Constraints on Mantle Convection from Seismic Tomography and Flow Modeling

by

Hrafnkell Kárason
C.S. Mechanical Engineering
University of Iceland, 1994

Submitted to the Department of Earth, Atmospheric and Planetary Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2002

Author .....
Department of Earth, Atmospheric and Planetary Sciences
March 4, 2002

Certified by .....
Robert D. van der Hilst
Professor
Thesis Supervisor

Certified by .....
Bradford H. Hager
Professor
Thesis Co-Supervisor

Accepted by .....
Ronald G. Prinn
Department Head, Department of Earth, Atmospheric and Planetary Sciences

Abstract

In this thesis I combine high resolution seismic tomography and realistic flow modeling to constrain mantle convection. The bulk of the data used in the tomographic imaging are millions of P, pP and pwP travel time residuals, read from high frequency seismic records. The distribution of earthquakes and stations results in uneven data coverage and to improve the model I use core phases (PKP, P_{diff}) for the deep mantle and surface reflected phases (PP) for the shallow mantle. Since narrow rays are not adequate for low frequency measurements, I construct broad 3-D sensitivity kernels to relate some of the added data to earth's structure. Furthermore, the parameterization of the tomographic model is adapted to data density and the model shows, among other details, the varying style of subduction in the shallow mantle and complex flow patterns around the transition zone between the upper and lower mantle. I develop a novel and efficient method of modeling buoyancy driven mantle flow in spherical geometry. Here, the linear Stokes equation is solved using a Green's function approach and 3-D surfaces, representing the boundaries of dense material, such as subducting slabs, are tracked through time. I describe the basis of the approach, represent tests, and investigate the thickening and slowing of subducting slabs for different viscosity contrasts between the upper and lower mantle. I compare seismic tomography, flow modeling and reconstructed plate motions associated with the collision of India to mainland Asia, in particular, the south and south-west motion of the Sunda block. This way I can evaluate subduction scenarios and make quantitative comparison between tomography and plate reconstructions. I conclude that a viscosity increase of $\sim$200 in the lower mantle, resulting in $\sim$10 times higher viscosity than previously estimated, best fits the observations. Juxtaposing the tomography, the flow models and inversion tests using the flow models as input, indicates that the results are robust.

Acknowledgments

Having spent more than 5 years at MIT I have much to acknowledge. First, I would like to thank Professor Rob van der Hilst for accepting a mechanical engineer with little experience in geosciences to the doctoral program in geophysics to work on seismic tomography. Rob has been and remains an insightful advisor and a very good friend. I thank Professor Brad Hager, for introducing me to Stokeslets and for our many discussions where his intuition has quickly led me to the right path.

Frederik Simons and I were the first two students of Rob at MIT. We have been officemates and friends for five and a half years, which not only has made the long days in front of the computer bearable, but also brought us to San Francisco, Paris, Reykjavík, Düsseldorf, Antwerp and Berlin.

I treasure deep, as well as shallow, discussions with my friends Oded Aharonson and Sergei Lebedev at different hours of the day. I thank my fellow students Mark Behn, Margaret Boettcher, Jun Korenaga, Maureen Long, Laurent Montesi, Eliza Richardson, Rebecca Saltzer and Phil Tracadas for their inspiration and friendship over the years.

When Brynhildur and I moved to Boston we were expecting our first child and while leaving our families behind was hard, it was a great comfort to have Snorri and Þórdís, my wife's brother and his wife, living within an hours drive. Erla was born just before our first Christmas in the U.S. and two years later, our younger daughter, Iðunn was born. Then followed the first words, bipedalism, countless hours at the playground and, most recently, phonetics, all of which I am very thankful for.

Elsku Brynhildur, ástin mín, ástin mín, þakka þér allt og allt ... Þinn Hrafnkell

Contents

List of Figures

Introduction

Mantle convection is the mechanism by which heat is transferred from Earth's deep interior to its surface. The upper thermal boundary layer of the convective mantle is a rigid, but broken shell, made up of different sized tectonic plates. As a result of the convective flow, these plates move with respect to each other, with most of the translation or rotation being accommodated at the plate boundaries [*Davies and Richards*, 1992]. At ridges, plates diverge, exposing hot upwellings from the interior and creating new sea-floor. Where plates move past each other they create strike-slip faults, which commonly are very seismically active. The plates are destroyed as the old, cold and negatively buoyant oceanic lithosphere (i.e. slabs) sinks at convergent margins in a process called subduction. The plates are not passive and their geometry and mechanical strength may exert an influence on the mode of convection within the mantle [*Conrad and Hager*, 1999; *Anderson*, 2001]. Furthermore, the negative buoyancy of the cold downwellings is the dominant driving force behind plate motion and mantle flow in general [*Hager and O'Connell*, 1981], which is manifested in the rapid motion of plates that are "attached" to a slab vs. other plates [*Forsyth and Uyeda*, 1975]. These plate motions at the surface can be reconstructed back in time, giving valuable constraints on the sinking rates of slabs and mantle flow in general. The highly viscous silicates that constitute Earth's mantle deform very slowly in response to the large convective forces, but on short timescales they behave as a solid and their elastic response allows seismic waves to propagate through the mantle. Since the elastic moduli are sensitive to temperature, seismic waves can be used to map the lateral temperature gradients caused by the convection. This is done by a technique known as seismic tomography, which produces

the most direct depictions of mantle convection inside the Earth [*Montagner*, 1994]. During the last three decades, advances in seismological theory, combined with the buildup of larger data sets and increasing computing power, have allowed researchers to make ever more detailed images of Earth's interior structure [*Kárason and van der Hilst*, 2000, see Chapter 1]. However, because of the very slow flow the tomographic images only reveal the present-day flow patterns. The main objective of this thesis is to constrain better the state and nature of mantle convection, by improving the tomographic models themselves as well as their interpretation, by relating the images to plate motion history, using geodynamical modeling.

A travel-time residual is the deviation between a predicted and observed arrival time of a seismic phase (e.g. P and pP). We invert millions of such residuals from the catalogue of *Engdahl, van der Hilst, and Buland* [1998], to create a tomographic model. However, due to the uneven distribution of earthquakes and stations and the ray geometry of the seismic phases involved, the sampling of Earth's interior structure is very uneven using these data. We improve our model of the deep mantle by using differential travel times of core-refracted and -diffracted phases (PKP_{DF}-P_{diff} [*Wysession*, 1996] and PKP_{AB}-PKP_{DF} [*McSweeney*, 1995; *Engdahl et al.*, 1998]) along with the travel time residuals of mantle P and pP waves. The P_{diff} data used are measured at low frequencies, which means that high frequency approximations (i.e. rays) used for the other data are not valid. Instead, to adequately relate those data to Earth's structure, we construct broad sensitivity kernels. The resulting tomographic model shows different convective patterns at $\sim$1200 km depth and at the base of the mantle. This is further confirmed by calculating global radial correlation, which shows diminished continuity around mid-mantle depths. These and other observations prompted *van der Hilst and Kárason* [1999] and *Kellogg et al.* [1999] to propose the existence of a deep dense layer, perhaps with a highly undulating boundary [*Kárason and van der Hilst*, 2001, see Chapter 2].

However, P, PKP, P_{diff} do not help in large regions of shallow mantle that are devoid of sampling. This is especially true in the upper mantle beneath the seismically inactive intra-plate regions. To improve our global model we used ~20,000 PP-P differential travel times from Guy Masters [*pers. comm.*]. The low frequency (40 mHz) PP-P data are relatively insensitive to structure beneath source and receiver, but have a wide and complex sensitivity zone near the bounce point, where upper mantle heterogeneity is generally strong. For low frequency PP-P we account for sensitivity to structure away from the optical ray path with 3-D Fréchet derivatives (sensitivity kernels) estimated from single forward scattering. Inversion tests, spectral analysis, and comparison with geology indicate that the large-scale upper mantle structure is better constrained with the addition of PP-P [*Kárason et al.*, 2002, see Chapter 3].

As mentioned earlier, the sampling by seismic rays within the mantle is very uneven, and consequently the length scale at which structure can be resolved is spatially variable. In our most recent P-wave tomographic model, the mantle is imaged at varying length scales, i.e. the grid size of our block parameterization adapts to the density of data coverage (see also *Bijwaard et al.* [1998]). As before, we use broad sensitivity kernels, which are projected onto the fine grid, for the low frequency data, which this way can constrain long wavelength heterogeneity without keeping the short period data, for which we use linear raytracing, from mapping details in densely sampled regions. Non-linear re-raytracing through the model [e.g. *Bijwaard and Spakman*, 2000] might perhaps enhance the structure locally, but our focus has been on the data. The combination of data sets, integrated with each through appropriate kernels, along with the use of the adaptive grid arguably results in one of the best P-wave models currently available. In particular, detailed images of complex trajectories of convective flow in the upper mantle, across the transition zone and, in some cases, down to mid-mantle depths are obtained beneath many of the seismically active subduction zones. [*Rogers et al.*, 2001; *Kárason and van der Hilst*, 2002, see

Chapter 4].

To make realistic flow models that can relate subduction history and our detailed tomographic images we have developed a new code to simulate buoyancy-driven mantle flow in three dimensions, in spherical geometry and with high spatial ($<$15 km) and temporal resolution. Here, Stokes equation, which governs the viscous mantle flow, is solved efficiently using a Green's functions approach. Surfaces representing the boundaries of dense material, i.e. slabs, are tracked through time to build a flow model with depth dependent viscosity profiles of the mantle. At the surface, geometrically complex and/or time-varying boundary conditions are easily applied, facilitating integration with subduction history. The key advantage of this approach is that these simulations are very flexible and can be made relatively cheaply, using much less computer power than the traditional finite element based 3-D studies. The major drawback of this method is that viscosity has to be laterally uniform (i.e. no strong slabs). (see Chapter 5).

One important factor of mantle convection is the sinking rate of subducting slabs. The sinking rate decreases when the viscosity increases and assuming an approximately constant rate of subduction at the surface, the slab will thicken in a more viscous lower mantle if mass flux is to be preserved. Initially, the slab will thicken above the higher viscosity layer and slow down. Gradually, more mass will accumulate, the slab will thicken and the sinking velocity of the deepest part will increase until a balance between the mass flux in the upper and lower mantle is reached. This balance of thickening and sinking rate and their transient nature is is subject of one of our studies (see Chapter 6).

Comparing tomography to flow modeling inspired by subduction history and vice versa gives an opportunity to constrain both mantle structure and plate reconstructions. We apply this approach to tectonics in Southeast-Asia, i.e. the subduction associated with the South and South-West motion of the Sunda-block, caused by the collision of India with mainland Asia. The flow modeling allows us to dynamically connect plate reconstructions [*Replumaz*

and Tapponnier, 2001], which constrain surface position of the convergent margin at different times in the past, and the tomography, which constrains the morphology of the slabs at depth. We find that in order for our modeling to fit the observed shape of the slab in the tomography, we need a viscosity contrast of 200 between the upper and lower mantle. In addition, our estimate for lower mantle viscosity is about an order of magnitude higher than recent estimates from post-glacial rebound (see Chapter 7).

Chapter 1

Constraints on Mantle Convection From Seismic Tomography[1]

Abstract

Since the advent of global seismic tomography some 25 years ago, advances in technology, seismological theory, and data acquisition have allowed spectacular progress in our ability to image seismic heterogeneity in Earth's mantle. Using examples from some recently published tomographic models, we briefly review some concepts of seismic tomography, such as parameterization, and summarize how this class of imaging has contributed to our knowledge of the scale of mantle convection. With the presently available data seismologists can map the subduction of former oceanic lithosphere beneath most island arcs. Tomographic imaging has revealed that many slabs extend below the deepest earthquakes of the Wadati-Benioff zones and sink deep into the lower mantle, although intense deformation of flow trajectories has been detected in the upper mantle transition zone beneath several convergent margins (in particular in the western Pacific). We illustrate this with examples of our recent 3-dimensional model for mantle P-wavespeed, which is inferred from arrival time data of P, pP, PKP, and P_{diff} waves. In combination with constraints from computational geodynamics the seismic images render mantle stratification at 660 km depth unlikely and show that surface plate motions are tied to large scale convective circulation. However, the ultimate fate of the slabs that penetrate into the lower mantle is still enigmatic and in some

[1]This chapter was published as: H. Kárason and R. D. van der Hilst, Constraints on Mantle Convection From Seismic Tomography, in *History and Dynamics of Plate Motion*, Geophys. Monogr. Ser., vol. 121, 277–288, 2000.

recent tomographic models the subduction related heterogeneity pattern vanishes at very large depth (2000 km or so). These observations have inspired models for mantle convection that have the potential for reconciling geophysical views and geochemical constraints, but much work remains to be done to establish the detailed pattern of mantle flow.

1.1 Introduction

The relative motion of lithospheric plates at the surface of the Earth, which is directly related to natural hazards such as earthquakes and volcanoes and can cause long-term variations in climate, is the surface expression of slow, large scale deformation of rock in the deep interior of our planet. On a time scale of millions of years the mantle silicates flow in a process known as 'mantle convection', but on short time scales mantle material behaves as a solid. This viscoelastic behavior allows the propagation of longitudinal (P) and transverse (S) seismic waves, which speed up in "cold" downwellings and slow down in "hot" upwellings so that recordings of these waves can be used for the mapping of mantle flow. Our objective here is not to provide a rigorous review of seismic tomography but to highlight developments in seismic imaging pertinent to the scope of this special volume.

In the past half century, the scale of mantle convection, which relates critically to compositional stratification and the thermal and chemical evolution of our planet, has been one of the big puzzles in Earth Sciences. A major challenge has been to evaluate and reconcile the range of observations and constraints provided by different scientific disciplines [cf. *Albarède and Van der Hilst*, 1999]. Earth's heat budget (the balance between heat production and heat loss) and geochemical analyses of ocean floor basalts suggest that distinct mantle reservoirs have retained their identity for 2 billion years or more. One reservoir boundary is typically placed at 660 km depth, that is, between the upper and lower mantle. Seismic imaging and computational geodynamics indicate, however, that this interface is not an effective barrier to mantle flow and suggest that convective circulation occurs at a larger scale.

1.2 Three Decades of Seismic Tomography

Since the pioneering studies in the mid 1970's [see, for instance, *Julian and Sengupta*, 1973; *Sengupta and Toksöz*, 1976; *Aki et al.*, 1977; *Dziewonski et al.*, 1977; *contributions to the 1975 AGU Fall Meeting* (*EOS Trans. of the Am. Geophys. Un.*, 56, 393-396, 1975)], advances in technology (e.g., three to five orders of magnitude increase in computer processing speed, mass storage, and memory), inverse theory, and data quality and volume have vastly improved the tomographic imaging of Earth's deep interior structure. In a parallel development, our understanding of the images has improved dramatically because spectacular advances in computational geodynamics have facilitated the integration of the geological constraints on past plate motion at Earth's surface and the results of experimental and theoretical mineral physics with the snapshots of convection provided by seismic imaging.

1.2.1 Different inversion strategies

For the imaging of global structure two methods have become popular, each with specific benefits and shortcomings. The first represents lateral variations in seismic properties by superposition of global basis functions, such as spherical harmonics. This class of tomography [for reviews see *Woodhouse and Dziewonski*, 1989; *Romanowicz*, 1991; *Montagner*, 1994; *Ritzwoller and Lavely*, 1995; *Ekström*, this volume; *Mégnin and Romanowicz*, this volume] is attractive for imaging structure at a long wavelength, λ, because the number of model parameters, which scales as l^2 — with harmonic degree l inversely proportional to wavelength ($\lambda = 2\pi r/l$), is then small enough to resolve the coefficients by direct inversion of carefully selected and processed waveform data. Figure 1a depicts long wavelength variations in shear wavespeed at 1300 km depth according to the l=12 model of Su et al. [1994]; at this depth, the (half wavelength) resolution is 1350 km.

The large number of global basis functions required to describe structure at length

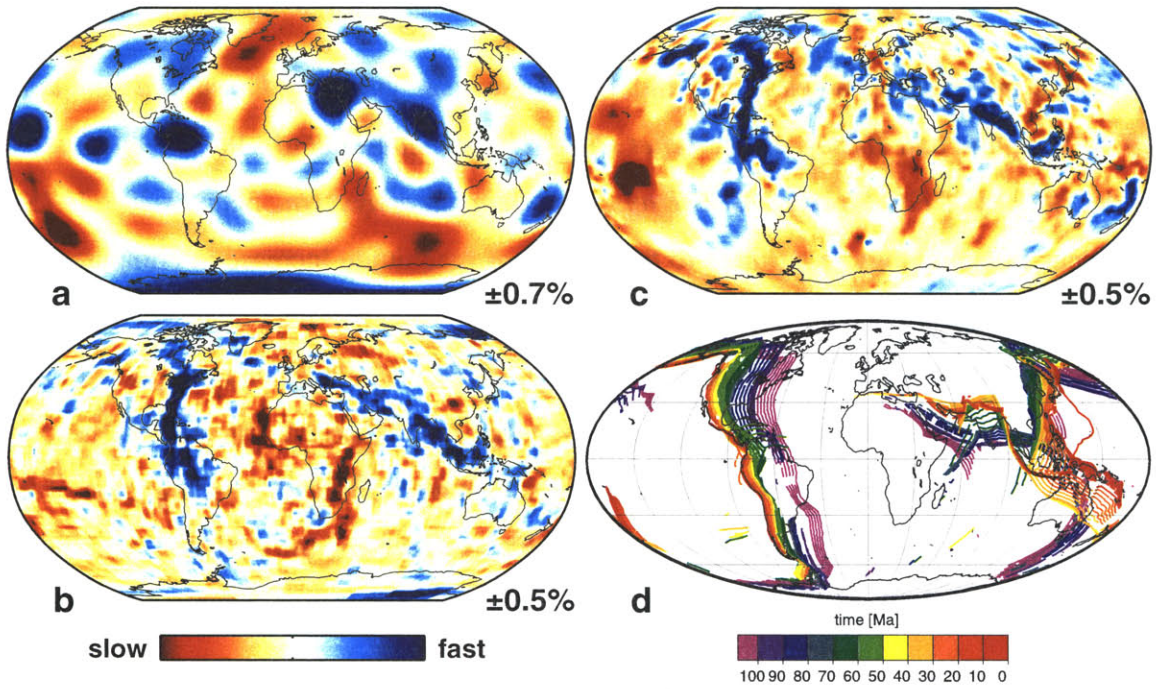


Figure 1-1: Different types of data and parameterization. Robinson projections of lateral variation of seismic wavespeed at approximately 1300 km depth. (a) *S* wavespeed expanded by global basis functions (spherical harmonics up to degree and order 12) by *Su et al.* [1994]; (b) *S* wavespeed represented by regular blocks as local basis functions [*Grand, personal communication*, 1999; see *Grand et al.*, 1997; *Grand*, 1994]; (c) *P* wavespeed represented by blocks [Kárason and Van der Hilst, 2000]. The relative amplitude of the perturbations is given at the bottom right of each panel. (d) Mollweide projection of the history of plate convergence [cf. *Lithgow-Bertelloni and Richards*, 1998]; color indicates time interval in which subduction occurred.

scales of several hundred km or less, such as slabs of subducted lithosphere in the upper mantle, would prohibit direct inversion. Moreover, the coefficients can no longer be determined accurately owing to uneven data coverage on the relevant length scales, and artifacts can be introduced in regions of poor coverage [*Boschi and Dziewonski*, 1999]. It then becomes attractive to represent wavespeed variations by local basis functions, such as non-overlapping constant-slowness volumes (e.g., tetrahedrons, voronoi cells, rectangular blocks) or cubic splines or wavelets that interpolate between grid values. In this class of

tomography, the large number of model parameters necessitates the use of iterative solvers such as *LSQR* [*Paige and Saunders*, 1982; *Nolet*, 1985] or *SIRT* [*Humphreys and Clayton*, 1988], with solution selection and resolution assessment less elegant than for direct inversions, but regularization (damping) can be used to avoid artifacts in regions of limited data coverage [*Spakman and Nolet*, 1988]. In the past decade the cell size used in regular grid inversions has decreased from about 6°x6° [*Inoue et al.*, 1990; *Pulliam et al.*, 1993] to 2°x2° [*Van der Hilst et al.*, 1997]. Figure 1b depicts S wavespeed variations in constant slowness blocks of 2.5°x2.5°x200km [*Grand, personal communication*, 1999; see *Grand et al.*, 1997], and Figure 1c depicts lateral variations in P wavespeed [*Kárason and Van der Hilst*, 2001] in 3°x3°x150km blocks. These map views reveal long linear features of faster-than-average wave propagation that reflect past episodes of plate convergence at Earth's surface. Even though the wavespeed variations may at first seem rather different, many structures in Figure 1a resemble the low-pass filtered version of those in Figures 1b and 1c, and if one considers the difference in nominal resolution the three models are, in fact, fairly consistent with each other in regions of adequate data coverage.

1.2.2 Flexible parameterization

Uneven source and receiver distribution results in significant spatial variations in data coverage. Global basis functions offer no flexibility for regionalization but a local basis can be adapted to lateral variations in data coverage. Small blocks can be used in densely sampled regions without unnecessary overparameterization of poorly sampled regions [e.g., *Bijwaard et al.*, 1998]. Ideally such irregular grids reflect the spatial variation of resolution, but since this is difficult to quantify they have been designed on the basis of sampling [*Gudmundsson and Sambridge*, 1998; *Bijwaard et al.*, 1998] or regional interest [*Abers and Roecker*, 1991; *Widiyantoro and Van der Hilst*, 1996]. Figure 2 illustrates the effect of different parameterizations.

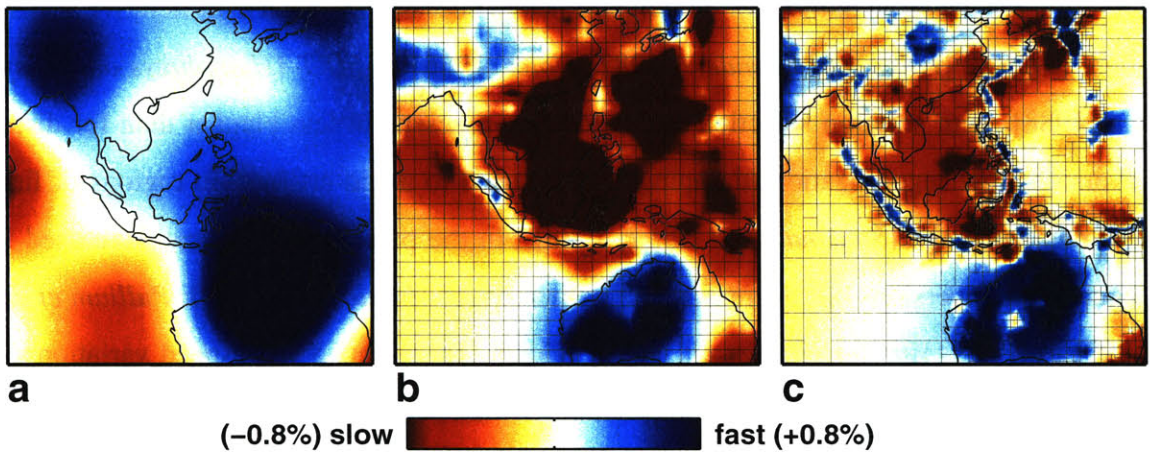


Figure 1-2: Recent progress in tomographic imaging. The images depict the lateral variation in seismic wavespeed at about 300 km depth beneath Australiasia according to different global models published in the last decade. (a) spherical harmonics up to degree and order 12 [*Su et al.*, 1994]; (b) *P* wavespeed on a regular 3°x3° grid [*Kárason and Van der Hilst*, 2000], which is similar to the technique used by Van der Hilst et al. [1997] but with slightly larger cells; (c) *P* wavespeed on an irregular grid [*Kárason and Van der Hilst*, 1999]. The image based on spherical harmonics is dominated by the high wavespeeds of the continental cratons in Australia and central Asia, and in the regular grid inversion the signature of the narrow slabs is overwhelmed by the low wavespeed in the back arc regions. Irregular gridding [e.g., *Bijwaard et al.*, 1998; *Kárason and Van der Hilst*, 1999] gives the most satisfactory rendering of the actual structure. We remark that results similar to Plate 2c can be obtained by reducing the cell size in regular grid inversions [e.g., *Widiyantoro and Van der Hilst*, 1996] but for global inversions the number of model parameters would then become prohibitively large.

1.2.3 Data quality and coverage

The success of any tomographic study depends critically on data quality. Most shear wave studies have been based on waveform data that were carefully selected and processed by individual investigators [e.g., *Woodhouse and Dziewonski*, 1984; *Masters et al.*, 1996; *Li and Romanowicz*, 1996; *Grand et al.*, 1997; *Van Heijst and Woodhouse*, 1999; *Ekström*, this volume; *Mégnin and Romanowicz*, this volume], but for global *P*-wave inversions such data sets are only now being constructed. With almost 15 million entries, the largest sin-

gle data source available for tomography consists of the travel-time residuals processed and published by the International Seismological Centre (ISC). This data set is noisy but its size and redundancy allows the extraction of structural signal. Many researchers have processed the ISC data prior to inversion, but probably the most rigorous effort was made by *Engdahl et al.* [1998]. Using non-linear procedures for earthquake relocation and seismic phase re-identification they improved hypocenter parameters and travel time residuals for a large range of seismic phases. The striking agreement between images based either on travel times inferred from careful waveform processing (Figure 1b) or from routinely reported and processed phase arrivals (Figure 1c) demonstrates the value of the data processing by Engdahl and co-workers. It also demonstrates that — at least at the depth shown — compressional and shear wavespeed are highly correlated with each other.

Uneven data coverage continues to be one of the most persistent problems in travel time tomography, although the use of irregular grids can reduce some of its detrimental effects on imaging. For a given source-receiver distribution the sampling of Earth's structure can be improved by considering data and ray paths not only of direct P or S but of other phases as well. For P-wave imaging, seismologists have experimented successfully with depth phases (such as pP), which also helps constraining focal depth [*Engdahl et al.*, 1998], surface reflections (such as PP and PPP), and core reflected (PcP) and refracted (PKP) waves [cf. *Van der Hilst et al.*, 1991; *Van der Hilst and Engdahl*, 1991; *Obayashi and Fukao*, 1997; *Vasco and Johnson*, 1998; *Bolton and Masters*, 1998; *Van Heijst and Woodhouse*, 1999]. For S travel time tomography, arrival times of direct S, the multiple surface reflections such as SS, SSS, $SSSS$, core reflections and refractions ScS, SKS, and surface (Love) wave phase velocities have been used [cf., *Grand*, 1994; *Liu and Dziewonski*, 1998; *Widiyantoro et al.*, 1998]. Large gaps in data coverage remain, however. In regions where mantle structure cannot yet be resolved by body wave data one can impose zero perturbations from the reference model or use independent constraints on long wavelength variations, such as

frequency shifts of Earth's free oscillations [*Kárason and Van der Hilst*, 1998; *Ishii and Tromp*, 1999].

1.2.4 A new model for mantle P-wavespeed

Because of our interest in the deep mantle [*Van der Hilst and Kárason*, 1999; *Kellogg et al.*, 1999], we made an effort to improve the sampling of lower mantle structure by incorporation of travel time data of core refracted *PKP* and diffracted P_{diff} waves [*Kárason and Van der Hilst*, 2001]. Figure 3 illustrates the ray geometry for these phases (NB. we used 3-D sensitivity kernels based on calculations by *Zhao et al.* [2000] for the back projection of the P_{diff} data but those are not shown here). In fact, we used differential times PKP_{DF}-PKP_{AB}, PKP_{BC}-PKP_{AB}, and PKP_{DF}-P_{diff} to reduce effects of hypocenter mislocation and to facilitate the extraction of signal from structure in the deep mantle. Moreover, the requirement that several readings be made from a single record works as an excellent quality criterion [*Kárason and Van der Hilst*, 2001]. For these phases we used accurate travel times inferred from waveform cross correlation by *McSweeney* [1997] (1383 *PKP* differential times) and *Wysession* [1996] (542 P_{diff} differential times) as well as some 27,412 carefully selected and processed *PKP* differential data from the data base of Engdahl and co-workers. These data were used along with nearly 8 million *P* and *pP* data from the Engdahl catalog; see *Kárason and Van der Hilst* [2001] for details about the data integration. This augmented data set provides much better data coverage than the *P* data alone (Figure 4), in particular in the southern hemisphere.

Figure 5 illustrates how the lateral variation of *P*-wavespeed changes with increasing depth in the mantle. The signature of cratonic parts of ancient continents and the narrow linear anomalies associated with plate subduction characterize structure in the upper mantle. (We remark that in this model the thickness of continental lithosphere is not well resolved since for most continental receivers these the body waves used sample shallow

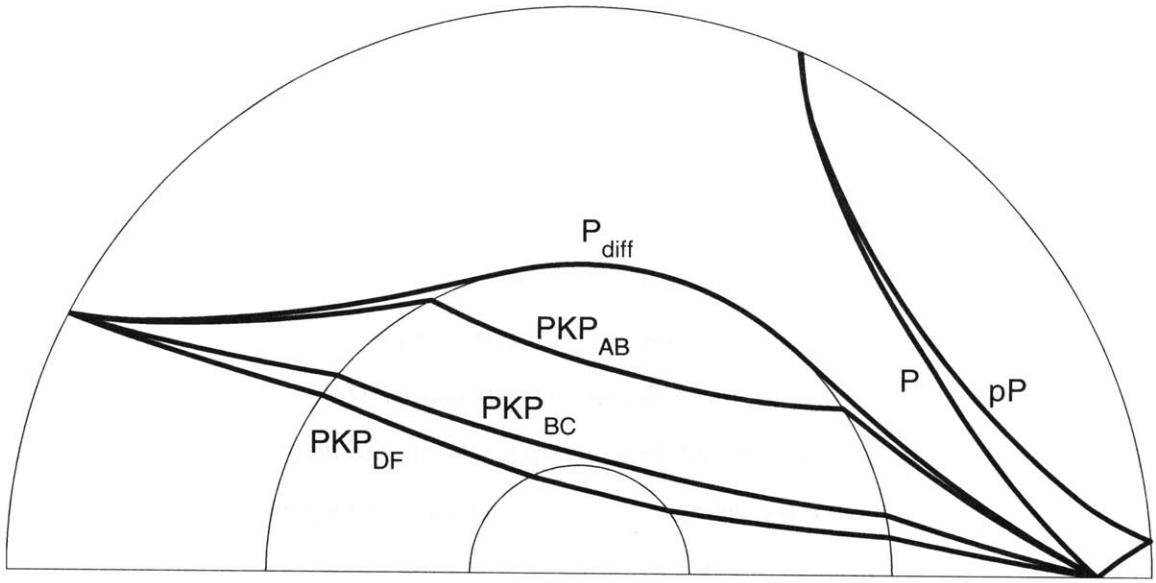


Figure 1-3: Ray geometry for the phases used in the *P*-wave study by Kárason and Van der Hilst [2000]. The core refracted *PKP* and diffracted P_{diff} waves provide better sampling of the deep mantle than *P* waves alone. We remark that instead of narrow rays we used 3-D sensitivity kernels to account for the fact that the P_{diff} travel times were measured from long period waveforms [*Wysession*, 1996]. The sensitivity kernels were based on mode summation [*Zhao et al.*, 1999].

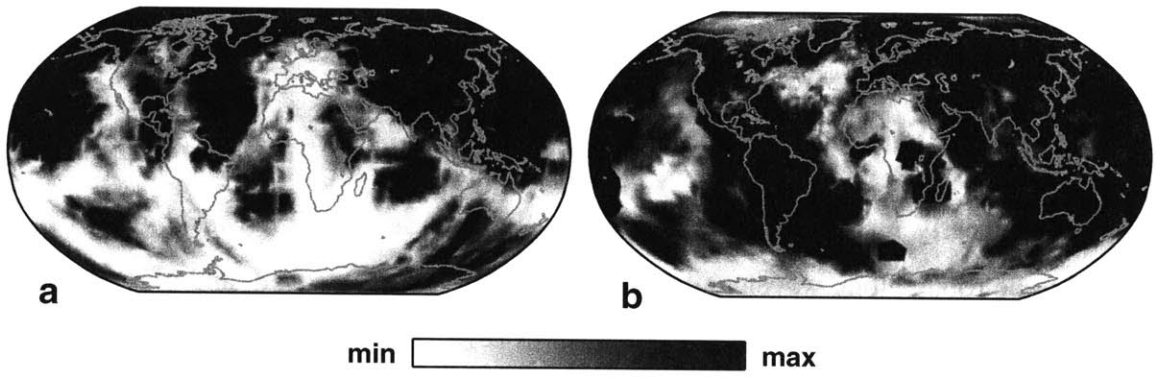


Figure 1-4: (a) sampling by *P* and *pP* and (b) *PKP* and P_{diff}. Shown here are the column sums of the sensitivity matrix **A** used by *Kárason and Van der Hilst* [2000].

mantle structure in only a small range of incident angles.) In the mid-mantle the amplitude of wavespeed variations is significantly smaller than in the upper mantle, but a pattern of long, tabular structural features emerges beneath the major convergent plate boundaries (see also Figure 1). The thickness of these tabular structures is much larger than the subducted slabs in the shallow mantle, which can perhaps be attributed to radial changes in viscosity [*Fischer et al.*, 1990; *Bunge et al.*, 1996]. Toward the base of the mantle the peak-to-peak amplitudes increase, albeit not as much as reported for shear wave perturbations [*Masters et al.*, 1996]. Figure 6 depicts an estimate of spatial resolution at selected depths. As expected from the uneven data coverage, the image recovery also reveals large geographical variations, but the changes in the character of heterogeneity from mid-mantle to the base of the mantle can not be attributed to sampling alone. The detail provided by the fine parameterization in regions of dense sampling is lost in the global maps of Figure 5 and Figure 6 but can be appreciated once we zoom into a particular region (Figure 2) or display structure by means of vertical sections across selected plate margins (Figure 7).

1.3 Results of Imaging Pertinent to the Issue of Mantle Convection

While many early results [*Dziewonski*, 1984; *Woodhouse and Dziewonski*, 1984] have proved to be robust, long wavelength models continue to be improved, and in recent studies the wavespeed variations are expanded up to l=24 [*Ekström*, this issue; *Mégnin and Romanowicz*, this issue]. This class of modeling does not, however, resolve trajectories of mantle flow in sufficient detail to determine unequivocally whether or not convection is stratified at 660 km. Indeed, its results can be used to argue either way [*Richards and Engebretson*, 1992; *Wen and Anderson*, 1995]. In the 1990's, travel time tomography with local basis functions has made several seminal contributions to our understanding of mantle

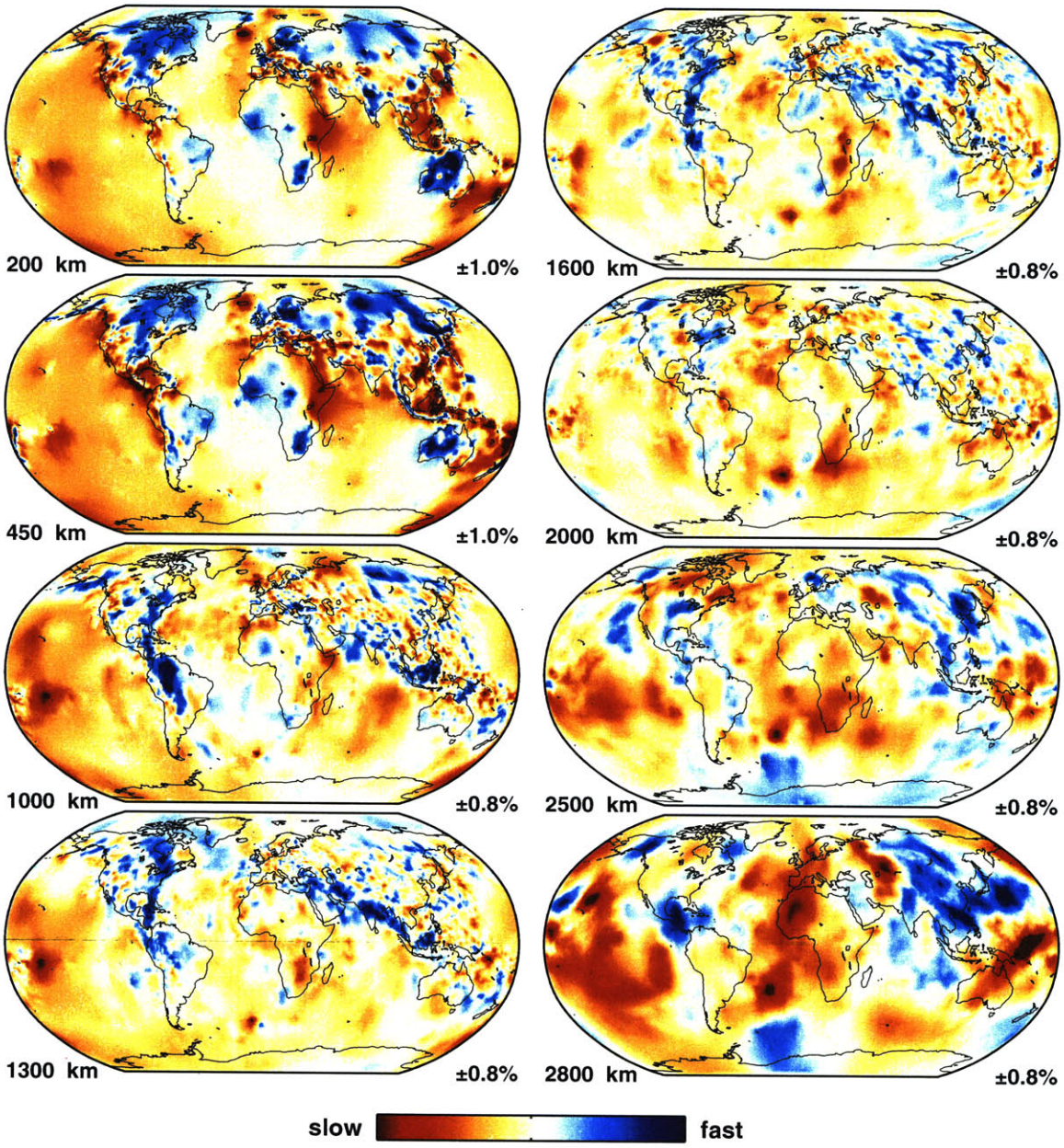


Figure 1-5: Robinson projections of the lateral variation in *P* wavespeed at several depths in Earth's mantle according to *Kárason and Van der Hilst* [1999]. This model is based on data from the seismic phases displayed in Figure 1, and thus benefits from improved sampling in the deep mantle (Figure 2), and the flexible parameterization illustrated in Plate 2 (although the effect of the irregular grid is hardly appreciated at this scale).

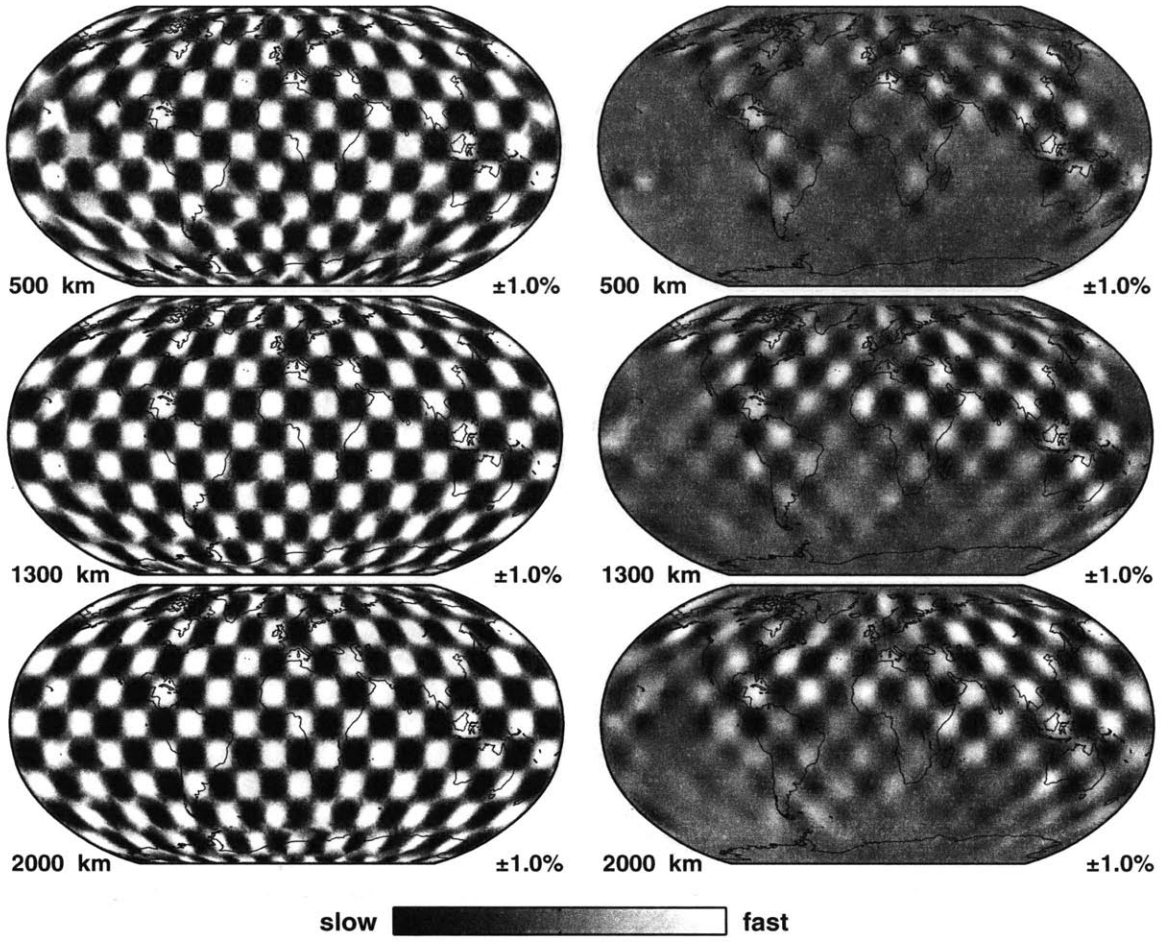


Figure 1-6: Estimate of spatial resolution at 500, 1300, and 2000 km depth. From the input model shown on the left we calculated synthetic travel time residuals by ray tracing and we then inverted them to test if (and where) we can resolve the know input pattern with the data coverage used. The input anomaly is $\pm 2\%$ in the center of each cell. This exercise shows that there are large regions in the mantle were resolution is poor owing to inadequate sampling, in particular in the upper mantle. Notice also that image recovery at 2000 km depth is equal to or better than at 1300 km, which indicates that the dramatic change in heterogeneity, compare, for instance, Plates 3d and e, is not simply related to sampling.

convection because it has enabled the mapping of flow trajectories in unprecedented detail. First, some slabs of subducted lithosphere penetrate into the lower mantle while others appear trapped in the upper mantle. This was first borne out by regional studies [e.g., *Zhou*

and Clayton, 1990; *Van der Hilst et al.*, 1991; *Fukao et al.*, 1992] but has since been confirmed by other studies, including high resolution global inversions [*Widiyantoro*, 1997; *Bijwaard et al.*, 1998; *Kárason and Van der Hilst*, 1999]. Figure 7 provides examples of both styles of subduction. Slab deflection may occur beneath Izu Bonin and the southern Kuriles (Figure 7b,c), the Banda arc, and beneath the Thyrrenian Sea; deep slabs, sometimes severely deformed in the transition zone, have been detected beneath the Mariana, Tonga-Kermadec (Figure 7e), Sunda (4d), and northern Kurile arcs, the Philippines, the Aegean Sea (4a), and Central and South America (4f). Inspired by *Kincaid and Olson* [1987], *Van der Hilst and Seno* [1993] and *Van der Hilst* [1995] argued that the observed complexity does not imply stratification at 660 km depth but can be caused by interplay between relative plate motion (i.e., lateral trench migration) and the deformation of slabs when they encounter resistance (e.g., higher viscosity or a depressed phase boundary). Experimental [*Griffiths et al.*, 1995; *Guiltier-Frottier et al.*, 1995] and numerical [*Zhong and Gurnis*, 1995; *Davies*, 1995; *Christensen*, 1996] fluid dynamical experiments support this view, but selective weakening of the descending plate by grainsize reduction upon phase transformation may also contribute [*Riedel and Karato*, 1997]. The complexity caused by interaction of downwellings with the upper mantle transition zone persists to near 1000 km depth [*Van der Hilst and Kárason*, 1999] but the significance of this depth is not yet established and many slabs sink to even larger depth (Figure 7).

Second, independent *P* and *S* studies have begun to agree on structure as small as several hundred km [*Grand et al.*, 1997; *Van der Hilst et al.*, 1997] - a development that has increased the credibility of this class of imaging and may prove to be one of the milestones of tomography; they revealed a relatively simple pattern of narrow high-wavespeed structures in the lower mantle beneath plate boundaries with a long history of subduction (Figure 1) and indicated that these deep structures often connect to seismogenic slabs in the upper mantle (Figure 7). Several slab structures disappear from tomographic view in the bottom

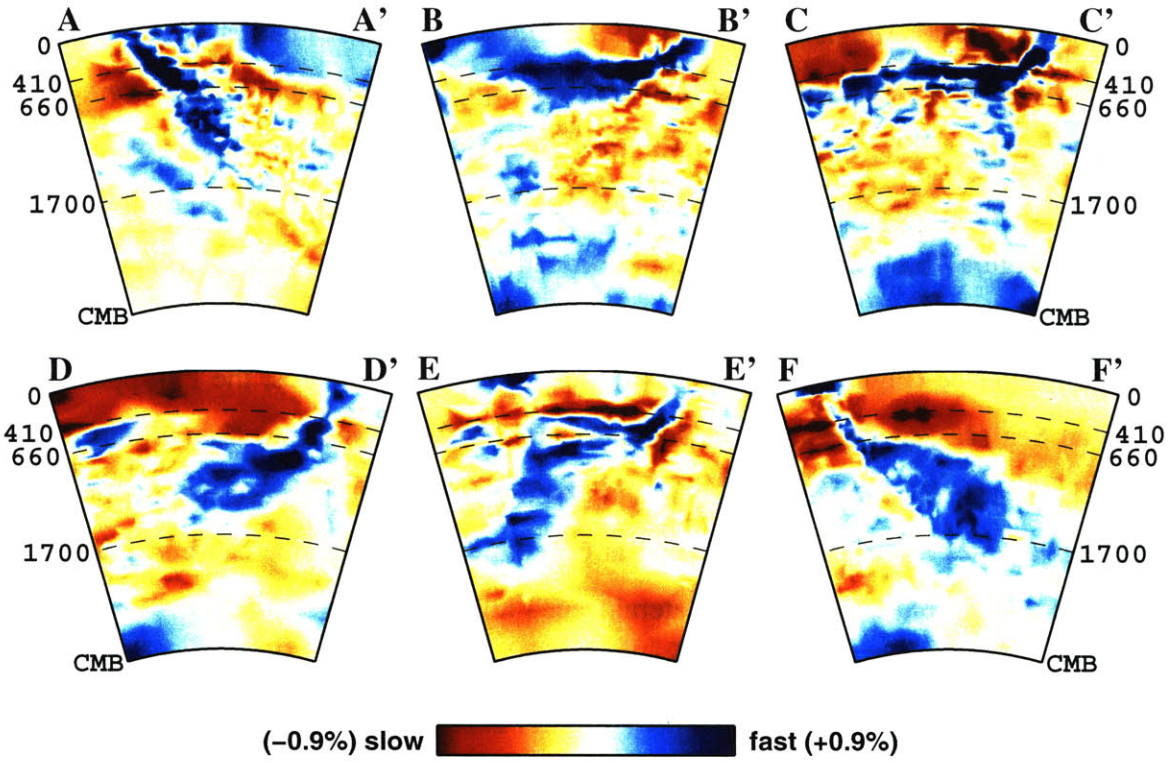


Figure 1-7: Slab structure illustrated by vertical mantle sections across (a) the Hellenic (Aegean), (b) southern Kurile, (c) Izu Bonin, (d) Sunda (Java), and (e) the northern Tonga island arcs, and (f) central America. See map inset for cross section locations.

1000 km of the mantle (e.g., Figures 2a, d-f), but some fragments seem to connect to D" heterogeneity, for instance beneath eastern Asia and central America. Also the correlation between P and S images, which persists to large depth, may break down near the base of

the mantle [*Grand et al.*, 1997; *Su and Dziewonski*, 1997; *Kennett et al.*, 1998].

Third, despite theoretical and practical difficulties significant progress has been made in the mapping of seismically slow anomalies, and several studies now suggest that mantle upwellings are continuous over a large depth range, supporting the view that 'plumes' originate (at a boundary layer) below the 660 km discontinuity [*Wolfe et al.*, 1997; *Bijwaard and Spakman*, 1999; *Goes et al.*, 1999; *Ritsema et al.*, 1999]. Image resolution — in particular in the radial direction — continues to form a formidable obstacle, however.

1.4 Scale of Mantle Convection

These observations render untenable the conventional end-member views of either convective stratification at 660 km or undisturbed whole mantle flow. The 660 km discontinuity distorts mantle flow, occasionally resulting in local and transient layering, but many slabs penetrate to at least 1700 km depth in the mantle. The change in structure and heterogeneity spectrum between 1700 and 2300 km depth is probably real, but its origin is not yet known. It may point to stratification of some sort in the deep mantle [*Van der Hilst and Kárason*, 1999; *Kellogg et al.*, 1999], but it may also reflect changes in the nature of global plate motion in the distant past [*Richards and Engebretson*, 1992]. Alternatively, *Anderson* [1999] explains the slab-like lower mantle structures and their apparent continuity to upper mantle subduction zones by thermal coupling and coincidental alignment of structures in separately convecting upper and lower mantles. We take issue with this. It would be very fortuitous indeed if slowly changing structures in the sluggish lower mantle align with rapidly changing structures in the less viscous upper mantle. Thermal coupling can work [*Nataf*, 1988; *Čížková et al.*, 1999] but is too slow to explain the large depth of slab penetration beneath young convergent margins and is inconsistent with the constant dip angle inferred for several slabs. Moreover, if only heat is exchanged, thousands of kilometers of subducted lithosphere must have accumulated in the upper mantle beneath margins with a

long subduction record, for which there is no observational evidence.

The geophysical evidence against stratification at 660 km depth is strong, and if isolated and seismically visible 'reservoirs' exist they are likely to reside at a larger depth. A mantle convection scenario that is then worth considering is one in which — apart from the upper and lower boundary layers (lithosphere and D", respectively) — three domains are identified. In the view postulated by *Kellogg et al.* [1999] and *Van der Hilst and Kárason* [1999], and discussed by *Albarède and Van der Hilst* [1999], undegassed and enriched material in the bottom 1000 km of the mantle has not (yet) mixed with the part of the mantle — the top 2000 km or so — that is involved in the recycling of slab material, has a relatively uniform major element composition, and represents the depleted and outgassed source of mid-ocean ridge basalts. The upper mantle transition zone (400-1000 km depth) divides the depleted part of the mantle in a well mixed, low viscosity domain and a deeper one with high viscosity and slower transients; here, mantle flow is distorted by viscosity stratification, effects of phase transitions in the mantle silicates, and changes in plate motion at Earth's surface.

For long term survival the deep domain must have a slightly higher intrinsic density than the overlying depleted mantle [*Tackley*, 1998; *Kellogg et al.*, 1999]. The interface between these domains would be close to isopycnic (i.e., compositional and thermal effects on buoyancy are in balance) and significant dynamic topography can develop with some slabs penetrating to near the CMB [*Kellogg et al.*, 1999]; this resembles the 'penetrative' convection proposed previously for stratification at 660 km depth [*Silver et al.*, 1988]. The anomalous deep mantle "layer" — if it indeed exists — could represent long-lived differences in composition or phase chemistry [*Van der Hilst and Kárason*, 1999]. Alternatively, in an evolutionary sense, it may not yet have been churned by subduction. It takes many tens of millions of years (or more) before changes in the plate configuration at Earth's surface affect the heterogeneity structure of the deep mantle. Moreover, as cold slabs sink, their negative buoyancy can be diminished (or enhanced) by phase reactions in the upper

mantle transition zone and may be neutralized by compositional effects well above the very base of the mantle [*Kesson et al.*, 1998]. In this transient system recycling may involve only the upper and mid-mantle, and few subduction systems may have operated long enough to produce (sufficiently 'cold') slabs that reach the CMB and churn the deep layer. This can be tested by rigorous integration of plate reconstructions and tomographic images of mantle structure [*Besse, personal communication*, 1999] by means of numerical modeling [e.g., *Bunge et al.*, 1998].

1.5 Directions for Future Research

Seismic tomography did not settle the debate about the scale of mantle convection just by demonstrating the deep penetration of many slabs, but it has moved research away from conventional — but inadequate — end-member convection models. Many aspects of mantle convection remain enigmatic [e.g., *Albarède and Van der Hilst*, 1999]. Seismic imaging will continue to play a central role in constraining the pattern of convective flow, but several issues must be sorted out. However dramatic the "red-and-blue" images, our actual understanding of them is unsatisfactory. We need better constraints on the amplitude of the changes in P and S-wavespeed, and their ratio $\delta lnV_S/\delta lnV_P$, and we need to integrate them with results from experimental and theoretical mineral physics to quantify changes in temperature, phase, and bulk composition. In addition, the trajectories of mantle convection must be delineated in even more detail in order to resolve outstanding issues, such as the ultimate fate of the slabs that penetrate across the upper mantle transition zone and the source and morphology of the return flow (plumes?). The latter requires a significant effort to improve data coverage beneath regions far away from seismically active plate boundaries. Seismic anisotropy can also be used as a tool for delineating flow trajectories [*Montagner*, 1998]. Continued imaging of mantle structure beneath convergent margins and better integration with the record of plate motions through computational geodynamic modeling will

further our understanding of subduction and its partnership with mantle convection.

Acknowledgments. We thank Wei-jia Su and Steve Grand for providing us with their models, Li Zhao for the mode calculations that we used to approximate the sensitivity kernels for the PKP_{DF}-P_{diff} *travel time differentials, Bernard Steinberger for producing the diagram used as Figure 1d, and Mark Richards and Richard Gordon for constructive reviews. This research is supported by the National Science Foundation (EAR-9909492).*

Chapter 2

Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (*PKP*, P_{diff})[1]

Abstract

The mapping of variations in *P* wave speed in the deep mantle is restricted by the uneven sampling of *P* waves, in particular beneath the Southern Hemisphere. To enhance data coverage, we augmented the ~1.6 million summary rays of P, pP, and pwP that we used in previous studies with differential travel times of diffracted and refracted core phases. For the core-refracted differential travel time residuals (PKP_{AB}-PKP_{DF} and PKP_{AB}-PKP_{BC}) we used 1383 cross-correlated digital waveforms as well as ~27,000 routinely processed bulletin data. We used the waveform data to define quality criteria for the selection and reduction of the bulletin PKP data. For PKP_{DF}-P_{diff} we only considered 543 records derived from waveform cross correlation. No PcP data were used in this study.

[1]This chapter was published as: H. Kárason and R. D. van der Hilst, Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (*PKP*, P_{diff}), in *Journal of Geophysical Research*, *106*, 6,569–6,588, 2001.

We used optical ray theory to calculate the ray paths associated with the P, pP, pwP, and PKP data, which are measured at 1 Hz. However, to account for the large Fresnel zones of the low-frequency (~50 mHz) PKP_{DF}-P_{diff} data we estimated the three-dimensional shape of the Fréchet sensitivity kernels from kernels calculated by normal mode summation. The use of these kernels allows us to properly distribute the sensitivity for a given seismic phase over a large mantle volume while allowing the high-frequency data to constrain small-scale structure. The differential times are relatively insensitive to source mislocation and to structure in the shallow mantle beneath source and receivers, and they have previously been interpreted exclusively in terms of lateral structure directly above the core mantle boundary (CMB). However, images thus obtained can be contaminated by effects of small scale structure elsewhere in the mantle. Here, we do not make a priori assumptions about the mantle source of anomalous time differentials. From test inversions we conclude that (both upper and lower) mantle structures that are poorly resolved by P data can be mapped into the core along PKP paths but that the effect of outer core structures, if any, on the mantle model is small. Compared to the inversion of the P, pP, and pwP alone, the inclusion of the PKP_{AB}-PKP_{DF} and PKP_{AB}-PKP_{BC} and PKP_{DF}-P_{diff} data improves the resolution of structure beneath 2200 km depth. In particular, the joint inversion puts better constraints on the long-wavelength variations in the very deep mantle and yields an increase in the amplitude of velocity perturbations near the CMB that is in agreement with but still smaller than inferences from shear wave studies. Resolution tests indicate that in some regions the enhanced definition of structure is significant, but in most regions the improvements are subtle and structure remains poorly resolved in large regions of the mantle.

2.1 Introduction

In the last decade, continuous improvement of seismic imaging techniques, data quality, and computer power has resulted in increasingly detailed images of Earth's interior structure [*Su et al.*, 1994; *Li and Romanowicz*, 1996; *Masters et al.*, 1996; *van der Hilst et al.*, 1997; *Grand et al.*, 1997; *Vasco and Johnson*, 1998; *Bijwaard et al.*, 1998; *Kárason and van der Hilst*, 2000]. There is growing agreement on the large-scale (>4000 km) structure of the upper and lowermost mantle [*Dziewonski et al.*, 1999]. Studies based on body and surface waves have discovered high wave speed anomalies beneath the stable continents and low wave speeds beneath tectonically active regions and marginal basins and travel

time tomography has revealed relatively narrow structures of higher-than-average seismic wave speed beneath convergent margins [i.e., *Grand*, 1994; *van der Hilst et al.*, 1997; *Bijwaard et al.*, 1998; *Vasco and Johnson*, 1998]. These narrow features have been interpreted as subducted slab. Recent studies suggest that some of these slabs continue into the lower mantle, which renders unlikely convection models with two distinct layers separated by the seismic discontinuity at 660 km depth.

Statistical analysis of International Seismological Centre (ISC) residuals suggests an increase in the amplitude of long-wavelength structure in the lowermost mantle, reaching a maximum at the core-mantle-boundary (CMB) [*Gudmundsson et al.*, 1990], which is in accord with tomographic imaging that revealed relatively slow and fast propagation beneath Africa and the Pacific and in large regions beneath the Americas and eastern Asia, respectively [e.g., *Su et al.*, 1992]. Other studies reported a decreased correlation between S and P wave speed perturbations in the deep mantle which may in part be due to changes in bulk composition [e.g., *Robertson and Woodhouse*, 1995; *Su and Dziewonski*, 1997; *Kennett et al.*, 1998; R.L. Saltzer and R.D. van der Hilst, submitted manuscript, 2000]. These and other observations inspired *van der Hilst and Kárason* [1999] to propose the existence of compositionally distinct domains in the deep mantle, which could host heat-producing elements and certain noble gases [*Kellogg et al.*, 1999; *Albarède and van der Hilst*, 1999]. There is also evidence for small-scale structures near the base of the mantle, including ultralow-velocity zones and rapid lateral variations in anisotropy (see *Garnero* [2000] for a review), but this complexity cannot be resolved with the techniques discussed here.

