Supporting Information for

GRAIL gravity constraints on the vertical and lateral density structure of the lunar crust

Jonathan Besserer*, Francis Nimmo, Mark A. Wieczorek, Renee C. Weber, Walter S. Kiefer, Patrick J. McGovern, Jeffrey C. Andrews-Hanna, David E. Smith, and Maria T. Zuber

Geophysical Research Letters

* To whom correspondence should be addressed; E-mail: jbessere@ucsc.edu

This document contains:

- Supporting Information text
- Tables S1 to S3
- Figures S1 to S11
- Supplementary references

Contents

S1 Modeling approach

Consider a vertically layered crust, with a surface topography $h(x)$ and a density distribution $\rho(z)$, with x an arbitrary horizontal coordinate and z the depth below the surface (see Fig. S1).

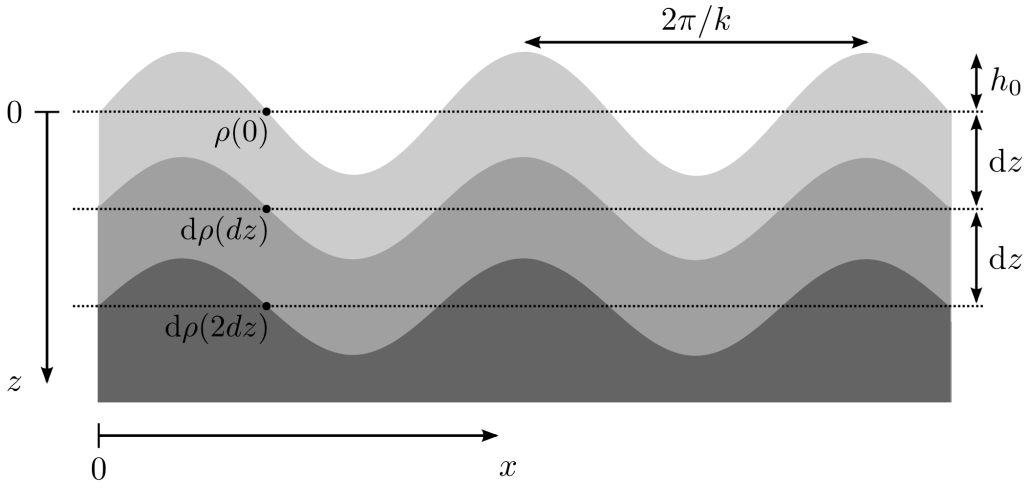

Figure S1: Definition sketch for the theoretical admittance calculations. The crust is decomposed into a series of (thin) constant density layers. Each one is characterized by a density contrast $\mathrm{d}\rho = \frac{\partial \rho}{\partial z}\,\mathrm{d}z$ with the overlying (upper) layer, whereas the first layer's density contrast corresponds to its own density. Here, only the first three layer interfaces are schematically pictured, for a given wavelength $2\pi/k$. $\mathrm{d}z$ is the infinitesimal distance between consecutive layers and h_0 is the topographic amplitude.

We will use a Cartesian geometry for simplicity, and will later argue that this induces a negligible error in comparison to a more rigorous approach in spherical geometry. The resulting free-air gravity anomaly measured at $z = 0$ is given by the "mass-sheet" approximation [e.g., *Turcotte and Schubert*, 2002] and by summing the contributions of each layer of topography $h(x) = h_0 \sin(kx)$ for a given wavenumber k:

$$\Delta g(x) = 2\pi G h_0 \left[\rho_{\mathrm{surf}} + \int_{0^+}^{+\infty} \frac{\partial \rho}{\partial z}\, e^{-kz}\, \mathrm{d}z\right] \sin(kx), \tag{S1}$$

where $\rho_{\text{surf}} = \rho(0)$ is the surface density. The exponential term is an attenuation factor and comes from the fact that the gravitational potential associated with each layer's mass anomaly at depth must satisfy Laplace's equation. Note that each layer is assumed to have the same amplitude (h_0).

The free-air admittance Z is given by the ratio of the Fourier components (here sans-serif characters) of the gravity anomaly and topography [e.g., *Audet*, 2014, and references in it]:

$$Z(k) = \frac{\Delta \mathsf{g}(k)}{\mathsf{h}(k)}. \tag{S2}$$

If ρ is independent of depth, then $Z(k) = 2\pi G\rho$. By analogy, when density varies with depth, we can define the effective density spectrum:

$$\rho_{\text{eff}}(k) = \frac{Z(k)}{2\pi G}. \tag{S3}$$

Combined with Eq. (S1), this yields

$$\rho_{\text{eff}}(k) = \rho_{\text{surf}} + \int_{0^+}^{+\infty} \frac{\partial \rho}{\partial z}\, e^{-kz}\, \mathrm{d}z. \tag{S4}$$

Finally, by converting the wavenumber to its spherical harmonic equivalent,

$$k(\ell) = \frac{\sqrt{\ell(\ell+1)}}{R}, \tag{S5}$$

where ℓ is spherical harmonic degree and R is the radius of the reference surface, the effective density is given by:

$$\rho_{\text{eff}}(\ell) = \rho_{\text{surf}} + \int_{0^+}^{+\infty} \frac{\partial \rho}{\partial z}\, e^{-\frac{\sqrt{\ell(\ell+1)}}{R} z}\, \mathrm{d}z. \tag{S6}$$

In practice, the integral can be performed over a restricted depth range (e.g., crustal thickness, z_c). Though the density anomalies (e.g., porosity) could extend down to the upper mantle, we will consider only high degrees, i.e. shallow sampled depths in this study, as discussed below. We do not consider lower degrees in this study, because of the influence of complicating effects such a flexure. The effect of the crust-mantle discontinuity itself will be discussed in Sec. S3.

Note that we are here neglecting finite-amplitude topography effects. The justification for doing so is that these effects are automatically accounted for in an approximate way by the way in which the observed effective density is obtained (Sec. S4). We also verified our approach by carrying out tests using synthetic gravity data and comparing the synthetic effective density spectra obtained with our analytical spectra (see below).

The assumed in-phase nature of the density layers (see above) is reasonable for situations in which the main source of near-surface topography is impacts, or where gravitational and/or thermal compaction of the pore space has occurred, though it might be questionable for lava-filled regions (e.g., mare). However, the simple linear model below, derived from Eq. (S4), is able to detect mare regions, so this assumption does not prevent simple mare *detection*. The in-phase assumption at short wavelengths is also strongly supported by the near-perfect (i.e. close to $+1$) correlation between the observed gravity and the gravity predicted from topography (i.e. Bouguer correction).

S2 Simple density depth-dependencies

Eq. (S4) or (S6) can be combined in general with a given expression for the depth-dependence $\rho(z)$, and a simple expression of the effective density at high degrees (i.e. short wavelengths) is obtained by considering the crustal thickness z_c to be large compared with the other lengthscales of interest. The case of low crustal thickness, where the crust-mantle interface's signal should be taken into account, will be detailed in Sec. S3.

In the case of a linear density increase with depth [$\rho(z) = \rho_{\rm surf} + az$], the theoretical effective density spectrum is given by:

$$\rho_{\rm eff,\,lin}(k) = \rho_{\rm surf} + \frac{a}{k}, \tag{S7}$$

where a is the density gradient. If the density is limited by a maximum value $\rho_{\rm max}$, a saturation term will appear in Eq. (S7) from the modified density profile

$$\rho(z) = \begin{cases} \rho_{\rm surf} + az & \text{if} \quad z \leq z_{\rm crit} \\ \rho_{\rm max} & \text{if} \quad z \geq z_{\rm crit} \end{cases}, \tag{S8}$$

with $z_{\rm crit} = (\rho_{\rm max} - \rho_{\rm surf})/a$ the critical depth below which the density is constant:

$$\rho^{\mathrm{sat}}_{\mathrm{eff},\,\mathrm{lin}}(k) = \rho_{\mathrm{surf}} + \frac{a}{k}\left(1 - e^{-kz_{\mathrm{crit}}}\right). \tag{S9}$$

Because of the saturation term, this case may be more appropriate, for instance, if one considers the density increase to be due to porosity closure alone. Eq. (S7) or (S9) can be fit to the observed effective density signal to estimate the spatial variation of the density gradient a and the surface density ρ_{surf} (e.g., Fig. 1).

More realistic density profiles are given by an asymptotic exponential increase of the density [*Athy*, 1930; *Carrier III et al.*, 1991; *Han et al.*, 2014] toward a maximal, deep value $\rho_0 = \rho_{\mathrm{surf}} + \Delta\rho$, in which case $\rho(z) = \rho_0 - \Delta\rho\, e^{-z/d}$, where d is an e-folding depth scale. The corresponding effective density spectrum is:

$$\rho_{\mathrm{eff},\,\exp}(k) = \rho_{\mathrm{surf}} + \frac{\Delta\rho}{1 + kd}. \tag{S10}$$

Note that if ρ_0 is the grain (i.e. intact rock) density ρ_g and if one considers density variations to be due to porosity closure only, then d is the e-folding depth scale of the porosity profile $\phi(z) = 1 - \rho(z)/\rho_g$. If d is large or k is large, then the effective density is close to the surface value, while if d is small or k is small, then the effective density is that at depth, ρ_0.

We note that the spatial distributions of the characteristic depth scale obtained by fitting the observations to a model are very similar irrespective of whether the model adopted is described by Eq. (S9) (i.e. depth scale is z_{crit}) or Eq. (S10) (depth scale is d). Similar distributions also arise through a simple model with two constant-density layers when one maps the thickness of the top (low-density) layer, fixing the bottom ("infinite") layer at the density ρ_g (see Sec. S6).

