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

A.P. Valentine[1] **and D.R. Davies**[1]

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

Key Points:

- We present a new method for spectral analysis of sparse point data and apply this to measurements of residual topography within the oceans;
- Using a smaller, more reliable dataset than earlier studies, we confirm a power spectrum that is dominated by long-wavelength components;
- We also confirm that the spectrum contains significant power at shorter wavelengths, 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 0.46–0.76 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;

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

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

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

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

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

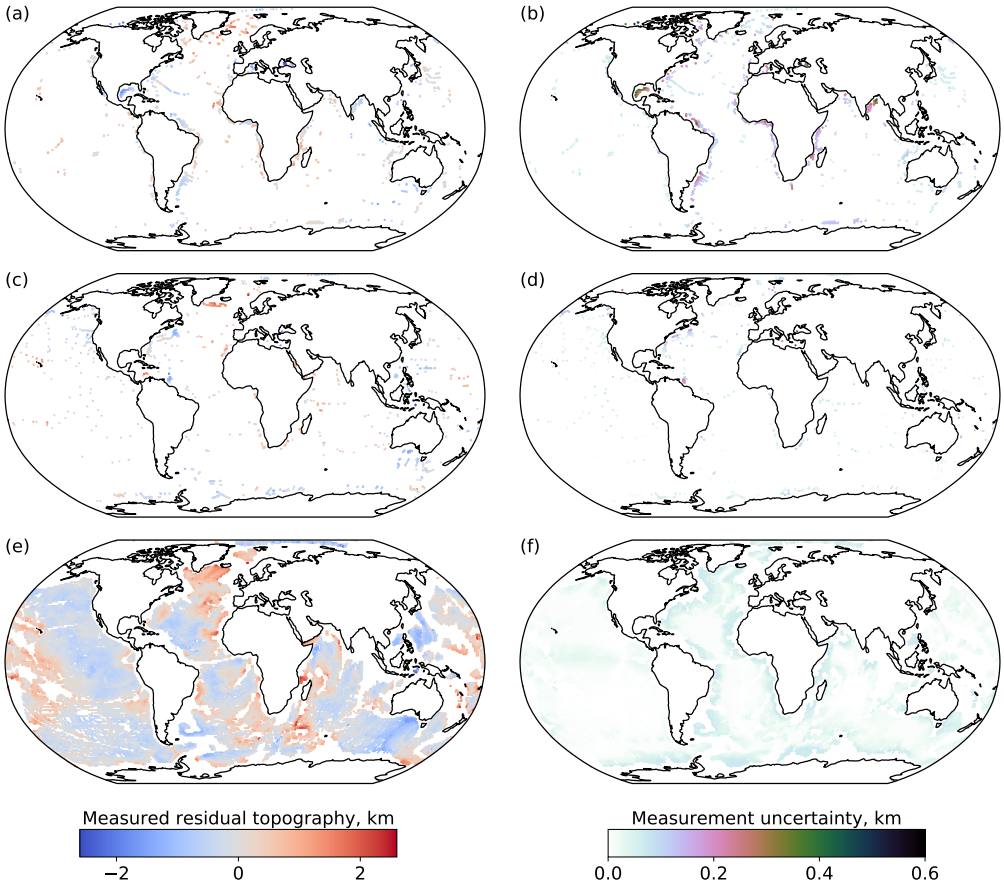

Figure 1. Measurements of residual topography. 1160 high-accuracy spot measurements are shown in (a), with associated uncertainties in (b). A further 870 spot measurements that lack crustal corrections are shown in (c), with uncertainties in (d): these are shown without the additional 200 m uncertainty 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.

2 Datasets

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

2.1 High-accuracy spot measurements

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

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

2.2 Lower-accuracy spot measurements

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

2.3 Shiptrack-derived measurements

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

3 Methodology

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

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

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

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

3.1 A Gaussian Process model for residual topography

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

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

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

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

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

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

$$d(\theta, \varphi; \theta', \varphi') \equiv \Theta(\theta, \varphi; \theta', \varphi') = \arccos\left(\sin\theta \sin\theta' + \cos\theta \cos\theta' \cos(\varphi - \varphi')\right) . \tag{3}$$

Note that this can be converted into the great-circle distance on the surface of the (assumed spherical) Earth by multiplying by Earth's radius. The quantities σ_1, σ_2 and ν in eq. (2) are hyperparameters controlling the detailed shape of the covariance function. Specifically, σ_1 governs its amplitude, σ_2 represents the spatial length-scale over which we expect an observation to be informative, and ν is an order parameter that, essentially, determines the smoothness of the interpolation.

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

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

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

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

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

Now, suppose we are able to measure residual topography at a discrete set of points, $\hat{\mathbf{x}}_{1\ldots N}$ (as in Valentine and Sambridge (2020a), a 'hat' is used to distinguish quantities associated with the observational dataset). Our set of N measurements can be represented as the N-dimensional vector $\hat{\mathbf{d}}$, such that $\hat{d}_i$ represents our estimate of $h(\hat{\mathbf{x}}_i)$, and we assume that all measurement errors are Gaussian, described overall by the $N \times N$ covariance matrix, $\mathbf{C}_{\hat{\mathbf{d}}}$. In the present case, we regard this as a function of the unknown contribution to uncertainty arising from a lack of crustal corrections, Δ. According to the assumptions we have made in setting up our Gaussian Process model, this specific set of observations have likelihood

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

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

Having selected optimal values for $\boldsymbol{\sigma}$, we are now in a position to predict residual topography at unseen locations. To do this, we note that eq. (4) holds for *any* set of locations: in particular, it must hold if we consider the N observed locations and one additional point, $\mathbf{x}$. We can then condition this distribution on the values we actually observed, exploiting the correlations between locations to refine our knowledge about $h(\mathbf{x})$. As explained in Valentine and Sambridge (2020a), our posterior estimate of residual topography is therefore given by

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

where

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

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

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

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

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

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

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

3.2 The power spectrum of residual topography

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

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

Moreover, the spherical harmonics are normalised to each have unit power, and are orthogonal, in the sense that

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

Naïvely, one might anticipate evaluating these numerically. However, this is impractical, 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 analytic reductions of these integrals.

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

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

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

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

and this integral can be evaluated numerically for any l.

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

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

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

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

$$g_{lml'm'} = 2\pi \delta_{ll'} \delta_{mm'} \int_0^\pi k(\cos\Theta) P_l(\cos\Theta) \sin\Theta \, \mathrm{d}\Theta \tag{18b}$$

We are now in a position to evaluate $\mathbf{y}$ and $\mathbf{\Sigma}$, and, thus, characterise the distribution describing $\mathbf{c}$. We highlight that no approximations have been made in determining these expressions, and our results are not influenced by spectral leakage caused by truncation of a spherical harmonic expansion at some particular degree.

