The Annals of Applied Statistics
2013, Vol. 7, No. 1, 102–125
DOI: 10.1214/12-AOAS574

VARYING COEFFICIENT MODEL FOR MODELING DIFFUSION TENSORS ALONG WHITE MATTER TRACTS[1]

By Ying Yuan, Hongtu Zhu, Martin Styner, John H. Gilmore and J. S. Marron

St. Jude Children's Research Hospital, University of North Carolina at Chapel Hill, University of North Carolina at Chapel Hill, University of North Carolina at Chapel Hill and University of North Carolina at Chapel Hill

Diffusion tensor imaging provides important information on tissue structure and orientation of fiber tracts in brain white matter in vivo. It results in diffusion tensors, which are 3×3 symmetric positive definite (SPD) matrices, along fiber bundles. This paper develops a functional data analysis framework to model diffusion tensors along fiber tracts as functional data in a Riemannian manifold with a set of covariates of interest, such as age and gender. We propose a statistical model with varying coefficient functions to characterize the dynamic association between functional SPD matrix-valued responses and covariates. We calculate weighted least squares estimators of the varying coefficient functions for the log-Euclidean metric in the space of SPD matrices. We also develop a global test statistic to test specific hypotheses about these coefficient functions and construct their simultaneous confidence bands. Simulated data are further used to examine the finite sample performance of the estimated varying coefficient functions. We apply our model to study potential gender differences and find a statistically significant aspect of the development of diffusion tensors along the right internal capsule tract in a clinical study of neurodevelopment.

1. Introduction. Diffusion Tensor Imaging (DTI), which measures the effective diffusion of water molecules, can provide important information on the microstructure of fiber tracts and the major neural connections in white matter [Basser, Mattiello and LeBihan (1994a, 1994b)]. It has been widely

Received September 2011; revised April 2012.
[1]Supported in part by NIH Grants RR025747, CA142538, MH086633, GM70335, CA74015, HD03110, EB002779, EB00514901, MH064065, HD053000 and MH070890. The readers are welcome to request reprints from Dr. Hongtu Zhu (hzhu@bios.unc.edu).
Key words and phrases. Confidence band, diffusion tensor imaging, global test statistic, varying coefficient model, log-Euclidean metric, symmetric positive matrix.

used to assess the integrity of anatomical connectivity in white matter. In DTI, a 3×3 symmetric positive definite (SPD) matrix, called a diffusion tensor (DT), and its three eigenvalue-eigenvector pairs $\{(\lambda_k, \mathbf{v}_k) : k = 1, 2, 3\}$ with $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are estimated to quantify the degree of diffusivity and the directional dependence of water diffusion in each voxel (volume pixel). Multiple fiber tracts in white matter can be constructed by consecutively connecting the estimated principal directions ($\mathbf{v}_1$) of the estimated DTs in adjacent voxels [Basser et al. (2000)]. Subsequently, some tensor-derived scalar quantities, such as fractional anisotropy (FA) and mean diffusivity (MD), are commonly estimated along these white matter fiber tracts for each subject. Specifically, $\mathrm{MD} = (\lambda_1 + \lambda_2 + \lambda_3)/3$ describes the amount of diffusion, whereas FA describes the relative degree of anisotropy and is given by

$$\mathrm{FA} = \sqrt{\frac{3\{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2\}}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}. \tag{1.1}$$

In the recent DTI literature, there is an extensive interest in developing fiber-tract based analysis for comparing DTIs in population studies [Goldsmith et al. (2011), Goodlett et al. (2009), O'Donnell, Westin and Golby (2009), Smith et al. (2006), Yushkevich et al. (2008), Zhu et al. (2010, 2011)]. The reason is that the region-of-interest (ROI) method primarily computes averages of diffusion properties in some manually drawn ROIs, generates various summary statistics per ROI, and then carries out statistical analysis on these summary statistics. This method suffers from identifying meaningful ROIs, particularly the long curved structures common in fiber tracts, the instability of statistical results obtained from ROI analysis, and the partial volume effect in relative large ROIs [Zhu et al. (2011)]. The fiber-tract based analysis usually consists of two major components, including DTI atlas building and a follow-up statistical analysis [Goodlett et al. (2009), Smith et al. (2006), Zhu et al. (2010)]. The DTI atlas building is primarily to extract DTI fibers and to establish DTI fiber correspondence across all DTI data sets from different subjects. The key steps of the DTI atlas building include DTI registration, atlas fiber tractography and fiber parametrization. Finally, we get a set of individual tracts with the same corresponding geometry but varying DTs and diffusion properties. Some statistical approaches have been developed for the analysis of scalar tensor-derived quantities along fiber tracts [Goldsmith et al. (2011), Goodlett et al. (2009), Smith et al. (2006), Yushkevich et al. (2008), Zhu, Li and Kong (2010), Zhu et al. (2010, 2011)], but little has been done on the analysis of whole DTs along fiber tracts, which is the focus of this paper.

There is a growing interest in the DTI literature in developing statistical methods for the direct analysis of DTs in the space of SPD matrices [Dryden, Koloydenko and Zhou (2009)]. Schwartzman, Mascarenhas and Taylor

(2008) proposed parametric models for analyzing SPD matrices and derived the distributions of several test statistics for comparing differences between the means of the two (or multiple) groups of SPD matrices. Kim and Richards (2011) developed a nonparametric estimator of the density function of a random sample of SPD matrices. Zhu et al. (2009) developed a semiparametric regression model with SPD matrices as responses and covariates in a Euclidean space. Barmpoutis et al. (2007) and Davis et al. (2010) developed nonparametric methods, including tensor spline methods and local constant regression, to interpolate diffusion tensor fields. However, no one has ever developed statistical methods for functional analysis of DTs along fiber tracts.

In this paper, we propose a *varying coefficient model for DT-valued functions* (VCDF). We use varying coefficient functions to characterize the varying association between diffusion tensors along fiber tracts and a set of covariates. Here, the varying coefficients are the parameters in the model which vary with location. Since the impacts of the covariates of interest may vary spatially, it would be more sensible to treat the covariates as functions of location instead of constants, which leads to varying coefficients. In addition, we explicitly model the within-subject correlation among multiple DTs measured along a fiber tract for each subject. To account for the curved nature of the SPD space, we employ the log-Euclidean framework in Arsigny (2006) and then use a weighted least squares estimation method to estimate the varying coefficient functions. We also develop a global test statistic to test hypotheses on the varying coefficient functions and use a resampling method to approximate the p-value. Finally, we construct a simultaneous confidence band to quantify the uncertainty of each estimated coefficient function and propose a resampling method to approximate its critical points. To the best of our knowledge, this is the first paper for developing a statistical framework for modeling functional manifold-valued responses with covariates in Euclidean space.

There are several advantages of the analysis of DTs over the analysis of scalar diffusion properties along fiber tracts. The first one is that it can avoid the statistical artifacts, including biased parameter estimates and incorrect test statistics and p-values for hypotheses of interest, created by comparing the biased diffusion properties along fiber bundles. This is because the real DT data estimated from the diffusion weighted images (DWIs) using weighted least squared methods are almost unbiased [Zhu et al. (2007b)], whereas the diffusion properties, which are nonlinear and linear functions of three eigenvalues of DT data, may be substantially different from the true diffusion properties [Anderson (2001), Pierpaoli and Basser (1996), Zhu et al. (2007b)]. In addition, as shown in Yuan et al. (2012), directly modeling DTs along fiber bundles as a smooth SPD process allows us to incorporate a smoothness constraint to further reduce noise in the estimated DTs along

the fiber bundles. This leads to the further reduction of noise in estimated scalar diffusion properties along the fiber bundles and less biased estimators of diffusion properties as shown in Figure 4 in Section 3. Moreover, the sole use of diffusion properties, which ignores the directional information of DT, can decrease the statistical power in detecting the difference in DTs oriented in different directions.

The rest of the paper is organized as follows. Section 2 presents VCDF and related statistical inference. Section 3 examines the finite sample performance of VCDF via a simulation study. Section 4 illustrates an application of VCDF in a clinical study of neurodevelopment. Section 5 presents concluding remarks.

2. Data and methods.

2.1. *Early brain development study of white matter tracts.* We consider 96 healthy infants (36 males and 60 females) from the neonatal project on early brain development led by Dr. Gilmore at the University of North Carolina at Chapel Hill. The mean gestational age of these infants is 245.6 days with SD: 18.5 days (range: 192–270 days). A 3T Allegra head only MR system was used to acquire all the images. The system was equipped with a maximal gradient strength of 40 mT/m and a maximal slew rate of 400 mT/(m $\cdot$ msec). The DTIs were obtained by using a single shot EPI DTI sequence (TR/TE = 5400/73 msec) with eddy current compensation. The six noncollinear directions at the b-value of 1000 s/mm^2 with a reference scan ($b = 0$) were applied. The voxel resolution was isotropic 2 mm, and the in-plane field of view was set to 256 mm in both directions. To improve the signal-to-noise ratio of the DTIs, a total of five repetitions were acquired and averaged.

We processed the DTI data set as follows. We used a weighted least squares estimation method [Basser, Mattiello and LeBihan (1994a), Yuan et al. (2008), Zhu et al. (2007b)] to construct the diffusion tensors. We used a DTI atlas building pipeline [Goodlett et al. (2009), Zhu et al. (2010)] to register DTIs from multiple subjects to create a study-specific unbiased DTI atlas, to track fiber tracts in the atlas space, and to propagate them back into each subject's native space by using registration information. Then, we calculated DTs and their scalar diffusion properties at each location along each individual fiber tract by using DTs in neighboring voxels close to the fiber tract. Since the description of the DTI atlas building has been described in detail [Goodlett et al. (2009), Zhu et al. (2010)], we do not include these image processing steps here for the sake of simplicity. Figure 1(a) displays the fiber bundle of the right internal capsule fiber tract (RICFT), which is an area of white matter in the brain. The internal capsule, which lies between the lenticular and caudate nuclei, consists of a group of myelinated

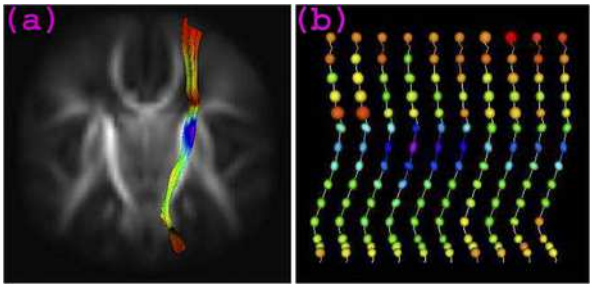

FIG. 1. (a) *The fiber bundle of the right internal capsule fiber tracts in the atlas space.* (b) *The ellipsoidal representations of full tensors along a representative right internal capsule fiber tract obtained from each of* 10 *selected subjects, colored with fractional anisotropy (FA) values. The rainbow color scheme is used with red corresponding to low FA value and purple corresponding to high FA value.*

fiber tracts including axons of pyramidal and extrapyramidal upper motor neurons that connect the cortex to the cell bodies of lower motor neurons. Although the internal capsule ends within the cerebrum, the axons that pass through it continue down through brain stem and spinal cord. It was found that neonatal microstructural development of the internal capsule tract correlates with severity of gait and motor deficits [Rose et al. (2007)]. Figure 1(b) presents DTs along a representative RICFT obtained from each of 10 subjects, in which each DT is geometrically represented by an ellipsoid. In this ellipsoidal representation, the lengths of the semiaxes of the ellipsoid equal the square root of the three eigenvalues of a DT, while the three eigenvectors define the direction of the three axes.

