Bayesian Image-on-Scalar Regression with a Spatial Global-Local Spike-and-Slab Prior

Zijian Zeng[*], Meng Li[†] and Marina Vannucci[‡]

Abstract. In this article, we propose a novel spatial global-local spike-and-slab selection prior for image-on-scalar regression. We consider a Bayesian hierarchical Gaussian process model for image smoothing, that uses a flexible Inverse-Wishart process prior to handle within-image dependency, and propose a general global-local spatial selection prior that broadly relates to a rich class of well-studied selection priors. Unlike existing constructions, we achieve simultaneous global (i.e., at covariate-level) and local (i.e., at pixel/voxel-level) selection by introducing *participation rate* parameters that measure the probability for the individual covariates to affect the observed images. This along with a hard-thresholding strategy leads to dependency between selections at the two levels, introduces extra sparsity at the local level, and allows the global selection to be informed by the local selection, all in a model-based manner. We design an efficient Gibbs sampler that allows inference for large image data. We show on simulated data that parameters are interpretable and lead to efficient selection. Finally, we demonstrate performance of the proposed model by using data from the Autism Brain Imaging Data Exchange (ABIDE) study (Di Martino et al., 2014).

Keywords: nonparametric regression, variable selection, spike-and-slab prior, smoothing, mean-covariance estimation.

1 Introduction

With the explosive growth in the amount of image data collected for various medical research there comes an increasing interest in discovering the relation between the image data and potential covariates measured on the same set of subjects. Image-on-scalar regression models have drawn increasing attention for this purpose, see Worsley et al. (2004); Zhu et al. (2014); Li et al. (2020); Zhang et al. (2020); Yu et al. (2021), among others. These models present several challenges: spatial dependency in the image data can be highly complex and hard to model; image data can be composed by a large number of pixels/voxels and lead to extremely large covariance matrices with heavy computational burden; covariates can have partial influence on the image responses, i.e., they can affect only a few pixels/voxels in the image, making it hard to distinguish such covariates from noisy ones.

Conventional approaches for image-on-scalar regression are based on mass univariate analysis (MUA), for example by running pixel/voxel-wise independent general linear models to generate maps for statistics of interest, and then applying methods for post-inference (Worsley et al., 2004; Groppe et al., 2011). These methods are computationally

[*]Department of Statistics, Rice University, Houston, TX, zz57@rice.edu
[†]Department of Statistics, Rice University, Houston, TX, meng@rice.edu
[‡]Department of Statistics, Rice University, Houston, TX, marina@rice.edu

efficient and have well-studied theoretical properties. However, they completely ignore the spatial dependency within the images and are generally not optimal in regards to statistical power (Chumbley and Friston, 2009). To address these shortcomings, recent approaches consider the image data as realizations of functions on a given domain and apply functional data analysis (FDA) methods that use basis expansions and spatially-varying coefficients to account for dependency within and across images (Zhu et al., 2014; Li et al., 2020; Yu et al., 2021). Joint uncertainty quantification for all model parameters, however, is difficult to achieve for these methods in the frequentist literature. Here, we consider a Bayesian hierarchical Gaussian process (GP) model for image smoothing that avoids assumptions on functional forms and that uses a flexible Inverse-Wishart process to handle within-image dependency. This modeling structure extends an approach proposed by Yang et al. (2016) for longitudinal data to the case of image data.

An important aspect in image-on-scalar regression is the selection and interpretation of influential covariates. Ideally, one may want a covariate to be influential for the whole image. In practice, however, the covariate can only partially affect the image, i.e., being influential only for a few pixels/voxels. We refer to the aspect of selecting whether a covariate is influential for the images as "global", and to the aspect of selecting which pixels/voxels are affected by the covariate as "local". In the Bayesian framework, a global selection prior was proposed by Reich et al. (2010) and allows coefficients to be non-zero constant, spatially-varying function, or zero constant. For local selection, a common way of selecting pixels/voxels uses a two-component mixture prior, which models the spatially-varying coefficient via a latent continuous process and a binary selection indicator process, see Smith and Fahrmeir (2007); Scheel et al. (2013); Goldsmith et al. (2014); Li et al. (2015); Choi and Lawson (2018). Recently Kang et al. (2018) proposed a soft-threshold Gaussian process prior that does not make use of the indicator process but rather achieves local sparsity by thresholding. This idea can be traced back to the earlier research of Nakajima and West (2013), who used a hard-threshold prior for longitudinal data to introduce sparsity at each time point. Overall, none of these prior constructions achieve simultaneous global- and local-level selection of the coefficients. Finally, in the more general framework of function-on-scalar regression, in which we consider images as 2/3-dimensional functions, Bayesian approaches employ basis functions and functional principal component analysis to model the within-function dependency, see for example Chen et al. (2016b); Kowal and Bourgeois (2020). Built on the basis functions domain, this framework can be computationally more efficient. However, the covariates' effects are assumed on the basis functions, instead of directly on the observed functions. As a consequence, selection relies on the choice of the basis functions, especially for high-dimensional functional responses, and a two-level selection becomes less intuitive, as the local level selection would require all basis functions to be set to 0 at some specific pixels/voxels.

We propose a spatial global-local spike-and-slab process prior for image-on-scalar regression that broadly relates to a rich class of well-studied local selection priors. We achieve simultaneous global and local selection by introducing *participation rate* parameters, that measure the probability for the individual covariates to affect the observed

images, and employing hard thresholding. The proposed prior performs bi-level selection, allowing the global selection to be informed by the local selection. We design an efficient Gibbs sampler that allows inference for large image data and use simulated data to show that prior parameters are interpretable and lead to efficient selection. We also demonstrate the performance of the proposed model with respect to MUA methods. Finally, we apply our method to data from the Autism Brain Imaging Data Exchange (ABIDE) study (Di Martino et al., 2014). Results show that modeling dependency in the data leads to more localized selection.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed method, the prior construction and the sampler procedure. In Section 3, we conduct simulations and compare the proposed approach with widely used MUA methods. In Section 4, we apply the method to image data from the ABIDE study.

2 Methods

2.1 Bayesian Image-on-Scalar Regression

Suppose n images $Y_i(\cdot)$ are observed on a K-dimensional common domain $\boldsymbol{S} \subseteq \mathbb{R}^K$, each associated with a q-dimensional covariate vector $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iq})^T$, for $i = 1, \ldots, n$. We begin with a Bayesian hierarchical model for image responses and scalar covariates

$$Y_i(\boldsymbol{s}) = Z_i(\boldsymbol{s}) + \epsilon_{i,\boldsymbol{s}}, \quad \epsilon_{i,\boldsymbol{s}} \stackrel{i.i.d}{\sim} N\left(0, \sigma_\epsilon^2\right), \quad \boldsymbol{s} \in \boldsymbol{S} \tag{2.1}$$

$$Z_i(\cdot) \sim \mathcal{GP}\left(\mu_i(\cdot), \Sigma\left(\cdot, \cdot\right)\right), \quad \mu_i(\cdot) = \beta_0(\cdot) + \sum_{j=1}^{q} x_{ij}\beta_j(\cdot), \quad i = 1, \ldots n, \tag{2.2}$$

where the noise-free mean surface of $Y_i(\cdot)$, $Z_i(\cdot)$, is modeled by a Gaussian process (GP) with covariate-dependent, subject-specific mean $\mu_i(\cdot)$ and a common covariance surface $\Sigma(\cdot, \cdot)$ for all images, $\beta_0(\cdot)$ is the intercept coefficient image and $\{\beta_j(\cdot)\}_{j=1}^{q}$ are the coefficient images linking covariates $\boldsymbol{x}_i$ with $\mu_i(\cdot)$. Here we assume Gaussian errors $\epsilon_{i,\boldsymbol{s}} \sim N\left(0, \sigma_\epsilon^2\right)$ independently across both location $\boldsymbol{s}$ and subject i, with $\sigma_\epsilon^2 \sim$ Inverse-Gamma(a_ϵ, b_ϵ). Throughout this article, we assume $\boldsymbol{S}$ to be a compact set.

Equations (2.1)-(2.2) define a Bayesian hierarchical model for image data. This model has considerable virtues: it enables simultaneous smoothing of individual observations and borrowing of information across observations, while being flexible through Bayes nonparametrics and interpretable. It has been widely used to analyze functional data. For example, in the case of longitudinal data (i.e., $K = 1$), Yang et al. (2016) focus on a mean-covariance structure in the absence of covariates, with a common mean function, Liu et al. (2020) study conditional quantiles of $Y_i(\cdot)$ by altering the Gaussian error to asymmetric Laplace, and Shamshoian et al. (2020) consider a sparse Bayesian infinite factor model for $Z_i(\cdot)$.

2.2   Spatial Global-Local Spike-and-Slab Prior

We are interested in performing selection at both image and location (pixel/voxel) levels, while estimating the model coefficients and the covariance structure. We achieve this via a novel spatial global-local spike-and-slab (SGLSS) prior for the coefficient images $\{\beta_j(\cdot)\}_{j=1}^{q}$. Here, selection at the global level represents the covariate selection, as eliminating a covariate would zero out the entire coefficient image, while the local level refers to individual locations (pixels/voxels) in the coefficient image. The proposed SGLSS prior is a three-component process consisting of a continuous process, a local-level discrete selection process, and a global-level indicator:

$$
\begin{aligned}
\beta_j(\cdot) &= \tilde{\beta}_j(\cdot) \times \tau_j(\cdot) \times I\left(\pi_j \geq d\right), \\
&= \left[\tau_j(\cdot)I\left(\pi_j \geq d\right)\right]\tilde{\beta}_j(\cdot) + \left[1 - \tau_j(\cdot)I\left(\pi_j \geq d\right)\right]\delta_0,
\end{aligned}
\tag{2.3}
$$

with δ_0 a point mass distribution at 0, $\tilde{\beta}_j(\cdot)$ the continuous process, $\tau_j(\cdot)$ the discrete local selection process and $I(\pi_j \geq d)$ the global indicator. Construction (2.3) is completed by choosing priors for $\tilde{\beta}_j(\cdot)$, $\tau_j(\cdot)$, and π_j. The high dimensionality of image data poses substantial challenges to computational efficiency, particularly for Markov Chain Monte Carlo methods. To achieve computational scalability, we adopt the following prior setting that exploits conjugacy:

$$
\begin{aligned}
\tilde{\beta}_j(\boldsymbol{s})|\tau_j(\boldsymbol{s}) &\sim \tau_j(\boldsymbol{s})N\left(\mu_{0j}(\boldsymbol{s}), \sigma_{0j}^2(\boldsymbol{s})\right) + (1 - \tau_j(\boldsymbol{s}))\delta_0, \quad \boldsymbol{s} \in \boldsymbol{S} \\
\tau_j(\boldsymbol{s})|\pi_j &\sim \text{Bernoulli}\left(\pi_j\right), \\
\pi_j &\sim \text{Beta}(a_\pi, b_\pi).
\end{aligned}
\tag{2.4}
$$

The SGLSS prior for $\beta_j(\cdot)$ with the specification of Equation (2.4) can be re-written into a classic spike-and-slab prior as

$$
\begin{aligned}
\beta_j(\boldsymbol{s})|\tau_j(\boldsymbol{s}), \pi_j &\sim \left[\tau_j(\boldsymbol{s})I\left(\pi_j \geq d\right)\right]N\left(\mu_{0j}(\boldsymbol{s}), \sigma_{0j}^2(\boldsymbol{s})\right) + \left[1 - \tau_j(\boldsymbol{s})I\left(\pi_j \geq d\right)\right]\delta_0, \\
\tau_j(\boldsymbol{s})|\pi_j &\sim \text{Bernoulli}\left(\pi_j\right), \\
\pi_j &\sim \text{Beta}(a_\pi, b_\pi).
\end{aligned}
$$

As for the intercept image $\beta_0(\boldsymbol{s})$, we fix $\tau_0(\boldsymbol{s})I(\pi_0 \geq d) = 1$ for all $\boldsymbol{s} \in \boldsymbol{S}$ since this term is typically always included in the model. The proposed SGLSS prior construction has rich connections with a wide range of existing priors, as we point out in the section below.

