Efficient Fitting of Bayesian Regression Models with Spatio-temporally Varying Coefficients

Mark Bass[1], Sujit Sahu[1]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 3

[1]Univeristy of Southampton, Southampton Statistical Sciences Research Institute (S3RI), Southampton, UK
mrb2g11@soton.ac.uk
sks@soton.ac.uk

Abstract

Bayesian regression models with spatio-temporally varying coefficients are gaining popularity among researchers who are looking to model the spatio-temporal processes that are ubiquitous in the environmental and physical sciences. The fitting of these highly overparameterised and non-stationary models is challenging and computationally expensive.

Typically, the coefficients are given a Gaussian process prior, and, with the assumption of a normal likelihood, can be marginalised over to reduce the dimension of the parameter space. By considering different parameterisations of the model we obtain numerous marginalisation schemes, each of which define a new fitting method.

We use the MCMC output to compare the fitting methods in terms of convergence rates and effective sample sizes per second, and thus identify the most efficient fitting strategy for models of this type.

Implementation of the optimal strategy achieves faster convergence rates and significant savings in computation time, illustrated here with a simulation example and also a real data example modelling daily ozone concentration data.

Keywords: Bayesian estimation; spatio-temporal modelling; parameterisation; marginalisation.

1 Introduction

Given that large data sets are now prevalent in many areas of statistics, it is important to efficiently implement any Markov Chain Monte Carlo (MCMC)

algorithm. It has long been understood that the parameterisation of a hierarchical model affects the performance of the MCMC method used for inference. In particular, high posterior correlations between model parameters can lead to poor mixing and slow convergence. Papaspiliopoulos, Roberts and Sköld [1] develop a framework for the parameterisation of hierarchical models, applied to a wide range of statistical contexts. They focus on two parameterisations, namely *centred* and *non-centred*, terms introduced by Gelfand, Sahu and Carlin [2].

In this paper we aim to develop the most efficient way to fit a non-linear, stochastic process-based model to spatio-temporal data. Motivated by, but not restricted to, modelling ground-level ozone concentrations, we look at a class of models sometimes referred to as *downscaler* models. A downscaler model regresses the observed point referenced data upon a covariate, or set of covariates, given as averages over gridded cells, see [3] for details. In our application the covariate is the output from a numerical model.

To respect the non-linearity of the surface that we are attempting to model, we allow the intercept and slope to be site specific. This is done by introducing locally varying adjustments to the coefficients that are realisations of Gaussian processes. Consequently we have a different model at every point in the spatial domain. The inclusion of spatially varying coefficients introduces non-stationarity into the model. By extending the model to allow for data collected over time we further increase the computational burden. These highly overparameterised models are very challenging to fit as is using them to produce forecast maps, which is the target for spatio-temporal modelling of air pollution.

2 Model Specifications

Suppose that the data has been transformed such that we can assume normality of the errors. Let $Z(\boldsymbol{s},t)$ denote the suitably transformed observed ozone concentration at site $\boldsymbol{s}$ and time t, $(t = 1,\ldots,T)$, and let $x(\boldsymbol{s},t)$ be the similarly transformed numerical model output for the grid cell containing site $\boldsymbol{s}$. Consider the following spatio-temporal model in it's non-centred form:

$$Z(\boldsymbol{s}_i,t) = \beta_0 + w(\boldsymbol{s}_i,t) + \{\beta_1 + \beta_1(\boldsymbol{s}_i,t)\}x(\boldsymbol{s}_i,t) + \epsilon(\boldsymbol{s}_i,t), \tag{1}$$

for i=1,...,n, t=1,...,T, where $\epsilon(\boldsymbol{s}_i,t) \overset{ind}{\sim} N(0,\sigma^2_\epsilon)$.

Under this parameterisation we interpret β_0 as the fixed intercept and β_1 as the fixed regression coefficient. These are locally perturbed by $w(\boldsymbol{s}_i,t)$ and $\beta_1(\boldsymbol{s}_i,t)$ respectively, which are both modelled as zero mean Gaussian processes with separable covariance structure, given by:

$$Cov\{w(\boldsymbol{s}_i,t_k), w(\boldsymbol{s}_j,t_l)\} = \sigma^2_w \rho_s(|\boldsymbol{s}_i - \boldsymbol{s}_j|;\phi_{w(s)})\rho_t(|t_k - t_l|;\phi_{w(t)}),$$

and

$$Cov\{\beta_1(\boldsymbol{s}_i,t_k), \beta_1(\boldsymbol{s}_j,t_l)\} = \sigma^2_\beta \rho_s(|\boldsymbol{s}_i - \boldsymbol{s}_j|;\phi_{\beta(s)})\rho_t(|t_k - t_l|;\phi_{\beta(t)}),$$

where ρ_s and ρ_t are valid isotropic covariance functions from the Matérn family. The vectors $\boldsymbol{w}$ and $\boldsymbol{\beta}$ contain the nT realisations of $w(\boldsymbol{s},t)$ and $\beta_1(\boldsymbol{s},t)$ respectively such that:

$$\boldsymbol{w} = (w(\boldsymbol{s}_1,1),\ldots,w(\boldsymbol{s}_n,T))' \sim N(0,\boldsymbol{\Sigma}_w),$$

$$\boldsymbol{\beta} = (\beta_1(\boldsymbol{s}_1,1),\ldots,\beta_1(\boldsymbol{s}_n,T))' \sim N(0,\boldsymbol{\Sigma}_\beta).$$

