Bayesian Mixed Models for Functional Data

Jeffrey S. Morris

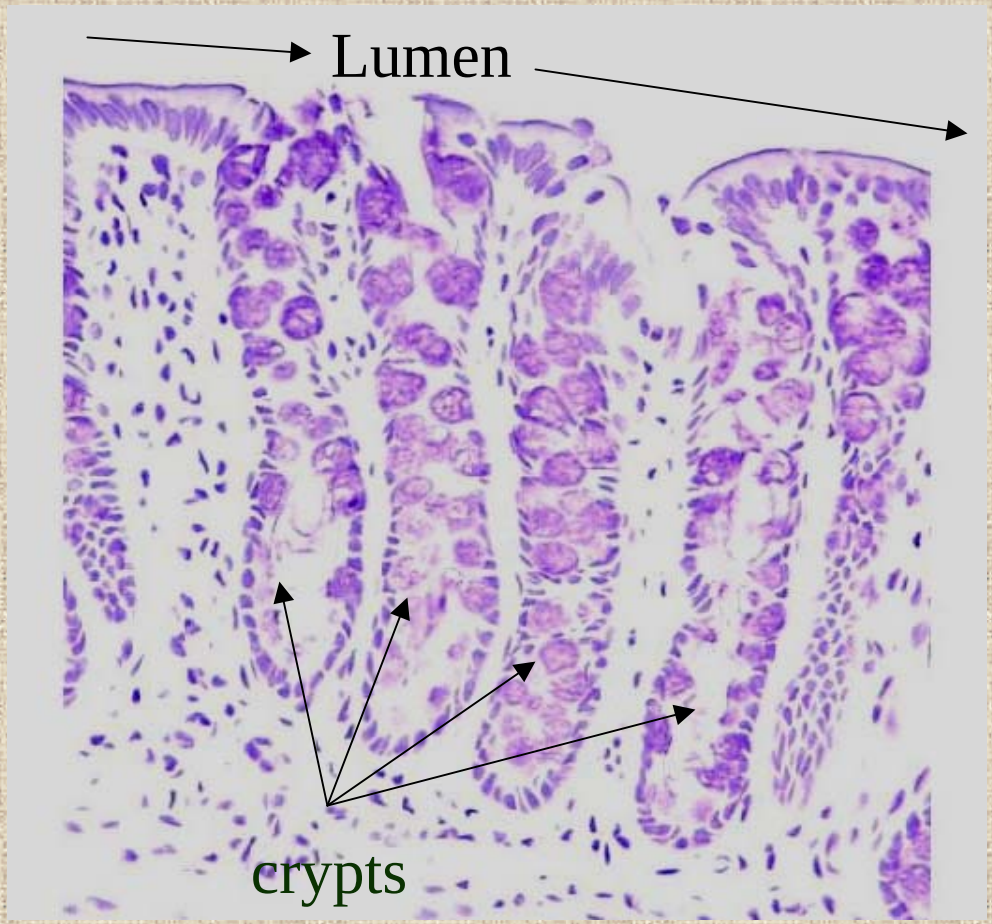
UT MD Anderson Cancer Center

Houston, Texas

Functional Data

- **Functional Data:**
  - **Ideal units of observation: curves**
  - **Observed data: curves sampled on fine grid**
- **Increasingly encountered in biomedical research with new technologies taking automated measurements**
- **Present unique challenges:**
  - **Extremely large data sets (>100s-1000s per curve)**
  - **Curves may be complex and irregular, spatially heterogeneous with many local features**
- **Our focus: Functional regression with functional response and scalar predictors.**