The parameter π_j, which we call *'participation rate'*, has the interpretation that π_j percent of the jth coefficient image is expected to be non-zero, and can also be interpreted as the probability that x_j has an influence on the observed images. The parameter d defines the threshold at which we include covariate x_j, i.e., if $d \times 100$ percent of its corresponding coefficient images are expected to be non-zero. Therefore, the threshold parameter d controls selection at the "global" level, leading to the exclusion of those covariates with low participation rates. Here, without a priori information, we use a common d for all covariates. When a priori information is available, covariate-dependent parameters d_j's can be specified. The parameter π_j, along with the hard-thresholding structure, introduces extra sparsity at local level, as it can be seen by

calculating the expectation of the global-local selection indicator

$$
\begin{aligned}
E\left[I(\pi_j \geq d)\tau_j(\boldsymbol{s})\right] &= E_{\pi_j}\left[E\left[I(\pi_j \geq d)\tau_j(\boldsymbol{s})\right]|\pi_j\right] = E_{\pi_j}\left[I(\pi_j \geq d)\pi_j\right] \\
&= \int_d^1 \pi_j \frac{1}{B(a_\pi, b_\pi)} \pi_j^{a_\pi - 1}(1 - \pi_j)^{b_\pi - 1} d\pi_j \\
&= \frac{a_\pi}{a_\pi + b_\pi}\left[1 - F_{\text{Beta}}(d)\right] = E\left[\tau_j(\boldsymbol{s})\right]\left[1 - F_{\text{Beta}}(d)\right],
\end{aligned}
\tag{2.5}
$$

where $F_{\text{Beta}}(\cdot)$ is the cumulative distribution function of $\text{Beta}(a_\pi + 1, b_\pi)$. Hence, the extra factor $[1 - F_{\text{Beta}}(d)] \leq 1$ in Equation (2.5), which is strictly decreasing in d, introduces more sparsity in the coefficient images on average from a prior perspective. The participation rate parameter π_j and the threshold parameter $d \in [0, 1]$ establish a bridge between global and local level selection, favorably endowing existing local level selection prior with simultaneous two-level selection, allowing global selection to be *informed* by the selection at the local level.

In practice, image data are typically observed at a grid of discretized locations. Given a vector of p locations of interest $\vec{\boldsymbol{s}} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p)$, we use $Z_i(\vec{\boldsymbol{s}}) = (Z_i(\boldsymbol{s}_1), \ldots, Z_i(\boldsymbol{s}_p))$ to denote a p-by-1 vector of the process values at location $\boldsymbol{s} \in \vec{\boldsymbol{s}}$ and $\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}}) = \{\Sigma(\boldsymbol{s}, \boldsymbol{s}'); \boldsymbol{s}, \boldsymbol{s}' \in \vec{\boldsymbol{s}}\}$ a p-by-p matrix of the within-image covariance. We remark that the proposed method is applicable to discretized locations in a general K-dimensional domain, i.e., the vector $\vec{\boldsymbol{s}}$ is not limited to integers nor needs to be equidistant.

Relationship to Existing Literature

The proposed SGLSS process prior broadly relates to a wide range of existing priors that can be obtained for different choices of the parameter d and prior choices for $\tilde{\beta}_j(\cdot)$ and $\tau_j(\cdot)$. In particular, for $d = 1$ the use of continuous priors on π_j leads to $I(\pi_j \geq 1) = 0$ with probability 1, thus no covariate will be included in the model almost surely. This degenerate SGLSS process prior results in a high-dimensional extension of the mean-covariance smoothing model of Yang et al. (2016) for the following choice of prior

$$
\beta_0(\cdot) = \tilde{\beta}_0(\cdot) \sim \mathcal{GP}\left(\mu_0(\cdot), \frac{1}{c}\Sigma(\cdot, \cdot)\right),
$$

extending the mean-covariance smoothing structure from one-dimensional time-series to high-dimensional images.

When $d = 0$ is specified, the three-component mixture prior in Equation (2.3) degenerates to a two-component mixture of the type

$$
\beta_j(\cdot) = \tilde{\beta}_j(\cdot) \times \tau_j(\cdot) \times 1.
$$

This degenerate construction naturally relates to the prior constructions used in scalar-on-image regression, and easily accommodates spatially dependent priors on $\tilde{\beta}_j(\cdot)$ and $\tau_j(\cdot)$ (Smith and Fahrmeir, 2007; Scheel et al., 2013; Goldsmith et al., 2014; Li et al., 2015; Choi and Lawson, 2018). We note, however, that incorporating spatially-correlated priors for the regression coefficients within our general global-local construction poses

substantial computational challenges and a more careful interpretation of the prior parameters. See also the Conclusion section. Thresholding priors can also be accommodating. For example, by using an autoregressive process for $\tilde{\beta}_j(\cdot)$ and setting the hard threshold, $\tau_j(\cdot) = I(\tilde{\beta}_j(\cdot) > d_j)$, the prior in Nakajima and West (2013) can be obtained, and by setting $\tilde{\beta}_j(\cdot) = \text{sgn}\,(z_j(\cdot))\,(|z_j(\cdot)| - \lambda_j)$, with threshold $\tau_j(\cdot) = I\,(|z_j(\cdot)| > \lambda_j)$ and $z_j(\cdot) \sim \mathcal{GP}$, the prior in Kang et al. (2018) can be obtained. Another partially reproducible prior is the global level selection prior of Reich et al. (2010),

$$\beta_j(\cdot) = \gamma_{1j}\,(\beta_0 + \gamma_{2j} z_j(\cdot))\,, \tag{2.6}$$

that sets $\beta_j(\cdot)$ to a constant non-zero coefficient, a spatially varying process $(\beta_0 + z_j(\cdot))$ or a zero image. Since the SGLSS prior does not distinguish various types of included coefficient images, such as a non-zero constant versus a spatially varying process, it cannot fully recover the prior of Reich et al. (2010). However, it can distinguish zero images, when $\pi_j = 0$ with all indicators $\tau_j(\vec{s}) = 0$, and full images, when $\pi_j = 1$ with all indicators $\tau_j(\vec{s}) = 1$.

We also mention the closely related bi-level selection priors for covariates with group structure, see Stingo et al. (2011); Xu and Ghosh (2015); Chen et al. (2016a); Liquet et al. (2017). These priors also deal with a two-level selection, but the global level is a group of covariates and the local level is the single covariate in the group. For example, Stingo et al. (2011) proposed a linear regression model for identifying pathways, i.e, groups of genes, related to a particular phenotype, and a two-layer selection of pathways and genes. The two selection priors, however, are not linked and this leads to the necessity of constraining the prior set of possible configurations, to avoid selection of an empty group. Chen et al. (2016a) addressed this issue by adding an indicator correction procedure to do post-inference for group selection. Xu and Ghosh (2015) and Liquet et al. (2017) followed the idea of Bayesian group lasso and conducted inference based on posterior median estimators of coefficients. Unlike these constructions, the proposed SGLSS prior construction naturally leads to dependency between selections at the global and local levels, via the participation rate parameter, π, therefore preventing the selection of empty "groups", i.e., covariates with no effects on the images. Furthermore, with the bi-level selection priors, a group is selected in the regression if at least one of its members has non-zero effect, while in the proposed SGLSS prior construction the selection at global level is based on the probability of a covariate to affect the image, as measured by the participation rate.

2.3 Inverse-Wishart Process Prior

For the covariance surface $\Sigma(\cdot, \cdot)$, pre-specified parametric kernels such as the Matérn or squared exponential kernel lack flexibility, and the possible misspecification may introduce considerable bias that hampers inference. We employ a flexible process, called the *Inverse-Wishart process* (IWP), to mitigate this concern. As a nonparametric generalization of the finite-dimensional Inverse-Wishart (IW) distribution, the IWP has been used in time series to capture time-varying *volatility* and *co-volatility* (Philipov and Glickman, 2006; Gouriéroux et al., 2009; Wilson and Ghahramani, 2011; Heaukulani

and van der Wilk, 2019), and as a flexible prior for covariance kernels in functional data analysis (Yang et al., 2016). Existing literature, such as Zhu et al. (2016); Yang et al. (2016), often uses a one-dimensional support; we instead define an IWP for a general K-dimensional index set $\boldsymbol{S}$ in the following sense.

Definition 1. An Inverse-Wishart process is a stochastic process $\Sigma = (\Sigma(\boldsymbol{s}, \boldsymbol{s}') : (\boldsymbol{s}, \boldsymbol{s}') \in \boldsymbol{S} \times \boldsymbol{S})$ indexed by $\boldsymbol{S} \times \boldsymbol{S}$ such that the random matrix $\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}}) = ((\Sigma(\boldsymbol{s}_i, \boldsymbol{s}_j)))_{i,j}$ possesses an Inverse-Wishart distribution for any $\vec{\boldsymbol{s}} = (\boldsymbol{s}_1, \dots, \boldsymbol{s}_p)$ and $p \in \mathbb{N}$ with $\boldsymbol{s}_i \in \boldsymbol{S}$ for $i, j = 1, \dots, p$.

The matrix-valued $\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})$ in the definition are finite-dimensional marginals evaluated on $\vec{\boldsymbol{s}}$. An IW distribution is determined by two parameters: the degrees of freedom and a scale matrix that is symmetric and positive semi-definite. However, we shall follow the parameterization in Dawid (1981), denoted by $\text{IW}(\delta, \Psi(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}}))$ with δ a positive integer and scale matrix $\Psi(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})$. This parameterization guarantees a crucial consistency property of IW after marginalization. Let $\Psi : \boldsymbol{S} \times \boldsymbol{S} \to \mathbb{R}$ be a symmetric and positive semi-definite mapping, i.e., the matrix $\Psi(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}}) = ((\Psi(\boldsymbol{s}_i, \boldsymbol{s}_j)))_{i,j=1,\dots,p}$ is symmetric and positive semi-definite for any $\vec{\boldsymbol{s}} = (\boldsymbol{s}_1, \dots, \boldsymbol{s}_p)$. By the Kolmogorov extension theorem, there exists an IWP for integer $\delta > 0$ and Ψ, which we denote by $\text{IWP}(\delta, \Psi(\cdot, \cdot))$; see Lemma 2 in the Appendix of Zhu et al. (2016) for an elaborate proof when the index set is $\mathbb{N}$ and Proposition 1 in Yang et al. (2016) for a related discussion. We typically choose $\delta > 4$ to ensure marginals of an IWP have finite second moments.

