

JGR Solid Earth

RESEARCH ARTICLE

10.1029/2018JB016593

Key Points:

- A data-driven approach is presented to determine crustal structure and its uncertainty on a global scale
- Uncertainty of crustal thickness is less than 4 km in well-studied areas and reaches 12 km in poorly studied regions

Supporting Information:

- Supporting Information S1
- Data Set S1
- Data Set S2
- Data Set S3

Correspondence to:

W. Szwilus,
wolfgang.szwilus@ifg.uni-kiel.de

Citation:

Szwilus, W., Afonso, J. C. C., Ebbing, J., & Mooney, W. D. (2019). Global crustal thickness and velocity structure from geostatistical analysis of seismic data. *Journal of Geophysical Research: Solid Earth*, 124, 1626–1652.
<https://doi.org/10.1029/2018JB016593>

Received 21 AUG 2018

Accepted 8 JAN 2019

Accepted article online 14 JAN 2019

Published online 6 FEB 2019

Global Crustal Thickness and Velocity Structure From Geostatistical Analysis of Seismic Data

Wolfgang Szwilus¹ , Juan Carlos Afonso² , Jörg Ebbing¹ , and Walter D. Mooney³

¹Department of Geosciences, Kiel University, Kiel, Germany, ²Department of Earth and Planetary Sciences, Macquarie University, Sydney, New South Wales, Australia, ³USGS, Menlo Park, CA, USA

Abstract

Active source seismology provides a critical constraint on the global crustal structure. However, the heterogeneous data coverage means that interpolation is necessary to fill the gap between seismic profiles. This has the potential to cause large uncertainties especially if the data are interpolated over a large distance. In previous models, geological intuition was often employed to ensure reasonable results. To investigate crustal model uncertainty, we apply geostatistical analysis to a database of active seismic investigations. Unlike previous models, our workflow in the construction of the crustal model is completely transparent. Apart from the points from the database, we only use an a priori separation in oceanic and continental domains. We calculate global maps of Moho depth and average P wave velocity in the crystalline crust. Additionally, we obtain the interpolation error and error covariance. Overall, our results agree with previous global crustal models such as Crust1.0. Our uncertainty estimates show that the Moho depth uncertainty in the most well studied areas such as North America and Europe is less than 4 km but can reach 10 km or more in frontier regions such as most of Africa. P wave velocity shows the same pattern, but is less accurate overall, due to more small-scale variation. We demonstrate the benefit of having a numerical estimate of uncertainty by propagating the uncertainty to the residual topography. We see two main uses for our crustal model in the geophysical research community: (1) as a starting model for inversions focusing on the crust and upper mantle and (2) as a starting point for including other pointwise information about crustal structure, for example, from passive seismology.

1. Introduction

Global models that describe the seismic velocity and density structure of the crust play a crucial role for many geophysical applications. For instance, crustal models are needed to determine crustal corrections for mantle seismic tomography (Schaeffer & Lebedev, 2015) or for calculating residual mantle gravity and the crustal contribution to isostatic topography (Kaban et al., 1999). Gravity inversions often use crustal models as additional constraints or as a starting model (e.g., Reguzzoni et al., 2013).

The most common global crustal models are those based primarily on seismic information. Specifically, active (controlled-source) seismic studies are carried out along regional-scale profiles and use refracted or reflected seismic waves to determine thickness and velocity of the crustal layers (Prodehl & Mooney, 2012). Although active source studies provide the most accurate information, the use of so-called passive methods (based on seismic energy generated by earthquakes) has become routine in the past decade. Receiver functions (Kind et al., 2012) and the ambient noise technique (Shapiro & Campillo, 2004) are two of the most popular techniques.

Creating a two-dimensional model of crustal properties from seismic information given along profiles or at individual points requires some form of interpolation. Most authors use geological a priori information during the interpolation process to ensure that a geologically plausible result is obtained. For instance, the Crust5.1 model (Mooney et al., 1998) uses tectonic domains based on broad age categories ranging from Archean to Proterozoic (Artemieva & Mooney, 2001) and crustal domains (oceans, rifts, orogen, etc.). Crustal thickness from active seismology is then interpolated separately within each category. The P wave velocity structure is described by three layers, although not all layers need to exist at every point. Each category is assigned an archetypical vertical velocity profile that is assigned to all points of this category. In this way no explicit interpolation of velocity information from individual points is necessary. The more recent models Crust2.0 (Bassin et al., 2000) and Crust1.0 (Laske et al., 2013) have essentially used the same idea.

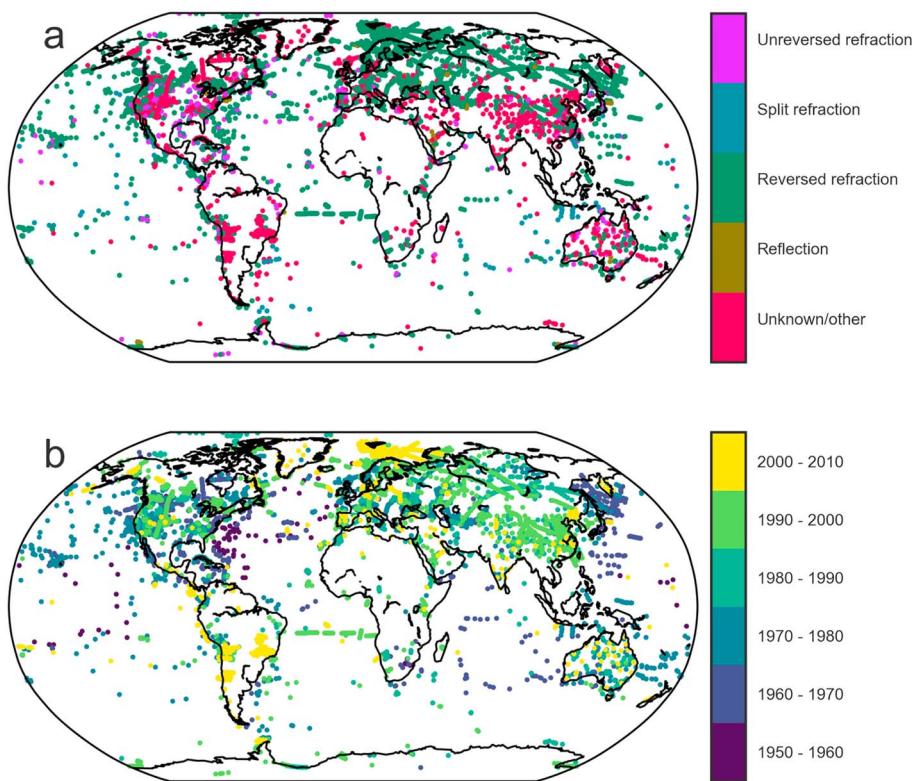


Figure 1. Moho depth points from the U.S. Geological Survey Global Seismic Catalog database. (a) Seismic method used to determine Moho depth and P wave velocity. (b) Year of publication of results. Unknown/other contains points derived from earthquake models, waveform models, tomography, sonobuoy measurements, and the time term method.

Continental-scale models that use variations of this technique have also been developed (e.g., Chulick et al., 2013; Stolk et al., 2013; Tesauro et al., 2014).

A major drawback of including predefined domains is that the uncertainty of the resulting crustal model is hard, if not impossible, to estimate. In fact, the plausible appearance of the model might be misleading in areas with sparse data coverage. Furthermore, the boundaries of the domains might be uncertain as well and contribute to total crustal model uncertainty.

In this work, we use a geostatistical interpolation approach to derive a crustal model and quantitative estimates of its uncertainty from a database of seismic points (from the U.S. Geological Survey [USGS] Global Seismic Catalog [GSC] database). The only additional data are the age of the ocean floor, which is needed to separate continental and oceanic domains. Since we do not include other additional data sets (e.g., gravity anomalies), we expect our results to be ambiguous in regions where data coverage is nil or sparse. However, we do not consider this a drawback of the method but an illustration of the limitations of the current database. The interpolation itself is a form of nonstationary kriging, which allows lateral variation of covariance parameters expected on a global scale (Risser & Calder, 2017).

We apply this method to the latest version of the GSC collected by the USGS (Mooney, 2015). The GSC contains mainly active seismic measurements. However, the method is completely general and can be applied to any compilation of seismic points. We obtain global estimates of the depth to the Moho boundary and average crustal P wave velocity on a global grid with 1° spacing. Estimates of crustal model uncertainty can be helpful for any geophysical application that relies on such models.

To demonstrate some of the advantages of having representative estimates of uncertainty, we apply our results to the estimation of residual topography. Here we consider residual topography as that part of the surface topography which is not explained by crustal structure (Kaban et al., 1999; Molnar et al., 2015). By propagating the uncertainties of the crustal model to residual topography, we can derive a global map of residual topography uncertainty.

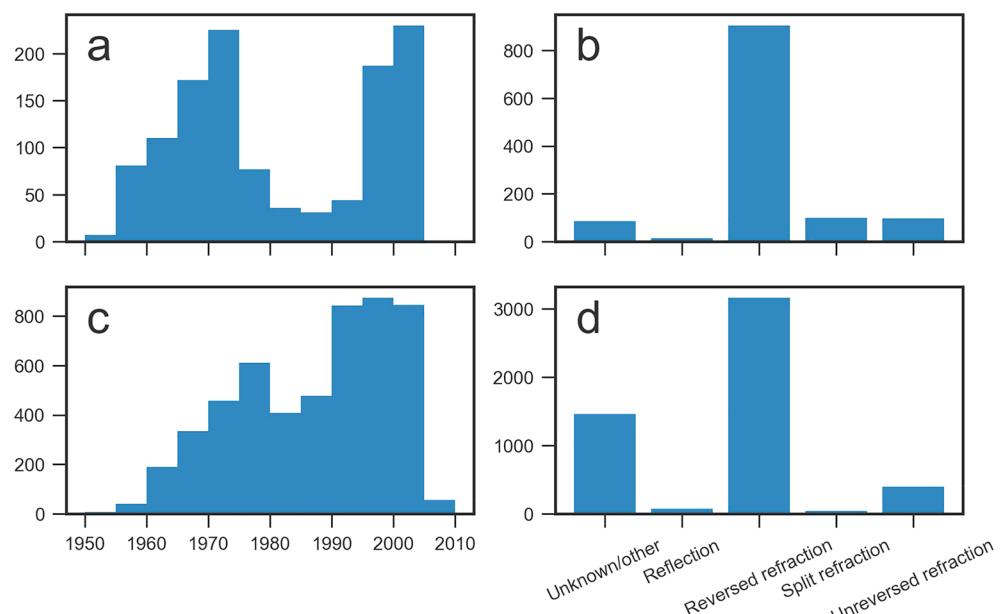


Figure 2. Frequency of used seismic methods and year of acquisition. (a, b) The frequency of acquisition year and seismic method used for oceanic points. (c, d) The frequencies for continental points.

2. Data

2.1. Crustal Database

Each data record in the USGS GSC database provides information about crustal structure at a geographical location based on seismic surveys. From this database we select 6,340 seismic observations that have detected the Moho boundary. The vertical structure at each point is described as layers of different types (sediments, crust, or mantle) and their associated thickness and seismic velocities. Note that information about *S* wave velocities is very sparse, and hence, we limit ourselves to *P* wave velocities. In addition, the type of survey (mostly refraction) and the year when the results were published are given (Figure 1).

About 75% of oceanic and 60% of continental points are reversed refraction surveys (Figure 2), which are generally considered to be the most reliable method (Mooney et al., 1998). About 7% of the oceanic and 28% of the continental points are from Unknown/Other methods. This encompasses some passive seismological methods like earthquake models, earthquake tomography, and/or waveform models.

The data were acquired over a period from 1950 to 2005. The distribution for oceanic points shows two distinct maxima in the 1970s and after 1995. There are relatively few data points acquired between 1980 and 1995. The distribution is more uniform for continental points and also shows an increase in the 1970s and late 1990s (Figure 2).

Many crustal models give velocities and thicknesses separately for the upper, middle and lower crust. Interpolating this layering from the seismic points is challenging, because the number of layers is not consistent over all points. Thus, it is not clear how the layers at different points relate to each other. To circumvent this issue, we interpolate the average velocity of the crystalline crust directly.

The average velocity of the crystalline crust is calculated as the weighted average

$$\bar{v} = \frac{\sum_i v_i t_i}{\sum_i t_i}$$

where t_i is the thickness of layer i and v_i is its velocity.

After calculating the average velocities, we find that the database contained some points with very low average crustal velocities. At these points, low-velocity sediments are incorrectly classified as part of the crystalline crust in the database. Following Chulick et al. (2013), all points with average velocities of 5.8 km/s or less are excluded from the interpolation. In addition, we remove all oceanic points with average velocities

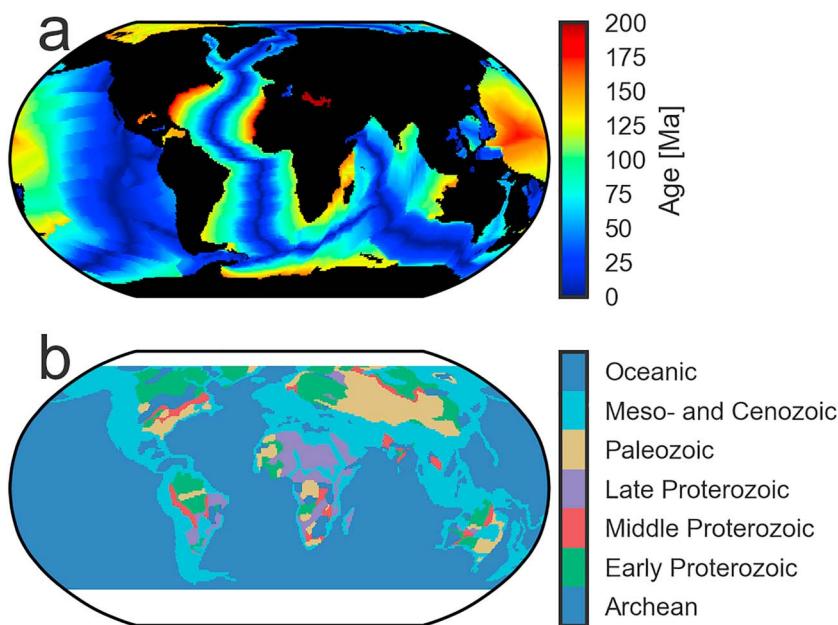


Figure 3. (a) Ocean floor age (Müller et al., 2008). (b) Tectonothermal/ Tectonic age categories from Goutorbe et al. (2011). Note that the tectonic age map stops at 60° latitude.

