Statistics and Computing (2024) 34:70
https://doi.org/10.1007/s11222-024-10380-1

ORIGINAL PAPER

Global–local shrinkage multivariate logit-beta priors for multiple response-type data

Hongyu Wu[1] · **Jonathan R. Bradley**[1]

Received: 20 February 2023 / Accepted: 3 January 2024 / Published online: 3 February 2024

Abstract
Multiple-type outcomes are often encountered in many statistical applications, one may want to study the association between multiple responses and determine the covariates useful for prediction. However, literature on variable selection methods for multiple-type data is arguably underdeveloped. In this article, we develop a novel global–local shrinkage prior in multiple response-types settings, where the observed dataset consists of multiple response-types (e.g., continuous, count-valued, Bernoulli trials, etc.), by combining the perspectives of global–local shrinkage and the conjugate multivaraite distribution. One benefit of our model is that a transformation or a Gaussian approximation on the data is not needed to perform variable selection for multiple response-type data, and thus one can avoid computational difficulties and restrictions on the joint distribution of the responses. Another benefit is that it allows one to parsimoniously model cross-variable dependence. Specifically, our method uses basis functions with random effects, which can be presented as known covariates or pre-defined basis functions, to model dependence between responses and dependence can be detected by our proposed global–local shrinkage model with a sparsity-inducing model. We provide connections to the original horseshoe model and existing basis function models. An efficient block Gibbs sampler is developed, which is found to be effective in obtaining accurate estimates and variable selection results. We also provide a motivating analysis of public health and financial costs from natural disasters in the U.S. using data provided by the National Centers for Environmental Information.

Keywords Bayesian hierarchical model · Gibbs sampler · Markov Chain Monte Carlo · Multiple response-types · Spike and slab

1 Introduction

Statistical models often assume the response variable is a single type (e.g., see Bradley 2022, for a discussion). By "type", we mean continuous, count-valued, or binary, among others. However, in many statistical applications, there are observed datasets consisting of multiple response-types. As an example, we consider data made available by NOAA's NCEI. In particular, one might be interested in finding useful covariates in predicting the responses: the number of natural disasters in the U.S., the financial costs (in billions of U.S. dollars) of natural disasters, and the mortality associated with natural disasters. Solving this variable selection problem is particularly important considering the rising cases of natural disasters due to climate change (e.g., Thuiller 2007, among others). However, this problem is particularly difficult considering that these variables likely exhibit multivariate dependence, and are mixed continuous and discrete-valued. Joint modeling of multi-type responses can provide a way to produce more precise estimates and predictions by leveraging dependence and including more (types of) data. Statistical approaches that are suitable for jointly modeling multiple response-type data include generalized linear mixed effects models (Christensen and Amemiya 2002; Schliep and Hoeting 2012), Markov models (Yang et al. 2014), copulas-based models (Liu et al. 2009; Dobra and Lenkoski 2011; Xue and Zou 2012), multi-task learning models (Argyriou et al. 2007; Kim and Xing 2009; Yang et al. 2009), random forests (Fellinghauer et al. 2013), and transformations (Bradley 2022). However, some of these approaches require one to make substantial modifications based on the types of

✉ Hongyu Wu
hongyu.sam.wu@gmail.com

Jonathan R. Bradley
jrbradley@fsu.edu

[1] Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306-4330, USA

responses, model transformations instead of the data directly, and create computational difficulties (e.g., see Schliep and Hoeting (2012), for a Metropolis-Hasting Gibbs sampler). Multiple response-type modeling is a relatively new area, and hence, it is of growing and important interest to develop new frameworks and tools for statistical analysis in this area. In particular, in this article, we focus on variable selection for multiple response-type data in a Bayesian framework.

To perform variable selection in a Bayesian context we make use of shrinkage global–local priors to enforce sparsity, which is widely used for estimating high-dimensional sparse parameters (Carvalho et al. 2010; Griffin and Brown 2010; Polson and Scott 2011). Specifically, in continuous shrinkage global–local priors, a "global" prior should enforce substantial shrinkage towards zero while a "local" prior with heavy tails should prevent over-shrinkage and capture local variation or "signals". One of the most well-known methods is the horseshoe prior (Carvalho et al. 2010), which can be defined as a Gaussian scale mixture of a global shrinkage parameter and a local shrinkage parameter for a regression coefficient. The horseshoe prior has been shown to have high performance in variable selection, estimation, and prediction. The concept of the global–local shrinkage has given rise to many different variants of horseshoe-like priors in a great amount of recent literature (e.g., see Bhadra et al. (2019), among others). However, the original horseshoe prior and most of the existing variants of the global–local shrinkage method are designed for the Gaussian settings. Analysis can be difficult when the data is non-Gaussian as the marginal likelihood of the mixed effects is not analytically available for non-Gaussian likelihoods. Thus, the extension to the non-Gaussian case is a growing topic in Bayesian variable selection. For instance, Datta and Dunson (2016) have extended the use of global–local priors to generalized linear models for count data, Kundu and Dunson (2014) consider semiparametric linear models where the errors are modeled non-parametrically, and a Gaussian approximation approach is developed by Piironen and Vehtari (2017) to approximate the posterior of the mixed effect coefficients. However, in the literature on Bayesian variable selection, most existing methods assume all responses are of the same type, either Gaussian or non-Gaussian, and the literature on variable selection for multiple response-type data is still relatively underdeveloped.

The multivariate logit-beta distribution (MLB), a special case of the conjugate multivariate distribution (CM), has been developed to facilitate Bayesian inference of non-Gaussian data from the natural exponential family (Bradley et al. 2019; Xu et al. 2023; Gao and Bradley 2019; Bradley et al. 2020; Bradley 2022). In this article, we are motivated to combine the horseshoe prior with this recent development in the non-Gaussian literature and propose a Bayesian variable selection for the multiple response-type data. The implementation of our approach is straightforward given the fact that the multivariate logit-beta prior is the conjugate prior for several members from the natural exponential family of distributions, which leads to the binomial/beta and negative binomial/beta hierarchical models. Specifically, we assume continuous data follows the MLB distribution, categorical data follows the binomial distribution, and count data follows the negative binomial distribution. What is unique with this strategy is that we do not need to model the data given a transformation or a Gaussian approximation, and thus avoid restrictions on the joint distribution of the responses. Furthermore, in our model, the dependence across response-types can be explicitly modeled. To do so, we assume dependence between responses can be modeled parsimoniously through basis functions (Wikle 2010) and our proposed global–local shrinkage model can be used to detect cross-variable dependence where the coefficients of the basis functions are interpreted as random effects. The use of soft-thresholding based Bayesian methods to select basis functions for dependent data has been used in the past in spatio-temporal models (Wikle and Holan 2011) and in the horseshoe method literature (Carvalho et al. 2010), among others. We also show that our novel use of the MLB prior has an explicit connection to the horseshoe prior under a Pólya-Gamma augmentation scheme (Polson et al. 2013), which provides evidence that the MLB distribution is reasonable for both Gaussian and non-Gaussian data. Additionally, we obtain an easy-to-implement block Gibbs sampler for our proposed model, which is similar to that of Hu and Bradley (2018), Bradley et al. (2019), and Bradley et al. (2020).

The remaining sections of this article are organized as follows. In Sect. 2, we introduce our proposed statistical model, develop connections to the existing horseshoe model, and describe the implementation. In Sect. 3, we present simulation studies that contains numerical comparisons with competing methods. Section 4 provides a data analysis of multiple response-type data for natural disasters in the U.S. using our proposed method. Section 5 contains a discussion.

2 Methodology

As discussed in the Introduction, we assume continuous data (e.g., financial costs of a particular type of natural disasters in billions of U.S. dollars) follows the MLB distribution, categorical data (e.g., number of a particular type of natural disasters) follows the binomial distribution, and count data (e.g., mortality) follows the negative binomial distribution. The MLB distribution is not a standard choice, and consequently, we provide a review in Sect. 2.1. Then in Sect. 2.2, we introduce the Bayesian hierarchical model for these aforementioned multiple response-type data and the model specifications. In Sects. 2.3 and 2.4, we establish a connection between the continuous-only setting of our model

and the traditional horseshoe model. Specifically, in Sect. 2.3, we first provide a connection between an MLB distribution and a multivariate Gaussian distribution using the Pólya-Gamma data-augmentation strategy introduced by Polson et al. (2013), and then in Sect. 2.4 we show that the shrinkage factor in our model also shares the same shrinkage feature as well as the one in the traditional horseshoe model. Section 2.5 provides details on model specifications.

2.1 Review of the multivariate logit-beta distribution

Define the n-dimensional random vector $\mathbf{w} = (w_1, ..., w_n)'$, where the n elements of this random vector are mutually independent logit-beta random variables. From Bradley et al. (2020), the ith element $w_i = \log\left(\frac{p}{1-p}\right)$, where $p \sim \text{Beta}(\alpha_i, \kappa_i - \alpha_i)$ and the corresponding shape parameters $\alpha_i > 0$ and $\kappa_i > \alpha_i$ for $i = 1, ..., n$. A logit-beta random variable w_i has probability density function (pdf) as follows:

$$f(w_i|\alpha_i, \kappa_i) = K(\alpha_i, \kappa_i)\exp\{\alpha_i w_i - \kappa_i \phi(w_i)\}, \tag{1}$$

where $K(\alpha_i, \kappa_i) = \frac{\Gamma(\kappa_i)}{\Gamma(\alpha_i)\Gamma(\kappa_i - \alpha_i)}$ is the normalizing constant and $\phi(q) = \log(1+\exp(q))$ for real q is the unit log partition function (Lehmann 1998).

