arXiv: arXiv:0000.0000

The Spatial Wishart Process Model and Its Application on Diffusion Tensor Images

Zhou Lan[1], **Brian J Reich**[1], **Joseph Guinness**[1] **and Dipankar Bandyopadhyay**[2]

[1] *Department of Statistics, North Carolina State University*

[2] *VCU Medical Center*

Abstract: Diffusion Tensor Imaging (DTI) is an MRI-based neuroimaging technique used to measure the diffusion process of water molecules in the brain. Data from DTI scan are often summarized by a 3×3 positive definite matrix for each voxel. To make full use of the matrix-valued data, we propose a spatial Wishart process which captures spatial dependence between nearby matrices while assuming diffusion tensors marginally follow a Wishart distribution. We propose a spatial Wishart process with varying coefficients to model spatial dependence and test for covariate effects. Because the spatial Wishart process has a complicated density function, we develop approximation based on Cholesky decomposition. Due the computational problem caused by massive MRI data, we further adopt Nearest Neighbor Gaussian Processes (NNGP) for fast computation. In simulations, we demonstrate the improved performance compared to standard methods for detecting regions of the brain affected by covariates. We apply our method on cocaine users data and controls to detect regions of difference.

Keywords and phrases: Diffusion Tensor Imaging, Spatial Wishart Process, Cholesky Decomposition, Nearest Neighbor Gaussian Processes, Cocaine Use.

1. Introduction

Diffusion Tensor Imaging (DTI) is an MRI-based neuroimaging technique used to measure the diffusion process of water molecules in the brain (Soares et al., 2013). Water diffusions at each location in the brain can be summarized by an estimated 3×3 positive definite matrix representing the covariance matrix for a 3-dimensional brownian motion (Schwartzman, 2006; Dryden, Koloydenko and Zhou, 2009; Basser, Mattiello and LeBihan, 1994). Since DTI has been used extensively to map white matter tractography in the brain, it has an advantage over other MRI-based techniques in revealing abnormal topological organization in the brain (Lo et al., 2010).

Currently, the statistical analysis of DTI can be categorized as: 1) Fractional anisotropy (FA)-based and 2) Random matrix-based. FA is a scalar ranging from 0 to 1 that describes the degree of anisotropy of a diffusion process, larger FA indicates anisotropic diffusion. The univariate representation is feasible for

a statistical analysis due to the difficulty of modeling dependent field of 3×3 positive definite matrices. While univariate representations are convenient, projecting a 3×3 matrix onto a univariate quantity comes with some loss of information. In this regard, Random matrix-based methods are developed subsequently: Schwartzman (2006) and Lee and Schwartzman (2017) used a random ellipsoid model for diffusion tensors, which created matrix-based probability distributions for modelling positive definite matrices.

In this paper, we propose a novel statistical model for modeling DTI. It is obvious that the spatial correlation of the Gaussian Brownian motions among voxels is non-negligible. Hence diffusion tensors generated by those Brownian motions are also spatially correlated. In order to incorporate this spatial dependence in the analysis and prevent loss of information from projecting onto a univariate quantity, we propose a spatial Wishart process as a spatially correlated field of 3×3 positive definite matrices. Comparing to other methods/models, spatial Wishart process provides an appropriate statistical description of DTI for the consideration of spatial correlation and nature of observations.

In clinical research, researchers are interested in detecting which regions of the brain are statistically different among groups (Froeling, Pullens and Leemans, 2016). Therefore, based on spatial Wishart process, we further propose a spatial Wishart model with varying coefficients for voxel-wise statistical tests for differences among groups.

Besides the applications to DTI, we explore some statistical properties of spatial Wishart processes such as its probability density function, in this paper. Also, the assumption of correlated Wishart matrices is widely adopted in areas other than DTI, such as multiple-input-multiple-output (MIMO) systems (Smith and Garth, 2007) and polarimetric image modeling (Ferro-Famil, Pottier and Lee, 2001). The exploration provides some theoretical support for expanding the application of models with assumption of correlated Wishart matrices.

2. Description of Diffusion Tensor Imaging (DTI) Data

This work is also motivated by data from subjects with cocaine use disorder (Ma et al., 2017): a data set recording diffusion tensors from each subject i at each voxel $\boldsymbol{s}$ as well as the subjects' cocaine use status and other information such as age, weight and gender. We have $\boldsymbol{A}_i(\boldsymbol{s})$ represents a 3×3 diffusion tensor of subject i located at voxel $\boldsymbol{s}$ and a $1\times p$ sub design matrix $\boldsymbol{X}_i$ for covariates; Also, i is for subjects from 1 to N, $\boldsymbol{s}$ is for voxels in $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ and p is the number of predictors.

The scientific objective is to identify regions of the brain that are different between groups (e.g. cocaine users and non-users), which motivates our statistical objective of building an appropriate statistical model for DTI data and

accompanying methodology.

Current methods for analyzing DTI either neglect the existence of spatial correlation among diffusion tensors or consider the spatial correlation inappropriately. In this regard, we build a spatial model that appropriately takes the spatial correlation into account. To identify regions that are statistically different, we adopt the spatial varying coefficients process Gelfand et al. (2003) for detection. In next section, we will present the details of spatial Wishart process Model with varying coefficients (SWVC).

3. Spatial Wishart Process

We use $\boldsymbol{A}_i(\boldsymbol{s})$ to denote the 3×3 positive definite matrix representing the diffusion tensor (DT) for subject i at voxel $\boldsymbol{s} \in \mathbb{R}^3$. DTs are 3×3 positive definite matrices and are often modelled as Wishart matrices (Dryden, Koloydenko and Zhou, 2009). We model each DT as a Wishart matrix with scale parameter $\boldsymbol{\Sigma}_i(\boldsymbol{s})/\nu$ and degree of freedom ν, denoted as $\boldsymbol{A}_i(\boldsymbol{s}) \sim \mathcal{W}(\boldsymbol{\Sigma}_i(\boldsymbol{s})/\nu, \nu)$. The expectation $\mathbb{E}\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{\Sigma}_i(\boldsymbol{s})$, depends on the voxel location $\boldsymbol{s}$, and p-vector of predictors, $\boldsymbol{X}_i$.

In the remainder of this section, we first provide the description of the model for scale matrix $\boldsymbol{\Sigma}_i(\boldsymbol{s})$ in terms of $\boldsymbol{X}_i$, and then describe the model for the spatial dependence between responses at nearby voxels that cannot be explained by the covariates.

3.1. Population Mean Matrix

Rather than modelling on $\boldsymbol{\Sigma}_i(\boldsymbol{s})$ directly, we build our model on its Cholesky decomposition. Let $\boldsymbol{L}_i(\boldsymbol{s})$ be the lower triangle Cholesky matrix of $\boldsymbol{\Sigma}_i(\boldsymbol{s})$ with $\boldsymbol{\Sigma}_i(\boldsymbol{s}) = \boldsymbol{L}_i(\boldsymbol{s})\boldsymbol{L}_i(\boldsymbol{s})^T$. The (k, l) element of $\boldsymbol{L}_i(\boldsymbol{s})$ is denoted as $\boldsymbol{L}_{ikl}(\boldsymbol{s})$ and is regressed on covariates as

$$\begin{aligned} \log \boldsymbol{L}^2_{ikk}(\boldsymbol{s}) &= \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}) \\ \boldsymbol{L}_{ikl}(\boldsymbol{s}) &= \boldsymbol{X}_i\boldsymbol{\beta}_{kl}(\boldsymbol{s}) \quad \text{for } k > l \end{aligned} \tag{3.1}$$

where $\boldsymbol{\beta}_{kl}(\boldsymbol{s})$ is a $p \times 1$ spatially-varying coefficient vector at the voxel located at $\boldsymbol{s}$. This formulation ensures that $\boldsymbol{\Sigma}_i(\boldsymbol{s})$ is a positive definite matrix for all $\boldsymbol{X}_i$. If there is a single binary covariate, i.e, $\boldsymbol{X}_i$ is either $[1, 0]$ or $[1, 1]$, the model spans all possible mean matrices of $\boldsymbol{\Sigma}_i(\boldsymbol{s})$; otherwise, the linear assumption should be verified.

The $6p$-vector of covariate effects $\boldsymbol{\beta}(\boldsymbol{s}) = [\boldsymbol{\beta}_{11}(\boldsymbol{s})^T, \boldsymbol{\beta}_{21}(\boldsymbol{s})^T, ..., \boldsymbol{\beta}_{33}(\boldsymbol{s})^T]^T$ are modelled using a multivariate spatial Gaussian process (Gelfand et al., 2003). The mean vector is $\mathbb{E}\boldsymbol{\beta}(\boldsymbol{s}) = 0$ and the separable covariance function is $\mathcal{C}(\boldsymbol{\beta}(\boldsymbol{s})^T, \boldsymbol{\beta}(\boldsymbol{s}')^T) = \mathcal{C}(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\Phi}_\beta)\boldsymbol{T}$, where $\mathcal{C}(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\Phi}_\beta)$ is a spatial correlation function that depends

on correlation parameters $\boldsymbol{\Phi}_\beta$ and T is a $6p \times 6p$ cross-covariance matrix. We denote this process as

$$\boldsymbol{\beta} \sim \mathcal{MGP}(0, \boldsymbol{\Phi}_\beta, \boldsymbol{T}). \tag{3.2}$$

3.2. Spatial Wishart Process For Residual Dependence

We introduce the spatial Wishart process as a means of modelling spatial dependence between DTs.