The theoretical effective density spectra derived in this section, for various density profiles $\rho(z)$, have been verified against numerically derived synthetic spectra. The latter were calculated using the same approach as for real data as described in Sec. S4. The synthetic gravity spherical harmonic coeffcients were calculated with the finite-amplitude method of *Wieczorek and Phillips* [1998], using power series of topography up to order 20 and for a crustal density profile $\rho(z)$ as specified by an analytical expression. Though the exact approach will be explained in Sec. S4, it is worth noting that, as compared to Eqs. (S2)-(S3), the idea is here to compute the synthetic effective density as follows:

$$\rho_{\mathrm{eff}}(k) = \frac{\Delta \mathrm{g}(k)}{\Delta \hat{\mathrm{g}}(k)}, \tag{S11}$$

where the numerator is the computed gravity for a given density profile $\rho(z)$, and the denominator is the gravity predicted from surface topography (i.e. Bouguer correction) per unit density. When the topography amplitude is small, we simply have $\Delta\hat{\mathrm{g}}(k) = 2\pi G\,\mathrm{h}(k)$, and Eqs. (S2)-(S3) are retrieved. In practice, Eq. (S11), valid in Cartesian geometry, is realized through Eq. (1) of the main text (spherical geometry), which also appears below as Eq. (S18).

Figure S2A shows the excellent agreement between the analytical spectra and the synthetic ones by mean of examples with exponential density profiles. Errors in our standard, high-degree range ($250 \leq \ell \leq 550$) are typically less than 0.1 % (see Fig. S2B). Similar agreement is obtained for the other types of density profiles considered above. This justifies a posteriori the use of Cartesian geometry and the neglect of the finite-amplitude correction in our analytical treatment, as long as the real effective density spectrum that will be fit with various analytical spectra is calculated by taking finite-amplitude effects into account (see Sec. S4).

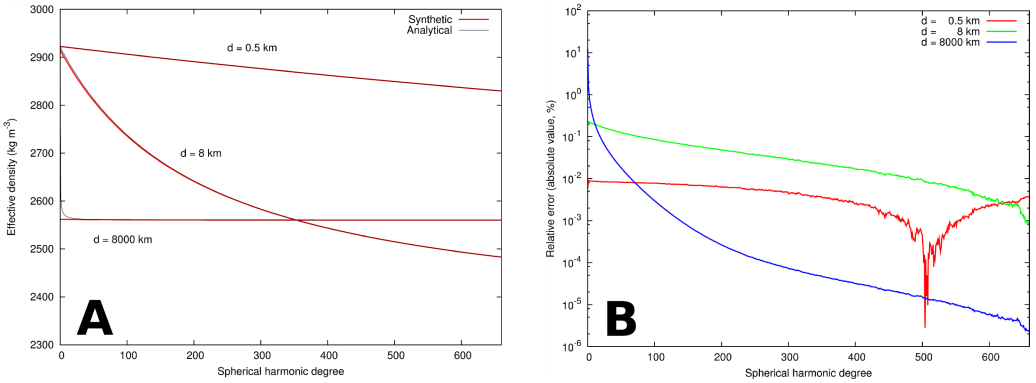

Figure S2: Comparison between analytical and synthetic effective density spectra. Three cases are displayed (**A**) for an exponential porosity profile: two cases with shallow main density variations ($d = 0.5$ km, $d = 8$ km; $\rho_0 = 2923$ kg m^{-3}, $\Delta\rho = 0.20\rho_0$), and a case with a quasi-constant density ($d = 8000$ km, $\Delta\rho = 0.12\rho_0$). The match is very good in the wavelength range of this study ($250 \leq \ell \leq 550$) and significant differences only appear at low degrees. Relative errors are displayed in panel **B**; note that the sharp slope transitions in the signal (e.g., at $\ell \sim 500$ for $d = 0.5$ km) correspond to a change of sign for the error.

Although equations for ρ_{eff} corresponding to Eqs. (S7), (S9), and (S10) can be derived directly in spherical geometry [e.g., *Han et al.*, 2014], the Cartesian approximation remains sufficient in our short-wavelength range.

S3 Effect of crustal thickness variations

The density constrast at the lunar crust-mantle boundary might contribute to the slope of the observed effective density spectra observed. Similarly, sharp transitions of porosity at depth (e.g., viscous compaction-induced porosity interface) could also affect that slope. This is especially true in regions with low crustal thickness, as in the South Pole-Aitken impact basin area. To investigate the effect of such density contrasts, we shall use an admittance approach in the same spirit as previously used for the shallow crustal density contribution. Consider two sinusoidal interfaces – here the lunar surface and the crust-mantle interface – separated by a mean distance z_u, and with a phase offset Φ. The density constrast at the upper and lower interfaces are ρ_u and $\Delta\rho_l$, respectively, the topographic amplitudes are h_0 and h_1. Neglecting sphericity, as our study focuses on short wavelength, and finite-amplitude corrections, the free-air gravity anomalies due to the upper and lower interfaces are respectively

$$\Delta g_u = 2\pi G \rho_u h_0 \cos(kx) \tag{S12}$$

and

$$\Delta g_l = 2\pi G \Delta\rho_l h_1 e^{-kt_u} \cos(kx + \Phi), \tag{S13}$$

where G is the gravitational constant, x is a horizontal coordinate, and k is the wavenumber.

The ratio of the net surface gravity $\Delta g = \Delta g_u + \Delta g_l$ to the surface topography h, referred to as the admittance Z, can be simply calculated from the formula [e.g., *Forsyth*, 1985; *Audet*, 2014]

$$Z(k) = \frac{\overline{\mathsf{h}(k)\Delta\mathsf{g}(k)}}{\overline{\mathsf{h}^2(k)}}, \tag{S14}$$

where the right-hand side functions of k are the Fourier components (here sans-serif characters) of the corresponding spatial functions, and overlines represent the average over some range of wavenumbers.

Then, assuming that h_0 and h_1 are statistically independent of Φ and of each other, and that Φ is independent of k [e.g., *Audet*, 2014], Eq. (S14) reduces to the simple formula

$$Z = 2\pi G\rho_u(1 + f_u \cos\Phi), \tag{S15}$$

with

$$f_u = \frac{h_1}{h_0}\frac{\Delta\rho_l}{\rho_u}e^{-kz_u}. \tag{S16}$$

If the two interfaces are in phase ($\cos\Phi = 1$) then the gravity contributions add, while if they are in anti-phase, they subtract. The quantity f_u compares the contribution of the lower interface to that of the upper interface: upward attenuation reduces the contribution of the lower interface.

At short wavelengths, the surface topography (which is caused mainly by impacts) is not expected to be correlated with topography on the crust-mantle interface. Furthermore, the topography of this interface may be modified by lower crustal viscous flow (relaxation), which is more rapid at short wavelengths. The distribution of Φ can therefore be reasonably be assumed to be random. If the relief along the crust-mantle interface is uncorrelated with surface topography, the resulting ensemble average of Eq. (S15) over Φ yields the same answer as Eq. (S3). The effective density will therefore be unbiased by the topography on the crust-mantle interface.

This argument is further supported by the lack of spatial correlation between the SP-A lower-d feature (Fig. 2) and the crustal thickness variations inside SP-A (see crustal thickness models of *Wieczorek et al.* [2013a]). Furthermore, in the non-SP-A farside, our model best-fit solution displayed in Fig. 2 exhibits low-d features in regions of higher crustal thickness and, conversely, high-d features in low-crustal thickness regions (e.g., at $\sim$ 0°N, 90°E or at $\sim$ 20°N, 180°E). We therefore conclude that variations in crustal thickness are unlikely to be affecting our results.

S4 Data windowing procedure – Multitaper approach

In this study, we use localized estimates of the spherical harmonic coefficients of the lunar gravity and topography fields in order to retrieve the crustal effective density spectrum. For this purpose, a multitaper windowing procedure was applied (see below). The original, global spherical harmonic coefficients for each field come from the data files listed in Tab. S1.

Table S1: GRAIL and LOLA data used in this study (spherical harmonic coefficients). The spatial fields were reconstructed up to spherical harmonic degree $\ell = 660$. Note that the data fits (see Sec. S5) were performed over a short-wavelength range ($250 \leq \ell \leq 550$). The Bouguer correction (b) was calculated in the same way as in the study of *Wieczorek et al.* [2013a], i.e. using the finite-amplitude method of *Wieczorek and Phillips* [1998]. In practice, only g and b are explicitly needed in our approach; see Eq. (S18) and (S20).

Data type	File	Comment
Free-air gravity (g)	`GRGM900B_SHA.TAB`	NASA GSFC model
Topography	`LRO_LTM01_PA_1080_SHA.TAB`	Principal axis referenced

The observed effective density was calculated from the spherical harmonic coefficients of the free-air (GRAIL) gravity, and of the gravity predicted from (LOLA) topography (Bouguer correction), evaluated assuming a unit surface density and taking the finite-amplitude correction into account. The procedure is summarized below.

