GlobalMass project glossary

The GlobalMass Team*

August 3, 2017

1 The Bayesian hierarchical model

In this section, we introduce the Bayesian hierarchical model for predicting the global GIA process.

We assume that the true GIA process is a real-valued spatial process continues on the sphere and denote it by $\boldsymbol{Y} : \mathbb{S}^2 \mapsto \mathbb{R}$. We use one of the GIA solution, say from one of the *ice6g* models, as the prior mean of the true process and denote it by $\boldsymbol{\mu} : \mathbb{S}^2 \mapsto \mathbb{R}$. Then the residuals between the true process and forward model solutions can be modelled as a stationary Gaussian process on the sphere

$$\boldsymbol{X} := \boldsymbol{Y} - \boldsymbol{\mu} \sim \mathcal{GP}(\mathbf{0}, \kappa(\boldsymbol{\theta})) \tag{1.1}$$

where $\kappa(\boldsymbol{\theta})$ defines the covariance function with hyper parameters $\boldsymbol{\theta}$.

In order to assess the bias and uncertainties in the *ice6g* solutions, we use the GPS observations to update the GIA process. The GPS data are the yearly trends of vertical movements in millimetre at the observed locations. These observations can be regarded as a linear map of the GIA process with measurement errors.

1.1 From geophysical processes to observations

With the geophysical processes defined and data processed, we need to link the processes to various observations according to their contributions. These links are usually linear translators that average the process values at the different spatial scales of the data. This is often called *change of support problem (COSP)* in the literature of spatial statistics.

In our framework, we define the term spatial *foot print* Ω_i of an observation O_i as the spatial unit where the geophysical process has a contribution to the

*Z. Sha contributes to most part of the Bayesian Hierarchical model.

observed value. Define $\boldsymbol{\mathcal{A}_i}$ to be a linear operator that maps from the process value over Ω_i to the observation value O_i over its own spatial unit δ_i. Then $\boldsymbol{\mathcal{A}_i}$ can usually be a map from point to point, area to point or area to area. Figure 1 shows the footprint of the GPS and GIA data.

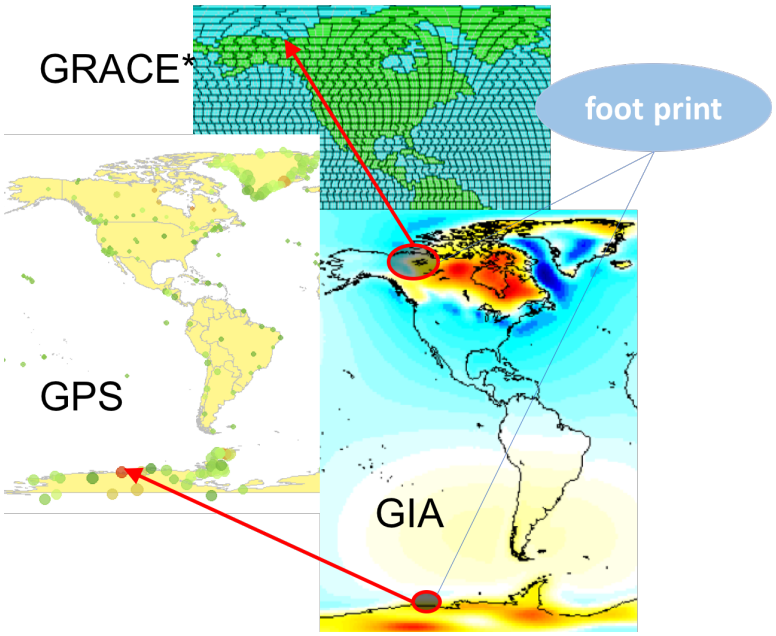

Figure 1: Examples of foot prints of GPS and GRACE data on the GIA process.

For example, the GPS signal can be affected by GIA process over a large area with decaying weights along distance to the GPS location. In this first test of framework, we simple assume the foot print of a GPS data point to be a single point of the GIA process at the same location. Thus, $\mathcal{A}_i$ is a point to point map and we have the following observation equation

$$Z_i = \boldsymbol{\mathcal{A}_i} \boldsymbol{Y} + \varepsilon_i, \ i = 1, \ldots N. \tag{1.2}$$

where Z_i measures the uplift rate of the bedrock at the i^{th} GPS station and the measurement errors ε_is are assumed to be independent Gaussian errors $\mathcal{N}(0, e_i^2)$. In practice, Z_i is estimated from the raw GPS time series signals and e_i is fixed to be the standard error Z_i.

