HAL
open science

Moho depth of the British Isles: a probabilistic perspective

Andrea Licciardi, R W England, N Piana Agostinetti, Kerry Gallagher

To cite this version:

Andrea Licciardi, R W England, N Piana Agostinetti, Kerry Gallagher. Moho depth of the British Isles: a probabilistic perspective. Geophysical Journal International, 2020, 221 (2), pp.1384-1401. 10.1093/gji/ggaa021 . insu-02506722

HAL Id: insu-02506722

https://insu.hal.science/insu-02506722v1

Submitted on 12 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Geophysical Journal International

Geophys. J. Int. (2020) **00,** 1
GJI General Geophysical Methods

doi: 10.1093/gji/ggaa021

Moho depth of the British Isles: a probabilistic perspective

A. Licciardi [1,*] R.W. England,[2] N. Piana Agostinetti [3] and K. Gallagher[1]

[1] *Géosciences, Université de Rennes* 1, *Rennes, France. E-mail: alicciardi@cp.dias.ie*
[2] *School of Geography, Geology and the Environment, University of Leicester, Leicester, UK*
[3] *Department of Geodynamics and Sedimentology, University of Vienna, Vienna, Austria*

Accepted 2020 January 6. Received 2020 January 6; in original form 2019 March 18

SUMMARY

We present a new Moho depth model of the British Isles and surrounding areas from the most up-to-date compilation of Moho depth estimates obtained from refraction, reflection and receiver function data. We use a probabilistic, trans-dimensional and hierarchical approach for the surface reconstruction of Moho topography. This fully data-driven approach allows for adaptive parametrization, assessment of relative importance between different data-types and uncertainties quantification on the reconstructed surface. Our results confirm the first order features of the Moho topography obtained in previous work such as deeper Moho (29–36 km) in continental areas (e.g. Ireland and Great Britain) and shallower Moho (12–22 km) offshore (e.g. in the Atlantic Ocean, west of Ireland). Resolution is improved by including recent available data, especially around the Porcupine Basin, onshore Ireland and Great Britain. NE trending features in Moho topography are highlighted above the Rockall High (about 28 km) and the Rockall Trough (with a NE directed deepening from 12 to about 20 km). A perpendicular SE oriented feature (Moho depth 26–28 km) is located between the Orkney and the Shetland, extending further SW in the North Sea. Onshore, our results highlight the crustal thinning towards the N in Ireland and an E–W oriented transition between deep (34 km) and shallow (about 28 km) Moho in Scotland. Our probabilistic results are compared with previous models showing overall differences around ±2 km, within the posterior uncertainties calculated with our approach. Bigger differences are located where different data are used between models or in less constrained areas where posterior uncertainties are high.

Key words: Spatial analysis; Statistical methods; Crustal imaging; Crustal structure.

1 INTRODUCTION

The Mohorovicic (Moho) discontinuity is a major feature of the Earth and defines the boundary between the crust and the upper mantle. As such, the depth of the Moho is a key parameter for many aspects of Earth Sciences, such as seismic tomography, geodynamic reconstructions, crust–mantle compositional models, isostatic compensation and gravity studies, seismic hazard mitigation and exploration for Earth's resources. Given the distinctive seismic signature of the Moho (*P*-wave velocities increase from a range of 6.7–7.2 km s^{-1} to a range of 7.6–8.6 km s^{-1}), seismic data provide fundamental information to build models of Moho topography at various scales. For more than 50 yr, controlled source seismology has represented the primary tool to study the depth of the Moho (see Prodehl *et al.* 2013, and references therein). Over the last two decades, passive seismic methods and in particular the receiver function (RF) approach (Langston 1977) have relied on the ever increasing number of relatively low cost seismic stations to produce abundant estimates of Moho depth. Results from RF studies have been used alone (e.g. Piana Agostinetti & Amato 2009; Li *et al.* 2014) or combined with refraction and reflection data to construct Moho models on both global (Prodhel & Mooney 2012; Laske *et al.* 2013) and regional scales (e.g. Kennett *et al.* 2011).

Regional Moho depth models are usually constructed by the spatial interpolation of irregularly distributed point estimates. This 2-D spatial interpolation process, or surface reconstruction, is in fact an inverse problem in which all the available but sparse data are used to infer a continuous surface describing the Moho topography over the specified region, including areas of no data. Interpolation of spatial data is routinely performed in many Earth Science applications with different techniques. The choice of interpolation method varies depending on user experience and on the type of data to be interpolated. Some widely used methods for spatial interpolation include kriging (Stein 1999), near neighbour, triangulation, spline-in-tension (Smith & Wessel 1990).

The majority of these methods suffer from well known limitations. First, the level of smoothness of the reconstructed surface

* Now at Université Côte d'Azur, Géoazur, Sophia Antipolis, France.

© The Author(s) 2020. Published by Oxford University Press on behalf of The Royal Astronomical Society.

is typically specified by the user over the whole 2-D domain of interest. As the data are irregularly distributed and have spatially varying uncertainties, a constant smoothness approach cannot adapt to changes in spatial resolution. In addition, the parameters controlling the smoothness need to be tuned. This tuning is often performed with a trial and error approach which requires subjective choices and is user-time consuming. For example, the results given by the gridding with tension approach (Smith & Wessel 1990) rely on the choice of the tension parameter of the splines used in the interpolation, which is completely user-subjective. Similarly, the search radius parameter in near neighbour interpolation needs to be set based on experience and strongly influences the results.

A second limitation arises when multiple data-types are involved. In this case, weights need to be assigned to each type of data. If done arbitrarily, this may possibly introduce a bias in the reconstructed surface. Finally, estimation of the uncertainty associated with a reconstructed surface is difficult to obtain with the majority of these methods. To overcome these limitations, a Bayesian approach for the surface reconstruction problem has been proposed (Bodin *et al.* 2012a). In this approach, based on a reversible jump Markov chain Monte Carlo (rjMcMC) algorithm, both the number of parameters in the reconstructed surface and the magnitude of data noise are treated as unknowns in the interpolation. The main result is a self-adaptive parametrization in which both the surface complexity (i.e. smoothness) and the weights of different data-types are directly inferred from the data themselves. This algorithm has been successfully applied to the surface reconstruction of Moho topography (Bodin *et al.* 2012a), coastal sea level rise (Choblet *et al.* 2014) and geodetic uplift rates (Husson *et al.* 2018).

In this work, we develop a similar algorithm and apply it to the most complete database of seismic Moho depth data of the British Isles. We show that the Bayesian approach provides a framework for dealing with multiple data-types with different resolution, spatial distribution and errors. The probabilistic solution and quantification of uncertainties provided by our Bayesian approach gives new insights into the Moho topography across the British Isles and surroundings and opens new perspectives for future studies in the area.

1.1 Moho depth of the British Isles: previous studies

The crustal structure of the British Isles (Fig. 1) has been extensively explored over the last few decades through a combination of active and passive seismic studies driven by both scientific and economic interests (e.g. hydrocarbon exploration). This has resulted in one of the most seismically surveyed regions in the world with a high density of good quality data, both onshore and offshore.

The first maps of Moho depth were based on the BIRPS deep seismic reflection data (Klemperer & Hobbs 1991) and mainly cover the offshore domain of Great Britain (Meissner *et al.* 1986; Chadwick & Pharaoh 1998). Building on an earlier compilation (Clegg & England 2003), a comprehensive database of seismic wide-angle reflection and refraction data was assembled by Kelly *et al.* (2007) to construct velocity and Moho depth models of the British Isles. This compilation is also included in the Moho model for Europe, Greenland and the North Atlantic of Artemieva & Thybo (2013). More recently, passive seismic studies based on the RF method, have provided new sets of Moho depth measurements onshore Ireland and Great Britain (see Davis *et al.* 2012; Licciardi *et al.* 2014, and references therein). Some of these results were used to build a recent Moho model of the NE Atlantic Ocean (Funck *et al.* 2016). The most recent Moho depth model of the British Isles is that of Baykiev *et al.* (2018) using a compilation of results from refraction and RF data and also including the estimates from offshore deep reflection profiles provided by Davis *et al.* (2012).

The available Moho depth models of the British Isles have been obtained with classical interpolation algorithms. In particular, kriging was used to obtain the Moho models of Kelly *et al.* (2007) and Funck *et al.* (2016), while a combination of 'nearest neighbour' and kriging was used by Artemieva & Thybo (2013). Finally, the gridding with tension approach (Smith & Wessel 1990) was used by Baykiev *et al.* (2018). Despite the abundance of Moho depth models of the study area, no efforts have been focused on quantify the uncertainties on them.

In this work, we aim at filling this gap by using a probabilistic approach to the surface reconstruction problem. In order to deal with the most up-to-date compilation of Moho depth estimates, we update the most complete available database with recently published results (Watremez *et al.* 2016; Prada *et al.* 2017; Chen *et al.* 2018), which have not been used in the construction of previous models.

2 DATA

At the core of this study, there is a database of 4581 Moho depth estimates comprised between -20° and 7° of longitude and between 48° and 63° of latitude (Fig. 2) and assembled from published results using active and passive seismic data. Based on the type of data and the different techniques used to estimate the Moho depth, we subdivide the database into four categories for the surface reconstruction problem: refraction/wide-angle, reflection, broad-band RFs and short-period RFs. Alternative and/or additional partitions of the database are possible. For example, the type of processing applied to RF data for obtaining Moho depth values (e.g. from $H - \kappa$ stacking or forward/inverse modelling of RF data) could also be considered for defining additional categories in our inversion. Another choice could be grouping the active seismic data based on the year of acquisition, to take in consideration the development of the technology across 40 yr. Here, we have opted for a simple data partitioning to facilitate the interpretation of the results.

The data have been preprocessed as little as possible to avoid user-induced bias in our data-driven Bayesian approach. In particular, no spatial resampling or smoothing has been performed on the data. Therefore, the final density of data points is highly variable over the region of interest, with different sampling intervals between different categories.

Refraction data (circles in Fig. 2) constitute the primary source of information with 4146 Moho depth estimates both offshore and onshore. Wide-angle/refraction data provide high quality estimates of Moho depth as they directly contain information on velocity at depth and require few *a priori* constraints. We use the compilation of refraction data presented in Kelly *et al.* (2007), where the reader may find the complete list of experiments with references. In addition, we complement the refraction data with recently published results from experiments carried out in the Porcupine Basin (Watremez *et al.* 2016; Prada *et al.* 2017; Chen *et al.* 2018).

Reflection data (triangles in Fig. 2) contribute with 294 point estimates. We use the data compiled by Davis *et al.* (2012) from the work of Chadwick & Pharaoh (1998) in the British Isles. The majority of the points within this group are located offshore, although some sparse information is available onshore Great Britain. Moho depth estimates from reflection data are obtained from two-way-traveltime (twtt) measurements, which are converted to depth using

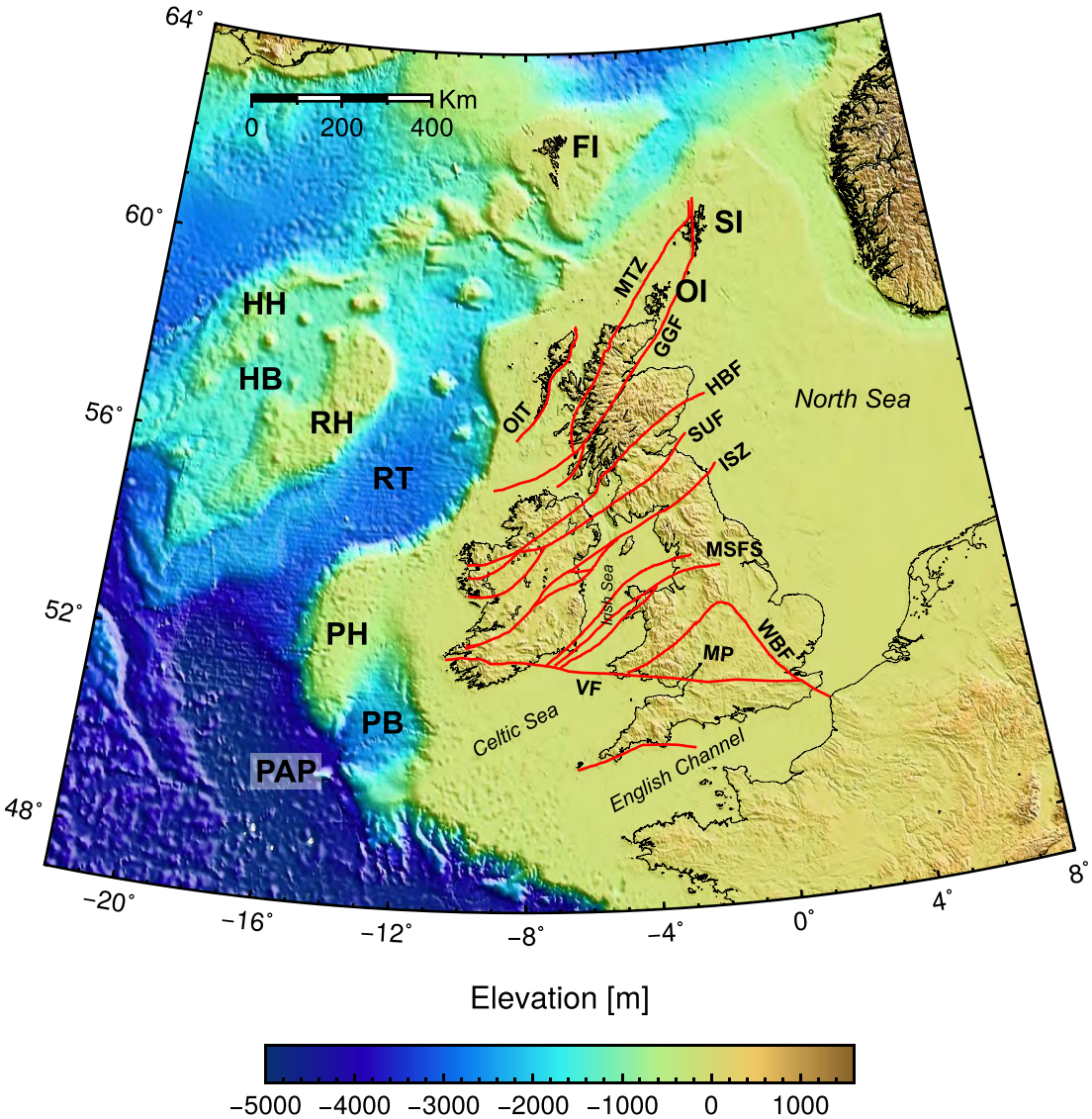

Figure 1. Physiographical map of the study area according the the ETOPO 1 Global Relief Model (Amante & Eakins 2009) and main tectonic boundaries (red lines) modified from Tomlinson *et al.* (2006). FI, Faroe Islands; GGF, Great Glen Fault; HB, Hatton Basin; HBF, Highland Boundary Fault; HH, Hatton High; ISZ, Iapetus Suture Zone; MP, Midland Platform; MSFS, Menai Straits Fault System; MTZ, Moine Thrust Zone; OI, Orkney Islands; OIT, Outer Islands Fault; PAP, Porcupine Abyssal Plain; PB, Porcupine Basin; PH, Porcupine High; RH, Rockall High; RT, Rockall Through; SI, Shetland Islands; SUF, Southern Uplands Fault; VF, Variscan Front; WBF, Welsh Borderland Fault zone.

an assumed crustal velocity model, as described by Chadwick & Pharaoh (1998).

