Computationally Efficient Spatiotemporal Generalized Linear Modeling

Ellen Graham

Faculty Advisors: Professor Brianna Heggeseth, Professor Alicia Johnson

Department of Mathematics, Statistics, and Computer Science

April 26, 2020

Abstract

Spatiotemporal data is a common occurrence in a variety of fields such as ecology and epidemiology. However, observations are often spatially and temporally correlated, leading naive models of such data to underestimate variance in parameter estimates. Furthermore, methods that do account for this spatial and temporal dependence often take prohibitively long to estimate due to their computational complexity. This paper extends a computationally efficient method for spatial modeling to the spatiotemporal domain while retaining its computational efficiency. We implement this method and examine its effectiveness using a simulation study and by applying it to Carolina Wren population counts in the United States between 1990 and 2010. We find that it performs favorably compared to the naive approach and is significantly more computationally efficient compared to the full spatiotemporal model. Additionally, it requires much less expert knowledge to specify compared to comparable methods, making this method an attractive approach for users with less experience with spatiotemporal modeling.

Acknowledgments

First and foremost I would like to thank Professor Brianna Hegesseth and Professor Alicia Johnson for guiding me through this project and through my college career. I would also like to thank Professor Holly Barcus for her presence on my honors committee. Thank you as well to Suzanne Burr and the MSCS department as a whole for creating a wonderful community to learn and grow in. Finally, I would like to thank my all of friends for supporting me through this project, with special thanks to Conor Broderick for struggling through the honors process with me!

List of Figures

1 Distribution of point estimates for fixed effects across 100 simulated data sets with $\phi = \psi = 0.2$. Black lines show the true value. 9
2 Distribution of point estimates for fixed effects across 100 simulated data sets with $\phi = \psi = .5$. Black lines show the true value. 10
3 Distribution of point estimates for fixed effects across 100 simulated data sets with $\phi = \psi = 1$. Black lines show the true value. 11
4 Distribution of point estimates for random effect hyperparameters across 100 simulated data sets with $\phi = \psi = 0.2$. Black lines show the true value. 12
5 Distribution of point estimates for random effect hyperparameters across 100 simulated data sets with $\phi = \psi = 0.5$. Black lines show the true value. 13
6 Distribution of point estimates for random effect hyperparameters across 100 simulated data sets with $\phi = \psi = 1$. Black lines show the true value. 14
7 Observed Carolina Wrens at Selected Sites for 1990 Survey 16
8 Observed Carolina Wrens at Selected Sites for 1991 Survey 17
9 Posterior Predictions for the Generalized Linear Model. True distribution shown in dark blue, predicted distributions shown in light blue. 18
10 Posterior Predictions for the Spatiotemporal Generalized Linear Mixed Model. True distribution shown in dark blue, predicted distributions shown in light blue. 19
11 Posterior estimates for the Coastal Flatwoods stratum coefficient. The shaded region is 95% central posterior density, the dark blue line is the coefficient point estimate using the median of the draws. 20
12 Credible intervals for β_0 across simulations for each model when $\phi = \psi = 0.2$. 24
13 Credible intervals for β_1 across simulations for each model when $\phi = \psi = 0.2$. 25
14 Credible intervals for β_2 across simulations for each model when $\phi = \psi = 0.2$. 26
15 Credible intervals for β_3 across simulations for each model when $\phi = \psi = 0.2$. 27
16 Credible intervals for β_0 across simulations for each model when $\phi = \psi = 0.5$. 28
17 Credible intervals for β_1 across simulations for each model when $\phi = \psi = 0.5$. 29
18 Credible intervals for β_2 across simulations for each model when $\phi = \psi = 0.5$. 30
19 Credible intervals for β_3 across simulations for each model when $\phi = \psi = 0.5$. 31
20 Credible intervals for β_0 across simulations for each model when $\phi = \psi = 1$. β_0 . . . 32
21 Credible intervals for β_1 across simulations for each model when $\phi = \psi = 1$. 33
22 Credible intervals for β_2 across simulations for each model when $\phi = \psi = 1$. 34
23 Credible intervals for β_3 across simulations for each model when $\phi = \psi = 1$. 35
24 Simulated Data GLM Markov chain trace plots 36
25 Simulated Data STGLMM Markov chain trace plots for fixed effects 37
26 Simulated Data STGLMM Markov chain trace plots for hyperparameters 38
27 Simulated Data STGLMM Markov chain trace plots for spatial effects (1-25) 39
28 Simulated Data STGLMM Markov chain trace plots for spatial effects (26-50) 40
29 Simulated Data STGLMM Markov chain trace plots for temporal effects 41
30 Carolina Wren GLM Markov chain trace plots 42
31 Carolina Wren STGLMM Markov chain trace plots for fixed effects 43
32 Carolina Wren STGLMM Markov chain trace plots for hyperparameters 44
33 Carolina Wren STGLMM Markov chain trace plots for spatial effects (1-25) 45
34 Carolina Wren STGLMM Markov chain trace plots for spatial effects (26-50) 46
35 Carolina Wren STGLMM Markov chain trace plots for temporal effects 47

List of Tables

1 Parameter Sets for Simulated Data 8
2 Proportion of times true value is within 50% and 90% credible intervals across 100 simulated datasets 15

Contents

1 Introduction **1**

2 Existing Methods **2**
2.1 Spatial Models 2
2.2 Spatiotemporal Models 5

3 Random Projections for the Spatiotemporal Generalized Linear Mixed Model **6**
3.1 Model Estimation 7

4 Applications **7**
4.1 A Simulation Study 7
4.2 Bird Population Counts 15

5 Discussion and Future Work **20**

References **22**

R Packages **23**

A Appendix **24**
A.1 Conditional Likelihood Equations 24
A.2 Simulation Coverage Results 24
A.3 Markov Chain Diagnostics 35

1 Introduction

Spatial data are of interest in a wide range of domains. For example, it is important to take into account patient location when considering disease prevalence in epidemiology. In ecology, studying and accounting for physical location is crucial to modeling animal populations. When working with such data, researchers often must take into account the potential spatial dependence; observations that are close in space are likely to be more similar than observations that are far apart in space. This violates the independence assumption critical to many common statistical models, such as generalized linear models. Modeling spatially dependent data using models that assume independence leads to incorrect standard error calculations, which can, for example, provide misleading results when performing hypothesis tests. Data collected over time also have the same feature, where observations close in time are likely to be more similar than observations far apart in time, such as in longitudinal data or time series. Furthermore, many data contexts have data collected both over time and space. This means we must not only account for spatial and temporal dependence, but may also need to account for the ways in which these dependence structures change over time, adding further complexity to modeling such data.

With higher model complexity often comes computational expense. Because each location and time can be viewed as additional variables observed on a single data point, spatiotemporal data is often high dimensional. Models that account for correlation in this way are called random effects models, and consist of fixed effects of covariates and random effects for each time and location. Such high dimensionality can make estimating models for even a moderate number of subjects, in the range of 1000 to 10000 observations, take prohibitively long. This computational complexity includes large matrix operations as well as strong correlations between random effects, which slows mixing of Markov Chain Monte Carlo (MCMC) algorithms when using Bayesian inference [1, 2]. Additionally, many methods that attempt to make such models computationally tractable do so by requiring the modeler to specify model features such as assumed basis functions that can make spatiotemporal approaches inaccessible to those with less statistical training or less computational resources. Developing models that avoid these shortcomings are a promising path forward to increasing the positive and effective impact of statistics on peoples lives.

One prominent example of the difficulties of spatial methods in practice is the work of the United Church of Christ's seminal paper on racism in toxic waste facility siting decisions [3, 4]. In the original study, performed in 1987, and in their follow-up study in 2007, the researchers were unable to appropriately account for the spatial dependence in both facility locations and populations that live near them due to lack of methods accessible to social scientists and activists not trained in statistics. In their 1987 study, this caused them to significantly underestimate the correlation between minority populations and toxic waste facility sites and so underestimate the effects of systemic racism on environmental hazard siting. While the 2007 study improved on this, they were still unable to come up with robust estimates of the uncertainty of this correlation.

One recent technique that attempts to account for spatial dependence while prioritizing both computational efficiency and specification simplicity is the work of Guan and Haran [1] in their paper "A Computationally Efficient Projection-Based Approach for Spatial Generalized Linear Mixed Models." Here, the covariance structure of the data is projected onto a lower-dimensional space, allowing for speedier computation of matrix operations, while additionally decorrelating random effects, allowing for faster mixing of MCMC methods. In section 2 we discuss existing methods for modeling spatial and spatiotemporal data. The method from Guan and Haran is discussed in 2.1.4. In section 3, I extend this method into the temporal domain while retaining many of the computational advantages as well as a relatively simple model specification, making the model more accessible to non-experts. Section 4 applies this model to both simulated and real-world data to demonstrate its computational efficiency as well as its accuracy.

2 Existing Methods

2.1 Spatial Models

2.1.1 Spatiotemporal Gaussian Processes

One common approach to spatial modeling treats the data as a Gaussian process realized at observed locations. A Gaussian process is a random process such that every finite set of random variables drawn from it, here variables at observed locations, follow a multivariate normal distribution. We use the notation example in Banerjee, Carlin, and Gelfand [5] and Zhang [6] . Let $Y(s)$ be the response random variable and $\boldsymbol{x(s)}$ be a $p \times 1$ vector of explanatory variables at spatial location s. Assuming a linear model between x and Y, we let $Y(s)$ be modeled by

$$Y(s) = \boldsymbol{x}(s)^T \boldsymbol{\beta}(s) + e(s)$$

where $\boldsymbol{\beta}(s)$ is a vector of length p of coefficients, which are often assumed to be identical for each location, such that $\boldsymbol{\beta(s)} = \boldsymbol{\beta}$. The residual term of the model $e(s)$ can capture spatial dependence if we let

$$e(s) = w(s) + \epsilon(s)$$

where $\epsilon(s)$ is a Gaussian noise process and $w(s)$ is a spatial Gaussian process realized at location s. If these processes are realized at a set of locations $\boldsymbol{s}$, then $\boldsymbol{w} \sim N(0, \boldsymbol{\Sigma}(\sigma_w^2, \phi))$ and $\epsilon \sim N(0, \sigma_\epsilon^2 \boldsymbol{I})$. This lets $\boldsymbol{w}$ be a mean zero random effect with variance parameter σ_w^2 and an $n \times n$ covariance matrix $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$ for n spatial locations. The correlation of the errors across spatial location is assumed to be a function of distance between spatial points, d, and parameter ϕ, $\rho(d; \phi)$. For the ith and jth locations s_i and s_j, the distance between them is $d_{ij} = d(s_i, s_j)$ and

$$\boldsymbol{\Sigma}(\sigma_w^2, \phi)_{ij} = \sigma_w^2 \rho(d_{ij}; \phi)$$

The parameter ϕ in the correlation function controls the properties of the assumed correlation function, such as the strength of correlation between two locations. If points were independent and the correlation between any two distinct locations were zero, ie. $\rho(d_{ij}; \phi) = 0$ for $i \neq j$, our covariance matrix would be $\sigma_w^2 \boldsymbol{I}$, and our model would simplify to a linear model. For the spatial case, a common choice for $\rho(\cdot; \phi)$ is the Matérn covariance function. This function has two parameters, ϕ and ν. Only one of these parameters is identifiable, and we often choose to estimate ϕ instead of ν [5]. The equation for the Matérn covariance simplifies significantly for values of ν of the form $\nu = n + \frac{1}{2}$ for integers n. In fact, when ν is 0.5, the Matérn covariance function simplifies to the exponential covariance function. In practice, we often choose ν to be one of 0.5, 1.5, or 2.5 to take advantage of the simplified equations.

2.1.2 SGLMM

This model is referred to as a spatial linear mixed model due to the combination of fixed effects $\boldsymbol{\beta}$ and spatial random effects $w(s)$ [6]. This spatial model can be extended for non-Gaussian outcomes $Y(s)$ for some link function $g(\cdot)$,

$$g(E[Y(s)|\boldsymbol{\beta}, w(s)]) = \boldsymbol{x}(s)^T \boldsymbol{\beta} + w(s)$$
$$\boldsymbol{w} \sim N(0, \boldsymbol{\Sigma}(\sigma_w^2, \phi))$$

This model is a spatial generalized linear mixed model, or the SGLMM. One issue with the SGLMM is the high dimension of the random effects $\boldsymbol{w}$. With the model above, we have one random effect for each location. This can lead to computational issues when the number of locations is large.

For example, to calculate the likelihood function, we must invert the $n \times n$ covariance matrix. Matrix inversion is a computationally expensive operation, with time complexity $O(n^3)$. For large n, this quickly becomes infeasible. Additionally, when fitting the model using Bayesian methods, the correlation among the high dimensional random effects lowers the efficiency of MCMC methods, increasing the number of iterations needed to get an accurate estimate of the posterior distribution of parameters.

2.1.3 Existing Simplifications of the SGLMM

Many approaches have been developed to address the computational issues of fitting SGLMMs to data. In the Gaussian case, it is possible to integrate out the spatial effects $w(s)$, which is referred to as marginalization. This allows for simpler model fitting, however it cannot be used for the non-Gaussian case [7].

One popular choice for dealing with the computational issues encountered with non-Gaussian outcomes is the predictive process model [8]. The predictive process replaces $w(s)$ with $\tilde{w}(s)$, where for some small set of locations S^* (called knots) that may or may not be disjoint with the observed locations, we let $\boldsymbol{w}^*$ be the realization of the spatial Gaussian process $\tilde{w}(s)$ at the points in S^*. We are then left with $\boldsymbol{w}^* \sim N(0, Cov^*(\phi))$, which has a much smaller dimension covariance matrix. Let $c(s_0; \phi)$ be the evaluation of the covariance function used to generate $Cov^*(\phi)$ at some point s_0 with all of the knots in S^* and, we have

$$\tilde{w}(s_0) = c^T(s_0; \phi) Cov^{*-1}(\phi) \boldsymbol{w}^*$$

Thus, when the size of the set of knots is much smaller than the original number of locations, a much smaller dimension covariance matrix must be inverted. This method effectively uses a small set of locations, spaced over a similar area as the original observed locations, to estimate the parameter ϕ for the full Gaussian process. In this sense, it is a low-rank method, as the rank of its covariance matrix is much smaller than that of the original process. Choosing the appropriate knots is not a straightforward process. Finley et al. [9] propose a sequential search over the observed locations to select knots which is itself is an intensive process.

Another choice for reducing computational complexity in the SGLMM comes from Hughes and Haran [10]. They propose using the Moran I operator, which is derived from a common measure of spatial dependence, the Moran I statistic, to make a low-rank approximation of the spatial dependence structure. They do this by applying this operator to their observed data, then taking its eigendecomposition to identify the subset of vectors that best explain the variation in the data. While this method works well for areal data, it does not apply to situations where data are observed at points. However, its methodology is very similar to the random projections approach discussed in 2.1.4.

2.1.4 Random Projections for the SGLMM

The random projections based approach to dimension reduction of the spatial effects in an SGLMM from Guan and Haran [1] circumvents the issues of selecting appropriate knots or basis functions. This low-rank model relies on decorrelating the random effects using the eigendecomposition of their covariance matrix. However, calculating the eigendecomposition of a matrix is a computationally expensive operation, so they rely on the random projections algorithm to quickly calculate an approximation of the eigendecomposition. To see how this works, recall the SGLMM framework:

$$g(E[Y(s)|\boldsymbol{\beta}, w(s)]) = \boldsymbol{x}(s)^T \boldsymbol{\beta} + w(s)$$
$$\boldsymbol{w} \sim N(0, \boldsymbol{\Sigma}(\sigma_w^2, \phi))$$

Note that the covariance matrix of $\boldsymbol{w}$, $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$, is symmetric and positive semi-definite. This is true for all covariance matrices, and means that the eigendecomposition $\boldsymbol{\Sigma}(\sigma_w^2, \phi) = \sigma_w^2 V \Lambda V^T$ exists for

real valued diagonal $n \times n$ matrix Λ consisting of the eigenvalues of $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$ arranged in descending magnitude, and real valued orthonormal $n \times n$ matrix V, consisting of the eigenvectors of $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$. The properties of orthonormal matrices allow us to create an independent random variable $\boldsymbol{\delta}$ from the correlated spatial effects $\boldsymbol{w}$. This $\boldsymbol{\delta}$ acts as a tool to ease model fitting as we no longer need to invert the covariance matrix of the spatial effects $\boldsymbol{w}$. Consider

$$\boldsymbol{\delta} = (\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2})^T \boldsymbol{w}$$

Then, leveraging the fact that because $\boldsymbol{V}$ is orthonormal, $\boldsymbol{V}^{-1} = \boldsymbol{V}^T$, the covariance matrix of $\boldsymbol{\delta}$ a diagonal matrix:

$$\begin{aligned} Cov(\boldsymbol{\delta}) &= Cov((\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2})^T \boldsymbol{w}) \\ &= \boldsymbol{\Lambda}^{-1/2}\boldsymbol{V} Cov(\boldsymbol{w}) \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2} \\ &= \boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}^T \sigma_w^2 \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2} \\ &= \sigma_w^2 \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} \\ &= \sigma_w^2 \boldsymbol{I} \end{aligned}$$

Hence, $\boldsymbol{\delta} \sim N(0, \sigma_w^2 I)$, meaning it is easy to calculate the likelihood of $\boldsymbol{\delta}$. Then, we have transferred the computational complexity of this likelihood calculation into of the eigendecomposition for $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$. However, the eigenvalue decomposition is just as computationally complex as matrix inversion. This issue is resolved through the use of the random projections algorithm, which creates a low rank approximation of the eigendecomposition of $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$. The details of the method can be found in Banerjee [11], but broadly consist of using a rank $k << n$ matrix whose elements are randomly drawn from a Gaussian distribution to project $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$ to be rank k, taking the singular value decomposition of this reduced matrix, then using those values to approximate the eigendecomposition of the full matrix.

To apply this approximation to the SGLMM, let $\boldsymbol{\Sigma}(\sigma_w^2, \phi) \approx \tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{V}}^T$ where $\tilde{\boldsymbol{V}}$ is the $n \times k$ matrix of the approximated leading eigenvectors and $\tilde{\boldsymbol{\Lambda}}$ is the $k \times k$ diagonal matrix of associated eigenvalues of $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$. If we now let $\boldsymbol{\delta} = (\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}^{1/2})^T \boldsymbol{w}$, we can model the SGLMM using the following specification

$$g(E[Y(s)|\boldsymbol{\beta}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{\Lambda}}, \boldsymbol{\delta}]) = \boldsymbol{x}(s)^T\boldsymbol{\beta} + \left[(\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}^{1/2})\boldsymbol{\delta}\right]_s$$
$$\boldsymbol{\delta}|\phi, \sigma_w^2 \dot{\sim} N(0, \sigma_w^2 \boldsymbol{I})$$

Notice how $(\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}^{1/2})\boldsymbol{\delta} = (\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}^{1/2})(\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Lambda}}^{-1/2})^T\boldsymbol{w} \approx \boldsymbol{w}$, so we are still modeling our original spatial random effect $\boldsymbol{w}$, only using the decorrelated random vector $\boldsymbol{\delta}$. This model has a number of advantages. Decorrelation of random effects works to increase the rate of MCMC mixing. The model interpretation is almost unchanged from the original spatial Gaussian process model, only now our random effects are approximations. And, as opposed to choosing basis functions to approximate the random effects or knot locations to use, we simply need to specify the rank of the approximation. If the rank is sufficiently small, computational issues of calculating the likelihood disappear because the random projections algorithm has time complexity on the order of $O(nk^2)$ as opposed to the original $O(n^3)$ time complexity of matrix inversion. The results are less sensitive to the choice of k than compared to choices of basis functions or knots, making this an attractive model for users with less expertise in spatial modeling [1].

