Stats 295
Final Project

David Mwakima
Student ID: 11083916

12/05/2022

1   Background

A typical approach to modeling a point-referenced spatial process $Y(\mathbf{s})$ for $\mathbf{s} \in \mathcal{D} \subset \mathbb{R}^2$ is using a Bayesian hierarchical model consisting of:

- **Data Model**

$$Y(s_i)\,|\,\tilde{\beta}, \eta', \tau^2, \tilde{\theta} \overset{ind}{\sim} N(\mu(s_i), \tau^2 \mathbf{I}_n)$$

    for $i = 1, 2, \ldots, n$. Here $\mu(s_i) = \mathbf{X}(s_i)\tilde{\beta} + \eta'(s_i)$ models the mean structure of $Y(\mathbf{s})$. $\mathbf{X}(s_i)$ is an $n \times p$ design matrix with covariates at the observation locations and $\tilde{\beta}$ is a $p \times 1$ vector of regression coefficients.

- **Spatial Process Model**

$$\eta'(s_i)|\tilde{\theta} \sim N(\mathbf{0}, \Sigma(\tilde{\theta}))$$

    The $\eta'(s_i)$ are spatial random effects, interpreted as capturing the effect of unmeasured or unobserved covariates with spatial correlation. The spatial process model is a mean-zero Gaussian process with an $n \times n$ covariance matrix $\Sigma(\tilde{\theta})$ induced by a stationary and isotropic covariance function $C(s, s'; \tilde{\theta})$ that depends on a set of parameters $\tilde{\theta}$.

- **Parameter Model**

    This consists of a prior distribution for $\tilde{\beta}$, a prior distribution for each of the components of the vector of parameters $\tilde{\theta}$ and a prior distribution for τ^2.

With this hierarchical model, inference proceeds using Markov chain Monte Carlo (MCMC) methods by sampling from the posterior distribution, which is proportional to the joint distribution of the product of these models. Computationally, the product of the data model and the spatial process model is a mixture of two Gaussian distributions. Specifically, this product requires one to compute the inverse of the square root of the determinant of $\Sigma(\tilde{\theta})$ and the inverse of $\Sigma(\tilde{\theta})$. But because there is spatial correlation in the $\eta'(\mathbf{s})$ process, the matrix $\Sigma(\tilde{\theta})$ is not necessarily diagonal. This means that computing both the inverse and the determinant can be both computationally expensive and time intensive, especially for large matrices. This is the issue of dimensionality or the "big n" problem that forms the background of my paper.[1] It is not just a problem for Bayesian hierarchical modeling; the same problem needs to be faced in Frequentist statistics since the likelihood also involves computing both the determinant and the inverse of the covariance matrix $\Sigma(\tilde{\theta}) + \tau^2 \mathbf{I}_n$.

Although several proposals have been offered in the literature (See (Higdon 2002), (Johannesson and Cressie 2004) and (Guan and Haran 2018) for an overview) to address this problem; in this paper, I will be interested specifically in *predictive process models* due to (Banerjee et al. 2008). With these models, the "big n" problem is addressed by projecting the higher dimensional $\eta'(\mathbf{s})$ process onto a much lower dimensional process $\tilde{\eta}'(\mathbf{s})$ at a specified set of locations ("knots"). The basic idea is to predict (in the sense of "kriging") what the spatial process is at these knots based on the original process and then use this lower dimensional predictive process for parameter estimation and inference about the original process. Since the knots can be set to have dimensions orders of magnitude smaller than the number of dimensions of the parent process $\eta'(\mathbf{s})$, the computational task can be simplified significantly.

However, one problem with this approach is the following. The predictive process $\tilde{\eta}'(\mathbf{s})$ is a conditional expectation, given a realization of $\eta'(\mathbf{s})$ at the knots. But the mean is always smoother than the original

---

[1] See (Higdon 2002), (Johannesson and Cressie 2004), (Banerjee et al. 2008), (Cressie and Johannesson 2008), (Katzfuss 2017) and (Guan and Haran 2018) for further background

process. So, by using knots on coarser-grained regions we lose information on what happens on smaller distances in the original process. This is a problem if in the original process the correlation, in fact, decays faster on much smaller/finer distances. Consequently, one might worry that we are not *truly* capturing the spatial correlation of the random effects $\eta'(s_i)$.

