A Centered Bivariate Markov Random Field Model for Mixed Response Distributions

Steve Harms

January 21, 2019

Abstract

The class of Markov random field models known as auto-models provides a flexible and highly-interpretable model structure for analysis of lattice data, which may be of particular interest for research in agriculture, forestry, image construction, or genetics. This paper proposes a Markov random field model with a bivariate Binary-Gaussian response, which provides a foundation for reliable statistical inference with multivariate response distributions of mixed types in the presence of spatial dependence. Pseudo-likelihood-based estimation and inference are investigated via simulation study. the model adequately captures both covariate and spatial dependence relationships under a not-too-strong spatial dependence regime, however the definition of "strong spatial dependence" is a difficult question when distributions are mixed. The model is implemented in 2 cases: data from wheat field trials and data on observations of prairie chickens. The full bivariate model outperforms univariate and non-spatial models in capturing spatial dependence and covariate relationships together.

Contents

List of Figures

List of Tables

1   Introduction

Data collected on a spatial field have become more ubiquitous with the increased use of satellite imagery and other geographic mapping systems, often at discrete-indexed (rather than continuous) locations such as a grid. In some applications (e.g. weather forecasting), prediction may be of interest. However, many areas of research focus on using statistical inference about the spatial correlation to answer a scientific question of interest, such as ecology, epidemiology, image analysis, and marketing. In these cases, a clearly specified model with reliable and interpretable parameter estimates is important to the analysis.

Occasionally there may be more than one different quantity of scientific interest collected at each location. Examples might include species presence and soil properties, crop type and crop yield, or air temperature and rain presence. When the response is bi- or multivariate, an effective model would capture both the spatial dependence and the correlation between response variables while still allowing for reliable inference on covariates. Models for capturing spatial dependence on a grid include spatial generalized linear mixed models, copula-based models, and Markov random field models (auto-models). An effective model would combine a spatial model with a multivariate regression model to capture all of the hypothesized covariance structure in the available data.

In Caragea and Berg (2014), a bivariate auto-logistic model is proposed for presence-absence data for two tree species collected on a grid. The model allows for spatial correlation for each tree species and also models cross-dependence between the two different species. The model proposed in this work is an extension of the bivariate auto-logistic model where the response variables are both Bernoulli distributed, however in our work the distribution of the two response variables need not be of the same type. This extension is intended to both (i) provide a model that can answer a scientific question of interest; and (ii) begin building a foundation to higher-dimensional (p>2) Markov random field models with response variables

coming from more than one family of distributions.

Prior research has considered inference for multivariate Markov random fields, primarily when the response is all Gaussian, such as Mardia (1988) and Rue and Held (2005). There is much additional recent work in the network analysis literature on constructing probabilistic models for structure detection. The model proposed in this work combines these two related fields. Another work that considers mixed spatially-correlated random variables is Hardouin and Cressie (2018), in which an underlying Gaussian fine-grid process is used to capture background variability on an observed coarse-scaled binary quantity (in their case, the presence/absence of corn borers). The model proposed here is similar in many ways, except the Gaussian and binary quantities are on the same discrete-indexed scale.

The rest of the work is organized as follows. In Section 2, the general graph structure is outlined and some foundational work from literature on Markov random fields and mixed graphical models is briefly reviewed. Section 3 builds the centered bivariate mixed Markov random field model based on Section 2, and explores some of the favorable and unfavorable properties of the model. Section 4 outlines the procedure for estimation and inference of model parameters, and Section 5 investigates the reliability of the procedure under different spatial dependence conditions. Sections 6 and 7 illustrate model implementation with brief examples using re-purposed agricultural field trial data and observations of Greater Prairie Chickens, respectively. The final section is a conclusion and discussion of possible avenues for future work.

2 Preliminaries

2.1 Bivariate Graph Structure

First, we define the basic graph structure that necessitates a bivariate mixed model. Consider n locations, denoted as $\{s_i : i = 1, ..., n\}$, arranged on a (not necessarily regular) lattice,

with 2 quantities of interest at each location. One variable, $Y(s_i)$, is distributed as a binary variable and the other, $Z(s_i)$, is Gaussian. Let each quantity at each location to be a node (vertex), such that entire graph has $2n$ nodes. This work only considers a four-nearest neighbor edge structure, with a given interior node being connected to its four nearest neighbors of the *same type* and the node at the same location of the other type. Thus any interior node has an edge connecting it to 5 other nodes, with nodes on the edge/corner of the lattice having fewer than 5. For visualization purposes, it may be easier to view the structure as two separate lattices that are stacked on top of each other on the grid, and then linked via a cross-dependence parameter. This structure is similar to the illustrated multivariate lattice structure in the context of a hierarchical spatial GLMM from section 3 of Sain, et al.(2011). Here, the number of stacked lattices is 2 and the variables are denoted as $Y(s_i)$ and $Z(s_i)$ to distinguish the differing variable distribution types. The neighborhood and distribution structure is defined as above for simplicity and brevity. For notational purposes throughout the rest of this work, we define $\mathbf{z}(\bar{s}_i) : \{z(s_j) : i \neq j\}$ and $\mathbf{y}(\bar{s}_i) : \{y(s_j) : i \neq j\}$ to be all of the nodes of the same type at all locations except s_i.
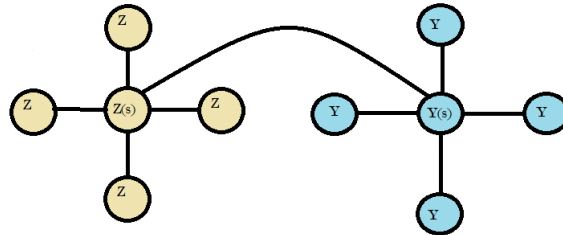
2.2   Markov Random Field Models

Consider univariate versions of the graph structure defined above, such that we have n Gaussian nodes and n binary nodes arranged on a lattice, with no edges connecting nodes of different types. Besag (1974) developed a class of auto-models, in which the probability distribution of each node conditional on all other nodes can be expressed as a one parameter exponential family. For a quantity at location s_i, a general form for the conditional distribution

is

$$f(x(s_i)|x(s_j) : i \neq j) = \exp \{A_i(x(s_j))x(s_i) - B(x(s_j)) + C_i(x(s_i)))\} \tag{1}$$

where $A_i(x(s_j)) = \mu_i + \sum_{j=1}^{n} \eta_{ij} x(s_j)$ is the natural parameter function and $B_i(\cdot)$ is a function of neighboring sites $x(s_j)$. The conditional mean is specified through A_i, where the explanatory variables $x(s_j)$ are themselves observations from the same process. The term $\sum_{j=1}^{n} \eta_{ij} x(s_j)$ is also known as the autocovariate or autoregression term and is thus the reason for the name "auto-" model.

For Gaussian nodes, the model in (1) has specific form

$$f(z(s_i)|z(s_j) : i \neq j) = (2\pi\sigma^2)^{-1} \exp \left[\frac{-1}{2\sigma^2} \left(z(s_i) - \mu_i - \sum_{j=1}^{n} \eta_{ij} z(s_j) \right)^2 \right]$$

and for binary nodes (1) has specific form

$$f(y(s_i)|y(s_j) : i \neq j) = \frac{\exp \left[y(s_i)(\delta_i + \sum_{j=1}^{n} \eta_{ij} y(s_j)) \right]}{1 + \exp \left[\delta_i + \sum_{j=1}^{n} \eta_{ij} y(s_j) \right]}$$

In both cases, if $\eta_{ij} = 0$ for all i and j, the model reduces to the familiar Gaussian and logistic regressions at each node. Additional covariate information is added to the model through the mean function: for covariate vector $\mathbf{x}_i = [x_1(s_i), ..., x_p(s_i)]$ and regression parameter vector $\boldsymbol{\beta} = [\beta_1, ..., \beta_p]$, we can set $\delta_i = \mathbf{x}_i' \boldsymbol{\beta}$ and $\mu_i = \mathbf{x}_i' \boldsymbol{\beta}$.

2 key assumptions are required for this class of auto-models to be well-defined. The first- from Besag (1974)- is a Markov assumption with clique size of at most 2 locations (i.e., a pairwise dependence model). This implies that $x(s_i)$ depends conditionally on neighboring quantity $x(s_j)$ only through the dependence parameter η_{ij}, with no indirect influence by other quantities. For the neighborhood set N_i for location s_i, $\eta_{ij} = 0$ if $s_j \notin N_i$, so that any two observations are conditionally independent if they are not in the same neighborhood. A second

assumption is that the autoregressive function $\sum_{j=1}^n \eta_{ij}x(s_j)$ is a linear function of $x(s_i)$.

2.3 Mixed Multivariate Markov Random Fields

The Markov random field model of Section 2.2 defines a spatial dependence structure for observations of the same type. We can extend that model to define a dependence structure for random variables of different types. Lee and Hastie (2015) consider a pairwise graphical model with node conditional distributions specified as Gaussian and logistic regressions. With mean parameters μ and δ, and dependence parameters η_z, η_y, and ρ, these conditional distributions take the form of

$$p(z(s_i)|\mathbf{z}(\bar{s}_i), \mathbf{y}(s_j)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-\sigma^{-2}}{2} \left[\frac{\mu_i + \sum_{i=1}^n \rho_{ij}y(s_j) + \sum_{i=1}^n \eta_{z,ij}z(s_j)}{\sigma^{-2}} - z(s_i) \right]^2 \right\} \tag{2}$$

and

$$p(y(s_i)|\mathbf{y}(\bar{s}_i), z(s_j)) = \frac{\exp\left\{ y(s_i)\left(\delta_i + \sum_{i=1}^n \rho_{ij}z(s_j) + \sum_{i=1}^n \eta_{y,ij}y(s_j) \right) \right\}}{1 + \exp\{ y(s_i)\left(\delta_i + \sum_{i=1}^n \rho_{ij}z(s_j) + \sum_{i=1}^n \eta_{y,ij}y(s_j) \right) \}} \tag{3}$$

The forms in (2) and (3) are extended from Section 2.2 with additional autoregressive parameters ρ_{ij} for dependence between nodes of different types. (2) and (3) lead to a full joint distribution up to a normalizing constant. A form of this joint distribution adapted from Halsbeck and Waldorp (2015) is

$$p(\mathbf{y}, \mathbf{z}|\boldsymbol{\eta^y}, \boldsymbol{\eta^z}, \boldsymbol{\rho}) \propto \exp\{ \sum_{i=1}^n [\frac{-z(s_i)^2}{2\sigma^2} + \frac{\mu_i}{\sigma^2}z(s_i) + \delta_i y(s_i)]$$

$$+ \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\sigma^2}\eta_{z,ij}z(s_i)z(s_j) + \sum_{i=1}^n \sum_{j=1}^n \eta_{y,ij}y(s_i)y(s_j) + \sum_{i=1}^n \sum_{j=1}^n \frac{\rho_{ij}}{\sigma^2}z(s_i)y(s_j)\} \tag{4}$$

Table 1 in Chen et al. (2014) provides restrictions for the parameters $\boldsymbol{\eta^y}, \boldsymbol{\eta^z}, \boldsymbol{\rho}$ that allow for valid joint distributions formed from conditional specifications from various one-parameter exponential family distributions, such as (4). In particular, (4) requires that the log-normalization term be finite so that the distribution is normalizable. For the model considered in this work,

9

with only Gaussian and binary nodes, the only explicit restriction is that, for all i and j, $\sum_{i=1}^{n} \eta_{ij}^{z}$ is small enough to ensure positive-definiteness of the Gaussian covariance matrix. In a four-nearest neighbor model, this constraint is $\sum_{i=1}^{n} \eta_{ij}^{z} < 4$. There are no restrictions on the binary-binary or Gaussian-binary dependence parameters that would violate the log-normalization requirement above. For models with other types of mixed distributions (i.e., Poisson or gamma-family), details on parameter space restrictions for models can be found in Yang, et al. (2014) or Chen, et al. (2014).

According to Chen, et al. (2014), a conditionally-specified graphical model as in (2) and (3) leads to a valid full joint distribution (4) if and only if, for any $z(s_i)$ and/or $y(s_j)$, the conditional distributions for any set of nodes can be recovered from the full joint distribution. That is, using the definition of conditional probability,

$$p(y(s_i), z(s_j)|\mathbf{y}(\bar{s}_i), \mathbf{z}(\bar{s}_j)) = \frac{p(\mathbf{y}, \mathbf{z}|\boldsymbol{\eta_y}, \boldsymbol{\eta_z}, \boldsymbol{\rho})}{\int p(\mathbf{y}, \mathbf{z}|\boldsymbol{\eta_y}, \boldsymbol{\eta_z}, \boldsymbol{\rho})dy(s_i)dz(s_j)} \tag{5}$$

holds for any choices of nodes $z(s_i)$ and/or $y(s_j)$.

A more general method for constructing valid joint distributions from conditionally-specified models for single-distribution types is provided in Kaiser and Cressie (2000). After first specifying the distribution, neighborhood and mean structure for each node individually, the conditional distributions are stitched together into a full joint distribution for all nodes. Determining a valid joint distribution in many cases requires restriction of the marginal supports, which can be done through the examination of the *negpotential* function for the model.

Define the negpotential function for a pairwise Markov random field model as $\mathbf{Q}(\mathbf{x}) \equiv \sum_{i=1}^{n} H_i(x(s_i)) + \sum\sum_{1 \le i < j \le n} H_{ij}(x(s_i), x(s_j))$, with the H functions specified so that $\mathbf{Q}(\mathbf{x})$ looks like the exponentiated expression in (4). In order to define a valid full joint distribution, Theorem 3 of Kaiser and Cressie(2000) requires that (i) our H functions are invariant to index permutation and (ii) the negpotential functions satisfies $\int_{\Omega} \exp[\mathbf{Q}(\mathbf{t})]dt < \infty$, where Ω is the

support of the joint distribution of the variables $\mathbf{x}(s_i)$. Condition (i) generally holds when we consider undirected graphical models with similar neighborhood structures for each node. Condition (ii) requires careful consideration parameter spaces of spatial dependence parameters (e.g., η_{ij}s and ρ_{ij}s) that specify Ω such that the integral is finite.

Determining the model by first specifying the full joint distribution- then deriving the marginal or conditional distributions of the variables- is not straightforward in most cases. If we instead specify the conditionals and then derive the joint distribution, we need only to check the 2 conditions above to justify the valid joint joint distribution. Theorem 3 of Kaiser and Cressie (2000) leads to a valid full joint distribution as

$$p(\mathbf{y}, \mathbf{z}) = \frac{\exp[\mathbf{Q}(\mathbf{y}, \mathbf{z})]}{\int_\Omega \exp[\mathbf{Q}(\mathbf{t})]dt}$$

which is just a generalization of the condition in (5) for specially constructed node-conditional distributions. In Section 3, we show that both conditions from hold for the binary-Gaussian model.

2.4 Centered Parameterizations for Markov Random Field Models

The class of mixed Markov random field models described in Section 2.3 are primarily utilized to learn the structure of the graphical data using n i.i.d. replicates of the multivariate model. In the bivariate mixed model of this work, the structure is considered fixed and known to a four-nearest neighbor model with one cross-dependence parameter, such that $\eta_{ij} = 0$ for $j \notin N_i$ and $\rho_{ij} = 0$ for $i \neq j$, where N_i is the neighborhood of location s_i. The primary focus here is on reliable inference on mean parameters and covariate effects that allow us to appropriately analyze a scientific question of interest. To do so, we want to re-parameterize the model in (2)-(4) that allows us to separate the effects of mean structure from dependence structure.