Definition 3.1. ***Spatial Wishart Process***: Suppose there are ν spatial multivariate Gaussian process with mean $\mathbf{0}$, cross-covariance matrix $\boldsymbol{\Sigma}$ and spatial correlation parameters $\boldsymbol{\Phi}$, denoted as

$$\boldsymbol{Z}_\jmath \overset{i.i.d}{\sim} \mathcal{MGP}(\mathbf{0}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \tag{3.3}$$

For $\jmath = 1, 2, ..., \nu$,let $\boldsymbol{U}(\boldsymbol{s}) = \frac{1}{\nu}\sum_{\jmath=1}^{\nu} \boldsymbol{Z}_\jmath(\boldsymbol{s})\boldsymbol{Z}_\jmath(\boldsymbol{s})^T$. By construction, the marginal distribution of $\boldsymbol{U}(\boldsymbol{s})$ is $\mathcal{W}(\boldsymbol{\Sigma}/\nu, \nu)$ and $\boldsymbol{U}(\boldsymbol{s})$ are spatially dependent with dendence controlled by $\boldsymbol{\Phi}$. Hence, we define this as *spatial Wishart process*, denoted as

$$\boldsymbol{U} \sim \mathcal{SWP}(\nu, \boldsymbol{\Phi}, \boldsymbol{\Sigma}/\nu) \tag{3.4}$$

Our model in Section 3.1 allows the scale parameter to vary by voxel. To allow for this, we decompose $\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{L}_i(\boldsymbol{s})\boldsymbol{U}_i(\boldsymbol{s})\boldsymbol{L}_i(\boldsymbol{s})^T$ where $\boldsymbol{U}_i(\boldsymbol{s})$ is the 3×3 residual with spatial dependence and distributed as $\boldsymbol{U}_i \sim \mathcal{SWP}(\nu, \boldsymbol{\Phi}_u, \boldsymbol{I}/\nu)$. This model preserves the population mean $\mathbb{E}\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{\Sigma}_i(\boldsymbol{s}) = \boldsymbol{L}_i(\boldsymbol{s})\boldsymbol{L}_i(\boldsymbol{s})^T$ (Gupta and Nagar, 1999) and accounts for spatial dependence via dependence among the $\boldsymbol{U}_i(\boldsymbol{s})$.

3.3. Model Summary

The model given in Section 3.1 and 3.2 can be written as

$$\begin{aligned} \boldsymbol{A}_i(\boldsymbol{s}) &= \boldsymbol{L}_i(\boldsymbol{s})\boldsymbol{U}_i(\boldsymbol{s})\boldsymbol{L}_i(\boldsymbol{s})^T \\ \log \boldsymbol{L}^2_{ikk}(\boldsymbol{s}) &= \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}) \\ \boldsymbol{L}_{ikl}(\boldsymbol{s}) &= \boldsymbol{X}_i\boldsymbol{\beta}_{kl}(\boldsymbol{s}) \quad \text{if } k > l \\ \boldsymbol{U}_i &\sim \mathcal{SWP}(\nu, \boldsymbol{\Phi}_u, \boldsymbol{I}/\nu) \end{aligned} \tag{3.5}$$

for subjects $i = 1, 2, ..., N$ and voxels $s \in \{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_n\}$. For Bayesian inference, the prior of $\boldsymbol{\beta}(\boldsymbol{s})$ are defined as in (3.2) and the prior of degree of freedom is assigned as a discrete uniform distribution ranging from 3 to B, denoted as $\nu \sim \mathcal{U}(3, B)$. We use Matérn correlation function Stein (2012) for $\boldsymbol{\beta}$ and $\boldsymbol{U}_i$. The Matérn has two unknown parameters $\boldsymbol{\Phi} = \{\rho, \upsilon\}$, where ρ controls the smoothness of the process and ρ controls the range of spatial correlation. We define rho_u and ν_u for range and smoothness as Matérn parameters in $\boldsymbol{\Phi}_u$ and their priors as $\log \rho_u \sim \mathcal{N}(\mu_\rho, \sigma^2_\rho)$ and $\log \upsilon_u \sim \mathcal{N}(\mu_\upsilon, \sigma^2_\upsilon)$. We also ρ_β and ν_β

for range and smoothness as Matérn parameters in $\boldsymbol{\Phi}_\beta$ and defined based on *priori* knowledge. $\boldsymbol{T}$ is a fixed identity matrix.

We define this model as the spatial Wishart process with varying coefficients (SWVC) and describe its properties in Section 4.

3.4. Bayesian Testing for Local Covariate Effects

Our primary objective is to detect regions affected by covariates. The test whether covariate j affects the mean DT at voxel $\boldsymbol{s}$ is

$$\begin{aligned} &\mathscr{H}_{oj}(\boldsymbol{s}) : \beta_{klj}(\boldsymbol{s}) = 0, \forall k, l \\ &\mathscr{H}_{aj}(\boldsymbol{s}) : \beta_{klj}(\boldsymbol{s}) \neq 0, \text{for some } k, l. \end{aligned} \tag{3.6}$$

Although other methods are possible, we use marginal posterior credible sets of $\beta_{klj}(\boldsymbol{s})$ to test $\mathscr{H}_{oj}(\boldsymbol{s})$. We reject the null hypothesis at voxel $\boldsymbol{s}$ if for any k, l

$$Q^{klj}_{\alpha'/2}(\boldsymbol{s}) > 0 \text{ or } Q^{klj}_{1-\alpha'/2}(\boldsymbol{s}) < 0 \tag{3.7}$$

where $\alpha' = \alpha/m$ is the Bonferroni-controlled significance level, $m = 6$ is the number of comparisons and $Q^{klj}_\alpha(\boldsymbol{s})$ is the posterior α-quantile of $\boldsymbol{\beta}_{klj}(\boldsymbol{s})$. Although Bonferroni correction is derived from the assumption of independence of the tests, it is conservative under dependent tests (Abdi, 2007).

4. Model Properties

In the next two subsections, we describe the properties of the marginal distribution at one site and the joint distribution at multiple sites. The proofs of the theorems and corollaries in this section are summarized in the Appendix.

4.1. Marginal Distributions

As discussed in Section 3, the marginal distribution of $\boldsymbol{A}_i(\boldsymbol{s})$ follows a Wishart distribution. The Cholesky decomposition of a Wishart matrix also has nice properties given in Lemma 8.10 and Proposition 8.11 of Eaton (1983), and summarized in Result 4.1:

Result 4.1. *Suppose* $\boldsymbol{A} \sim \mathcal{W}(\boldsymbol{\Sigma}, \nu)$ *and* $\boldsymbol{T}$ *is* $\boldsymbol{A}$*'s Choleksy matrix so that* $\boldsymbol{T}\boldsymbol{T}^T = \boldsymbol{A}$ *and* $\{t_{kl}\}$ *are the elements of* $\boldsymbol{T}$*, then*

1. ***Diagonals***: $t_{kk}^2 \sim \mathcal{GA}(\frac{\nu-(k-1)}{2}, 2\sigma_{kk}^2)$, *independent over* k, *where* σ_{kk}^2 *is the* k*-th diagonal element of* $\boldsymbol{\Sigma}$;
2. ***Off-diagonals***: *if we partition* $\boldsymbol{A}$, $\boldsymbol{T}$ *and* $\boldsymbol{\Sigma}$ *as*

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}, \boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{0} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \tag{4.1}$$

where $\boldsymbol{A}_{kk}$, $\boldsymbol{T}_{kk}$ *and* $\boldsymbol{\Sigma}_{kk}$ *are* $p_k \times p_k$ *square matrices, then* $\boldsymbol{T}_{21}^T|\boldsymbol{T}_{11} \sim \mathcal{MN}_{p_1,p_2}(\boldsymbol{T}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}, \boldsymbol{I}_{p_1}, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$.

In Result 4.1, $\mathcal{GA}(\alpha, \beta)$ is the gamma distribution with shape parameter α and scale parameter β; $\mathcal{MN}_{p_r,p_c}(\mu, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c)$ is the $p_r \times p_c$ matrix-variate normal distribution (Dawid, 1981) with mean matrix $\boldsymbol{\mu}$, row covariance matrix $\boldsymbol{\Sigma}_r$ and column covariance matrix $\boldsymbol{\Sigma}_c$.

To apply this result to the spatial Wishart process, let $\boldsymbol{T}_i(\boldsymbol{s})$ be the Cholesky decomposition of $\boldsymbol{A}_i(\boldsymbol{s})$ with $\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{T}_i(\boldsymbol{s})\boldsymbol{T}_i(\boldsymbol{s})^T$ and $t_{kli}(\boldsymbol{s})$ be the (k,l)-th element of $\boldsymbol{T}_i(\boldsymbol{s})$. Since $\boldsymbol{A}_i(\boldsymbol{s})$ can also be decomposed as $\boldsymbol{L}_i(\boldsymbol{s})\boldsymbol{U}_i(\boldsymbol{s})\boldsymbol{L}_i(\boldsymbol{s})^T$, we give Corollary 4.1:

Corollary 4.1. *Suppose* $\boldsymbol{A} \sim \mathcal{W}(\boldsymbol{\Sigma}, \nu)$ *has the Cholesky decomposition* $\boldsymbol{T}\boldsymbol{T}^T = \boldsymbol{A}$. *Also* $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}\boldsymbol{L}^T$ *where* $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{\Sigma}$ *and* $\boldsymbol{U} \sim \mathcal{W}(\boldsymbol{I}, \nu)$ *with Cholesky decomposition* $\boldsymbol{D}\boldsymbol{D}^T = \boldsymbol{U}$. *If* $\{t_{kl}\}$ *and* $\{d_{kl}\}$ *are the Cholesky decomposition elements of* $\boldsymbol{A}$ *and* $\boldsymbol{U}$, *respectively, then*

1. ***Diagonals***: $t_{kk}^2 = d_{kk}^2\boldsymbol{L}_{kk}^2 \sim \mathcal{GA}(\frac{\nu-(k-1)}{2}, 2\boldsymbol{L}_{kk}^2)$, *independent over* k;
2. ***Off-diagonals***: *if we partition* $\boldsymbol{L}$, $\boldsymbol{T}$ *and* $\boldsymbol{D}$ *as*

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix}, \boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{0} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{11} & \boldsymbol{0} \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{bmatrix} \tag{4.2}$$

where $\boldsymbol{L}_{kk}$, $\boldsymbol{T}_{kk}$, $\boldsymbol{D}_{kk}$ *are* $p_k \times p_k$ *square matrices, then* $\boldsymbol{T}_{21}|\boldsymbol{T}_{11} \sim \mathcal{MN}_{p_2,p_1}(\boldsymbol{L}_{21}\boldsymbol{D}_{11}, \boldsymbol{L}_{22}\boldsymbol{L}_{22}^T, \boldsymbol{I}_{p_1})$.

