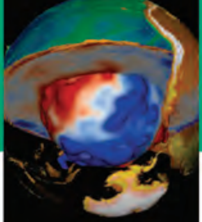
Contents



1 An Exploration of Basic Solid Earth Structure and Dynamics

The Earth sciences have always been interdisciplinary and are founded on interpreting the rich geological record of our planet. Much of the dramatic progress over the last century has been driven by discoveries based on observations in solid Earth geophysics, from gravity, to magnetic field measurements, to global seismology, and more recently to space geodetic techniques. However, the biggest advances usually arise when diverse observations from different disciplines are combined in mechanically consistent ways to establish a comprehensive theory for why the Earth works this way.

The establishment of the integrative model of plate tectonics is such an occasion and *the* big story in the solid Earth geosciences. As conceived originally, plate tectonics is a kinematic description for how the surface of the Earth deforms and moves horizontally, without specifying anything about the driving forces (*McKenzie & Parker*, 1967; *Morgan*, 1968; *Le Pichon*, 1968). However, those forces were quickly reconnected to the fact that plate tectonics is the surface expression of mantle convection, as had been suggested roughly 40 years earlier (*Holmes*, 1931).

Because of sustained radiogenic heat production within the mantle and the sluggish nature of mantle convection, our planet is still cooling and its surface continues to be shaped and reworked by its internal dynamics. Mantle convection controls the evolution of the solid systems of our planet, encompassing nearly all tectonic and magmatic activity, including the earthquakes and volcanic eruptions that affect society in sometimes catastrophic ways. Moreover, mantle convection and plate tectonics also interact with the exosphere, comprising the ocean and atmospheric layers, and with that the climate and the biosphere we are part of. However, there are still many open questions as to exactly how those interactions occur, and there are a number of important observations, such as those relating to tectonic transients, continental dynamics, and plate boundary evolution, that have so far resisted integration into a proper, comprehensive dynamic theory.

We proceed to review some of the fundamental geophysical and geological observations for Earth, with a focus on global scales, as if we were exploring an alien planet remotely. We then discuss the Earth's thermochemical structure, often explored jointly with seismological constraints, and then turn to an overview of a dynamic description of tectonic activities, which we follow up on throughout the remainder of the book.

1.1 Topography

When viewed from space, the outer terrestrial planets are strikingly different. The famous, Earth-as-a-blue-marble photo taken during the Apollo 17 mission emphasized our special, perhaps precarious, state with water oceans covering much of the planet, and a temperate atmosphere with clouds and prevailing wind currents generated by the interaction of solar heat influx, rotation, and the hydrological cycle, along with the atmospheric signatures of life (*Sagan et al.*, 1993). There is evidence for water having been available in fluid form on Earth's surface over billions of years (*Mojzsis et al.*, 2001), yet why and how this stable condition was maintained is not entirely clear.

The early Sun was much fainter than at present day, such that our oceans would have likely been permanently frozen in the Archean (*Feulner*, 2012). While there were periods of near-global glaciation in Earth's history (*Hoffman et al.*, 2017), those episodes were temporarily restricted, and for the most part, a self-regulatory climate feedback mechanism appears to have been at work, thanks in part to plate tectonics (*Walker et al.*, 1981): If the atmosphere gets too hot, rock is more readily eroded by surface processes. Weathering transports carbon into the oceans, where it forms sediments that are subducted back into the mantle. This pathway draws down atmospheric CO_2, and so reduces the atmospheric greenhouse effect. This reduces temperatures, and with it the strength of weathering, and as a consequence, less carbon gets transported into the mantle. Due to the sustained convective and tectonic outflux of CO_2 from the interior of the Earth, e.g., from volcanism, carbon dioxide concentrations increase in the atmosphere, leading to warming.

Such long-term climate regulation, and the existence of oxygen or any other atmospheres, are not guaranteed for terrestrial planets, as our sister planet Venus shows. The atmosphere of Venus is $\sim$340 K hotter than Earth's, thanks to runaway greenhouse effects due a high CO_2 and sulfur-rich atmosphere that also leads to surface pressures $\approx$ 92 times those of Earth. There are no oceans or other surface water on Venus, and plate tectonics is absent, perhaps because of rheological effects due to the absence of water, the high surface temperatures, or both. Heat from Venus's mantle is transported mainly through a more or less stagnant surface that exhibits volcanism and appears to show episodic reworking. Mars has a more temperate atmosphere but is likewise frozen tectonically, indicating that Earth's style of tectonics is unique in our solar system (§14.4). Surface conditions in terms of temperature, availability of water, and the carbon cycle are thus linked with the internal evolution of a planet, and somehow interact to lead to habitable conditions, at least for Earth.

If a spacecraft remotely exploring planets in our solar system were to use a radar instrument, an explorer could then peek through the atmosphere to reveal the planets' surface elevation with respect to some reference. Figure 1.1 shows topography for the three Earthlike planets using the same type of colorscale, centered on the reference (sea level for Earth). Thanks to space exploration, we have learned much about the surface geology of Mars, which shows the remnants of fluid water and wind-driven transport processes, but is at present characterized by a hemispheric dichotomy with southern highlands and northern lowlands. This feature is related to the internal dynamics of Mars's mantle, and there are remnants of extensive volcanism, including Olympus Mons, the highest topographic feature of all terrestrial planets. We know that Mars's surface is relatively static because many large-impact craters are preserved, and the relative decrease in typical impact size over solar system evolution can be used for dating.

Venus's surface is comparatively younger and shows features of large-scale reorganizations, in particular near-circular features, *coronas*, as well as indications of active volcanism. Earth shares the general planetary age of $\sim$4.6 Gyr with the terrestrial planets that formed from silicic- and iron-type materials in similar fashion during the formative eon of the solar system. The last planet-shattering event for Earth was the impact of a Mars-sized object that led to the formation of Earth's moon (§8.1.2.1). However, compared to Mars, Earth's surface is much younger. While there are rocks of $\sim$3.8 Gyr age preserved in Earth's continents, there are few signs of impact craters, indicating tectonic reworking and active surface processes. In particular, Earth's oceans are underlain by seafloor which, for the most part, is at present nowhere older than $\sim$200 Ma.

One way to analyze topography is by computation of fractional or cumulative topography histograms, *hypsometric curves* (Fig. 1.2). Such analysis brings out the aforementioned global features; for example, Venus has a slightly skewed, but overall nearly normally distributed topography. Earth shows clear peaks at the average continental and oceanic elevations of $\sim$100 m and ~ -4500 m,

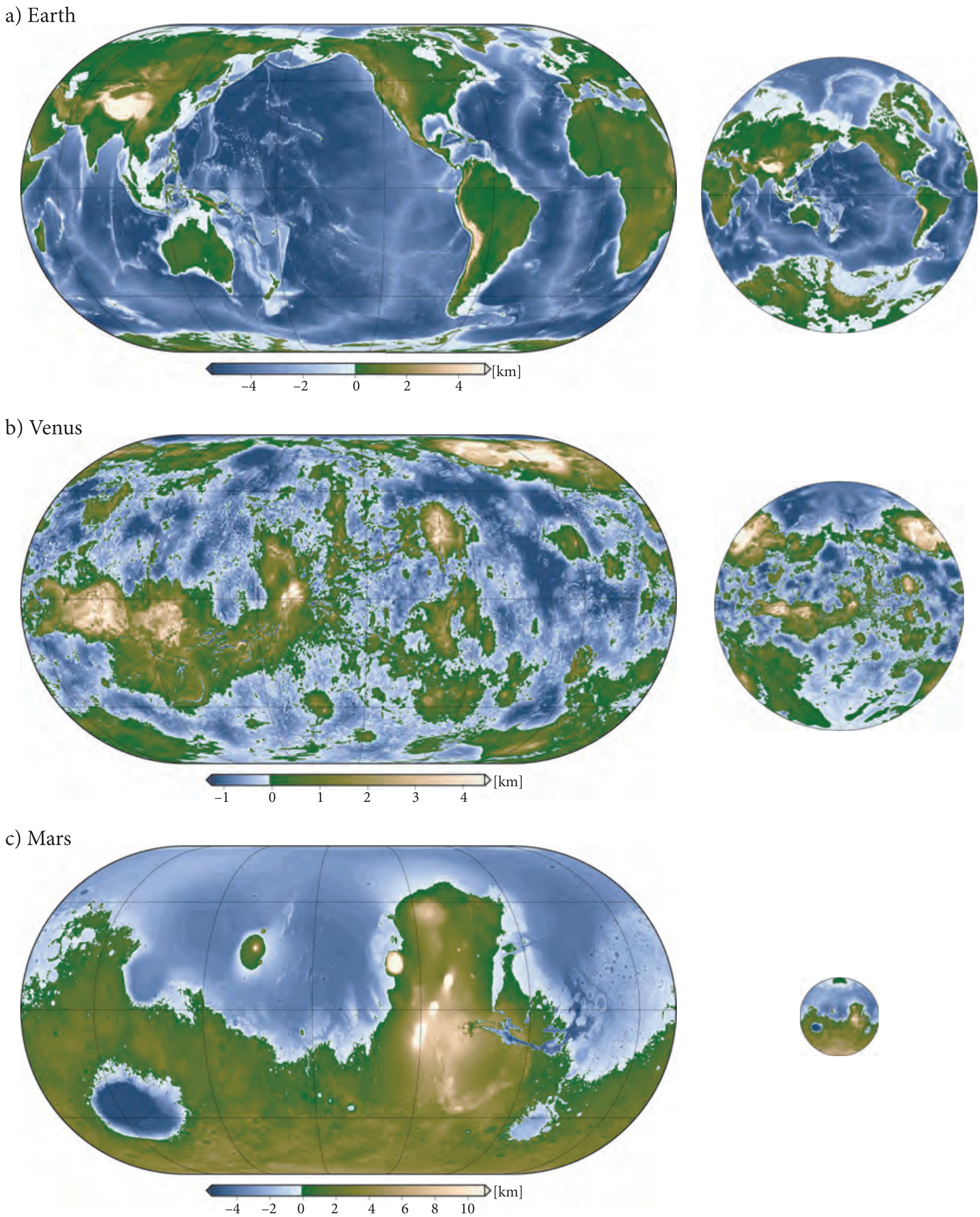


Figure 1.1: Topography of Earth (**a**; bedrock elevation, removing ice for Greenland and Antarctica; *Amante & Eakins*, 2009), Venus (**b**; Magellan, Pioneer Venus, and Venera missions), Mars (**c**; Mars Global Surveyor), where datasets for b) and c) were expanded from spherical harmonics (*Wieczorek*, 2015) (up to degree 720, when available). On left, Eckert-IV projection; on right, van der Grinten projection, scaled with the square of the object radius. See Fig. A.5 for power spectra and geoid anomalies.

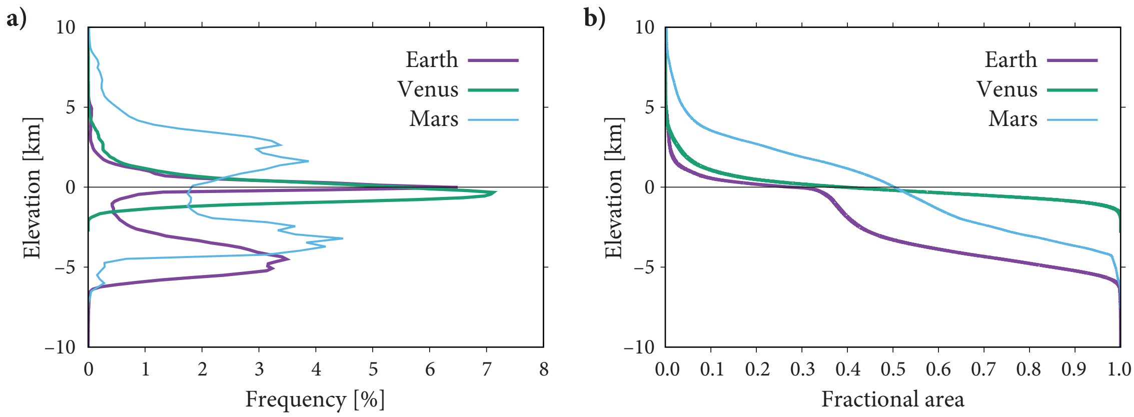


Figure 1.2: **a)** Fractional histogram of elevation for planetary topography from Fig. 1.1. **b)** Cumulative distribution function of elevation in area (i.e., hypsometric curve) (y axes clipped) for objects as in a).

respectively. This dichotomy is profound, and is the result of thermochemical mantle convection. Oceanic plates are continuously generated and recycled with a lifetime of ∼100 Ma. Continents are created mainly through fractionation at subduction zone arc volcanism, and subduction-driven accumulation of protocontinental fragments, e.g., created by plumes.

1.1.1 Isostasy

The bulk difference in oceanic and continental elevation of Fig. 1.2 can be explained by the floating equilibrium of different crustal columns. *Airy* (1855) first proposed that

> *Earth's crust lying upon the lava can be compared with perfect correctness to the state of a raft timber floating upon water.*

This idea was based on 1730s survey results by Pierre Bouguer from the Andes, and in the Nineteenth century in India, including by George Everest close to the south Himalayas. Both studies showed that the plumb line deflection at the base of the mountain was less than expected from the extra mass above sea level.

One way to explain this is there was a negative mass anomaly at the base of the mountain due to the "substitution of a certain volume of light crust for heavy lava" (*Airy*, 1855), as for a floating iceberg. This is one way to state the principle of *isostasy*, which posits that a crustal density column should be in an (Archimedes) floating equilibrium over long time and length scales (Fig. 1.3). If this isostatic equilibrium holds, the *pressure* due to the overburden of a crustal column, p, has to add up to the same everywhere at some *compensation depth*, Z, i.e., $p(Z) = const.$ everywhere. Pressure arises from the forces pushing inward on an object from all directions, normalized by the surface area of the object.

How can we link this assumption about force balance to topographic or structural observables? When considering forces, it is often useful to start with Newton. His second law says that force, expressed as a *vector* (§C.3) pointing in some direction, $\boldsymbol{F}$ (units N, Newton), is mass m (units kg) times *acceleration*, $\boldsymbol{g}$,

$$\boldsymbol{F} = m\boldsymbol{g}, \tag{1.1}$$

where acceleration, $\boldsymbol{g}$, is a temporal change in velocity (units of m/s^2). Eq. (1.1) means that it takes a force to change the rate of motion of an object due to inertia.

