Modeling High-Dimensional Functional and Image Data

Jeffrey S. Morris

The University of Texas MD Anderson Cancer Center

Houston, TX USA

BASS XIX, November 5, 2012

Savannah, GA

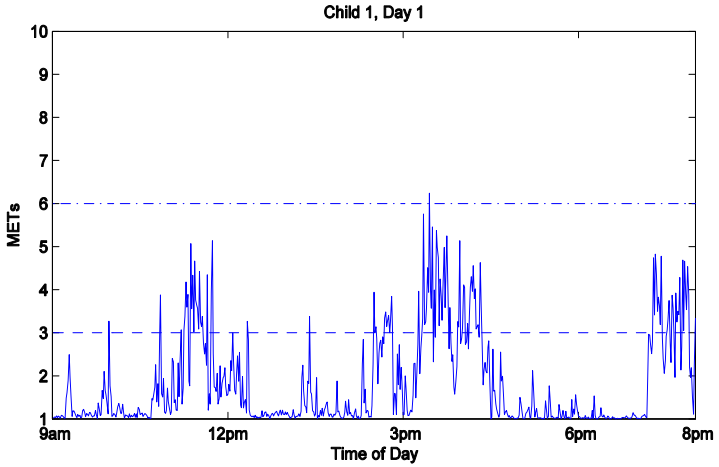
Object Data

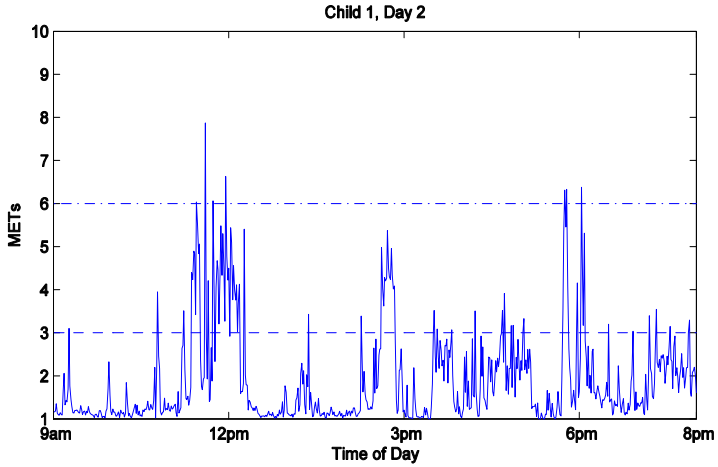
- Technological innovations have produced modern biomedical data that are increasingly complex and high-dimensional
- Object Data: data consisting of multiple (often many) measurements on some type of structured space.
  - **Functional Data**: Time series, measurements on spatial grids, spectral data; e.g. accelerometers, copy number, mass spectra
  - **Quantitative Image Data**: pixel intensities represent some quantitative measure; e.g. fMRI, 2DGE proteomics, LC-MS/GC-MS
  - **Functions on other Manifolds**: spheres or closed surfaces; e.g. ophthalmological data; corticol surface thickness
  - **Other Objects**: shapes, trees, graphs (pathways)
  - **Multi-way Objects**: time-space data, spatial functional data, longitudinal functional data; e.g. ERP, fMRI
  - **Genomic Data**: View entire genome as single structured object
- Internal structure can be simple and driven by basic geometry (space/time proximity), or can be more complex and driven by underlying biology (functional connectivity/pathways)
  - Efficient statistical methods should account for this structure in the modeling. (structure~correlation)

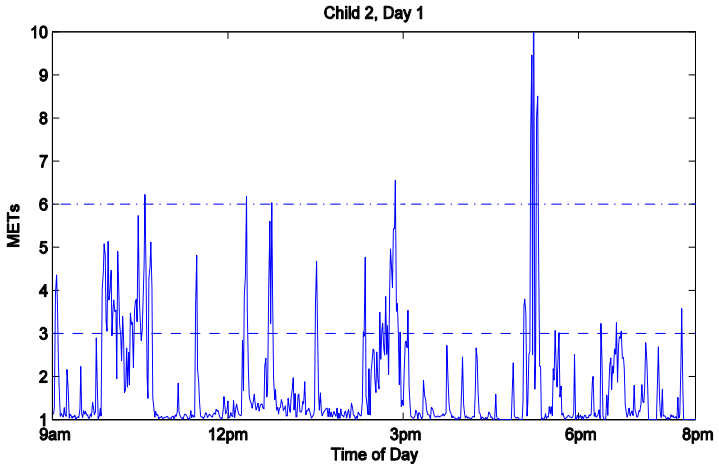
2-D Gel Electrophoresis (2dGE)

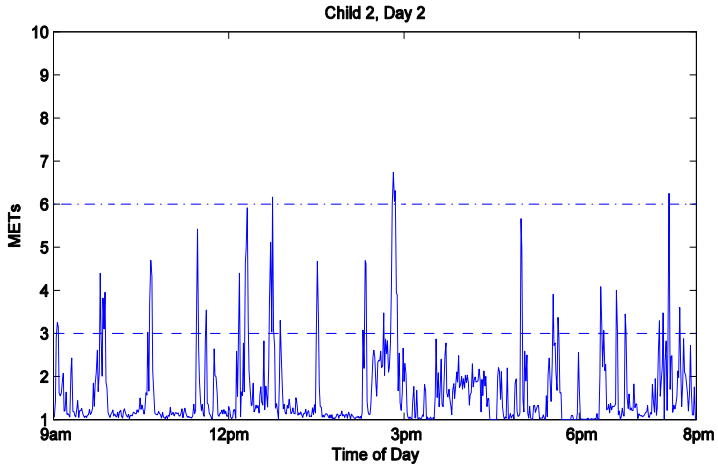
- Each gel scanned, resulting in 2d image (1 MP)
- Frequently have multiple gels per individual
- **Goal**: Associate proteins with factors of interest
- Other assays: MS, lipids + metabolites

