

Mantle variation within the Canadian Shield: Travel times from the portable broadband Archean-Proterozoic Transect 1989

Götz H. R. Bokelmann¹

Department of Geophysics, Stanford University, Stanford, California

Paul G. Silver

Department of Terrestrial Magnetism, Carnegie Institution of Washington, DC

Abstract. We report travel times from the Archean-Proterozoic Transect 1989. This type of data set recorded by a transect of portable broadband instruments allows us to make inferences about mantle structure in the region between the Wyoming Craton and the Superior Province of the Canadian Shield. With station separations of 50 to 100 km and frequencies up to 5 Hz the resolution of lateral changes is increased by nearly an order of magnitude over previous studies to a scale that allows us to study the relation between velocity variation in the continental upper mantle, surface geology, tectonic features and age provinces. Travel times of direct and later phases are obtained from waveform matching. The values are corrected for crustal contribution and inverted for the vertical path upper mantle delay δt_{UM} under each station as well as the azimuthal dependence of this quantity. The prominent feature in the upper mantle delays δt_{UM} is the variation by at least 1.5 s for S , much of which occurs over a narrow zone of just a few hundred kilometers width. This suggests a major lateral upper mantle transition which does not coincide with the surface geological edge of the Canadian Shield but is located within the shield. This same transition is also observed in shear wave splitting delay times. Surprisingly, however, the P delays do not exhibit a corresponding variation. We address this apparent contradiction and show how it may be explained in conjunction with anisotropy in the subcontinental lithosphere. A simplified thermal model of the lithospheric transition zone, in which temperature controls the degree of crystallographic alignment and thus seismic anisotropy, predicts this phenomenon.

1. Introduction

The North American continent is notable for the variety of its tectonic environments, from the stability of the ancient Canadian Shield to the active tectonics of the western margin. These tectonic differences are clearly manifest in the mantle as shown by velocity-depth profiles obtained in these two regions [Grand and Helmberger, 1984], as well as three-dimensional velocity models (e.g., Romanowicz, 1979; Woodhouse and Dziewonski, 1984; Grand, 1987, 1994; Zhang and Tanimoto, 1990; van der Lee, 1996; van der Lee and Nole, 1997). Grand [1994] showed that the transition

from the low velocities of the western United States to faster mantle occurs over a length scale of not more than 1000 km, with the strongest variation apparently occurring beneath the Great Plains. These results are intriguing in that they suggest a general relationship between surface geology and mantle structure. However, to fully understand this relationship requires significantly higher resolution, since geologic transitions can occur over 100 km or even less, such as the boundary between the Archean Western Superior Province of the Canadian Shield and the Proterozoic Trans-Hudson (Figure 1). Indeed, a transportable seismic network was deployed to study this boundary [Silver *et al.*, 1993]. It was primarily motivated by observed large differences in mantle anisotropy from shear wave splitting observations at two stations: RSON, in the Western Superior Province and RSSD, just west of the Trans-Hudson on the edge of the Wyoming Craton [Silver and Chan, 1988, 1991] (Figure 1). This suggested a significant change in mantle properties across a distance of 1000 km. To determine where this transition actually takes

¹Previously at Department of Geosciences, Princeton University, Princeton, New Jersey, and at Institut für Geophysik, Universität Bochum, Bochum, Germany.

Copyright 2000 by the American Geophysical Union.

Paper number 1999JB900387.
0148-0227/00/1999JB900387\$09.00

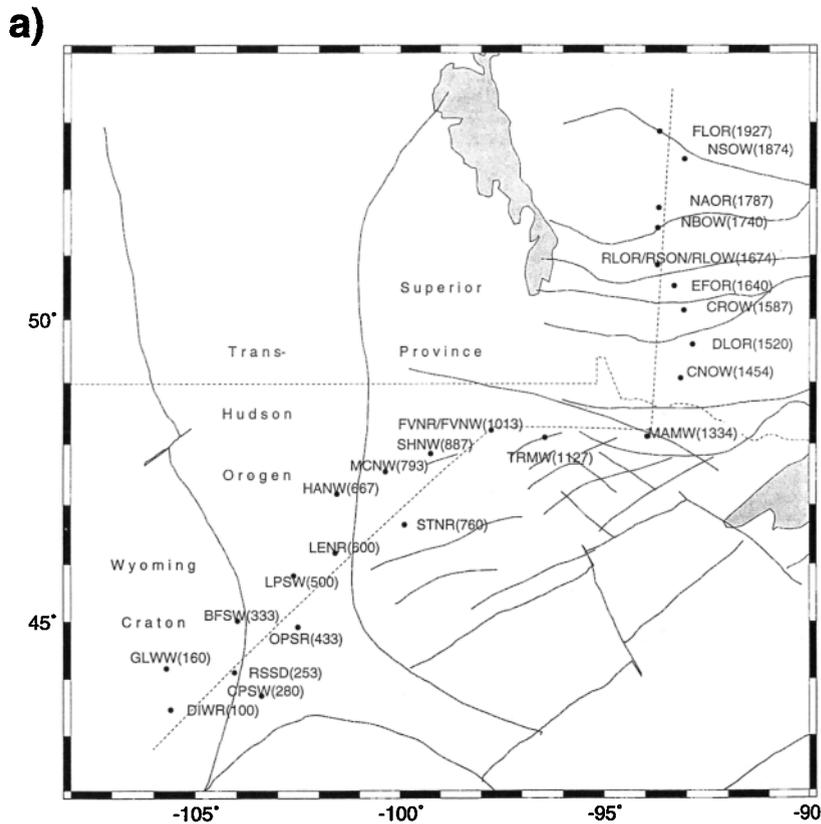


Figure 1a. Location of major geologic basement divisions in the central U.S./Canadian region (Superior Province, Wyoming Craton, and Trans-Hudson Orogen) and portable station locations during the APT89 experiment. Stations ending on R are equipped with 5-s instruments and those ending on W are equipped with 1-s instruments. RSON and RSSD are permanently installed stations. Projections onto the dotted transect line give distances from the western end used in subsequent figures.

place and what its relation to geologic boundaries is, the Archean-Proterozoic Transect (APT89) was conducted in 1989. As seen in Figure 1, the instrument spacing was between 50 and 100 km, so that the lateral resolution is increased by an order of magnitude over permanent station networks. Previous work on the data from this array [Silver and Kaneshima, 1993] demonstrated that there are indeed strong variations in mantle anisotropy, in particular, shear wave splitting, but that they are not clearly associated with the edge of the Superior Province. Instead, they appear to occur within the Canadian Shield.

In this paper, we analyze P and S travel time residuals from the APT89 experiment in conjunction with the shear wave splitting data. Our primary focus is the bulk upper mantle delay δt_{UM} . Variations of this delay across the transect will be used to constrain lateral variations in mantle velocity.

We use data from 43 events (Figure 2 and Table 1) for this purpose. As we will show, the travel times suggest a significant change in mantle structure within the Western Superior Province. An important aspect of our technique is a crustal correction based on P -to- S conver-

sions at crustal interfaces. Using this technique, we find that S wave delays show substantial variation within the mantle, whereas P delays hardly vary. To explain this effect, we study the effect of an anisotropic medium on "isotropic" P and S delays (i.e., delays measured under the assumption of an isotropic medium) and find that shear wave splitting has little effect on isotropic S delays given the set of ray geometries and initial polarizations in our study. Considering olivine, the mineral that dominates mantle anisotropy, in more detail we find that its effects on P delays can be rather large, however. We suggest a simple lithospheric model of coupled thermal and anisotropic variation which explains the observations.

2. Data

Figure 1a shows locations of 25 portable stations of the Archean-Proterozoic Transect Experiment 1989 (APT89) effectively connecting the Regional Seismic Test Network (RSTN) stations RSSD and RSON (at distances 253 and 1674 km along the transect). Also shown are boundaries between major geological units

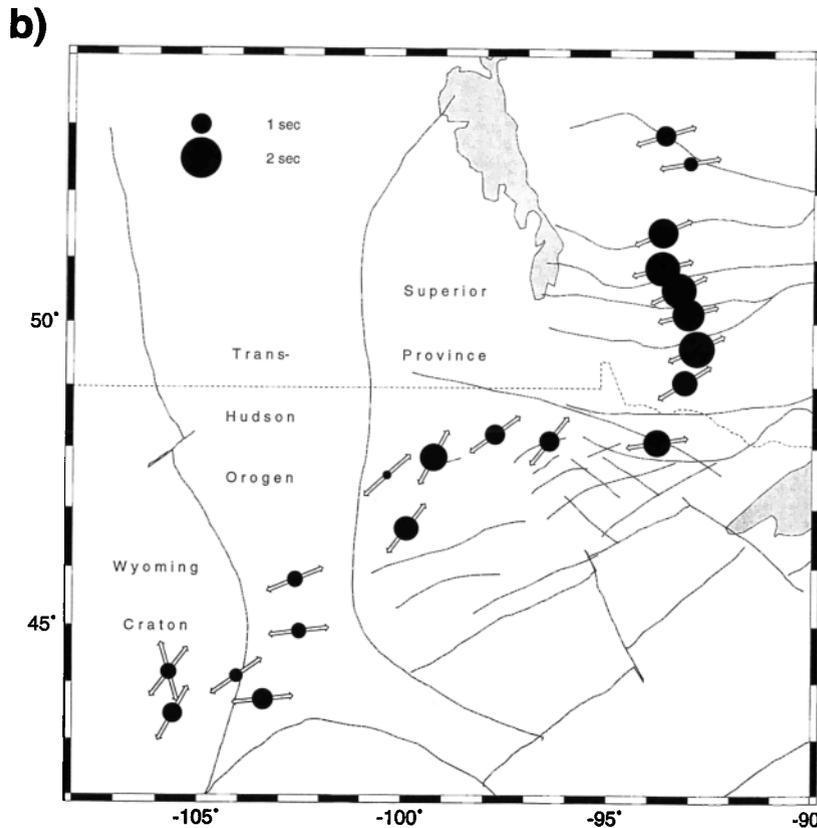


Figure 1b. Values of shear wave splitting recorded at the stations [taken from *Silver and Kaneshima, 1993*]. The time delay between the two shear wave phases is given by the size of the solid circle; polarization of the faster shear wave is given by the double arrow.

along the transect, namely, Superior Province (Archean), and Wyoming Craton (Archean) which are separated by the Trans-Hudson Orogen (Proterozoic). The APT89 experiment consisted of a 4-month deployment of portable broadband instruments between Wyoming Craton and Superior Province [*Silver et al., 1993*]. Two instrument types, 5-s (station names ending in R in Figure 1) and 1-s sensors (W) were used along the transect. Figure 1b shows results from a shear wave splitting study [*Silver and Kaneshima, 1993*] as polarization direction of the fast shear wave and the time delay between the two split shear waves. The latter reaches its maximum value of 1.8 s at the three collocated stations RSON, RLOR, and RLOW in the Superior Province.

From the intermediate- and short-period waveform data we selected 43 earthquakes for the travel time study (Table 1). Examples of these waveform data are given as seismogram sections in Figure 2, where for consistency, all traces simulate 5-s instruments. Both *P* waveforms in Figure 2a and *SKS* waveforms in Figure 2b are aligned to the predicted arrival times for the model resulting in this study (dotted line). Comparison of that model with arrival times through a laterally homogeneous Earth (crosses) shows that there are substantial lateral variations in travel time across the

transect, both for *P* and *SKS*, amounting to 1.1 s for *P* and 2.5 s for *SKS*.

The *P* waveform data are from event 89242:11:38:12.8 (yr day: hr: min: sec). They are used in the inversion for the laterally heterogeneous model. The *SKS* waveforms in Figure 2b were taken from a deep-focus event in the Solomon Islands (89233:18:25:40.7, latitude -4.09° , longitude 154.45°) and were not included in the data set used for the inversion. These independent data confirm the large size of the variation across the transect. In fact, the true variation may even be larger than the model due to damping in the inversion.

Figure 2 also shows the residual variation after correcting for crustal delays (thick lines). For Figure 2b, there is a clear residual change from the northeast of the transect (top of the section) to the central part. For *P* this transition appears to be missing. The explanation of this surprising discrepancy is a major focus of the later part of this paper.

The criterion for selecting events for the travel time study was the presence of a discrete *P* phase. Synthetic seismograms were then computed using the Complete Ordered Ray Expansion (CORE) synthetics approach of *Clarke and Silver [1991]*, with source wavelets estimated from *P* arrivals as a reference phase. While

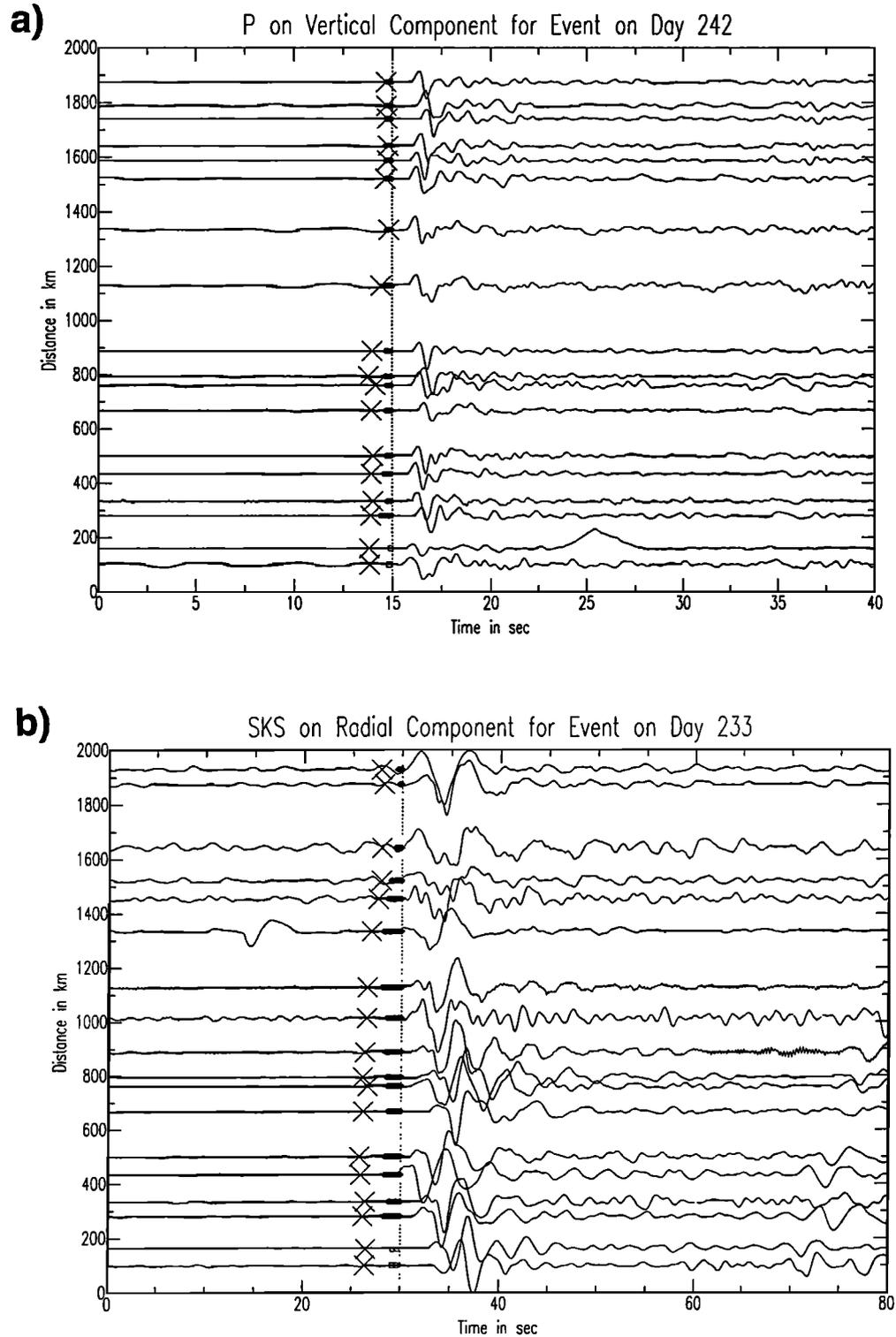


Figure 2. Waveform sections for (a) *P* and (b) *SKS* waves. Each section is aligned with respect to the resulting model of this study (dotted lines). Crosses give arrival times through the laterally homogeneous reference model. The difference between crosses and the dotted line indicates strong lateral variation across the transect. We also give the residual variation after correcting for crustal delays (thick lines; the two thick open lines are considered unreliable). Thus lateral variation in the mantle under the transect amounts to almost 2 s for *SKS*, but only a few tenths for *P*. Even though the inferred lateral variation is large, the *SKS* waveforms suggest that the true effect may be even larger due to damping of the inversion (see text).