Our final data set includes DTs and diffusion properties sampled along the RICFT and a set of covariates of interest from all $n = 96$ subjects. Specifically, let $\mathrm{Sym}^+(3)$ be the set of 3×3 SPD matrices and $x_j \in [0, L_0]$ be the arc length of the jth point on the RICFT relative to a fixed end point for $j = 1, \ldots, n_G$, where L_0 is the longest arc length and n_G is the number of points on the RICFT. For the ith subject, there is a diffusion tensor at the jth point on the RICFT, denoted by $S_i(x_j) \in \mathrm{Sym}^+(3)$, for $i = 1, \ldots, n$. Let $\mathbf{z}_i$ be an $r \times 1$ vector of covariates of interest. In this study, we have two specific aims. The first one is to compare DTs along the RICFT between the male and female groups. The second one is to delineate the development of fiber DTs across time, which is addressed by including the gestational age at MRI scanning as a covariate. Finally, our real data set can be represented as $\{(\mathbf{z}_i; (x_1, S_i(x_1)), \ldots, (x_{n_G}, S_i(x_{n_G}))) : i = 1, \ldots, n\}$.

2.2. *Varying coefficient model for SPD matrix-valued functional data.* In this section we present our VCDF. The code for VCDF written in Matlab along with its documentation and a sample data set will be freely accessi-

ble from http://www.bios.unc.edu/research/bias/software.html. To make the code easily accessible, we developed a Graphical User Interface (GUI), also freely downloadable from the same website.

To proceed, we need to introduce some notation. Let Sym(3) be the set of 3×3 symmetric matrices with real entries. For any $A = (a_{kl}) \in \mathrm{Sym}(3)$, we define $\mathrm{vecs}(A) = (a_{11}, a_{21}, a_{22}, a_{31}, a_{32}, a_{33})^T$ to be a 6×1 vector and

$$\mathrm{vec}(A) = (a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33})^T$$

to be a 9×1 vector. Let $\mathrm{Ivecs}(\cdot)$ be the inverse operator of $\mathrm{vecs}(\cdot)$ such that $\mathrm{Ivecs}(\mathrm{vecs}(A)) = A$ for any $A \in \mathrm{Sym}(3)$. The matrix exponential of $A \in \mathrm{Sym}(3)$ is given by $\exp(A) = \sum_{m=0}^{\infty} A^m/m! \in \mathrm{Sym}^+(3)$. For any 3×3 SPD matrix S, there is a logarithmic map of S, denoted as $\log(S) = A \in \mathrm{Sym}(3)$, such that $\exp(A) = S$. Let $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector or matrix $\mathbf{a}$.

Since the space of SPD matrices is a curved space, we use the log-Euclidean metric [Arsigny (2006)] to account for the curved nature of the SPD space. Specifically, we take the logarithmic map of the DTs $S_i(x) \in \mathrm{Sym}^+(3)$ to get $\log(S_i(x)) \in \mathrm{Sym}(3)$, which has the same effective dimensionality as a six-dimensional Euclidean space. Thus, we only model the lower triangular portion of $\log(S_i(x))$ as follows:

$$\mathrm{vecs}(\log(S_i(x))) = B(x)\mathbf{z}_i + \mathbf{u}_i(x) + \boldsymbol{\varepsilon}_i(x), \tag{2.1}$$

where $B(x)$ is a $6 \times r$ matrix of varying coefficient functions for characterizing the dynamic associations between $S_i(x)$ and $\mathbf{z}_i$, $\mathbf{u}_i(x)$ is a 6×1 vector characterizing the within-subject correlation between the log-transformed DTs, and $\boldsymbol{\varepsilon}_i(x)$ is a 6×1 vector of measurement errors. It is also assumed that $\boldsymbol{\varepsilon}_i(x)$ and $\mathbf{u}_i(x)$ are independent and identical copies of $\mathrm{SP}(\mathbf{0}, \Sigma_{\boldsymbol{\varepsilon}})$ and $\mathrm{SP}(\mathbf{0}, \Sigma_{\mathbf{u}})$, respectively, where $\mathrm{SP}(\mathbf{0}, \Sigma)$ denotes a stochastic process with mean $\mathbf{0}$ and covariance matrix function $\Sigma(x, x')$ for any $x, x' \in [0, L_0]$. Let $\mathbf{1}(\cdot)$ be an indicator function. Assume that $\boldsymbol{\varepsilon}_i(x)$ and $\boldsymbol{\varepsilon}_i(x')$ for $x \neq x'$ are independent and, thus, $\Sigma_{\boldsymbol{\varepsilon}}(x, x') = \Sigma_{\boldsymbol{\varepsilon}}(x, x)\mathbf{1}(x = x')$. It follows that the covariance structure of $\mathrm{vecs}(\log(S_i(x_j)))$, denoted by $\Sigma_S(x, x')$, is given by

$$\Sigma_S(x, x') = \Sigma_{\mathbf{u}}(x, x') + \Sigma_{\boldsymbol{\varepsilon}}(x, x)\mathbf{1}(x = x'). \tag{2.2}$$

Model (2.1) is a multivariate varying coefficient model with a 6×1 vector response and, thus, it can be regarded as a generalization of univariate varying coefficient models, which have been widely studied and developed for longitudinal, time series and functional data [Fan, Yao and Cai (2003), Fan and Zhang (1999, 2008) Wang, Li and Huang (2008), Wu and Chiang (2000)].

2.3. *Weighted least squares estimation.* Before estimating the varying coefficient functions in $B(x)$, we need to introduce a few facts about the log-Euclidean metric for the space of SPDs [Arsigny (2006)]. The use of the log-Euclidean metric results in classical Euclidean computations in the domain of matrix logarithms. Particularly, under the log-Euclidean metric,

the geodesic distance between S_1 and S_2 in $\mathrm{Sym}^+(3)$ is uniquely given by

$$d(S_1, S_2) = \sqrt{\mathrm{tr}[\{\log(S_1) - \log(S_2)\}^{\otimes 2}]}, \tag{2.3}$$

which equals the Euclidean distance between $\log(S_1)$ and $\log(S_2)$ in Euclidean space Sym(3). However, there is a subtle, but important, difference between regarding $S(x)$ as a single point in $\mathrm{Sym}^+(3)$ and treating $\log(S(x))$ as a vector in Euclidean space. By regarding $S(x)$ as a point in $\mathrm{Sym}^+(3)$, we treat all elements of $S(x)$ as a single unit and use a single bandwidth to smooth DTs. In contrast, by treating $\log(S(x))$ as a vector in Euclidean space, traditional smoothing methods smooth each element of $\log(S(x))$ independently [Fan and Gijbels (1996), Wand and Jones (1995), Wu and Zhang (2006)].

We use the local linear regression method and the weighted least squares estimation to estimate $B(x)$ [Fan and Gijbels (1996), Ramsay and Silverman (2005), Wand and Jones (1995), Welsh and Yee (2006), Wu and Zhang (2006), Zhang and Chen (2007)]. Since the local linear regression method adapts automatically at the boundary points [Fan and Gijbels (1992)], it is ideal for dealing with DTs and scalar diffusion properties along fiber tracts with two ends (see Figure 1). Let $h^{(1)}$ be a given bandwidth, $\dot{B}(x) = dB(x)/dx$ be a $6 \times r$ matrix, and I_r be the $r \times r$ identity matrix. Using Taylor's expansion, we can expand $B(x_j)$ at x to obtain

$$B(x_j) \approx B(x) + \dot{B}(x)(x_j - x) = B_{h^{(1)}}(x)\{I_r \otimes \mathbf{y}_{h^{(1)}}(x_j - x)\}, \tag{2.4}$$

where $\mathbf{y}_h(x_j - x) = (1, (x_j - x)/h)^T$ and $B_{h^{(1)}}(x) = [B(x), h^{(1)}\dot{B}(x)]$ is a $6 \times 2r$ matrix. Based on (2.1) and (2.4), $B(x_j)\mathbf{z}_i$ can be approximated by $B_{h^{(1)}}(x)\{I_r \otimes \mathbf{y}_{h^{(1)}}(x_j - x)\}\mathbf{z}_i$. For a fixed bandwidth $h^{(1)}$, we can calculate a weighted least squares estimate of $B_{h^{(1)}}(x)$, denoted by $\hat{B}_{h^{(1)}}(x) = [\hat{B}(x; h^{(1)}), h^{(1)}\dot{B}(x; h^{(1)})]$, by minimizing an objective function given by

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x) \\ &\qquad \times d(\log(S_i(x_j)), \mathrm{Ivecs}(B_{h^{(1)}}(x)\{I_r \otimes \mathbf{y}_{h^{(1)}}(x_j - x)\}\mathbf{z}_i))^2, \end{aligned} \tag{2.5}$$

where $K_{h^{(1)}}(\cdot) = K(\cdot/h^{(1)})/h^{(1)}$ is rescaling of the kernel function $K(\cdot)$, such as the Gaussian or uniform kernel [Fan and Gijbels (1996), Wand and Jones (1995)]. The explicit form of $\hat{B}(x; h^{(1)})$ can be found in Appendix C.

