

1 Spatial Clustering Regression of Count Value Data via Bayesian 2 Mixture of Finite Mixtures

3 Anonymous Author(s)

4 ABSTRACT

5 Investigating relationships between response variables and covariates in areas such as environmental science, geoscience, and public health is an important endeavor. Based on a Bayesian mixture of finite mixtures model, we present a novel spatially clustered coefficients regression model for count value data. The proposed method detects the spatial homogeneity of the Poisson regression coefficients. A Markov random field constrained mixture of finite mixtures prior provides a regularized estimator of the number of clusters of regression coefficients with geographical neighborhood information. As a by-product, we also provide the theoretical properties of our proposed method when the Markov random field is exchangeable. An efficient Markov chain Monte Carlo algorithm is developed by using multivariate log gamma distribution as a base distribution. Simulation studies are carried out to examine the empirical performance of the proposed method. Additionally, we analyze Georgia's premature death data as an illustration of the effectiveness of our approach. The supplementary materials are provided anonymously at Anonymous GitHub https://anonymous.4open.science/r/MLG_MFM_anonymous-6F3B.

6 CCS CONCEPTS

- 7 • Mathematics of computing → Bayesian computation; Cluster analysis; Metropolis-Hastings algorithm.

8 KEYWORDS

9 Spatial data, count data, MCMC, Bayesian hierarchical model, mixture model

10 ACM Reference Format:

11 Anonymous Author(s). 2023. Spatial Clustering Regression of Count Value
12 Data via Bayesian Mixture of Finite Mixtures. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 14 pages. <https://doi.org/XXXXXX.XXXXXXXX>

13 1 INTRODUCTION

14 Spatial regression models have been universally used in many different fields such as environmental science [16, 39, 40], biological science [42], and econometrics [5, 41] to explore the relationship between a response variable and a set of predictors over a region. One of the most important tasks for a spatial regression model is to

15 Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

16 Conference acronym 'XX, Aug 06–10, 2023, Long Beach, CA

17 © 2023 Association for Computing Machinery.

18 ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

19 <https://doi.org/XXXXXX.XXXXXXXX>

20 capture the spatial dependence structure for the response variable. Spatial random effects are accounted for by the intercepts, and the regression coefficients are assumed to be constant over space under both linear models [7] and generalized linear models [11]. [5] proposed a geographically weighted regression (GWR) to capture spatially varying patterns in regression coefficients. The idea of GWR has been subsequently extended to various works by Hu & Huffer [17], Ma et al. [27], Xue et al. [38]. Furthermore, Gelfand et al. [14] incorporated spatial Gaussian process to linear regressions to build a spatially varying coefficients regression model. The aforementioned works all assume that each location has its own set of regression parameters, which sometimes leads to overfitting. The detection of clustered covariate effects has significant benefits in many various fields, including environmental science, spatial econometrics, and disease mapping. For instance, different parts of a country may have different economic conditions and development patterns. From a modeling perspective, grouping more advanced regions and less developed regions into separate clusters produces a more parsimonious model.

21 1.1 Related Work and Challenges

22 Spatial cluster detection methods, such as the scan statistic based method [19, 20], provide a remedy for spatial heterogeneity detection. Another important approach for spatial heterogeneity detection is to use the Bayesian framework to pursue spatial clusters [6, 25]. These two important approaches mainly focus on estimating cluster configurations of spatial responses. Recently, methods for cluster detection of spatial regression coefficients have been proposed to detect the homogeneity of the covariate effects among sub-areas [22, 23] under spatial scan statistics. From a graph theory perspective, Li & Sang [24] incorporated spatial neighborhood information based on minimum spanning trees in a penalized approach to detect spatially clustered coefficients. The existing literature focuses on Gaussian data under the linear model framework. For many social and environmental applications, Poisson regression for count response plays an important role [4].

23 **Developing clustering algorithms for regression coefficients under Poisson models poses several major challenges.** First, some specific spatial contiguity constraints need to be imposed on the clustering configuration to facilitate interpretations in the spatially clustered coefficients regression. Furthermore, in many regional science applications, spatially contiguous constraints should not dominate the global clustering configuration. In other words, the clustering results should contain the spatially contiguous pattern and spatially disconnected pattern. While the aforementioned methods [22–24] guarantee spatial contiguity, they fail to obtain globally discontiguous clusters that allow two clusters with long geographical distance to belong to the same cluster. Furthermore, [1] discusses Poisson regression with a spatially clustered intercept and a spatially clustered slope, but does not impose a spatial contiguity constraint.

Another critical consideration in clustering algorithms is to estimate the number of clusters. Bayesian inference provides a probabilistic framework for simultaneous inference of the number of clusters and the clustering configurations. Nonparametric Bayesian approaches, such as the Dirichlet process mixture (DPM) model [12], offer choices to estimate the number of clusters and the clustering configurations simultaneously. Ma et al. [28] proposed a Bayesian clustered regression for spatially dependent data based on Dirichlet process mixture model. However, their methods do not contain a consistent estimator of the number of the clusters due to inconsistency of the Dirichlet process mixture model [29]. To solve this over clustering problem of DPM, rich literature [26, 30, 37] propose several different ideas to obtain consistent estimators of the number of clusters. While existing works try to mitigate the over-clustering problem, no spatial information, such as neighborhood relationships, is utilized, while these have great potential for improving the clustering performance. From Tobler's first law of geography [34], "Everything is related to everything else, but near things are more related than distant things," it is reasonable to consider a similar pattern of the data due to similar environmental circumstances. Although it is hopeless to incorporate an arbitrary type of spatial information with consistent guarantee of the number of clusters, it is still interesting to see what specific types of dependency structure can be applied without any harm of consistency.

1.2 Contributions

To address these challenges, in this paper, we develop a Markov random field (MRF) constrained MFM (MRF-MFM) model to capture the spatial homogeneity in regression coefficients for Poisson model. Specifically, we develop a new Bayesian method for spatially clustered coefficients Poisson regression which leverages geographical information based on Markov random field constrained MFM model. The proposed methods leverage geographical information in a Bayesian model-based clustering algorithm for Poisson regression, enabling the capture of both locally spatially contiguous and globally discontiguous clusters simultaneously. We develop a Gibbs sampler that enables efficient full Bayesian inference on the number of clusters, mixture probabilities as well as other modeling parameters for Poisson regression with the help of multivariate log gamma (MLG) process [4]. We demonstrate the excellent numerical performance of proposed mixture models through simulations and an analysis of the premature deaths data in the state of Georgia. Finally, we establish a consistency result for the posterior estimates of the cluster number and the associated modeling parameters obtained from our proposed method when the Markov random field is assumed to be exchangeable.