≤ 6.4 km/s. The oceanic crust consists mainly of gabbroic rocks with velocities ≥ 6.6 km/s, underlying a slower (< 5 km/s) layer of basalts and sheeted dikes (White et al., 1992). Therefore, points with anomalously low velocities likely represent local features such as fracture zones that are not representative of large-scale oceanic crust. After removing these anomalous sites, 5,197 points are retained.

2.2. Ocean Floor Age Map

During the interpolation procedure, oceans and continents are treated separately. This is necessary due to their different nature and evolution. The thickness and velocity structure of unaltered oceanic crust mainly depends on the conditions at the mid-ocean ridge where it was formed (cf. McKenzie & Bickle, 1988). By contrast, the properties of continental crust are the result of a much longer evolution. Accordingly, continental and oceanic domains are defined based on the ocean floor age (Müller et al., 2008; Figure 3a). A point where an ocean floor age is defined is considered oceanic, whereas a point with no ocean floor age is considered continental. In total there are 554 oceanic and 4643 continental points.

This binary ocean-continent classification results in a sharp transition between these domains at the continental margin. For simplicity, we use the binary ocean-continent transition, but we acknowledge that the real crustal structure is certainly more complicated in the transitional region.

The ocean-continent classification employed here is a simplified version of the tectonic classification used in other crustal models. For example, the thermotectonic age map from Mooney et al. (1998; as digitized by Goutorbe et al., 2011) introduces additional spatial structures in the continental domains (Figure 3b). At the same time, the average crustal thickness for the different continental ages is quite similar, ranging from ~36 to 41 km (Figure 4). By comparison, the standard deviation within age categories is more than 6 km. Thus, giving up tectonic regionalization leads only to an acceptable loss of information.

3. Geostatistical Analysis and Interpolation

We use a nonstationary kriging method based on Gaussian fields to interpolate the Moho depth and average P wave velocity. Let $Z(x)$ be the spatial field of interest (either the Moho depth or average crustal P wave velocity), where x is the two-dimensional location in terms of geographical coordinates. Z is separated into three parts (Cressie, 2015)

$$Z(x) = \mu(x) + Y(x) + \epsilon(x)$$

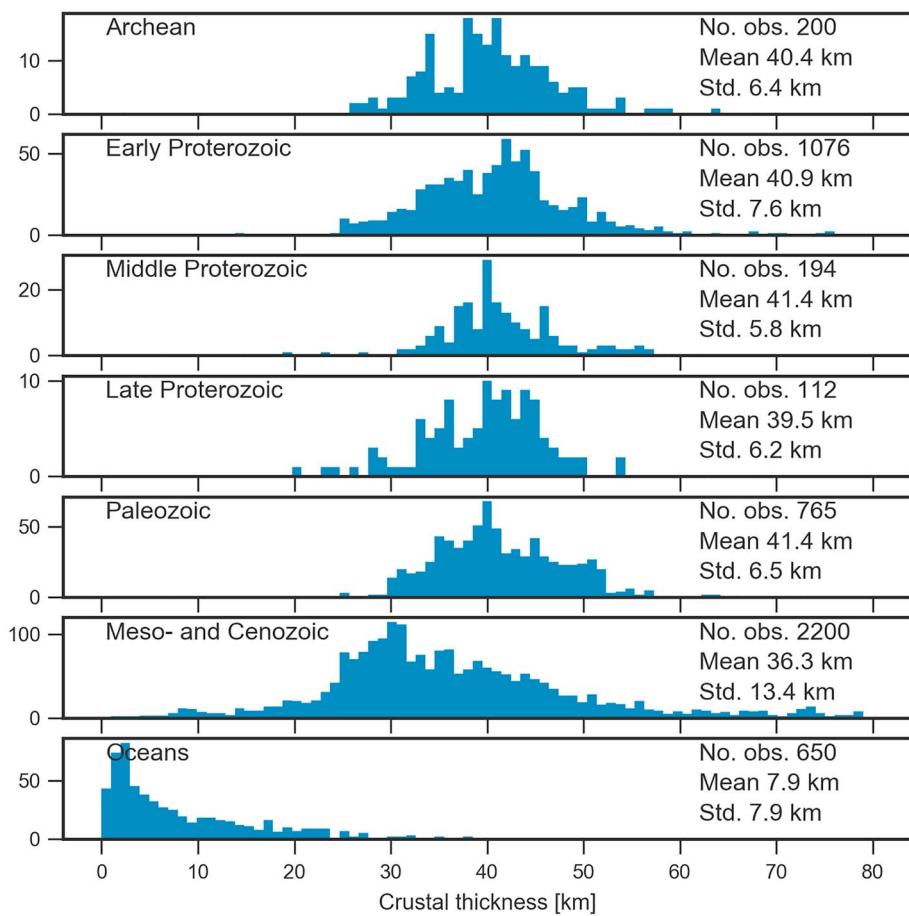


Figure 4. Crustal thickness of points from the Global Seismic Catalog database falling in each of the thermotectonic age categories. Note that we only use points that lie inside the thermotectonic age map of Goutorbe et al. (2011).

Here μ is a deterministic component (trend), which describes the first-order spatial structure, Y is a Gaussian field, and ϵ describes the spatially independent errors, which are also assumed to be Gaussian.

The trend component μ depends only on the classification as oceanic or continental

$$\mu(x) = \begin{cases} \mu_{\text{ocean}} & \text{if } x \text{ in ocean domain} \\ \mu_{\text{continent}} & \text{if } x \text{ in continent domain} \end{cases}$$

The Gaussian field Y has zero mean and is described by a covariance function $C(x_1, x_2; \vec{\theta})$, which in turn depends on a set of parameters $\vec{\theta}$. Often, it is assumed that C is stationary, which implies that C only depends on the distance between x_1 and x_2 . $\text{Cov}(x_1, x_2) = C(x_1, x_2) = C_s(d)$, where $d = \|x_1 - x_2\|$ denotes the **great-circle distance** between the points x_1 and x_2 .

Equivalently, the covariance structure can be described by the semivariogram function $\gamma(d)$, which is given by $\gamma(d) = C(0) - C(d)$, where $C(0)$ denotes the variance of the quantity in question.

The covariance structure is typically described by at least three parameters: the nugget τ^2 , the sill σ^2 , and the range ρ . The nugget is often due to unsampled small-scale variability or noise in the data. In our case, the nugget effect is contained in ϵ . The sill describes the overall variance of the quantity of interest, and the range is a measure of correlation distance. Points that are more distant than the range are assumed to be uncorrelated (or very nearly).

We chose to use a spherical covariance function, because it has a number of beneficial properties. It is an admissible covariance function on the sphere using great-circle distances (Guinness & Fuentes, 2016), it does not impose differentiability on the interpolated field, and it is computationally efficient, because it gives exactly zero covariance for points which are more separated than the range, which allows us to use sparse

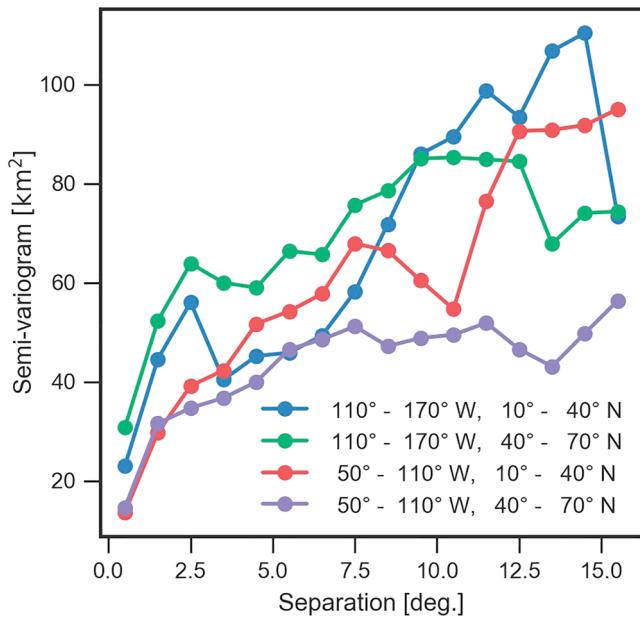


Figure 5. Empirical semivariogram of Moho depth calculated of four quadrants of North America. The bin size for estimating the semivariogram is 1° great-circle distance. The four quadrants show differences in terms of nugget, sill, and range.

covariance matrices. The covariance function is given by Cressie (2015)

$$C_s(d; \tau^2, \sigma^2, \rho) = \begin{cases} \sigma^2 \left(1 - \frac{3d}{2\rho} + \frac{1}{2} \left(\frac{d}{\rho} \right)^3 \right) + \delta_{d,0} \tau^2 & \text{if } d < \rho \\ 0 & \text{if } d \geq \rho \end{cases} \quad (1)$$

Here $\delta_{d,0}$ is zero unless $d = 0$.

3.1. Nonstationarity

Stationarity might be a poor approximation, because the parameters of the covariance function represent a measure of roughness and scale of variability (Cressie, 2015). Thus, it is likely that different geologic regions are characterized by different parameters of the covariance function, and their variation needs to be considered.

To test how much the semivariogram properties vary spatially, we calculate semivariograms for different portions of North America. We use North America, because the database contains enough points to permit calculating informative semivariograms even for smaller subsets of data. The North American continent is split into four quadrants, where each quadrant has a size of $60^\circ \times 30^\circ$. Inside each quadrant, we estimate an empirical semivariogram using only continental points (Figure 5). We find pronounced differences between the quadrants, which motivates us to use a nonstationary approach.

In the nonstationary case, the covariance function explicitly depends on the spatial locations x_1 and x_2 . One way to derive a nonstationary covariance function from a stationary covariance function is to allow τ^2 , σ^2 , and ρ to vary spatially (cf. Paciorek & Schervish, 2004). They are then given by three functions $\tau^2(x)$, $\sigma^2(x)$, and $\rho(x)$, which leads to a nonstationary covariance function

$$C^{NS}(x_1, x_2; \theta) = \frac{\rho_{\text{eff}}}{\sqrt{\rho(x_1)\rho(x_2)}} C_s(|x_1 - x_2|; \tau(x_1)^2, \sqrt{\sigma(x_1)\sigma(x_2)}, \rho_{\text{eff}}) \quad (2)$$

where

$$\rho_{\text{eff}} = \sqrt{2} \frac{\rho(x_1)\rho(x_2)}{\sqrt{\rho(x_1)^2 + \rho(x_2)^2}}$$

is the effective range between the points x_1 and x_2 .

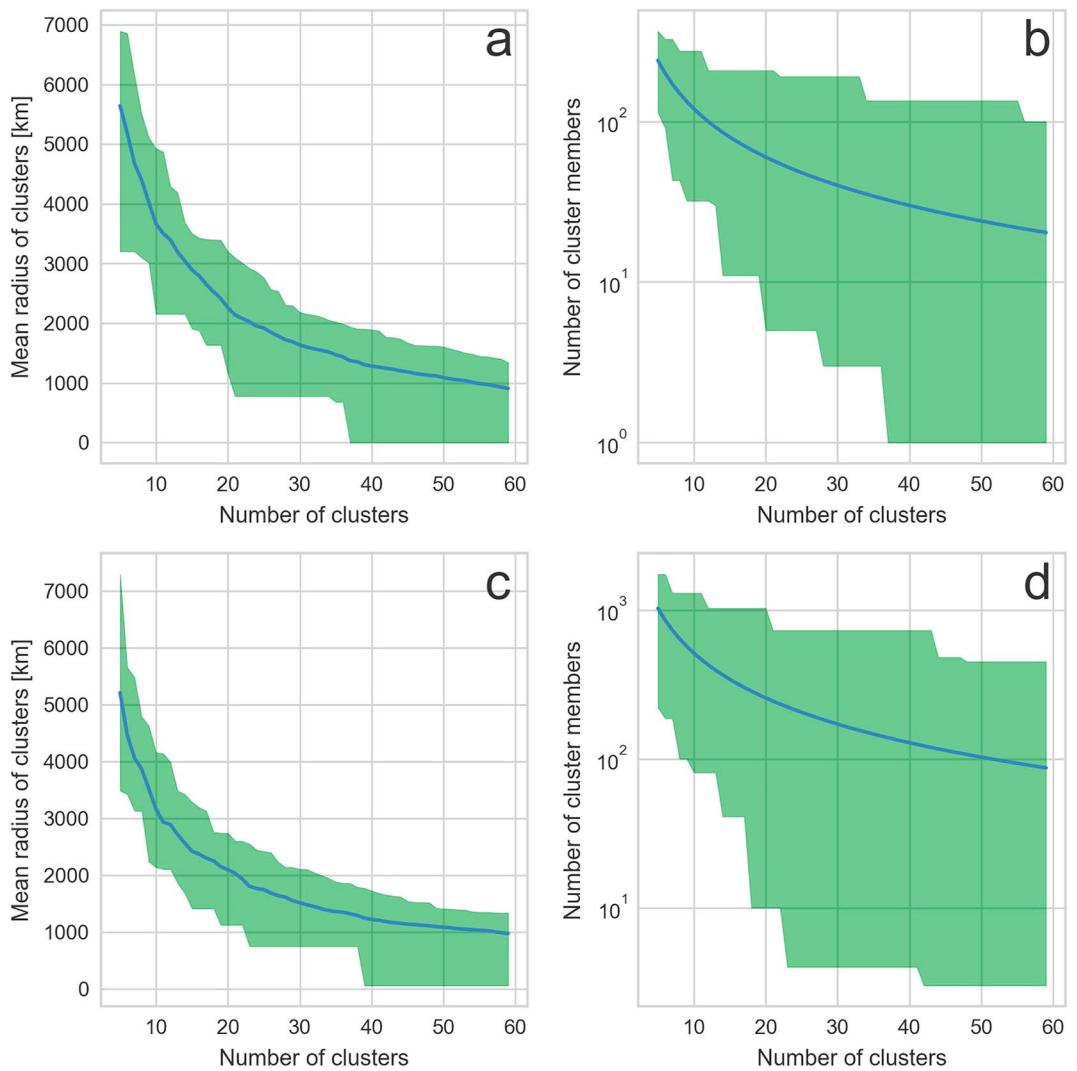


Figure 6. Results of clustering. Radius of clusters (a, c) and number of points in each cluster (b, d). The shaded area gives the range over all clusters, while the blue line gives the mean. Oceanic areas (a, b) were treated separately from continents (c, d). For oceans we selected 35 clusters, because this prevents single point clusters from appearing (b) and gives a mean radius of about 1,100 km (corresponding to 10°). For the continents we used 50 clusters, because this gives the desired mean radius of 1,100 km.