In this paper, I investigate whether this is a genuine problem for this approach. One might think that this problem depends on how we set up the knots. So my question is: does the estimation of the covariance function parameters $\tilde{\theta}$ using the predictive process model depend on how the knot locations are chosen or specified?

2 Simulations

To address this question, I simulated some data on 900 locations on $\mathcal{D} = [0, 10] \times [0, 10] \subset \mathbb{R}^2$ regular grid with known values of $\tilde{\theta}$.
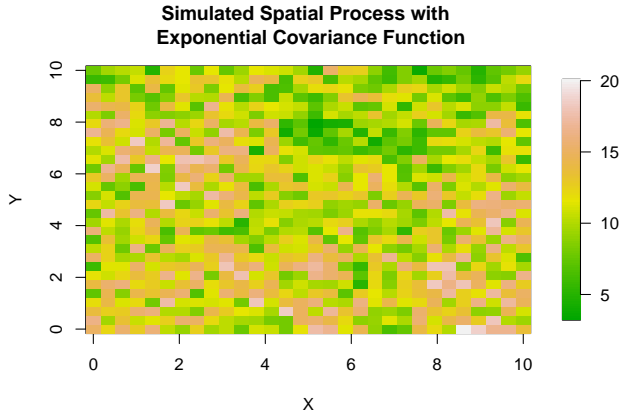
- **Data Model**

$$Y(s_i) \,|\, \tilde{\beta}, \eta', \tau^2, \tilde{\theta} \overset{ind}{\sim} N(\mu(s_i), \tau^2 \mathbf{I}_n)$$

    for $i = 1, 2, \ldots, 900$. Here $\mu(s_i) = 4.5 + 3.8X_1(s_i) + 8X_2(s_i) + \eta'(s_i)$ models the mean structure of $Y(\mathbf{s})$.

- **Spatial Process Model**

$$\eta'(s_i)|\tilde{\theta} \sim N(\mathbf{0}, \Sigma(\tilde{\theta}))$$

    Here $\Sigma(\tilde{\theta})$ is induced by a stationary and isotropic exponential covariance function $C(s, s'; \tilde{\theta})$ that depends on a set of parameters $\tilde{\theta} = (\sigma^2 = 3, \phi = 2, \tau^2 = 1)$.

Simulated Spatial Process with Exponential Covariance Function

I then fit a Bayesian hierarchical model to estimate the parameters using the `spLM` function with three different knot settings to investigate whether the estimates of $\tilde{\theta}$ depends on the knot settings. These are the knot settings:

- A regular grid which is a 6×6 partition of the original grid, i.e., 36 knots.

- A regular grid which is a 12×12 partition of the original grid, i.e., 144 knots.

- Randomly selected 324 locations in $\mathcal{D}$, i.e, 324 knots.

Here is the hierarchical model specification of the parameter model:

- **Parameter Model**

$$\beta_j \propto 1 \text{ for } j = 0, 1, \ldots, 3$$
$$\sigma^2 \sim \text{InvGamma}(shape = 1, scale = 2)$$
$$\tau^2 \sim \text{InvGamma}(shape = 2, scale = 0.5)$$
$$\phi \sim \text{Unif}(0.15, 3)$$

Here is my justification for these priors on the parameters. I specify an inverse gamma prior for σ^2 with shape parameter = 2 and scale parameter = 1. This achieves infinite variance. For the range, I used a uniform distribution such that the maximum distance between villages is contained in it. Since the maximum distance among points in my simulated dataset is approximately 14.14214, I will specify a uniform prior distribution for ϕ such that the effective range can range between 1 and 20. Given that the spBayes package parametrizes the exponential function as $\frac{1}{\phi}$, this means that when the effective range is 1, $\phi = 3/1 = 3$ and when the effective range is 20, $\phi = 3/20 = 0.15$.

2.1 Knot setting 1

Here I used a regular grid with 36 knots over $\mathcal{D}$ and ran the MCMC for 30000 iterations with a burn-in of 12000. The run-time was 3.75 minutes. See the traceplots below in Figure 1 and Figure 2.

