ of the matrix $\mathbf{A}$ consists of all vectors of the form $\mathbf{Ax}$, and therefore must also contain the predicted data vector $\mathbf{A}\delta\mathbf{v}$ for the Earth $\delta\mathbf{v}$. We want this vector to be as close as possible to the observed data vector $\Delta\mathbf{T}$. This closest approximation is obtained by projecting the observed data onto $\mathcal{R}$ The difference vector $\mathbf{A}\delta\mathbf{v} - \Delta\mathbf{T}$ (broken line) must then be orthogonal to $\mathcal{R}$.

is a reasonable model. The minimum norm model can then be interpreted as the minimum excursion away from the reference model that is needed to satisfy the data. For an interpretation of velocity anomalies in terms of mineralogy, this not a very useful approach, though. For example, damping may result in a velocity anomaly of 5% to be reduced to 3%. In section 4 we shall see that such inaccuracies prohibit us to distinguish between effects of temperature and changes in the abundance of various minerals.

ITERATIVE SOLUTIONS

In practice, computers are not powerful enough to solve (7) exactly if N and M are larger than, say, 10^3. We use iterative techniques that refine a solution at every step to give an improvement to the fit. There are a number of such iterative techniques. At each step, an improvement to the model is sought that further reduces the misfit to (7). Suppose that after k iterations we have arrived at a model $\delta\mathbf{v}_k$. We can then set up a new system

$$\mathbf{A}(\delta\mathbf{v}_{k+1} - \delta\mathbf{v}_k) = \Delta\mathbf{T} - \mathbf{A}\delta\mathbf{v}_k = \mathbf{r}_k, \tag{11}$$

where $\mathbf{r}_k$ is the residual after k iterations. This equation is not likely to be solvable exactly, so we take a least squares approach to find the best fit.

Most schemes work directly on (11) without explicitly computing $\mathbf{A}^T\mathbf{A}$. This is done by searching in a sequence of directions $\mathbf{p}_0, \mathbf{p}_1, ..., \mathbf{p}_k$. For the new updated model we then use: $\delta\mathbf{v}_{k+1} = \delta\mathbf{v}_k + \alpha\mathbf{p}_k$, and solve:

$$\mathbf{A}(\alpha_k\mathbf{p}_k) = \mathbf{r}_k.$$

We now need to find the scalar α_k that minimizes $|\mathbf{A}(\alpha_k\mathbf{p}_k) - \mathbf{r}_k|^2$. Differentiating this expression and setting it to zero yields α_k:

$$\alpha_k = \frac{(\mathbf{A}\mathbf{p}_k, \mathbf{r}_k)}{(\mathbf{A}\mathbf{p}_k, \mathbf{A}\mathbf{p}_k)}$$

We see that only multiplications with $\mathbf{A}$ are needed, so these schemes are very useful for large matrices.

The simplest iterative methods take the rows of $\mathbf{A}$ as search directions. It is easily seen that this constructs a solution outside of the nullspace (since the product of a row with itself as a search direction gives the squared norm of that row, which is nonzero). However, this slows down convergence since an update at iteration k may partly undo the reduction in misfit obtained at an earlier stage. Conjugate gradient methods, of which LSQR (Paige and Saunders, 1982) is judged the most stable, search in orthogonal directions outside of the nullspace, to avoid interference with earlier iterations.

Even for very large systems with $N \sim 10^5$, ten or twenty iterations are often sufficient to bring $\mathbf{A}\delta\mathbf{v}$ close enough to $\Delta\mathbf{T}$, i.e. within a tolerance that is reasonable in view of the accuracy of the residuals.

RESOLUTION AND ARTIFACTS

While the minimum norm solution is useful for some studies, it may be a misleading solution when we use the tomographic image for structural studies, as it forces the model back to the reference velocity whenever possible, e.g. when there is little information. This introduces false structure, or 'artifacts' into the model. One obvious example is for the case of rays. A search direction parallel to row i of the matrix will only change the model cells visited by the seismic ray belonging to the ith datum. Experience shows that models tend to concentrate around the regions well sampled by raypaths. Efforts to correct for this effect have been made, but easily result in unstable images. See Papazachos and Nolet (1995a) for a recent study on this topic.

