Bayesian Wavelet-Based Mixed Models for Functional Data

Jeffrey S. Morris

Department of Biostatistics

The University of Texas
MD Anderson Cancer Center

Houston, Texas

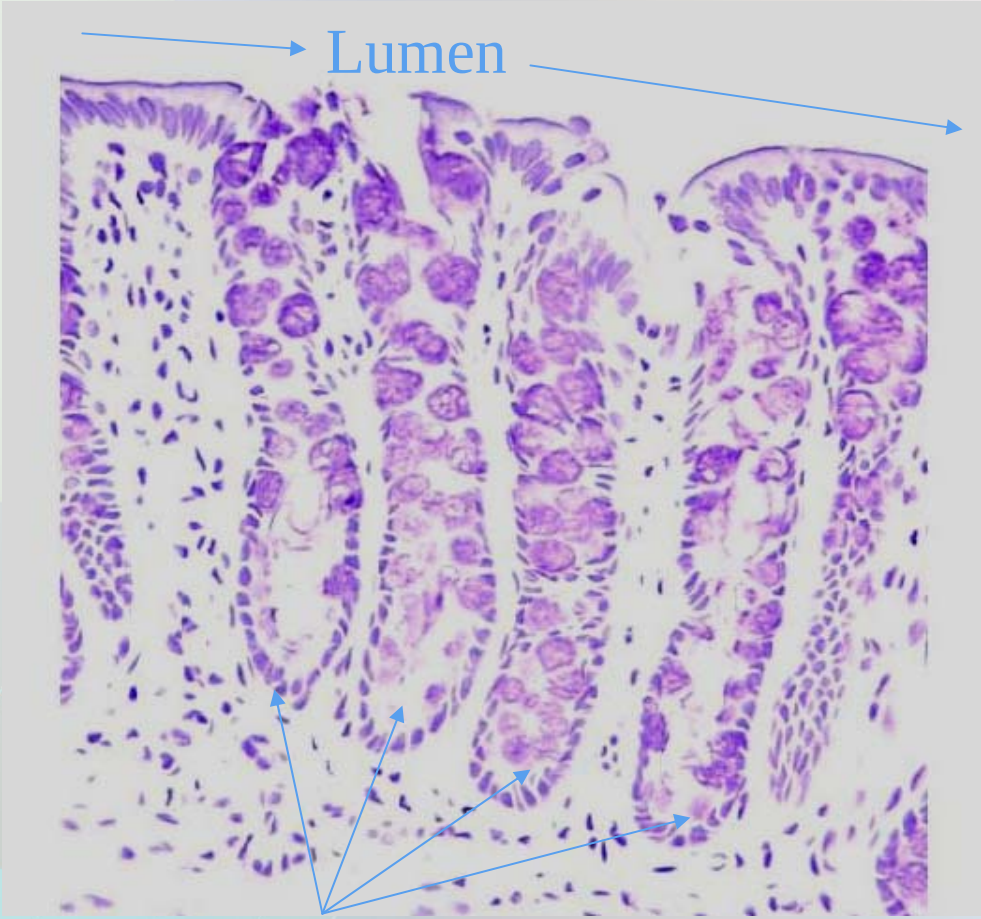
Functional Data

- **Functional Data:**
  - Ideal units of observation: **curves**
  - Observed data: **curves sampled on fine grid**
- Increasingly encountered in practice with new technologies taking automated measurements

Functional Data

- **Functional Data:**
  - Ideal units of observation: **curves**
  - Observed data: **curves sampled on fine grid**
- Increasingly encountered in practice with new technologies taking automated measurements
- **Challenge of Functional Data Analysis** (Ramsay and Silverman 1997): Must simultaneously consider **regularization** and **replication**
  - ***Regularization:*** take advantage of functional structure to borrow strength from adjacent observations within curve
  - ***Replication:*** combining information across sample curves to make inferences on population from which they came

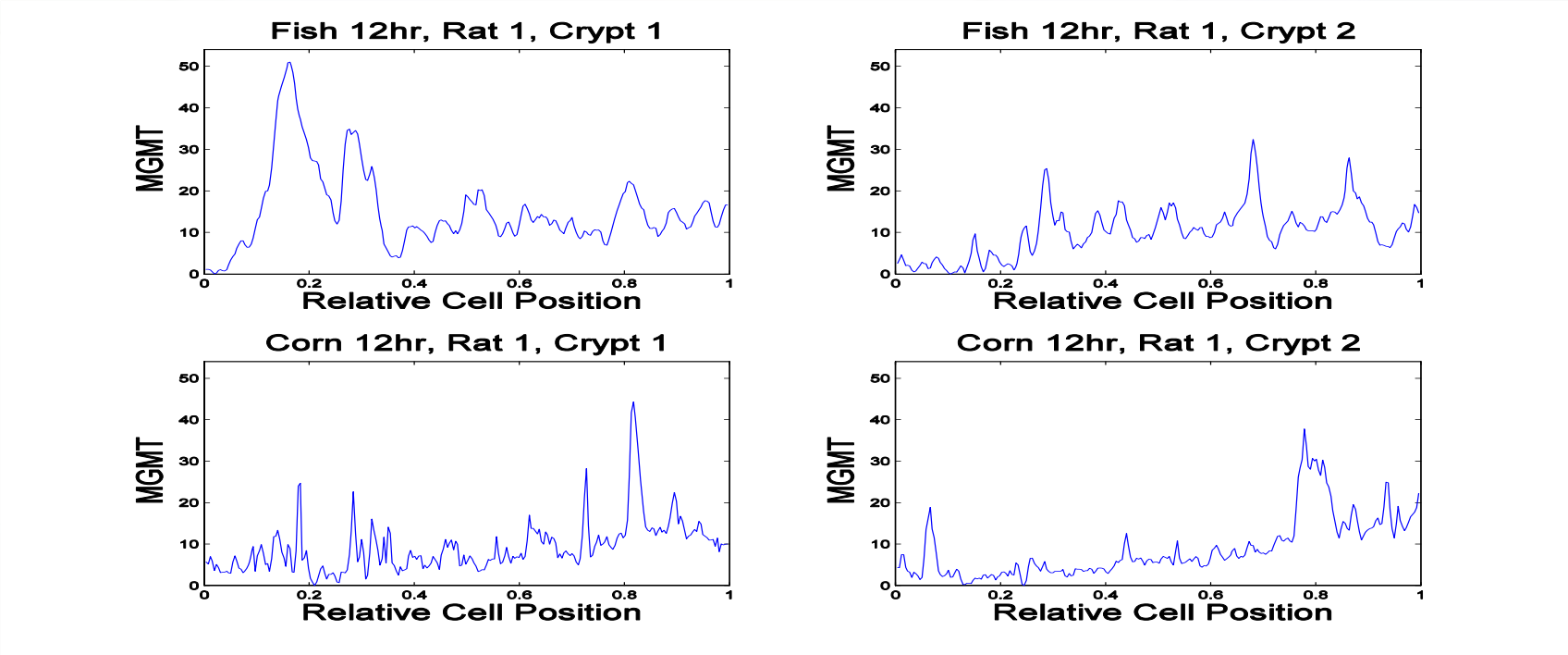
Example: Colon Carcinogenesis



- **Stem Cells:** Mother cells near crypt base
- Depth in crypt ~ age of cells
- ***Relative Cell Position:*** depth within crypts

  $t \in (0,1)$

Colon Carcinogenesis Data

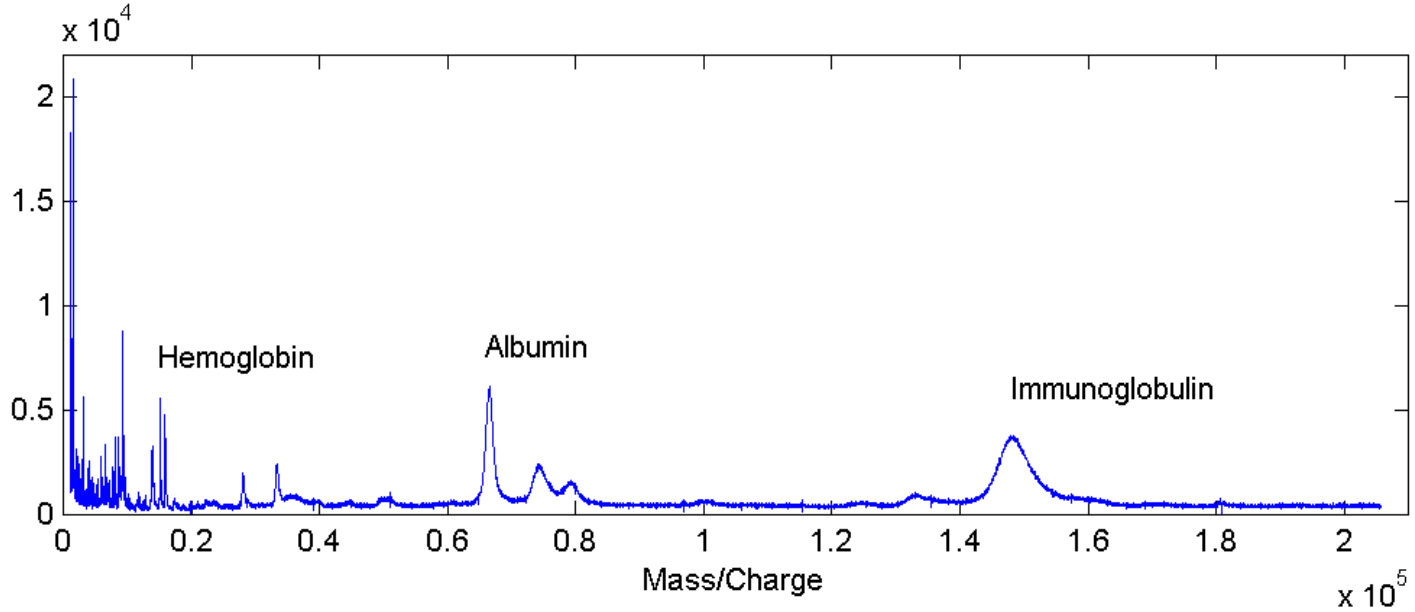


- 30 rats fed 1 of 2 diets, exposed to carcinogen, euthanized after 1 of 5 times after carcinogen exposure (0h,3h,6h,9h,12h)
- MGMT measured via IHCS for 15 crypts/rat, ~250 obs/crypt
- Diet effect? Vary by time and/or crypt depth? Related to other covariates (DNA adduct/apoptosis)? Relative variability from rat-to-rat vs. crypt-to-crypt?

Example: Mass Spectrometry Proteomics

- **Central dogma:** DNA → mRNA → protein
- **Microarrays:** measure expression levels of 10,000s of genes in sample (amount of mRNA)
- **Proteomics:** look at proteins in sample.
  - Gaining increased attention in research
    - Proteins more biologically relevant than mRNA
    - Can use readily available fluids (e.g. blood, urine)
- **MALDI-TOF:** mass spectrometry instrument that can see 100s or 1000s of proteins in sample

Sample MALDI-TOF Spectrum



- **MALDI-TOF Spectrum:** observed function
- **g(*t*) =** intensity of spectrum at m/z value *t*
- Intensity at peak (roughly) estimates the abundance of some protein with molecular weight of *t* Daltons

Example: MALDI

- 16 mice had 1 of 2 cancer cell lines injected into 1 of 2 organs (lung or brain)
- **Cell lines:**
  - **A375P:** human melanoma, low metastatic potential
  - **PC3MM2:** human prostate, highly metastatic
- Blood serum extracted and placed on SELDI chip
- Run at 2 different laser intensities (low/ high)
- Total of 32 spectra (observed functions), 2 per mouse
- Observations on equally-spaced grid of 7985

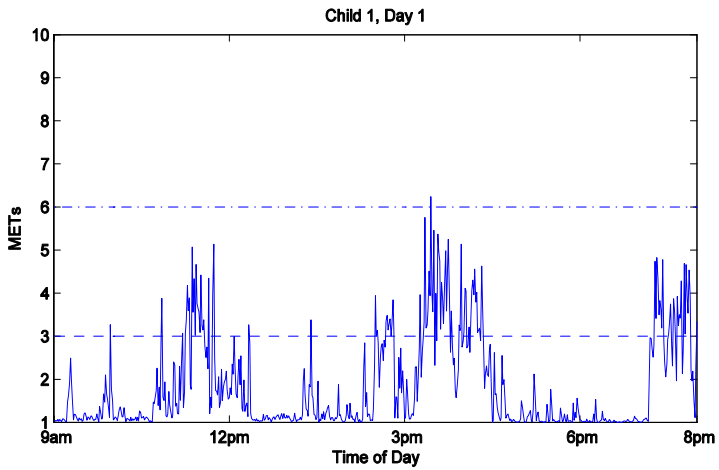
Example: MALDI

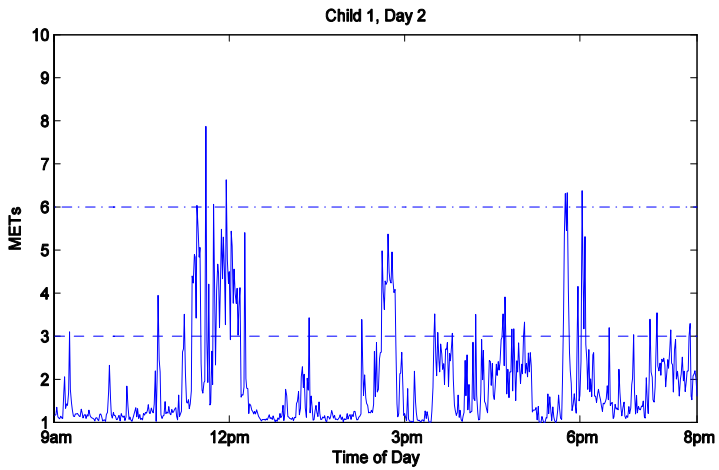
- **Goal:** Find proteins differentially expressed by:
  - Host organ site (lung/brain)
  - Donor cell line (A375P/PC3MM2)
  - Organ-by-cell line interaction
- **Combine information across laser intensities:** Requires us to include in modeling:
  - Functional laser intensity effect
  - Random effect functions to account for correlation between spectra from same mouse