The high-degree gravity contribution of the topography being a non-linear function of the topography, it can be written, in terms of spherical harmonic coefficients $f_{\ell m}$ of a function f:

$$g_{\ell m} = \rho_{\text{eff}}(\ell)\, b_{\ell m} + \nu_{\ell m}, \tag{S17}$$

where g and b refer to the observed free-air gravity and to the gravity predicted from surface topography (Bouguer correction) assuming unit density and applying the finite-amplitude correction [*Wieczorek et al.*, 2013a], respectively, and ν is the component of the observed signal that is not due to the surface topography. The term $\rho_{\text{eff}}(\ell)$ is an effective density that is sampled at a degree ℓ. Assuming ν to be a random variable uncorrelated with b, then, by multiplying both sides of Eq. (S17) by $b_{\ell m}$, summing over all angular orders m, and taking the expectation, we get an unbiased estimate of the crustal effective density at each degree:

$$\rho_{\text{eff}}(\ell) = \frac{S_{gb}(\ell)}{S_{bb}(\ell)}, \tag{S18}$$

with S_{fg} the cross-power spectrum of two functions f and g on the sphere:

$$S_{fg}(\ell) = \sum_{m=-\ell}^{\ell} f_{\ell m}\, g_{\ell m}. \tag{S19}$$

Another quantity of interest, that contains complementary information (see Sec. S5), is the correlation between g and b:

$$\gamma(\ell) = \frac{S_{gb}(\ell)}{\sqrt{S_{gg}(\ell) S_{bb}(\ell)}}. \tag{S20}$$

In order to map lateral variation in the crustal shallow density structure, we apply a windowing of the gravity and topography data. Localized estimates of the admittance and correlation spectra were obtained using the SHTOOLS package [*Wieczorek*, 2012]. The local analysis was performed in a series of regions of interest (ROIs) centered on each node of a quasi-equal-area distribution. A similar kind of distribution of ROIs was originally used in the study of *Wieczorek et al.* [2013a]. Here, we employ a coarser (though sufficient, see Sec. S5) distribution of 400 nodes, corresponding to an angular separation of $9.5 \pm 0.9^\circ$ (i.e. spatial resolution of $\sim$290 km). For each ROI, we calculated localized spectral estimates of the gravity and topography using the multitaper spectral analysis approach developed by *Wieczorek and Simons* [2005, 2007]. The localization windows are constructed to have their energy optimally concentrated in the region of interest for a specified spectral bandwidth of the windows. The optimization problem yields a family of orthogonal windows (spherical Slepian functions), or tapers, each with a specified concentration factor λ. In our analysis, the radius of the spherical cap is fixed at $\theta_0 = 15^\circ$ and the spectral bandwidth of the tapers is set to $L = 58$. With this bandwidth, there are $n = 30$ tapers whose concentration factors are such that $\lambda \geq 0.99$.

For each of these n tapers, the gravity and topography-predicted gravity fields are multiplied by the corresponding subwindow in the spatial domain. The resulting windowed fields are then decomposed into spherical harmonics, from which the admittance and correlation are estimated. Each individual tapered spectral estimate is then combined into a simple arithmetic average at each degree ℓ. Thus, for each ROI, an averaged spectrum for the effective density $\rho_{\text{eff}}(\ell)$ and the correlation $\gamma(\ell)$ is obtained, together with their associated standard deviations. The latter reflects the variability of the spectral estimates obtained with the various tapers associated with the ROI's window. It is therefore a way of characterizing

spatial variability inside each window, because the various tapers correspond to various spatial coverages over the window [1]. An alternative, and in principle more appropriate approach would be to compute the cross-power spectra, in Eq. (S18), associated with each individual taper, and then use the average of these cross-power spectra to calculate the effective density. In our high-degree range, given that we use a large number of tapers (see below), both approaches give very similar results. For example, over $\ell = 250 - 550$, the difference between the two kinds of effective density spectra for the regions depicted in Fig. 3 are less than 0.14 % (Fig. 3A) and less than 0.18 % (Fig. 3B).

We also employed a broad multitaper window ($\theta_0 = 82.5°$, $L = 11$, $n = 38$; Shannon number of $\sim$62.6) centered at 180°E,0°N to estimate the average characteristics of the farside; Fig. S3 shows the window's location over the lunar topography, the associated effective density spectrum, and the corresponding fit parameter space.

The high resolution gravity data obtained from the GRAIL mission are particularly well-suited for performing multitaper analyses: the spectral smoothing associated with our bandwidth L allows for an interpretation of a significant range of spherical harmonic degrees. In particular, as we use gravity and topography fields up to degree $\ell = 660$, the localized fields are reliable between degrees L and $660 - L$ [*Simons et al.*, 1997; *Wieczorek and Simons*, 2005, 2007]: for $L = 58$, we fit effective densities in the degree range $\ell = 250 - 550$.

In some cases, comparison of theoretical effective admittance spectra (which effectively correspond to a window that evenly weights data everywhere on the sphere) with observed spectra (derived using a finite window) can produce misleading results [e.g., *Pérez-Gussinyé et al.*, 2004]. We checked that this issue was not a problem for our results by generating a synthetic gravity field using a theoretical effective density profile as decribed below.

In order to obtain this synthetic gravity field, consider Eq. (S18) and our theoretical expressions for the effective density spectrum for a given depth-dependence of the density, $\rho_{\rm eff}^{\rm th}$ (see Sec. S2). We have:

$$S_{\breve{g}b}(\ell) = \rho_{\rm eff}^{\rm th}(\ell)\ S_{bb}(\ell), \tag{S21}$$

where $\breve{g}$ represents the synthetic gravity and b refers to the Bouguer correction for a unit density. The synthetic gravity coefficients therefore may simply be written

[1] See, e.g., Fig. 5.4 of *Simons et al.* [2006] for the spatial patterns associated with the various tapers.

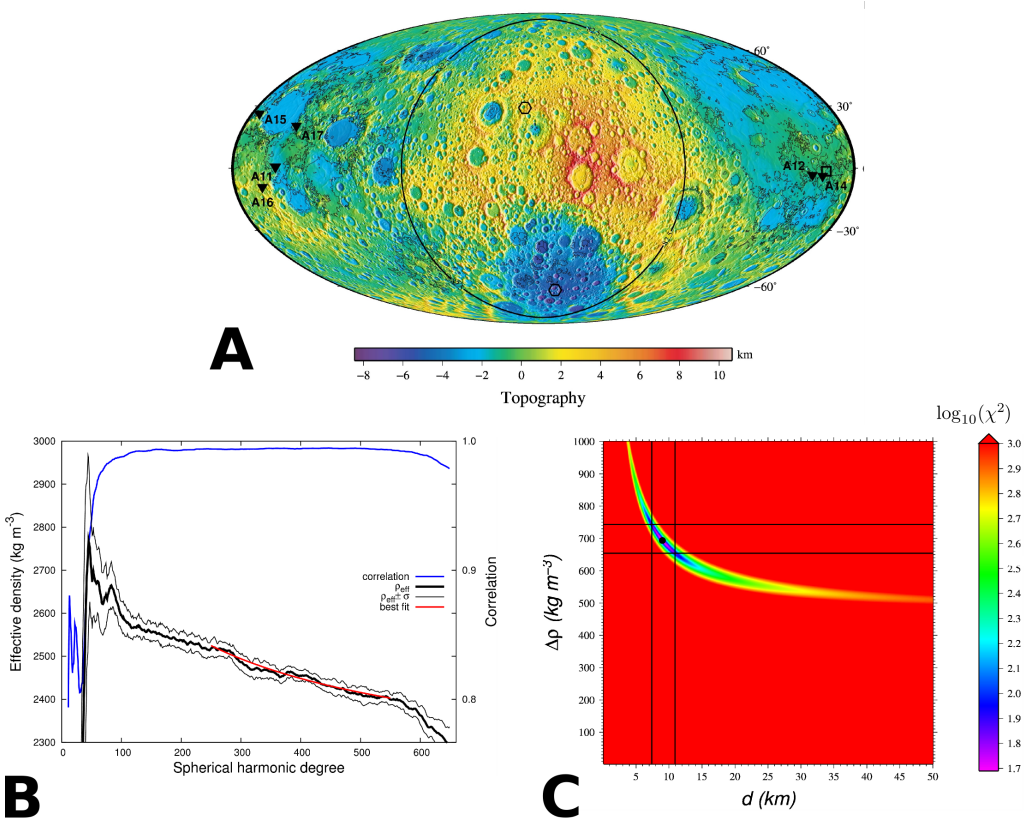

Figure S3: Average farside characteristics. (**A**) Broad ROI employed to extract farside average characteristics ($\theta_0 = 82.5°$, $L = 11$, $n = 38$), draped over shaded relief LOLA topography. Associated effective density spectrum and best-fit estimate for the exponential model ($d = 9.0^{+2.2}_{-1.8}$ km, $\rho_{\rm surf} = 2223^{+44}_{-54}$ kg m^{-3}) and correlation spectrum (**B**), and associated misfit map (**C**); format as in Figs. 1-3.

$$\check{g}_{\ell m} = \rho_{\text{eff}}^{\text{th}}(\ell)\, b_{\ell m}. \tag{S22}$$

This synthetic gravity field can then be localized (together with the Bouguer correction) in the same way as we do for the observed gravity field and the resulting effective density should ideally be identical to the input density model $\rho_{\text{eff}}^{\text{th}}$.

In the high-degree range considered for data fitting in this study ($\ell = 250 - 550$), the difference is very small (typically lower than 0.3 %). This is explained by the fact that, given that the amplitudes of the cross-power spectra are slowly varying at high degrees, the windowed power spectral estimates are not significantly biased from their global equivalents. Our results are thus not affected by the use of non-localized theoretical density spectra, i.e. the simple analytical expressions derived in section S2.

