Matlab Tool: Functional Regression Analysis of DTI tract Statistics

1 Introduction

1.1 FRATS summary

Diffusion tensor imaging (DTI) provides important information on the structure of white matter fiber bundles as well as detailed tissue properties along these fiber bundles in vivo. A functional regression framework, called FRATS and implemented by Matlab, was presented for the analysis of multiple diffusion properties along fiber bundle as functions in an infinite dimensional space and their association with a set of covariates of interest, such as age, diagnostic status and gender, in real applications. The functional regression framework consists of four integrated components: (1) the local polynomial kernel method for smoothing multiple diffusion properties along individual fiber bundles, (2) a functional linear model for characterizing the association between fiber bundle diffusion properties and a set of covariates, (3) a global test statistic for testing hypotheses of interest, and (4) a resampling method for approximating the p-value of the global test statistic. The resulting analysis pipeline can be used for understanding normal brain development, the neural bases of neuropsychiatric disorders, and the joint effects of environmental and genetic factors on white matter fiber bundles.

1.2 Motivation

Diffusion Tensor Imaging (DTI), which can track the effective diffusion of water in the human brain in vivo, has been widely used to map the structure and orientation of the white matter fiber tracts of the brain (Basser et al., 1994b,a). In the current literature, three major approaches to the group analysis of diffusion imaging data are region-of-interest (ROI) analysis, voxel based analysis, and fiber tract based analysis (Smith et al., 2006; O'Donnell et al., 2009; Snook et al., 2007). The ROI analysis used in some neuroimaging studies (Bonekam et al., 2008; Gilmore et al., 2008) primarily suffers from the difficulty in identifying meaningful ROIs. Voxel based analysis is used more commonly than ROI analysis in neuroimaging studies (Chen et al., 2009; Focke et al., 2008; Camara et al., 2007; Snook et al., 2005). The major drawbacks of voxel based analysis include the issues of alignment quality and the arbitrary choice of smoothing extent (Hecke et al., 2009; Ashburner and Friston, 2000; Smith et al., 2006; Jones et al., 2005). With the drawbacks mentioned of the ROI and voxel based analysis, there is a growing interest in the DTI literature in developing fiber tract based analysis

of diffusion properties (Smith et al., 2006; O'Donnell et al., 2009; Yushkevich et al., 2008; Goodlett et al., 2009; Zhu et al., 2010). Statistically, diffusion properties along fiber bundles are functional data and its analysis requires advanced functional data analysis methods (Li and Hsing, 2010; Yao and Lee, 2006; Hall et al., 2006; Ramsay and Silverman, 2005, 2002). Functional data analysis methods for the statistical analysis of diffusion properties along fiber tracts, a "smoothing first, then estimation" procedure, was also developed (Goodlett et al., 2009). Their method is limited to a univariate diffusion property and cannot control for other covariates of interest, such as age, gender and behavioral variables. Moreover, the permutation test used there ignores substantial noise in the original data and can lead to misleading results.

What these three methods do not account for is the comparison of fiber bundle diffusion properties across groups and the development of fiber bundle diffusion properties along time, while controlling for other covariates of interest, such as gender (Chen et al., 2009; Bonekam et al., 2008; Smith et al., 2006; Focke et al., 2008; Camara et al., 2007; Snook et al., 2005). Making these comparisons requires a regression modeling framework for the analysis of fiber bundle diffusion properties and a set of covariates of interest, such as age, diagnostic status and gender. This tool presents a functional regression analysis of DTI tract statistics, called FRATS, for modeling the relationship between fiber bundle diffusion properties and covariates of interest.

1.3 FRATS description

Compared with (Goodlett et al., 2009) and other existing literature, literature, there are four methodological contributions in this paper. First, the local polynomial kernel method is used to regularize multiple diffusion properties along individual fiber bundles. Second, a functional linear model is developed to characterize the association between fiber bundle diffusion properties and any covariate of interest. Third, a global test statistic is proposed for testing hypothesis of interest. Fourth, a resampling method is developed for estimating the p-value of the global test statistic. A schematic overview of FRACTS is given in Figure 1. We describe each of these components briefly below. Detailed description can be found in Zhu et al. (2010).

1. Nonparametric Model

For the i-th subject, we consider a $m \times 1$ vector of diffusion properties, denoted by $\mathbf{y}_{i,j} = (y_{ij,1}, \cdots, y_{ij,m})^T$, and its associated arc length s_j for the j-th location grid point on the fiber bundle for $j = 1, \cdots, L_0$ and $i = 1, \cdots, n$, where L_0 and n denote the numbers of grid points and subjects, respectively. The nonparametric model is given by

$$\mathbf{y}_{i,j} = \mathbf{f}_i(s_j) + \epsilon_{i,j}, \tag{1}$$

where $\mathbf{f}_i(s) = (f_{i,1}(s), \cdots, f_{i,m}(s))^T$ is an $m \times 1$ vector of continuous functions with second-order continuous derivative, $E[\mathbf{y}_{i,j}|\mathbf{f}_i(s_j)] = \mathbf{f}_i(s_j)$, and $\mathrm{Cov}[\mathbf{y}_{i,j}|\mathbf{f}_i(s_j)] = \Sigma(s_j)$. Using Taylor's expansion, we can expand $\mathbf{f}_i(s_j)$ at s to obtain

$$\mathbf{f}_i(s_j) = \mathbf{f}_i(s) + \dot{\mathbf{f}}_i(s)(s_j - s) = \mathbf{A}_i \mathbf{z}_j, \tag{2}$$

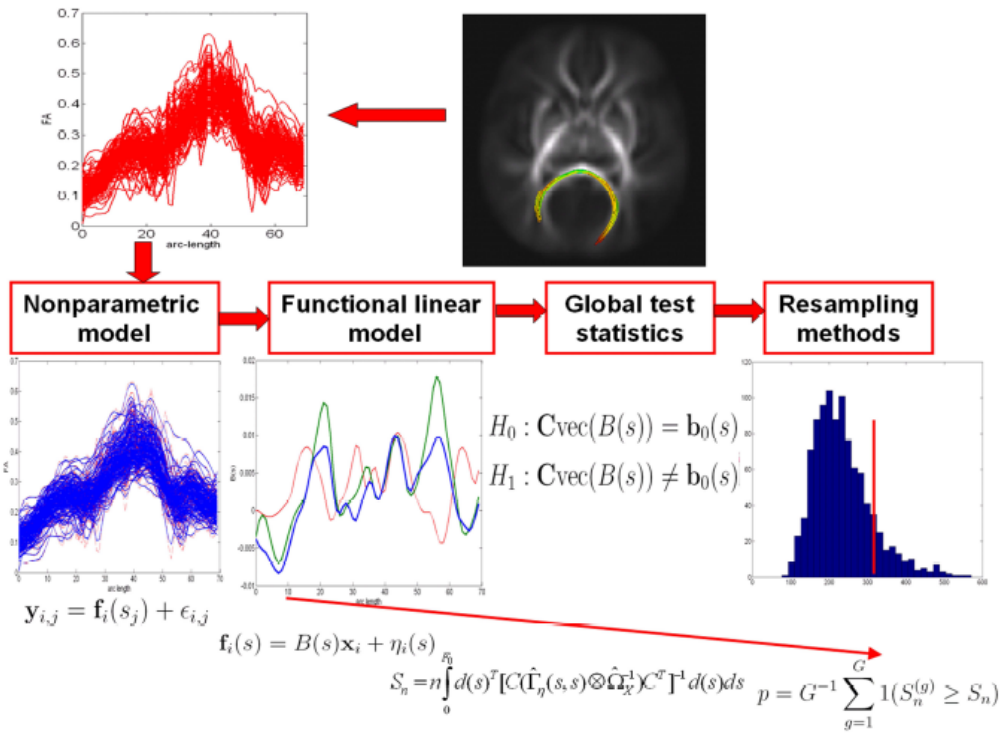

Figure 1: A schematic overview of FRATS: a nonparametric model for regularizing individual tracts, a functional linear model, a global test statistic for hypothesis testing, and a resampling method for estimating the p-value of the global test statistic.