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

$$p_l = \sum_{m=-l}^{l} c_{lm}^2 \, . \tag{19}$$

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

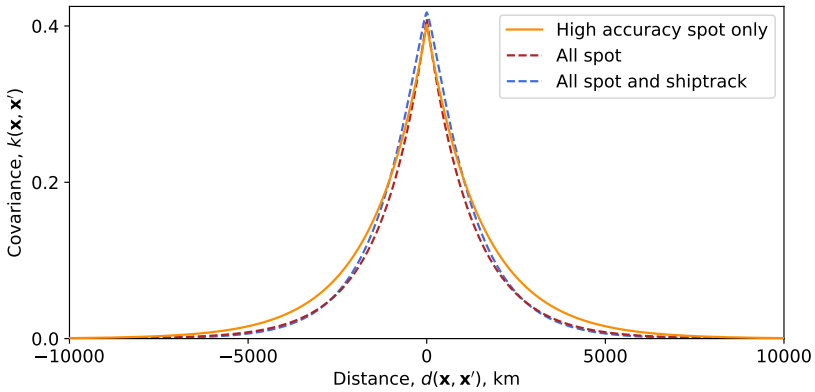

Figure 2. Covariance properties of datasets. Transects through the function $k(\mathbf{x}, \mathbf{x}')$, plotted as a function of the distance between $\mathbf{x}$ and $\mathbf{x}'$. All three datasets display similar properties.

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

4 Results

4.1 Dataset properties and models of residual topography

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

We then condition the Gaussian Process on the observed data, to obtain a model for residual topography. The posterior mean elevation for this high-accuracy dataset is mapped in Fig. 3(a), with uncertainties shown in Fig. 3(b). We observe a spatial pattern of residual topography that is consistent with that reported by Davies et al. (2019), dominated by broad topographic highs within the Pacific, African and North Atlantic regions, separated by a band of topographic lows extending from Antarctica, through the Americas to the Arctic, broadening beneath the Eurasian continent and extending south of Australia. Unsurprisingly, given the sparse and uneven data coverage, this model has large uncertainties in many regions of the globe, including continental interiors and substantial sections of the oceans.

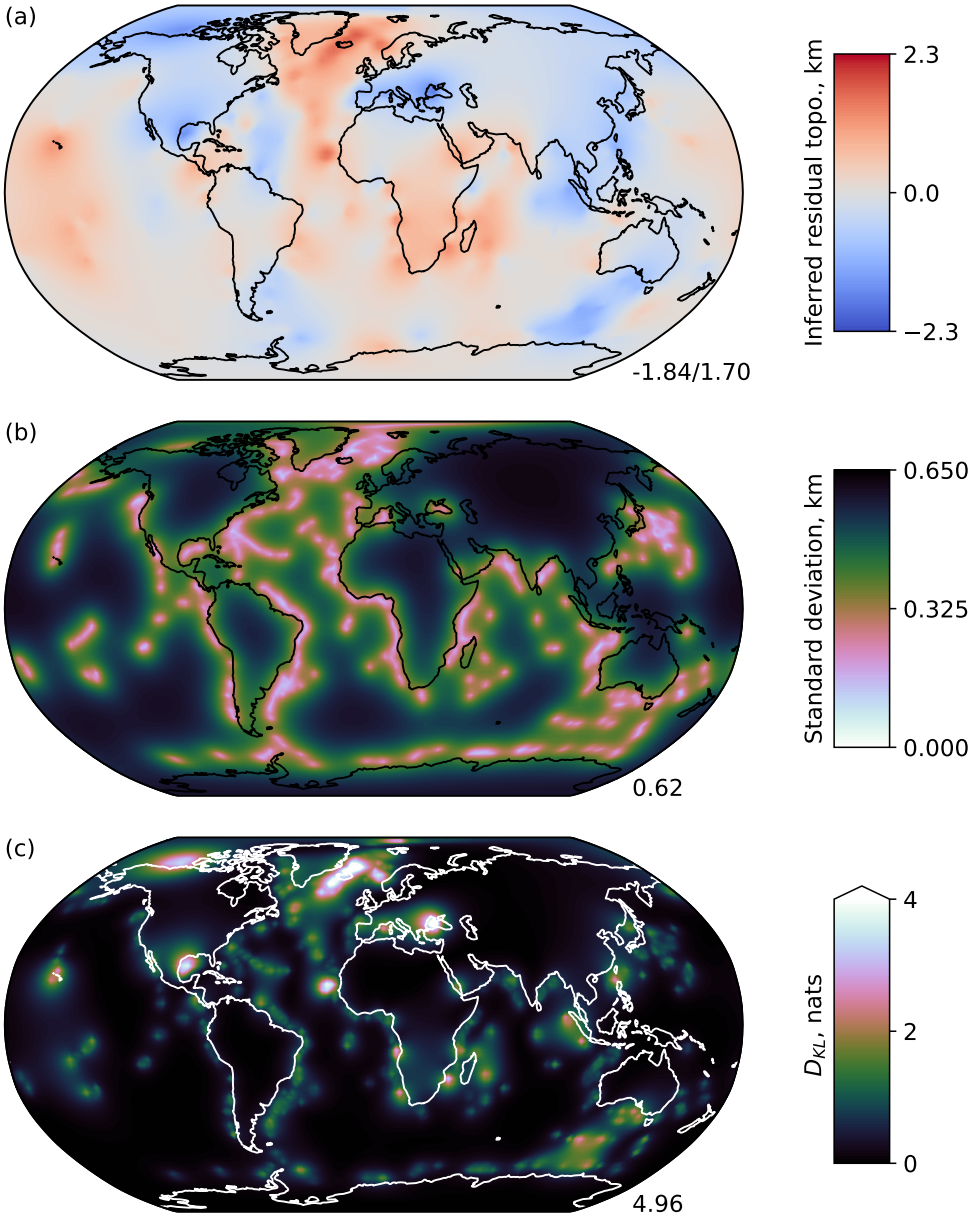

Figure 3. Residual topography determined from high-accuracy spot measurements only. Panel (a) shows the posterior mean point-wise estimate of residual topography, $\tilde{\mu}(\mathbf{x})$, with (b) mapping the associated standard deviation, $\sqrt{\tilde{k}(\mathbf{x},\mathbf{x})}$. In (c), we show 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.

