Honors Thesis

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 (citation! Wikle 2019 probably works). 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. Such high dimensionality can make estimating models for even moderately sized data, 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 (Guan and Haran, 2018; Bradley et al, 2015). 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.

One prominent example of this in practice is the work of the United Church of Christ's seminal paper on racism in toxic waste facility siting decisions. In the original study, performed in 1987, and in their follow-up study in 2007, they 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. These problems persist in more modern methods, such as the necessity of specifying the basis functions used to approximate spatial dependence in Spatio-Temporal Statistics with R book (Wikle et al, 2019?).

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 (2018) 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. 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. I then apply 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 Gaussian Processes and the SGLMM

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 (2015) and Zhang (2002). 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 w be a mean zero random effect with variance parameter σ_w^2 and an $n \times n$ covariance matrix $\Sigma(\sigma_w^2, \phi)$ 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 $\rho(;\phi)$ is the Matérn covariance function. This function has two parameters, ϕ and ν. Only one of these parameters is identifiable (citation? its somewhere in Banerjee 2015), and because of this we often set ν to be 1.5 or 2.5, which works well in most applications.

This model is referred to as a spatial linear mixed model due to the combination of fixed effects $\boldsymbol{\beta}$ and spatial random effects $w(s)$ (Zhang 2002). This spatial model can be extended for non-Gaussian outcomes $Y(s)$ for some link function $g(\cdot)$

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

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 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 iterations needed to get an accurate estimate of the posterior distribution of parameters.

2.2 Existing Approaches to the SGLMM

Many approaches have been developed to address these computational issues. 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 this approach fails for the non-Gaussian case (Wikel 2019).

One popular choice for dealing with the computational issues that works in non-Gaussian cases is the predictive process model (Banerjee et al, 2008). The predictive process replaces $w(s)$ with $\tilde{w}(s)$, where for some small subset of locations S^* (called knots), we let $w^*(s)$ be the realization of the full spatial Gaussian process at the points in $S*$. We are then left with $w^* \sim N(0, Cov^*(\phi))$, which has a much smaller dimension covariance matrix. Letting $c(s_0; \phi)$ be the evaluation of the covariance function used to generate Cov^* at some point s_0 with all of the knots, we have

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

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.

TODO: Talk about drawbacks of predictive process

TODO: Talk about other low rank methods. (Basis functions, Moran's I, displaced centroids in Prates?)

2.3 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 hierarchically as

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

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^2 is the variance parameter of the spatial effects and $\Sigma(\sigma^2, \phi)$ 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 letting $\eta(s,t) = w(s)\alpha(t)$ and assume 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 (Banerjee 2015 or Wikel 2019 or the original paper?). 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, and for Bayesian fitting correlated mixed effects lead to a lower MCMC mixing efficiency.

2.4 Existing approaches to Spatiotemporal Gaussian Processes

Many approaches have been developed to circumvent the computational issues. For example, to ease MCMC mixing, Bradley et al. (2018) develop a model for count data which uses strategic choices of prior distributions and parameter distributions to allow the construction of a Gibbs sample for the model. This sidesteps difficulties that arise when picking proposal distributions for the high dimensional mixed effects that occur when fitting models using the Metropolis Hasting algorithm (Rue 2009). However, computational issues around calculating the likelihood for the high dimensional random effects remain. Bradley et al.'s (2018) 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 (2015) 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 mixed effects remains (idk the citation for this). (Could also talk about INLA here).

2.5 Random Projections for the SGLMM

One approach of that circumvents the issues of selecting appropriate knots or basis functions is the random projections based approach to dimension reduction of the spatial effects in an SGLMM from Guan and Haran (2018). 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 (citation?) to quickly calculate an approximation of the eigendecomposition. To see how this works, recall the SGLMM framework:

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

Note that the covariance matrix of $\boldsymbol{w}$, $\boldsymbol{\Sigma}(\sigma^2, \phi)$, is symmetric and positive semi-definite. This is true for all covariance matrices, and means that the eigendecomposition $\boldsymbol{\Sigma}(\sigma^2, \phi) = \sigma^2 V \Lambda V^T$ exists for real valued diagonal $n \times n$ matrix Λ consisting of the eigenvalues of $\boldsymbol{\Sigma}(\sigma^2, \phi)$ arranged in descending magnitude, and real valued orthonormal $n \times n$ matrix V, consisting of the eigenvectors of $\boldsymbol{\Sigma}(\sigma^2, \phi)$. The properties of orthonormal matrices allow us to create a independent random variable $\boldsymbol{\delta}$ from the correlated spatial effects $\boldsymbol{w}$. This δ 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 $\boldsymbol{\delta}$ is the identity matrix:

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

Hence, $\boldsymbol{\delta} \sim N(0, \sigma^2 I)$, 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 $\boldsymbol{\Sigma}(\sigma^2, \phi)$. 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 $\boldsymbol{\Sigma}(\sigma^2, \phi)$. The details of the method can be found in banerjee (2012 I think), but broadly consist of using a rank $k << n$ random matrix to project $\boldsymbol{\Sigma}(\sigma^2, \phi)$ 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 $\boldsymbol{\Sigma}(\sigma^2, \phi) \approx \tilde{\boldsymbol{V}} \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{V}}^T$ 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 $\boldsymbol{\Sigma}(\sigma^2, \phi)$. 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

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

$$\boldsymbol{\delta} | \phi, \sigma^2 \overset{approx}{\sim} N(0, \sigma^2 \boldsymbol{I})$$

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 variable $\boldsymbol{\delta}$. 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. Choosing a rank close to 50 seems to work well in most applications while maintainting computational efficiency.

3 Contributions

The Guan and Haran (2018) model can naturally be extended into the spatiotemporal case. Recall the spatiotemporal model with independent spatial and temporal effects given by

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

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

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

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 $\boldsymbol{\gamma} = (UK^{-1/2})^T\boldsymbol{\alpha}$, where the $m \times m$ orthonormal matrix U and $m \times m$ diagonal matrix 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, 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 $l \times l$. Again, to model the generalized linear mixed model version, replace the first equation with $g(E[Z(s,t)|\beta, \eta(s,t)]) = \boldsymbol{x}(s)^T\boldsymbol{\beta} + \eta(s,t)$. We refer to this model as the spatiotemporal generalized linear mixed model, or STGLMM.

TODO: Include likelihood equations for above model: Should this go in an appendix instead of the paper?

TODO: Discuss extension to other seperable spatiotemporal effect specifcations, and maybe mention usage in dynamic spatiotemporal model

4 Applications

4.1 A Simulation Study

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 (cite website?). For this survey, researches 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 1: Observed Carolina Wrens at Selected Sites for 1990 Survey

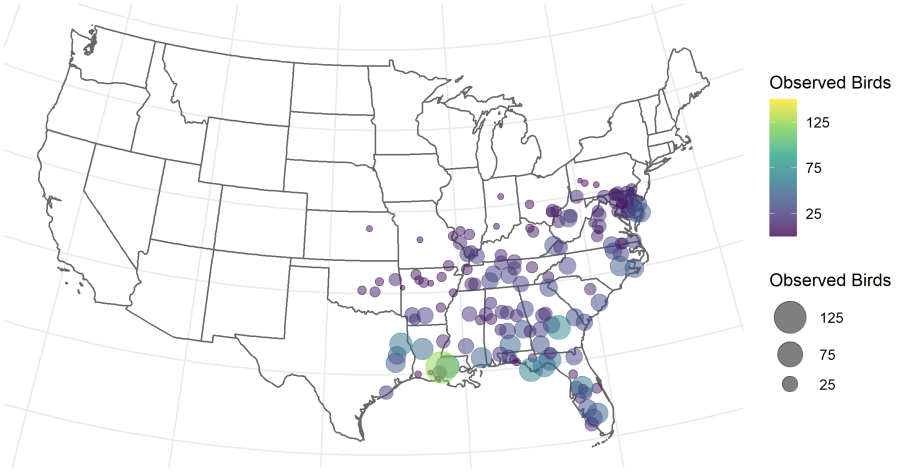
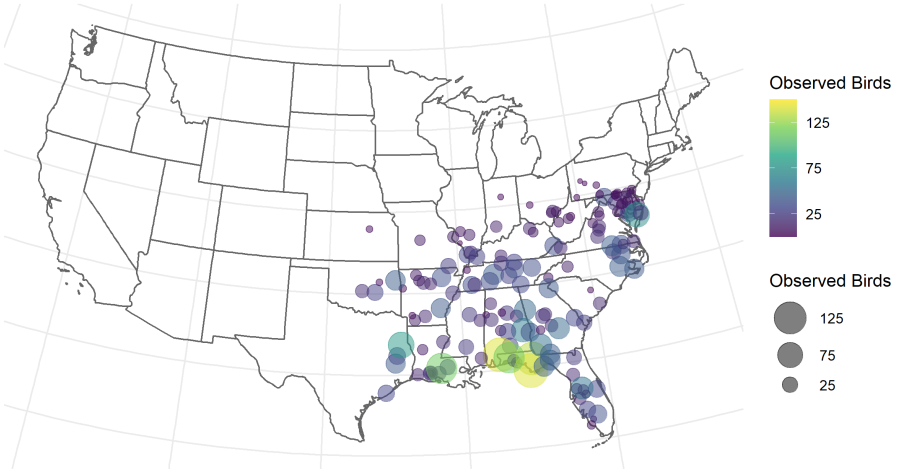