2.2 Spatiotemporal Models

2.2.1 Spatiotemporal Gaussian Processes

The spatial Gaussian process model can be extended into the spatiotemporal domain. Let $Y(s,t)$ be the response variable at location s and time t. We assume that each location is observed at the same times as each other location for convenience. Let there be n locations and m times for each location. Given an explanatory vector $\boldsymbol{x}(s,t)$ of length p, we have

$$Y(s,t) = \boldsymbol{x}(s,t)^T\boldsymbol{\beta}(s,t) + e(s,t)$$
$$e(s,t) = \eta(s,t) + \epsilon(s,t)$$

As before, we let $\boldsymbol{\beta}(s,t) = \boldsymbol{\beta}$ to simplify the model. There are several ways to handle the spatiotemporal process variable $\eta(s,t)$. The simplest choice is to let $\eta(s,t) = w(s) + \alpha(t)$ and assume our spatial effects $w(s)$ are additive and independent of our temporal effects $\alpha(t)$. This simplifies the model fitting process. If we let both $w(s)$ and $\alpha(t)$ to be mean zero Gaussian processes, the full model can be written out as

$$Y(s,t) = \boldsymbol{x}(s,t)^T\boldsymbol{\beta} + \eta(s,t) + \epsilon(s,t)$$
$$\eta(s,t) = \alpha(t) + w(s)$$
$$\boldsymbol{w} \sim N(0, \boldsymbol{\Sigma}(\sigma_w^2, \phi))$$
$$\boldsymbol{\alpha} \sim N(0, \boldsymbol{\Theta}(\theta^2, \psi))$$
$$\boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2\boldsymbol{I})$$

where θ^2 is the variance parameter of the temporal effects and $\boldsymbol{\Theta}(\theta^2, \psi)$ is the $m \times m$ temporal covariance matrix created by an assumed correlation function $\upsilon(\cdot;\psi)$. As before, σ_w^2 is the variance parameter of the spatial effects and $\boldsymbol{\Sigma}(\sigma_w^2, \phi)$ is the $n \times n$ spatial covariance matrix created by an assumed spatial correlation function $\rho(\cdot;\phi)$.

As in the spatial case, if $Y(s,t)$ are non-Gaussian, we can extend this to the general linear mixed model formulation by replacing the first equation with

$$g(E[Y(s,t)|\boldsymbol{\beta}, \eta(s,t)]) = \boldsymbol{x}(s)^T\boldsymbol{\beta} + \eta(s,t)$$

for some link function $g(\cdot)$. Other choices for $\eta(s,t)$ include setting $\eta(s,t) = w(s)\alpha(t)$ and assuming independence of spatial and temporal effects. We might also let $\eta(s,t) = w_t(s)$ so that for each time t we model different spatial random effects $w_t(s)$, or analogously $\eta(s,t) = \alpha_s(t)$, so that for each location s we model different temporal random effects $\alpha_s(t)$. These two models allow either the spatial or the temporal effect to vary across time or space respectively, but still preclude spatial and temporal interaction. An attractive approach for spatiotemporal interaction is the dynamic spatiotemporal model [5, 7]. However, this model is beyond the scope of this paper. All of these models suffer from similar computational difficulties as in the spatial case, where likelihood evaluation is computationally intensive due to matrix inversion, and for Bayesian fitting correlated random effects lead to a lower MCMC mixing efficiency.

2.2.2 Existing Simplifications for the Spatiotemporal Gaussian Processes

Similar to the spatial Gaussian process, developing computationally efficient methods for spatiotemporal Gaussian processes is an active area of research. For example, to ease MCMC mixing, Bradley et al. [12] develop a model for count data which uses strategic choices of prior distributions and parameter distributions to allow the construction of a Gibbs sampler for the model. This sidesteps difficulties that arise when picking proposal distributions for the high dimensional random effects as required by the Metropolis Hasting algorithm [13]. However, computational issues

around calculating the likelihood for the high dimensional random effects remain. Bradley et al.'s [12] model can be extended to use basis function approximations popular in spatial modeling, but this still leaves the difficulty of determining the appropriate basis functions for a given application. Other methods focus on low-rank methods analogous to those in the spatial frameworks to reduce computational load, such as Bradely et al.'s [2] use of the Moran's I basis functions approach. However, the Moran's I approach only works for areal data. The predictive process can be extended to the spatiotemporal domain, allowing for modeling of point referenced data, but then the issues of selecting the appropriate knots (locations) from which to approximate the random effects remains.

3 Random Projections for the Spatiotemporal Generalized Linear Mixed Model

We propose extending the Guan and Haran [1] model to the spatiotemporal case. Then, the gains in computational efficiency realized in the SGLMM could be applied to spatiotemporal data, increasing the size of data that can feasibly be analyzed. Recall the spatiotemporal model with independent spatial and temporal effects given by

$$\begin{aligned} Y(s,t) &= \boldsymbol{x}(s,t)^T\boldsymbol{\beta}(s,t) + \eta(s,t) + \epsilon(s,t) \\ \eta(s,t) &= \alpha(t) + w(s) \\ w &\sim N(0, \boldsymbol{\Sigma}(\sigma_w^2, \phi)) \\ \alpha &\sim N(0, \boldsymbol{\Theta}(\theta^2, \psi)) \end{aligned}$$

Because the spatial and temporal effects are independent, we can apply the random projections algorithm to one or both of them, depending on the dimension of the temporal and spatial data. Then, our model would become

$$\begin{aligned} Y(s,t) &= \boldsymbol{x}(s,t)^T\boldsymbol{\beta} + \eta(s,t) + \epsilon(s,t) \\ \eta(s,t) &= \left[(\boldsymbol{U}\boldsymbol{K}^{1/2})\boldsymbol{\gamma}\right]_t + \left[(\boldsymbol{V}\boldsymbol{\Lambda}^{1/2})\boldsymbol{\delta}\right]_s \\ \boldsymbol{\delta}|\phi, \sigma_w^2 &\dot{\sim} N(0, \sigma_w^2 I) \\ \boldsymbol{\gamma}|\psi, \theta^2 &\dot{\sim} N(0, \theta^2 I) \end{aligned}$$

where $\boldsymbol{\delta}, \boldsymbol{V}, \boldsymbol{\Lambda}$ are the same as above. The temporal effects are decorrelated in the same way as the spatial effects using $\boldsymbol{\gamma} = (\boldsymbol{U}\boldsymbol{K}^{-1/2})^T\boldsymbol{\alpha}$, where the $m \times m$ orthonormal matrix $\boldsymbol{U}$ and $m \times m$ diagonal matrix $\boldsymbol{K}$ come from the eigendecompositon of the temporal covariance matrix such that $\theta^2\boldsymbol{\Theta}(\boldsymbol{\psi}) = \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^T$. Again, to lower the dimension of these effects, we would would choose $k << n$ and $l << m$ such that $\tilde{\boldsymbol{V}}$ is $n \times k$, $\tilde{\boldsymbol{\Lambda}}$ is $k \times k$, $\tilde{\boldsymbol{U}}$ is $m \times l$, $\tilde{\boldsymbol{K}}$ is $l \times l$. To model the generalized linear mixed model version, replace the first equation with $g(E[Y(s,t)|\beta, \eta(s,t)]) = \boldsymbol{x}(s)^T\boldsymbol{\beta} + \eta(s,t)$. We refer to this model as the spatiotemporal generalized linear mixed model, or STGLMM. When $k \ll n$ and $l \ll m$, we refer to this model as the low rank STGLMM, as the rank of the covariance matrices is significantly reduced. When $k = n$ and $l = m$, we refer to this as the full ranks STGLMM, which we analyze to understand the computationally efficiency of the low rank STGLMM.

To choose k and l, we suggest following the recommendation of Guan and Haran [1]. For the purely spatial case, they propose selecting the k that minimizes the BIC of a non-spatial generalized linear model using the decorrelated spatial effects δ as predictors. To choose the temporal rank l, we choose to follow the same process, choosing l such that it minimizes the BIC. In practice, due to MCMC issues, we often choose values that are larger than the values returned by this process, as the spatial variance tends to grow without bound for small values of k, and increasing them minimizes this issue.

3.1 Model Estimation

To implement this model, we wrote code to estimate the STGLMM in R. We estimate the model using the Bayesian framework due to its relative ease in working with hierarchical models. We estimate posteriors for our parameters using Gibbs sampling with Metropolis Hastings steps. We block update the fixed effects $\boldsymbol{\beta}$ using spherical Normal proposals, and similarly block update the spatial ($\boldsymbol{\delta}$) and temporal ($\boldsymbol{\gamma}$) random effects using spherical Normal proposals. We let the priors for ϕ and ψ each follow a Uniform distribution and σ_w^2 and θ^2 each follow an Inverse Gamma distribution [5]. Each of the hyperpriors is updated separately. The eigendecompositions of the assumed spatial and temporal covariance matrices, $\Sigma(\sigma_w^2, \phi)$ and $\Theta(\theta^2, \psi)$ respectively, are recalculated when a new proposal for ϕ or ψ is accepted. We assume the spatial random effects have Matérn covariance with $\nu = 2.5$. Then, for two locations i and j with distance between them d_{ij},

$$\Sigma(\sigma_w^2, \phi)_{ij} = \sigma_w^2 \left(1 + \frac{\sqrt{5}d_{ij}}{\phi} + \frac{5d_{ij}^2}{3\phi^2}\right) \exp\left(-\frac{\sqrt{5}d_{ij}}{\phi}\right)$$

and we assume the temporal random effects have squared exponential covariance. Then, for two times i and j with time difference between them h_{ij},

$$\Theta(\theta^2, \psi)_{ij} = \theta^2 \exp\left(-\frac{h_{ij}^2}{2\psi^2}\right)$$

The conditional likelihoods we use to accept or reject each Metropolis Hastings step can be found in appendix A.1. R Code to estimate this particular model is available at this GitHub repo. This code can estimate models with Poisson or Binomial likelihoods for the above covariance structure. The code is written using Base R [20] version 4.0.3, and the packages Rcpp [21] and RcppArmadillo [22] are used to compute matrix decompositions in C++.

4 Applications

In this section we evaluate the STGLMM in two ways. In section 4.1, we present a simulation study were we compare coverage and point estimates of parameters between the naive GLM and the low rank STGLMM. We find that the low rank STGLMM has superior coverage, especially for the intercept parameters, and has more accurate point estimates for low and moderate levels of spatial and temporal dependence. In section 4.2, we compare the low rank STGLMM to the GLM for population modeling of Carolina Wrens in the United States between 1990 and 2010. We find that while neither model provides an accurate fit to the data, likely due to quirks in the data discussed later, the low rank STGLMM has a better predictive accuracy for the data set. Our analysis only explores outcomes that are recorded as counts, and the efficacy of the model may differ for other data that could be modeled using the STGLMM. Several packages were used to produce and visualize these results. [23, 24, 25, 26, 27, 28, 29, 30]

4.1 A Simulation Study

To assess the quality of the low rank STGLMM, we simulated data from the spatiotemporal Gaussian process with independent and additive spatial and temporal random effects. The spatial and temporal random effects follow the same covariance structures as are used in fitting the low rank STGLMM, Matérn for spatial effects and squared exponential for temporal effects. We let the rank of the approximation of the spatial correlation be 50, and the rank of the approximation of the temporal correlation be 10. We find that for these values MCMC issues are minimized, and the BIC is small for generalized linear models using the decorrelated spatial and temporal effects relative to other values that minimized MCMC issues. We also fit the same full rank STGLMM, so the approximations are exact, the understand the speedup of the low rank STGLMM. For these data sets, the low rank model

fit in approximately half the amount of time as the full rank STGLMM. For each parameter set, we simulate 100 data sets, where each one consists of 5000 location time pairs. Outcomes are Poisson, and each outcome has 3 covariates, $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}$, which are independently drawn from $N(0, 0.05)$. There are 100 locations, and each location is observed at 50 different times spaced uniformly between 0 and 1. Each location s is a 2-dimensional vector such that $s = (s_1, s_2)$ and $s_1, s_2 \overset{iid}{\sim} Unif(0, 1)$. Each location is observed at the same times, with the difference between each consecutive time being $\frac{1}{50}$. We simulate for 3 different parameter sets. The values of the parameters from which we simulate are given in Table 1. The values of ϕ and ψ provide 3 situations from which to simulate. When $\phi = \psi = 0.2$, we have little correlation between observations as the scale of distances in time and space is large relative to the parameter values. When $\phi = \psi = 1.0$, we have extremely high correlations between observations as the scale of distances is approximately equal to the parameters. $\phi = \psi = 0.5$ represents a situation where there is relatively moderate correlation between neighboring observations.

For each data set, we fit the low rank STGLMM as described earlier in Section 3. We also fit a GLM that assumes independence of observations to each data set to understand what we gain by accounting for spatial and temporal dependence. For the GLM, we assume the outcomes follow a Negative Binomial distribution as assuming a Poisson distribution performs exceedingly poorly. This is likely because the Poisson likelihood is overly restrictive given the additional structure provided by the spatial and temporal effects, while the Negative Binomial likelihood is better able to account for this structure. Each GLM is estimated using `rstanarm` with 2 chains of 5000 iterations each, the first 2500 of which are thrown out for the burn-in period.`rstanarm` uses Hamiltonian Monte Carlo sampling as opposed to Gibbs sampling with Metropolis Hastings steps used for sampling the low rank STGLMM model, but the interpretation of outcomes is identical. For the low rank STGLMM, we run 2 chains of 30000 iterations each, the first 15000 of which are thrown out for the burn-in period. We fit these models for each of the 100 simulated data sets for each parameter set, and record the 50% and 90% credible intervals, as well as the median estimate for each parameter, which we use as our point estimate.

Table 1: Parameter Sets for Simulated Data

Parameter Set	β_0	β_1	β_2	β_3	σ_w^2	θ^2	ϕ	ψ
1	1	1	1	-1	1	1	0.2	0.2
2	1	1	1	-1	1	1	0.5	0.5
3	1	1	1	-1	1	1	1.0	1.0

Figure 1: Distribution of point estimates for fixed effects across 100 simulated data sets with $\phi = \psi = 0.2$. Black lines show the true value.

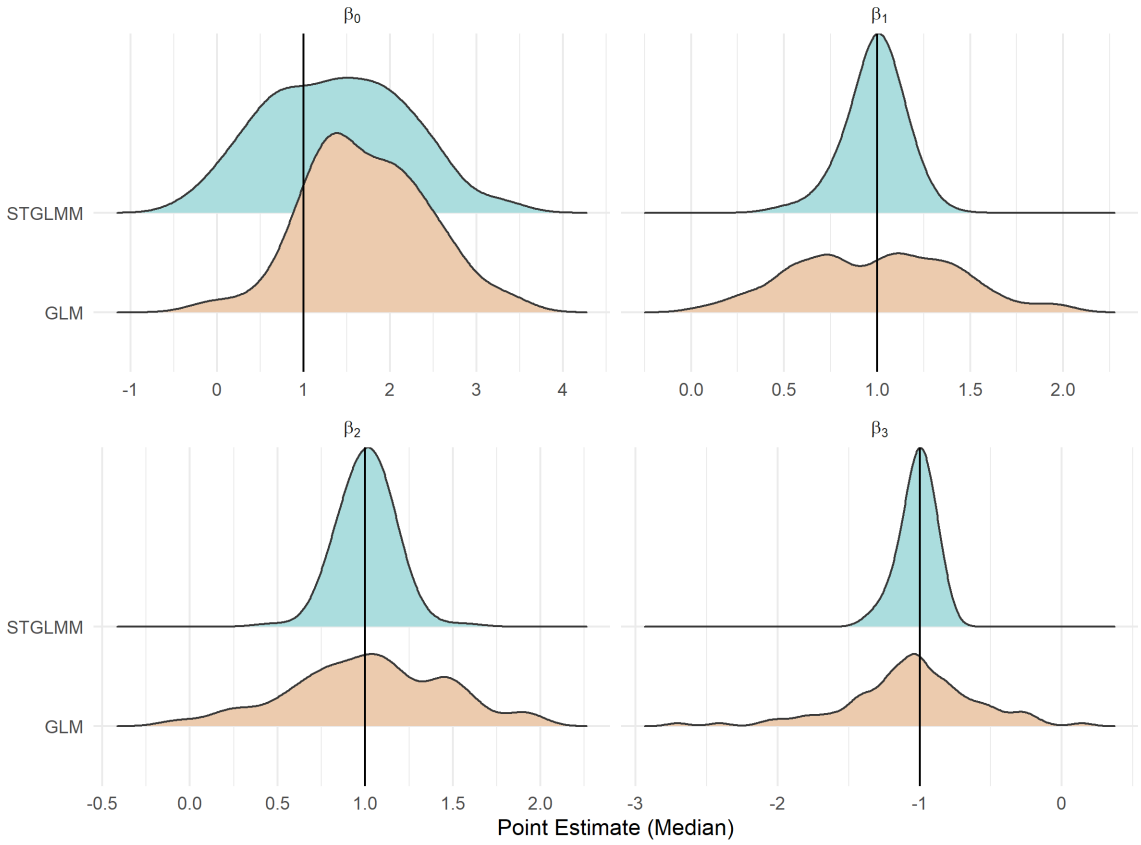


Figure 2: Distribution of point estimates for fixed effects across 100 simulated data sets with $\phi = \psi = .5$. Black lines show the true value.

