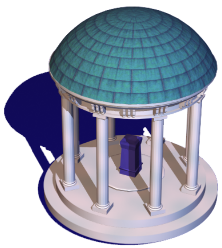


Spatial and Adaptive Models for Neuroimaging Data

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

UNC
SCHOOL OF
PUBLIC HEALTH

The UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Motivation: Neuroimaging Data

Large Neuroimaging Data

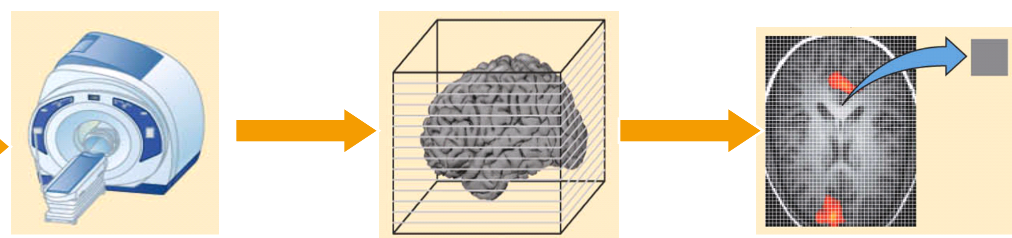
NIH normal brain development

1000 Functional Connectome Project

Alzheimer's Disease Neuroimaging Initiative

National Database for Autism Research (NDAR)

Human Connectome Project



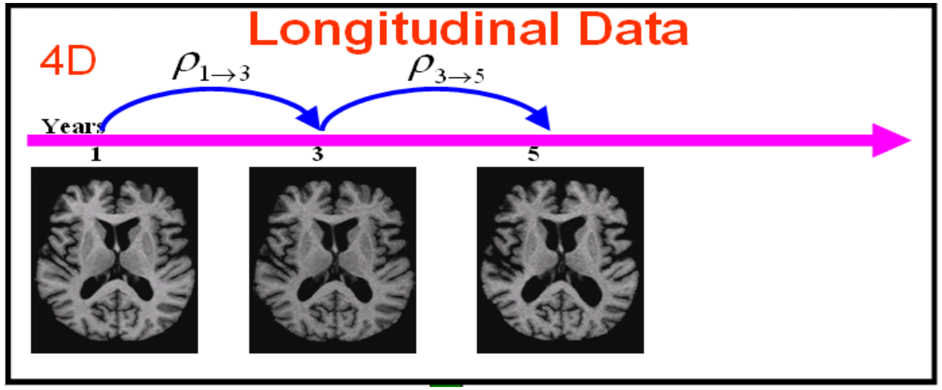
www.guysandstthomas.nhs.uk/.../T/Twins400.jpg

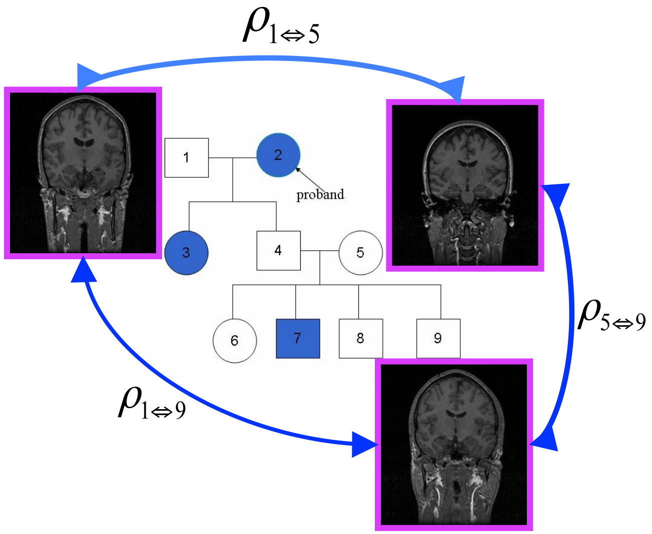
Complex Study Design

cross-sectional studies;

clustered studies including

longitudinal and twin/familial studies;





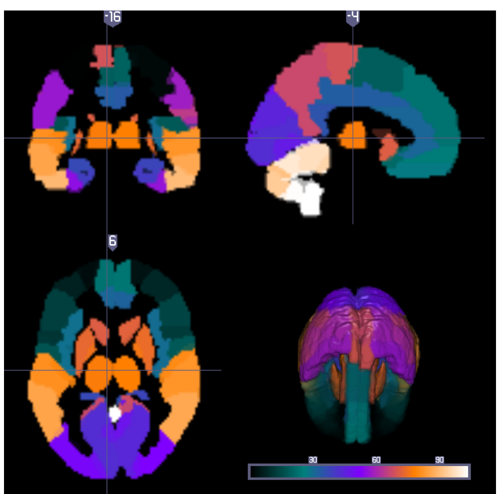
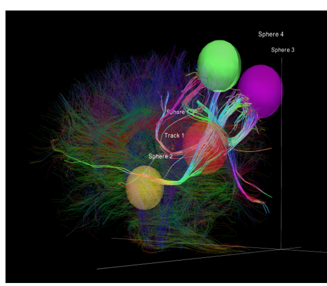
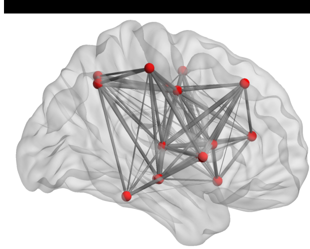
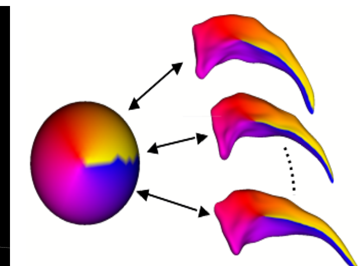
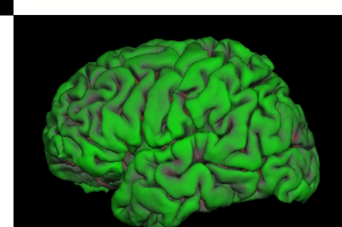
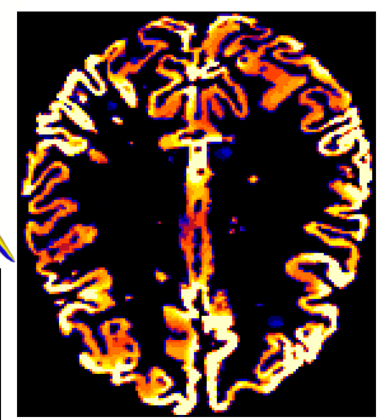
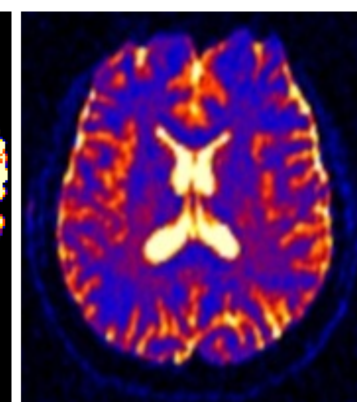
Complex Data Structure

Multivariate Imaging Measures

Smooth Functional Imaging Measures

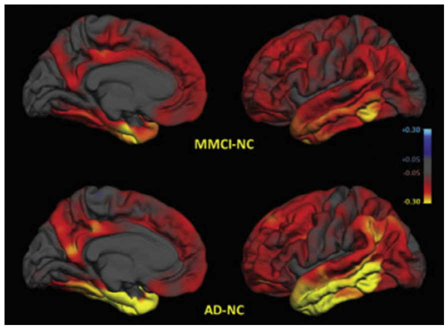
Whole-brain Imaging Measures

4D-Time Series Imaging Measures

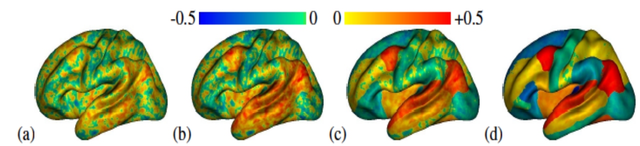
Group Analysis Applications

Group Differences



Longitudinal/Family Brain

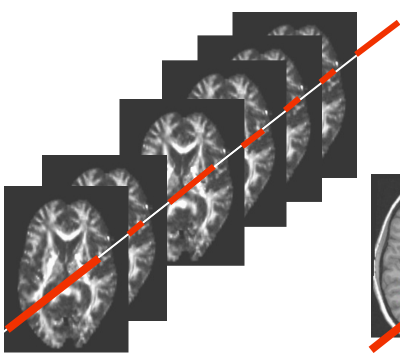
Prediction



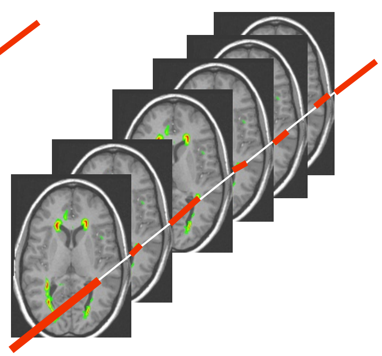
NC/Diseased

Multimodal Analysis

DTI



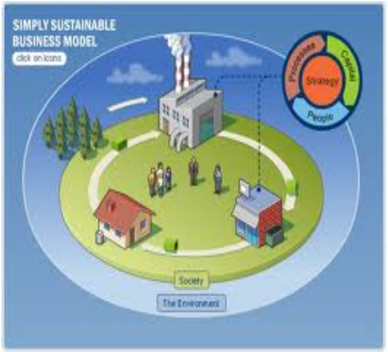
FLAIR

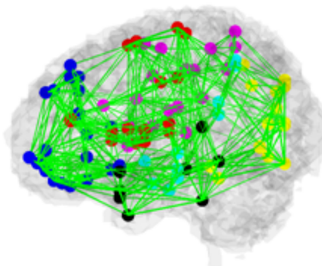


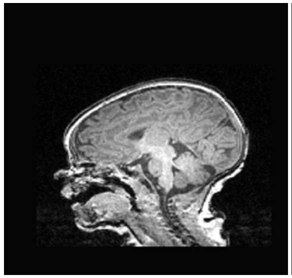
Imaging Genetics

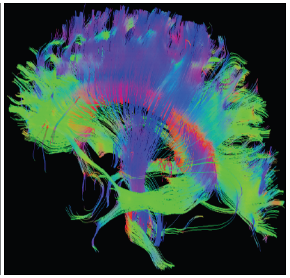
Imaging / Genetics	Candidate ROI	Many ROI	Voxelwise
Candidate SNP	Imager	Imager	Imager
Candidate Gene	Geneticist		
Genome-wide SNP	Geneticist		
Genome-wide Gene	Geneticist		

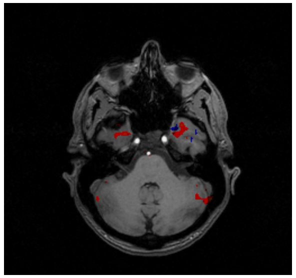
Directed Acyclic Graphs for Imaging Genetic Studies

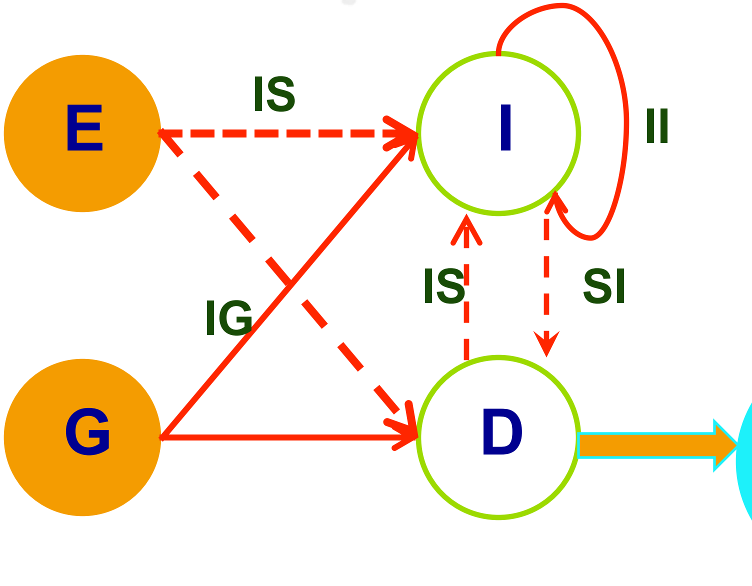


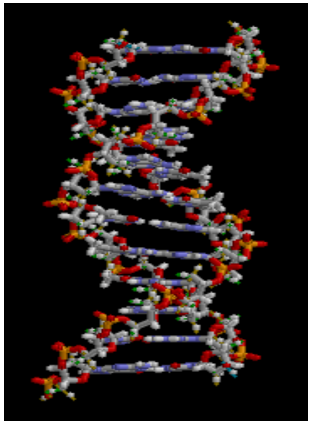


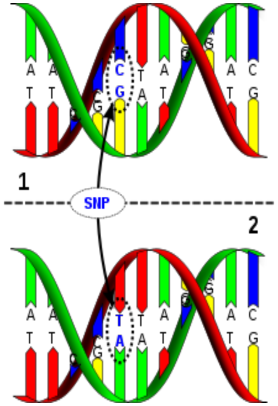












http://en.wikipedia.org/wiki/DNA_sequence

Roles of Imaging Data

Image-on-scalar (IS) model:
Image data as response, clinical variables as predictors.

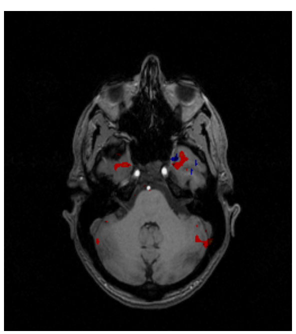
Scalar-on-image (SI) model:
Clinical variables as response, image data as predictors.

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.

Challenges in Image Data

Identify brain regions associated with covariates of interest



Raw Images

Image Reconstruction

Image Registration

Image Smoothing

Multiple Comparisons

Statistical Modelling

Statistical Analysis

Cons

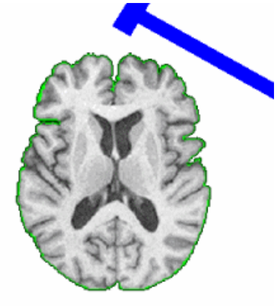
Independently and sequentially run each step.

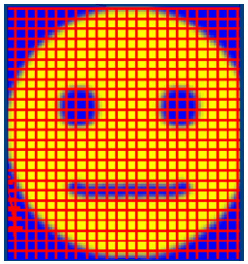
Each step has profound effects on the final statistical results and scientific findings.

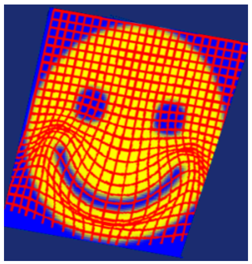
Most existing statistical methods ignore the effects of image registration and inherent spatial feature on statistical analysis.

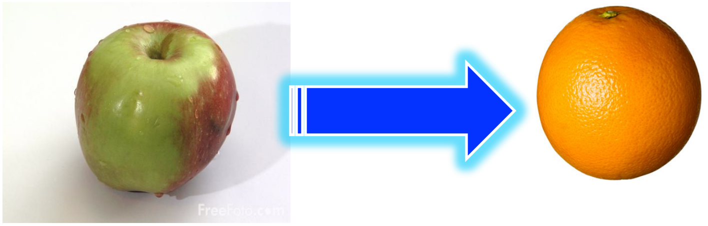
Image Registration

Image registration is the process of transforming different sets of data into ***one coordinate system***. Given a reference image R and a template image T, find a reasonable transformation Y, such that the transformed image T[Y] is similar to R.

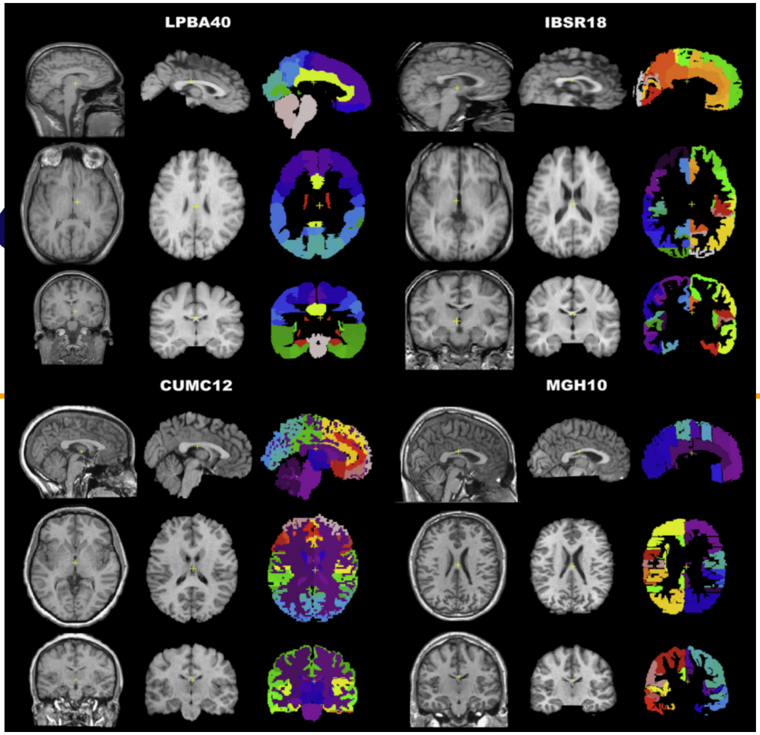








Registration ErrorS



Brain image dataset with manually labeled ROIs

Method	LPBA40	IBSR18	CUMC12	MGH10
FLIRT	59.29±11.94	39.71±13.00	39.63±11.51	46.24±14.03
AIR	65.23±10.72	41.41±13.35	42.52±11.90	47.99±14.10
ANIMAL	66.20±10.17	46.31±13.51	42.78±11.95	50.40±15.21
ART	71.85±9.59	51.54±14.42	50.54±12.16	56.10±15.33
D. Demons	68.93±9.23	46.83±13.37	46.45±11.46	52.28±14.94
FNIRT	70.07±9.80	47.63±14.15	46.53±12.26	49.54±14.58
IRTK	70.02±10.26	52.09±14.97	51.75±12.45	54.90±15.70
JRD-fuild	70.02±9.83	48.95±13.87	46.37±12.06	52.33±14.81
ROMEO	68.49±10.12	46.48±13.91	44.49±13.04	51.23±14.55
SICLE	60.41±16.21	44.53±13.03	42.08±12.19	48.36±14.31
SyN	71.46±10.86	52.81±14.85	51.63±12.60	56.83±15.81
SPM_N[1]	66.97±10.14	42.10±13.25	36.70±12.43	49.77±14.54
SPM_N[2]	57.13±14.95	37.18±14.11	42.93±11.75	43.16±15.88
SPM_US[3]	68.62±9.00	45.29±12.60	44.81±11.35	49.61±14.08
SPM_D[4]	67.15±18.34	54.02±14.70	51.98±13.91	54.31±16.05
S-HAMMER	**72.48±8.46**	**55.47±11.27**	**53.74±9.82**	**58.20±15.03**

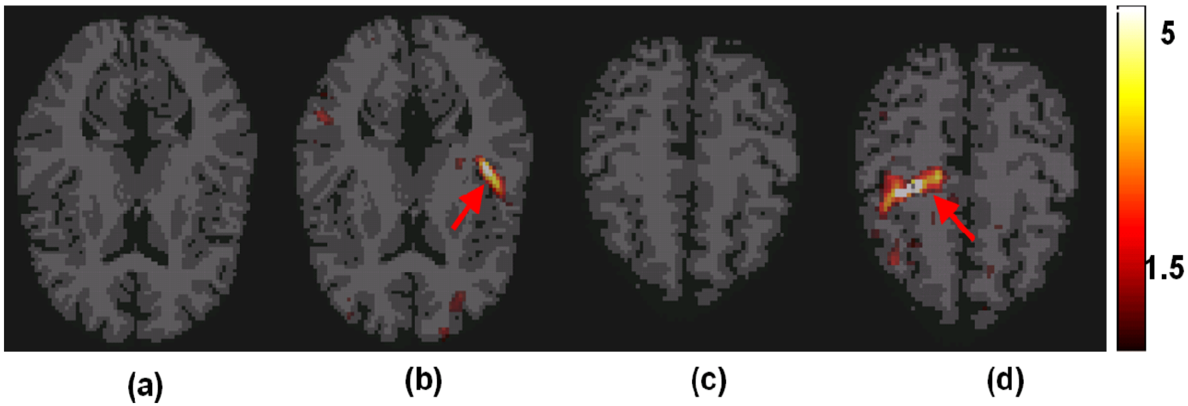
[1] SPM 5 ("SPM2-type" Normalization)
[2] SPM 5 (Normalization) [3] SPM 5 (Unified Segmentation) [4] SPM 5 (DARTEL Toolbox)

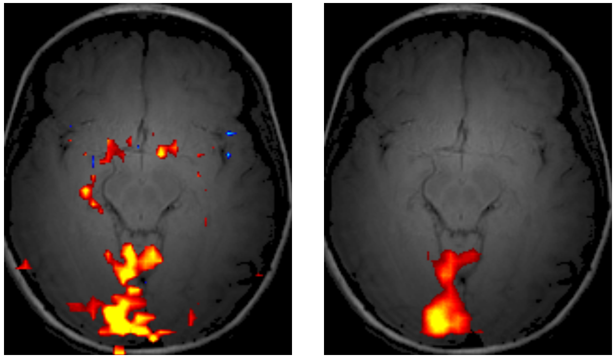
[1] Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46, 786-802.

[2] Wu, G., Kim, M., Wang, Q., Shen, D.: Hierarchical Attribute-Guided Symmetric Diffeomorphic Registration for MR Brain Images. MICCAI 2012, Nice, France (2012)

Smoothing ErrorS

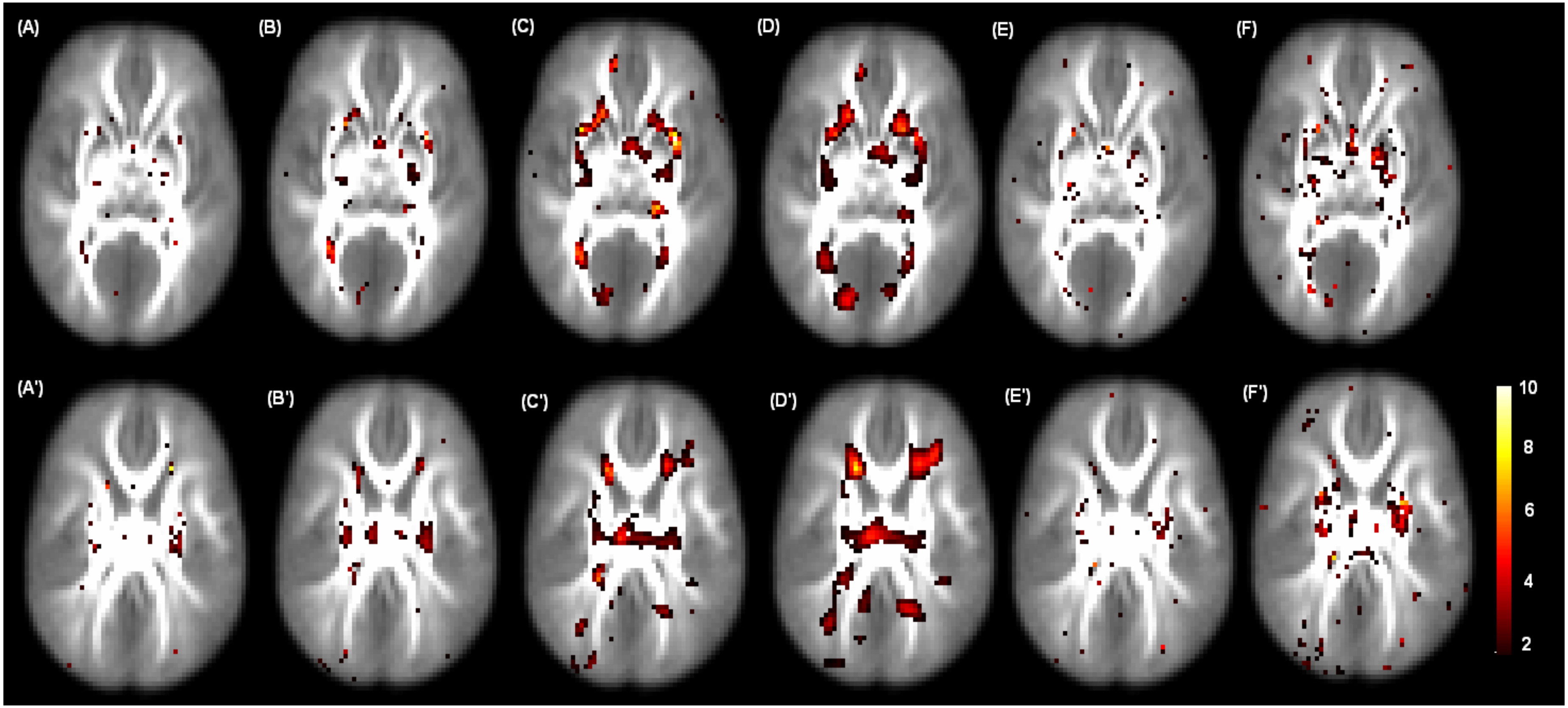
- Smoothing method is independent of data
- Degree of smoothness is arbitrary
- Effect of smoothness is profound
- The relationship between smoothing method and study design is unknown