With 36 knots, I obtained median estimates for $\tilde{\beta}$ of 4.7, 3.76 and 8.0, which are very close to the true values $\tilde{\beta} = (4.5, 3.8, 8)$ (See Table 1). The covariance function parameters estimates, however, are not as good (See Table 2). σ^2, whose median estimate is 3.47, is the only parameter whose estimate is close to the true value of 3. The median estimate of τ^2 is 0.254, which is quite far from the true value of 1. Similarly, the median estimate for $\frac{1}{\phi}$ is 0.253. This implies a median estimate for ϕ of $1/0.253 = 3.95$, which is very far from the true value of 2.
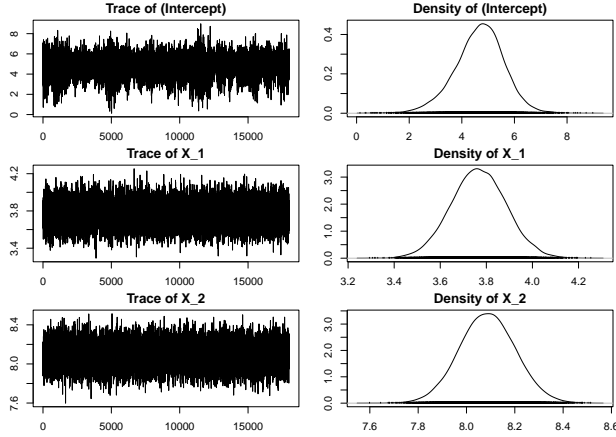
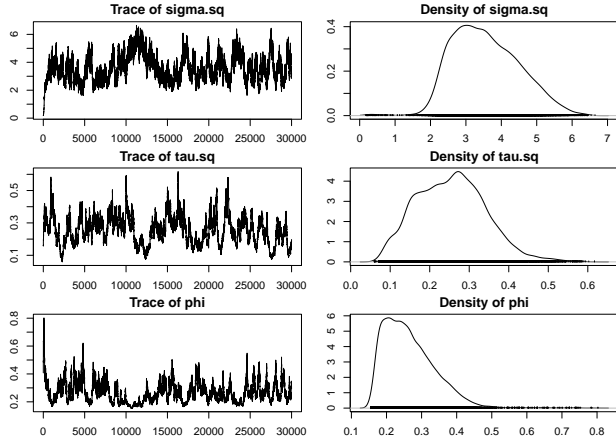Figure 1: Trace Plots of Mean Model Parameters Setting 1

Table 2: Bayesian Estimates of Covariance Function Parameters Under Knot Setting 1

	sigma.sq	tau.sq	phi
Mean	3.566	0.255	0.266
2.5%	2.099	0.101	0.170
Median	3.471	0.254	0.253
97.5%	5.514	0.435	0.431

2.2 Knot Setting 2

Here I used a regular grid with 144 knots over $\mathcal{D}$ and ran the MCMC for 30000 iterations with a burn-in of 12000 and total run time of 38.5 minutes. See the traceplots below (Figure 3 and Figure 4).
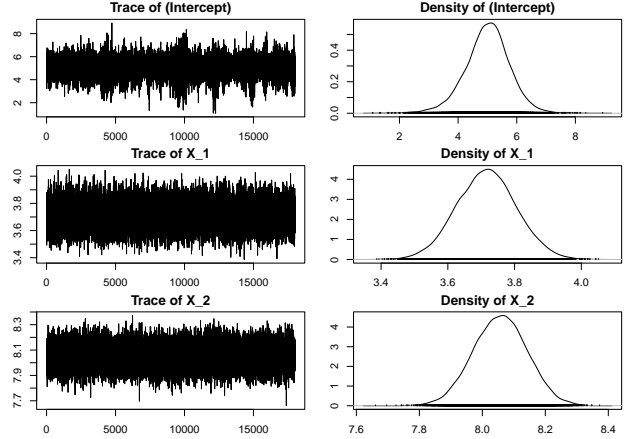

Figure 2: Trace Plots of the Covariance Parameters Setting 1

Figure 3: Trace Plots of Mean Model Parameters Setting 2

Table 1: Bayesian Estimates of Mean Model Parameters Under Knot Setting 1

	Intercept	X_1	X_2
Mean	4.671	3.766	8.087
2.5%	2.687	3.533	7.863
Median	4.721	3.766	8.087
97.5%	6.407	4.004	8.315