Table 1: Events Used in This Study

Day	Date	Time, UT	Latitude	Longitude	Depth, km	m_s
165	June 14,1989	0035:58.3	51.55°	-174.32°	33.0	5.3
165	June 14,1989	1806:38.3	34.29°	26.05°	13.0	5.2
167	June 16,1989	0718:35.5	13.22°	145.15°	67.0	5.5
167	June 16,1989	1051:21.6	57.75°	-153.99°	58.0	5.8
167	June 16,1989	2342:35.1	31.80°	137.98°	359.0	5.9
168	June 17,1989	1457:48.4	-31.41°	-67.55°	27.0	5.3
168	June 17,1989	1828:08.6	-40.49°	-74.67°	33.0	5.6
169	June 18,1989	1406:28.8	17.76°	-68.81°	62.0	5.9
170	June 19,1989	1600:47.9	-22.11°	-67.56°	188.0	5.5
175	June 24,1989	1258:39.1	-28.34°	-66.31°	21.0	5.4
176	June 25,1989	1115:59.4	32.91°	-39.61°	10.0	5.2
176	June 25,1989	2037:32.5	1.13°	-79.62°	15.0	5.9
177	June 26,1989	0327:04.0	19.36°	-155.08°	9.0	5.8
177	June 26,1989	1038:39.5	39.11°	-28.24°	11.0	5.7
184	July 3,1989	1709:55.8	51.62°	-175.21°	33.0	5.7
189	July 8,1989	0931:57.1	52.84°	159.86°	30.0	5.5
201	July 20,1989	1209:53.3	-18.87°	-175.53°	240.0	5.4
205	July 24,1989	0327:48.8	36.08°	71.07°	95.0	5.8
218	Aug. 6,1989	0819:56.1	-23.16°	-68.32°	114.0	5.3
218	Aug. 6,1989	2253:56.5	42.80°	145.11°	44.0	5.7
220	Aug. 8,1989	2344:04.4	-22.72°	-68.48°	102.0	5.3
227	Aug. 15,1989	1004:22.3	-38.31°	-93.82°	10.0	5.4
232	Aug. 20,1989	1832:29.9	37.28°	21.20°	10.0	5.4
235	Aug. 23,1989	2025:22.4	52.35°	-168.02°	33.0	5.3
242	Aug. 30,1989	0306:55.1	54.61°	162.79°	31.0	5.5
242	Aug. 30,1989	1138:12.8	55.60°	161.36°	73.0	5.8
243	Aug. 31,1989	0817:22.5	-41.85°	-71.68°	154.0	5.4
243	Aug. 31,1989	1104:58.7	-0.17°	-17.80°	10.0	5.4
245	Sept 2,1989	1420:59.0	-17.82°	-178.55°	613.0	5.3
247	Sept 4,1989	1314:58.3	55.54°	-156.83°	11.0	6.5
252	Sept 9,1989	0140:35.8	2.43°	-79.76°	6.0	6.0
252	Sept 9,1989	1038:06.9	51.31°	-175.80°	33.0	5.3
258	Sept 15,1989	0948:09.2	51.57°	-173.37°	33.0	5.4
258	Sept 15,1989	1834:13.0	53.23°	159.72°	51.0	5.6
259	Sept 16,1989	0149:15.9	-0.59°	-77.47°	10.0	5.4
259	Sept 16,1989	0205:08.9	40.34°	51.53°	54.0	6.4
260	Sept 17,1989	0053:39.8	40.20°	51.75°	51.0	6.1
263	Sept 20,1989	1319:32.0	51.18°	178.82°	33.0	5.5
276	Oct. 3,1989	2133:34.8	-24.10°	-66.89°	154.0	5.4
277	Oct. 4,1989	1217:39.0	46.84°	153.96°	37.0	5.5
280	Oct. 7,1989	1548:29.1	51.31°	-179.03°	19.0	6.1
282	Oct. 9,1989	1003:19.5	-4.29°	-77.56°	35.0	5.4
282	Oct. 9,1989	1801:07.9	51.78°	171.87°	26.0	6.0

Locations and source depths are from PDE.

it may be difficult to determine onset times for later phases visually, the travel times can be more stably extracted using waveform fitting [Bokelmann and Silver, 1991b]. This requires finding an absolute travel time only for the reference phase. Since the reference phase is the first arriving phase for all events in this study, we can obtain absolute travel times in a stable manner. The waveform fitting is done using matched filtering

and a simple set of constraints based on the smoothness of differential travel time residuals across the spatially confined transect (Figure 1).

The model used in generating the synthetics was the Preliminary reference Earth model (PREM) [Dziewonski and Anderson, 1981] at 1 Hz, augmented to have a 40-km-thick continental crust and with the quality factor Q increased throughout the whole Earth model by an empirically determined factor of 1.7 to better match the observed waveforms. This is consistent with the frequency dependence of Q in Sipkin and Jordan [1979]. For each travel time a confidence interval is obtained from an F test similar to the procedure of estimating shear wave splitting parameters of Silver and Chan [1991]. Depending on the data quality, the a priori uncertainty is typically near ± 0.15 s for P and ± 0.45 s for S at the 95% confidence level. In a number of cases we obtained multiple “acceptable” travel time intervals. This ambiguity is usually removed by the smoothness constraint across the transect.

For this study we use 43 events on up to 26 three-component stations for which direct and later phases were extracted separately for the vertical, radial and transverse component. We used 590 P phases (P , pP and PcP) on the vertical component and 302 S phases (S , sS , ScS and $sScS$) on the transverse. However, data from the radial component are not used in order to reduce interference of converted phases near the receivers. For each event and phase we have a set of nearby ray paths to the stations along the transect. In order to characterize the size of the variations across the array we have defined for each event an array anomaly as the difference of maximum and minimum travel time residuals across the transect. This array anomaly is shown for each event in Figure 3. If there were no lateral variation of velocity along the paths (and no noise in the travel time data set), the symbols would shrink to zero. We see that this anomaly reaches maximum values of up to 4 s for P (Figure 3a) and 11 s for S (Figure 3b) in northern South America (see scale). The transect also shows large variation for events in Kasachstan, the Aleutians, and the Kuriles.

3. Inversion for Bulk Station Delays

We separate the total travel time residual into terms related to station corrections, event corrections, and contributions from the path, similar to the method of time-term analysis [Herrin et al., 1968; Dziewonski and Anderson, 1983]. For the k th phase we describe the residual recorded for the i th station and the j th event as

$$\delta t_{ijk} = \epsilon_{ijk} + a_{ijk} + b_{jk} + d_{ijk}, \quad (1)$$

where ϵ_{ijk} is a known ellipticity correction obtained in a procedure based on the work of Dziewonski and Gilbert [1976]. Equation (1) includes a phase-dependent

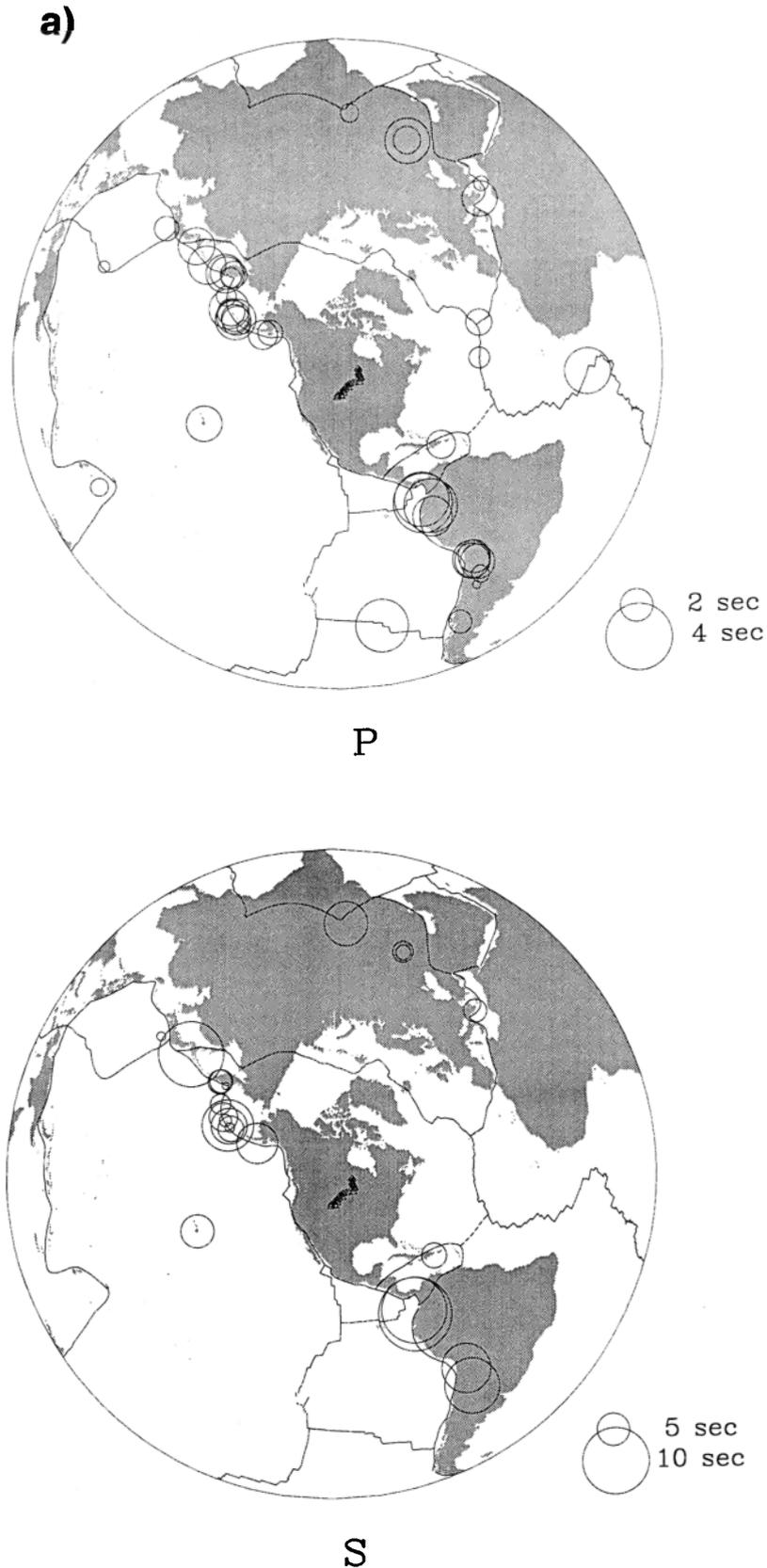


Figure 3. Magnitude of near-receiver delay variation across the transect obtained for the direct phases (a) *P* and (b) *S* shown at the respective event location. The size of the circle gives the maximum minus the minimum travel time residual observed among the stations. In this fashion the size of the circle directly gives an indication of the lateral heterogeneity in the “triangular” path region connecting the source with the transect stations. Also shown are the major oceanic plate boundaries. Note that this heterogeneity is particularly strong for paths from northern South America.

event term b_{jk} and a path term a_{ijk} . The latter may be used for large-scale tomography [Dziewonski *et al.*, 1977] or specified as a purely distance-dependent term $\sum a_i(\Delta_{ij})$ [Herrin *et al.*, 1968; Dziewonski and Anderson, 1983]. Since our experimental geometry produces steeply incident ray paths for teleseismic body waves, we choose the latter and implement a distance dependence by non overlapping distance segments. We focus on the station corrections d_{ijk} given as

$$d_{ijk} = c_{i0} \cos^{-1} \phi_{ijk} + \sum_{n=1}^D [c_{in} \cos(n\xi_{ij}) + s_{in} \sin(n\xi_{ij})] \quad (2)$$

with the back azimuth ξ and the near-receiver incidence angle ϕ_{ijk} .

This approach allows for variation of travel times with back azimuth ξ , which may be due to lateral heterogeneity and/or anisotropy in the near-receiver lithosphere. Truncating the series expansion in (2) represents an approximate treatment for anisotropy [see Sayers, 1994]; in practice, only low-order terms can be stably extracted from a set of noisy observations. This is also the motivation for using (2) in connection with S waves: first, we search for stable low-order azimuthal dependencies, which are later interpreted with the help of known initial shear wave polarizations.

The static terms c_{i0} may be regarded as the bulk travel time through the upper mantle, crust, and sediments. We account for nonvertical incidence by including the near-receiver incidence angle ϕ_{ijk} , thus removing a bias in station delays.

We utilize (1) as the basis for a least squares inversion for the station term d_{ijk} . The inversion procedure and error estimates are discussed in Appendix A. We consider various subsets of the data along with the full data set as a way of characterizing potentially contaminating effects of structure near the source or turning point of the array. The data set contains travel time residuals for P , pP , and PcP on the vertical component and S , sS , ScS and $sScS$ on the transverse component. From Figure 3 it is clear that most events occur near subduction zones. Several ray azimuths are near the strike direction of a near-source slab. Such near-source

heterogeneity, which may influence travel times [Sleep, 1973; Creager and Jordan, 1984], is not included in our model equations. To guard against influences from such heterogeneity, we generate a subset of the full data set (“all”) which contains only data with azimuths more than 20° off the estimated strike direction of the near-source slab. In this data set, potential contributions from the Caribbean lower mantle are also excluded, which have strong effects on certain paths from South to North America [e.g., Bokelmann and Silver, 1993; Grand, 1987; van der Hilst *et al.*, 1997]. This subset of the data is referred to as “allnosub” (see Table 2). Another subset, “dirall”, contains only observations of direct P and S phases for the purpose of studying potential errors in the later phase travel times. Finally, the data set “dirnosub” contains only direct phases and no potential contributions from subduction zones or the Caribbean lower mantle anomaly. This set is the smallest but probably the least contaminated by unwanted effects. These four data sets are inverted with station corrections depending on the back azimuth, with maximum order of either $D=0$ or $D=2$. A value of $D=0$ only allows for azimuthally independent station corrections. As shown in Appendix A, we utilized a damping parameter that simultaneously minimizes data misfit and model roughness. The optimum trade-off parameter is chosen by inspection of the trade-off curve, which was generated by performing a set of inversions for different values. Variance reductions for the optimum damping parameter under joint minimization of variance and model norm are typically in the range 86-97% for P and 55-75% for S , indicating that we are able to explain most of the variance in the data. Including azimuthal dependence up to $D=2$ increases the variance reduction of the optimal models for P by only $\sim 3\%$ but for S by up to 15%. The azimuthal effect is clearly negligible for P but may be important for S .

3.1. Total Near-Station Travel Time Delay Across the Transect

In this section we consider the near-station travel time delay, $\delta t_{\text{station}}$, resulting from the inversion (i.e.,

Table 2: Data Subsets

Dataset	Contained Events (N)	Contained Phases	Number of Data (P/S)
all	all (43)	$P, pP, PcP, S, sS, ScS, sScS$	578/302
allnosub	no known near-source heterogeneity nor lower mantle heterogeneity (30)	$P, pP, PcP, S, sS, ScS, sScS$	355/238
dirall	all (43)	P, S	458/192
dirnosub	no known near-source heterogeneity nor lower-mantle heterogeneity (30)	P, S	306/135

N is the number of events.

c_{i0} in equation (2)), which gives the travel time residual accumulated on a vertical path through upper mantle, crust, and sediments. Later we will consider the results after a crustal correction is applied. In (2), nonvertical incidence is taken into account through the predicted incidence angle taken at a depth of 150 km. The error from not taking curved near-receiver ray paths into account should be a few percent at most. To derive azimuthally independent station corrections $\delta t_{\text{station}} (=c_{i0})$, we use $D=0$ in (2), effectively averaging over azimuth. Inversions for the four data sets using several damping parameters were performed to insure that model features are stable with respect to variation in data and parameters. Figure 4a shows the station corrections, $\delta t_{\text{station}}$, determined for each of the data subsets and three damping parameters. For convenience, the delay is given as a function of distance along the transect instead of a two-dimensional map. In each of the panels the intermediate damping value (solid line) is the optimum one, whereas the other two values test the model dependence on γ (see Appendix A). Figure 4a (top) shows results for the full data set all, P in Figure 4a (left) and S in Figure 4a (right). The station delays for P vary across the transect by ~ 1.3 s, with the eastern part (larger distances) of the transect (Superior Province) appearing faster than the western part. The three curves are very similar with the inversion damped 2.5 times more (dashed line) than the optimal model (solid line) showing smaller values in the west. The variation of S delay across the transect is ~ 3 s, with ~ 2.5 s of the change being associated with a transitional zone between the fast eastern and slow western region. The P curve has ~ 1 s associated with a similar transition zone in the center of the transect.

According to the required stability criterion, the westernmost part of the transect (for stations GLWW and DIWR) is less resolved, since it shows variation depending on the data set; e.g., the result is different for the data sets all and allnosub. This shows a possible contamination of the data set in the westernmost region due to additional heterogeneity.

While the overall shapes of the S curves are very similar for the different data sets, the size of the total variation across the transect depends on the data set and is somewhat less for the smaller data sets (2 s instead of 3 s). The data set dirnosub consisting out of 306 and 135 data points for P and S gives roughly the same result for lower values of the damping parameter. For stronger damping, the overall variation of the S delay shrinks to ~ 2.3 s whereas the P effect is less influenced. This smaller effect as compared to data set all may either be due to (1) artifacts from known heterogeneity contaminating the data sets all and dirall or (2) an underconstrained problem for which the damping term in (A2) causes the model to be small.

