Evidence for small-scale mantle convection in the upper mantle beneath the Baikal rift zone

Stephen S. Gao and Kelly H. Liu
Department of Geology, Kansas State University, Manhattan, Kansas, USA

Paul M. Davis and Phillip D. Slack
Department of Earth and Space Sciences, University of California, Los Angeles, California, USA

Yuliy A. Zorin, Valentina V. Mordvinova, and Vladimir M. Kozhevnikov
Institute of the Earth’s Crust, Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia

Received 18 June 2002; revised 2 October 2002; accepted 28 January 2003; published 11 April 2003.

[1] Inversion of teleseismic P wave travel time residuals collected along a 1280-km-long profile traversing the Baikal rift zone (BRZ) reveals the existence of an upwarped lithosphere/asthenosphere interface, which causes a travel time delay of about 1 s at the rift axis (“central high”). An area with early arrivals relative to the stable Siberian platform of up to 0.5 s is observed on each side of the rift, about 200 km from the rift axis (“flank lows”). While the location of the central high is approximately fixed in the vicinity of the rift axis, those of the flank lows vary as much as 200 km with the azimuth of the arriving rays. We use three techniques to invert the travel time residuals for velocity anomalies beneath the profile. Two of the techniques assume an isotropic velocity structure, and one of them considers a transversely isotropic velocity model with a vertical axis of symmetry. We use independent geophysical observations such as gravity, active source seismic exploration, and crustal thickness measurements to compare the applicability of the models. Other types of geophysical measurements suggest that the model involving transverse isotropy is a plausible one, which suggests that the central high and flank lows are caused by the combined effects of an upwarped asthenosphere with a 2.5% lateral velocity reduction, and a velocity increase due to transverse isotropy with a vertical axis of symmetry. We consider the anisotropy to be the result of the vertical component of a lithosphere/asthenosphere small-scale mantle convection system that is associated with the rifting.

INDEX TERMS: 7218 Seismology: Lithosphere and upper mantle; 7203 Seismology: Body wave propagation; 8109 Tectonophysics: Continental tectonics—extensional (0905); 8120 Tectonophysics: Dynamics of lithosphere and mantle—general; KEYWORDS: Baikal rift, tomography, anisotropy, lithosphere, asthenosphere


1. Introduction

[2] A continental rift is a region where the lithosphere is extending and is usually marked by a rift valley. Continental rifting is the first stage in a Wilson cycle, although some rifts never evolve into an oceanic basin [e.g., Turcotte and Schubert, 1982]. Consequently, the seismic velocity and thermal structures beneath a continental rift show a remarkable resemblance to those of an oceanic rift [Bjarnason et al., 1996].

[3] The mechanisms of continental rifting can be separated into two end-members called passive and active rifting. Passive rifting is the result of extensional or shearing stresses that originate beyond the vicinity of the resulting rift (see, e.g., Tapponnier and Molnar [1979] for the Baikal rift zone (BRZ)), and active rifting is the result of active intrusion of an asthenospheric diapir (see, e.g., Turcotte and Emerman [1983], Logatchev and Zorin [1992], and Gao et al. [1994a] for BRZ). Geodynamic modeling has been used to suggest that a small-scale mantle convection system would develop beneath a rift formed by either mechanism [Turcotte and Emerman, 1983; Steckler, 1985; Anderson, 1994; King and Anderson, 1995, 1998; Huismans et al., 2001]. Major continental rifts often form at the edge of cratons in regions that have undergone transpressional tectonics in the past. It has been suggested that the associated juxtaposition of cold cratonic and warm orogenic lithosphere would cause small-scale convection in the mantle, and the subsequent rifting could lead to formation.
of flood basalts [Anderson, 1994; King and Anderson, 1995, 1998]. Recently, Huismans et al. [2001] calculated how passive extension can destabilize the mantle lithosphere resulting in upward doming of the asthenosphere and adjacent down warping of the lithosphere into the asthenosphere. Such models can be tested using teleseismic tomography. Because the finite strains associated with the convection are expected to generate seismic anisotropy in the mantle from lattice preferred orientation (LPO) of olivine crystals (see, e.g., Blackman et al. [1996] for mid-ocean ridge flows), modification of the traditional isotropic tomographic inversions to include anisotropy is required. In this paper, we analyze teleseismic P wave travel time residuals obtained across the Baikal rift and compare isotropic and anisotropic tomographic inversions with the structures expected from small-scale convection. We conclude that small-scale convection probably exists beneath the rift and has given rise to anisotropic structure in the underlying asthenosphere.

