Spatial Autoregressive Model for von-Mises Fisher Distributed Principal Diffusion Directions

Zhou Lan∗
Yale University School of Medicine
Arkaprava Roy
Department of Biostatistics, University of Florida

July 19, 2022

Abstract

The principal diffusion directions are one of the most important statistics derived from diffusion tensor imaging (DTI). It is directional data that depict the anatomical structures of brain tissues. However, only a few approaches are available for covariate-dependent statistical modeling of principal diffusion directions. We thus propose a novel spatial autoregressive model by assuming that the principal diffusion directions are von-Mises Fisher (vMF) distributed directional data. Using a novel link function relying on transformation between Cartesian coordinates and spherical coordinates, we regress the vMF distributed principal diffusion directions on the subject’s covariates, measuring how the clinical factors affect the anatomical structures. The spatial residual dependence along fibers is captured by an autoregressive model. Key statistical properties of the model and a comprehensive toolbox for Bayesian inference of the directional data with applications to medical imaging analysis are thoroughly developed. The numerical studies based on synthetic data demonstrate that our model has more accurate estimation of the effects of clinical factors. Applying our regression model to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) data, we obtain new insights.

∗Corresponding Author: Zhou Lan (zhou.lan@yale.edu)
Keywords: Alzheimer’s Disease Neuroimaging Initiative; Diffusion Tensor Imaging; Directional Statistics; Spatial Autogressive Model; von-Mises Fisher Distribution
1 Introduction

Tissue micro-structure of human brain is an important medical characteristic in clinical and surgical anatomy. With the development of neuroimaging techniques, diffusion tensor imaging (DTI) has become a powerful tool to measure the structures in a non-invasive way (Soares et al., 2013). The rich use of DTI in brain science triggers many meaningful interdisciplinary studies. For example, using DTI, we measure the voxel-by-voxel white matter tracts within a tissue of the human brain. The depiction at the left panel of Figure 1 is the anatomical structure measured at a voxel. The movement of water molecules can be characterized by giving their respective diffusion ellipsoids (see the right panel Figure 1). Diffusivity of the water molecule is quantified by the diffusion coefficient, a $3 \times 3$ positive definite matrix. The eigenvectors of the diffusion coefficients ($[E_1 \ E_2 \ E_3]$) reveals 3D orthogonal directions of water diffusion; The squared roots of the eigenvalues $[l_1 \ l_2 \ l_3]$ are associated with corresponding semi-axis lengths (Zhou, 2010, Section 1.2.3). The diffusion tensor coefficient $D = [E_1 \ E_2 \ E_3]\text{diag}(l_1, l_2, l_3)[E_1 \ E_2 \ E_3]^T$ describes the anatomical structure at the voxel. In this way, we obtain an image whose voxel-wise variables are positive definite diffusion tensors, describing the anatomical structure of a tissue. In addition, there are several other summary features (e.g., fractional anisotropy) that are also derived from the diffusion tensors for different clinical applications.

There are a variety of clinical usage of DTI data. In recent years, several methodological approaches are introduced to handle the statistical randomness of the diffusion tensors (e.g., Schwartzman et al., 2008; Zhu et al., 2009; Yuan et al., 2012; Lee and Schwartzman, 2017; Lan et al., 2021). For example, Schwartzman et al. (2008) adopted Gaussian distribution for symmetric matrices to model diffusion tensor and develop statistical inference tools for eigenvalues and eigenvectors when samples are drawn from that distribution. Following
Schwartzman et al. (2008), Zhu et al. (2009), Yuan et al. (2012) further introduce intrinsic regression models and polynomial regression models. Adopting matrix-Gamma distribution or Wishart distribution as an alternative random distribution for diffusion tensors, Lee and Schwartzman (2017) further provide the inference for eigenvalues and eigenvectors. Recently, Lan et al. (2021) proposed a spatial random process based on Wishart distribution to characterize spatial modeling of diffusion tensors.

White matter alignment of the brain is investigated in several clinical applications. Among various markers for white matter fibers, one of the important marker is the DTI-derived principal diffusion direction which is the principal eigenvector, $E_1$. They are interpreted as tangent directions along fiber bundles at the corresponding voxel (Figure 2). The estimated diffusion directions are then used as an input for tractography algorithms to reconstruct fiber tracts (Wong et al., 2016) and obtain white matter structural connectivity profiles. For clinical applications, low-resolution summary of networks, called connectome are usually constructed using these connectivity profiles considering some parcellation of the brain and their associations with subject’s covariates are studied (Roy et al., 2019).
Zhang et al. (2019); Guha and Rodriguez (2020). However, such summarization may often be crude, which could lead to inefficient statistical inference. In this paper, we thus study the association between the principal diffusion directions and subject level covariates (e.g., age, gender, disease status) directly to reveal the factors driving the brain’s anatomical structures. To our best knowledge, only a limited number of methods are developed to investigate such association.

Figure 2: The left panel gives a human brain in terms of fiber tracts. The middle panel is the fiber tract of interest. The right panel gives how fiber tracts are constructed. The principal eigenvalues of diffusion tensors are given in each voxel. The voxels with smallest separation angles are considered as a fiber tract, e.g., the voxels linked by the blue curve.

The challenges in modeling principal diffusion directions are similar to the existing works on modeling diffusion tensors, as they are both on manifold. In multivariate statistics, the principal diffusion directions belong to the class of directional data statistics (Mardia, 2014). The directional data of dimension $p$ is on the $(p-1)$-dimensional sphere, denoted as $S^{p-1}$ (Mardia and Jupp, 2009, Section 9.3.2). Since the most commonly used multivariate distributions (e.g., multivariate normal distribution) are supported on the Euclidean space, they cannot adequately characterize directional data, and thus it sacrifices both geometric interpretability and statistical reliability. To tackle this issue, we propose our statistical model relying on von Mises–Fisher (vMF) distribution (Mardia, 1975). This is a classical

---

1Since we focus on the applications to DTI in this paper, we use $p = 3$ in the rest of the paper.
distribution to model directional data, providing parsimonious parameterization to quantify
the mode direction and its corresponding variation. Relying on a vMF distributed error
model, we propose a spatial generalized linear model to inferring the effects of covariates on
diffusion directions. We consider a so-called scaled Cartesian-spherical link function relying
on transformation between Cartesian and spherical coordinates. This allows us to project
the diffusion direction on the Euclidean space. This innovative link function also enjoys
the desired monotonic property and provides a platform to regress the diffusion direction
on the subject-specific covariates.

Furthermore, spatial dependence is a very important characteristic in neuroimaging
applications (e.g., [Reich et al. 2018], [Lan et al. 2021]). It helps to capture more potential
variations. The key step can be to construct a spatial correlation function on the diffusion
direction space, as discussed in [Kang and Li 2016]. In the previous works (e.g., [Wong
et al. 2016], [Lan et al. 2021]), the spatial dependence based on Euclidean distance was
implemented. One important feature of our proposed methodology is that we incorporate
the fiber tractography information into consideration. Thus, the spatial variation of the
principal diffusion directions is assumed auto-correlated along a fiber from the beginning
to the end. To accommodate this special type of spatial dependence, we consider an
autoregressive model to induce the sequential dependence along the fiber, suggested in
several discussions ([Zhu et al. 2011], [Kang and Li 2016]).

Our methodological development is primarily motivated by Alzheimer’s Disease Neu-
roimaging Initiative (ADNI) ([Mueller et al. 2005]) study. Based on their cognitive perfor-
mance, the subjects are categorized into different clinical groups, namely, healthy controls
(CN), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI),
and Alzheimer’s Disease (AD). The subject specific covariates (e.g., age, gender, and ge-
etic information) may have heterogeneous effects on the response. We thus specify the
regression model with group-specific coefficients. Furthermore, due to the manifold nature of the data, we develop a novel toolbox for Bayesian angular inference of principal diffusion direction, addressing the key questions in a clinical study. We validate the performance of our model by using both real ADNI data and the synthetic data which is generated mimicking the ADNI dataset. Our proposed model registers an overwhelmingly better performance than other competing models. This demonstrates the importance of our proposal. Our application to ADNI data also reveals several insightful scientific findings using our toolbox of Bayesian angular inference for principal diffusion direction.