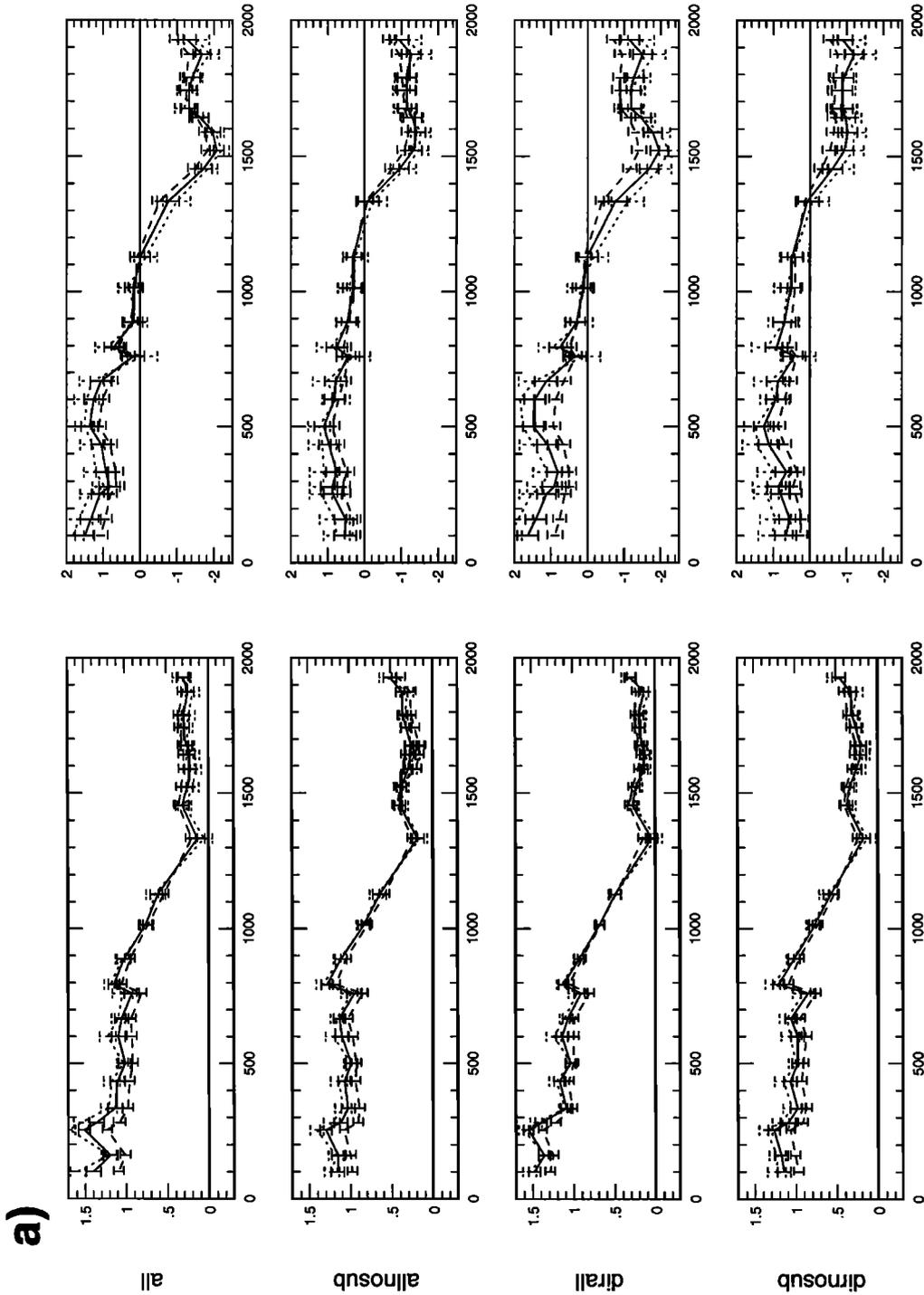
We tested the cause of the decrease via a synthetic example of a simple lithospheric transition in the mid-

dle of the transect. For this test we chose the same geometry and the same random error distribution as in our data set. That synthetic test, however, did not show the decrease for the smaller data sets. Hence we conclude that the larger data sets are indeed affected by additional effects, particularly the near-source or known lower mantle heterogeneity. Preferentially, the smaller data set allnosub and dirnosub should be used, which then leaves us with a smaller overall effect across the transect and earlier station delays for the two stations in the west.

The station corrections, $\delta t_{\text{station}}$, for allnosub and dirnosub are actually quite similar. For S the strongest variation is in the eastern part of the transect within the Superior Province north of station MAMW at 1334 km (Figure 1), while $\delta t_{\text{station}}$ for P changes most in the central region of the transect between 800 and 1300 km. Cleary and Hales [1966] estimated station corrections for six long-range seismic measurement (LRSM) stations that were located within 200 km of the APT89 transect. Their trend is very similar, with a maximum variation of 1.3 s for P , which is precisely the value obtained in this study. Another group of LRSM stations 100 to 600 km to the south still shows a similar trend but with more complexity. For S , literature values are slightly more diverse. The general west-slow to east-fast pattern is apparent in the work by Hales and Roberts [1970], Romanowicz [1979], Wickens and Buchbinder [1980], and Grand [1987, 1994]. However, there are some variations between the studies: e.g., Hales and Roberts [1970] show an anomalously fast station (RCD) in South Dakota, which is not confirmed here, nor by the other studies. Figure 4b shows results for similar inversions as in Figure 4a, but here the inversions were performed after crustal corrections had been applied for P and S . Interestingly, the residual upper mantle effect appears quite different for P and S . This will be extensively discussed below.

3.2. Inversion Parameters

In addition to the damping parameter γ , the inversion also depends on the correlation lengths $L_{c,s}$, which enforce a smoothness constraint of station correction parameters, and L_{path} , which enforces a distance dependent correlation (see Appendix A). We choose a value for $L_{c,s}$ of 200 km which is not overly restrictive and affords sufficient smoothing. For the path term a where $|\mathbf{r}' - \mathbf{r}|$ is defined as the difference in epicentral distance we choose a correlation length $L_{\text{path}} = 5^\circ$. This corresponds to an assumption that the medium has a vertical correlation length of a few hundred kilometers, which is not overly restrictive. The event locations in this study are from the Preliminary Determination of Epicenters (PDE) catalogue, namely, from relocation of global P wave travel times using the Jeffreys-Bullen Earth model. This is slower than PREM at teleseismic



Distance along the transect

Distance along the transect

Figure 4a. Station corrections δt_{stat} for the four data subsets (Table 2) shown as a function of distance along the transect (Figure 1) for (left) P and (right) S . The values are displayed for the optimum (γ_{opt}) and two further choices of the damping parameters with $\gamma = 0.4$ and $2.5 \cdot \gamma_{\text{opt}}$ (dotted line). A correlation length of 200 km for the station terms and 5° for the path terms is used. Results for the bulk near-receiver heterogeneity are shown.

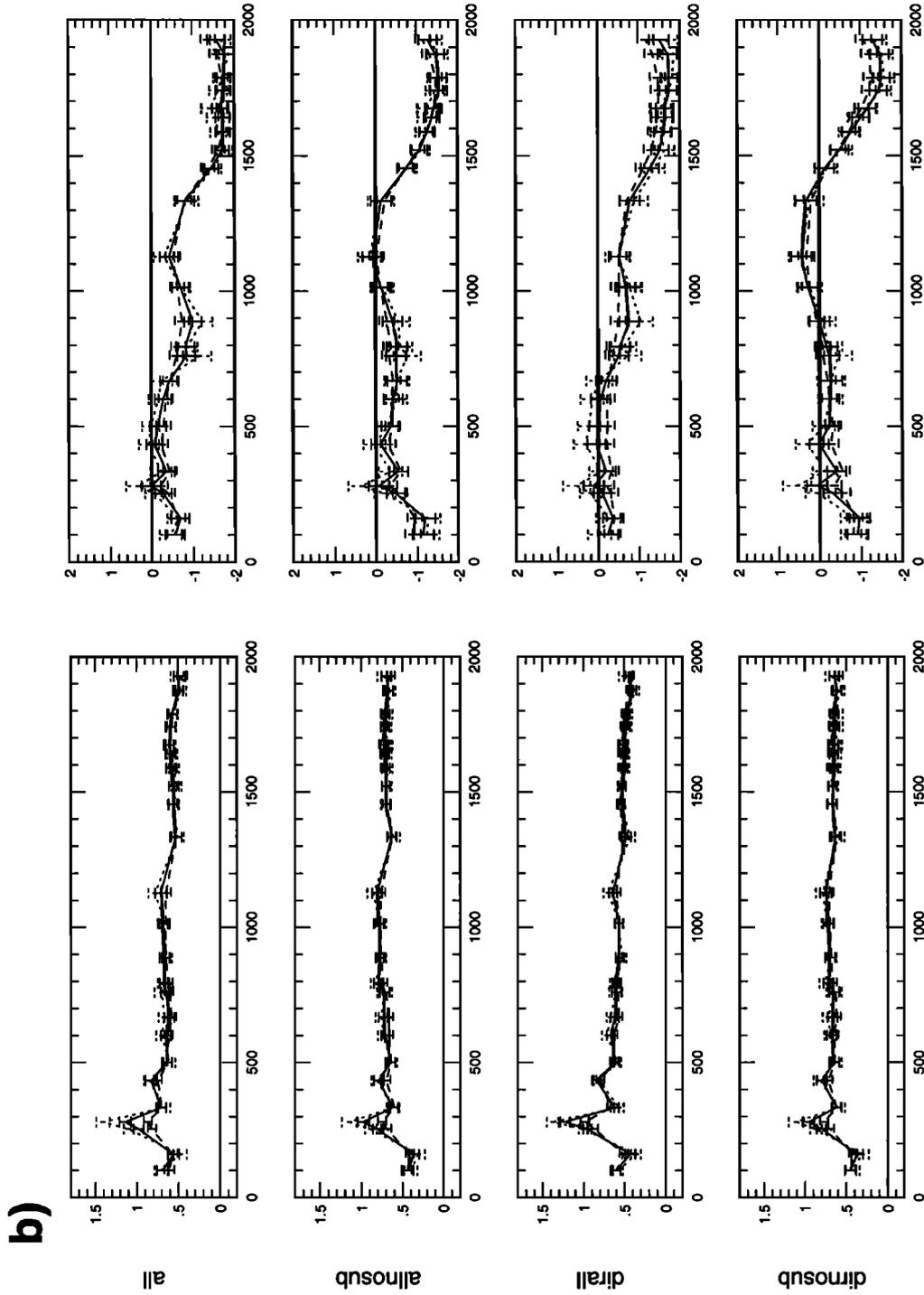


Figure 4b. Same as Figure 4a, except results are for the upper mantle alone (observed travel time residuals were crustal-corrected prior to the inversion). Interestingly, the variation of P delays in the mantle across the transect almost vanishes, while the S delays show a consistent decrease from large (slow) values in the center to small (fast) values in the northeast.

distances by a few seconds. Consequently, a consistent offset from zero for the P travel times is found in this study, which for small correlation lengths is correctly incorporated into the path term. For larger correlation lengths this offset is preferentially incorporated into the absolute level of $\delta t_{\text{station}}$ for P . The size of the variation of $\delta t_{\text{station}}$ across the transect is not influenced however. A systematic dependence of the offset with correlation length is not apparent for S . This is expected as no S travel times are used in locating the events. This average $\delta t_{\text{station}}$ for S is near zero for all correlation lengths. Clearly, a trade-off exists between path term and station delay offset. Our experimental geometry is not well-suited to determine the absolute station effect, but the variation of the station delay across the transect is well-determined.

4. Crust-Corrected Upper Mantle Delay Across the Transect

In order to focus on the mantle variations in travel time we must apply a crustal correction to the data. This requires knowledge of the total travel time delay through the crust, which is available either from previous experiments or directly from our data. Refraction profiles in the region [Braile *et al.*, 1989] and drilling [American Association of Petroleum Geologists (AAPG), unpublished basement map, 1967] suggest a cross section across the crust like the one shown in Figure 5 from which a general increase of sedimentary cover in southwestern direction is inferred, while the whole crust apparently thins to the northeast in this region.

While information from refraction data is available for correction purposes, it is much more desirable to use information which is taken directly from our data and is therefore more likely to be valid for the proper station locations. This technique which uses P -to- S conversions in our three-component data is described in detail in Appendix C.

Crustal correction applied to inversion results ($D=0$, no azimuthally dependent terms) produces delays arising from the near-receiver part of the mantle which we will call upper mantle delay δt_{UM} in the following. The study of near-surface structure shows that it may, in fact, have considerable effect on travel times. This effect can vary laterally and may therefore violate the smoothness assumption of c and s in (2). Since the inversion may have some smoothing effect itself, it is best to correct for crustal structure and subsequently to invert for the residual upper mantle structure directly. Using this procedure we obtain an upper mantle residual structure directly which better justifies an a priori smoothness assumption. This upper mantle delay across the transect is shown in Figure 4b for the different data and parameter sets. For S (Figure 4b, right), the dominant gradient between 1100 and 1700 km is present for all data and parameter sets. Results for data sets allnosub and dirnosub. In contrast, the more conservative data sets give a larger transition effect of 1.5 s, although we would expect a smaller effect for less constraining data. This suggests that the difference in size may be due to potential near-source

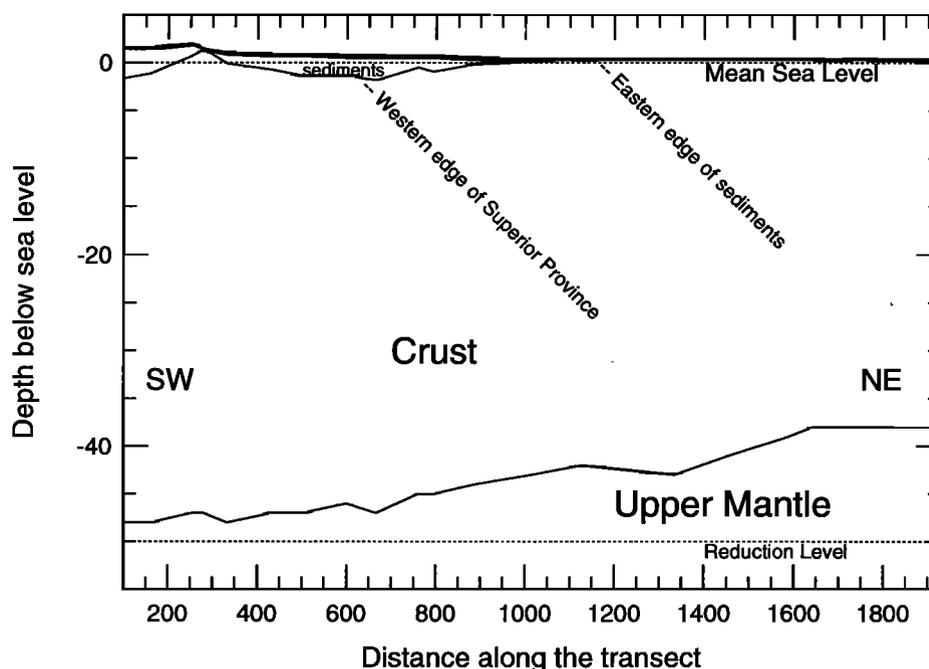


Figure 5. A model of the crust along the transect obtained from refraction seismology (after Braile *et al.* [1989] and drilling [AAPG, unpublished basement map, 1967]). For purposes of crustal correction we reduce to a level at 50 km depth.

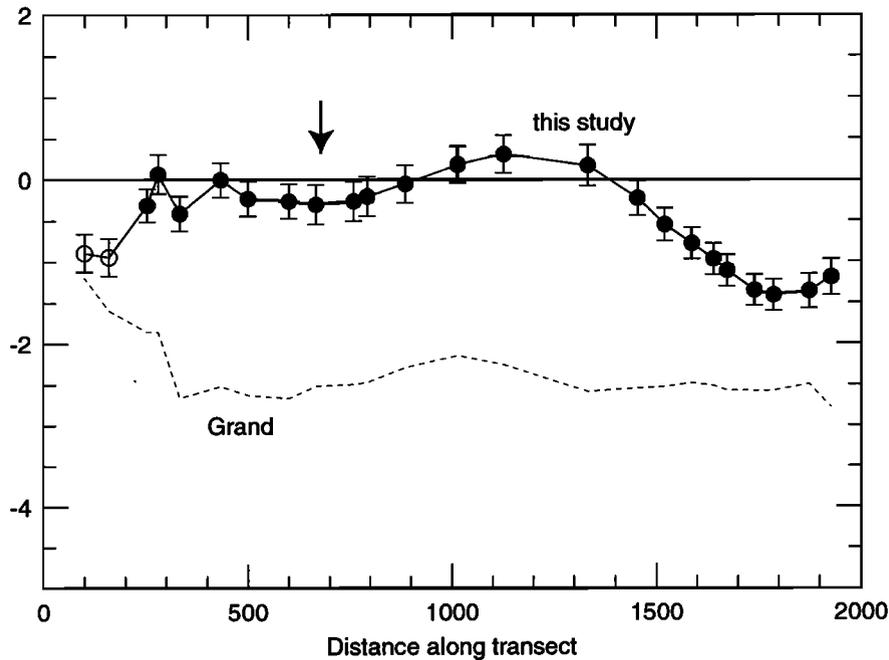


Figure 6. S wave upper mantle delay (symbols with error bars) from data set dirnosub compared with prediction computed from *Grand's* [1987] model (dashed line) at the same locations. Solid circles show the inversion results which are considered reliable. In comparison with the permanent station study of Grand, the transition to fast shear velocities occurs ~ 1000 km farther to the northwest. This difference may be caused by the coarse station spacing of about 1000 km in Grand's study. With a station spacing of 50 to 100 km in the present study the resolution is increased by one order of magnitude. Note that the mantle transition does not agree with the edge of the shield (arrow).

contributions contaminating the data sets all and dirall. The two western stations on the Wyoming Craton are comparatively fast for the conservative data sets. As reasoned above, the results for allnosub and dirnosub are more credible than for all and dirall. Consequently, the upper mantle under the Wyoming Craton appears intermediately fast for S with values between the slow and the fast region.