To estimate the functions τ^2 , σ^2 , and ρ from data points, the functions need to be given a discrete representation. We use the mixture approach of Risser and Calder (2017), in which K master locations b_k with corresponding parameters τ_k^2 , σ_k^2 , and ρ_k are defined. The parameter at an arbitrary location is then found as a weighted sum of the mixture components $\sigma(x)^2 = \sum_{k=1}^K w_k(x)\sigma_k^2$, where each weight function w_k decreases exponentially with great-circle distance from the master location

$$w_k(x) \propto \exp \frac{-||x - b_k||^2}{2\lambda_w}$$

under the condition that $\sum_{k=1}^K w_k(x) = 1$.

The parameter λ_w controls the behavior of the interpolation. If λ_w is small, the interpolation effectively becomes nearest neighbor interpolation, whereas large values of λ_w lead to a smooth distribution of the parameter. Thus, the parameter needs to be adjusted to the scale of the domain in question and the desired interpolation behavior.

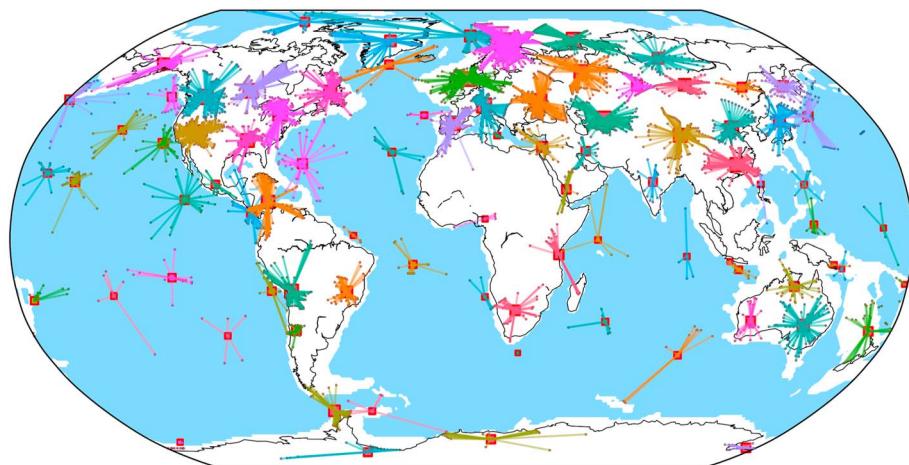


Figure 7. Results of clustering for oceanic and continental points. Small dots mark the location of points from the database. Colors indicate to which cluster each point belongs. The centroid of each cluster is indicated as a larger square, where the size is proportional to the number of points contained in each cluster. The lines connect each point with the centroid of the cluster to which it belongs.

We chose the discrete mixture approach, because it is directly applicable to a spherical geometry and can effectively adapt to a very heterogeneous distribution of points. Furthermore, the locations of the master locations can be determined using clustering techniques semiautomatically, which means no additional information is needed.

3.1.1. Clustering Analysis

Let the value of Z be known at N locations x_i . We first construct the master locations using agglomerative clustering of the points based on their great-circle distance (see Appendix A for algorithm details). The clustering will then be used to estimate the nugget, sill, and range from the data. How the points are assigned to clusters is controlled by the choice of the number of clusters to create. Here we separately cluster oceanic and continental points.

We tested different numbers of clusters to optimize the spatial radius and the number of points in each cluster. The radius of a cluster is defined here as half the maximum distance of any two points that are a member of that cluster. We consider the following factors to choose the number of clusters. First, the radii of the clusters should be large enough to detect the long-range correlation between points. Second, there should be enough clusters to detect spatial variations of the covariance parameters. Third, the radius of all clusters should be similar, because otherwise the clusters might appear to possess very different ranges ρ , even though this is just a result of the different cluster radius. Lastly, each cluster should contain enough points to reliably estimate the covariance parameters. Fulfilling these conditions is not possible everywhere, and some compromise needs to be made.

After numerous tests, we decided to use 35 oceanic and 50 continental clusters. For the oceans this is mainly motivated by preventing clusters containing only a single point (Figure 6). For both oceanic and continental domains, we try to choose the number of clusters so that the mean cluster radius is ca. 10° (ca. 1,100 km). The interpolation parameter λ_w is set to $(1,100 \text{ km})^2$, which is the squared mean cluster radius.

The master locations b_k are defined as the spherical average location of all members of each cluster. This is calculated by converting the location of the points to an Earth Centered Earth Fixed Cartesian coordinate system, taking the average location of all points in 3-D space and then projecting back to the surface of the Earth. This avoids issues that arise when taking the average of points with positive and negative longitude or near the poles. The spatial structure of the clusters (Figure 7) is reasonably adapted to the heterogeneous distribution of the data points.

3.1.2. Inference of Covariance Parameters

The nonstationary covariance function depends on the covariance parameters at each master location. These parameters have to be estimated from the points of the database. By assuming that the Gaussian field Y is

Table 1
Parameters of the Prior Covariance Parameters

Parameter	Moho		<i>P</i> wave velocity	
	Mean	Standard deviation	Mean	Standard deviation
(cont.)				
Nugget τ^2	(4.0 km) ²	(4.0 km) ²	(0.2 km/s) ²	(0.2 km/s) ²
Sill σ^2	(80.0 km) ²	(80.0 km) ²	(0.2 km/s) ²	(0.2 km/s) ²
Range ρ	10°	10°	10°	10°
Parameter (ocean)				
Nugget τ^2	(1.0 km) ²	(3.0 km) ²	(0.2 km/s) ²	(0.2 km/s) ²
Sill σ^2	(40.0 km) ²	(40.0 km) ²	(0.2 km/s) ²	(0.2 km/s) ²
Range ρ	10°	10°	10°	10°

Note. We assign different parameters to continental and oceanic points for the Moho depth to account for the overall much less variable nature of Moho depth in the oceans. The parameters of the velocity prior are identical for ocean and continent.

stationary in the neighborhood of each master location b_k , the parameters can be efficiently estimated by using the local maximum likelihood estimation technique (Tibshirani & Hastie, 1987).

Let \hat{x}_i be the points belonging to an arbitrary cluster k and let $\hat{Z}_i = Z(\hat{x}_i)$ be the values at each location. Since Z is a Gaussian field, \hat{Z}_i are elements of a multivariate normal distribution with covariance matrix Σ and mean μ . The entries of the covariance matrix are given by evaluating the stationary covariance function $C_s(d; \tau_k^2, \sigma_k^2, \rho_k)$, for all pairs of points inside the cluster (equation (4)).

The mean value inside each cluster is estimated using the sample mean of all cluster members. Our method is therefore slightly biased, because the Z_i are correlated.

The log-likelihood function is the probability density function of a multivariate Gaussian distribution with constant mean μ and covariance matrix Σ :

$$\log L(\vec{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \det \Sigma(\vec{\theta}_k) - \frac{1}{2} (\vec{Z} - \mu)^T \Sigma^{-1}(\vec{\theta}_k) (\vec{Z} - \mu),$$

where n is the number of points in \vec{Z} .

We numerically maximize the value of the log-likelihood to obtain $\vec{\theta}$ using the L-BFGS-B algorithm (Byrd et al., 1995; Zhu et al., 1997) as implemented in the scipy package (Jones & Oliphant, 2001).

During first test runs, where we maximized the above likelihood function directly, the nugget and sill values were systematically very small in regions with few data points. This would lead to overly optimistic uncertainty estimates in regions with little data points. To somewhat alleviate this issue, we use a Bayesian approach and put independent inverse gamma priors on the σ_k^2 , τ_k^2 , and ρ_k . The inverse gamma distribution has a log probability density function given by

$$\log f(x; \alpha, \beta) \propto -(\alpha + 1) \log x - \frac{x}{\beta}$$

The mean and standard deviation are, respectively, Mean = $\frac{\beta}{\alpha-1}$ Std.Dev. = $\frac{\beta}{(\alpha-1)\sqrt{\alpha-2}}$

The hyperparameters α and β were chosen for each covariance parameter such that the prior distribution agrees with a specified mean and standard deviation as given in Table 1. Note that we set the standard deviation of all properties to be at least as large as the mean value; thus, the prior is relatively weak and only has an effect on clusters with very little data points. The nugget parameter is chosen according to previous accuracy estimates of active seismic studies (Christensen & Mooney, 1995; Mooney et al., 1998). The sill parameter choice is guided by the standard deviation of Moho depth and *P* wave velocity in the database. For the range, we use an average value of 10° in agreement with the mean size of the clusters.

The modified posteriori likelihood becomes, after application of Bayes theorem (Tarantola, 2005): $\log \tilde{L}(\theta) = \log L(\theta) + \sum_{\varphi \in \{\tau_k^2, \sigma_k^2, \rho_k\}} \log f(\varphi; \alpha_\varphi, \beta_\varphi)$

$\tilde{L}(\theta)$ is optimized using the L-BFGS-B algorithm as above.

3.1.3. Interpolation and Prediction

After the sill, nugget, and range belonging to all clusters have been determined, the value of Z can be predicted at an arbitrary set of locations x_j^* . Let \vec{Z} be the vector of known values and let \vec{Z}^* be the vector of unknown values at locations x_j^* . The complete covariance between two arbitrary points x_i and x_j is found by combining equations (4) and (2)

$$C(x_i, x_j; \vec{\theta}) = \begin{cases} \sigma(x_i)\sigma(x_j) \frac{\rho_{\text{eff}}}{\sqrt{\rho(x_i)\rho(x_j)}} \left(1 - \frac{3||x_i - x_j||}{2\rho_{\text{eff}}} + \frac{1}{2} \left(\frac{||x_i - x_j||}{\rho_{\text{eff}}} \right)^3 \right) + \delta_{ij}\tau(x_i)^2 & \text{if } d < \rho_{\text{eff}} \\ 0 & \text{if } d \geq \rho_{\text{eff}} \end{cases}$$

Evaluating the equation above for all pairs of points (both observation and prediction points) gives the complete covariance matrix. The complete covariance matrix naturally separates into three block matrices: Σ_{ZZ} is the covariance between the observed points, Σ_{ZZ^*} is the cross-covariance between the observed and prediction locations, and $\Sigma_{Z^*Z^*}$ is the covariance between prediction locations.

The prediction of Z^* is then given by (Risser & Calder, 2017)

$$\vec{Z}^* = \mu(x^*) + \Sigma_{ZZ^*} \Sigma_{ZZ}^{-1} (\vec{Z} - \vec{\mu})$$

where $\vec{\mu}$ contains the mean value at each of the prediction locations, depending on its classification as oceanic or continental.

The posteriori covariance matrix of the prediction error at all prediction locations is

$$\Sigma_{Z^*|Z} = \Sigma_{Z^*Z^*} - \Sigma_{ZZ^*} \Sigma_{ZZ}^{-1} \Sigma_{Z^*Z} \quad (3)$$

The diagonal of $\Sigma_{Z^*|Z}$ contains the error variances, that is, the square of the estimated uncertainty. The off-diagonal entries of $\Sigma_{Z^*|Z}$ correspond to the error covariances at different locations.

3.1.4. Outlier Removal and Duplicate Selection

We test all points for consistency with neighboring points using a hold-one-out cross-validation approach similar to Stolk et al. (2013). The quantity of interest is predicted at each observation location but without using that observation itself. If the difference between the measured and predicted value is larger than twice the estimated interpolation error and the difference is larger than 5 km (Moho depth) or 0.2 km/s (P wave velocity), it is flagged as an outlier and not used in prediction. In addition, there are some locations with duplicate values. Here we predict the value at the location of the duplicates, without using the duplicate values, and choose the value that is closest to the interpolated value at that location.

The estimation of parameters σ_k^2 , τ_k^2 , and ρ_k is then repeated using only the chosen points. Next, a second round of outlier detection follows, which only uses the points not flagged as outliers and the chosen duplicate points. This gives the final set of points for the interpolation.

Apart from detecting outliers, the hold-one-out procedure also provides insights into the accuracy of the data set and our method. The difference between an observation z_j and the prediction \hat{z}_j if that observation is held out is called the *prediction error*. The root-mean-squared prediction error gives an estimate of the overall accuracy of the prediction procedure and the internal consistency of the database

$$RMSPE = \sqrt{\frac{1}{N} \sum_{j=1}^N (z_j - \hat{z}_j)^2}$$

Furthermore, we can also test the error estimation. Specifically, the prediction error should be similar to the estimated error (from equation (3)). To quantify this, we use the deviation ratio

$$\text{DR}_j = \left| \frac{z_j - \hat{z}_j}{\hat{\sigma}_j} \right|$$

where $\hat{\sigma}_j$ is the estimated error corresponding to \hat{z}_j . Thus, the optimal value of the deviation ratio should be one. If the deviation ratio is larger than 1, the error has been underestimated, whereas values below 1 indicate an overestimation of error.

Note that under the assumption of normality, the deviation ratio should follow a folded normal distribution (Leone et al., 1961) with standard deviation one. If this assumption holds, the median deviation ratio should be $\frac{\operatorname{erf}^{-1}(\frac{1}{2})}{\sqrt{2}} \approx 0.67$. Around 68% of the points should have a deviation ratio of 1 or less (i.e., the error is overestimated), and 95% of the points should have a deviation ratio of less than 2.

4. Results

4.1. Outlier Detection

Applying the hold-one-out cross-validation technique to the database, 310 points were flagged as outliers for interpolating the Moho depth. Another 321 points were removed due to the selection of duplicates. This leaves 4,566 points for the interpolation of Moho depth, corresponding to about 14% removed points. Applying the same approach to P wave velocity leads to a similar number of removed points: 255 points as outliers and 316 due to the selection of duplicate points.

4.2. Covariance Parameters

The spatial distribution of the three covariance parameters nugget, sill, and range has a strong impact on the behavior of the interpolation, particularly the estimated uncertainties. The nugget τ^2 at a point determines the minimum uncertainty at that location, and the maximum uncertainty tends toward the sum of the nugget and the sill $\tau^2 + \sigma^2$. The range ρ controls how quickly the uncertainty increases with distance from the nearest prediction point. Moreover, larger values of ρ lead to a smoother interpolated field (Chiles & Delfiner, 1999).

The estimated covariance parameters of the Moho depth points display a strong dichotomy between oceans and continents (Figure 8). Nuggets and sills are higher in the continents than in the oceans, whereas the range tends to be higher in the oceans. This is due to the relatively uniform crustal thickness in the oceans, which does not change significantly, unless affected by magmatism (e.g., White et al., 1992).