As explained by Hughes and Haran (2010), the regression parameters original automodels from Section 2.2 are confounded and uninterpretable. The *marginal* mean structures (μ_i and δ_i) are not separated from the conditional autoregressive structures given through $\sum_{j=1}^{n} \eta_{ij} z(s_j)$ and $\sum_{j=1}^{n} \eta_{ij} y(s_j)$. Interpretation of η_{ij}s is thus difficult because we have not specified whether we are considering them *after* or *together* with the marginal mean parameters. For binary variables, dependence on neighboring values can only increase the conditional expectations for each node, an unfavorable characteristic for the model. Moreover, confounding also makes the dependence parameters η_{ij} and the marginal mean parameters more difficult to consistently recover in the uncentered model, especially when the spatial dependence is non-negligible.

Kaiser, Caragea, and Furakawa (2012) proposed centered re-parameterizations of the models in Section 2.2, wherein the marginal mean structure is subtracted in the autocovariate. Define $\tau_x^{-1}(\cdot)$ as the function that maps expected values onto exponential family natural parameters, such that $\tau_z^{-1}(\mu_i) = \mu_i/\sigma^2$ and $\tau_y(\delta_i) = (\tau_y^{-1}(\delta_i))^{-1} = \frac{\exp(\delta_i)}{1+\exp(\delta_i)}$. The natural parameter functions $A_i(\cdot)$ then become

$$A_i(z(s_i)) = \tau_z^{-1}(\mu_i) + \sum_{j=1}^{n} \eta_{ij} \left(z(s_j) - \mu_j \right) \tag{6}$$

and

$$A_i(y(s_i)) = \delta_i + \sum_{j=1}^{n} \eta_{ij} \left(y(s_j) - \tau_y(\delta_j) \right) \tag{7}$$

The centered re-parameterization alleviates the confounding and allows a convenient interpretation of the model parameters, because we have now isolated the marginal mean structure such that dependence parameters can only affect the covariance and conditional mean structure. For weak to moderate spatial dependence, we can interpret the parameters μ_i and δ_i as approximately reflecting the marginal or *global* structure of the model, while the η_{ij} parameters reflect the conditional or *local* mean structure of the model. Behavior and interpretation under

strong dependence is considered later.

3 Centered Bivariate Mixed Markov Random Field Model

3.1 Model Development

The bivariate mixed Markov random field model builds on foundations in the previous sections. Recall the undirected bivariate graph structure from Section 2.1, where we have n locations with 2 mixed nodes at each location connected by an edge, and a four-nearest neighbor edge structure among nodes of the same type. The Markov property as stated in the beginning of Section 2.4 governs the conditional distribution of each node, and any two nodes are conditionally independent if they are not in the same neighborhood. For simplicity consider only an isotropic model ($\eta_{ij} = \eta$ and $\rho_{ij} = \rho$ for all i, j) and only interior locations (number of neighbors $= M = 4$ for all i). The latter 2 conditions are not necessary assumptions for a valid model, but allow us to simplify the scope of our investigation in this work. With these assumptions, we combine the mixed conditional distribution specification of (2) and (3) and the centered natural parameter functions of (6) and (7), with an additional centering for the cross-dependence parameter ρ, and build the model from conditionally-specified distributions as in Kaiser and Cressie (2000).

For a binary node at location s_i, we specify the conditional distribution as

$$p(y(s_i)|\mathbf{y}(\bar{s}_i), z(s_i)) = \frac{\exp\left\{y(s_i)\left(\phi(s_i)\right)\right\}}{1 + \exp\{y(s_i)\left(\phi(s_i)\right)\}} \tag{8}$$

where

$$\phi(s_i) = \delta + \frac{\eta_y}{M} \sum_{j \in N_i} \left[y(s_j) - \frac{\exp(\delta)}{1 + \exp(\delta)} \right] + \frac{\rho}{\sigma^2}(z(s_i) - \mu) \tag{9}$$

In canonical exponential family form from (1), this corresponds to $A_i\{\mathbf{y}(\bar{s}_i), z(s_i)\} = \phi(s_i)$ and $B_i\{\mathbf{z}(\bar{s}_i), y(s_i)\} = log[1 + \exp\{A_i\{\mathbf{y}(\bar{s}_i), z(s_i)\}\}]$. The conditional mean of $y(s_i)$ is then

$E[y(s_i)|\mathbf{y}(\bar{s}_i), z(s_i)] = \frac{\exp(\phi(s_i))}{1+\exp(\phi(s_i))}$. In (9), $M = 4$ is the number of neighbors of the same type

for interior node $y(s_i)$ and $\mathbf{y}(\bar{s}_i)$ is the set of binary nodes at all locations s_j, $i \neq j$. As opposed

to the models of Section 2.3, here we divide our dependence parameter η_y by the number of

neighbors M to allow for a convenient interpretation of the conditional mean as an average of

neighboring values.

For a Gaussian node at location s_i, the conditional distribution is

$$p(z(s_i)|\mathbf{z}(\bar{s}_i), y(s_i)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2}[z(s_i) - \alpha(s_i)]^2\right\} \tag{10}$$

where

$$\alpha(s_i) = \mu + \frac{\eta_z}{M} \sum_{j \in N_i} (z(s_j) - \mu) + \rho\left(y(s_i) - \frac{\exp(\delta)}{1 + \exp(\delta)}\right) \tag{11}$$

In (11), $M = 4$ is the number of neighbors of the same type for interior node $z(s_i)$ and $\mathbf{z}(\bar{s}_i)$ is

the set of Gaussian nodes at all locations s_j, $i \neq j$. Under this parameterization, the conditional

mean is $\alpha(s_i)$ and the conditional variance is σ^2. This is the "traditional" parameterization in

terms of the conditional mean following from Besag (1974).

An alternative parameterization for the Gaussian nodes follows from Kaiser (2007) and

Kaiser, Caragea, and Furukawa (2012). We can reparameterize the Gaussian nodes directly

into natural exponential family form as in Section 2.4, such that the sufficient statistic is

$z(s_i)$. Under this parameterization, the structure from (10) corresponds to $A_i\{\mathbf{z}(\bar{s}_i), y(s_i)\} =$

$\frac{\mu}{\sigma^2} + \frac{\eta_z}{M} \sum_{j \in N_i} (z(s_j) - \mu) + \rho\left(y(s_i) - \frac{\exp(\delta)}{1+\exp(\delta)}\right)$, $B_i\{\mathbf{z}(\bar{s}_i), y(s_i)\} = (\sigma^2/2)A_i\{\mathbf{z}(\bar{s}_i), y(s_i)\}$, and

$C_i(z(s_i)) = \frac{-1}{2\sigma^2} z(s_i)$. From this parameterization, the conditional mean of $z(s_i)$ is $E[z(s_i)|\mathbf{z}(\bar{s}_i), y(s_i)] =$

$\sigma^2[A_i\{\mathbf{z}(\bar{s}_i), y(s_i)\}]$, and the conditional variance is σ^2 (assumed to be constant across loca-

tions). We refer to this parameterization in our discussion in Sections 2.4 and 3.3 due to its

connection to the centered parameterization literature as well as its use in Kaiser (2007). A

resulting valid joint distribution is possible for either parameterization, but the induced covari-

ance matrix does differ in each case. This leads to different parameter spaces for η_z depending on the model: we need $\eta_z < 1$ in (10), which we use in the following joint distribution, but $\eta_z < 1/\sigma^2$ in the second parameterization. The resulting model for either parameterization is largely the same and produces generally similar results, but leads to different interpretations of η_z.

Stitching together the conditional distributions, we adapt the negpotential function for the full joint distribution for the centered bivariate Markov random field model from equation 7 in Lee and Hastie (2015):

$$Q(\mathbf{y},\mathbf{z}) = \sum_{i=1}^{n} \left[\frac{\mu_i}{\sigma^2} z(s_i) - \frac{1}{2\sigma^2} z(s_i)^2 + \delta_i y(s_i) \right] + \frac{\rho}{\sigma^2} \sum_{i=1}^{n} \left[(z(s_i) - \mu_i) \left(y(s_i) - \frac{\exp(\delta_i)}{1 + \exp(\delta_i)} \right) \right]$$
$$- \frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{j \in N_i} \left[\frac{\eta_z}{M} (z(s_i) - \mu_i)(z(s_j) - \mu_j) \right]$$
$$+ \sum_{i=1}^{n} \sum_{j \in N_i} \left[\frac{\eta_y}{M} \left(y(s_i) - \frac{\exp(\delta_i)}{1 + \exp(\delta_i)} \right) \left(y(s_j) - \frac{\exp(\delta_j)}{1 + \exp(\delta_j)} \right) \right] \quad (12)$$

The full joint probability distribution is

$$p(\mathbf{y},\mathbf{z}) = \frac{\exp(Q(\mathbf{y},\mathbf{z}))}{\int_{\Omega} \exp(Q(\mathbf{y},\mathbf{z})) d\mathbf{y} d\mathbf{z}} \quad (13)$$

where $\Omega_{Y,Z} = \{\{0,1\} \times \Re\}$ is the full joint support.

It remains to show that the full joint distribution is valid. It is clear that our H functions that define the negpotential function are not affected by permutation of indices, because clique size is restricted to 2 in this case and the model is only centered based on the marginal mean at the specific node. When generalizing beyond pairwise-only dependence to irregular and/or anisotropic structures, the permutation condition is not necessarily met and requires careful consideration. Using the definition of conditional probability and some calculus, it is possible to show that the requirement in (5) and thus also Theorem 3 of Kaiser and Cressie (2000)

holds for this model, and thus $\exp\{Q(\mathbf{y},\mathbf{z})\}$ is a valid generating function.

A more involved proof of the validity of the joint distribution- requiring us to show that the denominator in (13) is finite- is based on the proof for (4) from Chen et al. (2014). One important key in that proof is that the distribution of the Gaussian variables when conditioned on the binary variables is a Gaussian Markov random field. This holds for both the parameterization from Lee and Hastie (2015) (see equations 4-6) and the slightly different parameterization of Halsbeck and Waldorp (2015). We provide the details in Appendix A.1.

3.2 Model Properties

In (8) and (10), the *conditional* (local) mean structure is determined by the parameter functions $\phi(s_i)$ and $\alpha(s_i)$. In accordance with the centering ideas of Section 2.4, the *marginal* (global) mean structure is reflected in parameters μ and δ for cases where the dependence is not too strong. The univariate spatial dependence parameters η and cross-dependence parameter ρ only determine the covariance structure and local expectations of the model, so that we can more reliably estimate both the marginal mean structure and spatial structure.

Extending to a non-constant marginal mean structure, additional covariate information can be added to the global structure, such that $\mu = \mu_i = \boldsymbol{X_z}\boldsymbol{\beta_z}$ and $\delta = \delta_i = \boldsymbol{X_y}\boldsymbol{\beta_y}$ for n x p covariate matrices $\mathbf{X}$ and p x 1 parameter vectors $\boldsymbol{\beta}$. The dependence parameters η_y, η_z, and ρ thus account for additional variation in the responses after accounting for $\mathbf{X}\boldsymbol{\beta}$ structure. Of note in (8) and (10) is that when $\rho = 0$, the model reduces to two univariate autoregression models, and for $\eta_y = \eta_z = 0$, the model reduces to familiar linear and logistic regressions at each node.

An alternative to the bivariate spatial model is to consider two univariate spatial regression models, one for each distribution type. The relationship between $\mathbf{y}$ and $\mathbf{z}$ could then be modeled as a covariate in each regression, by including $\mathbf{y}$ as a covariate in the $\mathbf{X_z}\boldsymbol{\beta_z}$ structure and vice versa. A particular advantage of the bivariate model is that it takes advantage of

knowledge about the spatial structure in the joint likelihood while estimating the relationship between response variables. Later investigation on simulated data shows that when the spatial dependence is relatively weak, two univariate models performs reasonably well when compared to the more complicated bivariate model. For larger values of η_y and η_z, the bivariate model is more effective at capturing the relationship.

3.3   Model Breakdown

Several mentions of model behavior *strong* and *weak* spatial dependence have been made thus far without quantifying or defining what the terms actually mean. A discussion of what constitutes strong dependence means and how it affects the properties of the bivariate mixed model follows.

Even when the model parameters are within the appropriate parameter space to define a valid joint distribution, the convenient interpretation of marginal and conditional expectations as in Section 2.4 depends on the ability to discern between global and local structure. When the dependence is strong, the effect of the local structure through the autocovariate η overwhelms the effect of the global structure in $\boldsymbol{X\beta}$. The observed marginal mean of the response values is a reflection of neighboring values rather than global mean structure, thus estimation of the global regression parameter vector $\boldsymbol{\beta}$ is difficult. With this in mind, we can define *strong* dependence as the values of η_y, η_z, and ρ such that the marginal expectations under dependence are no longer approximately the same as those under independence ($\eta_y = \eta_z = \rho = 0$).

Define the upper bound for a dependence parameter to be the smallest (in absolute value) value of η_y, η_z, and ρ such that the marginal expectations under dependence are approximately equal to the marginal expectations under independence. Kaiser (2007) proposed that the upper bound for a univariate model could be found as

$$|\eta| \leq \left[\sup_{\tau(A_i) \in \Theta} \frac{\tau(A_i) - \kappa_i}{A_i - \tau^{-1}(\kappa_i)} \right]^{-1} \tag{14}$$

17
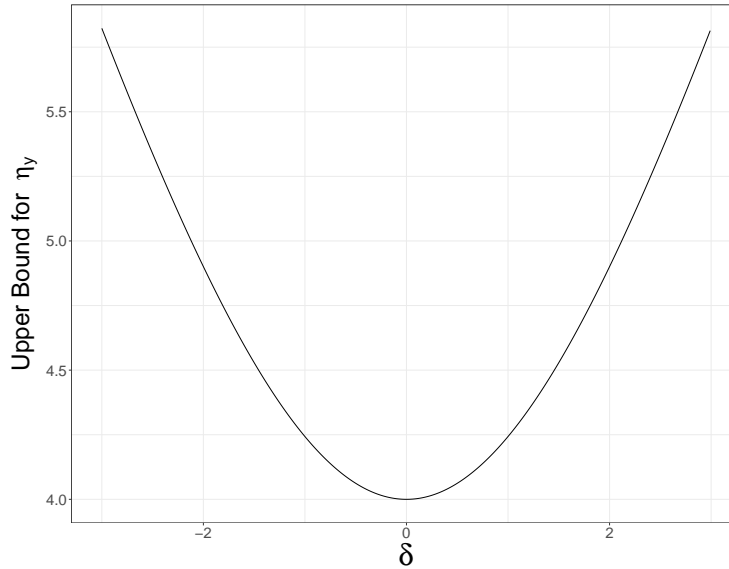
Figure 1: Upper Bounds for η_y vs. δ_i as reproduced from Kaiser (2007)

where τ and A_i are as in (6) and (7) of Section 2.4, $\kappa_i = \mu_i$ or δ_i depending on the model, and Θ is the range of possible values for the conditional expectations. For Gaussian auto-model, Kaiser plugs in the natural parameters and link functions on the right hand side of (14) to get the upper bound for η_z as $1/\sigma^2$. Using the Gaussian parameterization of the same model as in (10), this would correspond to an upper bound of 1. For binary univariate models, the bound must be estimated numerically and depends on the value of δ_i. The smallest upper bound for a four-nearest neighbor model is $|\eta_y| = 4$ when $\delta_i = 0$, with the bound increasing symmetrically as δ_i gets further from 0. A plot of these upper bounds is provided in Figure 1. With $\rho = 0$, binary data generated from the model (8) tends to toward a field of all 0s or 1s as η_y goes beyond its upper bound.

