

Global models from sparse data: A robust estimate of Earth’s residual topography spectrum

Andrew P Valentine^{1,1} and D. Rhodri Davies^{1,1}

¹Australian National University

November 30, 2022

Abstract

A significant component of Earth’s surface topography is maintained by stresses induced by underlying mantle flow. This ‘dynamic’ topography cannot be directly observed, but it can be approximated — particularly at longer wavelengths — from measurements of residual topography, which are obtained by removing isostatic effects from the observed topography. However, as these measurements are made at discrete, unevenly-distributed locations on Earth’s surface, inferences about global properties can be challenging. In this paper, we present and apply a new approach to transforming point-wise measurements into a continuous global representation. The approach, based upon the statistical theory of Gaussian Processes, is markedly more stable than existing approaches — especially for small datasets. We are therefore able to infer the spatial pattern, wavelength and amplitude of residual topography using only the highest-quality oceanic spot measurements within the database of Hoggard et al. (2017). Our results indicate that the associated spherical harmonic power spectrum peaks at $l=2$, with power likely in the range $0.46\text{--}0.76\text{ km}^2$. This decreases by over an order of magnitude to around 0.02 km^2 at $l=30$. Around 85% of the total power is concentrated in degrees 1–3. Our results therefore confirm previous findings: Earth’s residual topography expression is principally driven by deep mantle flow, but shallow processes are also crucial in explaining the general form of the power spectrum. Finally, our approach allows us to determine the locations where collection of new data would most impact our knowledge of the spectrum.

1 **Global Models from Sparse Data: A Robust Estimate**
2 **of Earth's Residual Topography Spectrum**

3 **A.P. Valentine¹ and D.R. Davies¹**

4 ¹Research School of Earth Sciences, The Australian National University,
5 142 Mills Road, Acton ACT 2601, Australia.

6 **Key Points:**

- 7 • We present a new method for spectral analysis of sparse point data and apply this to
8 measurements of residual topography within the oceans;
9 • Using a smaller, more reliable dataset than earlier studies, we confirm a power spec-
10 trum that is dominated by long-wavelength components;
11 • We also confirm that the spectrum contains significant power at shorter wavelengths,
12 reflecting the multiscale nature of global mantle flow.

Corresponding author: Andrew Valentine, andrew.valentine@anu.edu.au

Abstract

A significant component of Earth’s surface topography is maintained by stresses induced by underlying mantle flow. This ‘dynamic’ topography cannot be directly observed, but it can be approximated — particularly at longer wavelengths — from measurements of residual topography, which are obtained by removing isostatic effects from the observed topography. However, as these measurements are made at discrete, unevenly-distributed locations on Earth’s surface, inferences about global properties can be challenging. In this paper, we present and apply a new approach to transforming point-wise measurements into a continuous global representation. The approach, based upon the statistical theory of Gaussian Processes, is markedly more stable than existing approaches — especially for small datasets. We are therefore able to infer the spatial pattern, wavelength and amplitude of residual topography using only the highest-quality oceanic spot measurements within the database of Hoggard et al. (2017). Our results indicate that the associated spherical harmonic power spectrum peaks at $l = 2$, with power likely in the range $0.46\text{--}0.76\text{ km}^2$. This decreases by over an order of magnitude to around 0.02 km^2 at $l = 30$. Around 85% of the total power is concentrated in degrees 1–3. Our results therefore confirm previous findings: Earth’s residual topography expression is principally driven by deep mantle flow, but shallow processes are also crucial in explaining the general form of the power spectrum. Finally, our approach allows us to determine the locations where collection of new data would most impact our knowledge of the spectrum.

Plain Language Summary

As the mantle flows, it induces deformation at Earth’s surface, pushing it up at some locations and pulling it down elsewhere. This deformation can be quantified by measuring so-called ‘residual topography’, at specific locations. However, we only have a small number of such measurements, and they are clustered in particular areas on Earth’s surface: inferring a global representation of residual topography is therefore a challenge. In this paper, we develop a new method for deriving robust global maps from such data, and apply this to a set of measurements of residual topography. An advantage of our method is that it remains effective for small datasets, enabling a more conservative approach to data selection. Our results complement recent studies on residual topography, highlighting how mantle dynamics shapes the surface of our planet across a range of scales.

1 Introduction

Earth’s surface topography arises as a balance between processes that create elevation, such as tectonic convergence, and those that destroy it, including erosion, sediment transport and deposition. Most topography is *isostatic*, being maintained by thickness and density variations within the crust and lithospheric mantle. However, there is growing consensus that a substantial proportion is controlled by mantle convection: as the mantle flows, it transmits normal stresses to the lithosphere, and these are balanced by gravitational stresses arising through topographic deflections of Earth’s surface (e.g. Pekeris, 1935; Parsons & Daly, 1983; Hager et al., 1985; Hager & Richards, 1989; Mitrovica et al., 1989; Gurnis, 1993; Lithgow-Bertelloni & Silver, 1998; Gurnis et al., 2000; Conrad & Husson, 2009; Braun, 2010; Shephard et al., 2010; Flament et al., 2013; Hoggard et al., 2016; Yang & Gurnis, 2016; Rubey et al., 2017; Eakin & Lithgow-Bertelloni, 2018). This so-called *dynamic* topography is transient, varying both spatially and temporally in response to underlying mantle flow. As a result, it is challenging to isolate — but doing so can provide important constraints upon the flow regime, planform and intensity of mantle convection.

Observational constraints on dynamic topography are indirect: estimates of its spatial pattern, wavelength and amplitude are generally equated to so-called *residual* topography, which is calculated by removing the isostatic contribution of sediments, ice, crust and lithosphere from the observed topography (e.g. Panasyuk & Hager, 2000; Kaban et al., 2003;

63 Flament et al., 2013; Guerri et al., 2016; Hoggard et al., 2016, 2017). In practice, residual
64 topography provides an upper bound on flow-related dynamic topography (e.g. Hoggard
65 et al., 2017). Indeed, whilst the long-wavelength components of residual topography are
66 principally dynamic in origin, a significant proportion of shorter-wavelength oceanic ‘resid-
67 ual’ topography arises from isostatic effects associated with local variations in lithospheric
68 thickness and density, that cannot be accounted for through a simple plate cooling model
69 (Davies et al., 2019). Despite this, measurements of residual topography are currently our
70 best source of information on dynamic topography.

71 Accurate measurements of residual topography depend upon a careful synthesis of
72 information from a variety of sources — including models for surface elevation, sediment
73 cover, crustal properties, and thermal effects. Inevitably, the availability and quality of such
74 information is not uniform across Earth’s surface. A recent database of residual topography
75 within the world’s oceans, compiled by Hoggard et al. (2016, 2017), provides one of the
76 most comprehensive datasets currently available. This has underpinned a series of studies
77 into the spectral character of residual topography and its relationship to underlying mantle
78 dynamics (e.g. Hoggard et al., 2016, 2017; Steinberger, 2016; Yang & Gurnis, 2016; Yang
79 et al., 2017; Watkins & Conrad, 2018; Steinberger et al., 2019). However, the conclusions
80 from these studies have often appeared contradictory. At least in part, this has been due
81 to authors making different subjective choices within their analyses, with the debate being
82 further complicated by a lack of consistency around the spherical harmonic normalisation
83 conventions employed by different groups.

84 In an effort to reduce subjectivity in the analysis, a recent study by Davies et al.
85 (2019) considered the database of Hoggard et al. (2017) using a novel hierarchical Bayesian
86 approach (Valentine & Sambridge, 2018). In particular, a strategy of ‘Automatic Relevance
87 Determination’ (ARD) was employed to avoid the need to impose any pre-determined notion
88 of ‘smoothness’ upon the residual topography. Results indicate that the spectrum was
89 dominated by long-wavelength ($\sim 10^4$ km) features, with power dropping by about an order
90 of magnitude at shorter wavelengths ($\sim 10^3$ km). Based upon predictions from instantaneous
91 models of global mantle dynamics, Davies et al. (2019) demonstrate that both deep mantle
92 flow and shallow, lithosphere-controlled processes are important in generating this surface
93 response.

94 However, one drawback in the ARD approach employed by Davies et al. (2019) is that
95 it becomes unstable as the number of data points is reduced. Within the database compiled
96 by Hoggard et al. (2016, 2017), a subset of around 5% of the measurements are considered
97 markedly more robust than the rest. Ideally, it would be possible to analyse those points in
98 isolation — but this is not possible using the ARD strategy. The present paper, therefore,
99 develops an alternative approach to analysis, building on the statistical theory of Gaussian
100 Processes (e.g. Rasmussen & Williams, 2006) and the work of Valentine and Sambridge
101 (2020a, 2020b). This method remains stable for even the smallest datasets, and allows us to
102 convincingly demonstrate that the conclusions of Davies et al. (2019) remain valid when only
103 the highest-quality data points are used. We are also able to obtain maps showing where
104 new measurements would be most valuable in constraining the residual topography profile.
105 While the present paper focusses only on residual topography, our underlying method has
106 much wider application, and may be useful for any attempt to characterise continuous Earth
107 properties from discrete samples.

108 We begin by providing a brief summary of the observational datasets used in this study,
109 which are described more fully elsewhere. We then summarise how Gaussian processes can be
110 used to represent functions on Earth’s surface, building on the presentation in Valentine and
111 Sambridge (2020a, 2020b), and develop the mathematical results necessary for converting the
112 recovered model into a spherical harmonic representation. Finally, we show that application
113 of this theory to the observational data yields results that are self-consistent across different
114 data subsets, and in general agreement with earlier studies.

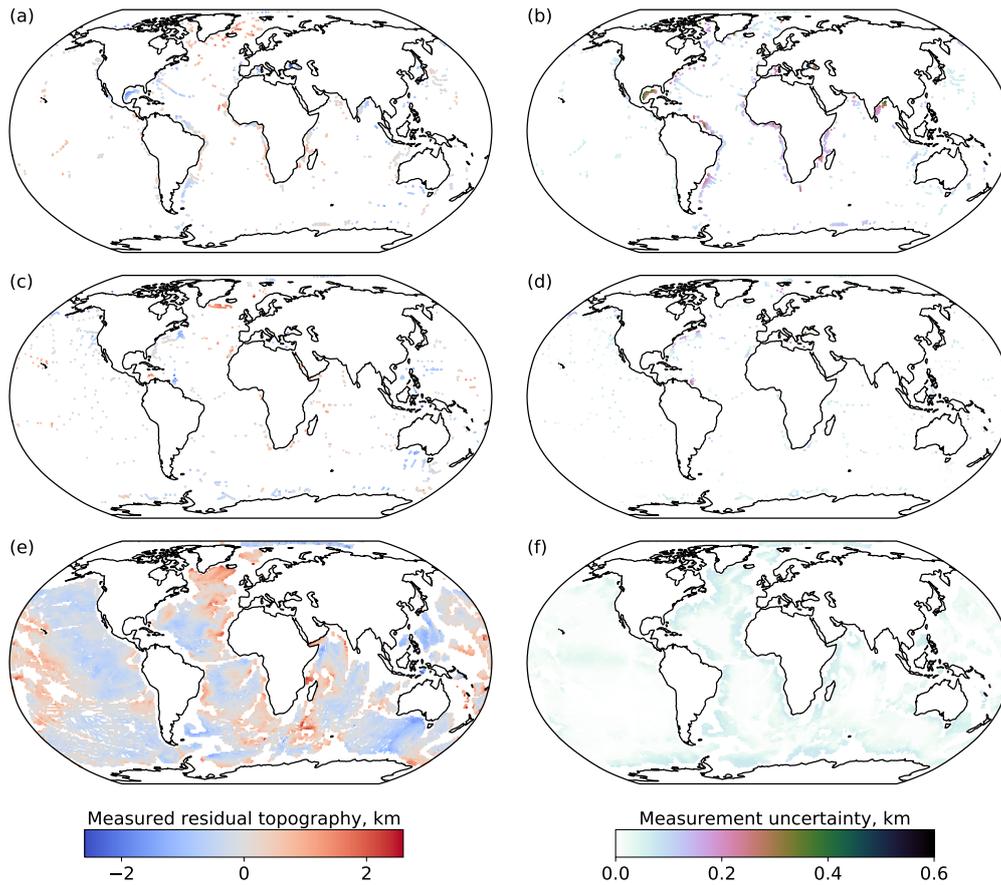


Figure 1. Measurements of residual topography. 1160 high-accuracy spot measurements are shown in (a), with associated uncertainties in (b). A further 870 spot measurements that lack crustal corrections are shown in (c), with uncertainties in (d): these are shown without the additional 200 m uncertainty assessed by Hoggard et al. (2017) to reflect the absence of crustal information. Finally, (e) shows 20767 gridded measurements derived from shiptrack bathymetry; again, the uncertainties depicted in (f) do not include any additional uncertainty to reflect the lack of crustal corrections.