Our proposed method is unique in the following aspects. First, the idea of introducing Markov random field into mixture of finite mixtures model for solving spatial cluster coefficients regression is novel. In fact, this idea and our proposed approach are widely applicable to spatial statistics applications such as environmental science and geographical analysis, and provide a valuable alternative to the existing literature that mainly relies on penalized regression or scan statistics. We provide a detailed comparison to the related literature (e.g., [1, 24]) in Section D in the supplement. Secondly, by

adopting a full Bayesian framework, the clustering results yield useful probabilistic interpretation. Moreover, the developed posterior sampling scheme also renders efficient computation. Thirdly, our theoretical result is among the first of its kind for mixture models under exchangeable assumption. The posterior consistency result not only provides a theoretical justification for the excellent empirical performance (e.g., regularization on the number of clusters), but also connects to the existing theoretical findings on mixture models in general.

2 METHODOLOGY

2.1 Clustered Poisson Regression and Mixture of Finite Mixtures

Consider a Poisson regression model with spatially varying coefficients as follows

$$y(s_i) \sim \text{Poisson}(\exp(X(s_i)\beta_{z_i})), i = 1, \dots, n, \quad (1)$$

where $X(s_i)$ is a $n \times p$ covariates matrix, $z_i \in \{1, \dots, k\}$ are labels of clusters, $\beta_{z_i} = \beta(s_i)$ is a p dimensional regression coefficients at location s_i . From [14], a Gaussian process prior can be assigned on regression coefficients to obtain spatially varying pattern. Compared with spatially varying pattern, heterogeneity pattern of covariate effects over subareas is also universally discussed in many different fields, such as real estate applications, spatial econometrics, and environmental science.

In the popular Chinese restaurant process, $z_i, i = 2, \dots, n$ are defined through the following conditional distribution [12]:

$$P(z_i = c | z_1, \dots, z_{i-1}) \propto \begin{cases} |c|, & \text{at an existing table labeled } c \\ \gamma, & \text{if } c \text{ is a new table} \end{cases}, \quad (2)$$

where $|c|$ is the size of cluster c , and γ is a concentration parameter of Dirichlet Process.

While CRP has a very attractive feature of simultaneous estimation on the number of clusters and the cluster configuration, a striking consequence of this has been recently discovered [29] where it is shown that the CRP produces extraneous clusters in the posterior, leading to inconsistent estimation of the *number of clusters* even when the sample size grows to infinity. A modification of the CRP called Mixture of finite mixtures (MFM) model is proposed to circumvent this issue [30].

$$\begin{aligned} k &\sim p(\cdot), \quad (\pi_1, \dots, \pi_k) | k \sim \text{Dir}(\gamma, \dots, \gamma), \\ z_i | k, \pi &\sim \sum_{h=1}^k \pi_h \delta_h, \quad i = 1, \dots, n, \end{aligned} \quad (3)$$

where $p(\cdot)$ is a proper probability mass function on $\{1, 2, \dots\}$, γ is a concentration parameter of Dirichlet Process and δ_h is a point-mass at h . Compared to the CRP, the introduction of new tables is slowed down by the factor $V_n(t+1)/V_n(t)$, which allows a model-based pruning of the tiny extraneous clusters.

The coefficient $V_n(t)$ is precomputed as:

$$V_n(t) = \sum_{k=1}^{+\infty} \frac{k(t)}{(\gamma k)^{(n)}} p(k), \quad (4)$$

where $k_{(t)} = k(k-1) \dots (k-t+1)$, and $(\gamma k)^{(n)} = \gamma k(\gamma k+1) \dots (\gamma k+n-1)$. $z_i, i = 2, \dots, n$ under (3) can be defined in a Pólya urn scheme similar to CRP:

$$\begin{aligned} P(z_i = c \mid z_1, \dots, z_{i-1}) \\ \propto \begin{cases} |c| + \gamma, & \text{at an existing table labeled } c. \\ \gamma V_n(t+1)/V_n(t), & \text{if } c \text{ is a new table.} \end{cases} \end{aligned} \quad (5)$$

where t is the number of existing clusters.

2.2 Introducing Dependency on the Base Measure

Recall that the full model for MFM is

$$\begin{aligned} K &\sim p_K, \text{ where } p_K \text{ is a p.m.f. on } \{1, 2, \dots\} \\ (\pi_1, \dots, \pi_k) &\sim \text{Dirichlet}_k(\gamma, \dots, \gamma) \\ z_1, \dots, z_n &\stackrel{\text{iid}}{\sim} \pi, \text{ given } \pi \\ \beta_1, \dots, \beta_k &\stackrel{\text{iid}}{\sim} H, \text{ given } K = k \\ y_j &\sim f_{\beta_{z_j}} \text{ independently for } j = 1, \dots, n, \end{aligned} \quad (6)$$

where H is the base distribution for β . The main insight of MFM is introducing a prior on the length of the Dirichlet distribution, and thus renders some regularization on the number of clusters created. However, the fourth step in the model, where i.i.d. samples are obtained from a base measure, fails to incorporate any dependency structure.

Inspired by [32], we apply the pairwise MRF in the level of coefficients to bring in interactions. With the assistance of Markov random field modeling, our MRF-MFM can incorporate more broad types of base measures. Consider an undirected random graph $G = (V, E, W)$, where $V = \{v_1, \dots, v_n\}$ is the vertex set while E is the set of graph edges, with weights W on the corresponding edges. Each vertex v_i is associated with a random variable β_i for $i = 1, 2, \dots, k$. The pairwise MRF model is defined as

$$\begin{aligned} \Pi(\beta_1, \dots, \beta_k) &= \exp \left\{ \sum_{i \in E} H_i(\beta_i) + \sum_{(i,j) \in E, j \neq i} H_{ij}(\beta_i \beta_j) \right\} \\ -A(W) &= \frac{1}{Z_H} \exp(H(\beta_1, \dots, \beta_k)), \end{aligned} \quad (7)$$

where Z_H is the normalizing constant. For example, for a Gaussian MRF, $H_i(\beta_i) = -W_{ii}\beta_i^2/2$ and $H_{ij}(\beta_i \beta_j) = -W_{ij}\beta_i \beta_j/2$; while for a binary MRF, i.e., the celebrated Ising model, $H_i(\beta_i) = W_{ii}\beta_i$ and $H_{ij}(\beta_i \beta_j) = W_{ij}\beta_i \beta_j$. We can then decompose the pairwise MRF into a vertex-wise term P and an interaction term M , then

$$\begin{aligned} \Pi(\beta_1, \dots, \beta_k) &\propto P(\beta_1, \dots, \beta_k) M(\beta_1, \dots, \beta_k), \text{ with} \\ P(\beta_1, \dots, \beta_k) &:= \frac{1}{Z_P} \exp \left\{ \sum_i H_i(\beta_i) \right\}; \\ M(\beta_1, \dots, \beta_k) &:= \frac{1}{Z_M} \exp \left\{ \sum_{C \in C_2} H_C(\beta_C) \right\}, \end{aligned} \quad (8)$$

where $C_2 := \{C \in C \mid \text{s.t. : } |C| = 2\}$ and C is the set of all cliques for the random graph (V, E, W) . For the spatial clustered coefficient regression, we study the component P defined in equation (8)

with a MFM prior. Our next theorem provides the generalized urn-model induced by MRF-MFM, thus a collapsed Gibbs sampler can be applied.