Then by stacking $\boldsymbol{\mathcal{A}_i}$ together as

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} \boldsymbol{\mathcal{A}}_1 \\ \vdots \\ \boldsymbol{\mathcal{A}}_N \end{bmatrix},$$

we can write equation (1.2) into vector form

$$\boldsymbol{Z} = \boldsymbol{\mathcal{A}}\boldsymbol{Y} + \boldsymbol{\varepsilon} \tag{1.3}$$

Now combining equations (1.1) and (1.3), the system simplifies to

$$\tilde{\boldsymbol{Z}} = \boldsymbol{Z} - \boldsymbol{\mathcal{A}}\boldsymbol{\mu} = \boldsymbol{\mathcal{A}}\boldsymbol{X} + \boldsymbol{\varepsilon} \tag{1.4}$$

Hence, our final model is

$$\begin{cases} \tilde{\boldsymbol{Z}} = \boldsymbol{\mathcal{A}}\boldsymbol{X} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(e_1^2, e_2^2, \ldots, e_N^2)) \\ \boldsymbol{X} \sim \mathcal{GP}(\mathbf{0}, \kappa(\boldsymbol{\theta})) \\ \boldsymbol{\theta} \sim \boldsymbol{p}(\boldsymbol{\theta}) \end{cases} \tag{1.5}$$

where $\boldsymbol{\pi}(\boldsymbol{\theta})$ is the prior distribution for the hyper parameters.

We call the above approach of linking the process to the observation *forward modelling*. To be differentiated from forward model, the forward modelling approach describes the functional relationship between the process and the data rather than the underlying physical mechanism of the latent process.

1.2 GMRF Approximation

Suppose we would like to predict the GIA process on a set of grid points $\boldsymbol{S} = \{s_i : i = 1, \ldots, m\}$ with a given resolution. The Gaussian process model can be computationally expensive for large scale inference since the Bayesian update scales as $\mathcal{O}(m^3)$ mainly due to the inverse of a dense covariance matrix.

At the same time, Gaussian Markov random field is often used for modelling discrete spatial unit. The covariance structure is defined through its inverse, the precision matrix, which is usually sparse and thus has nice computational properties.

The Gaussian process with Matérn covariance function can be treated as solutions to a class of stochastic partial differential equations (Lindgren et al., 2011) which can then be approximated by GMRF using finite element methods.

Denote by $\tilde{\boldsymbol{X}}$ the GMRF approximation of $\boldsymbol{X}$ on a given triangulation of the sphere with piecewise linear local basis functions $\{\boldsymbol{\phi}_i\}_{i \in \mathbb{N}}$, then given any location $s \in \mathbb{S}^2$

$$\boldsymbol{X}(s) \approx \boldsymbol{\phi}_i(s)^T \tilde{\boldsymbol{X}} \tag{1.6}$$

and for a given set $\boldsymbol{S}$ of locations, we have

$$\boldsymbol{X}(\boldsymbol{S}) \approx \boldsymbol{C}(\boldsymbol{S})\tilde{\boldsymbol{X}} \tag{1.7}$$

where the matrix $\boldsymbol{C}$ contains basis functions for all locations.

Now with the GMRF approximation, our model becomes

$$\begin{cases} \tilde{\boldsymbol{Z}} = \boldsymbol{\mathcal{A}}\boldsymbol{C}\tilde{\boldsymbol{X}} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \mathrm{diag}(e_1^2, e_2^2, \ldots, e_N^2)) \\ \tilde{\boldsymbol{X}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta})) \\ \boldsymbol{\theta} \sim \boldsymbol{p}(\boldsymbol{\theta}) \end{cases} \tag{1.8}$$

where $\boldsymbol{Q}$ is the precision matrix of the GMRF approximation.

The GMRF approximation requires generating a mesh which is often a triangulation of a given area of interest. Then $\tilde{\boldsymbol{X}}$ are located at the vertices of the mesh. For a mesh that is approximately regular, we define the resolution of the mesh to be the average area of the triangles converted to degree by degree; and for irregular mesh, the resolution is calculated by using the smallest triangle.

1.3 Bayesian Inference and prediction

The Bayesian inference requires finding the posterior distributions of the hyper parameters and the latent process. The Markov chain Monte Carlo (MCMC) can be used in general to sample from the posteriors and the integrated nested Laplace approximation (INLA) is a fast approximation method. In the following, we use $\boldsymbol{\pi}$ for a general pdf and define a few terms used in our Bayesian inference.