Accelerometer Data



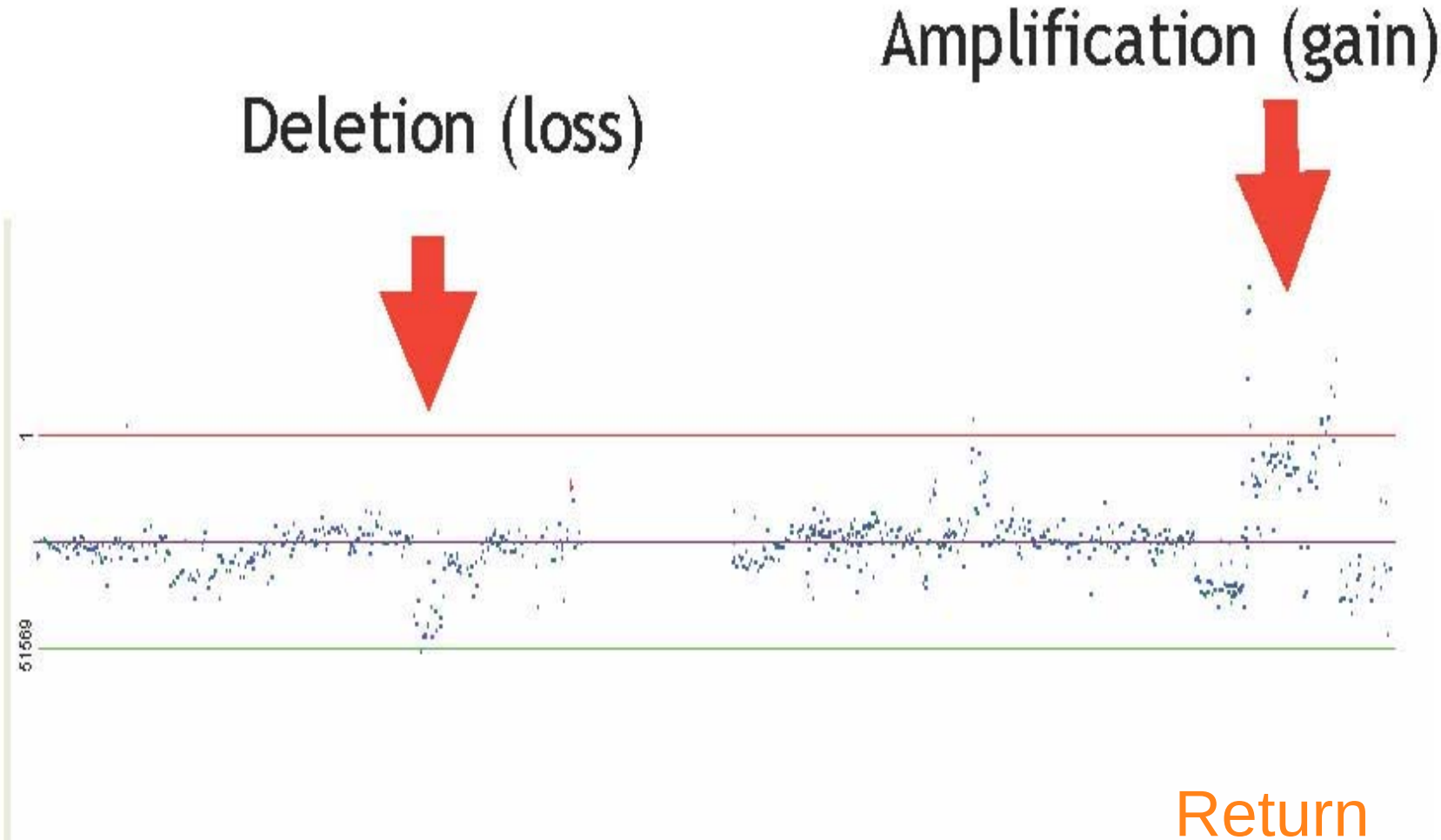






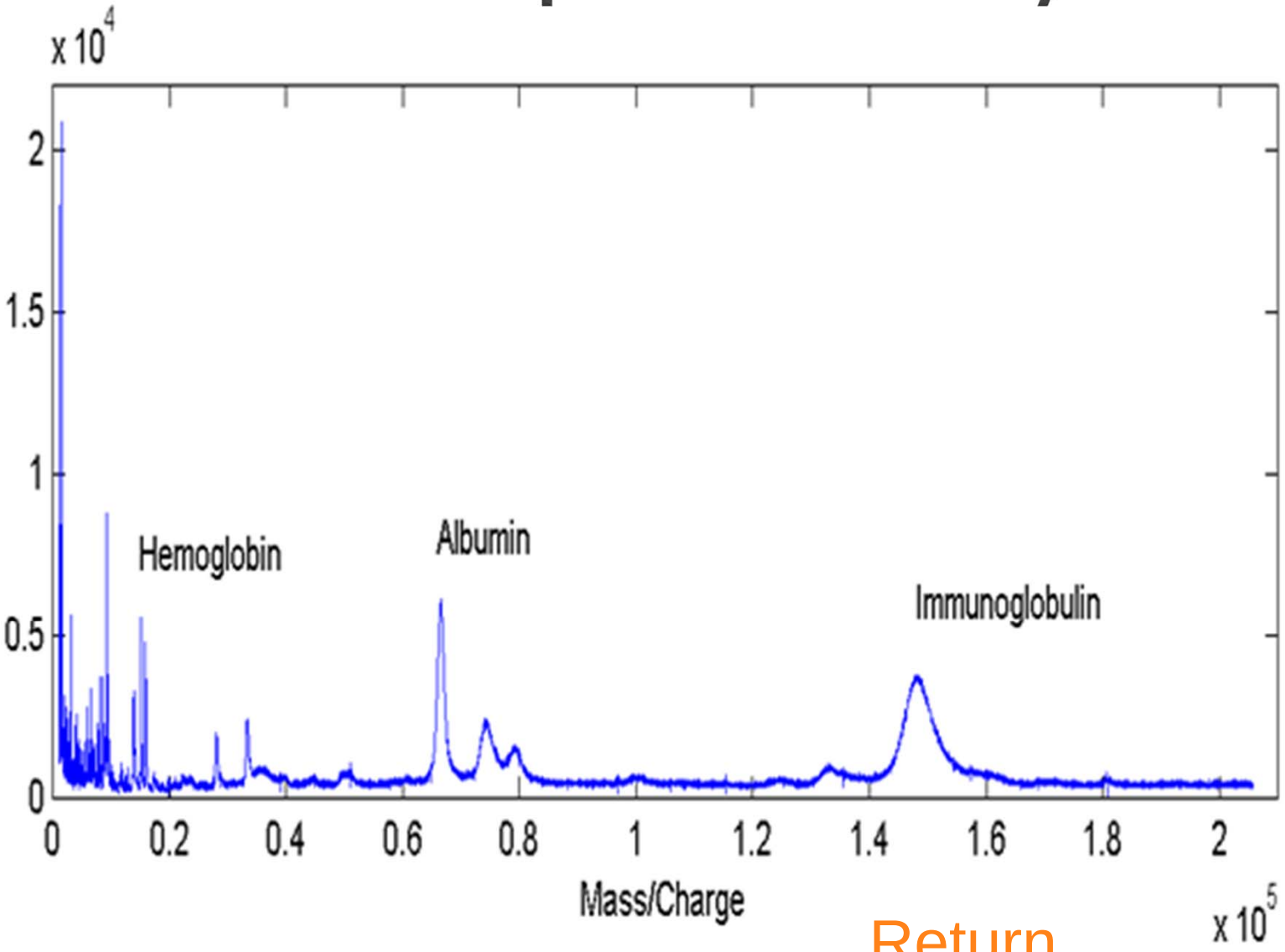
Return

DNA Copy Number Cancer Genomics Study



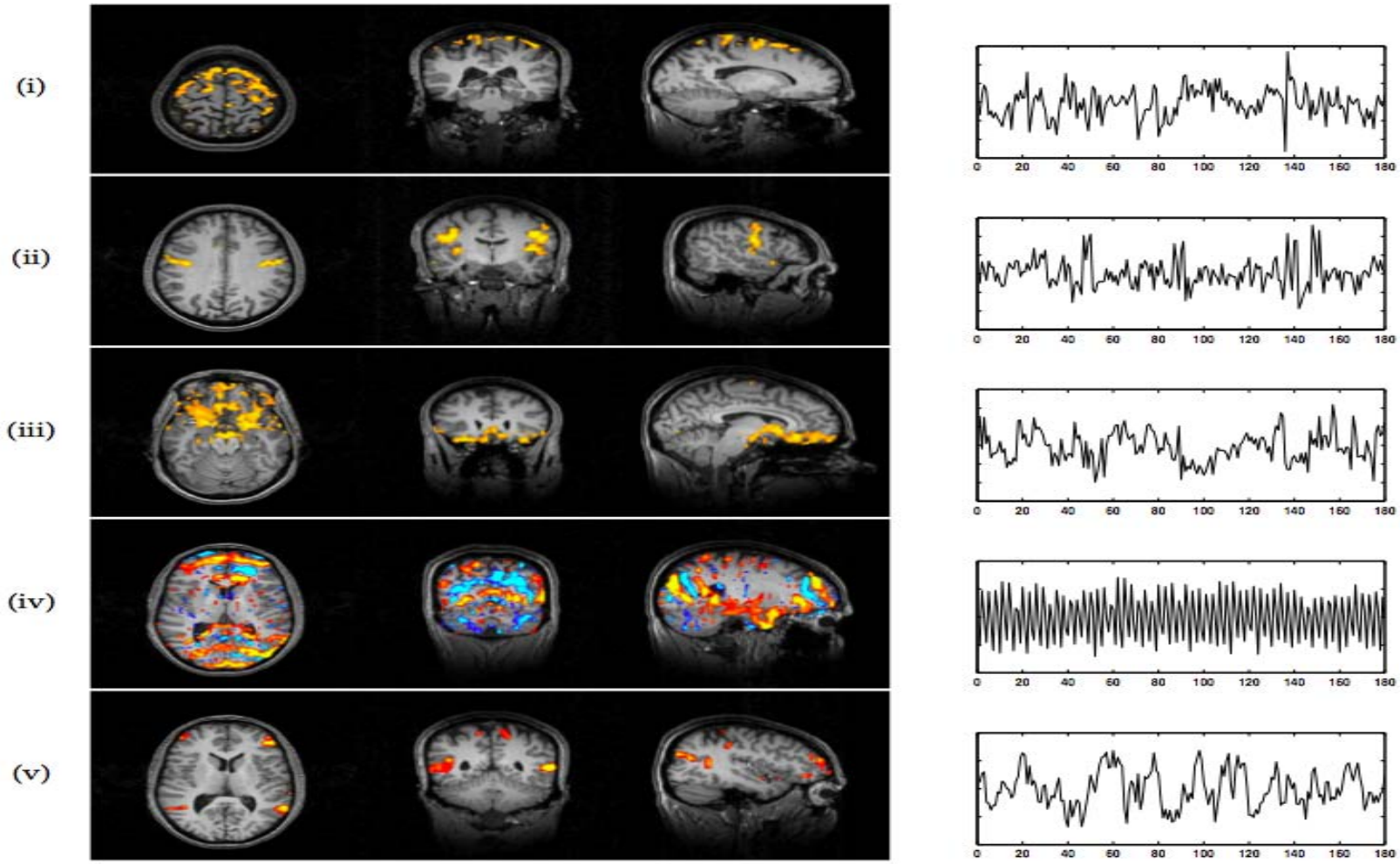
Return

Mass Spectrometry



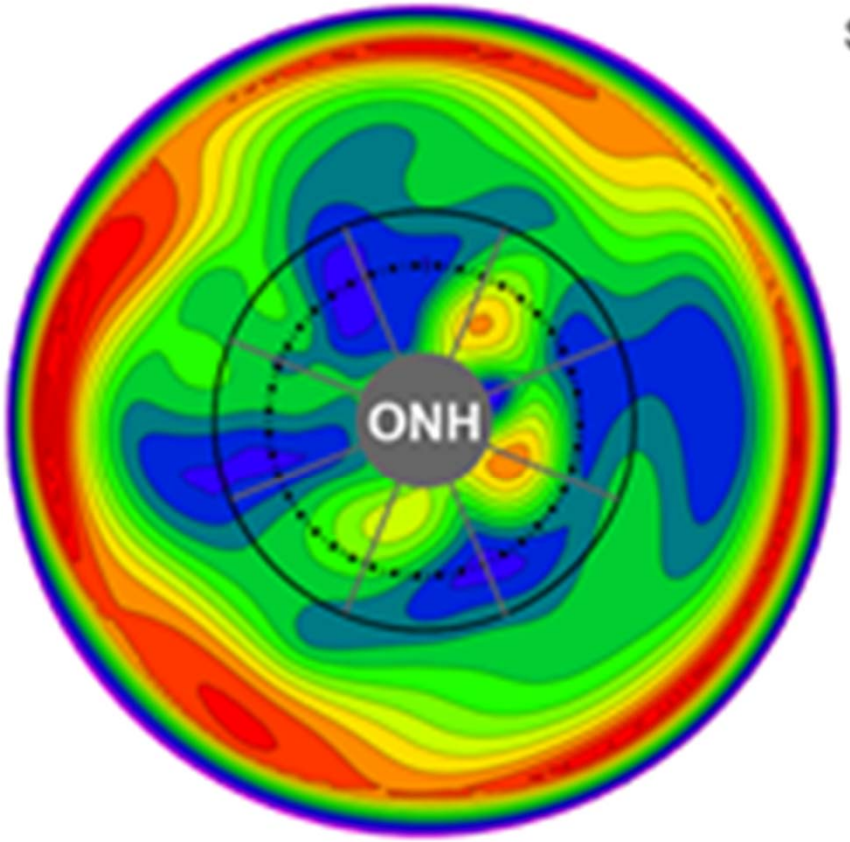
Return

fMRI



Return

Ophthalmological Data: Scleral Surface Tension

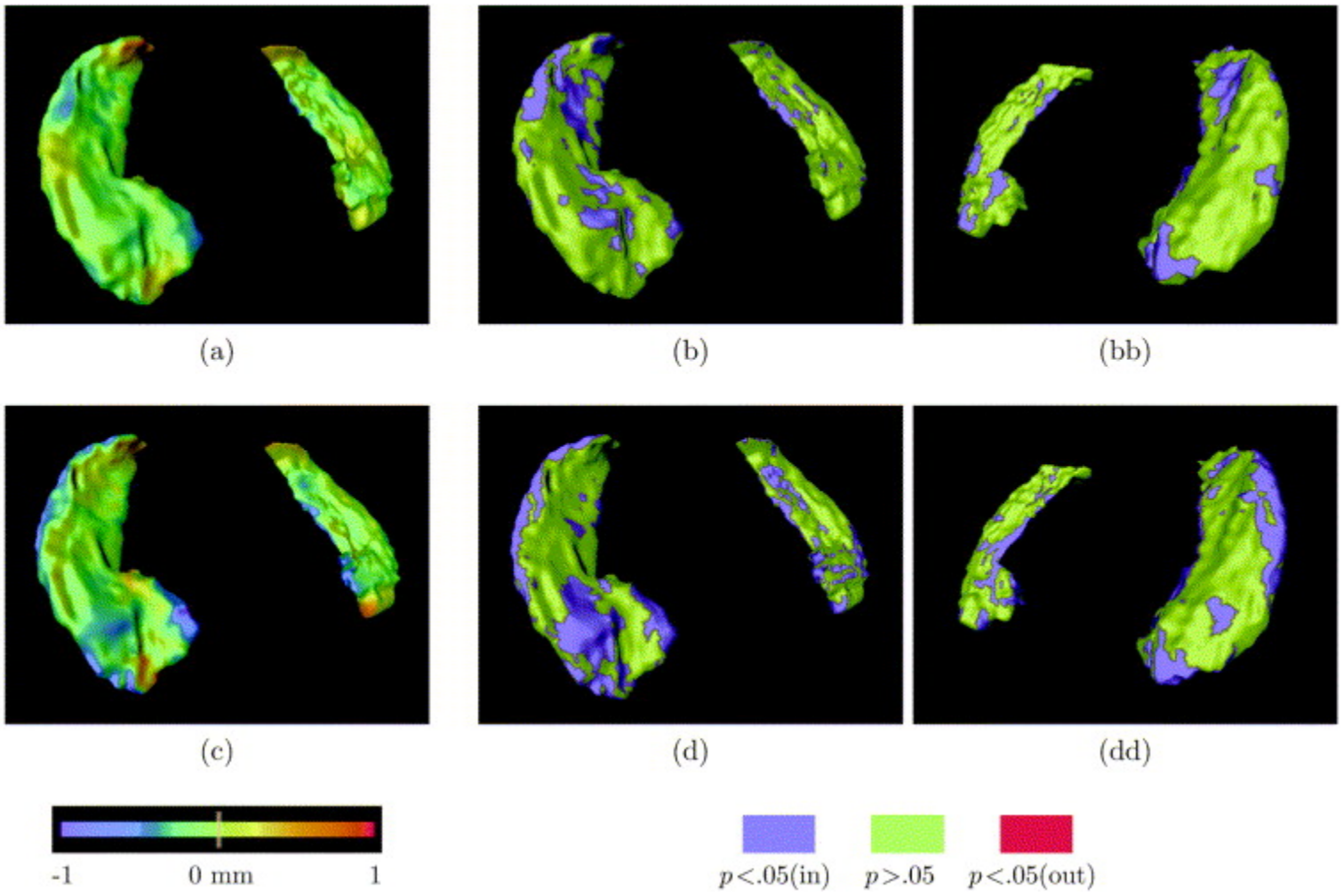


Max. Principal Strain

Return

Fazio MA, et al. (2012). Age-related changes in Human Peripapillary Scleral Stiffness. Submitted.