The upper bounds for η_y and η_z as derived from Kaiser (2007) provide suitable starting point limitations for the dependence parameters in the bivariate mixed MRF model. When $\rho \neq 0$, the upper bounds of $\eta_z = 1$ and $\eta_y = 4$ are too large. When generating data from the bivariate model, the cross-dependence can effectively "transmit" the spatial dependence from the Gaussian field to the binary field or vice versa. Hardouin and Cressie (2018) use the term "reinforce" when referring to the connected spatial dependence in their bivariate model. A

suitably large combination of ρ and η_z could cause the generated binary data to be far from its true marginal expectations, even if η_y is near 0 (i.e., no spatial dependence among binary nodes). As a result, we need suitable upper bounds for ρ and ηs *together* in order to determine if the proposed model is viable for inference in a particular dataset.

One suggestion is to consider the nodes together in an isotropic model as in Section 7 of Kaiser (2007), which would suggest bounds of $|\eta_z| + |\rho/\sigma^2| \leq 1$ or $|\eta_y| + |\rho/\sigma^2| \leq 4$. Under spatial independence ($\eta_z = \eta_y = 0$) the bounds for ρ would typically not agree, with one being too small and the other too large. We note that the Gaussian upper bound is determined by the necessity of a positive definite covariance matrix, which is unaffected by ρ in the model specification in (11) with only one cross-dependent neighbor. However, the binary upper bound is determined through the conditional mean specification, and thus we might consider taking $|\eta_y| + |\rho/\sigma^2| \leq 4$ as the very conservative upper bound for ρ. Unfortunately, our simulations often produced degenerate data well before this boundary was reached, thus it appears be too conservative for any practical use.

The preceding discussion is intended to provide caution about interpretation and estimation of model parameters in the presence of strong dependence. The difficulties in deriving a reasonable upper bound are due to the differences in scale and support for Gaussian and binary distributions. Moreover, degenerate auto-normal models and degenerate auto-logistic models do not behave in similar ways under strong dependence. A careful analytical derivation of upper bounds in multivariate models is an avenue for future work. We relegate some speculative initial discussion to Appendix A.2.

4 Estimation and Inference

4.1 Likelihood-based Estimation

In this work we consider the estimation of model parameters via maximum pseudo-likelihood, although Bayesian estimation methods exist, such as using an auxillary spatial field in the prior as in Sun and Clayton (2008) or Liang (2010). We prefer the traditional maximum pseudo-likelihood estimation here due to relatively simple implementation and faster computation.

Estimates of the parameters can be found by maximizing the full log-likelihood or minimizing the negative of the log-likelihood $\boldsymbol{\ell}$, given by

$$\boldsymbol{\ell} = \log \sum_{i=1}^{n} \{ \mathbf{Q}(\mathbf{y}, \mathbf{z}) - \log \mathbf{D}(\mathbf{y}, \mathbf{z}, \eta_{\mathbf{z}}, \eta_{\mathbf{y}}, \rho) \} \tag{15}$$

where $\mathbf{Q}(\mathbf{y}, \mathbf{z})$ is the negpotential function from (12) and $\mathbf{D}(\mathbf{y}, \mathbf{z}, \mu, \delta, \sigma^2, \eta_{\mathbf{z}}, \eta_{\mathbf{y}}, \rho)$ is the normalizing constant from the denominator in (13). The form of $\mathbf{D}(\mathbf{y}, \mathbf{z}, \mu, \delta, \sigma^2, \eta_{\mathbf{z}}, \eta_{\mathbf{y}}, \rho)$ requires a high-dimensional integral for estimation, and is intractable for more than a few locations. To avoid computation issues with the normalizing constant, Besag (1975) instead proposed maximum pseudo-likelihood estimation (MPLE), in which we may take advantage of the conditional specification of the model. Instead of using the full log-likelihood, we consider the likelihood of each individual node as a separate regression model and then stitch together each conditional distribution into a pseudo-likelihood, which is then maximized to obtain parameter estimates. The log-pseudo-likelihood to maximize is given by

$$\tilde{\boldsymbol{\ell}} = \log \left(\sum_{i=1}^{n} \{ p(z(s_i) | \mathbf{z}(\bar{s}_i), y(s_i)) \} \right) + \log \left(\sum_{i=1}^{n} \{ p(y(s_i) | \mathbf{y}(\bar{s}_i), z(s_i)) \} \right) \tag{16}$$

where $p(z(s_i) | \mathbf{z}(\bar{s}_i), y(s_i))$ and $p(y(s_i) | \mathbf{y}(\bar{s}_i), z(s_i))$ are the node-conditional distributions from (8) and (10). While there is some loss in efficiency because the MPLE is not the same as the

full likelihood, Hughes et al. (2011) show that the loss is small for large lattices. When the number of sites is small, the estimates for both MLE and MPLE are typically so inaccurate that efficiency is not the primary concern.

4.2 Simulation and Inference

Simulated realizations of mixed Markov random fields are useful for various inference and model assessment procedures. Conveniently, the specification of conditional distributions allows a ready implementation of a Gibbs sampling procedure. The procedure generates realizations of a bivariate MRF by simulating from the node-conditional distributions and iterating until convergence, then keeping samples from the Markov chain. For pre-specified parameter set $\boldsymbol{\Theta} = (\mu, \delta, \sigma^2, \eta_z, \eta_y, \rho)$ and n locations ($2n$ total nodes) for the model in Section 3, the simulation procedure to generate M independent samples is as follows:

1. Generate n x 1 vectors of starting values $\mathbf{y^{(0)}} \sim Ber(\frac{\exp{(\delta)}}{1+\exp{(\delta)}})$ and $\mathbf{z^{(0)}} \sim N(\mu, \sigma^2)$ for locations $s(i)$, $i = 1, ..., n$.

2. For $j = 1, ..., B + M \cdot S$:

    (a) For locations $i = 1, ..., N$:

    - Generate a new vector of binary values $\mathbf{y^{(j+1)}} = \{y^{(j+1)}(s_i) : i = 1, ..., n\}$ by generating from the node-conditional distributions, a simulation from the distribution in (8):

      $p(y^{(j+1)}(s_i)|\mathbf{y}^{(j)}(\bar{s}_i), z^{(j)}(s_i))$, where

      $\mathbf{y}^{(j)}(\bar{s}_i) = \{y^{(j+1)}(s_k) : k = 1, ..., i-1\} \cap \{y^{(j)}(s_q) : q = i+1, ..., n\}$

    (b) For locations $i = 1, ..., N$:

    - Generate a new vector of Gaussian node values $\mathbf{z^{(j+1)}} = \{z^{(j+1)}(s_i) : i = 1, ..., n\}$ by generating from the node-conditional distributions, a simulation from the distribution in (10):

$$p(z^{(j+1)}(s_i)|\mathbf{z}^{(j)}(\bar{s}_i), y^{(j+1)}(s_i)), \text{ where}$$

$$\mathbf{z}^{(j)}(\bar{s}_i) = \{z^{(j+1)}(s_k) : k = 1, ..., i-1\} \cap \{z^{(j)}(s_q) : q = i+1, ..., n\}$$

3. For $j > B$, keep every Sth collection of observations $\{\mathbf{y}^{(j)}, \mathbf{z}^{(j)}\}$, discarding others, until desired sample size $M \geq 1$ is reached.

Here B is the number of iterations needed for convergence, varying on starting values and lattice size. For a 30 x 30 lattice, $B = 150$ is reasonable but little time is lost in increasing B to 500 or more to ensure convergence. Likewise for the thinning value, $S = 10$ works well but little time is lost in increasing to $S \geq 50$ to ensure independence between simulations. Examination of the autocorrelations between samples for simulated data in Section 5 indicated no concerns about independence for $S \geq 10$ for the schemes used.

As in Caragea and Berg (2014), we obtain measures of uncertainty about parameter estimates for the bivariate mixed Markov random field model via parametric bootstrap. Given maximum pseudo-likelihood estimates $\hat{\mathbf{\Theta}}$ as starting values, we generate bootstrap samples by keeping every Sth simulated dataset after an initial burn-in size of B, until a suitable bootstrap sample size R is obtained for reliable inference. Here we need B large enough to ensure convergence of the first replicate we keep, and we need S large enough to ensure independence between samples.

For each of the R bootstrap replicates, we obtain MPLE estimates and use the percentiles of the resulting empirical distribution to obtain confidence intervals for model parameters. We use a basic bootstrap comparison function to ensure the intervals do not depend on the unknown model parameters. The lower and upper bounds for the $100(1 - \alpha)\%$ bootstrap confidence intervals are $L_\alpha = 2\hat{\theta} - \theta^*_{1-\alpha/2}$ and $U_\alpha = 2\hat{\theta} - \theta^*_{\alpha/2}$, where $\hat{\theta}$ is the initial MPLE of parameter $\theta \in \mathbf{\Theta}$ over the data and θ^*_q is the quantile of the corresponding bootstrap empirical distribution of θ. In Section 5.2, we investigate the empirical coverage of the basic parametric bootstrap under different model parameter conditions.

Alternatives to the basic parametric bootstrap exist. One example is the simple percentile bootstrap, in which we only take the percentiles of the empirical distribution as our bounds. Based on exploratory simulations, we found that the percentile bootstrap tends to give confidence intervals that are much too wide. A normal confidence interval based on large sample properties of the likelihood estimates and the Fisher information matrix is not easily derived from the pseudo-likelihoods. Hughes, et al. (2010) show how to estimate the necessary information for a higher-dimensional approximation to the full likelihood in a univariate autologistic model, although this is much more computationally demanding. Another consideration to improve efficiency would be the double bootstrap, in which the standard error is estimated by bootstrapping each initial bootstrap estimate. Implementing such a procedure in any practical scenario would be prohibitively expensive due to computation time. As we will illustrate in Section 5.2, the basic bootstrap coverage is reasonably close to the nominal coverage and there is not much efficiency to be gained.

4.3 Model Assessment

In addition to the typical measures of model goodness-of-fit, we use the estimated conditional distribution of each node to compare the *in-sample* fit of various models. The continuous ranked probability score (Gneiting and Raftery (2007)) can be used to assess the sharpness (precision) and calibration (bias) of a probabilistic forecast. When comparing the fit of several models, the model that minimizes average CRPS over the data is considered "best". The CRPS is computed based on conditonal distributions, hence it reflects how well the model fits the (conditionally-specified) spatial structure of a given model. The CRPS is a generalization of the Brier score that is illustrated in Caragea and Berg (2014).

For a node $x(s_i)$ with conditional cumulative distribution function $F(x(s_i)|\mathbf{x}(\bar{s}_i), \mathbf{\Theta})$, and

an observation x_i the CRPS is given by

$$CRPS(F(x(s_i)|\mathbf{x}(\bar{s}_i), \mathbf{\Theta}), x_i) = \int_{-\infty}^{\infty} (F(x(s_i)) - \mathbb{1}_{[x_i,\infty)}(t))^2 dx(s_i) \qquad (17)$$

For a specified model, each of n Gaussian nodes has a unique conditional CDF $F(\cdot)$ from a Gaussian distribution as in (10), and each of n binary nodes has a unique conditional CDF $F(\cdot)$ from a Bernoulli distribution as in (8). With our observed data, we calculate the CRPS for each node and then take the average to obtain a measure of in-sample model fit. For some classes of distributions the integral in (17) is difficult to evaluate analytically or numerically, however for only binary and Gaussian conditionals the form is well-known and can be evaluated quickly using the `scoringRules` package in `R` (Jordan, et al. (2017)).

The CRPS is often modified into a "skill score", in which the model's fit is compared to the performance of a reference model. The skill of a model is defined as

$$Skill = 100 \cdot \left[1 - \frac{CRPS(M)}{CRPS(c)} \right] \qquad (18)$$

where $CRPS(M)$ is the mean CRPS of the model over all points as above, and $CRPS(c)$ is the score of a constant mean model where every observation is predicted to be the sample mean. For other scenarios, it may be necessary to use a different reference model as the baseline. The skill of the model is between 0 to 100 and is the percent improvement in a model over the constant model. The optimal model is then the model that maximizes the skill.

5   Simulation Study

5.1   Parameter Estimation

The efficiency of the maximum pseudo-likelihood estimation procedure from Section 4 varies based on the model, lattice size, and especially strength of the dependence, and we investigate

the latter here. In this study, we simulated data for 3 different combinations of dependence. The estimation procedure for the full bivariate mixed model was compared to reduced models on a 30 x 30 square lattice. The reduced models are univariate spatial and non-spatial models, as well as bivariate models without covariates or spatial parameters. The three combinations of dependence parameters (ρ, η_y, η_z) are given in Table 1 and correspond to weak, moderate, and strong dependence, respectively. In particular, the "strong" dependence combination simulates data from a model with dependence parameters close to the edge of the well-behaved parameter spaces discussed in Section 3.3.

For the univariate models, the opposite variable is included as a centered covariate (i.e., in the $\boldsymbol{X}\boldsymbol{\beta}$ structure) rather than in the autocovariate (dependence structure), and the regression coefficient from the univariate Gaussian model is kept as an estimate of cross-correlation ρ for that model. The intuition for this approach in the separate univariate models is that, in a practical situation, we would not ignore available information on the other response variable. Even if we had no knowledge on how to model the variables in a dependence structure as introduced here, we would still like to take advantage of the relationship between the data.

Additionally, we include 2 covariates (one for each response variable) in the model as an illustration of the flexible modeling capabilities for the bivariate mixed MRF model. The covariate for the binary response is a normal random variable with visible but weak spatial correlation generated using the simulation algorithm from Section 4.2, and the covariate for the Gaussian response is a randomly generated $Gamma(3, 4)$ random variable with no spatial correlation. Images of the covariates used are presented in Figure 2. We keep the corresponding regression coefficients $(\beta_{1,y}, \beta_{1,z})$, intercepts $(\beta_{0,y}, \beta_{0,z})$ and variance (σ^2) constant across the 3 combinations of dependence parameters. For each combination of parameters in Table 1, we simulated 1000 data sets and obtained maximum pseudo-likelihood estimates of the parameters for each of the 5 models. Means and 5th and 95th percentiles of parameter estimates are available
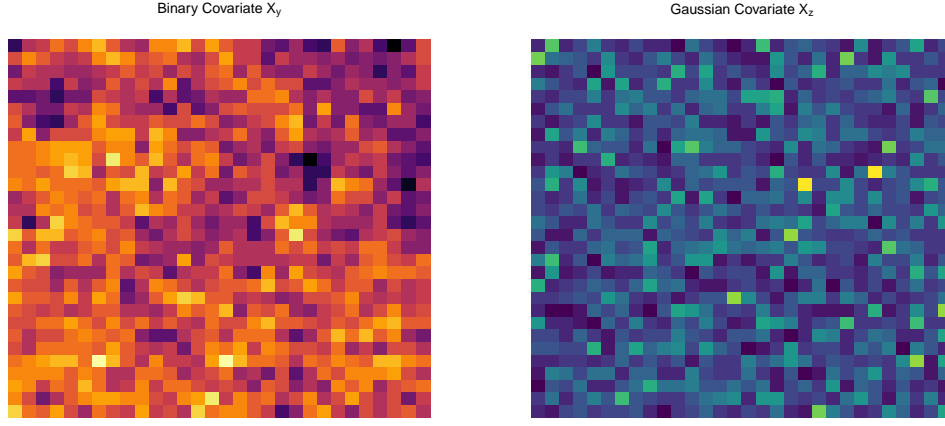
25

Figure 2: Covariates used to simulate data. The covariate for the binary variable (left) exhibits a slight trend and weak spatial dependence.

