Statistical Challenges for Neuroimaging Data Analysis
Lecture 3. Functional Data Analysis

WNAR 2018 @ Edmonton
Linglong Kong
lkong@ualberta.ca

UNIVERSITY OF ALBERTA

"uplifting the whole people"
— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Outline

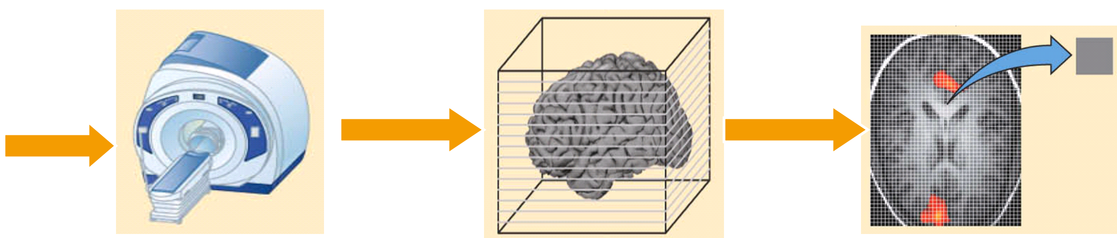
- Big Neuroimaging Data
- Several Key Features
- Varying Coefficient Models
- Multiscale Adaptive Regression Models
- Scalar-on-Image Models

Big Neuroimaging Data

Is it really big data or pig data?

Big Neuroimaging Data

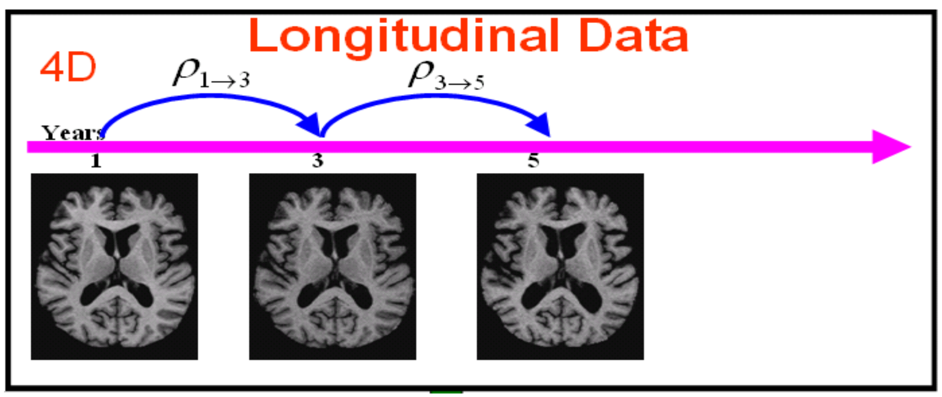
NIH normal brain development
1000 Functional Connectome Project
Alzheimer's Disease Neuroimaging Initiative
National Database for Autism Research (NDAR)
Human Connectome Project

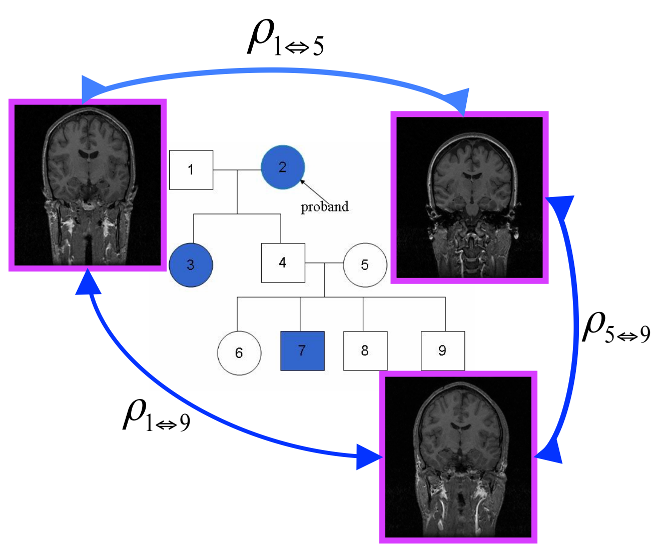


www.guysandstthomas.nhs.uk/.../T/Twins400.jpg

Complex Study Design

cross-sectional studies;
clustered studies including
longitudinal and twin/familial studies;





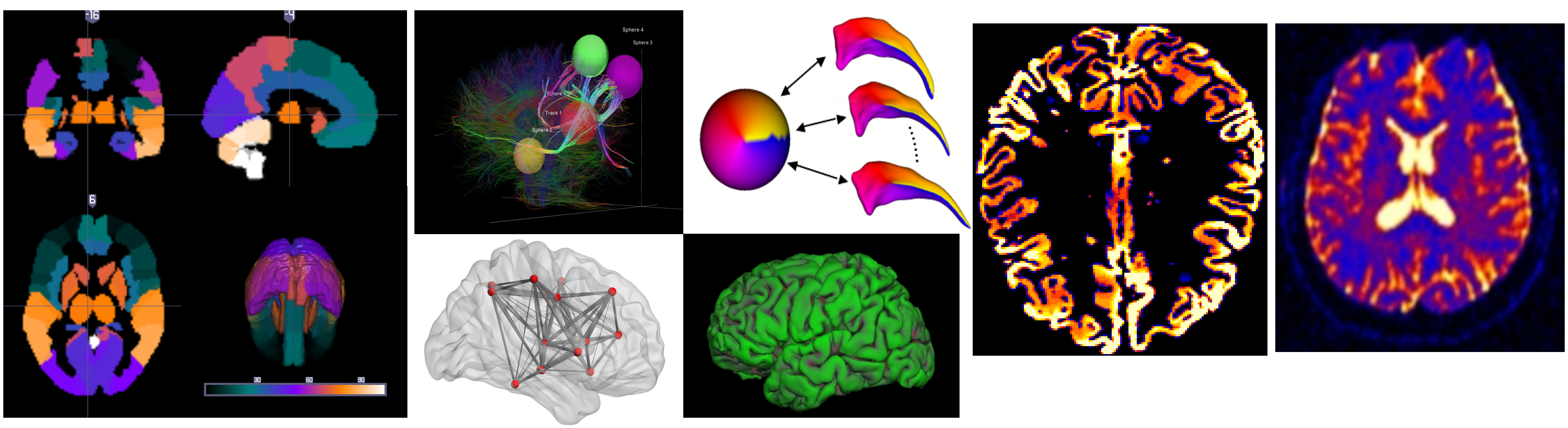
Complex Data Structure

Multivariate Imaging Measures

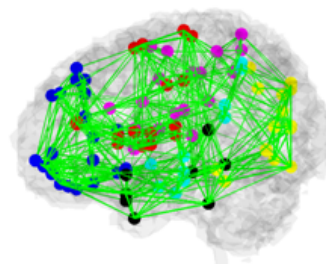
Smooth Functional Imaging Measures

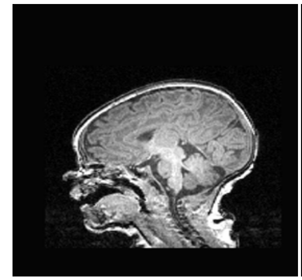
Whole-brain Imaging Measures

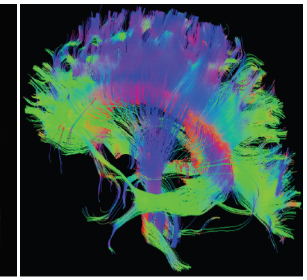
4D-Time Series Imaging Measures

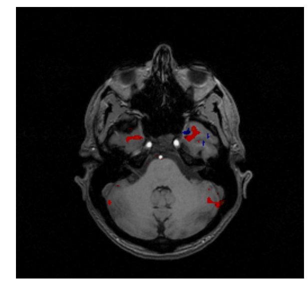


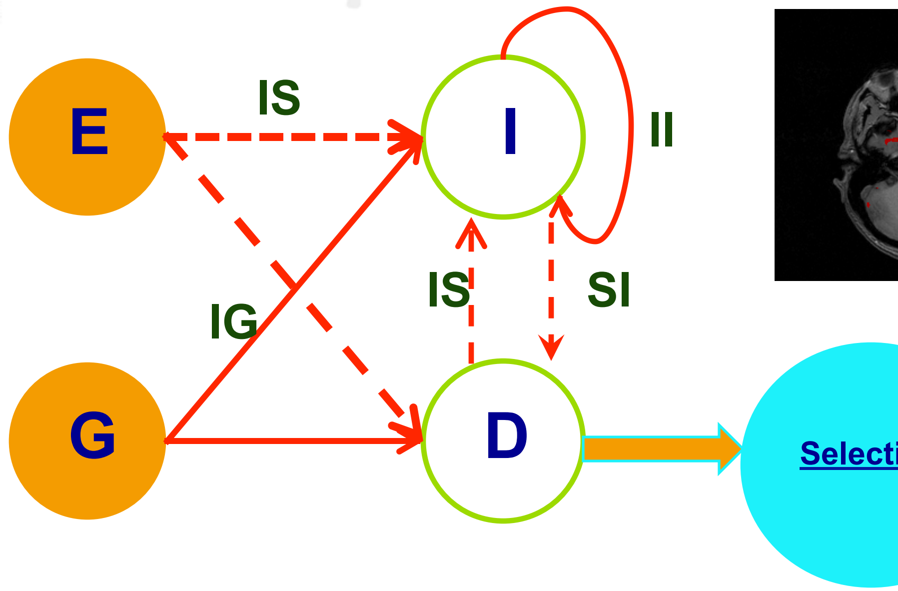
Big Data Integration

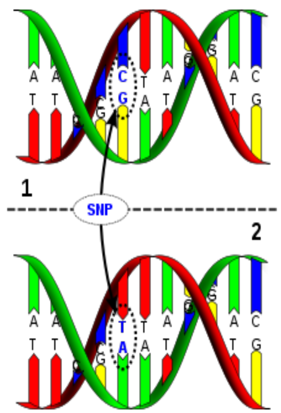












http://en.wikipedia.org/wiki/DNA_sequence

Models for Big Data Integration

Image-on-Scalar (IS) model

Image data as response, clinical variables as predictors.

Scalar-on-Image (SI) model

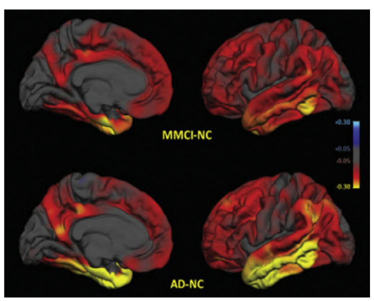
Clinical variables as response, image data as predictors

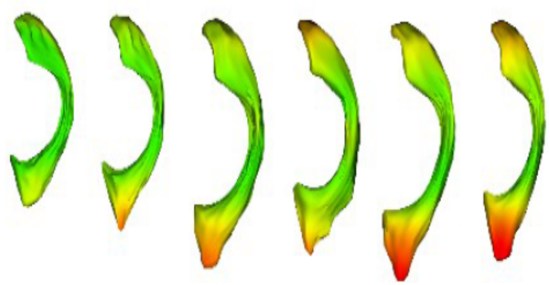
Image-on-Genetic (IG) model

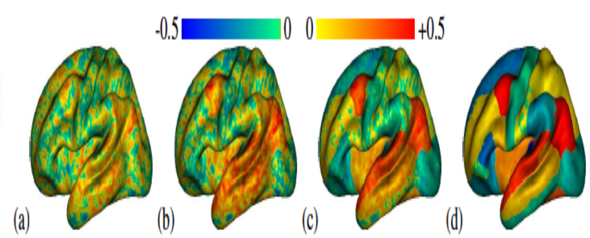
Image data as response, genetic data as predictors

Image-on-Image (II) model

Image data as response, image data as predictors





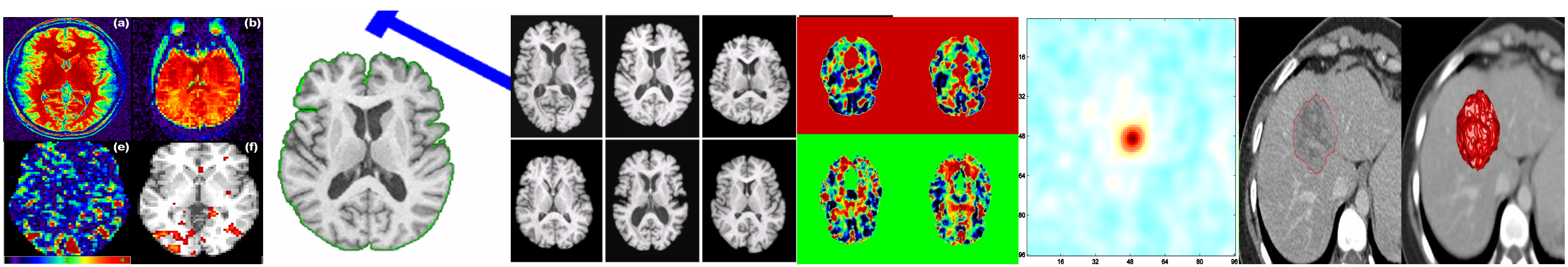


Imaging / Genetics	Candidate ROI	Many ROI	Voxelwise
Candidate SNP	Imager	Imager	Imager
Candidate Gene	Geneticist		
Genome-wide SNP	Geneticist		
Genome-wide Gene	Geneticist		