115 2 Datasets

116 The database employed for this study is described in Davies et al. (2019), and is a
 117 compilation of residual topography measurements within the world’s oceans. It represents
 118 an incremental update to the database of Hoggard et al. (2017), which itself builds on several
 119 previous studies (e.g. Winterbourne et al., 2009; Czarnota et al., 2013, 2014; Winterbourne
 120 et al., 2014; Hoggard et al., 2016). In comparison to Hoggard et al. (2017), the version
 121 accompanying Davies et al. (2019) incorporates information from additional marine seismic
 122 surveys, and has undergone further quality control. For the purposes of this paper, it is
 123 appropriate to divide the dataset into three classes, and we discuss each in turn.

124 2.1 High-accuracy spot measurements

125 These measurements are derived from marine seismic surveys, allowing direct observa-
 126 tion of the topography associated with the oceanic crust. Careful processing allows isostatic
 127 effects associated with sedimentary loading and crustal thickness to be removed, and age-

128 depth cooling is accounted for using a simple analytical plate model (Hoggard et al., 2017).
 129 A total of 1160 residual topography observations are obtained (see Fig. 1a), with associated
 130 uncertainty (Fig. 1b). We consider these points to be the most robust within the database,
 131 and the primary goal of this paper is to ascertain the extent to which the power spectrum
 132 of residual topography can be constrained using only these measurements.

133 2.2 Lower-accuracy spot measurements

134 A further 870 points, shown in Fig. 1(c), were obtained in a similar manner — but a
 135 lack of information prevented isostatic correction for variations in crustal thickness. These
 136 observations must therefore be assumed to be less accurate. To reflect this, Hoggard et
 137 al. (2017) increases the uncertainties reported for these data points by 0.2 km, a somewhat
 138 arbitrary figure chosen based on the median crustal correction associated with the high-
 139 accuracy data points. In the present paper, we replace this additional 0.2 km component by
 140 an unknown correction, Δ . We can then determine the value Δ should take to ensure sta-
 141 tistical consistency with information from the high-accuracy data points. The un-corrected
 142 uncertainties (i.e. without the additional 0.2 km component) are shown in Fig. 1(d).

143 2.3 Shiptrack-derived measurements

144 Finally, the database contains 20767 measurements derived from a global shiptrack
 145 bathymetry grid (Smith & Sandwell, 1997), with sedimentary corrections based on the
 146 model of Laske and Masters (1997). Again, Hoggard et al. (2017) increases the uncertain-
 147 ties associated with these measurements by 0.2 km, to reflect the lack of detailed crustal
 148 information; again, we replace this with the unknown correction Δ . These data points are
 149 illustrated in Fig. 1(e), with uncorrected uncertainties in Fig. 1(f). Clearly, these points
 150 provide excellent spatial coverage within the oceans, but the use of global datasets, which
 151 are themselves derived from a variety of sources, raises the possibility of significant un-
 152 quantified systematic biases within this portion of the dataset. We therefore regard the
 153 shiptrack-derived measurements as being the least robust within the database.

154 3 Methodology

155 The fundamental challenge in this work — and in many other geoscience questions
 156 — is to infer a continuous spatial function (i.e. the residual topography at any position on
 157 Earth’s surface) from a finite set of observations made at discrete locations. A variety of
 158 approaches are possible, but one common strategy is to assume that the spatial function
 159 can be expanded in terms of a finite set of basis functions. This transforms the inference
 160 task into one of determining the expansion coefficients relative to this basis, typically as a
 161 least-squares inversion problem. This is the approach adopted by Hoggard et al. (2016) and
 162 Davies et al. (2019), using a spherical harmonic basis. However, two key difficulties emerge.
 163 First, it is likely that the true spatial function contains features that cannot be represented
 164 using the finite basis. This can lead to an aliasing-like effect known as ‘spectral leakage’
 165 (Trampert & Snieder, 1996), biasing the recovered expansion coefficients away from their
 166 correct values. While it is possible to correct for this, it is computationally expensive to
 167 do so. Instead, Davies et al. (2019) adopted a pragmatic approach: since spectral leakage
 168 mainly affects the shortest-wavelength components within the expansion, these terms were
 169 omitted from further analysis.

170 The second difficulty arises because with sparse, unevenly-distributed data, not all
 171 components of the basis function expansion are equally-well constrained. Typically, some
 172 expansion coefficients can be varied significantly without affecting the fit to data, which can
 173 cause the interpolated function to behave unrealistically in regions of low data coverage.
 174 To address this, one typically introduces ‘regularisation’: additional constraints on the ba-
 175 sis function expansion, designed to ensure that a well-behaved solution can be found. In

176 Hoggard et al. (2016), the chosen constraints amounted to a requirement that the residual
 177 topography field be ‘small’ and ‘smooth’. In general, these are reasonable requirements—
 178 but they carry implications for the spectral characteristics of the recovered field. Thus, one
 179 could question whether the power spectrum from Hoggard et al. (2016) was truly mandated
 180 by the data, or if it had arisen as a consequence of the assumptions inherent to their analy-
 181 ses. As noted above, to overcome this potential shortcoming, Davies et al. (2019) employed
 182 a novel approach to regularisation, developed in Valentine and Sambridge (2018), based on
 183 the idea of ‘Automatic Relevance Determination’ (ARD) (Mackay, 1992). In the present
 184 context, this enables appropriate ‘smoothness’ characteristics to be determined from the
 185 data, rather than being imposed from the outset. However, application of this approach
 186 requires a relatively large dataset and, as a result, the ARD analysis can only be applied to
 187 the full residual topography dataset (i.e. that incorporating all spot- and shiptrack-derived
 188 measurements); attempts to apply it to only the high-accuracy spot measurements prove
 189 unstable.

190 Subsequently, Valentine and Sambridge (2020a, 2020b) set out an alternative frame-
 191 work for inferring continuous functions from point data, based on the statistical theory
 192 of Gaussian Processes. This avoids a number of the difficulties associated with the basis-
 193 function–expansion approach, and remains stable for small datasets. Thus, unlike ARD, it
 194 can be used to analyse the high-accuracy spot data from Hoggard et al. (2017) in isolation,
 195 as well as their dataset incorporating ship-track derived constraints.

196 3.1 A Gaussian Process model for residual topography

197 Gaussian Processes are a class of statistical model that have become an increasingly
 198 popular tool for machine learning. They are discussed at length in, for example, Rasmussen
 199 and Williams (2006) or Murphy (2012), and we do not attempt to provide a comprehensive
 200 introduction here. Instead, we outline the key features, building on the development and
 201 notation adopted in Section 2 of Valentine and Sambridge (2020a). Readers are encouraged
 202 to consult that paper if further details are required.

203 We wish to develop a model for the residual topography field, which we denote $h(\theta, \varphi)$,
 204 as a function of latitude (θ) and longitude (φ). To simplify notation, we introduce the posi-
 205 tion vector $\mathbf{x} = (\theta, \varphi)$, and will use $h(\mathbf{x})$ interchangeably with $h(\theta, \varphi)$. Whereas expressing
 206 $h(\mathbf{x})$ using a basis function expansion requires us to choose the set of basis functions and
 207 an approach to regularisation, expressing it as a Gaussian Process requires us to specify a
 208 ‘mean function’ and a ‘covariance function’. The mean function, $\mu(\mathbf{x})$, describes our *a priori*
 209 statistical assumptions about the expected value of $h(\mathbf{x})$; for present purposes, we assume
 210 that this has a constant value everywhere,

$$\mu(\mathbf{x}) = \mu_0. \quad (1)$$

211 The covariance function, $k(\mathbf{x}, \mathbf{x}')$ describes our *a priori* assumptions about the covariance
 212 between $h(\mathbf{x})$ and $h(\mathbf{x}')$. Put in more intuitive terms: $k(\mathbf{x}, \mathbf{x}')$ describes how learning the
 213 value of h at some point, \mathbf{x} , would influence our beliefs about the value of h at some other
 214 point, \mathbf{x}' . In the present case, it is reasonable to assume that the degree of influence should
 215 diminish as the distance between \mathbf{x} and \mathbf{x}' grows: we might expect residual topography to
 216 be similar for two localities 1 km apart, but knowledge of residual topography in the North
 217 Atlantic would not help us constrain residual topography in Australia. In this paper, we
 218 adopt the Matérn family of covariance functions, for which

$$k(\mathbf{x}, \mathbf{x}') = \sigma_1^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} d(\mathbf{x}, \mathbf{x}')}{\sigma_2} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} d(\mathbf{x}, \mathbf{x}')}{\sigma_2} \right), \quad (2)$$

219 where Γ denotes the Gamma function, and K_ν is a modified Bessel function of the second
 220 kind. Here, $d(\mathbf{x}, \mathbf{x}')$ is some measure of the distance between the points \mathbf{x} and \mathbf{x}' ; for reasons
 221 which will later become apparent, we employ the epicentral angle, Θ , defined such that

$$d(\theta, \varphi; \theta', \varphi') \equiv \Theta(\theta, \varphi; \theta', \varphi') = \arccos(\sin \theta \sin \theta' + \cos \theta \cos \theta' \cos(\varphi - \varphi')). \quad (3)$$

222 Note that this can be converted into the great-circle distance on the surface of the (assumed
 223 spherical) Earth by multiplying by Earth’s radius. The quantities σ_1 , σ_2 and ν in eq. (2)
 224 are hyperparameters controlling the detailed shape of the covariance function. Specifically,
 225 σ_1 governs its amplitude, σ_2 represents the spatial length-scale over which we expect an
 226 observation to be informative, and ν is an order parameter that, essentially, determines
 227 the smoothness of the interpolation. This Matérn family arises as a generalisation of the
 228 Gaussian (or squared-exponential) covariance function, which corresponds to the case $\nu =$
 229 ∞ , and is widely-used in machine learning and spatial statistics in cases where a general,
 230 localised covariance function is required. Readers seeking greater intuition on the role played
 231 by the covariance function may wish to consult Figs. 1 & 2 of Valentine and Sambridge
 232 (2020a) and the associated discussion.

233 When we represent residual topography using a Gaussian Process, which we denote by
 234 writing $h(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, we effectively assert that whenever we consider the value
 235 of h at any N distinct locations, $\mathbf{x}_{1\dots N}$, our overall state of knowledge should be described
 236 by an N -dimensional Gaussian distribution,

$$\begin{pmatrix} h(\mathbf{x}_1) \\ h(\mathbf{x}_2) \\ \vdots \\ h(\mathbf{x}_N) \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}) . \quad (4)$$

237 The elements of $\boldsymbol{\mu}$ are obtained by evaluating the mean function, $\mu_i = \mu(\mathbf{x}_i)$, and the
 238 elements of \mathbf{K} are derived from the covariance function, $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. In particular, if
 239 we consider a single point \mathbf{x} , we find that

$$h(\mathbf{x}) \sim \mathcal{N}(\mu_0, \sigma_1^2) . \quad (5)$$

240 This represents our prior probability distribution for residual topography at a point. Thus,
 241 before observing any data, our presumptions about residual topography, irrespective of
 242 location, can be characterised as $h = \mu_0 \pm \sigma_1$.

243 Now, suppose we are able to measure residual topography at a discrete set of points,
 244 $\hat{\mathbf{x}}_{1\dots N}$ (as in Valentine and Sambridge (2020a), a ‘hat’ is used to distinguish quantities as-
 245 sociated with the observational dataset). Our set of N measurements can be represented as
 246 the N -dimensional vector $\hat{\mathbf{d}}$, such that \hat{d}_i represents our estimate of $h(\hat{\mathbf{x}}_i)$, and we assume
 247 that all measurement errors are Gaussian, described overall by the $N \times N$ covariance matrix,
 248 $\mathbf{C}_{\hat{\mathbf{d}}}$. In the present case, this has entries on the leading diagonal only (so that all measure-
 249 ments are assumed independent), and is treated as a function of the unknown contribution
 250 to uncertainty arising from a lack of crustal corrections, Δ . According to the assumptions
 251 we have made in setting up our Gaussian Process model, this specific set of observations
 252 have likelihood

$$\mathbb{P}(\hat{\mathbf{d}} | \boldsymbol{\sigma}) = \frac{1}{(2\pi)^{N/2} |\hat{\mathbf{K}}_{\boldsymbol{\sigma}} + \mathbf{C}_{\hat{\mathbf{d}}}(\Delta)|^{1/2}} \exp \left\{ -\frac{1}{2} (\hat{\mathbf{d}} - \hat{\boldsymbol{\mu}}_{\boldsymbol{\sigma}})^{\text{T}} (\hat{\mathbf{K}}_{\boldsymbol{\sigma}} + \mathbf{C}_{\hat{\mathbf{d}}}(\Delta))^{-1} (\hat{\mathbf{d}} - \hat{\boldsymbol{\mu}}_{\boldsymbol{\sigma}}) \right\} \quad (6)$$

253 where $\hat{\boldsymbol{\mu}}_{\boldsymbol{\sigma}}$ and $\hat{\mathbf{K}}_{\boldsymbol{\sigma}}$ are obtained by evaluating the mean and covariance functions at the
 254 sample points $\hat{\mathbf{x}}$, and where the subscript $\boldsymbol{\sigma}$ is used to emphasise that these quantities them-
 255 selves depend on specific choices for the hyperparameters, $\boldsymbol{\sigma} = (\mu_0, \sigma_1, \sigma_2, \nu, \Delta)$. Plainly,
 256 we should not adopt assumptions that make our observations inherently ‘unlikely’, and so
 257 it is sensible to select hyperparameter values that maximise eq. (6). This is straightforward
 258 to implement using standard computational tools for optimisation, such as the L-BFGS-B
 259 algorithm (Byrd et al., 1995), available through Scipy. We remark that this concept can be
 260 framed more formally as an approximation to a hierarchical Bayesian estimation of $\boldsymbol{\sigma}$; see
 261 Valentine and Sambridge (2018).