By Corollary 4.1, the diagonal elements $t_{ikk}^2(\boldsymbol{s})$ follow gamma distributions

$$t_{ikk}^2(\boldsymbol{s}) \sim \mathcal{GA}\left(\frac{\nu-(k-1)}{2}, \frac{2e^{\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s})}}{\nu}\right). \tag{4.3}$$

Via the log link function for the gamma distribution (Myers and Montgomery, 1997), $\log \mathbb{E}t_{ikk}^2(\boldsymbol{s}) = \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}) + \log\frac{\nu-(k-1)}{\nu}$ and thus $\boldsymbol{\beta}_{kk}(\boldsymbol{s})$ is the linear effect

on the logarithm of expectation of $t^2_{ikk}(\boldsymbol{s})$. The off-diagonals are

$$\begin{aligned} t_{i21}(\boldsymbol{s})|t_{i11}(\boldsymbol{s}) &\sim \mathcal{N}(t_{i11}(\boldsymbol{s})e^{-\frac{1}{2}\boldsymbol{X}\boldsymbol{\beta}_{11}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{21}(\boldsymbol{s}), e^{\boldsymbol{X}_i\boldsymbol{\beta}_{22}(\boldsymbol{s})}/\nu) \\ t_{i31}(\boldsymbol{s})|t_{i11}(\boldsymbol{s}), t_{i21}(\boldsymbol{s}), t_{i22}(\boldsymbol{s}) &\sim \mathcal{N}(t_{i11}(\boldsymbol{s})e^{-\frac{1}{2}\boldsymbol{X}\boldsymbol{\beta}_{11}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{31}(\boldsymbol{s}) + d_{i21}(\boldsymbol{s})\boldsymbol{X}_i\boldsymbol{\beta}_{32}(\boldsymbol{s}), e^{\boldsymbol{X}_i\boldsymbol{\beta}_{33}(\boldsymbol{s})}/\nu) \\ t_{i32}(\boldsymbol{s})|t_{i22}(\boldsymbol{s}) &\sim \mathcal{N}(t_{i22}(\boldsymbol{s})e^{-\frac{1}{2}\boldsymbol{X}\boldsymbol{\beta}_{22}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{32}(\boldsymbol{s}), e^{\boldsymbol{X}_i\boldsymbol{\beta}_{33}(\boldsymbol{s})}/\nu) \end{aligned} \tag{4.4}$$

where $d_{i21}(\boldsymbol{s}) = \frac{e^{\frac{1}{2}\boldsymbol{X}_i\boldsymbol{\beta}_{11}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{21}(\boldsymbol{s}) - t_{11}(\boldsymbol{s})e^{-\frac{1}{2}\boldsymbol{X}_i\boldsymbol{\beta}_{11}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{21}(\boldsymbol{s})}{e^{\frac{1}{2}\boldsymbol{X}_i\boldsymbol{\beta}_{22}(\boldsymbol{s})}}$. Thus $\boldsymbol{\beta}_{kl}(\boldsymbol{s})$ is interpreted as the linear effect on the expectation of $t_{ikl}(\boldsymbol{s})$.

4.2. Joint Spatial Distribution

As discussed in Section 4.1, a Wishart random matrix at one site can be written in terms of conditionally independent elements of its Cholesky decomposition.

Below we show that the joint distribution of Choleksy decomposition elements of spatial Wishart process does not have such a convenient representation.

Theorem 4.2. *Let $\boldsymbol{U}$ follows a standard spatial Wishart process, denoted as $\frac{1}{\nu}\sum_{\jmath=1}^{\nu}\boldsymbol{Z}_\jmath(\boldsymbol{s})\boldsymbol{Z}_\jmath(\boldsymbol{s})^T = \boldsymbol{U}(\boldsymbol{s})$ and $\boldsymbol{U} \sim \mathcal{SWP}(\nu, \boldsymbol{\Phi}, \boldsymbol{I}/\nu)$; $\{d_{kl}(\boldsymbol{s})\}$ are Choleksy decomposition elements of $\boldsymbol{U}(\boldsymbol{s})$ such that $\boldsymbol{D}(\boldsymbol{s})\boldsymbol{D}(\boldsymbol{s})^T = \boldsymbol{U}(\boldsymbol{s})$, then*

1. ***Diagonals**: $d^2_{kk} \sim \mathcal{MGA}(\frac{\nu-(k-1)}{2}, \frac{2}{\nu}, \boldsymbol{\Phi})$;*
2. ***Off-diagonals**: if we partition $\boldsymbol{U}(\boldsymbol{s})$ and $\boldsymbol{D}(\boldsymbol{s})$ as*

$$\boldsymbol{U}(\boldsymbol{s}) = \begin{bmatrix} \boldsymbol{U}_{11}(\boldsymbol{s}) & \boldsymbol{U}_{12}(\boldsymbol{s}) \\ \boldsymbol{U}_{21}(\boldsymbol{s}) & \boldsymbol{U}_{22}(\boldsymbol{s}) \end{bmatrix}, \boldsymbol{D}(\boldsymbol{s}) = \begin{bmatrix} \boldsymbol{D}_{11}(\boldsymbol{s}) & \boldsymbol{0} \\ \boldsymbol{D}_{21}(\boldsymbol{s}) & \boldsymbol{D}_{22}(\boldsymbol{s}) \end{bmatrix} \tag{4.5}$$

where $\boldsymbol{U}_{kk}$ and $\boldsymbol{D}_{kk}$ are $p_k \times p_k$ square matrices. Also $\frac{1}{\nu}\sum_{\jmath=1}^{\nu}\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})^T = \boldsymbol{U}_{11}(\boldsymbol{s})$, then $\boldsymbol{D}_{21}(\boldsymbol{s})|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \jmath, s$ is a matrix variate Gaussian process with mean zero and conditional covariance as $\boldsymbol{cov}(\boldsymbol{D}_{21}(\boldsymbol{s}), \boldsymbol{D}_{21}(\boldsymbol{s}')|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \jmath, s) = \rho(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})\boldsymbol{I}_{p_2} \otimes \frac{1}{\nu}\boldsymbol{D}_{11}(\boldsymbol{s})^{-1}\left[\sum_{\jmath=1}^{\nu}\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}')^T\right][\boldsymbol{D}_{11}(\boldsymbol{s}')^{-1}]^T$, where $\rho(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})$ is the spatial correlation between site $\boldsymbol{s}$ and $\boldsymbol{s}'$ defined by parameter $\boldsymbol{\Phi}$.

In Theorem 4.2, $\mathcal{MVG}(\alpha(\boldsymbol{s}), \beta(\boldsymbol{s}), \boldsymbol{\Phi})$ is the multivariate gamma distribution defined in Krishnaiah and Rao (1961), where $\alpha(\boldsymbol{s}), \beta(\boldsymbol{s})$ are shape and scale parameter for marginal gamma distribution, respectively, and $\boldsymbol{\Phi}$ is the spatial correlation parameters of the latent Gaussian process.

By applying Theorem 4.2, we have the generalization to the model for the Choleky decomposition elements of $\boldsymbol{A}_i(\boldsymbol{s})$, $t_{ijk}(\boldsymbol{s})$ and give the explicit expression in the Appendix.

5. Approximations

The spatial Wishart process provides a reasonable starting point for a spatial extension of standard statistical methods that rely on the Wishart distribution. However, the distributions of both diagonals and off-diagonals pose computational challenges: the multivariate gamma distribution for the diagonals has a closed form but it is too complicated to be applied in high dimensions; covariance of the off-diagonals involves the latent Gaussian process $\boldsymbol{Z}_j$ which are not observed in practice. In this section, we discuss approximations that make computations feasible.

5.1. Diagonals Elements

As given in Theorem 4.2, the joint distribution of diagonal elements is multivariate gamma. Viraswami (1991) provides a comprehensive review of the multivariate gamma distribution. An analytically tractable expression can be obtained via traces of Wishart matrices (Jensen et al., 1970), but the expression is complicated and requires consideration calculation. Smith and Garth (2007) provided an expression for density function of bi-correlated gamma variables, however, this does not apply for data on $\mathbb{R}^3$.

5.1.1. Gaussian Copula Approximation

The Gaussian copula provides an approximation for the joint distribution of distribution of $\boldsymbol{t}^2_{ikk} = [t^2_{ikk}(\boldsymbol{s}_1), ..., t^2_{ikk}(\boldsymbol{s}_n)]$. By retaining the gamma marginal distribution, we model spatial dependence using a multivariate normal distribution on the transformed residuals. Let $\boldsymbol{u}_i = [\Phi^{-1}(F_{\Gamma,\boldsymbol{s}_1}(t^2_{ikk}(\boldsymbol{s}_1)), ..., \Phi^{-1}(F_{\Gamma,\boldsymbol{s}_n}(t^2_{ikk}(\boldsymbol{s}_n))]^T$ where $F_{\Gamma,\boldsymbol{s}_1}$ is the cumulative distribution function (CDF) of a gamma distribution with corresponding parameters and Φ^{-1} is the inverse CDF of standard normal distribution. This results the joint density function of $\boldsymbol{t}^2_{ikk}$ as

$$g(t^2_{ikk}(\boldsymbol{s}_1), ..., t^2_{ikk}(\boldsymbol{s}_n)) = \frac{1}{|\boldsymbol{R}_C|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\boldsymbol{u}_i^T(\boldsymbol{R}_C^{-1} - \boldsymbol{I})\boldsymbol{u}_i\right] \prod_{\boldsymbol{s}} f_{\Gamma,\boldsymbol{s}}\left(t_{ikk}(\boldsymbol{s})\right) \tag{5.1}$$

, where $f_{\Gamma,\boldsymbol{s}}$ is the probability density function (PDF) of a gamma distribution with corresponding parameters.

Although Gaussian copula approximation provides a way for modeling gamma variables with correlations, disadvantages such as the problem of copula dependency is handled in a different way from that of the default process itself are summarized in Donnelly and Embrechts (2010).

5.1.2. Gaussian Approximation via Logarithm Gamma

Multiple normal approximations for gamma-distributed variables are studied and compared by Peizer and Pratt (1968) and Pratt (1968). In our paper, we apply logarithm of gamma-distributed variables for normal approximation.