Noisy Imaging Data

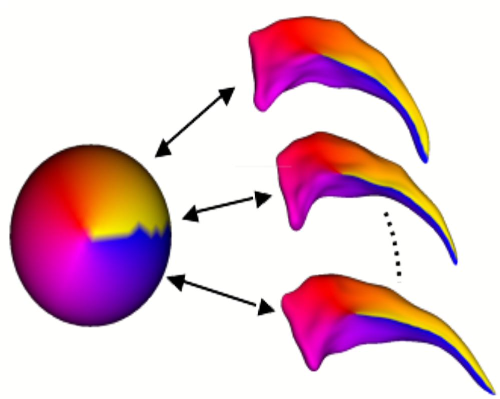
Key Features

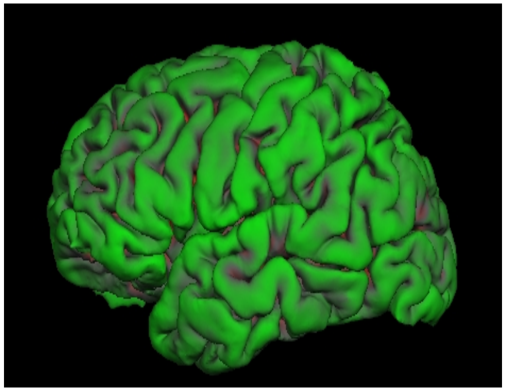
- Infinite Dimension
- Spatial Smoothness
- Spatial Correlation
- Spatial Heterogeneity

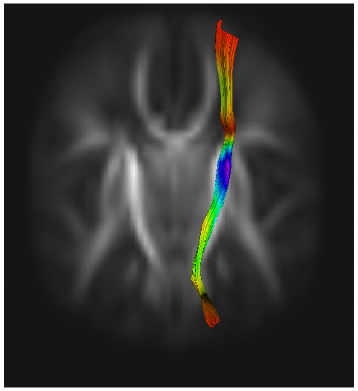


Varying Coefficient Models

Smoothed Functional Data

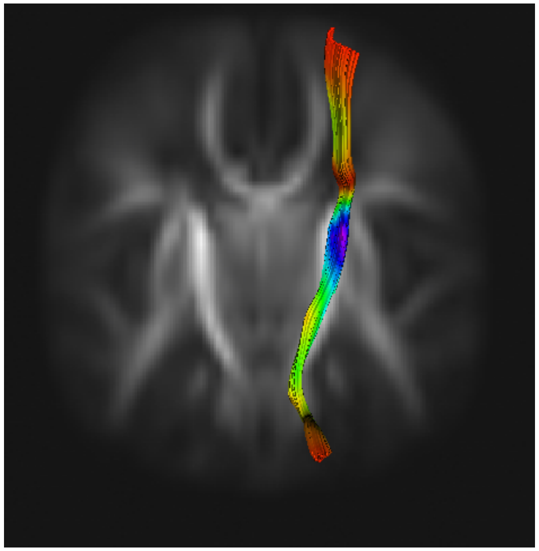






Covariates (e.g., age, gender, diagnostic)

DTI Fiber Tract Data



Data

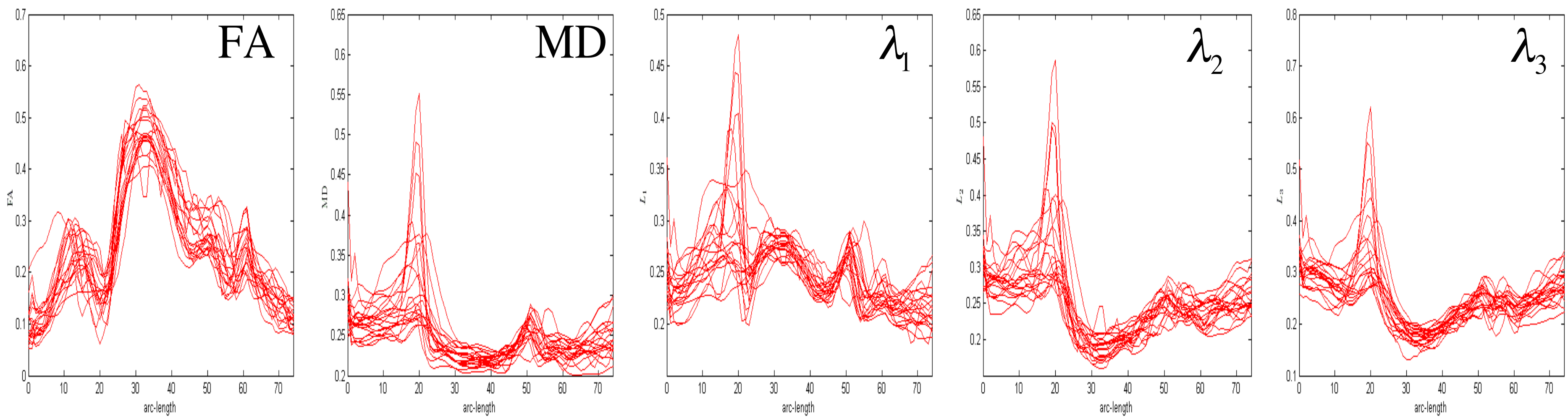
- **Diffusion properties (e.g., FA, RA)**

$$Y_i(s_j) = (y_{i,1}(s_j), \cdots, y_{i,m}(s_j))^T$$

- **Grids** $\{s_1, \cdots, s_{n_G}\}$

- **Covariates (e.g., age, gender, diagnostic)**

$$x_1, \cdots, x_n$$



MVCM

Decomposition:

$$y_{i,k}(s) = x_i^T B_k(s) + \eta_{i,k}(s) + \varepsilon_{i,k}(s)$$

Coefficients $x_1, \cdots, x_n$

Long-range Correlation $\eta_{i,k}(\bullet) \sim SP(0, \Sigma_\eta)$

Short-range Correlation $\varepsilon_{i,k}(\bullet) \sim SP(0, \Sigma_\varepsilon),$

Covariance operator: $\Sigma_y(s,s') = \Sigma_\eta(s,s') + \Sigma_\varepsilon(s,s')$

$$\sqrt{n}\{\mathrm{vec}(\hat{B}(d) - B(d) - 0.5O(H^2)) : d \in D\} \xrightarrow{L} G(0, \Sigma_B(d,d'))$$

Zhu, Li, and Kong (2012). AOS

MVCM

$$\min_{B_k(s)} \sum_{i=1}^{n} \sum_{j=1}^{n_G} K_h(s-s_j)[y_{i,k}(s_j) - x_i^T B_k(s_j)]^2$$

$$\sqrt{n}\{\mathrm{vec}(\hat{B}(s) - B(s) - 0.5O(H^2)) : s \in [0, L_0]\} \xrightarrow{L} G(0, \Sigma_\eta(s,s') \otimes \Omega_X^{-1})$$

Key Advantage $\Rightarrow$ **Low Frequency Signal**

MVCM

Smooth individual functions

$$\min_{\eta_{i,k}(s)} \sum_{j=1}^{n_G} K_h(s_j - s)[y_{i,k}(s_j) - x_i^T \hat{B}_k(s_j) - \eta_{i,k}(s_j)]^2$$

Estimated covariance operator

$$\hat{\Sigma}(s,t) = \sum_{i=1}^{n} \hat{\eta}_i(s)\hat{\eta}_i(t)^T$$

Estimated eigenfunctions

$$\{(\hat{\lambda}_{k,l}, \hat{\psi}_{k,l}(s)) : l = 1, \cdots, \infty\}$$

Functional Principal Component Analysis

MVCM

Testing Linear Hypotheses

$$H_0 : C\text{vec}(B(s)) = b_0(s) \quad \text{versus} \quad H_1 : C\text{vec}(B(s)) \neq b_0(s)$$

Grid Point

Whole Tract

$$S_n(s_j) = nd(s_j)^T [C(\Sigma_\eta(s_j, s_j) \otimes \Omega_X^{-1}) C^T]^{-1} d(s_j)$$

Global Test Statistics

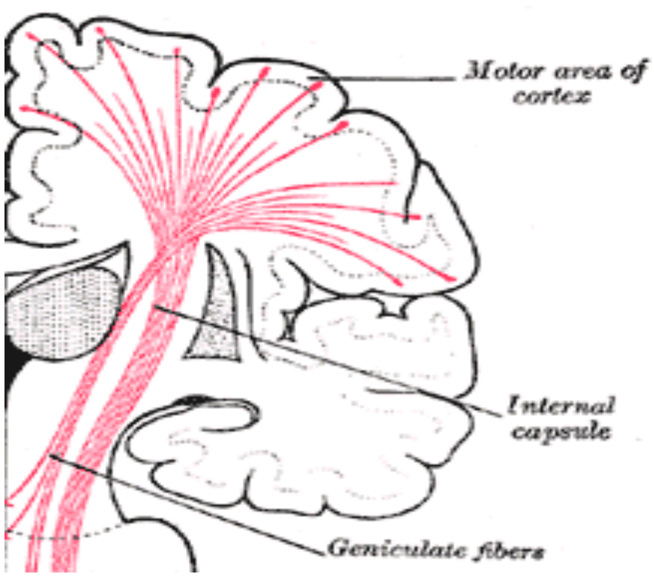
Local Test Statistics

$$S_n = n \int_0^{L_0} d(s)^T [C(\Sigma_\eta(s,s) \otimes \Omega_X^{-1}) C^T]^{-1} d(s) ds$$

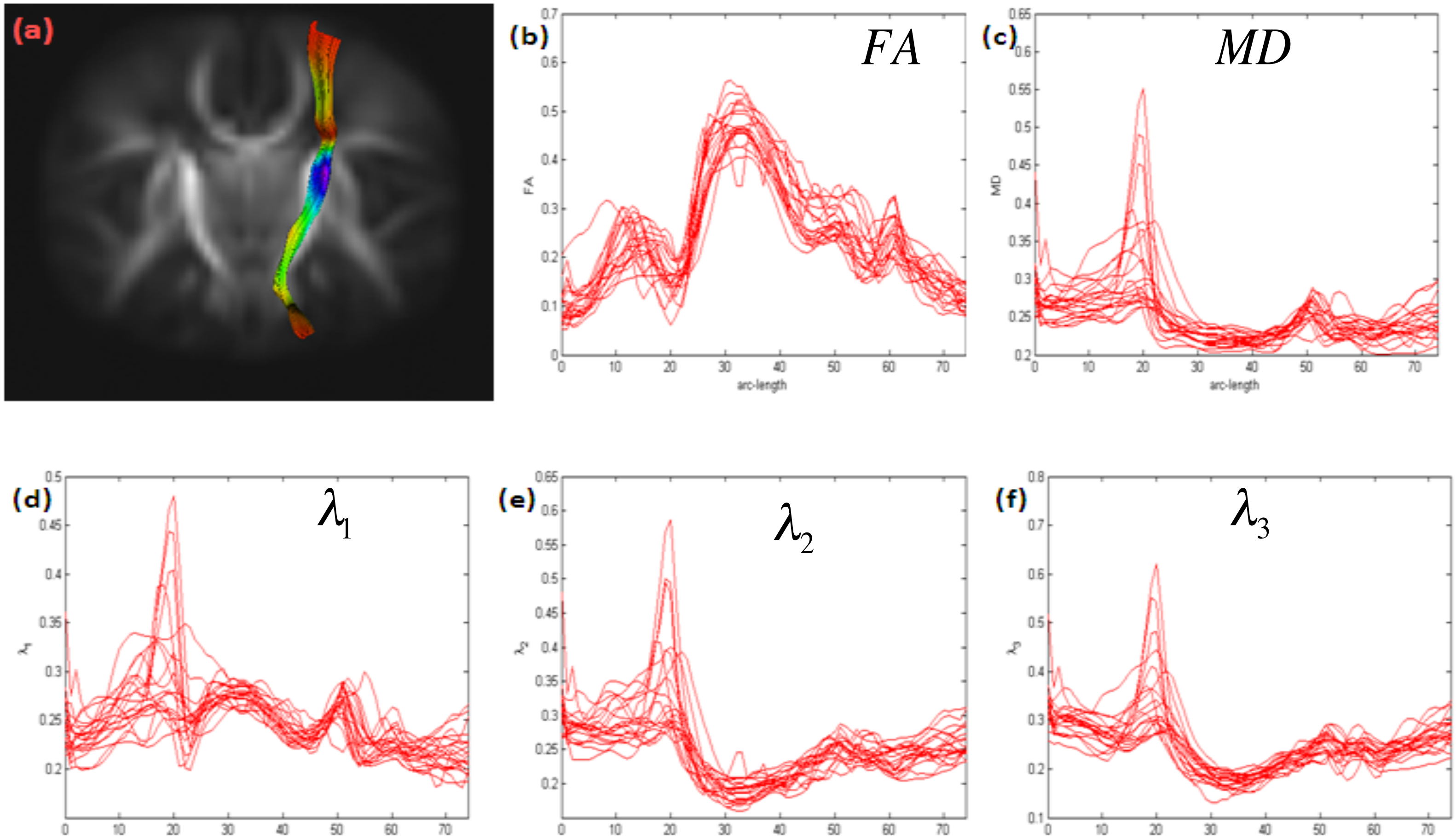
$$S_n(s_j) \Rightarrow \chi_k^2(m) \quad \text{and} \quad S_n \Rightarrow \sum_{k=1}^{K} w_k \chi_k^2(1),$$