Theorem 2.1. Suppose the data generating process follows equation (6) with H replaced by the Markov random field $\Pi(\beta_1, \dots, \beta_k)$ in equation (8). If P is a continuous distribution and $n_0 > 1$, the distributions of β_{n_0} given $\beta_1, \dots, \beta_{n_0-1}$ is proportional to

$$\frac{V_{n_0}(t+1)\gamma}{V_{n_0}(t)} P(\beta) + \sum_{i=1}^t \exp(H_{i|-i}(\beta_i \mid \beta_{-i})) (n_i + \gamma) \delta_{\beta_i^*},$$

with

$$\begin{aligned} V_{n_0}(t) &= \sum_{k=1}^{\infty} \frac{k_{(t)}}{(\gamma k)^{(n_0)}} p_K(k); \\ H_{i|-i}(\beta_i \mid \beta_{-i}) &= \sum_{\{j: (i,j) \in E, j \neq i\}} H_{ij}(\beta_i \beta_j), \end{aligned} \quad (9)$$

where $\beta_1^*, \dots, \beta_t^*$, $t \leq n_0-1$ are the distinct values taken by $\beta_1, \dots, \beta_{n_0-1}$ and $n_i = \#\{j \in \{1, 2, \dots, n_0-1\} : \beta_j = \beta_i^*\}$, $x^{(m)} = x(x+1) \cdots (x+m-1)$ and $x_{(m)} = x(x-1) \cdots (x-m+1)$.

This theorem shows how the MRF constraints directly affect the urn sampling scheme compared with MFM. Consider the pairwise interactions, we model the conditional cost functions as

$$H_{i|-i}(\beta_i \mid \beta_{-i}) = \lambda \sum_{j \in \partial(i)} I(\beta_i = \beta_j), \quad (10)$$

where λ is the smoothness parameter, $\partial(i)$ denotes the set of the neighbors of observation i . The spatial smoothness can be controlled by the magnitude of λ . When $\lambda = 0$, the MRF-MFM reduces to MFM [30]. The conditional cost function in (10) is used in the data analysis of the paper.

2.3 Spatial Clustered Coefficient Regression for Count Value Data

In the MRF-MFM, a natural choice for the base distribution of β_1, \dots, β_k is the multivariate normal distribution. However, since the multivariate normal distribution is not a conjugate prior for Poisson regression, if it is to be used as the base distribution, it must be updated with Metropolis-Hastings or auxiliary parameters such as [31] in Gibbs sampling algorithms. Furthermore, the multivariate normal distribution has a very thin tail, which is not a good candidate for estimation of long-tail probability, such as the Poisson distribution. Bradley et al. [4] constructed a multivariate log-gamma distribution (MLG) which is conjugate with a Poisson distribution. The multivariate log-gamma distribution was formulated starting from Demirhan & Hamurkaroglu [10], who defined a multivariate gamma distribution that was then transformed to the log-scale. However, this transformation leads to complications for Gibbs sampling, as it requires component-wise updating to obtain known full-conditional distributions. Instead, the approach proposed by Bradley et al. [4] leads to block-wise full-conditional distributions that are easier to simulate from. It also exhibits asymptotic properties with respect to the multivariate normal distribution. A brief review on MLG is presented in the next section. We also refer readers to see [4] for details. In Hu & Bradley [16], it is demonstrated that MLG has a better long-tail probability property than

the multivariate normal distribution. In other words, the multivariate log-gamma distribution (MLG) not only serves as a conjugate for the Poisson distribution but also has the ability to handle long-tailed probabilities. The ultimate goal is to propose an MRF-MFM for spatial clustered coefficients in Poisson regression based on the MLG prior.

2.4 Probability Density Function for Multivariate Log-Gamma Distribution

We first review the multivariate log-gamma distribution from [4]. We define the n -dimensional random vector $\phi = (\phi_1, \dots, \phi_n)'$, which consists of n mutually independent log-gamma random variables with shape and scale parameters organized into the n -dimensional vectors $\alpha \equiv (\alpha_1, \dots, \alpha_n)'$, and $\kappa \equiv (\kappa_1, \dots, \kappa_n)'$, respectively. Then define the n -dimensional random vector q as follows

$$q = \mu + V\phi, \quad (11)$$

where the matrix $V \in \mathcal{R}^n \times \mathcal{R}^n$ and $\mu \in \mathcal{R}^n$. [4] called q the multivariate log-gamma random vector. The random vector q has the following probability density function:

$$f(q | c, V, \alpha, \kappa) = \frac{1}{\det(V)} \left(\prod_{i=1}^n \frac{\kappa_i^{\alpha_i}}{\Gamma(\alpha_i)} \right) \times \exp[\alpha' V^{-1}(q - \mu) - \kappa' \exp\{V^{-1}(q - \mu)\}]; \quad q \in \mathcal{R}^n, \quad (12)$$

where “det” represents the determinant function. As a shorthand we use the notation, $\text{MLG}(\mu, V, \alpha, \kappa)$, for the probability density function in (12).

2.5 Conditional Distributions for Multivariate Log-Gamma Random Vectors

Gibbs sampling from full-conditional distributions will require simulating from conditional distributions of multivariate log-gamma random vectors. Here, we provide a review for the technical results needed to simulate from these conditional distributions.

We first look at the Proposition 1 from [4]. Let $q \sim \text{MLG}(c, V, \alpha, \kappa)$, and let $q = (q'_1, q'_2)'$, where q'_1 is g -dimensional and q'_2 is $(n-g)$ -dimensional. In a similar manner, partition $V^{-1} = [H \ B]$ into an $m \times g$ matrix H and an $m \times (m-g)$ matrix B . Then, the conditional pdf of q'_1 is given by

$$f(q'_1 | q'_2 = d, c, V, \alpha, \kappa) = M \exp(\alpha' H q'_1 - \kappa'_{1,2} \exp(H q'_1)). \quad (13)$$

where $\kappa'_{1,2} \equiv \exp(Bd - V^{-1}c - \log(\kappa))$ and the normalizing constant M is

$$M = \frac{1}{\det(VV')^{\frac{1}{2}}} \left(\prod_{i=1}^n \frac{\kappa_i^{\alpha_i}}{\Gamma(\alpha_i)} \right) \frac{\exp \alpha' Bd - \alpha' V^{-1}c}{\left[\int f(q | c, V, \alpha, \kappa) dq_1 \right]_{q_2=d}}, \quad (14)$$

so the $\text{cMLG}(H, \alpha, \kappa'_{1,2})$ is equal to the pdf in equation (17), where “cMLG” stands for “conditional multivariate log-gamma.” In [4], it indicates that cMLG does not fall within the same class of pdfs given in (13). This is primarily due to the fact that the real-valued matrix H , within the expression of cMLG, is not square. Thus, we require an additional result that allows us to simulate from cMLG.