262 Having selected optimal values for $\boldsymbol{\sigma}$, we are now in a position to predict residual
 263 topography at unseen locations. To do this, we note that eq. (4) holds for *any* set of

264 locations: in particular, it must hold if we consider the N observed locations and one
 265 additional point, \mathbf{x} . We can then condition this distribution on the values we actually
 266 observed, exploiting the correlations between locations to refine our knowledge about $h(\mathbf{x})$.
 267 As explained in Valentine and Sambridge (2020a), application of a standard result for the
 268 conditioning of Gaussian distributions results in a posterior estimate of residual topography
 269 given by

$$\tilde{h}(\mathbf{x}) \sim \mathcal{GP} \left(\tilde{\mu}(\mathbf{x}), \tilde{k}(\mathbf{x}, \mathbf{x}') \right), \quad (7a)$$

270 where

$$\tilde{\mu}(\mathbf{x}) = \mu(\mathbf{x}) + \hat{\mathbf{k}}(\mathbf{x})^T \left(\hat{\mathbf{K}} + \mathbf{C}_a \right)^{-1} \left(\hat{\mathbf{d}} - \hat{\boldsymbol{\mu}} \right) \quad (7b)$$

$$\tilde{k}(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \hat{\mathbf{k}}(\mathbf{x})^T \left(\hat{\mathbf{K}} + \mathbf{C}_a \right)^{-1} \hat{\mathbf{k}}(\mathbf{x}') \quad (7c)$$

271 and where $\hat{\mathbf{k}}(\mathbf{x})$ is the vector-valued function that evaluates the covariance function between
 272 point \mathbf{x} and each of the observation locations, such that

$$\left[\hat{\mathbf{k}}(\mathbf{x}) \right]_i = k(\mathbf{x}, \hat{\mathbf{x}}_i). \quad (7d)$$

273 Thus, within the Gaussian Process framework, the data allow us to infer that

$$\tilde{h}(\mathbf{x}) = \tilde{\mu}(\mathbf{x}) \pm \sqrt{\tilde{k}(\mathbf{x}, \mathbf{x})}. \quad (8)$$

274 This allows us to map the spatial pattern and amplitude of residual topography, which we
 275 return to in Section 4.

276 3.2 The power spectrum of residual topography

277 To understand the power spectrum of the recovered residual topography field, and to
 278 compare this with previously-published results, we must express $h(\mathbf{x})$ in terms of spherical
 279 harmonics. As in Hoggard et al. (2016) and Davies et al. (2019), we employ the real surface
 280 spherical harmonics, $\mathcal{Y}_{lm}(\theta, \varphi)$, defined as in Section B6 of Dahlen and Tromp (1998). The
 281 infinite set of spherical harmonics is complete and, thus, we can express our Gaussian Process
 282 model exactly as an infinite sum

$$\tilde{h}(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{lm} \mathcal{Y}_{lm}(\theta, \varphi). \quad (9)$$

283 Moreover, the spherical harmonics are normalised, and orthogonal, in the sense that

$$\int_{\Omega} \mathcal{Y}_{lm}(\theta, \varphi) \mathcal{Y}_{l'm'}(\theta, \varphi) d\Omega = \delta_{ll'} \delta_{mm'}, \quad (10)$$

284 where δ_{ij} is a Kronecker delta and integration is over a spherical surface. Hence, any
 285 coefficient c_{lm} can be found by evaluating

$$c_{lm} = \int_{\Omega} \tilde{h}(\theta, \varphi) \mathcal{Y}_{lm}(\theta, \varphi) d\Omega. \quad (11)$$

286 It is important to recognise that, whereas Hoggard et al. (2016) and Davies et al. (2019)
 287 sought to directly fit a spherical harmonic expansion to point data, our approach is different:
 288 we have already obtained a global representation of $\tilde{h}(\theta, \varphi)$, and we simply need to express
 289 this within the spherical harmonic basis. This distinction underpins the stability of our
 290 approach.

291 Because $\tilde{h}(\theta, \varphi)$ is a Gaussian Process, the integration in eq. (11) results in a probability
 292 distribution for c_{lm} . As discussed in Valentine and Sambridge (2020a), c_{lm} is normally-
 293 distributed, and the coefficients associated with different values of l and m are jointly normal.
 294 If we use \mathbf{c} to denote some set of these coefficients, then

$$\mathbf{c} \sim \mathcal{N}(\mathbf{y}, \boldsymbol{\Sigma}) \quad (12a)$$

295 where

$$y_i = \int_{\Omega} \tilde{\mu}(\theta, \varphi) \mathcal{Y}_{l_i m_i}(\theta, \varphi) d\Omega \quad (12b)$$

$$\Sigma_{ij} = \iint_{\Omega^2} \mathcal{Y}_{l_i m_i}(\theta, \varphi) \tilde{k}(\theta, \varphi; \theta', \varphi') \mathcal{Y}_{l_j m_j}(\theta', \varphi') d\Omega d\Omega' \quad (12c)$$

296 and where we have used (l_i, m_i) to indicate the spherical harmonic degree and order appropriate to the i th element of \mathbf{c} .
297

298 Substituting the expressions for $\tilde{\mu}$ and \tilde{k} from eq. (7) into these integrals, we obtain

$$y_i = \int_{\Omega} \mu(\mathbf{x}) \mathcal{Y}_{l_i m_i}(\mathbf{x}) d\Omega + \sum_{pq} f_{l_i m_i}(\hat{\mathbf{x}}_p) \left[\left(\hat{\mathbf{K}} + \mathbf{C}_{\hat{\mathbf{a}}} \right)^{-1} \right]_{pq} (\hat{d}_q - \mu(\hat{\mathbf{x}}_q)) \quad (13a)$$

$$\Sigma_{ij} = g_{l_i m_i l_j m_j} - \sum_{pq} f_{l_i m_i}(\hat{\mathbf{x}}_p) \left[\left(\hat{\mathbf{K}} + \mathbf{C}_{\hat{\mathbf{a}}} \right)^{-1} \right]_{pq} f_{l_j m_j}(\hat{\mathbf{x}}_q) \quad (13b)$$

299 where we have introduced f and g to represent certain integrals of the covariance function,

$$f_{lm}(\theta, \varphi) = \int_{\Omega} k(\theta, \varphi; \theta', \varphi') \mathcal{Y}_{lm}(\theta', \varphi') d\Omega' \quad (14a)$$

$$g_{lm l' m'} = \int_{\Omega} f_{lm}(\theta, \varphi) \mathcal{Y}_{l' m'}(\theta, \varphi) d\Omega \quad (14b)$$

300 Naïvely, one might anticipate evaluating these numerically. However, this is impractical,
301 since integration over the surface of the sphere requires a double integral, and thus g represents a quadruple integral. For our approach to be tractable, it is essential that we find
302 analytic reductions of these integrals.
303

304 By design, we adopted epicentral angle as the distance measure within our covariance
305 function. As a result, the function k can be regarded as a function of epicentral angle,
306 $k(\theta, \varphi; \theta', \varphi') \rightarrow k(\cos(\Theta))$. Recognising this allows us to write it as an expansion in terms
307 of Legendre polynomials,

$$k(\cos(\Theta)) = \sum_{l=0}^{\infty} a_l P_l(\cos(\Theta)) \quad (15)$$

308 where, again, no approximation is involved provided the sum is infinite. Any coefficient in
309 this expansion can be computed,

$$a_l = \frac{2l+1}{2} \int_0^{\pi} k(\cos \Theta) P_l(\cos \Theta) \sin \Theta d\Theta \quad (16)$$

310 and this integral can be evaluated numerically for any l .

311 However, by the spherical harmonic addition theorem, we can also express the Legendre
312 polynomials as a sum of spherical harmonics,

$$P_l(\cos(\Theta)) = \frac{4\pi}{2l+1} \sum_{m=-l}^l \mathcal{Y}_{lm}(\theta, \varphi) \mathcal{Y}_{lm}(\theta', \varphi'). \quad (17)$$

313 Combining eqs. (14–17) and exploiting the orthogonality properties given in eq. (10), we
314 therefore have

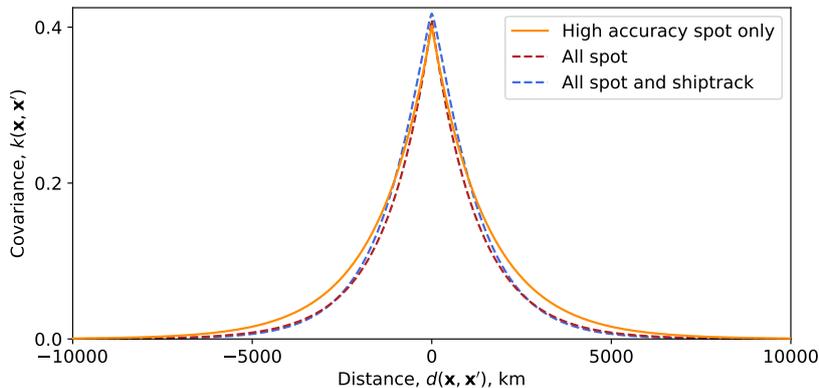
$$f_{lm}(\theta, \varphi) = 2\pi \mathcal{Y}_{lm}(\theta, \varphi) \int_0^{\pi} k(\cos \Theta) P_l(\cos \Theta) \sin \Theta d\Theta \quad (18a)$$

$$g_{lm l' m'} = 2\pi \delta_{ll'} \delta_{mm'} \int_0^{\pi} k(\cos \Theta) P_l(\cos \Theta) \sin \Theta d\Theta \quad (18b)$$

315 We are now in a position to evaluate \mathbf{y} and Σ , and, thus, characterise the distribution
316 describing \mathbf{c} . We highlight that no approximations have been made in determining these

Table 1. Optimal hyperparameters determined for each dataset. Quantities are defined as in the main text.

	μ_0 (km)	Δ (km)	σ_1 (km)	σ_2 (rad)	ν
High accuracy spot only	-0.03	–	0.63	0.24	0.49
All spot	-0.06	0.08	0.64	0.20	0.49
All spot and shiptrack	-0.01	0.09	0.65	0.20	0.69

**Figure 2.** Covariance properties of datasets. Transects through the function $k(\mathbf{x}, \mathbf{x}')$, plotted as a function of the distance between \mathbf{x} and \mathbf{x}' . All three datasets display similar properties.

317 expressions, and our results are not influenced by spectral leakage caused by truncation of
 318 a spherical harmonic expansion at some particular degree.

319 Finally, to obtain a spectrum for any random realisation of \mathbf{c} , we follow Hoggard et
 320 al. (2016) and define the power at degree l to be

$$p_l = \sum_{m=-l}^l c_{lm}^2. \quad (19)$$

321 Since the coefficients c_{lm} are Gaussian-distributed, p_l follows a generalised χ^2 -distribution.
 322 It should be noted that this has a number of counter-intuitive properties, including the fact
 323 that the expected value of the power is typically higher than the power of the expected set
 324 of coefficients.

325 4 Results and Discussion

326 4.1 Dataset properties and models of residual topography

327 We begin by considering the 1160 high-accuracy spot measurements only. We deter-
 328 mine the hyperparameters μ_0 , σ_1 , σ_2 and ν that maximise the inherent likelihood of this
 329 dataset (subject to the assumptions underpinning our approach), as discussed in Section 3.1.
 330 Values are given in Table 1; in particular, we find the characteristic length-scale over which
 331 each data-point is informative to be around 1500 km. The resulting covariance function
 332 $k(\mathbf{x}, \mathbf{x}')$ is shown in Fig. 2.

333 We then condition the Gaussian Process on the observed data, to obtain a model for
 334 residual topography. The posterior mean elevation for this high-accuracy dataset is mapped

335 in Fig. 3(a), with uncertainties shown in Fig. 3(b). We observe a spatial pattern of residual
 336 topography that is consistent with that reported by Davies et al. (2019), dominated by
 337 broad topographic highs within the Pacific, African and North Atlantic regions, separated
 338 by a band of topographic lows extending from Antarctica, through the Americas to the
 339 Arctic, broadening beneath the Eurasian continent and extending south of Australia. Un-
 340 surprisingly, given the sparse and uneven data coverage, this model has large uncertainties
 341 in many regions of the globe, including continental interiors and substantial sections of the
 342 oceans.

