Bayesian Multi-object Data Integration in the Study of Primary Progressive Aphasia

Rene Gutierrez*‡ Aaron Scheffler*§ Rajarshi Guhaniyogi*¶

May 31, 2022

Abstract

Clinical researchers often collect multiple images from separate modalities (sources) to investigate fundamental questions of human health that are inadequately explained by considering one image source at a time. Viewing the collection of images as multiple objects, the successful integration of multi-object data produces a sum of information greater than the individual parts, but this integration can be challenging due to the complexity induced through different topological structure of the objects. This article focuses on a multi-modal imaging application where structural/anatomical information from grey matter (GM) and brain connectivity information in the form of a brain connectome network from functional magnetic resonance imaging (fMRI) are available for a limited number of subjects with different degrees of primary progressive aphasia (PPA), a neurodegenerative disorder (NDs) measured through a score on language loss. The clinical/scientific goal in this application becomes the identification of brain regions significantly related to the language score to gain insight into ND pathways. This article develops a flexible Bayesian regression framework exploiting network information of the brain connectome, while leveraging linkages among connectome network

‡Rene Gutierrez, Postdoctoral Researcher, Department of Statistics, Texas A & M, TAMU 3143, College Station, Texas 77843-3143 (E-mail: renegutierrez@tamu.edu).

§Aaron Scheffler, Assistant Professor, Department of Epidemiology and Biostatistics, SOE2, UC San Francisco, 550 16th Street, San Francisco, CA 94158 (E-mail: Aaron.Scheffler@ucsf.edu).

¶Rajarshi Guhaniyogi, Associate Professor, Department of Statistics, Texas A & M University, TAMU 3143, College Station, Texas 77843-3143 (E-mail: rajguhaniyogi@tamu.edu).

and anatomical information from GM to draw inference on brain regions significantly related to the language score. The principled Bayesian framework allows precise characterization of the uncertainty in ascertaining a region being actively related to the language score. Our framework is implemented using an efficient Markov Chain Monte Carlo algorithm. Empirical results with simulated data illustrate substantial inferential gains of the proposed framework over its popular competitors. Our framework yields new insights into the relationship of brain regions with PPA, also providing the uncertainty associated with the scientific findings.

Key Words: Bayesian inference; brain connectome; functional magnetic resonance imaging; grey matter; multi-modal imaging; primary progressive aphasia; spike and slab prior.

1 Introduction

Aided by technological advances in both biomedical hardware and software, neuroscientists routinely collect high-dimensional imaging data from multiple sources (modalities) to interrogate the human brain (Sui *et al.*, 2012). Inspection of multiple brain images produces complementary cross information that can be leveraged to combat Alzheimer's disease (AD) and other neurodegenerative disorders (NDs) by advancing foundational cognitive theory, models of disease progression, and biomarker development (Ossenkoppele *et al.*, 2016). For example, ND progression is best tracked via disruptions of brain structure and networks, and are detected by stitching together information across these images (Mandelli *et al.*, 2016; Gorno-Tempini *et al.*, 2008; Brown *et al.*, 2019). This article is motivated by a clinical application on a limited number of patients suffering from brain loss due to Primary Progressive Aphasia (PPA), an ND with similar pathology to Alzheimer's disease and fronto-temporal dementia. Multi-modal imaging data are available for each of these PPA patients which include: (a) *brain network information* and (b) *brain structural information*. Brain network envisions each region of interest (ROI) as a network node and quantifies connectivity between the pairs of network nodes using functional magnetic resonance imaging (fMRI). Brain structural information is obtained using structural magnetic resonance imaging (sMRI), e.g., grey matter (GM) images obtained over brain volumetric pixels (voxels). Both images are collected on a common brain atlas which segments a human brain into different regions of

interest (ROI), with each ROI containing a number of voxels. Along with these images, a scalar-valued cognitive score is available for all these patients to measure the degree of their language loss due to PPA.

There are two major inferential objectives of the clinical study. First, neuroscientists are often interested in identifying regions of the brain significantly related to the cognitive score (influential regions), to gain more insight into neuro-degeneration pathway due to PPA. Uncertainty quantification regarding the inference on influential regions becomes crucial since the analysis only involves a limited number of PPA patients. Second, it is of practical interest to examine the inferential advantage of exploiting cross-information from multiple modalities over a single modality.

When inference on influential brain regions is of interest, one may fit a regression model with cognitive score as the response and brain network and structural images as predictors, extending the popular literature of scalar-on-object regressions (Guhaniyogi *et al.*, 2017; Fan *et al.*, 2019; Guha and Rodriguez, 2020; Goldsmith *et al.*, 2014; Li *et al.*, 2015; Feng *et al.*, 2019) to scalar-on-multi-object regressions. However, limited sample size in our application precludes drawing satisfactory inference from these models. As an alternative, we address both inferential objectives by formulating a multi-modal regression framework with the brain network (defined over brain ROIs) and structural images (defined over brain voxels) as two sets of responses and the language score as the predictor. We adopt a hierarchical Bayesian modeling approach in fitting the multi-modal regression method since it can naturally share information between different objects through careful construction of a joint prior structure on regression coefficients and quantifies uncertainty in the inference. In particular, we develop a joint prior structure on coefficients of the cognitive score corresponding to the structural and network objects (hereon referred to as the structural and network coefficients, respectively) to account for both sets of topology in multi-object data and to allow the information in separate image objects to complement and reinforce each other in their relation to the scalar predictor to support our inferential objectives. To elaborate on it, we begin with a common brain atlas for both image objects to ensure that it provides an organizing principle that links together structural and network information via a shared set of ROIs and a group of voxels within each ROI. The joint prior on network and structural coefficients are

constructed respecting the *hierarchical* constraint in multi-object topology which ensures all voxels within an ROI is uninfluential if the ROI is un-influential (see Section 3.2). While this article does not explicitly make use of the structural information in the GM images by careful spatial modeling of GM image coefficient, it partially exploits the structural information by imposing the *hierarchical* constraint in the prior construction step. The problem of identifying influential ROIs is cast under a nonlinear variable selection framework, wherein binary latent indicator corresponding to all ROIs are shared among both sets of coefficients to enforce that all voxels from a particular ROI and all network edges connected to that particular ROI have no relationship with the predictor when the binary latent indicator corresponding to the ROI is zero. As a byproduct of our construction, *symmetry* and *transitivity* property of the un-directed network object is preserved, as discussed in Section 3.2. The prior construction achieves efficient computation, identifies influential ROIs which are key to study neuronal atrophy, and produces well-calibrated interval estimates for the multi-modal regression coefficients. Moreover, our framework attaches uncertainty in identifying these ROIs and offers improved inference over regression methods with a single imaging modality.

There is a dearth of principled Bayesian literature addressing the inferential objectives of the motivating application, and our proposal is arguably the first Bayesian multi-object regression approach to answer the inferential questions stated before. We now provide a brief overview of the available literature to contrast them with our proposal. In the course to determine the association between network or structural objects and the cognitive score, the most popular approach estimates the association between each network edge or GM voxel and the cognitive score independently, providing a p-value "map" (Friston, 2003; Lazar, 2008, 2016). The p-values can be adjusted for multiple comparisons to identify "significant" ROIs and voxels. These approaches have been mainly used for regressions with a single-object response, but can be suitably extended to ascertain the hierarchical constraint between brain ROIs and voxels, as we discuss and implement in Section 5. Although appealing due to their simplicity, such independent screening approaches have key disadvantages relative to methods that take into account the joint impact of all network edges and voxels of the GM images simultaneously, as demonstrated in Section 5.

To identify joint association of brain network matrix and GM images with the cognitive

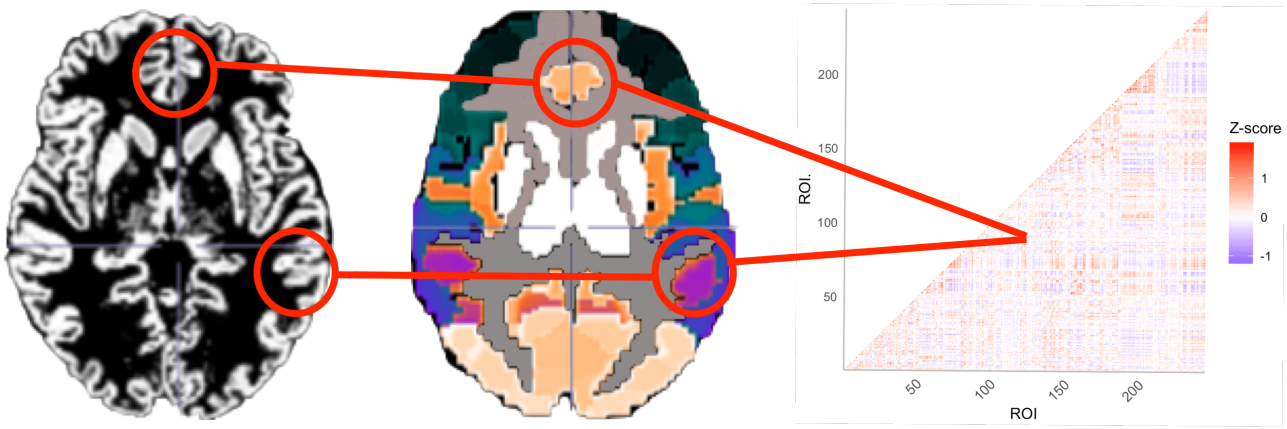
score, one can possibly proceed to vectorize both objects and regress them jointly on the cognitive score, leading to a high dimensional vector regression problem. This approach can take advantage of the recent developments in high dimensional multivariate reduced rank sparse regression literature, consisting of both penalized optimization (Yuan *et al.*, 2007; Rothman *et al.*, 2010; Chen and Huang, 2012) and Bayesian shrinkage (Goh *et al.*, 2017). However, they treat the edges of the network coefficients as if they were fully exchangeable, ignoring the fact that coefficients that involve common network nodes can be expected to be correlated a priori. Additionally, vectorization of both sets of object responses during analysis ignores their individual topology (e.g., the symmetry and transitivity in the network) and linkage between the topology, e.g., the hierarchical constraint. Both these issues may lead to inadequate inference on predictor coefficients. Moreover, this architecture does not allow identification of influential ROIs with uncertainty.