We put this IWP prior on $\Sigma(\cdot, \cdot)$,

$$\Sigma(\cdot, \cdot) \sim \text{IWP}(\delta, \Psi(\cdot, \cdot)), \tag{2.7}$$

and choose the Matérn covariance function for $\Psi(\cdot, \cdot)$,

$$\begin{aligned} \Psi(\boldsymbol{s}, \boldsymbol{s}') &= \text{Matérn}\left(||\boldsymbol{s} - \boldsymbol{s}'||_{l_2}; \sigma_s^2, \rho, \nu\right), \quad \boldsymbol{s}, \boldsymbol{s}' \in \boldsymbol{S} \\ &= \frac{\sigma_s^2}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu}\frac{||\boldsymbol{s} - \boldsymbol{s}'||_{l_2}}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{||\boldsymbol{s} - \boldsymbol{s}'||_{l_2}}{\rho}\right), \end{aligned} \tag{2.8}$$

where $||\cdot||_{l_2}$ is the l_2 norm. For a given vector of locations $\vec{\boldsymbol{s}}$, the prior leads to

$$\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}}) \sim \text{IW}(\delta, \Psi(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})). \tag{2.9}$$

In the applications reported below, we fix $\nu = 5/2$, following Yang et al. (2016), to have the analytical forms for both the Matérn kernel and its gradients, facilitating computation for large covariance matrices in image data, and choose the other two hyperparameters (σ_s^2, ρ) by minimizing the mean square error between an empirical covariance estimate, obtained as the MUA estimate, and the Matérn $(\sigma_s^2, \rho, 5/2)$ kernel.

2.4 Posterior Inference

We derive an efficient Gibbs sampler for the proposed hierarchical model with SGLSS prior. Posterior sampling proceeds in three main steps as follows, with detailed derivations provided in the Supplement (Zeng et al., 2022).

- **Update the BHM parameters** $\{Z_i(\vec{s})\}_{i=1}^n$ **and** σ_ϵ^2 **conditional on** $\{\beta_j(\vec{s})\}_{j=1}^q$ **and** $\Sigma(\vec{s}, \vec{s})$:

  Evaluated on locations $\vec{s}$, Equations (2.1) and (2.2) yield

  $$Y_i(\vec{s})|Z_i(\vec{s}), \sigma_\epsilon^2 \sim \text{MVN}(Z_i(\vec{s}), \sigma_\epsilon^2 I_p), \quad \sigma_\epsilon^2 \sim \text{Inverse-Gamma}(a_\epsilon, b_\epsilon),$$

  $$Z_i(\vec{s})|\{\beta_j(\vec{s})\}_{j=1}^q, \Sigma(\vec{s}, \vec{s}) \sim \text{MVN}(\mu_i(\vec{s}), \Sigma(\vec{s}, \vec{s})), \quad \mu_i(\vec{s}) = \beta_0(\vec{s}) + \sum_{j=1}^q x_{ij}\beta_j(\vec{s}).$$

  In view of conjugacy, we sample

  $$Z_i(\vec{s})|Y_i(\vec{s}), \{\beta_j(\vec{s})\}_{j=1}^q, \Sigma(\vec{s}, \vec{s}), \sigma_\epsilon^2 \sim MVN(\mu_{Z_i}, V_{Z_i}),$$

  with $V_{Z_i} = \left(\sigma_\epsilon^{-2} I_p + \Sigma(\vec{s}, \vec{s})^{-1}\right)^{-1}$ and $\mu_{Z_i} = V_{Z_i}\left(\sigma_\epsilon^{-2} Y_i(\vec{s}) + \Sigma(\vec{s}, \vec{s})^{-1}\mu_i(\vec{s})\right)$ independently for $i = 1, \ldots, n$, and

  $$\sigma_\epsilon^2|\{Y_i(\vec{s})\}_{i=1}^n, \{Z_i(\vec{s})\}_{i=1}^n \sim$$
  $$\text{Inverse-Gamma}\left(a_\epsilon + \frac{np}{2}, b_\epsilon + \frac{1}{2}\sum_{i=1}^n \left(Y_i(\vec{s}) - Z_i(\vec{s})\right)^T \left(Y_i(\vec{s}) - Z_i(\vec{s})\right)\right).$$

- **Update the SGLSS prior parameters** $\{\beta_j(\vec{s}), \tau_j(\vec{s}), \pi_j\}_{j=0}^q$ **conditional on** $\{Z_i(\vec{s})\}_{i=1}^n$ **and** $\Sigma(\vec{s}, \vec{s})$:

  In this step, we first sample the indicators $\{\tau_j(\vec{s})\}_{j=1}^q$ and update $\{\pi_j\}_{j=1}^q$ to obtain selection indicators at both global and local levels. This is achieved via a blocked Gibbs strategy as in Reich et al. (2010) and location-wise Gibbs updates similar to Brown et al. (1998) and Smith and Fahrmeir (2007), adapted to the SGLSS prior.

  We use a blocked Gibbs sampler with respect to each feature $j \geq 1$. Denote $\tilde{Z}_{ij}(\vec{s}) = Z_i(\vec{s}) - \sum_{j' \neq j} x_{ij'}\beta_{j'}(\vec{s})$. Equations (2.2) and (2.4) lead to a location-wise model,

  $$\tilde{Z}_{ij}(s)|\tilde{\beta}_j(s), \tau_j(s) = 1, \Sigma(s, s) \sim N(x_{ij}\tilde{\beta}_j(s), \Sigma(s, s)),$$
  $$\tilde{\beta}_j(s)|\tau_j(s) = 1 \sim N(\mu_{0j}(s), \sigma_{0j}^2(s)), \tag{2.10}$$
  $$\tilde{Z}_{ij}(s)|\tau_j(s) = 0, \Sigma(s, s) \sim N(0, \Sigma(s, s)).$$

  The location-wise Bayes factor can be obtained by integrating out $\tilde{\beta}_j(s)$,

  $$\theta_j(s)$$
  $$= \frac{\prod_{i=1}^n p\left(\tilde{Z}_{ij}(s)|\tau_j(s) = 0, \Sigma(s, s), \pi_j\right) p(\tau_j(s) = 0|\pi_j)}{\left\{\int \prod_{i=1}^n p\left(\tilde{Z}_{ij}(s)|\tilde{\beta}_j(s), \tau_j(s) = 1, \Sigma(s, s), \pi_j\right) p(\tilde{\beta}_j(s))d\tilde{\beta}_j(s)\right\} p(\tau_j(s) = 1|\pi_j)}$$
  $$= \frac{1 - \pi_j}{\pi_j \times \left(\sigma_{0j}^2(s)\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mu_{0j}^2(s)/\sigma_{0j}^2(s)\right)\right\} \times (\tilde{\nu}_j(s))^{\frac{1}{2}} \exp\left\{\frac{1}{2}\tilde{m}_j^2(s)\tilde{\nu}_j(s)\right\}},$$
  $$\tag{2.11}$$

with

$$\tilde{\nu}_j(\boldsymbol{s}) = \left[\sum_{i=1}^{n} x_{ij}^2/\Sigma(\boldsymbol{s}, \boldsymbol{s}) + 1/\sigma_{0j}^2(\boldsymbol{s})\right]^{-1},$$

$$\tilde{m}_j(\boldsymbol{s}) = \sum_{i=1}^{n} x_{ij}\tilde{Z}_{ij}(\boldsymbol{s})/\Sigma(\boldsymbol{s}, \boldsymbol{s}) + \mu_{0j}(\boldsymbol{s})/\sigma_{0j}^2(\boldsymbol{s}).$$

This Bayes factor allows us to sample local selection indicators $\tau_j(\vec{\boldsymbol{s}})$ and participation rates π_j from the conditional posterior distributions

$$\tau_j(\boldsymbol{s})|\{\boldsymbol{\beta}_{j'}(\vec{\boldsymbol{s}})\}_{j' \neq j}, \{Z_i(\vec{\boldsymbol{s}})\}_{i=1}^n, \Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}}), \pi_j \sim \text{Bernoulli}\left(\frac{1}{1 + \theta_j(\boldsymbol{s})}\right),$$

$$\pi_j|\tau_j(\vec{\boldsymbol{s}}) \sim \text{Beta}\left(a_{\pi_j} + \sum_{\boldsymbol{s} \in \vec{\boldsymbol{s}}} \tau_j(\boldsymbol{s}), b_{\pi_j} + p - \sum_{\boldsymbol{s} \in \vec{\boldsymbol{s}}} \tau_j(\boldsymbol{s})\right), \tag{2.12}$$

which also gives samples of global selection indicators $I(\pi_j \geq d)$. When $j = 0$, both $\tau_0(\vec{\boldsymbol{s}})$ and π_j are fixed at 1.

Conditional on selection indicators at the two levels, we sample the coefficient image $\beta_j(\vec{\boldsymbol{s}})$ for $j \geq 0$ as follows. If $\tau_j(\boldsymbol{s}) \times I(\pi_j \geq d) = 0$, we set $\beta_j(\boldsymbol{s}) = 0$; otherwise, using Equation (2.10) we sample $\tilde{\beta}_j(\boldsymbol{s})$ from

$$\tilde{\beta}_j(\boldsymbol{s})|\{\boldsymbol{\beta}_{j'}(\vec{\boldsymbol{s}})\}_{j' \neq j}, \{Z_i(\vec{\boldsymbol{s}})\}_{i=1}^n, \Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}}) \sim N(\tilde{\nu}_j(\boldsymbol{s})\tilde{m}_j(\boldsymbol{s}), \tilde{\nu}_j(\boldsymbol{s})),$$

and set $\beta_j(\boldsymbol{s}) = \tilde{\beta}_j(\boldsymbol{s})$. Note that the covariate $x_{i0} = 1$ when sampling the intercept $\beta_0(\vec{\boldsymbol{s}})$. The joint update of coefficient images and selection indicators avoids reversible jump (Savitsky et al., 2011).

- **Update the IWP prior parameter $\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})$ conditional on $\{Z_i(\vec{\boldsymbol{s}})\}_{i=1}^n$ and $\{\beta_j(\vec{\boldsymbol{s}})\}_{j=0}^q$:**

Equations (2.2) and (2.9) lead to the conditional posterior distribution,

$$\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})|\{Z_i(\vec{\boldsymbol{s}})\}_{i=1}^n, \{\beta_j(\vec{\boldsymbol{s}})\}_{j=0}^q \sim$$
$$\text{IW}\left(n + \delta, \sum_{i=1}^n (Z_i(\vec{\boldsymbol{s}}) - \mu_i(\vec{\boldsymbol{s}}))(Z_i(\vec{\boldsymbol{s}}) - \mu_i(\vec{\boldsymbol{s}}))^T + \Psi(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})\right),$$

where $\mu_i(\vec{\boldsymbol{s}}) = \beta_0(\vec{\boldsymbol{s}}) + \sum_{j=1}^q x_{ij}\beta_j(\vec{\boldsymbol{s}})$.

