Reparameterized SGLMM (Spatial Generalized Linear Mixed Model): A Review

Frances Lin

June 2022

1. Introduction

The SGLMM (spatial generalized linear mixed model) for areal data is a hierarchical model that introduces spatial dependence through a GMRF (Gaussian Markov random field, Besag et al., 1991). It was initially proposed for prediction for count data, which can be reframed as problem of image restoration. It has later been applied to estimation and prediction for other types of data (e.g. binary) and found applications in many fields (e.g. ecology, geology, epidemiology, image analysis, and forestry).

The SGLMM has been the dominant model for areal data because of its flexible hierarchical specification, the availability of the software `WinBUGS` for data analysis (Lunn et al., 2000) and various theoretical and computational advantages over the competitive model named automodel. However, SGLMMs suffer from two major shortcomings: i) variance inflation due to spatial confounding and ii) computational challenges posed by high dimensional latent variables (or random effects).

On the other hand, while another model named RHS model seeks to alleviate confounding by introducing synthetic predictors that are orthogonal to the fixed-effects predictors (Reich et al., 2006), failing to account for the underlying graph, the RHS model can result in random-effects structure with negative spatial dependence (i.e. repulsion), which is not typically applicable in practice.

In this paper, a sparse, reparameterized model is proposed, and it is able to alleviate confounding, and, at the same time, include patterns of positive spatial dependence (i.e. attraction) while excluding patterns of repulsion. Other methods have been proposed (Higdon, 2002; Cressie & Johannesson, 2008; Banerjee et al., 2008; Furrer et al., 2006; Rue & Tjelmeland, 2002), but these methods focus on dimension reduction for spatial point-referenced models. The proposed model is thus considered one of the first dimension reduction techniques for spatial areal models.

2. Traditional SGLMM (Spatial Generalized Linear Mixed Model)

The traditional SGLMM for areal data (sometimes referred to as the BYM model) is a hierarchical model (Besag, et al., 1991).

Undirected graphs

Undirected graphs G are typically used to represent the conditional independence properties. Let $G = (V, E)$ be an undirected, labelled graph, where $V = \{1, 2, ..., n\}$ is a set of vertices (nodes) and $E = \{i, j\}$ is a set of edges, where $i, j \in V$, $i \neq j$. Each vertex v represents an area of interest and each edge e represents the proximity of areas i and j. The graph G is represented using an adjacency matrix A, which is a nxn matrix with $diag(A) = 0$ and entries $A_{ij} = 1\{(i, j) \in E, i \neq j\}$, where $1(\cdot)$ is an indicator function. That is, if the entry a_{ij} is $= 1$, then vertex v_i and v_j are adjacent, and the entry is $= 0$, otherwise.

The traditional SGLMM

Further let $Z = (Z_1, ...Z_n)^T$ be the random field of interest, where Z_i is the random variable associated with vertex i. Then, the first stage of the model is given by

$$g(E(Z_i|\beta, W_i)) = X_i\beta + W_i, \tag{1}$$

where g is a link function, X_i is the ith row of the design matrix X, β is a p-vector of regression parameters and W_i is a spatial random effect associated with vertex i. Different types of data require different canonical choices of the link function g. For example, for binary spatial data, the link is the logit function, for count data, it is the natural logarithm function, and for normal data, it is the identity function (Nelder & Wedderburn, 1972).

GMRF prior for the random effects

The field of random effects $W = (W_1, ..., W_n)^T$, through which spatial dependence is incorporated, is most commonly assumed to follow the so-called intrinsic conditionally autoregressive (ICAR, Besag & Kooperberg, 1995) or an improper Gaussian Markov random field (GMRF, Rue & Held, 2005) prior

$$p(W|\tau) \propto \tau^{rank(Q)/2} \exp(-\frac{\tau}{2} W^T Q W), \tag{2}$$

where τ is a smoothing (or precision) parameter and $Q(= Q(\tau)) = diag(A1) - A$ is a precision matrix. The precision matrix Q is the matrix inverse of the covariance matrix Σ (i.e., $Q = \Sigma^{-1}$), the precision parameter is the inverse of the variance parameter ($\tau = 1/\sigma^2$), and A is an adjacency matrix. The precision matrix Q incorporates both dependence and prior uncertainty. That is, W_i and W_j, $i \neq j$, are independent given their neighbors iff $Q_{ij} = Q_{ji} = 0$ iff $(i, j) \notin E$. Uncertainty about W_i is inversely proportional to the degree of vertex i ($Q_{ii} = A_i 1$, where A_i is the ith row of A).