We pool the data from all n subjects and develop a cross-validation method to select an estimated bandwidth $h^{(1)}$, denoted by $\hat{h}_e^{(1)}$. Let $\hat{B}(x; h^{(1)})^{(-i)}$ be the weighted least squares estimator of $B(x)$ for the bandwidth $h^{(1)}$ based on the observations with the ith subject excluded. We define a

cross-validation score, denoted by $\mathrm{CV}_1(h^{(1)})$, as follows:

$$\mathrm{CV}_1(h^{(1)}) = (nn_G)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G} d(\log(S_i(x_j)), \mathrm{Ivecs}(\hat{B}(x;h^{(1)})^{(-i)}\mathbf{z}_i))^2. \tag{2.6}$$

We select $\hat{h}_e^{(1)}$ by minimizing $\mathrm{CV}_1(h^{(1)})$. In practice, within a given range of $h^{(1)}$, the value of $\hat{h}_e^{(1)}$ can be approximated by computing $\mathrm{CV}_1(h^{(1)})$ through a series of $h^{(1)}$. Finally, we can calculate a weighted least squares estimate of $B(x)$, denoted by $\hat{B}_e(x) = \hat{B}(x;\hat{h}_e^{(1)})$.

2.4. *Smoothing individual functions and estimating covariance matrices.* To simultaneously construct the individual function $\mathbf{u}_i(x)$, we also employ the local linear regression method. Let $\dot{\mathbf{u}}_i(x) = d\mathbf{u}_i(x)/dx$. Taylor's expansion of $\mathbf{u}_i(x_j)$ at x gives

$$\mathbf{u}_i(x_j) \approx \mathbf{u}_i(x) + \dot{\mathbf{u}}_i(x)(x_j - x) = U_i(x)\mathbf{y}_{h^{(2)}}(x_j - x), \tag{2.7}$$

where $U_i(x) = [\mathbf{u}_i(x), h^{(2)}\dot{\mathbf{u}}_i(x)]$ is a 6×2 matrix. For each fixed x and each bandwidth $h^{(2)}$, the weighted least square estimator of $U_i(x)$, denoted by $\hat{U}_i(x;h^{(2)}) = [\mathbf{u}_i(x;h^{(2)}), h^{(2)}\dot{\mathbf{u}}_i(x;h^{(2)})]$, can be calculated by minimizing an objective function given by

$$\sum_{j=1}^{n_G} K_{h^{(2)}}(x_j - x)d(\log(S_i(x_j)), \mathrm{Ivecs}(\hat{B}_e(x_j)\mathbf{z}_i + U_i(x)\mathbf{y}_{h^{(2)}}(x_j - x)))^2.$$

Let R_i be an $n_G\times 6$ matrix with the jth row $\mathrm{vecs}(\log(S_i(x_j))) - \hat{B}_e(x_j)\mathbf{z}_i$ and $\mathcal{S}$ be an $n_G\times n_G$ smoothing matrix with the (i,j)th element $\tilde{K}^0_{h^{(2)}}(x_j - x_i, x_i)$, where $\tilde{K}^0_{h^{(2)}}(\cdot,\cdot)$ is the empirical equivalent kernel [Fan and Gijbels (1996)]. It can be shown that

$$(\hat{\mathbf{u}}_i(x_1), \ldots, \hat{\mathbf{u}}_i(x_{n_G}))^T = \mathcal{S}R_i. \tag{2.8}$$

We pool the data from all n subjects and select an estimated bandwidth of $h^{(2)}$, denoted as $\hat{h}_e^{(2)}$. We define a generalized cross-validation score, denoted by $\mathrm{GCV}(h^{(2)})$, as follows:

$$\mathrm{GCV}(h^{(2)}) = n^{-1}\frac{\sum_{i=1}^{n}\mathrm{tr}\{(R_i - \mathcal{S}R_i)^{\otimes 2}\}}{\{1 - n^{-1}\mathrm{tr}(\mathcal{S})\}^2}. \tag{2.9}$$

We select $\hat{h}_e^{(2)}$ by minimizing $\mathrm{GCV}(h^{(2)})$. Like the bandwidth selection in Section 2.3, the value of $\hat{h}_e^{(2)}$ can be approximated by computing $\mathrm{GCV}(h^{(2)})$ through a series of $h^{(2)}$. Finally, by substituting $\hat{h}_e^{(2)}$ into (2.8), we can calculate a weighted least squares estimate of $\mathbf{u}_i(x)$, denoted by $\hat{\mathbf{u}}_{i,e}(x)$.

After obtaining $\hat{\mathbf{u}}_{i,e}(x)$, we can estimate the mean function $\mathbf{u}(x)$ and the covariance function $\Sigma_{\mathbf{u}}(x,x')$. Specifically, we estimate $\mathbf{u}(x)$ and $\Sigma_{\mathbf{u}}(x,x')$ by

using their empirical counterparts based on the estimated $\hat{\mathbf{u}}_{i,e}(x)$ as follows:

$$\hat{\mathbf{u}}_e(x) = n^{-1}\sum_{i=1}^{n}\hat{\mathbf{u}}_{i,e}(x) \quad \text{and} \quad \hat{\Sigma}_{\mathbf{u}}(x,x') = (n-6)^{-1}\sum_{i=1}^{n}\hat{\mathbf{u}}_{i,e}(x)\hat{\mathbf{u}}_{i,e}(x')^T.$$

We construct a nonparametric estimator of the covariance matrix $\Sigma_{\boldsymbol{\varepsilon}}(x,x)$ as follows. Let $\hat{\boldsymbol{\varepsilon}}_i(x_j) = \operatorname{vecs}(\log(S_i(x_j))) - \hat{B}_e(x_j)\mathbf{z}_i - \hat{\mathbf{u}}_{i,e}(x_j)$ be the estimated residuals for $i = 1,\ldots,n$ and $j = 1,\ldots,n_G$. We consider the kernel estimate of $\Sigma_{\boldsymbol{\varepsilon}}(x,x)$ given by

$$\hat{\Sigma}_{\boldsymbol{\varepsilon}}(x,x;h^{(3)}) = (n-6)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G}\frac{K_{h^{(3)}}(x_j-x)\hat{\boldsymbol{\varepsilon}}_i(x_j)^{\otimes 2}}{\sum_{j=1}^{n_G}K_{h^{(3)}}(x_j-x)}. \tag{2.10}$$

We pool the data from all n subjects and select an estimated bandwidth of $h^{(3)}$, denoted as $\hat{h}_e^{(3)}$. Let $\tilde{\Sigma}_{\boldsymbol{\varepsilon}}(x_j,x_j) = (n-6)^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\varepsilon}}_i(x_j)\hat{\boldsymbol{\varepsilon}}_i(x_j)^T$ be an estimate of $\Sigma_{\boldsymbol{\varepsilon}}(x_j,x_j)$ and $\hat{\Sigma}_{\boldsymbol{\varepsilon}}(x,x;h^{(3)})^{(-i)}$ be the leave-one-out weighted least squares estimator of $\hat{\Sigma}_{\boldsymbol{\varepsilon}}(x,x)$. We define a cross-validation score, denoted by $\mathrm{CV}_2(h^{(3)})$, as follows:

$$(nn_G)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_G}\operatorname{tr}\{[\hat{\boldsymbol{\varepsilon}}_i(x_j)^{\otimes 2} - \hat{\Sigma}_{\boldsymbol{\varepsilon}}(x_j,x_j;h^{(3)})^{(-i)}]^{\otimes 2}\tilde{\Sigma}_{\boldsymbol{\varepsilon}}(x_j,x_j)^{-1}\}.$$

We select $h^{(3)}$ by minimizing $\mathrm{CV}_2(h^{(3)})$. In practice, within a given range of $h^{(3)}$, the value of $\hat{h}_e^{(3)}$ can be approximated by computing $\mathrm{CV}_2(h^{(3)})$ through a series of $h^{(3)}$. Finally, by substituting $\hat{h}_e^{(3)}$ into (2.10), we can calculate a weighted least squares estimate of $\Sigma_{\boldsymbol{\varepsilon}}(x,x)$, denoted by $\hat{\Sigma}_{\boldsymbol{\varepsilon},e}(x,x)$.

2.5. *Asymptotic properties.* We will use the following theorems to make statistical inference on varying coefficient functions. The detailed assumptions of these theorems can be found in Appendix A and their proofs are similar to those in Zhu, Li and Kong (2010). Thus, we omit them for the sake of space. We need some notation. Let $\ddot{B}(x) = d^2B(x)/dx^2$ and $G(\mathbf{0},\Sigma)$ be a Gaussian process with zero mean and covariance matrix function $\Sigma(x,x')$ for any $x,x' \in [0,L_0]$.

THEOREM 1. *If the assumptions* (C1)–(C6) *in the Appendix A hold, then* $\sqrt{n}\{\operatorname{vec}(\hat{B}(x;h^{(1)}) - B(x) - 0.5u_2\ddot{B}(x)h^{(1)2}[1+o_p(1)]) : x \in [0,L_0]\} \Rightarrow X_B(x)$, *where* $\Rightarrow$ *denote weak convergence of a sequence of stochastic processes,* $X_B(\cdot)$ *follows a Gaussian process* $G(\mathbf{0},\Sigma_{\mathbf{u}}\otimes\Omega_{\mathbf{z}}^{-1})$, *and* $\Omega_{\mathbf{z}}$ *is the limit of* $n^{-1}\sum_{i=1}^{n}\mathbf{z}_i^{\otimes 2}$ *as* $n\to\infty$.

Theorem 1 establishes weak convergence of $\hat{B}(x;h^{(1)})$ as a stochastic process indexed by $x \in [0,L_0]$ and forms the foundation for constructing a global test statistic and simultaneous confidence bands for $\{B(x) : x \in [0,L_0]\}$.

THEOREM 2. *If the assumptions* (C1)–(C7) *in the Appendix A hold, then*

$$\sup_{(x,x')\in[0,L_0]^2} |\hat{\Sigma}_{\mathbf{u}}(x,x';h^{(3)}) - \Sigma_{\mathbf{u}}(x,x')| = o_p(1).$$

Theorem 2 shows the uniform convergence of $\hat{\Sigma}_{\mathbf{u}}(x,x';h^{(3)})$. This is useful for constructing global and local test statistics for testing the covariate effects.