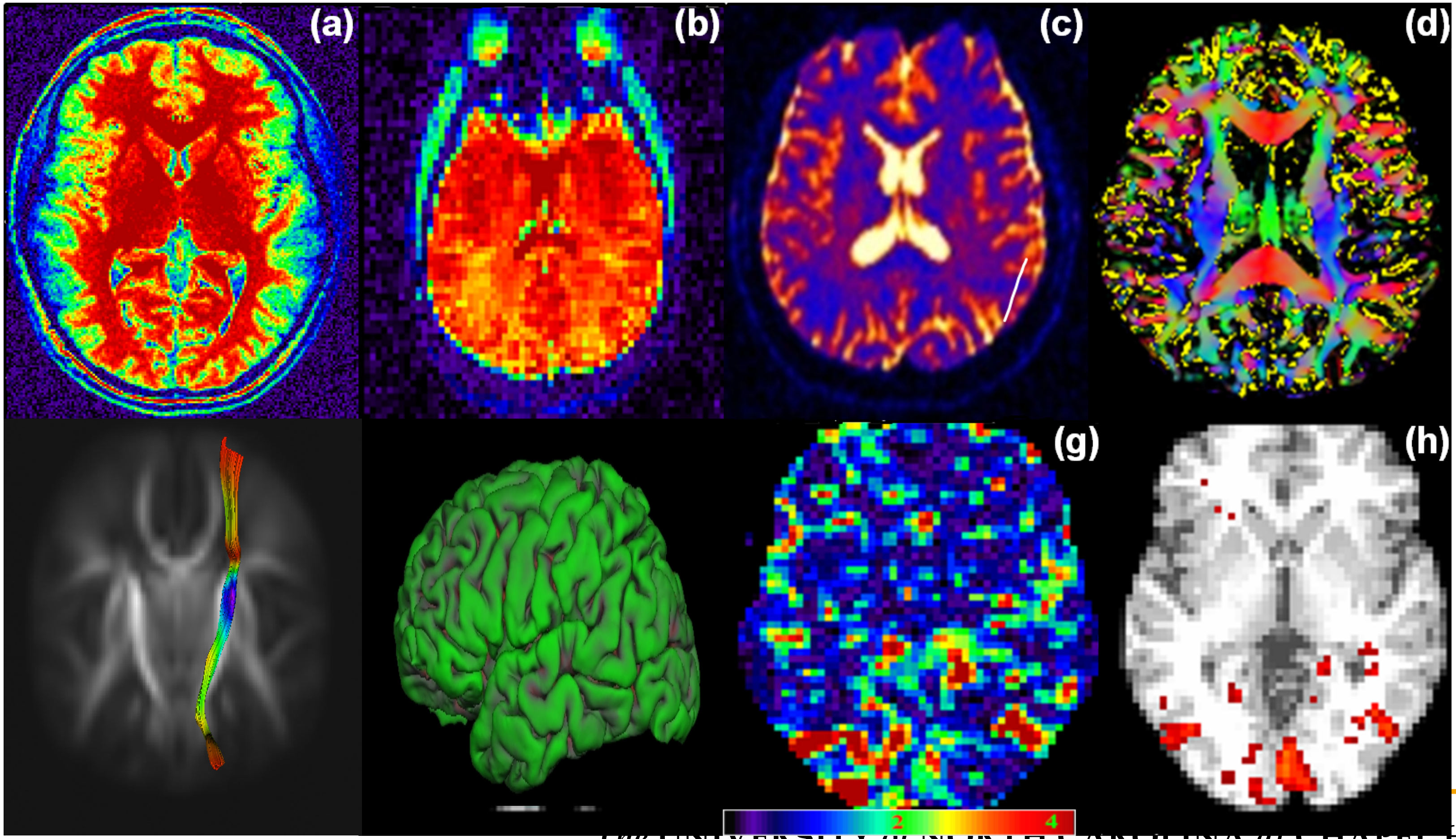
Jones et al. (2006),
Yue et al. (2010)

Real Twin Data



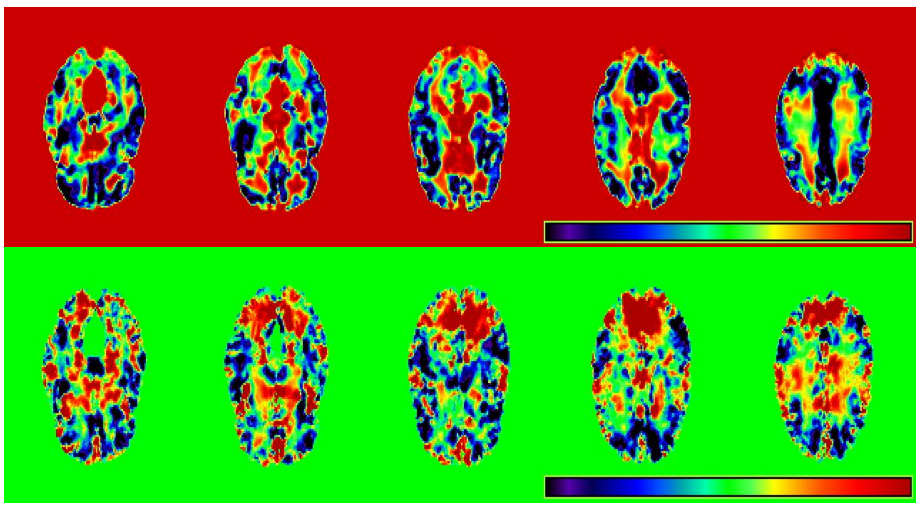
Gaussian Smoothing **Twin-MARM**

Spatial Pattern



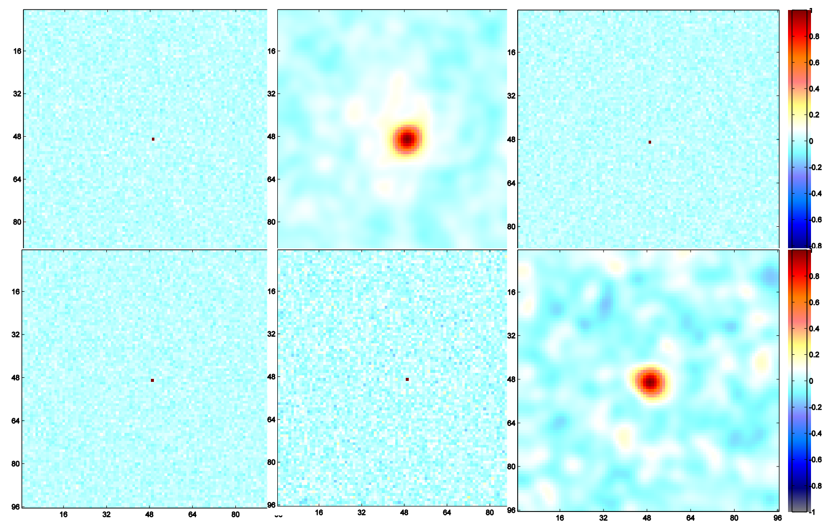
Spatial Correlation

Long-range Correlation



“Unmodeled effects”

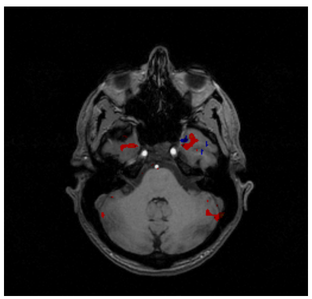
Short-range Correlation

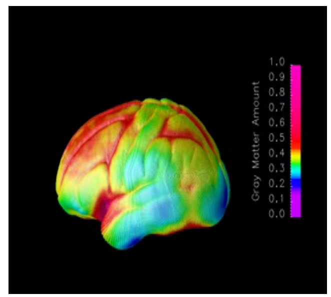


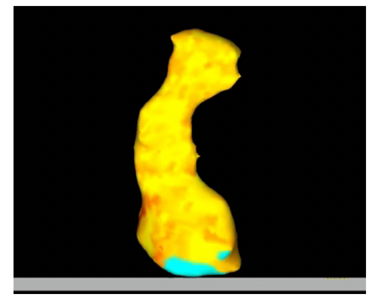
“Signal Processing”

Data types

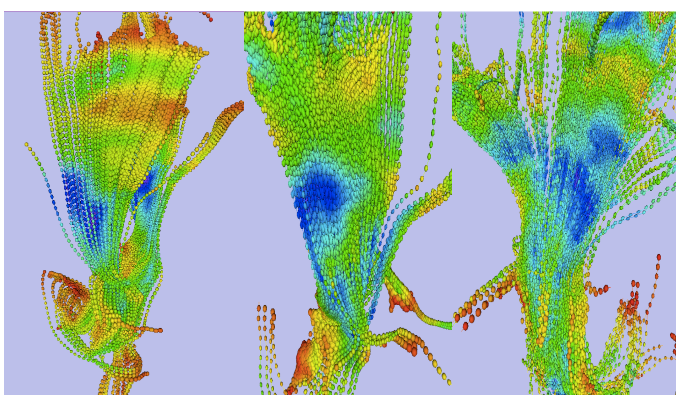
Euclidean-valued data (non-normal distributed data)

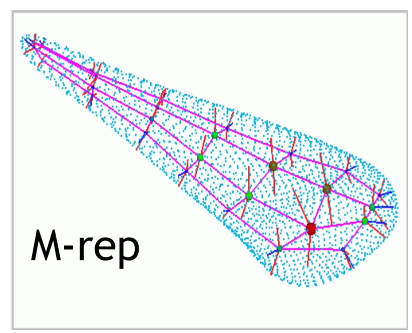


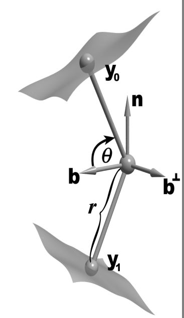




Manifold-valued data







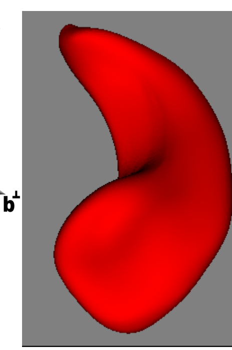


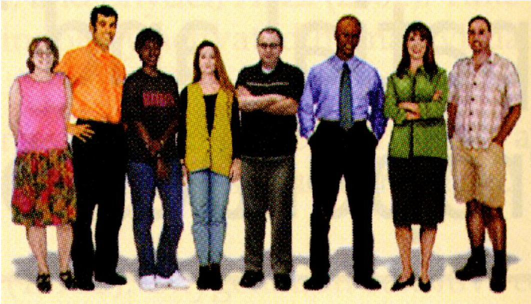
Image-on-Scalar: Voxel-based Analysis

Reading materials:

1. D.O. Siegmund, K.J. Worsley (1995). Testing for a signal with unknown location and scale in a stationary gaussian random field. Ann. Stat., 23, pp. 608–639.
2. TE Nichols and S Hayasaka. Controlling the Familywise Error Rate in Functional Neuroimaging: A Comparative Review. Statistical Methods in Medical Research, 12:419–446, 2003.
3. WL Luo and TE Nichols. Diagnosis & Exploration of Massively Univariate Neuroimaging Models. NeuroImage, 19:1014–1032, 2003.
4. K.J. Worsley, J.E. Taylor, F. Tomaiuolo, J. Lerch (2004). Unified univariate and multivariate random field theory. Neuroimage, 23, pp. 189–195.
5. Penny, Flandin, and Trujillo-Bareto (2005). Bayesian comparison of spatially regularised general linear models. *Human Brain Mapping*, 28: 275-293.
6. Harrison, Penny, Ashburner, Trujillo-Bareto, and Friston. (2007). Diffusion-based spatial priors for imaging. *NeuroImage*, 38: 677-695.
7. Bowman, F. D., Caffo, B. A, Bassett, S., and Kilts, C. (2008). Bayesian Hierarchical Framework for Spatial Modeling of fMRI Data. *NeuroImage* 39: 146–156.
8. L Xu, TD Johnson, TE Nichols, DE Nee. Modeling inter-subject variability in fMRI activation location: A Bayesian hierarchical spatial model. Biometrics, 65(4):10410–51, 2009.
9. Chumbley, K.J. Worsley, G. Flandin, K.J. Friston (2010). Topological fdr for neuroimaging Neuroimage, 49 (4), pp. 3057–3064.
10. Y Yue, JM Loh, MA Lindquist. (2010). Adaptive spatial smoothing of fMRI images. *Statistics and its Interface* 3, 3-13.
11. G Salimi-Khorshidi, SM Smith, TE Nichols. Adjusting the effect of nonstation- arity in cluster-based and TFCE inference. *Neuroimage*, 54(3):2006-19, 2011.
12. TE Nichols. Multiple testing corrections, nonparametric methods, and random field theory. *NeuroImage*, 62(2):811-815, 2012.
13. Zhao, L., Boucher, M., Rosa-Neto, P., Evans, A., (2013). Impact of scale space search on age- and gender-related changes in mri-based cortical morphometry. *Human Brain Mapping,* in press.
14. Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Junghoon Kim, John Whyte, James C. Gee and James R. Stone. (2013). Logical circularity in voxel-based analysis: normalization strategy may induce statistical bias. *Human Brain Mapping,* In press.
15. Michelle F. Miranda, *Hongtu Zhu*, and Joseph G. Ibrahim. (2013). Bayesian Analysis of Spatial Transformation Models with Applications in Neuroimaging Data. *Biometrics,* in press.

Voxel Based Analysis (VBA)

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



VBA

Stage 0: Gaussian Kernel Smoothing

Stage 1: Model Fitting

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

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 $p(\theta) = p(\{\theta(d) : d \in D\})$

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

Pro:

- **Computationally straightforward;**
- **Bayesian inference based on MCMC samples**

Con:

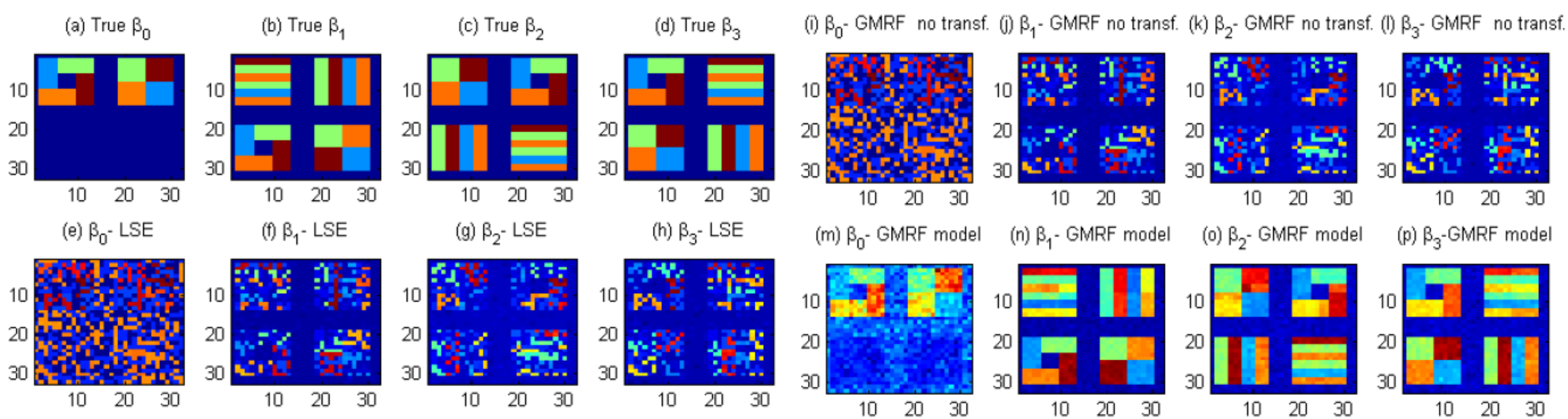
- **Computationally heavy;**
- **Lack of understanding for Bayesian inference tools.**

Example

Spatial Transformation Model

$$T(Y_i(d),\lambda(d)) = x_i^T \beta(d) + \sigma(d)\varepsilon_i(d)$$

where $T(.,\lambda(d))$ **is a Box-Cox transformation function at d.**



VBA Frequentist Extensions

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

Spatial Correlation
Spatial Smoothness

Pro:

- **Computationally easy and fast;**

Con:

- **Derive all inference tools.**

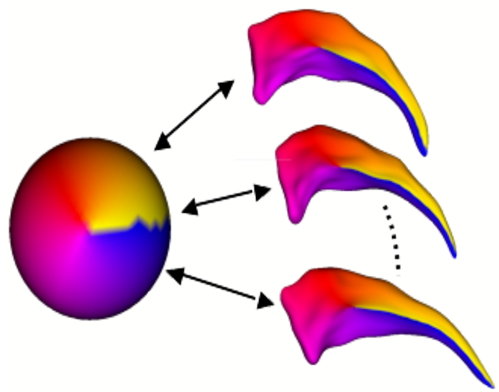
Image-on-Scalar: Varying Coefficient Models

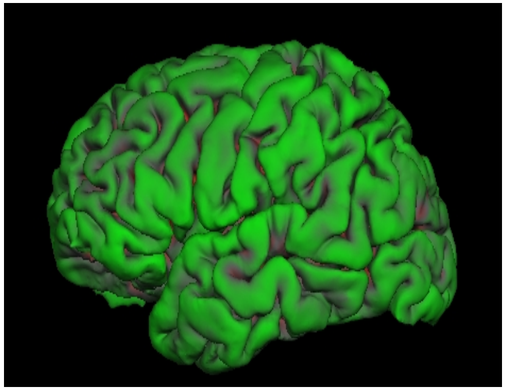
Reading materials:

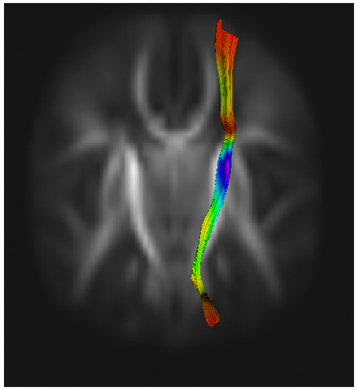
1. Yuan, Y., Gilmore, J., Geng, X. J., Styner, M., Chen, K. H., Wang, J. L., and *Zhu, H.T.* (2013). A longitudinal functional analysis framework for analysis of white matter tract statistics. *NeuroImage*, in press.
2. Yuan, Y., Zhu, H.T., Styner, M., J. H. Gilmore., and Marron, J. S. (2013). Varying coefficient model for modeling diffusion tensors along white matter bundles. *Annals of Applied Statistics.* 7(1):102-125..
3. Zhu, H.T., Li, R. Z., Kong, L.L. (2012). Multivariate varying coefficient models for functional responses. *Ann. Stat.* 40, 2634-2666.
4. 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.
5. 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.
6. 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.
7. Greven, S., Crainiceanu, C., Caffo, B., Reich, D. (2010). Longitudinal principal component analysis. E.J.Statist. 4, 1022-1054.
8. 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.
9. Yushkevich, P. A., Zhang, H., Simon, T., Gee, J. C. (2008). Structure-specific statistical mapping of white matter tracts. NeuroImage, 41, 448-461.
10. Ramsay, J. O., Silverman, B. W. (2005). *Functional Data Analysis*, Springer-Verlag, New York.

Smooth Neuroimaging Data

Smooth Functional Data

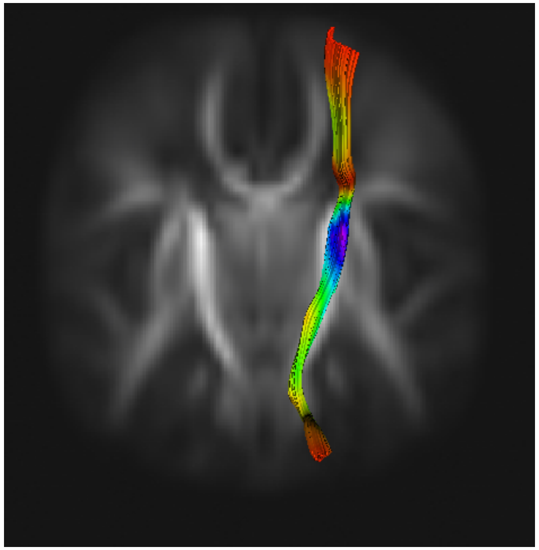






Covariates (e.g., age, gender, diagnostic)

DTI Fiber Tract Data



Data

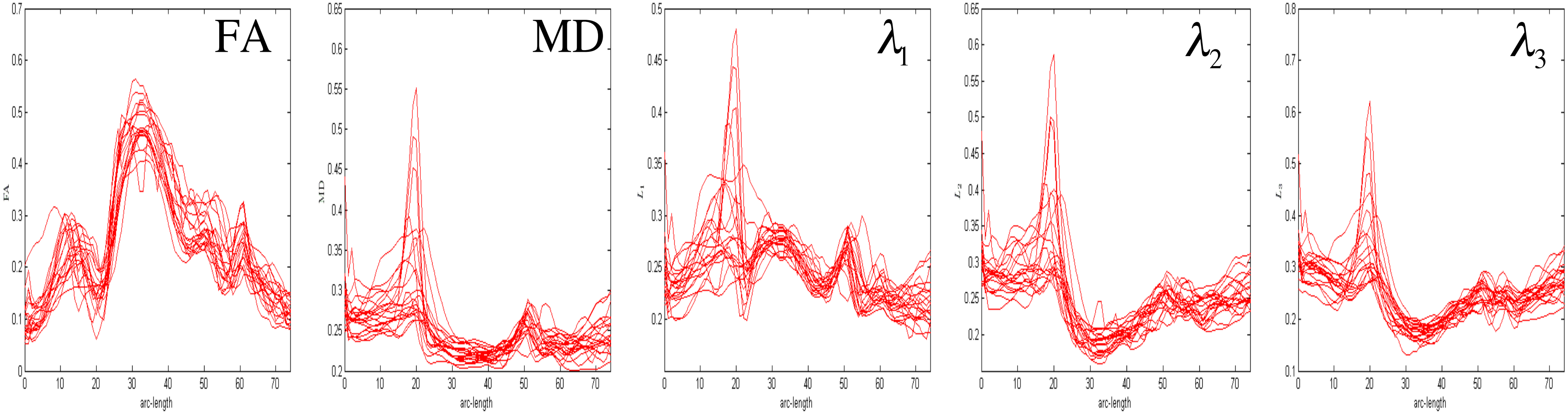
- **Diffusion properties (e.g., FA, RA)**

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

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

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

$$x_1, \cdots, x_n$$



Longitudinal Extensions

Longitudinal Data

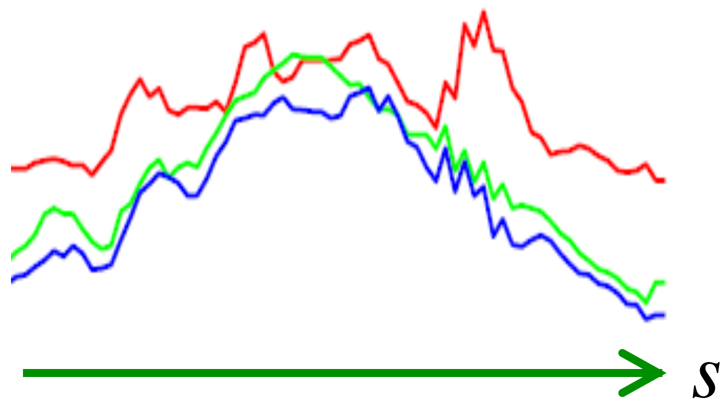
Spatial-temporal Process

t

$y_i(s,t_3)$

$y_i(s,t_2)$

$y_i(s,t_1)$

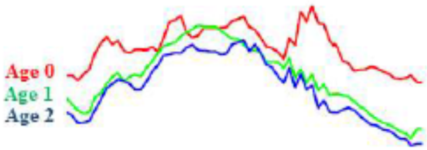
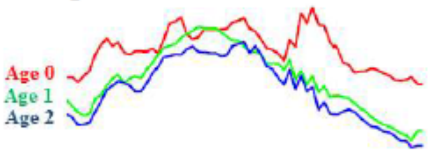


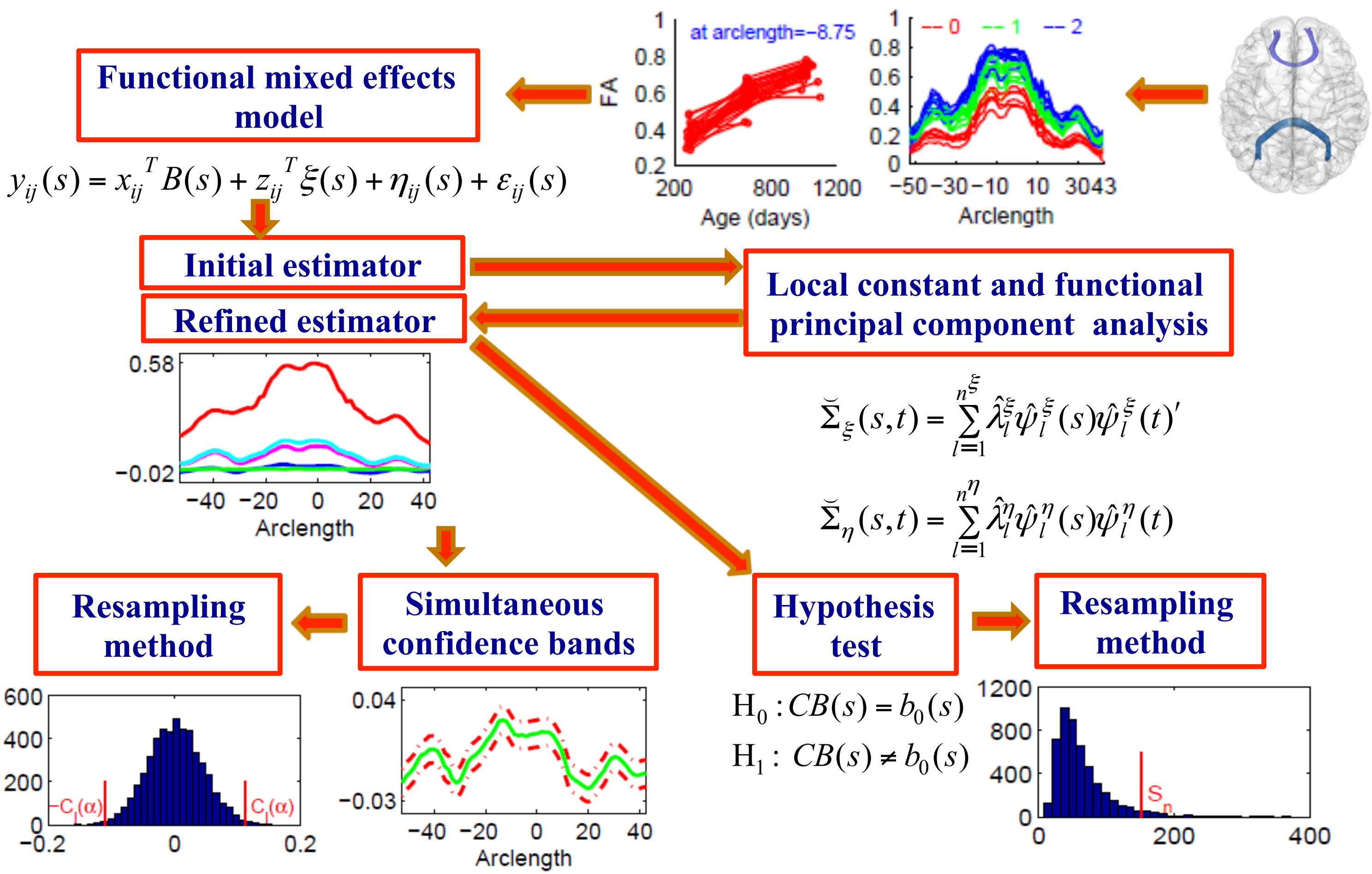
Functional Mixed Effect Models

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

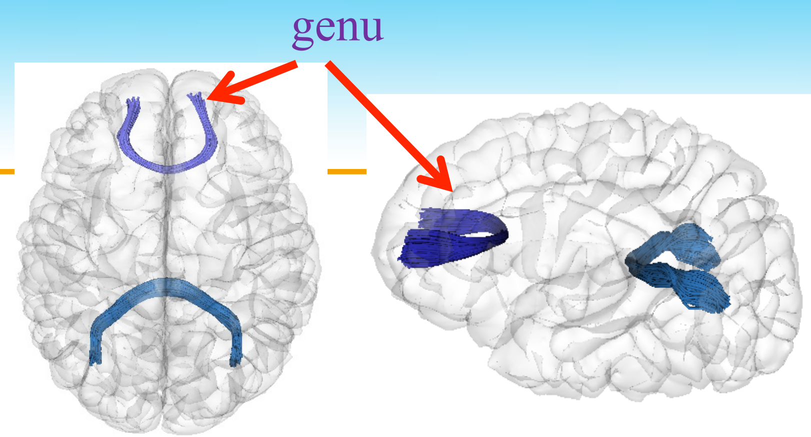
Objectives:

Dynamic functional effects of covariates of interest on functional response.