The nugget and sill in the oceans are relatively uniform with values of around 1 and 9 km², respectively. The range in the oceans is more variable. It is more than 20° in the Western Pacific and Northern Atlantic but can be less than 5° in the Southern Atlantic and Indian Ocean. Overall, the range tends to be higher in well-studied regions compared to more sparsely surveyed areas. This suggests that the observed differences do not reflect differences in oceanic crustal structure but differences in sampling density.

The continents are also characterized by relatively uniform nugget, sill, and range. The most notable anomaly is the high sill over the Andes, with values up to (12 km)². Note that other major orogens such as the Himalaya do not show this increased sill. In addition, the Andes have a reduced range. The different covariance parameters in the Andes might be an artifact caused by the large north-south extent of the Andes. Thus, in north-south directions the sill should be less than in east-west directions and vice versa for the range. It might hence be possible that the anomalous values in the Andes are the result of the inability of an isotropic covariance function to describe an anisotropic structure. As a result of the high sill, the uncertainty estimates in South America might be too high.

The covariance function parameters of the average P wave velocity differ substantially from those of the Moho depth (Figure 9). Over both continents and oceans, sill and nugget are often of the same magnitude. This implies that small-scale variations or measurement errors are of similar importance as large-scale structural features. As a result, direct interpolation of the velocity might not be possible in all regions.

Over the continents nugget and sill are relatively uniform. Exceptions are Central Eurasia, Northern Europe, parts of North America, and Brazil, since these regions are characterized by increased sill of up to (0.3 km/s)², relatively low nugget, and increased range. In these regions large-scale velocity anomalies exist that should be well suited for interpolation.

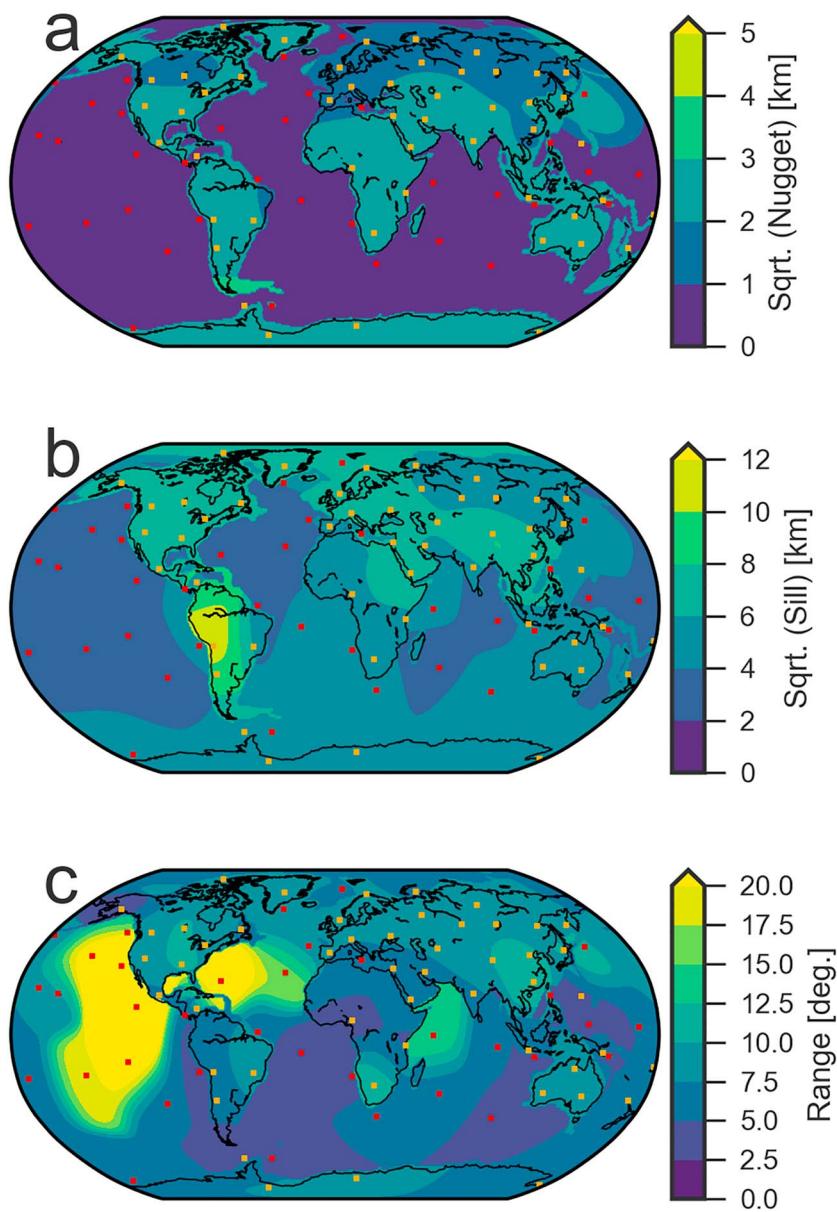


Figure 8. (a–c) Estimated parameter distribution for Moho depth. The red and orange squares give the centers of the clusters used for interpolation in the oceans and continents, respectively. Note that we give the square root of nugget and sill.

By contrast, the oceans possess almost everywhere similar nugget and sill values of about $(0.1 \text{ km/s})^2$ and typical range between 5° and 7.5° . Thus, interpolation of velocity in the oceans will probably not be able to construct an extensive model, especially since the distribution of survey points is in many regions extremely sparse.

4.3. Interpolated Moho Depth and Average P wave Velocity

The interpolated depth of the Moho discontinuity below the geoid (Figure 10a) shows well-known large-scale crustal features. The data are available in supporting information Data Sets S1 and S2. We will discuss the continental features in section 4.5, when comparing with Crust1.0.

The oceans are mostly flat with some single spots corresponding to individual data points. This occurs in areas with few points and estimated low range ρ . Structure is best resolved in the Western Pacific Ocean, where the trend of the pacific ridge is clearly visible. In the other oceans hardly any structure could be interpolated.

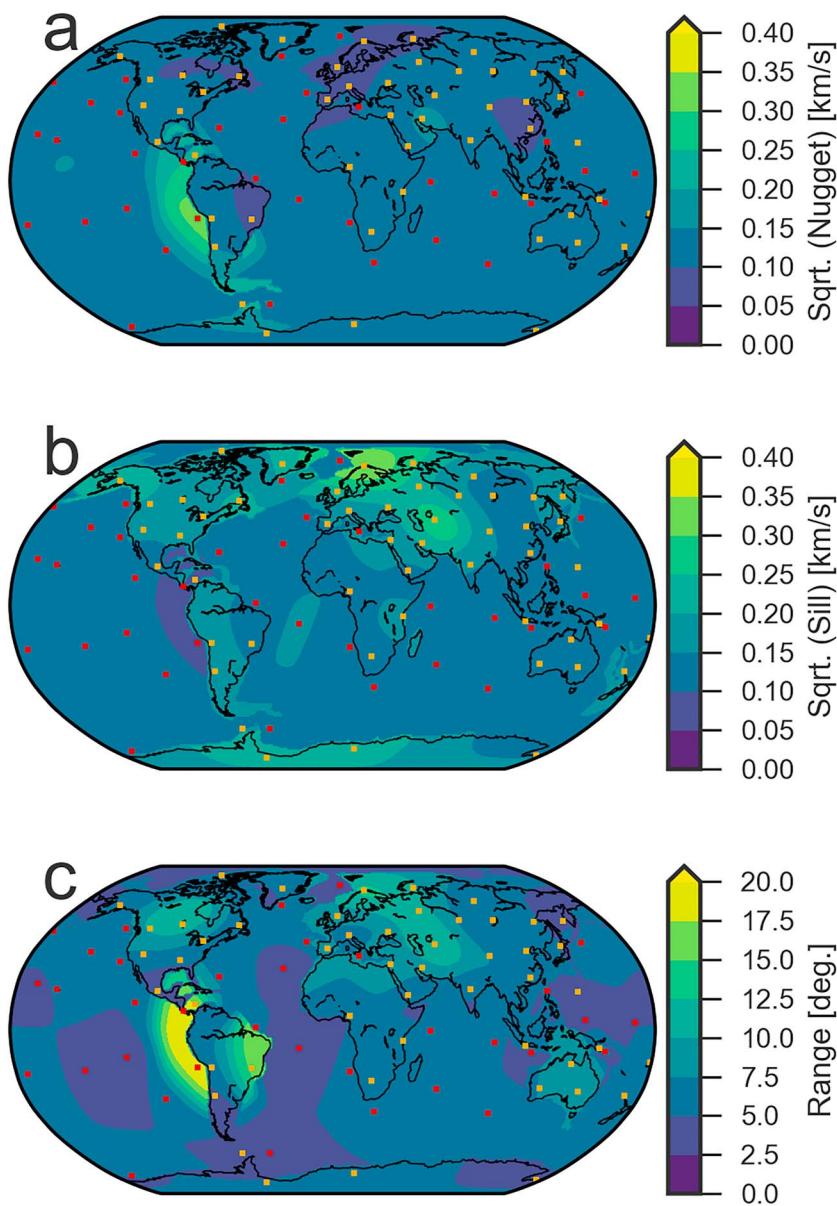


Figure 9. (a–c) Distribution of covariance parameters for the average P wave velocity of the crystalline crust.

Our interpolation method produces 1-sigma estimates of uncertainty due to the underlying assumption of a Gaussian field. Furthermore, the uncertainties measure the internal consistency of the database and cannot describe the external accuracy of the database. The estimated uncertainties range from about 1 km to more than 12 km (Figure 10b). In agreement with the estimated distribution of nugget and sill (section 4.4.2), uncertainties in the oceans are systematically lower than in the continents. Unsurprisingly, uncertainty is negatively correlated with point density. However, the highest uncertainties are estimated for South America, even though Africa contains even less points. This is a result of the extremely high sill value for South America, which might not be realistic. The lowest uncertainties are estimated for North America, most of Eurasia and Australia.

The interpolated P wave velocity of the crystalline crust and its uncertainty is shown in Figure 11. In the oceans hardly any structure is visible in the interpolation results except for some very localized velocity anomalies. The magnitude of these anomalies is typically less than 0.2 km/s, which is comparable to the typical estimated uncertainty. Hence, these localized features are probably of much smaller scale than the resolution of the data points or are the results of velocity inaccuracies.

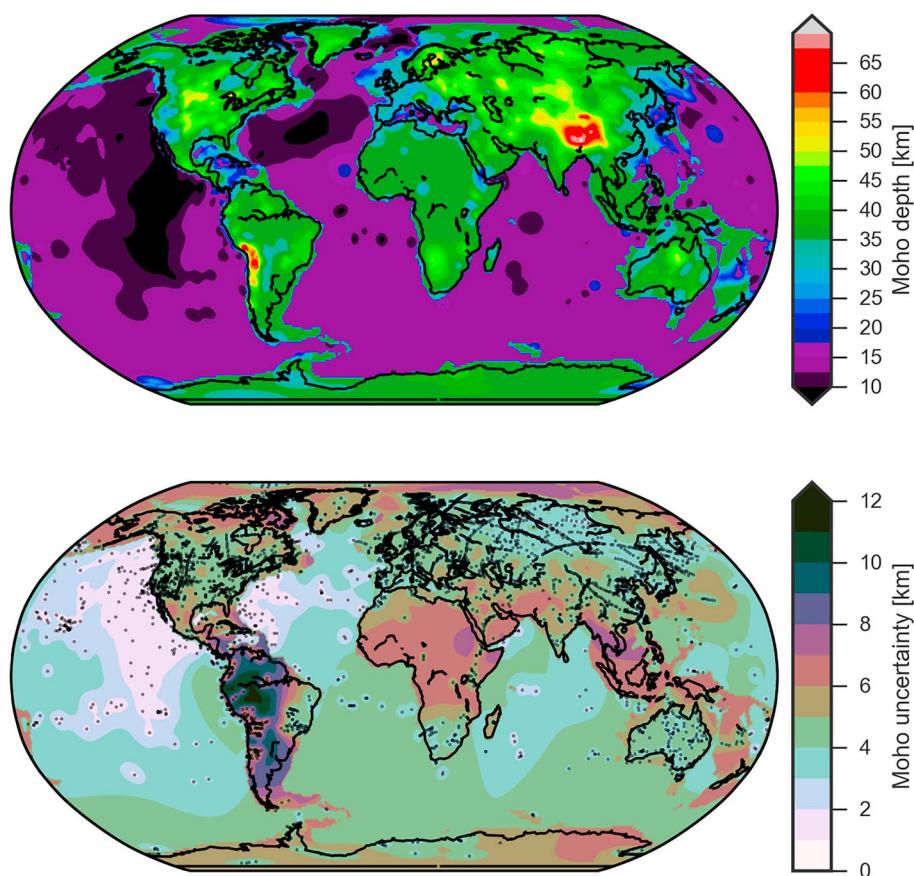


Figure 10. Global map of Moho boundary depth below the geoid and its estimated uncertainty.

4.4. Quality of Error Estimation

4.4.1. Hold-One-Out Cross Validation

The hold-one-out cross validation allows testing the prediction accuracy and error estimate. In the case of Moho depth, the Root-Mean-Square (RMS) prediction errors for oceans and continents were 3.5 and 4.7 km, respectively. The median absolute deviation ratio is 0.53 for oceans and 0.48 for continents, which is less than the theoretically expected value of 0.67. Deviation ratios above one occur for 30% of the oceanic points and 25% of the continental points, which is comparable to the theoretical expectation of 32%. Thus, overall, the error estimate is adequate, albeit slightly too pessimistic.

The prediction error as a function of survey year reflects the increasing quality of instrumentation and processing over time. For continents (Figure 12b), newer surveys are clearly characterized by smaller prediction errors and deviation ratios than older surveys. The prediction error decreases from a median value of 3 km for surveys from the 1950s to less than 2 km for surveys after the year 2000. For surveys acquired in oceanic domains, the prediction error stays more or less constant at about 1 km (Figure 12a). The deviation ratio shows the same trend (Figures 12b and 12d).

Thus, the interpolation results are more consistent with newer—more accurate—data than with older data. Note that the survey year was not explicitly included in the interpolation process in any way, and this is a natural outcome of the geostatistical analysis and interpolation.

Repeating the same analysis for the survey type shows a similar pattern (Figure 13): Reversed refraction profiles have the lowest prediction errors, followed by split refraction and unreversed refraction. Reflection profiles are on average least accurate, followed by the unknown/other category, which is also highly variable. Thus, overall, the most accurate type of survey is most consistent with the interpolation results.