At each iteration of the MCMC algorithm, for each covariate, the local level informs the selection at the global level, as the local selection indicator is sampled first and used to calculate the participation rate. The covariate is then selected if the participation rate is greater than d. Meanwhile, at the next iteration, the participation rate serves as the prior in the Binomial-Beta conjugate update of the local level indicators, providing feedback from the local level at the previous iteration. At convergence, global-level selection is done by calculating the marginal posterior probabilities of inclusion (MPPIs)

of $I(\pi_j \geq d)$. Following Barbieri and Berger (2004), we use the median probability model and include the covariate if MPPI > 0.5, i.e., if more than half of the posterior samples give $I(\pi_j \geq d) = 1$. Similarly, local-level selection for covariate j is determined by thresholding the MPPIs of $I(\pi_j \geq d)\tau_j(\boldsymbol{s})$ at 0.5. Commonly used values can be specified for the sparsity parameter d, such as $d = 0.05$ or $d = 0.1$, to induce a desired level of sparsity. See results from the applications below and the supplementary material. Given the selected covariates and locations, we estimate the corresponding $\beta_j(\boldsymbol{s})$ via posterior means obtained from the MCMC samples. We also estimate $\{Z_i(\vec{\boldsymbol{s}})\}_{i=1}^n$, σ_ϵ^2, and $\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})$ via the posterior means.

We recommend setting the initial value of $\{\beta_j(\vec{\boldsymbol{s}})\}_{j=0}^q$ at $\hat{\boldsymbol{\beta}}_{\mathrm{MUA}}(\vec{\boldsymbol{s}})$, and the initial value of $\Sigma(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})$ at $\Psi(\vec{\boldsymbol{s}}, \vec{\boldsymbol{s}})$. Aside from the initial values, the sampler needs very little tuning. We provide a python implementation where we have optimized the sampling of $Z_i(\cdot)$ and $\Sigma(\cdot, \cdot)$ by utilizing various matrix decompositions to avoid redundant matrix inversions through equivalent formulations, and further building upon pytorch, which allows automated efficient large matrix operations. We remark here that the independent prior specified in Equation (2.4) allows a parallel update for the $(\beta_j(\boldsymbol{s}), \tau_j(\boldsymbol{s}))$ parameters at each local pixel/voxel $\boldsymbol{s} \in \boldsymbol{S}$ in each iteration of the MCMC, therefore reducing the computational cost from a multivariate Gaussian, roughly $O(|\boldsymbol{S}|^3)$ for the inverse of covariance matrix, to $|\boldsymbol{S}|$ univariate Gaussian.

3   Simulation Study

In this section we conduct simulations to assess the performances of the proposed method, which we call BHM, and perform comparisons with alternative approaches.

We set the number of covariates to $q = 15$ and the sample size to $n = 100$ and generate image data (i.e., $K = 2$) from model (2.1)-(2.2) using a 30-by-30 grid, i.e., $p = 900$. We sample the coefficient images $\{\beta_j(\vec{\boldsymbol{s}})\}_{j=1}^{15}$ and the intercept $\beta_0(\vec{\boldsymbol{s}})$ similarly to Li et al. (2020). At covariate level, we induce sparsity by sampling $\tilde{\beta}_j(\vec{\boldsymbol{s}})$ for $j \in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ from a $\mathcal{GP}(\boldsymbol{0}, \Sigma_\beta)$, with Σ_β specified by the Matérn kernel, and setting the remaining $\beta_j(\boldsymbol{s}) = \tilde{\beta}_j(\boldsymbol{s}) = 0, \forall \boldsymbol{s} \in \vec{\boldsymbol{s}}, j \in \{9, 10, 11, 12, 13, 14, 15\}$. At location level, we first rescale the images as

$$\beta_j(\vec{\boldsymbol{s}}) = \frac{\tilde{\beta}_j(\vec{\boldsymbol{s}}) + \mathrm{sign}(\tilde{\beta}_j(\boldsymbol{s}')) \left|\tilde{\beta}_j(\boldsymbol{s}')\right|}{2\left|\tilde{\beta}_j(\boldsymbol{s}')\right|}, \quad \boldsymbol{s}' = \operatorname*{argmax}_{s \in \vec{\boldsymbol{s}}} \left|\tilde{\beta}_j(\boldsymbol{s})\right|,$$

which excludes zeros introduced by randomness and then consider two different scenarios to introduce sparsity. In the first scenario we set 10% randomly chosen elements of $\beta_2(\vec{\boldsymbol{s}})$ and $\beta_7(\vec{\boldsymbol{s}})$ to zero, 20% randomly chosen elements of $\beta_3(\vec{\boldsymbol{s}})$ and $\beta_8(\vec{\boldsymbol{s}})$ to zero, 30% randomly chosen elements of $\beta_4(\vec{\boldsymbol{s}})$ to zero and 40% randomly chosen elements of $\beta_5(\vec{\boldsymbol{s}})$. The second scenario addresses a more realistic and challenging case, in which signals are clustered and influential covariates only affect a small portion of the images, hardly distinguishable from the noise covariates. In this case, we randomly select a square with π percent pixels/voxels being non-zero. We consider two settings, $\pi_j \approx 10\%$ and $\pi_j \approx 20\%$ ($j = \{1, 2, 3, 4, 5, 6, 7, 8\}$). Figures 1, 2 and 3 show some example images of both noise-free images and coefficient images generated from the three scenarios.

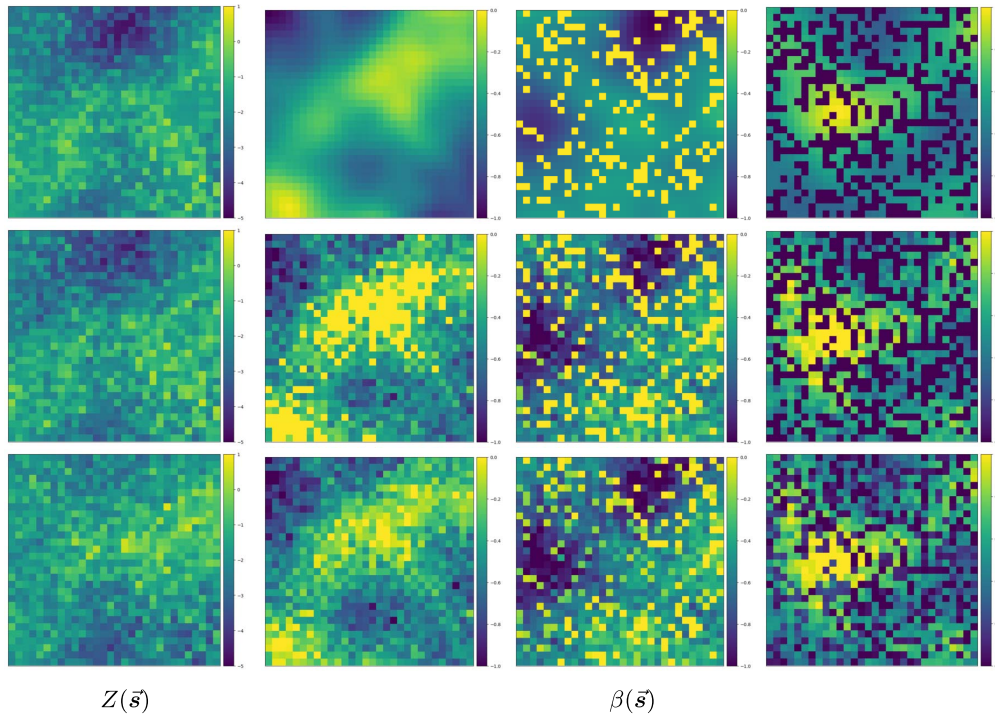$Z(\vec{s})$ $\beta(\vec{s})$

Figure 1: First simulated scenario: Example images of $Z(\vec{s})$ and $\beta(\vec{s})$, where 1st row is generated data; 2nd row is the estimates of BHM; and 3rd row is the estimates of MUA.

Next, we generate the covariates $x_{i,j}$'s, including both continuous and discrete variables. We generate $x_{i,j}$ with $j = 1, 2, 3, 4, 5$ from $N(0,1)$, to obtain continuous features, and $x_{i,j}$, with $j = 6, 7, 8$, from Bernoulli (0.5), for the discrete features. We add noisy features generated from $N(0,1)$, for $j = 9, \ldots, 15$. Finally, we sample the noise-free mean surface $Z_i(\vec{s})$ from Equation (2.2) using a Matérn kernel for the covariance matrix Σ, and the image data $Y_i(\vec{s})$ from Equation (2.1) with $\epsilon_{i,s}$ sampled from $N(0,1)$ for $i = 1, \ldots, n$. Below we report results using Matérn kernels of the type $\Sigma = \Sigma_\beta =$ Matérn $(1, 1/4, 5/2)$.

3.1 Prior Specification

For the prior specification, we use a weakly-informative prior on the participation rate parameters $\{\pi_j\}_{j=1}^q$ in Equation (2.4), by setting $a_\pi = b_\pi = 1$. For the threshold d, we report the results for a conventional sparsity level, $d = 0.05$, and then discuss sensitivity in the supplementary material. We center the slab distribution in Equation (2.4) at $\{\mu_{0j}(\vec{s})\}_{j=0}^q = 0$, as commonly done with spike-and-slab priors, and set $\{\sigma_{0j}^2(\vec{s})\}_{j=0}^q = 1$ (see supplementary material for a sensitivity analysis). As previously discussed, we derive empirical estimates of the Matérn parameters (σ_s^2, ρ) in Equation (2.8) by mini-
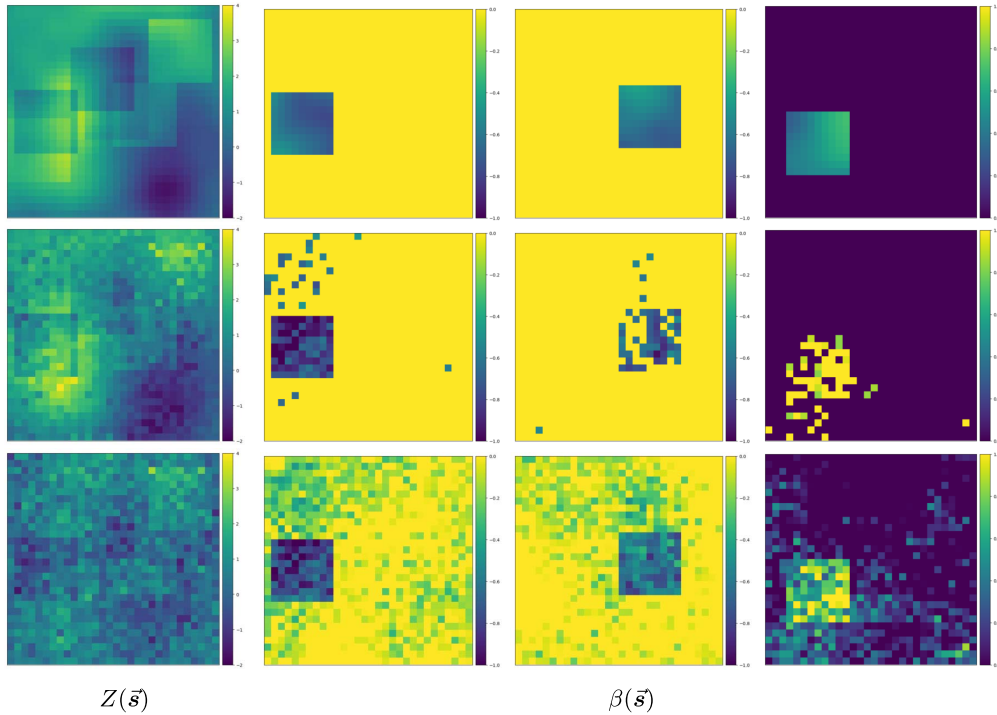
Figure 2: Second simulated scenario ($\pi = 9\%$): Example images of $Z(\vec{s})$ and $\beta(\vec{s})$, where 1st row is generated data; 2nd row is the estimates of BHM; and 3rd row is the estimates of MUA.

mizing the mean square error between the sample covariance estimate and the Matérn kernel. This provides a prior with the closest kernel to the empirical covariance matrix by Frobenius norm and prevents singularity issues caused by $n < p$. We set $\delta = 5$ for the IWP prior in Equation (2.7), following Yang et al. (2016). Finally, we set a weakly-informative Inverse-Gamma prior on the noise variance $\sigma_{\epsilon,\boldsymbol{s}}^2$ by setting $a_\epsilon = b_\epsilon = 1$.