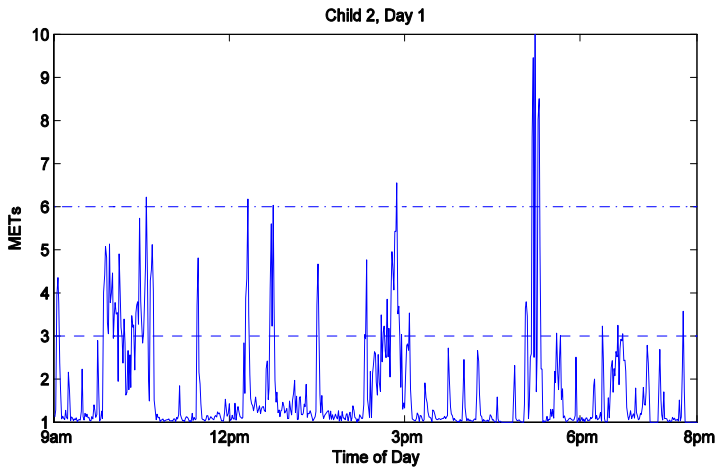
Example: Accelerometer Data

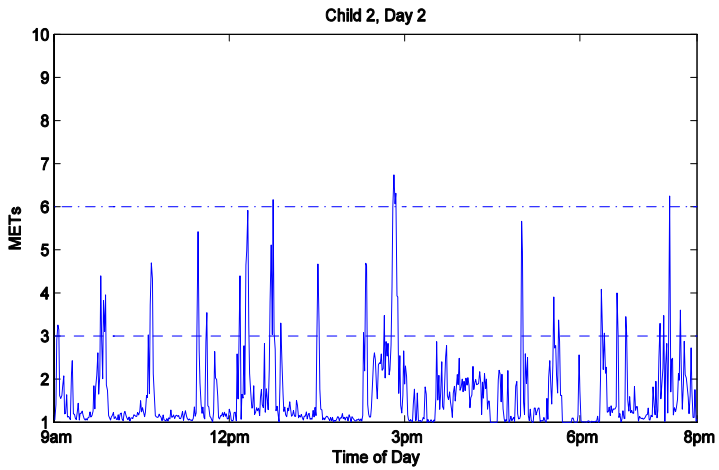
- **Accelerometers:** small motion sensors that digitally record minute-by-minute activity levels
  - Increasingly used in surveillance and intervention studies
- **TriTrac-R3D:** sensor worn on hip
  - Minute-by-minute record of motion in 3 axes
  - Condensed into single activity level measurement/minute
  - Activity “profile” for each day

Accelerometer Data





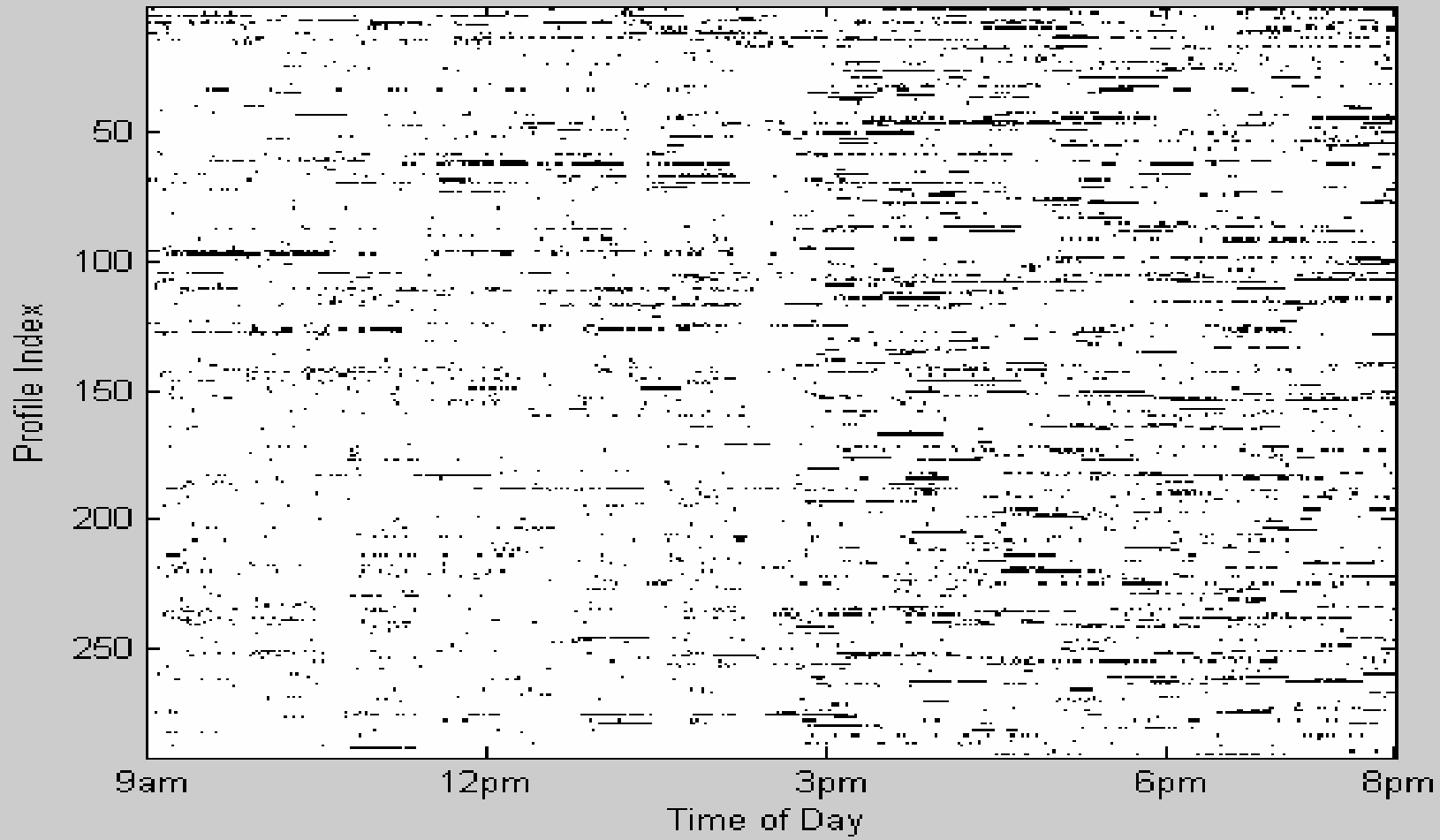




Accelerometer Data

- **Planet Health Study (Harvard University):**
  - Intervention study investigating activity levels of middle school children in Boston area schools
  - Children's activity levels objectively monitored using TriTrac-R3D activity monitor for one or two 4-day sessions
  - **Data considered:** 292 daily profiles/103 children/5 schools, 660 measurements/profile (every minute from 9am-8pm)
- **Primary Goal:** Assess how activity levels vary across child-level and other covariates, and assess whether these effects vary by time-of-day.

Heatmap of Missingness
(Black=missing)



Functional Data

Common characteristics of these data:

- **Extremely high dimensional data**
  (dozens to 100's of curves, 100's to 1000's obs/curve)
- **Complex, irregular, spatially heterogeneous** functions with many local features: requires adaptive methods incorporating nonparametric representation of curves
- **Functional response,** relate to scalar covariates
- **Functions not i.i.d.** – need to model correlation structure between curves
- Want to perform **inference**, not just estimation.

Linear Mixed Models

Linear Mixed Model (Laird and Ware, 1982):

$$\underbrace{Y}_{N\times 1} = \underbrace{X}_{N\times p}\overbrace{\beta}^{p\times 1} + \underbrace{Z}_{N\times m}\overbrace{u}^{m\times 1} + \underbrace{e}_{N\times 1}$$

$$\boldsymbol{u} \sim N(\mathbf{0}, \overbrace{\boldsymbol{D}}^{m\times m})$$

$$\boldsymbol{e} \sim N(\mathbf{0}, \underbrace{\boldsymbol{R}}_{N\times N})$$

- **Fixed effects** part, $X\beta$, accommodate a broad class of mean structures, including main effects, interactions, and linear coefficients.
- **Random effects** part, Zu, provides a convenient mechanism for modeling correlation among the N observations.

Functional Mixed Model (FMM)

Suppose we observe a sample of N curves, $Y_i(t)$, i=1, ..., N, on a closed interval $\mathcal{T}$

$$\underbrace{Y(t)}_{N \text{ functions}} = \overbrace{X}^{N\times p} \underbrace{B(t)}_{p \text{ functions}} + \overbrace{Z}^{N\times m} \underbrace{U(t)}_{m \text{ functions}} + \underbrace{E(t)}_{N \text{ functions}}$$

Functional Mixed Model (FMM)

Suppose we observe a sample of N curves, $Y_i(t)$, i=1, ..., N, on a closed interval $\mathcal{T}$

$$\underbrace{Y(t)}_{N \text{ functions}} = \overbrace{X}^{N\times p} \underbrace{B(t)}_{p \text{ functions}} + \overbrace{Z}^{N\times m} \underbrace{U(t)}_{m \text{ functions}} + \underbrace{E(t)}_{N \text{ functions}}$$

- **DEFN**: $U(t) \sim MGP(P,Q)$ implies the rows of $P^{-1/2}U(t)$ are ind. mean zero Gaussian Processes with covariance surface $Q(t_1,t_2)$.
  - Functional generalization of Matrix Normal (Dawid, 1981).
  - Implies Cov$\{U_i(t_1),U_j(t_2)\}=P_{ij} * Q(t_1,t_2)$

Functional Mixed Model (FMM)

Suppose we observe a sample of N curves, $Y_i(t)$, i=1, ..., N, on a closed interval $\mathcal{T}$

$$\underbrace{Y(t)}_{N \text{ functions}} = \overbrace{X}^{N\times p} \underbrace{B(t)}_{p \text{ functions}} + \overbrace{Z}^{N\times m} \underbrace{U(t)}_{m \text{ functions}} + \underbrace{E(t)}_{N \text{ functions}}$$

$$U(t) \sim MGP(P,Q)$$
$$E(t) \sim MGP(R,S)$$

- **DEFN**: $U(t) \sim MGP(P,Q)$ implies the rows of $P^{-1/2}U(t)$ are ind. mean zero Gaussian Processes with covariance surface $Q(t_1,t_2)$.
  - Functional generalization of Matrix Normal (Dawid, 1981).
  - Implies Cov$\{U_i(t_1),U_j(t_2)\}=P_{ij} * Q(t_1,t_2)$

Functional Mixed Model (FMM)

Suppose we observe a sample of N curves, $Y_i(t)$, i=1, ..., N, on a closed interval $\mathcal{T}$

$$\underbrace{Y(t)}_{N \text{ functions}} = \overbrace{X}^{N\times p} \underbrace{B(t)}_{p \text{ functions}} + \overbrace{Z}^{N\times m} \underbrace{U(t)}_{m \text{ functions}} + \underbrace{E(t)}_{N \text{ functions}}$$

$$U(t) \sim MGP(P,Q)$$
$$E(t) \sim MGP(R,S)$$

- **DEFN**: $U(t) \sim MGP(P,Q)$ implies the rows of $P^{-1/2}U(t)$ are ind. mean zero Gaussian Processes with covariance surface $Q(t_1,t_2)$.
  - Functional generalization of Matrix Normal (Dawid, 1981).
  - Implies Cov$\{U_i(t_1),U_j(t_2)\}=P_{ij} * Q(t_1,t_2)$
- P and R are covariance matrices (between-curve)
- Q(t_1,t_2) and S(t_1,t_2) are covariance surfaces on $\mathcal{T}\times\mathcal{T}$

Model: MALDI Example

Let $Y_i(t)$ be the MALDI spectrum i

$$\log_2\{Y_i(t)\} = B_0(t) + \sum_{j=1}^{4} X_{ij} B_j(t) + \sum_{k=1}^{16} Z_{ik} U_k(t) + E_i(t)$$

- $X_{i1} = 1$ for lung, -1 brain. $X_{i2} = 1$ for A375P, -1 for PC3MM2

  $X_{i3} = X_1 * X_2$ $X_{i4} = 1$ for low laser intensity, -1 high.
- $B_0(t)$ = overall mean spectrum $B_1(t)$ = organ main effect function

  $B_2(t)$ = cell-line main effect $B_3(t)$ = org x cell-line int function

  $B_4(t)$ = laser intensity effect function
  - $Z_{ik} = 1$ if spectrum i is from mouse k ($k=1, \ldots, 16$)
  - $U_k(t)$ is random effect function for mouse k.

Discrete Version of FMM

Suppose each observed curve is sampled on a common equally-spaced grid of length T.

$$\underbrace{Y}_{N\times \mathrm{T}} = \overbrace{X}^{N\times p}\ \underbrace{B}_{p\times \mathrm{T}} + \overbrace{Z}^{N\times m}\ \underbrace{U}_{m\times \mathrm{T}} + \underbrace{E}_{N\times \mathrm{T}}$$

$$U \sim MN(P,Q)$$
$$E \sim MN(R,S)$$

- U and E follow the Matrix Normal distn.
  - $U \sim \mathrm{MN}(P,Q)$ implies $\mathrm{Cov}\{U_{ij}, U_{i'j'}\} = P_{ii'} * Q_{jj'}$
- P and R are covariance matrices ($m \times m$ & $N \times N$)
- Q and S are within-curve covariance matrices ($T \times T$)

Functional Mixed Models

- **Key feature of FMM:** Does not require specification of parametric form for curves
- **Basis function approach:** $y(t)=\Sigma d_{jk}\psi_{jk}(t)$

Functional Mixed Models

- **Key feature of FMM:** Does not require specification of parametric form for curves
- **Basis function approach:** $y(t)=\Sigma d_{jk}\psi_{jk}(t)$
- **<u>Benefits of Using Wavelet Bases</u>**
  1. **Compact support** allows efficient representations of local features
  2. **Whitening property** allows parsimonious yet flexible representations of Q and S
  3. Decomposes function in both **frequency** and **time** domains
     - Enables mechanism for *adaptive regularization* of functions
  4. Orthonormal transformation has **linear representation**, and special structure allows **fast calculation** of coefficients.

Wavelet-Based FMM:

General Approach

1. Project observed functions Y into wavelet space.
2. Fit FMM in wavelet space.
   (Use MCMC to get posterior samples)
3. Project wavelet-space estimates (posterior samples) back to data space.

Wavelet-Based FMM:

General Approach

1. **Project observed functions Y into wavelet space.**
2. Fit FMM in wavelet space
   (Use MCMC to get posterior samples)
3. Project wavelet-space estimates (posterior samples) back to data space.