How does one judge what is real and what is artifact? Seismologists have come up with a number of 'sensitivity tests' to investigate the reliability of the solution to (7). One may test the sensitivity of the system to data errors by feeding it random noise with zero average. In other words, instead of solving (7) for the observed data vector $\Delta\mathbf{T}$, we substitute an artificially generated data vector consisting of random numbers. If the data averaging involved in the normal equations (10) is adequate, the final model should remain very close to 0. This experiment can be repeated with different random data to identify regions in the model that are sensitive to propagating errors. Another common test is to generate a synthetic data vector from a known model (by multiplying $\mathbf{A}$ with the known model, and then solving (7) to see how well we can reconstruct our model. Figure 10 shows how we reconstruct a model consisting of spikes. Often, we find that tomographic images do a good job of reconstructing the locations of such spikes, but may only reconstruct 50% or less of the amplitude. This, of course, is a direct consequence of our choice of minimum norm for the model. Thus, while such cross-sections are a useful tool for delineating subsurface structure, extreme care must be taken when trying to interpret the amplitudes of the perturbations in terms of local temperature or composition.

We often find that sensitivity tests give better results as we increase the size of the objects in the test model. This is a well known result originally derived by Backus and Gilbert in their pioneering studies on inversion. For large systems such as we encounter in seismic tomography, no one has yet come up with a feasible method to determine analytically how large the averaging volumes should be, but by experimenting with different spike sizes we can usually get a decent idea.

MINERAL COMPOSITION AND SEISMIC VELOCITIES

Can we interpret the measured seismic velocities in terms of the Earth's chemistry and mineralogy? The question goes a long way back, but usually Birch (1952) is credited with the first systematic attempt to extrapolate laboratory measurements at room pressure and temperature to ambient mantle conditions. In the forty years since then, many geophysicists have tried to test global Earth models against the

Table 1: Velocity/density derivatives (1200°C,4 GPa)

$\partial V_P/\partial T$	-5.1×10^{-4}	km/s/°K
$\partial V_S/\partial T$	-3.5×10^{-4}	km/s/°K
$\partial\rho/\partial T$	-0.6×10^{-4}	g/cm^3/°K
$\partial V_P/\partial X_{opx}$	-0.36	km/s
$\partial V_S/\partial X_{opx}$	-0.02	km/s
$\partial\rho/\partial X_{opx}$	-0.03	g/cm^3
$\partial V_P/\partial X_{cpx}$	-0.11	km/s
$\partial V_S/\partial X_{cpx}$	-0.11	km/s
$\partial\rho/\partial X_{cpx}$	0.03	g/cm^3
$\partial V_P/\partial X_{gt}$	0.66	km/s
$\partial V_S/\partial X_{gt}$	0.34	km/s
$\partial\rho/\partial X_{gt}$	0.37	g/cm^3
$\partial V_P/\partial\gamma$	-0.10	km/s
$\partial V_S/\partial\gamma$	-0.03	km/s
$\partial\rho/\partial\gamma$	-0.06	g/cm^3

available data on elastic properties of minerals, but - as noted by Duffy and Anderson (1989) who also give a large number of references - a consensus has failed to emerge. In this section we illustrate some of the difficulties.

At seismic frequencies, the relevant elastic moduli are the adiabatic ones. If the geotherm was adiabatic from the surface, the pressure derivatives $\partial\kappa/\partial P$ and $\partial\mu/\partial P$ for the bulk and shear modulus would suffice. In the real Earth, the true surface temperature is far from the value one would get by extrapolating an adiabatic geotherm all the way up to the surface (this *potential temperature* is often estimated to lie between 1200° and 1400°C), and consequently we need both the pressure *and* the temperature derivative. As long as the pressure in the Earth is a small fraction of the incompressibility, a linear extrapolation with pressure is permissible (Stacey et al., 1981). Since the pressure at the bottom of the upper mantle (24 GPa) is less than 10% of the value of κ we do not have to be picky in choosing a finite strain theory, which takes the nonlinear behaviour with pressure into account. The extra uncertainty that the adoption of a particular finite strain theory entails plays no role for the upper mantle.