S5 Error analysis and statistical significance

To estimate the best fit parameters, we employ a simple chi-square (χ^2) minimization approach. Using a grid search-based parameter space exploration, every observed, windowed (multitapered) effective density spectrum $\rho_{\text{eff}}^{\text{obs}}$ is compared to a theoretical prediction $\rho_{\text{eff}}^{\text{th}}$ (see Sec. S2) to calculate the chi-squared misfit:

$$\chi^2 = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \left[\frac{\rho_{\text{eff}}^{\text{obs}}(\ell) - \rho_{\text{eff}}^{\text{th}}(\ell)}{\sigma(\ell)} \right]^2. \tag{S23}$$

Here, $\sigma(\ell)$ is the standard deviation at each degree ℓ of the observed effective density spectrum obtained from the multitaper windowing of the data (see Sec. S4). The short-wavelength range considered in this study corresponds to data from degree $\ell_{\min} = 250$ to $\ell_{\max} = 550$. In particular, this avoids the effects of flexure and/or crustal thickness variations at lower degrees, and ensures the correlation between the free-air gravity (g) and topography (or, equivalently, the Bouguer correction, b) is close to unity [e.g., *Han*, 2013, see also Fig. S3B], which is a requirement of our model. The Bouguer anomaly ($g - b$) spectrum shows a loss of fidelity for increasing degrees beyond $\ell > 550$, as in the studies of *Han* [2013], *Han et al.* [2014], and *Konopliv et al.* [2014]. However, the exact values of $\ell_{\min}$ and $\ell_{\max}$ are not critical to our results, as will be shown below (Fig. S9). We note also that a more rigorous approach would be to define χ^2 in Eq. (S23) with a sum

over a *subset* of the range $\{\ell_{\min} .. \ell_{\max}\}$, in order to take into account the fact that only estimates at degrees separated by $2L$ are statistically independent, L being the tapers' spectral bandwidth (see Sec. S4). However, we will show in Sec. S6 that this alternative approach does not affect our main conclusions.

For each individual analysis region window, uncertainties in estimated parameters are likely mostly due to the combination of regions with different geological characteristics (e.g., maria vs. non-maria) into the same analysis region. The error due to the gravity coefficients, estimated from a series of clones of the GSFC (see Tab. S1) gravity field [Mazarico, pers. comm.], are much smaller and represent a minor contribution. The main source of error on our estimated best-fit parameter is therefore of geological (hence a priori unknown) origin.

The magnitude of the signal that is not explained by our model (here referred to as geologic "noise") can be estimated from the correlation γ between gravity (g) and topography (i.e., here, Bouguer correction per unit density, b) [Eq. (S20)]. Assuming that the gravity signal contains some random signal ν that is not predicted by the model admittance and that is uncorrelated with b, it can be shown that the observed correlation $\gamma^{\rm obs}$ is given by

$$\gamma^{\rm obs}(\ell) = \frac{\gamma^{\rm th}(\ell)}{\sqrt{1 + \frac{S_{\nu\nu}(\ell)}{S_{\check{g}\check{g}}(\ell)}}}, \tag{S24}$$

where $\gamma^{\rm th}$ is the correlation that would be observed for the synthetic gravity $\check{g}$ predicted from a given model density profile [Eq. (S22)]. The second term under the square root in Eq. (S24) can be viewed as the inverse of a signal-to-noise ratio[2]: the absence of noise would make $\gamma^{\rm obs}$ equal to $\gamma^{\rm th}$. Therefore, using Eq. (S17), the noise power spectrum can be estimated from the ratio of the coherences γ^2:

$$S_{\nu\nu}(\ell) = \rho_{\rm eff}^{\rm th}(\ell)^2 \; S_{bb}(\ell) \left\{ \left[\frac{\gamma^{\rm th}(\ell)}{\gamma^{\rm obs}(\ell)} \right]^2 - 1 \right\}. \tag{S25}$$

Because of the implied in-phase nature of the density layers in our model (see Sec. S1), $\gamma^{\rm th}$ is unity, and the variance of the noise at each degree can be estimated from

$$\sigma_\nu^2(\ell) = \frac{S_{\nu\nu}(\ell)}{2\ell + 1}. \tag{S26}$$

[2]Note that $S_{gg}(\ell) = S_{\check{g}\check{g}}(\ell) + S_{\nu\nu}(\ell)$.

To determine an appropriate confidence interval on χ^2, and therefore error bars on our best-fit parameters, we performed Monte Carlo simulations as follows. At a given location, the best-fit effective density spectrum $\rho_{\rm eff}(\ell)$ was used to generate a series of synthetic gravity fields by means of Eq. (S17): at each degree ℓ, the noise coefficients $\nu_{\ell m}$ are distributed in a Gaussian random fashion of standard deviation $\sigma_\nu(\ell)$. The latter is computed from Eqs. (S25)-(S26), using the localized, observed correlation ($\gamma^{\rm obs}$). The localized effective density and correlation spectra of the resulting synthetics are then computed (see Sec. S4). Each realization is characterized by its χ^2 value [eq. (S23)] which quantifies the misfit to the input (best-fit theoretical) model.

Fig. S4 shows the synthetics obtained from a set of 1000 random realizations, the associated probability density function of χ^2, and the inferred admissible parameters range. Our usual, 15-degree radius localization window (see Sec. S4) was here centered on the farside. Panel A shows that the synthetics reasonably reproduce the characteristics of the observed spectrum. The associated 1-σ confidence interval is determined from the cumulative distribution function of χ^2. In this case we find that $\chi^2 < 1.23\chi^2_{\rm best}$ encompasses 68.27 % of the values, where $\chi^2_{\rm best}$ denotes the minimum misfit obtained by fitting the actual observed spectrum. For another, different example region, corresponding to panel **B** of Fig. 3 we found the normalized bound on χ^2 was $\sim$1.45. On this basis, we take a conservative, single bound of the acceptable χ^2 value of 1.5 times the local minimal (i.e. best-fit) value, assuming (rough) stationarity, and keeping in mind that the main assumption is that of random noise uncorrelated with topography. The fact that $\chi^2_{\rm best}$ (observed) is close to the expected value derived from the synthetics (see Fig. S4B) suggests that our approach is a reasonable approximate way to derive uncertainties on the local effective density (see also Sec. S6).

The region of acceptable models in the parameter space (e.g., Fig. S4C) enables us to define appproximations of the uncertainties on each parameters, as illustrated in Fig. S5.

Fig. S6 displays the minimum misfit (chi-square, $\chi^2_{\rm best}$) and the corresponding relative uncertainties on the parameters for the exponential model (i.e. Fig. 2). The displayed uncertainties on a parameter $\bullet$ are simply $\sigma_\bullet = \frac{\sigma^-_\bullet + \sigma^+_\bullet}{2}$.

In order to quantify the statistical difference between the SP-A region and the rest of the farside (see Fig. 2), a Monte Carlo analysis was performed as follows. Two broad regions of 30° radius (see Fig. 2) that each contain 26 window centers were considered. For each Monte Carlo realization we randomly assigned values

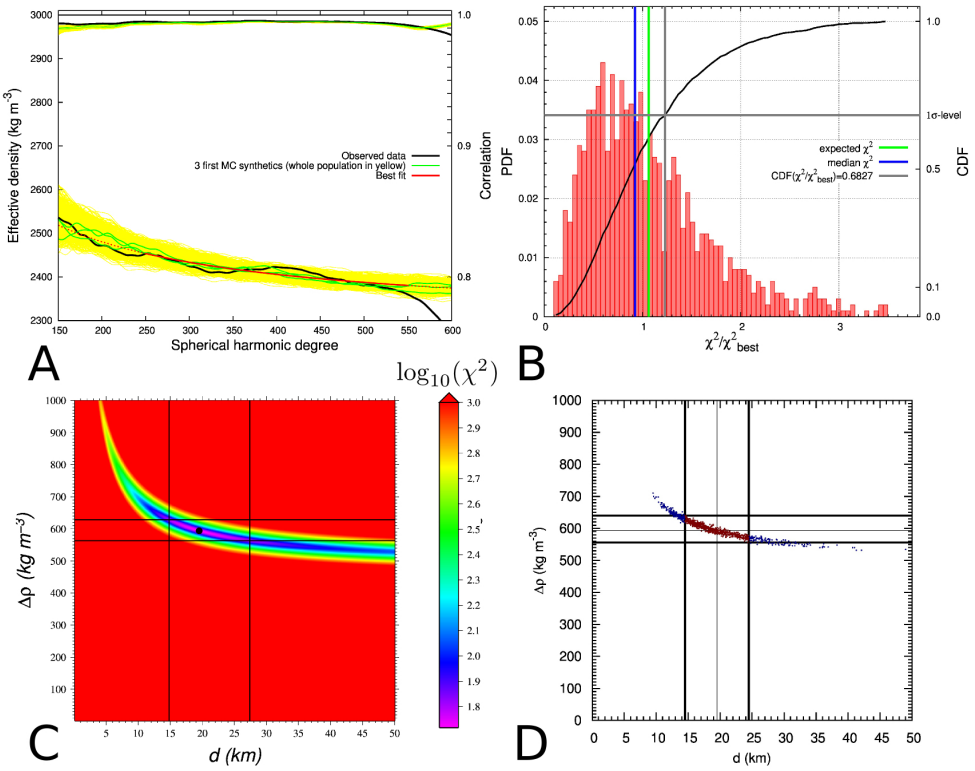

Figure S4: Example of a population of synthetic density spectra obtained from a set of 1000 random Monte Carlo realizations of a synthetic gravity field. The perturbation (noise) is estimated from the observed correlation (see text). The density model profile is exponential [Eq. (S10), with here $\rho_0 = \rho_g = 2892$ kg m^{-3}]. (**A**) Effective density (bottom curves) and corresponding correlation (top curves) spectra. The 3 first synthetics are shown (green curves), as well as the entire population (yellow curves). The observed data are displayed in black. The best-fit to the observed effective density over $\ell = 250-550$ is shown as the red curve. The regular, 15-degree radius localization window was here centered on the farside. (**B**) Probability density function (PDF, histogram) and associated cumulative distribution function (CDF, blue curve) of χ^2 of the population population with respect to the best-fit model. The median and expectation are indicated for reference. The horizontal axis shows χ^2 values normalized to the χ^2 obtained for the best-fit model to the observed data ($\chi^2_{\text{best}} \approx 52.17$). The 1-$\sigma$ confidence interval is determined from the CDF, and is $\chi^2_{\text{max}} \approx 1.23\chi^2_{\text{best}}$. (**C**) Associated misfit map (same format as in Figs. 3 & S3: dot locates best fit, lines depict admissible parameters range). (**D**) Distribution of the synthetics in the parameter space: each synthetics' pair of best-fit parameters is located by a dot. The red dots correspond to the 683 closest to the observation's best fit (thin lines), depicting an alternative admissible parameters range (thick lines) in good agreement with that shown in panel **C**.