Real Data

- PI: Dr. John H. Gilmore from Dept of Psychiatry at UNC-CH
- 128 healthy full-term infants: 75 males and 53 females
- Mean gestational age: 298±17.6 days, range: 262 - 433 days
- 5 diffusive outcomes: FA, MD,
- Internal Capsule

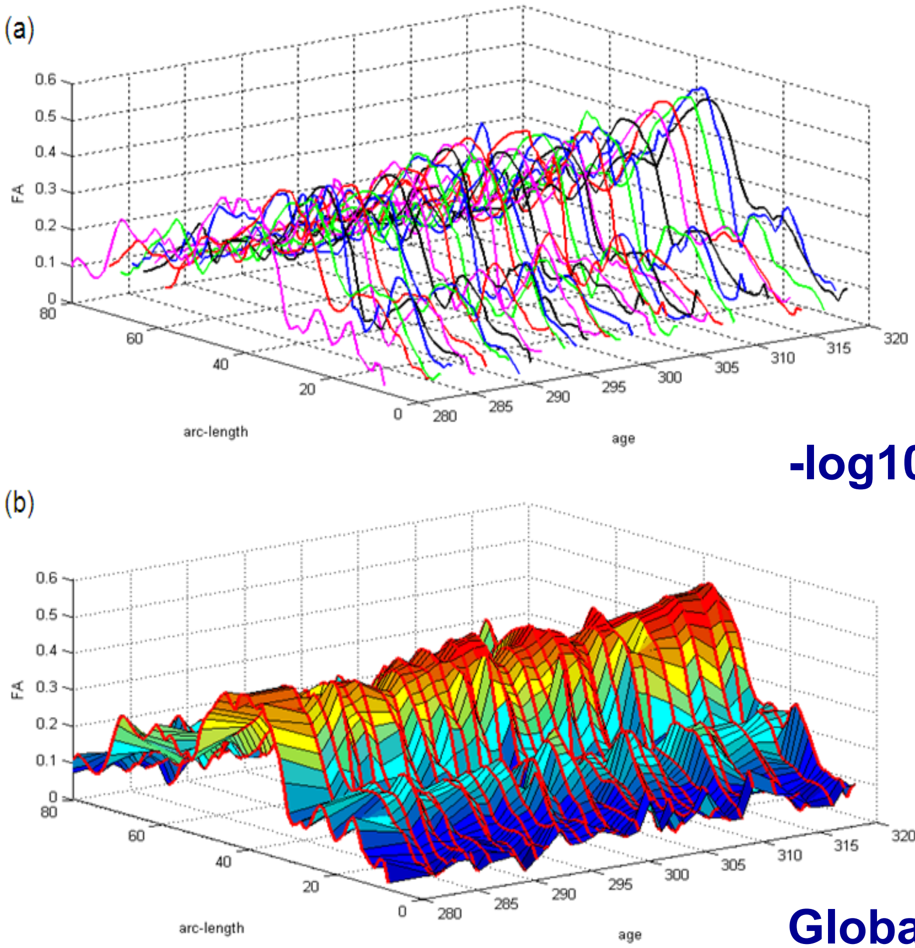


Real Data



Diffusion properties = Gender + Gestational age

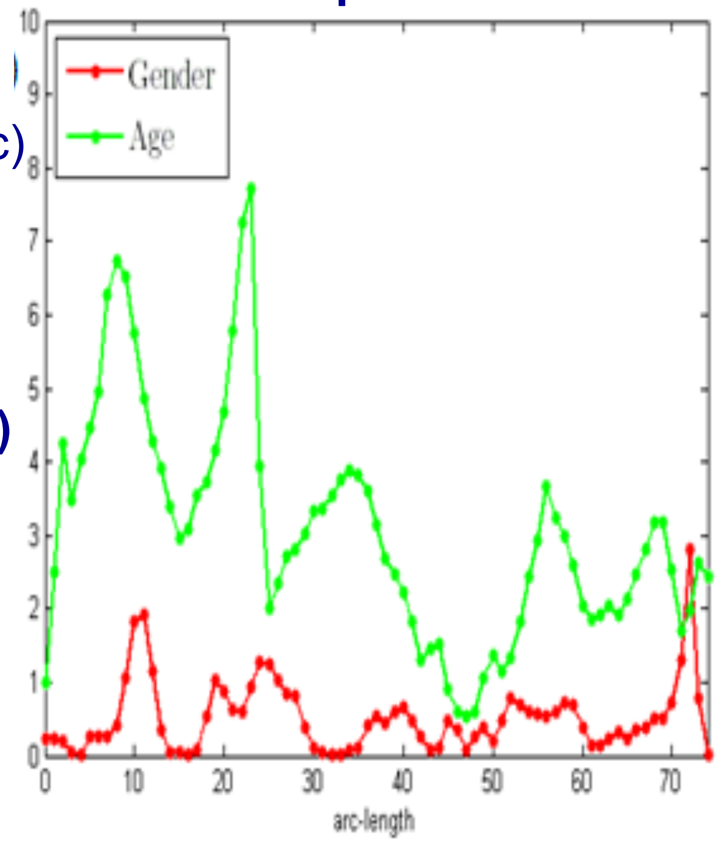
Real Data



Local p-value

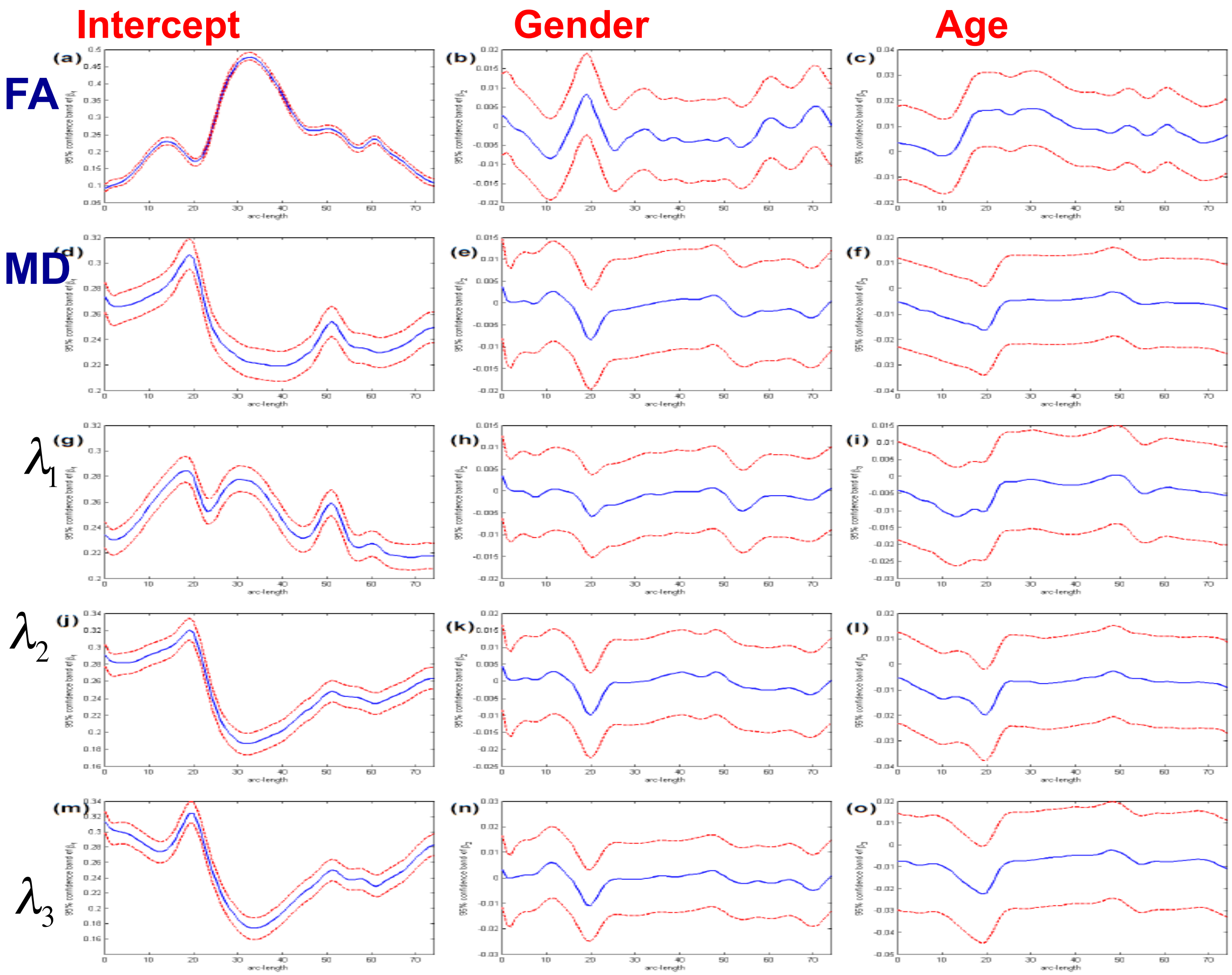
(c)

-log10(p)

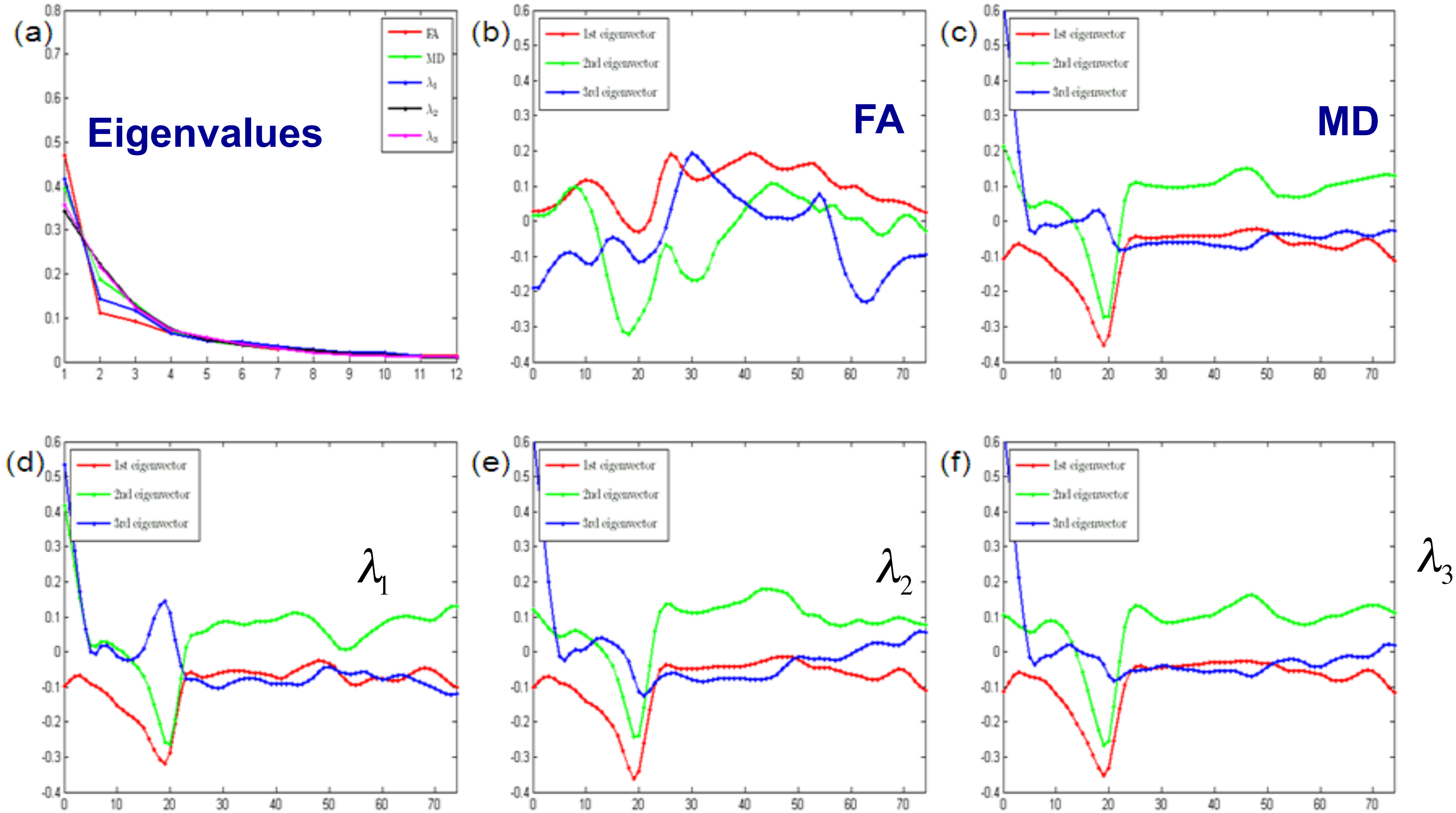


Global p-value: Age (<0.001), Gender (0.341)