This, however, is about the only good news. The interpretation of laboratory data is hampered by a lack of data on pressure and temperature derivatives for the elastic moduli. Moreover, the temperature in the upper mantle is uncertain by several hundred degrees. Finally, the question has been raised whether the data on elastic moduli, which are determined at ultrasonic frequencies, are really applicable at seismic frequencies (Karato, 1993).

Table 1 lists a few important derivatives. These were computed using third order

finite strain theory and the laboratory data listed in Duffy and Anderson (1989), supplemented with some new and revised values (Duffy, pers. comm., 1994). We used a fairly standard composition for the upper mantle with 57% Olivine (ol), 17% Orthopyroxene (opx, Enstatite), 12% Clinopyroxene (cpx, a 1:1 mix of Diopside and Jadeiite), 14% Garnet (gt), and the ratio Fe/Mg=0.1. For this standard composition the laboratory data predict at a potential temperature of 1200°C that $V_P = 8.19$ km/s, $V_S = 4.63$ km/s and $\rho = 3.40$ g/cm^3. These values are high with respect to globally averaged seismic models like PREM, because they do not reflect any effects of partial melt or shear modulus weakening in the asthenosphere. The compositional derivatives are at the expense of Olivine, e.g. if X_{gt} (garnet) increases, X_{ol} decreases by similar amount. The values in the table are valid at pressure of 4 GPa (120 km depth). Changes with depth are not fully negligible but the values given here are sufficiently accurate for schematic analyses. The last parameter in this table is a depletion factor γ, which is explained in more detail in section (4.1).

EXAMPLE: THE TECTOSPHERE

We shall first give a simple example to illustrate the difficulties we encounter when trying to make a mineralogical interpretation of seismic data. Jordan (1975, 1978) proposed the existence of deep, cold keels or 'tectospheres' under the older parts of the continents to explain the large differences observed in the travel times of S waves. The increase in density because of cooling is offset by a decrease in density due to the depletion of basaltic components - otherwise the Archean cratons would sink several km below sea level.

Figure 7 shows histograms for Δt_S peaking at -1.7 s for the older cratons, and at +0.7 s for younger continental regions, a difference of 2.4 s. How does this constrain the properties of the tectosphere? There are obviously too many parameters in Table 1 to arrive at a unique answer. To reduce the number of parameters, we model the depletion with a parameter γ, such that $\gamma = 0$ yields the standard composition, and $\gamma = 1$ corresponds to a removal of all clinopyroxene and garnet from this model:

$$\begin{aligned} X_{ol} &= 0.57 + 0.20\gamma, \\ X_{opx} &= 0.17 + 0.06\gamma, \\ X_{cpx} &= 0.12 - 0.12\gamma, \\ X_{gt} &= 0.14 - 0.14\gamma. \end{aligned}$$

For a tectosphere of thickness H km, the gravitational equilibrium demands that:

$$\frac{\partial \rho}{\partial T} H \Delta T + \frac{\partial \rho}{\partial \gamma} H \gamma = 0, \tag{12}$$

while the travel time discrepancy should satisfy:

$$\Delta t_S = \Delta \left(\frac{H}{V_S} \right) = -\frac{H}{V_S^2} \left(\frac{\partial V_S}{\partial T} \Delta T + \frac{\partial V_S}{\partial \gamma} \gamma \right). \tag{13}$$

The first equation gives us directly a ratio between the temperature difference and the depletion factor: $\Delta T/\gamma = -0.06/0.6 \times 10^{-4} = 1000$. In other words, for a temperature difference of 500°, the depletion must be about halfway. If we use the numbers of Table 1, we find the following solution for the products of ΔT and γ with H that occur in the constraints (12) and (13):

$$\begin{aligned} H\Delta T &= -1.61 \times 10^5 \\ H\gamma &= 161 \end{aligned}$$