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


Functional mixed effects model
$y_{ij}(s) = x_{ij}^T B(s) + z_{ij}^T \xi(s) + \eta_{ij}(s) + \varepsilon_{ij}(s)$
at arclength=−8.75
FA
Age (days)
-- 0 -- 1 -- 2
Arclength
Initial estimator
Refined estimator
Local constant and functional principal component analysis
$\breve{\Sigma}_{\xi}(s,t) = \sum_{l=1}^{n^{\xi}} \hat{\lambda}_l^{\xi} \hat{\psi}_l^{\xi}(s) \hat{\psi}_l^{\xi}(t)'$
$\breve{\Sigma}_{\eta}(s,t) = \sum_{l=1}^{n^{\eta}} \hat{\lambda}_l^{\eta} \hat{\psi}_l^{\eta}(s) \hat{\psi}_l^{\eta}(t)$
Arclength
Resampling method
Simultaneous confidence bands
Hypothesis test
Resampling method
$H_0 : CB(s) = b_0(s)$
$H_1 : CB(s) \neq b_0(s)$
$-C_l(\alpha)$
$C_l(\alpha)$
Arclength
S_n

Real Data



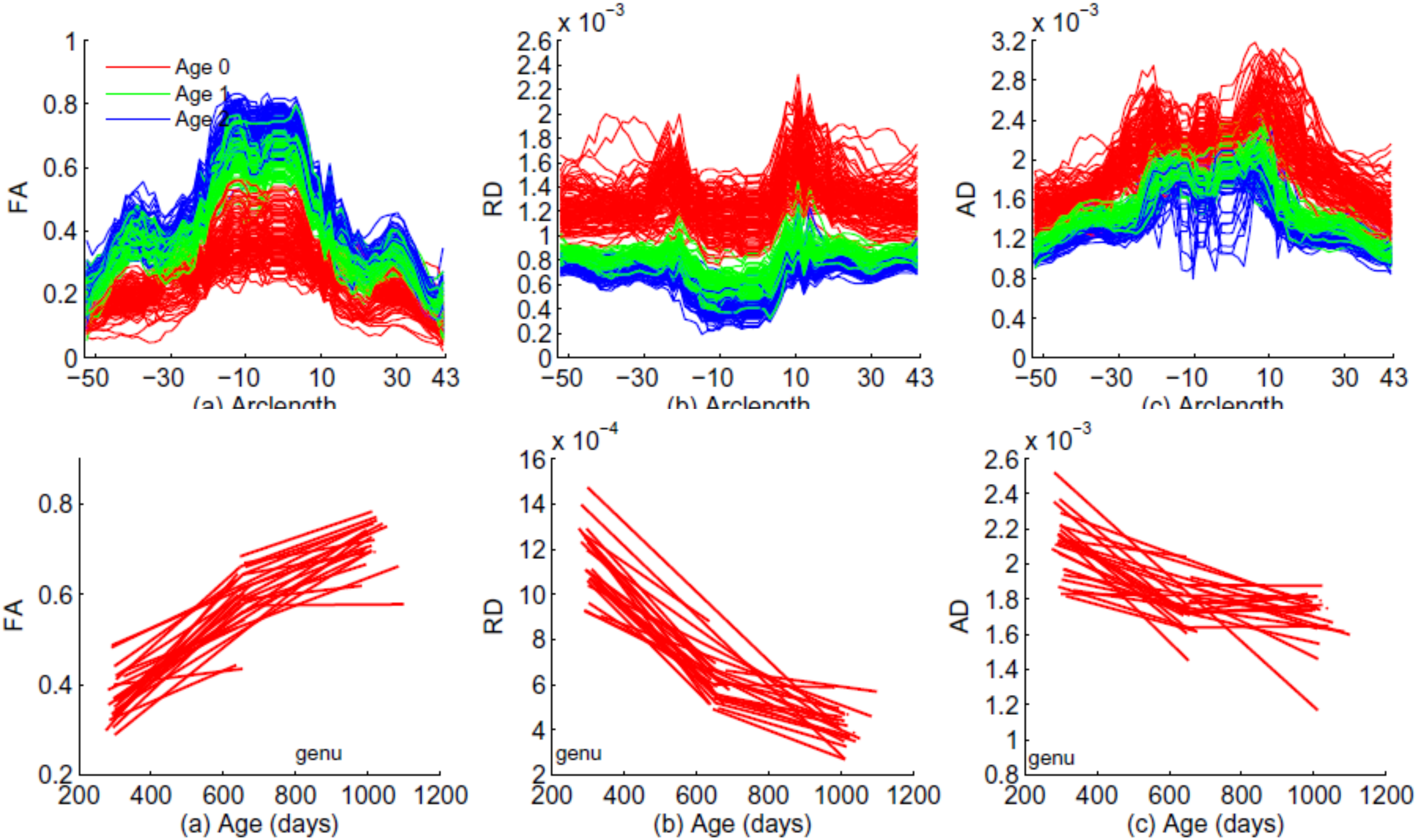
Gender: Male/Female	83/54
Gestational age at birth (weeks)	38.67 ± 1.74
Age at scan 1 (days)	297.89 ± 13.90
Age at scan 2 (days)	655.34 ± 24.00
Age at scan 3 (days)	1021.70 ± 28.26
Number of Gradient directions	
dir6/dir42 at scan 1	80/24
dir6/dir42 at scan 2	59/44
dir6/dir42 at scan 3	42/49

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

DTImaging parameters:

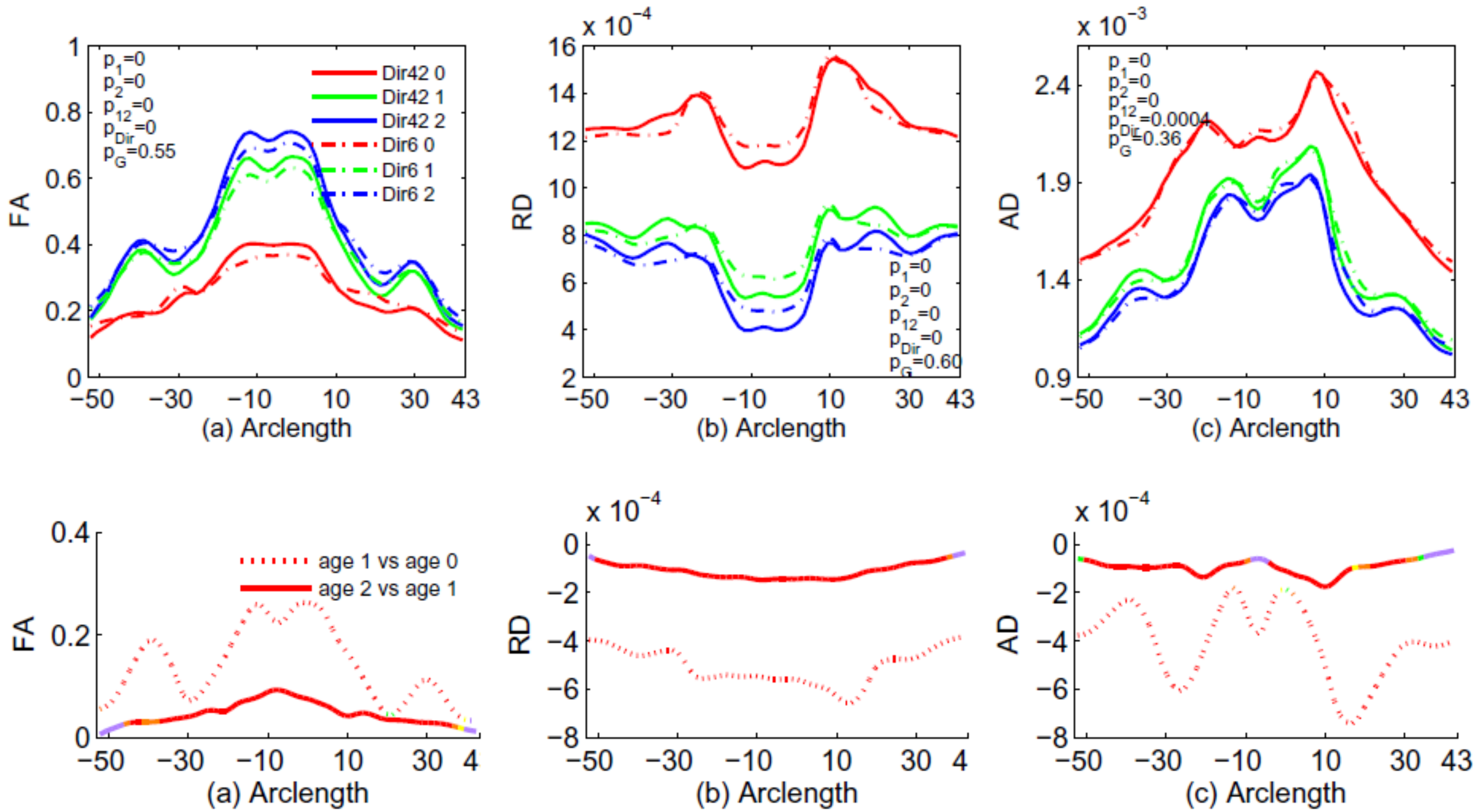
- **TR/TE = 5200/73 ms**
- **Slice thickness = 2mm**
- **In-plane resolution = 2x2 mm^2**
- **b = 1000 s/mm^2**
- **One reference scan b = 0 s/mm^2**
- **Repeated 5 times when 6 gradient directions applied.**

Real Data



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

Real Data Analysis Results



Real Data Analysis Results

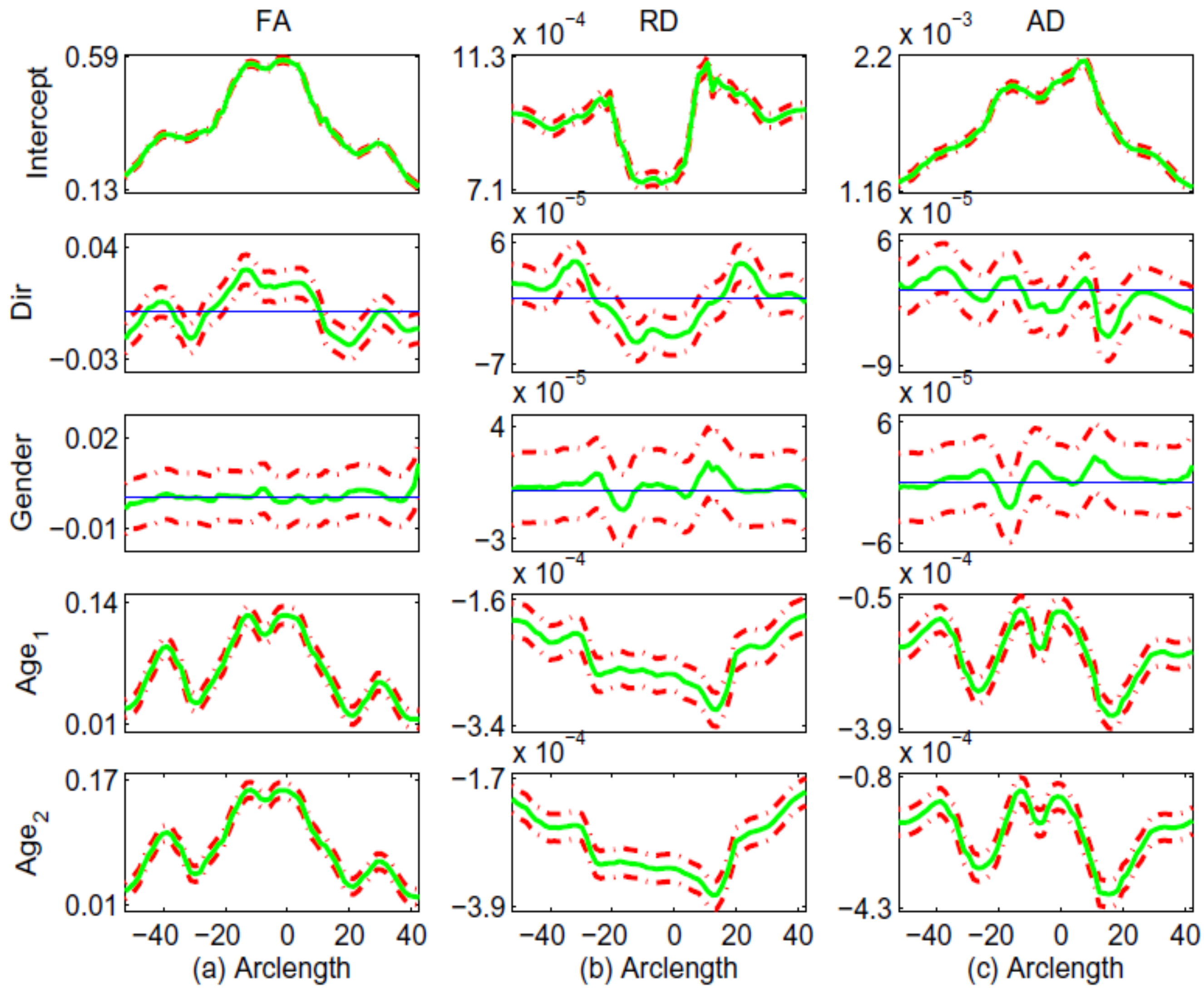


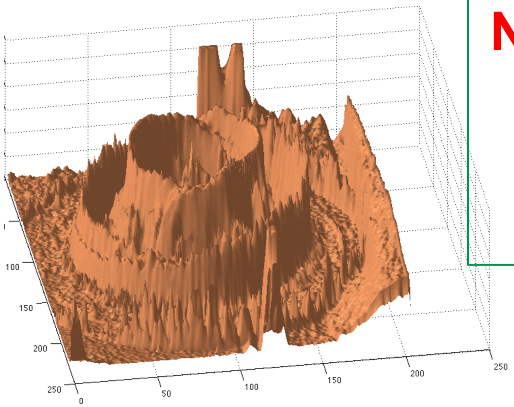
Image-on-Scalar: Multiscale Adaptive Regression Models

Reading materials:

1. Zhu, HT., Fan, J.Q., and Kong, L. (2013). Spatial varying coefficient model and its applications in neuroimaging data with jump discontinuity. in submission.
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. 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.***

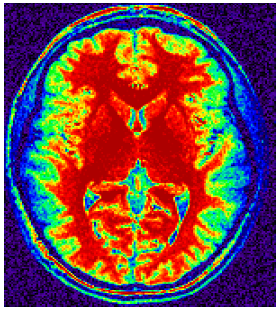
- Ω is a compact set. $\tilde{x} \in \Omega \subseteq R^k$

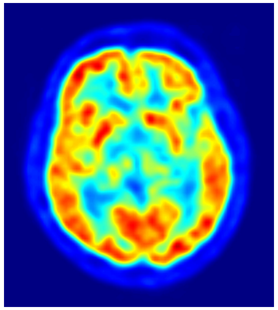
➡ $f(\tilde{x}) \in M \subseteq R^m$ $\qquad f : \Omega \to M \subseteq R^m$

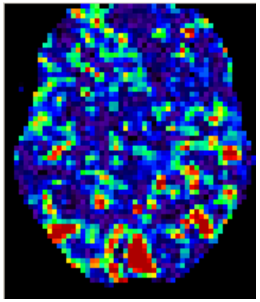
★ $\int_\Omega \| f(\tilde{x}) \|^k \, d\tilde{x} < \infty$ for some k>0

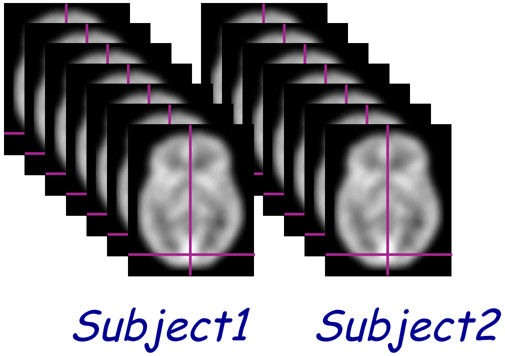
Neuroimaging Data with Discontinuity

Noisy Piecewise Smooth Function with Unknown Jumps and Edges









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

Multiscale Adaptive Regression Model

Voxel-wise Approach

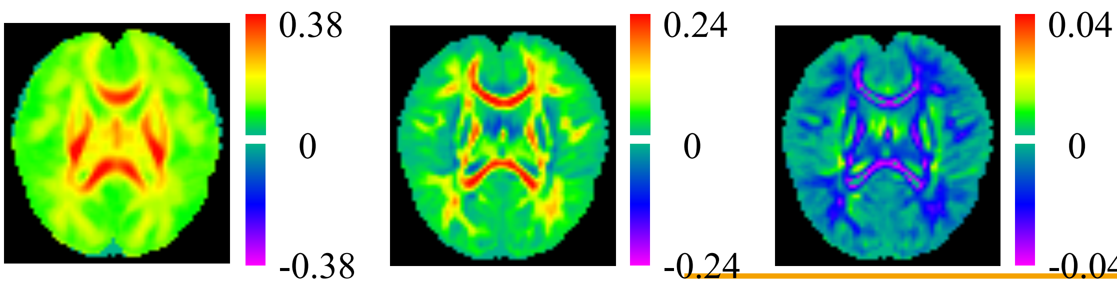
$$p(\mathbf{Y}_{i,\mathcal{D}}|\mathbf{X}_i) = \prod_{d\in\mathcal{D}} p(Y_i(d)|\mathbf{x}_i, \boldsymbol{\theta}(d)),$$

MARM

Being Spatial

$$p(\mathbf{Y}_{i,\mathcal{D}}|\mathbf{X}_i) \approx \prod_{D_k} p(\{Y_i(d') : \quad d' \in D_k\}|\mathbf{x}_i)$$

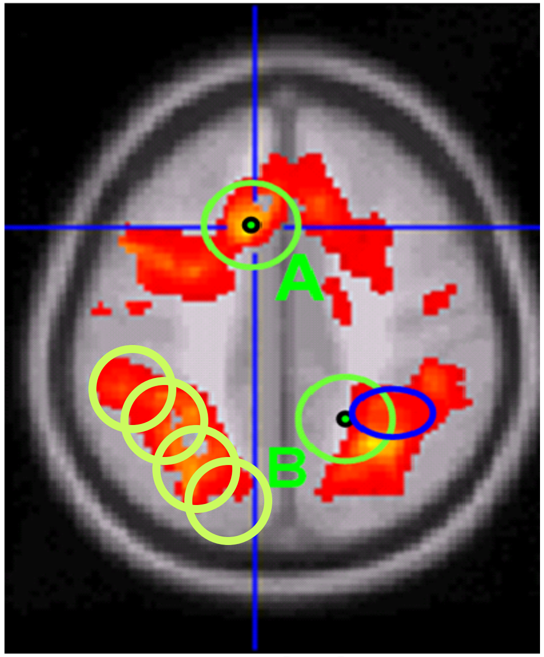
D_k **denotes the set of all voxels in a homogeneous region**



Multiscale Adaptive Regression Model

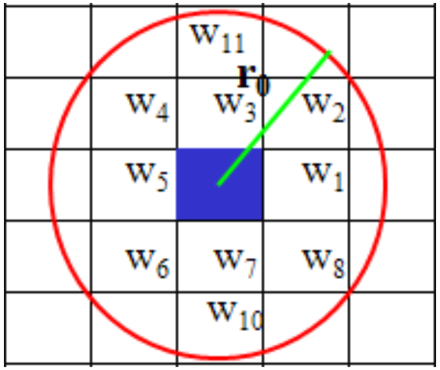
Identifying homogeneous regions

D_k



Drawing a sphere with radius r0 at each voxel

Calculating the similarities between the current voxel and its neighboring voxels.



Multiscale Adaptive Regression Model

Model Specification

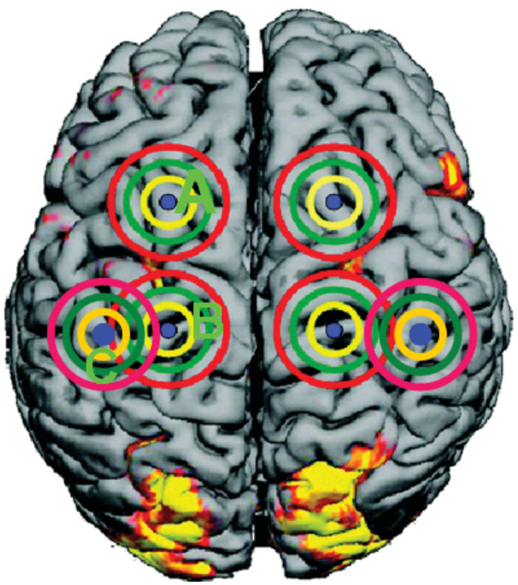
$$p(\mathbf{Y}_{i,\mathcal{D}}|\mathbf{X}_i) \approx \prod_{d\in\mathcal{D}} p(\{Y_i(d'): \quad d' \in B(d,r_0)\}|\mathbf{x}_i),$$

$$p(\{\mathbf{Y}_i(d'): \quad d' \in B(d,r_0)\}|\mathbf{x}_i) \approx \prod_{d'\in B(d,r_0)} p(Y_i(d')|\mathbf{x}_i, \boldsymbol{\theta}(d))^{\omega(d,d';r_0)}$$

$\omega(d, d'; r_0)$ is a **weight** function for characterizing the **similarity** between the data in voxels d and d'.