To our best knowledge, this is the very first paper to propose a spatial generalized regression model for the unit vector-valued response. The rest of the paper is organized as follows. In Section 2 we provide the details of our motivating data. Driven by the motivating data, we provide our methodology in Section 3. The model comparison is given in Section 5. Finally, we apply our proposed model to the ADNI dataset in Section 6 and end with some concluding remarks in Section 7. All the supplementary sections are summarized in Supplementary Materials.

2 ADNI Data

Our proposed methodology is motivated by ADNI data (Mueller et al., 2005). ADNI is a multi-site study that aims to improve clinical trials for the prevention and treatment of Alzheimer’s disease. Scientists at 63 sites in the US and Canada track the progression of AD in the human brain, and diffusion tensor imaging is one of their measures. In this paper, we focus on ADNI-2 which starts on September 2011 lasts five years (Aisen et al., 2010). We randomly selected 30 subjects from the groups of CN, EMCI, LMCI, and AD to create our study cohort. Let $i = 1 : N_g$ denote the number of subjects in the clinical
group $g$, where $N_g = 30$ and $g \in \{\text{CN, EMCI, LMCI, AD}\}$ for our data analysis. For each subject, we collect the subject-level information including age, gender, mini mental state examination (MMSE) score, and Apolipoprotein E (APOE) information. Age and gender are basic demographics. The MMSE is a performance based neuropsychological test score whose value ranges between 0-30; a subject with Alzheimer’s disease is usually associated with lower MMSE. APOE is polymorphic with three major alleles ($\epsilon$-2, $\epsilon$-3, and $\epsilon$-4). Generally, the $\epsilon$-4 variant was the largest known genetic risk factor for Alzheimer’s disease in a variety of ethnic groups (Sadigh-Eteghad et al., 2012). Therefore, we use APOE-4 as a binary indicator showing whether a subject has $\epsilon$-4 variant or not.

![Fornix and Corpus Callosum Attractography Atlases](image)

Figure 3: Tractography atlase of fornix (left) and and corpus callosum (right).

Many recent studies reveal the brain fiber tracts associated with cognitive performance play importance roles in the progression of Alzheimer’s disease. Among these fiber tracts, fornix (Oishi et al., 2012; Nowrangi and Rosenberg, 2015) and corpus callosum (Teipel et al., 2002; Di Paola et al., 2010) play important roles. In terms of brain anatomy, the fornix is a C-shaped bundle of nerve fibers in the brain; the corpus callosum is thick nerve tract, consisting of a flat bundle of commissural fibers, beneath the cerebral cortex in the brain. In this paper, we focus on these two tracts and consider the tractography atlases based on Yeh et al. (2018) (see Figure 3). In each fiber tract, there are $K$ fibers tracked.
by a fiber tracking algorithm. From the tractography atlases of a given fiber tracts (e.g., fornix or corpus callosum), we identify \( j = 1 : J_k \) voxels from a fiber starting one end of the \( k \)-th fiber to the other end.

3 vMF Regression for Principal Diffusion Directions

For the \( i \)-th subject of \( g \)-th clinical group, we use \( E_{gikj} \) to denote the corresponding principal diffusion direction measured at \( j \)-th voxel on the \( k \)-th fiber. We let \( X_{ig} \) to denote the design matrix containing covariate information (including intercept) of the \( i \)-th subject of \( g \)-th clinical group. To tackle the clinical problem that how the covariate effects drive the variation of principal diffusion directions, we provide the methodology in this section which regresses the principal diffusion direction on in sphere space on the subject’s covariates in Euclidean space.

3.1 vMF Distribution

The principal diffusion directions \( E_{gikj} \) are on \( S^2 \) by definition. To accommodate this, we let \( E_{gikj} \) to follow vMF distribution, a popular probability distribution to characterize the randomness of the directional data (Mardia, 1975). The probability density function of vMF distributed \( E_{gikj} \) (see Mardia and Jupp, 2009, Equation 9.3.4) is

\[
\int_{E_{gikj}} e_{gikj} | \mu_{gikj}, \kappa \rangle = C_3(\kappa) \exp \left( \kappa \mu_{gikj}^T e_{gikj} \right),
\]

\[
C_3(\kappa) = \frac{\kappa}{2 \pi \left( e^\kappa - e^{-\kappa} \right)} \parallel \mu_{gik}(j) \parallel = 1,
\]

(1)
denoted as \( E_{gikj} \sim \text{vMF}(\mu_{gikj}, \kappa) \) and \( \parallel \cdot \parallel \) stands for the \( \ell_2 \) norm.

In the above density function, the principal diffusion direction \( e_{gikj} \) only contributes to the term \( \mu_{gikj}^T e_{gikj} \). This term is essentially \( \cos \delta(\mu_{gikj}, e_{gikj}) \) where \( \delta(\mu_{gikj}, e_{gikj}) \) is the
separation angle between unit vectors $\mu_{igkj}$ and $e_{gikj}$. This implies that $\mu_{igkj}$ is the mode direction since $e_{gikj} = \mu_{igkj}$ maximizes the likelihood. The likelihood increases as $\mu_{igkj}$ and $e_{gikj}$ has a smaller angle. The density function (Equation 1) is maximized at $\mu_{gikj}$ and minimized at $-\mu_{gikj}$. The concentration parameter $\kappa$ controls the concentration of the distribution around the mode direction $\mu_{gikj}$. To be specific, the tangential component $(I - \mu_{gikj}\mu_{gikj}^T)e_{gikj}$ of $E_{gikj}$ is a vector describing whether $E_{gikj}$ is concentrated at the mode direction $\mu_{gikj}$ closely or not. It converges in probability to 0 as $\kappa \to \infty$ (Mardia and Jupp 2009, Equation 9.3.15) (see Figure 4).

Figure 4: The illustration of concentration parameter $\kappa$. In Panel (a), the yellow arrow represents the mode direction $\mu_{igkj}$; the dashed blue arrows represent the confidence regions $C$ where $Pr(E_{gikj} \in C) = 1 - \alpha$; As $\kappa \to \infty$, the region $C$ becomes narrow. In Panel (b), the yellow arrow represents the mode direction $\mu_{gikj}$, the blue arrow represents the random vector $E_{igkj}$, and the green arrow represents the $R_{igkj} \in S^2$. $R_{igkj}$ and $\mu_{igkj}$ are orthogonal to each other. the tangential component $(I - \mu_{gikj}\mu_{gikj}^T)e_{gikj}$ of $E_{gikj}$, a vector describing whether $\mu_{gikj}$ is concentrated around the mode direction $\mu_{gikj}$ closely.

---

2In many textbooks (e.g., Mardia and Jupp 2009), the term $\mu_{gikj}$ in the vMF distribution is named as mean direction. However, we think mode direction can be more appropriate in describing the nature of $\mu_{gikj}$. 