2. Baikal Rift Zone

The Baikal rift zone in Siberia is a major continental rift zone. The 1500-km-long en echelon system of rift depressions, which originated about 30 Ma along the Paleozoic suture between the Siberian and Amurian microplates, is the most seismically active continental rift in the world (Figure 1). Previous studies revealed that it has all the common features of a typical continental rift. These features [e.g., Turcotte and Schubert, 1982] include (1) a subsided central valley and uplifted adjacent blocks; (2) flanking normal faults; (3) negative Bouguer gravity anomalies [Zorin et al., 1989]; (4) higher than normal heat flow [Lysak, 1984]; and (5) shallow, tensional and higher than normal seismicity [Doser, 1991]. Another feature of most continental rifts is the thinning of the crust beneath the rift valley [Davis, 1991]. Deep seismic sounding experiments reveal that beneath the BRZ, the thinning is no more than 5 km [Puzyrkov, 1993], which is significantly smaller than that beneath other major rifts. A recent study from stacking of teleseismic receiver functions [Zachary et al., 2000] reveals a dramatic change in Moho depth, from about 37 km beneath the Siberian Craton to about 45 km beneath the fold belt south of the rift. The change takes place over a distance of less than 20 km. The measurements suggest that the Baikal rift zone was formed near this zone of sudden change in Moho depth, which, along with a change in lithospheric thickness and age, is probably a zone of weakness.

Over the last 30 years, the deep-seated structure beneath the Baikal rift zone and its adjacent regions has been studied by using various geophysical techniques, such as deep seismic sounding [Puzyrkov et al., 1978], gravimetric investigations [Zorin et al., 1989], modeling of heat flow [Lysak, 1984, 1987; Zorin and Osokina, 1984; Zorin and Lepina, 1985], seismic spectral ratio methods [Mordvinova, 1983, 1988], magnetotelluric measurements [Popov, 1990], teleseismic travel time tomography [Gao et al., 1994b; Gao, 1995], and shear wave splitting [Gao et al., 1994a, 1997, 1999].

Gravity and seismic studies suggest that the lithospheric thickness beneath the rift zone is about 40–50 km; beneath the Siberian platform it increases to 200 km; and in the Mongolian foldbelt it ranges from 75 to 160–175 km [Zorin et al., 1989; Logatchev and Zorin, 1992; Egorkin et al., 1984]. Magnetotelluric experiments indicate that in the southern part of the Baikal rift zone, the depth of a mantle conductive layer, which was inferred to be the asthenosphere, was found to be at about 110 km depth [Popov, 1990; Kiselev and Popov, 1992]. Using teleseismic data recorded along an E-W profile across the central part of Lake Baikal, Gao et al. [1994b] suggest that the asthenospheric upwarp has an asymmetric shape with the NW edge being steeper. They also find that at a distance of about 300 km NW of the rift axis, the thickness of the lithosphere beneath the Siberian platform is about 100 km thick and appears to be continuously increasing at the end of the profile.

Currently there is a debate on the extent of thermal modification of the rifted lithosphere beneath Lake Baikal. A broad asymmetrical region of low Bouguer gravity extending well beyond the boundaries of the lake was interpreted as arising from thinned lithosphere, with greatest thinning to the southeast [Zorin et al., 1989] beneath the Mongolian fold belt. Our earlier study [Gao et al., 1994b] reported teleseismic travel time delays which correlate with the gravity, suggesting an asymmetric asthenospheric upwarp that peaks under Lake Baikal but is greater under the Mongolian fold belt than beneath the Siberian craton to the northwest. A subsequent study [Petit et al., 1998] using regional events rather than teleseismic finds diametrically opposite results, with high velocities beneath the lake and fold belt, and low velocities in the mantle beneath the Siberian craton. Those authors suggest a narrow mantle plume reaches the bottom of the craton and follows its border in the Baikal area. A number of other workers argue for cold strong lithosphere beneath the rift based on the presence of a deep seismogenic zone [Deverchere et al., 2001], or large values of the effective elastic thickness from flexure calculations using the gravity field [Diamant and Kogan, 1990; Ruppel et al., 1993; Petit et al., 1997]. In particular, Petit et al. [1997] find an elastic thickness of approximately 30–50 km, and infer the mantle is strong to a depth of about 85 km. In this report we confirm the earlier result [Gao et al., 1994b] that the lowest mantle velocities lie directly beneath the rift, and are at shallow depth. We find no evidence for low velocity associated with a plume under the Siberian craton. Given the low velocity, low gravity, low Q [Gao et al., 1994b] and limited volcanism [Kiselev, 1987], our new results are compatible with sub-solidus, thermally modified mantle.