If $X \sim \mathcal{GA}(\alpha, \beta)$, then $\mathbb{E}(\log X) = \psi^{(0)}(\alpha) + \log(\beta)$ and $\mathbb{V}(\log X) = \psi^{(1)}(\alpha)$, where $\psi^{(k)}$ is the Psi (polygamma) function. Therefore, we assume the approximation as $\log X \dot{\sim} \mathcal{N}(\psi^{(0)}(\alpha) + \log(\beta), \psi^{(1)}(\alpha))$, where $\dot{\sim}$ is for approximately distributed. The variance is stabilized and only dependent on ν after log-transformation. Then the marginal distribution of $t^2_{ikk}(\boldsymbol{s}) \sim \mathcal{GA}\left(\frac{\nu-(k-1)}{2}, \frac{2e^{\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s})}}{\nu}\right)$ has approximated distribution as

$$\log t^2_{ikk}(\boldsymbol{s}) \dot{\sim} \mathcal{N}\left(\psi^{(0)}(\frac{\nu-(k-1)}{2}) + \log(\frac{2e^{\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s})}}{\nu}), \psi^{(1)}(\frac{\nu-(k-1)}{2})\right). \tag{5.2}$$

The multivariate gamma distribution can be approximated as a Gaussian process with mean function as $\psi^{(0)}(\frac{\nu-(k-1)}{2}) + \log\frac{2}{\nu} + \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s})$, spatial covariance function as $\psi^{(1)}\left(\frac{\nu-(k-1)}{2}\right)\mathcal{C}(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\Phi}_\Xi)$, where $\mathcal{C}(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\Phi}_\Xi)$ is the spatial correlation function defined by $\boldsymbol{\Phi}_\Xi$, denoted as

$$\log t^2_{ikk} \dot{\sim} \mathcal{GP}\left(\psi^{(0)}(\frac{\nu-(k-1)}{2}) + \log\frac{2}{\nu} + \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}), \psi^{(1)}\left(\frac{\nu-(k-1)}{2}\right)\mathcal{C}(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\Phi}_\Xi)\right). \tag{5.3}$$

Logarithm gamma provides a simple normal approximation. However, the exact spatial correlation of the log gamma process $\mathbb{E}[\log t^2_{ikk}(\boldsymbol{s}) \log t^2_{ikk}(\boldsymbol{s}')] - \mathbb{E}[\log t^2_{ikk}(\boldsymbol{s})]\mathbb{E}[\log t^2_{ikk}(\boldsymbol{s}')]$ depends on $\boldsymbol{\beta}_{kk}(\boldsymbol{s})$, but we ignore this to avoid a complex covariance expression.

5.2. Off-Diagonal Elements

In Theorem 4.2, the covariance of the off-diagonal process depends on $\tilde{\boldsymbol{Z}}(\boldsymbol{s}, \boldsymbol{s}') = \sum_{j=1}^{\nu} \boldsymbol{Z}_{j1}(\boldsymbol{s})\boldsymbol{Z}_{j1}(\boldsymbol{s}')$ which is not observed in practice. Motivated by Cauchy-Schwarz inequality with a matrix extension (Tripathi, 1999), we can approxi-

mate the conditional covariance in Theorem 4.2 as

$$\begin{aligned}\boldsymbol{cov}(\boldsymbol{D}_{21}(\boldsymbol{s}),\boldsymbol{D}_{21}(\boldsymbol{s}')|\boldsymbol{Z}_{j1}(\boldsymbol{s})\forall_j,s) &= \rho(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi})\boldsymbol{I}_{p_2}\otimes\left[\frac{1}{\nu}\boldsymbol{D}_{11}(\boldsymbol{s})^{-1}\left[\tilde{\boldsymbol{Z}}(\boldsymbol{s},\boldsymbol{s}')\right][\boldsymbol{D}_{11}(\boldsymbol{s}')^{-1}]^T\right]\\ &\approx \rho(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi})\boldsymbol{I}_{p_2}\otimes\left[\frac{1}{\nu}\mathbb{E}\left[\boldsymbol{U}_{11}(\boldsymbol{s})\right]^{-\frac{1}{2}}\mathbb{E}\left[\tilde{\boldsymbol{Z}}(\boldsymbol{s},\boldsymbol{s}')\right][\mathbb{E}\left[\boldsymbol{U}_{11}(\boldsymbol{s})\right]^{-\frac{1}{2}}]^T\right]\\ &= \rho(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi})\boldsymbol{I}_{p_2}\otimes\frac{1}{\nu}\rho(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi})\boldsymbol{I}_{p_1}.\end{aligned} \tag{5.4}$$

This approximation can be adopted for obtaining the approximated distribution of off-diagonal cholesky decomposition elements $\{t_{ikl}(\boldsymbol{s})\}$ of $\boldsymbol{A}_i(\boldsymbol{s})$ and their explicit expression is as

$$\begin{aligned}t_{i21}|[t_{i11}(\boldsymbol{s})\forall\boldsymbol{s}\ \forall\boldsymbol{s},j] &\sim \mathcal{GP}(t_{i11}(\boldsymbol{s})e^{-\frac{1}{2}\boldsymbol{X}_i\boldsymbol{\beta}_{11}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{21}(\boldsymbol{s}),\\ &\quad \frac{1}{\nu}\mathcal{C}(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi}_u)^2e^{\sum_{\boldsymbol{w}\in\{\boldsymbol{s},\boldsymbol{s}'\}}\boldsymbol{X}_i\boldsymbol{\beta}_{22}(\boldsymbol{w})\frac{1}{2}})\\ t_{i31}|t_{i11}(\boldsymbol{s}),t_{i21}(\boldsymbol{s}),t_{i22}(\boldsymbol{s})\ \forall\boldsymbol{s},j] &\sim \mathcal{GP}(t_{i11}(\boldsymbol{s})e^{-\frac{1}{2}\boldsymbol{X}_i\boldsymbol{\beta}_{11}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{31}(\boldsymbol{s})+d_{i21}(\boldsymbol{s})\boldsymbol{X}_i\boldsymbol{\beta}_{32}(\boldsymbol{s}),\\ &\quad \frac{1}{\nu}\mathcal{C}(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi}_u)^2e^{\sum_{\boldsymbol{w}\in\{\boldsymbol{s},\boldsymbol{s}'\}}\boldsymbol{X}_i\boldsymbol{\beta}_{33}(\boldsymbol{w})\frac{1}{2}})\\ t_{i32}|[t_{i22}(\boldsymbol{s})\forall\boldsymbol{s}] &\sim \mathcal{GP}(t_{i11}(\boldsymbol{s})e^{-\frac{1}{2}\boldsymbol{X}_i\boldsymbol{\beta}_{11}(\boldsymbol{s})}\boldsymbol{X}_i\boldsymbol{\beta}_{32}(\boldsymbol{s}),\\ &\quad \frac{1}{\nu}\mathcal{C}(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi}_u)^2e^{\sum_{\boldsymbol{w}\in\{\boldsymbol{s},\boldsymbol{s}'\}}\boldsymbol{X}_i\boldsymbol{\beta}_{22}(\boldsymbol{w})\frac{1}{2}}).\end{aligned} \tag{5.5}$$

6. Fast Computation via Nearest Neighbor Gaussian Process (NNGP)

Massive spatial data such as MRI data poses a computational challenge because the spatial covariance matrix. we use the Nearest Neighbor Gaussian Process (NNGP) (Datta et al., 2016) approximation for Fast Computation.

The basic idea of NNGP is similar to Vecchia's method (Vecchia, 1988) to approximate the density function of Gaussian process as products of conditional densities. That is, for a spatial Gaussian process $[\boldsymbol{w}_{\boldsymbol{s}_1},...,\boldsymbol{w}_{\boldsymbol{s}_n}]$ with n locations, we consider those k locations are on a directed acyclic graph with k nodes and each node ($\boldsymbol{s}_i$ e.g.) as a source node has $m \ll k$ direct edges towards sink nodes, denoted as $N(\boldsymbol{s}_i)$. Then the density function is approximated as

$$p(\boldsymbol{w}_{s_1},...,\boldsymbol{w}_{s_n}) \approx \prod_{i=1}^{n} p(\boldsymbol{w}_{s_i}|\boldsymbol{w}_{N(\boldsymbol{s}_i)}) \tag{6.1}$$

6.1. Application of NNGP on SWVC model

In Section 5.1 and 5.2, we have described the method for approximating off-diagnol elements and diagnoal elements in *SWVC-Log Gamma* as spatial Gaussian process, where NNGP can be directly applied on those. In *SWVC-Gaussian copula*, the density function of Diagnols approximated by Gaussian copula can be approximated via the idea of NNGP as well by conditioning all nearest neighbors:

$$g(t^2_{ikk}(\boldsymbol{s}_1), ..., t^2_{ikk}(\boldsymbol{s}_n)) \approx \prod_{\imath=1}^{n} g(t^2_{ikk}(\boldsymbol{s}_\imath)|t^2_{ikk}(\boldsymbol{s}_\jmath); \boldsymbol{s}_\jmath \in \boldsymbol{N}(\boldsymbol{s}_\imath)) \tag{6.2}$$

and the conditional Gaussian copula density (Käärik and Käärik, 2009) can be written as

$$g(t^2_{ikk}(\boldsymbol{s}_\imath)|t^2_{ikk}(\boldsymbol{s}_\jmath); \boldsymbol{s}_\jmath \in \boldsymbol{N}(\boldsymbol{s}_\imath)) = f_\Gamma(t_{ikk}(\boldsymbol{s}_\imath)|\frac{\nu-(k-1)}{2}; \frac{2e^{\boldsymbol{X}\boldsymbol{\beta}_{kk}(\boldsymbol{s}_\imath)}}{\nu})\frac{c[\Phi(u(\boldsymbol{s}_{\imath'})), \boldsymbol{s}_{\imath'} \in \boldsymbol{s}_\imath \cup \boldsymbol{N}(\boldsymbol{s}_\imath); \boldsymbol{R}_c]}{c[\Phi(u(\boldsymbol{s}_{\imath'})), \boldsymbol{s}_{\imath'} \in \boldsymbol{N}(\boldsymbol{s}_\imath); \boldsymbol{R}_c]} \tag{6.3}$$

where $c[\Phi(u(\boldsymbol{s}_{\imath'}))$ some $\boldsymbol{s}_{\imath'}; \boldsymbol{R}_c]$ is copula density of correspondent sub-samples.

