CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Explore Bristol Research

Sha, Z., Rougier, J., Schumacher, M., & Bamber, J. (2019). Bayesian model-data synthesis with an application to global Glacio-Isostatic Adjustment. *Environmetrics*, *30*(1), [e2530]. https://doi.org/10.1002/env.2530

Peer reviewed version

Link to published version (if available):
10.1002/env.2530

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2530 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms

Environmetrics 00, 1–25

Research Article, DOI: 10.1002/env.000

Bayesian model-data synthesis with an application to global Glacio-Isostatic Adjustment

Z. Sha[a,b], J. C. Rougier[a,*], M. Schumacher[a] and J. L. Bamber[a]

Summary: Glacial isostatic adjustment (GIA) is a crucial component in evaluating sea level change. The GIA process has been simulated globally from various physical forward models and it can also be measured locally at some GPS stations. In this paper we combine the physical model simulations and GPS measurements in a Bayesian hierarchical modelling framework to update global GIA. In common with many large-scale spatial modelling applications, there are two major challenges. One is the scale of the update, which is too large for naïve Gaussian conditioning. The other is the need to represent non-stationarity in the prior. We address the first challenge with the now well-established stochastic partial differential equations (SPDE) and integrated nested Laplace approximation (INLA) approach. For non-stationary global process, we propose two general models that accommodate commonly-seen geospatial patterns. We present and compare the GIA result for the two models, alongside the default option of assuming stationarity.

Keywords: spatial statistics; non-stationarity; stochastic partial differential equations; Bayesian hierarchical model; geophysical processes

1. INTRODUCTION

Glacial isostatic adjustment (GIA) is the slow viscous response of the solid Earth (lithosphere) due to the unloading of ice-sheets that covered much of the global land surface at the last glacial maximum (about 20 thousand years Before Present). At that time, North

[a] *School of Geographical Sciences, University of Bristol, Bristol, UK*
[b] *School of Mathematics, University of Bristol, Bristol, UK*
[*] *Correspondence to: J.C. Rougier, School of Mathematics, University of Bristol, Bristol, BS8 1TW.*
E-mail: J.C.Rougier@bristol.ac.uk

This paper has been submitted for consideration for publication in *Environmetrics*

America was covered by an ice-sheet that was up to 4km thick and global mean sea level was around 130 m lower than present due to the additional ice on land. The solid Earth response to changes in ice loading has important implications for many geophysical processes. GIA affects the Earth's shape, gravity field and axis of rotation via redistribution of mantle material. All of these changes, in turn, affect global sea level. To understand contemporary changes in sea level it is necessary, therefore, to have a reliable estimate of GIA. In general, this has been achieved using geophysical forward models (Guo *et al.*, 2012). These forward models require knowledge of past ice loading history and Earth structure (in particular lower mantle viscosity) to estimate the lithospheric response. Both these quantities are relatively poorly known. In addition, mantle viscosity varies laterally, which has been challenging to incorporate into forward models of GIA (Peltier *et al.*, 2015). As a consequence, large uncertainties in the present-day global GIA signal exist, evidenced by a comparison of 14 forward model solutions, which show significant regional differences between them (Guo *et al.*, 2012), disagreeing even on the sign of vertical land motion in some areas.

This paper presents a model-data synthesis method based on Bayesian hierarchical modelling (BHM, see, e.g., Banerjee *et al.*, 2004; Cressie and Wikle, 2011) for updating global GIA using global positioning system (GPS) data. The main feature of our approach is to use observations to adjust a model-based solution ('simulation'), by modelling explicitly the discrepancy between the simulation and the true process. This approach follows a long tradition in the field of computer experiments, in which such a discrepancy is a key part of the statistical model linking simulations, reality, and observations (see, e.g., Kennedy and O'Hagan, 2001; Craig *et al.*, 2001; Goldstein and Rougier, 2004, 2009). In contrast, a popular alternative is an approach in which the true process is modelled explicitly, and the simulation and observations are both treated as measurements; see Rougier *et al.* (2013) and the references therein, for the use of this approach in climate science. There are also hybrid approaches, such as Zammit-Mangion *et al.* (2014).

In spatial applications there are several advantages to modelling the discrepancy explicitly, as opposed to modelling the underlying process. First, it is more defensible to treat the discrepancy in a parsimonious way as an expectation-zero isotropic stochastic process. This presumes that the physical model simulation gets most of the large-scale 'non-stationary' and spatial correlation features right, in which case the discrepancy concerns itself mainly with smaller-scale local features. Second, the discrepancy is often the primary concern of researchers, who are interested in where their simulation performs poorly, and how large these regions are. Likewise, in comparing simulations, discrepancy maps can shed light on how different representations of process components cause different types of error. Stationarity is still a strong assumption, however, there will be situation where a non-stationary prior process is more appropriate. We introduce two general approaches to model such non-stationarity.

The outline of the paper is as follows. Section 2 summarizes the data, including forward simulations and observations of GIA, as well as some prior information on GIA. Section 3 describes our general approach to synthesizing a simulated field and observations, and the challenges of performing this operation at scale. We also review the stochastic partial differential equation (SPDE) approach to exploiting the sparsity of Gaussian Markov random fields (GMRF). Section 4 describes two different approaches to modelling non-stationary prior beliefs, and some useful variants. Section 5 shows the result of applying our approach to updating global GIA on a one degree map, and Section 6 is a brief conclusion.

2. DATA

In this paper, we model the GIA mean annual trend for the period 2005 to 2015. Over time-intervals as short as decades, the trend can be treated as a time-invariant spatial process and it is measured as the vertical land motion in mm/year. Our task is to assimilate the

forward simulation and the GPS data into an updated assessment of GIA, with predictive uncertainties.

The forward simulation we used is from the ICE6G-VM5 model (Peltier *et al.*, 2015), which we treat as our prior expectation of the GIA process. The ICE6G GIA solution is provided on a standard one-degree longitude-latitude.

The GIA process, as the vertical land motion, can be observed geodetically from the GPS stations (King *et al.*, 2010). Therefore, we prepared our GPS data from a global network of 4072 GPS stations for the same period (Schumacher *et al.*, 2018). The GPS observations are processed to include only the vertical bedrock movement so that the majority of non-GIA effects and artefacts are removed. Then an annual trend is estimated for each station to represent the observed vertical movement rate. The estimated standard errors of these trends are used as GPS measurement errors.

One important feature of GIA is that although the process affects the entire earth surface, there is negligible change in the middle of low latitude oceans. However, for various reasons, most simulations from physical models have deviations from zero in these regions, and there is little reliable GPS data available to correct these systematic errors. Therefore, to adjust these errors in our model, we created "zero-GIA regions" based on a further 7 of the 14 GIA solutions compared in Guo *et al.* (2012); other solutions are ignored since they are out-dated or have obvious flaws. The zero-GIA regions are generated by the following procedure.

1. Calculate the ensemble means and standard errors for each 1 degree pixel from the eight GIA simulations.
2. Retain the pixels that have values smaller than 0.3 mm/year, after exploring a few different thresholds with our experts.
3. Identify and remove dubious pixels by setting threshold values for the ensemble standard deviation. We remove pixels with standard deviations larger than 0.4 mm/year.

4. Connect the remaining pixels into polygons and remove polygons that have area smaller than $200\,\mathrm{km}^2$, which we regard as too small to be defined as a region.

[Figure 1 about here.]

Figure 1 shows the result of this procedure. We call the union of these polygons the 'zero-region' and the complement the 'region of interest'. The zero-region contains most of the ocean basins, plus some low-latitude lands, as expected. We also removed the GPS observations inside the zero-region since they are less reliable. Thus, the final GPS dataset contains 2515 observations, all inside the region of interest.