Again, if $\gamma = 0.5$, we would get $H = 322$ km, not an unreasonable thickness in view of the tomographic results by Su et al. (1994). However, the resolving power of global tomography studies are such that the depth of the cratonic roots cannot be determined within 100-200 km, and more local studies are limited as well. If we add Δt_P, we merely add a third constraint for the variables, and have to solve the system using least squares (adding $\Delta t_P = -0.7$, as observed by Nolet et al. (1996), changes $H\gamma$ to 154). We are therefore forced to look for other evidence if we wish to determine H more precisely.

Such evidence is likely beyond the scope of seismology. For example, the pyroxene geotherm of Boyd (1973) is located about 500° below the solidus for dry peridotite. If we impose this as a maximum for ΔT over the full depth range H (which is questionable), the 322 km for H is a minimum thickness for the tectosphere. If the magnitude of ΔT is smaller, H has to increase. This is in fact the more likely scenario, but it leads to a paradox: only for low T is the viscosity of the rock low enough to allow it to resist shear tractions and translate with the continent. To resolve the paradox one might try to change the composition of the standard model, only to find that the influence of composition is fairly minor. To explain the magnitude of Δt_S, we must take recourse to other effects that influence the velocity, such as partial melt, shear modulus weakening with temperature or anisotropy.

EXAMPLE: THE HELLENIC ARC

Figure 9 shows a cross-section through the structure of the Hellenic arc as derived by Papazachos and Nolet (1995b). The dark regions denote high P velocity, and a (thin) slab can be seen subducting beneath Peloponnesus. The acretionary prism is characterized by P velocities that are 4% lower than those in the background model, while the high velocity of the slab exceeds background V_P by 3-4%. Because this image is damped (Figure 10), we can take this difference of 7-8%, or about 0.50 km/s, as a minimum. We could wonder if this order of magnitude is caused by the effects of temperature alone. Table 1 shows that this requires a temperature difference of the order of $0.50/5.1 \times 10^{-4} \approx 1000°$ between the prism and the slab, if caused by effects of temperature alone. This would bring the temperature in the crust well above 1000°C, which is unrealistic.

The positive velocity anomaly in the slab of about 0.25 km/s is explainable by a negative temperature anomaly of 500° with respect to background model temperature

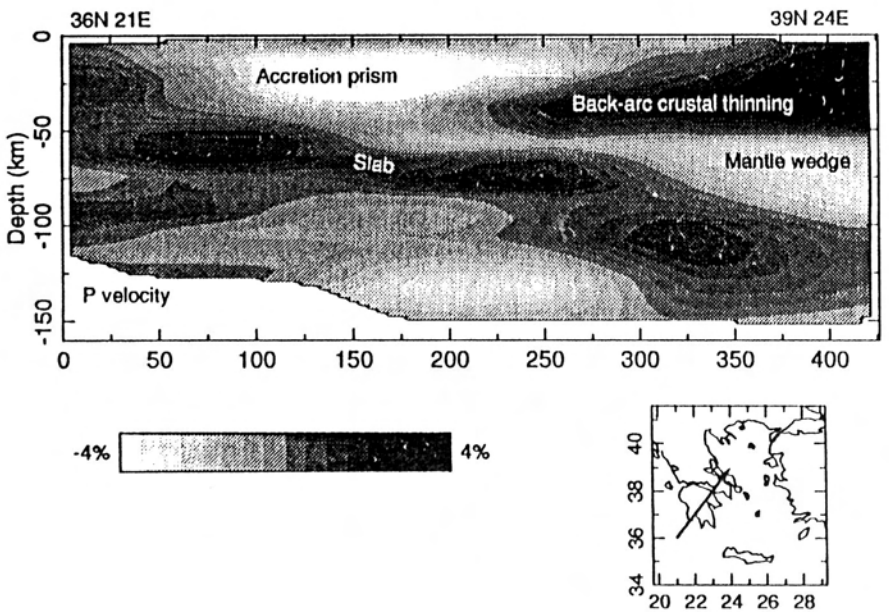

Figure 9: Cross section through the P-velocity model for the Hellenic Arc derived by Papazachos and Nolet (1995b). The velocity difference between the low velocities in the acrretionary prism and the slab are of the order of 6% and cannot be explained by a simple temperature difference (see text).