Figure 1 shows the locations of routes as well as the number of Carolina Wrens observed at each route during the 1990 survey. As can be seen, Carolina Wrens live primarily in the American South-East. Notice how there appears to be a spatial pattern to these data. For example, we tend to see larger numbers

of birds observed in and around Northern Florida, and 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 2: Observed Carolina Wrens at Selected Sites for 1991 Survey



In figure 2 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 autocorrelation are present, it is appropriate to use a spatiotemporal model to capture the correlation structure.

To compare between a model that doesn't take into account spatial or temporal autocorrelation, we'll fit a generalized linear model to the data and compare it to the 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, with the stratum as a predictor. The `stratum` represents the type of ecosystem the route is in as defined by ecologists at the Breeding Bird Survey. Because the outcome represents counts, we'll use a log link function for the general linear model. Here, we use β_1 to represent the set of coefficients representing the different possible values of `stratum`. There were 18 different observed strata,

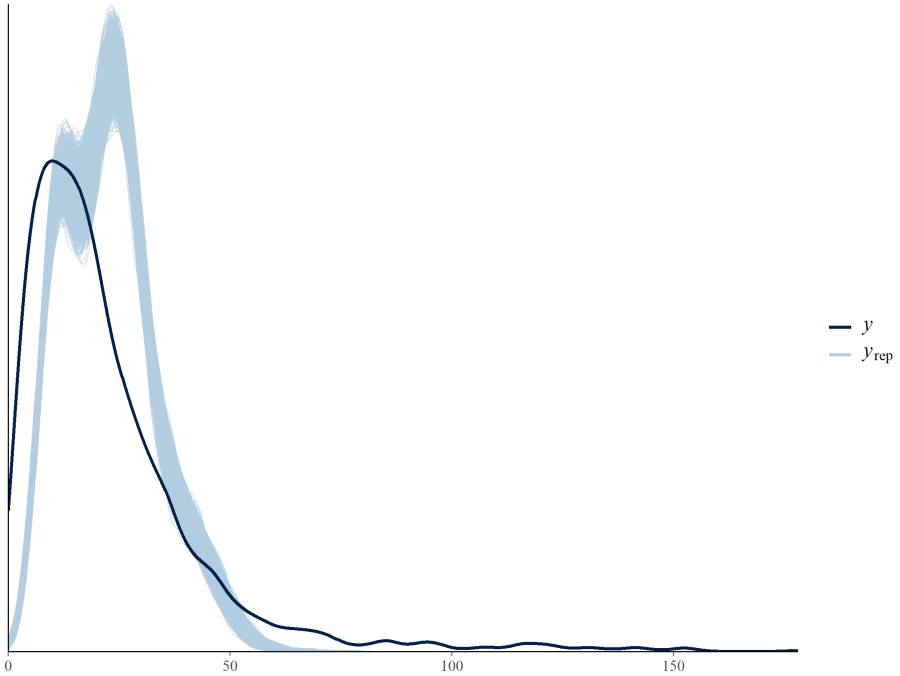
so β_1 actually represents 17 different coefficents.

$$\begin{gathered} log(E[\text{Observed Birds}|\beta_0, \boldsymbol{\beta_1}]) = \beta_0 + \boldsymbol{\beta_1 X} \\ \beta_0, \boldsymbol{\beta_1} \overset{iid}{\sim} N(0, 10^2) \end{gathered} \tag{1}$$

Model 1 was fit using the `rstanarm` package for Bayesian modeling. `rstanarm` uses Hamiltonian Monte Carlo sampling as opposed to Markov chain Monte Carlo sampling used for sampling the STGLMM model, but the interpretation of outcomes is identical. We ran 2 chains with 10000 iterations each, and removed the first 5000 iterations as the burnout period.