Parameter Regimes Used to Simulate Data								
Dependence	ρ	η_z	η_y	β_{0y}	β_{1y}	β_{0z}	β_{1z}	σ^2
Weak	0.2	0.3	1	-1	0.5	1	0.5	1
Moderate	1	0.3	1	-1	0.5	1	0.5	1
Strong	0.5	0.9	3.5	-1	0.5	1	0.5	1
Models								
Model	ρ	η_z	η_y	β_{0y}	β_{y1}	β_{z0}	β_{z1}	σ^2
Full (1)	✓	✓	✓	✓	✓	✓	✓	✓
Constant Mean (2)	✓	✓	✓	✓	x	x	✓	✓
Univariate Spatial (3)	*	✓	✓	✓	✓	✓	✓	✓
Univariate Non-Spatial (4)	*	x	x	✓	✓	✓	✓	✓
Bivariate Non-Spatial (5)	✓	x	x	✓	✓	✓	✓	✓

Table 1: Parameter sets and reduced models compared in the simulation study.

in Table 2. Boxplots of the estimates are available in Figures 3 and 4.

Estimates of the dependence parameters in Figure 3 show that for the "weak" dependence combination, estimates from the reduced models are still approximately unbiased for η_z and η_y. As the dependence strength increases, only the full model is able capture all of the dependence structure effectively. The univariate models (M3 and M4) are able to capture some of the cross-dependence ρ without the autocovariate structure at weaker levels of dependence, al-

though the cost is more uncertainty in estimates of the other dependence parameters η_y and η_z. This provides evidence that the more complex full bivariate mixed model is most appropriate when the total spatial dependence (combined ρ, η_y, and η_z) is strong, but may not be necessary under weaker dependence regimes. Moreover, the inability of the constant mean model (M2) to capture spatial dependence indicates the need to first correctly specify the global structure through covariates before any attempt at unbiased estimation inference on the spatial parameters is possible. As one might expect, the variance parameter σ^2 is smallest for the full model in all regimes, and is also the closest to the true simulated value of 1.

Estimates of the regression coefficients (Figure 4) indicate that under weak dependence regimes, the simpler models perform almost as well as the full model. The approximate gain in covariate estimation efficiency for the full model under the weak regime is approximately 10% over the simpler models. Intuitively, the simpler models can only account for variability due to spatial dependence by adding uncertainty to the covariate structure and process variance (σ^2). When the dependence is weak, there is less variation due to dependence that causes bias in the estimates of covariates.

For the strongest combination (far right in Figure 4), none of the models effectively capture the intercepts, and only the bivariate models (M1 and M5) are able to capture the covariate relationships in β_{1y} and β_{1z} (even if unreliably). Because the "strong" dependence regime is beyond the limits of what we would consider a nicely-behaved model, the simulated data reflect local structure rather than global structure. The observed data does not accurately reflect the covariate relationships and expected marginal mean structure because it is overwhelmed by local spatial effects, thus we would not expect any model to effectively estimate those parameters. Figure 4 indicates that our best hope is to use a bivariate model that accounts for as much of the local dependence is possible.

Weak

Model	ρ	η_z	η_y	β_{0y}	β_{1y}	β_{0z}	β_{1z}	σ^2
1	0.2062 (0.0678 , 0.3495)	0.3017 (0.1429 , 0.4559)	1.0727 (0.234 , 1.9024)	-1.0104 (-1.1829 , -0.844)	0.5123 (0.3977 , 0.6372)	1.006 (0.8729 , 1.1321)	0.0997 (0.0912 , 0.1086)	0.9968
2	0.1998 (0.0447 , 0.3574)	0.1973 (0.0472 , 0.3563)	1.4093 (0.6741 , 2.1326)	-0.8144 (-0.9686 , -0.6722)	-	2.2606 (2.1934 , 2.328)	-	1.4841
3	0.5475 (0.274 , 1.0181)	0.3029 (0.1467 , 0.4568)	0.9912 (0.173 , 1.8122)	-1.0018 (-1.1602 , -0.8519)	0.504 (0.3899 , 0.6232)	1.0023 (0.8771 , 1.1293)	0.0999 (0.0912 , 0.1084)	1.003
4	0.6909 (0.3616 , 1.0746)	-	-	-0.9859 (-1.1463 , -0.8382)	0.4982 (0.3854 , 0.6119)	1.0016 (0.8778 , 1.1217)	0.0999 (0.0916 , 0.1084)	1.0295
5	0.2123 (0.0694 , 0.3587)	-	-	-0.9967 (-1.1624 , -0.8474)	0.5035 (0.392 , 0.6242)	1.0018 (0.8753 , 1.1228)	0.0999 (0.0917 , 0.1082)	1.0194

Moderate

Model	ρ	η_z	η_y	β_{0y}	β_{1y}	β_{0z}	β_{1z}	σ^2
1	1.0115 (0.8755 , 1.1387)	0.2989 (0.1706 , 0.4469)	1.1712 (0.3512 , 2.0575)	-0.907 (-1.2885 , -0.4525)	0.5113 (0.389 , 0.6463)	1.0551 (0.8436 , 1.2943)	0.0997 (0.0899 , 0.109)	1.001
2	0.9601 (0.816 , 1.1019)	0.2111 (0.083 , 0.3409)	1.3912 (0.6957 , 2.0989)	-0.7456 (-0.9768 , -0.5441)	-	2.2837 (2.1799 , 2.3844)	-	1.5052
3	1.3286 (1.1061 , 1.5817)	0.3987 (0.2548 , 0.5483)	1.3046 (0.562 , 2.0314)	-0.7135 (-0.8753 , -0.5573)	0.3924 (0.2789 , 0.5049)	1.0738 (0.9187 , 1.2244)	0.1 (0.0907 , 0.1094)	1.2159
4	1.4553 (1.2142 , 1.7302)	-	-	-0.6907 (-0.8445 , -0.5393)	0.3801 (0.2736 , 0.4922)	1.0763 (0.9305 , 1.229)	0.0998 (0.0906 , 0.1096)	1.2721
5	1.0563 (0.9223 , 1.1811)	-	-	-0.8982 (-1.107 , -0.6914)	0.4966 (0.3757 , 0.6224)	1.0472 (0.8972 , 1.2067)	0.0997 (0.0901 , 0.1088)	1.028

Strong

Model	ρ	η_z	η_y	β_{0y}	β_{1y}	β_{0z}	β_{1z}	σ^2
1	0.5142 (0.3418 , 0.693)	0.9117 (0.8444 , 0.9672)	3.4981 (1.8553 , 5.0357)	2.6214 (1.0745 , 3.3464)	0.4372 (0.2468 , 0.6738)	3.9697 (2.2577 , 4.681)	0.0994 (0.0905 , 0.1086)	0.9975
2	0.5565 (0.3166 , 0.7836)	0.79 (0.7034 , 0.8682)	3.5541 (1.9407 , 5.0138)	2.7552 (2.4245 , 3.133)	-	5.3429 (5.1198 , 5.5862)	-	1.5643
3	1.4778 (0.992 , 2.0845)	0.9397 (0.8753 , 0.9951)	4.152 (2.2188 , 5.5685)	2.5053 (2.1888 , 2.865)	0.3368 (0.164 , 0.5297)	4.0064 (3.6027 , 4.3635)	0.0996 (0.0905 , 0.1085)	1.0177
4	1.1773 (0.8635 , 1.4944)	-	-	2.2789 (1.976 , 2.6012)	0.2504 (0.0721 , 0.4537)	3.9488 (3.6971 , 4.1967)	0.1013 (0.0915 , 0.1117)	1.7
5	1.0222 (0.7024 , 1.3476)	-	-	2.4555 (2.1372 , 2.8032)	0.4646 (0.2689 , 0.6668)	3.9585 (3.7029 , 4.2055)	0.1009 (0.0909 , 0.1112)	1.6084

Table 2: Summary of parameter estimates for weak, moderate, and strong dependence regimes. Quantities in parentheses are the 5th and 95th percentiles
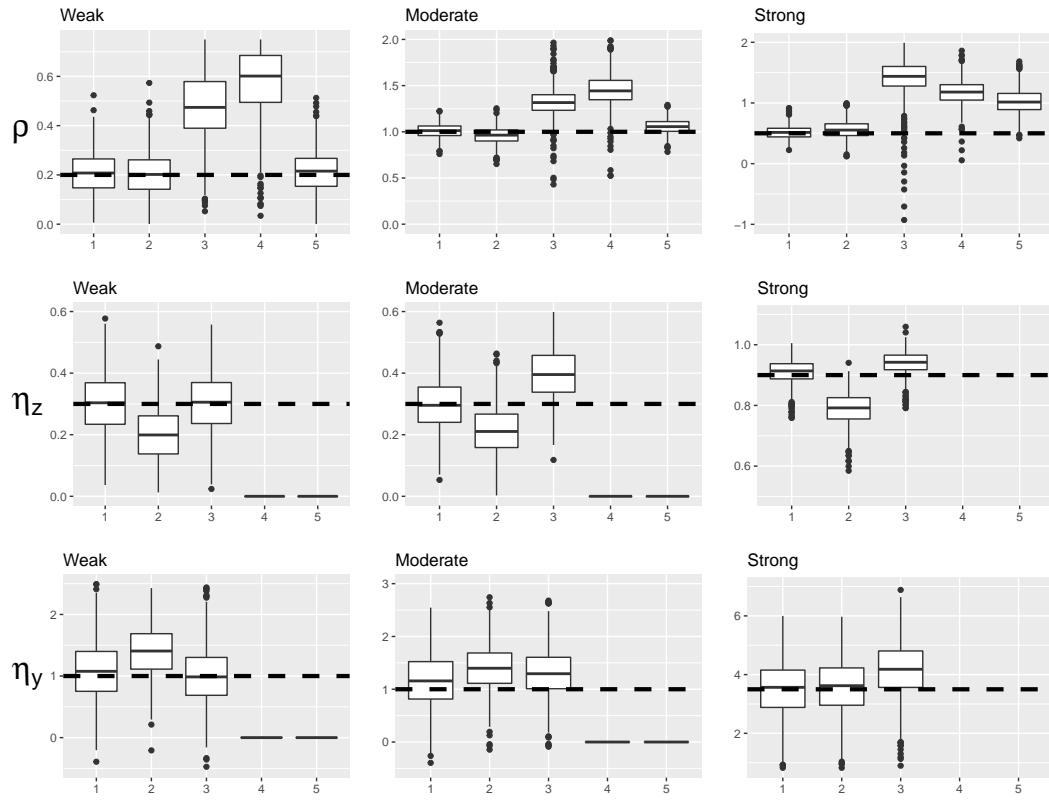
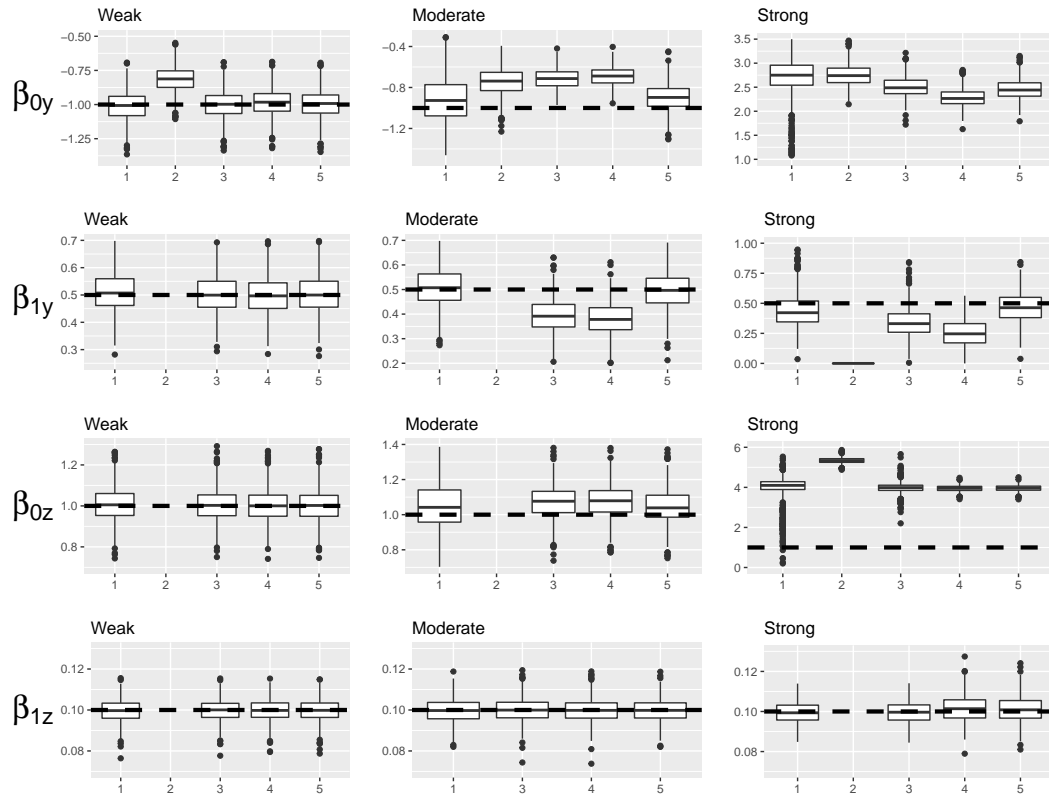Figure 3: Spatial Parameter Estimates for Simulated Data.

Figure 4: Regression Coefficient Estimates for Simulated Data.

5.2 Confidence Interval Coverage

The supplemental material in Caragea and Berg (2014) concluded that the confidence intervals from the parametric bootstrap procedure were slightly too narrow in the bivariate model, and similar results hold for the model considered in this work. For the bivariate mixed model, we investigated the confidence interval construction method for two sets of parameters in an intercept-only model. The parameters we used to simulate data are listed in the second column of Table 3, with one weak dependence regime and one strong dependence regime. For each of the two parameter sets, we simulated M=1,000 data sets and constructed a 95% parametric bootstrap confidence interval was for each. Due to the intensive computation required, we limited the bootstrap sample size to 500 for each interval, and kept every 20th estimate after a burn-in of 300 to ensure independence among samples.

Table 3 displays the means and standard errors of the 1,000 parameter estimates and the proportion of parametric bootstrap intervals that contained the true value used to simulate the data. For the weak regime, the Monte Carlo estimates are approximately unbiased, and the simulated data reflects the true value closely. Only η_y had a notably large standard error, which aligns with the estimation study results from Section 5.1 and the simulation study of the bivariate auto-logistic model from Caragea and Berg (2014). All of the parameters had empirical coverage less than the 95% nominal coverage, which agrees with the result from Caragea and Berg (2014) indicating that the bootstrap intervals are slightly too narrow. The largest under-coverage for the 6 estimates are ρ and σ^2 at -2.1%, which is still relatively close to the desired coverage.