Analyzing the P wave velocity in terms of survey year and type yields the same conclusions: newer surveys systematically have lower prediction errors and deviation ratios. Likewise, reversed refraction profile is the

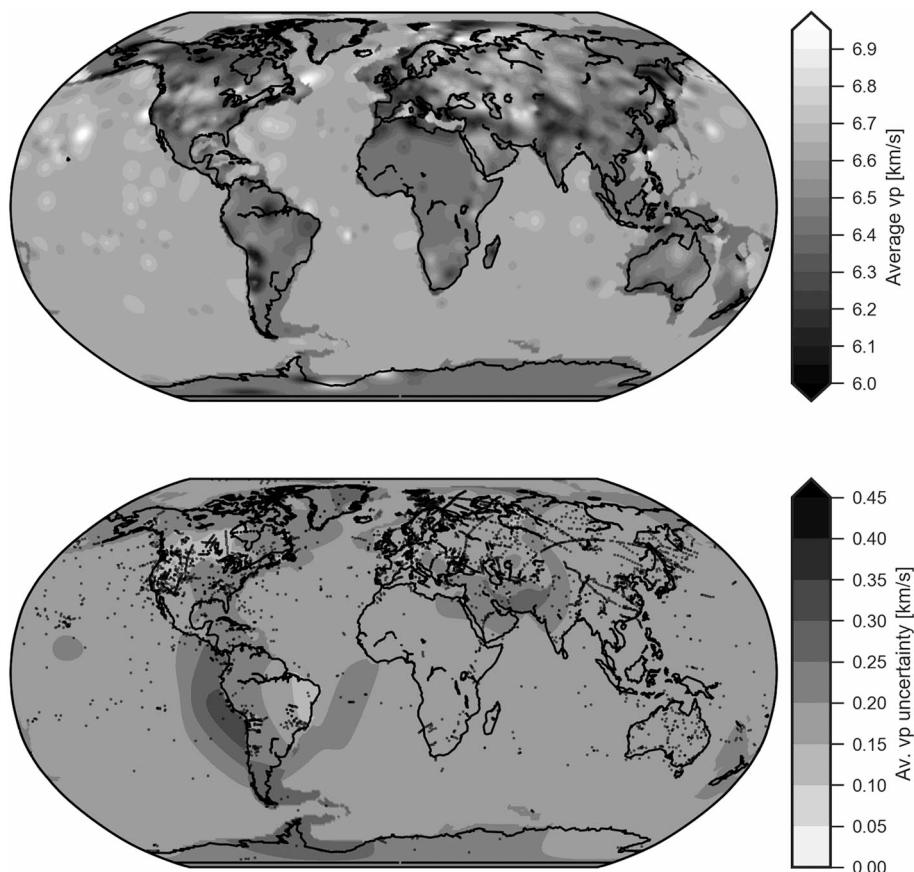


Figure 11. Interpolated distribution of P wave velocity in the crystalline crust.

type of survey with the lowest prediction errors and deviation ratios. The root-mean-squared prediction error is about 0.18 km/s for oceans and continents. The median absolute deviation ratios are 0.58 for oceans and 0.46 for continents. About 75% of the points stay within the estimated error bounds.

Overall, the prediction errors are within the range of reported error estimates for crustal scale seismic profiles, which are about 10% (Christensen & Mooney, 1995). Thus, our interpolation method and the chosen parameters are adequate for reproducing the seismic measurements.

However, holding out only a single observation might lead to optimistic error estimates, due to the very uneven distribution of points. For instance, many points are taken from continuous profiles, so that it should be relatively easy to reconstruct that single point using the neighboring points from the same profile. For this reason we also used a more challenging cross-validation test of the data set.

4.4.2. Median Split Cross Validation

We split the data set into a validation and prediction set. The points from the prediction set will be used to predict the points from the validation set. In this test, the older half of the data is used to predict the Moho depth at the points of the newer half. The split between older and newer half occurs at the median survey year 1989. Most parameters of the interpolation were left unchanged, with the exception of the number of clusters, which had to be reduced to 35 continental and 30 oceanic clusters, in order to prevent single-point clusters and to maintain an average cluster radius of 10° .

The root mean prediction errors is 5.6 km for oceans and 8.1 km for continents, which is roughly twice as large as the prediction errors of the hold-one-out cross validation. While the oceanic error estimates are on average adequate (median deviation ratio of 0.68), the estimated error for continents is systematically too low (median deviation ratio of 0.83). In addition, the distribution of continental deviation ratios becomes skewed toward high values, because about 15% of the continental points have deviation ratios of 2 or more, greatly surpassing the theoretical expectation of 5 %.

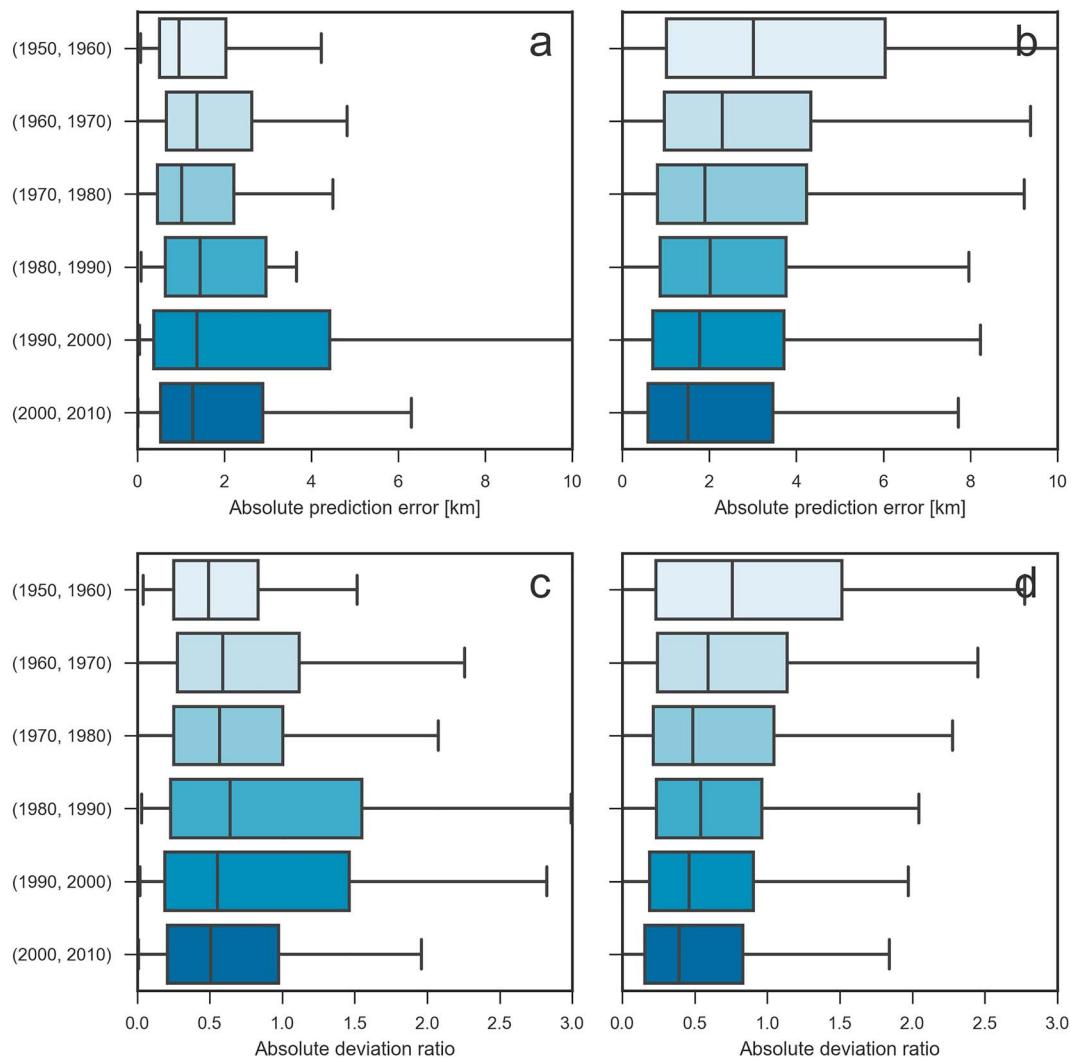


Figure 12. Prediction accuracy for points grouped according to decade in which the survey was carried out. Panels (a) (b) Give the absolute prediction error, while (c) and (d) give the absolute deviation ratio. (a) and (c) refer to oceanic and points, while (b) and (d) refer to continental points. In all figures, the colored boxes cover the 25% to 75% percentile, while the median is marked by a vertical line inside each box. The extent of the whiskers gives 1.75 of the interquartile range relative to the median.

In general, the error estimate can be seen as adequate for this test case, because most points from the validation set stay within the error bounds. However, the interpolation error estimate does not and cannot account for the possible presence of unexpected or rapid changes that do not conform to the underlying assumptions of the interpolation. For example, in sparsely surveyed area, the Moho depth distribution might appear to be more smooth than it actually is. This leads to an underestimation of sill and range and ultimately to an underestimation of uncertainty. This should be considered when interpreting the uncertainties over large poorly covered regions.

4.5. Comparison With Crust1.0

4.5.1. Global

The model Crust1.0 (Laske et al., 2013) is the de facto standard for global crustal model, as evidenced in its widespread use in geophysics. We will focus on Crust1.0 in this paper, but there are also crustal models based on surface wave inversion (e.g., Meier et al., 2007; Pasquano et al., 2014) or satellite gravity inversion (Reguzzoni & Sampietro, 2015). Crust1.0 is also based on a compilation of active seismic surveys; thus, we expect large overlap between our database and that of Crust1.0. On a global scale, our interpolated Moho depth distribution agrees well with the Moho map of Crust1.0 (Laske et al., 2013), with a median absolute

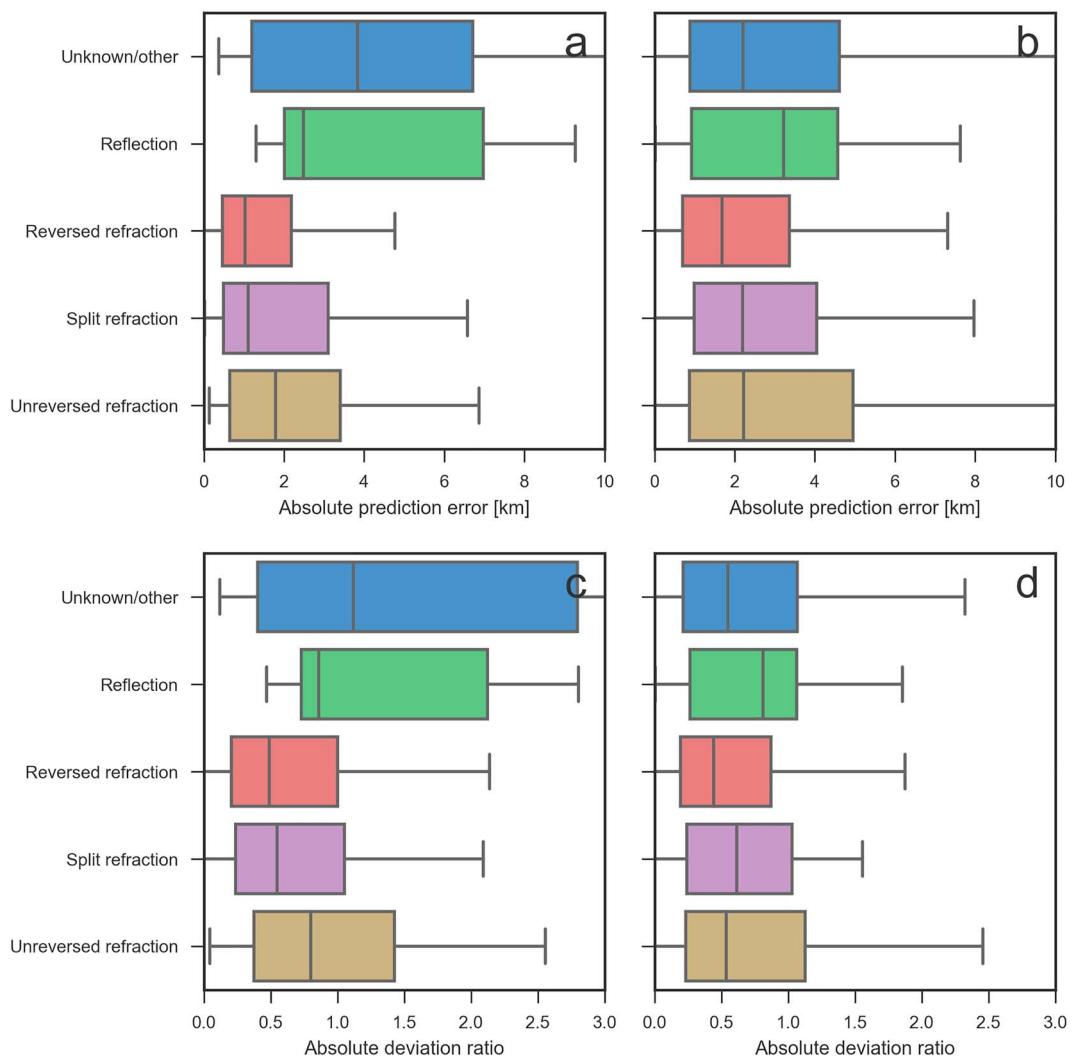


Figure 13. Prediction accuracy grouped according to survey type. Panels (a) and (b) give the absolute prediction error, while (c) and (d) give absolute deviation ratio. (a) and (c) refer to oceanic points, while (b) and (d) correspond to continental points. For an explanation of the figure see caption of Figure 12.

difference of 2.9 km. The spatial pattern of the disagreement between our model and Crust1.0 (Figure 14) shows that the largest differences occur in South America and Asia. While the overall patterns are similar, our results appear less sharp than Crust1.0 due to the tectonic regularization applied in Crust1.0.

In the oceans, Crust1.0 shows significantly more structure, since a constant crustal thickness in the unaltered oceans was assumed and the Moho depth was calculated from ocean floor depth. We have directly interpolated Moho depth and is thus not using the topographic information, leading to a flat Moho in most oceans. The continent-ocean transition is modeled explicitly in Crust1.0 as a smooth transition over the continental margin, whereas we have used a sharp jump from continental to oceanic crust. This leads to a systematic difference between our results and Crust1.0 on the edges of the continents.