---

10
3.2 Linking to Predictors

Modeling covariate effect and spatial dependence are easier in the Euclidean space. However, directional data lies on a manifold. Therefore, it is not optimal to represent directional data using the Cartesian coordinate system. Instead, we use spherical coordinates to obtain a transparent representation of the directions where the new set of parameters (Azimuth angle and elevation angle) are supported on a Euclidean space. Two new parameters, Azimuth angle \( \theta \in [-\pi, \pi] \) and Elevation angle \( \phi \in [-\frac{\pi}{2}, \frac{\pi}{2}] \) are the two inputs in the spherical coordinates (see Figure 5), where Azimuth angle \( \theta \) is the counterclockwise angle in the x-y plane measured from the positive x-axis and the Elevation angle \( \phi \) is the elevation angle from the x-y plane. Let \((x, y, z)\) be the three inputs in the Cartesian coordinates, the transformation between Cartesian and spherical coordinates is

\[
\theta = \arctan2(y, x) \\
\phi = \arctan2(z, \sqrt{x^2 + y^2}) = \arctan2(z, \sqrt{1 - z^2}),
\]

where \( \arctan2(y, x) = \lim_{z \to x + \pi} \left( \frac{y}{z} + \frac{\pi}{2} \text{sgn}(y) \text{sgn}(x)(\text{sgn}(x) - 1) \right) \). We use \( u(x, y, z) = [\theta, \phi] \in [-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] \) to denote this projection. By scaling the two radians (\( \theta \) and \( \phi \)) to (0, 1), we further use the logit function to project the values to the real line, denoted as, \( g(\theta, \phi) = \left[ \logit\left( \frac{\theta + \pi}{2\pi} \right), \logit\left( \frac{\phi + \pi/2}{\pi} \right) \right] = [\tilde{\theta}, \tilde{\phi}] \in \mathbb{R} \times \mathbb{R} \). We use \( \ell(x, y, z) = [\tilde{\theta}, \tilde{\phi}] \) to denote this innovative function which project the directional data in \( S^2 \) to \( \mathbb{R} \times \mathbb{R} \).

We implement \( \ell(\cdot) \) as the link function in generalized linear models (Dobson and Barnett, 2018, Section 3.4), that is \( \ell\left( \mu_{gikj} \right) = [\tilde{\theta}_{gikj}, \tilde{\phi}_{gikj}] \). The term \([\tilde{\theta}_{gikj}, \tilde{\phi}_{gikj}]\) is referred to as prediction terms, and it is specified as a function of the covariate \( X_{ig} \):

\[
E[\tilde{\theta}_{gikj}, \tilde{\phi}_{gikj}]^T = [X_{ig} \alpha_{gkj}, X_{ig} \beta_{gkj}]^T
\]

to capture the group-specific covariate effects. Lemma 1 shows that our proposed link function \( \ell(\cdot) \) is a bijection function, and this bijection link function provides the correspondent relationship between the linear predictor \([\tilde{\theta}_{gikj}, \tilde{\phi}_{gikj}]^T\) and the mode direction \( \mu_{gikj} \). Like any other generalized linear models, \( \ell(\cdot) \) is a link function which allows learning the covariate effects to the response variables.
Lemma 1. The function $\ell(\cdot) : S^2 \rightarrow \mathbb{R} \times \mathbb{R}$ is bijective.

Proof. We know $\ell(\cdot) := u(\cdot) \circ g(\cdot)$ where $\circ$ is function composition. Since $u(\cdot)$ and $g(\cdot)$ are both bijective, $\ell(\cdot)$ is bijective.

3.3 Autoregressive Modeling

Incorporating spatial dependence while analyzing neuroimaging data is usually important to achieve efficient statistical inference. Different from other neuroimaging applications (e.g., Reich et al., 2018), the key step here is to construct a spatial correlation function supported on the principal diffusion direction space, as discussed in Kang and Li (2016). In the previous works, the spatial dependence based on Euclidean distance was employed. These works enjoy traditional geostatistical modeling of spatial statistics (Wong et al., 2016; Lan et al., 2021), but may be suboptimal if the analysis is within the fiber tracts-based regions of interests.
The spatial profiling of the DTI statistics along a fiber tract reveals that the spatial dependence depends on their geodesic distance along a fiber, but not their Euclidean distance (Wong et al., 2007; Goodlett et al., 2009; Zhu et al., 2011). For example, Wong et al. (2007) shows that there is a spatial dependency of diffusion parameters along the corticospinal tract in healthy individuals. Goodlett et al. (2009) and Zhu et al. (2011) further propose to induce spatial dependence considering arc length distances, computed relative to a fixed end point of the fiber bundle while modeling the scalar diffusion properties (e.g., fractional anisotropy, mean diffusivity). In Figure A.1 in Section A of supplementary materials, we visualize $\ell(E_{gikj})$ of typical fibers to endorse this dependence. Following these works, we also incorporate the spatial dependence depending on their geodesic distance along a fiber.

By giving additive spatial residual terms, the link function $\ell(\cdot)$ further eases to induce residual spatial dependence to capture spatial variation. Preserving $E[\tilde{\theta}_{gikj}, \tilde{\varphi}_{gikj}]^T = [X_{ig\alpha_{gkj}}, X_{ig\beta_{gkj}}]^T$ but inducing the possible spatial dependence, we decompose $[\tilde{\theta}_{gikj}, \tilde{\varphi}_{gikj}]^T$ by adding two spatial residuals ($\epsilon_{gikj}$ and $\xi_{gikj}$) such as $[\tilde{\theta}_{gikj}, \tilde{\varphi}_{gikj}]^T = [X_{ig\alpha_{gkj}} + \epsilon_{gikj}, X_{ig\beta_{gkj}} + \xi_{gikj}]^T$. The sequential dependence via spatial profiling of principal diffusion direction motivates us to induce the spatial dependence via an autoregressive characterization on the two residual terms $\epsilon_{gikj}$ and $\xi_{gikj}$ to capture the corresponding spatial variations, where the joint densities of $[\epsilon_{gik1}, ..., \epsilon_{gikj}, ..., \epsilon_{gikJ_k}]$ and $[\xi_{gik1}, ..., \xi_{gikj}, ..., \xi_{gikJ_k}]$ are

$$f\left(\epsilon_{gik1}, ..., \epsilon_{gikj}, ..., \epsilon_{gikJ_k}\right) = f\left(\epsilon_{gik1}\right) \prod_{j=2}^{J_k} f\left(\epsilon_{gikj} | \epsilon_{gik(j-1)}, ..., \epsilon_{gik\max(j-1)}\right)$$

$$f\left(\xi_{gik1}, ..., \xi_{gikj}, ..., \xi_{gikJ_k}\right) = f\left(\xi_{gik1}\right) \prod_{j=2}^{J_k} f\left(\xi_{gikj} | \xi_{gik(j-1)}, ..., \xi_{gik\max(j-1)}\right)$$

Through this construction, the residual terms ($\epsilon_{gikj}$ and $\xi_{gikj}$) only rely on the previous $q$ terms along the fiber. Such characterization can be viewed as a special case of Vecchia’s method (Vecchia 1988).
We characterize the conditional densities of the Vecchia decomposition (Equation 2) using an autoregressive model-based framework (Hamilton 2020). For clarity, we first introduce the definitions and notations related to the autoregressive modeling approach. An autoregressive process of order \( q \), AR-\( q \) can be stated formally as in Definition 1:

**Definition 1 (AR-\( q \) Process).** Given \( X_t = X_{t-1}\phi_1 + \ldots + X_{t-q}\phi_q + \epsilon_t \) and \( \epsilon_t \sim \mathcal{N}(0, \sigma^2) \), for \( t \in \mathbb{Z} \). For \( t \in \mathbb{Z} \), we define that \( \{X_t : t \in \mathbb{Z}\} \) follows a \( \text{AR}(q) \) process with correlation parameters \( \Phi(q) = [\phi_1, \ldots, \phi_q] \) and variance \( \sigma^2 \), denoted as \( \{X_t : t \in \mathbb{Z}\} \sim \text{AR}(\Phi(q), \sigma^2) \), where \( \Phi(q) = [\phi_1, \ldots, \phi_q] \).