3. Data

The teleseismic data set used in the study was collected by 28 short-period, three-component seismographs deployed along a 1280 km profile traversing the Siberian platform, Baikal rift zone, and the Mongolian fold
Our analysis is based on tomographic inversion of teleseismic $P$ wave travel time residuals. In order to obtain those residuals, we first filter the seismograms in the 0.2–1.5 Hz frequency band, and manually pick the onset of the first arrival. Only seismograms with a clear onset are used, and those from an event are not used if the number of high-quality travel time picks is less than 14 (out of a maximum of 28). A total of 1370 travel time picks from 77 events (Figure 2) in the epicentral distance from $28^\circ$ to $93^\circ$ are used to invert for velocity structures beneath the profile. We then correct the observed arrival times with theoretical arrival times calculated using the IASP91 Earth model [Kennett and Engdahl, 1991] and the EHB catalog [Engdahl et al., 1995]. Relative residuals are obtained by subtracting the event’s mean residual from the raw residuals.

We further correct the relative residuals by removing the slope on the residuals for each event, which is most likely caused by mislocation of the events instead of velocity gradients in the crust or upper mantle. The main evidence for this conclusion is that while most of the events located in non-subduction zone areas such as the western Unite States show near-zero slopes, events with slopes of larger than 0.2 ms/km are all located in subduction zones, where the near-source stations used to locate these events are almost all on one side of the event, and hence a systematic mislocation is plausible [Dziewonski and Anderson, 1981]. The error in the position and the origin time must be compensated by an equivalent error in depth, which can lead to a location-dependent tilt of a derived travel time curve [Dziewonski and Anderson, 1981]. We calculated that for an earthquake of intermediate focal depth, at $40^\circ$ from the center of the array, a 50-km depth mislocation causes a slope error in the travel time residual curves of 0.3 ms/km [Gao et al., 1994b]. Anisotropy in the subduction area can also cause systematic mislocation of events when an isotropic model is used for earthquake location [e.g., Kendall and Thomson, 1993]. The final relative travel time residuals after these corrections are shown in Figure 3.

To study event location dependence of the relative travel time residuals, we group the events by their back azimuth ($\phi$) and epicentral distance ($\Delta$) relative to station 13,
which is located near the south shore of Lake Baikal. We divide the source regions on the Earth’s surface into 24 areas. Events within area \((i, j)\) satisfy
\[
i/C_0^1 \times 30/C_1^1/C_2^0 < i \times 30/C_1^1;
\]
\(i = 1, 2, \ldots, 12\). \((1)\)

and
\[
(j - 1) \times 60^\circ < \Delta < j \times 60^\circ, (j = 1, 2). \quad (2)
\]

The 77 events used in the study occurred within 12 of the 24 areas (Figure 2). The number of events within each group ranges from 1 to 20. To minimize the dominance of the SE event groups in the final velocity model, we calculate the average relative travel time residuals from each group (Figure 4) and use those averaged residuals in the inversions below (except for the block inversion which uses the individual residuals). Most of the groups show a positive residual (“central high”) of 0.2–1.0 s in the distance range –50 to 50 km (Figure 4). In addition, all but group C, which is the closest event group to the profile, show an area of relatively early arrivals (“flank lows”) on each side of the central high. The location of the lowest point in the southern flank lows ranges from 100 km to 300 km from station 13, and that for the northern ones ranges from –350 to –100 km (Figure 4c). A systematic pattern can be observed when the locations are plotted against the mean back azimuth of the event groups. Figure 4c indicates that the flank lows shift northwestward for events from the southeast, and southeastward for those from the northwest. In spite of the systematic variation of the locations with back azimuth, the distance between the two flank lows remains approximately constant regardless of the back azimuth (Figure 4).

4. Inversion of Travel Time Residuals

We used three techniques to invert the travel time residuals to obtain P wave velocity structures beneath the profile. The first technique assumes that the residuals are mostly caused by the lithosphere/asthenosphere interface and solves for the spatial variation of that interface. The second technique is the standard ACH block inversion scheme [Aki et al., 1977], and the third technique is basically a modification of the first one by introducing transverse isotropy with a vertical symmetry axis in the asthenosphere.

4.1. Lithosphere/Asthenosphere Interface: Isotropic Model

Under the assumption that the travel time residuals are primarily caused by the spatial variation of the depth of the lithosphere/asthenosphere interface, we construct an
The best fitting parameters are $a_1$, the horizontal location relative to station 13, $d$ is the depth of the cylinder, $\theta$ is the angle of incidence, $\phi$ is the back azimuth relative to the SE strike of the rift axis. The two flank lows on travel time residual curves on each side of the rift, the approximately constant distance between them, and the large and systematic location shift for events from different azimuths (Figure 4) can be caused by two high-velocity structures (among other possibilities as discussed below). To estimate the scale of such structures, we search for the optimal depth ($d$) and location ($x_0$) of a rift-parallel high-velocity cylinder on each side of the rift by fitting the observed azimuthal dependence of the locations (Figure 4c) using $x = x_0 - d \tan(\theta) \cos(\phi)$, where $x$ is the observed location of the travel time lows associated with the cylinders, $x_0$ is the horizontal location relative to station 13, $d$ is the depth of the cylinder, $\theta$ is the angle of incidence, $\phi$ and is the back azimuth relative to the SE strike of the profile. The best fitting parameters are $x_0 = -220$ km and $d = 235$ km for the northern cylinder, and $x_0 = +205$ km and $d = 160$ km for the southern cylinder. The calculated locations using those parameters are well matched by the observed ones (Figure 5).