Statistical analysis of the differences between our results and Crust1.0 (Figure 15a) shows that the differences between the two models are typically 3 km on all continents except South America, where typical differences are 4.5 km. Importantly, better surveyed continents are not associated with less disagreement between our model and Crust1.0 (i.e., Africa and North America are similarly different). However, our estimated uncertainties (Figure 15b) differ substantially for the different continents and tend to be higher than the average differences between our model and Crust1.0. This suggests that taking the difference between

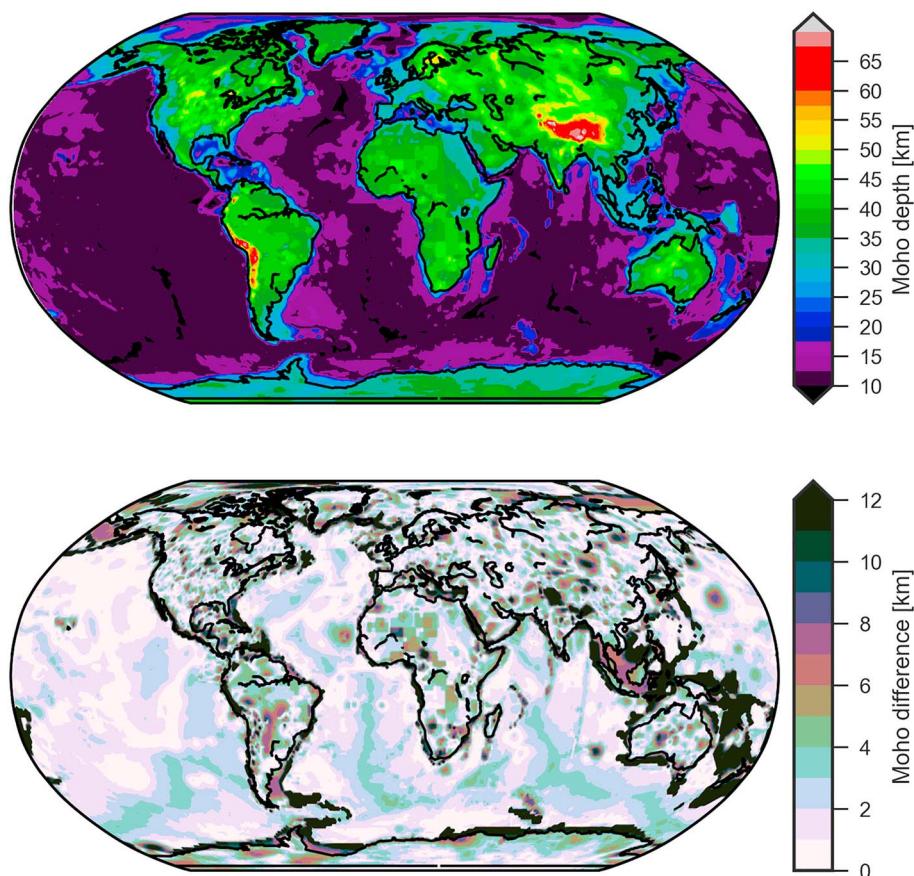


Figure 14. Depth to Moho according to Crust1.0 model and the differences with respect to our model.

different crustal models is an overly optimistic indicator of model uncertainty, if the models are based on similar methods and/or database.

Our interpolation results for average P wave velocity differ from Crust1.0 substantially (see Figure 16). Overall, the scale of velocity variations in Crust1.0 is much less, since velocities are average values for complete domains. In most of the oceans the two models differ mostly by a shift, because our average velocity in the oceans is 6.65 km/s, whereas Crust1.0 uses 6.75 km/s.

In the continents there is qualitative agreement between the two models in many areas. Examples include the separation between slow Western Europe and faster Central Europe visible in both models or the reduced velocity in the Western United States. However, even in areas with qualitative agreement, the results of the direct interpolation are less defined spatially but are characterized by larger velocity variations.

Disagreement between the models can be up to 0.5 km/s and the RMS difference between the models is 0.03 km/s. Thus, on average the models again agree better than expected based on the estimated interpolation uncertainty.

4.5.2. South America

On the global map, South America stands out as the area with the highest uncertainty, reaching values of 12 km or more (Figure 17). This does not seem entirely reasonable, because South America contains much more data points than, for example, Africa.

The uncertainty values of more than 10 km occur only in areas that are not covered by any points. In regions with points, the uncertainty is decreased to realistic values of around 5 km. However, even in the densely studied areas in the Andes, the uncertainty is ~8 km and the thick crust underneath the Andes is not interpolated along the entire mountain chain. There are two explanations for this.

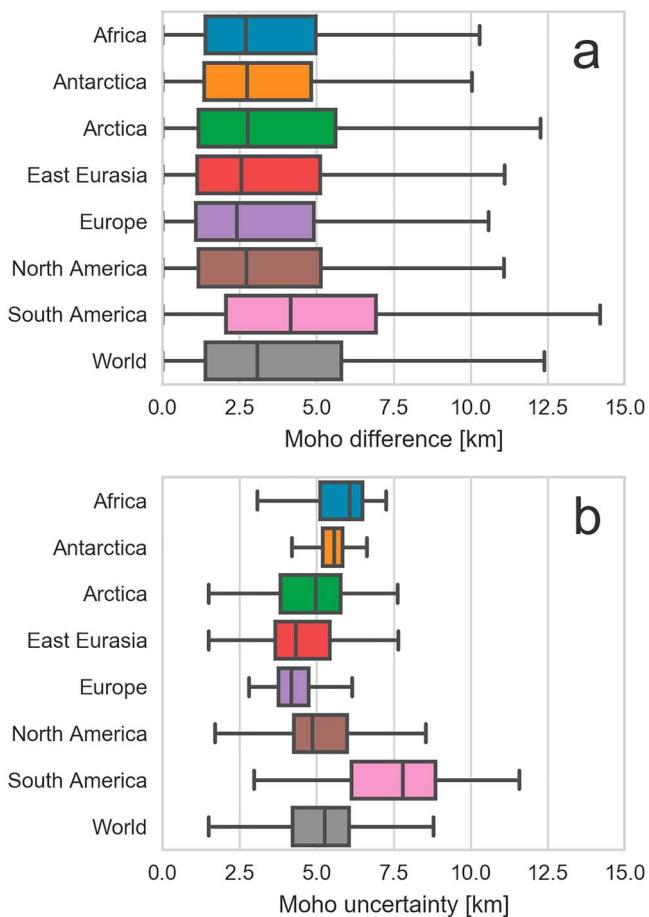


Figure 15. (a) Statistical distribution of differences between our model and Crust1.0. (b) Distribution of uncertainties in our model. The two models are more similar than the uncertainties would suggest, because they rely on the same data source. Note that the frequencies are calculated from the grid points and do not take into account the different sizes of grid cells at higher latitudes. For an explanation of the figure see caption of Figure 12.

First, the Moho depth is extremely variable, even for relatively short distances. The Moho boundary depth goes from up to 70 km in the Andean Cordillera to less than 30 km in the hinterland basins. This leads to high estimated sill values and low estimated range values. Second, Moho depth variations are highly anisotropic, as variations along the orogen (roughly N-S) are much less than perpendicular to the mountain chain (E-W). Thus, an anisotropic covariance function would be required to adequately model the spatial structure in this region.

This is compounded by the fact that the clustering is also essentially isotropic, since it is based on great-circle distances. Thus, points belonging to the high Andes are grouped together with points in the hinterlands (see Figure 17). To test the impact of the clustering, we varied the number of clusters used globally for the continents, (see supporting information) and find that the extremely high uncertainties disappear, when more than 80 clusters are used globally for the continents. Thus, the high uncertainties are also a result of suboptimal clustering. The main difference with respect to previous models (Crust1.0 or the model of Chulick et al., 2013) is the very thin crust in the hinterlands of the Andean orogeny. There are only a few points that show this thinning, and this trend is interpolated along the entire north-south extent of the mountain chain. However, this feature is also associated with the highest uncertainties of 12 km. We note that Van der Meijde et al. (2013) and Reguzzoni and Sampietro (2015) detected a similar area of thinned crust using gravity inversion.

4.5.3. Africa

We chose Africa to illustrate the behavior of the interpolation algorithm in a data sparse environment (Figure 18). Data points are concentrated in Southern Africa, the East-African rift system, and the Medi-

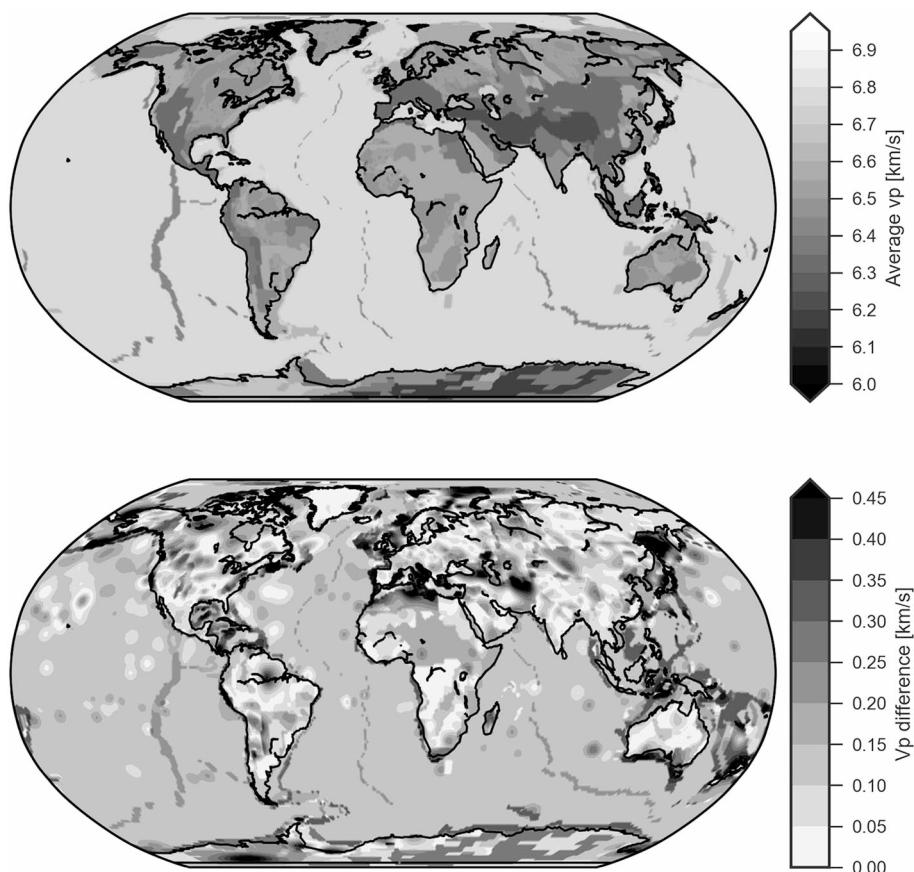


Figure 16. Average crustal P wave velocity for Crust1.0 and the difference with respect to our results.

ranean coast. The values of the range ρ in most of Africa are very low, so that the results are not interpolated very far away from the points. These few scattered points provide little constraint on the overall crustal structure, which can therefore be regarded as virtually unknown, as reflected in the high uncertainties of around 7 km.

Note that this pessimistic view only applies to our database, which does not contain important new results derived from passive seismic methods (Fishwick & Bastow, 2011). However, the crustal structure of Africa remains unclear, and nonseismic methods are still important to better determine it (Globig et al., 2016).

Crust1.0 shows a tile-like pattern of Moho depth variation (Figure 14) that reflects the predefined crustal types. Over South Africa, Crust1.0 disagrees with our results as well as the points in the GSC database.

4.5.4. Europe

Europe is among the most well surveyed areas in the world. Thus, comparing our results with Crust1.0 in this area highlights differences caused by different methodologies, because neither method is adversely affected by sparse data.

In this region, there is substantial agreement of the Moho depth between our results and Crust1.0 (Figure 19), apart from the general differences arising from the different treatment of the continent-ocean boundary (see, e.g., the Mediterranean coast of Africa). On the continent, substantial large-scale differences mostly occur where no data are available in our database (e.g., Baltic States; Figure 19d). There are some punctual large differences, for example, in the Alps. However, given the coarse parameterization of both models, neither model is capable of resolving the crustal variations on the scale of the Alps satisfactorily. Overall, higher differences between the models are also associated with higher estimated uncertainties (Figure 19c and 19d).

While the main patterns are resolved by both for the Moho depth, this does not hold for average P wave velocity. In continental Europe there tends to be agreement in areas covered by seismic points. However, in

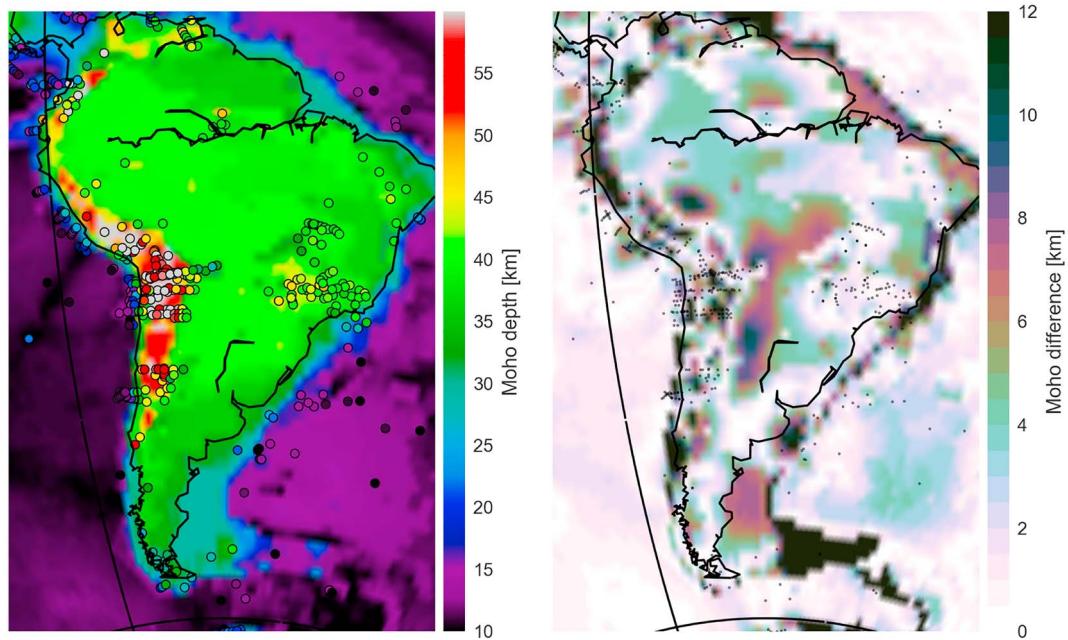


Figure 17. Interpolated Moho depth for South America and the estimated uncertainty from interpolation. The dots give the location of interpolation points.

some areas, for example, Finland, there are differences of up to 0.2 km/s, despite a sufficient number of data points. In less well covered areas, there are larger discrepancies.