Confidence Bands



FPCA



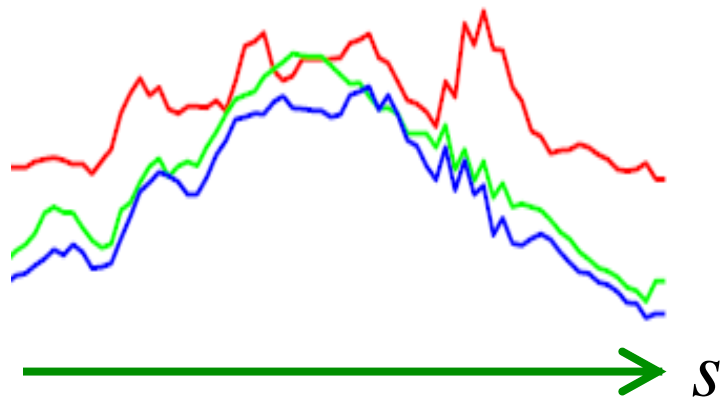
(a)
Eigenvalues
FA
MD
λ_1
λ_2
λ_3
(b)
1st eigenvector
2nd eigenvector
3rd eigenvector
FA
(c)
MD
(d)
λ_1
(e)
λ_2
(f)
λ_3

Longitudinal Extensions

Longitudinal Data

$t \uparrow$ $y_i(s,t_3)$, $y_i(s,t_2)$, $y_i(s,t_1)$

Spatial-temporal Process



Functional Mixed Effect Models

$$y_i(s,t) = x_i(t)^T B(s) + z_i(t)^T \xi_i(s) + \eta_i(s,t) + \varepsilon_i(s,t)$$

Objectives:

Dynamic functional effects of covariates of interest on functional response.

FMEM

Decomposition:

$$y_i(d,t) = x_i(t)^T B(d) + z_i(t)^T \xi_i(d) + \eta_i(d,t) + \varepsilon_i(d,t)$$

Global Noise Components **Local Correlated Noise**

$$\eta_i(\bullet,\bullet) \sim SP(0,\Sigma_\eta), \quad \xi_i(\bullet) \sim SP(0,\Sigma_\xi) \qquad \varepsilon_i(\bullet) \sim SP(0,\Sigma_\varepsilon),$$

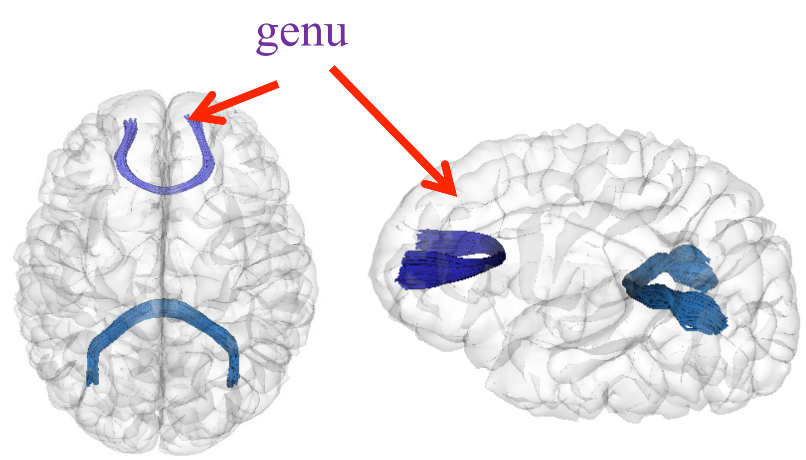
$$\sqrt{n}\{\mathrm{vec}(\hat{B}(d) - B(d) - 0.5O(H^2)) : d \in D\} \xrightarrow{L} G(0,\Sigma_B(d,d'))$$

Yuan et al. (2014). NeuroImage.
Zhu, Chen, Yuan, and Wang (2014). Arxiv.

Real Data

Gender: Male/Female	83/54
Gestational age at birth (weeks)	38.67 ± 1.74
Age at scan 1 (days)	297.89 ± 13.90
Age at scan 2 (days)	655.34 ± 24.00
Age at scan 3 (days)	1021.70 ± 28.26
Number of Gradient directions	
dir6/dir42 at scan 1	80/24
dir6/dir42 at scan 2	59/44
dir6/dir42 at scan 3	42/49

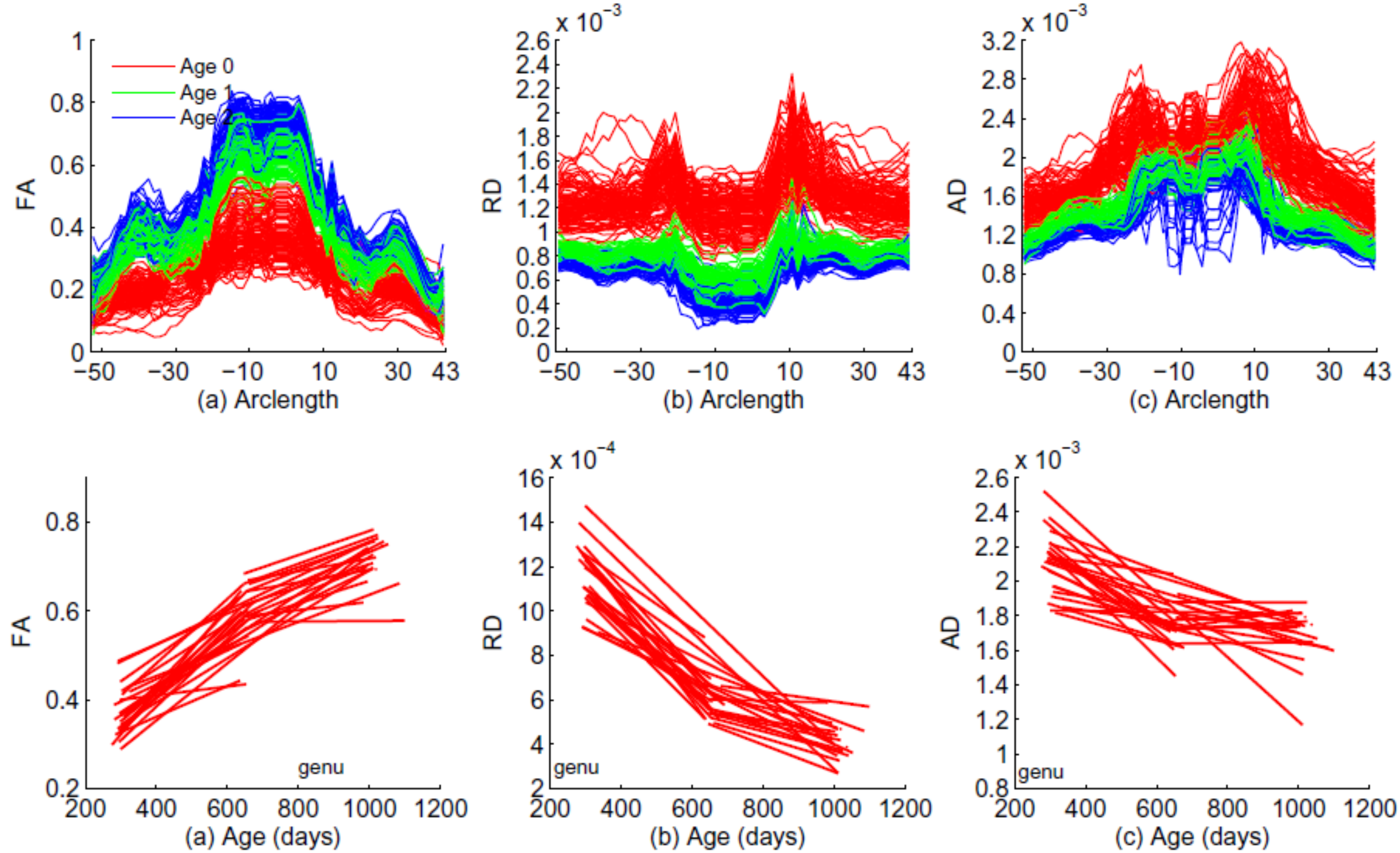
Available scans	N
Neonate scan only	1
1 year scan only	2
2 year scan only	3
Neonate + 1 year scan	43
Neonate + 2 year scan	30
1 year + 2 year scan	28
Neonate + 1 year + 2 year scan	30



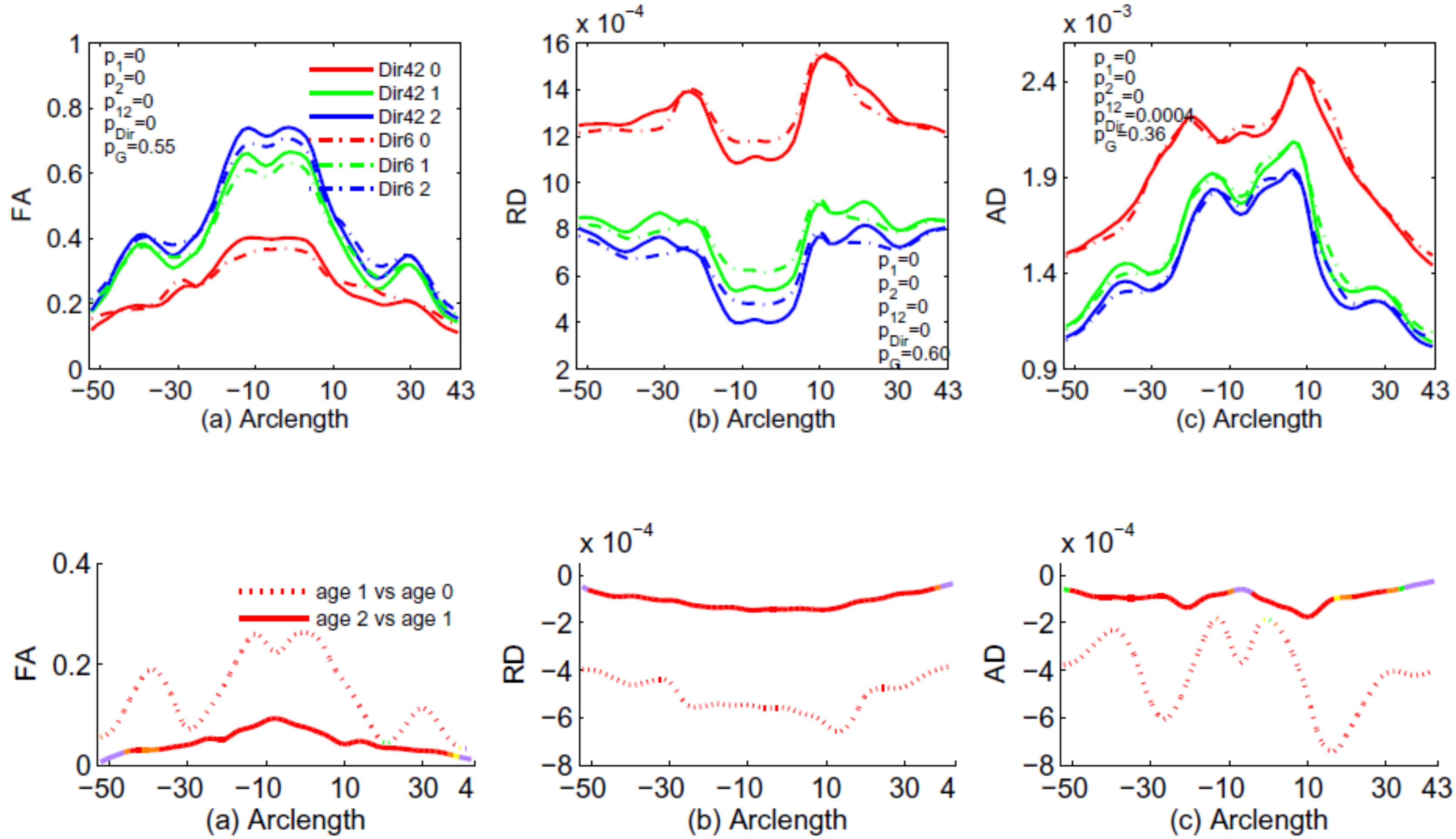
DTImaging parameters:

- **TR/TE = 5200/73 ms**
- **Slice thickness = 2mm**
- **In-plane resolution = 2x2 mm^2**
- **b = 1000 s/mm^2**
- **One reference scan b = 0 s/mm^2**
- **Repeated 5 times when 6 gradient directions applied.**