Although the literature on multi-object response regression is extremely sparse, there have been some recent efforts to build regression models with a single-object response, mainly for the functional and structural neuro-imaging data. For example, Guhaniyogi and Spencer (2021); Spencer *et al.* (2020) formulate a Bayesian approach with a tensor response and scalar predictors to jointly identify activated brain regions due to a task, and connectivity between different brain regions from the fMRI data. In the same vein, Roy *et al.* (2019); Zeng *et al.* (2022) develop regression with a structural image object and scalar predictors allowing identification of brain voxels significantly associated with a predictor. A recent Bayesian approach presents a network on scalar regression for the diffusion tensor imaging (DTI) data with a tool to identify network nodes (brain ROIs) significantly associated with creativity (Guha and Guhaniyogi, 2021). While these regression approaches establish the importance of preserving the structure and/or network topology in image objects for better inference, the referenced works mainly regard object topology for a single image object but principled linkages among image objects are not made (e.g., through the hierarchical constraint) and thus inference on scalar outcomes is made without regard to valuable information that are shared across these objects. For instance, in our clinical case study of language dysfunction in PPA patients, existing methods (Mandelli *et al.*, 2014, 2016; Gorno-Tempini *et al.*, 2011, 2008) do not directly combine structural information relating to neuronal atrophy with network

information on brain connectivity to jointly model deficits in language comprehension scores. Failure to consider the structure and cross information from multiple images have generally a negative impact on ND research in terms of lower detection power (Li *et al.*, 2018), bias in estimated effects (Dai and Li, 2021), statistical inefficiency (Dai and Li, 2021), and sensitivity of results to noise (Calhoun and Sui, 2016). Additionally, the existing approaches involving network objects consider low-dimensional structure for the network coefficient, while the approach we propose does not rely on such restrictive assumptions.

Our proposed approach is considerably different from the existing statistical literature on multi-modal data integration. In particular, there have been a class of unsupervised multi-modal analysis built on matrix or tensor factorization (Lock *et al.*, 2013), or methods exploiting structural connectivity information from diffusion tensor imaging (DTI) in the prior construction for the functional connectivity analysis from functional MRI (fMRI) data (Xue *et al.*, 2015). In contrast, we focus on the supervised analysis with multi-modal image and a scalar predictor. To this end, Xue *et al.* (2018) proposes regression on disease status on low-frequency fluctuation (fALFF) from resting-state fMRI scans, voxel based morphometry (VBM) from T1-weighted MRI scans, and fractional anisotropy (FA) from DTI scans. In the same vein Li and Li (2021) develops a factor analysis-based linear regression model, and Dai and Li (2021) extends this framework to account for non-linear association between a scalar response and multi-modal predictors. While these supervised approaches do form linkages among image modalities, they do not properly model within image correlations and thus are not able to address our inferential goals of jointly modeling information across images while maintaining within image topology. Moreover, all these approaches do not naturally offer identification of influential ROIs with uncertainty.

The rest of the manuscript proceeds as follows. Section 2 provides a description of the multi-modal data we analyze in this article. Section 3 describes the model development and prior framework to draw inference from multi-modal images and Section 4 discusses posterior computation of the proposed model. Empirical investigations with data generated under various simulation settings are reported in Section 5. Section 6 analyzes the multi-modal dataset, offering scientific findings on influential ROIs. Finally, Section 7 summarizes the idea laid out in this article and highlights some of the extensions of our model to be



(a) Structural image (b) Atlas (c) Network image

Figure 1: Schematic of the multi-object brain imaging data structure for a PPA patient. (a) Structural image encoding voxel-level gray matter (GM) probability, (b) Brainnetome atlas parcellation of the brain into anatomical ROIs, (c) Network image obtained by calculating the pairwise Pearson correlation Z-score for the average fMRI signal in each ROI. Red circles and lines connect (a) structural and (c) network information from images via the (b) parcellated atlas. Thus, the atlas provides an organizing hierarchy that links together structural information (GM) at the voxel level with network information indexed by pairs of ROIs (fMRI).

explored in the near future.

2 Clinical Case Study on Nonfluent Primary Progressive Aphasia (nfvPPA) Neurodegenerative Disorder

We focus on a clinical application derived from multi-modal image studies conducted on patients with PPA, a form of ND characterized by damage to the frontal and/or temporal lobes leading to loss of language ability. Our interest lies in the nonfluent/aggramatic variant of PPA (nfvPPA) characterized by motor speech and grammar loss and left inferior frontal atrophy (Gorno-Tempini *et al.*, 2008). To investigate the neural underpinnings of disruption to motor speech/fluency in nfvPPA patients, clinical images from multiple modalities was collected as detailed below.

Clinical images: Imaging data is acquired on 26 nfvPPA patients during the course of clinical research activity. Data is collected from the following imaging modalities: sMRI derived gray matter (GM) **(Figure 1a)** which measures the likelihood a voxel containing neuronal cell bodies; and task-free resting state functional magnetic resonance imaging (fMRI) to measure brain activation via neuronal oxygen consumption in subjects at rest. All images

are registered to the Montreal Neurological Institute (MNI) template space with voxels parcellated into 246 ROIs using the Brainnetome atlas such that images across modalities and subjects can be directly compared and each voxel is nested in an anatomically defined ROI **(Figure 1b)** (Fan *et al.*, 2016). For each subject, a 'brain network' represented by a symmetric adjacency matrix is obtained from the fMRI image by considering rows and columns of this matrix corresponding to different ROIs and entries corresponding to the Z-scores obtained by transforming the Pearson correlation between average fMRI data of two ROIs **(Figure 1c)**.

Language loss in nfvPPA patients is driven by neurodegeneration in the *left inferior frontal* region but the dual role of structural damage and brain connectivity in language loss is not well characterized (Mandelli *et al.*, 2016). To better understand the neural underpinnings of language dysfunction in 26 nfvPPA patients, measures of language deficiency must be associated with sophisticated multi-modal images, specifically GM maps which capture focal neurodegeneration, and, fMRI brain connectivity networks which capture disruptions of brain connectivity. We focus our analysis on pause in speech time (PTIS), a time measure of the silent interval between phonations which captures the fluidity of speech. This measure is automatically extracted from recorded speech via SALT software (https://www.saltsoftware.com/), a software platform used to automatically extract language features from recorded speech. Our modeling objective includes identifying ROIs influential in related to the language loss by associating the PTIS with multi-modal imaging data. Findings from the prior clinical studies have identified 38 regions of interest (ROIs) which are related to motor speech/fluency, known as the motor speech/fluency speech production network (SPN) (Mandelli *et al.*, 2016). One of our primary goals to extend the multi-modal study of motor speech/fluency beyond this established SPN to the whole brain. The next section describes a novel regression framework to achieve these scientific goals.

3 Bayesian Regression with Multiple Imaging Responses and Cognitive Score as a Predictor

This section presents model development and prior formulation, including the hyper-parameter specification.

3.1 Model Framework

For the ith subject, let $y_i \in \mathbb{R}$ denote the language score and $\boldsymbol{A}_i$ represent the weighted brain network object. We assume that the network of all subjects are defined on a common set of nodes, with elements of $\boldsymbol{A}_i$ encoding the strength of network connections between different nodes for the i-th subject. In particular, the network object $\boldsymbol{A}_i$ is expressed in the form of a $P \times P$ matrix with the (p, p')-th entry of the matrix $a_{i,(p,p')}$ signifying the strength of association between the pth and p'th node, where $p, p' = 1, ..., P$ and P is the number of network nodes. This paper specifically focuses on networks that contain no self relationship, i.e., $a_{i,(p,p)} \equiv 0$, and are un-directed ($a_{i,(p,p')} = a_{i,(p',p)}$). Such assumptions hold for the data application pertaining to Section 2, where $\boldsymbol{A}_i$ represents the brain connectome network matrix obtained from the fMRI scan, with each node representing a specific brain region of interest (ROI) and edges signify correlations between fMRI signals in two regions. Let $\boldsymbol{g}_{i,1}$,...,$\boldsymbol{g}_{i,P}$ denote the $V_1, ..., V_P$ dimensional structural objects in regions $\mathcal{R}_1$,...,$\mathcal{R}_P$, respectively. In the context of our data application, they represent volumetric elements (voxels) of the GM image from the P ROIs. This multi-object characterization of fMRI and GM data allows structural information across these images to share a common set of ROIs, where each voxel is nested within an ROI. The nested structure of voxels within ROIs provides biologically plausible organization and is instrumental for variable selection and computation as detailed in the upcoming methodological development.