Next, we look at the Theorem 2 from [4]. Let $q \sim \text{MLG}(0_n, V, \alpha, \kappa)$, and partition this n -dimensional random vector so that $q = (q'_1, q'_2)'$,

where q'_1 is g -dimensional and q'_2 is $(n-g)$ -dimensional. Additionally, consider the class of MLG random vectors that satisfy the following:

$$V^{-1} = [Q_1 \ Q_2] \begin{bmatrix} R_1 & 0_{g,n-g} \\ 0_{n-g,g} & \frac{1}{\sigma_2} I_{n-g} \end{bmatrix} \quad (15)$$

where in general $0_{r,t}$ is an $r \times t$ matrix of zeros; I_{n-g} is an $(n-g) \times (n-g)$ identity matrix;

$$H = [Q_1 \ Q_2] \begin{bmatrix} R_1 \\ 0_{n-g,g} \end{bmatrix} \quad (16)$$

is the QR decomposition of the $n \times g$ matrix H ; the $n \times g$ matrix Q_1 satisfies $Q'_1 Q_1 = I_g$, the $n \times (n-g)$ matrix Q_2 satisfies $Q'_2 Q_2 = I_{n-g}$, and $Q'_2 Q_1 = 0_{n-g,g}$; R_1 is a $g \times g$ upper triangular matrix; and $\sigma_2 > 0$. Hence, the marginal distribution of the g -dimensional random vector q'_1 is given by

$$f(q'_1 | H, \alpha, \kappa) = M_1 \exp(\alpha' H q'_1 - \kappa' \exp(H q'_1)). \quad (17)$$

where the normalizing constant M_1 is

$$M_1 = \det([H \ Q_2]) \left(\prod_{i=1}^n \frac{\kappa_i^{\alpha_i}}{\Gamma(\alpha_i)} \right) \frac{1}{\int f(q | 0_n, V = [H \ Q_2]^{-1}, \alpha, \kappa) dq_1}. \quad (18)$$

And, the g -dimensional random vector q'_1 is equal in distribution to $(H'H)^{-1} H' \omega$, where the n -dimensional random vector $\omega \sim \text{MLG}(0_n, I_n, \alpha, \kappa)$.

In [4], it is evident that this particular class of marginal distributions (defined in Theorem 2 in [4]) falls into the same class of distributions as the conditional distribution of q'_1 given q'_2 . And Theorem 2 in [4] provides a way to simulate from cMLG. Furthermore, it shows that it is (computationally) easy to simulate from cMLG provided that $g \ll n$. Recall that H is $n \times g$, which implies that computing the $g \times g$ matrix $(H'H)^{-1}$ is computationally feasible when g is “small.” We refer the readers to [4] for a comprehensive discussion.

3 FULL CONDITIONAL DISTRIBUTIONS AND ALGORITHM

We adapt the MRF-MFM in conjunction with MLG to a spatial Poisson regression setting, focusing on the clustering of spatially-varying coefficients $\beta(s_1), \dots, \beta(s_n)$, where $\beta(s_i)$ is the p -dimensional coefficient vector for location s_i . In our setting, we assume that the n parameter vectors can be clustered into k groups, i.e., $\beta(s_i) = \beta_{z_i} \in \{\beta_1, \dots, \beta_k\}$. Hence, the hierarchical model can be expressed as follows

Data Model: $y(s_i) | \beta(s_i) \sim \text{Poisson}(\exp(X(s_i)\beta(s_i)))$

$$\text{MRF: } (\beta(s_1), \dots, \beta(s_n)) \sim M(\beta(s_1), \dots, \beta(s_n)) \prod_{i=1}^n G(\beta(s_i)) \quad (453)$$

$$\text{MLG: } \beta_1, \dots, \beta_k \sim \text{MLG}(\mu, V, \alpha, \kappa) \quad (455)$$

$$\text{MFM: } G(\beta(s_i)) = \sum_{j=1}^k \pi_j \beta_j, \pi_1, \dots, \pi_k \mid k \sim \text{Dirichlet}(\gamma, \dots, \gamma), \quad (457)$$

$$k \sim p(\cdot), \text{ where } p(\cdot) \text{ is a p.m.f on } \{1, 2, \dots\}. \quad (459)$$

The full conditional distributions in Markov chain Monte Carlo (MCMC) sampling of MRF-MFM are given in Algorithm 1 and

Table 1: Parameters of the full conditional distribution

Parameter	Form
H_β	$\begin{bmatrix} V^{-1} \\ X(s_i) \end{bmatrix}$
α_β	$\begin{bmatrix} \alpha \\ \sum_{z_i=r} y(s_i) \end{bmatrix}$
κ_β	$\begin{bmatrix} \kappa \\ \sum_{z_i=r} I_{(z_i=r)} \end{bmatrix}$

Table 1 where the detailed derivations are given in Section A in the supplement.

Algorithm 1 Collapsed sampler for MRF-MFM

Initialize: $z = (z_1, \dots, z_n)$ and $\beta = (\beta_1, \dots, \beta_k)$
for each iteration = 1 **to** B **do**

 Update $\beta = (\beta_1, \dots, \beta_k)$ conditional on z in a closed form as

$$f(\beta_r | -) \sim \text{cMLG}(H_\beta, \alpha_\beta, \kappa_\beta)$$

where,

$$H_\beta = \begin{bmatrix} V^{-1} \\ X(s_i) \end{bmatrix}, \alpha_\beta = \begin{bmatrix} \alpha \\ \sum_{z_i=r} y(s_i) \end{bmatrix}, \kappa_\beta = \begin{bmatrix} \kappa \\ \sum_{z_i=r} I_{(z_i=r)} \end{bmatrix}$$

Update $z = (z_1, \dots, z_n)$ conditional on $\beta = (\beta_1, \dots, \beta_k)$ for each i in $(1, \dots, n)$, we can get closed form expression for $P(z_i = c | z_{-i}, \beta)$:

$$\propto \begin{cases} P(z_i = c | z_{-i}) d\text{Poisson}(y(s_i), \exp(X(s_i)\beta_c)), & \text{at an existing table labeled } c \\ \frac{V_n(|C_{-i}|+1)}{V_n(|C_{-i}|)} \gamma m(y(s_i)), & \text{if } c \text{ is a new table} \end{cases}.$$

where C_{-i} denotes the partition obtained by removing z_i and

$$m(y(s_i)) = \frac{1}{\det(VV')^{\frac{1}{2}}} \left(\prod_{i=1}^p \frac{\kappa_i^{\alpha_i}}{\Gamma(\alpha_i)} \right) \frac{1}{M_1}$$

where,

$$M_1 = \det([H_\beta \ Q_2]) \left(\prod_{i=1}^{n+p} \frac{\kappa_i^{\alpha_i}}{\Gamma(\alpha_i)} \right) \frac{1}{\int f(y(s_i) | 0, V = [H_\beta, Q_2]^{-1}, \alpha, \kappa)}$$

end for

3.1 Theoretical Properties under the exchangeable structure

In this section, we assume the covariates $X(s_i)$ are generated from random homogenous distribution so it is marginalized. The incorporation of proper dependency structures into the estimation process and assessing uncertainty is always an interesting subject. However, complex dependency structures may destroy the consistency of MFM. Therefore, to maintain theoretical consistency, this paper considers the case in which samples from the base measure are a subset of an infinite sequence of exchangeable variables.