Between 1000 and 2000 km depth the amplitude of the wave speed variations is small, the scatter in the pertinent data is high [*Gudmundsson et al.*, 1990], and the spectrum of heterogeneity, white, at least up to degree 12 [e.g., *Su and Dziewonski*, 1992]. Global tomography suggests the presence of deep slabs of subducted lithosphere [*van der Hilst et al.*, 1997, 1998; *Grand et al.*, 1994], but in some models a slab signature is not evident

beneath 1700 ± 300 km depth [e.g.,*van der Hilst and Kárason*, 1999; and *Kárason and van der Hilst*, 2000). A transition from middle to lowermost mantle structure has been proposed [*Su et al.*, 1994; *van der Hilst et al.*, 1997; *Liu and Dziewonski*, 1998; *van der Hilst and Kárason*, 1999], but uncertainty remains about its existence, the depth of onset, and its impliciations for mantle flow.

While models of mantle S wave speed can be constrained by body waves, long period surface waves, and observed frequency shifts of Earth's free oscillations [e.g., *Masters et al.*, 1996], global P models have traditionally been derived from travel times of body wave phases that are confined to the mantle, such as P [*Dziewonski et al.*, 1977; *Inoue et al.*, 1990; *Pulliam et al.*, 1993], pP [*van der Hilst et al.*, 1997; *Bijwaard et al.*, 1998], or PP [*Vasco et al.*, 1995a]. The spatial coverage by these phases is, however, very uneven owing to the distribution of earthquakes and stations and to the geometry of the ray paths (Figure 1a).

Large regions in the lower mantle, especially in the Southern Hemisphere, are not adequately sampled owing to the scarcity of stations. In such regions, tomographic imaging will underestimate the true amplitude of the wave speed variations or not reveal them at all. *Song and Helmberger* [1997] showed that the model by *van der Hilst et al.* [1997] explains only $\sim$20% of the variance of observed PKP travel times and concluded that the amplitudes of *P* velocity variations in the lower mantle were underestimated by up to 1 order of magnitude. *Van der Hilst et al.* [1998] showed that the pattern (but not the magnitude) of PKP differential time residuals measured by *McSweeney* [1995], mapped at the exit and entry points at the CMB, is consistent with their tomographic model. Assuming that they sense structure near the very base of the mantle, *Wysession* [1996] argued that PKP_{DF}-P_{diff} differential travel time residuals require higher levels of velocity perturbations than is typically inferred from mantle tomography.

To improve sampling, Vasco and co workers [*Pulliam et al.*, 1993; *Vasco et al.* 1993,

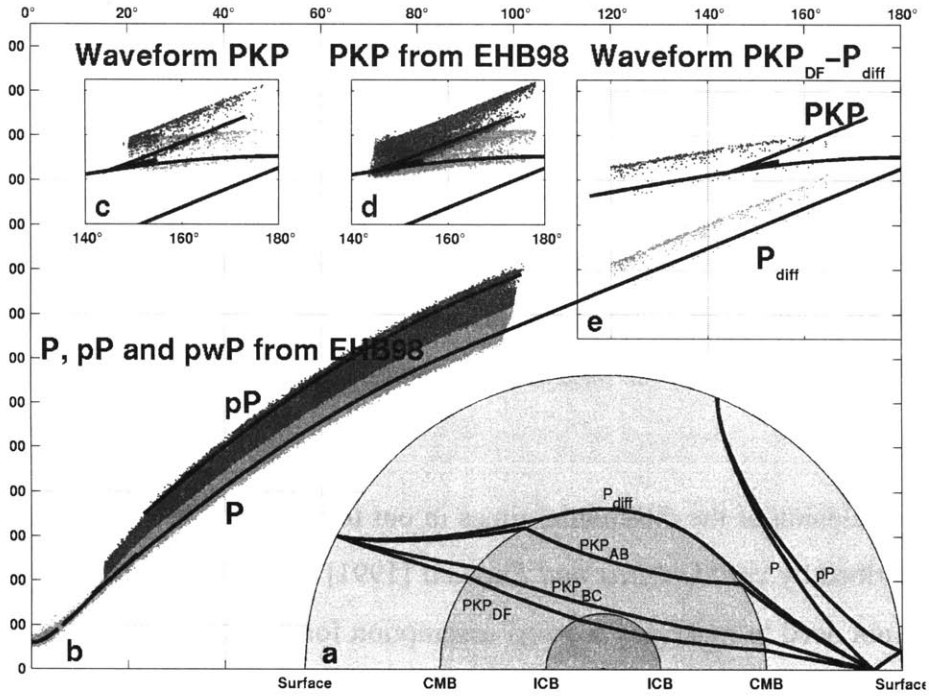


Figure 2-1: (a) The ray paths of the phases used in this study. P is a direct P wave that does not travel through the core, and pP is reflected once off the surface. $PKP_{\rm AB}$ and $PKP_{\rm BC}$ travel through the outer core and $PKP_{\rm DF}$ travels both through the outer and the inner core. For $P_{\rm diff}$, the diffracted wave along the core-mantle boundary, we account for finite frequency effects by using 3-D kernels. (b) Travel times versus distance, (solid lines) theoretical curves for a earthquake at 500 km depth with respect to the ak135 model. Also shown are scatter plots for P (light shading, $\sim$7,000,000) and pP (dark shading, $\sim$430.000) from EHB98. (c) Same as Figure 1b except waveform based $PKP_{\rm AB}$-$P_{\rm DF/BC}$ (dark and light shading, respectively, 1383). (d) Same as Figure 1b except $PKP_{\rm AB}$-$P_{\rm DF/BC}$ from EHB98 (dark and light shading, respectively, $\sim$ 27.000). (e) Same as (b) except waveform based $PKP_{\rm DF}$-$P_{\rm diff}$ (dark and light gray, respectively, 543). For detailed discussion of data sets, see text.

1994, 1995a] used ISC data from a large range of seismic phases. Their most recent P wave model is based on P, PP, $PKP_{\rm AB}$, $PKP_{\rm BC}$, $PKP_{\rm DF}$, and PcP (absolute) travel time residuals [*Vasco and Johnson*, 1998]. The addition of core-reflected and -refracted phases enhances data coverage and improves the resolution. However, estimates or formal calculations of resolution depend on data coverage only, and it is not well known to what extend the high noise level in later arriving phases, in particular PcP, degrades the model.

Here, we aim to improve models of the deep mantle by using differential travel time times of core-refracted and -diffracted phases along with (absolute) travel time residuals of mantle P and pP waves. Instead of relying on routinely processed data we use PKP_{DF}-P_{diff} data from *Wysession* [1996] and PKP_{AB}-PKP_{DF} and PKP_{AB}-PKP_{BC} residuals from *McSweeney* [1995] (see Figure 1a for ray geometry). These differential travel time residuals have been determined using waveform cross correlation. We also used the large number of PKP times from the data set of *Engdahl et al.* [1998] to construct differential travel time residuals.

For the inclusion of the differential times in our tomographic scheme we used the approach described by *van der Hilst and Engdahl* [1991] for $pP - P$ and $PP - P$ data so that we do not need to make any a priori assumption for the localization of the source of the structural signal. Including the diffracted P waves in travel time tomography for three-dimensional mantle structure is, however, not trivial because (1) it is not a ray geometrical phase (it is evanescent and dispersive so that the observed travel time and its sensitivity to structure are frequency dependent) and (2) they were measured at much longer period than both the routinely processed (P, PKP) and waveform-based (PKP) data. Approximating the sensitivity by infinitely narrow rays is thus not appropriate, and, instead, we used 3-D sensitivity kernels, inferred from mode summation [*Zhao and Jordan*, 1998; *Zhao et al.*, 2000]. We did not use P_{diff} data from the *Engdahl et al.* [1998] catalog. We evaluated $PcP - P$ differential travel times, but because the data are noisy and not readily consistent with our mantle models, we decided not to use them for the construction of the models presented here.

We discuss the construction and use of 3-D sensitivity kernels and the selection, reduction, and combination of the waveform-based and routinely processed differential travel time residuals. We discuss the resolution of structure in the deep mantle and the potential trade off between mantle and core structure.

2.2 Data

2.2.1 P, pP, and pwP

Most of our data are selected from the routinely processed data set by *Engdahl et al.* [1998]. These, in turn, are based on bulletin data from the ISC and the U.S. Geological Survey's National Earthquake Information Center (NEIC). However, in an iterative nonlinear procedure, earthquakes were relocated, and for more than 60 different body wave phases the picks were reidentified using travel times according to the ak135 [*Kennett et al.*, 1995] reference model for P wave speed. The data set is updated almost every year and the catalog used here comprises a total of nearly 14,000,000 phase arrival times from ~100,000 earthquakes that occurred between January 1, 1964, and December 31, 1998. The data reported to ISC and NEIC are measured from short-period (~1 Hz) records. Hereinafter this data set will be referred to as EHB98.

From the EHB98 data we used almost 7,000,000 *P* and 430,000 *pP* and *pwP* travel time residuals (see Figure 1a for ray paths and Figure 1b for distance ranges), many of which have also been used in previous studies [e.g. *van der Hilst et al.* 1997; *Widiyantoro*, 1997; *Bijwaard et al.*, 1998]. We remark that the EHB98 data file contains more entries for these phases, but we have rejected those that were labeled as inaccurate by *Engdahl et al* [1998]. Hereinafter we refer to the EHB98 P, pP, and pwP, simply as EHB98 $P+$ data.

An advantage of the EHB98 $P+$ data is that data redundancy facilitates the extraction of structural signal, but a disadvantage of this and (most) other body wave data sets is the uneven sampling of the mantle. Upper mantle structure is well sampled near belts of high seismic activity and beneath dense receiver arrays, but large mantle volumes are poorly sampled, for instance in the upper mantle beneath intraplate regions and above the core mantle boundary beneath the Southern Hemisphere (Figure 2a).

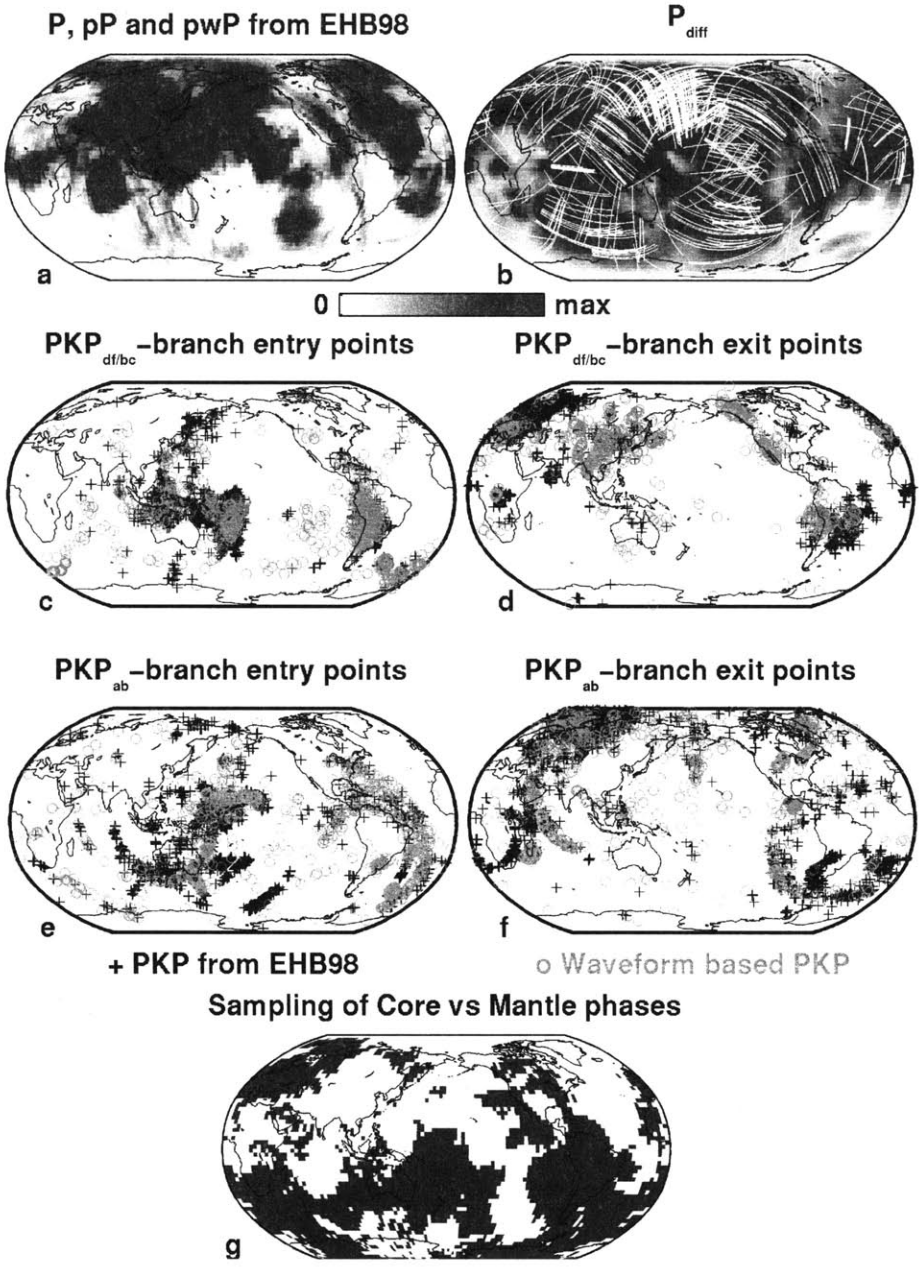


Figure 2-2: (a) The sampling of structure just above the CMB in terms of the (normalized) sum of ray segment lengths within each cell for EHB P, pP, and pwP. (b) Same as Figure 2a except $P_{\rm diff}$ data. Note that for the construction of this figure we used the 3-D sensitivity kernels (see section 3.5). Also shown is the part of the optical ray that grazes the CMB (see Figure 1a). (c)-(f) Points of entry to and exit from the core for different branches of PKP for both PKP data sets. (g) Binary plot, showing as solid where the sum of the ray segments from all core phases exceeds the sum of EHB98 $P+$ ray segments.

2.2.2 Core Phases (PKP and $P_{\rm diff}$)

To improve our ability to map the lowest few hundred km of the mantle we consider core refracted (PKP) and diffracted ($P_{\rm diff}$) waves (Figure 1). $PKP_{\rm AB}$ and $PKP_{\rm BC}$ travel through the outer core and $PKP_{\rm DF}$ goes both through the outer and the inner core. For $P_{\rm diff}$ the incidence angle is critical at the CMB so that the waves are diffracted along the core mantle boundary. PKP data were used for tomographic imaging by *Vasco and Johnson* [1998]. $P_{\rm diff}$ data have been used to constrain lateral variations of *P* wave speed [*Wysession*, 1996] but not in published 3-D studies.

Differential travel time residuals.

For the core phases we consider differences in arrival time of two phases in a single seismogram. A differential travel time residual is the observed (obs) minus predicted (ak135) difference in travel time (T) of two phases in a single seismogram, which can be obtained from reported travel time residuals of the individual phases. For example, a $PKP_{\rm AB}$-$PKP_{\rm DF}$ differential travel time residual

$$\begin{aligned} &\delta t_{PKP_{\rm AB}-PKP_{\rm DF}} \\ &= (T_{PKP_{\rm AB}} - T_{PKP_{\rm DF}})_{\rm obs} - (T_{PKP_{\rm AB}} - T_{PKP_{\rm DF}})_{\rm ak135} \\ &= \delta t_{PKP_{\rm AB}} - \delta t_{PKP_{\rm DF}}. \end{aligned} \tag{2.1}$$

Differential times can also be measured by waveform cross correlation. In a single seismogram the waveform of one arrival can be used to predict that of another by accounting for attenuation and phase shifts. For example, the Hilbert transform of $PKP_{\rm BC}$ resembles the $PKP_{\rm AB}$ arrival since the latter passes through an internal caustic [*Jeffreys and Lapwood*, 1957; *McSweeney*, 1995]. Alternatively, the observed record can be cross-correlated

with synthetic seismograms calculated by mode summation [*Woodward and Masters*, 1991; *Wysession et al.*, 1995] or ray theory [e.g., *Grand*, 1994]. In contrast to most travel time picks, such waveform cross correlations do not necessarily measure the first onset, but for measuring time differences this is not as serious a problem as for measuring absolute times.

Differential times can not be formed at all distances from the source, as can be seen from the travel time curves in Figures 1c-1e. This limits the number of data that can be used for imaging. However, in the distance range beyond which differential times can be formed all PKP waves are of type DF, that is, they propagate through the inner core, and they have shorter ray segments in the deep mantle than any of the other PKP phases, which makes them less attractive for the imaging of mantle structure. The use of differential travel time residuals has several important advantages. First, they are relatively insensitive to uncertainties in earthquake location and origin time and to heterogeneity in the shallow mantle, so that the contribution of the structural signal from the deep mantle can be isolated. However, the success of this isolation depends on ray geometry (and thus epicentral distance), on the (frequency dependent) width of the sensitivity zone around the ray, and on the length scale of aspherical earth structure. For instance, at midmantle depths the paths of PKP and P_{diff} can differ by hundreds of kilometers (Figure 1a), which is more than the width of a Fresnel zone for a 1-Hz wave at that depth. Small scale structure in the mid mantle can thus be sampled preferentially along one of the paths, which can lead to bias if the differential times are interpreted as due to deep mantle structure only. Second, it serves as a data quality criterion. Multiple phase identification is less prone to phase misidentification, is only possible from records with high signal-to-noise ratio, and is probably only done by experienced seismic analysts.

PKP.

For PKP we use residuals determined both from the EHB98 catalog and from short-period digital seismograms using waveform cross correlation. *McSweeney* [1995] made 1383 short-period measurements and the EHB98 catalog yielded ~15,000 of PKP_{AB}-PKP_{BC} and PKP_{AB}-PKP_{DF} differential residuals.

From the EHB98 PKP data we also from differential times The restriction to specific epicentral distances (Figure 1d) necessarily reduces the number of data that can be used; in our case from a total of $\sim 1.5 \times 10^6$ to the ~54,000 that are used to construct ~27,000 differential times. Assuming that they are of higher quality than the routinely processed EHB data, we used the measurements by *McSweeney* [1995] to reduce the EHB98 differential travel time residuals. For a direct comparison of the two data sets we searched the EHB98 catalog for stations and events that were located in caps with a radius of 0.5° around the earthquake and station location in the waveform data set. Subsequently, we binned the qualifying EHB98 PKP data and calculated the median of the residuals. Figure 3 shows the correlation (as a function of the number of EHB98 data within each bin) between the waveform based data and the medians of the EHB98 residuals thus selected. Good correlation is achieved for bins containing more than three paths. We processed all EHB98 PKP data in a similar way: for every event-station pair we searched for other earthquake-station pairs within 0.5° caps and determined the median value of the selected residuals. Effectively, this creates a spatial moving average filter of the PKP data. The use of bins with three or more hits significantly reduced the scatter (standard deviation decreased from 2.7 to 1.4 s) in the EHB98 PKP differential times, but it further reduced the number of bundles of differential pairs used in the inversion to ~15,000.

For the data set used in our inversions we show in Figures 2c-2f the core entry and exit points for the two PKP data sets. The potential for improving models of lower mantle P wave speed with PKP data is greatest in the Southern Hemisphere, e.g. beneath South

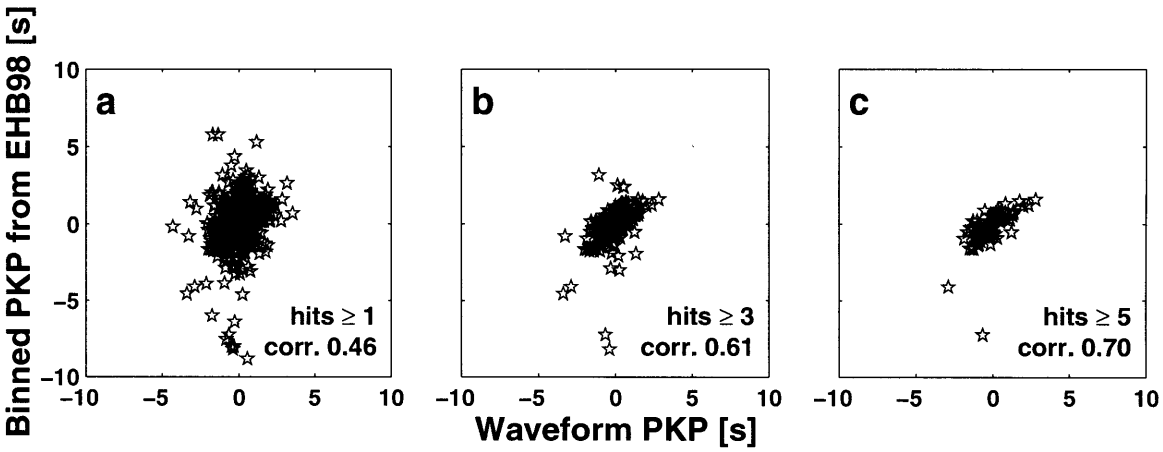


Figure 2-3: Correlation between raw residuals from waveform based PKP data and the (medians of) binned PKP residuals from EHB98. All medians (a) and medians with minimum three (b) and five (c) hits per bin.

America and the southwest Pacific, but many regions in the Northern Hemisphere benefit also, in particular beneath northern America, western Europe, and the Atlantic.

P_{diff}.

for the diffracted waves we do not use the EHB98 catalog. Instead, we use 543 PKP_{DF}-P_{diff} measurements obtained by *Wysession* [1996] from long-period (~50 mHz) digital seismograms by waveform cross correlation. For these data, the epicentral distance range considered is 110° to 165° (Figure 1e). Diffraction cannot be described by ray theory, and a significant fraction of the work for this study was devoted to the design and implementation of 3-D sensitivity kernels (see section 3.5).

Figure 2b illustrates the effective data coverage in the deep mantle for long period P_{diff}. We also show the part of the corresponding optical P_{diff} ray that grazes the CMB. Although the number of PKP and P_{diff} data is small compared to the number of EHB98 $P+$ data they are given a larger weight in the inversions (see below) and in many regions of the deep mantle they provide the only direct constraints.

Effects of core structure.

PKP residuals are sensitive to wave speed variations in the outer core, if any, and, for the DF branch, to inner core structure [e.g., *Creager*, 1992; *Song and Helmberger*, 1995]. We omit $PKP_{\mathrm{BC}} - PKP_{\mathrm{DF}}$ differential times because they have small sensitivity to mantle structure. We further reduced the sensitivity of our data to inner core structure by correcting all PKP_{DF} travel times for the inner core model by *Su and Dziewonski* [1995], although test inversions suggest that with the differential pairs used here the contamination of mantle structure by inner core signal would be small (see section 5.1).

The conventional view is that there is no significant heterogeneity in the outer core [e.g. *Jellinek et. al*, 1999]. *Kohler and Tanimoto* [1992] investigated velocity variations within the outermost 200 km of the core using SKS and $SKKS$ waveforms and concluded that effects from mantle structure are hard to separate from core signal. *Vasco and Johnson* [1998] discussed the possibility of aspherical structure in the Earth's liquid outer core. Since such structure would effect the PKP data used in our study, we investigated the trade-off between mantle and core structure with a series of inversion tests (see section 5.2). We conclude that explaining the core data do not require outer core heterogeneities. Therefore, except in specific tests we do not include the outer core in the inversions discussed here.

2.3 Methodology

2.3.1 Tomographic Problem

The purpose of travel time tomography is to interpret differences between observed and predicted (theoretical) travel times (that is, the "residuals") in terms of aspherical variations in wave speed (that is, "structure" or the "model") in Earth's interior. The inversion technique that we use, including the parameterization and the use of composite rays, is now

fairly standard and is described in detail by, for instance, *Nolet* [1987] and *Spakman and Nolet* [1988]. Here we merely mention some basic aspects and discuss the integration of different data sets and the implementation of approximated 3-D sensitivity kernels.

Without the terms for regularization, the system of normal equations that describe the (linearized) relationship between model parameters and the data is given by

$$\mathbf{L}\Delta\mathbf{s} = \Delta\mathbf{t}. \tag{2.2}$$

The sensitivity matrix that contains the Fréchet derivatives, $\mathbf{L}$, contains the length of the ray segments in different cells and the partial derivatives for the simultaneous source location, $\Delta\mathbf{s}$ (the model vector) represents the slowness perturbation in the model space relative to an assumed reference model (ak135) and the change in hypocenter (origin time, longitude, latitude, depth), and $\Delta\mathbf{t}$ (the data vector) contains the travel time residuals. Using the iterative LSQR algorithm [*Paige and Saunders*, 1982; *Nolet*, 1985; *van der Sluis and van der Vorst*, 1987], we invert $\Delta\mathbf{t}$ simultaneously for structure and for effects on source mislocation by minimizing the variance of the solution, $\sigma^2 = \boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = \Delta\mathbf{t} - \mathbf{L}\Delta\mathbf{s}$ is the misfit between data and model predictions. The results shown here were obtained after 200 iterations, but with LSQR most of the convergence is achieved within a small number of iterations: $\sim$98% of the final variance reduction was obtained after only 25 iterations.

There are many more data than model parameters, but the problem is underdetermined since a large part of the model space cannot be constrained because of uneven sampling. Moreover, the system of equations is inconsistent owing to data noise. Consequently, there is no unique solution and we need to use regularization (damping). We apply norm damping, which suppresses anomalies in poorly constrained cells and thus biases to a lower amplitude solution, and gradient (first derivative) minimization, which enforces smooth variations in latitudinal and longitudinal directions. No regularization is used in radial direction.

2.3.2 Parameterization

The model (Δs) is usually described by a linear combination of orthogonal basis functions that span the model space of interest. Many global studies have used global basis functions, such as spherical harmonics [e.g. *Su et al.*, 1994; *Li and Romanowicz*, 1996; *Masters et al.*, 1996]. Here, we follow *Inoue et al.* [1990], *Pulliam et al.* [1993], *Vasco et al.* [1994, 1995a], *Grand et al.* [1997], *van der Hilst et al.* [1997], *Vasco and Johnson* [1998], *Bijwaard et al.* [1998] and many others and use a cellular representation with nonoverlapping constant wave speed blocks in longitude, latitude, and radius. Both representations have specific advantages and disadvantages. A local basis facilitates the suppression of artifacts in regions of no data coverage [*Boschi and Dziewonski*, 1999] and allows efficient regionalization [*Abers and Roecker*, 1991; *Bijwaard et al.*, 1998; *Kárason and van der Hilst*, 2000]. However, the number of parameters often is too large for exact inversions so that such approaches have to resort to iterative methods, which involve somewhat subjective criteria for convergence and regularization and do not allow the calculation of a resolution matrix.

Since the objective here is to improve the constraints on deep mantle structure by inclusion of core phases, no attempt was made to optimize the solution for small scale structure, such as slabs, in the upper mantle. For the parameterization of the mantle we use a regular 3-D grid consisting of 20 layers, each $\sim$150 km thick, which are divided into 7200 $3^\circ \times 3^\circ$ latitude and longitude cells. For the assessment of the trade-off between mantle and core structure (see section 5) we performed test inversions with a $10^\circ \times 10^\circ$ lateral grid in 6 (2) layers in the outer (inner) core.

2.3.3 Composite Rays and Variance Reduction of EHB98 $P+$ data

To suppress random noise in ISC picks we use the redundancy in the almost 7×10^6 EHB98 $P+$ residuals and construct "composite rays" [*Spakman and Nolet*, 1988]. We trace the

ray for each earthquake that is located in a mantle volume of $1^\circ \times 1^\circ \times 75$ km to a station located in a receiver cluster, but the path is always added to the same row of the sensitivity matrix $\mathbf{L}$. Such ray bundles, or "composite rays", better represent the sampling than single "summary rays" that connect an earthquake and station cluster. However, a comparison between the results based on either approach reveals little difference [*Widiyantoro*, 1997]. For the inversions we used almost 1.5×10^6 P and almost 220,000 pP and pwP composite rays. For the processing of the EHB98 differential PKP times see section 2.2.2.

We do not restrict the number of individual rays that form a composite ray. Some bundles comprise more than 1000, but the average is about four. To suppress effects of outliers we use the median of the residuals as the datum associated with such a ray bundle. In order not to bias the distribution, we do not discard data prior to calculating the medians; however, composite rays with medians that have absolute values in excess of 5 s are excluded to reduce further the effects of outliers (this concerns $<1\%$ of the original data set). Increasing the number of rays that constitute a composite ray results in a decrease of the spread of the associated residuals (Figure 4, solid line). We scale each matrix row so its total path length equals the square root of the sum of path lengths for all rays that constitute the composite ray and weight the residual accordingly. This is roughly equivalent to using the individual rays but repeating the median residual in a least squares inversion. However, beyond the first 10 or so rays the net gain of adding more measurements diminishes because effects of, for instance, earthquake mislocation, instrumentation, near receiver structure, and imperfect theoretical approximations begin to overshadow the random reading errors. Therefore we apply a constant weight ($\sqrt{10}$) to all composites with 10 or more rays. Although only 10% of the composite rays get the highest weighting they comprise more than 50% of the data.

Figure 4 shows that the variance reduction upon tomographic inversion, defined as $\mathbf{1} - \mathrm{Var}\,(\boldsymbol{\epsilon})/\mathrm{Var}\,(\boldsymbol{\Delta t})$, (dashed line) is small for composite rays with very few entries. If all

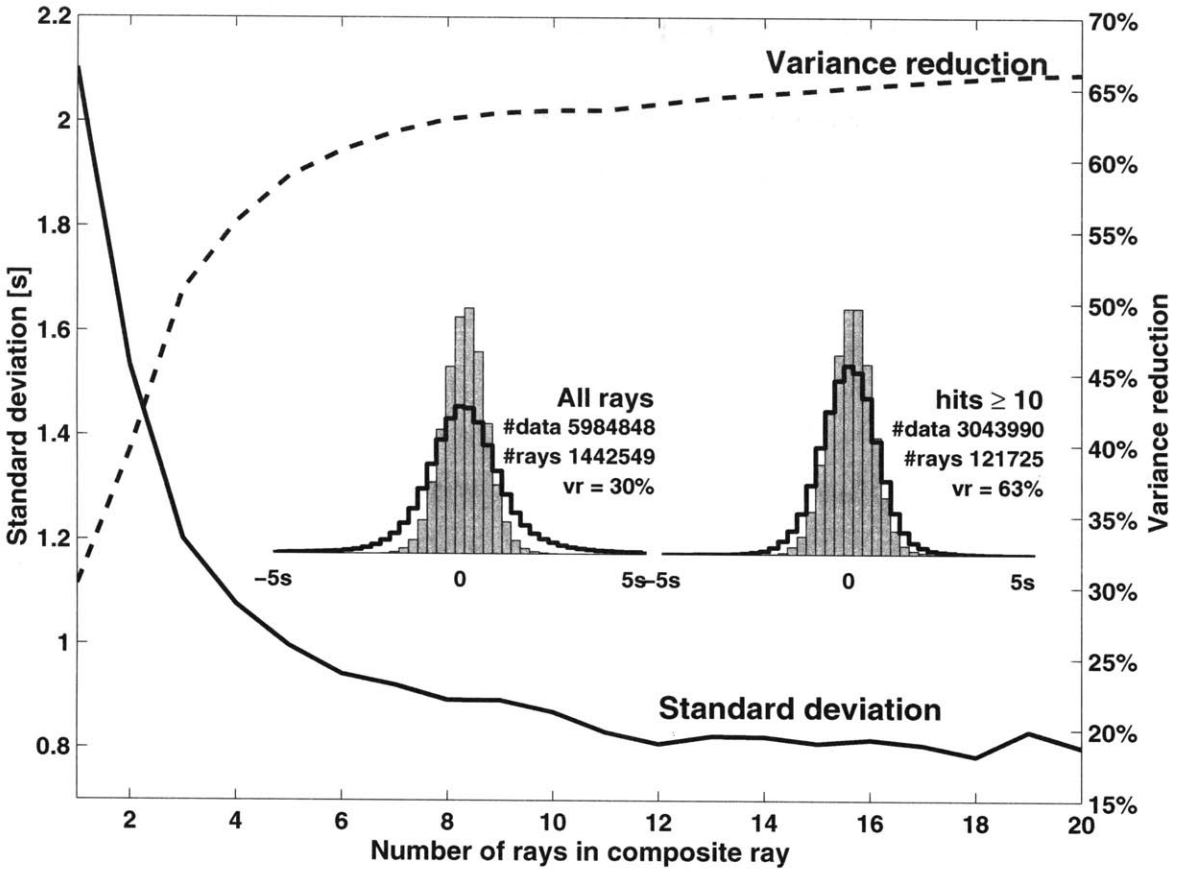


Figure 2-4: Standard deviation (solid line) of the median of travel time residuals from individual rays in a composite ray as function of number of rays in a composite ray. Variance reduction (vr) (dashed line) as function of minimum number of rays per composite ray used in the calculation. Histograms (inset plots) showing the distribution of original residuals (solid line) and model produced residuals (shaded bars) for all composite rays and for composite rays with 10 or more hits.

composite rays are included, irrespective of the number of rays within them, the variance reduction for the (composite ray) residuals is $\sim$30%. However, our weighting increases the influence of ray bundles with a larger number of rays, and for the composite rays with more than 10 hits (which involves about 50% of the data) the variance reduction is 63% and the frequency distributions of $\mathbf{L\Delta s}$ versus $\mathbf{\Delta t}$ are similar (see Figure 4, inset). The variance reduction of the 7×10^6 or so individual data is thus somewhere between 30% and 63%.

2.3.4 Weighting of Waveform Data

The regularization of the inversion problem depends large-ly on the sampling and the noise level of the largest data set, and small data sets, even high-quality ones, can only exert influence if given more weight than individual entries in the routinely processed EHB98 catalog. Each PKP and P_{diff} waveform measurement gets a weight equal to a composite ray containing 10 or more rays (i.e., $\sqrt{10}$).

2.3.5 Fresnel Zones and Sensitivity Kernels

The volume about the geometrical ray path between source and receiver that contributes significantly to the recorded seismogram is related to the Fresnel zone. In homogeneous media the maximum width (w_m) of the Fresnel zone can be approximated by ray geometry [*Nolet*, 1987]:

$$w_m = \sqrt{\frac{\lambda S}{2}}, \tag{2.3}$$

where λ is the wavelength and S is the total path length from source to receiver. In general the width of the Fresnel zone thus increases with decreasing frequency (increasing λ) and with increasing distance between source and receiver. With few exceptions [e.g., *Vasco et al.*, 1995b] such Fresnel zones are not used in travel time tomography. Instead, one usually calculates the Fréchet derivatives (i.e., the elements of matrix $\mathbf{L}$) by simple ray tracing and applies damping or imposes some spatial correlation length to obtain a smooth result. This avoids a major computational effort, it is adequate when high frequency waves are used since then the Fresnel zones are relatively small. For a frequency of $\sim$1 Hz the width of a Fresnel zone for a P wave that bottoms just above the CMB is $\sim$200 km, which is slightly larger than the cell size used in our inversions but less than the length scale of structure we expect to resolve. Moreover, the use of composite ray bundles also results

in the distribution of the sensitivity over neighboring cells and further smoothing can be accomplished by regularization.

For the interpretation of the PKP_{DF}-P_{diff} residuals the ray theoretical approximation is not satisfactory. They are deduced from records with a central frequency of $\sim$50 mHz, which implies a $\sim$1000 km Fresnel zones for waves that just touch the CMB and much more for the diffractions around the core. Furthermore, P_{diff} is not a ray geometrical phase: it is an evanescent and dispersive wave at the CMB so that the travel time and its sensitivity to structure are frequency-dependent.

Zhao and Jordan [1998] and *Zhao et al.* [2000] calculated frequency-dependent sensitivity kernels with a stationary phase approximation and normal mode summation with cross-branch mode coupling. An example of a P_{diff} kernel is shown in Plate 1a. The sensitivity is concentrated near the ray theoretical path. However, the cancellation of energy away from the ray (resulting from destructive interference of the modes) is not perfect, which is mostly due to the fact that the limited number of the modes and the finite frequency bandwidth used in the summation prevents the complete isolation of the low amplitude P_{diff} phase from other, often more energetic arrivals, such as PP. Plate 1b shows the sensitivity kernel for the travel time of PKP. For this epicentral distance the difference in arrival time for the three branches of PKP is small and at a frequency of 50 mHz it is not possible to isolate them from each other. The sensitivity away from the ray is suppressed more effectively than in the case of P_{diff} because PKP is a distinct high energy arrival.

To illustrate further the complex morphology of sensitivity kernels of core diffracted waves, we show in Plate 1c a great circle section through a 3-D kernel of S_{diff}, calculated for the transverse (*SH*) component, and in Plates 1d and 1e we depict the kernel in sections perpendicular to the ray path (at locations d and e in Plate 1c). For this calculation, *Zhao et al.* [2000] used an earth model based on 1066A [*Gilbert and Dziewonski*, 1975]. Interestingly, most of the sensitivity for diffracted phases is located above the ray; energy in waves

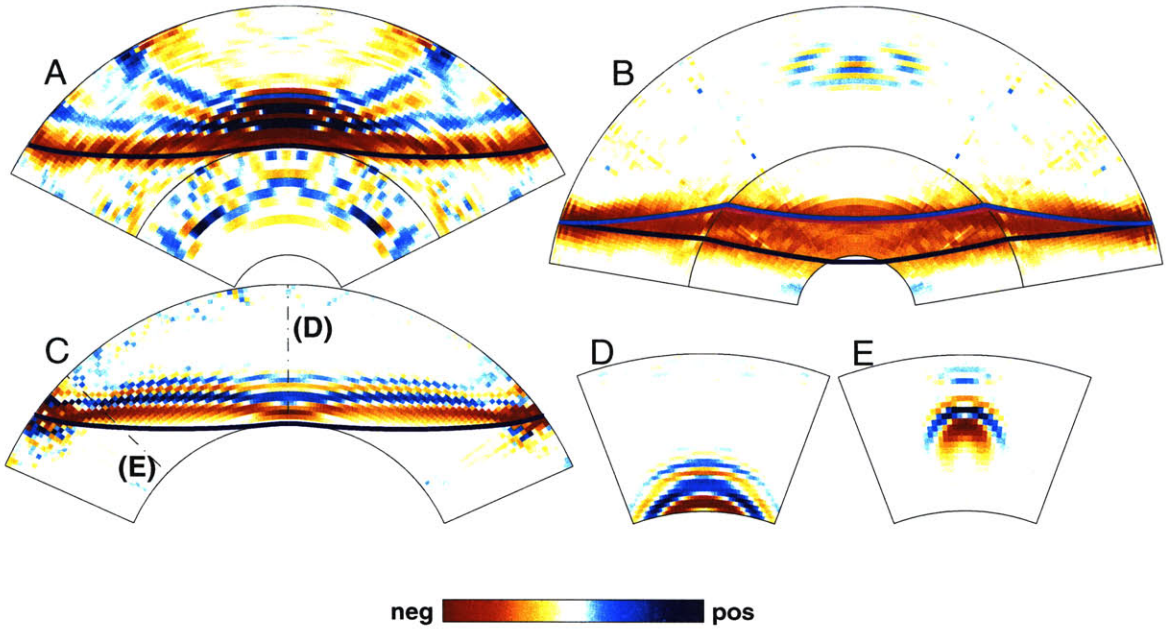


Figure 2-5: (a) Fréchet kernel in the event-station great circle plane showing sensitivity to wave speed variations within the earth for P_{diff} phase. Alternating red and blue indicate the first, second, etc, fresnel zone. (b) Same as Plate 1a except for PKP. (c) Same as Plate 1a except S_{diff} through a smoothed earth model. (d) and (e) Fréchet kernels along lines shown in Plate 1c.

heading deeper than the critical ray is refracted away from this ray path and does not contribute to the record as observed at the receiver. (Note that this should make the long-period diffracted waves virtually insensitive to topography of the core mantle boundary.) Plates 1d and 1e illustrate the broadening of the sensitivity kernel above the CMB; the maximum width of the kernel increases with increasing epicentral distance and is largest for Δ=180° when it covers almost the entire surface of the core.

The calculation of 3-D Fréchet kernels for each data point is not practical for large data sets. For computational efficiency we developed analytical expressions that match the 'true' kernels (calculated for three epicentral distances only) and designed an algorithm to project them onto the basis functions that define the model space. Plate 2 illustrates such an approximated kernel. The sensitivity is distributed over many cells around the ray and is scaled so that the total travel time is correct for a given 1-D Earth model. Plate 2a

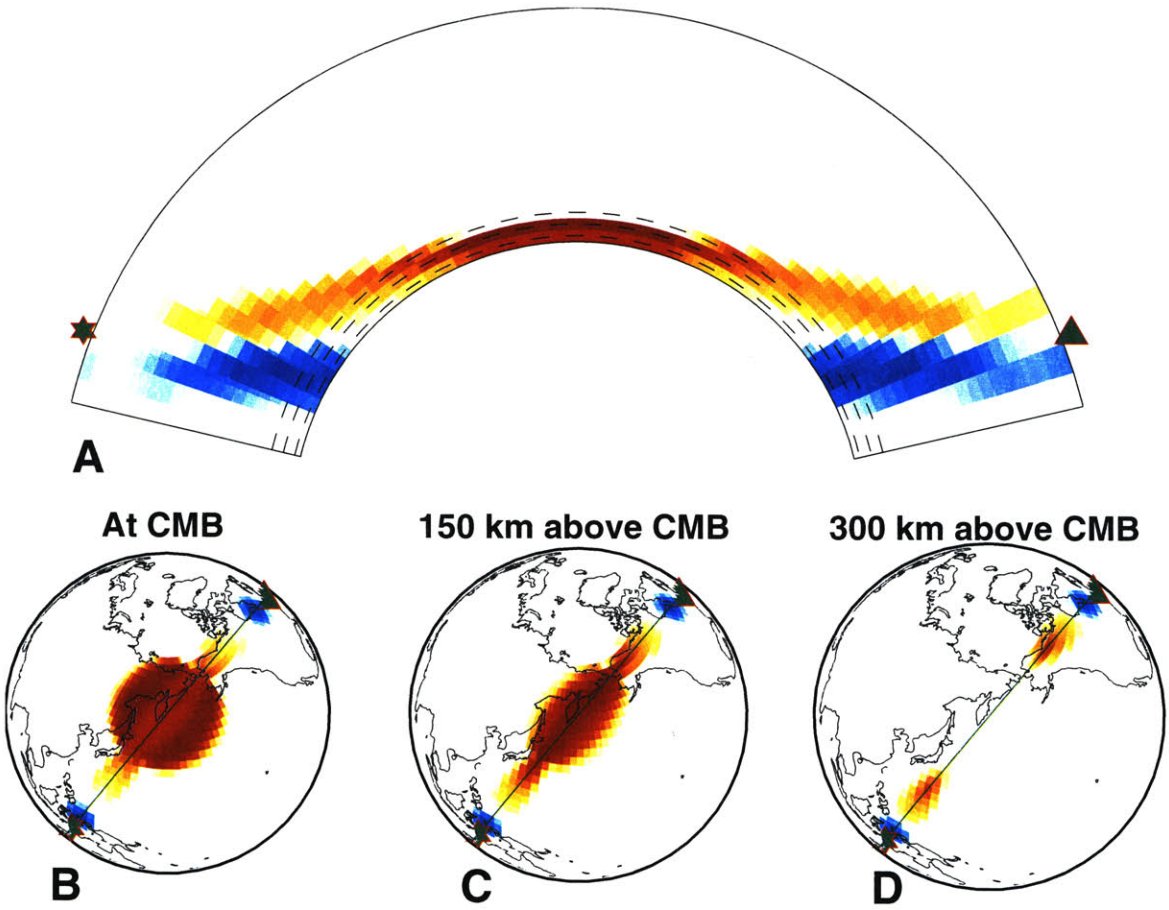


Figure 2-6: Sensitivity to slowness variations within the Earth using differential times. Red and blue indicate negative and positive sensitivity as one kernel is subtracted from the other to give sensitivity to differential times. (a) Cross section along the great circle connecting event (star) and station (triangle) from P_{diff} data file. (b)-(d) Three deepest layers of sensitivity kernel in map view.

shows that the sensitivities of PKP_{DF} and P_{diff} cancel in the shallow mantle and that the maximum sensitivity of the travel time residual is reached near the base of the mantle. Notice, however, that even for these long-period differential times the sensitivity to midmantle structure along both the P_{diff} and the PKP path is not negligible. Plates 2b-2d show the change in sensitivity with increasing distance from the CMB.

Although our analytical functions are mere approximations, we prefer them over narrow ray paths because they allow us to deal with the frequency dependance of P_{diff} sensitivity

kernels and to account for the differences in Fresnel zone width in the data space instead of accomplishing the required smoothing in the model space which would also prevent the short-period data from constraining small-wavelength structure.

2.4 Results

2.4.1 Models From Different Data Sets

We use the layer just above the CMB to illustrate the influence of the individual data sets on the images (Plate 3). Plate 3a depicts the solution based on EHB98 $P+$ data only (KH2000PM). The inclusion of the PKP_{DF}-P_{diff} significantly enhances the slow anomalies beneath the Pacific and west Africa and introduces high wave speed anomalies beneath southern Africa and the Tasman Sea (Plate 3b). We remark, however, that the deepest mantle beneath southern Africa is not well sampled by either of the data sets used (see also section 4.3). When the PKP data are inverted along with EHB98 $P+$ data, the fast structures beneath the Caribbean and at high latitude in the Southern Hemisphere become stronger and so do the slow anomalies beneath Africa and the western Pacific toward Australia (Plate 3c). The solution based on all data (KH2000PC) (Plate 3d) is in good agreement with the result based on EHB98 $P+$ alone, but several anomalies are now more prominent, in particular, beneath the Southern Hemisphere, as expected from the complementary nature of data coverage (Figure 2), and the overall amplitude of the anomalies has increased.

2.4.2 Compatibility of the Different Data Sets

We assessed the success of the integration of the different data sets by quantifying how well a particular subset of data is explained by the 3-D mantle models constructed with or without the data under evaluation. In Figure 5 we compare for different data sets the observations ($\mathbf{\Delta t}$) and model predictions ($\mathbf{L\Delta s}$), with $\mathbf{L}$ the sensitivity matrix for the data

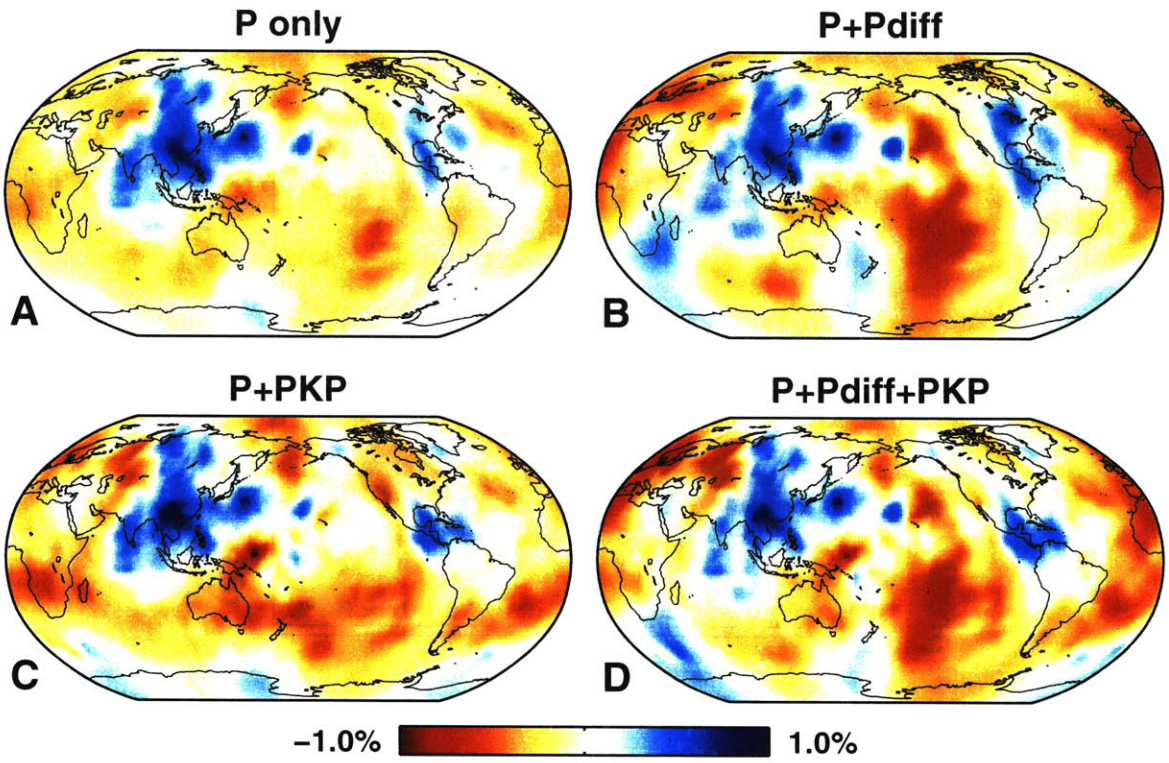


Figure 2-7: Deepest layer, centered around 2819 km depth for models based on different combinations of data sets. (a) KH2000PM: Model based on EHB $P+$. (b) P+$P_{\rm diff}$: EHB $P+$ and $P_{\rm diff}$ data. (c) P+PKP: EHB $P+$ and both waveform based PKP and PKP from EHB98 data. (d) KH2000PC: Model based on all data sets.

set to be tested. In Figure 5 the left column shows the predictions by KH2000PM based on EHB98 $P+$ only but all data were used for KH2000PC that render the fits in the right column. We show only residuals for composite rays with 10 or more rays to allow a direct comparison with waveform data. The models fit the EHB98 $P+$ data well, even though there is still a significant fraction of the signal that is not explained [see also *Bijwaard et al.*, 1998].

Similar to our previous models [e.g., *van der Hilst et al.*, 1997], the model based only on EHB98 $P+$ significantly underestimates the structural signal in the core phases (Figure 5, left). From PKP data fit, *Song and Helmberger* [1997] argued that the model by *van der Hilst et al.* [1997] underestimates the amplitude of deep mantle structure by up to an order

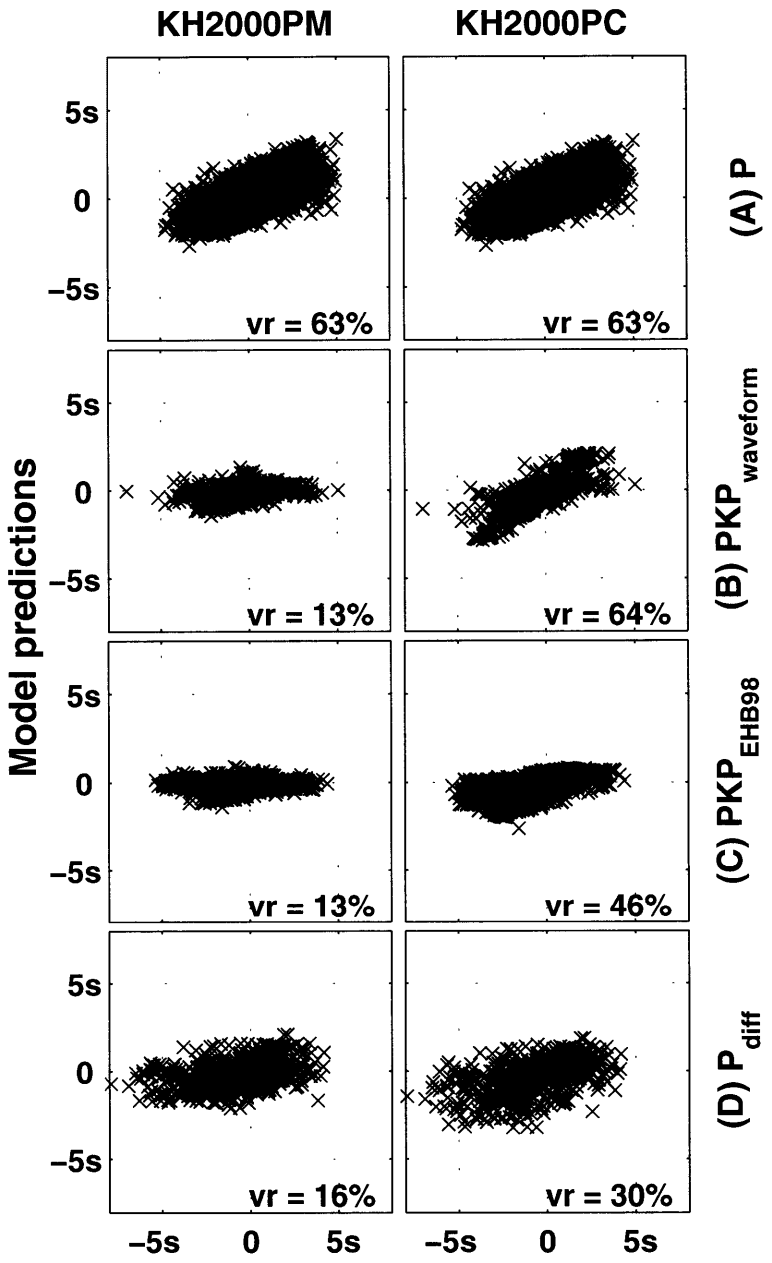


Figure 2-8: Scatterplots, original residuals vs model predicted residuals. (left) Model KH2000PM. (right) Model KH2000PC. (a) EHB98 $P+$ residuals from EHB. (b) residuals from waveform based PKP. (d) PKP residuals from EHB98. (d) P_{diff} residuals.

of magnitude. *Van der Hilst et al.* [1998] confirmed this conclusion but also showed that the pattern of PKP residuals measured by *McSweeney* [1995] is consistent with the model

by *van der Hilst et al.* [1997], which argues against data errors as the prime explanation for the amplitude discrepancy. Our model probably underestimates the amplitude of the wave speed perturbations owing to parameterization by means of constant (i.e., average) wave speed blocks and to regularization, in particular in regions of poor data coverage. For the model, in general, the amplitude amplification implied by *Song and Helmberger* [1997] is not realistic, but a substantial increase is possible in areas sampled by the PKP used in their analysis but not by our EHB98 $P+$ data. By including the core phases in our tomographic inversion we improve the fit (Figure 5, right), in particular of the waveform based PKP differential travel time residuals (Figures 5b and 5c).

Figure 5 demonstrates that the joint inversion has significantly improved the fit of the core phases without degrading the fit of P data itself. The smallest variance reduction is achieved for P_{diff} (Figure 5d): 16% when they are not included and 30% when they are. Since the P_{diff} signal is averaged over large Fresnel zones the small variance reduction suggests that we still underestimate the amplitude of long-wavelength structure in the lowest mantle. However, we decided not to weight this data set further because effects on the sensitivity kernels of CMB topography, earthquake mechanisms, and radiation patterns, and unmodeled finite-frequency effects may combine to a larger uncertainty.

2.4.3 Resolution

In addition to quantifying the data fit we have investigated how the addition of the extra data increases our ability to reconstruct Earth's structure. Calculation of even an approximate resolution matrix is computationally very expensive [*Nolet et al.*, 1999; *Vasco et al.*, 1999]. With our inversion method we performed test inversions with synthetic data calculated from different input models by multiplying the sensitivity matrix with the input structure, $\mathbf{\Delta t_{syn}} = \mathbf{L \Delta s_{syn}}$. We used the same regularization as for the inversion of the observed data. We performed two sets of inversion tests. In both cases the input model is a layer

consisting of sinusoidal wave speed variations with a peek wave speed perturbation of 2% (see scalebar and input patterns at the bottom of Plate 4). The first set has half wavelengths ranging from ~1800 km at the CMB up to ~2800 km at 1000 km depth. The second test has half the wavelength of the first one. Plates 4a-4d show, at the same scale as the input models, the recoveries of such a pattern placed at depths of 1000, 1500, 2000 km, and just above the CMB. For the CMB we show the result from the longer-wavelength test since structures with characteristic length ~900 km are on the margin of our resolving power.

These results were obtained using all data, that is the EHB98 $P+$ and the PKP and P_{diff} differential travel time residuals. With the caveat that such tests primarily depend on ray coverage and do not account properly for data error, these results suggest that the lateral resolution of structure at wavelengths is excellent beneath most geographic regions, except beneath the central Pacific and the higher latitudes of the Southern Hemisphere. We remark that lateral resolution does not seem to change much between 1000 and 2000 km depth, but it deteriorates toward the base of the mantle (beneath 2500 km depth).

Of our primary interest is to what extent the added data improve our ability to recover structure in the deep mantle. Plate 4e displays the result of a test inversion with EHB98 $P+$ data for a target just above the CMB using the longer wavelength pattern. Plate 4f shows the difference between this result and the recovery obtained when the core phases are included (Plate 4d). Visual inspection of Plates 4d and 4e shows that overall the recovery of both the heterogeneity pattern as well as the amplitude improves when the core phases are included. In many regions the improvement is small, but, as expected from the differential sampling (Figure 2g), the recovery is improved beneath South and Central America, Europe, north Asia, and the Australia-western Pacific region (Plate 4f). In these regions the structure, the lateral gradients, and the amplitudes are better reproduced. We remark, however, that even after addition of the new data the recovery remains problematic south of South Africa and in the southern Pacific.

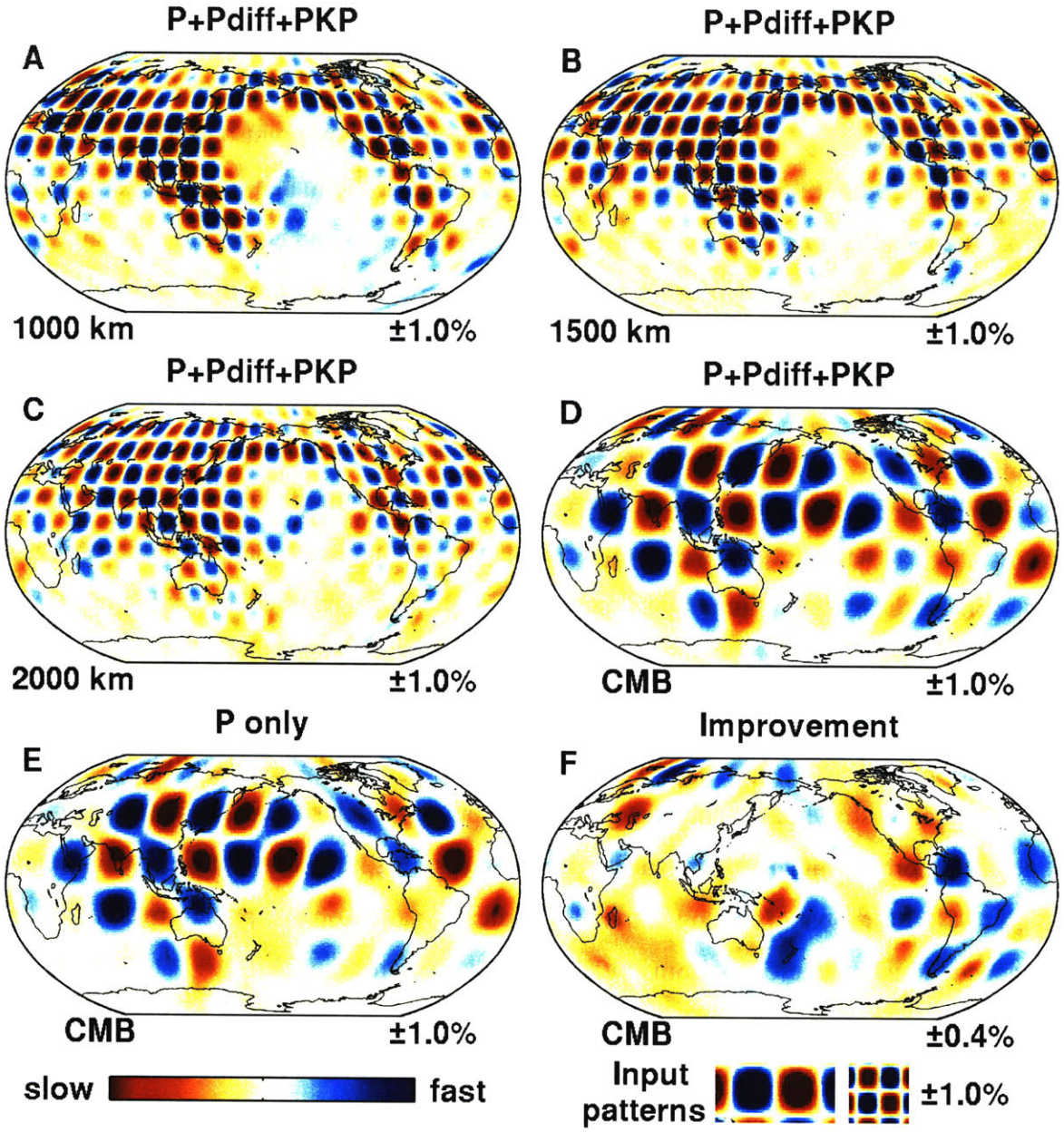


Figure 2-9: Inversion tests using harmonic input patterns with peek values ±2% (see scale bar and input patterns). Recoveries are shown for different data sets. (a)-(d) Same data sets and damping as used for model KH2000PC. (e) Same data sets and damping as used for model KH2000PM. (f) Improvement at the CMB, i.e., the difference between the two recoveries in Plates 4d and 4e.