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

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

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

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

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

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

A residual topography model derived from the full dataset is shown in Fig. 5. Unsurprisingly, given the ten-fold increase in the number of measurements used to construct the model, we see fine-scale detail that is not visible using the spot data alone. Uncertainties are markedly reduced throughout the oceans, particularly in the North Atlantic. As expected, the data is informative throughout the oceans, but the lack of onshore data results in significant uncertainties within continental interiors.

4.2 Power spectra

Having obtained residual topography models based on the three data subsets, we next express each in terms of spherical harmonics and compute power spectra. These are shown in Fig. 6, presented in a format consistent with Hoggard et al. (2016) and Davies

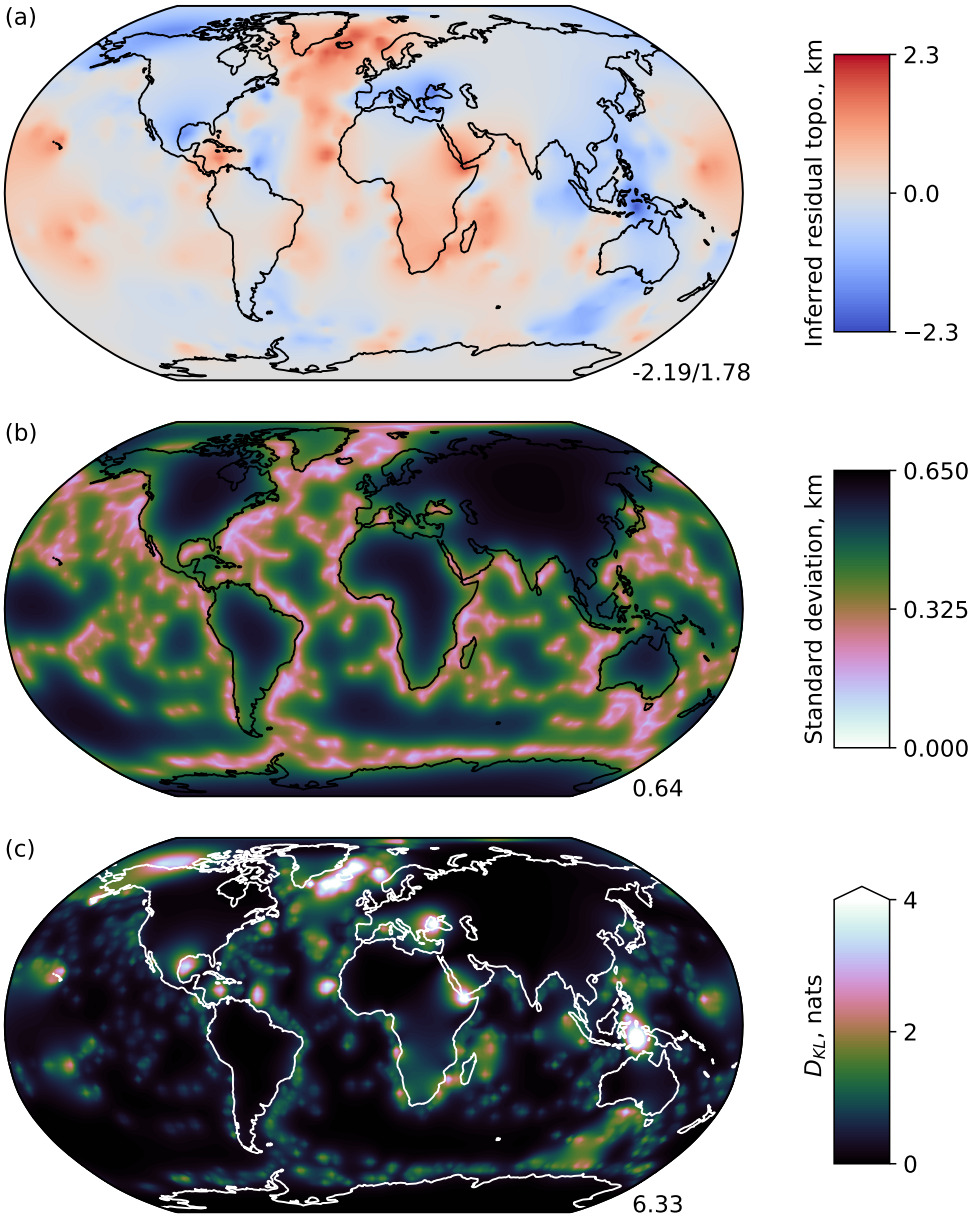

Figure 4. Residual topography determined from spot data, including the 1160 high-accuracy points and a further 870 measurements lacking a crustal correction, for which an additional uncertainty has been estimated. As in Fig. 3, we show (a) the mean residual topography, (b) the standard deviation in this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