Finally, to incorporate our knowledge of the zero-region, we generate two sets of pseudo observations, one with 50 observations spread evenly across the zero region and the other with 50 points evenly spaced on the boundaries of the zero regions. These pseudo observations are set to be zero with an error of 0.1 mm/yr, which is about the same size as the smallest GPS measurement errors. Although one may argue that using pseudo observations can introduce bias in estimating the parameters, this is true only when there are a large number of them with inappropriate values. In this application, the number is relatively small compared to true observations. Moreover, there is also strong scientific support for the value chosen for the pseudo-observations: if there were any reliable GPS measurements in these regions, they should be very close to zero with tiny measurement error. Therefore, the effect on the parameter estimation over the region of interest is rather small.

3. MODEL AND METHODS

3.1. The Bayesian model-data synthesis approach

The outline of our approach is easily stated, although, for computation at scale, the devil is in the details. Without loss of generality, the process domain is taken to be the surface of the Earth, denoted $\mathbb{S}^2$. The simulation and the true process are both functions $\mathbb{S}^2 \to \mathbb{R}$; we denote

the simulation as m, with a lower-case letter because it is known, and the process as X, with a capital letter because it is unknown. For clarity, we will refer to X throughout as the 'latent process'. The stochastic process representing uncertainty about X is induced through the specification of a stochastic process for the discrepancy $\widetilde{X} := X - m$. The observations $y := (y_1, \ldots, y_n)^T$ are measurements on known linear maps of X, possibly made with error. The simulation/observations synthesis involves conditioning $\widetilde{X}$ on y. After conditioning, the updated discrepancy process $\widetilde{X} \mid y$ can be summarised in terms of point values, areal averages, or differences, as required, each quantified in terms of a posterior expectation, with uncertainty summarized in terms of a posterior standard deviation. If necessary, the latent process X can be reconstructed by adding m back to $\widetilde{X} \mid y$.

In this approach, the most demanding aspect is the specification of the stochastic process for the discrepancy $\widetilde{X}$. In contrast, the relationship between the latent process X and the observations y is typically well-understood, and the observation error reasonably well-quantified. We tackle the challenge of specifying a stochastic process for $\widetilde{X}$ in two steps. First, we treat $\widetilde{X}$ as expectation-zero, isotropic, and Gaussian. These strong modelling assumptions are very common in spatial statistics, but vary in their defensibility. In our approach, we defend them on the basis that we are modelling the discrepancy between simulation and latent process, for which we can expect the stochastic structure to be far simpler than the approach where we are modelling the latent process explicitly. Essentially, we are placing our faith in the simulation m, to have got the large-scale features of the latent process about right. Second, we do not commit to a specific choice of expectation-zero, isotropic, and Gaussian process, but allow its two parameters, the marginal variance σ^2 and the range parameter ρ to vary within some parametric family. We write $\theta := (\sigma^2, \rho)$, and term these the 'hyperparameters'. The marginal variance is the variance the process at any given location $s \in \mathbb{S}^2$. The range parameter, also called correlation length, is the spatial distance at which the process is almost uncorrelated, say correlation under 0.1.

Now we can summarize our statistical model in the standard hierarchical form, starting at the 'top', with the observation equation. We link each observation with the latent process via a specified linear map $\mathcal{A}_i : \mathbb{S}^2 \to \mathbb{R}$, so that

$$Y_i = \mathcal{A}_i X + E_i, \qquad i = 1, \ldots, n, \tag{1}$$

where E_i is measurement error. Thus, the observations are modelled as

$$Y = \mathcal{A} X + E, \tag{2}$$

where $\mathcal{A} := (\mathcal{A}_1, \ldots, \mathcal{A}_n)^T$, and the measurement error vector E is treated as probabilistically independent of X, and $E \sim \mathrm{N}_n(0, Q_Y^{-1})$, and Q_Y is a known precision matrix, typically sparse and often diagonal.

The linear map $\mathcal{A}_i$ encodes the 'footprint' of observation Y_i, with spatial locations outside this footprint not contributing to the value of Y_i. In our GIA application, the GPS observations are point measurements made at the known locations of the GPS stations. Therefore, the footprint of each GPS observation is just a point. In other work we are using measurements from the Gravity Recovery and Climate Experiment (GRACE) satellite (Luthcke *et al.*, 2013), where each observation corresponds to a 'mascon', a known spatial polygon, and we treat the GRACE observations as integrals over their mascons.

The latent process is modelled as

$$X = m + \widetilde{X}$$

where the stochastic process $\widetilde{X}$ is parametrized by the hyperparameters θ. After rearranging, the final hierarchical form of the model is

$$\widetilde{Y} \mid \widetilde{X}, \theta \sim \mathrm{N}_n(\mathcal{A}\widetilde{X}, Q_Y^{-1}) \tag{3a}$$

$$\widetilde{X} \mid \theta \sim \mathrm{GP}(\theta) \tag{3b}$$

$$\theta \sim \pi \tag{3c}$$

where $\widetilde{Y} := Y - \mathcal{A}m$, 'GP' denotes an expectation-zero isotropic Gaussian process on $\mathbb{S}^2$, and π is a prior distribution for the hyperparameters. In this formulation, the simulation m is used to adjust the observation Y prior to update the discrepancy process $\widetilde{X}$. If necessary, the hyperparameters can be extended to include parameters for $\mathcal{A}$ and Q_Y.

In this approach, we have the option to model the latent process as $\mathbb{E}(X; \alpha, \beta) = \alpha + \beta m$ and to include α and β among the hyperparameters, possibly with strong priors, and this has often been proposed in the computer experiment literature (e.g., Kennedy and O'Hagan, 2001). Our decision not to do so in this case is due to the strong belief that the simulation m is suitable without modification as the prior expectation. The simulation m is itself a product of a very complex modelling procedure designed to reflect both global and local physical constraints, and our primary interest is in the local deviation between m and the true latent process.

In most geophysical applications, the final output is usually in the form of a gridded map or aggregation over predefined regions. To summarize the updated discrepancy process we introduce additional known linear maps $\mathcal{B} := (\mathcal{B}_1, \ldots, \mathcal{B}_m)^T$, for which $\widetilde{Z} := \mathcal{B}\widetilde{X}$ are the quantities of interest. Figure 2 shows a graphical illustration of the model outline.

[Figure 2 about here.]

In the GIA application, we want to update the process on a one degree global map, therefore $\mathcal{B}$ maps the point values within each pixel to the arithmetic mean. However we simplify this

operation by using the centre of the pixel instead of the pixel average. This is because the grid resolution is sufficiently large to ensure that there is little variation within each pixel. Therefore, the linear operator $\mathcal{B}_i$ maps the point value of the GIA process at the same location as the pixel centre as the predicted value of that grid.

Then the marginal posterior distribution for $\widetilde{Z}$ factorizes as

$$\begin{aligned} \mathrm{p}^*(\tilde{z}) &= \int \mathrm{p}^*(\tilde{z}, \theta)\, \mathrm{d}\theta \\ &= \int \mathrm{p}(\tilde{z} \mid \tilde{y}\,; \theta)\ \mathrm{p}(\theta \mid \tilde{y})\, \mathrm{d}\theta. \end{aligned} \tag{4}$$

The first term in the integrand is a standard Gaussian update (Rue and Held, 2005, Ch 2) and will be discussed further below, in Section 3.2. The second term in (4) is unlikely to have a closed-form expression. This type of integration can be approximated using the method of Integrated Nested Laplacian Approximations (INLA, see Rue *et al.*, 2009; Martins *et al.*, 2013).

3.2. Computational issues

We now review the challenge of computing $\mathrm{p}^*(\tilde{z})$, and recent developments. Putting aside the integration over θ, the challenge for computing (4) is that $\mathrm{p}(\tilde{z} \mid \tilde{y}\,; \theta)$ involves factorising a $k \times k$ precision matrix, where k is the number of basis functions representing the latent process X. This is the dominant cost as it requires $\mathrm{O}(k^3)$ flops to compute. Current desktop technology tends to be computationally expensive for k much larger than a thousand. This is not nearly enough for some applications, particularly those where the latent process is defined on the whole of the surface of the Earth. Long computing time is also a major bottleneck during code development.