Finally, we compiled Moho depth estimates from RF data obtained over the last decade onshore Ireland and Great Britain. This technique is based on the analysis of teleseismic converted waves below a single station and is mainly sensitive to the impedance contrasts located at the Moho (Zhu & Kanamori 2000). We subdivide these estimates based on the type of sensor used (broadband or short period, Fig. 2) to obtain the Moho depth values. The reason for this choice is that, depending on the depth and nature of the crust–mantle boundary (shallow/deep, gradational/sharp), the frequency content in RF data determines the resolution at which the Moho signal can be detected, with broad-band signal carrying more information and higher resolution to identify gradient at the Moho transition (Bodin *et al.* 2012a). A total of 116 Moho depth estimates from broad-band RF and 25 from short period RFs have been included. For broad-band RF, the main sources of data are those from Landes *et al.* (2006), Shaw Champion *et al.* (2006), Tomlinson *et al.* (2006), Di Leo *et al.* (2009), Davis *et al.* (2012) and Licciardi *et al.* (2014). Short period sensors were used only in the work of Tomlinson *et al.* (2006). In some cases, for a single station multiple estimates of Moho depth are available. Each individual value is considered as an independent point in our surface reconstruction.

In our approach, the magnitudes of the data errors are treated as unknown in the inversion, but the relative importance between data points of the same category is fixed (see the next section). For this reason, good relative knowledge of the uncertainties is required for each point. Whenever possible, information about uncertainties on each Moho depth value has been included. For refraction data, we use the empirical estimates of uncertainties given in Kelly *et al.* (2007) and the Monte Carlo error estimates for the data in the Porcupine Basin. For RF data, we used published uncertainties with a caveat: most published results rely on a common RF technique

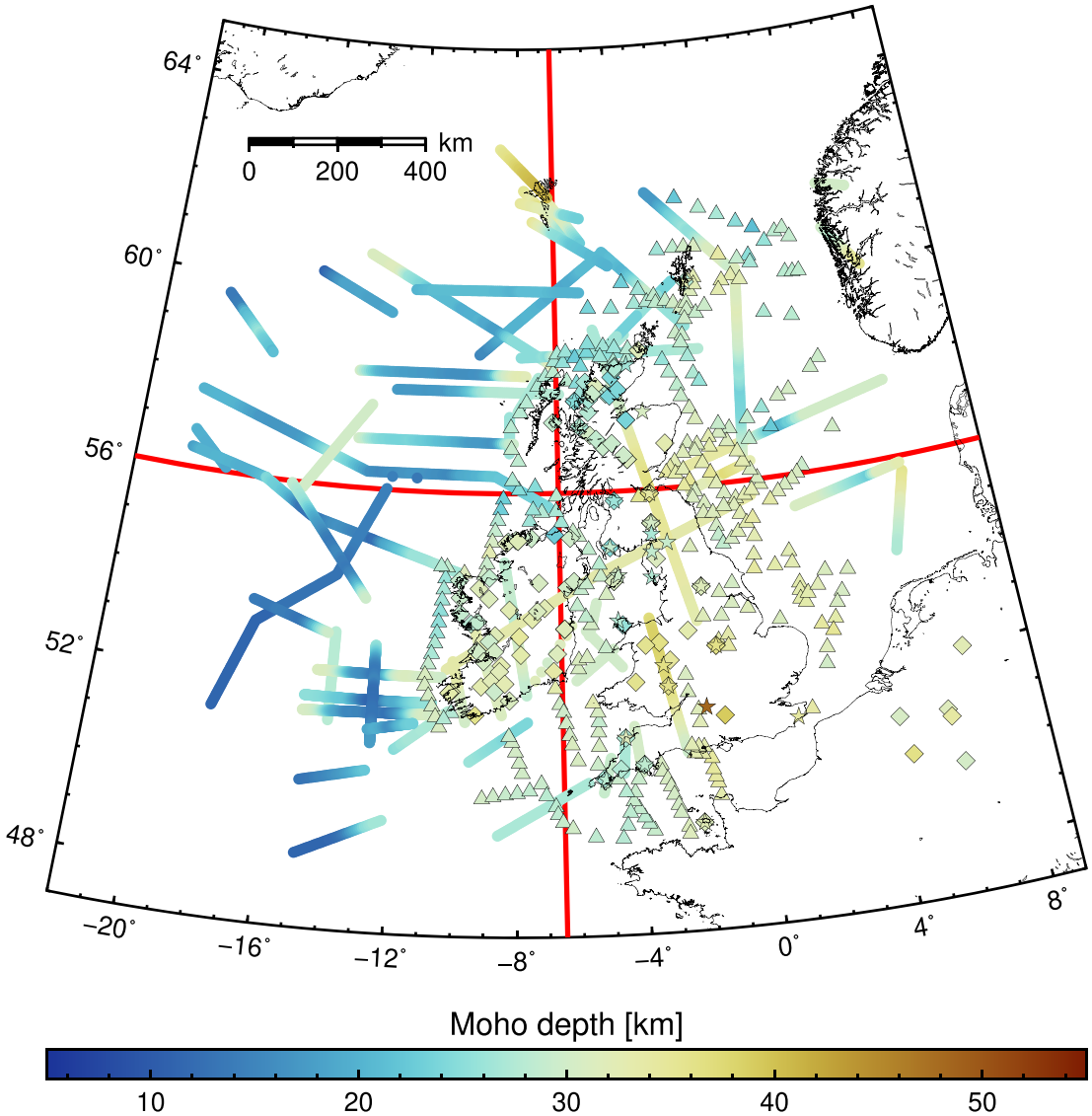

Figure 2. Location of the Moho depth estimates used in this study. Four different data-types are considered: refraction/wide angle (coloured lines), reflection (triangles), broad-band (squares) and short period (stars) receiver functions. Red lines mark the orientation of the two profiles shown in Fig. 6.

known as $H - \kappa$ stacking (Zhu & Kanamori 2000), which may underestimate uncertainties on the retrieved Moho depth, depending on the number of data and the particular approach used (Ogden *et al.* 2019). For some of the estimates used in this study, the authors reported uncertainties as low as 0.3 km, which is below the resolution of the RF data. For this reason, we set a minimum value (0.8 km) for the standard deviation associated with these estimates. This value is somehow arbitrary, but it better represents the smallest achievable errors when using the $H - \kappa$ technique (e.g. Licciardi *et al.* 2014).

Finally, uncertainties on reflection data were not available. For this category, we set the standard deviation for each point estimate equal to 1 km. In principle, uncertainties on Moho depth values estimated from reflection data could be differentiated according to the age, the source frequency and/or fold of the original seismic lines (e.g. Bodin *et al.* 2012a). In practice, the main factor controlling the uncertainties on Moho depth estimates from reflection lines is the type of reflectivity at the Moho that can affect the picking. The greatest uncertainty on Moho depth are found in areas where there is no lower crustal reflectivity. In addition, the amount of reflectivity can vary along individual lines. Finally, the twtt picks are converted to depth with a velocity-depth function, and this step itself introduces errors which are difficult to track. The Moho depth values used in this study are those obtained by (Chadwick & Pharaoh 1998) by depth-converting the twtt values using a different layered velocity model for each seismic line (table 1 of Chadwick & Pharaoh 1998). In this way, the errors are at least internally consistent between different lines. Therefore, although assigning a constant error to the whole reflection data set might not be the optimal solution, it is the only viable option as a differentiation of the errors between reflection lines is not possible with the data in our possession. A conservative approach is used in this regard as a subjective differentiation of data error may introduce a bias which would be difficult to quantify.

3 METHOD

In this work, we use a Bayesian probabilistic approach for the surface reconstruction algorithm. In particular, we use a hierarchical approach (Malinverno & Briggs 2004) combined with a rjMcMC algorithm (Green 2003) in order to sample the unknown posterior probability distribution (PPD) of the model parameters (the unknown surface) given prior information on the model parameters and a measure of data fit (the likelihood function). Here, we give a description of the features characterizing our specific implementation, while more details are given in the Supporting Information. For general description of the probabilistic approach to inverse problem refer to Mosegaard & Tarantola (2002). For reviews of McMC and rjMcMC methods in geosciences refer to Gallagher *et al.* (2009) and Sambridge *et al.* (2013).

Our approach is based on that described by Bodin *et al.* (2012a), in which both the number of parameters (k) used to describe the reconstructed surface and the magnitude of data noise are considered unknown in the inversion. At each iteration of the rjMcMC algorithm, the current model is perturbed probabilistically according to a specific set of rules (see the Supporting Information). The proposed model is accepted or rejected based on the Metropolis–Hasting acceptance criterion (Metropolis *et al.* 1953; Hastings 1970) which we implemented following Mosegaard & Tarantola (1995) and Piana Agostinetti & Malinverno (2010), as described in the Supporting Information. The final ensemble of collected models will have a distribution that approximate the PPD.

In the rjMcMC algorithm, the dimensions of the problem can vary and this is classically referred to as a trans-dimensional approach (see Sambridge *et al.* 2013, for a review). One of the recognized advantages of a trans-dimensional approach is its implicit parsimony (Malinverno 2002; Mackay 2003). Given two models with similar likelihood the simplest one (with lower k) is preferred. This represents an advantage in our application as: (i) overfitting the data is avoided and (ii) the level of complexity of the reconstructed surface will be determined directly from the data. In other words, there is no need to specify a smoothing term for the interpolant surface as the degree of smoothing is self-adapted for different portions of the map based on the information contained in the data, their uncertainties and the spatial distribution of data points.

3.1 Model specification

Following Bodin *et al.* (2012a), the surface to be inferred is parametrized with a k number of Voroni cells (Sambridge *et al.* 1995). Although alternative parametrizations have been recently suggested for this kind of problem (Hawkins *et al.* 2019), the Voronoi scheme represents a common choice in trans-dmensional inference due to its flexibility and the relative low number of free parameters (Bodin *et al.* 2012a; Piana Agostinetti *et al.* 2015).

Each Voronoi cell is described by the location coordinates of its centre $\mathbf{c} = [x, y]$ and by the value of the parameter to be interpolated (Moho depth), z which is considered constant within the cell. By construction, a surface made up of k Voronoi cells is uniquely determined by specifying the two vectors: $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_k]$ and $\mathbf{z} = [z_1, z_2, ..., z_k]$. In addition, the magnitude of data error associated with each of the 4 data-type is considered as an unknown and is specified by a set of so called hyperparameters $\mathbf{h} = [h_1, h_2, h_3, h_4]$. Therefore, the full set of model parameters can be expressed as $\mathbf{m} = [k, \mathbf{C}, \mathbf{z}, \mathbf{h}]$.

3.2 Likelihood and hierarchical approach

The likelihood function plays a fundamental role in the rjMcMC algorithm by driving the sampling of the PPD. In this work, the likelihood of a given model is the combination of the likelihoods of four independent data-types. This leads to the following expression for the likelihood function:

$$p(\mathbf{d} \mid \mathbf{m}) = \prod_{i=1}^{4} \frac{1}{\left[(2\pi)^{N_i} |\mathbf{C}_{\mathbf{e},i}|\right]^{1/2}} \exp\left[-\frac{1}{2}\mathbf{e}_i^T \mathbf{C}_{\mathbf{e},i}^{-1} \mathbf{e}_i\right], \quad (1)$$

where $\mathbf{C}_{\mathbf{e},i}$ is the covariance matrix of the data error and $\mathbf{e}_i = \mathbf{d}_i - \mathbf{g}_i(\mathbf{m})$ is the error vector for the ith data-type. The forward operator $\mathbf{g}_i(\mathbf{m})$ simply expresses the Moho depth values of the current model for the ith data-type.