One can allow for possible dependence by defining an n-dimensional random vector $\mathbf{Y} \equiv (Y_1, ..., Y_n)'$ such that

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{V}\mathbf{w}, \tag{2}$$

where $\mathbf{Y} \in \mathbb{R}^n$, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\mathbf{V}$ is a lower-triangle $n \times n$ invertible matrix. We call $\mathbf{Y}$ in Eq. (2) the multivariate logit-beta (MLB) random vector. The MLB random vector $\mathbf{Y}$ has the following pdf (Bradley et al. 2020):

$$\begin{aligned} &f(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \\ &\quad = \det(\mathbf{V}^{-1})\left\{\prod_{i=1}^{n} K(\alpha_i, \kappa_i)\right\}\exp\Big[\boldsymbol{\alpha}'\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \\ &\qquad - \boldsymbol{\kappa}'\log\{\mathbf{J}_{n,1} + \exp(\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu}))\}\Big] I(\mathbf{Y} \in \mathbb{R}), \end{aligned} \tag{3}$$

where "det" denotes the determinant function, $\mathbf{J}_{n,1}$ is a n-dimensional vector of 1's, $I(\cdot)$ is the indicator function, $\boldsymbol{\alpha} \equiv (\alpha_1, ..., \alpha_n)'$ and $\boldsymbol{\kappa} \equiv (\kappa_1, ..., \kappa_n)'$. We use $\text{MLB}(\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ as a shorthand for the pdf of a MLB distribution, where $\boldsymbol{\mu}$ is a location parameter, the invertible $\mathbf{V}$ is a covariance parameter, and $\boldsymbol{\alpha}$ and $\boldsymbol{\kappa}$ are the shape parameters.

As Bayesian inference requires simulating from a conditional distribution, we review the conditional logit-beta distribution (CMLB). Consider $\mathbf{Y} \sim \text{MLB}(\mathbf{V}\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ and partition this n-dimensional random vector into $(\mathbf{Y}_1', \mathbf{Y}_2')'$, where $\mathbf{Y}_1$ is r-dimensional and $\mathbf{Y}_2$ is $(n - r)$ dimensional. Similarly, partition $\mathbf{V}^{-1} = [\mathbf{H}, \mathbf{B}]$ into an $n \times r$ matrix $\mathbf{H}$ and an $n \times (n - r)$ matrix $\mathbf{B}$.

The CMLB $\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{d}, \boldsymbol{\mu}^*, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\kappa}$ is given by

$$\begin{aligned} &f(\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{d}, \boldsymbol{\mu}^*, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \\ &\quad = \frac{\Big[f(\mathbf{Y}|\mathbf{V}\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa})\Big]_{\mathbf{Y}_2 = d}}{\Big[f(\mathbf{Y}|\mathbf{V}\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa})d\mathbf{Y}_1\Big]_{\mathbf{Y}_2 = d}} \\ &\quad \propto \exp\Big[\boldsymbol{\alpha}'\left(\mathbf{H}, \mathbf{B}\right)\begin{pmatrix}\mathbf{Y}_1 \\ \mathbf{d}\end{pmatrix} \\ &\qquad - \boldsymbol{\kappa}'\log\{\mathbf{J}_{r,1} + \exp(\left(\mathbf{H}, \mathbf{B}\right)\begin{pmatrix}\mathbf{Y}_1 \\ \mathbf{d}\end{pmatrix} - \boldsymbol{\mu})\}\Big] \\ &\quad \propto \exp\Big[\boldsymbol{\alpha}'\mathbf{H}\mathbf{Y}_1 - \boldsymbol{\alpha}'\boldsymbol{\mu}^* \\ &\qquad - \boldsymbol{\kappa}'\log\{\mathbf{J}_{r,1} + \exp(\mathbf{H}\mathbf{Y}_1 - \boldsymbol{\mu}^*)\}\Big], \end{aligned} \tag{4}$$

where $\boldsymbol{\mu}^* = \boldsymbol{\mu} - \mathbf{B}\mathbf{d}$, $\mathbf{d} \in \mathbb{R}^{n-r}$. We define $\text{CMLB}(\boldsymbol{\mu}^*, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ as a shorthand for the pdf in Eq. (4).

We do not know how to simulate from (4) directly; however, we do know how to simulate from the MLB in (3) through (2). This motivated Bradley et al. (2019), Gao and Bradley (2019), and Bradley et al. (2020) to introduce a nuisance term into the model so that the implied full-conditional distribution of the mixed effects and nuisance parameter is MLB as opposed to a CMLB. This allows one to update fixed and random effects using the MLB. The nuisance parameters are marginalized from the posterior distribution and estimated to be zero. In Sect. 2.2, we adopt the same strategy.

2.2 The Bayesian hierarchical model for multiple response-type data

Let Y_{ij} be the multiple response-type data for subject i of response-type j, where $i = 1, ..., I_j$ and $j = 1, 2, 3$. For each subject i, Y_{i1} is continuous-valued, Y_{i2} is integer-valued, and Y_{i3} is count-valued. For ease of notation, we denote the vector of each type of response variable with $\mathbf{Y}_j = (Y_{1j}, ..., Y_{I_j j})'$, and the entire $(\sum_{j=1}^{3} I_j)$-dimensional vector $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \mathbf{Y}_3')'$. Let $\mathbf{Z}_j$ be an $I_j \times p$ observable matrix of predictor variables, where $j = 1, 2, 3$, and the columns of $\mathbf{Z}_j$ consist of both covariates and basis functions. That is, $\mathbf{Z}_j = \left[\mathbf{X}_j, \mathbf{G}_j\right]$, where $\mathbf{X}_j$ is a known $I_j \times p_x$ matrix of covariates and $\mathbf{G}_j$ is an $I_j \times p_g$ matrix of basis functions such that $p = p_x + p_g$. For each pair of response-types, we define a pair of random effects associated with $\mathbf{G}_{jj'}^{[j]}$ and $\mathbf{G}_{jj'}^{[j']}$, where $j = 1, 2, 3$, $j' = 1, 2, 3$, and $j \neq j'$. For example, $\mathbf{G}_{12}^{[1]}$ and $\mathbf{G}_{12}^{[2]}$ are $I_1 \times r$ and $I_2 \times r$, and both have the same random effects in the models for $\mathbf{Y}_1$ and $\mathbf{Y}_2$. We let the interaction random effects consist of pre-defined basis functions or known covariates (see Sect. 2.5 for discussion).

Define the design matrices

$$\begin{aligned}\mathbf{X}_1^* &= \left[\mathbf{Z}_1\ \mathbf{0}_{I_1,p}\ \mathbf{0}_{I_1,p}\ \mathbf{G}_{12}^{[1]}\ \mathbf{G}_{13}^{[1]}\ \mathbf{0}_{I_1,r}\right],\\ \mathbf{X}_2^* &= \left[\mathbf{0}_{I_2,p}\ \mathbf{Z}_2\ \mathbf{0}_{I_2,p}\ \mathbf{G}_{12}^{[2]}\ \mathbf{0}_{I_2,r}\ \mathbf{G}_{23}^{[2]}\right],\\ \mathbf{X}_3^* &= \left[\mathbf{0}_{I_3,p}\ \mathbf{0}_{I_3,p}\ \mathbf{Z}_3\ \mathbf{0}_{I_3,r}\ \mathbf{G}_{13}^{[3]}\ \mathbf{G}_{23}^{[3]}\right],\end{aligned} \tag{5}$$

where $\mathbf{0}_{m,n}$ is a $m \times n$ matrix of zeros. We assume $\mathbf{Y}_j$ has a linear mixed effects representation on the appropriate link scale denoted with $\mathbf{X}_j^*\boldsymbol{\beta}$. The latent representation $\mathbf{X}_j^*\boldsymbol{\beta}$ has overlapping (across j) random effects to enforce dependence across these three data types, where $\mathbf{X}_j^* \in \mathbb{R}^{I_j} \times \mathbb{R}^{3p+3r}$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\eta}_1', \boldsymbol{\beta}_2', \boldsymbol{\eta}_2', \boldsymbol{\beta}_3', \boldsymbol{\eta}_3', \boldsymbol{\eta}_{12}', \boldsymbol{\eta}_{13}', \boldsymbol{\eta}_{23}')'$, $\boldsymbol{\beta}_j \in \mathbb{R}^{p_x}$, $\boldsymbol{\eta}_j \in \mathbb{R}^{p_g}$, and $\boldsymbol{\eta}_{jj'} \in \mathbb{R}^r$. Stacking fixed and random effects into a single vector $\boldsymbol{\beta}$ simplifies Gibbs sampling to have fewer steps. However, it is important to emphasize that $\boldsymbol{\beta}$ can contain both fixed and random effects. That is,

$$\mathbf{X}_j^*\boldsymbol{\beta} = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{G}^{(j)}\boldsymbol{\eta}^{(j)}, \tag{6}$$

where $\mathbf{G}^{(j)} = \left[\mathbf{G}_j, \mathbf{G}^{[j]}_{\min(j,m),\max(j,m)}, \mathbf{G}^{[j]}_{\min(j,k),\max(j,k)}\right]$ and $\boldsymbol{\eta}^{(j)} = (\boldsymbol{\eta}_j', \boldsymbol{\eta}'_{\min(j,m),\max(j,m)}, \boldsymbol{\eta}'_{\min(j,k),\max(j,k)})'$ for $j \neq m$, $j \neq k$, $k \neq m$, and $j, m, k \in \{1, 2, 3\}$. The implied covariance is given by,

$$\begin{aligned}&\text{cov}(\mathbf{X}_j^*\boldsymbol{\beta}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) = \mathbf{G}^{(j)}\text{cov}(\boldsymbol{\eta}^{(j)}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)\mathbf{G}^{(j)'}\\ &\text{cov}(\mathbf{X}_j^*\boldsymbol{\beta}, \mathbf{X}_{j'}^*\boldsymbol{\beta}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)\\ &\quad = \mathbf{G}^{[j]}_{jj'}\text{cov}(\boldsymbol{\eta}_{jj'}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)\mathbf{G}^{[j']'}_{jj'};\ j < j',\end{aligned} \tag{7}$$

where the first Eq. in (7) is of the same form as covariance parameterizations using basis function expansions (e.g., see Cressie and Johannesson (2006), Wikle (2010), and Cressie and Wikle (2011) for standard references), and the second equation uses the parameterization for multi-type cross-covariances based on basis function expansions found in more recent multi-type data literature (Bradley 2022; Xu et al. 2023). The cross-dependence between data types is modeled via basis function expansions where the coefficients are interpreted as random effects, and the matrix $\mathbf{G}^{[j]}_{jj'}$ are predefined basis functions. In general, there are several choices for $\mathbf{G}^{[j]}_{jj'}$ and $\mathbf{G}^{[j']}_{jj'}$ and can be specified to be different. In practice, we suggest the use of complete basis functions so that one can flexibly model any dependence structure, and set $\mathbf{G}^{[j]}_{jj'} = \mathbf{G}^{[j']}_{jj'}$. See Sect. 2.5 for more details on the available choices for basis functions.