There are some simple work-arounds. Thinning or aggregating the observations down to less than a thousand is one possibility. Another is splitting the update into separate regions, each containing less than a thousand observations. These are valuable pragmatic approximations,

but can be risky when the range parameter of the discrepancy is uncertain, because the value of the range parameter affects the accuracy of thinning, aggregating, or splitting. Lindgren *et al.* (2011, sec. 1) discuss other more technical approaches.

Therefore we look for a different type of approximation, which reduces the dominant cost to below $\mathrm{O}(k^3)$. Gaussian Markov random fields (GMRFs) are one answer. As discussed in Rue and Held (2005), GMRFs defined on a finite-dimensional random vector can exploit sparsity in the precision matrix. This sparsity is represented by an undirected graph where the vertices are the elements of the vector, and the absence of an edge between two vertices indicates a conditional independence, usually induced by a neighbourhood structure. The difficulty of applying a GMRF approach directly is that $\mathbb{S}^2$ would need to be discretized into a finite number of pixels. Having done this, though, the natural neighbourhood structure would link pixels with a common boundary (a first-order neighbourhood scheme), for which the computational cost of factorizing the precision matrix Q_{X_K} is $\mathrm{O}(k^{3/2})$. However, it is not straightforward to configure the resulting precision matrix to approximate an isotropic Gaussian process with variance and range parameters. It is also somewhat arbitrary to pixelate the continuous domain $\mathbb{S}^2$, and wasteful to do so if some regions of $\mathbb{S}^2$ are more interesting and more highly-observed than others.

The breakthrough came with Lindgren *et al.* (2011); see also Simpson *et al.* (2012), although be warned that the notation in these two papers is at variance. Whittle (1954) had shown that an isotropic Gaussian process with a Matérn covariance function arose as a solution to a particular stochastic partial differential equation (SPDE). Lindgren *et al.* were able to use this insight to construct a finite-dimensional approximation to a Matérn Gaussian process for $\widetilde{X}$ of the form

$$\widetilde{X}(s) \approx \sum_{j=1}^{k} W_j \, \psi_j(s), \tag{5}$$

in which the ψ_j are specified according to a triangulation of $\mathbb{S}^2$, which can be adapted to the

needs of the application. Lindgren *et al.*'s particular choice of ψ as piecewise linear functions induces a GMRF specification for the vector $W := (W_1, \ldots, W_k)^T$, with a sparse precision matrix Q_W with a simple parametrization in terms of (σ^2, ρ), where ρ is now interpreted as the range parameter for a Matérn covariance function. Without this simple specification in terms of the hyperparameters, most of the computational advantages would be lost. Lindgren *et al.* (2011) and Simpson *et al.* (2012) show that the approximation error in (5) is $\mathrm{O}(h)$ where h is a measure of the triangle size, such as the radius of the largest inscribed circle, or the longest edge. Thus the hierarchical model from (3) becomes, under the approximation in (5),

$$\widetilde{Y} \mid W, \theta \sim \mathrm{N}_n(AW, Q_Y^{-1}) \tag{6a}$$

$$W \mid \theta \sim \mathrm{N}_k\big(\mathbf{0}, Q_W^{-1}(\theta)\big) \tag{6b}$$

$$\theta \sim \pi \tag{6c}$$

where A is $(n \times k)$ with $A_{ij} := \mathcal{A}_i\, \psi_j$, and $Q_W(\theta)$ is a sparse precision matrix with a simple parametrization in terms of $\theta = (\sigma^2, \rho)$.

When combined with sparsity in Q_W, sparsity in A and Q_Y (which is often diagonal) implies sparsity in the precision matrix of $W \mid \{Y; \theta\}$, reducing the cost of conditioning from $\mathrm{O}(k^3)$ to something more like $\mathrm{O}(k^{3/2})$. Sparsity in A arises naturally when the observations have small spatial footprints, because the ψ_j are localized in the triangulation, and thus also have small spatial footprints. Any observation Y_i whose footprint $\mathcal{A}_i$ does not overlap ψ_j has $A_{ij} = 0$. The interplay between the small footprints in $\mathcal{A}$ and the basis functions ψ is discussed in more detail in Simpson *et al.* (2012, sec. 3.6). The more complex interplay between $\mathcal{A}$, $\mathcal{B}$, and ψ is discussed in Zammit-Mangion and Rougier (2018).

4. MODELLING NON-STATIONARITY

In many applications, the latent process is only locally stationary and this may also be true for the discrepancies between the process and the simulations. The nature of the non-stationarity in geophysical processes usually falls into one of two categories: (1) the process is defined only on a subset of the domain; (2) prior knowledge of the process varies by region; the latter usually arise from strong regional differences. The SPDE approach outlined in Section 3.2 can be extended to non-stationary processes, notably those defined by regions with different behaviours.

Before introducing the models for non-stationary processes, we need to consider the triangulation for the SPDE approximation. Let the domain be partitioned as regions $\{\Omega_i\}_{i=1}^p$, where we assume that the boundaries between the regions are made up of linear segments. If we respect the boundaries by using their segments as triangle edges, then automatically-generated triangulations can have short edges and sharp angles, which is inefficient. On the other hand, if we triangulate the whole domain first in an efficient manner, and assign each triangle to a region according to its centroid, then we distort the boundaries between regions, and have less control over the resolution in each region.

The approach we adopt below is a compromise, in which we tolerate a small amount of distortion in order to derive an efficient triangulation of the whole domain, and variable resolution in the regions. We create a vertex set for each region (varying the resolution as appropriate), and then merge these sets, to cover the entire domain. Then we re-triangulate based on the vertices, to control for short edges and sharp angles. Finally, we assign each triangle to a region according to its centroid, and in this way we distort the regions slightly. If there is too much distortion we can modify the approach, for example by increasing resolution near the boundaries. The `R` codes for triangulation and building the corresponding precision matrices is based on Bakka *et al.* (2016) and more extended discussion can be found in (Bakka, 2016, pp.96-116). The `R` script and functions that produce all the plots and results

in this paper are included in the supplementary materials. We also provide a document with detailed comments and instructions for using the functions and scripts. Figure 3 shows an example of this approach to create a triangulation of the Earth with high resolution over the oceans and low resolution over land.

[Figure 3 about here.]

4.1. Subset model

Some processes are defined only on particular regions. For example, sea level change is only meaningful over the ocean and around the coastal regions. The domain of this process is connected globally by the oceans but separated locally by the lands. For sea level changes at any two points separated by land, their correlation is more likely to depend on the path connecting the two points along the coastline rather than the Euclidean distance across the land. Thus the land introduces non-stationarity.

In this case, it is natural to model the process only on the subset of interest, and we call this approach the *subset model*. Denote by $\Omega_s \subset \mathbb{S}^2$ the subset of the domain where the process is defined. The GMRF approximation of the process can be built in the same way as the global stationary model using a triangulation over Ω_s only. Then subset model retains the same form as the BHM in (6).

The left panel of Figure 4 provides an illustration of the subset model on the plane. The domain is a square with side length equal to 5. Suppose the process has a correlation length equal to 2.5, then for an adequate finite element approximation, the maximum edge length of the mesh triangles within the study region should be much smaller than 2.5: we choose the maximum edge length to be 0.5. The process is defined over the whole square except for the blank region Ω_0 in the middle. Consider the correlation between two pairs of points AB and AC as shown in the plot. The Euclidean distances between them are the same, so in a stationary model they both have correlations around 0.61. However, in the subset model,

the correlation between AB is smaller than 0.01, which reflects the fact that A and B are separated by Ω_0; informally, the path in this region is far longer. The correlation between AC is 0.77. This is slightly higher than the stationary model (0.61) because after removing Ω_0, the rest of the triangles get more weights in building up the precision matrix.

[Figure 4 about here.]

