Layered mantle convection: A model for geoid and topography

Lianxing Wen *, Don L. Anderson
Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

Received 28 May 1996; accepted 4 October 1996

Abstract

The long-wavelength geoid and topography are dynamic effects of a convecting mantle. The long-wavelength geoid of the Earth is controlled by density variations in the mantle and has been explained by circulation models involving whole mantle flow. However, the relationship of long-wavelength topography to mantle circulation has been a puzzling problem in geodynamics. We show that the dynamic topography is mainly due to density variations in the upper mantle, even after the effects of lithospheric cooling and crustal thickness variation are taken into account. Layered mantle convection, with a shallow origin for surface dynamic topography, is consistent with the spectrum, small amplitude and pattern of the topography. Layered mantle convection, with a barrier about 250 km deeper than the 670 km phase boundary, provides a self-consistent geodynamic model for the amplitude and pattern of both the long-wavelength geoid and surface topography.

Keywords: geoid; topography; mantle; convection

1. Introduction

Geoid and topography are connected dynamic effects of a convecting mantle. The dynamic topography, caused by deep mass anomalies, is quite different from the actual observed topography, which is dominated by variations in crustal thickness and thermal subsidence of oceanic plates. It is difficult to correct for these variations so it is uncertain exactly how much of the Earth’s long-wavelength topography is actually due to the mass heterogeneities in the mantle. Several studies indicate that the smoothed topography in oceans deviates only slightly from thermal conduction models (about ~ 500 m at spherical harmonic degree l = 2) [1–4]. The small gravity signal related to cratonic regions indicates that the topographic signal related to cratonic “roots” is also weak. The small amplitude of dynamic topography is consistent with the rise and fall of continents inferred from flooding records [5]. The residual topography, which is the residual after removal of the topographic components resulting from near-surface density contrasts and seafloor subsidence, can be expected to represent the topographic response of Earth’s surface to internal loads of the mantle. Fig. 1a,c,e shows degree l = 2–3 components of residual topography and nonhydrostatic geoid [6]. The residual topography was provided by A. Cazenave. We use the residual topography corrected by the subsidence laws in oceanic regions: (1) by Stein and Stein [7] (plate model), shown in Fig. 1a; (2) by Marty and Cazenave [8] (half-space cooling model), shown in Fig. 1c. Sediment loading is corrected as explained in Cazenave et al. [9]. Different subsidence laws give basically the same pattern and range up to 30%
Fig. 1. The 1–2–3 components of residual topography corrected for crustal thickness variation by assuming Airy compensation in continental regions and sediment loads and thermal subsidence based on: (a) plate model [7]; and (c) half-space cooling model [8] in oceans. Dynamic topography predicted by: (b) model WA1; and (d) model WA2. (e) Nonhydrostatic geoid and (f) geoid predicted by model WA1. Topography and geoid are predicted by assuming layered mantle flow stratified at 920 km. Topography and geoid lows are shaded. Contours: (a)–(d) 100 m; (e)–(f) 20 m.
higher for the half-space cooling model. Over continental regions, topography is corrected for crustal thickness variation [10] assuming local Airy compensation. The largest source of uncertainties for the continental corrections is due to assumed crustal density. A constant density of 2800 kg/m$^3$ and a reference crustal thickness of 35 km are assumed. The mean residual elevation over continents is subtracted to avoid any baseline difference. Cazenave et al. [1] have performed several other treatments for the continental correction: (1) continental elevations were set to zero; (2) plate boundary regions and ice sheets were excluded. They obtained very stable patterns at long wavelengths ($\geq 5000$ km) and concluded that continental areas contribute negligibly to the very long wavelength residual topography.

The long-wavelength geoid can be explained by whole mantle flow models [11–14], although there are problems with the long-wavelength dynamic topography [15–19]. It is worth noting that, in the traditional geodynamic modeling of the geoid, the predicted geoid anomalies are actually the summation of the contribution of mass heterogeneities in the mantle and mass anomalies due to the dynamic topography caused by those mass heterogeneities in the mantle [20]. Both the geoid and topography must be explained by a mantle flow model before we can claim that we have a self-consistent model. Dynamic topography is difficult to model because the relations between seismic velocity variations and density variations are non-unique, particularly in the upper mantle [21], and are not entirely thermal in nature [22,23]. Cratonic roots have high seismic velocity, but do not necessarily have high density because they are chemically distinct from the surrounding mantle [22,23]. Most previous modelings of topography either exclude the shallow structure of the mantle [17,18], which is an important contributor to the dynamic surface topography, or put a theoretical slab model into the upper mantle and ignore other density anomalies [15,20,24]. Here, we infer mantle density from seismic tomography [25] in the lower mantle and residual tomography [26] in the upper mantle. The residual tomography, the residual after removal of cratonic roots and the effects of conductive cooling of oceanic plates, is used since we want to isolate the dynamic response of the Earth’s surface to internal loads. Detailed procedures are presented elsewhere [26]. With this approach, realistic subduction effects can also be included. We assume incompressible, self-gravitating Newtonian mantle flow [27]. The large-scale structure ($l = 2–3$) will be studied for two reasons: (1) most of the power of the geoid and topography is concentrated at $l = 2–3$; and (2) mode coupling becomes important at shorter wavelengths, due to lateral variation in viscosity [28,29]. We first re-examine whole mantle flow models and then propose a layered mantle flow model.