Wavelet Space Representation

$$y(t) = \sum_{j,k \in \Im} d_{jk} \psi_{jk}(t)$$

$$d_{jk} = \int y(t) \psi_{jk}(t) dt$$

$$\psi_{jk}(t) = 2^{-j/2} \psi(2^{-j/2} t - k)$$

Wavelet Space Representation

$$y(t) = \sum_{j,k \in \Im} d_{jk} \psi_{jk}(t)$$

$$d_{jk} = \int y(t) \psi_{jk}(t) dt$$

$$\psi_{jk}(t) = 2^{-j/2} \psi(2^{-j/2} t - k)$$

Linear Representation:

$$\overbrace{\mathbf{y}}^{1 \times T} = \underbrace{\mathbf{d}}_{1 \times T} \overbrace{\mathbf{W}}^{T \times T}$$

DWT Design Matrix $\mathbf{W} = [\psi_{11}(\mathbf{t})\ \psi_{12}(\mathbf{t})\ \ldots\ \psi_{JK}(\mathbf{t})]$

Wavelet Space Representation

$$y(t) = \sum_{j,k \in \Im} d_{jk} \psi_{jk}(t)$$

$$d_{jk} = \int y(t) \psi_{jk}(t) dt$$

$$\psi_{jk}(t) = 2^{-j/2} \psi(2^{-j/2} t - k)$$

Linear Representation:

$$\overbrace{\mathbf{y}}^{1 \times T} = \underbrace{\mathbf{d}}_{1 \times T} \overbrace{\mathbf{W}}^{T \times T}$$

$$\overbrace{\mathbf{d}}^{1 \times T} = \underbrace{\mathbf{y}}_{1 \times T} \overbrace{\mathbf{W}}^{T \times T}{}'$$

DWT Design Matrix $\mathbf{W} = [\psi_{11}(\mathbf{t})\ \psi_{12}(\mathbf{t})\ \ldots\ \psi_{JK}(\mathbf{t})]$

Wavelet Space Representation

$$y(t) = \sum_{j,k \in \Im} d_{jk} \psi_{jk}(t)$$

$$d_{jk} = \int y(t) \psi_{jk}(t) dt$$

$$\psi_{jk}(t) = 2^{-j/2} \psi(2^{-j/2} t - k)$$

Linear Representation:

$$\overbrace{\mathbf{y}}^{1 \times T} = \underbrace{\mathbf{d}}_{1 \times T} \overbrace{\mathbf{W}}^{T \times T}$$

$$\overbrace{\mathbf{d}}^{1 \times T} = \underbrace{\mathbf{y}}_{1 \times T} \overbrace{\mathbf{W}}^{T \times T}{}'$$

DWT Design Matrix $\mathbf{W} = [\psi_{11}(\mathbf{t})\ \psi_{12}(\mathbf{t})\ \ldots\ \psi_{JK}(\mathbf{t})]$

Given *T*-vector **y** consisting of function sampled on equally-spaced grid, a pyramid-based algorithm for DWT (Mallat) can be used to obtain **d**, *T*–vector of wavelet coefficients, in ***O(T)*** operations (converse also true)

Wavelet-Based FMM

1. **Project observed functions Y to wavelet space**
   - Apply DWT to rows of Y to get wavelet coefficients corresponding to each observed function

$$\underbrace{D}_{N\times T} = \underbrace{Y}_{N\times T}\ \underbrace{W'}_{T\times T}$$

   - Projects the observed curves into the space spanned by the wavelet bases.

Wavelet-Based FMM:

General Approach

1. Project observed functions Y into wavelet space.
2. **Fit FMM in wavelet space (Use MCMC to get posterior samples)**
3. Project wavelet-space estimates (posterior samples) back to data space.

Wavelet Space FMM

$$\underbrace{\boldsymbol{Y}}_{N \times \mathrm{T}} = \overbrace{\boldsymbol{X}}^{N \times p} \underbrace{\boldsymbol{B}}_{p \times \mathrm{T}} + \overbrace{\boldsymbol{Z}}^{N \times m} \underbrace{\boldsymbol{U}}_{m \times \mathrm{T}} + \underbrace{\boldsymbol{E}}_{N \times \mathrm{T}}$$

Wavelet Space FMM

$$\underbrace{\boldsymbol{Y}}_{N \times \mathrm{T}} = \overbrace{\boldsymbol{X}}^{N \times p} \underbrace{\boldsymbol{B}}_{p \times \mathrm{T}} + \overbrace{\boldsymbol{Z}}^{N \times m} \underbrace{\boldsymbol{U}}_{m \times \mathrm{T}} + \underbrace{\boldsymbol{E}}_{N \times \mathrm{T}}$$

Wavelet Representations

Y=DW

B=B*W

U=U*W

E=E*W

Wavelet Space FMM

$$\underbrace{DW}_{N\times T} = \overbrace{X}^{N\times p}\ \underbrace{B^{*}W}_{p\times T} + \overbrace{Z}^{N\times m}\ \underbrace{U^{*}W}_{m\times T} + \underbrace{E^{*}W}_{N\times T}$$

Wavelet Representations

Y=DW

B=B*W

U=U*W

E=E*W

Wavelet Space FMM

$$\underbrace{DW}_{N\times T}W' = \overbrace{X}^{N\times p}\underbrace{B^{*}W}_{p\times T}W' + \overbrace{Z}^{N\times m}\underbrace{U^{*}W}_{m\times T}W' + \underbrace{E^{*}W}_{N\times T}W'$$

Wavelet Representations

Y=DW

B=B*W

U=U*W

E=E*W

WW'=I

Wavelet Space FMM

$$\underbrace{D}_{N \times T} = \overbrace{X}^{N \times p} \underbrace{B^*}_{p \times T} + \overbrace{Z}^{N \times m} \underbrace{U^*}_{m \times T} + \underbrace{E^*}_{N \times T}$$

Wavelet Representations

YW'=D

BW'=B*

UW'=U*

EW'=E*

WW'=I

Wavelet Space FMM

D : empirical wavelet coefficients for observed curves
Row *i* contains wavelet coefficients for observed curve *i*
Each column double-indexed by wavelet scale *j* and location *k*

$$\underbrace{D}_{N\times \mathrm{T}} = \overbrace{X}^{N\times p}\underbrace{B^*}_{p\times \mathrm{T}} + \overbrace{Z}^{N\times m}\underbrace{U^*}_{m\times \mathrm{T}} + \underbrace{E^*}_{N\times \mathrm{T}}$$

- *B*=BW'* & *U*=UW'*: Rows contain wavelet coefficients for the fixed and random effect functions, respectively
- *E*=EW'* is the matrix of wavelet-space residuals

Wavelet Space FMM

D : empirical wavelet coefficients for observed curves
Row *i* contains wavelet coefficients for observed curve *i*
Each column double-indexed by wavelet scale *j* and location *k*

$$\underbrace{D}_{N\times \mathrm{T}} = \overbrace{X}^{N\times p}\underbrace{B^*}_{p\times \mathrm{T}} + \overbrace{Z}^{N\times m}\underbrace{U^*}_{m\times \mathrm{T}} + \underbrace{E^*}_{N\times \mathrm{T}}$$

$$U^* \sim MN(P, Q^*)$$
$$E^* \sim MN(R, S^*)$$

- *B*=BW'* & *U*=UW'*: Rows contain wavelet coefficients for the fixed and random effect functions, respectively
- *E*=EW'* is the matrix of wavelet-space residuals

Wavelet Space FMM

D : empirical wavelet coefficients for observed curves
Row *i* contains wavelet coefficients for observed curve *i*
Each column double-indexed by wavelet scale *j* and location *k*

$$\underbrace{D}_{N\times T} = \overbrace{X}^{N\times p}\underbrace{B^*}_{p\times T} + \overbrace{Z}^{N\times m}\underbrace{U^*}_{m\times T} + \underbrace{E^*}_{N\times T}$$

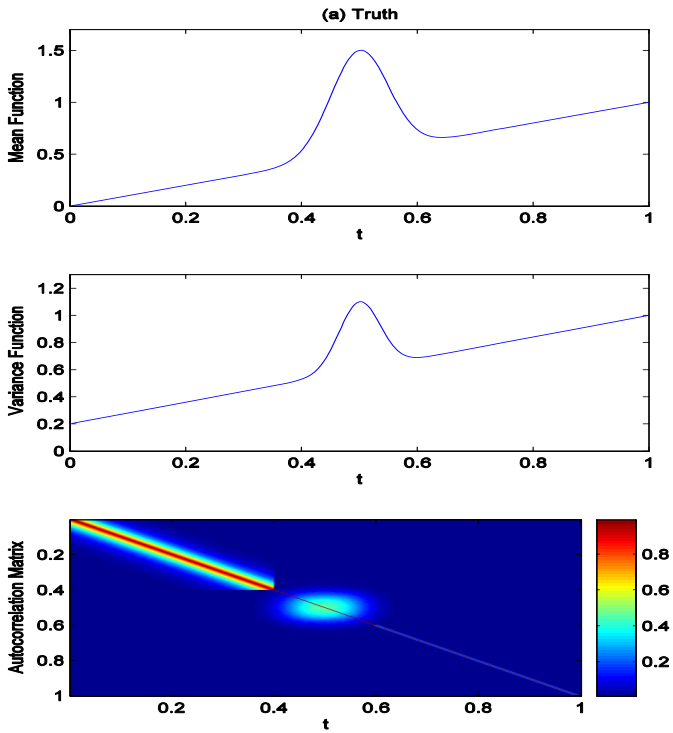
$$U^* \sim MN(P, Q^*)$$
$$E^* \sim MN(R, S^*)$$

- *B*=BW'* & *U*=UW'*: Rows contain wavelet coefficients for the fixed and random effect functions, respectively
- *E*=EW'* is the matrix of wavelet-space residuals
- *Q*=WQW'* and *S*=WSW'* model the covariance structure between wavelet coefficients for a given function.
- *P, Q*, R* and *S** are typically too large to estimate in an unstructured fashion.

Covariance Assumptions

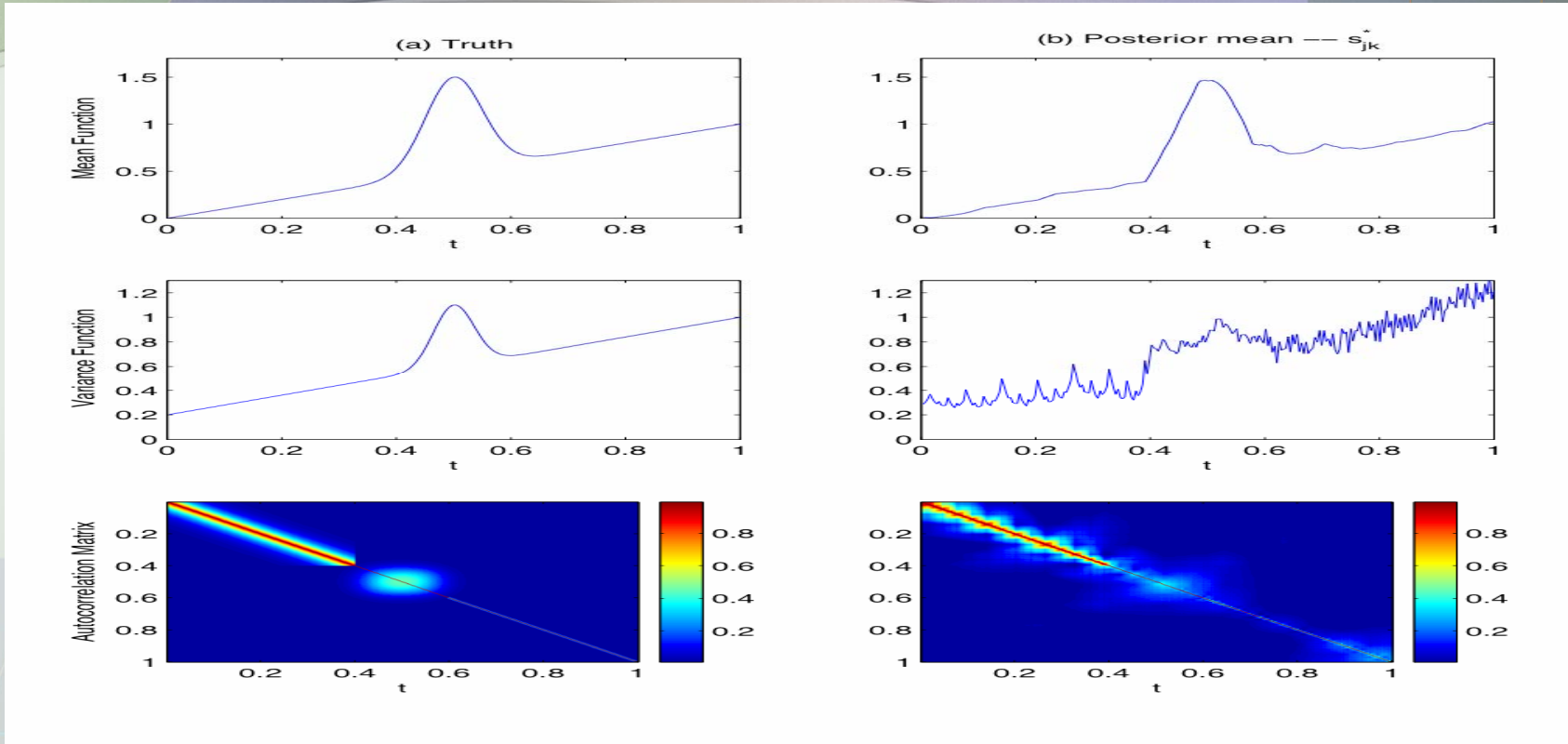
- We choose parametric structures for P and R to model the covariance structure between the curves.
  - Based on the experimental design
  - As in linear mixed models.
- We assume the between-wavelet covariance matrices Q^* and S^* are diagonal ($Q^* = \text{diag}\{q_{jk}\}, S^* = \text{diag}\{s_{jk}\}$).
  - Wavelet coefficients within given function modeled as independent
  - Heuristically justified by whitening property of DWT
  - Common working assumption in wavelet regression settings
  - Is parsimonious in wavelet space (T parameters), yet leads to flexible class of nonstationary covariance structures in data space
  - Key: We allow variances to vary by scale j and location k