For spatial data over a continuous domain, i.e. point-referenced data or geostatistical data, a SGLMM can be formulated by replacing the GMRF with a GP (Gaussian process, Diggle et al., 1998; De Oliveira, 2000; Christensen & Waagepetersen, 2002).

Since prior (2) is improper (Q is singular), the SGLMM is restricted to a Bayesian or restricted maximum likelihood (REML) analysis. In contrast, the models in Section 3 and 4 have invertible precision matrices, so both classical and Bayesian analyses can be considered, but Bayesian terminology are used throughout the paper.

3. Spatial confounding

The automodel

The automodel, SGLMM's closest competitor, is a Markov random field (MRF) model that incorporates dependence directly and can be defined as

$$g(E(Z_i|\beta,\eta,Z_{-i})) = X_i\beta + \eta \sum_{(i,j)\in E} Z_j^*, \qquad (3)$$

where g is a canonical link function, η is the dependence parameter ($\eta > 0$ implies attraction; $\eta < 0$ implies repulsion), and Z_{-i} is the field excluding the ith observation. For the centered automodel, $Z_j^* = Z_j - \mu_j$, where μ_j is the independence expectation of Z_j (i.e., $\mu_j = E(Z_j|\beta,\eta=0) = g^{-1}(X_j\beta)$). For the uncentered automodel, $Z_j^* = Z_j$.

The sum term $\sum_{(i,j)\in E} Z_j^*$ is called the autocovariate, and it is considered as a synthetic predictor since it uses the observations or observations along with the regression component $X\beta$ of the model. The autocovariate of the centered automodel makes the dependence parameter easily interpretable (η captures the relativity of an observation to its neighbours, conditional on the hypothesized regression component). On the other hand, the autocovariate of the uncentered automodel not only poses conceptual challenge but also shows spatial confounding. That is, the uncentered autocovariate is difficult to interpret, and β and η also tend to be strongly correlated.

For the autologistic model, ML (maximum likelihood) and Bayesian inference are complicated by an intractable normalizing function (Hughes, 2014). The **R** package `ngspatial` resolves this issue through (1) composite likelihood inference and (2) auxiliary-variable MCMC for Bayesian inference, which allows the normalizing function to be cancelled from the Metropolis-Hastings acceptance probability or ratio (Hughes, 2014).

The RHS model

The traditional SGLMM can also shown to be confounded. More specifically, introduction of the random effects can inflate the variance of the posterior distribution of β. This is because the traditional model contains predictors that are collinear with X, and this collinearity causes the variance inflation (Reich et al., 2006). The RHS model (later referred to as the restricted spatial regression) not only seeks to alleviate confounding but also speeds computing (Reich et al., 2006).

Consider P be a projection onto $C(X)$

$$P = X(X^TX)^{-1}X^T$$

and let $P^\perp$ be the projection onto $C(X)$'s complement such that $P^\perp = I - P$, then equation (1) can be rewritten as

$$g(E(Z_i|\beta,W_i)) = X_i\beta + K_i\gamma + L_i\delta,$$

where K and L are orthogonal bases for $C(X)$ and $C(X)^\perp$ respectively and γ and δ are random coefficients. K and X now share the same column space, and this is the source of the spatial confounding.

Since K has no practical meaning, setting $\gamma = 0$, the first stage of the RHS model is given by

$$g(E(Z_i|\beta, \delta)) = X_i\beta + L_i\delta,$$

and, compared to equation (2), the prior for the random effects δ becomes

$$p(\delta|\tau) \propto \tau^{(n-p)/2} \exp(-\frac{\tau}{2}\delta^T Q_R \delta),$$

where $Q_R = L^T Q L$.

The traditional SGLMM and RHZ model both share very similar approach to that of the automodel. The traditional SGLMM model makes use of synthetic predictors K and L, and the RHZ model makes use of L. Furthermore, the traditional SGLMM is analogous to the uncentered automodel since both models include predictors — K and the uncentered autocovariate $\sum Z_j^*$, respectively — that lead to spatial confounding. The RHZ model is analogous to the centered automodel since both models are designed to fit only residual structure of the data.

4. Sparse reparameterization of the areal SGLMM

The RHZ model does not allow parsimonious fitting of the residual clustering. Because the corresponding geometry of the projection $P^\perp$ fails to account for the underlying graph G, the RHZ model permits structure of negative spatial dependence (i.e. repulsion) in the random effects. However, neighboring observation tends to be similar, rather than dissimilar, so patterns of attraction are more useful in practice. The proposed reparameterized model considers an alternative projection that captures the geometry of the models, thus allowing only patterns of positive spatial dependence (i.e. attraction). Utilizing the geometry of the models also leads to dimension reduction of the random effects naturally.