343 In any Bayesian approach, the posterior distribution arises as a compromise between
 344 information obtained from the data, and information imposed by the prior distribution. To
 345 help quantify where the dataset is informative, it is useful to measure the difference between
 346 prior and posterior distributions at any point in our model. To do this, we employ a common
 347 similarity measure for probability distributions, the Kullback-Leibler divergence (Kullback
 348 & Leibler, 1951),

$$D_{KL}(h(\mathbf{x})\|\tilde{h}(\mathbf{x})) = \int_{-\infty}^{\infty} \rho_h(u) \log \frac{\rho_{\tilde{h}}(u)}{\rho_h(u)} du \quad (20)$$

349 where ρ_h and $\rho_{\tilde{h}}$ represent the probability density functions associated with the distribu-
 350 tions $h(\mathbf{x})$ and $\tilde{h}(\mathbf{x})$, respectively. Since the natural logarithm is used, information gain is
 351 expressed in units of ‘nats’ (compare ‘bits’ for base-2 information measures). Where the
 352 Kullback-Leibler divergence is low, our posterior distribution remains rather similar to our
 353 prior distribution, and so our inference has not taught us much. This is mapped for the
 354 high-accuracy spot measurements in Fig. 3(c), and in many regions — particularly conti-
 355 nental interiors, the Pacific Ocean and the Southern Ocean — there is little information
 356 available from data. This must be borne in mind when interpreting the residual topography
 357 maps.

358 We repeat this analysis, extending the dataset to include the additional 870 spot
 359 measurements for which crustal corrections are not available. As discussed in Section 3.1, we
 360 increase the uncertainties associated with these data points by an amount Δ , and determine
 361 $\Delta = 0.08$ km as part of the hyperparameter optimisation process. This is somewhat less
 362 than the correction adopted by Hoggard et al. (2017) (who proposed $\Delta = 0.2$ km, based on
 363 the median crustal correction applied to the high-accuracy points). Other hyperparameter
 364 values are as listed in Table 1: these are generally similar to those obtained using the
 365 high-accuracy points, with the characteristic length-scale determined to be ~ 1300 km. The
 366 resulting covariance function is depicted in Fig. 2, and is consistent with that obtained using
 367 the high-accuracy data alone.

368 Conditioning the Gaussian Process on the observed data, we obtain the residual to-
 369 pography model shown in Fig. 4. Again, this is generally similar to that obtained using the
 370 high-accuracy data, although the additional constraints reveal new features in some regions
 371 (e.g. the Western Pacific, where some shorter wavelength structure is apparent). Again, we
 372 report uncertainties and the information gain between prior and posterior, and these should
 373 be considered carefully before any detailed interpretation of model features.

374 Finally, we extend the dataset once more, incorporating the 20767 measurements de-
 375 rived from shiptrack bathymetry. Again, to reflect the fact that no crustal corrections are
 376 applied to these data points, we increase measurement uncertainties by the amount Δ . We
 377 redetermine a value for this using the expanded dataset, obtaining $\Delta = 0.09$ km, and other
 378 hyperparameters as given in Table 1. Given the significant expansion of the dataset, these
 379 are remarkably consistent with those determined from the spot data alone. Only the order
 380 parameter ν changes appreciably. However, Fig. 2 illustrates that this has minimal effect
 381 on the form of the covariance function.

382 A residual topography model derived from the full dataset is shown in Fig. 5. Unsur-
 383 prisingly, given the ten-fold increase in the number of measurements used to construct the
 384 model, we see fine-scale detail that is not visible using the spot data alone. Uncertainties

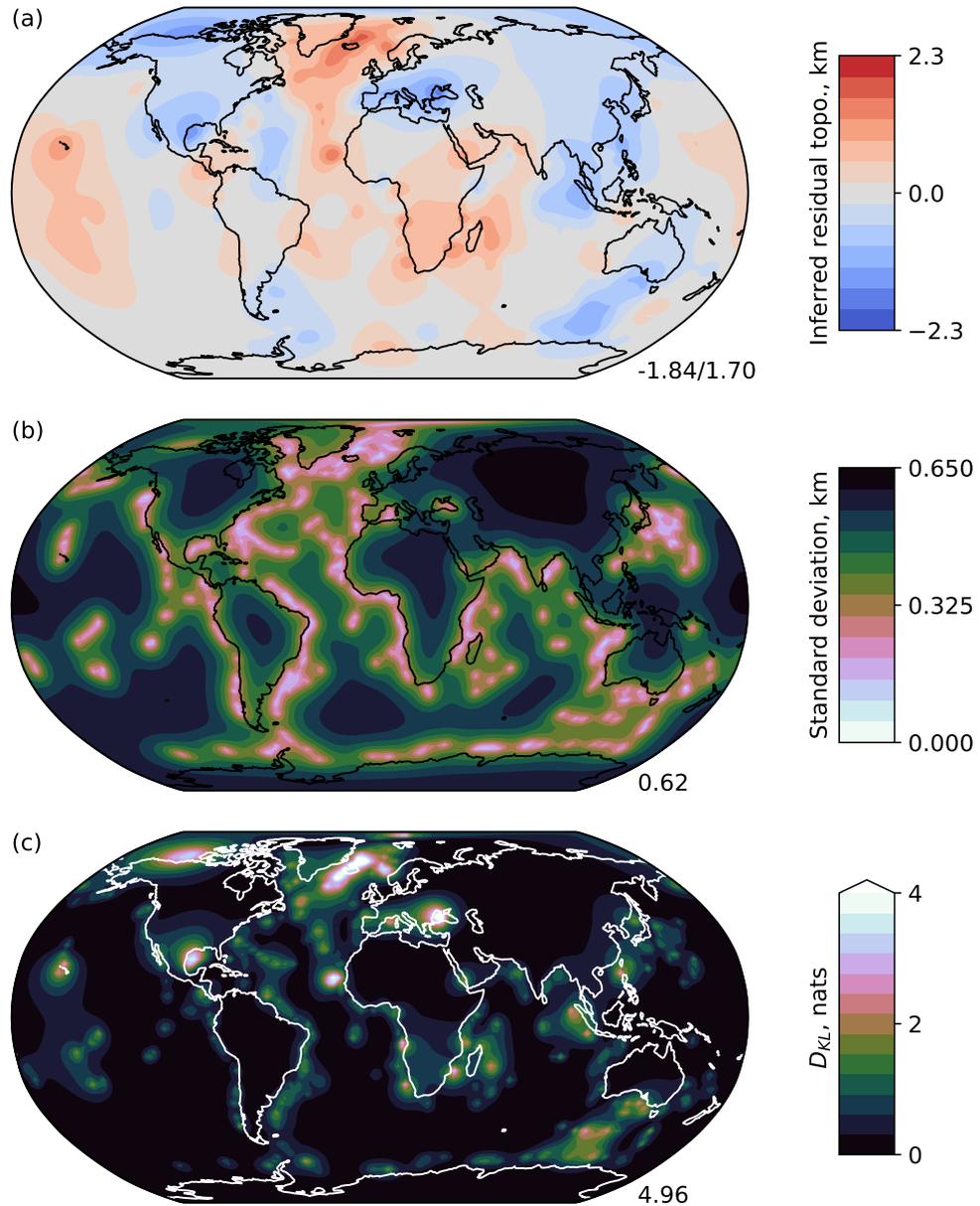


Figure 3. Residual topography determined from high-accuracy spot measurements only. Panel (a) shows the posterior mean point-wise estimate of residual topography, $\tilde{\mu}(\mathbf{x})$, with (b) mapping the associated standard deviation, $\sqrt{\tilde{k}(\mathbf{x}, \mathbf{x})}$. In (c), we show the point-wise information gain from prior to posterior, $D_{KL}(\mathbf{x})$, in units of ‘nats’ (see text). Colour scales are chosen for consistency with subsequent figures; the maximum values for each map are given in the lower-right corner. Note that uncertainties grow rapidly away from measurement locations, and there are large regions (including all continental interiors) where the spot data is uninformative.

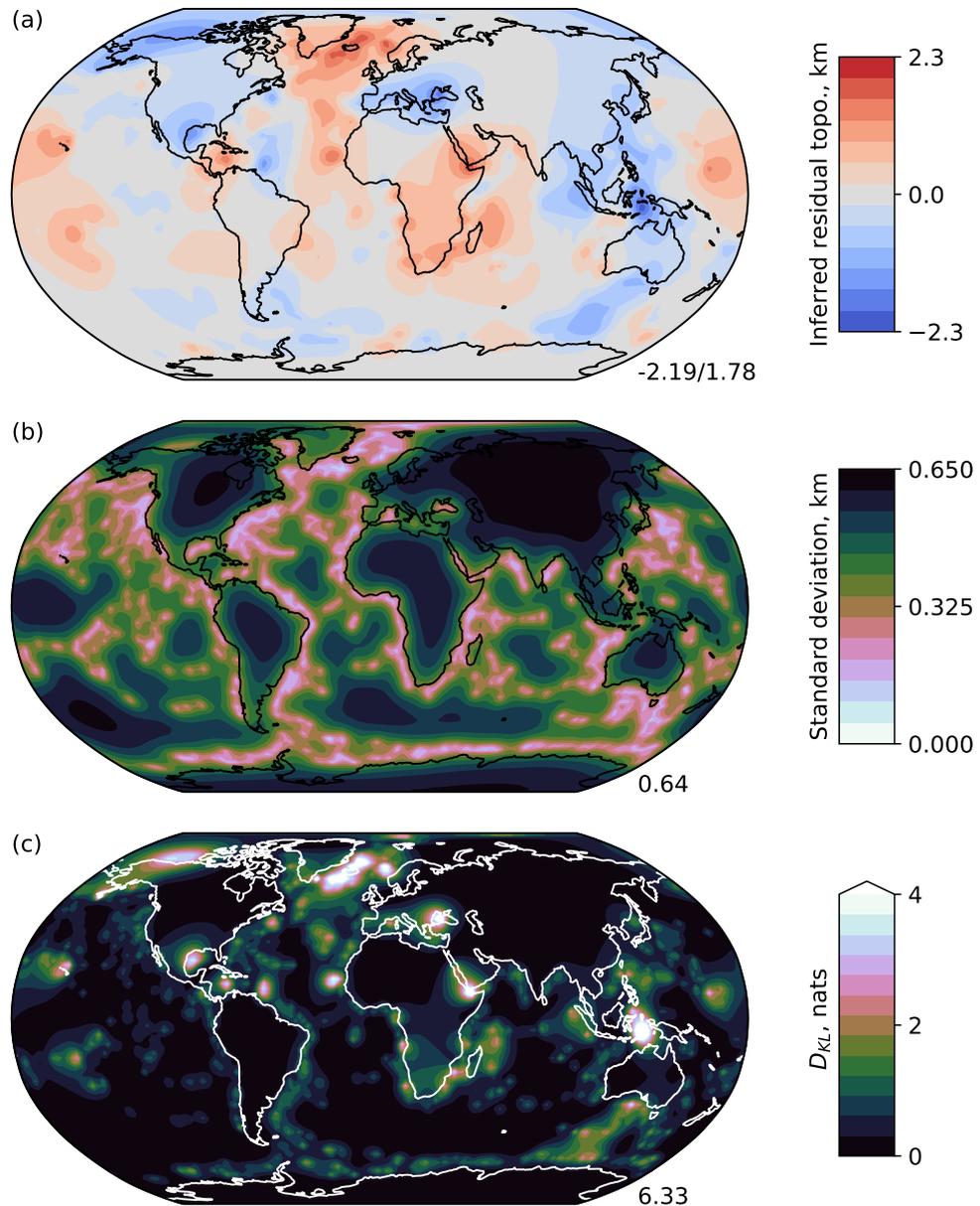


Figure 4. Residual topography determined from spot data, including the 1160 high-accuracy points and a further 870 measurements lacking a crustal correction, for which an additional uncertainty has been estimated. As in Fig. 3, we show (a) the mean residual topography, (b) the standard deviation in this estimate, and (c) the pointwise information gain.