Simulation: Covariance Structure



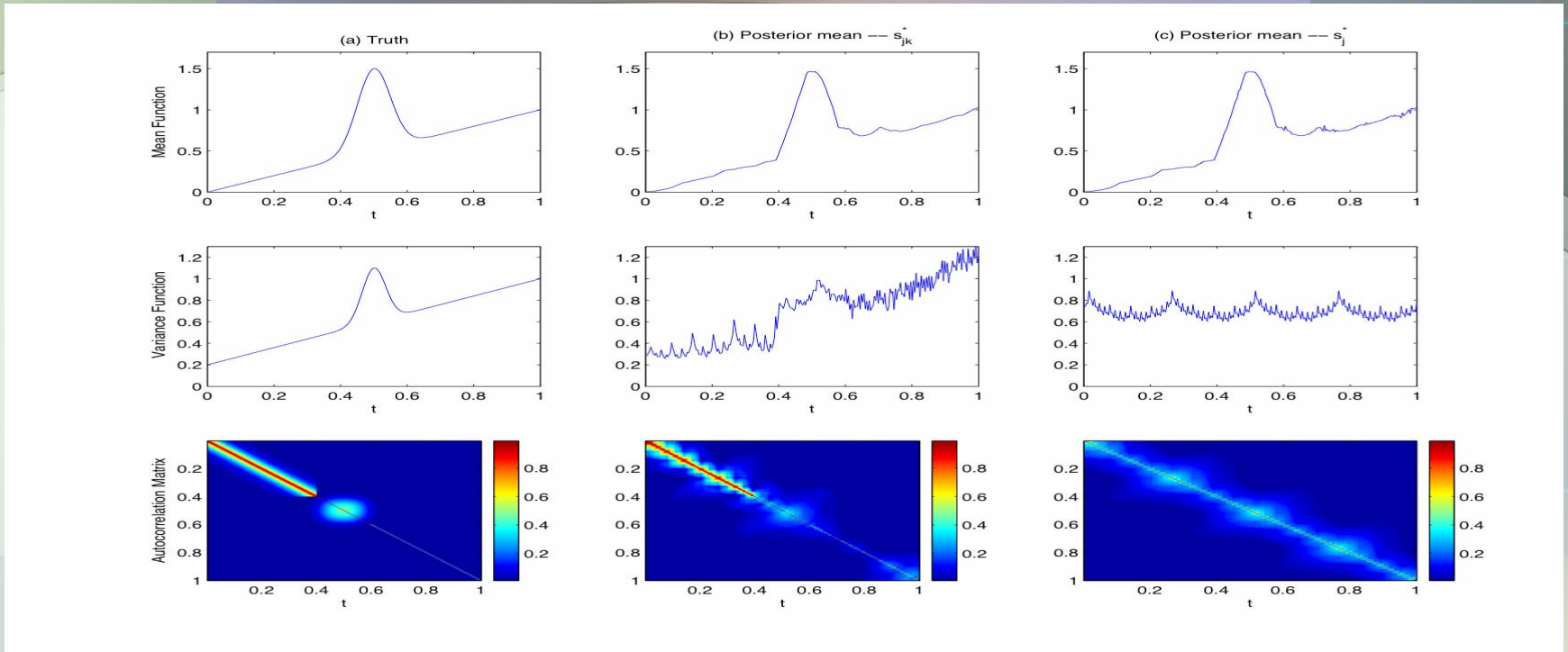
- True mean: line plus peak
- True variance: increasing in t, with extra var at peak
- True autocorrelation: Strong autocorrelation (0.9) at left, weak autocorrelation (0.1) right, extra at peak

Simulation: Covariance Structure



- Independence in wavelet space accommodates varying degrees of autocorrelation in data space
- Allowing variance components to vary across scale *j* and location *k* accommodates nonstationarities

Simulation: Covariance Structure



- Most wavelet regression methods (Fan and Lin 1998, Morris, et al. 2003, Abramovich and Angelini 2006, Antoniadis and Sapatinas 2006) only index variances by scale *j*, but not location *k*.
- Not flexible enough to capture nonstat. covariance features
- Unnecessary restriction in multiple function case, since replicate functions allow estimation of separate VC for each (*j*,*k*)

Model Each Column Separately

$$\underbrace{d_{jk}}_{N\times 1} = \overbrace{X}^{N\times p}\underbrace{B^*_{jk}}_{p\times 1} + \overbrace{Z}^{N\times m}\underbrace{u^*_{jk}}_{m\times 1} + \underbrace{e^*_{jk}}_{N\times 1}$$

$$u^*_{jk} \sim N(0, P\cdot q^*_{jk})$$

$$e^*_{jk} \sim N(0, R\cdot s^*_{jk})$$

Prior Assumptions

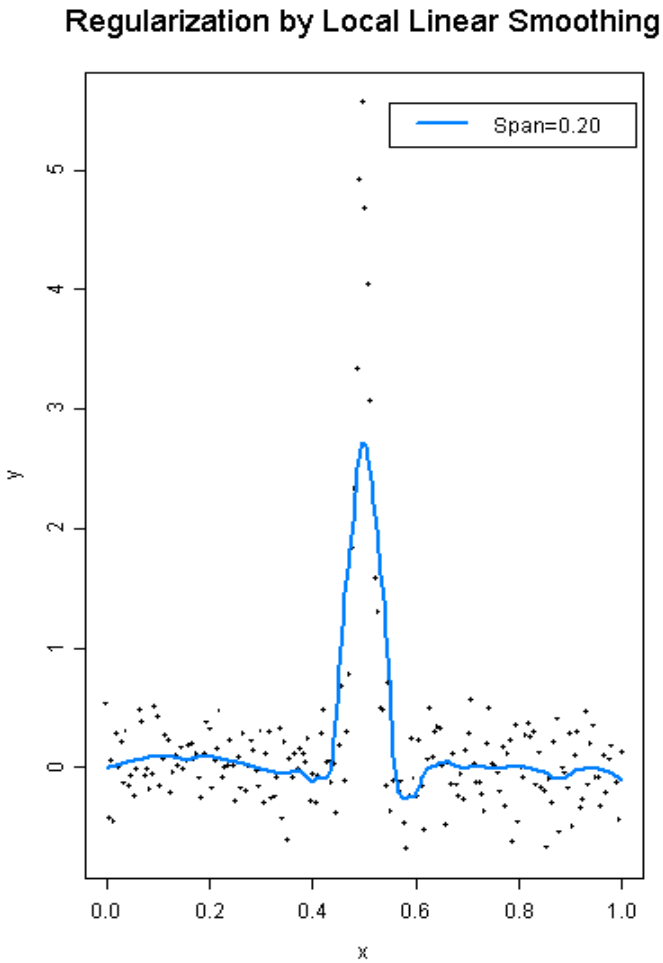
Mixture prior on B_{ijk}^*:

$$B_{ijk}^* = \gamma_{ijk}^* N(0, \tau_{ij}) + (1 - \gamma_{ijk}^*)\delta_0$$

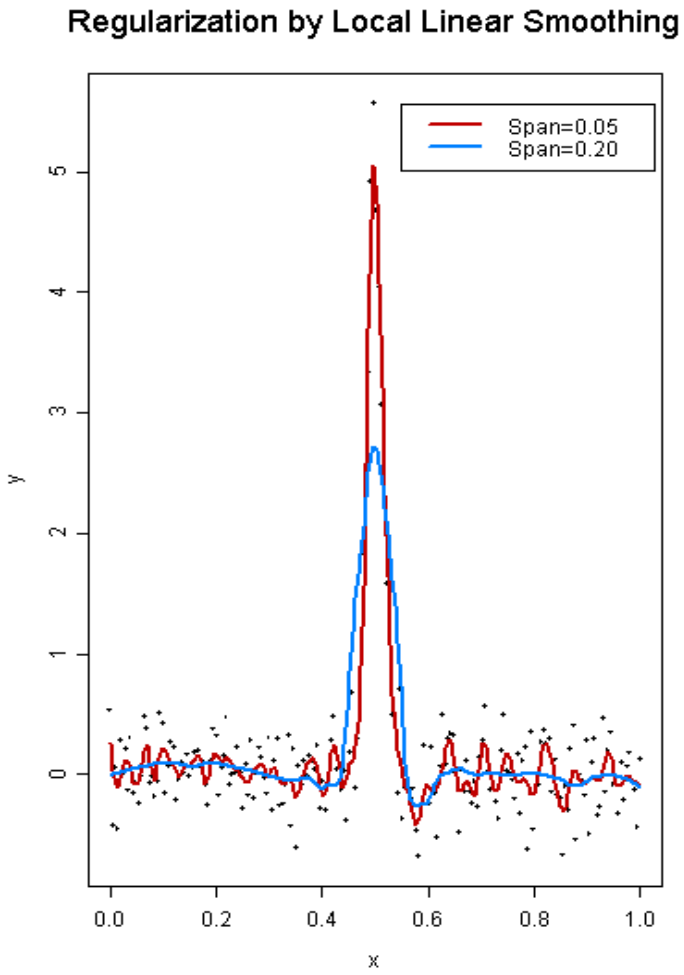
$$\gamma_{ijk}^* = \text{Bernoulli}(\pi_{ij})$$

- Nonlinearly shrinks B_{ijk}^* towards 0, leading to ***adaptively regularized*** estimates of $B_i(t)$.
- τ_{ij} **&** π_{ij} **are regularization parameters**
- They can be elicited, or estimated from the data using empirical Bayes approach (extending Clyde & George 1999 to FMM)

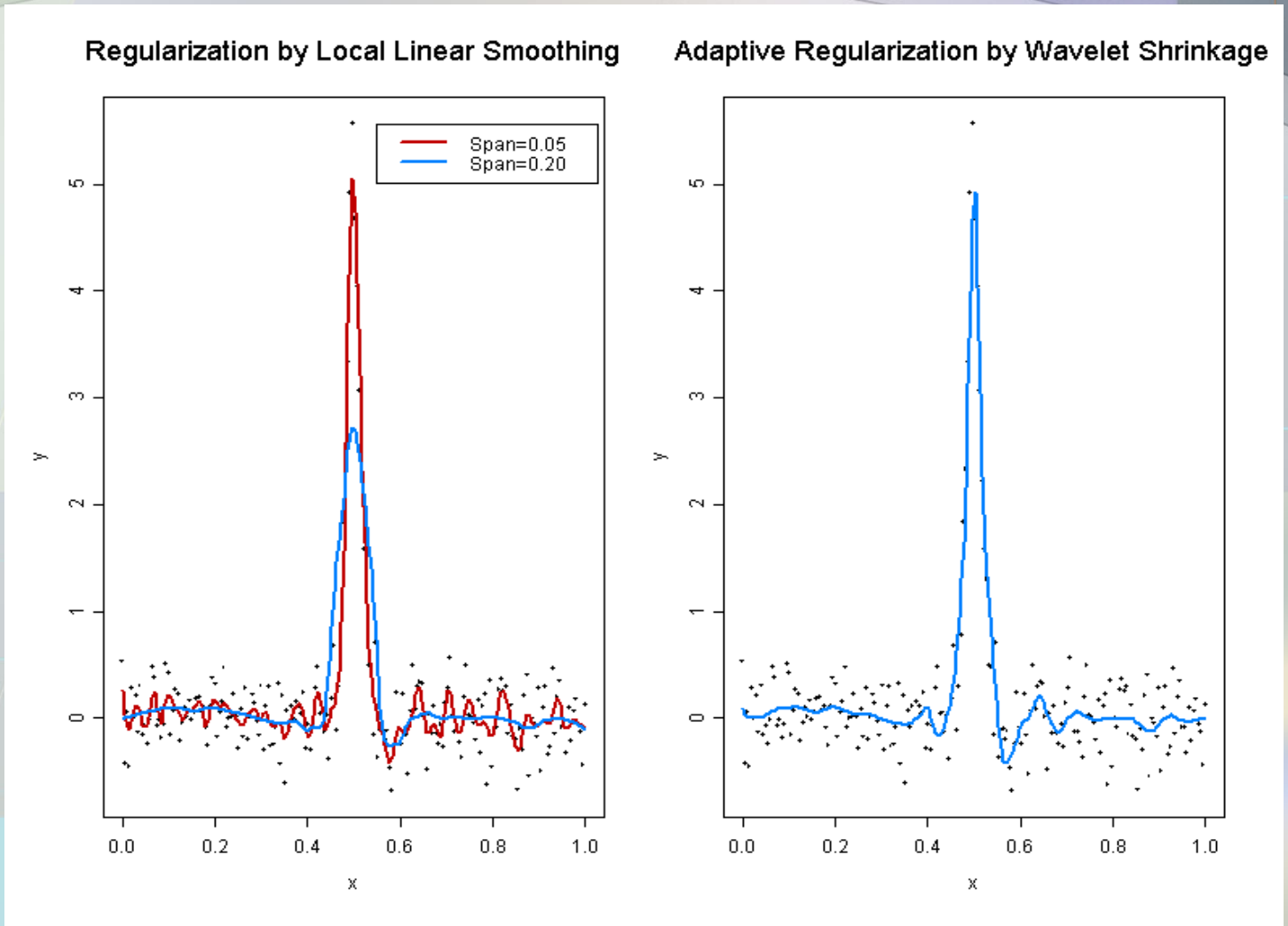
Adaptive Regularization



Adaptive Regularization



Adaptive Regularization



Model Fitting

- MCMC to obtain posterior samples of model quantities
  - Work with marginal likelihood; U* integ. out;
- Let Ω be a vector containing ALL covariance parameters (i.e. P,R,Q^*,S^*).

MCMC Steps

1. Sample from $f(B^*|D,\Omega)$:

   Mixture of normals and point masses at 0 for each i,j,k.