where $\mathbf{z}_j = (1, s_j - s)^T$ and $\mathbf{A}_i = [\mathbf{f}_i(s) \;\; \dot{\mathbf{f}}_i(s)]$ is an $m \times 2$ matrix with $\dot{\mathbf{f}}_i(s) = d\mathbf{f}_i(s)/ds$. We develop an algorithm to estimate $\mathbf{A}_i$ as follows.

Step (1.1). Step (1.1) is to construct an initial estimate of $\mathbf{f}_i(s)$ for each i. Let $\mathbf{a}_{i;k}$ be the k-th row of $\mathbf{A}_i$ and $K(\cdot)$ be a kernel function. For each k and a fixed bandwidth h_k, we estimate $\mathbf{a}_{i;k}$ by minimizing an objective function given by

$$\sum_{j=1}^{L_0} (y_{ij,k} - \mathbf{a}_{i;k}^T \mathbf{z}_j)^2 K_{h_k}(s_j - s), \tag{3}$$

where $K_{h_k}(\cdot) = K(\cdot/h_k)/h_k$ is a rescaled kernel function. With some calculation, it can be shown that

$$\hat{\mathbf{a}}_{i;k} = [\sum_{j=1}^{L_0} \mathbf{z}_j K_{h_k}(s_j - s)\mathbf{z}_j^T]^{-1} \sum_{j=1}^{L_0} K_{h_k}(s_j - s)\mathbf{z}_j y_{ij,k}. \tag{4}$$

Let $\mathbf{e}_{1,2} = (1, 0)^T$. Then,

$$\hat{f}_{i,k}(s) = \mathbf{e}_{1,2}^T \hat{\mathbf{a}}_{i;k} = \sum_{j=1}^{L_0} \tilde{K}_{h_k}^0(s_j - s, s) y_{ij,k}, \tag{5}$$

where $\tilde{K}^0_{h_k}(\cdot,\cdot)$ are the empirical equivalent kernels Fan and Gijbels (1996). Thus, $\hat{\mathbf{f}}_{i,k} = (\hat{f}_{i,k}(s_1), \cdots, \hat{f}_{i,k}(s_{L_0}))^T = S_{i,k}\mathbf{y}_{i:,k}$, where $\mathbf{y}_{i:,k} = (y_{i1,k}, \cdots, y_{iL_0,k})^T$ and $S_{i,k}$ is the smoother matrix for the k-th measurement of the i-th subject. For each k, we pool the data from all n subjects and select the optimal bandwidth h_k, denoted by $\hat{h}^{(1)}_{k,opt}$, by minimizing the generalized cross-validation score given by

$$\mathrm{GCV}_k(h_k) = (n)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{L_0}\frac{[y_{ij,k} - \hat{f}_{i,k}(s_j)]^2}{1 - L_0^{-1}\mathrm{tr}(S_{i,k})}. \tag{6}$$

Based on the optimal $\hat{h}^{(1)}_{k,opt}$, we can estimate $\hat{f}_{i,k}(s)$ for all i.

Step (1.2). Step (1.2) is to construct an estimator of the covariance matrix $\Sigma(s_j)$ at s_j. Specifically, we consider the unbiased sample covariance matrix at s_j given by

$$\hat{\Sigma}(s_j) = (n-m)^{-1}\sum_{i=1}^{n}[\mathbf{y}_{i,j} - \hat{\mathbf{f}}_i(s_j)]^{\otimes 2}, \tag{7}$$

where $\hat{\mathbf{f}}_i(s) = (\hat{f}_{i,1}(s), \cdots, \hat{f}_{i,m}(s))^T$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector $\mathbf{a}$. It can be shown that $\hat{\Sigma}(s_j)$ converges to the true $\Sigma(s_j)$ in probability as both n and L_0 go to infinity.

Step (1.3). Step (1.3) is to compute an adaptive estimator of $\mathbf{f}_i(s)$ for all i using the initial results from Steps (1.1) and (1.2). For all k and a fixed bandwidth h, we estimate $\mathbf{A}_i$ by minimizing an objective function given by

$$\sum_{j=1}^{L_0}(\mathbf{y}_{ij} - \mathbf{A}_i\mathbf{z}_j)^T\hat{\Sigma}(s_j)^{-1}(\mathbf{y}_{ij} - \mathbf{A}_i\mathbf{z}_j)K_h(s_j - s). \tag{8}$$

Let $\mathbf{Z}_j = \text{block diagonal}(\mathbf{z}_j^T, \mathbf{z}_j^T, \cdots, \mathbf{z}_j^T)$ be an $m \times 2m$ matrix and $\mathbf{B}_i = (\mathbf{a}_{i;1}^T, \cdots, \mathbf{a}_{i;m}^T)^T$. It can be shown that

$$\hat{\mathbf{B}}_i = [\sum_{j=1}^{L_0} K_h(s_j - s)\mathbf{Z}_j^T\hat{\Sigma}(s_j)^{-1}\mathbf{Z}_j]^{-1}\sum_{j=1}^{L_0} K_h(s_j - s)\mathbf{Z}_j^T\hat{\Sigma}(s_j)^{-1}\mathbf{y}_{ij}, \tag{9}$$

which leads to a new estimator of $f_{i,k}(s)$, denoted by $\hat{f}_{i,k}(s)^{sec}$ for each i and k. Let $\tilde{S}_{i,k}$ be the smoother matrix for the k-th measurement of the i-th subject such that $\hat{\mathbf{f}}^{sec}_{i,k} = (\hat{f}_{i,k}(s_1)^{sec}, \cdots, \hat{f}_{i,k}(s_{L_0})^{sec})^T = \tilde{S}_{i,k}\mathbf{y}_{i:,k}$. We pool the data from all n subjects and m measurements and select the optimal bandwidth h, denoted by $\hat{h}_{opt}$, by minimizing the generalized cross-validation score given by

$$\mathrm{GCV}(h) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{L_0}\frac{[\mathbf{y}_{ij} - \hat{\mathbf{f}}_i(s_j)^{sec}]^T\hat{\Sigma}(s_j)^{-1}[\mathbf{y}_{ij} - \hat{\mathbf{f}}_i(s_j)^{sec}]}{1 - (mL_0)^{-1}\sum_{k=1}^{m}\mathrm{tr}(\tilde{S}_{i,k})}, \tag{10}$$

where $\hat{\mathbf{f}}_i(s)^{sec} = (\hat{f}_{i,1}(s)^{sec}, \cdots, \hat{f}_{i,m}(s)^{sec})^T$. Based on the optimal $\hat{h}_{opt}$, we can estimate $\hat{f}_{i,k}(s)^{sec}$ for all i and k. Similar to the arguments in Welsh and Yee (2006), it can be shown

that when $\Sigma(s)$ varies across s, $\hat{f}_{i,k}(s)^{sec}$ based on the optimal $\hat{h}_{opt}$ is more accurate than $\hat{f}_{i,k}(s)$ obtained from Step (1.1).