The P variation is comparatively small. Near the strong transition in S delay the corresponding P delays vary by < 0.2 s. In the westernmost (less well-constrained) part the shape of the P delays is more similar to the S delays. The scales of Figure 4b of 2 s in P to 4 s in S are such that the same percent variation between S and P should show similar-sized variations. Yet the P variation is much smaller. The residual upper mantle delays are also shown in Figure 2.

Clearly, the large transition in S delays is not affected by the crustal correction, since it is located in an area without sedimentary coverage. This important feature of the data is addressed in the following sections.

4.1. Location of the Mantle Transition

For S and data set dirnosub (Figure 6; symbols with error bars) we compare the upper mantle delay with predictions from the tomographic model of *Grand* [1987,

dashed line]. Both our models and the prediction from Grand essentially show slow mantle in the west and fast mantle in the east. However, whereas our observed transition is near 1400 km the predictions from Grand's model at the locations of the APT89 stations show an approximately constant delay on most of the transect and a transition to larger delays in the far west. With typical separations of permanent stations of 1000 km the resolution in that study is not expected to be better than 1000 km. The difference between the two results may be due simply to the different resolution.

The transition region in S in our data is not located near the boundary between Superior Province and Trans-Hudson Orogen, which is interpreted as the shield boundary based on basement rock age. Instead, it is located several hundred kilometers inside the shield.

4.2. Relative Variation of P and S Delays

The observation of much larger variation in S delays compared with P delays is not unique to this study. The same apparent discrepancy between P and S delays has been found in other data sets. For example, the S station delays of *Wickens and Buchbinder* [1980] show a well defined S anomaly within the shield but only a weak P anomaly. Figure 7 shows the P and S residuals from that study together with our results.

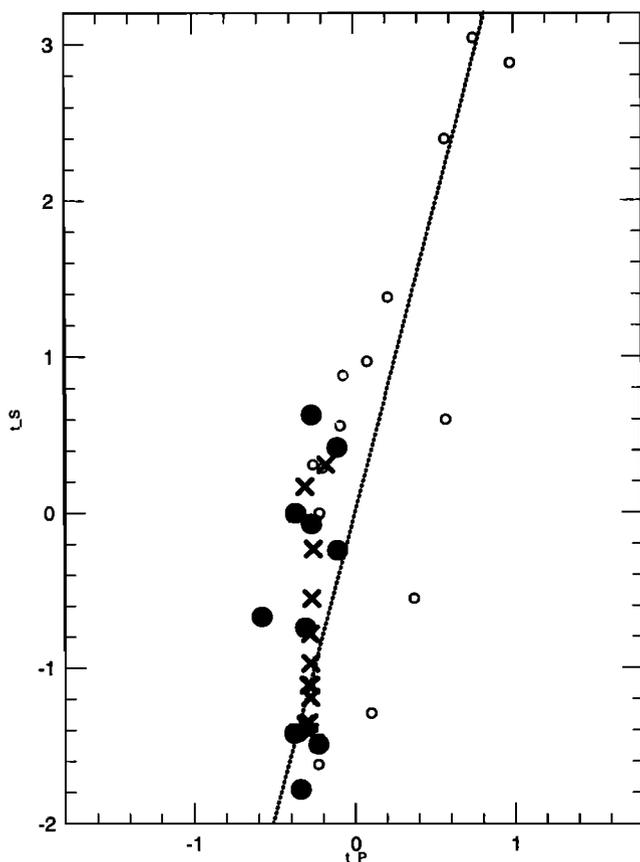


Figure 7. Comparison of older P and S travel time residuals [Wickens and Buchbinder, 1980, Buchbinder and Poupinet, 1977] (circles) with our results from Figure 4b (crosses, baseline-shifted to account for the different reference Earth model). Solid circles show the stations located on or near the edge of the Precambrian shield (as defined by Bally *et al.*, [1989]). They are consistent with the large slope of $\Delta t_s/\Delta t_p$ found in our data. Note that this slope is not consistent with thermal variation (the largest consistent slope is shown by the dotted line for $d\ln v_S/d\ln v_P \leq 2.25$).

The stations, which are located on the shield or near its edge show a similarly large variation in S but small variation in P residuals. Taking only these stations results in essentially no P delay variation with S , corresponding to much higher values of $d\ln v_S/d\ln v_P$ than the value of 2.25 which is the value that roughly fits the whole data set (dotted line). Such results imply values of $d\ln v_S/d\ln v_P$ substantially larger than 2.25 under the shield. While values near 2.25 have been argued to indicate the presence of partial melt, there is no simple explanation for values larger than 2.25. Data from Romanowicz [1979] for North America also indicate large values. A similar effect may perhaps exist in eastern Europe. While S velocity variation associated with the western edge of Baltic Shield is large [Zielhuis and Nolet, 1994], P velocity variations are small or not discernable although there is little resolution within the

Baltic Shield [G. Nolet, personal communication, 1992; Spakman *et al.*, 1993].

4.3. Azimuthal Dependence of Upper Mantle Delays

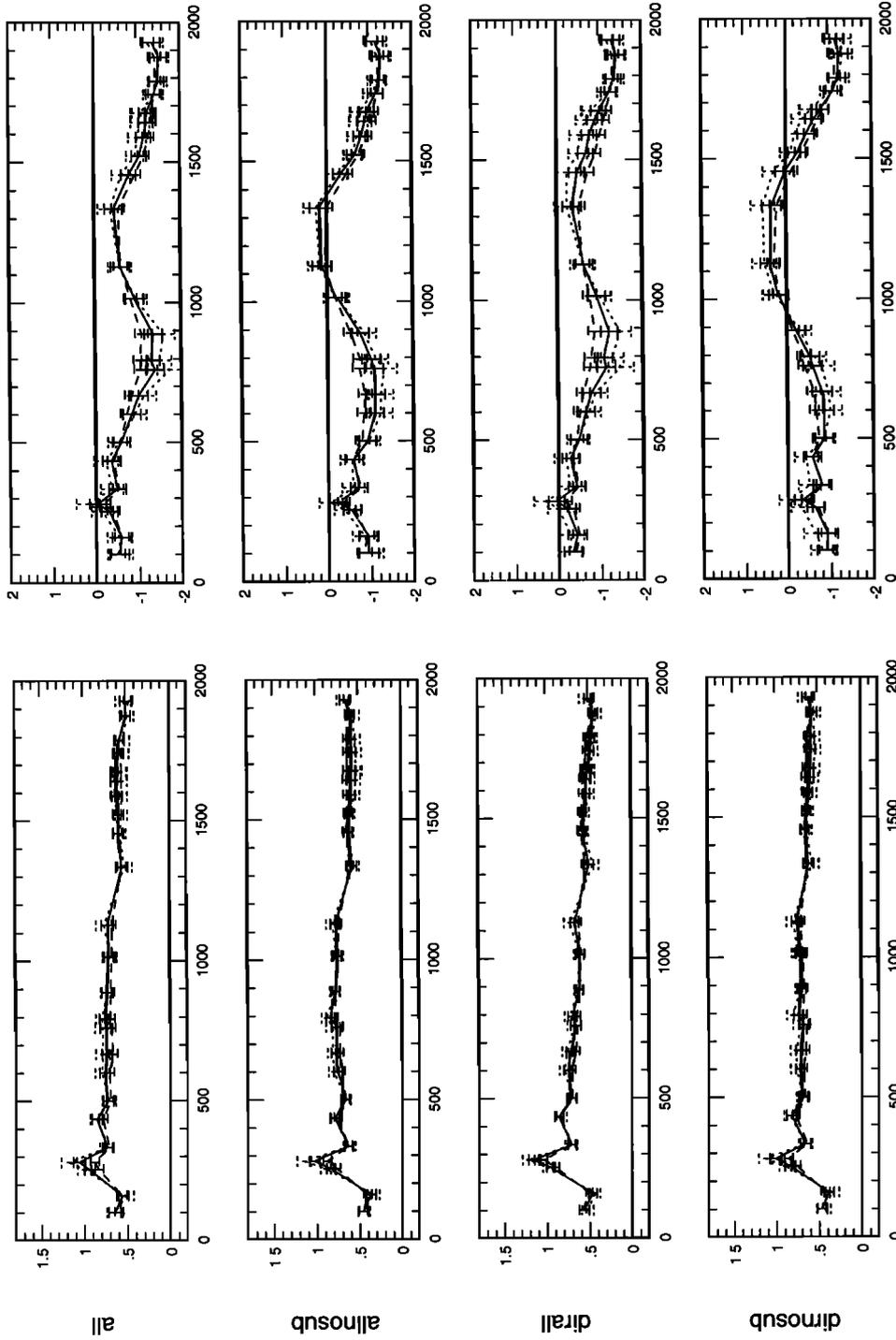
At the beginning of section 4 we estimated the upper mantle delay at each station from inversions with $D=0$ in (2), which effectively averages over azimuth. The azimuthal dependence of δt_{ijk} is obtained by extending this sum up to $D=2$. In this case we have five parameters for each station and phase type P and S . One criterion for accepting results can be based on whether the azimuthally independent terms (c_{i0}) obtained from these expanded inversions (Figure 8) are consistent with upper mantle delays δt_{UM} obtained from the azimuthally independent inversion or whether they tradeoff with the azimuthally dependent terms. If these terms are not similar, we should not expect to recover the azimuthally dependent terms correctly.

For P (on Figure 8, left) the models are very similar to Figure 4b. For S , however, the curves do not generally agree with Figure 4b, indicating a trade-off of azimuthally dependent terms with the c_{i0} terms.

While our criterion allows the further usage of azimuthal terms for P , but not for S , the variance improvement for P is small. The low variance ratio improvement for P of $< 4\%$ shows that the inversion does not require inclusion of azimuthal terms for P . We show in section 5.3 that this is reasonable in the light of an olivine upper mantle model. On the other hand, the variance improvement for S is much larger ($\sim 10\%$). Even though we cannot constrain the correct values of the azimuthal terms for S according to our criterion, the data set appears to contain azimuthal variation. That can also be understood in conjunction with anisotropy (see section 5.3).

5. Interpretation of Travel Time Delays

Any model for the upper mantle in this region must explain three basic observations: the 1.5 s change in S wave delay between the northern stations (e.g., station NAOR) and the stations in the Western Superior Province (e.g., station TRWM), the near-absence of a P wave change at the same location, and the observation that the splitting delays change in roughly the same locations as the S delays. Additional observations are that the azimuthal terms for P are small, while they are relatively large for S . Assuming that the S delay is due to a change in lithospheric thickness, it would require a change of 160 km, assuming that S wave velocity is related to temperature by $-3.6 \times 10^{-4} \text{ km s}^{-1} \text{ }^\circ\text{C}^{-1}$ [Anderson and Isaak, 1995]. Yet such a model would yield a corresponding P delay of ~ 0.9 s, assuming the laboratory values of $d\ln v_S/d\ln v_P \approx 1$. The absence of a detectable P delay, however, requires an unrealistically high value of $d\ln v_S/d\ln v_P \gg 2$. Another way of gener-



Distance along the transect

Distance along the transect

Figure 8. Contribution c_0 (equation (2), azimuthally independent term) for the four data subsets in an inversion including azimuthal effects up to order $n=2$ for the upper mantle delay. While the results for P are virtually the same for each data set and also agree with the inversion up to order $n=0$ (Figure 4b), the results for S vary between the different data set. This suggests that the data sets for S are too small to independently constrain parameters in an inversion up to order $n=2$.

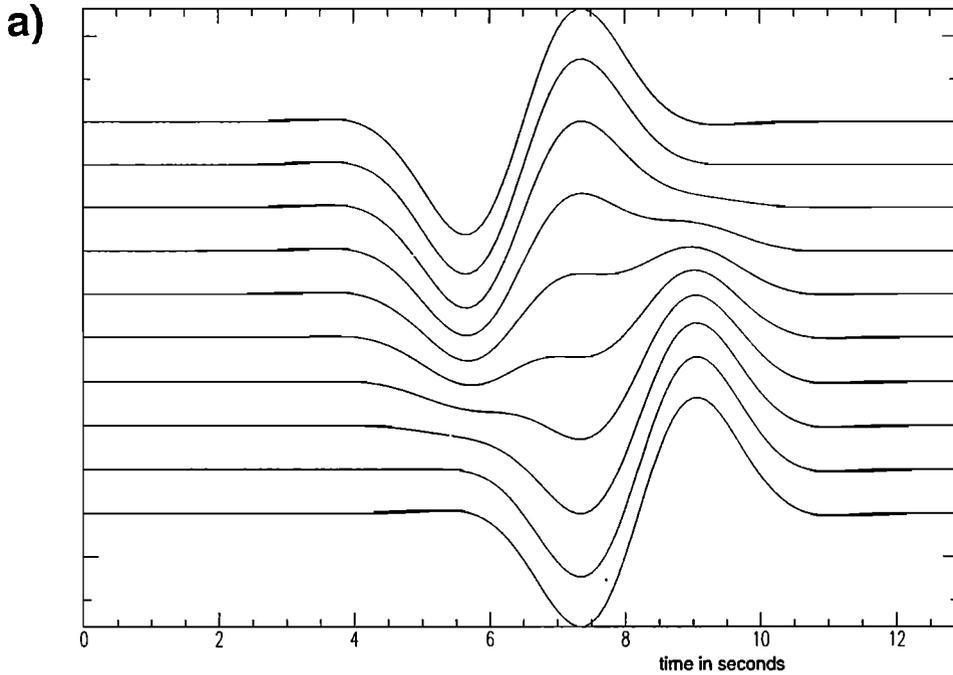


Figure 9a. Illustration of a transverse component shear waveform after propagation through an anisotropic layer. The signal consists of two separate phases. The initial shear wave polarization is in transverse direction. For an azimuth parallel to the slow direction the polarization direction is parallel to the fast axis and the waveform consists only of the faster *S* phases (top trace). The azimuth is varied by 10° up to the slow polarization direction in the bottom trace.

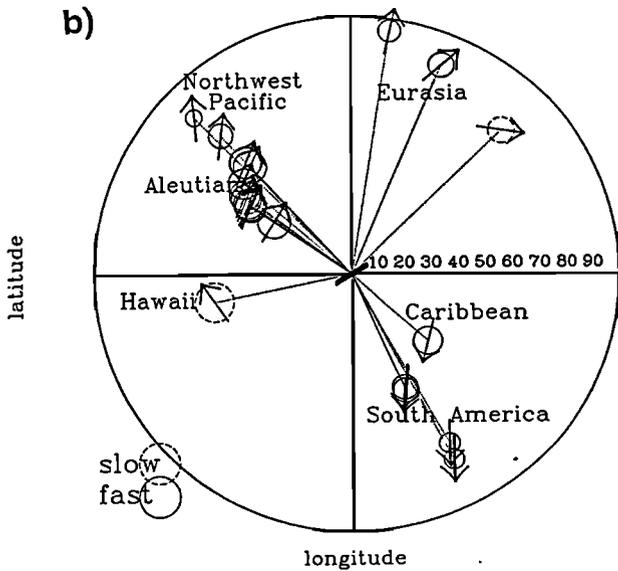


Figure 9b. The initial polarization angles for all events (from the Harvard Centroid moment tensor (CMT) solutions) in a lower hemisphere projection for station DLOR. An arrow parallel to the great circle (dotted line) would thus indicate pure *SV* polarization and perpendicular pure *SH* initial polarization. The bar in the center shows the polarization direction (70°) obtained from shear wave splitting for DLOR. Circles with solid (dashed) lines show events for which the faster (slower) shear wave has the stronger amplitude. The amplitude of that stronger wave is given by the size of the circle; maximum obtainable values for pure fast or slow phases are given by the scale on the lower left.

ating this *P/S* difference is by considering the role of anisotropy.