It is possible to re-weight the parameters near the subset boundaries so that the parameter stays the same in the entire region of interest. This can be done by introducing varying parameter values near the boundaries, and falls into the class of partition models that are to be introduced in the next subsection. To differentiate from the partition models, we keep the parameters un-normalised in the subset model. In some applications, local variation of the variance and correlation length may be desirable, for example, if the land acts as a reflecting boundary for ocean processes.

4.2. Partition models

There are also processes which are well-defined over the whole domain, but which have varying spatial characteristics. A typical example is an atmospheric process that shows very different behaviours across the oceans, coasts and land. We propose *partition models* to capture such spatial heterogeneity. First the domain is partitioned into p regions such as oceans, coastal regions and land. As before, denote this partition by $\{\Omega_i\}_{i=1}^p$. Then there are two approaches. The first is a decomposition of the process according to the region, hence we call it a *process partition model*. The second uses a single process over the domain but varies the hyperparameters according to the region, and therefore we call it a *parameter partition model*.

In a process partition model, the process is decomposed into independent sub-processes defined on each $\Omega_i, i = 1, \ldots, p$. If each observation is associated with exactly one Ω_i, then

independence in the prior distributions for each Ω_i implies independence in the posterior distribution as well.

In contrast, the parameter partition model allows correlation between the regions in the prior distribution, and therefore in the posterior distribution as well. The latent process over the domain is modelled as a single Matérn Gaussian process, but the hyperparameters θ become a function of the location s:

$$\theta(s; \theta_1, \ldots, \theta_p) = \theta_i, \qquad i : s \in \Omega_i. \tag{7}$$

The parameter partition model takes exactly the same form as the global stationary model in (6) but there are p sets of θ_i in the covariance function to be estimated. The precision matrix for such locally stationary process is guaranteed to be positive-definite by the SPDE construction (Lindgren *et al.*, 2011). For numerical reason, though, it can be beneficial to have the hyperparameters change smoothly across the boundary, and to increase the density of the triangulation near the boundary.

The right panel of Figure 4 illustrates the spatial correlation in a parameter partition model. The four points $ABCD$ are the vertices of a square. Instead of removing Ω_0 in the middle, we assume the process has a correlation length $\rho_0 = 3$ within this region, and $\rho_s = 2.5$ outside it, as before. The resolution of the triangulation within this region can remain the same as the outside region since the correlation length is longer; however we choose to use larger triangles to reduce the computational cost. The correlation between AD is 0.61 and the correlation between BC is 0.56. For AB and CD , which cross the boundary, the correlations are 0.60 and 0.59.

4.3. Constrained partition models

In the above examples, the partition of the spatial domain is determined by physical boundaries. It is also possible to define the regions based on our knowledge and interests. In

fact the latter is quite common as human activities and studies are not evenly distributed over the Earth. Therefore we might be more certain about the behaviour of a process in particular regions, and we can use this knowledge to our advantage, to reduce the number of hyperparameters.

We consider one common case in more detail, which will also feature in the illustration in Section 5. Consider a region where the process is known to high accuracy: for simplicity, and without loss of generality, suppose the process is known to be near-zero throughout the region, which we therefore term the 'zero-region', and denote as Ω_0. We can use a combination of hyperparameters and 'pseudo-observations' to enforce both the near-zero value in Ω_0, and also the continuity of the process across the boundary of Ω_0. First, we set the correlation length for Ω_0, so that it is at least half the length of the longest diagonal. Then we introduce pseudo-observations each with value zero and very small error, roughly equally-spaced inside Ω_0. When combined with the long correlation length for Ω_0, conditioning on these observations has the effect of constraining the process inside Ω_0 to be near-zero, and constraining the process just outside Ω_0 to be close to zero. This holds regardless of the variance for Ω_0, and therefore a common variance hyperparameter can be used both inside and outside Ω_0.

This approach is more attractive than the alternative of only modelling the process outside the zero-region, for two reasons. The first has already been noted: continuity across the boundary of the zero-region. The second is that the zero-region is incorporated within the inference, and does not have to be treated separately. This simplifies the code and the packaging of results, and reduces the possibility of error. The additional cost of more vertices is slight, because the triangulation in the zero-region can be low-resolution.

5. APPLICATION TO GIA AND RESULTS

5.1. Outline of the calculation

First we need to generate the triangulation used in the GMRF representation of GIA through the SPDE appraoch. The triangulation needs to have at least the same resolution as the prediction required, and for efficiency the triangles should have similar sizes and shapes. The final presentation of the updated GIA is on a one degreee map. The grid is non-regular in shape because it has a pixel size of about 100 km at the equator and less than 10 km at high latitude; therefore it does not have the same resolution globally. To achieve a uniform representation of GIA, we generate an equal area lattice on the sphere using 30,000 Fibonacci points (see, e.g., González, 2009) as the starting vertices. Euler's formula then implies that this gives about 60,000 triangles with an average spacing of 100 km. This is the desired prediction resolution and the one degree map is projected from these triangle vertices.

Then we need to adjust the observations by the prior mean field m as explained after (3). Thus, the ICE6G gridded values are mapped to the GPS station locations by linear interpolation, and then subtracted from the corresponding GPS values. This gives values for $\widetilde{Y}$ in (3). We use INLA to update the discrepancy $\widetilde{X}$ on the triangulation by conditioning on $\widetilde{Y}$ and integrating out the hyperparameters θ according to our prior given below in Section 5.2. The updated discrepancy field is then mapped back to the ICE6G grid, and the GIA process is reconstructed by adding back the simulation m. The result is an object similar to a GIA simulation, but representing the posterior expectation of GIA. Additionally, each pixel in the result is accompanied by a measure of uncertainty, the predicted standard deviation.

Given the nature of our *a priori* non-stationarity, we prefer to use the constrained partition model described in Section 4.3. For the zero-region, we choose a correlation length equal to the diameter of the Earth, which is long enough for a sparse representation of the zero-region and also keeps the precision matrix of the latent process from being singular. We

use 50 pseudo-observations with an error of 0.1 mm/yr, which is about the same size as the smallest GPS measurement errors, spread evenly through the zero-region.

By way of contrast, we also present two other approaches: modelling the process on the entire Earth as *a priori* stationary, and modelling only the region of interest, as *a priori* stationary. In the latter, we use 50 pseudo-observations spread evenly around the boundary of the zero region, to enforce continuity. In all three approaches, there are just two hyperparameters: the variance σ^2 and the correlation length ρ.

5.2. Prior distribution for the hyperparameters

In our application, $d = 2$, the dimension of the domain, and we choose $\nu = 1$, the smoothness parameter of the Matérn covariance function. This implies $\alpha = 2$, where α is a parameter of the SPDE used to induce the precision matrix Q_W in (6). In the equations below we will use these explicit values of d, ν, and α to simplify some expressions.

We use the `R-INLA` package for computation; see Lindgren and Rue (2015). This package expects the Matérn hyperparameters (σ, ρ) to be specified in terms of the alternative parameters (κ, τ), where

$$\sigma^2 = \frac{1}{4\pi\,\kappa^2\,\tau^2}, \quad \rho = \frac{\sqrt{8}}{\kappa}. \tag{8}$$

Note that under this representation, ρ is the distance at which the correlation function has fallen to about 0.13 rather than the commonly used values 0.05 and 0.1. `R-INLA` represents the prior for (τ, κ) as

$$\log \kappa = \kappa_0 + a_1\,\theta_1 + a_2\,\theta_2 \tag{9a}$$

$$\log \tau = \tau_0 + b_1\,\theta_1 + b_2\,\theta_2 \tag{9b}$$

where all terms on the right-hand side bar are specified, and (θ_1, θ_2) is a Gaussian vector

with specified expectation and precision. Thus we must convert our beliefs about (σ, ρ) into values for (τ_0, κ_0), (a_1, a_2), (b_1, b_2), and the expectation and precision of (θ_1, θ_2), which are modelled as log-normal.