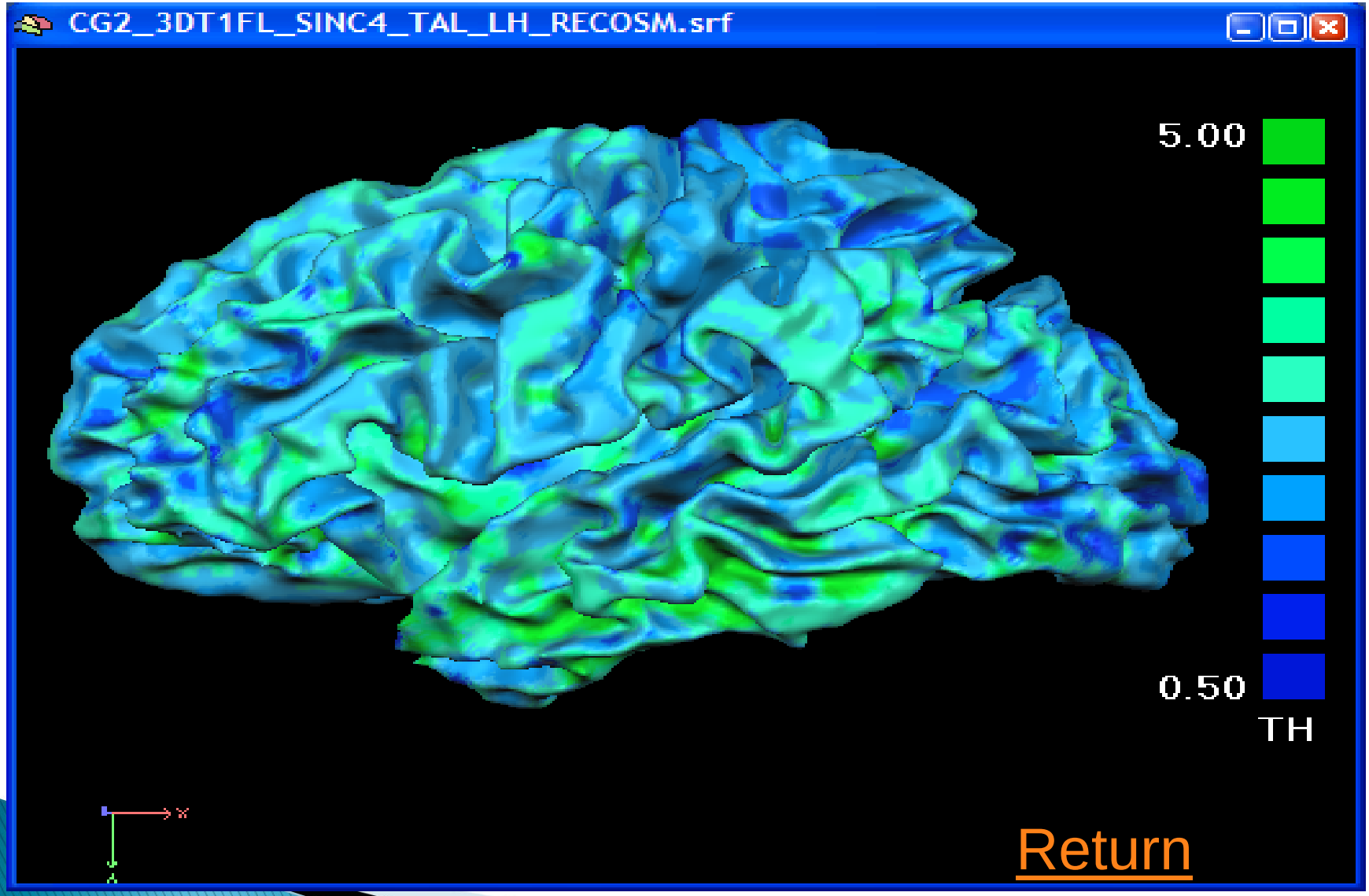
Longitudinal Shape Data



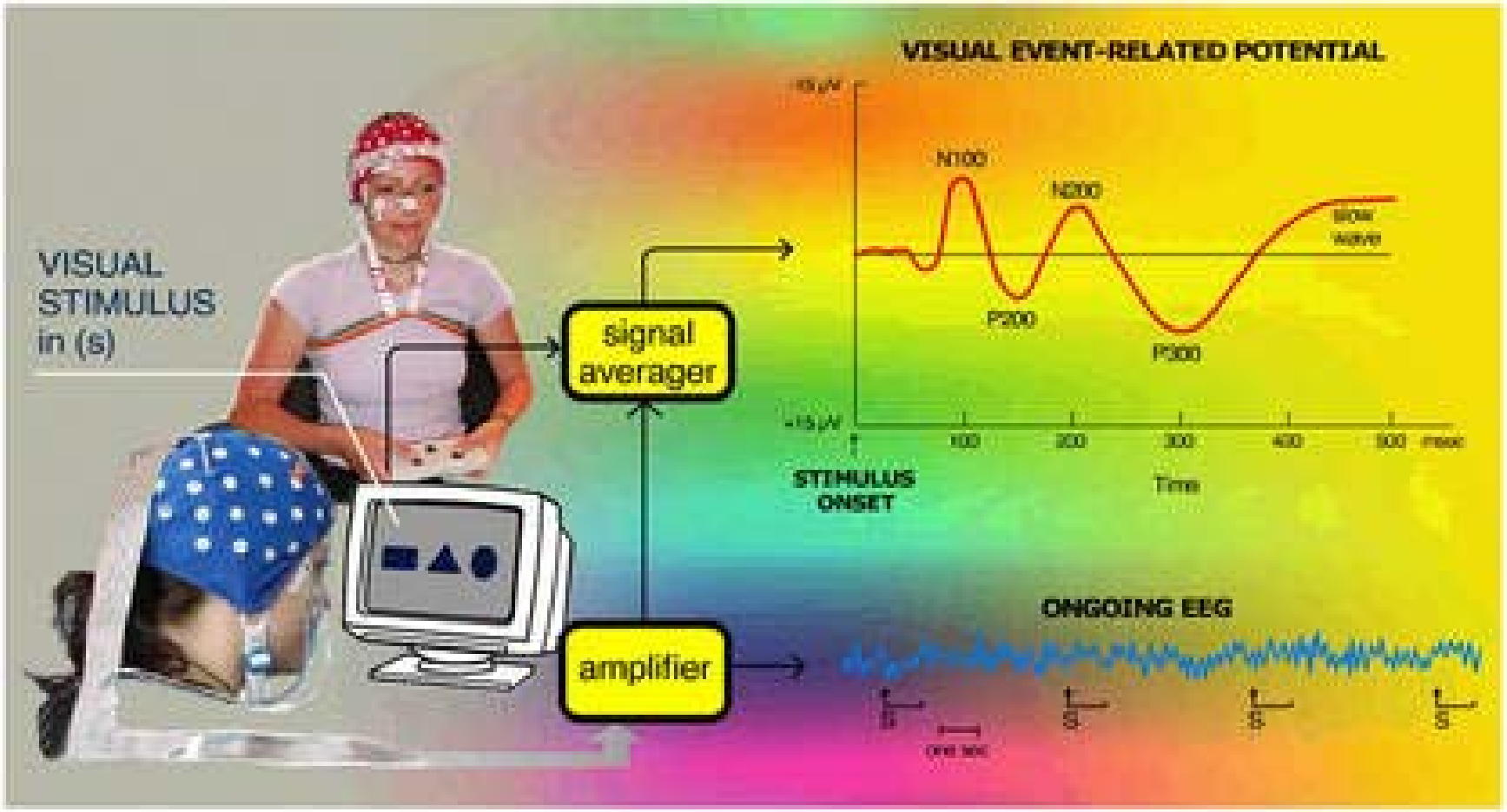
- Wei L, et al. (2003). Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy Aging. Neuroimage 20(2): 667-682.

Return

Corticol Surface Thickness



Event-related Potentials (ERP/EEG)



Return

Key Questions of Interest in Object Data Analysis

1. **Group Comparison**: regression of objects on scalar class predictors to assess which "parts" of the object differ across groups

   Object ~ Class + Covariates

2. **Group Discrimination**: classify subjects into groups based on their object data; e.g. by regressing class on object.

   Class ~ Object + Covariates

3. **Clustering**: unsupervised clustering of subjects based on their object data.

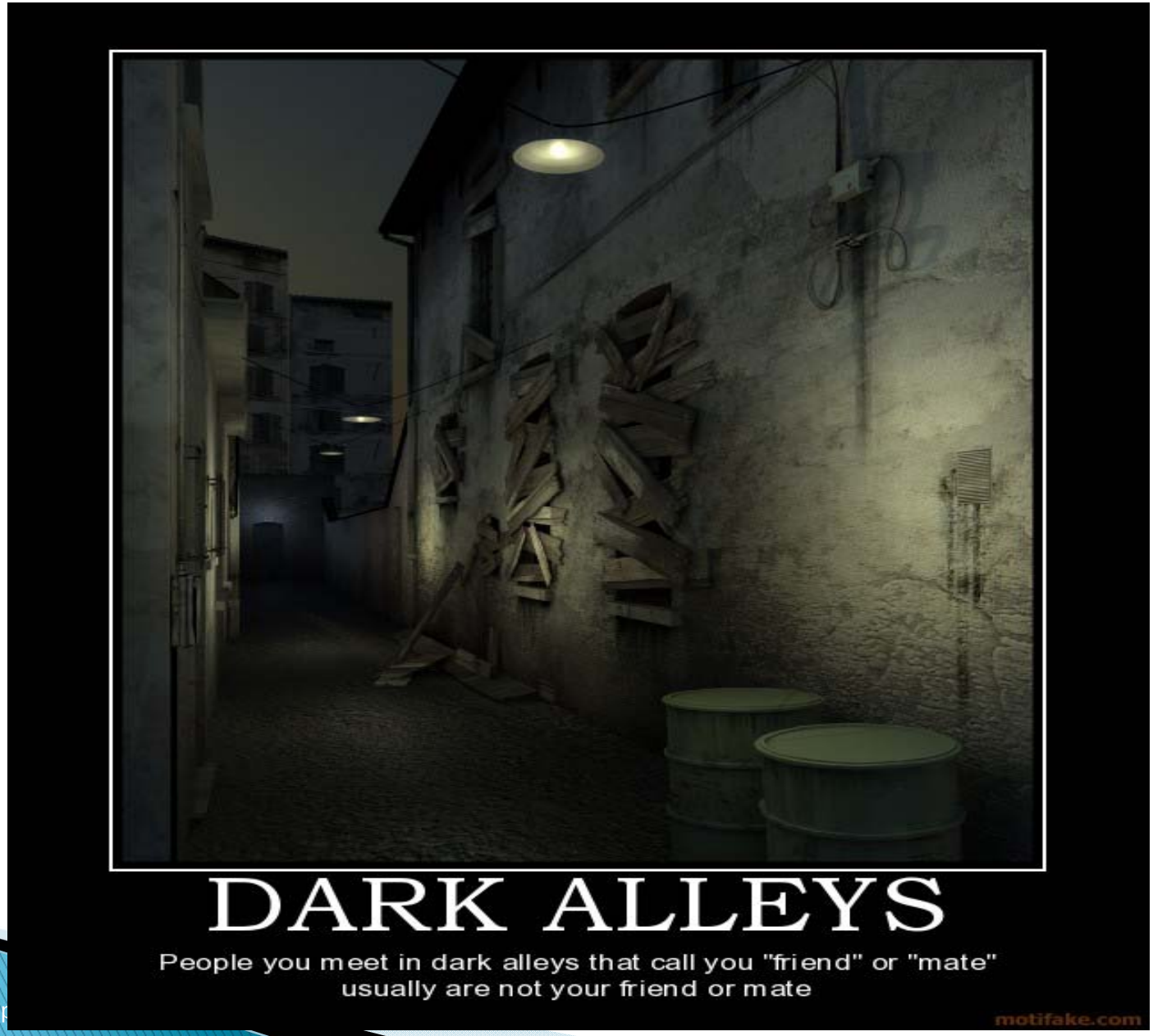
Standard Approach 1: Feature Extraction

- Compute summary statistics from object and then perform standard analyses on the summaries
- **Examples**:
  - *Accelerometers*: average daily levels, % above threshold
  - *Mass spectra*: detect peaks, then analyze by peak
  - *2dGE*: detect spots, then analyze by spot
  - *fMRI*: integrate within known brain regions (ROI)
  - *Copy number*: segments of gain/loss on individual array
- **Benefits**: reduces dimensionality, can incorporate biological information about objects, easy to use
- **Drawbacks**: loses information not in summaries

Looking under the Lamp Post

"I'm searching for my keys."

Complex Data are Scary!!!



Standard Approach 2: Element-wise Modeling

- Apply standard statistical tests on each element of the object, treating them as independent
- **Examples**:
  - *Time series*: separate analyses at each time point
  - *fMRI*: separate analyses for each voxel in the brain
  - *ERP*: separate analyses for each EEG sensor
- **Benefits**: retains all information, easy to implement
- **Drawbacks**: doesn't borrow strength across measurements within an object; ignores internal structure; inefficient; may give misleading inference

Element-wise Modeling

- "No object element is an island"
- They have neighbors, and should be allowed to share with and borrow from their neighbors

Approach 3:
Statistical Modeling of Object Data

- Increasingly, statisticians are developing innovative statistical models to account for internal structure in objects, e.g. functional data analysis (FDA).
- **Note**: Dimensionality typically precludes modeling within-object correlation in unstructured fashion
- Alternative: use **basis functions**/**frames** to parsimoniously capture the local (splines, kernels, wavelets) or global (PCA) internal structure within the objects.
- **Challenges:** Scale up to extremely large data sets
  - Provide unified inference that accounts for sources of var.
  - Handle multiple types of objects with different structure
  - Be able to model common types of between-object structure from experimental design (subsamples; nested designs; longitudinal objects)

MaTaDOR: MulTi-Domain Object Regression

General suite of methods for object data analysis

- Flexible enough for broad class of objects
- Models object~scalar, scalar~object, object~object
- Can account for various types of between-object structure induced by experimental design
- Yields unified Bayesian inference, including pointwise and joint intervals and FDR thresholds
- Automated code that scales up to EXTREMELY large data sets (up to 100s of GBs)
- Can incorporate **biological knowledge** as well as uncover **unknown structure**
- Modular approach: extendible in many ways as we keep building on the core method/code

Model 1: GLOMM
Generalized Linear Object Mixed Models

$$g\{E(X_i)\} = \sum_{a=1}^{p_Y} \int_{\mathrm{T}} Y_{ia}(t) B_a(t) dt + V_i \boldsymbol{\beta} + \sum_{h=1}^{H} Z_{ih} \mathbf{U}_h + \varepsilon_i$$

- X_i: exponential family response, g: link function
- $Y_{ia}(t)$: object predictor of type a, subject i
- t: index for elements w/in object (may be multi-dim)
- $B_a(t)$: object regression coefficient for type a
- V_i and β: vectors of scalar predictors and coeffs.
- Z_{ih} and $\mathbf{U}_h$: random predictors/coeffs at level h
- ε_i: residual error in latent space

Model 2: ORMM
Object Response Mixed Models

$$Y_i(t) = \sum_{a=1}^{p_X} X_{ia} B_a(t) + \sum_{h=1}^{H} \sum_{b=1}^{p_h} Z_{hib} U_{hb}(t) + E_i(t)$$