2.6. *Hypothesis tests.* In neuroimaging studies, many scientific questions of interest require the comparison of fiber bundle diffusion tensors along fiber bundles across two (or more) diagnostic groups and the assessment of the development of fiber bundle diffusion tensors along time. Such questions can often be formulated as linear hypotheses of $B(x)$ as follows:

$$H_0 : C\,\mathrm{vec}(B(x)) = \mathbf{b}_0(x) \quad \text{for all } x \quad \text{vs.} \quad H_1 : C\,\mathrm{vec}(B(x)) \neq \mathbf{b}_0(x), \tag{2.11}$$

where C is a $c \times 6r$ matrix of full row rank and $\mathbf{b}_0(x)$ is a given $c \times 1$ vector of functions.

We propose both local and global test statistics. The local test statistic can identify the exact location of a significant location on a specific tract. At a given point x_j on a specific tract, we test the local null hypothesis

$$H_0(x_j) : C\,\mathrm{vec}(B(x_j)) = \mathbf{b}_0(x_j) \quad \text{v.s.} \quad H_1(x_j) : C\,\mathrm{vec}(B(x_j)) \neq \mathbf{b}_0(x_j).$$

We use a local test statistic $T_n(x_j)$ defined by

$$T_n(x_j) = n\mathbf{d}(x_j)^T \{C(\hat{\Sigma}_{\mathbf{u}}(x_j,x_j) \otimes \hat{\Omega}_{\mathbf{z}}^{-1})C^T\}^{-1}\,\mathbf{d}(x_j), \tag{2.12}$$

where $\hat{\Omega}_{\mathbf{z}} = n^{-1}\sum_{i=1}^n \mathbf{z}_i^{\otimes 2}$ and $\mathbf{d}(x) = C\,\mathrm{vec}(\hat{B}_e(x) - \mathrm{bias}(\hat{B}_e(x))) - \mathbf{b}_0(x)$. Following Fan and Zhang (2000), a smaller bandwidth leads to a smaller value of $\mathrm{bias}(\hat{B}_e(x))$. Moreover, according to our simulation studies below, we have found that the effect of dropping $\mathrm{bias}(\hat{B}_e(x))$ is negligible and, therefore, we drop it from now on.

To test the null hypothesis $H_0 : C\,\mathrm{vecs}(B(x)) = \mathbf{b}_0(x)$ for all x, we propose a global test statistic $\mathbf{T}_n$ defined by

$$\mathbf{T}_n = \int_0^{L_0} T_n(x)\,dx. \tag{2.13}$$

Let $G_C(\cdot)$ be a Gaussian process with zero mean and covariance matrix function $\Sigma_C(x,x')$, which is the limit of

$$\{C(\hat{\Sigma}_{\mathbf{u}}(x,x) \otimes \hat{\Omega}_{\mathbf{z}}^{-1})C^T\}^{-1/2}\{C(\hat{\Sigma}_{\mathbf{u}}(x,x') \otimes \hat{\Omega}_{\mathbf{z}}^{-1})C^T\} \times \{C(\hat{\Sigma}_{\mathbf{u}}(x',x') \otimes \hat{\Omega}_{\mathbf{z}}^{-1})C^T\}^{-1/2}.$$

It follows from Theorem 1 that $\sqrt{n}\{C(\hat{\Sigma}_{\mathbf{u}}(x,x)\otimes\hat{\Omega}_{\mathbf{z}}^{-1})C^T\}^{-1/2}\mathbf{d}(x)$ converges weakly to $G_C(x)$. Therefore, it follows from the continuous mapping theorem that as both n and n_G converge to infinity, we have

$$\mathbf{T}_n \Rightarrow \int_0^{L_0} G_C(x)^T G_C(x)\,dx. \tag{2.14}$$

Based on the result (2.14), we develop a wild bootstrap method to approximate the p-value of $\mathbf{T}_n$. The detailed steps of the wild bootstrap method are given in Appendix B.

2.7. *Confidence band.* Based on model (2.16), we construct a confidence band for $S(B(x),\mathbf{z}) = \exp(\mathrm{Ivecs}(B(x)\mathbf{z})) \in \mathrm{Sym}^+(3)$ over $x \in [0, L_0]$ for a fixed $\mathbf{z}$. Specifically, at a given significance level α, we construct a simultaneous confidence region in the space of SPD matrices for each $\mathbf{z}$ based on the critical value $C_z(\alpha)$ such that

$$P(d(S(B(x),\mathbf{z}), S(\hat{B}(x),\mathbf{z})) \le C_z(\alpha) \text{ for all } x \in [0, L_0]) = 1-\alpha. \tag{2.15}$$

Note that $d(S(B(x),\mathbf{z}), S(\hat{B}(x),\mathbf{z})) = \sqrt{\mathrm{tr}([\mathrm{Ivecs}(\{\hat{B}_e(x) - B(x)\}\mathbf{z})]^{\otimes 2})}$. By using Theorem 1, we have that as $n \to \infty$,

$$\sqrt{n}d(S(B(x),\mathbf{z}), S(\hat{B}(x),\mathbf{z})) \Rightarrow \sqrt{\mathrm{tr}[\{\mathrm{Ivecs}(X_B(x)\mathbf{z})\}^{\otimes 2}]}.$$

We develop an efficient resampling method [Kosorok (2003), Zhu et al. (2007a)] to approximately draw random samples from $\{X_B(x) : x \in [0, L_0]\}$, denoted by $\{X_B(x)^{(g)} : x \in [0, L_0]\}$ for $g = 1, \ldots, G$. The detailed steps of such a resampling method can be found in Appendix C. Subsequently, we can calculate $\sqrt{\mathrm{tr}[\{\mathrm{Ivecs}(X_B(x)^{(g)}\mathbf{z})\}^{\otimes 2}]}$ for all g and use them to approximate $C_z(\alpha)$ for any given α.

Moreover, for $B(x) = (\beta_{kl}(x))$, we can construct confidence bands for its individual varying coefficient function $\beta_{kl}(x)$ for all (k,l), $k = 1, \ldots, 6$ and $l = 1, \ldots, r$. Specifically, at a given significance level α, we construct a confidence band for each $\beta_{kl}(x)$ such that

$$P(\hat{\beta}_{kl}^{L,\alpha}(x) < \beta_{kl}(x) < \hat{\beta}_{kl}^{U,\alpha}(x) \text{ for all } x \in [0, L_0]) = 1-\alpha, \tag{2.16}$$

where $\hat{\beta}_{kl}^{L,\alpha}(x)$ and $\hat{\beta}_{kl}^{U,\alpha}(x)$ are the lower and upper limits of the confidence band. Let $\mathbf{e}_{kl}$ be a $6r \times 1$ vector with the $(l-1)r + k$th element equal to 1 and all others equal to 0. It follows from Theorem 1 and the continuous mapping theorem that

$$\sup_{x\in[0,L_0]} |\sqrt{n}\{\hat{\beta}_{kl,e}(x) - \beta_{kl}(x)\}| \Rightarrow \sup_{x\in[0,L_0]} |\mathbf{e}_{kl}^T X_B(x)|.$$

We define the critical point $C_{kl}(\alpha)$ to satisfy $P(\sup_{x\in[0,L_0]}|\mathbf{e}_{kl}^T X_B(x)| \le C_{kl}(\alpha)) = 1-\alpha$. Thus, a $1-\alpha$ simultaneous confidence band for $\beta_{kl}(x)$ is given by

$$\left(\hat{\beta}_{kl,e}(x) - \frac{C_{kl}(\alpha)}{\sqrt{n}}, \hat{\beta}_{kl,e}(x) + \frac{C_{kl}(\alpha)}{\sqrt{n}}\right). \tag{2.17}$$

Similar to $C_z(\alpha)$, the critical point $C_{kl}(\alpha)$ can be approximated as the $1-\alpha$ empirical percentile of $\sup_{x\in[0,L_0]}|\mathbf{e}_{kl}^T X_B(x)^{(g)}|$ for all $g = 1,\dots,G$.

3. Simulation study. We conducted a Monte Carlo simulation study to examine the finite sample performance of VCDF. At each point x_j along the RICFT, the noisy diffusion tensors are simulated according to the following model:

$$S_i(x_j) = \exp(\text{Ivecs}(B(x_j)\mathbf{z}_i + \tau_i\hat{\mathbf{u}}_i(x_j) + \tau_i(x_j)\hat{\boldsymbol{\varepsilon}}_i(x_j))), \tag{3.1}$$

where τ_i and $\tau_i(x_j)$ were independently generated from a $N(0,1)$ random generator for $i = 1,\dots,n$ and $j = 1,\dots,n_G$. Specifically, we set $n = 96$, $n_G = 112$ and $\mathbf{z}_i = (1, \text{G}_i, \text{Gage}_i)$ for $i = 1,\dots,96$, where G_i and Gage_i, respectively, denote gender and gestational age. To mimic real imaging data, we applied our proposed VCDF method to DTs along the RICFT from all 96 infants in our clinical data to estimate $B(x)$ by $\hat{B}_e(x)$, $\mathbf{u}_i(x)$ by $\hat{\mathbf{u}}_{i,e}(x)$ via (2.8), and $\boldsymbol{\varepsilon}_i(x)$ by $\hat{\boldsymbol{\varepsilon}}_i(x) = \text{vecs}(\log(S_i(x)) - \hat{B}_e(x)\mathbf{z}_i - \hat{\mathbf{u}}_{i,e}(x))$. The curves of the varying coefficient functions of $\hat{B}_e(x)$ are presented in Figure 5. According to our real data analysis in Section 4, the gestational age effect is significant for our clinical data. So we fixed all functions in $B(x)$ at their corresponding functions in $\hat{B}_e(x)$ except that the third column of $B(x)$, denoted by $(\beta_{13}(x),\dots,\beta_{63}(x))^T$, was set as c times the third column of $\hat{B}_e(x)$ where c is set at different values in order to study the Types I and II error rates of our global test statistic in testing the gestational age effect. Figure 2(a) displays the simulated diffusion tensors along the RICFT at $c = 1$.