With 144 knots, I obtained median estimates for $\tilde{\beta}$ of 5.0, 3.71 and 8.0, which are still quite close to the true values $\tilde{\beta} = (4.5, 3.8, 8)$ (See Table 3). Although the estimates of the covariance function parameters here are better than those obtained with 36 knots; they are still not that good (See Table 4). The median estimate for σ^2 is 3.1, which is better than that obtained from 36 knots and very close to the true value of 3. The median estimate of τ^2 is 0.077, which is still quite far from the true value of 1. Similarly, the median estimate for $\frac{1}{\phi}$ of 0.353 implies a median estimate for ϕ of $1/0.353 = 2.83$, which is somewhat far from the true value of 2; albeit a better estimate than the estimate with 36 knots.
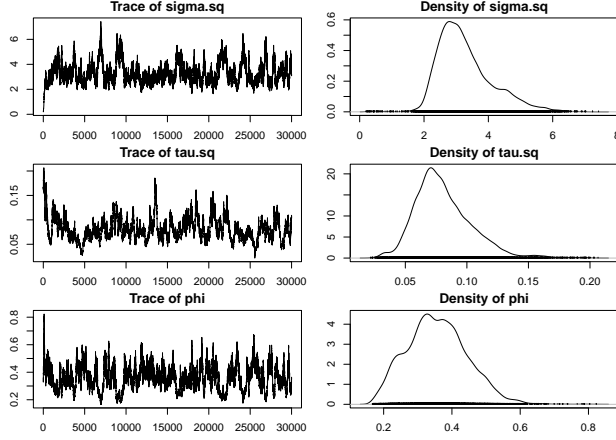
Figure 4: Trace Plots of the Covariance Parameters Setting 2

Table 3: Bayesian Estimates of Mean Model Parameters Under Knot Setting 2

	Intercept	X_1	X__2
Mean	4.998	3.716	8.057
2.5%	3.397	3.541	7.890
Median	5.028	3.716	8.057
97.5%	6.459	3.892	8.221

Table 4: Bayesian Estimates of Covariance Function Parameters Under Knot Setting 2

	sigma.sq	tau.sq	phi
Mean	3.274	0.081	0.357
2.5%	2.102	0.045	0.204
Median	3.103	0.077	0.353
97.5%	5.324	0.134	0.535

2.3 Knot Setting 3

For this knot setting I randomly picked 324 locations in $\mathcal{D}$ and ran the MCMC for 30000 iterations with a burn-in of 12000. The goal here was not just to see whether more knots improved the quality of the estimates (which is what we would expect ((Banerjee et al. 2008))), but whether the arrangement of the knots matters (regular grid vs. irregular (chosen at random)). The run-time for this setting was 4.69 hours. See the traceplots below (Figure 5 and Figure 6).
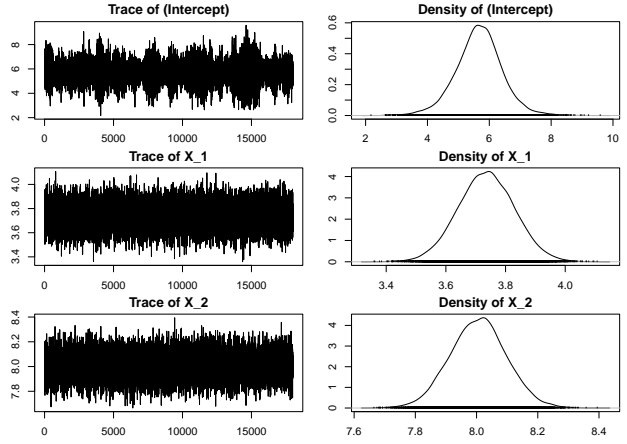