7. Simulation Studies

In this section, we will perform the simulation studies to demonstrate the performance of spatial Wishart process models, computational improvements via NNGP and robustness to model assumption.

7.1. Simulation 1: Methods Comparison

In Simulation 1, we compared the SWVC model to other methods for data generated from the SWVC model. We generate data on the 3D $5 \times 5 \times 5$ grid with spacing 1 between adjacent grid points and a $2 \times 2 \times 2$ region of different in the middle for imitating a brain image with a region region of different between groups. 10 subjects of each group are simulated with spatial correlation parameters for correlated residual matrices as $\{\rho_u = 1, \nu_u = 0.5\}$, $\{\rho_u = 2, \nu_u = 0.5\}$ and $\{\rho_u = 3, \nu_u = 0.5\}$, named as *Low Spatial Correlation*, *Median Spatial Correlation* and *High Spatial Correlation*, respectively. We set the Wishart matrix degree of freedom $\nu = 10$. For voxels $\boldsymbol{s}$ with signals, we set $\beta_{kk}(\boldsymbol{s}) = 0.5$ and $\beta_{kl}(\boldsymbol{s}) = 0.25, \forall k > l$; otherwise, $\beta_{kl}(\boldsymbol{s}) = 0, \forall k, l$. Without loss of generality, we set intercepts as 0. We set prior of ν as $\mathcal{U}(3, 20)$, $\boldsymbol{\Phi}_\beta = \{\sigma_\beta = 0.1, \rho_\beta = 0.1, \upsilon_\beta = 0.5\}$, and $\log \rho_u \sim \mathcal{N}(0, 1)$ and $\log \upsilon_u \sim \mathcal{N}(-1, 1)$.

We choose *Gaussian Symmetrical Matrix* model (GSM) (Schwartzman, Mascarenhas and Taylor, 2008), a matrix-based model without considering spatial correlation and *Univariate Spatial Varying Coefficients Model* (Gelfand et al., 2003) (USVC), a univariate spatial model with logit transformation of FA as response, where $\text{FA} = \sqrt{\frac{1}{2}}\frac{\sqrt{(\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_3-\lambda_1)^2}}{\sqrt{\lambda_1^2+\lambda_2^2+\lambda_3^2}}$ and

$\{\lambda_1, \lambda_2, \lambda_3\}$ are the eigenvalues of a DT. We compare *SWVC-Gaussian copula* and *SWVC-Log Gamma* with 5000 samples after 1000 burn-ins with GSM and USVC. We use the `Matlab` codes of GSM and `R` package `spTDyn` for implementation.

We have summarized true positive rate (TPR), false positive rate (FPR), false discovery rate (FDR) and the typical computational times from 50 simulations with different seeds in Table 1. The selection of all methods are under significance of 0.05 with Bonferroni correction.

TABLE 1

Selection Accuracy of Simulation 1: Multivariate spatial model: The spatial Wishart varying coefficients (SWVC) with Gaussian copula and log-gamma approximation are compared with univariate spatial varying coefficients (USVC) and Gaussian Symmetrical Matrix model (GSM) in terms of true positive rate (TPR), false positive rate (FPR), false discovery rate (FDR).

	SWVC Gaussian copula	SWVC Log Gamma	USVC	GSM
Low Spatial Correlation $\rho_u = 1$				
TPR	0.93	0.65	0.63	0.51
FPR	0.003	0.0002	0.03	0.0005
FDR	0.007	0.02	0.05	0.03
Median Spatial Correlation $\rho_u = 2$				
TPR	0.96	0.90	0.125	0.49
FPR	0.007	0.002	0.03	0.0005
FDR	0.001	0.009	0.09	0.03
High Spatial Correlation $\rho_u = 3$				
TRP	0.99	0.91	0.125	0.50
FPR	0.02	0.003	0.06	0.0005
FDR	0.02	0.009	0.10	0.03
Time (s)	1200	1000	800	10

We find although the computational time is longer for the inference of SWVC model is via MCMC, SWVC has a powerful test with good control of FPR and FDR when compared to other methods.

7.2. Simulation 2: NNGP for Massive Data

In simulation 2, we use a larger $10\times10\times10$ grids with spacing 1 between adjacent grid points and a $3 \times 3 \times 3$ region of difference in the middle for demonstrating the implementation of NNGP can facilitate the computing with comparable performance on accuracy of selection. Similar to Simulation 1, 10 subjects of each group are simulated with spatial correlation as $\{\rho_u = 1, \nu_u = 0.5\}$. We compare *SWVC-Gaussian copula* model and *SWVC-Log Gamma* model with their NNGP model, named as *SWVC-Gaussian copula-NNGP* and *SWVC-Log Gamma-NNGP*, respectively. All methods generate 5000 samples after 1000 burn-ins. We choose 10 nearest neighbors for NNGP, which is recommended by Datta et al. (2016). The prior settings, the values of coefficent $\boldsymbol{\beta}$, Wishart

matrix degree of freedom ν are set same as Simulation 1.

We have summarized TPR, FPR, FDR and the typical computational time from 10 simulations with different seeds in Table 2. The selection of all methods are under significance of 0.05 with Bonferroni correction. Also, we summarized the selection accuracy in Table 2. From the simulation study, the computational

TABLE 2
Selection Accuracy of Simulation 2: The spatial Wishart varying coefficients (SWVC) with Gaussian copula and log-gamma approximation are compared with their NNGP model: SWVC-Gaussian copula-NNGP and SWVC-Log Gamma-NNGP, respectively, in terms of true positive rate (TPR), false positive rate (FPR), false discovery rate (FDR).

	SWVC Gaussian copula	SWVC Gaussian copula NNGP	SWVC Log Gamma	SWVC Log Gamma NNGP
TPR	0.7778	0.7593	0.7130	0.7010
FPR	0.0044	0.0075	0.0010	0.0013
FDR	0.0130	0.0137	0.0088	0.0204
Time (hr)	12	5	12	5

time is largely reduced due to implementation of NNGP and the accuracy in regions selection is not changed.

7.3. Simulation 3: Occam's Razor

Although spatial Wishart process model provides a good approach for modeling spatially correlated DTs, using the whole matrix information, taking the full matrix for modeling is sometimes considered as a violation of Occam's Razor if a univariate model is sufficient for modeling DTI.

For demonstrating the SWVC test is still powerful for testing signals when FA based model is sufficient, we design a simulation where all coefficient are zeros except $\beta_{11}(\boldsymbol{s})$ for some $\boldsymbol{s}$ located at voxels with regional effect. This simulation is to imitating the situation that a univariate model is sufficient.

We reuse the $5 \times 5 \times 5$ grids with spacing 1 between adjacent grid points and a $2 \times 2 \times 2$ region of different in the middle. We implemented *SWVC-Gaussian copula* and *SWVC-Log Gamma* with 5000 samples after 1000 burn-ins. FA are computed and logit transformed for USVC fitting with 5000 samples after 1000 burn-ins. We have summarized TPR, FPR and FDR from 20 simulations with different seeds in Table 3. The selection of all methods are under significance of 0.05 with Bonferroni correction.

From the simulation result, we find the SWVC is still necessary and USVC is not powerful enough for detecting regions of difference at this scenario as well.

TABLE 3
Selection Accuracy of Simulation 3: The univariate spatial varying coefficient model (USVC) is compared to the spatial Wishart varying coefficients (SWVC) with Gaussian copula and log-gamma approximation in terms of true positive rate (TPR), false positive rate (FPR), false discovery rate (FDR).

	SWVC Gaussian copula	SWVC Log Gamma	USVC
TPR	0.9812	0.8250	0.1562
FPR	0.0030	0.0026	0.1594
FDR	0.0040	0.0136	0.2035

8. Application: Cocaine Use Disorders

8.1. Data Source and Description

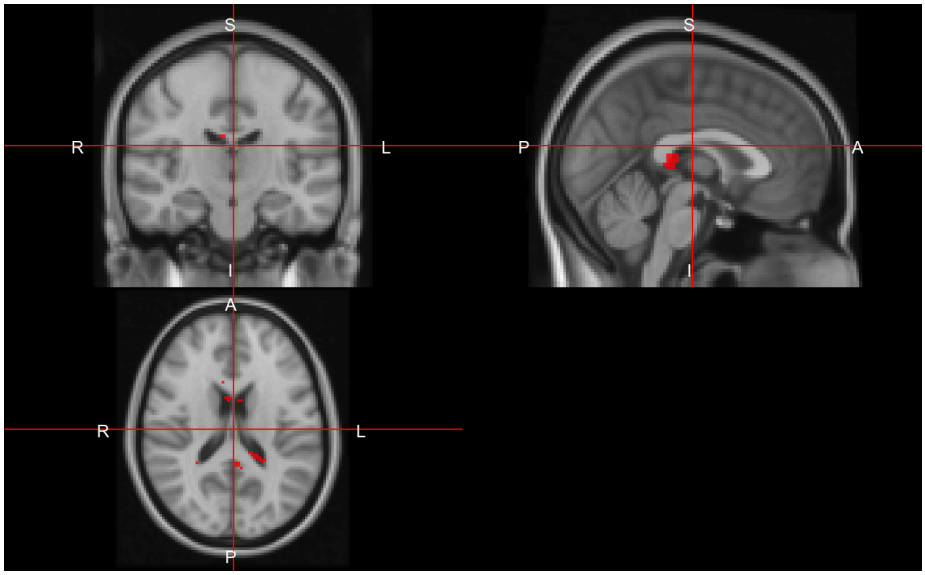
We apply SWVC on the data collected as part of a study of cocaine use on DTI (Ma et al., 2017). The data are from *the Institute for Drug and Alcohol Studies of Virginia Commonwealth University* (VCU). The study recruited 11 cocaine users and 11 controls to test for areas with white matter alteration.

The corpus callosum, the major white-matter bundle connecting the right and left hemispheres of the brain is a region of interest in studies of regional effects of cocaine (Ma et al., 2009) and so we focused on the corpus callosum. The corpus callosum region contains $15,273$ voxels.