Step (1.4). Step (1.4) is to estimate the mean function $\mathbf{f}(s)$ and the covariance function $\Gamma(s,t)$ of $\mathbf{f}_i(s)$. Specifically, following Zhang and Chen (2007a); Ramsay and Silverman (2005), we can estimate $\mathbf{f}(s)$ and $\Gamma(s,t)$ by using their empirical counterparts of the estimated $\hat{\mathbf{f}}_i(s)^{sec}$ as

$$\hat{\mathbf{f}}(s) = n^{-1}\sum_{i=1}^{n}\hat{\mathbf{f}}_i(s)^{sec} \quad \text{and} \tag{11}$$

$$\hat{\Gamma}(s,t) = n^{-1}\sum_{i=1}^{n}[\hat{\mathbf{f}}_i(s)^{sec}-\hat{\mathbf{f}}(s)][\hat{\mathbf{f}}_i(t)^{sec}-\hat{\mathbf{f}}(t)]^T. \tag{12}$$

The diagonal of $\Gamma(s,s)$ reflects the variance of $\mathbf{f}_i(s)$ at the location s.

2. Functional Linear Model

We develop a functional linear model to characterize the relationship between all diffusion properties along fiber tracts and a set of covariates of interest, such as age, group, and gender. We assume that

$$\mathbf{f}_i(s) = B(s)\mathbf{x}_i + \eta_i(s), \quad i = 1,\cdots,n, \tag{13}$$

where $B(s)$ is a $m \times p$ matrix of functions of s, $\mathbf{x}_i$ is a $p \times 1$ vector of covariates of interest, and $\eta_i(s)$ satisfies $E[\eta_i(s)|\mathbf{x}_i] = 0$ and $\mathrm{Cov}[\eta_i(s),\eta_i(t)|\mathbf{x}_i] = \Gamma_\eta(s,t)$. $B(s)$ characterizes the association between fiber bundle diffusion properties and the covariates of interest $\mathbf{x}_i$. We develop an estimation algorithm to estimate $B(s)$ and $\Gamma_\eta(s,t)$ as follows.

Step (2.1). Step (2.1) is to estimate $B(s)$. Let $B_k(s)$ be the kth row of $B(s)$. Then, we calculate the least-squares estimator of $B(s)$, denoted by $\hat{B}(s)$, by minimizing an objective function given by

$$\sum_{i=1}^{n}[\hat{\mathbf{f}}_i(s)^{sec}-B(s)\mathbf{x}_i]^T[\hat{\mathbf{f}}_i(s)^{sec}-B(s)\mathbf{x}_i]. \tag{14}$$

Specifically, the least-squares estimator of $B_k(s)$, denoted by $\hat{B}_k(s)$, is given by

$$\hat{B}_k(s)^T = (\sum_{i=1}^{n}\mathbf{x}_i^{\otimes 2})^{-1}\sum_{i=1}^{n}\mathbf{x}_i\hat{f}_{i,k}(s)^{sec}, \quad \text{for} \quad k = 1,\cdots,m. \tag{15}$$

Step (2.2). Step (2.2) is to estimate $\Gamma_\eta(s,t)$. Let $\hat{\eta}_i(s) = \hat{\mathbf{f}}_i(s)^{sec} - \hat{B}(s)\mathbf{x}_i$. Then, the covariance matrix $\Gamma_\eta(s,t)$ can be estimated by

$$\hat{\Gamma}_\eta(s,t) = (n-m)^{-1}\sum_{i=1}^{n}\hat{\eta}_i(s)\hat{\eta}_i(t)^T. \tag{16}$$

The covariance matrix $\Gamma_\eta(s,t)$ characterizes the variation of $\eta_i(s)$, which is different from $\Gamma(s,t)$.

3. Global Test Statistic

We develop a global test statistic to test linear hypotheses of $B(s)$ in order to answer various scientific questions involving a comparison of fiber bundle diffusion properties along fiber bundles across two (or more) diagnostic groups and the development of fiber bundle diffusion properties along time. We can formulate these questions as linear hypotheses of $B(s)$ as follows:

$$H_0 : \mathbf{C}\text{vec}(B(s)) = \mathbf{b}_0(s) \ \text{ for all } s \ \text{ vs. } \quad H_1 : \mathbf{C}\text{vec}(B(s)) \neq \mathbf{b}_0(s), \tag{17}$$

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

We test the null hypothesis $H_0 : \mathbf{C}\text{vec}(B(s)) = \mathbf{b}_0(s)$ using a global test statistic S_n defined by

$$S_n = n \int_0^{F_0} \mathbf{d}(s)^T [\mathbf{C}(\hat{\Gamma}_\eta(s,s) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1} \mathbf{d}(s) ds, \tag{18}$$

where $\hat{\Omega}_X = n^{-1} \sum_{i=1}^n \mathbf{x}_i^{\otimes 2}$, $\mathbf{d}(s) = \mathbf{C}\text{vec}(\hat{B}(s)) - \mathbf{b}_0(s)$ and F_0 is the whole arc length of a specific fiber bundle. In order to use S_n as a test statistic, we need an asymptotic result. Specifically, similar to the arguments in Zhang and Chen (2007a), we can show that under some conditions and H_0, $\sqrt{n}\mathbf{d}(s)$ and S_n converge weakly to $N(\mathbf{0}, \mathbf{C}(\Gamma_\eta(s,s) \otimes \Omega_X^{-1})\mathbf{C}^T)$ and a weighted χ^2 distribution, respectively, as $n \to \infty$.

In addition, at a given grid point s_j on a specific tract, we can also test the local null hypothesis $H_0(s_j) : \mathbf{C}\text{vec}(B(s_j)) = \mathbf{b}_0(s_j)$ using a local test statistic $S_n(s_j)$ defined by

$$S_n(s_j) = n\mathbf{d}(s_j)^T [\mathbf{C}(\hat{\Gamma}_\eta(s_j,s_j) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1} \mathbf{d}(s_j). \tag{19}$$

Under some conditions and $H_0(s_j)$, $\sqrt{n}\mathbf{d}(s_j)$ and $S_n(s_j)$ converge weakly to $N(\mathbf{0}, \mathbf{C}(\Gamma_\eta(s_j,s_j) \otimes \Omega_X^{-1})\mathbf{C}^T)$ and a weighted χ^2 distribution with r degrees of freedom, respectively, as $n \to \infty$.

4. Resampling Method

We develop a resampling method (or wild bootstrap method) to approximate the p-value of S_n (Zhu et al., 2007; Lin, 2005). The resampling method has four key steps as follows.

Step (3.1): Fit the functional linear model $\hat{\mathbf{f}}_i(s) = B(s)\mathbf{x}_i + \eta_i(s), \quad i = 1, \cdots, n$, under the null hypothesis H_0, which yields $\hat{B}^*(s)$ and $\hat{\eta}_i^*(s) = \hat{\mathbf{f}}_i(s) - \hat{B}^*(s)\mathbf{x}_i$.

