The Bayesian Hierarchical model and glossary

Zhe Sha

July 14, 2017

1 Forward modelling and spatial support

2 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

$$\boldsymbol{X} := \boldsymbol{Y} - \boldsymbol{\mu} \sim \mathcal{GP}(\mathbf{0}, \kappa(\boldsymbol{\theta})) \tag{2.1}$$

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

$$Z_i = \boldsymbol{\mathcal{A}}_i \boldsymbol{Y} + \varepsilon_i, \ i = 1, \dots N. \tag{2.2}$$

where $\boldsymbol{\mathcal{A}}_i$ is the linear operator that maps the GIA process to the i^{th} GPS observation and ε_i are assumed to be independent Gaussian errors $\mathcal{N}(0, e_i^2)$. In practice, e_i^2 can usually be estimated from raw GPS data and therefore we set them to be fixed values from prior information.

Denote the linear operator for the GPS observation vector $\boldsymbol{Z}$ by

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

Then we can write equation 2.2 into the vector form

$$\boldsymbol{Z} = \boldsymbol{\mathcal{A}}\boldsymbol{Y} + \boldsymbol{\varepsilon} \tag{2.3}$$

Now combining equations 2.1 and 2.3, the system simplifies to

$$\tilde{\boldsymbol{Z}} = \boldsymbol{Z} - \boldsymbol{\mathcal{A}}\boldsymbol{\mu} = \boldsymbol{\mathcal{A}}\boldsymbol{X} + \boldsymbol{\varepsilon} \tag{2.4}$$

Hence, our final model is

$$\begin{cases} \tilde{\boldsymbol{Z}} = \boldsymbol{\mathcal{A}}\boldsymbol{X} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \mathrm{diag}(e_1^2, e_2^2, \ldots, e_N^2)) \\ \boldsymbol{X} \sim \mathcal{GP}(\boldsymbol{0}, \kappa(\boldsymbol{\theta})) \\ \boldsymbol{\theta} \sim \boldsymbol{p}(\boldsymbol{\theta}) \end{cases} \tag{2.5}$$

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

3 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 (**?**) 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 basis functions $\{\boldsymbol{\phi}_i\}_{i \in \mathbb{N}}$, then given any location

$s \in \mathbb{S}^2$

$$\boldsymbol{X}(s) \approx \boldsymbol{\phi}_i(s)^T \tilde{\boldsymbol{X}} \tag{3.1}$$

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

$$\boldsymbol{X}(\boldsymbol{S}) \approx \boldsymbol{C}(\boldsymbol{S})\tilde{\boldsymbol{X}} \tag{3.2}$$

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

Now with the GMRF approximation, our model becomes

$$\begin{cases} \tilde{\boldsymbol{Z}} = \boldsymbol{\mathcal{A}}\boldsymbol{C}\tilde{\boldsymbol{X}} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(e_1^2, e_2^2, \ldots, e_N^2)) \\ \tilde{\boldsymbol{X}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta})) \\ \boldsymbol{\theta} \sim \boldsymbol{p}(\boldsymbol{\theta}) \end{cases} \tag{3.3}$$

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

4 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

$$\boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}|\tilde{\boldsymbol{Z}}) = \frac{\boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}, \tilde{\boldsymbol{Z}})}{\boldsymbol{\pi}(\tilde{\boldsymbol{Z}})} = \frac{\boldsymbol{\pi}(\tilde{\boldsymbol{Z}}|\boldsymbol{X}, \boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{X}|\boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{\theta})}{\int_{\boldsymbol{\mathcal{X}},\boldsymbol{\Theta}} \boldsymbol{\pi}(\tilde{\boldsymbol{Z}}|\boldsymbol{X}, \boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{X}|\boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{X}\,\mathrm{d}\boldsymbol{\theta}} \tag{4.1}$$

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

$$\boldsymbol{\pi}(\boldsymbol{\theta}|\tilde{\boldsymbol{Z}}) = \int_{\boldsymbol{\mathcal{X}}} \boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}|\tilde{\boldsymbol{Z}})\,\mathrm{d}\boldsymbol{X} \tag{4.2}$$

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

$$\boldsymbol{\pi}(\boldsymbol{X}|\tilde{\boldsymbol{Z}}) = \int_{\boldsymbol{\Theta}} \boldsymbol{\pi}(\boldsymbol{X}, \boldsymbol{\theta}|\tilde{\boldsymbol{Z}})\, \mathrm{d}\boldsymbol{\theta} \tag{4.3}$$

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

$$\boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}}) = \int_{\boldsymbol{\mathcal{X}_{-i}}} \boldsymbol{\pi}(\boldsymbol{X}|\tilde{\boldsymbol{Z}})\, \mathrm{d}\boldsymbol{X_{-i}} \tag{4.4}$$

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 $\boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}})$

$$X_i^* = \mathbb{E}(X_i|\tilde{\boldsymbol{Z}}) = \int_{\mathbb{R}} X_i \boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}})\, \mathrm{d}X_i \tag{4.5}$$

$$\mathrm{u}(X_i) = \sqrt{\mathrm{Var}(X_i|\tilde{Z})} = \sqrt{\int_{\mathbb{R}} (X_i - X_i^*)^2 \boldsymbol{\pi}(X_i|\tilde{\boldsymbol{Z}})\, \mathrm{d}X_i} \tag{4.6}$$

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

GIA glacial isostatic adjustment. 1, 2

GMRF Gaussian Markov random field. 2, 3

INLA integrated nested Laplace approximation. 3, 4

MCMC Markov chain Monte Carlo. 3, 4

pdf probability density function. 3

SPDE stochastic partial differential equation. 2

Glossary

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

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, 2

predicted mean the mean with respect to the predictive distribution. 4

predicted uncertainty the standard deviation with respect to the predictive distribution. 4

predictive distribution the posterior marginal distribution of a given location X_i that integrate out the unknown parameters and $\boldsymbol{X}_{-i}$. 4

prior distribution the probability distribution for the (hyper) parameters or latent processes. 2

prior information information for setting up the model, including the values for the parameters, constraints, etc.. 1