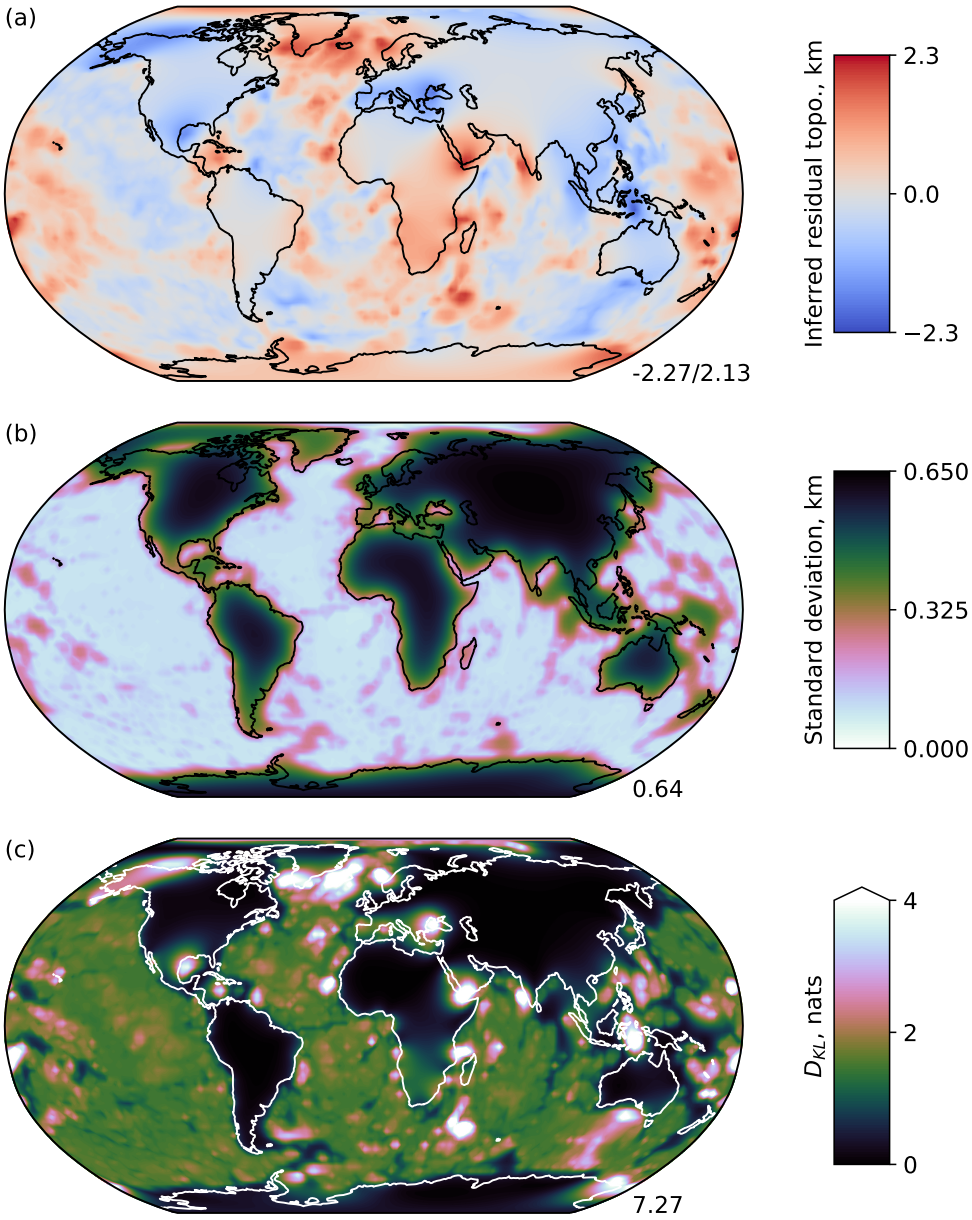

Figure 5. Residual topography determined from all available data, including both spot- and shiptrack-derived measurements. As in Figs 3 and 4, we show (a) the mean residual topography, (b) the standard deviation in this estimate, and (c) the pointwise information gain. The use of shiptrack-derived data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

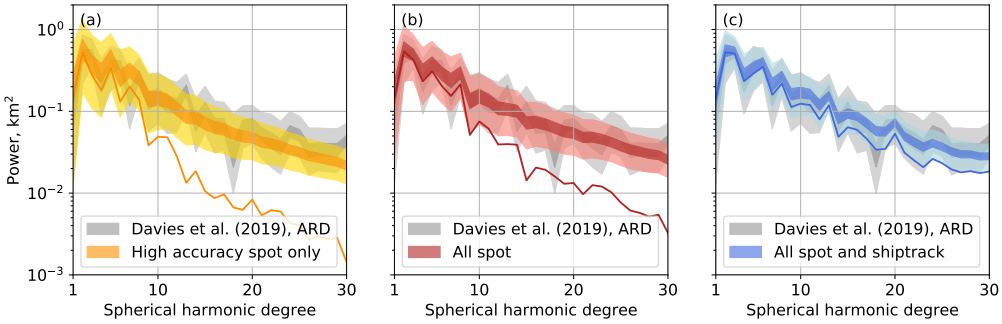

Figure 6. The power spectrum of residual topography: (a) spectrum obtained using high-accuracy spot measurements only; (b) spectrum obtained using all spot measurements; and (c) spectrum obtained using spot and shiptrack data. All panels also show results from Davies et al. (2019) (derived from spot 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).

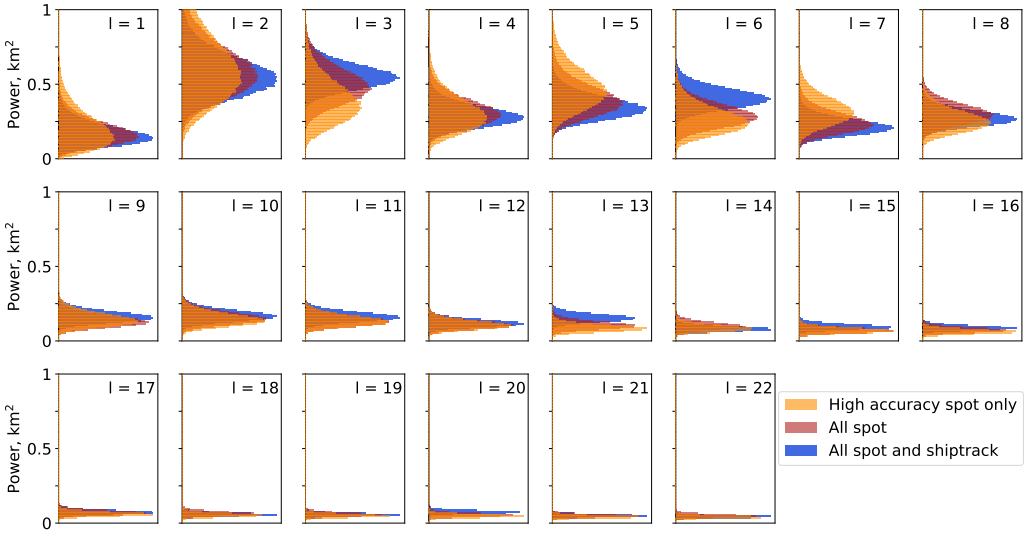

Figure 7. Uncertainties in the power spectrum of residual topography. An alternative representation of the data in Fig. 6: histograms of the power per degree for 100,000 random samples from the posterior distribution of spherical harmonic coefficients obtained using each dataset. Note that the vertical (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.

Table 2. Residual topography power, km^2.

	Degree 2			Degree 5			Degree 10			Degree 20		
	MP[a]	50%[b]	99%[c]	MP[a]	50%[b]	99%[c]	MP[a]	50%[b]	99%[c]	MP[a]	50%[b]	99%[c]
High accuracy spot only	0.53	0.46–0.76	0.17–1.32	0.34	0.38–0.56	0.19–0.89	0.05	0.12–0.18	0.06–0.28	0.01	0.04–0.06	0.03–0.08
All spot	0.54	0.47–0.70	0.23–1.12	0.31	0.32–0.46	0.18–0.71	0.08	0.13–0.18	0.07–0.29	0.01	0.05–0.06	0.03–0.09
All spot & shiptrack	0.52	0.47–0.65	0.27–0.96	0.29	0.29–0.40	0.18–0.58	0.12	0.15–0.20	0.09–0.28	0.05	0.07–0.08	0.05–0.10

[a]Most probable model, as depicted in Figs. 3(a)–5(a).
[b]Range spanned by central 50% of 10^6 samples from the posterior distribution, i.e. the interquartile range.
[c]Range spanned by central 99% of 10^6 samples from the posterior distribution.