One of the main goals of this article is to perform variable selection for multiple response-type data by putting a sparsity-inducing prior (i.e., the global–local shrinkage prior) on the mixed effects coefficients $\boldsymbol{\beta}$ to allow one to choose a subset of the given covariates/basis functions for a prediction model. However, a traditional linear model (e.g., the horseshoe prior) does not allow for both different response-types and explicitly modeling the cross-variable dependence. Thus, we consider the following specifications of the hierarchical model for multiple response-type data:

$$\begin{aligned}&\mathbf{Y}_1|\boldsymbol{\beta}, \tau, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \mathbf{q}_1 \sim \text{MLB}(\mathbf{X}_1^*\boldsymbol{\beta} + \tau^{-1}\mathbf{q}_1, \tau^{-1}\mathbf{I}_{I_1}, \boldsymbol{\alpha}, \boldsymbol{\kappa});\\ &\mathbf{Y}_2|\boldsymbol{\beta}, \mathbf{n}, \mathbf{q}_2 \sim \text{Binomial}\{\mathbf{n}, \text{logit}^{-1}(\boldsymbol{X}_2^*\boldsymbol{\beta} - \mathbf{q}_2)\};\\ &\mathbf{Y}_3|\boldsymbol{\beta}, \mathbf{r}, \mathbf{q}_3 \sim \text{NB}\{\mathbf{r}, \text{logit}^{-1}(\boldsymbol{X}_3^*\boldsymbol{\beta} - \mathbf{q}_3)\};\\ &\boldsymbol{\beta}|\tau_\beta, \boldsymbol{\Lambda}, \boldsymbol{\alpha}_\beta, \boldsymbol{\kappa}_\beta, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_\beta\\ &\quad \sim \text{CMLB}\left\{\begin{pmatrix}\mathbf{q}_2\\ \mathbf{q}_3\\ \mathbf{q}_\beta\end{pmatrix}, \begin{pmatrix}\mathbf{X}_2^*\\ \mathbf{X}_3^*\\ \tau_\beta\boldsymbol{\Lambda}\end{pmatrix}, \begin{pmatrix}0.5\mathbf{J}_{I_2}\\ 0.5\mathbf{J}_{I_3}\\ \boldsymbol{\alpha}_\beta\end{pmatrix}, \begin{pmatrix}\mathbf{J}_{I_2}\\ \mathbf{J}_{I_3}\\ \boldsymbol{\kappa}_\beta\end{pmatrix}\right\};\\ &\lambda_j|a, b \sim \text{IB}(a, b);\\ &\tau|a_\tau, b_\tau \sim \text{IB}(a_\tau, b_\tau);\\ &\tau_\beta|a_{\tau_\beta}, b_{\tau_\beta} \sim \text{IB}(a_{\tau_\beta}, b_{\tau_\beta}),\end{aligned} \tag{8}$$

where logit^{-1} denotes the inverse-logit function, i.e., $\text{logit}^{-1}(p) = \frac{\exp(z)}{1+\exp(z)}$ for a real z; Binomial(n, p) is a shorthand for the binomial distribution with $n > 0$ number of trials and probability of success $p \in (0, 1)$; NB(r, p) is a shorthand for the negative binomial distribution with $r > 0$ number of failures until the experiment is stopped and probability of success $p \in (0, 1)$; and IB(a, b) is a shorthand for the inverted-beta distribution with shape parameters $a > 0$ and $b > 0$. Let $\mathbf{I}_m$ be an $m \times m$ identity matrix and $\mathbf{J}_n$ be an n-dimensional vector of ones. Let τ be the global shrinkage parameter for $\mathbf{Y}_1$. An unknown vector $\boldsymbol{\beta} = (\beta_1, ..., \beta_{3p+3r})' \in \mathbb{R}^{3p+3r}$ contains the mixed effects (i.e., fixed and random) coefficients for all the multiple response-type data. The global shrinkage parameter for $\boldsymbol{\beta}$ is labeled as τ_β. The diagonal matrix $\boldsymbol{\Lambda}$ is $(3p+3r) \times (3p+3r)$, and the l^{th} diagonal element corresponds to λ_l, the local shrinkage parameter for β_l, where $l = 1, ..., 3p + 3r$. In the model for $\boldsymbol{\beta}$, the global shrinkage parameter τ_β is used to shrink all the coefficients of $\boldsymbol{\beta}$ towards zero, while the local shrinkage parameter λ_l for β_l gives the coefficient the flexibility to transcend that shrinkage as needed. This allows the model to conduct variable selection and detect cross-variable dependence. We use an inverted beta distribution as prior on each global–local shrinkage parameter. The inverted beta distribution is also known as a beta prime distribution, which belongs to a class of hypergeometric inverted-beta family that generalizes various priors in the Bayesian literature (Polson and Scott 2012) and hence can be used as a prior for the Bernoulli, binomial, negative binomial, and geometric distributions. One can consider using a half-Cauchy prior $C^+(0, 1)$ as a weakly informative prior (Carvalho et al. 2009). There is a probabilistic transformation that links the inverted beta and half-Cauchy distributions and both of them, as priors, can result in a horseshoe-shaped

shrinkage profile in a global-local shrinkage setting. In terms of variable selection, the distribution has the flexibility to be heavy-tailed, which allows strong signals to escape the global shrinkage. Define the shape parameters for $\mathbf{Y}_1$ to be $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{I_1})'$ and $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_{I_1})'$, and let each value be strictly positive values such that $\boldsymbol{\kappa} > \boldsymbol{\alpha}$. The parameters $\boldsymbol{\alpha}_\beta = (\alpha_{\beta,1}, ..., \alpha_{\beta,3p+3r})'$ and $\boldsymbol{\kappa}_\beta = (\kappa_{\beta,1}, ..., \kappa_{\beta,3p+3r})'$ represent shape parameters for $\boldsymbol{\beta}$. All the shape parameters are given improper priors.

The shape parameters of logit-beta distribution need, for example, $\kappa > \alpha$. Adding "0.5" and "1" to the elements of shape parameters of the prior of $\boldsymbol{\beta}$ ensures shape parameters that satisfy this constraint. This is similar to adding a random effect to normal data to guarantee that the posterior density does not allow for parameters on the boundary of the parameter space. By doing so, the prior of $\boldsymbol{\beta}$ can be written as a CMLB distribution. The global–local shape parameters are updated via a slice sampling scheme and a latent variables approach to obtain full-conditional distributions that have the same form as the likelihoods.

To sample directly from the full-conditional distribution, we introduce nuisance location parameters $\mathbf{q}_1 \in \mathbb{R}^{I_1}$, $\mathbf{q}_2 \in \mathbb{R}^{I_2}$, $\mathbf{q}_3 \in \mathbb{R}^{I_3}$, $\mathbf{q}_\beta \in \mathbb{R}^{3p+3r}$, and $\mathbf{q}_B \equiv (\mathbf{q}_1', \mathbf{q}_2', \mathbf{q}_3', \mathbf{q}_\beta')'$ that lead to straightforward block Gibbs sampling. In particular, when the nuisance location parameters are given an improper prior as described in Supplementary Appendix A, we obtain that $(\boldsymbol{\beta}, \mathbf{q}_B)$ follows an MLB distribution, which can be sampled using the form in Eq. (2). Following Bradley et al. (2019), Gao and Bradley (2019), and Bradley et al. (2020), since $\mathbf{q}_\beta$ is interpreted as a nuisance parameter, it is marginalized out and estimated to be a zero vector. Details are provided in Supplementary Appendix A to update $\boldsymbol{\beta}$ in this way.

2.3 Pólya-gamma distributed precision parameters

One can establish a connection between an MLB distribution and a multivariate Gaussian distribution using the Pólya-Gamma data-augmentation strategy from Polson et al. (2013). We say a random variable Z has a Pólya-Gamma distribution, denoted $Z \sim \text{PG}(c, d)$, if

$$Z \overset{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k-1/2)^2 + d^2/(4\pi^2)} \tag{9}$$

where $c > 0, d \in \mathbb{R}$ and $g_k \sim \text{Gamma}(c, 1)$ are independent gamma random variables.

Consider $\mathbf{w} \sim \text{MLB}(\boldsymbol{\mu}, \tau^{-1}\mathbf{I}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$, then the integral identity for a w_i underlying the Pólya-Gamma augmentation scheme (Polson et al. 2013) can be used as:

$$\begin{aligned} f(w_i|\mu_i, \tau, \alpha_i, \kappa_i) &\propto \frac{\{\exp(\tau(w_i - \mu_i))\}^{\alpha_i}}{\{1 + \exp(\tau(w_i - \mu_i))\}^{\kappa_i}} \\ &= 2^{-\kappa_i} \exp\left(\frac{t_i(w_i - \mu_i)}{\tau^{-1}}\right) \int_0^\infty \exp\left(\frac{-\rho_i(w_i - \mu_i)^2}{2\tau^{-2}}\right) \\ &\quad p(\rho_i|\kappa_i, 0) d\rho_i, \end{aligned} \tag{10}$$

which implies

$$\begin{aligned} &f(w_i|\mu_i, \tau, \alpha_i, \kappa_i, \rho_i) \\ &\quad \propto \exp\left(\frac{t_i(w_i - \mu_i)}{\tau^{-1}}\right) \exp\left(\frac{-\rho_i(w_i - \mu_i)^2}{2\tau^{-2}}\right) \\ &\quad \propto \exp\left(-\frac{\rho_i \tau^2}{2}\left(w_i - \left(\mu_i + \frac{t_i}{\tau\rho_i}\right)\right)^2\right), \end{aligned} \tag{11}$$

where $t_i = (\alpha_i - \kappa_i/2)$ and $p(\rho_i|\kappa_i, 0)$ is denoted as the density of $\rho_i \sim \text{PG}(\kappa_i, 0)$. The term outside the integral part in Eq. (10) contains the shape parameters α_i and κ_i, while the integrand is the kernel of a Gaussian distribution. If we condition on ρ_i, by completing the square and a few steps of algebra, the integral leads to a Gaussian in w_i, that is, $w_i|\mu_i, \tau, \alpha_i, \kappa_i, \rho_i \sim \text{N}(\mu_i + \frac{t_i}{\tau\rho_i}, (\rho_i\tau^2)^{-1})$.