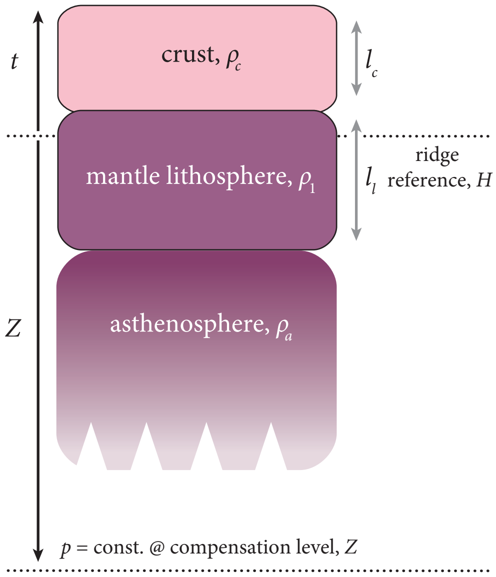


Figure 1.3: Isostatic balance column for load components from a crustal layer with thickness l_c and density ρ_c, and a mantle lithosphere with thickness l_l and density ρ_l (total lithospheric thickness $L = l_c + l_l$), floating over an asthenosphere of density ρ_a. The pressure p at the compensation depth Z is constant for a floating equilibrium, where the overburden pressure p_L, eq. (1.3), is computed for a column of height $h = Z + t$. Isostatic topography, t, is relative to a hypothetical spreading-center reference height H, where $l_c = l_l = 0$.

If $\boldsymbol{g}$ is the *gravitational acceleration*, i.e., the pull due to gravity, e.g., of our planet, then eq. (1.1) allows us to infer the weight of an object, and $\boldsymbol{g}$ points more or less vertically down toward the center of the Earth. If that m object is our mantle column considered for the floating equilibrium, with area A and volume V on top of the compensation depth is $m = \rho V = \rho h A$, with *density*

$$\rho = \frac{m}{V}, \tag{1.2}$$

and topography z and $h = t + Z$ (Fig. 1.3). The weight of that crustal column is thus $F = mg = \rho g h A$, with the amplitude, or length/norm, of $\boldsymbol{g}$, $g = |\boldsymbol{g}|$ (eq. C.22). The pressure (force per area, units of N/m^2 or Pa) is therefore (ignoring atmospheric pressure)

$$p_l = \rho g h = \int \mathrm{d}z \rho(z) g(z), \tag{1.3}$$

where the integral over depth, $\int \mathrm{d}z$, means sum up the area under a function (§C.2.4), and replaces the product if ρ and/or $g(z)$ are functions of depth. This lithostatic, or overburden, pressure is identical to the hydrostatic pressure in a fluid at rest, and the depth derivative of eq. (1.3),

$$\frac{\mathrm{d}p}{\mathrm{d}z} = -\rho g, \tag{1.4}$$

is called the *hydrostatic equilibrium equation*.

Let us assume that the crust is part of the thicker *lithosphere* (Greek *lithos* (λίθος) = stone, here: strong), which is a mechanical term that encompasses the relatively slowly deforming, cold surface regions of the Earth. It sits on top of the *asthenosphere* (Greek *asthenos* (ἀσθενής) = without strength, weak) which can flow so that things equilibrate. The timescales for this equilibration after load removal, e.g., due to erosion or deglaciation at the surface, or removal of dense layers by tectonic processes, depends on the creep behavior of the underlying mantle and the wavelength of perturbations (§6.7.1).

Assuming that density is constant with each layer (following Airy), this yields a balance for a column floating at topography z_{iso} under a cover of air (Fig. 1.3, and see exercise 2). Comparing the lithospheric and reference column in Fig. 1.3, we can write the pressure at compensation depth Z and asthenospheric density ρ_a as

$$g\left(\rho_c l_c + \rho_l l_l + (Z + t - l_c - l_l)\rho_a\right) = gZ\rho_a$$
$$(\rho_c - \rho_a)\, l_c + (\rho_l - \rho_a)\, l_l + t\rho_a + Z\rho_a = Z\rho_a$$
$$t = \frac{\rho_a - \rho_c}{\rho_a} l_c + \frac{\rho_a - \rho_l}{\rho_a} l_l. \tag{1.5}$$

Any asthenospheric layer below l_l thus cancels out irrespective of the actual choice for Z. However, this is only the case under the assumption that there are no asthenospheric density variations. Those complicate analysis of the nonisostatic residuals and considerations of the origin of topography (§9.2.1).

Allowing for water coverage, we can write the isostatic elevation of the continental lithosphere relative, e.g., to a hypothetical spreading center reference level, H, with no oceanic crust, $l_c = 0$, and zero plate thickness, $l_l = 0$, by definition, as

$$z_{\mathrm{iso}}^{\mathrm{land}} = f_1 l_c + f_2 l_l - f_3 H \quad \text{if} \quad z \geq 0 \text{ and} \tag{1.6}$$
$$z_{\mathrm{iso}}^{\mathrm{water}} = \frac{z_{\mathrm{iso}}^{\mathrm{land}}}{f_3} \quad \text{if} \quad z < 0,$$

with

$$f_1 = \left(1 - \frac{\rho_c}{\rho_a}\right), \quad f_2 = \left(1 - \frac{\rho_l}{\rho_a}\right), \quad \text{and} \quad f_3 = \left(1 - \frac{\rho_w}{\rho_a}\right) \approx \frac{1}{1.44},$$

where ρ_w is the density of water, with $H \approx 2.6 \ldots 2.7$ km (§7.3.2, §10.3.5). With $l_l = 0$ or no lithospheric variations, eq. (1.6) shows that high topography at constant crustal density has to be balanced by thicker crust, such that the depth to the base of the crust is variable (*Airy*, 1855). The factors relating crustal and lithospheric layer thickness and topography are of order $f_1 \sim 0.1 \ldots 0.2$ and $f_2 \sim -0.02 \ldots -0.01$, respectively (Fig. 1.4). This means that variations in crustal thickness have ~ 10 times the effect of variations in lithospheric thickness, albeit with larger uncertainties for l_l (§9.2.1).

Alternatively, the depth from sea level to the base of the crust or lithosphere, P, can be chosen to be the same as the compensation depth, $P = Z$. In this case, floating equilibrium for variable surface topography can also be achieved if the crustal density, ρ_P, is variable, e.g., reduced underneath high topography. This sort of compensation is attributed to Pratt, who in *Pratt* (1855) attempted to explain the Himalayan plumb line measurements with variable crustal density, akin to this line of reasoning. If compensation were to happen at a constant crustal bottom depth Z, the same kind of balance shows that the required density is

$$\rho_P = \rho_0 \frac{Z}{Z + t}, \tag{1.7}$$

where ρ_0 is the reference density at the spreading center. If a lithospheric column is perfectly in *Airy isostasy* or *Pratt isostasy*, it is called *fully compensated* (§3.2.2.2).

Granitic continental crust is lighter, i.e., less dense, than the basalt (extrusive) and gabbroic (intrusive) oceanic crust, which is also underlain by a relatively denser, colder convective boundary layer, and thus mainly thermally controlled, lithosphere. These bulk differences explain much of

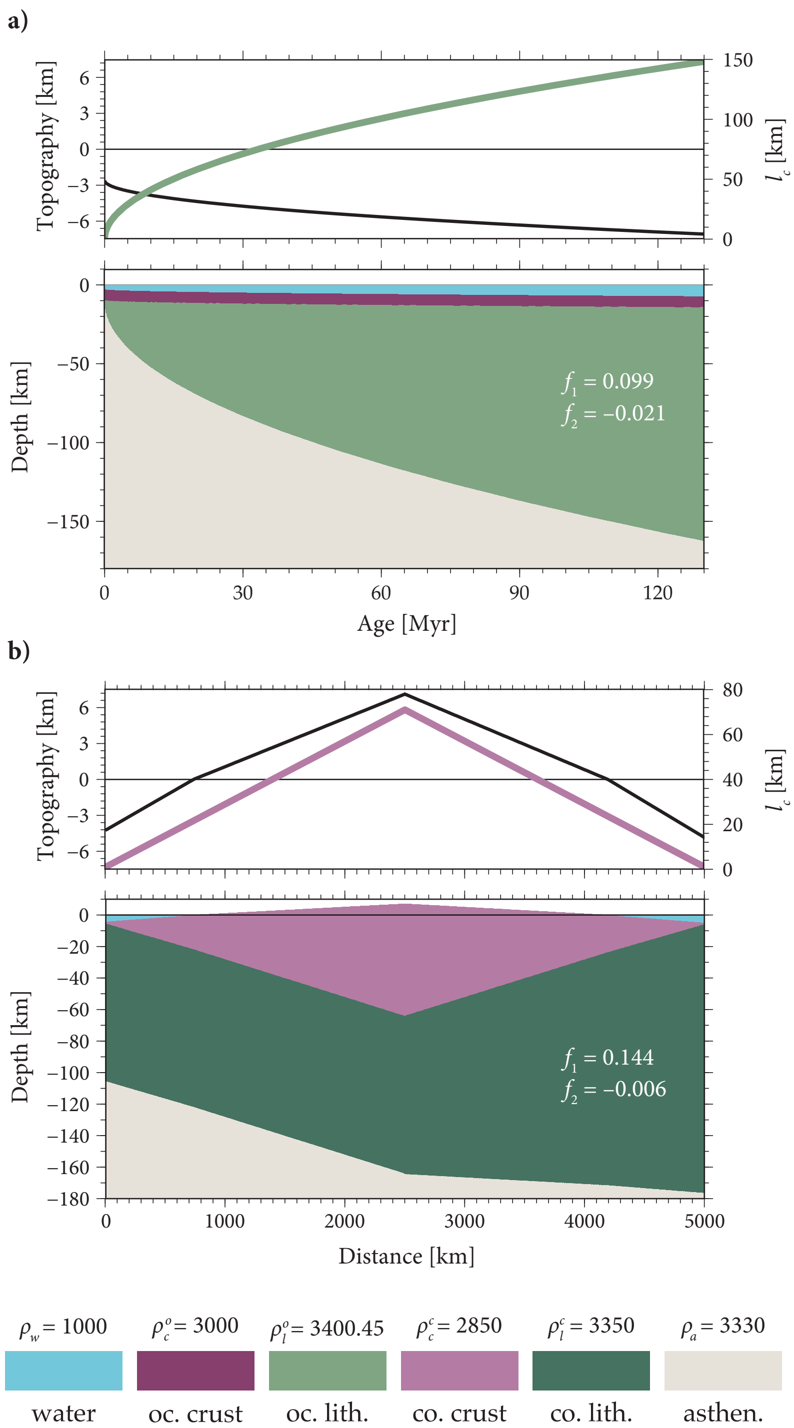


Figure 1.4: Nominal oceanic (a) and continental (b) lithospheric columns in isostatic balance, eq. (1.6), using the density values as provided in the color legend (units of kg/m^3). Top subplots show topography (black) and lithospheric, l_l^o, and crustal, l_c^c, thickness, respectively. The oceanic plate is represented by constant crustal thickness, $l_c^o = 7$ km, and a half-space cooling (§7.3.2) dependence of lithospheric thickness, l_l^o, following eq. (7.45), with constant lithospheric density, ρ_l^o, as per eq. (7.51), but can also be viewed as having constant l_l^o and variable ρ_l^o in a Pratt sense. In order to have the zero age, $l_l^o = 0$, plate rest at realistic $H \sim 2.6$ km water depth for nonzero l_c^o at the ridge, eq. (1.5) was corrected by adding $\Delta H l_c^o(\rho_c^o - \rho_a)/(\rho_w - \rho_a)$ to $\hat{H} = H + \Delta H \approx 3.6$ km. Continental crustal thickness is for a nominal orogen where the left and right half have constant l_l^c and constant $l_c^c + l_l^c$, respectively. Inset f_1 and f_2 values are the thickness multipliers of eq. (1.6).

the topography of Earth (Fig. 1.2) as a mix of mostly Airy isostasy for the continents and Airy and Pratt isostasy for the oceanic lithosphere (Fig. 1.4). There, we can apply the half-space cooling model to explain much of the deepening of the seafloor away from midoceanic spreading centers, one of the major achievements of geodynamics (§7.3.2).

Besides plate-scale features in the oceans and the continent-ocean difference, curious alien observers of our planet's topography might further identify linear chains of islands on the seafloor (Fig. 1.1a), and might make the connection to hotspot volcanism and moving plates (§10.1). They would also see that the suboceanic mountains show midoceanic ridges in the Atlantic but more subdued slopes within the Pacific. They might suggest different types of oceanic crust production, which we know to be due to slow and fast spreading, respectively (§10.3). Along the spreading centers, there are characteristic offsets by transform faults, another hallmark of Earth's style of plate tectonics and indicative of the generation of strike-slip-type motion by some means of strain-localizing rheologies (§8.5). Continents reflect a much longer geological record of more complex processes, with the high plateau of Tibet and the Andes being prominent topographic features, and other orogens, such as the North American Cordillera being more subdued, with evidence for water and wind-driven surface transport processes and erosion.

Even from such global-scale comparative, qualitative assessments of the outer terrestrial planets, it is thus clear that not only the surface morphology of planets but also their mass and energy transport processes, including climate and perhaps life, depend on interior dynamics through mantle convection. The latter is expressed, within our solar system only for Earth, as plate tectonics at the surface. However, the internal feedbacks are not limited to those expressed by the mantle at the surface. At its base, mantle convection controls core cooling, and hence magnetic field generation through the convective geodynamo operating in the liquid outer core (§1.4.1). The heat flow imposed on the core by the mantle may matter directly for the types and stability of the magnetic field, and indirectly through regulating inner core freezing. The field in turn may play a role for sustaining an atmosphere by shielding planets from solar radiation, another example of systems-level feedbacks. Our sister planet Venus is lacking a magnetic field at present, which is another indication of different planetary evolutionary trajectories (§9.4.1).