5.1. Effect of Shear Wave Splitting on *S* Delays

Strain-induced anisotropy in aggregates of olivine is the most likely cause of upper mantle anisotropy. This will perturb both *P* and *S* velocities and give rise to shear wave splitting. As a first step in addressing the role of anisotropy, we consider the effect of shear wave splitting on the measurement of *S* wave delays. This issue has received surprisingly little attention in the context of mantle tomography, even though splitting delay times can be comparable in size to isotropic delays. We define apparent travel times as those made under the assumption of an isotropic medium, when, in fact, there is an anisotropic medium with two distinct split shear waves. In this case, the waveforms are composites of both split shear waves. This is illustrated in Figure 9a for a vertically propagating *S* wave with initial transverse polarization shown on the transverse component. For back azimuths nearly parallel (perpendicular) to the fast polarization direction the slower (faster) wave dominates the signal. The actual travel time will depend on their relative amplitudes and sign. This in turn depends on the geometry of the arrival (back azimuth and incidence angle), initial polarization ϕ_i and the fast polarization direction ϕ intrinsic to the medium. Figure 9b gives the initial polarization directions for the events used in this study. If we have a large number of events

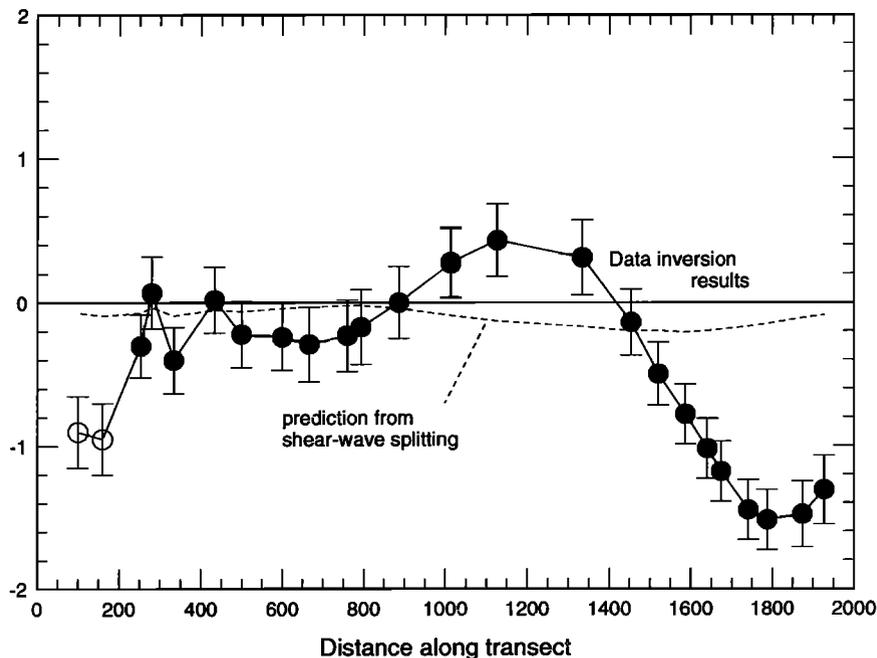


Figure 10. Prediction of travel time delays from shear wave splitting using equation (3). Comparison with the inversion results for data set dirnosub (Figure 4b) shows that the prediction from shear wave splitting alone is too small to explain the observed travel time delays. This suggests that the S delays in our study represent the delay average between the fast and the slow shear wave.

with randomly distributed initial polarizations ϕ_i , we expect to find a station delay for S which is the average of the two polarization directions. However, for smaller and less diverse data sets, that average is not reached and anisotropy does have an effect on isotropic travel times. How S delays depend on the properties of an anisotropic medium is shown in Appendix D. For fixed frequency ω the apparent travel time delay δT_a due to splitting parameters (fast polarization direction ϕ and delay time δt) is given by

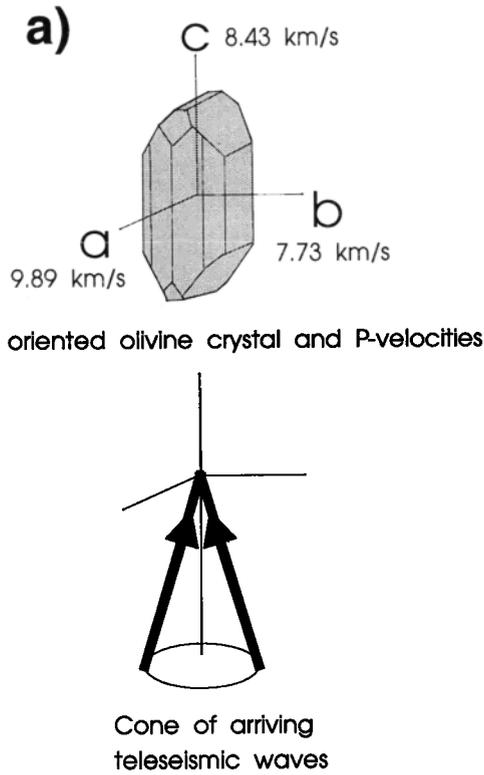
$$\tan(\omega\delta T_a) = -\tan\left(\omega\frac{\delta t}{2}\right) \left[\frac{\cos(2\phi_h + \delta\phi)}{\cos(\delta\phi)} \right], \quad (3)$$

where ϕ_h is the azimuth of the observing component (usually transverse or radial) with respect to the fast direction. We have $\delta\phi = \phi_p - \phi_h$ with the initial polarization ϕ_p relative to the fast direction. The parameter ω is chosen to correspond to the dominant period of the data, taken to be ~ 5 s. Thus, for given geometry and splitting parameters it is possible to calculate the apparent delay. We have done this for one of the data sets that we have used, dirnosub, and inverted these values in the same way as the data. The dashed line in Figure 10 shows the result of that inversion in comparison with the data inversion result. We find that there is a 0.2 s change in delay time across the array that can be attributed to this effect. This is nearly an order of magnitude smaller than the observed S wave delay, and thus shear wave splitting alone does not appear to ac-

count for the delay. On the other hand, this suggests that S delays in our study are indeed close to the average between fast and slow delay times for each station (to within 0.2 s) due to a sufficiently diverse path and station geometry.

5.2. Effect of Olivine in the Continental Mantle

We next consider more general models in which there is an anisotropic and an isotropic component to the model, such as a variable-thickness anisotropic lithosphere. The origin of upper mantle anisotropy is reasonably well understood. Olivine is thought to be a major constituent of the upper mantle. Mantle xenoliths contain large amounts of olivine, typically $> 60\%$ [Nicolas and Christensen, 1987]. The rest is mostly orthopyroxene. The anisotropy of the rock is typically dominated by olivine, since it has large single-crystal anisotropy and shows a strong tendency to develop lattice preferred orientation (LPO) under deformation [Nicolas and Christensen, 1987; Kern and Wenk, 1990]. Admixture of pyroxene and other relevant minerals may decrease the bulk anisotropy of the rock [Christensen, 1984; Mainprice and Silver, 1993]. Early refraction work in the oceans [Hess, 1964] and also on continents [Bamford, 1977] suggested that large observed azimuthal travel time variations of P_n could be explained by preferential alignment of olivine within the mantle. In order to have a strong effect on P_n either the proportion of aligned olivine must be very high or the a axis



must be nearly horizontal. For a large range of temperatures, *Nicolas and Christensen* [1987] showed that the a axis aligns parallel to the flow line (see *Silver*, [1996] for discussion).

Figure 11a shows a single olivine crystal [after *Nicolas and Christensen*, 1987] and the *P* velocities for propagation along the three crystallographic axes. Note that depending on the direction, the velocity variation may be substantial, up to 25%. The fastest velocity occurs for propagation along the a axis. The events in this study have a range of incidence angles between 10 and 35° from the vertical. Our ray path coverage below the receivers is thus essentially a cone with an average opening angle of about 20°. For such a cone, we inspect the velocity of *P* and *S* waves depending on back azimuth for three orientations (vertical a, b, c). We note that the *P* velocity depends primarily on which of the three axes is vertical. Shear wave splitting is large both for vertical b and c axes, but not for a vertical a axis.

5.3. Variable Thickness Anisotropic Lithosphere

In the following we restrict our attention to the part of the shield between 1100 and 2000 km along the transect. In that region we found the strong variation of shear wave delay associated with the change in shear wave splitting, but with no or only small *P* wave delay variations. Inversion results for that region are not

Figure 11a. Olivine model and *P* wave velocities for propagation along the *a*, *b* and *c*-axes (elastic constants after *Kumazawa and Anderson*, [1969]). Most events in this study have incidence angles in the upper mantle near 20°.

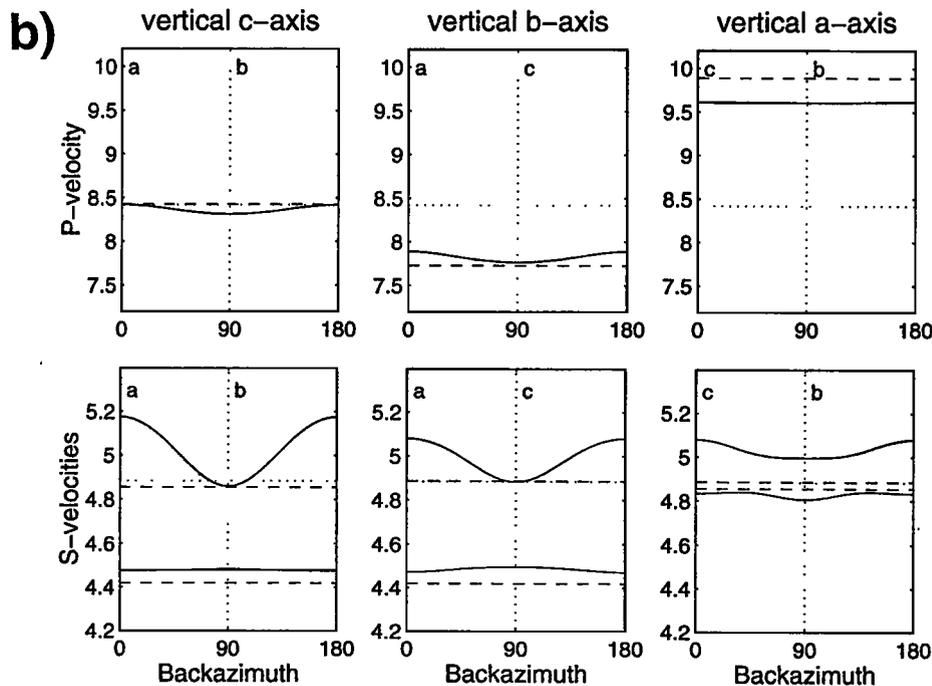


Figure 11b. Velocities (in km/s) of *P* and the two *S* velocities (solid lines) for this cone of teleseismic waves. We allow each *a*, *b* and *c* axis to be vertical. Note that the *P* velocity depends strongly on which axis is vertical. The splitting is nearly the same for vertical *b* and *c*-axes orientations. Dashed lines show the velocities for vertical propagation; dotted lines show the velocities for the isotropic Voigt-Reuss-Hill averages [*Hill*, 1952]. The *S* velocity depends strongly on which of the split shear wave phases dominates the composite waveform rather than on the azimuthal velocity variation of the individual shear wave branch.

Lithospheric Transition Model

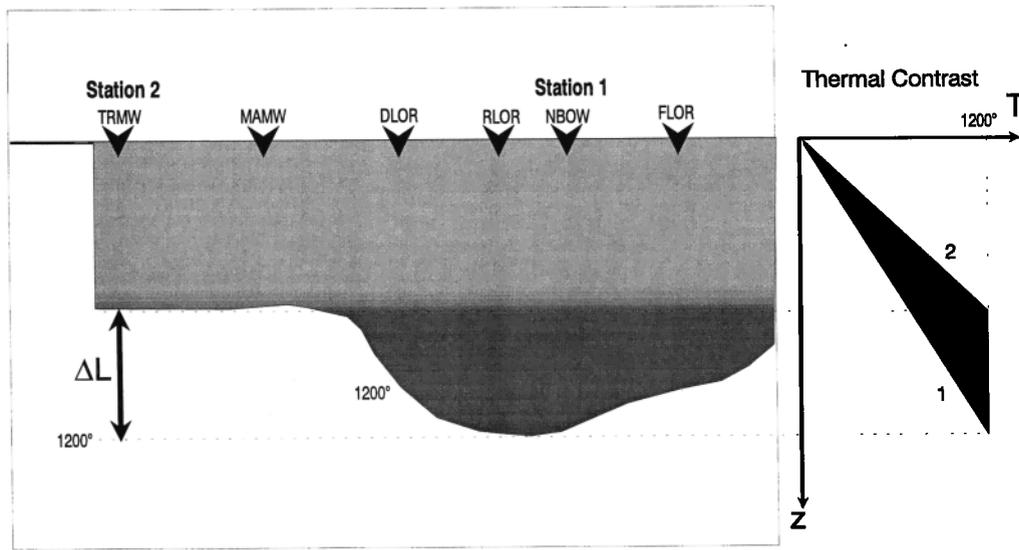


Figure 12. (left) Model for the lithospheric transition at the western edge of the Canadian Shield. The lithospheric root ΔL is defined by the 1200°C isotherm (thermal boundary layer). Our simplified model assumes a transition from the lithospheric root to a hotter mantle which for simplicity is assumed to be 1200°C . (right) The lithosphere has a constant thermal gradient. The shaded area indicates the temperature excess of the normal mantle relative to the lithospheric root. The root is also characterized by deformation-related anisotropy, while the hotter mantle is assumed to be isotropic (see text).

affected by the trade-off with the sedimentary delay discussed in Appendix C due to the lack of sediments there.

Near the edge of a craton, we may expect to find variations in seismic structure of the mantle due to changes in composition, temperature, and seismic anisotropy. Temperatures in the mantle are thought to be lower under the craton compared with the surrounding mantle at the same depth [Jordan, 1988]. We may also expect a contrast in chemical composition across the edge with the material under the shield being more depleted than the surrounding [Jordan, 1981].

Lateral temperature variations at 150-200 km depth are generally thought to be $\sim 200^{\circ}$ - 400° with the lithospheric mantle under shield regions having lower temperatures than the surrounding mantle, especially oceanic mantle [Pollack and Chapman, 1977]. With temperature variations that large, it is generally concluded that thermal variations are likely to dominate compositional changes [see e.g., Nolet and Zielhuis, 1994].

The seismic contrast may also be influenced substantially by a contrast in seismic anisotropic properties, for example, due to thickness changes of the deformed subcratonic lithosphere [Silver and Chan, 1988, 1991]. The effect of anisotropy on travel times can, in fact, be as large as the effect of temperature variations or even exceed them. We will thus consider the combination of thermal and anisotropic changes below and regard the smaller effect of composition separately later.

From Figure 11b we see that anisotropy may indeed affect P and S delays quite differently. For instance, a perfectly aligned vertical b axis would slow down P waves by 8% compared with the Voigt-Reuss-Hill (VRH) average, which we would get for a random distribution of olivine. S phases split into two phases, one faster and one slower than the average. The average between the two, however, is slower than the VRH-average by 3%. The amount of oriented olivine in the lithospheric mantle thus slows down P phases much more than S phases. This may be a mechanism for explaining why the P delays hardly vary across the transect. We seek to explain these observations by anisotropy and a thermal change $(\partial v/\partial T)\Delta T$ with a temperature difference ΔT .

We will consider anisotropy localized within the lithosphere [Silver and Chan, 1988, 1991], which may be coupled to temperature variations through thickness changes of the anisotropic layer. The alternative case of anisotropy confined within the asthenosphere will be discussed later in the discussion section. Our simplified lithospheric transition model (Figure 12) includes a thermal boundary layer, the lithosphere, which we define using the 1200°C isotherm.

The lithospheric material may be deformed and thus give rise to seismic anisotropy [e.g., Silver and Chan, 1988, 1991]. The changes of P and S wave delay within the shield enable us to make inferences about

Table 3: Observed and Predicted Delay Changes

	P Delay $t_2^p - t_1^p$	S Delay $t_2^s - t_1^s$
Observation	0.1 s	1.5 s
Effect of temperature	$-(\frac{\partial v_P}{\partial T}) \frac{\Delta L}{(v_{\text{RH}}^p)^2} 600^\circ\text{C}$	$-(\frac{\partial v_S}{\partial T}) \frac{\Delta L}{(v_{\text{RH}}^s)^2} 600^\circ\text{C}$
Effect of anisotropy	$\frac{v_{\text{ol}}^p - v_{\text{RH}}^p}{(v_{\text{RH}}^p)^2} c \Delta L$	$\frac{v_{\text{ol}}^s - v_{\text{RH}}^s}{(v_{\text{RH}}^s)^2} c \Delta L$
Resulting lithospheric root thicknesses ΔL , km:		
for $d \ln v_S / d \ln v_P \approx 1.0$	100	230
for $d \ln v_S / d \ln v_P \approx 2.25$	100	100
Symbols	Values	
ΔL	excess thickness of lithospheric root	
c	volume fraction of oriented olivine in the mantle	
$\frac{\partial v_P}{\partial T}$	temperature derivative of P velocity (-5.4×10^{-4} km/sK)	
v_{RH}^p	Voigt-Reuss-Hill average of P velocities	
v_{RH}^s	Voigt-Reuss-Hill average of S velocities	
v_{ol}^p	azimuthal P velocity average (olivine, vertical b axis)	
v_{ol}^s	azimuthal S velocities average (olivine, vertical b axis)	
δt_{split}	difference in splitting delay between stations NAOR and TRMW (0.95 s)	
$\delta \hat{v}_{\text{ol}}$	expected shear wave splitting in olivine (0.1048)	

the lithospheric root region only, where we compare the deformed cratonic root containing partially oriented olivine with the hotter mantle around it, which we assume to be isotropic (randomly oriented olivine). For simplicity, we assume that the latter is isothermal at 1200°C , since the thermal gradient is expected to be much smaller there than within the thermal boundary layer. We will determine the excess root thickness ΔL , which is required to satisfy the data.