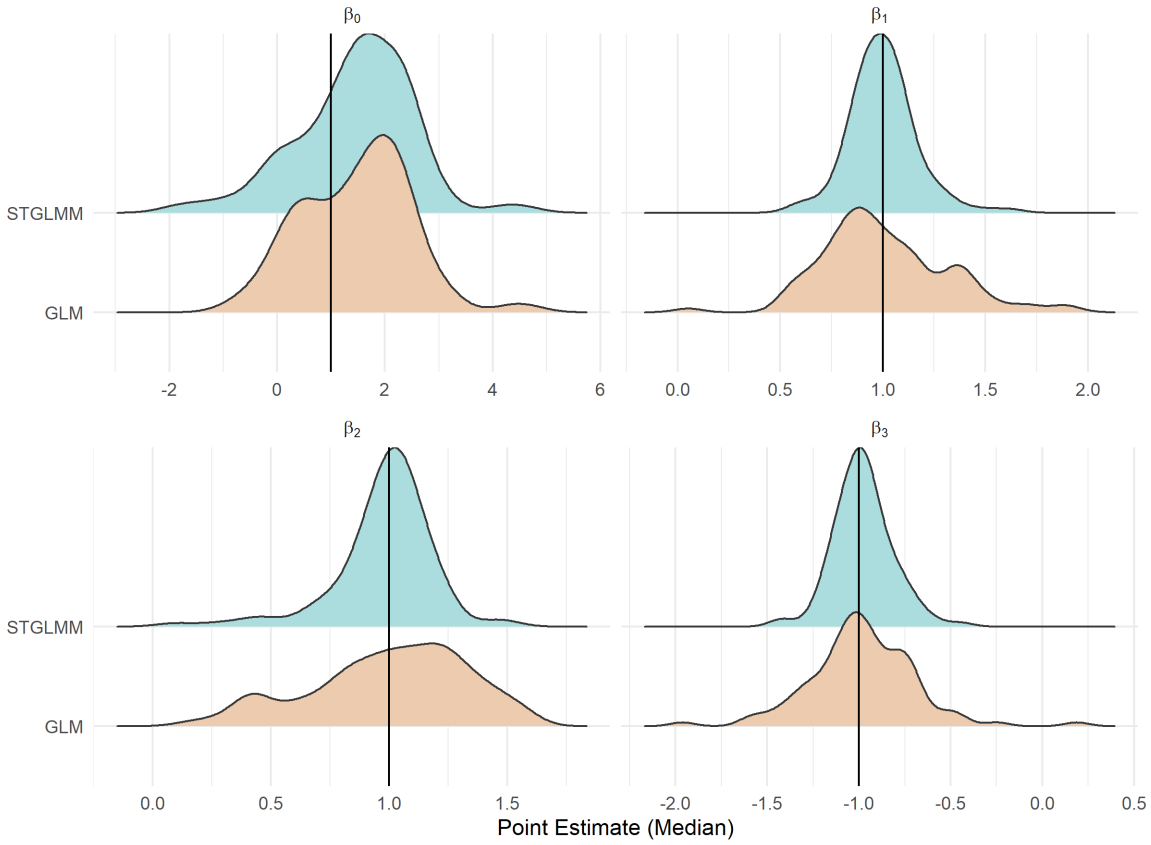
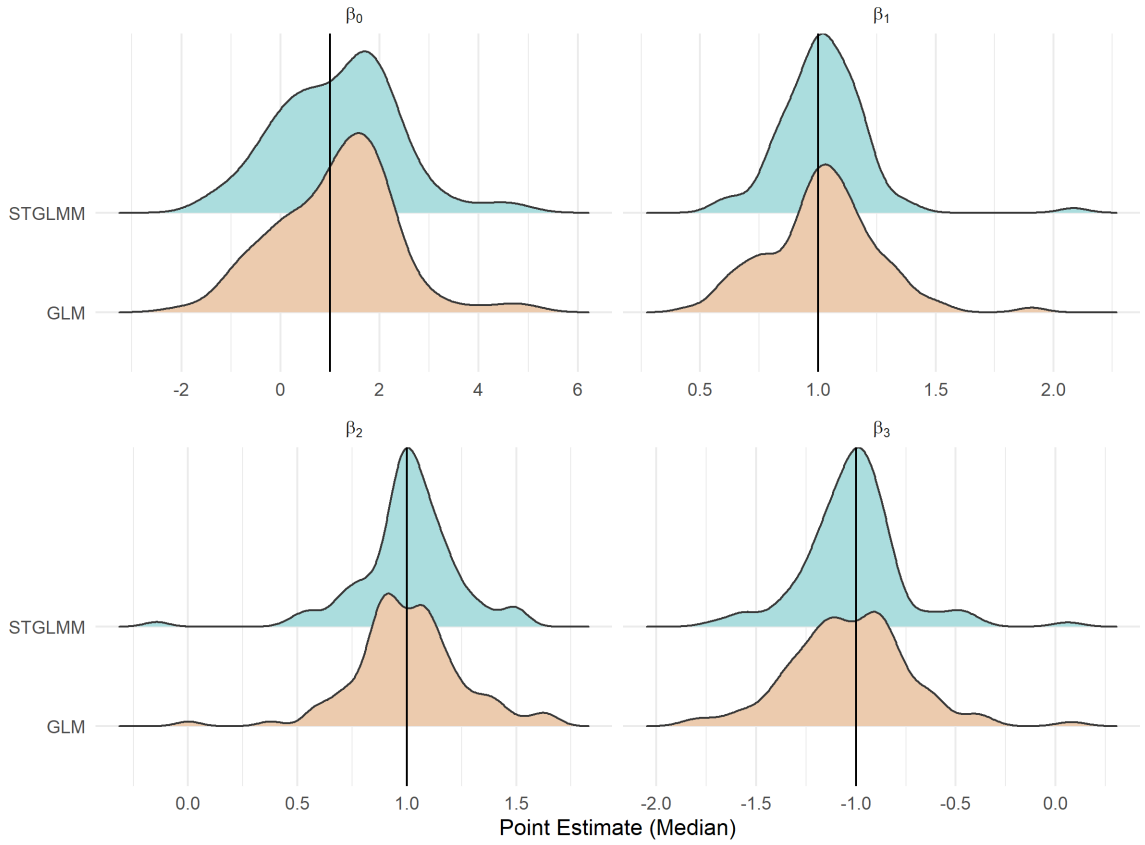


Figure 3: Distribution of point estimates for fixed effects across 100 simulated data sets with $\phi = \psi = 1$. Black lines show the true value.



Figures 1, 2, and 3 show the distribution of point estimates for the fixed effects coefficients β_0, β_1, β_2, β_3 under the 3 sets of correlation parameter values. The estimates of both the lwo rank STGLMM and the GLM appear to be unbiased estimates for the true values of fixed effect coefficients other than the intercept β_0. However the low rank STGLMM's estimates have a much lower standard error compared to the GLM. Interestingly, the GLM actually performs better compared to itself with respect to its point estimates variance as ϕ and ψ increase, corresponding to higher spatial and temporal correlation. Both models provide poor estimates of the intercept β_0 as there is increased variability in the intercept due to the random effects. In the low rank STGLMM, this is likely due to the random effects being correlated with the intercepts. Examining the chains for an example simulated dataset shows that often the chains of some of the synthetic random effects δ and γ are highly correlated with the intercept chains, leading to slower mixing. This could be contributing to the relatively poor estimate of the intercept for the low rank STGLMM. For the GLM the error in the intercept likely occurs because the true model is not a GLM with Negative Binomial outcomes.

Figure 4: Distribution of point estimates for random effect hyperparameters across 100 simulated data sets with $\phi = \psi = 0.2$. Black lines show the true value.

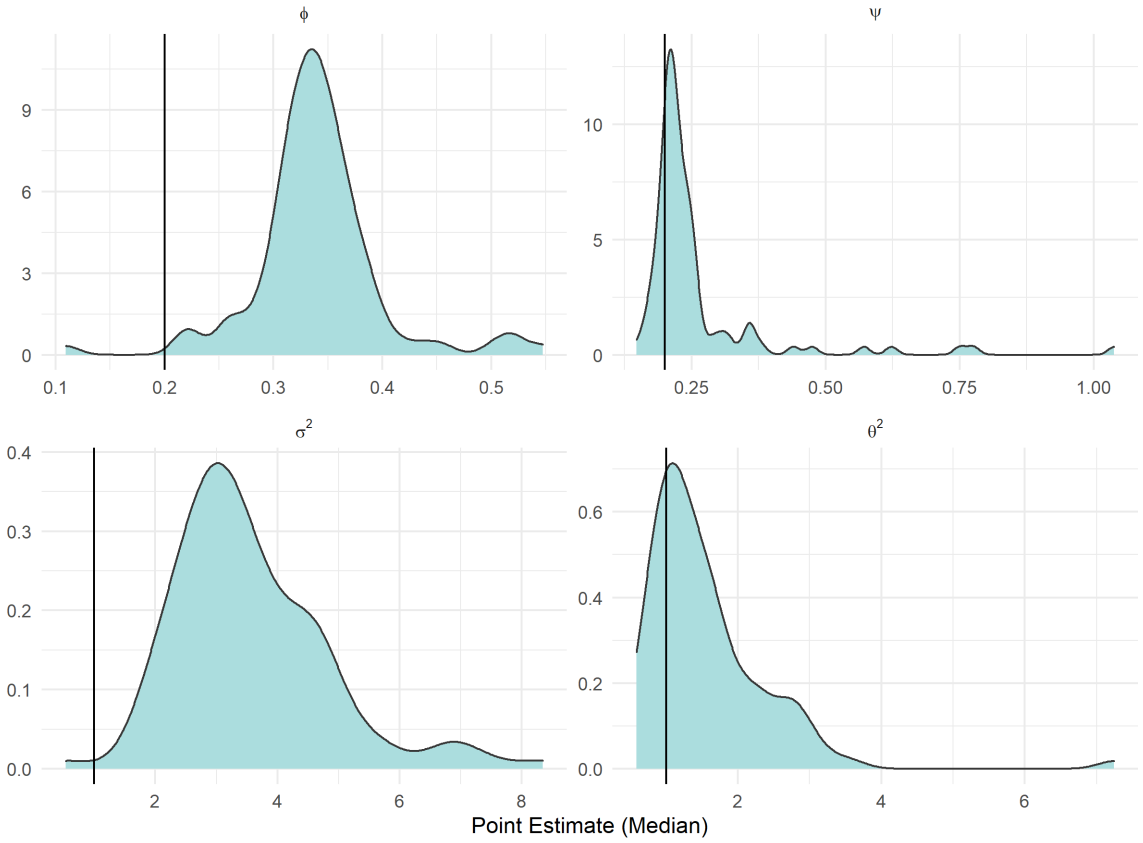


Figure 5: Distribution of point estimates for random effect hyperparameters across 100 simulated data sets with $\phi = \psi = 0.5$. Black lines show the true value.

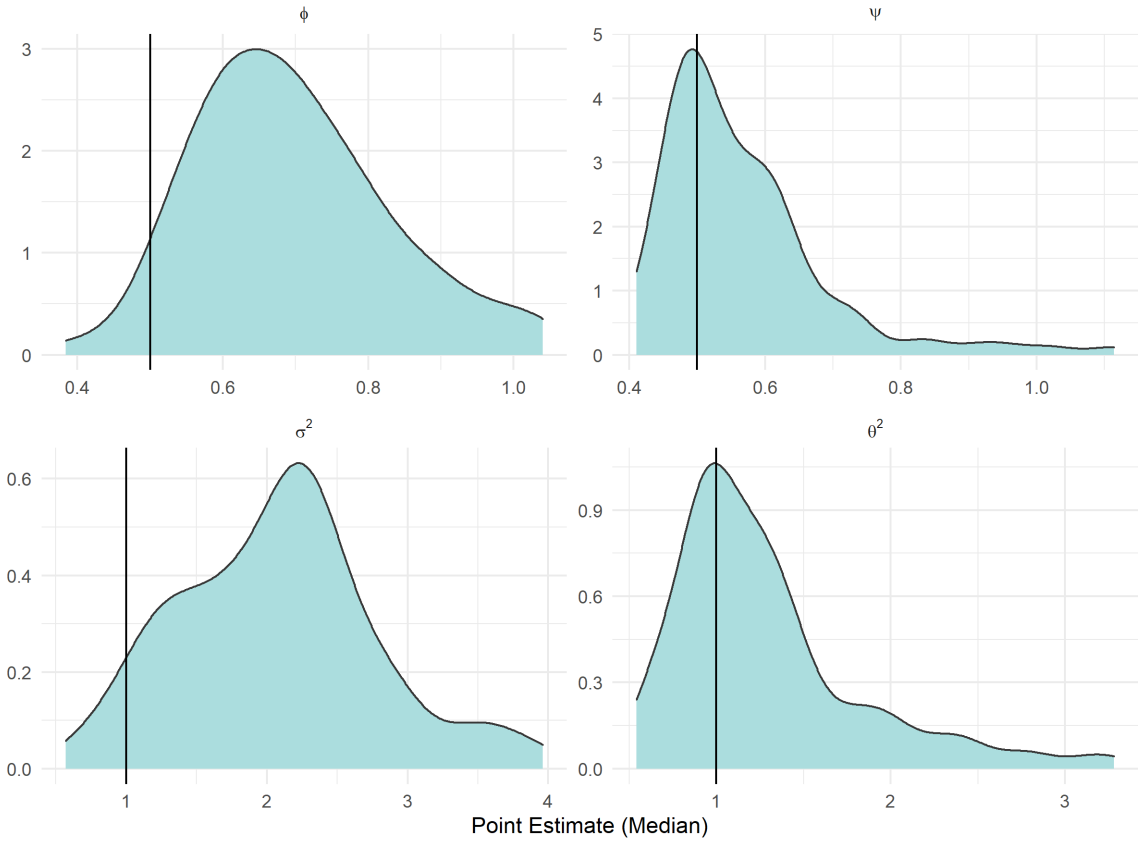
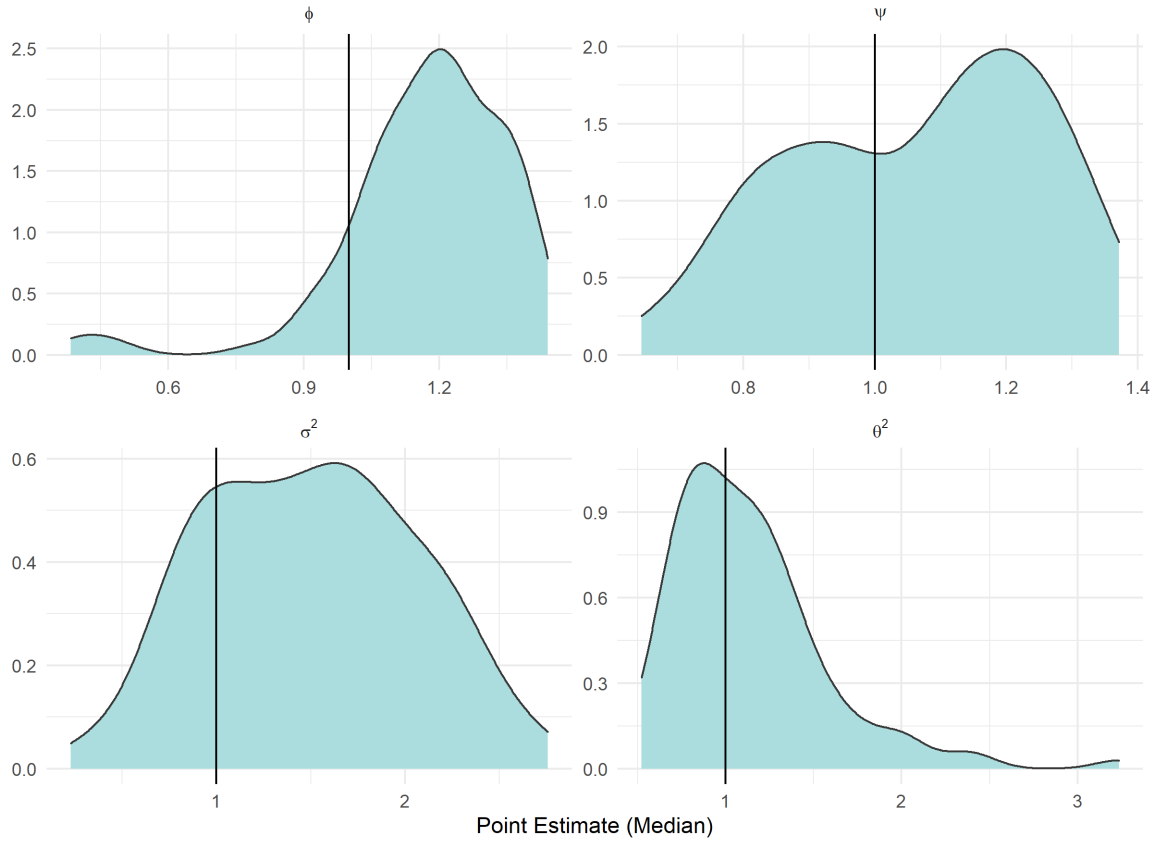


Figure 6: Distribution of point estimates for random effect hyperparameters across 100 simulated data sets with $\phi = \psi = 1$. Black lines show the true value.



Figures 4, 5, and 6 show the distributions of the point estimates of the model hyperparameters controlling the random effects of the low rank STGLMM. The low rank STGLMM consistently overestimates the amount of spatial covariance between locations. This is to be expected, as the low-rank of the approximated spatial effects means it is only able to represent the most dominant components of the true spatial correlation structure. In order to account for the missing spatial correlation from less dominant components of the correlation structure, it inflates the estimates of ϕ and σ_w^2. This does not occur in the temporal hyperparameters ψ and θ^2. This is likely the case because the true correlation structure is much less complex, as times are evenly spaced between 0 and 1 instead of randomly located on a unit square. The Markov chains used to estimate σ_w^2 when $\phi = 0.2$ also show a consistent and accelerating trend towards large values as the length of the chain increases, meaning the posterior for σ_w^2 is likely unreliable. There are also issues in some of the random effects estimates, however the most prominent, separation of the chains, can be explained by noting that they tend to converge towards the same absolute value. In this case, it's likely that the sign of the singular values has flipped between chains, causing them to converge to opposite values. The Markov chains for one of the simulated data sets for $\phi = \psi = 0.2$ can be viewed in Appendix A.3 for reference.

Because the GLM ignores correlation between observations, we would expect it to underestimate the standard error of its parameter estimates. One way to understand how accurate a model's parameter estimates are is to examine their coverage, or how often their credible intervals of parameters contain the true values of those parameters. If the models perfectly captured the truth, we would expect the true parameters to be in the 50% credible intervals in 50 of the 100 simulations, and in the 90%

Table 2: Proportion of times true value is within 50% and 90% credible intervals across 100 simulated datasets

	Model	CI	β_0	β_1	β_2	β_3	σ_w^2	θ^2	ϕ	ψ
$\phi = \psi = 0.2$	GLM	50%	0.00	0.35	0.43	0.51	NA	NA	NA	NA
		90%	0.02	0.82	0.83	0.79	NA	NA	NA	NA
	STGLMM	50%	0.27	0.62	0.52	0.67	0.04	0.48	0.04	0.31
		90%	0.49	0.90	0.95	0.96	0.07	0.84	0.05	0.57
$\phi = \psi = 0.5$	GLM	50%	0.00	0.49	0.38	0.53	NA	NA	NA	NA
		90%	0.00	0.91	0.86	0.91	NA	NA	NA	NA
	STGLMM	50%	0.22	0.55	0.48	0.49	0.13	0.43	0.15	0.68
		90%	0.35	0.90	0.87	0.90	0.21	0.74	0.31	0.90
$\phi = \psi = 1.0$	GLM	50%	0.00	0.58	0.58	0.41	NA	NA	NA	NA
		90%	0.00	0.90	0.94	0.88	NA	NA	NA	NA
	STGLMM	50%	0.28	0.56	0.50	0.53	0.33	0.39	0.34	0.71
		90%	0.46	0.95	0.92	0.90	0.57	0.83	0.60	0.98

credible intervals in 90 of the 100 simulations. Looking at Table 2, we see that in general the low rank STGLMM has more accurate coverage compared to the GLM. This difference is especially stark when comparing the intercepts. When $\phi = \psi = 0.2$, the true value is in the 50% credible interval in 0 of the simulations for the GLM, and it is in the 90% credible interval in only 2 of the simulations. These credible intervals can be viewed in Appendix A.2. While the low rank STGLMM still underestimates the standard error of its parameter estimates, it far outperforms the GLM, capturing the true value in its 50% and 90% credible intervals for the intercept in 27 and 49 of the simulations respectively. The GLM performs significantly better when estimating the fixed effects other than the intercept. This is because the random effects are random intercepts and not random slopes, so while each data point provides less information about the intercept than expected, they provide exactly as much information about the other coefficients as expected. Across both the GLM and the low rank STGLMM, coverage improves as ϕ and ψ increase. Mirroring the results from the point estimates of the hyperparamters, the low rank STGLMM has superior coverage for the temporal parameters than the spatial parameters.

4.2 Bird Population Counts

One commonly used spatiotemporal data set comes from the Breeding Bird Survey, which is a collaboration between the United States Geological Service and the Canadian Wildlife Service to monitor the status of North American bird populations [14]. For this survey, researchers drive along approximately 25 mile routes, stopping numerous times and counting the number of birds of each species observed at that stop. The survey began in 1966 and has steadily expanded to include over 4100 routes and 420 different bird species. For the purposes of this paper, we consider observations of the Carolina Wren at 171 routes between 1990 and 2010. These routes were chosen as they have complete data for this time span.