Given a finite set of indices, \( t = 1, \ldots, T \), the random variables \([X_1, \ldots, X_t, \ldots, X_T]\) in an AR-\( q \) process (Definition 1) follows a mean-zero multivariate normal distribution with variance-covariance matrix as \( \sigma^2 V(\Phi(q)) \), where \( V(\Phi(q)) \) is a positive definite matrix whose \((i,j)\)-th is \( \gamma_{|i-j|} = \text{cov}(X_i, X_j)/\sigma^2 \). The specification of \( \gamma_g \) is (Hamilton 2020, Equation 3.4.36)

\[
\gamma_g = \begin{cases} 
\phi_1\gamma_{g-1} + \phi_2\gamma_{g-2} + \ldots + \phi_q\gamma_{g-q} & \text{for } g = 1, 2, \ldots \\
\phi_1\gamma_1 + \phi_2\gamma_2 + \ldots + \phi_q\gamma_q + \sigma^2 & \text{for } g = 0 
\end{cases}
\]

(3)

Therefore, the AR-\( q \) process (Definition 1) can induce a mean-zero multivariate normal distribution with variance-covariance matrix as \( V(\Phi(q)) \), and we continue use the process notation to denote the induced distribution such as \([X_1, \ldots, X_t, \ldots, X_T] \sim \mathcal{N}(0, \sigma^2 V(\Phi(q))) \).

Due to the autoregressive process construction, the joint likelihood can be written similar to the formulation of the Vecchia’s method (Equation 2). By applying the AR-\( q \) process to our residual terms \( \epsilon_{gikj} \) and \( \xi_{gikj} \), we finally have

\[
[\xi_{gik1}, \ldots, \xi_{gikj}, \ldots, \xi_{gikJk}] \sim \text{AR}(\Phi(q), \sigma^2_\xi),
\]

(4)
where \((\tau_\epsilon^2, \Phi_\epsilon^{(q)})\) and \((\tau_\xi^2, \Phi_\xi^{(q)})\) are the corresponding parameters of the AR-\(q\) models. The above two vectors are independent over \(g = 1 : G, i = 1 : I_g,\) and \(k = 1 : K\).

### 3.4 Prior Settings

To proceed with our Bayesian inference, we put priors on the model parameters. To further induce the spatial dependence of the mean effects and follow the spatial modeling with spatially varying coefficient (Gelfand et al., 2003), we assign an autoregressive process of order \(q\) as the prior on the coefficients, denoted as \([\alpha_{gkj}(c), \ldots, \alpha_{gkj}(C)], \ldots, \alpha_{gkj}(C)\] and \([\beta_{gkj}(c), \ldots, \beta_{gkj}(C)], \ldots, \beta_{gkj}(C)\] and \(C+1\) is the number of covariates (including intercept). The variances are assigned with a weakly informative inverse-gamma priors with shape and rate parameters set to 0.1, denoted as \(\tau_\epsilon^{-2}, \tau_\xi^{-2}, \sigma_\alpha^{-2}, \sigma_\beta^{-2} \sim \mathcal{G}\mathcal{A}(0.1, 0.1)\). We put a diffuse prior for the concentration parameter, \(\kappa\), which can be viewed as nugget effect, denoted as \(\kappa \propto 1\). To guarantee the specification based on autoregressive processes are stationary processes, we utilize Yule-Walker equations (Hamilton, 2020, Equation 3.4.37) to specify the uniform priors ranging from \(-1\) to \(1\) to partial autocorrelations \((\rho_{\epsilon,|j|}, \rho_{\xi,|j|}, \rho_{\alpha,|j|}, \rho_{\beta,|j|})\) instead, that is \(\rho_{\epsilon,|j|} = \phi_{\epsilon,1}\rho_{\epsilon,|j|-1} + \phi_{\epsilon,2}\rho_{\epsilon,|j|-2} + \ldots + \phi_{\epsilon,q}\rho_{\epsilon,|j|-q}\) instead, that is \(\rho_{\epsilon,|j|} = \phi_{\epsilon,1}\rho_{\epsilon,|j|-1} + \phi_{\epsilon,2}\rho_{\epsilon,|j|-2} + \ldots + \phi_{\epsilon,q}\rho_{\epsilon,|j|-q}\sim \mathcal{U}(-1, 1), \rho_{\xi,|j|} = \phi_{\xi,1}\rho_{\xi,|j|-1} + \phi_{\xi,2}\rho_{\xi,|j|-2} + \ldots + \phi_{\xi,q}\rho_{\xi,|j|-q}\sim \mathcal{U}(-1, 1), \rho_{\alpha,|j|} = \phi_{\alpha,1}\rho_{\alpha,|j|-1} + \phi_{\alpha,2}\rho_{\alpha,|j|-2} + \ldots + \phi_{\alpha,q}\rho_{\alpha,|j|-q}\sim \mathcal{U}(-1, 1), \rho_{\beta,|j|} = \phi_{\beta,1}\rho_{\beta,|j|-1} + \phi_{\beta,2}\rho_{\beta,|j|-2} + \ldots + \phi_{\beta,q}\rho_{\beta,|j|-q}\sim \mathcal{U}(-1, 1)\) for \(j = 1 : q\), and \(\rho_{\epsilon,0} = \rho_{\xi,0} = \rho_{\alpha,0} = \rho_{\beta,0} = 1\).

### 3.5 Model Summary

By breaking the above model construction into three generic stages, i.e., data model, process model, and prior model (Gelfand et al., 2010 Chapter 7: Hierarchical Modeling with Spatial...
Data), we can wrap up the proposed spatial autoregressive model for vMF distributed principal diffusion directions for \( g = 1 : G, i = 1 : N_g, k = 1 : K, \) and \( j = 1 : J_k \) as below:

**Data Model:**

\[
f_{E_{gikj}}(\mathbf{e}_{gikj}|\boldsymbol{\mu}_{gikj}, \kappa) = C_3(\kappa) \exp \left( \kappa \mathbf{e}_{gikj}^T \mathbf{u}_{gikj} \right),
\]

\[
C_3(\kappa) = \frac{\kappa}{2\pi (e^\kappa - e^{-\kappa})} \| \mathbf{u}_{gikj} \| = 1,
\]

**Process Model:**

\[
\ell(\boldsymbol{\mu}_{gikj}) = \tilde{\theta}_{gikj}, \tilde{\phi}_{gikj}^T
\]

\[
[\tilde{\theta}_{gikj}, \tilde{\phi}_{gikj}]^T = [X_{ig} \alpha_{gkj} + \epsilon_{gikj}, X_{ig} \beta_{gkj} + \xi_{gikj}]^T
\]

\[
[\epsilon_{gik1}, ..., \epsilon_{gikj}, ..., \epsilon_{gikJ_k}] \sim \mathcal{AR}(\Phi_\epsilon^{(q)}, \sigma_\epsilon^2)
\]

\[
[\xi_{gik1}, ..., \xi_{gikj}, ..., \xi_{gikJ_k}] \sim \mathcal{AR}(\Phi_\xi^{(q)}, \sigma_\xi^2)
\]