[16] Studies in the BRZ [e.g., Logatchev and Zorin, 1992; Gao et al., 1994b] and in other rift zones [e.g., Parker et al., 1984; Dahlin, 1979; Davis, 1991; Davis et al., 1984, 1993; Slack et al., 1994, 1996] reveal that the velocity contrast between the upwarped asthenosphere and the velocity of the surrounding areas is about 3–12%. In addition, the location of the highest point of the upwarped asthenosphere, which might have reached the Moho, may or may not be directly beneath the rift axis.

[17] Given the above a priori information, we represent the geometry of the lithosphere/asthenosphere interface using the combination of a cosine and a Gaussian function, i.e.,

$$f(x) = -h_0 + f_1(x) \ast f_2(x),$$

where

$$f_1(x) = \left\{ \begin{array}{ll} a_1 \cos(2\pi x/\lambda) & |x| \leq 3\lambda/4 \\ 0 & \text{elsewhere} \end{array} \right.$$ \hspace{1cm} (4)

and

$$f_2(x) = \exp(-0.5x^2/\sigma^2).$$ \hspace{1cm} (5)

where $h_0$ is the depth of the lithosphere outside the anomalous region, $a_1$ is the magnitude of the upwarp, $\lambda$ is the wavelength of the cosine function, $\sigma$ and is the standard deviation of the Gaussian function.

[18] Numerical tests show that $f(x)$ is a function with great flexibility. Some of the features of $f(x)$ include (1) the flank lows occur at $|x| \leq \lambda/2$; (2) when $\lambda/\sigma > 6$ the magnitude of the two flank lows reduces to nearly zero; and (3) the magnitude of the central high $\geq$ that of the flank lows.

[19] To allow for the possible asymmetric shape of the upwarp, we give each side an independent $\lambda$ and $\sigma$. The strike and location of the vertex line of the two-dimensional structure are also treated as unknown parameters.

[20] In summary, there are nine unknown parameters to be found through nonlinear inversion. They are (1) $a_1$, magnitude of the upwarp; (2) $\lambda_1$, wavelength of the left cosine function; (3) $\sigma_1$, standard deviation of the left Gaussian function; (4) $\lambda_2$, wavelength of the right cosine function; (5) $\sigma_2$, standard deviation of the right Gaussian function; (6) $\gamma$, asthenospheric-lithospheric velocity contrast; (7) $\phi_0$, strike of the structure measured anticlockwise from the east; (8) $b$, location of the vertex line of the 2-D structure; and (9) $h_0$, depth of normal lithosphere.

[21] We use a three-dimensional downward projection method [Davis et al., 1984; Gao et al., 1994b] and a nonlinear Bayesian inversion technique [Jackson and Mat-su’ura, 1985; Jackson, 1972] to estimate the parameters. The method assumes straight rays and a plane wave approximation. The resulting parameters are (1) $a_1 = 153 \pm 10$ km; (2) $\lambda_1 = 518 \pm 15$ km; (3) $\sigma_1 = 171 \pm 10$ km; (4) $\lambda_2 = 505 \pm 25$ km; (5) $\sigma_2 = 132 \pm 10$ km; (6) $\gamma = (2.4 \pm 0.05)\%$; (7) $\phi_0 = 31.6 \pm 0.7^\circ$; (8) $b = 14 \pm 25$ km; and (9) $h_0 = 198 \pm 25$ km. The resulting lithosphere/asthenosphere interface is shown in Figure 6.

[22] Some main features of the interface include the following: (1) the asthenosphere reaches to $h_0 - a_1 = 45 \pm 15$ km depth; (2) the magnitude of the flank lows is about 60 km for the left one, and 40 km for the right one; (3) the velocity contrast between the lithosphere and the asthenosphere is 2.4%; (4) the strike of the structure is $31.6^\circ$ measured anticlockwise from the east, which is smaller than the general direction of the strike of the BRZ ($55^\circ$), but it is close to the local strike of the rift; and (5) the vertex line of the 2-D structure is about 14 km south of station 13, i.e., close to the rift axis.