Figure 7: Observed Carolina Wrens at Selected Sites for 1990 Survey

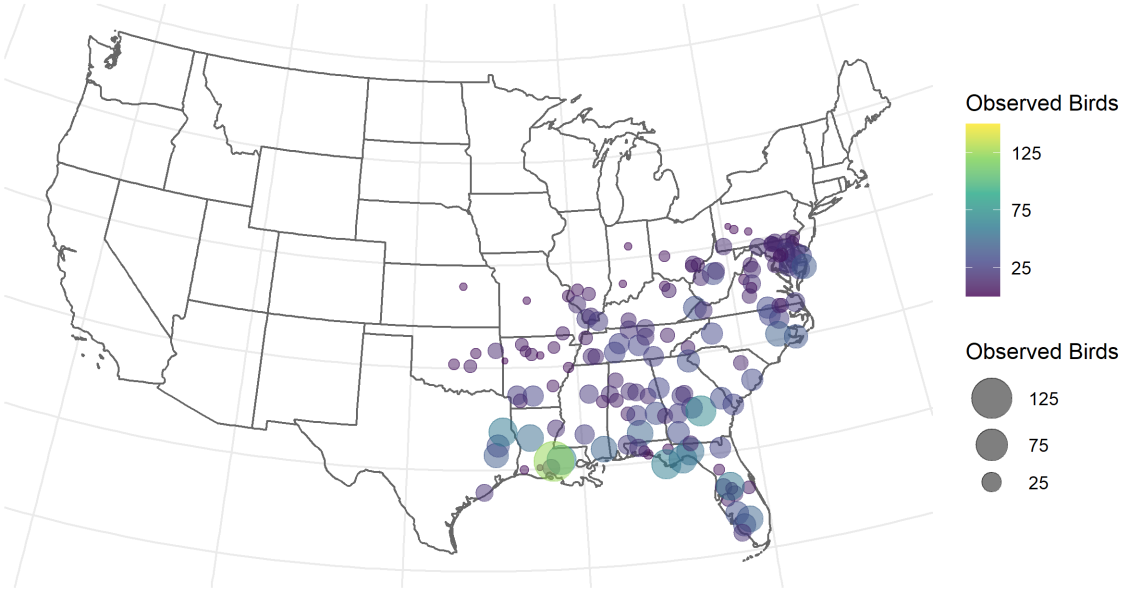
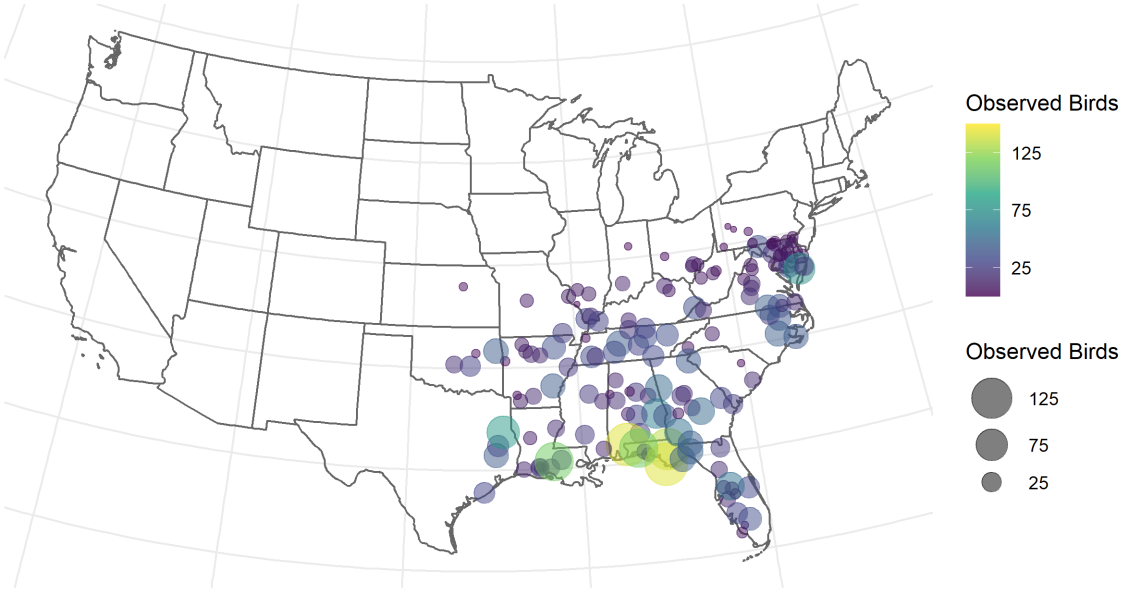


Figure 7 shows the locations of routes as well as the number of Carolina Wrens observed at each route during the 1990 survey. As can be seen, Carolina Wrens live primarily in the American South-East. Notice how there appears to be a spatial pattern to these data. For example, we tend to see larger numbers of birds observed in and around Northern Florida and Louisiana, and fewer birds further West in the area around Arkansas. This spatial pattern suggests that the data are spatially correlated, meaning methods explicitly taking spatial correlation into account may be warranted.

Figure 8: Observed Carolina Wrens at Selected Sites for 1991 Survey



In Figure 8 we see data for the 1991 survey at the same routes as the 1990 survey. While the data have changed, broad patterns remain similar, suggesting there may be temporal correlation as well. If both spatial and temporal dependence are present, it is appropriate to use a spatiotemporal model to capture the correlation structure.

4.2.1 GLM

For comparison to a model that doesn't take into account spatial or temporal dependence, we'll fit a generalized linear model to the data and compare it to the low rank spatiotemporal generalized linear mixed model outlined in Section 3. We will use the observed number of birds on a route in that year as the outcome $\boldsymbol{Y}(s,t)$ for a location s and year t, with the stratum as a predictor. The stratum represents type of ecosystem the route is in as defined by ecologists at the Breeding Bird Survey. Examples include Upper Coastal Plain and Coastal Flatwoods, the later of which we show coefficient estimates for. We'll use a Poisson likelihood for the model to be comparable to the low rank STGLMM with a log link function $g(\cdot) = log(\cdot)$.

Here, we use $\boldsymbol{\beta}_1$ to represent the set of coefficients representing the different possible values of stratum. There were 18 different observed strata, so $\boldsymbol{\beta}_1$ actually represents 17 different coefficients. A row $\boldsymbol{x}(s,t)$ of our design matrix $\boldsymbol{x}$ consists of dummy variables representing the stratum the observation was located in. The mean model we fit is

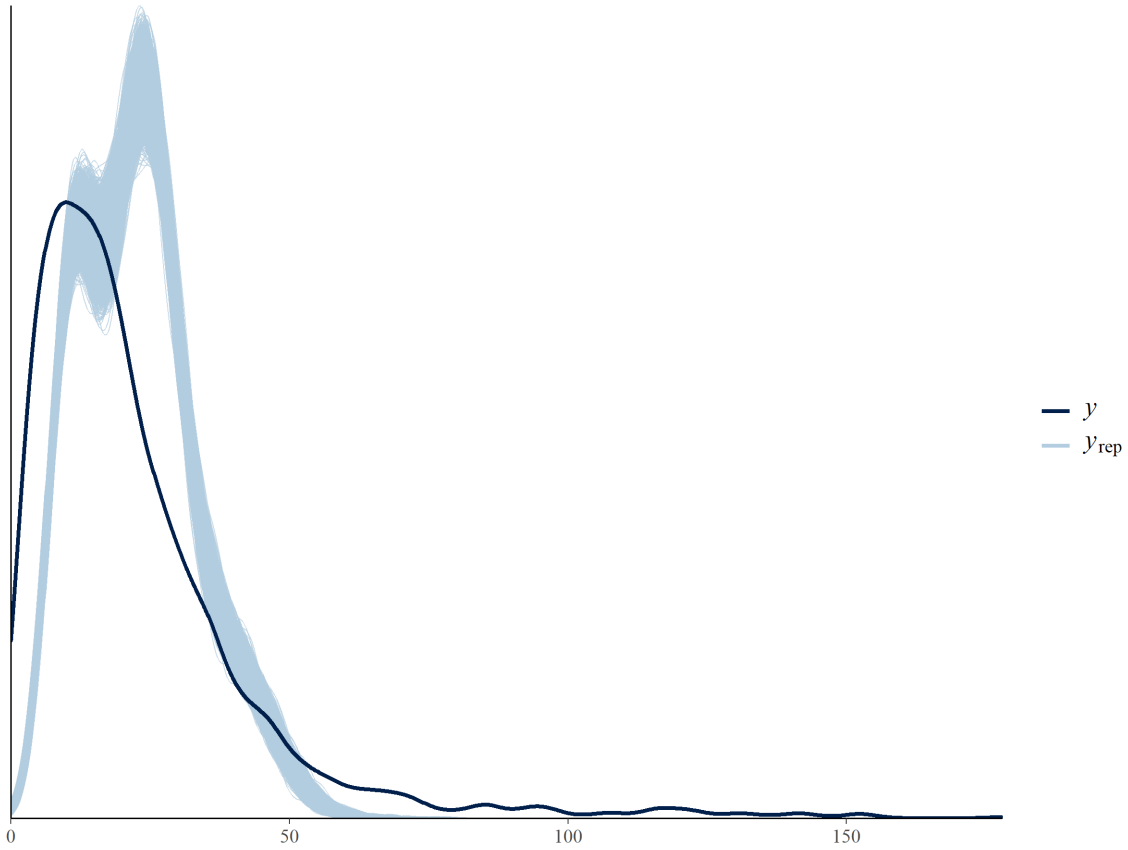
$$log(E[Y(s,t)|\beta_0, \boldsymbol{\beta_1}]) = \beta_0 + \boldsymbol{x}(s,t)^T\boldsymbol{\beta}_1$$

$$\beta_0, \boldsymbol{\beta}_1 \overset{iid}{\sim} N(0, 10^2)$$

The GLM was fit using the `rstanarm` package [31] for Bayesian modeling. We ran 2 chains with 10000 iterations each, and removed the first 5000 iterations as the burnout period.

Figure 9 shows the posterior predictions across each location and time. The predictions are far from the observed value, suggesting this model doesn't fully capture the dynamics surrounding bird populations in our data. There is also relatively little variance in predicted estimates, suggesting that this model may be overconfident in its predictions. This makes sense given spatiotemporal correlation. Recall that assuming independence when the data are correlated has the effect of overestimating the amount of data the model has access to, deflating the variance estimates.

Figure 9: Posterior Predictions for the Generalized Linear Model. True distribution shown in dark blue, predicted distributions shown in light blue.



4.2.2 STGLMM

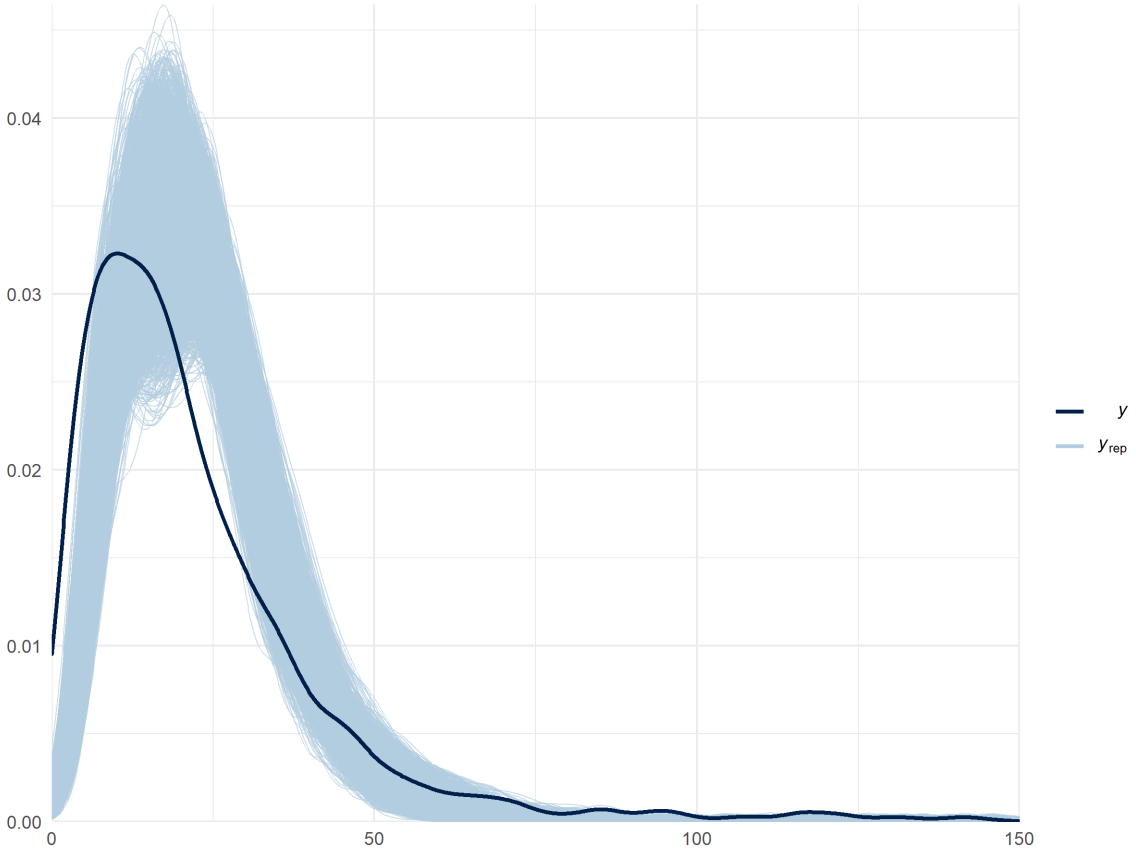
Now we consider the low rank STGLMM defined in section 3. Again we'll use a log link function, but this time we will include our spatial and temporal random effects. In full, the low rank STGLMM can be written

$$log(E[\text{Observed Birds}|\beta_0, \boldsymbol{\beta_1}, \eta(s,t)]) = \beta_0 + \boldsymbol{x}(s,t)^T\boldsymbol{\beta}_1 + \eta(s,t)$$
$$\eta(s,t) = \Big[(\boldsymbol{U}\boldsymbol{K}^{1/2})\boldsymbol{\gamma}\Big]_t + \Big[(\boldsymbol{V}\boldsymbol{\Lambda}^{1/2})\boldsymbol{\delta}\Big]_s$$
$$\boldsymbol{\delta}|\phi, \sigma_w^2 \dot{\sim} N(0, \sigma_w^2\boldsymbol{I})$$

$$\boldsymbol{\gamma}|\psi, \theta^2 \dot{\sim} N(0, \theta^2 \boldsymbol{I})$$
$$\beta_0, \boldsymbol{\beta}_1 \sim N(0, 10^2 \boldsymbol{I})$$
$$\phi \sim Unif(0.01, 1.5)$$
$$\psi \sim Unif(0.01, 1.5)$$
$$\sigma_w^2 \sim Gamma(2, 2)$$
$$\theta^2 \sim Gamma(2, 2)$$

We adopt Uniform priors for ϕ and ψ and Gamma priors for σ_w^2 and θ^2 following the recommendations of Banerjee et al. [5]. We again ran 2 chains with 10000 iterations each, and removed the first 5000 iterations as the burnout period. The data were identical to those used to fit the generalized linear model. Figure 10 shows the posterior predictions across each location and time.

Figure 10: Posterior Predictions for the Spatiotemporal Generalized Linear Mixed Model. True distribution shown in dark blue, predicted distributions shown in light blue.



Note how the low rank STGLMM produces a much wider spread of predictions than the GLM. This reflects the higher variance in parameter estimates we would expect when data are correlated. However, while the low rank STGLMM has a better coverage of the true outcomes than the GLM, both are biased. This is likely due in part to the structure of the data. There are no routes in the dataset for which there were 0 observations of birds on the route that year. It may be the case that some routes aren't missing each year because they weren't surveyed but rather because 0 birds were observed and so the route was excluded from the data. If this is the case, it might be

more appropriate to use a model where 0 values are disallowed, such as a zero-truncated Poisson distribution as opposed to a standard poisson distribution. However, this is beyond the scope of this paper.

Figure 11: Posterior estimates for the Coastal Flatwoods stratum coefficient. The shaded region is 95% central posterior density, the dark blue line is the coefficient point estimate using the median of the draws.

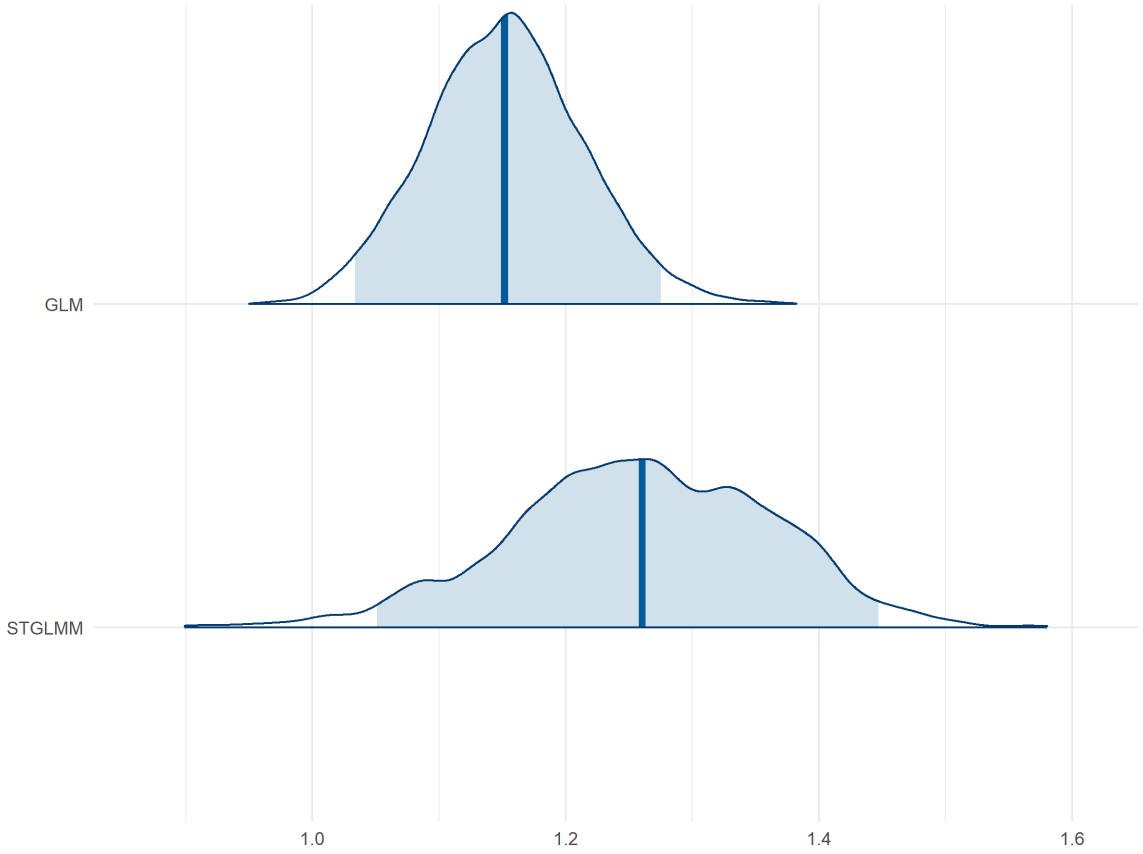


Figure 11 displays the posterior density for the stratum representing the coefficient for the Coastal Flatwoods ecosystem. There is a much larger variance in plausible estimates for the parameter in the low rank STGLMM compared to the GLM. Also note how the low rank STGLMM has a greater deviation from the expected Normal distribution. This is because the low rank STGLMM sampler is more slowly mixing, leading to a smaller effective sample size. While both of these models took similar amounts of time to estimate, the quality of posterior estimates from the generalized linear model are superior. Running the low rank STGLMM for a larger number of iterations may minimize this issue.

5 Discussion and Future Work

As expected, the proposed low rank STGLMM has much more accurate estimates of coverage for fixed effect parameters compared to the GLM when data are spatially and temporally correlated. Somewhat surprisingly, the low rank STGLMM's point estimates of these parameters are more accurate when spatial and temporal correlation is low to moderate. These gains also come at a relatively reduced computational cost compared to the full rank model, taking approximately half as

long to estimate the parameters compared to the full model. When data are larger, we can expect these computational gains to increase. While the model takes significantly longer than the naive GLM approach, the improved accuracy in credible intervals is likely worth the additional computational burden in most cases. The low rank STGLMM also performs well on the Carolina Wren population data compared to the GLM providing a better predictive accuracy even under mispecification of the likelihood model.