In a continuous-only setting, we simply consider:

$$\begin{aligned} &\mathbf{Y}_1|\boldsymbol{\beta}, \tau, \boldsymbol{\alpha}, \boldsymbol{\kappa} \sim \text{MLB}(\mathbf{X}_1^*\boldsymbol{\beta}, \tau^{-1}\mathbf{I}, \boldsymbol{\alpha}, \boldsymbol{\kappa}); \\ &\boldsymbol{\beta}|\tau_\beta, \boldsymbol{\Lambda}, \boldsymbol{\alpha}_\beta, \boldsymbol{\kappa}_\beta \sim \text{MLB}(\mathbf{0}_{p,1}, \tau_\beta^{-1}\boldsymbol{\Lambda}^{-1}, \boldsymbol{\alpha}_\beta, \boldsymbol{\kappa}_\beta), \end{aligned} \tag{12}$$

where for discussion, we have set the nuisance parameters to zero. When applying Eqs. (10)–(12), we have $\mathbf{Y}_1$ and $\boldsymbol{\beta}$ become:

$$\begin{aligned} &\mathbf{Y}_1|\boldsymbol{\beta}, \tau, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \mathbf{V} \sim \text{N}(\mathbf{X}_1^*\boldsymbol{\beta} + \mathbf{m}, \mathbf{V}^{-1}); \\ &\boldsymbol{\beta}|\tau_\beta, \boldsymbol{\Lambda}, \boldsymbol{\alpha}_\beta, \boldsymbol{\kappa}_\beta, \mathbf{V}_\beta \sim \text{N}(\mathbf{m}_\beta, \mathbf{V}_\beta^{-1}), \end{aligned}$$

where $\mathbf{m} = \left(\frac{\alpha_1 - \kappa_n/2}{\tau\rho_1}, ..., \frac{\alpha_n - \kappa_n/2}{\tau\rho_n}\right)$, $\mathbf{V} = \text{diag}\left(\rho_1\tau^2, ..., \rho_n\tau^2\right)$, $\mathbf{m}_\beta = \left(\frac{\alpha_{\beta,1} - \kappa_{\beta,1}/2}{\tau_\beta\rho_{\beta,1}}, ..., \frac{\alpha_{\beta,p} - \kappa_{\beta,p}/2}{\tau_\beta\rho_{\beta,p}}\right)$, $\mathbf{V}_\beta = \text{diag}(\rho_{\beta,1}\tau_\beta^2\lambda_1^2, ..., \rho_{\beta,p}\tau_\beta^2\lambda_p^2)$, $\rho_i \sim PG(\kappa_i, 0)$, and $\rho_{\beta,j} \sim \text{PG}(\kappa_{\beta,j}, 0)$. This conditionally conjugate augmentation scheme leads to a connection between the MLB distribution and the Gaussian distribution used in the horseshoe model after integrating out the extra local Pólya-Gamma distributed precision $\{\rho_i\}$ and $\{\rho_{\beta,j}\}$. Furthermore, the shape parameters in $\mathbf{m}$ and $\mathbf{m}_\beta$ allow one to model continuous non-Gaussian data, since upon marginalization of $\{\rho_i\}$ and $\{\rho_{\beta,j}\}$, we obtain a possibly skewed logit-beta distribution. The key point of this section is that our proposed MLB model is equivalent to a horseshoe model after introducing and marginalizing $\{\rho_i\}$ and $\{\rho_{\beta,j}\}$. The motivation for the use of the logit-beta distribution instead of the more traditional Gaussian distribution

is that the logit-beta distribution is conjugate with the logit-beta, binomial, and negative binomial likelihoods. This will allow us to model multiple response-type data in an efficient way.

2.4 The shrinkage factor

In the continuous-only setting, the predictive mean for the coefficients $\boldsymbol{\beta}$ can be derived as (See Supplementary Appendix A for details):

$$\begin{aligned} &\mathrm{E}(\boldsymbol{\beta}|\mathbf{Y}_1, \mathbf{X}_1^*, \tau_\beta, \boldsymbol{\Lambda}, \tau) \\ &= (\tau^2 \mathbf{X}_1^{*'} \mathbf{X}_1^* + \tau_\beta^2 (\boldsymbol{\Lambda}' \boldsymbol{\Lambda}))^{-1} (-\tau \mathbf{X}_1^{*'} \mathbf{w}_1 + \tau_\beta \boldsymbol{\Lambda} \mathbf{w}_2 \\ &\quad + \tau^2 \mathbf{X}_1^{*'} \mathbf{Y}_1), \end{aligned} \tag{13}$$

where I_1-dimensional vector $\mathbf{w}_1 \sim \mathrm{MLB}(\mathbf{0}, \mathbf{I}_{I_1}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ and p-dimensional vector $\mathbf{w}_2 \sim \mathrm{MLB}(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\alpha}_\beta, \boldsymbol{\kappa}_\beta)$. As discussed in the introduction, this modeling framework can model both symmetric and skewed data by specifying the shape parameters. It is worth noting that, when $\boldsymbol{\kappa} = 2\boldsymbol{\alpha}$ and $\boldsymbol{\kappa}_\beta = 2\boldsymbol{\alpha}_\beta$, we can have $E(\mathbf{w}_1) = 0$ and $E(\mathbf{w}_2) = 0$ (see Supplementary Appendix B for a discussion). Hence, when $I_1 = p, \mathbf{X}_1^* = \mathbf{I}$ is the identity, $\tau = \tau_\beta = 1, \boldsymbol{\kappa} = 2\boldsymbol{\alpha}$, and $\boldsymbol{\kappa}_\beta = 2\boldsymbol{\alpha}_\beta$, the conditional posterior mean for the j^{th} component of $\boldsymbol{\beta}$, denoted β_j, becomes

$$\mathrm{E}(\beta_j | y_j, \lambda_j^2) = (1 - k_j) y_j, \tag{14}$$

where

$$k_j = \frac{1}{1 + \lambda_j^{-2}} \tag{15}$$

can be defined as the shrinkage factor for $\beta_{1,j}$ in our method, which shares the same shrinkage feature with the one for the horseshoe prior model (See Supplementary Appendix C for a review of the horseshoe priors). Choosing $\lambda_j \sim \mathrm{IB}(a, b)$ can result in different patterns of k_j. The density function of k_j lacks a closed-form representation, but it can behave similarly to a beta distribution. The parameters a and b in the inverted-beta distribution are analogous to those of the beta distribution, allowing the probability density function of k_j to place more mass either close to 0 or 1. For example, a special case occurs when setting $a = b = 1/2$, which yields a horseshoe-shaped profile of k_j (see Fig. 1) and develops the same feature of distinguishing between noise and signal as the traditional horseshoe model.

2.5 Specification of basis functions

We now give further discussion on choices for basis functions to use in our model. One immediate choice is to use known covariates, including functions of covariates, pairwise-interactions among covariates, and even the original covariates. The downside of this choice is that it can cause confounding among the mixed effect coefficients (Greenland et al. 1999). An option to alleviate confounding is to define an orthogonal projection of the covariates (Griffith 2000, 2002, 2004). That is, one can let $\mathbf{X}_{jj'}^{[j]}$ contain linear combinations of columns from $\mathbf{I} - \mathbf{X}_j (\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j'$. An example is Moran's I basis functions (Hughes and Haran 2013), which are defined to be the eigenvectors of an orthogonal projection of the covariates and are considered as reduced-dimensional basis functions aiding in analyzing large-dimensional data. The other responses are allowed to act as covariates as well provided that the hierarchical model is well defined (e.g., $\mathbf{Y}_1|\mathbf{Y}_2, \mathbf{Y}_3$ and $\mathbf{Y}_2|\mathbf{Y}_3$). In general, the choices are not limited to the use of covariates and may include the radial basis functions, the spline basis functions, wavelet basis functions, Fourier basis functions among many others (see Bradley et al. 2017; Cressie and Johannesson 2008; Wahba 1990; Donoho and Johnstone 1994, for examples). For instance, a radial basis function depending on equally spaced knot locations, referred to as a bisquare function (Cressie and Johannesson 2008), is often used in analyzing multi-resolutional point-referenced data in spatial settings.

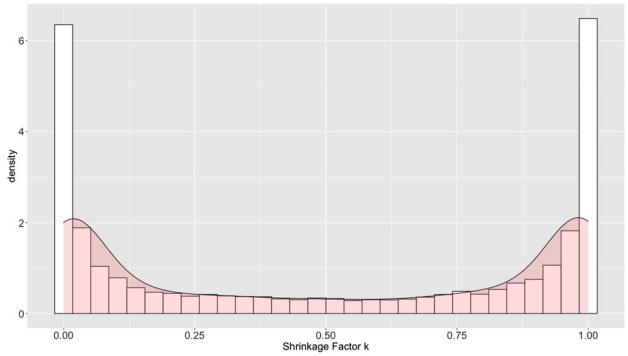

Fig. 1 Density plot for the shrinkage factor $k_j \in [0, 1]$

In general, a complete class of basis functions allows one to approximate any random function (with any dependence structure) in L_2 provided enough basis functions are used. Well-known results, like the Karhunen-Loéve expansion and Mercer's theorem show that any orthogonal basis can approximate an unknown function and unknown covariance function well (Karhunen 1946; Riesz and Nagy 2012; Huang et al. 2001; Daw et al. 2022). Similar results have been developed for non-orthogonal basis matrices (Obled and Creutin 1986; Bradley et al. 2017). In practice, these expansions can not be evaluated with infinite basis functions, leading to reduced rank basis expansions, and as such, variable selection methods are an invaluable tool in this literature. In our application, we make use of Fourier basis functions, which are L_2 complete. These choices are common in the spatial

random effects, spatio-temporal, and time series literature (e.g., see Cressie and Wikle 2011, pg 102).