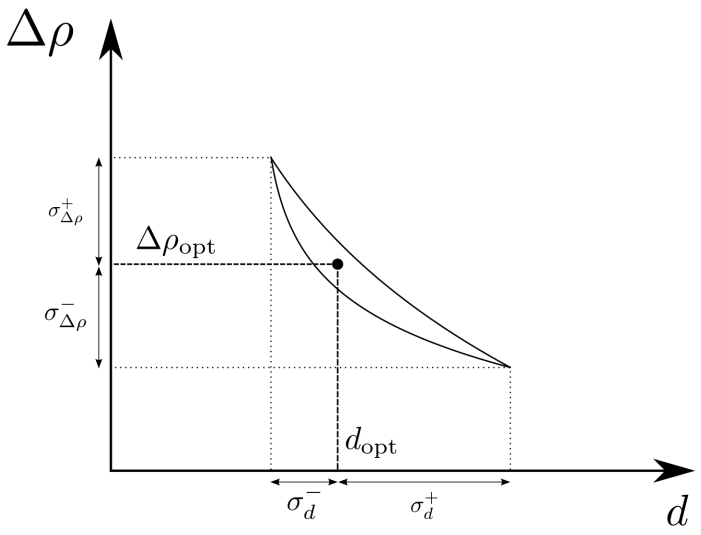

Figure S5: Definition of the error bars on the best-fit parameters. For each window's effective density profile, a grid search is performed over a range of value for the density increase depth scale d and density contrast $\Delta\rho$ (see text). A best-fit solution is found (dot), together with a range of admissible solutions (crescent-shape region) such that the associated misfit χ^2 is lower than $1.5\chi^2_{\mathrm{best}}$ (see text). Because of the skewness of the admissible region, we define asymmetric uncertainties $\sigma^+_\bullet$ and $\sigma^-_\bullet$ from the endmember solutions (dotted lines). Conservative Monte Carlo simulations of Fig. S7 below (p. 18) were peformed using half Gaussian random deviates from the best-fit values with corresponding standard deviations $\sigma^\pm_\bullet$.

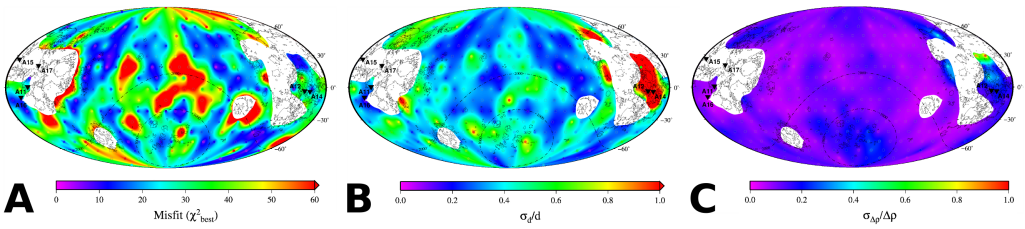

Figure S6: Spatial variability of the fit quality for the exponential model density profile and associated estimated uncertainties on best-fit parameters. (**A**) Minimum misfit (chi-square, χ^2_{best}) and associated relative uncertainties $\sigma_\bullet/\bullet$ on the (**B**) density increase depth scale (d; parameter mapped in Fig. 2) and (**C**) density contrast ($\Delta\rho$). For each individual window acceptable models (hence uncertainties) are considered to be such that $\chi^2 \leq 1.5\chi^2_{\mathrm{best}}$ (see text). Similar format as Fig. 2.

of d and $\Delta\rho$ to each window based on the associated $\sigma_\bullet^+$ and $\sigma_\bullet^-$ and assuming asymmetric Gaussian distributions (see Fig. S5). Fig. S7 displays the outcome of such Monte Carlo simulations for the exponential model density profile (i.e. Fig. 2).The results shown are the mean values $< d >$ and $< \Delta\rho >$ taken over the 26 windows, computed from a set of 10^6 realizations of each population. The final outcome are Gaussian-like, slightly asymmetric distributions. The width at half maximum for each of these distributions gives an estimate of the statistical uncertainties over d and $\Delta\rho$. The SP-A region and the typical non-SP-A farside appear to be significantly different with respect to the depth scale d (or, equivalently, with respect to the low-density layer's thickness; see Fig. S7): $d = 8.1 \pm 1.0$ km for SP-A, and $d = 15.0 \pm 1.5$ km for the non-SP-A farside. Note, however, that these are only approximate estimates as the uncertainties $\sigma_\bullet^\pm$ are underestimated in some places (as in Fig. 3B, where the upper bound of $\Delta\rho$ is not constrained through our approach).

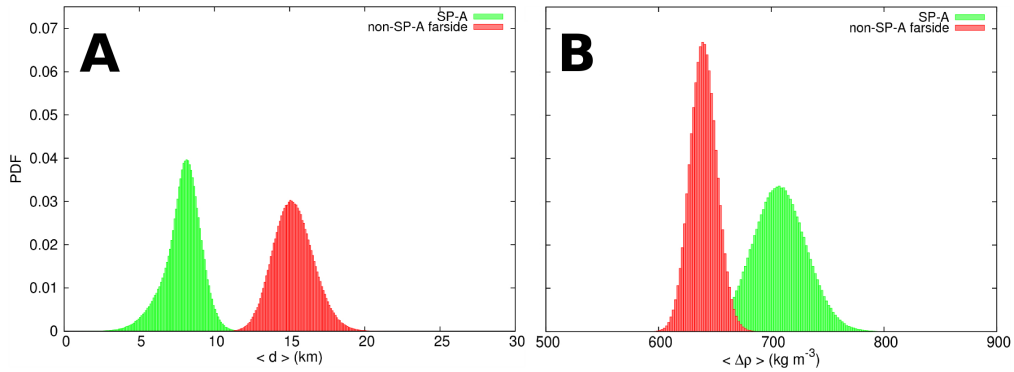

Figure S7: Statistical difference between the low density (porous) layers of SP-A and the rest of farside. Distribution of average values for the density increase depth scale d (**A**) and density contrast $\Delta\rho$ (**B**) for the model with an exponential density profile, from Monte Carlo simulations. For each of the two regions, the probability density function (PDF) of the mean value $< \ >$ of each parameter over 26 windows is computed from sets of 10^6 random draws, which were performed using a Gaussian perturbation of standard deviation $\sigma_d^\pm$ or $\sigma_{\Delta\rho}^\pm$ around the best fit value – see text for detail. Histograms' bin widths correspond to our parameters grid-search steps (see caption of Tab. S2, p. 23). A robust difference does exist between SP-A and the rest of the farside in terms of the density increase depth scale.

S6 Robustness to models and fitting parameters and sensitivity

All our standard calculations were made on a set of 400 nodes quasi-regularly distributed over the sphere (see Sec. S4) as a trade-off between too coarse, poorly resolved but computationally cheap grids and finer but computationally expensive grids. We checked that most of the spatial detail is captured with our nominal grid resolution when compared to a higher resolution case (e.g., 1000 nodes).

We also examined the effect of changing the window size (i.e. radius θ_0). Too broad a window strongly smoothes the inferred spatial features. Analyses using $\theta_0 = 15^\circ$ were found to give results very similar to those obtained with $\theta_0 = 10^\circ$ (i.e. approximately our node spacing, see Sec. S4). The window cannot be too small in order to keep a sufficiently large degree range for the spherical harmonic analysis.

The spatial resolution is also controlled by the spectral bandwidth L of the multitaper window in a trade-off with the spectral resolution. In most cases in this paper we favor spatial resolution over spectral resolution by taking a large value for L, implying a signifiant number n of optimally concentrated tapers (see Sec. S4: we take[3] $n = 30$ for $\lambda \geq 0.99$ with $L = 58$). This choice is mainly motivated by the importance of detecting spatial variations in crustal properties together with the theoretical expectation that the effective density curves will be spectrally smooth (see, e.g., Fig. 3 & S3).

