UNC SCHOOL OF PUBLIC HEALTH

Functional Data Analysis of Big Neuroimaging Data

Hongtu Zhu, Ph.D
Department of Biostatistics† and Biomedical Research Imaging Center‡
The University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599, USA

Acknowledgement: Some pictures were copied from multiple resources including Suetens (2009), Fass (2008), Dr. Niethammer, Drs. Lindquist, Rowe, Huettel, Wiki, google, gustaf@cb.uu.se, etc.

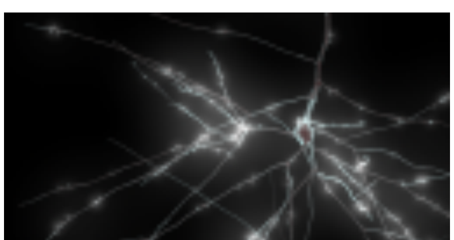
Big Neuroimaging Data

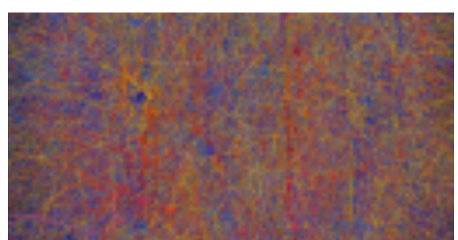
Is it really big?

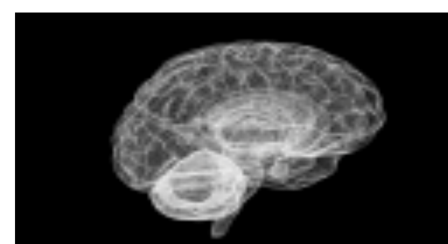
Human Brain Project

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

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

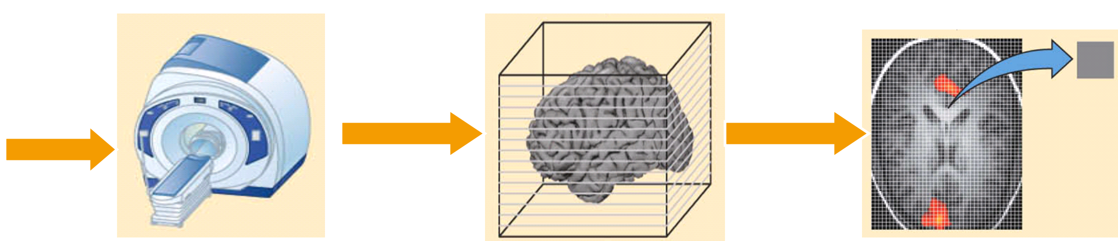
NIH normal brain development

1000 Functional Connectome Project

Alzheimer's Disease Neuroimaging Initiative

National Database for Autism Research (NDAR)

Human Connectome Project



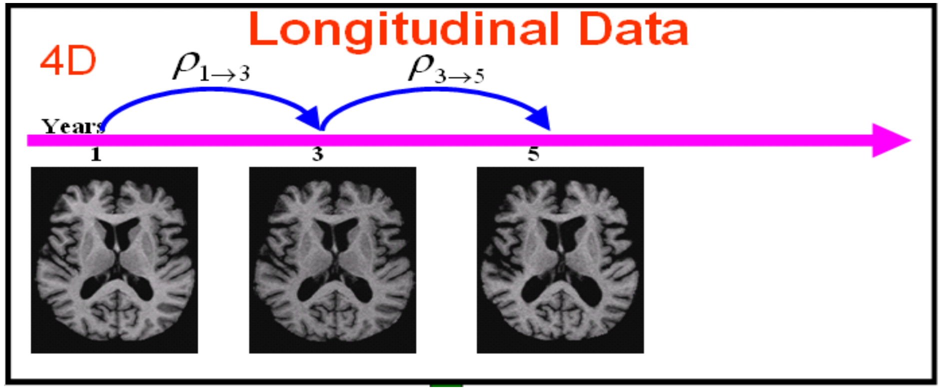
www.guysandstthomas.nhs.uk/.../T/Twins400.jpg

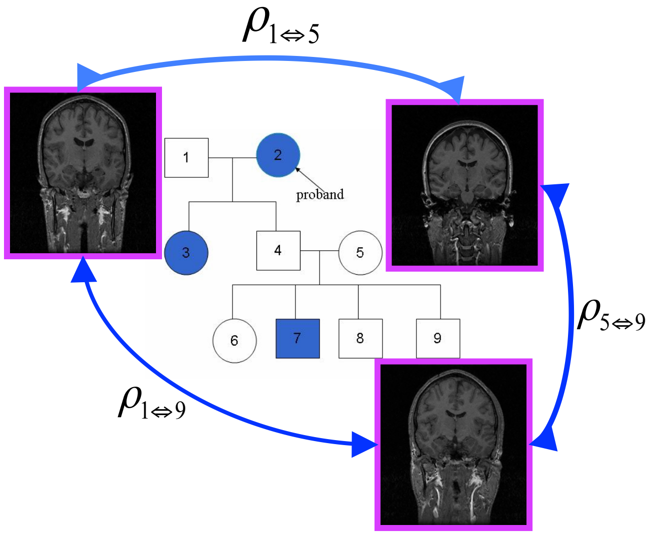
Complex Study Design

cross-sectional studies;

clustered studies including

longitudinal and twin/familial studies;





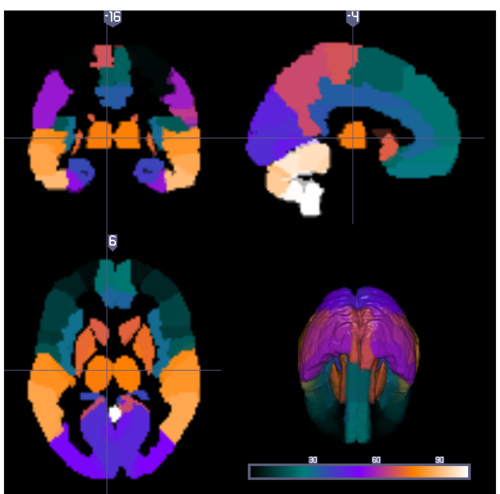
Complex Data Structure

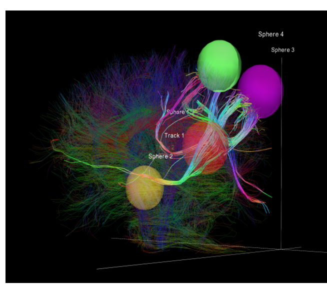
Multivariate Imaging Measures

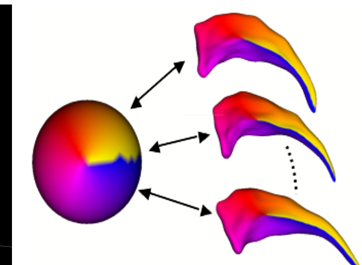
Smooth Functional Imaging Measures

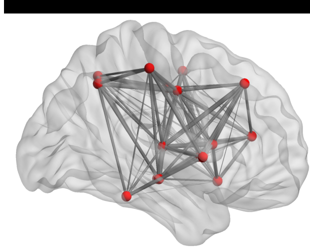
Whole-brain Imaging Measures

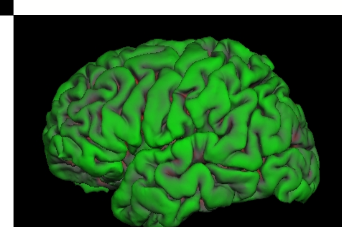
4D-Time Series Imaging Measures

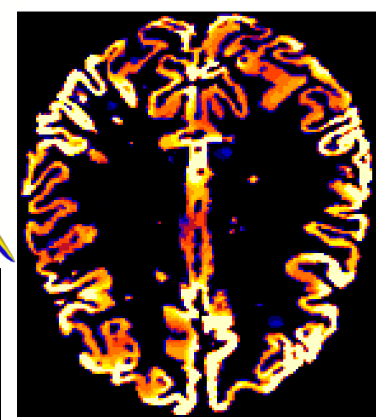


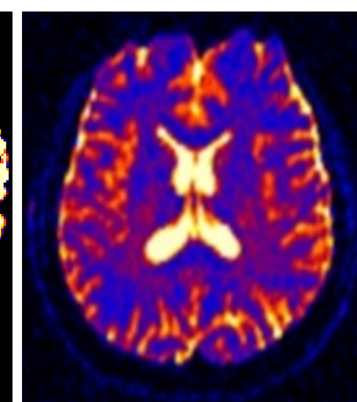




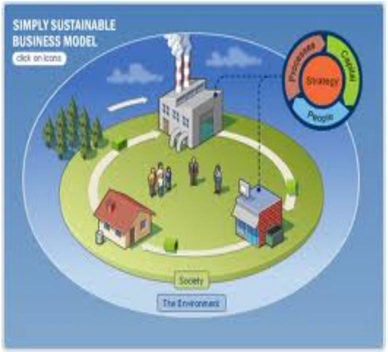


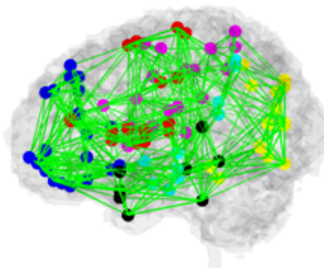


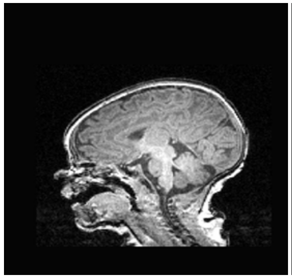


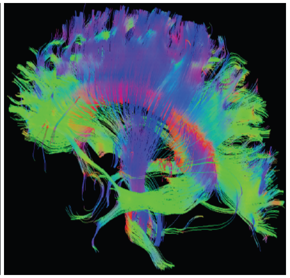


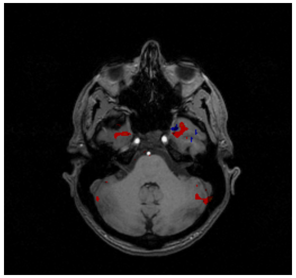
Big Data Integration

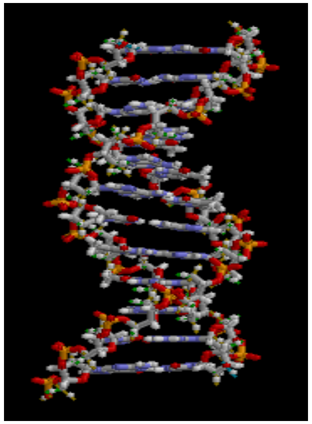


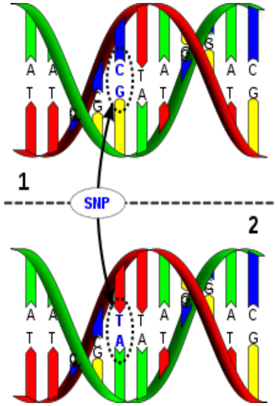


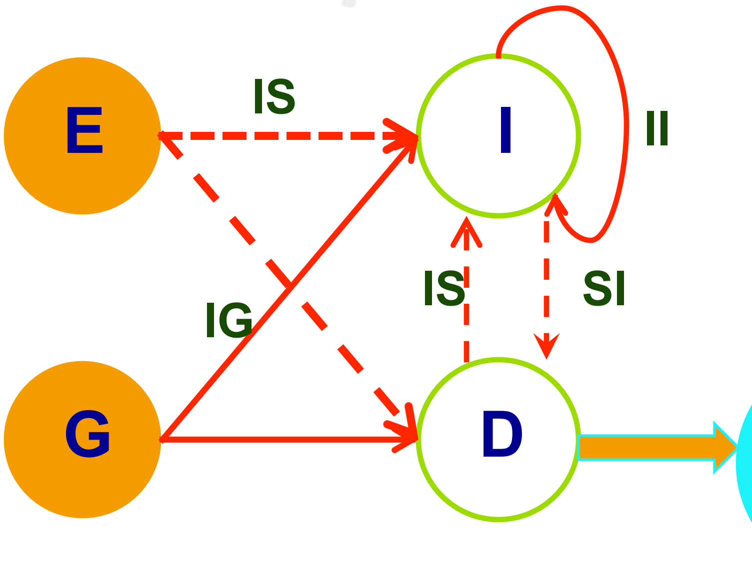












http://en.wikipedia.org/wiki/DNA_sequence

Models for Big Data Integration

Image-on-Scalar (IS) model

Image data as response, clinical variables as predictors.

Scalar-on-Image (SI) model

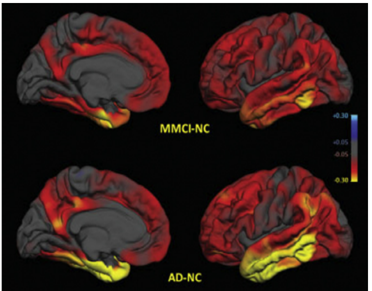
Clinical variables as response, image data as predictors

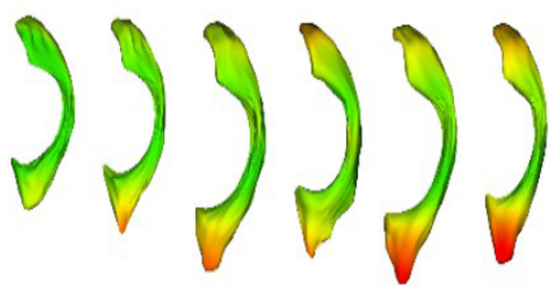
Image-on-Genetic (IG) model

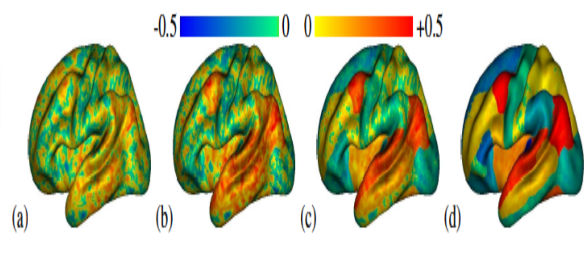
Image data as response, genetic data as predictors

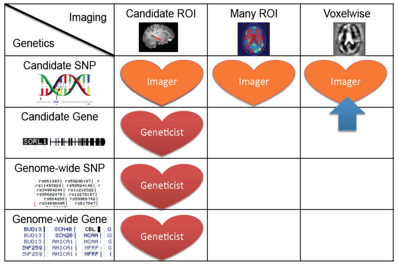
Image-on-Image (II) model

Image data as response, image data as predictors









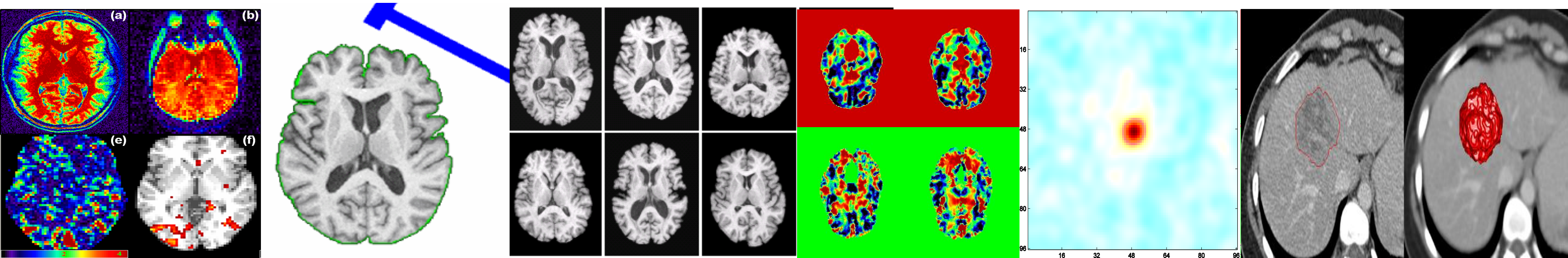
Noisy Imaging Data

- Spatial Maps
- `Registration'
- `Smoothing'
- `Correlation'
- `Spatial Heterogeneity'

Estimation

Inference

Prediction



Imaging-on-Scalar Regression

VBA versus FDA

Data $\{(x_i, Y_i) : i = 1, \cdots, n\}$ $\quad Y_i = \{Y_i(d_0) : d_0 \in D_0\}$

- **Intrinsic Discrete Approach (VBA)**

$$Y_i = \{Y_i(d_0) : d_0 \in D_0\}$$

- **Intrinsic Functional Approach (FDA)**

$$Y_i(\bullet) = \{Y_i(d) : d \in D\}$$

Functional Data Analysis (FDA)

Big data

$$y_i(d) = x_i^T B(d) + \eta_i(d) \qquad \eta_i(\bullet) \sim SP(0, \Sigma_\eta)$$

Hotelling-type Test Statistics T_n^2

Pro:

- Incorporate spatial smoothness and spatial correlation

Con:

- Computational and theoretical difficulties

$$\bar{Y} = \sum_{i=1}^{n} Y_i / n \qquad S_Y = \sum_{i=1}^{n} (Y_i - \bar{Y}) \otimes (Y_i - \bar{Y}) / n \qquad T_n^2 = n \sup_{\|u\|=1} \frac{<\bar{Y}, u>^2}{<u, S_Y u>}$$

$$\boxed{P(T_n^2 = \infty) = 1} \qquad S_Y \to \alpha(S_Y) \qquad S_Y \to diag(S_Y)$$

High-dimensional Regression Models

Big data

$$Y_i = BX_i + \eta_i \qquad \eta_i(\bullet) \sim SP(0, \Sigma_\eta) \qquad \dim(Y_i) >> n$$

Hotelling-type Test Statistics T_n^2

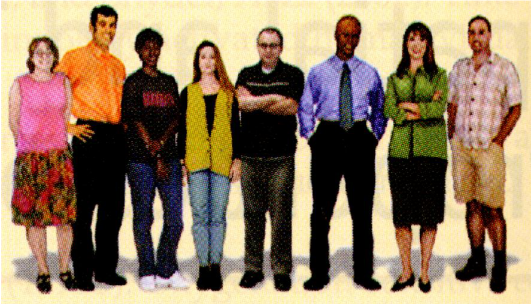
$$\bar{Y} = \sum_{i=1}^{n} Y_i / n \qquad S_Y = \sum_{i=1}^{n} (Y_i - \bar{Y})(Y_i - \bar{Y})^T / n \qquad T_n^2 = n \sup_{\|u\|=1} \frac{< \bar{Y}, u >^2}{< u, S_Y u >}$$

$$\boxed{P(T_n^2 = \infty) = 1} \qquad (\bar{Y}, S_Y) \to \alpha(Y, S_Y)$$

?? ??

Voxel Based Analysis (VBA)

Data $\{(x_i, Y_i) : i = 1, \cdots, n\}$ $\quad Y_i = \{Y_i(d_0) : d_0 \in D_0\}$



VBA

Stage 0: Gaussian Kernel Smoothing

Stage 1: Model Fitting

$$\prod_{i=1}^{n} p(Y_i \mid x_i) = \prod_{i=1}^{n} \prod_{d \in D_0} p(Y_i(d_0) \mid x_i, \theta(d_0))$$

Ignore spatial smoothness

Stage 2: Hypothesis Testing

$H_0 : \theta(d) = \theta_*(d)$ for all voxels

$H_1 : \theta(d) \neq \theta_*(d)$ for some voxels

Random Field Theory: functional data and local smoothness

FDR

VBA

Cons

Potential large smoothing errors.

Treat voxels as independent units/images as a collection of independent voxels.

Ignore spatial correlation and smoothness in statistical analysis.

Inaccurate for both Prediction and Estimation.

Decrease statistical power.

VBA Bayesian Extensions

Bayesian Modeling

Spatial smooth prior (MRF) $p(\theta) = p(\{\theta(d_0) : d_0 \in D_0\})$

$$p(\theta \mid Y) \propto \{\prod_{i=1}^{n} p(Y_i \mid x_i, \theta)\} p(\theta) = \{\prod_{i=1}^{n} \prod_{d \in D_0} p(Y_i(d) \mid x_i, \theta(d))\} p(\theta)$$

Pro:

- Computationally straightforward;
- Bayesian inference based on MCMC samples

Con:

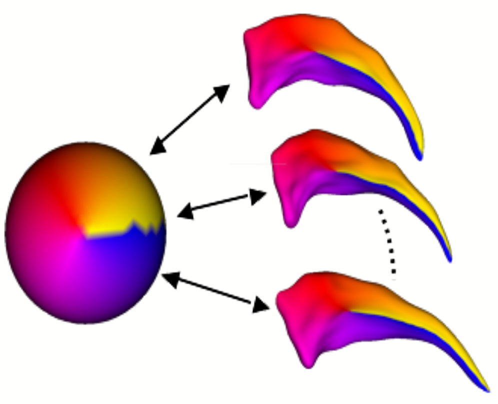
- Computationally heavy;
- Lack of understanding for Bayesian inference tools.

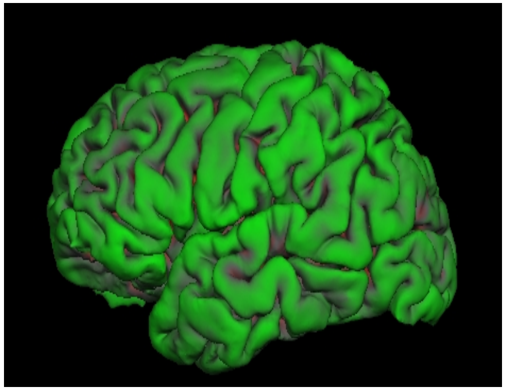
Varying Coefficient Models

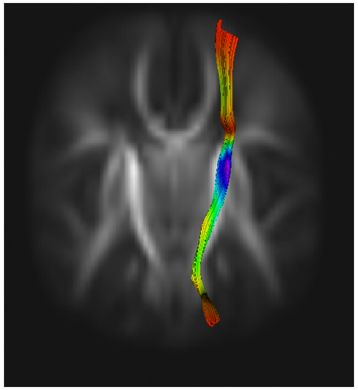
Reading materials:

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.

Smoothed Functional Data

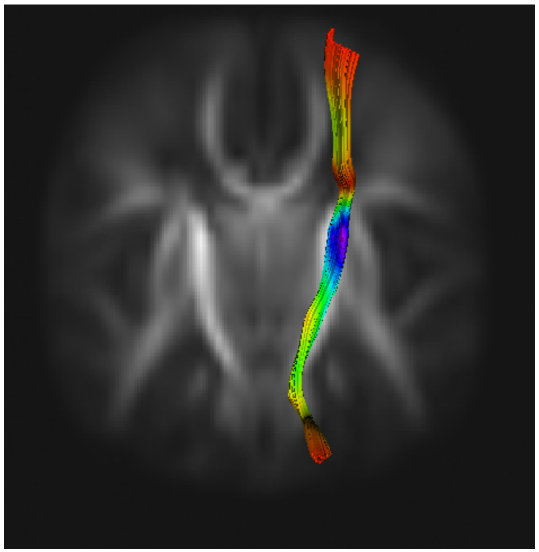






Covariates (e.g., age, gender, diagnostic)

DTI Fiber Tract Data



Data

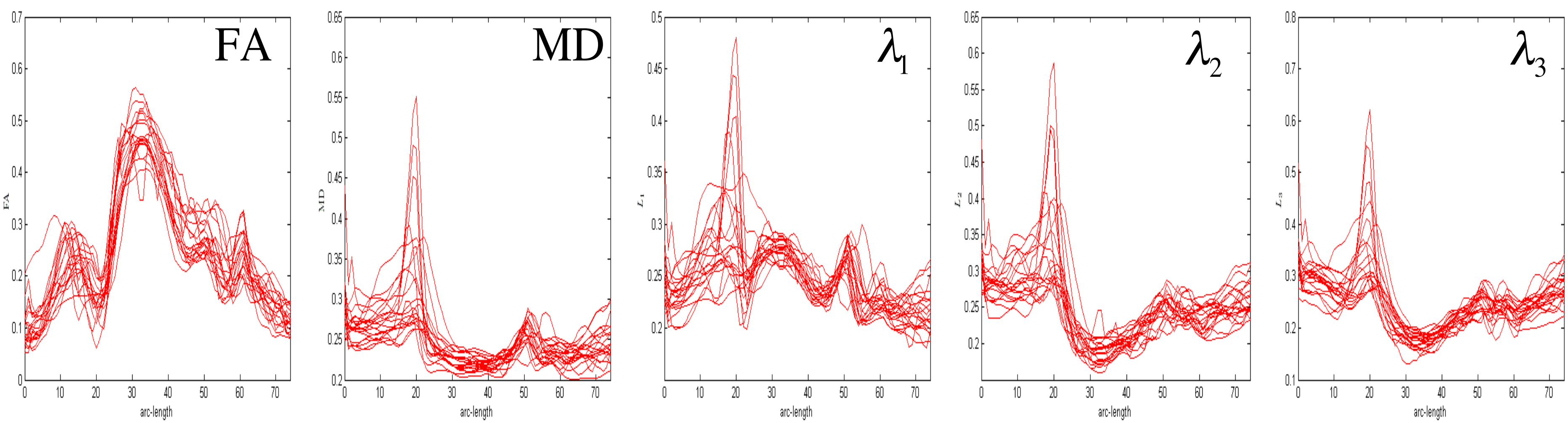
- **Diffusion properties (e.g., FA, RA)**

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

- **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: $\Sigma_y(s, s') = \Sigma_\eta(s, s') + \Sigma_\varepsilon(s, s')$

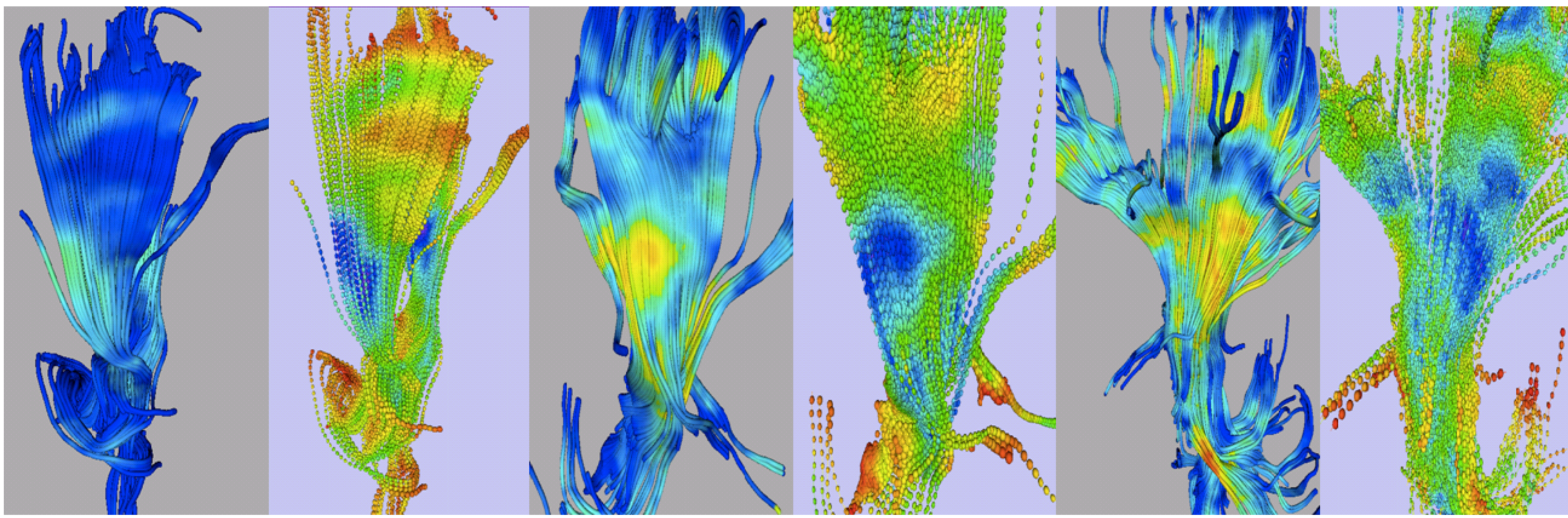
$$\sqrt{n}\{\mathrm{vec}(\hat{B}(d) - B(d) - 0.5O(H^2)) : d \in D\} \xrightarrow{L} G(0, \Sigma_B(d, d'))$$

Zhu, Li, and Kong (2012). AOS

Motivation

Diffusion Tensor Tract Statistics

FA **Tensor**

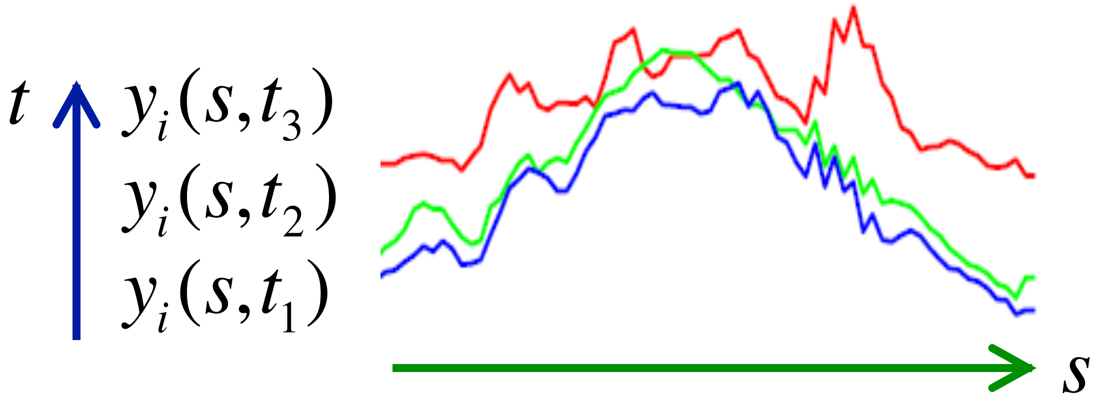


(a1)	(b1)	(a2)	(b2)	(a3)	(b3)
2 week	**2 week**	**1 year**	**1 year**	**2 year**	**2 year**

Longitudinal Extensions

Longitudinal Data

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

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

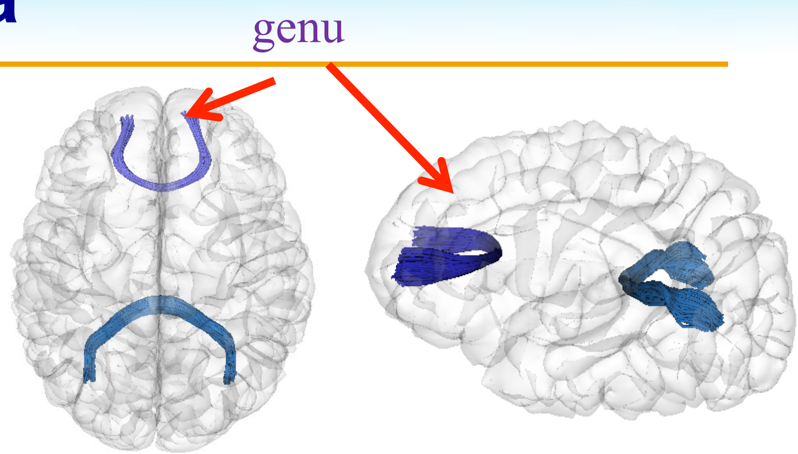
$$\sqrt{n}\{\mathrm{vec}(\hat{B}(d) - B(d) - 0.5O(H^2)) : d \in D\} \xrightarrow{L} G(0,\Sigma_B(d,d'))$$

Ying 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

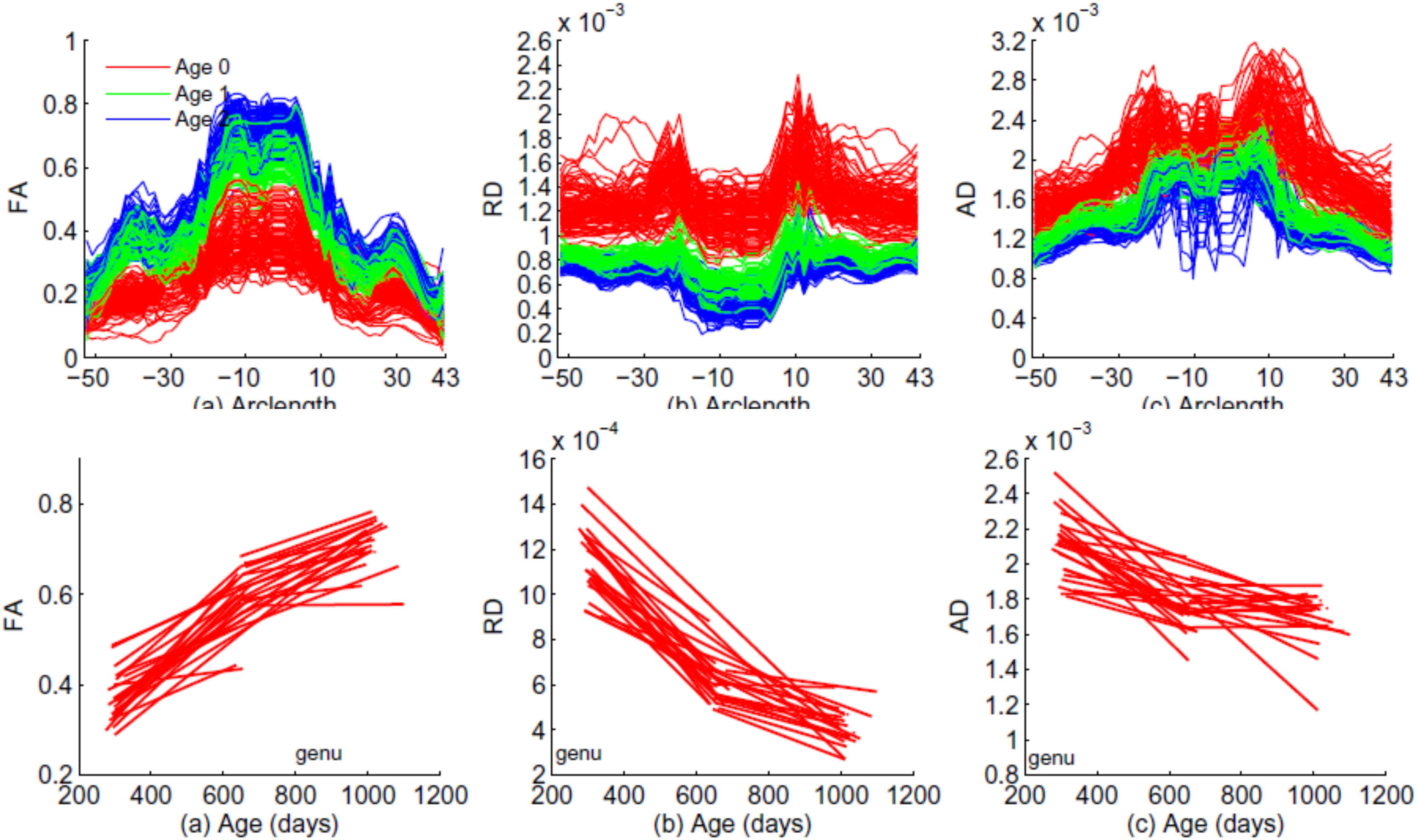
Available scans	N
Neonate scan only	1
1 year scan only	2
2 year scan only	3
Neonate + 1 year scan	43
Neonate + 2 year scan	30
1 year + 2 year scan	28
Neonate + 1 year + 2 year scan	30



DTImaging parameters:

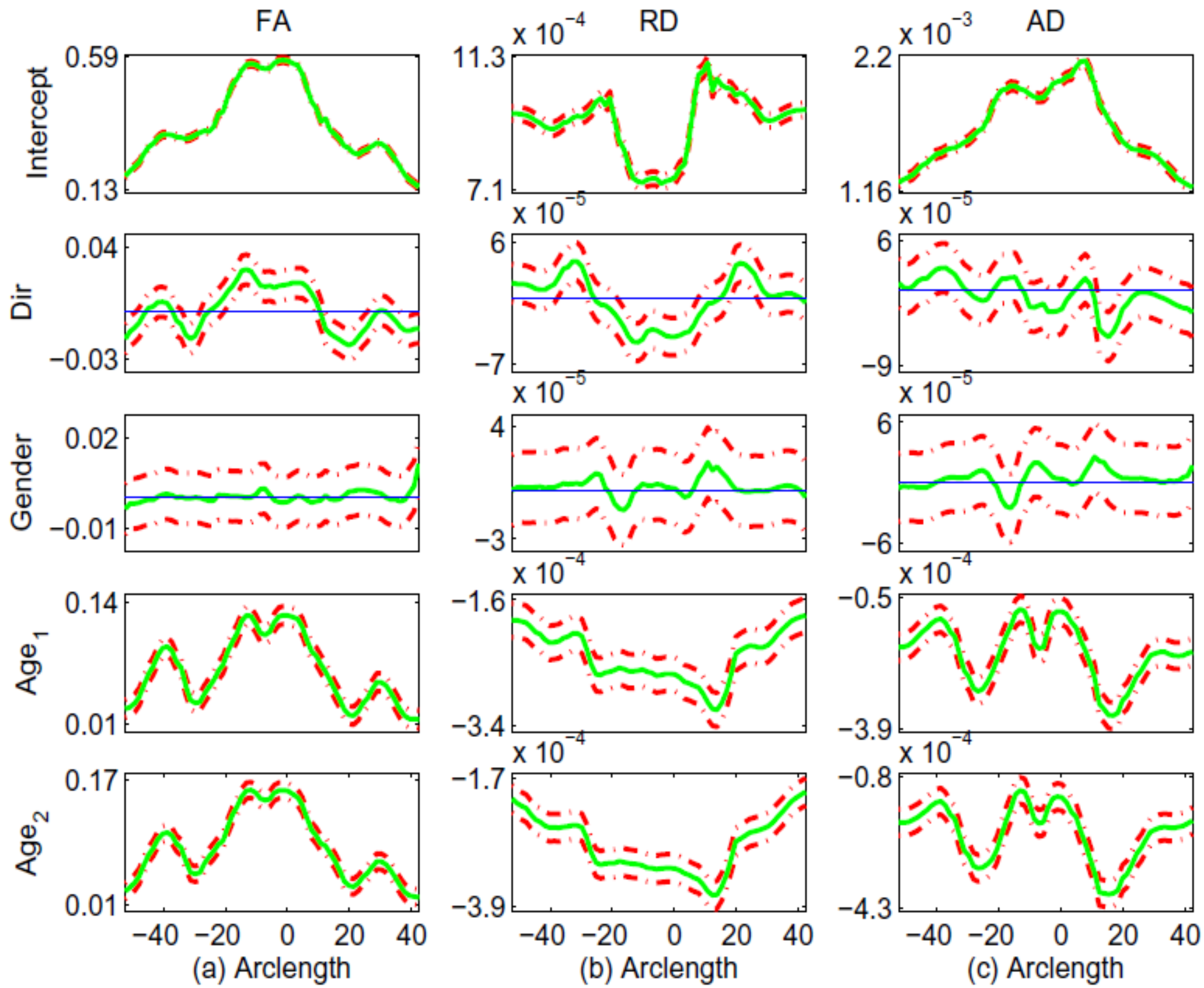
- TR/TE = 5200/73 ms
- Slice thickness = 2mm
- In-plane resolution = 2x2 mm^2
- b = 1000 s/mm^2
- One reference scan b = 0 s/mm^2
- Repeated 5 times when 6 gradient directions applied.

Real Data



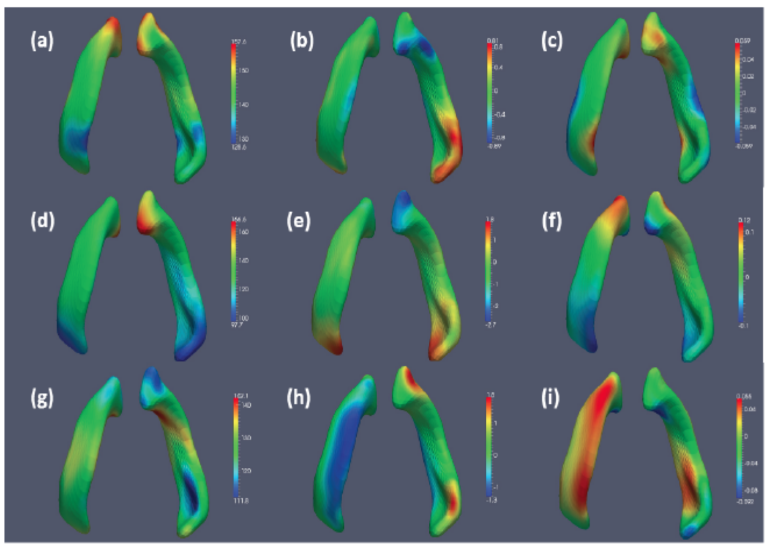
Age 0
Age 1
Age 2
FA
RD
AD
x 10^-3
x 10^-4
(a) Arclength
(b) Arclength
(c) Arclength
genu
(a) Age (days)
(b) Age (days)
(c) Age (days)

Real Data Analysis Results



Prediction

$$y_{i,k}(s) =$$



$$f(x_i\ , B_k(s) + \eta_{i,k}(s))$$

Long-range Correlation

$$+\,\varepsilon_{i,k}(s)$$

Small-range Correlation

Missing Big Data???

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

Real Data

Table 3: 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

Prediction Accuracy is much improved

Multiscale Adaptive Regression Models

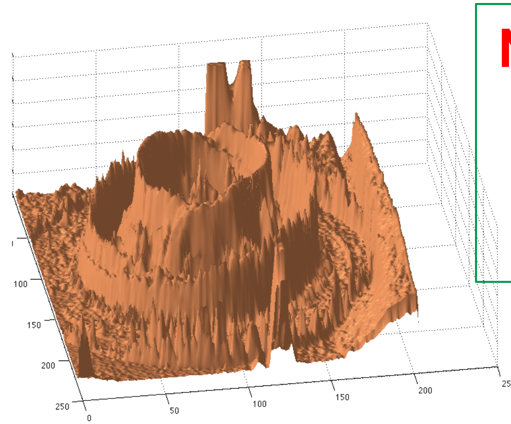
Reading materials:

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.

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.***

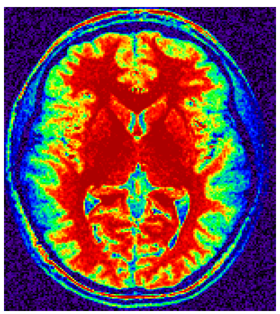
- Ω is a compact set. $\tilde{x} \in \Omega \subseteq R^k$

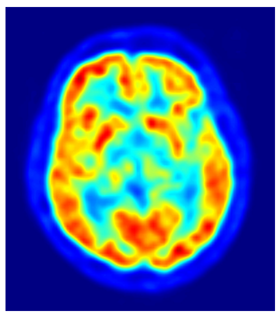
$\Rightarrow$ $f(\tilde{x}) \in M \subseteq R^m$ $\quad f : \Omega \to M \subseteq R^m$

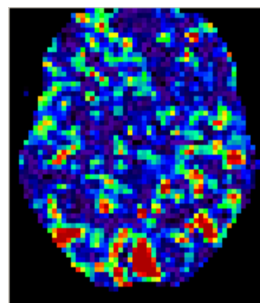
★ $\int_\Omega \| f(\tilde{x}) \|^k \, d\tilde{x} < \infty$ for some k>0

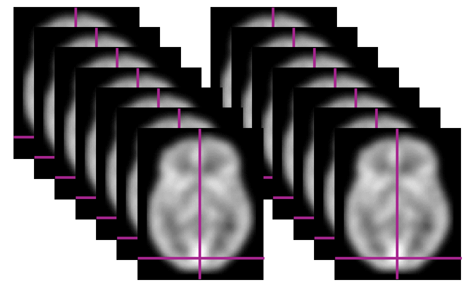
Neuroimaging Data with Discontinuity

Noisy Piecewise Smooth Function with Unknown Jumps and Edges









Subject1 *Subject2*

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

SVCM

Decomposition:

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

<u>Piecewise Smooth Varying Coefficients</u>

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

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:

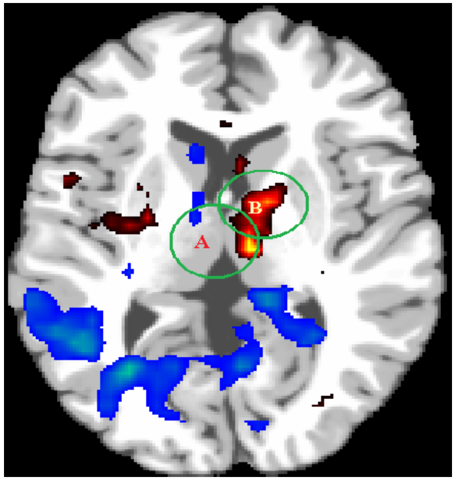
$$\Sigma_y(d, d') = \Sigma_\eta(d, d') + \Sigma_\varepsilon(d, d)$$

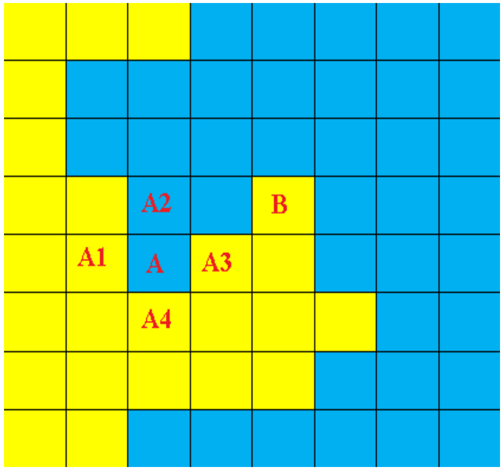
SVCM

Cartoon Model

$$B_k(d)$$

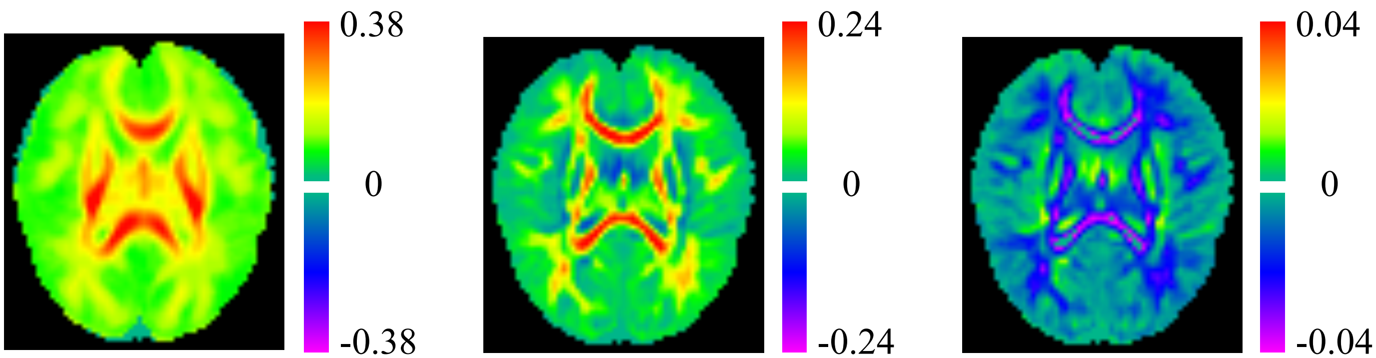
- **Disjoint Partition** $D = \cup_{l=1}^{L} D_l$ and $D_l \cap D_{l'} = \phi$
- **Piecewise Smoothness: Lipschitz condition**
- **Smoothed Boundary**
- **Local Patch**
- **Degree of Jumps**





Challenging Issues

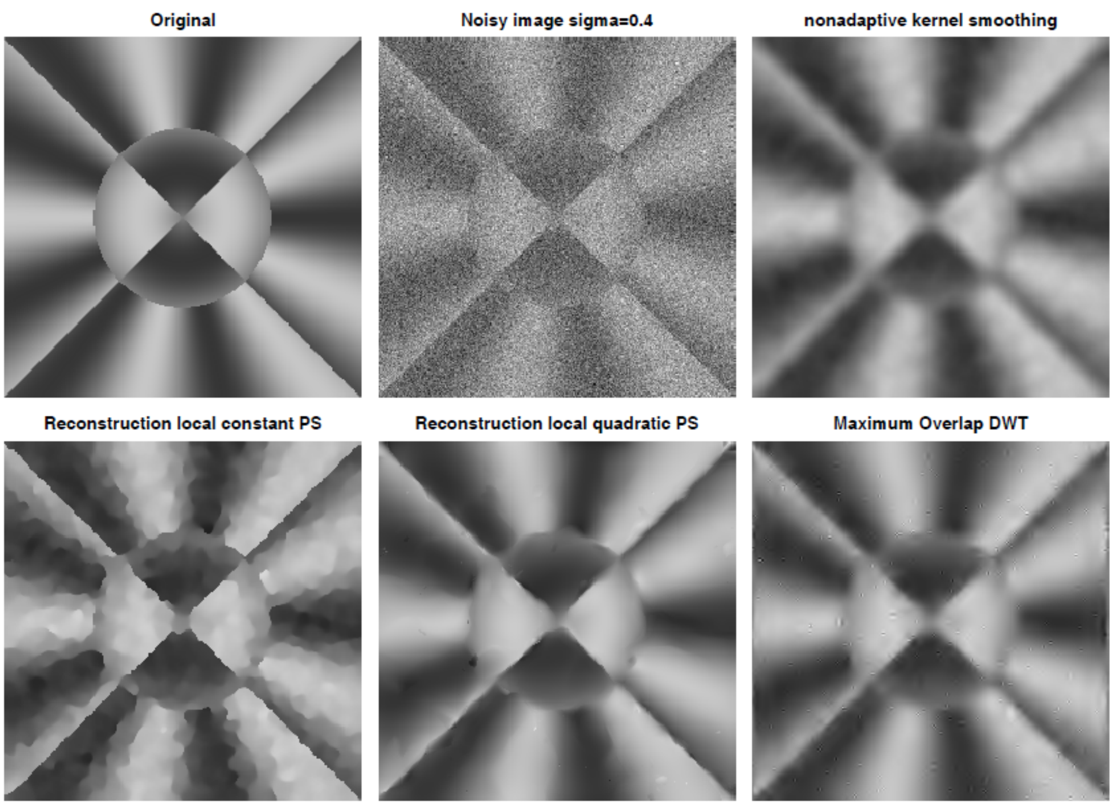
- Smoothing coefficient images, while preserving unknown boundaries
- Different patterns in different coefficient images
- Calculating standard deviation images
- Asymptotic theory



Smoothing Methods

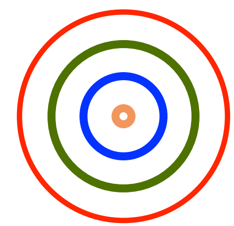
Propogation-Seperation Method

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



Features

- Increasing Bandwidth



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

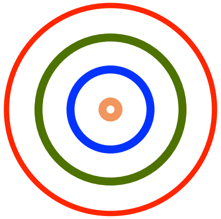
- Adaptive Weights

- Adaptive Estimates

Smoothing Methods

MARM

At each voxel d



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

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

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

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

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

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

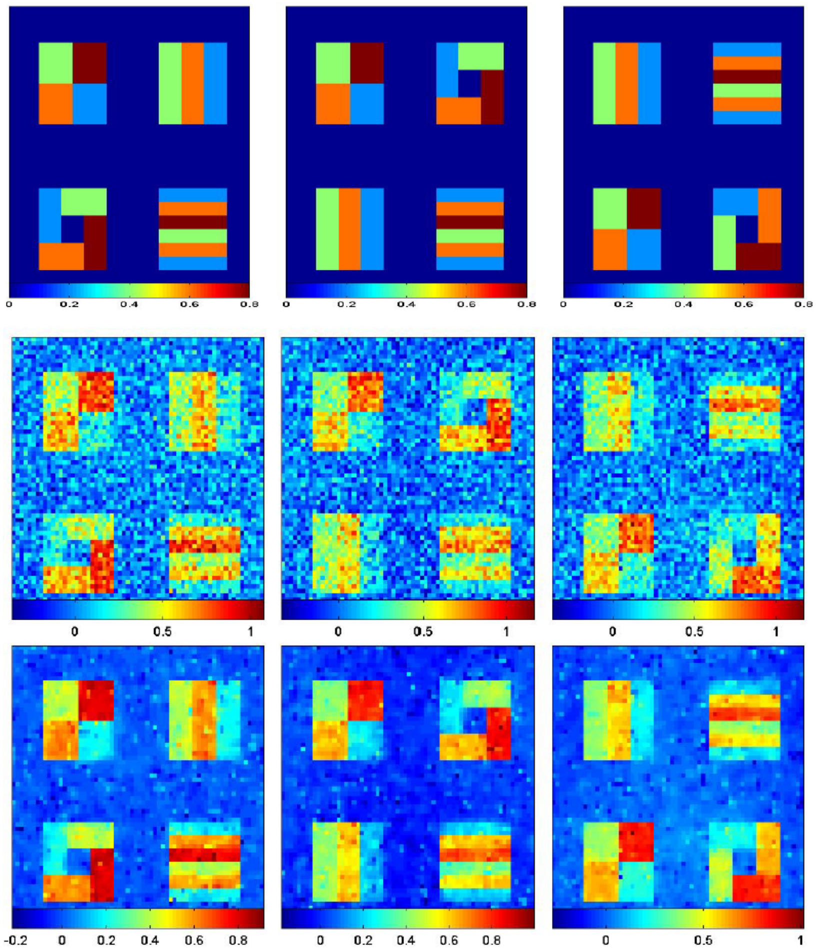
$\hat{\mu}(d;h_S)$

Stopping Rule

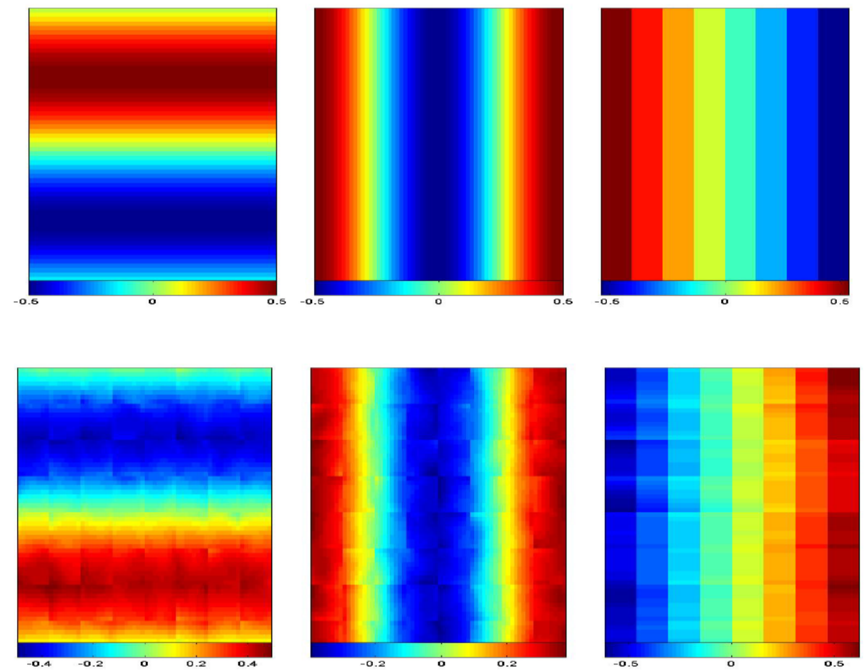
$$\omega(d,d';h_s) = K_{loc}(\| d - d' \| / h_s) K_{st}(D_\mu(d,d';h_{s-1}) / C_n)$$

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

Simulation

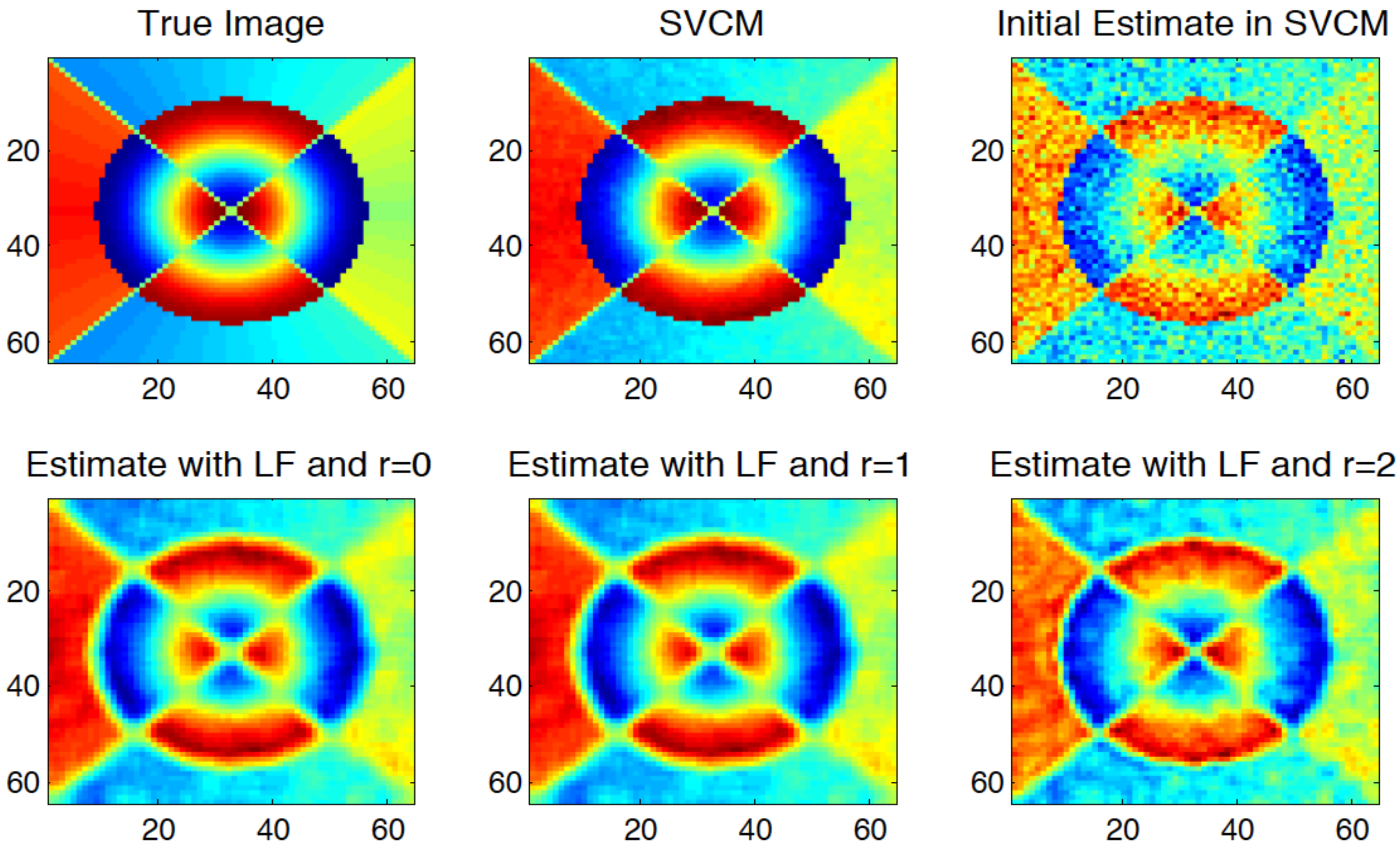


From up to down: initial and adaptive estimates; left to right: $\beta_1(d)$, $\beta_2(d)$, and $\beta_3(d)$.



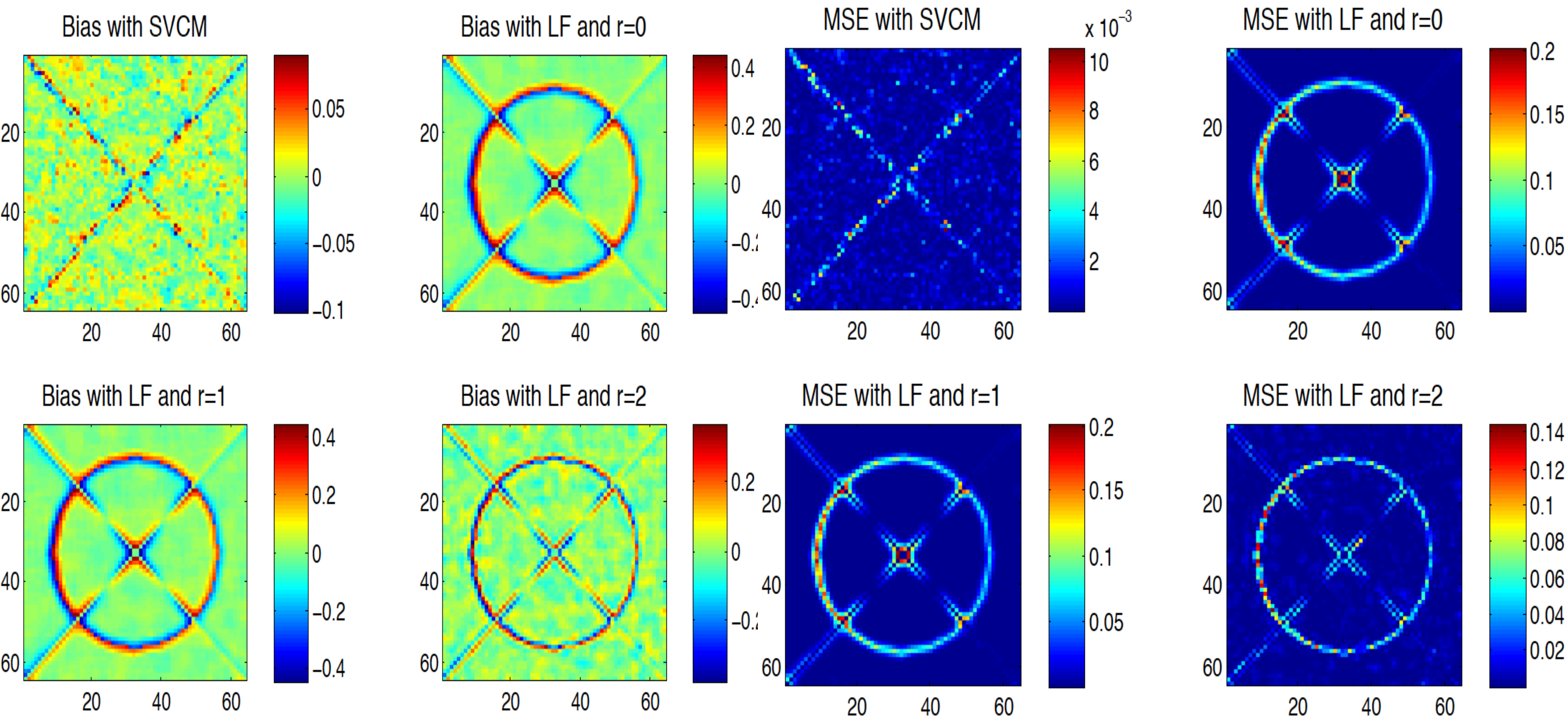
From left to right: $\psi_1(d)$, $\psi_2(d)$, and $\psi_3(d)$.

Simulation



True Image
SVCM
Initial Estimate in SVCM
Estimate with LF and r=0
Estimate with LF and r=1
Estimate with LF and r=2
20
40
60
20
40
60

Simulation


Bias with SVCM
Bias with LF and r=0
MSE with SVCM
MSE with LF and r=0
Bias with LF and r=1
Bias with LF and r=2
MSE with LF and r=1
MSE with LF and r=2
x 10^-3

Scalar-on-Imaging Regression

Scalar-on-Image Models

Reading materials:

1. Miranda, M. F., Zhu, H.T. and Ibrahim, J. G. (2014). Bayesian partial supervised tensor decomposition with applications in Neuorimaging data analysis.
2. Yang, H., Zhu, H. T., and Ibrahim, J. G. (2014). Multiscale projection model in RKHS.
3. Zhu, H. T. and Shen, D. (2014). Multiscale Weighted PCA for Imaging Prediction.
4. Guo, R.X., Ahye M., and Zhu, H. (2014). Spatially weighted PCA for imaging classification. *JCGS.* In revision.
5. Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in Neuorimaging data analysis. *JASA*. In press.

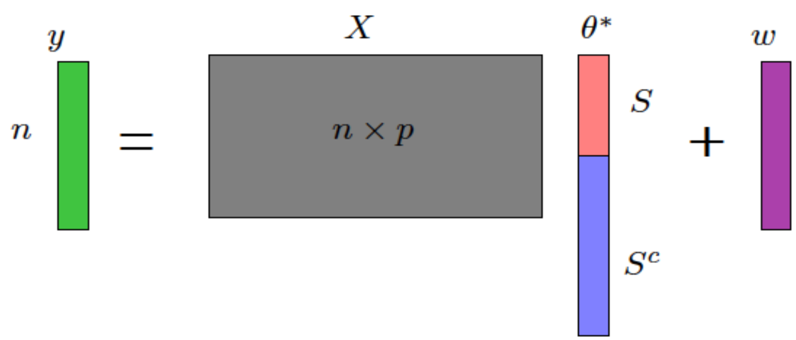
6. Cuingnet, R., Glaunes, J. A., Chupin, M., Benali, H., Colliot, O., and ADNI. (2012). Spatial and anatomical regularization of SVM: a general framework for neuroimaging data. IEEE PAMI. In press.
7. Cai, T. & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. JASA. 107,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

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))

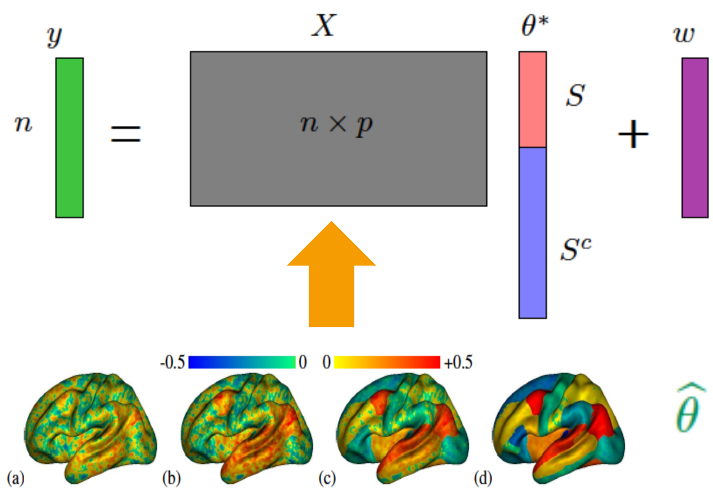


Strategy 2: Functional Regression Model (FRM)

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

High-dimension Regression Model

Approach 1: Regularization Methods



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

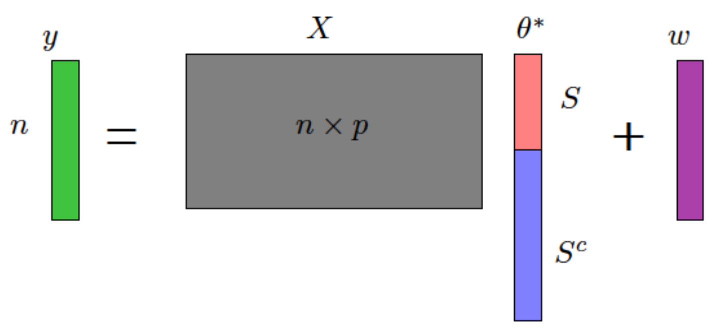
Key Conditions:

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

High-dimension Regression Model

Tensor Structure:

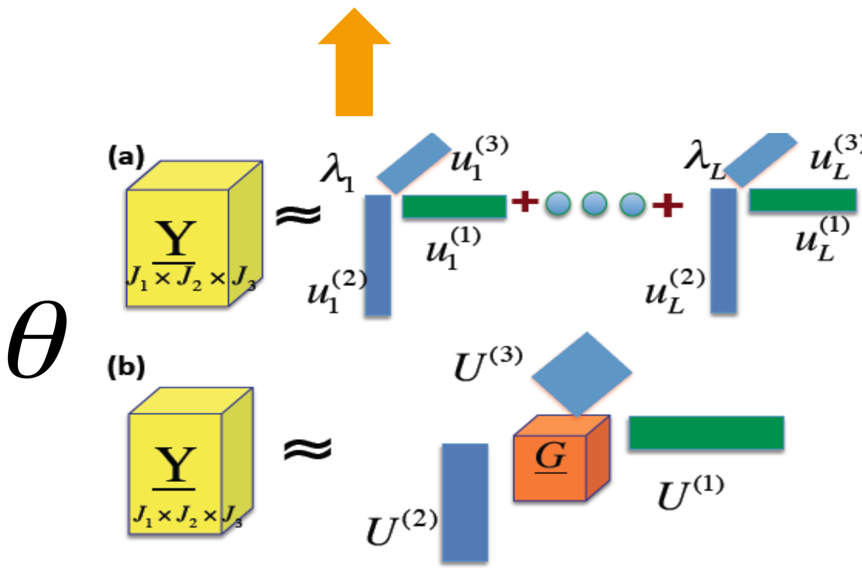
- Ultra-high dimensionality (256^3)
- Spatial structure



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

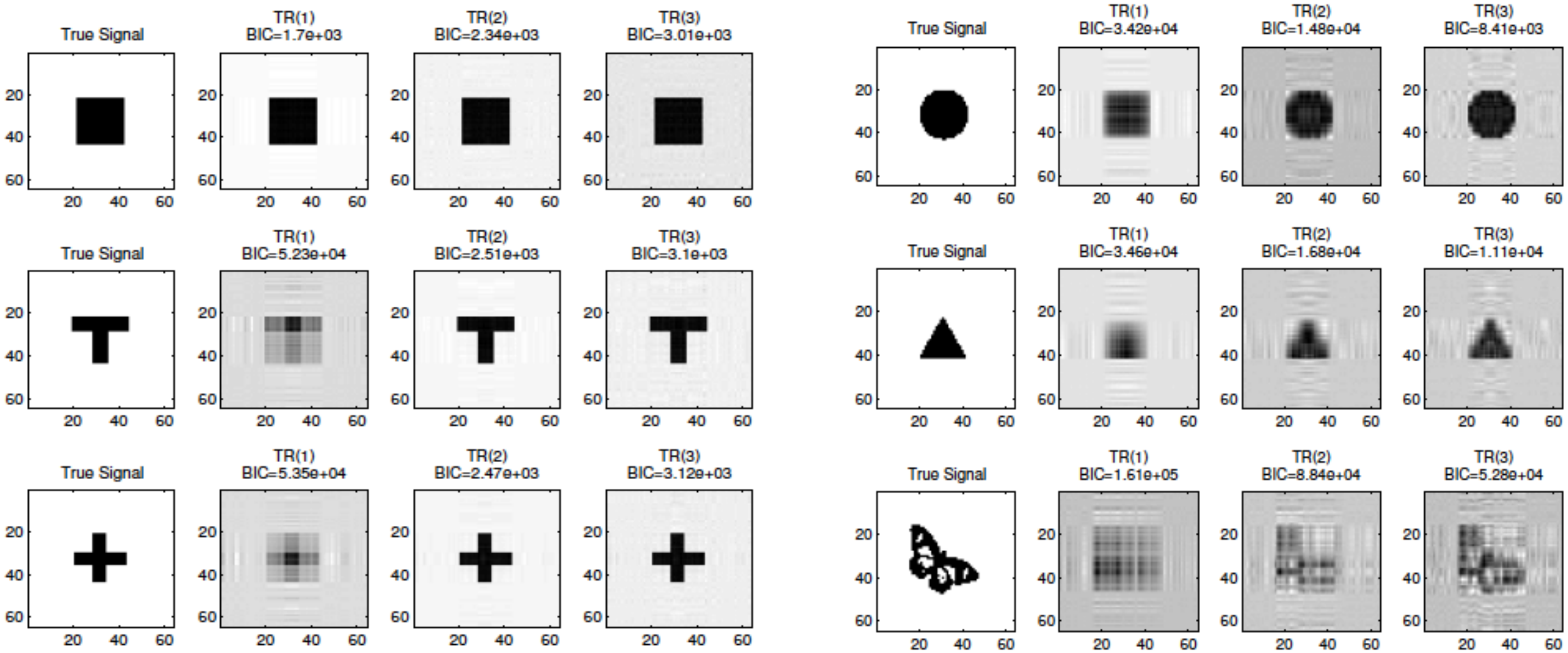
CP decomposition

Tucker decomposition



Scalar-on-Image Models

Simulations



Key Conditions:

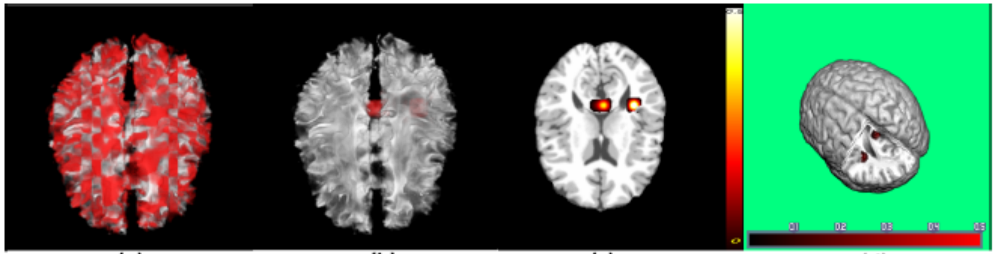
- Tensor Approximation B
- Restricted space property for X and B

Scalar-on-Image Models

ADHD 200

Attention Deficit and Hyperactivity Disorder (ADHD) data (http://fcon_1000.projects.nitrc.org/indi/adhd200/)

- 776 subjects: 491 normal controls and 285 combined ADHD subjects
- 442 males (average age: 11.98, sd: 3.14 years) and 287 females (average: 11.86, sd: 3.49)
- T1-weighted images were acquired and preprocessed by standard steps
- Segmentation: grey matter (GM), white matter (WM), ventricle (VN), and cerebrospinal fluid (CSF)



(a) (b) (c) (d)

Figure: Panel (a) is the unpenalized estimate overlaid on a randomly selected subject; (b) is the regularized estimate; (c) is a selected slice of the regularized estimate overlaid on the template; and (d) is a 3D rendering of the regularized estimate.

- Two regions of interest: left temporal lobe white matter and the splenium in the corpus callosum

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

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

Key Conditions

Key Conditions: an excellent set of basis functions

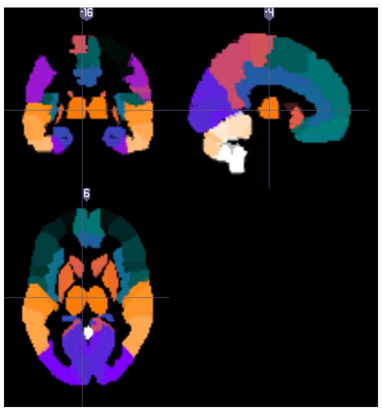
- **Sparsity of** $\{\theta_k : k = 1, \cdots\}$
- **Decay rate of spectral of** $C(d,d') = Cov(X(d), X(d'))$

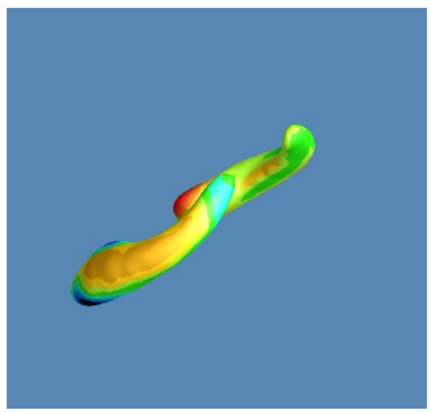
$$\theta(d) \approx \sum_{k=1}^{K} \theta_k \psi_k(d) \qquad K << n$$

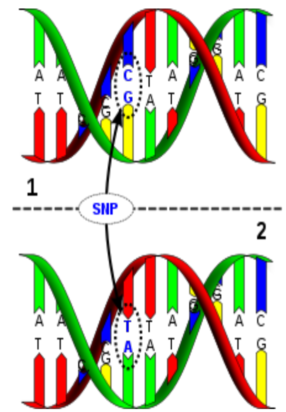
The ADNI data

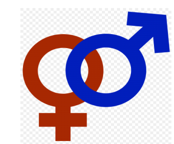
Mild Cognitive Impairment subjects

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

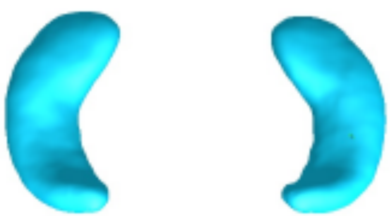
Partial Model: AUC=0.82

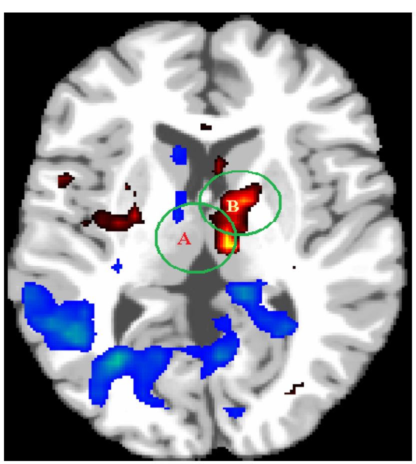
Limitations

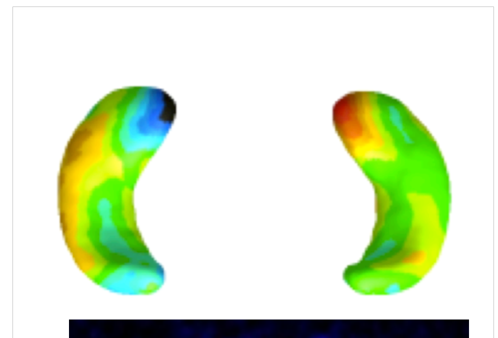
Is this the right space for statistical inference?

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

$$\{\int_D \psi_k(d) X_i(d) m(d) : d \in D\}$$





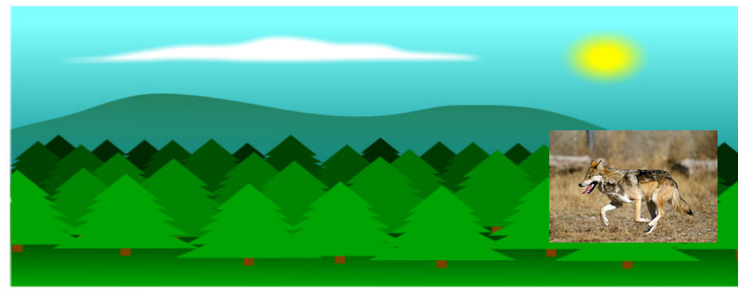


Rabbit and Wolf Story

$$\{y_i : i = 1, \cdots, n\}$$

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

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



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

$$G_i = G\{\tilde{X}_i(d) : d \in D_0\}$$

$$\tilde{X}_i = \{\tilde{X}_i(d) : d \in D_0\}$$

Feature Space Determination

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

Splitting/Weighting methods

$$\tilde{X}_i = \{\tilde{X}_i(d) : d \in D_0\}$$

Level sets
Factor Models

$$G_i = G\{\tilde{X}_i(d) : d \in D_0\}$$

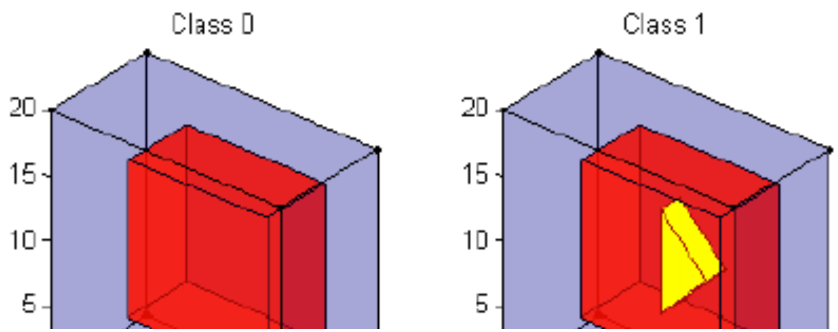
Splitting

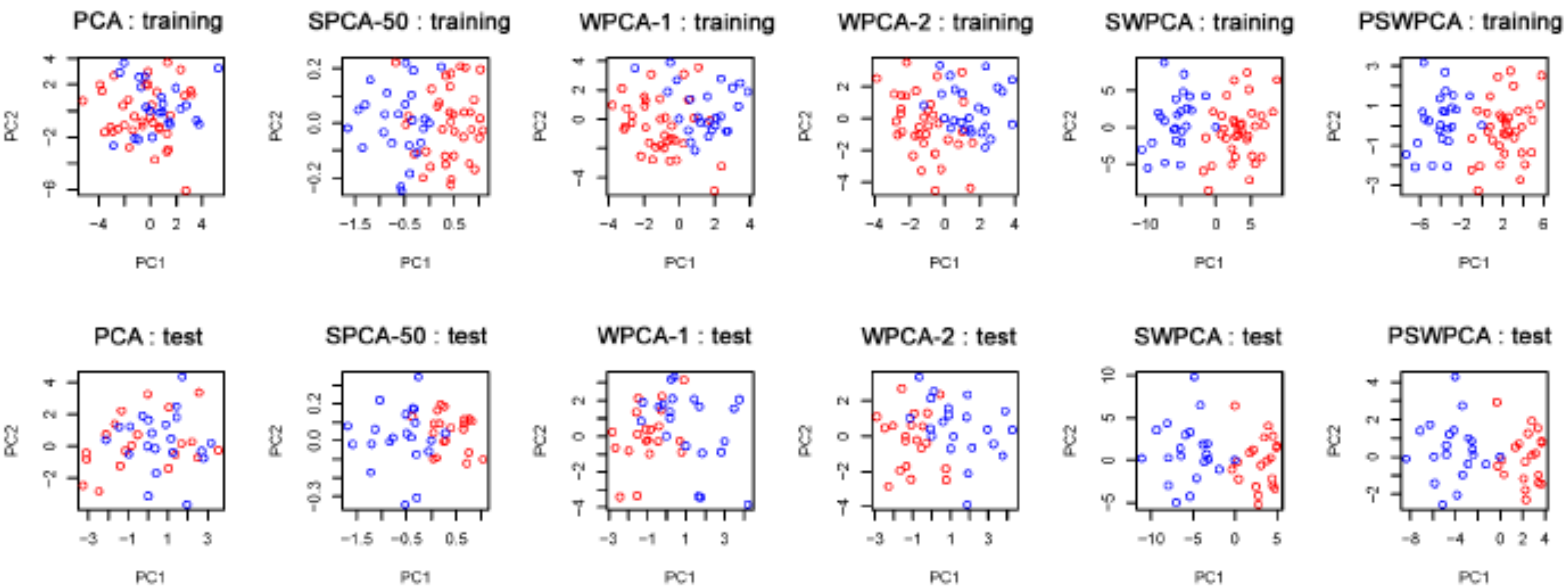
$$y_i = \theta_0 + \sum_k \int_{D_k} \theta(d) X_i(d) m(d) + \varepsilon_i$$

Weighting

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

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 (.078)	.126 (.052)	.132 (.052)	.142 (.055)	.162 (.057)	.205 (.064)	.199 (.064)	.130 (.056)	**.026** (.025)	**.025** (.024)
k-NN	.338 (.071)	.135 (.049)	.141 (.049)	.152 (.050)	.182 (.053)	.225 (.071)	.186 (.055)	.156 (.059)	**.030** (.029)	**.027** (.025)
SVM	.327 (.078)	.140 (.054)	.147 (.055)	.159 (.055)	.183 (.059)	.226 (.072)	.215 (.067)	.152 (.055)	**.033** (.029)	**.028** (.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 (.052)	.139 (.056)	.156 (.066)	.128 (.050)	.120 (.050)

Standard deviations are in parenthesis.

Multiscale Factor Prediction Model

Hippocampal Surfaces Data Analysis

- Hippocampal surface data consist of the vector with the length $30,000$ at the baseline for each subject.
- The first $15,000$ parts of the vector were from the left location and the rest parts of it were from the right location.
- We use the diagnostic covariate (Alzheimer's disease VS Normal), gender and age as demographic information.
- We also use the APOE genotype variables since relevant studies have shown that the APOE4 genotype has significant effect on the subject.

Multiscale Factor Prediction Model

Hippocampal Surfaces Data Analysis

- Our goal is to predict the behavior score of the subject at the time when it is five year after the baseline.
- We had 406 individulas, 226 individuals for training and 180 individuals for test set, respectively.
- First, we estimated the number of the nonzero gobal coordinates as 500, 700 and 1000 cases where for each case, the half of each case was selected from the left and the rest was selected from the right.
- Next, we extracted the local information from the sequence of the correlation matrices.

Multiscale Factor Prediction Model

Table: Predictions for behavior score in hippocampal surfaces data.

ID	Y_{true}	$\hat{Y}_{1000}$	$\hat{Y}_{700}$	$\hat{Y}_{500}$	Age	$Status$	$Gender$	$AP1$	$AP2$
270	27.33	27.934	27.746	27.559	85.2	1	1	3	4
268	13.33	14.895	14.733	15.238	82.8	1	1	3	4
318	30.67	32.029	32.237	31.742	88.2	1	1	3	3
283	20	18.348	18.631	19.043	80.1	1	1	3	3
304	24	22.873	23.128	22.894	80.1	1	0	3	4
307	17.67	16.163	16.126	16.267	76.2	1	0	3	4
312	17.33	17.484	17.532	17.42	72.3	1	0	3	4
280	20.67	22.537	20.688	20.092	72	1	0	3	4
302	17.67	17.947	17.425	17.506	80.1	1	0	2	4
345	10.33	10.132	11.517	12.233	80.7	0	1	3	4
343	3.67	4.035	3.502	3.977	73.8	0	1	3	4
337	7.33	6.809	7.012	7.758	72.6	0	1	3	4
361	9.67	10.466	10.794	10.992	85.8	0	1	3	3
344	13	14.677	14.689	14.61	70.8	0	1	3	3
364	11	11.557	11.718	11.742	70.8	0	1	3	3
401	0.67	1.12	1.175	1.082	74.1	0	1	2	3
380	7	6.866	5.476	5.961	72.9	0	0	3	4
353	8.67	8.123	7.431	7	83.7	0	0	3	3

ASA: Statistics in Imaging Section

SAMSI

2013 Neuroimaging Data Analysis
2015-2016 Challenges in Computational Neuroscience

Acknowledgement