3.2   Results

All results we report were obtained by running MCMC chains with 2000 iterations and 500 burn-in. A single chain took around 8 minutes to run on a 6-core 2.6GHz Intel(R) core i7 CPU. For each chain, convergence was assessed by inspecting the MCMC traces, and more formally using the Geweke test (Geweke, 1992) to check for signs of non-convergence of the individual parameters. As an example, the z-scores from the Geweke test were 0.9603 for $\{\tau_j(\vec{s})\}_{j=1}^{15}$ and 1.0581 for $\{\pi_j\}_{j=1}^{15}$, clearly indicating that the MCMC chains were run for a sufficient number of iterations.

We evaluated performance for variable selection and parameter estimation. For vari-

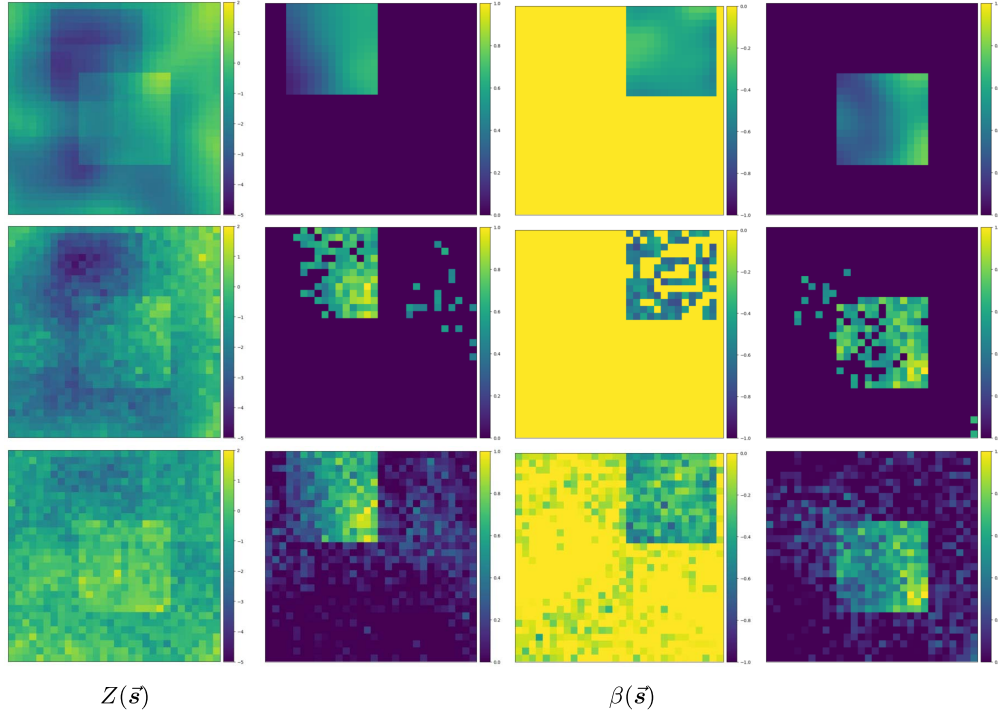$$Z(\vec{s}) \qquad\qquad\qquad\qquad \beta(\vec{s})$$

Figure 3: Second simulated scenario ($\pi \approx 18.8\%$): Example images of $Z(\vec{s})$ and $\beta(\vec{s})$, where 1st row is generated data; 2nd row is the estimates of BHM; and 3rd row is the estimates of MUA.

able selection, we calculated

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad \text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}.$$

For parameter estimation, we evaluated performances by calculating mean squared errors (MSEs) as

$$\text{MSE} = \frac{1}{|\boldsymbol{A}|} \sum_{\boldsymbol{a} \in \boldsymbol{A}} (F(\boldsymbol{a}) - \hat{F}(\boldsymbol{a}))^2,$$

where $F(\cdot)$ and $\hat{F}(\cdot)$ represent the true and estimated parameters, respectively, and $\boldsymbol{a}$ represents the vector of the related indices and/or locations, e.g. $i = 1, \ldots, n; j = 1, \ldots, q; \boldsymbol{s}, \boldsymbol{s}' \in \vec{\boldsymbol{s}}$. We report the accumulated MSE of all coefficient images $\{\beta_j(\vec{s})\}_{j=0}^{15}$ as a summary measure of performance.

We first showcase inference from our BHM model on one simulated data set and then perform comparisons on 50 replicated data sets. Figures 1, 2 and 3 show estimates from BHM of the example images for one data set from each of the three simulated scenarios. The MUA estimates are also shown, for comparison. Results show that the

First simulated scenario									
	Global				Local				
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
Precision	1	1	0.901	0.973	0.985	0.981	1	0.949	0.958
Recall	1	0.821	1	0.983	0.917	0.944	0.742	0.823	0.769
F_1 scores	1	0.902	0.948	0.978	0.950	0.962	0.852	0.882	0.854
Second simulated scenario ($\pi = 9\%$)									
	Global				Local				
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
Precision	1	0.743	0.673	0.714	0.551	—	0.505	0.276	—
Recall	0.75	1	0.432	0.617	0.728	0	0.568	0.840	0
F_1 scores	0.857	0.853	0.526	0.662	0.628	—	0.535	0.416	—
Second simulated scenario ($\pi \approx 18.8\%$)									
	Global				Local				
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
Precision	1	0.806	0.923	0.858	0.753	0.890	0.839	0.793	0.480
Recall	1	0.663	0.355	0.609	0.959	0.769	0.615	0.840	0.852
F_1 scores	1	0.727	0.513	0.713	0.844	0.825	0.710	0.816	0.614

Table 1: BHM ($d = 0.05$): Global-local selection for a representative dataset.

coefficient images estimates can capture the pixel-level information relatively well, even with the location-wise independent priors (2.3)-(2.4) for spatially-dependent coefficient images.

Tables 1 reports precision, recall and F_1 scores for both global and local selections and Table 2 reports the MSEs for the parameters of interests. The proposed method performs well at the global level selection, leading to precision, recall and F_1 scores all relatively high. At the local level selection, some differences are noted among the different simulated scenarios, in particular in the second scenario with the sparser case $\pi = 9\%$, as in this scenario influential covariates are closer to noisy ones. Also, results vary with the covariates' types, with coefficient images for discrete covariates being challenging for local selection, as shown by the lower recalls and F_1 scores. The MSEs of all parameters of interests are relatively small, demonstrating that BHM can estimate those parameters relatively well. Results for BHM with $d = 0.01$ and $d = 0.1$ are reported in the supplementary material. As expected, as d increases, precision tends to increase and recall tends to decrease. However, when there exists a 'good separation' between the influential and noisy covariates, like in the first scenario and the second scenario with $\pi = 18.8\%$, good performances overall can be observed for different choices of d.

3.3   Performance Comparisons

Next, we simulate 50 replicated data sets, according to the same settings described above, and compare the performance of BHM with MUA methods. MUA approaches

MSE				
Scenarios	$\{Z_i(\vec{s})\}_{i=1}^{100}$	$\Sigma(\vec{s}, \vec{s})$	$\{\beta_j(\vec{s})\}_{j=0}^{15}$	σ_ϵ^2
First	0.1816	0.0143	0.0308	0.0015
Second ($\pi = 9\%$)	0.1428	0.0113	0.0345	0.0014
Second ($\pi \approx 18.8\%$)	0.1639	0.0102	0.0363	0.0025

Table 2: BHM ($d = 0.05$): MSEs for a representative dataset.

Averaged Precision									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	0.944	1	0.937	0.949	0.962	0.952	1	0.953	0.931
MUA (SBH)	0.964	1	0.963	0.952	0.954	0.947	1	0.982	0.968
MUA (BH)	0.993	1	0.996	0.988	0.985	0.979	1	0.993	0.988
MUA (BY)	1	1	0.999	0.998	0.998	0.996	1	0.999	0.999
Averaged Recall									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	1	0.970	0.954	0.920	0.885	0.865	0.764	0.752	0.690
MUA (SBH)	0.992	0.957	0.936	0.900	0.871	0.846	0.522	0.506	0.422
MUA (BH)	0.982	0.817	0.806	0.811	0.790	0.782	0.257	0.291	0.253
MUA (BY)	0.948	0.656	0.628	0.654	0.632	0.622	0.069	0.080	0.070
Averaged F_1 scores									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	0.970	**0.983**	0.943	**0.932**	**0.919**	**0.903**	**0.837**	**0.814**	**0.771**
MUA (SBH)	0.977	0.975	**0.947**	0.921	0.906	0.888	0.642	0.622	0.555
MUA (BH)	**0.987**	0.892	0.884	0.884	0.868	0.862	0.379	0.440	0.386
MUA (BY)	0.972	0.775	0.755	0.775	0.751	0.749	0.145	0.154	0.143

Table 3: First simulated scenario: Global and local selection for 50 replicates.

fit independent linear regressions at each location s, to estimate coefficient images, and
rely on post-inference to do variable selection and smoothing. For global selection, we
first use Simes test (Simes, 1986) to convert multiple p-values at each location to one
single p-value for the whole coefficient image, and then control the False Discovery Rate
(FDR) at 0.05 for the 15 coefficient images. We implement three different FDR control
procedures, the Benjamini–Hochberg (BH) procedure (Benjamini and Hochberg, 1995),
the Benjamini–Yekutieli (BY) procedure (Benjamini and Yekutieli, 2001), both imple-
mented in R via the function '*p.adjust*', and another Benjamini-Hochberg procedure
described in Strimmer (2008), implemented in the R package '*fdrtool*', which estimates
the proportion of null features from data. We denote the third procedure by MUA
(SBH). Local level selection is achieved by applying these three procedures to control
the FDR at 0.05 for each coefficient image.

Table 3 reports precision, recall and F_1 scores for both global-level selection and
local-level selection, averaged over 50 replicates from the first scenario, for MUA (BH),
MUA (BY), MUA (SBH) and BHM with SGLSS prior and $d = 0.05$. For the global level
selection, all the methods achieve similarly high values for all three metrics, indicating

that the influential covariates can potentially be well distinguished from the noisy ones. For the local level selection, we observe BHM ($d = 0.05$) and MUA (SBH) have similar performance with respect to the continuous covariates, $\tau_{1,2,3,4,5}(\vec{s})$. Meanwhile, when it comes to the discrete covariates $\tau_{6,7,8}(\vec{s})$, BHM ($d = 0.05$) achieves higher averaged F_1 scores than the other methods, due to a relatively better balance between precision and recall. As for the other MUA approaches, MUA (BH) and MUA (BY) have higher precision but much lower recall, leading to lower F_1 scores, especially for the discrete covariates.