1.2 Geopotentials: Shape, spin, and geoid

If our aliens are able to put satellites into orbit around Earth and track their orbits precisely, as we have done for Earth, Mars, and Venus (Fig. A.5), they would be able to determine models of the spatial variation of the gravitational attraction one experiences in different places. Those variations arise due to internal density anomalies (Fig. 1.5) and lead to subtle fluctuations in the distance between the satellite and the planet, as exploited by terrestrial spacecraft missions of the GRACE series by having two satellites follow each other through the ups and downs of their orbits.

1.2.1 Gravitational potential

Newton's first law states that the gravitational force, $\boldsymbol{F}$, exerted by a mass M, e.g., a planet, on a mass m, e.g., a satellite, is given by

$$\boldsymbol{F} = -G\frac{mM}{r^2}\hat{\boldsymbol{e}}_r \quad \text{or} \quad F = |\boldsymbol{F}| = G\frac{mM}{r^2}, \tag{1.8}$$

where r is the distance between the objects (their centers of mass, assumed as points), $\hat{\boldsymbol{e}}_r$ is a unity vector, i.e., $|\hat{\boldsymbol{e}}_r| = 1$, pointing in from the M object to the m object's center of mass, and G is the *universal gravitational constant* (one of nature's less well constrained fundamental constants;

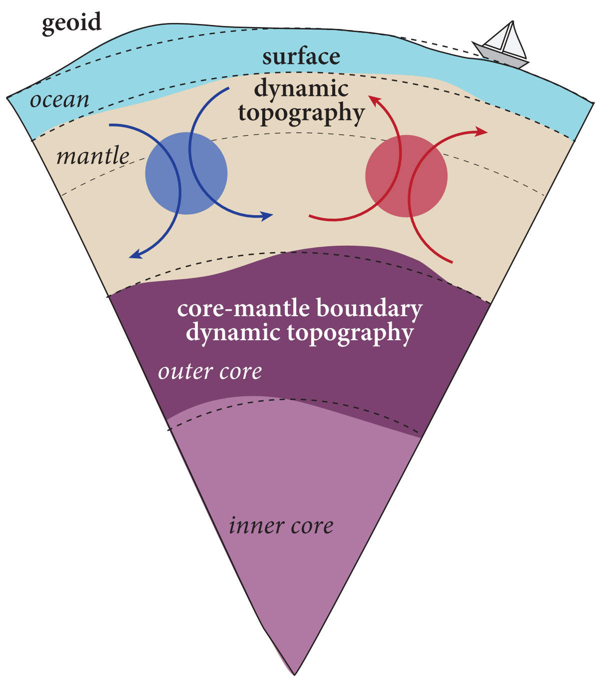


Figure 1.5: Drawing illustrating the definition of the geopotential surface of the actual geoid and its deflection from a reference geoid of ellipsoidal form shown with dashed lines (Fig. A.3). Internal density anomalies, such as due to a dense and light anomaly (blue and red, respectively), have a static and dynamic response on the geoid. Static, since the excess mass of a positive geoid anomaly means one has to move further away to experience the same gravitational pull, eq. (1.35). If the anomaly introduces flow, the resulting stresses also deflect the surface and core-mantle boundary, and those undulations lead to additional, dynamic effects on the geoid on timescales of ∼ Myr (§9.4). Here, we assume a higher viscosity lower mantle which leads to reduction of the surface relative to the bottom topographic deflection, and an effective positive geoid anomaly for a mid-mantle slab sinker (§9.4). Another time-dependent effect on the geoid is due to glacial isostatic adjustment, on timescales of melting, ∼ 26,000 yr for the last glacial maximum (§6.7).

Table A.1). If we consider the total work, W, involved in moving an object m from far away, at infinite distance, toward another with mass M at R, say, the surface of the Earth, we need to sum up, i.e., again integrate, the spatially dependent force that is required to do that. Using eq. (1.8) and integrating against that force (note sign),

$$W = \int_{\infty}^{R} F(r)\,\mathrm{d}r = GMm \int_{\infty}^{R} r^{-2}\,\mathrm{d}r = GMm\left[r^{-1}\right]_{\infty}^{R} = -GMm\left[\frac{1}{R} - \frac{1}{\infty}\right] = -\frac{GMm}{R}, \qquad (1.9)$$

which defines the *gravitational energy* of the two masses, $E_p = W$. The energy is negative because work has to be done to separate them to infinite distances such that energy is conserved, and energy, or work, has units of force × length, i.e., Nm, or J (joules).

Since the gravitational force is also given by Newton's second law, eq. (1.1),

$$\boldsymbol{g} = -\frac{GM}{r^2}\hat{\boldsymbol{e}}_r, \quad \text{or} \quad |\boldsymbol{g}| = \frac{GM}{r^2} \qquad (1.10)$$

holds, which defines the gravitational acceleration, $\boldsymbol{g}$, for the satellite, or for any observer at the surface of a homogeneous sphere with radius R such that $g = \frac{GM}{R^2}$, and the negative sign indicates that the acceleration is downward from the normal on the surface of the M object. Eq. (1.10) means that the gravitational attraction of the planet is fully described locally by $\boldsymbol{g}$; it is only the actual force (eqs. 1.1 and 1.8) that depends on the test mass m, e.g., of the satellite, that we are considering.

After things have settled down following planetary formation, any planet's mass distribution will mainly be layered, since denser material sinks to the bottom in the gravitational field, e.g., to form a predominantly iron core and silicic-type mantle. However, things get more complicated because there might be lateral variations of density, e.g., due to internal convective dynamics (Fig. 1.5). To capture these complexities, along with the possibility of having a rotating planet, we can express the gravitational force through a *potential*, U, which at a distance r from the center of the planet is defined as

$$U(r) = U_g + U_{\mathrm{rot}} = -G \int_V \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}|} \, \mathrm{d}V' - \frac{1}{2}\Omega^2 r_0^2 \cos^2 \lambda \tag{1.11}$$

(see Expanded details 1), where $\int_V$ indicates integration over mass density at $\boldsymbol{r}$ from the mass center of the planet. In contrast to the vector $\boldsymbol{g}$, U is a scalar; the value of U depends on location, but there is only a single number defining it at that location, which is part of the motivation for introducing U. U is hence a *scalar field*, like topography expressed on a topo map.

For a spherical planet,

$$U_g(r) = -\frac{GM}{r} \tag{1.12}$$

is the *point mass potential* equivalent to a mass M at the center, and recovers eq. (1.9) for a unit test mass $m = 1$. The second term on the right-hand side (RHS) of eq. (1.11) is the rotational component of the potential, U_{rot}, and accounts for the centrifugal force due to a spin. The angular velocity of the Earth at the present day is $\Omega_E = 7.292,211,5 \cdot 10^{-5}$ rad/s as of the *World Geodetic System* (WGS-84; *NIMA*, 1984).

U_{rot} increases from zero at the pole of the rotation axes (latitude $\lambda = 90°$) to maximum at the equator ($\lambda = 0$). Also in contrast to eq. (1.1), eq. (1.12) has a $\frac{1}{r}$ dependence, and bringing things back from U to g shows why. It takes work to move an object from $U(r_0)$ to $U(r_0 + \mathrm{d}r)$ radially within the potential; that change is $\mathrm{d}U = -g_r \, \mathrm{d}r$, and we can recover g from the radial derivative of U (§C.2.2):

$$g_r = -\frac{\partial U}{\partial r} \quad \text{or, more generally,} \quad \boldsymbol{g} = -\begin{pmatrix} \frac{\partial U}{\partial x} \\ \frac{\partial U}{\partial y} \\ \frac{\partial U}{\partial z} \end{pmatrix} = -\boldsymbol{\nabla} U, \tag{1.13}$$

i.e., the gravitational attraction is a vector that aligns with the gradient of the potential field, U. The $\boldsymbol{g}$ vector points away from the steepest gradient, measured by the *gradient operator*, $\boldsymbol{\nabla}$ (which is a vector; §C.5). For a nonrotating, homogeneous sphere, eq. (1.12) holds, which also follows from integration of eq. (1.10) or the stationary part of eq. (1.11). One can show that eq. (1.13) is a sufficient condition for the gravitational force, eq. (1.1), to be conservative.

EXPANDED DETAILS 1: Moments of inertia and geopotentials of an ellipsoidal Earth

Consider the *moment of inertia* for rotation for a location at $\boldsymbol{r}$ given a spin vector $\boldsymbol{\omega}$,

$$I(\boldsymbol{r}) = \int (r' \sin\theta)^2 \, \mathrm{d}M \overset{\text{point mass}}{=} mr^2$$

(units: kgm^2), where $r' \sin\theta$ is the distance between $\boldsymbol{r}$ and $\boldsymbol{\omega}$, and $\mathrm{d}M$ indicates integration over all mass elements, hence weighted by distance squared. The *angular momentum* is $\boldsymbol{L} = \boldsymbol{I}\boldsymbol{\omega}$ with a moment of inertia tensor, $\boldsymbol{I}$,

$$I_{xx} = \int (y^2+z^2)\,\mathrm{d}M, \quad I_{yy} = \int (x^2+z^2)\,\mathrm{d}M, \quad I_{xy} = -\int xy\,\mathrm{d}M, \quad \text{and similar,}$$

and for the point mass

$$L = mr^2\omega, \tag{1.14}$$

which implies that since L is conserved for a closed system, r has to vary with any changes in ω, and vice versa, as for ice skaters extending their arms, for example.

In a reference frame with $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$, $\boldsymbol{I}$, the moment of inertia depends on direction cosines i_i, with, e.g., $i_3 = \cos\theta$, where θ is the angle between $I(\boldsymbol{r})$ and the $\hat{\boldsymbol{z}}$ axis, the colatitude, and $I = i_1^2 A + i_2^2 B + i_3^2 C$ with $i_1^2 + i_2^2 + i_3^2 = 1$ and

$$A = \int \left(\hat{y}^2 + \hat{z}^2\right)\mathrm{d}M, \quad B = \int \left(\hat{x}^2 + \hat{z}^2\right)\mathrm{d}M, \quad \text{and} \quad C = \int \left(\hat{x}^2 + \hat{y}^2\right)\mathrm{d}M. \tag{1.15}$$

If two moments of inertia are equal, $A = B$, then $I = A + i_1^2(C - A)$.

Let us move our mass integration for the potential into the center of mass, and further align the $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ so that they are the eigenvectors of $\boldsymbol{I}$; then $\boldsymbol{I}$ has no off-diagonal elements (§2.4.4). This makes A, B, and C the *principal moments of inertia*. In this moment of inertia frame, the general U_g term of eq. (1.11) is given by *MacCullagh's formula*,

$$U_g(r) = -\frac{GM}{r} - \frac{G}{2r^3}\left(A + B + C - 3I(\boldsymbol{r})\right). \tag{1.16}$$

For a spherically symmetrical body, $A = B = C = I$ holds, meaning that eq. (1.16) reduces to eq. (1.12).

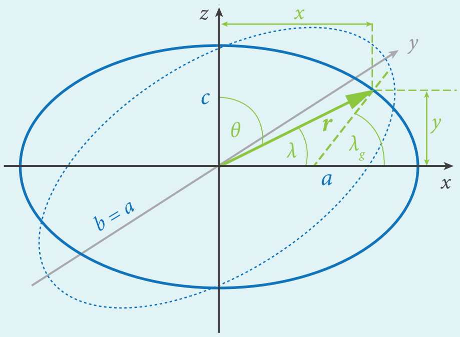


Figure 1.6: Coordinates and geometry for an oblate ellipsoid, or spheroid, with axes $a = b > c$. A location on the surface in geocentric coordinates $\boldsymbol{r} = x, y, z$ is described by longitude and *geocentric latitude* λ and colatitude $\theta = 90° - \lambda$, or *geographic latitude* λ_g, defined with respect to the normal to the surface.

Now consider an Earthlike object, which is as an approximation an oblate, squashed *spheroid*. This is a special case of an even, long-axes, *$a = b$ ellipsoid*, with long axis a along the equator (distance from the center), and shorter axis c to the pole aligned with the rotational axis (Fig. 1.6). If we then define C to be the *polar moment* and A and B, to be *equatorial moments*, such that $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ are aligned with the a, b, and c axes of the ellipsoid, then

$$I(\boldsymbol{r}) = A + (C - A)\cos^2\theta.$$

One can then show that

$$U_g(r) = -\frac{GM}{r} + \frac{G}{r^3}(C-A)\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) = -\frac{GM}{r} + \frac{GMa^2}{r^3}J_2\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right), \quad (1.17)$$

with a as the equatorial radius and J_2 as the *dynamic flattening parameter*,

$$J_2 = \frac{C-A}{Ma^2} \approx 1.081 \cdot 10^{-3}, \quad (1.18)$$

where J_2 can be determined in different ways. The relevance of this equation is that our aliens (or ourselves for other planets) can determine J_2 from remote observations of the gravitational potential, and eq. (1.18) links this to properties that depend on the internal density distribution of the object, e.g., the fractionation into a core and mantle once the geometrical property a is determined.

The potential on the reference ellipsoid best matching the spheroid form of our planet with axes a and c, U_{ref}, which we might use for navigation, defines the *reference geoid*, where a geoid is a surface of constant gravitational potential U. Thus U_{ref} is the closest approximation of the geoid (Fig. 1.5) that can be fit by a spheroid. Even though the ellipticity inherent in U_g from eq. (1.17) results from a planet's spin, that potential does not include the effects of rotation. That motion around an axis, $\hat{\boldsymbol{e}}_z$, contributes

$$U_{\text{rot}}(\boldsymbol{r}) = -\frac{1}{2}\Omega^2 r^2 \sin^2\theta = -\frac{1}{2}|\boldsymbol{\omega} \times \boldsymbol{r}|^2, \quad (1.19)$$

with $\boldsymbol{\omega} = \Omega\hat{\boldsymbol{e}}_z$, where $\boldsymbol{\omega} \times \boldsymbol{r}$ denotes the cross product (§C.3.3.3). Eq. (1.19) follows from an integration like eq. (1.9) with unity mass, but instead of using the gravitational acceleration, $\boldsymbol{g}$, we use the *centrifugal acceleration*, $\boldsymbol{p} = \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r})$.