**Prior Model:**

\[
[\alpha_{gk1}(c), ..., \alpha_{gkj}(c), ..., \alpha_{gkJ_k}(c)] \sim \mathcal{AR}(\Phi_{\alpha}^{(q)}, \sigma_\alpha^2), \text{ for } c = 0 : C,
\]

\[
[\beta_{gk1}(c), ..., \beta_{gkj}(c), ..., \beta_{gkJ_k}(c)] \sim \mathcal{AR}(\Phi_\beta^{(q)}, \sigma_\beta^2), \text{ for } c = 0 : C,
\]

\[
\alpha_{gkj} = [\alpha_{gkj}(0), ..., \alpha_{gkj}(c), ..., \alpha_{gkj}(C)]
\]

\[
\beta_{gkj} = [\beta_{gkj}(0), ..., \beta_{gkj}(c), ..., \beta_{gkj}(C)]
\]

\[
\tau_\epsilon^{-2}, \tau_\xi^{-2}, \sigma_\alpha^{-2}, \sigma_\beta^{-2} \sim \mathcal{GA}(0.1, 0.1)
\]

\[
\rho_{\epsilon,|j|} = \phi_{\epsilon,1}\rho_{\epsilon,|j-1|} + \phi_{\epsilon,2}\rho_{\epsilon,|j-2|} + ... + \phi_{\epsilon,q}\rho_{\epsilon,|j-q|} \sim \mathcal{U}(-1, 1), \text{ for } j = 1 : q
\]

\[
\rho_{\epsilon,|j|} = \phi_{\xi,1}\rho_{\epsilon,|j-1|} + \phi_{\xi,2}\rho_{\epsilon,|j-2|} + ... + \phi_{\xi,q}\rho_{\epsilon,|j-q|} \sim \mathcal{U}(-1, 1), \text{ for } j = 1 : q
\]

\[
\rho_{\alpha,|j|} = \phi_{\alpha,1}\rho_{\alpha,|j-1|} + \phi_{\alpha,2}\rho_{\alpha,|j-2|} + ... + \phi_{\alpha,q}\rho_{\alpha,|j-q|} \sim \mathcal{U}(-1, 1), \text{ for } j = 1 : q
\]

\[
\rho_{\beta,|j|} = \phi_{\beta,1}\rho_{\beta,|j-1|} + \phi_{\beta,2}\rho_{\beta,|j-2|} + ... + \phi_{\beta,q}\rho_{\beta,|j-q|} \sim \mathcal{U}(-1, 1), \text{ for } j = 1 : q
\]

\[
\rho_{\epsilon,0} = \rho_{\xi,0} = \rho_{\alpha,0} = \rho_{\beta,0} = 1
\]

We use Markov chain Monte Carlo (MCMC) algorithm for sampling from the posterior. The MCMC scheme consists of Metropolis Hastings and Gibbs steps. Based on the posterior samples, we infer the properties of an underlying data mechanism in a Bayesian paradigm. However, the principal diffusion direction data is directional on the sphere,
making the traditional Bayesian inference suboptimal. In the next section, we thus introduce a novel Bayesian angular inference framework, which are motivated to tackle the important scientific questions related to diffusion directions. In all the numerical studies presented in the following sections, we collect 3,000 MCMC samples after discarding 2,000 as burn-in for Bayesian inference. The implementation codes are attached in Section C of supplementary materials.

4 Bayesian Angular Inference for Principal Diffusion Directions

In this section, we introduce our novel Bayesian angular inference for principal diffusion directions. The proposed inference framework not only enriches the toolbox for statistical inference of directional data, but also tackles the scientific questions related to DTI analysis. In Section 4.1, we introduce a novel angular expectation, which provides a more appropriate metric for inferring directional statistics. Based on the convenient tool of angular expectation, we further introduce tangent-normal decomposition of covariate effects (Section 4.2) and regions of differences characterized by separation angle (Section 4.3) to efficiently quantify covariate effects and regions of differences.

4.1 Angular Expectation

In general, for a random unit vector \( X \), the angular expectation is defined as \( \mathbb{A} X = \arg \min_{x \in S^{p-1}} \int \delta(x, X) \pi(X) dX \), where \( \mathbb{A} \) is introduced as an operator returning the angular expectation, \( p \) is the dimension of \( X \), and \( \pi(X) \) stands for the distribution of \( X \).
scheme, providing a more reasonable inferential route. Given a typical subject in group $g$ with design matrix $X_{g0}$, the mode directions i.e., $\mu_{g0} = \{\mu_{g0kj} : k = 1 : K, j = 1 : J_k\}$ are of primary interest and the corresponding posterior distribution can be expressed as

$$\mu_{g0}|X_{g0}; \text{data} = \int \mu_{g0}|\alpha_g, \beta_g, \tau^2_{\epsilon}, \tau^2_{\xi}, \Phi_\epsilon, \Phi_\xi; X_{g0} \alpha_g, \beta_g, \tau^2_{\epsilon}, \tau^2_{\xi}, \Phi_\epsilon, \Phi_\xi|\text{data},$$

where $\alpha_g$ and $\beta_g$ are vectors of all corresponding coefficients $\alpha_{gkj}$ and $\beta_{gkj}$ of group $g$ in Model 5, respectively.

The posterior distribution can be approximately learned using the MCMC outputs. For summarization, traditional posterior mean is sub-optimal as it does not guarantee that the estimates are always on a unit sphere. We thus empirically estimate the angular expectation using the MCMC samples $\{\mu_{g0}^{(t)} : t = 1 : T\}$ as

$$\tilde{\mu}_{g0} = \arg \min_{x \in S^2} \int \delta(\mu_{g0}, x)\pi(\mu_{g0}|X_{g0}; \text{data})d\mu_{g0} \approx \arg \min_{x \in S^2} \frac{1}{T} \sum_{t=1}^{T} \delta(x, \mu_{g0}^{(t)}) = \frac{\sum_{t=1}^{T} \mu_{g0}^{(t)}/T}{\|\sum_{t=1}^{T} \mu_{g0}^{(t)}/T\|}.$$

Relating to ADNI data analysis, the angular expectation $\tilde{\mu}_{g0}$ profiles the anatomical structure after model-based adjustment with the covariate $X_{g0}$.

### 4.2 Tangent-Normal Decomposition of Covariate Effects

A transparent inference framework to illustrate the covariate effects is essential for clinical studies. In our proposed model, the coefficients $\alpha_g$ and $\beta_g$ do not offer a straightforward illustration of the covariate effects on the mode directions. Therefore, we incorporate the tangent-normal decomposition \cite{Mardia2009} to summarize the covariate effects. Specific steps are described below.

Let $X_g$ be the design matrix of a typical subject in a clinical group $g$, where the continuous predictors are all set to the corresponding sample means and the categorical predictors are the corresponding sample modes of the subjects within the clinical group $g$. Now, let $A$ be a covariate whose effect is to be investigated. We let $\tilde{X}_g(A)$ be the design
matrix of an *hazard* subject in clinical group $g$ if $A$ is of a more hazard condition (e.g., one year older, one point decrement in MMSE, or having $\epsilon - 4$ variant). The corresponding predictive posterior of the mode directions, i.e., $\mu_g = \{\mu_gk : k = 1 : K, j = 1 : J_k\}$ can be expressed as follows and the empirical posterior predictive distribution can be obtained using the MCMC outputs as well. We use $\{\mathbf{\mu}_g^t : t = 1 : T\}$ and $\{\mathbf{\mu}_g(A)^t : t = 1 : T\}$ to denote the corresponding samples obtained from the MCMC outputs:

**Typical Mode Direction:**

$$\mathbf{\bar{\mu}}_g := [\mu_g|X_g; data] = \int [\mu_g|\alpha_g, \beta_g, \epsilon, \xi, \Phi_\epsilon, \Phi_\xi; X_g] [\alpha_g, \beta_g, \epsilon, \xi, \Phi_\epsilon, \Phi_\xi|data].$$  

(7)

**Harzard Mode Direction:**

$$\mathbf{\tilde{\mu}}_g := [\mu_g|X_g; data] = \int [\mu_g|\alpha_g, \beta_g, \epsilon, \xi, \Phi_\epsilon, \Phi_\xi; X_g] [\alpha_g, \beta_g, \epsilon, \xi, \Phi_\epsilon, \Phi_\xi|data]$$

To understand the covariate effect, we decompose $\mathbf{\mathbf{\mu}_gk}$ as in Figure 6 where $\mathbf{\mu}_gk$ is the tangent direction and $\mathbf{R}_gk$ is the normal direction. In terms of principal diffusion directions, the term $\mathbf{R}_gk$ can be interpreted as an *deviation* direction that the covariate $A$ exerts. The scalar $m_gk \in (0, 1) \frac{1}{2}$ which controls the magnitude of this effect can be used to quantitatively describe us the importance of this covariate.