The impact of the different methodologies is strikingly visible in the way the separation between West and Central Europe is represented in the Crust1.0 and our results. In Crust1.0 the velocity changes abruptly, due to the tectonic localization, whereas our results show a more diffuse boundary (Figure 20).

The diffuse appearance does not reflect the nature of the boundary itself but is a result of the interpolation. The underlying assumption of a Gaussian field with slowly varying covariance parameters is not well suited to represent abrupt parameter jumps. Furthermore, the precise location of the boundary is not completely constrained from our data set, which contributes to a diffuse boundary.

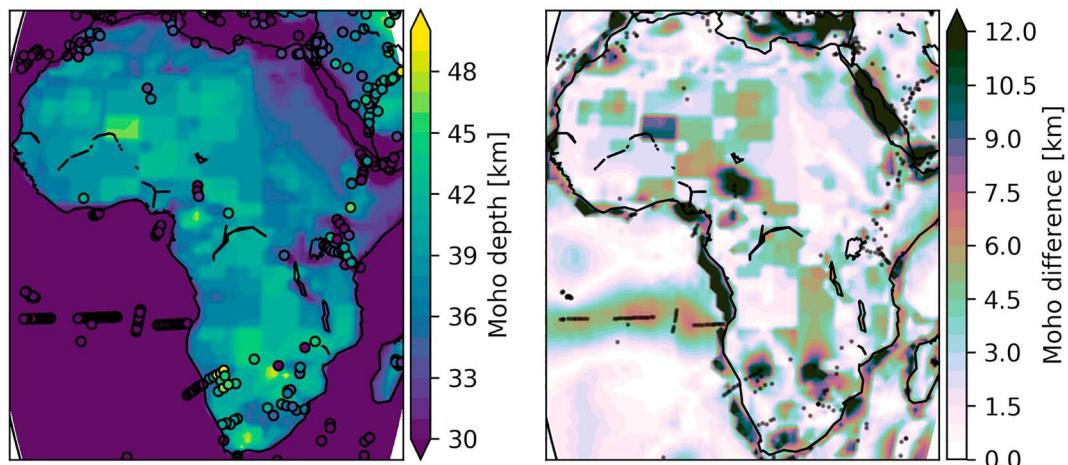


Figure 18. Interpolated Moho depth for Africa and the estimated uncertainty from interpolation.

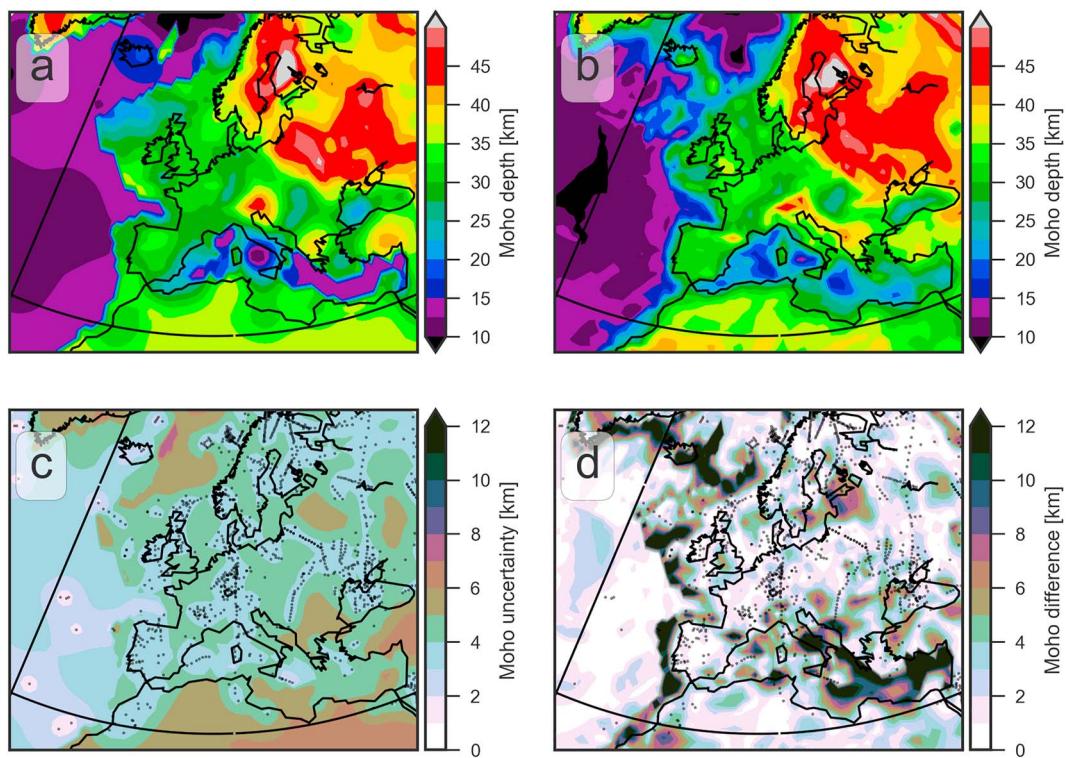


Figure 19. Moho depth interpolation results for Europe. (a, b) Moho depth from our results and Crust1.0. (c) Estimated uncertainty. (d) Difference between our results and Crust1.0.

5. Application of Global Maps to Residual Topography

We will use the calculation of residual topography as an example of how the uncertainties can be used in geophysical applications. Isostatic topography t_{iso} is given by

$$t_{\text{iso}} = \frac{M - M_{\text{ref}}}{\rho_t} (\rho_m - \rho_c)$$

where M is Moho depth (positive downward) and M_{ref} is Moho depth in the reference model, $\rho_t = 2,670 \frac{\text{kg}}{\text{m}^3}$ is the density of topography, ρ_c is the mean crustal density, and $\rho_m = 3,200 \frac{\text{kg}}{\text{m}^3}$ is the reference mantle density. For simplicity, we will neglect the contribution of sediments and ice here.

Residual topography is defined as $t_{\text{equiv}} - t_{\text{iso}}$ (Kaban et al., 1999), where equivalent topography accounts for the reduced density of water

$$t_{\text{equiv}} = \begin{cases} t & \text{if } t > 0 \\ t \frac{1640}{2670} & \text{if } t < 0 \end{cases}$$

Residual topography is possibly due to four causes: errors in the crustal model, density variations in the lithospheric mantle, elastic flexure of the lithosphere, and viscous stresses resulting from mantle convection that act on the base of the lithosphere (e.g., Molnar et al., 2015).

To find the mean crustal density ρ_c , a velocity-density relationship is required. Since the typical $v_p - \rho$ relations are pressure and thus depth dependent (Brocher, 2005; Christensen & Mooney, 1995), we cannot apply them immediately to the mean crustal velocity. Instead, we use linear regression to the mean velocities and densities from Crust1.0 to derive an empirical relation between average crustal v_p and density. This leads to the following relations for oceans and continents

$$\begin{aligned} \rho \left[\frac{\text{kg}}{\text{m}^3} \right] &= 350 + 385 v_p \left[\frac{\text{km}}{\text{s}} \right] \\ \rho \left[\frac{\text{kg}}{\text{m}^3} \right] &= 590 + 346 v_p \left[\frac{\text{km}}{\text{s}} \right] \end{aligned} \quad (4)$$

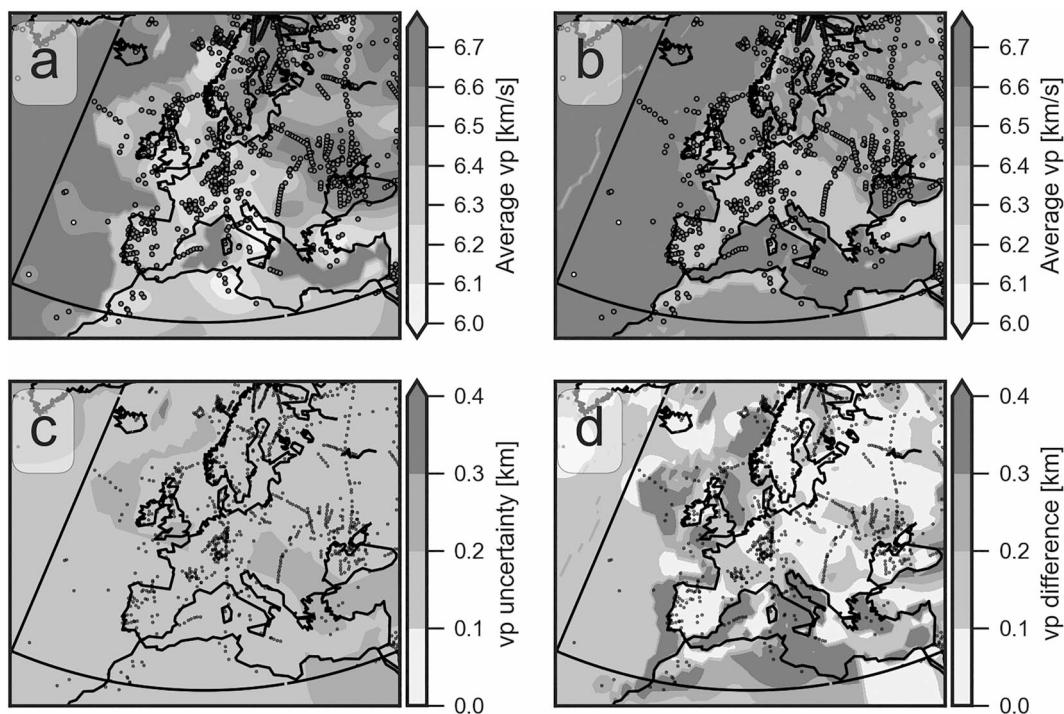


Figure 20. P wave velocity of the crystalline crust for Europe. (a, b) Comparison of our results and Crust1.0. The circles give the original points from the database. (c) Estimated uncertainty. (d) Difference between our results and Crust1.0.

We propagate the uncertainties of Moho and P wave velocity to the isostatic topography using Gaussian error propagation. Isostatic topography is a simple 1-D calculation, so spatial covariances between errors at different locations can be ignored. The contribution of a Moho uncertainty of δM is then

$$\delta t_M = \left| \frac{\rho_m - \rho_c}{\rho_t} \right| \delta M$$

A velocity uncertainty of δv contributes

$$\delta t_v = \left| \frac{M - M_{ref}}{\rho_t} \right| \frac{\partial \rho_c}{\partial v_p} \delta v$$

where the partial derivative of density with respect to velocity corresponds to the linear term in the empirical velocity-density relation (equation (4)).

The total uncertainty is the sum of the uncertainty variances resulting from Moho depth and P wave velocity estimates, $\delta t = \sqrt{\delta t_M^2 + \delta t_v^2}$, assuming that velocity and Moho depth errors are uncorrelated. Note that velocity and depth errors are in reality correlated because velocity is used to convert measurements of travel time to depth. This would tend to reduce the overall uncertainty, because depth and density have opposing effects on isostatic topography.

The calculated residual topography (Figure 21a) has values between -4 and 6 km. On a large scale the patterns are as would be expected: Positive anomalies at mid-oceanic ridges that systematically decrease as the oceanic plate gets older. Continents are generally associated with negative anomalies that are probably related to thick, cold sublithospheric mantle, whose cold thick lithosphere pulls down topography (Zoback & Mooney, 2003)

In general, the uncertainties coming from Moho depth interpolation are larger than the errors coming from velocity in the continents (Figures 21b and 21c). Only in the regions with the thickest crust, velocity contributes more than a few 100 m to the uncertainty. The oceans have a very high velocity uncertainty, leading to systematic uncertainty of almost 1 km. Maximum total uncertainty is 2 km, which only occurs in South America and Himalayas (Figure 21d). Note that, as discussed above, the uncertainties for South America are probably too pessimistic in the Andean region.

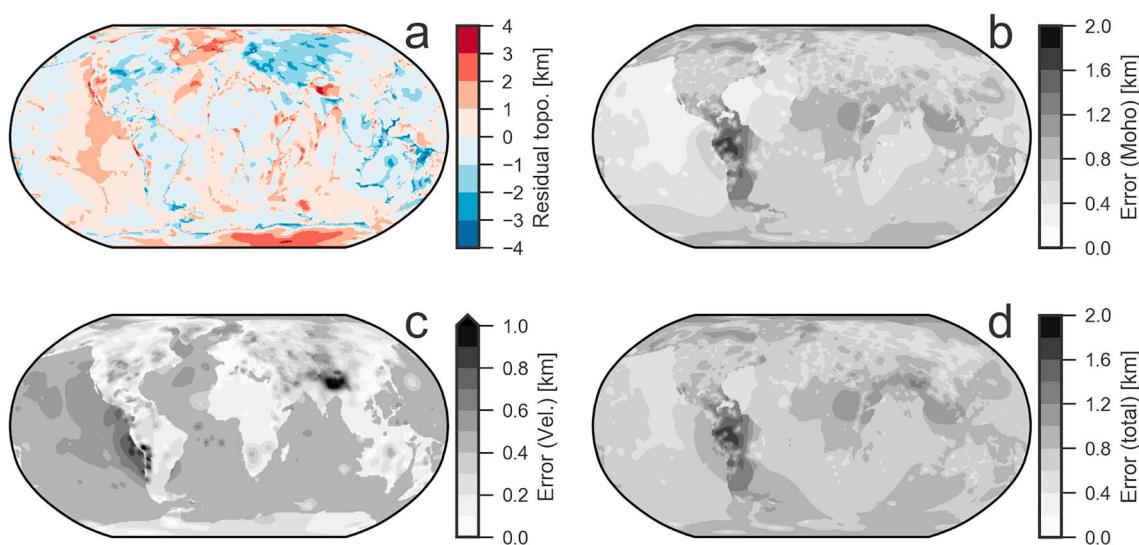


Figure 21. (a) Residual topography calculated from our crustal models. (b) Uncertainty of residual topography resulting from uncertainty of Moho depth. (c) Uncertainty caused by P wave velocity uncertainty. (d) Total uncertainty assuming Moho depth and P wave velocity uncertainties are uncorrelated.