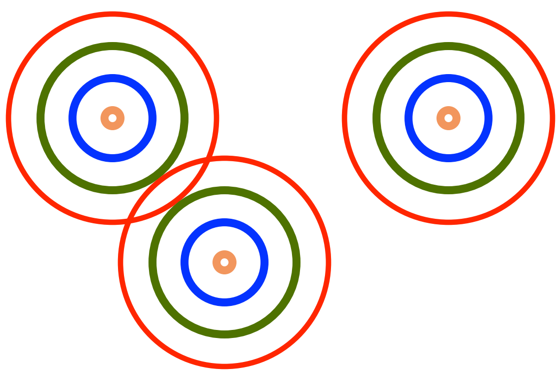
Multiscale Adaptive Regression Model

Being Hierarchical



Drawing nested spheres with increasing radiuses at each voxel

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

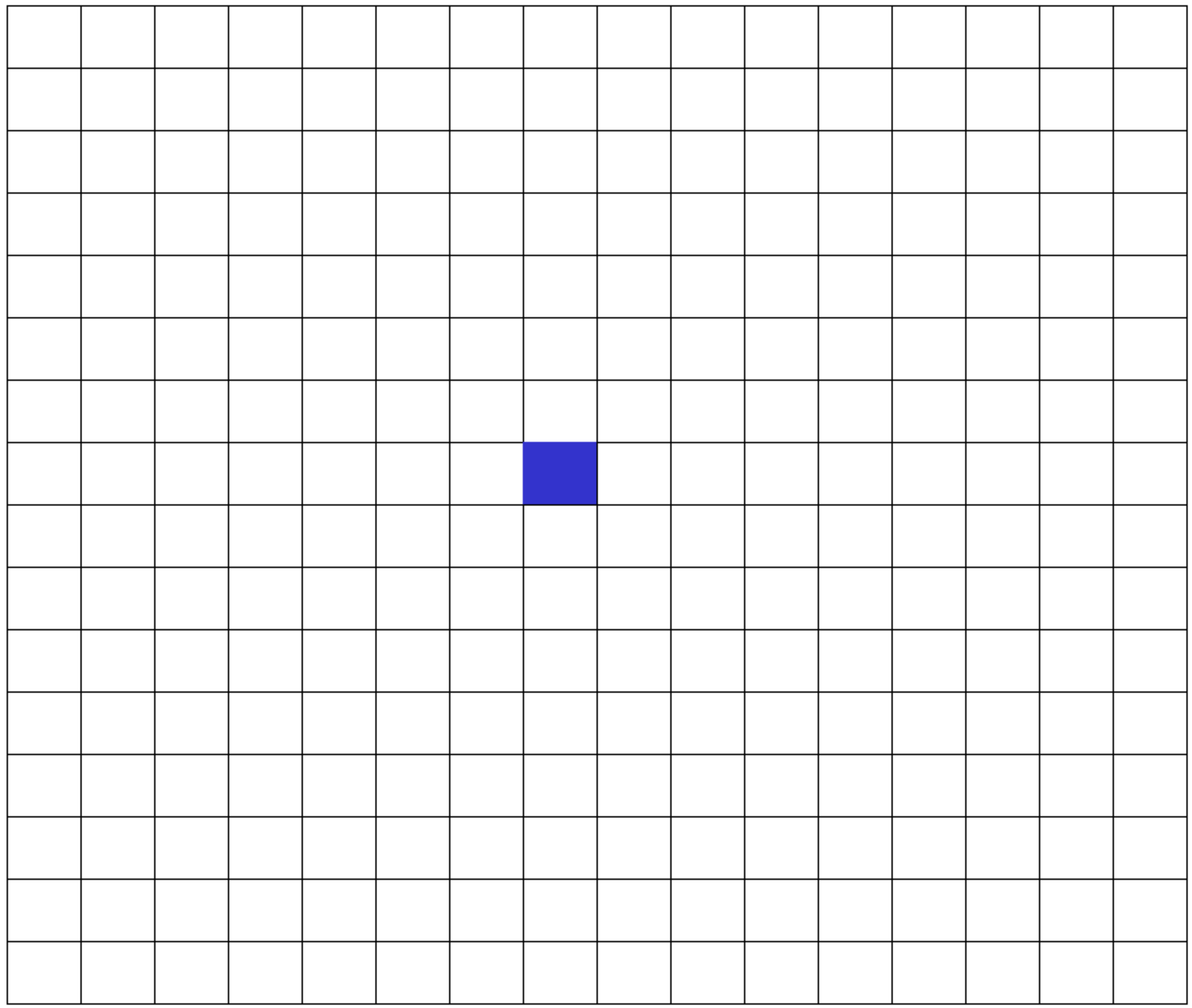


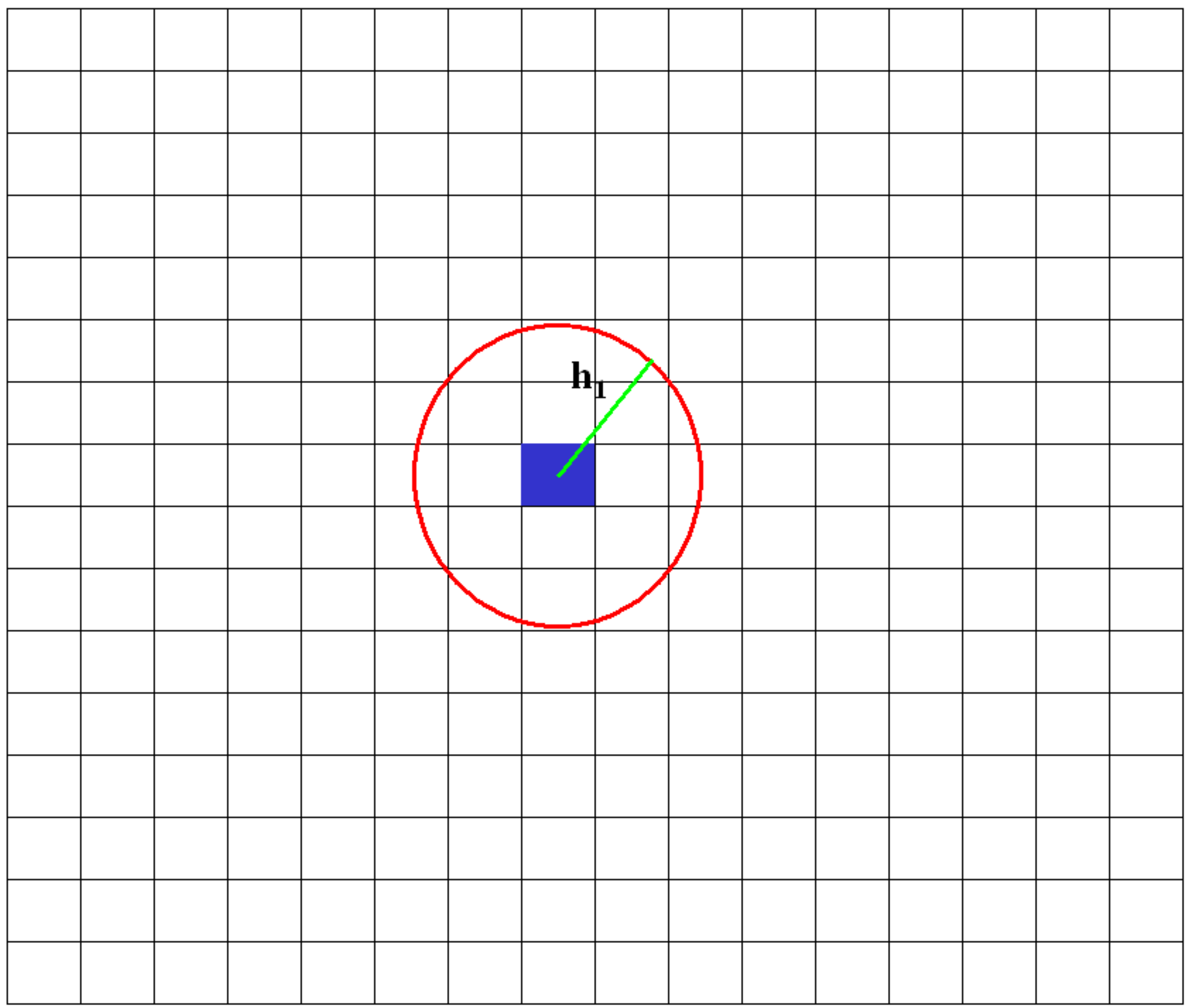
Multiscale Adaptive Regression Model

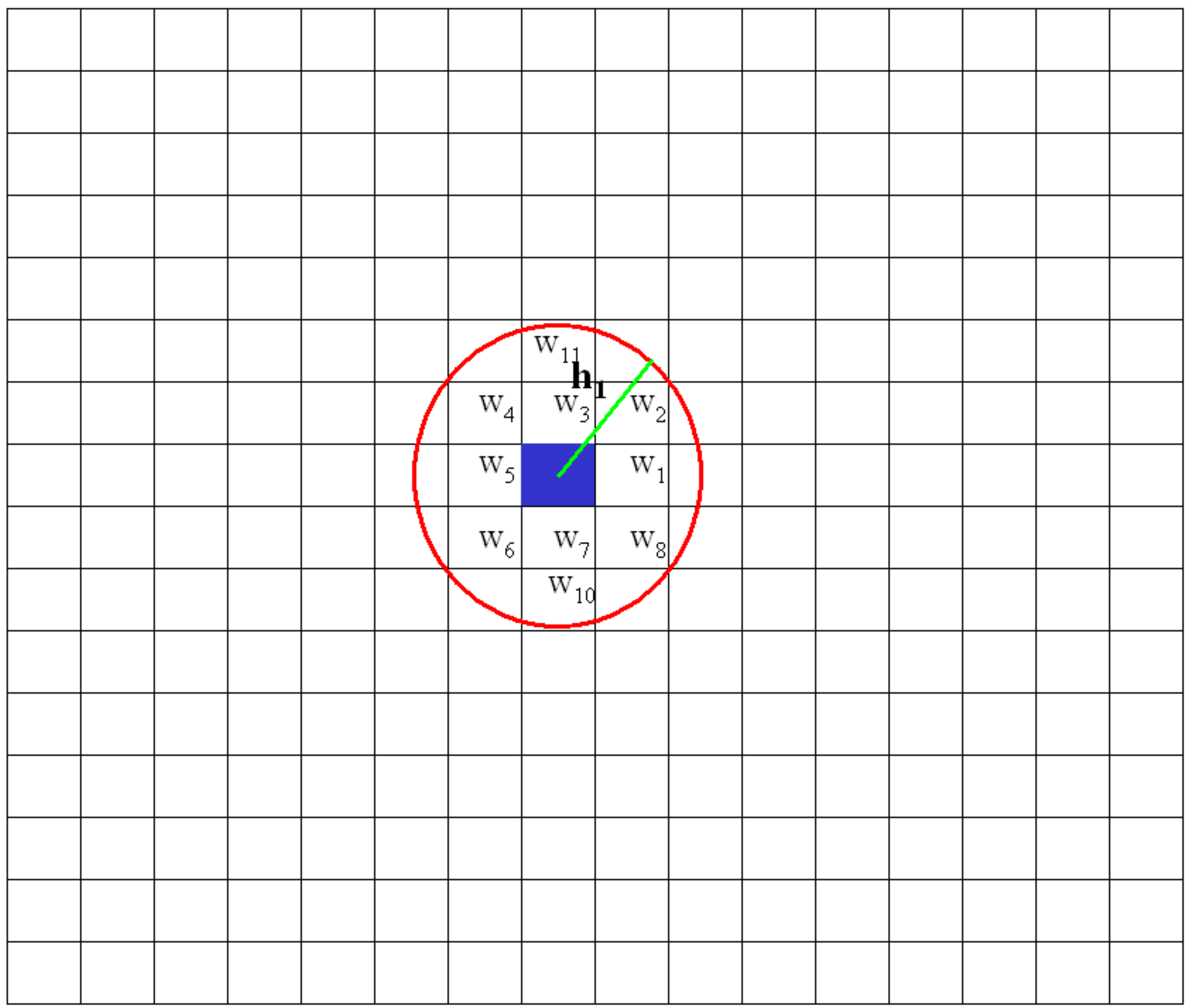
Being Adaptive

Sequentially determi $\omega(d, d'; h)$ **and adaptively updat** $\hat{\boldsymbol{\theta}}(d, h)$

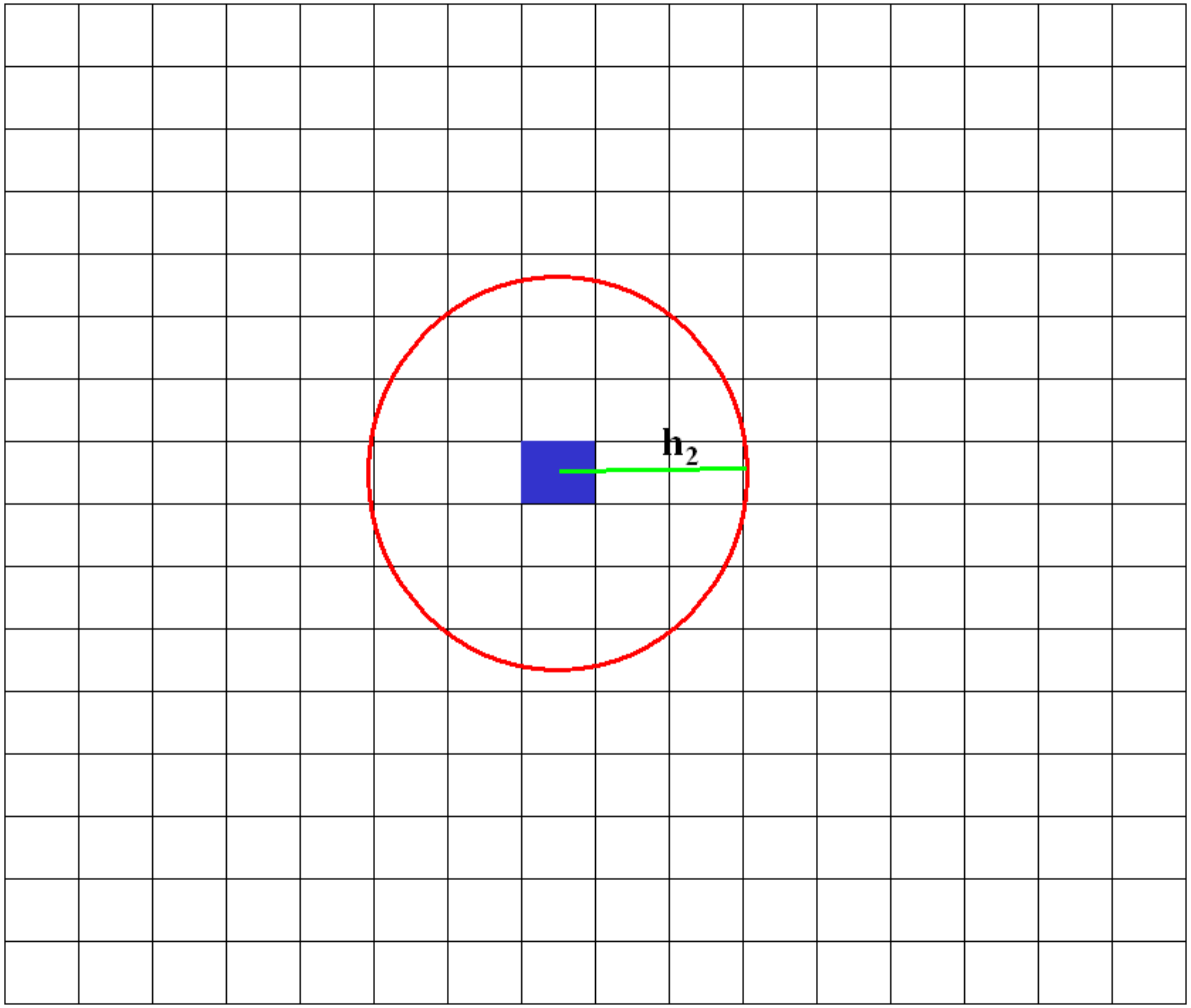
$$
\begin{array}{ccccccc}
\omega(d, d'; h_0) & & \omega(d, d'; h_1) & & \cdots & & \omega(d, d'; h_S = r_0) \\
\Downarrow & \nearrow & \Downarrow & \nearrow & \cdots & \nearrow & \Downarrow \\
\hat{\boldsymbol{\theta}}(d; h_0) & & \hat{\boldsymbol{\theta}}(d; h_1) & & \cdots & & \hat{\boldsymbol{\theta}}(d; h_S) \\
\Downarrow & & \Downarrow & & \cdots & & \Downarrow \\
W_\mu(d; h_0) & & W_\mu(d; h_1) & & \cdots & & W_\mu(d; h_S)
\end{array}
$$