Our algorithm makes use of the so-called Hierarchical approach (Malinverno & Briggs 2004), in which the magnitude of data errors is treated as an unknown. This is implemented through the use of the hyperparameters h_i according to:

$$\mathbf{C}_{\mathbf{e},i} = \theta_i \times \mathbf{C}^*_{\mathbf{e},i}, \quad (2)$$

where $\mathbf{C}^*_{\mathbf{e},i}$ is the covariance matrix of the observed errors and $\theta_i = 10^{h_i}$. With this type of implementation, the relative error between data points of the same type of data is fixed while θ_i is a scaling parameter that controls the absolute magnitude of $\mathbf{C}_{\mathbf{e},i}$. In this work, all data points are assumed to be uncorrelated (i.e. $\mathbf{C}^*_{\mathbf{e},i}$ is diagonal). In addition, as explained in Section 2, we use the same value for the observed standard deviation for all reflection data points (1 km). In this case, $\mathbf{C}^*_{\mathbf{e},i}$ further simplifies to a scalar, whose value is directly inferred in the inversion through the corresponding hyperparameter. We note that in our formulation, the hyperparameters will also include the contribution of errors due to the approximation implied by the Voronoi model (i.e. an ensemble of 2-D step functions with a constant Moho depth value). In practice, in order for the model to fit the data where this approximation is less valid (e.g. for dipping Moho), the hyperparameters will tend to increase.

Eqs (1) and (2) show how the hyperparameters contribute to the likelihood. Therefore, they play a role in driving the sampling of the PPD. The effect of the hierarchical approach is twofold. First, it provides a posterior estimate of the data errors, which may be poorly known *a priori*. Secondly, the hyperparameters act as weighting factors in our joint inversion of multiple data-types. They are automatically adjusted by the algorithm, based solely on the information contained in the data, thus reducing the influence of subjective choices imposed by the user (Bodin *et al.* 2012b).

3.3 Prior distributions

In the Bayesian approach, prior distributions (or simply priors) plays a key role in posterior sampling as they limit the possible values of the parameters in the Markov chain. In this work, we use uniform prior distributions for all the model parameters as they are easy to deal with and do not impose strong constraints on the solution. For the number of Voronoi cells, $k \sim U(1,350)$. For longitude, latitude and Moho depth we set respectively $x \sim U(-22, 9)^\circ$, $y \sim U(47, 65)^\circ$ and $z \sim U(5, 55)$ km. For all hyperparameters $h \sim U(0, 1)$ resulting in a scale factor θ of the observed errors between 1 and 10.

Working with prior distributions offers a flexible framework to define the level of prior knowledge on each model parameter. These distributions can be easily modified according to previous studies or any kind of scientific knowledge. Although our Bayesian approach aims at reducing the user-induced bias, it still requires a number

of assumptions to be made. These assumptions are involved in the choice of parametrization, in the type of data noise correlation, and in the data selection process. We call these 'implicit priors' and Appendix A show that all these assumptions impose strong constraints on the interpolation.

4 RESULTS

The rjMcMC algorithm is applied to the data set of Moho depths described in Section 2. We use 300 parallel chains, each of them running for 5,000,000 million iterations, discarding the first half of them and collecting models every 500 iterations to reduce potential correlations between successively sampled parameter values. Overall, the posterior inference is made on an ensemble of 1.5 million models.

The probabilistic Moho depth surface resulting from our algorithm is conditional on the number of Voronoi cells and on the values of $\mathbf{C_e}$ that depend on the hyperparameters (Fig. 3). Therefore, in order to better understand results, an assessment of the relative importance of each data type in the inversion is needed. This is done by considering the posterior marginal distribution of θ multiplied by $\langle\sqrt{diag(\mathbf{C_e^*})}\rangle$ (i.e. the average value of the observed standard deviations of the data noise) for each category (Fig. 3b). Moho depth estimates from reflection data have the lowest posterior standard deviation of data noise (1.5–1.8 km) and therefore the highest importance in the joint inversion. The results for refraction data partially overlap with reflection data with values less than 2 km, but show a more skewed distribution towards higher values. This reflects the high number of measurements belonging to this group, with observed standard deviation spanning a higher range of values. Finally, Moho depth estimates from RF data have higher posterior errors, with those from broad-band RF (≈2.8 km) being more important than those from short period RF, which shows a broad distribution centred at about 3.5–4.0 km. The inferred relative importance of each data type is supported by experience: seismic data from active sources carry information with higher resolution than those from passive seismic data and therefore, are given more importance in the inversion. In addition, the absolute values of the posterior standard deviation of the data noise fall into the usually assumed range for each category (Zelt & Smith 1992; Zhu & Kanamori 2000).

It is important to bear in mind that the results shown in Fig. 3(b) are tightly constrained by our choices regarding, density of data points, error correlation and fixed relative importance of each point within a given data type. For example, it is not surprising that the two most densely spaced, internally consistent data types (i.e. refraction and reflection data) yield the lowest posterior standard deviations, given that the used algorithm prefers parsimonious solutions. Moreover, the assumption of uncorrelated errors is probably adequate for RF data as data points can safely be considered as independent, as they are individually obtained by a single station analysis. On the other hand, refraction and reflection data sets consist of dense points along oriented lines that come from the interpretation of seismic images or velocity models. This interpretation process is likely to introduce a certain degree of correlation between nearby points along the same refraction/reflection line. Therefore, our assumption may lead to an underestimation of the posterior errors for this type of data, whose importance may be consequently overestimated in the joint inversion. With two additional tests (Figs A1 and A2, Appendix A1), we show that decreasing the data density and/or the strength of data noise correlation results in larger posterior estimates of the data noise. These tests suggest that, given our assumptions, the results of Fig. 3(b) should be taken as a rough estimate of the relative importance of each data type rather than an exhaustive one. In any case, this represents a clear advantage over classical interpolation techniques which do not provide such information.

The full solution of our interpolation scheme is represented by a multidimensional PPD which is difficult to visualize in a single figure. However, some representative statistics of the PPD can be extracted from the ensemble of collected models for simpler analysis and interpretation. The mean Moho depth map is shown in Fig. 4. Although each individual model sampled in the Markov chain is a single Voronoi tessellation, and so the depth map is discontinuous, the process of averaging over the whole ensemble of collected models produces a smooth mean surface (Fig. 4a). A visual comparison between the mean model and the data (Fig. 2) supports the robustness of the first order features of the mean model. Deeper Moho (between 30 and 40 km depth) is found in continental areas (e.g. Ireland and Great Britain), and shallower Moho (12–22 km) offshore (e.g in the Atlantic Ocean, west of Ireland). Maximum values of Moho depth are found in Wales (36–37 km), SE England and near the Faroe Islands (38–40 km). On the other hand, minimum values of Moho depth (11–12 km) are located in the Porcupine Abyssal Plain (PAP).

A spot of anomalously thick crust of about 40 km is present in SE Great Britain. This is controlled by a single data point of short-period RFs. However, as we shall see in the next section, the probabilistic solution obtained with our algorithm provides robust guidelines for the interpretation of this Moho depth feature.

Nevertheless, in the Supporting Information we show and discuss the results of an outlier-robust inversion in which the the influence of outliers is reduced by replacing the likelihood of eq 1 with an equivalent expression that uses an L1-norm measure of data misfit (e.g. Poggiali *et al.* 2019).

4.1 Moho depth uncertainties and resolution

The Bayesian approach at the core of our algorithm readily allows for a quantification of uncertainties on the retrieved Moho map. Posterior uncertainties depend on the combined effect of observed data errors, values of the hyperparameters, spatial distribution of measurements, data consistency and model complexity. In order to understand how uncertainties are obtained and how they should be interpreted, we plot the map of posterior standard deviation of the sampled depth values at each pixel, σ, together with the posterior density of Voronoi nuclei in Fig. 5. The σ map (Fig. 5a) indicates how well resolved is the mean of the posterior distribution shown in Fig. 4. The nuclei density map (Fig. 5b) is a proxy for model complexity: high density of Voronoi nuclei is associated with small Voronoi cells and is found where high complexity in the reconstructed surface is required by the data. On the other hand, low density values indicate the presence of larger cells. The non-homogeneous distribution of nuclei density values expresses the adaptive nature of our Voronoi parametrization.

The highest values of σ (above 10 km) are located at the borders of the map, where no measurements are available. Here, the solution is not constrained by the data: the mean Moho model and its standard deviation are equal to the equivalent statistics of the prior uniform distributions (mean 30 km, $\sigma \approx 14.5$ km), while the density of nuclei is homogeneous. Within the region delimited by the data, σ tends to increase away from data points and decrease in regions with consistent measurements. In particular, higher σ

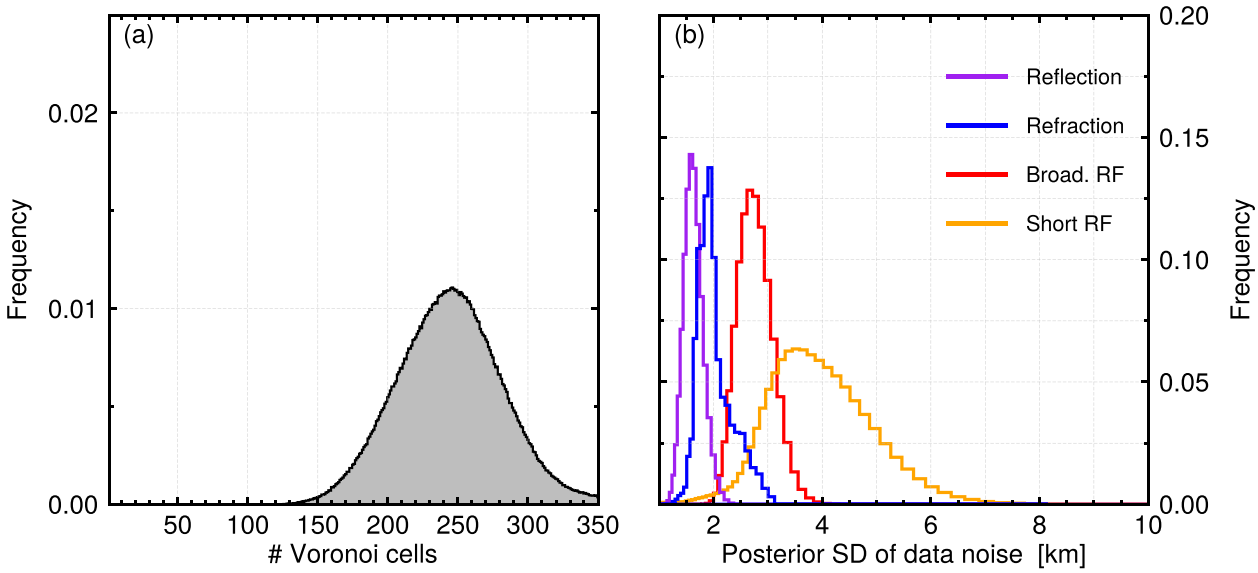

Figure 3. Marginal posterior distributions for (a) number of Voronoi cells and (b) posterior standard deviation (*SD*) of data noise. See text for details.

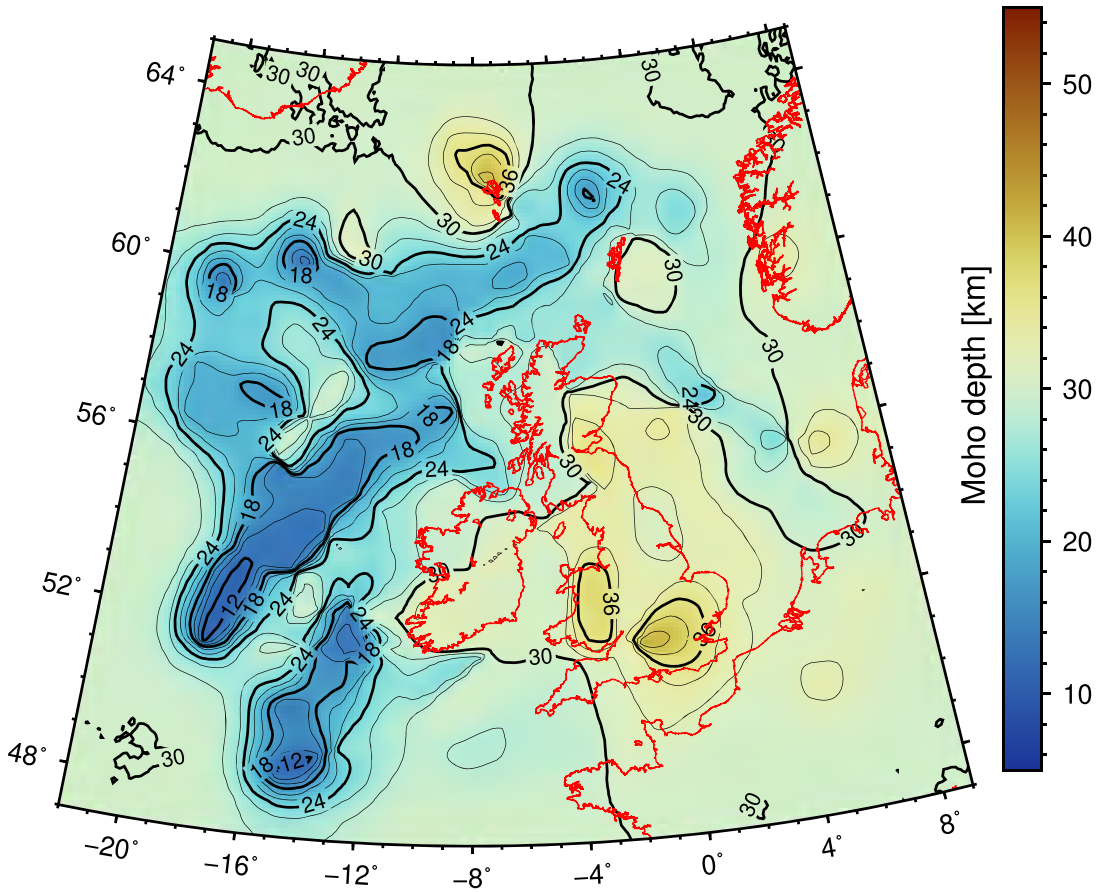

Figure 4. Results of the probabilistic Moho depth surface reconstruction. Mean surface computed from the ensemble of models belonging to the posterior probability distribution. Contours of Moho depth are drawn every 2 km and labelled every 6 km.