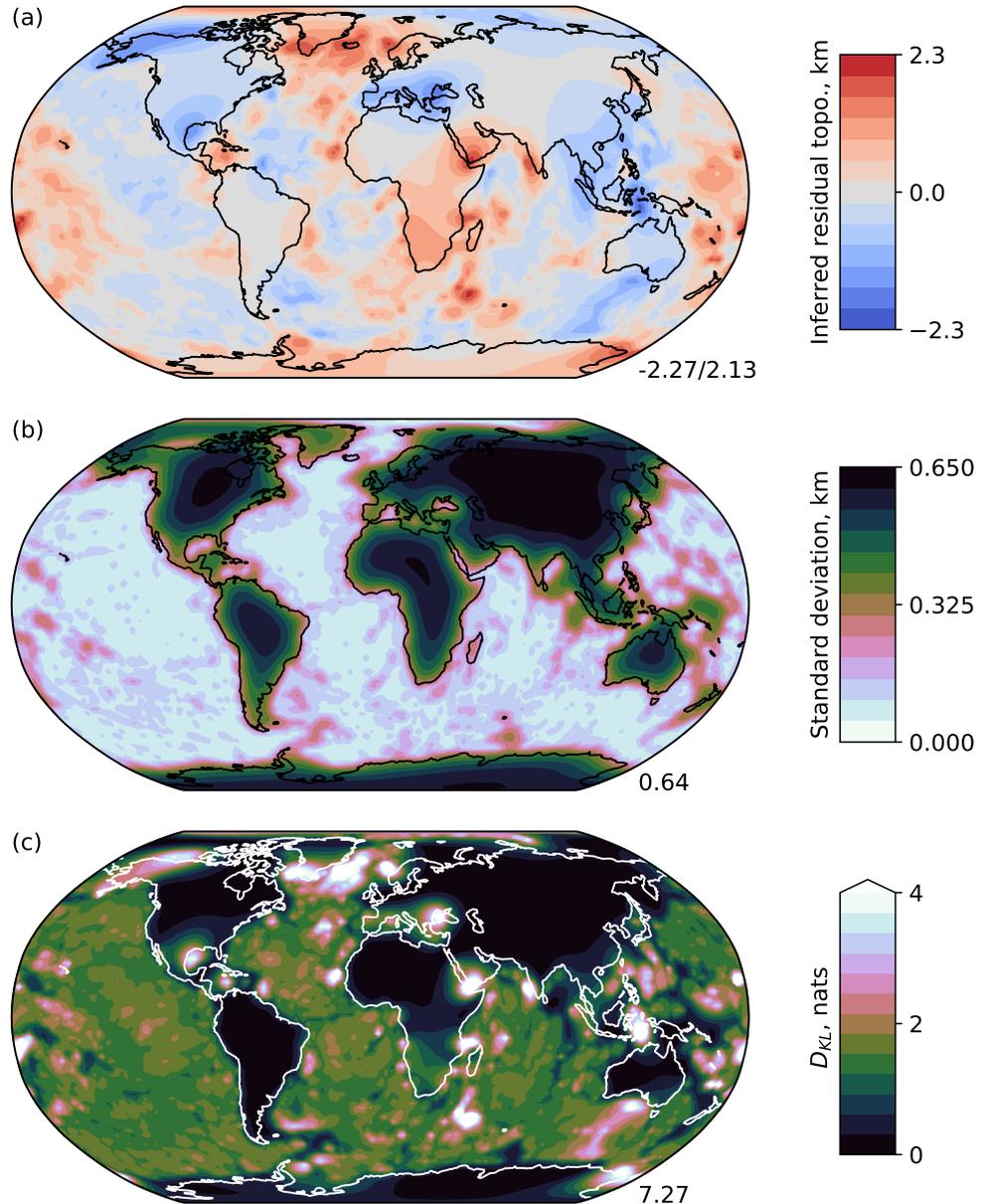


Figure 5. Residual topography determined from all available data, including both spot- and shiptrack-derived measurements. As in Figs 3 and 4, we show (a) the mean residual topography, (b) the standard deviation in this estimate, and (c) the pointwise information gain. The use of shiptrack-derived data substantially expands the area in which the model is informative.

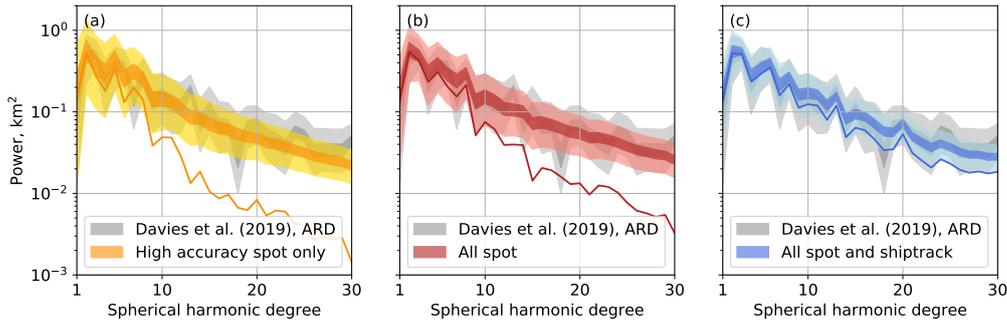


Figure 6. The power spectrum of residual topography: (a) spectrum obtained using high-accuracy spot measurements only; (b) spectrum obtained using all spot measurements; and (c) spectrum obtained using spot and shiptrack data. All panels also show results from Davies et al. (2019) (derived from spot and shiptrack data) for comparison. Solid lines represent the power spectrum of the mean spherical harmonic coefficients; coloured bands represent the ranges spanned by the central 50% and 99% of spectra computed for 100,000 random samples from the distribution of spherical harmonic coefficients. In general, the results obtained in this paper are in agreement with those of Davies et al. (2019).

385 are markedly reduced throughout the oceans, particularly in the North Atlantic. As expected,
 386 the data is informative throughout the oceans, but the lack of onshore data results
 387 in significant uncertainties within continental interiors.

388 4.2 Power spectra

389 Having obtained residual topography models based on the three data subsets, we next
 390 express each in terms of spherical harmonics and compute power spectra. These are shown in
 391 Fig. 6, presented in a format consistent with Hoggard et al. (2016) and Davies et al. (2019),
 392 allowing for straightforward comparison. To represent the (non-Gaussian) uncertainties as-
 393 sociated with these spectra, we generate power spectra for 100,000 random samples from the
 394 posterior residual topography models, and Fig. 6 depicts the ranges spanned by the central
 395 99% and 50% of samples. An alternative presentation of the same data is shown in Fig. 7,
 396 allowing direct comparison between the three data sets at each spherical harmonic degree,
 397 and some representative numerical values are provided in Table 2. Files containing the ex-
 398 pansion coefficients for each data set are included in the software repository accompanying
 399 this paper; see the ‘data availability statement’ for details.

400 In general, all three datasets tell a similar story, which is consistent with that reported
 401 by Davies et al. (2019) (results also shown in Fig. 6). The spectrum of residual topography
 402 is relatively flat, peaking at or around degree-2 (wavelength ~ 16000 km), with steadily-
 403 declining power at shorter length scales. Based on the high-accuracy spot data, the most-
 404 probable model has degree-2 power 0.53 km^2 , although the data could support power up to
 405 $\sim 1.3 \text{ km}^2$. By degree-10 (wavelength ~ 4000 km) and degree-20 (~ 2000 km), the power is
 406 likely in the ranges $0.12\text{--}0.18 \text{ km}^2$ and $0.04\text{--}0.06 \text{ km}^2$, respectively. In general, the additional
 407 information available in the expanded datasets (all spot data; all spot and shiptrack data)
 408 enables a modest reduction in the spectral uncertainty, but does not substantially alter the
 409 most-probable power. Based on our random samples, we find that degrees 1–3 account for
 410 about 85% of the total power in the residual topography field below degree-30 (high
 411 accuracy spot data only: 85.9%; all spot data: 83.9%; spot and shiptrack data: 84.2%).

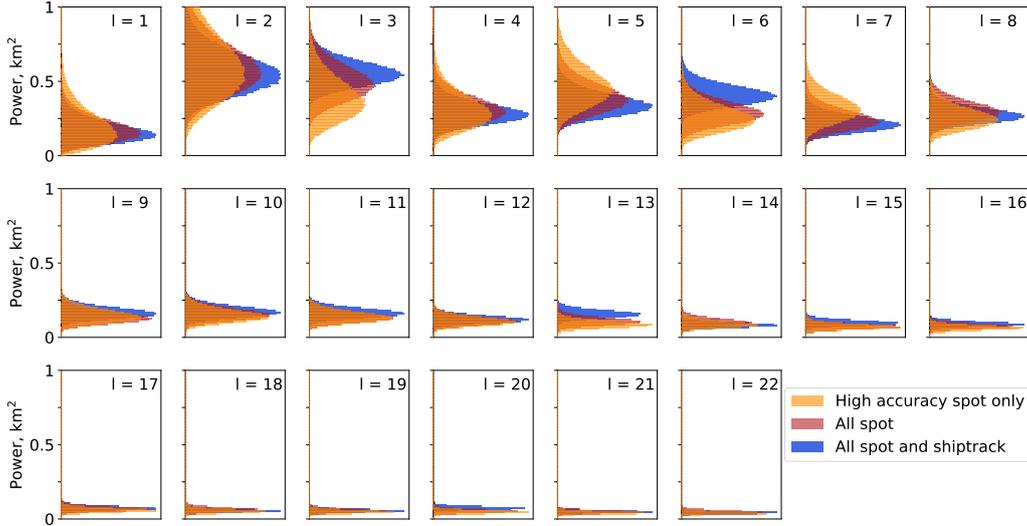


Figure 7. Uncertainties in the power spectrum of residual topography. An alternative representation of the data in Fig. 6: histograms of the power per degree for 100,000 random samples from the posterior distribution of spherical harmonic coefficients obtained using each dataset. Note that the vertical (power) scale here is linear, whereas in Fig. 6 a logarithmic scale is used, to match figures elsewhere in the literature. Beyond $l = 22$ differences between the two datasets are negligible on the scale of this figure.

412 By generating random models consistent with the data, we can also quantify the
 413 typical height variations associated with residual topography. Some representative figures
 414 are given in Table 3. Based on the high-accuracy data, we see that the maximum amplitude
 415 of residual topography is likely to be in the range 2.12–2.39 km, with an upper limit of
 416 around 3 km. This is consistent with the range of measurements depicted in Fig. 1. When
 417 filtered at the longest wavelengths (spherical harmonic degrees 1–3), models consistent with
 418 the data typically have maximum amplitudes in the range 680–850 m, with an upper limit of
 419 ~ 1.2 km. The expanded datasets reduce these ranges somewhat: the full spot and shiptrack
 420 data suggests maximum amplitudes at these wavelengths of 740–820 m, and an upper limit
 421 of ~ 1 km.

422 4.3 A synthetic test

423 The consistency of results across datasets is a strong indication that we are obtaining
 424 models that provide a useful representation of the truth. To further validate our approach,
 425 we perform a synthetic test. We take a synthetic residual topography field generated from
 426 a global model of instantaneous mantle flow (fully-described in Davies et al. (2019); we here
 427 use the simulation ‘with shallow structure’, which incorporates shallow-density heterogeneity
 428 and thermal structure based upon an estimate of lithospheric thickness, and has been shown
 429 to have characteristics that are consistent with the observational data). This simulated field
 430 is first sampled at the 1160 locations corresponding to our high-accuracy measurements. We
 431 use this dataset to construct a global Gaussian Process model, which — as already discussed
 432 — enables us to generate the mean and standard deviation characterising our prediction
 433 for residual topography at any point on Earth’s surface. We then sample our synthetic
 434 residual topography field at 116062 points drawn from a regular mesh within the oceans,
 435 and compare this synthetic ground truth data with our Gaussian Process predictions for
 436 these 116062 locations. For each location, we compute the difference between the mean

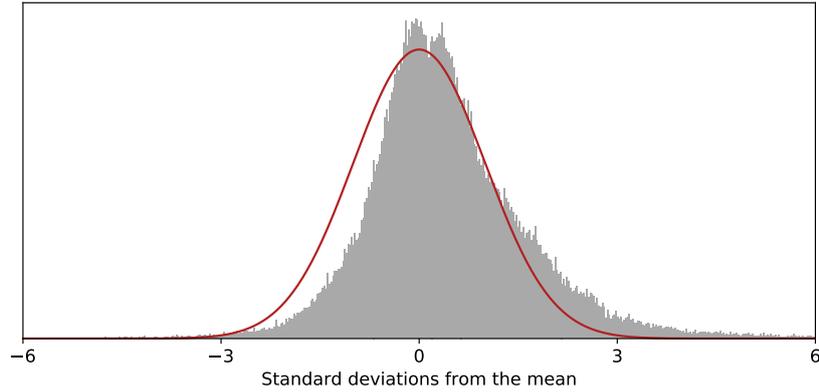


Figure 8. A synthetic test: predictive performance of Gaussian Process residual topography model constructed using simulated data. Output from a global simulation of instantaneous mantle flow from Davies et al. (2019) is used to generate a synthetic residual topography dataset at the same 1160 locations as in our ‘high accuracy’ dataset. Applying our Gaussian Process approach, we can then predict the residual topography on a regular mesh of 116062 locations throughout the oceans, and associated uncertainties. We then compare these to the ‘true’ simulation-derived value at each point, expressing any prediction error relative to the uncertainty. In the ideal case, results (grey bars) should follow a standard normal distribution (red line).