Table 3 lists the observed changes of P and S delay (0 and 1.5 s) between stations NAOR (1787 km) and TRMW (1127 km). These are to be compared with the combined effect of temperature and anisotropy. The temperature effect on P delay for instance is $dt_P = -(\partial v_P / \partial T)(\Delta L / v_P^2) dT$ which, integrated over the lithosphere, gives a travel time difference between the two stations (for the temperature profile in Figure 12) of approximately $-(\partial v_P / \partial T)(1200^\circ / 2v_P^2) \Delta L$. Anisotropy, on the other hand, depends on $c \Delta L$ with the concentration c of anisotropy. That dependence (trade-off) can be eliminated by using shear wave splitting observations since the difference of shear wave splitting delay time δt_{split} between the two stations gives

$$c \Delta L = \frac{v_{\text{ol}}^s}{\delta \hat{v}_{\text{ol}}} \delta t_{\text{split}}$$

with the known relation between splitting delay and anisotropic properties for olivine. The relative magnitude of shear wave splitting in a pure (single-crystal)

olivine mantle is $\delta \hat{v}_{\text{ol}}$. Equating the “observed” and predicted effects in Table 3, we obtain estimates of ΔL independently for P and S . The results, however, depend (again) on the relation $d \ln v_S / d \ln v_P$ in the mantle. Isotropic values range from $d \ln v_S / d \ln v_P \approx 1$ to $d \ln v_S / d \ln v_P \approx 2.25$ [Anderson, 1989]. For the P delays we get excess root thicknesses ΔL of 100 km in both cases for the (relatively uncertain) value for $(\partial v_P / \partial T) = 5.4 \times 10^{-4}$ km s $^{-1}$ K $^{-1}$ [Anderson and Isaak, 1995]. For the S delays we get an agreement with the P delay result only if we choose $d \ln v_S / d \ln v_P$ to be relatively large (2.25). This model also predicts the behavior of the observed azimuthal terms: The azimuthal velocity variation for P is small ($\approx 0.5\%$, similar to the behavior of pure olivine in Figure 11b but smaller in size), while the S azimuthal terms may be affected by the shear wave splitting and are thus likely to be much larger, as observed. Generally, this model predicts a concentration of $c \approx 0.4$, which is close to the observed value [Silver, 1996].

Although the estimate of the excess root thickness is relatively uncertain due to the large uncertainty in the temperature derivatives of velocity, this indicates that the data may indeed be explained by the simple model of a lithospheric transition involving thermal changes and accompanying changes in lithospheric anisotropy. The effect of compositional heterogeneity, namely, that of depleted peridotite in the lithospheric root versus undepleted material around it, is to increase P and S velocities in the root similar to the effect of lower temperatures but by a smaller factor. Allowing for such compositional heterogeneity would thus require anisotropy to balance a slightly larger effect. This would give a slightly larger thickness ΔL of the lithospheric root.

6. Discussion

The most robust and intriguing feature of the travel time data set that we have presented is the 1.5-s change in S delay times within the Superior Province. This variation roughly corresponds to a similar variation in shear wave splitting delay times [Silver and Kaneshima, 1993], but is not accompanied by a detectable change in P delay times.

Such results imply large values of $d \ln v_S / d \ln v_P$, larger than 2.25. Values near 2.25 have been argued to indicate the presence of partial melt. No purely thermal model (and probably no purely compositional model) can be found to satisfy values larger than 2.25. However, as we have seen, very large values of $d \ln v_S / d \ln v_P$ can be produced by not adequately accounting for the effect of anisotropy.

The mantle transition within the Superior Province, which is apparent in both upper mantle delay and shear wave splitting, is not clearly related to any of the most prominent surface geological suture zones (Figure 1), which have been interpreted as crustal divisions. Instead, it occurs within the Superior Province of the

Canadian Shield located ~ 500 to 800 km east and 200 km north of the boundary between Archean and Proterozoic age basement. The western geological boundary is constrained by drilling [*Klasner and King, 1986*] and has been interpreted as the western edge of the shield. The mantle variation found in this study is accommodated over not more than ~ 500 km in lateral distance and appears to represent a major lateral mantle discontinuity. The model is based on lateral variations in temperature corresponding to variations in lithospheric thickness. We suggest that the higher temperature under the westernmost part of the Superior Province may have destroyed the fabric in the rock, thus weakening the bulk anisotropy of the mantle under this region. Alternatively, there may not have been a similar lithospheric root under the Western Superior Province in the first place. Whether the mantle in that region is chemically distinct from the mantle under the cratonic root cannot be answered based on these data alone. However, the uplift of the cratonic region suggests a lower density of the subcratonic root [*Jordan, 1981*]. A lower density may help to explain why the cratonic root has apparently been dynamically stable for considerable time despite its negative buoyancy from lower temperatures. One might perhaps expect to find variations in heat flow in the region. However, given the scatter in the heat flow data [e.g., *Roy et al., 1972*], there appear to be no systematic variations within the shield region, which gave rise to the large changes in S and splitting delay.

An essential feature of the anisotropic lithosphere model is that the 'excess' anisotropy (i.e., the difference in the splitting delay times between the two stations), presumably located in the bottom part of the thickened lithosphere, will only cancel the thermal effect on P delays if the b axis concentration of olivine is vertical, rather than the c axis. While a vertical c axis is expected for transpressional deformation associated with orogenic events, a vertical b axis is expected for horizontal shearing of the base of the lithosphere. The anisotropic properties of continental lithosphere generally appear to be compatible with a vertical c axis, which would be the case for coherent transpressional deformation of the lithosphere during orogenic events. Indeed, the close correspondence between fast polarization direction and geologic fabric suggests the dominance of this form of anisotropy globally produced by present and past orogenic events [*Silver, 1996*]. For the case of extremely thick lithosphere, as found in the Western Superior Province (and other regions such as the Baltic Shield), and/or strong lateral gradients in lithospheric thickness, it may be that the bottom of the lithosphere is sheared by its interaction with general mantle circulation.

Potential asthenospheric anisotropy would also be associated with horizontal foliation planes [e.g., *Silver, 1996*]. However, it is difficult to reconcile our obser-

variations with asthenospheric anisotropy since it would predict larger splitting delay times for regions of thicker asthenosphere, namely, where the lithosphere is thinner. Just the opposite is observed.

While teleseismic delay and shear wave splitting data do not have strong depth resolution, studies of regionally propagating surface and body waves do. Regional studies of the Australian continent [e.g., *Gaherty and Jordan, 1995*] suggest that the anisotropy is indeed confined to the shallower part of the upper mantle (above the Lehmann discontinuity at about 250 km depth) rather than to the lower part. *Gaherty and Jordan [1995]* speculate that the Lehmann discontinuity marks the transition to a more mobile part of the continental "tectosphere". Our results for the Canadian Shield seem to represent a layer within the anisotropic lithosphere (a possible subdivision). Our crude thickness estimate would argue for an extended zone of distributed deformation rather than a localized shear zone.

The model that explains the data requires a contrast in temperature between the central and northern part of the transect with the central part having higher temperatures than the north. We now address the question of which process may have caused excess temperatures under the central part of the transect. The geological structure of the region has been shaped and altered most profoundly in the Precambrian. Geological events younger than Precambrian cannot be easily related to the observations. The Mesozoic Laramide orogeny [*Bird, 1984*], while having a pronounced effect on the western United States, is not likely to have affected the lithosphere so far to the east. Its easternmost extent is thought to be the Black Hills, South Dakota, at the western end of our seismic line. The one large Paleozoic feature in the region is the Williston basin. While the subsidence history of the basin is quite well known [*Fowler and Nisbet, 1984; Ahern and Mrkvicka, 1984*], its underlying cause remains obscure. One model for the basin is the heating of the mantle, thinning of the crust, and subsequent subsidence upon recooling of the mantle. There is, however, no direct evidence for such a thermal event associated with basin formation. In addition, the center of the basin is located over the Trans-Hudson, ~ 500 km to the west of the observed seismic change.

Thus it appears that Precambrian tectonic activity resulted in the observed lateral change in the mantle. This explanation requires that the thermal changes have been present since then. If lateral heat transport is solely by conduction, then we expect the heat excess of a thermal anomaly produced in the Proterozoic or Late Archean to have propagated laterally ~ 200-300 km. Thus lateral temperature anomalies originating in the Precambrian should still be present in the mantle if the thermal transport is dominated by conduction.

The Precambrian history of the region is complex [*Hoffman, 1989*]. Correlating mantle structure with

these ancient events must therefore remain somewhat speculative. Two processes which are known to give rise to lateral changes of temperature in the mantle, however, are subduction and rifting.

Whether and in which fashion subduction occurred during the assemblage in the western part of the Canadian Shield is controversial. Reflection surveys found prominent eastward-dipping crustal reflectors [Hajnal *et al.*, 1996], which have been interpreted as indications of subduction to the east under the Superior Province. There are, however, no indications of major subduction-related volcanism within the Western Superior Province (P.F. Hoffman, personal communication, 1992) rendering a subduction explanation unlikely.

The Precambrian history of the region may also have involved rifting. There are two sets of mafic dike swarms in the Minnesota area, indicating at least two periods of extension in the Proterozoic. The earlier swarm trends northwestward and predates the Trans-Hudson Orogen (age data 2.1 to 2.2 Ga from Sims, [1996]). This may perhaps be associated with rifting near the western edge of the craton, although there is no definitive geological evidence in this matter due in part to sedimentary cover in crucial areas.

The later swarm of mafic dikes is apparently associated with the Midcontinental Rift System MCRS (Early Proterozoic, 1.1 Ga), which occurred several hundred kilometers south of our transect, in the Great Lakes region. Large amounts of volcanic material (e.g., Duluth Complex) were emplaced during this episode [Cannon and Nicholson, 1996]. It is conceivable that a thermal anomaly associated with that rifting produced hotter mantle under the closest stations MAMW, TRMW, and CNOW. The closest station MAMW, where the slowest shear wave velocities occur, is only 130 km from the Duluth Complex. A thermal mantle anomaly from the Proterozoic would have had just enough time to warm up the mantle under the stations closest to the MCRS.

Which of these causes gives rise to the mantle change must remain somewhat speculative given the incomplete geological data at hand. However, Proterozoic rifting (the MCRS) appears to be a promising candidate. It has been suggested previously that igneous activity weakens anisotropic fabric in the mantle lithosphere. This has been invoked to account for weak shear wave splitting along the east coast of North America for stations within regions of both low mantle velocities and Mesozoic diking associated with continental breakup [Barruol *et al.* 1997]. In addition, the 2 Ga Buschveld complex, the worlds largest igneous intrusion, is coincident with anomalously weak anisotropy in southern Africa [Silver *et al.* 1998]. The occurrence of very large values of $d\ln v_S/d\ln v_P$ at shield boundaries suggests that the model proposed in this paper may actually represent a more general feature of shield boundaries, namely, a lateral variation in mantle temperature associated with a change in anisotropy.

Appendix A: Inversion Strategy

Equation (1) gives rise to a linear problem

$$\mathbf{Ax} = \delta\mathbf{t} \quad (A1)$$

with a vector of unknowns \mathbf{x} which contains the event mislocation residuals \mathbf{b} , path residuals \mathbf{a} , and station correction parameters \mathbf{c} and \mathbf{s} . We perform the inversion separately for the sets of P (P , pP , PcP) and S (S , sS , ScS , $sScS$) travel times, as the two sets may be affected by different types of errors. Also the scaling between t_P and t_S is still controversial [Dziewonski and Woodhouse, 1987; Bokelmann and Silver, 1993]. On the other hand, we may expect that station corrections as well as azimuthal delays and fast directions vary smoothly across the transect. This was suggested from a preliminary study [Bokelmann and Silver, 1991a], which showed that travel time variation across the transect was large but occurred rather smoothly. A smoothness constraint on relative travel time variation across the transect [Bokelmann, 1992] has also been used for obtaining the travel time residuals, used as data in this study. If azimuthal station corrections vary smoothly across the transect, the model parameters \mathbf{c} and \mathbf{s} in (2) will also vary smoothly. This can be imposed as a constraint in the inversion by minimizing an objective function which consists of a data misfit term and a model misfit term

$$(\mathbf{Ax} - \delta\mathbf{t})^T \mathbf{C}_d^{-1} (\mathbf{Ax} - \delta\mathbf{t}) + \gamma \mathbf{x}^T \mathbf{C}_x^{-1} \mathbf{x} \quad (A2)$$

with the data covariance matrix \mathbf{C}_d^{-1} and the model covariance matrix \mathbf{C}_x^{-1} . Elements of the latter can be chosen depending on distance which incorporates the smoothness constraint. From (A2) follows the stochastic inverse

$$\mathbf{x} = (\mathbf{A}^T \mathbf{C}_d^{-1} \mathbf{A} + \gamma \mathbf{C}_x^{-1})^{-1} \mathbf{A}^T \mathbf{C}_d^{-1} \delta\mathbf{t} \quad (A3)$$

In this stochastic inversion [Jordan and Franklin, 1971] the parameter γ trades off the data misfit and model misfit such that for larger γ the latter is preferentially minimized. We obtain a model closer to our a priori model which in this case is simply the reference Earth model.

Waveform fitting [Bokelmann and Silver, 1991b] produces travel time residuals δt_l each with associated confidence intervals bounded by δt_l^{\min} and δt_l^{\max} . Owing to noise and waveform asymmetry in time, these confidence bounds are not necessarily symmetric around the best fit δt_l . We take the (initial) data covariance matrix as $\mathbf{C}_{d_0} = \text{diag}(\sigma_l^2)$ which assumes uncorrelated travel time residuals. The waveform fitting in principle allows multiple acceptable confidence regions. In this case we estimate the variance σ_l^2 of the l th data point from summing ($n=1$ to N) over these n confidence regions as $\sigma_l^2 = [\sum \max(\delta t_{ln}^{\max} - \delta t_{ln}, \delta t_{ln} - \delta t_{ln}^{\min})]^2$. Such cases of multiple detection typically represent the pres-

ence of multiple phases or lower signal-to-noise ratio. The corresponding travel time data are then effectively downweighted by this procedure.

Median values of the uncertainties are near $(\sigma_{\text{med}})_P = 0.15$ s for the P set and $(\sigma_{\text{med}})_S = 0.45$ s for the S set. It is important to note, however, that these uncertainties depend on several assumptions (see below). The main assumption is that for any earthquake the determination of the reference time (start time of the reference phase P) is consistent for all stations. If this is not so, station-dependent time shifts are introduced which are the same for all phases. These consequently do not influence relative travel times.

The model covariance matrix \mathbf{C}_x is used to impose the smoothness constraint by setting

$$\mathbf{C}_x(\mathbf{r}, \mathbf{r}') = \sigma_x^2 \mathbf{C}_{x_0}(\mathbf{r}, \mathbf{r}') + \tau^2 \mathbf{I}, \quad (\text{A4})$$

where a spatial correlation length L is implemented by

$$\mathbf{C}_{x_0}(r, r') = e^{-\frac{1}{2} \frac{|r-r'|^2}{L^2}} \quad (\text{A5})$$

Such a spatial correlation was used by *Tarantola and Nercessian* [1984]. While they used the constraint for the spatial variation of a slowness model, we use it for the spatial variation of average lithospheric properties, where $|\mathbf{r}' - \mathbf{r}|$ is the distance between the two stations along the surface. This correlation is imposed independently for path terms \mathbf{d} and each of the station correction terms \mathbf{c} and \mathbf{s} giving a block tridiagonal matrix \mathbf{C}_{x_0} . The second term in (A4) relaxes the spatial correlation imposed by the first term, such that small-scale variation is not entirely prohibited. It also serves to stabilize the inverse model covariance matrix which for $|\mathbf{r}' - \mathbf{r}| = 0$ is degenerate. This case exists for two locations in the experiment, at Red Lake, Ontario (RSON, RLOR, RLOW), and Fordville, North Dakota (FVNR, FVNW), where two or more stations are collocated. A value of $\tau = \sigma_x \times 10^{-2}$ was used in the study. All parameters have units of seconds and roughly similar size. We choose a constant value $\sigma_x = 1$ and vary the trade-off parameter γ instead.