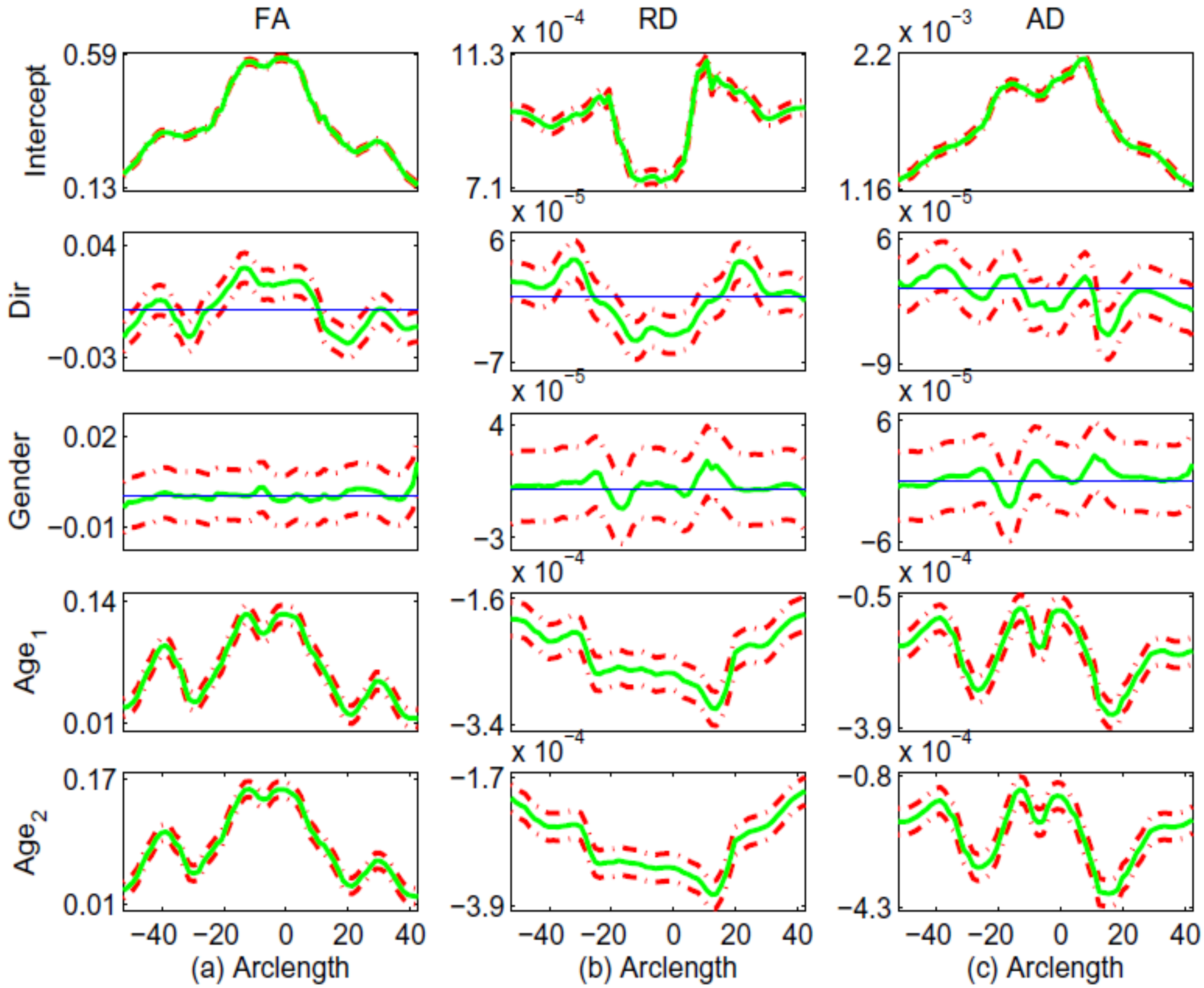
Real Data


Age 0
Age 1
Age 2
FA
(a) Arclength
RD
x 10^-3
(b) Arclength
AD
x 10^-3
(c) Arclength
FA
genu
(a) Age (days)
RD
x 10^-4
genu
(b) Age (days)
AD
x 10^-3
genu
(c) Age (days)

Real Data Analysis Results



Real Data Analysis Results



Functional Nonlinear Mixed Effects Model

Decomposition:

$$y_{i,j}(s) = f(\phi_i(s), x_{i,j}) + \varepsilon_{i,j}(s), \quad \phi_i(s) = \beta(s) + b_i(s)$$

Nonlinear Function   Mixed Effect   Fixed Effect   Random Effect

Asymptotic Normality:

$$\sqrt{n}\{\mathrm{vec}(\tilde{\beta}(s) - \beta(s) - O(h^2)) : d \in D\} \xrightarrow{L} G(0, \Sigma_\beta(s, s'))$$

Luo, Zhu, Kong, and Zhu (2015). IPMI

Estimation Procedure

MLE for each grid point s_m:

$$y_{i,j}(s_m) = f(\phi_i(s_m), x_{i,j}) + \varepsilon_{i,j}(s_m) \longrightarrow \hat{\beta}(s_m)$$

Smoothing:

$$\tilde{\beta}(s) = \sum_{m=1}^{M} \tilde{K}_h(s_m - s)\hat{\beta}(s_m)$$ **with kernel function**

$$\tilde{K}_h(s_m - s) = K_h(s_m - s) \Big/ \sum_{m=1}^{M} K_h(s_m - s)\hat{\beta}(s_m),$$

$$K_h(\cdot) = K(\cdot / h) / h$$

Inference Procedure

Hypothesis Test:

$$H_0 : R\beta(s) = b_0(s) \text{ for all } s \quad vs. \quad H_1 : R\beta(s) \neq b_0(s)$$

Simultaneous Confidence Bands:

$$P\left(\hat{\beta}_l^{L,\alpha}(s) < \beta_l(s) < \hat{\beta}_l^{U,\alpha}(s) \text{ for all } s\right) = 1 - \alpha$$

Real Data

- We analyzed a data set taken from a national database for autism research (NDAR) (http: // http://ndar.nih.gov/), an NIH-funded research data repository,
- that aims to accelerate progress in autism spectrum disorders (ASD) research through data sharing, data harmonization, and the reporting of research results.
- 416 high quality MRI scans are available for 253 children (126 males and 127 females) with 45 grid points.

Model

Gompertz Function:

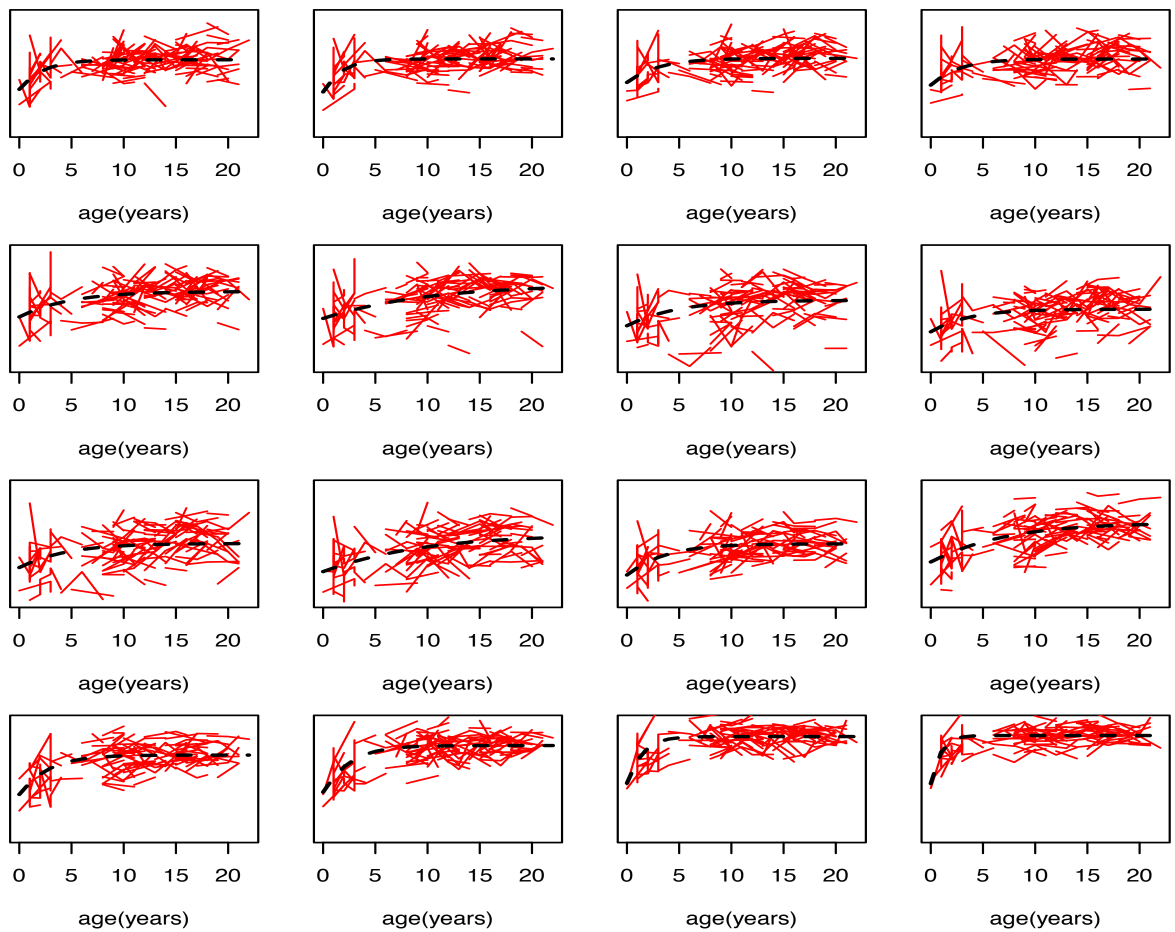
$$y = \text{asymptote} \cdot \exp(-\text{delay} \cdot \exp(-\text{speed} \cdot t))$$

used to characterize longitudinal white matter development during early childhood.

Functional Version:

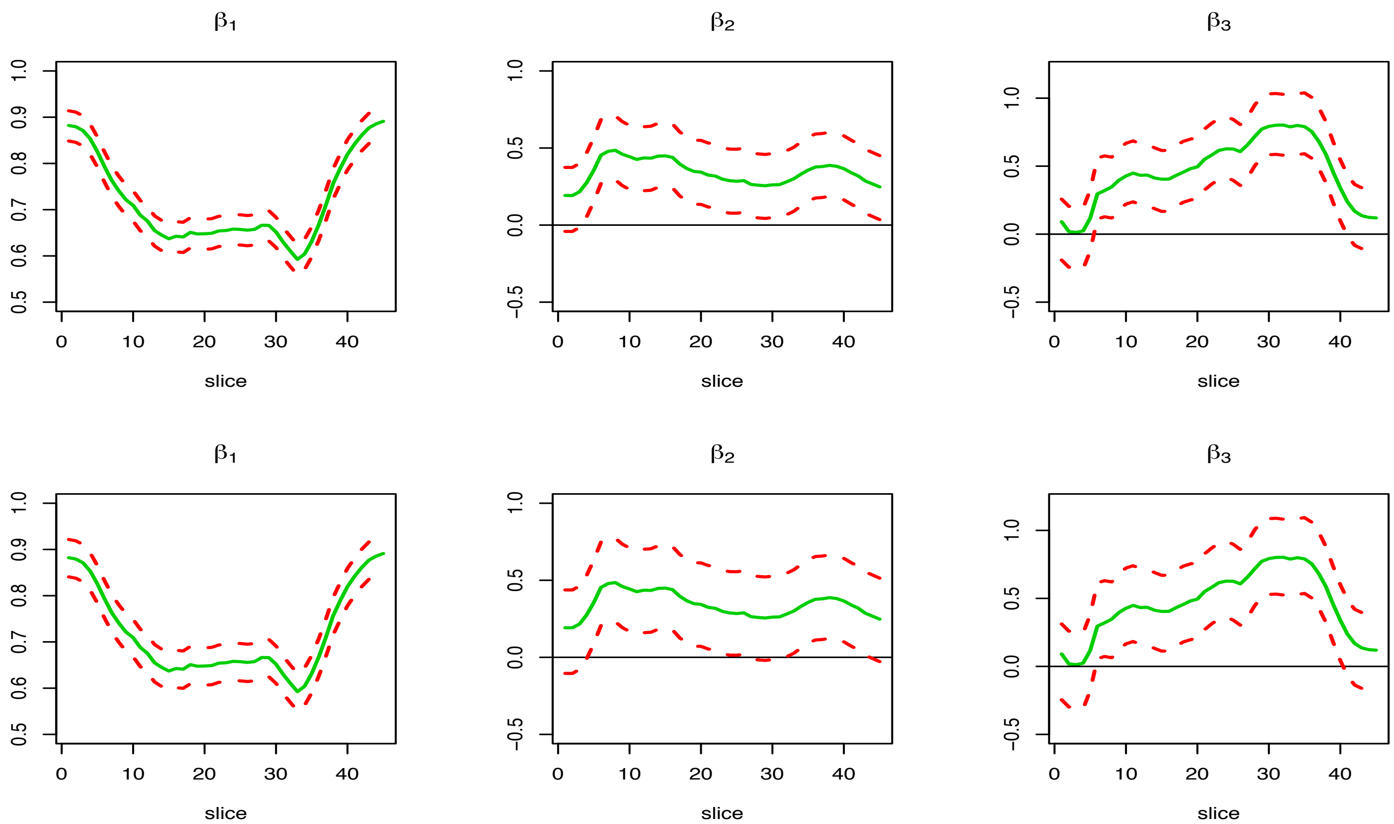
$$y_{i,j}(s) = \phi_{1i}(s) \cdot \exp\left(- \phi_{2i}(s) \cdot \phi_{3i}(s)^{t_{i,j}} \right) + \varepsilon_{i,j}(s)$$

Real Data Analysis Results



Tract (red solid lines) varying as a function of age for grid points from 25 to 40, the black dash curves are estimated curves.