437 predicted topography and the ‘true’ value, and then divide this by the standard deviation
 438 associated with that prediction. Treating performance at each location as independent, this
 439 quantity should follow a standard normal distribution (i.e., one with zero mean and unit
 440 variance) if our model were perfect.

441 A histogram of the resulting data is shown in Fig. 8, overlain by a standard normal
 442 distribution. As one might anticipate, the two do not match exactly, with the histogram
 443 somewhat skewed. Nevertheless, it is clear that the Gaussian Process model provides good
 444 predictive performance: the truth lies within one standard deviation of the predicted mean
 445 in 65.4% of cases, and within three standard deviations in 96.2%. For a true normal distri-
 446 bution, these figures should be 68.3% and 99.7% respectively.

447 A number of factors can be identified that contribute to this discrepancy. First, it
 448 is not strictly correct to treat the performance at adjacent locations as independent; we
 449 ought to account for the fact that they are correlated. However, neglecting this greatly
 450 simplifies the analysis, and likely better-represents the manner in which users will approach
 451 maps such as those shown in Figs. 3–5. Second, in treating the residual topography field as
 452 being representable using a Gaussian Process, we are making a strong assumption about its
 453 statistical properties. This is pragmatically justifiable — one cannot analyse any dataset
 454 without making assumptions — but is almost certainly imperfect.

455 A third issue may lie within the hyperparameter optimisation procedure. Implicit
 456 in this is an assumption that the available data is representative of the field as a whole.
 457 Unfortunately, as with most geoscience datasets, our sample locations are not randomly-
 458 distributed around the globe: they are influenced by a variety of practical, economic and
 459 scientific constraints. In the present case, the points within the high-accuracy dataset are
 460 governed by the availability of seismic surveys. Inevitably, such surveys are targeted towards
 461 particular geological environments — resulting in a bias towards older oceanic crust that
 462 abuts continental margins. As a result, there is a significant possibility that our observations

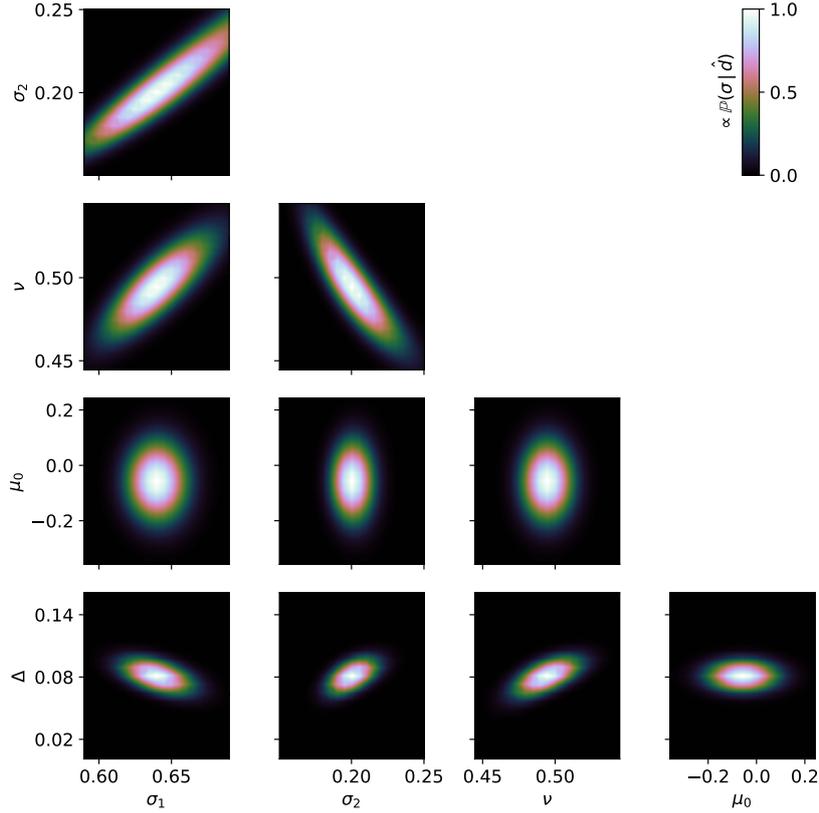


Figure 9. Hyperparameter tradeoffs, computed using all spot data. Plots show the unnormalised likelihood of residual topography measurements, eq. (6), as pairs of hyperparameters are varied in the vicinity of the optimal set (see Table 1); in each plot, all other hyperparameters are fixed to their optimal value. All hyperparameters are fairly tightly-constrained. Relationships between σ_1 , σ_2 , ν and μ_0 computed using the high-accuracy spot data are very similar to those seen here; the parameter Δ is not relevant to that dataset.

463 present a biased picture of the overall field. This is the case in our synthetic example:
 464 although the average global residual topography is zero, the average residual topography at
 465 our 1160 sample points is -422 m (globally, the simulation has amplitudes that are somewhat
 466 larger than we expect for the Earth, lying in the range -4.95 – 3.39 km). This sampling bias
 467 likely explains (at least partially) the skewness seen in Fig. 8. It is impossible to fully assess
 468 whether similar biases are present in the real datasets used elsewhere in this paper (although
 469 the consistency across all three datasets suggests that the issue is not as pronounced as in
 470 this synthetic test). We stress that any such biases depend heavily on the detail of the
 471 residual topography field, and these synthetic results should not be taken to indicate that
 472 the models in Figs. 3–5 are similarly skewed towards underestimating topography.

473 **4.4 Hyperparameter tradeoffs and interpretability**

474 In Fig. 9, we show slices through the hyperparameter likelihood function, computed
 475 using all spot data. We see that the maximum-likelihood point is well-constrained, implying
 476 that any uncertainty associated with the values presented in Table 1 is small. There are
 477 clear correlations: σ_1 is correlated with both σ_2 and ν , while σ_2 is anti-correlated with

478 ν . The mean function hyperparameter, μ_0 , appears to be largely independent of others.
 479 Essentially identical patterns are seen with the high-accuracy spot data alone; we use all
 480 spot data here since this also illuminates the uncertainty hyperparameter, Δ . This is also
 481 seen to have some weak trade-offs with the hyperparameters σ_1 , σ_2 and ν . It is not feasible
 482 to make similar plots using the full spot and shiptrack dataset, due to the computational
 483 costs involved.

484 To what extent are these hyperparameters interpretable? In particular, it is notable
 485 that the uncertainty correction, Δ , is markedly less than that identified Hoggard et al.
 486 (2017), who adopted $\Delta = 0.2$ km. This raises the question: is our value ‘more correct’?
 487 In addressing this, it is essential to remember that our hyperparameters are optimal in a very
 488 specific sense: they maximise eq. 6 for the available data, and are predicated on the various
 489 assumptions inherent to our approach. These inevitably represent a simplified version of
 490 reality, and one can readily propose alternative assumptions that are — at a minimum —
 491 equally plausible. End users must carefully consider whether our choices are appropriate to
 492 their application, and thus whether it makes sense to rely on our hyperparameter values or
 493 other results.

494 In the present work, we effectively assume that we have no prior knowledge about the
 495 hyperparameters, and that all values are equally plausible. If desired, it would be straight-
 496 forward to modify the hyperparameter determination procedure to incorporate an explicit
 497 ‘hyperprior’, expressing *a priori* expectations or preferences. This is discussed in more de-
 498 tail in, for example, Valentine and Sambridge (2018). However, for the present problem,
 499 we do not see a route to constructing a hyperprior that is objectively more justifiable than
 500 the uniform hyperprior implicit in our approach. Nevertheless, we emphasise that the pri-
 501 mary goal of the present work is the analysis of the high-accuracy spot data, for which the
 502 uncertainty hyperparameter Δ is not relevant.

503 4.5 The future: obtaining new measurements

504 Although the first order characteristics of the power spectra obtained here are consis-
 505 tent across all data sets, there is no doubt that tighter constraints on the exact spatial pat-
 506 tern, wavelength and amplitude of residual topography require extending the high-accuracy
 507 spot measurements across the globe: our knowledge of residual topography is limited by the
 508 amount of data available to us. However, obtaining high-quality observations is expensive
 509 and time-consuming. A pertinent question is therefore: where would new data be most
 510 useful? This question takes us into the field of ‘optimal experimental design’ (e.g. Curtis,
 511 1999), and the Gaussian Process approach allows it to be addressed straightforwardly.

512 As discussed in Section 3.1, our knowledge of residual topography after seeing the data
 513 is given by $\tilde{h}(\mathbf{x}) \sim \mathcal{GP}(\tilde{\mu}(\mathbf{x}), \tilde{k}(\mathbf{x}, \mathbf{x}'))$. Suppose we subsequently obtain one more data point
 514 at some location \mathbf{x}_0 , measuring value $d_0 \pm \sigma_d$. Incorporating this additional information into
 515 our model results in an updated estimate of topography, $\bar{h}(\mathbf{x}) \sim \mathcal{GP}(\bar{\mu}(\mathbf{x}), \bar{k}(\mathbf{x}, \mathbf{x}'))$, where

$$\bar{\mu}(\mathbf{x}) = \tilde{\mu}(\mathbf{x}) + \alpha(\mathbf{x}_0)\tilde{k}(\mathbf{x}, \mathbf{x}_0) [d_0 - \tilde{\mu}(\mathbf{x}_0)] \quad (21a)$$

$$\bar{k}(\mathbf{x}, \mathbf{x}') = \tilde{k}(\mathbf{x}, \mathbf{x}') - \alpha(\mathbf{x}_0)\tilde{k}(\mathbf{x}, \mathbf{x}_0)\tilde{k}(\mathbf{x}', \mathbf{x}_0) \quad (21b)$$

516 and where we have introduced

$$\alpha(\mathbf{x}_0) = \frac{1}{\tilde{k}(\mathbf{x}_0, \mathbf{x}_0) + \sigma_d^2}. \quad (21c)$$

517 We can then express this updated model in terms of spherical harmonics, as before. This
 518 yields a new distribution for the coefficients, $\bar{\mathbf{c}} \sim \mathcal{N}(\bar{\mathbf{y}}, \bar{\mathbf{\Sigma}})$, where

$$\bar{y}_i = y_i + \alpha(\mathbf{x}_0)z_i(\mathbf{x}_0) [d_0 - \tilde{\mu}(\mathbf{x}_0)] \quad (22a)$$

$$\bar{\Sigma}_{ij} = \Sigma_{ij} - \alpha(\mathbf{x}_0)z_i(\mathbf{x}_0)z_j(\mathbf{x}_0) \quad (22b)$$

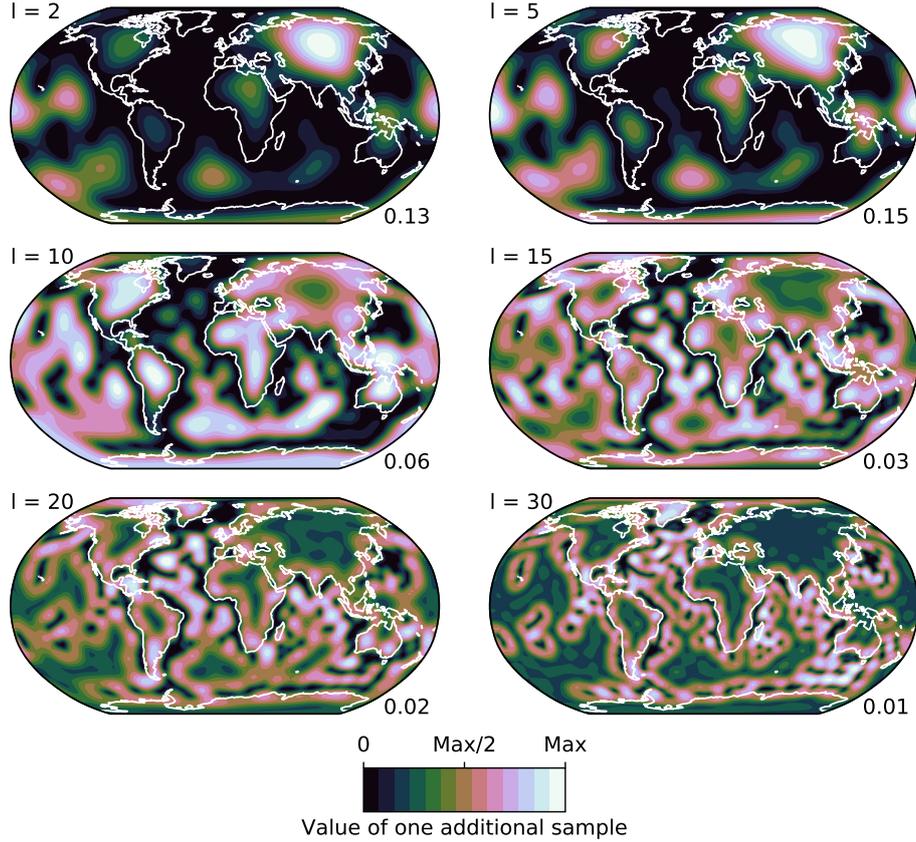


Figure 10. Where should data collection efforts be focused? Map of expected information gained (see eq. 24) by collecting one additional sample at that point, broken down by spherical harmonic degree, for the 1160-point high-accuracy spot dataset. Each map is normalised by its maximum value, given in the lower-right corner. Compare with Fig. 3: unsurprisingly, the most-useful places to collect new information correspond to regions of high uncertainty in the recovered topography model.