Step (3.2): Generate a random sample $\tau_i^{(g)}$ from a $N(0,1)$ generator for $i = 1, \cdots, n$ and then construct $\hat{\mathbf{f}}_i(s)^{(g)} = \hat{B}^*(s)\mathbf{x}_i + \tau_i^{(g)}\hat{\eta}_i^*(s)$. Then, based on $\hat{\mathbf{f}}_i(s)^{(g)}$, we calculate

$$\hat{B}_k(s)^{(g)T} = (\sum_{i=1}^n \mathbf{x}_i^{\otimes 2})^{-1} \sum_{i=1}^n \mathbf{x}_i \hat{\mathbf{f}}_{i,k}(s)^{(g)}, \quad k = 1, \cdots, m, \tag{20}$$

where $\hat{B}_k(s)^{(g)T}$ and $\hat{\mathbf{f}}_{i,k}(s)^{(g)}$ are, respectively, the kth row of $\hat{B}(s)^{(g)}$ and $\hat{\mathbf{f}}_i(s)^{(g)}$. Finally, let $\mathbf{d}(s)^{(g)} = \mathbf{C}\text{vec}(\hat{B}(s)^{(g)}) - \mathbf{b}_0(s)$, we compute

$$\begin{aligned} S_n^{(g)} &= n \int_0^{F_0} \mathbf{d}(s)^{(g)T} [\mathbf{C}(\hat{\Gamma}_\eta(s,s) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1} \mathbf{d}(s)^{(g)} ds, \\ S_n(s_j)^{(g)} &= n\mathbf{d}(s_j)^{(g)T} [\mathbf{C}(\hat{\Gamma}_\eta(s_j,s_j) \otimes \hat{\Omega}_X^{-1})\mathbf{C}^T]^{-1} \mathbf{d}(s_j)^{(g)} \ \text{ for } \ j = 1, \cdots, L_0. \end{aligned} \tag{21}$$

Step (3.3): Repeat Step (3.2) G times to obtain $\{S^{(g)}_{n,\max} = \max_{1\le j\le L_0} S_n(s_j)^{(g)} : g = 1, \cdots, G\}$ and calculate

$$p(s_j) = G^{-1}\sum_{g=1}^{G} 1(S^{(g)}_{n,\max} \ge S_n(s_j))$$

for each s_j. The $p(s_j)$ is the corrected p-value at the location s_j.

Step (3.4): Repeat Step (3.2) G times to obtain $\{S^{(g)}_n : g = 1, \cdots, G\}$ and calculate

$$p = G^{-1}\sum_{g=1}^{G} 1(S^{(g)}_n \ge S_n).$$

If p is smaller than a pre-specified value α, say 0.05, then we reject the null hypothesis H_0.

2 Matlab functions

We implemented the FRACTS pipeline in Matlab. The following is the description of the functions in FRACTS Matlab tool. We first give an overview of the Matlab function and then explain each of the function in terms of function name, input, output, the function goal, and remarks if desired. Examples and results will be given in the next section.

2.1 Function overview

fiberSTATHT1_read: read raw data and generate, arc length, standardized design and response matrices, and related dimension parameters.

fiberSTATHT2_MVreg: read arc length, standardized design and response matrices, related dimension parameters, and design matrices of a hypothesis testing; using wald-test and mvregress function to generate the test statistics, p-values, and some other statistics by pooling data from all grid points.

myFDR: read a vector of p-values and False Discovery Rate level; and generate p-value threshold based on independence or positive dependence and Nonparametric p-value threshold.

fiberSTATHT3_1_Local: it is to implement Zhang and Chen (2007b) method to each of the fiber bundle diffusion properties.

chi2D: it is to implement Zhang and Chen (2007b) method to hypothesis tests about functional data.

fdaflm: it is to implement Zhang and Chen (2007b) method to fit a functional linear model using local polynomial kernel (LPK) with plot options and other parameter estimations.

fdafit: it is to implement Zhang and Chen (2007b) method to fit a functional linear model using local polynomial kernel (LPK) with plot options.

lpsfit: it is to implement a local polynomial kernel (LPK) fit.

fiberSTATHT3_LocalLM: it is to implement Zhang and Chen (2007b) method to each of the fiber bundle diffusion properties and generate various p values.

fiberSTATHT3_ZHCHann: it is to implement Zhang and Chen (2007b) method to each of the fiber bundle diffusion properties and generate test statistics and its corresponding p value.

fiberSTATHT4_LocalLM: it is to implement Zhu et al. (2010) method to all the fiber bundle diffusion properties and generate the estimation of functional coefficients, global test statistics and their p values.

fiberSTATHT5_multiLPSfit: it is to implement Zhu et al. (2010) method **step (1.3)** in nonparametric model and generate fitted curves and GCV value.

fiberSTATHT6_LocalLM: it is to implement Zhu et al. (2010) method to all the fiber bundle diffusion properties and generate the estimation of functional coefficients, global test statistics and their p values given smoothed curves.

fiberSTATHT7plot: it is to implement Zhu et al. (2010) method to plot related graphs.

2.2 Function description

fiberSTATHT1_read

Function [NoSetup, arclength, Xdesign, Ydesign, scalediffusion]=fiberSTATHT1_read(tractdata, designdata, diffusionFiles, nofeatures, featuresname)

Input `tractData:` the text file containing (x, y, z) coordinates of all locations on a given fiber tract. The data set should start from one end to the other end.
tractData is a $L_0 \times 3$ matrix, where L_0 denotes the number of locations. 3 denotes the three coordinates.

`designData:` the text file containing covariates of interest. Please always include the intercept in the first column. designData is a $n \times p$ matrix, where n denotes the number of subjects and p denotes the number of covariates.

`diffusionFiles:` a $m \times 1$ cell containing the names of all fiber diffusion properties files. Each fiber bundle diffusion properties should contain a $L_0 \times n$ matrix. Rows correspond to the columns in tractData, while columns correspond to the columns in designData.

`nofeatures:` the number of diffusion properties, denoted by m.

`featurenames:` a $m \times 1$ cell of property names.

Output `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`arclength:` a $L_0 \times 1$ column vector of the arclength from one end to the other end.

`Xdesign:` a $n \times p$ normalized design matrix.

`Ydesign:` a $n \times L0 \times m$ matrix.

`scalediffusion:` a $m \times 1$ vector of scales for each properties.

Remark to avoid unnecessary errors, please use this function to preprocess the raw data before you go to the nonparametric model.

fiberSTATHT2_MVreg

Function [rawpvalue, pvalue, waldtest, CorrFiber, Cpvalue, Npvalue]=fiberSTATHT2_MVreg(NoSetup, arclength, Xdesign, Ydesign, Cdesign, B0vector)

Input `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`arclength:` a $L_0 \times 1$ column vector of the arclength from one end to the other end.

`Xdesign:` a $n \times p$ normalized design matrix.

`Ydesign:` a $n \times L0 \times m$ matrix.

`Cdesign:` a $k \times mp$ matrix for characterizing the k linear constraints among mp parameters.

`B0vector:` a $k \times 1$ vector for hypothesis testing.

Output `rawpvalue:` a $L0 \times 1$ vector of uncorrect p values.

`pvalue:` a $L0 \times 1$ vector of sorted uncorrect p values.

`waldtest:` a $L0 \times m * (m-1)/2$ matrix of correlations.

`CorrFiber:` a $L0 \times 1$ vector of Wald test values.

`Cpvalue:` a threshold for FDR.

`Npvalue:` a threshold method nonparametric method.