For $i = 1, \ldots, n$, we assume that the relationship between the cognitive score y_i varies with every network edge and every GM voxel and propose a set of conditionally independent generalized linear models for every network edge and GM voxel, given by

$$E[a_{i,(p,p')}] = H_1^{-1}(\theta_{p,p'}\, y_i), \;\; E[g_{i,v,p}] = H_2^{-1}(\beta_{v,p}\, y_i), \;\; v = 1, \ldots, V_p; \;\; p = 1, \ldots, P, \qquad (1)$$

where H_1 and H_2 are the link functions, $\theta_{p,p'}$ is the (p,p')th element of the $P \times P$ matrix $\boldsymbol{\Theta}$ and $\beta_{v,p}$ is the vth element of the V_p dimensional vector of coefficients $\boldsymbol{\beta}_p$. Considering the symmetry and zero diagonal constraint in the network object $\boldsymbol{A}_i$, we set $\theta_{p,p'} = \theta p', p$ and $\theta_{p,p} = 0$, for all $1 \leq p < p' \leq P$. When both sets of responses follow a normal linear model with an identity link function, (1) becomes

$$a_{i,(p,p')} = \theta_{p,p'}\, y_i + e_i^{(p,p')},\;\; g_{i,v,p} = \beta_{v,p}\, y_i + w_i^{(v,p)},\;\; v = 1, \ldots, V_p;\;\; p = 1, \ldots, P, \tag{2}$$

where $e_i^{(p,p')}, w_i^{(v,p)}$ represent the idiosyncratic error in both regression models. While carefully constructed spatial covariance structure can be imposed on $e_i^{(p,p')}$ and $w_i^{(j,p)}$, we relegate it as a future work and offer discussion in Section **??**. Instead, this article focuses on joint learning of the mean structure for two sets of models and assumes $e_i^{(p,p')}, w_i^{(v,p)} \overset{ind.}{\sim} N(0, \tau^2)$ (for $1 \leq p < p' \leq P$ and $v = 1, \ldots, V_p$) for simplicity, following literature on multivariate linear response regression model (Goh *et al.*, 2017). Consistent with the symmetry and zero diagonal entry in $\boldsymbol{A}_i$, we assume $e_i^{(p,p')} = e_i^{(p',p)}$ and $e_i^{(p,p)} = 0$. Therefore, stacking over elements of the network matrix and elements of the GM voxels over each region, (2) can be written as

$$\begin{aligned} \boldsymbol{A}_i &= \boldsymbol{\Theta}\, y_i + \boldsymbol{E}_i \\ \boldsymbol{g}_{i,p} &= \boldsymbol{\beta}_p\, y_i + \boldsymbol{w}_i^{(p)},\;\; p = 1, \ldots, P, \end{aligned} \tag{3}$$

where $\boldsymbol{E}_i \in \mathbb{R}^{P \times P}$ is the symmetric error matrix with zero diagonal entries corresponding to the network object and $\boldsymbol{w}_i^{(p)}$ is the P-dimensional error vector corresponding to the GM image at the pth ROI. The key to joint learning of the multi-modal data lies in the development of a joint prior structure on $\boldsymbol{\Theta}$ and $\boldsymbol{\beta}_p$'s, as described in the next section.

3.2 Prior Distribution on Multi-Modal Coefficients

Our joint prior construction on coefficients $\boldsymbol{\beta}_p$'s and $\{\theta_{p,p'} : p < p'\}$ for multi-modal predictors is fundamental to exploiting topology of the image objects and cross-information among them by forming principled linkages among images at the ROI level. The prior construction is aimed at: (a) identification of influential ROIs with uncertainty; (b) shrinkage

of unimportant voxel coefficients to zero within an influential ROI; and (c) guaranteeing efficient computation of the posterior for the proposed prior. We cast the problem of identifying influential ROIs from the multi-modal images as a high-dimensional variable selection problem and formulate prior distributions on multi-modal object coefficients building upon the existing literature on prior constructions for high-dimensional regression coefficients.

To this end, two classes of prior distributions on coefficients are typically employed in an ordinary high dimensional regression literature. The traditional approach is to develop a discrete mixture of prior distributions (George and McCulloch, 1993, 1997; Scott and Berger, 2010). These methods enjoy the advantage of inducing exact sparsity for a subset of parameters, but may face computational challenges when the number of coefficients is large. As an alternative to this approach, continuous approximation to the discrete mixture priors (Carvalho *et al.*, 2010; Armagan *et al.*, 2013) have emerged which induce approximate sparsity in high-dimensional parameters. Such prior distributions can mostly be expressed as global-local scale mixtures of Gaussians (Polson and Scott, 2010), are computationally efficient and offer an approximation to the operating characteristics of discrete mixture priors.

The direct application of variable selection prior on multi-modal coefficients are unappealing for multiple reasons. First, an ordinary variable selection prior on coefficients $\boldsymbol{\Theta}$ and $\boldsymbol{\beta}_p$'s identifies cells in $\boldsymbol{A}_i$ and $\boldsymbol{g}_{i,p}$ (which in our application refers to the network edges and GM voxels) significantly related to the predictor, rather than identifying influential ROIs. Second, we seek to impose an additional restriction on the prior construction of $\boldsymbol{\Theta}$ motivated by the neuro-scientific application, that is, if any of the pth and p'th ROIs are un-influential in predicting the response, the edge coefficient $\theta_{p,p'}$ corresponding to the edge between pth and p'th nodes is unimportant. This restriction is relevant due to the hierarchical arrangement of voxels and ROIs and is referred to as the *hierarchical constraint*. Finally, we expect the matrix of coefficients $\boldsymbol{\Theta}$ (which itself can be regarded as describing a weighted network) to exhibit *transitivity effects*, that is, we expect that if the interactions between regions p and p' and between regions p' and p'' are both related to the cognitive score, the interaction between regions p and p'' will likely associated with the cognitive score (see, e.g., Li *et al.* (2013)). An ordinary variable selection prior on multi-modal coefficients does not necessarily conform to all these requirements.

We offer a prior construction exploiting the literature on both discrete and continuous mixture variable selection priors to fulfill our inferential goals. To elaborate on it, let $\xi_1, ..., \xi_P$ denote the binary inclusion indicators corresponding to the P ROIs taking values in $\{0, 1\}$, with $\xi_p = 0$ determining no effect of the pth ROI on the response from all covariates. The network edge coefficient $\theta_{p,p'}$ is then endowed with a variable selection prior given by

$$\theta_{p,p'}|\lambda_{p,p'}, \tau^2, \sigma_\theta, \xi_p, \xi_{p'} \overset{ind.}{\sim} \xi_p\xi_{p'}N(0, \tau^2\sigma_\theta^2\lambda_{p,p'}^2) + (1 - \xi_p\xi_{p'})\delta_0, \;\; p < p', \tag{4}$$

where δ_0 corresponds to the Dirac-delta function, $\lambda_{p,p'}$ is a local shrinking parameter corresponding to the (p, p')th edge and σ_θ is the global shrinking parameter for the network coefficient. The prior closely mimics the spike-and-slab variable selection structure with an important difference. While an ordinary spike-and-slab prior introduces a binary inclusion indicator corresponding to each variable, (4) enforces $\theta_{p,p'} = 0$ when either $\xi_p = 0$ or $\xi_{p'} = 0$. Such a formulation is sensible from a network perspective as it implies that the edge connecting two network nodes is insignificant in predicting the response when at least one of the network nodes is not influential. Additionally, the formulation naturally incorporates transitivity effects in the network coefficient $\mathbf{\Theta}$. We further assign half-Cauchy distributions on $\sigma_\theta \sim C^+(0, 1)$ and $\lambda_{p,p'} \overset{ind.}{\sim} C^+(0, 1)$ to complete prior specification on network coefficient. Integrating out σ_θ and $\lambda_{p,p'}$ in (4), $\theta_{p,p'}|\tau^2, \xi_p = 1, \xi_{p'} = 1$ follows the popular horseshoe prior (Carvalho *et al.*, 2010) which offers a flexible prior structure for precise estimation of nonzero network edge coefficients a posteriori.

The GM coefficient $\beta_{v,p}$ in the pth ROI is modeled using $\beta_{v,p} = \xi_p\gamma_{v,p}$, to ensure all voxel coefficients in the pth ROI become unrelated to the predictor if the pth ROI is uninfluential (i.e., $\xi_p = 0$). To estimate voxel level effects in the pth ROI on the predictor, each $\gamma_{v,p}$ is assigned a horseshoe shrinkage prior which takes the following scale-mixture representation,

$$\gamma_{v,p}|\phi_{v,p}, \eta_p^2, \tau^2 \sim N(0, \tau^2\eta_p^2\phi_{v,p}^2), \;\; \phi_{v,p} \overset{i.i.d.}{\sim} C^+(0, 1), \;\; \eta_p \overset{i.i.d.}{\sim} C^+(0, 1), \tag{5}$$

for $v = 1, ..., V_p$; $p = 1, ..., P$. The prior structure (5) induces approximate sparsity in voxel-level GM coefficients $\gamma_{v,p}$ by shrinking the components which are less influential toward zero while retaining the true signals (Polson and Scott, 2010). Finally, the binary inclusion

indicators are assigned Bernoulli prior distribution $\xi_p \overset{i.i.d.}{\sim} Ber(\nu)$ with $\nu \sim Beta(a_\nu, b_\nu)$ to account for multiplicity correction (Scott and Berger, 2010). Notably, an estimate of the posterior probability of the event $\{\xi_p = 1\}$ shows the uncertainty in identifying the pth ROI to be influential. Thus, $P(\xi_p = 1|\text{Data})$ close to 1 or 0 signifies strong evidence in favor of identifying the pth ROI to be active or inactive, repectively. The prior specification is completed by assigning an inverse Gamma $IG(a_\tau, b_\tau)$ prior on the error variance τ^2 and an $IG(a_\theta, b_\theta)$ prior on the error variance σ_θ.