The choice of r is also important. When the random effects consist of known covariates/functions/interactions, the value of r is a feature of the observed dataset (see Sect. 3 for examples). However, when a pre-defined basis function is used, the value of r has to be selected. The selection is often determined by certain selection criteria, such as Akaike information criterion and deviance information criterion (e,g, see Bradley et al. 2011, among others). In this article, we, of course, use our proposed model to select covariates and basis functions, and consider r as large as possible such that we obtain a reasonable diagnostic measure of goodness-of-it of the model, such as posterior predictive p-value (Meng 1994; Gelman et al. 2003; Gelman 2013), which suggests the model provides a fit to the dataset.

3 Simulation

The goal of the simulation is to illustrate that our model provides reasonable performances in terms of variable selection and estimation for multiple response-type data. In addition, we compare our method with a version of the horseshoe model that transforms multiple response-type data (Bradley 2022). We consider several specifications of a simulation model and apply the model in Eq. (8) to analyze the simulated data.

We are particularly interested in the case when p grows. It is well known that variable selection methods tend to perform worse as p increases, as the inclusion of unnecessary covariates and basis functions creates a very large parameter space that is difficult to search through and affects the MCMC properties (Chen et al. 2011; Garcia-Donato and Martinez-Beneito 2013; Griffin et al. 2021). As such, we provide metrics for both estimation and variable selection when applying our algorithm with various choices of p.

3.1 Simulation setup

To generate $\mathbf{Z}_j$, we specify the I_j-dimensional vector $\mathbf{u}_{j,k} \sim$ Uniform(0, 1), where Uniform(0, 1) is a shorthand for the uniform distribution over the interval $[0, 1]$, $j = 1, 2, 3$, and $k = 1, ..., p$. We let the first two vectors $\mathbf{z}_{j,1}$ and $\mathbf{z}_{j,2}$ be equal to $\sin(\pi \mathbf{u}_{j,1}\mathbf{u}_{j,2})$ and $(\mathbf{u}_{j,3} - 0.5)^2$ respectively to include non-linear terms and let $\mathbf{z}_{j,k} = \mathbf{u}_{j,k}$ for $k = 3, ..., p$. This specification is similar to the simulation design in Friedman (1991). Each element in $\mathbf{z}_{j,k}$ is independent and $\mathbf{Z}_j = (\mathbf{z}_{j,1}, ..., \mathbf{x}_{j,p})$. To specify the random effects in the simulation model, we consider r randomly selected pairwise interactions among covariates (i.e., the element-wise product of two vectors in $\mathbf{Z}_j$, where these two vectors are randomly selected) to define $\mathbf{G}^{[j]}_{jj'}$ and $\mathbf{G}^{[j']}_{jj'}$. This allows us to have the design matrix $\mathbf{X}_j$ in Eq. (5). This simulation has $I_1 = 35$, $I_2 = 35$, $I_3 = 30$, and $r = 6$. We choose a $(3p + 3r)$-dimensional coefficient vector $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \boldsymbol{\beta}'_3, \boldsymbol{\eta}'_{12}, \boldsymbol{\eta}'_{13}, \boldsymbol{\eta}'_{23})'$, where we fix the first 6 elements of each main effect coefficient vector, that is,

$$\begin{aligned} \boldsymbol{\beta}_1 &= (10, 20, 10, 10, 10, 10, 0, ..., 0)', \\ \boldsymbol{\beta}_2 &= (0, 1, 1, 0, 0, 1, ..., 0)', \\ \boldsymbol{\beta}_3 &= (2, 2, 2, 0, 0, 0, ..., 0)', \end{aligned} \tag{16}$$

and the interaction effect coefficient vectors are set equal to the following realizations, $\boldsymbol{\eta}_{12} = (1, 1, 0, 0, 0, 0)'$, $\boldsymbol{\eta}_{13} = (0, 1, 1, 0, 0, 0)'$, and $\boldsymbol{\eta}_{23} = (0, 0, 1, 1, 0, 0)'$. Thus, every non-zero (zero) element of $\boldsymbol{\beta}$ corresponds to a covariate that is (not) useful for prediction and should (not) be selected. The multiple response-type data are simulated according to the following specifications for the distribution associated with the data $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2, \mathbf{Y}'_3)'$:

$$\begin{aligned} \mathbf{Y}_1 &\sim \mathrm{N}(\mathbf{X}_1\boldsymbol{\beta}, \mathbf{I}_{I_1}), \\ \mathbf{Y}_2 &\sim \text{Binomial}(40\mathbf{J}_{I_2}, \text{logit}^{-1}(\mathbf{X}_2\boldsymbol{\beta})), \\ \mathbf{Y}_3 &\sim \mathrm{NB}(20\mathbf{J}_{I_3}, \text{logit}^{-1}(\mathbf{X}_3\boldsymbol{\beta})), \end{aligned} \tag{17}$$

where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a shorthand for the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$.

The data models in (17) represent the more traditional data model assumptions in the Bayesian hierarchical modeling literature (Cressie and Wikle 2011). The model in (17) differs from our model in two ways: (1) our model assumes $\mathbf{Y}_1$ is MLB instead of normal; and (2) we consider fitting our model with incorrect choices of p. As we discuss in Sect. 2.3 the MLB itself is a scale mixture of normals, and hence is different from the normal distribution (see Supplementary Appendix B for general details on the MLB). Considering the ever-present nature of normal data that arise due to central limit theorems, in Supplementary Appendix D we provide a sensitivity study to assess the robustness of our model to the setting when the data are generated from the normal distribution. In particular, we compare to the horseshoe model, which assumes the data are normal, when the data are generated solely from a normal distribution (i.e., is not of mixed type). In Supplementary Appendix D, We found that we perform nearly identical to the horseshoe in terms of variable selection. The correctly specified horseshoe model has better estimation performance for an extremely small signal-to-noise ratio, and the difference between the models became practically negligible as the signal-to-noise ratio increased.

The Poisson distribution is an alternative choice to the negative binomial specification in (17), however, it is well known that the Poisson distribution is a limiting case of the negative binomial. The binomial distribution is by far the standard dis-

tributional assumption for the number of success outcomes out of a finite number of trials (McCullagh 2019). Other alternatives to the simulation data models in (17) could include zero-inflated models (Hall 2000) and the Conway-Maxwell Poisson (COM-Poisson) distribution (Sellers 2023). However, our model can not be immediately adapted to these alternatives due to our use of conjugacy, which highlights a limitation of our approach. One should not consider our model if zero-inflation is present or the assumption of COM-Poisson appears appropriate. In practice, posterior predictive p-values can be used to assess how appropriate our model assumptions are (Meng 1994; Gelman et al. 2003; Gelman 2013). To aid with comparison, we also use the same link function between the simulated data and our proposed model (i.e., the logit) so that the estimated coefficients are on the same scale as the true coefficients.

We consider 20 replicate datasets. In each replicate, our method in Sect. 2.2 is implemented using the block Gibbs sampler in Supplementary Appendix A to generate 5,000 iterations with a burn-in of 2,000. We choose $a = b = 0.5$ for the local shrinkage parameter to produce a horseshoe-shaped shrinkage pattern. We let $a_\tau = b_\tau = a_{\tau_\beta} = b_{\tau_\beta} = 10$ for the global shrinkage parameter. We denote the estimated posterior mean of $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$ and define the root mean squared error (RMSE) to be

$$\left\{ \frac{(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}})}{I_1 + I_2 + I_3} \right\}^{1/2} \tag{18}$$

to evaluate the estimation performance, where $\mathbf{X} = (\mathbf{X}_1^{'}, \mathbf{X}_2^{'}, \mathbf{X}_3^{'})'$. We also report the average false negative rate (FNR) and the average false positive rate (FPR) to assess the model's performance of selecting variables. Specifically, FPR (FNR) are the proportion out of the zero (non-zero) elements of $\boldsymbol{\beta}$ such that the method suggests the variable is non-zero (zero). We provide a plot of receiver operating characteristic (ROC) curve to assess our ability to do hard-threshold. Recall that soft variable selection methods can be used to select variables by specifying a threshold value c, such that $|\beta_l| < c$ suggests the variable is zero. Here we choose $c = 0.5$. When threshold is changing, the ROC curve presents the trade-off between sensitivity (i.e., true positive rate) and specificity (i.e., 1−false positive rate), where both vary from 0 to 1. Generally, when the method suggests more elements as non-zero (positive), the true positives will increase, but this will also cause the false positives to increase. We also calculate the area under the curve (AUC), which is equivalent to the two sample Wilcoxon rank-sum statistic (Mann and Whitney 1947). In practice, the higher the AUC, the better the model is at selecting variables. To compute the FPR and FNR in the simulation study, we apply credible set. Specifically, we see whether a 95% credible interval of a posterior mean of β_l contains zero or not to determine if the mixed effect coefficient is a zero or not.

Table 1 Average metrics over 20 independent replicates by p

p	10	20	30	40
RMSE	0.554	0.580	0.592	0.614
FNR	0	0	0.014	0.106
FPR	0.002	0	0	0

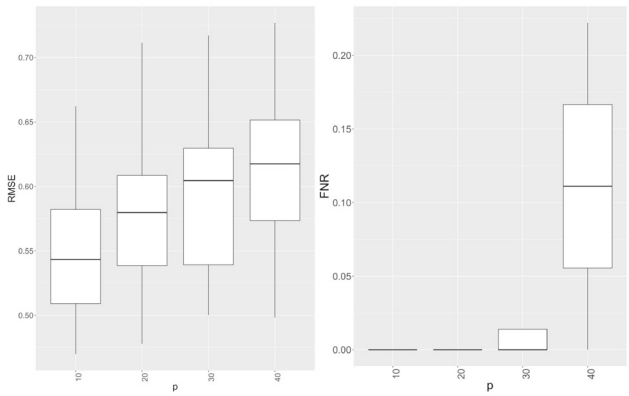

Fig. 2 In the left panel we plot a boxplot of the RMSE (y-axis) by the choice of p (x-axis) over 20 independent replicate data sets. In the right panel we plot FNR. The data is simulated as described in Sect. 3.1

3.2 Simulation results

When computing the AUC when plotting the true positive rate (TPR), which is equal to 1− FNR, against the FPR for $p = 20$ we obtain a value of 98.9%. This result is consistent across replicates. This AUC value shows that our method has a strong ability to do hard-threshold and can select the correct predictive covariates while avoid selecting non-predictive covariates.