In Bayesian Statistics, the infinite sequence of exchangeable random variables is an important concept. When β_1, \dots are infinite exchangeable, for any finite k ,

$$\beta_1, \dots, \beta_k \stackrel{D}{=} \beta_{\pi(1)}, \dots, \beta_{\pi(k)} \text{ for all } \pi \in S(k), \quad (20)$$

where $S(k)$ is the set of all permutations for the index set $\{1, \dots, k\}$. If β_1, \dots are i.i.d. sampled from a distribution $P(\beta)$, then they are exchangeable, but the reverse is not always true. Some widely used models are based on exchangeable random variables that are not independent, like the Pólya's Urn [3] and Gaussian random variables that have the same marginal distribution and the same correlation between any two of them.

The famous de Finetti's Theorem [9] reveals the intrinsic characterization of exchangeable random variables: there is a latent random variable θ , such that β_1, \dots, β_n are a subset of a infinite sequence of exchangeable variables sampled from $\Pi(\beta_1, \dots)$. It is summarized into the following sampling procedure:

$$\theta \sim \Theta, \quad \beta_1, \dots, \beta_k \stackrel{i.i.d.}{\sim} \Pi(\beta | \theta), \quad (21)$$

where Θ only depends on $\Pi(\beta_1, \dots)$. In other words, a subset of an infinite sequence of exchangeable variables are conditionally i.i.d. given their latent labels. We refer to [2] for more details on exchangeable sequences.

Theorem 3.1. Suppose the data generating process follows equation (6) with H replaced by the hierarchical distribution in equation (21), and the distribution is correctly specified. If $p_K(1), \dots, p_K(k) > 0$, denote T as the random variable for the number of clusters and t is all the possible values T will take in the true data generating process. Then we have

$$|p(T = t | \mathbf{y}) - p(K = k | \mathbf{y})| \longrightarrow 0 \quad (22)$$

as $n \rightarrow \infty$.

Theorem 3.1 provide some insight into our proposed MRF-MFM, compared to Dirichlet process mixture model with the above Markov random fields [DP-MRF; 32]. For DP-MRF, there could be a lot of small spurious clusters due to inconsistency of the Dirichlet process mixture even in the i.i.d. case [29]. Due to the fact that we specify a prior distribution for the number of components, the number of components in the posterior is appropriately regularized. Even though the consistency result only holds for the exchangeable structure, we believe that the regularization effect holds for all types of structures. Theorem 3.1 is an extension of Theorem 5.2 in [30] to the case of an exchangeable base measure. The limitation of the above theorem is that it does not explore the frequentist property of the posterior, where the number of clusters is assumed to be a fixed truth.

4 SIMULATION

4.1 Settings

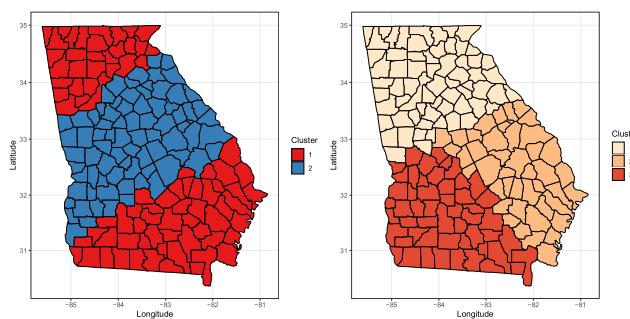
Our goal is to sample from the posterior distribution of the unknown parameters $k, z = (z_1, \dots, z_n) \in \{1, \dots, k\}$ and $\beta = (\beta_1, \dots, \beta_k)$. We choose $k - 1 \sim \text{Poisson}(1)$ and $y = 1$ in (19), $\mu = 0_n$, $V = 100I_n$ and $\alpha = \kappa = 100001_n$ for all the simulations and real data analysis, where 0_n is an n -dimensional vector with 0, 1_n is an n -dimensional vector of 1's, and I_n is an n -dimensional identity matrix. The computing algorithm and full conditional distributions are presented in Appendix 3, which efficiently cycles through the full conditional distributions of $z_i | z_{-i}$ for $i = 1, 2, \dots, n$ and β , where $z_{-i} = z \setminus z_i$. The marginalization over k can avoid complicated reversible jump MCMC algorithms or even allocation samplers. The

581 posterior sampling algorithm is given in Algorithm 1 in Appendix
 582 3. The details of the deviations of full conditionals are also given in
 583 Appendix 3.

584 It is not appropriate to use the posterior mean or median of
 585 clustering configurations [z]. Dahl's method [8] provides a remedy
 586 for posterior inference of clustering configurations based on the
 587 squared error loss. There are also alternative loss functions in [35]
 588 that do not involve squared errors such as those in [8]. The Rand
 589 Index (RI) [33] is used to measure the accuracy of clustering. The
 590 tuning parameter in Markov random fields needs to be selected
 591 in our proposed model. The Logarithm of the Pseudo-Marginal
 592 Likelihood (LPML) [18] is applied for tuning parameter selection,
 593 where a model with a larger LPML value is preferred.

594 4.2 Simulation Setting and Evaluation Metrics

595 Our analysis is based on the spatial structure of the state of Georgia,
 596 which contains 159 counties. Using the county-level data, we build
 597 the graph using an adjacency matrix among different counties. 159
 598 counties represent 159 vertices in this graph, and if a county shares
 599 a boundary with another county, then v_i and v_j are connected. This
 600 graph is used for both simulation studies and real data analysis. We
 601 consider two different spatial cluster designs shown in Figure 1.
 602 The first design consists of two disjoint parts located in the top and
 603 bottom parts of Georgia. A second cluster comprises the counties
 604 in the middle. The second design comprises three major spatial
 605 clusters. It is designed to mimic a common premature death pattern
 606 in which geographically distant areas can share a similar distribution
 607 pattern, and geographical proximity is not considered the only
 608 factor responsible for homogeneity in premature death rates.