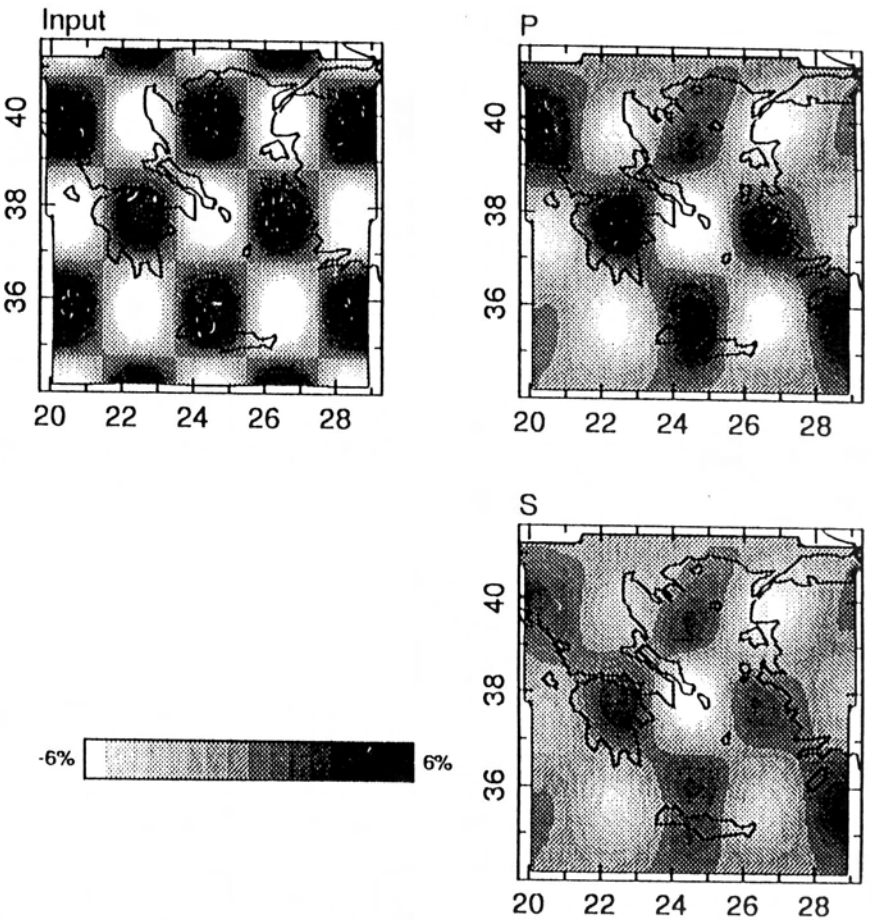

Figure 10: A sensitivity test for the data set of P and S waves for the shallow structure in the Hellenic Arc used by Papazachos and Nolet (1995b). A 'synthetic' data vector ΔT^{syn} was created by multiplying the input model (upper left) with A. This was then inverted for a model δv. cross sections at a depth of 70 km for the P and S solutions are shown in the figures on the right. One can see that the velocity deviations are well reproduced for the P velocity along the western part of the arc. Elsewhere the amplitudes are strongly reduced, as are the amplitudes of the S velocity perturbations.

at 60 km depth. We cannot rule out, however, that compositional differences cause part of the anomaly. Eclogite is rich in high velocity garnet. A similar uncertainty is involved in the interpretation of the low velocity anomaly of about -0.25 km/s in the prism. Here, a pure temperature effect (500°) seems unlikely.