4 Posterior Computation

Although summaries of the posterior distribution cannot be computed in closed form, full conditional distributions for all the parameters are available and mostly correspond to standard families (available in the supplementary materials). Thus, posterior computation can proceed through a Markov chain Monte Carlo algorithm.

The MCMC sampler is run for 5000 iterations, with the first 1000 discarded as burn-in. All posterior inference is based on post-convergence samples suitably thinned. The average effective sample size as a fraction of the total post-convergence iterations averaged over all $\boldsymbol{\Theta}$ and $\boldsymbol{\beta}_p$'s is 3500. All simulation scenarios show an average effective sample size over 3143, indicating fairly uncorrelated post burn-in MCMC samples.

We have implemented our code in R (without using any C++, Fortran, or Python interface) on a cluster computing environment with three interactive analysis servers, 56 cores each with the Dell PE R820: 4x Intel Xeon Sandy Bridge E5-4640 processor, 16GB RAM and 1TB SATA hard drive. Different replications of the model are implemented under a parallel architecture by making use of the packages `doparallel` and `foreach` within R. The computation times of running 5000 MCMC iterations with $P = 100$ and $V_1 = ... = V_P = 50$ is given by 142 min on average across all simulations.

L (suitably thinned) post-convergence MCMC samples $\xi_p^{(1)}, ..., \xi_P^{(L)}$ of the binary indicator ξ_p are used to empirically assess if the pth ROI is significantly associated with the response. In particular, the pth ROI $\mathcal{R}_p$ is related to the response if $\sum_{l=1}^{L} \xi_p^{(l)}/L > t$, for $0 < t < 1$. The ensuing simulation section computes the true positive rates (TPR) and true negative rates (TNR) for $t = 0.5$ to decide which ROIs are influential in predicting the response.

5 Simulation Studies

In this section we compare our proposed approach, referred to as Bayesian multi-object response regression (BMRR), to that of a few representative frequentist and Bayesian competitors in terms of identifying influential regions, and in terms of drawing inference on regression coefficients.

5.1 Data Generation

In all our simulations, we first simulate $y_1, \ldots, y_n$ from N(0,1) and then generate responses $\boldsymbol{A}_i$ and $\boldsymbol{g}_{i,p}$ from model (3) with the true network coefficient $\boldsymbol{\Theta}_t$ and true structural coefficient $\boldsymbol{\beta}_{p,t}$ for the pth ROI. The subscript t indicates the true data generating parameters. In all simulations we set the sample size to $n = 16$ to assess performance of our approach in a limited sample setting which matches the scenario in our application. We also assume equal number of voxels per ROI, i.e., $V_1 = \cdots = V_P = V$.

Simulating true coefficients $\boldsymbol{\Theta}_t$ and $\boldsymbol{\beta}_{p,t}$. To simulate the true coefficients $\boldsymbol{\Theta}_t$ and $\boldsymbol{\beta}_{p,t}$, we first generate binary variables $\xi_{1,t}, ..., \xi_{P,t} \overset{i.i.d.}{\sim} Ber(\nu_t)$ with $\xi_{p,t} = 1$ sets the p-th region influentially related to the scalar predictor. Since $(1 - \nu_t)$ is the probability of a region not being "influential," it is referred to as the node sparsity parameter. For our simulation studies we consider two sparsity levels, $(1-\nu_t) = 0.85$ and $(1-\nu_t) = 0.70$. For each sparsity level, the coefficient corresponding to the edge connecting the p-th and p'-th region is drawn from the following mixture distribution,

$$\theta_{(p,p'),t} | \xi_{p,t}, \xi_{p',t} \sim \xi_{p,t}\xi_{p',t} N(\mu_{\theta,(p,p')}, \sigma^2_\theta) + (1 - \xi_{p,t}\xi_{p',t})\delta_0, \quad \theta_{(p,p'),t} = \theta_{(p',p),t}; \;\; p < p'. \tag{6}$$

(6) ensures that any edge connecting to the p-th region in the network response is unrelated to the predictor if the p-th region is un-influential, i.e., $\xi_{p,t} = 0 \Rightarrow \theta_{(p,p'),t} = 0$ for all $p' \in \{1, .., P\}$. Similarly, corresponding to each un-influential region $\mathcal{R}_p$, the $V \times 1$ dimensional GM coefficient $\boldsymbol{\beta}_{p,t}$ is set at $\mathbf{0}$. When $\xi_{p,t} = 1$, i.e., the p-th region is influential, we randomly choose $\upsilon_t = 0.4$ proportion of cell coefficients in the p-th region to be nonzero and rest are set at zero. These nonzero coefficients within $\boldsymbol{\beta}_{p,t}$ are simulated from $N(\mu_{\beta,(p,p')}, \sigma^2_\beta)$, where the values of $\mu_{\theta,(p,p')}$ and $\mu_{\beta,(p,p')}$ are drawn from $Unif(0.25, 1)$. All simulations fix $\sigma^2_\theta = 1$,

$\sigma_\beta^2 = 1$ and the error variance $\tau_t^2 = 1$.

Finally, for each of the two node sparsity levels, different number of ROIs and different number of voxels within an ROI are considered. Specifically, we consider two cases, (a) $P = 20$ and $V = 10$, and (b) $P = 100$ and $V = 50$. Cases (a) and (b) are referred to as the "small dimensional example" and "high dimensional example." In both these cases we have approximately similar number of parameters to estimate from the network and the structural coefficients.

5.2 Competitors and Metrics of Comparison

The simulated data will be used to assess the performance of: (A) identifying influential regions; (B) estimating the true network coefficient $\boldsymbol{\Theta}_t$ and structural coefficients $\boldsymbol{\beta}_{1,t}, ..., \boldsymbol{\beta}_{P,t}$; and (C) quantifying uncertainty in the point estimation of network and structural coefficients. We construct a series of competitors of BMRR to assess (A)-(C) as described below.

5.2.1 Frequentist competitors

As frequentist competitors, we implement popularly used approaches wherein each network edge and each cell of the structural image is independently regressed on the predictor to obtain p-values corresponding to the point estimates of $\theta_{p,p'}$ and $\beta_{v,p}$, denoted by p-value($\hat{\theta}_{p,p'}$) and p-value($\hat{\beta}_{v,p}$), respectively, for $v = 1, ..., V_p$ and $p = 1, ..., P$. These p-values will be compared to a threshold to declare if a region is influential, after accounting for the multiple comparison issue through Bonferroni correction. Given the structured nature of our problem, we will consider several implementations of this correction, either separately or jointly on the structural and network objects as described below. We set $\alpha_0 = 0.05$ throughout this discussion.

Global and Regional Bonferroni's correction on structural images. The first competitor, referred to as the **Global Bonferroni** on Grey Matter, focuses on $\{\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_P\}$ together and declares a region p to be influential if at least one voxel coefficient in region p turns out to be significant after adjusting for multiplicity, i.e., p-value($\hat{\beta}_{v,p}$) $\leq \frac{\alpha_0}{\sum_{p=1}^P V_p}$ for at least one $v \in \{1, ..., V_p\}$. As discussed in Genovese *et al.* (2002), applying Bonferroni's correction at a regional level may improve performance. Thus we consider a second competitor called **Regional Bonferroni** on Grey Matter, which selects a region p as influential

if p-value$(\hat{\beta}_{v,p}) \leq \frac{\alpha_0}{V_p}$ for at least one $v \in \{1, ..., V_p\}$.

Global and Regional Bonferroni's correction on the network object. Similar to the structural images, we construct a global approach or a regional approach using the network coefficients. Specifically, the **Global Bonferroni** on the network object focuses on $\{\hat{\theta}_{p,p'} : p \neq p'\}$ and identifies region p as influential if at least one edge connecting to this region is influential after the Bonferroni correction on all the network edge coeffcients, i.e., if p-value$(\theta_{p,p'}) \leq \frac{\alpha_0}{P(P-1)/2}$ for at least one $p' \neq p$. In contrast, the **Regional Bonferroni** on the pth region implements the correction only on the edges connecting to the pth network node and identifies region p to be influential if p-value$(\hat{\theta}_{p,p'}) \leq \frac{\alpha_0}{P-1}$ for at least one $p' \neq p$.

Global and Regional Bonferroni's correction on all objects jointly. We also consider global and regional Bonferroni's correction focusing on both sets of coefficients jointly. Following the similar logic as above, **Global Bonferroni** on structural and network objects jointly determines region p to be influential if at least one edge connecting to the region or at least one voxel coefficient in region p is significant, i.e., p-value$(\hat{\theta}_{p,p'}) \leq \frac{\alpha_0}{P(P-1)/2+\sum_{p=1}^{P} V_p}$ for at least one $p' \neq p$ or p-value$(\hat{\beta}_{p,v}) \leq \frac{\alpha_0}{P(P-1)/2+\sum_{p=1}^{P} V_p}$ for at least one $v \in \{1, .., V_p\}$. The **Regional Bonferroni** for objects jointly identifies pth region as influential if p-value$(\hat{\theta}_{p,p'}) \leq \frac{\alpha_0}{(P-1)+V_p}$ for at least one $p' \in \{1, ..., P\}$ or p-value$(\hat{\beta}_{p,v}) \leq \frac{\alpha_0}{(P-1)+V_p}$ for at least one $v \in \{1, .., V_p\}$.

