

## Statistica Sinica Preprint No: SS-2023-0312

<b>Title</b>	Bayesian Inference of Spatially Varying Correlations via the Thresholded Correlation Gaussian Process
<b>Manuscript ID</b>	SS-2023-0312
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202023.0312
<b>Complete List of Authors</b>	Moyan Li, Lexin Li and Jian Kang
<b>Corresponding Authors</b>	Jian Kang
<b>E-mails</b>	<a href="mailto:jiankang@umich.edu">jiankang@umich.edu</a>

# Bayesian Inference of Spatially Varying Correlations via the Thresholded Correlation Gaussian Process

Moyan Li<sup>1</sup>, Lexin Li<sup>2</sup> and Jian Kang<sup>1</sup>

<sup>1</sup>*University of Michigan, Ann Arbor* and <sup>2</sup>*University of California, Berkeley*

**Abstract:** A central question in multimodal neuroimaging analysis is to understand the association between two imaging modalities and to identify brain regions where such an association is statistically significant. In this article, we propose a Bayesian nonparametric spatially varying correlation model to make inference of such regions. We build our model based on the thresholded correlation Gaussian process (TCGP). It ensures piecewise smoothness, sparsity, and jump discontinuity of spatially varying correlations, and is well applicable even when the number of subjects is limited or the signal-to-noise ratio is low. We study the identifiability of our model, establish the large support property, and derive the posterior consistency and selection consistency. We also develop a highly efficient Gibbs sampler and its variant to compute the posterior distribution. We illustrate the method with both simulations and an analysis of functional magnetic resonance imaging data from the Human Connectome Project.

**Key words and phrases:** Bayesian modeling; Gaussian process; Multimodal correlation analysis; Neuroimaging analysis.

## 1. Introduction

Multimodal neuroimaging is now prevailing in neuroscience research, where different types of brain images are collected for a common set of subjects. Common imaging modalities include anatomic magnetic resonance imaging (MRI), resting-state or task-based functional MRI (fMRI), diffusion tensor imaging (DTI), positron emission tomography (PET), among many others. Multimodal neuroimaging analysis aggregates such diverse but often complementary information, consolidates knowledge across different modalities, and produces improved understanding of neurological disorders. See Uludağ and Roebroeck (2014) and Zhu et al. (2023) for some excellent reviews on multimodal neuroimaging analysis.

A central question in multimodal neuroimaging analysis is to understand the association between two imaging modalities and to identify brain regions where such an association is statistically significant. This question is of great scientific interest. For instance, Zhu et al. (2014) surveyed and showed joint analysis of fMRI and DTI reveals important interplays between brain functions and structures. Cavaliere et al. (2018) showed fMRI and PET together improve the characterization of patients with consciousness disorder. Li et al. (2019) jointly analyzed two PET modalities with different nuclear tracers, and identified brain regions where the tau protein and glucose metabolism are strongly correlated to facilitate the understanding of Alzheimer's disease pathology. Harrewijn et al. (2020) studied resting-state and task-based fMRI, and found that functional connectivities during the rest and the dot-probe task are positively correlated, which conforms to and further extends the current

studies of human cognitive behaviors.

In this article, we propose a Bayesian nonparametric spatially varying correlation model to address the question of estimation and inference of spatial regions where two imaging modalities are significantly correlated. We build our model based on the thresholded correlation Gaussian process (TCGP), which ensures piecewise smoothness, sparsity, as well as jump discontinuity of spatially varying correlations, and works well even when the number of subjects is limited or the signal-to-noise ratio is small. We study the identifiability of our model, establish the large support property, and derive the posterior consistency and selection consistency. We derive the full conditional distributions, propose a Gibbs sampling algorithm that is highly efficient, and propose a hybrid mini-batch MCMC to further improve the computational efficiency. We apply our proposed method to jointly analyze the resting-state and working memory task-based fMRIs from a study of the Human Connectome Project, and identify a number of scientifically meaningful brain regions that offer useful insights for cognitive neuroscience research.

Our proposal is related to but also substantially different from the existing literature on multimodal correlation analysis and Bayesian inference.

For multimodal correlation analysis, there are, broadly speaking, three categories of solutions. The first category is voxel-wise analysis, which estimates the correlation at each voxel separately, then conducts massive voxel-wise significance tests with false discovery control. This approach is computationally easy to implement, but it does not incorporate

any spatial or scientific knowledge into statistical inference. Besides, the number of voxels, and thus the number of tests, is huge, whereas the number of subjects in most studies is limited. As a result, voxel-wise analysis often suffers from a particularly low detection power. Although the random field theory has been suggested for multiple testing correction so to improve voxel-wise analysis (Worsley et al., 2004), it does not fully address the low power issue, and is also not directly applicable in our problem due to the complex structure of spatially varying correlations. The second category is region-wise analysis, which first summarizes, usually by averaging, the imaging signals within each brain region defined by some pre-specified brain atlas, then carries the correlation analysis at the region level. Although region-wise analysis generally enjoys a better power than voxel-wise analysis, it is sensitive to the choice of brain atlas. More importantly, the voxels within the same region may not always share the same correlation patterns. Averaging the signals by regions may weaken or cancel out significant correlations. The third category merges voxel-wise and region-wise analysis. In particular, Li et al. (2019) adapted the spatially varying coefficient model, which is widely used in neuroimaging analysis but generally for a different purpose (e.g., Zhu et al., 2014; Li et al., 2017, 2021), to the problem of multimodal correlation analysis. They proposed a multi-step procedure, which first fits a spatially varying coefficient model and obtains a smoothed correlation estimate at the voxel level, then applies a graph clustering algorithm to partition the brain into regions with homogeneous correlations, and finally carries out a likelihood ratio test at the region level to identify the regions where two

imaging modalities are significantly correlated. However, this procedure involves multiple tuning parameters, and the testing results may be sensitive to their choices. In addition, due to multiple steps of estimation, it is difficult to establish the theoretical guarantees for the final inference method.

For Bayesian modeling and inference, our proposal also makes a number of useful contributions. First of all, we propose a new Bayesian nonparametric prior, i.e., the thresholded correlation Gaussian process, for spatially varying correlation coefficients that are sparse and piecewise smooth over the space. It is constructed under a Bayesian hierarchical model, by thresholding a Gaussian process of the variances for another two correlated Gaussian processes. Our model targets the second-order correlations between two modalities. Relatedly, Bhattacharya and Dunson (2011) proposed a multiplicative Gamma process shrinkage prior with latent factors to model high-dimensional covariance matrices. Nevertheless, their method places the sparsity on the individual latent factors, whereas we need to impose the sparsity at the voxel level, and the sparsity on the latent factors does not lead to the sparsity on the voxels. In addition, our model hinges on the idea of thresholding a Gaussian process. A similar strategy has been adopted in prior constructions for modeling sparse regressions or spatially varying functions, i.e., either thresholding Gaussian random variables (Nakajima and West, 2013; Ni et al., 2019; Cai et al., 2020), or thresholding Gaussian processes (Kang et al., 2018; Wu et al., 2024). However, none of those priors are readily applicable for Bayesian analysis of spatially varying correlations as in our setting. Recently, there are

several literature focuses on Bayesian brain connectivity analysis. For instance, Chen et al. (2016) introduced a novel Bayesian hierarchical model designed to infer brain connectivity. This model unifies voxel-level and region-level connectivity by acknowledging the distribution of voxel connectivity between regions. Lukemire et al. (2021) developed an integrative Bayesian approach for jointly modeling multiple brain networks, which provides a systematic inferential framework for network comparisons. However, our proposal addresses a completely different problem, which aims to infer spatially varying correlations between two image modalities.

Second, we contribute to posterior computations for Bayesian models with thresholding type priors. Most existing solutions resort to gradient based MCMC algorithms (Roberts and Rosenthal, 1998; Girolami and Calderhead, 2011), where a smooth approximation of the thresholding function is required to get the analytically tractable first derivative (Cai et al., 2020; Wu et al., 2024). There have also been recent advances in developing new sampling algorithms (e.g., Ahn et al., 2012; Chen et al., 2014; Nishimura et al., 2020). However, these algorithms usually converge relatively slowly, and require multiple tuning parameters. By contrast, instead of using a gradient-based MCMC, we derive the full conditional distributions, and propose a Gibbs sampler algorithm that is highly efficient. Besides, the proposed posterior computation algorithm is general, and can be applied to other Bayesian models with thresholding priors as well.

Finally, we are among the first to study the theoretical properties of Bayesian analysis

of spatially varying correlations. Particularly, we show that the proposed TCGP has a large prior support on a wide class of sparse, piecewise smooth, and spatially varying correlation functions. We establish the posterior consistency based upon the foundational work of Choi (2005); Ghosal and Roy (2006); Tokdar and Ghosh (2007). However, it is far from a simple extension, as it involves a two-level Bayesian hierarchical model, multiple Gaussian processes, as well as some thresholding functions. To address these challenges, we propose an equivalent model representation for the transformed data, where the spatially varying correlation coefficients become model parameters that specify the mean of the transformed data. This equivalent formulation substantially simplifies the theoretical analysis in the original model. In light of the sparsity, we further establish the selection consistency of activation regions with nonzero correlation coefficients.

The rest of the article is organized as follows. We develop our spatially varying correlation model in Section 2. We derive the theoretical properties in Section 3, and the Gibbs sampling algorithm in Section 4. We carry out the simulations in Section 5, and analyze the fMRI data in Section 6. We relegate all technical proofs to the Supplementary Material.

## 2. Spatially Varying Correlation Model

In this section, we first propose our Bayesian spatially varying correlation model and the correlation Gaussian process. We then present an equivalent model formulation.