519 with

$$z_i(\mathbf{x}) = f_{l_i m_i}(\mathbf{x}) - \sum_{pq} f_{l_i m_i}(\hat{\mathbf{x}}_p) \left[\left(\hat{\mathbf{K}} + \mathbf{C}_a \right)^{-1} \right]_{pq} k(\mathbf{x}, \hat{\mathbf{x}}_q). \quad (22c)$$

520

521 Using the Kullback-Leibler divergence (cf. eq. 20), we can now quantify how much the
 522 single additional observation has revealed about the spherical harmonic coefficients, given
 523 by

$$D_{KL} = \frac{1}{2} \left[\text{Tr}(\bar{\Sigma}^{-1} \Sigma) - D + \log \frac{\det \bar{\Sigma}}{\det \Sigma} \right] + \frac{1}{2} \alpha^2(\mathbf{x}_0) \mathbf{z}(\mathbf{x}_0)^T \bar{\Sigma}^{-1} \mathbf{z}(\mathbf{x}_0) [d - \tilde{\mu}(\mathbf{x}_0)]^2, \quad (23)$$

524

525 where D represents the dimension of the vector $\bar{\mathbf{c}}$. We therefore see that the amount we learn
 526 depends on the value of d : if our additional observation happens to match what we predicted
 527 based on the original dataset ($\tilde{\mu}(\mathbf{x}_0)$) we learn less than if the observation reveals unexpected
 528 new features. Using $\tilde{h}(\mathbf{x})$, we can determine that the expected value (in a statistical sense)
 of $(d - \tilde{\mu})^2$ is $\tilde{k}(\mathbf{x}_0, \mathbf{x}_0)$ and, hence, we can quantify the expected information gain due to

529 an additional observation at \mathbf{x}_0 by

$$G(\mathbf{x}_0) = \frac{1}{2} \left[\text{Tr}(\bar{\Sigma}^{-1}\Sigma) - D + \log \frac{\det \bar{\Sigma}}{\det \Sigma} \right] + \frac{1}{2} \alpha^2(\mathbf{x}_0) \tilde{k}(\mathbf{x}_0, \mathbf{x}_0) \mathbf{z}(\mathbf{x}_0)^T \bar{\Sigma}^{-1} \mathbf{z}(\mathbf{x}_0). \quad (24)$$

530 This quantity is straightforward to compute, and it is therefore possible to produce maps
 531 showing the expected value of one additional sample at any point on the globe. Alternatively,
 532 for a more conservative perspective, one might map only the first, parenthesised term in
 533 eq. (24): this would represent the minimum information gain from a sample at any point.
 534 In particular, one can make such maps considering only the spherical harmonic coefficients
 535 corresponding to a particular angular order l , to yield an understanding of how samples in
 536 different regions might help constrain different aspects of the residual topography spectrum.

537 In Fig. 10, we map the expected information gain (i.e., the complete form of $G(x_0)$)
 538 based on the high-accuracy spot data. We assume that the one additional data point is
 539 measured with an uncertainty of $\sigma_d = 0.1$ km. We clearly see that our knowledge of the
 540 low-degree spherical harmonic components can be improved through addition of samples
 541 at a relatively small number of locations: continental interiors (especially in Russia), the
 542 Southern Ocean, and the South and West Pacific. To constrain shorter wavelengths, our
 543 analysis indicates that the most useful sampling locations are in regions close to existing
 544 data points: this makes sense, as our approach is predicated on collecting only a single
 545 sample.

546 5 Conclusions

547 In this paper, we have successfully demonstrated a new approach for obtaining a con-
 548 tinuous, global model for residual topography given a finite set of point-wise measurements.
 549 Our method is based upon the statistical theory of Gaussian Processes (Valentine & Sam-
 550 bridge, 2020a, 2020b), and has a number of important advantages over those employed
 551 in previous studies, particularly: (i) it does not presuppose that the residual topography
 552 field is well-represented using a particular set of basis functions, avoiding potential issues
 553 such as ‘spectral leakage’; (ii) prior information (i.e. the covariance function) is defined
 554 within physical space, rather than in an abstract ‘model space’ where its effects may be
 555 poorly-understood; and (iii) the method remains robust for even the smallest datasets.

556 These properties enable us to construct a residual topography model that is based
 557 only on the 1160 high-accuracy spot measurements in the database compiled by Hoggard et
 558 al. (2017). These points are believed to be particularly robust, since they are obtained via
 559 bespoke analysis of seismic and other datasets, and incorporate location-specific information
 560 about crustal properties. The resulting model is broadly in agreement with those reported in
 561 earlier studies, which have typically relied on a much larger dataset derived from automated
 562 processing of shiptrack bathymetry. Applying our method to this larger dataset continues
 563 to yield self-consistent results.

564 The global pattern of residual topography is dominated by broad topographic highs
 565 in the North Atlantic, African and central Pacific regions, with the largest amplitudes adja-
 566 cent to Iceland. These are separated by broad topographic lows, extending from Antarctica,
 567 through the Americas, beneath the Eurasian continent and to the south of Australia. Based
 568 on just the 1160 high-accuracy measurements, the associated spherical harmonic power spec-
 569 trum peaks at $l = 2$, with power likely to lie in the range 0.46–0.76 km². This decreases by
 570 over an order of magnitude to the range 0.02 km² at $l = 30$. The overall form of the spectrum
 571 is consistent across all data subsets analysed. In light of the sensitivity kernels linking topo-
 572 graphic deflections to density anomalies within the mantle at different depths and spherical
 573 harmonic degree (e.g. Richards & Hager, 1984; Colli et al., 2016), and as discussed in detail
 574 by Davies et al. (2019), generating this spectrum requires Earth’s topographic response to be
 575 driven by multi-scale convective processes. The long-wavelength portion of the spectrum is

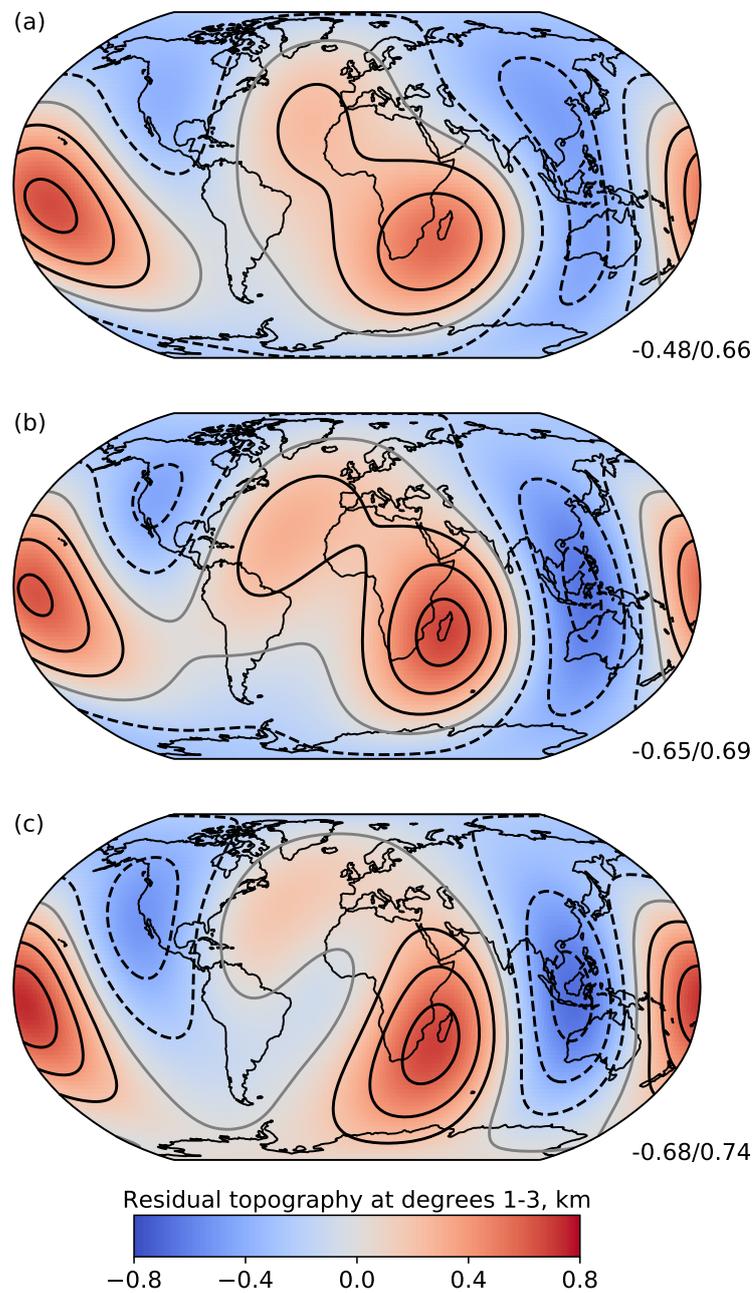


Figure 11. Long-wavelength residual topography. Models (as in Figs. 3–5) filtered to include only spherical harmonic degrees 1–3. We see consistent patterns across all three datasets: (a) high-accuracy spot measurements; (b) all spot measurements; and (c) all spot and shiptrack measurements. Contours are drawn at 200 m intervals, and the amplitude range in each plot is shown, bottom right. Negative contours are dashed. Zero contour is in grey.

576 primarily attributable to deep mantle flow, while shorter-wavelength features are supported
577 by flow associated with structural heterogeneity in the shallow mantle and lithosphere.

578 Given the link between the long-wavelength components of residual topography and
579 deep mantle flow, the patterns and amplitudes associated with low-order spherical harmonics
580 provide an independent constraint on the long-wavelength components of global seismic to-
581 mography images and the dynamical interpretation of these images. Accordingly, in Fig. 11,
582 we show our best-fitting models for each data set, filtered to contain only degrees 1–3. The
583 spatial pattern is fairly consistent across all three datasets, with highs centred off South-East
584 Africa that extend into the North Atlantic, and in the central and western Pacific Ocean.
585 As set out in Table 3, models consistent with the observations may have maximum residual
586 topography heights up to around 1.2 km at these wavelengths, although a figure in the range
587 680–850 m is most likely. The residual topography highs strongly resemble the shape and
588 extent of the large low velocity provinces that have been consistently imaged in the deep
589 mantle across a number of global seismic tomography studies (e.g. Becker & Boschi, 2002;
590 Ritsema et al., 2011; Bozdağ et al., 2016), and likely represent concentrations of hot (and
591 potentially chemically-distinct) material that drives upwelling flow towards Earth’s surface
592 (e.g. McNamara & Zhong, 2005; Schubert et al., 2009; Bull et al., 2009; Davies et al.,
593 2012; Flament et al., 2013; Davies, Goes, & Sambridge, 2015; Garnero et al., 2016). These
594 long-wavelength residual topography highs are separated by a band of residual topography
595 lows, extending from Antarctica, through the Americas, the Arctic, Eurasia and Australia.
596 These are likely related to downwelling slabs, associated with former subduction zones (e.g.
597 Richards & Engebretson, 1992; Bunge et al., 2002; Flament et al., 2013; Davies, Goes, &
598 Lau, 2015; Rubey et al., 2017).

599 Finally, from a methodological perspective, the Gaussian Process-based approach for
600 inferring global models from point data appears attractive. It is mathematically-elegant and,
601 as we have demonstrated, produces results that are consistent with basis-function expansion
602 strategies. Indeed, as discussed in detail in Valentine and Sambridge (2020b), the two have
603 close links, becoming equivalent under certain conditions as the number of basis functions
604 tends to infinity. We highlight that our approach can be straightforwardly translated to the
605 analysis of other datasets and problems, and suggest that it may prove powerful in cases
606 where models must be constrained using spatially-sparse data.

607 Data availability statement

608 The specific data files used for this study are as reported in Davies et al. (2019), and
609 may be obtained from https://github.com/drhodrid/Davies_etal_NGeo_2019_Datasets.
610 Code to analyse this data and produce the results presented here may be obtained from
611 <https://github.com/valentineap/DynamicTopographyGP> (doi: 10.5281/zenodo.3895317).
612 This repository also includes files containing the spherical harmonic expansion coefficients
613 associated with our mean models.