Table 3. Residual topography maximum amplitudes, km.

	Full model			Degrees 1–3			Degrees 3–10			Degrees 11–30		
	MP[a]	50%[b]	99%[c]	MP[a]	50%[b]	99%[c]	MP[a]	50%[b]	99%[c]	MP[a]	50%[b]	99%[c]
High accuracy spot only	1.84	2.12–2.39	1.86–2.95	0.66	0.68–0.85	0.49–1.16	1.24	1.27–1.48	1.05–1.87	0.68	1.01–1.14	0.89–1.41
All spot	2.19	2.27–2.54	2.00–3.09	0.69	0.71–0.82	0.57–1.01	1.32	1.35–1.56	1.12–1.93	0.95	1.11–1.26	0.96–1.61
All spot and shiptrack	2.27	2.31–2.51	2.12–3.02	0.74	0.74–0.82	0.65–0.97	1.38	1.42–1.63	1.17–2.03	1.30	1.35–1.48	1.19–1.78

[a]Most probable model, as depicted in Figs. 3(a)–5(a).
[b]Range spanned by central 50% of 10^6 samples from the posterior distribution, i.e. the interquartile range.
[c]Range spanned by central 99% of 10^6 samples from the posterior distribution.

et al. (2019), allowing for straightforward comparison. To represent the (non-Gaussian) uncertainties associated with these spectra, we generate power spectra for 100,000 samples from the posterior residual topography models, and Fig. 6 depicts the ranges spanned by the central 99% and 50% of samples. An alternative presentation of the same data is shown in Fig. 7, allowing direct comparison between the three data sets at each spherical harmonic degree, and some representative numerical values are provided in Table 2.

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

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

4.3 The future: obtaining new measurements

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

As discussed in Section 3.1, our knowledge of residual topography after seeing the data 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 at some location $\mathbf{x_0}$, measuring value $d_0 \pm \sigma_d$. Incorporating this additional information into 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

$$\bar{\mu}(\mathbf{x}) = \tilde{\mu}(\mathbf{x}) + \alpha(\mathbf{x_0})\tilde{k}(\mathbf{x}, \mathbf{x_0})\,[d_0 - \tilde{\mu}(\mathbf{x_0})] \tag{21a}$$

$$\bar{k}(\mathbf{x}, \mathbf{x}') = \tilde{k}(\mathbf{x}, \mathbf{x}') - \alpha(\mathbf{x_0})\tilde{k}(\mathbf{x}, \mathbf{x_0})\tilde{k}(\mathbf{x}', \mathbf{x_0}) \tag{21b}$$

and where we have introduced

$$\alpha(\mathbf{x_0}) = \frac{1}{\tilde{k}(\mathbf{x_0}, \mathbf{x_0}) + \sigma_d^2}\,. \tag{21c}$$

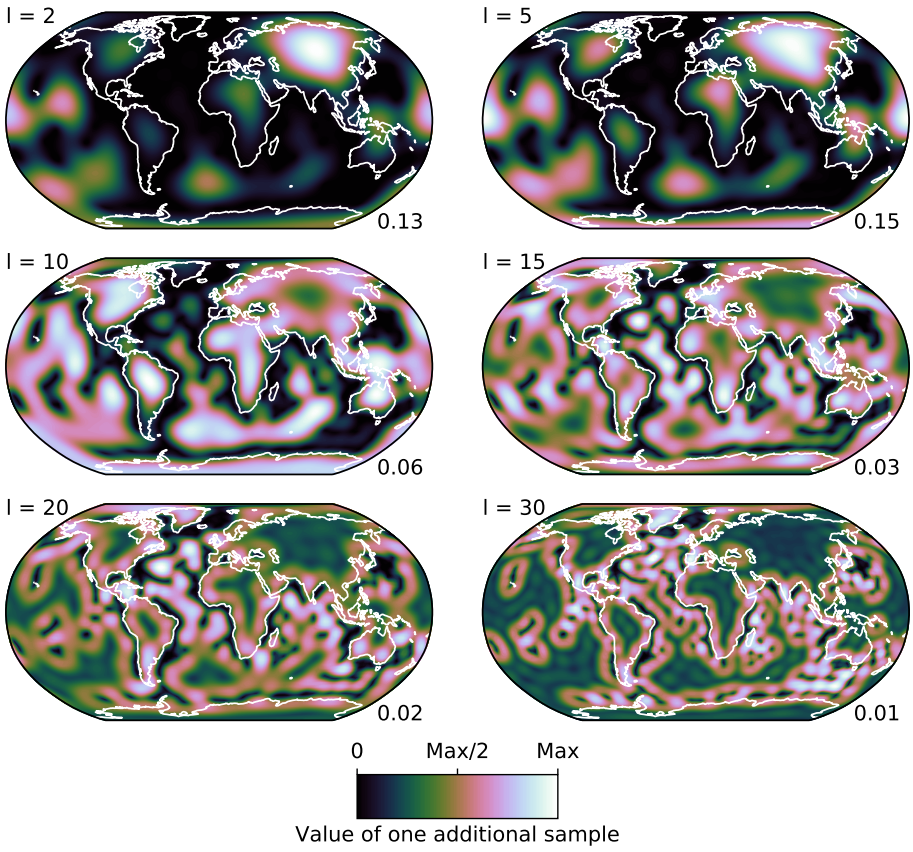

Figure 8. Where should data collection efforts be focused? Map of information gained by collecting one additional sample at that point, broken down by spherical harmonic degree, for the 1160-point high-accuracy spot dataset. Each map is normalised by its maximum value, given in the lower-right corner. Compare with Fig. 3: unsurprisingly, the most-useful places to collect new information correspond to regions of high uncertainty in the recovered topography model.