Computationally, the tangent-normal decomposition can be proceeded for each MCMC sample $t$, thus we obtain the empirical posterior distribution of $\mathbf{R}_gk$ and $m_gk$, denoted as $\{R_gk^t : t = 1 : T\}$ and $\{m_gk^t : t = 1 : T\}$. In practice, we can use $A[R_gk|data]$ as the Bayesian empirical estimate to spatially profile the *deviation* directions caused by the covariate effects. Similarly, $\mathbb{E}[m_gk|data]$ can be used to quantify the magnitude of the covariate effect. We describe the implementation of tangent-normal decomposition due to covariate effects in our ADNI application (Section 6.1).
The yellow arrow is the *typical* mode direction; the blue arrow is the *hazard* mode direction; the green arrow is the *deviation* direction; The scalar $m_{gkj} \in (0, 1)$ which controls the magnitude of this effect.

### 4.3 Regions of Differences Characterized by Separation Angle

Especially in imaging analysis, characterizing regions of differences between clinical groups is critical. For principal diffusion directions, we propose a novel separation angle based method to characterize regions of difference. For each clinical group $g$, we get the angular expected mean for each subject at each voxel $A[\mu_{gikj} | X_{gikj}; \text{data}]$. We get the *sample* angular mean of the subjects in group $g$, defined as $\hat{\mu}_{gkj} = \arg \min_{x \in S^{p-1}} \sum_{i=1}^{N_g} \delta(x, A[\mu_{gikj} | X_{gikj}; \text{data}])$

In this way, the regions of differences among two groups can be characterized by separation angle defined as $\Delta_{g,g'}(v) = \delta(\hat{\mu}_{gkj}, \hat{\mu}_{g'kj})$ The results of $\Delta_{g,g'}(v)$ can be used to depict the regions of differences between any two clinical groups. We describe the implementation of regions of differences characterized by separation angle in our ADNI application (Section 6.2).
5 Model Comparison

In this section, we conduct numerical studies on the ADNI principal diffusion direction data (Section 2) and synthetic principal diffusion direction data to demonstrate the performance of our proposed vMF regression in comparison to other traditional alternatives. The codes and relevant data files are attached in Section B of the supplementary materials for reproducing the results. The basic assumption of our proposal is that the principal diffusion directions are the realizations of vMF distribution as $E_{gikj} \sim vMF(\mu_{gikj}, \kappa)$. This assumption guarantees that $E_{gikj} \in S^2$. However, there is a limiting equivalence between vMF distribution and multivariate Gaussian for large $\kappa$ as presented in Lemma 2 (Song and Dunson, 2022, Lemma 1).

Lemma 2. If $E \sim vMF(\mu, \kappa)$, as $\kappa \rightarrow \infty$, then $\sqrt{\kappa}[E - \mu] \rightarrow N(0, (I - \mu\mu^T))$ and $\sqrt{\kappa}[(E - \ell(\mu)] \rightarrow N(0, \nabla\ell(\mu)^T(I - \mu\mu^T)\nabla\ell(\mu))$, where $\nabla\ell(\mu)$ is the gradient of $\ell(\mu)$ with respect of $\mu$.

Hence, we introduce two benchmark methods based on multivariate Gaussian distribution. To make the benchmark methods parsimonious, the two methods are specified as follows: 1) Gaussian Regression 1: the multivariate regression model of diffusion directions, i.e., $E_{gikj}$, 2) Gaussian Regression 2: the multivariate regression model of transformed means, i.e., $\ell(E_{gikj})$. The multivariate regression model of diffusion directions is simply to treat the diffusion direction $E_{gikj}$ as a normal distribution $E_{gikj} \sim N(\mu_{gikj}, \Sigma_1)$, where $\mu_{gikj}(p) = X_{ig}u_{gkj}(p)$, $\mu_{gikj}(p)$ is the p-th element of $\mu_{gikj}$ for $p = 3$, and $u_{gkj}(p)$ is the corresponding coefficient. Alternatively, the multivariate regression model of transformed means is to treat the transformed means $\ell(E_{gikj}) = [A_{gikj}, B_{gikj}]^T$ as a normal distribution $\ell(E_{gikj}) = [A_{gikj}, B_{gikj}]^T \sim N([\tilde{\theta}_{gikj}, \tilde{\phi}_{gikj}]^T, \Sigma_2)$, where $\tilde{\theta}_{gikj} = X_{ig}a_{gkj}$ and $\tilde{\phi}_{gikj} = X_{ig}b_{gkj}$. $a_{gkj}$ and $b_{gkj}$ are corresponding coefficients. The Gaussian Regression 1
enjoys simplicity but does not guarantee that the support of principal diffusion directions to be in $S^2$. The Gaussian Regression 2 takes the advantage of proposed novel link function but ignores the randomness caused by the vMF distribution.

First, we apply the models to the motivating ADNI Data to measure the performances of the models. We randomly partition the subjects by their clinical groups. Among all the subjects, 50% of the subjects are treated as training data and 50% of the subjects are treated as validation data. Based on the model fitting results from the training data, we obtain the Bayesian posterior expectation $\hat{\mu}_{gikj}$ on the validation data. For Gaussian Regression 1, we obtain the posterior mean of $\mu_{gikj}$; For Gaussian Regression 2 and our proposed vMF regression, we use the angular expectation as introduced in Section 4. For the vMF regression, we set $P = 1, \ldots, 5$. To compare the methods, we measures their prediction error on the validation data. Two measures are used to calculate the prediction error: the separation angle $\delta(\hat{\mu}_{gikj}, E_{gikj})$ and the root square error $\|\hat{\mu}_{gikj} - E_{gikj}\|$.

In Figure 7 and 8, we visualize the prediction errors on fornix and corpus callosum, respectively. Given the mean of errors over subjects and voxels, our proposal provides overwhelmingly better performance than the benchmark methods. The advantages of our proposal become more transparent when inspecting the distribution of the prediction errors: the proposed vMF regression has a good stability and reliability since they are with fewer outlying larger errors. This implies that our proposal provides a more convincing result in analyzing ADNI data.

To further validate our method, we create synthetic principal diffusion direction data. To ensure that the synthetic data adequately mimics the ADNI data, the design matrix containing patients’ characteristics are borrowed directly from our real data, where 5 subjects’ covariate information of each clinical group are used for the training data and another

---

3For Gaussian Regression 1, we standardize $\mu_{gikj}$ to be unit vectors.
Figure 7: Prediction error within fornix: The violin plots show the prediction error of methods. The x-axis gives methods and the y-axis gives prediction error.
Figure 8: Prediction error within corpus callosum: The violin plots show the prediction error of methods. The x-axis gives methods and the y-axis gives prediction error.
5 subjects’ covariate information of each clinical group are used for the validation data. The coefficients are constructed accordingly to make the principal diffusion direction follow templates in the middle panel Figure 2. We simulate the data following our proposed full model (Model 5). When simulating the data, we set $P = 5$ and generate the autocorrelation coefficients from $\mathcal{U}(-1, 1)$. The variances are set as $\tau^2 = \tau_\xi^2 = \sigma^2 = \sigma_\xi^2 = 1$. We also consider multiple choices for the concentration parameter as $\kappa = 10, 30, 50$ to validate Lemma 2 that the vMF distribution and the Gaussian distribution become equivalent when $\kappa$ is sufficiently large. We generate 50 replicated datasets in total for each setting.

In Figure A.2 in Section A of supplementary materials, we illustrate the prediction errors over all the subjects, voxels, and replications. Similar to the real data results, our proposal again registers an overwhelmingly better performance than the Gaussian alternative. We also find that larger $\kappa$ produce better results for the normal distribution-based methods. This is an expected result due to Lemma 2. Furthermore, the vMF regression with $P = 5$ performs best among $P = 1, \ldots, 9$ while the data is generated with $P = 5$. Thus, this is our default approach to choose the best $P$ in general.

