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?

Human Brain Project

aims to simulate the complete human brain on Supercomputers to better understand how it functions.

BRAIN Funding Opportunities

The Brain Research through
    Advancing Innovative Neurotechnologies or BRAIN,
aims to reconstruct the activity of every single neuron as they fire simultaneously in different brain circuits, or perhaps even whole brains.
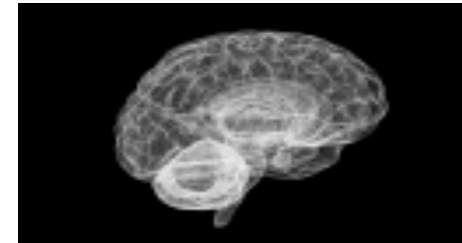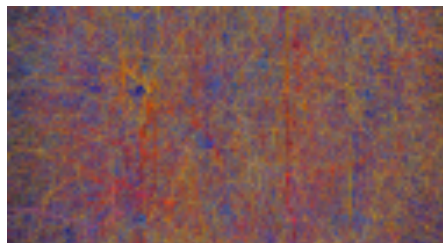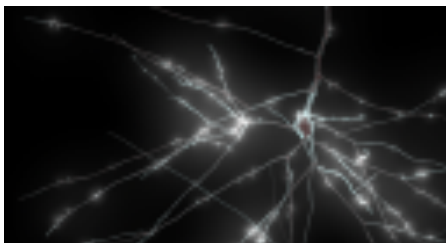
Big Neuroimaging Data

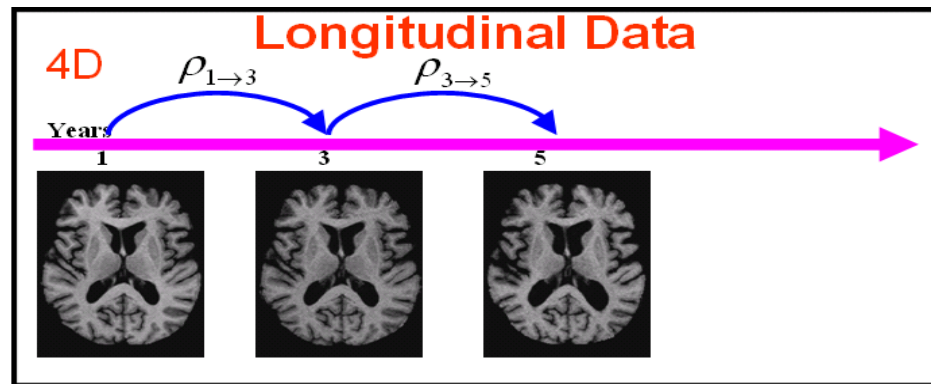**NIH normal brain development**
 1000 Functional Connectome Project
  Alzheimer's Disease Neuroimaging Initiative
   National Database for Autism Research (NDAR)
   Human Connectome Project
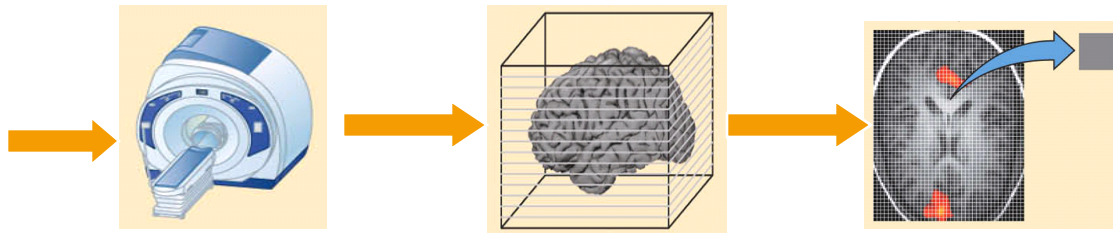
www.guysandstthomas.nhs.uk/.../T/Twins400.jpg
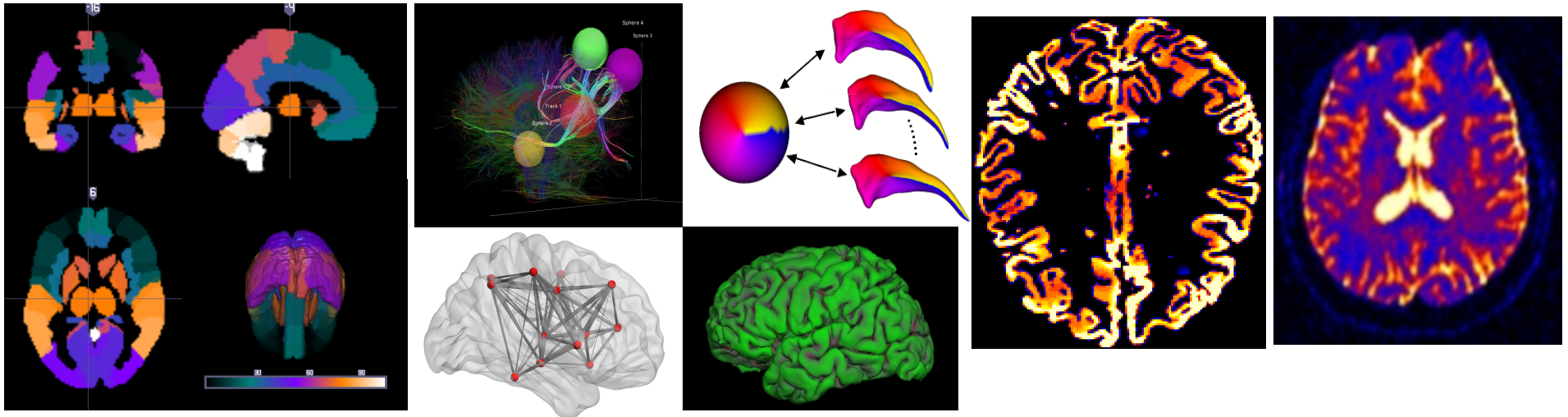
Complex Study Design

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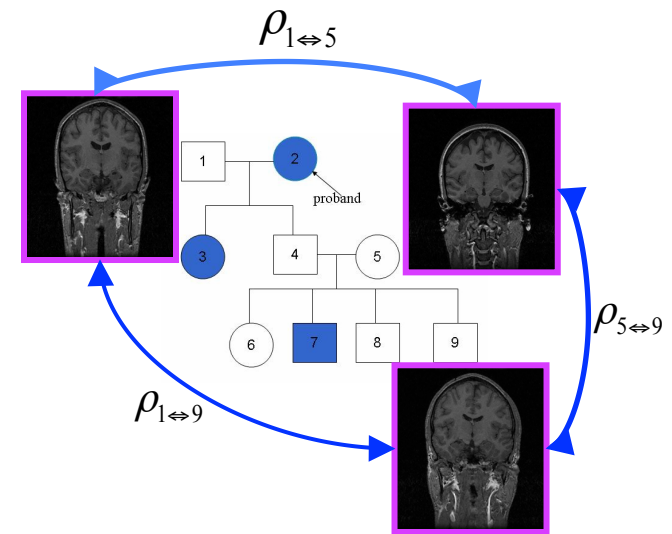
Complex Data Structure

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.

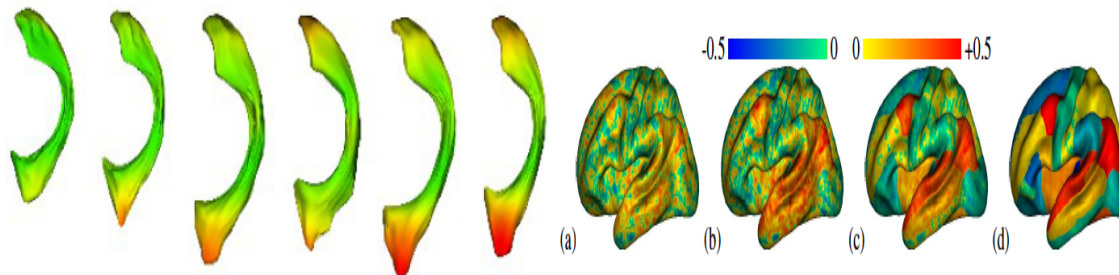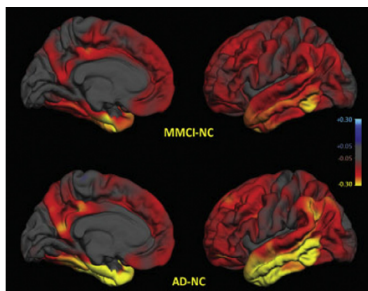## Scalar-on-Image (SI) model
Clinical variables as response, image data as predictors

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

Noisy Imaging Data

Spatial Maps

- `Reconstruction'

  `Segmentation'

- `Registration'

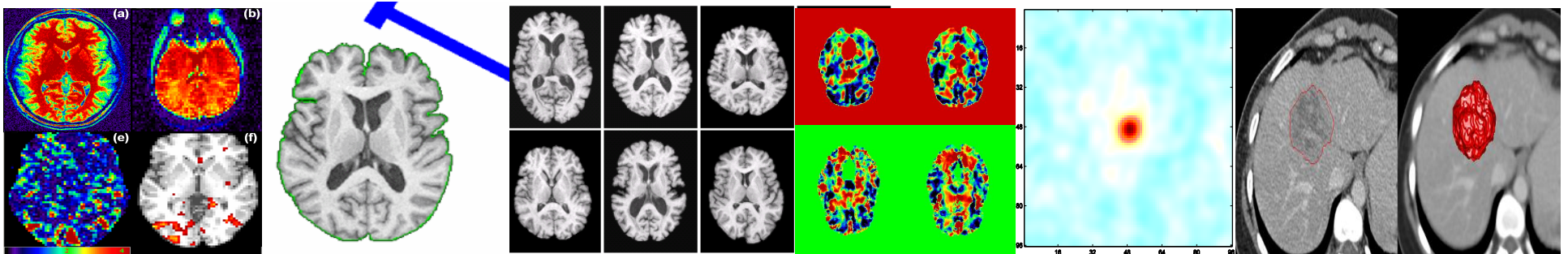 `Smoothing'

Estimation

Inference

Prediction
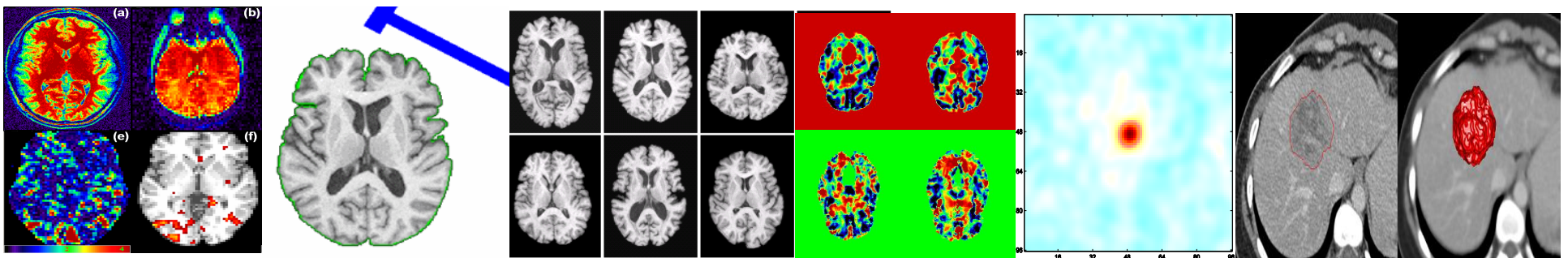
Noisy Imaging Data

Key Features

- **Infinite Dimension**

- **Spatial Smoothness**

- **Spatial Correlation**

- **Spatial Heterogeneity**

UNIVERSITY OF ALBERTA

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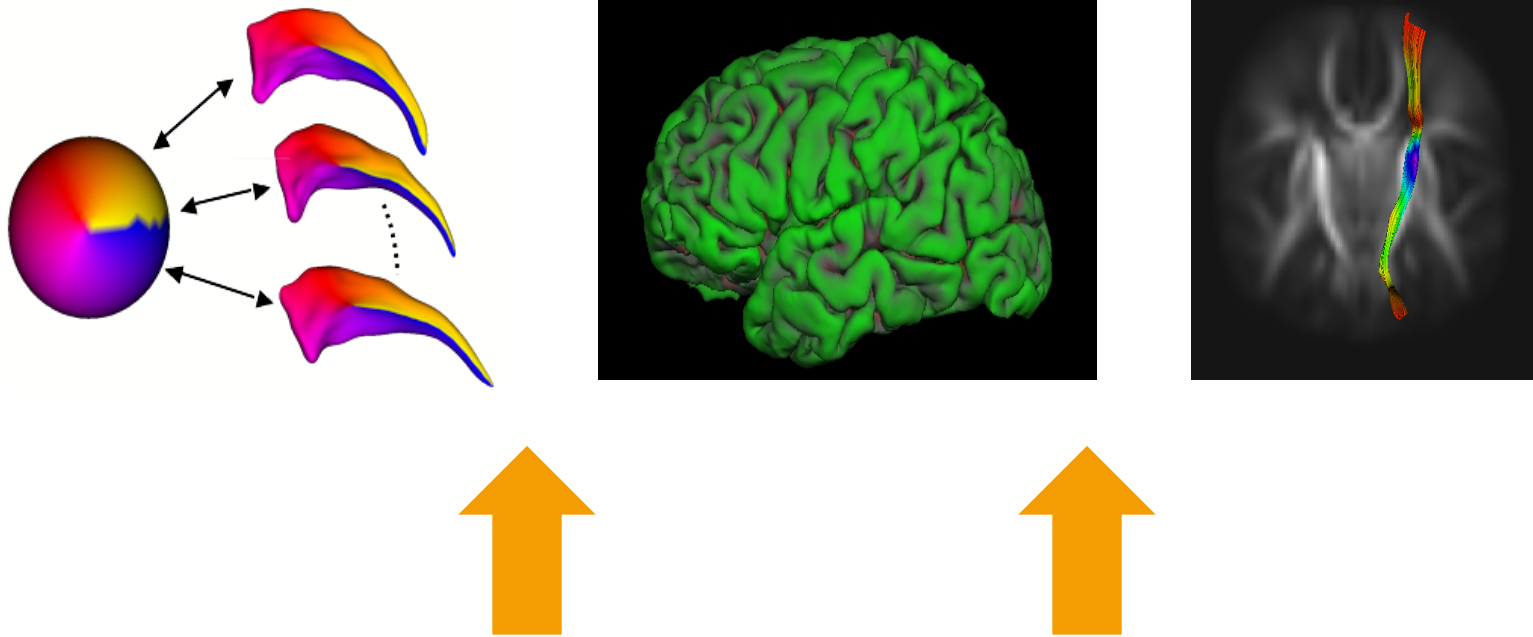— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Varying Coefficient Models
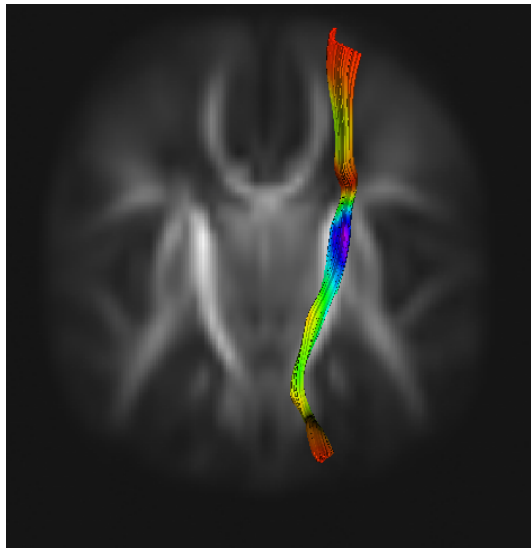
Varying Coefficient Models

Reading materials:

1. **Zhu, H. T.**, Chen, K. H., Yuan, Y. and Wang, J. L. (2014). Functional Mixed Processes Models for Repeated Functional Data. In submission.
2. Liang, J. L., Huang, C., and **Zhu, H.T** (2014). Functional single-index varying coefficient models. In submission.
3. Yuan, Y., Gilmore, J., Geng, X. J., Styner, M., Chen, K. H., Wang, J. L., and **Zhu, H.T** (2013). A longitudinal functional analysis framework for analysis of white matter tract statistics. *NeuroImage*, in press.
4. Yuan, Y., **Zhu, H.T.**, Styner, M., J. H. Gilmore., and Marron, J. S. (2013). Varying coefficient model for modeling diffusion tensors along white matter bundles. *Annals of Applied Statistics.* 7(1):102-125..
5. Zhu, H.T., Li, R. Z., Kong, L.L. (2012). Multivariate varying coefficient models for functional responses. *Ann. Stat.* 40, 2634-2666.
6. Hua, Z.W., Dunson, D., Gilmore, J.H., Styner, M., and **Zhu, HT.** (2012). Semiparametric Bayesian local functional models for diffusion tensor tract statistics. *NeuroImage,* 63, 460-674.
7. **Zhu, HT.,** 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.
8. **Zhu, H.T.**, Styner, M., Tang, N.S., Liu, Z.X., Lin, W.L., Gilmore, J.H. (2010). FRATS: functional regression analysis of DTI tract statistics. *IEEE Transactions on Medical Imaging*, 29, 1039-1049.

9. Greven, S., Crainiceanu, C., Caffo, B., Reich, D. (2010). Longitudinal principal component analysis. E.J.Statist. 4, 1022-1054.
10. Goodlett, C.B., Fletcher, P. T., Gilmore, J. H., Gerig, G. (2009). Group analysis of dti fiber tract statistics with application to neurodevelopement. NeuroImage, 45, S133-S142.
11. Yushkevich, P. A., Zhang, H., Simon, T., Gee, J. C. (2008). Structure-specific statistical mapping of white matter tracts. NeuroImage, 41, 448-461.
12. Ramsay, J. O., Silverman, B. W. (2005). *Functional Data Analysis,* Springer-Verlag, New York.

UNIVERSITY OF ALBERTA

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

Smoothed Functional Data

Covariates (e.g., age, gender, diagnostic)
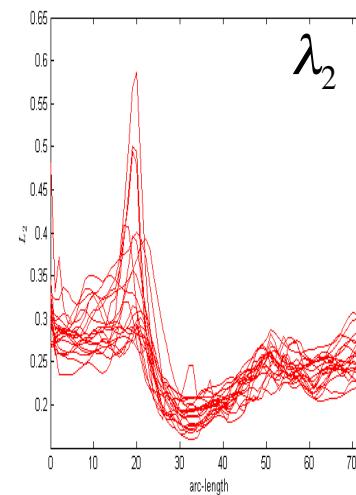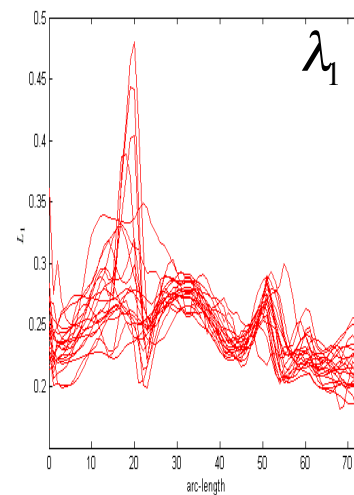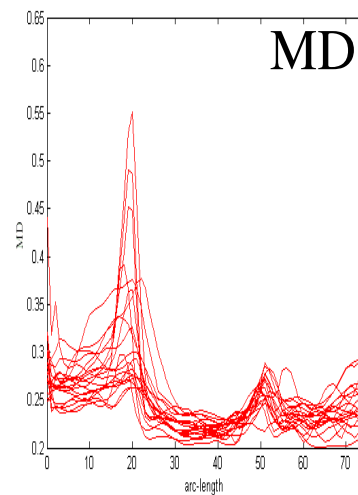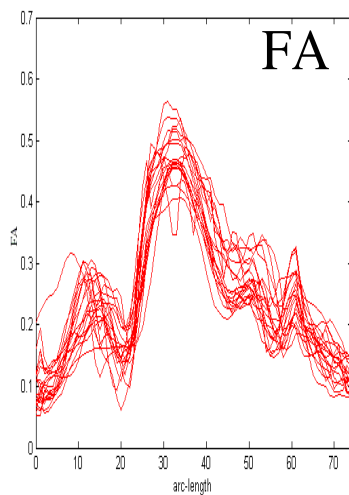
DTI Fiber Tract Data

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)**

$$x_1, \cdots, x_n$$

FA   MD λ_1 λ_2 λ_3
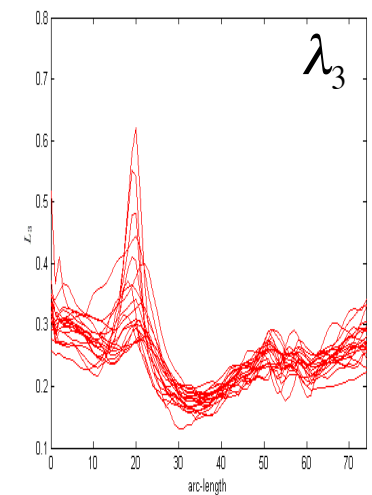
MVCM

Decomposition:

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

Coefficients   **Long-range Correlation**   **Short-range Correlation**

$$x_1, \cdots, x_n \qquad \eta_{i,k}(\bullet) \sim SP(0, \Sigma_\eta) \qquad \varepsilon_{i,k}(\bullet) \sim SP(0, \Sigma_\varepsilon),$$

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

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

Zhu, Li, and Kong (2012). AOS

MVCM

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

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

Key Advantage ➡ **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

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

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

MVCM

Asymptotics

$$\sqrt{n}[\hat{b}_{k,l}(s) - b_{k,l}(s) - \text{bias}(\hat{b}_{k,l}(s))] \Rightarrow G_{k,l}(\bullet)$$

Critical point

$$P(\sup_{s \in [0, L_0]} | G_{k,l}(s) | \leq C_{k,l}(\alpha)) = 1 - \alpha$$
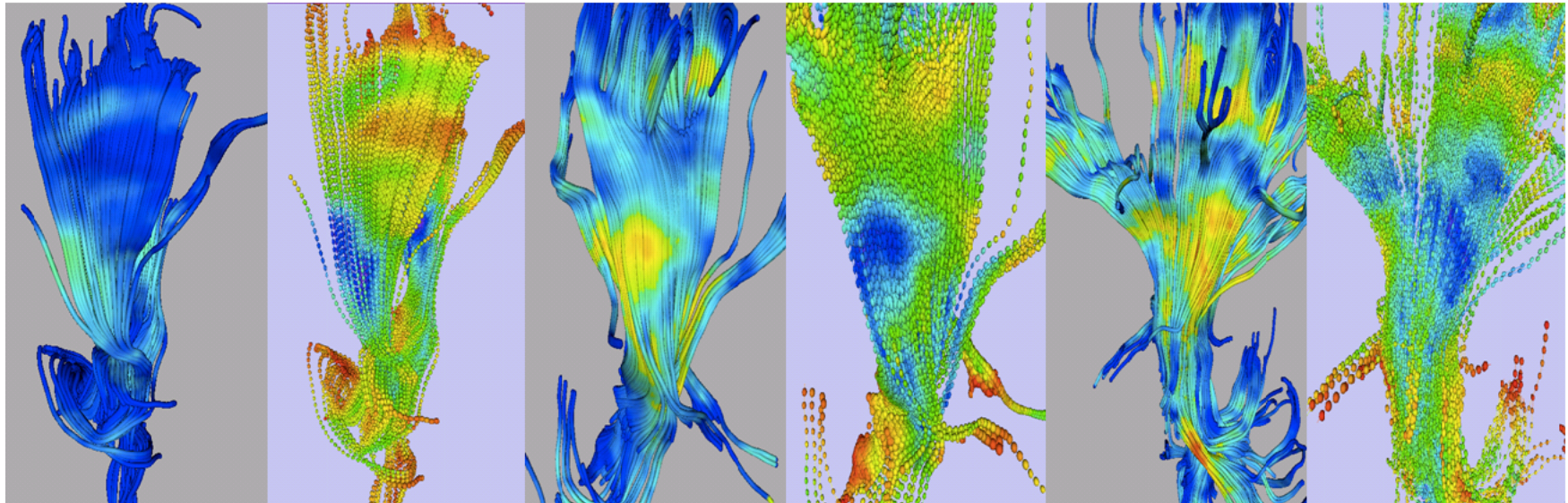
Confidence band

$$(\hat{b}_{k,l}(s) - \frac{C_{k,l}(\alpha)}{\sqrt{n}}, \ \hat{b}_{k,l}(s) + \frac{C_{k,l}(\alpha)}{\sqrt{n}})$$

Motivation

Diffusion Tensor Tract Statistics

FA   **Tensor**

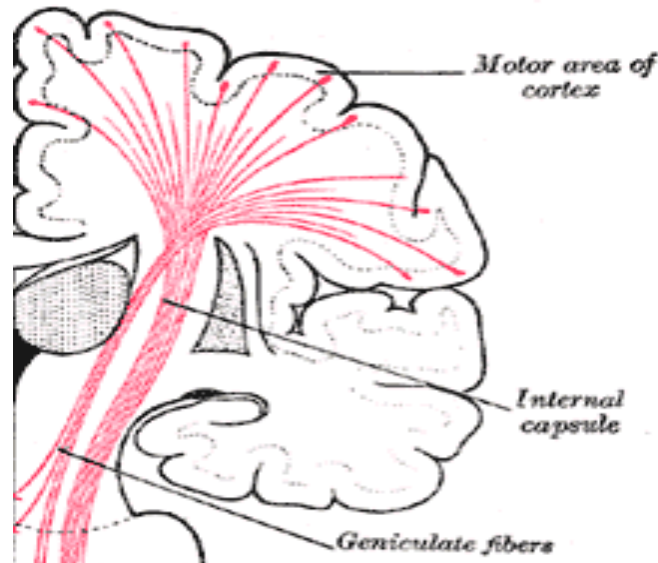(a1)   (b1)   (a2)   (b2)   (a3)   (b3)

2 week   **2 week**   **1 year**   **1 year**   **2 year**   **2 year**

UNIVERSITY OF ALBERTA

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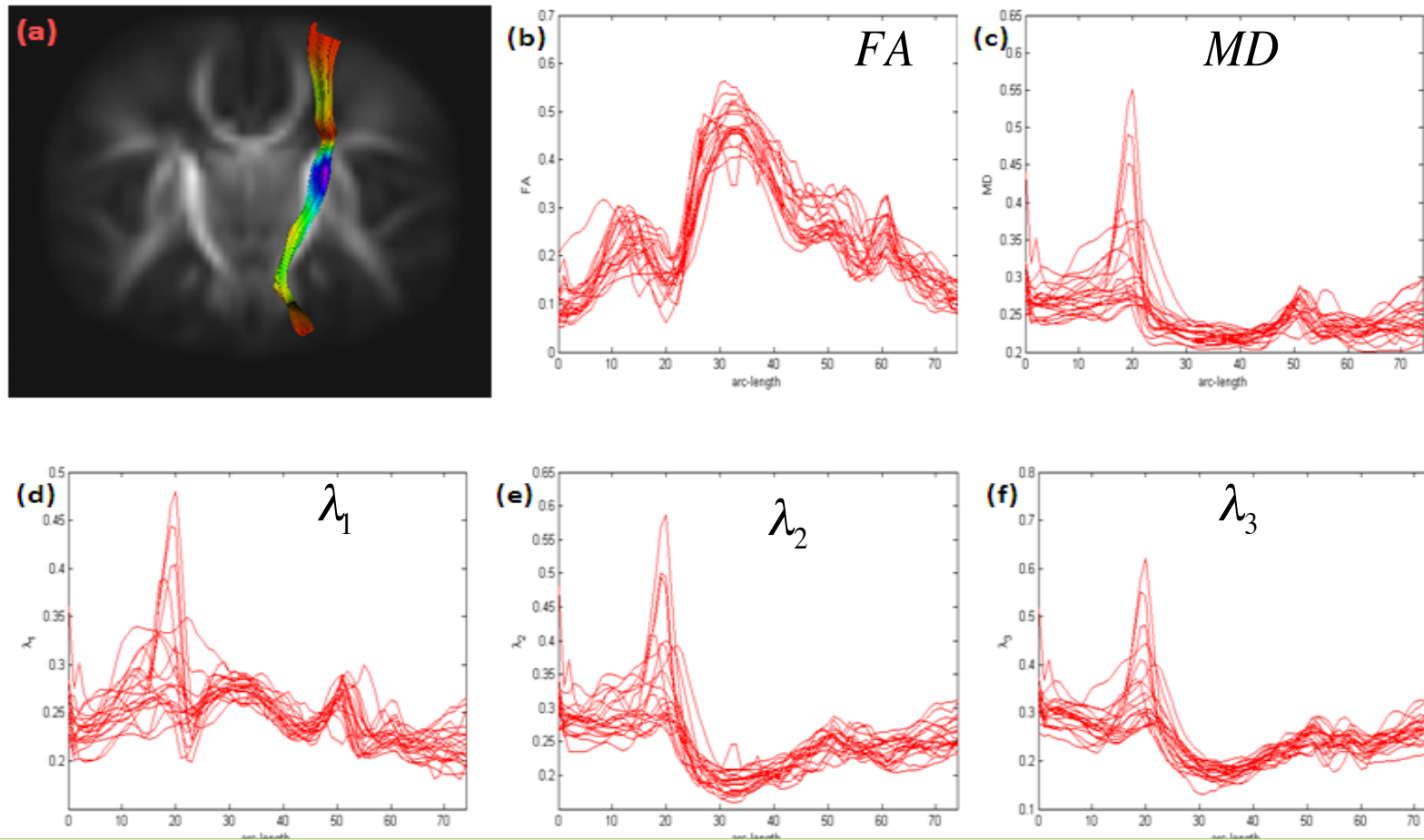— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

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
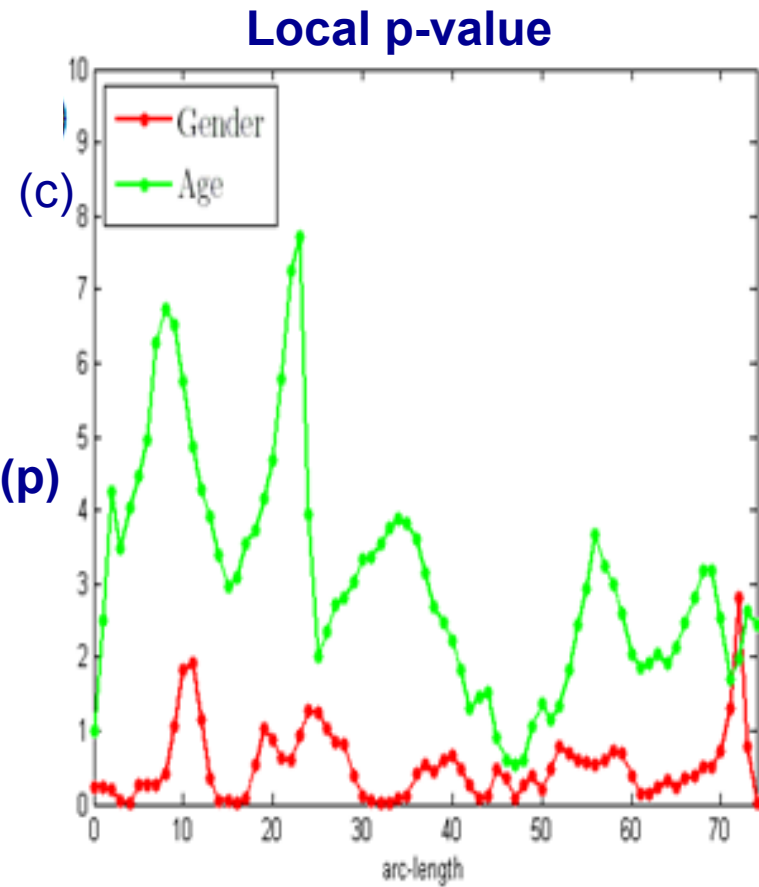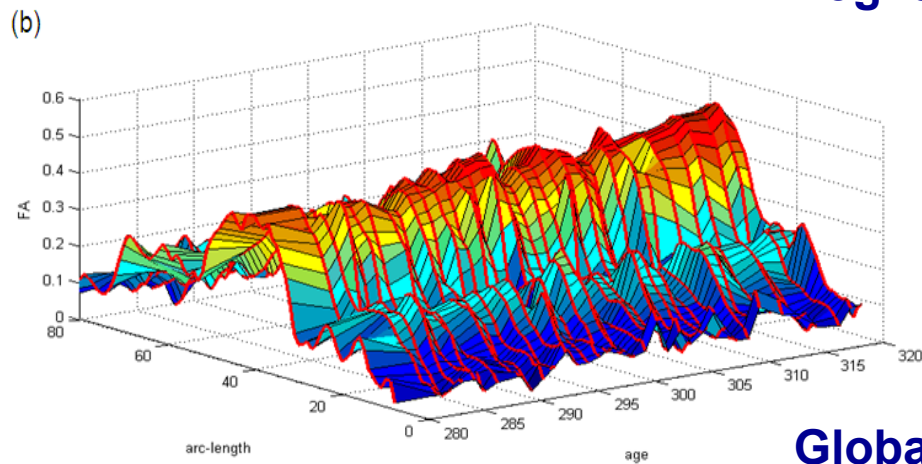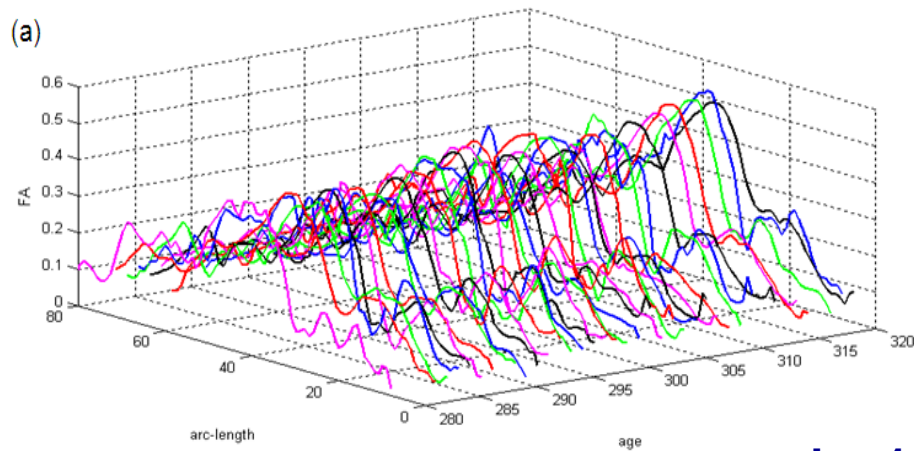- 5 diffusive outcomes: FA, MD,
- Internal Capsule

Real Data

Diffusion properties = Gender + Gestational age

Real Data

(a)

(b)

-log10(p)

Local p-value

(c)

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

Confidence Bands
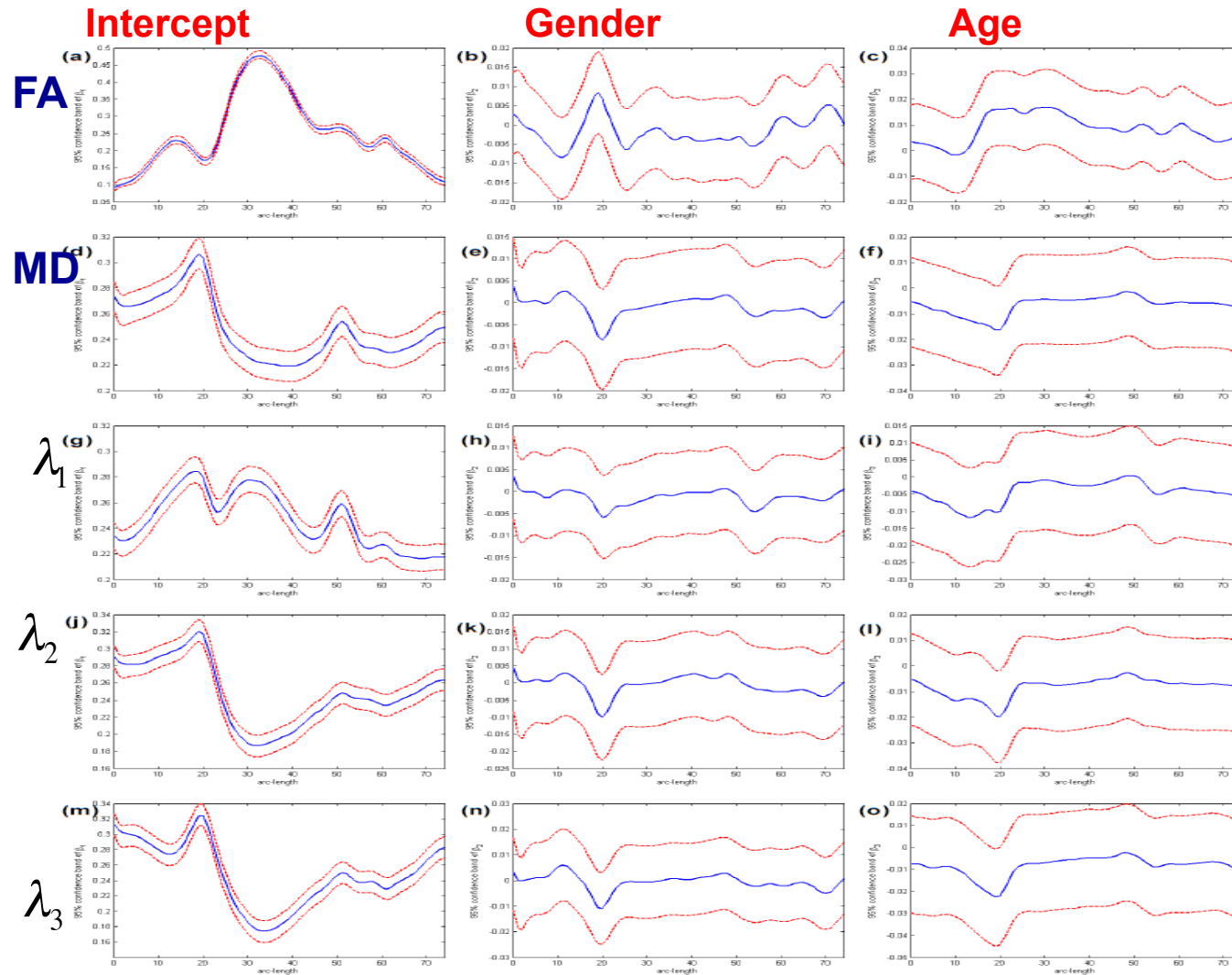
FPCA
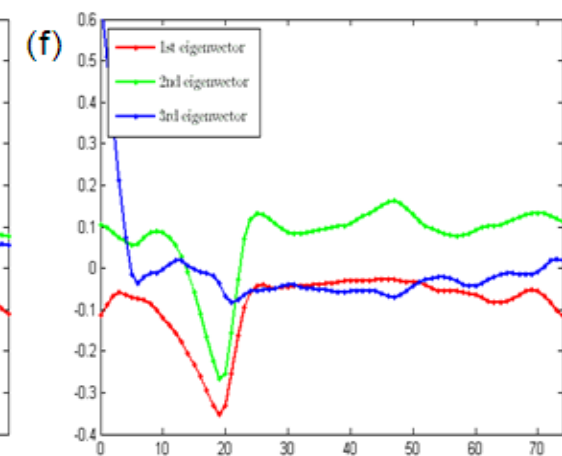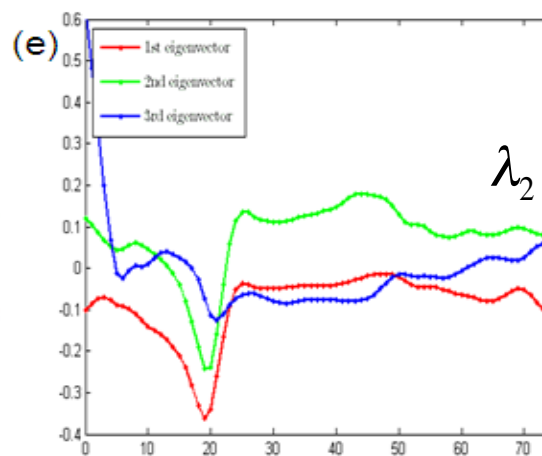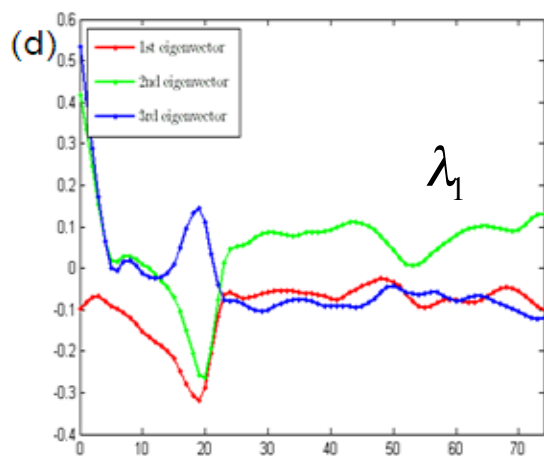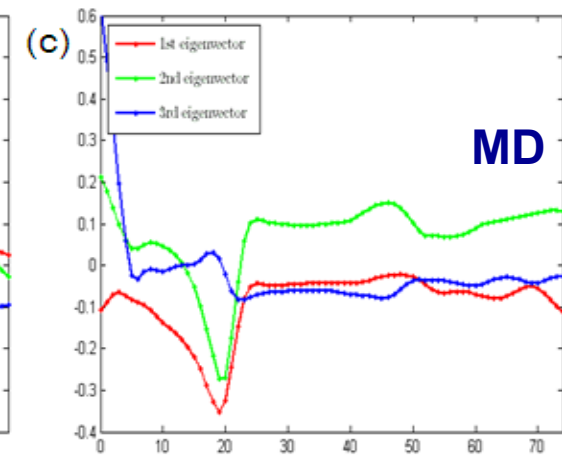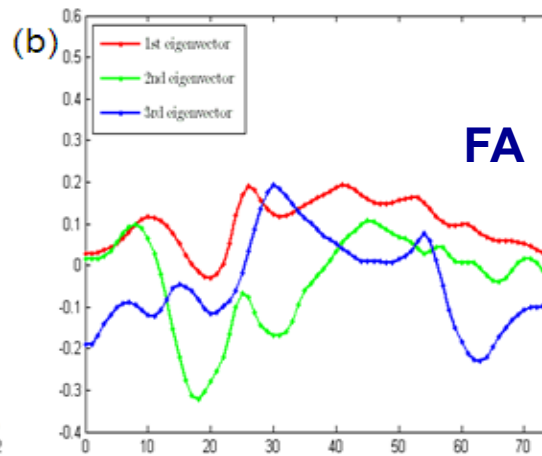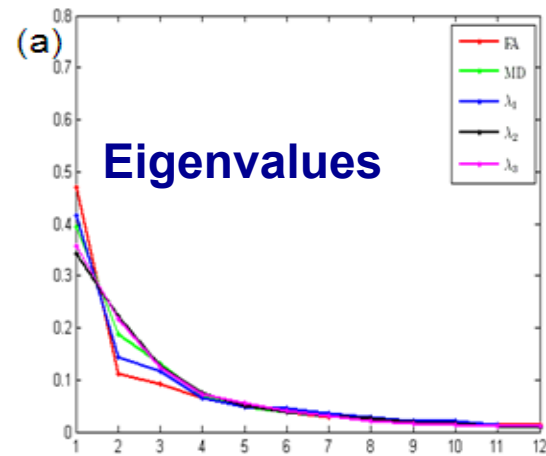
Longitudinal Extensions

Longitudinal Data

Spatial-temporal Process

$t \uparrow$

$y_i(s, t_3)$

$y_i(s, t_2)$

$y_i(s, t_1)$

$\longrightarrow s$
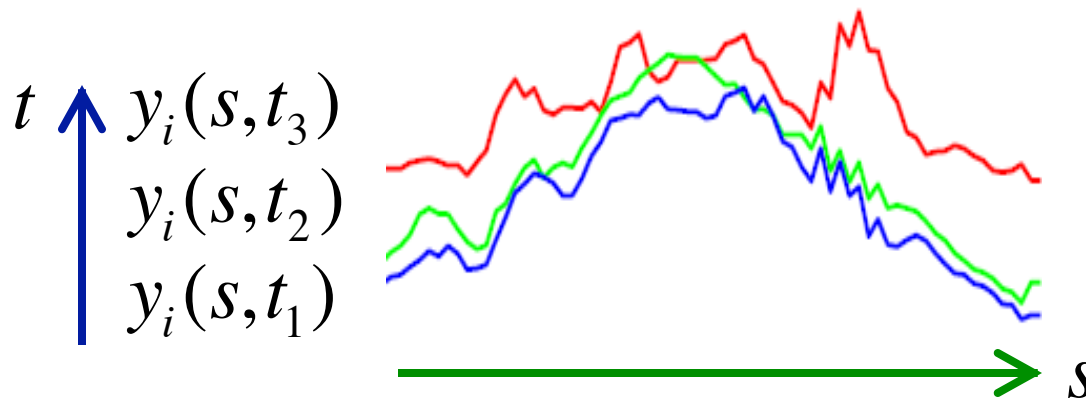
Functional Mixed Effect Models

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

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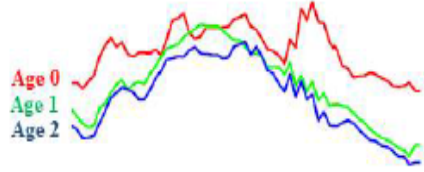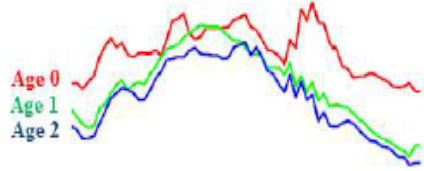
$$\eta_i(\bullet,\bullet) \sim SP(0,\Sigma_\eta), \quad \xi_i(\bullet) \sim SP(0,\Sigma_\xi) \qquad \varepsilon_i(\bullet) \sim SP(0,\Sigma_\varepsilon),$$

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

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

UNIVERSITY OF
ALBERTA

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

Comparison

	Data format	Estimation	Inference tool
Linear mixed effects models	Standard longitudinal data y_{ij}	Fixed effects	Test statistics
Guo (2002)'s method	One-time-measured curves $y_i(s)$	Covariance Fixed effect functions	Test statistics
Greven et al. (2010)'s method	Multiple-time-measured curves $y_{ij}(s)$	Random effect functions Fixed effect functions	
FMEM	Multiple-time-measured curves $y_{ij}(s)$	Covariance functions Fixed effect functions	Test statistics
		Covariance functions	

Real Data

genu

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
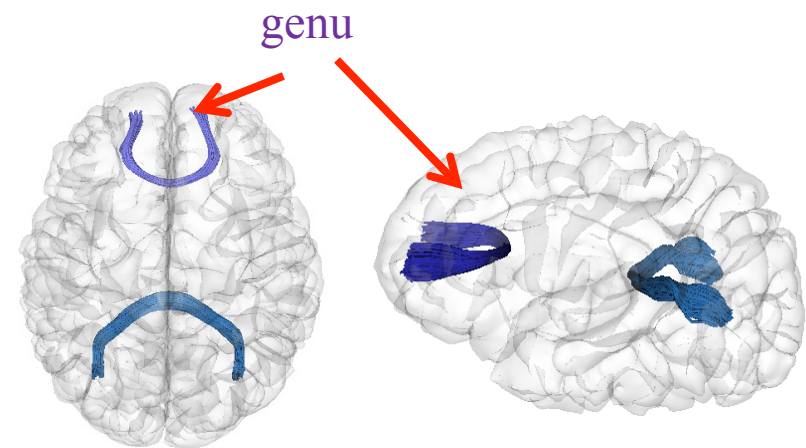
DTImaging parameters:

- **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
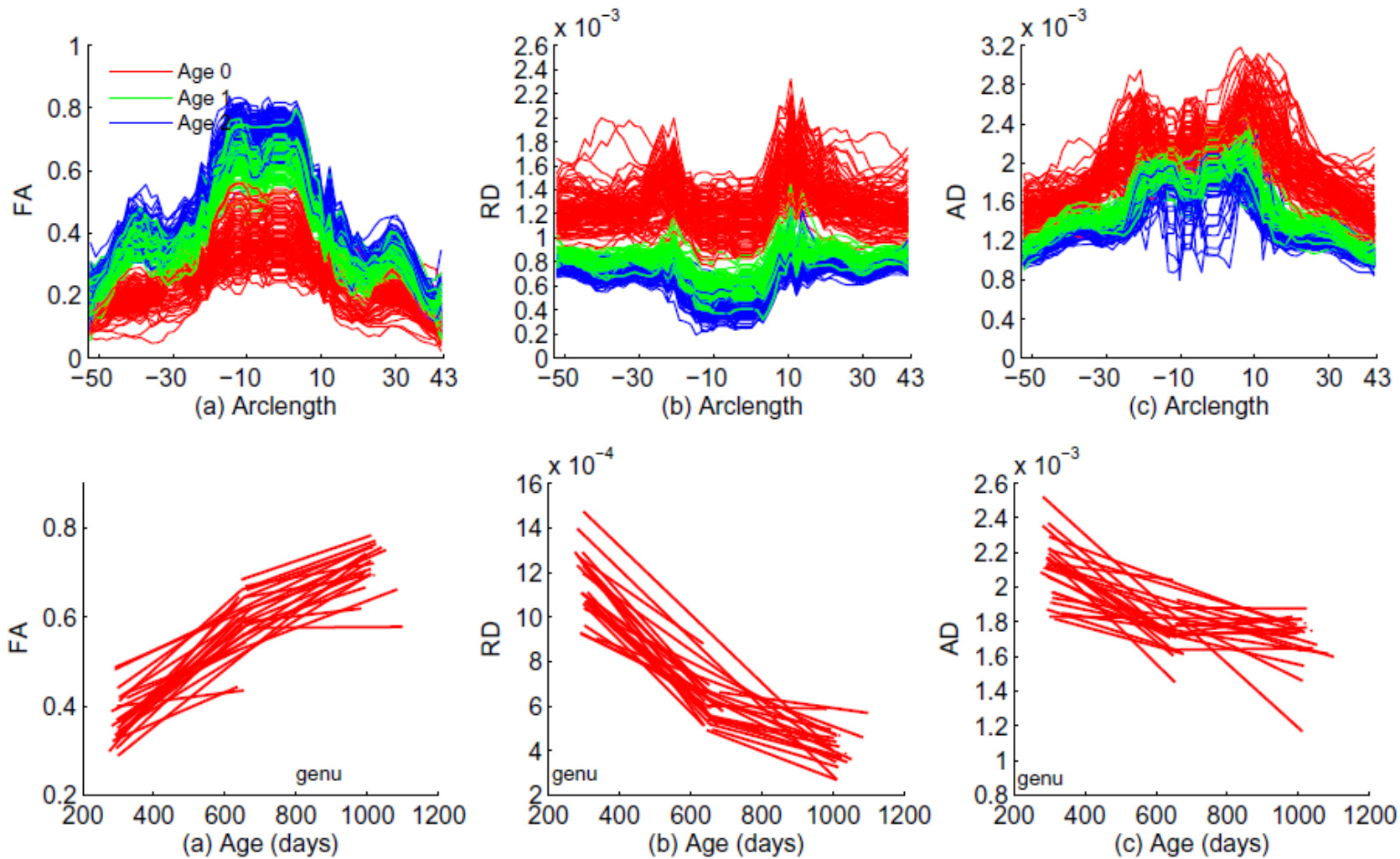
Real Data Analysis Results
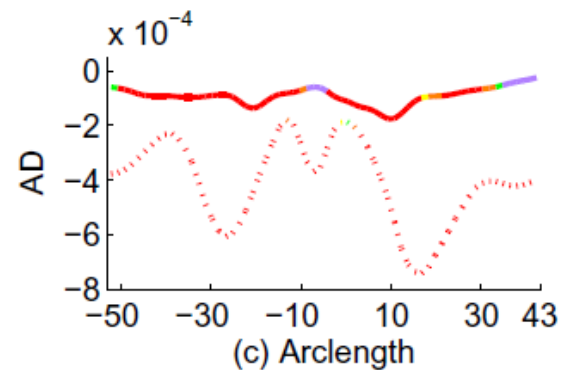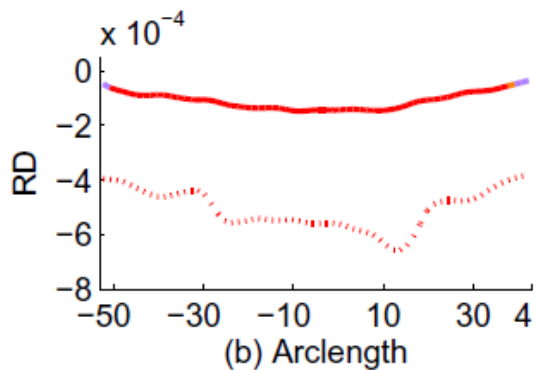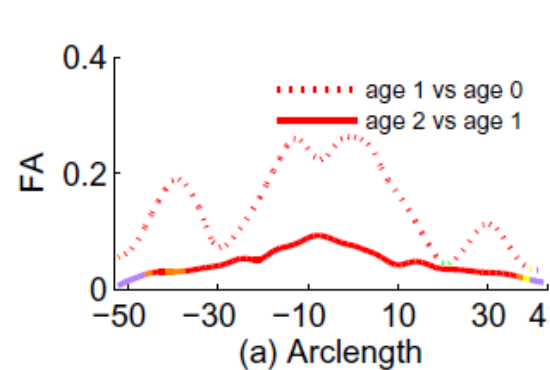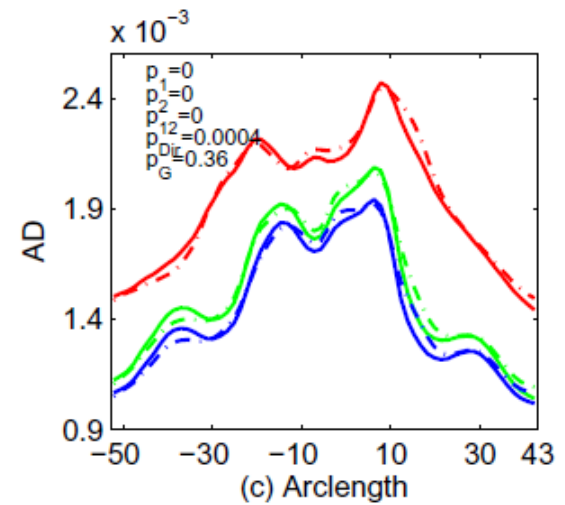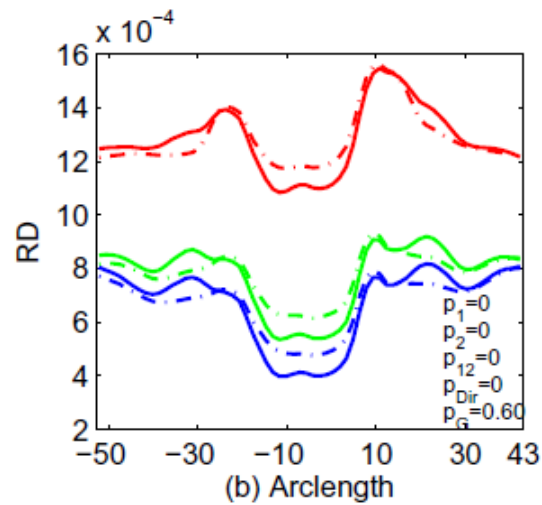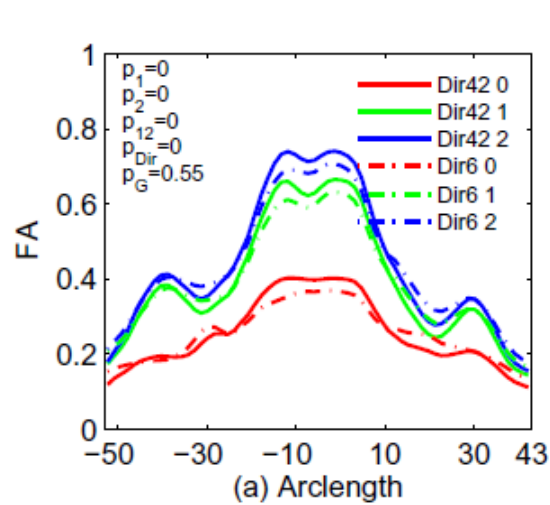
Real Data Analysis Results
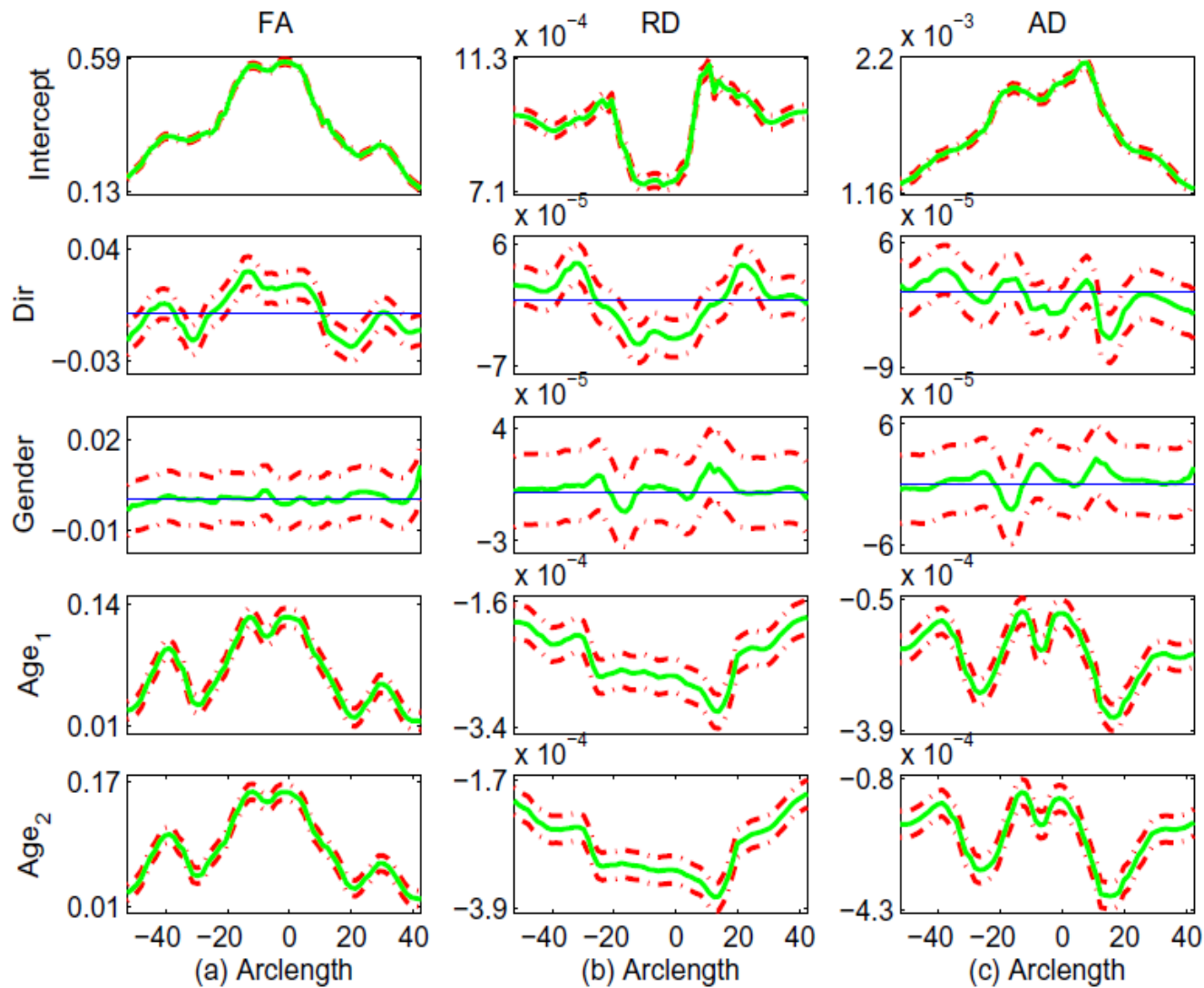
Functional Nonlinear Mixed Effects Model

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

UNIVERSITY OF ALBERTA

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

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:

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

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

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

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.

Real Data

Table 2. Demographic information for participants.

Visit	Number of subjects	Age(years)	Range(years)
1	58	10.53(5.96)	[0, 18]
2	148	12.25(4.62)	[0, 21]
3	160	12.29(5.14)	[1, 22]
4	19	1.84(1.42)	[1, 6]
5	7	1.57(0.79)	[1, 3]
6	10	2.70(0.67)	[2, 4]
7	6	3.17(0.75)	[2, 4]
8	5	3.40(1.14)	[2, 5]
9	3	3.67(1.15)	[3, 5]
Gender	Male/Female		126/127

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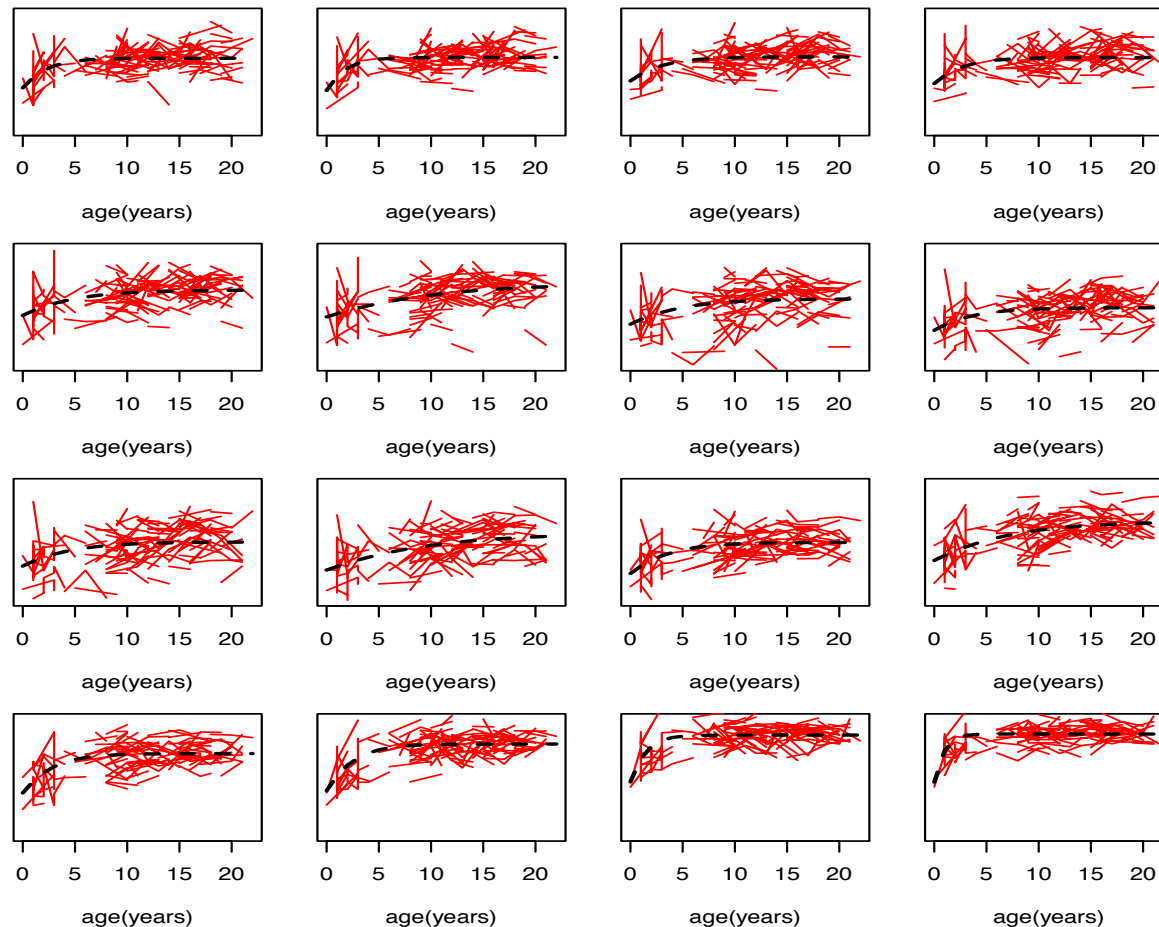
Tract (red solid lines) varying as a function of age for grid points from 25 to 40, the black dash curves are estimated curves.

UNIVERSITY OF ALBERTA

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

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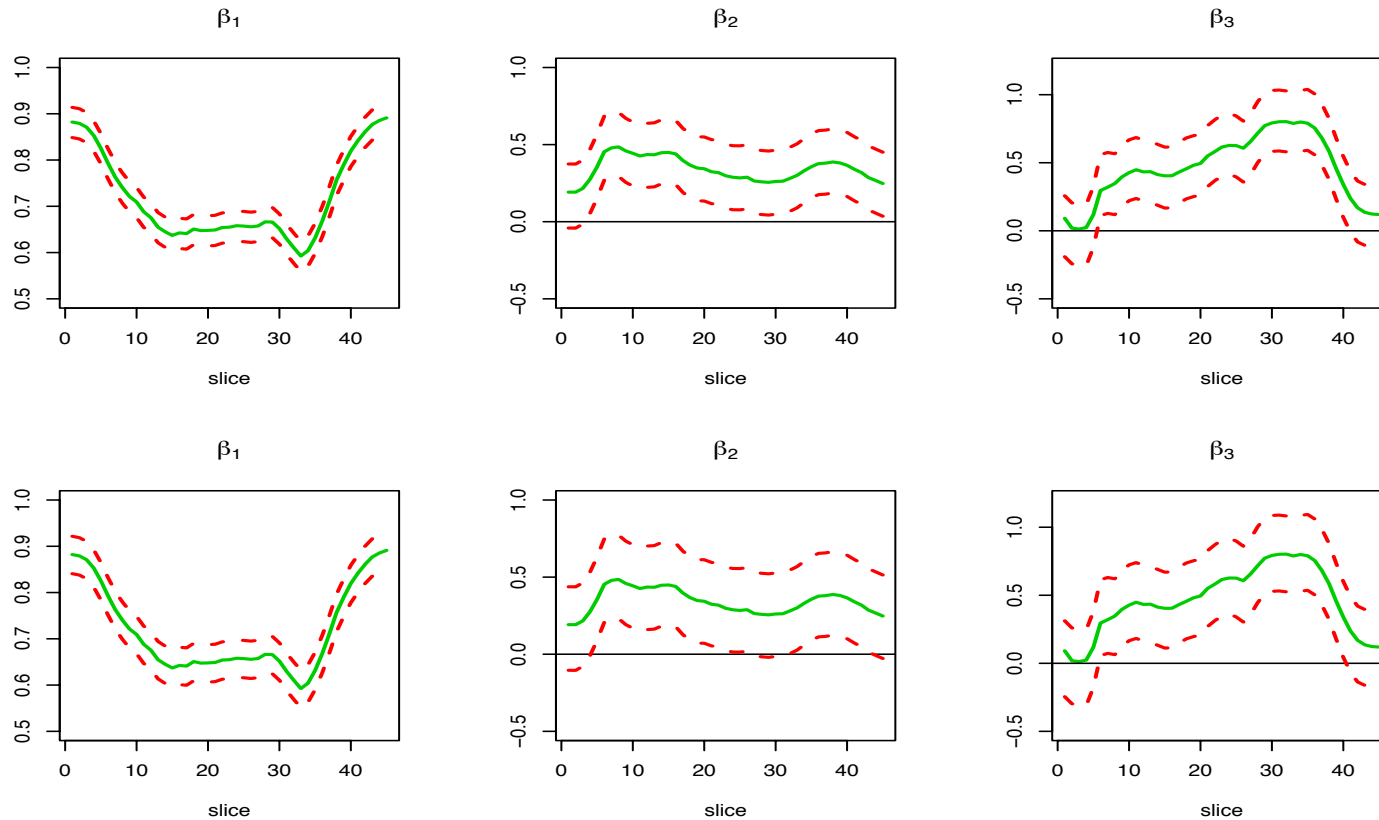
The 100(1 − α)% simultaneous confidence bands of parameters for α = 0.05 (the first row) and α = 0.01 (the second row). The green solid and red dash curves are, respectively, the estimated curves and their corresponding 95% and 99% simultaneous confidence bands.

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
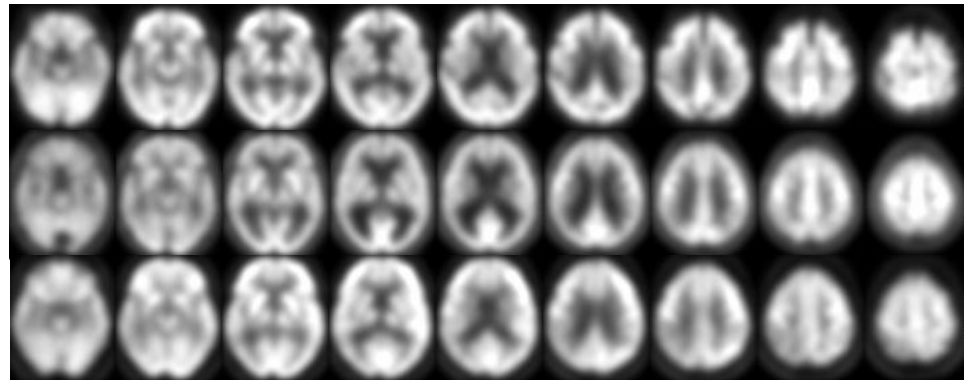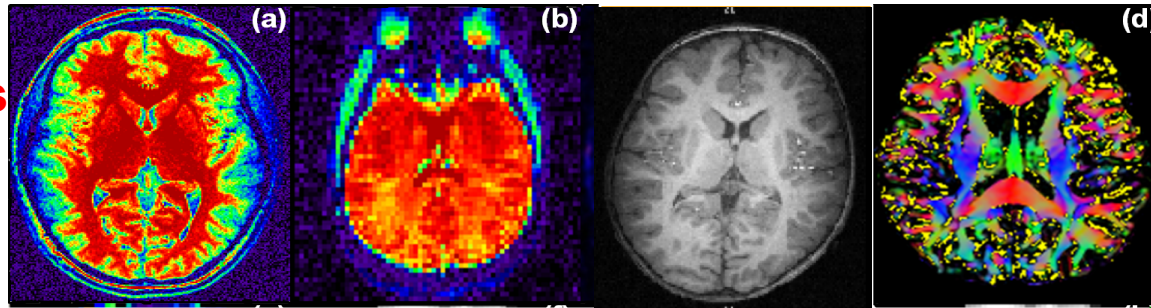
Across modalities

Local spatial smoothness
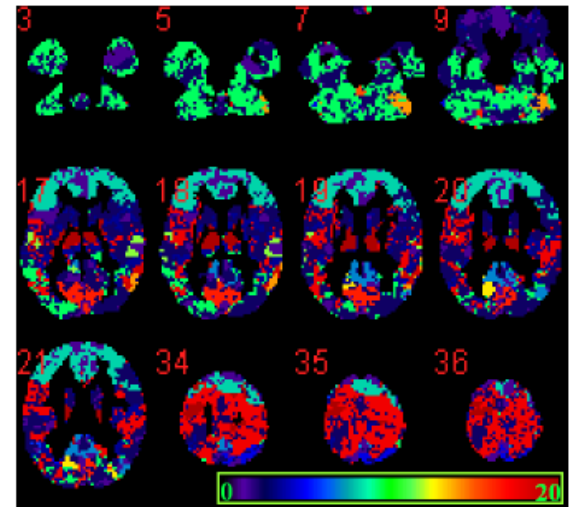
Across subjects

SGPP Model

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

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

UNIVERSITY OF ALBERTA

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

fPCA+SAR

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

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

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) $\hat{\beta}_{33}(d)$.
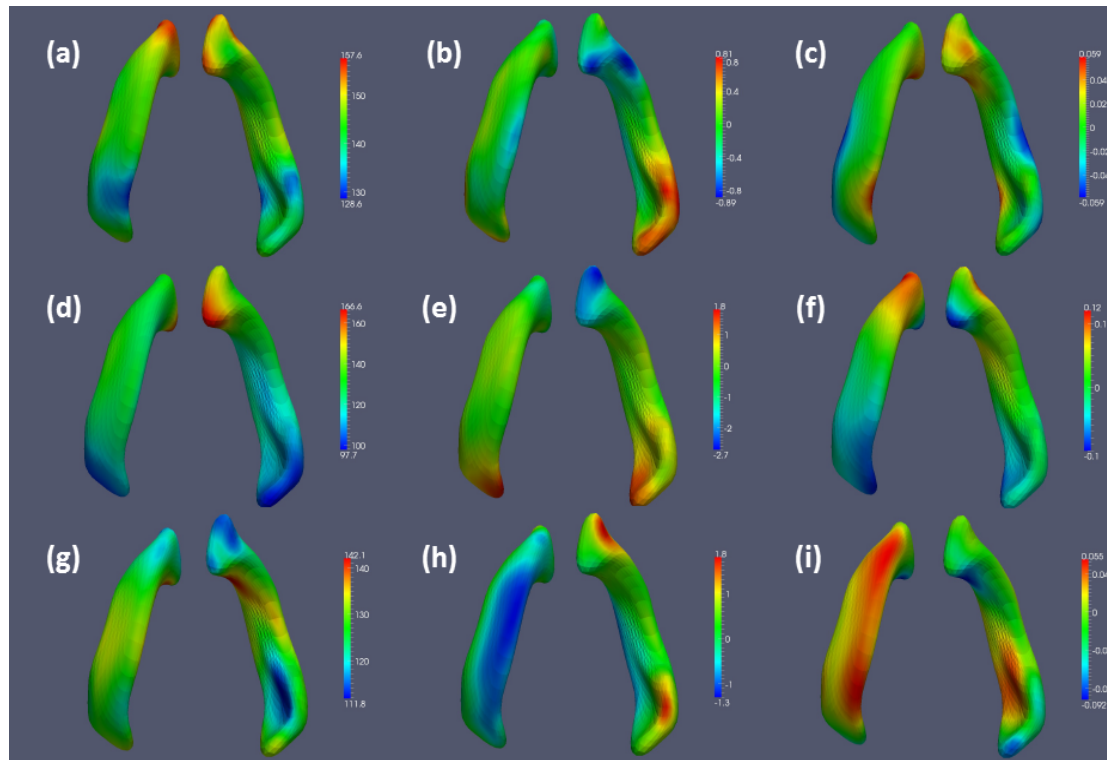
Ventricle Surface Data
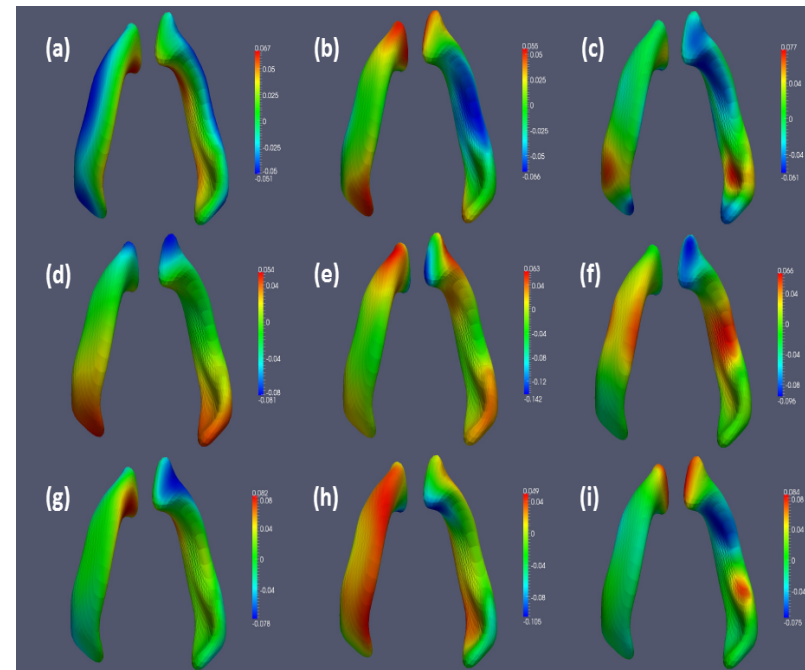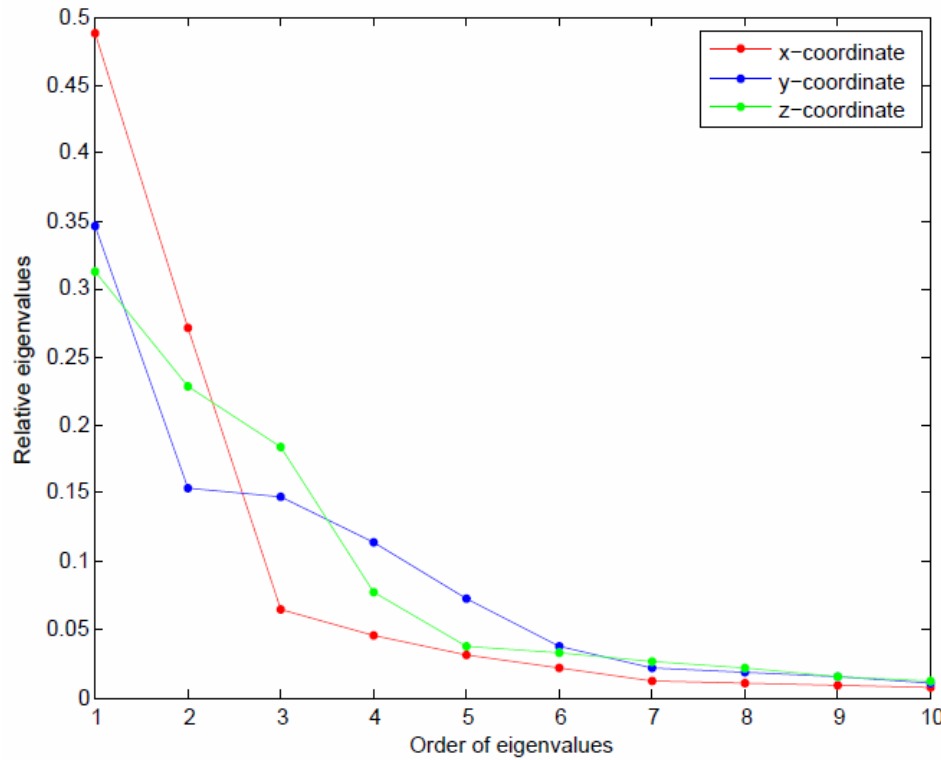
Figure : Results from the surface data of the left lateral ventricle: (a) $\hat{\psi}_{1,1}(d)$; (b) $\hat{\psi}_{1,2}(d)$; (c) $\hat{\psi}_{1,3}(d)$; (d) $\hat{\psi}_{2,1}(d)$; (e) $\hat{\psi}_{2,2}(d)$; (f) $\hat{\psi}_{2,3}(d)$; (g) $\hat{\psi}_{3,1}(d)$; (h) $\hat{\psi}_{3,2}(d)$; and (i) $\hat{\psi}_{3,3}(d)$.

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

Multiscale Adaptive Regression Models

Reading materials:

1. Zhu, HT., Fan, J., and Kong, L. (2014). Spatial varying coefficient model and its applications in neuroimaging data with jump discontinuity. *JASA,* in press.
2. Li, YM, John Gilmore, JA Lin, Shen DG, Martin, S., Weili Lin, and *Zhu, HT.* (2013). Multiscale adaptive generalized estimating equations for longitudinal neuroimaging data. *NeuroImage,* 72, 91-105.
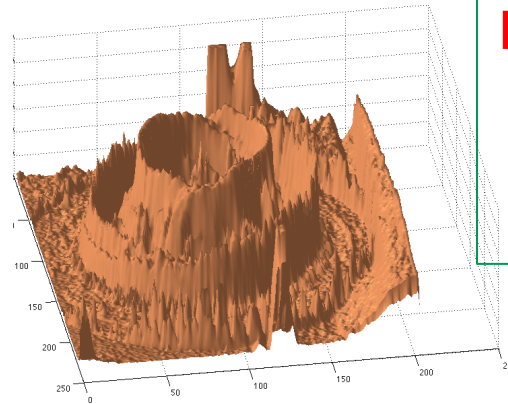3. Li, YM, John Gilmore, JP Wang, M. Styner, Weili Lin, and *Zhu, HT.* (2012). Two-stage spatial adaptive analysis of twin neuroimaging data. *IEEE Transactions on Medical Imaging.* 31, 1100-12.
4. Skup, M., Zhu, H.T., and Zhang HP. (2012). Multiscale adaptive marginal analysis of longitudinal neuroimaging data with time-varying covariates. *Biometrics*, 68(4):1083-1092.
5. Shi, XY, Ibrahim JG, Styner M., Yimei Li, and *Zhu, HT.* (2011). Two-stage adjusted exponential tilted empirical likelihood for neuroimaging data. *Annals of Applied Statistics*, 5, 1132-1158.
6. Li, YM, *Zhu HT,* Shen DG, Lin WL, Gilmore J, and Ibrahim JG. (2011). Multiscale adaptive regression models for neuroimaging data. *JRSS, Series B*, 73, 559-578.

7. Polzehl, Jörg; Voss, Henning U.; Tabelow, Karsten. Structural adaptive segmentation for statistical parametric mapping. *NeuroImage, 52 (2010) pp. 515--523.*
8. Polzehl, J. and Spokoiny, V. G. (2006). Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135, 335-362.
9. J. Polzehl, V. Spokoiny, (2000) *Adaptive Weights Smoothing with applications to image restoration*, J. R. Stat. Soc. Ser. B Stat. Methodol., 62 pp. 335--354.

UNIVERSITY OF ALBERTA

"uplifting the whole people"

— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Piecewise Smooth Data

Mathematics.

Noisy Piecewise Smooth Functions with Unknown Jumps and Edges

Image *is the point or set of points in the range corresponding to a designated point in the domain of a given function.*

▲ Ω is a compact set. $\quad \tilde{x} \in \Omega \subseteq R^{k}$

➡ $f(\tilde{x}) \in M \subseteq R^{m} \qquad f : \Omega \to M \subseteq R^{m}$

⭐ $\int_{\Omega} \| f(\tilde{x}) \|^{k} \, d\tilde{x} \ < \infty \ $ for some k>0

Neuroimaging Data with Discontinuity

Noisy Piecewise Smooth Function with Unknown Jumps and Edges
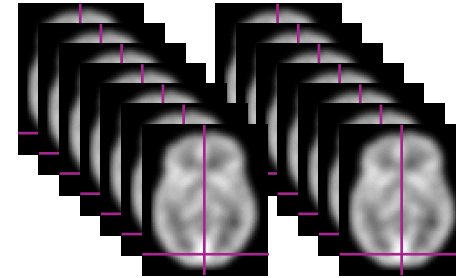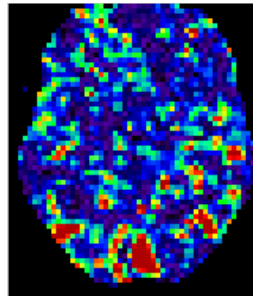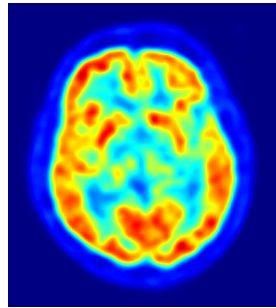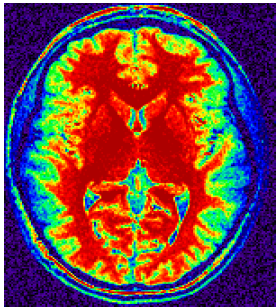
Subject1   Subject2

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

3D volume/ 2D surface

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

Covariance operator:

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

SVCM

Cartoon Model

$$B_k(d)$$

- **Disjoint Partition** $D = \cup_{l=1}^{L} D_l$ and $D_l \cap D_{l'} = \phi$

- **Piecewise Smoothness: Lipschitz condition**

- **Smoothed Boundary**

- **Local Patch**

- **Degree of Jumps**
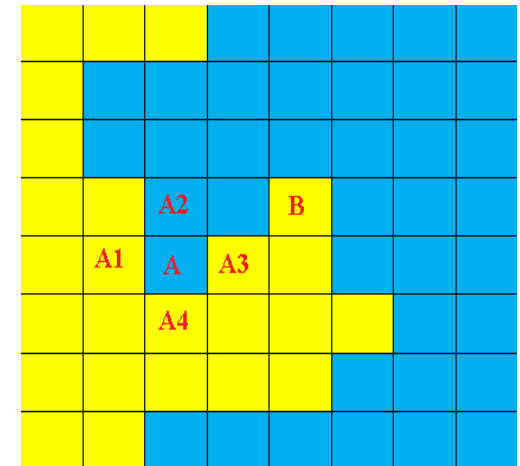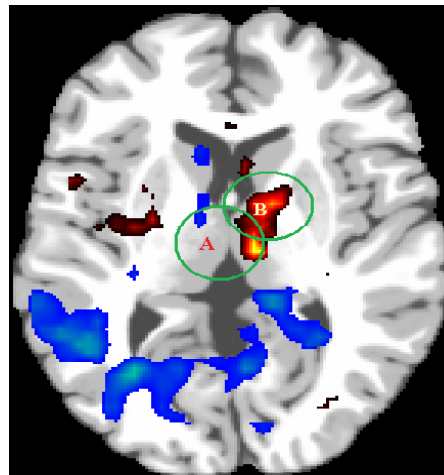
Kernel-based Smoothing Methods

Observed image y = Underlying scene f + Noise ε
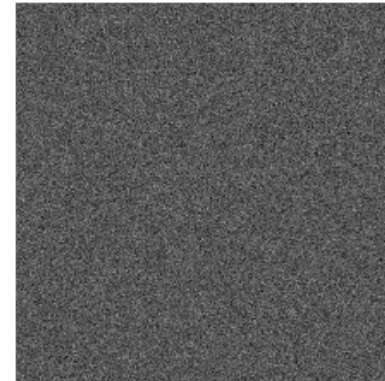
$$y = f + \varepsilon; \qquad \varepsilon \text{ uncorrelated, mean}=0, \text{ var}=\sigma^2$$

Estimate f_i as a weighted average of the noisy pixels:

$$\widehat{f_i} = \sum_j w_{i,j} y_j$$

Arias-Casto, Salmon, Willett (2011)
- **Local constant/linear**
- **Yaroslavsky/Bilateral Filter**
- **Nonlocal Means**
- **PS**

Smoothing Methods

Propogation-Seperation Method
J. Polzehl and V. Spokoiny, (2000,2005)

Original | Noisy image sigma=0.4 | nonadaptive kernel smoothing

Reconstruction local constant PS | Reconstruction local quadratic PS | Maximum Overlap DWT
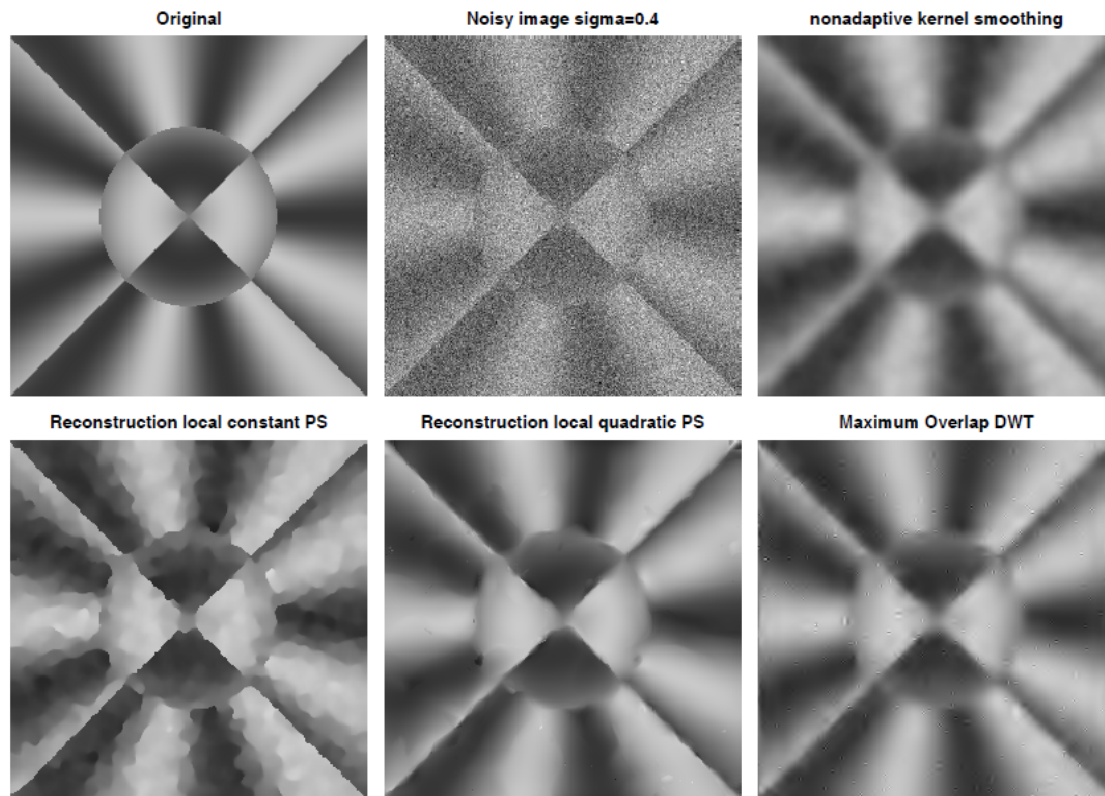
Features

- **Increasing Bandwidth**
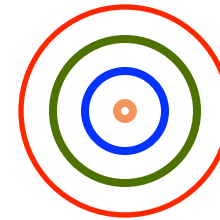
$$0 < h_0 < h_1 < \cdots < h_S = r_0$$

- **Adaptive Weights**

- **Adaptive Estimates**

Smoothing Methods

MARM

At each voxel d
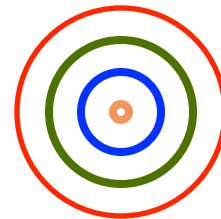
- **Increasing Bandwidth**

- **Adaptive Weights**

- **Adaptive Estimates**

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

$\omega(d,d';h_1)$

$\omega(d,d';h_2)$

$\omega(d,d';h_2)$

$\hat{\mu}(d;h_0)$ $\hat{\mu}(d;h_1)$

$\hat{\mu}(d;h_S)$
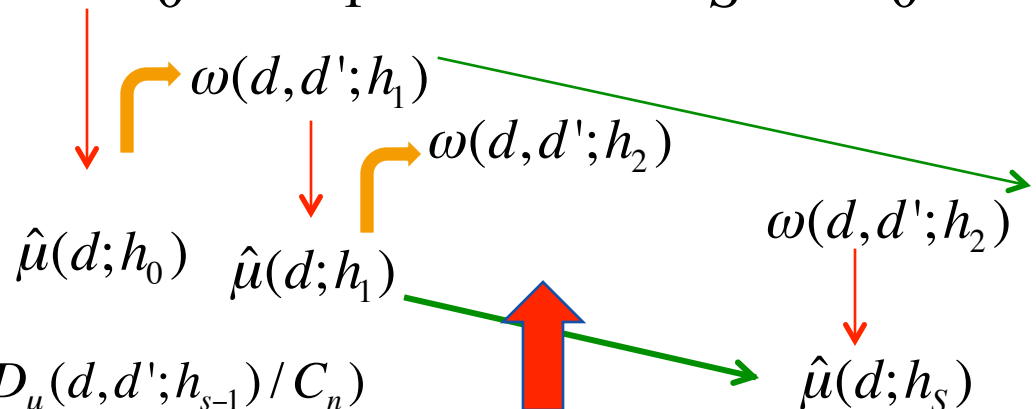
Stopping Rule

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

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

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

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

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**
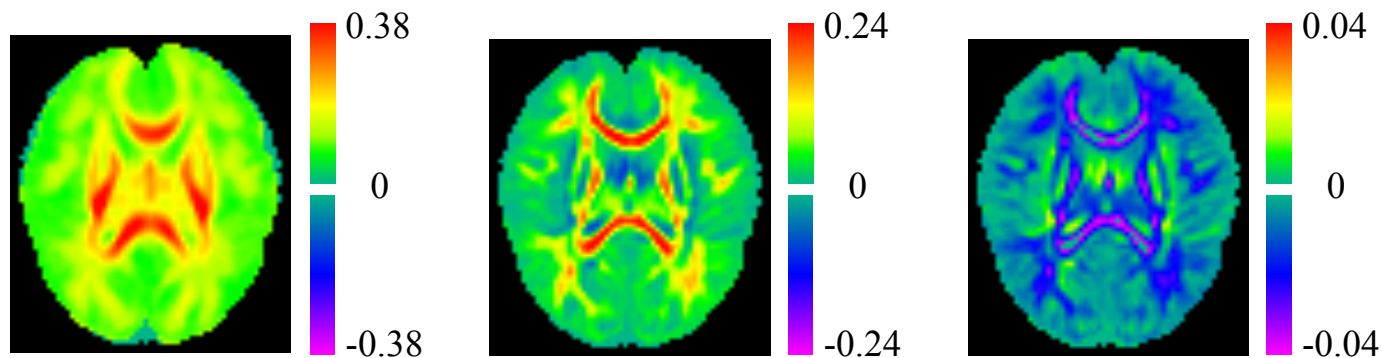
Real Data

- 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

Eigenimages
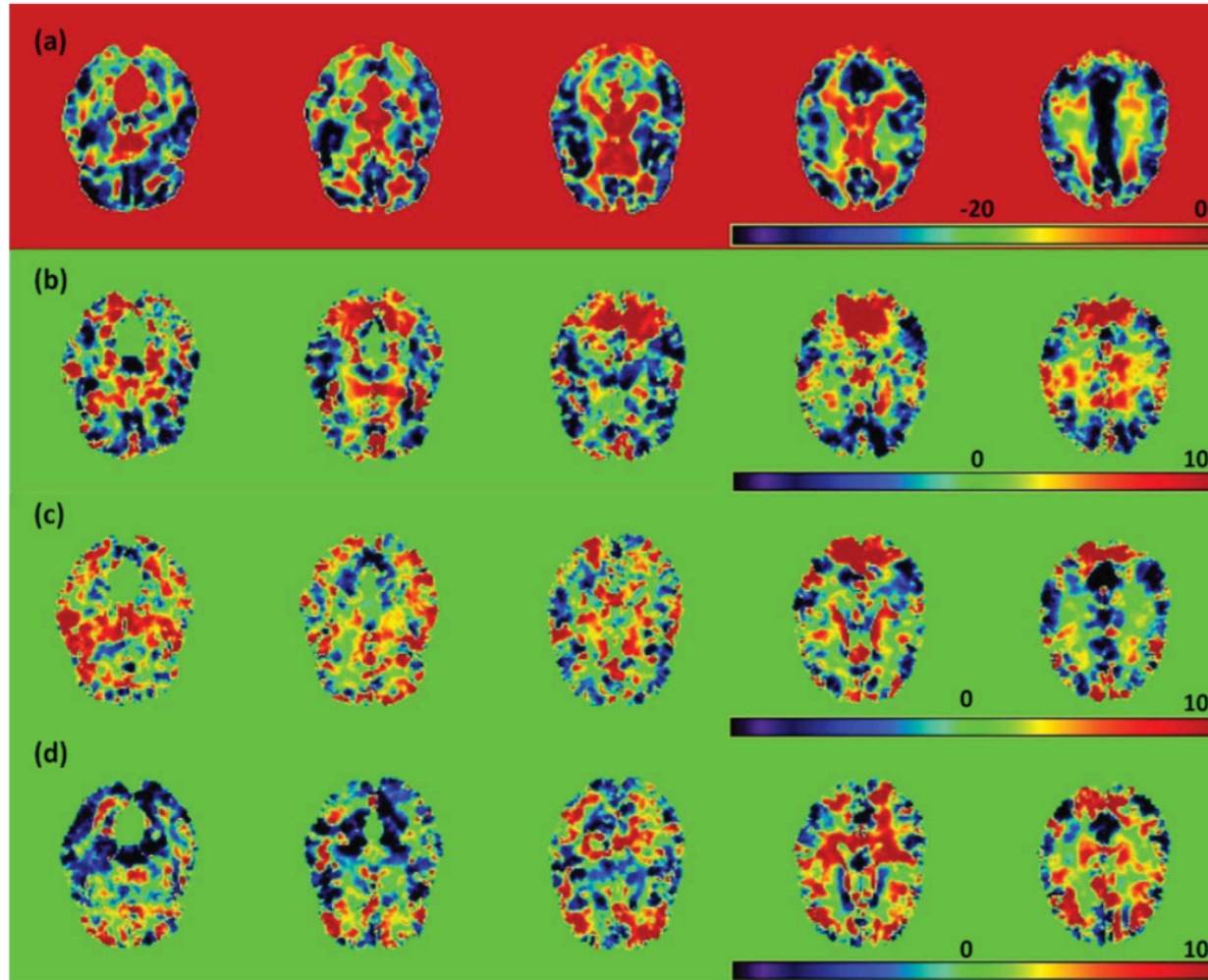
Real Data
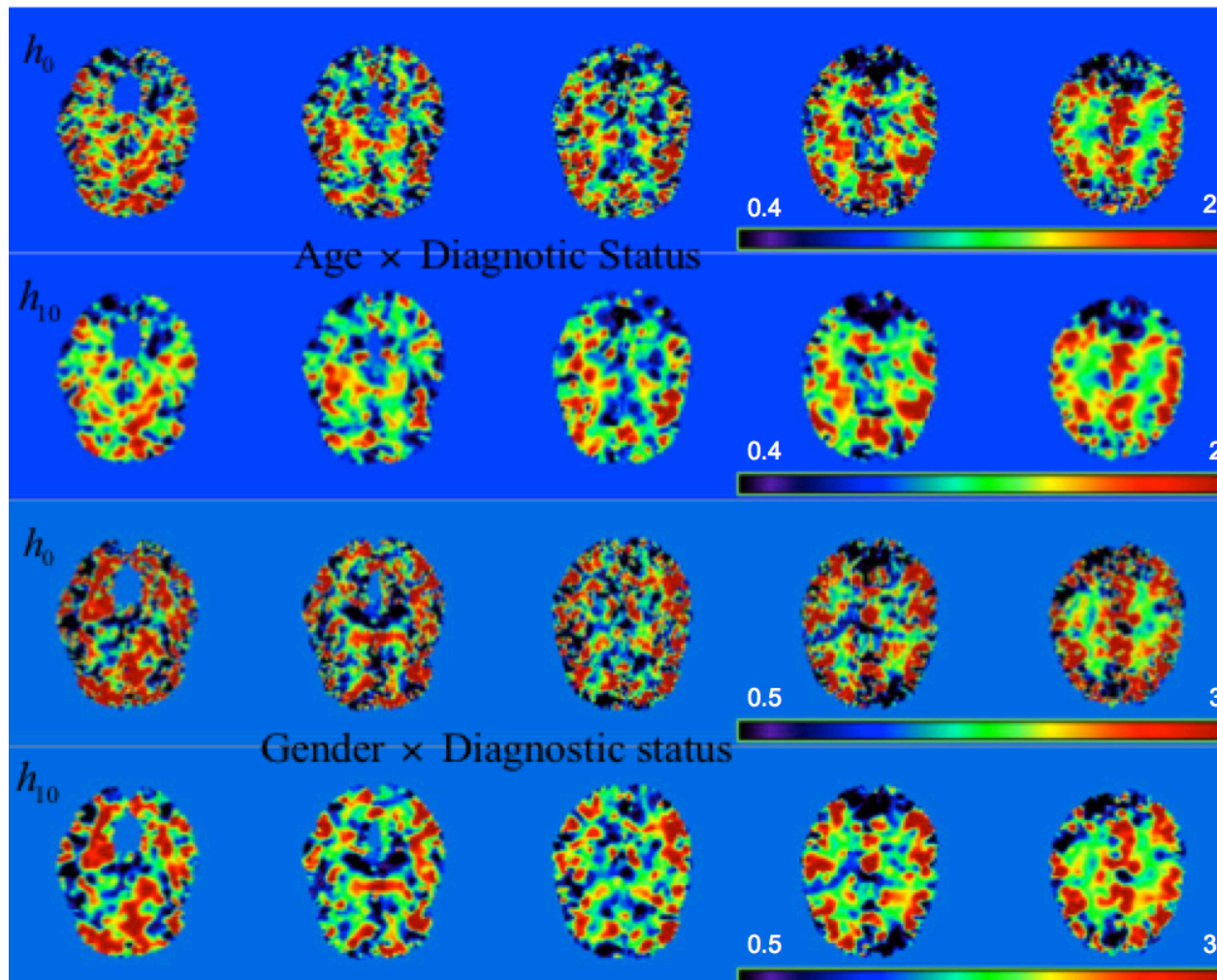
Significant regions overlayed on template

Age × Diagnotic Status
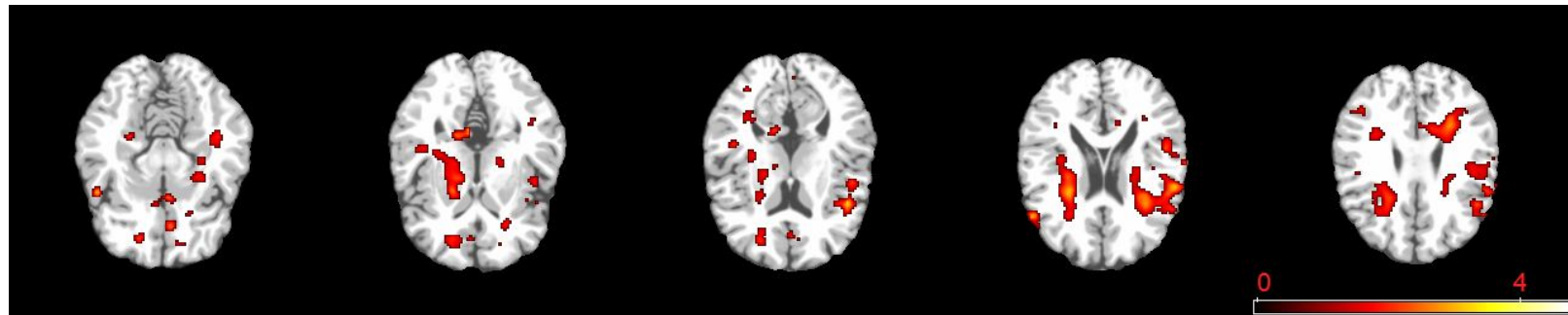
Gender × Diagnostic status
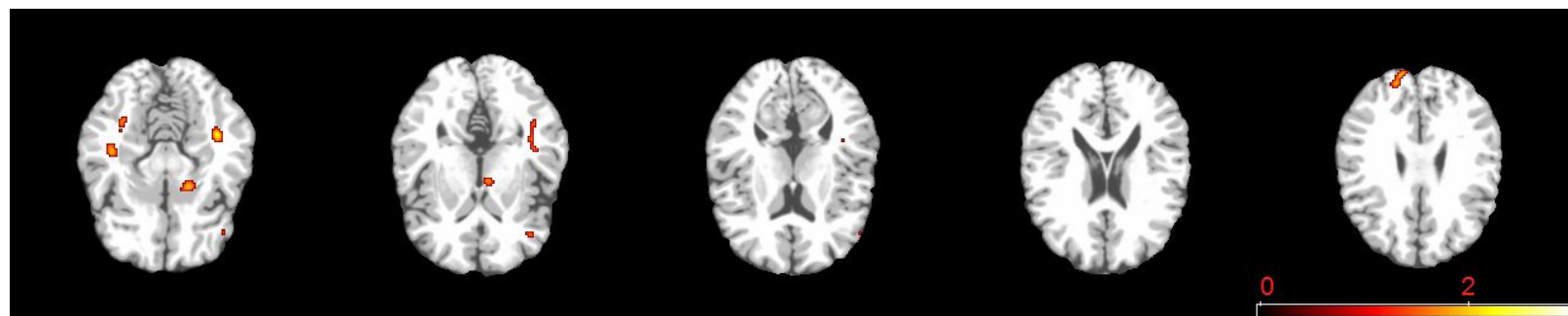
Significant regions

Table 3: The first two largest significant regions of the first three largest significant blocks for hypothesis tests $H_0 : \beta_6(d) = 0$ and $H_0 : \beta_7(d) = 0$ with block and region voxel sizes. WM, L and R, respectively represent white matter, left and right.

	block	size	1st largest ROI		2nd largest ROI	
			ROI label	size	ROI label	size
A × D	1	3954	frontal lobe WM L	1567	frontal lobe WM R	455
	2	2065	frontal lobe WM R	900	anterior limb of internal capsule R	220
	3	1642	nucleus accumbens L	1019	frontal lobe WM L	213
G × D	1	228	temporal lobe WM L	184	middle temporal gyrus L	22
	2	216	frontal lobe WM R	163	superior frontal gyrus L	33
	3	95	temporal lobe WM R	66	lateral occipitotemporal gyrus R	21

Scalar-on-Imaging Regression

Scalar-on-Image Models

Reading materials:

1. Miranda, M. F., <u>Zhu, H.T.</u> and Ibrahim, J. G. (2014). Bayesian partial supervised tensor decomposition with applications in neuorimaging data analysis.
2. Wang, X. and <u>Zhu, H.T.</u> (2014) Scalar-to-Image Regression via Total Variation.
3. Zhong, C., Liu, Y., Wang, J. H., and *Zhu, H. T.* (2014). Reunforced angled multicategory support vector machines.
4. Yang, H., <u>Zhu, H. T.</u>, and Ibrahim, J. G. (2014). Multiscale projection model in RKHS.
5. <u>Zhu, H. T</u>. and Shen, D. (2014). Multiscale Weighted PCA for Imaging Prediction.
6. Guo, R.X., Ahye M., and <u>Zhu, H.</u> (2014). Spatially weighted PCA for imaging classification. *JCGS.* In press.
7. Zhou, H., Li, L., and <u>Zhu, H.</u> (2013). Tensor regression with applications in Neuorimaging data analysis. *JASA.* 108(502), 540-552.

8. Cuingnet, R., Glaunes, J. A., Chupin, M., Benali, H., Colliot, O., and ADNI. (2012). Spatial and anatomical regularization of SVM: a general framework for neuroimaging data. IEEE PAMI. In press.
9. Cai, T. & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. JASA. 107, 1201-1216.
10. Fan, J. and Lv, J. (2010). A selective overview of variable selection in high-dimensional feature space. Statistica Sinica, 20, 101-148.
11. Li, B., Kim, M. K., and Altman, N. (2010). On dimension folding of matrix or array valued statistical objects. Ann. Stat., 38, 1097-1121.
12. Reiss, P. T., and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. Journal of the American Statistical Association 102, 984–996.
13. Bair, E., Hastie, T., Paul, D. and Tibshirani, R. (2006). Prediction by supervised principal components. JASA. 101, 119-137.
14. Ramsay, J.O. and Silverman, B.W. (2005). Functional Data Analysis, 2nd Edition. Springer, New York.
15. James, G. (2002). Generalized linear models with functional predictor variables. JRSSB. 64, 411-432.
16. Tibshirani, Robert (1996). Regression shrinkage and selection via the lasso. JRSSB. 58, 267-288.

UNIVERSITY OF ALBERTA

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

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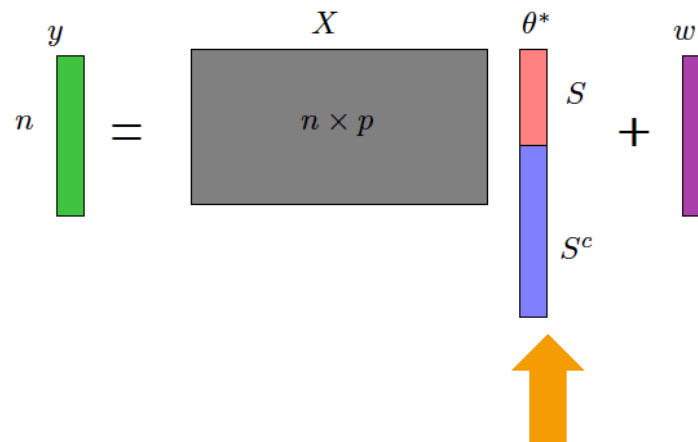
Zhou, Li, and Zhu (2013)
Li, Zhou, and Li (2013)

CP decomposition

θ
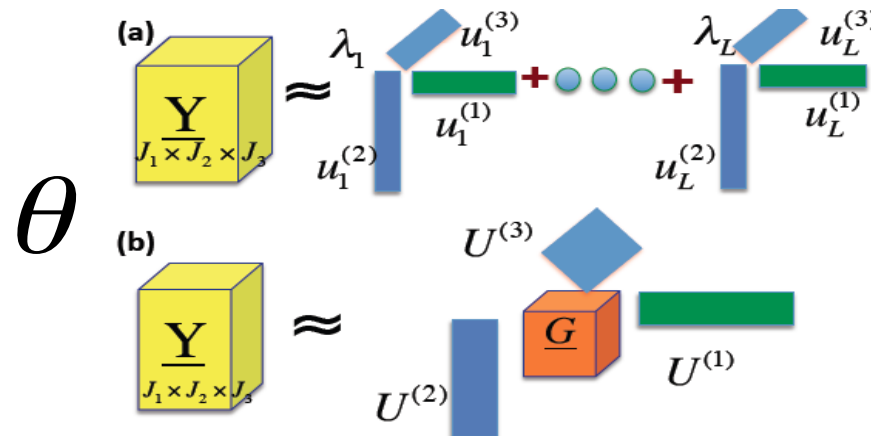
Tucker decomposition

Scalar-on-Image Models

Simulations
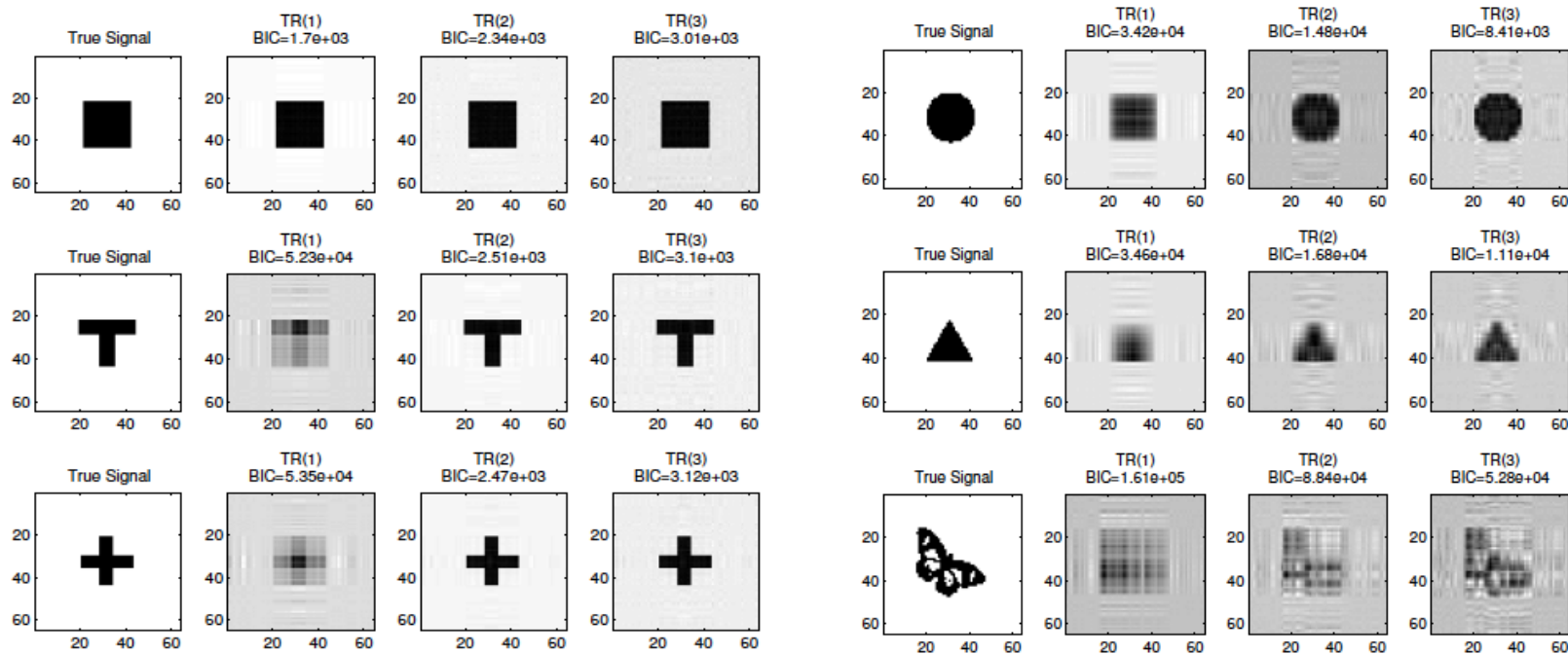
Key Conditions:

- Tensor Approximation B
- Restricted space property for X and B

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

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

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** $K^{1/2}CK^{1/2}$

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

Extensions

- **Functional linear Cox regression models**

- **Generalized scalar-to-image regression models**

- **Multiscale Functional Linear models**

Functional Linear Cox Regression Model

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

$$y_i = \min(T_i, C_i) \quad T_i : \text{ failure time}; C_i : \text{censored time}$$

Model

☀ $\quad h(t) = f(t)/S(t) = h_0(t)\exp(z_i^T \gamma + \int_S X_i(s)\beta(s)ds)$

☀ $\quad X_i(s) = \mu(s) + \sum_{j=1}^{\infty} \xi_{ij}\phi_j(s) + \varepsilon_i(s)$

- **Consistency**

- **Asymptotic distribution of score test**

Functional Linear Cox Regression Model

Mild Cognitive Impairment subjects

Interested in predicting the timing of an MCI patient that converts to AD by integrating the imaging data, the clinical variables, and genetic covariates.

Full Model: AUC=0.96

Partial Model: AUC=0.82

Generalized scalar-to-image regression models

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

Model $y \sim$ exponential family(μ, ϕ)

$$g(\mu) = \theta_0^T Z + <X, \beta_0>$$

Total Variation

Estimation: $\displaystyle\sum_{i=1}^{n} \ell(y_i; \mu(X_i; \gamma, \beta(\bullet))) + \lambda \parallel \beta \parallel_{TV}$

Non-asymptotic Error Bound:

$$\mathcal{R}_{2n} = \left\{ \mathbb{E}^* \left(\left\langle X^{(n+1)}, \hat{\beta} - \beta_0 \right\rangle \right)^2 \right\}^{1/2},$$

Generalized scalar-to-image regression models

Triangle | Oval | T-shape | checkerboard
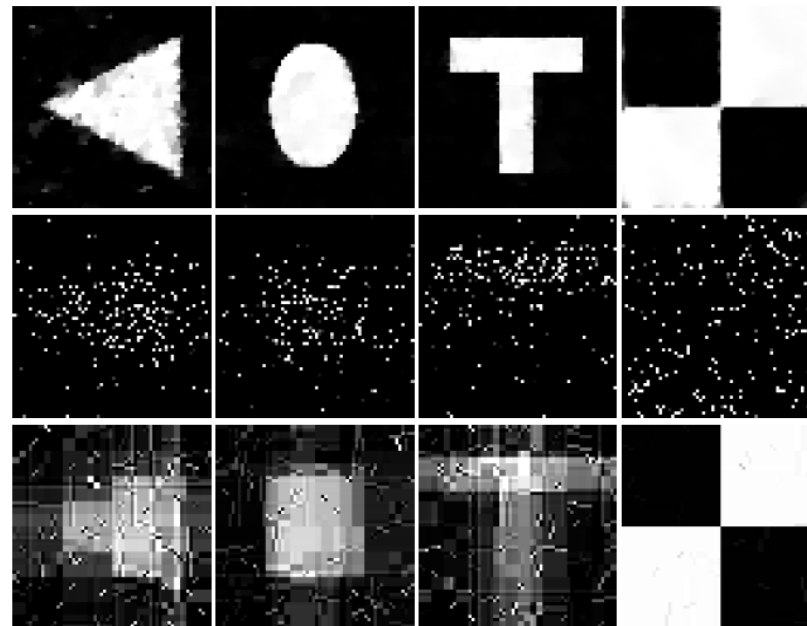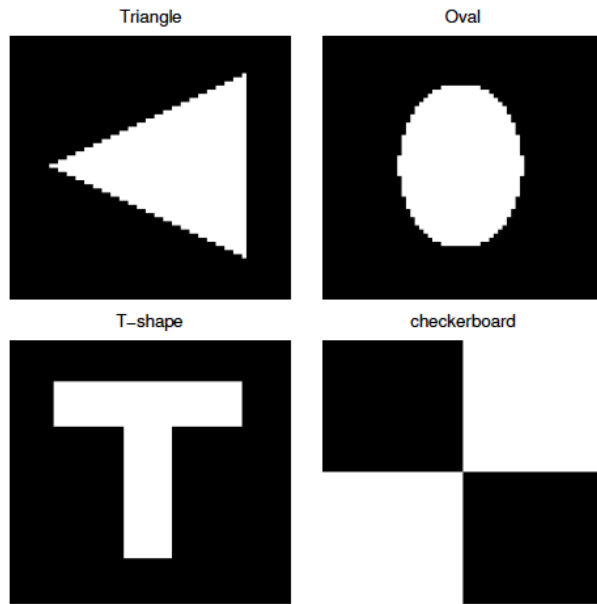
TV

Lasso

Lasso on WL

Multiscale Functional Linear models

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

Models

(A1) $D = (\bigcup_{k=1}^{K} D_k) \bigcup D_0$ • **Informative sets + Irrelevant set**

(A2) $y \perp \{X(d) : d \in D_0\}$

(A3) $y \sim p(\{X(d) : d \in D_1\}, \cdots, \{X(d) : d \in D_K\})$
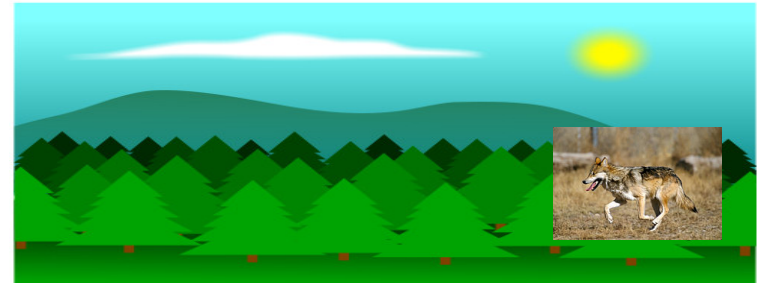
Rabbit and Wolf Story

$$y_i = 1(\text{death})$$

$$y_i = f_0(X_i) + \varepsilon_i$$

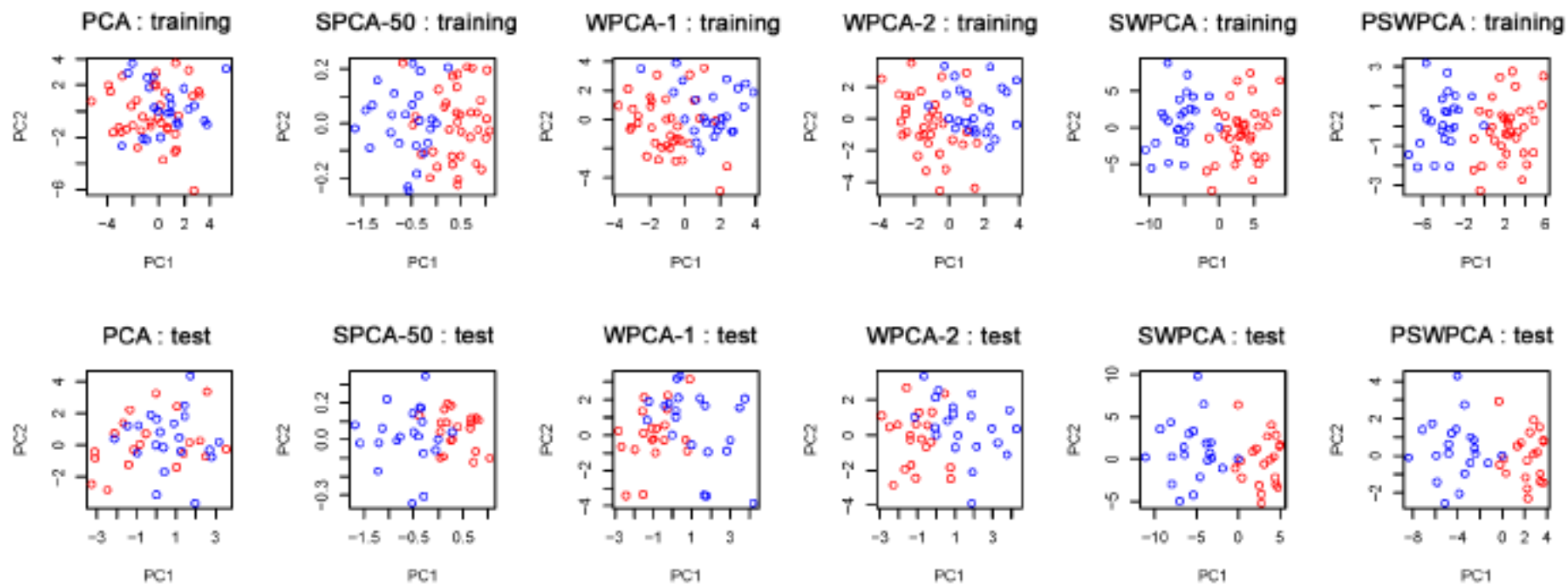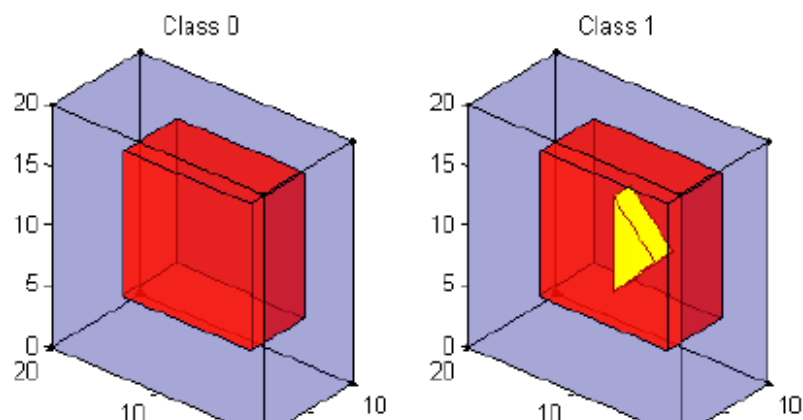$$X_i = \{X_i(d) : d \in D\}$$

$$y_i = f(G_i) + \varepsilon_i$$

$$G_{i,k} = G\{\tilde{X}_i(d) : d \in D_k\}$$

$$\tilde{X}_i = \{\tilde{X}_i(d) : d \in D_k\}_{k \geq 1}$$

Spatially Weighted PCA

Spatially Weighted PCA

Table 1: Average Misclassification Percentage for Simulation I

	PCA	SPCA					WPCA-1	WPCA-2	SWPCA	PSWPCA
	ALL	50	100	200	400	1000	ALL	ALL	ALL	ALL
REG	.302	.126	.132	.142	.162	.205	.199	.130	**.026**	**.025**
	(.078)	(.052)	(.052)	(.055)	(.057)	(.064)	(.064)	(.056)	(.025)	(.024)
k-NN	.338	.135	.141	.152	.182	.225	.186	.156	**.030**	**.027**
	(.071)	(.049)	(.049)	(.050)	(.053)	(.071)	(.055)	(.059)	(.029)	(.025)
SVM	.327	.140	.147	.159	.183	.226	.215	.152	**.033**	**.028**
	(.078)	(.054)	(.055)	(.055)	(.059)	(.072)	(.067)	(.055)	(.029)	(.026)

Standard deviations are in parenthesis. For SPCA, the number of "top" selected voxels used in the algorithm are considered to be 50, 100, 200, 400, and 1000.

Table 2: Average Misclassification Percentage for Simulation I (Non-PCA Methods)

SPLS-REG	SPLS-kNN	SPLS-SVM	SPLS	SDA
.130	.139	.156	.128	.120
(.052)	(.056)	(.066)	(.050)	(.050)

Standard deviations are in parenthesis.

Simulation I: Classification

Class 0 Class 1

0 White
1 Green
2 Red
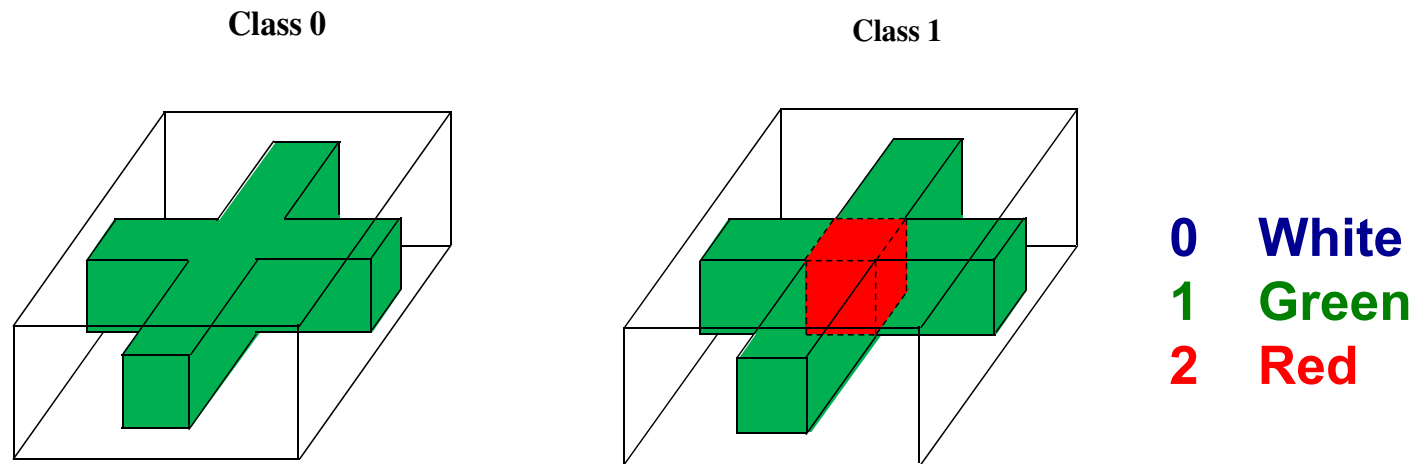
$$X_i(d) = \beta_0(d) + \beta_1(d)y_i + \varepsilon_i(d)$$

Type I **Type II** **Type III**

$N(0,4)$ Short-range correlation Long-range correlation

Simulation I: Classification

Table 1: Misclassification rates for PCA and SWPCA under the different number of PCs.

Noise	Number of PCs	PCA	SWPCA1	SWPCA2	SWPCA3
Type I	5	0.40	0.11	0.09	0.10
	7	0.40	0.13	0.11	0.10
	10	0.40	0.13	0.11	0.10
Type II	5	0.40	0.04	0.08	0.03
	7	0.39	0.03	0.09	0.04
	10	0.38	0.03	0.07	0.04
Type III	5	0.40	0.13	0.10	0.09
	7	0.41	0.13	0.10	0.10
	10	0.41	0.13	0.10	0.10

Simulation I: Classification

Noise	sLDA	sPLS	SLR	SVM	ROAD	PCA	SWPCA
Type I	0.28	0.43	0.45	0.38	0.36	0.36	0.10
Type II	0.27	0.08	0.18	0.26	0.08	0.45	0.03
Type III	0.52	0.30	0.61	0.60	0.50	0.35	0.09

sLDA: sparse discriminant analysis
sPLS: sparse partial least squares analysis
SLR: sparse logistic regression
SVM: support vector machine
ROAD:

ADNI

- Alzheimer's Disease (AD) is a neurodegenerative brain disorder which leads to progressive loss in memory and cognition, and eventual death

- A definitive diagnosis of AD can only be made postmortem after examining brain tissue for the presence of neuritic β-amyloid plaques and neurofibrilatory tangles in certain brain regions.

- Effective and well-targeted treatment necessitates early diagnosis of AD

- Magnetic Resonance Imaging (MRI): a non-invasive neuroimaging technique provides detailed images of brain structures in vivo
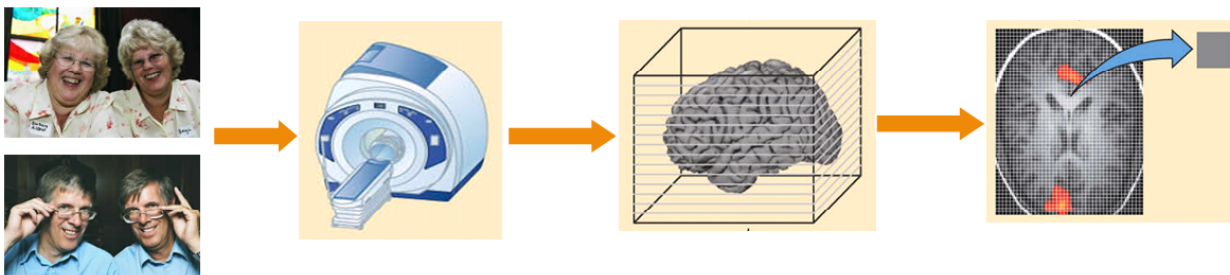
ADNI

- Alzheimer's Disease Neuroimaging Initiative (ADNI) Database (http://adni.loni.ucla.edu/)

- $N = 100$ subjects with 43 AD patients, and 57 CN (Cognitively Normal) subjects

- Randomly split into training (2/3) and test set (1/3), and repeat the split for 20 times.

- Image resolution: $128 \times 128 \times 128$ (3D), $p = 2,097,152$.

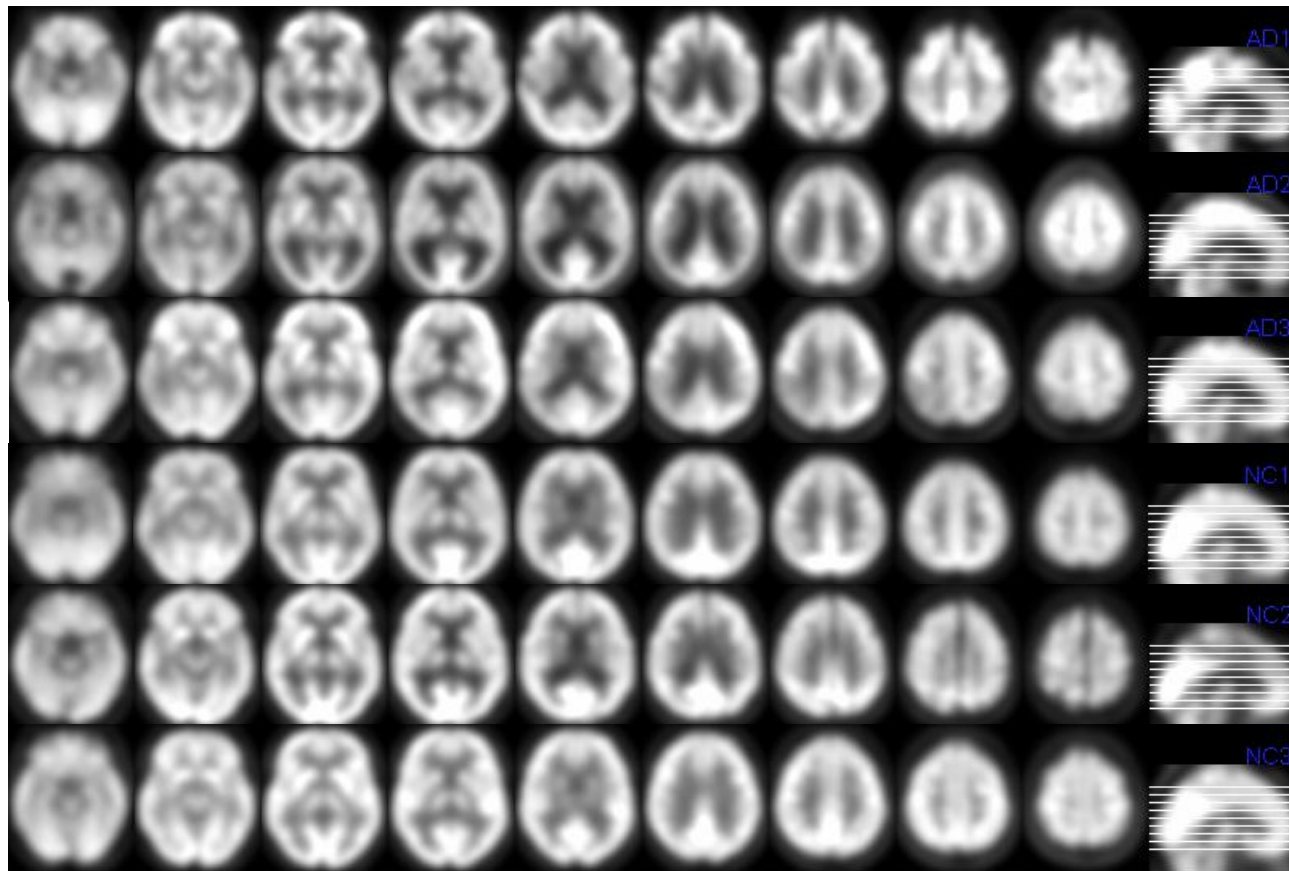- Sample image (RAVENS Map, Goldszal et al., 1998; Shen & Davatzikos, 2003):

ADNI

PET
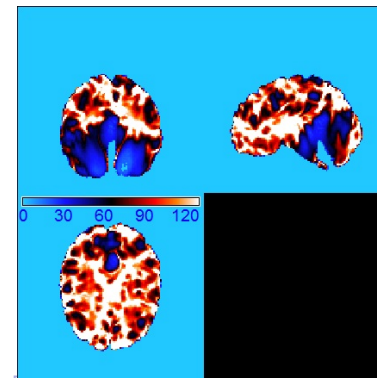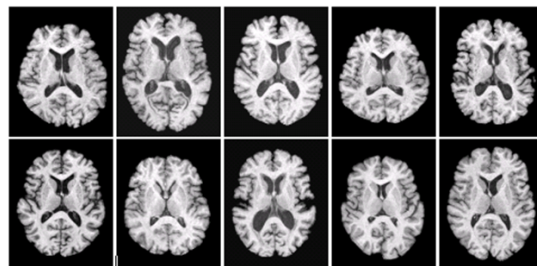
94 AD subjects and 104 NC subjects

Table 3: Results of Real Data: average misclassification rates.

sLDA	sPLS	sLogistic	SVM	ROAD	PCA	SWPCA
0.255	0.163	0.179	0.168	0.189	0.194	0.117