In Table 1, we provide the mean RMSE, FPR, and FNR over the 20 replicate datasets. We find that, as p increases, an increasing trend of RMSE is presented. It shows that the estimation performance of our method is related to the dimension of the coefficient vector (i.e., $3p+3r$). A better performance is observed when $3p + 3r$ is smaller than the number of objects n. The increasing trend in FNR along with p suggests that our method performs better in selecting the useful covariates when $3p+3r$ is not larger than n. The FPR, which is almost 0 in each case, indicates that our method have a robust performance in correctly labeling an irrelevant element regardless of the amount of zero elements. Boxplots of the RMSE and FNR in Fig. 2 (FPR is not shown since it's close to 0 in each case) indicate the added variability in RMSE and FNR as p increases and the high performance of our method in terms of these metrics.

In Supplementary Appendix E, we provide a plot of the estimates versus the truth for a single replicate dataset over different choices of p (i.e., $p = 10, 20, 30, 40$). We can notice that our method has a consistent estimation performance for the continuous-valued data regardless of the value of p. For the integer-valued data and count-valued data, the estimated value points below the blue line in the bottom panels indicate that our method tends to underestimate the coefficients. The value points around one suggests that our method has the ability to capture the enforced interaction effect (dependence) across the response-types.

3.3 Comparison with the hierarchical generalized transformation

The most natural comparison would be the horseshoe model. However, the horseshoe model has yet to be developed for multiple response-type data. Other methods, such as generalized linear models for count data in Datta and Dunson (2016), semiparametric linear models where the errors are modeled non-parametrically (Kundu and Dunson 2014), and a Gaussian approximation approach (Piironen and Vehtari 2017) may also be considered for comparison, but again, these methods are not immediately comparable in a multiple response-types setting. Another path to handle multiple response data is to transform the multi-type data so that the transformations can be reasonably modeled using the horseshoe method. A recent fully Bayesian framework allows the transformations of multiple response-type data to be unknown, and the uncertainty in the choice of transformation is accounted for in a fully Bayesian framework. This approach is called the hierarchical generalized transformation (HGT) model (Bradley 2022). The HGT treats posterior replicates of the unknown transformation as data for use in the "preferred model". Thus, instead of comparing our method to the horseshoe model, we compare it to the HGT-horseshoe model to account for multiple response-types.

To implement HGT-horseshoe, we obtain posterior replicates of the transformed multiple response-type data from a latent conjugate multivariate model, which are continuous-valued. Consequently, the continuous transformed data is suitable for the horseshoe model to perform variable selection. Generally, the transformation is treated as unknown and the transformed data is set to over-fit the original data so that it can be viewed as a "proxy" for the original data. See Supplementary Appendix F for a technical review of the HGT model.

We repeat the simulation 10 times (which is large enough to gain significant difference) and provide the boxplots of the RMSE. We conduct this comparison for all members of response-types combining together, as well as each one individually. Figure 3 shows that our method outperforms the HGT-horseshoe model in terms of RMSE for the combination

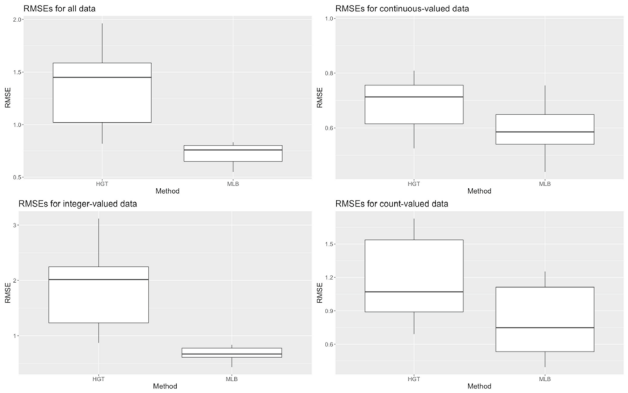

Fig. 3 We plot boxplots of the RMSE by our proposed method (labeled as "MLB") and the competing method HGT-horseshoe (label as "HGT") for comparison over 10 independent replicates of the simulated data. Top left: comparison for the combination of all the data; top right: comparison only for the continuous-valued data; bottom left: comparison only for the integer-valued data; bottom right: comparison only for the count-valued data

of all the data, the integer-valued data, and the count-valued data. A supportive evidence is that the pairwise p-values for paired t-tests are 0.0007, 0.0004, and $3.929e-05$, respectively for each case. For the case of the continuous-valued data, the paired t-test results in a p-value of 0.2471, which suggests that the performance between our method and the competing method in terms of RMSE is not significantly different in this case. This is expected because both methods technically use the same global–local shrinkage framework designed for continuous data.

We also compare the performance of variable selection by both methods using FNR and FPR. The FNR is quite large (between 0.2 and 0.4) over 10 replicates for the HGT-horseshoe model, while our method has smaller FNR. Both approaches give near zero FPR rates over 10 replicate data sets.

4 Data analysis

4.1 The data set: multiple response-type data on natural disasters in the U.S.

The National Centers for Environmental Information (NCEI), part of the National Oceanic and Atmospheric Administration (NOAA), is an important governmental program that is responsible for preserving, monitoring, assessing, and providing public access to severe weather and climate events both in the United States and internationally with their historical perspective. In particular, NCEI tracks and evaluates climate events in the U.S. and globally that have great economic and societal impacts. The U.S. Billion-dollar Weather/Climate Disaster report by NCEI (https://www.ncei.

noaa.gov/access/billions/) provides readers with an aggregated loss perspective for weather and climate events with costs equaling or exceeding $1 billion in damages (adjusting for inflation) from 1980 to 2022. This report assesses numerous weather and climate disasters, including tropical cyclones, floods, droughts, severe local storms, wildfires, crop freeze events and winter storms, and estimates the loss reflecting direct effects of the aforementioned weather and climate disasters. While each region of the United States faces a unique combination of weather and climate events, every state in the country has been impacted by at least one billion-dollar disaster since 1980. The U.S. has endured a total of 332 weather and climate disasters since 1980, where the total public costs have exceeded $2.278 trillion and the total associated deaths have exceeded 15, 355. The number and cost of disasters are noticeably increasing over time and that climate change is increasing the frequency of some types of extremes that lead to billion-dollar disasters. This gives motivation to use our method to conduct a data analysis on discovering what features that significantly influence the number of a climate event every year, alongside with the public costs and public health cost in terms of deaths due to the natural disaster in a multiple response-type data setting. Our interests also lie in estimating the dependence across public costs, event counts, and associated deaths due to the most frequent disaster event, local severe storms.

Among natural weather and climate disasters listed in the report, in this real data analysis, we focus on local severe storms (e.g., tornado, hail, straight-line wind damage), which have caused the highest number of billion-dollar disaster events in the U.S. with a percent frequency of 48.2% from 1980 to 2022. We consider using our proposed Bayesian variable selection method for the multiple response-type data associated with the local severe storms. We model the disaster costs with an MLB distribution since it is continuous-valued. The event counts are non-negative and integer-valued, which are bounded by the total disaster counts each year, thus these counts are modeled by a binomial distribution with sample size of the total number of disaster events. Let $\mathbf{Y}_1$ denote the CPI-adjusted costs of local severe storms in billions of US Dollars, $\mathbf{Y}_2$ be the severe storm counts, and $\mathbf{Y}_3$ represent the deaths in which the severe storm resulted. The data on reported deaths are count-valued and are modeled as negative binomial. We analyze the data across the 1981–2020 period and remove the data from 1987 as zero storm event was recorded this year (i.e., $I_1 = I_2 = I_3 = 39$). See Fig. 4 for a plot of public costs, events counts, and deaths due to local severe storms from 1981 to 2020 (excluding the year of 1987).

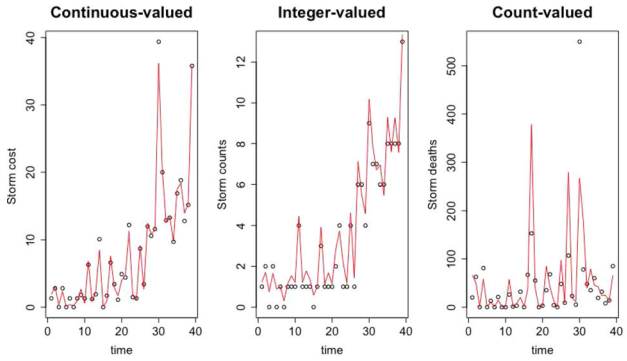

Fig. 4 We plot the posterior means of cost (left), event counts (middle), and associated deaths (right) due to severe storms versus the observed values, respectively, with the use of the Fourier basis functions. The red lines represent the predicted values from our model and the black circles represent the observed values. The title of each panel indicates the response-type of data

4.2 Setup and analysis

We consider the design matrices for each response-type, respectively, to be

$$\begin{aligned}
\mathbf{X}_1^* &= \left[\mathbf{1}\ \mathbf{Y}_2\ \mathbf{G}_{12}^{[1]}\ \mathbf{G}_{13}^{[1]}\ \mathbf{0}_{I_1,r}\right], \\
\mathbf{X}_2^* &= \left[\mathbf{1}\ \mathbf{0}_{I_2,1}\ \mathbf{G}_{12}^{[2]}\ \mathbf{0}_{I_2,r}\ \mathbf{G}_{23}^{[2]}\right], \\
\mathbf{X}_3^* &= \left[\mathbf{1}\ \mathbf{0}_{I_3,1}\ \mathbf{0}_{I_3,r}\ \mathbf{G}_{13}^{[3]}\ \mathbf{G}_{23}^{[3]}\right],
\end{aligned} \tag{19}$$

where we specify the random effects as the Fourier basis functions (Konidaris et al. 2011). The Fourier basis function can be seen as functional covariate as it exploits properties of sine and cosine to recover the amplitude and phase of each sinusoid, which allows one to capture the variability depending on time. The linear dependence among the columns of $(\mathbf{X}_1^{*'}, \mathbf{X}_2^{*'}, \mathbf{X}_3^{*'})'$ can be investigated by conducting a QR decomposition, that is, $\mathbf{X}^* = \mathbf{QR}$, where $\mathbf{X}^* = (\mathbf{X}_1^{*'}, \mathbf{X}_2^{*'}, \mathbf{X}_3^{*'})'$, $\mathbf{Q}$ is an orthogonal matrix, and $\mathbf{R}$ is an upper triangular matrix. In particular, we choose the value of r to be 36 to ensure that the matrix $\mathbf{R}$ from the QR decomposition of $\mathbf{X}^*$ is a full rank matrix. Consequently, the columns in $\mathbf{X}^*$ are linearly independent, which alleviates confounding among the mixed effect coefficients. Severe storm counts $\mathbf{Y}_2$ is introduced as a predictor for severe storm cost $\mathbf{Y}_1$ as there appears that a linear relationship exists between $\mathbf{Y}_1$ and $\mathbf{Y}_2$ (See Sect. 2.5 for a discussion). We additionally include an intercept for each matrix as a fixed effect. The design matrices, except the intercepts, are centered and rescaled to be between 0 and 1. We let $a = b = 0.5$ for the local shrinkage