In the presence of the typical trade-off problem, namely, that in general for better data fits, larger model variances have to be tolerated, the typical criterion for choosing an optimum damping parameter γ_{opt} is to use a value which simultaneously minimizes both [Menke, 1984]. Such a criterion is also used in this paper, but we recognize the specific types of errors which are expected for our data. *Parker* [1994] suggests that the choice of γ be made to enforce consistency of the problem, meaning that the data variance χ_ν^2 (first term in equation (A2) for an optimal model \mathbf{x}_{opt} be equal to the number of degrees of freedom $\nu = M - N$ (number of data points M minus the number of model parameters N). This implies a multiplicative rescaling of the data errors [Gubbins and Bloxham, 1985]. In general, deviations from

$\chi_\nu^2 = \nu$ may be caused by (1) underestimated a priori uncertainties in the travel times, (2) inadequate model assumptions, or (3) non-Gaussian noise or more likely a combination of these three. In our case, one prominent type of error arises from errors in the reference time (due to an inappropriate attenuation operator or perhaps to uncorrected instrument timing errors). This would suggest a rescaling of the problem by $\mathbf{C}_d = \mathbf{C}_{d_0} + \rho^2 \mathbf{I}$ with an additional noise variance ρ^2 from the reference time uncertainty. For simplicity, we assume that this ρ^2 is independent of the instrument type. However, we want to retain the important relative error information from \mathbf{C}_{d_0} . This can be done empirically by summing over two terms for which we require that both terms have roughly similar size. We use $\mathbf{C}_d = 1/2\zeta[\mathbf{C}_{d_0} + \sigma_{\text{med}}^2 \mathbf{I}]$ with the median data variance σ_{med}^2 and a parameter ζ needed to render the problem consistent, that is, to make χ^2 approximately equal to the number of degrees of freedom.

For finding the optimum damping parameter γ_{opt} , we use criteria based on the simultaneous minimization of misfit and model norm [Gubbins and Bloxham, 1985]. This eliminates the need for choosing a specific resolution spread function. We use the shape of the trade-off curve to determine the optimum damping parameter interactively which simultaneously minimizes data variance and model norm. Subsequently, we study sensitivity of the results with respect to parameter variation by monitoring the model variation with varying γ . For this purpose, we perform two additional inversions with γ chosen as $\gamma/\gamma_{\text{opt}} = 0.4$ and 2.5 . We regard features in the model as stable features of the data set if they are insensitive against such variation. This condition is equivalent to requiring stability of the model for a change of half an order of magnitude in the relative importance of data and misfit term.

Appendix B: Instrument Bias

As this experiment consisted of two types of instrumentation, it was necessary to check for possible biases in the estimation of travel times. Figure 1 shows that instruments were essentially alternated so that this bias would never lead to large-scale variations, but only an alternating pattern. For two locations (FVNR, RLOR in Figure 1), instruments of different type (1 s and 5 s eigenperiod) were positioned together to test for an instrument bias in the determination of the reference time. In fact, the 5-s instruments give faster P travel times by 0.44 s in the average. For consistency, we therefore reduce all travel time residuals to the level of the 5-s broadband instruments. The S delay is consistent with P but has larger uncertainty due to the smaller number of data. A residual bias would give rise to short-wavelength variations along the transect. We are the most confident in long-wavelength variations across the transect.

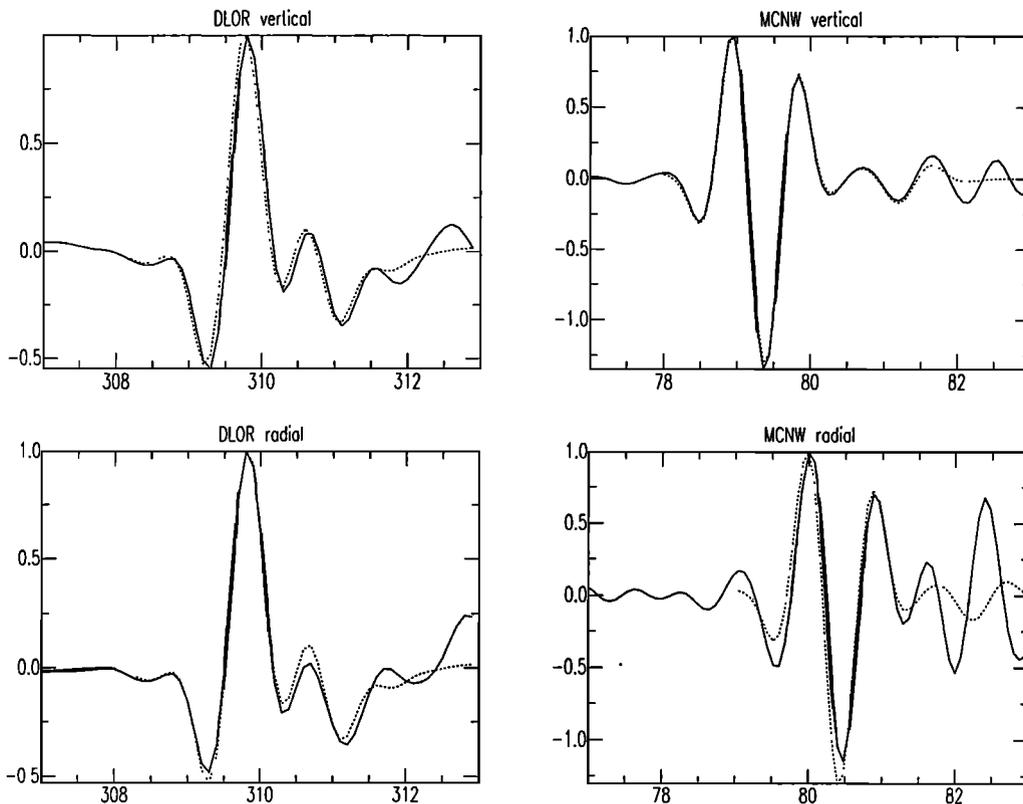


Figure C1. Systematic delay of arrival on radial versus vertical component studied for stations (left) DLOR (no sedimentary cover) and (right) MCNW (~ 2.7 km sedimentary thickness). There is no apparent delay for DLOR, while for MCNW the radial appears to be delayed by ~ 1 s. Observed waveform are shown by solid line and fitted synthetics are shown by dotted lines.

Appendix C: Crustal and Sedimentary Delays

In this appendix we illustrate our hybrid procedure of crustal correction, which uses sedimentary delay estimates from converted waves together with estimates of crustal thickness and average velocity from *Braile et al.* [1989, Figure 5] for the remnant crustal effect. Together these allow recovery of the total P and S transit times through the crust. Due to their large size, it is of particular importance to determine the contributions from the sedimentary layer correctly. Our technique for extracting the sedimentary delay is illustrated in Figure C1, which shows the P wave time window on vertical (Figure C1, top) and radial (Figure C1, bottom) components for two stations. Station DLOR is known to have no sedimentary cover. The arrival time of the P phases determined from fitting of the synthetic seismogram (dotted line) is the same on both components to within one sample interval. On the other hand, station MCNW is located on ~ 2.7 km of sediments. Comparing the vertical and the radial components in Figure C1 (right), we see that the dominant waveform on the radial is delayed by ~ 1 s with respect to the vertical component. For this teleseismic event (day 170) the ra-

dial component waveform is dominated by the P -to- S conversion at the crust-sediment boundary. From this differential time, the sedimentary transit times for P and S phases can be obtained for a given Poisson's ratio. Laboratory measurements of Williston basin sediments [Meissner, 1976] give sedimentary P velocities between 2 and 3 km/s. With sedimentary velocities that low, it is clear that errors in sedimentary thickness affect travel times ~ 5 times more than errors in crustal thickness. For each station except the two westernmost ones, differential travel time estimates through the sedimentary layer are very consistent for different teleseismic events (Figure C2), while the consistency suffers for shallow incident waves. Also the change of sedimentary delay along the transect is remarkably similar to the (approximately known) sedimentary thicknesses. For stations DIWR and GLWW on the Wyoming Craton, there are conflicting data. The reason for this is unclear but may reflect the complexity of its geological history [Egglar and Furlong, 1991]. While this strategy can be successfully applied for the sedimentary delay, total crustal delays are less stable probably due to a weaker material contrast.

To correct travel times for the crustal and sedimentary effects, reduction to a reference level h_{red} is carried

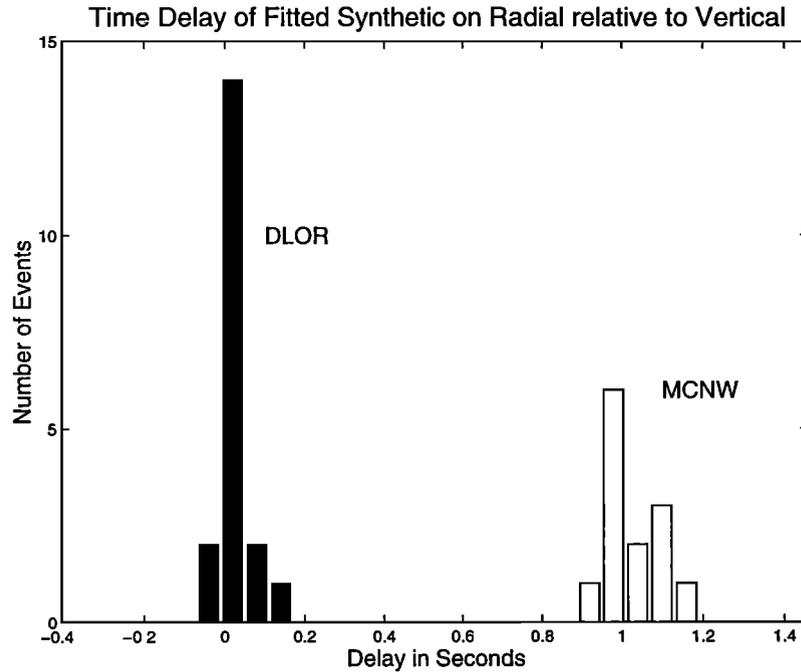


Figure C2. Estimated delay (radial minus vertical) for a large number of distant events, consistently giving delays near 0 for DLOR and near 1 s for MCNW, suggesting that the delay is caused by the P - S conversion at the lower boundary of the near-surface layer. The radial component shows mainly the generated S wave.

out, chosen to lie below the crust-mantle interface, 50 km depth under the mean sea surface. Elevation h_{el} , sediment thickness h_{sed} , crustal thickness h_{cr} , and remnant upper mantle h_{um} (above the reduction level) give rise to individual delay terms. From the crustal average velocity \bar{v}_P [Braille *et al.*, 1989] at each station the total correction can be obtained for the nonsedimentary part of the crust. The correction for P (and similarly for S) is therefore

$$t_{red} = t_{um} + \frac{h_{ns}}{\bar{v}_P} + t_{sed} + t_{el}; \quad (C1)$$

t_{um} is the time spent in the remnant upper mantle with a v_P/v_S ratio of $\sqrt{3}$.

Since for refraction seismological results the reference level is typically the local surface, the thickness of the nonsedimentary crust is $h_{ns} = h_{cr} - h_{sed} - h_{el}$. For a sedimentary Poisson's ratio of 0.30 the crustal correction varies across the transect by ~ 1 s for P and ~ 2.5 s for S . Variation of the Poisson's ratio can affect these correction substantially; that effect is studied further below.

The crustal transit times and in particular the time delay due to the sedimentary layer may have considerable effect on the total delay. Thus we test the dependence of the total P and S wave delay changes across the transect on the Poisson's ratio of the sedimentary rock column. For this purpose, we compute residual mantle delay differences between the fast and the slow regions (stations NAOR and LENR) for a number of

crustal v_P/v_S values for the sediments and display the resulting overall relative velocity variation $d\ln v_S/d\ln v_P$ for the change of upper mantle properties between the shield and the Trans-Hudson Orogen (Figure C3). The larger v_P/v_S is, the smaller the upper mantle variation becomes. For the typically assumed value of $\sqrt{3}$, we get an unrealistically large value of $d\ln v_S/d\ln v_P$ near 3.7. It has been shown in a number of studies addressing this issue that fluid content in rocks usually results in higher values of v_P/v_S of up to 2 and in extreme cases even larger than 2 [Castagna *et al.*, 1985]. The effect is particularly pronounced if the fluid is overpressured. Indeed, sediments under the Williston basin frequently show considerable fluid content and pressure [Meissner, 1976] and are thus in general agreement with the observations.

However, that trade-off exists only for regions with substantial sediment cover. Namely, the inversion for the mantle under the eastern part of the transect, where the strongest transition in shear wave delay (and small transition in P delays) occurs, is not affected by that trade-off, since it has no sediment cover. We regard this discrepancy as the most important feature of the data set.

Appendix D: Effect of Shear Wave Splitting on Travel Times

The effect of shear wave splitting on a waveform $U(\omega)$ can be written as

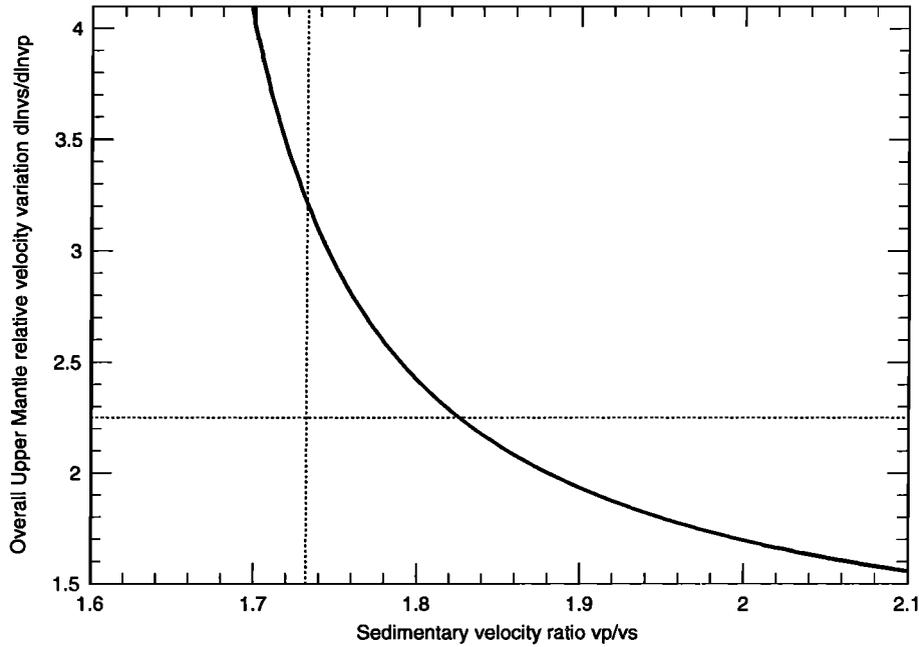


Figure C3. Illustration of the influence of the sedimentary Poisson's ratio on an overall upper mantle relative velocity variation $d\ln v_S/d\ln v_P$ (not considering the upper mantle transition in the Canadian Shield alone). To be within the range explicable by thermal effects ($1 \geq d\ln v_S/d\ln v_P \leq 2.25$) requires an elevated v_P/v_S value for the sediments. That is consistent with expected partial fluid saturation in the sediments. No sediments exist under the eastern part of the transect. Therefore such a trade-off does not exist for the transition within the shield.

$$\mathbf{U}(\omega) = \mathbf{\Gamma}(\omega) \cdot \hat{\mathbf{p}} \quad (D1)$$

using the splitting operator $\mathbf{\Gamma}(\omega)$ and the polarization vector $\hat{\mathbf{p}}$ [Silver and Chan, 1991]. We consider the signal on the observing component $\hat{\mathbf{h}}$ and determine its amplitude r and phase θ :

$$\mathbf{U}_h(\omega) = \hat{\mathbf{h}} \cdot \mathbf{\Gamma}(\omega) \cdot \hat{\mathbf{p}} = r e^{i\omega\tau} \equiv r e^{i\theta}. \quad (D2)$$

Using the splitting operator

$$\mathbf{\Gamma}(\omega) = \hat{\mathbf{f}}\hat{\mathbf{f}}e^{-i\omega\delta t/2} + \hat{\mathbf{s}}\hat{\mathbf{s}}e^{i\omega\delta t/2} \quad (D3)$$

with fast and slow direction vectors $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ and the splitting delay δt , we obtain

$$U_h(\omega) = (\hat{\mathbf{h}} \cdot \hat{\mathbf{f}})(\hat{\mathbf{p}} \cdot \hat{\mathbf{f}})e^{-i\omega\delta t/2} + (\hat{\mathbf{h}} \cdot \hat{\mathbf{s}})(\hat{\mathbf{p}} \cdot \hat{\mathbf{s}})e^{i\omega\delta t/2}. \quad (D4)$$

Defining

$$(\hat{\mathbf{h}} \cdot \hat{\mathbf{f}}) = \cos\phi_h$$

$$(\hat{\mathbf{p}} \cdot \hat{\mathbf{f}}) = \cos\phi_p$$

$$(\hat{\mathbf{h}} \cdot \hat{\mathbf{s}}) = -\sin\phi_h$$

$$(\hat{\mathbf{p}} \cdot \hat{\mathbf{s}}) = -\sin\phi_p$$

we get

$$U_h(\omega) = \cos\phi_h \cos\phi_p e^{-i\omega\delta t/2} + \sin\phi_h \sin\phi_p e^{i\omega\delta t/2}. \quad (D5)$$

With $\theta_o = \omega\delta t/2$, this becomes

$$\begin{aligned} & \cos\phi_h \cos\phi_p [\cos\theta_o - i\sin\theta_o] + \sin\phi_h \sin\phi_p [\cos\theta_o + i\sin\theta_o] \\ &= \cos\theta_o [\cos\phi_h \cos\phi_p + \sin\phi_h \sin\phi_p] \\ & - i \sin\theta_o [\cos\phi_h \cos\phi_p - \sin\phi_h \sin\phi_p]. \end{aligned} \quad (D6)$$

The phase on the observing component is thus given through

$$\begin{aligned} \tan\theta &= -\tan\theta_o \left[\frac{\cos\phi_h \cos\phi_p - \sin\phi_h \sin\phi_p}{\cos\phi_h \cos\phi_p + \sin\phi_h \sin\phi_p} \right] \\ &= -\tan\theta_o \left[\frac{\cos(\phi_h + \phi_p)}{\cos(\phi_h - \phi_p)} \right]. \end{aligned} \quad (D7)$$

Defining $\phi_p = \phi_h + \delta\phi$, this can be rewritten

$$\tan\theta = -\tan\theta_o \left[\frac{\cos(2\phi_h + \delta\phi)}{\cos(\delta\phi)} \right]. \quad (D8)$$

Note that $\theta_o = 0$ ($\delta t = 0$) corresponds to the time midway between the two split shear waves. In the special case where $\delta\phi = 0$, and for small θ_o , the apparent travel time delay due to splitting corresponds to a weighted sum of the delays of the two split shear waves. For more general nonzero values of $\delta\phi$, there is not such a simple relation. The travel time delay can even exceed the time delays between the split shear waves.