628 **Figure 1: Simulation design with two and three cluster ass-
 629 signments**

630 Two different scenarios are considered for each design. The
 631 first scenario does not take into account spatial random effects,
 632 while in the second scenario, spatial random effects are included
 633 for each design. The spatial random effects are assumed to follow a
 634 multivariate normal distribution with a mean zero and exponential
 635 covariogram. Our simulation study consists of four scenarios in
 636 total. The details of the data generation process are given as

- (1) $y(s_i) \sim \text{Poisson}(X_1(s_i)\beta_{1z_i} + X_2(s_i)\beta_{2z_i})$, where $X_1(s_i), X_2(s_i) \stackrel{\text{ind}}{\sim} \text{Unif}(1, 2)$, $i = 1, \dots, n$, $(\beta_{11}, \beta_{21}) = (1, 1)$, $(\beta_{12}, \beta_{22}) = (1.5, 1.5)$. 639
- (2) $y(s_i) \sim \text{Poisson}(X_1(s_i)\beta_{1z_i} + X_2(s_i)\beta_{2z_i} + w(s_i))$, where 640
- $X_1(s_i), X_2(s_i) \stackrel{\text{ind}}{\sim} \text{Unif}(1, 2)$, $i = 1, \dots, n$, $(\beta_{11}, \beta_{21}) = (1, 1)$, 641
- $(\beta_{12}, \beta_{22}) = (1.5, 1.5)$. $\omega \sim N(0, \sigma_\omega^2 H(\phi))$, where $H(\phi) = 642$
- $\exp(-\phi \|s_i - s_j\|)$, we set $\sigma_\omega^2 = 0.3$ and $\phi = 0.05$. 643
- (3) $y(s_i) \sim \text{Poisson}(X_1(s_i)\beta_{1z_i} + X_2(s_i)\beta_{2z_i})$, where $X_1(s_i), X_2(s_i) \stackrel{\text{ind}}{\sim} \text{Unif}(1, 2)$, $i = 1, \dots, n$, $(\beta_{11}, \beta_{21}) = (0.5, 0.5)$, $(\beta_{12}, \beta_{22}) = 644$
- $(1, 1)$, $(\beta_{13}, \beta_{23}) = (1.5, 1.5)$. $\omega \sim N(0, \sigma_\omega^2 H(\phi))$, where $H(\phi) = 645$
- $\exp(-\phi \|s_i - s_j\|)$, we set $\sigma_\omega^2 = 0.3$ and $\phi = 0.05$. 646
- (4) $y(s_i) \sim \text{Poisson}(X_1(s_i)\beta_{1z_i} + X_2(s_i)\beta_{2z_i} + w(s_i))$, where 647
- $X_1(s_i), X_2(s_i) \stackrel{\text{ind}}{\sim} \text{Unif}(1, 2)$, $i = 1, \dots, n$, $(\beta_{11}, \beta_{21}) = (0.5, 0.5)$, 648
- $(\beta_{12}, \beta_{22}) = (1, 1)$, $(\beta_{13}, \beta_{23}) = (1.5, 1.5)$. $\omega \sim N(0, \sigma_\omega^2 H(\phi))$, 649
- where $H(\phi) = \exp(-\phi \|s_i - s_j\|)$, we set $\sigma_\omega^2 = 0.3$ and $\phi = 0.05$. 650

651 The four scenarios are for two cluster design without spatial random 652 effect, two cluster design with spatial random effect, three cluster 653 design without spatial random effect, and three cluster design with 654 spatial random effect, respectively. In the three clusters design, 655 the original regression coefficients are set to be 0.5, 1 and 1.5 for 656 each cluster correspondingly. On the other hand, in two clusters 657 design, the original regression coefficient set to be 1 and 1.5 for 658 each cluster, respectively. For each case, we add the spatial random 659 effect with the intensity. We use the centroid coordinate in each 660 county to represent that county then construct the spatial random 661 effect. Also, the range parameter and spatial variance parameter 662 are both fixed in each simulation. In each case, we avoid the zero 663 count value to prevent numerical instability. Based on the estimated 664 number of clusters and Rand Index (RI), the clustering performance 665 is evaluated. Each replicate is also used to calculate the final number 666 of clusters estimated. A total of 100 sets of data are generated under 667 different scenarios. We run 5000 iterations of the MCMC chain and 668 burn-in the first 1000 for each replicate. 669

670 4.3 Simulation Results

671 For each replicated data set, we fit MFM and MRF-MFM with differ- 672 ent values of the smoothness parameter and select the best smooth- 673 ness parameter for each replicate based on LPML. We see that our 674 model outperforms the MFM model in terms of LPML in all four 675 different scenarios. We also evaluate the performance in terms of 676 estimation results of the number of clusters. We report the propor- 677 tion of times the true cluster recovered among the 100 replicates. 678 For the two-cluster without spatial random effect design, we find 679 out our model can recover the true number of clusters 100% of the 680 replicates. And the MFM model can recover 85% of the replicates. 681 In this case, both models perform well in the number of clusters 682 estimation. But our model outperforms the MFM model in terms of 683 LPML value. For the two-cluster design with spatial random effects, 684 we see that our model can recover the true number of clusters 97% 685 of the replicates, but the MFM model did not recover the true clus- 686 ter for any replicates. For the three-cluster without spatial random 687 effect design, we find out our model can recover the true number of 688 clusters 88% of the replicates. On the other hand, MFM recovers 62% 689 of the replicates. Finally, for the three-cluster design with spatial 690 random effects, we find out our model can recover the true number 691 692 693 694 695 696

of clusters 73% of the replicates. However, MFM did not recover the true cluster for all replicates.

The results of the comparison of LPML, Rand index, and estimation of the number of clusters for each design can be found in Table 2. Our method can effectively estimate the true number of clusters based on the results shown in Table 2. However, if spatial random effects exist, MFM will overestimate the number of clusters. Our proposed method also outperforms vanilla MFM with respect to model fitness and clustering, as demonstrated by the LPML values and Rand index.

Table 2: Simulation Results for Four Scenarios including LPML, Rand Index (RI), and number of true cluster cover rate by MRF-MFM (optimal) model and MFM model. We provide mean and standard deviation for both LPML and RI.

Scenario	Method	LPML	RI	Cover Rate
Scenario 1	Optimal	-544.29 (12.06)	0.9970 (0.0062)	100%
	MFM	-1146.32 (593.33)	0.9901 (0.0233)	85%
Scenario 2	Optimal	-690.91 (34.36)	0.9875 (0.0129)	97%
	MFM	-7632.18 (1947.31)	0.8348 (0.0597)	0%
Scenario 3	Optimal	-752.76 (235.91)	0.9470 (0.0389)	88%
	MFM	-2201.69 (830.66)	0.9570 (0.0231)	62%
Scenario 4	Optimal	-1297.58 (278.33)	0.8469 (0.0434)	73%
	MFM	-8890.92 (2028.92)	0.8350 (0.0431)	0%

Furthermore, we show the average mean square error (AMSE) of our proposed method and MFM in Table 3. We see that in all four different scenarios, our proposed method outperforms MFM in terms of coefficients estimations. The improvement of our proposed methods is evident for the data generated from the model with spatial random effect.