Solving (8) in logs,

$$\begin{pmatrix} \log \kappa \\ \log \tau \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \log 8 \\ -\log(4\pi) - \log 8 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \log \sigma \\ \log \rho \end{pmatrix}. \tag{10}$$

This expression identifies the terms in (9), with $\theta_1 = \log \sigma$ and $\theta_2 = \log \rho$.

We treat σ and ρ as *a priori* independent. Our starting point are prior expectations $\mathbb{E}(\sigma) = 1.5\,\text{mm/yr}$ and $\mathbb{E}(\rho) = 1000\,\text{km}$. When implemented in `R-INLA`, the distance between any two points is represented by the great circle distance on a unit ball; hence ρ need to be scaled by the Earth radius 6371 km, and becomes $\mathbb{E}(\rho) = 1000/6371 \approx 0.16$. For our prior uncertainty, we set the prior standard deviations to be twice the prior expectations, i.e. a coefficient of variation of 2. If $\log Z \sim \mathrm{N}(m, s^2)$, then $\mathbb{E}(Z) = \exp(m + s^2/2)$ and $\mathrm{CV}(Z) = \sqrt{\exp(s^2) - 1}$. Hence

$$\mathbb{E} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \log(1.5) - \log \sqrt{5} \\ \log(2000/6371) - \log \sqrt{5} \end{pmatrix}, \quad \mathbb{V} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \log 5 & 0 \\ 0 & \log 5 \end{pmatrix}, \tag{11}$$

which together with (10) and (11) completes the specification of the prior distribution for the hyperparameters.

5.3. Results

We present the results from the globally stationary model and the two approaches for modelling non-stationarity: the subset model and the constrained parameter partition model. Figure 5 shows the marginal posterior distributions for the two hyperparameters. Clear differences between these marginal distributions indicate that our different ways of treating

non-stationarity are practically as well as theoretically different; although these differences will not necessarily translate into differences in the updated discrepancy.

[Figure 5 about here.]

Nevertheless, some clear differences are seen in the posterior expectation and standard deviation, shown in Figures 5and 7. In the expectation, the two non-stationary models have higher resolution features relative to the stationary model, e.g., in North America where there are lots of GPS stations: this might reflect the shorter correlation lengths shown in the right panel of Figure 5. As might be expected, there are larger differences in the standard deviations. Comparing the two non-stationary models, the presence of an explicit zero-region in the parameter partition model is marked by close-to-zero standard deviations inside the zero-region, and larger standard deviations in the region of interest. The subset model has an interesting 'Gibbs effect' of a ridge of slightly raised standard deviations just outside the boundary of the zero region, which is completely absent from the parameter partition model.

[Figure 6 about here.]

[Figure 7 about here.]

Looking at both the expectation in Figure 6 and the standard deviation, we are much happier presenting the results of the parameter partition model to geoscientists, than the other two models.

6. CONCLUSION

In this paper we provide an update of the global GIA using a Bayesian hierarchical model that synthesizes physical model output and imperfect GPS observations, over a spatial domain in which the discrepancy between the model and the true GIA process has a systematic component. We model the discrepancy with an isotropic Gaussian random field.

In this application, we have also addressed two important challenges. First, the challenge of large-scale computation, which is increasingly common in environmental statistics, where many interesting questions concern global behaviour, and many interesting datasets are dense and global in their coverage. We have provided a review of the key issues, and the benefits of the SPDE approach proposed by Lindgren *et al.* (2011). This approach uses a bespoke spatial triangulation, which can be adapted to the dataset and the needs of the inference.

Second, we have proposed a variety of methods for modelling non-stationarity. Non-stationarity will often be a feature in practice, especially over very large spatial domains which encompass several different types of region. In this paper we have used what we term a 'parameter partition' model. We have used this model to impose a zero-region on the update of GIA, which also involves the introduction of pseudo-observations. In forthcoming work, where it is important to separate land and ocean effects, we will use what we term a 'process partition' model.

We have also provided practical guidance, including how to construct triangulations over regions, and how to parameterize the prior distribution of the hyperparameters in the R-INLA package.

The next step for us is the use of multiple latent processes, with more complex observation operators. Some of these observations have large spatial footprints (e.g., gravitation measurements from the GRACE satellite), and, at their native resolution, non-zero measurement error covariances. This combination of multiple processes and large footprints will push our current computing resources to the limit, and likely require some further approximations.

SUPPORTING MATERIAL

Additional information and supporting material for this article is available online at the journal's website.

ACKNOWLEDGMENTS

The authors are grateful for the fnancial support provided by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 694188. We would like to thank H. Bakka and H. Rue for providing technical support on using the R-INLA package. J. L. Bamber was also supported by a Leverhulme Trust Fellowship (RF-2016-718) and a Royal Society a Royal Society Wolfson Research Merit Award.

REFERENCES

Bakka H, 2016. *Modeling Spatial Dependencies using Barriers and Different Terrains*. Ph.D. thesis, Norwegian University of Science and Technology.

Bakka H, Vanhatalo J, Illian J, Simpson D, Rue H, 2016. Accounting for physical barriers in species distribution modeling with non-stationary spatial random effects.

Banerjee S, Carlin B, Gelfand A, 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, Florida: Chapman & Hall/CRC.

Craig P, Goldstein M, Rougier J, Seheult A, 2001. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association* **96**: 717–729.

Cressie N, Wikle C, 2011. *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Inc., Hoboken NJ, USA.

Goldstein M, Rougier J, 2004. Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM Journal on Scientific Computing* **26**(2): 467–487.

Goldstein M, Rougier J, 2009. Reified Bayesian modelling and inference for physical systems. *Journal of Statistical Planning and Inference* **139**: 1221–1239, with discussion, pp. 1243–1256.

González Á, 2009. Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices. *Mathematical Geosciences* **42**(1): 49.

Guo J, Huang Z, Shum C, van der Wal W, 2012. Comparisons among contemporary glacial isostatic adjustment models. *Journal of Geodynamics* **61**(Supplement C): 129 – 137.

Kennedy M, O'Hagan A, 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B* **63**: 425–450, with discussion, pp. 450–464.

King MA, Altamimi Z, Boehm J, Bos M, Dach R, Elosegui P, Fund F, Hernández-Pajares M, Lavallee D, Mendes Cerveira PJ, Penna N, Riva REM, Steigenberger P, van Dam T, Vittuari L, Williams S, Willis P, 2010. Improved constraints on models of glacial isostatic adjustment: A review of the contribution of ground-based geodetic observations. *Surveys in Geophysics* **31**(5): 465–507.

Lindgren F, Rue H, 2015. Bayesian spatial modelling with r-inla. *Journal of Statistical Software, Articles* **63**(19): 1–25.

Lindgren F, Rue H, Lindström J, 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4): 423–498.

Luthcke SB, Sabaka T, Loomis B, Arendt A, McCarthy J, Camp J, 2013. Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *Journal of Glaciology* **59**(216): 613–631.

Martins T, Simpson D, Lindgren F, Rue H, 2013. Bayesian computing with INLA: new features. *Computational Statistics and Data Analysis* **67**: 68–83.

Peltier WR, Argus DF, Drummond R, 2015. Space geodesy constrains ice age terminal deglaciation: The global ICE-6G_C (VM5a) model. *Journal of Geophysical Research: Solid Earth* **120**(1): 450–487.

Rougier J, Goldstein M, House L, 2013. Second-order exchangeability analysis for multi-model ensembles. *Journal of the American Statistical Association* **108**: 852–863.

Rue H, Held L, 2005. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton FL, USA.

Rue H, Martino S, Chopin N, 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(2): 319–392.

Schumacher M, King MA, Rougier J, Sha Z, Khan SA, Bamber JL, 2018. A new global gps data set for testing and improving modelled gia uplift rates. *Geophysical Journal International* **214**(3): 2164–2176.

Simpson D, Lindgren F, Rue H, 2012. Think continuous: Markov Gaussian models in spatial statistics. *Spatial Statistics* **1**: 16–29.