However, the low rank STGLMM is not without limitations. In the simulation study, the low rank STGLMM consistently overestimates parameters defining the spatial covariance structure. This is a well known problem in low rank models for spatial covariance structures, as they only capture the largest components of spatial covariance, and occasionally magnify the components they capture to account for loss of covariance structure from smaller components. As can be seen from the coverage estimates of the intercept in table 2 however, the low rank STGLMM still underestimates the true variance caused by correlated outcome. An important area of future work for this model would be developing bias corrections to these estimates of spatial covariance. Finley et al. [9] propose such a bias correction for the predictive process model, however their bias correction limits the model to have Gaussian outcomes [5]. It may be that an analogous bias correction for the low rank STGLMM would also limit it to model Gaussian outcomes.

Another important limitation to the low rank STGLMM is the issue of spatial and temporal confounding. This occurs when the fixed effect parameters are correlated with the spatial or temporal random effects. An example of this comes in the Carolina Wren data, where it is highly likely that the stratum, which represent ecosystems, are correlated with the spatial random effects. This is a well studied problem which can lead to biased or deflated variance estimates of parameters of interest [15, 16]. Several methods have been developed in an attempt to circumvent spatial and temporal confounding. Notably, Guan and Haran [1] follow the approach used by Hughes and Haran [10] as well as Hanks et al. [16] called restricted spatial regression. Here, we restrict the spatial and temporal random effects to be perpendicular to the fixed effects. However, after years of use, Khan and Calder [17] found that this approach has counterintuitive effects that cause it to provide worse estimates than the naive non-spatial model in many cases. For this reason, we chose not to include restricted spatial regression in the low rank STGLMM. However, it still suffers from the issues of spatial and temporal confounding, and it may be prudent to attempt to incorporate other methods for addressing confounding, such as the work of Prates et al. [18].

Finally, it is important to note that this paper does not compare the low rank STGLMM to other computationally efficient models for spatiotemporal modeling. This was in part due to the fact that very few models have public implementations that work with non-Gaussian outcomes, and so comparison with newer approaches is infeasible. However, there are approaches, such as integrated nested Laplace approximations (INLA), that have shown promise in spatial and spatiotemporal applications [19]. This approach is an alternative to MCMC methods for Bayesian inference that is often more computationally efficient, meaning it may outperform methods based on MCMC for spatiotemporal modeling. An interesting future direction of work on the low rank STGLMM would be using INLA as opposed to MCMC in order to provide still faster model estimation for very large data. One advantage the low rank STGLMM does hold compared to virtually every comparable method is the easy of implementation. While the model's MCMC requires tuning, the model specification itself requires the user to choose only two parameters, the rank of the spatial approximation and the rank of the temporal approximation. When compared to require the user to specify basis functions to approximate the covariance structure or select a set of knots, this model becomes very attractive for users who are relatively inexperienced with spatiotemporal modeling.

References

[1] Yawen Guan and Murali Haran. A Computationally Efficient Projection-Based Approach for Spatial Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics*, 27(4):701–714, October 2018. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10618600.2018.1425625.

[2] Jonathan R. Bradley, Scott H. Holan, and Christopher K. Wikle. Multivariate spatio-temporal models for high-dimensional areal data with application to Longitudinal Employer-Household Dynamics. *Annals of Applied Statistics*, 9(4):1761–1791, December 2015. Publisher: Institute of Mathematical Statistics.

[3] United Church of Christ and Commission for Racial Justice. *Toxic wastes and race in the United States: a national report on the racial and socio-economic characteristics of communities with hazardous waste sites*. Public Data Access : Inquiries to the Commission, New York, N.Y., 1987. OCLC: 16985343.

[4] Robert D. Bullard, Paul Mohai, Robin Saha, and Beverly Wright. TOXIC WASTES AND RACE AT TWENTY: WHY RACE STILL MATTERS AFTER ALL OF THESE YEARS. *Environmental Law*, 38(2):371–411, 2008. Publisher: Temporary Publisher.

[5] Sudipto Banerjee, Bradley Carlin, and Alan Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, New York, 2 edition, September 2014.

[6] Hao Zhang. On Estimation and Prediction for Spatial Generalized Linear Mixed Models. *Biometrics*, 58(1):129–136, 2002. Publisher: [Wiley, International Biometric Society].

[7] Christopher K. Wikle, Andrew Zammit-Mangion, and Noel Cressie. *Spatio-Temporal Statistics with R*. Chapman and Hall/CRC, Boca Raton, 1 edition, February 2019.

[8] Sudipto Banerjee, Alan E. Gelfand, Andrew O. Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848, 2008. _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2008.00663.x.

[9] Andrew O. Finley, Huiyan Sang, Sudipto Banerjee, and Alan E. Gelfand. Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis*, 53(8):2873–2884, June 2009.

[10] John Hughes and Murali Haran. Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models. *arXiv:1011.6649 [math, stat]*, November 2010. arXiv: 1011.6649.

[11] Anjishnu Banerjee, David B. Dunson, and Surya T. Tokdar. Efficient Gaussian process regression for large datasets. *Biometrika*, 100(1):75–89, March 2013.

[12] Jonathan R. Bradley, Scott H. Holan, and Christopher K. Wikle. Computationally Efficient Multivariate Spatio-Temporal Models for High-Dimensional Count-Valued Data (with Discussion). *Bayesian Analysis*, 13(1):253–310, March 2018. Publisher: International Society for Bayesian Analysis.

[13] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2008.00700.x.

[14] K.L Pardieck, D.J. Ziolkowski Jr., M. Lutmerding, V.I. Aponte, and M-A.R. Hudson. *North American Breeding Bird Survey Dataset 1966 - 2019*. U.S. Geological Survey, 2020.

[15] James S. Hodges and Brian J. Reich. Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love. *The American Statistician*, 64(4):325–334, 2010. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[16] Ephraim M. Hanks, Erin M. Schliep, Mevin B. Hooten, and Jennifer A. Hoeting. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4):243–254, 2015. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.2331.

[17] Kori Khan and Catherine A. Calder. Restricted Spatial Regression Methods: Implications for Inference. *Journal of the American Statistical Association*, pages 1–13, July 2020. Publisher: Taylor & Francis.

[18] Marcos Oliveira Prates, Renato Martins Assunção, and Erica Castilho Rodrigues. Alleviating Spatial Confounding for Areal Data Problems by Displacing the Geographical Centroids. *Bayesian Analysis*, 14(2):623–647, June 2019. Publisher: International Society for Bayesian Analysis.

[19] Marta Blangiardo, Michela Cameletti, Gianluca Baio, and Håvard Rue. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, 4:33–49, March 2013.

R Packages

[20] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[21] Dirk Eddelbuettel and Romain François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

[22] Dirk Eddelbuettel and Conrad Sanderson. Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, March 2014.

[23] Jonah Gabry and Tristan Mahr. bayesplot: Plotting for bayesian models, 2021. R package version 1.8.0.

[24] Lionel Henry and Hadley Wickham. *purrr: Functional Programming Tools*, 2020. R package version 0.3.4.

[25] Hadley Wickham and Jim Hester. *readr: Read Rectangular Text Data*, 2020. R package version 1.4.0.

[26] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2020. R package version 1.0.2.

[27] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[28] Edzer Pebesma. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446, 2018.

[29] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.

[30] Claus O. Wilke. *ggridges: Ridgeline Plots in 'ggplot2'*, 2021. R package version 0.5.3.

[31] Ben Goodrich, Jonah Gabry, Imad Ali, and Sam Brilleman. rstanarm: Bayesian applied regression modeling via Stan., 2020. R package version 2.21.1.

A Appendix

A.1 Conditional Likelihood Equations

A.2 Simulation Coverage Results

Here we show the 50% and 90% credible intervals for each of the fixed effects coefficients for every simulated dataset and parameter set. Notice how the low rank STGLMM in general has superior coverage, especially for the intercept. This is exactly what is to be expected when there is significant correlation between outcomes, as the GLM overestimates the amount of new information in each observation. For each plot, the inner boxes are the 50% credible intervals, and the whiskers are the 90% credible intervals. Small lines within boxes are the point estimates of the parameter. Horizontal black lines are the true value of the parameter.

A.2.1 Parameter Set 1: $\phi = \psi = 0.2$

Figure 12: Credible intervals for β_0 across simulations for each model when $\phi = \psi = 0.2$.

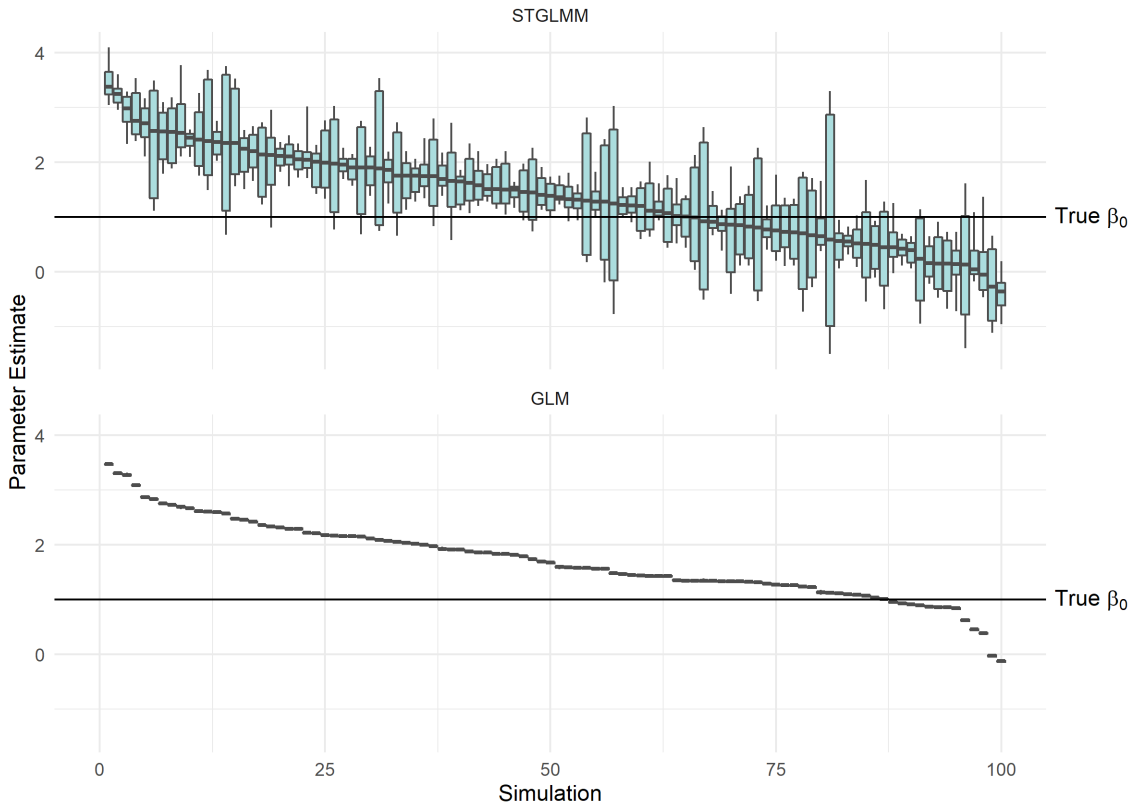


Figure 13: Credible intervals for β_1 across simulations for each model when $\phi = \psi = 0.2$.

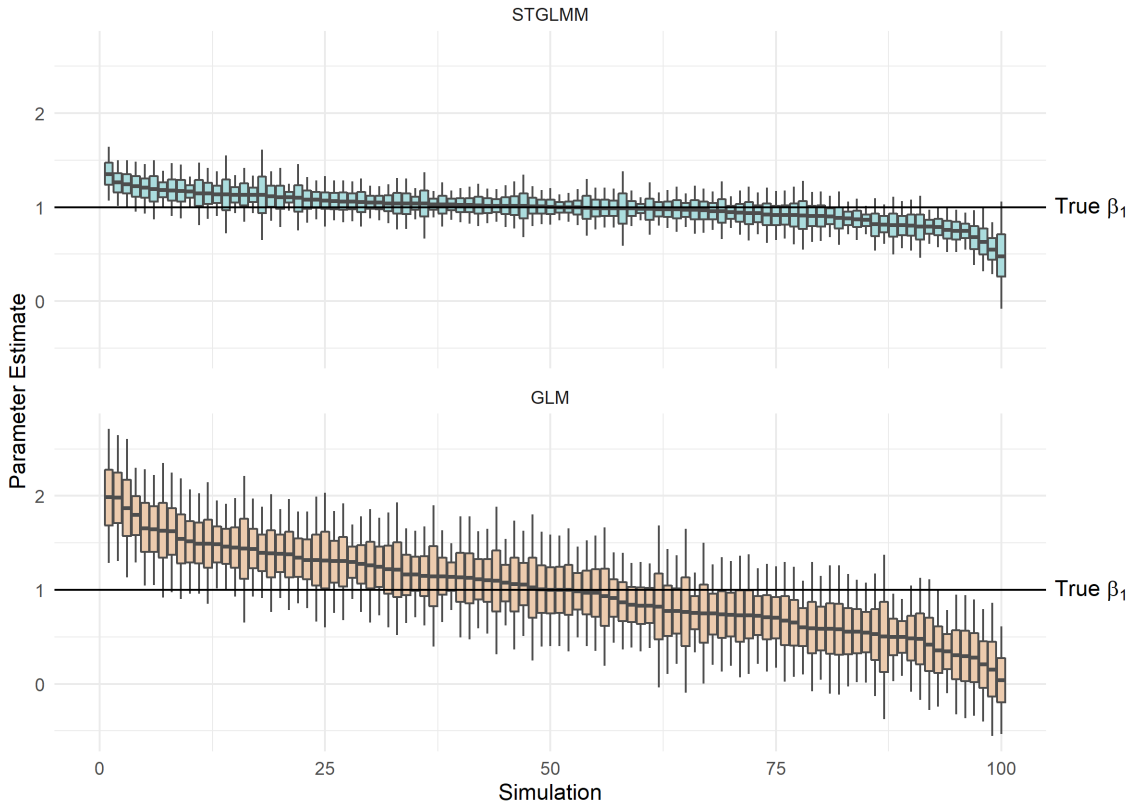


Figure 14: Credible intervals for β_2 across simulations for each model when $\phi = \psi = 0.2$.

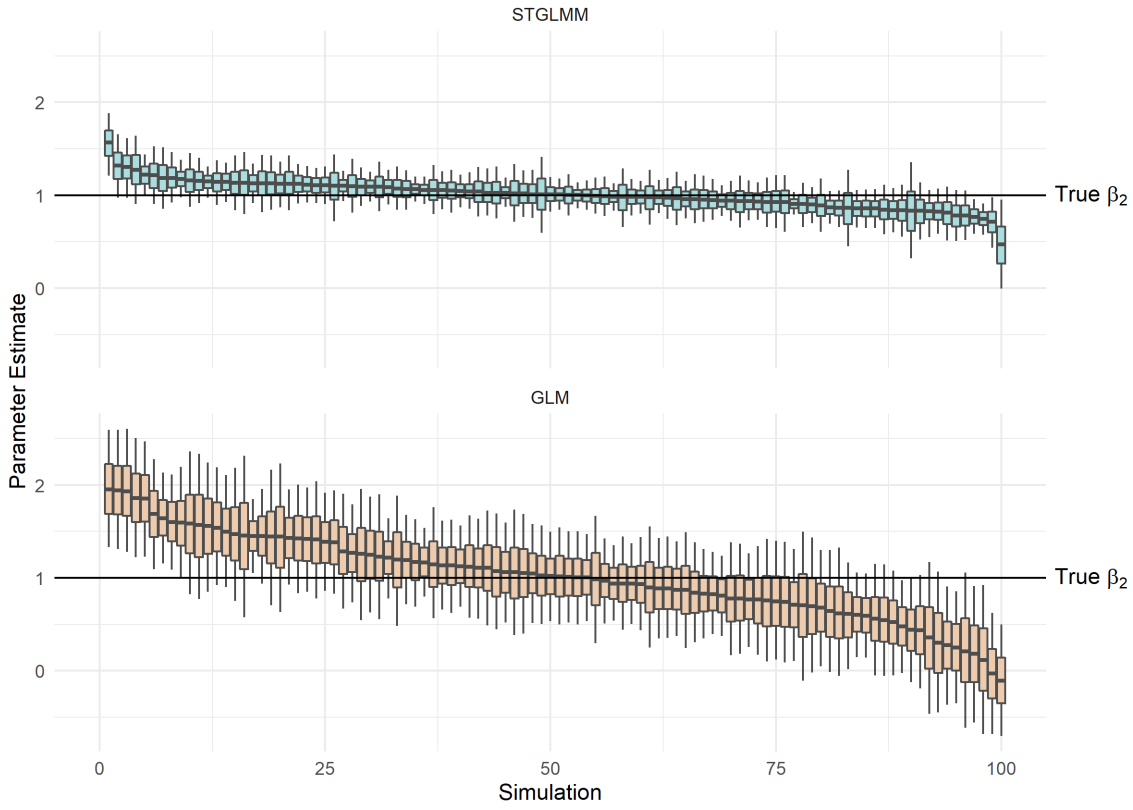
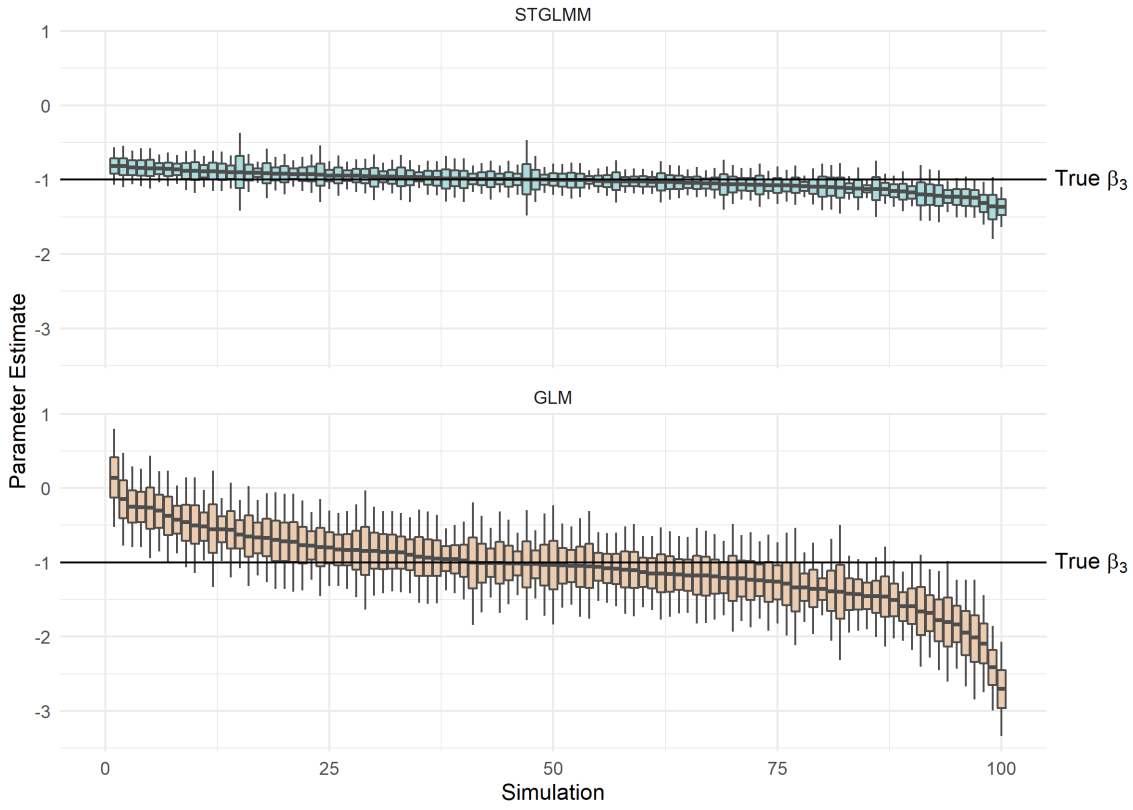


Figure 15: Credible intervals for β_3 across simulations for each model when $\phi = \psi = 0.2$.