Real Data Analysis Results



The 100(1 − α)% simultaneous confidence bands of parameters for α = 0.05 (the first row) and α = 0.01 (the second row). The green solid and red dash curves are, respectively, the estimated curves and their corresponding 95% and 99% simultaneous confidence bands.

Big Missing Data Problem

- **Missing imaging data caused by design**

  Add new imaging techniques in the middle of studies (ADNI),
  Drop out,

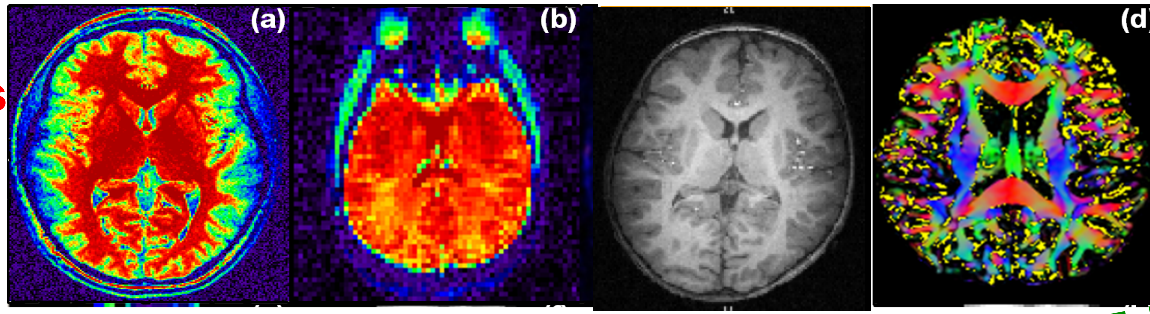
- **Missing imaging data caused by acquisition**

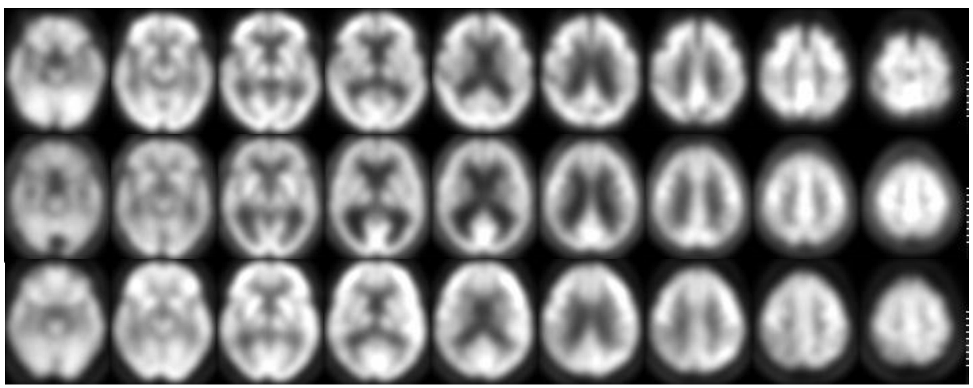
  Head motion, Physiological fluctuations, Artifact-induced problem,
  Susceptibility artifact.

- **Predicting neural activity or brain development longitudinally**

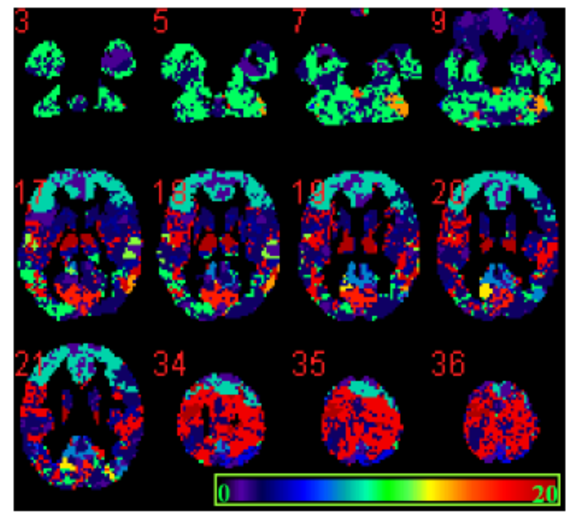
Data Structure

Across modalities





Local spatial smoothness



Across subjects

SGPP Model

$$y_{i,j}(d) = \mathbf{x}_i^T \boldsymbol{\beta}_j(d) + \eta_{i,j}(d) + \epsilon_{i,j}(d) \text{ for } i = 1, \ldots, n;\ j = 1, \ldots, J, \quad (1)$$

where

Across subjects

$\boldsymbol{\beta}_j(d) = (\beta_{j1}(d), \ldots, \beta_{jp}(d))^T$ is a $p \times 1$ vector of regression coefficients at voxel d,

Across modality

$\eta_{i,j}(d)$ characterizes both individual image variations from $\mathbf{x}_i^T \boldsymbol{\beta}_j(d)$ and the medium-to-long-range dependence of imaging data between $y_{i,j}(d)$ and $y_{i,j}(d')$ for any $d \neq d'$,

Local smoothness

$\epsilon_{i,j}(d)$ are spatially correlated errors that capture the local (or short-range) dependence of imaging data.

Hyun,J.W., Li, Y. M., J. H. Gilmore, Z. Lu, M. Styner, H. Zhu (2014) SGPP. NeuroImage

fPCA+SAR

Combining a functional principal component model and a multivariate simultaneous autoregressive model, we obtain an approximation of model (1) given by

$$\begin{aligned} y_{i,j}(d) \quad \approx \quad & \mathbf{x}_i^T \boldsymbol{\beta}_j(d) + \sum_{l=1}^{L_0} \xi_{ij,l} \psi_{j,l}(d) \\ & + \rho \frac{1}{|N(d)|} \sum_{d' \in N(d)} \left(y_{i,j}(d') - \mathbf{x}_i^T \boldsymbol{\beta}_j(d') - \sum_{l=1}^{L_0} \xi_{ij,l} \psi_{j,l}(d') \right) \\ & + e_{i,j}(d). \end{aligned}$$

Ventricle Surface Data

- The surface data set of the left lateral ventricle consists of 43 infants (23 males and 20 females).
- The gestational ages of the 43 infants range from 234 to 295 days and their mean gestational age is 263 days with standard deviation 12.8 days.
- The left lateral ventricle surface of each infant is represented by 1002 location vectors with each location vector consisting of the spatial x, y, and z coordinates of the corresponding vertex on the SPHARM-PDM surface.
- We randomly splitted the data set into a training set (70%) and a test set (30%).
- We fitted the SGPP model to the training set and predicted the multiple measurements at the hold-out voxels, based on the measurements at other voxels and the fitted model, for each subject in the test set.

Ventricle Surface Data

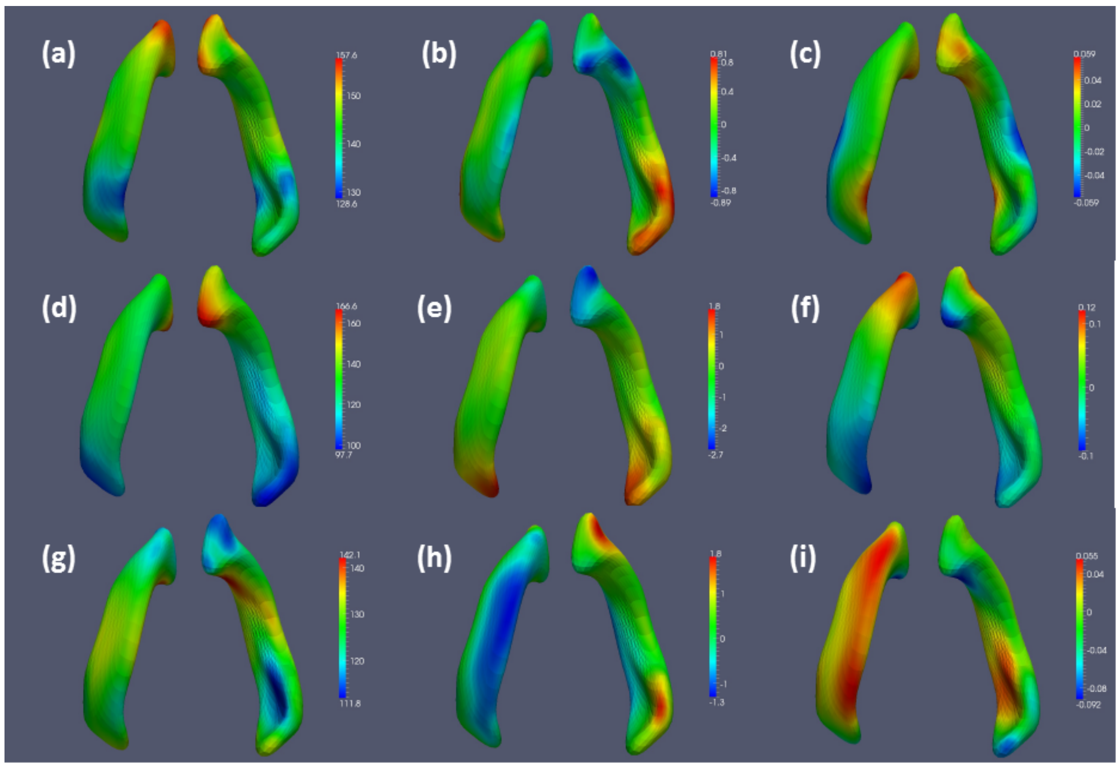
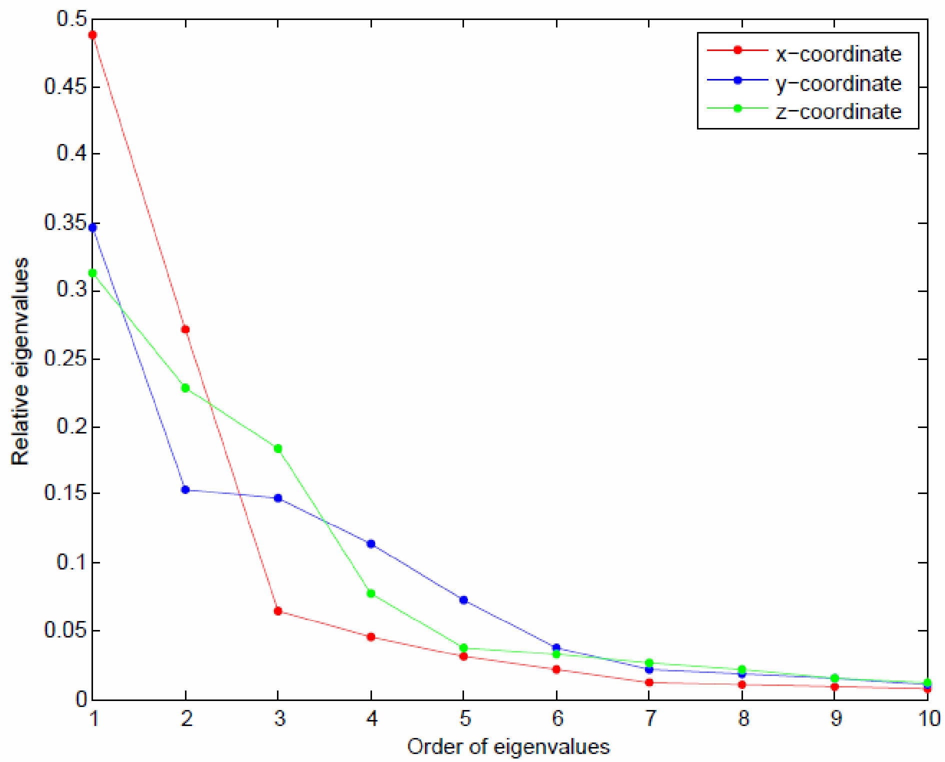


Figure : Results from the surface data of the left lateral ventricle: (a) $\hat{\beta}_{11}(d)$; (b) $\hat{\beta}_{12}(d)$; (c) $\hat{\beta}_{13}(d)$; (d) $\hat{\beta}_{21}(d)$; (e) $\hat{\beta}_{22}(d)$; (f) $\hat{\beta}_{23}(d)$; (g) $\hat{\beta}_{31}(d)$; (h) $\hat{\beta}_{32}(d)$; and (i) $\hat{\beta}_{33}(d)$.