2. Whole mantle flow models

Three radial mantle viscosity structures and related velocity–density scalings, inferred from whole mantle flow models, by Hager and Richards (HR) [30], King and Masters (KM) [13] and Forte et al. (FPDW) [16] are shown in Fig. 2a,b. Those models are also used in their analysis of heat flow by Phipps Morgan and Shearer [17]. These models have an increase in viscosity of about 10–30 times between the upper and lower mantle. The geoid kernels from these models are similar (Fig. 2c). These show the effect on the geoid for a mass anomaly at a given depth. Geoid kernels of models KM and FPDW peak in the transition region, whereas model HR is more sensitive to the upper mantle. All geoid kernels are negative in the deep lower mantle. The geoid correlates with the seismic structure positively in the transition region and negatively in the deep lower mantle. The viscous flow model cannot be uniquely determined by geoid modeling. However, the characteristics of these geoid responses may be intrinsic. A successful model, whether it assumes whole mantle flow or layered mantle flow, should have a geoid response similar to those shown, in order to fit the geoid. A layered mantle flow model, stratified near the 670 km discontinuity, would have geoid kernels insensitive to mass anomalies in the transition zone region. These three models predict the geoid well from mass anomalies derived from seismic tomography. However, the magnitudes of predicted topography are much larger than observed (about 2.5–3.5 km in peak–peak amplitude). There is some correlation at $l = 2$ and no correlation at $l = 3$ between observed and predicted topography for these whole mantle flow models. This is consistent with the
results of other recent studies [17–19]. A large-amplitude dynamic topography is also predicted for the density model inferred from past subduction [24].

It is obvious from effective topography kernels (Fig. 2d) that current whole mantle convection models cannot predict both the geoid and residual topography simultaneously. The contribution from lower mantle heterogeneity, by itself, already exceeds the observed residual topography for these whole mantle convection models. This is evident also in previous work. We have attempted to find a whole mantle flow model that satisfies both the geoid and the dynamic topography. We use the above models for viscosity and assume that the velocity–density scalings are constants in the depth intervals 0–400 km, 400–670 km and 670 km–CMB. We attempted a least-squares fit to the geoid by searching over a range of velocity–density scaling constants. We were unable to reduce the amplitude of the predicted topography. We were unable to obtain a satisfactory fit to both the geoid and the topography at l = 2–3 with whole mantle convection by performing a least-squares fit to both the geoid and topography. There is no correlation between residual topography and predicted dynamic topography at l = 3 in any case.

3. The origin of dynamic topography

The spectra of residual topography and geoid are shown in Fig. 3a. The amplitude of the geoid de-
Fig. 4. The comparison of $l = 3$ components among the residual topography models by plate model (a1) and half-space cooling model (a2), averaged residual tomography [26] in 0–400 km (b), 400–670 km regions (c) and averaged seismic tomography [25] (d) in the lower mantle. Note that, unlike $l = 2$, strong correlations are only found between residual topography and shallow structure. This also confirms the shallow origin of residual topography. Contours: (a) 100 m; (b)–(d) 0.15\%.
increases much faster toward short wavelengths than that of residual topography. The long-wavelength signal senses deeper than short wavelengths. For potential fields, the faster the spectrum decreases with inverse wavelength, the deeper the origin of anomalies. For example, the magnetic field comes from the core and the spectrum decreases much faster than the geoid and topography spectra. The different behaviors of the spectra of geoid and residual topography imply that dynamic topography is controlled by density variations in the shallow mantle. Thoroval et al. [18] recently proposed that the undulation of the 670 km discontinuity is responsible for the excessive topography at the surface produced by whole mantle flow models. Spectral analysis is useful in discussing this possibility since it depends only on the viscosity model. Let us assume that whole mantle flow models predict twice as much dynamic topography as observed. In this case, half confirms the observation and we will attribute the other half to the topographic effect of the 670 km discontinuity. In order to produce the excess topography that we have assumed, we can quantitatively calculate the spectral behavior of the undulation of the 670 km discontinuity for the various viscosity structure models. Fig. 3b shows examples of the spectral behavior of the 670 km discontinuity for viscosity models of Thoroval et al. [18] and Hager and Richards [30]. This indicates that, if the undulation of the 670 km discontinuity is responsible for the “excess topography” predicted by whole mantle flow models, then the signal increases at short wavelengths for the topography at 670 km. Geoid, topography, subduction history and seismic tomography show that large-scale features dominate mantle convection. It is hard to believe that small scale dominates for topography at the 670 km discontinuity. The spectrum of the seismologically inferred topography of the 670 km discontinuity [31] is also shown for comparison. This quantitative analysis indicates that, if a certain boundary is responsible for the “excess topography” at the surface produced by whole mantle flow models, this boundary cannot be as deep as 670 km. The incorporation of the topography of the 410 km and 670 km discontinuities into the geodynamical modeling also produces large amplitude surface dynamic topography [17]. It is worth noting that heterogeneities in the upper 300 km were not included in those models [17,18]. These heterogeneities, however, have a significant influence on dynamic topography.