6 Bayesian Angular Inference for ADNI Data

In this section, we continue with our ADNI data application and implement Bayesian Angular Inference to lay out some scientific investigations on the ADNI data, revealing the underlying mechanism of Alzheimer’s disease. In Section 6.1, we illustrate the analytic results of covariate effects through our proposed tangent-normal decomposition. In section 6.2, we use separation angle metric to detect regions of differences. Both of the two approaches provide new insights on the principal diffusion direction data.
6.1 Tangent-Normal Decomposition of Covariate Effects

We now analyze the covariate effects applying the tangent-normal decomposition. As defined in Section 4, the tangent-normal decomposition requires a definition of \( \mathbf{X}_g \), the design matrix of a typical subject in a clinical group \( g \). Given our data, we define clinically meaningful \( \mathbf{X}_g \) as follows. For each clinical group, we average the continuous covariates which are age and MMSE of all the subjects in the group. As the APOE \( \epsilon_4 \) variant has been largest known genetic risk factor for AD in a variety of ethnic groups (Sadigh-Eteghad et al., 2012), we make non-APOE \( \epsilon_4 \) variant with average age and MMSE-score as typical. To follow the state of the art in reducing gender bias (e.g., Chilet-Rosell, 2014; Labots et al., 2018), we take either male or female in the \( \mathbf{X}_g \) to investigate either male’s or female’s covariate effects, respectively. Given \( \mathbf{X}_g \), we define \( \tilde{\mathbf{X}}_g(A) \) be a design matrix of an hazard subject in clinical group \( g \) as follows: 1) \( \text{Age} \): Add 1 unit to investigate the effect if a subject gets older; 2) \( \text{MMSE} \): Reduce 1 unit to investigate the effect if a subject gets worse cognitive score.; 3) \( \text{APOE} \): Adjust to \( \epsilon_4 \) variant to investigate if a subject holds a genetic risk factor. For conciseness, we only report the results of female in the main context. To efficiently visualize the result, we provide two visualize: 1) We visualize the \( \mathbb{E}[m_{gkj}|\text{data}] \) which quantify the magnitude of the covariate effect; 2) For the voxels where \( \mathbb{E}[m_{gkj}|\text{data}] \) is larger than 9-th decile of all applicable voxels within the regions of interest, we visualize the posterior angular means of typical direction and deviation direction. The \( \mathbb{E}[m_{gkj}|\text{data}] \) further helps us to quantify the importance of each predictor.

The comprehensive results of fornix and corpus callosum are given in the Section C of the supplementary materials, and we illustrate only the most interesting ones here for space. The results provide more insightful and meaningful results by inferring the principal diffusion directions. Based on the range of \( \mathbb{E}[m_{gkj}|\text{data}] \)-values for each predictors, we conclude that APOE has the largest effect among all, followed by MMSE, and Age.
Thus, we illustrate APOE-based results here. Like many fMRI studies (Trachtenberg, Filippini, Ebmeier, Smith, Karpe and Mackay, 2012; Trachtenberg, Filippini, Cheeseman, Duff, Neville, Ebmeier, Karpe and Mackay, 2012), APOE-effects in fornix exhibit heterogeneities on the left and right hemisphere of the brain, and it becomes more heterogeneous with increasing disease severity (see Figure 9). Such heterogeneous effect profiles are also observed for MMSE and Age. However, the magnitudes, quantified by $E[m_{gkj}|data]$, are very small in case Age for both fornix and corpus callosum. The effect magnitudes for fornix are in general larger than those in corpus callosum. It will thus be interesting to run similar analysis as ours for other structural MRI-based markers in the future. More importantly, our novel covariate-dependent analysis on principal diffusion directions allows us to understand how the covariates affects is present in terms of directional statistics. For example, via Figure 10, we can learn how the important covariate effects of APOE is present since our modeling approach applies to the principal diffusion directions directly. By inspecting the pattern of the deviate direction (green arrows), it helps to reveal the physiological disruption of white matters, providing more insightful information for in-depth investigation (e.g., Desikan et al., 2010) of the effects of human traits on brain’s structural anatomy.

6.2 Regions of Differences Characterized by Separation Angle

In this section, we investigate the regions of differences across different clinical groups. Here the differences are characterized by separation angles. For better visualization of the changes, we only show the separation angles in Figure 11 with the voxels whose values larger than 5°. In general, when we compare the clinical groups to the healthy controls, the separation angle increases with increasing severity of the cognitive impairment. This is a basic finding as we expected. In corpus callosum, anterior middle regions are apparently different between EMCI-to-LMCI and LMCI-AD. These results are consistent with many
Figure 9: Tangent-normal decomposition of APOE effects in fornix is given. For each subfigure, the upper panel gives $\mathbb{E}[m_{gkj}|data]$ which quantify the magnitude of the covariate effect where the color bar below indicates the corresponding values; Among the voxels where $\mathbb{E}[m_{gkj}|data]$ is larger than 9-th decile of all applicable voxels within the regions of interest, the red and green arrows are the posterior angular means of typical direction and deviation direction, respectively. The yellow lines connects the voxels which are in the same fiber.
Figure 10: In the left panel, we provide the AD group’s tangent-normal decomposition of APOE effects in fornix and outline a region in a blue box; In the right panel, we zoom in to have a clearly visualization of the blue box region. The red and green arrows are the posterior angular means of typical direction and deviation direction, respectively. The yellow lines connects the voxels which are in the same fiber. Among the voxels where $\mathbb{E}[m_{gkj}|data]$ is larger than 9-th decile of all applicable voxels within the regions of interest, the red and green arrows are the posterior angular means of typical direction and deviation direction, respectively.

recent scientific reports regarding (Walterfang et al., 2014; Bachman et al., 2014).

7 Conclusion and Discussion

In this paper, we develop a novel spatial generalized linear regression framework for modeling diffusion directions using a vMF-distributed error model. The regression model is shown to accurately capturing the local variation when random diffusion tensors are supported on a sphere. Given the nature of fiber tract-based data, the spatial variations are subsequently captured by an autoregressive framework. The numerical evaluation on the real data and the synthetic data demonstrate that our proposal has an overwhelming better performance. Important scientific findings are given through our proposed Bayesian
Figure 11: Regions of differences characterized by separation angle. In each panel, the rows and the columns give the combinations to be compared. The color bar below indicates the corresponding values of regions of difference shown for each voxel. The voxels whose values larger than 5° are labelled with colors.
angular inference.

In this paper, cross-sectional data is used for the analysis. However, the longitudinal data analysis of ADNI data is becoming more popular these years (e.g., [Wang and Guo 2019], [Kundu et al. 2019]). In ADNI data, the longitudinal data analysis may be more challenging given the involvement of asynchronous observations, i.e., the neuroimages and the covariates are not measured at the same time. As a future work, it is an appealing avenue to extend our generalized vMF regression to allow longitudinal analysis.

Another important direction will be to analyze the changes in fiber orientations among the converted subjects. These subjects are the ones whose disease status changed during their follow-up visits, with either increasing or decreasing severity. It is thus interesting to identify the factors fundamental to these changes. Our generalized vMF regression may be modified to analyze this important scientific problem.

References

Aisen, P. S., Petersen, R. C., Donohue, M. C., Gamst, A., Raman, R., Thomas, R. G., Walter, S., Trojanowski, J. Q., Shaw, L. M., Beckett, L. A. et al. (2010), ‘Clinical core of the Alzheimer’s disease neuroimaging initiative: progress and plans’, *Alzheimer’s & Dementia* 6(3), 239–246.