4.2. Block Inversion

We next invert the travel time residuals using the 3-D ACH block inversion method [Aki et al., 1977]. The layer thicknesses are chosen so that the time a ray spends in each

Figure 5. Observed (circles) and fitted (triangles) locations of the two flank lows. $X_0$ and depth are the optimal locations and depths of the imaginary rift-parallel high-velocity cylinders.
layer is approximately the same for all layers. Because the mean station spacing is about 50 km, the horizontal dimensions of the blocks are chosen as 80 km. To determine the optimum depth for the tomography model, we ran a series of inversions with models of different depth. We then plotted the misfits, which are defined as 

\[ \xi = \sum_{i}^{N} (t_i - t^m_i)^2 / (N - M), \]

where \( N \) is the number of data points, \( M \) is the number of parameters, and \( t_i \) and \( t^m_i \) are the observed and fitted travel time residuals, against the depths, and we found that the misfits stopped decreasing significantly when depth \( \geq 400 \) km. Thus we choose 410 km, which is the mean global depth of the bottom of the upper mantle, as the bottom of the area with lateral velocity variations. The optimum damping parameter, which controls the smoothness of the resulting velocity model, was determined by running a series of models with damping parameters from 10 to 200. Obviously, a larger damping parameter corresponds to a larger misfit. We chose the optimum damping parameter as the one at which the slopes of the misfits versus damping parameters curve changes significantly. It turns out that the value is about 56.

[24] The resulting velocity slices are shown in Figures 7 and 8. A low-velocity body is observed in the vicinity of the rift axis in the top 300 km of the Earth, and a high-velocity body is observed on each side of the rift in the top 210 km. In the 125–210 km depth range, the width of the low-velocity body is about 200 km. The centers of the high-velocity bodies are located at about 200 km from the rift axis. The maximum velocity contrast is about 3%. These results are consistent with those obtained under the assumption that most of the travel time residuals are caused by the spatial variation in the depth of the lithosphere/asthenosphere interface (Figure 6).

[25] Also shown on Figures 7 and 8 are the resolution and uncertainty contours from the inversion, which indicate the regions of the model that are well resolved and modeled by the data. For each layer the blocks within the solid lines have resolution values of 0.7 or greater. Hence these regions of the model are well resolved by the data. The blocks within the dotted lines have standard errors of 0.6% or less. For the upper two layers of the model, 40 to 125 km and 125 to 210 km deep, the velocity contrasts modeled by the data vary from \(-2 \pm 0.6\%\) to \(+1 \pm 0.6\%\) in the well resolved region of the model. This two to three percent velocity contrast is significantly greater than the uncertainty. For the lower two layers, deeper than 210 km, the magnitude of the velocity contrast is significantly smaller. The lower layers have velocity contrasts on the order of 1%. Given the 0.6% uncertainty for these blocks, the inversion results are consistent with a smooth asthenosphere or limited velocity variations in the asthenosphere beneath the rift.

4.3. Lithosphere/Asthenosphere Interface: Transversely Isotropic Model

[26] In this section, we use a transversely isotropic asthenosphere with vertical axis of symmetry to interpret the travel time residuals. The use of an anisotropic model is mainly motivated by the two flank lows observed on the travel time residual curves (Figure 4) and by the fact that other studies such as deep seismic sounding and gravity, among others, do not support the existence of a high-velocity body in the areas beneath where the flank lows are observed (see section 5.1). Furthermore, there is growing evidence that both the lithosphere and asthenosphere are anisotropic with a fabric that is controlled by shear deformation associated with the movement of mantle flow [Kendall, 1994; Blackman et al., 1996; Montagner and Guillot, 2000]. The shear associated with mantle flow in a localized convection cell [e.g., Huismans et al., 2001, Plate 1] from the upward and downward flanking flows will give a complex distribution of anisotropy that is beyond the ability of teleseismic residuals to resolve. We thus consider a gross simplification of the true situation. Near vertically incident \( P \) waves will be most influenced by fabrics caused by vertical shears. We use a model consisting of a low-velocity asthenospheric upwarp and an underlying flat asthenospheric layer of laterally variable anisotropy. We assume that the lithosphere is isotropic for steeply incident teleseismic \( P \) waves, and both the upwarped and flat parts of the asthenosphere are transversely isotropic with a vertical axis of symmetry. As discussed below, the observed travel time residuals require an anisotropy of about 3% in the mantle to the depth of about 400 km. While most studies suggest that anisotropy is mostly limited in the top 250 km of the Earth, some other studies find radial anisotropy at about 400 km depth beneath large areas such as the western Pacific Ocean [Boschi and Dziewonski, 2000], and azimuthal anisotropy in the mantle transition zone [Trampert and van Heijst, 2002].