Whittle P, 1954. On stationary processes in the plane. *Biometrika* **41**(3-4): 434–449.

Zammit-Mangion A, Rougier J, 2018. A sparse linear algebra algorithm for fast computation of prediction variances with Gaussian Markov random fields. *Computational Statistics and Data Analysis* **123**: 116–130.

Zammit-Mangion A, Rougier J, Bamber J, Schön N, 2014. Resolving the Antarctic contribution to sea-level rise: A hierarchical modelling framework. *Environmetrics* **25**: 245–264.

LIST OF FIGURES

1 The polygons where the GIA is very likely to be zero (coloured in blue). . . . 26

2 Framework of the Bayesian model-data synthesis method. 27

3 Separating triangles over the land and oceans. Here we have used a lower resolution triangulation over the land. 28

4 An example of GMRF approximation for non-stationary models. Left panel (subset model): the Euclidean distances of AB and AC are the same and the two correlations are both 0.61 in the stationary model. However in the subset model, the correlation is smaller than 0.1 for AB, and 0.77 for AC. Right panel (partition model): the shaded region in the middle has $\rho_0 = 3$ and the outside has $\rho_s = 2.5$. The four points $ABCD$ are the vertices of a square. The correlation between AB is about 0.60, between AD is 0.61, between BC is 0.56. 29

5 Posterior distribution of the hyperparameters. Part of the region below the prior densities are shown as the grey polygons in the plots. 30

6 Predicted GIA mean field. From top to bottom are the predicted GIA mean field from the constrained partition model, the subset model and the stationary model. The subset model has no prediction in the zero-region. All models performs similarly in the region of interest but show difference in the zero-region and near the boundary. 31

7 Predicted GIA uncertainty. The subset model has no prediction in the zero-region and thus no predicted uncertainties either. The partition model shows much lower uncertainty in the zero-region as expected but the uncertainties are pushed in to the region with less information and thus resulting higher uncertainties in the region of interest compared to the other models. 32

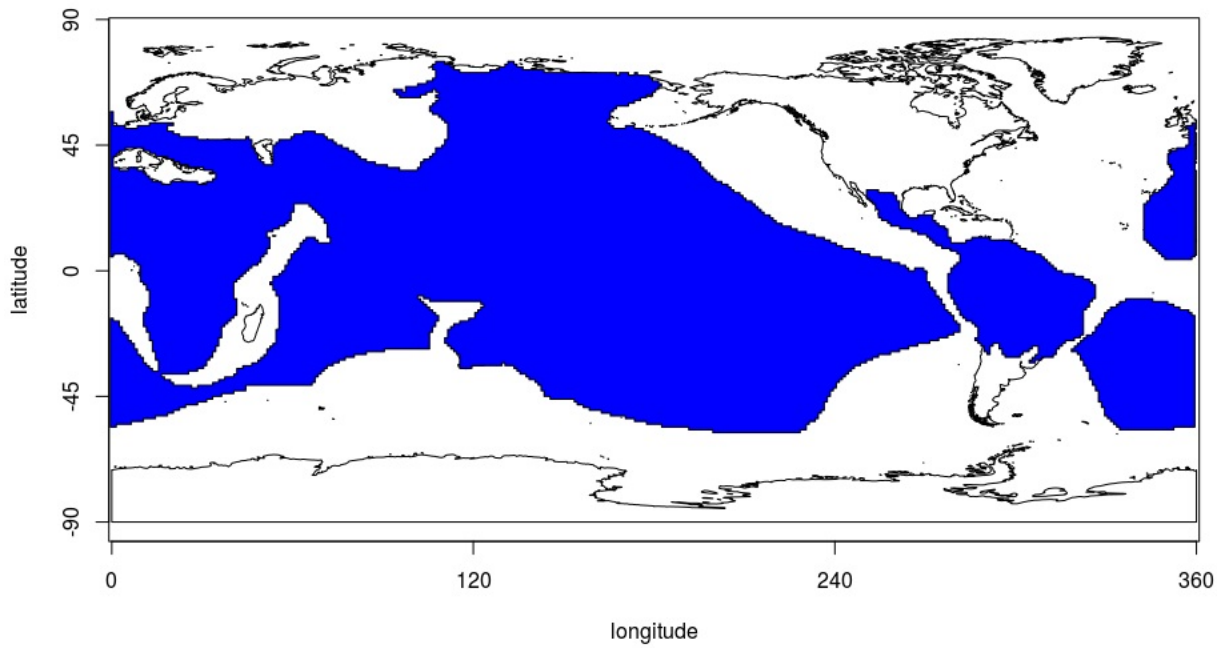

Figure 1. The polygons where the GIA is very likely to be zero (coloured in blue).

FIGURES

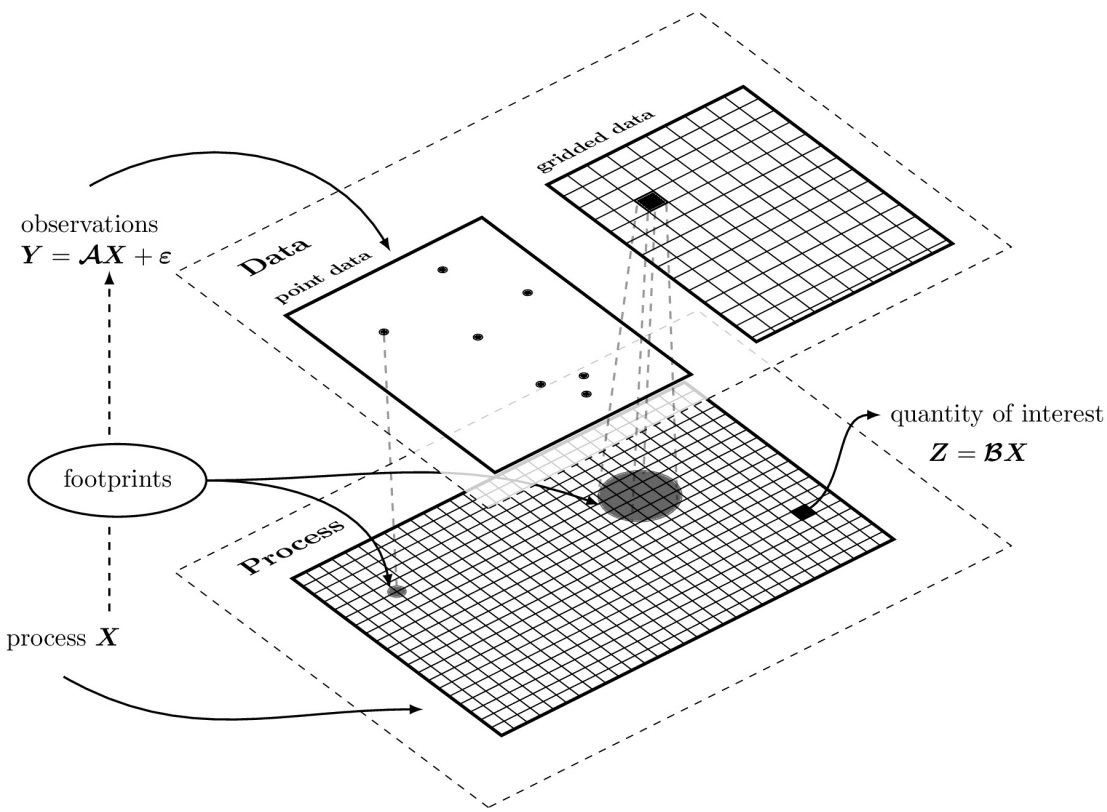

Figure 2. Framework of the Bayesian model-data synthesis method.