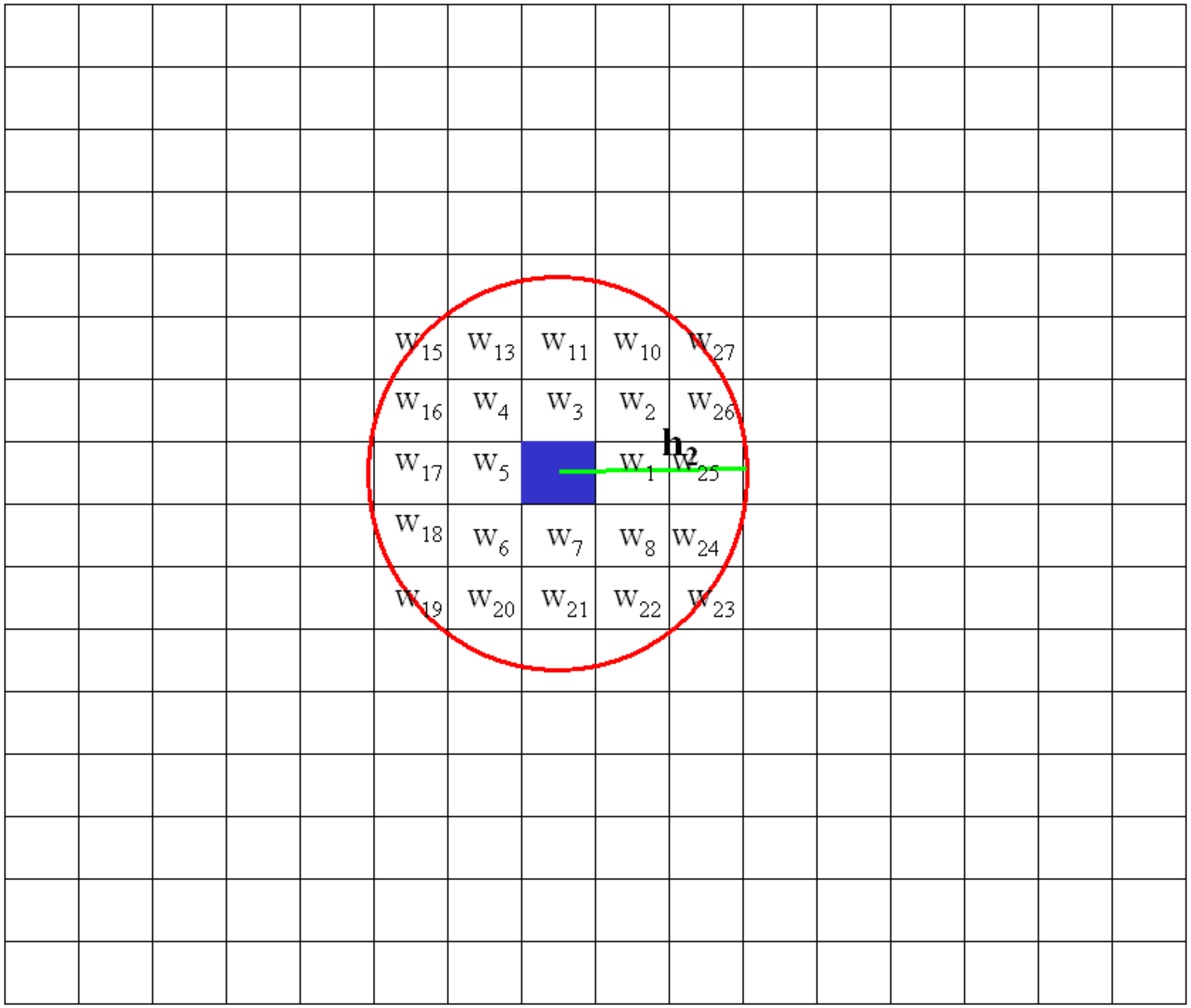



h1

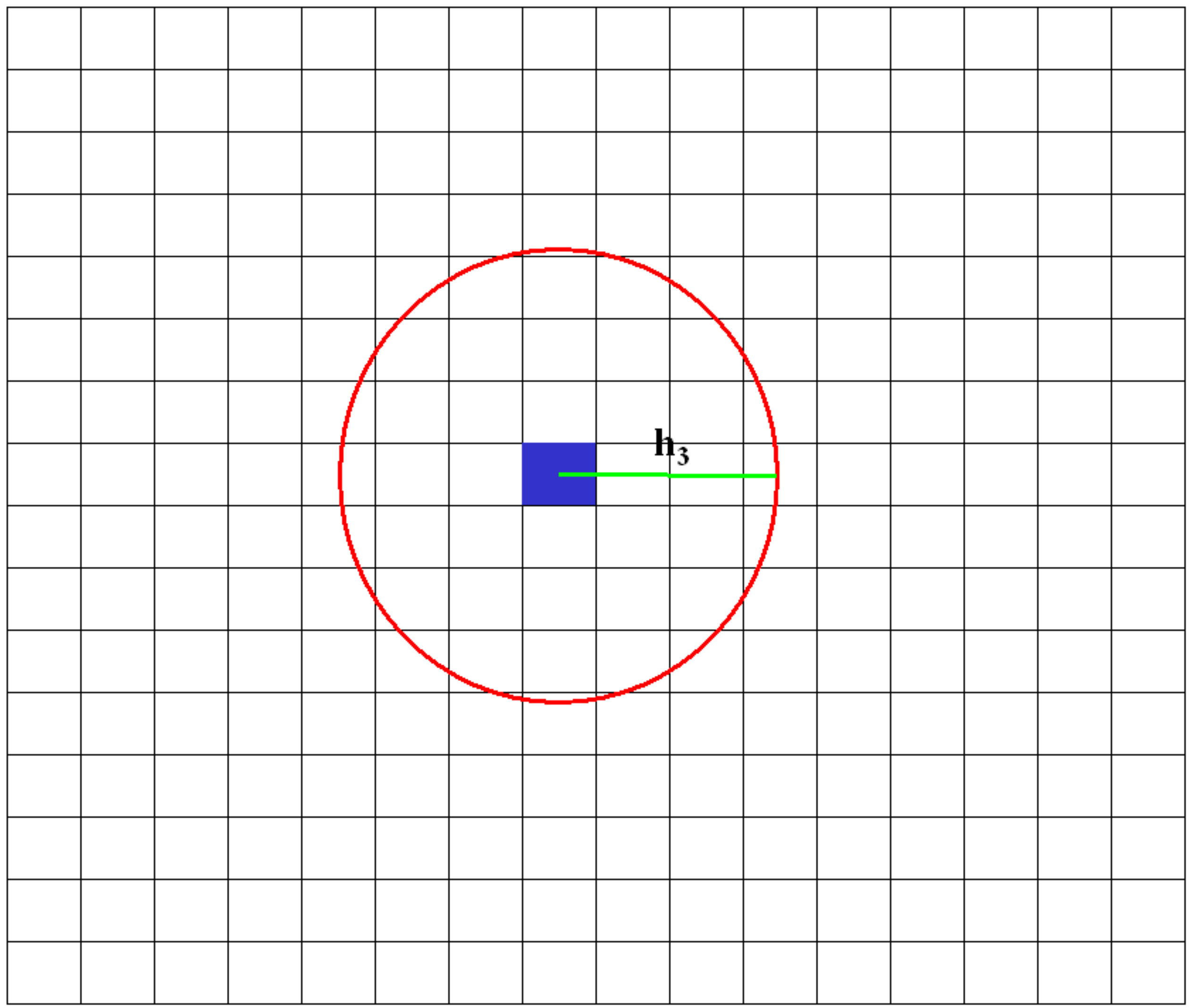


W11
h1
W4
W3
W2
W5
W1
W6
W7
W8
W10

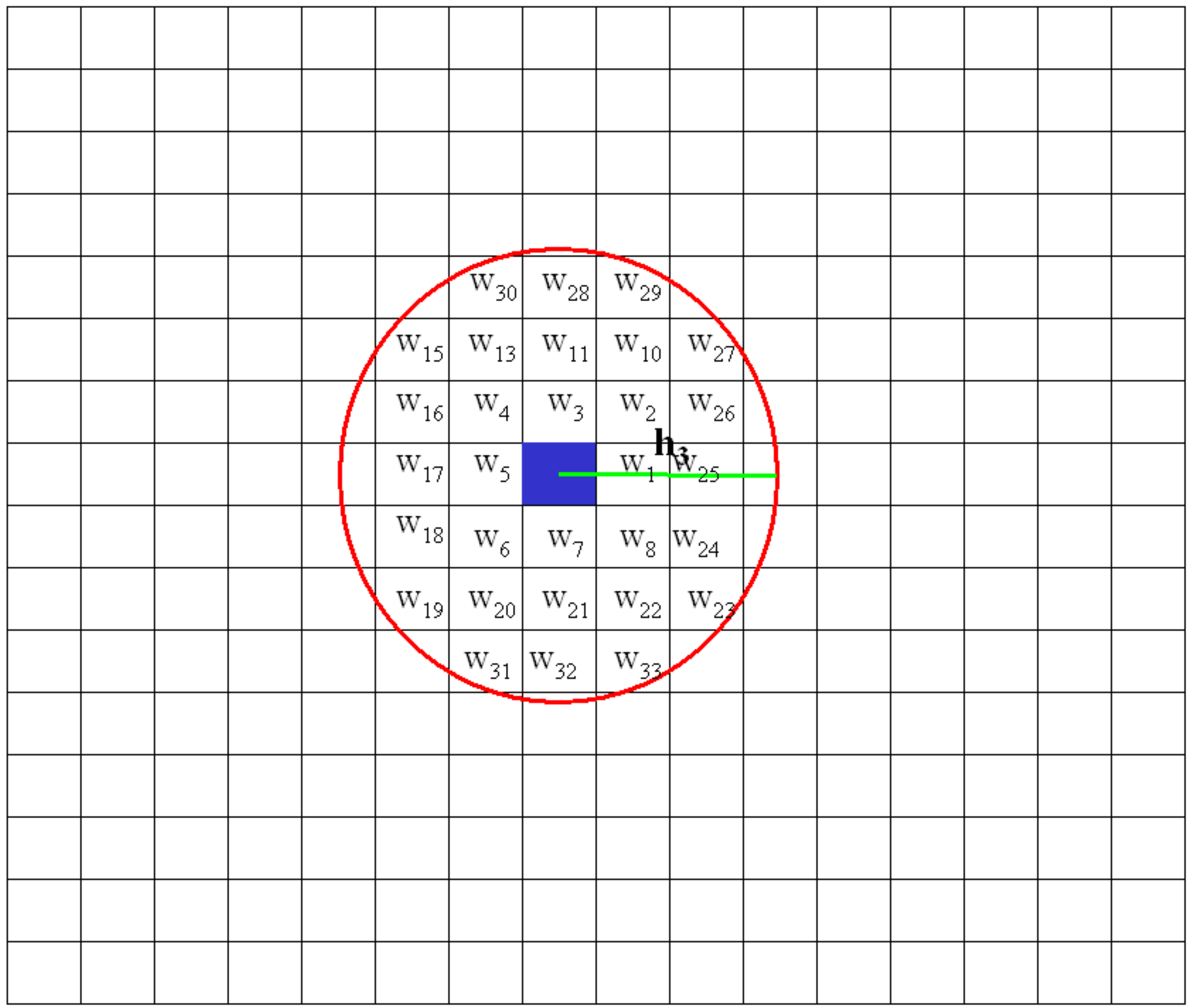

h2



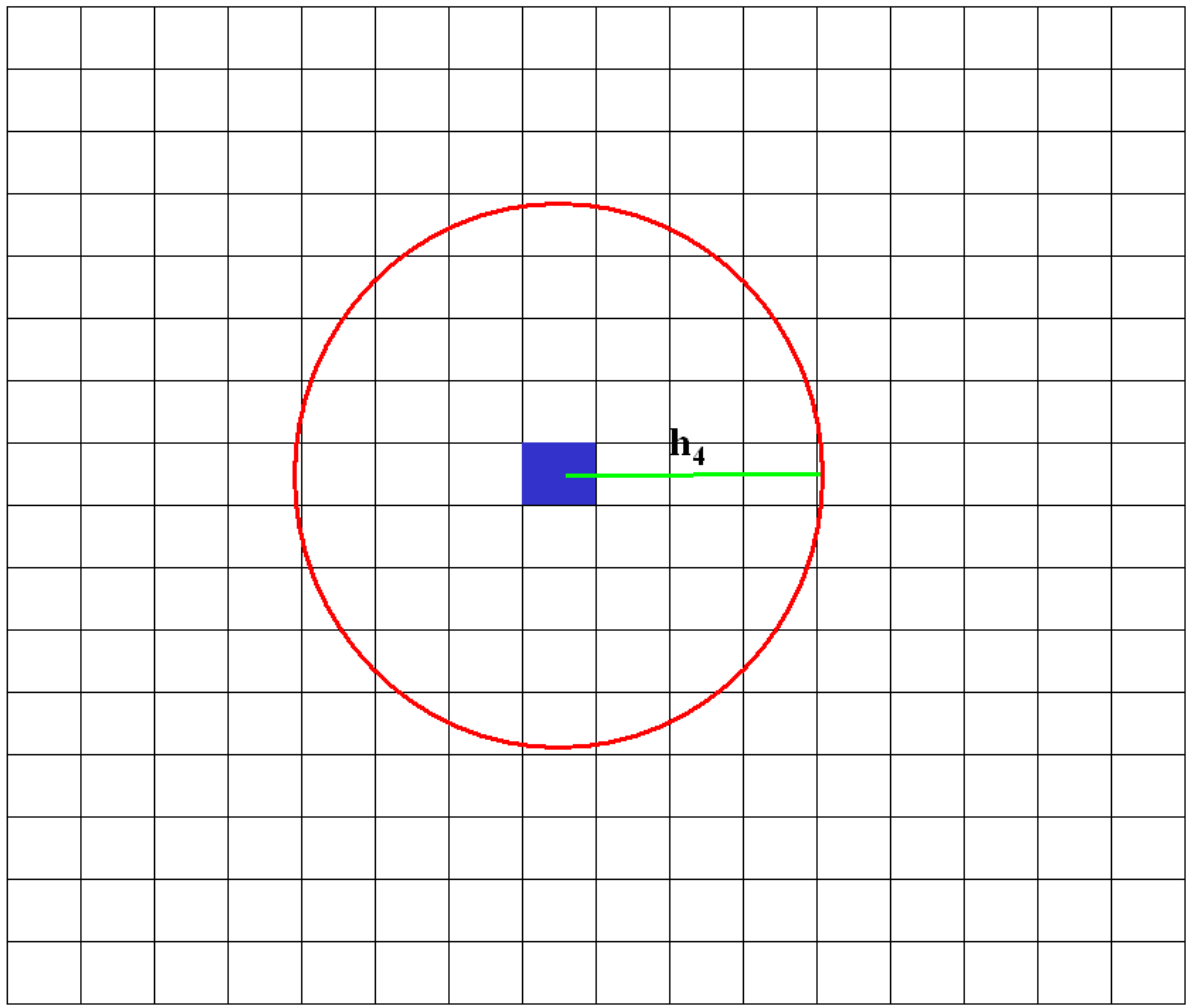
W_{15} W_{13} W_{11} W_{10} W_{27}
W_{16} W_4 W_3 W_2 W_{26}
h_2
W_{17} W_5 W_1 W_{25}
W_{18} W_6 W_7 W_8 W_{24}
W_{19} W_{20} W_{21} W_{22} W_{23}



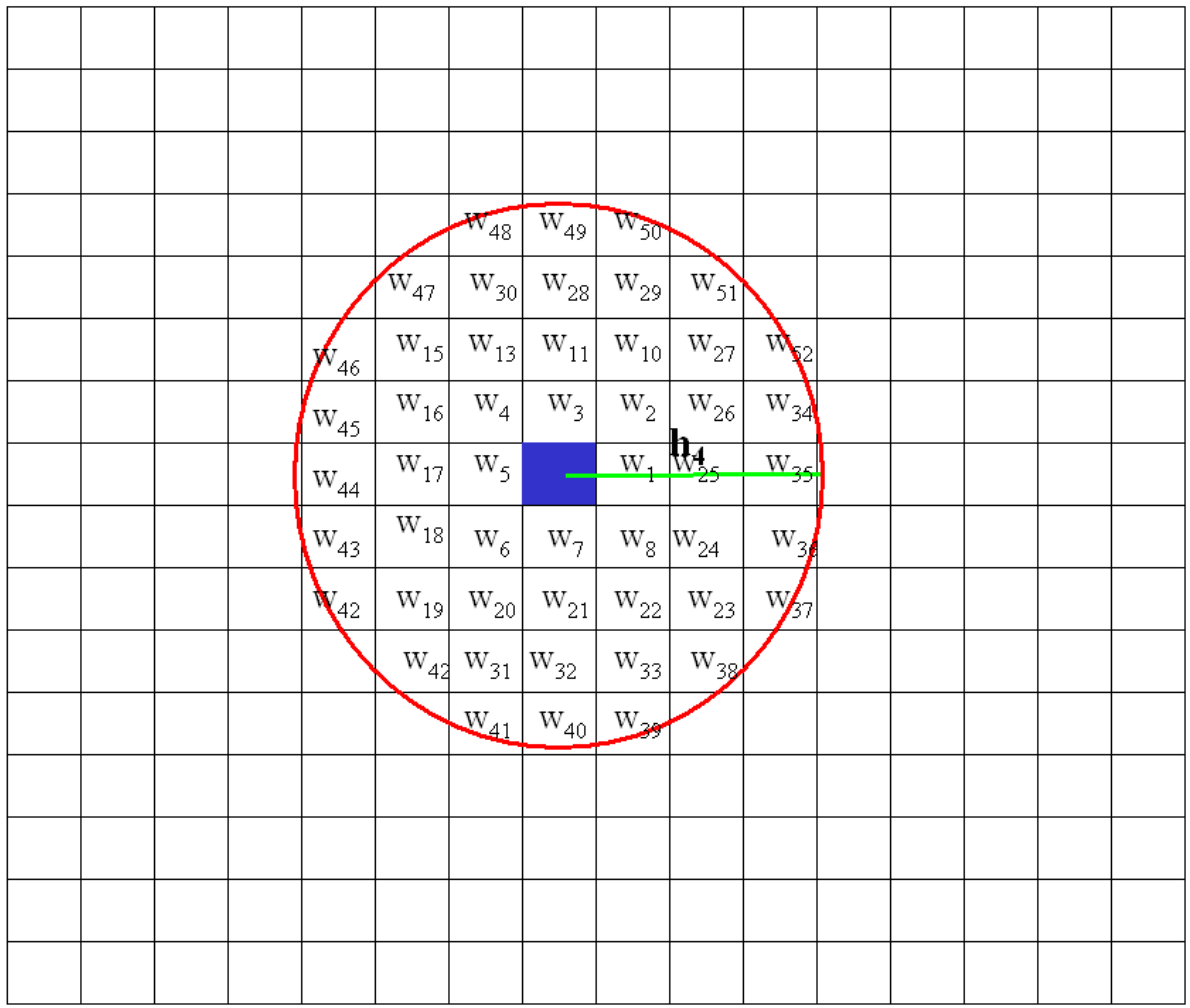
h_3



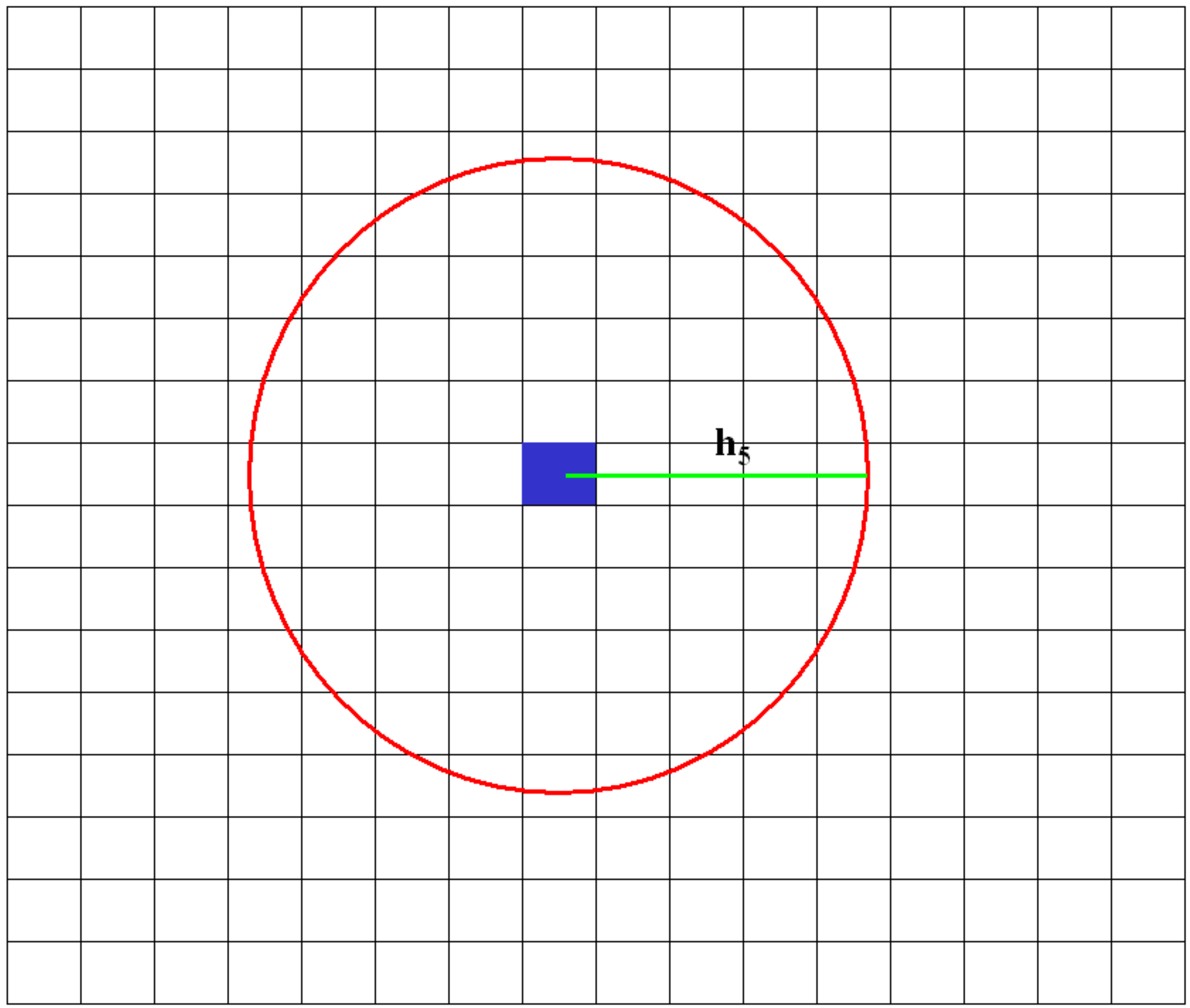
W30 W28 W29
W15 W13 W11 W10 W27
W16 W4 W3 W2 W26
h3
W17 W5 W1 W25
W18 W6 W7 W8 W24
W19 W20 W21 W22 W23
W31 W32 W33



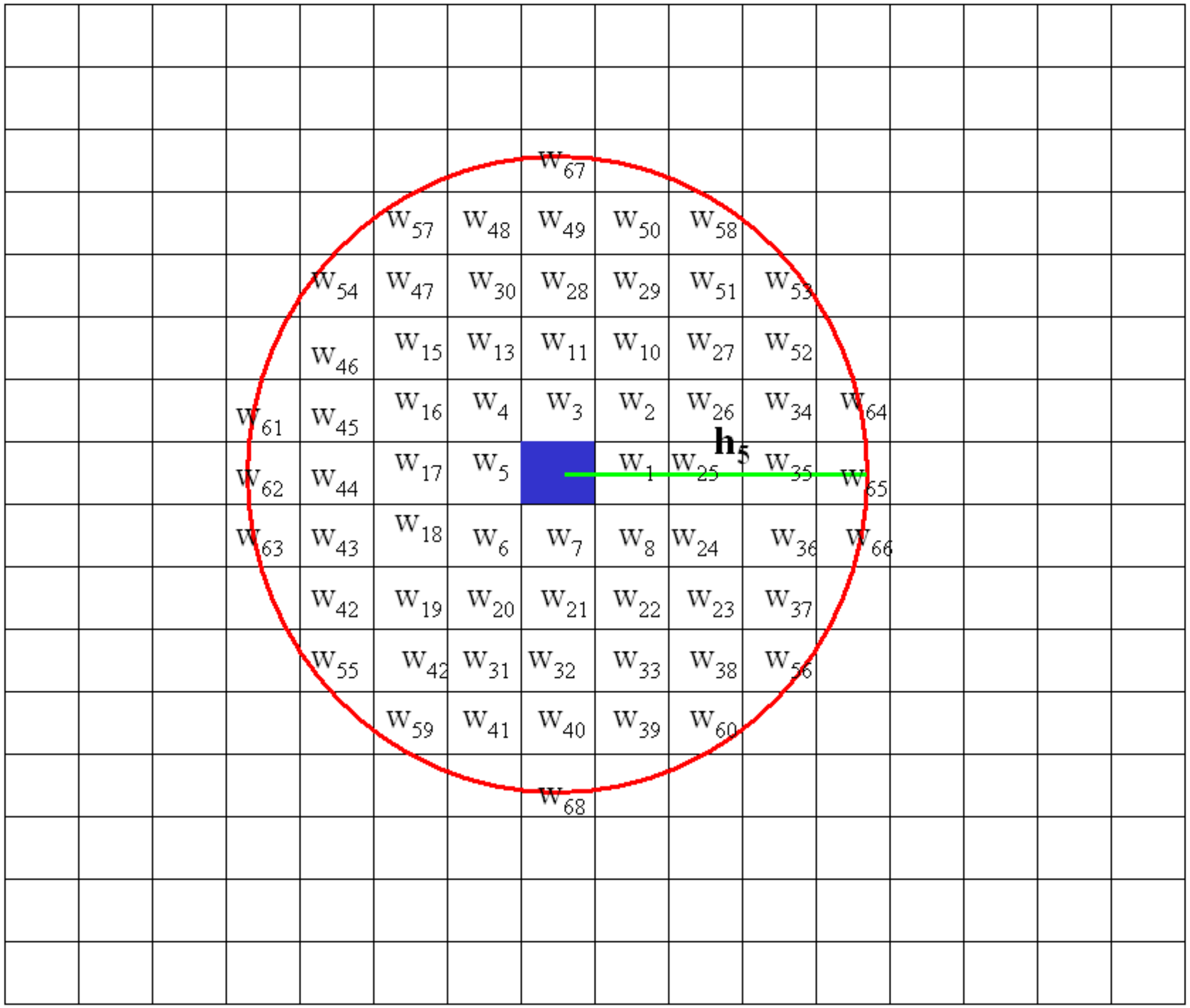
h_4



W48 W49 W50
W47 W30 W28 W29 W51
W46 W15 W13 W11 W10 W27 W52
W45 W16 W4 W3 W2 W26 W34
h4
W44 W17 W5 W1 W25 W35
W43 W18 W6 W7 W8 W24 W36
W42 W19 W20 W21 W22 W23 W37
W42 W31 W32 W33 W38
W41 W40 W39



h_5



W_{67}
W_{57}
W_{48}
W_{49}
W_{50}
W_{58}
W_{54}
W_{47}
W_{30}
W_{28}
W_{29}
W_{51}
W_{53}
W_{46}
W_{15}
W_{13}
W_{11}
W_{10}
W_{27}
W_{52}
W_{61}
W_{45}
W_{16}
W_{4}
W_{3}
W_{2}
W_{26}
W_{34}
W_{64}
h_5
W_{62}
W_{44}
W_{17}
W_{5}
W_{1}
W_{25}
W_{35}
W_{65}
W_{63}
W_{43}
W_{18}
W_{6}
W_{7}
W_{8}
W_{24}
W_{36}
W_{66}
W_{42}
W_{19}
W_{20}
W_{21}
W_{22}
W_{23}
W_{37}
W_{55}
W_{42}
W_{31}
W_{32}
W_{33}
W_{38}
W_{56}
W_{59}
W_{41}
W_{40}
W_{39}
W_{60}
W_{68}

Multiscale Adaptive Regression Model

MARM

- Learning Voxel Feature
- Local Feature Adaptation
- Adaptive Estimation and Testing
- Automatic Stop

Multiscale Adaptive Regression Model

Learning Voxel Feature

Set bandwidth h_0 =0 and run voxel-wise approach.

Generate a geometric series $\{\mathrm{h}_s = c_h^s : s = 1,\ldots S\}$

Multiscale Adaptive Regression Model

Local Feature Adaptation

- For any radius $h_s > h_0$, define

$$\omega(d,d';h_s) = K_{loc}(||d-d'||_2/h_s)K_{st}(D_{\boldsymbol{\theta}}(d,d';h_{s-1})/C_n)$$

- $K_{loc}(u)$ and $K_{st}(u)$ are two decreasing kernel functions
- Smoothing kernel: $K_{loc}(u) = (1-u^2)_+$
- Similarity kernel: $K_{st}(u) = \exp(-u)\mathbf{1}\left(u \leq \frac{s+2}{s(\log s+2)}\right)$
- Dissimilarity measure:

$$D_{\boldsymbol{\theta}}(d,d';h_{s-1}) = [\hat{\boldsymbol{\theta}}(d;h_{s-1}) - \hat{\boldsymbol{\theta}}(d';h_{s-1})]^T \hat{\Sigma}(\hat{\boldsymbol{\theta}}(d;h_{s-1}))^{-1}[\hat{\boldsymbol{\theta}}(d;h_{s-1}) - \hat{\boldsymbol{\theta}}(d';h_{s-1})].$$

Multiscale Adaptive Regression Model

Adaptive Estimation and Testing

Weighted quasi-likelihood

$$\ell_n(\boldsymbol{\theta}(d); h, \tilde{\boldsymbol{\omega}}) = \sum_{i=1}^{n} \sum_{d' \in B(d,h)} \tilde{\omega}(d, d'; h) \log p(Y_i(d') | \mathbf{x}_i, \boldsymbol{\theta}(d))$$

MWQLE

$$\hat{\boldsymbol{\theta}}(d, h) = \operatorname{argmax}_{\boldsymbol{\theta}(d)} n^{-1} \ell_n(\boldsymbol{\theta}(d); h, \tilde{\boldsymbol{\omega}})$$

Newton-Raphson Algorithm

$$\hat{\boldsymbol{\theta}}(d,h)^{(t+1)} = \hat{\boldsymbol{\theta}}(d,h)^{(t)} + \{-\partial^2_{\boldsymbol{\theta}(d)} \ell_n(\hat{\boldsymbol{\theta}}(d,h)^{(t)}; h, \tilde{\boldsymbol{\omega}})\}^{-1} \partial_{\boldsymbol{\theta}(d)} \ell_n(\hat{\boldsymbol{\theta}}(d,h)^{(t)}; h, \tilde{\boldsymbol{\omega}})$$

Expectation-Maximization Algorithm

Multiscale Adaptive Regression Model

Adaptive Estimation and Testing

Sandwich Estimator

$$\mathrm{Cov}[\hat{\boldsymbol{\theta}}(d,h)] \approx \Sigma_n(\hat{\boldsymbol{\theta}}(d,h)) = [\Sigma_{n,1}(\hat{\boldsymbol{\theta}}(d,h))]^{-1}\Sigma_{n,2}(\hat{\boldsymbol{\theta}}(d,h))[\Sigma_{n,1}(\hat{\boldsymbol{\theta}}(d,h))]^{-1}$$

$$\Sigma_{n,1}(\boldsymbol{\theta}(d)) = -\partial^2_{\boldsymbol{\theta}(d)}\ell_n(\boldsymbol{\theta}(d); h, \tilde{\boldsymbol{\omega}}) \text{ and}$$

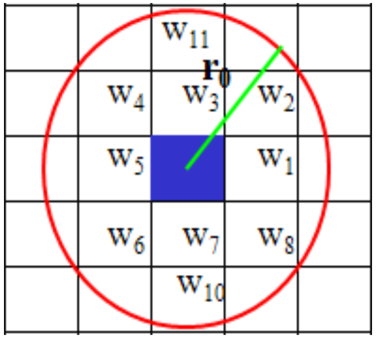
$$\Sigma_{n,2}(\boldsymbol{\theta}(d)) = \sum_{i=1}^{n}[\sum_{d' \in B(d,h)} \tilde{\omega}(d,d';h)\partial_{\boldsymbol{\theta}(d)} \log p(Y_i(d')|\mathbf{x}_i, \boldsymbol{\theta}(d))]^{\otimes 2}$$

Wald Test Statistic

$$[R(\hat{\boldsymbol{\theta}}(d;h)) - \mathbf{b}_0]^T[\partial_{\boldsymbol{\theta}(d)}R(\hat{\boldsymbol{\theta}}(d;h))\hat{\Sigma}_n(\hat{\boldsymbol{\theta}}(d;h))\partial_{\boldsymbol{\theta}(d)}R(\hat{\boldsymbol{\theta}}(d;h))^T]^{-1}[R(\hat{\boldsymbol{\theta}}(d;h)) - \mathbf{b}_0]$$

Multiscale Adaptive Regression Model

Automatic Stop



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

$$K_{loc}(u) = (1 - u^2)_+$$

$$K_{st}(u) = \exp(-u)\mathbf{1}\left(u \le \frac{s+2}{s(\log s + 2)}\right)$$

$$\omega(d, d'; h_s) = K_{loc}(||d - d'||_2 / h_s) K_{st}(D_\theta(d, d'; h_{s-1}) / C_n)$$

As S increases, the first kernel gets larger for any voxel pairs, whereas the second kernel penalizes more and more for the voxel pairs with distinctive features.