values are found where the data coverage is lower (e.g. $\sigma \approx 8$ km in SE England and 6 km off the NW coast of Ireland) or inconsistent (as with the anomalously thick crust in SE Great Britain). In addition, we note that around well constrained anomalies in Moho depth (as the 36-km-thick crust in Wales with low σ) a transition region characterized by a loop of higher σ develops. Similar features (called uncertainty loops) were observed in previous studies using the trans-dimensional Bayesian approach (Galetti *et al.* 2015; Zhang *et al.* 2018) as a consequence of the Voronoi cell parametrization

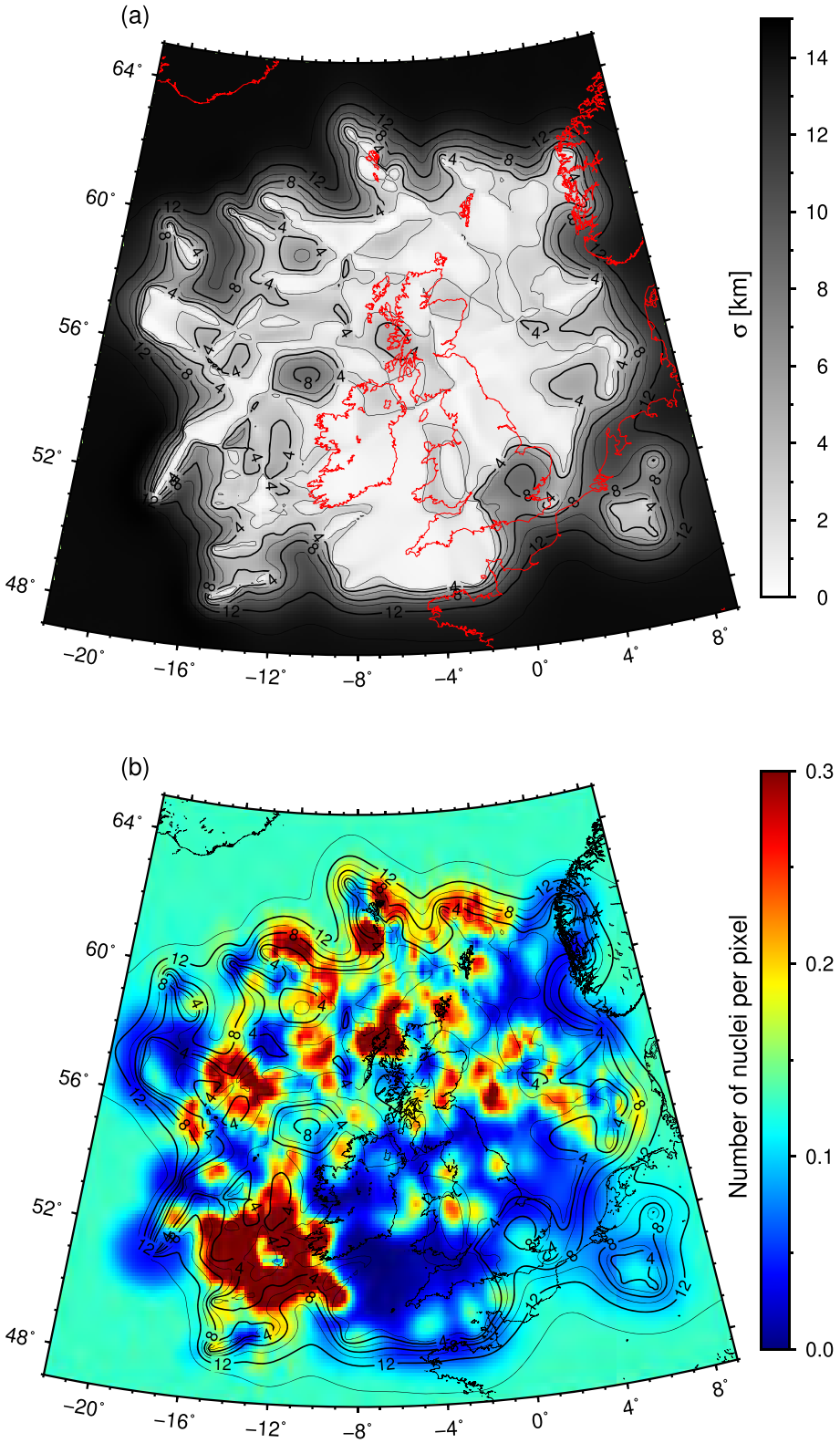

Figure 5. Results of the probabilistic Moho depth surface reconstruction. (a) Standard deviation surface. (b) Voronoi nuclei density computed from the ensemble of models belonging to the posterior probability distribution. In both plots, contours of standard deviation are drawn every 2 km and labelled every 4 km.

(Hawkins *et al.* 2019). In Appendix A2 (Fig. A3d) we show that uncertainty loops develop in a dataless transitional domain between well constrained and neighbouring regions even in a simple 1-D interpolation problem.

Spatial variations of σ and nuclei density are strictly related to the information contained in the data and to the complexity of the model. Taken jointly, the maps in Fig. 5 provide the key to understand the spatial resolution of the data. High frequency variations of σ are present where highly informative data require gradients in Moho depth and are usually associated with high posterior density of the Voronoi nuclei. For example, around the Porcupine Basin and in the NNW portion of the study area, refraction data are dense, coherent along the same line and with small posterior data noise (see Fig. 3b). In this case, σ tends to be low in close proximity of the refraction lines and high in between them, if the required variation in Moho depth is outside the range allowed by the data noise. As described in Appendix A2 (Figs A3, a–c), this results in high nuclei density, as our constant-value Voronoi parametrization requires more and smaller cells in order to create steeper gradients and accommodate small scale variations in Moho depth. On the other hand, large regions with low σ and low nuclei density (as in southern Ireland, eastern Celtic Sea and eastern Great Britain) correspond to areas in which spatially coherent data are fit (within uncertainties), with a few large Voronoi cells. This is the effect of our parsimonious approach: simpler geometries are preferred if the data do not justify higher complexity. The resulting small value of σ suggests that the data are sensitive to the long wavelength mean Moho depth, which is robustly constrained in these areas.

As in other inverse problems, the solution of our surface reconstruction is unlikely to be unique. Non-uniqueness can be investigated by plotting the entire marginal distribution for each point in the map along different profiles (Fig. 6). With this type of representation, the shape of the marginal PPD for each point can be directly observed. The mean value and 95 per cent credible intervals are plotted to better illustrate the lateral variations of Moho depth and associated uncertainties. Profiles are constructed along arcs of fixed longitude (-6.5°, Fig. 6a) and latitude (56°, Fig. 6b). Their location is shown with red lines in Fig. 2 for reference. It is clear that the PPD in some portions of the profiles is far from Gaussian and possibly multimodal. For example, the profile in Fig. 6(a) has a broad PPD at around 55.5° of latitude, where two distinct peaks can be distinguished. At this location on the map, the 1-D marginal posterior distribution for Moho depth (Fig. 6c) indicates values of Moho depth of about 29 and 24 km for the two peaks. Recently, it has been suggested that alternative choices of parametrization can alleviate the problem of multimodality in surface reconstruction (Hawkins *et al.* 2019).

We note that the smooth Moho structures obtained in areas where there is little data constraint do not necessarily preclude the existence of complex Moho structures. In such poorly constrained areas, our trans-dimensional parsimonious algorithm only encourages the simplest interpretations, which is not necessarily the closest solution to the real Earth structure. Consequently, as described in Appendix A2 (Fig. A3d), the final posterior standard deviation in regions with poor/absent data coverage is strongly dependent on implicit priors (choice of parametrization and the parsimonious nature of the algorithm), therefore it should be taken with caution. More generally, how to use posterior uncertainties from a Bayesian inversion for subsequent studies is a matter of on-going research (e.g. Visser & Markov 2019) and a deep discussion about this topic is beyond the scope of this work. However, as we have stressed thorough this paper, the outputs of a Bayesian inversion should not be considered alone without considering the 'prior belief' (implicit priors), that is how such outputs have been created. Possible developments along this line may involve the use of the full family of models sampled from the PPD which, implicitly, contains all the 'prior belief' used to generate such models.

4.2 Additional model statistics

Further insights on the statistical properties of our solution can be gained through the analysis of higher moments of the posterior distribution such as skewness and kurtosis (Fig. 7), which give additional information about the shape of the PPD. Skewness is the third standardized moment of a distribution and measures the asymmetry of a given distribution with respect to its mean. Positive (negative) values of skewness indicates that the distribution has longer tail on the right (left) side, therefore towards large (small) Moho depth values. For unimodal distributions this implies that the mean is therefore shifted from the mode towards the corresponding tail of the distribution. A skewness value of zero indicates a symmetric distribution such as the boundaries of our model where the PPD is equal to the uniform (and symmetric) prior distribution, or the majority of the Irish landmass where the posterior distribution appears symmetric.

Kurtosis is the fourth standardized moment of a distribution and its value is mainly influenced by the presence of outliers in the distribution. Higher kurtosis indicates infrequent extreme deviations (or outliers), as opposed to frequent modestly sized deviations. It has been usually interpreted as a measure of 'peakedness' or 'bimodality' of a given distribution although it does not provide a direct measure of these properties.

Given maps of skeweness and kurtosis it is possible to picture the shape of the posterior distribution at each point in our domain of interpolation. For example, high kurtosis is located along two of the refraction profiles on the westernmost part of our model. At the same location, skewness is strongly negative. This indicates that the posterior distribution here, is very peaked and with low amplitude tails towards small values of Moho depth.

As opposed to optimization methods which produce a single reconstructed surface, our approach provides a full posterior distribution of surfaces. We have already discussed the mean model and its standard deviation maps (Figs 4 and 5a). However, our probabilistic solutions allow to gain further insights on the Moho depth variations by examining additional statistics. In Fig. 8, we show the median and mode models of Moho depth. The median model is constructed by selecting the middle value of Moho depth from its marginal distribution at each point in the map. The median is less affected by outliers than the mean and provides a better measure of central tendency in case of skewed distributions. We note that regions of the map with strongly skewed distributions, such as the refraction profiles discussed before, the gradients at the Moho are stronger (with shorter wavelength) in the median model compared to the mean model which is smoother.

On the other hand, the mode (maximum *a posteriori*, MAP) model corresponds to the Moho depth value associated with the maximum of the marginal distributions at each location. The mode model tends to be rougher and preserves some features of the Voronoi parametrization. However, it indicates the most probable value of Moho depth at each point. Not only this could give a better idea about the transition of the Moho between regions but in this case, it also reduces the influence of the anomalously thick crust in SE Great Britain.

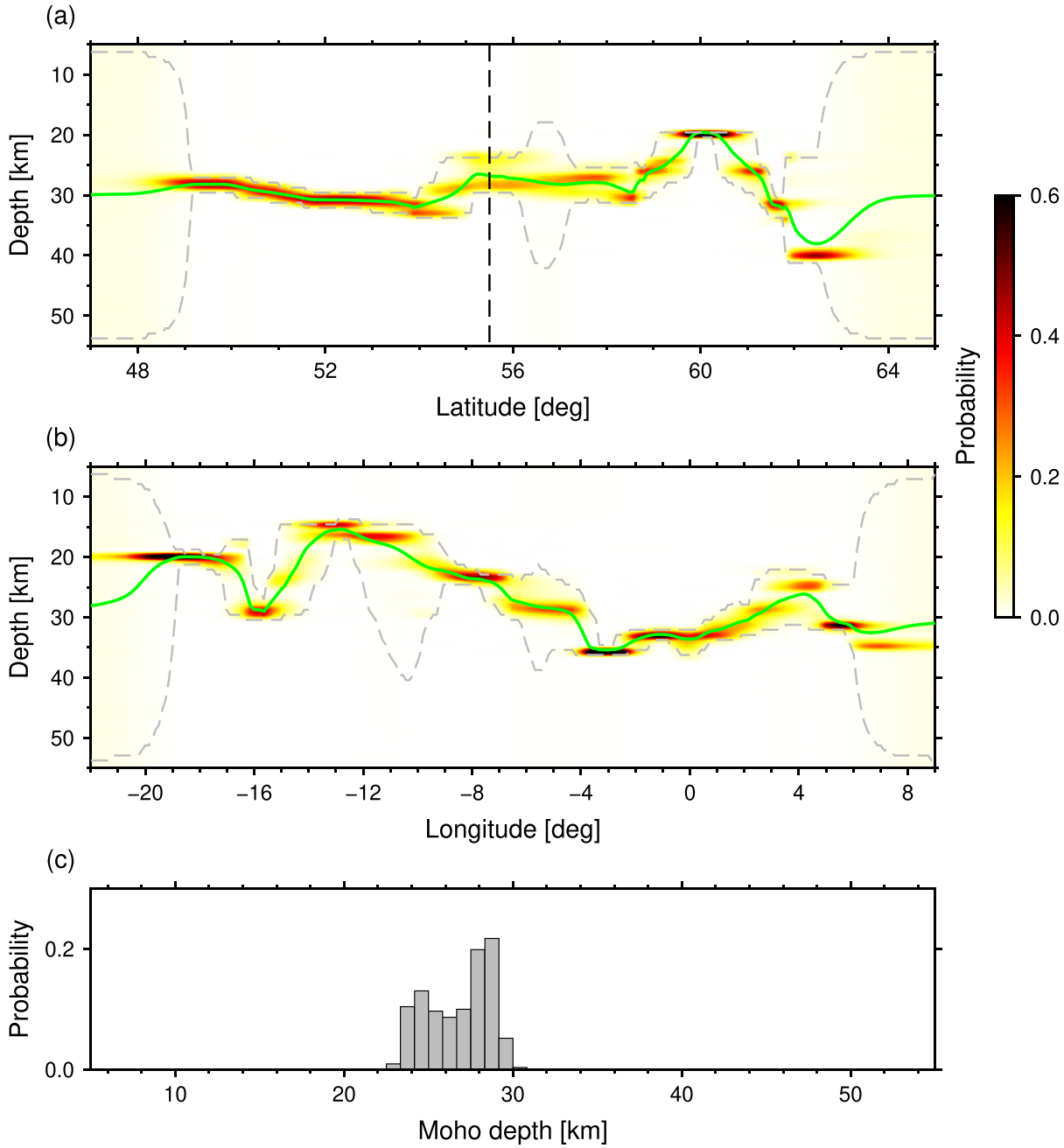

Figure 6. Full posterior probability distributions along the profiles indicated by the red lines in Fig. 2. These are cross sections at (a): fixed longitude of –6.5° and (b): fixed latitude of 56°. The green line is the mean and the grey dashed lines are the 95 per cent credible intervals of the distributions. In (c) the 1-D marginal distribution along the dashed line in (a) is shown. This corresponds to the full solution at –6.5° (lon) and 55.5° (lat).