5 ILLUSTRATION: PREMATURE DEATHS IN GEORGIA

5.1 Data Description

In this study, the proposed methods are used to analyze the factors that influence the number of premature deaths in Georgia. The objective of this study is to investigate the relationship between premature deaths and environmental factors such as PM 2.5 and food environment index. The dataset is available at www.countyhealthrankings.org with 159 observations corresponding to the 159 counties in state of Georgia in 2015. For each county, the dependent variable is the number of the premature death in each county. The premature death is the death that occurs before the average age of death in a certain population. In the United States,

Table 3: AMSE for β Estimation under All Scenarios

Method	No Spatial Random effect	
	Two Clusters	Three Clusters
MRF-MFM-MLG	$\hat{\beta}_1$ $\hat{\beta}_2$	0.0848 0.0839
		0.2508 0.2435
MFM-MLG	$\hat{\beta}_1$ $\hat{\beta}_2$	0.1170 0.1164
		0.2841 0.2781
With Spatial Random effect		
	Two Clusters	Three Clusters
	$\hat{\beta}_1$ $\hat{\beta}_2$	0.0966 0.0967
MFM-MLG	$\hat{\beta}_1$ $\hat{\beta}_2$	0.3918 0.3814
		0.6996 0.6898

the average age of death is about 75 years. The dependent variable is the number of lives lost per 100,000 population before age 75 in each county. The two covariates we consider in this paper are PM 2.5 (X_1) and food environment index (X_2). PM 2.5 is the average daily density of fine particulate matter in micrograms per cubic meter. The food environment index is the index of factors that contribute to a healthy food environment, 0 (worst) to 10 (best). Figures 2a and 2b present a visualization of the response and two covariates on the Georgia map.

5.2 Data Analysis

In this section, we apply the proposed methodology to present a detailed analysis of premature death data in the state of Georgia. First, we rescale the data to a decent range as the variance in the Poisson distribution is equal to the mean. The count of the premature death is scaled to hundreds. We run 25,000 MCMC iterations and burn-in the first 15,000 iterations. The smoothing parameter is tuned over the grid $\{0.1, 0.2, \dots, 1\}$. All other parameters are set to be consistent with the simulation study. The final clustering result corresponds to the largest LPML [18], hence we choose the smoothing parameter equal to 0.3. The 159 counties turned out to be put into four clusters as illustrated in Figure 3. The number of the counties in each cluster are 150, 3, 5 and 1, respectively. We also compare our model with the best LPML to vanilla MFM, Latent Gaussian Process (LGP) [15], conditional autoregressive (CAR) [21] models and Bayesian spatially varying coefficient models (SVC) [13, 14, 36]. The LPML values for candidate models are: -2221.45 (MRF-MFM), -3614.38 (MFM), -2461.31 (LGP), -5015.93 (CAR), -3123.47 (SVC). Based on the LPML results, our proposed model outperforms other models. In contrast, there are 15 different clusters identified by vanilla MFM. From the estimation results shown in Table 4, we see that all the counties with higher PM 2.5 will have higher premature deaths. For Cobb County, PM 2.5 has the largest effect on premature death. An extensive analysis could be conducted to investigate why the majority of counties are grouped into one cluster while the other three clusters only contain a couple of counties. For instance, one possible approach is to use log likelihood ratio test (LRT) to detect spatial cluster signals.

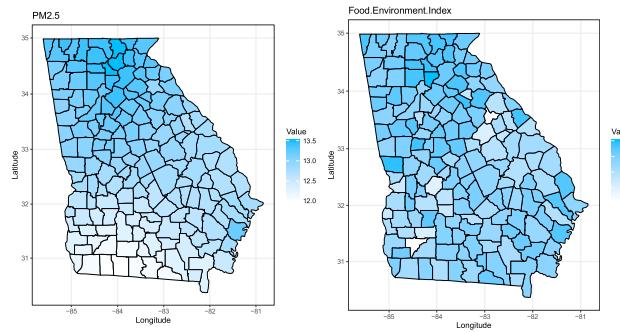
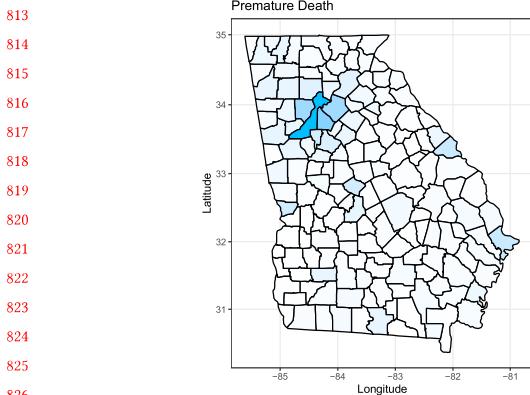


Table 4: Dahl's method estimates for the four clusters of Georgia Data

Cluster	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
1	-1.134	0.077	0.209
2	-3.644	0.060	1.222
3	-1.325	0.476	-0.249
4	-0.188	1.446	-2.093

6 DISCUSSION

Some topics beyond the scope of this paper are worth further investigations. First, in our MCMC algorithm, one numerical integration is required for Gibbs sampling. Proposing an efficient calculation algorithm of the numerical integration will broaden the applications of our proposed methods. Furthermore, the proposed algorithm is numerically unstable when zero counts are observed, which should be addressed in the future. Furthermore, different clusters may have different sparsity patterns of the covariates. Incorporating spatial clustered sparsity structure of regression coefficients into the model will enable the selection and identification of the most important covariates. One parameter of the Markov random field is required to be selected. Proposing a hierarchical model for the tuning parameter is also an interesting future work. The frequentist property

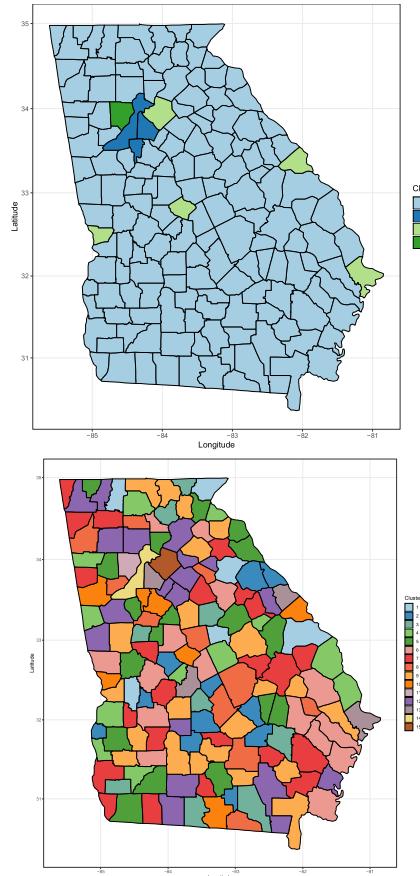


Figure 3: Top: Illustration of 4 clusters identified by the proposed method for counties. Bottom: Illustration of 15 clusters identified by vanilla MFM for counties.

of the posterior distribution is also expected to be explored in the future.