Multiscale Adaptive Regression Model

- Voxel size is much larger than the sample size
- Sample size increases to infinity, whereas voxel size is fixed
- A multiscale adaptive procedure
- Propagation-separation conditions do not work

Multiscale Adaptive Regression Model

log(Voxel size)<<Cn << sample size

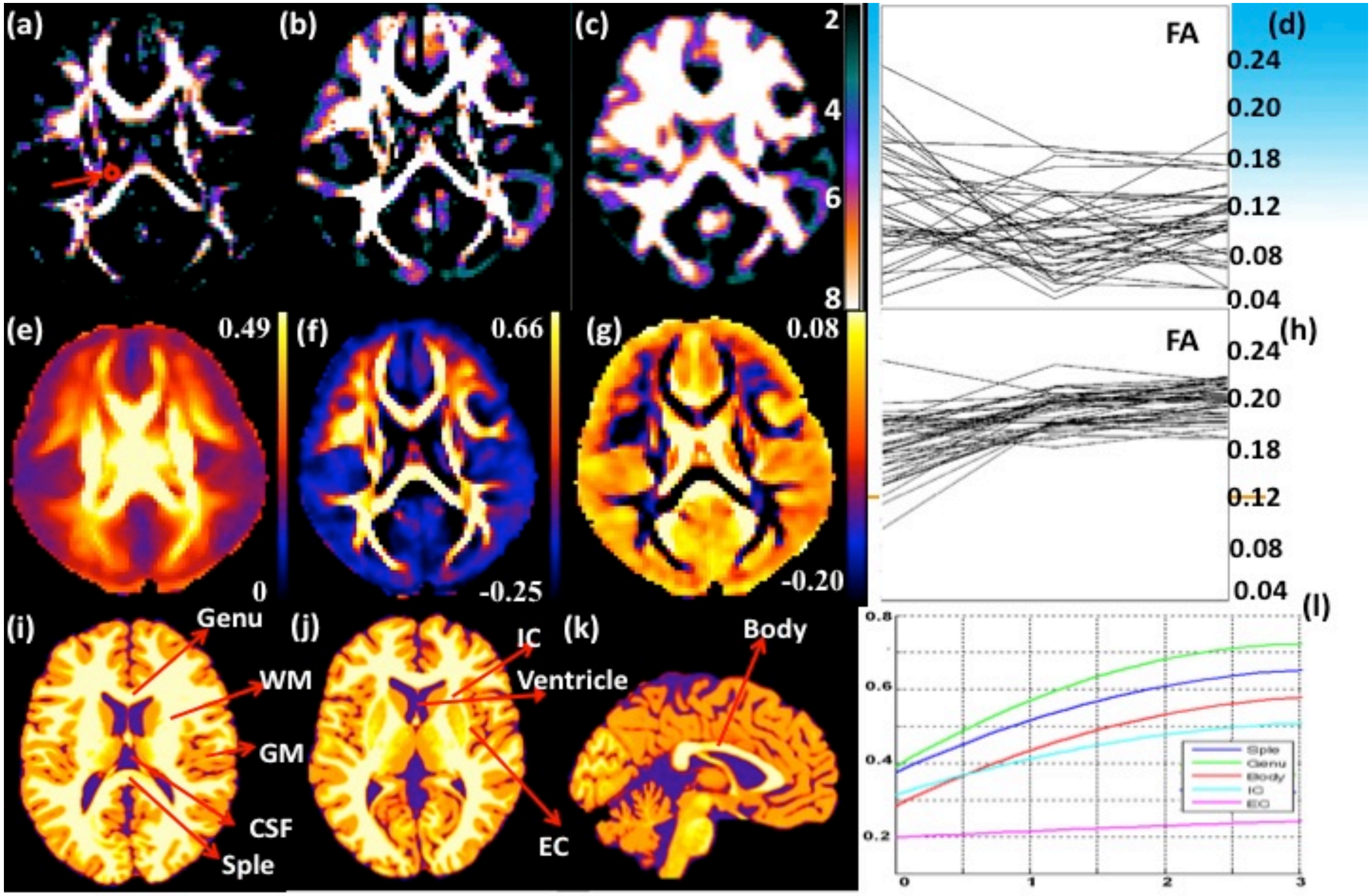
Kernel functions

Conditions for M-estimators hold uniformly

- Weak Consistency
- Asymptotical Normality
- Asymptotically Chi-squared distribution

Infant Brain Development Data

- **Objective:** We want to assess the brain structure change in the early brain development.
- **Subject:** 38 infants.
- **Image:** Diffusion-weighted images and T1 weighted images were acquired for each subject at 2 weeks, 1 and 2 years old.
- **Method:** Voxel-wise imaging analysis and MARM.


(a)
(b)
(c)
2
4
6
8
FA
(d)
0.24
0.20
0.18
0.12
0.08
0.04
(e)
0.49
0
(f)
0.66
-0.25
(g)
0.08
-0.20
FA
(h)
0.24
0.20
0.18
0.12
0.08
0.04
(i)
Genu
WM
GM
CSF
Sple
(j)
IC
Ventricle
EC
(k)
Body
(l)

New Developments

Adaptive Neighhoods

Adaptive Weights

Cross-sectional, longitudinal, twin and family studies

Robust Procedure

Parametric and Nonparametric Components

SVCM

Decomposition:

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

<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

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

$$\Sigma_\varepsilon(d, d') = \Sigma_\varepsilon(d, d) 1(d = d')$$

3D volume/
2D surface

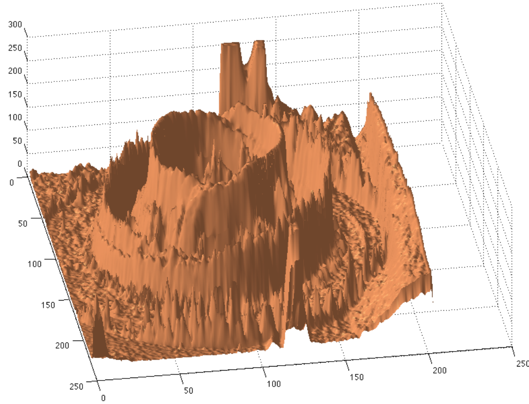
Covariance operator:

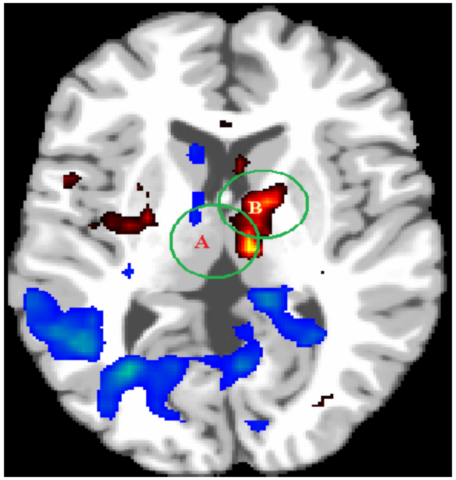
$$\Sigma_y(d, d') = \Sigma_\eta(d, d') + \Sigma_\varepsilon(d, d) 1(d = d')$$

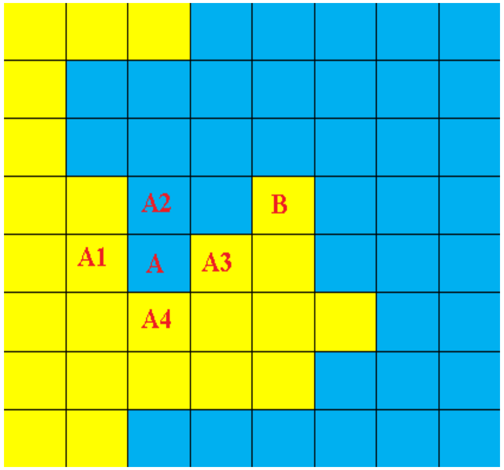
SVCM

Piecewise Smoothness Condition $B_k(d)$

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



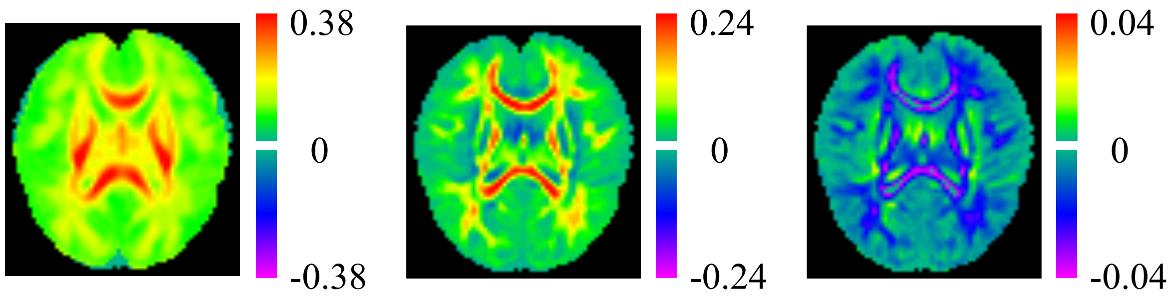




Challenging Issues

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

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



SVCM

Least Squares Estimates

$$\hat{B}(d;h_0) = \left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1} \sum_{i=1}^{n} x_i y_i(d)$$

Smoothing residual images

$$\hat{\eta}_i(d) = S(y_i(d) - x_i^T \hat{B}(d;h_0))$$

Estimate covariance operator

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

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

Adaptively Smoothing LSEs

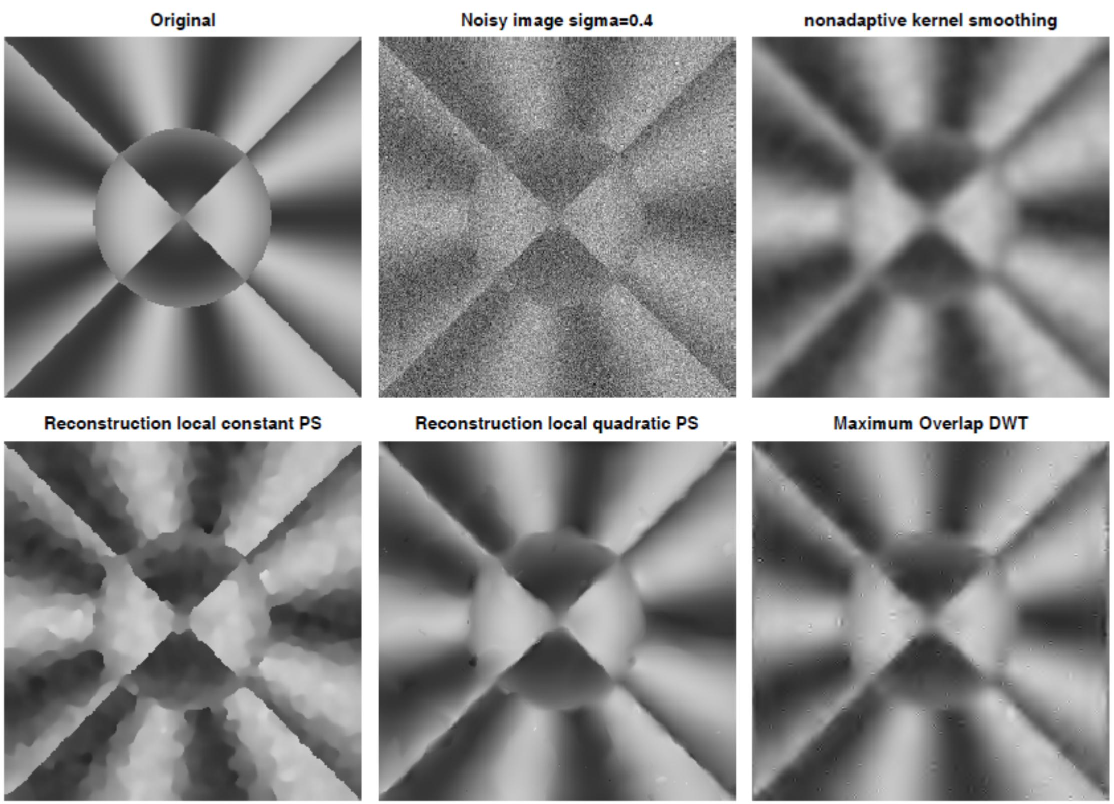
$$\hat{\beta}_j(d;h_s) = \sum_{d' \in B(d,h_s)} w_j(d,d';h_s)\hat{\beta}_j(d;h_0) \Big/ \sum_{d' \in B(d,h_s)} w_j(d,d';h_s)$$

Calculate standard deviation

Smoothing Methods

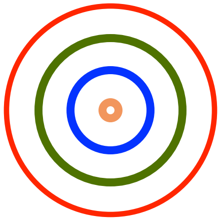
Propogation-Seperation Method

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



Features

- Increasing Bandwidth



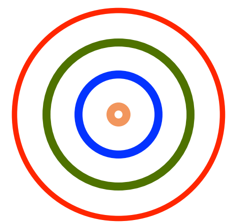
$$0 < h_0 < h_1 < \cdots < h_S = r_0$$

- Adaptive Weights
- Adaptive Estimates

Smoothing Methods

Propogation-Seperation Method

At each voxel d



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

- Increasing Bandwidth
- Adaptive Weights
- Adaptive Estimates

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

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

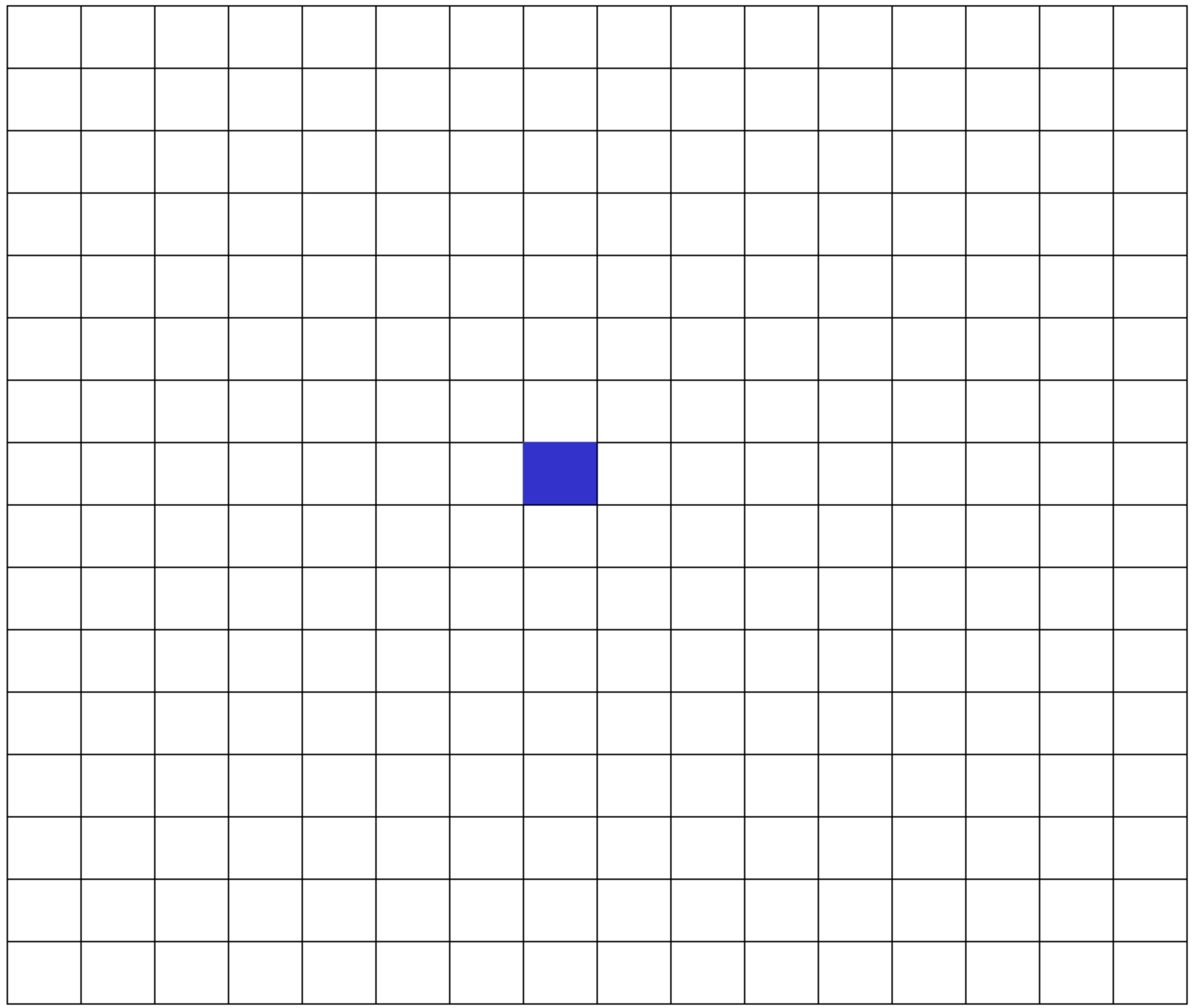
$\omega(d,d';h_2)$

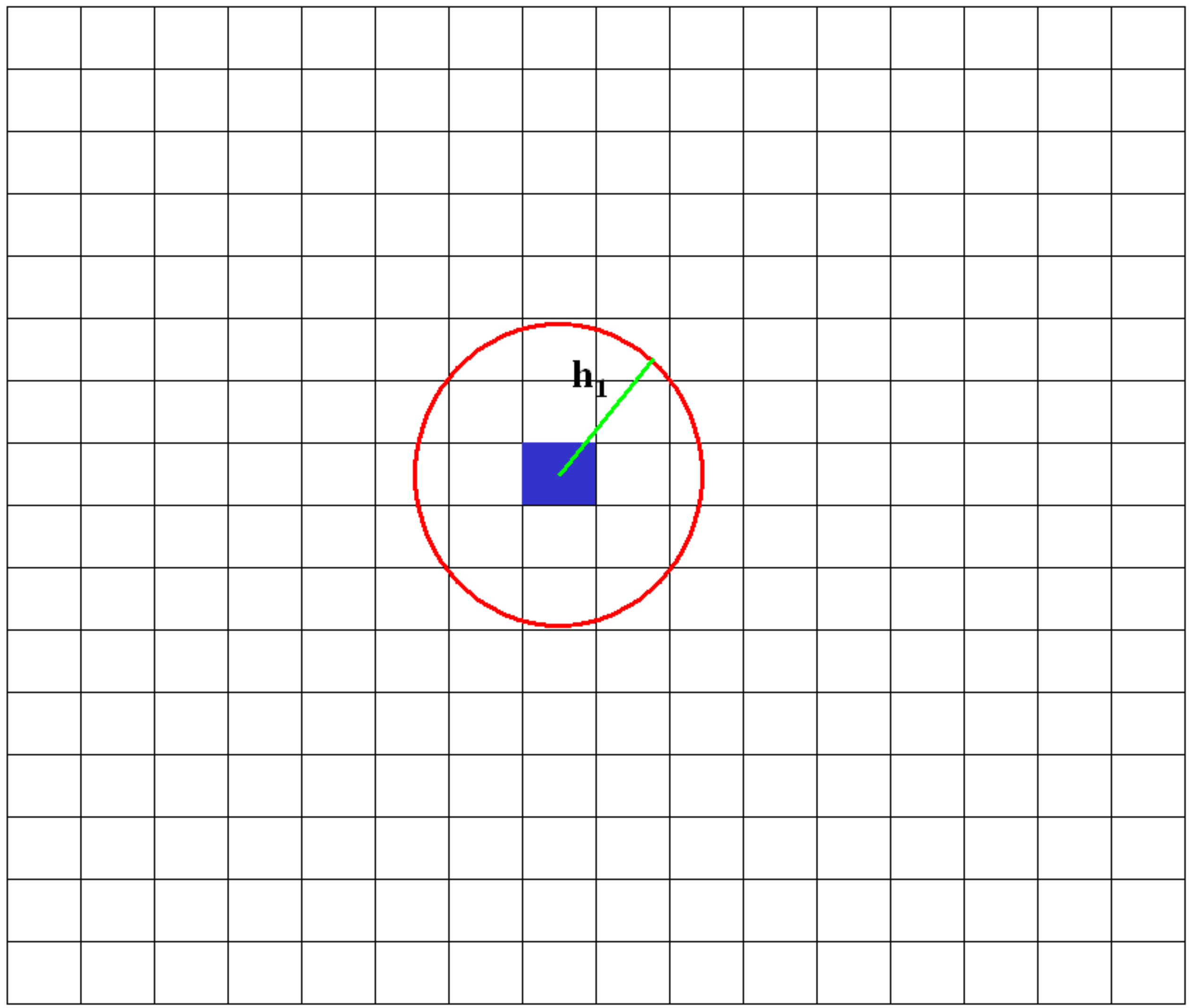
$\hat{\mu}(d;h_0)$ $\hat{\mu}(d;h_1)$ $\hat{\mu}(d;h_S)$

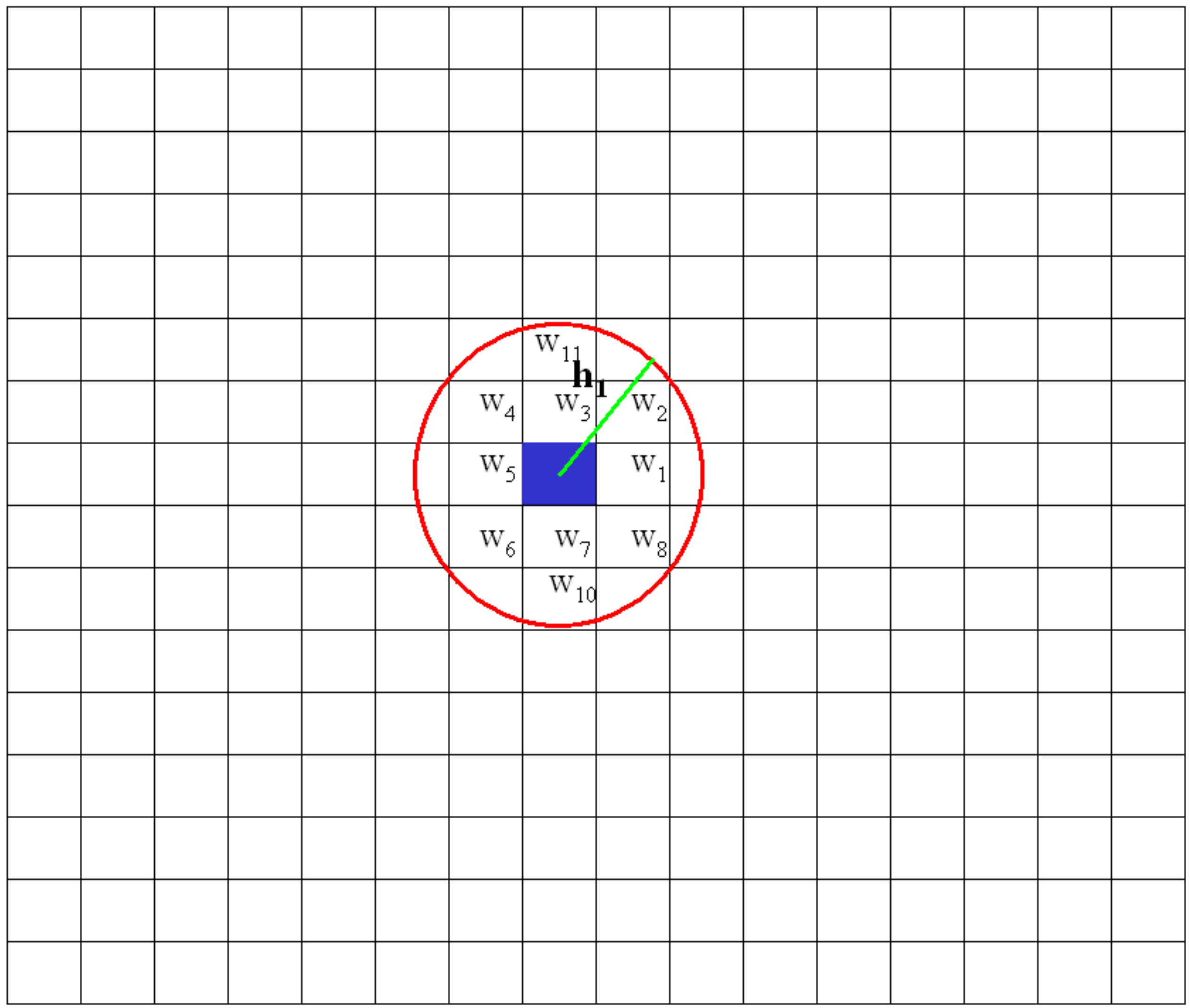
Stopping Rule

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

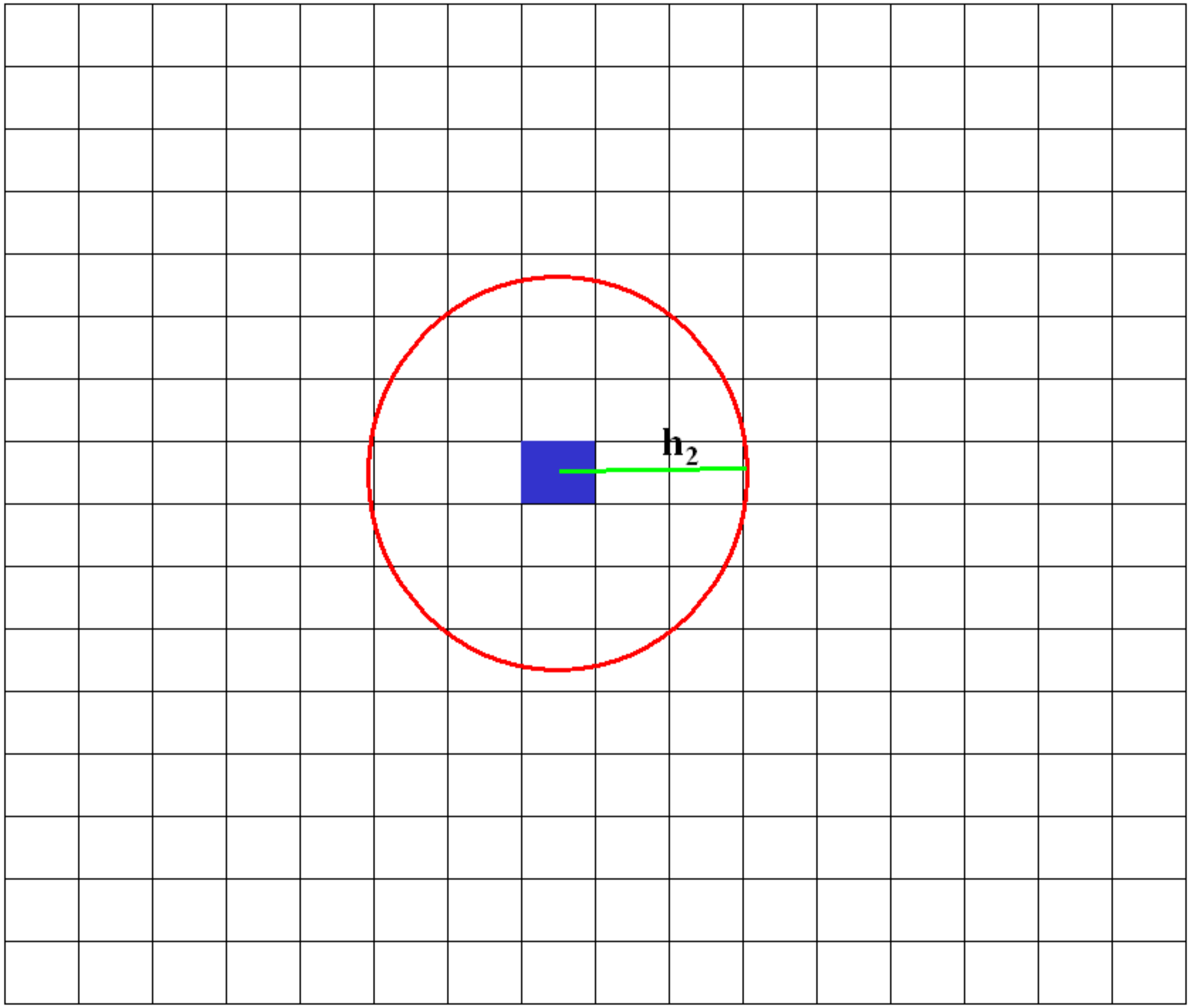
$$D_\mu(d,d';h_{s-1}) = \rho(\hat{\mu}(d;h_{s-1}), \hat{\mu}(d';h_{s-1}))$$