## 2.1 Nonparametric model and correlation Gaussian process

Suppose the observed data consist of  $n$  subjects, each with two imaging modalities. Suppose these two imaging modalities are well aligned in a  $d$ -dimensional compact spatial space  $\mathcal{B} \subset \mathbb{R}^d$ , which is generally true for multimodal neuroimaging. Suppose each image consists of measurements at  $m$  voxel locations  $\mathcal{B}_m = \{v_1, \dots, v_m\} \subseteq \mathcal{B}$ , and we often use  $v, v' \in \mathcal{B}$  to denote some generic voxel locations in  $\mathcal{B}$ . Let  $Y_{1,i}(v)$  and  $Y_{2,i}(v)$  denote the two imaging measures at location  $v$ , for subject  $i = 1, \dots, n$ . We consider the following model:

$$Y_{k,i}(v) = \mu_{k,i}(v) + \varepsilon_{k,i}(v), \quad \varepsilon_{k,i}(v) \sim N(0, \tau_k^2(v)), \quad \text{for } k = 1, 2, \quad (2.1)$$

where  $\mu_{k,i}(v)$  are the spatially varying functions that represent the expected values of  $Y_{k,i}(v)$ ,  $\varepsilon_{k,i}(v)$  are the random noises that are mutually independent over  $k, i, v$ , and follow a normal distribution  $N(\cdot, \cdot)$  with mean zero and variance  $\tau_k^2(v)$ ,  $k = 1, 2$ .

We next propose a novel prior model for  $\mu_{1,i}(v)$  and  $\mu_{2,i}(v)$ :

$$\begin{aligned} \mu_{1,i}(v) &= \eta_{+,i}(v) + \eta_{-,i}(v), \quad \mu_{2,i}(v) = \eta_{+,i}(v) - \eta_{-,i}(v), \quad \text{for } i = 1, \dots, n, \\ \eta_{+,i} &\sim \text{GP}(0, \kappa_+), \quad \eta_{-,i} \sim \text{GP}(0, \kappa_-), \end{aligned} \quad (2.2)$$

where  $\eta_{+,i}$  and  $\eta_{-,i}$  are two independent Gaussian processes given  $\kappa_+$  and  $\kappa_-$ , and  $\eta_{+,i}$  and  $\eta_{-,i}$  capture the positive and negative correlations between the two modalities, respectively.

We further assume that  $\kappa_+$  and  $\kappa_-$  are of the form,

$$\kappa_+(v, v') = \sigma_+(v)\sigma_+(v')\kappa(v, v'), \quad \kappa_-(v, v') = \sigma_-(v)\sigma_-(v')\kappa(v, v'), \quad (2.3)$$

where  $\sigma_+^2(v)$  and  $\sigma_-^2(v)$  are the spatially varying variance functions for  $\eta_{+,i}(v)$  and  $\eta_{-,i}(v)$ , respectively, and  $\kappa(\cdot, \cdot)$  is a stationary correlation kernel function. There are various choices for the kernel  $\kappa(\cdot, \cdot)$ ; for instance, we employ a Matérn kernel in our implementation,

$$\kappa(v, v') = \frac{2^{1-\gamma_1}}{\Gamma(\gamma_1)} \left\{ \sqrt{(2\gamma_1)} \frac{\|v - v'\|}{\gamma_2} \right\}^{\gamma_1} B_{\gamma_1} \left\{ \sqrt{(2\gamma_1)} \frac{\|v - v'\|}{\gamma_2} \right\}, \quad (2.4)$$

where  $\Gamma(\cdot)$  is the gamma function,  $B_{\gamma_1}(\cdot)$  is the modified Bessel function of the second kind, and  $\gamma_1$  and  $\gamma_2$  are two positive hyperparameters that can be determined by the Bayes factor. To impose the sparsity as well as to ensure the identifiability, we require that  $\sigma_+^2(v)\sigma_-^2(v) = 0$ . In other words, only one of the two terms  $\sigma_+^2(v)$  and  $\sigma_-^2(v)$  is nonzero.

Finally, we impose that the variance functions are of the form,

$$\sigma_+(v) = G_\omega\{\xi(v)\}, \quad \sigma_-(v) = G_\omega\{-\xi(v)\}, \quad \xi(v) \sim \text{GP}(0, \kappa), \quad (2.5)$$

where  $G_\omega(x) = xI(x > \omega)$  is a thresholding function with the thresholding parameter  $\omega \geq 0$  and  $I(\cdot)$  the indicator function,  $\xi(v)$  is a spatially varying function that determines both  $\sigma_+(v)$  and  $\sigma_-(v)$  through  $G_\omega(x)(\cdot)$ . As a prior specification, we assume  $\xi(v)$  follows another Gaussian process with mean zero and correlation kernel  $\kappa(\cdot, \cdot)$ , and  $\kappa(\cdot, \cdot)$  is the same as that in (2.3). Note that the construction in (2.5) ensures  $\sigma_+^2(v)$  and  $\sigma_-^2(v)$  are uniquely determined by  $\xi(v)$ , and  $\sigma_+^2(v)\sigma_-^2(v) = 0$ .

Following the prior specifications (2.2) to (2.5), and integrating out  $\mu_{1,i}(v)$  and  $\mu_{2,i}(v)$  in (2.1), we obtain the spatially varying correlation function between  $Y_{1,i}(v)$  and  $Y_{2,i}(v)$  as,

$$\begin{aligned} \rho(v) &= \text{Corr}\{Y_{1,i}(v), Y_{2,i}(v) \mid \xi(v), \tau_1^2(v), \tau_2^2(v)\} \\ &= \frac{G_\omega^2\{\xi(v)\} - G_\omega^2\{-\xi(v)\}}{[G_\omega^2\{\xi(v)\} + G_\omega^2\{-\xi(v)\} + \tau_1^2(v)]^{1/2} [G_\omega^2\{\xi(v)\} + G_\omega^2\{-\xi(v)\} + \tau_2^2(v)]^{1/2}}. \end{aligned} \quad (2.6)$$

We say that  $\rho(v)$  in (2.6) follows a *thresholded correlation Gaussian Process*, as formally defined below.

**Definition 1** Given any nonzero spatially varying variance functions  $\tau_1^2(v)$  and  $\tau_2^2(v)$ , and the thresholding parameter  $\omega \geq 0$ , suppose  $\xi(v) \sim GP(0, \kappa)$ , then  $\rho(v)$  in (2.6) follows a *thresholded correlation Gaussian process*, denoted as  $\rho \sim TCGP(\omega, \kappa, \tau_1^2, \tau_2^2)$ .

Under this construction, with probability one, a correlation Gaussian process is between -1 and 1, and enjoys both piecewise smoothness and sparsity.

We make a few remarks regarding our model setup.

First, our proposed model encompasses a large class of spatially varying functions that are piecewise smooth, sparse, and jump discontinuous, the features that we commonly encounter in neuroimaging data (Zhu et al., 2014). Moreover, instead of specifying a voxel-wise prior, our TCGP incorporates the spatial information of the image, leading to potentially more accurate detection. We note that, our choice to model spatial correlation as a piecewise smoothness function is rooted in the observation that spatially contiguous voxels often exhibit similar correlation patterns (Li et al., 2019). Due to the anatomical and func-

tional connections between different brain regions, neighboring regions often show similar patterns of structural or functional organizations. Consequently, the correlation between two modalities tends to vary smoothly within these contiguous regions, resulting in piecewise smoothness in the correlation image. In addition, the correlation image derived from brain images of two modalities may also display jump discontinuities, attributed to abrupt changes or transitions in the underlying relations between the two modalities. Jump discontinuities frequently manifest at boundaries between different brain regions or tissue types. For instance, abrupt shifts in correlation values may occur at boundaries between grey matter and white matter, or between cortical and subcortical structures. These boundaries signify distinct anatomical or functional transitions, leading to discontinuities in the correlation image.

Second, our prior model focuses on the correlation between the two image modalities, and does not directly specify the correlation between the two modalities at two different voxels, mainly because it is not our primary target of interest. Nevertheless, our prior specifications implicitly take into account the cross-voxel correlations through Gaussian process.

Third, the introduction of  $\xi(v)$  ensures that at least one of  $\sigma_+^2(v)$  and  $\sigma_-^2(v)$  equals 0. If  $\sigma_+^2(v) = 0$  and  $\sigma_-^2(v) \neq 0$ , then  $\eta_+(v) = 0$  and  $\eta_-(v) \neq 0$ , indicating a negative correlation between the two imaging modalities at this voxel. Conversely, if  $\sigma_+^2(v) \neq 0$  and  $\sigma_-^2(v) = 0$ , then a positive correlation exists. Furthermore, if  $-\omega \leq \xi(v) \leq \omega$ ,  $\sigma_+^2(v) = \sigma_-^2(v) = 0$ . In this case,  $\eta_+(v) = \eta_-(v) = 0$ , which imposes sparsity and indicates that there is no significant correlation between these two imaging modalities at this specific location. On

the other hand, if  $\xi(v)$  were not introduced, the spatial correlation is given by:

$$\rho(v) = \frac{\sigma_+^2(v) - \sigma_-^2(v)}{\{\sigma_+^2(v) + \sigma_-^2(v) + \tau_1^2(v)\}^{1/2} \{\sigma_+^2(v) + \sigma_-^2(v) + \tau_2^2(v)\}^{1/2}}.$$

In the instances where  $\rho(v) = 0$ , it is possible for  $\sigma_+^2(v) = \sigma_-^2(v)$  to take very large values, suggesting that two sets of  $\sigma_+^2(v)$  and  $\sigma_-^2(v)$  could yield the same distribution of  $Y_{1i}$  and  $Y_{2i}$ . This unidentifiability of  $\sigma_+^2(v)$  and  $\sigma_-^2(v)$  poses serious challenges for subsequent MCMC convergence. At some locations,  $\sigma_+^2(v)$  and  $\sigma_-^2(v)$  may escalate significantly while  $\rho(v)$  remains zero. Therefore, it is crucial to introduce  $\xi(v)$  to ensure the identifiability.

## 2.2 Equivalent model representation

To facilitate both theoretical investigation and posterior computation, we next derive an equivalent representations of model (2.1) under the prior specifications (2.2) to (2.5).