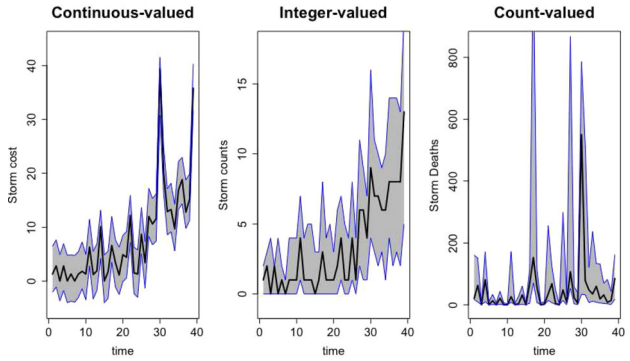

Fig. 5 We plot the 95% point-wise credible predictions of cost (left), event counts (middle), and associated deaths (right) due to severe storms versus the observed values, respectively, with the use of the Fourier basis functions. The black lines represent the observed values. The blue lines are the 95% point-wise credible intervals. The title of each panel indicates the response-type of data

parameter to produce a horseshoe-shaped shrinkage pattern. We choose $a_\tau = b_\tau = a_{\tau_\beta} = b_{\tau_\beta} = 10$ for the global shrinkage parameter. We run our block Gibbs sampler in Appendix A of the Supplementary Materials over 10,000 iterations with a burn-in of 2,000. Standard diagnostics suggests no issues with lack of convergence.

To assess the in-sample error, we denote the posterior mean of predictive replicates with $\hat{\mathbf{Y}}_j$ and calculate the in-sample root mean square prediction error (RMSPE) for each response-type data,

$$\left\{\frac{(\mathbf{Y}_j - \hat{\mathbf{Y}}_j)'(\mathbf{Y}_j - \hat{\mathbf{Y}}_j)}{I_j}\right\}^{1/2}, \tag{20}$$

where $\hat{\mathbf{Y}}_j$ is the posterior mean of the predictive data of a response type and $j = 1, 2, 3$. The RMSPEs for each response-type are as follows: 0.78 (costs), 0.65 (event counts), and 65.90 (associated deaths). The relatively small quantity in the value of each RMSPE, compared to the range of each response-type data (see Fig. 4 for reference), suggests that the in-sample error is small in each response-type data and that our model provides reasonable predictions that are close to the data. In Fig. 4 we plot the posterior means of cost, event counts, and associated deaths due to severe storms versus the observed values, respectively. In Fig. 5 we plot the 95% point-wise credible predictions versus the observed values, respectively. We see that, in Fig. 4, the posterior means are reasonably close to the data, and, in Fig. 5, that the observed data are generally contained within the 95% point-wise intervals. The visual evidences in both plots suggest that the predictions reflect the general patterns of the data and our model has the ability to capture most of the functional variabilities.

Additionally, we use the posterior predictive p-value to access the out-of-sample goodness-of-fit. Specifically, the posterior predictive p-value is defined as

$$P[(T(y_j^{rep}) \geq T(y_j))|H, \theta], \tag{21}$$

where replication y_j^{rep} is the posterior predictive data of a response-type, y_j is the observed data of a response-type, H is the assumed model (i.e., our proposed model), θ is the unknown model parameter and random effects (i.e., $\boldsymbol{\beta}$), and $j = 1, 2, 3$. The test statistics T is chosen to be chi-square statistic, that is,

$$T(y_j) = \frac{\{y_j - E(y_j|y_1, y_2, y_3)\}^2}{E(y_j|y_1, y_2, y_3)}. \tag{22}$$

To interpret the posterior predictive p-values, it is commonly known that p-values being close to zero suggests that the estimates produced by the assumed model are not close to the true data. Conversely, p-values close to one show that the estimates are too close to the true data. Thus, a desired p-value should be close to 0.5. The posterior predictive p-values for each response-type data are as follows: 0.60 (costs), 0.40 (event counts), and 0.42 (associated deaths), which show that our proposed model has a reasonable out-of-sample fit to the data in the sense that the values are subjectively close to 0.5.

We construct joint credible regions associated with the coefficients $\boldsymbol{\beta}$ to determine the significance of elements of $\boldsymbol{\beta}$. Table 2 presents point-wise the credible intervals that do not contain zero for elements of $\boldsymbol{\eta}_{12}$. Similar tables for $\boldsymbol{\eta}_{13}$ and $\boldsymbol{\eta}_{23}$ are given in Appendix E of the Supplementary Materials. For the mixed effect coefficients corresponding to storm costs, we have a 95% credible interval of [8.29, 10.11], which does not contain zero and is thus deemed significant. It is not surprising that the number of occurrences of severe storms has a relationship with the public cost associated directly with the disaster. The significant elements among $\boldsymbol{\eta}_{12}$, $\boldsymbol{\eta}_{13}$,

Table 2 Joint credible regions for elements of $\boldsymbol{\eta}_{12}$ that do not contain zero

Element in $\boldsymbol{\eta}_{12}$	Joint credible region
1	[−1.49 −0.21]
5	[0.37, 1.51]
6	[−1.14, −0.07]
7	[0.29, 1.34]
10	[0.10, 1.23]
16	[−1.42, −0.26]
17	[0.45, 1.57]
18	[0.21, 1.30]
23	[0.03, 1.15]
24	[−1.44, −0.34]
27	[0.32, 1.39]
31	[−1.27, −0.19]

and η_{23} indicate the dependence across public costs, event counts, and associated deaths due to local severe storms, since these coefficients are associated with Fourier basis functions that are shared across responses-types. Recall that the coefficients associated with the basis functions are interpreted as random effects, which is standard in the spatial, time series, and spatio-temporal statistics literature (Cressie and Wikle 2011), and as such, these random effects that appear to be present in our study are interpreted as left-over terms after the means have been modeled via covariates.

5 Discussion

Multiple-type outcomes are often encountered in many statistical applications, one may want to study the association between multiple responses and determine the covariates useful for prediction. However, the literature on variable selection methods for multiple-type data is arguably underdeveloped. In this article, we have proposed a novel Bayesian variable selection approach in a global–local shrinkage prior framework that makes use of the MLB distribution. One benefit of our model is that a transformation or a Gaussian approximation on the data is not needed to perform variable selection for multiple response-type data, and thus one can avoid computational difficulties and restrictions on the joint distribution of the responses. Another benefit is that it allows one to parsimoniously model cross-variable dependence. Specifically, our method uses basis functions and random effects to model dependence between responses and dependence can be detected by our proposed global–local shrinkage model with a sparsity-inducing model. Finally, the use of the MLB distribution aids with a computationally efficient Gibbs sampling algorithm.

In the simulation study, our model provides reasonable estimation and variable selection performance for multiple response-type data. In a comparison with the hierarchical generalized transformation version of the horseshoe method, an approach that incorporates unknown transformations of multiple response-type data to a global–local shrinkage framework, our method outperforms the competing method in terms of overall estimation and selecting the relevant covariates. Finally, we apply our model to study the public health and financial costs of natural disasters dataset provided by NCEI. Using the Fourier basis functions with random effects, we discovered what features are significant for predicting the number of a climate event (i.e., local severe storms) every year, alongside with the public costs and deaths due to the natural disaster in a multiple response-type data setting. With the unique feature of our model, we find there appears to be dependence across response-types. Also, the prediction performance is reasonable in terms of in-sample fit and out-of-sample fit.

As discussed in the introduction, conjugate modeling and variable selection are both extremely active areas in statistics that show no signs of slowing down. The methodological contributions described above may offer strategies for those developing their own methods in these areas. In this new era of "big data" observing multi-type data is becoming more and more common, and as such, this approach offers a tool to those who want to use an existing method and do not want to compromise by performing variable selection for single-type data. To aid those interested in using this methodology, the first author has provided R code at https://github.com/hwuatfsu/MLBpriorsMRT.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-024-10380-1.

Acknowledgements This research was partially supported by the U.S. National Science Foundation (NSF) under NSF Grants SES-1853099 and DMS-2310756. I'd like to thank the editor, associate editor, and reviewers for their helpful comments that improved this paper.