To quantify the gain from the additional data, we calculated the change in model variance by subtracting the recovery from the input model; this measure is mostly sensitive to amplitude differences. Furthermore, we performed a point-by-point correlation between the input and output model, which provides a measure of the match in heterogeneity pattern. The laterally averaged values of the variance reduction and correlation are shown in Figure 6a as a function of depth for both tests. These results suggest that on average, and in most of the depth range of the lower mantle, we recover $\sim$70% of the true variance and match $\sim$90% of the long wavelength heterogeneity pattern (solid lines) if the core phases are included. Near the base of the model these values drop to $\sim$55% and $\sim$85% for amplitude and pattern, respectively, suggesting that we can map the geographic distribution of wave speed variations rather well, but even with the core phases added, we underestimate, on average, the true amplitude by as much as a factor of 2. The effect of adding the PKP and P_{diff} becomes noticable at $\sim$2000 km depth, but it remains small until we reach the bottom 300 or so km of the mantle; without the core phases, only 45% of the variance is recovered and only 75% of the pattern. It is harder to resolve the smaller scale structure (Figure 6a shaded lines) and the gradient damping, which suppresses sudden changes, becomes more dominant as the wavelength decreases. Still the correlation remains fairly high ($\sim$70% for most of the lower mantle), indicating a reasonable pattern match, but the variance reduction decreases from $\sim$40% $\rightarrow$ $\sim$20% towards the CMB.

The input models have structures only in the specific layer that is being tested but during the inversion some signal is "smeared" to other layers. We investigated this radial smearing by calculating the root-mean-square (rms) of the recoveries, layer by layer, as can be seen in Figure 6b for the longer-wavelength test with all phases. All the values shown here, including the input (shaded bars), are relative to the recovered rms of the layer being tested. As is to be expected with steeply descending rays the ability to constrain the radial extent of structures is smallest in the top few hundred kilometers. The radial smearing is however

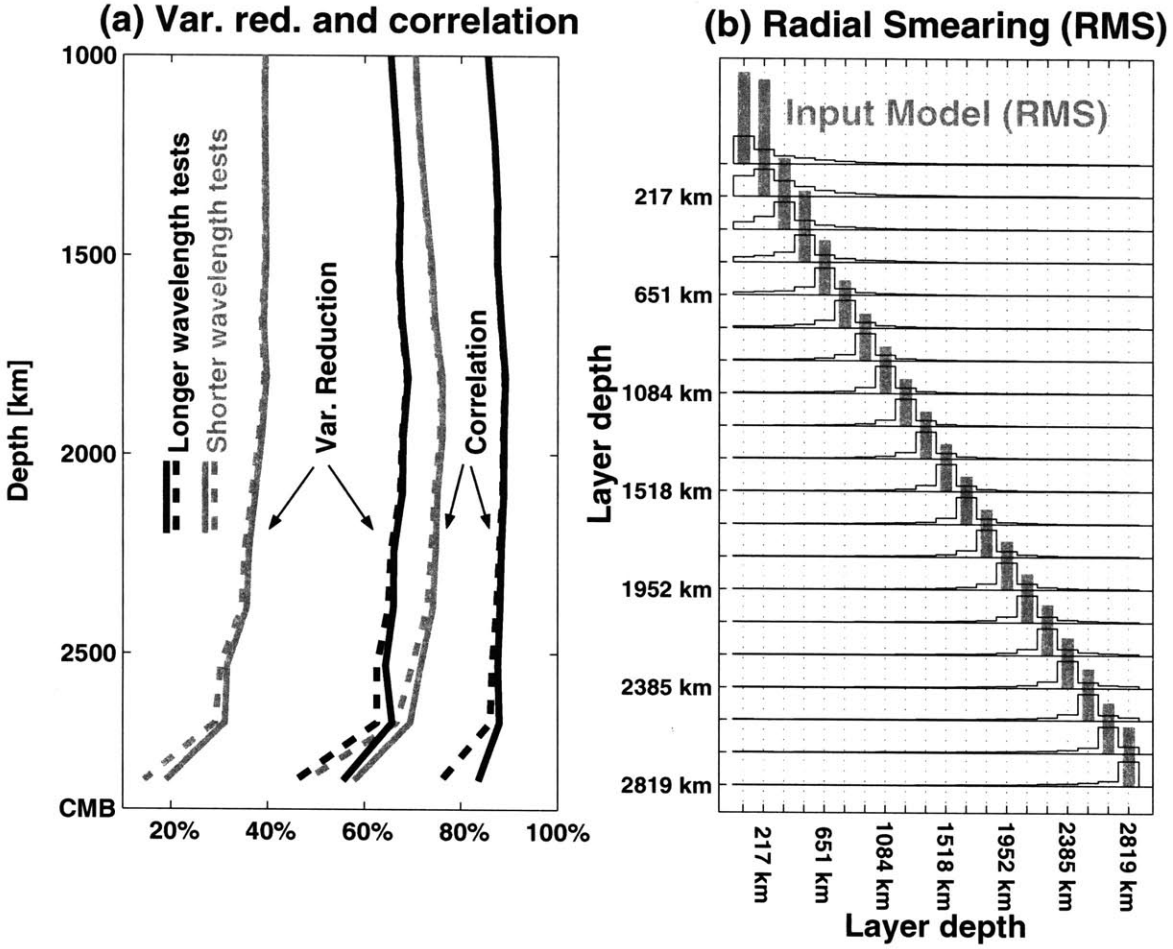


Figure 2-10: The two pairs of curves on the right depict correlation as function of depth. Solid unbroken line: Correlation between the longer wavelength input pattern (Plate 4, inset) and the recovery using mantle and core phases (e.g. Plate 4d). Solid dashed line: Same except using mantle phases only (e.g. Plate 4e). Shaded lines: Same as solid lines except for the shorter wavelength pattern (Plates 4a-4c and inset). The two pairs of curves on the left depict variance reduction. Solid unbroken line: The reduction of variance when subtracting the recovery from the input pattern using mantle and core phases. Solid dashed line: Same except using mantle phases only. Shaded lines: Same as solid lines except for the shorter wavelength pattern. (b) Vertical smearing as inferred from resolution tests with the longer wavelength structure (shaded bars) placed in one layer at a time. The recovery reveals some "smearing" to adjacent layers (solid lines). Values are scaled relative to the maximum rms of the recovery.

similar over the most of the lower mantle and no significant deviations in the depth interval 1000 km to 2500 km are apparent; both the rms of the recovery in the input layer and the

level that is smeared to adjacent layers are fairly constant. In the deepest of mantle the smearing is still similar, but the recovered rms is slightly lower, consistent with Figure 6a. The short wavelength test gives a similar result except for lower rms in the recovery.

These inversion tests should, however, be interpreted with caution [*van der Hilst et al.*, 1993]. First, such tests primarily reflect data coverage but do not assess the influence of data quality. Second, we do not know the true noise level of the data, which makes it hard to judge what level of damping is needed for such experiments and how (systematic) errors propagate into the model. Third, resolution assessed this way depends on the length scale of the input heterogeneity [*Lévêque et al.*, 1993]. Ideally one would calculate the resolution kernel for each model parameter but that is computationally prohibitive and even approximations take a long time to calculate [*Nolet et al.*, 1999; *Vasco et al.*, 1999]. Finally, implicit in the way the test inversions are conducted is that we know the propagation paths and the earthquake locations so that the nonlinearity of the problem is much less severe. In particular, we use the same (linear) theory for the synthesis and the back-projection of the data, and for both the forward and inverse problem the Earth is assumed to be isotropic.

2.4.4 Lower Mantle P Wave Speed

Selected layers from the final model are shown in Plate 5. The long and narrow features that emerge in the upper mantle transition zone are prominent around 1500 km depth beneath North and South America and beneath the southern margin of Asia. These have been associated with plate motion history and are thought to be the remnants of old subducted slabs [*Woodhouse and Dziwonski*, 1989; *Grand*, 1994; *van der Hilst et al.*, 1997; *Grand et al.*, 1997]. At larger depth these structures seem to disintegrate and at about 2000 km no such large coherent fast structures are visible. Our resolution tests (e.g., Figure 6b) indicate that these observations can not simply be explained by changes in data coverage or "smearing" in radial direction. Indeed, *Kárason and van der Hilst* [2000] show that

beneath many of the major convergent plate boundaries slabs of subducted lithosphere lose their wave speed signature (or disappear althogether) near 1,800 km depth. Judging from Plates 4a-4d and Figure 6 those changes cannot easily be attributed to differences in data coverage. In accord with other models [e.g., *Liu and Dziewonski*, 1998; *Castle et al.*, 2000] we find that the pattern of heterogeneity near the base of the mantle persists to several hundreds of kilometers above the CMB; it does not connect to the midmantle pattern but gradually breaks down and is absent at depths $<\sim$2300 km. However, some slow anomalies probably extend to much larger distances above the CMB [e.g. *Ritsema et al.*, 1998, 1999; *Mégnin and Romanowicz*, 2000].

Near the base of the mantle the most prominent features are the fast anomalies beneath Central America and Asia and the slow anomalies beneath the Pacific, west Africa, and the South Atlantic. This structure is generally consistent with results of shear wave tomography (although differences exist, notably beneath Australia and Alaska where shear wave speed is high and P wave speed low) but has not been borne out so well by our previous P inversions. In a spectral analysis the structures of higher-than-average P wave speed around the Pacific would strongly contribute to a spherical harmonics degree 2 signal, but it is probably not a continuous structure.

Figure 7 shows the rms of the models without and with core phases (KH2000PM and KH2000PC, respectively) and the correlation between them as function of depth. Comparison of the images (Plate 5) and analysis of model correlation (Figure 7, shaded dashed line) and radial changes in rms amplitude (Figure 7, black lines) reveals that the models are nearly identical in structure and amplitude to a depth of about 1500 km. At larger depths the models slowly become more distinct and, in particular, the rms amplitude in model KH2000PC increases much more rapidly with depth than in model KH2000PM (Figure 7), especially in the slow regions. The correlation between the two models (Figure 7) begins to decrease at $\sim$1500 km depth, whereas the lateral resolution is the same to $\sim$2000 km

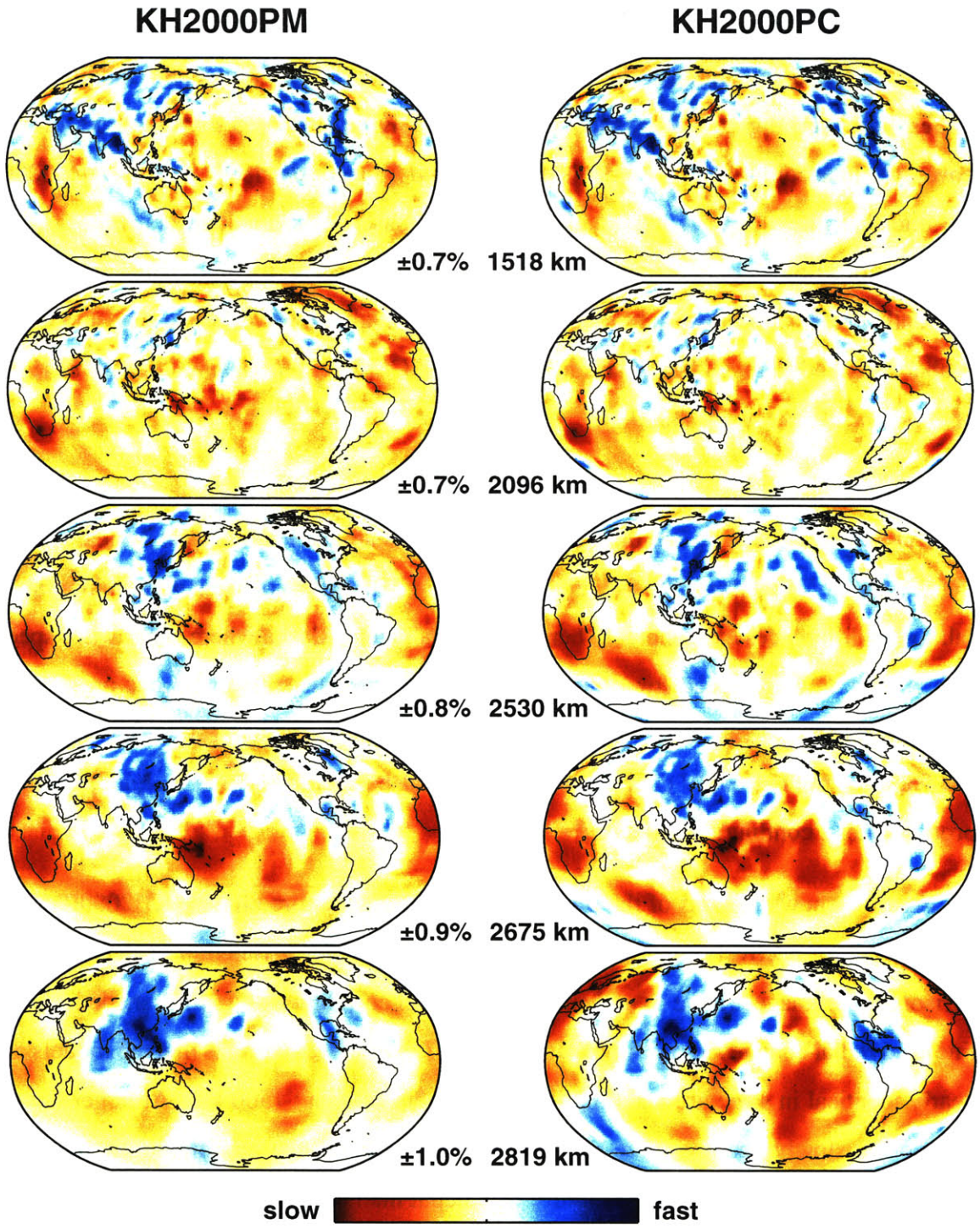


Figure 2-11: Selected layers from model based only on EHB98 $P+$ (KH2000PM) and the final model based on all data (KH2000PC).

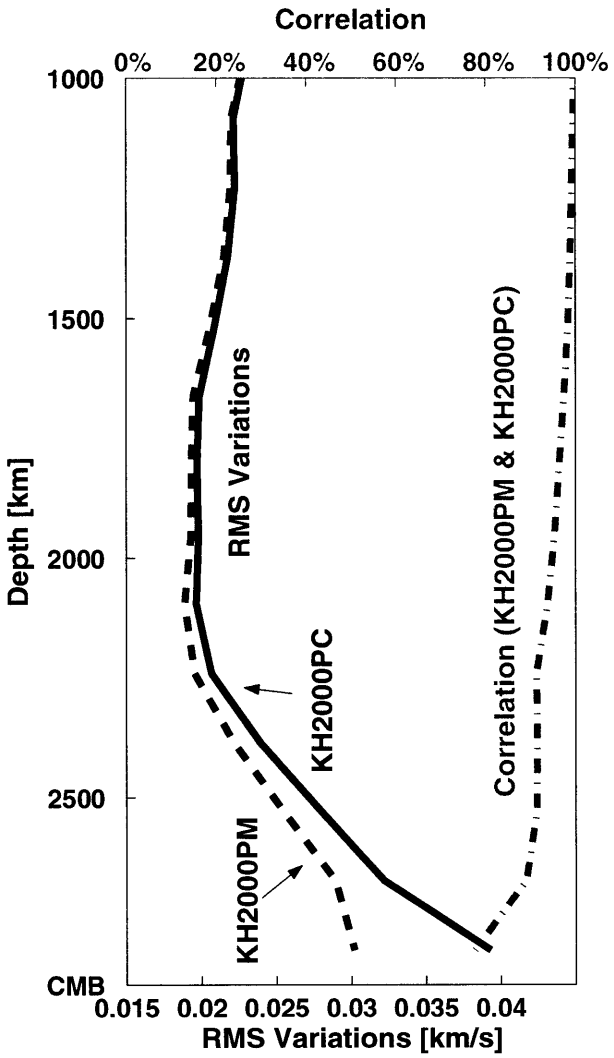


Figure 2-12: The rms value (solid lines) of wave speed amplitudes for models KH2000PM (dashed line) and KH2000PC (unbroken line). Correlation between the models is depicted by the shaded line.

depth (Figure 6a). This discrepancy is probably due to small scale (few hundred kilometers) structure that was not represented in the checkerboard models used and to smearing along the mantle legs of PKP and P_{diff}.

2.5 Trade-off With Core Structure

2.5.1 Inner Core

Differential times involving PKP_{DF} can map signal from the inner core into the mantle and vice versa. For the results presented in Plate 5 we have corrected the travel times of PKP_{DF} branch according to a model for inner core anisotropy by *Su and Dziewonski* [1995]. However, there is no consensus on inner core anisotropy and this correction does not account for any isotropic heterogeneity. Therefore, we carried out three test inversions to assess the dependence of our model on inner core structure. In the first we excluded near polar PKP_{DF} paths since they are most sensitive to axial anisotropy. In the second we simply omitted the correction for inner core anisotropy. In the third we adapted the model parameterization and introduced $\sim$1300 cells in the inner core and allowed them to absorb as much signal in the PKP_{DF} travel time residuals as they can. The models resulting from tests two and three can be considered as end-member cases regarding the influence of inner core structure on our mantle model results. Comparison with the result discussed above (Plates 3d and 5) shows that in general the pattern of deep mantle structure is, for our set of differential times, virtually independent of our treatment of inner core structure. However, the amplitude of the high wave speed anomalies in the deep mantle beneath Alaska and between South Africa and Antarctica should be interpreted with caution.

2.5.2 Outer Core

Vasco and Johnson [1998] suggested the existence of aspherical wave speed variations in the liquid outer core with an rms amplitude of $\sim$50% of that of mantle structure. We investigated the trade-off between mantle and outer core structure with a series of test inversions and discuss the result by means of radial profiles of rms wave speed perturbations (Figure 8).

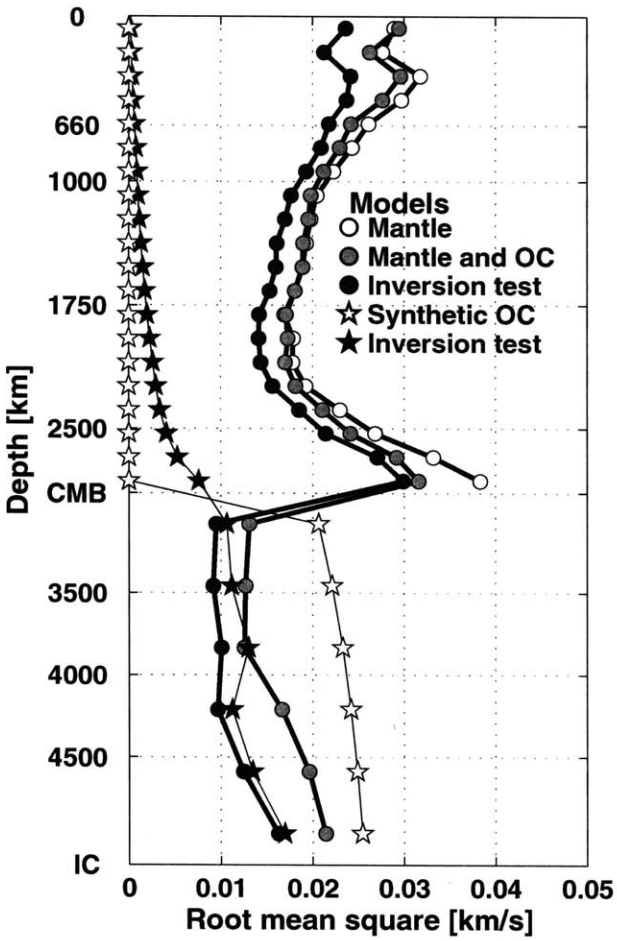


Figure 2-13: The rms of wave speed variations within each layer for models based on EHB98 $P+$ and both PKP data sets. Open circles indicate model confined to the mantle. Shaded circles indicate model of mantle and outer core. Solid circles indicate recovery of inversion test using the model confined to the mantle as an input. Open stars indicate synthetic structure with $\pm 1\%$ wave speed variations on $30° \times 30°$ cells within the outer core. Solid stars indicate recovery of synthetic outer core structure.

First, we adapted the parameterization (see section 3.2) and inverted our data set for aspherical structure in both the mantle and the outer core. Figure 8 (shaded circles) suggests that the data allow significant rms variations in the liquid outer core, albeit at lower amplitudes than in the mantle, which is consistent with *Vasco and Johnson* [1998]. Comparison with the result from inversion for mantle structure only (open circles, Figure 8) suggests

that the trade-off concerns the strong heterogeneity in the mantle beneath 2300 km and, surprisingly, above 1000 km depth.

Second, we inverted synthetic data calculated using our model KH2000PC (i.e., no wave speed perturbations in the outer core) for both mantle and core structure. No minimization of the radial component of the gradient across the CMB was applied. Figure 8 (solid dots) demonstrates that potentially the mapping of mantle structure into the outer core is significant. Apart from a slight drop in amplitude, which is typical in such tests, the resulting radial rms profile is rather similar to the one obtained from the inversion of all original data (shaded circles). This suggests that the structure inferred from tomographic inversion of P and PKP data (Figure 8, shaded circles) can be explained by contamination of mantle structure and that outer core structure itself may not contribute much, if anything, to the structural signal in the PKP data used in our study.

Third, we inverted synthetic data calculated for a model with structure in the outer core only (Figure 8, open stars). In this case, the mantle structure is set to zero so that all synthetic P data are zero, which results in a very significant bias to small amplitude in the inversion result. Despite this severe condition, some outer core structure leaks to lower mantle depths (Figure 8, solid stars), presumably along PKP branches that sample regions devoid of EHB98 $P+$ paths. However, even if we take the rms of core structure comparable to that of the lower mantle the rms of the deep mantle artifacts is almost 1 order of magnitude less than the lateral variations inferred from the real data (Plate 5).

2.6 Summary and Conclusions

We have improved the constraints on variations in lower mantle P wave speed by incorporating PKP and P_{diff} travel time information. These phases sample the deep mantle in a different way than the P phases used in our previous global inversions, and resolution tests suggest that addition of these phases slightly improves the images of structure in the

deep mantle. However, integrating data sets that differ substantially in size, quality, and frequency content is neither trivial nor unique. The integration of the different data was guided by the improvement of data fit for all data involved. We have inverted the high-quality differential travel time residuals for PKP and P_{diff} phases and also used them as the reference for the reduction of the PKP residuals from the data set by *Engdahl et al.* [1998]. The long period PKP_{DF}-P_{diff} data have large Fresnel zones, and back projection of the data along infinitesimally narrow ray paths is not appropriate. To account for finite frequency effects we have developed a technique to calculate realistic estimates of 3-D sensitivity kernels, based on exact Fréchet kernels calculated by L. Zhao (personal communication, 1999) using normal mode summation with consideration of cross-branch mode coupling [*Zhao and Jordan*, 1998; *Zhao et al.*, 2000]. The use of these kernels allows us to properly distribute the sensitivity for a given seismic phase over a large mantle volume while allowing the high-frequency data to constrain small-scale structure. Finally, we investigated the trade-off between mantle and core structure and the possible contamination of our mantle model with signal pertaining to inner core anisotropy. From these investigations we reach the following conclusions:

1. The inversion of the core phases along with the short period P, pP, and pwP data resulted in an aspherical model for compressional wave speed that explains a large fraction of the structural signal in the core phases without degrading the fit to the other data used.

2. On the basis of the improved data fit and results of our resolution tests, we conclude that the new data have enhanced the lower mantle part of our model. In particular beneath the Southern Hemisphere the amplitude of the wave speed perturbations has increased by a factor of 2 and the spatial definition of the wave speed anomalies improved. However, many features in the "old" models have proved to be robust. Our tests reveal that spatial resolution does not change much from the middle to the lowermost mantle which suggests that the transition near 2000 km depth as described by *van der Hilst et al.* [1997, 1998];

van der Hilst and Kárason [1999], and *Kárason and van der Hilst* [2000] is not an artifact of variations in data coverage.

3. Differential travel time residuals are not as effective as commonly assumed in isolating the source region of the structural signal [i.e., *Wysession*, 1996; *Castle et al.*, 2000]. Specifically, the differential pairs used in our study (PKP_{AB}-PKP_{DF}, PKP_{AB}-PKP_{BC}, and PKP_{DF}-P_{diff}) are sensitive to small-scale structure anywhere in the bottom half of the lower mantle, and significant smearing can occur in the lower mantle along the subvertical PKP and P_{diff} paths.

4. Even with differential times it seems very difficult to constrain outer core structure independent of the (stronger) heterogeneity in the mantle. Test inversions of observed data or of synthetic residuals that do not contain core signal give similar rms amplitudes for the outer core. This is consistent with the presumption that the contribution of the outer core to signal in the PKP data is small and all or most of the apparent outer core structure could be explained by mapping ("smearing") along PKP paths of mantle structure that is not sufficiently well constrained by the P data. This may also apply to the inner core but we did not test that explicitly. The strong heterogeneity near the base of the mantle, in the so called D" region is not the only source of the signal that is mapped into the core. Indeed, structure anywhere in the mantle, but in particular in the top and bottom 1000 km, can produce spurious core structure. We cannot rule out outer core heterogeneities, but our data are consistent with the conventional view [*Stevenson*, 1987; *Jellinek et. al*, 1999] that the outer core is effectively homogeneous.

Acknowledgments.This work would not have been possible without the waveform data generously provided by Michael Wysession, Tom McSweeney, and Ken Creager, and without the 3-D sensitivity kernels calculated for us by Li Zhao. We are also grateful to E.R. Engdahl for processing the extended data base of routinely reported phase arrivals used in this study and for the constructive comments by Jeroen Ritsema, an anonymous reviewer,

and the Associate Editor. This research is supported by the National Science Foundation under grants EAR-9627087 and EAR-9969492.

Chapter 3

Three-Dimensional Mantle Structure From Low Frequency PP-P Differential Travel Times[1]

Abstract

We used ~20,000 PP-P differential travel times to improve our global P-wave model, in particular for the upper mantle beneath the intra-plate regions poorly sampled by the (P, PKP, P_{diff}) used previously. The added data were measured by cross-correlation of the PP arrival with the *Hilbert* transform of P, corrected for attenuation. The low frequency (40 mHz) PP-P data are relatively insensitive to structure beneath source and receiver but have a wide and complex sensitivity zone near the bounce point where upper mantle heterogeneity is generally strong. For data measured at 1 Hz (P, PKP) we use a high frequency approximation and trace rays through a fine grid of constant slowness cells to invert for mantle structure. For low frequency PP-P we account for sensitivity to structure away from the optical ray path with 3-D Fréchet derivatives (sensitivity kernels) estimated from single forward scattering and projected onto basis functions (constant slowness blocks) used for model parameterization. With such kernels the low frequency data can constrain

[1]This paper is under review as: H. Kárason, R. D. van der Hilst and G. Masters, Three-Dimensional Mantle Structure From Low Frequency *PP-P* Differential Travel Times, in *Geophysical Research Letters* (submitted Dec. 2001).

long wavelength heterogeneity without keeping the short period data from mapping details in densely sampled regions. Inversion tests, spectral analysis, and comparison with geology indicate that the large-scale upper mantle structure is better constrained with the addition of PP-P.

3.1 Introduction

Surface wave dispersion and free oscillation data have produced robust constraints on long wavelength shear structure in Earth's interior (see, for instance, *Masters et al.*, [2000]). Tomography with travel times of high-frequency body-waves provides complementary information and has been producing increasingly detailed images of trajectories of convective flow in Earth's mantle [e.g., *Grand et al.*, 1997, *Van der Hilst et al.*, 1997, *Bijwaard et al.*, 1998], revealing complex structures of subducted slabs in the upper mantle transition zone [e.g., *Van der Hilst et al.*, 1991, 1997] and mid-mantle [*Kárason and Van der Hilst*, 2000]. However, the spatially variable data coverage by body waves, due to uneven source and receiver distribution and the 3D path geometry of the seismic phases considered, has left large mantle volumes unexplored and our understanding of mantle structure and dynamics incomplete.

Kárason and Van der Hilst [2001], hereon *KH2001*, improved the sampling of lower mantle structure by including carefully measured PKP and P_{diff} data; here we aim to improve the P-wave imaging of the shallow mantle. Unfortunately, the fundamental and higher-mode surface wave dispersion that has been used so successfully for upper mantle S-wave imaging has limited sensitivity to variations in compressional wave speed. Surface wave models could be used to constrain upper mantle P-wave models by assuming a scaling between P- and S-wavespeed but this would prevent a meaningful analysis of wavespeed ratios.

The P-wave sampling of the shallow mantle can be improved with multiple surface reflections (i.e., PP, PPP, $PPPP$ etc., Figure 1a). Here we use PP-P differential travel

time residuals. Since they are relatively insensitive to structure close to the source and receiver, PP-P residuals are often mapped onto bounce point bins at Earth's surface. Maps thus produced reveal good correlation with large scale geological features (e.g., *Woodward and Masters* [1991]) but cannot be integrated self-consistently with other data to construct a 3-D model. Moreover, such an approach ignores the sensitivity of the PP-P residuals to lower mantle structure sampled along the P paths and to small scale structure in the transition zone. Therefore, we follow *Van der Hilst and Engdahl* [1991] and use the PP-P data in a tomographic inversion for 3-D mantle structure. For the proper integration of the low-frequency PP-P data with the short period P and PKP data used by *KH2001* we use 3-D kernels to account for sensitivity away from the optical ray paths. Here we explain the data integration and discuss the effects of adding the PP-P data on resolution and on the images of upper mantle P-wave speed. The model itself will be described in more detail elsewhere [*Kárason and Van der Hilst*, in prep.].

3.2 Data

The surface-reflected PP waves pass through an internal caustic [*Jeffreys and Lapwood*, 1957], which has two important implications for seismic imaging that aggravate with decreased frequency (see also *Van der Hilst and Engdahl* [1991]). The waves undergo a 90° phase shift, which results in a waveform distortion (compared to the direct wave) that has to be accounted for in waveform and travel time analyses. The PP waveform can be described as the Hilbert transform of the direct P-wave [*Choy and Richards*, 1975]. Furthermore, it is a mini-max travel time phase and the travel times describe a saddle plane: the travel time increases (decreases) with increasing distance between the actual and the ray geometrical bounce point in a direction perpendicular (parallel) to the great circle connecting source and receiver.

For short-period data used by, e.g., *Van der Hilst and Engdahl* [1991], *Vasco et al.*,

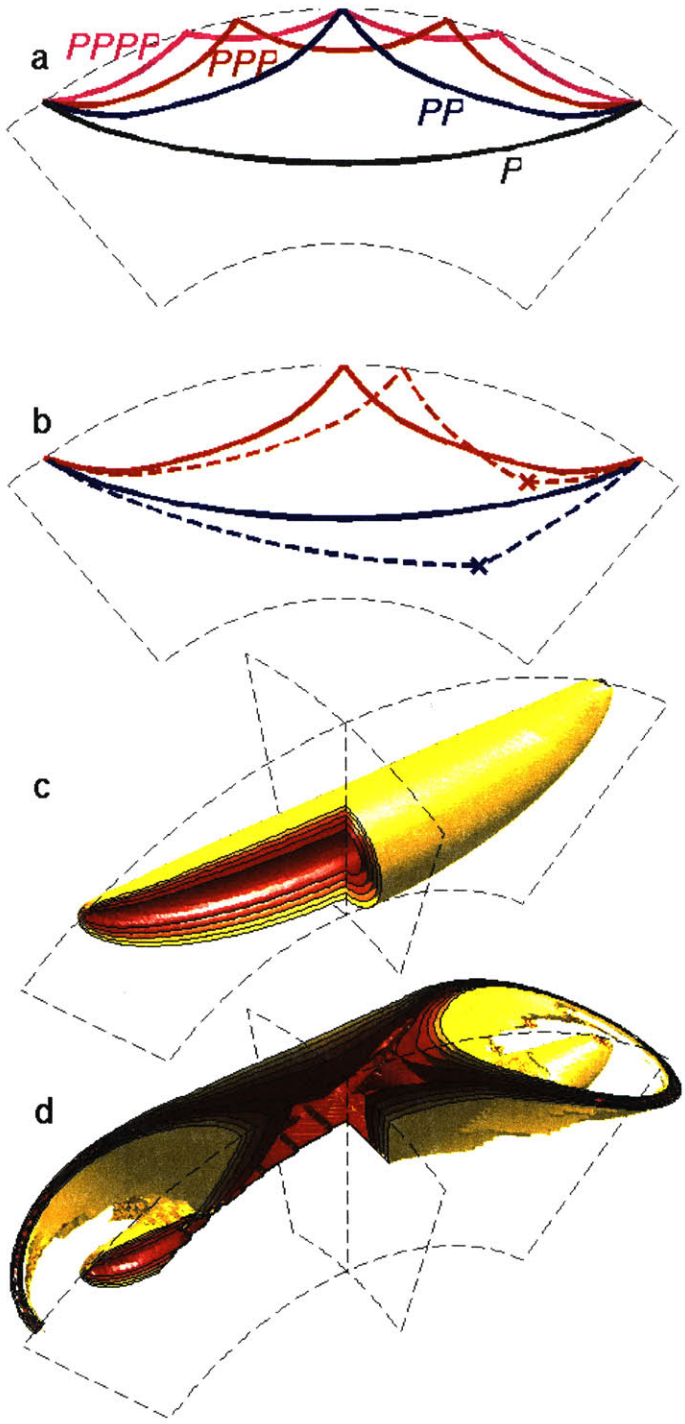


Figure 3-1: (a) Ray paths for P, PP, PPP and $PPPP$. (b) Direct (solid lines) and scattered (broken lines) ray paths for P and PP. (c) Travel time surfaces for P and (d) PP (see text).

[1998], and *Gorbatov et al.* [2001] it is reasonable to ignore the waveform distortion and the saddle plane geometry of the travel time sensitivity. However, since the routinely processed PP data from ISC (International Seismological Centre) are rather noisy, we chose to rely on PP-P times that were accurately measured from digital data, and for the frequencies considered (40 mHz) the waveform distortion is significant and has to be accounted for.

Following *Woodward and Masters* [1991], the differential travel time residuals were determined by cross-correlating the Hilbert transform of the P arrival with PP. A t^* operator is used to correct for the difference in average attenuation along the P and PP paths, and the differential times are corrected for crustal structure at the PP bounce point using the high resolution crustal model, CRUST2.0, due to *Bassin et al.* [2000]. PP is analyzed for epicentral distances beyond $48.5°$ to avoid triplications due to upper mantle discontinuities, and PP-P times between -8.0 and +7.5 s from the ak135 [*Kennett et al.*, 1995] reference values were kept for further processing. The records for earthquakes between 1978 and 1999 yielded 20,329 picks. Assuming that most of the signal is due to structure near the bounce point, the data are further reduced by removing the 63 residuals that differ by more than 5 s from the average residual in circular bounce-point caps of $3°$ radius, resulting in data set of 20,266 picks.

The PP-P thus determined were inverted along with the data used by *KH2001*: P, pP and pwP residuals from the catalogue by *Engdahl et al.* [1998], hereon EHB, measured at 1 Hz, and the differential travel-time residuals of core refracted (PKP, at 1 Hz) and diffracted (P_{diff}, at 50 mHz) waves. To balance them against the larger (and more noisy) EHB data set, the PP-P data are weighted according to the number and quality of the data and the variance reduction that one wants to achieve (see *KH2001* for a discussion pertinent to PKP and P_{diff} data). The model based on all data will hereinafter be referred to as KHM2001.

3.3 Method

Another consequence of using low frequency PP-P data is that near the ray geometrical bounce point the Fresnel zone can be many thousand km across [e.g., *Neele and Snieder*, 1992]. If PP-P were to be used alone this could be accounted for by regularization and smoothing, but in the joint inversion envisaged here this would degrade the ability of the short-period data (e.g., P, pP, PKP) to resolve finer details in regions of dense data coverage. Like *KH2001*, we use narrow rays for the short period data and design realistic 3-D sensitivity kernels for the low frequency data; both types of kernel are then projected onto the constant slowness cells (2.8° by 2.8° by 90 km) that are used to parameterize the inversions. The smoothing in the model space is thus largely guided by the frequency dependent Fréchet derivatives themselves instead of by regularization (e.g., *norm* or *gradient* minimization).

For the construction of the finite frequency kernels we assume that the travel time sensitivity to structure is zero outside the first Fresnel zone, which is commonly defined using single scattering (Figure 1b). If T, T_1, and T_2 are the source-receiver, source-scatterer, and scatterer-receiver travel-times, respectively, the travel-time deviation ΔT can be defined as $\Delta T = (T_1 + T_2) - T$. The first Fresnel zone is then estimated as the locus of all scatterers for which $0 \leq f\,|\Delta T| \leq \frac{1}{2}$ where f is the dominant frequency of the waveform [*Kravtsov and Orlov*, 1990]. The PP-P data have a dominant frequency of 40 mHz so that $\Delta T \sim$12.5 s is the limit of the first Fresnel zone. Figure 1c and 1d depicts the absolute travel-time surfaces for P and PP, respectively, with one quadrant removed. The yellow surface is for ΔT=12.5 s and the darker colors correspond to smaller ΔT. The P kernel has a relatively simple Śbananaš shape, but the PP kernel is more complicated because of its mini-max travel time nature [see also *Neele et al.*, 1992]. The differential PP-P kernel is obtained by subtracting the P from the PP kernel [*Van der Hilst and Engdahl* 1991, *Dahlen et al.* 2000].

The variation of the sensitivity within the kernel is subject to debate [*Marquering et al.*, 1998, *Zhao et al.*, 1998, *Hung et al.*, 2000]. We took a pragmatic view and investigated the effect on the tomographic images of three distributions of the sensitivity as a function of ΔT: (i) constant sensitivity over first Fresnel zone (i.e., for $\Delta T \leq 12.5$ s), (ii) maximum sensitivity at the geometrical ray (constant to ΔT=6.25 s and a cosine squared taper to zero at ΔT=12.5 s), and (iii) zero sensitivity at the ray (i.e., a Śbanana-doughnutŠ type of kernel). For each we constructed a tomographic model with the PP-P data. The similarity of the results suggests that for the data coverage considered here the distribution of sensitivity over the kernels is not the most critical parameter, and for the results presented we use a sensitivity function as in (ii). The precise shape of the kernels is likely to become more important when denser sampling by high quality data requires less regularization to stabilize the inversion.

3.4 Results

Figure 2 illustrates different aspects of the PP-P data and their inclusion in the tomographic inversions, both for the shallow and deep mantle.

Consistent with *Woodward and Masters* [1991], the bounce-point cap averages of the PP-P data (Figure 2a) correlate well with known geological features (e.g., old seafloor and continental areas are generally fast and geologically active regions slow).

Tomographic inversion of the PP-P data for 3-D structure yields smoothly varying structures in the shallow mantle (Figure 2b) that are largely consistent with the bounce-point projection (Figure 2a). Noteworthy differences include the slow zone along the North American west coast, which has narrowed upon formal inversion, and the continent, which now stands out by faster wave propagation. The low wave speeds beneath the Himalayas have become more pronounced and localized upon inversion. For the shallow mantle, the result of PP-P inversion (Figure 2b) is smoother than but otherwise consistent with the

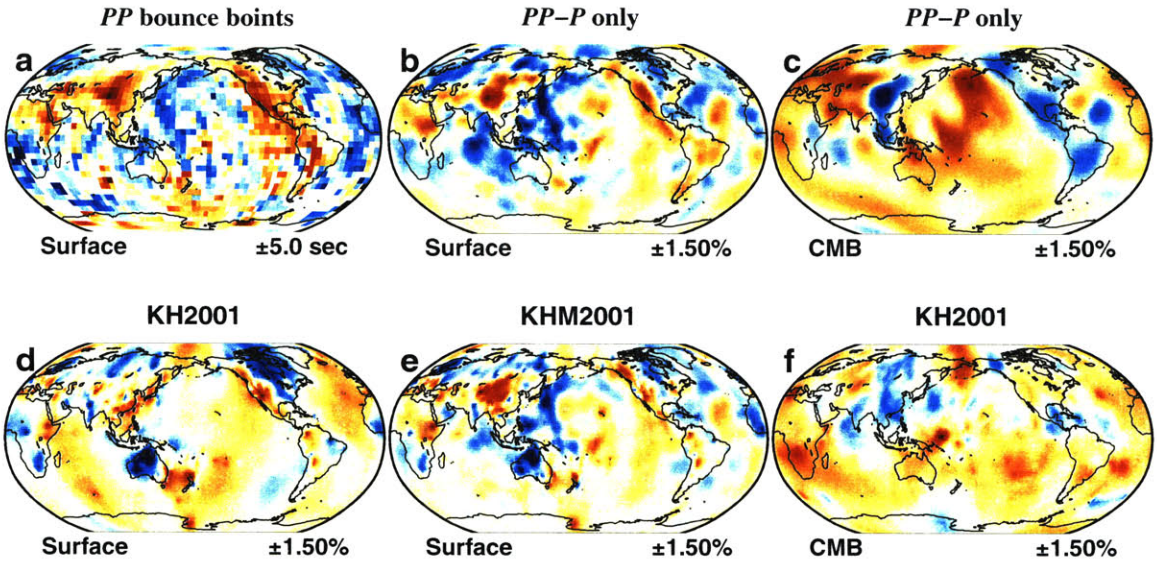


Figure 3-2: (a) Bounce-point cap (5° radius) averages of PP-P differential travel-times residuals. (b) Model based on PP-P at 100 km depth and (c) above the CMB. (d) KH2001 model based on same data as *KH2001* at 100 km. (e) KHM2001 model based on all data at 100 km depth (d) KH2001 above the CMB.

structure according to model KH2001 (Figure 2d), but the latter lacks definition of structure in oceanic and other aseismic intra-plate regions. For instance, the PP-P data depict South Africa as fast, but can not resolve the individual cratons. Lateral wavespeed variations beneath Australia and the slow P wave propagation beneath western North America are present in both models.

Joint inversion of all data sets (that is, model KHM2001) produces the image of upper mantle P-wavespeed shown in Figure 2e. The features revealed by KH2001 remain visible but the PP-P data constrain better the long wavelength variations away from seismic zones. Most notable is fast wave propagation in the old parts of the Pacific and North Atlantic oceans and the slow anomaly beneath the Himalayas. Spectral analysis of KH2001 and KHM2001 shows that inclusion of the PP data enhances the power in the upper mantle part at degrees $l = 5$ to $l = 15$, which corresponds to a half wave-length resolution near

1,500 km to 4,000 km. Even though the effect of including PP-P data is most pronounced in the top few hundred km, the PP-P data are also sensitive to deep mantle structure (along the P leg) and it is reassuring that the deep mantle features constrained by PP-P alone (Figure 2c) are compatible with KH2001 (Figure 2f).

3.5 Discussion

Demonstrating that the model based on the combination of data sets is actually better than the results from separate inversions is not trivial. Checkerboard tests suggest that the resolution of shallow large-scale structure is much improved (Figure 3), especially in the oceanic areas. While encouraging, one should realize that such resolution tests always suggest improved resolution when path coverage is improved, and that the evaluation of data quality and the accuracy of inverse and forward modeling is not trivial. As regards the latter, such resolution tests would have yielded satisfactory results even if our kernels - or the distribution of sensitivity within them - were incorrect.

Another indicator of data compatibility is the predictive power of models. Multiplying the sensitivity matrix based on the PP-P kernels with model KH2001 and comparing the result with the observed data (Table 1) suggests that KH2001 explains only $\sim$10% of the observed PP-P signal variance. This mainly reflects the fact that much of the upper mantle KH2001 sampled by PP is not sampled adequately by the data used by *KH2001*. The variance reduction for PP-P ($\sim$ 50%) upon joint inversion (i.e., resulting in model KHM2001) is similar to the data fit of the other data used to construct KHM2001, which is not degraded compared to the fit by KH2001. The fact that we Ślonly Š explain some 50% of the data may suggest that our forward theory is not correct, which compounds effects of measurement error and regularization. We also remark that such factors as the slope of the seafloor near the bounce point, the wave propagation through sediments and the water column, and the possible reflection at interfaces other than the free surface (see *Van der*

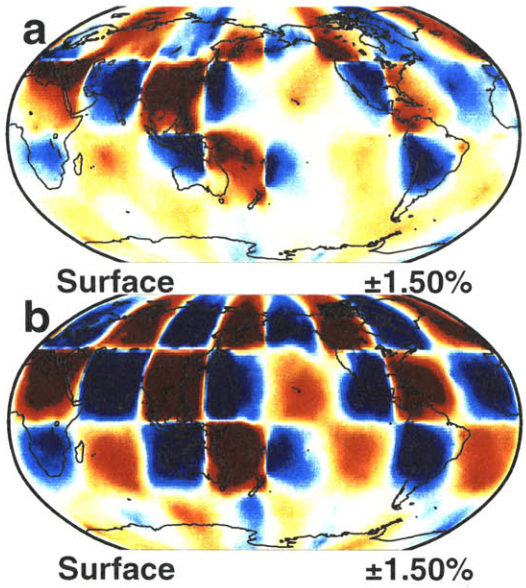


Figure 3-3: Recovery of input pattern at 100 km depth using same data coverage and inversion parameters as in (a) KH2001 and (b) KHM2001.

Hilst and Engdahl [1991]) are not accounted for in the interpretation of the PP-P data.

3.6 Concluding remarks

The data sets used in our joint inversions have specific strengths and weaknesses. Our recent research has focused on joining data sets that complement each other (this paper, *KH2001*). For that purpose we carefully select and process data sets and devise appropriate

Table 3.1: Variance Reduction

	KH2001	KHM2001
P:	53%	53%
PP-P:	10%	51%

forward theories (i.e., sensitivity kernels). While there is no unique way of integrating data, upon inversion we weight the different contributions - using data quality, amount, and variance reduction as guidelines - to create a balanced model.

Surface reflected P-waves (PP, PPP etc.) can overcome one shortcoming of our previous models, that is, the lack of resolution in the shallow mantle away from seismic sources and stations. By using the low-frequency waveform-derived PP-P, creating realistic kernels and projecting them onto a fine grid, we can simultaneously constrain detailed structure with high frequency data and long wave-length structure with the long period PP-P data. The resulting model, KHM2001 explains about 50% of the variance in the PP-P data. Spectral analysis confirms that the model contains small-scale structures, which will be detailed elsewhere [*Kárason and Van der Hilst*, in prep.], but that the PP-P enhances the power of long wavelength heterogeneity in the upper mantle. The improved resolving power of our augmented data set is further demonstrated with checkerboard resolution tests. The smooth features introduced by PP-P correlate well with old and young seafloor and geologically active regions which indicates that the improvements are real.

Chapter 4

A new model of mantle P-wave speed; advances in methodology and data handling

Abstract

We present a model that is the combined result of earlier studies. The mantle is imaged at varying length scales according to data coverage by our high frequency travel times. Our basis functions are cubes in latitude, longitude and radial space, the size of which reflects the quality of sampling. Superimposed onto this grid are broad three-dimensional sensitivity kernels that account for the sensitivity of low frequency data to large mantle volumes. The bulk of the data is P, pP and pwP from the catalogue by *Engdahl et al.* [1998], which is based on phase arrival information reported to the International Seismological Centre from January 1964 through August 1999 with phases re-identified and earthquakes relocated. To complement this data set in the deepest mantle we use short wavelength differential times of core phases, PKP_{DF}-PKP_{AB} and PKP_{BC}-PKP_{DF}, [*McSweeney*, 1995; *Engdahl et al.*, 1998] and a low frequency data set of P_{diff}-PKP_{DF} [*Wysession*, 1996]. To improve the resolution of shallow large-scale structures under oceanic regions we use low frequency PP-P differential times [*Kárason et al.*, 2002; *Woodward and Masters*, 1991, see also Chapter 3]. In accordance with earlier results, this model reveals relatively narrow fast structures around mid-mantle depths that have been linked to ancient subduction. They can in some places be traced to the surface but a connection to the very deep mantle is

only observed locally. The benefits of the irregular grid are illustrated with cross sections from nearly all active subduction zones, showing, among other details, the varying style of subduction in the shallow mantle and complex flow patterns around the transition zone between the upper and lower mantle.

4.1 Introduction

Seismic tomography provides direct information about mantle structure and the nature of mantle convection [*Montagner*, 1994]. In the past decade, an encouraging agreement has been established between the the long-wavelength structure of most global models [*Becker and Boschi*, 2002; *Boschi and Dziewonski*, 1999]. However, at wavelength less than 3000 km to 4000 km significant differences still exist. These differences are due to different types of data (e.g. P or S, surface waves, free oscillations), different parameterization, and different inversion techniques. In most cases data coverage is very uneven, due to the geographical distribution of earthquakes and stations, and consequently the length scale at which structure can be resolved is highly variable. The frequency of the seismic record from which data is derived also affects the length scales to which they are sensitive and lower frequency data typically constrain larger scale structure [*Marquering et al.*, 1998].

The use of different inversion methods has an effect on the appearance of tomographic models, in particular, the use of either a global or a local basis for the model parameterization. Using a global basis, such as spherical harmonics laterally and splines in radial direction, generally produces a smooth model [e.g. *Dziewonski*, 1984; *Su et al.*, 1994; *Li and Romanowicz*, 1996; *Masters et al.*, 1996; *Ritsema and van Heijst*, 2000]. Models based on constant wave speed blocks [e.g. *Inoue et al.*, 1990; *Pulliam and Stark*, 1993; *Vasco et al.*, 1994, 1995a; *Grand et al.*, 1997; *van der Hilst et al.*, 1997; *Vasco and Johnson*, 1998; *Bijwaard et al.*, 1998; *Kárason and van der Hilst*, 2001, and many others] can also be smooth if the regularization used in the inversion favors smooth lateral variations

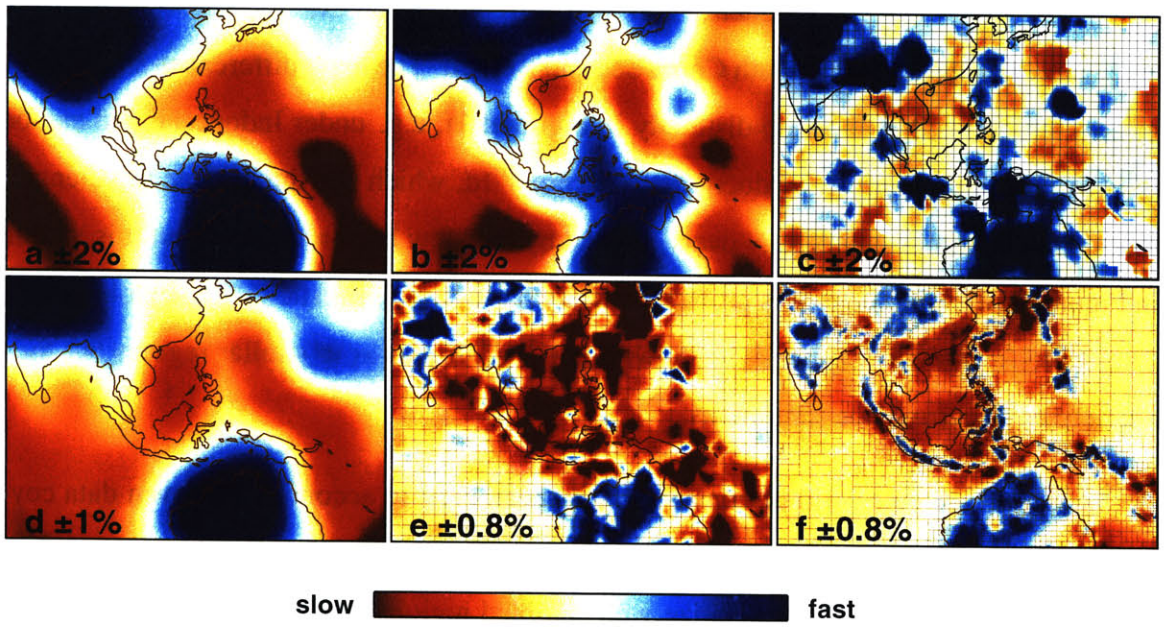


Figure 4-1: Horizontal cross section through six different models at 275 km depth. (a) S model *sks12wm13* by *Liu and Dziewonski* [1994]. (b) S model *s20rts* by *Ritsema and van Heijst* [2000]. (c) S model by *Grand et al.* [1997]. (d) P model *p16b30* by *Bolton* [1996]. (e) This study, P model, uniform parameterization. (f) This study, P model, adaptive parameterization.

[*Boschi and Dziewonski*, 1999] but they may also have sharper contrasts. Figure 4-1 depicts lateral cross sections at 275 km depth through Southeast Asia and Australia through six models. These models have been derived using different methods and they illustrate the differences that are inherent to the inversion technique used.

Provided that the number of coefficients is relatively small, the use of spherical harmonics (Figures 4-1a, b and d) allows for an elegant explicit solution of the inverse problem with formal resolution estimates. The local basis functions (Figures 4-1c, 4-1e and 4-1f) ensure a sparse but very large sensitivity matrix, which in turn means that iterative solvers, like LSQR [*Paige and Saunders*, 1982], must be used to find an approximate solution of the inversion problem, and a formal resolution estimate becomes impractical.

Although at first glance the models in Figure 4-1 may be quite different, a closer look reveals significant similarity. The Australian craton and the high wave speed beneath the

Himalayas are present in all models, albeit with different lateral extents, and so are the slow back arc region beneath the South China Sea and Borneo. The main differences are the very narrow structures that show up distinctly in Figure 4-1f, e.g. under Indonesia, and can be directly linked to ongoing subduction. In Figure 4-1d, which has the longest wavelength, this slab signature is completely overwhelmed by the slow back arc region but it is visible in Figure 4-1b even though it is probably smeared due to the limitations of the global basis. The S block model in Figure 4-1c shows some fast structures in similar regions but they are neither as narrow nor continuous.

We use an irregular grid of constant wave-speed cells to account for uneven data coverage, which allows the resolution of fine structures in regions of dense sampling. Effectively we choose to over-parameterize well sampled regions and use a much coarser grid elsewhere (see also *Bijwaard et al.* [1998]). An advantage of this approach, i.e. embedding the fine grid into a global model, over purely regional studies, is that large scale faraway structures are accounted for and the risk for such signal leaking into the local model is reduced. Figures 4-1e and 4-1f, which were made using exactly the same data, demonstrate the difference. The regular grid does not represent structures that are clearly contained in the data, but on the other hand, shallow oceanic regions are oversampled. Like any other modeling technique, the adaptive grid does not solve the problem of heterogeneous sampling, but it reduces its adverse effect on the images. Furthermore, it allows high resolution where one has data without a massive matrix inversion.

To better constrain the long wavelength component in the upper mantle we use low frequency PP-P differential times [*Kárason et al.*, 2002; *Woodward and Masters*, 1991, see also Chapter 3] and, to increase the resolution in the very deep mantle we use core phases from both high and low frequency records [*Kárason and van der Hilst*, 2001]. We use broad sensitivity kernels to adequately relate the low frequency data, i.e. PP-P and P_{diff}-PKP_{DF}, to mantle structure [*Vasco et al.*, 1995b; *Zhao and Jordan*, 1998; *Zhao et al.*,

2000; *Dahlen et al.*, 2000; *Hung et al.*, 2000; *Kárason and van der Hilst*, 2001]. When inverting all the data for mantle structure we adjust the regularization in such a way that we get the longer wavelength component in areas of less sampling along with the small scale structure where sampling (and thus grid size) allow.

Figure 4-1 demonstrates the motivation for our study. Although there are certain differences, the similarity between the models is encouraging and suggests that data measured at different frequencies can be used together to improve data coverage and, hence, the tomographic models.

In this paper we present this model. By zooming in on seismically active regions, details of small scale flow and complex slab morphology are revealed, in particular in the transition zone between the upper and lower mantle. We investigate to which depths subduction can be traced, how smaller and larger scale structures are connected, and how our model implies different flow regimes within the mantle.

4.2 Data

The data used in constructing this model can be divided into two basic categories, based on how the travel time residuals are measured. Here, we refer to them either as *routinely processed* or *waveform based* data. The reference earth model is *ak135* from *Kennett et al.* [1995].

The routinely processed data comes from the catalogue of *Engdahl, van der Hilst and Buland* [1998] (hereinafter referred to as the EHB data set). *Engdahl et al.* used arrival times reported to the International Seismological Centre and calculated travel-time residuals in a non-linear process, which included phase re-identification and earthquake re-location. The EHB data set is dominated by data of teleseismic first arrivals, i.e. P or PKP_{DF}. Along with the geometry of the associated rays, the heterogeneous geographic distribution of earthquakes and stations produces uneven data coverage. The ray coverage

can be improved by including later arriving phases from EHB, but since most of them are harder to read in the coda, they are typically much more noisy and the improved coverage would merely give a false impression of improved resolution. For this reason we do not include PP and PcP from the EHB data. Instead, to complement our EHB data set we use waveform based differential times that were carefully measured by cross correlation and other numerical methods from digital records. Here we describe the data sets briefly. For a detailed discussion of these data sets and how they are incorporated into the inversion we refer to *Kárason and van der Hilst* [2001] (see Chapter 2) and to *Kárason et al.* [2002] (see Chapter 3).