For the strong dependence regime, the simulated data does not reflect the true marginal mean structure. Just as in Section 5.1, δ and μ are not close to the true value used to generate the data due to strong local effects. However, the estimates for the dependence parameters and σ^2 are still approximately unbiased. The empirical 95% interval coverage indicates that

30

the intervals for ρ and η_y are slightly too wide, and the intervals for η_z and σ^2 are too narrow. However, the average empirical coverage for the three dependence parameters is 93.8%, which better aligns with the results for the weak regime.

We can see the breakdown in data generation in Figure 5. The histogram of the 1,000 estimates of μ shows a clearly bimodal distribution. In the right panel, we observe a strong linear relationship between our estimates of δ and ρ (R-squared in a simple linear regression was 0.9223), indicating an influence from the cross-dependence on data generation: the data reflect the conditional distributions rather than the marginal distributions. Additionally, we can see that the 95% bootstrap confidence intervals that failed to contain the true parameter value $mu = 1$ were clustered, while those that were successful in capturing the true value varied over the entire range of simulated data. This relationship is not observed in the simulated data under the weak dependence regime.

In general, the confidence intervals constructed via parametric bootstrap are too narrow, however not prohibitively so. All of the empirical coverages are within approximately 2% of the desired nominal rate when the model is not degenerate. If our interest is in both the covariate and dependence relationships, we have evidence that using the full bivariate spatial model provides reasonable estimates of uncertainty.
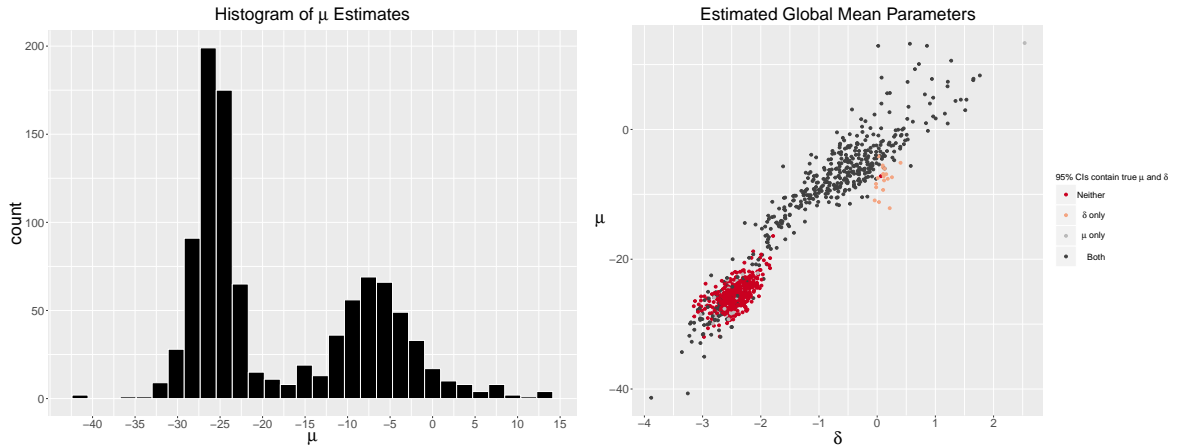

Figure 5: Estimated mean parameters for the 1,000 simulated datasets under the strong dependence regime.

Weak Dependence				
Parameter	**True Value**	**MC Estimate**	**MC Std. Error**	**95% CI Coverage**
ρ	0.2	0.202	0.073	92.9%
η_z	0.5	0.493	0.083	93.2%
η_y	1	0.959	0.419	94.5%
δ	0.5	0.508	0.090	94.1%
μ	1	1.001	0.052	93.6%
σ^2	1	0.996	0.053	92.9%
Strong Dependence				
Parameter	**True Value**	**MC Estimate**	**MC Std. Error**	**95% CI Coverage**
ρ	5	5.024	0.911	96.9%
η_z	0.9	0.901	0.042	86.5%
η_y	3	3.031	0.710	98.0%
δ	0.5	-1.683	1.103	49.4%
μ	1	-17.845	10.419	50.0%
σ^2	100	99.622	5.582	92.6%

Table 3: Bootstrap confidence interval coverage for simulated data

5.3 Marginal Behavior and Cross-Dependence

Here we investigate some of the troubles with interpretation and breakdown due to the cross-dependence ρ discussed in Section 3.3. Because ρ is a dependence parameter, its interpretation and governing behavior might be at least partially influenced by the Gaussian variance σ^2. The goal is to find the relationship such that, for different values of σ^2, similar data can be generated by scaling ρ appropriately.

Mean and dependence parameters were fixed at $\delta = 0.5$, $\mu = 1$, $\rho = 1$, $\eta_z = 0.5$, and

$\eta_y = 1.5$. For the fixed set of mean and dependence parameters, σ^2 was increased from 1 to 96 by 5 (20 values). For each value of σ^2, ρ was multiplied by $\sqrt{\sigma^2}$ or kept as is. For each value of ρ at each σ^2, we simulated 50 datasets ($50 \cdot 20 \cdot 2 = 2000$ total datasets) on a 30 x 30 lattice. For each dataset, we calculated the marginal mean of the binary values and the observed S-value (a moment estimator of the univariate spatial dependence). For further details on the S-value, see Kaiser and Caragea (2009).

Figure 6 displays the results of the simulations. In the left panel, it is clear that the marginal mean of Y is not constant when ρ is kept constant (solid line), and reverts to its marginal mean under independence of 0.6224 as σ^2 increases. When ρ is scaled by multiplying by $\sqrt{\sigma^2}$ (dotted line), the observed mean of the data is relatively constant. In the right panel, the data simulated with $\rho \cdot \sqrt{\sigma^2}$ generates data with relatively constant spatial dependence. The data with ρ held constant at 1 (solid line) is clearly not constant. This indicates that the cross-dependence scales at $\sqrt{\sigma^2}$, and as the conditional variance σ^2 increases with ρ fixed, the cross-dependence goes to 0. The relationship does not appear quite perfect: the observed spatial dependence is not quite constant across values of σ^2 for cross-dependence scaled to $\rho* = \rho \cdot \sqrt{\sigma^2}$.

Next, we investigate the idea of transmission of spatial dependence via ρ. The mean parameters are as before, except now $\sigma^2 = 1$ is fixed and we vary the cross-dependence parameter ρ. For two combinations of η_z and η_y, (0.7, 0) and (0, 2.5), data were simulated across values of ρ from 0 to 1.2. For each value of ρ, 50 datasets were simulated and the observed S-values were recorded for the Gaussian and binary nodes on the grid. Figure 7 displays the results from this simulation.

We expected that spatial dependence might be "transmitted" from the Gaussian to binary field and vice versa via the cross-dependence parameter ρ. In the right panel of Figure 7, we observe that increasing the cross-dependence can increase the observed univariate spatial dependence for binary data, even if the true univariate dependence is 0 (solid line). The
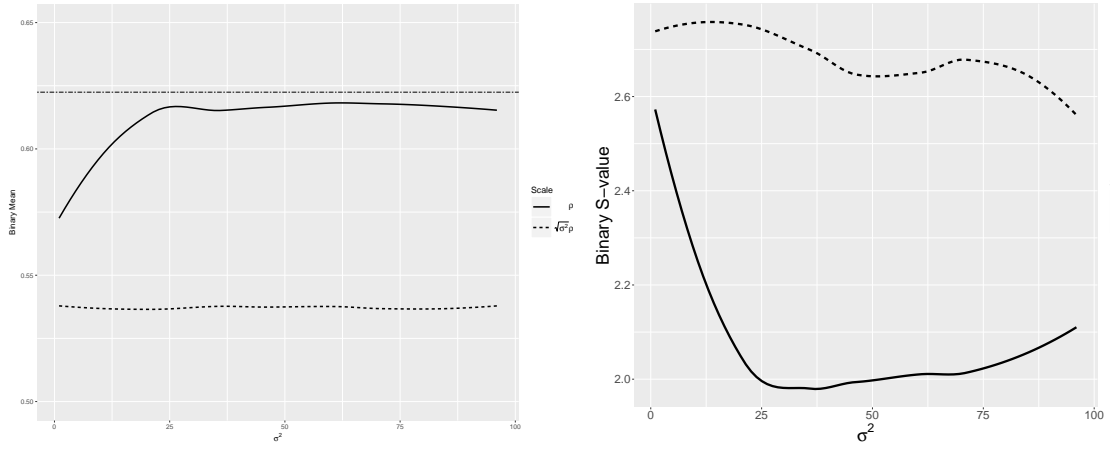
Figure 6: Observed simulated mean (left) and spatial dependence of the binary (right) nodes for increasing σ^2
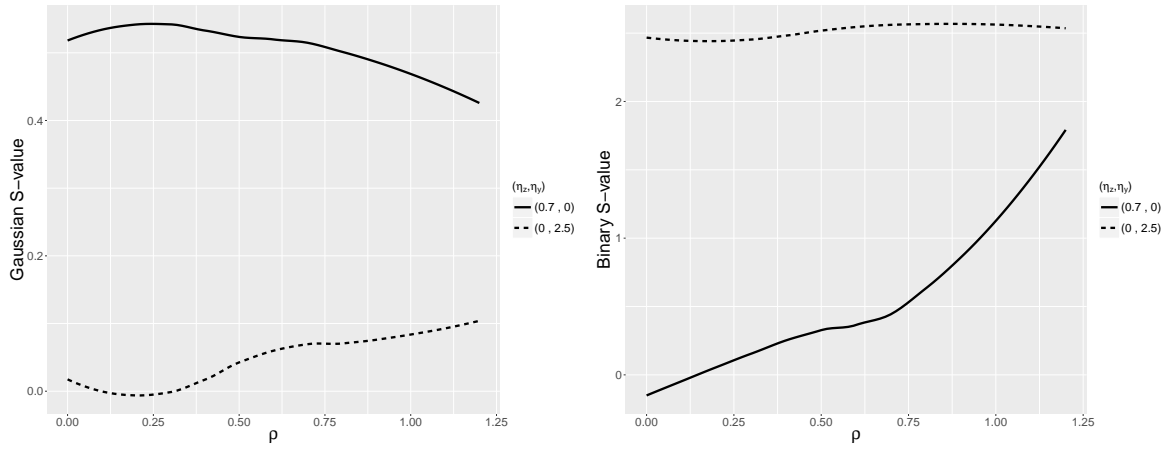

Figure 7: Observed mean spatial dependence of simulated Gaussian (left) and binary (right) nodes for increasing ρ

observed S-value for the binary nodes is 1.5 when $\rho = 1.2$ and $\eta_z = 0.7$. This means that the observed spatial dependence of the $Y(s_i)$s is a moderate 1.5, even though the true value used to simulate the data is actually 0. A similar phenomenon observed for the Gaussian variables in the dotted line in the left panel. Spatial dependence can be created by crossing the dependence structure from one variable type to the other, indicating that the dependence parameters might be at least partially confounded. While confounding would affect interpretation of the spatial structure, it should be noted from Section 5.1 that recovery of the true values is still possible in most cases and the model is still useful for inference.

6   Application to Agricultural Field Trials

The continuous ranked probability score as outlined in Section 4.3 is illustrated on a small dataset taken from an agricultural field trial study. Lado, et al. (2013) provide data on the performance of 384 types of wheat under three different water regimes. While the original study concerned genomic selection of the wheat variations after spatial adjustment, for illustrative purposes this paper only concerns data from the yields. The data from the "Mild Water Stress" regime was used.

The wheat fields used were set on a regular lattice with 40 rows and 20 columns. Variables available were Treatment, Row, Column, Days to Head, Gross Yield, Thousand-Kernel Weight, and Number of Kernels per Spike. Gross Yield (in Thousands of Bushels) is used as the continuous Gaussian variable. We transformed Days To Head to a binary variable by setting the value to 1 if greater than 78 days (the wheat is slow to head) and 0 otherwise. The purpose of the analysis is to explain variation in Gross Yield and Days to Head using spatial and covariate information. After exploratory analysis, only Thousand-Kernel Weight appeared to be significantly related to either response variable after accounting for spatial effects. Thousand-Kernel Weight was used as a covariate on both sides of the model, and other variables were ignored. An image of the two response variables and the covariate is displayed in Figure 8. S-values give moment estimators of the spatial dependence as $S_z = 0.6345$ for the Gaussian data (Gross Yield) and $S_y = 1.1145$ for the binary data (Days to Head).

As in Section 5.1, we estimated parameters for five different models and compared the full model to various reduced forms. In order to avoid possible edge effects, only interior nodes were used for estimation. For each model, we estimated parameters using the MPLE estimation procedure and constructed parametric bootstrap confidence intervals using a bootstrap sample size of 500. We computed the continuous-ranked probability score and corresponding skill separately for the Gross Yield and Days to Head for each model.
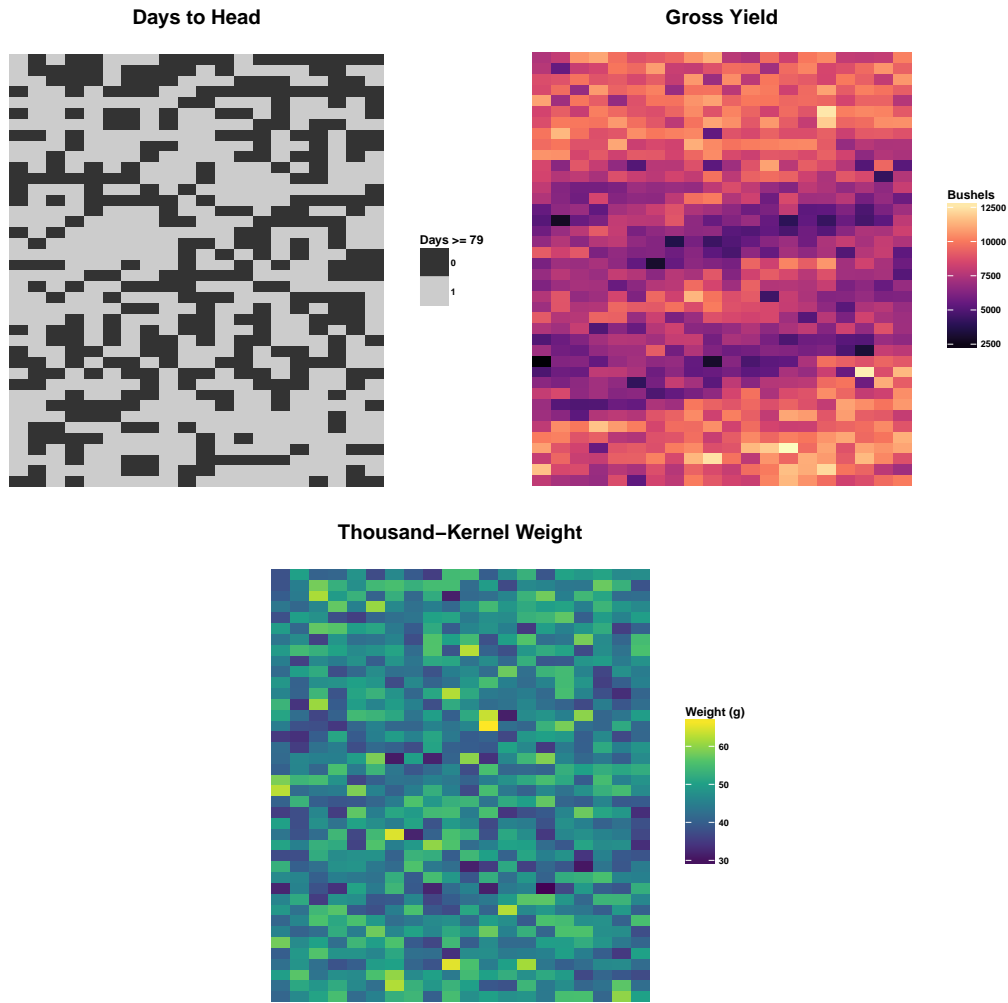
Figure 8: Response Variables and Covariate Used in Wheat Field Trial Study

Table 4 summarizes the results from model estimation. The 95% bootstrap confidence interval for η_y is quite wide and contains 0 in the full model, indicating an uncertain estimate of (possibly non-existent) spatial dependence among the Days to Head response variables on the lattice after accounting for covariates and cross-dependence. The spatial dependence estimates of 0.888 and 0.846 correspond to strong and weak/moderate spatial dependence in the Gaussian and binary variables, respectively. For an interior location, predicted Gross Yield is partially determined by taking $\frac{0.888}{4} = .222$ of the centered Gross Yield of each neighboring location. Estimation of cross-dependence ρ varies widely across the considered models, and is insignificant for the intercept-only model with no covariates. For the full model, $\rho = 0.299$ indicates a moderate cross-dependence between the response variables.