- $Y_i(t)$: object response for subject i for element t
- X_{ia} and $B_a(t)$: predictor and fixed effect object a
- Z_{hib} and $U_{hb}(t)$: predictor and random effect objects
- $E_i(t)$: residual for subject i at object element t
- $\mathrm{Cov}\{U_{hb}(t_1), U_{hb'}(t_2)\} = \boldsymbol{P_{(h)bb'}} \boldsymbol{Q_h(t_1,t_2)}$; $\mathrm{Cov}\{E_i(t_1), E_{i'}(t_2)\} = \boldsymbol{R_{ii'}} \boldsymbol{S(t_1,t_2)}$
- $\boldsymbol{P_{(h)}}$, $\boldsymbol{R}$: between-object covariance matrices
- $\boldsymbol{Q_h(t_1,t_2)}$, $\boldsymbol{S(t_1,t_2)}$: within-object covariance surfaces (form reflects internal structure of objects)

Multi-Domain Modeling of Object Data

- $\boldsymbol{Y}$=matrix ($N \times T$) of object data, on same T elements
- Write out basis function expansion: $Y=Y^*\phi$
  - ϕ =matrix ($T^* \times T$) of basis functions on grid of size T
  - Y^*=matrix ($N \times T^*$) of basis coefficients (T^* coefficients)
- Compute basis coefficients $Y^*=Y\phi^-$
- ϕ^-= ***transformation matrix*** (data space Y to basis space Y^*)
- ϕ = ***inverse transform matrix*** (basis space Y^* to data space Y)
- Multi-Domain Modeling Approach

1. Transform objects into alternative domain ($Y \rightarrow Y^*$)
2. Fit alternative-domain object regression models (ORMM/GLOMM)
3. Transform object coefficients back to data domain $\{B^* \rightarrow B(t)\}$
4. Perform Bayesian inference in data domain

Basis Function Modeling

- Types of Basis Functions
  - Local: splines, Fourier, wavelets, needlets (sphere)
  - Empirical: PC, fPC, sPC, gPC, IC, PLS, GLRAM
  - Biological: ROI, peak templates, pathway bases
- For many basis functions, special fast algorithms exist for computing Y^* from Y or Y from Y^*
  - E.g., wavelets O(T), Fourier O($T \log T$), PC, IC
- For many other basis functions, the transform and inverse transform matrices ϕ and ϕ^- are sparse, and only need be computed once.
- Many bases yield parsimonious representations , i.e. $T^* \ll T$, greatly reducing dimensionality while retaining most all information in data

Alternative Domain ORMM

$$Y^*_{ik} = \sum_{a=1}^{p_X} X_{ia} B^*_{ak} + \sum_{h=1}^{H} \sum_{b=1}^{p_h} Z_{hib} U^*_{hbk} + E^*_{ik}, k = 1, \ldots, T^*$$

- Y^*_{ik}: basis coefficient k for subject i
- B^*_{ak}, U^*_{hbk}, E^*_{ik} : basis space fixed, random, residual
- $\mathrm{Cov}\{U_{hbk}, U_{hb'k}\} = P^h_{bb'} q_{hk}$ $\mathrm{Cov}\{E_{ik}, E_{i'k}\} = R_{ii'} s_k$
- Computing is parallelizable and linear in T^*
- Form of $Q_h(t_1,t_2)$ and $S(t_1,t_2)$ defined by T^* dimensional manifold: $Q_h(t_1,t_2) = \phi' Q^*_h \phi$ and $S(t_1,t_2) = \phi' S^* \phi$ $Q^*_h = \mathrm{diag}_k\{q_{hk}\}$ $S^* = \mathrm{diag}_k\{s_k\}$
- Covariance of dimension T^* but flexibly capturing internal structure of object given suitably choice of basis
- Random effects U^*_{hbk} and residuals E^*_{ik} assumed Gaussian or for robust regression, heavier tailed distribution (DE)

Why Bayesian Modeling?

- Could fit ORMM without using Bayesian modeling
  - Mixed model (perhaps with penalty for regularization)
- So why do we use a Bayesian approach?
  - Our Bayesian approach does not require subjective priors
  - Automatically obtain pointwise/joint inference all model quantities
    - Posterior probabilities of effect sizes: connection to FDR
  - Unified modeling approach integrates over all variability
  - MCMC can be challenging in some high dimensional contexts, but here we have stable, automated algorithm.
  - Straightforward approach to handle measurement error and missing data. *(Morris, et al. 2006 JASA)*
  - Natural way classification.*(Zhu, Brown, Morris, 2012 Biom)*
  - Extendability: can make other distributional assumptions and sparsity priors and get improved performance (e.g, robust FMM; *Zhu, Brown, Morris 2011 JASA*)

Sparsity priors for fixed effects B^*_{ak}

- Spike_0/Gaussian, DE, Spike_0/DE, NG, NEG, NMIG, HShoe
- Performs variable selection/nonlinear shrinkage on B^*_{ak}
  - E.g. Gaussian = ridge regression, DE = Lasso
  - Induces **structured regularization** of $B_a(t)$, which is a type of smoothing within manifold defined by basis functions that should take internal structure of the object into account.
- Amount of regularization depends on set of **sparsity hyperparameters**, which can be estimated from data or given their own hyperpriors.
- This regularization should lead to improved estimation and inference on fixed effect functions $B_a(t)$

Model Fitting and Inference

- Model fit by automated MCMC (MH w/in Gibbs)
  - Parallelizable in MCMC iterations and/or basis coeffs. k
- Inverse transform ϕ used to transform posterior samples back to data space, e.g. $B_a(t)$, for inference
- Useful types of Bayesian Inference
  - Pointwise posterior credible intervals
  - Joint posterior credible intervals
  - Posterior probability (*pp*) of minimal effect size δ (posterior probability maps on object space)
  - Can find threshold on *pp* that corresponds to average Bayesian FDR of α. This approach takes both statistical and practical significance into account (Morris, et al. 2008 *Biometrics*)

Brain proteomics addiction study (Gutstein, MDACC)

- **Goal:** Find brain proteins related to cocaine addiction
- **Animal Model:**
  - Mice trained to obtain cocaine by pressing lever.
  - 21 mice, 6 short access (1hr), 7 long access (12hr), 8 ctrl
  - Mice euthanized, brain tissue harvested, microdissected
- **2d Gels**
  - Total of 53 gels from 21 rats, run on central nucleus of amygdala region (CeA) of brain
- **Analysis objective:**
  - Find proteins that are overexpressed/underexpressed in cocaine exposure group relative to controls.

Brain proteomics addiction study

- Standard analysis approach: **Spot-based**
  - Detect spots, quantify spot volumes, then analyze
  - Many flaws in existing commercial spot detection algorithms (*Gutstein and Clark 2009*)
  - *Pinnacle*: improved spot detection/quantification (*Morris, Clark, Gutstein 2008*, *Morris, et al. 2010*)
- Still limited in ability to detect and resolve all protein spots; e.g., **co-migrating proteins**
- Can we build models suitable for the scanned images themselves and flag significant regions?
- Would such an approach find more proteins and better separate effects of co-migrating proteins?
  - Morris (2011 Statistics and Its Interface): summary of work in statistics for proteomics data

Brain proteomics addiction study

- 53 gels, 21 mice, 3 groups (C/SA/LA), run in blocks
- (*Morris, et al. 2011 AOAS*): with Gutstein, Baladan.
- MODEL:

$$\underbrace{\log_2\{Y_i(t_1,t_2)\}}_{\text{response images}} = \sum_{j=0}^{2} \underbrace{X_{ij}}_{\text{group indicators}} \overbrace{B_j(t_1,t_2)}^{\text{group } j \text{ mean image}} + \sum_{k=1}^{21} \underbrace{Z_{ik}}_{\text{rat indicators}} \overbrace{U_k(t_1,t_2)}^{\text{Rat } k \text{ random effect image}} + \underbrace{E_i(t_1,t_2)}_{\text{residual error images}}$$

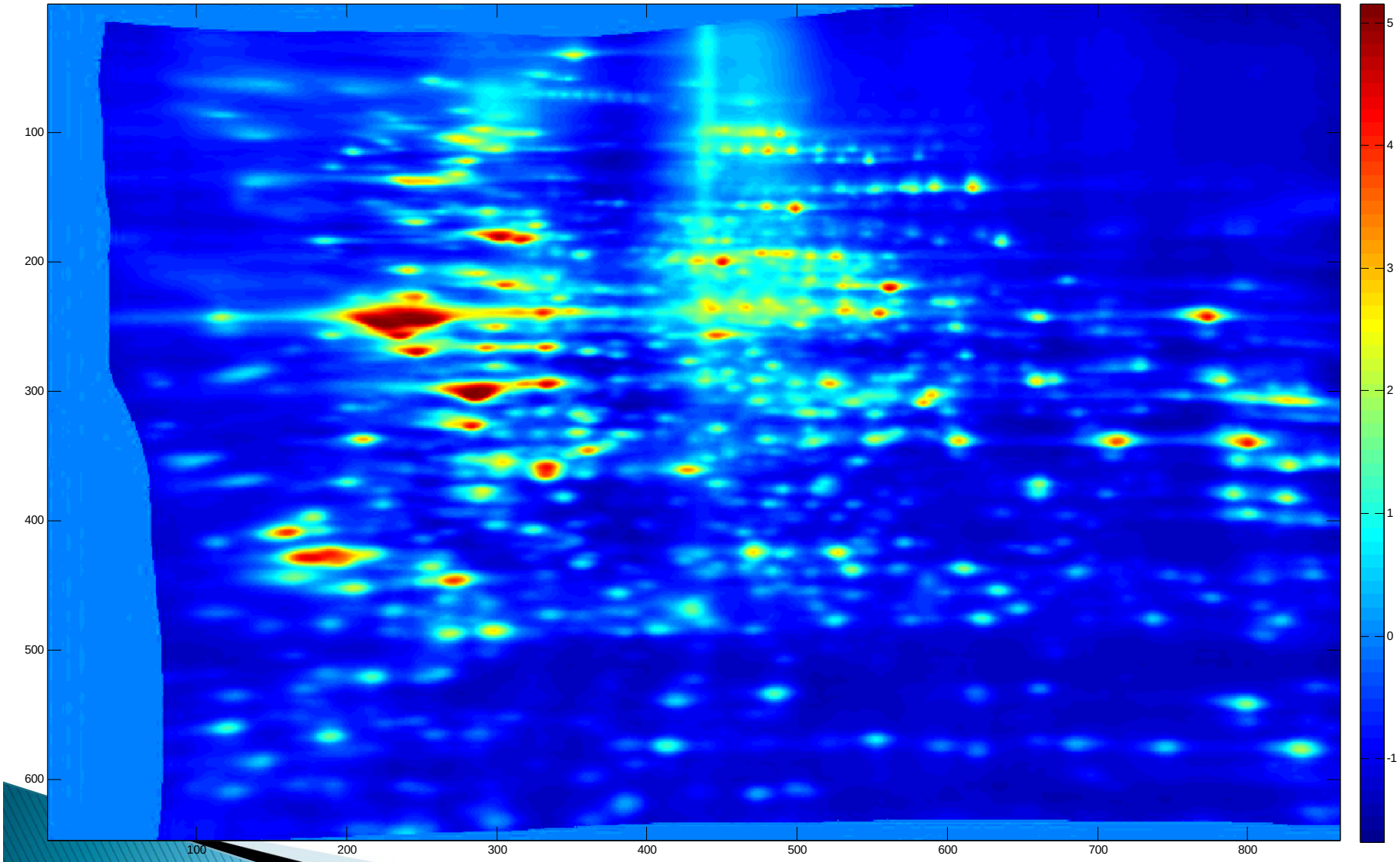
- Construct overall mean, case-control images:

  Mean Image: $M(t_1,t_2)=1/3\ \{B_0(t_1,t_2) + B_1(t_1,t_2) + B_2(t_1,t_2)\}$

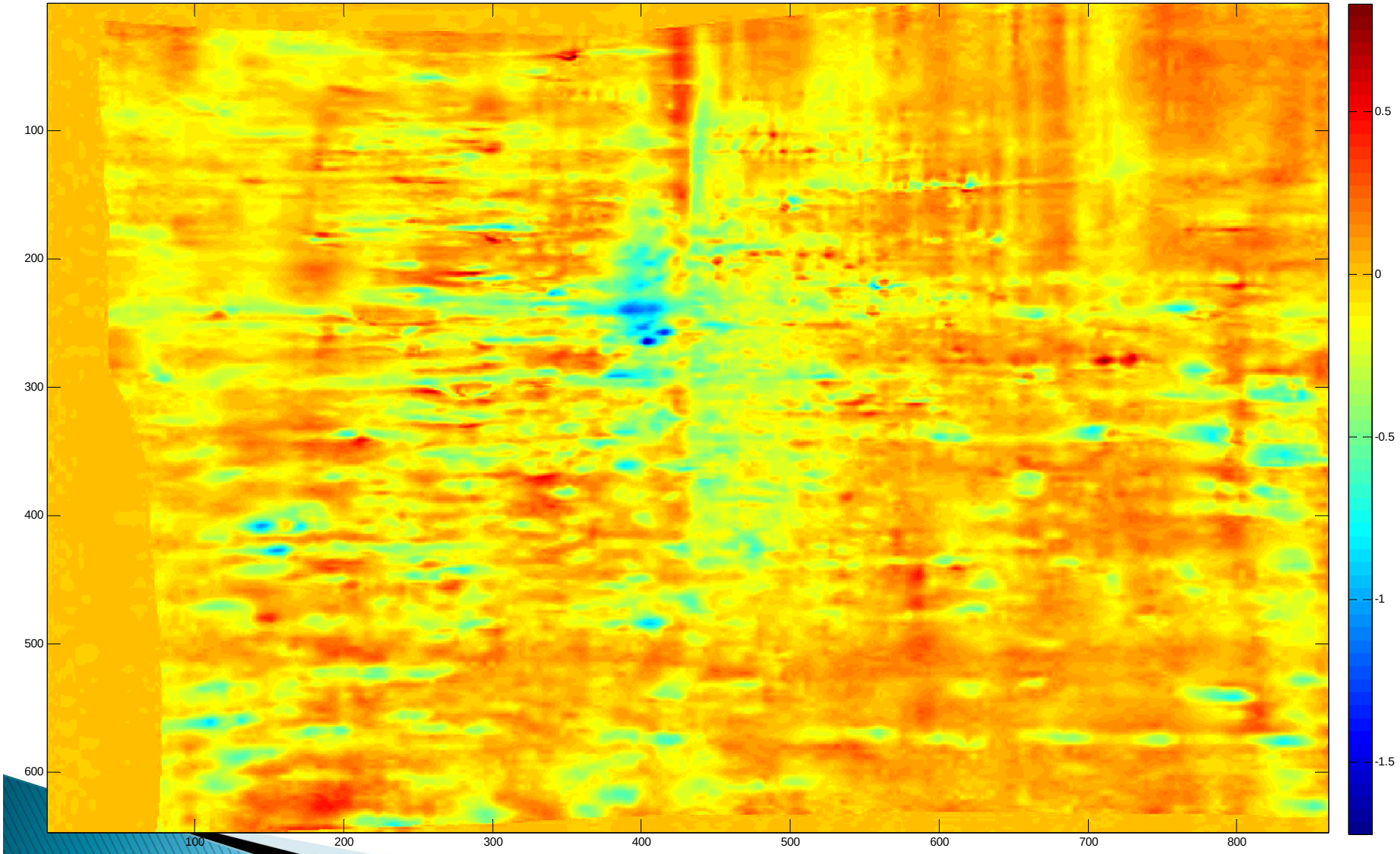
  Case-Control : $C(t_1,t_2)= B_1(t_1,t_2) - B_0(t_1,t_2)$

- **Goal**: Find regions of gel for which $C(t_1,t_2)$ is "significant" (significant evidence of at least 1.5-fold case/control ratio)

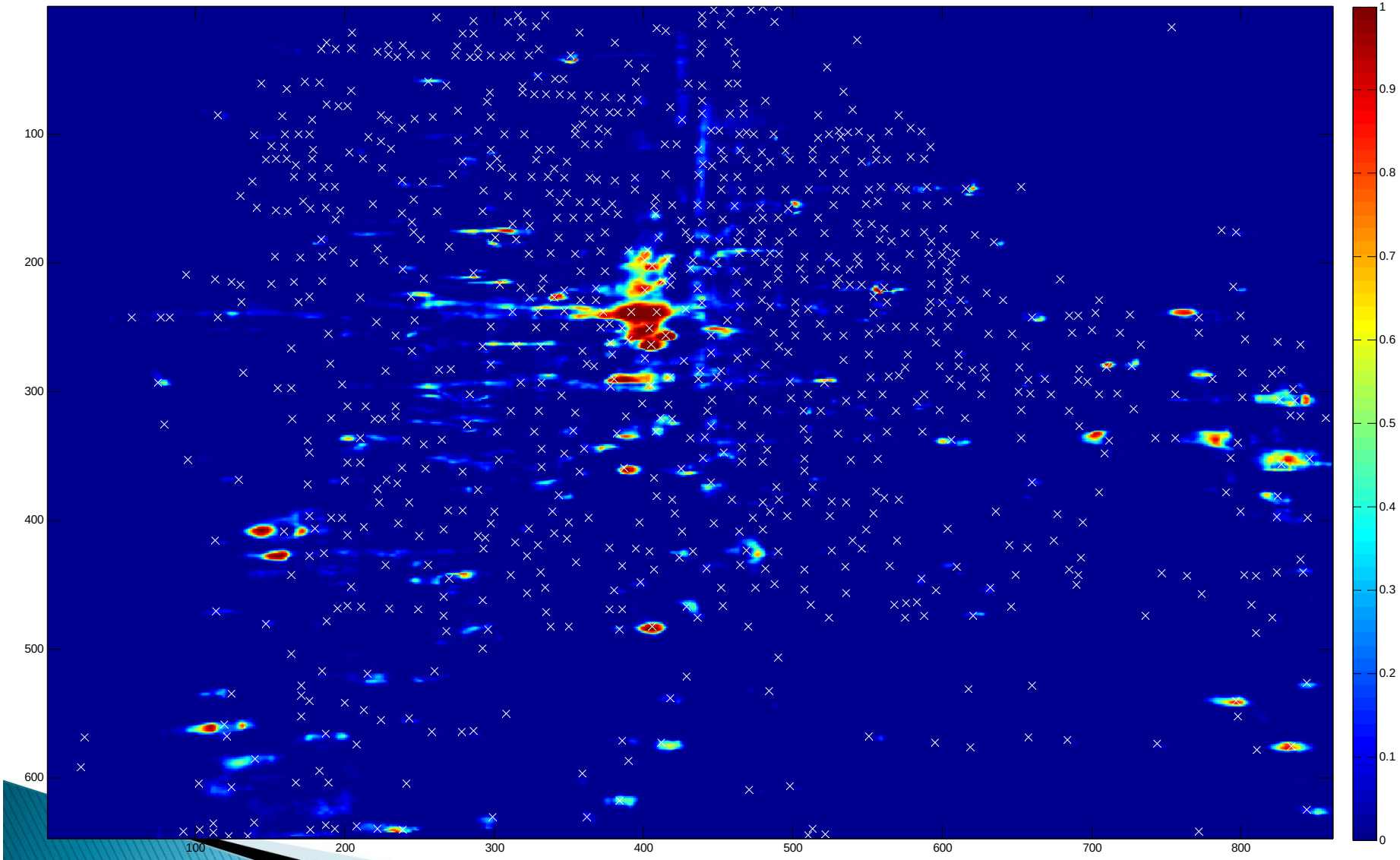
Model-Based Mean Gel : $M(t_1,t_2)$



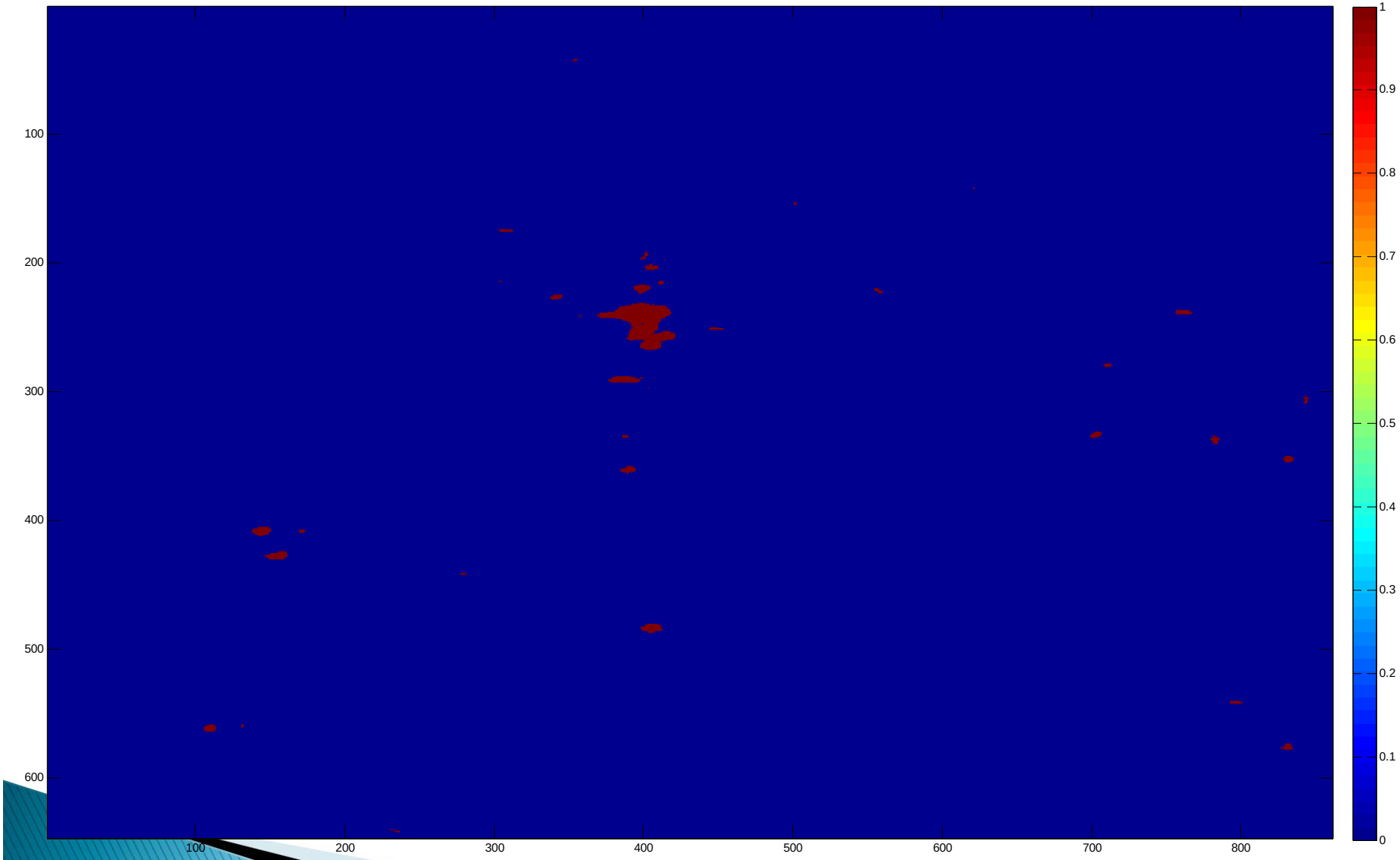
Case-Control Effect Image : $C(t_1,t_2)$



$p(t_1,t_2)$, with spots



Regions 1.5-fold different (FDR=0.10)