The second group of frequentist competitors are based on the mass univariate analysis (MUA) approach and tries to control the False Discovery Rate (FDR) (Genovese *et al.*, 2002). MUA relies on multiplicity correction for p-values of the coefficients but sets the thresholds in a different way than the above mentioned approaches. Similar to the Bonferroni's correction, we implement this method on the structural objects, the network object, or jointly on both objects.

Global and Regional MUA on the network and structural objects separately. The global implementation of MUA on the Network Matrix proceeds by ordering p-values corresponding to the $P(P-1)/2$ edge coefficients, denoted by p-value$_{(1)}, ...,$ p-value$_{(\frac{P(P-1)}{2})}$. We consider p-value$_{(i^*)}$ to be the threshold where $i^* \in \{1, .., \frac{P(P-1)}{2}\}$ is the largest index such that p-value$_{(i)} \leq \frac{i}{\frac{P(P-1)}{2}}\alpha_0$. A region p is considered influential if p-value$(\hat{\theta}_{p,p'}) \leq$ p-value$_{(i^*)}$ for at least one $p' \neq p$. This approach is referred to as the **Global MUA** approach for

the network object only. The **Regional MUA** for only the network object focuses on p-values for coefficients of edges attached to a node/region. Let p-value$_{p,(1)}, ...,$ p-value$_{p,(P-1)}$ be the ordering of such p-values for region p and i_p^* be the largest index in region p such that p-value$_{p,(i_p^*)} \leq \frac{i_p^*}{P-1}\alpha_0$. **Regional MUA** of the network object selects a region p if p-value$(\hat{\theta}_{p,p'}) \leq \max($p-value$_{(i_p^*)},$ p-value$_{(i_{p'}^*)})$ for at least one $p' \neq p$. **Global MUA** and **Regional MUA** for structural objects are similarly defined.

Global and Regional MUA on the network and structural objects jointly. The **Global MUA** on structural and network objects jointly proceeds by ordering p-value$(\hat{\beta}_{v,p})$ and p-value$(\hat{\theta}_{p,p'})$ for all $v = 1, ..., V_p$, $1 \leq p < p' \leq P$ in ascending order. Let p-value$_{(1)}, ...,$ p-value$_{\left(\frac{P(P-1)}{2}+\sum_{p=1}^{P} V_p\right)}$ be the ordered p-values and let i^* be the largest index such that p-value$_{(i^*)} \leq \frac{i^*\alpha_0}{\frac{P(P-1)}{2}+\sum_{p=1}^{P} V_p}$. A region p is identified as influential if p-value$(\hat{\theta}_{p,p'}) \leq$ p-value$_{(i^*)}$ for at least one $p' \neq p$, or p-value$(\hat{\beta}_{v,p}) \leq$ p-value$_{(i^*)}$ for at least one $v \in \{1, ..., V_p\}$. To implement **Regional MUA** jointly for both objects, the p-values corresponding to all cell coefficients for the pth region and edge coefficients connecting to the pth node/region are ordered in an ascending manner, p-value$_{p,(1)} \leq \ldots \leq$ p-value$_{p,(P-1+V_p)}$. Let i_p^* be the largest index such that p-value$_{p,(i_p^*)} \leq \frac{i_p^*\alpha_0}{P-1+V_p}$. The region p is considered influential if p-value$(\hat{\theta}_{p,p'}) \leq \max($p-value$_{(i_p^*)},$ p-value$_{(i_{p'}^*)})$ for at least one $p' \neq p$, or p-value$(\hat{\beta}_{v,p}) \leq \max($p-value$_{(i_p^*)},$ p-value$_{(i_{p'}^*)})$ for at least one cell $v \in \{1, ..., V_p\}$ in the pth region.

5.2.2 Bayesian competitors

While the frequentist competitors implemented here rely on univariate analysis and adjustment of p-values, we implement two Bayesian competitors, both capturing joint effects of all network edges and cells of the structural images on the predictor, but do not acknowledge the network topology or the connection between the topology of two sets of objects through the hierarchical constraint. The first Bayesian competitor, referred to as the Spike & Slab, applies an ordinary spike & slab prior (George and McCulloch, 1993) on each $\theta_{p,p'}$ and $\beta_{v,p}$. A voxel or an edge will be identified as influential if the posterior probability of corresponding coefficient equals to zero exceeds 0.5 (Barbieri and Berger, 2004). To assess joint modeling of objects vs. modeling them individually, we fit only the network on scalar regression (i.e., the first equation in (3)), or the structural images on scalar regression (i.e.,

the second equation in (3), or both of them jointly. For the network on scalar regression Spike & Slab competitor identifies a region to be influential if at least an edge connecting to that region is influential. Spike & Slab competitor for structural images on scalar regression identifies a region as influential if it has at least one influential voxel in that region. When both objects are used jointly, Spike & Slab identifies an influential region by observing if any voxel in that region or any edge connected to that region is influential.

Our second Bayesian competitor is based on the shrinkage literature in high dimensional regression. More specifically, it applies a Horseshoe shrinkage prior (Carvalho *et al.*, 2010) on each $\theta_{p,p'}$ and $\beta_{v,p}$. Since the Horseshoe prior does not result in exact zeros for the post burn-in iterates of coefficients, we perform a post-processing approach following Li and Pati (2017) to determine which voxel coefficients and network edges are related to the predictor. We identify influential regions using these estimates either on both sets of objects separately, or jointly, following the similar strategy outlined in the Spike & Slab competitor.

Similar to frequentist competitors, none of the Bayesian competitors of BMRR is able to provide uncertainty in identifying a region to be influential which may be crucial for small sample, as we encounter here. Comparison with the two Bayesian competitors will highlight the advantage of exploiting the object topology in BMRR.

5.2.3 Metrics of comparison

To assess performance of the competitors in terms of correctly identifying important regions, we will present True Positive Rate (TPR) and True Negative Rate (TNR), corresponding to all simulation cases. Since there is a natural trade-off between TPR and TNR, we will present a single measure F_1 score that combines the Precision and TPR, to balance the performance between true positives and true negatives for all competitors.

The point estimation of every competitor is assessed using mean squared errors (MSE) of estimating the network coefficient $\boldsymbol{\Theta}_t$ and the structural coefficients $\boldsymbol{\beta}_{1,t}, ..., \boldsymbol{\beta}_{P,t}$. Since both the fitted $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}_t$ are symmetric with zero diagonals, the MSE for the network coefficient is given by $2\sum_{p<p'}(\theta_{p,p',t} - \widehat{\theta}_{p,p'})^2/P(P-1)$, where $\widehat{\theta}_{p,p'}$ is the point estimate of $\theta_{p,p'}$. Similarly, we compute and present MSE for the structural coefficients given by $\sum_{p=1}^{P} ||\boldsymbol{\beta}_{p,t} - \widehat{\boldsymbol{\beta}}_p||^2/VP$, with $\widehat{\boldsymbol{\beta}}_p$ representing the point estimate of $\boldsymbol{\beta}_p$. MSE for both sets of coefficients jointly is

given by $[\sum_{p<p'}(\theta_{p,p',t}-\widehat{\theta}_{p,p'})^2+\sum_{p=1}^{P}||\boldsymbol{\beta}_{p,t}-\widehat{\boldsymbol{\beta}}_p||^2]/(P(P-1)/2+VP)$. The point estimates are taken to be the posterior median for the Bayesian competitors.

While all three Bayesian competitors BMRR, Spike & Slab and Horseshoe provide automatic characterization of uncertainty, the resulting confidence intervals may not have the correct frequentist coverage in high dimensional regressions (Szabó *et al.*, 2015). Thus, in order to assess uncertainty in estimating $\boldsymbol{\Theta}_t$ from Bayesian competitors, we evaluate the length and coverage of 95% credible intervals averaged across coefficients in $\boldsymbol{\Theta}$ and present them for all cases. Similar quantities for $\boldsymbol{\beta}_{p,t}$'s are also presented. All results presented in each simulation scenario are averaged over 500 simulated data.

5.3 Results

The Table 1 shows the results for different competitors in terms of influential region identification. All frequentist competitors identify influential nodes using a two-stage approach. Hence, their performances are evaluated applying the second stage on the structural and network objects jointly or separately, as described in Section 5.2. Bayesian competitors Horseshoe and Spike & Slab are fitted on structural and network objects separately as well as jointly. Hence TPR and TNR for these methods are recorded when they are applied on individual objects or they are applied jointly. In contrast, identification of influential regions is obtained as part of the inference from BMRR. Hence, we just show the TPR and TNR results for BMRR under the "joint object" column in Table 1. A few interesting patterns emerge from the table. First, both Regional MUA and Regional Bonferroni perform considerably better than Global MUA and Global bonferroni, respectively, in terms of TPR. In contrast, the global implementations of these methods perform marginally better than regional implementation in terms of TNR. This is due to the fact that the global implementations of both MUA and Bonferroni are more conservative than the corresponding regional implementations. Second, all frequentist and Bayesian competitors show significantly better TPR when the a method considers both objects jointly as opposed to using them separately. No significant differences are found for TNR in this aspect. Third, the simulation cases under the smaller dimensional example show significantly better performance for all competitors than the higher dimensional examples. Indeed, the higher dimensional examples

Table 1: True Positive Rate (TPR) and True Negative Rate (TNR) for competing methods under smaller dimensional and high dimensional examples. Both smaller and higher dimensional examples include two different node sparsity levels. We compute TPR and TNR separately for grey matter and network matrix, as well as for them together, as explained in Section 5.2. The best performing method under every metric of comparison in each simulation case is boldfaced.