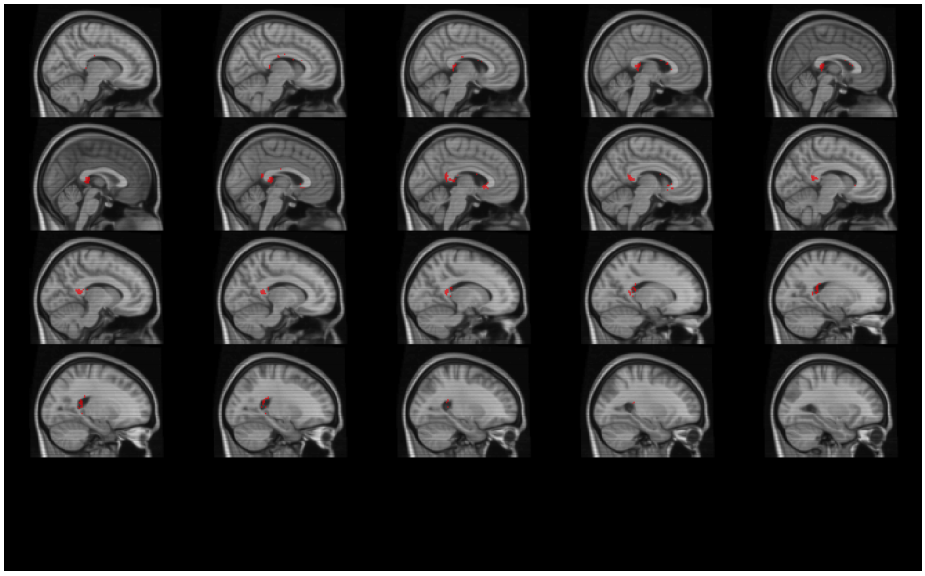
8.2. Data Analysis

We applied *SWVC-Gaussian copula* with 5000 MCMC iterations with the first 2000 discarded burn-in. Spatially continuous region of difference (in red) in the brain is detected and displayed using an orthographic projection and multiple slides at sagittal plane (Figure 1). The result is consistent with previous studies on cocaine users (Lane et al., 2010), where a region of difference is detected in splenium, which is located on the posterior end of the corpus callosum. However, different from the findings using other methods, the regions detected by SWVC is a spatial continuous region and some regions of difference outlined by Lane et al. (2010) may be neglected since SWVC considers those as a cause of spatial residuals.

Splenium is an important region in corpus callosum bridging the left and right hemispheres of the brain (See Figure 2) with essential effect on cognition. We believe the region of difference detection can assist scientists on brain science to have further diagnostics on the regional effect of cocaine.

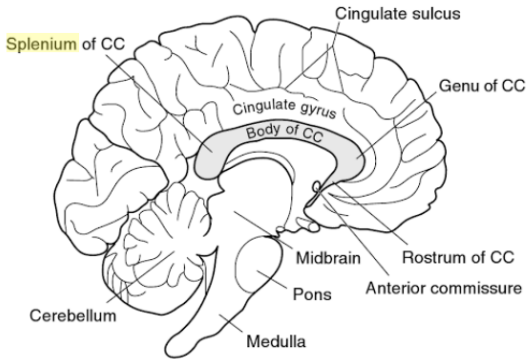


(a) Orthographic projection

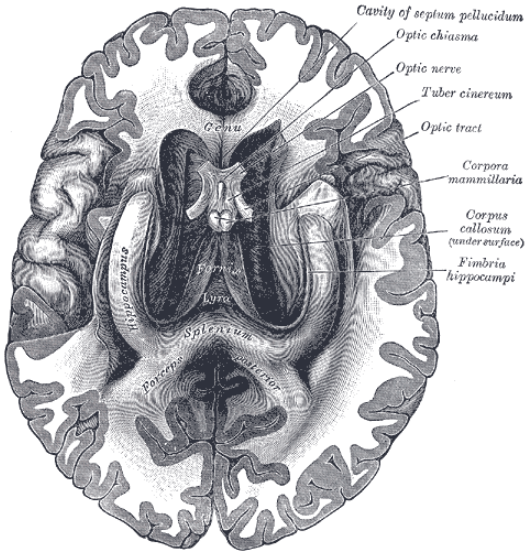


(b) Multiple slices at sagittal plane

Figure 1: Region of difference detection: a spatially continuous region of difference (in red) in the brain between cocaine users and healthy ones; the regions are presented using orthographic projection and multiple slices at sagittal plane.



(a) Sagittal plane: adopted from Clark, Boutros and Mendez (2010)



(b) Axial plane: adopted from Gray (1918)

Figure 2: Anatomy of corpus callosum: the anatomies of corpus callosum are presented at sagittal and axial plane; splenium is also marked, which is located on the posterior end of the corpus callosum and the left and right hemispheres of the brain

9. Discussion

In this article, we proposed a spatial model for modeling DTI data. The model captures the spatial dependence among DTs using the spatial Wishart process. By applying the Choleksy decomposition and other approximation meth-

ods, inference and computation of spatial Wishart process become feasible. The computational problem caused by massive MRI data is mitigated by NNGP approximation.

In our simulation studies, we find the proposed method is powerful and has good error control when compared to other methods, and the NNGP facilitates the computation for large data-set with little impact on the performance in accuracy. The application to the data set of cocaine users demonstrates the usage of SWVC for detecting meaningful regions of difference.

Appendix

Proofs in Model Properties

Proof of Corollary 4.1

Proof.

1. **diagonals**: It is easy to show $t_{kk}^2 = d_{kk}^2 \boldsymbol{L}_{kk}^2$ and $d_{kk}^2 \sim \mathcal{GA}(\frac{\nu-(k-1)}{2}, 2)$, hence $t_{kk}^2 = d_{kk}^2 \boldsymbol{L}_{kk}^2 \sim \mathcal{GA}(\frac{\nu-(k-1)}{2}, 2\boldsymbol{L}_{kk}^2)$;
2. **off-diagonals**: From Result 4.1 and Theorem 2.3.1 of Gupta and Nagar (1999), we have $\boldsymbol{D}_{21}|\boldsymbol{D}_{11} \sim \mathcal{MN}_{p_2,p_1}(\boldsymbol{D}_{11}, \boldsymbol{I}_{p_2}, \boldsymbol{I}_{p_1})$. We know

$$\begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{0} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_{11} & \boldsymbol{0} \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_{11}\boldsymbol{D}_{11} & \boldsymbol{0} \\ \boldsymbol{L}_{21}\boldsymbol{D}_{11} + \boldsymbol{L}_{22}\boldsymbol{D}_{21} & \boldsymbol{L}_{22}\boldsymbol{D}_{22} \end{bmatrix} \tag{9.1}$$

Given Theorem 2.3.10 of Gupta and Nagar (1999), we have $\boldsymbol{T}_{21}|\boldsymbol{T}_{11} \sim \mathcal{MN}_{p_2,p_1}(\boldsymbol{L}_{21}\boldsymbol{D}_{11}, \boldsymbol{L}_{22}\boldsymbol{L}_{22}^T, \boldsymbol{I}_{p_1})$

□

Proof of Theorem 4.2

Proof. 1. **diagonals**: It is not difficult to show that $\boldsymbol{U}_{22.1}(\boldsymbol{s}) = \boldsymbol{D}_{22}(\boldsymbol{s})\boldsymbol{D}_{22}(\boldsymbol{s})^T \sim \mathcal{W}(\boldsymbol{I}_{p_2}/\nu, \nu-p_1)$ and $\boldsymbol{U}_{22.1}(\boldsymbol{s}) = \frac{1}{\nu}[\boldsymbol{Q}(\boldsymbol{s})\boldsymbol{Z}_2(\boldsymbol{s})]^T[\boldsymbol{Q}(\boldsymbol{s})\boldsymbol{Z}_2(\boldsymbol{s})]$ where $\boldsymbol{Q}(\boldsymbol{s}) = \boldsymbol{I}_\nu - \boldsymbol{Z}_1(\boldsymbol{s})[\boldsymbol{Z}_1(\boldsymbol{s})^T\boldsymbol{Z}_1(\boldsymbol{s})]^{-1}\boldsymbol{Z}_1(\boldsymbol{s})^T$ and $\boldsymbol{Z}_1(\boldsymbol{s}) = [\boldsymbol{Z}_{11}(\boldsymbol{s}), ..., \boldsymbol{Z}_{\nu 1}(\boldsymbol{s})]^T$, a data matrix, a definition which is often used in Mardia, Kent and Bibby (1980). Let $\boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{Q}(\boldsymbol{s})\boldsymbol{Z}_2(\boldsymbol{s})$. By given Proposition 8.5 of (Eaton, 1983) and its proof, we have $\boldsymbol{U}_{22.1}(\boldsymbol{s}) = \boldsymbol{Y}(\boldsymbol{s})^T \begin{bmatrix} \boldsymbol{I}_{\nu-p_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{Y}_1(\boldsymbol{s})^T\boldsymbol{Y}_1(\boldsymbol{s})$, where $\boldsymbol{Y}_1(\boldsymbol{s})$ is the first $n-\nu$ rows of $\boldsymbol{Y}(\boldsymbol{s})$ and $\boldsymbol{Y}_1(\boldsymbol{s}) \sim \mathcal{MN}_{n-\nu,p_2}(\boldsymbol{0}, \boldsymbol{I}_{n-\nu}, \boldsymbol{I}_{p_2}/\nu)$. Hence it is equivalent to have $\boldsymbol{U}_{22.1}(\boldsymbol{s}) = \frac{1}{\nu}\sum_{j=1}^{\nu} \boldsymbol{Z}_{j22.1}(\boldsymbol{s})\boldsymbol{Z}_{j22.1}(\boldsymbol{s})^T$ and $\boldsymbol{Z}_{j22.1}(\boldsymbol{s}) \sim \mathcal{GP}(\boldsymbol{0}, \boldsymbol{\Phi}, \boldsymbol{I}_{p_2})$ since $\boldsymbol{Y}_1(\boldsymbol{s})$ is a normal data matrix with spatial correlation with parameter $\boldsymbol{\Phi}$.