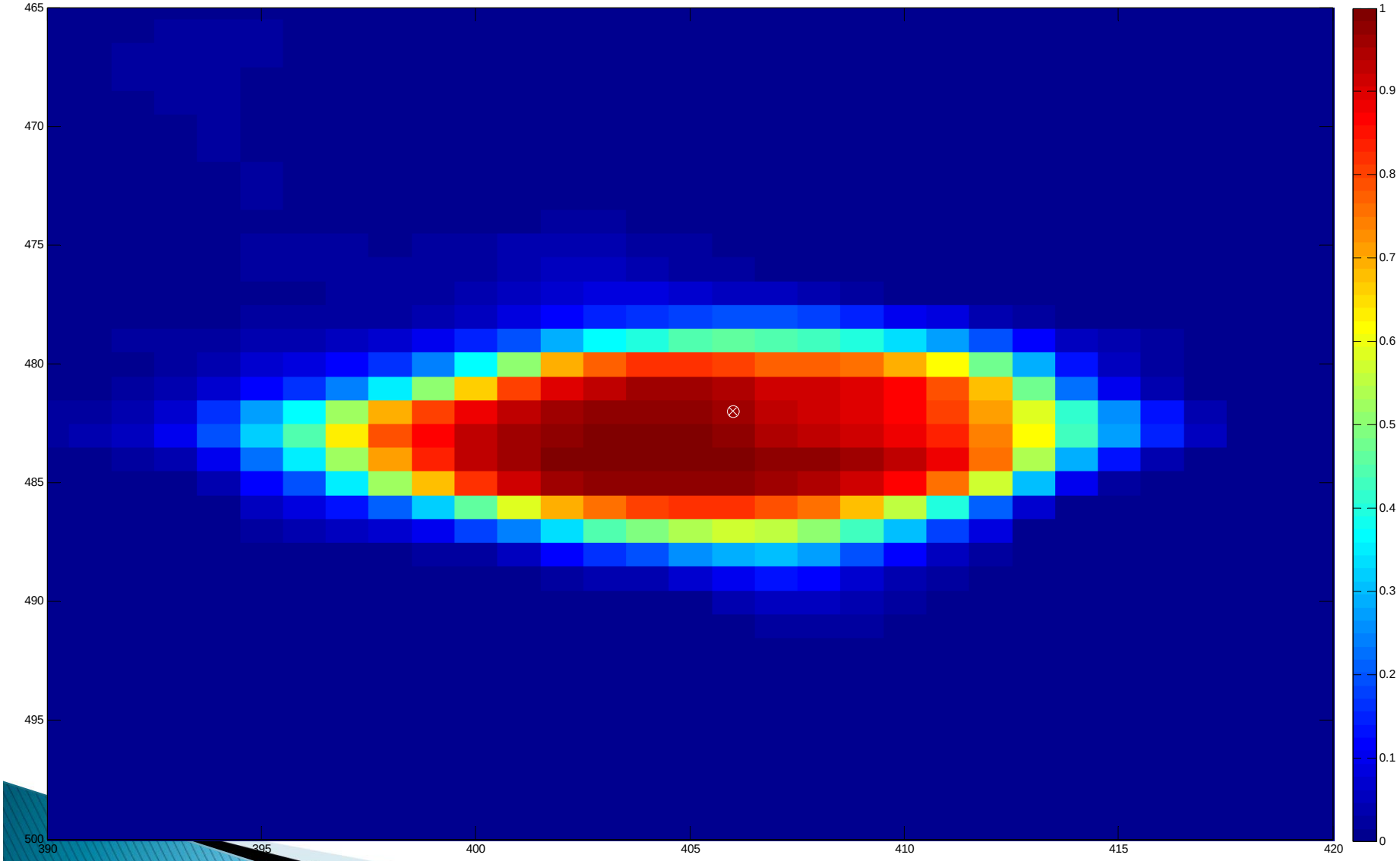


Brain proteomics addiction study

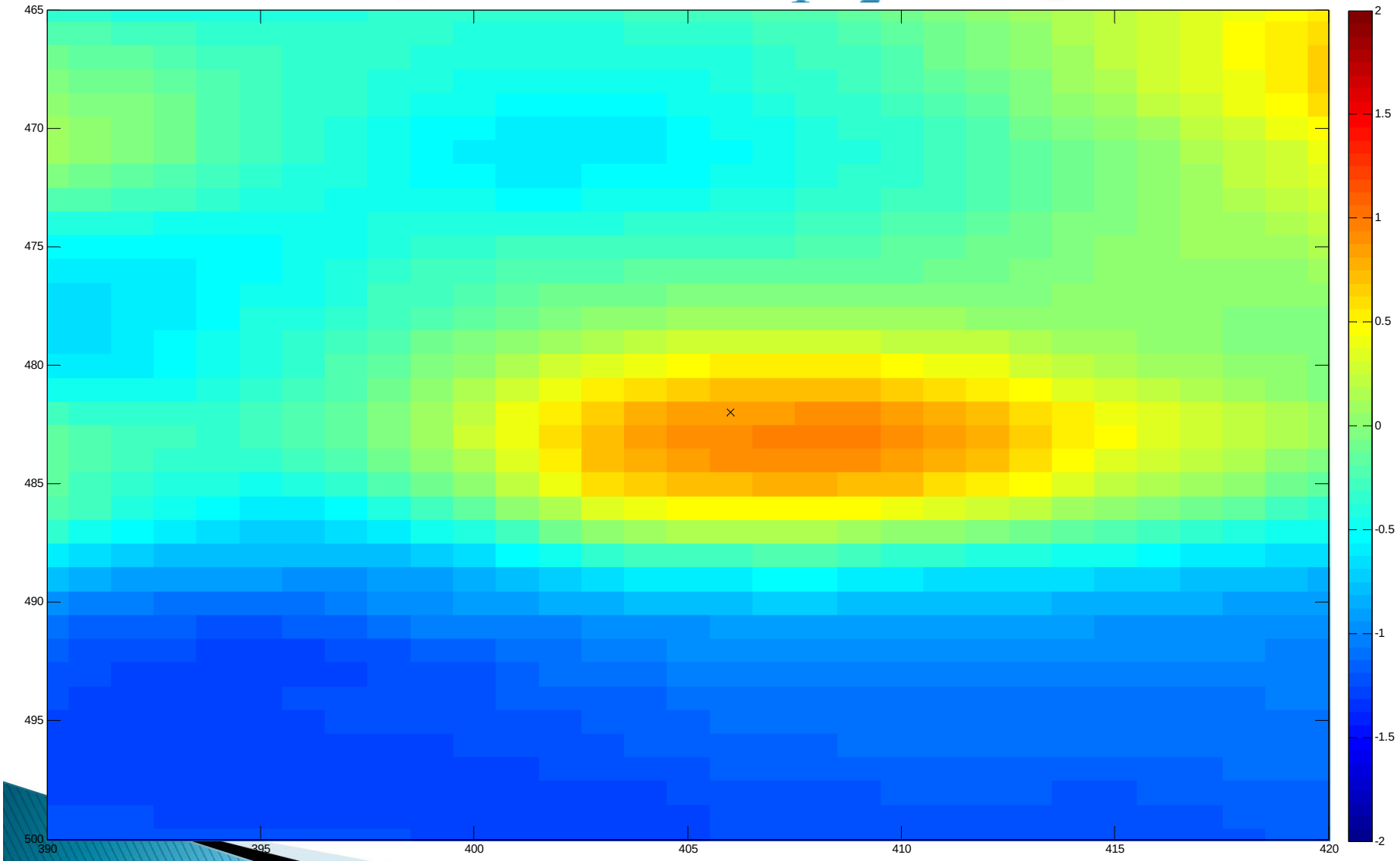
- **Results**:
  - WFMM Flagged a total of 27 contiguous regions as significant for cocaine vs. control
  - Spot-based method (*Pinnacle*) found only 17 spots.
- It appeared that WFMM was able to find essentially all results found by spot-level analysis, plus many additional results
  - Many of these were found in the tail of an abundant spot, and may correspond to co-migrating proteins
  - This suggests that perhaps there is more measurable proteomic information on 2D gels than thought, and image-based analyses can extract more of that information than spot-based approaches