A.2.2 Parameter Set 1: $\phi = \psi = 0.5$

Figure 16: Credible intervals for β_0 across simulations for each model when $\phi = \psi = 0.5$.

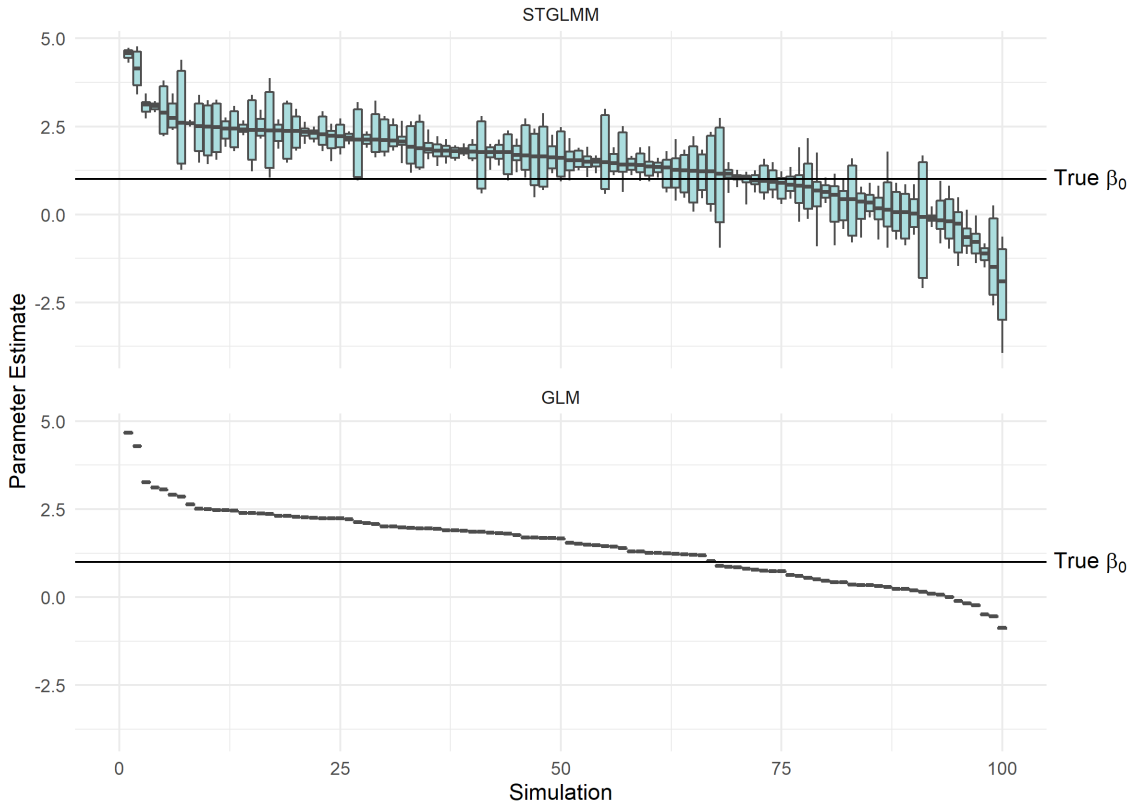


Figure 17: Credible intervals for β_1 across simulations for each model when $\phi = \psi = 0.5$.

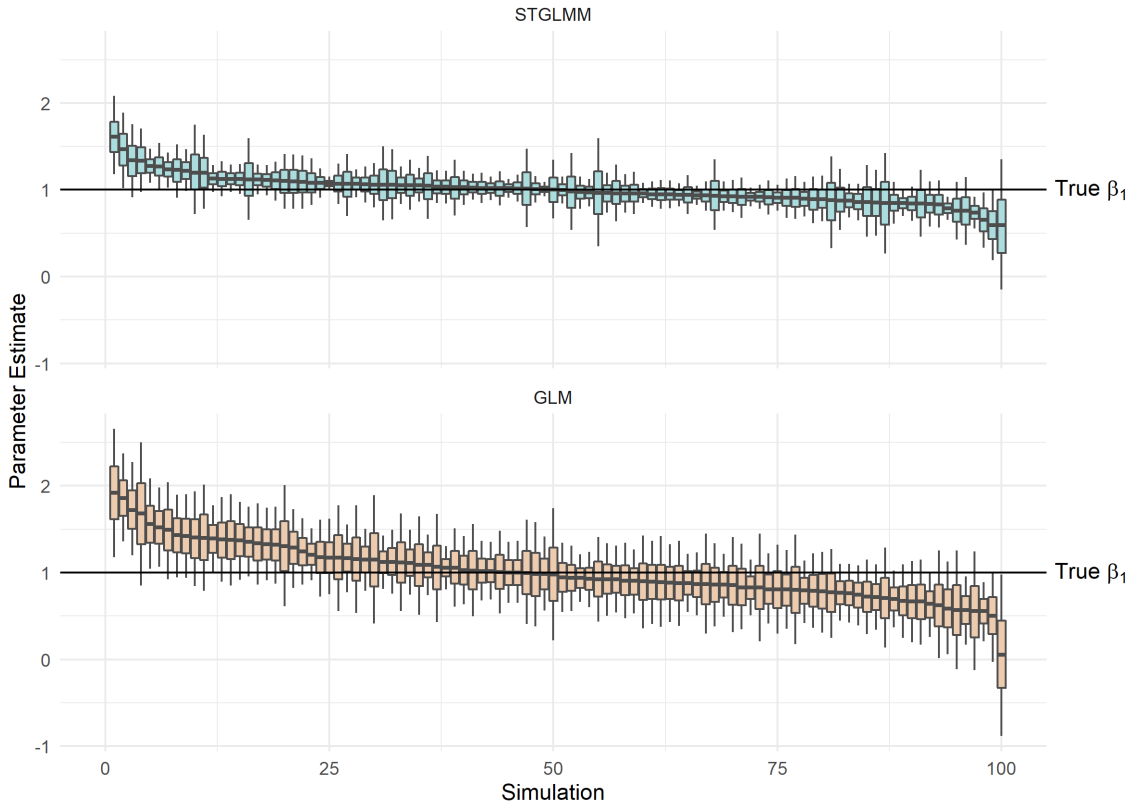


Figure 18: Credible intervals for β_2 across simulations for each model when $\phi = \psi = 0.5$.

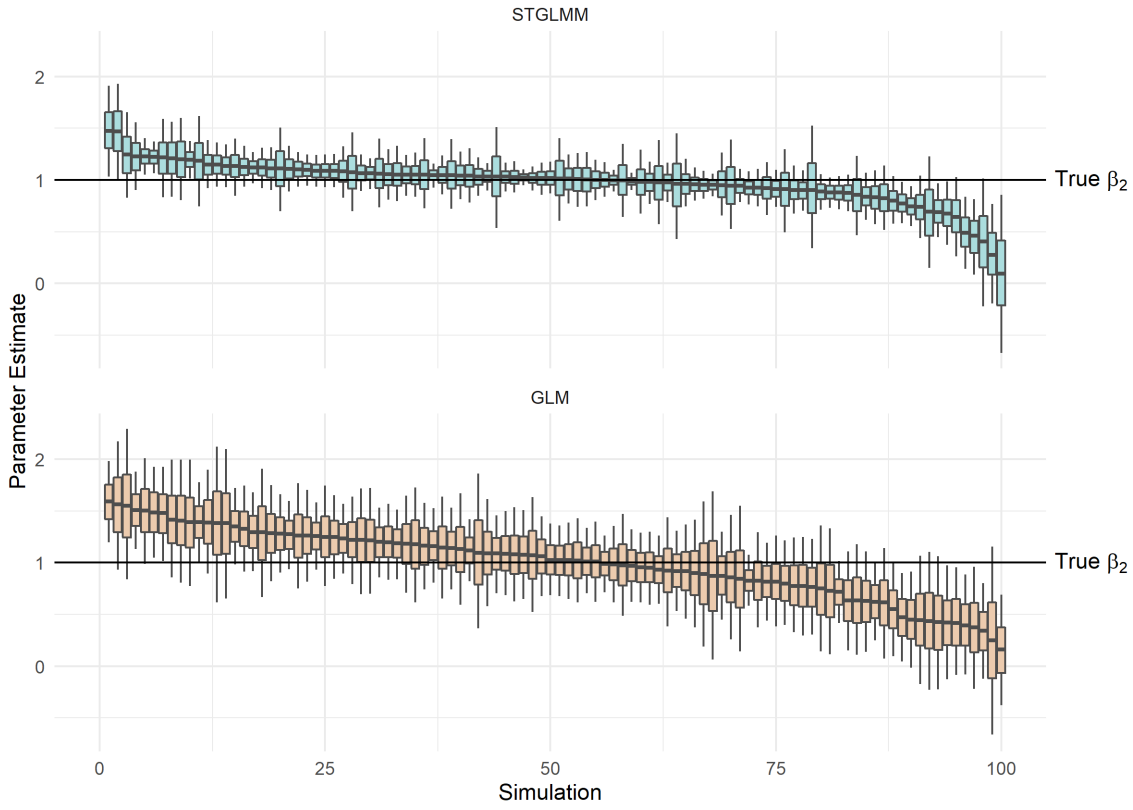
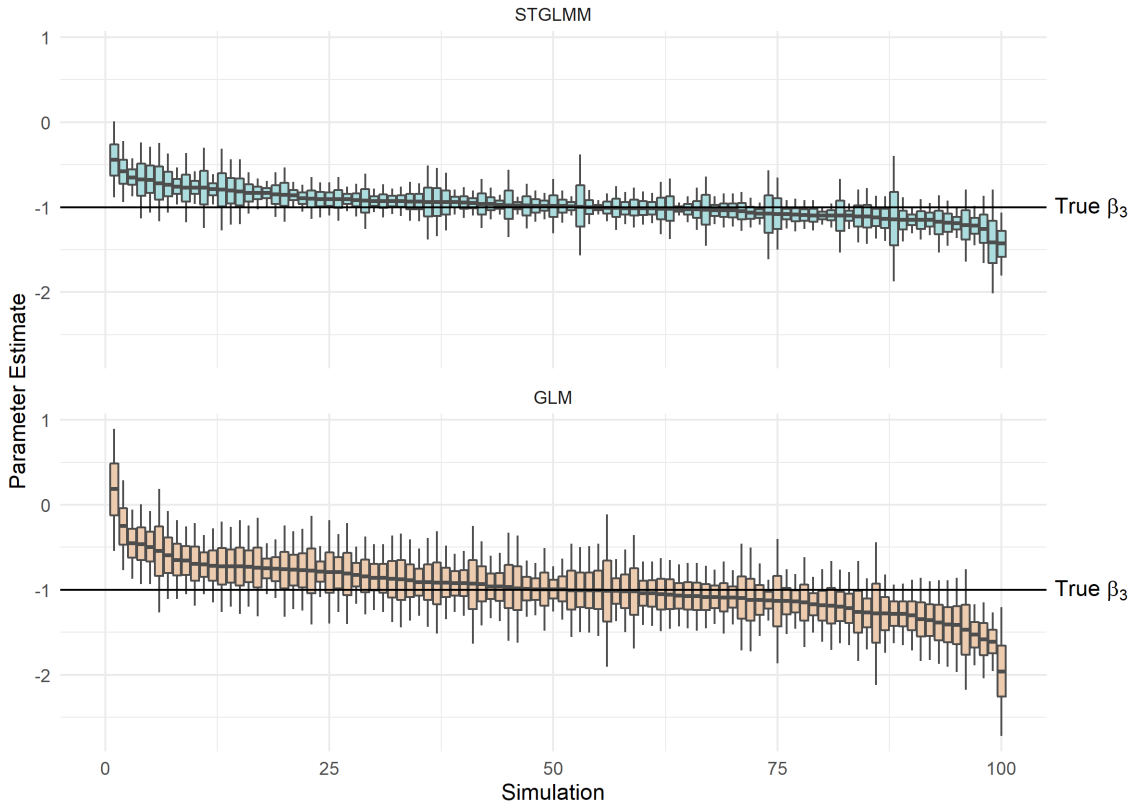


Figure 19: Credible intervals for β_3 across simulations for each model when $\phi = \psi = 0.5$.



A.2.3 Parameter Set 1: $\phi = \psi = 1$

Figure 20: Credible intervals for β_0 across simulations for each model when $\phi = \psi = 1$. β_0

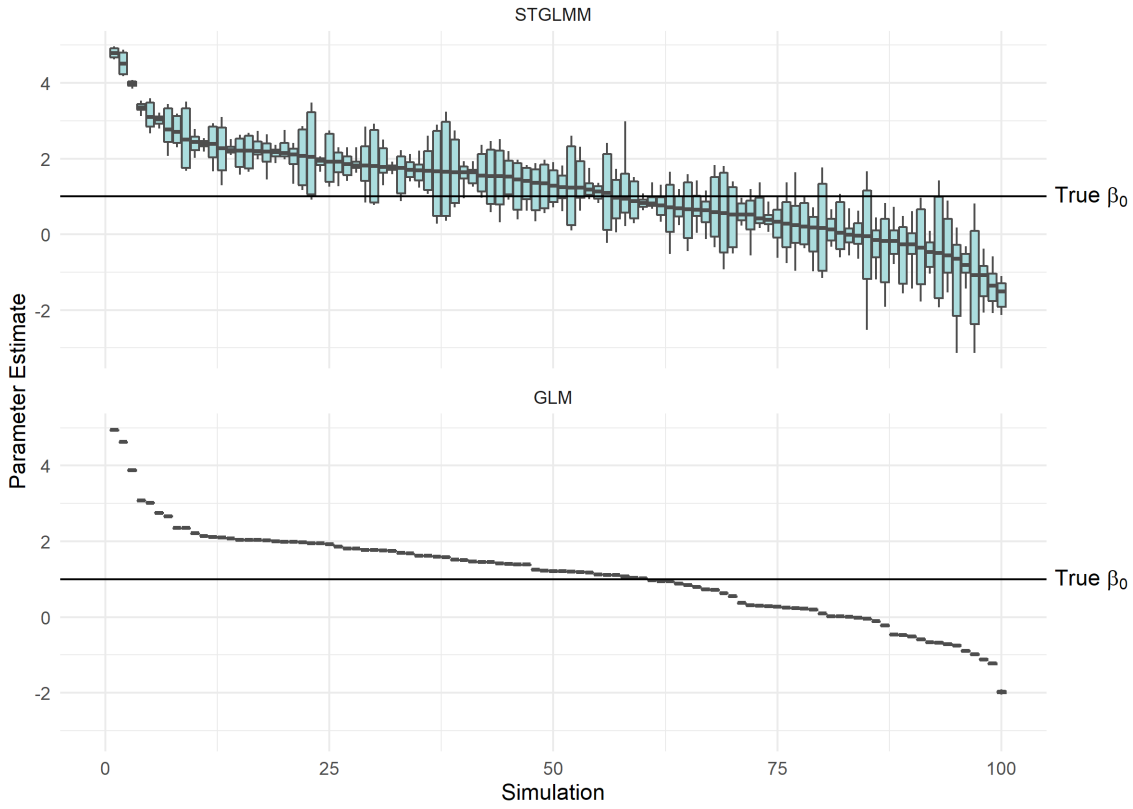


Figure 21: Credible intervals for β_1 across simulations for each model when $\phi = \psi = 1$.

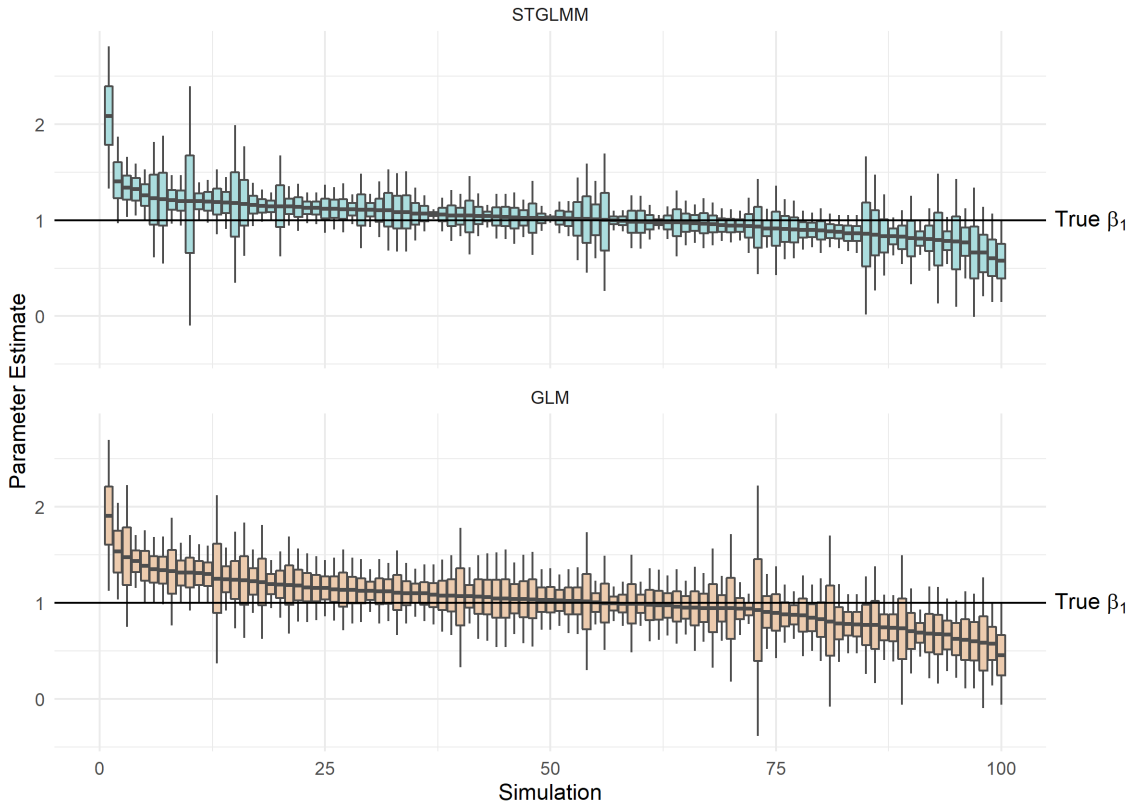


Figure 22: Credible intervals for β_2 across simulations for each model when $\phi = \psi = 1$.