The total potential outside a rotating, Earthlike reference spheroid is then

$$U(r) = -\frac{GM}{r} + \frac{GMa^2}{r^3}J_2\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) - \frac{1}{2}\Omega^2 r^2 \sin^2\theta. \quad (1.20)$$

From eq. (1.20), at the poles where $r = c$ and at the equator where $r = a$,

$$U_{\text{ref}}^{\text{pole}} = -\frac{GM}{c} - \frac{GMa^2}{c^3}J_2 = U_{\text{ref}}^{\text{equator}} = -\frac{GM}{a} - \frac{GM}{2a}J_2 - \frac{1}{2}a^2\Omega^2. \quad (1.21)$$

The *geometric flattening*, or ellipticity, of the reference geoid can then be derived as

$$f = \frac{a-c}{a} = J_2\left(\frac{c}{2a} + \frac{a^2}{c^2}\right) + \frac{1}{2}\frac{a^2 c\Omega^2}{GM} \approx \frac{1}{2}(3J_2 + m), \quad f_{\text{WGS-84}} = \frac{1}{298.257,223,563}, \quad (1.22)$$

with what turns out to be the ratio between centrifugal to total gravity at the equator:

$$m = \frac{a^3\Omega^2}{GM} = 3.461,391,899 \cdot 10^{-3},$$

with $a = 6{,}378{,}137.0$ m and $GM = 3.986,004,418 \cdot 10^{14}$ m^3/s^2 of the World Geodetic System (WGS-84; *NIMA*, 1984). Eq. (1.22) yields another approximation for $J_2 = \frac{(2f-m)}{3} = 1.08141 \cdot$

10^{-3}. We get $U_{\text{ref}} = 62,636,860.8497\ \text{m}^2/\text{s}^2$, and then $c = 6356.7523$ km; thus the difference between equatorial and polar radial axes is $a - c \approx 21$ km.

The geodetically observed flattening as in, e.g., WGS-84 is actually slightly larger than that expected theoretically for a rotating, hydrostatic Earth, with expected values of $f_H = \frac{1}{299.829}$ (*Nakiboglu*, 1982) and $f_H = \frac{1}{299.981}$ (*Chambat et al.*, 2010) for different theoretical approaches. The difference between f and f_H becomes significant when interpreting geopotential field anomalies for glacial isostatic adjustment (Expanded details 18) or as a constraint for mantle convection (§9.4). The observed *hydrostatic bulge*, $a - c$, is 113 m larger than the $a - c$ that is expected from a hydrostatic Earth (*Chambat et al.*, 2010). Different corrections are shown in Fig. A.3, but it is clear that J_2 is dynamically supported by mantle convection where it evolves over timescales of $\sim$ 50 Myr (*Ricard et al.*, 1984, 1993; *Steinberger & O'Connell*, 1997). On even longer, $\sim$ Gyr timescales, the slowdown of Earth's rotation through tidal dissipation will decrease J_2 (§1.2.2), and deglaciation since the last glacial maximum can be constrained by secular, present-day $\dot{J}_2$ (*Yuen & Sabadini*, 1985; *Mitrovica & Peltier*, 1993). The deglaciation trend is of order $\dot{J}_2 \sim -3 \cdot 10^{-11}\ \text{yr}^{-1}$ and needs to be accounted for to understand present-day contributions due to melt water influx and climatic cycles (Expanded details 18; *Cazenave & Nerem*, 2004; *Cheng et al.*, 2013).

For our purposes here, $\dot{J}_2$ is super tiny ($\sim$ 0.02 m of reduced bulge per 100 yrs) to require adjusting WGS-84 over decadal scales. Using the relationship between $\boldsymbol{g}$ and U from eq. (1.13) and eq. (1.20) and only considering the radial component of $\boldsymbol{g}$,

$$g(r) \approx \frac{GM}{r^2} - 3\frac{GMa^2}{r^4} J_2 \left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) - \Omega^2 r \sin^2\theta.$$

The distance from the center of the Earth to the spheroidal reference geoid, or the approximation of r for an ellipse, is

$$r_{\text{ref}}(\theta) \approx a\left(1 + \frac{f(2-f)}{(1-f)^2}\cos^2\theta\right) \approx a(1 - f\cos^2\theta). \tag{1.23}$$

Averaging of eq. (1.23) leads to the mean radius of the Earth, R_E (Table A.1). We can also derive a *reference gravitational acceleration* on the reference geoid at r_{ref} as

$$\begin{aligned} g_{\text{ref}}(\theta) &\approx \frac{GM}{a^2}\left(1 + 2f\cos^2\theta - 3J_2\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) - m\sin^2\theta\right) \\ &\approx \frac{GM}{a^2}\left(1 + \frac{3}{2}J_2\sin^2\theta - m + 2m\cos^2\theta\right), \end{aligned} \tag{1.24}$$

which can be written as a function of latitude, $\lambda = 90° - \theta$:

$$g_\lambda = 9.780{,}327(1 + 0.005{,}302{,}4\sin^2\lambda - 0.000{,}005{,}9\sin^2 2\lambda)\ \text{m/s}^2. \tag{1.25}$$

We can define a *gravity flattening* factor, analogous to eq. (1.22), as

$$f^* = \frac{g_{\text{pole}} - g_{\text{equator}}}{g_{\text{equator}}} = \frac{-\frac{3}{2}J_2 + 2m}{1 + \frac{3}{2}J_2 - m}, \tag{1.26}$$

and eq. (1.24) can then be rewritten like eq. (1.23) as another ellipse:

$$g_{\text{ref}}(\theta) = g_{\text{equator}}\left(1 + f^* \cos^2\theta\right).$$

Within these approximations, *Clairaut's theorem* (from 1743) describes this interesting relationship between geometrical and dynamic properties,

$$f^* = \frac{5m}{2} - f,$$

i.e., f can be inferred from gravity observations alone, for example. The links of J_2 with the satellite geoid are further explored in Expanded details 2.

1.2.2 Reference geoid, spin, and the Earth-Moon system

On global scales, we can then use the geoid to describe variations in gravitational pull. The *geoid* is a constant gravitational potential surface, defined to coincide with the average sea surface height in the absence of $\lesssim$ yr timescale oceanographic fluctuations, i.e., *sea surface equipotential*, or *sea surface height* (Fig. 1.5). In continents, the geoid would be the level of water in an imaginary set of interconnected canals, ignoring any extra pull from any surrounding mountains.

We call the best-fitting spheroidal approximation to the geoid the *reference geoid*, based on evaluating the gravitational potential on the *reference ellipsoid*, which describes the oblate approximation to the shape of the Earth (Expanded details 1). Since the ellipsoidal geometry dominates the actual geoid, one typically makes a correction before analysis of *geoid anomalies*. The best-fitting spheroidal shape that is observed for Earth is that of an ellipsoid where the equator is ≈ 21 km further from the center of the Earth than the poles. While this shape is mainly due to the spin, the degree of ellipticity is slightly larger than that expected for a spinning planet with a 1-D Earth structure, the purely *hydrostatic geoid* (§A.4). This means that there is a dynamic, mantle-convection-related contribution to the equatorial bulge itself (§9.4). It has long been recognized that the Moon's ellipticity is likewise too large (*Laplace*, 1823), in fact much too large, at ~ 20 times the hydrostatic value (*Keane & Matsuyama*, 2014). Unlike for the Earth where dynamic processes sustain a slight extra ellipticity, the Moon has a *fossil bulge* (*Jeffreys*, 1915), frozen in during the early planetary evolution, perhaps at ~ 4 Ga (*Lambeck & Pullan*, 1980; *Qin et al.*, 2018). While that bulge is now static, its formation is one way in which the Earth and Moon are coupled through their orbits, internal dissipation, and angular momentum.

In particular, sedimentary deposits on Earth record tidal cycles, and such constraints indicate that our planet experienced ≈ 22 h days at the end of the Neoproterozoic ($\sim$620 Ma; *Williams*, 1989), and ≈ 19 h days at 1.4 Ga (*Meyers & Malinverno*, 2018), i.e., a $\approx 25\%$ faster spin rate. The effects of the Coriolis forces due to planetary rotation are important for the geodynamo as well as for organizing atmospheric and ocean circulation (Expanded details 15), and hence possibly paleoclimate and the biosphere (*Walker*, 1982; *Olson et al.*, 2020). Perhaps less exciting, faster spin also implies a $\sim 60\%$ larger ellipticity at that time, since the hydrostatic bulge $\propto \Omega^2$ (eq. 1.11). The geological timescale slowdown of Earth's spin toward the present day is caused by tidal dissipation, which depends on the internal structure of the planet for solid tides, and on ocean circulation dynamics as affected, e.g., by plate-tectonics-dependent bathymetry and continental distribution. Oceanic processes make up the bulk of Earth's tidal dissipation at present, but both contributions have evolved over planetary history (*Ross & Schubert*, 1989; *Bills & Ray*, 1999; *Tyler*, 2021). Since the total angular momentum of any closed system has to be conserved, the Moon has moved away from Earth at ≈ 2.2 cm/yr on average since 620 Ma (*Williams*, 2000) (eq. 1.14), and this separation rate has sped up to ≈ 3.8 cm/yr at present (*Dickey et al.*, 1994), prolonging our aliens' journey.

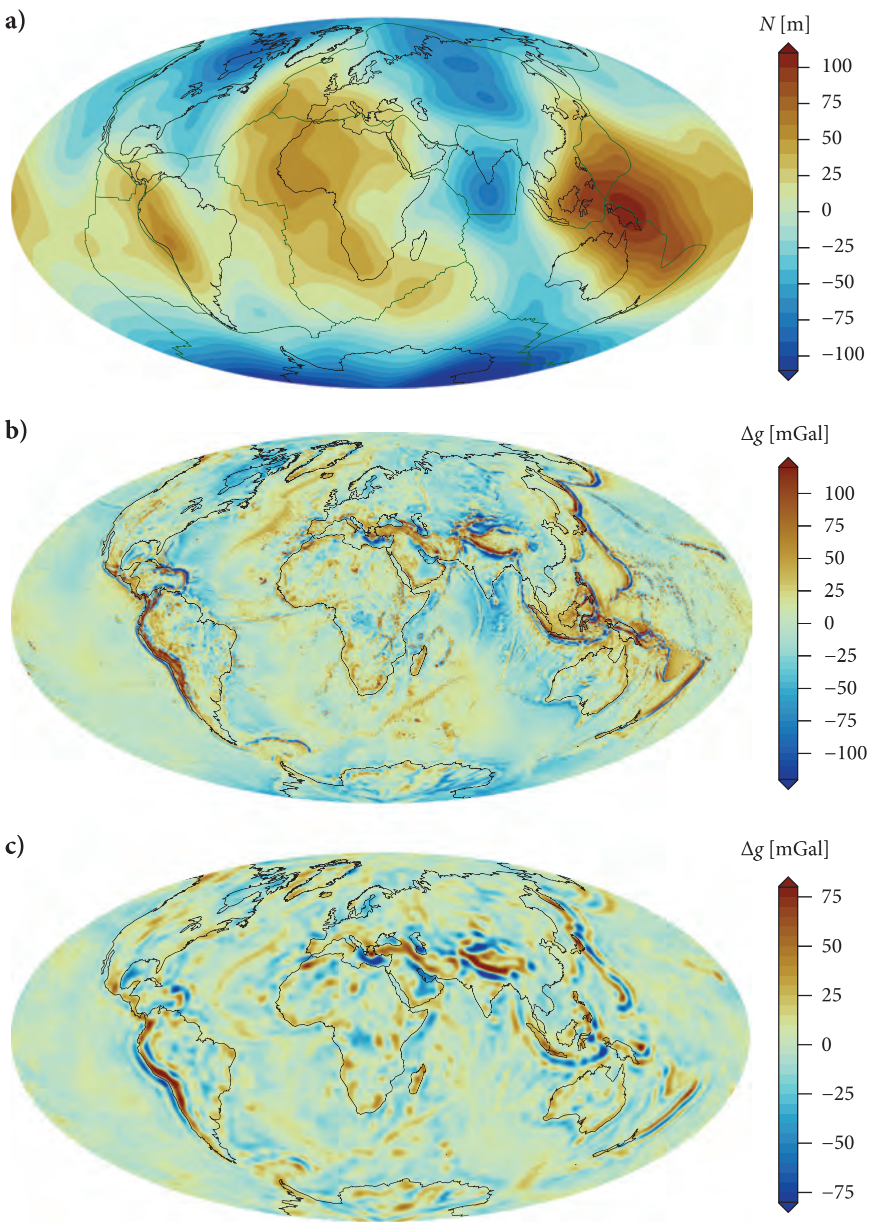


Figure 1.7: EGM 2008 (*Pavlis et al.*, 2012) geopotential model up to spherical harmonic degree $L = 512$ (§C.6.2). **a)** Nonhydrostatic geoid height anomalies, N of eq. (1.31), corrected following *Chambat et al.* (2010) (cf. Fig. A.3). **b)** Equivalent free-air gravity anomalies (cf. Fig. A.3d). **c)** Medium wavelength ($\ell = 10 \ldots 80$, $\cos^2$-smoothed), band-pass-filtered free-air gravity anomalies as often used to infer admittance-based dynamic topography, δz (e.g., $\delta z \sim \delta g/50$ km/mGal; *Craig et al.*, 2011). See Fig. A.5 for a comparison of topography-geoid spectra for terrestrial planets.

1.2.3 Geoid and gravity anomalies

Figure 1.7a shows geoid anomalies with respect to the expected hydrostatic shape of the Earth (Expanded details 2). This satellite-based geoid anomaly map reveals that the geoid is in reality quite a bumpy landscape for Earth. Latitudinally, most positive anomalies globally are found around the equator, and the poles show negative anomalies. This pattern might be expected if the planet were to continuously reorient its excess density in the mantle and crust with respect to the spin axis such that the major moment of inertia, eq. (1.15), is oriented with excess mass moved to the equator. This way of shifting the entire mass of the Earth, including its surface, with respect to the rotation axis is called *true polar wander* (TPW). Mantle convection as well as glacial isostatic adjustment further change the geoid on a range of timescales (§6.7 and §9.4).