We conclude this section by noting that the full PPD is made available in the Supporting Information, so that the interested readers can use the complete probabilistic solution according to their particular needs.

5 COMPARISON WITH PREVIOUS MODELS

In this section, we compare our results with the Moho depth models presented in previous works. As our complete solution is represented by an ensemble of models, a single model needs to be chosen to perform the comparison. The best data fitting model is discontinuous, being represented by a set of Voronoi cells. In contrast, the mean Moho depth model of Fig. 4 is smooth and an obvious candidate for this test.

Fig. 9 shows the comparison between the mean Moho depth model obtained in this study and the models presented in Kelly *et al.* (2007) and Baykiev *et al.* (2018). The corresponding difference maps are shown in Fig. 10 for a more quantitative comparison. Here, blue (red) values indicate that our model has a shallower (deeper) Moho than the one which it is compared with. The spatial domain for the comparison is determined by the geographical extent of each model, considering only regions for which the posterior σ of our results is less than 14 km. In addition, we have performed a

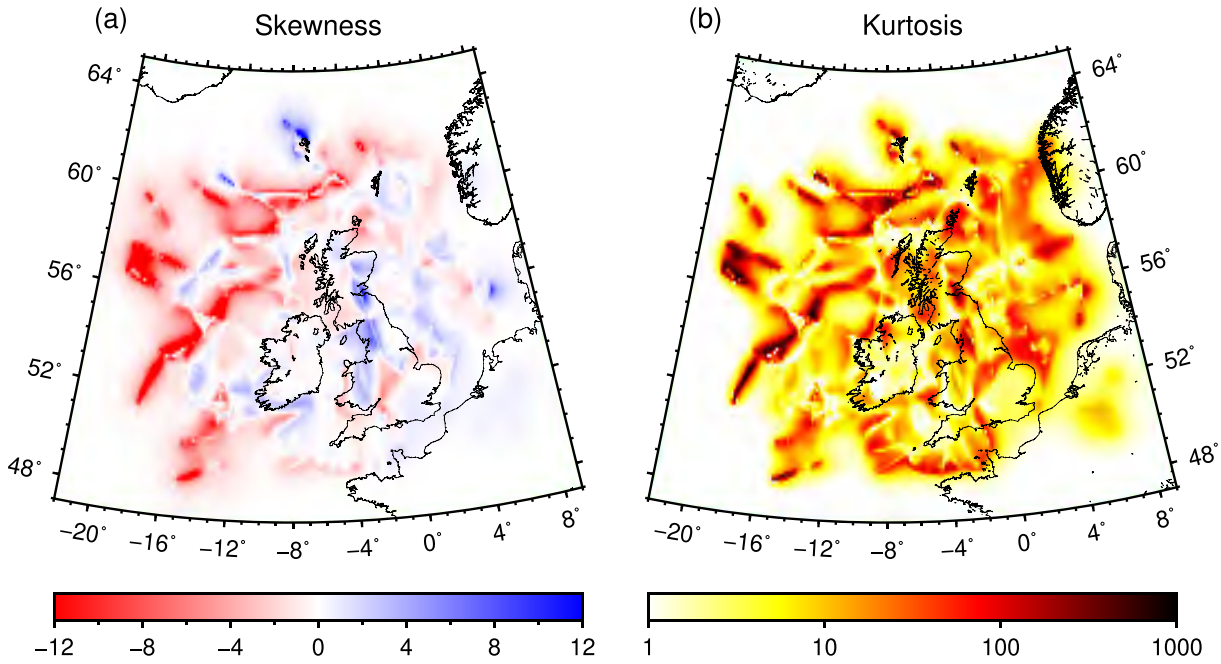

Figure 7. (a) Map of skewness of the posterior distribution. Negative (positive) values indicates that the distribution has left (right) tails. A value of zero indicate a symmetric distribution. (b) Map of kurtosis of the posterior distribution. High kurtosis is a proxy for highly peaked distributions.

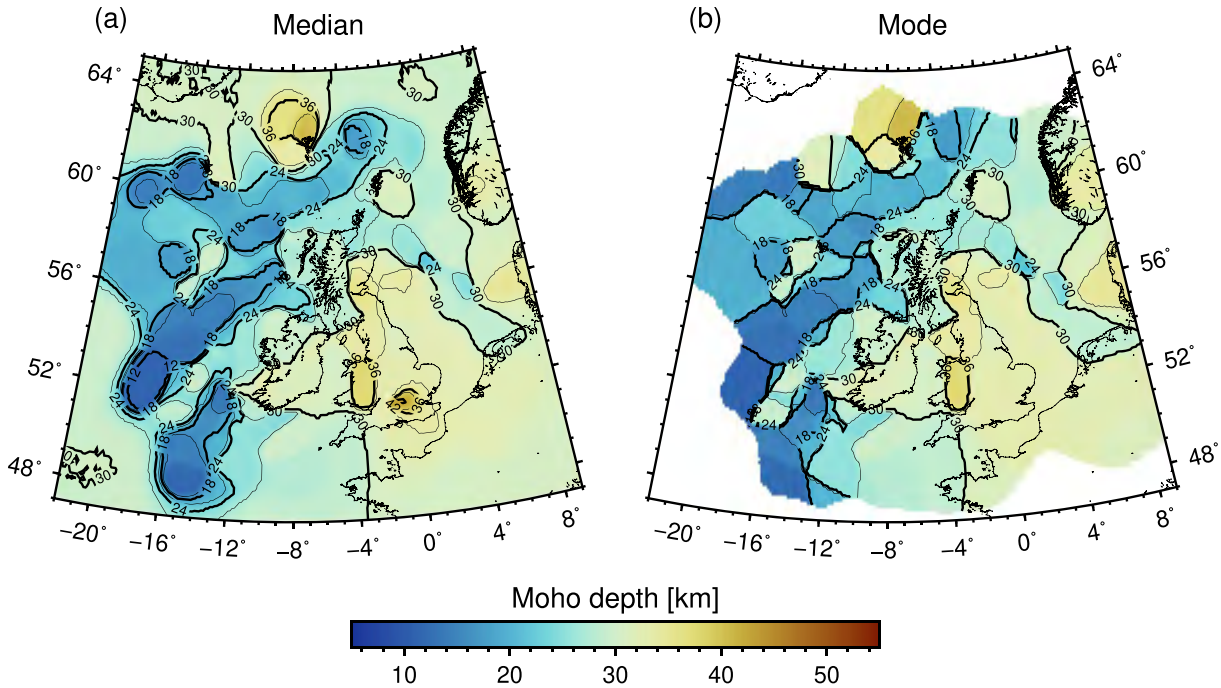

Figure 8. Posterior median (a) and mode (b) Moho depth models. For the purposes of visualization, the modal model is masked out in regions were the posterior standard deviation is above 14 km. Picking the maximum value of an almost flat histogram can introduce misleading artefacts in the map. Contours of Moho depth are drawn every 3 km and labelled every 6 km.

direct comparison between the model of Kelly *et al.* (2007) and the solution of our algorithm using exactly the same data set as in Kelly *et al.* (2007). The results of this comparison are shown and discussed in the Supporting Information.

The model of Kelly *et al.* (2007, Fig. 9c) relies exclusively on refraction data and has been generated using ordinary kriging using the code KT3D (Deutsch & Journel 1998). On the other hand, the Moho topography map of Baykiev *et al.* (2018) has been constructed from the same data set as in this study (without the refraction data in the Porcupine Basin), interpolated with the spline-in-tension approach of the GMT package *surface* (Wessel *et al.* 2013). From a visual comparison, both these models show shorter wavelength variations of Moho depth with respect to our mean model (Fig. 9). In any case, the differences in Moho depth related to these high frequency variations are generally comprised between ±2 km (Fig. 10) which is within the posterior σ estimated by our approach (Fig. 9b).

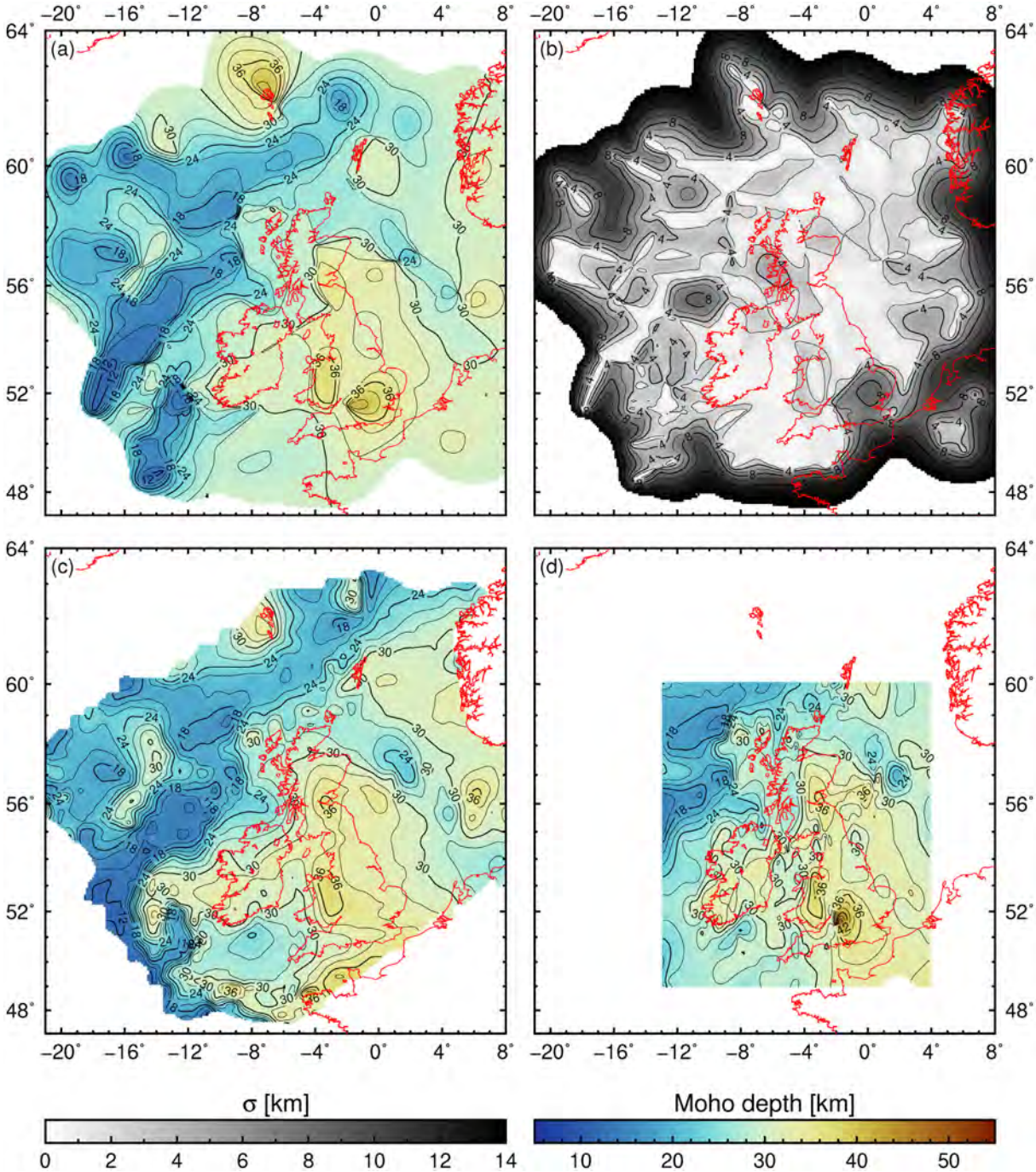

Figure 9. Comparison of our mean Moho depth model with previous studies. Panels (a) and (b) mean Moho depth and standard deviation of this study (same as Fig. 4). (c) and (d) Moho models from (Kelly *et al.* 2007) and (Baykiev *et al.* 2018), respectively. In all maps, regions for which the standard deviation of the posterior distribution is higher than 14 km are masked out. Contours are drawn every 2 km and labelled every 4 km in (b) and 6 km in (a), (c) and (d).