On average the total uncertainty is 1.1 km. For around 65% of the points, the magnitude of estimated residual topography is larger than the estimated error and for around 40% of the points, the value of residual topography is twice as high as the estimated error. This demonstrates that mantle contributions to topography above the uncertainty threshold exist. However, the depth at which the mantle contribution originates cannot be constrained with this method. The sources might lie in the lithospheric or in the deeper mantle. Differentiating between lithospheric and deeper sources requires additional information, for example, from gravity modeling (e.g., Afonso et al., 2016).

6. Conclusions

We have constructed a global crustal model based on a USGS database of crustal seismic studies. Unlike previous studies, we have tried to include as little additional information as possible, following a design philosophy that tries to respect the data as much as possible. We interpolated Moho depth and average P wave velocity of the crystalline crust using a nonstationary kriging algorithm (Risser & Calder, 2017). A major advantage of our approach is that it is relatively straightforward to derive uncertainties.

The uncertainties we derived should be seen as a worst-case estimate, since we have only included a subset of seismic methods that sample the crust. For instance, short-period surface waves, ambient noise tomography (Shapiro & Campillo, 2004), and receiver functions (Kind et al., 2012) also provide information about crustal structure. However, we chose not to include them in our analysis, mainly because the combination of information from different sources requires good knowledge of the uncertainties affecting either data source. Thus, the kind of analysis we carried out here will certainly benefit a combined analysis of several types of data. Furthermore, active seismic surveys, which make up the bulk of the GSC database, are generally seen as the most reliable way of imaging the crust.

However, uncertainties remain even in areas where a large number of profiles have been acquired with estimated Moho depth errors of around or less than 4 km over large areas. These uncertainties stem both from errors of individual data points, which are reflected in the nugget effect and the uncertainty introduced by interpolating. We find that the nugget effect of the Moho depth is generally less than the estimated accuracy of 10 % (4 km) given by Mooney et al. (1998), which is probably due to better accuracy of newer studies, as confirmed using hold-one-out cross validation.

Interpolation of average crustal v_p proved to be more challenging than interpolating the Moho depth. To some extent, this is a bit enigmatic, since depths are in all seismic methods calculated from the velocities in some way. However, the relative accuracy of velocities is still better than the relative accuracy of Moho depth, which partly explains this apparent contradiction.

Are domain-averaged velocity profiles and other methods of avoiding direct interpolation of v_p preferable over direct interpolation? On the one hand this depends on the importance given to small-scale variations of velocity. Are they noise or incompletely imaged structure? In the latter case, direct interpolation would at least partially reflect them instead of averaging them out. On the other hand it depends on what a crustal model should represent: Is it one plausible model or is it a mean of all plausible models? Our method aims for the latter. Thus, a sharp velocity transition between velocity domains would be represented as a smooth transition, as long as the precise location of that transition is unknown. Both variants of crustal models are valid, but it needs to be clear what type a specific crustal model is. Ultimately, it might be worthwhile to find ways to incorporate the tectonic regularization into the interpolation to combine advantages of both approaches.

Combining the results of Moho depth and velocity interpolation, we could derive a map of residual topography that shows the expected features. In addition, we could also estimate the uncertainty of our residual topography map. These uncertainties can be up to 2 km and are on average 1 km. Given the expected range of dynamic topography of a few hundred meters (Molnar et al., 2015), this raises the question if global statements about dynamic topography are even possible at this stage. Of course, crustal models incorporating additional information reduce this uncertainty. But even in already well-studied areas, uncertainties can easily reach 1 km. A more comprehensive approach that includes gravity in addition to topography is probably the most effective way to constrain residual topography, since gravity and topography are both sensitive to the density structure.

Future research about interpolation of crustal structure could improve our results in different ways. Constructing a layered model of crustal structure instead of a one-layer model might be a logical next step. In addition, the ocean-continent classification could be derived not only from additional data but also using the crustal thickness data itself. The mathematical model underlying our interpolation can be improved by including anisotropy. Furthermore, the clustering algorithm we used could be improved by including the Moho depth and P wave velocity values during the clustering process.

Since new seismic data are continually acquired, any crustal model is already outdated at the time of publication. In addition, global and regional catalogs of seismic determinations already exist (Assumpção et al., 2013; Baranov & Morelli, 2013) that rely on different selections of data. For this reason, the interpolation software is freely available here (supporting information Data Set S3) to allow users to apply them to any database of seismic studies.

Appendix A: Clustering Algorithm

We use an agglomerative clustering algorithm (Hastie et al., 2009) as implemented in sci-kit learn (Pedregosa et al., 2011).

Let D be the distance matrix, such that D_{ij} is the great-circle distance between points i and j . The algorithm begins with every point in its own singleton cluster. Larger clusters are then created iteratively by repeatedly merging the two most similar clusters. The distance between two clusters of points G and H is in our case defined as the maximum distance between any points of pairs in G and H .

$$d(G, H) = \min_{i \in G, j \in H} d_{ij} \quad (\text{A1})$$

This procedure is repeated until the number of remaining groups is equal to the number specified by the user. When using equation (A1), the algorithm is called complete linkage agglomerative clustering. We prefer to use complete linkage over other methods, because it tends to create compact clusters (Hastie et al., 2009).

References

- Afonso, J. C., Rawlinson, N., Yang, Y., Schutt, D. L., Jones, A. G., Fullea, J., & Griffin, W. L. (2016). 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle: III. Thermochemical tomography in the Western-Central US. *Journal of Geophysical Research: Solid Earth*, 121, 7337–7370. <https://doi.org/10.1002/2016JB013049>
- Artemieva, I. M., & Mooney, W. D. (2001). Thermal thickness and evolution of Precambrian lithosphere: A global study. *Journal of Geophysical Research*, 106(B8), 16,387–16,414. <https://doi.org/10.1029/2000JB900439>
- Assumpção, M., Feng, M., Tassara, A., & Julià, J. (2013). Models of crustal thickness for South America from seismic refraction, receiver functions and surface wave tomography. *Tectonophysics*, 609, 82–96. <https://doi.org/10.1016/j.tecto.2012.11.014>

Acknowledgments

This study was supported by the Deutsche Forschungsgemeinschaft (DFG) Priority Program DynamicEarth, SPP 1788, an Australia-Germany Joint Research Cooperation Scheme funded by the German Academic Exchange Service (DAAD) and Universities Australia, and European Space Agency (ESA) Support to Science Element 3D-Earth. We thank the two anonymous reviewers and the Editors for their thorough review and constructive comments. The global mode of Moho depth and P wave velocity and its uncertainty are available in the supporting information (Figures S1 and S2). The python code used to create the crustal model is available in the supporting information (Data Set S3).

- Baranov, A., & Morelli, A. (2013). The Moho depth map of the Antarctica region. *Tectonophysics*, 609, 299–313. <https://doi.org/10.1016/j.tecto.2012.12.023>
- Bassin, C., Laske, G., & Masters, T. G. (2000). The current limits of resolution for surface wave tomography in North America. *Transactions of the AGU*.
- Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America*, 95(6), 2081–2092.
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5), 1190–1208. <https://doi.org/10.1137/0916069>
- Chiles, J.-P., & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty*. Wiley series in probability and statistics: Applied probability and statistics section. New York: Wiley.
- Christensen, N. I., & Mooney, W. D. (1995). Seismic velocity structure and composition of the continental crust: A global view. *Journal of Geophysical Research*, 100(B6), 9761–9788. <https://doi.org/10.1029/95JB00259>
- Chulick, G. S., Detweiler, S., & Mooney, W. D. (2013). Seismic structure of the crust and uppermost mantle of South America and surrounding oceanic basins. *Journal of South American Earth Sciences*, 42, 260–276. <https://doi.org/10.1016/j.jsames.2012.06.002>
- Cressie, N. A. C. (2015). *Statistics for spatial data*. Hoboken, NJ: John Wiley Sons Inc.
- Fishwick, S., & Bastow, I. D. (2011). Towards a better understanding of African topography: A review of passive-source seismic studies of the African crust and upper mantle. *Geological Society, London Special Publications*, 357(1), 343–371. <https://doi.org/10.1144/SP357.19>
- Globig, J., Fernández, M., Torne, M., Vergés, J., Robert, A., & Faccenna, C. (2016). New insights into the crust and lithospheric mantle structure of Africa from elevation, geoid, and thermal analysis. *Journal of Geophysical Research: Solid Earth*, 121, 5389–5424. <https://doi.org/10.1002/2016JB012972>
- Goutorbe, B., Poort, J., Lucazeau, F., & Raillard, S. (2011). Global heat flow trends resolved from multiple geological and geophysical proxies. *Geophysical Journal International*, 187(3), 1405–1419. <https://doi.org/10.1111/j.1365-246X.2011.05228.x>
- Guinness, J., & Fuentes, M. (2016). Isotropic covariance functions on spheres: Some properties and modeling considerations. *Journal of Multivariate Analysis*, 143, 143–152. <https://doi.org/10.1016/j.jmva.2015.08.018>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer eBook collection Mathematics and statistics (2nd ed.). New York: Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Jones, E., & Oliphant, T. (2001). Peterson, P. SciPy: Open source scientific tools for Python.
- Kaban, M., Schwintzer, P., & Tikhotsky, S. A. (1999). A global isostatic gravity model of the Earth. *Geophysical Journal International*, 136(3), 519–536. <https://doi.org/10.1046/j.1365-246x.1999.00731.x>
- Kind, R., Yuan, X., & Kumar, P. (2012). Seismic receiver functions and the lithosphere–asthenosphere boundary. *Tectonophysics*, 536, 25–43.
- Laske, G., Masters, T. G., Ma, Z., & Pasyanos, M. E. (2013). Update on CRUST1.0—A 1-degree global model of Earth's crust. *Geophysical Research Abstracts*, 15.
- Leone, F. C., Nelson, L. S., & Nottingham, R. B. (1961). The folded normal distribution. *Technometrics*, 3(4), 543. <https://doi.org/10.2307/1266560>
- McKenzie, D., & Bickle, M. J. (1988). The volume and composition of melt generated by extension of the lithosphere. *Journal of Petrology*, 29(3), 625–679.
- Meier, U., Curtis, A., & Trampert, J. (2007). Global crustal thickness from neural network inversion of surface wave data. *Geophysical Journal International*, 169(2), 706–722. <https://doi.org/10.1111/j.1365-246X.2007.03373.x>
- Molnar, P., England, P. C., & Jones, C. H. (2015). Mantle dynamics, isostasy, and the support of high terrain. *Journal of Geophysical Research: Solid Earth*, 120, 1932–1957. <https://doi.org/10.1002/2014jb011724>
- Mooney, W. D. (2015). Global crustal structure. In G. Schubert (Ed.), *Treatise on geophysics* (2nd ed.). Amsterdam, Netherlands: Elsevier.
- Mooney, W. D., Laske, G., & Masters, T. G. (1998). CRUST 5.1: A global crustal model at $5 \times 5^\circ$. *Journal of Geophysical Research*, 103(B1), 727–747. <https://doi.org/10.1029/97JB02122>
- Müller, R. D., Sdrolias, M., Gaina, C., & Roest, W. R. (2008). Age, spreading rates, and spreading asymmetry of the world's ocean crust. *Geochemistry, Geophysics, Geosystems*, 9, Q04006. <https://doi.org/10.1029/2007GC001743>
- Paciorek, C., & Schervish, M. J. (2004). A nonstationary covariance functions for Gaussian process regression. In S. Thrun & L. Saul (Eds.), *Advances in Neural Information Processing Systems 16* (pp. 16). Schölkopf: MIT Press.
- Pasyanos, M. E., Masters, T. G., Laske, G., & Ma, Z. (2014). LITHO1.0: An updated crust and lithospheric model of the Earth. *Journal of Geophysical Research: Solid Earth*, 119, 2153–2173. <https://doi.org/10.1002/2013JB010626>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Prodehl, C., & Mooney, W. D. (2012). *Exploring the Earth's crust: History and results of controlled-source seismology*. Memoir / Geological Society of America (Vol. 208). Boulder, CO: Geological Soc. of America.
- Reguzzoni, M., & Sampietro, D. (2015). GEMMA: An Earth crustal model based on GOCE satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 35(Part A), 31–43. <https://doi.org/10.1016/j.jag.2014.04.002>
- Reguzzoni, M., Sampietro, D., & Sanso, F. (2013). Global Moho from the combination of the CRUST2.0 model and GOCE data. *Geophysical Journal International*, 195(1), 222–237. <https://doi.org/10.1093/gji/gjt247>
- Risser, M. D., & Calder, C. A. (2017). Local likelihood estimation for covariance functions with spatially-varying parameters: The convoSpat package for R. Retrieved from <https://arxiv.org/abs/1507.08613>
- Schaeffer, A. J., & Lebedev, S. (2015). Global heterogeneity of the lithosphere and underlying mantle: A seismological appraisal based on multimode surface-wave dispersion analysis, shear-velocity tomography, and tectonic regionalization. In A. Khan & F. Deschamps (Eds.), *The Earth's heterogeneous mantle* (pp. 3–46). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-15627-9_e1
- Shapiro, N. M., & Campillo, M. (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31, L07614. <https://doi.org/10.1029/2004GL019491>
- Stolk, W., Kaban, M., Beekman, F., Tesauro, M., Mooney, W. D., & Cloetingh, S. (2013). High resolution regional crustal models from irregularly distributed data: Application to Asia and adjacent areas. *Tectonophysics*, 602, 55–68. <https://doi.org/10.1016/j.tecto.2013.01.022>
- Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. SIAM - Soc. for Industrial and Applied Math, Philadelphia, PA.
- Tesauro, M., Kaban, M. K., Mooney, W. D., & Cloetingh, S. (2014). NACr14: A 3D model for the crustal structure of the North American continent. *Tectonophysics*, 631, 65–86. <https://doi.org/10.1016/j.tecto.2014.04.016>
- Tibshirani, R., & Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82, 398–559. <https://doi.org/10.2307/2289465>

- van der Meijde, M., Julià, J., & Assumpção, M. (2013). Gravity derived Moho for South America. *Tectonophysics*, 609, 456–467. <https://doi.org/10.1016/j.tecto.2013.03.023>
- White, R. S., McKenzie, D., & O'Nions, R. K. (1992). Oceanic crustal thickness from seismic measurements and rare Earth element inversions. *Journal of Geophysical Research*, 97(B13), 19683. <https://doi.org/10.1029/92JB01749>
- Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4), 550–560. <https://doi.org/10.1145/279232.279236>
- Zoback, M. L., & Mooney, W. D. (2003). Lithospheric buoyancy and continental intraplate stresses. *International Geology Review*, 45(2), 95–118. <https://doi.org/10.2747/0020-6814.45.2.95>