Phases	# records	# comp. rays	Cent. freq.	Forw. mod.	Var. red.	Source
P	7.2×10^6	2.2×10^6	1 Hz	Rays	~60%	EHB
pP and pwP	508×10^3	288×10^3	1 Hz	Rays	~30%	EHB
PKP_{AB}-$PKP_{DF/BC}$	29×10^3	15×10^3	1 Hz	Rays	~50%	EHB
PKP_{AB}-$PKP_{DF/BC}$	1383	N/A	1 Hz	Rays	~60%	McSweeney
PKP_{diff}-PKP_{DF}	543	N/A	50 mHz	Kernels	~30%	Wysession
$PP - P$	20.266	N/A	40 mHz	Kernels	~50%	Masters

Table 4.1: Summary of different data sets showing the phases they include and the number of records along with number of composite rays, central frequency of the measurements, type of forward modeling and the variance reduction upon inversion. Sources of data are *Engdahl et al.* [1998] (EHB), *McSweeney* [1995], *Wysession* [1996] and Guy Masters (pers. comm.). For details see sections 4.2 and 4.3

4.2.1 Waveform based differential times

For improved sampling of deep mantle structures we use differential times from the PKP-branch along with P_{diff}, and for the shallow mantle beneath oceanic regions we add measurements of PP-P differential times. Except for PKP, all of these travel times are measured at longer wavelengths, which in turn means that the mantle volume that influences the arrival times is much larger than in the case of the high frequency data (see below, section 4.3.2).

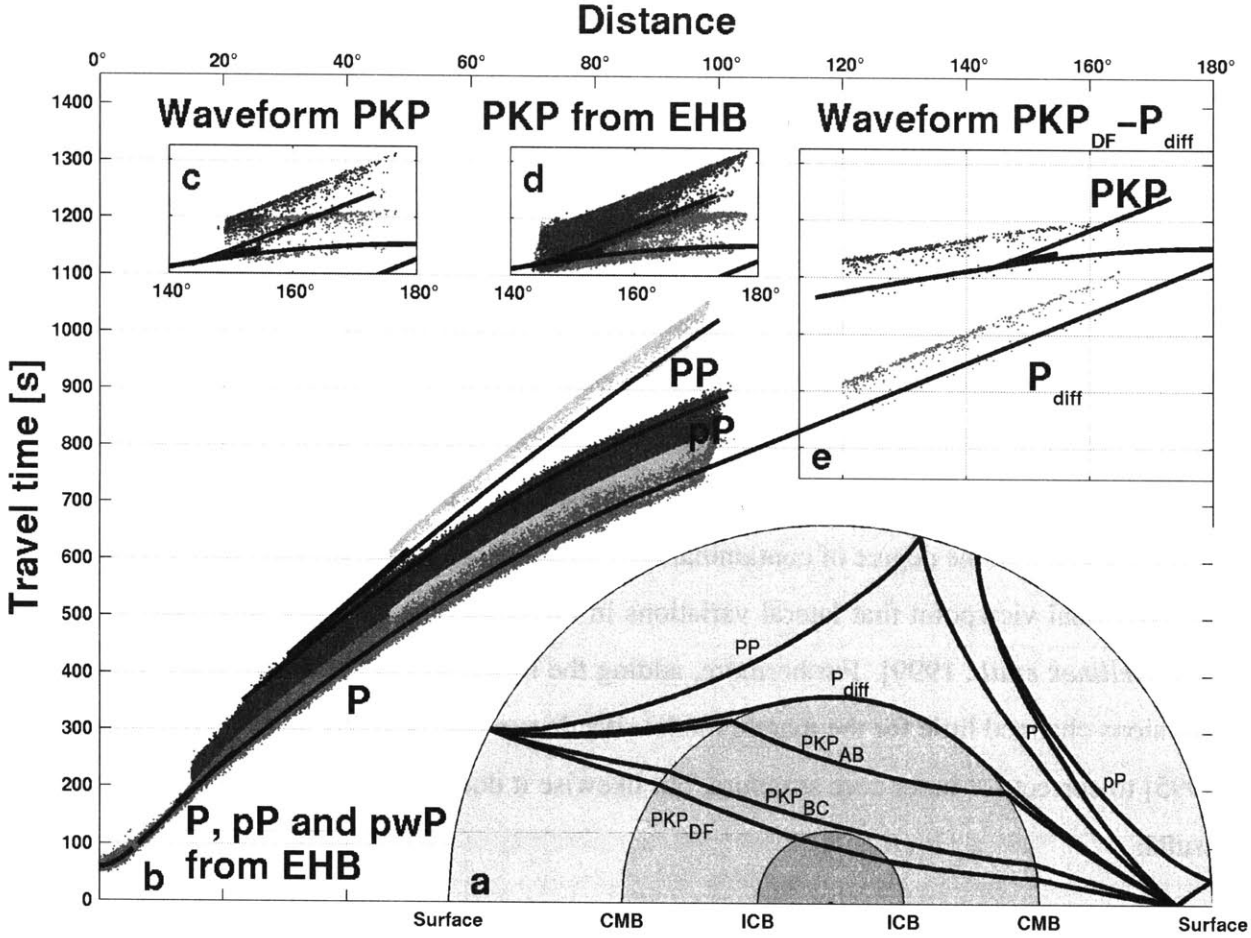


Figure 4-2: (a) The ray paths of the phases used in this study. P is a direct P wave that does not travel through the core. pP is up-going from the source while PP is down-going, both reflect once off Earth's surface. PKP_{AB} and PKP_{BC} travel through the outer core and PKP_{DF} travels both through the outer and the inner core, while P_{diff} grazes the core. For P_{diff} and PP we account for finite frequency effects by using 3-D kernels. (b) Theoretical travel time versus distance curves (solid lines) for an earthquake at 500 km depth calculated using the the *ak135* model. Also shown are scatter plots for ~7,000,000 P (light shading) and ~430.000 pP (dark shading) from EHB and ~20,000 PP (light shading). (c) Same as Figure 1b except 1383 waveform based PKP_{AB}-$P_{\mathrm{DF/BC}}$ (dark and light shading, respectively). (d) Same as Figure 1b except ~27.000 PKP_{AB}-$P_{\mathrm{DF/BC}}$ from EHB (dark and light shading, respectively). (e) Same as (b) except waveform based 543 PKP_{DF}-P_{diff} (dark and light gray, respectively). Using a deep earthquake moves the travel time curves down except for pP, which goes up, separating it from P. For detailed discussion of data sets, see text.

McSweeney [1995] measured PKP differential travel times by cross correlation of the observed PKP waveform with a synthetic signal calculated from theoretical predictions. We also use low frequency P_{diff}-PKP_{DF} differential times with central frequency around 50 mHz from *Wysession* [1996]. The P_{diff} phase is diffracted along the core mantle boundary and these differential times are very sensitive to structure near the base of the mantle (see Figure 4-2a and 4-2d and Table 4.1). To increase our ability to resolve structure in the upper mantle, in intraplate regions with few earthquakes and stations, we use long period PP-P data (Figure 4-2a and 4-2d, Table 4.1). Using waveforms with a central frequency around 40 mHz, the measurement is made by cross correlating the Hilbert transform of the P arrival with the PP arrival while accounting for attenuation (see *Woodward and Masters* [1991]; *Kárason et al.* [2002] and Chapter 3 for details).

The signal in the core phases might partly stem from structure in the outer and inner core. *Kárason and van der Hilst* [2001] studied the possible effects on the mantle model and concluded that the degree of contamination is minimal and found no reason to renounce the traditional viewpoint that lateral variations in the outer core are very small and short lived [*Jellinek et al.*, 1999]. Furthermore, adding the inner core to the model space in test inversions changed little for the mantle model. We do use the model by *Su and Dziewonski* [1995] to correct for inner core structure, but likewise it does not have a big impact on our results.

4.2.2 Routinely processed travel times (EHB data)

From the EHB high frequency data set we use P, pP and pwP as well as PKP. The depth phase, pP bounces off Earth's surface but pwP propagates also through water and bounces off the surface of the ocean. See Table 4.1 and Figure 4-2a for ray paths and Figure 4-2b for distance ranges.

The EHB data set is regularly updated by Dr. E. R. Engdahl and smaller versions of

the data file have been used in previous regional [e.g., *van der Hilst and Engdahl*, 1991] and global studies, [e.g., *van der Hilst et al.*, 1997; *Widiyantoro*, 1997; *Bijwaard et al.*, 1998; *Kárason and van der Hilst*, 2000, 2001]. This time the data have been processed in a slightly different way, relaxing some of the conditions regarding teleseismic recordings, which retains data from smaller earthquakes, that are well recorded locally. This version also includes one additional year, and now contains data from January 1964 through August 1999. The combined result is an increase in the number of earthquakes from ~100,000 in the old version to ~316,000 in the new version. The increase in number of records with higher quality P arrivals is not as dramatic, 6.2×10^6 to 7.2×10^6, since these smaller earthquakes generally have fewer recordings.

We also use PKP data extracted from the large EHB data file (see Figure 4-2a) but only records that have two or more PKP arrivals are used, so that the construction of differential times (PKP_{DF}-PKP_{AB} or PKP_{BC}- PKP_{AB}) is possible. Consequently, only a fraction of the available EHB PKP data is used, but much of the data left out are PKP_{DF}, which do not add much to the sampling of the deepest mantle and, moreover, can be sensitive to inner core structure and anisotropy [*Kárason and van der Hilst*, 2001]. Records with similar earthquake and station locations are then grouped together to form summary rays, and the median of the associated data is used in the inversions. Using the median residual increases the consistency with the waveform based PKP travel times [*Kárason and van der Hilst*, 2001].

4.3 Methodology

4.3.1 Ray tracing, composite rays and weighting

The center frequency of the short period data is ~1 Hz, and for the calculation of a sensitivity matrix, we trace the optical ray, which is a high frequency approximation. For the

EHB data we use weighted composite rays, to reduce the size of the sensitivity matrix, which relates travel times and mantle structure. These methods are now fairly standard, see for instance *Nolet* [1987] and *Spakman and Nolet* [1988]. To balance the small, but high quality waveform data sets against the much larger and noisier EHB data we give them extra weight. The level of weight is guided by the variance reduction of the different data sets (see Table 4.1) and by visual inspection of the model (see *Kárason and van der Hilst* [2001] for details).

4.3.2 Sensitivity kernels

The core-diffracted P_{diff} is an evanescent wave and the magnitude of the associated particle motion (and, as a corollary, the sensitivity to structure) decreases with increasing distance from the interface (i.e. CMB). This decay is frequency dependent. To account for this in the inversion we distribute the sensitivity over a finite mantle volume. The approximate sensitivity kernels in Figure 4-3a show that the sensitivity is distributed over a large irregularly shaped volume, which suggests that high frequency approximations are not suitable. Toward the CMB the kernel broadens even more, in particular, for large epicentral distances. Li Zhao (personal communication, 1999) calculated sensitivity of P_{diff} to mantle structure using normal mode theory [*Zhao and Jordan*, 1998; *Zhao et al.*, 2000]. Using cross sections, showing sensitivity along and normal to the great circle path between the source and the receiver, the shape of the kernel was approximated. For each record the appropriate kernel, adjusted for distance between source receiver is projected onto the basis function used for model parameterization. For a detailed discussion of the construction and implementation of these kernels into the inversion we refer to *Kárason and van der Hilst* [2001].

For the low frequency PP-P differential times, high frequency approximations (e.g. rays) are not appropriate either. Instead we used kernels like the ones shown in Figure 4-3b

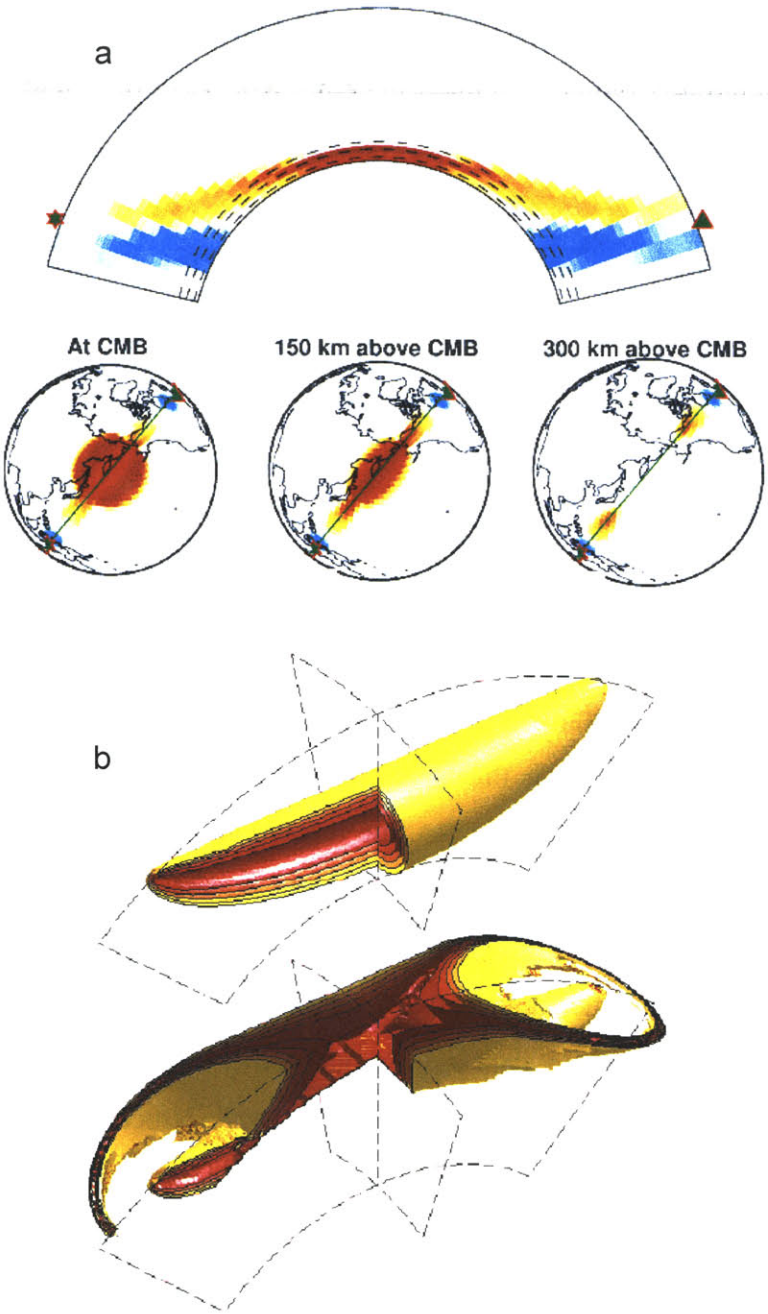


Figure 4-3: (a) Sensitivity to slowness variations within the Earth for low frequency PKP_{DF}-P_{diff} differential times. Red and blue indicate negative and positive sensitivity, as one kernel (PKP) is subtracted from the other (P_{diff}) to yield the sensitivity to differential times. Top: Vertical section through station and earthquake. Bottom: Map views of the sensitivity at different distances above the CMB. (b) Surfaces, representing constant travel time deviation using single scattering for P and PP. One quadrant is cut out to allow insight into the 3-D kernel. Darker colors correspond to smaller travel time deviations, see e.g. *Marquering et al.* [1999]. Assuming zero sensitivity

These kernels are calculated using single scattering theory, and only include the 1st Fresnel zone. From a series of inversions we concluded that, with the data coverage available to us, the actual distribution of sensitivity over the Fresnel zone is of little importance. The details of the implementation of the PP-P data and the associated changes to the model are discussed in *Kárason et al.* [2002] (see Chapter 3).

The use of 3-D kernels effectively smoothes the part of the model space constrained by the low frequency data while allowing the resolution of fine structures in regions of dense sampling by high frequency waves.

4.3.3 Adaptive parameterization

The uneven sampling of mantle structures by seismic waves renders significant lateral variations in resolution. The use of a regular grid would either over-parameterize poorly sampled regions (and be computationally inefficient) or average out small scale structure. A benefit of local basis functions is that their size can be adapted to reflect the resolution. Ideally, such an adaptive grid reflects the true resolution, but for large systems it is hard to quantify the true resolution, and we use sampling density (hit count) instead. Following *Abers and Roecker* [1991], *Gudmundsson and Sambridge* [1998] and *Spakman and Bijwaard* [2001], we constructed an adaptive parameterization scheme on the basis of the sampling by the high frequency data used in our study. The sampling by low frequency data is not considered since they do not resolve small scale structure anyway.

The first step in making the variable resolution grid is to make a base grid. The cells of this grid are used as building blocks when making the variable resolution grid and therefore they represent the maximum possible resolution within the variable grid. The base grid is approximately 0.6° in latitude and longitude and 45 km in depth throughout the mantle. Rays are traced through this grid to produce a sensitivity matrix from which irregular grid matrices are derived.

Each cell within the adaptive grid is assembled of one or more cells from the base grid. The total ray length of such a combined cell is simply the sum of the individual entries from the base grid cells. Even though nothing prohibits combining base cells in a very complicated way, we only allowed the variable grid to contain cells that are cubes in latitude, longitude and radius space. Also, in order to avoid very thin and long cells, the aspect ratio of the cells is kept low, i.e. each assembled cell can only have twice the number of base cells along one side vs. the other sides. Irrespective of sampling density, the minimum width of the cells in the irregular grid is set by the approximate size of the Fresnel zone for 1 Hz P-waves; this minimum size increases from 45 km at the surface to 100 km near the CMB [*Kárason and van der Hilst*, 2001].

The variable grid is constructed by first creating large cells, each containing 16x16x16 base cells. These large cells are then divided in a recursive algorithm, until the desired distribution is obtained (e.g. Figure 4-4a). The largest cells are mostly found in oceanic regions in the upper mantle, whereas the smallest cells are near regions of dense seismicity or receiver distributions. In the middle to deep mantle, the grid is fairly regular due to more homogeneous sampling. To reduce imaging artifacts owing to very large gradients in sampling (especially in the upper mantle) - and thus cells size - we smooth the grid, i.e we subdivide large cells that are adjacent to small cells, see Figure 4-4b.

4.3.4 Regularization and inversion

As in our previous studies, we apply a combination of *norm* damping, which tends to minimize the amplitude of the model, and *gradient* damping, which produces smooth variations, both laterally and radially. We also account for effects of earthquake mislocation, i.e. we allow earthquakes to move coherently within a block of a certain size, and that way absorb some of the signal. For the numerical inversion we use LSQR [*Paige and Saunders*, 1982], which is a computationally efficient version of the conjugate gradient method.

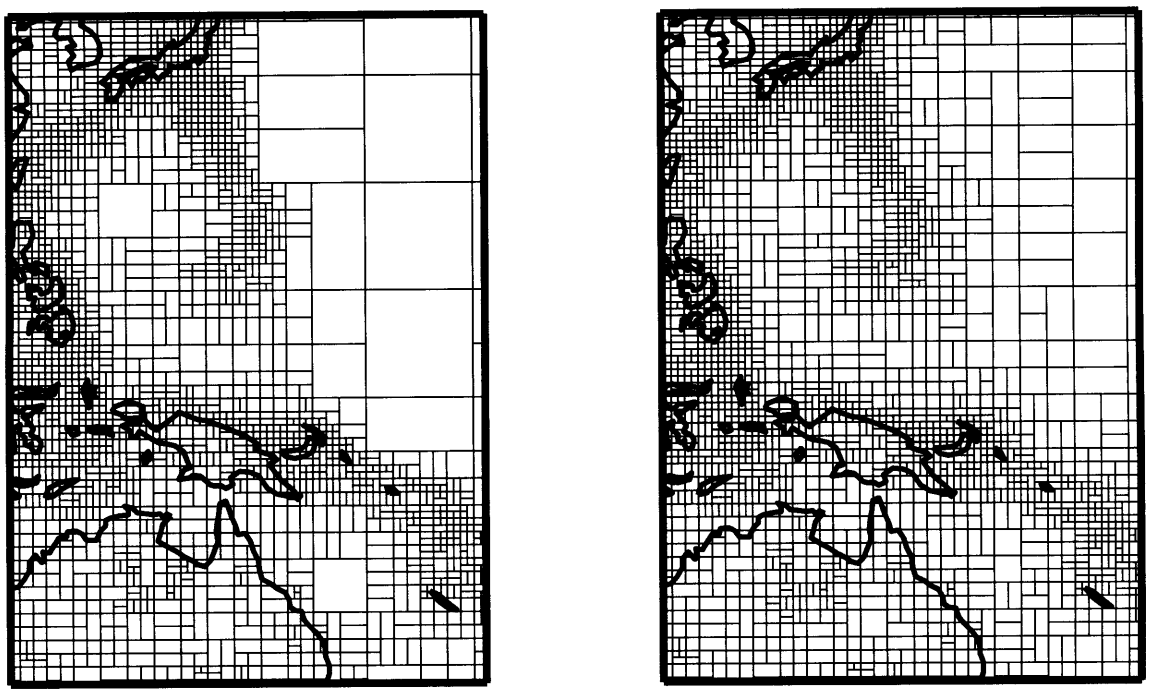
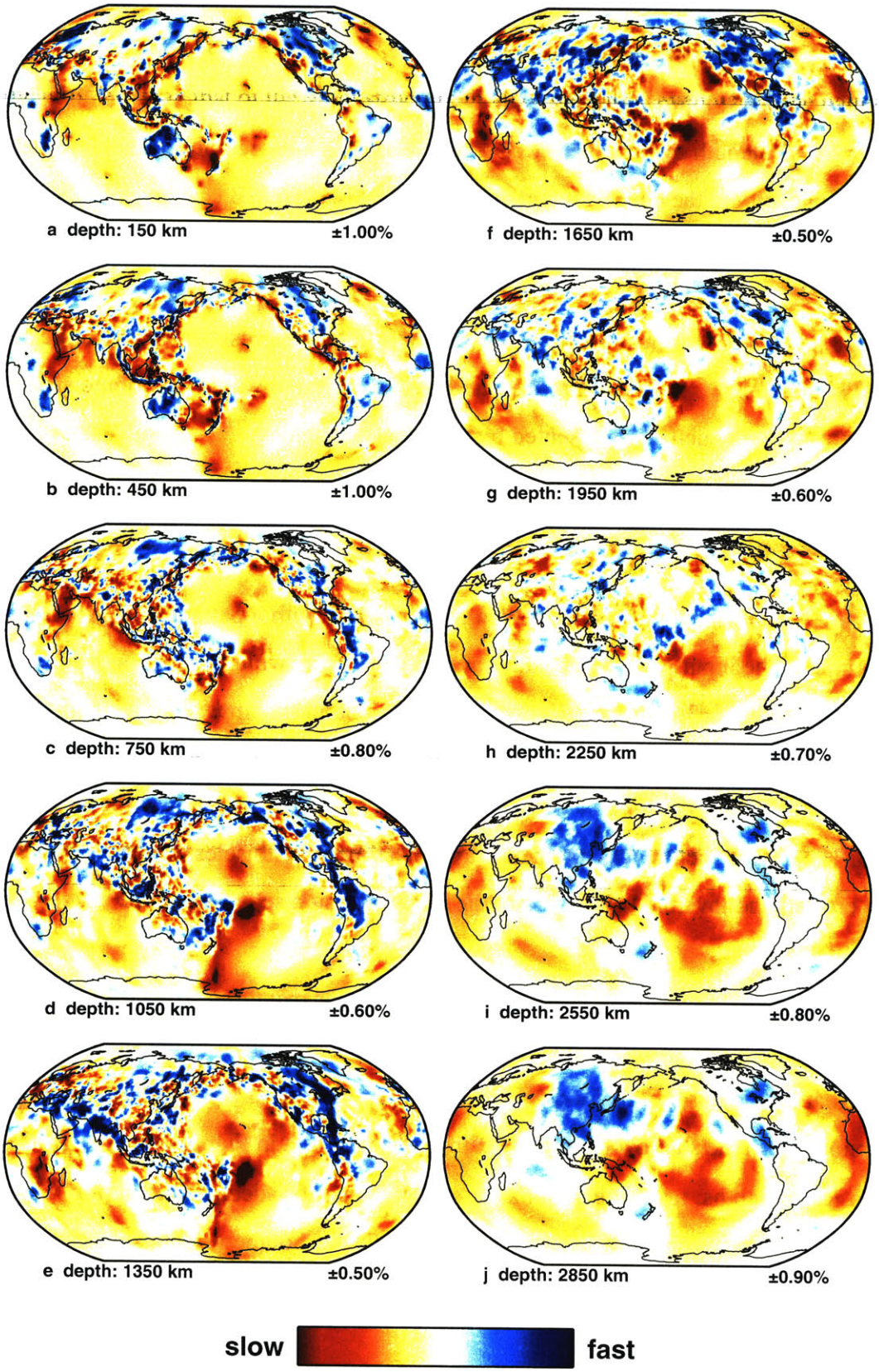


Figure 4-4: (a) Adaptive grid before spatial smoothing in is applied and (b) after.

4.4 Results

4.4.1 Global structure

Figure 4-5 depicts wave speed variations according to the model at selected depths in the mantle. From a global perspective, the results are consistent with our former studies using regular grids, i.e. *Kárason and van der Hilst* [2001] and *van der Hilst et al.* [1997], and our analysis and interpretation of this part of the model will therefore be brief. Slow back arc regions and fast subduction zones along with craton signatures are prominent at shallow depths. Deeper, the model is marked by long and narrow traces of fast material from the upper mantle transition zone to mid-mantle depths beneath North and South America and beneath southern Asia. These structures have been associated with plate motion history and are thought to be the remnants of old subducted slabs [*van der Hilst et al.*, 1997; *Grand*


a depth: 150 km ±1.00%
f depth: 1650 km ±0.50%
b depth: 450 km ±1.00%
g depth: 1950 km ±0.60%
c depth: 750 km ±0.80%
h depth: 2250 km ±0.70%
d depth: 1050 km ±0.60%
i depth: 2550 km ±0.80%
e depth: 1350 km ±0.50%
j depth: 2850 km ±0.90%
slow
fast

et al., 1997]. At larger depth these structures begin to disintegrate and around 2000 km depth no large distinctive structures are noticeable.

A few hundred kilometers above the CMB, a new pattern emerges, which is generally of longer wavelength than at shallower depths. There is a degree of resemblance between the heterogeneity in the deepest mantle with the large-scale CMB structure of other models [*Liu et al.*, 1998; *Castle et al.*, 2000], i.e. the regions beneath the Pacific, and the Atlantic are generally slow. Differences still remain, as here the Pacific is not surrounded by fast material, but a discontinuous band of fast material stretches across the north Pacific and the west Pacific is slow. Nonetheless, the pattern observed here would contribute strongly to the degree 2 component if represented by spherical harmonics.

4.4.2 Resolution of large scale structure

In Figure 4-6 we show the results from a test inversion with a pattern of alternating signs as an input. The input pattern has half wavelength of $\sim 5^{\circ}$, which corresponds to spherical harmonic degree, l=36, and a peak amplitude of 1.5% in the upper half of the mantle. In the lower half of the mantle, the half wavelength is $\sim 10^{\circ}$, or l=18, which gives similar real length (500 km-600-km for half wavelength) in the test at the surface and for the test at the CMB. The input structure was put in one layer at a time, and noiseless synthetic travel times were created and inverted using exactly the same inversion scheme as used in constructing the model.

The resolution is spatially most limited in the upper mantle, restricted by the location of stations and zones of active seismicity. Locally, the resolution in the shallow mantle is better than borne out by this long wavelength input pattern, as shown in the smaller scale inversion test in the shallow mantle beneath Southeast Asia in Figure 4-7.

At larger depths, rays spread out and so do the areas of adequate resolution. However, due to the scarcity of stations, large areas beneath the Southern Hemisphere are poorly

resolved at all depths. Throughout most of the lower mantle the quality of the recoveries and their radial smearing stays approximately constant. The amplitude of the recovery is significantly less than that of the input structure in all of the cases, even where the pattern is retrieved well. This is typical for the type of inversion we perform. The data sets are large and noisy and the level of damping therefore needs to be high (for more detailed analysis of the resolution on a global scale and further tests see *Kárason and van der Hilst* [2001]).

4.4.3 Regional structures: Subduction zones

In many regions the data resolve structure on much smaller scales than can be appreciated from the global view in Figure 4-5, especially in the shallow mantle. Ray coverage is dense where earthquakes are common, and as a demonstration of the detailed structure within our model we focus on slabs of subducted lithosphere that mark tangible trajectories of mantle flow. In Figures 4-8 through 4-10 we show examples from several currently active subduction systems and Figure 4-11 contains resolution tests for the cross sections displayed. Our objective here is to give an overview of different subduction styles and an indication of the resolution. Detailed discussion of particular subduction systems and the implications of our imaging results of their tectonic evolution over Earth's history is beyond the scope of this paper.

The choice of cross sections is arbitrary, but we constructed most of them in the dip direction of the seismic zones as constrained by the EHB earthquake locations. For this purpose we used the Wadati-Benioff zone rendering of *Gudmundsson and Sambridge* [1998]. We stress, however, that two dimensional cross sections can be misleading and that three dimensionality must be accounted for.

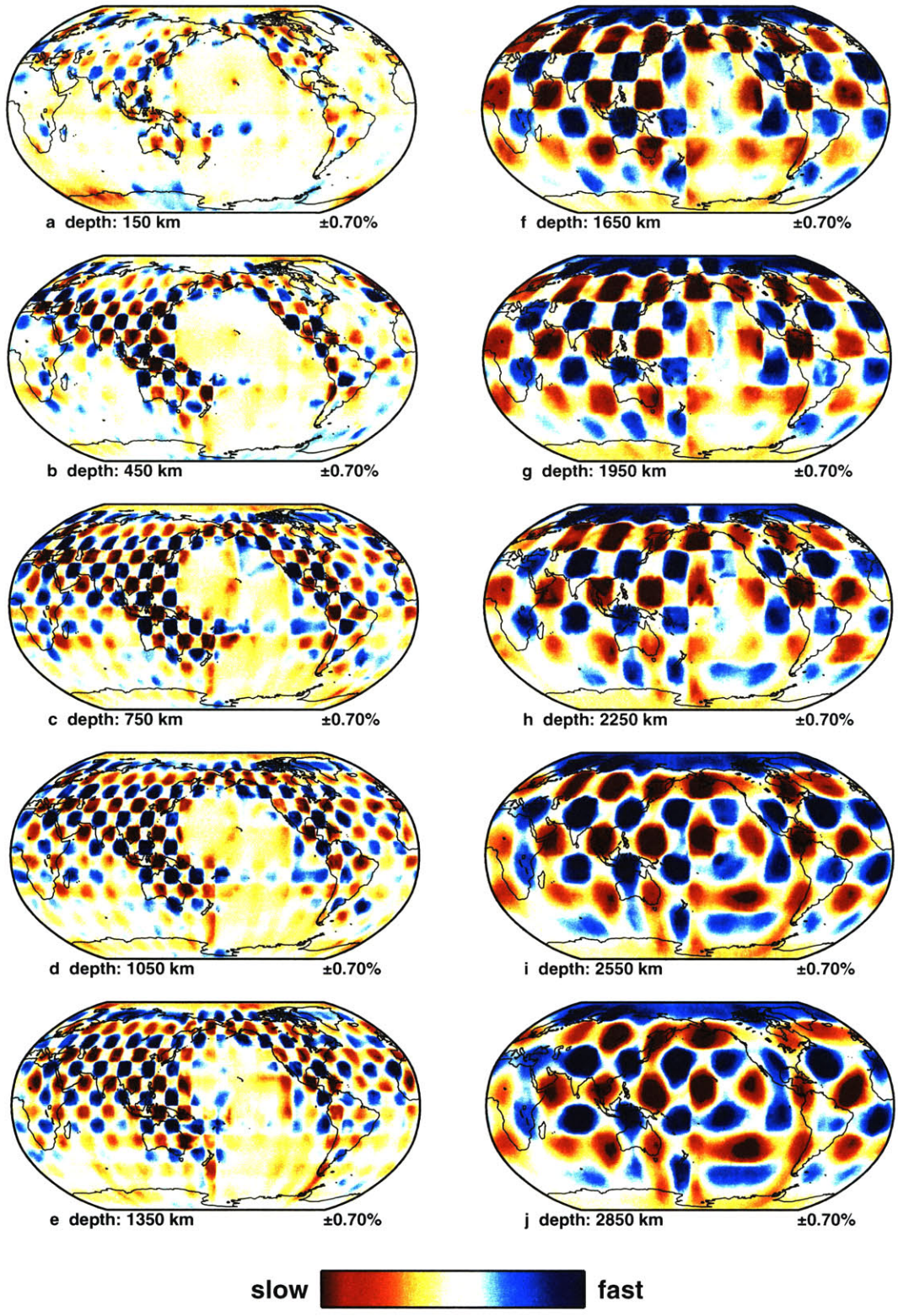


Figure 4-6: Recovery of global inversion tests using harmonic input patterns with peek values $\pm 1.5\%$ throughout the mantle. (a-e) Half-wavelength of $\sim 5^{\circ}$ (equivalent to spherical harmonic

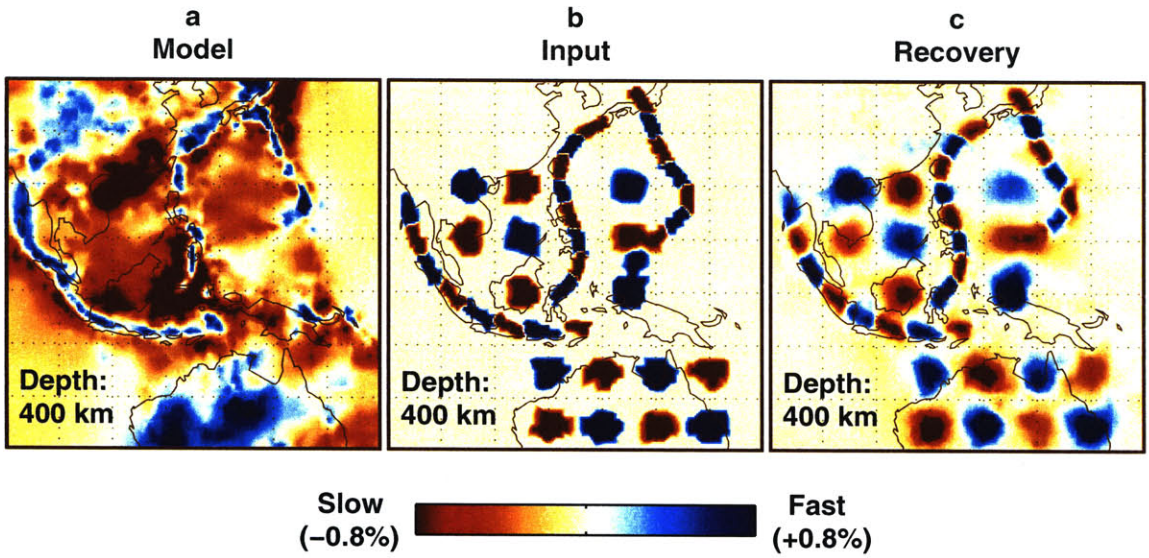


Figure 4-7: (a) Regional model at 400 km depth in Southeast Asia. (b) Small scale artificial input structure for inversion test. (c) Recovery of input structure.

Mediterranean convergent plate boundary:

Sections 1 to 5 through the Hellenic subduction zone (Figure 4-8) underscore how deluding 2-D cross sections can be. Here, perhaps no single cross section can capture the essence of the complex three dimensional morphology. Cross section 1 suggests that the subducting slab continuous directly into the lower mantle with little deformation, while section 3 suggests stagnation in the transition zone. These seemingly conflicting pictures stem from cutting through the structure at varying angles and from choosing cross sections that are much larger than the radius of curvature of the subduction system being considered. Cross section 4, which is drawn perpendicular to the strike of the Mediterranean convergence zone at large, instead of the local dip direction of the seismic zone, still shows a stagnant slab in the transition zone, but the angle of subduction is shallower. Fast material is present in the lower mantle but, in the west, there is no direct connection to upper mantle structures and, in the east, the connection is vague, in contrast to cross section 1. The deep anomaly

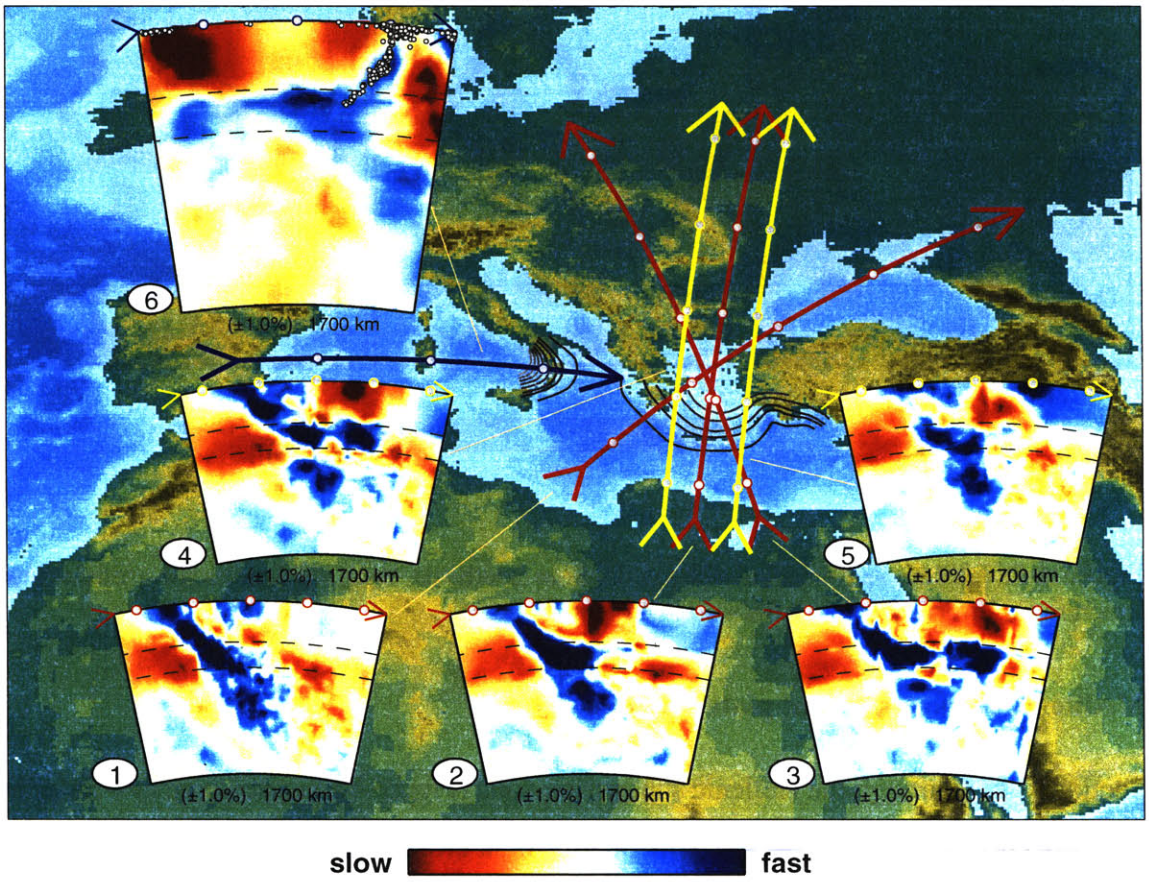


Figure 4-8: Cross sections through active subduction zones around the Mediterranean. Gray scale contours show outlines of subduction zones from *Gudmundsson and Sambridge* [1998]. Location of cross sections is indicated and circular markers are placed 5° apart. Limits of the color-scale and the depth extent are indicated at the bottom of each section.

could be related to closing/reopening events of the Tethys sea, as it seems connected to the much larger anomaly around mid-mantle depths along the southern margin of Asia as seen in Figure 4-5d-e. For cross sections 1 through 5, the inversion tests (Figure 4-11) suggest excellent resolution and indicate that, for instance, the apparent disconnection between the lower and upper mantle structures is real. Cross section 6 shows subduction in southern Italy and across the Tyrrhenian sea. The resolution is worse, but horizontal smearing in the

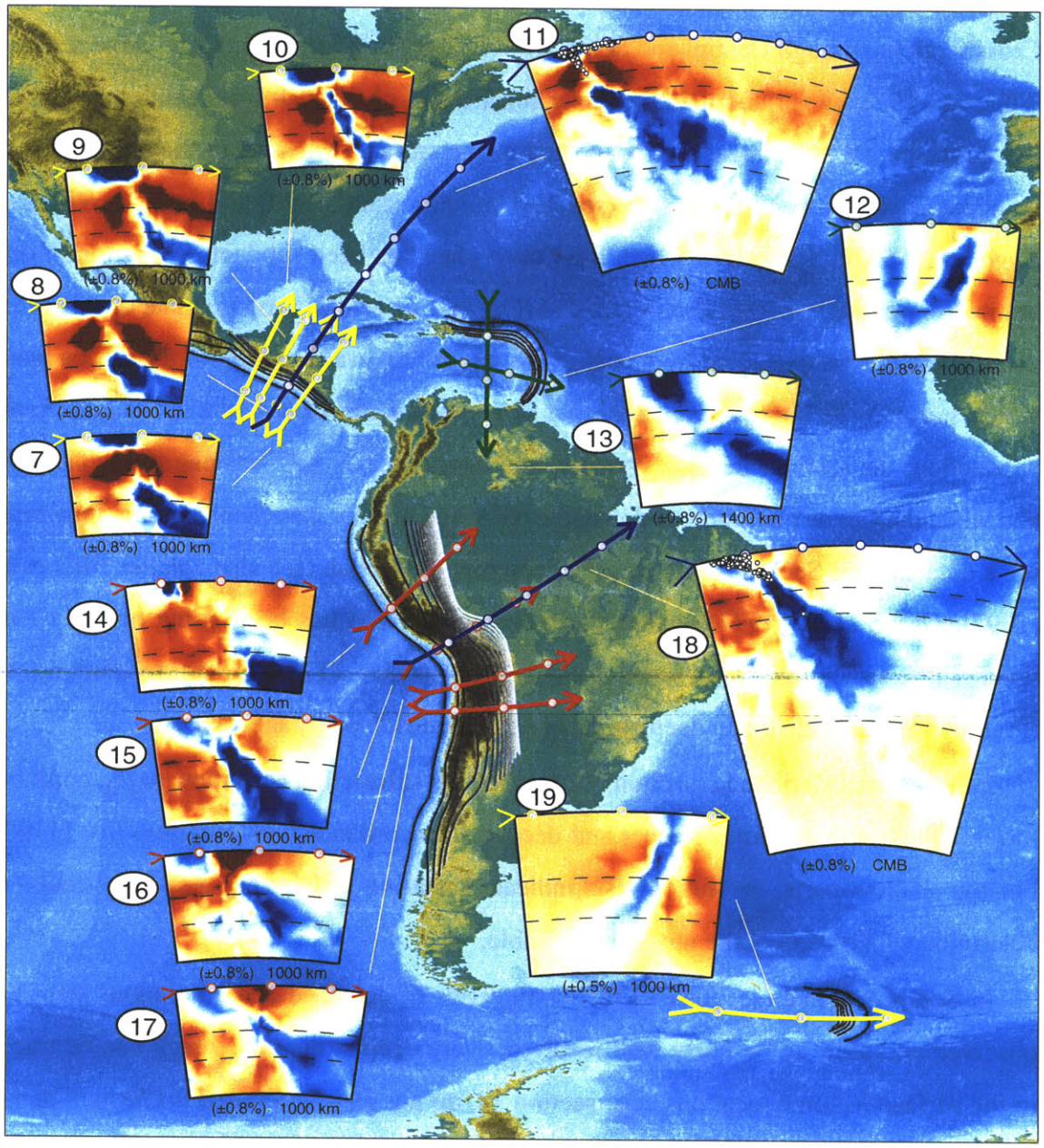


Figure 4-9: Same as figure 4-8 for Central and South America

transition zone seems small, suggesting that the accumulation above the 660 km depth may be real [e.g. *Spakman et al.*, 1993; *Faccenna et al.*, 2001]. Judging from the resolution test, the connection to the lower mantle may well result from smearing.

Central America and the Andean subduction system:

Cross sections through Central and South America are shown in Figure 4-9. Cross sections 7 through 10 show variations in the upper mantle part of the slab from south to north. In the north the deep subduction is connected to the surface, whereas in the south there is an apparent detachment of the slab (see also *van der Hilst* [1990] and *van der Hilst and Engdahl* [1991]). The rate at which the Cocos plate has been subducting has varied dramatically over the last 30 Million years and *Rogers et al.* [2001] proposed that the observed detachment can be explained by a change from rapid eastward motion of the Cocos plate, which started $\sim$28 million years ago, to a slow down some 10 Ma ago. During the rapid subduction, the seafloor became younger, and hence weaker and less negatively buoyant, and when the Cocos plate slowed down, the older and heavier parts kept sinking, thus causing a detachment. Section 11 reveals the deep part of the slab (see also *van der Hilst et al.* [1997]; *Grand et al.* [1997]) and suggests that subduction has been going on for a long time. These structures do not seem to extend deeper than $\sim$1700 km but, resolution tests (Figure 4-11) indicate that a deep anomaly could be resolved by the data used.

In the eastern Caribbean (sections 12 and 13) the subduction is shallow to the north but towards the east it becomes steeper and deeper, in accordance with *van der Hilst* [1990]. For the shallow parts, resolution tests indicate smearing and further tests are needed to determine which parts of the model are reliable.

In South America (sections 14 to 17), intervals of detached and continuous slabs are observed and seem to be fairly well resolved. Very deep structures, which can be traced to the surface in some locations, e.g. section 18, do not seem to exceed $\sim$1700 km depth.

The resolution test indicates that a connection to the lower-most mantle should probably be resolved with the data used. As is the case for the Southern Hemisphere in general, the resolution deteriorates when going south and the depth extents of the subducting slab beneath South Sandwich Islands (section 19) is probably not constrained.

North-West Pacific:

To the west of the Pacific, continuous subduction stretches from the Kamchatka Peninsula in the north, through the Kuril Islands and Japan, ending south in the Marianas Trench (sections 20 to 25). The style of subduction varies dramatically along this trench. In the north, the slab is rather steep and seems directly connected to lower mantle structures. Farther south, the dip angle decreases to about 30° for Central Japan and the slab appears largely stagnant in the transition zone. The seismicity displayed in section 22 outlines the slab down to about 500 km depth but the stagnant part is mostly aseismic. The contrast in the style of subduction between sections 22 and 23 is stark, very shallow subduction stagnant in the upper mantle to vertical subduction that seems to be connected almost to 1700 km depth [e.g. *van der Hilst and Engdahl*, 1991; *Fukao et al.*, 1992]. Farther towards the Marianas trench (sections 24 and 25) the slab remains steep but becomes gradually shallower and seems to be limited to the upper mantle in section 25. For this seismically active region the resolution is, in general, acceptable, but some smearing is observed, for example in section 21 and the amplitude is reduced in section 23.

In the New Britain subduction zone (sections 26 and 27) the sinking slab seems to have accrued above the 660 km discontinuity with little, if any, connection to the lower mantle. Some smearing is observed, but larger trends are resolved.

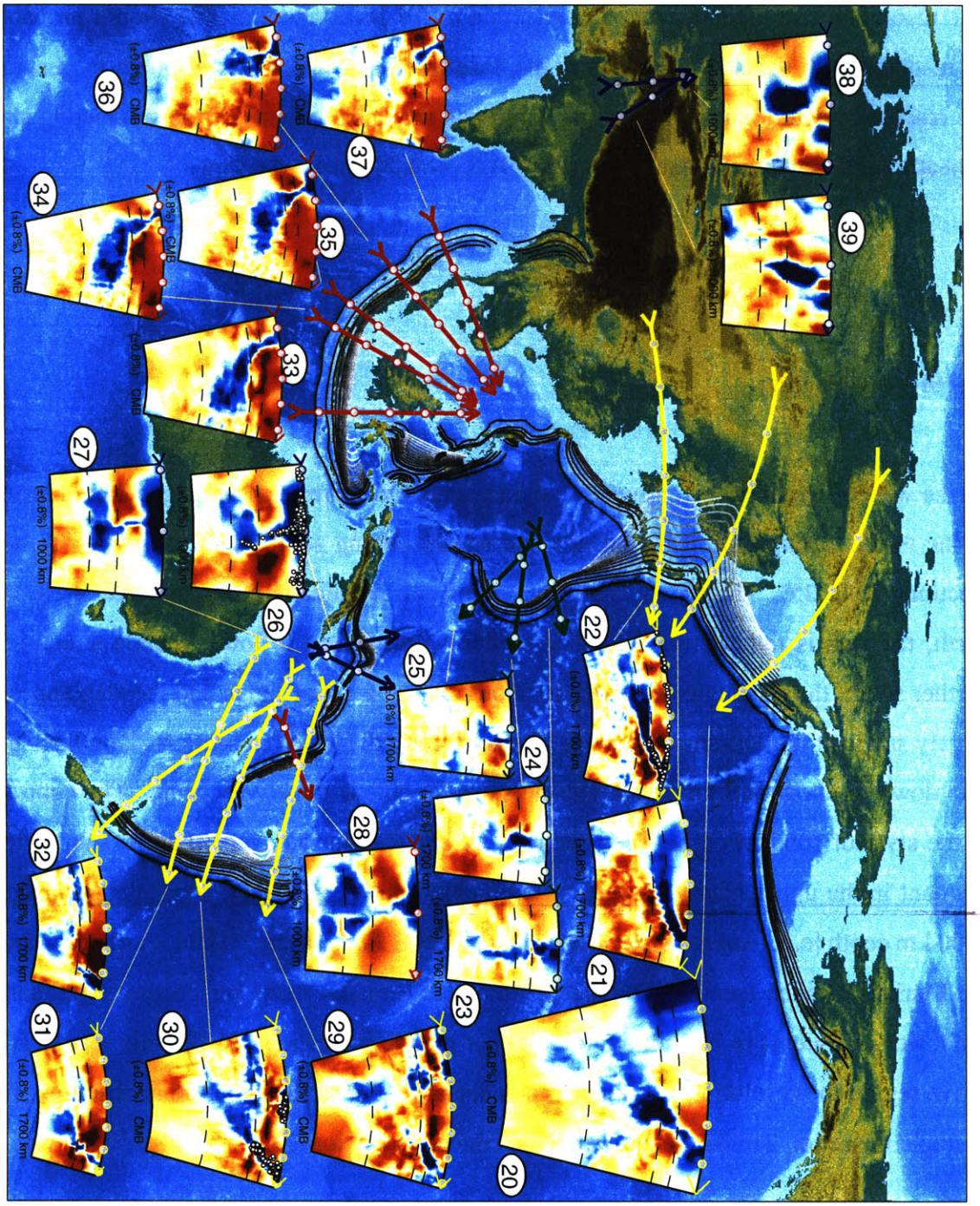


Figure 4-10: Same as figure 4-8 for the northern, western and southern Pacific

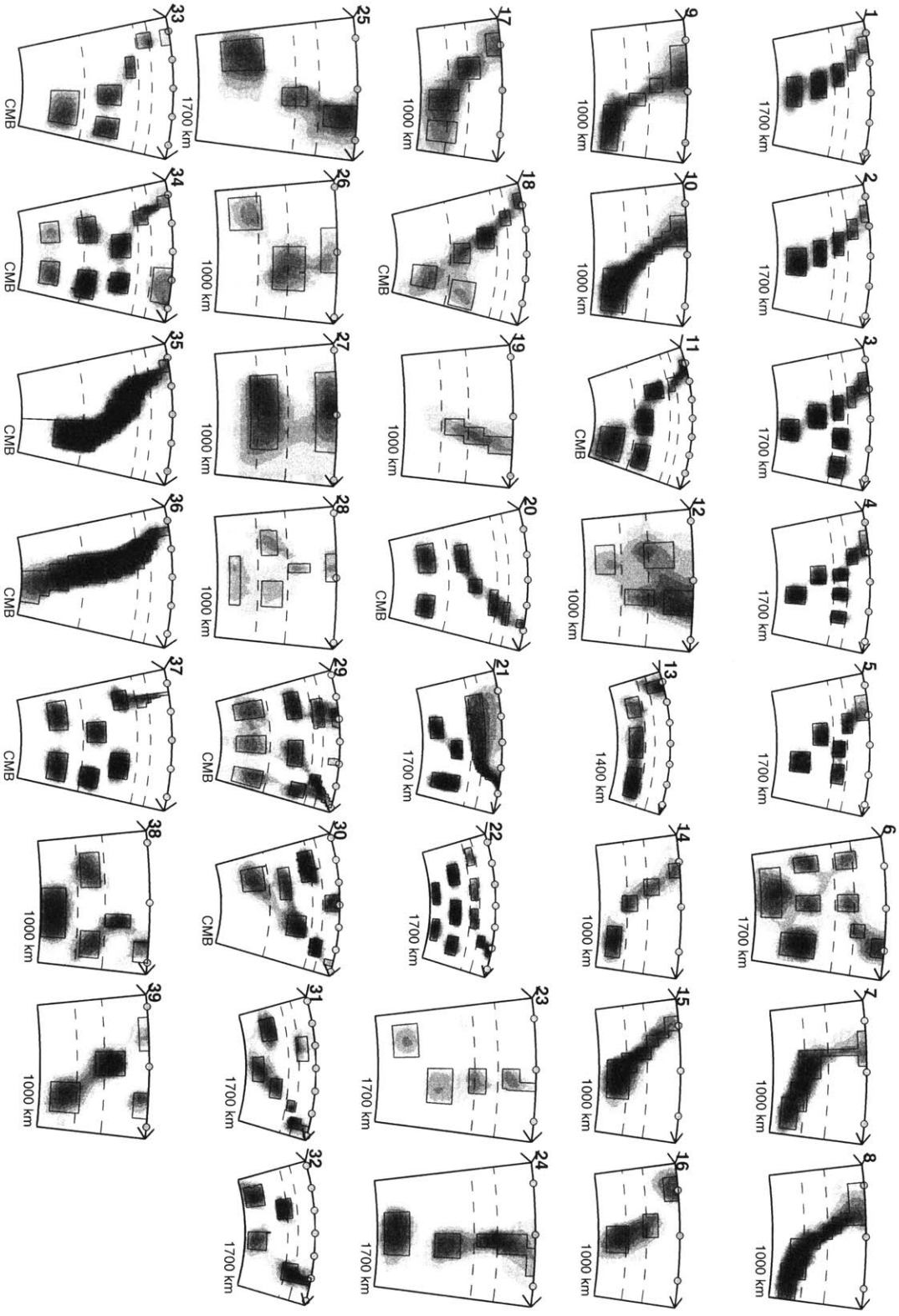


Figure 4-11: Inversion tests for cross sections. Outlines of fast *input* structure (+1%) of fast material is indicated with black lines. The *recovery* is shown in gray scale. Numbers refer to the cross sections shown in Figures 4-8, 4-9 and 4-10

Tonga-Kermadec:

Of all subduction zones, the Tonga-Kermadec region (section 29) has the most deep seismicity [*Isacks and Molnar*, 1971; *Vassiliou et al.*, 1984]. *Gurnis et al.* [2000] proposed that the deep seismicity was due to resistance against the sinking of the cold heavy downwelling from a deeper hot upwelling. Our model is in agreement with their findings, showing slow material from ~1700 km up to ~660 km depth. Farther to the south (section 37), the subduction seems to be much deeper, reaching towards ~1700 km depth, but it gradually becomes shallow again towards New Zealand in sections 31 and 32. Resolution in this area is generally good and major trends are resolved. For most parts, these results are consistent with *van der Hilst* [1995] who explained the changes in morphology of the subducting slab along the Tonga and Kermadec trenches in terms of recent tectonic evolution of the subduction system.

Indonesia and South Asia:

In the south eastern part of Indonesia the subduction continuous to large depth (section 33), with an up to 1000 km thick mass of fast material in the lower mantle and a rather sharp cut-off to the north. In this region the major slab structure is related to the subduction of the Indian Ocean beneath Indochina and the collision of India with mainland Asia, which caused the south and south-west rotation of the Sunda block [*Widiyantoro and van der Hilst*, 1996; *Replumaz and Tapponnier*, 2001]. The thickening of the slab in the lower mantle is consistent with lower sinking rates due to increased viscosity (see Chapter 7). To the west and north the accumulation in the lower mantle becomes gradually less and the slab more steep and eventually limited to the upper mantle (sections 34 to 37). Resolution is in general good and, in particular, shows that with our data coverage we should be able to discern between a slab reaching the CMB or ending at shallower depths. The two Hindu Kush subduction zones (sections 38 and 39) are more clearly defined in the model and

reach down to the 660 km discontinuity. Resolution is good and the main features of the subducting slab are well resolved.

4.5 Discussion

4.5.1 Large scale structure and layering within the mantle

As mentioned earlier, the global structure in Figure 4-5 is consistent with our earlier models [*van der Hilst et al.*, 1997; *Kárason and van der Hilst*, 2001; *Kárason et al.*, 2002, see Chapters 2 and 3]. Some differences do exist, but they are clearly mostly due to different combinations of data sets used in each case. All these earlier studies used regular parameterization and, while the fine details of this model are not borne out at this scale, no detoriation of the large scale structure is observed. For example, the observed change in flow patterns around mid-mantle depths is as pronounced as before (See also *Su et al.* [1994]; *van der Hilst et al.* [1997]; *Liu et al.* [1998]; *van der Hilst and Kárason* [1999]) As discussed in detail by *Kárason and van der Hilst* [2001], many changes seem to occur around this depth interval, e.g. between 1000 and 2000 km depth the amplitude of the wave speed variations is small, the scatter in the pertinent travel time data is high [*Gudmundsson et al.*, 1990], and the spectrum of heterogeneity is white, at least up to degree 12 [e.g., *Su and Dziewonski*, 1992]. Resolution over this depth range seems fairly constant according to Figure 4-6 and *Kárason and van der Hilst* [2001]. Changes in resolution can therefore not readily explain the change in patterns at mid-mantle depths.

In addition to inferences from specific examples of flow patterns, as shown earlier and discussed later, some global diagnostics are also indicative of changes in flow with depth. For example, the root-mean-square (*rms*) of the amplitudes within the model is probably related to the intensity of the thermal or chemical anomalies that cause them, and the radial correlation is sensitive to the radial continuity of structures within the model [*Puster and*

Jordan, 1997; *van der Hilst and Kárason*, 1999]. The latter is therefore a good indicator of layering within the mantle and, along with the *rms*, provides a straightforward means for comparison between tomographic models and convection simulations [*Puster and Jordan*, 1997]. The downside of the adaptive grid is that it is hard both to define and interpret model averages because the wavelength is spatially variable, both laterally and radially and, for example, the definition of a 'layer' becomes ambiguous. Therefore, the radial correlation and *rms* of amplitudes, are shown in Figure 4-12 as function of depth for a model based on a uniform grid. This model is created using exactly the same data but the mantle is divided into $\sim 2.8^\circ$ by $\sim 2.8^\circ$ cells and 32 layers of constant thickness. From a global perspective the two models are very similar except for the very small scale structure that the adaptive grid is able to represent (see Figures 4-1e and 4-1f).

In the upper mantle the *rms* of the amplitudes is generally high but changes rapidly in the transition zone. We are hesitant to put too much weight onto the detail of these curves and we do not try to explain the increase in amplitude that coincides exactly with the two upper mantle discontinuities. It is also unclear if it is significant that the radial correlation has troughs where the *rms* has peaks and vice versa. However, the dip in the correlation around 410 km and 660 km is consistent with the cross sections in Figures 4-8-4-10, which show, in many instances, accumulation of material in the transition zone.

In the shallow part of the lower mantle the opposite trends continue; the *rms* decreases when the radial correlation increases. In contrast, below 1400 km depth the trends are the same; the *rms* continues to decrease but so does the radial correlation down to about 2000 km depth, where they both start to increase, reaching a maximum at the CMB. Again, these variations are not easy to interpret, but the fact that radial correlation and *rms* are either anti-correlated or correlated might have bearing on different convection regimes in the upper and lower half of the mantle.