Bachman, A. H., Lee, S. H., Sidtis, J. J. and Ardekani, B. A. (2014), ‘Corpus callosum shape and size changes in early Alzheimer’s disease: A longitudinal MRI study using the OASIS brain database’, *Journal of Alzheimer’s Disease* 39(1), 71–78.

Chilet-Rosell, E. (2014), ‘Gender bias in clinical research, pharmaceutical marketing, and the prescription of drugs’, *Global health action* 7(1), 25484.
Desikan, R. S., Sabuncu, M. R., Schmansky, N. J., Reuter, M., Cabral, H. J., Hess, C. P., Weiner, M. W., Biffi, A., Anderson, C. D., Rosand, J. et al. (2010), ‘Selective disruption of the cerebral neocortex in Alzheimer’s disease’, PloS one 5(9), e12853.

Di Paola, M., Spalletta, G. and Caltagirone, C. (2010), ‘In vivo structural neuroanatomy of corpus callosum in Alzheimer’s disease and mild cognitive impairment using different MRI techniques: a review’, Journal of Alzheimer’s disease 20(1), 67–95.

Dobson, A. J. and Barnett, A. G. (2018), An introduction to generalized linear models, Chapman and Hall/CRC.

Gelfand, A. E., Diggle, P., Guttorp, P. and Fuentes, M. (2010), Handbook of spatial statistics, CRC press.

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(462), 387–396.

Goodlett, C. B., Fletcher, P. T., Gilmore, J. H. and Gerig, G. (2009), ‘Group analysis of dti fiber tract statistics with application to neurodevelopment’, Neuroimage 45(1), S133–S142.

Guha, S. and Rodriguez, A. (2020), ‘Bayesian regression with undirected network predictors with an application to brain connectome data’, Journal of the American Statistical Association pp. 1–13.

Hamilton, J. D. (2020), Time series analysis, Princeton university press.

Kang, J. and Li, L. (2016), ‘Discussion of “fiber direction estimation in diffusion mri”’, The Annals of Applied Statistics 10(3), 1162.
Kundu, S., Lukemire, J., Wang, Y. and Guo, Y. (2019), ‘A novel joint brain network analysis using longitudinal Alzheimer’s disease data’, *Scientific reports* **9**(1), 1–18.

Labots, G., Jones, A., de Visser, S. J., Rissmann, R. and Burggraaf, J. (2018), ‘Gender differences in clinical registration trials: is there a real problem?’, *British journal of clinical pharmacology* **84**(4), 700–707.

Lan, Z., Reich, B. J., Guinness, J., Bandyopadhyay, D., Ma, L. and Moeller, F. G. (2021), ‘Geostatistical modeling of positive definite matrices: An application to diffusion tensor imaging’, *Biometrics* p. In Press.

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.

Mardia, K. V. (1975), ‘Statistics of directional data’, *Journal of the Royal Statistical Society: Series B (Methodological)* **37**(3), 349–371.

Mardia, K. V. (2014), *Statistics of directional data*, Academic press.

Mardia, K. V. and Jupp, P. E. (2009), *Directional statistics*, Vol. 494, John Wiley & Sons.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W. and Beckett, L. (2005), ‘The Alzheimer’s disease neuroimaging initiative’, *Neuroimaging Clinics* **15**(4), 869–877.

Nowrangi, M. A. and Rosenberg, P. B. (2015), ‘The fornix in mild cognitive impairment and Alzheimer’s disease’, *Frontiers in aging neuroscience* **7**, 1.
Oishi, K., Mielke, M. M., Albert, M., Lyketsos, C. G. and Mori, S. (2012), ‘The fornix sign: a potential sign for Alzheimer’s disease based on diffusion tensor imaging’, *Journal of Neuroimaging* 22(4), 365–374.

Reich, B. J., Guinness, J., Vandekar, S. N., Shinohara, R. T. and Staicu, A.-M. (2018), ‘Fully bayesian spectral methods for imaging data’, *Biometrics* 74(2), 645–652.

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

Sadigh-Eteghad, S., Talebi, M. and Farhoudi, M. (2012), ‘Association of apolipoprotein epsilon 4 allele with sporadic late onset Alzheimer’s disease. a meta-analysis’, *Neurosciences Journal* 17(4), 321–326.

Schwartzman, A., Mascarenhas, W. F., Taylor, J. E. et al. (2008), ‘Inference for eigenvalues and eigenvectors of gaussian symmetric matrices’, *The Annals of Statistics* 36(6), 2886–2919.

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

Song, H. and Dunson, D. B. (2022), ‘Curved factor analysis with the ellipsoid-gaussian distribution’, *arXiv preprint arXiv:2201.08502*.

Teipel, S. J., Bayer, W., Alexander, G. E., Zebuhr, Y., Teichberg, D., Kulic, L., Schapiro, M. B., Möller, H.-J., Rapoport, S. I. and Hampel, H. (2002), ‘Progression of corpus callosum atrophy in Alzheimer disease’, *Archives of Neurology* 59(2), 243–248.
Trachtenberg, A. J., Filippini, N., Cheeseman, J., Duff, E. P., Neville, M. J., Ebmeier, K. P., Karpe, F. and Mackay, C. E. (2012), ‘The effects of apoe on brain activity do not simply reflect the risk of Alzheimer’s disease’, Neurobiology of aging 33(3), 618–e1.

Trachtenberg, A. J., Filippini, N., Ebmeier, K. P., Smith, S. M., Karpe, F. and Mackay, C. E. (2012), ‘The effects of apoe on the functional architecture of the resting brain’, Neuroimage 59(1), 565–572.

Vecchia, A. V. (1988), ‘Estimation and model identification for continuous spatial processes’, Journal of the Royal Statistical Society: Series B (Methodological) 50(2), 297–312.

Walterfang, M., Luders, E., Looi, J. C., Rajagopalan, P., Velakoulis, D., Thompson, P. M., Lindberg, O., Östberg, P., Nordin, L. E., Svensson, L. et al. (2014), ‘Shape analysis of the corpus callosum in Alzheimer’s disease and frontotemporal lobar degeneration subtypes’, Journal of Alzheimer’s Disease 40(4), 897–906.

Wang, Y. and Guo, Y. (2019), ‘A hierarchical independent component analysis model for longitudinal neuroimaging studies’, NeuroImage 189, 380–400.

Wong, J. C., Concha, L., Beaulieu, C., Johnston, W., Allen, P. S. and Kalra, S. (2007), ‘Spatial profiling of the corticospinal tract in amyotrophic lateral sclerosis using diffusion tensor imaging’, Journal of Neuroimaging 17(3), 234–240.

Wong, R. K., Lee, T. C., Paul, D., Peng, J., Alzheimer’s Disease Neuroimaging Initiative et al. (2016), ‘Fiber direction estimation, smoothing and tracking in diffusion mri’, The Annals of Applied Statistics 10(3), 1137–1156.
Yeh, F.-C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J. C., Vettel, J. M. and Verstynen, T. (2018), ‘Population-averaged atlas of the macroscale human structural connectome and its network topology’, Neuroimage 178, 57–68.

Yuan, Y., Zhu, H., Lin, W. and Marron, J. S. (2012), ‘Local polynomial regression for symmetric positive definite matrices’, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74(4), 697–719.

Zhang, Z., Allen, G. I., Zhu, H. and Dunson, D. (2019), ‘Tensor network factorizations: Relationships between brain structural connectomes and traits’, Neuroimage 197, 330–343.

Zhou, D. (2010), Statistical analysis of diffusion tensor imaging, PhD thesis, University of Nottingham.

Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y., Hall, C. and Lin, W. (2009), ‘Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging’, Journal of the American Statistical Association 104(487), 1203–1212.

Zhu, H., Kong, L., Li, R., Styner, M., Gerig, G., Lin, W. and Gilmore, J. H. (2011), ‘Fadtts: functional analysis of diffusion tensor tract statistics’, NeuroImage 56(3), 1412–1425.