In the univariate spatial model, including Thousand-Kernel Weight as a covariate for Gross Yield indicates a relationship between the response variables that is larger than in the bivariate model, although the estimate is also more variable. For all models estimated, the covariate Thousand-Kernel Weight is a significant predictor for Days to Head (Y), and is only not significant for Gross Yield in the univariate spatial model. In the full model, a 1 gram increase in Thousand-Kernel Weight is associated with a 0.14 day decrease in number of days before heading and a .022 x 1000 = 22 bushel increase in Gross Yield, after accounting for other dependence.

The estimated Gaussian conditional variance σ^2 is minimized for the full model, indicating that the full model with both spatial and covariate information accounts for the most variation in Gross Yield. Using the continuous ranked probability score, the full model has the best in-sample performance among the five models. The non-spatial models perform significantly worse than the reduced spatial models for prediction of Gross Yield, but significantly better for the binary variable Days to Head. The full model is the only of the five options that exhibits non-negligible improvement over the constant mean model, thus we would select the full model using this criterion. The relative similarity between the full and univariate spatial models indicates that cross-dependence does not have a drastic effect on the model. While the full model is one of the best by the CRPS criterion, skill scores of 16.74 and 34.23 represent only modest improvements over the constant mean model.

7   Greater Prairie Chicken Presence and Vegetative Cover

The following implementation of the mixed bivariate MRF model illustrates some of the model's benefits in estimation of spatial dependence in the presence of strong dependence.

Model	ρ	η_z	η_y	Binary (Y) Intercept	1000-Kernel Wt (Y)	Gaussian (Z) Intercept	1000-Kernel Wt (Z)	σ^2
Full	0.299 (0.127 , 0.474)	0.888 (0.795 , 1.008)	0.846 (0.000 , 1.947)	6.591 (5.109 , 7.794)	-0.137 (-0.162, -0.106)	6.842 (5.915 , 7.606)	0.022 (0.008 , 0.036)	1.387
Intercept Only Spatial	0.126 (-0.026 , 0.264)	0.900 (0.809 , 0.996)	1.001 (0.135 , 2.001)	0.336 (0.130 , 0.534)	-	8.014 (7.496 , 8.384)	-	1.449
Univariate	-	0.618 (0.634 , 0.921)	1.232 (0.622 , 2.710)	6.600 (4.319 , 7.104)	-0.135 (-0.145 , -0.087)	7.065 (6.969 , 8.370)	0.019 (-0.011 , 0.019)	1.448
Univariate Non-Spatial	-	-	-	6.561 (5.177 , 7.818)	-0.134 (-0.161 , -0.105)	5.330 (4.435 , 6.250)	0.057 (0.038 , 0.077)	2.907
Bivariate Non-Spatial	0.233 (0.282 , 0.487)	-	-	6.873 (5.328 , 8.266)	-0.141 (-0.171 , -0.110)	5.442 (4.485 , 6.507)	0.055 (0.032 , 0.076)	2.822

Model	$CRPS_y$	$Skill_y$	$CRPS_z$	$Skill_z$
Full	0.20	16.74	0.66	34.23
Intercept-Only Spatial	0.24	1.62	0.66	33.46
Univariate	0.21	14.82	0.70	29.57
Univariate Non-Spatial	0.22	8.11	0.97	2.59
Bivariate Non-Spatial	0.22	10.44	0.96	3.37

Table 4: Results of Model Estimation for Wheat Field Trial Data

7.1 Data Description

The greater prairie chicken, *Tympanuchus cupido*, is an endangered species of bird endemic to the tallgrass prairies of the central United States. Greater prairie chickens prefer to make nests in areas with large amounts of vegetative cover and do not migrate, thus we might be able to predict greater prairie chicken presence by measuring the condition of vegetation in the area. Moreover, presence of the greater prairie chicken might spark species conservation efforts or other causes that could encourage the preservation of the native tallgrass prairie ecosystems against the alternative of conversion to farmland or developments. Thus we hypothesize a positive relationship between vegetative cover and greater prairie chicken presence in an area.

We constructed a dataset to test this hypothesis by aggregating data from 3 online sources. Data on sightings of greater prairie chickens in North America were obtained from the Global Biodiversity Information Facility (GBIF) via the Cornell Lab of Ornithology's eBird database. Data on vegetation condition, specifically the Normalized Difference Vegetation Index (NDVI), is collected via satellite and are available from the USDA National Agricultural Statistics Service's VegScape explorer. This is a measure of the relative greenness of the biome, and is thus a good indicator of the amount of vegetation cover. In addition to these, we collected data on the soil conditions from the National Resource Conservation Service's Web Soil Survey

database for use as possible covariates in the model.

In order to keep the illustration to a reasonable sample size, we focus on a portion of Lincoln County in western Nebraska. The satellite NDVI data is collected on a fine scale, so we aggregated the data onto a 31 x 31 grid and overlaid the soil data. The grid corresponds to areas of approximately 2 square miles for each point. The observations of the prairie chickens are limited to those since 1990 and the spring mating season months of March, April, and May, in which they are more active and more likely to be spotted. We used a snapshot of the NDVI from April 2018 as a representative example of the vegetation cover during this time of year.
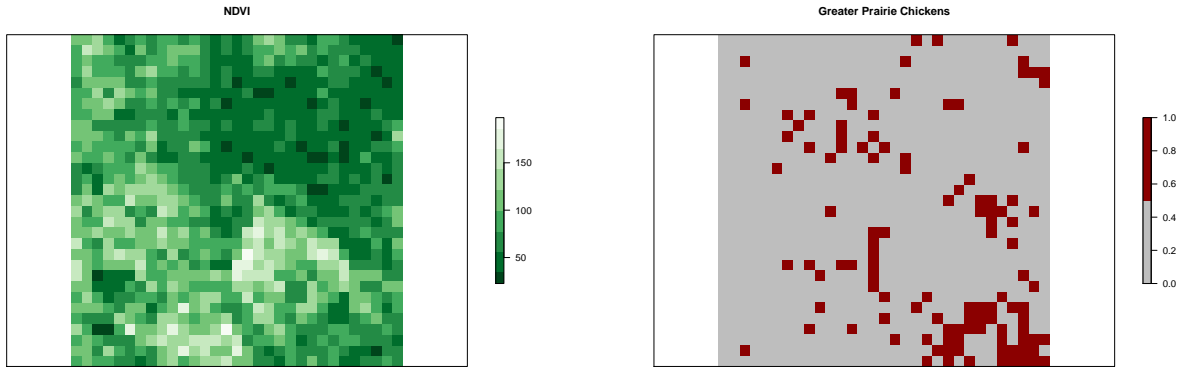

Figure 9: NDVI (left) and Greater Prairie Chicken observations (right), after aggregation

7.2 Data Analysis

We modeled the relationship between NDVI (Gaussian response) and greater prairie chicken presence (binary response, with 1 indicating an observation at that location since 1990) as a bivariate spatial model. Exploratory analysis of the data indicates strong dependence for both variables, with S-values of $S_z = 0.994$ for the NDVI and $S_y = 3.308$ for the observed greater prairie chickens. The correlation between the response variables is 0.07, indicating a weak but possibly non-zero relationship between the response variables. Naive (non-spatial) linear regression with the covariates indicated significant linear relationships with several of the soil property covariates. We considered soil variables wind erodibility index, soil pH, percent sand

Parameter	Estimate	Std. Err.	90% Bootstrap C.I.
Y-Intercept($\beta_{0,y}$)	-0.417	1.962	(-2.343 , 4.183)
Wind Erodibility Index ($\beta_{1,y}$)	-0.005	0.006	(-0.017 , 0.002)
% Sand ($\beta_{2,y}$)	0.024	0.018	(0.003 , 0.063)
Water Supply ($\beta_{3,y}$)	0.019	0.045	(-0.061 , 0.083)
Z-Intercept($\beta_{0,z}$)	204.543	15.970	(-43.172 , 66.560)
σ^2	16.765	0.949	(15.246 , 18.245)
η_y (Presence-Absence)	3.613	0.697	(2.356 , 4.614)
η_z (NDVI)	0.991	0.017	(0.955 , 1.012)
ρ (Cross-dependence)	0.446	.260	(0.000 , 0.844)

Table 5: Summary of model estimates for greater prairie chicken / vegetation data

composition, and water supply rating. All were significant in a multiple linear regression with NDVI as the response, however none were significant in a multiple logistic regression with prairie chicken presence as the response.

We estimated the model with several different combinations of the soil covariates, implementing the pseudo-likelihood estimation and parametric bootstrap (sample size M=500) for each. Results from model estimation and bootstrap inference for our selected final model are presented in Table 5. We found that after considering spatial dependence, the only significant relationships with the response variables were between percent sand and the presence of greater prairie chickens. This was contrary to our initial analysis of non-spatial multiple linear and logistic regressions, indicating that the presence of spatial dependence has a significant influence on the conclusions of the analysis. Given the results on simulated data in Section 5.1, the strong dependence in this data should make us hesitant to draw too many conclusions about the covariate relationships. Moreover, even the statistically significant covariate relationship does not appear to be practically significant.

The estimate of binary spatial dependence was $\eta_y = 3.613$, indicating strong dependence in the binary field but well below the upper bound as outlined in Section 3.3. The estimate of spatial dependence in the continuous variable is $\eta_z = 0.991$, which is very near the boundary of the parameter space. This indicates very strong dependence among the NDVI values, which we might expect given the aggregation of the original fine-scale measurements in the data. The

cross-dependence among the response variables is estimated to be $\rho = 0.446$, which is quite strong when we consider the presence of the other strong spatial dependence parameters. Here we note that the estimates of dependence are close to the strong parameter set used to study estimation and confidence interval procedures in Section 5.

The estimate of ρ suggests a significant cross-dependence, thus we conclude that there is a benefit to using a bivariate spatial model in this analysis. Just as in the initial exploratory analysis, there is a positive correlation between our two response variables: it is more likely that we observe greater prairie chickens in areas with heavier vegetation cover. This (mildly) supports the hypothesis that greater prairie chickens prefer heavier vegetative cover to make their nests.

The fitted model leads to a $CRPS_y$ of 0.1254 and a corresponding skill score of 21.81 for the binary data. For the NDVI data, we have $CRPS_z$ of 2.25 with corresponding skill score of 55.76 for the Gaussian data. This is a very notable improvement over a constant mean model.

8   Conclusion

This work proposes a Markov random field model with bivariate response variables that belong to different distributions within the exponential family. The model follows from previous work in univariate Markov random field models and network models, with a constructed joint density function that follows the familiar form of one-parameter exponential families often used in generalized linear models. This joint density builds from conditionally-specified distributions on the nodes of the graph. Estimation and inference relies on maximum pseudo-likelihood and parametric bootstrap, following the node-conditional specification. The MPLE estimation and parametric bootstrap are relatively simple to implement, but can be computationally demanding and may not be reliable when the spatial dependence is very strong.

There are many readily apparent extensions of the rather simple model proposed in Section 3. Additional dependence structures, such as extended neighborhoods or temporal depen-

dence beyond that described in Section 2.1, can be modeled by adding additional terms to the node-conditional distributions as in (8) and (10) or generalizing the indices of the dependence parameters. The work in mixed graphical structures discussed in Section 2.3 could also allow the extension of the bivariate model to higher-dimensional responses with perhaps more than 2 different distribution types, such as nodes with Poisson or even Gamma-family distributions. While the initial stage of specifying node-conditional distributions is relatively straightforward, caution must be taken to ensure that the conditional distributions can then form a valid joint distribution.

Even if an extended model satisfies a valid joint distribution, model degeneracy poses a larger obstacle to the proposed possible extensions. With each additional spatial or temporal neighbor added to the model, the conditional structure has another constraint that must be satisfied in order to ensure observed marginal behavior matches expected marginal behavior. As discussed in Section 3.3, determining parameter spaces for the spatial dependence structure that ensures reliable estimation and interpretation is an open question. Moreover, different structures require different centering schemes to allow for interpretable model parameters. Adding different distribution types to the model requires significant analytical work that is not considered here. We encountered difficulties with the four-nearest neighbor with one cross-dependent neighbor structure here, which is perhaps the simplest case of this type. More complex models certainly require even more caution, despite our flexible method for specifying valid distributions.

Even though we have some open issues with model behavior, the multivariate Markov random field model still has many favorable properties. The flexible conditionally-specified model structure, which reduces to node-wise regression models, allows easy modeling of any one-parameter exponential family distribution with separated structures for covariates and dependence parameters. The computation required for estimation and inference is relatively simple

when compared to alternatives, while still allowing for easy-to-interpret model parameters. When the purpose of an analysis is prediction at unobserved locations, other models may more adequately capture variation due to spatial dependence. The bivariate mixed Markov random field model is most appropriate when quantifying and interpreting the spatial dependence structure is integral to understanding the scientific question of interest.

References

[1] Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. Series B (Methodological), 36(2), 192-236.

[2] Besag, J. (1975). Statistical Analysis of Non-Lattice Data. Journal of the Royal Statistical Society. Series D (The Statistician), 24(3), 179-195.

[3] Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 43: 1.

[4] Caragea, P.C. and Berg, E. (2014). A centered bivariate spatial regression model for binary data with an application to presettlement vegetation data in the Midwestern United States. J. Agric. Biol. Environ. Stat. 19, 453471.

[5] Chen, S., Witten, D. M., and Shojaie, A. (2014). Selection and estimation for mixed graphical models. Biometrika, 102(1), 47-64.

[6] de Leon, A. R. and Wu, B. (2011), Copula-based regression models for a bivariate mixed discrete and continuous outcome. Statist. Med., 30: 175-185.