To check how our results were affected if spectral resolution is favored over spatial resolution, we performed a windowing of the data with our nominal grid resolution (400 nodes) and a value $\theta_0 = 15^\circ$, and for $L = 20$ (Shannon number of $\sim$7.5) with $n = 8$ eigenvalue-weighted[4] tapers. With such a set of parameters, the spectral estimates at degrees $\ell >> L$ should exhibit variances that are about as good ($\sim$99 %) as if data were available everywhere on the sphere (see Fig. 10 of *Dahlen and Simons* [2008]). The resulting effective density spectra were then fit to theoretical spectra corresponding to linear and exponential density profiles, as described in Sec. S5. However, in this specific case, we applied a more rigorous calculation of χ^2 in the best-fit approach to take into account the fact that

[3] In general, we found that our results (best-fit parameters) were almost unaffected for $n \gtrsim 15$, although the spectral estimates exhibit a higher variance as n decreases [e.g., *Dahlen and Simons*, 2008].

[4] In our nominal case, we instead use only the best-concentrated tapers, that are then equally weighted. Both approaches should be equivalent [*Dahlen and Simons*, 2008].

only estimates $\rho_{\rm eff}(\ell)$ at degrees ℓ separated by $2L$ are statistically independent. Following *Lewis and Simons* [2012], who applied a similar local analysis for the Martian crustal magnetic field, we compute χ^2 given by Eq. (S23) with a sum over the degree set $\mathcal{S} = \{\ell_{\rm min}, \ell_{\rm min} + 2L + 1, \ell_{\rm min} + 2 \times (2L + 1), ...\}$. As our nominal fit degree range is between $\ell_{\rm min} = 250$ and $\ell_{\rm max} = 550$, $L = 20$ implies $\mathcal{N} = 8$ statistically independent spectral estimates. To correctly take into account the local spectral leakage bias at a given ℓ (correlation with adjacent degrees ℓ' such that $|\ell - \ell'| \leq L$), we computed the expected value of the multitaper estimate at each degree $\ell \in \mathcal{S}$, using the coupling matrix M detailed in *Dahlen and Simons* [2008] (see also *Wieczorek and Simons* [2005, 2007]):

$$\langle \rho_{\rm eff}(\ell) \rangle = \sum_{\ell'=\ell-L}^{\ell+L} M_{\ell\ell'} \, \rho_{\rm eff}(\ell'). \tag{S27}$$

The previous estimates and corresponding standard deviations over adjacent degrees $|\ell - \ell'| \leq L$ for each of the $\mathcal{N}$ spectral bands are then used in the calculation of χ^2 [Eq. (S23)]. To compute M, we use the analytical expression provided by *Dahlen and Simons* [2008], in its approximate form which is valid for degrees $\ell, \ell' \gg L$ [see their Eq. (146)]. Note that, for a given distance $|\ell - \ell'|$, the expression for $M_{\ell\ell'}$ depends only on L ; see Eqs. (146) & (6) of *Dahlen and Simons* [2008]. Finally, the coupling (S27) should in principle be applied to S_{gb} and S_{bb} separately [see Eq. (S18)], although this should only mildly affect the value of χ^2.

Fig. S8 displays the outcome of the windowing and fitting procedures decribed just above. It shows both the spatial distribution of the best-fitting density increase depth scale and the mare detection (masked regions) from the linear model (Fig. S8A), together with the corresponding minimum misfit (Fig. S8B). When comparing with corresponding Figs. 2 and S6A, the smoothing effect of the poorer spatial resolution implied by the chosen higher spectral resolution ($L = 20$) can be appreciated. In particular, only the strongest mare regions are detected (i.e. broadest and/or thickest). However, although the local values of the best-fit parameters may slightly differ (see also examples below), the main characteristics of our nominal model are conserved; in particular, the SP-A signature (thinner low-density layer) is still strong. Note that the misfit is here displayed as the reduced chi-square χ^2/ν: with the more rigorous definition of χ^2 adopted here (see above), and because the number of free (fitted) parameters is 2, the (maximal) number of degrees of freedom is here exactly $\nu = \mathcal{N} - 2 = 6$.

In addition, the above reduced χ^2 approach yields error estimates on param-

eters that are in general of comparable magnitude (or smaller) to those derived through the correlation-based Monte Carlo approach described in Sec. S5. For instance, for the two cases depicted in Fig. 3 (see also Tab. S2), the ranges $\rho_{\rm surf} \leq 2206$ kg m^{-3} and $d = 2.3^{+1.1}_{-0.2}$ km for the region within SP-A become $\rho_{\rm surf} \leq 2030$ kg m^{-3} and $d = 3.5^{+0.3}_{-0.1}$ km ($\chi^2_{\rm best}/\nu \approx 1.68$). For the other, non-SP-A farside region, the ranges $\rho_{\rm surf} = 2308^{+34}_{-40}$ kg m^{-3} and $d = 28.4^{+19.8}_{-9.3}$ km become $\rho_{\rm surf} = 2325^{+28}_{-32}$ kg m^{-3} and $d = 29.2^{+12.9}_{-7.7}$ km ($\chi^2_{\rm best}/\nu \approx 1.66$). These new error estimates were derived from the 1-σ confidence level associated with the regular reduced χ^2, i.e. from the condition $\chi^2/\nu \leq \chi^2_{\rm best}/\nu + \sqrt{2/\nu}$.

Fig. S8B shows that our approach provides a reasonable fit to many regions (χ^2/ν around 1), although care should be taken regarding the exact value of an "ideal" χ^2/ν with such a low number of degrees of freedom. In regions where larger misfit values are obtained, our theoretical model is probably too simple to account for the detail of the observed signal. However, it should be remembered that our goal is to investigate spatial variations and trends in the density stratification of the lunar crust, rather than determining an exact model at some particular region.

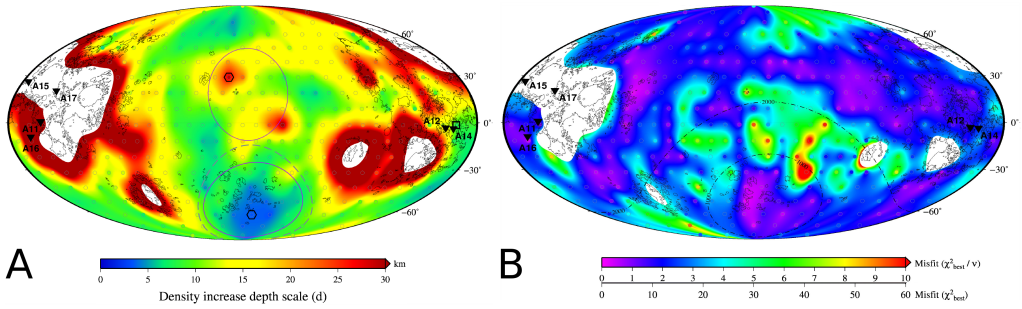

Figure S8: Effect of higher spectral resolution at the expense of a lower multitaper spatial resolution. The data windowing parameters are here $\theta_0 = 15°$ with $L = 20$ and $n = 8$ (see text for more detail). Results are shown through the distribution of the best-fit density increase depth scale d (**A**) and the corresponding misfit map (**B**). The misfit is here displayed as the best-fit reduced chi-square χ^2/ν (see also scale for χ^2), ν being the (maximal) number of degrees of freedom (see text). Masked areas correspond to the mare regions as detected from the linear model (i.e. $a < 5$ kg m^{-3} km^{-1} criterion; see Figs. 1A & 2). Panel **A** should be compared with Fig. 2 (same format).

The significant difference between SP-A and the rest of the farside, depicted in Figs. 2-3 & S7, is robust to different theoretical models. Mapping the critical depth $z_{\rm crit}$ associated with the "saturated" linear model [Eq. (S9)] gives spatial patterns that are very similar to those shown in Fig. 2; $z_{\rm crit}$ is the depth at which the maximal (grain) density will be reached using the best-fit surface value ($\rho_{\rm surf}$) and gradient (a). An equivalent result is obtained by mapping the best-fit top layer thickness of a two-layer (each of constant density) model (fixing the bottom layer's density at the local grain density).

As can be seen from Fig. S9, the SP-A/non-SP-A farside difference is also robust to the details of the fitting procedure, namely the chosen deep density ρ_0 (Fig. S9B) or the fit's degree range (Fig. S9C-D); though local values of the best-fit parameters can vary (e.g., Tab. S2), the spatial patterns are conserved.

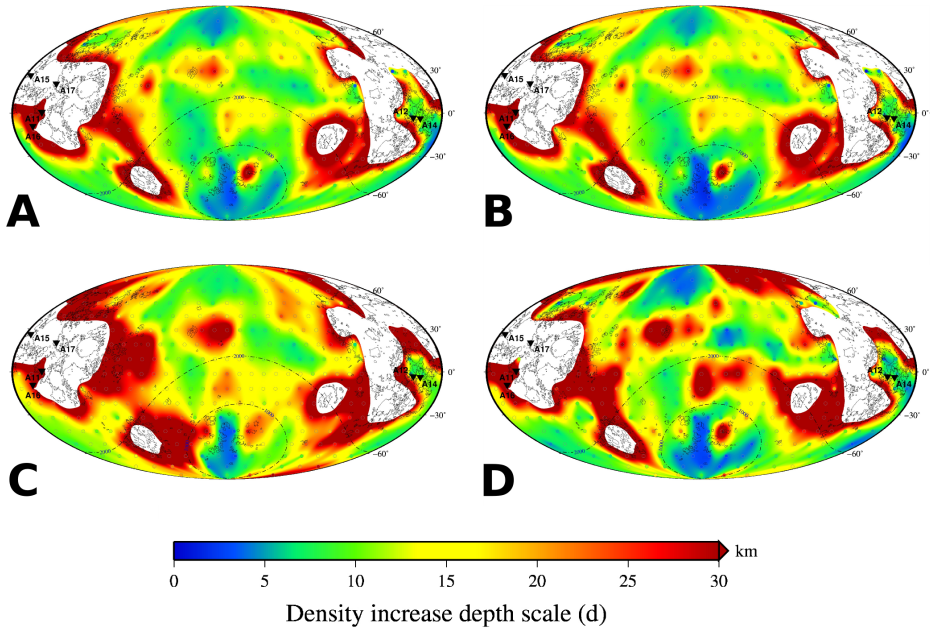

Figure S9: Effect of a constant deep density ρ_0 and of the values of the upper and lower bounds for the fit's degree range on the spatial patterns of the best-fit parameters – example with the exponential density model profile. (**A**) Best-fit solutions for our canonical model ($\ell = 250 - 550$ and $\rho_0 =$ local grain density, i.e. Fig. 2); (**B**) Same case with ρ_0 fixed at 2917 kg m^{-3}. The bottom row is for the same ρ_0 as in panel **A**, but with degree ranges $\ell = 150 - 550$ (**C**) or $\ell = 250 - 450$ (**D**). Format is similar to Fig. 2.