2. Sample from $f(\Omega|D,B^*)$:

   Metropolis-Hastings steps for each j,k

3. If desired, sample from $f(U^*|D,B^*,\Omega)$:

   Multivariate normals

Wavelet-Based FMM:

General Approach

1. Project observed functions Y into wavelet space.
2. Fit FMM in wavelet space
   (Use MCMC to get posterior samples)
3. **Project wavelet-space estimates (posterior samples) back to data space.**

Wavelet-Based FMM

3. Project wavelet-space estimates (posterior samples) back to data space.

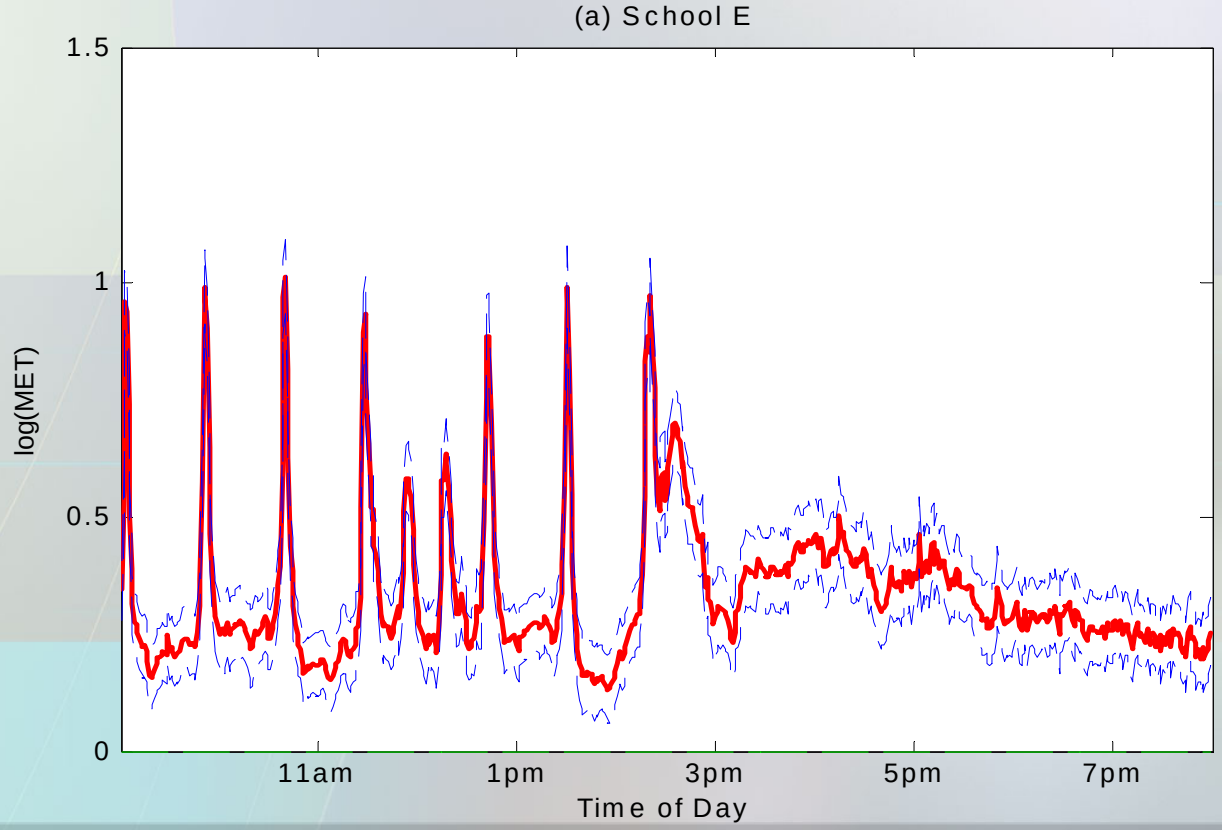
- Apply IDWT to posterior samples of **B*** to get posterior samples of fixed effect functions $B_j(t)$ for $i=1,\ldots, p$, on grid **t**.

$$\mathbf{B}=\mathbf{B}^*\mathrm{W}$$

- Posterior samples of $U_k(t)$, Q, and S are also available, if desired.

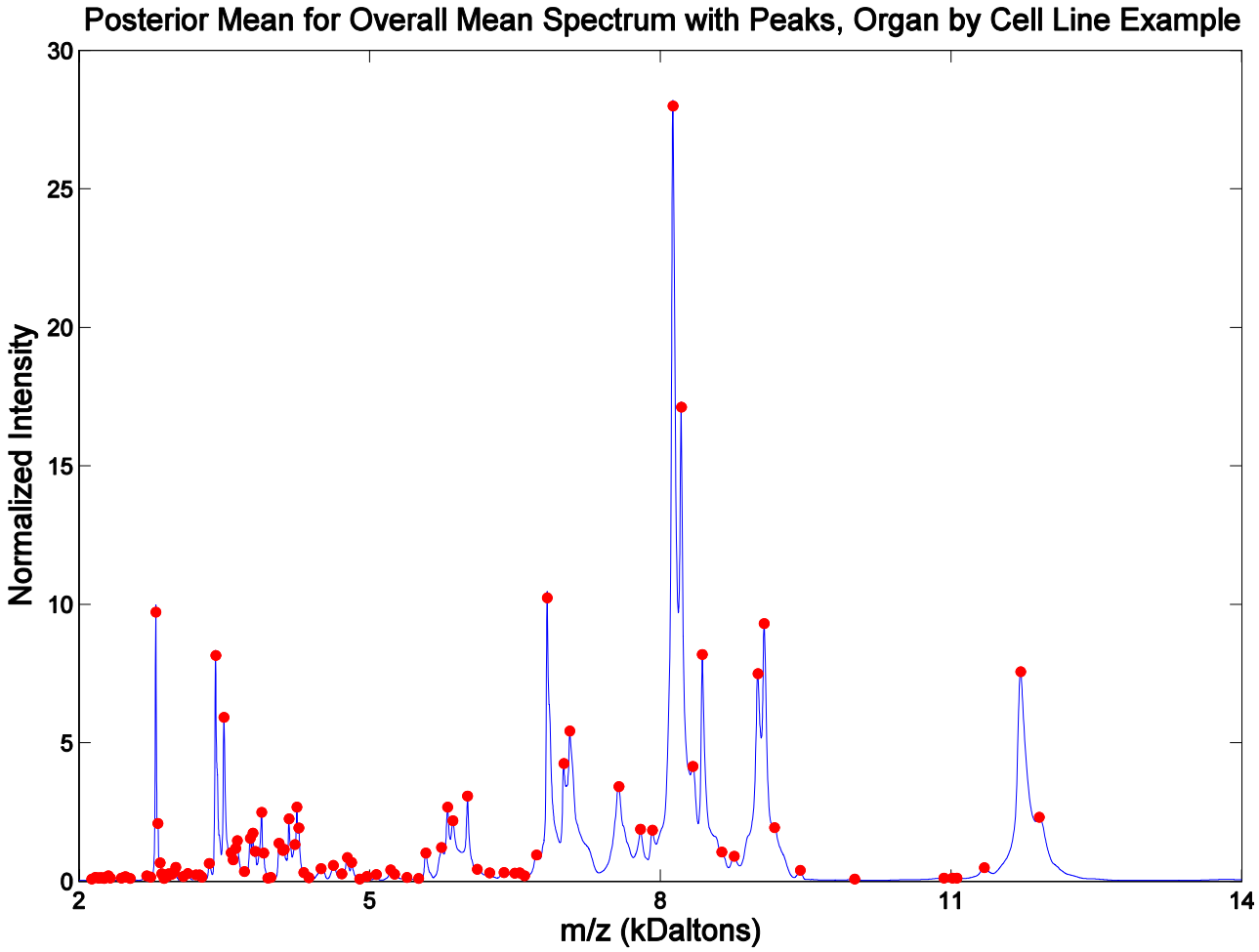
Adaptive Regularization

- Posterior samples/estimates of fixed effect functions $B_i(t)$ ***adaptively regularized*** as a result of shrinkage prior applied to wavelet coefficients.



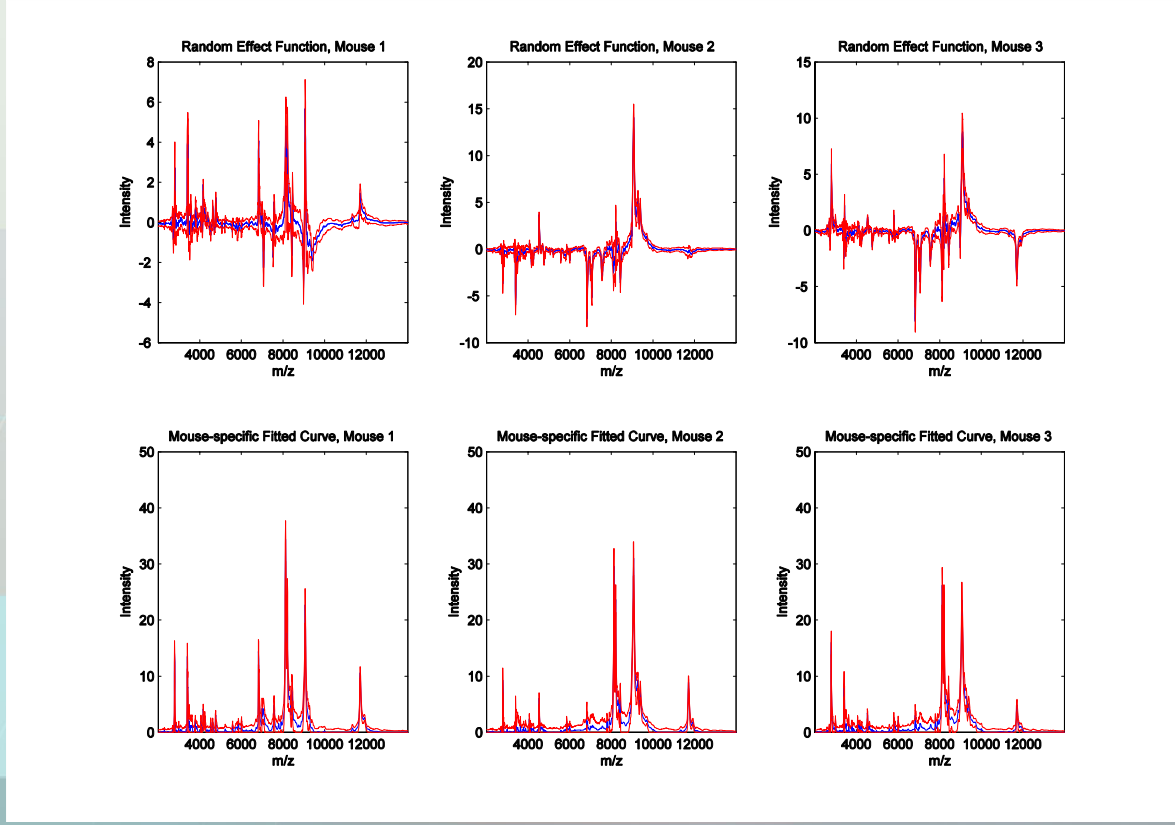
- Able to preserve dominant spikes in mean curves, if present

Adaptive Regularization



Adaptive Regularization

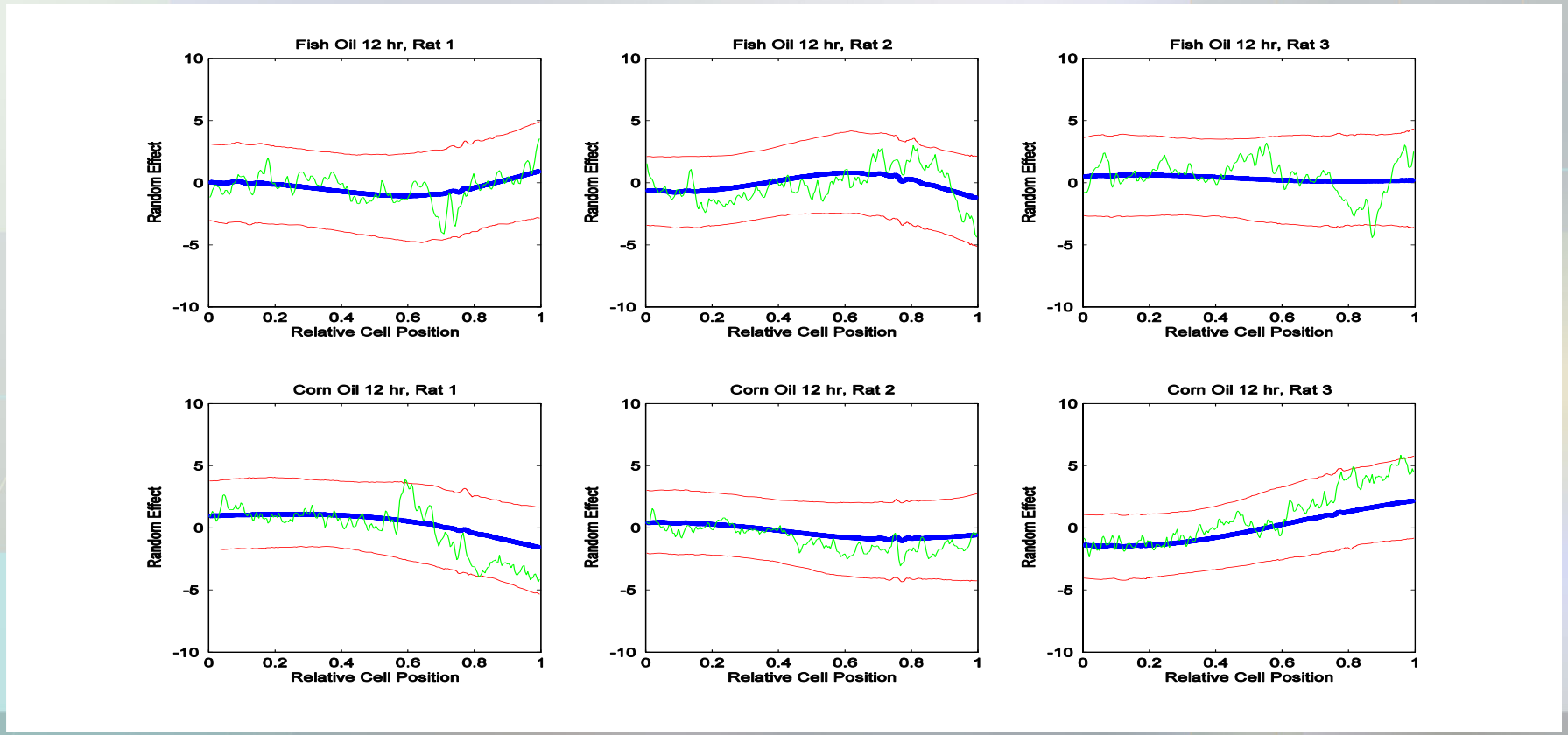
- Posterior samples/estimates of random effect functions $U_j(t)$ are also ***adaptively regularized*** from Gaussian prior, since each wavelet coefficient has its own random effect & residual variance



- Able to preserve spikes in random effect functions, as well
- Important for estimation of random effect functions AND for valid inference on fixed effect functions.

Adaptive Regularization

- While adaptive to irregularity, this framework can also yield relatively smooth effect functions when the data supports smooth representations.



Bayesian Inference

Given posterior samples of all model quantities, we can perform desired Bayesian inference or prediction:

1. Pointwise posterior **credible intervals** for functional effects
2. **Posterior probabilities** of interest – either pointwise, joint, or aggregating across locations within the curve.
3. Can account for multiple testing in identifying significant regions of curves by controlling the **expected Bayesian FDR**
4. Can compute **posterior predictive distributions**, which can be used for model-checking or other purposes.

Bayesian Inference:

Identifying Significant Regions of Curves

Procedure (desired effect size $\geq\delta$, FDR α)

1. Compute pointwise posterior probabilities of effect size of interest being at least δ
   $p_{il}=\Pr\{|B_i(t_l)|>\delta|Y\}$ for $l=1, \ldots, T$
2. Sort in descending order of p_{il} $\{p_{i(l)}, l=1, \ldots, T\}$
3. Identify cutpoint φ_α on posterior probabilities that controls expected Bayesian FDR to be $\leq \alpha$ $\varphi_\alpha = p_{i(\lambda)}$, where

$$\lambda = \max\left[l^* : \sum_{l=1}^{l^*} \{1 - p_{i(l)}\} \leq l^* \alpha \right]$$

4. Flag the set of locations $\{t_l : p_{il} \leq \varphi_\alpha\}$ as significant (According to model, expect only α to be false pos.)