Acknowledgments. We would first like to acknowledge the DTM and University of Wisconsin field crews, especially Randy Kuehnel and Hans Meyer, who made this data set possible. Logistical support is acknowledged by the Ontario Geological Survey (Greg Stott) and the University of North Dakota at Grand Forks (Will Gosnold). The first author thanks Bob Phinney for support under NSF grant EAR-8904050 and DTM under grant EAR-91050607. We also thank Cecily Wolfe for help in deriving the expressions in Appendix D. We wish to thank Steve Shirey, Stephen Grand, George Helffrich, Satoshi Kaneshima, Bob Phinney, Tim Clarke, and Cecily Wolfe for helpful discussions and Randy Kuehnel for help in processing the data. The manuscript was greatly improved through the numerous comments of Ray Russo, Susan van der Lee, and an anonymous reviewer.

References

- Ahern, J. L., S.R. Mrkvicka, A mechanical and thermal model for the evolution of the Williston Basin, *Tectonics*, **3**, 79-102, 1984.
- Anderson, D.L., *Theory of the Earth*, Blackwell Sci., Malden, Lass, 1989.
- Anderson, O.L. and D.G. Isaak, Elastic constants of mantle minerals at high temperature, in *Mineral Physics and Crystallography: A Handbook of Physical Constants*, AGU Ref. Shelf, vol. 2, edited by T.J. Ahrens, pp. 64-97, AGU, Washington, D.C., 1995.
- Bally, A.W., C.R. Scotese and M.I. Ross, North America: Plate-tectonic setting and tectonic elements, in *The Geology of North America* vol. A.,: *The Geology of North America-An Overview*, edited by A. W. Bally and A. R. Palmer, pp. 1-15, Geol. Soc. of Am., Boulder, Colo., 1989.
- Bamford, D., P_n velocity anisotropy in a continental upper mantle, *Geophys. J. R. Astron. Soc.*, **49**, 29-48, 1977.
- Barruol, G., P.G. Silver and A. Vauchez, Seismic anisotropy in the eastern United States: Deep structure of a complex continental plate, *J. Geophys. Res.*, **102**, 8329-8348, 1997.
- Bird, P., Laramide crustal thickening event in the Rocky Mountain foreland and Great Plains, *Tectonics*, **3**, 741-758, 1984.
- Bokelmann, G.H.R., Upper and lower mantle small-scale heterogeneity studied by systematic analysis of portable broadband waveforms and traveltimes, Ph.D. thesis, Princeton Univ., Princeton, N.J., 1992.
- Bokelmann, G.H.R. and P.G. Silver, Later-phase body wave traveltimes, *Eos Trans. AGU*, **72** (17), Spring Meet. Suppl., 194, 1991a.
- Bokelmann, G.H.R. and P.G. Silver, A procedure for the extraction of multiple body wave traveltimes from seismological waveform data, in *Proceedings of the CP90 European Conference on Computational Physics*, edited by A. Tenner, pp. 276-283, World Sci., River Edge, N.J., 1991b.
- Bokelmann, G.H.R. and P.G. Silver, The Caribbean anomaly: Short-wavelength lateral heterogeneity in the lower mantle, *Geophys. Res. Lett.*, **20**, 1131-1134, 1993.
- Braile, L.W., W.J. Hinze, R.R.B. von Frese and G.R. Keller, Seismic properties of the crust and upper-most mantle of the conterminous United States and adjacent Canada, in *Geophysical Framework of the continental United States*, edited by L.C. Pakiser and W.D. Mooney, *Memoir Geol. Soc. Am.* **172**, pp. 655-680, 1989.
- Buchbinder, G.G.R. and G. Poupinet, P -wave residuals in Canada, *Can. J. Earth Sci.*, **14**, 1292-1304, 1977.
- Cannon, W.F. and S.W. Nicholson, Middle Proterozoic rift system, in *Archean and Proterozoic Geology of the Lake Superior Region, U.S.A.*, edited by P.K. Sims and L.M.H. Carter, *U.S. Geol. Surv. Prof. Pap.*, **1556**, pp. 60-67, 1996.
- Castagna, J.P., M.L. Batzle and R.L. Eastwood, Relationships between compressional-wave and shear wave velocities in clastic silicate rocks, *Geophysics*, **50**, 571-581, 1985.
- Christensen, N.I., The magnitude, symmetry, and origin of upper mantle anisotropy based on fabric analysis of ultramafic tectonics, *Geophys. J. R. Astron. Soc.*, **76**, 89-112, 1984.
- Clarke, T.J. and P.G. Silver, A procedure for the systematic interpretation of body wave seismograms, 1, Application to moho depth and crustal properties, *Geophys. J. Int.*, **104**, 41-72, 1991.
- Cleary, J. and A.L. Hales, An analysis of the travel times of P waves to North American stations, in the distance range 32° to 100° , *Bull. Seismol. Soc. Am.*, **56**, 467-489, 1966.
- Crcager, K.C. and T.H. Jordan, Slab penetration into the lower mantle, *J. Geophys. Res.*, **89**, 3031-3049, 1984.
- Dziewonski, A.M. and D.L. Anderson, Preliminary reference Earth model (PREM), *Phys. Earth Planet. Inter.*, **25**, 297-356, 1981.
- Dziewonski, A.M. and D.L. Anderson, Travel times and station corrections for P waves at teleseismic distances, *J. Geophys. Res.*, **88**, 3295-3314, 1983.
- Dziewonski, A.M. and F. Gilbert, The effect of small, aspherical perturbations on travel times and a re-examination of the corrections for ellipticity, *Geophys. J. R. Astron. Soc.*, **44**, 7-7, 1976.
- Dziewonski, A.M. and J.H. Woodhouse, Global images of the Earth's interior, *Science*, **236**, 37-48, 1987.
- Dziewonski, A.M. and B.H. Hager, R.J. O'Connell, Large-scale heterogeneities in the lower mantle, *J. Geophys. Res.*, **82**, 239-255, 1977.
- Eggler, D.H. and K.P. Furlong, Destruction of subcratonic mantle keel: The Wyoming Province, paper presented at the fifth International Kimberlite Conference, Araxa, Brazil, 1991.
- Fowler, C. M. R. and E.G. Nisbet, The subsidence of the Williston Basin, *Can. J. Earth Sciences*, **22**, 408-415, 1985.
- Gaherty, J.B. and T.H. Jordan, Lehmann discontinuity as the base of an anisotropic layer beneath continents, *Science*, **268**, 1468-1471, 1995.
- Grand, S.P., Tomographic inversion for shear velocity beneath the North American plate, *J. Geophys. Res.*, **92**, 14,065-14,090, 1987.
- Grand, S.P., Mantle shear structure beneath the Americas and surrounding oceans, *J. Geophys. Res.*, **99**, 11,591-11,621, 1994.
- Grand, S.P. and D.V. Helmberger, Upper-mantle shear structure of North America, *Geophys. J. R. Astron. Soc.*, **76**, 399-438, 1984.
- Gubbins, D. and J. Bloxham, Geomagnetic field analysis, III, magnetic fields on the core-mantle boundary, *Geophys. J. R. Astron. Soc.*, **80**, 695-713, 1985.
- Hajnal, Z., S. Lucas, D. White, J. Lewry, S. Bezdan, M.R. Stauffer and M.D. Thomas, Seismic reflection images of high-angle faults and linked detachments in the Trans-Hudson Orogen, *Tectonics*, **15**, 427-439, 1996.
- Hales, A.L. and J.L. Roberts, The travel times of S and SKS , *Bull. Seismol. Soc. Am.*, **60**, 461-489, 1970.
- Herrin, E.W., J. Tucker, J. Taggart, D.W. Gordon and J.L. Lobdell, Estimation of surface focus P travel times, *Bull. Seismol. Soc. Am.*, **58**, 1273-1291, 1968.
- Hess, H.H., Seismic anisotropy of the uppermost mantle under oceans, *Nature*, **203**, 629-631, 1964.

- Hill, L., The elastic behavior of a crystalline aggregate, *Proc. Phys. Soc. London, Ser. A*, 65, 349-354, 1952.
- Hoffman, P.F., Precambrian Geology and tectonic history of North America, in *The Geology of North America* vol. A: *The Geology of North America-An Overview*, edited by A. W. Bally and A. R. Palmer, pp. 447-512, Geol. Soc. of Am., Boulder, Colo., 1989.
- Jordan, T.H., Mineralogies, densities and seismic velocities of garnet lherzolites and their geophysical implications, in *Proceedings of the Second International Kimberlite Conference*, edited by F.R. Boyd and H.O.A. Meyer, pp. 1-14, AGU, Washington, D.C., 1981.
- Jordan, T.H., Structure and formation of the continental tectosphere, *J. Petrol.*, 1, 11-37, 1988.
- Jordan, T.H. and J.N. Franklin, Optimal solutions to a linear inverse problem in geophysics, *Proc. Nat. Acad. Sci. USA*, 68, 291-293, 1971.
- Kern, H. and H.R. Wenk, Fabric related velocity anisotropy and shear wave splitting in rocks from the Santa Rosa mylonite zone, California, *J. Geophys. Res.*, 95, 11,213-11,223, 1990.
- Klasner, J.S. and E.R. King, Precambrian basement geology of North and South Dakota, *Can. J. Earth Sci.*, 23, 1083-1102, 1986.
- Kumazawa, M. and O.I. Anderson, Elastic moduli, pressure derivatives, and temperature derivatives of single-crystal olivine and single-crystal forsterite, *J. Geophys. Res.*, 74, 5961-5972, 1969.
- Mainprice, D. and P.G. Silver, Interpretation of SKS waves using samples from the subcontinental lithosphere, *Phys. Earth Planet. Inter.*, 78, 257-280, 1993.
- Meissner, F.F., Abnormal electric resistivity and fluid pressure in Bakken formation, Williston Basin, and its relation to petroleum generation, migration, and accumulation, *AAPG Bull.*, 60, 1403-1404, 1976.
- Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic, San Diego, Calif., 1984.
- Nelson, K.D., D.J. Baird, J.J. Walters, M. Hauck, L.D. Brown, J.E. Oliver, J.L. Ahern, Z. Hajnal, A.G. Jones and L.L. Sloss, Trans-Hudson Orogen and Williston basin in Montana and North Dakota: New COCORP deep profiling results, *Geology*, 21, 447-450, 1993.
- Nicolas, A. and N.I. Christensen, Formation of anisotropy in upper mantle peridotites-A review, in *Composition, Structure and Dynamics of the Lithosphere-Asthenosphere System, Geodyn. Ser.*, vol. 16., edited by K. Fuchs, pp. 111-123, AGU, Washington, D.C., 1987.
- Nolet, G. and A. Zielhuis, Low *S* velocities under Tornquist-Teisseyre zone: Evidence for water injection into the transition zone by subduction, *J. Geophys. Res.*, 99, 15,813-15,820, 1994.
- Parker, R.L., *Geophysical Inverse Theory*, Princeton Univ. Press, Princeton, N.J., 1994.
- Pollack, H.N. and D.S. Chapman, On the regional variation of heat flow, geotherms, and lithospheric thickness, *Tectonophysics*, 38, 279-296, 1977.
- Romanowicz, B.A., Seismic structure of the upper mantle beneath the United States by three-dimensional inversion of body wave arrival times, *Geophys. J. R. Astron. Soc.*, 57, 479-506, 1979.
- Roy, R.F., D.D. Blackwell and E.R. Decker, Continental heat flow, in *The Nature of the Solid Earth*, edited by J.F. Hays and L. Knopoff, pp. 506-543, McGraw-Hill, New York, 1972.
- Sayers, C.M., *P*-wave propagation in weakly anisotropic media, *Geophys. J. Int.*, 116, 799-805, 1994.
- Silver, P.G., Seismic anisotropy beneath the continents: Probing the depths of geology, *Annu. Rev. Earth Planet. Sci.*, 24, 385-432, 1996.
- Silver, P.G. and W.W. Chan, Implications for continental structure and evolution from seismic anisotropy, *Nature*, 335, 34-39, 1988.
- Silver, P.G. and W.W. Chan, Shear wave splitting and subcontinental mantle deformation, *J. Geophys. Res.*, 96, 16,429-16,454, 1991.
- Silver, P.G. and S. Kaneshima, The APT89 seismic experiment: Shear wave splitting results, *Geophys. Res. Lett.*, 20, 1127-1130, 1993.
- Silver, P.G., R.P. Meyer and D.E. James, Intermediate-scale observations of the Earth's deep interior from the APT89 transportable teleseismic experiment, *Geophys. Res. Lett.*, 20, 1123-1126, 1993.
- Silver, P.G. and S. Gao, Kaapvaal Working Group, Mantle anisotropy beneath southern Africa from shear wave splitting: Preliminary results from the Southern Africa Seismic Experiment, *EOS Trans. AGU*, 79, (45), Fall Meet. Suppl., F579, 1998.
- Sims, P.K., Mafic dike swarms, in *Archean and Proterozoic Geology of the Lake Superior Region, U.S.A.*, edited by P.K. Sims and L.M.H. Carter, U.S. Geol. Surv. Prof. Paper, 1556, pp. 71-74, 1996.
- Sipkin, S.A. and T.H. Jordan, Frequency dependence of Q_{SCS} , *Bull. Seismol. Soc. Am.*, 69, 1055-1079, 1979.
- Sleep, N., Teleseismic *P*-wave transmission through slabs, *Bull. Seismol. Soc. Am.*, 63, 1349-1373, 1973.
- Spakman, W., S. van der Lee, R. van der Hilst, Travel-time tomography of the European-Mediterranean mantle down to 1400 km, *Phys. Earth. Planet. Inter.*, 79, 3-74, 1993.
- Tarantola, A. and A. Necessian, Three-dimensional inversion without blocks, *Geophys. J. R. Astron. Soc.*, 76, 299-306, 1984.
- van der Hilst, R.D., S. Widiyantoto and E.R. Engdahl, Evidence for deep mantle circulation from global tomography, *Nature*, 386, 578-584, 1997.
- van der Lee, S., The Earth's upper mantle: its structure beneath North America and the 660-km discontinuity beneath northern Europe, Ph.D. thesis, Princeton Univ., N.J., 1996.
- van der Lee, S. and G. Nolet, Upper mantle *S* velocity structure of North America, *J. Geophys. Res.*, 102, 22815-22838, 1997.
- Wickens, A.J. and G.G.R. Buchbinder, *S*-wave residuals in Canada, *Bull. Seismol. Soc. Am.*, 70, 809-822, 1980.
- Woodhouse, J.H. and A.M. Dziewonski, Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, 89, 5953-5986, 1984.
- Zhang Y.-S. and T. Tanimoto, Global Love wave phase velocity variation and its significance to plate tectonics *Phys. Earth Planet. Inter.*, 66, 160-202, 1990.
- Zielhuis, A. and G. Nolet, Shear wave velocity variations in the upper mantle beneath central Europe, *Geophys. J. Int.*, 117, 695-715, 1994.

G. H. R. Bokelmann, Department of Geophysics, Stanford University, Stanford, CA 94305-2215. (goetz@pangea.stanford.edu)

P. G. Silver, Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, Washington, DC 20015. (silver@pssl.ciw.edu)

(Received December 16, 1998; revised July 29, 1999; accepted October 27, 1999.)