Without constraints about the internal structure of the planet, our aliens could not make a hydrostatic correction, but they could nonetheless explore, e.g., the correlation between topography and the geoid (§A.5), and muse as to the mechanisms of generation of the anomalies at different wavelengths (§3.2.2). When viewed in light of the interpretation of the oceanic topography, the aliens may note that there is no clear relationship between spreading centers and geoid highs, for example, which we now interpret as implying that spreading is predominantly passive, i.e., not driven by active, hot upwellings. There are large, positive geoid anomalies in places that we now associate with the main drivers of plate motions, subducted slabs, and deep mantle anomalies, but their proper interpretation relies on more sophisticated models of internal dynamics (§9.4).

Given that the geoid is predominantly ellipsoidal, the gravitational pull also mainly varies with latitude, as we are experiencing different degrees of the centrifugal force, eq. (1.25). We therefore consider anomalies from that reference gravity (Expanded details 2). Given that geoid anomalies are typically expanded as spherical harmonics, eq. (1.30), it is straightforward to manipulate the field, and, e.g., *free-air gravity anomalies*, eq. (1.34), at the reference level can be obtained from the geoid deflection's (N of eq. 1.31), spherical harmonic coefficients by multiplying them, at a given spherical harmonic degree ℓ, by an $(\ell-1)$-dependent factor (Expanded details 2). The resulting free-air anomalies (Fig. 1.7b) therefore look much rougher, or more detailed, than the geoid (Fig. 1.7a), since the derivative operation of eq. (1.13) leads to an $\ell-1$ enhancement of higher spatial frequency power, eq. (C.67) (Fig. A.4). Tectonic analysis often uses gradients of the free-air anomalies to further enhance local variations, but note that the true information content in all those gravity "products," as in Fig. 1.7, is identical.

Expressing gravity in a harmonic basis also allows filtering by wavelength. At the shortest wavelengths, $\ell \gtrsim 100$, a strong lithosphere is able to support surface topography by flexure rather than isostasy, leading to large free-air gravity anomalies that are well correlated with topography (§3.2.2). On longer than flexural wavelengths, isostatic adjustment over geologic timescales leads to small gravity anomalies and decorrelation with topography, and for $\ell \lesssim 20$, we expect mantle dynamics to matter (§9.4).

While any direct interpretation of gravity anomalies is highly dependent on a range of assumptions, mid-wavelength free-air gravity (Fig. 1.7c) is sometimes used to identify "anomalous," that is, uncompensated, or "dynamic" topography. There are a number of features in this map that we return to (§9.2), such as the W-E tilting of Arabia, indication of mantle support of the Atlas Mountains of NE Africa, a marked depression in central Africa, and a characteristic pattern of free-air highs and lows along subduction arcs.

EXPANDED DETAILS 2: Geoid, spherical harmonics, and gravity anomalies

If we apply the divergence theorem, eq. (C.55), on $\boldsymbol{g}$, and use eqs. (1.10) and (1.13) considering a small volume V of homogeneous mass with surface area $4\pi r^2$, we can heuristically derive *Poisson's equation* for the gravitational potential,

$$\nabla^2 U(\boldsymbol{r}) = 4\pi G\rho(\boldsymbol{r}), \tag{1.27}$$

where ρ is density and ∇^2 is the *Laplacian* operator, in Cartesian $\left\{\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right\}$ (§C.5). Eq. (1.27) is called the *Laplace equation* if the RHS is zero, which is the case for the potential at and above the Earth's surface. Written in spherical coordinates with longitude ϕ and colatitude θ (eq. C.50),

$$\nabla^2 U = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial U}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial U}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 U}{\partial\phi^2} = 0. \tag{1.28}$$

The type of second-order partial differential equation (PDE) the Laplace equation, eq. (1.28), belongs to is called *elliptic*, and indeed it has an ellipsoidal, J_2-type lowest-order solution, eq. (1.21).

The general solution to associated Legendre functions is one of *solid spherical harmonics*; for the potential outside the mass distribution, U is given by the superposition of harmonic functions

$$U(r) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \left(\frac{1}{r}\right)^{\ell+1} \left(c'_{\ell m} \cos m\phi + d'_{\ell m} \sin m\phi\right) P_{\ell m}(\cos\theta),$$

where ℓ is degree, m is order, the $P_{\ell m}$ are associated Legendre functions, eq. (C.64) (Fig. 1.8), and c' and d' are spherical harmonic coefficients, where care needs to be taken as to different normalizations being in use (§C.6.2). Spherical harmonics are in general highly useful to characterize the spectral content of global features (e.g., §1.5, §C.6.2). The equivalent feature length, D, or half wavelength, λ, on the surface of the Earth (§C.6.2) is

$$D = \frac{\lambda}{2} = \frac{\pi R_E}{\sqrt{\ell(\ell+1)}} \approx \frac{\pi R_E}{\ell} \approx \frac{20,000 \text{ km}}{\ell}. \tag{1.29}$$

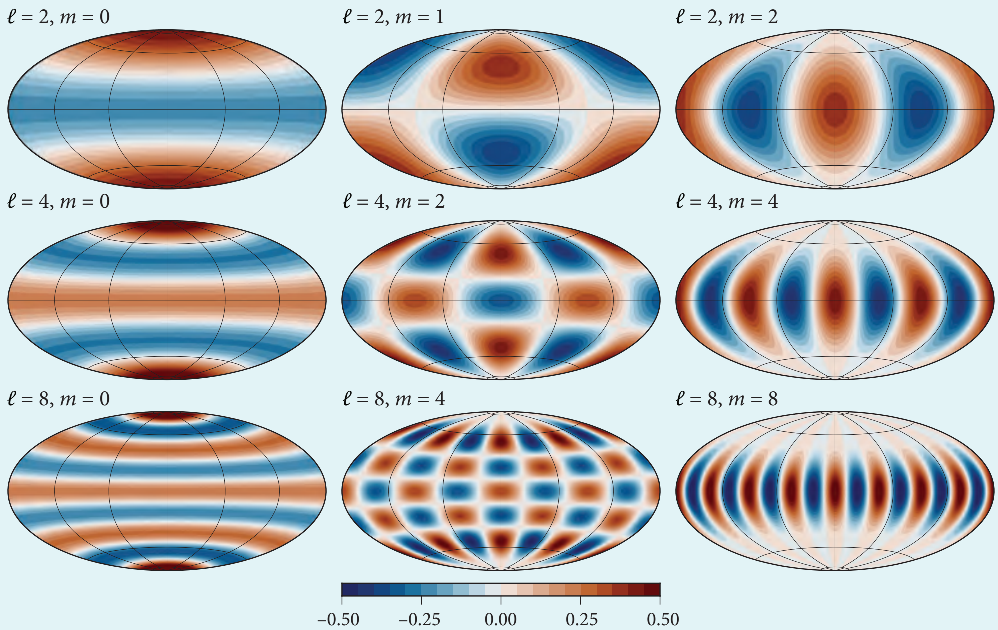


Figure 1.8: Examples of spherical harmonic basis functions (§C.6.2), with *degree*, ℓ, and *order*, m, showing the $\cos m\phi$–type of contributions, or the $\mathcal{Y}_{\ell\,-m}$ spherical harmonics, eq. (C.63), in a projection centered on longitude $\phi = \pi$. The $m = 0$–type functions are called *zonal*; there is no longitude dependence. Wavelengths of features $\lambda \sim 40{,}000$ km$/\ell$, eq. (1.29). Hammer equal area projection.

The actual geoid, e.g., as measured from satellite orbits for the Earth, is represented by a *band-limited* expansion with $\ell \leq L$ (cf. §C.6):

$$U(r) = -\frac{GM}{a} \sum_{\ell=0}^{L} \sum_{m=0}^{\ell} \left(\frac{a}{r}\right)^{\ell+1} \left(c_{\ell m} \cos m\phi + d_{\ell m} \sin m\phi\right) P_{\ell m}(\cos\theta). \tag{1.30}$$

The ellipsoidal U_g without rotation, eq. (1.17), is recovered by only considering $\ell = 0$ and $\ell = 2$ and $m = 0$ terms, for which $c_{20} = \frac{-J_2}{\sqrt{5}}$ for "fully normalized" harmonics (§C.6.2). At a certain location on the actual geoid, the difference between the value of the actual geoid when evaluated at the reference height, $U(r_{\text{ref}})$, and the reference geoid value, U_{ref}, is called a *potential anomaly*, and considering only verticals, is approximately

$$\Delta U = U(r_{\text{ref}}) - U_{\text{ref}}.$$

To first order, the difference in height away from the reference ellipsoid to where the actual U has the U_{ref} value is the *geoid undulation*, N, and *Bruns' formula* relates the two:

$$\Delta U = -g_{\text{ref}} N. \tag{1.31}$$

This allows inferring N from the reference gravity (eq. 1.25), and potential measurements, as well as computing the geoid anomalies of Fig. 1.7a from an expansion of the form eq. (1.30).

The gravity varies on the geoid, equipotential surface, and we can define *free-air* gravity anomalies from the gravity g at some point at height r on the actual geoid relative to the gravity at the projection of this point onto the reference geoid, $g_{\text{ref}}(r')$:

$$\Delta g(r) = g(r) - g_{\text{ref}}(r').$$

From eq. (1.10), gravity changes with radial distance to first order as

$$\frac{\partial g}{\partial r} = \frac{\partial}{\partial r}\frac{GM}{r^2} = -\frac{2GM}{r^3} \approx -\frac{2g}{r_{\text{ref}}}, \tag{1.32}$$

which means that we can correct $g_{\text{ref}}(r')$ to $g_{\text{ref}}(r)$ by adding $\approx -\frac{2g_{\text{ref}}(\theta)}{r_{\text{ref}} N(\boldsymbol{r})}$. Using eq. (1.31) and $g(\boldsymbol{r}) - g_{\text{ref}}(\boldsymbol{r}) \approx \frac{\partial \Delta U(\boldsymbol{r})}{\partial r}$, the *fundamental equation of geodesy* results in

$$\Delta g(r) = \frac{\partial \Delta U(\boldsymbol{r})}{\partial r} + \frac{2}{r_{\text{ref}}} \Delta U(\boldsymbol{r}), \tag{1.33}$$

where Δg are the measurements and the disturbing potential ΔU is the unknown, which has to be determined from eq. (1.28) since we are limited to measurements on the surface.

The most straightforward way of converting observed gravity values, g_{obs}, such as from a survey, collected at some height $h = r - r_{\text{ref}}$ above the reference geoid to some standard is thus by means of using the *free-air gravity anomaly*, Δg_{FA} (Fig. 1.7b). Close to the reference surface, we can write

$$\Delta g_{FA} = g_{\text{obs}} - g_\lambda + \frac{2g}{r} h = g_{\text{obs}} - g_\lambda + \gamma_{FA} h, \tag{1.34}$$

with $\gamma_{FA} \approx 0.3086$ mGal/m, where mGal is 10^{-5} m/s^2 in SI units. Therefore, Δg_{FA} corrects for the decrease of gravitational pull away from the reference height as if there were air in between r_{ref} and r.

Free-air gravity anomalies are also linked through eq. (1.33) if there is an expansion of ΔU given by some set of anomaly coefficients $\bar{c}$ and $\bar{d}$ akin to eq. (1.30), in which case the corresponding N expansion is given by using the same $\bar{c}$ and $\bar{d}$ but dividing by $-g_{\text{ref}}$ as per eq. (1.31), and the free-air anomalies by scaling $\bar{c}_{\ell m}$ and $\bar{d}_{\ell m}$ by $-\frac{(\ell-1)}{a}$ (Fig. A.4). The equivalent, local *gravity anomalies*, e.g., due to buried ore bodies or other lateral density variations, are tiny fractions of g_λ, and modern gravimeters used for surveying have relative μGal sensitivity, i.e., of order 10^{-9} of g.

The *Bouguer* gravity anomaly assumes that there is a laterally infinite slab of material with density ρ_B instead of air in between; with $\rho_B = 2670$ kg/m^3, an equivalent $\gamma_B = 0.1967$ mGal/m. We can compute the *surface gravity anomaly* dg of a buried density anomaly by integrating the incremental effect of all mass anomalies dM via eq. (1.9), $\mathrm{d}g = \frac{G}{r^2}\,\mathrm{d}m$. For example, a buried sphere of radius a with $\Delta\rho$ density anomaly yields

$$\Delta g = \frac{4\pi G a^3 \Delta\rho}{3r^2} \quad \text{or} \quad \Delta g_v = \frac{4\pi G a^3 \Delta\rho}{3} \frac{b}{\left(x^2+b^2\right)^{\frac{3}{2}}},$$

where Δg_v is the vertical gravity anomaly, x is the horizontal distance from the center of the sphere, and b is the burial depth to the center. Integration for complex geometries has to proceed numerically, and different-shaped anomalies may yield similar Δg expressions at the surface—a classic example of nonuniqueness of inversions, as expected from the solution of eq. (1.28) as required for eq. (1.33).

If a region is under isostasy, eq. (1.3) ensures that density anomalies down to compensation depth Z (Fig. 1.3) balance such that

$$\int_0^Z \Delta\rho(z)\,\mathrm{d}z = 0,$$

and this can be achieved in numerous ways. However, the geoid anomalies need not be zero since those depend on the *dipole moment of the density distribution*,

$$\Delta N = -\frac{2\pi G}{g}\int_0^Z z\,\Delta\rho(z)\,\mathrm{d}z = -\frac{3}{2\bar{\rho}R_E}\int_0^Z z\,\Delta\rho(z)\,\mathrm{d}z \tag{1.35}$$

(*Parsons & Richter*, 1980), where $\bar{\rho}$ is the mean density of Earth, and the integral arises in computations of gravitational potential energy (§6.6).

If a crustal column of density ρ_c, thickness l_c, and topography t is in Airy compensation floating in a mantle of ρ_a, then

$$\Delta N^A_{t>0} = \frac{\pi G}{g}\rho_c\left(2l_c t + \frac{\rho_a}{\rho_a-\rho_c}t^2\right) \quad \text{or} \quad \Delta N^A_{t<0} = \frac{\pi G}{g}(\rho_c-\rho_w)\left(2l_w t + \frac{\rho_a-\rho_w}{\rho_a-\rho_c}t^2\right) \tag{1.36}$$

(*Turcotte & Schubert*, 2014; sec. 5.13), where the second equation accounts for an ocean basin, with water, ρ_w, coverage of depth l_w. For Pratt isostasy at a constant crustal base, depth $Z = P$ and using eq. (1.7) is

$$\Delta N^P_{t>0} = \frac{\pi G}{g}\rho_0 Z l_c \quad \text{or} \quad \Delta N^P_{t<0} = \frac{\pi G}{g}(\rho_0-\rho_w)\,Z l_c. \tag{1.37}$$

Local topography variations of $\mathcal{O}$(1 km) of isostatic crust thus correspond to $\mathcal{O}$(10 m) geoid variations.