We first note that, from (2.5) and (2.6),  $\sigma_+(v)$  and  $\sigma_-(v)$  can be uniquely determined by  $\rho(v)$ , in that, given  $\tau_1^2(v)$  and  $\tau_2^2(v)$ ,

$$\begin{aligned}\sigma_+(v) &= G_\omega\{\xi(v)\} = s\{\rho(v); \tau_1^2(v), \tau_2^2(v)\}, \\ \sigma_-(v) &= G_\omega\{-\xi(v)\} = s\{-\rho(v); \tau_1^2(v), \tau_2^2(v)\},\end{aligned}\tag{2.7}$$

where

$$s(x; t_1, t_2) = \left[ \frac{2t_1 t_2}{\{(t_1 - t_2)^2 + 4x^{-2} t_1 t_2\}^{1/2} - (t_1 + t_2)} \right]^{1/2} I(x > 0),$$

for any  $x \in [-1, 1]$ ,  $t_1, t_2 > 0$ . We next consider a transformation of the observed images  $\{Y_{1,i}(v), Y_{2,i}(v)\}$ , the average  $Y_{+,i}(v) = \{Y_{1,i}(v) + Y_{2,i}(v)\}/2$ , and the contrast  $Y_{-,i}(v) = \{Y_{1,i}(v) - Y_{2,i}(v)\}/2$ . Denote

$$E_{+,i}(v) = \frac{\eta_{+,i}(v)}{\sigma_+(v)}, \quad E_{-,i}(v) = \frac{\eta_{-,i}(v)}{\sigma_-(v)}. \quad (2.8)$$

By (2.7), model (2.1) is equivalent to

$$\begin{aligned} Y_{+,i}(v) &= s \left\{ \rho(v); \tau_1^2(v), \tau_2^2(v) \right\} E_{+,i}(v) + \varepsilon_{+,i}(v), \\ Y_{-,i}(v) &= s \left\{ -\rho(v); \tau_1^2(v), \tau_2^2(v) \right\} E_{-,i}(v) + \varepsilon_{-,i}(v), \end{aligned} \quad (2.9)$$

where  $\varepsilon_{+,i}(v)$  and  $\varepsilon_{-,i}(v)$  are random noises that are independent over  $i, v$ , and follow a normal distribution with mean zero and variance  $\{\tau_1^2(v) + \tau_2^2(v)\}/4$ . The covariance between  $\varepsilon_{+,i}(v)$  and  $\varepsilon_{-,i}(v)$  is  $\{\tau_1^2(v) - \tau_2^2(v)\}/\{\tau_1^2(v) + \tau_2^2(v)\}$ .

Following the prior specifications (2.2) to (2.5), we have the equivalent prior specifications for  $E_{+,i}(v)$ ,  $E_{-,i}(v)$ , and  $\rho(v)$  as,

$$E_{+,i} \sim \text{GP}(0, \kappa), \quad E_{-,i} \sim \text{GP}(0, \kappa), \quad \rho \mid \tau_1^2, \tau_2^2 \sim \text{TCGP}(\omega, \kappa, \tau_1^2, \tau_2^2), \quad (2.10)$$

where  $\kappa(\cdot, \cdot)$  is the correlation kernel as specified in (2.3) and (2.5), and in our modeling process, we use the Matérn kernel as specified in (2.4) for  $\kappa(\cdot, \cdot)$ . Figure 1 gives a graphical illustration of our nonparametric Bayesian spatially varying correlation model. In our sub-

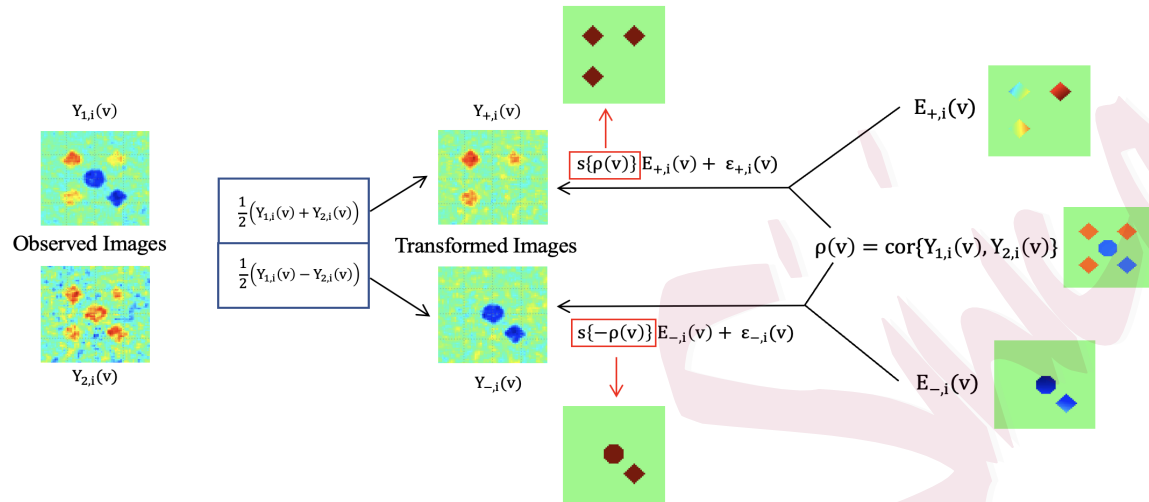


Figure 1: Graphical illustration of the proposed Bayesian spatially varying correlation model. The transformed image  $Y_{\pm,i}(v)$  are modeled based on (2.9).

sequent theoretical and numerical analysis, we focus on the equivalent transformed model.

Before the proposed correlation analysis, we normalize  $\{Y_{k,i}(v)\}_{i=1}^n$  across all the subject for each voxel, thus these two imaging modalities are in the same scale and variation. This ensures the validity of our analysis. In addition, we spatially align the two imaging modalities. The general idea is to map the two images into some common representation space. There are multiple ways to achieve such an alignment, including image registration (Chen et al., 2021; Gholipour et al., 2007), feature-based alignment (Toews et al., 2013; Calhoun and Adali, 2008; Zhu et al., 2014), and machine learning-based alignment (Wang et al., 2021; Liu et al., 2014). In Section 6, as a specific example, we give more details on how we align resting-state and task fMRI images.

### 3. Theory

In this section, we study the model identifiability, derive the large support property, and establish the posterior and selection consistency.

#### 3.1 Notations and definitions

We begin with some notations and definitions. For any vector  $v = (v_1, \dots, v_d)^T \in \mathbb{R}^d$ , let  $\|v\|_p = \left(\sum_{l=1}^d |v_l|^p\right)^{1/p}$  denote the  $L_p$ -norm,  $p \geq 1$ , and  $\|v\|_\infty = \max_{l=1}^d |v_l|$  the supremum norm. For any real function  $f$  on the region  $\mathcal{B}$ , let  $\|f\|_p = \left\{\int_{\mathcal{B}} |f(v)|^p dv\right\}^{1/p}$  denote the  $L_p$ -norm,  $p \geq 1$ , and  $\|f\|_\infty = \sup_{v \in \mathcal{B}} |f(v)|$  be the supremum norm. Suppose  $\mathcal{B}$  is a compact convex set. Recall there are  $n$  subjects, and  $m$  spatial locations for each image. Denote  $Y_{\pm} = \{Y_{\pm,1}^T, \dots, Y_{\pm,n}^T\}^T$ , where  $Y_{\pm,i} = \{Y_{\pm,i}(v_1), \dots, Y_{\pm,i}(v_m)\}^T$ . Furthermore, denote our parameter of interest as  $\theta(\cdot) = \{\rho(\cdot), E_+^T(\cdot), E_-^T(\cdot)\}^T$ , where  $E_{\pm}(\cdot) = \{E_{\pm,1}(\cdot), \dots, E_{\pm,n}(\cdot)\}^T$ , and the true parameter  $\theta_0(\cdot) = \{\rho_0(\cdot), E_{+,0}^T(\cdot), E_{-,0}^T(\cdot)\}^T$ .

**Definition 2** Define  $\mathcal{C}^q(\mathcal{B})$  as a set of differentiable functions of order  $q$  defined on  $\mathcal{B}$ , such that a function  $f \in \mathcal{C}^q(\mathcal{B})$  has the partial derivative,

$$D^b f(v) = \frac{\partial^{\|b\|_1} f}{v_1^{b_1} \dots v_d^{b_d}}(v) = \sum_{\|a\|_1 + \|b\|_1 \leq q} \frac{D^{b+a} f(u)}{a!} (v - u)^a + R_q(v, u),$$

where  $b = (b_1, \dots, b_d)^T \in \mathbb{Z}_+^d$ ,  $a \in \mathbb{Z}_+^d$ ,  $\mathbb{Z}_+$  denotes the set of non-negative integers,  $u \in \mathbb{R}^d$ , and the remainder  $R_q(v, u)$  satisfies the following properties: (i) Given any point  $v_0$  of  $\mathcal{B}$



and any constant  $\epsilon > 0$ , there is a constant  $\delta > 0$ , such that if  $v$  and  $u$  are any two points of  $\mathcal{B}$  with  $\|v - v_0\|_1 < \delta$  and  $\|u - v_0\|_1 < \delta$ , then  $|R_q(v, u)| \leq \|v - u\|_1^{q-\|b\|_1} \epsilon$ ; (ii) If  $\|D^b f\|_\infty \leq C < \infty$  for some constant  $C$ , then  $|R_q(v, u)| \leq (C\|v - u\|_1^{q+1})/(q+1)!$ .

**Definition 3** Define  $\Theta_\rho = \{\rho(v) \in (-1, 1) : v \in \mathcal{B}\}$  as a collection of spatially varying correlation functions that satisfy the following properties: (i) There exist two disjoint non-empty open sets  $\mathcal{R}_{-1}$  and  $\mathcal{R}_1$  with  $\overline{\mathcal{R}}_1 \cap \overline{\mathcal{R}}_{-1} = \emptyset$ , such that  $\rho(v)$  is smooth over  $\overline{\mathcal{R}}_{-1} \cup \overline{\mathcal{R}}_1$ , i.e.,  $\rho(v)I(v \in \overline{\mathcal{R}}_{-1} \cup \overline{\mathcal{R}}_1) \in \mathcal{C}^\alpha(\overline{\mathcal{R}}_{-1} \cup \overline{\mathcal{R}}_1)$ , with  $\alpha = \lceil d/2 \rceil + 1$ , the least integer greater than or equal to  $d/2$ ; (ii)  $\rho(v) = 0$  for  $v \in \mathcal{R}_0$ ,  $\rho(v) > 0$  for  $v \in \mathcal{R}_1$ , and  $\rho(v) < 0$  for  $v \in \mathcal{R}_{-1}$ , where  $\mathcal{R}_0 = \mathcal{B} - (\mathcal{R}_{-1} \cup \mathcal{R}_1)$  and  $\mathcal{R}_0 - (\partial\mathcal{R}_1 \cup \partial\mathcal{R}_{-1}) \neq \emptyset$ ; (iii)  $\rho(v)$  is a discontinuous function and is bounded away from zero for any  $v \notin \mathcal{R}_0$ , i.e.,  $\gamma = \inf_{v \notin \mathcal{R}_0} |\rho(v)| > 0$ .