7 SUPPLEMENTARY MATERIALS

The supplementary materials are provided anonymously at Anonymous GitHub https://anonymous.4open.science/r/MLG_MFM_anonymous. The supplementary materials contain a detailed comparison with related literature like [24] and [1], proofs of the main theorems, of derivations full conditional distributions and the MCMC algorithms, additional simulation using states of Georgia and Mississippi and reproducing codes for data analysis.

REFERENCES

- [1] Anderson, C., Lee, D., and Dean, N. Spatial clustering of average risks and risk trends in Bayesian disease mapping. *Biometrical Journal*, 59(1):41–56, 2017.
- [2] Bernardo, J. M. and Smith, A. F. *Bayesian Theory*, volume 405. John Wiley & Sons, 2009.
- [3] Blackwell, D., MacQueen, J. B., et al. Ferguson distributions via polya urn schemes. *The Annals of Statistics*, 1(2):353–355, 1973.
- [4] Bradley, J. R., Holan, S. H., Wikle, C. K., et al. Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). *Bayesian Analysis*, 13(1):253–310, 2018.

- [5] Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4):281–298, 1996.
- [6] Carlin, B. P., Gelfand, A. E., and Banerjee, S. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2014.
- [7] Cressie, N. Statistics for spatial data. *Terra Nova*, 4(5):613–617, 1992.
- [8] Dahl, D. B. Model-based clustering for expression data via a Dirichlet process mixture model. *Bayesian Inference for Gene Expression and Proteomics*, 4:201–218, 2006.
- [9] De Finetti, B. Funzione caratteristica di un fenomeno aleatorio. In *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, pp. 179–190, 1929.
- [10] Demirhan, H. and Hamurkaroglu, C. On a multivariate log-gamma distribution and the use of the distribution in the bayesian analysis. *Journal of Statistical Planning and Inference*, 141(3):1141–1152, 2011.
- [11] Diggle, P. J., Tawn, J. A., and Moyeed, R. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350, 1998.
- [12] Ferguson, T. S. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.
- [13] Finley, A. O., Banerjee, S., and Gelfand, A. E. spbayes for large univariate and multivariate point-referenced spatio-temporal data models. *arXiv preprint arXiv:1310.8192*, 2013.
- [14] Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396, 2003.
- [15] Hadfield, J. D. et al. Mcmc methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2):1–22, 2010.
- [16] Hu, G. and Bradley, J. A Bayesian spatial-temporal model with latent multivariate log gamma random effects with application to earthquake magnitudes. *Stat*, 7(1):e179, 2018. e179 stat:4.179.
- [17] Hu, G. and Huffer, F. Modified Kaplan–Meier estimator and Nelson–Aalen estimator with geographical weighting for survival data. *Geographical Analysis*, 52(1):28–48, 2019.
- [18] Ibrahim, J. G., Chen, M.-H., and Sinha, D. *Bayesian Survival Analysis*. Springer Science & Business Media, 2013.
- [19] Jung, I., Kulldorff, M., and Klassen, A. C. A spatial scan statistic for ordinal data. *Statistics in Medicine*, 26(7):1594–1607, 2007.
- [20] Kulldorff, M. and Nagarwalla, N. Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14(8):799–810, 1995.
- [21] Lee, D. CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software*, 55(13):1–24, 2013.
- [22] Lee, J., Gangnon, R. E., and Zhu, J. Cluster detection of spatial regression coefficients. *Statistics in Medicine*, 36(7):1118–1133, 2017.
- [23] Lee, J., Sun, Y., and Chang, H. H. Spatial cluster detection of regression coefficients in a mixed-effects model. *Environmetrics*, pp. e2578, 2019.
- [24] Li, F. and Sang, H. Spatial homogeneity pursuit of regression coefficients for large datasets. *Journal of the American Statistical Association*, pp. 1–21, 2019.
- [25] Li, P., Banerjee, S., Hanson, T. E., and McBean, A. M. Nonparametric hierarchical modeling for detecting boundaries in areally referenced spatial datasets. Technical report, Technical Report rr2010-014, Divison of Biostatistics, School of Public , 2010.
- [26] Lu, J., Li, M., and Dunson, D. Reducing over-clustering via the powered Chinese restaurant process. *arXiv preprint arXiv:1802.05392*, 2018.
- [27] Ma, Z., Xue, Y., and Hu, G. Geographically weighted regression analysis for spatial economics data: A Bayesian recourse. *arXiv preprint arXiv:2007.02222*, 2020.
- [28] Ma, Z., Xue, Y., and Hu, G. Heterogeneous regression models for clusters of spatial dependent data. *Spatial Economic Analysis*, pp. 1–17, 2020.
- [29] Miller, J. W. and Harrison, M. T. A simple example of Dirichlet process mixture inconsistency for the number of components. In *Advances in Neural Information Processing Systems*, pp. 199–206, 2013.
- [30] Miller, J. W. and Harrison, M. T. Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 113(521):340–356, 2018.
- [31] Neal, R. M. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.
- [32] Orbanz, P. and Buhmann, J. M. Nonparametric Bayesian image segmentation. *International Journal of Computer Vision*, 77(1–3):25–45, 2008.
- [33] Rand, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.
- [34] Tobler, W. R. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1):234–240, 1970.
- [35] Wade, S., Ghahramani, Z., et al. Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis*, 13(2):559–626, 2018.
- [36] Wheeler, D. C. and Calder, C. A. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, 9(2):145–166, 2007.
- [37] Xie, F. and Xu, Y. Bayesian repulsive Gaussian mixture model. *Journal of the American Statistical Association*, pp. 1–29, 2019.
- [38] Xue, Y., Schifano, E. D., and Hu, G. Geographically weighted Cox regression for prostate cancer survival data in Louisiana. *Geographical Analysis*, 2019. Forthcoming.
- [39] Yang, H.-C. and Bradley, J. R. Bayesian inference for big spatial data using non-stationary spectral simulation. *Spatial Statistics*, 43:100507, 2021.
- [40] Yang, H.-C., Hu, G., and Chen, M.-H. Bayesian variable selection for pareto regression models with latent multivariate log gamma process with applications to earthquake magnitudes. *Geosciences*, 9(4):169, 2019.
- [41] Yang, H.-C., Xue, Y., Geng, L., and Hu, G. Spatial weibull regression with multivariate log gamma process and its applications to china earthquake economic loss. *Statistics and Its Interface*, 15(1):29–38, 2022.
- [42] Zhang, J. and Lawson, A. B. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *Journal of Applied Statistics*, 38(3):591–603, 2011.