Differences of about ±4 km are found where different data are used across models. For example, the offshore south of Ireland and the western offshore of Norway are constrained by reflection data in our model. This produces about 4 km difference with respect to the model of Kelly *et al.* (2007). Regions with higher differences are represented by the offshore area west of Ireland and around the north, south and west edges of the map in Kelly *et al.* (2007) model (difference as big as ±10 km). In the first case, the results from new refraction data included in our database help to better constrain variations in Moho depth around the Porcupine Basin. In the second case, the observed differences are due to the fact that our model tends to converge to the mean value of Moho depth for the uniform prior distribution used (30 km) where no data are available.

As a consequence of our parsimonious algorithm, continuous structures in Moho topography are better highlighted in our model. One of these structures is the NE trend of shallow Moho offshore west of Ireland from the PAP through the Rockall Through (RT). From SW to NE along this trend, the Moho gets deeper from about 12 km to about 20 km. Following the structural orientation of the

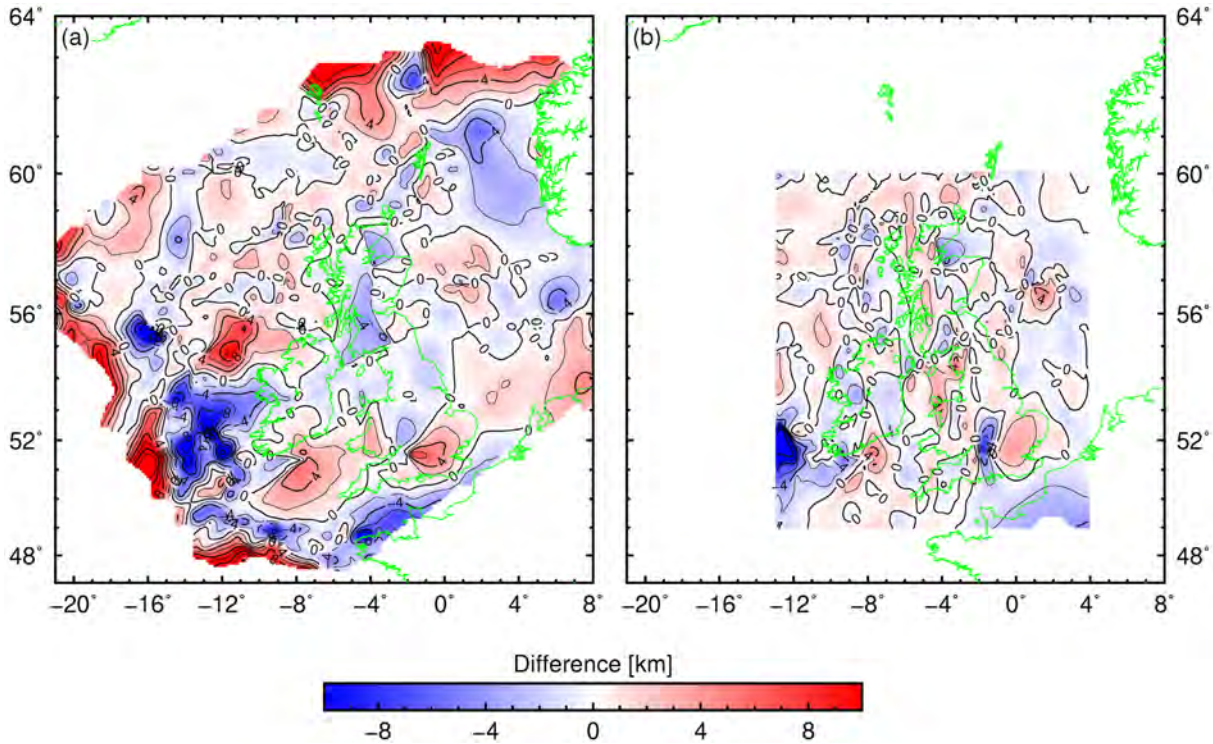

Figure 10. Comparison of our mean Moho depth model with previous studies. (a) and (b) Differences between our mean model and the models of (Kelly *et al.* 2007) and (Baykiev *et al.* 2018), respectively. In all maps, regions for which the standard deviation of the posterior distribution is higher than 14 km are masked out.

RT, this feature is displaced to the north where it continues, with a NE trend and Moho depth between 20 and 24 km, north of the Hebrides and between the Shetland and the Faroe Islands. Here, between the Orkney and the Shetland, another continuous feature in Moho depth (26–28 km) extends towards the North Sea with a SE trend. Both these features are present in the model of Kelly *et al.* (2007), but with lower lateral continuity.

Despite being smooth, our model robustly captures small scale features as well. For example, above the Rockall High, the NE trending structure characterized by a Moho depth of about 28 km is clearly defined, whereas in Kelly's model it appears smoothed out towards north. The posterior σ increases where the two models do not agree, indicating that the observed differences are within the estimated posterior uncertainties. The use of recent refraction data highlight the short wavelength Moho depth variations between the Porcupine Basin (16–18 km) and the Porcupine High (24–26 km).

Despite the major role of refraction and reflection data in building the interpolation surface, RF data help to determine variations of Moho depth onshore. In Ireland, for example, our model indicates a simple and smooth pattern of crustal thinning from south to north with the location of the main discontinuity (30 km contour line) that closely follows the surface trace of the Southern Uplands Fault. This confirms the results obtained by the analysis of RF data alone (Licciardi *et al.* 2014), indicating that RF data are constraining our model in this area. In both Kelly's and Baykiev's models, this feature is less clear due to high frequency variations in Moho depth. In the former, this is due to the lack of data, in the latter, to the choice of the interpolation algorithm. Onshore Great Britain, the main difference between our model and previous published models is located along the west coast of Scotland (Fig. 10). In this area, our mean model is between Kelly's and Baykiev's, which have deeper and shallower Moho respectively. The use of RF data in this region helps to better characterize roughly E-W oriented Moho variations across the Great Glen Fault, from about 34 km in the east to about 28 km to the west. This transition is smoother than in Baykiev's model and shows shallower Moho values than in Kelly's model, but it is associated with high posterior σ (more than 3 km).

6 CONCLUSION

In this work, we have presented the results of a Bayesian surface reconstruction algorithm to determine variations in Moho depth around the British Isles in a probabilistic sense. Our results are based on the most up-to-date compilation of Moho depth estimates in the region, assembled from a variety of different studies and obtained with different seismic methods. We have shown that our approach does not require user-subjective choices regarding spatial smoothing, nor an *ad hoc* weighting between different data-types. On the contrary, our trans-dimensional algorithm relies on adaptive parametrization and automatic weighting, based solely on the information contained in the data. Given its parsimonious nature, our probabilistic approach produces an ensemble of models for which the intrinsic trade-off between model complexity and data fit is addressed following Occam's principle of simplicity. In addition, it provides a suitable framework for uncertainty quantification. In this way, robust features are highlighted and uncertainties can be used for statistical testing of different scenarios such as lithospheric modelling or crustal studies.

Our results provide the first attempt to map the Moho topography of the British Isles in a probabilistic fashion. Despite the advantages provided by our Bayesian workflow, our approach still requires a number of assumptions to be made. These assumptions are involved in the choice of parametrization, in the type of data noise correlation,

and in the data selection process as highlighted in this work. In any case, we have shown that our Bayesian framework provides a flexible setting for testing and adapting the underlying assumptions ('implicit priors') if required.

Under the assumption of uncorrelated data errors, refraction and reflection data provides the main source of information to determine Moho depth variations. Although with larger posterior errors, the suite of RF data contained in our database yields constraints on crustal thickness onshore Ireland and Great Britain.

Moho depth variations of the mean model extracted from the PPD indicates deeper Moho (29–36 km) in continental areas(e.g. Ireland and Great Britain) and shallower Moho (12–22 km) offshore (e.g. in the Atlantic Ocean, west of Ireland). Maximum values of Moho depth are found in Wales (36–37 km), SE England (36–37 km) and near the Faroe Islands (38–40 km), while minimum values of Moho depth (11–12 km) are located in the Porcupine Abyssal Plain to the west of the Porcupine Basin and the Porcupine High.

Our results confirm the first order features of the Moho topography obtained in previous work, but we improve resolution by adding new constraints provided by recent available data, especially around the Porcupine Basin, onshore Ireland and Great Britain. In particular, NE trending features in Moho topography are highlighted above the Rockall High (about 28 km) and the Rockall Trough (with a NE directed deepening from 12 to about 20 km). A perpendicular SE oriented feature (Moho depth 26–28 km) is located between the Orkney and the Shetland, extending further SW in the North Sea. Onshore, our results highlight the crustal thinning toward the N in Ireland and an E–W oriented transition between deep (34 km) and shallow (about 28 km) Moho in Scotland. Differences with previously published models are in general around ±2 km, within the posterior uncertainties calculated with our approach. Bigger differences are located where different data are used between models or in less constrained areas where posterior uncertainties are high.

The mean and standard deviation maps presented in this work are available to be downloaded, together with the full PPD obtained with our Bayesian algorithm. The reader can access the results here: https://doi.org/10.6084/m9.figshare.11558976.v1. These probabilistic results open new perspective for future work. In particular, posterior errors on Moho depth can be used for statistical hypothesis testing or probabilistic modelling of the structure and/or composition of the crust and the lithosphere in the British Isles.

ACKNOWLEDGEMENTS

The authors would like to thank Ian Bastow, Andrew Curtis and an anonymous reviewer for their suggestions and comments, which helped to strongly improve the quality of the paper. We also thank Nicky White for providing the compilation of Moho depth measurements used in (Davis *et al.* 2012), Manel Prada and Brian O'Reilly for sharing the results in the Porcupine Basin. This research has been partially conducted with the financial support of Science Foundation Ireland under grant numbers 14/IFB/2742 and 13/RC/2092. N.P.A. research is funded by Austrian Science Foundation (FWF) under grant number M2218-N29.

REFERENCES

Amante, C. & Eakins, B.W., 2009. Etopo1 1 arc-minute global relief model: procedures, data sources and analysis, NOAA Technical Memorandum.

Artemieva, I.M. & Thybo, H., 2013. EUNAseis: a seismic model for Moho and crustal structure in Europe, Greenland, and the North Atlantic region, *Tectonophysics,* **609,** 97–153.

Baykiev, E., Guerri, M. & Fullea, J., 2018. Integrating gravity and surface elevation with magnetic data: mapping the Curie temperature beneath the British Isles and surrounding areas, *Front. Earth Sci.,* **6**(165), doi:10.3389/feart.2018.00165.

Bodin, T., Salmon, M., Kennett, B.L.N. & Sambridge, M., 2012a. Probabilistic surface reconstruction from multiple data sets: An example for the Australian Moho, *J. geophys. Res.,* **117**(B10), B10307, doi:10.1029/2012JB009547.

Bodin, T., Sambridge, M., Tkalčić, H., Gallagher, K., Arroucau, P. & Rawlinson, N., 2012b. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117**(B02301), doi:10.1029/2011JB008560.

Chadwick, R.A. & Pharaoh, T.C., 1998. The seismic reflection Moho beneath the United Kingdom and adjacent areas, *Tectonophysics,* **299,** 255–279.

Chen, C. *et al.*, 2018. From continental hyperextension to seafloor spreading: new insights on the Porcupine Basin from wide-angle seismic data, *J. geophys. Res.,* **123,** 8312–8330.

Choblet, G., Husson, L. & Bodin, T., 2014. Probabilistic surface reconstruction of coastal sea level rise during the twentieth century, *J. geophys. Res.,* **119,** 9206–9236.

Clegg, B. & England, R., 2003. Velocity structure of the uk continental shelf from a compilation of wide-angle and refraction data, *Geol. Mag.,* **140,** 4537–4567.

Davis, M.W., White, N.J., Priestley, K.F., Baptie, B.J. & Tilmann, F.J., 2012. Crustal structure of the British Isles and its epeirogenic consequences, *Geophys. J. Int.,* **190**(2), 705–725.

Deutsch, C.V. & Journel, A.G., 1998. *GSLIB - Geostatistical Software Library and Users Guide,* Oxford Univ. Press.

Di Leo, J., Bastow, I.D. & Helffrich, G., 2009. Nature of the Moho beneath the Scottish Highlands from a receiver function perspective, *Tectonophysics,* **479,** 214–222.

Funck, T., Geissler, W.H., Kimbell, G.S., Gradmann, S., Erlendsson, O., McDermott, K. & Petersen, U.K., 2016. Moho and basement depth in the NE Atlantic Ocean based on seismic refraction data and receiver functions, in *The NE Atlantic Region: A Reappraisal of Crustal Structure, Tectonostratigraphy and Magmatic Evolution,* eds Péron-Pinvidic, G., Hopper, J.R., Stoker, M.S., Gaina, C., Doornenbal, J.C., Funck, T. & Àrting, U.E., Geological Society London, Special Publications.

Galetti, E., Curtis, A. & Meles, G.A., 2015. Uncertainty loops in travel-time tomography from nonlinear wave physics, *Phys. Rev. Lett.,* **114,** 148501.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M. & Stephenson, J., 2009. Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for earth science problems, *Mar. Petrol. Geol.,* **26,** 525–535.

Green, P., 2003. Trans-dimensional Markov chain Monte Carlo, *Highly Struct. Stochast. Syst.,* **27,** 179–198.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrica,* **57,** 97.

Hawkins, R., Bodin, T., Sambridge, M., Choblet, G. & Husson, L., 2019. Trans-dimensional surface reconstruction with different classes of parameterization, *Geochem., Geophys., Geosyst.,* **20,** doi:10.1029/2018GC008022.