**Definition 4** Define  $\Theta_E = \{E(v) \in \mathbb{R}^n : \|E(v)\|_2^2 = C_v\}$  for some constant  $C_v < \infty$ .

In summary,  $\Theta_\rho$  is the collection of all piecewise smooth, sparse, and jump discontinuous correlation functions  $\rho(v)$  defined on  $\mathcal{B}$ , where  $\gamma$  in Definition 3 represents the minimum nonzero effect size of the correlation functions that have discontinuity jumps, and  $\Theta_E$  is the collection of the spatially varying functions  $E(v)$  that satisfy the second moment constraints.

### 3.2 Model identifiability and large support

We first show that model (2.9) is identifiable, then show that the prior specification in (2.10) has a large support. We begin with a regularity condition.

**Assumption 1** *The true correlation function  $\rho_0$  is piecewise smooth, sparse, and jump discontinuous, in that  $\rho_0 \in \Theta_\rho$ . In addition, the true functions  $E_{+,0}$  and  $E_{-,0}$  have constant second moments with respect to the location  $v$ , i.e.,  $E_{+,0} \in \Theta_E$  and  $E_{-,0} \in \Theta_E$ .*

Assumption 1 essentially specifies the class of true functions that we target. Denote  $\mathcal{V}(\rho) = \{v : \rho(v) \neq 0\}$ , and  $\mathcal{V}(\rho') = \{v : \rho'(v) \neq 0\}$ . The next proposition shows that model (2.9) is identifiable. Specifically,  $\rho(v)$  is identifiable for all  $v \in \mathcal{B}$ , and  $E_+(v), E_-(v)$  are identifiable for  $v \in \mathcal{V}(\rho) \cup \mathcal{V}(\rho')$ . The identifiability of  $E_+(v)$  and  $E_-(v)$  is constrained on  $\mathcal{V}(\rho) \cup \mathcal{V}(\rho')$  because when  $\rho(v) = 0$ ,  $s\{\rho(v)\} = 0$  in model (2.9).

**Proposition 1** (Identifiability) *Suppose Assumption 1 holds. Then model (2.9) is identifiable. That is, if the probability distributions of  $\{Y_+, Y_-\}$  under  $\theta = \{\rho, E_+^T, E_-^T\}^T$  and  $\theta' = \{\rho', E_+'^T, E_-'^T\}^T$  are equal, then we have  $\rho = \rho'$  for  $v \in \mathcal{B}_m$ , and  $\theta = \theta'$  for  $v \in \mathcal{V}(\rho) \cup \mathcal{V}(\rho')$ .*

To ensure the large-support property, we introduce another condition on the correlation kernel function  $\kappa(\cdot, \cdot)$ . The same condition was imposed in Ghosal and Roy (2006) as well.

**Assumption 2** *The correlation kernel  $\kappa(\cdot, \cdot)$  satisfies that, for any  $v \in \mathcal{B}$ ,  $\kappa(v, \cdot)$  has continuous partial derivatives up to order  $2\alpha + 2$ , where  $\alpha = \lceil d/2 \rceil + 1$ . In addition, suppose  $\kappa(v, v') = \prod_{l=1}^d \kappa_l(v_l - v'_l; \nu_l)$ , for any  $v = (v_1, \dots, v_d)^T$ , and  $v' = (v'_1, \dots, v'_d)^T \in [0, 1]^d$ , where  $\kappa_l(\cdot; \nu_l)$  is a continuous, nowhere zero, symmetric density function on  $\mathbb{R}$  with parameter  $\nu_l \in \mathbb{R}^+$ , for  $l = 1, \dots, d$ .*

The next theorem shows that our prior specification in (2.10) is desirable, in that it has support over a large class of sparse, piecewise smooth and jump discontinuous spatially varying correlation functions. That is, there is a positive probability that  $\theta = \{\rho, E_+^T, E_-^T\}^T$  concentrates on an arbitrarily small neighborhood of any true parameter in the parameter space  $\Theta = \Theta_\rho \times \Theta_E \times \Theta_E$ .

**Theorem 1** (*Large Support*) *Suppose Assumptions 1 and 2 hold. Under the prior specification in (2.10), for any  $\epsilon > 0$ ,  $pr(\|\theta - \theta_0\|_\infty < \epsilon) > 0$ , where  $pr(\cdot)$  denotes a probability measure on the Borel set of  $\Theta$ .*

### 3.3 Posterior consistency

Next, we establish the posterior consistency, then the selection consistency.

**Assumption 3** *There exist constants  $d/(2\alpha) < \nu_0 < 1$ ,  $C_0 > 0$ ,  $C_1 > 0$ , and  $N \geq 1$ , with  $\alpha = \lceil d/2 \rceil + 1$ , such that  $C_0 n^d \leq m \leq C_1 n^{2\alpha\nu_0}$  for all  $n > N$ .*

Assumption 3 imposes that the number of spatial locations  $m$  should be of the polynomial order of the sample size  $n$ . The lower bound indicates that  $m$  needs to be sufficiently large to ensure that the posterior distribution of the spatially varying coefficient function concentrates around the true value. The upper bound ensures that a sufficient amount of information is collected across subjects to identify the population level true parameters.

The next theorem shows that, under the proposed prior, the posterior distribution of  $\theta$

concentrates in an arbitrarily small neighborhood of the true parameter  $\theta_0$ , when the number of subjects  $n$  and the number of spatial locations  $m$  are sufficiently large.

**Theorem 2** (*Posterior Consistency*) Suppose Assumptions 1, 2 and 3 hold. Under model (2.9) and the prior specification in (2.10), for any  $\epsilon > 0$ , as  $m \rightarrow \infty$  and  $n \rightarrow \infty$ ,

$$pr\left(\{\theta \in \Theta : \|\theta - \theta_0\|_1 < \epsilon\} \mid Y_+, Y_-\right) \rightarrow 1 \text{ in } \mathbb{P}_{\theta_0}^{(m,n)}\text{-probability,}$$

where  $\mathbb{P}_{\theta_0}^{(m,n)}$  denotes the distribution of  $\{Y_+, Y_-\}$  given the true parameter  $\theta_0$ , and  $pr(\cdot \mid Y_+, Y_-)$  denotes the posterior probability measure on the Borel set of  $\Theta$  given data  $\{Y_+, Y_-\}$ .

The next theorem shows that, with probability tending to one, we can identify the true activation regions that have positive correlations, negative correlations, and no correlations, respectively, when both  $n$  and  $m$  tend to infinity.

**Theorem 3** Suppose the same conditions in Theorem 2 hold. Then, as  $m \rightarrow \infty$  and  $n \rightarrow \infty$ ,

$$pr\left(\text{sgn}\{\rho(v)\} = \text{sgn}\{\rho_0(v)\}, v \in \mathcal{B} \mid Y_+, Y_-\right) \rightarrow 1 \text{ in } \mathbb{P}_{\theta_0}^{(m,n)}\text{-probability,}$$

where  $\text{sgn}(x) = 1$  if  $x > 0$ ,  $\text{sgn}(x) = -1$  if  $x < 0$  and  $\text{sgn}(0) = 0$ .

## 4. Posterior Computation

In this section, we first adopt the Karhunen-Loève expansion to simplify the model to a finite number of parameters. We next derive the full conditional distributions of the model parameters, and develop an efficient Gibbs sampling algorithm. We also propose a hybrid mini-batch MCMC to further improve the computational efficiency.

### 4.1 Karhunen-Loève approximation

Model (2.9) involves three Gaussian processes, for  $E_{+,i}(v)$ ,  $E_{-,i}(v)$ , and  $\xi(v)$ , respectively, and all hinge on the infinite dimensional correlation kernel function  $\kappa(\cdot, \cdot)$ . We first adopt the usual strategy of Karhunen-Loève expansion to simplify the model to a finite number of parameters. Specifically, consider the spectral decomposition of the kernel function,  $\kappa(v, v') = \sum_{l=1}^{\infty} \lambda_l \psi_l(v) \psi_l(v')$ , where  $\{\lambda_l\}_{l=1}^{\infty}$  are the eigenvalues in descending order, and  $\{\psi_l(v)\}_{l=1}^{\infty}$  are the corresponding orthonormal eigenfunctions. By Mercer's Theorem (Mercer, 1909), we can represent the Gaussian processes in our model by the Karhunen-Loève (KL) expansion,  $E_{+,i}(v) = \sum_{l=1}^{\infty} e_{i,l,+} \psi_l(v)$ ,  $E_{-,i}(v) = \sum_{l=1}^{\infty} e_{i,l,-} \psi_l(v)$  and  $\xi(v) = \sum_{l=1}^{\infty} c_l \psi_l(v)$ , where  $c_l, e_{i,l,\pm}$  are Karhunen-Loève coefficients. We further truncate the above expansions by focusing on the leading  $L$  eigenvalues and eigenfunctions, where  $L$  can be determined following the usual practice of principal components analysis that retains a certain percentage of total variation. That is, we compute the variance percentage as  $R = \sum_{l=1}^L \lambda_l / \sum_{l=1}^{\infty} \lambda_l$ , where  $\lambda_l$  is the  $l$ th largest eigenvalue of the covariance kernel,

and we approximate the denominator  $\sum_{l=1}^{\infty} \lambda_l$  by a summation of truncated series  $\sum_{l=1}^{L'} \lambda_l$ . Following Wu et al. (2024), we choose the smallest  $L$  such that the variance percentage  $R$  exceeds 60%, while we set  $L' = 900$ . In our numerical analyses, we have found such a choice achieves a good balance between computational cost and model fitting performance.