Ventricle Surface Data



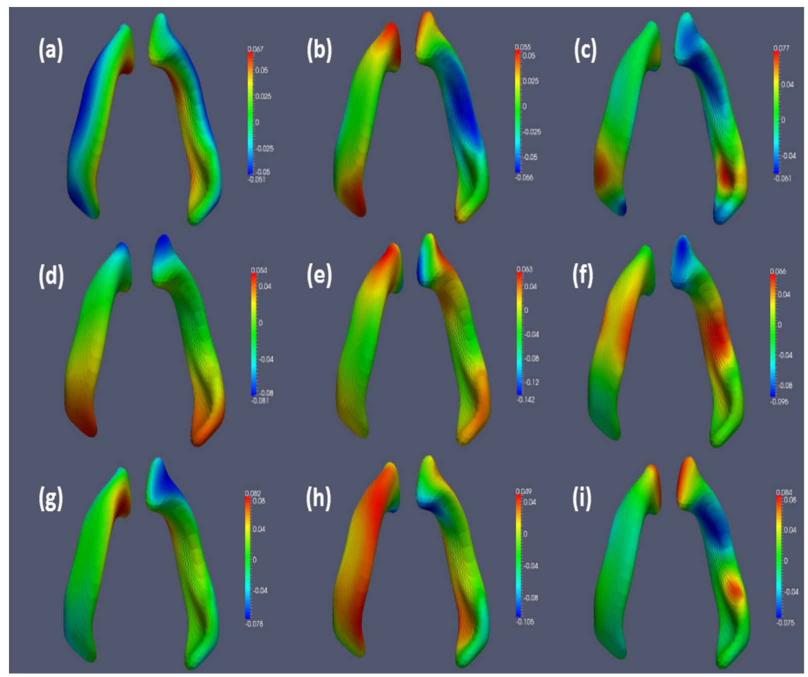


Figure : Results from the surface data of the left lateral ventricle: (a) $\hat{\psi}_{1,1}(d)$; (b) $\hat{\psi}_{1,2}(d)$; (c) $\hat{\psi}_{1,3}(d)$; (d) $\hat{\psi}_{2,1}(d)$; (e) $\hat{\psi}_{2,2}(d)$; (f) $\hat{\psi}_{2,3}(d)$; (g) $\hat{\psi}_{3,1}(d)$; (h) $\hat{\psi}_{3,2}(d)$; and (i) $\hat{\psi}_{3,3}(d)$.

Ventricle Surface Data

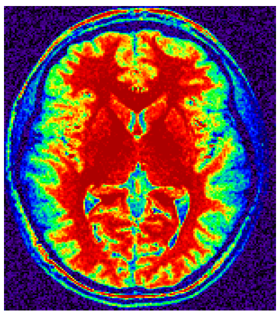
Table : rtMSPE for the surface data of the left lateral ventricle

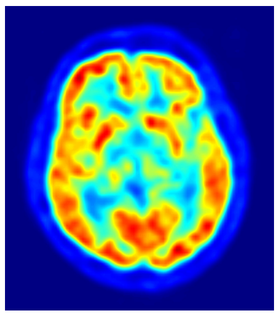
Missingness		VWLM	GLM+fPCA	SGPP
10%	x-coordinate	1.9272	0.9810	0.0738
	y-coordinate	2.2448	1.3455	0.1067
	z-coordinate	2.1554	1.1753	0.0926
30%	x-coordinate	1.9337	1.0197	0.1156
	y-coordinate	2.2655	1.3827	0.1657
	z-coordinate	2.1906	1.2069	0.1446
50%	x-coordinate	1.9263	1.0294	0.1615
	y-coordinate	2.2012	1.3471	0.2204
	z-coordinate	2.1862	1.1830	0.1924

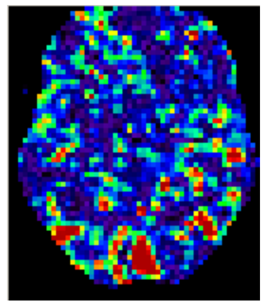
Multiscale Adaptive Regression Models

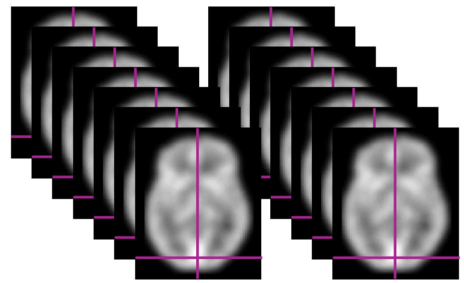
Neuroimaging Data with Discontinuity

Noisy Piecewise Smooth Function with Unknown Jumps and Edges









Subject1 *Subject2*

Covariates (e.g., age, gender, diagnostic, stimulus)

SVCM

Decomposition:

$$y_i(d) = f(x_i, B(d) + \eta_i(d)) + \varepsilon_i(d), d \in D$$

Piecewise Smooth Varying Coefficients

$$B(d) \in L^K$$

Long-range Correlation

$$\eta_{ij}(\bullet) \sim SP(0, \Sigma_\eta)$$

Short-range Correlation

$$\varepsilon_{ij}(\bullet) \sim SP(0, \Sigma_\varepsilon),$$

3D volume/ 2D surface

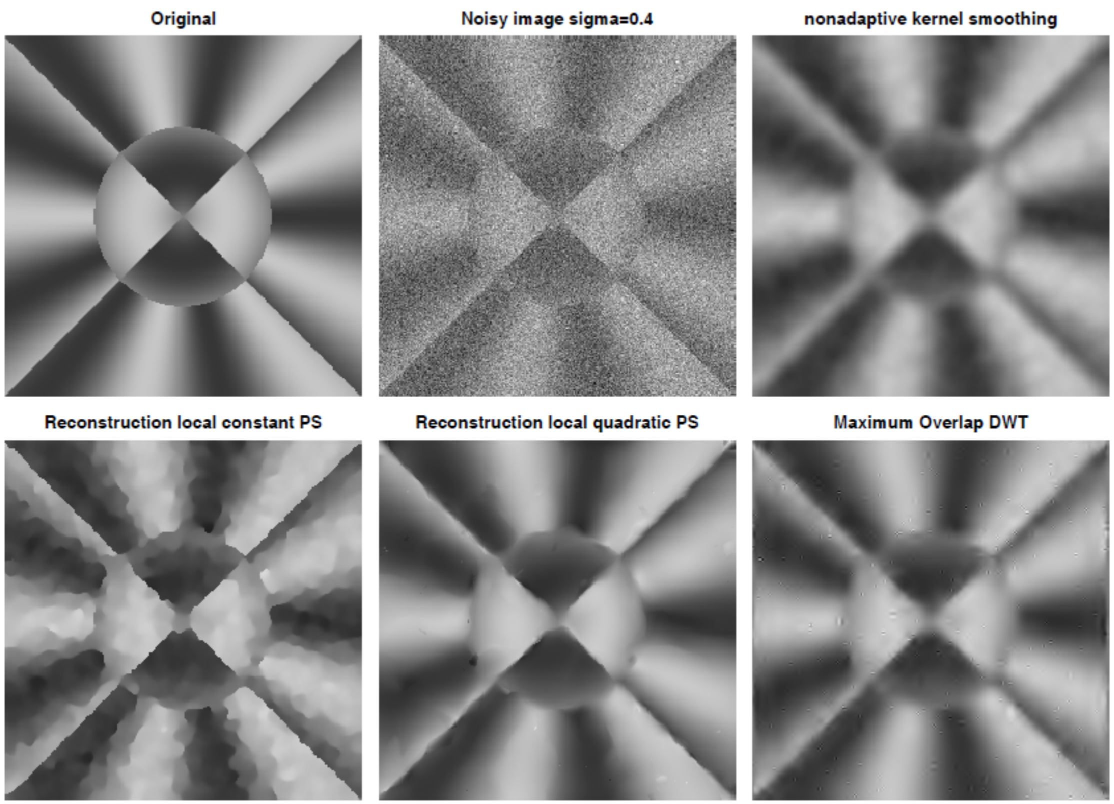
Covariance operator:

$$\Sigma_y(d, d') = \Sigma_\eta(d, d') + \Sigma_\varepsilon(d, d)$$

Smoothing Methods

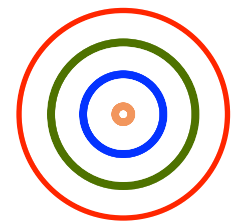
Propogation-Seperation Method

J. Polzehl and V. Spokoiny, (2000,2005)



Features

- Increasing Bandwidth



$$0 < h_0 < h_1 < \cdots < h_S = r_0$$

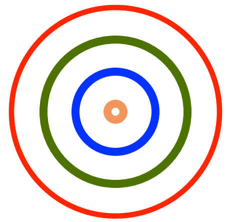
- Adaptive Weights

- Adaptive Estimates

Smoothing Methods

MARM

At each voxel d



- **Increasing Bandwidth**
- **Adaptive Weights**
- **Adaptive Estimates**

$$0 < h_0 < h_1 < \cdots < h_S = r_0$$

$$\hat{\mu}(d;h_0) \quad \omega(d,d';h_1) \quad \hat{\mu}(d;h_1) \quad \omega(d,d';h_2) \quad \omega(d,d';h_2) \quad \hat{\mu}(d;h_S)$$

Stopping Rule

$$\omega(d,d';h_s) = K_{loc}(\| d - d' \| / h_s) K_{st}(D_\mu(d,d';h_{s-1}) / C_n)$$

$$D_\mu(d,d';h_{s-1}) = \rho(\hat{\mu}(d;h_{s-1}), \hat{\mu}(d';h_{s-1}))$$

SVCM

Adaptively Smooth coefficients

$$\beta_j(d;h_1) = \sum_{d' \in B(d,h_1)} w(d,d';h_1)\beta_j(d;h_0) \Big/ \sum_{d' \in B(d,h_1)} w(d,d';h_1)$$

Estimated covariance operator

$$\hat{\Sigma}(d,d') = \sum_{i=1}^{n} \hat{\eta}_i(d)\hat{\eta}_i(d')^T$$

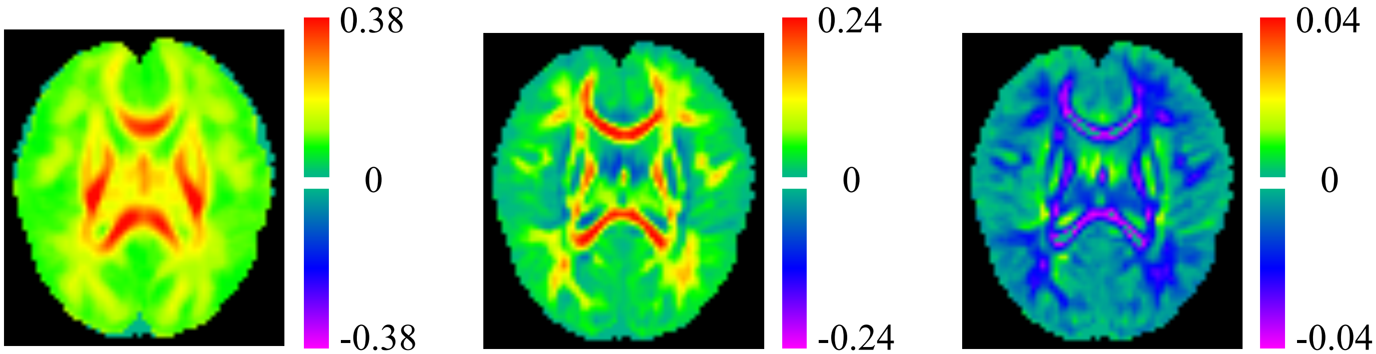
Estimated eigenfunctions

$$\{(\hat{\lambda}_{kl}, \hat{\psi}_{kl}(d)) : l = 1, \ldots, \infty\}$$

Functional Principal Component Analysis

Challenging Issues

- Smoothing coefficient images, while preserving unknown boundaries
- Different patterns in different coefficient images
- Calculating standard deviation images
- Asymptotic theory
- Risk theory