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

$$\bar{y}_i = y_i + \alpha(\mathbf{x_0}) z_i(\mathbf{x_0}) \left[d_0 - \tilde{\mu}(\mathbf{x_0})\right] \tag{22a}$$

$$\bar{\Sigma}_{ij} = \Sigma_{ij} - \alpha(\mathbf{x_0}) z_i(\mathbf{x_0}) z_j(\mathbf{x_0}) \tag{22b}$$

with

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

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

$$D_{KL} = \frac{1}{2} \left[\mathrm{Tr}(\bar{\mathbf{\Sigma}}^{-\mathbf{1}} \mathbf{\Sigma}) - D + \log \frac{\det \bar{\mathbf{\Sigma}}}{\det \mathbf{\Sigma}} \right] + \frac{1}{2} \alpha^2(\mathbf{x_0}) \mathbf{z}(\mathbf{x_0})^{\mathbf{T}} \bar{\mathbf{\Sigma}}^{-\mathbf{1}} \mathbf{z}(\mathbf{x_0}) \left[d - \tilde{\mu}(\mathbf{x_0})\right]^2 \,, \tag{23}$$

where D represents the dimension of the vector $\bar{\mathbf{c}}$. We therefore see that the amount we learn depends on the value of d: if our additional observation happens to match what we predicted based on the original dataset ($\tilde{\mu}(\mathbf{x_0})$) we learn less than if the observation reveals unexpected 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 an additional observation at $\mathbf{x_0}$ by

$$G(\mathbf{x_0}) = \frac{1}{2}\left[\mathrm{Tr}(\bar{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}) - D + \log\frac{\det\bar{\mathbf{\Sigma}}}{\det\mathbf{\Sigma}}\right] + \frac{1}{2}\alpha^2(\mathbf{x_0})\tilde{k}(\mathbf{x_0},\mathbf{x_0})\mathbf{z}(\mathbf{x_0})^{\mathbf{T}}\bar{\mathbf{\Sigma}}^{-1}\mathbf{z}(\mathbf{x_0})\,. \tag{24}$$

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

Such maps, based on the high-accuracy spot data alone, are shown in Fig. 8. We assume that the one additional data point is measured with an uncertainty of $\sigma_d = 0.1\,\mathrm{km}$. We clearly see that our knowledge of the low-degree spherical harmonic components can be improved through addition of samples at a relatively small number of locations: continental interiors (especially in Russia), the Southern Ocean, and the South and West Pacific. To constrain shorter wavelengths, our analysis indicates that the most useful sampling locations are in regions close to existing data points: this makes sense, as our approach is predicated on collecting only a single sample.

5 Discussion and Conclusions

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

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

The global pattern of residual topography is dominated by broad topographic highs in the North Atlantic, African and central Pacific regions, with the largest amplitudes adjacent to Iceland. These are separated by broad topographic lows, extending from Antarctica, through the Americas, beneath the Eurasian continent and to the south of Australia. Based on just the 1160 high-accuracy measurements, the associated spherical harmonic power spectrum peaks at $l = 2$, with power likely to lie in the range 0.46–0.76 km^2. This decreases by over an order of magnitude to the range 0.02 km^2 at $l = 30$. The overall form of the spectrum is consistent across all data subsets analysed. In light of the sensitivity kernels linking topographic deflections to density anomalies within the mantle at different depths and spherical

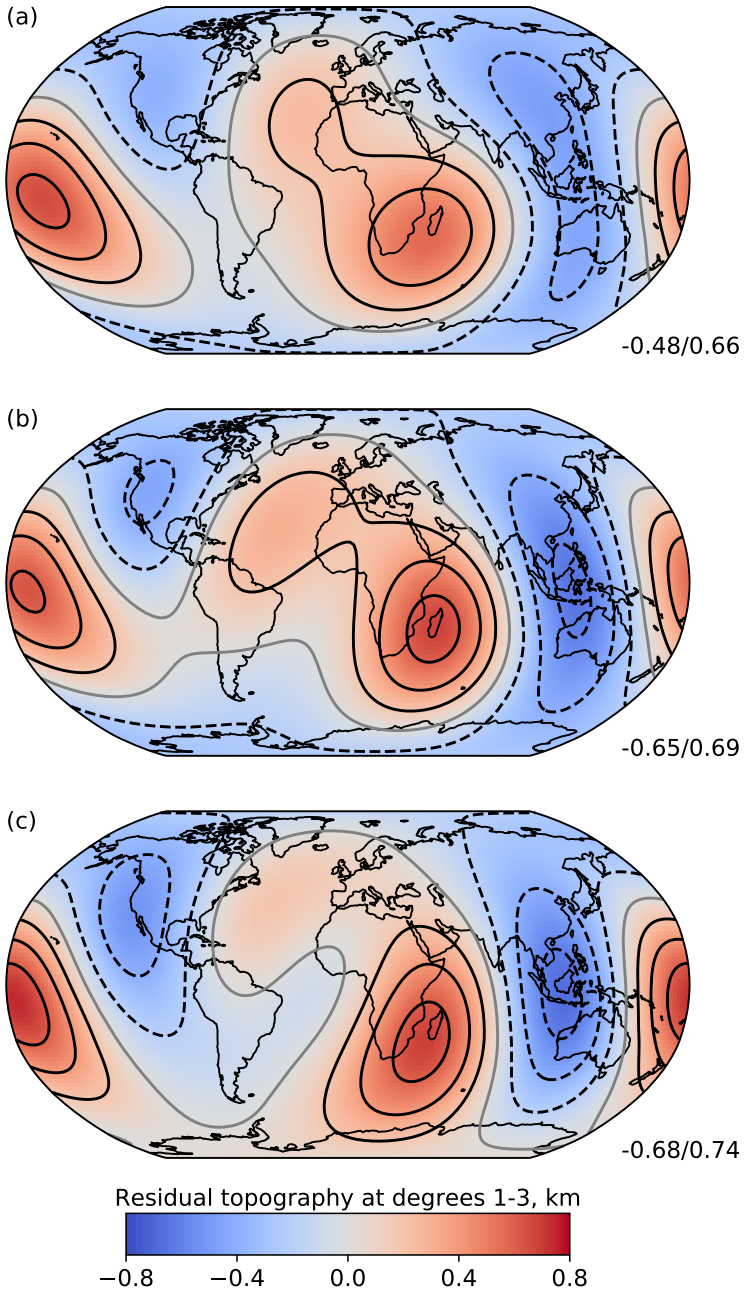

Figure 9. Long-wavelength residual topography. Models (as in Figs. 3–5) filtered to include only spherical harmonic degrees 1–3. We see consistent patterns across all three datasets: (a) high-accuracy spot measurements; (b) all spot measurements; and (c) all spot and shiptrack measurements. Contours 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.

harmonic degree (e.g. Richards & Hager, 1984; Colli et al., 2016), and as discussed in detail by Davies et al. (2019), generating this spectrum requires Earth's topographic response to be driven by multi-scale convective processes. The long-wavelength portion of the spectrum is primarily attributable to deep mantle flow, while shorter-wavelength features are supported by flow associated with structural heterogeneity in the shallow mantle and lithosphere.