Based on the Karhunen-Loève truncation, model (2.1) becomes,

$$\begin{aligned} Y_{+,i}(v) &= G_{\omega} \left\{ \sum_{l=1}^L c_l \psi_l(v) \right\} \left\{ \sum_{l=1}^L e_{i,l,+} \psi_l(v) \right\} + \varepsilon_{+,i}(v), \\ Y_{-,i}(v) &= G_{\omega} \left\{ - \sum_{l=1}^L c_l \psi_l(v) \right\} \left\{ \sum_{l=1}^L e_{i,l,-} \psi_l(v) \right\} + \varepsilon_{-,i}(v). \end{aligned} \quad (4.1)$$

Recall that  $\mathcal{B}_m = \{v_1, \dots, v_m\}$  denotes the set of locations where the imaging data are observed, and let  $Y = \{Y_{1,i}(v), Y_{2,i}(v), i = 1, \dots, n, v \in \mathcal{B}_m\}$  denote the imaging data observed at the set of voxels in  $\mathcal{B}_m$ . Then all the parameters in our model include:

$$\tilde{\Theta} = \left\{ \{c_l\}_{l=1}^L, \{e_{i,l,+}\}_{l=1}^L, \{e_{i,l,-}\}_{l=1}^L\}_{i=1}^n, \{\tau_1^2(v), \tau_2^2(v)\}_{v \in \mathcal{B}_m}, \omega \right\}. \quad (4.2)$$

We specify their prior distributions as,

$$c_l \sim N(0, \lambda_l), \quad e_{i,l,\pm} \sim N(0, \lambda_l), \quad \tau_1^2(v), \tau_2^2(v) \sim \text{IG}(a_{\tau}, b_{\tau}), \quad \omega \sim U(a_{\omega}, b_{\omega}), \quad (4.3)$$

That is, we impose a normal distribution for the Karhunen-Loève coefficients  $c_l, e_{i,l,\pm}$ , where  $\lambda_l$  is the eigenvalue of the kernel  $\kappa(v, v')$  as specified above. We impose an inverse Gamma

prior for the variance terms  $\tau_1^2(v)$ ,  $\tau_2^2(v)$ , with shape  $a_\tau$  and scale  $b_\tau$ , and we choose some small values for  $a_\tau, b_\tau$ , so that this prior is non-informative. We also impose a uniform prior for the thresholding parameter  $\omega$ , with range from  $a_\omega$  to  $b_\omega$ , and we choose  $a_\omega, b_\omega$  based on the quantiles of  $|\xi(v)|_{v \in \mathcal{B}_m}$ . It is also possible to consider other types of prior for  $\omega$ , e.g., an exponential distribution. Note that the conditional prior for  $\omega$  allows it to be adaptively learnt in a fully Bayesian way in our Gibbs sampling. This is different from the gradient based MCMC methods, which require a smooth approximation of the thresholding function.

## 4.2 Gibbs sampling

We first present a general result that is useful for deriving the full conditional distributions of some of our key parameters. We note that this result is both new and general, and can be applied to deriving the Gibbs sampler for other types of models involving thresholded Gaussian processes.

**Proposition 2** *Consider a random variable  $\theta$ , and two sets of functions  $f_p(\theta) = a_{1p}\theta^2 + a_{2p}\theta + a_{3p}$ , and  $h_k(\theta) = b_{1k}\theta^2 + b_{2k}\theta + b_{3k}$ , where  $a_{lp}, b_{lk}$  are some coefficients,  $l = 1, 2, 3$ ,  $p = 1, \dots, P, k = 1, \dots, K$ . Suppose the density of  $\theta$  is proportional to*

$$\exp \left\{ \sum_{p=1}^P f_p(\theta) I(\theta > L_p) + \sum_{k=1}^K h_k(\theta) I(\theta < U_k) \right\}, \quad (4.4)$$

where  $U_k, L_p$  are some thresholding coefficients. Then,

(i) *If at least one of  $\{a_{1p}, \dots, a_{1P}, b_{1k}, \dots, b_{1K}\}$  is not equal to 0, then  $\theta$  follows a mixture*

*of truncated normal distributions.*

- (ii) *If  $a_{1p} = b_{1k} = a_{2p} = b_{2k} = 0$  for all  $p, k$ , and at least one of  $\{a_{3p}, \dots, a_{3P}, b_{3k}, \dots, b_{3K}\}$  is not equal to 0, then  $\theta$  follows a mixture of uniform distributions.*
- (iii) *If  $a_{1p} = b_{1k} = 0$  for all  $p, k$ , and at least one of  $\{a_{2p}, \dots, a_{2P}, b_{2k}, \dots, b_{2K}\}$  is not equal to 0, then  $\theta$  follows a mixture of exponential distributions.*

We next derive the full conditional distributions of our model parameter  $\tilde{\Theta}$  in (4.2). Specifically, we first derive the full conditionals of  $\{c_l\}_{l=1}^L$  and  $\omega$ , both of which are based on Proposition 2. We then derive the full conditionals of  $\{e_{i,l,\pm}\}_{l=1,i=1}^{L,n}$  and  $\{\tau_1^2(v), \tau_2^2(v)\}_{v \in \mathcal{B}_m}$ , both of which have closed forms thanks to their conjugate priors. Let  $\tilde{\Theta}_{\setminus \theta}$  be the set of parameters in  $\tilde{\Theta}$  but without  $\theta$ .

The full conditional of  $c_l$  is a mixture of truncated normal distributions, as we show in the Supplementary Material, Section S3.2. This is because the density of  $c_l$  is of the form,

$$\pi(c_l | Y, \tilde{\Theta}_{\setminus c_l}) \propto \exp \left( \sum_{\substack{j=1 \\ \psi_l(v_j) > 0}}^m [g_+(c_l; v_j) I\{c_l > T_+(v_j)\} + g_-(c_l; v_j) I\{c_l < T_-(v_j)\}] \right. \\ \left. + \sum_{\substack{j=1 \\ \psi_l(v_j) < 0}}^m [g_+(c_l; v_j) I\{c_l < T_+(v_j)\} + g_-(c_l; v_j) I\{c_l > T_-(v_j)\}] \right).$$

Given the location  $v_j$ ,  $g_{\pm}(c_l; v_j)$  are two quadratic functions of  $c_l$ , and  $T_{\pm}(v_j)$  are two scalars, whose detailed forms are given in the Supplementary Material, Section S3.2. If we set



$f_p(\theta) = g_+(c_l; v_j)$ ,  $h_k(\theta) = g_-(c_l; v_j)$  for those locations  $v_j$  satisfying  $\psi_1(v_j) > 0$ , and set  $f_p(\theta) = g_-(c_l; v_j)$ ,  $h_k(\theta) = g_+(c_l; v_j)$  otherwise, then the density of  $c_l$  satisfies the condition of Proposition 2(i), and thus it follows a mixture of truncated normal distributions.

The full conditional of  $\omega$  is a mixture of uniform distributions, as we show in the Supplementary Material, Section S3.3. This is because the density of  $\omega$  is of the form,

$$\omega \mid Y, \tilde{\Theta}_{\setminus \omega} \sim \exp \left[ \sum_{a_{\omega} < \xi(v_j) < b_{\omega}}^m \sum_{j=1} C_+(v_j) I\{\omega < \xi(v_j)\} + \sum_{a_{\omega} < -\xi(v_j) < b_{\omega}}^m \sum_{j=1} C_-(v_j) I\{\omega < -\xi(v_j)\} \right].$$

Given the location  $v_j$ ,  $\xi(v_j) = \sum_{l=1}^L c_l \psi_l(v_j)$ ,  $C_{\pm}(v_j)$  are two scalars, whose detailed forms are given in the Supplementary Material, Section S3.3. If we set  $h_k(\theta) = C_+(v_j)$  for those  $v_j$  satisfying  $a_{\omega} < \xi(v_j) < b_{\omega}$ , and set  $h_k(\theta) = C_-(v_j)$  for those locations satisfying  $a_{\omega} < -\xi(v_j) < b_{\omega}$ , then the density of  $\omega$  satisfies the condition of Proposition 2(ii), and thus it follows a mixture of uniform distributions. We make two additional remarks. First, we specify the prior of  $\omega$  as  $U(a_{\omega}, b_{\omega})$ , where we choose  $a_{\omega}, b_{\omega}$  to have a non-informative prior. In practice, we may adopt the empirical Bayes idea, by running the Gibbs sampling once with a non-informative prior first, then using the quantile values of the sorted  $\{|\xi(v)|\}_{v \in B_m}$  to refine the range of the uniform distribution for  $\omega$ . The refinement is fully included in the posterior sampling procedure, where  $a_{\omega}$  and  $b_{\omega}$  are adaptively changed based on the value of  $\{|\xi(v)|\}_{v \in B}$  in each iteration. This can further improve the convergence behavior of the algorithm. Second, if we specify the prior of  $\omega$  as an exponential distribution, then we may apply Proposition 2(iii) to obtain the full conditional of  $\omega$ .

We present the derivations of the full conditionals of  $e_{i,l,\pm}$  and  $\tau_k^2(v)$  and summarize the Gibbs sampling for the TCGP in Algorithm S1 in the Supplementary Material, Section S3.

### 4.3 Hybrid mini-batch MCMC

The proposed Gibbs sampler is computationally efficient in general. Meanwhile, the complexity of computing the full conditional of  $c_l$  is  $O(m^2)$ , where  $m$  is the total number of voxels. When  $m$  is large, this step can be expensive. We next propose a hybrid mini-batch MCMC, with two key components, to further improve the computational efficiency.

The first component is to develop an adaptive proposal function in Gibbs sampling. We note that the Gibbs sampler can be viewed as a special case of Metropolis–Hastings, in which the newly proposed state is always accepted with probability one, and the proposal function in Metropolis–Hastings corresponds to the full conditional distribution in the Gibbs sampler. There have been some recent progress developing scalable MCMC methods (Wu et al., 2022). However, those algorithms mainly focus on how to more efficiently evaluate the ratio of the likelihood function at each iteration, instead of focusing on the proposal function. Moreover, their aims are not to perform Bayesian inference from the exact posterior, but rather to exploit the tempered posterior with an efficient MCMC sampler to obtain a better solution from the global optimization.