Table S2: Example of best-fit parameters at the two locations of Fig. 3 for different fit parameters of the exponential model. ρ_g is the surface grain density; its average over the farside (over a region defined in Fig. S3A) is $< \rho_g >= 2917$ kg m^{-3}. The surface density is $\rho_0 - \Delta\rho$. Our grid-search space is $\Delta\rho = 2 - 1000$ kg m^{-3} (steps of 2 kg m^{-3}) and $d = 0.1 - 50$ km (steps of 0.1 km). The lines in bold characters indicate our nominal case (i.e. Fig. 3).

Window (center coordinates)	$\ell_{\min}$	$\ell_{\max}$	$\Delta\rho$ (kg m^{-3})	ρ_0	d (km)	$\chi^2_{\rm best}$	$\frac{\chi^2_{\rm best}}{\ell_{\max}-\ell_{\min}+1}$
SP-A	**250**	**550**	**994**	ρ_g	**2.3**	**51.96**	**0.173**
	150	550	902	ρ_g	2.6	107.06	0.267
(191.25°E, -62.32°N)	250	450	988	ρ_g	2.3	37.47	0.186
$\rho_g = 2996$ kg m^{-3}	250	550	978	$< \rho_g >$	1.6	84.76	0.282
non-SP-A farside	**250**	**550**	**596**	ρ_g	**28.4**	**22.03**	**0.073**
	150	550	584	ρ_g	34.0	30.25	0.075
(168.39°E, 29.07°N)	250	450	590	ρ_g	30.4	17.34	0.086
$\rho_g = 2904$ kg m^{-3}	250	550	610	$< \rho_g >$	28.9	22.06	0.073

Fig. S10 below displays the average density/porosity maps obtained with a single-layer density model [i.e. $\rho(z)$ is constant]. Results are shown for both our high-degree range and the lower degree range employed in the study of *Wieczorek et al.* [2013a] (bulk porosity). The azimuthally averaged radial structure in bulk porosity and exponential density increase depth scale d within SP-A is diplayed in Fig. S11 (see main text).

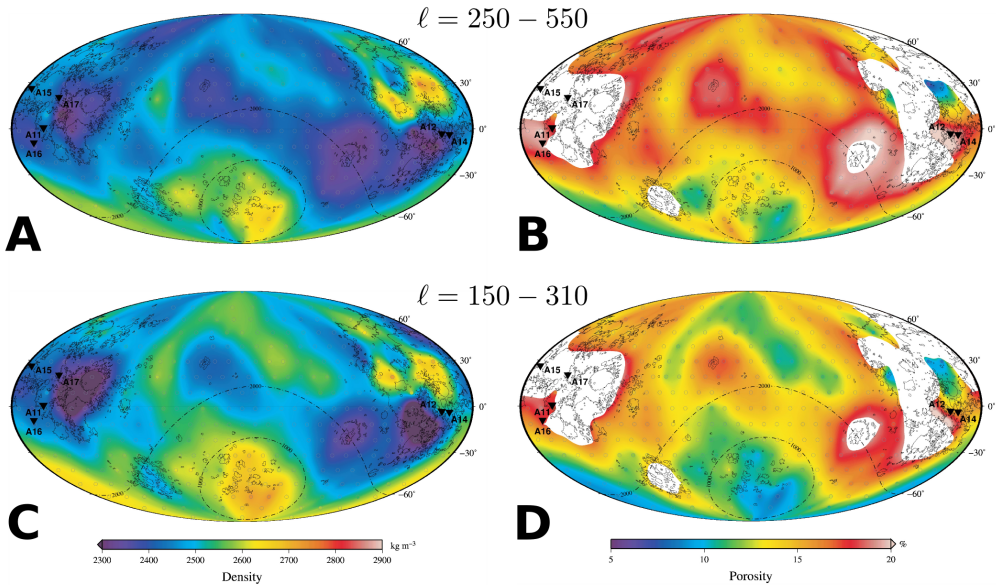

Figure S10: Column-averaged crustal density (**A**, **C**) and corresponding porosity (**B**, **D**) over degree range $\ell = \ell_{\min} - \ell_{\max}$. Top row corresponds to our nominal, high-degree range (shallow density/porosity); bottom row corresponds to a lower degree range (bulk density/porosity), employed in the study of *Wieczorek et al.* [2013a]: the bulk density/porosity distribution is consistent with that study (see their Figs. 1-2 & S7), though our spatial resolution is lower. Porosity values are generally smaller in **D** than in **B**. This is because the low degree range is more sensitive to greater depths, and porosity decreases with depth (see also *Han* [2013] and *Han et al.* [2014]). The "mare" regions [defined here as regions with shallow density gradients $a < 5$ kg m^{-3} (see Figs. 1A & 2)] have been masked on the porosity maps as the grain densities used (model of *Huang and Wieczorek* [2012], available at `http://www.ipgp.fr/%7Ewieczor`) are probably not relevant to the (bulk) crust, underneath the mare. Format as in Fig. 2.

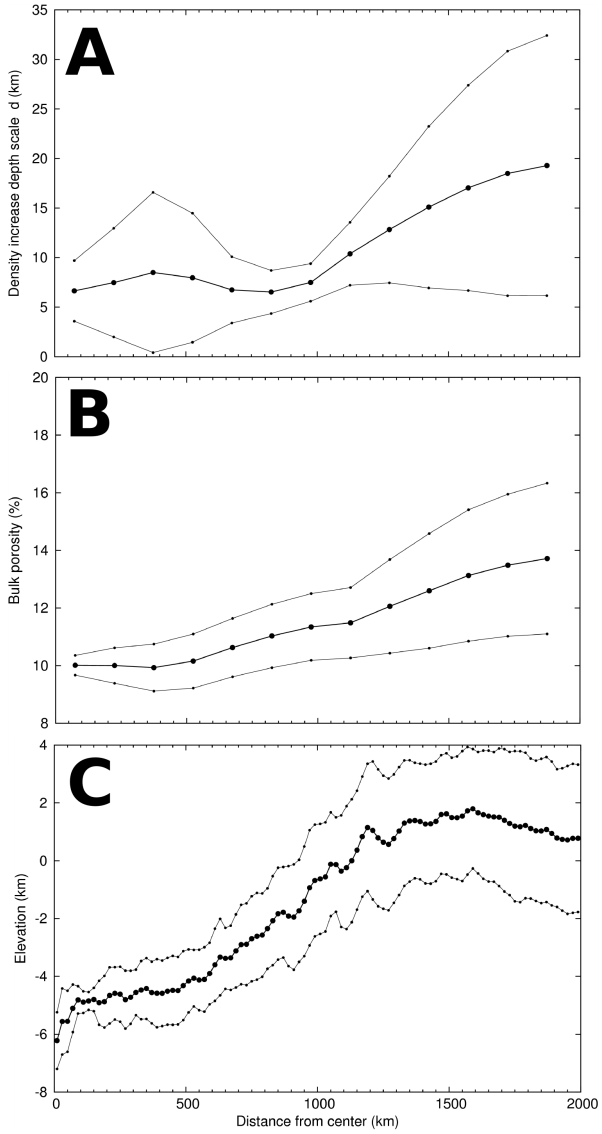

Figure S11: Averaged radial variation for various parameters within the South Pole-Aitken (SP-A) basin region. Best-fit density increase depth scale (d) of the exponential model (**A**), bulk porosity (**B**), and corresponding LOLA topographic profile (**C**). Data were azimuthally averaged in radial bins (dots) of about half our spatial resolution in extent (interpolated fields were used), except for the topography (higher spatial resolution), and correspond to the fields displayed in Figs. 2 & S10D. Thick lines denote the mean and thin line the ± 1 standard deviation levels in each bin. SP-A's center is taken as defined by *Garrick-Bethell and Zuber* [2009], i.e. at -55° N, 191.1° E.

S7 Theoretical vs. observed seismic velocities

In Fig. 4, two kinds of porosity dependencies were used: an empirical law calibrated on lunar anorthositic Apollo samples [*Sondergeld et al.*, 1979], and a purely theoretical law: the so-called three-phase model [*Smith*, 1974; *Christensen and Lo*, 1979], also referred to as the inclusion-matrix-composite model, or as the Generalized Self-Consistent Method (GSCM). Note that we do not consider here the effect of thermal expansion on the seismic velocity.