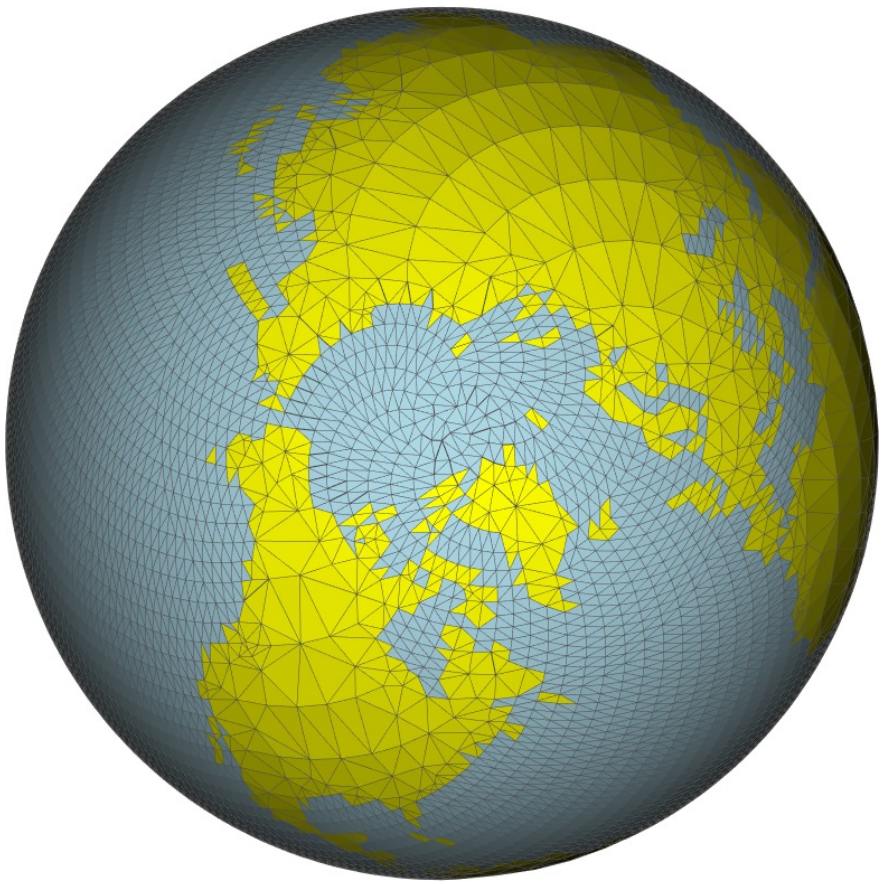

Figure 3. Separating triangles over the land and oceans. Here we have used a lower resolution triangulation over the land.

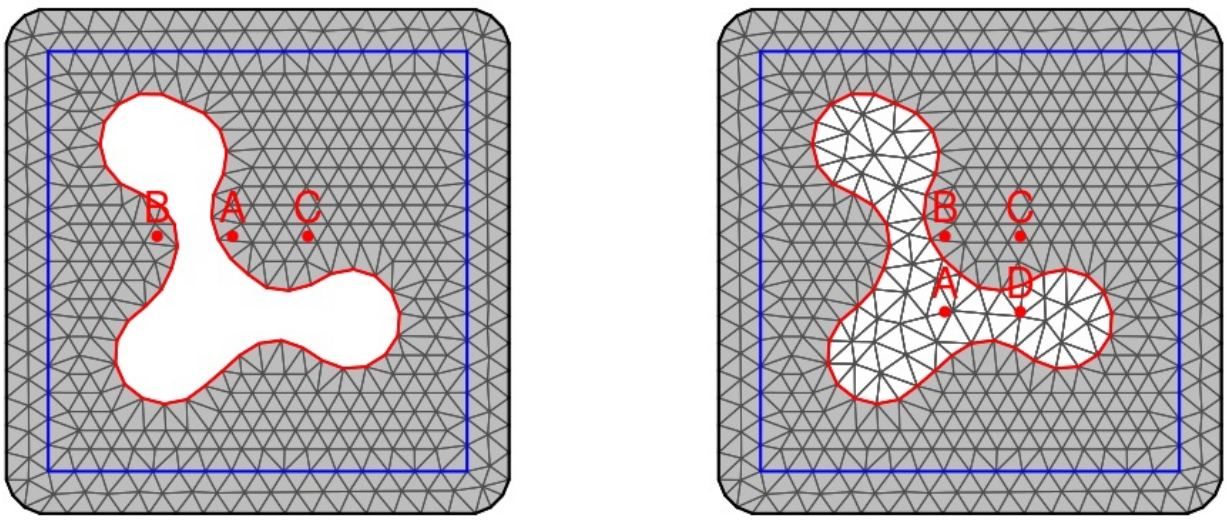

Figure 4. An example of GMRF approximation for non-stationary models. Left panel (subset model): the Euclidean distances of AB and AC are the same and the two correlations are both 0.61 in the stationary model. However in the subset model, the correlation is smaller than 0.1 for AB, and 0.77 for AC. Right panel (partition model): the shaded region in the middle has $\rho_0 = 3$ and the outside has $\rho_s = 2.5$. The four points $ABCD$ are the vertices of a square. The correlation between AB is about 0.60, between AD is 0.61, between BC is 0.56.

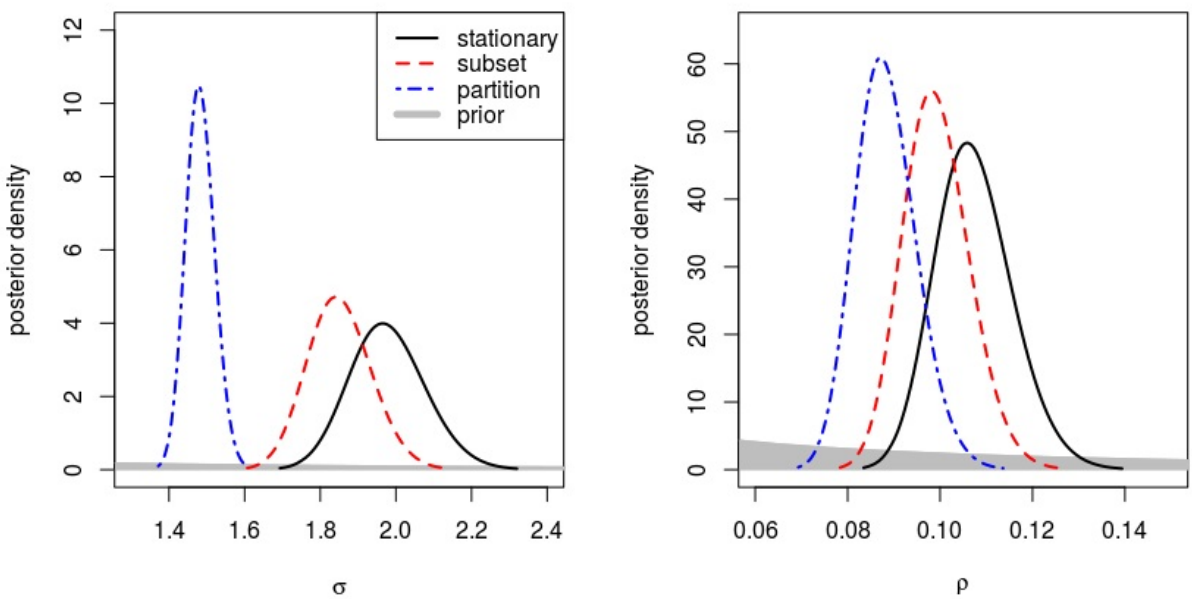

Figure 5. Posterior distribution of the hyperparameters. Part of the region below the prior densities are shown as the grey polygons in the plots.