Jump

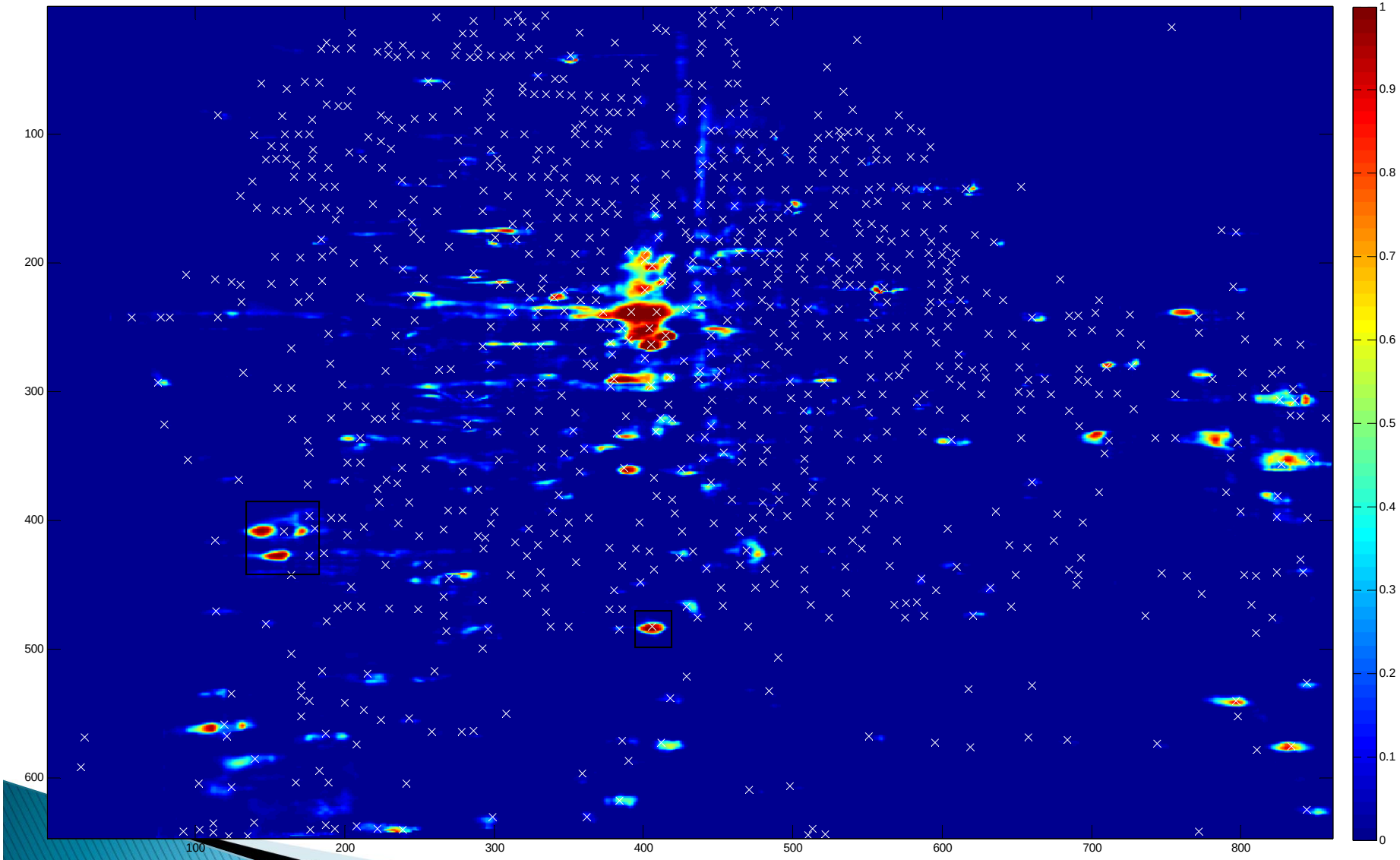
$p(t_1,t_2)$, region 1



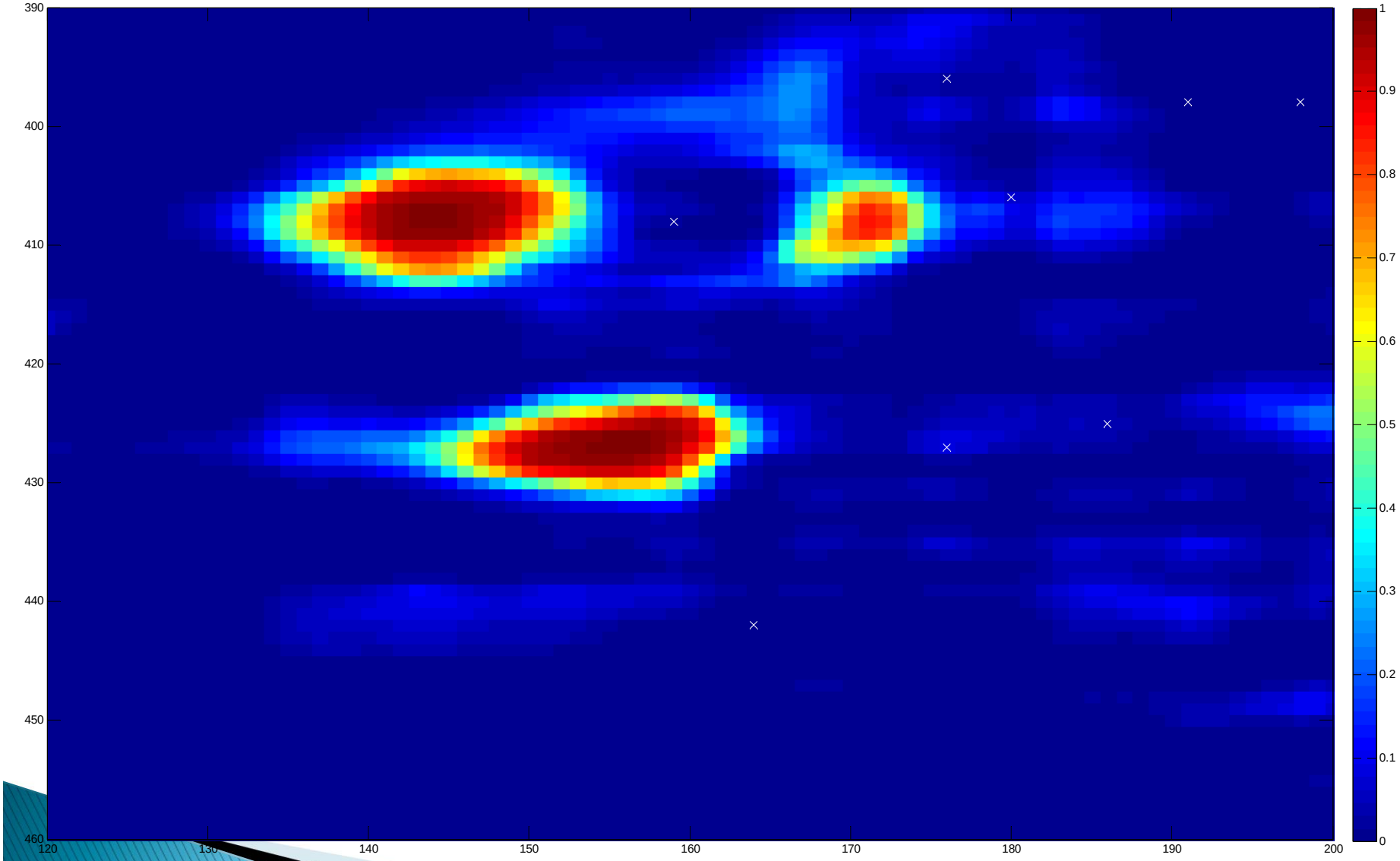
Average Gel $M(t_1,t_2)$, region 1



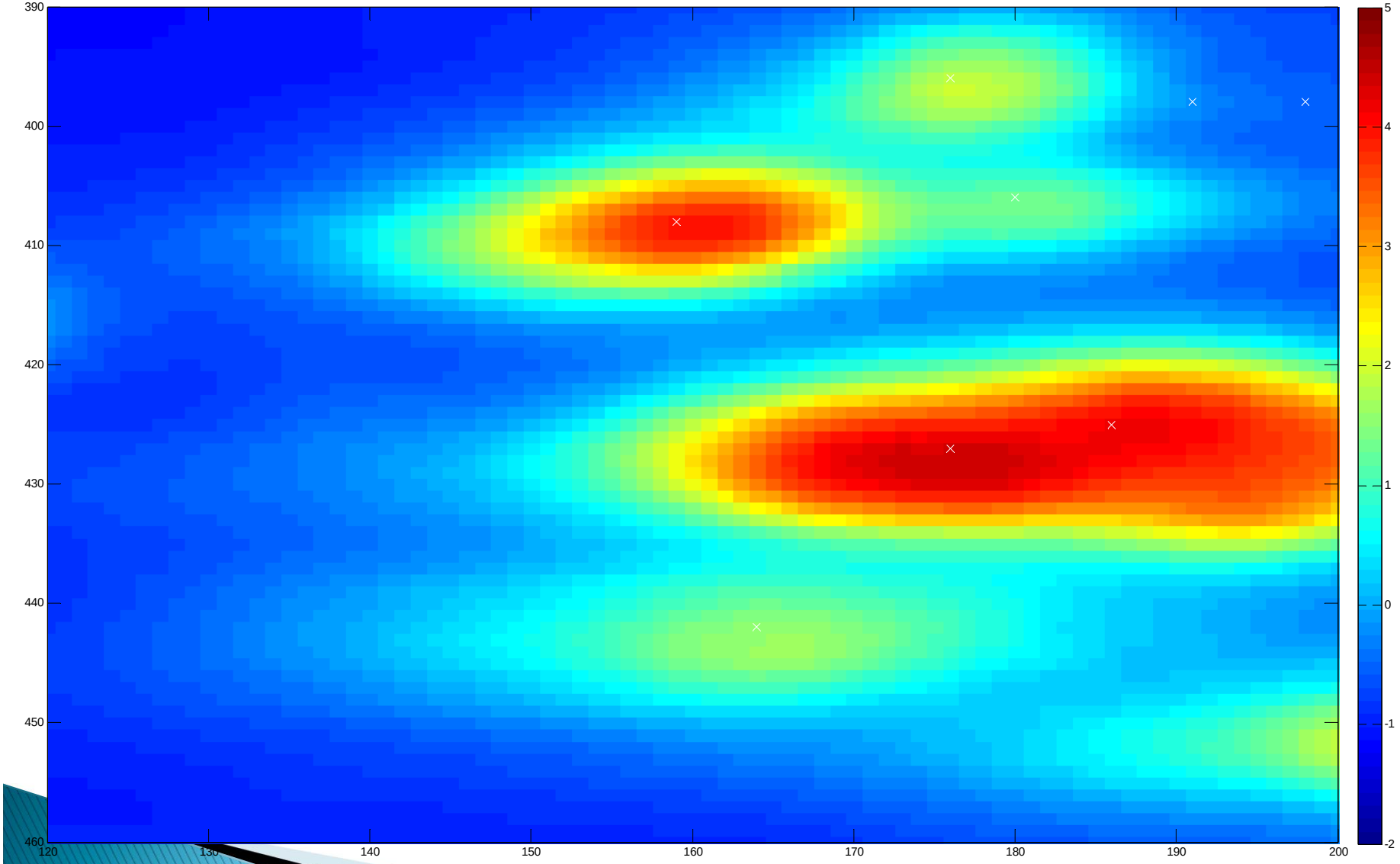
$p(t_1,t_2)$, with spots



$p(t_1,t_2)$, region 2



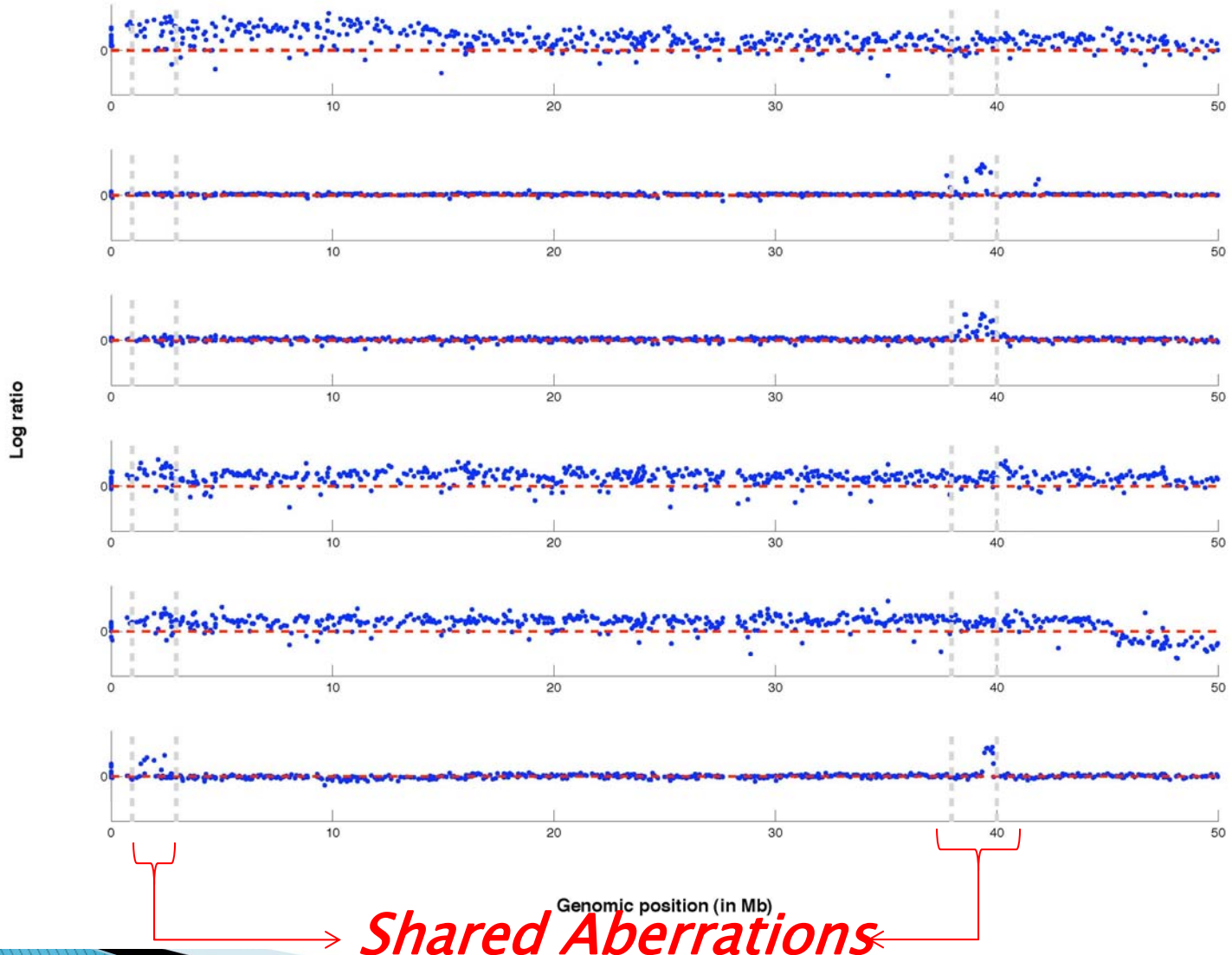
Average Gel $M(t_1,t_2)$, region 2



DNA Copy Number Cancer Genomics Study

- Cancer is characterized by various types of genomic instabilities, including in copy number
  - Discovery of prevalent copy number changes in given cancer can help better characterize a given cancer, and potentially provide markers for detection, prognosis, and prediction of response
- **Lung cancer array CGH data set** (*Coe, et al. 2006*)
  - Copy number arrays from 39 lung cancer cell lines, 4 types: small cell classical (SC) and variant (SV), non-small cell adenocarcinoma (NA) and squamous (NS)
  - **Goal**: Find ***shared aberrations*** within each of 4 lung cancer types, and assess differences between subtypes
  - ***Shared aberrations***: genomic regions with copy number changes that characterize a population

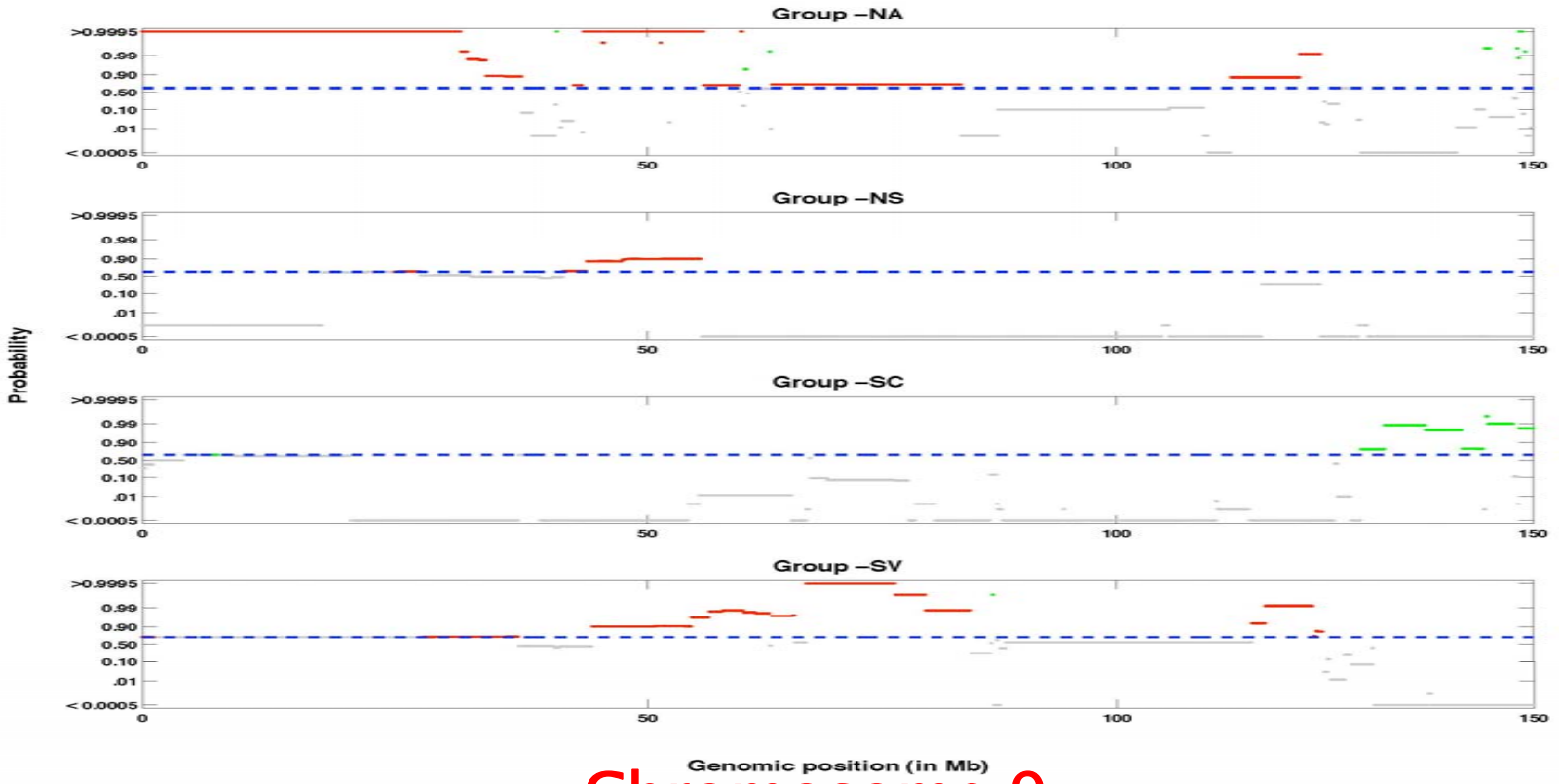
DNA Copy Number Cancer Genomics Study



DNA Copy Number Cancer Genomics Study

- Modeled using FMM with constant basis functions; prior: mixture with non-local alternatives
  - Baladandayuthapani, et al. (2010 JASA)
  - RJMCMC involving stochastically varying cut points defining regions of shared aberration
  - Unified model that borrows strength across/within arrays
  - Computed posterior probability of gain or loss for each group, and posterior probabilities of group differences
- Results:
  - Simulation study showed significant gains in sensitivity & specificity for detecting shared aberrations over alternative multi-step methods
  - Relative improvement was greater as:
    - Number of arrays in group increased
    - Noise level increased

DNA Copy Number Cancer Genomics Study



Chromosome 9

- Found 34 genes in flagged regions known to be related to lung cancer, many more than found in original multi-step analysis.