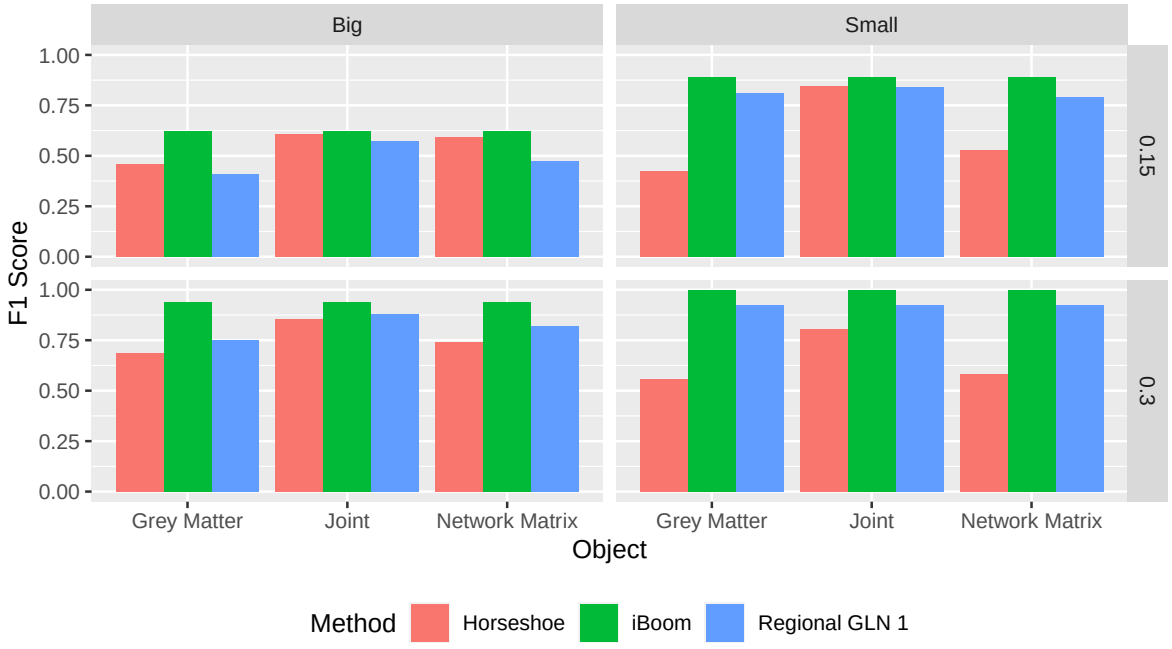
	Sparsity 85%						Sparsity 70%					
	TPR			TNR			TPR			TNR		
Method	Grey Matter	Network Matrix	Joint Objects	Grey Matter	Network Matrix	Joint Objects	Grey Matter	Network Matrix	Joint Objects	Grey Matter	Network Matrix	Joint Objects
High dimensional example												
Global Bonferroni	0.0947	0.1580	0.1633	**0.9985**	**0.9920**	**0.9951**	0.2443	0.3757	0.3920	**0.9950**	**0.9940**	**0.9960**
Regional Bonferroni	0.3667	0.4167	0.5180	0.9359	0.9528	0.9411	0.6743	0.7523	0.8430	0.9389	0.9610	0.9533
Global MUA	0.1313	0.1973	0.2213	0.9693	0.9547	0.9564	0.4030	0.5420	0.6777	0.9557	0.9493	0.9419
Regional MUA	0.3700	0.4447	0.5413	0.9355	0.9355	0.9274	0.6913	0.7810	0.8737	0.9373	0.9440	0.9400
Horseshoe	0.5320	**0.7400**	0.6960	0.7701	0.6956	0.7981	0.7170	0.8613	0.8593	0.8177	0.7249	0.9153
BMRR	–	–	**0.6967**	–	–	0.8666	–	–	**0.9703**	–	–	0.9460
Spike and Slab	0.3040	0.4007	0.5120	0.8866	0.8576	0.8524	0.7267	0.8323	0.9527	0.7727	0.7041	0.6161
Small dimensional example												
Global Bonferroni	0.1339	0.1853	0.1965	**0.9990**	**0.9988**	**0.9991**	0.2517	0.3484	0.3468	0.9991	0.9989	0.9994
Regional Bonferroni	0.8637	0.8393	0.9352	0.9244	0.9452	0.9345	0.9788	0.9759	0.9949	0.9258	0.9362	0.9261
Global MUA	0.4496	0.4285	0.6400	0.9667	0.9814	0.9648	0.9847	0.9787	0.9997	0.8875	0.9082	0.8186
Regional MUA	0.8997	0.8781	0.9663	0.9216	0.9276	0.9244	0.9935	0.9917	0.9995	0.9238	0.9241	0.9184
Horseshoe	0.9947	0.9875	0.9852	0.3739	0.5598	0.9187	1.0000	1.0000	1.0000	0.2855	0.3577	0.7693
BMRR	–	–	**0.9997**	–	–	0.9454	–	–	**1.0000**	–	–	**0.9996**
Spike and Slab	0.9996	0.9960	0.9997	0.0769	0.3616	0.0020	1.0000	1.0000	1.0000	0.0285	0.0960	0.0020

present much more challenging scenarios for the fitted methods, given that the sample size is limited. Fourth, Horseshoe prior applied on joint modeling of objects yield significantly better results than applying Horseshoe prior in the scenario where the two sets of objects are fitted separately. This also serves as a highlighting point of modeling objects jointly to draw better inference. Finally, our proposed approach BMRR demonstrates better TPR and TNR than Horseshoe. Spike & Slab approach performs similar to BMRR in terms of TPR, but massively underperforms in terms of TNR. In contrast, BMRR significantly outperforms frequentist competitors in terms of TPR and shows marginally inferior performance to them. Notably, BMRR is the only approach that identifies influential regions in a principled Bayesian way.

Since there are trade-offs between the TPR and the TNR performances, Figure 2 presents a single performance measure, F1-score for influential region identification. Given that the two most competing methods in terms of the F1-score are Horseshoe and Regional MUA, we show the plots for F1-score corresponding to BMRR, Horseshoe and Regional MUA. All other competitors show worse performance than these three. As expected, the performance of

BMRR along with all other competitors deteriorate for higher dimensional simulation cases. We also observe notable improvement in the performance of Horseshoe, especially in smaller dimensional cases, when it is used for joint model fitting as opposed to using Horseshoe for fitting structural and network object models separately. In contrast, Regional MUA improves only moderately when both set of objects are considered for multiplicity correction, as opposed to employing them separately. BMRR outperforms both its competitors in all simulation cases. The performance gap between BMRR and Horseshoe is more in low sparsity cases than in high sparsity cases.

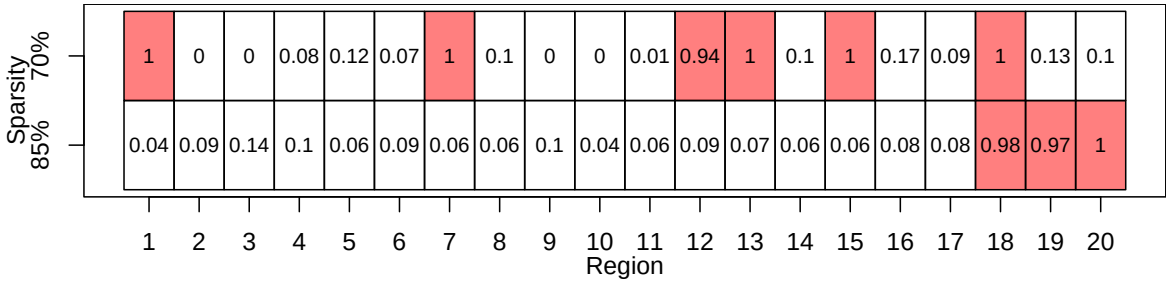
Figure 2: F1 Score for the Horseshoe, the selected MUA approach and BMRR for different simulation scenarios. Here "Big" and "Small" refer to high-dimensional and smaller dimensional examples.



One notable advantage of BMRR over its competitors is that it is able to provide characterization of uncertainty for each region being influential by offering posterior probability of $\{\xi_p = 1\}$. Figure 3 shows the posterior probability of the pth region being influential for small dimensional simulation example under two different node sparsities, with dark cells corresponding to the regions which are truly influential. The results suggest that the posterior probability of a region being influential is close to 1 or 0 if the region is truly influential and un-influential, respectively, suggesting minimal uncertainty in the region selection by

BMRR.

Figure 3: The figure presents posterior probability of each region being influential under the two simulation cases in smaller dimensional example. Each row corresponds to a simulation case. Dark cells corresponds to the truely influential regions in each row.



In terms of estimating the regression coefficients, Figure 4 shows significantly better performance offered by BMRR over its competitors. The significant improvement of point estimation by BMRR demonstrates the importance of exploiting multi-object topology and linkage between multiple objects to offer better point estimation when sample size is limited. The results also demonstrate more accurate point estimation for all competitors for high sparsity cases.