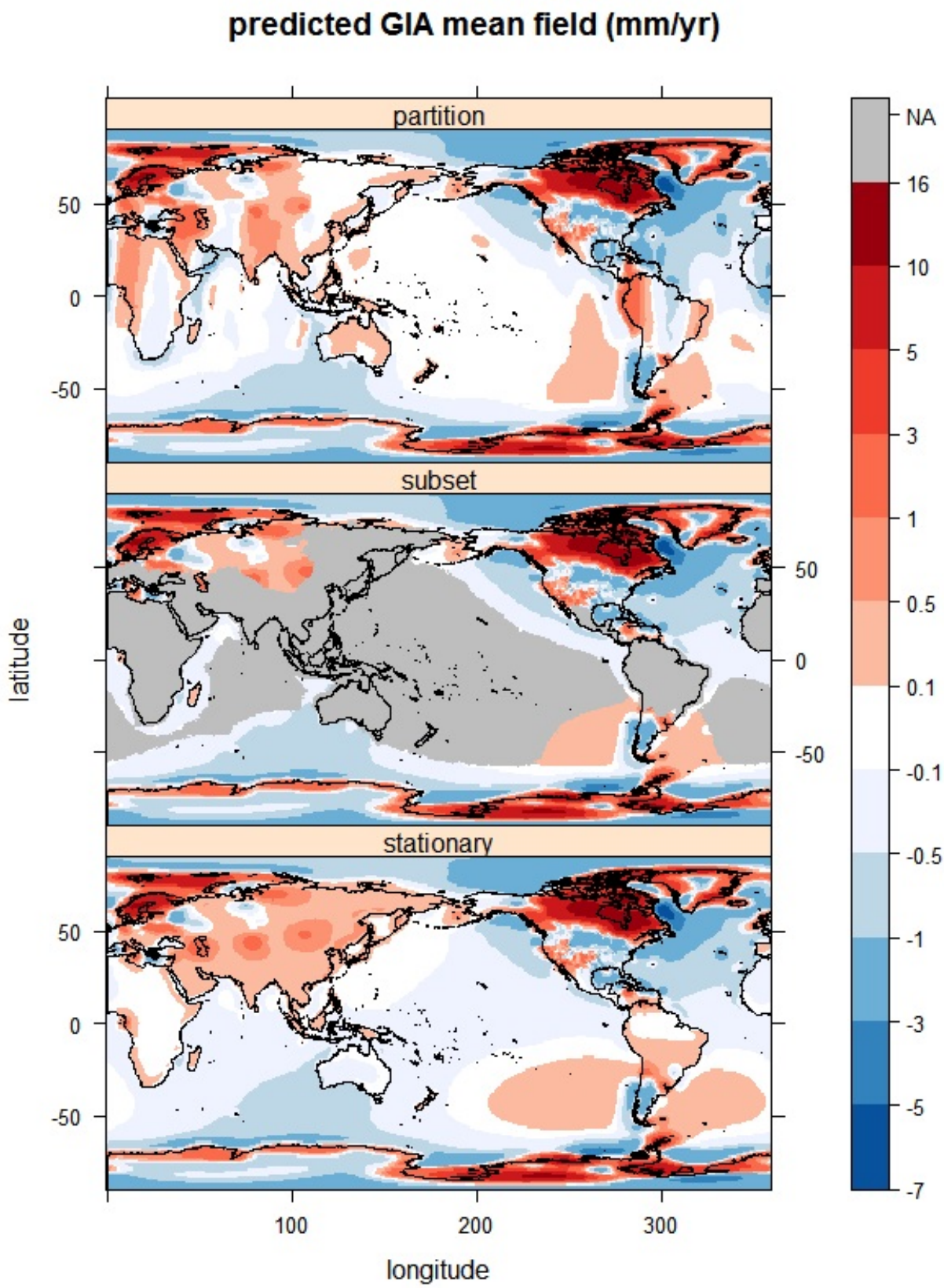

Figure 6. Predicted GIA mean field. From top to bottom are the predicted GIA mean field from the constrained partition model, the subset model and the stationary model. The subset model has no prediction in the zero-region. All models performs similarly in the region of interest but show difference in the zero-region and near the boundary.

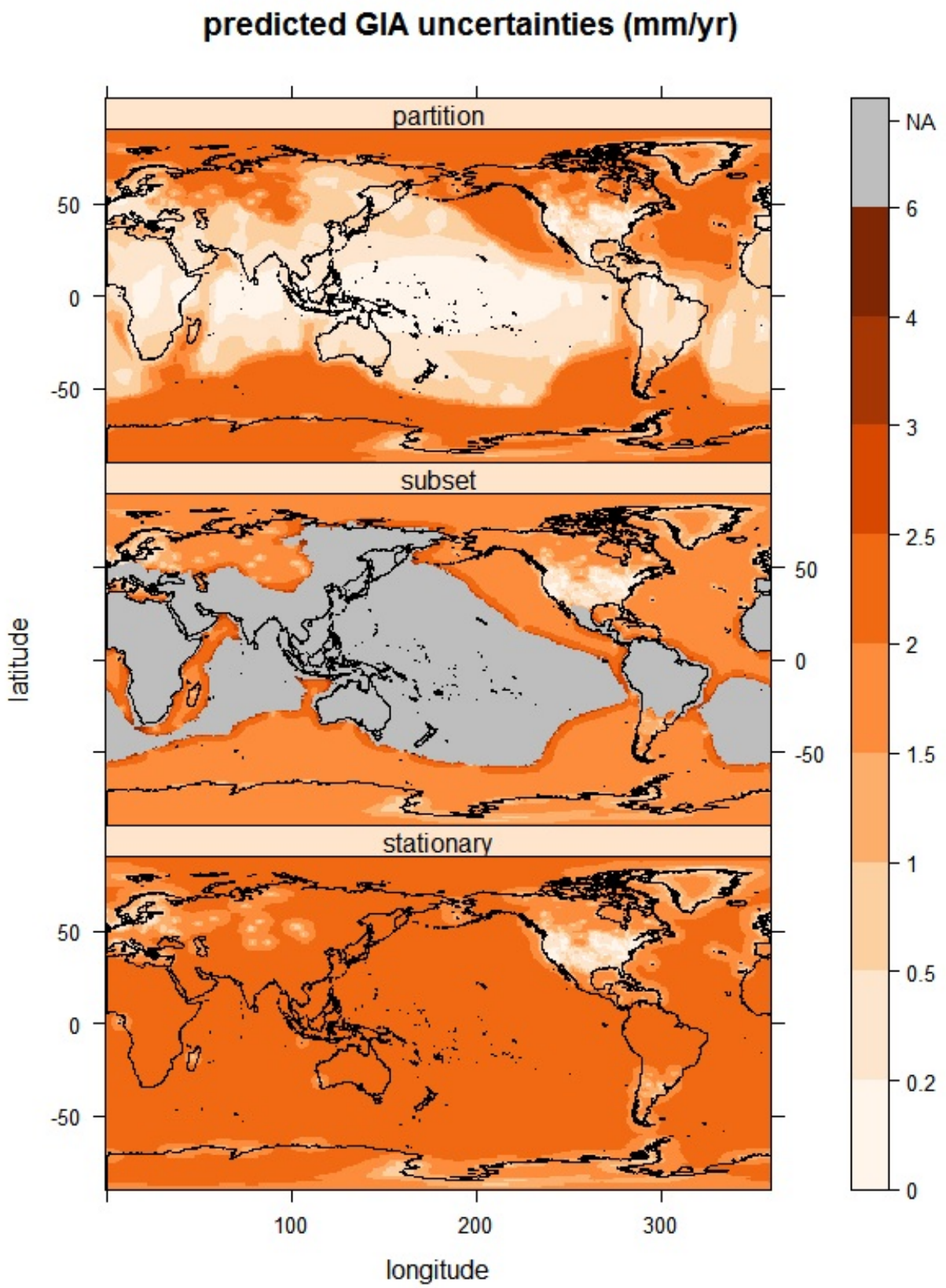

Figure 7. Predicted GIA uncertainty. The subset model has no prediction in the zero-region and thus no predicted uncertainties either. The partition model shows much lower uncertainty in the zero-region as expected but the uncertainties are pushed in to the region with less information and thus resulting higher uncertainties in the region of interest compared to the other models.