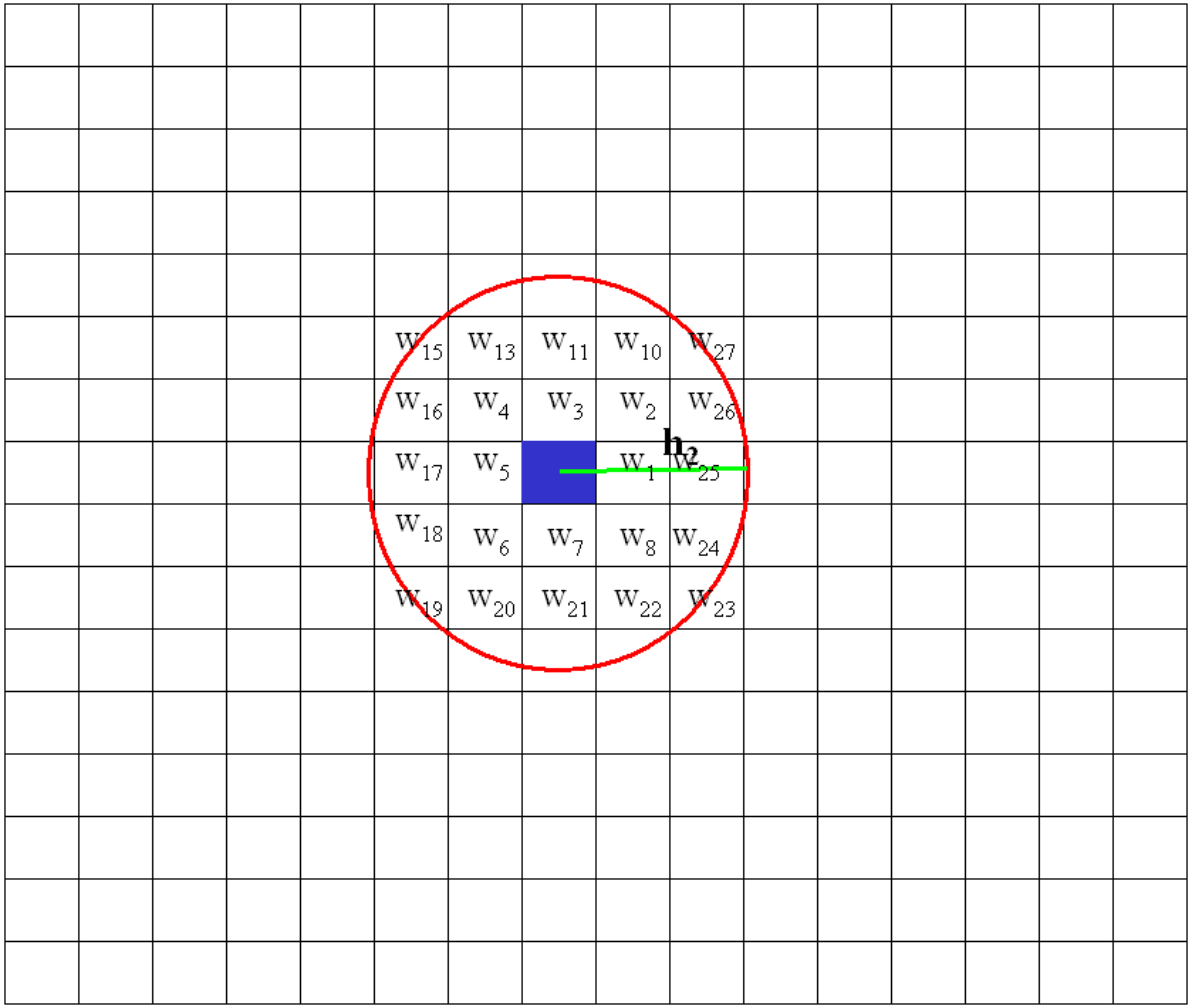



h_1



W11
h1
W4
W3
W2
W5
W1
W6
W7
W8
W10


h_2



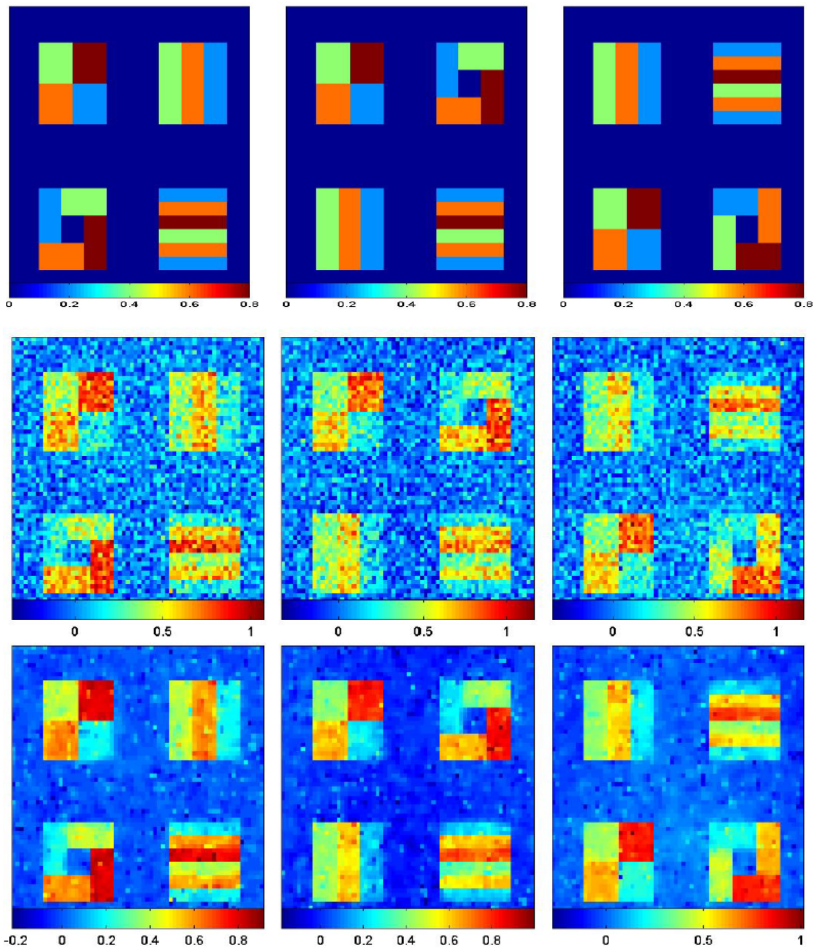
W15 W13 W11 W10 W27
W16 W4 W3 W2 W26
W17 W5 W1 W25
h2
W18 W6 W7 W8 W24
W19 W20 W21 W22 W23

Simulation

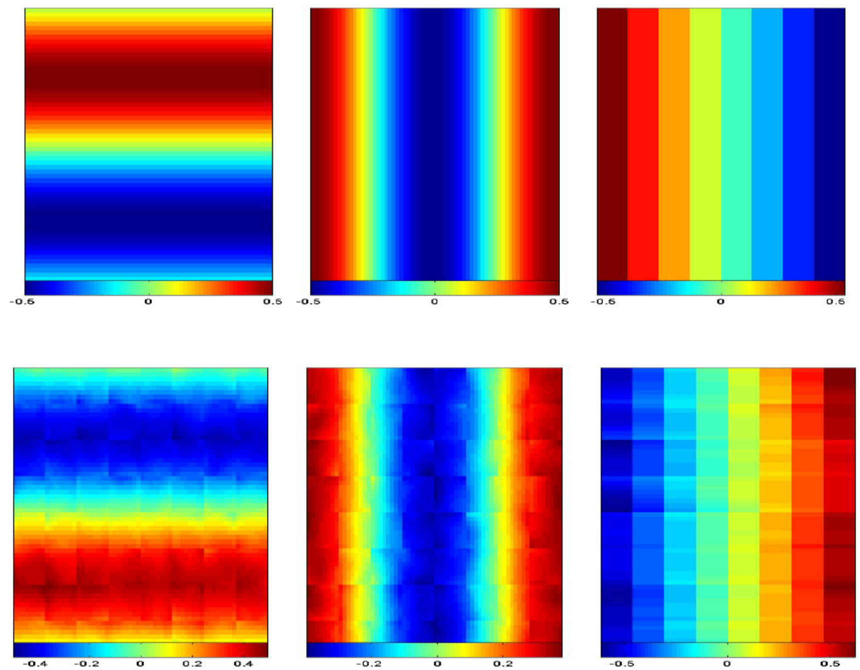
$$Y_i(d) = X_i^T B(d) + \sum_{j=1}^{3} \xi_{ij}\psi_j(d) + \epsilon_i(d);$$

- $d = (d_1, d_2, d_3)^T \in \mathcal{D}$, a $64 \times 64 \times 8$ 3D image
- $X_i = (x_{i1}, x_{i2}, x_{i3})^T$ with $x_{i1} = 1$, $x_{i2} \sim$ Bernoulli(0.5), $x_{i3} \sim$ Unif (1,2)
- $B(d) = (\beta_1(d), \beta_2(d), \beta_3(d))^T$ with $\beta_1(d)$, $\beta_2(d)$ and $\beta_3(d) \in \{0, 0.2, 0.4, 0.6, 0.8\}$
- $\xi_{i1} \sim N(0, 0.6), \quad \xi_{i2} \sim N(0, 0.4), \quad \xi_{i3} \sim N(0, 0.2)$ $\epsilon_i(d) \sim N(0, 1)$
- $\psi_1(d) = 0.5\sin(2\pi d_1/64), \quad \psi_2(d) = 0.5\cos(2\pi d_2/64)$ $\psi_3(d) = \sqrt{1/2.625}(9/8 - d_3/4)$

Simulation

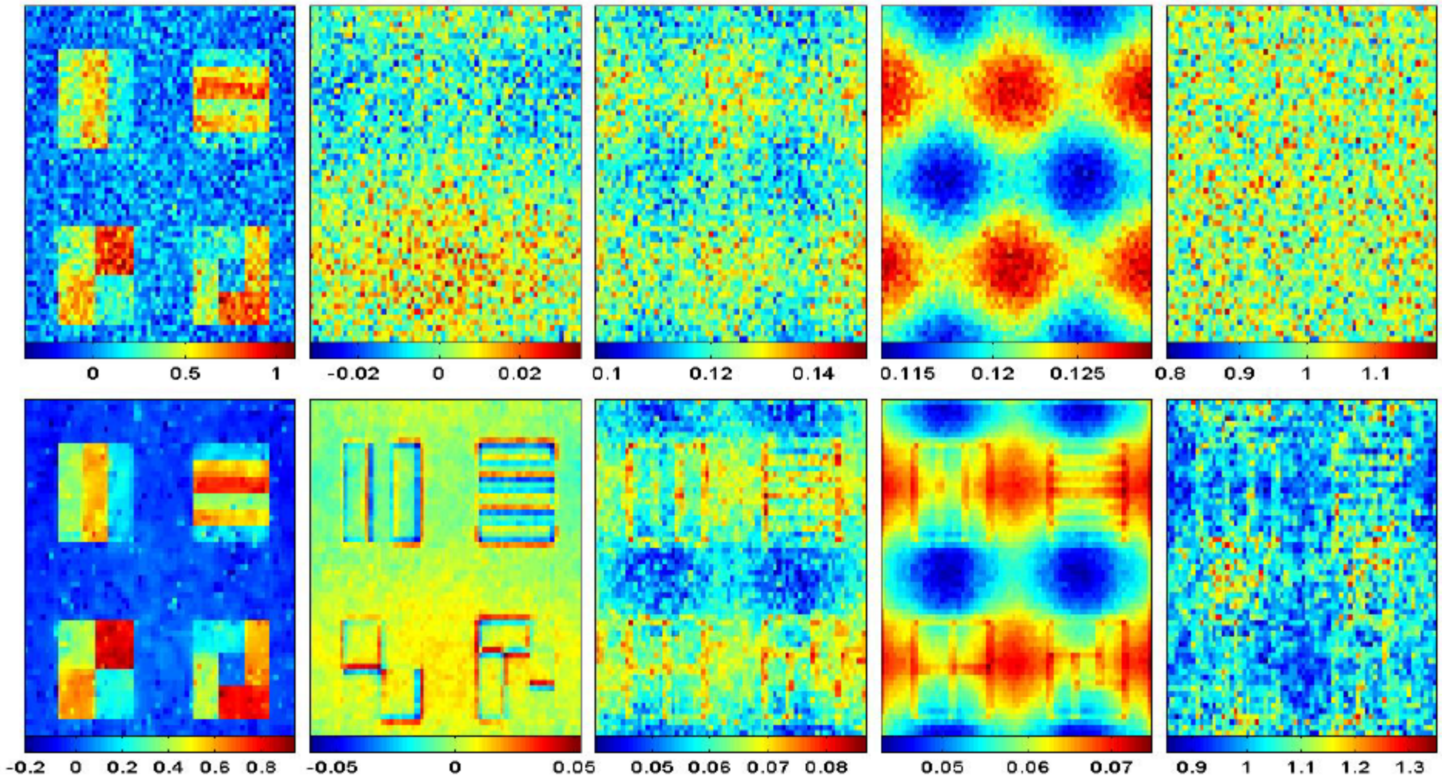


From up to down: initial and adaptive estimates; left to right: $\beta_1(d)$, $\beta_2(d)$, and $\beta_3(d)$.



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

Simulation

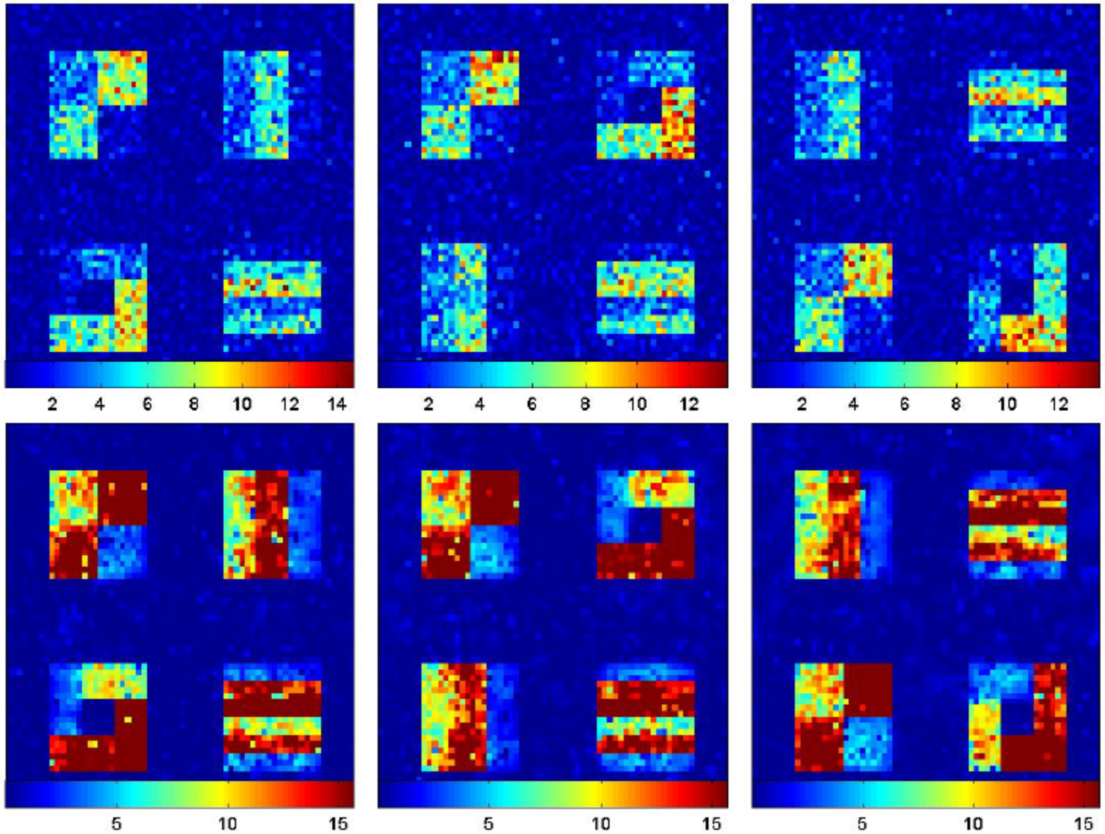


From up to down: initial and adaptive estimations; left to right: β, Bias, RMS, SD and RE (RMS/SD).

Simulation

$\beta_2(d)$		$n=60$ h_0	h_5	h_{10}	$n=80$ h_0	h_5	h_{10}
0	BIAS	0.000	0.002	0.002	0.000	0.002	0.002
	RMS	0.139	0.070	0.068	0.121	0.061	0.060
	SD	0.140	0.074	0.067	0.121	0.064	0.058
	RE	0.993	0.942	1.026	1.001	0.947	1.036
0.2	BIAS	0.000	-0.006	-0.007	0.001	-0.005	-0.006
	RMS	0.140	0.074	0.073	0.122	0.065	0.064
	SD	0.141	0.077	0.070	0.122	0.067	0.061
	RE	0.993	0.963	1.043	1.000	0.971	1.056
0.4	BIAS	0.000	0.001	0.001	-0.001	0.001	0.001
	RMS	0.140	0.075	0.074	0.122	0.066	0.065
	SD	0.141	0.078	0.071	0.122	0.068	0.062
	RE	0.992	0.962	1.041	1.001	0.973	1.055
0.6	BIAS	0.000	-0.006	-0.007	0.000	-0.004	-0.005
	RMS	0.139	0.073	0.072	0.121	0.063	0.063
	SD	0.140	0.075	0.069	0.121	0.066	0.059
	RE	0.994	0.969	1.052	0.999	0.967	1.053
0.8	BIAS	-0.001	-0.008	-0.010	0.000	-0.006	-0.008
	RMS	0.141	0.075	0.074	0.123	0.066	0.066
	SD	0.143	0.081	0.074	0.123	0.070	0.064
	RE	0.990	0.935	1.008	1.001	0.949	1.025

Simulation



From up to down: $-log_{10}(p)$ of initial and adative estimates; left to right: $\beta_1(d)$, $\beta_2(d)$, and $\beta_3(d)$.

Simulation

$\beta_2(d)$		$n=60$ ES	$n=60$ SE	$n=80$ ES	$n=80$ SE
0	h_0	0.048	0.015	0.050	0.016
	h_{10}	0.036	0.016	0.040	0.019
0.2	h_0	0.282	0.033	0.370	0.035
	h_{10}	0.777	0.107	0.870	0.081
0.4	h_0	0.794	0.030	0.895	0.024
	h_{10}	0.994	0.006	0.998	0.003
0.6	h_0	0.988	0.008	0.998	0.003
	h_{10}	1.000	0.001	1.000	0.000
0.8	h_0	1.000	0.001	1.000	0.000
	h_{10}	1.000	0.000	1.000	0.000

Estimates (ES) and standard errors(SE) of rejection rates

Real Data

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

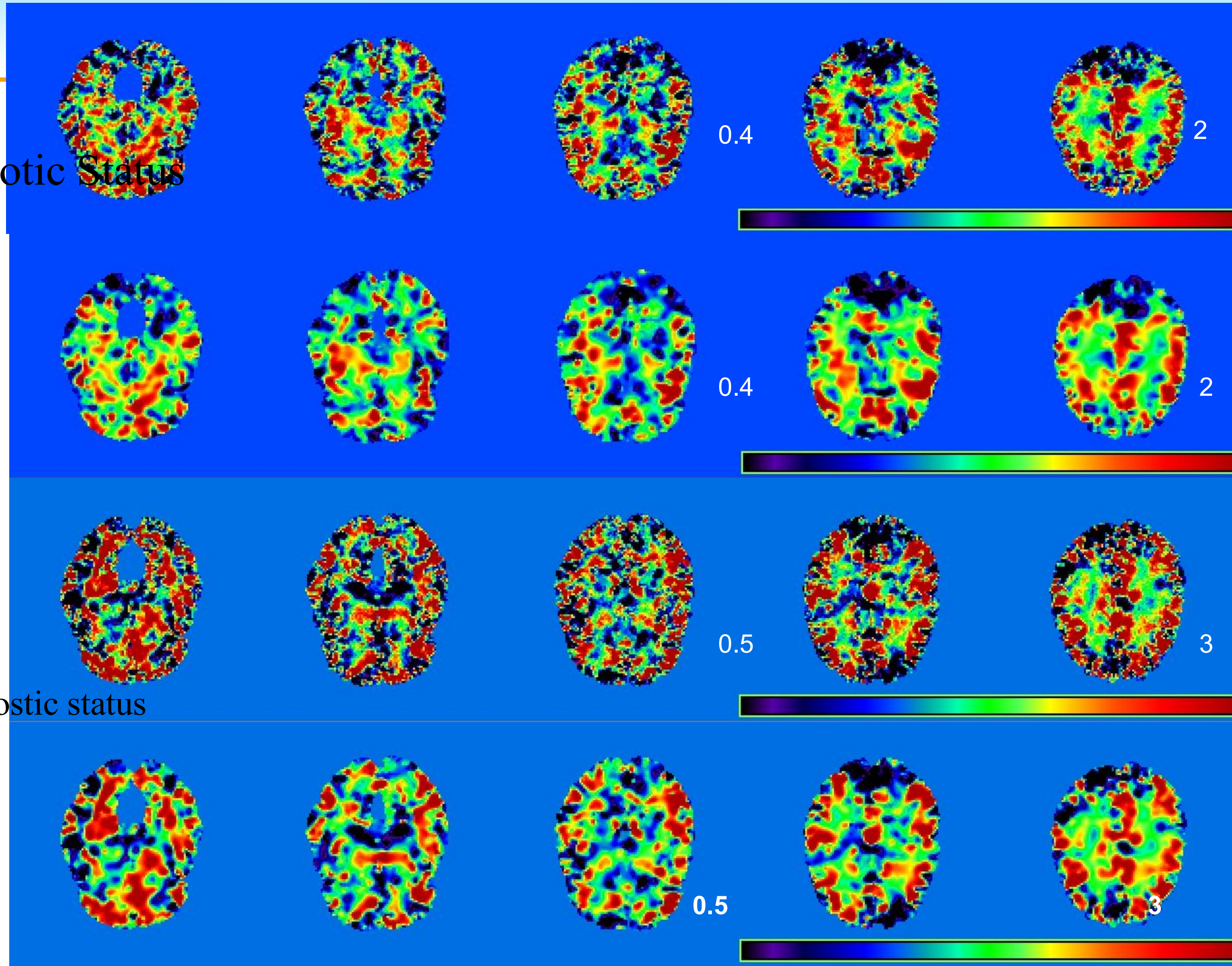
ADHD200 NYU Data

Subjects: 174 subjects, 99 normal and 75 ADHD-combined
Response: RAVEN map
Covariates: age, gender, group, G*Age, G*Gender and whole brain volume
Goal: Group*Age and Group*Gender