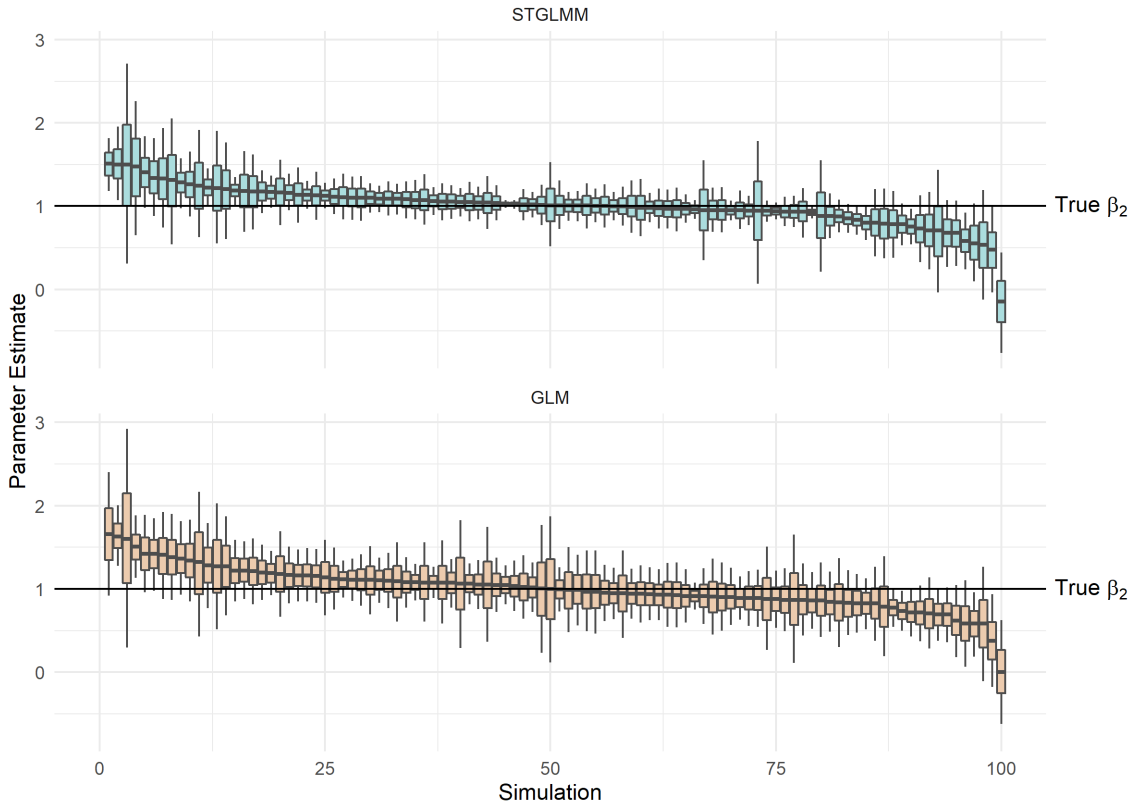
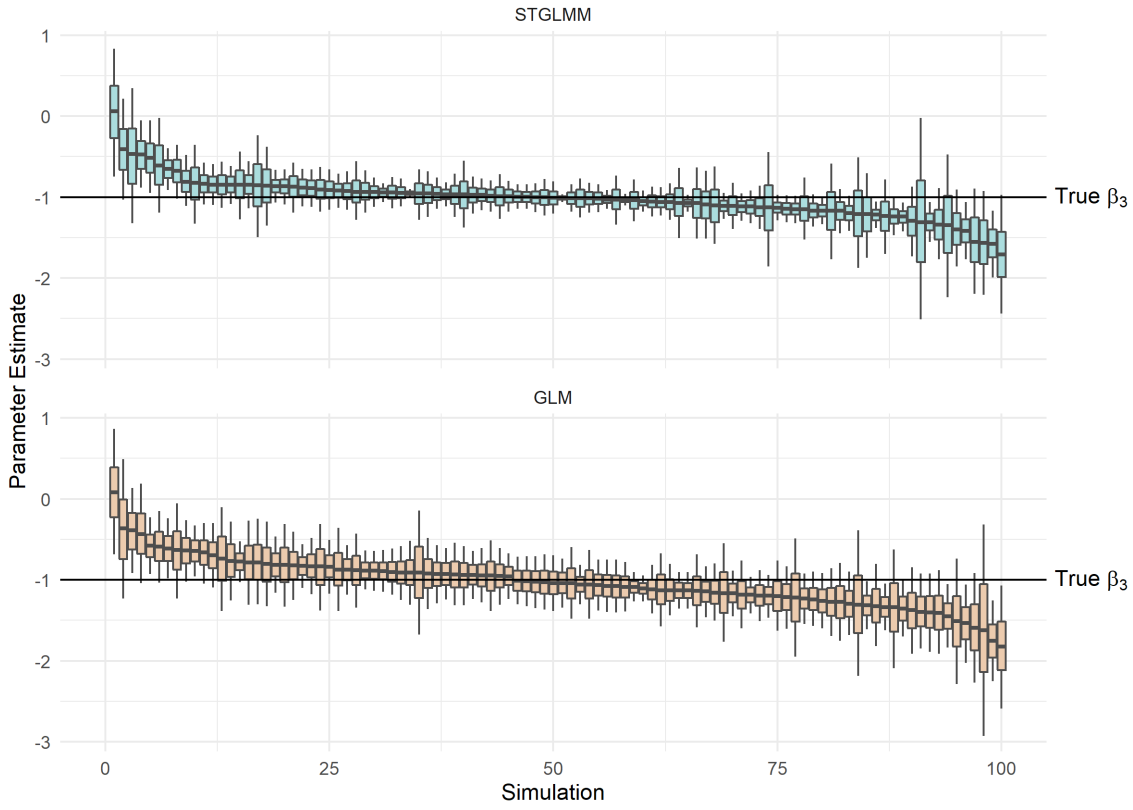


Figure 23: Credible intervals for β_3 across simulations for each model when $\phi = \psi = 1$.

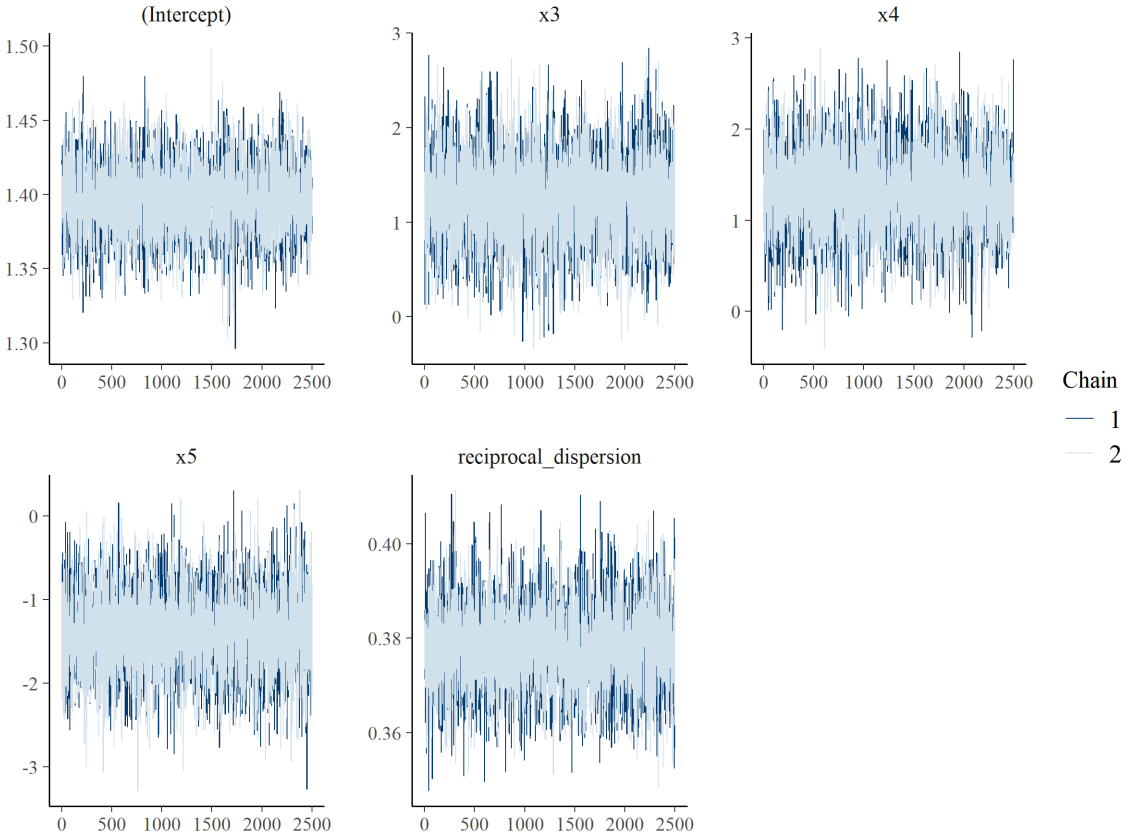


A.3 Markov Chain Diagnostics

Here we present the Markov chain trace plots for the GLM and low rank STGLMM estimated for an example simulated data set, as well as the GLM and the low rank STGLMM estimated for Carolina Wren data set.

A.3.1 Simulated Data Set: GLM

Figure 24: Simulated Data GLM Markov chain trace plots



A.3.2 Simulated Data Set: STGLMM

Figure 25: Simulated Data STGLMM Markov chain trace plots for fixed effects

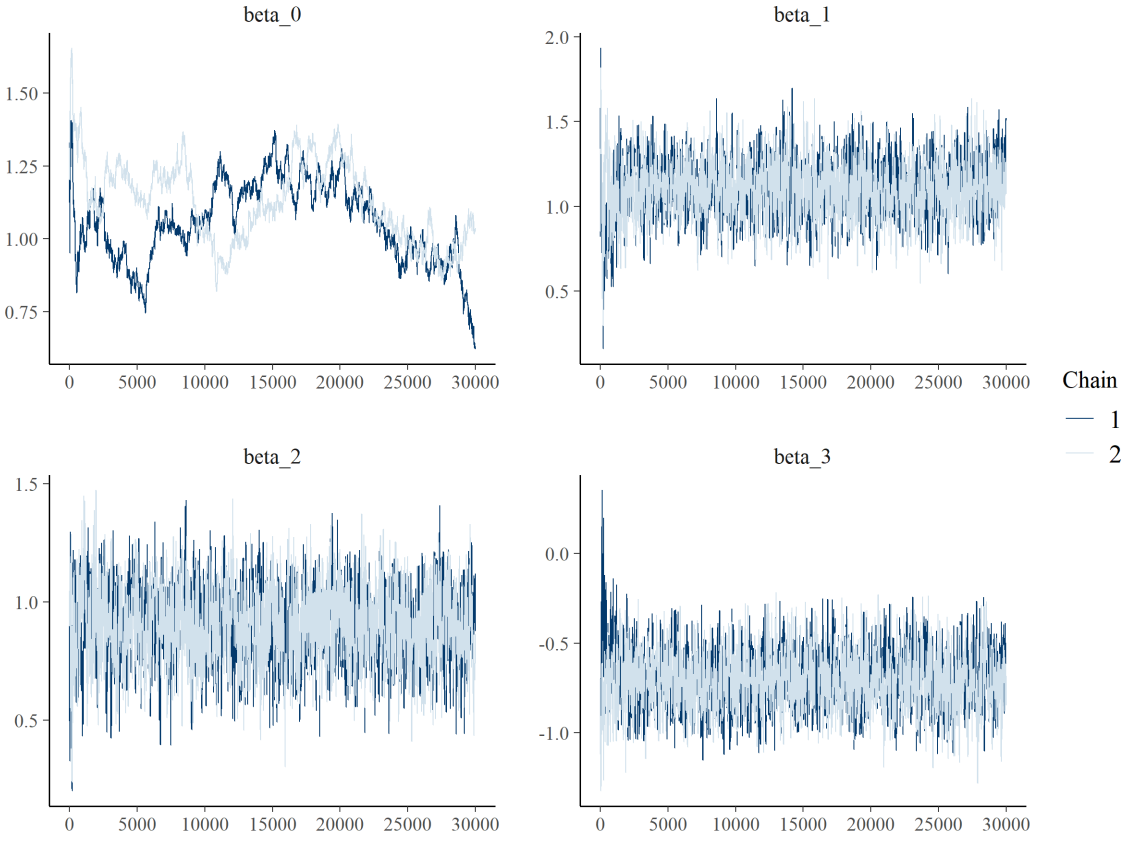


Figure 26: Simulated Data STGLMM Markov chain trace plots for hyperparameters

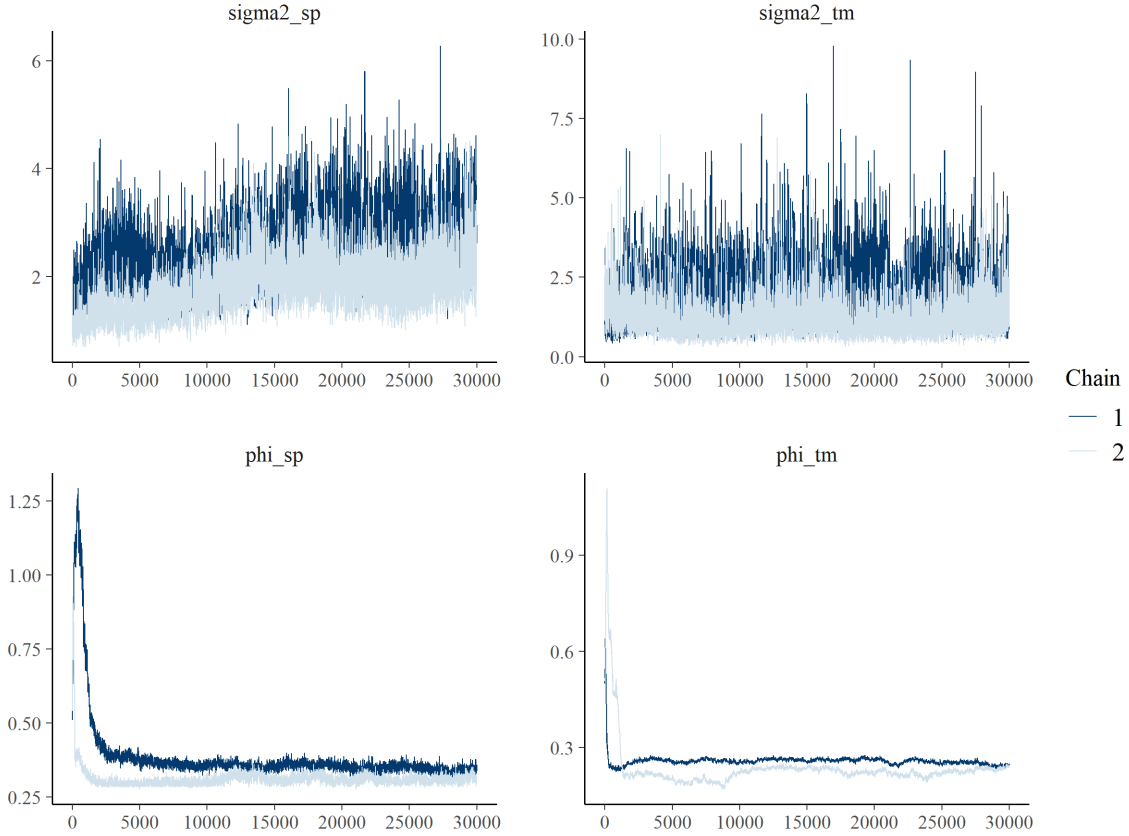


Figure 27: Simulated Data STGLMM Markov chain trace plots for spatial effects (1-25)

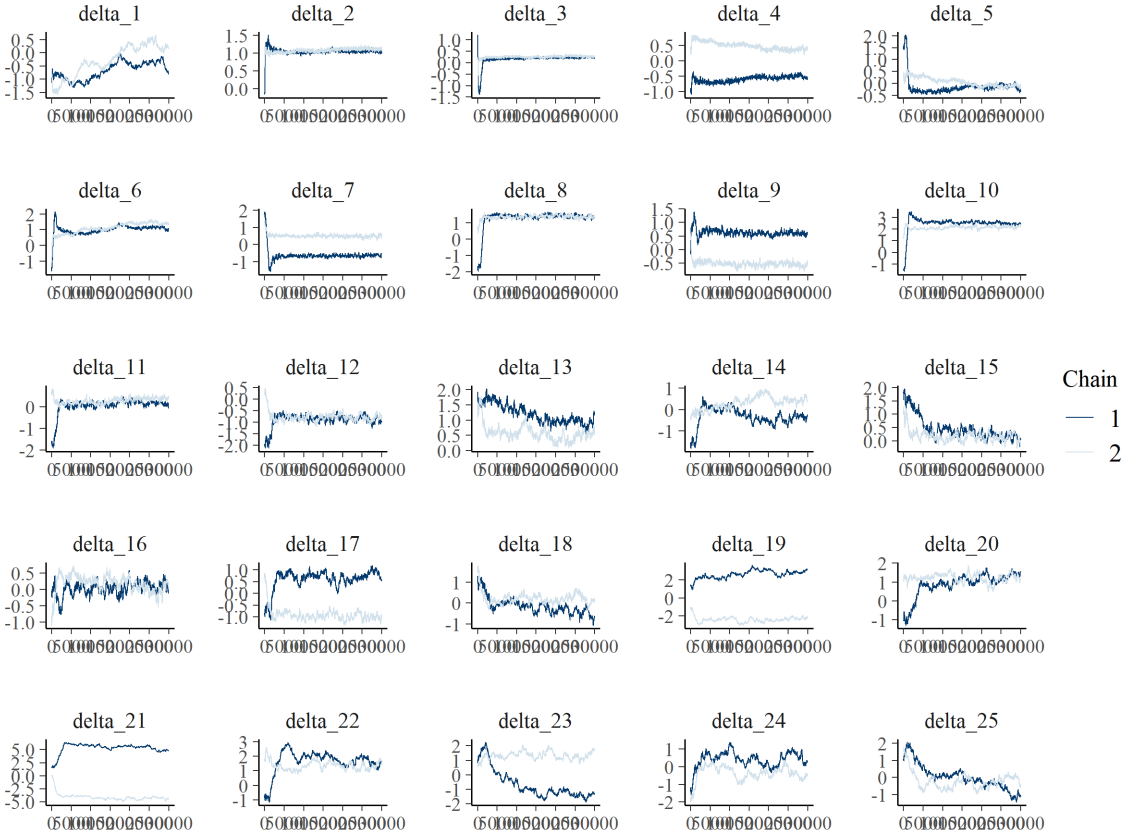


Figure 28: Simulated Data STGLMM Markov chain trace plots for spatial effects (26-50)