[27] A \( P \) wave with vertical incidence traveling through a transversely isotropic medium with vertical axis of symmetry has the highest velocity, and that with horizontal incidence has the lowest velocity. For nonvertical incidence, the velocity, \( V_p(\theta) \), can be calculated using the result of Backus [1965] and its modified forms. One of the most frequently used forms for transversely isotropic media is

\[ V_p^2(\theta) - V_0^2 = c_1 \cos(2\theta) + c_2 \cos(4\theta), \]

where \( V_p \) is the velocity of the \( P \) wave, \( V_0 \) is the velocity of the \( P \) wave at normal incidence, \( c_1 \) and \( c_2 \) are constants, and \( \theta \) is the angle of incidence.\(\square\)
where \( V_0 \) is the mean velocity, \( \theta \) is the angle between the symmetry axis and the ray direction, \( c_1 \) and \( c_2 \) are combinations of four elastic constants,

\[
c_1 = \frac{(c_{11} - c_{22})}{2c_0}; \quad c_2 = \frac{(c_{11} + c_{22})}{8} - \frac{c_{12}}{4} - \frac{c_{66}}{2} \quad \text{ [Bamford, 1977; Crampin and Bamford, 1977; Fuchs, 1984; Anderson, 1989].}
\]

[28] We assume that the magnitude of anisotropy is the highest beneath the rift axis and decreases exponentially outward; that is, \( c_1 \) and \( c_2 \) above are considered to be Gaussian functions of distance from the rift axis with the forms

\[
c_1(x) = c_{10} \exp(-0.5x^2/\sigma^2) \quad \text{(7)}
\]

and

\[
c_2(x) = c_{20} \exp(-0.5x^2/\sigma^2), \quad \text{(8)}
\]

where \( \sigma \) is the standard deviation of the Gaussian function. To include possible asymmetric decay of the magnitude of anisotropy, a different \( \sigma \) is used for \( x < 0 \) and \( x > 0 \).

[29] Figure 9 (top) shows travel time residuals in three scenarios for a ray with \( \theta = 10^\circ \). When the flat asthenospheric layer is isotropic, the residual curve consists of positive values with the peak near the center; when the upwarped asthenosphere is absent, the residuals are all negative with the minimum value near the center due to lateral variation of anisotropy in the underlying asthenosphere. When both parts of the asthenosphere are present and are anisotropic, the combined travel time residual curve has a valley on each side of the rift and a reduced peak near the center.

[30] Because of the largely 1-D configuration of our array (Figure 1), we assume that the structure is two-dimensional beneath the array. The surface of the upwarp is described using a parabola of the form

\[
z(x) = a_1 + \lambda x^2, \quad \text{(9)}
\]

where \( a_1 \) is the depth of the peak of the upwarp and \( \lambda \) is the coefficient of the parabola. To include a possible asymmetric shape of the upwarp, an independent \( \lambda \) is used for \( x < 0 \) and \( x > 0 \). The strike and location of the vertex line

---

**Figure 7.** Smoothed ACH block inversion of travel time residuals for the two upper mantle layers in the lithosphere. Open squares are the stations. The solid line indicates the contour of blocks with 0.7 resolution; blocks within the contour are resolved at 0.7 or greater. The dotted line indicates the contour of blocks with standard errors of 0.6%; blocks within the contour have errors of 0.6% or less. The region within the contours are well resolved and modeled by the data. HVR, high-velocity region.

**Figure 8.** Same as Figure 7 but for the two upper mantle layers below the lithosphere.
parameters from Bayesian inversion are (1) $a_1 = 45 \pm 15$ km; (2) $\lambda_1 = 0.00877 \pm 0.00065$ km$^{-1}$; (3) $\lambda_2 = 0.01363 \pm 0.00070$ km$^{-1}$; (4) $\gamma = 2.53 \pm 0.099$; (5) $c_1 = 1.573 \pm 0.053$ km$^2$/s$^2$; (6) $c_2 = -0.692 \pm 0.052$ km$^2$/s$^2$; (7) $d_1 = 350 \pm 125$ km; (8) $s_2 = 170 \pm 40$ km; (9) $\phi_0 = 51.8^\circ \pm 0.8^\circ$; (10) $b = 33 \pm 20$ km; (11) $h_1 = 170 \pm 20$ km; and (12) $h_0 = 230 \pm 30$ km.

[31] The resulting velocity model is shown in Figure 10, in which a localized velocity model [Egorkin et al., 1984] was used as the reference model. The results indicate that the low-velocity upwarp starts at a depth of 230 km and reaches 45 km depth. The velocity inside the upwarp is 2.5% lower than the outside velocity at the same depth. At its bottom the upwarp is about 260 km wide. The strike of the structure is $52^\circ$ measured anticlockwise from the east, and the vertex of the 2-D structure is 33 km south of station 13. The inversion indicates that the vertex of the 2-D structure is approximately parallel to the axis of the surficial manifestation of the rift.