[7] Gelfand, A., and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics, 4(1), 11-15.

[8] GBIF.org (31st October 2018) GBIF Occurrence Download https://doi.org/10.15468/dl.powetx

[9] Genest, C. and Neslehova, J. (2007). A Primer on Copulas for Count Data. ASTIN Bulletin, 37(2), 475-515.

[10] Gneiting, T. and Raftery, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association, 102:477, 359-378.

[11] Griffith, Daniel. (2002). A Spatial Filtering Specification for the Auto-Poisson Model. Statistics & Probability Letters. 58. 245-251. 10.1016/S0167-7152(02)00099-8.

[12] Guan, Y. and Haran, M. (2018). A Computationally Efficient Projection-Based Approach for Spatial Generalized Linear Mixed Models, Journal of Computational and Graphical Statistics.

[13] Hardouin, C. and Cressie, N. (2018). Two-scale Spatial Models for Binary Data. N. Stat Methods Appl (2018) 27: 1.

[14] Haslbeck, J.M.B, and Waldorp, L.J. (2015). Structure estimation for mixed graphical models in high dimensional data. arXiv:1510.05677.

[15] Hodges, J. and Reich, B. (2010). Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love. The American Statistician, 64:4, 325-334.

[16] Hughes, J. (2015) copCAR: A Flexible Regression Model for Areal Data. Journal of Computational and Graphical Statistics, 24:3, 733-755.

[17] Hughes, J., and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75, 139159.

[18] Hughes, J. , Haran, M. and Caragea, P. C. (2011), Autologistic models for binary data on a lattice. Environmetrics, 22: 857-871.

[19] Jordan, A., Kruger, F., and Lerch, S. (2016). scoringRules: Scoring Rules for Parametric and Simulated Distribution Forecasts. Available at https://cran.r-project.org/web/packages/scoringRules

[20] Kaiser, M.S. (2007). Statistical Dependence in Markov Random Field Models. Preprint 2007-1, Department of Statistics, Iowa State University, Ames, Iowa.

[21] Kaiser, M. S. and Caragea, P. C. (2009). Exploring Dependence with Data on Spatial Lattices. Biometrics, 65: 857-865.

[22] Kaiser, M., Caragea, P., and Furakawa, K. (2012). Centered Parameterizations and Dependence Limitations in Markov Random Field Models. Journal of Statistical Planning and Inference, Volume 142, Issue 7, 2012, 1855-1863, ISSN 0378-3758.

[23] Kaiser, M., and Cressie N. (2000). The Construction of Multivariate Distributions from Markov Random Fields. Journal of Multivariate Analysis. 73. 199-220. 10.1006/jmva.1999.1878.

[24] Lado, B., Matus, I., Rodrguez, A., Inostroza, L., Poland, J., Belzile, F., del Pozo, A., Quincke, M., Castro, M., von Zitzewitz, J. (2013). Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3 (Bethesda, Md.), 3(12), 2105-14. doi:10.1534/g3.113.007807

[25] Lee, J. and Hastie, T.J. (2015). Learning the Structure of Mixed Graphical Models. Journal of Computational and Graphical Statistics, 24:1, 230-253.

[26] Liang, F. (2010). A double Metropolis−Hastings sampler for spatial models with intractable normalizing constants. Journal of Statistical Computation and Simulation, 80:9, 1007-1022.

[27] Mardia, K. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. Journal of Multivariate Analysis, 24(2), 265-284.

[28] Musgrove, D. R., Hughes, J., and Eberly, L. E. (2016). Hierarchical copula regression models for areal data. Spatial Statistics, 17, 38-49.

[29] Brian J. Reich, Hodges, J., & Vesna Zadnik. (2006). Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models. Biometrics, 62(4), 1197-1206.

[30] Rue, H., Held, L., Reid, N., Keiding, N., Tibshirani, R., Louis, T., Tong, H., Isham, V. (2005). Gaussian Markov Random Fields. New York: Chapman and Hall/CRC.

[31] Sain, S., Furrer, R., and Cressie, N. (2011). A spatial analysis of multivariate output from regional climate models. The Annals of Applied Statistics, 5(1), 150-175.

[32] Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at the following link: https://websoilsurvey.sc.egov.usda.gov/. (Accessed: 31 October 2018).

[33] Sun, L., and Clayton, M. (2008). Bayesian Analysis of Crossclassified Spatial Data with Autocorrelation. Biometrics, 64(1), 74-84.

[34] Yang, Z., Yu, G., Di, L., Zhang, B., Han, W., and Mueller, R. (2013). Web service-based vegetation condition monitoring system - VegScape, Geoscience and Remote Sensing Symposium (IGARSS) 2013 IEEE International, 3638-3641.

[35] Wall, Melanie. (2004). A close look at the spatial structure implied by the CAR and SAR models. Journal of Statistical Planning and Inference. 121. 311-324. 10.1016/S0378-3758(03)00111-3.

[36] Yang, E., Baker, Y., Ravikumar, P., Allen, G. and Liu, Z.. (2014). Mixed Graphical Models via Exponential Families. Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, in PMLR 33:1042-1050

A Supplemental Model Development Discussion

A.1 Proof for a Valid Joint Distribution

Proof. The joint distribution $p(\mathbf{y}, \mathbf{z})$ is valid if it is integrable, i.e, that the denominator in (13) is finite over the paramter space.

Because any binary variable has only 2 outcomes, for a lattice with n locations we have 2^n possible outcomes for n binary variables, thus the sum over the binary nodes is finite. We then condition on the binary nodes and show that this joint conditional distribution is integrable is finite.

Conditioning on the binary nodes, the resulting conditional expectation for the remaining Gaussian nodes as in (10) conditional upon all other locations can then be written as $E[z(s_i)|\mathbf{z}(\bar{s}_i)] = \mu_i^* + \frac{\eta_z}{M_i}\sum_{j\in N_i}(z(s_j) - \mu_j*)$, where $\mu_i^* = \mu_i + \rho(y(s_i) - \frac{\exp(\delta_i)}{1+\exp(\delta_i)})$ and the full joint distribution for the Gaussian nodes conditioned on the binary values, with common variance σ^2, is multivariate Gaussian as in equation (4.10) of Besag (1974):

$$p(\mathbf{z}|\mathbf{y}) \propto (2\pi\sigma^2)^{-n/2}|\boldsymbol{H}|\exp\{\frac{-1}{2\sigma^2}(\mathbf{z} - \boldsymbol{\mu}^*)^T \boldsymbol{H}(\mathbf{z} - \boldsymbol{\mu}^*)\} \tag{19}$$

where $\mathbf{z} = (z(s_1), ..., z(s_n))^T$, $\boldsymbol{\mu}^* = (\mu_1*, ..., \mu_n*)$, and $\boldsymbol{H}$ is an n x n matrix with 1 on the diagonals and $\boldsymbol{H}_{ij} = \frac{-\eta_z}{M_i}$ for $j \in N_i$, $j \neq i$, all other entries being 0.

If $\sum_{i=1}^{M_i}\frac{\eta_z}{M_i} = \eta_z < 1$, then $\boldsymbol{H}$ is strictly diagonally dominant, and thus $\boldsymbol{H}$ is non-singular. By properties of multivariate Gaussian distributions, non-singular $\boldsymbol{H}$ implies that (19) is a valid joint probability density and is integrable. Because the binary values are defined only on 2^n outcomes, if $\eta_z < 1$ then we have $\int_{\Omega_{Y,Z}} p(\mathbf{y}, \mathbf{z})d\mathbf{y}d\mathbf{z} = \sum_{\Omega_Y}\int_{\Re} p(\mathbf{z}|\mathbf{y})d\mathbf{z}d\mathbf{y} < \infty$ and thus the full joint distribution is valid. $\square$

A.2 A Relationship Between Cross-dependence and Spatial Effects on Marginal Behavior

In what follows, we provide some speculative preliminary investigation between the various dependence parameters in the bivariate mixed MRF model. We borrow from the ideas of Kaiser (2007) and the interpret ρ as a covariance parameter between Y and Z to find a standard bound for the cross-dependence parameter. Since we have centered the data, ρ is then a reflection of a relationship between spatial parameters. For a fixed spatial dependence, we wish to find a function of η_y and η_z (i.e., find values of ρ such that the data generated look similar for varying η_y and η_z). We can use the definition of correlation, and fix the value of the correlation and solve for covariance:

$$Corr\left[z(s_i) - \mu, y(s_i) - logit^{-1}(\delta)\right] = \frac{Cov\left[z(s_i)-\mu, y(s_i)-logit^{-1}(\delta)\right]}{\sqrt{Var(z(s_i)-\mu)}\sqrt{Var(y(s_i)-logit^{-1}(\delta))}}$$

$$\approx \frac{\rho}{\sqrt{Var(z(s_i)-\mu)}\sqrt{Var(y(s_i)-logit^{-1}(\delta))}}$$

$$\approx \frac{\rho}{\sqrt{\eta_z}\sqrt{\eta_y}}$$

We can then hold the left side constant (at 1, for example), and then solve for ρ to get $\rho = \sqrt{\eta_z}\sqrt{\eta_y}$.

We now wish to find an "upper bound" for ρ, denoted here as $\ddot{\rho}$, such that $|\rho| < \ddot{\rho}$ ensures non-degenerate model behavior. We start at the bounds for our univariate spatial parameters and work backwards, so that under independence ($\eta_z = \eta_y = 0$) the marginal expectations of the model approximately match the conditional expectations. For univariate models, recall from Section 3.3 that the upper bound for η_z is $\ddot{\eta}_z = 1$ and the upper bound for η_y depends on δ with $\ddot{\eta}_y \geq 4$ (Figure 1). With these two values, we propose an upper bound for $\ddot{\rho}$:

$$\ddot{\rho}/\sqrt{\sigma^2} = \sqrt{(\ddot{\eta}_z)(\ddot{\eta}_y)}$$

$$\implies \ddot{\rho} = \sqrt{(\sigma^2 \ddot{\eta}_y)} \tag{20}$$

Under constant variance (say, $\sigma^2 = 1$), (20) becomes $\ddot{\rho} = \sqrt{\ddot{\eta}_z \ddot{\eta}_y} = \sqrt{\ddot{\eta}_y}$.

The standard bound given by (20) is a bound for ρ when the $\eta_z = \eta_y = 0$ (i.e., a non-spatial model), such that the model degenerates as ρ approaches or passes $\ddot{\rho}$. For the more useful case in which the spatial dependence parameters are non-zero, we can work backward from (20) to find a function for the upper bound for ρ. With $\ddot{\eta}_z$ and $\ddot{\eta}_y$ as defined previously, we simply take the upper bounds and subtract the actual parameter values to get our function:

$$\rho\left(\eta_z, \eta_2\right) = \sqrt{\sigma^2}\sqrt{(1 - \eta_z)(\ddot{\eta}_y - \eta_y)} \tag{21}$$

which for $\sigma^2 = 1$ simplifies to

$$\rho\left(\eta_z, \eta_y\right) = \sqrt{(1 - \eta_z)(\ddot{\eta}_y - \eta_y)} \tag{22}$$

With (21) and (22) we now have a useful result which allows us to quantify the dependence across our two variables of interest in the context of the independent spatial dependence parameters. For varying levels of η_z and η_y, our model will degenerate as the realized ρ approaches $\rho(\eta_z, \eta_y)$. Additionally, the proposed function in (21) also implies that we can generate similar realizations of bivariate spatial data for different values of η_z and η_y by adjusting ρ. We could generate data such that the $y(s_i)$ values behave as if strongly dependent (i.e., $\eta_y > 3.5$) even with a true value of $\eta_y = 0.1$ by varying our ρ and η_z accordingly through (22).

Because the smallest upper bound for η_y is $\ddot{\eta}_y = 4$, we speculate an uniform upper bound for ρ as a function of the other dependence parameters as $\ddot{\rho} = \sqrt{\sigma^2}\sqrt{(1 - \sigma^2 \eta_z)(4 - \eta_y)}$, or scaled/reparameterized as $\ddot{\rho} = \sqrt{(1 - \eta_z)(4 - \eta_y)}$. This bound works reasonably well when dependence is weak to moderate, in that we can simulate data that looks approximately the same for varying values of η_y and η_z by adjusting ρ such that the the equality above holds. This relationship breaks down for larger values of η_y and η_z, but appears a more reasonable

upper bound for ρ than the conservative suggestion in Section 3.3. The above derivation is only an outline and is not rigorous, but seems to be more plausible.

We provide some illustrations of the relationship here, by assessing the marginal (observed) behavior of simulated data for varying levels of the spatial dependence parameters. The simulation study process was as follows:

1. Fixed parameters $\eta_y = 1$, $\sigma^2 = 10000$, $\mu = 1$, and $\delta = 0.5$

2. Calculate a "baseline" relationship by plugging $\rho = 100$ and $\eta_z = 0$ to (21), to get $\rho_b = 1.749871$.

3. For each $\eta_{z*} \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$:

   - Generate 20 datasets for two different values of the cross-dependence parameter: $\rho_1 = 100$ and $\rho_2 = 100 \cdot \frac{\rho_*}{\rho_b}$, where ρ_* is the value from plugging the respective η_{z*} value into (21)

   - Calculate average observed sample means and S-values (observed spatial dependence)

If our relationship in (22) is reasonable, we would expect that the observed data approximately reflects the expected marginal behavior under independence. Figure 10 displays results from the simulations. We can see that, as we increase η_z, keeping ρ_1 constant causes the observed marginal mean and observed spatial dependence (solid lines) of the binary data to drift away from the expected mean under independence. The data generated using the transformed ρ_2 (dashed lines) produces approximately constant mean and spatial dependence across η_z values, indicating that for constant σ^2, we can generate data that looks the same for different combinations of spatial dependence parameters! We only need to adjust using (21).