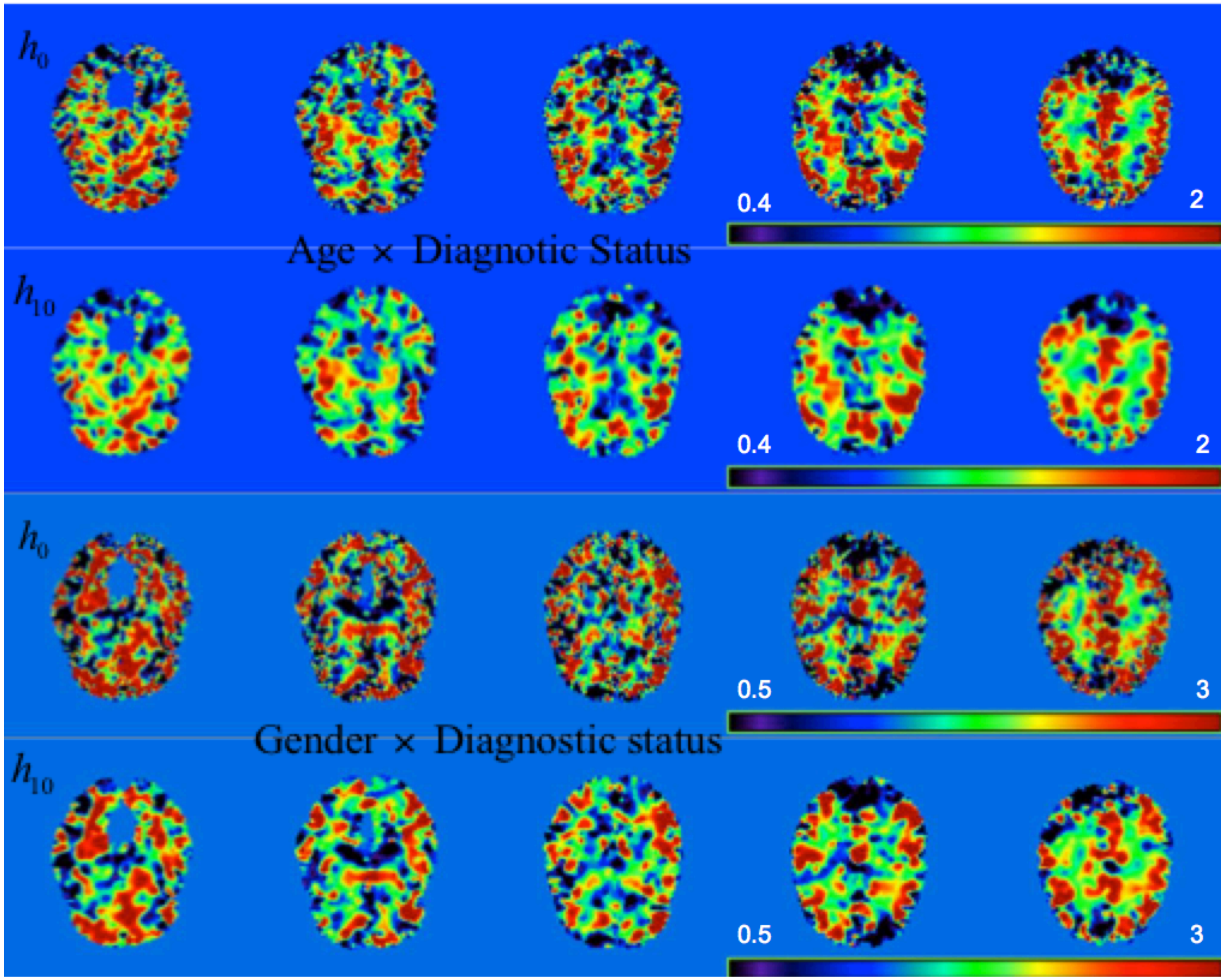
Real Data

- Attention deficit hyperactivity disorder (ADHD) is a developmental disorder.
- ADHD is the most commonly studied and diagnosed psychiatric disorder in children.
- It affects about 3 to 5 percent of children globally and diagnosed in about 2 to 16 percent of school aged children.
- It directly cost about $36 billion per year in US.
- ADHD-200 Global Competition is a grassroots initiative event to accelerate the understanding of ADHD.

Real Data

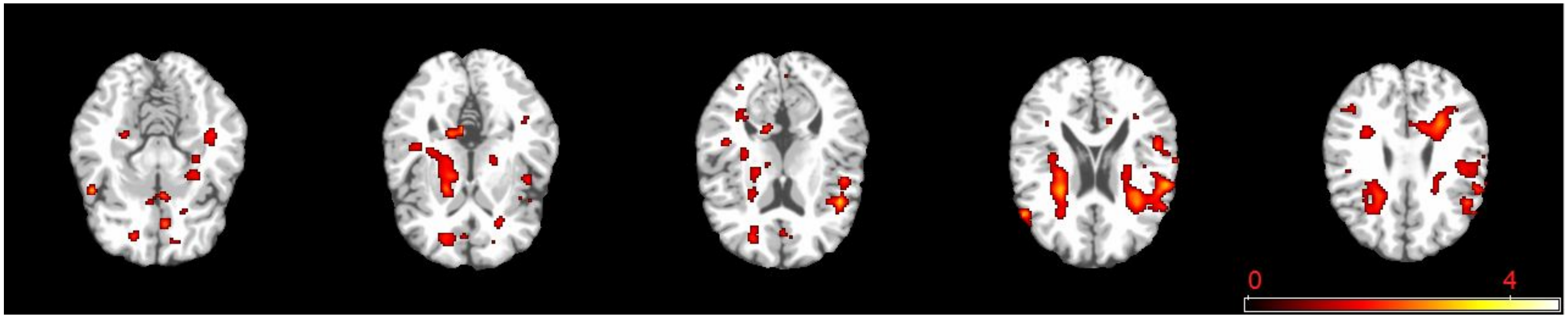
- 174 subjects, 99 normal, and 75 ADHD-combined
- Response: White Matter, orignial $256 \times 256 \times 198$, down size to $128 \times 128 \times 99$
- Covariate variables: age, gender, group (diagnosis status), and whole brain volume
- Goal: look at the group effects, including interaction effects with age and gender

Real Data

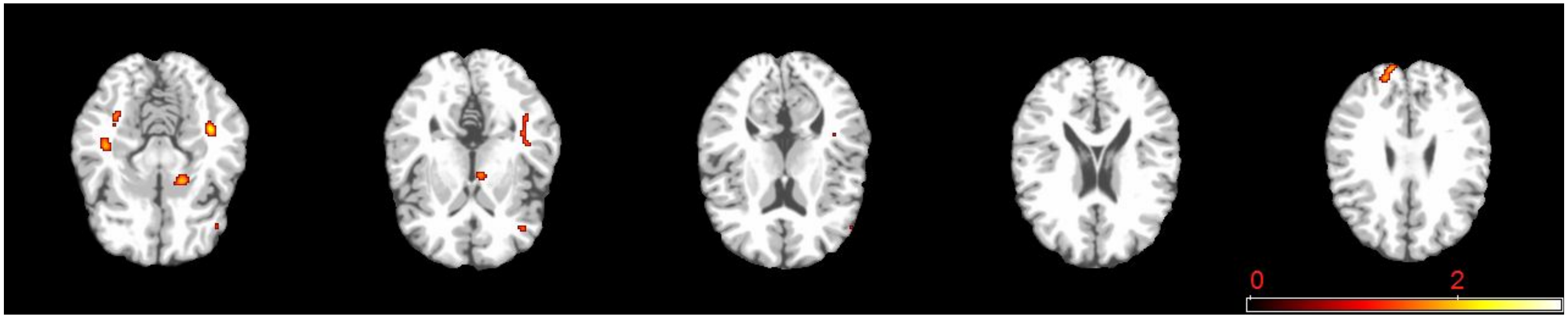

h_0
0.4
2
Age × Diagnotic Status
h_{10}
0.4
2
h_0
0.5
3
Gender × Diagnostic status
h_{10}
0.5
3

Significant regions overlayed on template

Age × Diagnotic Status



Gender × Diagnostic status



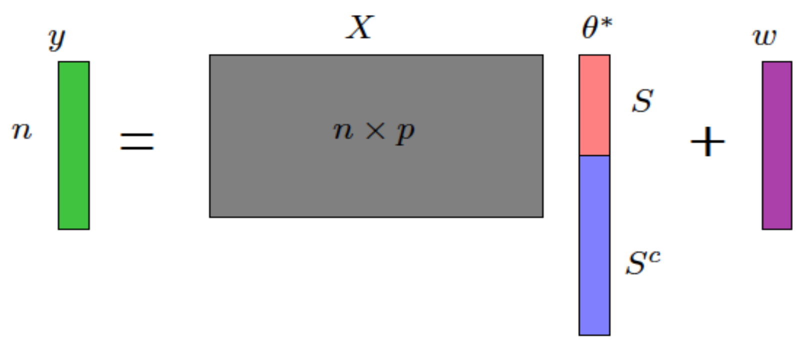
Scalar-on-Imaging Regression

HRM versus FRM

Data $\{(y_i, X_i) : i = 1, \cdots, n\}$ $\quad X_i = \{X_i(d) : d \in D\}$

$$y_i = < X_i, \theta > + \varepsilon_i$$

Strategy 1: Discrete Approach
(High-dimension Regression Model (HRM))

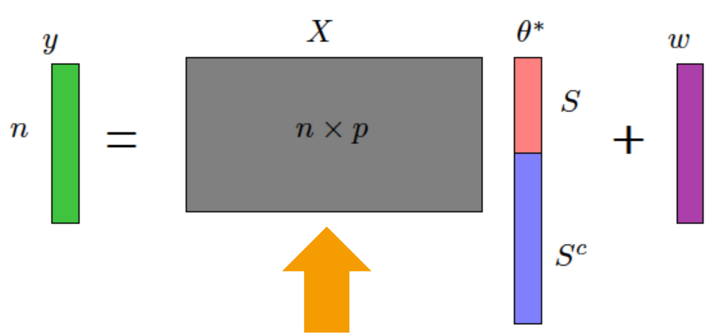


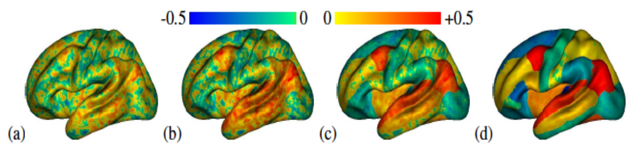
Strategy 2: Functional Regression Model (FRM)

$$y_i = \theta_0 + \int_D \theta(d) X_i(d) m(d) + \varepsilon_i$$

High-dimension Regression Model

Approach 1: Regularization Methods





$$\widehat{\theta} \in \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^T \theta)^2 + \lambda_n \sum_{j=1}^{p} |\theta_j|$$

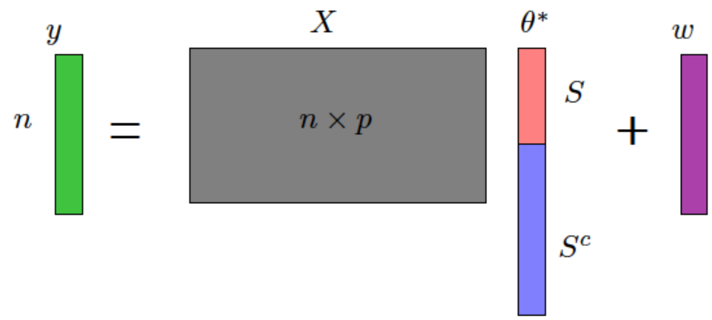
Key Conditions:

- **Sparsity of S**
- **Restricted null-space property for design matrix X**

High-dimension Regression Model

Tensor Structure:

- Ultra-high dimensionality (256^3)
- Spatial structure

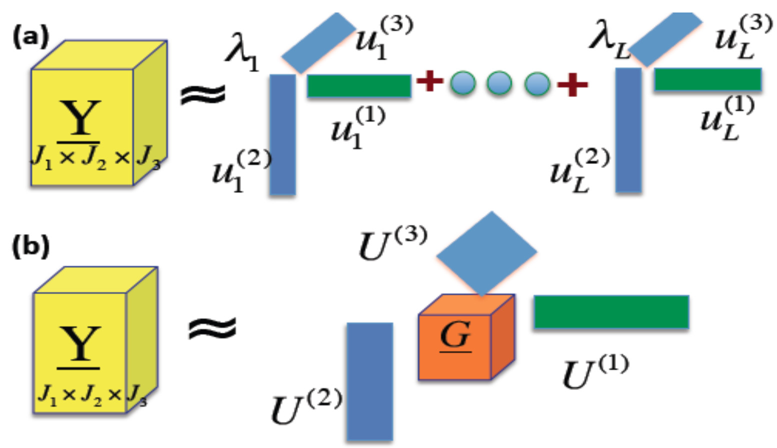


Zhou, Li, and Zhu (2013)
Li, Zhou, and Li (2013)

CP decomposition

θ

Tucker decomposition



Scalar-on-Image Models

Strategy 2: Functional Approach

$$y_i = \theta_0 + \int_D \theta(d) X_i(d) m(d) + \varepsilon_i$$

$$\theta(d) = \sum_{k=1}^{\infty} \theta_k \psi_k(d)$$

$$y_i = \theta_0 + \sum_{k=1}^{\infty} \theta_k \int_D \psi_k(d) X_i(d) m(d) + \varepsilon_i$$

Basis Methods: fixed and data-driven basis functions

$$K_\theta = \{\theta(d) = \sum_{k=1}^{\infty} \theta_k \psi_k(d) : (\theta_1, \cdots) \in \ell^2\} \Longleftrightarrow C(d,d') = Cov(X(d), X(d')) = \sum_{k=1}^{\infty} \lambda_k \zeta_k(d) \zeta_k(d')$$

Key Conditions

Key Conditions: an excellent set of basis functions

- **Sparsity of basis representation** $\{\theta_k : k = 1, \cdots\}$
- **Decay rate of spectral of** C **or** $K^{1/2} C K^{1/2}$

$$\theta(d) \approx \sum_{k=1}^{K} \theta_k \psi_k(d) \qquad K << n$$

Extensions

- Functional linear Cox regression models
- Generalized scalar-to-image regression models
- Multiscale Functional Linear models