While BMRR enjoys advantage in terms of point estimation, we also set out to investigate how well calibrated the point estimates are. To this end, we observe 95% confidence/credible intervals from BMRR attend close to the nominal level in all simulations, see Figure 5. Regional MUA also enjoys close to nominal coverage. While Horseshoe fitted with both objects jointly yields coverage similar to its competitors, severe under-coverage is observed when Horseshoe is applied on structural and network objects separately. Importantly, BMRR shows the similar coverage with much narrower credible intervals than its competitors, as demonstrated in Figure 6. The 95% credible intervals from Horseshoe fitted on objects jointly are $\sim 2-3$ times wider than BMRR in every simulations. The 95% confidence intervals estimated from Regional MUA are much wider than both its competitors, which is explained by the fact that it estimates every cell of the structural and network coefficient independently of each other.

Figure 4: MSE for the Horseshoe, the selected MUA approach and BMRR for different simulation scenarios. Here "Big" and "Small" refer to high-dimensional and smaller dimensional examples.

6 Application to the nfvPPA Multi-modal Imaging Data

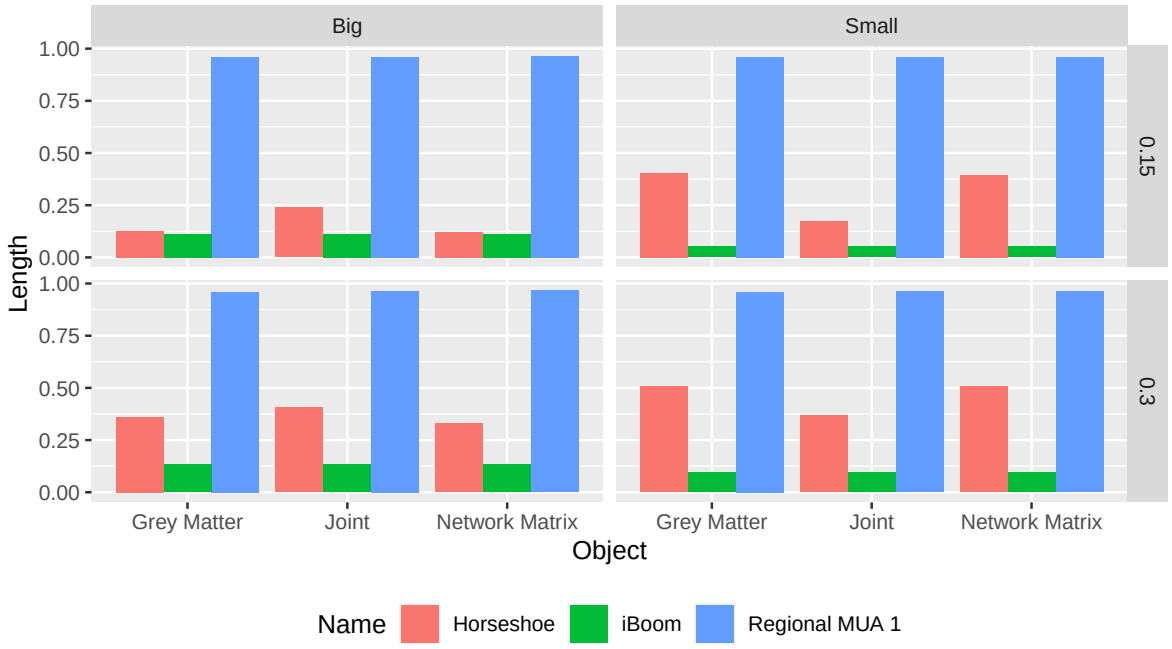
We present the results from our application of the proposed BMRR to the multi-modal imaging data. Recall, the goal of the analysis was to regress the multi-modal imaging data composed of sMRI GM maps and resting state fMRI connectivity networks on the automated measures of pause time in speech (PTIS). In total, 32 ROIs had a posterior probability $P(\xi_p = 1|\text{Data}) > 0.5$, indicating that a subset of the 245 total ROIs have a high posterior probability of association with PTIS in the multi-modal imaging data. Of these 32 ROIs, several were involved in the SPN network described in Section 2 which is previously known to be involved in motor speech and fluency. This included ROIs in the left medial, left and right dorsal, and left inferior frontal sulcus regions of the brain which are also generally associated with motor control and fluid speech. Thus, the BMRR model identifies regions not only consistent with existing hypotheses of brain regions associated with motor speech/fluency but also identifies new regions for further investigation.

Figure 7 provides a graphical representation of the application of the BMRR model to

Figure 5: Coverage of 95% confidence/credible intervals for the selected MUA approach and BMRR for different simulation scenarios. Here "Big" and "Small" refer to high-dimensional and smaller dimensional examples.

the multi-modal imaging data. Figure 7 (a) and (b) display the selection probability for each ROI ($P(\xi_p|\text{Data})$) from a 2D slice of the brain and network edge ($P(\xi_p\xi_{p'} = 1|\text{Data})$) from the whole brain, respectively. The selection probabilities for both the ROI and network edges show little variance, i.e. they tend to include or exclude ROIs and network edges with high certainty. While the BMRR model does not explicitly model spatial information, it is clear from the selection probability maps that there is some spatial continuity in the selected regions (Figure 7 (a)). Figure 7 (d) and (e) show the estimated gray matter voxel coefficients by ROI ($\boldsymbol{\beta}_p$) and network edge coefficients by ROI pair ($\boldsymbol{\Theta}$), respectively. Recall, the BMRR model estimates not only the selection probability for an ROI but also the estimated effect of that ROI, either in the form of voxel level coefficients for structural object or network edge coefficients for the network object. Therefore, even though certain ROIs were included with low probability (Figure 7 (a, b)), BMRR still provides estimates of the associated structural and network coefficients which we can visualize and inspect in conjunction with the selection probability maps - one example of the rich output provided by the BMRR approach to multi-modal image analysis. Specifically, in panels Figure 7 (d, e), the color of the estimated

Figure 6: Length of 95% confidence/credible intervals for the selected MUA approach and BMRR for different simulation scenarios. Here "Big" and "Small" refer to high-dimensional and smaller dimensional examples.



coefficients denotes whether an ROI or network edge have an selection probability greater than 0.5 but the shading denotes the magnitude of the coefficient as a measure of effect size in relation to the PTIS predictor. Thus, we can simultaneously inspect the certainty with which the model selects an ROI or network edge as well as the magnitude of the voxel or edge coefficients included with that ROI or network edge, respectively. Focusing our attention to the blue ROIs and network edges, it is clear there is variability in the effect of PTSI on the multi-modal imaging data both at the voxel level and edge level.

7 Conclusion and Future Work

Motivated by the structural and functional imaging data on patients with PPA, this article develops a regression approach with structural and network-valued objects and a scalar predictor. A novel prior structure is developed jointly on coefficients corresponding to different object responses which can simultaneously exploit topology of these objects and the linked information between the objects to draw inference on network nodes significantly related to the scalar predictor with uncertainty. The proposed approach is arguably the first

Bayesian multi-modal response regression approach equipped with identifying brain regions significantly related to a cognitive score measuring neuro-degeneration due to PPA. The analysis of PPA data leads to important understanding of neuro-degeneration pathway for PPA.

While we exploit the linkage between topology of structural and network objects through the hierarchical constraint, this article does not exploit the spatial correlation between GM image voxels. As an immediate future work, we will extend our approach to incorporate the spatial information in the GM image. We also plan to capture more complex non-linear dependence between network and structural objects and cognitive scores in our upcoming articles.

8 Acknowledgement

Rajarshi Guhaniyogi acknowledges funding from National Science Foundation Grant DMS-2220840 and Office of Naval Research Grant N00014-18-1-274.

References

Armagan, A., Dunson, D. B., and Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica*, **23**(1), 119–143.

Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *The annals of statistics*, **32**(3), 870–897.

Brown, J. A., Deng, J., Neuhaus, J., Sible, I. J., Sias, A. C., Lee, S. E., Kornak, J., Marx, G. A., Karydas, A. M., Spina, S., Grinberg, L. T., Coppola, G., Geschwind, D. H., Kramer, J. H., Gorno-Tempini, M. L., Miller, B. L., Rosen, H. J., and Seeley, W. W. (2019). Patient-tailored, connectivity-based forecasts of spreading brain atrophy. *Neuron*, **104**(5), 856–868.

Calhoun, V. D. and Sui, J. (2016). Multimodal fusion of brain imaging data: A key to finding the missing link (s) in complex mental illness. *Biological psychiatry: cognitive neuroscience and neuroimaging*, **1**(3), 230–244.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97**(2), 465–480.

Chen, L. and Huang, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *Journal of the American Statistical Association*, **107**(500), 1533–1545.

Dai, X. and Li, L. (2021). Orthogonal statistical inference for multimodal data analysis. *arXiv preprint arXiv:2103.07088*.

Fan, J., Gong, W., and Zhu, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of econometrics*, **212**(1), 177–202.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., and Jiang, T. (2016). The human brainnetome atlas: A new brain atlas based on connectional architecture. *Cereb Cortex*, **26**(8), 3508–3526.

Feng, X., Li, T., Song, X., and Zhu, H. (2019). Bayesian scalar on image regression with nonignorable nonresponse. *Journal of the American Statistical Association*, pages 1–24.

Friston, K. J. (2003). Statistical parametric mapping. In *Neuroscience databases*, pages 237–250. Springer.