2. **off-diagonals**: From Corollary 4.1, we have the marginals as $\boldsymbol{U}_{21}(\boldsymbol{s})|\boldsymbol{D}_{11}(\boldsymbol{s})\forall \boldsymbol{s} \sim \mathcal{MN}_{p_2,p_1}(\boldsymbol{0}, \boldsymbol{I}_{p_2}/\nu, \boldsymbol{I}_{p_1})$ or $\boldsymbol{U}_{21}(\boldsymbol{s})|\boldsymbol{D}_{11}(\boldsymbol{s})\forall \boldsymbol{s} \sim \mathcal{MN}_{p_2\times p_1}(\boldsymbol{0}, \boldsymbol{I}_{p_2}/\nu \otimes \boldsymbol{I}_{p_1})$.

Also we have

$$\begin{aligned}
&\mathbb{V}(\mathrm{vec}[\boldsymbol{U}_{21}(\boldsymbol{s})], \mathrm{vec}[\boldsymbol{U}_{21}(\boldsymbol{s}')|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \boldsymbol{s}, \jmath) = \\
&\qquad =\mathbb{V}(\mathrm{vec}[\sum_{\jmath=1}^{\nu} \boldsymbol{Z}_{\jmath 2}(\boldsymbol{s})\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})^T], \mathrm{vec}[\sum_{\jmath=1}^{\nu} \boldsymbol{Z}_{\jmath 2}(\boldsymbol{s}'\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}'^T]|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \boldsymbol{s}, \jmath) \\
&\qquad =\sum_{\jmath=1}^{\nu} \mathbb{V}(\mathrm{vec}[\boldsymbol{Z}_{\jmath 2}(\boldsymbol{s})\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})^T], \mathrm{vec}[\boldsymbol{Z}_{\jmath 2}(\boldsymbol{s}'\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}'^T]|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \boldsymbol{s}, \jmath) \\
&\qquad =\sum_{\jmath=1}^{\nu} \mathbb{V}((\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}) \otimes \boldsymbol{I}_{p_2})\boldsymbol{Z}_{\jmath 2}(\boldsymbol{s}), (\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}' \otimes \boldsymbol{I}_{p_2})\boldsymbol{Z}_{\jmath 2}(\boldsymbol{s}'|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \boldsymbol{s}, \jmath) \\
&\qquad =\sum_{\jmath=1}^{\nu} (\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}) \otimes \boldsymbol{I}_{p_2})\mathbb{V}(\boldsymbol{Z}_{\jmath 2}(\boldsymbol{s}), \boldsymbol{Z}_{\jmath 2}(\boldsymbol{s}'|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \boldsymbol{s}, \jmath)(\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}' \otimes \boldsymbol{I}_{p_2})^T \\
&\qquad =\rho(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi}) \sum_{\jmath=1}^{\nu} (\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}) \otimes \boldsymbol{I}_{p_2})(\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}' \otimes \boldsymbol{I}_{p_2})^T/\nu
\end{aligned} \tag{9.2}$$

Followed by the result that $\boldsymbol{U}_{21}(\boldsymbol{s}) = \boldsymbol{D}_{21}(\boldsymbol{s})\boldsymbol{D}_{11}(\boldsymbol{s})^T$, we have

$$\begin{aligned}
&\mathbb{V}(\mathrm{vec}[\boldsymbol{D}_{21}(\boldsymbol{s})], \mathrm{vec}[\boldsymbol{D}_{21}(\boldsymbol{s}')]|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \boldsymbol{s}, \jmath) = \\
&= \mathbb{V}(\boldsymbol{D}_{21}(\boldsymbol{s})\boldsymbol{D}_{11}(\boldsymbol{s})^T[\boldsymbol{D}_{11}(\boldsymbol{s})^T]^{-1}, \boldsymbol{D}_{21}(\boldsymbol{s}'\boldsymbol{D}_{11}(\boldsymbol{s}')^T[\boldsymbol{D}_{11}(\boldsymbol{s}')^T]^{-1}|\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\forall \boldsymbol{s}, \jmath) \\
&= \rho(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})(\boldsymbol{D}_{11}(\boldsymbol{s})^{-1} \otimes \boldsymbol{I}_{p_2}) \sum_{\jmath=1}^{\nu} (\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}) \otimes \boldsymbol{I}_{p_2})(\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}' \otimes \boldsymbol{I}_{p_2})^T(\boldsymbol{D}_{11}(\boldsymbol{s}'^{-1} \otimes \boldsymbol{I}_{p_2})^T/\nu \\
&= \rho(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})\boldsymbol{I}_{p_2} \otimes \frac{1}{\nu}\boldsymbol{D}_{11}(\boldsymbol{s})^{-1} \left[\sum_{\jmath=1}^{\nu} \boldsymbol{Z}_{\jmath 1}(\boldsymbol{s})\boldsymbol{Z}_{\jmath 1}(\boldsymbol{s}')^T\right] [\boldsymbol{D}_{11}(\boldsymbol{s}')^{-1}]^T
\end{aligned} \tag{9.3}$$

□

Joint Distribution of off-diagonals

$$
\begin{aligned}
t_{i21}|[Z_{\jmath 1}(\boldsymbol{s})\ \forall \boldsymbol{s},\jmath] &\sim \mathcal{GP}(d_{i11}(\boldsymbol{s})\boldsymbol{X}_i\boldsymbol{\beta}_{21}(\boldsymbol{s}),\\
&\mathcal{C}(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi}_u)\frac{1}{\nu}\frac{\sum_{\jmath=1}^{\nu} Z_{\jmath 1}(\boldsymbol{s})Z_{\jmath 1}(\boldsymbol{s}')}{d_{i11}(\boldsymbol{s})d_{i11}(\boldsymbol{s}')}e^{\sum_{\boldsymbol{w}\in\{\boldsymbol{s},\boldsymbol{s}'\}}\boldsymbol{X}_i\boldsymbol{\beta}_{22}(\boldsymbol{w})\frac{1}{2}})\\
t_{i31}|[Z_{\jmath 1}(\boldsymbol{s})\ \forall \boldsymbol{s},\jmath; Z_{\jmath 2}(\boldsymbol{s})\ \forall \boldsymbol{s},\jmath] &\sim \mathcal{GP}(d_{i11}(\boldsymbol{s})\boldsymbol{X}_i\boldsymbol{\beta}_{31}(\boldsymbol{s}) + d_{i21}(\boldsymbol{s})\boldsymbol{X}_i\boldsymbol{\beta}_{32}(\boldsymbol{s}),\\
&\mathcal{C}(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi}_u)\frac{1}{\nu}\frac{\sum_{\jmath=1}^{\nu} Z_{\jmath 1}(\boldsymbol{s})Z_{\jmath 1}(\boldsymbol{s}')}{d_{i11}(\boldsymbol{s})d_{i11}(\boldsymbol{s}')}e^{\sum_{\boldsymbol{w}\in\{\boldsymbol{s},\boldsymbol{s}'\}}\boldsymbol{X}_i\boldsymbol{\beta}_{33}(\boldsymbol{w})\frac{1}{2}})\\
t_{i32}|[Z_{\jmath 2}(\boldsymbol{s})\ \forall \boldsymbol{s},\jmath] &\sim \mathcal{GP}(d_{i22}(\boldsymbol{s})\boldsymbol{X}_i\boldsymbol{\beta}_{32}(\boldsymbol{s}),\\
&\mathcal{C}(\boldsymbol{s},\boldsymbol{s}'|\boldsymbol{\Phi}_u)\frac{1}{\nu}\frac{\sum_{\jmath=1}^{\nu} Z_{\jmath 2}(\boldsymbol{s})Z_{\jmath 2}(\boldsymbol{s}')}{d_{i22}(\boldsymbol{s})d_{i22}(\boldsymbol{s}')}e^{\sum_{\boldsymbol{w}\in\{\boldsymbol{s},\boldsymbol{s}'\}}\boldsymbol{X}_i\boldsymbol{\beta}_{22}(\boldsymbol{w})\frac{1}{2}})
\end{aligned}
\tag{9.4}
$$

Computational Details

In this section, we provide essential computational details.

MCMC

We use MCMC for sampling posteriors of interest. Except for $\boldsymbol{\beta}_{kl}(\boldsymbol{s}), k > l$ which are sampled by Gibbs sampling, all other posteriors are sampled using Metropolis Hasting. We just describe the details of sampling $\boldsymbol{\beta}_{kl}(\boldsymbol{s}), k > l$ and $\boldsymbol{\beta}_{kk}(\boldsymbol{s})$. Others are expected to be self-explored in codes provided.

Metropolis Hasting: $\boldsymbol{\beta}_{kk}(\boldsymbol{s})$,: We apply Metropolis hasting and the posteri-

ors are derived as below:

$$
\begin{aligned}
&\pi(\beta_{11}(s)|\beta_{11}(-s), \beta_{-(11)}(s)\forall s, t_{ij}(s)\forall\ i,j,s,\mathbf{\Phi}_w;\mathbf{X}) \propto \\
&\ \pi(\mathbf{t}_{11}|\nu,\beta_{11}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ \pi(\mathbf{t}_{21}|\mathbf{t}_{11},\nu,\beta_{11}(s)\forall s,\beta_{21}(s)\forall s,\beta_{22}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ \pi(\mathbf{t}_{31}|\mathbf{t}_{11},\nu,\beta_{11}(s)\forall s,\beta_{31}(s)\forall s,\beta_{33}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ p(\beta_{11}(s)) \\
\\
&\pi(\beta_{22}(s)|\beta_{2}(-s), \beta_{-(22)}(s)\forall s, t_{ij}(s)\forall\ i,j,s,\mathbf{\Phi}_w;\mathbf{X}) \propto \\
&\ \pi(\mathbf{t}_{22}|\nu,\beta_{22}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ \pi(\mathbf{t}_{21}|\mathbf{t}_{11},\nu,\beta_{11}(s)\forall s,\beta_{21}(s)\forall s,\beta_{22}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ \pi(\mathbf{t}_{32}|\mathbf{t}_{22},\nu,\beta_{22}(s)\forall s,\beta_{32}(s)\forall s,\beta_{33}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ \pi(\mathbf{t}_{31}|\mathbf{t}_{11},\nu,\beta_{11}(s)\forall s,\beta_{31}(s)\forall s,\beta_{33}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ p(\beta_{22}(s)) \\
\\
&\pi(\beta_{33}(s)|\beta_{33}(-s), \beta_{-(33)}(s)\forall s, t_{ij}(s)\forall\ i,j,s,\mathbf{\Phi}_w;\mathbf{X}) \propto \\
&\ \pi(\mathbf{t}_{33}|\nu,\beta_{33}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ \pi(\mathbf{t}_{31}|\mathbf{t}_{11},\nu,\beta_{11}(s)\forall s,\beta_{31}(s)\forall s,\beta_{33}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ \pi(\mathbf{t}_{32}|\mathbf{t}_{22},\nu,\beta_{11}(s)\forall s,\beta_{32}(s)\forall s,\beta_{33}(s)\forall s,\mathbf{\Phi}_w;\mathbf{X}) \\
&\ p(\beta_{33}(s))
\end{aligned}
\tag{9.5}
$$

where $\mathbf{t}_{pq}$ represents $\{t_{pq_i}(s), \forall i, s\}$. Some tricks can be adopted for facilitating the computing such as single-site Metropolis-Hasting sampling of posteriors of $\boldsymbol{\beta}_{kk}(\boldsymbol{s})$ by only computing correlated proportional parts for acceptance rate.