We propose an adaptive proposal function, by subsampling voxel locations, instead of individual subjects. More specifically, let  $\mathcal{B}_{m_s} \subset \mathcal{B}_m$  denote a random subset of all the

observed locations  $\mathcal{B}_m$ ,  $Y_{m_s}$  the corresponding imaging data observed at those voxels in  $\mathcal{B}_{m_s}$ , and  $m_s = |\mathcal{B}_{m_s}|$  the cardinality of  $\mathcal{B}_{m_s}$ . Recognizing that the Gibbs sampler is a special case of Metropolis–Hastings, the proposal function for the parameter  $\theta \in \tilde{\Theta}$  is  $P\{\theta|\tilde{\Theta}_{\setminus\theta}, Y\}$ , which is the full conditional distribution of  $\theta$ . Instead of using the entire imaging data  $Y$  to derive the full conditional distribution of  $\theta$ , we propose to use a mini-batch of data  $Y_{m_s}$  to obtain the proposal function  $P\{\theta|\tilde{\Theta}_{\setminus\theta}, Y_{m_s}\}$ . The acceptance ratio of  $\theta$  is,

$$\phi(\theta', \theta) = \min \left[ 1, \frac{\prod_{v \notin \mathcal{B}_{m_s}} P\{Y(v)|\theta', \tilde{\Theta}_{\setminus\theta}\}}{\prod_{v \notin \mathcal{B}_{m_s}} P\{Y(v)|\theta, \tilde{\Theta}_{\setminus\theta}\}} \right],$$

whose derivation is given in the Supplementary Material, Section S3.6. In this case, the computational complexity of sampling  $c_l$  is reduced from  $O(m^2)$  to  $O(m_s^2)$ .

The second component is to consider a hybrid version of mini-batch. This is because, when keeping using the mini-batch of voxels during the whole sampling process, the Markov chain may converge to local modes, and may also converge slowly. To overcome these issues, we propose to use the full dataset after, say, every  $T_0$  iterations of using the mini-batch data.

We summarize the hybrid mini-batch MCMC procedure in Algorithm S5 in the Supplementary Material, Section S3. In our implementation, we set  $m_s = m/16$  and  $T_0 = 20$ , which leads to a good empirical performance. We also carry out a sensitivity analysis in the Supplementary Material, Section S4.3, and find that the result is not sensitive to  $m_s$  and  $T_0$ , as long as they are in a reasonable range.

## 5. Simulations

We consider a  $d = 3$  example that mimics the Human Connectome Project data analyzed in Section 6. More specifically, we obtain the posterior means of  $c_l, e_{i,l,\pm}, \omega$  and  $\tau_k^2(v), k = 1, 2$  from our Human Connectome Project data analysis, then generate  $Y_+, Y_-$  following model (4.1). We simulate the noise  $\varepsilon_{k,i}$  from the normal distribution with mean zero and variance  $\zeta_k \tau_k^2(v), k = 1, 2$ .

To apply the proposed method, we employ the Matérn kernel in (2.4) in our data analysis. We set the prior hyperparameters  $a_\tau = b_\tau = 0.001$  to obtain a non-informative prior, and choose  $a_\omega$  and  $b_\omega$  adaptively as the minimum and the maximum of  $|\xi(v)|_{v \in \mathcal{B}_m}$ , respectively, from each iteration. We run the Gibbs sampler for 1000 iterations, with the first 200 iterations as the burn-in. We also run the hybrid mini-batch MCMC for 1200 iterations, with the first 400 iterations as the burn-in. We claim a voxel having a nonzero correlation by simply thresholding the posterior inclusion probability at 0.5, an approach commonly used in Bayesian analysis. We also compare with a number of alternative solutions, including the voxel-wise analysis, the region-wise analysis, and the integrated analysis method of Li et al. (2019) with two different thresholding values, 0.90 and 0.95, following the analysis in Li et al. (2019). We evaluate the performance of each method by the sensitivity, specificity, and false discovery rate.

We first set  $\zeta_k = 1$  which represents the same noise level as the Human Connectome Project data and vary the sample size. We consider two cases where  $n = 500$  and  $n = 900$ .

The image resolution is  $91 \times 109 \times 91$  with  $m = 117,293$  voxels in the brain region. Table 1 reports the results averaged over 100 data replications, and Figure 1 visualizes the result for one data replication when  $n = 900$ . The red, yellow and blue regions represent the true positive, the false negative, and the false positive regions, respectively. We see that our proposed method clearly outperforms the alternative solutions. In particular, the voxel-wise analysis suffers from a low detection power, the region-wise analysis yields a high false discovery rate, and the integrated method of Li et al. (2019) is sensitive to the thresholding parameter. With the 90% threshold, the integrated method enjoys a better sensitivity and specificity, but yields a larger false discovery rate, whereas with the 95% threshold, it can well control the false discovery rate, is not as powerful. We also observe that the proposed method TCGP performs the best across different values of  $n$ . Meanwhile, it maintains a competitive performance even when  $n$  is relatively small. Besides, in Figure 1, we only show the result for the positively correlated regions, as the result for the negative correlated regions is very similar. Finally, we briefly remark on the computational time of the two Gibbs samplers. On a laptop with 2 cores, 3.1GHz clock speed and 8GB memory, the Gibbs sampler algorithm took about 150 minutes for one data replication, while the hybrid algorithm took about 50 minutes, with the mean acceptance ratio around 0.3.

We then consider two noise levels, or equivalently the signal strengths, with  $\zeta_k = 5$  for a weak signal, and  $\zeta_k = 0.5$  for a strong signal when the sample size is set as  $n = 904$  following the Human Connectome Project data. We also conduct 2D simulations to

Table 1: Simulation results of the 3D image example with the varying sample size  $n$ . Reported are the average sensitivity, specificity, and FDR, with standard error in the parenthesis, based on 100 data replications. Six methods are compared: the voxel-wise analysis, the region-wise analysis, the integrated method of Li et al. (2019) with two thresholding values, 0.95 and 0.90, and the proposed Bayesian method with the Gibbs sampler and the hybrid mini-batch MCMC.

Sample size	Method	Positive Correlation			Negative Correlation		
		Sensitivity	Specificity	FDR	Sensitivity	Specificity	FDR
n=500	Voxel-wise	0.023 (0.001)	1.000 (0.000)	0.004 (0.001)	0.041 (0.005)	1.000 (0.000)	0.005 (0.002)
	Region-wise	0.307 (0.003)	0.944 (0.002)	0.330 (0.010)	0.485 (0.003)	0.966 (0.003)	0.398 (0.010)
	Integrated(0.95)	0.623 (0.005)	0.990 (0.005)	0.145 (0.005)	0.877 (0.004)	0.988 (0.004)	0.190 (0.009)
	Integrated(0.90)	0.897 (0.010)	0.950 (0.005)	0.305 (0.008)	0.905 (0.004)	0.948 (0.005)	0.320 (0.010)
	TCGP (Gibbs)	0.903 (0.005)	0.991 (0.001)	0.074 (0.007)	0.911 (0.002)	0.975 (0.003)	0.075 (0.003)
	TCGP (Hybrid)	0.901 (0.002)	0.990 (0.002)	0.076 (0.006)	0.910 (0.004)	0.973 (0.002)	0.081 (0.002)
n=900	Voxel-wise	0.120 (0.005)	1.000 (0.001)	0.002 (0.001)	0.182 (0.001)	1.000 (0.000)	0.002 (0.001)
	Region-wise	0.587 (0.003)	0.901 (0.002)	0.531 (0.010)	0.627 (0.006)	0.824 (0.005)	0.532 (0.003)
	Integrated(0.95)	0.731 (0.005)	0.995 (0.001)	0.020 (0.004)	0.892 (0.005)	0.970 (0.000)	0.092 (0.002)
	Integrated(0.90)	0.938 (0.010)	0.974 (0.003)	0.300 (0.005)	0.917 (0.008)	0.966 (0.002)	0.311 (0.003)
	TCGP (Gibbs)	0.931 (0.002)	0.994 (0.001)	0.047 (0.003)	0.925 (0.001)	0.987 (0.001)	0.058 (0.002)
	TCGP (Hybrid)	0.931 (0.003)	0.994 (0.001)	0.049 (0.003)	0.921 (0.003)	0.977 (0.001)	0.081 (0.002)

further illustrate the superiority of the proposed method. All the results are reported in the Supplementary Material, Section S4.1.

## 6. Analysis of Human Connectome Project Data

In this section, we further illustrate our method with an fMRI dataset from the Human Connectome Project. Our specific goal is to study the association between the resting-state fMRI and the memory task-related fMRI, and identify brain regions where the resting-state and task-related brain activities are strongly associated. This type of analysis is useful, as there has been increasing interest in recent years to predict task-related brain activations from resting-state fMRI (Tavor et al., 2016; Jones et al., 2017; Cohen et al., 2020). It also reveals numerous brain regions and offers useful insights to understand brain activities during rest

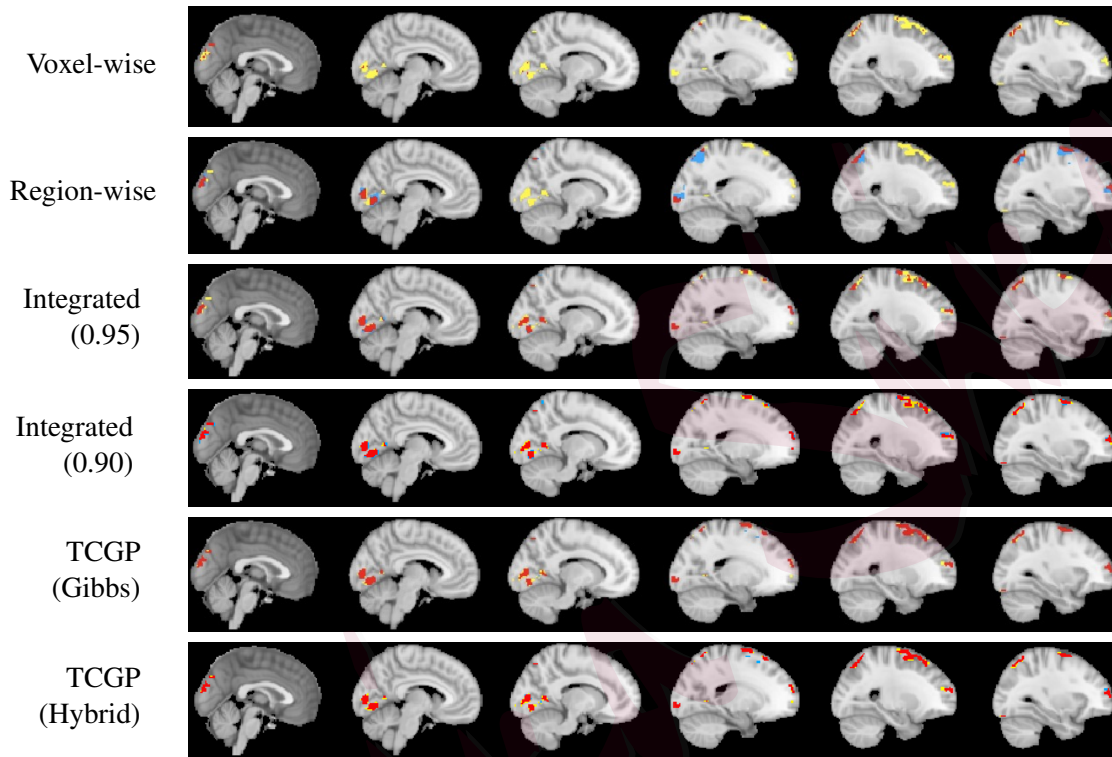


Figure 2: Simulation results of the 3D image example for  $n = 900$  with  $\zeta_k = 1$ . The 2D slices of positively correlated regions identified by the voxel-wise analysis, the region-wise analysis, the integrated method of Li et al. (2019) with two thresholding values, 0.95 and 0.90, and the proposed Bayesian method with the Gibbs sampler and the hybrid mini-batch MCMC. The red, yellow and blue regions represent the true positive, the false negative, and the false positive regions, respectively.

and working memory tasks.