Table 4 reports the three metrics for the second scenario with two settings, $\pi = 9\%$ and $\pi \approx 18.8\%$. As noted above, this scenario is more challenging since influential covariates are closer to the noisy ones, especially for the sparser case $\pi = 9\%$. However, results are relatively consistent with the previous setting. For the global level selection, all methods achieve comparably high metrics. We notice some precision-recall trade-offs, while high F_1 scores result from a relative balance between precision and recall. For the local level selection, BHM ($d = 0.05$) obtains similarly high F_1 scores as MUA (SBH) and MUA (BH) on the continuous covariates, and higher F_1 scores on the discrete covariates. Although the precision of BHM($d = 0.05$) is not as high as the other methods, its recall is relatively higher, leading to comparable F_1 scores. At the same time, although MUA (BY) has the lowest F_1 scores, it has the highest precision in both settings.

Table 5 report the MSEs for the parameters of interest and their standard errors (SE). The MUA estimators are best linear unbiased estimators (BLUE) at each location, and indeed lead to relatively accurate estimates for the coefficient images $\{\beta_j(\vec{s})\}_{j=0}^{15}$. Meanwhile, the proposed BHM with SGLSS prior can return comparably good estimates since the global level indicators can exclude noisy covariates, leading to zero errors when global-level selection is done correctly. In addition, given its hierarchical structure, BHM also produces estimates for noise-free mean surface $\{Z_i(\vec{s})\}_{i=1}^{100}$ and covariance surface $\Sigma(\vec{s}, \vec{s})$, which are shown to be relatively accurate.

4 Real Data Application

We demonstrate the proposed model using the Autism Brain Imaging Data Exchange (ABIDE) study of Di Martino et al. (2014). The study collected resting-state fMRI data from 17 experiment sites including 1112 subjects, with the aim of improving the understanding of neurophysiological mechanisms. For each subject, rs-fMRI data were recorded over time, along with the subject's information such as age, gender, intelligence quotient, etc. To aid computations, we reduced the image size by summarizing the fMRI data to voxel-level imaging statistics and then considered individual brain networks instead of the whole brain image. Following He et al. (2019) and Zhang et al. (2020), we used the pipeline of Cameron et al. (2013) to preprocess the data and then considered a parcellation of the brain as defined by the Automated Anatomical Labeling (Tzourio-mazoyer et al., 2002) to select networks. The selected networks are described in Table 6, and are known to be associated with cognitive ability based on previous research (van den Heuvel et al., 2009; Wu et al., 2013; Hearne et al., 2016; Hilger et al., 2017; Zhang et al., 2020). As for the covariates, those collected in the 17 experiment

Averaged Precision ($\pi = 9\%$)									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	0.919	0.708	0.754	0.671	0.726	0.762	0.484	0.386	0.454
MUA (SBH)	0.934	0.926	0.948	0.902	0.948	0.964	0.906	0.910	0.775
MUA (BH)	0.970	0.946	0.958	0.930	0.959	0.970	0.930	0.981	0.814
MUA (BY)	0.997	0.990	0.994	0.993	0.996	0.996	1	1	0.999
Averaged Precision ($\pi \approx 18.8\%$)									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	0.926	0.824	0.830	0.823	0.831	0.829	0.686	0.689	0.652
MUA (SBH)	0.945	0.949	0.952	0.953	0.956	0.959	0.927	0.901	0.918
MUA (BH)	0.977	0.960	0.963	0.965	0.966	0.967	0.949	0.913	0.928
MUA (BY)	0.998	0.991	0.996	0.997	0.995	0.997	0.997	0.995	0.993
Averaged Recall ($\pi = 9\%$)									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	0.742	0.681	0.579	0.570	0.592	0.680	0.272	0.232	0.253
MUA (SBH)	0.755	0.677	0.569	0.568	0.598	0.699	0.071	0.079	0.107
MUA (BH)	0.690	0.672	0.571	0.559	0.589	0.691	0.061	0.074	0.101
MUA (BY)	0.600	0.514	0.395	0.439	0.424	0.530	0.013	0.023	0.042
Averaged Recall ($\pi \approx 18.8\%$)									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	0.945	0.794	0.750	0.797	0.781	0.770	0.397	0.367	0.441
MUA (SBH)	0.870	0.760	0.697	0.762	0.743	0.73	0.153	0.161	0.175
MUA (BH)	0.818	0.738	0.682	0.748	0.730	0.712	0.131	0.148	0.147
MUA (BY)	0.732	0.553	0.532	0.586	0.589	0.530	0.033	0.053	0.039
Averaged F_1 scores ($\pi = 9\%$)									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	0.812	0.718	0.705	0.670	**0.716**	**0.754**	**0.450**	**0.387**	**0.412**
MUA (SBH)	**0.823**	0.753	0.714	0.680	0.699	0.752	0.227	0.267	0.345
MUA (BH)	0.795	**0.761**	**0.722**	**0.689**	0.693	0.748	0.195	0.266	0.330
MUA (BY)	0.735	0.670	0.577	0.611	0.571	0.682	0.124	0.154	0.241
Averaged F_1 scores ($\pi \approx 18.8\%$)									
Methods	Global	Local							
		$\tau_1(\vec{s})$	$\tau_2(\vec{s})$	$\tau_3(\vec{s})$	$\tau_4(\vec{s})$	$\tau_5(\vec{s})$	$\tau_6(\vec{s})$	$\tau_7(\vec{s})$	$\tau_8(\vec{s})$
BHM ($d = 0.05$)	**0.932**	0.795	0.771	0.796	0.801	0.793	**0.518**	**0.511**	**0.527**
MUA (SBH)	0.900	**0.816**	**0.774**	**0.813**	**0.811**	**0.794**	0.310	0.454	0.308
MUA (BH)	0.885	0.803	0.767	0.808	0.805	0.784	0.285	0.455	0.274
MUA (BY)	0.837	0.671	0.636	0.691	0.720	0.666	0.131	0.261	0.185

Table 4: Second simulated scenario: Global and local selection for 50 replicates.

sites include diagnostic, age, gender and full-scale intelligence quotient scores (FIQ). The FIQ scores were assessed differently across sites, including DAS-II, WASI, WISC, WAIS, RAVENS and STANFORD scales. After removing missing values, we ended up with 1001 subjects. We standardized the continuous variables, age and FIQ scores, to put them on the same scale with the discrete indicators, diagnostic and gender, which we left unchanged. We included a vector of ones as the intercept to account for those potentially influential covariates which are not available in the study. We then applied

	First simulated scenario					
Model	$\{Z_i(\vec{s})\}_{i=1}^{100}$		$\{\beta_j(\vec{s})\}_{j=0}^{15}$		$\Sigma(\vec{s},\vec{s})$	
	Mean	SE	Mean	SE	Mean	SE
BHM ($d = 0.05$)	0.180	(1.0×10^{-3})	0.472	(1.13×10^{-2})	0.014	(0.5×10^{-3})
MUA	-		0.678	(1.15×10^{-2})	-	
	Second simulated scenario ($\pi = 9\%$)					
Model	$\{Z_i(\vec{s})\}_{i=1}^{100}$		$\{\beta_j(\vec{s})\}_{j=0}^{15}$		$\Sigma(\vec{s},\vec{s})$	
	Mean	SE	Mean	SE	Mean	SE
BHM ($d = 0.05$)	0.140	(1.9×10^{-3})	0.540	(2.02×10^{-2})	0.013	(0.5×10^{-3})
MUA	-		0.671	(0.88×10^{-2})	-	
	Second simulated scenario ($\pi \approx 18.8\%$)					
Model	$\{Z_i(\vec{s})\}_{i=1}^{100}$		$\{\beta_j(\vec{s})\}_{j=0}^{15}$		$\Sigma(\vec{s},\vec{s})$	
	Mean	SE	Mean	SE	Mean	SE
BHM ($d = 0.05$)	0.162	(1.4×10^{-3})	0.597	(1.36×10^{-2})	0.013	(0.4×10^{-3})
MUA	-		0.664	(0.93×10^{-2})	-	

Table 5: MSEs for 50 replicates.

Network	Number of voxels	Regions included			
Visual	7946	Lingual L	Lingual R	Calcarine L	Cuneus R
Ventral Attention	9839	Temporal Mid L	Temporal Sup L	Temporal Sup R	-
Dorsal Attention	9600	Temporal Mid R	Postcentral L	Parietal Sup L	-
Default Mode	7440	Temporal Mid R	Frontal Med Orb R	Frontal Med Orb L	Occipital Mid R

Table 6: Networks of interests.

BHM with the SGLSS prior, separately, to the four selected networks.

We specified the threshold d at the conventional sparsity level $d = 0.05$. As for the slab prior specification, we set $\{\mu_{0j}(\vec{s})\}_{j=0}^{q} = 0$ and specified $\{\sigma_{0j}^2(\vec{s})\}_{j=0}^{q} = 1$. We ran MCMC chains with 2000 iterations and 500 burn-in. On average, the z-scores from the Geweke test were 0.9616 for $\{\tau_j(\vec{s})\}_{j=1}^{4}$ and 1.1858 for $\{\pi_j\}_{j=1}^{4}$, indicating that the MCMC chains were run for a sufficient number of iterations. The MCMC chain took around 35 seconds per iteration on two 20-core 2.4 GHz Intel(R) Xeon CPUs for networks with nearly $10,000$ voxels.

Table 7 shows selection results for BHM and the MUA methods. For local level selection, π denotes the ratio of selected voxels for MUA-based methods and the posterior mean of the participation rate π_j for BHM. The check marks denote whether the covariate is selected at global level. For global level selection, all methods agree on selecting the covariate *age* as influential for human cognitive ability, which makes sense because as people age the brain naturally changes, along with its cognitive functions. The difference is in the selection of the *FIQ* scores. The MUA-based methods tend to include the FIQ in the model, while the proposed BHM ($d = 0.05$) considers FIQ to be related only to Visual network and Dorsal network. At the local level selection, the MUA with Benjamini-Hochberg based procedures tend to select the covariates at almost all voxels. Meanwhile, MUA with Benjamini-Yekutieli procedure tends to select the covariates at fewer voxels than the MUA (BH) especially when it comes to the FIQ

		MUA (SBH)				BHM (d = 0.05)			
		diagnostic	age	gender	FIQ	diagnostic	age	gender	FIQ
Visual	$\pi(\%)$	0.0	100.0	0.0	100.0	0.01	77.2	0.01	9.24
	if selected		(✓)		(✓)		(✓)		(✓)
Ventral	$\pi(\%)$	0.0	100.0	0.0	100.0	0.01	83.6	0.01	1.38
	if selected		(✓)		(✓)		(✓)		
Dorsal	$\pi(\%)$	0.0	85.8	0.0	100.0	0.01	42.9	0.01	7.77
	if selected		(✓)		(✓)		(✓)		(✓)
Default	$\pi(\%)$	0.0	100.0	0.0	100.0	0.01	71.6	0.01	0.78
	if selected		(✓)		(✓)		(✓)		

		MUA (BH)				MUA (BY)			
		diagnostic	age	gender	FIQ	diagnostic	age	gender	FIQ
Visual	$\pi(\%)$	0.0	98.8	0.0	92.3	0.0	83.0	0.0	32.7
	if selected		(✓)		(✓)		(✓)		(✓)
Ventral	$\pi(\%)$	0.0	97.1	0.0	90.6	0.0	82.2	0.0	24.7
	if selected		(✓)		(✓)		(✓)		(✓)
Dorsal	$\pi(\%)$	0.0	74.9	0.0	88.2	0.0	54.3	0.0	22.7
	if selected		(✓)		(✓)		(✓)		(✓)
Default	$\pi(\%)$	0.0	90.4	0.0	89.6	0.0	74.5	0.0	24.4
	if selected		(✓)		(✓)		(✓)		(✓)

Table 7: Selection results for the four networks.

scores. The BHM-based methods tend to have similar results as the MUA (BY) on the selection of *age*. As for the selection of FIQ, however, BHM with $d = 0.05$ tends to select even fewer voxels. Furthermore, when fitting the BHM model we noticed that a more stringent threshold of $d = 0.1$ would exclude FIQ for all networks entirely, while a less stringent threshold of $d = 0.01$ would include FIQ for all networks but only for very few selected voxels. These results suggest that, although FIQ score may somehow be related to brain signals, this relationship can be hard to recover in this application, possibly because different experimental sites use different standards to measure this covariate.