5. Discussion

5.1. Applicability of the Velocity Models

[34] Three techniques were used to invert the same $P$ wave travel time residual data set for upper mantle velocities. Results from the two techniques under the assumption of isotropy (Figures 6–8) are consistent with each other. The major difference between the isotropy and anisotropy results is that in the isotropic models, the flank lows in the travel time curves (Figure 4) are caused by high-velocity bodies that are parallel to the rift axis, and in the anisotropic model, they are related to anisotropy with a vertical axis of symmetry.

[35] The final misfits are 0.046, 0.017, and 0.057 s$^2$ for the models presented in sections 4.1, 4.2, and 4.3, respectively. The small misfit (0.017) for the block inversion technique is probably the result of using a more realistic a 3-D model rather than a 2-D one. To test this hypothesis, we ran an ACH inversion using a 2-D model for the same data set, and found that the misfit is 0.037 s$^2$, which is comparable with those from the other two techniques. Thus it is impossible to distinguish between these models based on the goodness of fit of the $P$ wave travel time residuals alone. In addition, it is clear that a model with the best fit to data may not necessarily be the most reasonable one physically. Therefore other types of geophysical measurements that are independent from teleseismic $P$ wave travel time residuals are needed to determine the applicability of the models.

5.1.1. Results From Deep Seismic Sounding

[36] Deep seismic sounding experiments started in the former USSR in 1968. A total length of profiles of more than 4000 km covering an area of over 400,000 km$^2$ have been studied, and one profile extended across the southern Baikal area [Puzyrev et al., 1978]. The investigation discovered a low-velocity upper mantle layer in the area but did not find any indication of lithospheric downwarp. This technique uses near-horizontal refracted rays, which may be less sensitive to the vertical fabrics than teleseismic $P$ waves.

5.1.2. Bouguer Gravity Anomalies

[37] On the basis of Birch’s law, a 3% increase in seismic velocity at the base of the lithosphere leads to an increase of about 110 kg/m$^2$ in density. Given the geometry and depth of the two high-velocity downwarps (Figure 6), we calcu-
lated that a broad positive gravity anomaly with a maximum value of about 30 mGal would have been observed if the downwarps have a correlated density anomaly. As shown in Figure 11, such anomalies are not observed from existing gravity data in the study area [e.g., Zorin et al., 1989; Kaban et al., 1999]. In particular on the Siberian craton, the gravity is flat, whereas to the south large negative values are observed probably in part due to thickened crust in the Mongolian fold belt. Therefore the isotropic models are not supported by available gravity data.

5.1.3. Surface Topography
[38] To maintain isostatic balance, a broad high-density downwarp of the base of the lithosphere should be compensated by a broad depression of the surface of the Earth. Using the densities given in the IASP91 Earth model, we estimated that the magnitude of the depression would be as large as 1.4 km in order to reach compensation. Analysis of digital elevation data across the profile reveals that such depressions do not exist at the expected locations.

5.1.4. Crustal Thickness
[39] In the absence of a surface depression, a thickened crust in the area above the lithospheric downwarps could also compensate the excessive mass created by the lithospheric downwarps because the crust has a lower density than the lithosphere. We calculated that a thickening directly above the downwarps of about 10 km is needed, which is not observed by stacking P-to-S converted phases from the Moho [Zachary et al., 2000].

5.2. A Small-Scale Mantle Convection System Beneath the BRZ
[40] Independent geophysical measurements described in the above section suggest that an anisotropic model is a more plausible explanation for the travel times than the isotropic ones. Convective shear should cause lattice-preferred orientation (LPO) of crystallographic axes of anisotropic upper mantle minerals such as olivine, which is thought to comprise about 60% of the Earth’s uppermost

Figure 10. (bottom left) P wave velocity model derived from inversion of travel time residuals under the assumption of transverse isotropy. Both the flat and the upwarped parts of the asthenosphere are anisotropic, although the anisotropy in the latter cannot be visually observed. (top) P wave velocity at 300 km depth for rays with an angle of incidence of 28°. The velocity variations are the effect of anisotropy with a vertical axis of symmetry. (bottom right) Plot of isotropic V_p as a function of depth in the regions outside the upwarped area [Egorkin et al., 1984].

Figure 11. (bottom) Regional Bouguer gravity anomalies (averaged in 1° by 1° blocks) in the study area (data after Kaban et al. [1999]), and (top) gravity profile along the seismic line.
mantle. The a axis of olivine aligns in the flow direction under progressive simple shear [e.g., Hess, 1964; Karato, 1989; Babuska and Cara, 1991; Silver and Chan, 1991; Chastel et al., 1993; Silver, 1996]. For example, in a medium that is composed of pure olivine with perfectly aligned a axis and randomly aligned b and c axes in the plane perpendicular to a, seismic P waves traveling along the a axis have a velocity that is about 20% faster than P waves traveling orthogonal to it.