In summary, the large-scale structure observed around 1300 km depth cannot in general

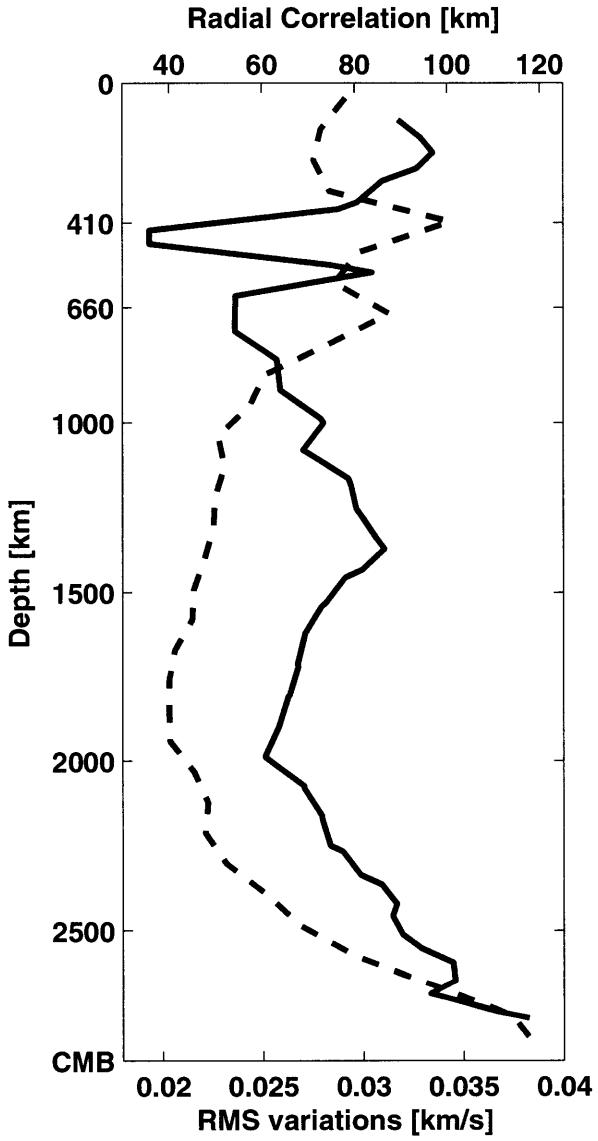


Figure 4-12: Changes in the root-mean-square of structure (broken line) and the radial correlation (solid line) for model based on a regular grid.

be traced to the core mantle boundary; some kind of a changeover in the convective pattern seems to take place around between 1700 km and 2000 km depth. Furthermore, in Figures 4-8-4-10, no slabs seem to reach the very deep mantle, but the inversion tests in Figure 4-11 suggest such a connection would have been observed. These observations confirm the

more tentative results that prompted *van der Hilst and Kárason* [1999] and *Kellogg et al.* [1999] to explore the hypotheses of a dense deep layer with an undulating boundary. This model seems to reconcile a vast array of observations [*Albarède and van der Hilst*, 1999], in particular, some previously contradicting geophysical data which, along with the heat budget of the Earth, advocate against whole mantle convection. An alternative explanation of these observations is that along with the break-up of Gondwana around 200 million years ago, the geometry of subduction around the globe was completely rearranged. Since no older subduction system may have survived, the current ones might all be too young to have reached deeper than mid-mantle depths. However, current evidence is not conclusive, and in the bottom 1000 km or so the nature of mantle convection remains enigmatic.

4.5.2 Different styles of subduction

As a result of the use of an adaptive grid, the connection from the intricate upper mantle structures to much larger remnants of subducted material in the lower mantle is much better represented within this model than our previous models using uniform parameterization.

The viscosity increase associated with the transition from upper to lower mantle and the negative Clapeyron of the postspinel phase transformation at 660 km depth impede the flow into the lower mantle. The effectiveness of this barrier varies greatly according to Figures 4-8-4-10. Some slabs seem to sink into the lower mantle, for example southern Java, Tonga, Kuril Islands, central and south America and show little distortion in the transition zone. In the Marianas, east Philippines, Vanuatu and Aegean the trajectories are severely deformed, but still connected to the lower mantle. Others seem confined to the upper mantle, i.e. New Britain and Hellas (depending on interpretation). Where subduction is recent, it may not have reached the lower mantle or not enough mass has accumulated to dynamically overcome the effects of the negative Clapeyron slope. Accumulation may in some cases be explained simply by the thickening of slabs as the rate of sinking decreases in the viscous

lower mantle. However, some slabs are clearly trapped in the upper mantle, if a relatively young and hot oceanic floor is stagnant for some time it may heat up and loose its excess negative buoyancy.

The fate of slabs remains uncertain. Most of the deep slabs reach no deeper than ~1700 km depth and an apparent connection to the CMB is only observed very locally. This does not seem to be related to any changes in data coverage, both according to tests in Figure 4-11, where most deep anomalies seem to be adequately resolved, and to more generic tests [i.e. Figure 4-6; *Kárason and van der Hilst*, 2001].

It is the interplay of many factors that determines the style of subduction [e.g. *van der Hilst and Seno*, 1993; *Zhong and Gurnis*, 1995; *Jarrard*, 1986; *Griffiths et al.*, 1995], for example the age of the subducting seafloor and therefore its strength, thickness, temperature and the magnitude of negative buoyancy, the characteristics of the overriding plate, the rate of subduction and rate at which the trench moves. The subduction history is important as any or all of these parameters may have changed on the time scale it takes the slab to sink a few hundred kilometers. The variety of how these parameters combine and the uncertainties in estimating them, especially in the past, make it hard single out one variable that might govern the flow pattern. Our model may help in resolving such ambiguities and many alike, for example if gaps in seismicity are due to slab detachment or related in another way to the state of stress within the subducting plate, depth and angle of subduction, and connection to mantle convection in general.

4.5.3 Future work

In this paper we have given an overview of many subduction zones but in future studies this model can be used to look in detail at different convergent margins; interpreting the morphology of subducted slabs in connection with subduction history and assessing the reliability of the model with hypotheses specific inversion tests. In particular, it is promising

to connect the ancient plate motions and margins to the tomographic images using realistic flow modeling, giving quantitative estimates of slab and mantle properties (see Chapter 7).

4.6 Conclusions

To improve tomographic models new data and advances in methodology are both needed. Therefore, to improve sampling we have added seismic phases that are sensitive to different parts of the mantle. Adding data with different ray paths than the P data set used so far might give the perception of improved resolution, but the actual image will only be as good as both the quality of the data and the accuracy of the forward problem calculations allow. We have therefore opted for rather small but high quality data sets, and to treat them properly we have adapted our methodology.

Variable resolution has allowed us to extract more from the signal in our current data. Although subtle, the improvements from the additional data sets are clear. Signal in the PP-P differential times is largely consistent with the other data, and the model based on them alone has many similar traits as model based on traditional travel times [*Kárason et al.*, 2002, see Chapter 3]. The core phases increase our resolution in the deepest mantle, [*Kárason and van der Hilst*, 2001, see Chapter 2 for details]. We believe that this model is well balanced with respect to the different data sets and, in the aggregate, a significant improvement over our previous models. It brings together details not previously present and links them to the larger scale structure which in many areas is better constrained than before.

This model shows the complexity of the flow in the transition zone, which can be related to the different boundary conditions at Earth's surface. However, this not simple, as the number of parameters that affect the manner of subduction is high. Comparing the detailed images obtained here to flow models while adjusting the aforementioned parameters using subduction history as an input can put stricter bounds on parameters that govern

mantle convection, for example, the viscosity contrast between the upper and lower mantle. Furthermore, our model does not reveal continuous structures that traverse the entire mantle. While the depth of the deepest penetration seems variable, most of the deep slabs do not sink deeper than some 1700 km-2000 km.,

This model will be of most value if used along with subduction history to constrain geodynamical flow modeling to investigate the state and nature of mantle convection.

Acknowledgments. This work would not have been possible without the waveform data generously provided by Michael Wysession, Tom McSweeney, Ken Creager, and Guy Masters, and without the 3-D sensitivity kernels calculated for us by Li Zhao. We are also grateful to E.R. Engdahl for processing the extended data base of routinely reported phase arrivals used in this study. This research is supported by the National Science Foundation under grants EAR-9627087 and EAR-9969492.

Chapter 5

Simulating mantle flow in three-dimension and spherical geometry using Green's functions

Abstract

The negative buoyancy associated with the descent of cold and dense oceanic plates into the mantle is the major driving force of mantle convection. Simulating the subduction process realistically is a formidable task since subducting slabs are very thin compared to the dimensions of the mantle and because of the complicated constitutive relationships that are involved. To make the modeling tractable, compromises in geometry, rheology, or resolution have to be made. Restricting the model to two spatial dimensions is the most common simplification, approximating either very long and thin (i.e slab) or axisymmetric (i.e. plume) structures, but allowing for complex rheologies to be tested at an acceptable cost. Global three dimensional models tend to be very expensive, and even when using large computer clusters, the resolution they achieve is on the order $\sim$100 km in the upper mantle, which is similar to the scale of subducting slabs, preventing study of the intricate details of the subduction process. Regional studies can achieve higher resolution in three dimensions, but at high computational cost, often in Cartesian geometry and with limited ability to account for faraway structures and boundaries. High cost may prohibit repeated simulations to validate or reject competing hypotheses.

We present an alternative method based on a Stokeslets approach. Flow at low Reynolds

numbers, like in the viscous mantle, is described by *Stokes* equation. Stokes equation is a linear differential equation, so any linear combination of solutions, or Green's functions, also satisfies the equation. We create axisymmetric Green's functions for small "Stokeslets" and arrange them to simulate three dimensional structures. Surfaces, which represent the boundaries of dense material, are tracked to advance the flow. This way, we can simulate buoyancy driven flow, i.e. subduction, with very high spatial (< 15 km) and temporal resolution in three dimensions, using realistic geometry and boundary condition. The key advantage of this approach is that these simulations are very flexible and can be made relatively cheaply, using much less computer power than the traditional finite-element-based 3-D studies. The major drawback of this method is that viscosity has to be laterally uniform (i.e. no strong slabs).

We introduce the theory behind the code and discuss the development of ideas that made it feasible. We compare the results to theoretical solutions and subject the code to a few consistency tests. Two examples are shown; subduction from a migrating trench with a kink, and the interactions between a subducting slab and a deep dense layer.

5.1 Introduction

Plate motions are not only the surface manifestations of mantle convection, but the subduction of plates is the dominant driving mechanism of mantle convection [*Davies and Richards*, 1992; *Bercovici et al.*, 2000; *Richards et al.*, 2000]. A more complete understanding of the descent of plates into the mantle will thus give valuable constraints on mantle dynamics. Because of the very slow rate of mantle flow, most geophysical observations pertain to the present state of the mantle. Therefore, many questions concerning the nature of mantle flow can only be addressed indirectly by comparing models constrained by subduction history to the few available observables, e.g. seismic tomography, the geoid and viscous response to loading.

The horizontal dimensions of lithospheric plates are much larger than their thickness. Consequently, the descending *slabs* of subducted lithosphere are relatively long and narrow features, with the cross sections of subduction zones often changing very little over large distances along the strike of the convergent margins. This is often used to justify the mod-

eling of subduction in two dimensions, that is, variations along the trench are ignored. This idealized approach has helped answer many fundamental questions concerning mantle flow [e.g. *Gurnis and Davies*, 1986; *Tao and O'Connell*, 1992, 1993; *King and Hager*, 1994].

However, in reality subducting slabs are finite length, 3-D features which vary along the trench. Furthermore, the geometry of subduction zones is often very complex [*van der Hilst and Engdahl*, 1991; *van der Hilst*, 1995], both when looking at the surface expressions of subduction today and when tracing slabs at depth. For large-scale flow the near-spherical shape of the mantle is also important.

In order to understand if some features of slabs as modeled in two dimensions are artifacts because of the unrealistic geometry, comparison between simple models of subduction in 2 and 3 dimensions is necessary. *van Keken and Gable* [1995] argued that 2-D or 3-D modeling of plumes show very different modes of flow as they ascend to the surface. In addition, continuous improvement of seismic imaging in recent years has begun to reveal details that can be tested against various hypotheses about the nature of Earth's mantle [e.g. *van der Hilst et al.*, 1997; *Kárason and van der Hilst*, 2000]. The precise shape of a subducting slab can, for example, be diagnostic of changes in viscosity with depth. For definitive results based on these and other observations from various branches of geosciences it is necessary to make realistic models that take both the exact geometry and tectonic history into account.

Modeling in three dimensions is a challange; the numerical size of the problem increases dramatically and the many degrees of freedom can make the experiments harder to design and the results more difficult to interpret. In addition, slabs are relatively fine detailed structures compared to the dimensions of the mantle and to represent the minutiae of them usually requires immense computer power. The first three dimensional models of mantle flow were made by *Hager and O'Connell* [1979, 1981]. They expanded the current plate motion into spherical harmonics and used them as boundary conditions as they solved

for large scale mantle flow with depth dependent viscosity. Later, a similar approach was used to estimate buoyancy driven flow using seismic heterogeneity from seismic tomography as a proxy for density anomalies [*Hager et al.*, 1985]. This approach becomes very expensive for small-scale flow as the number of the coefficients used to parameterize the flow field increases very rapidly with increased resolution.

The aforementioned calculations were only for the instantaneous flow field resulting from the current configuration. Other studies have used simplified mechanics to relate subduction history and the distribution of ancient slabs in the mantle [*Ricard et al.*, 1993; *Lithgow-Bertelloni and Richards*, 1998; *Richards et al.*, 2000]. In this approach *slablets* are assumed to sink straight down from migrating trenches with the same velocity as the 'attached' plate as estimated from subduction history. When entering the lower mantle the slablets are assumed to be slowed down by some *factor*. Expanding the resulting density anomalies into low degree spherical harmonics and calculating the correlation with low degree models based on seismic tomography gives the best fitting *slowing factor* which, is then interpreted in terms of viscosity contrast between the upper and lower mantle (See Chapter 6 for a detailed treatment of the slowing and thickening of subducting slabs). This approach only allows for comparison between plate reconstructions and seismic tomography on a very large scale, and, as acknowledged by the authors, errors on the order of 1000 km in slab location are not significant for their findings [*Lithgow-Bertelloni and Richards*, 1998].

Bunge et al. [1998] studied transient flow in a convection simulation, with a 40-fold increase in viscosity in the lower mantle and lithosphere. In this simulation, boundary conditions were abruptly switched from $\sim$110 Ma old to a modern day plate geometry and motion. Before switching, the model had been running for a long time and had reached a quasi-steady state; it took the flow $\sim$150 Ma to adjust to the present day plate motions. Comparison with tomography was encouraging, especially beneath the major subduction

zones of the older plate configuration [*Bunge and Grand*, 2000].

Another class of flow modeling uses temperature and depth dependent viscosity to create self consistent convection models in spherical [*Zhong et al.*, 2000] or Cartesian [*Tackley*, 2000] geometry. Here the comparison with subduction history is qualitative, e.g. length scales of plates, one-sided vs. two-sided subduction, and toroidal vs. poloidal energy in plate movements.

These recent global simulations all use a finite element scheme and have element spacing of approximately 50 to 100 km throughout the mantle [*Bunge et al.*, 1998; *Zhong et al.*, 2000], which is close to the thickness of the lithosphere and slabs in the upper mantle. In order to be able to resolve the cold downwellings adequately in such calculations, the Rayleigh number, which is measure of convective vigor, is kept lower than that estimated for the mantle, resulting in thicker boundary layers. Still, these calculations are expensive and time consuming [*Moresi et al.*, 2000]. For example, the study by *Bunge et al.* [1998] required over 10 million elements and computing resources from the Los Alamos National Laboratory. To double the resolution requires 8 times more elements and 2 times the number of time steps and the high computational cost prohibits repeating the calculations, necessary to test hypothesis against observation.

Here, we develop a method that requires less computational power but still can produce realistic flow patterns resulting from subduction in a three dimensional spherical medium with slabs of dimensions found in the Earth. The method is based on a Green's function approach to the linear Stokes equation which governs viscous flow. The resulting code depicts boundaries of anonymously buoyant material, i.e slabs, plumes or layers, with surfaces. Our methodology resembles *Manga* [1997] in that they used simple surfaces to represent a handful of buoyant drops in an infinite isoviscous medium. However, their way of solving Stokes equation, using surface integrals, is different from ours.

With our new code, subduction from a single trench a few thousand km long and with

spatial resolution of < 15 km can be run in a day on a small desktop computer. The simulations show in substantial detail the shape of the subducting material, complications associated with varying viscosity structure, and geometrically complex or time varying boundary conditions are easily implemented. Unlike full convection codes [e.g. *Tackley*, 2000; *Bunge and Grand*, 2000; *Zhong et al.*, 2000] it does not solve the energy equation.

Furthermore, in our modeling, slabs are inherently of the same viscosity as the surrounding mantle. Estimates of temperature dependent viscosity and thermal modeling imply that slabs might be many orders of magnitude more viscous than the surrounding mantle [*Moresi and Gurnis*, 1995]. However, since slabs are thin, their ability to guide mantle flow is less than their viscosity suggests. Moreover, the estimated viscosity increase is on the level of minerals, while the slab might be mechanically weakened due to faulting [*King*, 2001]. *Moresi and Gurnis* [1995] modeled in 3-D the subduction process in the Western Pacific and calculated an inferred geoid. They concluded that strong slabs in the upper mantle were not consistent with the observed geoid and that slabs somehow loose their strength at relatively shallow depths. Slab deformation inferred from seismicity, together with laboratory and numerical experiments, allowed *Becker et al.* [1999] to conclude that slabs are sculpted by the background flow in the mantle. They further deduced that a viscosity contrast of 500 between the slab and the surrounding mantle had insignificant effects on the flow. Furthermore, *Tao and O'Connell* [1993] modeled *weak* slabs, which only differed in density from the surrounding mantle, and were able to explain many inferences from seismology about the shape and state of slabs, such as deep seismicity. Toward the other limit, *Manga and Stone* [1993] showed that the dynamical difference between buoyant isoviscous and relatively inviscid drops is rather small. Although not conclusive, these studies all suggest that calculations where the higher viscosity of slabs is neglected can still provide useful insights about mantle dynamics. Also, in the current version of our code, the flow is incompressible. This is only a limitation of the two dimensional finite element

code currently used to create the two-dimensional Green's functions, as described later, but not a fundamental limitation of the method.

In this chapter the theory, development, and validation of the code is outlined and examples are shown. In Chapter 6 this code is applied to examine the thickening and slowing of subducting slabs with varying viscosity contrast between the upper and lower mantle, and in Chapter 7 it is used to bridge a gap between plate reconstructions and seismic tomography.

5.2 Theory

5.2.1 Basic equations

An incompressible steady flow with uniform viscosity is described by the *Navier-Stokes* equations which are in the absence of body forces [*Batchelor*, 1967]:

$$\frac{1}{\rho}\boldsymbol{\nabla} p = \frac{\mu}{\rho}\nabla^2 \mathbf{u} - \frac{D\mathbf{u}}{Dt} \tag{5.1}$$

$$\boldsymbol{\nabla} \cdot \mathbf{u} = 0 \tag{5.2}$$

Where ρ is density, p the pressure field, μ the dynamic viscosity and $\mathbf{u}$ the velocity. Eq. 5.1 is the *Navier-Stokes* equation of motion and eq. 5.2 is the conservation of mass for incompressible flow. The last term in eq. 5.1 is the local acceleration of a particle in the flow. If the flow is very slow and viscous the acceleration is negligible and this term is approximately zero:

$$\frac{D\mathbf{u}}{Dt} \approx 0 \tag{5.3}$$

Eq. 5.1 then reduces to the *Stokes* equation, which is linear [*Batchelor*, 1967]:

$$\frac{\boldsymbol{\nabla} p}{\mu} = \nabla^2 \mathbf{u} \tag{5.4}$$

The ratio between inertia forces and viscous forces is a measure of how good an approximation eq. 5.3 is. This ratio is the Reynolds number, Re.

$$\mathrm{Re} = \frac{\rho d u}{\mu} \tag{5.5}$$

Here, d is some characteristic width which governs the scale of the flow. For $\mathrm{Re} < 0.5$ eq. 5.4 is generally considered a valid approximation. The Earth's mantle is very viscous and it can easily be verified that the Reynolds number is very low. Some typical values for a slab sinking in the mantle are, $d \approx 100$ km, $\rho \approx 5000$ kg/m^3, $\mu \approx 5 \times 10^{21}$ Pa s, $u \approx 0.1$ m/year, which gives Re $\sim 10^{-22}$.

5.2.2 Analytical solutions

Tractable solutions of eq. 5.4 only exist in simple cases, but they can provide insight into the nature of *Stokes* flow. The most relevant one is the axisymmetric and irrotational flow resulting from a sphere moving in an infinite medium which is given in *Appendix A*. The sphere can have different viscosity than the surrounding material and eq. 5.4 is solved, both inside and outside the sphere. It is worth noting that the flow field is only dependent upon the ratio between the outer and inner viscosity but not the magnitude. From Eq. A.10 it can be seen that the outer solution has two terms both for u_r and u_θ. One terms falls as r^{-1} but the second term falls off as r^{-3}, i.e. close to the moving object rapid variations occur. These variations die out quickly and at farther distances the r^{-1} term takes over and governs the flow. At these distances the flow field is only dependent on the total excess density and the location of the sinking object. This makes calculations of *Stokes* flow in more complex situations troublesome. While needing high resolution close to the moving

objects, incorrect assumptions like inaccurate boundary conditions, will affect the results far away.

5.3 Methods

5.3.1 Linearity of Stokes equation

Since *Stokes* equation is linear, i.e if $\mathbf{u_1}$ and $\mathbf{u_2}$ both satisfy eq, 5.4, their sum will also a solution:

$$\frac{\boldsymbol{\nabla} p_1}{\mu} = \nabla^2 \mathbf{u_1} \text{ and } \frac{\boldsymbol{\nabla} p_2}{\mu} = \nabla^2 \mathbf{u_2} \implies \frac{\boldsymbol{\nabla}\,(p_1 + p_2)}{\mu} = \nabla^2(\mathbf{u_1} + \mathbf{u_2}) \tag{5.6}$$

Moreover, the flow due to an assemblage of loads will be equal to the sum of flow fields from each load by itself. This is the basis of our approach. We model the Earth's mantle as spherically symmetric in shape, boundary conditions and viscosity. This allows us to create *Green's functions* which are axisymmetric for axisymmetric loads, and therefore two dimensional, but can be combined in any way to construct the flow field from a three dimensional object.

An obvious approach would be to arrange point masses or *Stokeslets* into some configuration and then track each Stokeslet through time, calculating interactions with all the other Stokeslets (see Figure 5-1a). If L is some characteristic scale of the object being modeled, the number of calculations would scale as the number of Stokeslets squared, i.e. with the square of the volume, or L^6. Also, accounting for changes in distance between Stokeslets as the flow progresses is not trivial using this setup.

A more elegant solution is to have the Stokeslets stationary, but have a surface that moves with the flow (see Figure 5-1b). Whether a stationary Stokeslet is inside or outside

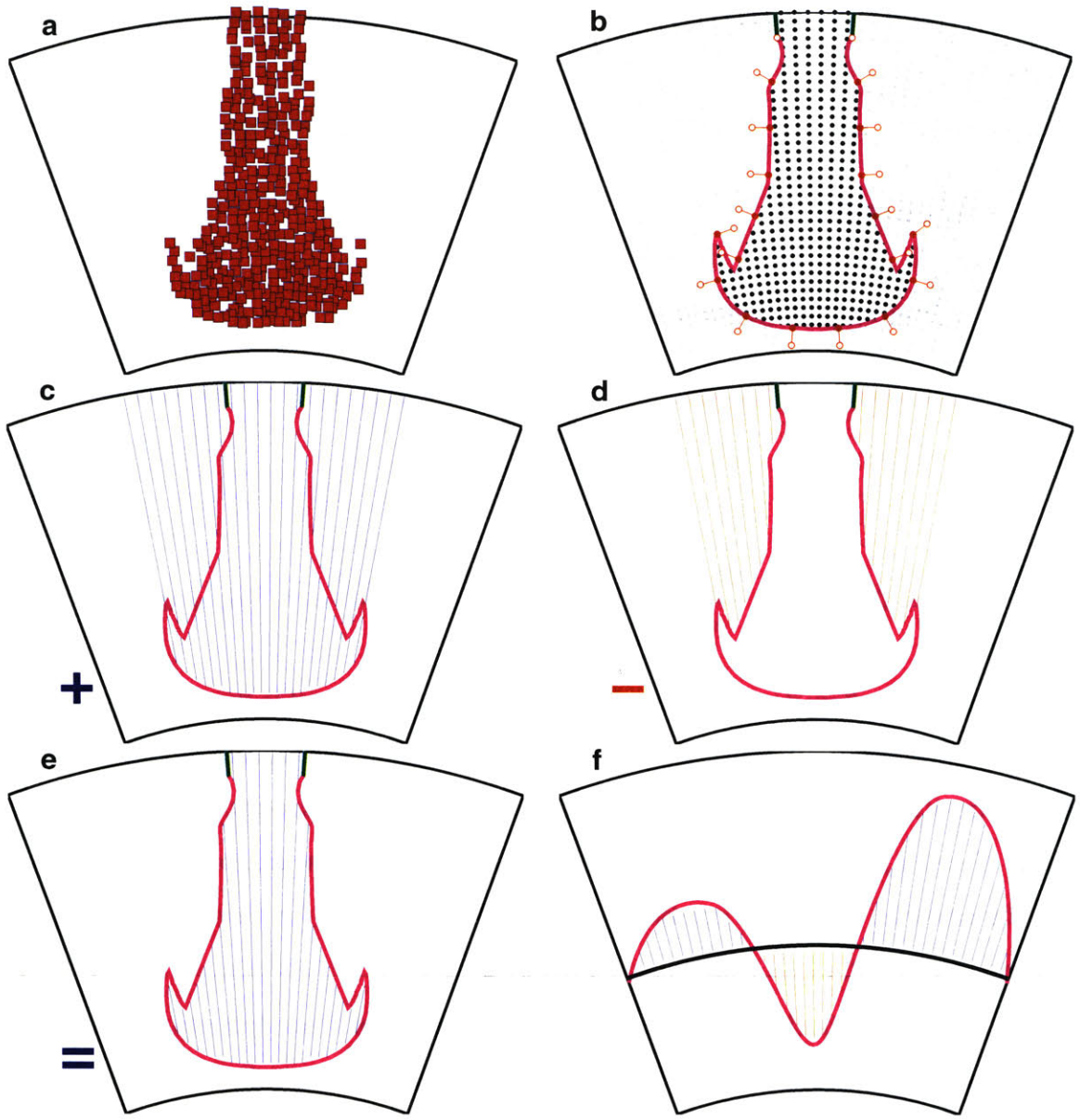


Figure 5-1: A 2-D rendering of different ideas behind the three dimensional flow calculations. (a) A collection of moving Stokeslets that interact and drive the flow. (b) The purple and green lines are a 2-D rendering of a 3-D bag which content is buoyant and drives the flow. Material can flow in or out of the bag through the green part which is stationary but the purple part of the bag moves with the flow. The flow is approximated by summing Green's functions from the Stokeslets that are inside the bag (black dots); other Stokeslets are inactive (gray dots). The purple surface is tracked by estimating the velocity at the nodes (solid red circles), selecting a small time step and moving them to their new position (hollow red circles). An equivalent but computationally efficient version, reducing drastically the number of Stokeslets involved, is to assume that Stokeslets start at the top of the mantle and extend either to the bottom of the bag (c) or to its surface (d). Subtracting the latter from the former results in (e) the same load and flow pattern as in (b). (f) A dense deep layer modeled by negatively buoyant (blue columns) and positively buoyant (red columns) of Stokeslets that extend from the layer's neutral position (black line) and the surface.

of the surface determines if it contributes to the flow. Only the surface is then tracked through time. An even more efficient version is to assume that the Stokeslets extend from the surface of the Earth as shown in Figures 5-1c-e. Here the number of calculation scales as surface area squared or L^4, and depth resolution is almost free. The surface is represented by a set of triangles whose corners are the nodes that are tracked through time.

Similarly, a dense layer can be modeled by a surface, see Figure 5-1f. When the layer is deflected downward, the material between the layer and its neutral position (black line) is positively buoyant (red columns) and conversely negatively buoyant (blue columns) if the layer moves up.

5.3.2 Green's functions

The Green's functions are axisymmetric and therefore two dimensional. To solve Stokes equation in this geometry we use the spherical version of *Conman*, a finite element code, developed by Scott King and others [*King et al.*, 1990; *Kellogg and King*, 1993]. Boundary conditions are no flow in and out of the mantle and no flow in the tangential direction on the axis of symmetry. Stokeslets are created by creating loads of a given radius on the axis of symmetry (i.e. north pole) and tabulated for different depths. The flow field consists of tangential, v_θ, and radial, v_r, velocities for all depths and distances from the north pole. Green's functions must be created for each viscosity profile that is to be modeled. In the studies presented here it is most commonly the viscosity contrast between the upper and lower mantle that is changed.

Other effects can also be accounted for in the Green's functions. The thermal expansion coefficient, α, is thought to decrease with increasing temperature and pressure within the mantle and at the CMB be it may be only 30% of the surface value [*Anderson*, 1989]. This change translates directly into the effective density that drives the flow, given the same thermal contrast between the slab and the surrounding mantle. The effective density

can be changed as function of depth during the simulation, but it is more effective and straightforward to account for this in the Green's functions themselves.

5.3.3 Advancing the flow

Given a surface of a particular shape and its mathematical description, the first step is to calculate intersection of this surface in radial direction and determine which combination of Stokeslets represents the material that is within this surface like illustrated in Figure 5-1c-e (or Figure 5-1f in the case of a layer). Although not indicated in Figure 5-1 the code can handle folded surfaces that have more than two intersections in depth for a given location.

For each node that defines the surface, contributions from all the Stokeslets are accrued. Effectively this means interpolating the Green's functions for v_θ and v_r, using the angle between the node and the Stokeslets and their depth. The radial velocity contributions can be summed directly, whereas the tangential velocity is a vector quantity with two components. Multiplying the velocities with a time step gives the new locations of the node. This step is selected so that the maximum shift of any node, in either tangential or radial direction, is kept within some distance, usually 10 km. Each node is then rotated and its radius adjusted. Additionally, in some runs, trench migration is imposed by moving the top of the surface (the green part in Figure 5-1) laterally at some fixed rate.

The most difficult part of a simulation is get the surface to remain stable using a limited number of nodes to represent it. As the surface expands due to the flow, new triangles and nodes are introduced where the surface has stretched the most. When the total number of nodes reaches a certain limit the surface is re-meshed using a smaller number of nodes. This step is achieved with the public domain software *qslim 1.0* by Michael Garland, see *Garland and Heckbert* [1997].

Once a certain process has been simulated, the whole simulation does not take much space, since the number of nodes is kept low, and can easily be stored on a disk for all

time steps. This means that inactive flow tracers can later be added and tracked through the simulation at a fractional cost of running the whole simulation again.

5.4 Tests

5.4.1 Comparison with theory

As mentioned before, tractable solutions of Stokes equation exist only in simple cases. Here we test our code against the theoretical solution for a sphere sinking in an infinite medium, with equal viscosity inside and outside the sphere (see appendix A for a derivation). The nominal velocity of the sphere is the velocity at its surface in the direction that it is sinking. Within the sphere, the material flows and velocity at the center is actually higher than at the surface. In Figure 5-2 we compare the theoretical solution for a sphere of equal viscosity to the surrounding material to a solution obtained by using the 2-D axisymmetric finite element code directly, which is possible since the flow field from the sphere is axially symmetric, and the 3-D flow code. The excess density of the sphere was adjusted such that the nominal velocity for the theoretical solution is exactly 1. The sphere is of radius 100 km and is placed in the middle of a uniform viscosity mantle, represented with a spherical surface for the 3-D code.

Even though the theoretical solution is not directly comparable to the flow field resulting from a sphere sinking within the finite mantlem they have have similar characteristics close to and inside the sphere, as is to be expected. The difference between the theoretical and 2-D solution is probably mostly real, i.e. the result of different boundary conditions. Flow due to a sphere constrained by infinite surfaces can be estimated by a combination of spheres. Such calculations predict that the theoretical solution in an infinite medium should be on the order of 5% higher than when constrained by walls at similar distances as the surface and the CMB are from the mid-mantle. The 3-D solution is made on a somewhat coarser grid

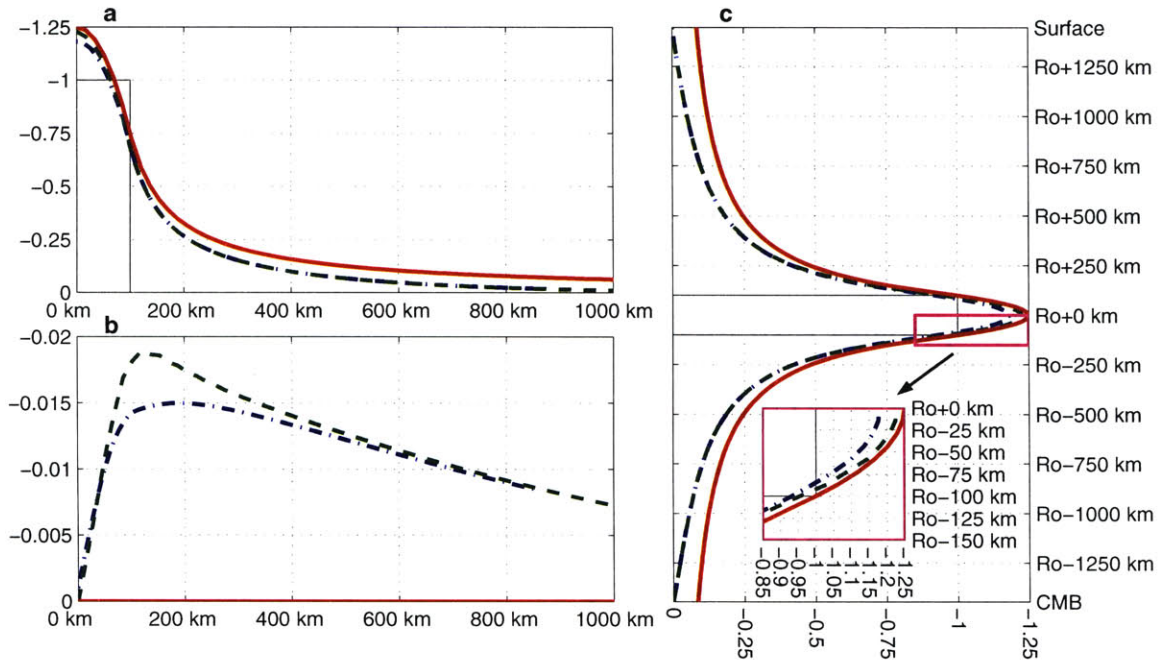


Figure 5-2: Red lines are the theoretical solution resulting from a sphere of radius 100 km (radius indicated with gray box) sinking in an infinite medium. The sphere and the surrounding medium are isoviscous. The blue and green lines show the solution for such a sphere, located at mid-mantle, sinking in a uniform viscosity mantle, obtained by using the axisymmetric 2D finite element code and the new 3-D code, respectively. Density is adjusted so that the nominal speed of the theoretical solution, i.e. the velocity in radial direction at the surface of the sphere, is 1. (a) Radial velocity and (b) tangential velocity as function of angular distance from the sphere. (c) Radial velocity as function of depth, Ro being the radius of the center of the sphere.

than the 2-D solution and most of the differences between these two solutions can probably be attributed to that. The highest relative difference between the two numerical solutions is for the tangential velocity, which is zero for the theoretical solution due to symmetry, but in magnitude that component is about 1/100 of the radial velocity. According to these tests, the consistency between the 3-D code and either theoretical calculations or the 2-D flow code is very acceptable.

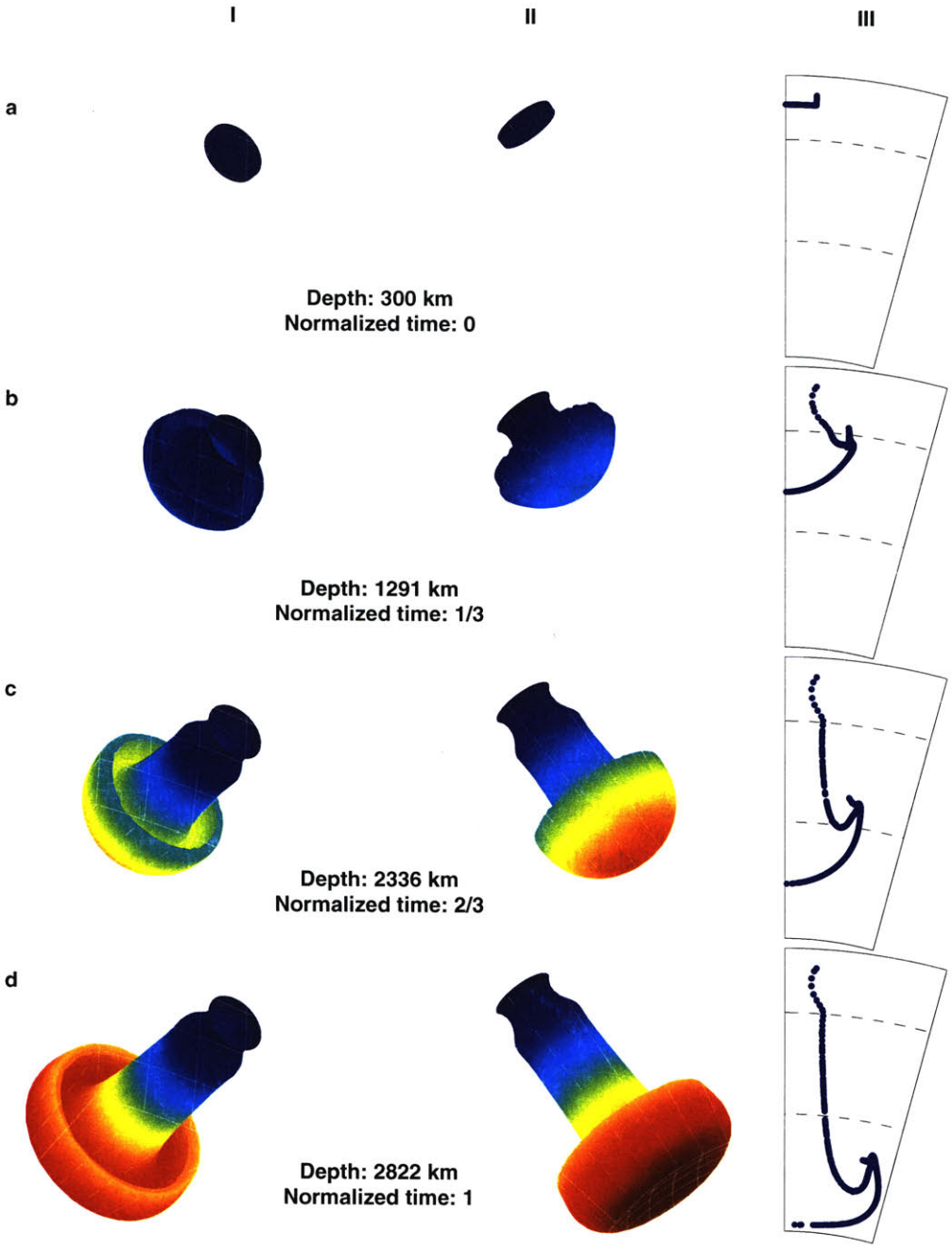


Figure 5-3: Results from a run starting with an axisymmetric surface (a) that sinks in a mantle (b-d) with 30-fold increase in viscosity at 660 km depth. Time is normalized with the total time of the simulation. Column I and II show different views of the slab as it sinks. In column III the distance from the initial axis of symmetry to all the nodes that define the surface is plotted as function of depth.

5.4.2 Preservation of symmetry and volume

The above comparison with theory is only possible for the instantaneous velocity field because a sphere of dense material deforms within the finite mantle as the flow progresses. It can therefore not be used to check if inconsistencies arise when the flow is tracked through time, e.g. due to programming errors or for too large time steps between iterations. Another diagnostic test that would reveal many internal inconsistencies and give a feeling for numerical accuracy is to start with an axisymmetric surface, let it evolve with time and see how the symmetry is preserved. In figure 5-3 the results from one such run are illustrated. The "slab" starts as a circular disk, which consequently sinks, develops a "plume" head and thickens in the 30-fold more viscous lower mantle. The symmetry remains intact, i.e. for a given depth the intersections with the surface are at almost the same distance from the initial axis of symmetry in all directions as can be seen in column III.

Furthermore, the flow field is incompressible and it can be tested how much material flows into the surface vs. how much the volume enclosed by the surface increases. The comparison in Figure 5-4 shows that the mismatch in volume, estimated by these two measures is only about 2% as the sinking slab expands to the bottom of the mantle. During this test the volume of the sinking slab has increased $\sim$75-fold and the surface area $\sim$30-fold, requiring it to go through series of re-meshings. Considering this, the results are very good and it can be said with confidence that numerical errors or other inconsistencies are not dominating the simulations.

5.5 Time scale

The flow calculations are made using dimensionless parameters. They can then by scaled by an arbitrary selection of real parameters to give mantle-like velocities and time. Which time scale to use when presenting results is likewise a matter of preference. We decided

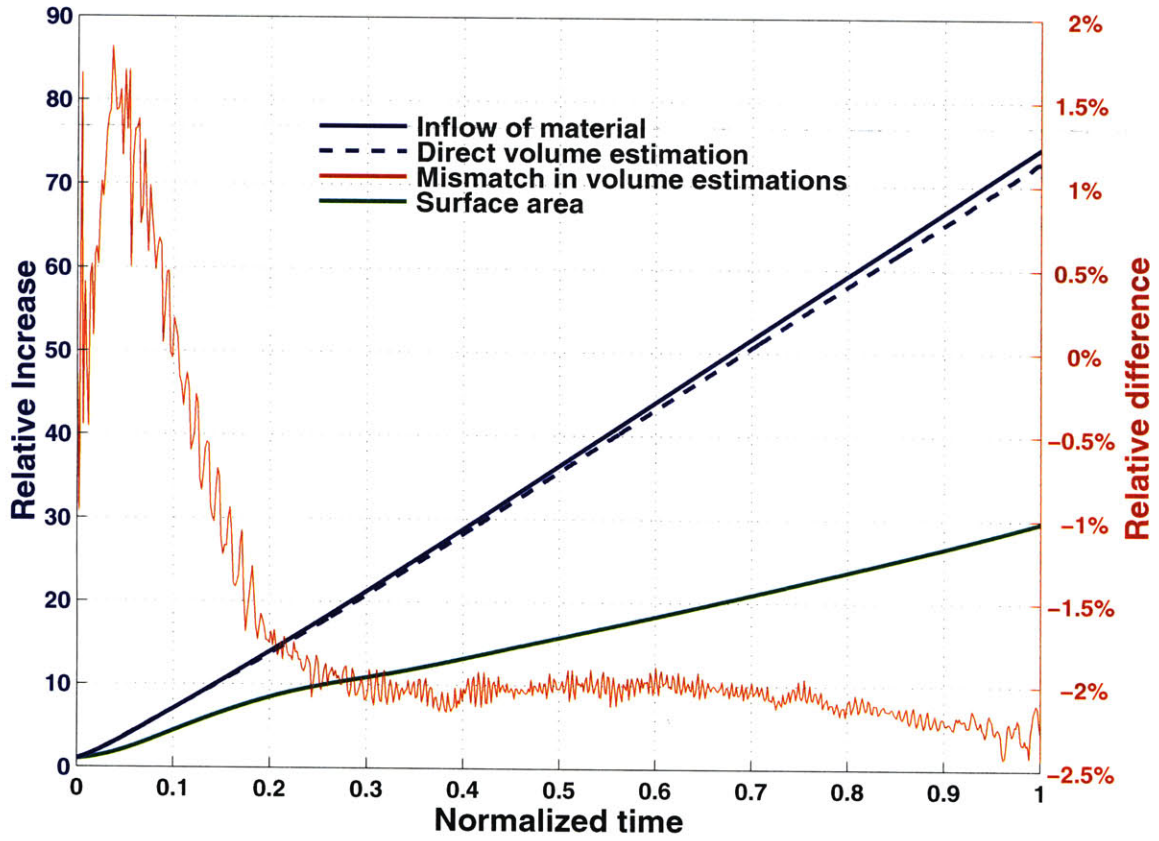


Figure 5-4: (Left axis) The relative increase in volume (normalized to be 1 at time 0) as a function of time using either the inflow of material (solid blue line) or a direct volume estimation from the simulation shown in Figure 5-3. Also shown is the relative increase in area of the surface containing the dense material. (Right axis) Relative difference between the two volume estimates (red line).

to reference our time to that of sinking sphere in an infinite medium, the reason being that it is a well understood theoretical solution and can give some insight into the effects of boundaries vs. a limitless medium. Our typical runs start with a slab extending to 300 km depth and they are usually terminated when the deepest part of the slab has reached 2850 km depth, which is ~40 km above the CMB. This length, $l_o = 2850\ \mathrm{km} - 300\ \mathrm{km} = 2550\ \mathrm{km}$ is the reference length for the sinking sphere. Its instantaneous velocity is given by *Batchelor* [1967]:

$$\mathbf{U_o} = \frac{4}{15}\frac{\Delta\rho_o a_o^2}{\mu_o}\,\mathbf{g} \tag{5.7}$$

where $\Delta\rho_o$ is the excess density of the sphere, a_o its radius, μ_o the viscosity of the sphere and the medium and g the gravitational acceleration. A sphere in an infinite medium, which flow field is governed by *Stokes* equation, is stable and does not deform, therefore this equation is valid for the whole length span, l_o. The reference time is given by:

$$T_o = \frac{l_o}{U_o} = \frac{15}{4}\frac{\mu_o}{\Delta\rho_o a_o^2 g} l_o \tag{5.8}$$

We will use values typical for the shallow mantle and as our slabs are typically 100 km thick we choose $a_o = 50$ km. For viscosity we choose $\mu_o = 3 \times 10^{20}$ Pa s which is the average upper mantle viscosity according to recent study by *Forte and Mitrovica* [2001]. The excess density is estimated according to the cooling of oceanic lithosphere using values for thermal expansion ($\alpha = 3.1 \times 10^{-5}$ K^{-1}), mantle density ($\rho = 3,330$ kg/m^3) and average cooling of the lithosphere ($\Delta T = 725$ K) from *Stein and Stein* [1992] which gives $\Delta\rho_o = \rho\alpha\Delta T = 75$ kg/m^3. Using $g = 10$ m/s^2 then gives $T_o \approx 49$ Ma and $U_o \approx 53$ km/Ma. This reference time and velocity can be changed to T and U using:

$$\begin{aligned} T &= \frac{\mu}{\mu_o}\frac{\Delta\rho_o}{\Delta\rho}\,T_o \\ U &= \frac{\mu_o}{\mu}\frac{\Delta\rho}{\Delta\rho_o}\,U_o \end{aligned} \tag{5.9}$$

where μ and $\Delta\rho$ are the new upper mantle viscosity and the new excess density. In figures that follow in this and later chapters these reference values are usually indicated as they emphasize that the time and velocity scales are somewhat arbitrary and illustrate how the results can be scaled and the relative importance of geometries and boundaries.

5.6 Examples

In Chapters 6 and 7 we show more detailed studies on different facets of the subduction process. Here we show two examples. The first example shows how geometrically complex and time varying boundary conditions can easily be implemented. The second example illustrates the interactions between two large bodies, subducting slab and a deep dense layer, and how they interact over large distances.

5.6.1 Migrating trench with kink

The initial geometry of the subducting slab modeled here is shown in Figure 5-5. The shape of the trench in this case can be described as a trench with a 120° kink in its center. The trench is migrating at a relatively fast pace, first to the left with respect to Figure 5-5a but after $\sim$27 million years, using the values from section 5.5 the migration direction is reversed. In this case the viscosity in the lower mantle increases 30-fold.

In Figure 5-6 the resulting slab is shown from above and from the side at 10 Ma intervals into the subduction process from 10 Ma to 40 Ma. The subducting kinked slab develops a complicated "plume" head, which broadens as it sinks, both upon entering the lower mantle but especially above the core-mantle boundary, where material starts to accumulate after $\sim$30 Ma. After the reversal a secondary "plume" head is created.

At the last time step in Figure 5-6 the parameterization of the surface of the slab with $\sim$10,000 triangles is indicated. The final surface is folded onto itself in a complicated way but nonetheless the features of the trench geometry survive. The edges are well preserved and can be traced from the trench to the deepest parts of the slab. Likewise the initial head is still clearly visible, although highly deformed. These runs show that although this surface has been re-meshed around twenty times it still retains all of its important characteristics.

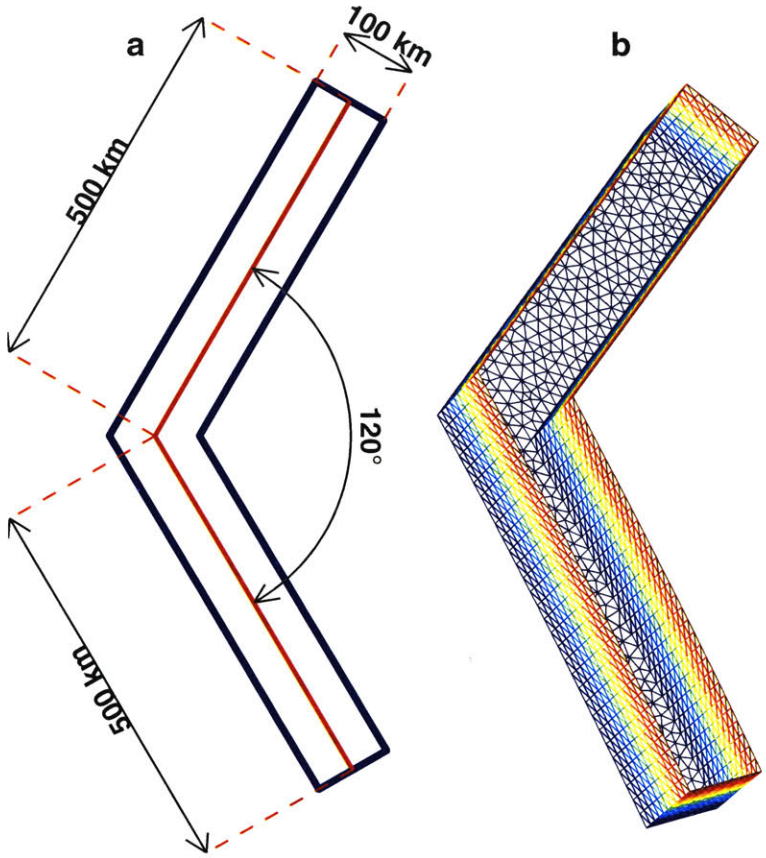


Figure 5-5: (a) The fixed geometry of the trench. (b) Meshed surface at time 0 before simulation starts.

5.6.2 Interactions with a deep layer

Recent studies have pondered the existence of a hot layer deep in Earth's mantle as a way of reconciling otherwise contradictory observations from different fields of geophysics and geochemistry [*van der Hilst and Kárason*, 1999; *Kellogg et al.*, 1999; *Anderson*, 2001]. This layer has to be convectively isolated from the above mantle to help explain apparent discrepancies in the geochemical budget and heat balance of the Earth. If the effective density of the hot layer is only slightly more than the average density of the overriding material its upper boundary may have significant topography, i.e. anomalously heavy material, such

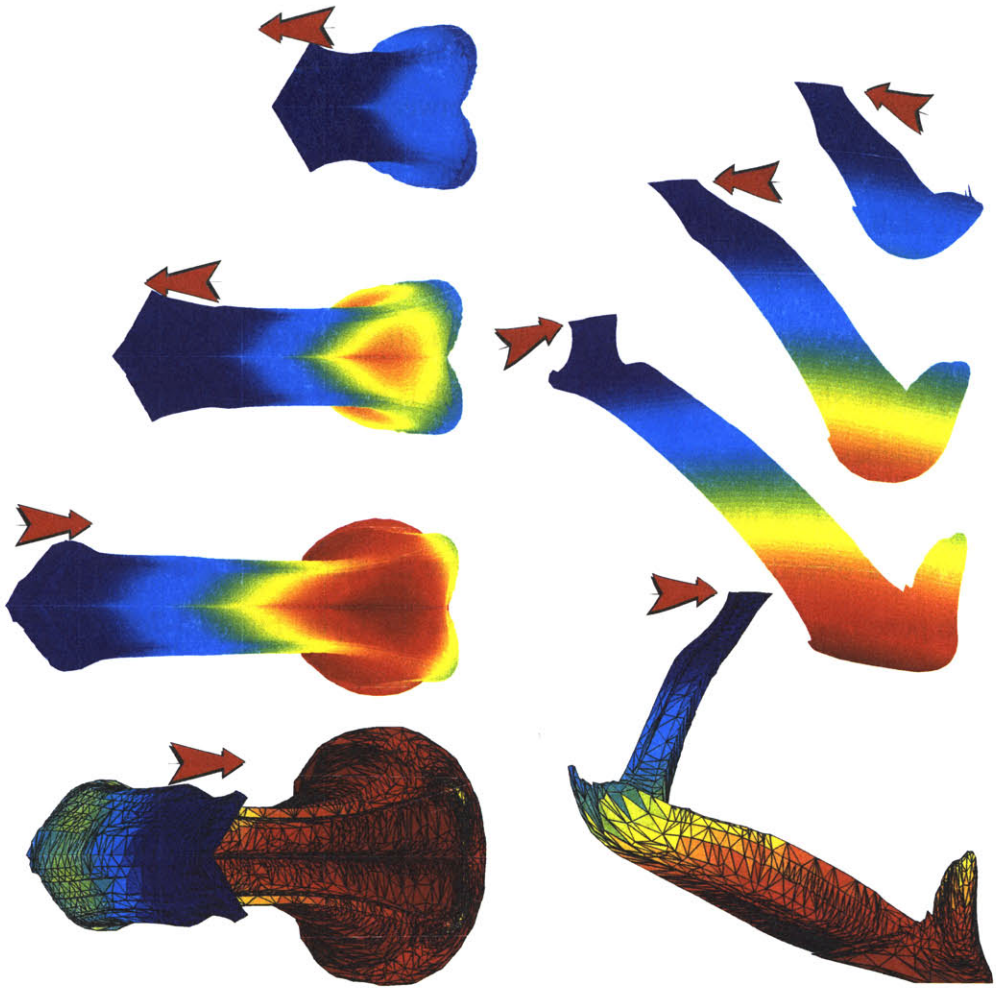


Figure 5-6: Simulation of a migrating (with a reversal after ~27 Ma) trench with a kink as shown in Figure 5-5 and 30-fold increase in viscosity in the lower mantle. Viewed from above (left column) and from the side (right column) after 10 (top), 20, 30 and 40 (bottom) Ma from the start of simulation. At the last time step the triangulation of the surface with ~10.000 triangles is indicated.

as cold sinking slabs, may be able to penetrate this layer and even reach the core mantle boundary. An interface atop such a layer has so far escaped detection but the tomographic model of the mantle represented in Chapter 4 provides some indirect evidence that support a different convective regime in the very deep mantle, i.e diminished radial continuity of structures around 1700 km depth. The convective patterns above and beneath these depths seem dissimilar and analysis of wave-speed ratios also suggests large-scale lateral varia-

tions in composition [*Saltzer et al.*, 2001].

To investigate which density contrasts between the sinking slab and the deep layer are needed to produce a significant level of topography on the boundary such a layer and a sinking slab, in a mantle with 30-fold increase in viscosity in the lower mantle, were simulated. Some of the results are illustrated in Figure 5-7. In Figure 5-7a the effective density of the layer (i.e. the density excess compared to the overriding mantle) is half the effective density of the slab. In this case the sinking slab is heavy enough to push the layer aside and finally reaches the core-mantle boundary. The slab however broadens above the layer and the plume "head" is much broader above the layer than within it. In Figure 5-7b the layer has twice the effective density of the slab. Here the slab makes a dent in the layer but is not able to penetrate it but instead broadens above the layer. Figure 5-7c shows that a rapidly migrating slab with 80% of the layers excess density makes a smaller dent in the layer.

In Figure 5-8a the depth of the tip of the slab and that of the deepest part of the layer is shown as function of time. For reference a simulation without a deep layer and the analytical reference values from Chapter 5.5 are also shown. In the upper mantle all the runs are almost the same, but as they enter the 30-fold more viscous lower mantle they begin to diverge. It should be noted that the slowing of the tip of slab for the migrating trench is partly due to the moving of the trench at the surface (see Chapter 6.4.3). The slabs are slowed down by the layer as soon as they enter the lower mantle and similarly, at that instance the layer starts to deflect noticeably. The heavier the layer, the more effect it has on the sinking rates, and the less it is deflected downward. The heaviest slab can reach the core-mantle boundary, but it takes $\sim$47 Ma instead of $\sim$ 27 Ma. How much the layer is deflected in the other cases depends on the density contrast, but also on how much mass is accumulated above the deflection.

It is obvious that a subducting slab can create significant topography on a deep dense layer, especially if the slab has similar or a little higher effective density than the layer. If

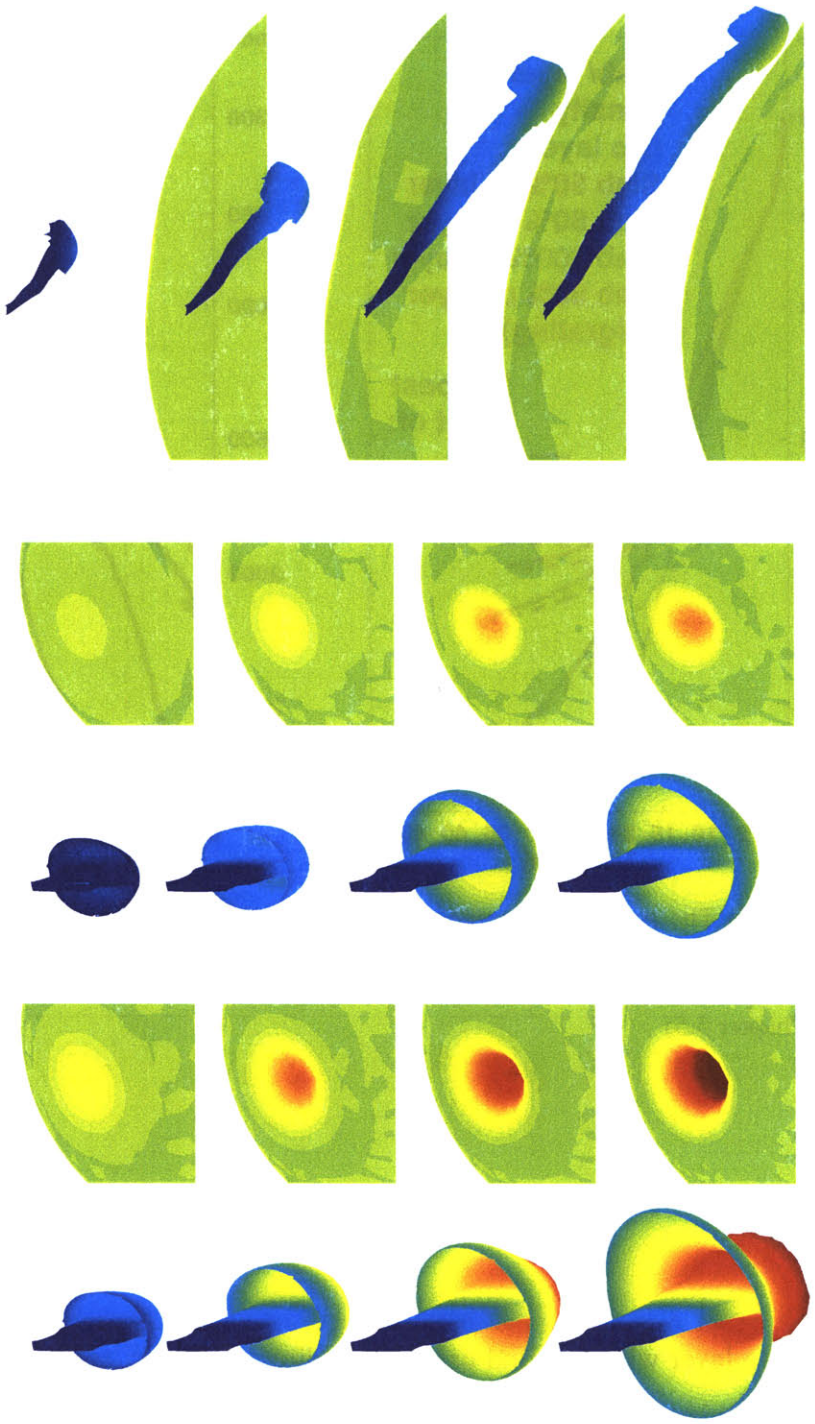


Figure 5-7: Interaction between a subducting slab and deep dense layer. (bottom) 3-D views of layer and slab, where the slabs excess density is 200% of the excess density of the layer. (middle) same as (bottom) except excess density ratio is 50%. (top) Same as (bottom) except ratio is 80% and the trench migrates at the surface.