Remark You need myFDR.m for correcting for p values using False Discovery Rate.

myFDR

Function [pID, pN]=myFDR(p,q)

Input `p:` a vector of p-values.

`q:` False Discovery Rate level.

Output `pID:` p-value threshold based on independence or positive dependence.

`pN:` Nonparametric p-value threshold.

Remark Based on FDR.m (1.4 Tom Nichols 02/07/02).

fiberSTATHT3_1_Local:

Function [pstat, efitBeta, efitYdata]=fiberSTATHT3_1_Local(NoSetup, arclength, Xdesign, Ydesign, Cdesign, B0vector)

Input `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`arclength:` a $L_0 \times 1$ column vector of the arclength from one end to the other end.

`Xdesign:` a $n \times p$ normalized design matrix.

`Ydesign:` a $n \times L0 \times m$ matrix.

`Cdesign:` a $k \times mp$ matrix for characterizing the k linear constraints among mp parameters.

`B0vector:` a $k \times L0$ vector for hypothesis testing.

Output `pstat:` a $m \times 1$ vector of p values for whole curve.

`efitBeta:` a $m \times L_0 \times p$ matrix of estimators.

`efitYdata:` a $m \times L_0 \times$ matrix of curves.

Remark You need myFDR.m for correcting for p values using False Discovery Rate and fdaflm.m for functional linear regression using local polynomial kernel.

chi2D

Function [pstat, pdf]=chi2D(stat, dd, params, q)

Input `stat:` stat=$sum_{r=1}^{m} d_r A_r$, $A_r \sim \chi^2_q$.

`dd:` dd=[d1,d2,...,dm].

`params:` params=[method,N,indfig]; method=0 Chisq approximation and 1 Simulation; indfig=1, plot the null density and 0 otherwise.

`q:` degrees of freedom of A_r, assume the same df for all A_r.

Output `pstat:` pstat=[stat,pvalue,df,M,Delta,alpha,beta].

`pdf:` density function.

Remark code wrote by Jin-Ting Zhang and need gpkde.m and related functions (Marron, 1996).

fdaflm

Function [efit, vfit, yfit, hgcv, vhgcv, Sig2]=fdaflm(data, params, xfit, labstr)

Input `data:` data=[subj, time, y, x]; a $N \times (3+p)$ matrix with p=# of the covariates.

`params:` params=[h, indfig, korder]; h-bandwidth; indfig=0 no plot and 1 plot; korder is the order of the polynomial.

`xfit:` a design matrix at which response is to be fitted; can be empty.

`labstr:` labels for x and y.

Output `efit:` efit=[xfit, eta, esig]; eta is mean curve and esig is its standard deviation.

`vfit:` vfit=[vfit1,vfit2,...,vfitn]; vfit is the error for each curve.

`yfit:` yfit=[f1,f2,...,fn] is the estimates of each curve.

`hgcv:` hgcv=[h,gcv] are the bandwidth candidates and their GCVs.

`vhgcv:` vhgcv=[vh,gcv] are the optimal bandwidth and its GCV.

`Sig2:` variance of yfit.

Remark code wrote by Jin-Ting Zhang and you need fdafit.m function to estimate parameters.

fdafit

Function [efit, yfit, hgcv, vhgcv]=fdafit(data, params, xfit, labstr)

Input `data:` data=[subj, time, y, x]; a $N \times (3+p)$ matrix with p=# of the covariates.

`params:` params=[h, indfig, korder]; h-bandwidth; indfig=0 no plot and 1 plot; korder is the order of the polynomial.

`xfit:` a design matrix at which response is to be fitted; can be empty.

`labstr:` labels for x and y.

Output `efit:` efit=[xfit, eta, esig]; eta is mean curve and esig is its standard deviation.

`yfit:` yfit=[f1,f2,...,fn] is the estimates of each curve.

`hgcv:` hgcv=[h,gcv] are the bandwidth candidates and their GCVs.

`vhgcv:` vhgcv=[vh,gcv] are the optimal bandwidth and its GCV.

Remark code wrote by Jin-Ting Zhang and you need lpsfit.m function to estimate parameters.

lpsfit

Function [fits, hgcv]=lpsfit(data, params, xfit, kstr)

Input `data:` data=[subj, time, y, x]; a $N \times (3+p)$ matrix with p=# of the covariates.

`params:` params=[h, indfig, korder]; h-bandwidth; indfig=0 no plot and 1 plot; korder is the order of the polynomial.

`xfit:` a design matrix at which response is to be fitted; can be empty.

`kstr:` kernel function.

Output `fits:` fits=[xfit, yfit, ysig]; yfit=fitted values at xfit.

`hgcv:` hgcv=[h,gcv] are the optimal bandwidth and its GCV.

Remark code wrote by Jin-Ting Zhang.

fiberSTATHT3_LocalLM:

Function [pvalue, Cpvalue, Npvalue]=fiberSTATHT3_LocalLM(NoSetup, arclength, Xdesign, Ydesign, Cdesign, B0matrix)

Input `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`arclength:` a $L_0 \times 1$ column vector of the arclength from one end to the other end.

`Xdesign:` a $n \times p$ normalized design matrix.

`Ydesign:` a $n \times L0 \times m$ matrix.

`Cdesign:` a $k \times mp$ matrix for characterizing the k linear constraints among mp parameters.

`B0vector:` a $k \times L0$ vector for hypothesis testing.

Output `pvalue:` a $L0 \times 1$ vector of sorted uncorrect p values.

`Cpvalue:` a threshold for FDR.

`Npvalue:` a threshold method nonparametric method.

Remark You need myFDR.m for correcting for p values using False Discovery Rate and fdaflm.m for functional linear regression using local polynomial kernel.

fiberSTATHT3_ZHCHann:

Function [pstat]=fiberSTATHT3_ZHCHann(NoSetup, arclength, Xdesign, Ydesign, Cdesign, B0matrix)

Input `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`arclength:` a $L_0 \times 1$ column vector of the arclength from one end to the other end.

`Xdesign:` a $n \times p$ normalized design matrix.

`Ydesign:` a $n \times L0 \times m$ matrix.

`Cdesign:` a $k \times mp$ matrix for characterizing the k linear constraints among mp parameters.

`B0vector:` a $k \times L0$ vector for hypothesis testing.

Output `pstat:` a vector of test statistic and associated p-value.

Remark You need myFDR.m for correcting for p values using False Discovery Rate and fdaflm.m for functional linear regression using local polynomial kernel.

fiberSTATHT4_LocalLM:

Function [eta, pstat, GTstat]=fiberSTATHT4_LocalLM(NoSetup, arclength, Xdesign, Ydesign, Cdesign, B0matrix)

Input `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`arclength:` a $L_0 \times 1$ column vector of the arclength from one end to the other end.

`Xdesign:` a $n \times p$ normalized design matrix.

`Ydesign:` a $n \times L0 \times m$ matrix.

`Cdesign:` a $k \times mp$ matrix for characterizing the k linear constraints among mp parameters.

`B0vector:` a $k \times L0$ vector for hypothesis testing.

Output `eta:` a $m \times L0 \times p$ array of estimated functional coefficients.

`pstat:` a vector of test statistic and associated p value.

`GTstat:` a vector of simulated global statistics.

Remark You need fiberSTATHT5_multiLPSfit.m for nonparametric fitting and fiberSTATHT6_LocalLM.m for resampling method.

fiberSTATHT5_multiLPSfit:

Function [FitYdesign, GCVnum]=fiberSTATHT5_multiLPSfit(NoSetup, Kmat, IcovYfit, Ydesign, Zxcell)

Input `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`Kmat:` a $L_0 \times L_0$ matrix related to kernel function.