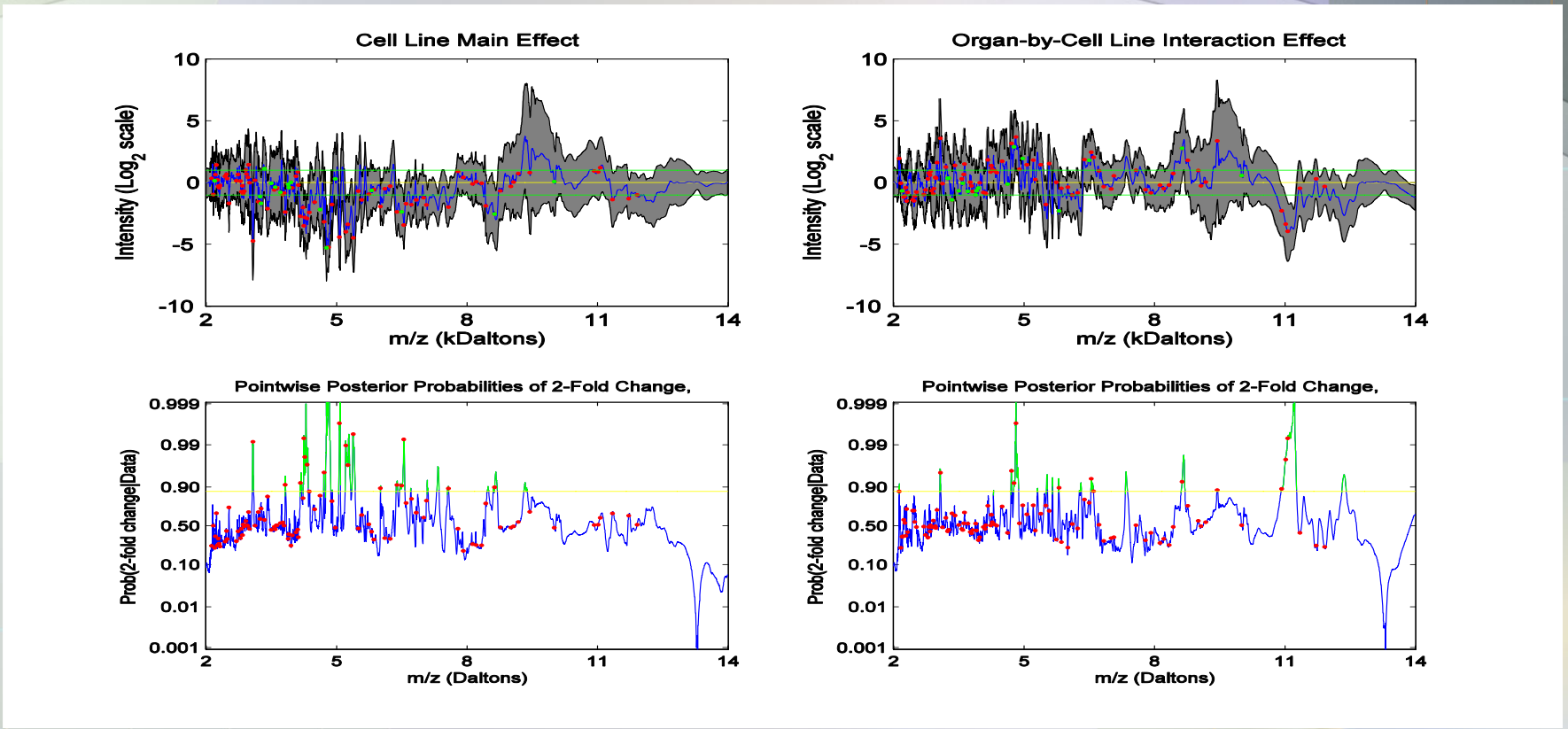
Bayesian Inference:

Identifying Significant Regions of Curves

Notes

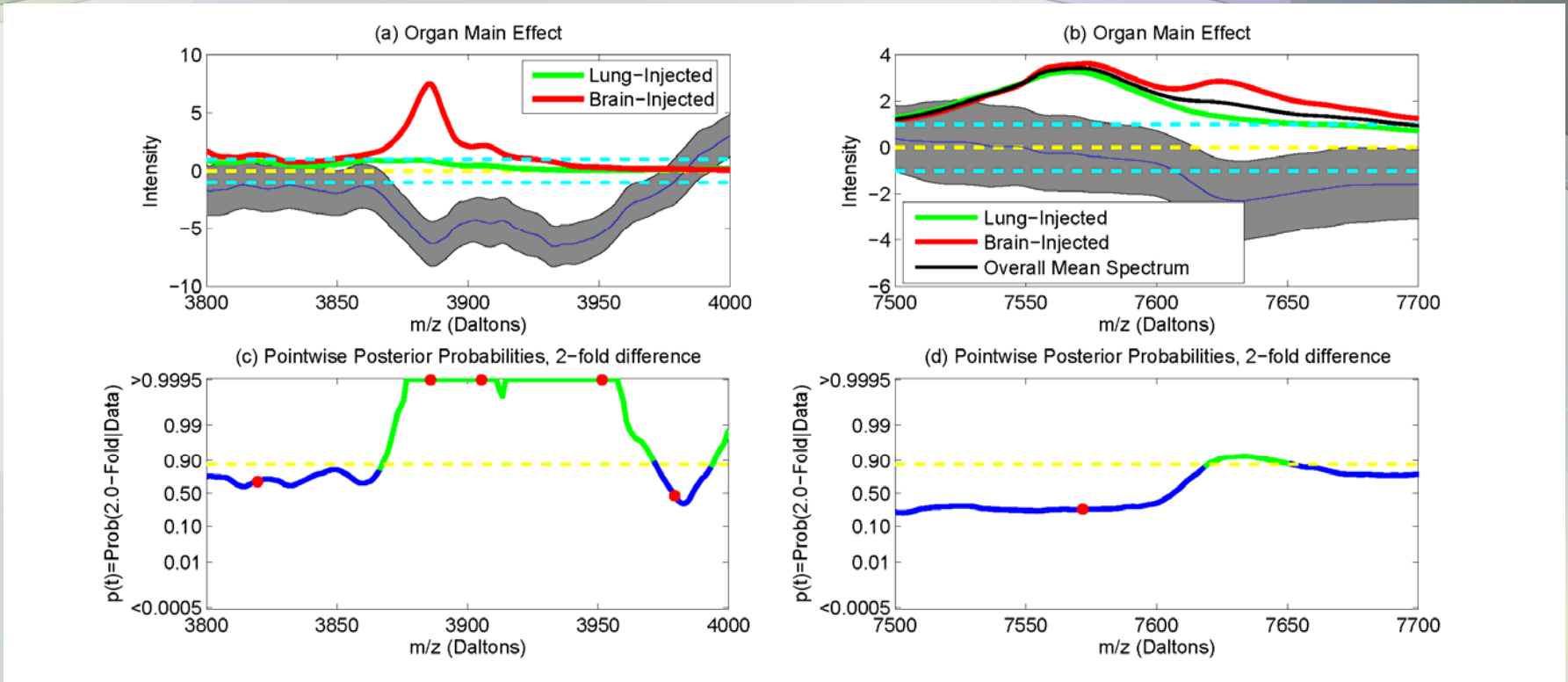
- This approach takes both statistical and practical significance into account when declaring differences significant.
- Given choice of cutpoint on posterior probabilities to flag significant regions, we can also estimate false negative rate (FNR), sensitivity, and specificity for declaring flagged regions significantly different.
- This approach can be used for any setting where estimated posterior probabilities are available for effect sizes (including but not limited to functional data setting)

Results: MALDI Example



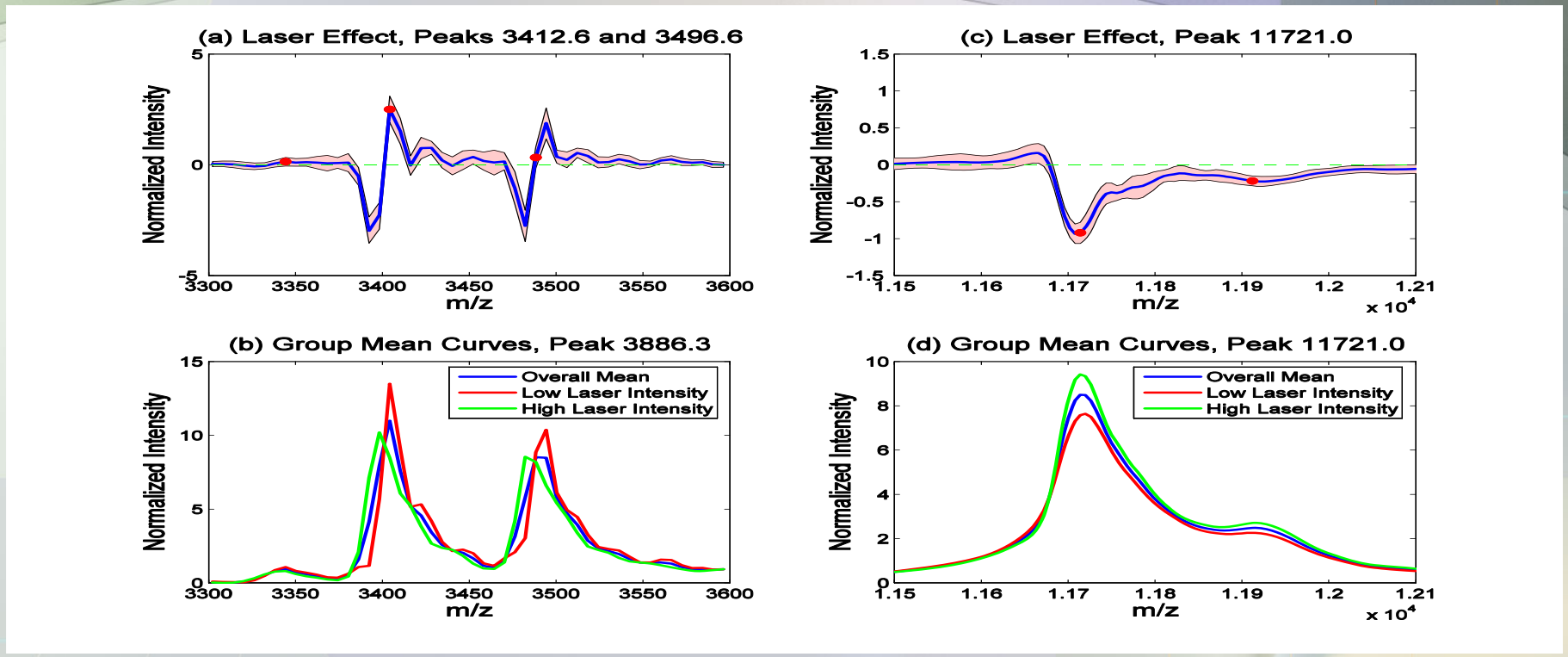
- Using α=0.05, δ=1 (2-fold expression on $\log_2$ scale), we flag a number of spectral regions.

Results: MALDI Example



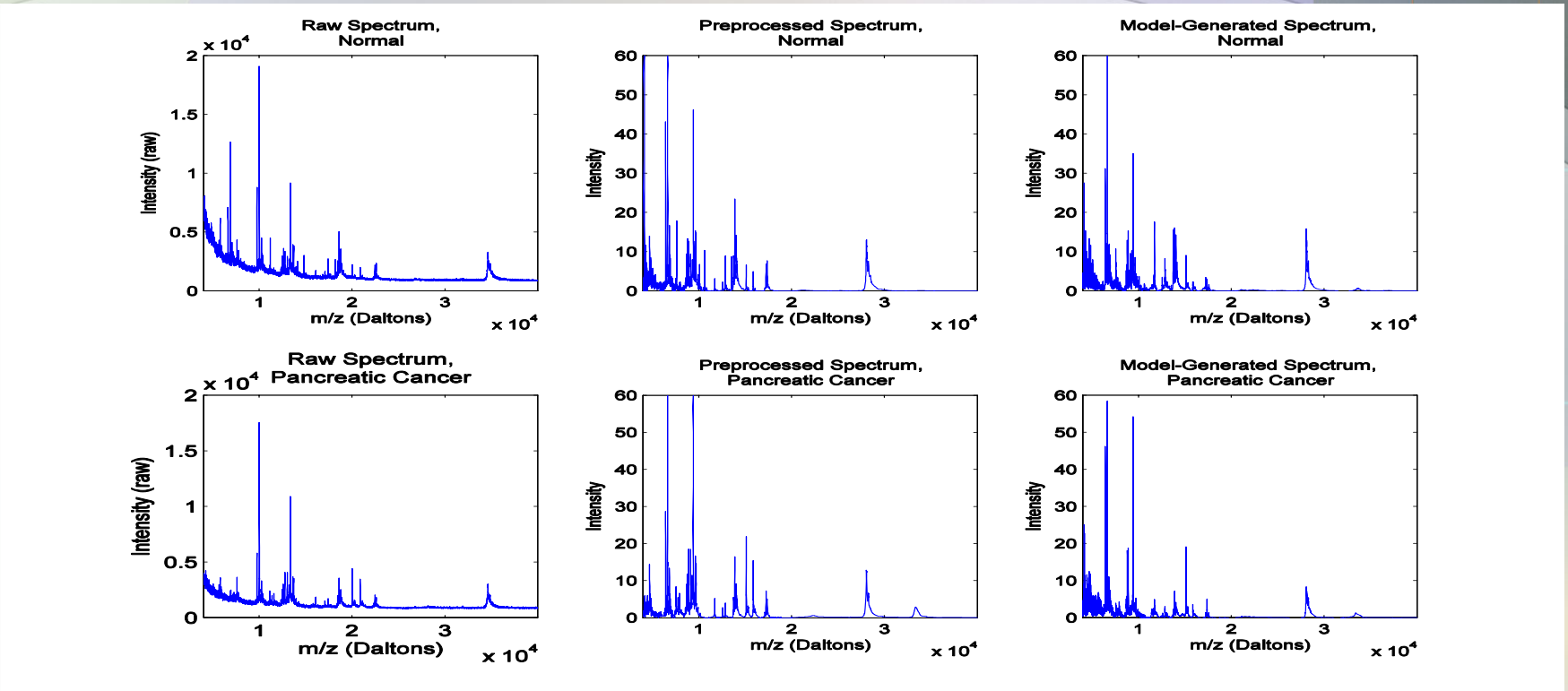
- 3900 D (~100-fold) (CGRP-II): dilates blood vessels in brain
- 7620 D (~5-fold) (neurogranin): active in synaptic modeling in brain (Not detected as peak)