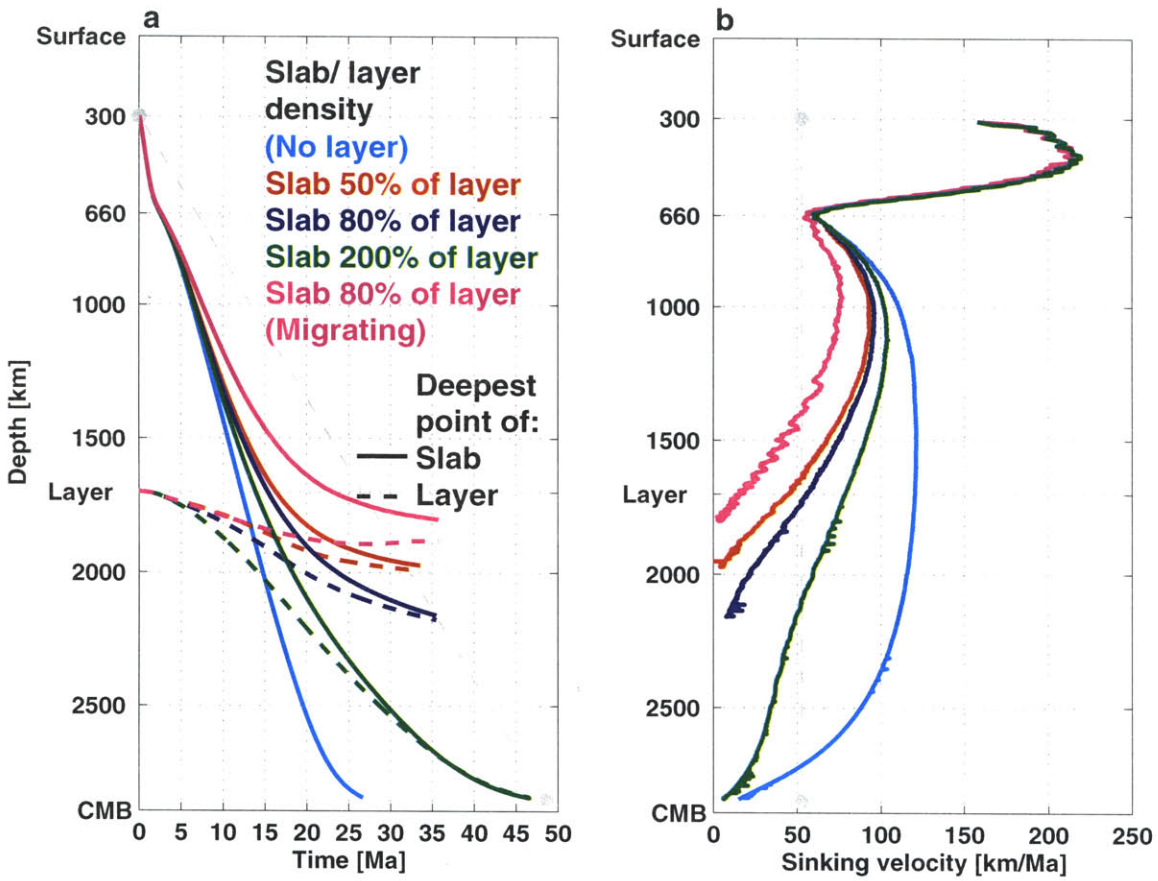


Figure 5-8: (a) The position of deepest point of slab (solid lines) and layer (broken lines) as function of time for different simulations as indicated. (b) The sinking rate of tip of slab as function of position for different runs. Reference values for tie and velocity are indicated with gray-dashed lines (see Chapter 5.5)

evidence for a layer will be found using more direct methods, such as seismic scattering, the subduction history along with modeling of the kind presented here might be able to predict where the layer should be deflected up or down.

5.7 Discussion

5.7.1 Applications to mantle flow

The code presented here has been proven to be an efficient and accurate way to solve Stokes equation. Examples of subduction from a geometrically complex and time varying trench and of interactions with a deep layer have been discussed; in Chapter 6 this code is used to assess in a systematic way the effects of viscosity on sinking rates and thickness of subducting slabs. Many other general questions relevant to mantle flow can be addressed. For example, more details of the subduction process can simulated, such as variation of density within the lithosphere or the effects of a delayed phase transition at the 410 km and 660 km discontinuities. Delamination of dense material beneath a rigid lid could be modeled by changing the boundary conditions in the Green's functions approprietely. Likewise, positively buoyant upwellings beneath either a free surface or a rigid lid could be compared. All these simulations could be performed with different viscosity profiles, examining, for example, return flow within the asthenosphere.

A limitation of the current version of the code is that flow in lateral direction can not be imposed. This has most relevance to the currently observed - and the inferred ancient - plate movements at the surface. A Green's function approach is still valid in addressing this but difficulties arise because the flow field in this case would be three-dimensional. This could be solved by a three-dimensional spherical finite element code but the most natural way is probably to use the spherical harmonics approach of *Hager and O'Connell* [1979]. Every sum of solutions to Stokes equation is still a solution, however they are obtained. By using a rigid lid in the Green's functions for the stokeslets, the calculations of the flow field resulting from the plate movements could be kept completely separate, except that the same viscosity profile would have to be used in both instances. This way, the preferred boundary conditions at the surface are easily applied.

5.7.2 Different geometries

An identical approach as outlined here can be used to solve Stokes equation in other than spherical geometries, but with the severe limitation that boundary conditions and viscosity structure need to be invariant to translation in two directions. As an example, flow in a limitless isoviscous medium can be simulated by replacing the current Green's functions with the theoretical solutions for a sinking sphere. They can then be calculated for a string of spheres, much similar to the columnar Green's functions mentioned before. A variant of this would be to have an endless impenetrable boundary in the direction of sinking by combining the theoretical solutions of two spheres moving head to head. The last version can be extended to flow within an infinite layer with rigid and flat boundaries, both perpendicular to gravity. This requires combining the flow field from, in theory, an infinite number of spheres, but numerically a limited number is sufficient. These versions outlined here could, for example, be used address interactions between buoyant drops.

5.7.3 Numerical considerations

The code lends itself very well to parallelization, which would make current simulations to be over in hours rather than days. At the core of the calculations are two nested for-loops, one loops over the Stokeslets and the other over the nodes that define the surfaces. The order in which they are performed does not matter and each of the loops can be arbitrarily divided up, to fit the number of processors, as the steps are completely independent. In the end the contributions from all Stokeslets to each node must be summed and the flow advanced. The only other significant calculation step, i.e. finding the combination of Stokeslets that fits the surfaces (Figure 5-1c-e), is also easily parallelized.

As can be inferred from eq. A.10 and further confirmed with the two dimensional finite element code, the flow field depends less and less on the details of the geometry with increasing distance from the sinking object; from faraway, only the total density excess and

its approximate location are important. When a single slab sinks, the details of the slab and how they affect the subduction process at that particular trench are important and densely spaced slablets (as depicted in Figure 5-1) are needed to resolve the minutiae of the process. However, when calculating interactions between slabs that are separated by some distance, it is safe to use larger slablets that represent the slabs in a coarser way. How much larger depends on the distance, but this can be tested with simple numerical experiments. This would make the code more efficient as the number of calculations scales as the product of the number of slablets and nodes that define the surfaces. The best way to implement this is probably to have a nested grid of slablets at the surface, where each large slablet is made up of several smaller slablets. Transferring the load description (Figure 5-1e) then becomes a simple summation. Something similar can be envisioned for the nodes themselves, using summary nodes as representative of a group of nodes, but this is not as straightforward and requires more numerical testing.

Both the number of nodes used to parameterize the surfaces and the number of Stokeslets scales approximately with trench length, assuming a relatively constant width. Currently, the descent of single slab from a 1500 km long trench down to the CMB can be simulated in about a day on a desktop computer. The total length of subduction zones at the surface of the earth is on the order of 40.000 km. To model them all would therefore require $\sim$25 times the number of nodes and Stokeslets and increase the number of days needed for the calculation to $\sim 25^2 = 625$. However, newest CPU's are at least 3 times as fast as the ones currently used. Therefore, a cluster of 10 PC's would give 30-fold increase in computer power, or $\sim$20 days for the simulations. How much the efficiency of the code will be improved by coarser Stokeslets as outlined previously is unclear, but it is almost certain that interactions between slabs that are subducting on different hemispheres will be numerically insignificant. It is therefore safe to assume that the calculation time could be further reduced to a few days or even less.

5.8 Conclusions

We have created a new flow code that accurately and efficiently solves the Stokes equation, using a Green's function approach. This code is versatile and can be used to examine many aspects of buoyancy driven flow in the mantle. The relative low computing cost allows for repeated simulations with systematic changes in the input parameters to map the effects on the subduction process. For example, the genaral behavior of subducting slabs can be mapped by systematically changing the viscosity profile, trench geometry or trench migration rate. More specific problems can also be addressed. The most promising is bridging the gap between plate reconstructions, which provide time dependent constraints, and seismic tomography which gives the three dimensional morphology of subducted slabs. Efficiency of the code will be improved and, as outlined, the ability to impose plate motions will be added. Simple estimates of calculation time indicate that when using a small cluster of PC's, the whole subduction history for the last 100 million years or so, can be simulated in a three-dimensional flow model of the mantle, while using the correct boundary conditions, in a matter of days.

Chapter 6

Sinking rates and thickening of subducting slabs

Abstract

The sinking velocity of a dense subducting slab decreases when the viscosity of the mantle increases. Assuming an approximately constant rate of subduction at the surface the slab will thicken in a more viscous lower mantle if mass flux is preserved. Initially, the slab will thicken above the higher viscosity layer and slow down. Gradually, more mass will accumulate, the slab will thicken and the sinking velocity of the deepest part will increase until a balance between the mass flux in the upper and lower mantle is reached. This balance of thickening and sinking rate and their transient nature is the subject of this study. We compare theoretical solutions for simple geometries to simulations of subduction made in spherical geometry and three dimensions using a new flow code. We show that a power-law relation best describes the thickening and slowing as function of viscosity contrast between the lower mantle. Relations often quoted in the literature seem to have weak foundations. However, due to geometry and the transient nature of the slowing, it is doubtful that an observed slowing factor, on a global scale, has a simple and predictable relation to the viscosity contrasts between the upper and lower mantle. Three dimensional flow modeling of subduction history directly compared to high resolution tomography models of mantle structure gives a more direct way of assessing the viscosity profile of the mantle.

6.1 Introduction

The upper thermal boundary layer of the convective mantle is a rigid, but broken shell, made up of plates of different sizes. These plates move with respect to each other with most of the translation or rotation being accommodated at the plate boundaries, i.e. with creation of new sea-floor at ridges where plates diverge, with strike-slip faults where the plates move past each other and with subduction of cold and negatively buoyant oceanic lithosphere (i.e. slabs) at convergent margins [*Davies and Richards*, 1992]. The geometry and mechanical strength of the plates may largely control the mode of convection within the mantle [*Conrad and Hager*, 1999; *Anderson*, 2001]. Furthermore, the negative buoyancy of the cold downwellings is the major driving force behind plate motion and mantle flow in general [*Hager and O'Connell*, 1981], which is manifested in the rapid motion of plates that are "attached" to a slab vs. other plates [*Forsyth and Uyeda*, 1975; *Gripp and Gordon*, 1990]. The rate at which plates move at the surface is therefore directly linked to the sinking rates of slabs and mantle flow in general.

As is evident from studies of glacial rebound [e.g *Walcott*, 1973; *Hager and Clayton*, 1989; *Simons and Hager*, 1997; *Forte and Mitrovica*, 2001] and by comparison of the observed and model inferred geoid [e.g. *Hager et al.*, 1985; *Moresi and Gurnis*, 1995] the viscosity increases dramatically with depth. A jump in viscosity is most commonly thought to coincide with the $\sim$660 km deep seismic discontinuity, at the boundary between the upper and lower mantle [*Hager*, 1984]. Depth dependent viscosity will affect the sinking rates and the shape of the subducting slabs [*Gurnis and Hager*, 1988]. Higher viscosity makes the flow more sluggish and, to maintain a constant mass-flux through high and low viscosity regions, the slab has to thicken.

The general characteristics of cold downwellings with respect to varying viscosity contrast between the upper and lower mantle has been the subject of many papers [e.g. *Gurnis and Hager*, 1988; *Tao and O'Connell*, 1992; *Moresi and Gurnis*, 1995; *Tao and O'Connell*,

1993; *Gaherty and Hager*, 1994]. However, few of these papers have addressed the sinking rate and thickening of slabs in a systematic manner.

The most comprehensive study to date linking subduction history to seismic tomography on a global scale has been presented in a string of papers by *Ricard et al.* [1993]; *Lithgow-Bertelloni and Richards* [1998]; *Richards et al.* [2000]. In these studies, simplified kinematics have been used to bridge the gap between the last 120 million years of subduction history and tomographic models; slabs are represented by discrete units (slablets) that are assumed to sink directly down from migrating trenches. The sinking rate in the upper mantle is kept the same as the inferred convergence rate of the subducting plate, but slowed down by a constant global factor in the lower mantle. The location of the slablets at the end of the simulation was expanded in terms of low degree spherical harmonics which guaranteed that errors up to $\sim$1000 km in trench location would not affect the results [*Lithgow-Bertelloni and Richards*, 1998]. Once expanded, the mass distribution was compared to tomographic models and the associated geoid to the observed geoid. The best fitting ratio of upper to lower mantle sinking rates (i.e. *slowing factor*) was then found by maximizing the correlation between the different spherical harmonics coefficients. A slowing factor of $\sim$4 best fit all the observations. In these papers, *Gurnis and Davies* [1986] and *Richards* [1991] are referenced as a source for a rule of thumb relating this slowing factor, u^*, to a viscosity contrast, μ^*, in a layered convection model, namely $u^* = ln\mu^*$. They then deduce that their results are consistent with a 40-fold increase in viscosity between the upper and lower mantle ($ln\ 40 \approx 3.69$).

In this chapter we address the slowing and the associated thickening of slabs. We first derive theoretical relationships assuming a constant mass-flux for a sphere and an infinite cylinder. Based on those we make a prediction for an infinite plate, which is most similar to a subducting slab. We then use the flow code introduced in Chapter 5 and make simulations for subduction from stationary trenches for different viscosity contrasts between the upper

and lower mantle. We also check the effects of varying the length, width and migration rate of the trench on the sinking rates. We compare our results to the theoretical relationships, to the results of *Gurnis and Davies* [1986] and to the logarithmic law mentioned above.

6.2 Theoretical calculations

Theoretical solutions to Stokes equation, which governs high viscosity flow, are only known in the most simple cases. Here we use two of them, the solution for a sphere and for an infinite cylinder, both sinking in a uniform viscosity, infinite medium. Using these two solutions we derive expressions for the thickening and slowing factor as a function of the viscosity contrast. Inspired by these solutions, we make a prediction for a sinking plate.

The basic assumption behind the calculations is that mass flux is independent of viscosity, i.e. as viscosity increases, the slower sinking rate is counteracted by an increase in size. Therefore, for two fluids of viscosity μ_1 and μ_2, the following must hold:

$$u_1 M_1 = u_2 M_2 \tag{6.1}$$

where u_1 and M_1 are velocity and mass of the sinking object in medium of viscosity μ_1. Assuming the same density of the objects in all cases, this translates to:

$$\begin{aligned} u_1 {a_1}^3 &= u_2 {a_2}^3 && \text{For a sphere} \\ u_1 {a_1}^2 &= u_2 {a_2}^2 && \text{For a cylinder} \\ u_1 a_1 &= u_2 a_2 && \text{For a plate} \end{aligned} \tag{6.2}$$

where a is the radius of the sphere or the cylinder or the thickness of the plate.

The terminal velocity of a sphere in an infinite medium changes with viscosity as $\sim a^2/\mu$ (see appendix A). Using this and eq. 6.2 we find for a sphere:

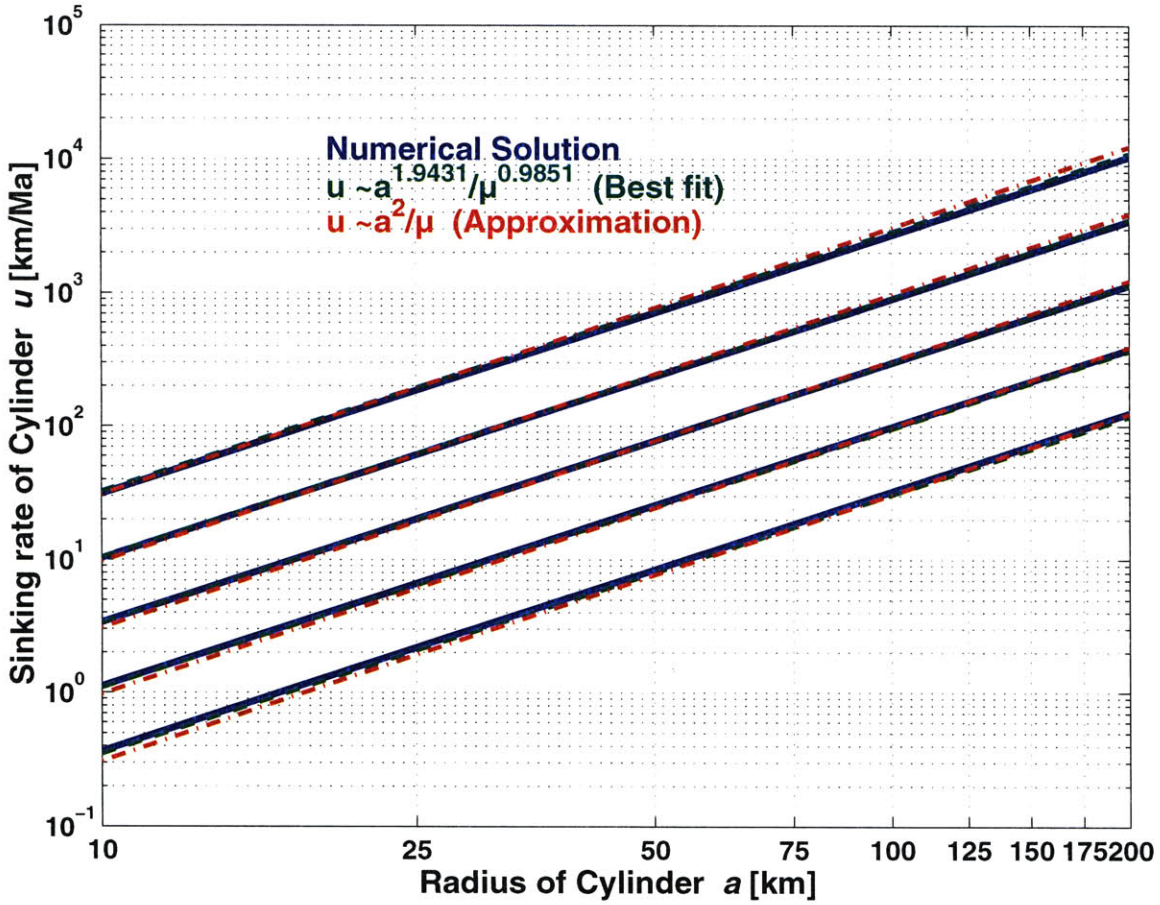


Figure 6-1: Theoretical solution to eq. 6.4 (blue lines) using $\rho_c = 5075\, kg/m^3$, $\rho_f = 5000\, kg/m^3$, $g = 10\, m/s^2$ and $\mu = 10^{21.0},\ 10^{21.5},\ 10^{22.0},\ 10^{22.5},\ 10^{23.0}\ Pa\ s$ (highest to lowest curve). Also shown are curves for the best fit and an approximate power-law relation between radius and sinking rate as indicated.

$$\frac{u_1}{u_2} = \frac{a_2^3}{a_1^3} = \frac{a_1^2/\mu_1}{a_2^2/\mu_2} \quad \Rightarrow \quad \frac{a_2}{a_1} = \left(\frac{\mu_2}{\mu_1}\right)^{\frac{1}{5}} \text{ Thickening}$$

$$\Rightarrow \quad \frac{u_1}{u_2} = \left(\frac{\mu_2}{\mu_1}\right)^{\frac{3}{5}} \text{ Slowing factor} \tag{6.3}$$

The terminal velocity for a cylinder is not as easily attainable since no solution using

the proper Stokes equation exists for the usual boundary conditions (velocity finite at the cylinder and zero at infinity). It turns out that independent of the Reynolds number ignoring inertia, which is the basis of Stokes equation, is not valid for an infinite cylinder. The reason for this *Stokes paradox* is that far enough from the cylinder, viscous and inertia forces are comparable. However, an improvement to Stokes equation that avoids this paradox was made by Oseen, and a solution for the drag on a moving cylinder exists [*Batchelor*, 1967]. Balancing the drag force against the gravitational acceleration gives the following implicit equation for the terminal velocity, u, of a rigid cylinder [*Ristow*, 1997].

$$\frac{u}{\ln\left(3.7\frac{u}{a\rho_f\mu}\right)} = \frac{1}{4}\frac{\rho_c - \rho_f}{\mu} g a^2 \tag{6.4}$$

Unlike the solution for a sphere eq. 6.4 is dependent on the absolute density of the fluid, ρ_f, not only the excess density of the cylinder, $\rho_c - \rho_f$, i.e. inertia matters. However, as shown in Figure 6-1, a numerical solution using mantle-like parameters is very close to have the same form as for a sphere, $u \sim a^{1\,9431}/\mu^{0.9851} \approx a^2/\mu$ (For other sets of parameters numerical solutions of eq. 6.4 can behave very differently). Using this approximation and following the same steps as in eq. 6.3 and using eq. 6.2 gives for a cylinder:

$$\begin{aligned} \frac{a_2}{a_1} &= \left(\frac{\mu_2}{\mu_1}\right)^{\frac{1}{4}} \quad \text{Thickening} \\ \frac{u_1}{u_2} &= \left(\frac{\mu_2}{\mu_1}\right)^{\frac{1}{2}} \quad \text{Slowing factor} \end{aligned} \tag{6.5}$$

Since the terminal velocity of a sphere and a cylinder seem to scale almost identically with viscosity and characteristic length, it is tempting to assume that the same relation is approximately valid for a plate. We do this here to compare with results later obtained by simulations; assuming that terminal velocity scales once again as $\sim a^2/\mu$, we get for a

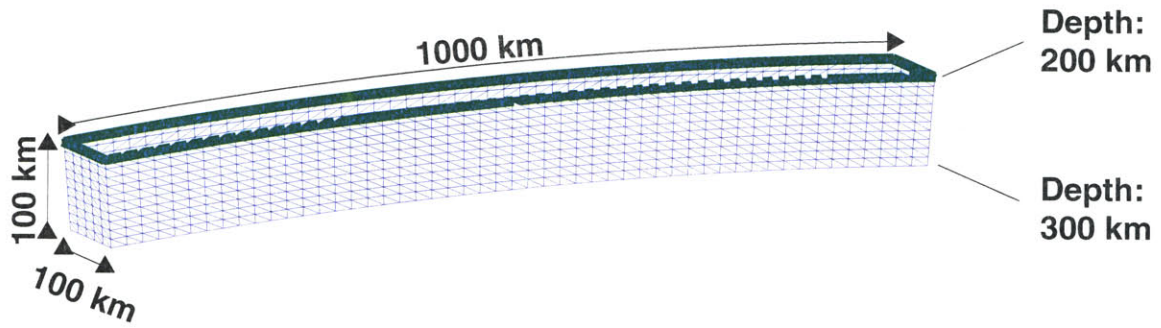


Figure 6-2: The meshed surface at time 0 before simulation starts. The green edge is anchored and stays at 200 km depth throughout the simulation but moves laterally in runs with a migrating trench.

plate:

$$\begin{aligned} \frac{a_2}{a_1} &= \left(\frac{\mu_2}{\mu_1}\right)^{\frac{1}{3}} \quad \text{Thickening} \\ \frac{u_1}{u_2} &= \left(\frac{\mu_2}{\mu_1}\right)^{\frac{1}{3}} \quad \text{Slowing factor} \end{aligned} \tag{6.6}$$

6.3 Simulations

To investigate the sinking rates and the velocity profiles of slabs, subducting within the Earth's mantle we use the new three dimensional spherical flow code described in Chapter 5. The initial setup at the beginning of the simulations is in most cases as shown in Figure 6-2, i.e. unless otherwise noted, the trench is 1000 km long and 100 km wide. The topmost line of the mesh is anchored and does not move, except laterally in runs with a migrating trench. The fluid inside the meshed surface and directly above its opening is buoyant and drives the flow. Material flows in through the top and the mesh expands. Time is given using the same assumptions as in Chapter 5.

In Figures 6-3 and 6-4, we show a typical result, a run with a 30-fold increase in vis-

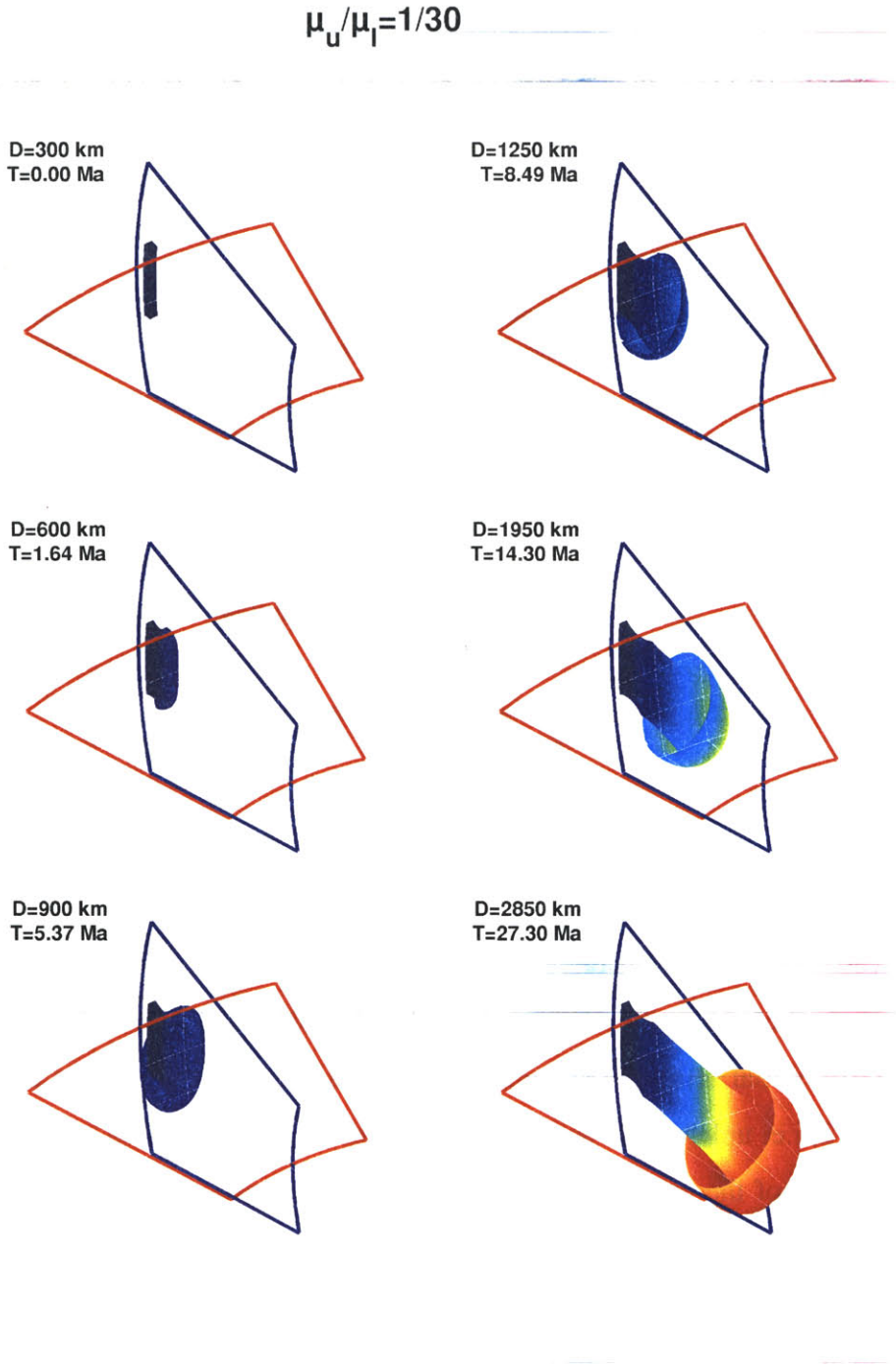


Figure 6-3: Views of simulations with 30-fold increase in viscosity in the lower mantle at selected depths of the tip of the slab.

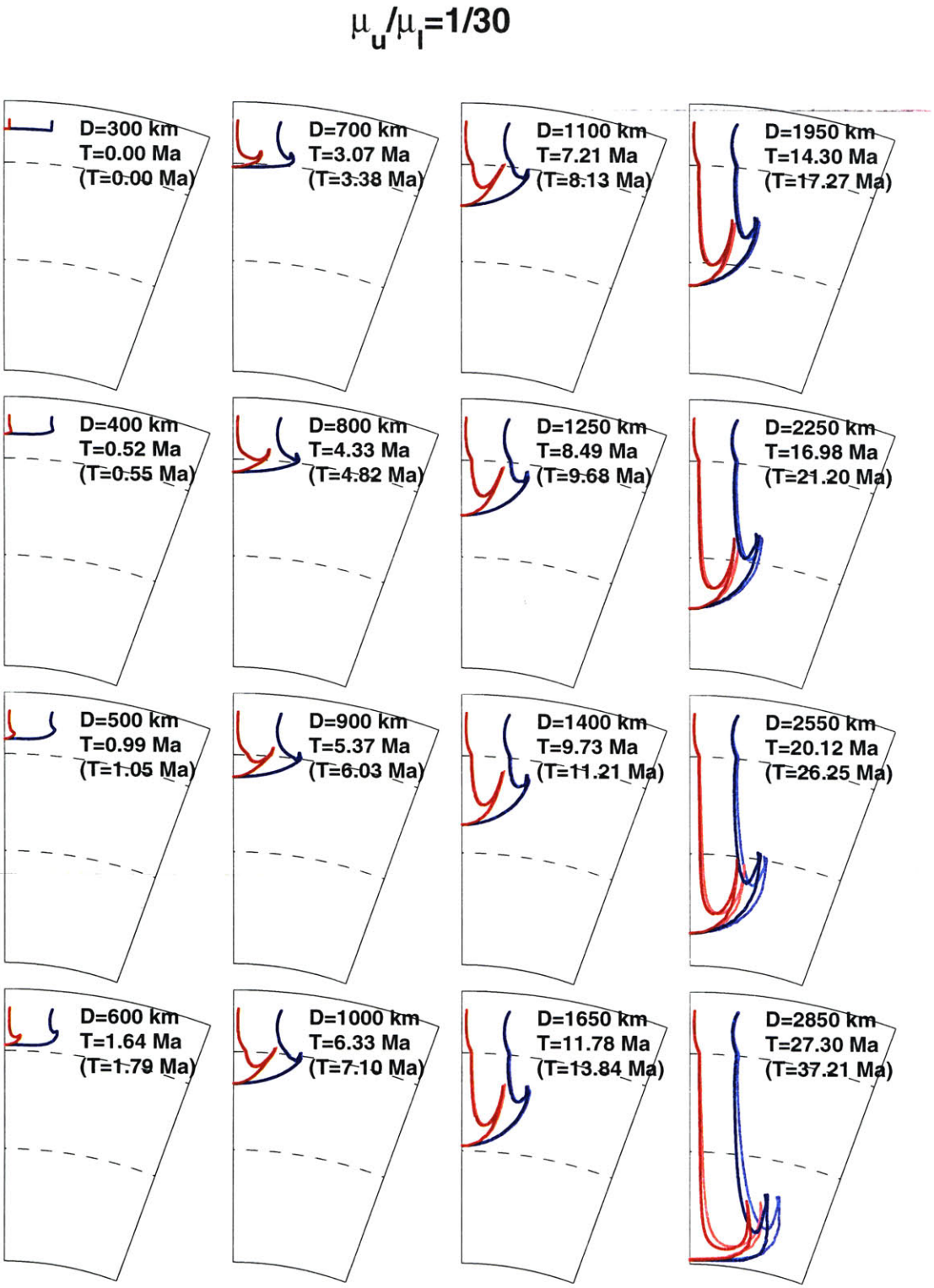


Figure 6-4: Cross sections at selected depths to the the leading tip of the slab along (blue lines) and perpendicular (red lines) to the trench for a simulation with 30-fold increase in viscosity in the lower mantle. Also shown are results from runs with reduced thermal expansion with depth (light blue and light red lines) as described in the text.

cosity in the lower mantle. The center of the slab starts sinking at the fastest rate. As the deepest part of the slab approaches 660 km depth where the viscosity jumps, the flow becomes more horizontal as material escapes the descending slab by flowing to the sides and up, creating a sort of a slab-head, analogous to heads of rising plumes, on the slab itself. The slab thickens and slows down in the lower mantle and the head grows. Deeper, the slab thickens as it feels more resistance from the lower boundary and the simulation is stopped when the deepest part of the slab is within ~50 km of the core-mantle boundary.

Theoretical thermodynamics suggests that the thermal expansion coefficient, α, decreases with depth and has only about 30% of its surface value close to the core-mantle boundary [*Anderson*, 1989]. Also shown in Figure 6-4 is a run mimicking the effects of reduced thermal coefficient. It has the same viscosity contrast as before but the excess density decreases linearly with depth, resulting in slower sinking and a thicker slab in the lower mantle. Here, we probably overestimate the effects of decreased thermal expansion coefficient as it is counteracted by increased absolute density with depth.

6.4 Results

6.4.1 Thickening

The thickening observed in Figures 6-3 and 6-4 is further analyzed in Figures 6-5 and 6-6 for reduced thermal expansion coefficient with depth. The thickness is measured at a few selected depths as a function of time for simulations with different viscosity contrasts. The height of each gray panel is proportional to the time it takes the slab to sink to 2850 km and the width is indicative of the thickness of the slabs in each case. As is to be expected from the theoretical relations in Chapter 6.2, the slabs slow down and thicken with increasing viscosity. The solid and broken lines show the minimum and the maximum thickness, respectively. The difference between the solid and the broken line is therefore indicative

of the size of the tails of the slab-head and how long it takes the entire head to go past a particular depth.

The green profiles that are measured at 410 km depth show the least variability, both with time and between simulations. This is not surprising since the upper mantle viscosity is the same in all the runs and the thickness is largely regulated by the initial conditions, which are shown in Figure 6-2 and do not change through time. However, due to increased resistance to downward flow, the upper mantle slab-head grows with a more viscous lower mantle.

Only 250 km lower at 660 km depth (blue lines), exactly at the viscosity jump,the thickness is much more sensitive to the lower mantle viscosity and, for the highest viscosities, almost all of the thickening occurs by that depth.

The inset plots in Figures 6-5 and 6-6 show the thickness at the mentioned depths when the tip of the slab has reached 2850 km depth. In all cases there is a thinning of the slab in the upper mantle, relative to its fixed thickness of 100 km at 200 km depth. This thinning reduces with increased viscosity contrast, i.e. at 410 km depth the thickness is affected by the thickening in the lower mantle. When the logarithm of slab thickness is plotted vs. the logarithm of viscosity contrast, the measurements seem to lie approximately on a line, indicating a power-law relationship. This is true both for simulations with and without reduced thermal expansion coefficient with depth.

6.4.2 Sinking rate

The time which has elapsed from the start of the simulation and the velocity of the leading tip of the slab are shown in Figure 6-7 as function of depth. Simulations with reduced thermal expansion are shown with dashed lines and for reference the analytical solution for a sinking sphere is indicated. As can be seen from the inset in Figure 6-7a, the time it takes the slab with a 100-fold increase in viscosity in the lower mantle to sink to 660km depth is

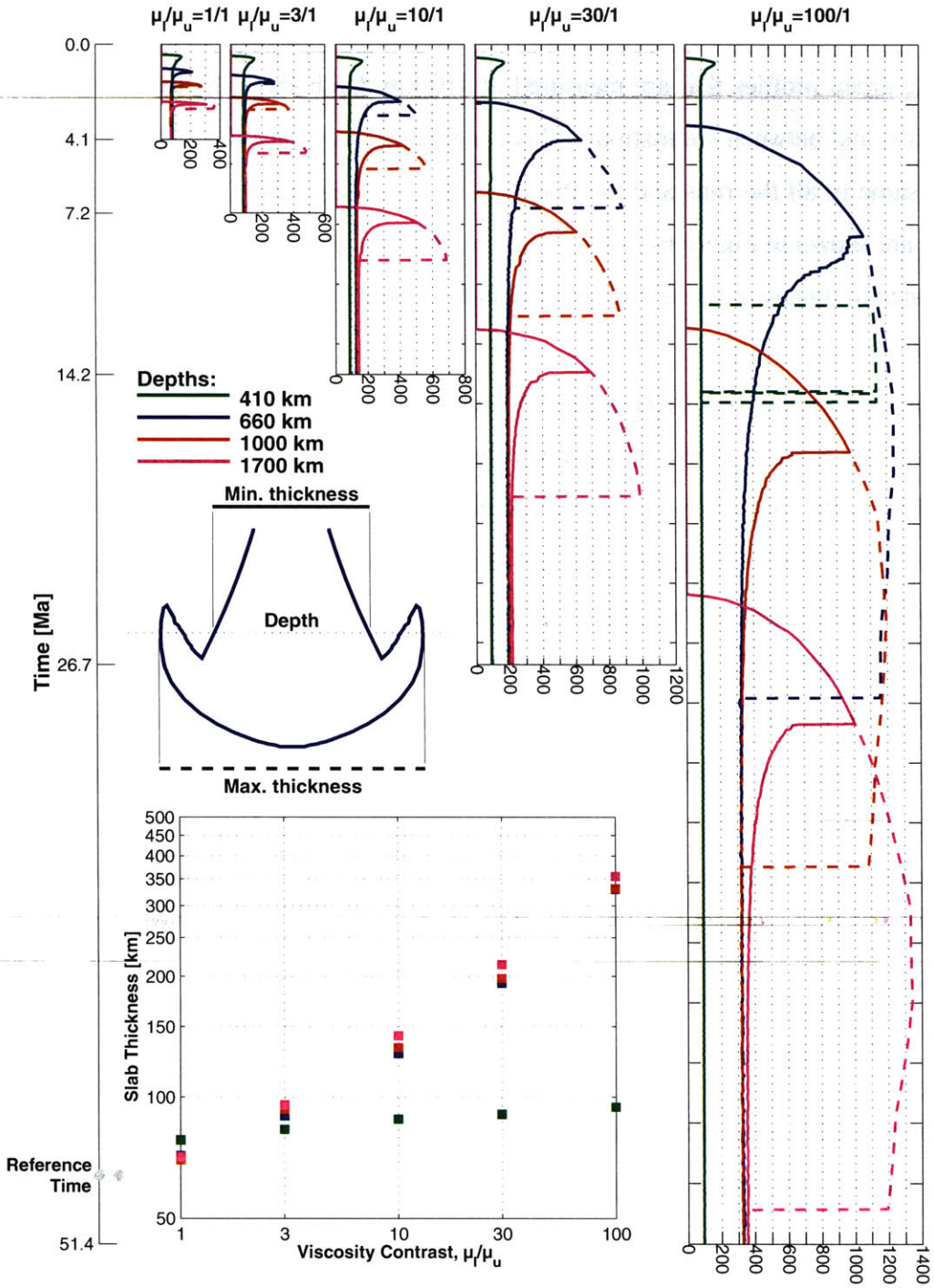


Figure 6-5: Thickness measured at different depths as function of time during simulation for different viscosity contrasts as indicated. Solid lines show the minimum thickness and the broken lines the maximum thickness. The inset plot shows the thickness at the different depths when the tip of the slab has reached 2850 km depth. The reference time for a sinking sphere as described in Chapter 5.5 is also indicated.

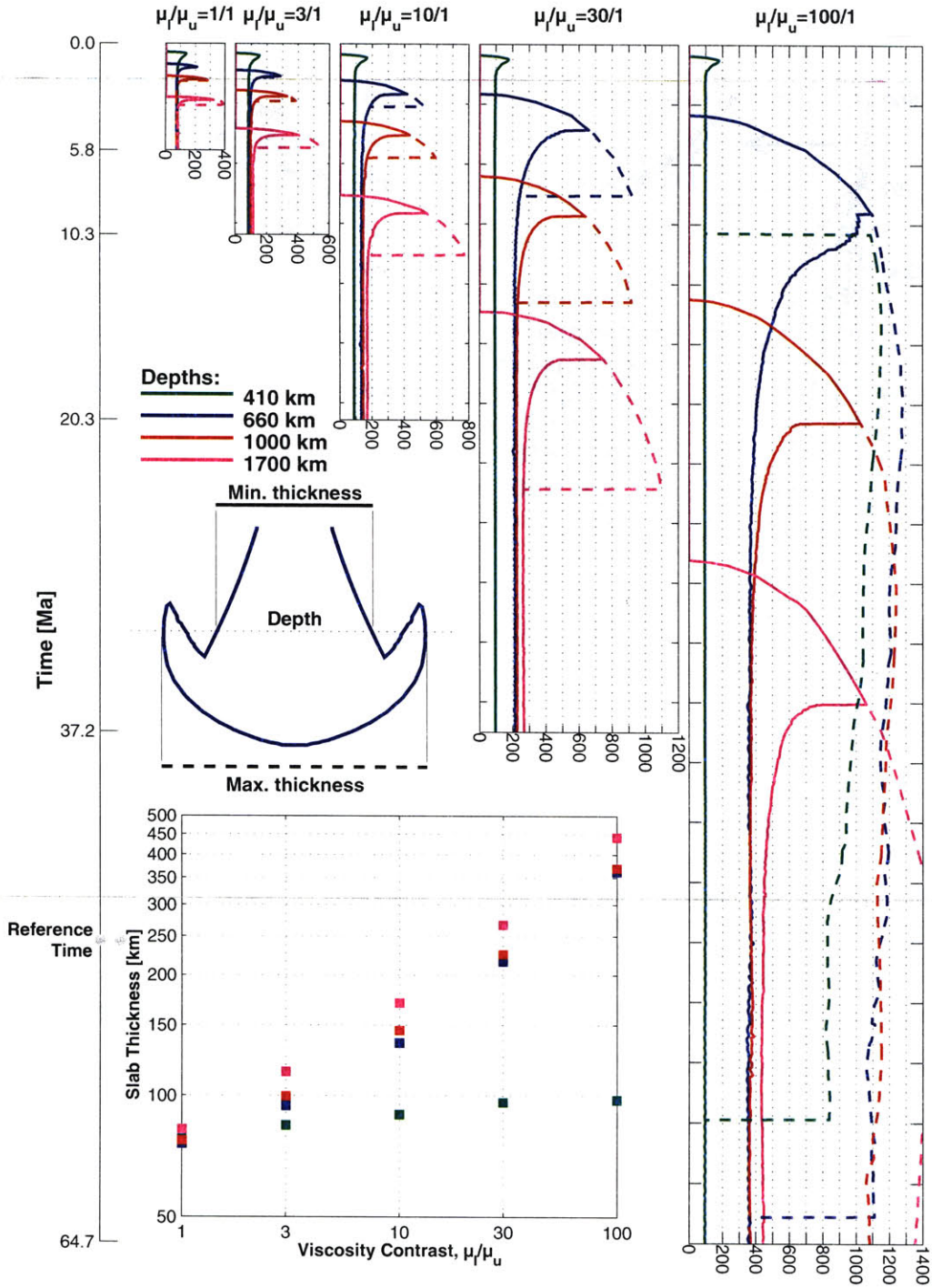


Figure 6-6: Same as Figure 6-5 except for runs with reduced thermal gradient.

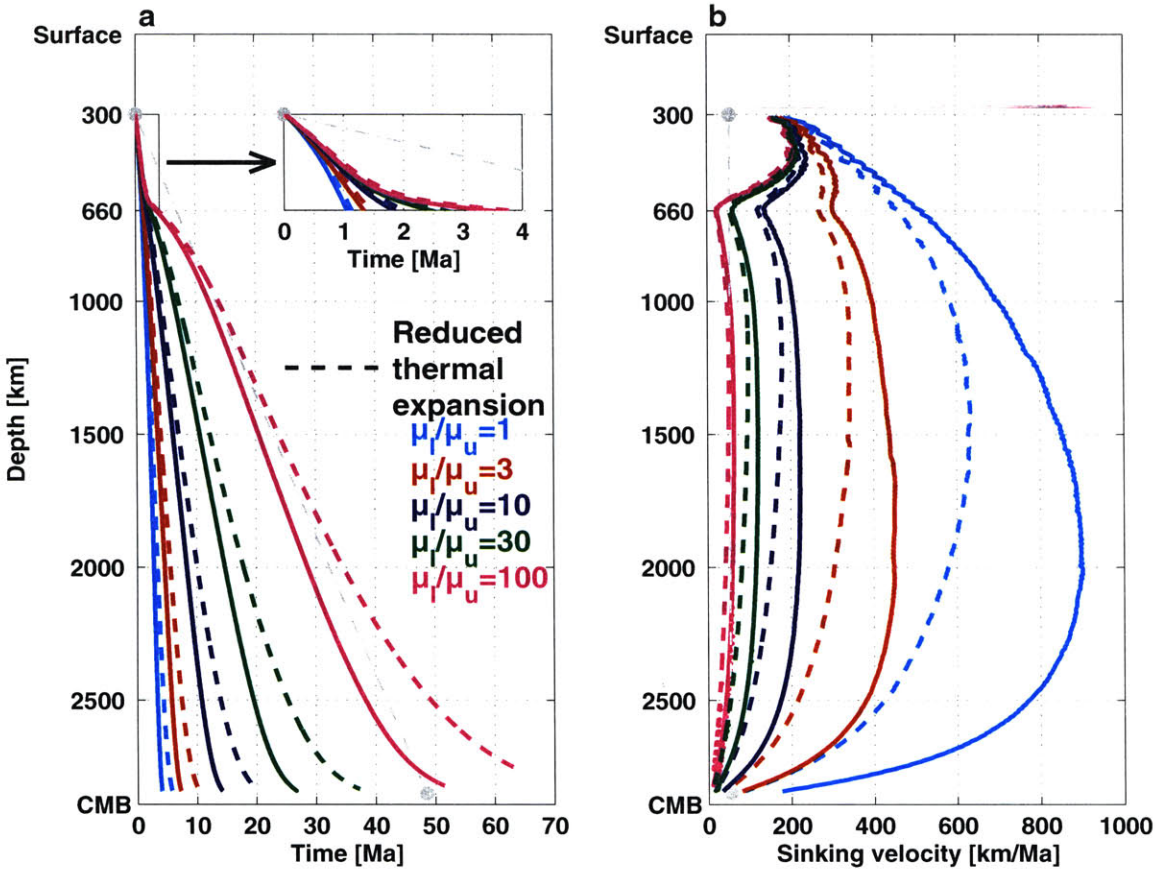


Figure 6-7: (a) The position of deepest point of slab for runs with (solid lines) and without (broken lines) reduced thermal gradient as function of time for different simulations as indicated. (b) The sinking rate of tip of slab as function of position for different runs. Theoretical solutions for a sphere sinking in an infinite medium are indicated with gray lines.

quadrupled compared to a uniform viscosity mantle. It takes the slab in a uniform viscosity mantle only 4 Ma to get within 50 km of the core-mantle boundary vs. over 50 Ma and over 60 Ma for slabs without and with reduced thermal exansion coefficient, respectively, in a mantle with viscosity contrast of 100.

For a uniform viscosity and a viscosity contrast of 3 the maximum velocity occurs in the lower mantle. With increased viscosity the lower mantle sinking rates are reduced but so is the velocity in the upper mantle. Sinking rates have a maximum in the lower mantle

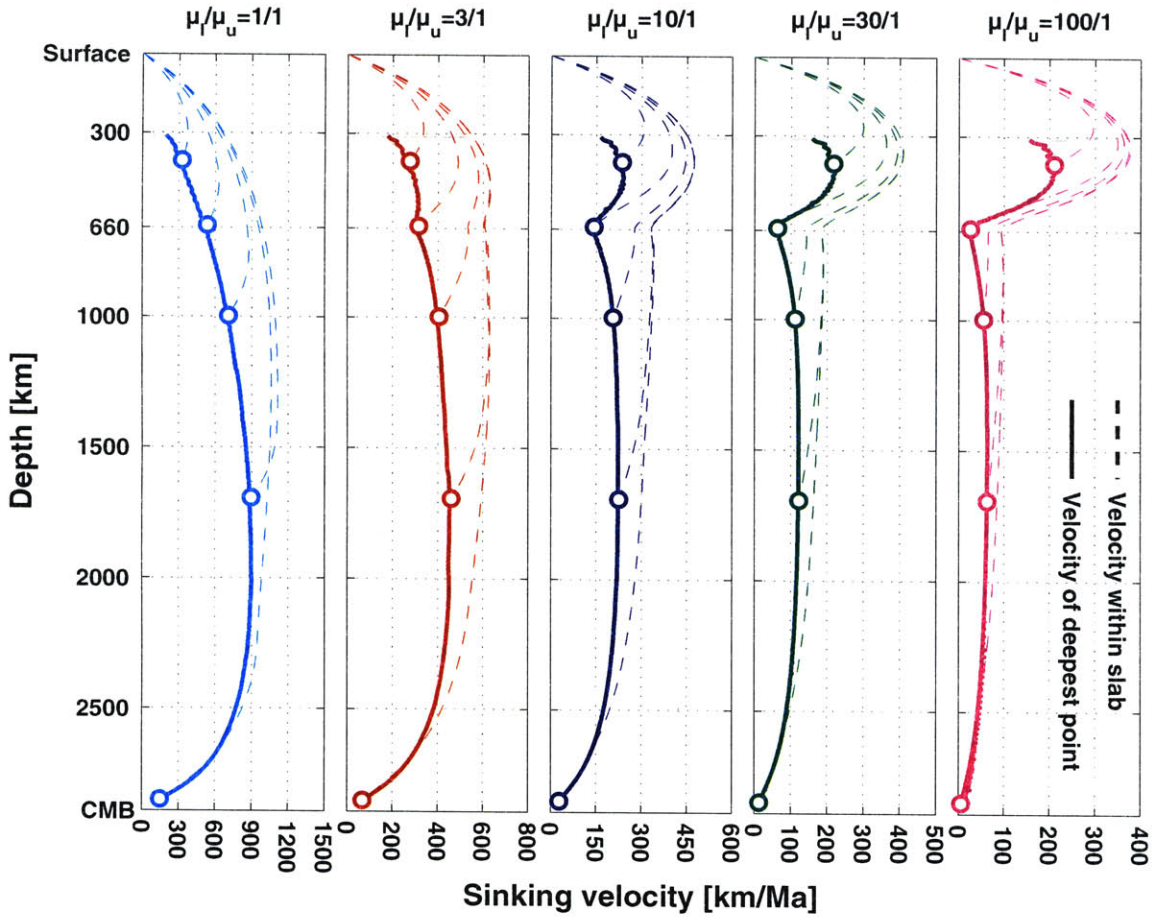


Figure 6-8: Solid lines show the sinking rates of the tip of slabs in different simulations as shown in Figure 6-7 for different simulations as indicated. The broken lines show the downward velocity in the center of the slab as function of depth at the instant the tip of the slab is at the depth indicated by the circles.

at around 1500 km to 2000 km depth for all runs, except when thermal expansion reduces with depth, where the maximum occurs at shallower depths.

As the tip of the slab is sinking, the whole slab is also moving downward. As shown in Figure 6-9, the velocity in the center of the slab is not constant, neither with respect to position nor time. The hollow circles are points common to the dashed lines that end in them and the solid line. They show the profile of vertical velocity within the slab, at its center, as a function of depth, at the moment the tip of the slab is at the hollow circle. As the tip sinks

deeper, the velocity within the slab constantly increases while the thickness stays approximately constant according to Figures 6-5 and 6-6. This means that the mass flux between the upper and lower mantle increases. However, as the tip sinks beyond ~2000 km, the upper mantle velocities stay approximately constant. This is probably not due to a decoupling between the upper and the lowest part of the mantle, rather the flow is being impeded by the lower boundary, barring the mass flux from the upper mantle fom increasing more. As mentioned earlier, plates that are attached to slabs move faster at the surface [*Forsyth and Uyeda*, 1975]. Assuming that plate velocity is directly linked to sinking rates somewhere in the range 100 km-300 km Figure 6-9 indicates that it should not change dramatically as the subduction progresses, especially for the higher viscosities. However, not modeled here is the viscous resistance of the rigid plate, which will be relatively more effective for a shallow slab, and therefore slow the flow at the start of the subduction process.

In Figure 6-9 the slowing factor, i.e. the rate of sinking in the upper mantle vs. the rate of sinking in the lower mantle, is quantified. For uniform viscosity and for viscosity contrast of 3, the slowing factor is actually less than one in most cases, i.e. the slabs speed up in the lower mantle. Another observation is that the slowing factor changes depending upon how it is measured. There is not a unique way of measuring it, but the green circles are calculated using the average velocity of the tip of the slab, which shows the least slowing. The other measurements are obtained using the inner velocity as the tip is at a certain position as shown in Figure 6-8. Measuring the slowing this way generally yields higher values than using the velocity of the tip but of those the measurements at 1700 km depth are the lowest as they give the maximum lower mantle sinking rate. However, when measured consistently between simulations the slowing factor seems to follow approximately a power law.

The measurements center approximately around 0.7, 1, 1.5, 2 and 3.5 for viscosity contrast of 1, 3, 10, 30 and 100, respectively. *Lithgow-Bertelloni and Richards* [1998]

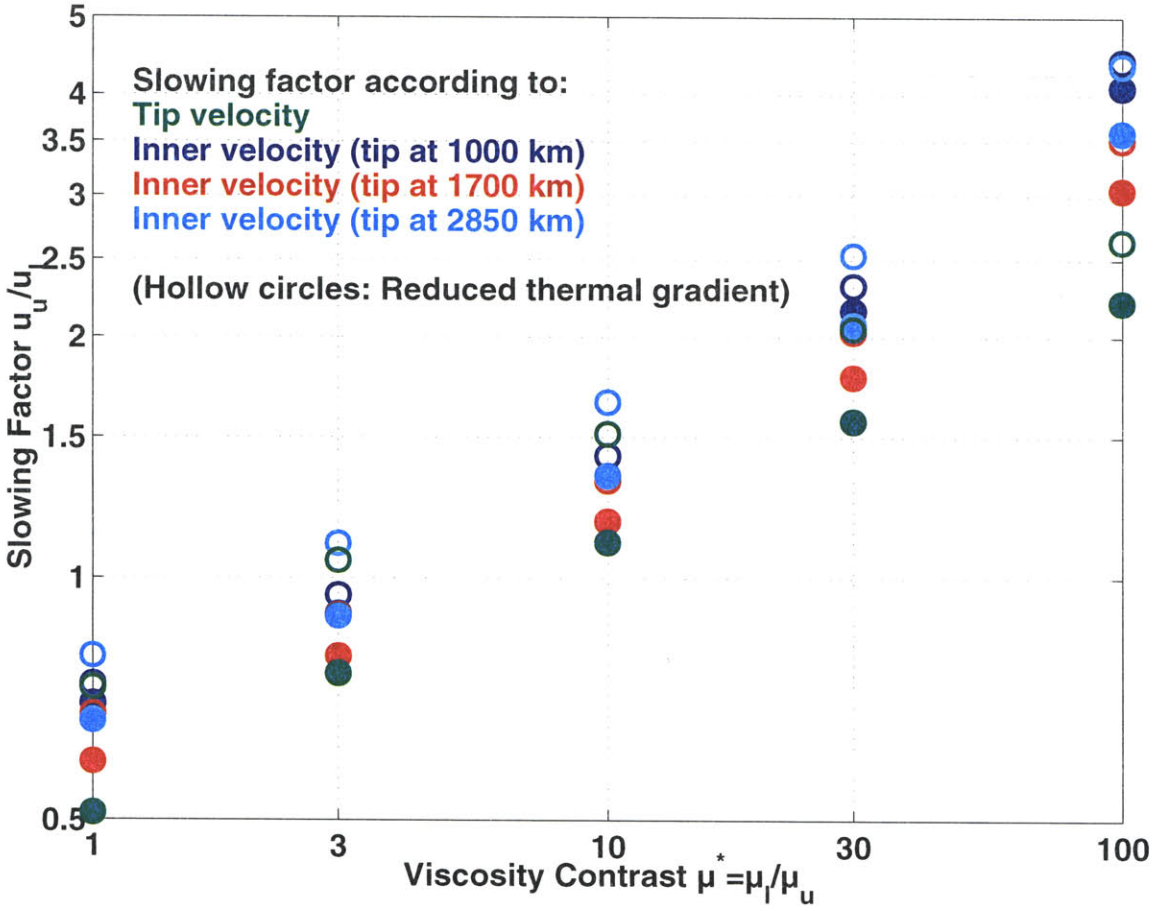


Figure 6-9: Slowing factor, i.e. the ratio of average lower mantle sinking rates to average upper mantle sinking rates, as function of viscosity contrast. Measurement are made using different velocity profiles (see text). Hollow symbols are for runs with decreased thermal expansion with depth.

found a slowing factor of $\sim$4 and interpreted that to be consistent with a viscosity contrast of 40, while the values shown in Figure 6-9 would indicate much higher value ($\geq$ 100).