The Bayesian inference draw conclusions from the posterior distribution

$$\boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}|\tilde{\boldsymbol{Z}}) = \frac{\boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}, \tilde{\boldsymbol{Z}})}{\boldsymbol{\pi}(\tilde{\boldsymbol{Z}})} = \frac{\boldsymbol{\pi}(\tilde{\boldsymbol{Z}}|\boldsymbol{X}, \boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{X}|\boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{\theta})}{\int_{\boldsymbol{\mathcal{X}},\boldsymbol{\Theta}} \boldsymbol{\pi}(\tilde{\boldsymbol{Z}}|\boldsymbol{X}, \boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{X}|\boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{X}\,\mathrm{d}\boldsymbol{\theta}} \tag{1.9}$$

This is the joint distribution of the latent process and the hyper parameters but in practice we are more interested in the posterior marginals for inference on parameters and prediction of the latent filed separately. For simplicity, we use posterior distribution for

$$\boldsymbol{\pi}(\boldsymbol{\theta}|\tilde{\boldsymbol{Z}}) = \int_{\boldsymbol{\mathcal{X}}} \boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}|\tilde{\boldsymbol{Z}})\,\mathrm{d}\boldsymbol{X} \tag{1.10}$$

the posterior marginal distribution for the hyper parameters. Inference on the parameters is not of crucial importance here but provides as sanity checks for the process property such as the length of spatial correlation. The aim of the

study is to predict the latent process and the corresponding uncertainty on a fine resolution map; hence define the joint predictive distribution of the latent process to be

$$\boldsymbol{\pi}(\boldsymbol{X}|\tilde{\boldsymbol{Z}}) = \int_{\boldsymbol{\Theta}} \boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}|\tilde{\boldsymbol{Z}})\, \mathrm{d}\boldsymbol{\theta} \tag{1.11}$$

the posterior marginal distribution of the latent process that integrate out the uncertainty of the hyper parameters. In practice, we are more interested in predicting the marginal means and variances at a given set of locations; hence we define the point-wise predictive distribution to be

$$\boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}}) = \int_{\boldsymbol{\mathcal{X}_{-i}}} \boldsymbol{\pi}(\boldsymbol{X}|\tilde{\boldsymbol{Z}})\, \mathrm{d}\boldsymbol{X_{-i}} \tag{1.12}$$

where X_i is the latent process at location i and $\boldsymbol{X_{-i}}$ are $\boldsymbol{X}$ elsewhere. The predicted mean X_i^* and predicted uncertainty $\mathrm{u}(X_i)$ are the expectation and standard deviation with respect to $\boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}})$

$$X_i^* = \mathbb{E}(X_i|\tilde{\boldsymbol{Z}}) = \int_{\mathbb{R}} X_i \boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}})\, \mathrm{d}X_i \tag{1.13}$$

$$\mathrm{u}(X_i) = \sqrt{\mathrm{Var}(X_i|\tilde{Z})} = \sqrt{\int_{\mathbb{R}} (X_i - X_i^*)^2 \boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}})\, \mathrm{d}X_i} \tag{1.14}$$

The INLA method directly approximate the $\boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}})$ and provide the predicted mean and uncertainty as summary statistics. For the MCMC approach, the predictive distribution can be approximate by the posterior sample distribution of X_i and the predicted mean and predicted uncertainty by the sample mean and standard error.

INLA is much faster and more efficient than MCMC in providing the predictive distribution but it provides limited information on the joint posterior distributions. The MCMC samples can be used in various ways for exploring the posteriors.

Acronyms

COSP change of support problem. 1

GIA glacial isostatic adjustment. 1, 3

GMRF Gaussian Markov random field. 3, 4

INLA integrated nested Laplace approximation. 4, 5

MCMC Markov chain Monte Carlo. 4, 5

pdf probability density function. 4

SPDE stochastic partial differential equation. 3

Glossary

foot print The spatial unit where the geophysical process makes a contribution to the observed data.. 1

forward model physical model, usually a (partial) differential equation system, used to solve certain geophysical process. 1, 3

forward modelling The approach that link the process contribution to the corresponding observation.. 3

hyper parameters parameters in defining the covariance structure of the latent process; in particular for a Gaussian process with Matérn covariance function, the hyper parameters are the lengthscale ρ and norminal variance σ^2. 1, 3

mesh A triangulation of a given area.. 4

predicted mean the mean with respect to the predictive distribution. 5

predicted uncertainty the standard deviation with respect to the predictive distribution. 5

predictive distribution the posterior marginal distribution of a given location X_i that integrate out the unknown parameters and $\boldsymbol{X}_{-i}$. 5

prior distribution the probability distribution for the (hyper) parameters or latent processes. 3

References

F. Lindgren, H. Rue, and J. Lindström. 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, 2011.