Figure 4 shows the selected voxels for the covariate *age* by MUA (SBH) (Left), MUA (BY) (Middle) and BHM ($d = 0.05$) (Right). We observe a decreasing number of voxels selected by the methods, and similar local selective results between MUA (BY) and BHM (0.05), both of which tend to have a more sparse selection with selected voxels mainly in the central portions of the regions in the functional networks. Table 8 reports the ratios of region included in the local selection, showing consistent selection results for BHM. We note that the final selection is determined by the posterior summary, i.e. the median rule, based on the posterior samples, while the threshold $I(\pi \geq d)$ takes effect at each iteration. Hence, although the global indicator can guarantee $\pi \geq d$ at each iteration, the final ratios of selected voxels/pixels are not necessarily greater than d, as it is evident from the results. We report results for BHM with $d = 0.01$ and $d = 0.1$ in the supplementary material and note here that BHM maintains highly consistent local selection results with different specification of d, i.e. when Ceneus R is considered to be affected by FIQ, BHM ($d = 0.01$) selects 1.6% of the region and BHM ($d = 0.05$) selects 1.77%; when Temporal Mid R is considered to be affected by FIQ, BHM ($d = 0.01$) selects 1.7% of the region and BHM ($d = 0.05$) selects 1.5%. These results also confirm the previous observation that FIQ scores, converted from different standards, may not
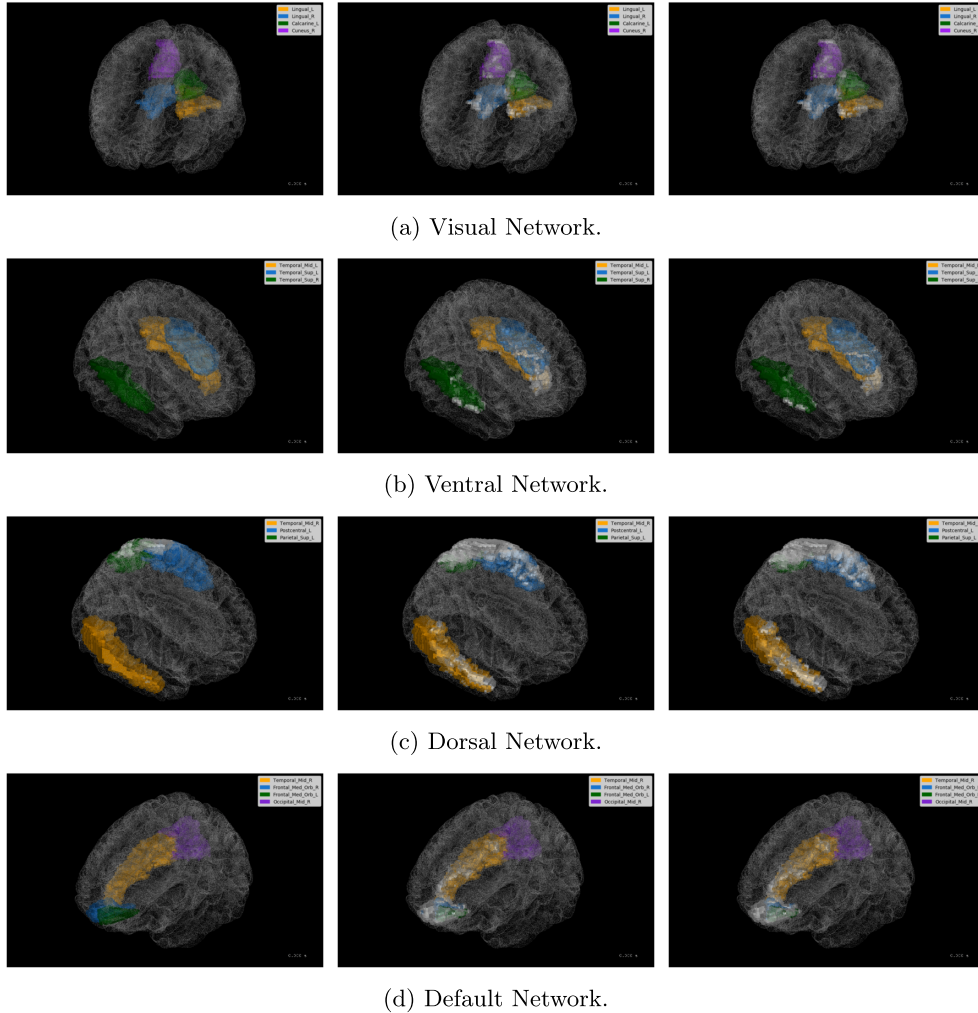
(a) Visual Network.

(b) Ventral Network.

(c) Dorsal Network.

(d) Default Network.

Figure 4: Selected voxels for covariate '*age*', by MUA with SBH (Left), MUA with BY (Middle) and BHM with SGLSS ($d = 0.05$) (Right). Figures are plotted using the R package *threeBrain* by Magnotti et al. (2020).

show strong relationship to the brain regions.

5   Concluding Remarks

In this article, we have extended to image data a Bayesian hierarchical Gaussian process (GP) model that uses a flexible Inverse-Wishart process prior to handle within-image dependency, and have proposed a novel spatial global-local spike-and-slab prior

Network	Methods	Covariates	Ratio of Region included (%)			
			Lingual L	Lingual R	Calcarine L	Cuneus R
Visual	BHM ($d = 0.05$)	age	82.1	81.4	83.5	88.9
		FIQ	5.79	7.00	3.29	1.77
	MUA (SBH)	age	100	100	100	100
		FIQ	100	100	100	100
	MUA (BH)	age	98.9	97.4	99.4	99.7
		FIQ	95.3	89.4	93.2	91.1
	MUA (BY)	age	81.8	81.5	82.4	88.2
		FIQ	34.2	34.7	29.1	32.9
			Temporal Mid L	Temporal Sup L	Temporal Sup R	-
Ventral	BHM ($d = 0.05$)	age	80.7	90.1	96.4	-
		FIQ	0.0	0.0	0.0	-
	MUA (SBH)	age	100	100	100	-
		FIQ	100	100	100	-
	MUA (BH)	age	94.5	99.7	99.8	-
		FIQ	84.8	96.9	95.8	-
	MUA (BY)	age	76.2	81.7	93.7	-
		FIQ	18.9	29.6	31.2	-
			Temporal Mid R	Postcentral L	Parietal Sup L	-
Dorsal	BHM ($d = 0.05$)	age	62.9	38.6	13.8	-
		FIQ	1.5	4.7	4.5	-
	MUA (SBH)	age	99.9	83.8	62.7	-
		FIQ	100	100	100	-
	MUA (BH)	age	96.9	74.0	34.5	-
		FIQ	89.6	91.7	79.1	-
	MUA (BY)	age	75.8	51.6	18.0	-
		FIQ	21.1	26.8	18	-
			Temporal Mid R	Frontal Med Orb R	Frontal Med Orb L	Occipital Mid R
Default	BHM ($d = 0.05$)	age	77.7	26.2	22.7	99.5
		FIQ	0.0	0.0	0.0	0.0
	MUA (SBH)	age	100	100	100	100
		FIQ	100	100	100	100
	MUA (BH)	age	97.6	62.0	57.7	100.0
		FIQ	89.7	70.1	84.1	99.2
	MUA (BY)	age	79.5	31.9	25.7	99.6
		FIQ	22.7	6.2	11.1	39.6

Table 8: Ratios of Region included within each networks.

that broadly relates to a rich class of well-studied selection priors. The proposed prior construction achieves simultaneous global (i.e, at covariate-level) and local (i.e., at pixel/voxel-level) selection via participation rate parameters that measure the probability for the individual covariates to affect the observed images. We have used hard-thresholding to decide whether a covariate should be included in the model and have shown on simulated data that parameters are interpretable and lead to efficient selection. The introduced participation rate and threshold parameters establish a bridge between global and local level selection, allowing global selection to be informed by the selection at the local level. This framework can be applied to more general functional data applications.

There are several interesting future directions to extend our model. Our rationale for choosing an independent prior on spatial coefficients $\beta_j(s)$ has been largely com-

putational. Our efficient Gibbs sampler takes advantage of this independent prior and only requires to invert the $|\boldsymbol{S}| - by - |\boldsymbol{S}|$ covariance matrix once for the noise-free mean surface $Z(\cdot)$, roughly $O(|\boldsymbol{S}|^3)$ at each Gibbs iteration. In the application to real data, our model is able to handle relatively large datasets, with S about 10,000 voxels, and n about 1,000 subject images. With a dependent prior, we would not be able to parallelize computations, which would result into having to calculate the inversion of at most $|\boldsymbol{S}| - by - |\boldsymbol{S}|$ covariance matrices for q covariates at each iteration, with roughly a $O(|\boldsymbol{S}|^3 q)$ complexity at each iteration. This would have been infeasible for our application. In addition, a construction with a dependent prior would require a more careful interpretation of the participation rate parameters π_j's, which measure the probability for the individual covariates to affect the observed image under the assumption of independence. Given these challenges, we have decided to leave the investigation of dependent priors to future work. We note, however, that, even though we do not explicitly account for dependency among the coefficients, our model borrows information across voxels via the use of the spatial Gaussian process prior $\mathcal{GP}\left(\mu(\cdot), \Sigma(\cdot, \cdot)\right)$ on Z_i.

In the applications of this paper, when investigating the role of the parameter d and the sensitivity of the results to the specification of this parameter, we found the case $d = 0$ interesting. In this degenerate case the model includes all the covariates at each iteration, to explain the observed images, and the traces of the parameters π_j inform us on the relative importance of the individual covariates. These trace plots provide an empirical tool that might be helpful in the choice of d, particularly in cases where a separation among the traces is observed. In the Supplementary Material we show these plots for one of the simulated scenarios used in this paper, along with comments on how the plots can guide the user in the choice of d. We remark, however, that this procedure is ad-hoc and cannot be used as a general method, in particular as the behavior of the trace plots is application-dependent and a clear separation of the traces might not always be observed. We leave further investigation of the role and properties of the parameter d to future work. In the absence of prior information, we recommend to view d as conventional sparsity parameter and use standard values, i.e. $d = 0.05$ or $d = 0.1$. Our sensitivity analyses in the simulations and real data application have shown good performances overall for different choices of d, with highly consistent local selection results.

Finally, our proposed global-local selection prior construction can be potentially useful for other modeling settings, such as function-on-scalar and network-on-scalar regressions.