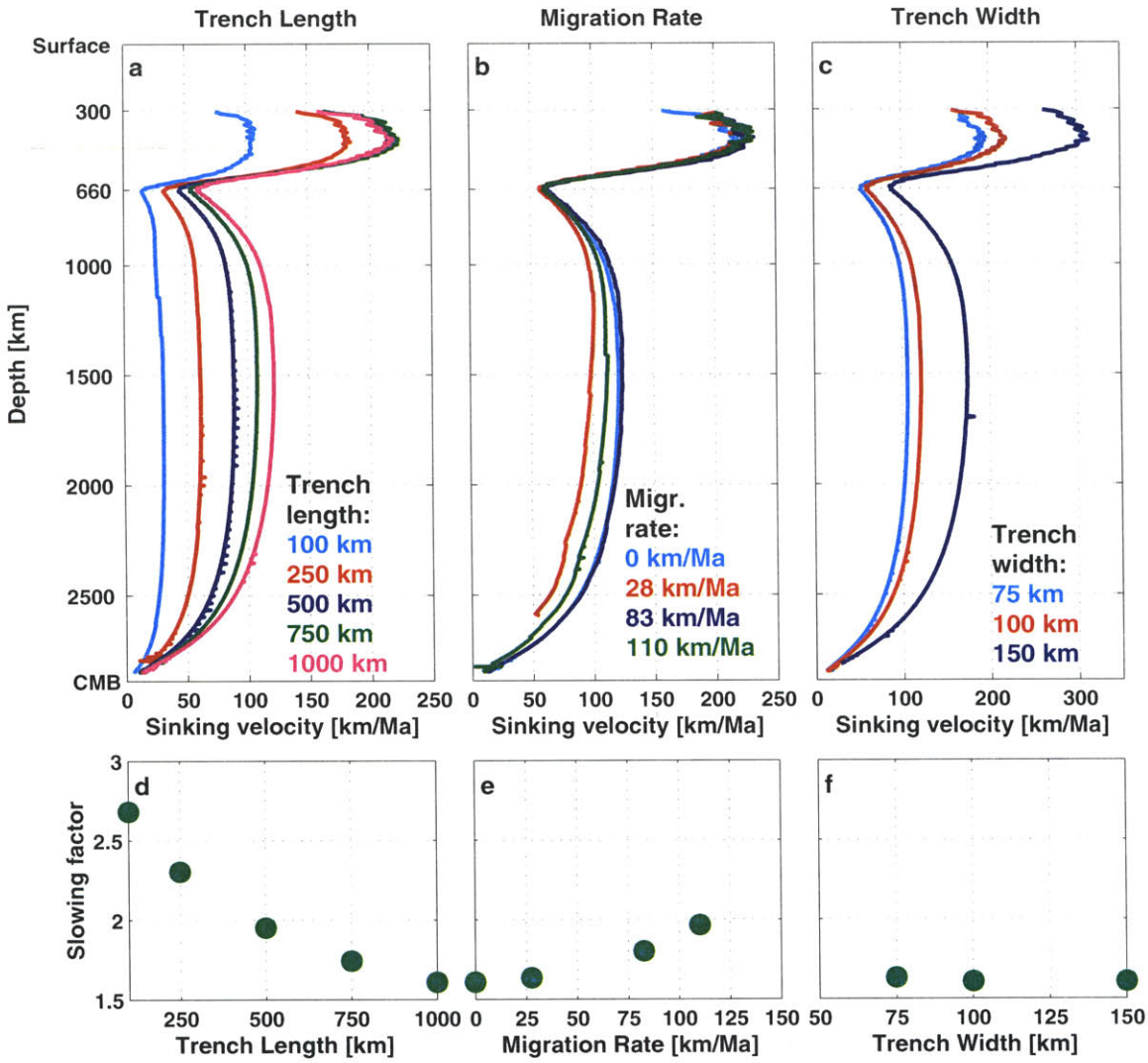


Figure 6-10: (a) The position of deepest point of slab for runs with varying trench lengths as indicated. (b) The sinking rate of tip of slab as function of position.

6.4.3 Geometrical effects

To investigate how robust the values for slowing factors displayed in Figure 6-9 we conducted three experiments where we check how conditions at the surface might influence the relative sinking rates in the upper and lower mantle.

In the previous simulations, the trench length was kept constant at 1000 km. In Figure 6-

10a we investigate how changing this length affects the sinking rates; all these simulations were run with a 30-fold increase in viscosity in the lower mantle. It is not surprising that that velocity in general decreases as the trench is shortened; reducing a characteristic dimension generally slows the flow, as is indicated by the theoretical relations in Chapter 6.2. In Figure 6-10a there is, however, an exception to this rule in the case of the 750 km and 1000 km long trenches, where the slab subducting at the shorter trench actually sinks faster in the upper mantle. This is due to the fact that it is relatively easier for the shorter trenches to push material away in the direction along the trench. There is an optimum length somewhere between 500 km and 1000 km where the sinking rate in the upper mantle is the fastest. When the slabs enter the lower mantle and the boundaries are much farther away, the usual trends are observed.

It is a consequence of upper and lower mantle sinking rates behaving differently that the slowing factor is sensitive to trench length, as is shown in Figure 6-10d. The slowing factor is higher for a short trench, more than 3 for 100 km length vs. $\sim$1.6 for 1000 km length.

For a migrating trench, the upper mantle velocities stay approximately constant, as is seen in Figure 6-10b. Upon entering the lower mantle, the migrating slab slows down, i.e. more slowing with faster migration rate, as is confirmed in Figure 6-10e. This is because, with increased migration rate, the slab is sinking more and more like a flat sheet which is less efficient than a vertical flow.

If the width of the trench is changed, as shown Figure 6-10c, the whole velocity profile is scaled. This scaling is approximately constant over the whole mantle, i.e. the slowing factor is relatively insensitive to changes in trench width, as shown in Figure 6-10c. This means that changes in width trade off with changes in viscosity or density, e.g. making the trench wider is almost equivalent to increasing the effective density of the buoyant material. The implication for geodynamical modeling is that errors in the absolute value

for the thickness of subducting lithosphere are not critical for the shape of slabs at depth.

6.5 Discussion

The theoretical relationships derived in Chapter 6.2 for the slowing factor all have the form $u_u/u_l = \alpha\,(\mu_l/\mu_u)^{\beta}$ with $\alpha = 1$. However, as the measurements in Figure 6-9 show, and as is to be expected within the finite mantle, the slowing factor is not unity for a uniform viscosity mantle. Furthermore, changes with viscosity seem otherwise consistent with a power-law relation, i.e. constant β but a varying α. In Figure 6-11 we normalize the measurements with the slowing factor in a uniform mantle, in effect eliminating α, which facilitates a direct comparison of the β for the different theoretical models and the measurements.

The plate model fits the data from the simulations very well and it is clear that $\beta = 1/3$ describes the relations between slowing factor and viscosity contrasts robustly, i.e independent of how this factor is measured or if the thermal gradient reduces with depth.

In Figure 6-12 we further analyze the thickening of slabs as obtained in Figures 6-5 and 6-6. The thickening is normalized with respect to the thickness at each depth in the case of a uniform viscosity mantle. This facilitates a direct comparison with the theoretical relations for thickening derived earlier. Another possibility would be to normalize by the thickness at 410 km depth within each run, but since that thickening takes place in part above the 660 km depth, it would skew the results.

When normalized in this way, there is little difference between the runs with or without reduced thermal expansion with depth. For the higher viscosities, the plate model gives an almost perfect fit. For a viscosity contrast of 10, the measurements could fit either the plate or the cylinder model, and for a contrast of 3, they are all possible. These results are therefore most consistent with a plate-like behavior, i.e the thickening grows as the cube root of the viscosity contrast.

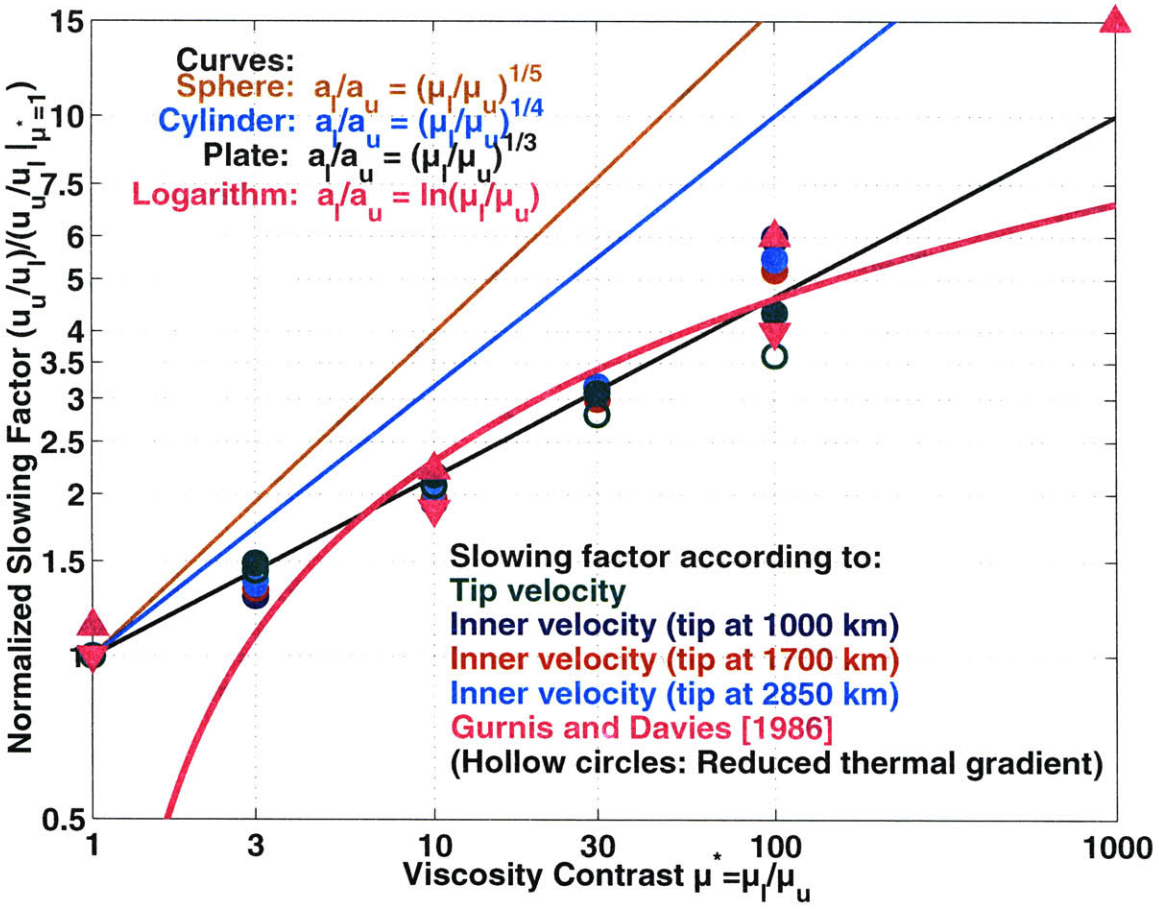


Figure 6-11: Normalized slowing factor measured as in Figure 6-9. Also shown are results from *Gurnis and Davies* [1986] (pink triangles). Also shown are theoretical solution as indicated.

In Figure 6-11 we also show results from *Gurnis and Davies* [1986], adopted from their Figure 8. They performed numerical simulations of convection in a two dimensional box with logarithmically increasing viscosity with depth. Velocity ratios are only shown in this paper on a plot of the ratio between average *horizontal* velocities at the top and and bottom of the box for four different viscosity profiles (uniform viscosity and 10-, 100- and 1000-fold increase). This paper, along with *Richards* [1991], which only quotes *Gurnis and Davies* [1986], are the referenced sources of a logarithmic relationship, $u_u/u_l = ln\mu^*$,

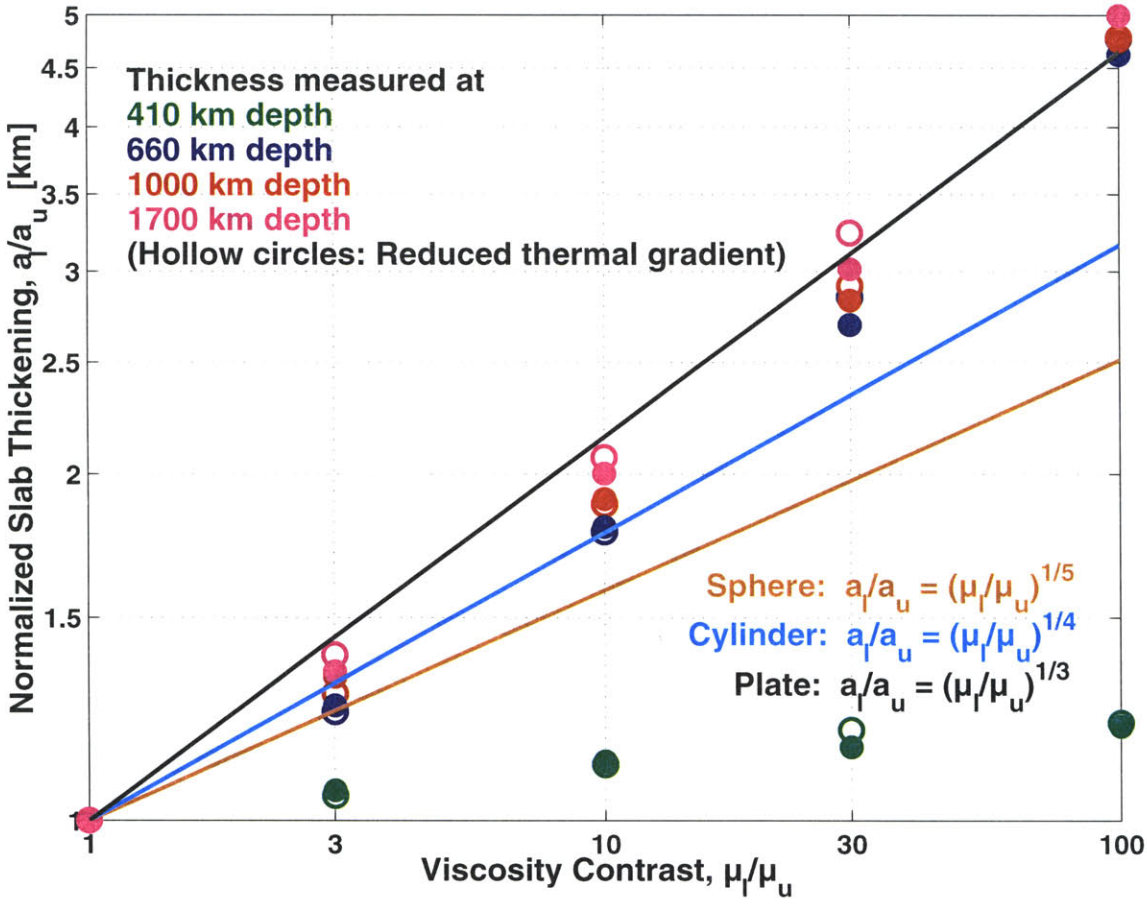


Figure 6-12: Normalized slab thickening measured at different depths as function of viscosity contrast. Also shown are curves for theoretical solutions.

also shown in Figure 6-11 [*Ricard et al.*, 1993; *Lithgow-Bertelloni and Richards*, 1998; *Richards et al.*, 2000]. This logarithmic relation is a decent fit to their two central measurements, although it is not pointed out in their paper. Surprisingly, a plate like relationship, with $\beta = 1/3$, is a very good fit to all of those measurements. The logarithmic relationship seems to have little theoretical basis, e.g. it will obviously never be valid for low viscosity contrasts since it gives 0 for uniform viscosity. It gives similar values as the plate model for viscosity contrasts between 10 and 100, but in neither case the geometrical effects (i.e.

α) are taken into account.

6.6 Conclusions

We have shown that, when measured consistently, the sinking factor and thickening of subducting slabs change systematically with viscosity contrast between the upper and lower mantle. The results are robust and agree with the analytical laws derived for slowing and thickening for an infinite plate. More precisely, both variables scale approximately as the cube-root of this contrast. In addition, this power-law relationship is scaled by a factor which is a strong function of geometry and trench migration rate. It is, however, relatively insensitive to trench width. Furthermore, two dimensional flow models cannot account for the geometrical effects, and the sometimes quoted logarithmic relationship between slowing and viscosity contrast seems to have little footing.

Relating subduction history through simple kinematics to seismic tomography and the geoid constrains, in a limited way, certain aspects of mantle flow. That way, the average sinking velocities in the upper and lower mantle can be estimated, structures observed in tomographic models can be related to past subduction events and different hypotheses for past plate motions can be evaluated. However, since the geometry of subduction zones is complex and time dependent, an observed slowing factor, on a local or global scale cannot be used to constrain the viscosity profile of the mantle. The direct applicability of the results from the simulations to the non-homogeneous and ever evolving mantle is therefore questionable.

The spatial and temporal variability of mantle convection calls for a fully three dimensional flow modeling to connect plate reconstructions and the complex morphology of subducted slabs as revealed by high resolution tomography.

Chapter 7

Integrating seismic tomography, 3-D spherical flow modeling, and plate reconstructions

Abstract

Using a novel geodynamical modeling approach to combine high resolution tomographic images of slabs of subducted lithosphere beneath Indochina with tectonic reconstructions of the Southeast Asia convergent margin, we investigate the development of mantle structure over geological time. Our tomographic model is based on P, pP, pwP, PP, PKP and P_{diff} data and has a variable parameterization which adapts to data density and allows high resolution in the region of interest. The geodynamical technique is based on solving Stokes equation, which governs mantle flow, in an efficient manner using a Green's functions approach. Specifically we investigate the subduction of the Indo-Australia plate beneath the Sunda-arc convergent margin. We model the mantle flow in three dimensions for different depth-dependent viscosity profiles of the mantle. At the surface we adjust the trench geometry to follow the south and southwest motion of the Sunda-block, which is related to the collision of India with mainland Asia. This way we can single out subduction scenarios, constrain the viscosity contrast between the upper and lower mantle and explain apparent discrepancies between tomography and plate reconstructions. Furthermore, juxtaposing the tomography, realistic flow models and inversion tests using the flow models as input, illustrates the biases built into tomographic inversions.

7.1 Introduction

Recent tomographic models show with increased detail the subduction of former oceanic lithosphere and provide evidence for the connection of structures with recent plate motion at Earth's surface. Regional models [e.g. *van der Hilst and Engdahl*, 1991] and global models that use adaptive parameterization [Chapter 4; *Bijwaard et al.*, 1998; *Kárason and van der Hilst*, 2000] have revealed complex flow trajectories in the transition zone, most likely the combined result of trench migration, the dynamical effects of phase changes and increased viscosity between the upper and lower mantle [*Griffiths et al.*, 1995].

In the past decade, many papers have related subduction history to tomographically derived slab structure in the mantle. For example, bands of fast material in the mid-mantle along the Americas and the southern margin of Asia are thought to be remnants of the subducted Farallon plate and Tethys ocean, respectively [*van der Hilst et al.*, 1997; *Grand et al.*, 1997]. On a regional scale qualitative inferences have been made, for example, for the Northwest Pacific [*van der Hilst and Seno*, 1993], Tonga [*van der Hilst*, 1995], Indochina [*Widiyantoro and van der Hilst*, 1996] and southern Europe [*Faccenna et al.*, 2001]. These studies commonly use simplified two-dimensional flow models, simple kinematics or cartoons to make their arguments. They have increased understanding of local geology, filled in gaps in the subduction history and given estimates for sinking rates. However, only a few studies have used dynamical modeling in three dimensions to investigate slab structures [e.g. *Bunge and Grand*, 2000; *Gurnis et al.*, 2000] and the simplified approaches cannot constrain the viscosity structure of the mantle or address if apparent inconsistencies might be the results of 3-dimensional flow in the mantle (see Chapter 6 for details).

We are well positioned to go beyond such kinematic approaches since we have developed a high resolution tomographic model of mantle P-wave speed as well as a code for 3-D flow modeling (see Chapters 4 and 5). This allows a more direct comparison between tomographic imaging and plate reconstructions than before. Here, we apply this approach

on a regional scale. As a study area we choose the south-eastern margin of Asia. Previous studies of this region, connecting seismic tomography to subduction history include *Widiyantoro and van der Hilst* [1996]; *Van der Voo et al.* [1999]; *Hafkenscheid et al.* [2001]; *Replumaz et al.* [2002] but none used flow modeling.

We use our new tomographic model (see Chapter 4) and the reconstruction by *Replumaz and Tapponnier* [2001] of the change in geometry of the plate boundaries associated with collision of India to mainland Asia. They synthesized the available tectonic data and reconstructed the continental deformation from the westernmost part of India to the eastern margin of the Sunda block. Guided by their reconstructed coastline location at different times we draw estimated convergent margins (see Figure 7-1). We first compare those margins to the tomographic model at different depths to evaluate how they agree (See *Replumaz et al.* [2002]). Having established their conformity we model the subduction along the Sunda-arc using the plate tectonic record and our tomographic model as a reference.

Furthermore, seismic tomography has a finite resolution which varies spatially according to ray geometry. In addition, due to the noise in the data, the tomographic inversions need to be regularized, which reduces the magnitude of the inferred wavespeed variations and smoothes out sharp gradients, even in the region under study, which is relatively well sampled. Using inversion tests we can estimate how well the tomography resolves synthetic mantle structure based on our flow modeling. This way we can assess to what level of detail we can carry the comparison between the tomography and the flow modeling.

7.2 Plate reconstructions and seismic tomography

7.2.1 Plate reconstructions

The plate reconstructions used in this study are from *Replumaz and Tapponnier* [2001], shown in Figure 7-1 (red lines). As India moved north the Sunda-block rotated clock-

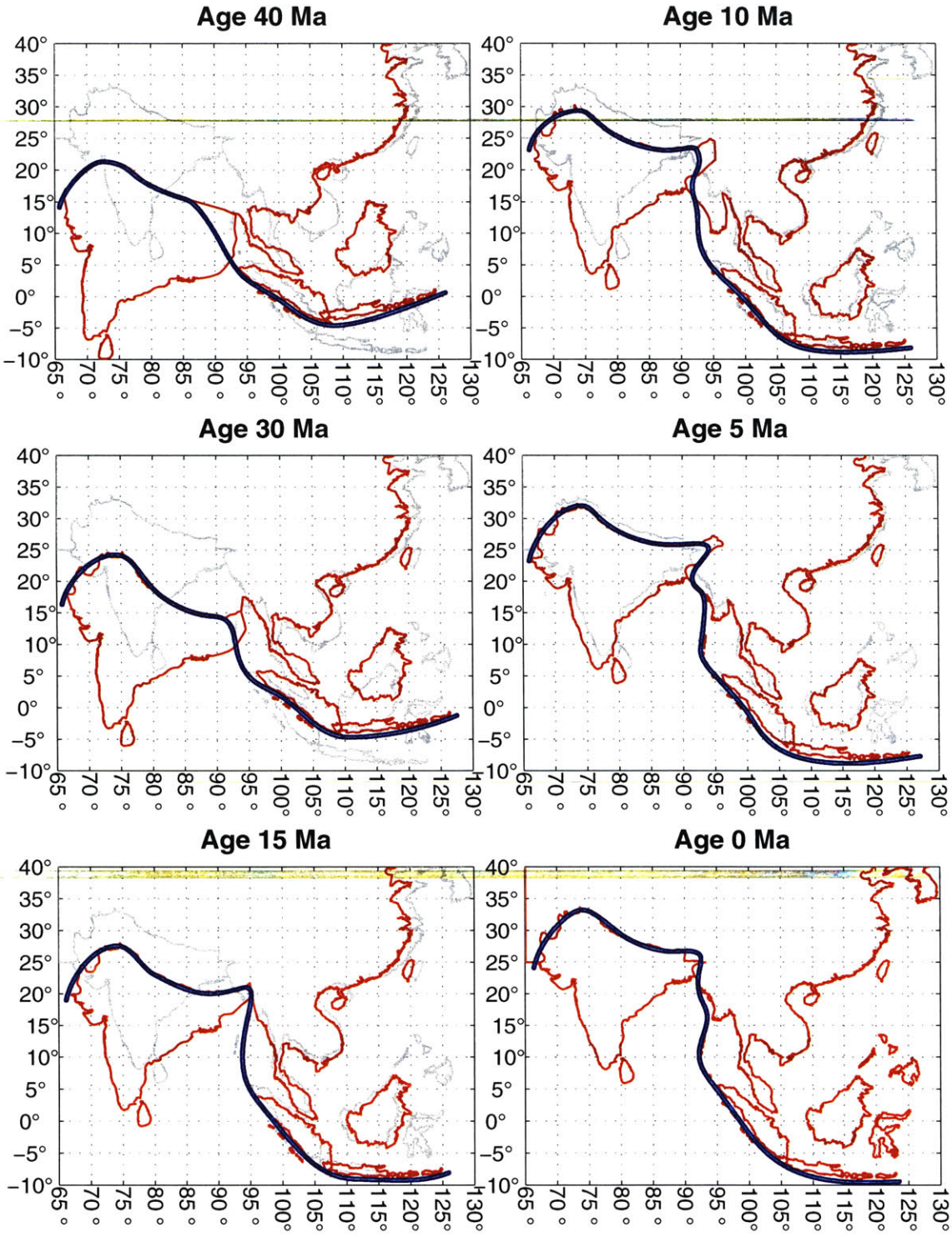


Figure 7-1: (red lines) Outlines of continents according to plate reconstruction by *Replumaz and Tapponnier* [2001] at different times as indicated. (gray lines) Current coastline. (blue lines) Plate-margins inferred from ancient coastlines.

wise toward the south and southwest. In their model, they assume that the deformation is accommodated along faults between relatively rigid blocks, and track the modern coastlines of those blocks back in time. We use use the locations at 5 Ma, 10 Ma, 15 Ma, 30 Ma and 40 Ma back. From these coastlines we estimated a plate margin for comparison with our tomographic models. To the west, the margin is between India and the Tethys ocean, or mainland Asia after the whole oceanic floor had subducted. To the east, the margin is between the southern edge of the Sunda block and the Indian Ocean. Although the margin is connected all the way from east to west, the motion of the plates in the center, i.e. near the Andaman sea and Northern Sumatra, resembles a strike-slip fault rather than convergence. Projecting the current coastline back in time may introduce substantial uncertainty in the location of the convergent margins, becoming larger the further we go back in time. There is therefore a substantial uncertainty in the exact location of the convergent margins, which increases with age. *Replumaz and Tapponnier* [2001] comment that beyond 30 Ma, the position of the blocks should be considered as a working hypotheses, and anything beyond 40 Ma would be an unconstrained speculation. In particular, it is not certain where and when subduction initiated along the eastern margin.

7.2.2 Plate margins and tomography

In Figure 7-2 we compare the plate margins obtained from the reconstructions to the global tomographic model of mantle P-wave speed that was presented in Chapter 4. Looking at the currently active subduction zone along the Sunda-arc, the margins and the top 300 km of the model show very good agreement with the reconstructed plate margin as far back as 15 Ma. Deeper, the slab is located farther to the north and to the east, which agrees with the 30 Ma plate margin. At 600 km or 700 km depth the subducting slab seems to be between the 15 Ma and 30 Ma margins. This gives 20 Ma-25 Ma as a very crude estimate of the time it takes the sinking material to reach the 660 km discontinuity, or around 3 cm/year. The

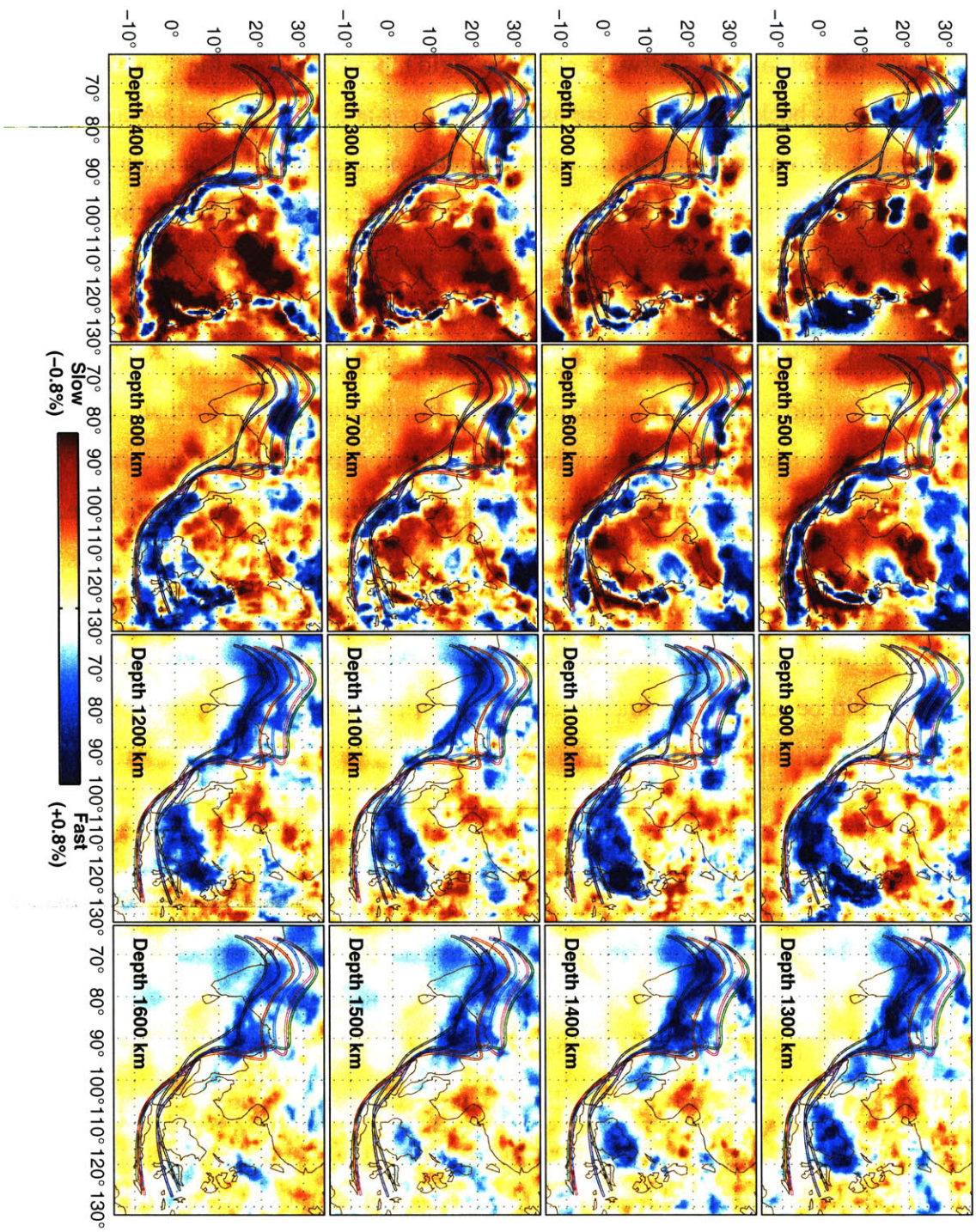


Figure 7-2: Horizontal slices through global tomographic P model showing the mantle beneath South-East Asia at 100 km depth intervals as indicated. Also drawn are ancient plate margins (see Figure 7-1) of age 40 Ma (black), 30 Ma (blue), 15 Ma (red), 10 Ma (cyan), 5 Ma (magenta) and current (green).

40 Ma margin has the closest, although not perfect, match with structure around 900 km depth, consistent with a sinking rate of 1-2 cm/year in the lower mantle. Still deeper, most of the fast structures are displaced to the north and seem consistent either with a more northerly margin or a northward advection. As mentioned before, the location of the plate margins is very speculative near and beyond 40 Ma. The discrepancy could therefore be explained either with the estimated margin placed too much to the south or with an earlier initiation of subduction, at a more northerly margin.

Around longitudes 90°E-100°E, the subduction seems much shallower, and mostly confined to the upper mantle. This is consistent with the plate reconstructions, as they suggest little convergence in this region.

In the shallow tomograms the mantle structure beneath India is overwhelmed by continental signal. This signal is smeared to large depths, resulting in artificially deep continental roots, as is the case for many continental areas that are only constrained by teleseismic rays. Around 600 km-700 km depth, structures consistent with the 5 Ma-10 Ma margins emerge. These structures widen and move south with depth, and the agreement between the oldest margins and images increases. For the entire depth range between 1000 km and 1600 km, the correlation with the oldest margins is good, indicating that much material was subducted, either over an extended time period at a near-stationary trench or over a brief period of very intense subduction.

Sinking rates of subducting slabs are influenced by many factors, including the age and thickness of the lithosphere at the trench, the relative plate motion and the background flow in the mantle. This is further underscored by the very different rates of plate movements at the surface [*Gripp and Gordon*, 1990]. There is therefore no reason to assume that the sinking rates in the upper mantle were the same at the east and west margins. Furthermore, boundary conditions, such as geometry and trench migration, will change the ratio between the sinking rates in the upper and lower mantle, as discussed in Chapter 6. Therefore, it

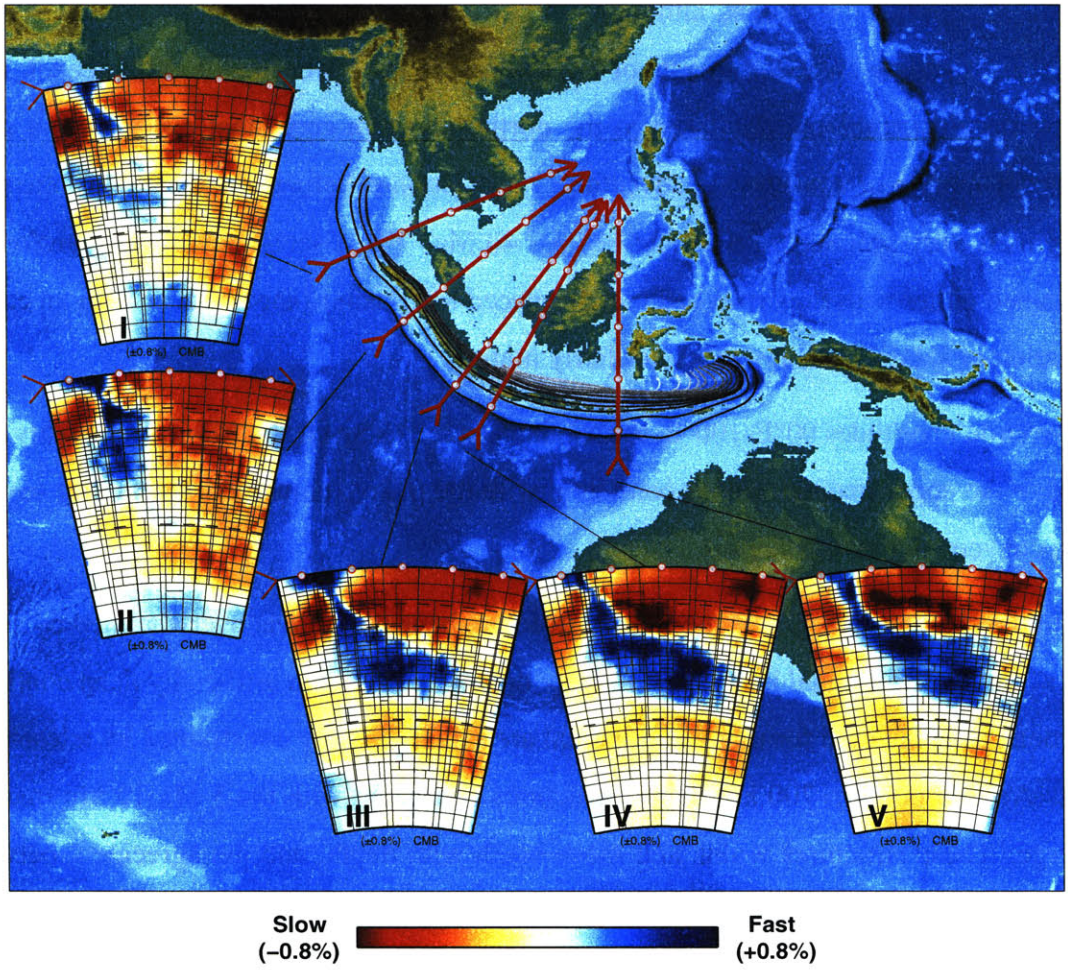


Figure 7-3: Cross sections through the Sunda-arc. Also shown are the outlines of the constant slowness ~~blocks in the adaptive~~ grid.

is not to be expected that structure at one particular depth fits the whole margin, from west to east, exactly. With this caveat in mind, the resemblance between the margins and the structures at various depths is striking and suggests that surface tectonics are clearly related to both upper and lower mantle structure.

7.2.3 Slab morphology

In this study we focus the flow modeling on the active subduction along the Sunda-arc. The cross sections in Figure 7-3 illustrate the slab morphology as inferred from our global tomographic imaging. To the east, in sections III to V, the subduction is steep but becomes much more horizontal and thicker in the lower mantle. This is consistent with much slower sinking rates in the lower mantle as discussed in Chapter 6. To the west, in sections I and II, the penetrating depth of the slab gradually decreases, which is most likely the combined effect of a younger trench and the large strike-slip component of the motion, which means that the convergence rate is small and subduction less active [*Widiyantoro and van der Hilst*, 1996]. The slab morphology seems to be pronounced and the edges sharp with little evidence of smearing.

Superimposed on the cross sections in Figure 7-3 are the outlines of the constant slowness blocks used in the inversion. The fine grid implies that in general the slab structure is well sampled, and, in particular, the sharp transition to the north in cross sections III to V seems well resolved. This is an important constraint, as the latitude of the initial trench is unknown.

7.3 Flow modeling

While designing the numerical experiments many parameters must be chosen, tested and adjusted. We kept simulations as simple as possible, while retaining the detail that allows a direct comparison with the tomography and the plate reconstructions.

7.3.1 Trench geometry and migration

As outlined in Chapter 5, the subducting slab is modeled with a 3-D surface which represents its boundaries. At 100 km depth, this surface has an opening which does not move

radially throughout the simulation. Material flows into this "mouth" of the 3-D surface, which in turn expands. The material inside the surface is assumed buoyant and it drives the flow. Designing the "trench" means to geometrically prescribe how this mouth of the surface migrates and stretches.

For the trench position we select a starting configuration and a final configuration. We then devise a scheme to migrate the trench between those positions. In Figure 7-4 we depict the time evolution of the Sunda-arc in our modeling. We model the trench as a *small circle* segment at all time steps. The center of the small circle, its radius and the azimuth from the center to the two endpoints of the arc representing the trench then *migrate* between those two end positions. The broken lines in Figure 7-4 indicate intermediate steps. We note that these steps are much larger than used in the simulation itself, which typically involves 250-500 steps. In the simulations presented here, the migration rate is kept constant with respect to the easternmost part of the trench, although this is not required.

The solid line to the north shows the location of the trench at the start of a simulation. As pointed out before, and can be seen in Figure 7-4, the fast structures extend to the north of the oldest plate margin in the mid-mantle. As the onset time and location of the subduction cannot be constrained by the plate reconstruction, we use the tomographic model as a reference and start the subduction farther to the north than the oldest plate margin in the model by *Replumaz and Tapponnier* [2001]. How much farther north depends on the viscosity contrast between the upper and lower mantle. The solid line to the south shows the final location, which is fixed and does not change between simulations. Plate reconstructions indicate that the location of the Sunda-arc has not changed significantly over the last 15 Ma. The simulations are therefore not stopped when the trench has reached its final configuration, and the final stage of the simulations assumes a stationary trench. The final model-trench is not a perfect fit to either the tomography or the plate margins, but for the purpose of our modeling, this misfit is not important.

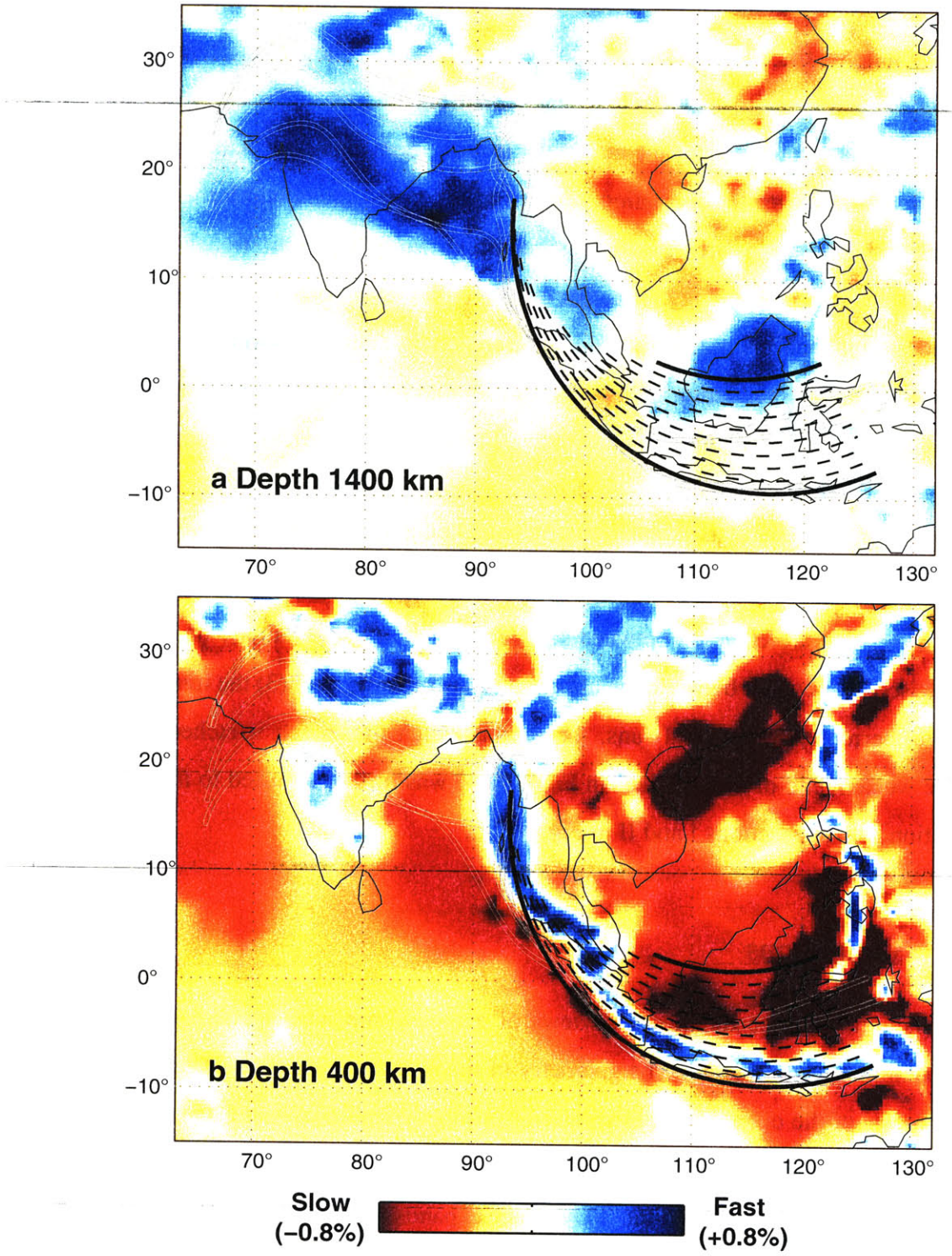


Figure 7-4: (a) The tomographic model at 1400 km depth with the initial (northerly solid lines) and final (southerly solid lines) with intermediate steps (broken lines). Plate margins are also shown (gray lines, see Figure 7-1). (b) Same as (a) except tomographic model at 400 km depth.

One important parameter that remains to be chosen is the trench width, i.e. the width of the opening of the 3-D surface. As shown in Chapter 6, the sinking rate scales fairly uniformly with the trench width. This means that trench width scales approximately with the excess density of the buoyant material. In turn, density scales with the velocity in general and hence with the migration rate. Changes in the mean width therefore mostly trade-off with migration rate. However, changes in width along the trench or with time will affect the slab morphology. It is, for example, not unreasonable to change the width according to the local migration rate along the trench, as faster migration rate would most likely signify faster subduction. This would mean that more material per unit time would sink at the eastern part of the trench, relative to the western part. However, to keep the simulations as simple as possible, we decided to have the thickness 100 km and uniform with time and over the whole trench. Only in a couple of test runs was the thickness varied, to verify that it would not significantly change our findings.

Variables that are changed between simulations are: (1) the latitude of initial trench; (2) the viscosity contrast between the upper and lower mantle; and (3) the trench migration rate. Furthermore, in the final runs, the rate at which the trench stretched to the west (i.e. the western end of dashed lines in Figure 7-4) was further adjusted to fit the depth of subduction in cross sections I and II in Figure 7-3.

7.3.2 Simulations

Figure 7-5 shows three dimensional views of a typical run with a 100-fold increase in viscosity in the lower mantle. The geometry of the trench as described in section 7.3.1 is the geometry of the opening of the surface as described earlier and in Chapter 5. Initially the surface extends from 100 km to 200 km depth but buoyant material flows in through the top, which is at fixed depth, and the 3-D surface expands. The highly viscous lower mantle is a formidable barrier to the slab which, in response, flows more to the sides creating a

sort of a slab "head". At the later stages, slab structure is fairly uniform to the east but becomes gradually shallower to the west, where subduction starts later. The thin "tails" of the slab head are real, from a fluid dynamics perspective, and are similar to the fine morphology of ascending thermal plumes [e.g. *Campbell et al.*, 1988]. However, as they contain little volume they can for most parts be considered as passive flow tracers, and, as will be demonstrated later, cannot be resolved by the tomography.

7.4 Results

7.4.1 Matching tomography and flow simulations

In Figure 7-6 we show cross sections through the tomographic model and through five simulations that are representative of the more than 50 runs that were made to explore the parameter space and find the best match to the tomography. We show these simulations at two instances: around the time they gave the best match, and at the moment the run started (in sections III-V that cut through the initial trench).

A viscosity increase of ~40 in the lower mantle is often quoted and used in simulations [e.g. *Lithgow-Bertelloni and Richards*, 1998; *Bunge and Grand*, 2000]. The initial simulations were run at a slightly smaller contrast of 30. This meant relatively little difference in sinking rates between the upper and lower mantle. In this case, the migration needed to be very rapid, as can be seen from the shallow subduction angle in Figure 7-5a, in order for the slab not to have sunk too deep at the end of the simulation. This is even true for a trench that starts further south compared to what is required to match the northern edge of the subducted slab in the lower mantle. Here, the shape of the slab is concave down, which is opposite to what is observed in the tomography; a viscosity contrast of 30 results in flow patterns that are not compatible with the tomography. After a handful of tries the simulations were not carried further; a more viscous lower mantle is clearly needed.

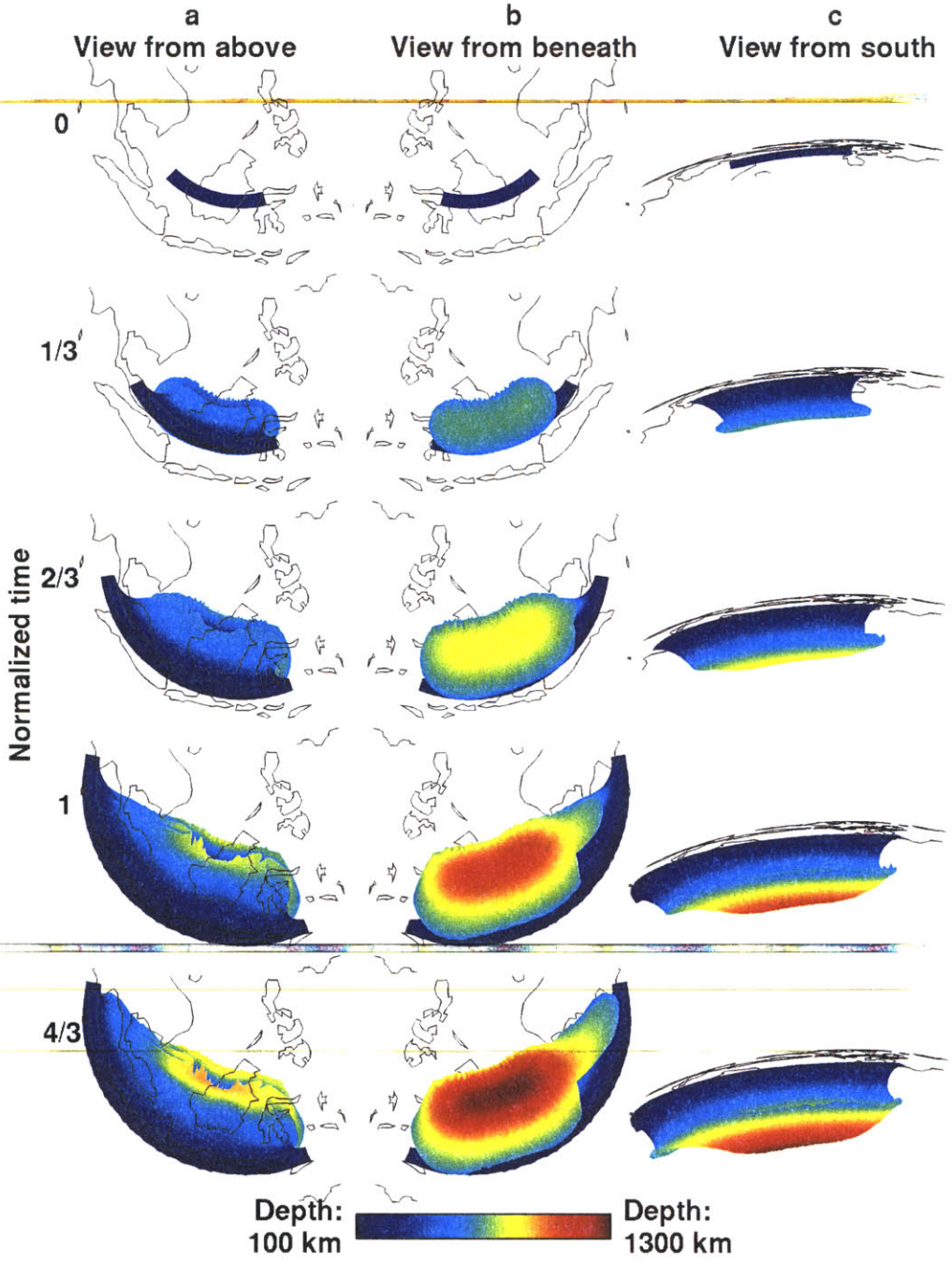


Figure 7-5: 3-D views of a typical simulation with 100-fold increase in viscosity in the lower mantle. (a) View from above. (b) View from beneath. (c) View from South. Time is normalized with the time it takes the trench to migrate from its initial position to the final position, where the simulation runs from a stationary trench.

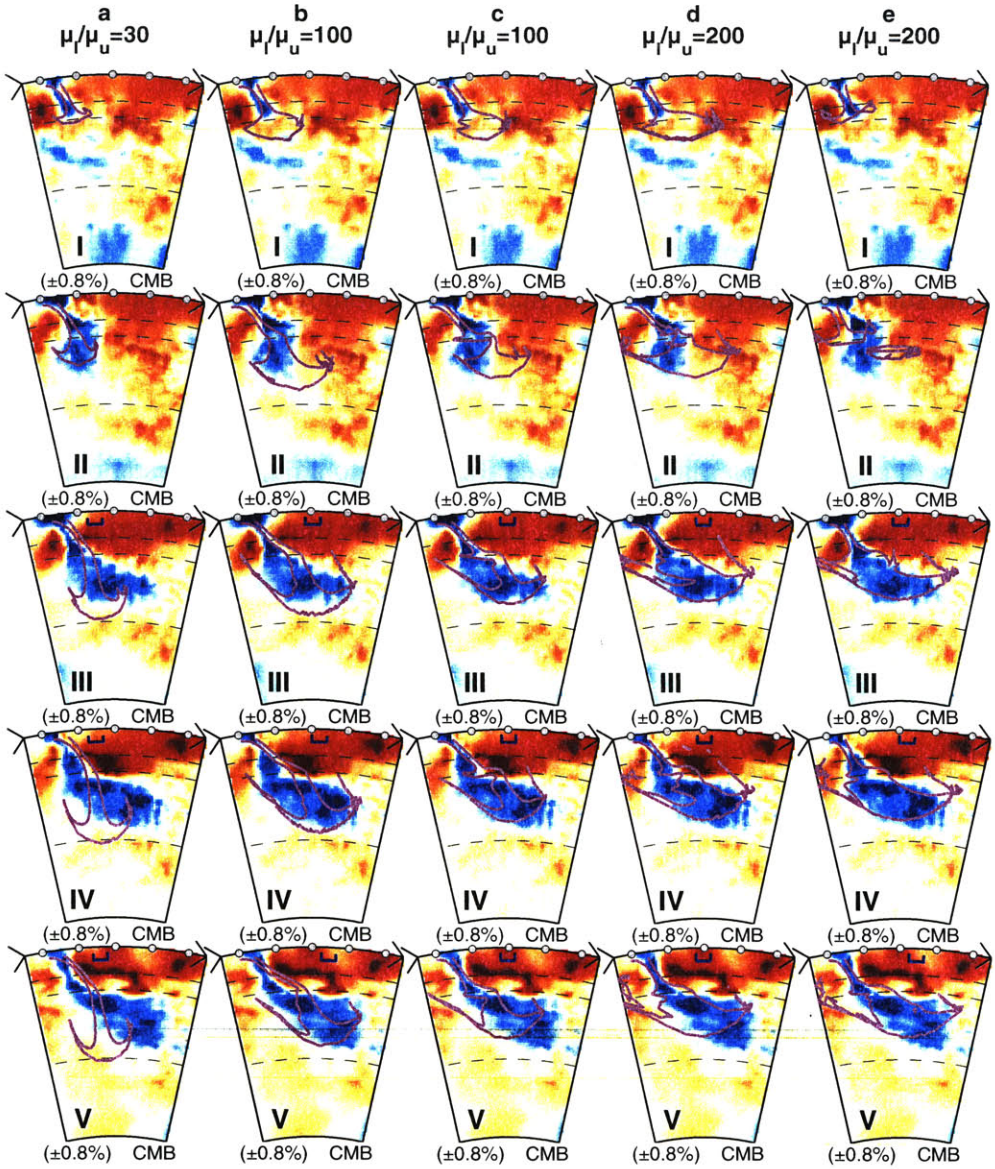


Figure 7-6: (a) Cross sections I-V through the tomographic model (see Figure 7-3) and through a simulation with a 30-fold increase in viscosity in the lower mantle, at a time step which gave the best match to the tomography (purple lines) and at the start of the simulation, where visible (sections III-V, blue lines). (b) Same as (a), except for 100-fold increase in viscosity in the lower mantle and a different initial trench position. (c) Same as (b), except with a different trench position and a faster migration rate followed by a longer phase of sinking from a stationary final trench. (d) and (e) Same as (a), except for a 200-fold increase in viscosity in the lower mantle and different positions of initial trench (The apparent disconnection in cross section II in (e) is created by cutting through the slab head that resulted from earlier subduction to the east).

More effort was made in trying to fit the slab structure using a 100-fold increase in viscosity in the lower mantle. The initial trench was moved north and south, the migration rate varied and the width of the trench changed to verify that it was not having an impact on our results. In Figures 7-6b and 7-6c we show two typical results from those runs. These results demonstrate that the initial trench does not need to be as far north as the northern edge of the subducted slab dictates. Still, matching the tomography was problematic. Sinking rates in the lower mantle were still too fast, as can be seen in Figure 7-6b, and the fast migration rates resulted in too shallow subduction in the upper mantle. To counteract this, we increased the migration rate and allowed the trench to adjust in the stationary phase at the end of the simulation for a longer time. In this case the trench rearranged itself as can be seen in Figure 7-6c, and started to disconnect from the body of the material that subducted earlier. Similar flow patterns have been observed in laboratory experiments [*Griffiths et al.*, 1995]. The run depicted in Figure 7-6c gives the best fit for a 100-fold increase, but not satisfactory in our judgment; although the angle of subduction in the upper mantle is similar the continuous connection along the slab is lost and the shape is quite different.

The last set of simulations were made with a lower mantle that was 200 times more viscous as the upper mantle. It became evident that it was much easier to make the match in this case, as can be seen in Figures 7-6d and 7-6e. Here, both the lateral extent and the angle of subduction in the upper mantle are convincingly reproduced. In the run in Figure 7-6e, the stretching of the trench to the west happened later to give a better fit to the depth penetration of the slab in sections I and II. The three dimensionality of the flow is evident in section II where the slab head, extending from deeper subduction in the east, appears detached from the proper subducting slab. We selected this simulation as our best fit to the tomography.