It is also worth noting that explaining the pattern of the dynamic topography is as important as matching the amplitude of dynamic topography for a convection model. At \( l = 2 \), the residual topography correlates with lower mantle seismic tomography [1]. However, the correlations are even better at shallow depths. No correlations are found in the transition zone region. The connection between deep mantle structure and shallow mantle structure at this degree \((l = 2)\) is unclear [26] and this confuses the situation. However, the pattern at \( l = 2,3 \) allows us to discriminate between the various proposals for the origin of dynamic topography. Fig. 4 shows the spherical harmonic degree \( l = 3 \) component of the residual topography corrected by different subsidence laws in the oceans, averaged residual seismic tomography in the top 400 km, transition zone region, and seismic tomography in the lower mantle. There are very good correlations between residual topography and shallow seismic residual tomography, with topographic lows corresponding with high velocities and topographic highs corresponding with low velocities. The shallow origin of residual topography can be seen clearly from the direct comparison of patterns of residual topography and seismic tomography. If one excludes the structure in the upper 300 km, it is hard to imagine that one can find a viscous flow model that predicts the correct pattern of dynamic topography from a density model inferred from seismic tomography for the rest of the mantle.

4. Layered mantle flow model

Shallow origin and small amplitude of dynamic topography have two possible explanations: (1) Whole mantle flow with a very high viscosity lower mantle. Such a model would be rejected by the geoid. (2) Layered mantle flow with the upper and lower part of the mantle mechanically decoupled. With a large viscosity jump from the upper to lower mantle, the Earth’s surface does not respond to the mass heterogeneities in the lower mantle for layered mantle flow model. Fleitout [32] pointed out that the magnitudes of observed surface topography and in-
tra-plate stress are more compatible with a two-layer convective mantle, with the lower mantle mechanically decoupled from the lithosphere and unable to induce tectonic stress.

Previous discussions about possible geodynamic barriers have focused on the 670 km discontinuity. Layered mantle convection models, stratified at 670 km, produce excessive topography at this boundary [30] or poor fits to the geoid [12]. The incorporation of the seismologically inferred topography on the 670 km discontinuity [33] seems to rule out this boundary as the dividing line between shallow and deep mantle convection. Many regional high resolution seismic tomography studies also reveal high velocity anomalies below this discontinuity beneath several subduction zones [34,35]. The 670 km discontinuity is primarily due to an endothermic phase change between the spinel and post-spinel forms of olivine [36]. There is no requirement that compositional difference must set in at this depth. For example, a chemical barrier may exist deeper and be unrelated to the present position of the spinel–post-spinel phase boundary.

There are now several independent lines of evidence for an important geodynamic boundary in the mid-mantle: seismic evidence for the presence and high relief of a 920 km discontinuity [37,38]; correlations between subduction history in the past 130 Ma and seismic tomography at about 800–1100 km depths in the mantle [39]; decorrelation of seismic tomographic models at 900–1000 km depth [40] and reversal of thermal fluctuations at a depth of about 850 km [41]. It is of interest to see if a barrier near these depths can satisfy the geoid and topographic data.

We applied rheological models, similar to that of Hager and Richards [30] (WA1 and WA2, Fig. 2a). The velocity–density scaling factors (\(\frac{\partial \ln \rho}{\partial \ln V_s}\)) are assumed to be constants for the depth intervals, 0–400 km, 400–670 km and 670 km–CMB. These three constants were obtained by a least-squares fit to both the geoid and residual topography, assuming a layered mantle convection with a boundary at the 920 seismic discontinuity. WA1 and WA2 were applied for residual topography models corrected by the plate model and half-space model separately.