Geoid anomalies relate to ridge push via the density moment, eq. (1.35), associated with half-space cooling (exercise 7). Fig. 1.7a shows that spreading centers have, in general, no strong, excess positive geoid anomaly, meaning that spreading is passive in general, and not due to an active, hot mantle push component (§8.1.1). However, there are long-wavelength, large ΔN variations of the geoid; those are dynamically supported by the deep mantle (§9.4).

1.3 Internal structure, temperature, and composition of Earth

Remarkably, the USSR landed a number of probes on Venus during the Venera program starting in 1970. Those short-lived missions gathered photos and determined rock composition, which indicates basaltic-type rock cover in the lowlands. Our geophysical analysis of Venus so far is based on topography, geopotential, and inertial dynamics. However, given the more temperate climate, our enterprising aliens might be tempted to link appearances with internal qualities and deploy sensors on Earth.

Besides gravity, deep sounding estimates typically depend on electromagnetic or acoustic wave propagation, with the deepest and most detailed imaging achieved by listening to seismic waves using seismometers. Outside our planet, astronauts have deployed four such instruments on the Moon (*Toksöz et al.*, 1974; *Nakamura*, 1983), and a single sensor has recently greatly advanced our knowledge of the interior of Mars (*Lognonné et al.*, 2020; *Khan et al.*, 2021; *Beghein et al.*, 2022). What can our aliens say based on seismic wave propagation?

1.3.1 1-D structure of Earth as seen from seismology

Seismic waves are vibrations of the Earth where mechanical and kinematic energy gets transported away from a source, such as an earthquake or explosion, through an outwardly propagating wave front where material gets locally and transiently deformed. While more complicated in practice, we can think of this seismic wave by considering a raypath that is orthogonal to the wave front. That raypath will find the fastest path between source and receiver. If velocity increases with depth in the planet, the ray will thus not go straight from one side of the object to a recorder on the other side, but instead dive downward first and then be bent upward. This *refraction* is governed by *Snell's law* if velocities are layered, and differences in arrival time from a known source can be used to infer velocity with depth. A wave may also be *reflected* at a sharp contrast, such as a compositional layer, where the energy of transmission through an interface depends on the *impedance*, the product of wave speed and density. This provides constraints on the depth of the interface and the nature of the contrast.

Acoustic waves through air are of the volumetric (bulk) compressional type, but in an elastic solid, compressional, P, and shear, S, waves exist (§B.1). If we think of them as propagating through a planet as rays, we consider them to be two types of *body waves*. Their respective wave speeds correspond to different types of solutions of the wave equation (§B.1) and depend on different mechanical (elastic) properties and density as

$$v_P = \sqrt{\frac{K + \frac{4}{3}\mu}{\rho}} \quad \text{and} \quad v_S = \sqrt{\frac{\mu}{\rho}}, \tag{1.38}$$

where μ, K, and ρ are the shear modulus, incompressibility (§3.2), and density, respectively. Because $K, \mu \geq 0, \rightarrow v_P > v_S$, which means that the P wave arrives first, hence the labels of primary and secondary. The different wave types provide complementary constraints about the subsurface, e.g., since K and μ have different dependence on properties such as temperature and melt (§4.4.3). For wave propagation, v_P is like light speed for remote detection, in that no information can arrive earlier than as carried by the P wave. Earthquake rupture propagation typically happens at $\lesssim v_S$, but the related coseismic deformation can be transmitted remotely in near-instantaneous fashion by gravity signals traveling at actual light speed (*Vallée et al.*, 2017).

Based on their understanding of wave propagation, the analysis of refraction and reflection from a large amount of body wave records would then allow our aliens to infer a 1-D model of v_P and v_S for the Earth (e.g., AK135 in Fig. 1.9). For Earth, the interpretation of wave speeds in terms of solid and fluid structure has an interesting history (*Brush*, 1980). *Wiechert* (1897) proposed the existence of an iron core in a rocky shell, but the discovery of the core can be attributed to

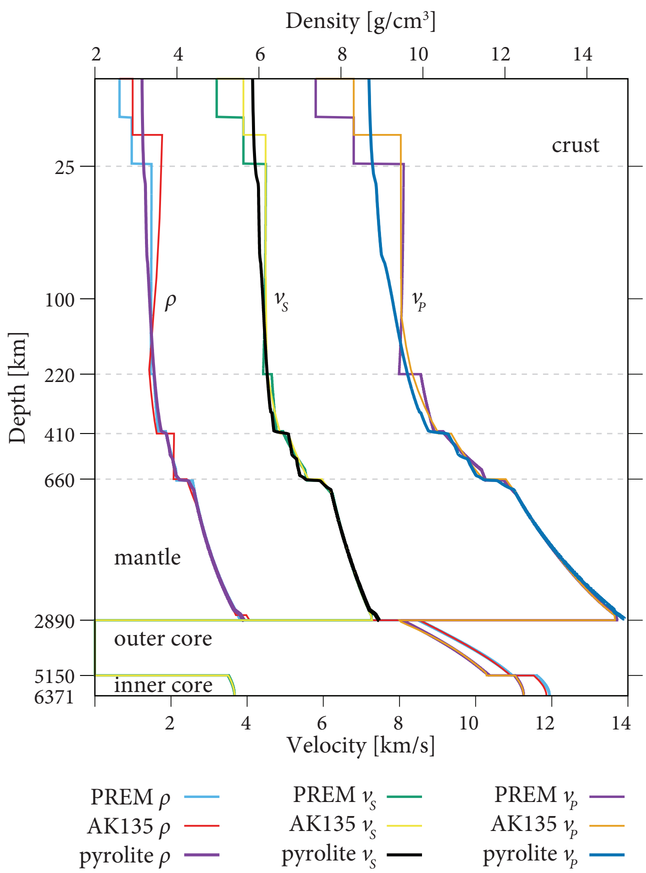


Figure 1.9: Layer averaged, 1-D Earth models showing density, ρ, as well as compressional, v_P, and shear wave, v_S, velocities from the long-period PREM (*Dziewoński & Anderson*, 1981) and shorter-period AK135 (*Kennett et al.*, 1995) seismic models, and for the mantle only from a mineral physics prediction for pyrolite (HeFESTo computation; *Stixrude & Lithgow-Bertelloni*, 2011, 2022) and the geotherm of Fig. 1.13. PREM layer depths for 25 km (nominal crustal thickness) and 220 km (not globally observed), as well as actual major phase transition depths of 410 km and 660 km (670 km in PREM) (Fig. 1.14) are indicated; note log-scale for depth.

Oldham (1906). The depth of the core-mantle boundary (CMB) was placed at 2900 km by *Gutenberg* (1913). The CMB is one example where the differences in K and μ dependence of wave speed, eq. (1.38), jump out: a fluid cannot sustain shear, and the absence of shear wave arrivals once bent raypaths reach a certain depth in the deep Earth lead to the discovery of the fluid outer core, where $\mu = 0 \rightarrow v_S = 0$ (*Oldham*, 1906; *Jeffreys*, 1926).

The inner core was discovered by *Lehmann* (1936; paper title: "P"), and the major, global layers as seen in Fig. 1.9 were established in the Earth model by *Bullen* (1947). This is not to say that we fully understand the internal structure and dynamics of the Earth even on large scales. For example, there are indications of an anisotropically distinct innermost inner core (*Ishii & Dziewoński*, 2002; *Pham & Tkalčić*, 2023), and such inner core heterogeneity remains to be fully linked to planetary evolution. The existence of the inner core is also not a static feature, but arose some time over the last ~2.5 Ga (*Buffett*, 2002; *Landeau et al.*, 2022), with revision of thermal conductivity estimates indicating a young, $\lesssim$700 Myr origin (*Labrosse*, 2015), perhaps consistent with anomalously low magnetic field strength at ~500 Ma (*Bono et al.*, 2019).

Properties in Fig. 1.9 are shown on a log scale for depth since the most interesting changes in velocities within the mantle happen in the upper ~700 km, which we focus on here. Figure 1.9 also shows the *preliminary reference Earth model* (PREM; *Dziewoński & Anderson*, 1981), which is mainly based on *normal modes*. Modes are global planetary oscillations that are excited by large earthquakes and can be described with spherical harmonics, like quantum mechanical orbitals, as the vibrations at various natural frequencies, *eigenfunctions*, combining to a range of possible oscillations. The surface displacements of spheroidal oscillations look like the examples of Fig. 1.8, but there are also toroidal types of oscillations, associated with twisting of the Earth.

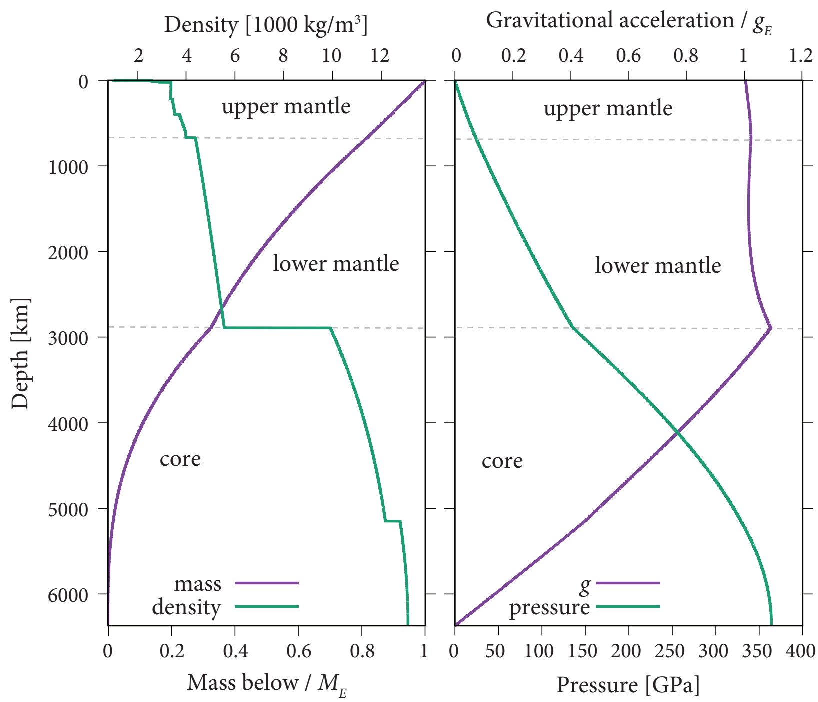


Figure 1.10: Left: Density ρ (mean density ρ_E) and mass, in terms of total, M_E, below a given depth, as a function of depth. Density jumps at phase transitions are important for mantle convection, and likely overestimated in the case of the 670 km for PREM (cf. §12.12; *Shearer & Flanegan*, 1999). Right: Gravitational acceleration relative to the surface, $g_E = 9.807$ m/s^2, and pressure. All based on integration of $\rho(z)$ from PREM (*Dziewoński & Anderson*, 1981) as in Fig. 1.9, cf. eqs. (1.8) and (1.39).

Higher-frequency (larger ℓ) normal modes localize deformation at the planet's surface, and their superpositions can also be identified as *surface waves*.

As the name implies, those waves require a free surface to vibrate. The displacement types of surface waves can be further divided into waves that oscillate in the horizontal, orthogonal to the propagation path (*Love waves*). Those waves require sphericity or a velocity increase with depth to exist, and consist of horizontally polarized shear, *SH*, waves and surface underside reflections of those (multiples). *Rayleigh waves* oscillate in the vertical and are a combination of vertically polarized shear, *SV*, and compressional, *P*, waves and can exist on any free surface, including a half-space. Love waves propagate faster than Rayleigh waves in the uppermost mantle, i.e., $v_{SH} > v_{SV}$, an example of the polarization dependence of wave speeds, or *seismic anisotropy* (§9.6).

1.3.2 Pressure and mass within the Earth

Since wave speeds depend on elastic moduli and density, eq. (1.38), the 1-D models of Fig. 1.9 provide further constraints for our aliens' assessments of Earth's internal mass distribution (Expanded details 3) and thus pressure conditions (Fig. 1.10). Whatever their density model, when ρ is integrated from the core up to radius r to obtain the mass underneath,

$$M(r) = 4\pi \int_0^r \rho(r) r^2 \, dr \quad \text{and} \quad M_E = M(R_E) = \frac{4\pi}{3} R_E^3 \rho_E, \tag{1.39}$$

the total mass of the Earth, M_E, needs to match the orbital constraints (§1.2). The mean density of Earth, $\rho_E = \langle \rho \rangle$, for an equivalent sphere is $\rho_E \approx 5513$ kg/m^3. Plotting against depth, $M(r)$, yields the first curve in Fig. 1.10, which tells the aliens that only $\sim \frac{1}{3}$ of our planet's total mass, $M_E \approx 5.972 \cdot 10^{24}$ kg, is within the core, even though its iron-dominated composition has roughly three times the density of mantle rocks, which is $\langle \rho_m \rangle \approx 4454$ kg/m^3. The mantle and crust contribute $M_m \approx 4 \cdot 10^{24}$ kg (67%) and $M_c \approx 2.8 \cdot 10^{22}$ kg (0.5%), respectively. These numbers matter for the assessment of composition and origin of the planet.

The gravitational acceleration due to the mass within an object is given by eq. (1.8). Figure 1.10 shows that $g(r)$ in the core roughly displays the linear increase that would be expected from $M \propto r^3$. However, the reduced densities of the mantle lead to a $\sim$ constant g within the mantle, with a volume-averaged mean value of $\langle g_m \rangle \approx 10.007$ m/s^2, close to the surface value. The aliens can then also use eq. (1.3) to find the pressure conditions throughout the Earth (Fig. 1.10). For the upper mantle, p increases by ~ 3.1 GPa per 100 km; depths z in km can be converted to p in GPa and vice versa as

$$p \approx -0.04809 + 0.03077z + 7.30972 \cdot 10^{-6} z^2 \quad \text{or} \quad z \approx 2.98129 + 31.698p - 0.15771p^2,$$

which are coarse approximations (e.g., at the surface, p should be zero).