Hybrid Basis Functions

- Each type of basis has strengths/weaknesses
  - Empirical: adaptive, global, may be inconsistent
  - Local: flexible for local structure, not global
  - Biological: based on science, may lose information
- Hybrid basis functions can be constructed that combine strengths and mitigate weaknesses of different bases
  - Sequential hybrids: Fit one type of basis, take null space of projection, then apply another basis transform to the null space
  - Composite hybrids: Fit one type of basis, apply second basis transform to basis coefficients from first transform
  - E.g., fMRI brain volumes: ROI+wavelets; PC(ROI)+wavelets
- Hybrid basis functions can be used on genomic data to simultaneously perform pathway/gene/exon level analyses while accounting for local, functional, and interactive structure in genome

Works with existing GLOMM and ORMM code

Nonparametric Additive Terms

- All regression terms in ORMM (and GLOMM) assume linear relationship between object $Y_i(t)$ and scalar X_{ia}

$$X_{ia}\, B_a(t)$$

- This linearity assumption can easily be relaxed to allow additive nonparametric relationships between $Y_i(t)$ & X_{ia}

$$\boldsymbol{f_a(x,t)}$$

- This is done in straightforward fashion using the existing ORMM code by specifying a design matrix $\boldsymbol{X}$ and specific sparsity priors on $\boldsymbol{B_a}$ that correspond to O'Sullivan splines (smoothing splines are special case).
- We can do any desired Bayesian inference on $\boldsymbol{f_a(x,t)}$
- A similar approach can be used with object predictor (GLOMM) model

Object-on-Object Regression

- The ORMM can also be straightforwardly extended to regress object type 1 $Y_i(t)$ on object type 2 $X_{ia}(s)$

$$Y_i(t) = \int X_{ia}(s)\, B_a(s,t)\, ds$$

- The multi-domain modeling approach is applied by transforming both $Y_i(t)$ and $X_{ia}(s)$ using respective bases and then fitting the alternative domain ORMM.
- Again, existing code can be used for the model fitting
- This approach allows us to investigate the relationships between different types of objects on same subject, e.g. *fMRI* and *ERP* data or different types of *genomics* data.
- Many other extensions of this framework are possible, e.g. to model complex multi-way object data in unprecedentedly flexible and efficient ways

ORMM Software: Current state

- Freely available standalone executable

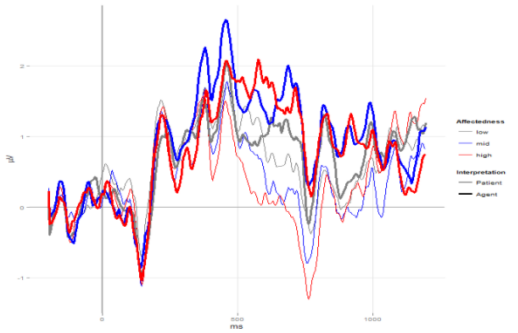
  https://biostatistics.mdanderson.org/SoftwareDownload

  - *Herrick and Morris (2006):* paper on computational issues
  - Automated: Can just specify $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{Z}$ and method will run if happy with default choices of basis, levels, priors
  - Produces posterior samples for all model parameters, plus standard summary statistics, including posterior means, variances, quantiles, probabilities of effect sizes
  - Can be used to flag regions of object related to outcomes of interest with effect size δ and FDR α
  - R wrapper for code, with plotting functions under development
  - Wavelet bases built in, can input trans. data Y^*; others to be added
- Our methods have been used for various object data types
  - Our analyses: colon carcinogenesis, accelerometer, MS, 2DGE, sonic data, copy number
  - Outside researchers: fMRI, ERP, tiling arrays, forestry data, ophthalmology data

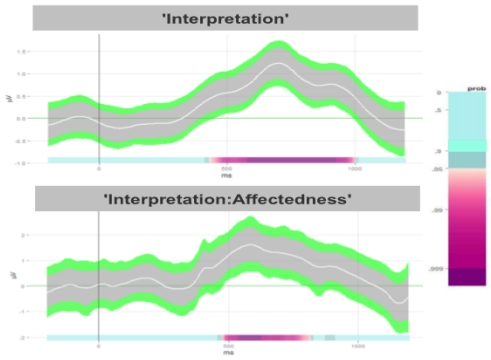
wrapfmm R package

wrapfmm

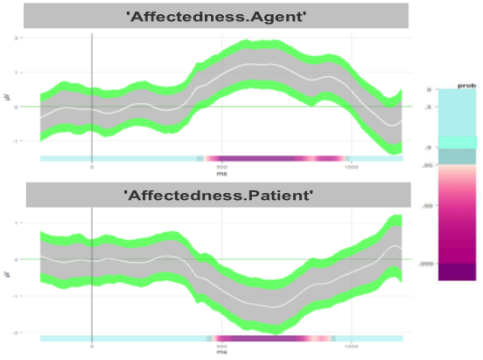
Data @ Fz split by 'Interpretation' levels & 3 'Affectedness' levels



Posterior fixed effect curves @ Fz (95% pointwise & simultaneous credible intervals)



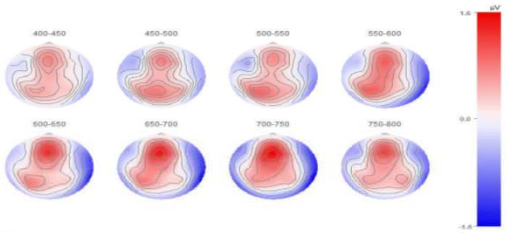
Simple effects of 'Affectedness' by 'Interpretation' (95% pointwise & simultaneous credible intervals)



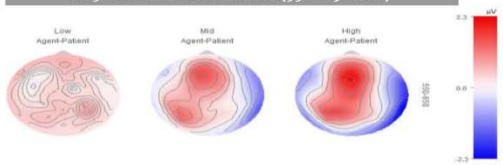
stepmom

Data plots

Difference maps for 'Interpretation' levels (400-800 ms)

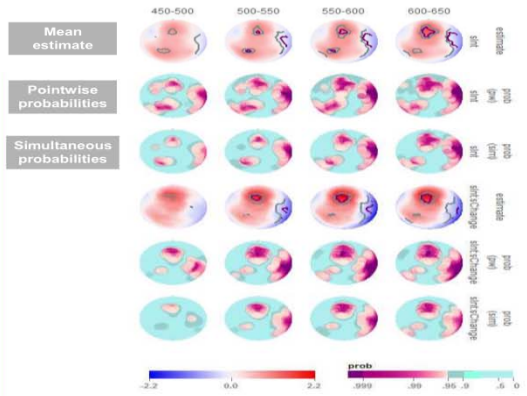


Difference maps for 'Interpretation' factor at 3 'Affectedness' levels (550-650 ms)



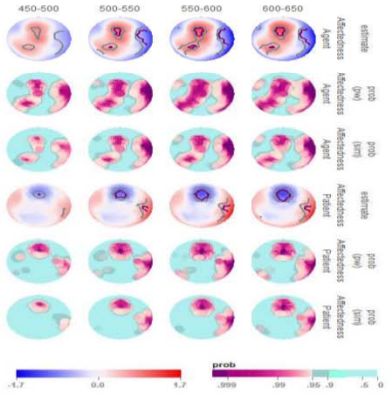
Model summaries

Model maps for 'Interpretation' & 'Interpretation:Affectedness' fixed effects (450-650 ms)



Custom posteriors

Model maps for simple effects of 'Affectedness' by 'Interpretation' levels (450-650 ms)



Conclusion

- Biomedical research experiencing an explosion of complex, high-dimensional data.
- **Object data**: general term encompassing many of these types of structured data.
- **MaTaDOR**: suite of object regression methods using multi-domain modeling approach
  - Handles object responses and/or predictors
  - Can capture *between-object structure* induced by design
  - Applies to a *broad class of object data*
  - Various internal structure captured by basis functions, local, empirical, biological, and hybrid basis functions
  - *Automated, parallelizable code*, linear in T^* (# of bases), and yields various types of unified Bayesian inference

    Framework *modular*, extendible in many ways.

The alley is much less scary in the light!

Key References

1. Baladandayuthapani V, Ji Y, Talluri R, Nieto-Barajas LE, and Morris JS (2010). Bayesian Random Segmentation Models to Identify Shared Copy Number Aberrations for Array CGH Data. *JASA*, 105(492): 1358-1375.
2. Clark BN and Gutstein HB (2008). The myth of automated, high-throughput two-dimensional gel analysis. *Proteomics*, 8(6): 1197-1203.
3. Coe, B.P., Lockwood, W. W., Girard, L., Charil, R., MacAulay, C., Lam S., Gazdar, A. F., Minna,J. D., and Lam W. L. (2006), Differential disruption of cell cycle pathways in small cell and non-small cell lung cancer, *British Journal of Cancer, 94, 1927-1935.*
4. Costafreda SG, Barker GJ, and Brammer MJ (2009). Bayesian wavelet-based analysis of functional magnetic resonance time series. *Magnetic Resonance Imaging*, 27, 460-469.
5. Davidson DJ (2009). Functional Mixed-Effect Models for Electrophysiological Responses. *Neurophysiology*, 41(1), 79-87.
6. Herrick RC, Morris JS (2006). Wavelet-based functional mixed model analysis: Computational Considerations. *Proceedings, Joint Statistical Meetings, ASA Section on Statistical Computing*, 2051-2053.
7. Morris JS and Carroll RJ (2006). Wavelet-Based Functional Mixed Models. *Journal of the Royal Statistical Society, Series B*, 68(2): 179-199.
8. Morris JS, Arroyo C, Coull B, Ryan LM, Herrick R, and Gortmaker SL (2006). Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: A case study. *JASA* 101: 1352-64.

Key References

13. Morris JS, Brown PJ, Herrick RC, Baggerly KA, and Coombes KR (2008). Bayesian Analysis of Mass Spectrometry Data using Wavelet Based Functional Mixed Models. *Biometrics*, 12, 479-489.
14. Morris JS, Baladandauthapani V, Herrick RC, Sanna PP, and Gutstein HG (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomic data. *Annals of Applied Statistics*, 5(2A), 894-923.
15. Morris JS, Clark BN and Gutstein HB (2008). Pinnacle: A Fast, Automatic Method for Detecting and Quantifying Protein Spots in 2-Dimensional Gel Electrophoresis Data. *Bioinformatics,* 24, 529-536.
16. Morris JS, Clark BN, Wei W, and Gutstein HB (2010): Evaluating the performance of new approaches to spot quantification and differential expression in 2-dimensional gel electrophoresis studies. *Journal of Proteome Research*, 9(1): 595-604.
17. Morris JS (2012) : Statistical Methods for Proteomic Biomarker Discovery using Feature Extraction or Functional Data Analysis Approaches. *Statistics and its Interface*,.
18. Subedi N and Sharma M (2011): Applying wavelet-based functional approach in modeling tree taper. *Annals of Forest Science*, published online 02 August 2011.
19. Zhu H, Brown PJ, and Morris JS (2011): Robust, Adaptive Functional Regression in Functional Mixed Model Framework. *JASA*, 106(495): 1167-1179.
20. Zhu H, Brown PJ, and Morris JS (2012): Robust Classification of Functional and Quantitative Image Data Using Functional Mixed Models. *Biometrics*.

A number of papers describing both feature extraction and functional mixed model methods, plus papers giving overviews of proteomics and proteomic data analysis are available on my website (http://works.bepress.com/jeffrey_s_morris)

Code for fitting Bayesian multi-domain FMM is also available on the web
http://biostatistics.mdanderson.org/SoftwareDownload/

Code will continue to be updated to make it more user friendly in the future, and to contain features from recent publications.

Acknowledgements

Statistical Collaborators

Raymond J. Carroll (Texas A&M)
Phil Brown (University of Kent)
Hongxiao Zhu (Duke University)
Veera Baladandayuthapani (MDACC)
Louise Ryan (CSIRO, Australia)
Brent Coull (Harvard University)
Betty Malloy (American University)
Marina Vannucci (Rice University)
Cassandra Arroyo (Georgia Southern)
Josue Martinez (Texas A&M University)
Naisyin Wang (University of Michigan)

Computing Collaborators

Richard C. Herrick (MDACC)
Andrew Dowsey (Imperial College, London)
Guang-Zhong Yang (Imperial College)
Philip Rausch (Humbodt University, Berlin)

Biomedical Collaborators

Joanne Lupton *Steve Gortmaker*
Rob Chapkin *Angie Cradock*
Meeyoung Hong *Paul Cinciripini*
Nancy Turner *Francesco Versace*
Howard Gutstein *Brittan Clark*
J. Crawford Downs *Massimo Fazio*
Werner Sommer *Scott Kopetz*

Grant Support

NCI: CA-107304, CA-160736, CA-57030, CA-48061, ES012044, ES000002, TCGA project
NIAA: AA-016157
NIEHS: P30-ES09106
NICHD: HD-30780
NEI: EY-18926, EY-19333
SAMSI Program on Object Data Analysis