Our best fit velocity–density scaling factors are similar to these for the whole mantle flow model of Forte et al. [12] (Fig. 2b). The effective geoid kernel is very similar to those of whole mantle convection models [12,13,30] (Fig. 2c). This indicates the lack of uniqueness of geoid modeling in discriminating between whole mantle and layered mantle flow. Remarkable differences between layered mantle and whole mantle flow models are evident in the topography kernels (Fig. 2d). For the whole mantle convection models, the topography kernels are sensitive to the whole mantle. For layered mantle convection models, with a viscosity jump of about a factor of 10–30 in the lower mantle (WA1 and WA2), and chemical stratification at 920 km, the topography kernel is only sensitive to the upper mantle; that is, only density anomalies in the upper mantle contribute to most of the dynamic topography. The lower mantle contributes little to the surface dynamic topography for layered mantle flow model WA2.

The predicted dynamic topography for layered mantle flow models, at \(l = 2–3\), are shown in Fig. 1b,d. The predicted and observed dynamic topography have highs in the central Pacific and in southeast Africa and large depressions in Eurasia, south Australia and South America. At \(l = 2,3\), the correlation coefficients between predicted and observed exceed the 95% confidence level for the topography corrected by the plate model, and exceed the 90% confidence level for topography corrected by the half-space cooling model. The predicted geoid is very similar for both models (WA1 and WA2). Therefore, only the prediction from model WA1 is presented in Fig. 1f. The correlation coefficients between observed and predicted geoid exceed the 95% confidence level at \(l = 3\) and the 93% confidence level at \(l = 2\). The predictions of geoid and topography reproduce not only the patterns, but also the amplitudes as well. The spectrum of predicted dynamic topography for model WA2 is shown in Fig. 3a. An example for the spectral behavior of dynamic topography predicted by the whole mantle flow model (HR) is also shown for comparison. The layered mantle convection flow model is more consistent with the spectral behavior of the observed topography.

The topographic relief at the 670 km endothermic phase change is a second-order effect in geoid and topography modeling. But it is important in predicting the dynamic topography at a deeper chemical discontinuity. We incorporate the topography at 670
Predicted CMB Topography (l=1-12)

Fig. 5. Dynamic topography at the CMB predicted by layered mantle flow model WA2. A density contrast of 4.5 g/cm³ across the CMB is assumed. Topography lows are shaded. Contours: 200 m.

km derived by Shearer and Masters [31] into the calculation and assume a density jump of 9% [42]. The predicted geoid changes by less than 10% in amplitude, and the correlations persist. There is almost no change in the computed topography. The predicted dynamic topography at 920 km has about 110 km of peak–peak variation of depth, assuming a density change of 0.2 g/cm³.

Similar to the prediction of surface topography, layered mantle flow models also predict much less amplitude of dynamic topography at the CMB than whole mantle flow models do. Fig. 5 shows predicted CMB dynamic topography for layered mantle flow model WA2. Model WA1 predicts a very similar pattern of the CMB topography with slightly different amplitude. Little topography is related to circum-Pacific regions. Most topography is related to upwelling regions. The peak–peak variation in predicted CMB topography is about 2 km. Whole mantle flow models predict at least twice as much [12,14,15]. We obtain a second zonal harmonic deviation from the hydrostatic equilibrium figure, with the peak-to-valley deviation being about 550 m, in excellent agreement with the analysis of nutation data [43]. Hager and Clayton [15] predicted 2 km of excess ellipticity of the CMB. However, our results are inconsistent with some seismological studies [44,45] but different studies give conflicting results [46]. It is also unclear how to relate dynamic topography to the actual topography at the CMB.

5. Discussion

The seismic discontinuity near 920 km has about one-half the velocity jump of the 670 km discontinuity [37] and it is less evident in seismic stacks [47], suggesting that it may exhibit greater topography. It should also be noted that the region of the mantle between 670 and about 900 km is radially inhomogeneous and has therefore not been included in equations of state fits to the lower mantle. Convection deforms chemical boundaries and the high expected relief (as well as its small velocity contrast) may explain why this discontinuity has a more checkered history of being found than the 400 and 670 km discontinuities. The extensive literature on a discontinuity near 900 km is reviewed by Anderson [48], Revenaugh and Jordan [47] and Kawakatsu and Niu [37]. Weak reflections are sometimes reported between 1000 and 1300 km depth. It is not yet clear if these are actually from a single, variable depth, discontinuity.