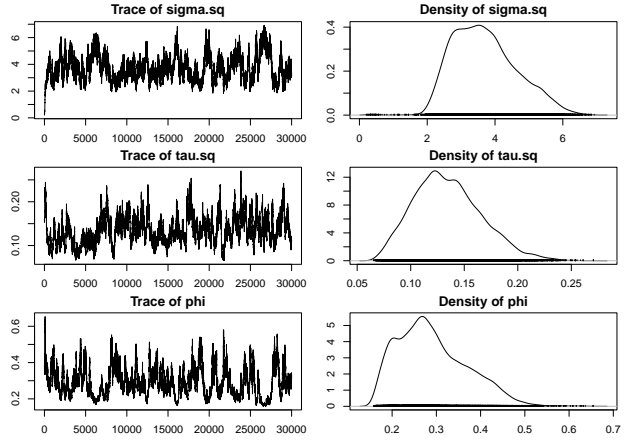Figure 5: Trace Plots of Mean Model Parameters Setting 3

Figure 6: Trace Plots of the Covariance Parameters Setting 3

With 324 randomly chosen knots, I obtained median estimates for $\tilde{\beta}$ of 5.6, 3.73 and 8.0; which, apart from the estimate for β_0, are still close to the true values $\tilde{\beta} = (4.5, 3.8, 8)$ (See Table 5). It is surprising that even with more knots, the estimates of the covariance function parameters here are worse than those

4

Table 5: Bayesian Estimates of Mean Model Parameters Under Knot Setting 3

	Intercept	X_1	X_2
Mean	5.663	3.735	8.006
2.5%	4.091	3.550	7.826
Median	5.674	3.736	8.007
97.5%	7.162	3.920	8.186

Table 6: Bayesian Estimates of Covariance Function Parameters Under Knot Setting 3

	sigma.sq	tau.sq	phi
Mean	3.699	0.135	0.291
2.5%	2.243	0.081	0.174
Median	3.596	0.132	0.279
97.5%	5.724	0.203	0.459

obtained with 36 knots and 144 knots (See Table 6). σ^2, whose median estimate is 3.596, is further from the true value of 3 on this knot setting than on the other previous two settings. The median estimate of τ^2, 0.132, is still quite far from the true value of 1. Similarly, the median estimate for $\frac{1}{\phi}$, 0.279, which implies a median estimate for ϕ of $1/0.279 = 3.58$; is very far from the true value of 2.

3 Discussion

Although, predictive process models provide very good estimates for the mean model parameters $\tilde{\beta}$ on all knot settings I have investigated; I have found that the estimates of some of the covariance function parameters, viz. ϕ and τ^2, are far from their true values. This is what we would expect given that the projected predictive process model for $Y(\mathbf{s})$ is coarser than the parent process. There was little improvement in my estimates even after I increased the number of knots. What was even more surprising is that an irregular grid of 324 knots (randomly selected) led to worse parameter estimates than regular grids with 36 knots and 144 knots respectively. The reason this is surprising is that (Banerjee et al. 2008), page 13 suggest that "estimation is more sensitive to the number of knots than to the underlying design." My work here suggests that a regular grid with fewer knots does better than an irregular grid with more knots!

One limitation of my study is that I am comparing apples and oranges. That is, in order to raise this surprising challenge to (Banerjee et al. 2008)'s suggestion, I need to compare: (1) an irregular grid with fewer knots to an irregular grid with more knots; or (2) a regular grid to an irregular grid where both have same number of knots but more knots than other regular grids. Future work can be directed towards making these comparisons.

References

Banerjee, Sudipto, Alan E Gelfand, Andrew O Finley, and Huiyan Sang. 2008. "Gaussian Predictive Process Models for Large Spatial Datasets." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (4): 825–48.

Cressie, Noel, and Gardar Johannesson. 2008. "Fixed Rank Kriging for Very Large Spatial Datasets." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (1): 209–26.

Guan, Yawen, and Murali Haran. 2018. "A Computationally Efficient Projection-Based Approach for Spatial Generalized Linear Mixed Models." *Journal of Computational and Graphical Statistics* 27 (4): 701–14.

Higdon, Dave. 2002. "Space and Space-Time Modeling Using Process Convolutions." In *Quantitative Methods for Current Environmental Issues*, 37–56.

Johannesson, Gardar, and Noel Cressie. 2004. "Finding Large-Scale Spatial Trends in Massive, Global, Environmental Datasets." *Environmetrics: The Official Journal of the International Environmetrics Society* 15 (1): 1–44.

Katzfuss, Matthias. 2017. "A Multi-Resolution Approximation for Massive Spatial Datasets." *Journal of the American Statistical Association* 112 (517): 201–14.