Author Contributions HW wrote the paper and the code; HW ran the numerical experiments; all authors analyzed the results; JRB devised and supervised the study; all authors read, edited, and approved the final manuscript.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Argyriou, A., Evgeniou, T., Pontil, M.: Multi-task feature learning. In: Schölkopf, B., Platt, J., Hoffman, T. (eds.) Advances in Neural Information Processing Systems. MIT Press, Cambridge (2007)

Bhadra, A., Datta, J., Polson, N.G., et al.: Lasso meets horseshoe: a survey. Statist. Sci. **34**(3), 405–427 (2019). https://doi.org/10.1214/19-STS700

Bradley, J.R.: Joint Bayesian analysis of multiple response-types using the hierarchical generalized transformation model. Bayesian Anal. **17**(1), 127–164 (2022). https://doi.org/10.1214/20-BA1246

Bradley, J.R., Cressie, N., Shi, T.: Selection of rank and basis functions in the spatial random effects model. In: Proceedings of the 2011 Joint Statistical Meetings, American Statistical Association Alexandria, VA, pp. 3393–3406. (2011)

Bradley, J.R., Wikle, C.K., Holan, S.H.: Regionalization of multiscale spatial processes by using a criterion for spatial aggregation error. J. R. Stat. Soc. Ser. B Stat. Methodol. **79**(3), 815–832 (2017)

Bradley, J.R., Wikle, C.K., Holan, S.H.: Spatio-temporal models for big multinomial data using the conditional multivariate logit-beta distribution. J. Time Ser. Anal. **40**(3), 363–382 (2019). https://doi.org/10.1111/jtsa.12468

Bradley, J.R., Holan, S.H., Wikle, C.K.: Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. J. Am. Stat. Assoc. **115**(532), 2037–2052 (2020). https://doi.org/10.1080/01621459.2019.1677471

Carvalho, C., Polson, N., Scott, J.: Handling sparsity via the horseshoe. J. Mach. Learn. Res. Proc. Track **5**, 73–80 (2009)
Carvalho, C.M., Polson, N.G., Scott, J.G.: The horseshoe estimator for sparse signals. Biometrika **97**(2), 465–480 (2010). https://doi.org/10.1093/biomet/asq017
Chen, R.B., Chu, C.H., Lai, T.Y., et al.: Stochastic matching pursuit for Bayesian variable selection. Stat. Comput. **21**, 247–259 (2011)
Christensen, W.F., Amemiya, Y.: Latent variable analysis of multivariate spatial data. J. Am. Stat. Assoc. **97**(457), 302–317 (2002). https://doi.org/10.1198/016214502753479437
Cressie, N., Johannesson, G.: Spatial prediction for massive data sets. In: Australian Academy of Science Elizabeth and Frederick White Conference, pp. 1–11. Australian Academy of Science, Canberra (2006)
Cressie, N., Johannesson, G.: Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **70**(1), 209–226 (2008). https://doi.org/10.1111/j.1467-9868.2007.00633.x
Cressie, N., Wikle, C.K.: Statistics for Spatio-Temporal Data. Wiley, Hoboken (2011)
Datta, J., Dunson, D.B.: Bayesian inference on quasi-sparse count data. Biometrika **103**(4), 971–983 (2016). https://doi.org/10.1093/biomet/asw053
Daw, R., Simpson, M., Wikle, C.K., et al.: An overview of univariate and multivariate karhunen loève expansions in statistics. J. Indian Soc. Prob. Stat. **23**(2), 285–326 (2022)
Dobra, A., Lenkoski, A.: Copula Gaussian graphical models and their application to modeling functional disability data. Ann. Appl. Stat. **5**(2A), 969–993 (2011). https://doi.org/10.1214/10-AOAS397
Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994)
Fellinghauer, B., Bühlmann, P., Ryffel, M., et al.: Stable graphical model estimation with random forests for discrete, continuous, and mixed variables. Comput. Stat. Data Anal. **64**, 132–152 (2013). https://doi.org/10.1016/j.csda.2013.02.022
Friedman, J.H.: Multivariate Adaptive Regression Splines. Ann. Stat. **19**(1), 1–67 (1991). https://doi.org/10.1214/aos/1176347963
Gao, H., Bradley, J.R.: Bayesian analysis of areal data with unknown adjacencies using the stochastic edge mixed effects model. Spat. Stat. **31**(100), 357 (2019). https://doi.org/10.1016/j.spasta.2019.100357
Garcia-Donato, G., Martinez-Beneito, M.A.: On sampling strategies in Bayesian variable selection problems with large model spaces. J. Am. Stat. Assoc. **108**(501), 340–352 (2013)
Gelman, A.: Two simple examples for understanding posterior *p*-values whose distributions are far from uniform. Electron. J. Stat. **7**(none), 2595–2602 (2013). https://doi.org/10.1214/13-EJS854
Gelman, A., Carlin, J., Stern, H., et al.: Bayesian Data Analysis. Chapman and Hall (2003). https://doi.org/10.2307/2988417
Greenland, S., Robins, J.M., Pearl, J.: Confounding and collapsibility in causal inference. Stat. Sci. **14**, 29–46 (1999)
Griffin, J.E., Brown, P.J.: Inference with normal-gamma prior distributions in regression problems. Bayesian Anal. **5**(1), 171–188 (2010). https://doi.org/10.1214/10-BA507
Griffin, J.E., Łatuszyński, K., Steel, M.F.: In search of lost mixing time: adaptive Markov chain monte Carlo schemes for Bayesian variable selection with very large p. Biometrika **108**(1), 53–69 (2021)
Griffith, D.A.: A linear regression solution to the spatial autocorrelation problem. J. Geogr. Syst. **2**(2), 141–156 (2000). https://doi.org/10.1007/PL00011451
Griffith, D.A.: A spatial filtering specification for the auto-poisson model. Stat. Prob. Lett. **58**(3), 245–251 (2002). https://doi.org/10.1016/S0167-7152(02)00099-8
Griffith, D.A.: A spatial filtering specification for the autologistic model. Environ. Plann. A Econ. Space **36**(10), 1791–1811 (2004). https://doi.org/10.1068/a36247
Hall, D.B.: Zero-inflated poisson and binomial regression with random effects: a case study. Biometrics **56**(4), 1030–1039 (2000)
Hu, G., Bradley, J.: A Bayesian spatial-temporal model with latent multivariate log-gamma random effects with application to earthquake magnitudes. Statistics **7**(1), e179 (2018). https://doi.org/10.1002/sta4.179
Huang, S.P., Quek, S.T., Phoon, K.K.: Convergence study of the truncated Karhunen-loeve expansion for simulation of stochastic processes. Int. J. Numer. Meth. Eng. **52**(9), 1029–1043 (2001). https://doi.org/10.1002/nme.255
Hughes, J., Haran, M.: Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **75**(1), 139–159 (2013)
Karhunen, K.: Zur spektraltheorie stochastischer prozesse, p. 34. Ann Acad Sci Fennicae, AI (1946)
Kim, S., Xing, E.P.: Statistical estimation of correlated genome associations to a quantitative trait network. PLoS Genet. **5**(8), 1–18 (2009). https://doi.org/10.1371/journal.pgen.1000587
Konidaris, G., Osentoski, S., Thomas, P.: Value function approximation in reinforcement learning using the Fourier basis. Proc. AAAI Conf. Artif. Intell. **25**(1), 380–385 (2011)
Kundu, S., Dunson, D.B.: Bayes variable selection in semiparametric linear models. J. Am. Stat. Assoc. **109**(505), 437–447 (2014). https://doi.org/10.1080/01621459.2014.881153. (**pMID: 25071298**)
Lehmann, C.: Theory of point estimation. Technometrics **41**(3), 274–274 (1998). https://doi.org/10.1080/00401706.1999.10485701
Liu, H., Lafferty, J., Wasserman, L.: The nonparanormal: semiparametric estimation of high dimensional undirected graphs. J. Mach. Learn. Res. **10**(80), 2295–2328 (2009)
Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. **18**(1), 50–60 (1947). https://doi.org/10.1214/aoms/1177730491
McCullagh, P.: Generalized linear models. Routledge (2019)
Meng, X.L.: Posterior predictive *p*-values. Ann. Stat. **22**(3), 1142–1160 (1994)
Obled, C., Creutin, J.: Some developments in the use of empirical orthogonal functions for mapping meteorological fields. J. Appl. Meteorol. Climatol. **25**(9), 1189–1204 (1986)
Piironen, J., Vehtari, A.: Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron. J. Stat. **11**(2), 5018–5051 (2017). https://doi.org/10.1214/17-EJS1337SI
Polson, N.G., Scott, J.G.: Shrink globally, act locally: sparse Bayesian regularization and prediction. Bayesian Stat. **9**(9), 501–538 (2011). https://doi.org/10.1093/acprof:oso/9780199694587.003.0017
Polson, N.G., Scott, J.G.: On the half-cauchy prior for a global scale parameter. Bayesian Anal. **7**(4), 887–902 (2012). https://doi.org/10.1214/12-BA730
Polson, N.G., Scott, J.G., Windle, J.: Bayesian inference for logistic models using Pólya-Gamma latent variables. J. Am. Stat. Assoc. **108**(504), 1339–1349 (2013). https://doi.org/10.1080/01621459.2013.829001
Riesz, F., Nagy, B.S.: Functional analysis. Courier Corporation (2012)
Schliep, E., Hoeting, J.: Multilevel latent gaussian process model for mixed discrete and continuous multivariate response data. J. Agric. Biol. Environ. Stat. (2012). https://doi.org/10.1007/s13253-013-0136
Sellers, K.F.: The Conway-Maxwell-Poisson distribution, vol. 8. Cambridge University Press (2023)
Thuiller, W.: Biodiversity: climate change and the ecologist. Nature **448**, 550–552 (2007). https://doi.org/10.1038/448550a
Wahba, G.: Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, https://books.google.com/books?id=ScRQJEETs0EC (1990)

Wikle, C.K.: Low-rank representations for spatial processes. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., et al. (eds.) Handbook of Spatial Statistics, pp. 107–118. Chapman & Hall/CRC Press, Boca Raton (2010)

Wikle, C.K., Holan, S.H.: Polynomial nonlinear spatio-temporal integro-difference equation models. J. Time Ser. Anal. **32**(4), 339–350 (2011)

Xu, Z., Bradley, J.R., Sinha, D.: Latent multivariate log-gamma models for high-dimensional multitype responses with application to daily fine particulate matter and mortality counts. Ann. Appl. Stat. **17**(2), 1175–1198 (2023)

Xue, L., Zou, H.: Regularized rank-based estimation of high-dimensional nonparanormal graphical models. Ann. Stat. **40**(5), 2541–2571 (2012). https://doi.org/10.1214/12-AOS1041

Yang, E., Ravikumar, P., Allen, G.I., et al.: A general framework for mixed graphical models. arXiv: Statistics Theory (2014)

Yang, X., Kim, S., Xing, E.: Heterogeneous multitask learning with joint sparsity constraints. In: Bengio, Y., Schuurmans, D., Lafferty, J., et al. (eds.) Advances in Neural Information Processing Systems. Curran Associates Inc (2009)