Although the 920 km discontinuity may represent a phase boundary, the garnet to perovskite transition being the most obvious candidate [49], it may also represent a chemical boundary. Chemical variations between the mesosphere and the lower mantle below about 920 km most plausibly involve variations in MgO, FeO, SiO₂, Al₂O₃ and CaO contents. These are the major controls on density and seismic velocity. Assuming that these oxides are distributed between perovskites and magnesiowustite, a reasonable intrinsic density difference may be of the order of 0.1 g/cm³. For example, a chondritic mantle, depleted of Al₂O₃, MgO and CaO to form the crust and a fertile upper mantle will have Mg/Si about 1 (perovskite) and low Ca and Al in the residue. The radioactive elements will also be depleted in the residual lower mantle. Long wavelength lateral temperature contrasts of about 200 K would yield 0.2% density variations, if the thermal expansivity at lower mantle conditions is 10⁻⁵ K⁻¹. This thermal density contrast is less than the intrinsic chemical contrast, at lower mantle pressures, inferred for upper mantle compared with plausible deep mantle compositions.
The density difference at lower mantle pressures, between various plausible chemical models of the upper and lower mantle, ranges from 2.6% to 5% [50]. This is enough to stratify mantle flow.

An increase in SiO$_2$ or a decrease in MgO content will tend to raise seismic velocities. The effect of FeO depends on its spin state and metallic nature. Assuming that Al$_2$O$_3$ stabilizes the garnet structure to about 1000 km depth [49,50], and assuming further that magma extraction has depleted the lower mantle (presumably during accretional melting) in Al$_2$O$_3$ and CaO, and that the original mantle was more chondritic than the current upper mantle, then a SiO$_2$-rich and CaO, Al$_2$O$_3$-poor, and U, Th, K-poor, lower mantle is expected. The FeO budget of the lower mantle depends not only on its melt extraction history but also its subsequent interaction with the core and core-forming material in its early history. It is, in fact, difficult to imagine how a homogeneous mantle may have formed, particularly considering the efficient extraction of the crust-forming elements upwards and the core-forming elements downwards. Thermal expansion is high at upper mantle pressures; this, plus the effects of partial melting, means that compositional layers can be breached and potentially mixed with surrounding mantle. Pressure suppresses thermal expansivity and chemical discontinuities can be expected to be more permanent in the deep mantle.

There are various arguments for and against stratified mantle convection. Small intrinsic density contrasts between layers can easily be overcome at lower pressures. The effect of pressure on thermal expansivity makes this more difficult at high pressures. Many of the arguments against stratified convection are actually arguments against the 670 km level being the boundary. These arguments do not apply if the convection interface occurs nearer to 920 km. The possibility of an important geodynamic boundary several hundred kilometers deeper than the major seismic discontinuity at 670 km is consistent with electrical conductivity and viscosity data, as well as with 1D and 3D seismic modeling. For example, the viscosity may rise rapidly below 800–900 km [51]. Layered mantle convection serves to insulate the lower mantle and is less efficient at heat removal than whole mantle convection [52]. Consequently, two-layer convection implies a lower mantle depleted of radioactive heat sources [52], consistent with the differentiation during accretion model [53].

6. Conclusion

The amplitude of dynamic topography predicted by previous viscosity models involving whole mantle flow is much larger than observed. We are unable to find reasonable velocity–density scalings to fit the geoid and residual topography for whole mantle flow models. This is consistent with recent studies on topography [17–19]. We show that the seismic tomography at shallow depths, once corrected for the chemical effects of cratonic roots and thermal cooling of oceanic lithosphere, exhibit slow velocities that are well correlated with uplifted regions at long wavelengths (l = 2,3). The spectrum of residual topography also reveals the shallow origin of dynamic topography. The long-wavelength dynamic topography is controlled by density variations in the upper mantle, whereas the long-wavelength geoid is controlled by density throughout the mantle. Layered mantle convection stratified at about 920 km provides a self-consistent geodynamic model in explaining the long-wavelength geoid and topography.

Acknowledgements

We thank Anny Cazenave for the residual topography data, Scott King for the viscosity structure and Michael Gurnis, David Stevenson and Craig Scrivner for reviews. This work was funded by NSF grant EAR 92-18390. Contribution No. 5598, Division of Geological and Planetary Sciences, California Institute of Technology. [RV]

References


[38] H. Kawakatsu and F. Niu, Depth variation of “the 920 km discontinuity” in the mid-mantle, EOS Trans. AGU 76, F382, 1995.