We have five aims in this simulation study. The first aim is to investigate the consequence of missing an important covariate. According to our real data analysis in Section 4, the Gage effect is significant, whereas the gender effect is not significant. We fitted two VCDF models, including three-covariate (intercept, gender and gestational age) and two-covariate (intercept and gender) models to smooth the DTs along the RICFT, and compare their performance in reconstructing the true DTs along the RICFT. Note that the two-covariate model does not include Gage_i as a covariate. Figure 2 presents the estimated diffusion tensors using the three-covariate model [Figure 2(c)] and the two-covariate model [Figure 2(d)]. Inspecting Figure 2(e) reveals that the three-covariate model leads a smaller mean geodesic distance between the true and estimated DTs compared with the two-covariate

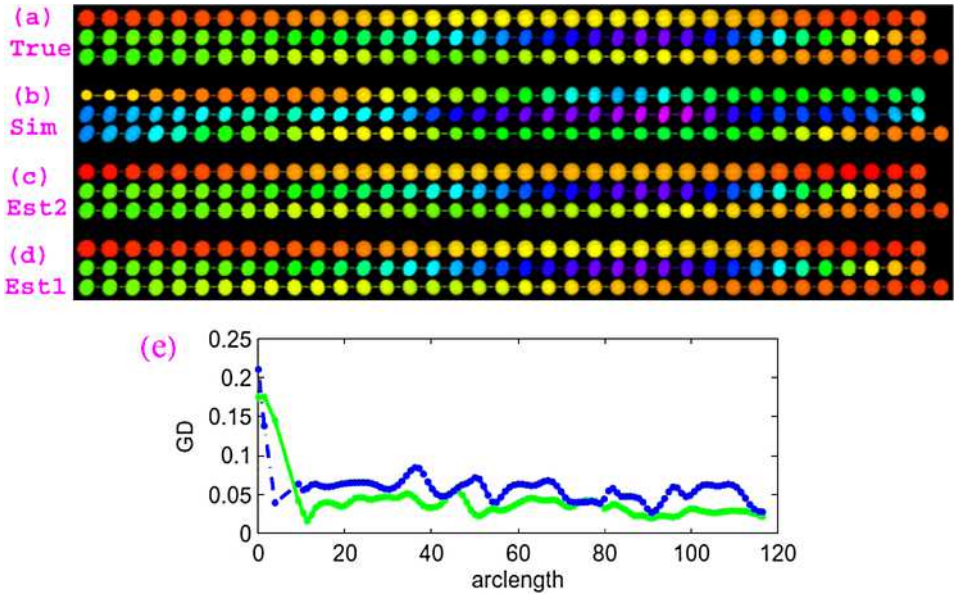

FIG. 2. *Ellipsoidal representations of the true* (a)*, simulated* (b) *and estimated* (c) *(based on three covariates) and* (d) *(based on two covariates) diffusion tensors along the RICFT, colored with their FA values. The rainbow color scheme is used with red corresponding to low FA value and purple corresponding to high FA value. Each set of* 3 *rows in* (a)–(d) *represents one tract of* 112 *DTs and the three rows are read from left to right in the top row, right to left in the middle row and then left to right in the bottom row.* (e) *Mean geodesic distances between the estimated and true diffusion tensors (green solid line based on three covariates and blue dash-dotted line based on two covariates) along the RICFT.*

model. Thus, the three-covariate model outperforms the two-covariate one in recovering the true DTs along the RICFT.

The second aim is to investigate the finite sample performance of the global test statistic $\mathbf{T}_n$ based on the whole DT. In neuroimaging studies, some scientific questions require the assessment of the development of diffusion tensors along fiber tracts across time. We formulated the questions as testing the null hypothesis $H_0 : \beta_{13}(x) = \cdots = \beta_{63}(x) = 0$ for all x along the RICFT. We first fixed $c = 0$ to assess the Type I error rates for $\mathbf{T}_n$, and then we set $c = 0.2, 0.4, 0.6,\ 0.8$ and 1.0 to examine the Type II error rates for $\mathbf{T}_n$ at different effect sizes.

We applied the estimation procedure of VCDF to the noisy DTs along the RICFT. We approximated the p-value of $\mathbf{T}_n$ by using the wild bootstrap method with $G = 1000$ described in Appendix B. For each c, we set the significance level α at both 0.05 and 0.01 and used 3000 replications to estimate the rejection rate of $\mathbf{T}_n$. At a fixed α, if the Type I rejection rate is smaller than α, then the test is conservative, whereas if the Type I rejection rate is greater than α, then the test is anticonservative, or liberal. Figure 3

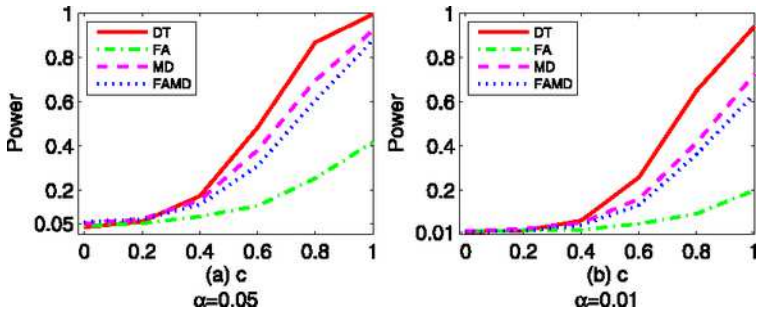

FIG. 3. *Simulation study: Types I and II error rates as functions of c. Rejection rates of T_n based on the wild bootstrap method are calculated at six different values of the effect size c for sample size* 96 *at the* (a) 0.05 *and* (b) 0.01 *significance levels using DTs, FA values, MD values, and joint values of FA and MD.*

presents the rejection rates of $\mathbf{T}_n$ across all effect sizes at the two significance levels ($\alpha = 0.05$ or 0.01) by using full diffusion tensors. It is observed that Type I error rates are well maintained at the two significance levels. In addition, the statistical power for rejecting the null hypothesis increases with the effect size and the significance level, which is consistent with our expectation.

The third aim is to demonstrate the power gain in using DTs compared with the sole use of diffusion properties. For each simulated diffusion tensor at $c = 0.2, 0.4, 0.6, 0.8$ and 1.0, we calculated its three eigenvalues λ_1, λ_2 and λ_3 and two well-known scalar diffusion properties MD and FA. To compare the power of our method based on DTs with other methods based on scalar diffusion properties, we applied an existing method for the analysis of diffusion properties in Zhu et al. (2011) to three different scenarios: (i) FA, (ii) MD and (iii) (FA, MD). Then we tested the gestational age effect in each scenario. Inspecting Figure 3 reveals that the statistical power for rejecting the null hypothesis increases with the effect size and the significance level in all scenarios. Moreover, compared with the sole use of diffusion properties, the use of DT dramatically increases the statistical power for rejecting the null hypothesis.

The fourth aim is to demonstrate the accuracy gain in estimating scalar diffusion properties along fiber tracts by directly modeling the DTs using VCDF. We compared two different methods for estimating FA's and MD's, here referred to as method A and method B, respectively. The method A first applies VCDF to estimate DT's and then calculates the FA or MD curve based on the estimated DT's. The method B first calculates the FA's or MD's from all SPD matrices and then uses varying coefficient methods in Euclidean space to estimate the FA's or MD's. We examined the finite sample performance of methods A and B by using the Mean Absolute Biases

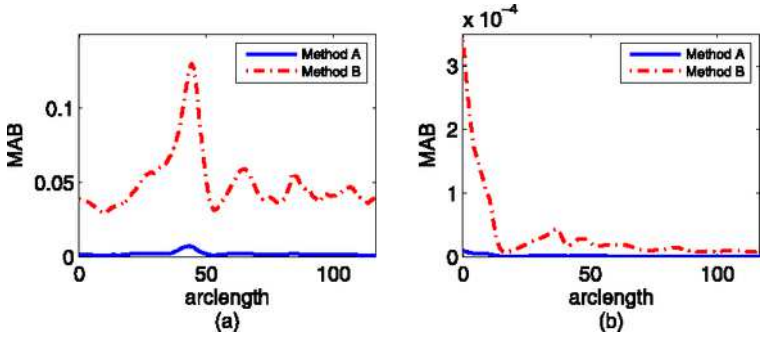

FIG. 4. *Plot of the MAB's of the estimated FA's* (a) *and MD's* (b) *using two methods A and B based on 3000 replications. The method A, which uses VCDF by directly modeling DT's, outperforms the method B in terms of smaller biases in estimating FA and MD values.*

(MAB) across all 112 locations, which is defined by

$$\text{MAB}_{Y,j} = 96^{-1} \sum_{i=1}^{96} \left| 3000^{-1} \sum_{s=1}^{3000} \hat{Y}_{sij} - Y_{ij} \right|, \tag{3.2}$$

where $\hat{Y}_{sij}$ is the estimator of Y_{ij}, which can be the estimated FA or MD value at the jth location for the ith subject and the sth simulation. Figure 4 reveals that method A has the smaller biases in estimating FA and MD values and the biases are negligible compared with those obtained using method B. This indicates the potential large improvement gained by directly modeling DT data over method B.

The fifth aim is to examine the coverage probabilities of the simultaneous confidence bands for all varying coefficient functions $\beta_{kl}(x)$ in $B(x)$ and $S(B(x), \mathbf{z})$. We only considered the generated diffusion tensor data at $c = 1$. We constructed the 95% and 99% simultaneous confidence bands for all $\beta_{kl}(x)$. Following Fan and Zhang (2000), we used a smaller bandwidth with a shrinkage factor 6 to improve the accuracy of the confidence bands.

Table 1 summarizes the empirical coverage probabilities based on 3000 replications for $\alpha = 0.01$ and 0.05. The coverage probabilities are quite close to the prespecified confidence levels. Figure 5 presents typical critical values of 95% simultaneous confidence regions for vectors of coefficient functions $\beta_{k\cdot}(x) = (\beta_{k1}, \ldots, \beta_{kr})^T, k = 1, \ldots, 6$. Figure 6 summarizes the empirical coverage probabilities for $S(B(x), \mathbf{z})$ based on 3000 replications at $\alpha = 0.01$ and 0.05. The coverage probabilities are quite close to the expected confidence levels.