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, **15**(4), 870–878.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**(423), 881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica sinica*, **7**, 339–373.

Goh, G., Dey, D. K., and Chen, K. (2017). Bayesian sparse reduced rank multivariate regression. *Journal of multivariate analysis*, **157**, 14–28.

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**(1), 46–64.

Gorno-Tempini, M. L., Brambati, S. M., Ginex, V., Ogar, J., Dronkers, N. F., Marcone, A., Perani, D., Garibotto, V., Cappa, S. F., and Miller, B. L. (2008). The logopenic/phonological variant of primary progressive aphasia. *Neurology*, **71**(16), 1227–1234.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., and Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, **76**(11), 1006–1014.

Guha, S. and Guhaniyogi, R. (2021). Bayesian generalized sparse symmetric tensor-on-vector regression. *Technometrics*, **63**(2), 160–170.

Guha, S. and Rodriguez, A. (2020). Bayesian regression with undirected network predictors with an application to brain connectome data. *Journal of the American Statistical Association*, (just-accepted (https://doi.org/10.1080/01621459.2020.1772079)), 1–34.

Guhaniyogi, R. and Spencer, D. (2021). Bayesian tensor response regression with an application to brain activation studies. *Bayesian Analysis*, **16**(4), 1221–1249.

Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017). Bayesian tensor regression. *The Journal of Machine Learning Research*, **18**(1), 2733–2763.

Lazar, N. (2008). *The statistical analysis of functional MRI data*. Springer Science & Business Media.

Lazar, N. A. (2016). Corrections for multiplicity in functional neuroimaging data. *Handbook of Neuroimaging Data Analysis*, page 355.

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.

Li, H. and Pati, D. (2017). Variable selection using shrinkage priors. *Computational Statistics & Data Analysis*, **107**, 107–119.

Li, L., Kang, J., Lockhart, S. N., Adams, J., and Jagust, W. J. (2018). Spatially adaptive varying correlation analysis for multimodal neuroimaging data. *IEEE transactions on medical imaging*, **38**(1), 113–123.

Li, Q. and Li, L. (2021). Integrative factor regression and its inference for multimodal data analysis. *Journal of the American Statistical Association*, pages 1–15.

Li, Y., Qin, Y., Chen, X., and Li, W. (2013). Exploring the functional brain network of alzheimer's disease: based on the computational experiment. *PloS one*, **8**(9), e73186.

Lock, E. F., Hoadley, K. A., Marron, J. S., and Nobel, A. B. (2013). Joint and individual variation explained (jive) for integrated analysis of multiple data types. *The annals of applied statistics*, **7**(1), 523.

Mandelli, M. L., Caverzasi, E., Binney, R. J., Henry, M. L., Lobach, I., Block, N., Amirbekian, B., Dronkers, N., Miller, B. L., Henry, R. G., and Gorno-Tempini, M. L. (2014). Frontal white matter tracts sustaining speech production in primary progressive aphasia. *J Neurosci*, **34**(29), 9754–9767.

Mandelli, M. L., Vilaplana, E., Brown, J. A., Hubbard, H. I., Binney, R. J., Attygalle, S., Santos-Santos, M. A., Miller, Z. A., Pakvasa, M., Henry, M. L., Rosen, H. J., Henry, R. G., Rabinovici, G. D., Miller, B. L., Seeley, W. W., and Gorno-Tempini, M. L. (2016). Healthy brain connectivity predicts atrophy progression in non-fluent variant of primary progressive aphasia. *Brain*, **139**(Pt 10), 2778–2791.

Ossenkoppele, R., Schonhaut, D. R., Schöll, M., Lockhart, S. N., Ayakta, N., Baker, S. L., O'Neil, J. P., Janabi, M., Lazaris, A., Cantwell, A., Vogel, J., Santos, M., Miller, Z. A.,

Bettcher, B. M., Vossel, K. A., Kramer, J. H., Gorno-Tempini, M. L., Miller, B. L., Jagust, W. J., and Rabinovici, G. D. (2016). Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain*, **139**(Pt 5), 1551–1567.

Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian Statistics*, **9**, 501–538.

Rothman, A. J., Levina, E., and Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, **19**(4), 947–962.

Roy, A., Ghosal, S., Prescott, J., and Choudhury, K. R. (2019). Bayesian modeling of the structural connectome for studying alzheimer's disease. *The Annals of Applied Statistics*, **13**(3), 1791–1816.

Scott, J. G. and Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, **38**(5), 2587–2619.

Spencer, D., Guhaniyogi, R., and Prado, R. (2020). Joint Bayesian estimation of voxel activation and interregional connectivity in fMRI experiments. *UCSC Technical Report*.

Sui, J., Adali, T., Yu, Q., Chen, J., and Calhoun, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of neuroscience methods*, **204**(1), 68–81.

Szabó, B., Van Der Vaart, A. W., and van Zanten, J. (2015). Frequentist coverage of adaptive nonparametric bayesian credible sets. *The Annals of Statistics*, **43**(4), 1391–1428.

Xue, W., Bowman, F. D., Pileggi, A. V., and Mayer, A. R. (2015). A multimodal approach for determining brain networks by jointly modeling functional and structural connectivity. *Frontiers in computational neuroscience*, **9**, 22.

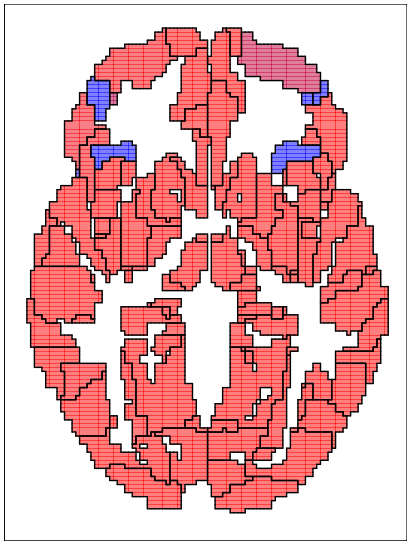
Xue, W., Bowman, F. D., and Kang, J. (2018). A bayesian spatial model to predict disease status using imaging data from various modalities. *Frontiers in neuroscience*, **12**, 184.

Yuan, M., Ekici, A., Lu, Z., and Monteiro, R. (2007). Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**(3), 329–346.

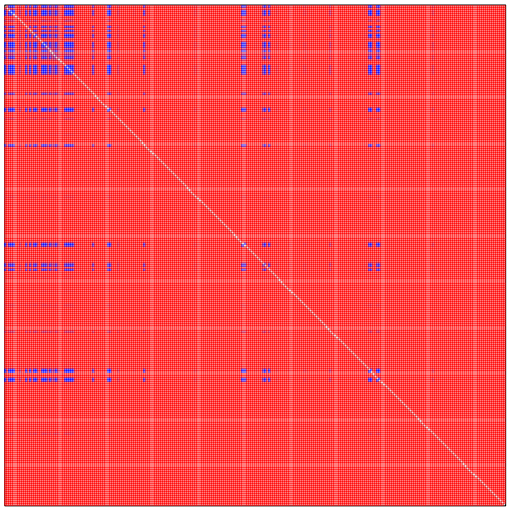
Zeng, Z., Li, M., and Vannucci, M. (2022). Bayesian image-on-scalar regression with a spatial global-local spike-and-slab prior.

Figure 7: Visual presentation of the estimated coefficients from the application of the BMRR to the multi-modal imaging data. Bluer tones indicate a higher selection probability for a region or network edge where applicable (i.e. panels a and b). (a) Selection probability by ROI ($P(\xi_p = 1|\text{Data})$) from a 2D slice of the brain. (b) Selection probability by network edge ($P(\xi_p\xi_{p'} = 1|\text{Data})$). (c) Selection probability color coding. (d) Gray matter voxel coefficients by ROI ($\boldsymbol{\beta}_p$) where shading corresponds to the magnitude of the coefficient. (e) Network edge coefficients by ROI ($\boldsymbol{\Theta}$) where shading corresponds to the magnitude of the coefficient.

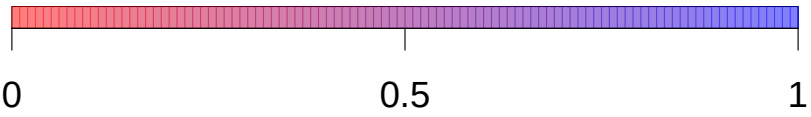
(a) Selection probability by ROI, $P(\xi_p = 1|\text{Data})$ (2D slice)



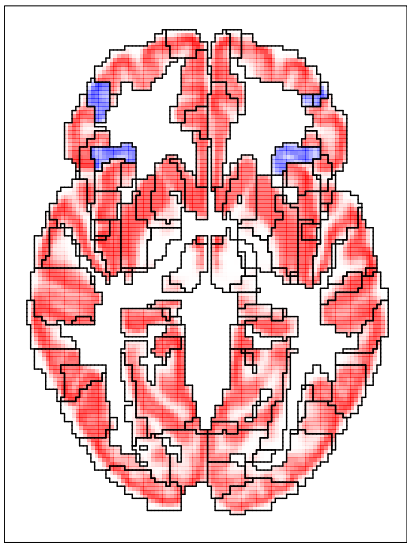
(b) Selection probability by network edge, $P(\xi_p\xi_{p'} = 1|\text{Data})$



(c) Selection probability color coding



(d) Gray matter voxel coefficients by ROI, ($\boldsymbol{\beta}_p$)



(e) Network edge coefficients by ROI, ($\boldsymbol{\Theta}$)