The dataset we analyze consists of  $n = 904$  subjects with both resting-state and task fMRI scans. We preprocess both types of images following the usual pipelines. The preprocessing of resting-state fMRI includes correction for distortions and head motion, removal of slowest temporal drifts and structured non-neuronal artifact (Smith et al., 2013), whereas the preprocessing of task fMRI includes gradient unwarping, motion correction,

distortion correction, and grand-mean intensity normalization (Barch et al., 2013). Moreover, we map both resting-state and task fMRI originally in 4D space onto some common 3D space. Specifically, for resting-state fMRI, we compute the fractional amplitude of low-frequency fluctuations (fALFF) at each voxel (Zou et al., 2008), which quantifies the amplitude of the low frequency oscillations in fMRI signals to reflect the local brain activities during the resting state. For task fMRI, we employ statistical parametric mapping (Penny et al., 2011), by fitting a generalized linear model for the time series at each voxel on the stimulus design matrix convoluted by the hemodynamic function, where the intensity values at individual voxels are the estimated regression coefficients that represent the strength of brain activation in response to the task. For both types of fMRI, the resulting images reside in a 3D spatial space, whereas the temporal dimension has been collapsed. In addition, we regress out potential confounding variables of age and sex using the image-on-scalar approach (Zhu et al., 2014). Finally, we register and align both images to the standard MNI space (Mazziotta et al., 2001), and the resulting image dimensions for both resting-state and task fMRI are  $91 \times 109 \times 91$ , with  $m = 117,293$  voxels in the brain mask.

We apply the proposed methods to this data. We employ the Matérn kernel in (2.4) in our data analysis. We choose the number of leading eigenvalues  $L$  as the smallest value such that the variance percentage exceeds 60% and obtain  $L = 540$ . We set the prior hyperparameters  $a_\tau = 0.001$ ,  $b_\tau = 0.001$ , and choose  $a_\omega$  and  $b_\omega$  as the 75% quantile and 100% quantile of  $\{|\xi(v)|\}_{v \in B}$ , respectively. The choice of  $a_\omega$  is based on the belief that at



Table 2: Results of Human Connectome Project data analysis. Reported are the activation regions containing more than 100 voxels that are declared having a nonzero correlation. AAL refers to the automatic anatomical labeling.

Regions with positive correlations					
AAL Regions	Cluster Size	Activation Center	Mean	Std.	PIP
Precentral-L	385	(45.4, 6.75, 42.95)	0.39	0.07	0.79
Frontal-Sup-R	141	(-25.6, 60.5, 19.2)	0.30	0.05	0.64
Frontal-Sup-R	329	(-26.4, 8.0, 65.2)	0.35	0.08	0.62
Frontal-Mid-L	643	(33.7, 32.9, 42.2)	0.35	0.04	0.59
Frontal-Inf-Tri-R	218	(-51.9, 28.0, 22.6)	0.21	0.06	0.64
Calcarine-R	200	(-13.8, -87.4, 3.64)	0.37	0.05	0.69
Cuneus-L	120	(-0.4, -87.4, 22.6)	0.33	0.04	0.65
Lingual-R	144	(-10.4, -75.3, -4.5)	0.35	0.05	0.62
Parietal-Sup-L	187	(20.0, -67.4, 53.7)	0.35	0.05	0.53
Parietal-Sup-L	108	(26.0, -52.8, 62.4)	0.40	0.06	0.58
Parietal-Sup-R	165	(-29.2, -20.9, 68.3)	0.30	0.08	0.76
Parietal-Inf-L	253	(47.2, -46.1, 49.7)	0.40	0.05	0.59
Angular-R	209	(-46.9, -60.2, 44.7)	0.43	0.03	0.70
Temporal-Sup-L	331	(54.3, -31.8, 18.0)	0.40	0.05	0.82
Temporal-Mid-L	104	(63.1, -25.7, 1.38)	0.41	0.07	0.56
Regions with negative correlations					
AAL Regions	Cluster Size	Activation Center	Mean	Std.	PIP
Precentral-L	115	(28.6, -23.1, 65.4)	-0.44	0.03	0.90
Precentral-R	183	(-54.4, 8.0, 36.0)	-0.40	0.08	0.59
Frontal-Mid-L	191	(28.2, 52.2, 12.7)	-0.39	0.06	0.78
Rolandic-Oper-L	186	(-45.6, -14.5, 15.9)	-0.36	0.14	0.58
Supp-Motor-Area-L	120	(1.1, -7.9, 66.1)	-0.36	0.05	0.71
Supp-Motor-Area-R	143	(-6.9, -13.3, 69.5)	-0.38	0.07	0.85
Calcarine-R	183	(-15.4, -68.8, 10.5)	-0.32	0.06	0.65
Lingual-L	292	(10.5, -75.0, -5.5)	-0.28	0.04	0.79
Lingual-R	286	(-21.2, -86.3, -9.0)	-0.38	0.05	0.80
Occipital-Sup-L	111	(16.0, -89.8, 25.0)	-0.40	0.06	0.80
Occipital-Sup-R	147	(-25.2, -89.9, 26.2)	-0.37	0.04	0.77
Occipital-Inf-R	122	(-38.8, -81.7, -3.2)	-0.44	0.05	0.56
SupraMarginal-L	191	(58.8, -25.7, 30.8)	-0.37	0.04	0.83
SupraMarginal-R	121	(-58.2, -36.9, 28.3)	-0.38	0.03	0.98
Paracentral-Lobule-R	147	(-5.7, -30.5, 70.4)	-0.33	0.04	0.63
Temporal-Mid-L	109	(58.6, -36.3, 7.7)	-0.43	0.05	0.55

most 25% voxels have non-zero correlations. We perform sensitivity analysis on the small changes of prior specification for the threshold parameters in the Supplementary Material, Section S4.4. We run the Gibbs sampler for 1000 iterations, with the first 200 iterations as the burn-in. We also run the hybrid mini-batch MCMC for 1200 iterations, with the first 400 iterations as the burn-in. We again claim a voxel having a nonzero correlation by simply

thresholding the posterior inclusion probability at 0.5.

Table 2 summarizes the estimated activation regions with strong positive or negative correlations. Here we only report those regions containing more than 100 voxels that are declared having a nonzero correlation. We also map those estimated regions to the automatic anatomical labeling (AAL) brain atlas, and report where each estimated activation region is located, the cluster size, the activation center coordinates, the mean and the standard deviation of the correlation in a specific cluster, and the posterior inclusion probability. We make the following observations. We identify a region in angular gyrus that has the highest positive mean correlation. This finding agrees with the literature, as intensive research has shown that angular gyrus is involved in cognitive processes related to language, number processing, spatial cognition, memory retrieval, and attention (Farrer et al., 2008; Seghier, 2013). We also identify a region with strong positive correlations in middle temporal gyrus and superior parietal gyrus. The former region is connected with numerous cognitive processes including recognition of known faces, audio-visual emotional recognition, and accessing word meaning while reading (Acheson and Hagoort, 2013), and the latter is critically involved in information manipulation in working memory (Koenigs et al., 2009). In addition, we identify two regions in lingual gyrus with strong negative correlation, while lingual gyrus is believed to play an important role in visual memory and word processing (Leshikar et al., 2012). Figure 3 shows the identified activation regions with significant correlations.

## Supplementary Material

In the Supplement Material, we first present the proofs of all the theoretical results in the paper, along with a number of useful lemmas. We next derive the full conditional distributions of the model parameters, and present some additional numerical results.

## Acknowledgments

This work was supported by NIH grants R01DA048993, R01MH105561, R01GM124061, and NSF grant IIS2123777. We thank the associate editor and anonymous reviewers for their constructive comments and insightful suggestions, which significantly improved the quality and clarity of the manuscript.

## References

- Acheson, D. J. and P. Hagoort (2013). Stimulating the brain's language network: syntactic ambiguity resolution after tms to the inferior frontal gyrus and middle temporal gyrus. *Journal of cognitive neuroscience* 25(10), 1664–1677.
- Ahn, S., A. Korattikara, and M. Welling (2012, July). Bayesian posterior sampling via stochastic gradient fisher scoring. In J. Langford and J. Pineau (Eds.), *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, ICML '12, New York, NY, USA, pp. 1591–1598. Omnipress.
- Barch, D. M., G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, et al. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80, 169–189.