Husson, L., Bodin, T., Spada, G., Choblet, G. & Kreemer, C., 2018. Bayesian surface reconstruction of geodetic uplift rates: mapping the global fingerprint of Glacial Isostatic Adjustment, *J. Geodyn.,* **122,** 25–40.

Kelly, A., England, R. & Maguire, K.H., 2007. A three-dimensional seismic velocity model for northwestern Europe, *Geophys. J. Int.,* **171,** 1172–1184.

Kennett, B., Salmon, M. & Saygin, E., AusMoho Working Group, 2011. AusMoho: the variation of Moho depth in Australia, *Geophys. J. Int.,* **187,** 946–958.

Klemperer, S. & Hobbs, R., 1991. *The BIRPS Atlas: Deep Seismic Reflection Profiles Around the British Isles,* Cambridge Univ. Press.

Landes, M., Ritter, J.R.R., O'Reilly, B.M., Readman, P.W. & Do, V.C., 2006. A N-S receiver function profile across the Variscides and Caledonides in SW Ireland, *Geophys. J. Int.,* **166,** 814–824.,

Langston, C.A., 1977. Corvallis, Oregon, crustal and upper mantle receiver structure from teleseismic P and S waves, *Bull. seism. Soc. Am.,* **67,** 713–724.

Laske, G., Masters, G., Ma, Z. & Pasyanos, M., 2013. Update on CRUST1.0—a 1-degree Global Model of Earth's Crust, *Geophys. Res. Abstracts,* **15,** Abstract EGU2013–2658.

Li, Y., Gao, M. & Wu, Q., 2014. Crustal thickness map of the Chinese mainland from teleseismic receiver functions, *Tectonophysics,* **611,** 51–60.

Licciardi, A., Piana Agostinetti, N., Lebedev, S., Schaeffer, A.J., Readman, P.W. & Horan, C., 2014. Moho depth and Vp/Vs in Ireland from teleseismic receiver functions analysis, *Geophys. J. Int.,* **199,** 561–579.

Mackay, D.J.C., 2003. *Information Theory, Inference, and Learning Algorithms,* Cambridge Univ. Press.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151,** 675–688.

Malinverno, A. & Briggs, V., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics,* **69,** 1005–1016.

Meissner, R., Matthews, D. & Wever, T., 1986. The Moho in and around Great Britain, *Ann. Geophys.,* **4**(B6), 659–664.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, N.M., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.,* **21**(6), 1087–1092.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Mosegaard, K. & Tarantola, A., 2002. Probabilistic approach to inverse problems, in *International Handbook of Earthquake & Engineering Seismology (Part A),* pp. 237–265, Academic Press.

Ogden, C., Bastow, I., Gilligan, A. & Rondenay, S., 2019. A reappraisal of the H–κ stacking technique: implications for global crustal structure, *Geophys. J. Int.,* **219,** 1491–1513.

Piana Agostinetti, N. & Amato, A., 2009. Moho depth and Vp/Vs ratio in peninsular Italy from teleseismic receiver functions, *J. geophys. Res.,* **114,** B06303.

Piana Agostinetti, N. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181,** 858–872.

Piana Agostinetti, N., Giacomuzzi, G. & Malinverno, A., 2015. Local 3D earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **201,** 1598–1617.

Poggiali, G., Chiaraluce, L., Stefano, R.D. & Agostinetti, N.P., 2019. Change-point analysis of VP/VS ratio time-series using a trans-dimensional McMC algorithm: applied to the Alto Tiberina Near Fault Observatory seismic network (Northern Apennines, Italy), *Geophys. J. Int.,* **217,** 1217–1231.

Prada, M. *et al.*, 2017. Crustal strain-dependent serpentinisation in the Porcupine Basin, offshore Ireland, *Earth planet. Sci. Lett.,* **474,** 148–159.

Prodehl, C., Kennett, B., Artemieva, I.M. & Thybo, H., 2013. 100 years of seismic research on the Moho, *Tectonophysics,* **609,** 9–44.

Prodhel, C. & Mooney, W.D., 2012. *Exploring the Earth's Crust History and Results of Controlled-Source Seismology,* Geol. Soc. Am. Memoir 208.

Sambridge, M., Braun, J. & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophys. J. Int.,* **122,** 837–857.

Sambridge, M., Bodin, T., Gallagher, K. & Tkačić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A,* **371,** 20110547.

Shaw Champion, M.E., White, N.J., Jones, S.M. & Priestley, K.F., 2006. Crustal velocity structure of the British Isles: a comparison of receiver functions and wide-angle seismic data, *Geophys. J. Int.,* **166,** 795–813.

Smith, W.H.F. & Wessel, P., 1990. Gridding with continuous curvature splines in tension, *Geophysics,* **55**(3), 293–305.

Stein, M., 1999. *Interpolation of Spatial Data: Some Theory for Kriging,* Springer.

Tomlinson, J.P., Denton, P., Maguire, P.K.H. & Booth, D.C., 2006. Analysis of the crustal velocity structure of the British Isles using teleseismic receiver functions, *Geophys. J. Int.,* **167,** 223–237.

Visser, G. & Markov, J., 2019. Cover thickness uncertainty mapping using Bayesian estimate fusion: leveraging domain knowledge, *Geophys. J. Int.,* **219,** 1474–1490.

Watremez, L. *et al.*, 2016. Deep structure of the Porcupine Basin from wide-angle seismic data, in *Petroleum Geology of NW Europe: 50 Years of Learning - Proceedings of the 8th Petroleum Geology Conference,* eds Bowman, M. & Levell, B., Geological Society London.

Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J.F. & Wobbe, F., 2013. Generic mapping tools: improved version released, *EOS, Trans. Am. Geophys. Un.,* **74,** 409–410.

Zelt, C.A. & Smith, R.B., 1992. Seismic Traveltime Inversion for 2-D Crustal velocity Structures, *Geophys. J. Int.,* **108,** 16–34.

Zhang, X., Curtis, A., Galetti, E. & de Ridder, S., 2018. 3-D Monte Carlo surface wave tomography, *Geophys. J. Int.,* **215,** 1644–1658.

Zhu, L. & Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions, *J. geophys. Res.,* **105,** 2969–2980.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

SOM.pdf

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

APPENDIX: IMPLICIT PRIORS

A1 Data noise correlation and data selection

One of the assumptions of our approach is that each individual value of Moho depth is considered as an independent data point. While this eliminates the subjective selection of an adequate sampling radius, this assumption acts as an implicit prior in the inversion and it may introduce a bias in assessing the relative importance of different data types and uncertainty. As described in the main text, the results of our surface reconstruction are obtained by using a data set which has been preprocessed as little as possible in order to limit any user-induced bias in data selection. The results indicate that the two most densely spaced data types (refraction and reflection data) produce the lowest values of the posterior standard deviation therefore, they contribute more to the final reconstructed surface. It is of interest to study the behavior of our algorithm as the density of data points for these two data set is reduced and how this affects the posterior estimates of standard deviation. In addition, by downsampling the data, the effect of data error correlation is decreased in an empirical fashion. We have performed two additional tests that should be taken as a qualitative measure of these two combined effects. In the first (Fig. A1) and second test (Fig. A2) both reflection and refraction data points are downsampled by a factor of two and four respectively yielding a spacing of ≈10 and ≈20 km between points for refraction data and of ≈70 and ≈140 km for reflection data. The results indicate that the posterior estimates of data error (and relative importance of each data type) are affected by data density: with fewer data points (and with less influence of data error correlation), the estimated posterior standard deviation of refraction and reflection data increases. At the same time, fewer Voronoi cells are needed as the data constraints on the Moho surface are strongly reduced. These results suggest that our estimates of the posterior standard deviation of the data error in the joint inversion need to be taken as a rough estimates of the relative importance of each data

type that is affected by our assumptions about data error correlation and data selection.

A2 Choice of parametrization and rjMcMC parsimony

Our algorithm implements a 2-D Voronoi parametrization with constant Moho depth value within each cell. The choice of the parametrization can be considered as an implicit prior imposed on the interpolation scheme. In particular, two cases are worth to consider to understand the effect of this implicit prior: (i) (the behaviour of the Voronoi parametrization when the surface to be interpolated is different from blocks with constant values and (ii) what happens in case of missing data (the actual interpolation process). The first case has been studied in detail by (Hawkins *et al.* 2019) for the surface reconstruction problem. In this appendix, we present equivalent results on synthetic tests for a 1-D change point problem which, we think, is more intuitive and easier to understand. In these tests, the goal is to reconstruct the objective function y, given some observed data points (x, y) at location x, generated from the true function and perturbed with Gaussian noise with a standard deviation of 0.2 (the same for all data points). A parametrization similar to the one adopted for the surface reconstruction of the Moho discussed in the main paper is used here. In 1-D, this parametrization is described by a series of segments with constant y. The discontinuity between two adjacent segments is called a change point.

First, we set up three synthetic data sets with regularly spaced datapoints (Figs A3, añc), in which the transition between two constant y values of 1 and 4 is: (1) a single jump (Fig. A3a) located at $x = 3$; (2) a linear function from $x = 2$ to $x = 4$ (Fig. A3b) and (3) the sigmoid function $y = 1 + 3/(1 + \exp\{-6(x - 3)\})$ (Fig. A3c). In the first case, only one change point is required by the algorithm and the objective function is correctly retrieved. In the second case, our simple parametrization requires additional parameters (between 3 and 9 with a maximum probability at 4) to explain the linear trend between $x = 2$ and $x = 4$. In this case, the mean of the posterior distribution will tend to a staircase geometry. In a similar fashion, the third case indicates that additional parameters are required as well when the transition is a smooth function. From test (2) and (3) we can conclude that the more the transition is similar to a step function, the fewer parameters will be required by our algorithm. In both cases, the information about the increasing trend of y is preserved in the posterior distribution, but the mean value of the PPD may result in geometries which are different from the true one. In any case, we note that the true models (green lines) in both Figs A3(b) and (c) are comprised within one standard deviation of the mean. Therefore, this does not change the interpretation of the Moho map presented in the paper. Rather, it gives further insights on the possible limitations of the chosen parametrization. Areas with high density of Voronoi cells in Fig. 5 may indicate the presence of complex Moho geometries which are difficult to represent with our simple Voronoi parametrization.

Finally, we examine the implicit prior constraints implied by our parametrization in the interpolation process itself, meaning in the reconstruction of a function across a data gap. This is shown in Fig. A3(d) which uses the same data set of Fig. A3(a) but with the addition of two data gaps in the x intervals [2,4] and [5,7]. In principle, because of the lack of data, an infinite number of solutions are possible. In practice, the results indicate that our parametrization imposes some constraints on the values of y in these cases. In particular, in the [2,4] interval, rather than sampling the prior distribution ($U(0, 6)$) across the data gap, the algorithm produces a symmetric bimodal distribution with peaks determined by the y values of the data points adjacent to the gap. This behavior has proven to be typical to Voronoi cell parametrization (Hawkins *et al.* 2019) and results in a so called uncertainty loop Galetti *et al.* (2015): the value of σ as a function of x (red line in the lower plot) indicates high uncertainty in the transition region and low uncertainties on either side where y is well constrained by the data. The resulting mean of the posterior distribution increases linearly from 1 (at $x = 2$) to 4 (at $x = 4$). Note that although the algorithm detects the presence of a change point across the data gap, it fails in determining its location which is equiprobable between $x = 2$ and $x = 4$.

In the [5,7] interval, the algorithm does not require the presence of a change point. Within the data gap, it assigns similar y values as those required by the data points on the left and right of the interval. This is a clear example showing parsimony in action: rather than adding complexity where no information are available, the algorithm prefers the simplest solution with fewer parameters. Two observations regarding posterior uncertainties across a data gap can be made from this test. First, posterior uncertainties are determined by the magnitude of the y jump and nearby data values and errors. Secondly, the combined effect of our parametrization and parsimony imposes prior constraints on the interpolation across a data gap: rather than sampling the prior distribution, the algorithm prefers the simplest solution. This contributes to the estimated uncertainties on the interpolation itself across a data gap.

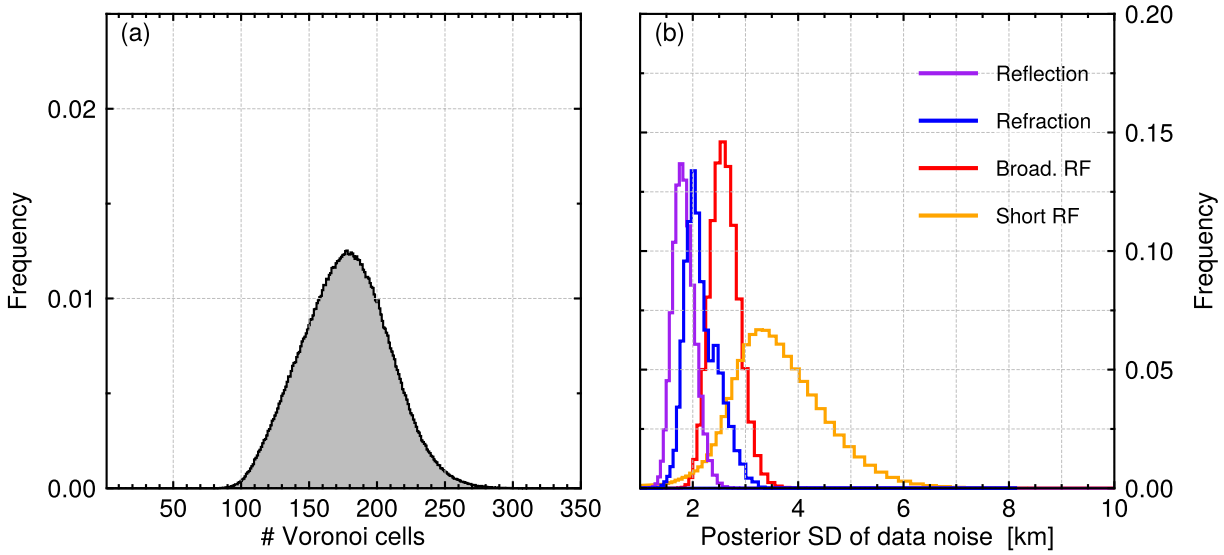

Figure A1. Marginal posterior distributions for (a) number of Voronoi cells and (b) posterior standard deviation (SD) of data noise. In the inversion, reflection and refraction data have been downsampled by a factor of two.