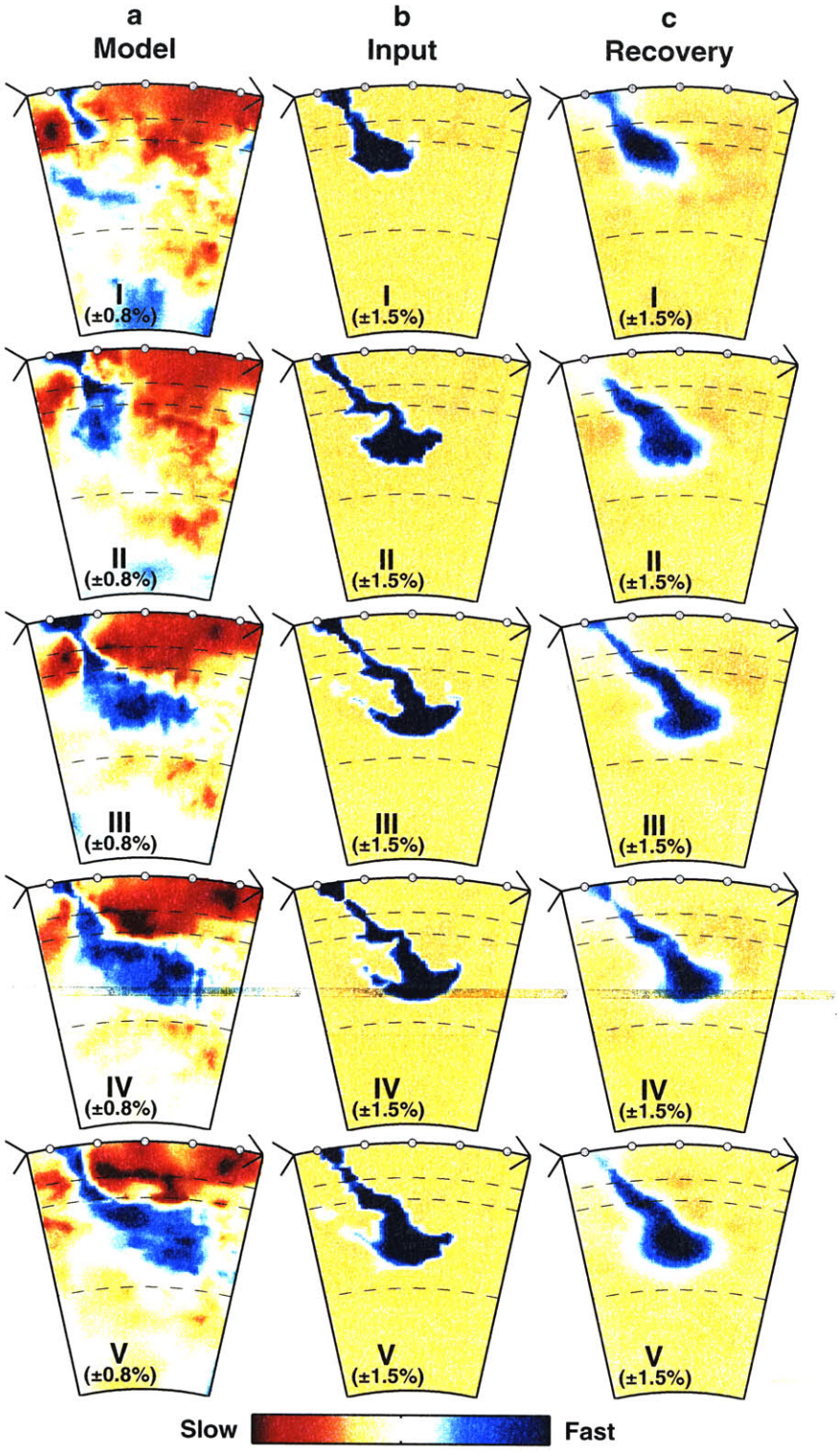


Figure 7-7: (a) Cross section I-V through the tomographic model (see Figure 7-3). (b) Cross section I-V through the input model for the inversion test, which is created by projecting the flow simulation depicted in Figure 7-6c with a 100-fold increase in viscosity in the lower mantle. (c)

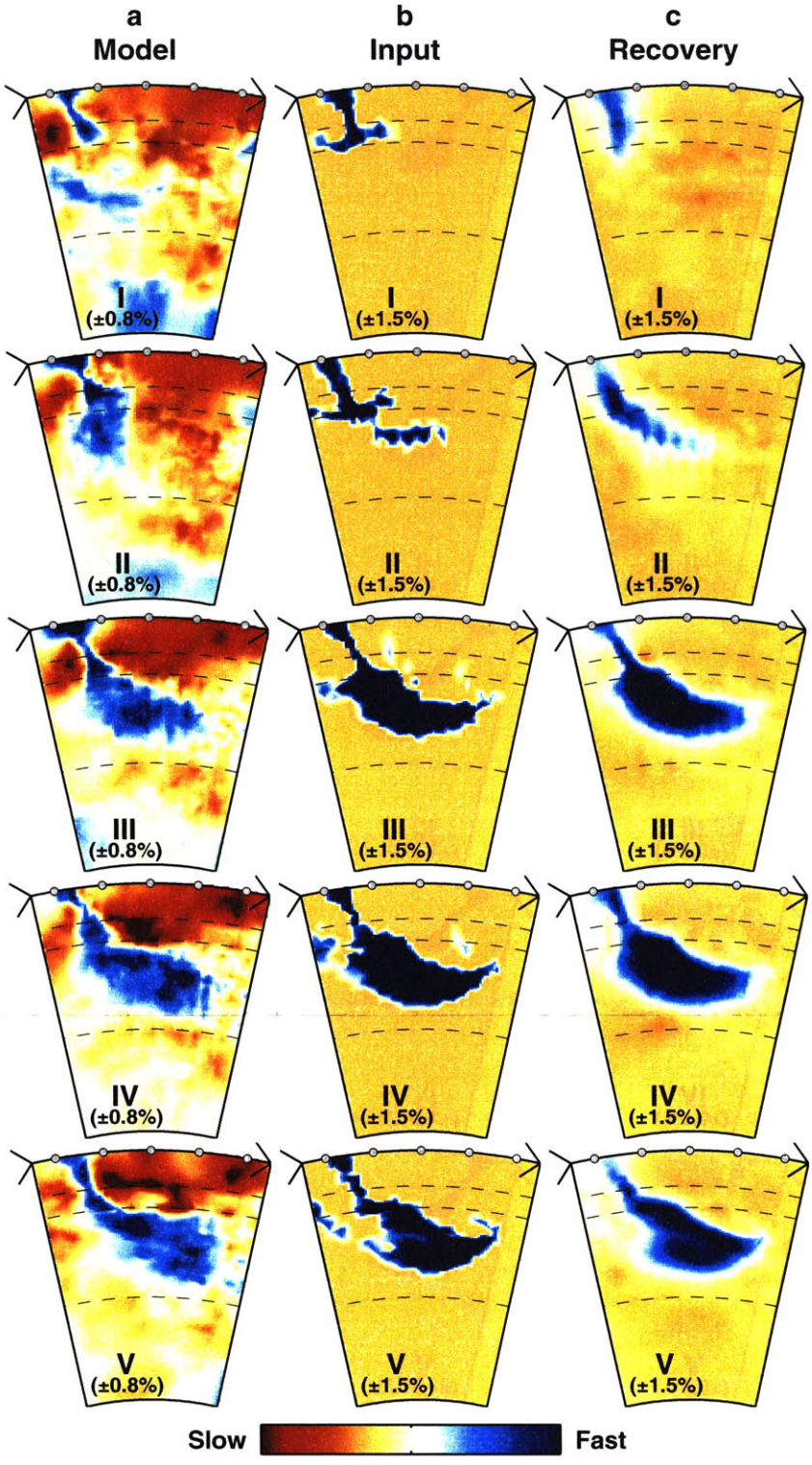


Figure 7-8: Same as Figure 7-7 except based on the flow simulation depicted in Figure 7-6e.

7.4.2 Resolution

Visually, the fit of the flow models depicted in Figure 7-6e to the tomography is good. However, tomography has shortcomings and the fine structure, such as is present in the flow models, is never completely resolved. Therefore, we tested if the flow models in Figure 7-6c and 7-6e as imaged by tomography would still lead us to the same conclusion, i.e. reject the former and accept the latter.

In Figures 7-7 and 7-8 we show the results of the two tests. The 3-D surface from the flow simulations marks the boundaries of negatively buoyant material, i.e cold and fast in our interpretation. This complicated three-dimensional volume is projected onto the basis functions used in the tomographic inversion (i.e. the adaptive grid) and assigned a velocity contrast of 2% compared to the reference velocity model. Similar to what is observed in the tomography, we put a background slow velocity contrast of -0.4% in this region. The parameterization of the tomography is much coarser than the basis function for the flow modeling and the projection onto this basis therefore already taints the results from the flow modeling; the finest details are somewhat lost as can be inferred from the comparison of Figures 7-7b and 7-8b with 7-6c and e.

Synthetic travel time data was then created by tracing exactly the same rays as used in creating the tomographic model through this artificial structure (in practise, multiplying the sensitivity matrix with the synthetic model). These travel times were then inverted with the same inversion scheme and damping parameters as before. The results from these inversion tests are shown in Figures 7-7c and 7-8c. In this region of good resolution, the only bias in the tomography is some smoothing and loss of amplitude, which was to be expected given the damping we apply in the inversion. The tomographic model should therefore reveal most characteristics of the subducting slab, making the comparison with the flow modeling meaningful. Notice, however, that the narrow flow structures off the slab "head" are not resolved.

7.5 Discussion

7.5.1 How viscous is the lower mantle?

According to the results presented here the lower mantle seems to be more viscous than often thought. Indeed, a recent estimate of mantle viscosity shows a complex viscosity profile with a viscosity high around 2000 km depth, which is more than 5 orders of magnitude higher than the low around 200 km depth and more than 3 orders of magnitude more viscous than the average of the upper mantle [*Forte and Mitrovica*, 2001]. Compared to these values, a 200-fold increase in viscosity is not unrealistic.

However, there are few other known ways of slowing down the subducting slab in the lower mantle, effectively decreasing the required viscosity contrast. A gradually increasing viscosity, even over large radial distance would give comparable results. A deep dense layer [*van der Hilst and Kárason*, 1999; *Kellogg et al.*, 1999] or other layering within the earth [*Anderson*, 2001] would slow the slab down. If the excess density of the deep layer is similar or higher than the excess density of the slab, the effects would be felt as soon as the slab enters the lower mantle (see Chapter 5). A reduced thermal expansion coefficient with depth will also reduce the effective density, slowing the slab. However, as shown in Chapter 6 common estimates of thermal expansion coefficient do not affect the sinking rates significantly down to ~1500 km depth. Similarly, heating up of the slab would decrease its effective density. Some details of the subduction process, such as a delayed phase transition as the slab sinks past the 660 km depth discontinuity would also hamper flow.

Other aspects of mantle flow that are not modeled will affect the flow, but in which way is not simply predictable. Plate motions at the surface will both affect the subduction directly and create a background flow. However, this surface motion will be largely decoupled from the highly viscous lower mantle. Temperature- and strain-dependent viscosity might critically change the slab behavior, especially at shallow depths. However, as out-

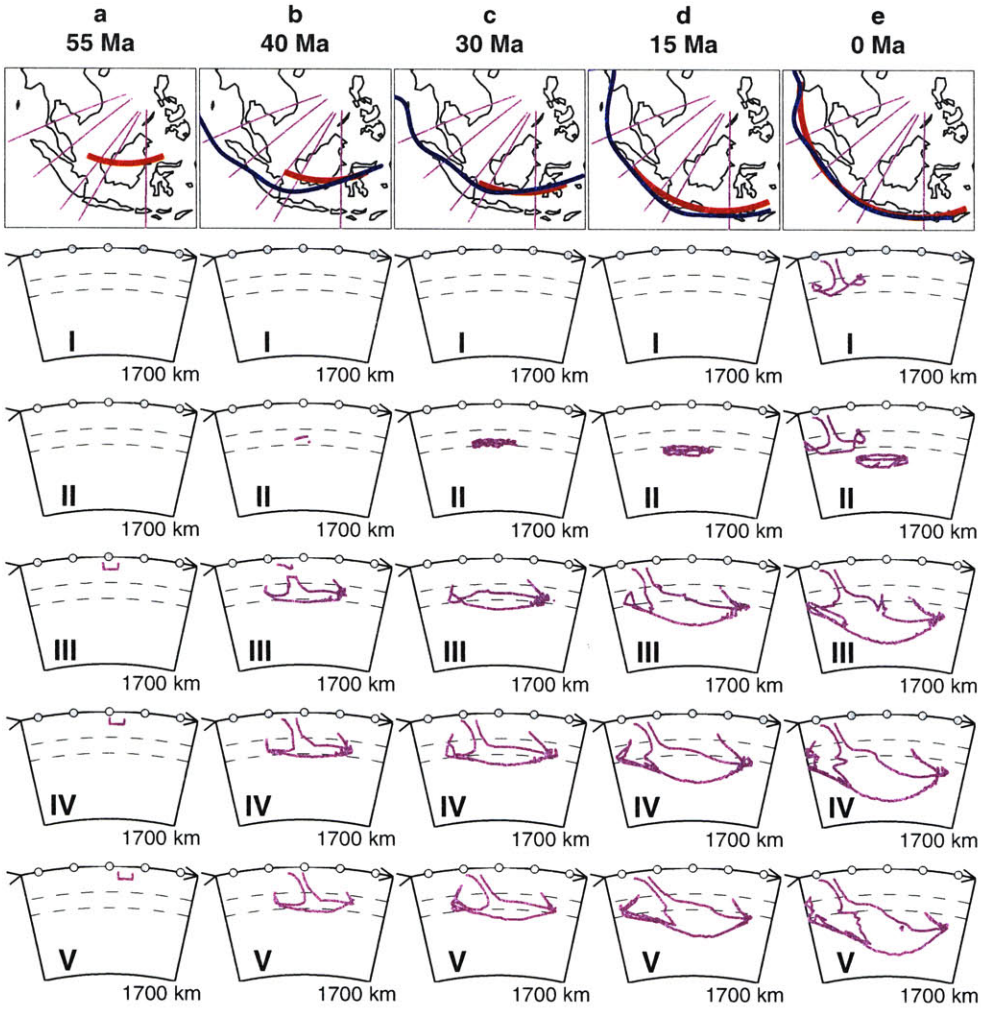


Figure 7-9: Maps of the current coastline (black lines), location of cross sections (purple lines), the ancient plate margins (blue lines) and the location of the trench (red lines) during simulations for different ages (a-e) as indicated. Also shown are cross sections I-V (see Figure 7-3 for reference) through the flow model (Figure 7-6e at the same time intervals.

lined in Chapter 5 many observations indicate that slabs are weak, so it is not obvious that this is an important effect in the Earth.

7.5.2 Initiation of subduction

In Figure 7-9 we show the time evolution of the subduction. By matching the 30 Ma old margin and the trench in our simulation, we can establish a time scale. This time scale is not very well constrained, e.g. it does not fit the 15 Ma old margin particularly well. Mindful of all the uncertainties in plate reconstruction, tomography, and the matching of flow modeling to either one, we find this to be consistent with an initiation of subduction 55 Ma ago along a trench that is ~200 km to the north of the oldest plate margin. According to *Replumaz and Tapponnier* [2001] the northern drift of India started to impact this area around 50 Ma or 60 Ma years ago, which is consistent with our findings. The initial trench is located several hundreds of kilometers to the south of the northernmost part of the subducted slab. As observed in some subduction zones, [*Fischer et al.*, 1988], the thickening above the lower mantle is very significant, as can be seen at 40 Ma. These early stages in the simulation bear resemblance to many images of deflected slabs shown in Chapter 4.

The depth to the deepest point of the slab is shown as function of time in Figure 7-10a. It only takes the slab ~5 Ma to reach the 660 km discontinuity but another ~30 Ma years to reach 1000 km depth. This is consistent with the thickening above 660 km in the early stages, i.e. not until enough mass has accumulated does the slab make significant advances into the lower mantle. This is further demonstrated in Figure 7-10b, where the sinking rate of the leading edge for different depths is depicted. Around 400 km depth the leading edge has a maximum sinking rate of ~13 cm/year, but slows down to less than 1 cm/year at 660 km. With time, as mass accumulates, the rate of sinking picks up and is close to 2.5 cm/year at the end of the simulation. The sinking is clearly transient in nature and, as shown in Chapter 6, the velocities within the slab might be quite different than for the leading edge but they are harder to quantify.

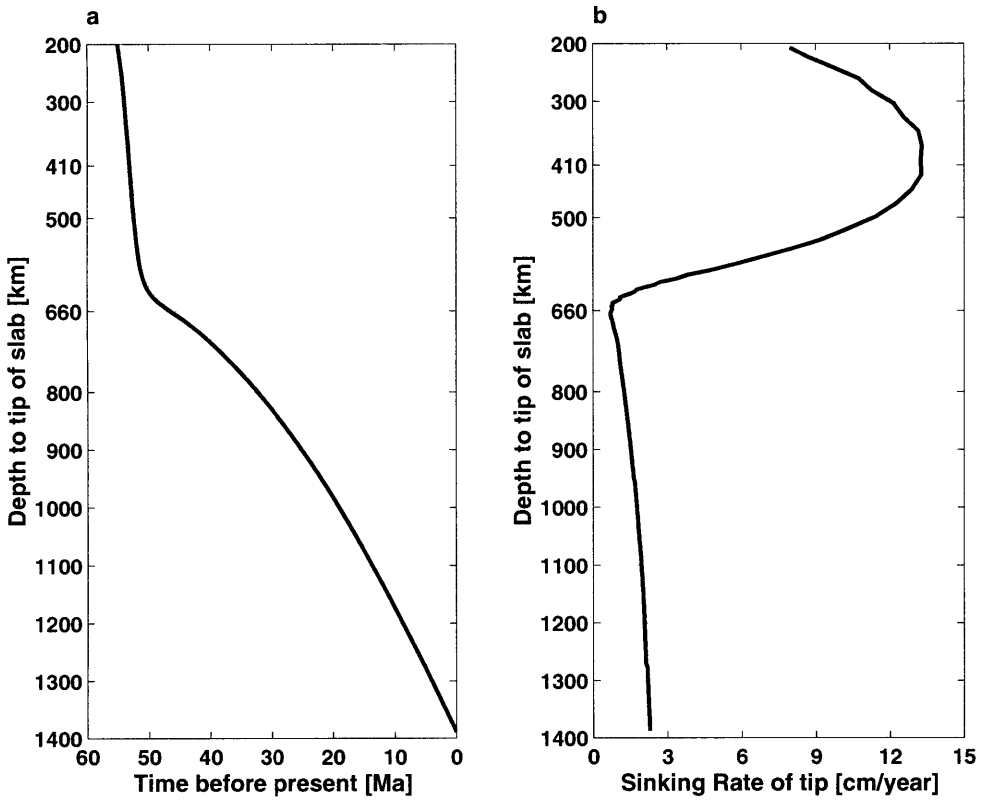


Figure 7-10: (a) Depth to the leading tip of the slab according flow simulation depicted in Figure 7-6e as function of time before present. (b) Sinking rate of the leading tip of the slab for different depths.

7.5.3 Absolute values of viscosity

The comparison of the flow modeling and the tomography can constrain the viscosity contrast between the upper and lower mantle. However, absolute values of time, buoyancy, and viscosity still trade off with each other as discussed in Chapter 5.5. Attuning the flow modeling to plate motions gives a timescale, but leaves the buoyancy and the absolute value of viscosity unresolved. Using lithospheric cooling models we can estimate the excess density of the slabs to 75 kg/m^3 (see Chapter 5.5). Although both the time scale and the excess density estimates have considerable uncertainties this gives an opportunity to constrain the absolute value of viscosity. According to Table 7.1 we get $\sim 6.7 \times 10^{20}$ Pa s and

$\sim 1.3 \times 10^{23}$ Pa s for upper and lower mantle viscosity, respectively, for our preferred model with a viscosity contrast of 200. This about an order of magnitude higher than reported as average lower mantle viscosity by both *Simons and Hager* [1997] (1.1×10^{22} Pa s) and *Forte and Mitrovica* [2001] (1.5×10^{22} Pa s). The upper mantle viscosity is also higher, about 2-4 times compared to what is reported in these studies. However, the viscosity in the shallow mantle beneath continental areas as measured by post-glacial rebound in these studies may very well be different from subduction zones. Furthermore, the sensitivity of post-glacial rebound to mantle viscosity diminishes with depth. Instead, *Forte and Mitrovica* [2001] constrain lower mantle viscosity by flow modeling based on long wavelength structure from seismic tomography. As evident from the Stokes flow calculations in Chapter 6, sinking rates are sensitive to the scale of the flow and ignoring the small dimensions of subducting slabs may lead to biases in viscosity estimation, probably toward lower values.

Table 7.1: Estimates of absolute values of viscosity of upper and lower mantle assuming 75 kg/m^3 excess density of the subducting material and a span of 55 Ma for each of the simulations. Values for the preferred flow model are shown in bold typeface.

Viscosity contrast	Upper mantle viscosity	Lower mantle viscosity	Reference
30	18×10^{20} Pa s	0.53×10^{23} Pa s	Figure 7-6a
100	9.0×10^{20} Pa s	0.89×10^{23} Pa s	Figure 7-6b
100	10×10^{20} Pa s	1.0×10^{23} Pa s	Figure 7-6c
200	7.3×10^{20} Pa s	1.5×10^{23} Pa s	Figure 7-6d
200	$\mathbf{6.7 \times 10^{20}}$ **Pa s**	$\mathbf{1.3 \times 10^{23}}$ **Pa s**	**Figure 7-6e**

To check how sensitive these estimates of viscosity are to different viscosity contrasts we also show in Table 7.1 the values for the other simulations displayed in Figure 7-6. We use the same time span and excess density as before. The depth to the tip of the slab is similar in all simulations but the shape is quite different, which means that average sinking rates and average viscosity are similar between simulations. Therefore, lowering the viscosity contrast increases the upper mantle viscosity and decreases the lower mantle

viscosity. Viscosity contrast of 30 results in 2-3 times more viscous upper mantle and 2-3 times less viscous lower mantle when compared to the preferred viscosity contrast of 200. The viscosity estimates are therefore not very sensitive to the contrast chosen and still higher than previous estimates for lower mantle viscosity in all cases.

7.6 Conclusions

Comparing seismic tomography and plate reconstructions in Southeast Asia reveals similarities, which provide strong evidence for a relation between plate motions at the surface and structures in the mantle, at least down to mid-mantle depths. Some discrepancies still exist, especially between the oldest margins and the very deep structure they should correspond to. Using flow modeling, we can largely explain these differences in the Sunda-arc region with a slightly more northerly margin at times prior to the oldest reconstructed margin. More specifically, our findings are consistent with an onset of subduction along a trench, $\sim$200 km northward of the 40 Ma margin around 55 million years ago.

The flow simulations also require a viscosity contrast of $\sim$200 between the upper and lower mantle to be consistent with the thickness and lateral extent of the slab in the lower mantle and its dip in the upper mantle. Although not unprecedented, this is higher than often quoted. Furthermore, our estimate of the absolute value of lower mantle viscosity is $\sim 1.3 \times 10^{23}$ Pa s which is about an order of magnitude higher than previous estimates. Compared to the fastest sinking rate in the upper mantle, the lower mantle sinking rates are 5 to 10 times slower. However, there are other means of slowing down the slab in the lower mantle which might lower the required viscosity. The resolution of the slab subducting at the Sunda-arc is quite good and the recovery of flow model structures in inversion tests show only some smoothing and loss of amplitude, but no other systematic biases are observed. The comparison between the flow models and the tomography is therefore valid.

Converging on a flow solution that matched the subducted slab of the Sunda-arc as imaged by tomography required more than 50 simulations. The advantages of a computationally efficient flow code are obvious.

Bibliography

Abers, G. A., and S. W. Roecker, Deep-structure of an arc-continent collision: Earthquake relocation and inversion for upper mantle P and S wave velocities beneath Papua New Guinea, *J. Geophys. Res.*, *96*, 6,379–6,401, 1991.

Aki, K., A. Christoffersson, and E. Husebye, Determination of the three-dimensional seismic structure of the lithosphere, *Geophys. J. R. astron. Soc.*, *82*, 277–296, 1977.

Albarède, F., and R. D. van der Hilst, New mantle convection model may reconcile conflicting evidence, *Eos Trans. AGU*, *80*, 535–539, 1999.

Anderson, D., Tomography, Gutenberg, and Freud: are color cross sections really Rorschach tests?, *Eos Trans. AGU 80*(46), Fall Meet. Suppl., F719, 1999.

Anderson, D., Top-down tectonics?, *Science*, *293*, 2016–2018, 2001.

Anderson, D. L., *Theory of the Earth*, Blackwell Scientific Publications, 1989.

Bassin, C., G. Laske, and G. Masters, The current limits of resolution for surface wave tomography in north america, *Eos Trans. AGU 81*(48), Fall Meet. Suppl., F897, 2000.

Batchelor, G., *An introduction to Fluid Dynamics*, Cambridge University Press, Great Britain, 1967.

Becker, T., and L. Boschi, A comparison of tomographic and geodynamic mantle models, *Geochemistry, Geophysics, Geosystems*, *3*, 2002.

Becker, T., S. Panasyuk, R. O'Connell, and C. Faccenna, The backward-bent Indonesia slab, *Eos Trans. AGU 80*(46), Fall Meet. Suppl., S18, 1999.

Bercovici, D., Y. Ricard, and M. A. Richards, The relation between mantle dynamics and plate tectonics: A primer, in *History and Dynamics of Plate Motion, Geophys. Monogr. Ser.,* vol. 121, edited by M. A. Richards, R. Gordon, and R. D. van der Hilst, pp. 5–46, AGU, Washington, D.C., 2000.

Bijwaard, H., and W. Spakman, Tomographic evidence for a narrow whole mantle plume below Iceland, *Earth Planet. Sci. Lett.*, *166*, 121–126, 1999.

Bijwaard, H., and W. Spakman, Nonlinear global P-wave tomography by iterated linearized inversion, *Geophys. J. Int.*, *141*, 71–82, 2000.

Bijwaard, H., W. Spakman, and E. R. Engdahl, Closing the gap between regional and global travel time tomography, *J. Geophys. Res.*, *103*, 30,055–30,078, 1998.

Bolton, H., Travel times and the structure of the mantle, Ph.D. thesis, University of California, San Diego, 1996.

Boschi, L., and A. M. Dziewonski, High- and low-resolution images of the Earth's mantle: Implications of different approaches to tomographic modeling, *J. Geophys. Res.*, *104*, 25,567–25,594, 1999.

Bunge, H., and S. Grand, Mesozoic plate-motion history below the northeast Pacific Ocean from seismic images of the subducted Farallon slab, *Science*, *280*, 91–95, 2000.

Bunge, H., M. Richards, C. Lithgow-Bertelloni, J. Baumgardner, S. Grand, and B. Romanowicz, Time scales and heterogeneous structure in geodynamic Earth models, *Science*, *280*, 91–95, 1998.

Bunge, H. P., M. A. Richard, and J. R. Baumgardner, Effect of depth-dependent viscosity on the planform of mantle convection, *Nature*, *379*, 436–438, 1996.

Campbell, A., et al., Chemistry of hot springs on the mid-atlantic ridge, *Nature*, *335*, 514–519, 1988.

Castle, J., K. C. Creager, J. P. Winchester, and R. D. van der Hilst, Shear wave speeds at the base of the mantle, *J. Geophys. Res.*, *105*, 21,543–21,557, 2000.

Choy, G. L., and P. G. Richards, Pulse distortion and Hilbert transformation in multiply reflected and refracted body waves, *Bull. Seismol. Soc. Am.*, *65*, 55–70, 1975.

Christensen, U., The influence of trench migration on slab penetration into the lower mantle, *Earth Planet. Sci. Lett.*, *140*, 27–39, 1996.

Conrad, C. P., and B. H. Hager, The thermal evolution of an earth with strong subduction zones, *Geophys. Res. Lett.*, *26*, 3041–3044, 1999.

Creager, K., Anisotropy of the inner core from differential travel times of the phases *PKP* and *PKIKP*, *Nature*, *356*, 309–314, 1992.

Dahlen, F., S. Hung, and G. Nolet, Fréchet kernels for finite-frequency traveltimes–I.Theory, *Geophys. J. Int.*, *141*, 157–174, 2000.

Davies, G., Penetration of plates and plumes through the mantle transition zone, *Earth Planet. Sci. Lett.*, *133*, 507–516, 1995.

Davies, G., and M. Richards, Mantle convection, *Journal of Geology*, *100*, 151–206, 1992.

Dziewonski, A., Mapping the lower mantle: determination of lateral heterogeneity up to degree and order 6, *J. Geophys. Res.*, *89*, 5929–5952, 1984.

Dziewonski, A. M., B. Hager, and R. J. O'Connell, Large scale heterogeneities in the lower mantle, *J. Geophys. Res.*, *82*, 239–255, 1977.

Dziewonski, A. M., L. Boschi, Y. Gu, G. Master, and G. Laske, Feasibility of construction of a 3-D reference Earth model, *Eos Trans. AGU 80*(46), Fall Meet. Suppl., F24, 1999.

Engdahl, E. R., R. D. van der Hilst, and R. Buland, Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seismol. Soc. Am.*, *88*, 722–743, 1998.

Faccenna, C., T. W. Becker, F. P. Lucente, L. Jolivet, and F. Rossetti, History of subduction and back-arc extension in the Central Mediterranean, *Geophys. J. Int.*, *145*, 809–820, 2001.

Fischer, K., T. Jordan, and K. Creager, Seismic constraints on the morphology of deep slabs, *J. Geophys. Res.*, *93*, 4773–4783, 1988.

Forsyth, D., and S. Uyeda, On the relative importance of driving forces of plate motion, *Geophys. J. Roy. astr. Soc*, *43*, 163–200, 1975.

Forte, A. M., and J. X. Mitrovica, Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data, *Nature*, *410*, 1049–1056, 2001.

Fukao, Y., M. Obayashi, H. Inoue, and M. Nenbai, Subducting slabs stagnant in the mantle transition zone, *J. Geophys. Res.*, *97*, 4809–4822, 1992.

Gaherty, J. B., and B. H. Hager, Compositional vs. thermal buoyancy and the evolution of subducted lithosphere, *Geophys. Res. Lett.*, *21*, 141–144, 1994.

Garland, M., and P. Heckbert, Surface simplification using quadric error metrics, *In SIGGRAPH97 proc.*, 1997.

Garnero, E. J., Heterogeneity of the lowermost mantle, *Annu. Rev. Earth Planet. Sci.*, *28*, 477–507, 2000.

Gilbert, F., and A. M. Dziewonski, An application of normal mode theory to the retrieval of structure parameters and source mechanisms from seismic spectra, *Philos. Trans. R. Soc. London*, *278*, 187–269, 1975.

Goes, S., W. Spakman, and H. Bijwaard, A lower mantle source for central european volcanism, *Science*, *286*, 1928–1931, 1999.

Gorbatov, A., Y. Fukao, and S. Widiyantoro, Application of a three-dimensional ray-tracing technique to global P, PP and Pdiff traveltime tomography, *Geophys. J. Int.*, pp. 583–593, 2001.

Grand, S. P., Mantle shear structure beneath the Americas and surrounding oceans, *J. Geophys. Res.*, *99*, 11,591–11,621, 1994.

Grand, S. P., R. D. van der Hilst, and S. Widiyantoro, Global seismic tomography: A snapshot of convection in the earth, *GSA Today*, *7*, 1–7, 1997.

Griffiths, R. W., R. Hackney, and R. D. van der Hilst, A laboratory investigation of effects of trench migration on the descent of subducted slabs, *Earth Planet. Sci. Lett.*, *133*, 1–17, 1995.

Gripp, A., and R. Gordon, Current plate velocities relative to the hotspots incorporating the NUVEL-1 global plate motion model, *Geophys. Res. Lett.*, *17*, 1,109–1,112, 1990.

Gudmundsson, O., and M. Sambridge, A regionalized upper mantle (RUM) seismic model, *J. Geophys. Res.*, *103*, 7121–7130, 1998.

Gudmundsson, O., J. H. Davies, and R. W. Clayton, Stochastic analysis of global traveltime data: Mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, *102*, 25–43, 1990.

Guillou-Frottier, L., J. Buttles, and P. Olson, Laboratory experiments of the structure of subducted lithosphere, *Earth Planet. Sci. Lett.*, *133*, 19–34, 1995.

Gurnis, M., and G. Davies, Numerical study of high rayleigh number convection in a medium with depth-dependent viscosity, *Geophys. J. Roy. astr. Soc*, *85*, 523–521, 1986.

Gurnis, M., and B. Hager, Controls of the structure of subducted slabs, *Nature*, *335*, 317–321, 1988.

Gurnis, M., J. Ritsema, H. van Heijst, and S. Zhong, Tonga slab deformation: The influence of a lower mantle upwelling on a slab in a young subduction zone, *Geophys. Res. Lett.*, *27*, 2,373–2,376, 2000.

Hafkenscheid, E., S. Buiter, M. Wortel, W. Spakman, and H. Bijwaard, Modelling the seismic velocity structure beneath Indonesia: a comparison with tomography, *Tectonophysics*, *333*, 35–46, 2001.

Hager, B., Subducted slabs and the geoid: Constraints on mantle rheology and flow, *J. Geophys. Res.*, *86*, 6,003–6,015, 1984.

Hager, B., and R. Clayton, Constraints on the structure of mantle convection using seismic observations, flow models, and the geoid, in *Mantle Convection, Plate Tectonics and Global Dynamics*, edited by W. Peltier, Gordon and Breach Science Publishers, 1989.

Hager, B., and R. O'Connell, Kinematic models of large-scale flow in the earth's mantle, *J. Geophys. Res.*, *84*, 1,031–1,048, 1979.

Hager, B., and R. O'Connell, A simple model of plate dynamics and mantle convection, *J. Geophys. Res.*, *86*. 4,843–4,867, 1981.

Hager, B., R. Clayton, M. Richards, P.Comer, and A. Dziewonski, Lower mantle heterogeneity, dynamic topography and the geoid, *Nature*, *313*, 541–545, 1985.

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

Humphreys, E., and R. Clayton, Adaption of backprojection tomography to seismic travel time problems, *J. Geophys. Res.*, *93*, 1073–1085, 1988.

Hung, S., F. Dahlen, and G. Nolet, Fréchet kernels for finite-frequency traveltimes–II. Examples, *Geophys. J. Int.*, *141*, 175–203, 2000.

Inoue, H., Y. Fukao, K. Tanabe, and Y. Ogata, Whole mantle P-wave travel time tomography, *Phys. Earth Planet. Inter.*, *59*, 294–328, 1990.

Isacks, B., and P. Molnar, Distribution of stresses in the descending lithosphere from a global survey of focal-mechanism solutions of mantle earthquakes, *Rev. Geophys. Space. Phys.*, *9*, 103–174, 1971.

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

Jarrard, R. D., Relations among subduction parameters, *Rev. of Geophys.*, *24*, 217–284, 1986.

Jeffreys, H., and E. R. Lapwood, The reflection of a pulse within a sphere, *Proc. R. Soc. London*, *241*, 455–479, 1957.

Jellinek, A. M., R. C. Kerr, and R. W. Griffiths, Mixing and compositional stratification produced by natural convection, 1. Experiments and their application to Earth's core and mantle, *J. Geophys. Res.*, *104*, 7,183–7,201, 1999.

Julian, B. R., and M. Sengupta, Seismic travel time evidence for lateral inhomogeneity in the deep mantle, *Nature*, *242*, 443–447, 1993.

Kárason, H., and R. D. van der Hilst, Improving seismic models of whole mantle P wave speed by the inclusion of data from differential times and normal modes, *Eos Trans. AGU 79*(45), Fall Meet. Suppl., F656, 1998.

Kárason, H., and R. D. van der Hilst, New constraints on 3d variations in mantle P wavespeed, *Eos Trans. AGU 80*(46), Fall Meet. Suppl., F731, 1999.

Kárason, H., and R. D. van der Hilst, Constraints on mantle convection from seismic tomography, in *History and Dynamics of Plate Motion, Geophys. Monogr. Ser.,* vol. 121, edited by M. A. Richards, R. Gordon, and R. D. van der Hilst, pp. 277–288, AGU, Washington, D.C., 2000.

Kárason, H., and R. D. van der Hilst, Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP,P_{diff}), *J. Geophys. Res.*, *106*, 6,569–6,588, 2001.

Kárason, H., and R. D. van der Hilst, Mantle P-wave speed from seismic tomography; advances in methodology and data integration, *in prep.*, 2002.

Kárason, H., R. D. van der Hilst, and G. Masters, Three-dimensional mantle structure from low frequency PP-P differential travel times, *subm. to Geophys. Res. Lett.*, 2002.

Kellogg, L. H., and S. D. King, Effect of mantle plumes on the growth of D" by reaction between the core and mantle, *Geophys. Res. Lett.*, *20*, 379–382, 1993.

Kellogg, L. H., B. H. Hager, and R. D. van der Hilst, Compositional stratification in the deep mantle, *Science*, *283*, 1881–1884, 1999.

Kennett, B. L. N., E. Engdahl, and R. Buland, Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, *122*, 108–124, 1995.

Kennett, B. L. N., S. Widiyantoro, and R. D. van der Hilst, Joint seismic tomography for bulk sound and shear wave speed in the Earth's mantle, *J. Geophys. Res.*, *103*, 12,469–12,493, 1998.

Kesson, S., J. F. Gerald, and J. Shelley, Mineralogy and dynamics of a pyrolite lower mantle, *Nature*, *397*, 252–255, 1998.

Kincaid, C., and P. Olson, An experimental study of subduction and slab migration, *J. Geophys. Res.*, *92*, 13,832–13,840, 1987.

King, S., and B. Hager, Subducted slabs and the geoid 1. numerical experiments with temperature-dependent viscosity, *J. Geophys. Res.*, *99*, 19,843–19,852, 1994.

King, S., A. Raefsky, and B. Hager, Vectorizing a finite element code for incompressive two-dimensional convection in the earth's mantle, *Phys. Earth Planet. Inter.*, *59*, 196–208, 1990.

King, S. D., Subduction zones: observations and geodynamic models, in *Processes and Consequences of deep subduction*, edited by D. Rubie and R. D. van der Hilst, Elseview, 2001.

Kohler, M., and T. Tanimoto, One-layer global inversion for outermost core velocity, *Phys. Earth Planet. Inter.*, *72*, 173–184, 1992.

Kravtsov, Y. A., and Y. I. Orlov, *Geometrical Optics of Inhomogeneous Media*, Springer-Verlag, New York, 1990.

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

Li, C. D., and B. Romanowicz, Global mantle shear velocity model developed using non-linear asymptotic coupling theory, *J. Geophys. Res.*, *101*, 22,245–22,272, 1996.

Lithgow-Bertelloni, C., and M. Richards, The dynamics of cenozoic and mesozoic plate motions, *Rev. Geophys.*, *36*, 27–78, 1998.

Liu, X.-F., and A. M. Dziewonski, Lowermost mantle shear wave velocity structure, *Eos Trans. AGU 75*Fall Meet. Suppl., T42F-06, 1994.

Liu, X. F., and A. M. Dziewonski, Global analysis of shear wave velocity anomalies in the lower-most mantle, in *The Core-Mantle Boundary Region,* Geodyn. ser. vol. 28, edited by M. Gurnis et al., pp. 21–36, AGU, Washington, D.C., 1998.

Liu, X. F., J. Tromp, and A. M. Dziewonski, Is there a first order discontinuity in the lowermost mantle?, *Earth Planet. Sci. Lett.*, *160*, 343–354, 1998.

Manga, M., Interactions between mantle diapirs, *Geophys. Res. Lett.*, *24*, 1,871–1,874, 1997.

Manga, M., and H. A. Stone, Buoyancy-driven interactions between deformable drops at low reynolds numbers, *J. Fluid Mech.*, *256*, 647–683, 1993.

Marquering, H., G. Nolet, and F. A. Dahlen, Three–dimensional waveform sensitivity kernels, *Geophys. J. Int.*, *132*, 521–534, 1998.

Marquering, H., F. A. Dahlen, and G. Nolet, Three–dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, *137*, 805–815, 1999.

Masters, G., G. Laske, and H. Bolton, A shear-velocity model of the mantle, *Philos. Trans. R. Soc. London, Ser. A, 354*, 1385–1411, 1996.

Masters, G., G. Laske, H. Bolton, and A. Dziewonski, The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale, Geophys. Monogr. Ser.,* vol. 117, edited by S.-I. Karato, A. M. Forte, R. C. Liebermann, G. Masters, and L. Stixrude, pp. 63–87, AGU, Washington, D.C., 2000.

McSweeney, T. J., Seismic constraints on core structure and dynamics, Ph.D. thesis, Univ. of Wash., Seattle, 1995.

Mégnin, C., and B. Romanowicz, The three-dimensional shear velocity structure of the mantle from the inversion of body, surface, and higher-mode waveforms, *Geophys. J. Int., 143*, 709–728, 2000.

Montagner, J. P., Can seismology tell us anything about convection in the mantle?, *Rev. Geophys., 32*, 115–137, 1994.

Montagner, J. P., Where can seismic anisotropy be detected in Earth's mantle? In boundary layers ...., *Pure Appl. Geophys., 151*, 223–156, 1998.

Moresi, L., and M. Gurnis, Constraints of the lateral strength of slabs from three-dimensional dynamic flow models, *Phys. Earth Planet. Inter., 138*, 15–28, 1995.

Moresi, L., M. Gurnis, and S. Zhong, Plate tectonics and convection in the earth's mantle: Toward a numerical simulation, *Comp. in Sci. and Eng.*, pp. 22–33, 2000.

Nataf, H.-C., What is responsible for thermal coupling in layered convection?, *J. Phys - Paris, 49*, 1707–1714, 1988.

Neele, F., and R. Snieder, Topography of the 400-km discontinuity from observations of long-period P400P phases, *Geophys. J. Int.*, *109*, 670–682, 1992.

Nolet, G., Solving or resolving inadequate and noisy tomographic systems, *J. Comput. Phys*, *61*, 463–468, 1985.

Nolet, G., Seismic wave propation and seismic tomography, in *Seismic Tomography, With Application, in Global Seismology and Exploration Geophysics*, edited by G. Nolet, pp. 1–27, D. Reidel, Norwell, Mass., 1987.

Nolet, G., R. Montelli, and J. Virieux, Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, *138*, 36–44, 1999.

Obayashi, M., and Y. Fukao, *P* and *PcP* travel time tomography for the core-mantle boundary, *J. Geophys. Res.*, *102*, 17,825–17,841, 1997.

Paige, C. C., and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *Trans. Math. Software*, *8*, 43–71, 1982.

Pulliam, R., and P. Stark, Bumps on the core-mantle boundary: Are they facts or artifacts?, *J. Geophys. Res.*, *98*, 1943–1955, 1993.

Pulliam, R. J., D. W. Vasco, and L. R. Johnson, Tomographic inversions for mantle *P* wave velocity structure based on the minimization of l2 and l1 norms of international seismological center travel time residuals, *J. Geophys. Res.*, *98*, 699–734, 1993.

Puster, P., and T. H. Jordan, How stratified is mantle convection?, *J. Geophys. Res.*, *102*, 7625–7646, 1997.

Replumaz, A., and P. Tapponnier, Reconstruction of the deformed zone between India and Asia by backward motion of lithospheric blocks, *submitted to J. Geophys. Res.*, 2001.

Replumaz, A., H. Kárason, P. Tapponnier, and R. D. van der Hilst, Comparison of plate reconstructions and seismic tomography, *In prep.*, 2002.

Ricard, Y., M. Richards, C. Lithgow-Bertelloni, and Y. Stunff, A geodynamic model of mantle density heterogeneity, *J. Geophys. Res.*, *98*, 21,895–21,909, 1993.

Richards, M., Hotspots and the case for a high-viscosity lower mantle, in *Glacial Isostasy, Sea Level and Mantle Rheology*, edited by R. Sabadini and K. Lambeck, pp. 571–588, Kluwer Acad., Norwell, Mass, 1991.

Richards, M. A., and D. Engebretson, Large-scale mantle convection and the history of subduction, *Nature*, *355*, 437–440, 1992.

Richards, M. A., H.-P. Bunge, and C. Lithgow-Bertelloni, Mantle convection and plate motion history: Toward general circulation models, in *History and Dynamics of Plate Motion, Geophys. Monogr. Ser.,* vol. 121, edited by M. A. Richards, R. Gordon, and R. D. van der Hilst, pp. 289–307, AGU, Washington, D.C., 2000.

Riedel, M. R., and S. Karato, Grain-size evolution in subducted oceanic lithosphere associated with the olivine-spinel transformation and its effects on rheology, *Earth Planet. Sci. Lett.*, *148*, 27–43, 1997.

Ristow, G. H., Wall correction for sinking cylinders in fluids, *Physical Review E*, *55*, 2808–2813, 1997.

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

Ritsema, J., S. Ni, D. V. Helmberger, and H. P. Crotwell, Anomalous shear velocity reductions and gradients in the lower mantle beneath Africa, *Geophys. Res. Lett.*, *25*, 4245–4248, 1998.

Ritsema, J., H. van Heijst, and J. Woodhouse, Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, *286*, 1925–1928, 1999.

Ritzwoller, M., and E. Lavely, 3-dimensional models of earth's mantle, *Rev. Geophys.*, *33*, 1–66, 1995.

Robertson, G. S., and J. H. Woodhouse, Evidence for proportionality of p and s heterogeneity in the lower mantle, *Geophys. J. Int.*, *123*, 85–116, 1995.

Rogers, R., H. Kárason, and R. D. van der Hilst, Three-phase tectonic model for Miocene-Recent subduction history of northern Central America, *submitted to Geology*, 2001.

Romanowicz, B., Seismic tomography of earth's mantle, *Ann. Rev. Earth Planet. Sci.*, *19*, 77–99, 1991.

Saltzer, R., R. van der Hilst, and H. Kárason, Comparing P and S wave heterogeneity in the mantle, *Geophys. Res. Lett.*, *28*, 1,335–1,338, 2001.

Sengupta, M. K., and M. Toksöz, Three-dimensional model of seismic velocity variations in the earth's mantle, *Geophys. Res. Lett.*, *3*, 84–86, 1976.

Shinbrot, M., *Lectures on Fluid Mechanics, Notes on Mathematics and its application*, Gordon and Breach, 174.

Silver, P. G., R. Carlson, and P. Olson, Deep slabs, geochemical heterogeneity, and the large-scale structure of mantle convection, *Ann. Rev. Earth Planet. Sci.*, *16*, 477–541, 1988.

Simons, M., and B. H. Hager, Localization of the gravity field and the signature of glacial rebound, *Nature*, *390*, 500–504, 1997.

Song, X., and D. V. Helmberger, Depth dependance of anistropy of earth's inner core, *J. Geophys. Res.*, *100*, 9805–9816, 1995.

Song, X., and D. V. Helmberger, *PKP* differential travel times: Implications for 3-d lower mantle structure, *Geophys. Res. Lett.*, *24*, 1863–1866, 1997.

Spakman, W., and H. Bijwaard, Optimization of cell parameterizations for tomographic inverse problems, *Pure Appl. Geophys.*, *158*, 1401–1423, 2001.

Spakman, W., and G. Nolet, Imaging algorithms, accuracy and resolution in delay time tomography, in *Mathematical Geophysics: A Survey of Recent Developments in Seismology and Geodynamics*, edited by N. J. Vlaar, pp. 155–188, D. Reidel, Norwell, Mass., 1988.

Spakman, W., S. van der Lee, and R. D. van der Hilst, Travel time tomography of the European Mediterranean mantle down to 1400 km, *Phys. Earth Planet. Inter.*, *79*, 3–74, 1993.

Stein, C. A., and S. Stein, A model for the global variation in oceanic depth and heat flow with lithospheric age, *Nature*, *359*, 123–129, 1992.

Stevenson, D. J., Limits on lateral density and velocity variations in the Earth's outer core, *Geophys. J. R. Astron. Soc.*, *88*, 311–319, 1987.

Su, W. J., and A. M. Dziewonski, On the scale of mantle heteroeneity, *Phys. Earth Planet. Inter.*, *74*, 29–54, 1992.

Su, W. J., and A. M. Dziewonski, Inner core anisotropy in three dimensions, *J. Geophys. Res.*, *100*, 9831–9852, 1995.

Su, W. J., and A. M. Dziewonski, Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth Planet. Inter.*, *100*, 135–156, 1997.

Su, W. J., R. L. Woodward, and A. M. Dziewonski, Deep origin of mid-oceanic ridge velocity anomalies, *Nature*, *359*, 149–152, 1992.

Su, W. J., R. Woodward, and A. M. Dziewonski, Degree-12 model of shear velocity heterogeneity in the mantle, *J. Geophys. Res.*, *99*, 6945–6980, 1994.

Tackley, P. J., Mantle convection and plate tectonics: Toward an integrated physical and chemical theory, *Science*, *288*, 2,002–2,007, 2000.

Tao, W., and R. O'Connell, Ablative subduction: A two-sided alternative to the conventional subduction model, *J. Geophys. Res.*, *97*, 8,877–8,904, 1992.

Tao, W., and R. O'Connell, Deformation of a weak subducted slab and variation of seismicity with depth, *Nature*, *361*, 626–628, 1993.

Čížková, H., O. Čadek, A. van den Berg, and N. Vlaar, Can lower mantle slab-like anomalies be explained by thermal coupling between the upper and lower mantles?, *Geophys. Res. Lett.*, pp. 1501–1504, 1999.

van der Hilst, R., and T. Seno, Effects of relative plate motion on the deep structure and penetration depth of slabs below the Izu-Bonin and Mariana island arcs, *Earth Planet. Sci. Lett.*, *120*, 395–407, 1993.

van der Hilst, R., E. Engdahl, W. Spakman, and G. Nolet, Tomographic imaging of subducted lithosphere below northwest Pacific island arcs, *Nature*, *353*, 37–43, 1991.

van der Hilst, R. D., Tomography with P, PP, and pP delay-time data and the three-dimensional mantle structure below the Caribbean region, Ph.D. thesis, Thesis Utrecht University, 1990.

van der Hilst, R. D., Complex morphology of subducted lithosphere in the mantle beneath the Tonga trench, *Nature*, *374*, 154–157, 1995.

van der Hilst, R. D., and E. R. Engdahl, On ISC *PP* and *pP* data and their use in delay-time tomography of the Caribbean region, *Geophys. J. Int.*, *106*, 169–188, 1991.

van der Hilst, R. D., and H. Kárason, Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: Toward a hybrid convection model, *Science, 283*, 1885–1888, 1999.

van der Hilst, R. D., E. R. Engdahl, and W. Spakman, Tomographic inversion of *P*-data and *pP*-data for aspherical mantle structure below the northwest Pacific region, *Geophys. J. Int., 115*, 264–302, 1993.

van der Hilst, R. D., S. Widiyantoro, and E. Engdahl, Evidence for deep mantle circulation from global tomography, *Nature, 386*, 578–584, 1997.

van der Hilst, R. D., S. Widiyantoro, and K. C. T. J. McSweeney, Deep subduction and aspherical variations in P-wavespeed at the base of Earth's mantle, in *The Core-Mantle Boundary Region, Geodyn. Ser.,* vol. 28, edited by M. Gurnis et al., pp. 5–20, AGU, Washington, D.C., 1998.

van der Sluis, A., and H. A. van der Vorst, Numerical solution of large sparse linear systems arising from tomographic problems, in *Seismic Tomography, With Application in Global Seismology and Exploration Geophysics*, edited by G. Nolet, pp. 53–87, D. Reidel, Norwell, Mass., 1987.

Van der Voo, R., W. Spakman, and H. Bijwaard, Tethyan subducted slabs under India, *Earth Planet. Sci. Lett., 171*, 7–20, 1999.

van Keken, P. E., and C. W. Gable, The interaction of a plume with a rheological boundary: A comparison between two- and three-dimensional models, *J. Geophys. Res., 100*, 20,291–20,302, 1995.

Vasco, D. W., and L. Johnson, Whole earth structure estimated from seismic arrival times, *J. Geophys. Res., 103*, 2633–2671, 1998.

Vasco, D. W., R. J. Pulliam, and L. R. Johnson, Formal inversion of ISC arrival times for mantle P-velocity structure, *Geophys. J. Int.*, *113*, 586–606, 1993.

Vasco, D. W., L. R. Johnson, R. J. Pulliam, and P. S. Earle, Robust inversion of iasp91 travel time residuals for mantle P and S velocity structure, earthquake mislocations, and station corrections, *J. Geophys. Res.*, *99*, 13,727–13,755, 1994.

Vasco, D. W., L. R. Johnson, and J. Pulliam, Lateral variations in mantle velocity structure and discontinuities determined from *P*, *PP*, *S*, *SS*, and *SS-SdS* travel time residuals, *J. Geophys. Res.*, *100*, 24,037–24,059, 1995a.

Vasco, D. W., J. E. Peterson, and E. L. Majer, Beyond ray tomography: Wavepaths and fresnel volumes, *Geophysics*, *60*, 1790–1804, 1995b.

Vasco, D. W., L. R. Johnson, and O. Marques, Global earth structure: inference and assassment, *Geophys. J. Int.*, *137*, 381–407, 1999.

Vassiliou, M., B. Hager, and A. Raefsky, The distribution of earthquakes with depth and stress in subducting slabs, *J. Geodyn.*, *1*, 11–28, 1984.

Walcott, R. I., Structure of the earth from glacio-isostatic rebound, *Annu. Rev. Earth Planet. Sci.*, *1*, 15–37, 1973.

Wen, L. X., and D. L. Anderson, The fate of slabs inferred from seismic tomography and 130 million years of subduction, *Earth Planet. Sci. Lett.*, *133*, 185–198, 1995.

Widiyantoro, S., Studies of seismic tomography on regional and global scale, Ph.D. thesis, 256 pp., Aust. Nat. Univ., Canberrra, 1997.

Widiyantoro, S., and R. D. van der Hilst, Structure and evolution of lithospheric slab beneath the Sunda arc, Indonesia, *Science*, *271*, 1566–1570, 1996.

Widiyantoro, S., B. Kennett, and R. D. van der Hilst, Extending shear-wave tomography for the lower mantle using S and SKS arrival-time data, *Earth, Planets, and Space*, *50*, 999–1012, 1998.

Widiyantoro, S., B. Kennett, and R. D. van der Hilst, Seismic tomography with P and S data reveals lateral variations in the rigidity of slabs, *Earth Planet. Sci. Lett.*, *173*, 91–100, 1999.

Wolfe, C., I. Bjarnason, J. VanDecar, and S. Solomon, Seismic structure of the Iceland mantle plume, *Nature*, *385*, 245–247, 1997.

Woodhouse, J. H., and A. Dziewonski, Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, *89*, 5953–5986, 1994.

Woodhouse, J. H., and A. M. Dziewonski, Seismic modeling of the Earth's large-scale 3-dimensional structure, *Philos. Trans. R. Soc. London*, *328*, 291–308, 1989.

Woodward, R. L., and G. Masters, Global upper mantle structure from long-period differential travel times, *J. Geophys. Res.*, *96*, 6351–6377, 1991.

Wysession, M. E., Large-scale structure at the core-mantle boundary from diffracted waves, *Nature*, *382*, 244–248, 1996.

Wysession, M. E., R. W. Valenzuela, A. N. Zhu, and L. Barko, Investigating the base of the mantle using differential travel times, *Phys. Earth Planet. Inter.*, *92*, 67–84, 1995.

Zhao, L., and T. H. Jordan, Sensitivity of frequency-dependent traveltimes to laterally heterogenous, anisotropic earth structure, *Geophys. J. Int.*, *133*, 683–704, 1998.

Zhao, L., T. H. Jordan, and C. H. Chapman, Three-dimensional Frèchet differential kernels for the seismic delay times, *Geophys. J. Int.*, *141*, 558–576, 2000.

Zhong, S., and M. Gurnis, Mantle convection with plates and mobile, faulted plate margins, *Science*, *267*, 838–843, 1995.

Zhong, S., M. T. Zuber, L. Moresi, and M. Gurnis, Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection, *J. Geophys. Res.*, *105*, 11,063–11,082, 2000.

Zhou, H., and R. Clayton, P and S wave travel time inversions for subducting slabs under the island arcs of the northwest pacific, *J. Geophys. Res.*, *95*, 6829–6851, 1990.

Appendix A

Analytical solutions to Stokes equation

For constant viscosity and incompressible flow the governing equations for *Stokes* flow are:

$$\frac{\boldsymbol{\nabla} p}{\mu} = \nabla^2 \mathbf{u} \tag{A.1}$$

and

$$\boldsymbol{\nabla} \cdot \mathbf{u} = 0 \tag{A.2}$$

Where p the pressure field, μ the dynamic viscosity and $\mathbf{u}$ the velocity. Eq. A.1 is the *Stokes* equation and eq. A.2 is the conservation of mass for incompressible flow.

Tractable solutions of eq. A.1 only exist in the most trivial cases, but they can provide insight into the nature of *Stokes* flow. The most relevant one is the axisymmetric and irrotational flow resulting from a sphere moving in an infinite medium. The sphere can have different viscosity than the surrounding material and eq. A.1 is solved, both inside and outside the sphere assuming that surface tension is strong enough to retain its shape.

We solve the problem in polar coordinates with origin at the center of the sphere. If a is the radius of the sphere, $\mathbf{U}$ its velocity, $\boldsymbol{\tau}$ the tractions, and the $\mathbf{u}$ and $\bar{\mathbf{u}}$ the outer and inner

flow field solution (overstrike denoting the properties of the inner solution in general), the appropriate boundary conditions are [*Shinbrot*, 174]:

$$\begin{aligned} &\mathbf{u} \longrightarrow \mathbf{0} \text{ and } p \longrightarrow 0 \text{ as } |\mathbf{r}| \longrightarrow \infty \\ &\bar{\mathbf{u}}(0,\theta) \text{ is finite} \\ &\mathbf{n}\cdot\mathbf{u} = \mathbf{n}\cdot\bar{\mathbf{u}} = \mathbf{n}\cdot\mathbf{U} \text{ at } r = a \\ &\boldsymbol{\tau}(a,\theta) = \bar{\boldsymbol{\tau}}(a,\theta) \end{aligned} \tag{A.3}$$

The right hand side of eq. A.1 can be worked on a little further:

$$\frac{\boldsymbol{\nabla} p}{\mu} = \nabla^2 \mathbf{u} = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{u}) - \boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{u}) = -\boldsymbol{\nabla}\times\boldsymbol{\omega} \tag{A.4}$$

Using eq. A.2 and the definition of vorticity, $\boldsymbol{\omega} = \boldsymbol{\nabla}\times\boldsymbol{u}$. Taking the divergence and curl of both sides in eq. A.4 results in [*Batchelor*, 1967]:

$$\frac{\nabla^2 p}{\mu} = -\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\boldsymbol{\omega}) = 0 \tag{A.5}$$

$$\frac{\boldsymbol{\nabla}\times\boldsymbol{\nabla} p}{\mu} = -\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{\omega}) = -\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{\omega}) + \nabla^2\boldsymbol{\omega} = 0 \tag{A.6}$$

Because $\boldsymbol{\nabla}\cdot\boldsymbol{\omega} = \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\boldsymbol{u}) = 0$ we get that both p and $\boldsymbol{\omega}$ obey Laplace equation:

$$\nabla^2 p = 0 \text{ and } \nabla^2\boldsymbol{\omega} = 0 \tag{A.7}$$

The solution to Laplace equation are series of spherical harmonics functions but the boundary conditions, eq. A.3, eliminate all but one term for each of the solutions. The stream function Ψ has the following relationships with the velocity:

$$u_r = \frac{1}{r^2 \sin\theta}\frac{\partial \Psi}{\partial \theta} \quad \text{and} \quad u_\theta = -\frac{1}{r \sin\theta}\frac{\partial \Psi}{\partial r} \tag{A.8}$$

Using identities for the vorticity $\boldsymbol{\omega}$ along with, eq. A.8, *Stokes* equation, A.1 can be rewritten as [*Batchelor*, 1967]:

$$\frac{\partial^2 \Psi}{\partial r^2} + \frac{\sin\theta}{r^2}\frac{\partial}{\partial \theta}\left(\frac{1}{\sin\theta}\frac{\partial \Psi}{\partial \theta}\right) = \begin{cases} -\frac{C\mathbf{U}\sin^2\theta}{r} & : \text{ outside} \\ \frac{\bar{C}\mathbf{U}r^2\sin^2\theta}{2} & : \text{ inside} \end{cases} \tag{A.9}$$

By assuming $\Psi = Uf(r)\sin^2\theta$ and solving for $f(r)$ and matching the boundary conditions eq. A.9 is solved.

For this application the outer solution is more interesting and it is given by [*Batchelor*, 1967]:

$$\begin{aligned} f(r) &= a\frac{1}{4}\frac{2\mu + 3\bar{\mu}}{\mu + \bar{\mu}} r + a^3 \frac{\frac{1}{2} - \frac{1}{4}\frac{2\mu+3\bar{\mu}}{\mu+\bar{\mu}}}{r} \\ u_r &= \frac{2U\cos\theta}{r^2} f(r) \\ u_\theta &= -\frac{U\sin\theta}{r} f'(r) \end{aligned} \tag{A.10}$$

To determine the velocity of a sinking or rising sphere under we need to balance driving and resisting forces. The driving force is [*Batchelor*, 1967]:

$$\mathbf{F} = \Delta\rho \frac{4\pi a^3}{3}\mathbf{g} \tag{A.11}$$

Where $\Delta\rho$ is the density difference between the outer and inner medium. The resisting force is calculated by integrating the tractions, $\boldsymbol{\tau}$, over the surface of the sphere. By balancing these two forces we get an expression for the terminal velocity of the sphere, **U**:

$$\mathbf{U} = \frac{1}{3}\frac{\Delta\rho a^2}{\mu}\mathbf{g}\,\frac{\mu + \bar{\mu}}{\mu + \frac{3}{2}\bar{\mu}} = \frac{1}{4\pi a\mu}\mathbf{F}\,\frac{\mu + \bar{\mu}}{\mu + \frac{3}{2}\bar{\mu}} \tag{A.12}$$

The viscosity of the outer, μ, medium has the most effects on the velocity. The factor on the right in eq. A.12 goes from 1 for a inviscous sphere to 2/3 for a rigid sphere. If the viscosity is the same in both medium the factor is 4/5.