4. Analysis of the right internal capsule fiber tract. We have two specific aims for the analysis of the right internal capsule fiber tracts. The first one is to compare DTs along the RICFT between the male and female

TABLE 1
Simulated coverage probabilities for varying coefficient functions in $B(x) = (\beta_{kl}(x))$ based on 3000 replications at the significance levels $\alpha = 0.01$ and 0.05

	$\alpha = 0.05$			$\alpha = 0.01$		
	Intercept $l = 1$	**Gender** $l = 2$	**Gage** $l = 3$	**Intercept** $l = 1$	**Gender** $l = 2$	**Gage** $l = 3$
$\beta_{1l}(x)$	0.9497	0.9420	0.9387	0.9867	0.9837	0.9810
$\beta_{2l}(x)$	0.9440	0.9443	0.9383	0.9843	0.9907	0.9857
$\beta_{3l}(x)$	0.9457	0.9383	0.9400	0.9870	0.9833	0.9807
$\beta_{4l}(x)$	0.9480	0.9457	0.9400	0.9880	0.9870	0.9850
$\beta_{5l}(x)$	0.9437	0.9350	0.9350	0.9870	0.9873	0.9823
$\beta_{6l}(x)$	0.9473	0.9400	0.9403	0.9860	0.9827	0.9797

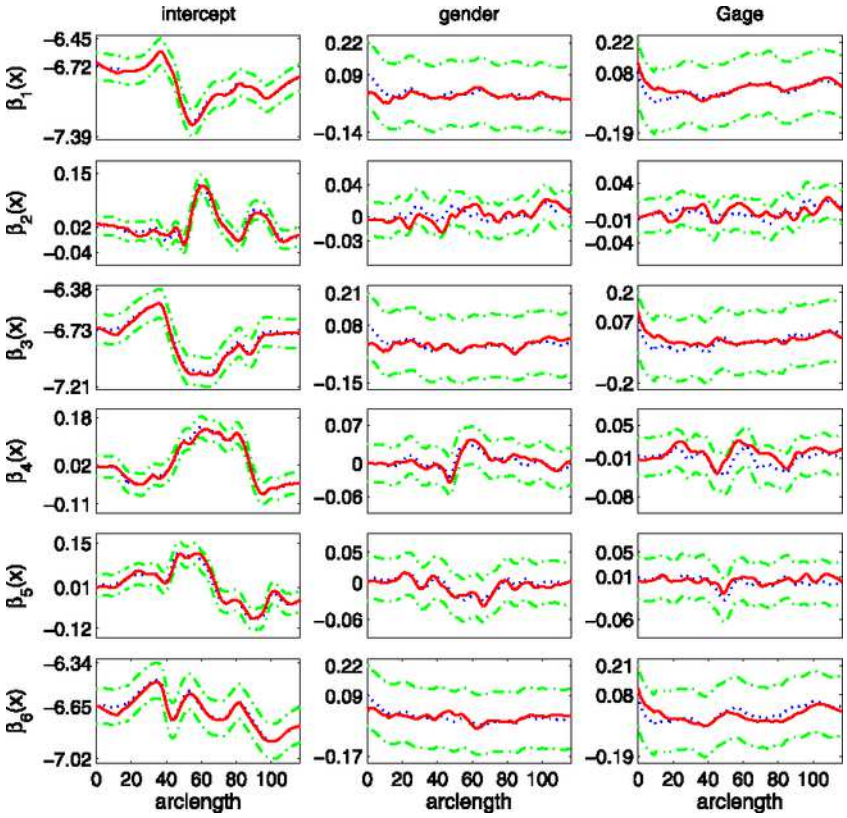

FIG. 5. *Typical 95% simultaneous confidence bands for varying coefficient functions $\beta_{kl}(x)$. The solid, dotted and dash-dotted curves are, respectively, the true curves, the estimated varying coefficient functions and their 95% confidence bands.*

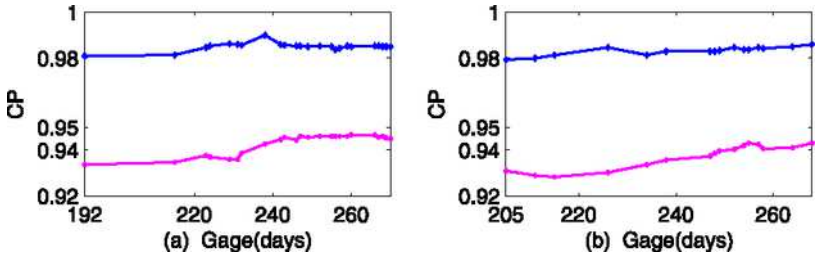

FIG. 6. *Simulated coverage probabilities for* $D(\mathbf{z}, \beta(x))$ *based on 3000 simulations for* $\alpha = 0.05$ *(solid lines with diamond markers) and* $\alpha = 0.01$ *(solid lines with circle markers),* (a) *for female* (b) *for male at different gestational ages, respectively.*

groups. The second one is to delineate the development of fiber DTs across time. To achieve these two aims, we fitted VCDF to DTs along the RICFT with gestational age at MRI scanning and gender as covariates. We applied the estimation procedure in Section 2 to estimate $B(x)$, $\Sigma_{\mathbf{u}}(\cdot,\cdot)$ and $\Sigma_{\boldsymbol{\varepsilon}}(\cdot,\cdot)$. Then, we constructed the global test statistics $\mathbf{T}_n$ and the local test statistics $T_n(x_j)$ to test the gender effect and the gestational age effect based on DTs along the RICFT. The p value of $\mathbf{T}_n$ was approximated by using the resampling method with $G = 5000$ replications. Finally, we constructed the 95% simultaneous confidence bands for the varying coefficient functions $\beta_{kl}(x)$.

To test the gender and gestational age effects, we calculated the local test statistics $T_n(x_j)$ and their corresponding p values across all points on the RICFT. It is shown in Figure 7(a) that most points do not have $-\log_{10}(p)$ values greater than 1.3 for testing the gender effect. Then, we also computed the global test statistic $\mathbf{T}_n = 797.65$ and its associated p-value $p = 0.3934$, indicating no gender effect. Inspecting Figure 7(b) reveals that the $-\log_{10}(p)$ values of $T_n(x_j)$ for testing the gestational age effect are extremely significant in the middle part of the RICFT. The global gestational age effect was also found to be highly significant with $\mathbf{T}_n = 5271.7$ and its p-value $p < 10^{-6}$. It indicates that DTs along the RICFT are significantly associated with the gestational age, even though there is no gender difference among DTs along the RICFT. In order to investigate the development of DTs across the gestational age, we chose a location at arclength $= 61.02$ and observed that the diffusion tensors become anisotropic and their sizes become smaller as gestational age increases [Figures 7(b) and (c)]. Recall that the three eigenvalues of a DT reflect the magnitude of the diffusion of water molecules along three directions parallel to its three eigenvectors and that MD reflects the total magnitude of the diffusion of water molecules. To show the decreasing trend of DT, we also plotted the curves of all three eigenvalues and MD values in Figures 7(e) and (g), respectively, both of which explicitly show that the first eigenvalue does not change much, whereas the second, third eigenvalues and MD values decrease with the gestational age. In addition, it is observed

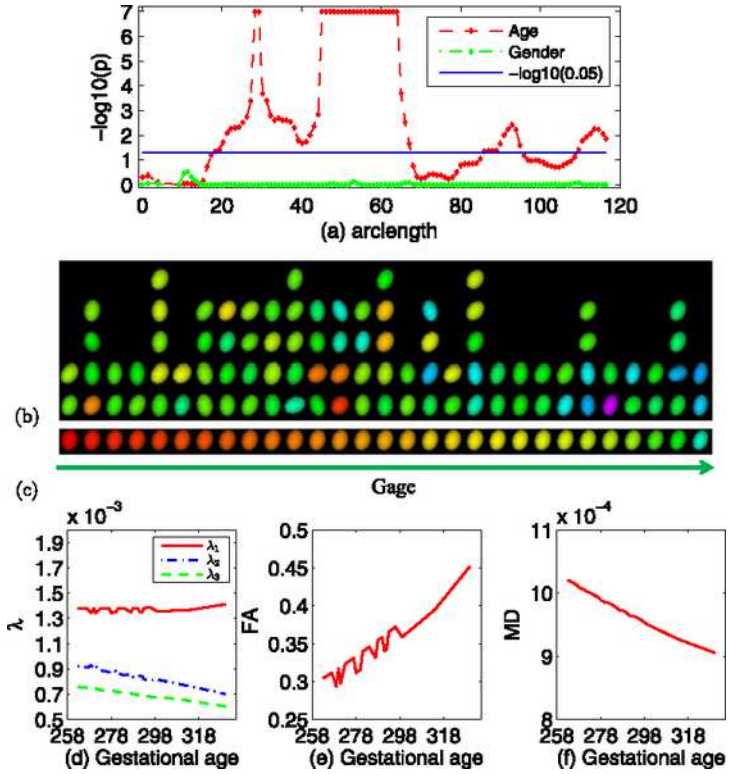

FIG. 7. (a) *The* $-\log_{10}(p)$ *values of test statistics* $T_n(x_j)$ *for testing gender or gestational age effect of diffusion tensors on the right internal capsule tract, which shows no significant gender effect and significant gestational age effect. The ellipsoidal representations of* (b) *raw and* (c) *smoothed diffusion tensors changing with the gestational age at one location (at* arclength $= 61.02$*) on the right internal capsule tract with significant gestational age effect, colored with FA values. The rainbow color scheme is used with red corresponding to low FA value and purple corresponding to high FA value. The plots of three eigenvalues* (d)*, FA* (e) *and MD* (f) *values at that location.*

from 7(f) that FA increases with gestational age, which indicates that DTs become more anisotropic as gestational age increases.

Figure 8 presents the estimated varying coefficient functions along with their 95% simultaneous confidence bands. In Figure 8 all simultaneous confidence bands contain the horizontal line crossing $(0, 0)$ for the gender effect, whereas the horizontal line is out of the 95% simultaneous confidence band for $\beta_{43}(x)$, which indicates the significant gestational age effect. This agrees with our previous analysis results based on the global and local test statistics for the gender and gestational age effects.