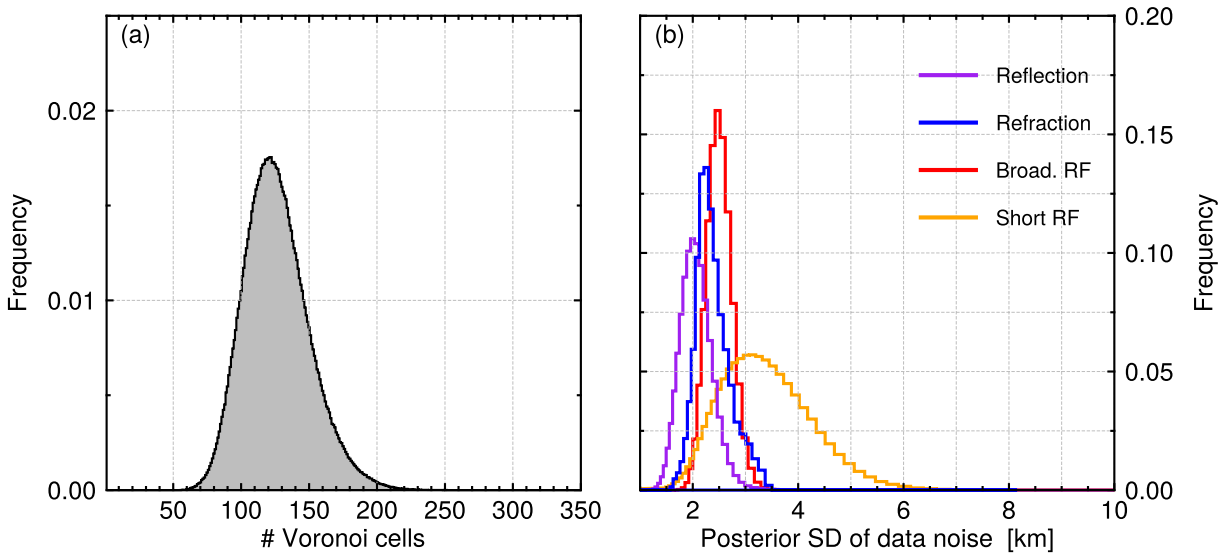

Figure A2. Marginal posterior distributions for (a) number of Voronoi cells and (b) posterior standard deviation (SD) of data noise. In the inversion, reflection and refraction data have been downsampled by a factor of four.

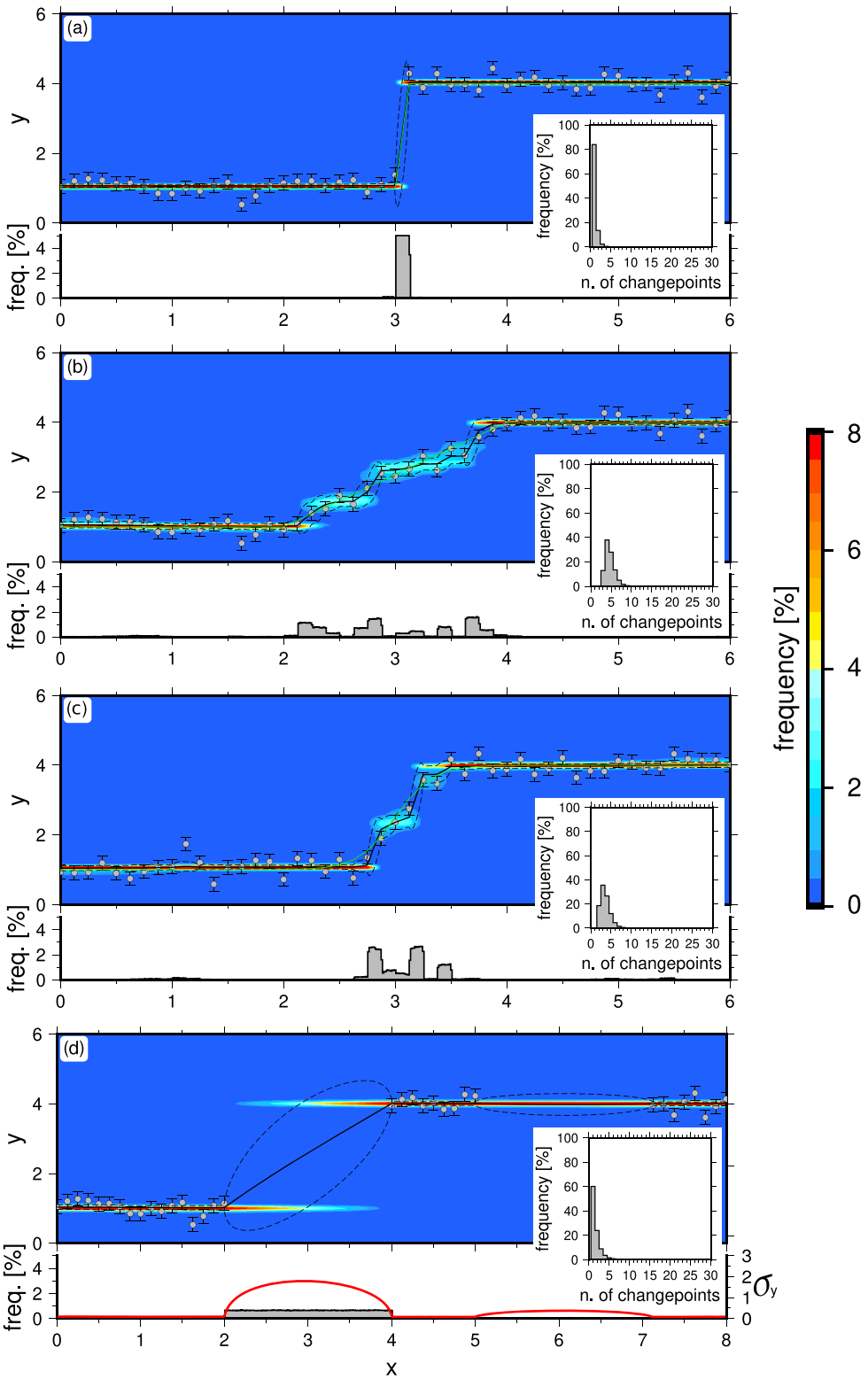

Figure A3. Results of the synthetic tests in 1-D to illustrate the effect of the chosen parametrization. (a–c) Results of three inversions with regularly spaced data points in which the nature of the discontinuity is changed from a simple jump in (a) to a linear function in (b) to a sigmoid function in (c). The 'observed' data (grey circles) are generated according to the true objective function (green line) and adding Gaussian noise with a standard deviation of 0.2 (error bars on each data point). d) Results of the inversion when portions of the data are missing. In all panels, the black solid (dashed) line is the mean (standard deviation) of the posterior distribution. The histograms of the number of change points and their probabilistic location along the x axis are plotted in each panel. In last panel we also show the standard deviation (red line) as a function of x.

Key words

Authors are requested to choose key words from the list below to describe their work. The key words will be printed underneath the summary and are useful for readers and researchers. Key words should be separated by a semi-colon and listed in the order that they appear in this list. An article should contain no more than six key words.

COMPOSITION and PHYSICAL PROPERTIES
Composition and structure of the continental crust
Composition and structure of the core
Composition and structure of the mantle
Composition and structure of the oceanic crust
Composition of the planets
Creep and deformation
Defects
Elasticity and anelasticity
Electrical properties
Equations of state
Fault zone rheology
Fracture and flow
Friction
High-pressure behaviour
Magnetic properties
Microstructure
Permeability and porosity
Phase transitions
Plasticity, diffusion, and creep

GENERAL SUBJECTS
Core
Gas and hydrate systems
Geomechanics
Geomorphology
Glaciology
Heat flow
Hydrogeophysics
Hydrology
Hydrothermal systems
Infrasound
Instrumental noise
Ionosphere/atmosphere interactions
Ionosphere/magnetosphere interactions
Mantle processes
Ocean drilling
Structure of the Earth
Thermochronology
Tsunamis
Ultra-high pressure metamorphism
Ultra-high temperature metamorphism

GEODESY and GRAVITY
Acoustic-gravity waves
Earth rotation variations
Geodetic instrumentation
Geopotential theory
Global change from geodesy
Gravity anomalies and Earth structure
Loading of the Earth
Lunar and planetary geodesy and gravity
Plate motions
Radar interferometry
Reference systems
Satellite geodesy
Satellite gravity
Sea level change
Seismic cycle
Space geodetic surveys
Tides and planetary waves
Time variable gravity
Transient deformation

GEOGRAPHIC LOCATION
Africa
Antarctica
Arctic region
Asia
Atlantic Ocean
Australia
Europe
Indian Ocean
Japan
New Zealand
North America
Pacific Ocean
South America

GEOMAGNETISM and ELECTROMAGNETISM
Archaeomagnetism
Biogenic magnetic minerals
Controlled source electromagnetics (CSEM)
Dynamo: theories and simulations
Electrical anisotropy
Electrical resistivity tomography (ERT)
Electromagnetic theory
Environmental magnetism
Geomagnetic excursions
Geomagnetic induction
Ground penetrating radar
Magnetic anomalies: modelling and interpretation
Magnetic fabrics and anisotropy
Magnetic field variations through time
Magnetic mineralogy and petrology
Magnetostratigraphy
Magnetotellurics
Marine electromagnetics
Marine magnetics and palaeomagnetics
Non-linear electromagnetics
Palaeointensity
Palaeomagnetic secular variation
Palaeomagnetism
Rapid time variations
Remagnetization
Reversals: process, time scale, magnetostratigraphy
Rock and mineral magnetism
Satellite magnetics

GEOPHYSICAL METHODS
Downhole methods
Fourier analysis
Fractals and multifractals
Image processing
Instability analysis
Interferometry
Inverse theory
Joint inversion
Neural networks, fuzzy logic
Non-linear differential equations
Numerical approximations and analysis
Numerical modelling
Numerical solutions
Persistence, memory, correlations, clustering
Probabilistic forecasting
Probability distributions
Self-organization
Spatial analysis
Statistical methods
Thermobarometry
Time-series analysis
Tomography
Waveform inversion
Wavelet transform

PLANETS
Planetary interiors
Planetary volcanism

SEISMOLOGY
Acoustic properties
Body waves
Coda waves
Computational seismology
Controlled source seismology
Crustal imaging
Earthquake dynamics
Earthquake early warning
Earthquake ground motions
Earthquake hazards
Earthquake interaction, forecasting, and prediction
Earthquake monitoring and test-ban treaty verification
Earthquake source observations
Guided waves
Induced seismicity
Interface waves
Palaeoseismology
Rheology and friction of fault zones
Rotational seismology
Seismic anisotropy
Seismic attenuation
Seismic instruments
Seismic interferometry
Seismicity and tectonics
Seismic noise
Seismic tomography
Site effects
Statistical seismology
Surface waves and free oscillations
Theoretical seismology

Tsunami warning
Volcano seismology
Wave propagation
Wave scattering and diffraction

TECTONOPHYSICS
Backarc basin processes
Continental margins: convergent
Continental margins: divergent
Continental margins: transform
Continental neotectonics
Continental tectonics: compressional
Continental tectonics: extensional
Continental tectonics: strike-slip and transform
Cratons
Crustal structure
Diapirism
Dynamics: convection currents, and mantle plumes
Dynamics: gravity and tectonics
Dynamics: seismotectonics
Dynamics and mechanics of faulting
Dynamics of lithosphere and mantle
Folds and folding
Fractures, faults, and high strain deformation zones
Heat generation and transport
Hotspots
Impact phenomena
Intra-plate processes
Kinematics of crustal and mantle deformation
Large igneous provinces
Lithospheric flexure
Mechanics, theory, and modelling
Microstructures
Mid-ocean ridge processes
Neotectonics
Obduction tectonics
Oceanic hotspots and intraplate volcanism
Oceanic plateaus and microcontinents
Oceanic transform and fracture zone processes
Paleoseismology
Planetary tectonics
Rheology: crust and lithosphere
Rheology: mantle
Rheology and friction of fault zones
Sedimentary basin processes
Subduction zone processes
Submarine landslides
Submarine tectonics and volcanism
Tectonics and climatic interactions
Tectonics and landscape evolution
Transform faults
Volcanic arc processes

VOLCANOLOGY
Atmospheric effects (volcano)
Calderas
Effusive volcanism
Eruption mechanisms and flow emplacement
Experimental volcanism
Explosive volcanism
Lava rheology and morphology
Magma chamber processes
Magma genesis and partial melting
Magma migration and fragmentation
Mud volcanism
Physics and chemistry of magma bodies
Physics of magma and magma bodies
Planetary volcanism
Pluton emplacement
Remote sensing of volcanoes
Subaqueous volcanism
Tephrochronology
Volcanic gases
Volcanic hazards and risks
Volcaniclastic deposits
Volcano/climate interactions
Volcano monitoring
Volcano seismology

© 2017 RAS, *GJI*