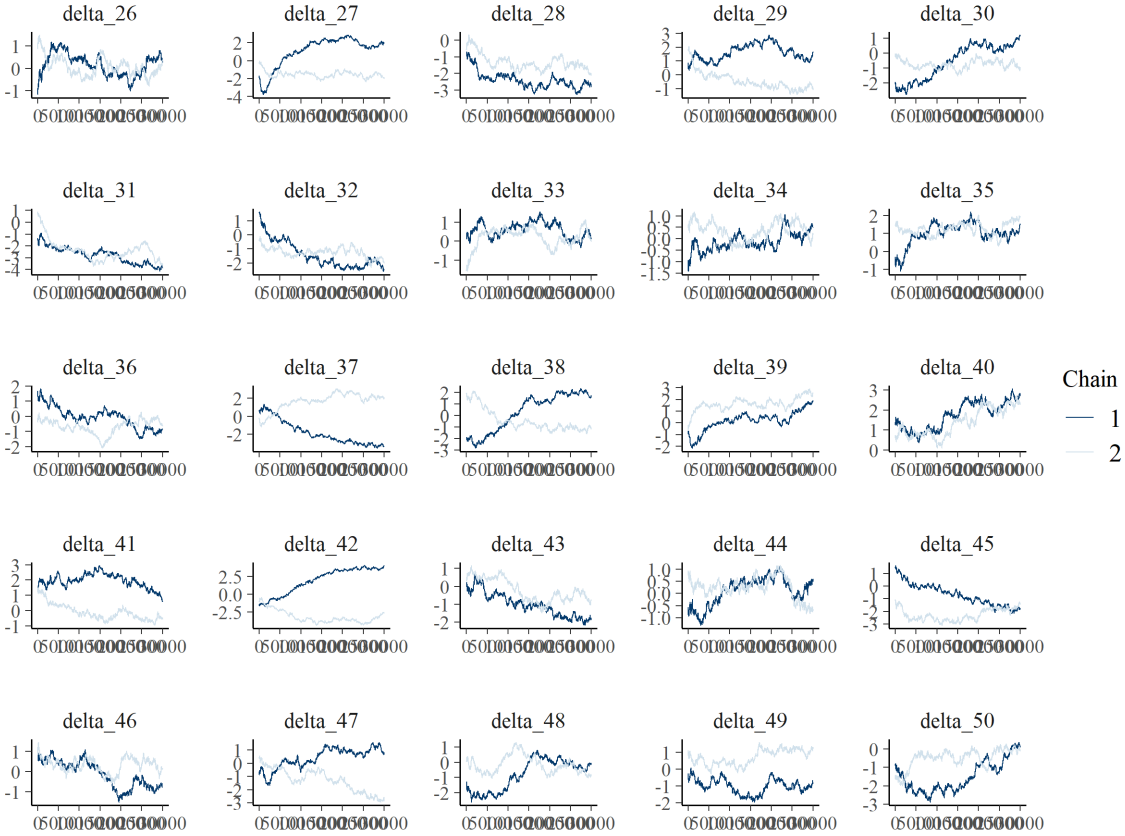
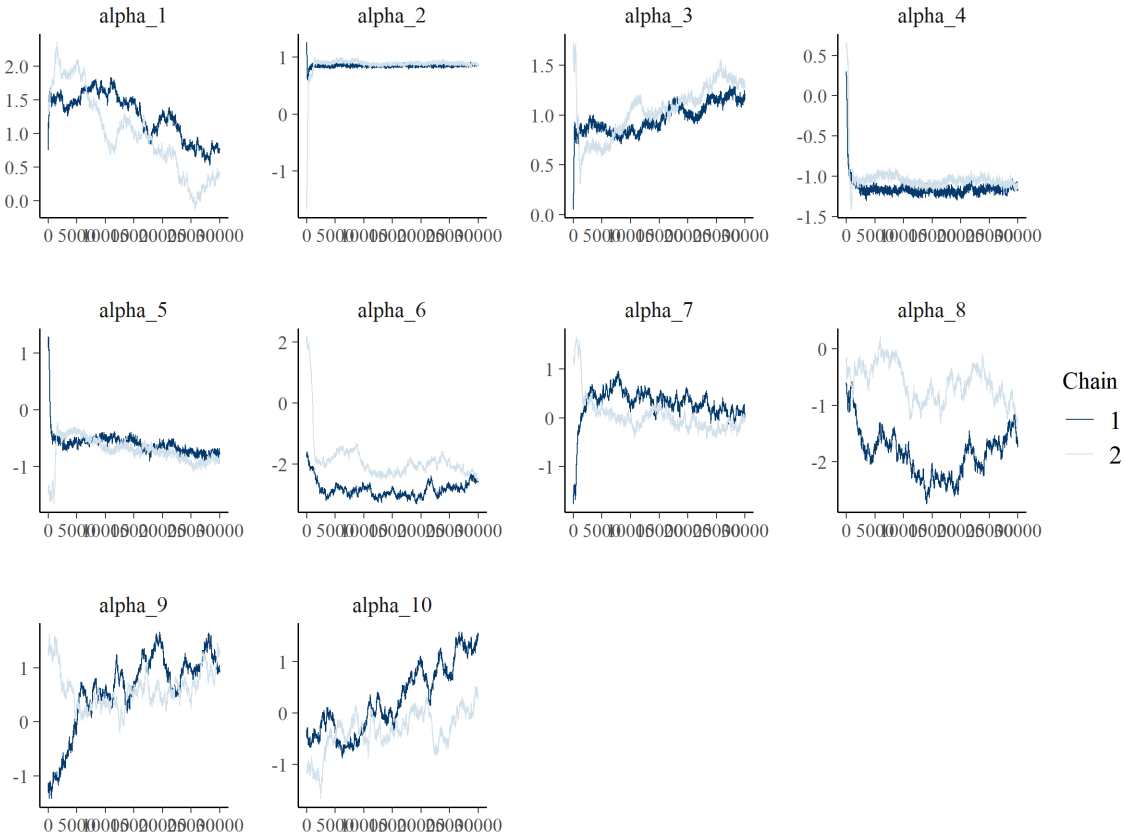
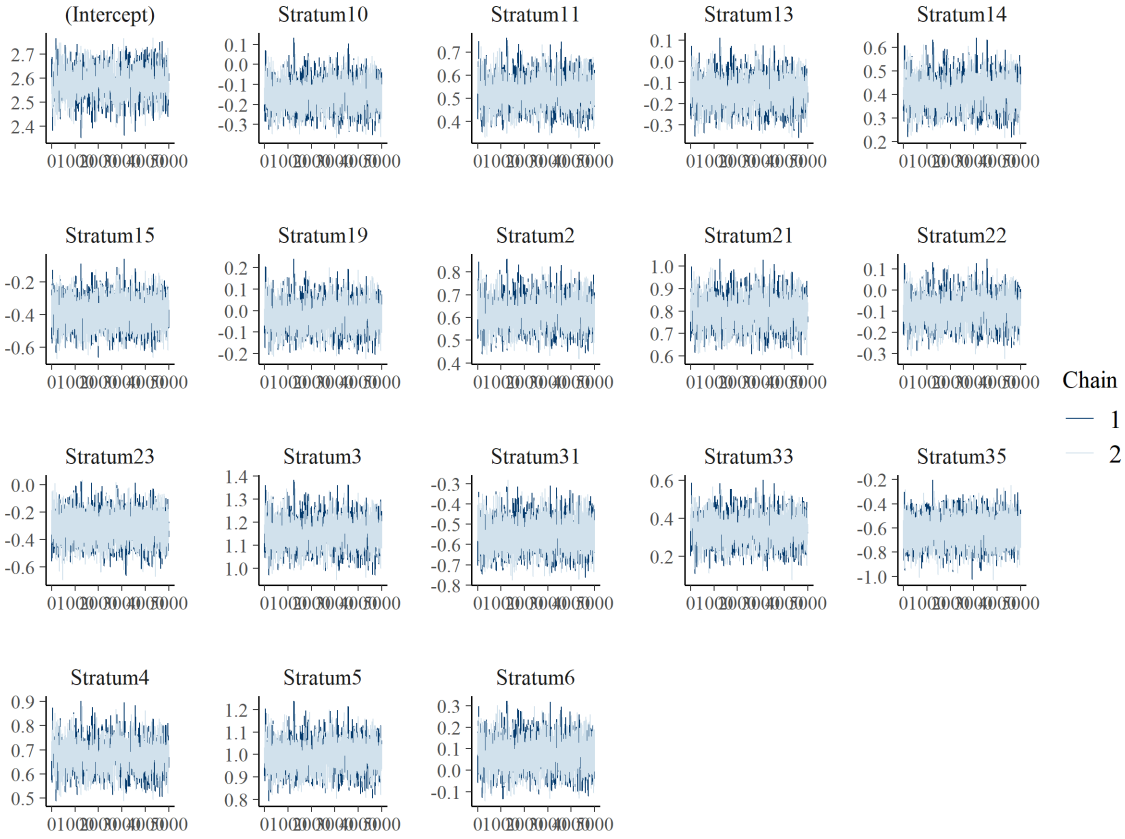


Figure 29: Simulated Data STGLMM Markov chain trace plots for temporal effects



A.3.3 Carolina Wren Data Set: GLM

Figure 30: Carolina Wren GLM Markov chain trace plots



A.3.4 Carolina Wren Data Set: STGLMM

Figure 31: Carolina Wren STGLMM Markov chain trace plots for fixed effects

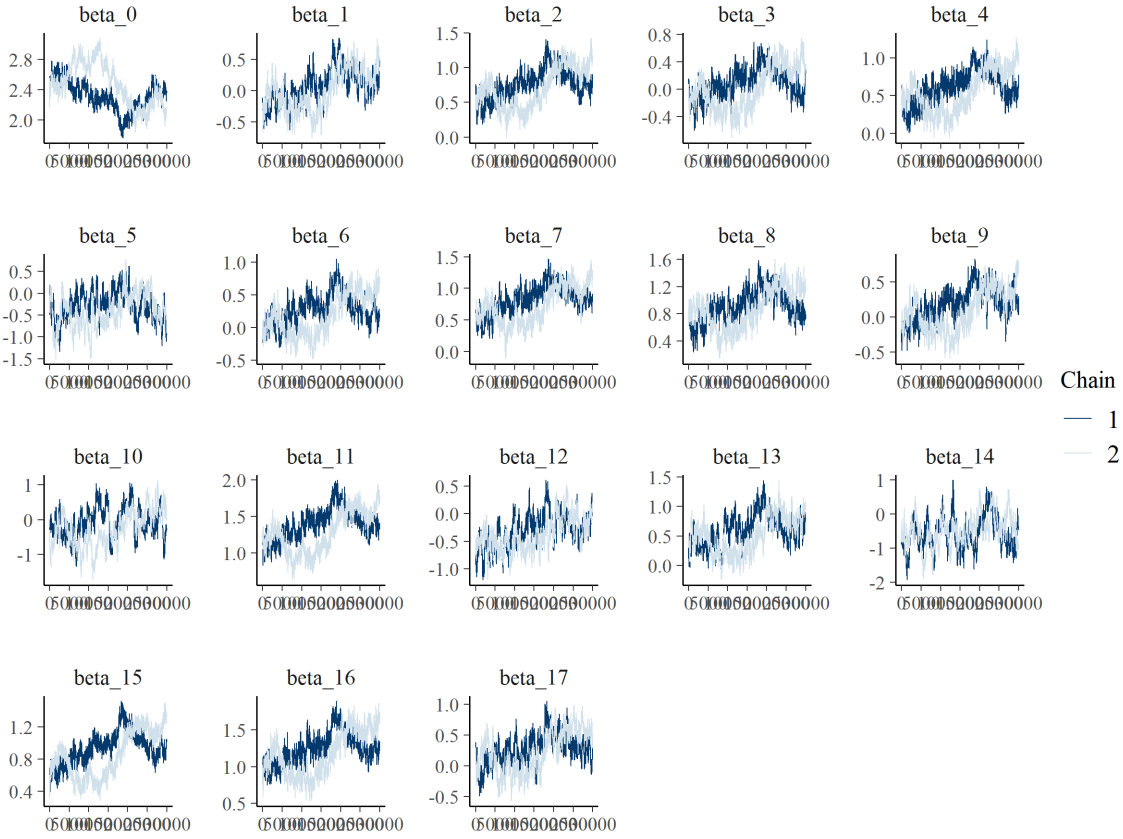


Figure 32: Carolina Wren STGLMM Markov chain trace plots for hyperparameters

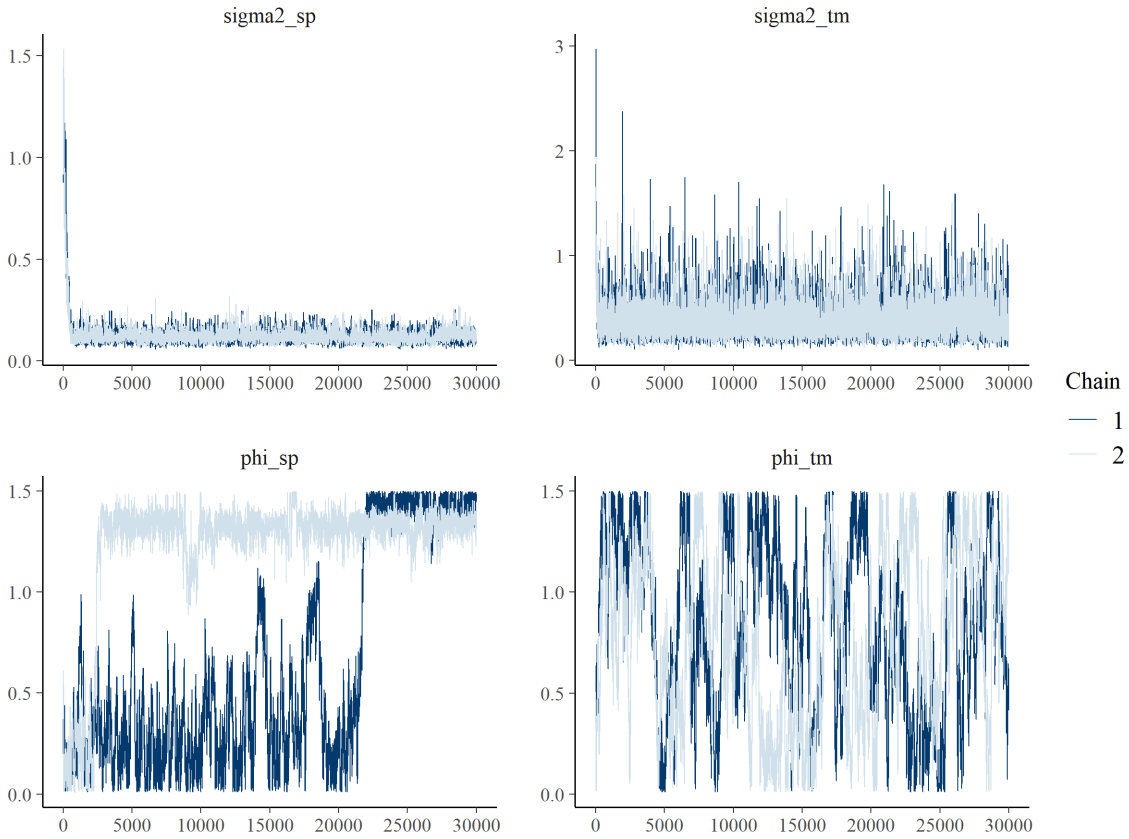


Figure 33: Carolina Wren STGLMM Markov chain trace plots for spatial effects (1-25)

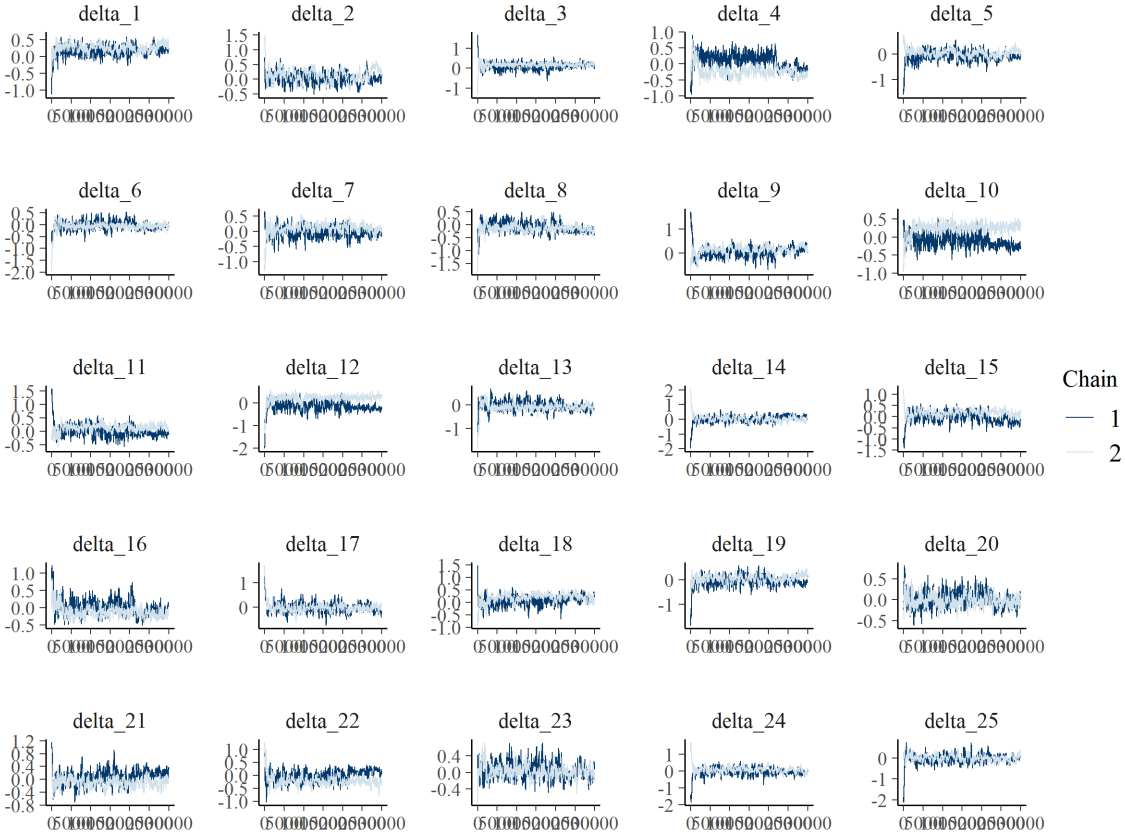


Figure 34: Carolina Wren STGLMM Markov chain trace plots for spatial effects (26-50)

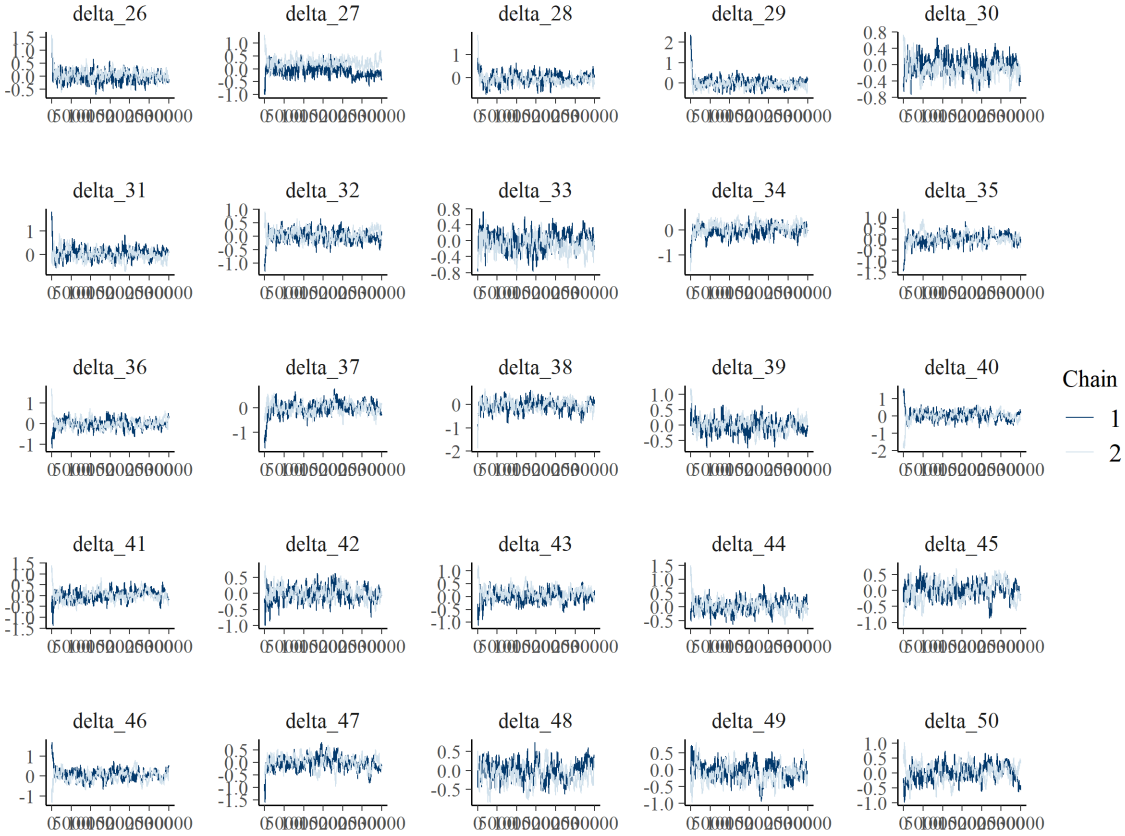


Figure 35: Carolina Wren STGLMM Markov chain trace plots for temporal effects