Empirical model

The empirical law of *Sondergeld et al.* [1979] we use can be written as

$$V_P(\phi) = V_{P_0}\,(1-\phi)^{-\frac{1}{2}}\,\exp\left[\frac{(\phi^2-\xi)\,\phi}{2(1-\phi)}\right], \tag{S28}$$

where V_P is the compressional velocity for a given porosity ϕ, $V_{P_0} = 7.15$ km s^{-1} is the unfractured rock velocity, and ξ is an empirical parameter [*Warren*, 1969]. Two endmember values of the empirical parameter ξ were used in Fig. 4: $\xi = 18$ – relevant to highly fractured lunar anorthosites, with long, elongated cracks – and $\xi = 1.97$ – relevant to spherical pores [*Warren*, 1969; *Sondergeld et al.*, 1979]. Eq. (S28) with $\xi = 18 - 24$ accounts for many published data on fractured lunar anorthosites (see references in *Sondergeld et al.* [1979]).

Theoretical model

This section describes a three-phase model approach [*Smith*, 1974; *Christensen and Lo*, 1979] that was used in Fig. 4, valid for spherical inclusions (i.e. here vacuum-filled pores). Below we use the following, standard form [e.g., *Mavko et al.*, 2009] for the effective compressional velocity of a rock with a porosity ϕ:

$$V_P(\phi) = \sqrt{\frac{K(\phi) + 4\mu(\phi)/3}{\rho(\phi)}}, \tag{S29}$$

where $K(\phi)$ and $\mu(\phi)$ are the effective bulk and shear modulus, respectively. The relevant density is $\rho(\phi) = \rho_0(1-\phi)$, with ρ_0 the unfractured rock (i.e. grain) density.

In the three-phase model, the effective bulk modulus is equivalent [*Christensen*, 1979] to the effective bulk modulus in the composite sphere model of *Hashin* [1962]. In the case with vacuum-filled pores, we have:

$$K(\phi) = K_0 \left(1 - \frac{\phi}{1 + (\phi - 1)\frac{K_0}{K_0 + \frac{4}{3}\mu_0}} \right), \tag{S30}$$

where the subscripts "0" denote matrix (grain) properties.

However, the effective shear modulus is implicitly given by the positive root of the following quadratic equation:

$$AX^2 + BX + C = 0, \tag{S31}$$

where $X = \mu(\phi)/\mu_0 - 1$, and A, B, and C are polynomial functions of the porosity and the matrix's Poisson ratio ν_0 [*Smith*, 1974]. The latter and the other matrix properties were chosen close to the typical values for anorthosite [*Williams and Jadwick*, 1980], and such that the values of $V_P(\phi = 0)$ given by Eqs. (S28) and (S29) are identical. Tab. S3 gives the chosen coefficients.

Table S3: Various anorthosite-like elastic properties [*Williams and Jadwick*, 1980] used to derive compressional seismic velocities from the three-phase model. Data are such that the intact (non-porous) rock velocity is the same as V_{P_0} in Eq. (S28). For the vacuum-filled pores, the moduli K and μ, and Poisson's ratio are considered equal to zero [e.g., *Christensen*, 1979].

Property	Value
K_0	96.33 GPa
μ_0	35 GPa
ν_0	0.338
ρ_0	2800 kg m^{-3}

Most of the effective properties obtained with the three-phase model we employ coincide [Mogilevskaya, pers. comm.] with those provided by the GSCM for cracks [e.g., *Huang et al.*, 1994]; the latter gives results that are very similar to those obtained with numerical simulations that accurately take pores' interactions into account[5] [*Mogilevskaya et al.*, 2007, see their Fig. 7].

Many different theoretical and/or empirical approaches could be used to predict seismic velocities as a function of porosity [e.g., *Mavko et al.*, 2009]. We note that, encouragingly, results that are very close to ours have recently been presented by *Schmerr and Han* [2014] in an independent, complementary study using Modified Biot-Gassman theory [*Lee*, 2008] (compare Fig. 1 of *Schmerr and Han* [2014] with our Fig. 4).

Selected seismic velocity determinations

Here, we provide complementary information about the six selected lunar velocity models used in Fig. 4.

The shallow model of *Cooper et al.* [1974] corresponds to a linear trend, $V_P(z) = az + b$, that excludes the biased Apollo 17 Lunar Module impact data [*Nakamura*, 2011], and extrapolated down to a depth of 2.5 km. With z being the depth in kilometers, the (best-fit) coefficients are $a = 0.778$ s^{-1} and $b = 0.395$ km s^{-1} (see their Fig. 19).

We also plot the pre-Apollo 17 era crustal model of *Kovach and Watkins* [1973] (see their Fig. 11), and two reference, post-Apollo era models: the first is the model of *Toksöz et al.* [1974], restricted to Oceanus Procellarum regions beneath Apollo 12 and 14 sites (see their Fig. 19; see also *Goins et al.* [1981]); the second is the model of *Nakamura* [1983] (see his Tab. 6), and includes the complete Apollo seismic data set.

The two more recent models that are included in Fig. 4 use the complete Apollo data set in a modern approach. The model of *Khan and Mosegaard* [2002] uses the same data set as in the model of *Nakamura* [1983] in a Bayesian approach (here inversion of artificial impacts, crustal thickness of $\sim$38 km; see their Fig. 15A). The model of *Lognonné et al.* [2003] relies on a complete reprocessing of the Apollo data; we use their model C (see their Fig. 5).

[5]The three-phase model only approximately takes cracks'/pores' (i.e. inclusions') interactions into account by embedding a single inclusion in a layer of matrix – to provide the required porosity – that is, in turn, surrounded by the material with the effective (unknown) property (i.e. that is subjected to the effective stress field).

Supplementary references

Athy, L. (1930), Density, porosity, and compaction of sedimentary rocks, *Bull. Amer. Assoc. Petrol. Geol.*, *14*, 1–24.

Audet, P. (2014), Toward mapping the effective elastic thickness of planetary lithospheres from a spherical wavelet analysis of gravity and topography, *Phys. Earth Planet. Int.*, *226*, 48–82, doi:10.1016/j.pepi.2013.09.011.

Carrier III, W. D., G. R. Olhoeft, and W. Mendell (1991), Physical properties of lunar surface, in *Lunar Sourcebook*, edited by G. H. Heiken, D. T. Vaniman, and B. French, chap. 9, pp. 475–594, Cambridge University Press, New York, USA.

Christensen, R. M., and K. H. Lo (1979), Solutions for effective shear properties in three phase sphere and cylinder models, *J. Mech. Phys. Sol.*, *27*, 315–330, doi:10.1016/0022-5096(79)90032-2.

Forsyth, D. W. (1985), Subsurface loading and estimates of the flexural rigidity of continental lithosphere, *J. Geophys. Res.*, *90*, 12,623–12,632, doi:10.1029/JB090iB14p12623.

Goins, N. R., A. M. Dainty, and M. N. Toksöz (1981), Lunar seismology - The internal structure of the moon, *J. Geophys. Res.*, *86*, 5061–5074, doi:10.1029/JB086iB06p05061.

Hashin, Z. (1962), The elastic moduli of heterogeneous materials, *J. Appl. Mech.*, *29*, 143–150, doi:10.1115/1.3636446.

Huang, Y., K. X. Hu, and A. Chandra (1994), A generalized self-consistent mechanics method for microcracked solids, *J. Mech. Phys. Sol.*, *42*, 1273–1291, doi:10.1016/0022-5096(94)90035-3.

Lee, M. (2008), Comparison of the modified Biot-Gassmann theory and the Kuster-Toksöz theory in predicting elastic velocities of sediments, *USGS Sci. Invest. Rep.*, *2008-5196*.

Mavko, G., T. Mukerji, and J. Dvorkin (2009), *The Rock Physics Handbook - 2nd Edition*, 511 pp., Cambridge University Press, Cambridge, UK.

Pérez-Gussinyé, M., A. R. Lowry, A. B. Watts, and I. Velicogna (2004), On the recovery of effective elastic thickness using spectral methods: Examples from synthetic data and from the Fennoscandian Shield, *J. Geophys. Res. (Solid Earth)*, *109*, B10409, doi:10.1029/2003JB002788.

Simons, F. J., F. A. Dahlen, and M. A. Wieczorek (2006), Spatiospectral concentration on a sphere, *SIAM Rev.*, *48*, 504–536, doi:10.1137/S0036144504445765.

Simons, M., S. C. Solomon, and B. H. Hager (1997), Localization of gravity and topography: Constraints on the tectonics and mantle dynamics of Venus, *Geophys. J. Internat.*, *131*, 24–31, doi:10.1111/j.1365-246X.1997.tb00593.x.

Turcotte, D. L., and G. Schubert (2002), *Geodynamics - 2nd Edition*, 456 pp., Cambridge University Press, Cambridge, UK.

Warren, N. (1969), Elastic constants versus porosity for a highly porous ceramic, perlite, *J. Geophys. Res.*, *74*, 713–719, doi:10.1029/JB074i002p00713.

Wieczorek, M. A. (2012), SHTOOLS, Software package freely available at `http://shtools.ipgp.fr`.

Wieczorek, M. A., and R. J. Phillips (1998), Potential anomalies on a sphere - Applications to the thickness of the lunar crust, *J. Geophys. Res.*, *103*, 1715–1724, doi:10.1029/97JE03136.

Williams, R. J., and J. J. Jadwick (1980), *Handbook of Lunar Materials*, 120 pp., NASA Ref. Public. 1057, NASA, Washington D.C., USA.