Figure 3 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 autocorrelation. 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 3: Posterior Predictions for Generalized Linear Model. True distribution shown in dark blue, predicted distributions shown in light blue.



Now we consider the 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 model can be written

$$
\begin{aligned}
log(E[\text{Observed Birds}|\beta_0, \boldsymbol{\beta_1}, \eta(s,t)]) &= \boldsymbol{x}(s,t)^T + \eta(s,t) \\
\eta(s,t) &= \Big[(\boldsymbol{U}\boldsymbol{K}^{1/2})\boldsymbol{\gamma}\Big]_t + \Big[(\boldsymbol{V}\boldsymbol{\Lambda}^{1/2})\boldsymbol{\delta}\Big]_s \\
\boldsymbol{\delta}|\phi, \sigma^2 &\overset{approx}{\sim} N(0, \sigma^2 I) \\
\boldsymbol{\gamma}|\psi, \theta^2 &\overset{approx}{\sim} N(0, \theta^2 I) \\
\beta &\sim N(0, 10^2) \\
\phi &\sim Unif(0.01, 1.5) \\
\psi &\sim Unif(0.01, 1.5) \\
\sigma^2 &\sim Gamma(2,2) \\
\theta^2 &\sim Gamma(2,2)
\end{aligned}
\tag{2}
$$

We adopt uniform priors for ϕ and ψ and gamma priors for σ^2 and $theta^2$ following the recommendations of Banerjee et al.(2015). 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 4 shows the posterior predictions across each location and time.

Note how model 2 produces a much wider spread of predictions than model 1. This reflects the higher variance in parameter estimates we would expect when data are correlated. However, while model 2 has a better coverage of the true outcomes than model 1, 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 5 displays the posterior density for the stratum representing the Coastal Flatwoods ecosystem. There is a much larger variance in plausible estimates for the parameter in the STGLMM compared to the GLM. Also note how although there are the same number of samples in both of these posterior estimates, the STGLMM has a greater deviation from the expected normal distribution. This is because the 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 STGLMM for a larger number of iterations may minimize this issue.

5 Discussion and Future Work

Figure 4: Posterior Predictions for Spatiotemporal Generalized Linear Model. True distribution shown in dark blue, predicted distributions shown in light blue.

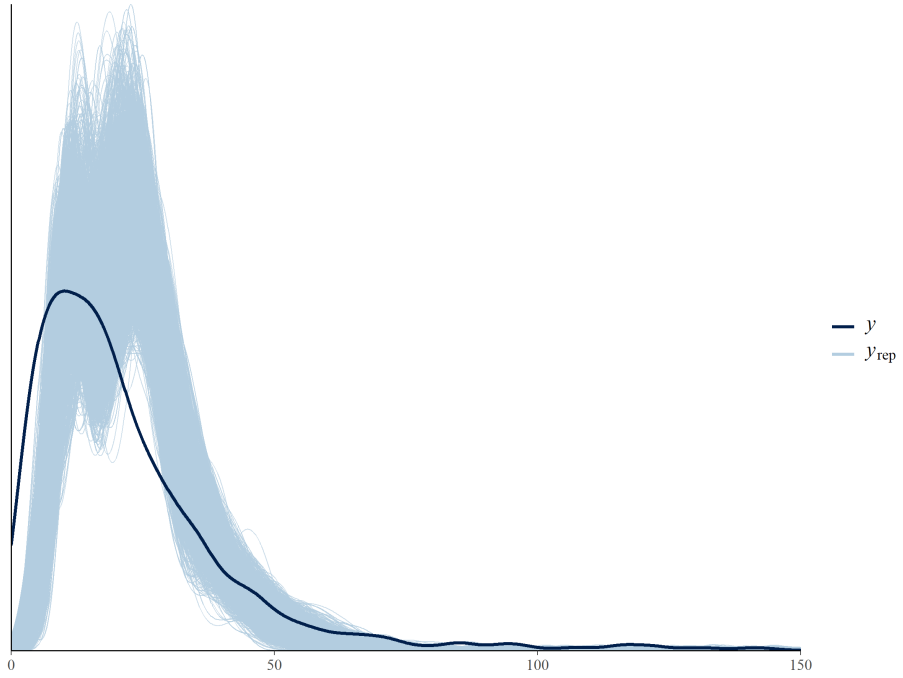


Figure 5: Posterior Estimates for Coastal Flatwoods Coefficient. THe shaded region is 95% central posterior density, the dark blue line is the coefficient point estimate using the median of the draws.