The reparameterized model

The random effects for the sparse, reparameterized areal SGLMM

i) allow patterns of positive spatial dependence (i.e. attraction) while excluding patterns of repulsion and

ii) have dimension much smaller than n.

Consider the operator $(I - 11^T/n)A(I - 11^T/n)$ that appears in the numerator of Moran's I-statistic (a commonly used for nonparametric method for spatial dependence)

$$I(A) = \frac{n}{1^T A 1} \frac{Z^T(I - 11^T/n)A(I - 11^T/n)Z}{Z^T(I - 11^T/n)Z},$$

where I is a nxn identity matrix and 1 is a n-vector of 1s. Next replace $I - 11^T/n$ with $P^\perp$, then the resulting operator called the Moran operator for X with respect to G, $P^\perp A P^\perp$, appears in the numerator of the generalized Moran's I-statistic

$$I_X(A) = \frac{n}{1^T A 1} \frac{Z^T P^\perp A P^\perp Z}{Z^T P^\perp Z},$$

where P again is a projection onto $C(X)$ so $P^{\perp}$ is the projection onto $C(X)$'s complement.

It has been shown that (Boots & Tiefelsdorf, 2000)

(a) the (standardized) spectrum of a Moran operator comprises the possible values for the corresponding $I_X(A)$ and

(b) the eigenvectors include all possible mutually distinct patterns of clustering residual to X while accounting for G.

The positive and negative eigenvalues correspond to varying degrees of positive and negative spatial dependence, respectively, and the associated eigenvectors are the patterns of spatial clustering of data.

Assume M to be a matrix that contains the first $q \ll n$ eigenvectors of the Moran operator. Further assume $\lambda_q > 0$ since neighboring observations tend to be similar, rather than dissimilar, in practice. At least half of the eigenvectors can be discarded as a result, making it possible to achieve a much greater dimension reduction.

Replacing L with M in the RHS model, the first stage of the reparameterized model is given by

$$g(E(Z_i|\beta, \delta_S)) = X_i\beta + M_i\delta_S,$$

and the prior for the random effects δ_S becomes

$$p(\delta_S|\tau) \propto \tau^{q/2} \exp(-\frac{\tau}{2}\delta_S^T Q_S \delta_S),$$

where $Q_S = M^T Q M$.

The sparse, reparameterized SGLMM model is more closely analogous to the centered automodel than it is to the RHZ model. Both the sparse SGLMM model and centered automodel account for X and the underlying graph. On the other hand, the RHZ model accounts for X but does not account for the underlying graph. The uncentered automodel does not fit residual structure to X but does account for the underlying graph by including (uncentered) autocovariate that sums the neighbors.

An alternative measure of spatial dependence is Geary's C, which uses the graph Laplacian Q in place of A (Geary, 1954). Geary's C is the spatial analogue of Durbin–Watson statistic for measuring autocorrelation in the residuals from a time series regression model (Durbin & Watson, 1950). The eigensystem of $P^{\perp}QP^{\perp}$ would then be an alternative to that of $P^{\perp}AP^{\perp}$.

Table 1 compares and contrasts the five models.

Table 1: Comparision of various SGLMMs

Model	Confounded	Account_G
Traditional SGLMM	Yes	No
Uncentered automodel	Yes	Yes
RHZ SGLMM	No	No
Centered automodel	No	Yes

Model	Confounded	Account_G
Sparse SGLMM	No	Yes

The centered autologistic model and the sparse SGLMM can be fitted using the **R** package `ngspatial`(Hughes, 2014).

5. Dimension reduction for spatial models

Fitting a GP (Gaussian process) model can be computationally expensive since it requires the repeated evaluation of expressions involving inversion of the covariance matrix $H(= H(\phi))$. One approach is to consider Cholesky decomposition of H. However, time complexity of the overall fitting algorithm is in $\mathcal{O}(n^3)$ since Cholesky decomposition of a typically dense H is in $\mathcal{O}(n^3)$. This computational expense makes the analyses of large-scale point-referenced data sets time consuming or indefeasible. Efforts to reduce the computational burden have resulted in approaches such as discrete process convolution (Higdon, 2002), fixed rank kriging (Cressie & Johannesson, 2008), Gaussian predictive process models (Banerjee et al., 2008), covariance tapering (Furrer et al., 2006) and approximation by a GMRF (Rue & Tjelmeland, 2002).

Fitting an areal mixed model can also be computationally demanding. Metropolis-Hastings algorithm for sampling from the posterior distribution of the field of random effects W is slow and may present convergence issue when components of W exhibit strong *a posteriori* dependence. Various methods such as MCMC block sampler have been proposed.