Obviously, if we could reliably measure ΔV_S we would be in a better position to discriminate since the compositional effects on V_S are less strong than the temperature effects. However, this approach meets with additional difficulties. As Figure 10 shows, the S velocity is not reliably determined. Papazachos and Nolet (1995b) find S velocity anomalies of -3% or -0.14 km/s, but this is probably a low value because of the damping, and may perhaps be only half the true anomaly, making a quantitative inference for compositional effects useless. In addition, we must realize that the values in Table 1 are only valid well below the solidus. Especially the S velocity decreases sharply when the temperature approaches the solidus. A rapid growth in crystal imperfections causes the wave to attenuate and slow down. The effect is dispersive, i.e. the velocity decrease depends on the frequency of the wave, and is strongest for the low frequency waves (Karato, 1993).

In some cases we observe a decrease in S velocity that cannot reasonably be explained as the effect of temperature alone, in which case we are able to conclude that the rock is close to the solidus, or even that partial melt is present. For example, Nolet and Zielhuis (1994) argue that dehydration in subduction regions is able to reduce the solidus of the ambient mantle at depths as large as 300 km. At the onset of partial melting the S velocity may decrease dramatically, especially if the melt is able to wet grain boundaries.

CONCLUSIONS

We have given a brief introduction into seismic tomography, and discussed one of its main weak points: while it may give a good impression of structural features at depth, the velocity anomaly amplitudes are often not well enough resolved. As two examples show, a purely seismological interpretation can never be definitive. Additional information on composition, temperature, or on petrological processes like basalt depletion, are needed to constrain the number of possible interpretations.

REFERENCES

Birch, F., Elasticity and constitution of the Earth's interior, *J. geophys. Res, 57*, 227-285, 1952.

Boyd, F.R., A pyroxene geotherm, *Geochim. Cosm. Acta, 37*, 2533-2546, 1973.

Duffy, T.S. and D.L. Anderson, Seismic velocities in mantle minerals and the mineralogy of the upper mantle, *J. geophys. Res., 94*, 1895-1912, 1989.

Dziewonski, A.M. and D.L. Anderson, Preliminary reference Earth model, *Phys. Earth Plan. Int., 25*, 297-356, 1981.

Gudmundsson, O., J.H. Davies and R.W. Clayton, Stochastic analysis of global travel time data: mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int., 102*, 25-44, 1990.

Jordan, T.H., The continental tectosphere, *Rev. Geophys. Space Phys., 13*, 1-12, 1975.

Jordan, T.H., Composition and development of the continental tectosphere, *Nature, 274*, 544-548, 1978.

Karato, S.-i., Importance of anelasticity in the interpretation of seismic tomography, *Geophys. Res. Lett., 20*, 1623-1626, 1993.

Morelli, A. and A.M. Dziewonski, Topography of the core-mantle boundary and lateral homogeneity of the liquid core, *Nature, 325*, 678-683, 1987.

Nolet, G., E. Richardson, K. Lindenberg and E. Zanzerkia, Global travel time determinations using crosscorrelation, m/s in preparation.

Nolet, G. and A. Zielhuis, Low S velocities under the Tornquist-Teisseyre zone: evidence for water injection into the transition zone by subduction, *J. Geophys. Res., 99*, 15813-15820, 1994.

Paige, C. C. and M. A. Saunders, LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Soft., 8*, 43-71 and 195-209, 1982.

Papazachos, C. and G. Nolet, Obtaining robust non-linear solutions in arrival time tomography, subm. for publ., 1995a.

Papazachos, C. and G. Nolet, P and S deep velocity structure of the Hellenic area obtained by robust non-linear inversion of travel times, subm. for publ., 1995b.

Stacey, F.D., B.J. Brennan and R.D. Irvine, Finite strain theories and comparisons with seismological data, *Geophys. Surveys, 4*, 189-232, 1981.

Su W.-J. , R.L. Woodward and A.M. Dziewonski, Degree 12 model of shear velocity heterogeneity in the mantle, *J. Geophys. Res. 99*, 6945, 1994

Willmore, P.L. and A.M. Bancroft, The time term approach to refraction seismology, *Geophys. J. R. astr. Soc., 3*, 419-432, 1960.