We then performed the same procedure for varying levels of η_y. Here, we fixed $\eta_z = 0.5$ and used the sequence of $\eta_{y*} \in \{0, 0.2, 0.4, \ldots, 3.8, 4\}$, with the other parameters as before. Again,

we generated 20 datasets for $\rho_1 = 100$ and $\rho_2 = 100 \cdot \frac{\rho_*}{\rho_b}$, where ρ_* values are determined by inputting each η_{y*} into (21). The results, displayed in Figures 11 and 12, were similar: the data generated using constant ρ_1 (solid lines) drifted away from expected marginal behavior. By transforming to ρ_2 (dashed lines), we were able to generate data that was approximately constant across all values of η_y*.
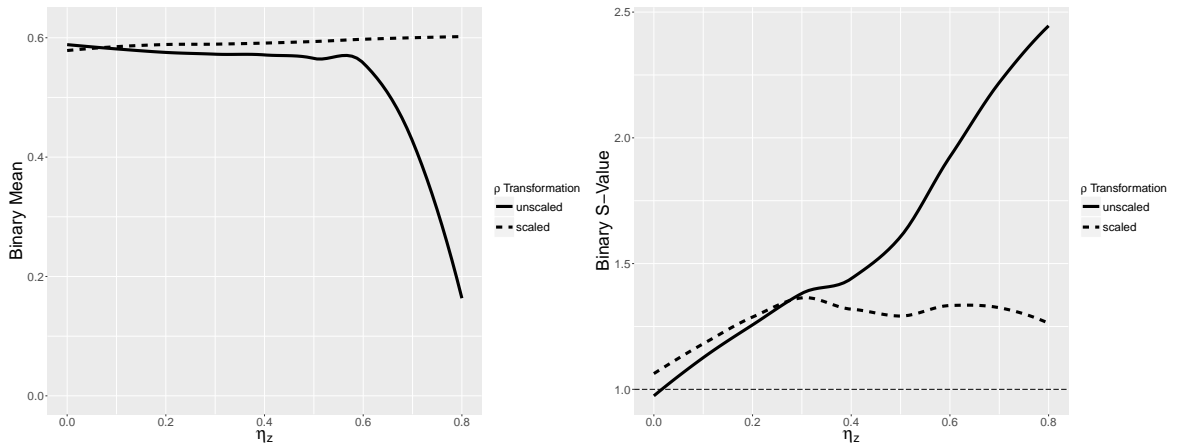

Figure 10: Observed simulated mean (left) and spatial dependence (right) of the binary nodes for increasing η_z, for two choices of ρ
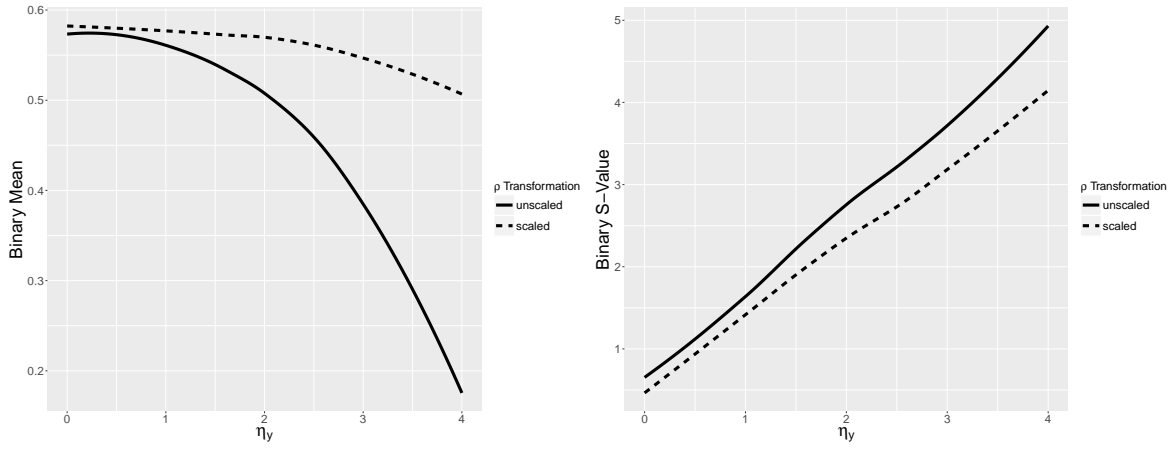
Figure 11: Observed simulated mean (left) and spatial dependence (right) of the binary nodes for increasing η_y, for two choices of ρ
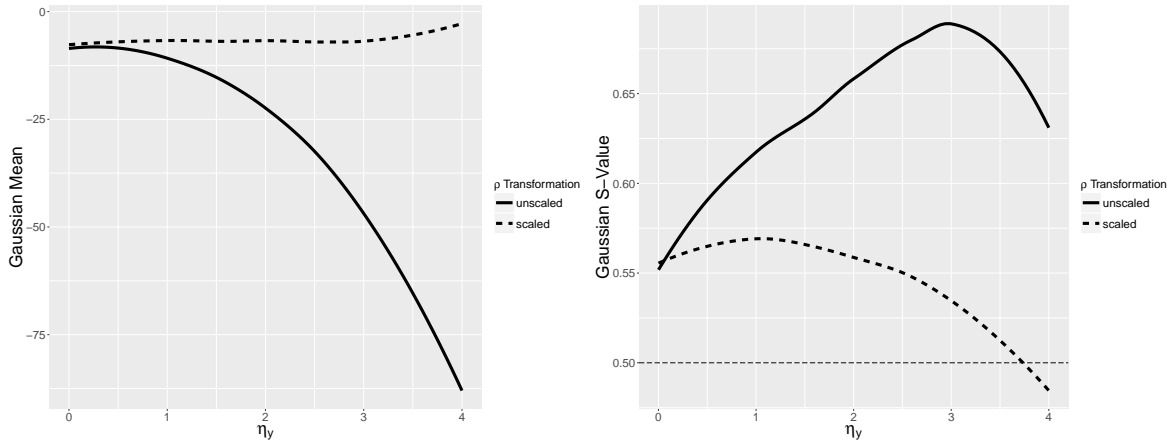

Figure 12: Observed simulated mean (left) and spatial dependence (right) of the Gaussian nodes for increasing η_z, for two choices of ρ

Unfortunately, we also see in the plots that the relationship is not perfectly linear, only approximate. Moreover, this relationship tends to break down when η_y or η_z are near their univariate bounds, as the within-type spatial dependence overwhelms the cross-dependence effects. Despite this and the rather non-rigorous derivation, it appears there is a reasonable approximation for the relationship.

A.3 Alternative Models for Lattice Data

Many alternatives to the auto-models have been considered for lattice data, the two most prominent being the spatial generalized linear mixed model and copula models. In the spatial generalized linear mixed model (SGLMM) as in Besag, York, and Mollie (1991), variability due to spatial dependence is modeled as spatial random effects $\boldsymbol{\psi} = \{\psi_1, ...\psi_n\}$. Typically, these

random effects are assumed to come from a zero-mean Gaussian random field as in Rue and Held (2005) and connected via an adjacency $\mathbf{A}$ or spatial weights matrix $\mathbf{W}$ that defines the graph structure. The distribution of the response variables conditional on these spatial random effects and possible fixed covariate effects is then modeled as an suitable distribution in the typical procedure of a generalized linear mixed model. This model can then be extended to a multivariate model as in Gelfand and Vounatsou (2003) with a Gaussian MRF of random effects. The specification of the SGLMM's covariance matrix $\mathbf{\Sigma}$ via design of $\mathbf{A}$ or $\mathbf{W}$ provides flexibility and leads to improved prediction capability over auto-models, but at a cost.

Estimation of the random effects $\boldsymbol{\psi}$ is computationally burdensome for large lattices due to the required repeated inversion of $\mathbf{\Sigma}$ or the (typically sparse) precision matrix $\mathbf{Q}$ in maximum likelihood estimation Hughes and Haran (2013). (Guan and Haran (2018)) or filtering via the addition of synthetic covariates based on spatial eigenvectors (Griffith (2002)). The dimension reduction methods for SGLMMs also partially alleviate the more troublesome property of confounding (Hughes and Haran (2013)), which inhibits the recovery of regression parameters (Hodges and Reich (2010)). In addition to this confounding, the modeling of spatial dependence via random effects can lead to counter-intuitive interpretations of spatial correlation (Wall (2004)).

Spatial dependence can also be incorporated into regression models via copulas. One method is the copCAR (Hughes (2015)), in which the CAR precision matrix $\mathbf{Q}$ (as in the SGLMM) or other desirable covariance structure (Musgrove, Hughes, and Eberly (2016)) is modeled as the Gaussian copula and then linked from the copula to a *marginal* distribution for the responses via the probability integral transform. The copula model is flexible in that it allows for positive spatial dependence for any type of response distribution. It is also a marginal regression model, in that regression coefficients have interpretations in terms of the marginal distribution of the response instead of conditional distributions (as in auto-models and SGLMMs). Another

convenient property of copula-based models, as in de Leon and Wu (2011), is the extension to the multivariate mixed response distributions considered in this work. One could conceivably use copulas to model a wide range of dependence and response structures without requiring the unique analytical work to justify each model (as in Section 3 here).

The class of auto-models considered in Sections 2 and 3 have many preferable properties despite the favorable characteristics of the models considered above. In 10 and 8, the spatial dependence parameters η_y, η_z, and ρ are modeled directly in the scale of the response distributions, rather than through the specification of the latent Gaussian random fields above. Genest and Neslehova (2007) point out that for discrete marginal distributions (e.g. binary), the joint distribution is only defined on the support for the marginals and thus interpretation of dependence parameters is subject to extreme caution. Because we have directly specified via conditional distributions in (8)-(10), we avoid the interpretability/identifiability issues of copula models. Even if dependence is now defined in terms of conditional instead of marginal means, spatial dependence in automodels has an intuitive interpretation in terms of neighboring values rather than the somewhat nebulous random effects of SGLMMs and copula models. Moreover, the regression coefficients (global structure) are not confounded with the dependence parameters (local structure) in the centered model, such that the regression coefficients $\boldsymbol{\beta}$ have convenient interpretations in terms of marginal means (like copula models) rather than the conditional means as in SGLMMs (Caragea and Berg (2014)).

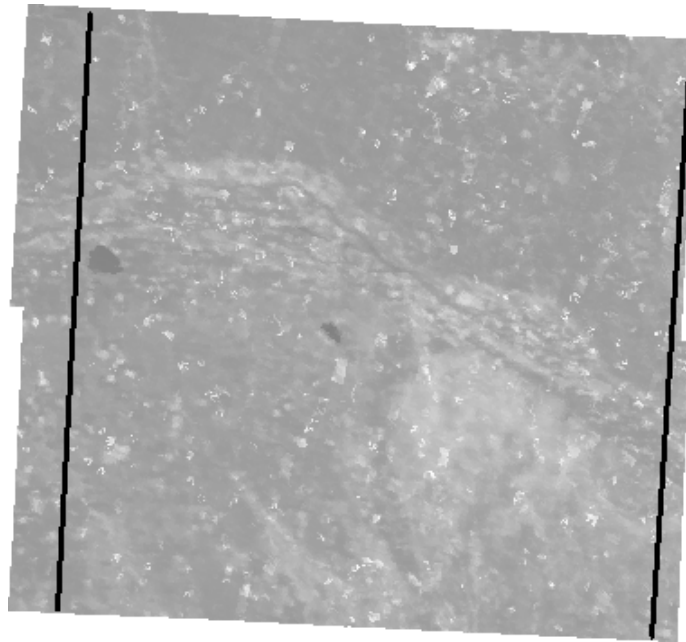# B   Supplementary Figures for Prairie Chicken Dataset

Figure 13: Satellite image of Lincoln County, Nebraska, downloaded from USDA's VegScape application. Area between the lines was used in the analysis
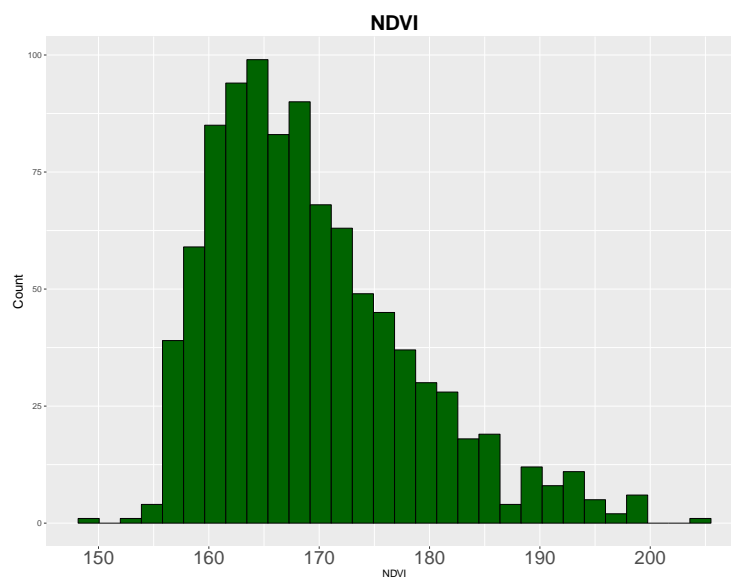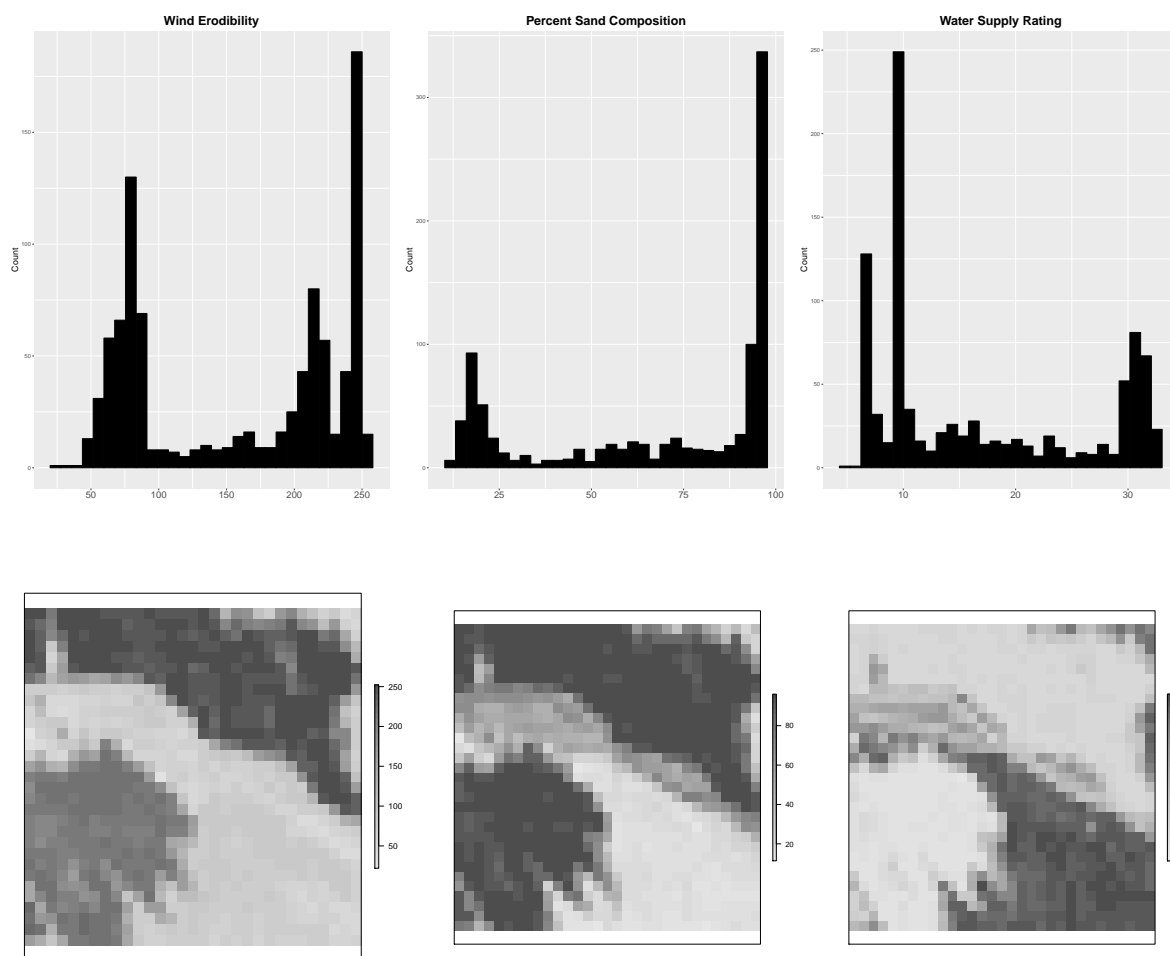

Figure 14: Histogram of NDVI values used in analysis

Figure 15: Histograms and raster images of covariates used in analysis