Gibbs Sampling: $\boldsymbol{\beta}_{kl}(\boldsymbol{s}), k > l$: The posteriors of off diagonals are derived as

below:

$$\begin{aligned}
&\left[\boldsymbol{\beta}_{21}(s_1)^T, ..., \boldsymbol{\beta}_{21}(s_n)^T\right]^T | \left[\boldsymbol{\Phi}_u, \nu, \boldsymbol{\beta}_{22}(s)\forall s; X_i \forall i, t_{11_i}(s), t_{21_i}(s) \forall i, s\right] \\
&\qquad \sim \mathcal{N}_{(p)s_n}\Bigg(\left[\sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{22}(s)\forall s)^{-1} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})\right]^{-1} \\
&\qquad\qquad \sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{22}(s)\forall s)^{-1} \boldsymbol{t}_{21_i}, \\
&\qquad\qquad \frac{\left[\sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{22}(s)\forall s)^{-1} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})\right]^{-1}}{\nu}\Bigg)
\end{aligned}$$

$$\begin{aligned}
&\left[\boldsymbol{\beta}_{31}(s_1)^T, ..., \boldsymbol{\beta}_{31}(s_n)^T\right]^T | \left[\boldsymbol{\Phi}_u, \nu, \boldsymbol{\beta}_{33}(s)\forall s; X_i \forall i, t_{11_i}(s), t_{21_i}(s), t_{22_i}(s), t_{31_i}(s) \forall i, s\right] \\
&\qquad \sim \mathcal{N}_{(p)s_n}\Bigg(\left[\sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{33}(s)\forall s)^{-1} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})\right]^{-1} \\
&\qquad\qquad \sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{33}(s)\forall s)^{-1} [\boldsymbol{t}_{31_i} - \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{21_i})\tilde{\boldsymbol{\beta}}_{32}], \\
&\qquad\qquad \frac{\left[\sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{33}(s)\forall s)^{-1} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})\right]^{-1}}{\nu}\Bigg)
\end{aligned}$$

$$\begin{aligned}
&\left[\boldsymbol{\beta}_{32}(s_1)^T, ..., \boldsymbol{\beta}_{32}(s_n)^T\right]^T | \left[\boldsymbol{\Phi}_u, \nu, \boldsymbol{\beta}_{33}(s)\forall s; X_i \forall i, t_{11_i}(s), t_{21_i}(s), t_{22_i}(s), t_{31_i}(s), t_{32_i}(s) \forall i, s\right] \\
&\qquad \sim \mathcal{N}_{(p)s_n}\Bigg(\left[\sum_{i=1}^{N} [\boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{22_i} + \boldsymbol{d}_{12_i})]^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{22}(s)\forall s)^{-1} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{22_i} + \boldsymbol{d}_{12_i})\right]^{-1} \\
&\qquad\qquad \sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{22_i} + \boldsymbol{d}_{12_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{22}(s)\forall s)^{-1} [\boldsymbol{t}_{31_i} + \boldsymbol{t}_{32_i} - \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{11_i})\tilde{\boldsymbol{\beta}}_{31}], \\
&\qquad\qquad \frac{2\left[\sum_{i=1}^{N} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{22_i} + \boldsymbol{d}_{12_i})^T \boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{22}(s)\forall s)^{-1} \boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{22_i} + \boldsymbol{d}_{12_i})\right]^{-1}}{\nu}\Bigg)
\end{aligned} \tag{9.6}$$

where

$$\boldsymbol{Q}(\boldsymbol{X}_i, \boldsymbol{d}_{kk_i}) = \begin{bmatrix} d_{kk_i}(s_1)\boldsymbol{X}_i & & & \\ & d_{kk_i}(s_2)\boldsymbol{X}_i & & \\ & & \dots & \\ & & & d_{kk_i}(s_n)\boldsymbol{X}_i \end{bmatrix}_{n \times n(p)}$$

, a block diagonal matrix,

$$\boldsymbol{W}(\boldsymbol{\Phi}_u; \boldsymbol{X}_i; \boldsymbol{\beta}_{kk}(s) \forall s)$$
$$= \text{diag}\left(e^{\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(s_1)\frac{1}{2}}, ..., e^{\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(s_n)\frac{1}{2}}\right) \ (\boldsymbol{R}(\boldsymbol{\Phi}_u) \circ \boldsymbol{R}(\boldsymbol{\Phi}_u)) \ \text{diag}\left(e^{\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(s_1)\frac{1}{2}}, ..., e^{\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(s_n)\frac{1}{2}}\right)$$

, $\boldsymbol{R}(\boldsymbol{\Phi}_u)$ is the covariance matrix defiend by $\boldsymbol{\Phi}_u$, $\{d_{ikl}\}$ are the Cholesky decomposition of $\boldsymbol{U}_i$ and can be computed by solving a function of $\{d_{ikl}\}$ and $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}_{kl}$ is the stack vector.

Transitional Probability, Parameter Tuning and Mixing

We use Gaussian random walk as transitional probability. The transition steps are tuned by guaranteeing the acceptance rate around 30%, at least for $\boldsymbol{\beta}$ which are important in inference.

For ensuring the good-mixing of the MCMC, we also visualize the MCMC trace and available at XXX

Codes

All codes are available at NCSU github: XXX

References

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics* **3** 103–107.

Basser, P. J., Mattiello, J. and LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B* **103** 247–254.

Clark, D. L., Boutros, N. N. and Mendez, M. F. (2010). *The brain and behavior: an introduction to behavioral neuroanatomy*. Cambridge University Press.

Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association* **111** 800–812.

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika* **68** 265–274.

Donnelly, C. and Embrechts, P. (2010). The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *ASTIN Bulletin: The Journal of the IAA* **40** 1–33.

Dryden, I. L., Koloydenko, A. and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics* 1102–1123.

Eaton, M. L. (1983). Multivariate statistics: A vector space approach. *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1983, 512*.

Ferro-Famil, L., Pottier, E. and Lee, J.-S. (2001). Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing* **39** 2332–2342.

Froeling, M., Pullens, P. and Leemans, A. (2016). Dti analysis methods: Region of interest analysis. In *Diffusion Tensor Imaging* 175–182. Springer.

Gelfand, A. E., Kim, H.-J., Sirmans, C. and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* **98** 387–396.

Gray, H. (1918). *Anatomy of the human body*. Lea & Febiger.

Gupta, A. K. and Nagar, D. K. (1999). *Matrix variate distributions* **104**. CRC Press.

Jensen, D. R. et al. (1970). The joint distribution of traces of Wishart matrices and some applications. *The Annals of Mathematical Statistics* **41** 133–145.

Käärik, E. and Käärik, M. (2009). Modeling dropouts by conditional distribution, a copula-based approach. *Journal of Statistical Planning and Inference* **139** 3830–3835.

Krishnaiah, P. and Rao, M. (1961). Remarks on a multivariate gamma distribution. *American Mathematical Monthly* 342–346.

Lane, S. D., Steinberg, J. L., Ma, L., Hasan, K. M., Kramer, L. A., Zuniga, E. A., Narayana, P. A. and Moeller, F. G. (2010). Diffusion tensor imaging and decision making in cocaine dependence. *PLoS One* **5** e11591.

Lee, H. N. and Schwartzman, A. (2017). Inference for eigenvalues and eigenvectors in exponential families of random symmetric matrices. *Journal of Multivariate Analysis* **162** 152–171.

Lo, C.-Y., Wang, P.-N., Chou, K.-H., Wang, J., He, Y. and Lin, C.-P. (2010). Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. *Journal of Neuroscience* **30** 16876–16885.

Ma, L., Hasan, K. M., Steinberg, J. L., Narayana, P. A., Lane, S. D., Zuniga, E. A., Kramer, L. A. and Moeller, F. G. (2009). Diffusion tensor imaging in cocaine dependence: regional effects of cocaine on corpus callosum and effect of cocaine administration route. *Drug and alcohol dependence* **104** 262–267.

Ma, L., Steinberg, J. L., Wang, Q., Schmitz, J. M., Boone, E. L., Narayana, P. A. and Moeller, F. G. (2017). A preliminary longitudinal study of white matter alteration in cocaine use disorder subjects. *Drug and Alcohol Dependence* **173** 39–46.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1980). Multivariate analysis (probability and mathematical statistics).

Myers, R. H. and Montgomery, D. C. (1997). A tutorial on generalized linear models. *Journal of Quality Technology* **29** 274.

Peizer, D. B. and Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I. *Journal of the American Statistical Association* **63** 1416–1456.

Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, II. *Journal of the American Statistical Association* **63** 1457–1483.

Schwartzman, A. (2006). Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data PhD thesis, Stanford University.

Schwartzman, A., Mascarenhas, W. F. and Taylor, J. E. (2008). Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *The Annals of Statistics* 2886–2919.

Smith, P. J. and Garth, L. M. (2007). Distribution and characteristic functions for correlated complex Wishart matrices. *Journal of Multivariate analysis* **98** 661–677.

Soares, J. M., Marques, P., Alves, V. and Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Frontiers in neuroscience* **7**.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging.* Springer Science & Business Media.

Tripathi, G. (1999). A matrix extension of the Cauchy-Schwarz inequality. *Economics Letters* **63** 1–3.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)* 297–312.

Viraswami, K. (1991). On multivariate gamma distributions.