Finally, Figure 9 presents the 95% critical values for $S(B(x), \mathbf{z})$ and the estimated $S(B(x), \mathbf{z})$ along the RICFT across gestational age for female and

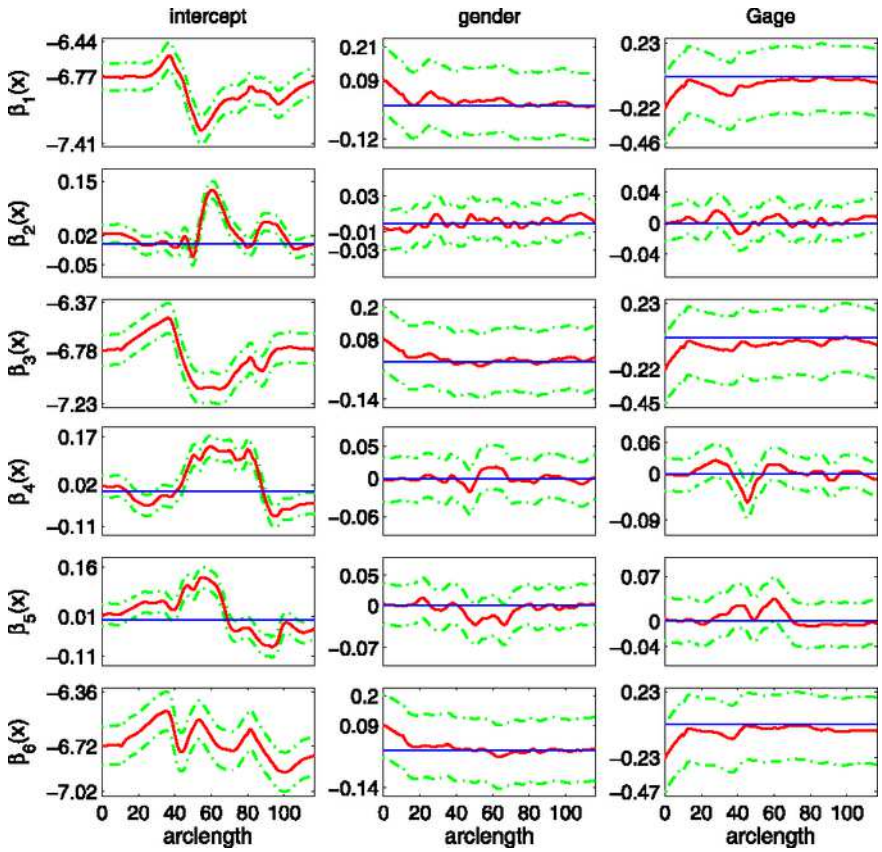

FIG. 8. *95% simultaneous confidence bands for coefficient functions. The solid curves are the estimated coefficient functions and the dashed curves are the 95% confidence bands. The thin horizontal line is the line crossing the origin.*

male groups, respectively. Inspecting Figure 9 reveals that the variation of $S(B(x), \mathbf{z})$ is larger on the two boundary points (especially on the right side) and smaller in the middle. In addition, the apparent trend of DT's changing with gestational age is shown at arc-length $= 61.02$ for both female and male groups.

5. Discussion. In this paper we have developed a functional data analysis framework, VCDF, for modeling diffusion tensors along fibber bundles in the Riemannian manifold of SPD matrices under the log-Euclidean metric with a set of covariates of interest. The most important characteristic of our method is that it is formulated based on the whole diffusion tensors instead of the DT derived scalar quantities and, thus, it can directly handle diffusion tensors. In addition, VCDF can characterize the dynamic association between functional DT-valued responses and covariates by using a set

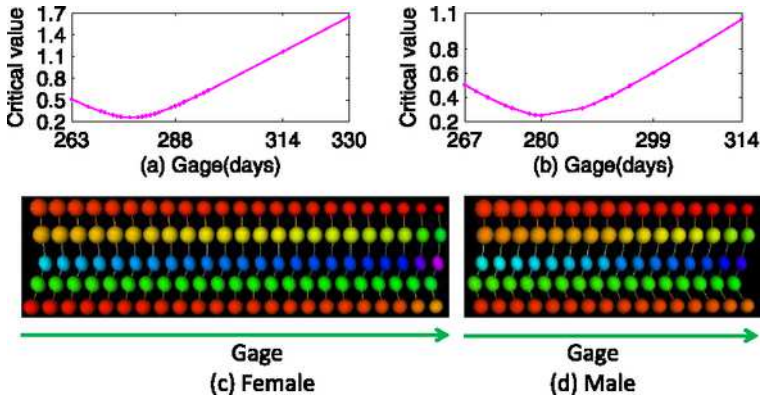

FIG. 9. *The* 95% *critical values for* $S(B(x), \mathbf{z})$ *across gestational ages for female* (a) *and male* (b) *groups, respectively. The ellipsoidal representation of the estimated* $S(B(x), \mathbf{z})$ *along the right internal capsule tract across gestational ages for female* (c) *and male* (d) *groups, respectively, colored with FA values. The rainbow color scheme is used with red corresponding to low FA value and purple corresponding to high FA value. The displayed four rows from the top to the bottom correspond to DTs at arclength* $0, 31.22, 61.02, 80.92$ *and* 116.47. *Specifically, the third row shows the apparent trend of DT's changing with gestational age for both female and male groups.*

of varying coefficient functions. Compared with the methods based on DT derived quantities, such as FA and MD, our method shows the apparent superiority in estimating DT derived quantities compared with those based on DT derived quantities (Figure 4). One reason is that the DT data which is estimated from DWIs is almost biased, whereas the DT derived quantities are linear and nonlinear functions of eigenvalues of DT data, which are very different from the ground truth. The other reason is that directly modeling DTs along fiber bundles as a smooth SPD process allows us to incorporate a smoothness constraint to further reduce noise in the estimated DTs along the fiber bundles. This leads to the further reduction of noise in estimated scalar diffusion properties along the fiber bundles. In addition, our method has the greater statistical power in detecting the effect of covariates of interest as is shown in Figure 3. One reason is that VCDF is less biased in parameter estimation. The other one is that our method accounts for all information contained in the DTs along the fiber bundles.

Several major issues remain to be addressed in future research. All fiber-tract-based methods including VCDF are only applicable to these prominent white matter tracts and do not account for the uncertainties of tracking these fiber tracts. It is important to develop new statistical methods to appropriately account for such uncertainties in fiber-tract analysis especially for inconspicuous fiber tracts. VCDF is based on the second-order diffusion tensor. It may be interesting to extend VCDF to the analysis of high angular

resolution diffusion imaging (HARDI), which is important for resolving the issue of fiber crossing [Assemlal et al. (2011)]. Furthermore, it would be of great interest to extend VCDF to longitudinal studies and family studies. Finally, we have treated DTs along fiber tracts as functional responses; it would be interesting to treat DTs along fiber tracts as varying covariate functions to predict a scalar outcome (e.g., diagnostic group) [Goldsmith et al. (2011)].

APPENDIX A: ASSUMPTIONS

ASSUMPTION C1. $\boldsymbol{\varepsilon}_i(x)$ and $\mathbf{u}_i(x)$ are identical and independent copies of $\mathrm{SP}(0, \Sigma_{\boldsymbol{\varepsilon}})$ and $\mathrm{SP}(0, \Sigma_{\mathbf{u}})$, respectively. $\boldsymbol{\varepsilon}_i(x)$ and $\boldsymbol{\varepsilon}_i(x')$ are independent for any $x \neq x' \in [0, L_0]$. $\boldsymbol{\varepsilon}_i(x)$ and $\mathbf{u}_i(x')$ are independent for any $x, x' \in [0, L_0]$. Moreover, with probability one, the sample path of $\mathbf{u}_i(x)$ has continuous second-order derivative on $[0, L_0]$ and $E[\sup_{x\in[0,L_0]} \|\mathbf{u}_i(x)\|_2^{r_1}] < \infty$ and $E\{\sup_{x\in[0,L_0]}[\|\dot{\mathbf{u}}_i(x)\|_2 + \|\ddot{\mathbf{u}}_i(x)\|_2]^{r_2}\} < \infty$ for all $r_1, r_2 \in (2, \infty)$, where $\|\cdot\|_2$ is the Euclidean norm.

ASSUMPTION C2. All components of $B(x)$ and $\Sigma_{\boldsymbol{\varepsilon}}(x, x)$ have continuous second-order derivatives on $[0, L_0]$. The fourth moments of $\boldsymbol{\varepsilon}_i(x)$ are continuous on $[0, L_0]$. All components of $\Sigma_{\mathbf{u}}(x, x')$ have continuous second-order partial derivatives with respect to $(x, x') \in [0, L_0]^2$. Moreover, $\Sigma_{\boldsymbol{\varepsilon}}(x, x)$ and $\Sigma_{\mathbf{u}}(x, x)$ are positive for all $x \in [0, L_0]$.

ASSUMPTION C3. The points $\mathcal{X} = \{x_j, j = 1, \ldots, n_G\}$ are independently and identically distributed with density function $\pi(x)$, which has the bounded support $[0, L_0]$. For some constants π_L and $\pi_U \in (0, \infty)$ and any $x \in [0, L_0]$, $\pi_L \leq \pi(x) \leq \pi_U$ and $\pi(x)$ has continuous second-order derivative.

ASSUMPTION C4. The kernel function $K(t)$ is a symmetric density function with a compact support $[-1, 1]$ and is Lipschitz continuous.

ASSUMPTION C5. The covariate vectors $\mathbf{z}_i$ are independently and identically distributed with $E\mathbf{z}_i = \mu_z$ and $E[\|\mathbf{z}_i\|_2^4] < \infty$ and that $E[\mathbf{z}_i^{\otimes 2}] = \Omega_Z$ is invertible.

ASSUMPTION C6. Both n and n_G converge to ∞, $h^{(1)} = o(1)$, $n_G h^{(1)} \to \infty$, and $h^{(1)-1}|\log h^{(1)}|^{1-2/q_1} \leq n_G^{1-2/q_1}$, where $q_1 \in (2, 4)$.

ASSUMPTION C7. $E[\|\boldsymbol{\varepsilon}_i(x_j)\|_2^{q_2}] < \infty$ for some $q_2 \in (4, \infty)$, $\max(h^{(2)}, h^{(3)}) = o(1)$, $n_G(h^{(2)} + h^{(3)}) \to \infty$, $(h^{(2)})^{-4}(\log n/n)^{1-2/q_2} = o(1)$, and $(h^{(3)})^{-2}(\log n/n)^{1-2/q_2} = o(1)$.

APPENDIX B: WILD BOOTSTRAP METHOD

We develop the four key steps of the wild bootstrap method for approximating the p-value of $\mathbf{T}_n$ as follows.