`IcovYfit:` a $n \times p$ normalized design matrix.

`Ydesign:` a $n \times L0 \times m$ matrix.

`Zxcell:` a L_0 vector of cells each Zxcelli is an array of $L0 \times m \times 2m$ related to arc length.

Output `eta:` a $m \times L0 \times p$ array of estimated functional coefficients.

`GTstat0:` a vector of global statistics.

Remark this function is to implement **step (1.3)** in nonparametric model.

fiberSTATHT6_LocalLM:

Function [eta, GTstat0]=fiberSTATHT6_FLM(NoSetup, arclength, Xdesign, FitYdesign, Cdesign, B0matrix, FigYes)

Input `NoSetup:` a column vector of $[n, L_0, p, m]$, where n is sample size, L_0 is the number of grid points, $p-1$ is the number of covariates and m is the number of features.

`arclength:` a $L_0 \times 1$ column vector of the arclength from one end to the other end.

`Xdesign:` a $n \times p$ normalized design matrix.

`FitYdesign:` a $n \times L0 \times m$ matrix.

`Cdesign:` a $k \times mp$ matrix for characterizing the k linear constraints among mp parameters.

`B0vector:` a $k \times L0$ vector for hypothesis testing.

`FigYes:` a scalar representing ¿0 for plot figure.

Output `eta:` a $m \times L0 \times p$ array of estimated functional coefficients.

`pstat:` a vector of test statistic and associated p value.

`GTstat:` a vector of simulated global statistics.

Remark this function is used for resampling method to simulate the p value.

3 Example

3.1 Data set and Model

We applied FRATS to the joint analysis of FA and MD values along the splenium tract as follows (Figure 2). We fitted the functional linear model (13) to these smoothed FA and MD functions from all 128 subjects, in which $\mathbf{x}_i = (1, \mathrm{g}_i, \mathrm{Gage}_i, \mathrm{age}_i)^T$ and $m = 2$, that is

$$(\mathrm{FA}_i(s_j), \mathrm{MD}_i(s_j))^T = (f_{i,1}(s_j), f_{i,2}(s_j))^T + \epsilon_{i,j}, \tag{22}$$

$$f_{i,1}(s) = \beta_{11}(s) + \beta_{12}(s) \times \mathrm{g}_i + \beta_{13}(s) \times \mathrm{Gage}_i + \beta_{14}(s) \times \mathrm{age}_i + \eta_{i1}(s),$$

$$f_{i,2}(s) = \beta_{21}(s) + \beta_{22}(s) \times \mathrm{g}_i + \beta_{23}(s) \times \mathrm{Gage}_i + \beta_{24}(s) \times \mathrm{age}_i + \eta_{i2}(s), \tag{23}$$

where $\eta_i(s) = (\eta_{i1}(s), \eta_{i2}(s))^T$ is a 2×1 vector of Gaussian process with zero mean and covariance matrix $\Gamma_\eta(s,t)$ and $\epsilon_{i,j}$ is a 2×1 vector of Gaussian random variables with zero mean and covariance matrix $\Sigma(s_j)$. Then we used equation (15) to estimate the function of regression coefficient vector $\hat{B}(s)$. Secondly, we constructed the global test statistic S_n to test the effects of all the three covariates for FA alone, MD alone, and joint FA and MD, respectively, and performed hypothesis testing on the whole splenium tract. For example, to test the effect of gender we have design matrices

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{b}_0(s) \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{for all} \quad s.$$

The p-value of S_n was approximated using the resampling method with $G = 10,000$. For more examples on simulation and real data, see Zhu et al. (2010).

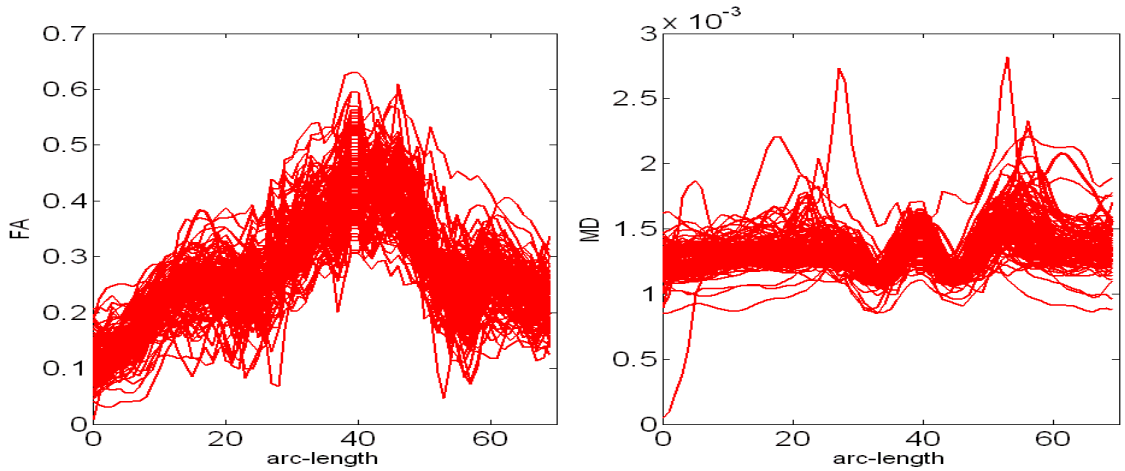

Figure 2: FA and MD along the splenium tract

3.2 Code and results

There are three data sets we need to import, namely, `tractdata, designdata, diffusionFiles`. The data set `tractdata` contains (x, y, z) coordinates of all locations on a given fiber tract. The data set should start from one end to the other end. `tractData` is a $L_0 \times 3$ matrix, where $L_0 = 22$ denotes the number of locations. 3 denotes the three coordinates. The following shows the first 4 rows of `tractData`,

```
-21 0 0
-20 0 0
-19 0 0
-18 0 0
...
```

the data set `designdata` contains covariates of interest. We always need to 1include the intercept in the first column. `designData` is a $n \times p$ matrix, where n denotes the number of subjects and $p = 4$ denotes the number of covariates, as we have intercept, gender, age and gage. Each covariate is listed in one column. The following shows the first 4 rows, where the columns are, respectively, intercept, gender, flu and age.

```
1 1 276 375
1 0 283 381
1 1 275 378
1 1 280 366
...
```

The data set `diffusionFiles` is a $m(= 2) \times 1$ cell containing the names of all fiber diffusion properties files. Each fiber bundle diffusion properties should contain a $L_0 \times n$ matrix. Rows correspond to the columns in `tractData`, while columns correspond to the columns in `designData`. In particular, we use

```
diffusionFiles=cell(2,1);
```

to define the cell structure. We then specify the first cell `diffusionFiles{1}` as FA values

and the second `diffusionFiles{2}` as MD values. Both are $n \times L_0$ matrices. The following are the first 4 rows of `diffusionFiles{1}`.

```
0.3286 0.2782 0.2172 0.2095 0.2939 0.1450 ···
0.2342 0.2104 0.2800 0.2621 0.2256 0.2578 ···
0.2954 0.2066 0.2514 0.3040 0.2378 0.2897 ···
0.3411 0.2922 0.3336 0.2956 0.2952 0.3268 ···
...
```