Results: SELDI Example



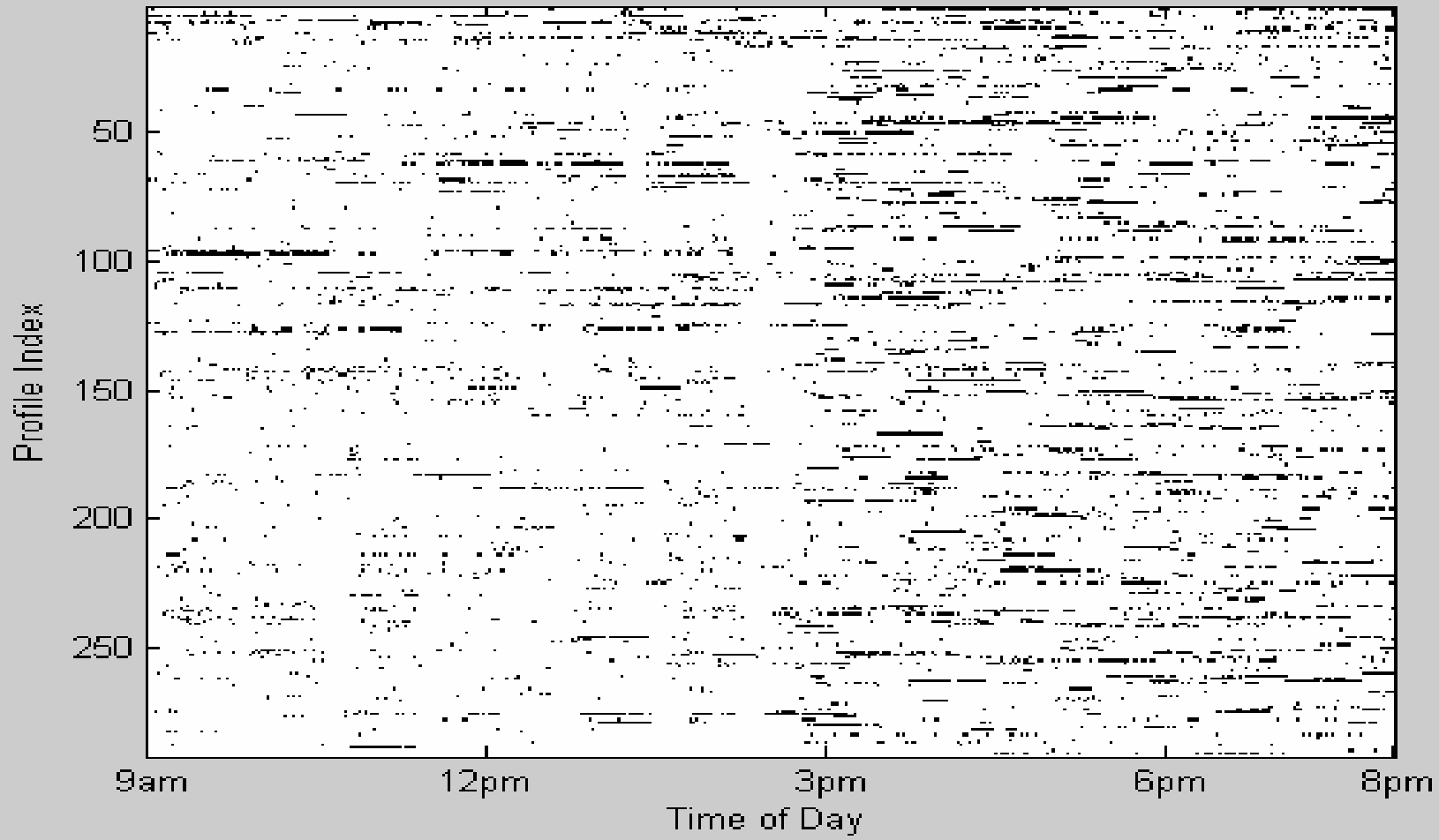
- Inclusion of nonparametric functional laser intensity effect is able to adjust for systematic differences in the *x* and *y* axes between laser intensity scans

Results: MALDI Example

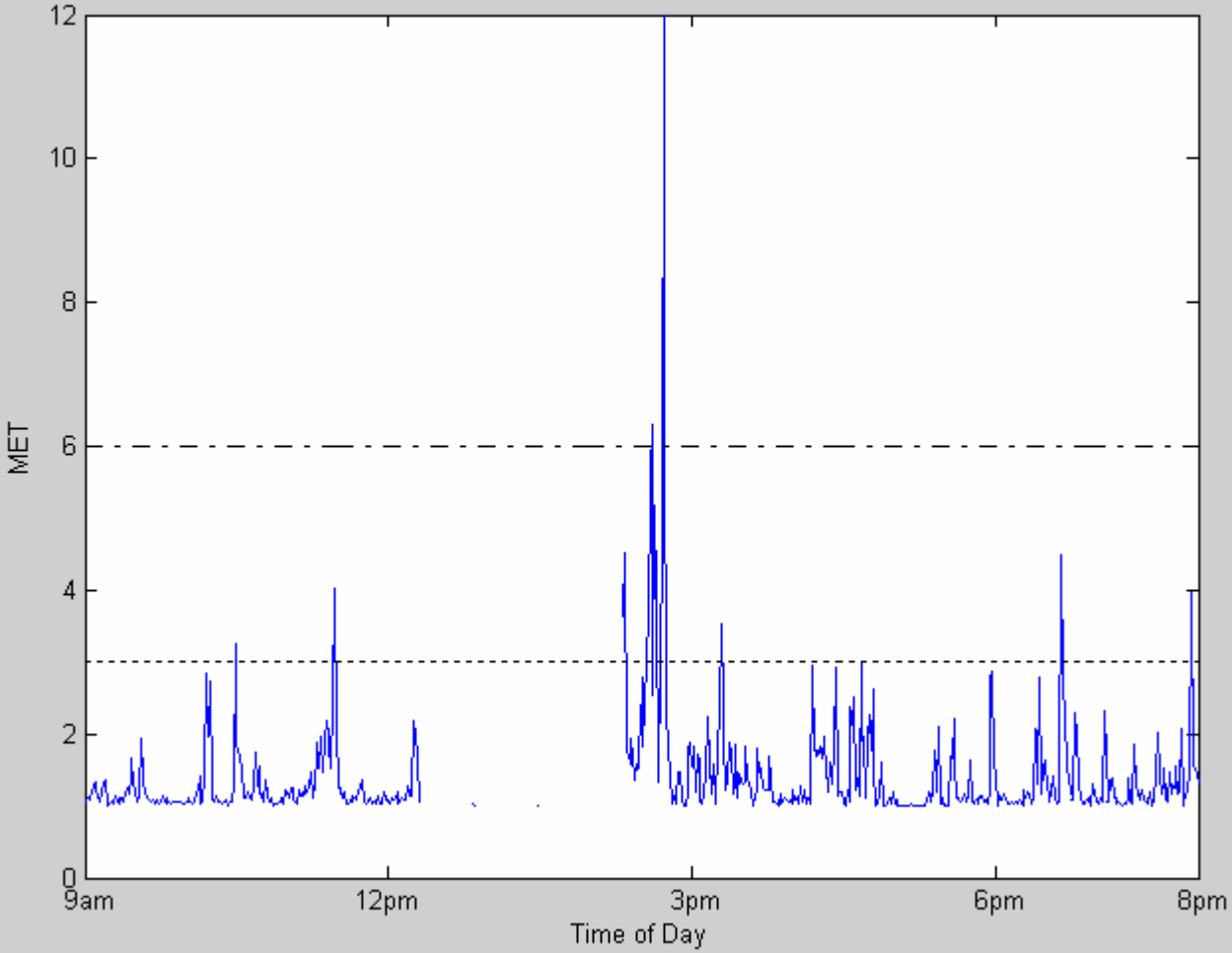


- Draws of spectra from posterior predictive distribution yield data that looks like real MALDI data (3rd column), indicating reasonable model fit.

Heatmap of Missingness
(Black=missing)



Incomplete Profile



Incomplete Profiles

- WFMM can only be applied to *complete profiles* (with no missing regions)
  - 95 of the 292 profiles *complete* from 9am-8pm
- How do we incorporate information from other 197 *incomplete profiles* ?

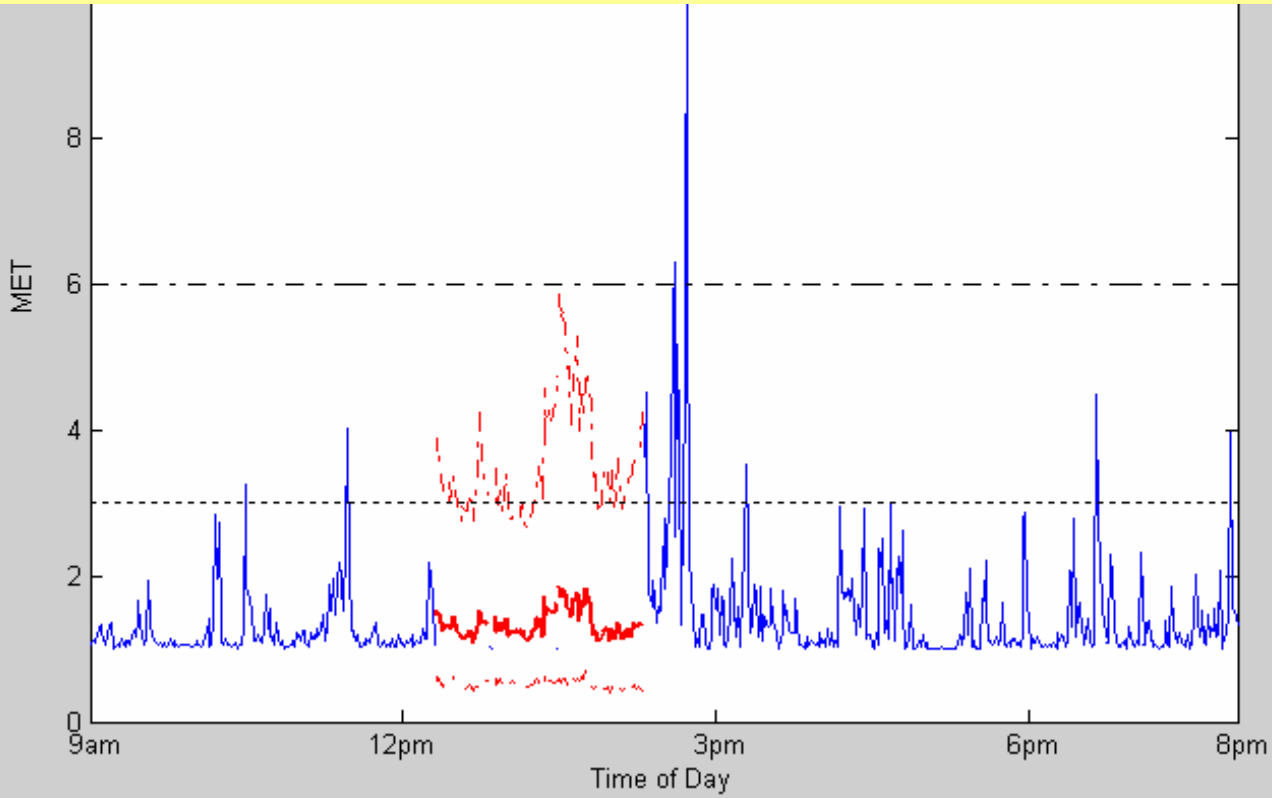
Approach: Incomplete Profiles

1. First fit model to *complete profiles*, get posterior distribution samples for model parameters.
2. Use these to estimate *predictive distributions* for the the incomplete profiles

Predictive Distribution

$$\mu_i(t) = E\{Y_i(t) \mid Y^C\} = \int Y_i(t)\, f\{Y_i(t) \mid X, Z, \Theta\}\, f(\Theta \mid Y^C)\, d\Theta$$

$$\Sigma_i(t_1, t_2) = COV\ \{Y_i(t_1), Y_i(t_2) \mid Y^C\}$$



Approach: Incomplete Profiles

1. First fit model to *complete profiles*, get posterior distribution samples for model parameters.
2. Use these to estimate *predictive distributions* for the the incomplete profiles
   - Borrow information about what the curves in these regions look like.
   - Account for child-specific and day-specific covariates.
   - Integrates over uncertainty in parameter estimation

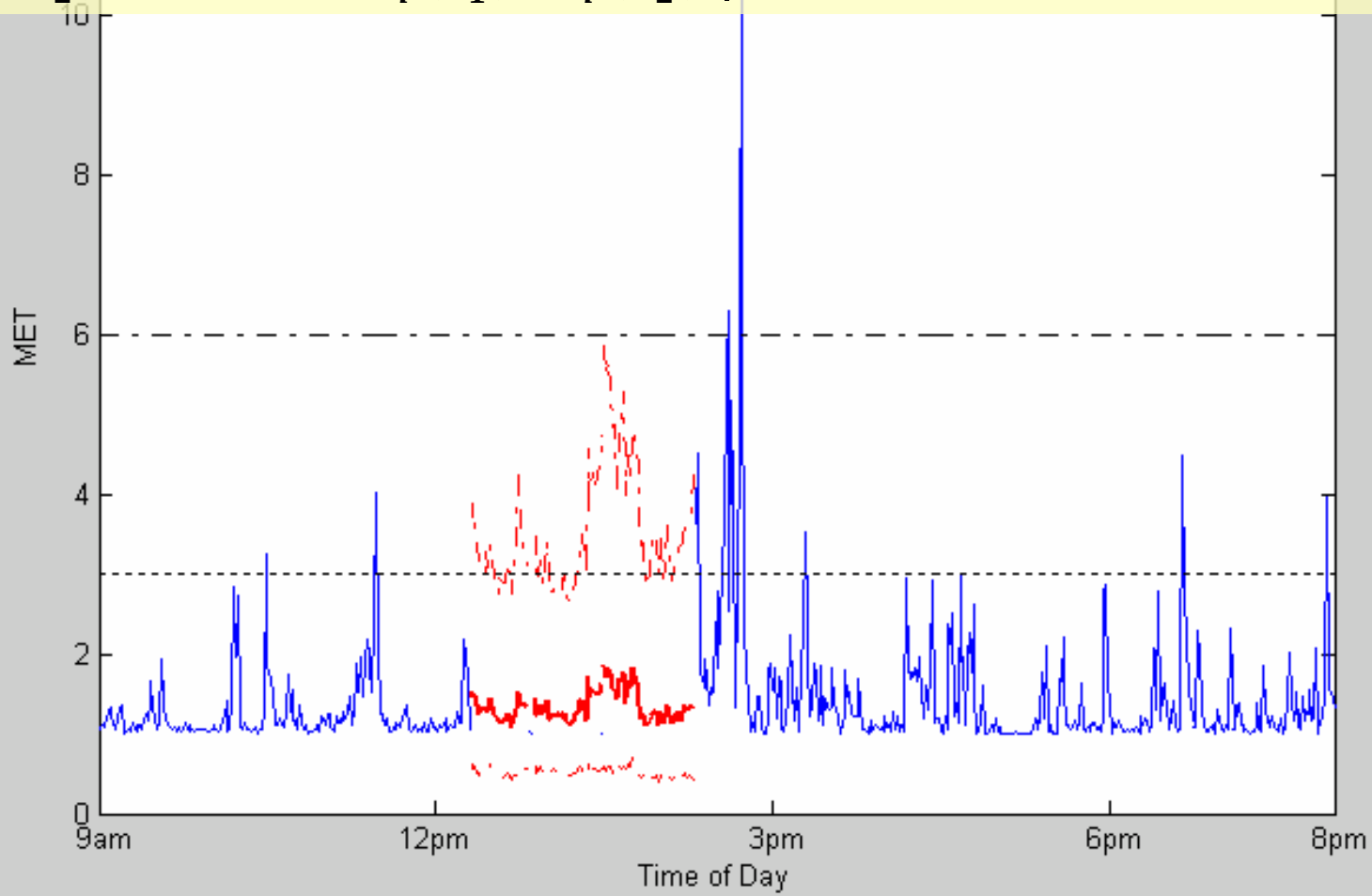
Approach: Incomplete Profiles

1. First fit model to *complete profiles*, get posterior distribution samples for model parameters.
2. Use these to estimate *predictive distributions* for the the incomplete profiles
   - Borrow information about what the curves in these regions look like.
   - Account for child-specific and day-specific covariates.
   - Integrates over uncertainty in parameter estimation
3. Regress missing data on the observed data to obtain *imputation distribution* for missing regions