The random effects for the RHZ and the sparse models, in contrast, are *a posteriori* uncorrelated (because of the orthogonality in the columns of L). The computational cost of fitting the RHZ or sparse model lies respectively in the evaluation of the quadratic form $\delta^T Q_R \delta$ or $\delta_S^T Q_S \delta_S$. However, because the random effects remain high dimensional, time complexity of the operation for the RHZ model is in $\mathcal{O}(n^2)$, which is still significant enough to discourage or prevent application to large data sets. By taking full advantage of G, evaluation of the quadratic form $\delta_S^T Q_S \delta_S$ for the sparse model can be $\mathcal{O}(1)$, which makes the model more suitable for large-scale data sets.

6. Simulated application

Since areal models are not directly related to my research area and simulation involves applying five of the aforementioned models to binary, count and normal data, this section is kept as a brief summary.

All three spatial models (e.g. Bernoulli, Poisson and Gaussian) are applied in a Bayesian setting. Several models (e.g. a non-spatial ordinary linear model, the centered autologistic, the traditional SGLMM, the RHZ model and the sparse model) are fitted.

. . . . . .

6.1 Binary data

For simulated binary data, the RHZ and sparse models perform comparably and reliably. The traditional SGLMM, in contrast, not only gives poor estimate but also results in wider CIs.

6.2 Count data

For the simulated count data, the traditional SGLMM gives good estimate but again results in wider CIs.

6.3 Normal data

For the simulated normal data, the RHZ and sparse models result in narrower CIs because the non-spatial model overestimates σ^2. The traditional SGLMM gives poor estimate and causes variance inflation so large that it causes a type II error.

The following figures are obtained from the paper.

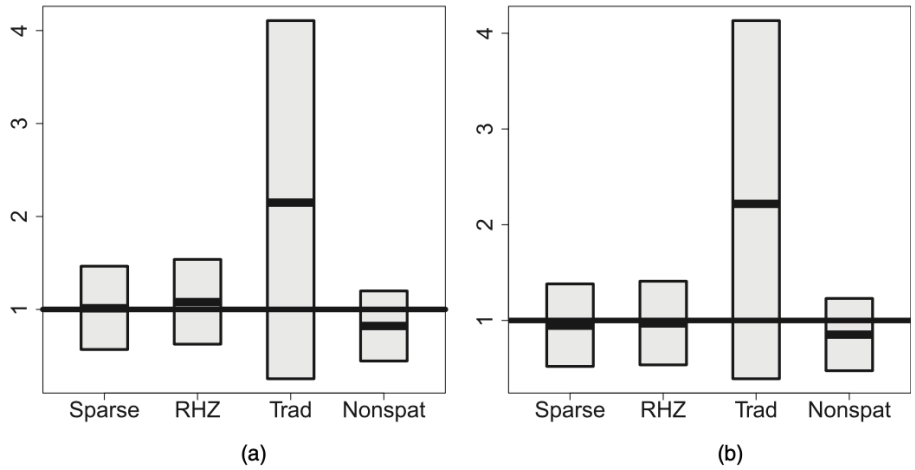

Fig. 4. Boxplots illustrating inference for β for the simulated binary data: (a) β_1; (b) β_2

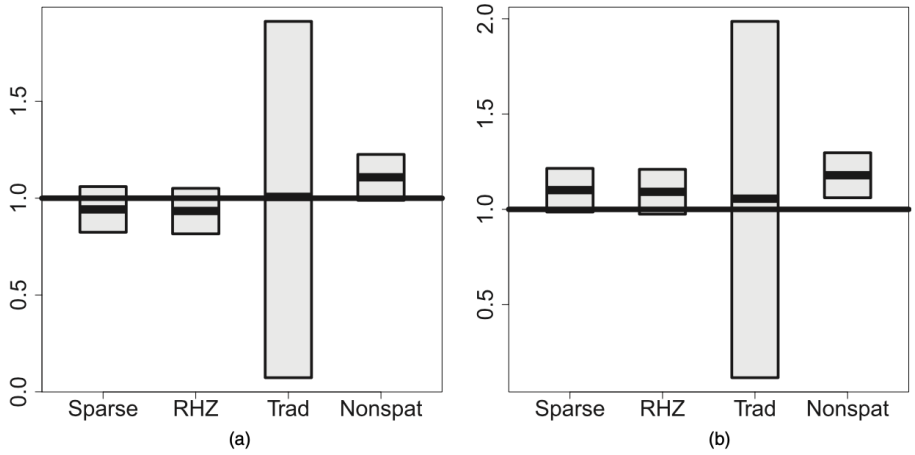

Fig. 8. Boxplots illustrating inference for β for the simulated count data: (a) β_1; (b) β_2