The ϕ parameters control the rates of decay of the spatial and temporal correlation between the random effects. Each of these are given a uniform prior. Variance parameters σ^2_ϵ, σ^2_w and σ^2_β are modelled on their inverse scales and are given gamma priors. Mean parameters β_0 and β_1 are given vague normal priors.

We denote the vector of length nT containing the numerical model output for the grid cells containing sites $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n$, at times $t = 1,\ldots,T$ by $\boldsymbol{x} = (x(\boldsymbol{s}_1,1),\ldots,x(\boldsymbol{s}_n,T))'$ and let $\boldsymbol{X} = diag(\boldsymbol{x})$. Then we can write the likelihood for model (1) as:

$$\boldsymbol{Z} \sim N(\beta_0\mathbf{1} + \boldsymbol{w} + \beta_1\boldsymbol{x} + \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{2}$$

where $\boldsymbol{\Sigma} = \sigma^2_\epsilon \boldsymbol{I}$, with $\boldsymbol{I}$ representing the $nT \times nT$ identity matrix. Given that $\boldsymbol{w}$ and $\boldsymbol{\beta}$ are given Gaussian priors, we can integrate them out of the full likelihood (2). Marginalising over $\boldsymbol{w}$ gives:

$$\boldsymbol{Z} \sim N(\beta_0\mathbf{1} + \beta_1\boldsymbol{x} + \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_w), \tag{3}$$

and achieves a reduction in the dimension of the parameter space of nT. Likewise, marginalising over $\boldsymbol{\beta}$ gives:

$$\boldsymbol{Z} \sim N(\beta_0\mathbf{1} + \boldsymbol{w} + \beta_1\boldsymbol{x}, \boldsymbol{\Sigma} + \boldsymbol{X}\boldsymbol{\Sigma}_\beta\boldsymbol{X}'). \tag{4}$$

Marginalising over both $\boldsymbol{w}$ and $\boldsymbol{\beta}$ reduces the the dimension of the parameter space by $2nT$, and gives:

$$\boldsymbol{Z} \sim N(\beta_0\mathbf{1} + \beta_1\boldsymbol{x}, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_w + \boldsymbol{X}\boldsymbol{\Sigma}_\beta\boldsymbol{X}'). \tag{5}$$

Re-writing model (1) in it's centred form we get

$$Z(\boldsymbol{s}_i,t) = w(\boldsymbol{s}_i,t) + \beta_1(\boldsymbol{s}_i,t)x(\boldsymbol{s}_i,t) + \epsilon(\boldsymbol{s}_i,t), \quad i = 1,\ldots,n, \quad t = 1,\ldots,T, \tag{6}$$

but now $w(\boldsymbol{s}_i,t)$ has mean β_0 and $\beta(\boldsymbol{s}_i,t)$ has mean β_1. For the centred paramterisation we get a full likelihood of:

$$\boldsymbol{Z} \sim N(\boldsymbol{w} + \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{7}$$

Marginalising over $\boldsymbol{w}$ gives:

$$\boldsymbol{Z} \sim N(\beta_0\mathbf{1} + \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_w), \tag{8}$$

and over $\boldsymbol{\beta}$ gives:

$$\boldsymbol{Z} \sim N(\boldsymbol{w} + \beta_1 \boldsymbol{x}, \boldsymbol{\Sigma} + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}'). \tag{9}$$

Finally, marginalising over both $\boldsymbol{w}$ and $\boldsymbol{\beta}$ gives:

$$\boldsymbol{Z} \sim N(\beta_0 + \beta_1 \boldsymbol{x}, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_w + \boldsymbol{X}\boldsymbol{\Sigma}_\beta \boldsymbol{X}'), \tag{10}$$

which is the same as likelihood (5).

Likelihoods (2)-(5) and (7)-(9) label seven ways of fitting the same model. Each parameterisation-marginalisation pair yields different full conditional distributions for the model parameters. In this paper we discuss the implications for convergence and effective sample size per second of the fitting method employed.

References

[1] O. Papaspiliopoulos, G. Roberts, M. Sköld. **A general framework for the parameterization of hierarchical models**. *Statistical Science*; 2007; 22(1); pp. 59-73.

[2] A. Gelfand, S. Sahu, B. Carlin. **Efficient parameterisations for normal linear mixed models**. *Biometrika*; 1995; 82(3); pp. 479-488.

[3] V. Berrocal, A. Gelfand, D. Holland. **A spatio-temporal downscaler for output from numerical models**. *Journal of Agriculture, Biological and Environmental Statistics*; 2010; 15(2); pp. 176-197.