614 Acknowledgments

615 We are grateful to Mark Hoggard for making his dataset available, and for many help-
616 ful discussions. APV acknowledges a Discovery Early Career Research Award and other
617 support from the Australian Research Council (ARC: grant numbers DE180100040 and
618 DP200100053). DRD also acknowledges support from the ARC under grants DP170100058
619 and DP200100053. This work has made use of computational resources operated by the
620 NCI National Facility in Canberra, which is supported by the Australian Commonwealth
621 Government.

622 References

623 Becker, T. W., & Boschi, L. (2002). A comparison of tomographic and geodynamic mantle

- 624 models. *Geochem. Geophys. Geosys.*, *3*, 2001GC000168. doi: 10.129/2001GC000168
- 625 Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., ... Pugmire,
626 D. (2016). Global adjoint tomography: first-generation model. *Geophysical Journal*
627 *International*, *207*, 1739–1766.
- 628 Braun, J. (2010). The many surface expressions of mantle dynamics. *Nature Geoscience*,
629 *3*, 826–833. doi: 10.1038/ngeo1020
- 630 Bull, A. L., McNamara, A. K., & Ritsema, J. (2009). Synthetic tomography of plume
631 clusters and thermochemical piles. *Earth Planet. Sci Lett.*, *278*, 152–156. doi: 10.1016/
632 j.epsl.2008.11.018
- 633 Bunge, H.-P., Richards, M. A., & Baumgardner, J. R. (2002). Mantle circulation models
634 with sequential data-assimilation: inferring present-day mantle structure from plate
635 motion histories. *Phil. Trans. R. Soc. London, Set. A*, *360*, 2545–2567. doi: 10.1098/
636 rsta.2002.1080.
- 637 Byrd, R., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound
638 constrained optimization. *SIAM Journal on Scientific Computing*, *16*, 1190–1208.
- 639 Colli, L., Ghelichkhan, S., & Bunge, H.-P. (2016). On the ratio of dynamic topography
640 and gravity anomalies in a dynamic Earth. *Geophys. Res. Lett.*, *43*, 2510–2516. doi:
641 10.1002/2016GL067929
- 642 Conrad, C., & Husson, L. (2009). Influence of dynamic topography on sea level and its rate
643 of change. *Lithosphere*, *1*, 110–120. doi: 10.1130/L32.1
- 644 Curtis, A. (1999). Optimal design of focused experiments and surveys. *Geophysical Journal*
645 *International*, *139*, 205–215.
- 646 Czarnota, K., Hoggard, M. J., White, N., & Winterbourne, J. (2013). Spatial and tempo-
647 ral patterns of Cenozoic dynamic topography around Australia. *Geochem. Geophys.*
648 *Geosys.*, *14*, 634–658. doi: 10.1029/2012GC004392
- 649 Czarnota, K., Roberts, G. G., White, N. J., & Fishwick, S. (2014). Spatial and temporal
650 patterns of Australian dynamic topography from river profile modelling. *J. Geophys.*
651 *Res.*, *119*, 1384–1419. doi: 10.1002/2013JB010436
- 652 Dahlen, T., & Tromp, J. (1998). *Theoretical global seismology*. Princeton: Princeton
653 University Press.
- 654 Davies, D. R., Goes, S., Davies, J. H., Schubert, B. S. A., Bunge, H., & Ritsema, J.
655 (2012). Reconciling dynamic and seismic models of Earth’s lower mantle: the dominant
656 role of thermal heterogeneity. *Earth Planet. Sci. Lett.*, *353*, 253–269. doi: 10.1016/
657 j.epsl.2012.08.016
- 658 Davies, D. R., Goes, S., & Lau, H. C. P. (2015). Thermally dominated deep mantle LLSVPs:
659 A review. In A. Khan & F. Deschamps (Eds.), *The Earth’s Heterogeneous Mantle*
660 (p. 441–477). Springer International Publishing. doi: 10.1007/978-3-319-15627-9_14
- 661 Davies, D. R., Goes, S., & Sambridge, M. (2015). On the relationship between volcanic
662 hotspot locations, the reconstructed eruption sites of large igneous provinces and deep
663 mantle seismic structure. *Earth Planet. Sci. Lett.*, *411*, 121–130. doi: 10.1016/j.epsl
664 .2014.11.052
- 665 Davies, D. R., Valentine, A., Kramer, S., Rawlinson, N., Hoggard, M., Eakin, C., & Wilson,
666 C. (2019). Earth’s multi-scale topographic response to global mantle flow. *Nature*
667 *Geoscience*, *12*, 845–850.
- 668 Eakin, C., & Lithgow-Bertelloni, C. (2018). An overview of dynamic topography: The
669 influence of mantle circulation on surface topography and landscape. *Mountains,*
670 *Climate and biodiversity*, *37*.
- 671 Flament, N., Gurnis, M., & Muller, R. D. (2013). A review of observations and models of
672 dynamic topography. *Lithosphere*. doi: 10.1130/L245.1
- 673 Garnero, E. J., McNamara, A. K., & Shim, S. (2016). Continent-sized anomalous zones with
674 low seismic velocity at the base of Earth’s mantle. *Nature Geosciences*, *9*, 481–489.
- 675 Guerri, M., Cammarano, F., & Tackley, P. J. (2016). Modelling Earth’s surface topography:
676 decomposition of the static and dynamic components. *Phys. Earth Planet. Int.*, *261*,
677 172–186. doi: 10.1016/j.pepi.2016.10.009

- 678 Gurnis, M. (1993). Phanerozoic marine inundation of continents driven by dynamic topog-
679 raphy above subducting slabs. *Nature*, *364*, 589-593. doi: 10.1038/364589a0
- 680 Gurnis, M., Mitrovica, J. X., Ritsema, J., & van Heijst, H. J. (2000). Constraining mantle
681 density structure using geological evidence of surface uplift rates: the case of the
682 African superplume. *Geochem. Geophys. Geosys.*, *1*, 1999GC000035.
- 683 Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P., & Dziewonski, A. M. (1985).
684 Lower mantle heterogeneity, dynamic topography and the geoid. *Nature*, *313*, 541-
685 545. doi: 10.1038/313541a0
- 686 Hager, B. H., & Richards, A. M. (1989). Long-wavelength variations in Earth's geoid:
687 Physical models and dynamical implications. *Phil. Trans. Roy. Soc. London, Ser. A.*,
688 *328*, 309-327.
- 689 Hoggard, M., White, N., & Al-Attar, D. (2016). Global dynamic topography observations
690 reveal limited influence of large-scale mantle flow. *Nature Geoscience*, *9*, 456-463.
- 691 Hoggard, M., Winterbourne, J., Czarnota, K., & White, N. (2017). Oceanic residual depth
692 measurements, the plate cooling model, and global dynamic topography. *Journal of*
693 *Geophysical Research*, *122*, 2328-2372.
- 694 Kaban, M. K., Schwintzer, P., Artemieva, I. M., & Mooney, W. D. (2003). Density of the
695 continental roots: Compositional and thermal contributions. *Earth Planet. Sci. Lett.*,
696 *209*, 53-69. doi: 10.1016/S0012-821X(03)00072-4
- 697 Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathe-*
698 *matical Statistics*, *22*, 79-86.
- 699 Laske, G., & Masters, G. A. (1997). A global digital map of sediment thickness. *EOS Trans.*
700 *Am. Geophys. Union*, *78*, F483.
- 701 Lithgow-Bertelloni, C., & Silver, P. G. (1998). Dynamic topography, plate driving forces
702 and the African superswell. *Nature*, *395*, 269-272. doi: 10.1038/26212
- 703 Mackay, D. (1992). Bayesian interpolation. *Neural Computation*, *4*, 415-447.
- 704 McNamara, A. K., & Zhong, S. (2005). Thermo-chemical structures beneath Africa and
705 the Pacific Ocean. *Nature*, *437*, 1136-1139. doi: 10.1038/nature04066
- 706 Mitrovica, J. X., Beaumont, C., & Jarvis, G. T. (1989). Tilting of continental in-
707 teriors by the dynamical effects of subduction. *Tectonics*, *8*, 1079-1094. doi:
708 10.1029/TC008i005p01079
- 709 Murphy, K. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT
710 Press.
- 711 Panasyuk, S. V., & Hager, B. H. (2000). Models of isostatic and dynamic topography,
712 geoid anomalies, and their uncertainties. *J. Geophys. Res.*, *105*, 28199-28209. doi:
713 10.1029/2000JB900249
- 714 Parsons, B., & Daly, S. (1983). The relationship between surface topography, gravity
715 anomalies, and temperature structure of convection. *J. Geophys. Res.*, *88*, 1129-1144.
716 doi: 10.1029/JB088iB02p01129
- 717 Pekeris, C. L. (1935). Thermal convection in the interior of the Earth. *Geophys. J.*, *3*,
718 343-367.
- 719 Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. Cambridge,
720 USA: MIT Press.
- 721 Richards, M. A., & Engebretson, D. C. (1992). Large-scale mantle convection and the
722 history of subduction. *Nature*, *355*, 437-440. doi: 10.1029/2007JB005155
- 723 Richards, M. A., & Hager, B. H. (1984). Geoid anomalies in a dynamic Earth. *J. Geophys.*
724 *Res.*, *89*, 5987-6002.
- 725 Ritsema, J., Deuss, A., van Heijst, H., & Woodhouse, J. (2011). S40RTS: a degree-40
726 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic
727 traveltimes and normal-mode splitting function measurements. *Geophysical Journal*
728 *International*, *184*, 1223-1236.
- 729 Rubey, M., Brune, S., Heine, C., Davies, D. R., Williams, S. E., & Muller, R. D. (2017).
730 Global patterns in Earth's dynamic topography since the Jurassic: the role of sub-
731 ducted slabs. *Solid Earth*, *8*, 899-919. doi: 10.5194/se-8-899-2017
- 732 Schubert, B., Bunge, H.-P., Steinle-Neumann, G., Moder, C., & Oeser, J. (2009). Thermal

- 733 versus elastic heterogeneity in high-resolution mantle circulation models with pyrolite
734 composition: high plume excess temperatures in the lowermost mantle. *Geochemistry,*
735 *Geophysics, Geosystems*, 10.
- 736 Shephard, G. E., Muller, R. D., Liu, L., & Gurnis, M. (2010). Miocene drainage reversal of
737 the Amazon River driven by plate–mantle interaction. *Nature Geoscience*, 3, 870–875.
738 doi: 10.1038/NNGEO1017
- 739 Smith, W. H. F., & Sandwell, D. T. (1997). Global sea floor topography from satellite
740 altimetry and ship depth soundings. *Science*, 277, 1956–1962.
- 741 Steinberger, B. (2016). Topography caused by mantle density variations: observation-based
742 estimates and models derived from tomography and lithosphere thickness. *Geophys.*
743 *J. Int.*, 205, 604–621. doi: 10.1093/gji/ggw040
- 744 Steinberger, B., Conrad, C. P., Osei Tutu, A., & Hoggard, M. J. (2019). On the amplitude
745 of dynamic topography at spherical harmonic degree two. *Tectonophysics*, In Press. doi:
746 10.1016/j.tecto.2017.11.032
- 747 Trampert, J., & Snieder, R. (1996). Model estimations biased by truncated expansions:
748 Possible artifacts in seismic tomography. *Science*, 271, 1257–1260.
- 749 Valentine, A., & Sambridge, M. (2018). Optimal regularization for a class of linear inverse
750 problem. *Geophysical Journal International*, 215, 1003–1021.
- 751 Valentine, A., & Sambridge, M. (2020a). Gaussian process models—I. A framework for
752 probabilistic continuous inverse theory. *Geophysical Journal International*, 220, 1632–
753 1647.
- 754 Valentine, A., & Sambridge, M. (2020b). Gaussian process models—II. Lessons for discrete
755 inversion. *Geophysical Journal International*, 220, 1648–1656.
- 756 Watkins, C. E., & Conrad, C. P. (2018). Constraints on dynamic topography from asym-
757 metric subsidence of the mid-ocean ridges. *Earth Planet. Sci. Lett.*, 484, 264–275. doi:
758 10.1016/j.epsl.2017.12.028
- 759 Winterbourne, J. R., Crosby, A. G., & White, N. J. (2009). Depth, age and dynamic
760 topography of oceanic lithosphere beneath heavily sedimented atlantic margins. *Earth*
761 *Planet. Sci. Lett.*, 287, 137–151.
- 762 Winterbourne, J. R., White, N. J., & Crosby, A. G. (2014). Accurate measurements of
763 residual topography from the oceanic realm. *Tectonics*, 33, 1–34.
- 764 Yang, T., & Gurnis, M. (2016). Dynamic topography, gravity and the role of lateral viscosity
765 variations from inversion of global mantle flow. *Geophysical Journal International*,
766 207, 1186–1202.
- 767 Yang, T., Moresi, L., Müller, R., & Gurnis, M. (2017). Oceanic residual topography agrees
768 with mantle flow predictions at long wavelengths. *Geophysical Research Letters*, 44.