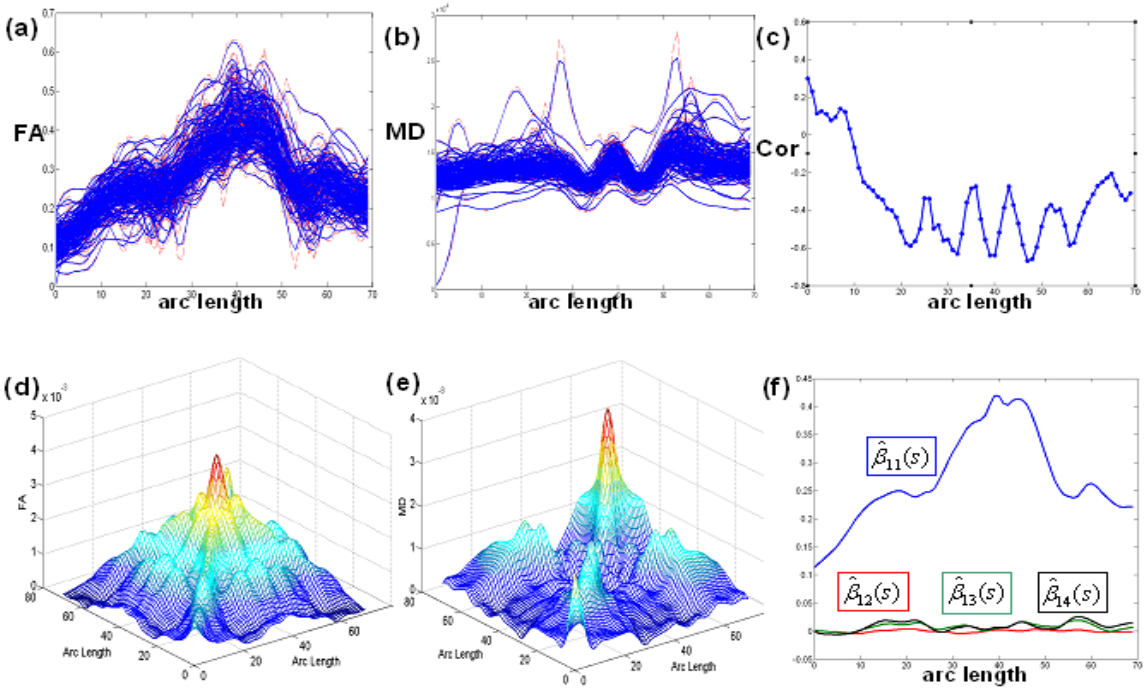

Figure 3: Results from the analysis of FA and MD on the splenium tract: reconstructed curves $\hat{\mathbf{f}}_i(s)^{sec}$ for FA in panel (a) and MD in panel (b); (c) estimated correlation between FA and MD along the tract; estimated covariance matrices $\hat{\Gamma}(s,t)$ for FA in panel (d) and MD in panel (e); (f) estimated regression coefficient functions for FA: $\hat{\beta}_{11}(s)$ for intercept (blue), $\hat{\beta}_{12}(s)$ for gender (red), $\hat{\beta}_{13}(s)$ for gestational age (green), and $\hat{\beta}_{14}(s)$ for age (black).

After load covariates, response and arc length data, we use fiberSTATHT1_read to transfer data into the format we want.

```
[NoSetup, arclength, Xdesign, Ydesign]
      =fiberSTATHT1_read(tractdata, designdata, diffusionFiles, nofeatures).
```

We then use function fiberSTATHT4_LocalLM to estimate the parameters, see Figure 3.2.

```
[eta, pstat, GTstat]
      =fiberSTATHT4_LocalLM(NoSetup, arclength, Xdesign, Ydesign, Cdesign,
```

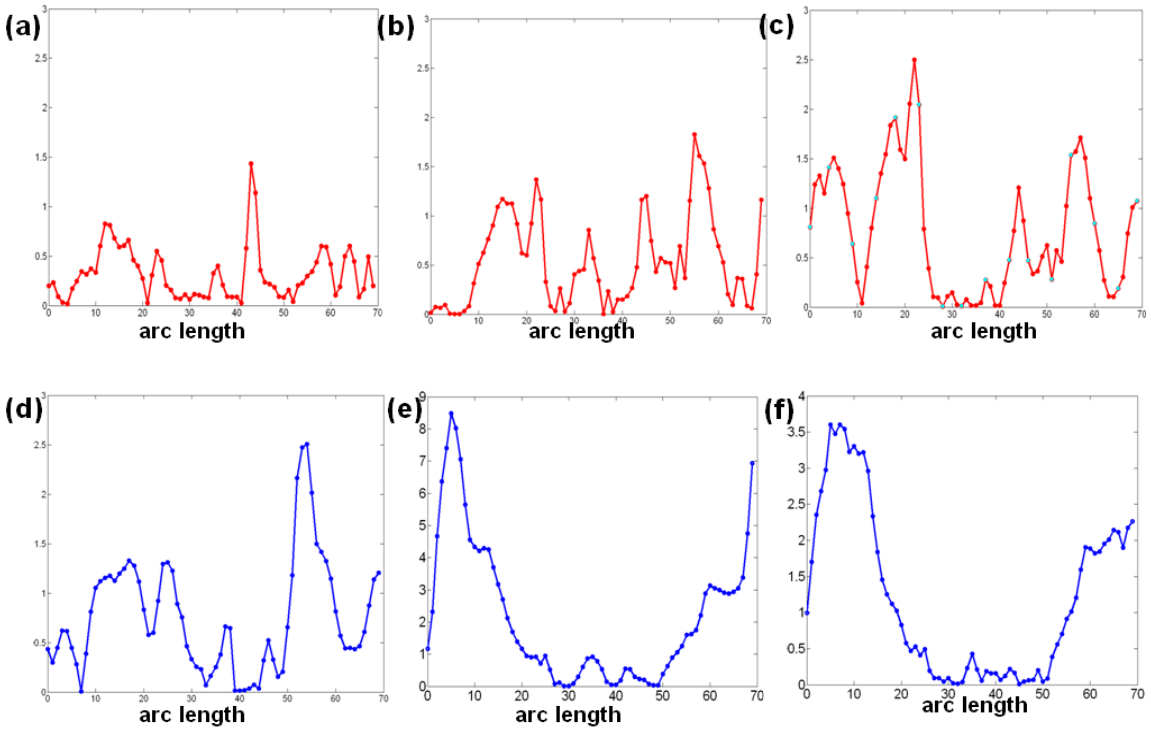

Figure 4: Results from the analysis of FA and MD on the splenium tract: the $-\log_{10}(p)$ values of test statistics $S_n(s_j)$ for testing gender effect in panel (a), gestational age effect in panel (b), and age effect in panel (c) on FA; the $-\log_{10}(p)$ values of test statistics $S_n(s_j)$ for testing gender effect in panel (d), gestational age effect in panel (e), and age effect in panel (f) on MD.

`B0matrix).`

We can also find the local p-values (see Figure 3.2.) by

`Lpvals=1-chi2cdf(pstat,m).`

Function fiberSTATHT7plot provides some useful example of plotting graphs.