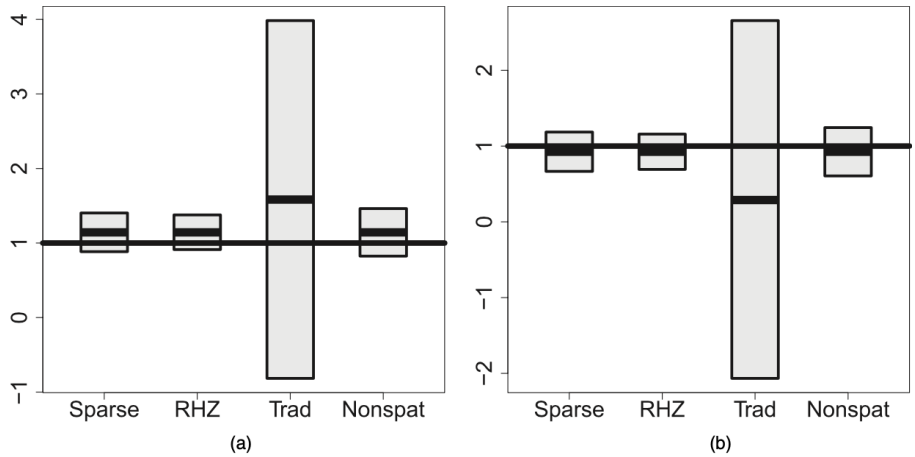

Fig. 9. Boxplots illustrating inference for β for the simulated Gaussian data: (a) β_1; (b) β_2

7. Application

The `infant` data set contains infant mortality rate for $n = 3071$ US counties from the 2008 area source file by the Bureau of Health Professions, Health Resources and Services Administration, US Department of Health and Human Services. The `infant` data set and its adjacency matrix A (respectively, a `data.frame` and `matrix` object) can be accessed from the **R** package `ngspatial`.

A sparse Poisson SGLMM is fitted

$$E(deaths_i|\beta, \delta_S) = births_i \exp(\beta_0 + \beta_1 low_i + \beta_2 black_i + \beta_3 hisp_i + \beta_4 gini_i + \beta_5 aff_i + \beta_6 stab_i + M_i\delta_S),$$

where death is the number of infant deaths, births is the number of live births, low is the rate of low birth weight, black is the percentage of black residents according to the 2000 US census, hisp is the percentage of Hispanic residents (2000 US census), gini is the Gini coefficient which measures income inequality, aff is a score of social affluence and stab is residential stability.

Results show that some estimates (e.g. `low_weight`, `gini`) differ from the result from the original study. The process of fitting the model takes approximately 52.47736 mins.

```
Coefficients:

              Estimate      Lower      Upper      MCSE
X           -4.987e+00 -5.166e+00  -4.807000 1.011e-03
Xlow_weight -1.069e+02 -1.905e+02 -24.930000 3.409e-01
Xblack       9.668e-03  8.516e-03   0.010840 9.032e-06
Xhispanic   -5.416e-03 -6.578e-03  -0.004273 8.165e-06
Xgini        6.227e-02 -3.803e-01   0.491400 2.547e-03
Xaffluence  -9.476e-02 -1.070e-01  -0.082480 5.893e-05
Xstability  -1.333e-02 -2.890e-02   0.002744 9.467e-05

DIC: 10280
```

```
Number of iterations: 1000000
```

8. Discussion

(1) The sparse SGLMM not only alleviates spatial confounding but also accounts for the underlying graph while achieving dimension reduction naturally. However, due to time constraint, some of the computational details are not well explored in this project. For example, MCMC methods and its variants. . . . . .

(2) The sparse SGLMM appear faster, compared to the traditional SGLMM or the RHZ model. However, MCMC methods may still take hours or days to run. INLA algorithm (a project that I am currently working on for the Bayesian Statistics course) is a fast alternative for fitting areal and more models and can be accessed through the **R** package `R-INLA`. It would be interesting to see comparisons of accuracy and speed.

Reference

Bolin, D. (2015). *Lecture 5: Intrinsic GMRFs Gaussian Markov random fields.* Personal Collection of D. Bolin, Chalmers University of Technology, Gothenburg Sweden.

Haran, M. (2011). Gaussian random field models for spatial data. Handbook of Markov Chain Monte Carlo, 449-478.

Hughes, J., & Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75(1), 139-159.

Hughes, J. (2014). ngspatial: A Package for Fitting the Centered Autologistic and Sparse Spatial Generalized Linear Mixed Models for Areal Data. R Journal, 6(2).