[41] We propose that vertical asthenospheric flow is responsible for generating the observed transverse isotropy beneath the BRZ and that the flow could be the ascending and descending branches of a small-scale convection system associated with the rifting [Huismans et al., 2001]. The presence of SKS splitting in the BRZ area [Gao et al., 1994a, 1997] implies anisotropy with horizontal symmetry axes are present. Because olivine is near-hexagonal with a axis being dominantly fast, we have used a transversely isotropic model and assumed that the P waves are sensitive to the vertical distribution of a axes, and the splitting to the horizontal distribution of a axes. In contrast to vertically traveling P waves, SKS splitting is more sensitive to the horizontal shears and therefore to the horizontal branches of the convection. Shear wave splitting measurements in the Baikal rift region [Gao et al., 1994a, 1997] reveal systematic spatial variations in the fast shear wave polarization directions. In the inner region of the BRZ, the fast directions are distributed in two orthogonal directions, NE and NW, approximately parallel and perpendicular to the NE strike of the rift. In the adjacent Siberian platform and northern Mongolian fold belt, only the rift-orthogonal fast direction is observed. The rift-parallel fast directions near the rift axes can be interpreted by oriented magmatic cracks in the mantle or small-scale mantle convection with rift-parallel flow [e.g., as described by Nicholas, 1993], and rift-orthogonal fast directions could be interpreted as the result of the horizontal component of a small-scale mantle convection system centered at the rift axis. Our results from the travel time residual inversion under the assumption of transverse isotropy (Figure 10) provide further supportive evidence for the existence of such a convection system.

5.3. Speculation on the Existence of Flank Lows in Other Rifts

[42] During the past 20 years several teleseismic experiments were conducted across the Rio Grande and East African rifts [Davis, 1991; Davis et al., 1993; Achauer et al., 1994, 1996; Ritsema et al., 1998]. While the area covered by late travel times is wider across both rifts than that across the BRZ (which implies that the asthenospheric upwarps are wider than that beneath BRZ) and the peak-to-peak anomalies are about twice as large, the rift-orthogonal dimensions of the seismic arrays deployed in those experiments were not as large as that of the BRZ. Therefore it is possible that similar flank lows, if they exist beneath those rifts, could be located outside the arrays. It is interesting to note that there is indeed an area with early arrivals of as large as 1 s located about 230 km east of the axis of the Rio Grande rift, and 270 km west of the eastern end of the E-W seismic profile [Davis, 1991, Figure 3]. This flank low is observed on events from the SE direction, but is not seen on events from the NW direction. This could suggest that the flank low shifts southeastward for events from the NW and consequently develops beyond the limits of the array. It seems that teleseismic profile of about 1500 km long is needed across both the Rio Grande and East African rifts in order to to study the existence and characterization of possible flank lows associated with those rifts.

6. Conclusions

[43] Nonlinear inversion of travel times and other geophysical measurements suggest that the travel time residuals observed along a 1280 km profile across the Baikal rift zone are the combined results of an upwarped asthenosphere and a vertical mantle flow centered at the rift axis. An isotropic inversion gives rise to asthenospheric upwarp beneath the rift and lithospheric downwarps on either side. However the downwarps are not seen in the gravity, or, if isostatically compensated, in the topography or crustal thickness. Further they have not been recorded in previous deep seismic sounding experiments carried out in the area. One way to have large velocity anomalies without gravity effects is the presence of anisotropy. Previous SKS splitting had shown laterally variable splitting in the Baikal rift zone and was interpreted as caused by small scale convection. The P wave residuals and gravity can be explained if the vertical shears associates with mantle upwelling and downwelling cause a axes of olivine to be oriented vertically, and consequently cause a localized high-velocity anomaly beneath the rift zone. This superimposed on the low-velocity anomaly of asthenospheric upwarp gives rise to the characteristic central high and flanking lows in the travel time patterns. The results lend weight to models of active rifting induced by instability of the mantle lithosphere that causes small-scale convection to develop, in which lithospheric extension is much greater than crustal extension.

[44] This study demonstrates the role that anisotropy can play in the inversion of seismic travel time residuals, and the significance of shear wave splitting and travel time modeling in detecting mantle convection systems. It also suggests the need for future seismic experiments along arrays that are significantly longer than most of the previous arrays across major continental rifts.

References


---

P. M. Davis and P. D. Slack, Department of Earth and Space Sciences, University of California, Los Angeles, CA 90095, USA. (pdavis@ess.ucla.edu; pslack@linex.com)

S. S. Gao and K. H. Liu, Department of Geology, Kansas State University, Manhattan, KS 66506, USA. (gao@ksu.edu; liu@ksu.edu)

V. M. Kozevnikov, V. V. Mordvinova, and Y. A. Zorin, Institute of the Earth’s Crust, Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia.