UNC
SCHOOL OF
PUBLIC HEALTH

Big Data Integration in Biomedical Studies: A Few Personal Views and Experiences

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

Outline

- Big Data
- BIAS and Big Data Integration
- Image-on-Scalar Models
- Image-on-Genetic Association Models
- Prediction Models

Big Data

What is 'Big Data'?

5V=Volume, Velocity, Variety, Value, and Veracity

The size of big data is beyond the ability of commonly used software tools to capture, manage, and process within a tolerable elapsed time.

- **Alzheimer's Disease Neuroimaging Initiative (US$134 millions)**
- **Philadelphia Neurodevelopmental Cohort (PNC)**
- **Human Connectome Project (HCP)**

Big Data in Boxes

How to promote statistics in ‘Big Data’ industry?

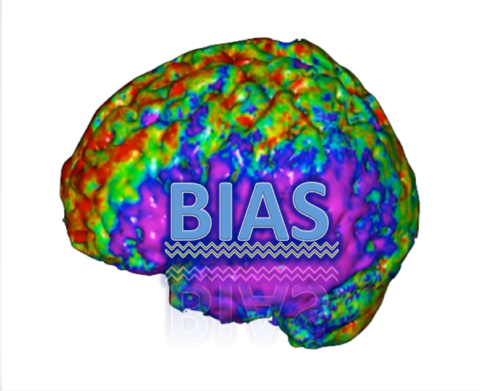
- Closely collaborate with people who are collecting ‘Big Data’
- Work as a team to develop new methods, packages, and textbooks with nice case studies
- Organize more workshops and short courses
- Train next-generation statisticians: training grants and new courses

a data scientist; an excellent programmer; an applied mathematician

How to make important contributions to 'Big Data'?

- Start with a few big data bases
- Start with a few methodological and clinical projects
- Develop a package with a set of good computational and statistical tools to efficiently extract important information from large Big data

BIAS: Biostatistics and Imaging AnalysiS and Big Data Integration



BIAS: Biostatistics and Imaging Analysis Lab

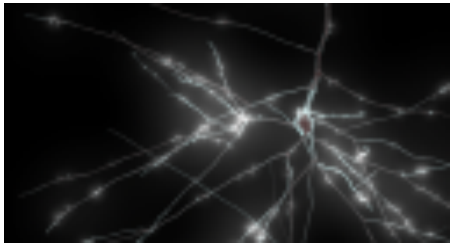
Man Power
Computer Power
Programming Power
Statistics
Mathematics

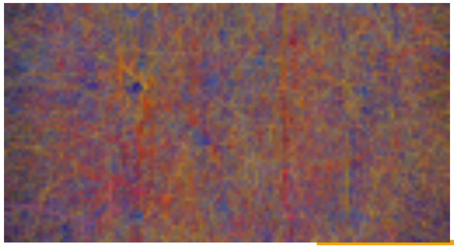
http://www.bios.unc.edu/research/bias/

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

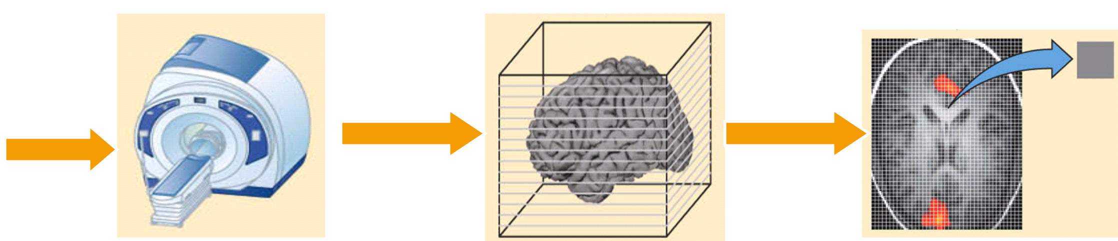
Alzheimer's Disease Neuroimaging Initiative

National Database for Autism Research (NDAR)

Human Connectome Project

Philadelphia Neurodevelopmental Cohort

Genome superstruct Project



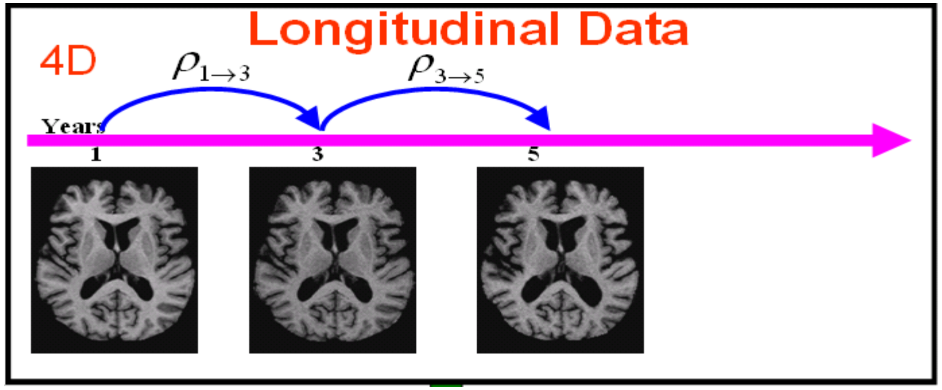
www.guysandstthomas.nhs.uk/.../T/Twins400.jpg

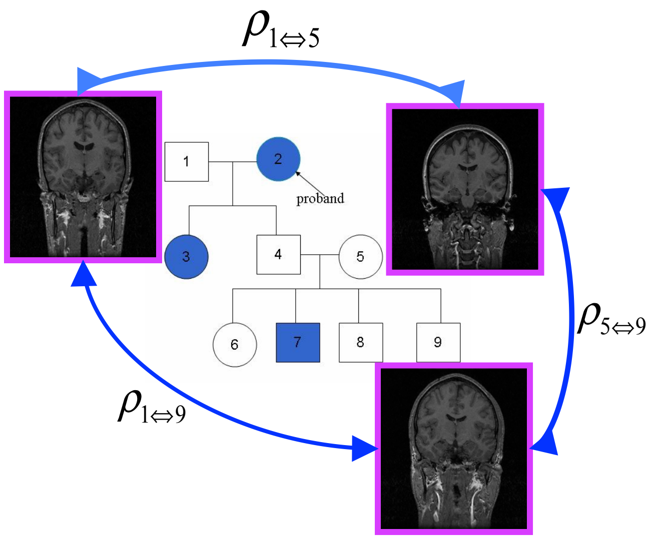
Complex Study Design

cross-sectional studies;

clustered studies including

longitudinal and twin/familial studies;





Neuroimaging Applications

Structural MRI

Functional MRI (task)

- Variety of acquisitions
- Measurement basics
- Limitations & artefacts
- Analysis principles
- Acquisition tips

Diffusion MRI

Functional MRI (resting)

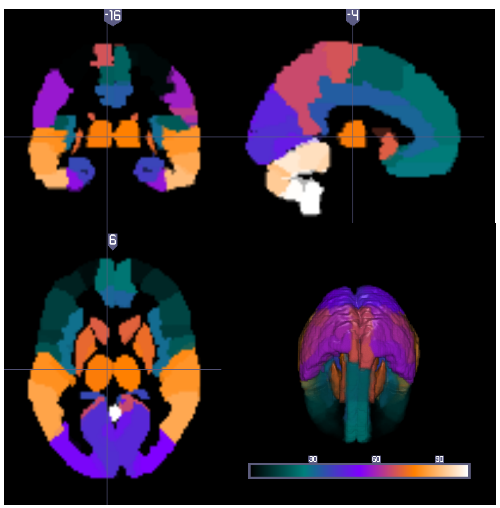
Complex Data Structure

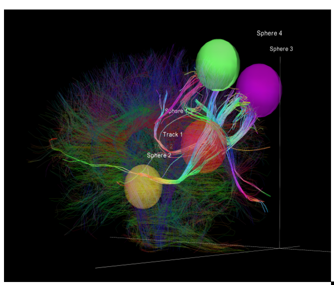
Multivariate Imaging Measures

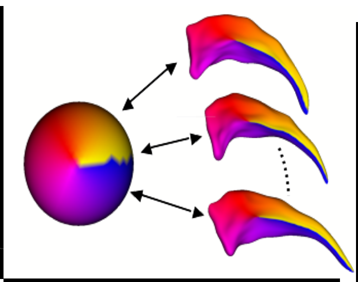
Smooth Functional Imaging Measures

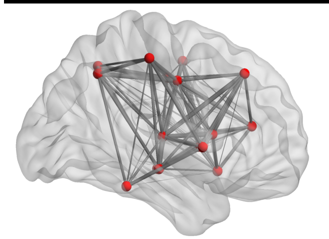
Whole-brain Imaging Measures

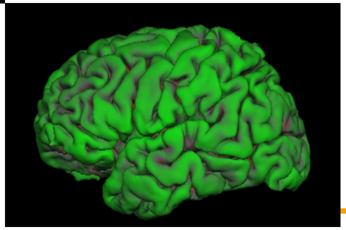
4D-Time Series Imaging Measures

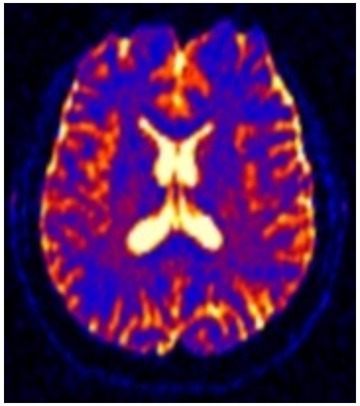




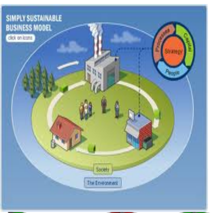


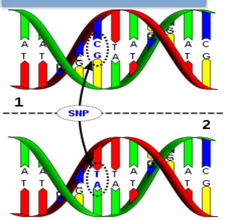


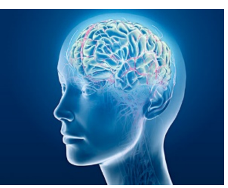




Big Data Integration







Medical Informatics & Management

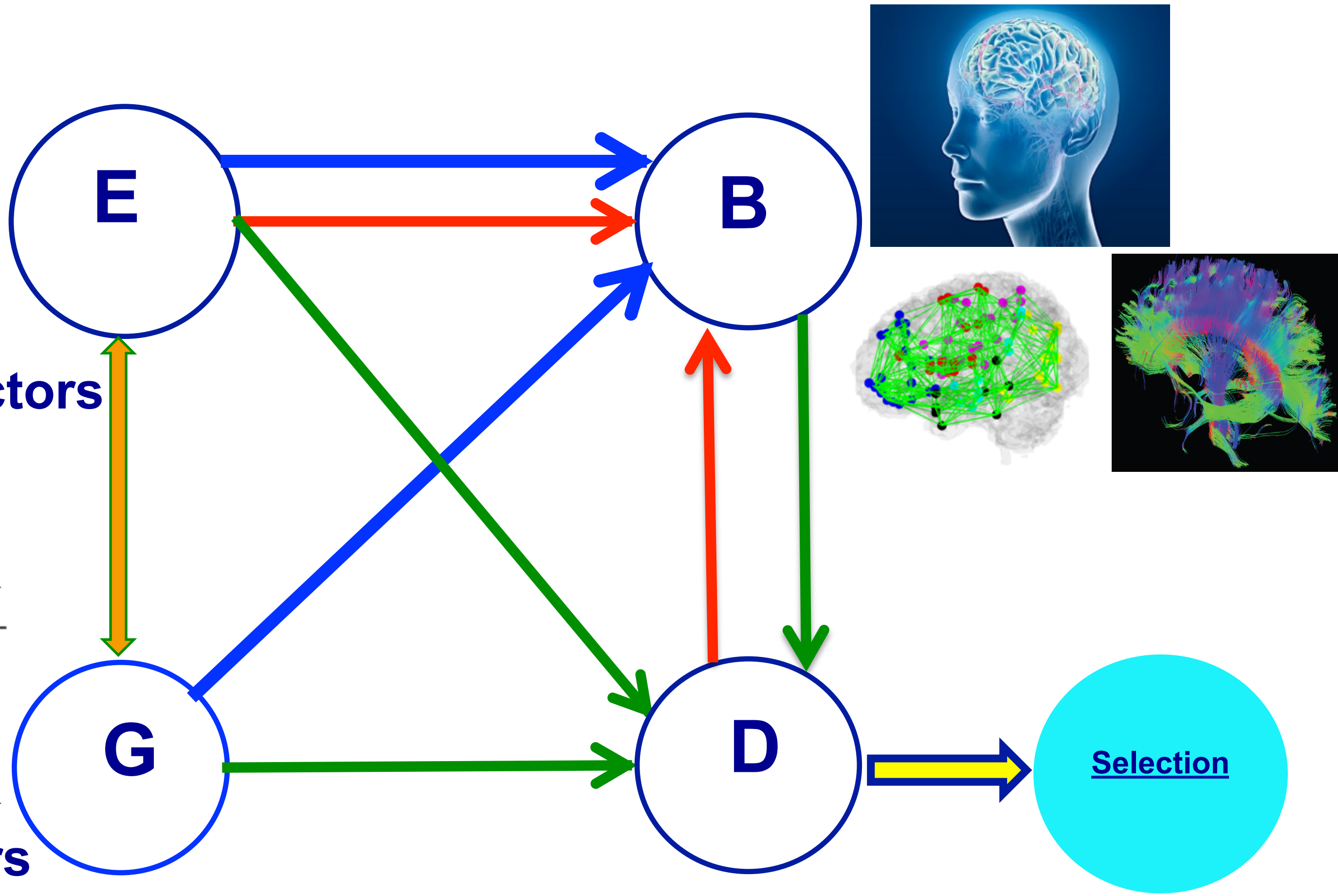
Disease

- Etiology
- Prevention
- Treatment

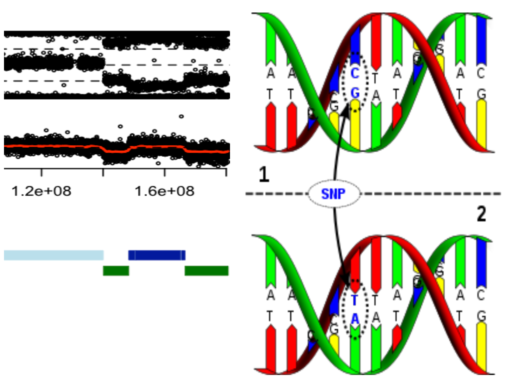
Medical Industry

- Care
- Policy
- System
- Science
- Insurance
- Economics
- Pharmaceutical

Big Data Integration



E: environmental factors



G: genetic markers

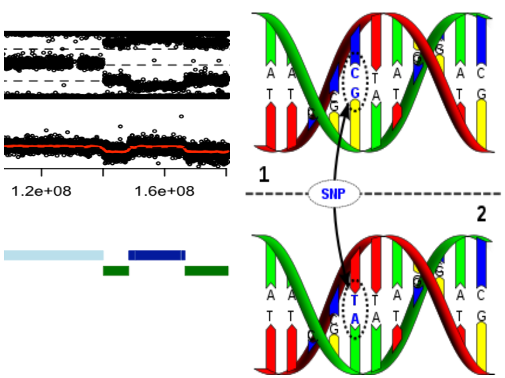
D: disease

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

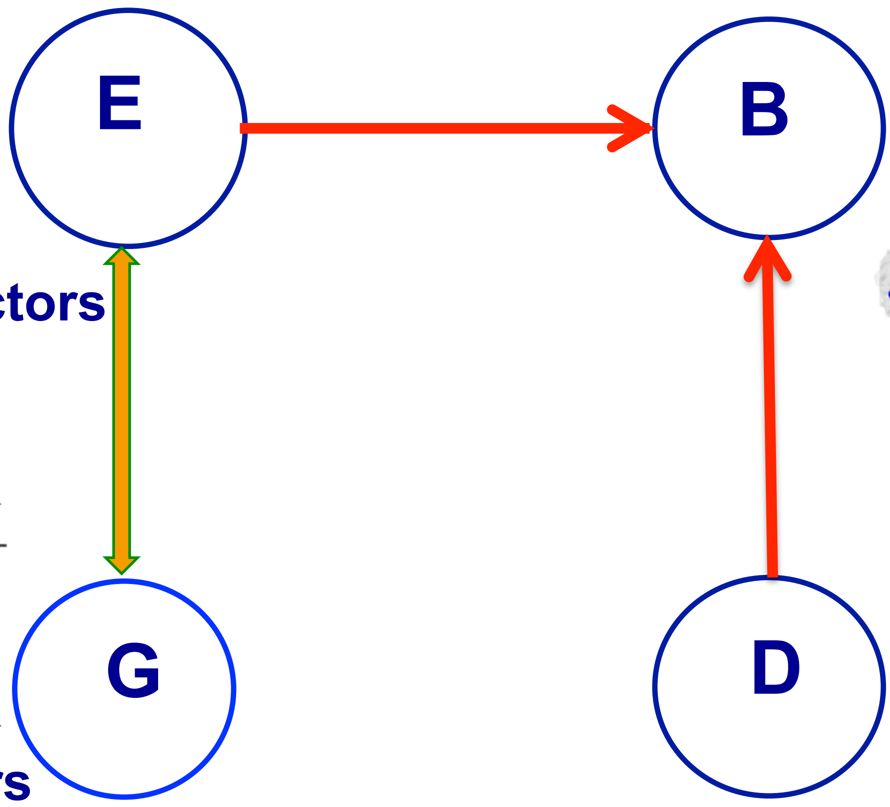
Image-on-Scalar Models

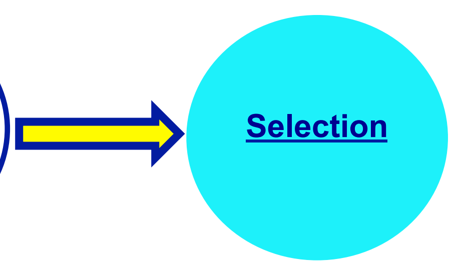
Big Data Integration

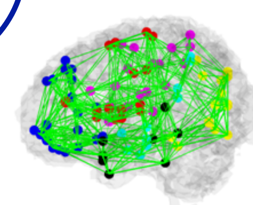
E: environmental factors

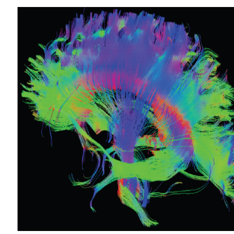


G: genetic markers









D: disease

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

The NIMH Strategic Plan

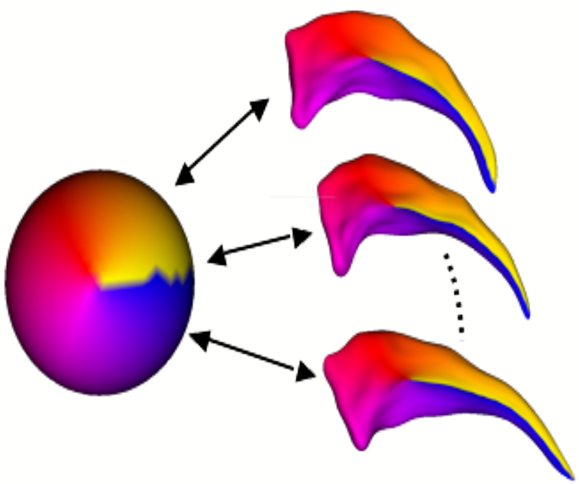
Strategic Objective 1: Promote Discovery in the Brain and Behavioral Sciences to Fuel Research on the Causes of Mental Disorders

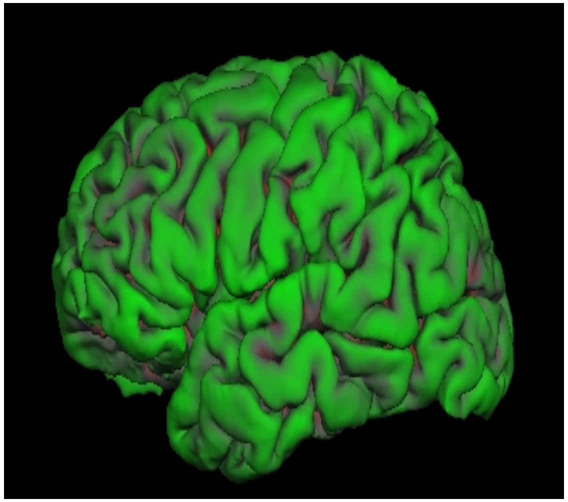
Identifying and validating high sensitivity and specificity biomarkers that define valid subtypes of the major mental illnesses.

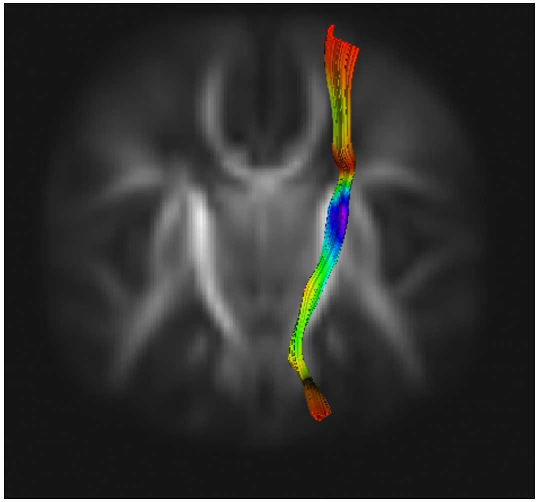
Strategic Objective 2: Chart Mental Illness Trajectories to Determine When, Where, and How to Intervene

Conducting longitudinal studies that track changes in behavior with brain structure, connectivity, and function, in order to characterize the progression from primary changes to subsequent clinical presentation, and to identify predictors of divergence from the typical trajectory.

Smoothed Functional Data



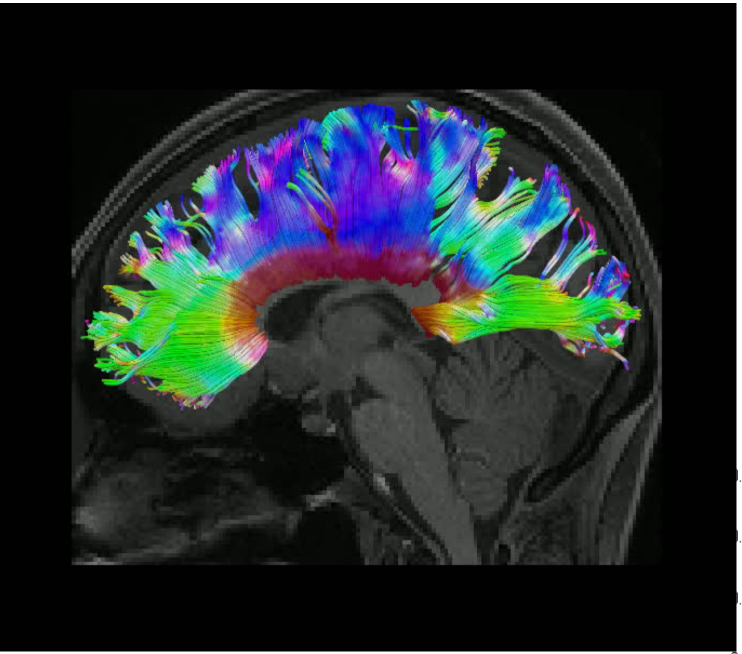




Covariates (e.g., age, gender, diagnostic)

Case 1: DTI Fiber Tract Data

Data



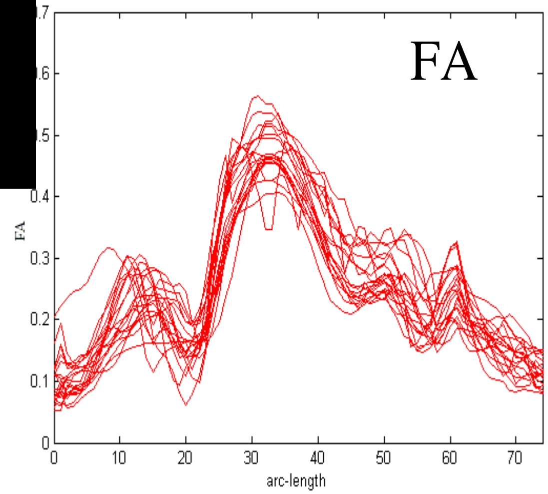
- Diffusion properties (e.g., FA, RA)

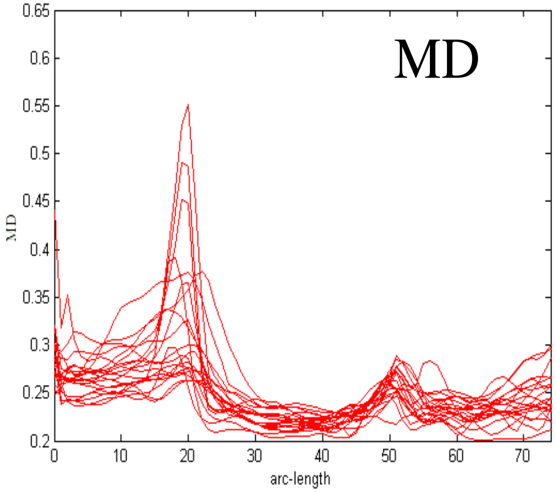
$$Y_i(s_j) = (y_{i,1}(s_j), \cdots, y_{i,m}(s_j))^T$$

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

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

$$x_1, \cdots, x_n$$

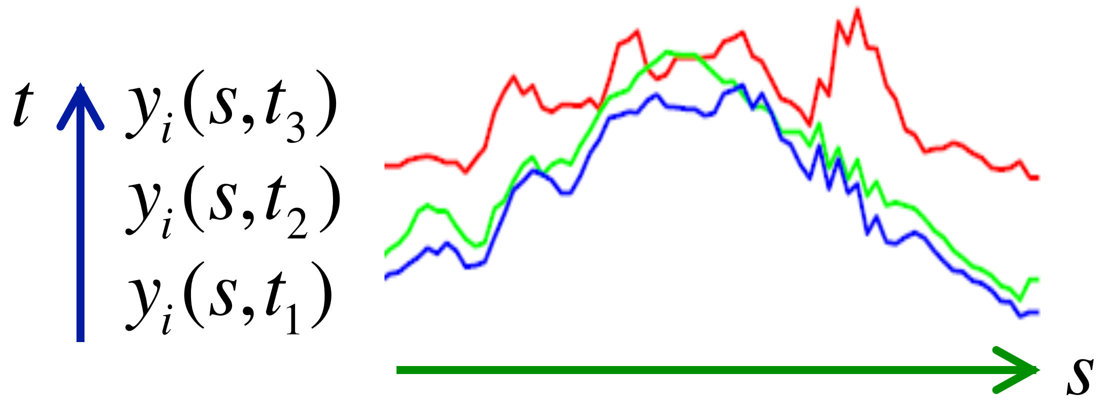




Longitudinal Tract Data

Longitudinal Data

Spatial-temporal Process



Functional Mixed Effect Models

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

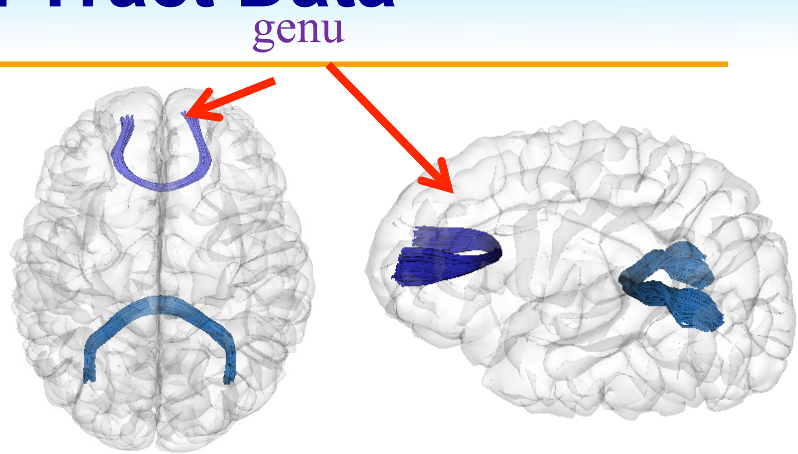
Objectives:

Dynamic functional effects of covariates of interest on functional response.

Ex 1: Longitudinal Tract 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

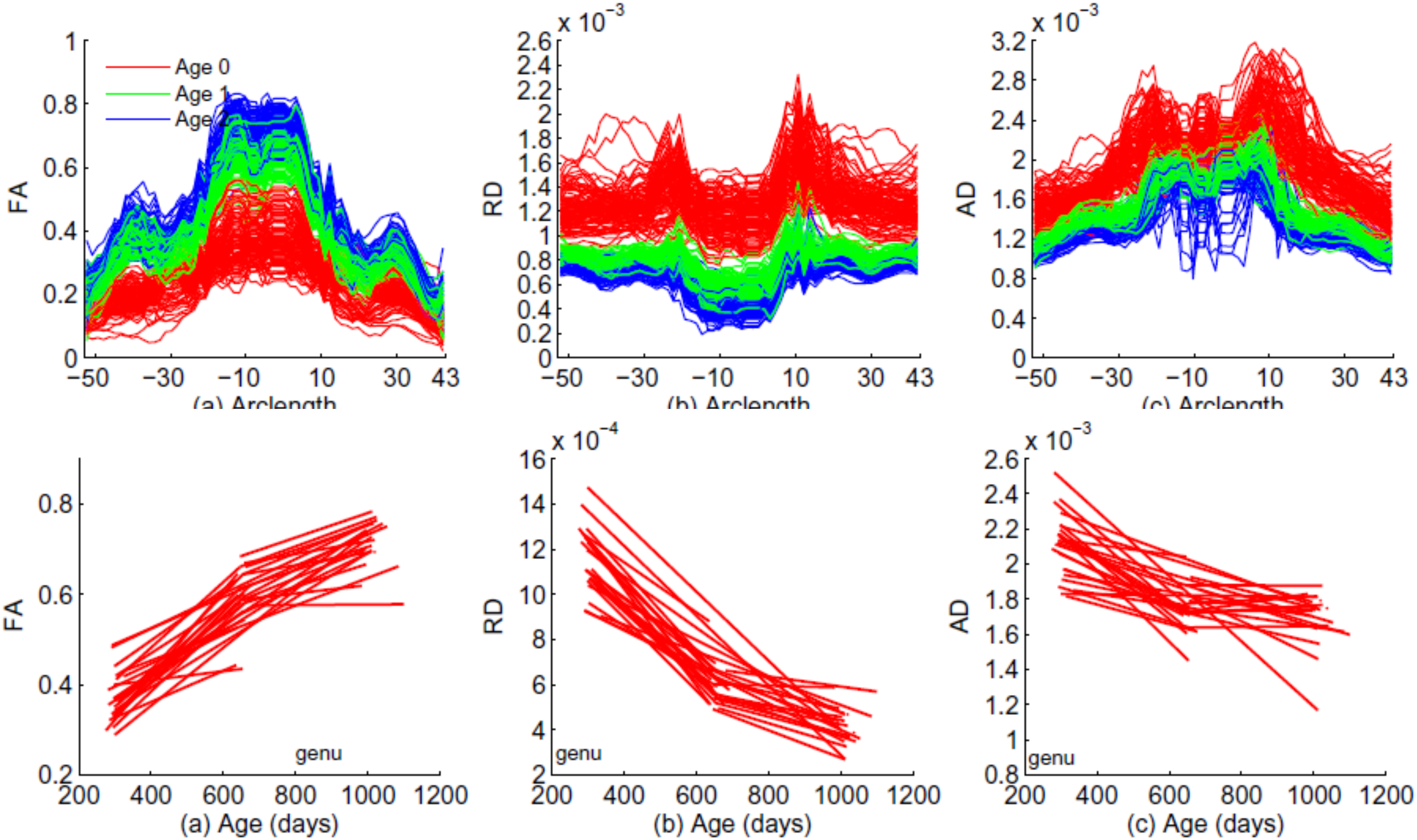
Available scans	N
Neonate scan only	1
1 year scan only	2
2 year scan only	3
Neonate + 1 year scan	43
Neonate + 2 year scan	30
1 year + 2 year scan	28
Neonate + 1 year + 2 year scan	30



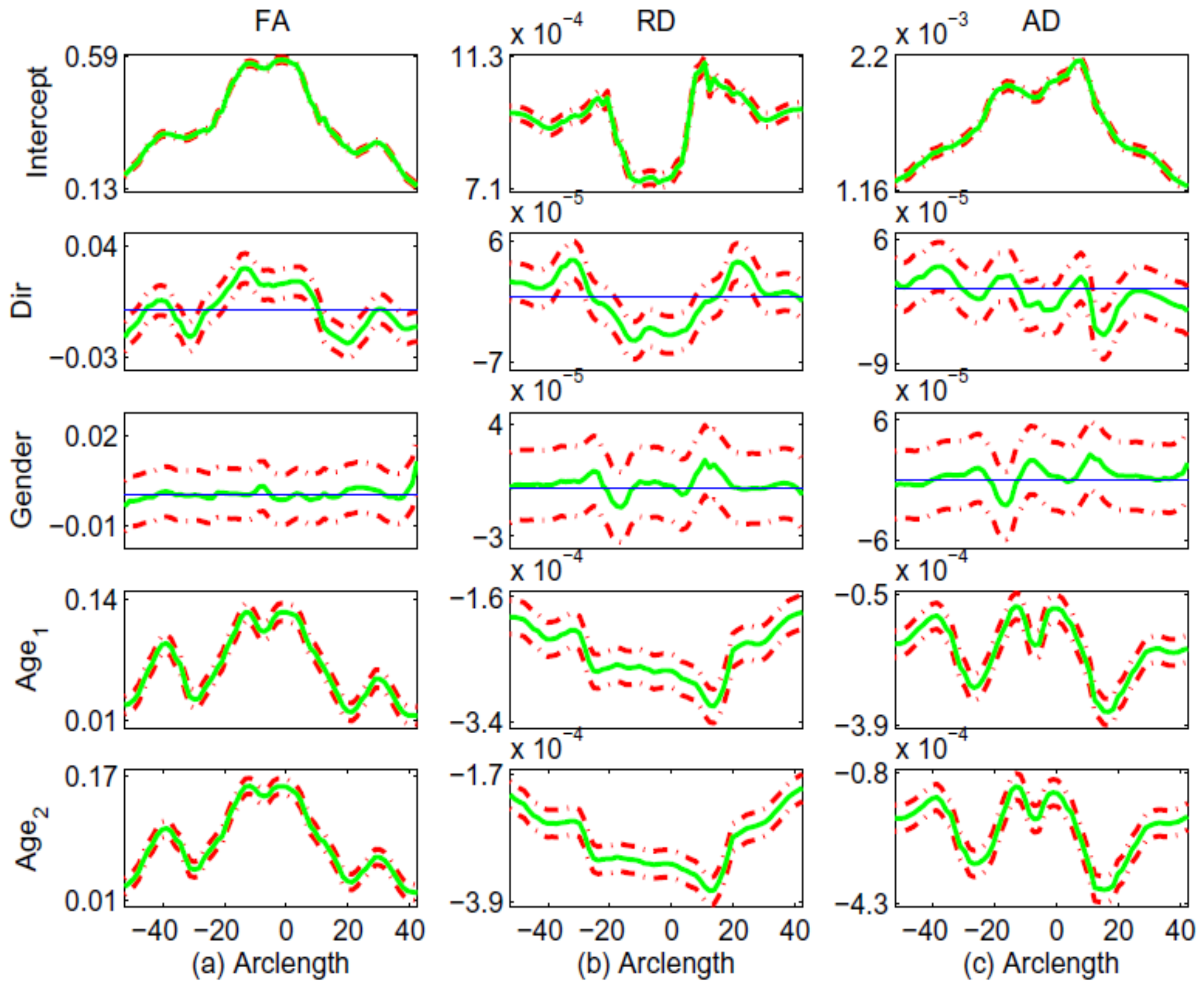
DTImaging parameters:

- 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.

Ex 1: Longitudinal Tract Data

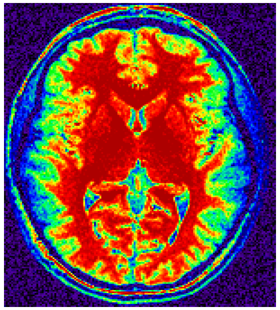
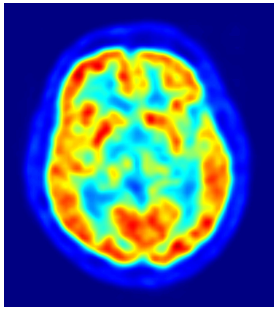
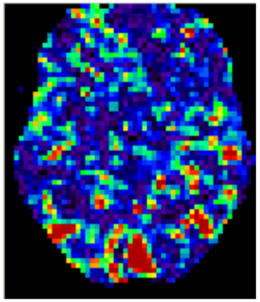
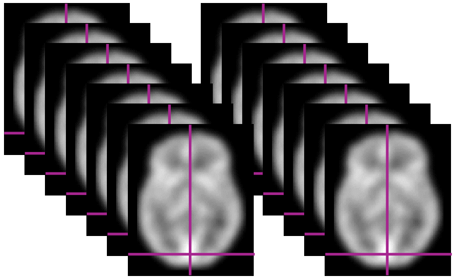


Ex 1: Longitudinal Tract Data



Neuroimaging Data with Discontinuity

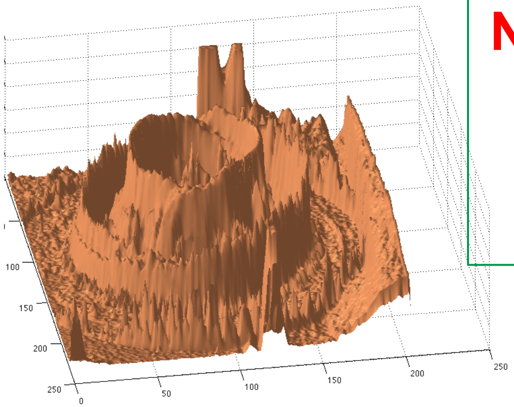
Noisy Piecewise Smooth Function with Unknown Jumps and Edges

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

Case 2: 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. $\tilde{x} \in \Omega \subseteq R^k$

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

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

M2: Spatial Varying Coefficient Model

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:

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

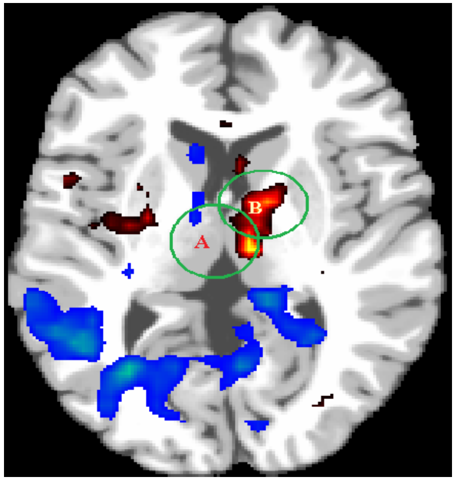
Li, Zhu, Shen, Lin, Gilmore, and Ibrahim (2011). JRSSB.
Zhu, Fan, and Kong (2014) JASA

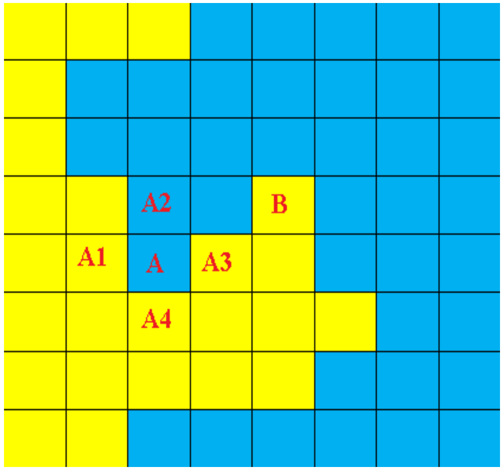
Spatial Varying Coefficient Model

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

$$y = f + \varepsilon; \qquad \varepsilon \text{ uncorrelated, mean=0, var=}\sigma^2$$

Estimate f_i as a weighted average of the noisy pixels:

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

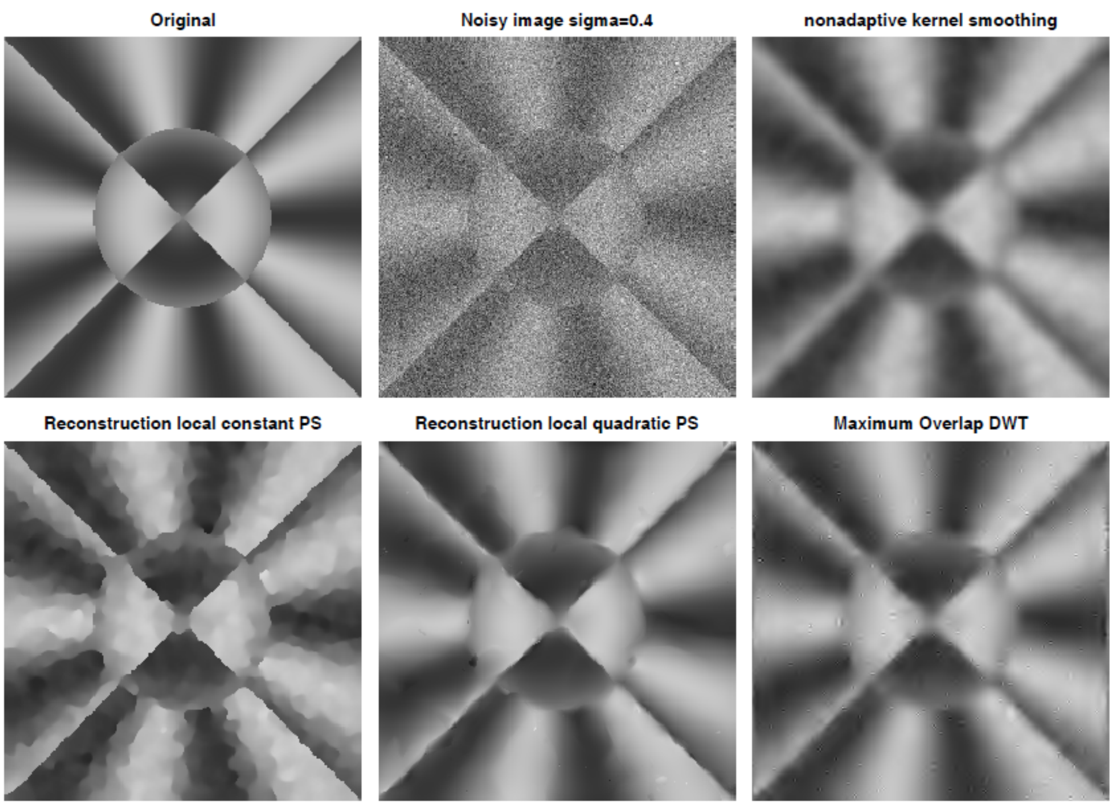
Arias-Casto, Salmon, Willett (2011)

- **Local constant/linear**
- **Yaroslavsky/Bilateral Filter**
- **Nonlocal Means**
- **PS**

Kernel-based Smoothing Methods

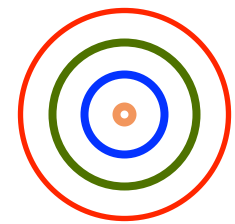
Propogation-Seperation Method

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



Features

- Increasing Bandwidth

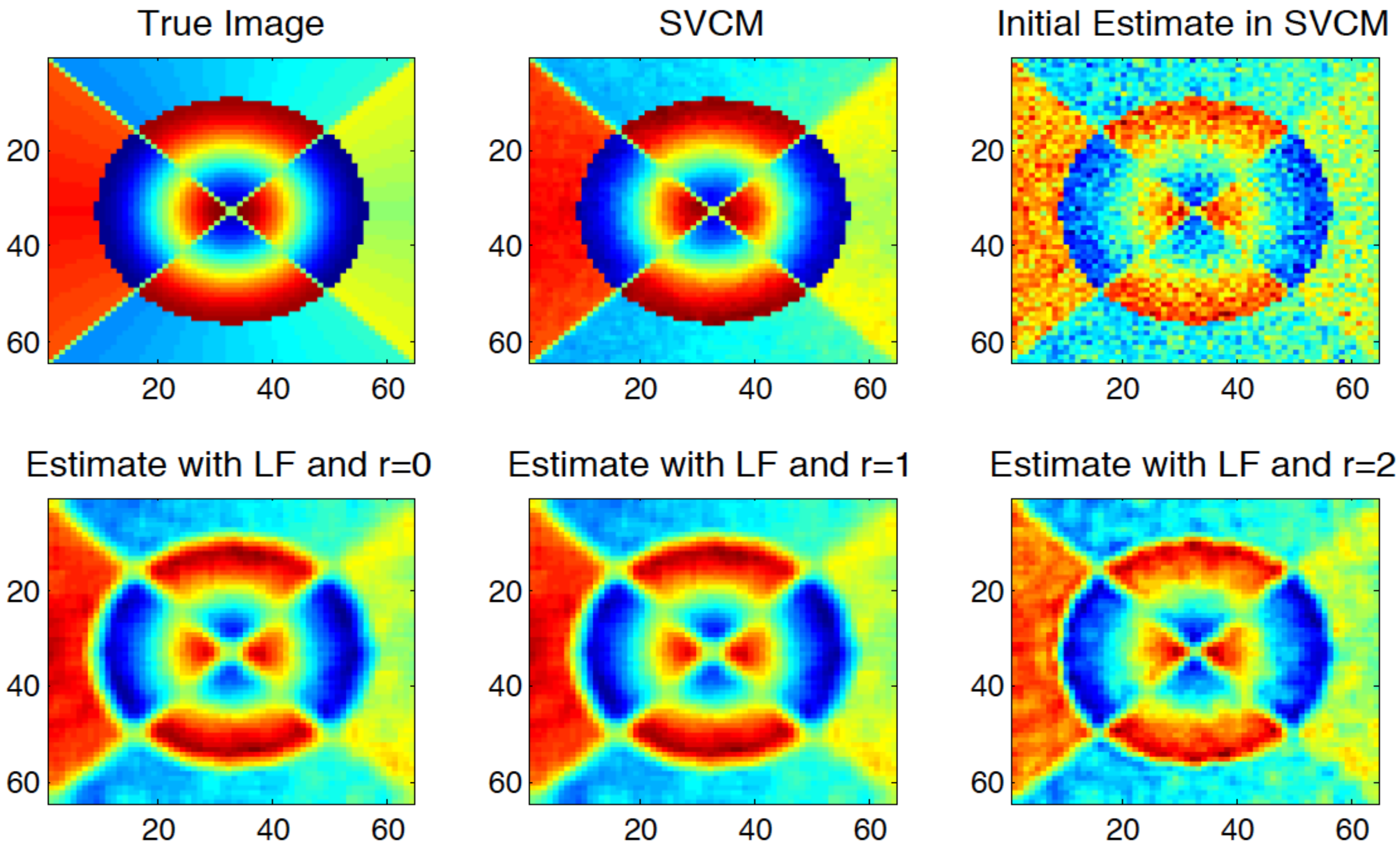


$$0 < h_0 < h_1 < \cdots < h_S = r_0$$

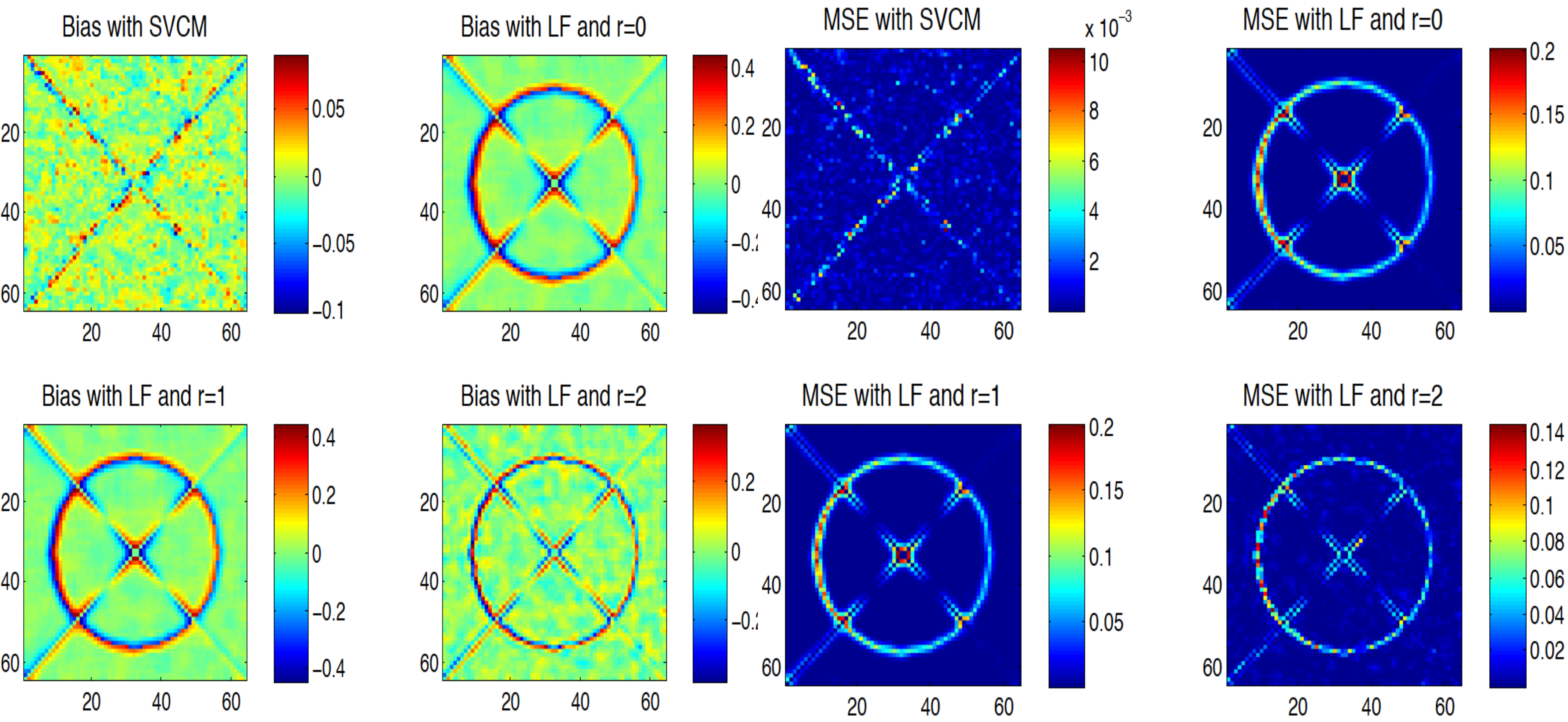
- Adaptive Weights

- Adaptive Estimates

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

EX2: ADNI PET Data

- Data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.
- We consider PET scans obtained at baseline, 6 months, and 12 months.
- Subjects are classified as having mild cognitive impairment (MCI), as AD patients, or as Normal Controls (NC).

Diagnostic status	age (years)	N male	N female
AD	75.9 ± 6.9	34	17
MCI	76.3 ± 7.3	33	25
NC	77.0 ± 4.2	30	20

- We randomly chose 80 subjects for the training set to develop the prediction model.
- We predicted the PET scans at month 12, based on the baseline and 6-month scans for 79 subjects in the test set.
- We used gender, diagnostic status (MCI, AD, NC), and age (55-90 years) as covariates for the semi-parametric model.

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

EX2: ADNI PET Data

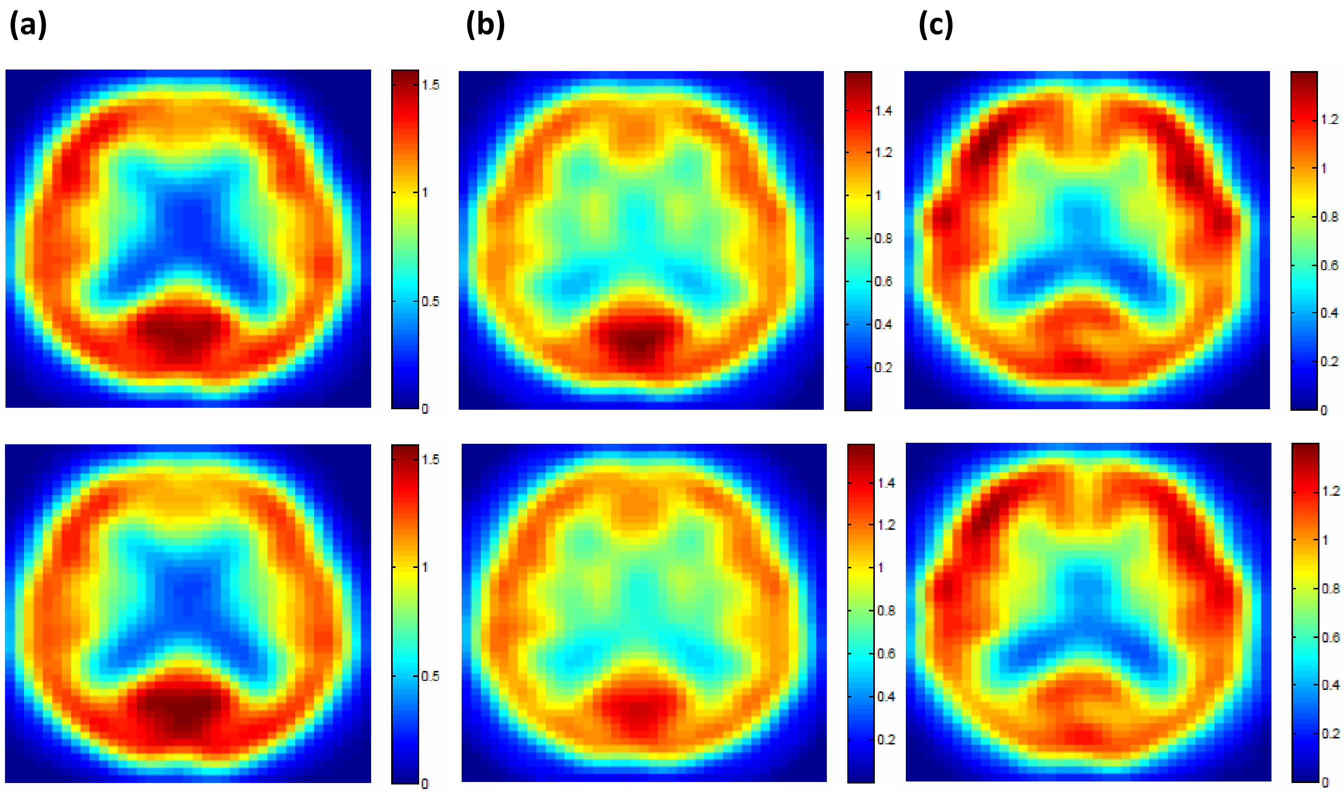


Figure : Observed (upper panel) and predicted (bottom panel) PET images at month 12 for (a) an AD patient, (b) an MCI subject, and (c) a NC subject. One selected slice is shown.

EX2: ADNI PET Data

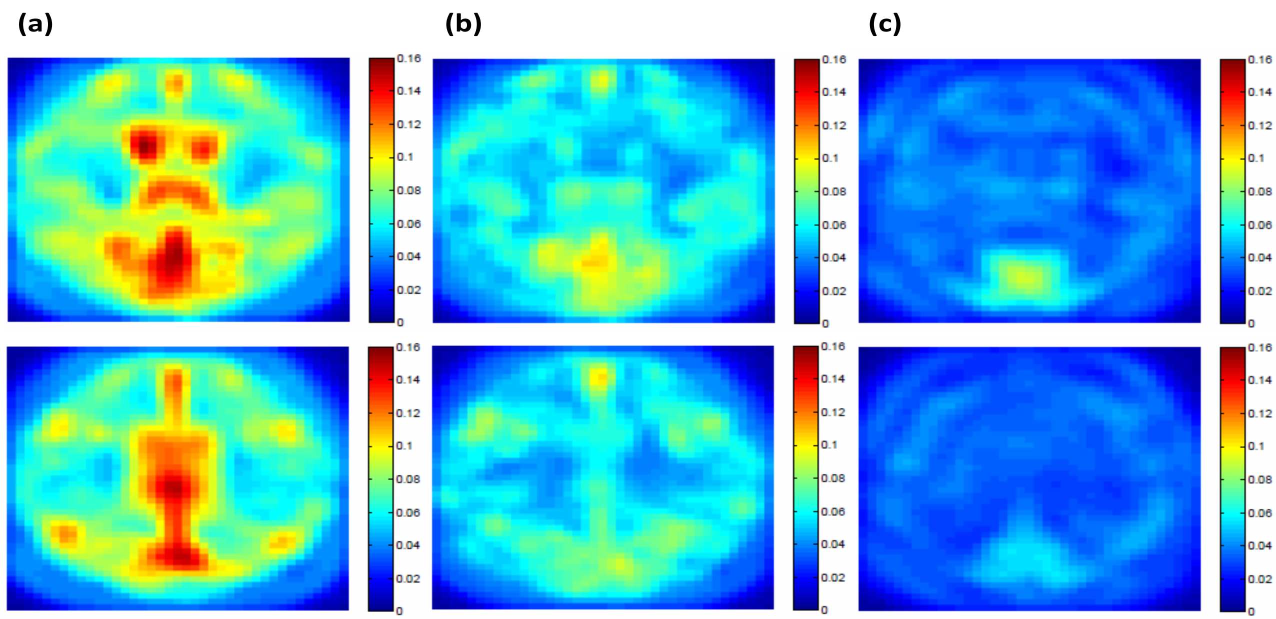


Figure : rtMSPE maps for prediction of ADNI PET images at month 12 for 79 test subjects. Selected slices are shown for (a) Semi-parametric model; (b) Semi-parametric model+FPCA; (c) Semi-parametric model+FPCA+Spatial-temporal model.

Semi-parametric model	0.0692
Semi-parametric mode+FPCA	0.0550
Semi-parametric model+FPCA+Spatial-temporal model	0.0354

Image-on-Genetic Association Models

The NIMH Strategic Plan

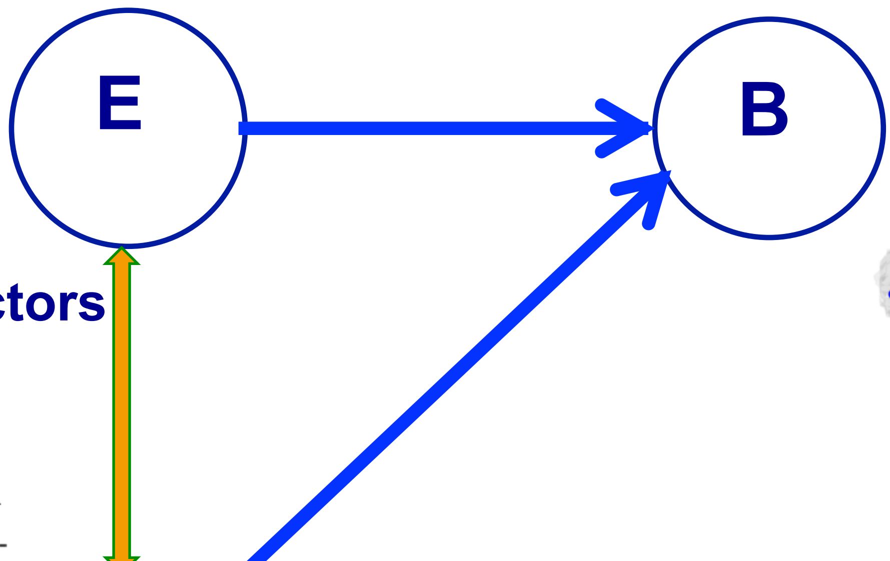
Strategic Objective 1: Promote Discovery in the Brain and Behavioral Sciences to Fuel Research on the Causes of Mental Disorders

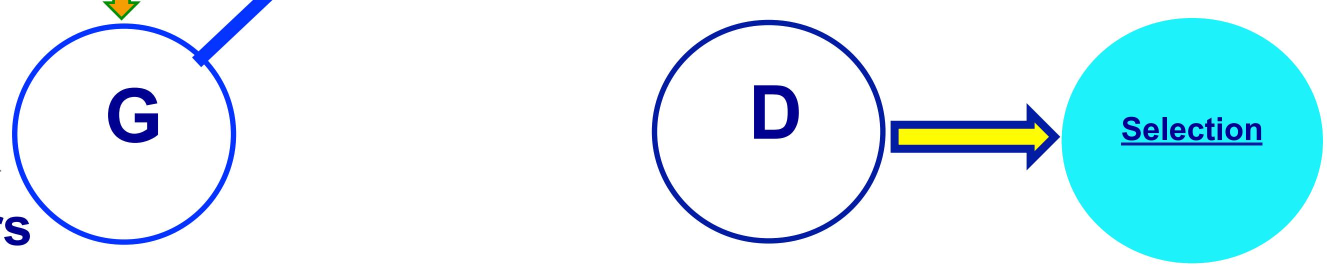
Identify the genetic and environmental factors associated with mental disorders.

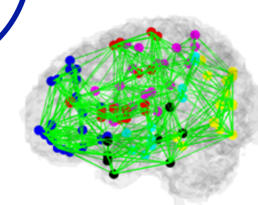
Strategic Objective 2: Chart Mental Illness Trajectories to Determine When, Where, and How to Intervene

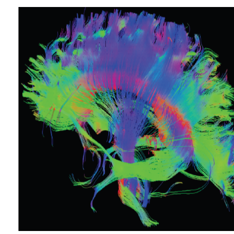
When identifying behavioral, neural, and/or genetic markers along the trajectory of illness, design the studies to consider variation in relation to age, sex, gender, race, ethnicity, and other important socio-demographic factors.

Big Data Integration

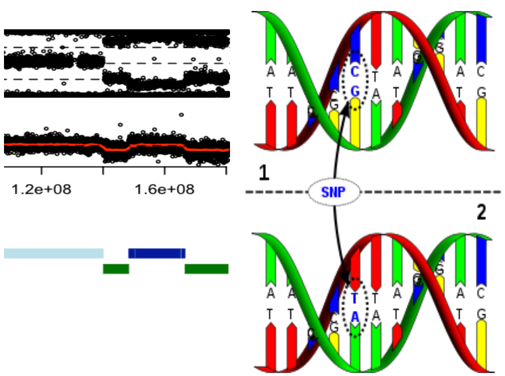








E: environmental factors

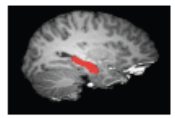
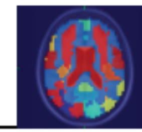
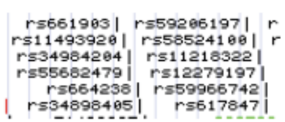
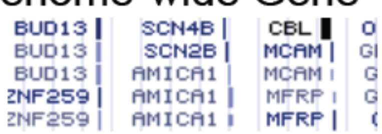


G: genetic markers

D: disease

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

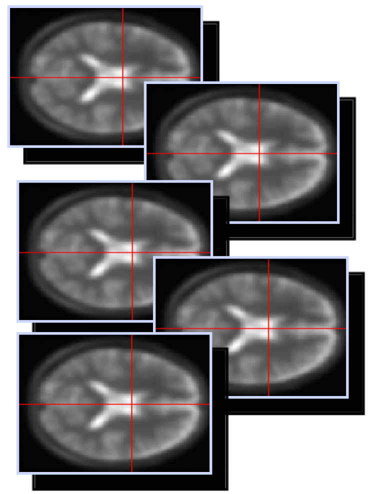
Statistical Methods

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

Hibar, et al. HBM 2012

Data Structure

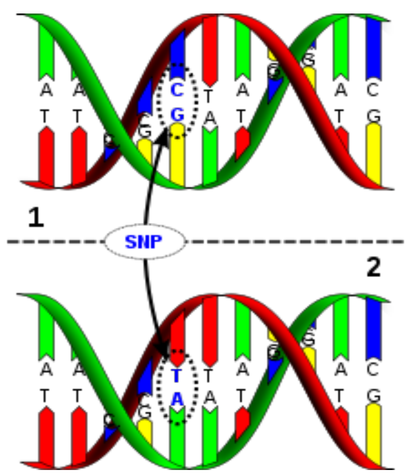
Imaging:



Person No.1 ······· Person No.1000

3D Matrix ······· 3D Matrix

Genetic：



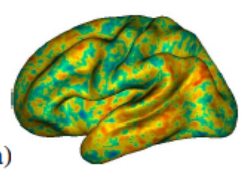
$$
\begin{array}{c} \\ \text{Person No. 1} \\ \cdot \\ \cdot \\ \cdot \\ \text{Person No. 100} \end{array}
\begin{array}{c}
\begin{array}{cccc} \text{SNP1} & \text{SNP2} & \ldots\ldots. & \text{SNP} \end{array} \\
\begin{bmatrix} 1 & 2 & \cdots & 0 \\ 0 & \ddots & & 1 \\ \vdots & & \ddots & \vdots \\ 1 & 0 & \cdots & 2 \end{bmatrix}
\end{array}
$$

M3: High Dimensional Regression Model

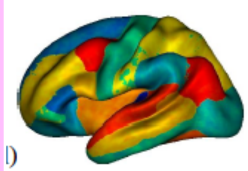
Data $\{(Y_i, X_i) : i = 1, \cdots, n\}$

$$Y_i = \{y_i(v) : v \in V\} \qquad \{X_i(g) : g \in G_0\}$$

Phenotype		Genotype			Error
Y		X	B		E
$n \times p_y$	=	$n \times p_x$	$p_x \times p_y$	+	$n \times p_y$



(a)



$$(p_x, n, p_y)$$

Sparse Projection Regression Model

- Let $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_k]$, then a projection regression model is given by:

$$\mathbf{W}^T y_i = (\mathbf{B}\mathbf{W})^T \mathbf{x}_i + \mathbf{W}^T \mathbf{e}_i = \beta_{\mathbf{w}}^T \mathbf{x}_i + \varepsilon_i$$

- Hypothesis problem reduces to:

$$H_{0W} : \mathbf{C}\beta_{\mathbf{w}} = \mathbf{b}_0 \quad v.s. \quad H_{1W} : \mathbf{C}\beta_{\mathbf{w}} \neq \mathbf{b}_0$$

$$\textit{where } \mathbf{C}\beta_{\mathbf{w}} = \mathbf{CBW} \textit{ and } \mathbf{b}_0 = \mathbf{B}_0\mathbf{W}$$

- How to determine an 'optimal' $\mathbf{W}$?

Sun, Zhu, Liu, and Ibrahim (2014) JASA

Sparse Projection Regression Model

- We show that this is achieved by optimizing the following generalized heritability ratio (GHR):

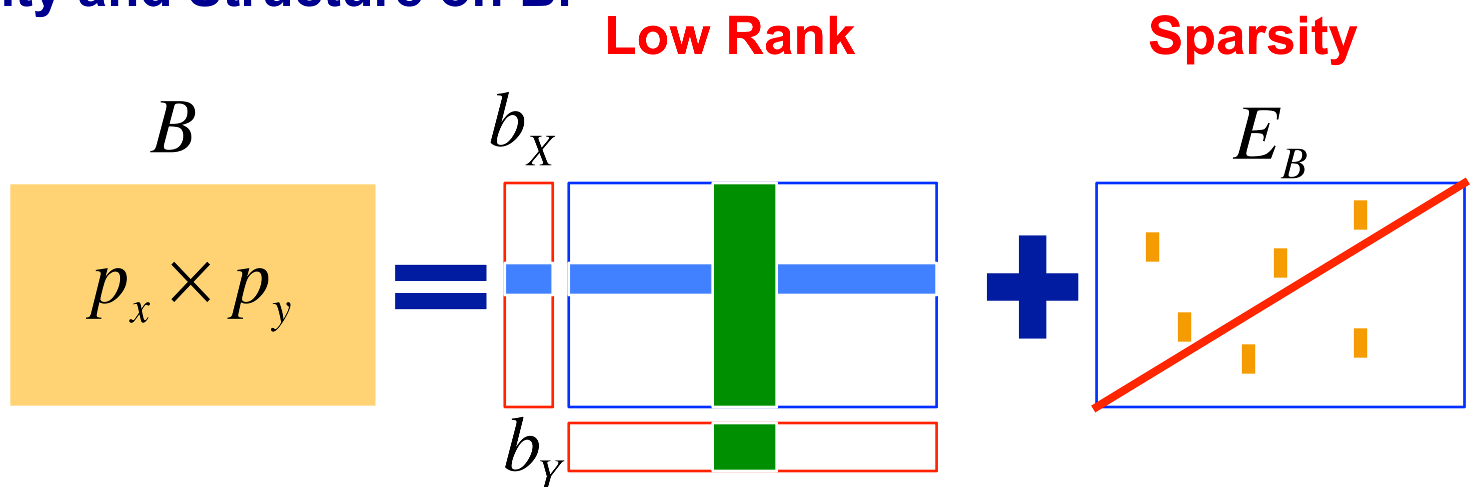
$$\mathrm{GHR}(\mathbf{w};\mathbf{C}) = \frac{\mathbf{w}^T(\tilde{\mathbf{B}}_1-\mathbf{B}_0)^T S_{\tilde{X}_1}(\tilde{\mathbf{B}}_1-\mathbf{B}_0)\mathbf{w}}{\mathbf{w}^T\Sigma_R\mathbf{w}} = \frac{\mathbf{w}^T\Sigma_C\mathbf{w}}{\mathbf{w}^T\Sigma_R\mathbf{w}}$$

- High Dimensional Setting
- noise accumulation
  - ill-conditioned sample covariance estimator: $\hat{\Sigma}_R$
- Sparse Projection Regression Model is proposed as following:

$$\mathrm{argmax}\{\frac{\mathbf{w}^T\hat{\Sigma}_C\mathbf{w}}{\mathbf{w}^T\tilde{\hat{\Sigma}}_R\mathbf{w}}\} \quad \text{s.t.} \quad ||\mathbf{w}||_1 \le t$$

Sparse and Low-rank Representation

Sparsity and Structure on B.



$$p_\lambda(B) \rightarrow p_\lambda(b_X) + p_\lambda(b_Y) + p_\lambda(E_B)$$

Regularization Methods

- **Lasso 1, 2, 3, ….**
- **SCAD, MCP, …..**

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

Shen, Shen, and Zhu (201?)

Factor Model

E

$n \times p_y$

Long-range Correlation

Short-range Correlation

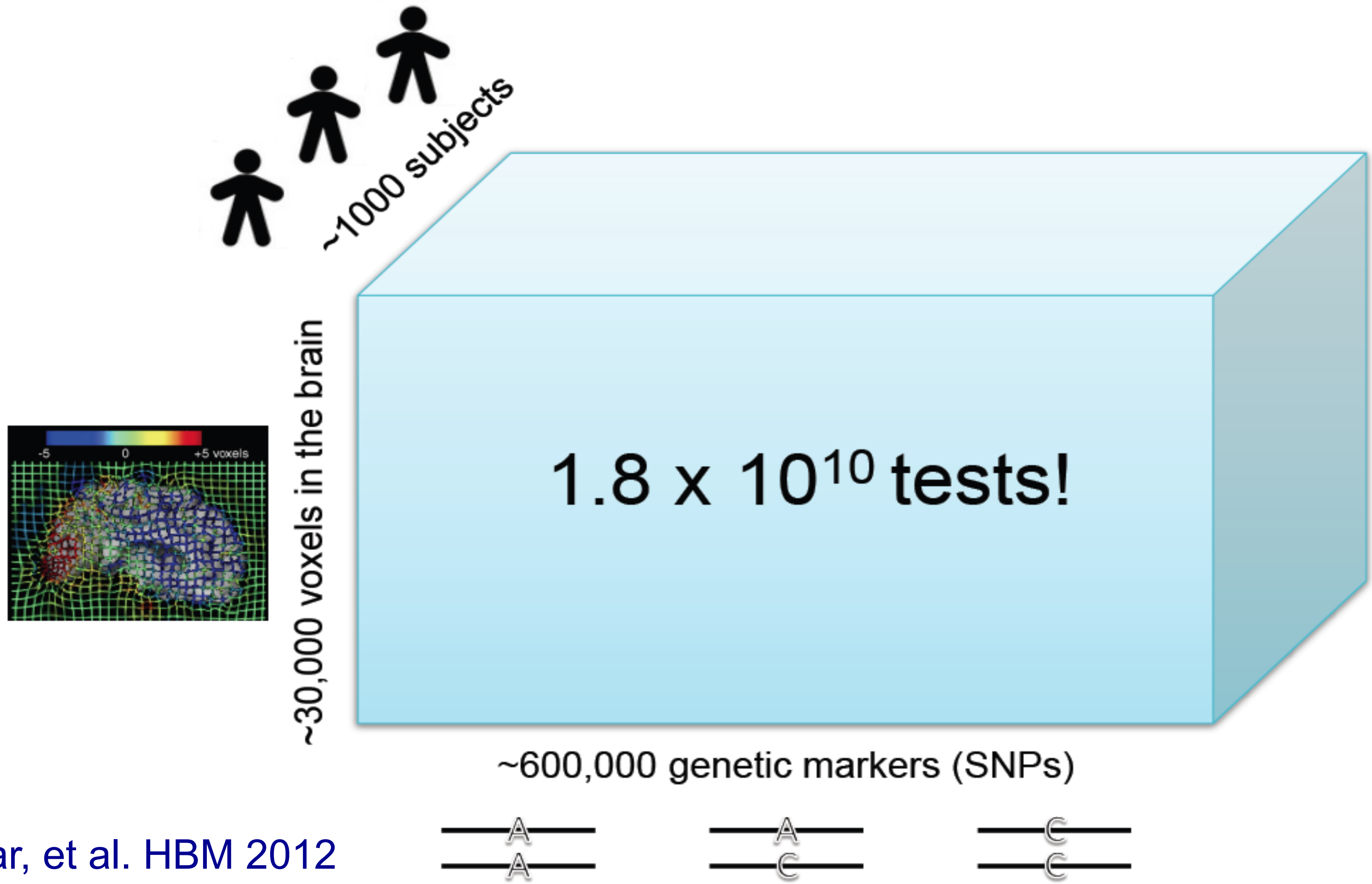
$$E_i = \Lambda \, \xi_i + \eta_i$$

$p_y \times 1 \qquad p_y \times q \qquad q \times 1 \qquad p_y \times 1$

$$\Sigma_E = \Lambda \, \Lambda^T + \Sigma_\eta$$

Zhu, Zakaria, Lu, and Ibrahim (2014) JASA

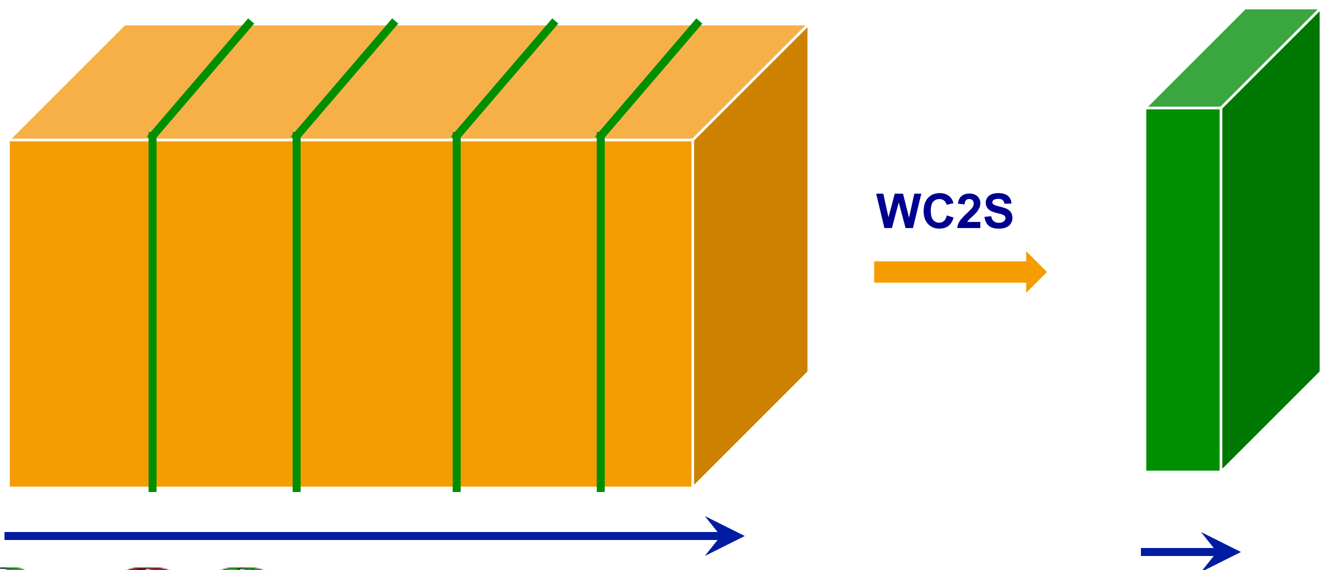
M4: Voxel-wise GWAS

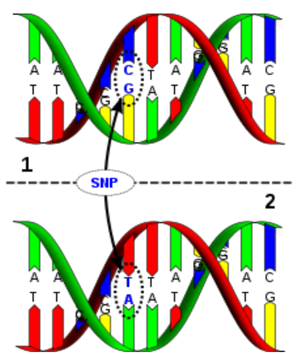


Hibar, et al. HBM 2012

M4: Voxel-wise GWAS

Fast Sure-Independence Screening Procedure



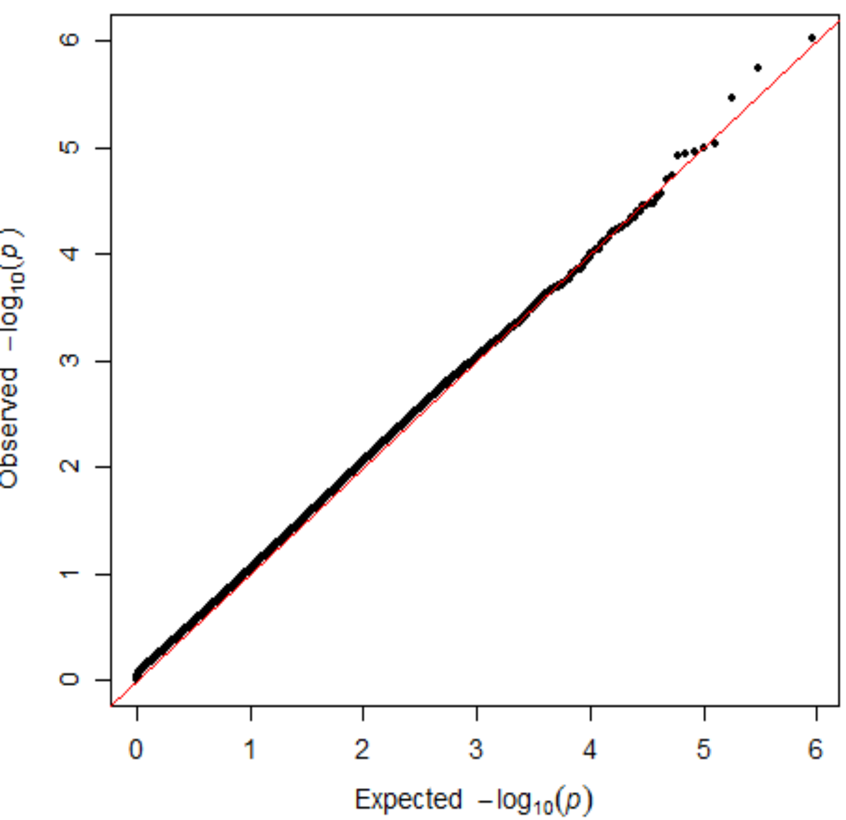


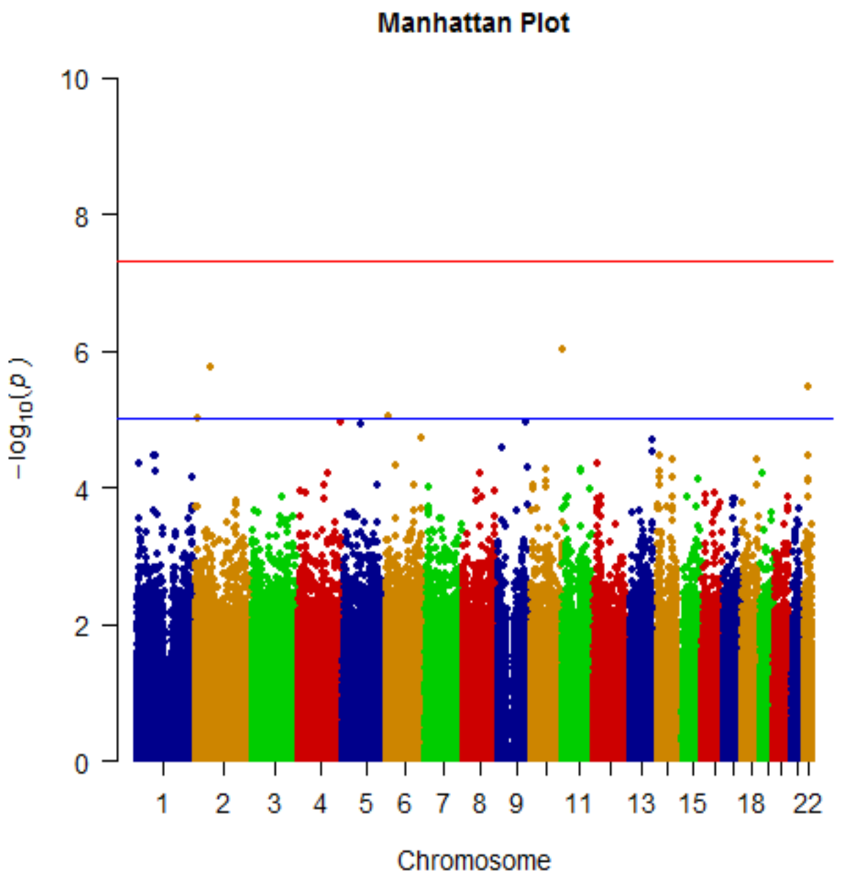
SNP

Each SNP

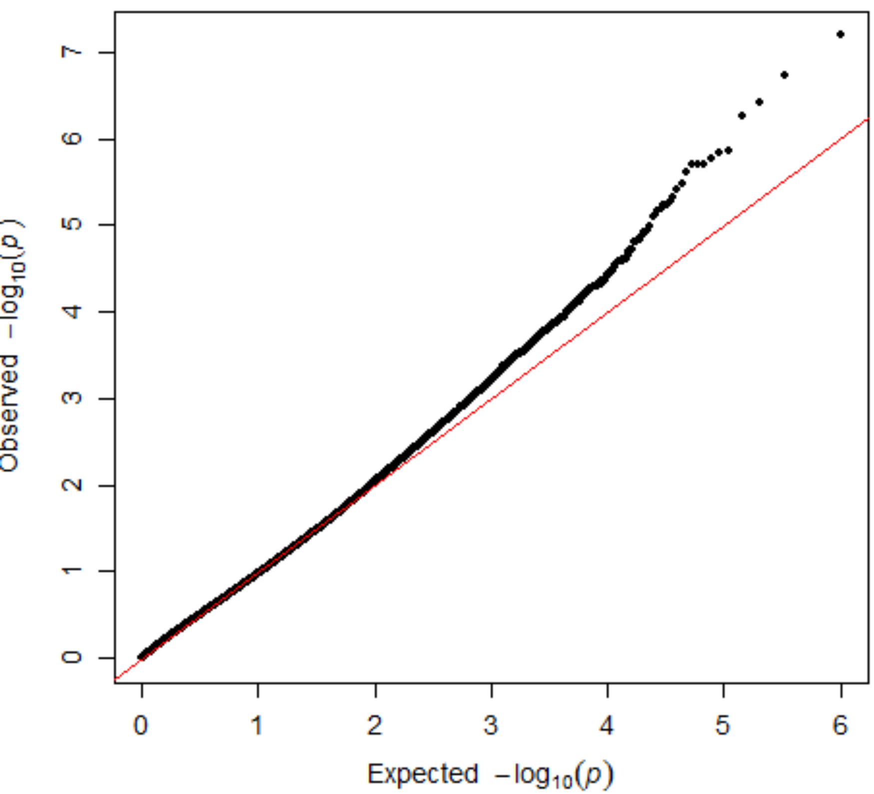
Huang,…. and Zhu (2014)

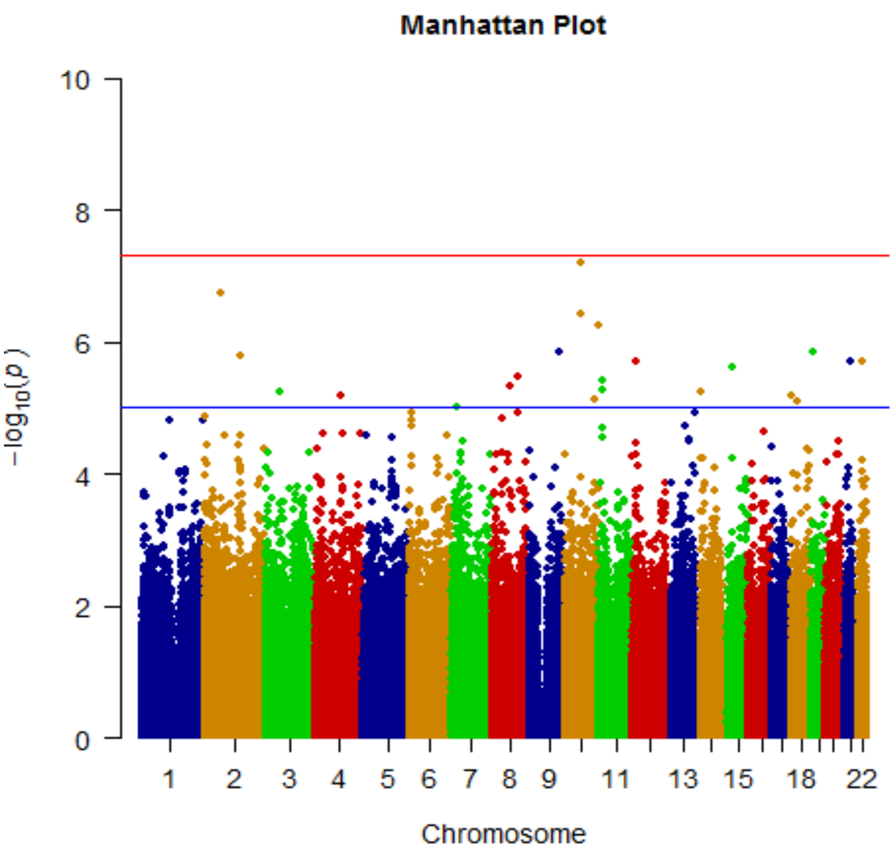
EX3: 93 ROI-GWAS





EX4: Whole Brain-GWAS





Prediction Models

Alzheimers Disease Big Data DREAM Challenge 1

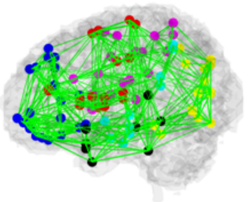
Its goal is to apply an open science approach to rapidly identify **accurate predictive AD biomarkers** that can be used by the scientific, industrial and regulatory communities to improve AD diagnosis and treatment.

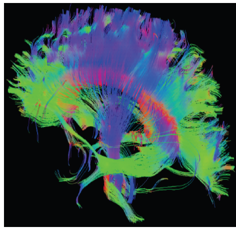
Sub 1: Predict the change in cognitive scores 24 months after initial assessment.

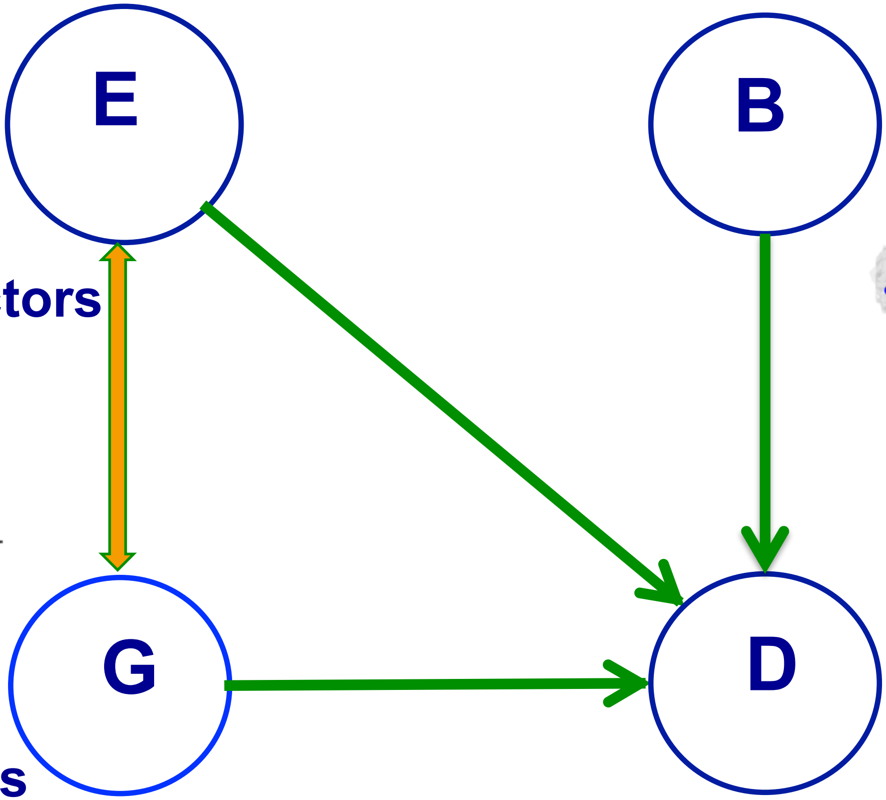
Sub 2: Predict the set of cognitively normal individuals whose biomarkers are suggestive of amyloid perturbation.

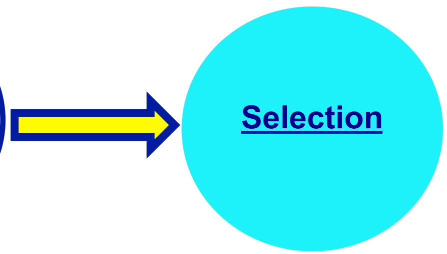
Sub 3: Classify individuals into diagnostic groups using MR imaging.

Big Data Integration

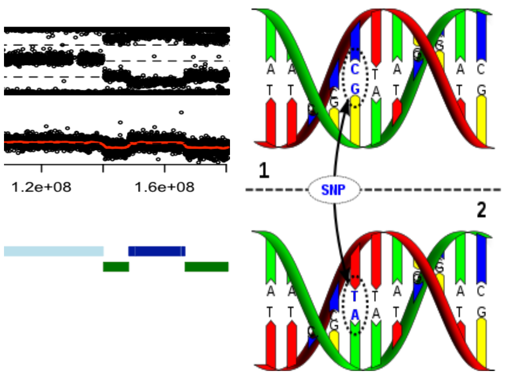








E: environmental factors



G: genetic markers

D: disease

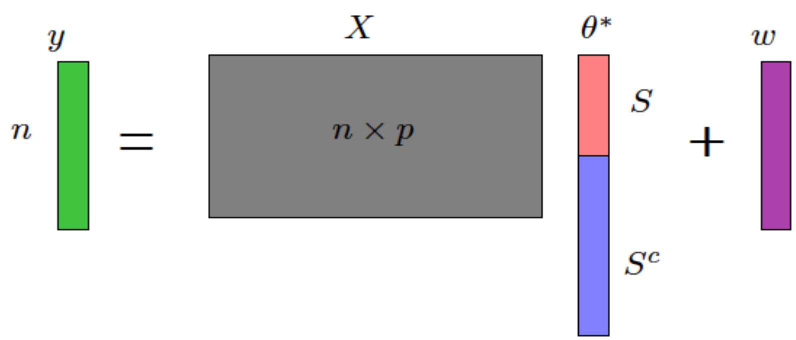
http://en.wikipedia.org/wiki/DNA_sequence

HRM versus FRM

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

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

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

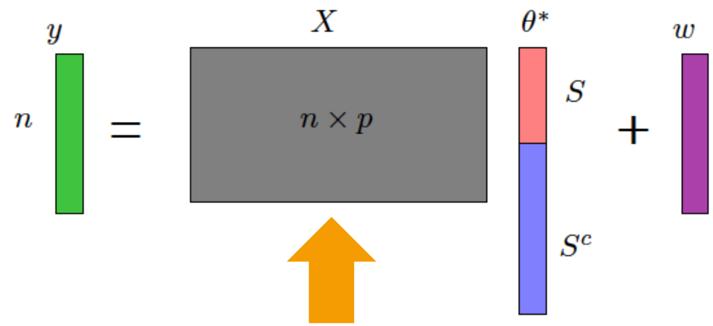


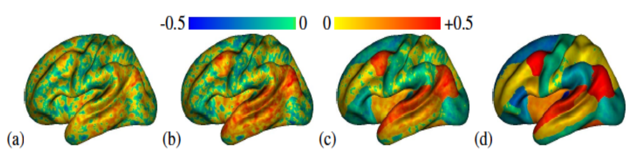
Strategy 2: Functional Regression Model (FRM)

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

High-dimension Regression Model

Approach 1: Regularization Methods





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

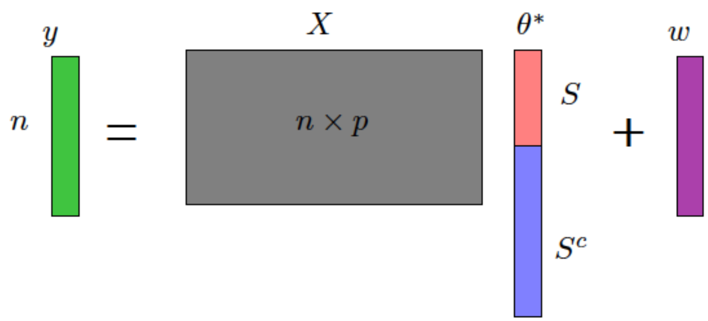
Key Conditions:

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

High-dimension Regression Model

Tensor Structure:

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

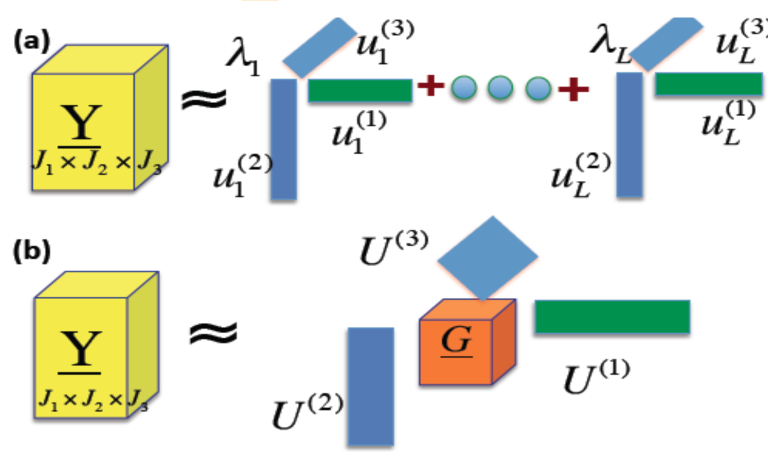


Zhou, Li, and Zhu (2013)
Li, Zhou, and Li (2013)

CP decomposition

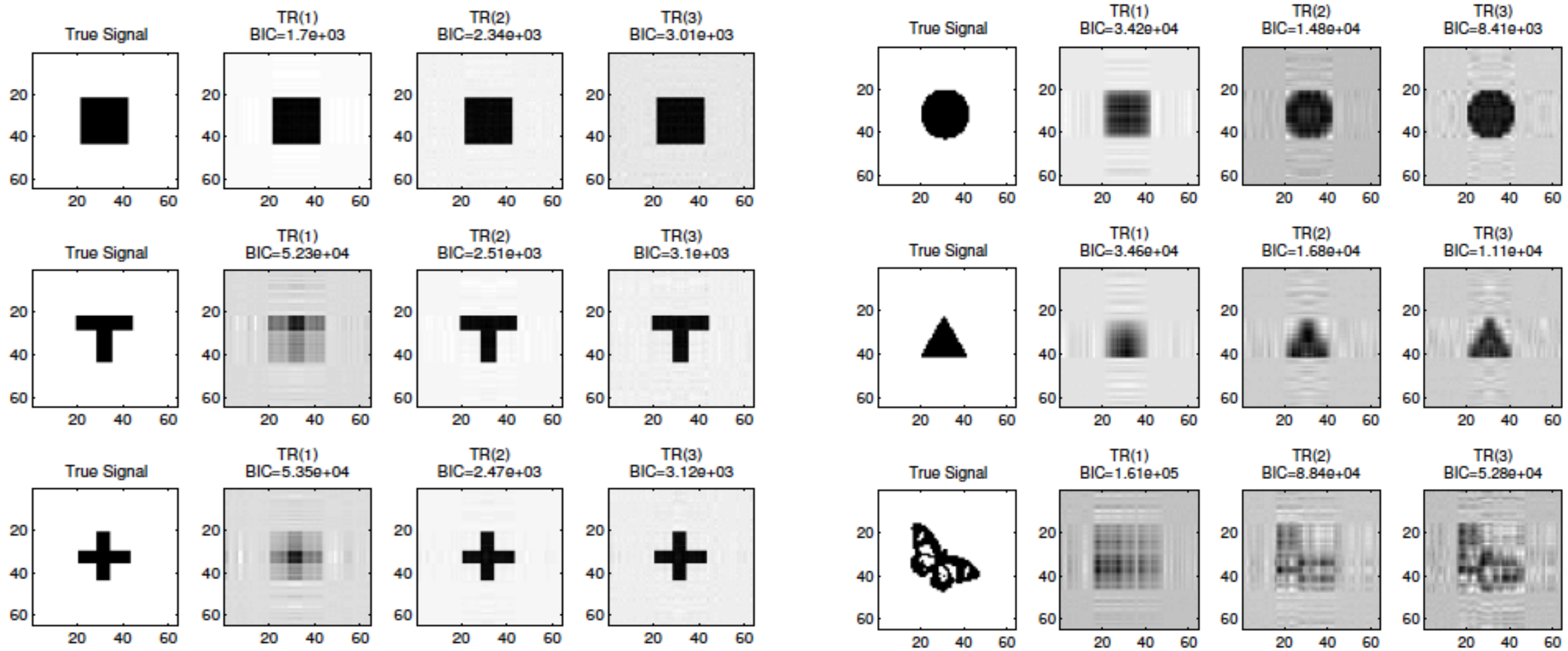
Tucker decomposition

θ



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

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

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

- M5: Functional linear Cox regression models
- M6: Generalized scalar-to-image regression models
- M7: Multiscale Functional Linear models

M5: Functional Linear Cox Regression Model

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

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

Model

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

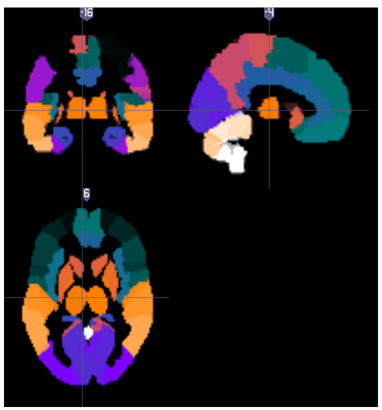
- $$X_i(s) = \mu(s) + \sum_{j=1}^{\infty} \xi_{ij}\phi_j(s) + \varepsilon_i(s)$$

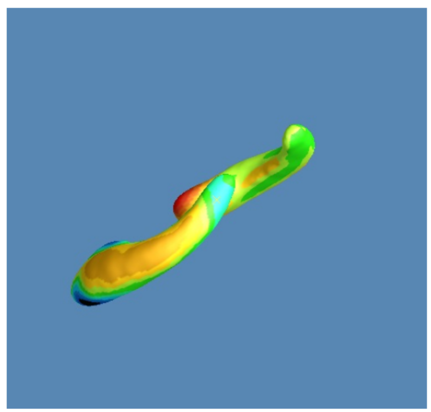
  - **Consistency**
  - **Asymptotic distribution of score test**

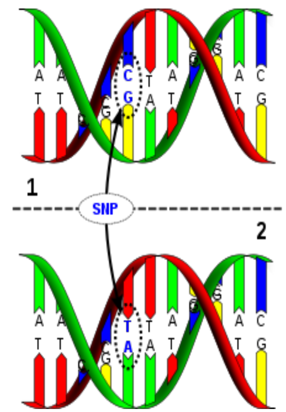
M5: Functional Linear Cox Regression Model

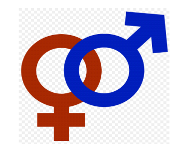
Mild Cognitive Impairment subjects

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

M6: 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 $y \sim \text{ exponential family}(\mu, \phi)$

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

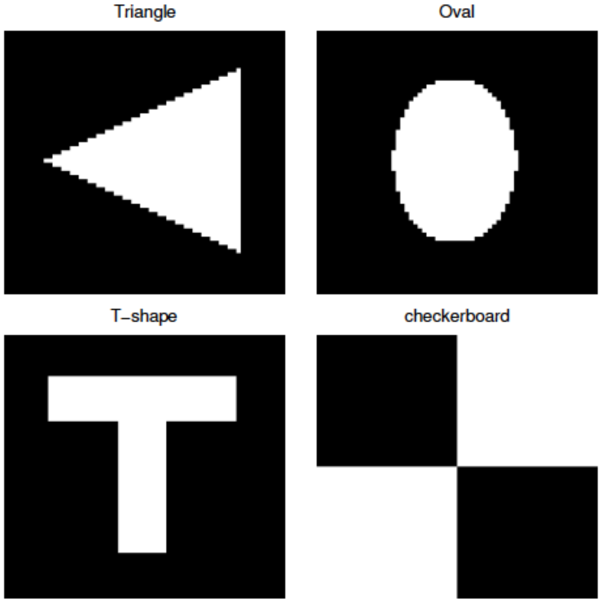
Total Variation

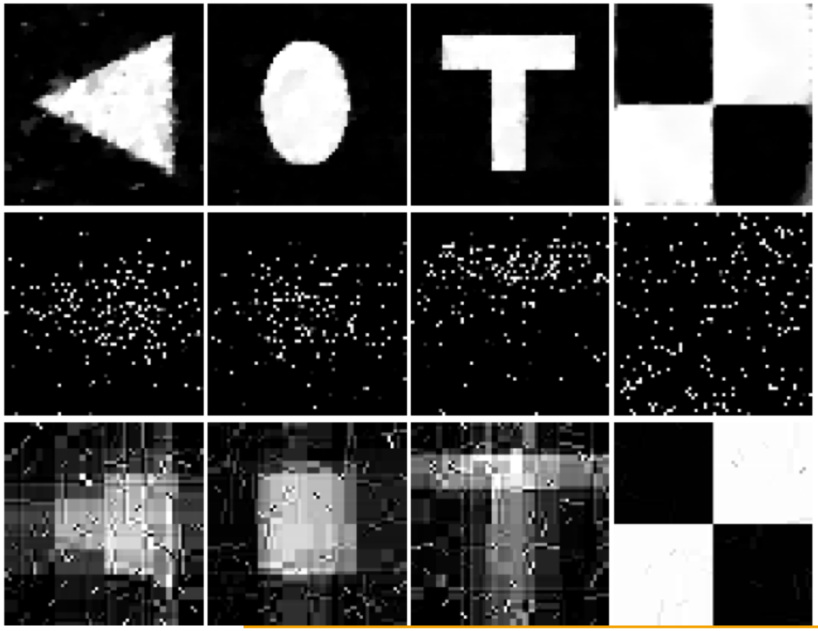
Estimation: $\sum_{i=1}^{n} \ell(y_i; \mu(X_i; \gamma, \beta(\bullet))) + \lambda \| \beta \|_{TV}$

Non-asymptotic Error Bound:

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

M6: Generalized scalar-to-image regression models





TV

Lasso

Lasso on WL

M7: Multiscale Functional Linear models

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

Models

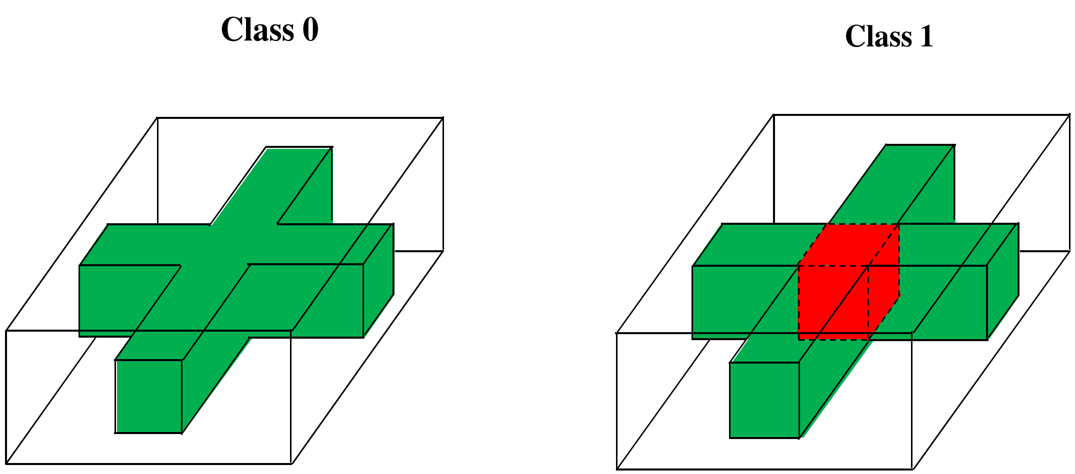
(A1) $D = (\bigcup_{k=1}^{K} D_k) \bigcup D_0$

- Informative sets + Irrelevant set

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

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

Simulation I: Classification



0 White
1 Green
2 Red

$$X_i(d) = \beta_0(d) + \beta_1(d)y_i + \varepsilon_i(d)$$

Type I

$N(0,4)$

Type II

Short-range correlation

Type III

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

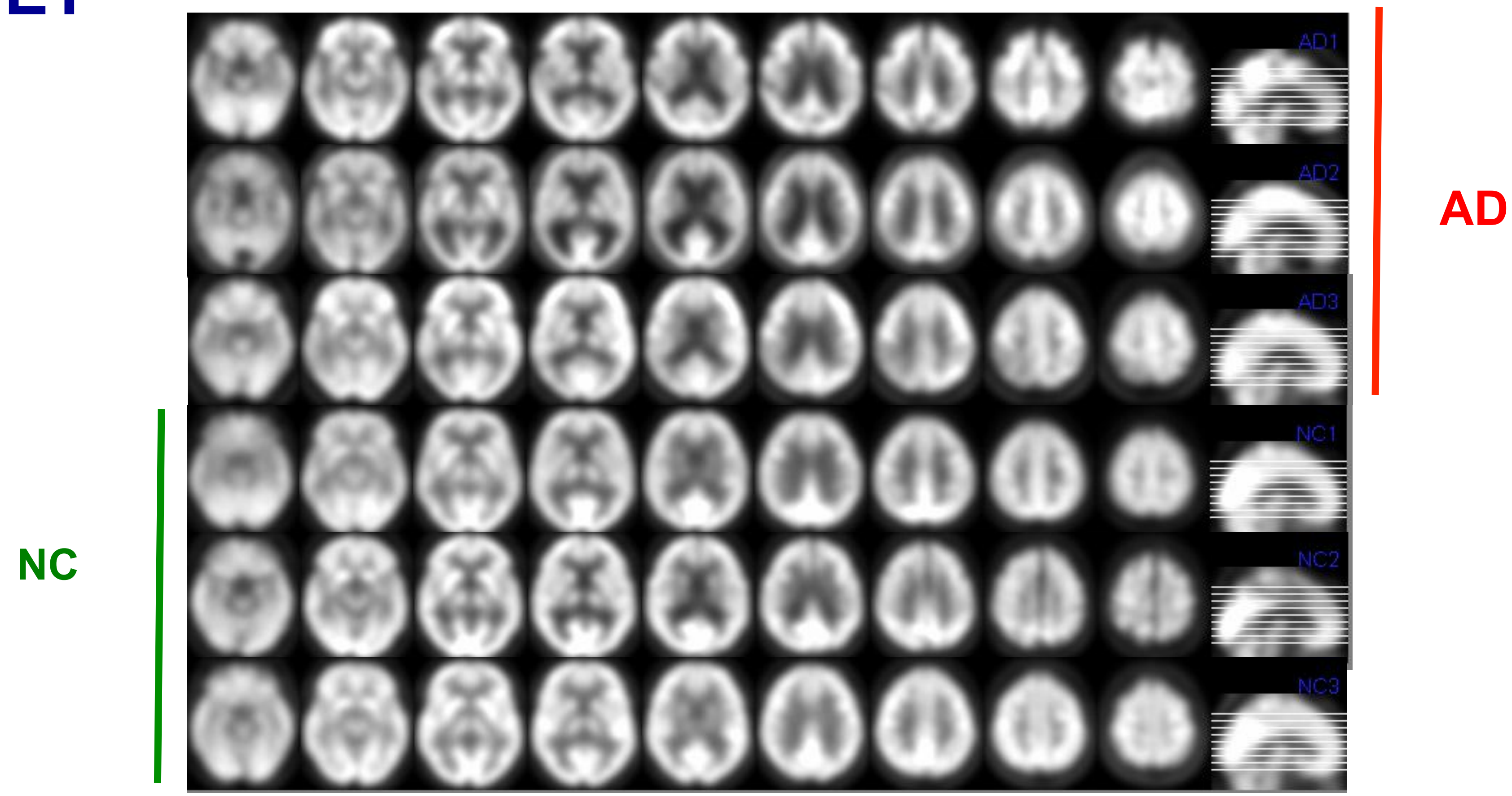
Simulation I: Classification

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:

EX5: ADNI-PET

PET



ADNI

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

Thank You!!

ASA: Statistics in Imaging Section

SAMSI

2013 Neuroimaging Data Analysis

2015-2016 Challenges in Computational Neuroscience

July 27-31 Summer School   August 17-21 Opening Workshop

- Shape Analysis
- Spike Train Analysis
- Big Data Integration
- Compressed Sensing
- Functional Data Analysis