4 FRATS: graphical user interface (GUI)

To make it easily accessible, we developed a Graphical User Interface (GUI) to pack the code. As shown in Figure 5, there are 4 button groups, which are supposed to be executed in order. The 4 groups are s Load Raw Data, *Basic Plots*, *Load Test Data*, and *P-value Plots*. There are 3 raw data sets, namely, tract data, design data and diffusion data. The test data sets include test design matrix and null hypothesis vector. All data sets must be in `.mat`. The package includes a sample matlab code `pre_address_data.m` on how to set up data. After loading all raw data, GUI will transfer the raw data and estimate the coefficients. Then you can plot the raw tract data or the coefficient functions by pushing the corresponding buttons.

If you want to do a test, you need to load the test design data. There are two types of test. One is to test individually and the other one is test all the diffusion properties together. Once you loaded the test design data, GUI will display what test type you requested. The test calculation may take a while. After matlab finishes the computation, GUI will report the global test statistics and p-values. You also have the option to plot the local p-values.

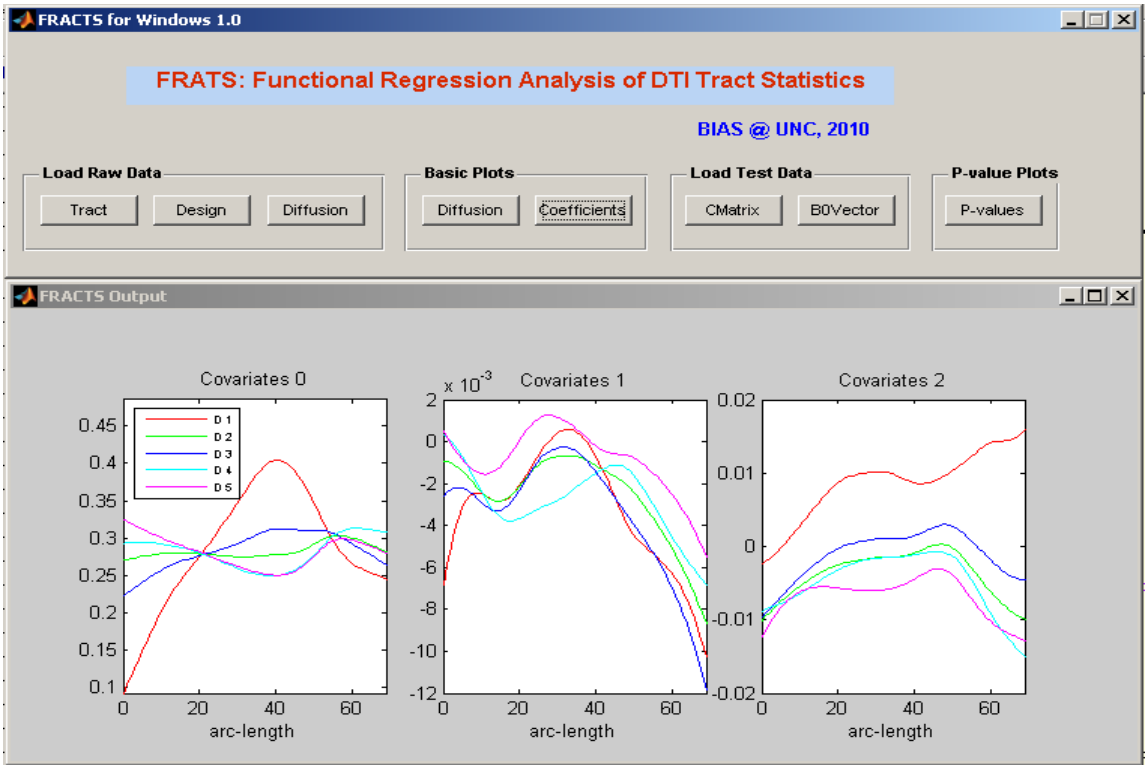

Figure 5: FRATS GUI

References

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometry: the methods. *Neuroimage*, 11:805–821.

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. *Biophysical Journal*, 66:259–267.

Bonekam, D., Nagae, L. M., Degaonkar, M., Matson, M., Abdalla, W. M., Barker, P. B., Mori, S., and Horská, A. (2008). Diffusion tensor imaging in children and adolescents: Reproducibility, hemispheric, and age-related differences. *NeuroImage*, 34:733–742.

Camara, E., Bodammer, N., Rodriguez-Fornells, A., and Tempelmann, C. (2007). Age-related water diffusion changes in human brain: A voxel-based approach. *NeuroImage*, 34:1588–1599.

Chen, Y. S., An, H. Y., Zhu, H. T., Stone, T., Smith, J. K., Hall, C., Bullitt, E., Shen, D. G., and Lin, W. L. (2009). White matter abnormalities revealed by diffusion tensor imaging in non-demented and demented hiv+ patients. *NeuroImage*, 47:1154–1162.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

Focke, N. K., Yogarajah, M., Bonelli, S. B., Bartlett, P. A., Symms, M. R., and Duncan, J. S. (2008). Voxel-based diffusion tensor imaging in patients with mesial temporal lobe epilepsy and hippocampal sclerosis. *NeuroImage*, 40:728–737.

Gilmore, J. H., Smith, L. C., Wolfe, H. M., Hertzberg, B. S., Smith, J. K., Chescheir, N. C., Evans, D. D., Kang, C., Hamer, R. M., Lin, W., and Gerig, G. (2008). Prenatal mild ventriculomegaly predicts abnormal development of the neonatal brain. *Biol Psychiatry*, 64:1069–1076.

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.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, 34(3):1493–1517.

Hecke, W. V., Sijbers, J., Backer, S. D., Poot, D., Parizel, P. M., and Leemans, A. (2009). On the construction of a ground truth framework for evaluating voxel-based diffusion tensor mri analysis methods. *NeuroImage*, 46:692–707.

Jones, D. K., Symms, M. R., Cercignani, M., and Howard, R. J. (2005). The effect of filter size on vbm analyses of dt-mri data. *NeuroImage*, 26:546–554.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, page in press.

Lin, D. Y. (2005). An efficient monte carlo approach to assessing statistical significance in genomic studies. *Bioinformatics*, 6:781–787.

Marron, J. S. (1996). A personal view of smoothing and statistics,. *in: W. Hürdle, M. Schimek (Eds.), Statistical Theory and Computational Aspects of Smoothing*, pages 1–9.

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

Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis*. Springer Series in Statistics. Springer-Verlag, New York. Methods and case studies.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer-Verlag, New York.

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

Snook, L., Paulson, L. A., Roy, D., Phillips, L., and Beaulieu, C. (2005). Diffusion tensor imaging of neurodevelopment in children and young adults. *NeuroImage*, 26:1164–1173.

Snook, L., Plewes, C., and Beaulieu, C. (2007). Voxel based versus region of interest analysis in diffusion tensor imaging of neurodevelopment. *NeuroImage*, 34:243–252.

Welsh, A. H. and Yee, T. W. (2006). Local regression for vector responses. *Journal of Statistical Planning and Inference*, 136:3007–3031.

Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(1):3–25.

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

Zhang, J. and Chen, J. (2007a). Statistical inference for functional data. *The Annals of Statistics*, 35:1052–1079.

Zhang, J. and Chen, J. (2007b). Statistical inferences for functional data. *Annals of Statistics*, 35:1052–1079.

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

Zhu, H. T., Styner, M., Tang, N. S., Liu, Z. X., Lin, W. L., and Gilmore, J. (2010). Frats: Functional regression analysis of dti tract statistics. *IEEE Transactions on Medical Imaging*, 29:1039–1049.