1.3.3 Complexities in the 1-D model

Considering shallow seismic wave speeds (Fig. 1.9), our aliens notice a sharp jump at ~ 25 km. This is the assumed depth of a global crust-mantle interface for PREM, the *Moho*, first inferred by *Mohorovičić* (1910) based on P wave refraction from a single earthquake. The Moho marks a compositional change, from the basaltic- and granitic-type crust of oceanic and continental lithosphere, respectively, to the peridotitic, i.e., olivine- and pyroxene-dominated, mantle. PREM also divides the crust between an upper and lower part, whereas AK135 has only a single crustal layer of 18 km thickness. For the lithosphere, such global averages mean little, as discussed next.

1.3.3.1 The lithosphere and its topography

Using detailed regional seismological constraints, one finds that oceanic crust is of roughly constant thickness of ~ 7 km (Fig. 1.11b). This indicates that the fractionation process at midoceanic ridges is sampling an underlying upper mantle of fairly constant temperature overall (§10.3.4.2). There are, however, important local deviations, e.g., due to volcanic pulses leading to oceanic plateaus, as well as regional or temporal fluctuations of upper mantle temperature.

Within the continents, crustal thickness can be highly variable between $\sim 10 \ldots 75$ km (Fig. 1.11b). Comparison of long-wavelength topography and crustal thickness estimates from seismology (Fig. 1.11a, b) leads our observers to expect that Airy isostasy does in fact broadly hold in continents. Particularly in continents, the depth of the lithosphere (or “plates”) is hard to define (Fig. 1.11c) since, to first order, the lithosphere is a mechanical layer. However, in oceanic lithosphere, the thermal, half-space cooling model can explain much of the surface topography through Pratt-type density variations (§7.3.2). A process difference between oceanic and continental lithosphere is that, once continental lithosphere is formed, it is then mainly floating at the surface because of compositional effects. This is partially because of its thick crust and partially because of melt depletion of its mantle-lithospheric part, in particular for old, cratonic regions (§10.4). Some continental material gets recycled into the mantle, but continental lithosphere is typically not subducted wholesale.

Partly as a consequence of such compositional heterogeneity, upon closer inspection, isostatic balance is often not exactly true. Figure 1.12 shows a global scatter plot for crustal thickness vs. *rock-equivalent topography*, z_{RET}. The latter removes the effect of water layers covering negative topography following eq. (1.6):

$$z_{\mathrm{RET}} = z_{\mathrm{bed}} + \frac{\rho_w}{\rho_c} \Delta W, \tag{1.40}$$

where z_{bed} is the bedrock elevation and ΔW the water layer thickness; it is the equivalent topography if water is replaced by an average crust.

There is an overall, broadly linear, isostatic trend (Fig. 1.12), but there is also large scatter and potentially different tectonic domains with different scalings (*Gvirtzman et al.*, 2016; *Ingalls et al.*, 2016). In particular, typical continental elevations of ~ 500 m correspond to a wide range of

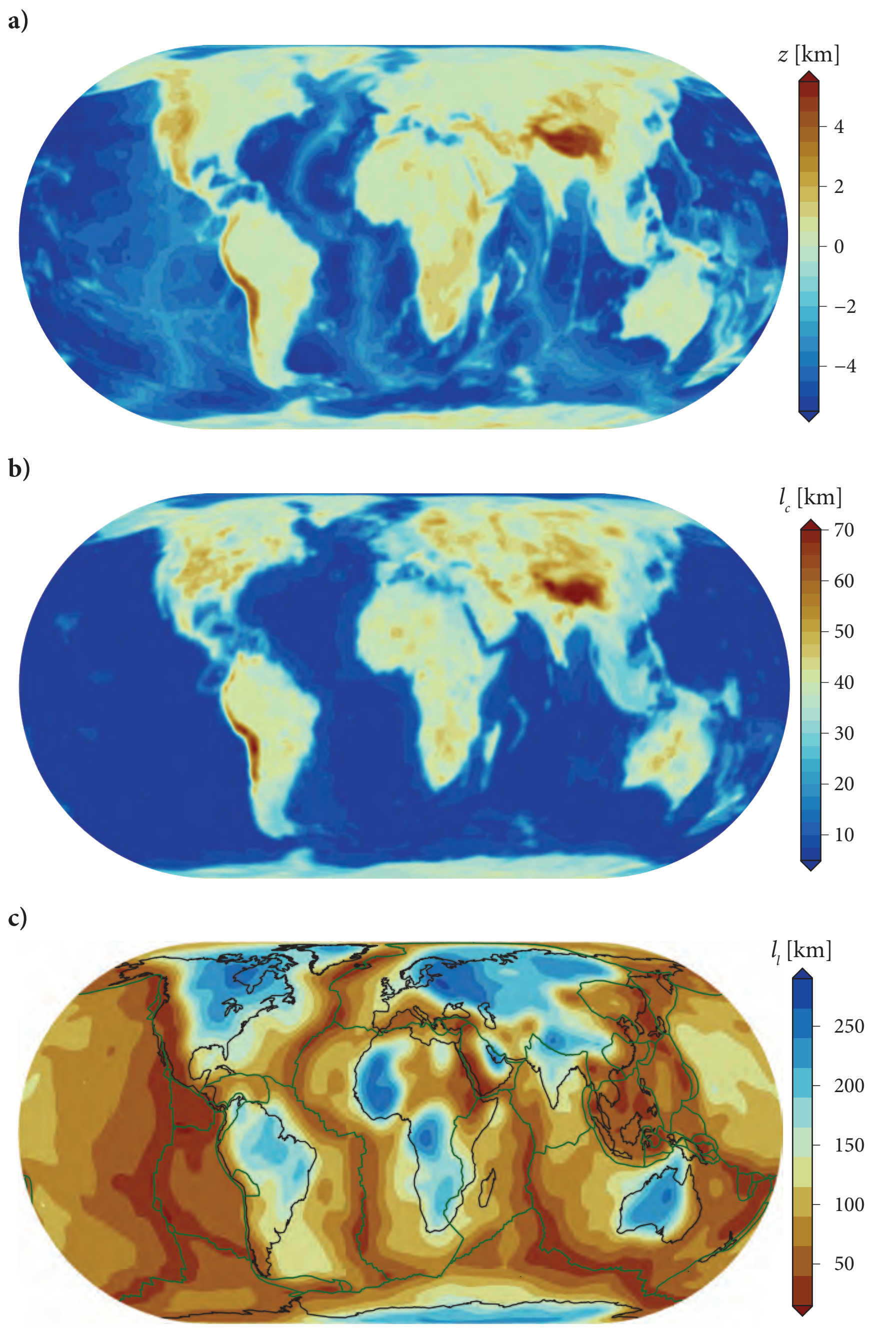


Figure 1.11: **a)** Bedrock-equivalent topography (*Hirt & Rexer*, 2015), eq. (1.40), up to degree $\ell = 200$, and smoothed by a $\cos^2$ filter. **b)** Crustal thickness from an update of the CRUST1 model (*Laske et al.*, 2013) with regional studies (from *Faccenna & Becker*, 2020), filtered as for b). **c)** Lithospheric thickness model based on seismic tomography and the method of *Steinberger & Becker* (2018), their mean model with an attempt to remove subduction zone effects (see *Steinberger & Becker*, 2018). Note that lithospheric thickness estimates are highly uncertain, in particular close to orogens.

crustal thicknesses, which motivates the exploration of the sources of nonisostatic elevation. We can compute the deviation between actual topography and an isostatic model topography, $z_{\rm iso}$; this difference here defines *residual topography*. As per eq. (1.6), a change in crustal and lithospheric thickness, respectively, is expected to lead to isostatic elevation changes according to

$$\Delta z_c = f_1 \Delta l_c \quad \text{or} \quad \Delta z_l = f_2 \Delta l_l, \tag{1.41}$$

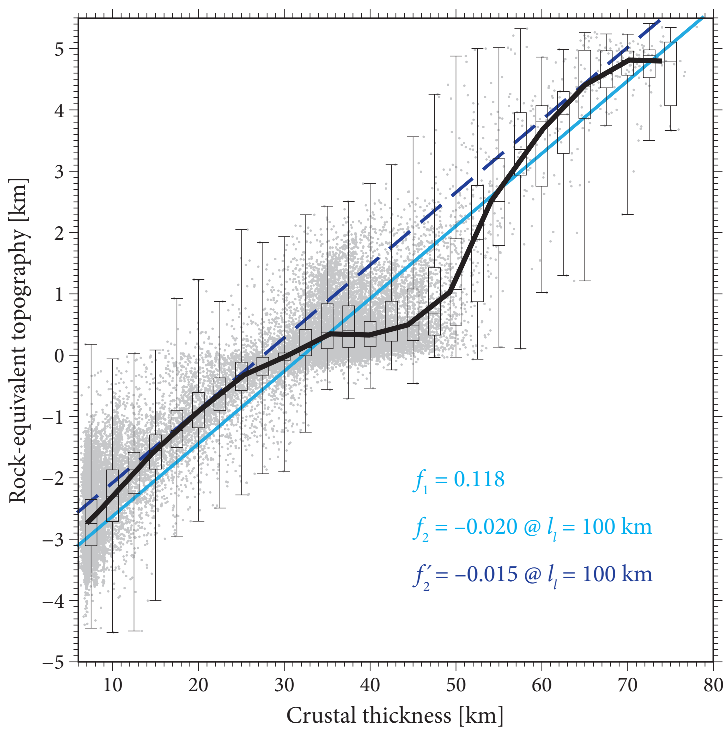


Figure 1.12: Relationship between crustal thickness (Fig. 1.11c) and rock-equivalent topography (eq. 1.40, Fig. 1.11a). Bar whisker plots (min and max as error bars, 25% and 75% quartiles as boxes) and black line are for the median at 2.5 and 5 km thickness bin width, respectively, and the blue line is a linear best fit. The $f_{1,2}$ values, eq. (1.6), are for an equivalent $H' = -1.8$ km and $l_l = 100$ km ($l_l = 97.5$ km is the mean of Fig. 1.11d). Dark blue dashed line corresponds to $f_2' = -0.015$, which better matches the roughly linear regions at $l_c \sim 20$ and ~ 60 km. Compare to the synthetic columns and $f_{1,2}$ values of oceanic and continental domains in Fig. 1.4.

and the best-fit values of Fig. 1.12 are $f_1 \approx 0.12 \ldots 0.17$ and $f_2 \approx -0.02$ and correspond to plausible, globally averaged density values (Fig. 1.4).

1.3.3.2 The asthenosphere

Below the Moho, the PREM model has another discontinuity at 220 km (Fig. 1.9): the *Lehmann discontinuity*. This depth and property jump is mainly a somewhat arbitrary choice of this particular model, and the velocity jump is not globally systematic. However, 1-D seismic velocities show a dip around $\sim 100 \ldots 300$ km depth, particularly underneath oceanic plates, and regionally, there are sharp velocity contrasts at similar depths. Those may in some way (e.g., thermally, by means of partial melt variations, or mechanically) be associated with the transition between the lithosphere and the asthenosphere, sometimes referred to as the *lithosphere-asthenosphere boundary* (LAB; *Fischer et al.*, 2010).

As mechanical terms, the L and the A are associated with relatively strong and weak parts of the mantle, deforming within mantle convection. Our aliens might thus study fluid dynamics and the deformation behavior of rocks in their laboratory by subjecting them to different deformation rates and varying pressures and temperatures. Depending on their patience, they might have to extrapolate their results over vast orders of magnitude since Earth's deformation is quite slow on experimental timescales. However, it is thought that the strength contrast between the asthenosphere and lithosphere is mainly achieved by temperature-dependent creep behavior of peridotitic mantle material (§5.1.6 and §7.3.2), with some contributions from composition and volatile variations. We can then associate the tectonic plates with the lithosphere, and the only compositionally distinct layer in Fig. 1.3 would be the crust. Figure 1.11c shows one seismology-based estimate of lithospheric thickness, based on the link between temperature and seismic velocity. However, there are many uncertainties, and the thickness estimate depends on which constraints are used to define the LAB (*Rychert et al.*, 2020).

The mantle deforms by means of solid-state, slowly creeping flow, and a fluid's resistance to shear is called the *viscosity*. There is a moderate increase of viscosity with depth within the mantle, but most of the interior is low viscosity compared to the strong lithospheric plates on top, by several orders of magnitude (§9). We could thus associate the asthenosphere with all of the mantle below the lithosphere. However, the term more commonly refers to the depths of ~ 100 to ~ 300 km, where seismic velocity actually decreases slightly with depth (see v_S in Fig. 1.9 at ~ 200 km), the *low-velocity zone* of the upper mantle.

Most of seismic wave propagation happens elastically with tiny, reversible, elastic deformation of the rock through which the wave passes. However, dissipative processes lead to a reduction of wave amplitudes with distance besides the purely geometric spreading of energy. Such damping, or *attenuation*, of wave motion is frequency dependent (it increases with shorter periods) and also depends on depth and material properties (§4.4). Attenuation estimates thus provide additional constraints on the subsurface state (e.g., temperature, fluids, and melting), but this requires interpreting amplitudes rather than travel times, which is more complex as an inverse problem. That said, asthenospheric depths are found to be highly attenuating. The origin of this and the decrease of velocities with depth in the asthenosphere remains debated; it is probably not purely the effects of temperature and pressure causing this change, but small fractions of partial melting and anelastic processes close to the melting temperature likely contribute (§10.3.4; *Stixrude & Lithgow-Bertelloni*, 2005; *Takei*, 2017).