Given the link between the long-wavelength components of residual topography and deep mantle flow, the patterns and amplitudes associated with low-order spherical harmonics provide an independent constraint on the long-wavelength components of global seismic tomography images and the dynamical interpretation of these images. Accordingly, in Fig. 9, we show our best-fitting models for each data set, filtered to contain only degrees 1–3. The spatial pattern is fairly consistent across all three datasets, with highs centred off South-East Africa that extend into the North Atlantic, and in the central and western Pacific Ocean. As set out in Table 3, models consistent with the observations may have maximum residual topography heights up to around 1.2 km at these wavelengths, although a figure in the range 680–850 m is most likely. The residual topography highs strongly resemble the shape and extent of the large low velocity provinces that have been consistently imaged in the deep mantle across a number of global seismic tomography studies (e.g. Becker & Boschi, 2002; Ritsema et al., 2011; Bozdağ et al., 2016), and likely represent concentrations of hot (and potentially chemically-distinct) material that drives upwelling flow towards Earth's surface (e.g. McNamara & Zhong, 2005; Schuberth et al., 2009; Bull et al., 2009; Davies et al., 2012; Flament et al., 2013; Davies, Goes, & Sambridge, 2015; Garnero et al., 2016). These long-wavelength residual topography highs are separated by a band of residual topography lows, extending from Antarctica, through the Americas, the Arctic, Eurasia and Australia. These are likely related to downwelling slabs, associated with former subduction zones (e.g. Richards & Engebretson, 1992; Bunge et al., 2002; Flament et al., 2013; Davies, Goes, & Lau, 2015; Rubey et al., 2017).

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

Acknowledgments

We are grateful to Mark Hoggard for making his dataset available, and for many helpful discussions. The specific data files used for this study are as reported in Davies et al. (2019), and may be obtained from `https://github.com/drhodrid/Davies_etal_NGeo_2019_Datasets`. Code to analyse this data and produce the figures presented here may be obtained from `https://github.com/valentineap/DynamicTopographyGP` (doi: 10.5281/zenodo.3895318). APV acknowledges a Discovery Early Career Research Award and other support from the Australian Research Council (ARC: grant numbers DE180100040 and DP200100053). DRD also acknowledges support from the ARC under grants DP170100058 and DP200100053. Calculations were undertaken on the NCI National Facility in Canberra, Australia, which is supported by the Australian Commonwealth Government.

References

Becker, T. W., & Boschi, L. (2002). A comparison of tomographic and geodynamic mantle models. *Geochem. Geophys. Geosys.*, *3*, 2001GC000168. doi: 10.129/2001GC000168

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., ... Pugmire, D. (2016). Global adjoint tomography: first-generation model. *Geophysical Journal*

International, *207*, 1739–1766.

Braun, J. (2010). The many surface expressions of mantle dynamics. *Nature Geoscience*, *3*, 826–833. doi: 10.1038/ngeo1020

Bull, A. L., McNamara, A. K., & Ritsema, J. (2009). Synthetic tomography of plume clusters and thermochemical piles. *Earth Planet. Sci Lett.*, *278*, 152–156. doi: 10.1016/j.epsl.2008.11.018

Bunge, H.-P., Richards, M. A., & Baumgardner, J. R. (2002). Mantle circulation models with sequential data-assimilation: inferring present-day mantle structure from plate motion histories. *Phil. Trans. R. Soc. London, Set. A*, *360*, 2545–2567. doi: 10.1098/rsta.2002.1080.

Byrd, R., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*, 1190-1208.

Colli, L., Ghelichkhan, S., & Bunge, H.-P. (2016). On the ratio of dynamic topography and gravity anomalies in a dynamic Earth. *Geophys. Res. Lett.*, *43*, 2510-2516. doi: 10.1002/2016GL067929

Conrad, C., & Husson, L. (2009). Influence of dynamic topography on sea level and its rate of change. *Lithosphere*, *1*, 110-120. doi: 10.1130/L32.1

Curtis, A. (1999). Optimal design of focused experiments and surveys. *Geophysical Journal International*, *139*, 205–215.

Czarnota, K., Hoggard, M. J., White, N., & Winterbourne, J. (2013). Spatial and temporal patterns of Cenozoic dynamic topography around Australia. *Geochem. Geophys. Geosys.*, *14*, 634–658. doi: 10.1029/2012GC004392

Czarnota, K., Roberts, G. G., White, N. J., & Fishwick, S. (2014). Spatial and temporal patterns of Australian dynamic topography from river profile modelling. *J. Geophys. Res.*, *119*, 1384–1419. doi: 10.1002/2013JB010436

Dahlen, T., & Tromp, J. (1998). *Theoretical global seismology.* Princeton: Princeton University Press.

Davies, D. R., Goes, S., Davies, J. H., Schuberth, B. S. A., Bunge, H., & Ritsema, J. (2012). Reconciling dynamic and seismic models of Earth's lower mantle: the dominant role of thermal heterogeneity. *Earth Planet. Sci. Lett.*, *353*, 253–269. doi: 10.1016/j.epsl.2012.08.016

Davies, D. R., Goes, S., & Lau, H. C. P. (2015). Thermally dominated deep mantle LLSVPs: A review. In A. Khan & F. Deschamps (Eds.), *The Earth's Heterogeneous Mantle* (p. 441-477). Springer International Publishing. doi: 10.1007/978-3-319-15627-9_14

Davies, D. R., Goes, S., & Sambridge, M. (2015). On the relationship between volcanic hotspot locations, the reconstructed eruption sites of large igneous provinces and deep mantle seismic structure. *Earth Planet. Sci. Lett.*, *411*, 121–130. doi: 10.1016/j.epsl.2014.11.052

Davies, D. R., Valentine, A., Kramer, S., Rawlinson, N., Hoggard, M., Eakin, C., & Wilson, C. (2019). Earth's multi-scale topographic response to global mantle flow. *Nature Geoscience*, *12*, 845–850.

Eakin, C., & Lithgow-Bertelloni, C. (2018). An overview of dynamic topography: The influence of mantle circulation on surface topography and landscape. *Mountains, Climate and biodiversity*, *37*.

Flament, N., Gurnis, M., & Muller, R. D. (2013). A review of observations and models of dynamic topography. *Lithosphere*. doi: 10.1130/L245.1

Garnero, E. J., McNamara, A. K., & Shim, S. (2016). Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle. *Nature Geosciences*, *9*, 481–489.

Guerri, M., Cammarano, F., & Tackley, P. J. (2016). Modelling Earth's surface topography: decomposition of the static and dynamic components. *Phys. Earth Planet. Int.*, *261*, 172-186. doi: 10.1016/j.pepi.2016.10.009

Gurnis, M. (1993). Phanerozoic marine inundation of continents driven by dynamic topography above subducting slabs. *Nature*, *364*, 589-593. doi: 10.1038/364589a0

Gurnis, M., Mitrovica, J. X., Ritsema, J., & van Heijst, H. J. (2000). Constraining mantle density structure using geological evidence of surface uplift rates: the case of the

African superplume. *Geochem. Geophys. Geosys.*, *1*, 1999GC000035.

Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P., & Dziewonski, A. M. (1985). Lower mantle heterogeneity, dynamic topography and the geoid. *Nature*, *313*, 541–545. doi: 10.1038/313541a0

Hager, B. H., & Richards, A. M. (1989). Long–wavelength variations in Earth's geoid: Physical models and dynamical implications. *Phil. Trans. Roy. Soc. London, Ser. A.*, *328*, 309–327.

Hoggard, M., White, N., & Al-Attar, D. (2016). Global dynamic topography observations reveal limited influence of large-scale mantle flow. *Nature Geoscience*, *9*, 456–463.

Hoggard, M., Winterbourne, J., Czarnota, K., & White, N. (2017). Oceanic residual depth measurements, the plate cooling model, and global dynamic topography. *Journal of Geophysical Research*, *122*, 2328–2372.

Kaban, M. K., Schwintzer, P., Artemieva, I. M., & Mooney, W. D. (2003). Density of the continental roots: Compositional and thermal contributions. *Earth Planet. Sci. Lett.*, *209*, 53-69. doi: 10.1016/S0012-821X(03)00072-4

Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*, 79–86.

Laske, G., & Masters, G. A. (1997). A global digital map of sediment thickness. *EOS Trans. Am. Geophys. Union*, *78*, F483.

Lithgow-Bertelloni, C., & Silver, P. G. (1998). Dynamic topography, plate driving forces and the African superswell. *Nature*, *395*, 269–272. doi: 10.1038/26212

Mackay, D. (1992). Bayesian interpolation. *Neural Computation*, *4*, 415–447.

McNamara, A. K., & Zhong, S. (2005). Thermo–chemical structures beneath Africa and the Pacific Ocean. *Nature*, *437*, 1136–1139. doi: 10.1038/nature04066

Mitrovica, J. X., Beaumont, C., & Jarvis, G. T. (1989). Tilting of continental interiors by the dynamical effects of subduction. *Tectonics*, *8*, 1079-1094. doi: 10.1029/TC008i005p01079

Murphy, K. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Panasyuk, S. V., & Hager, B. H. (2000). Models of isostatic and dynamic topography, geoid anomalies, and their uncertainties. *J. Geophys. Res.*, *105*, 28199-28209. doi: 10.1029/2000JB900249

Parsons, B., & Daly, S. (1983). The relationship between surface topography, gravity anomalies, and temperature structure of convection. *J. Geophys. Res.*, *88*, 1129–1144. doi: 10.1029/JB088iB02p01129

Pekeris, C. L. (1935). Thermal convection in the interior of the Earth. *Geophys. J.*, *3*, 343-367.

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. Cambridge, USA: MIT Press.

Richards, M. A., & Engebretson, D. C. (1992). Large-scale mantle convection and the history of subduction. *Nature*, *355*, 437–440. doi: 10.1029/2007JB005155

Richards, M. A., & Hager, B. H. (1984). Geoid anomalies in a dynamic Earth. *J. Geophys. Res.*, *89*, 5987-6002.

Ritsema, J., Deuss, A., van Heijst, H., & Woodhouse, J. (2011). S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophysical Journal International*, *184*, 1223–1236.

Rubey, M., Brune, S., Heine, C., Davies, D. R., Williams, S. E., & Muller, R. D. (2017). Global patterns in Earth's dynamic topography since the Jurassic: the role of subducted slabs. *Solid Earth*, *8*, 899-919. doi: 10.5194/se-8-899-2017

Schuberth, B., Bunge, H.-P., Steinle-Neumann, G., Moder, C., & Oeser, J. (2009). Thermal versus elastic heterogeneity in high-resolution mantle circulation models with pyrolite composition: high plume excess temperatures in the lowermost mantle. *Geochemistry, Geophysics, Geosystems*, *10*.

Shephard, G. E., Muller, R. D., Liu, L., & Gurnis, M. (2010). Miocene drainage reversal of

the Amazon River driven by plat e–mantle interaction. *Nature Geoscience*, *3*, 870–875. doi: 10.1038/NGEO1017

Smith, W. H. F., & Sandwell, D. T. (1997). Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, *277*, 1956-1962.

Steinberger, B. (2016). Topography caused by mantle density variations: observation-based estimates and models derived from tomography and lithosphere thickness. *Geophys. J. Int.*, *205*, 604–621. doi: 10.1093/gji/ggw040

Steinberger, B., Conrad, C. P., Osei Tutu, A., & Hoggard, M. J. (2019). On the amplitude of dynamic topography at spherical harmonic degree two. *Tectonophys.*, In Press. doi: 10.1016/j.tecto.2017.11.032

Trampert, J., & Snieder, R. (1996). Model estimations biased by truncated expansions: Possible artifacts in seismic tomography. *Science*, *271*, 1257–1260.

Valentine, A., & Sambridge, M. (2018). Optimal regularization for a class of linear inverse problem. *Geophysical Journal International*, *215*, 1003–1021.

Valentine, A., & Sambridge, M. (2020a). Gaussian process models—I. A framework for probabilistic continuous inverse theory. *Geophysical Journal International*, *220*, 1632–1647.

Valentine, A., & Sambridge, M. (2020b). Gaussian process models—II. Lessons for discrete inversion. *Geophysical Journal International*, *220*, 1648–1656.

Watkins, C. E., & Conrad, C. P. (2018). Constraints on dynamic topography from asymmetric subsidence of the mid-ocean ridges. *Earth Planet. Sci. Lett.*, *484*, 264-275. doi: 10.1016/j.epsl.2017.12.028

Winterbourne, J. R., Crosby, A. G., & White, N. J. (2009). Depth, age and dynamic topography of oceanic lithosphere beneath heavily sedimented atlantic margins. *Earth Planet. Sci. Lett.*, *287*, 137-151.

Winterbourne, J. R., White, N. J., & Crosby, A. G. (2014). Accurate measurements of residual topography from the oceanic realm. *Tectonics*, *33*, 1-34.

Yang, T., & Gurnis, M. (2016). Dynamic topography, gravity and the role of lateral viscosity variations from inversion of global mantle flow. *Geophysical Journal International*, *207*, 1186–1202.

Yang, T., Moresi, L., Müller, R., & Gurnis, M. (2017). Oceanic residual topography agrees with mantle flow predictions at long wavelengths. *Geophysical Research Letters*, *44*.