Supplementary Material

References

Barbieri, M. M. and Berger, J. O. (2004). "Optimal predictive model selection." *The Annals of Statistics*, 32(3): 870–897. MR2065192. doi: https://doi.org/10.1214/009053604000000238. 244

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 57: 289–300. MR1325392. 249

Benjamini, Y. and Yekutieli, D. (2001). "The control of the False Discovery Rate in multiple testing under dependency." *The Annals of Statistics*, 29(4): 1165–1188. MR1869245. doi: https://doi.org/10.1214/aos/1013699998. 249

Brown, P. J., Vannucci, M., and Fearn, T. (1998). "Multivariate Bayesian variable selection and prediction." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3): 627–641. MR1626005. doi: https://doi.org/10.1111/1467-9868.00144. 242

Cameron, C., Sharad, S., Brian, C., Ranjeet, K., Satrajit, G., Chaogan, Y., Qingyang, L., Daniel, L., Joshua, V., Randal, B., Stanley, C., Maarten, M., Clare, K., Adriana, D. M., Francisco, C., and Michael, M. (2013). "Towards Automated Analysis of Connectomes: The Configurable Pipeline for the Analysis of Connectomes (C-PAC)." *Frontiers in Neuroinformatics*, 7. 250

Chen, R.-B., Chu, C.-H., Yuan, S., and Wu, Y. N. (2016a). "Bayesian sparse group selection." *Journal of Computational and Graphical Statistics*, 25: 665–683. MR3533632. doi: https://doi.org/10.1080/10618600.2015.1041636. 240

Chen, Y., Goldsmith, J., and Ogden, R. T. (2016b). "Variable selection in function-on-scalar regression." *Stat*, 5: 101–88. MR3478799. doi: https://doi.org/10.1002/sta4.106. 236

Choi, J. and Lawson, A. B. (2018). "Bayesian spatially dependent variable selection for small area health modeling." *Statistical Methods in Medical Research*, 27: 234–249. MR3745667. doi: https://doi.org/10.1177/0962280215627184. 236, 239

Chumbley, J. R. and Friston, K. J. (2009). "False Discovery Rate revisited: FDR and topological inference using Gaussian random fields." *NeuroImage*, 44: 62–70. 236

Dawid, A. (1981). "Some matrix-variate distribution theory: Notational considerations and a Bayesian application." *Biometrika*, 68: 265–274. MR0614963. doi: https://doi.org/10.1093/biomet/68.1.265. 241

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., et al. (2014). "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism." *Molecular Psychiatry*, 19(6): 659–667. 235, 237, 250

Geweke, J. (1992). "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In *Bayesian Statistics*, 169–193. University Press. MR1380276. 246

Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014). "Smooth Scalar-on-Image Regression via Spatial Bayesian Variable Selection." *Journal of Computational and Graphical Statistics*, 23: 46–64. MR3173760. doi: https://doi.org/10.1080/10618600.2012.743437. 236, 239

Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). "The Wishart Autoregressive Process of Multivariate Stochastic Volatility." *Journal of Econometrics*, 150: 167–181. MR2535514. doi: https://doi.org/10.1016/j.jeconom.2008.12.016. 240

Groppe, D. M., Urbach, T., and Kutas, M. (2011). "Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review." *Psychophysiology*, 48 12: 1711–1725. 235

He, K., Xu, H., and Kang, J. (2019). "A selective overview of feature screening methods with applications to neuroimaging data." *WIREs Computational Statistics*, 11(2): e1454. MR3920631. doi: https://doi.org/10.1002/wics.1454. 250

Hearne, L. J., Mattingley, J., and Cocchi, L. (2016). "Functional brain networks related to individual differences in human intelligence at rest." *Scientific Reports*, 6. 250

Heaukulani, C. and van der Wilk, M. (2019). "Scalable Bayesian dynamic covariance modeling with variational Wishart and inverse Wishart processes." In *NeurIPS*. 240

Hilger, K., Ekman, M., Fiebach, C., and Basten, U. (2017). "Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence." *Intelligence*, 60: 10–25. 250

Kang, J., Reich, B. J., and Staicu, A.-M. (2018). "Scalar-on-image regression via the soft-thresholded Gaussian process." *Biometrika*, 105(1): 165–184. MR3768872. doi: https://doi.org/10.1093/biomet/asx075. 236, 240

Kowal, D. R. and Bourgeois, D. C. (2020). "Bayesian Function-on-Scalars Regression for High-Dimensional Data." *Journal of Computational and Graphical Statistics*, 29: 629–638. MR4153187. doi: https://doi.org/10.1080/10618600.2019.1710837. 236

Li, F., Zhang, T., Wang, Q., Gonzalez, M. Z., Maresh, E. L., and Coan, J. A. (2015). "Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression." *The Annals of Applied Statistics*, 9(2): 687–713. MR3371331. doi: https://doi.org/10.1214/15-AOAS818. 236, 239

Li, X., Wang, L., and Wang, H. J. (2020). "Sparse Learning and Structure Identification for Ultrahigh-Dimensional Image-on-Scalar Regression." *Journal of the American Statistical Association*, 1–15. MR4353728. doi: https://doi.org/10.1080/01621459.2020.1753523. 235, 236, 244

Liquet, B., Mengersen, K., Pettitt, A., and Sutton, M. (2017). "Bayesian Variable Selection Regression of Multivariate Responses for Group Data." *Bayesian Analysis*, 12: 1039–1067. MR3724978. doi: https://doi.org/10.1214/17-BA1081. 240

Liu, Y., Li, M., and Morris, J. S. (2020). "Function-on-scalar quantile regression with

application to mass spectrometry proteomics data." *The Annals of Applied Statistics*, 14(2): 521–541. MR4117818. doi: https://doi.org/10.1214/19-AOAS1319. 237

Magnotti, J. F., Wang, Z., and Beauchamp, M. S. (2020). "RAVE: Comprehensive open-source software for reproducible analysis and visualization of intracranial EEG data." *NeuroImage*, 223: 117341. 254

Nakajima, J. and West, M. (2013). "Bayesian Analysis of Latent Threshold Dynamic Models." *Journal of Business & Economic Statistics*, 31: 151–164. MR3055329. doi: https://doi.org/10.1080/07350015.2012.747847. 236, 240

Philipov, A. and Glickman, M. (2006). "Multivariate Stochastic Volatility via Wishart Processes." *Journal of Business & Economic Statistics*, 24: 313–328. MR2252482. doi: https://doi.org/10.1198/073500105000000306. 240

Reich, B. J., Fuentes, M., Herring, A. H., and Evenson, K. R. (2010). "Bayesian Variable Selection for Multivariate Spatially Varying Coefficient Regression." *Biometrics*, 66: 772–782. MR2758213. doi: https://doi.org/10.1111/j.1541-0420.2009.01333.x. 236, 240, 242

Savitsky, T. D., Vannucci, M., and Sha, N. (2011). "Variable Selection for Nonparametric Gaussian Process Priors: Models and Computational Strategies." *Statistical science: a review journal of the Institute of Mathematical Statistics*, 26 1: 130–149. MR2849913. doi: https://doi.org/10.1214/11-STS354. 243

Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O., Hinnerichsen, M., and Meze-Hausken, E. (2013). "A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1): 85–100. MR3042312. doi: https://doi.org/10.1111/j.1467-9876.2012.01039.x. 236, 239

Shamshoian, J., Şentürk, D., Jeste, S., and Telesca, D. (2020). "Bayesian analysis of longitudinal and multidimensional functional data." *Biostatistics*. Online ahead of print. MR4424984. doi: https://doi.org/10.1093/biostatistics/kxaa041. 237

Simes, R. J. (1986). "An improved Bonferroni procedure for multiple tests of significance." *Biometrika*, 73: 751–754. MR0897872. doi: https://doi.org/10.1093/biomet/73.3.751. 249

Smith, M. and Fahrmeir, L. (2007). "Spatial Bayesian Variable Selection with Application to Functional Magnetic Resonance Imaging." *Journal of the American Statistical Association*, 102(478): 417–431. MR2370843. doi: https://doi.org/10.1198/016214506000001031. 236, 239, 242

Stingo, F. C., Chen, Y. A., Tadesse, M. G., and Vannucci, M. (2011). "Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes." *The Annals of Applied Statistics*, 5(3): 1978–2002. MR2884929. doi: https://doi.org/10.1214/11-AOAS463. 240

Strimmer, K. (2008). "fdrtool: a versatile R package for estimating local and tail area-based False Discovery Rates." *Bioinformatics*, 24(12): 1461–1462. 249

Tzourio-mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain." *NeuroImage*, 15: 273–289.   250

van den Heuvel, M. P., Stam, C., Kahn, R., and Pol, H. H. H. (2009). "Efficiency of Functional Brain Networks and Intellectual Performance." *The Journal of Neuroscience*, 29: 7619–7624.   250

Wilson, A. and Ghahramani, Z. (2011). "Generalised Wishart processes." 736–744. 27th Conference on Uncertainty in Artificial Intelligence.   240

Worsley, K. J., Taylor, J. E., Tomaiuolo, F., and Lerch, J. P. (2004). "Unified univariate and multivariate random field theory." *NeuroImage*, 23: S189–S195.   235

Wu, K., Taki, Y., Sato, K., Hashizume, H., Sassa, Y., Takeuchi, H., Thyreau, B., He, Y., Evans, A. C., Li, X., Kawashima, R., and Fukuda, H. (2013). "Topological Organization of Functional Brain Networks in Healthy Children: Differences in Relation to Age, Sex, and Intelligence." *PLoS ONE*, 8.   250

Xu, X. and Ghosh, M. (2015). "Bayesian Variable Selection and Estimation for Group Lasso." *Bayesian Analysis*, 10: 909–936. MR3432244. doi: https://doi.org/10.1214/14-BA929.   240

Yang, J., Zhu, H., Choi, T., and Cox, D. D. (2016). "Smoothing and Mean–Covariance Estimation of Functional Data with a Bayesian Hierarchical Model." *Bayesian Analysis*, 11: 649–670. MR3498041. doi: https://doi.org/10.1214/15-BA967.   236, 237, 239, 241, 246

Yu, S., Wang, G., lian Wang, L., and Yang, L. (2021). "Multivariate Spline Estimation and Inference for Image-on-Scalar Regression." *Statistica Sinica*, 31: 1463–1487. MR4297712. doi: https://doi.org/10.5705/ss.202019.0188.   235, 236

Zeng, Z., Li, M., and Vannucci, M. (2022). "Supplementary Material for "Bayesian Image-on-Scalar Regression with a Spatial Global-Local Spike-and-Slab Prior"." *Bayesian Analysis*. doi: https://doi.org/10.1214/22-BA1352SUPP.   241

Zhang, D., Li, L., Sripada, C., and Kang, J. (2020). "Image-on-Scalar Regression via Deep Neural Networks." *arXiv e-prints arXiv:2006.09911*.   235, 250

Zhu, H., Fan, J., and Kong, L. (2014). "Spatially Varying Coefficient Model for Neuroimaging Data With Jump Discontinuities." *Journal of the American Statistical Association*, 109: 1084–1098. MR3265682. doi: https://doi.org/10.1080/01621459.2014.881742.   235, 236

Zhu, H., Strawn, N., and Dunson, D. (2016). "Bayesian Graphical Models for Multivariate Functional Data." *Journal of Machine Learning Research*, 17(204): 1–27. MR3580357.   241