1.3.4 Thermal background state of the mantle

Our aliens thus cannot understand seismic structure without also figuring out internal temperatures. Below the low-velocity zone from ~ 200 km to the core-mantle boundary, we are seeing a monotonous increase in both v_P and v_S. Why is not obvious from eq. (1.38). Pressure increases with depth because of the overburden of overlying rock layers (Fig. 1.10); this squeezes mass into a tighter volume. Temperature also increases because of this pressure effect by means of *adiabatic compression* (Expanded details 3), but as Fig. 1.9 shows, the effect of pressure overwhelms that of temperature, and density does increase with depth in the mantle. Since velocities increase rather than decrease, this means that the increase of the elastic parameters with depth due to pressure is even greater.

An *adiabatic process*, or *isentropic process*, is one where there is no exchange of heat with the surrounding medium. The *mantle adiabat* describes the background increase of temperature within the mantle with depth only due to the effects of pressure. Using thermodynamics, one can show that this gradient can be linked to seismological properties; it works out to be ~ 0.3°C/km (Expanded details 3). This approach allows our aliens to add temperature to their Earth model (Fig. 1.13). The adiabatic gradient alone would predict an increase of temperature throughout the mantle of ~ 1000 K.

Dynamically, the 1-D average temperature within the mantle, the *geotherm*, is also expected to reflect the *thermal boundary layer*s (TBLs) that arise due to mantle convection and show much stronger temperature gradients, ≈ 13.5°C/km for the top ~ 100 km (§7.4.3). In fact, lithospheric plates and their motions are part of the top, cold thermal boundary dynamics of convection. Assuming an asthenospheric temperature of 1350°C below the plates (Table 1.1), using this as the TBL temperature difference, and assuming symmetry between the top and bottom layers adds ~ 2700° C to the adiabatic increase, so the aliens expect a core-mantle boundary temperature of ~ 3700° C. Table 1.1 lists some additional constraints on the mantle geotherm, which overall are consistent with the high end of our convective plus adiabatic estimates. For the mantle, the symmetry between top and bottom TBL that purely bottom-heated, isoviscous convection would produce is an overestimate, and we expect the lower TBL to have a smaller temperature contrast because of significant internal heating (§8.1.1 and §10.1.4).

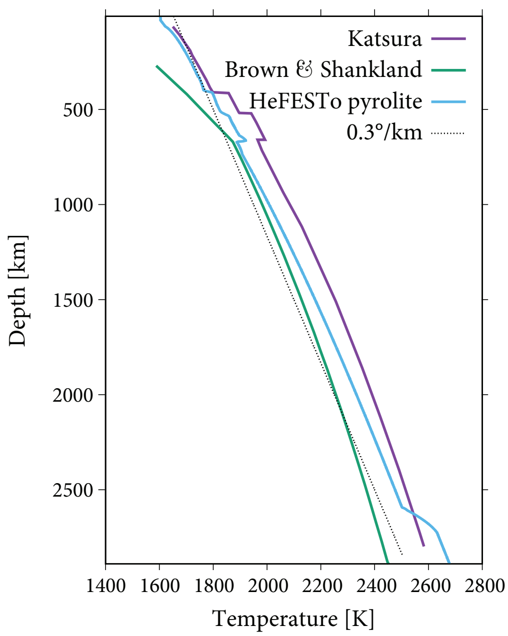


Figure 1.13: Adiabatic (i.e., constant entropy, no heat exchange) geotherm, accounting for latent heat due to phase transitions (Fig. 1.14), eq. (1.56), from *Katsura* (2022) for dry pyrolite, compared to seismology-based estimates from *Brown & Shankland* (1981), and pyrolite estimate as used for Fig. 1.9 (*Stixrude & Lithgow-Bertelloni*, 2011, 2022). For the latter, potential temperature, i.e., the intersection of the adiabat with the surface, is chosen as 1377°C. Also shown is a nominal adiabatic gradient, eq. (1.47), of 0.3°/km.

Boundary/region	Depth [km]	T [°C]	Reference
asthenospheric (potential) temperature			
half-space cooling (HSC)	~ 5	1365±10	*Carlson & Johnson* (1994)
petrological	50	1314±36(12)	*Brown Krein et al.* (2021)
petrological, v_S, HSC	~ 100	1383±40	*Dalton et al.* (2014)
phase transition anchor points (Fig. 1.14)			
olivine-wadsleyite	410	1490±45	*Katsura et al.* (2004)
post-spinel	660	1600±50	*Katsura et al.* (2004)
post-perovskite	2884	3730±200	*Hernlund et al.* (2005)

Table 1.1: Temperature constraints (modified from *Jaupart et al.*, 2007) with median ± one standard deviation recomputed for *Dalton et al.* and *Brown Krein et al.* for the same ridge segments (in parentheses: only considering the median cluster for *Brown Krein et al.*); cf. Fig. 1.13. For half-space cooling, see §7.3.2, eq. (7.53).

EXPANDED DETAILS 3: Thermodynamics of the adiabatic mantle temperature gradient

The *Maxwell equations* of thermodynamics can be written for the change of the *enthalpy*, H, type of energy, defined as internal energy, U, plus the product of pressure, p, and volume, V, $H = U + pV$. We can write the change in temperature, T (always in units of K for thermodynamics) due to change in pressure at constant *entropy*, S, as

$$\left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p. \tag{1.42}$$

With mass $m = \rho V$, the total heat produced, ΔQ, relates to the change in temperature and entropy as

$$\Delta Q = mc_p \Delta T = V\rho c_p \Delta T = T\Delta S \tag{1.43}$$

for a reversible process, with *heat capacity* at constant pressure, c_p, which is defined through the limit case of the amount of heat ΔQ needed to raise temperature by ΔT at fixed pressure:

$$c_p = \lim_{\Delta T \to 0} \left(\frac{\Delta Q}{\Delta T}\right)_p, \tag{1.44}$$

where $c_p \sim 1000$ J/K/kg. The change in volume relates to change in temperature as

$$\Delta V = \alpha V \Delta T,$$

with *thermal expansivity* α, which links temperature and volume, or density, variations at constant pressure as

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p = -\frac{1}{\rho}\left(\frac{\partial \rho}{\partial T}\right)_p; \quad \mathrm{d}\rho = -\rho\alpha\,\mathrm{d}T \quad \text{or} \quad \mathrm{d}\ln\rho = -\alpha\,\mathrm{d}T. \tag{1.45}$$

A typical value for the mantle is $\alpha_m \approx 2 \cdot 10^{-5}\ \mathrm{K}^{-1}$ (§4.4.3).

Plugging in those two relationships, we can rewrite eq. (1.42) for the increase of T with p as

$$\left(\frac{\partial T}{\partial p}\right)_{\mathrm{ad}} = \frac{\alpha T}{\rho c_p}. \tag{1.46}$$

For a hydrostatic fluid, $\mathrm{d}p = \rho g\,\mathrm{d}z$, such that the *adiabatic gradient* of temperature with depth is

$$\boxed{\left(\frac{\partial T}{\partial z}\right)_{\mathrm{ad}} = \frac{\alpha g T}{c_p}}. \tag{1.47}$$

This change in temperature is purely due to compressibility, and gradients are of order 0.3 K/km in much of the mantle, with values of ~ 0.6 K/km in the shallowest depths (Fig. 1.13).

Assuming homogeneity, we can integrate eq. (1.47) from the surface, *potential temperature*, $T_P(0)$, to a temperature at depth, $T(z)$, and simplify (§C.1.2, §C.2.4):

$$\int_{T_P}^{T} \frac{\mathrm{d}T'}{T'} = \frac{\alpha g}{c_p}\int_0^z \mathrm{d}z' \to \ln T - \ln T_P = \frac{\alpha g}{c_p} z \to \frac{T}{T_P} = \exp\left(\frac{g\alpha}{c_p} z\right). \tag{1.48}$$

Eq. (1.48) can be viewed as a way to correct some upper mantle temperature at depth, T, to its surface value, T_p, due to an adiabatic upwelling, without exchange of heat, i.e., constant entropy. Figure 1.13 shows that the actual adiabatic temperature profile does not follow this exponential shape because material parameters are depth dependent and phase transitions add complexity.

Eq. (1.47) is a combination of thermodynamic parameters and g, but we can further link the gradient to seismic parameters to build a planetary model. The *Grüneisen parameter*, γ, relates changes in volume to the properties of the crystal lattice and can be written as

$$\gamma = \frac{\alpha K_S}{\rho c_p} = \frac{K_S}{K_S^c} \quad \text{with} \quad K_S^c = \frac{\rho c_p}{\alpha}, \tag{1.49}$$

where K_S is the bulk modulus, or *incompressibility* (cf. eq. 3.6), denoted by subscript S for constant entropy:

$$K_S = \rho \left(\frac{\partial p}{\partial \rho}\right)_S = -V \left(\frac{\partial p}{\partial V}\right)_S . \tag{1.50}$$

The γ parameter is thus a nondimensional incompressibility with characteristic scale K_S^c, and $\gamma \approx 1 \ldots 1.1$ throughout the mantle. One can also define an *Anderson-Grüneisen parameter*, γ_T, based on the change of expansivity as a function of volume change,

$$\frac{\alpha}{\alpha_0} = \left(\frac{V}{V_0}\right)^{\gamma_T}, \tag{1.51}$$

where γ_T has been suggested to be ~ 5 (e.g., discussion in *Katsura et al.*, 2004).

Based on γ of eq. (1.49), we can then rewrite eqs. (1.46) or (1.47) as

$$\left(\frac{\partial T}{\partial p}\right)_{\text{ad}} = \frac{\gamma T}{K_S} \quad \text{or} \quad \left(\frac{\partial T}{\partial z}\right)_{\text{ad}} = \frac{\gamma g T}{\phi}, \tag{1.52}$$

where ϕ is the *seismic parameter*:

$$\phi = v_B^2 = \frac{K_S}{\rho} = v_P^2 - \frac{4}{3} v_S^2 \tag{1.53}$$

with the *bulk sound velocity*, v_B, cf. eq. (1.38). These relationships provide the desired link between seismic constraints (Fig. 1.9) and thermal properties (§4.4.3), and can be used to integrate temperature down the adiabat. The Grüneisen parameter is often assumed to be $\sim$ constant within the mantle (Fig. 1.13), and also arises in *Birch*'s (1952) *Adams-Williamson equation*, which, using eqs. (1.49) and (1.53), states that

$$\frac{\mathrm{d}\rho}{\mathrm{d}z} = -\frac{\rho g}{\phi} \quad \text{or} \quad \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}z} = -\frac{\alpha g}{c_p \gamma}, \tag{1.54}$$

where the LHS can be integrated to yield an estimate of ρ as a function of depth, given ρ_0 and depth-dependent g and the seismically constrained ϕ:

$$\frac{1}{\rho}\,\mathrm{d}\rho = -\frac{g(r)}{\phi(r)}\,\mathrm{d}r \quad \rightarrow \quad \log\left(\frac{\rho}{\rho_0}\right)(r) = -\int_{r_0}^{r} \frac{g(r)}{\phi(r)}\,\mathrm{d}r.$$

If we plug the definition of K_s into eq. (1.52), we can also relate temperature and density change as

$$\frac{\mathrm{d}T}{T} = \gamma \frac{\mathrm{d}\rho}{\rho}$$

or, after integration of both sides,

$$T = T_0 \left(\frac{\rho}{\rho_0}\right)^{\gamma},$$

where the subscript zero refers to the values at zero pressure.

1.3.5 Composition and mantle phase transitions

Given their knowledge of thermodynamics and seismic constraints, our aliens have thus assembled a density, pressure, and temperature model of our planet. What is Earth made out of, though, i.e., which compositions fit the isostatic topography variations, as well as the bulk mass and the depth dependence as seen in their reference model?

They could drill, but that is tough (since T and p increase significantly with depth), and none of our deep drilling efforts has penetrated the continental Moho yet. They could also conduct experiments using different compositions, but what speeds up the search is the consideration of comparable objects. For Earth, we consider two types of meteorites as representative: in very broad strokes, iron-type meteorites for the core, which is made out of iron and some lighter element (*Li & Fei*, 2005; *Hirose et al.*, 2013; sulfur, carbon, oxygen, and hydrogen have been suggested) and silicic-type meteorites (carbonaceous chondrites) to define the *bulk silicate Earth* (BSE; *McDonough & Sun*, 1995; *Palme & O'Neill*, 2005; §10.1.7).

Overall, mantle composition is like a peridotite rock. From the BSE, convection has extracted the oceanic and continental crust (§10.1.7), leaving a relatively depleted upper mantle component sampled by spreading centers, a midocean-ridge basalt (MORB) reservoir. From dredged samples, one can estimate the *depleted midocean-ridge mantle* (DMM; *Workman & Hart*, 2005), and estimate a hypothetical pyrolite material to match DMM, e.g., by mixing a fraction f_b of basalt produced with the $(1 - f_p)$ fraction of depleted harzburgite residuum, $f_b \approx 0.2$ (*Xu et al.*, 2008). Such f_b values are broadly consistent with seafloor creation rates and operation of plate tectonics over $\gtrsim$3 Gyr (*Phipps-Morgan & Morgan*, 1999).

Figure 1.9 compares thermodynamic estimates for pyrolite with the seismic models; there is a good fit overall, within uncertainties (*Cammarano et al.*, 2005). However, turned around, seismic models do not provide sufficient constraints to resolve some of the more subtle questions, e.g., the degree of compositional variations from a mean bulk composition (*Mattern et al.*, 2005; *Stixrude & Lithgow-Bertelloni*, 2012). There is also significant uncertainty about BSE composition itself. All of this means that we have a good overall understanding of the rocky mantle composition and what happens when fractionation modifies it. Yet, questions such as the degree of internal heating in the lower mantle and the degree to which compositional anomalies matter for midmantle convection remain open.

With these caveats in mind, we can return to the sharp increases in velocities and density seen in the transition zone (Fig. 1.9), typically defined to be below the asthenosphere, $\gtrsim$ 300 km and $\lesssim$ 700 km. In between, PREM identifies jumps at 410 km and 670 km, where the lower boundary is closer to 660 km according to more modern estimates. Those jumps are, to first order, not related to any compositional change, but rather a transition of the same pyrolitic material to closer packing due to *phase transitions* (Fig. 1.14). The pressure at which a phase transition occurs (and hence the depth) depends on the temperature via the *Clapeyron slope*,

$$\Gamma = \frac{\mathrm{d}p}{\mathrm{d}T}_{\text{transition}}, \tag{1.55}$$

(continued...)

Index