Interaction effect estimates

h_0

Age × Diagnotic Status



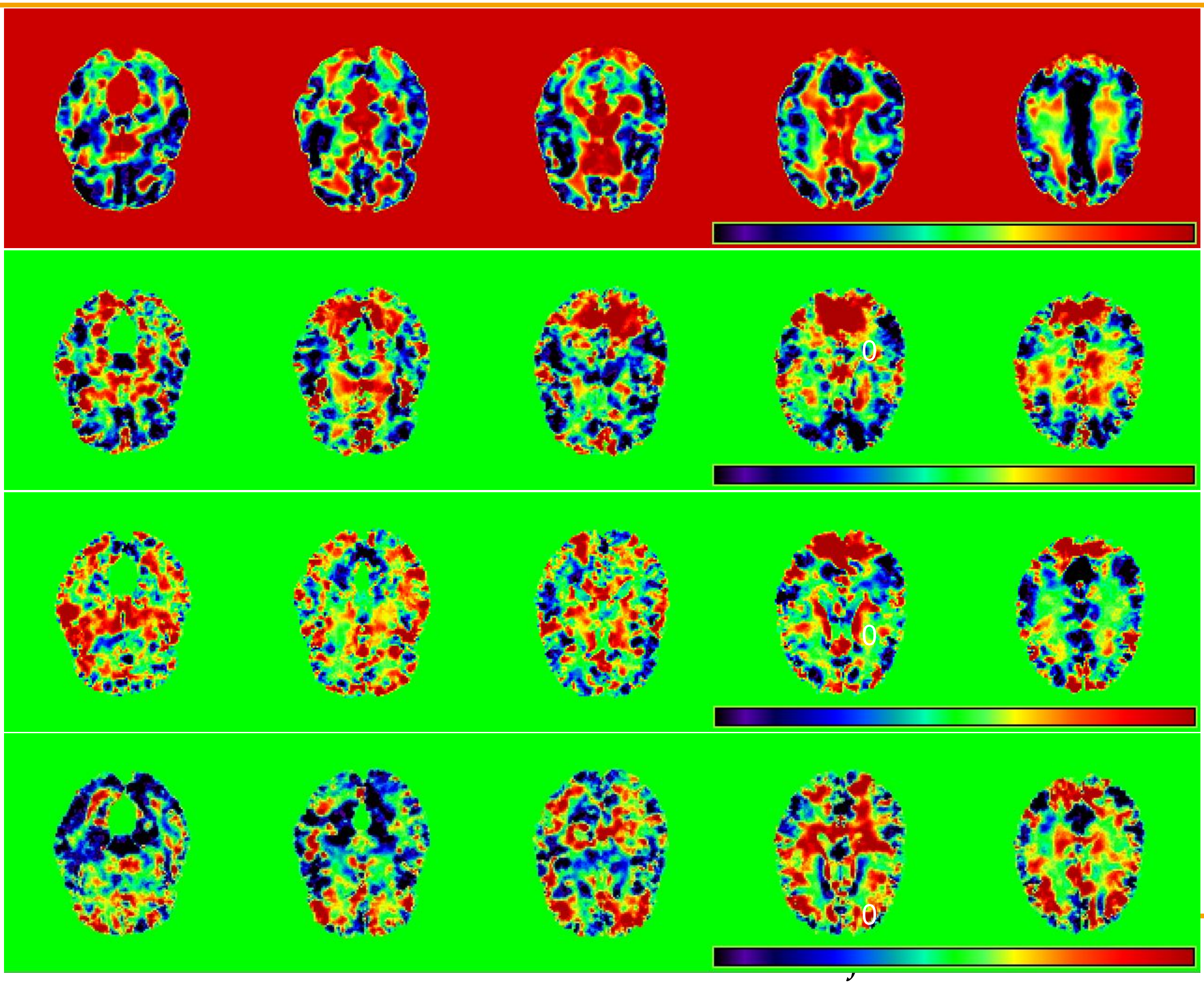
h_{10}

h_0

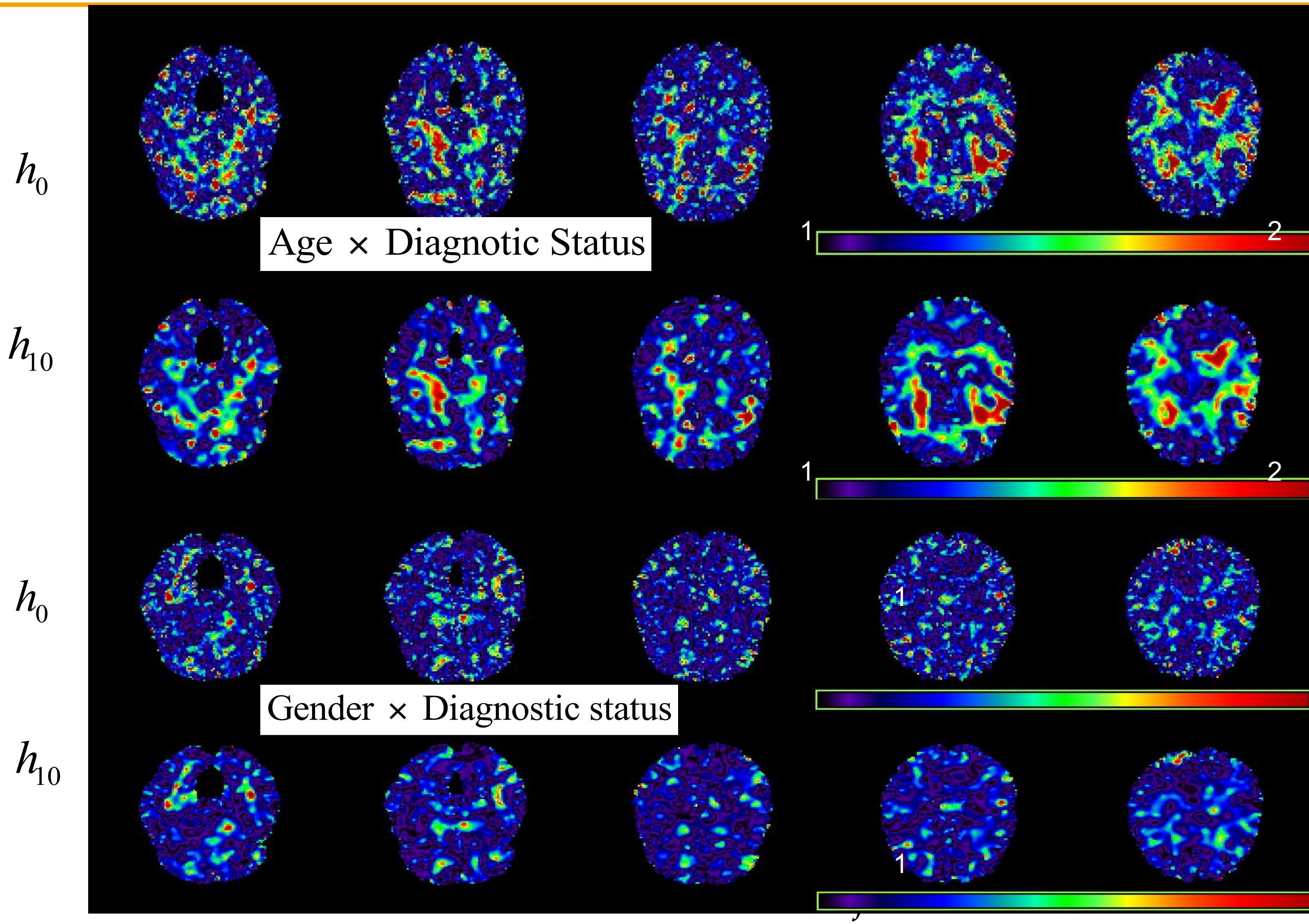
Gender × Diagnostic status

h_{10}

First Four Eigenfunctions

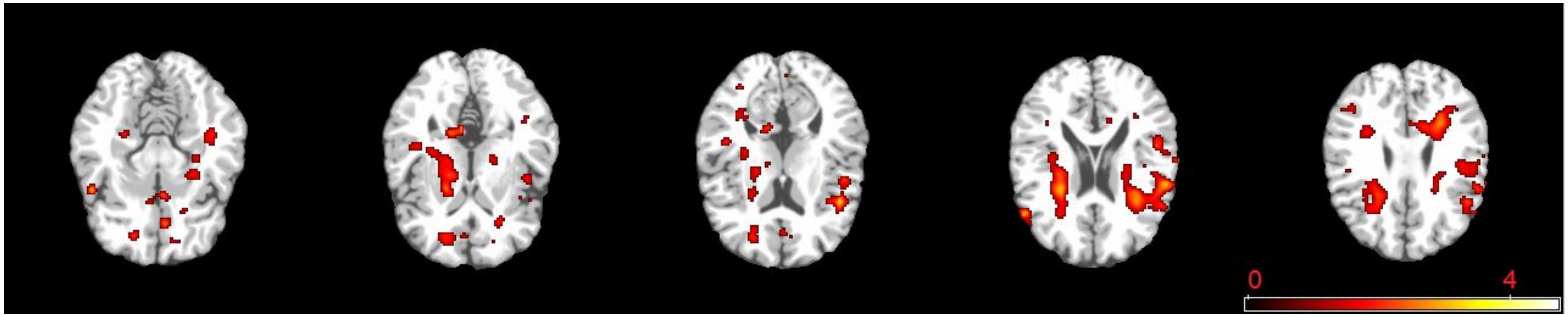


-log10(p) Maps

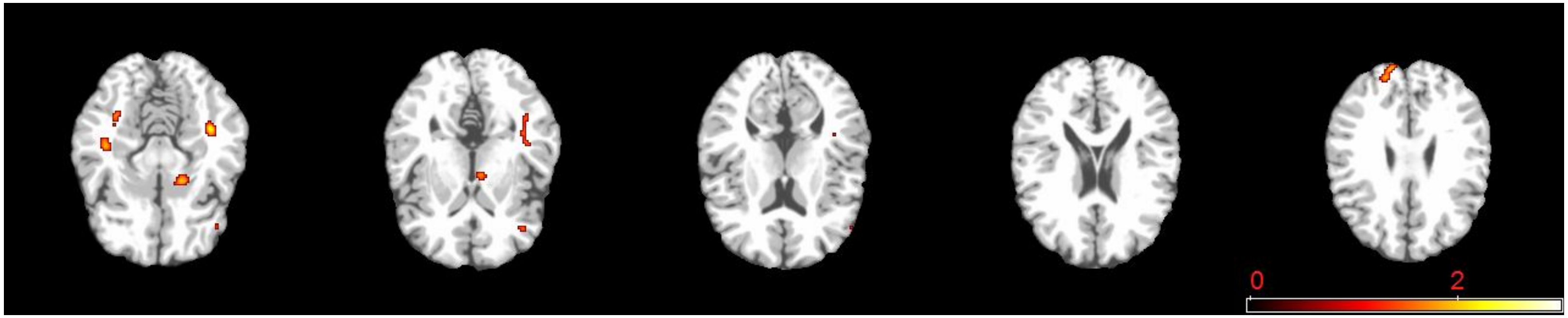

h_0
h_{10}
h_0
h_{10}
Age × Diagnotic Status
Gender × Diagnostic status
1
2
1
2
1
1

Significant Regions

Age × Diagnotic Status

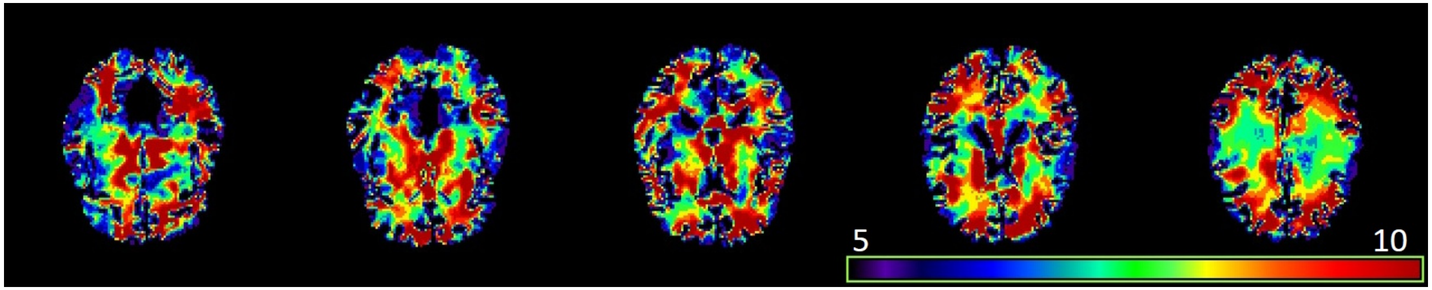


Gender × Diagnostic status

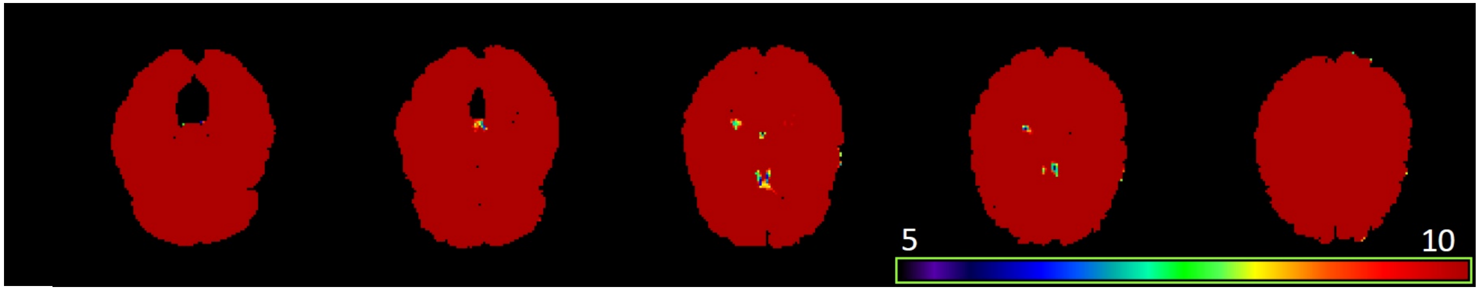


Prediction

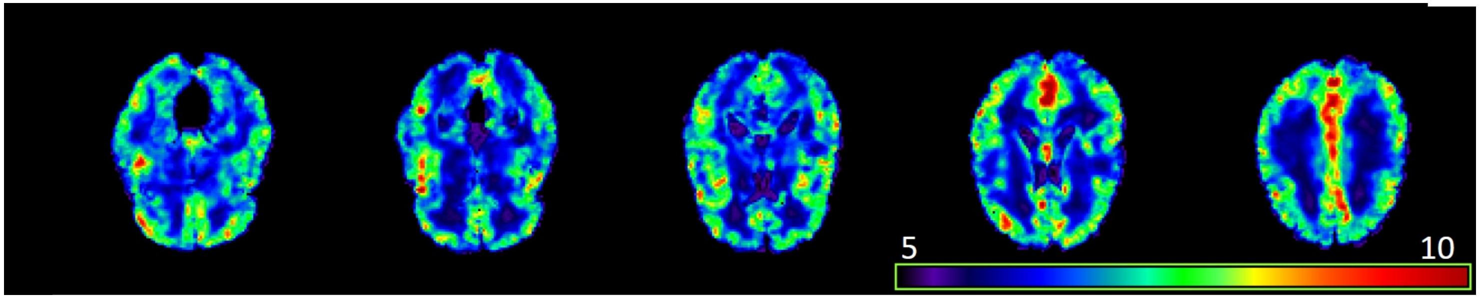
Raw Image



GLM



SVCM



Prediction

The ADNI data

Focus on the Mild Cognitive Impairment people

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

The imaging data: radial distance obtained from left and right hippocampus, 15000 dimensional vector each

The clinical covariates: Gender, Handedness, Marital Status, Education length, Retirement and Age.

The genetics covariates: APOE4 genotypes

Semiparametric functional linear Cox regression

$$h_i(t) = h_0(t)\exp\left(\sum_{k=1}^{p}\beta_k X_{ik} + \int Z_i(s)\gamma(s)ds + \int Z_i^*(s)\gamma^*(s)ds\right),$$

Using Functional Principal Component Analysis

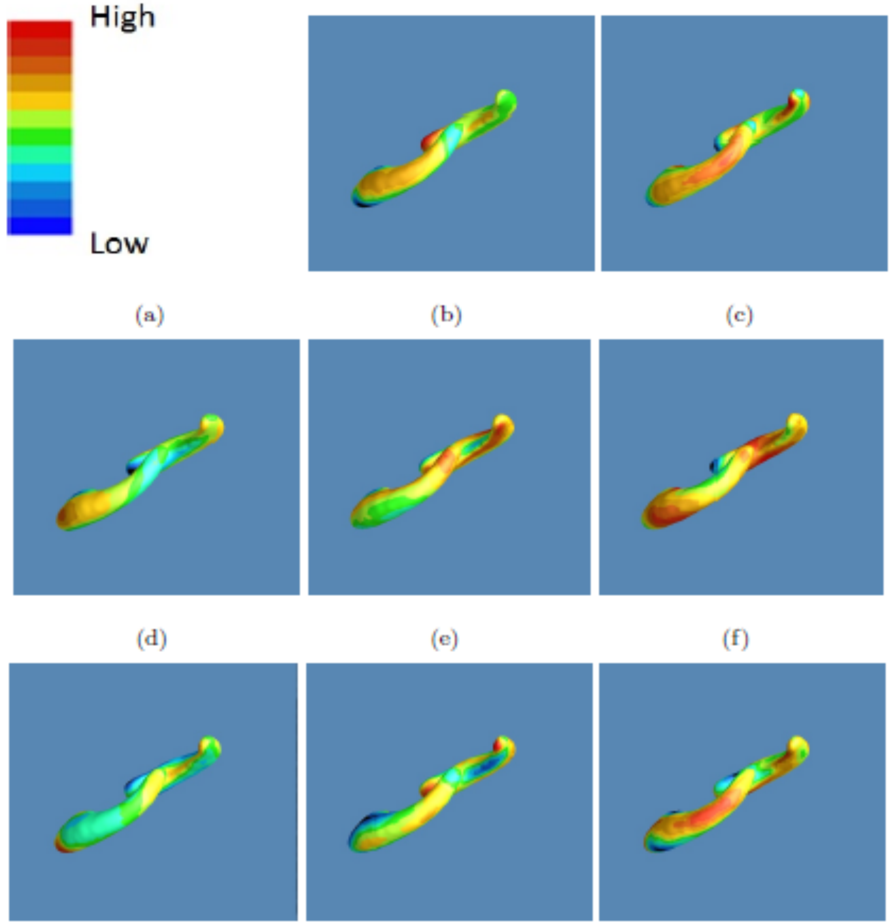
For the first functional predictor, the first three functional principal components are significant.

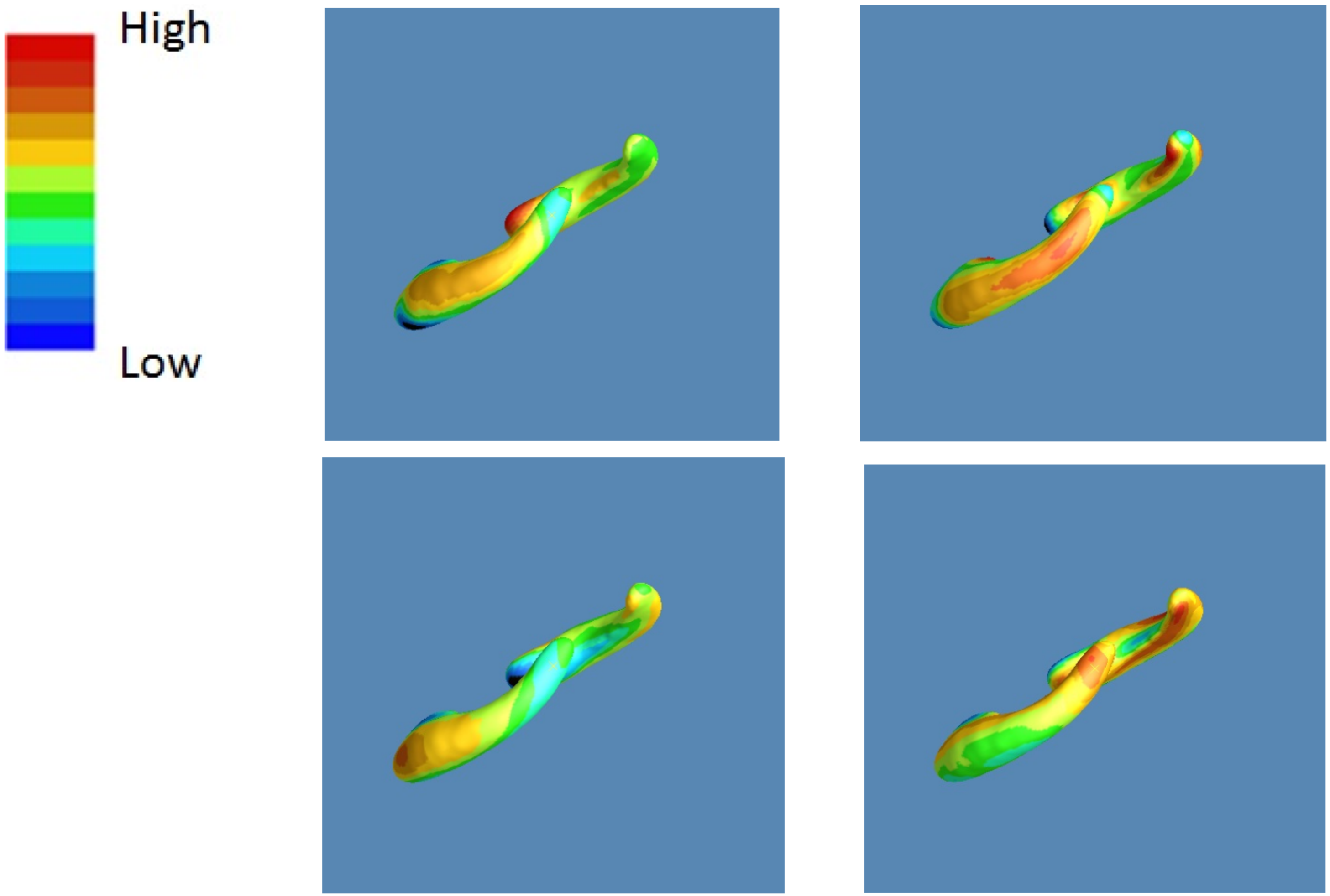
For the second functional predictor, the first and the fifth components are significant.

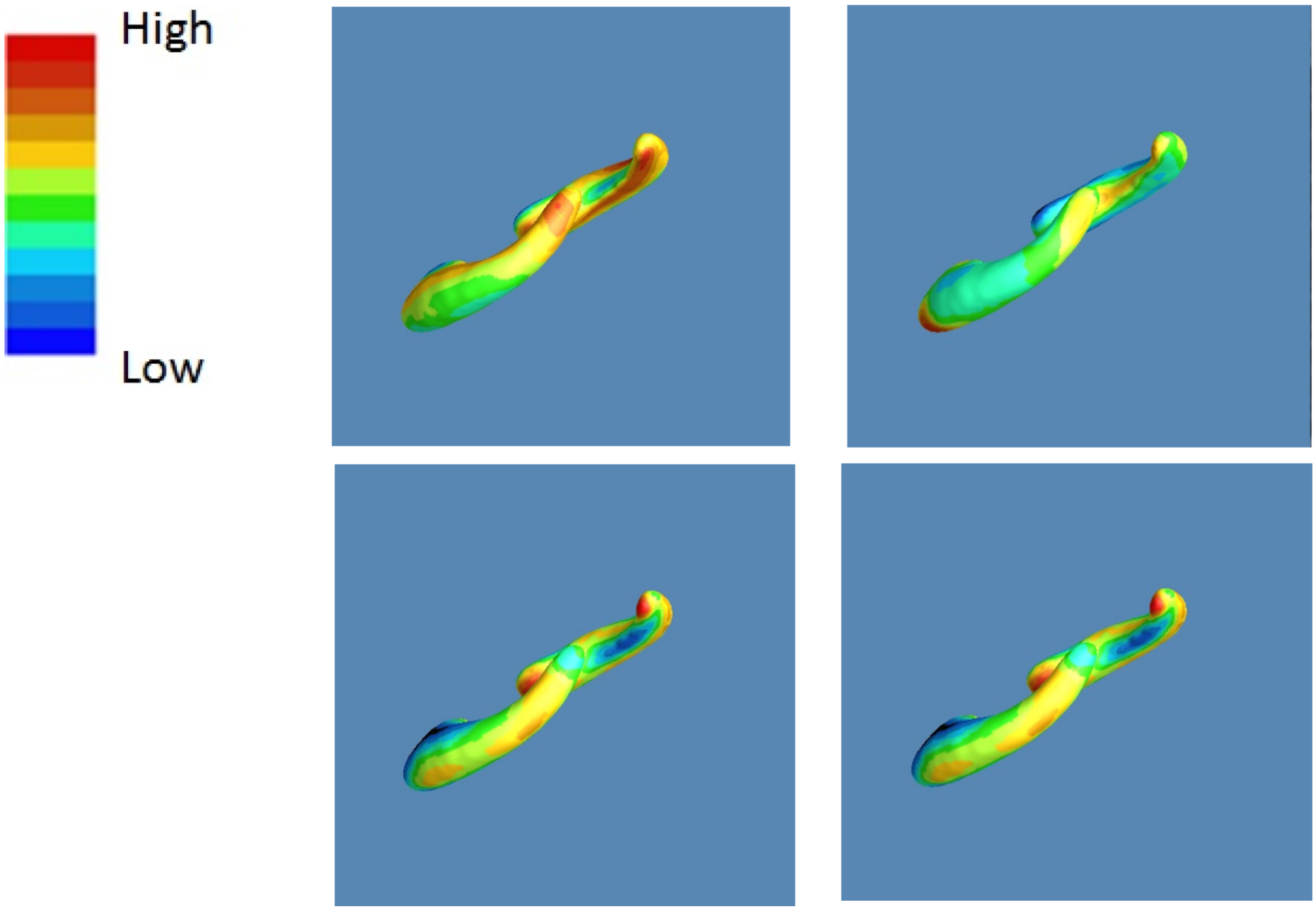
Indicates that both left hippocampus and right hippocampus have significant effect on the conversion.

For the clinical and genetics covariates, the gender, age and the genotype of the second allele in APOE4 are significant.

Panel (a) is the color bar illustration. Panel (b) are the estimated of the coefficient functions. Panel (c)-(i) represent the first seven estimated eigenfunctions for both predictors.




High
Low


High
Low

Acknowledgement