- Bhattacharya, A. and D. B. Dunson (2011). Sparse bayesian infinite factor models. *Biometrika* 98(2), 291–306.
- Cai, Q., J. Kang, and T. Yu (2020). Bayesian network marker selection via the thresholded graph laplacian gaussian prior. *Bayesian Analysis* 15(1), 79–102.
- Calhoun, V. D. and T. Adali (2008). Feature-based fusion of medical imaging data. *IEEE Transactions on Information Technology in Biomedicine* 13(5), 711–720.
- Cavaliere, C., S. Kandeepan, M. Aiello, D. Ribeiro de Paula, R. Marchitelli, S. Fiorenza, M. Orsini, L. Trojano, O. Masotta, K. St. Lawrence, V. Loreto, B. A. Chronik, E. Nicolai, A. Soddu, and A. Estraneo (2018). Multimodal neuroimaging approach to variability of functional connectivity in disorders of consciousness: A pet/mri pilot study. *Frontiers in Neurology* 9(861), doi: 10.3389/fneur.2018.00861.
- Chen, S., F. D. Bowman, and H. S. Mayberg (2016). A bayesian hierarchical framework for modeling brain connectivity for neuroimaging data. *Biometrics* 72(2), 596–605.
- Chen, T., E. Fox, and C. Guestrin (2014). Stochastic gradient hamiltonian monte carlo. In *International conference on machine learning*, pp. 1683–1691. PMLR.
- Chen, X., A. Diaz-Pinto, N. Ravikumar, and A. F. Frangi (2021). Deep learning in medical image registration. *Progress in Biomedical Engineering* 3(1), 012003.
- Choi, T. (2005). *Posterior consistency in nonparametric regression problems under Gaussian process priors*. Ph. D. thesis, Carnegie Mellon University.
- Cohen, A. D., Z. Chen, O. Parker Jones, C. Niu, and Y. Wang (2020). Regression-based machine-learning approaches to predict task activation using resting-state fMRI. *Human brain mapping* 41(3), 815–826.
- Farrer, C., S. H. Frey, J. D. Van Horn, E. Tunik, D. Turk, S. Inati, and S. T. Grafton (2008). The angular gyrus computes

action awareness representations. *Cerebral cortex* 18(2), 254–261.

Gholipour, A., N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath (2007). Brain functional localization: a survey of image registration techniques. *IEEE transactions on medical imaging* 26(4), 427–451.

Ghosal, S. and A. Roy (2006). Posterior consistency of gaussian process prior for nonparametric binary regression. *The Annals of Statistics* 34(5), 2413–2429.

Girolami, M. and B. Calderhead (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(2), 123–214.

Harrewijn, A., R. Abend, J. Linke, M. A. Brotman, N. A. Fox, E. Leibenluft, A. M. Winkler, and D. S. Pine (2020). Combining fmri during resting state and an attention bias task in children. *Neuroimage* 205, 116301.

Jones, O. P., N. Voets, J. Adcock, R. Stacey, and S. Jbabdi (2017). Resting connectivity predicts task activation in pre-surgical populations. *NeuroImage: Clinical* 13, 378–385.

Kang, J., B. J. Reich, and A.-M. Staicu (2018). Scalar-on-image regression via the soft-thresholded gaussian process. *Biometrika* 105(1), 165–184.

Koenigs, M., A. K. Barbey, B. R. Postle, and J. Grafman (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *Journal of Neuroscience* 29(47), 14980–14986.

Leshikar, E. D., A. Duarte, and C. Hertzog (2012). Task-selective memory effects for successfully implemented encoding strategies. *PloS one* 7(5), e38160.

Li, J., C. Huang, Z. Hongtu, and for the Alzheimer’s Disease Neuroimaging Initiative (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association* 112(519), 1169–1181.

- Li, L., J. Kang, S. N. Lockhart, J. Adams, and W. J. Jagust (2019). Spatially adaptive varying correlation analysis for multimodal neuroimaging data. *IEEE transactions on medical imaging* 38(1), 113–123.
- Li, X., L. Wang, H. J. Wang, and A. D. N. Initiative (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association* 116(536), 1994–2008.
- Liu, S., S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham, et al. (2014). Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer’s disease. *IEEE transactions on biomedical engineering* 62(4), 1132–1140.
- Lukemire, J., S. Kundu, G. Pagnoni, and Y. Guo (2021). Bayesian joint modeling of multiple brain functional networks. *Journal of the American Statistical Association* 116(534), 518–530.
- Mazziotta, J., A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, et al. (2001). A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356(1412), 1293–1322.
- Mercer, J. (1909). Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character* 209(441-458), 415–446.
- Nakajima, J. and M. West (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics* 31(2), 151–164.
- Ni, Y., F. C. Stingo, and V. Baladandayuthapani (2019). Bayesian graphical regression. *Journal of the American Statistical Association* 114(525), 184–197.
- Nishimura, A., D. B. Dunson, and J. Lu (2020). Discontinuous hamiltonian monte carlo for discrete parameters and

discontinuous likelihoods. *Biometrika* 107(2), 365–380.

Penny, W. D., K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols (2011). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

Roberts, G. O. and J. S. Rosenthal (1998). Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60(1), 255–268.

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist* 19(1), 43–61.

Smith, S. M., C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, et al. (2013). Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168.

Tavor, I., O. P. Jones, R. B. Mars, S. Smith, T. Behrens, and S. Jbabdi (2016). Task-free mri predicts individual differences in brain activity during task performance. *Science* 352(6282), 216–220.

Toews, M., L. Zöllei, and W. M. Wells (2013). Feature-based alignment of volumetric multi-modal images. In *International Conference on Information Processing in Medical Imaging*, pp. 25–36. Springer.

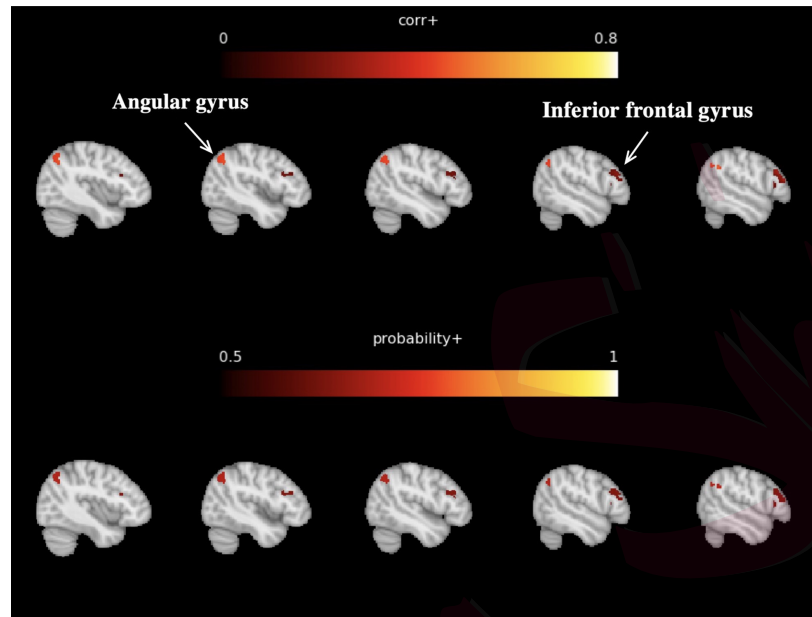
Tokdar, S. T. and J. K. Ghosh (2007). Posterior consistency of logistic Gaussian process priors in density estimation. *Journal of Statistical Planning and Inference* 137(1), 34–42.

Uludağ, K. and A. Roebroeck (2014). General overview on the merits of multimodal neuroimaging data fusion. *Neuroimage* 102, 3–10.

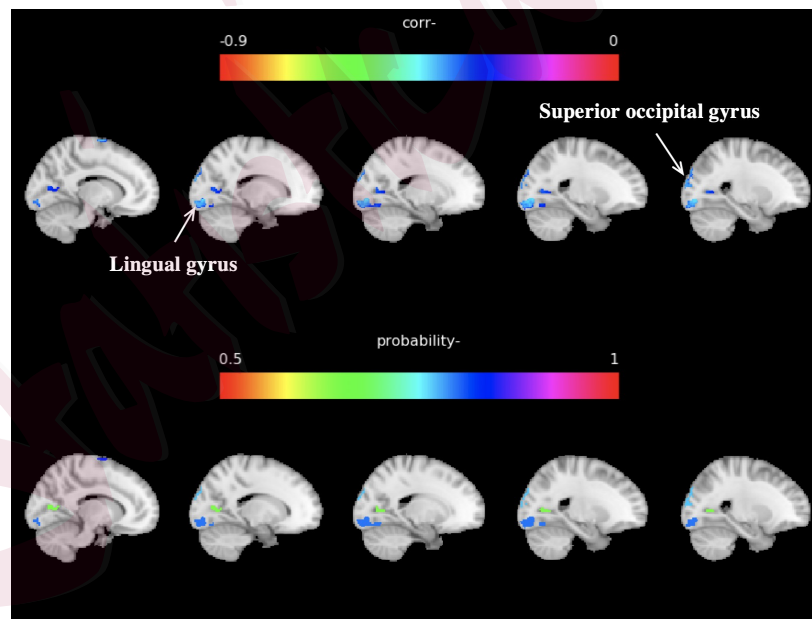
Wang, X., K. Shu, H. Kuang, S. Luo, R. Jin, and J. Liu (2021). The role of spatial alignment in multimodal medical image fusion using deep learning for diagnostic problems. In *Proceedings of the 2021 International Conference on Intelligent Medicine and Health*, pp. 40–46.

- Worsley, K. J., J. E. Taylor, F. Tomaiuolo, and J. Lerch (2004). Unified univariate and multivariate random field theory. *NeuroImage* 23, S189–S195.
- Wu, B., Y. Guo, and J. Kang (2024). Bayesian spatial blind source separation via the thresholded gaussian process. *Journal of the American Statistical Association* 119(545), 422–433.
- Wu, T.-Y., Y. Rachel Wang, and W. H. Wong (2022). Mini-batch metropolis–hastings with reversible sgld proposal. *Journal of the American Statistical Association* 117(537), 386–394.
- Zhu, D., T. Zhang, X. Jiang, X. Hu, H. Chen, N. Yang, J. Lv, J. Han, L. Guo, and T. Liu (2014). Fusing DTI and fMRI data: a survey of methods and applications. *NeuroImage* 102, 184–191.
- Zhu, H., J. Fan, and L. Kong (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association* 109(507), 1084–1098.
- Zhu, H., T. Li, and B. Zhao (2023). Statistical learning methods for neuroimaging data analysis with applications. *Annual Review of Biomedical Data Science* 6, 73–104.
- Zou, Q.-H., C.-Z. Zhu, Y. Yang, X.-N. Zuo, X.-Y. Long, Q.-J. Cao, Y.-F. Wang, and Y.-F. Zang (2008). An improved approach to detection of amplitude of low-frequency fluctuation (alff) for resting-state fMRI: fractional ALFF. *Journal of neuroscience methods* 172(1), 137–141.





(a) Positive correlation map



(b) Negative correlation map

Figure 3: Results of Human Connectome Project data analysis. The sagittal slices of activation regions with significant correlations. Panel (a) shows the slices of positive correlation map and the associated inclusion probability map. Panel (b) shows the slices of negative correlation map and the associated posterior inclusion probability map.