Predictive Distribution

$$\mu_i(t) = E\{Y_i(t) \mid Y^C\} = \int Y_i(t)\, f\{Y_i(t) \mid X, Z, \Theta\}\, f(\Theta \mid Y^C)\, d\Theta$$

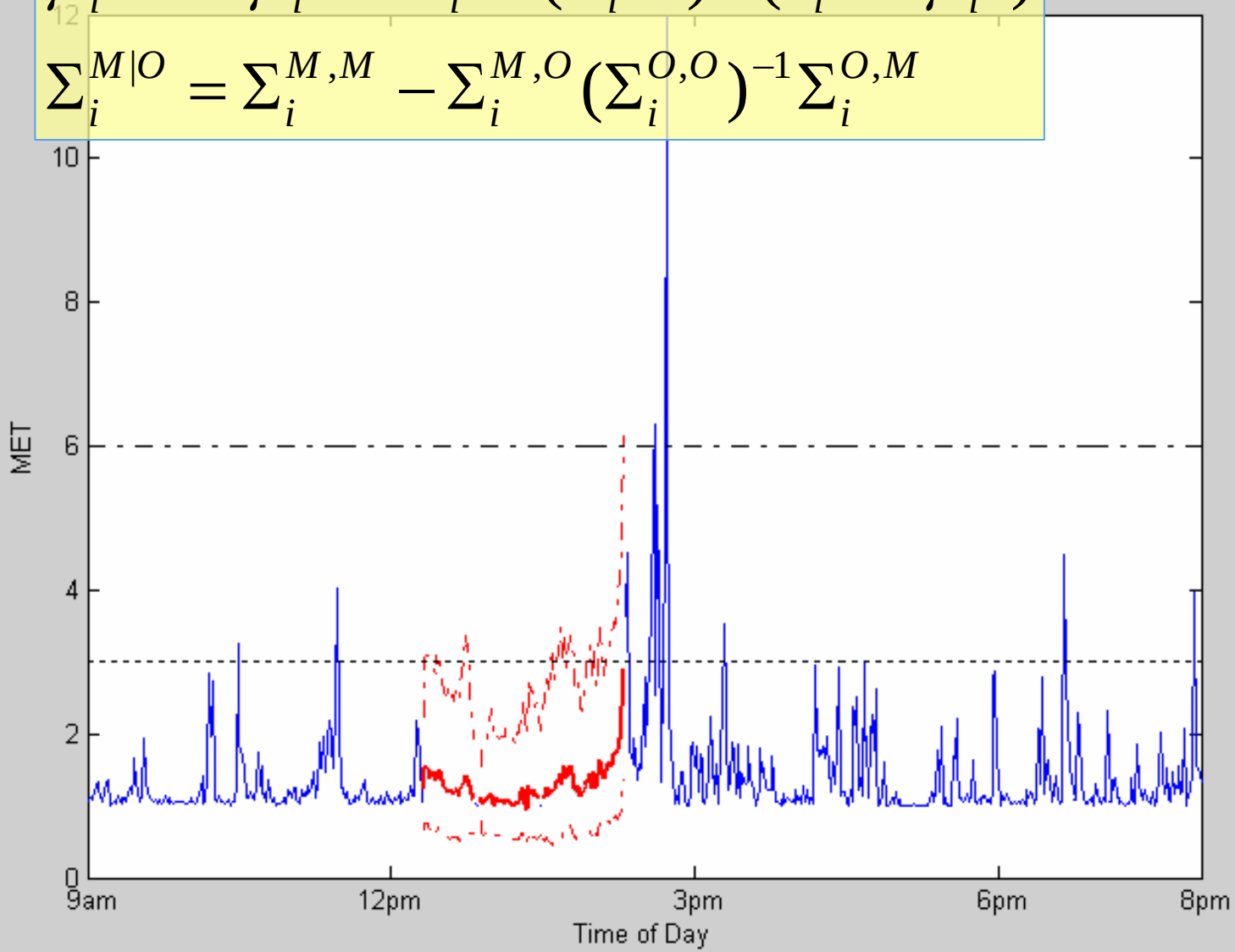
$$\Sigma_i(t_1, t_2) = COV\ \{Y_i(t_1), Y_i(t_2) \mid Y^C\}$$



Imputation distribution

$$\mu_i^{M|O} = \mu_i^M + \Sigma_i^{M,O}(\Sigma_i^{O,O})^{-1}(Y_i^O - \mu_i^O)$$

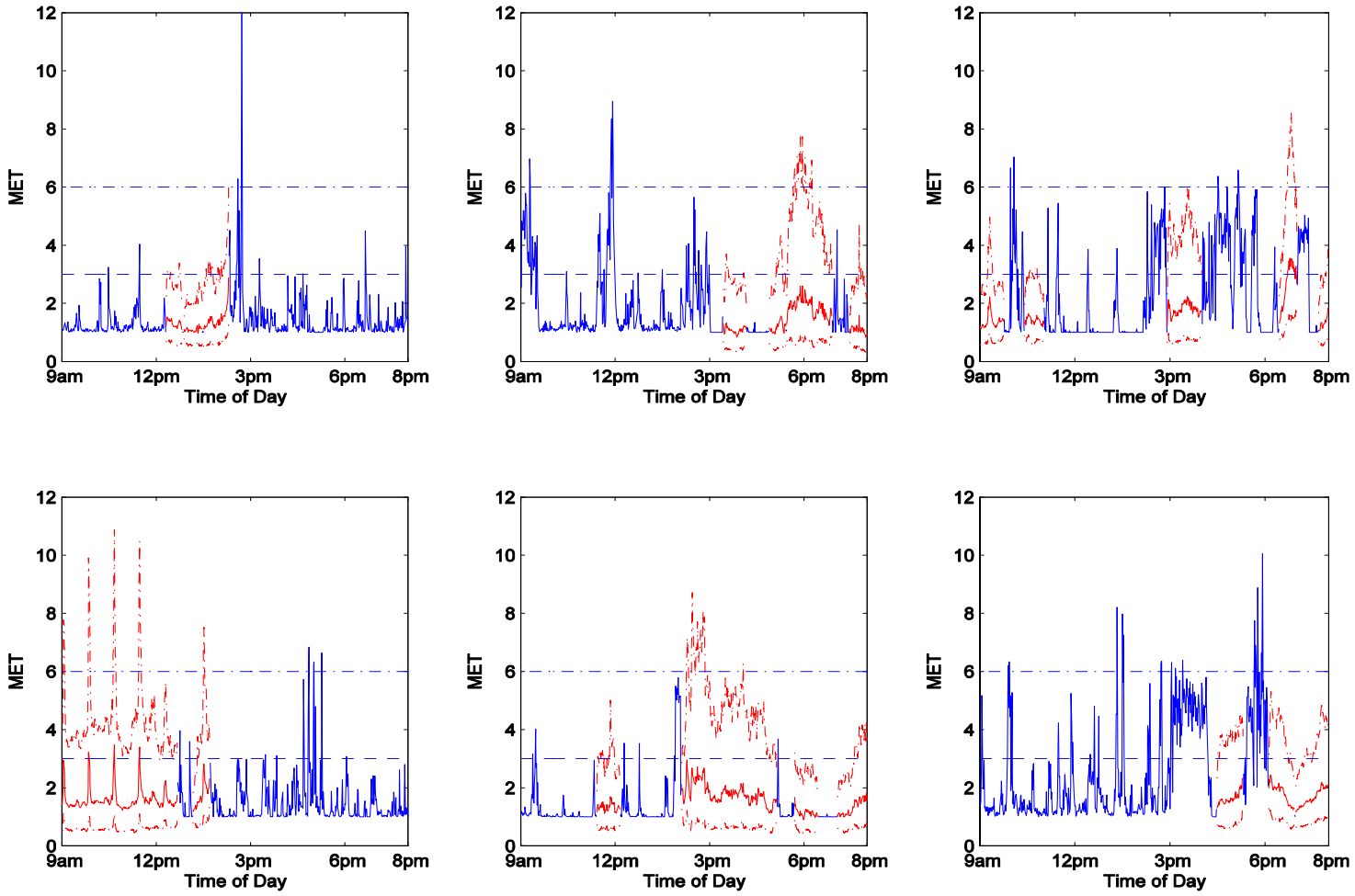
$$\Sigma_i^{M|O} = \Sigma_i^{M,M} - \Sigma_i^{M,O}(\Sigma_i^{O,O})^{-1}\Sigma_i^{O,M}$$



Approach: Incomplete Profiles

1. First fit model to *complete profiles*, get posterior distribution samples for model parameters.
2. Use these to estimate *predictive distributions* for the the incomplete profiles
   - Borrow information about what the curves in these regions look like.
   - Account for child-specific and day-specific covariates.
   - Integrates over uncertainty in parameter estimation
3. Regress missing data on the observed data to obtain *imputation distribution* for missing regions
   - Borrow information from nearby times in incomplete profiles.
   - Makes predictions for missing regions “connected” with observed.

Incomplete Profiles



Approach: Incomplete Profiles

1. First fit model to *complete profiles*, get posterior distribution samples for model parameters.
2. Use these to estimate *predictive distributions* for the the incomplete profiles
   - Borrow information about what the curves in these regions look like.
   - Account for child-specific and day-specific covariates.
   - Integrates over uncertainty in parameter estimation
3. Regress missing data on the observed data to obtain *imputation distribution* for missing regions
   - Borrow information from nearby times in incomplete profiles.
   - Makes predictions for missing regions "connected" with observed.
4. Supplement WFMM with step to stochastically impute values for missing data.
   - Inference appropriately accounts for uncertainty in imputation

Missing Data in the WFMM

- Problem: Imputation distribution in data space, modeling done in wavelet space
- Solution: Project imputation distributions into wavelet space

$$M_i(t) = \begin{cases} Y_i(t) \text{ if } t \text{ observed} \\ \mu_i^{M|O}(t) \text{ otherwise} \end{cases}$$

$$V_i(t_1, t_2) = \begin{cases} 0 \text{ if either } t_1 \text{ or } t_2 \text{ obs.} \\ \Sigma_i^{M|O}(t_1, t_2) \text{ otherwise} \end{cases}$$

$$M_i^* = M_i W'$$

$$V_i^* = W V_i W'$$

- Add step to MCMC whereby "missing" wavelet coefficients $D_{ijk} \sim N(M^*_{ijk}, V^*_{ijk})$

Discussion

- Presented unified modeling approach for FDA
  - Adaptive enough to handle irregularities in both mean structures and random effects (covariances)
- Method based on mixed models; is FLEXIBLE
  - Accommodates a wide range of experimental designs
  - Addresses large number of research questions
- Posterior samples allow Bayesian inference and prediction
  - Posterior credible intervals; pointwise or joint
  - Predictive distributions for future sampled curves
  - Predictive probabilities for classification of new curves
  - Bayesian functional inference can be done via Bayes Factors
- Since a unified modeling approach is used, all sources of variability in the model propagated throughout inference.

Discussion

- Approach is Bayesian. The only informative priors to elicit are regularization parameters, which can be estimated from data using empirical Bayes.
- Developed general-use code – reasonably fast and straightforward to use → minimum information to specify is Y, X, Z matrices.
- Can deal with missing data, i.e. partially observed functions
- Method can be generalized to higher dimensional functions, e.g. image data, space/time data
- The Gaussian/independence assumptions can be relaxed to robustify modeling

Acknowledgements

- **Work presented here is from 3 papers**

1. "*Wavelet-Based Functional Mixed Models*" (2006) Jeffrey S. Morris and Raymond J. Carroll, *JRSS-B*, 68(2): 179-199.
2. "*Using Wavelet-Based Functional Mixed Models to Characterize Population Heterogeneity in Accelerometer Profiles: A Case Study*" (2006) Jeffrey S. Morris, Cassandra Arroyo, Brent Coull, Louise Ryan, Richard Herrick, and Steve Gortmaker, *JASA*, 101(4): 1352-1364.
3. "*Bayesian Analysis of Mass Spectrometry Proteomics Data using Wavelet Based Functional Mixed Models*" (2007) Jeffrey S. Morris, Philip J. Brown, Richard Herrick, Keith A. Baggerly, and Kevin R. Coombes, *Biometrics*, to appear.

- **Supported by NIH Grant R01 CA107304**
- **Computer code/papers on web at** `http://biostatistics.mdanderson.org/Morris/papers.html`

Acknowledgements

Statistical Collaborators

Raymond J. Carroll

Philip J. Brown

Louise M. Ryan

Keith A. Baggerly

Kevin R. Coombes

Marina Vannucci

Brent A. Coull

Elizabeth J. Malloy

Cassandra Arroyo

Veera Baladandayuthapani

Richard C. Herrick **(computing)**

Biomedical Collaborators

Joanne Lupton

Meeyoung Hong

Nancy Turner

Steve Gortmaker

Stanley Hamilton

James Abbruzzesse

Ryuji Kobayashi

John Koomen

Nancy Shih

Howard Gutstein

Josh Fidler

Donghui Li