Step (i): Use the weighted least squares estimation to fit model (2.1) under the linear constraint specified in H_0, which yields $\hat{B}_e^*(x_j)$. Calculate $\hat{\mathbf{u}}_{i,e}^*(x_j)$ according to (2.8) and $\hat{\boldsymbol{\varepsilon}}_{i,e}^*(x_j) = \operatorname{vecs}(\log(S_i(x_j))) - \hat{B}_e(x_j)^*\mathbf{z}_i - \hat{\mathbf{u}}_{i,e}^*(x_j)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_G$.

Step (ii): Generate a random sample $\tau_i^{(g)}$ and $\tau_i(x_j)^{(g)}$ from a $N(0,1)$ random generator for $i = 1, \ldots, n$ and $j = 1, \ldots, n_G$ and then construct

$$\hat{S}_i(x_j)^{(g)} = \exp(\operatorname{Ivecs}(\hat{B}_e^*(x_j)\mathbf{z}_i + \tau_i^{(g)}\hat{\mathbf{u}}_{i,e}^*(x_j) + \tau_i(x_j)^{(g)}\hat{\boldsymbol{\varepsilon}}_{i,e}^*(x_j))).$$

Then, based on $\hat{S}_i(x_j)^{(g)}$, we recalculate $\hat{B}_e(x)^{(g)}$, and $\mathbf{d}(x)^{(g)} = C\hat{B}_e(x)^{(g)} - \mathbf{b}_0(x)$. We compute

$$\mathbf{T}_n^{(g)} = \int_0^{L_0} T_n(x)^{(g)}\,dx,$$

$$T_n(x_j)^{(g)} = n\,\mathbf{d}(x_j)^{(g)T}\{C(\hat{\Sigma}_{\mathbf{u}}(x_j, x_j) \otimes \hat{\Omega}_{\mathbf{z}}^{-1})C^T\}^{-1}\,\mathbf{d}(x_j)^{(g)}$$

for $j = 1, \ldots, n_G$.

Step (iii): Aggregate the results of Step (ii) over $g = 1, \ldots, G$ to obtain $\{T_{n,\max}^{(g)} = \max_{1 \le j \le n_G} T_n(x_j)^{(g)} : g = 1, \ldots, G\}$ and calculate $p(x_j) = G^{-1} \times \sum_{g=1}^{G} 1(T_{n,\max}^{(g)} \ge T_n(x_j))$ for each x_j. The $p(x_j)$ is the corrected p-value at the location x_j.

Step (iv): Aggregate the results of Step (ii) over $g = 1, \ldots, G$ to obtain $\{\mathbf{T}_n^{(g)} : g = 1, \ldots, G\}$ and calculate $p = G^{-1}\sum_{g=1}^{G} 1(\mathbf{T}_n^{(g)} \ge \mathbf{T}_n)$. If p is smaller than a prespecified significance level α, say, 0.05, then we reject the null hypothesis H_0.

APPENDIX C: RESAMPLING METHOD FOR APPROXIMATING GAUSSIAN PROCESS

Recall that $B_{h^{(1)}}(x) = [B(x), h^{(1)}\dot{B}(x)]$ in (2.4) is a $6 \times 2r$ matrix. It can be shown that $\hat{B}_{h^{(1)}}(x)^T$ is given by

$$\Sigma(h^{(1)}, x)^{-1} \sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)[\mathbf{z}_i \otimes \mathbf{y}_{h^{(1)}}(x_j - x)]\operatorname{vecs}(\log(S_i(x_j)))^T, \tag{C.1}$$

where $\Sigma(h^{(1)}, x) = \sum_{i=1}^{n}\sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x)[\mathbf{z}_i^{\otimes 2} \otimes \mathbf{y}_{h^{(1)}}(x_j - x)^{\otimes 2}]$. Thus, we can obtain $\hat{B}(x; h^{(1)})$ as follows:

$$\hat{B}(x; h^{(1)}) = [I_r \otimes (1, 0)]\hat{B}_{h^{(1)}}(x). \tag{C.2}$$

To approximately simulate from the Gaussian process $X_B(\cdot)$, we develop a resampling method as follows:

- Based on $\hat{B}(x_j; h^{(1)})$, we calculate $\hat{\mathbf{r}}_i(x_j) = \text{vecs}(\log(S_i(x_j))) - \hat{B}(x_j; h^{(1)})\mathbf{z}_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_G$.
- For $g = 1, \ldots, G$, we independently simulate $\{\tau_i^{(g)} : i = 1, \ldots, n\}$ from $N(0, 1)$.
- For $g = 1, \ldots, G$, we calculate a stochastic process $X_B(x)^{(g)}$ given by

$$\sqrt{n}[I_r \otimes (1,0)]\Sigma(h^{(1)}, x)^{-1} \sum_{i=1}^{n} \tau_i^{(g)} \sum_{j=1}^{n_G} K_{h^{(1)}}(x_j - x) C_i(x_j - x; h^{(1)}) \hat{\mathbf{r}}_{i,l}(x_j)^T,$$

where $C_i(x_j - x; h^{(1)}) = [\mathbf{z}_i \otimes \mathbf{y}_{h^{(1)}}(x_j - x)]$ is a $2r \times 1$ vector.

REFERENCES

Anderson, A. W. (2001). Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magn. Reson. Med.* **46** 1174–1188.

Arsigny, V. (2006). Processing data in lie groups: An algebraic approach. Application to non-linear registration and diffusion tensor MRI. Ph.D. thesis, Ecole Polytechnique.

Assemlal, H.-E., Tschumperlé, D., Brun, L. and Siddiqi, K. (2011). Recent advances in diffusion MRI modeling: Angular and radial reconstruction. *Med. Image Anal.* **15** 369–396.

Barmpoutis, A., Vemuri, B. C., Shepherd, T. M. and Forder, J. R. (2007). Tensor splines for interpolation and approximation of DT–MRI with applications to segmentation of isolated rat hippocampi. *IEEE Trans. Med. Imaging* **26** 1537–1546.

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

Basser, P. J., Mattiello, J. and LeBihan, D. (1994b). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* **66** 259–267.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J. and Aldroubi, A. (2000). In vivo fiber tractography using DT–MRI data. *Magn. Reson. Med.* **44** 625–632.

Davis, B. C., Bullitt, E., Fletcher, P. T. and Joshi, S. (2010). Population shape regression from random design data. *Int. J. Comput. Vis.* **90** 255–266.

Dryden, I. L., Koloydenko, A. and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. MR2750388

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20** 2008–2036. MR1193323

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. Chapman & Hall, London. MR1383587

Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 57–80. MR1959093

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **27** 715–731. MR1804172

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface* **1** 179–195. MR2425354

Goldsmith, A. J., Crainiceanu, C. M., Caffo, B. S. and Reich, D. (2011). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *NeuroImage* **57** 431–439.

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** S133–S142.

Kim, P. T. and Richards, D. S. (2011). Deconvolution density estimation on spaces of positive definite symmetric matrices. In *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P. Hettmansperger* 147–168. World Scientific Press, Singapore.

Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal.* **84** 299–318. MR1965224

O'Donnell, L. J., Westin, C.-F. and Golby, A. J. (2009). Tract-based morphometry for white matter group analysis. *NeuroImage* **45** 832–844.

Pierpaoli, C. and Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magn. Reson. Med.* **36** 893–906.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

Rose, J., Mirmiran, M., Butler, E., Lin, C., Barnes, P. D., Kermoian, R. and Stevenson, D. K. (2007). Neonatal microstructural development of the internal capsule on diffusion tensor imaging correlates with severity of gait and motor deficits. *Developmental Medicine and Child Neurology* **49** 745–750.

Schwartzman, A., Mascarenhas, W. F. and Taylor, J. E. (2008). Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann. Statist.* **36** 2886–2919. MR2485016

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M. and Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage* **31** 1487–1505.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing. Monographs on Statistics and Applied Probability* **60**. Chapman & Hall, London. MR1319818

Wang, L., Li, H. and Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J. Amer. Statist. Assoc.* **103** 1556–1569. MR2504204

Welsh, A. H. and Yee, T. W. (2006). Local regression for vector responses. *J. Statist. Plann. Inference* **136** 3007–3031. MR2281238

Wu, C. O. and Chiang, C.-T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist. Sinica* **10** 433–456. MR1769751

Wu, H. and Zhang, J.-T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Wiley-Interscience, Hoboken, NJ. MR2216899

Yuan, Y., Zhu, H., Ibrahim, J. G., Lin, W. and Peterson, B. S. (2008). A note on the validity of statistical bootstrapping for estimating the uncertainty of tensor parameters in diffusion tensor images. *IEEE Trans. Med. Imaging* **27** 1506–1514.

Yuan, Y., Zhu, H., Lin, W. and Marron, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74** 697–719. MR2965956

Yushkevich, P. A., Zhang, H., Simon, T. J. and Gee, J. C. (2008). Structure-specific statistical mapping of white matter tracts. *NeuroImage* **41** 448–461.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698

Zhu, H., Li, R. and Kong, L. (2010). Multivariate varying coefficient models for functional responses. Technical report, Univ. North Carolina at Chapel Hill.

Zhu, H., Ibrahim, J. G., Tang, N., Rowe, D. B., Hao, X., Bansal, R. and Peterson, B. S. (2007a). A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans. Med. Imaging* **26** 954–966.

Zhu, H., Zhang, H., Ibrahim, J. G. and Peterson, B. S. (2007b). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *J. Amer. Statist. Assoc.* **102** 1085–1102. MR2412530

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. *J. Amer. Statist. Assoc.* **104** 1203–1212. MR2750245

Zhu, H., Styner, M., Tang, N., Liu, Z., Lin, W. and Gilmore, J. H. (2010). FRATS: Functional regression analysis of DTI tract statistics. *IEEE Trans. Med. Imaging* **29** 1039–1049.

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** 1412–1425.

Y. Yuan
Department of Biostatistics
MS 768, Room 6009
St. Jude Children's Research Hospital
262 Danny Thomas Place
Memphis, Tennessee 38105-3678
USA
E-mail: ying.yuan@stjude.org

H. Zhu
M. Styner
J. H. Gilmore
J. S. Marron
University of North Carolina at Chapel Hill
3108-B McGavran Greenberg Hall
CB#7420
Chapel Hill, North Carolina 27599
USA
E-mail: htzhu@email.unc.edu
martin_styner@ieee.org
John_Gilmore@med.unc.edu
marron@email.unc.edu