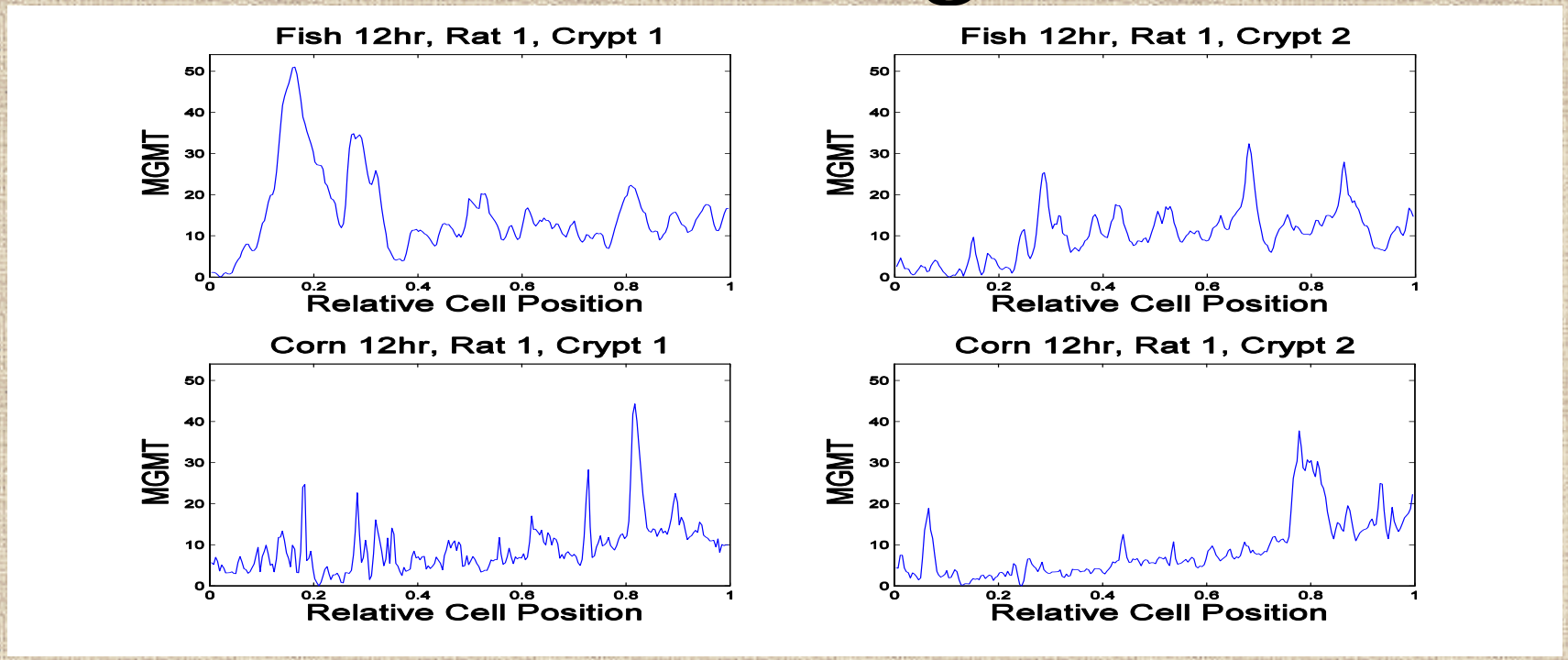
Example: Colon Carcinogenesis



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

$$t \in (0,1)$$

Colon Carcinogenesis Data

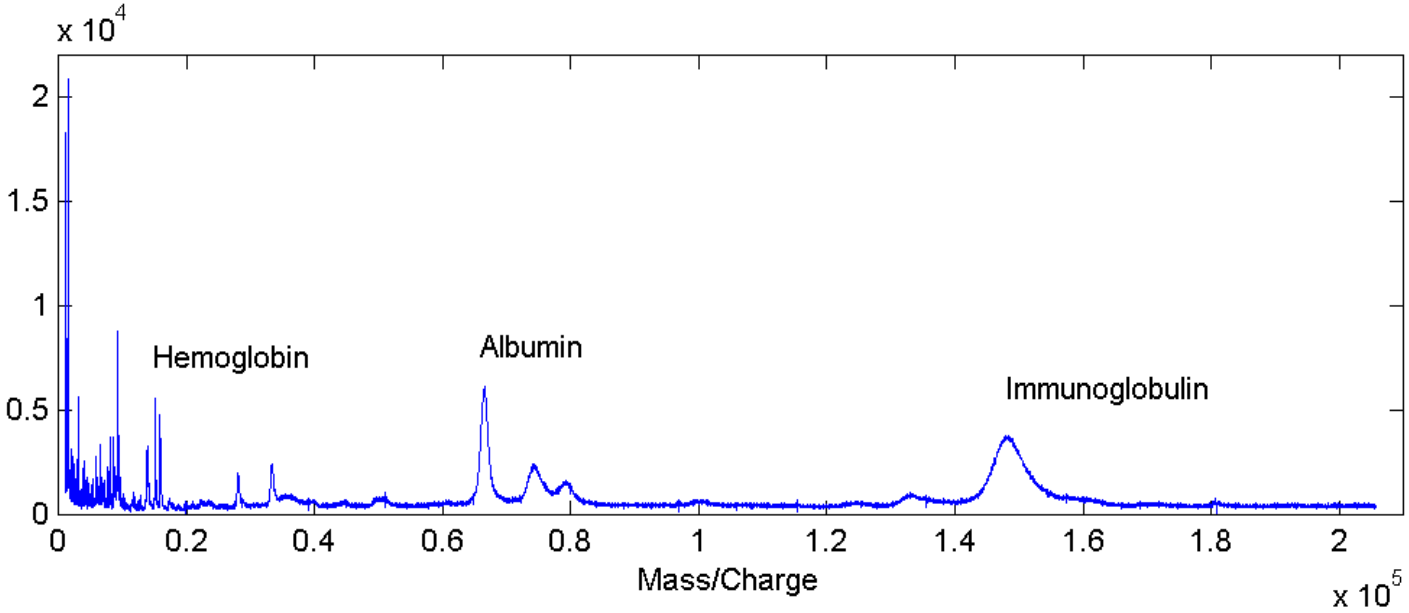


- **30 rats fed 1 of 2 diets, exposed to carcinogen, euthanized after 1 of 5 times after carcinogen exposure (0h,3h,6h,9h,12h)**
- **MGMT measured via IHCS for 15 crypts/rat, ~250 obs/crypt**
- **Diet effect? Vary by time and/or crypt depth? Related to other covariates (DNA adduct/apoptosis)? Relative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

- 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: Mouse proteomics study

- **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:**

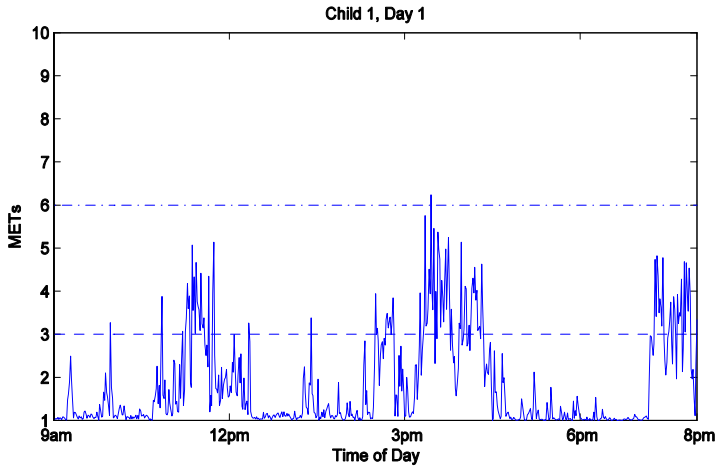
Requires us to include in modeling:

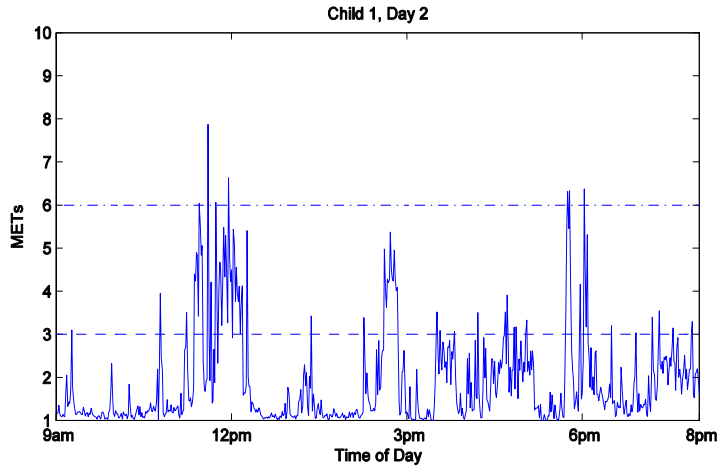
- **Functional laser intensity effect**
- **Random effect functions to account for correlation between spectra from same mouse**

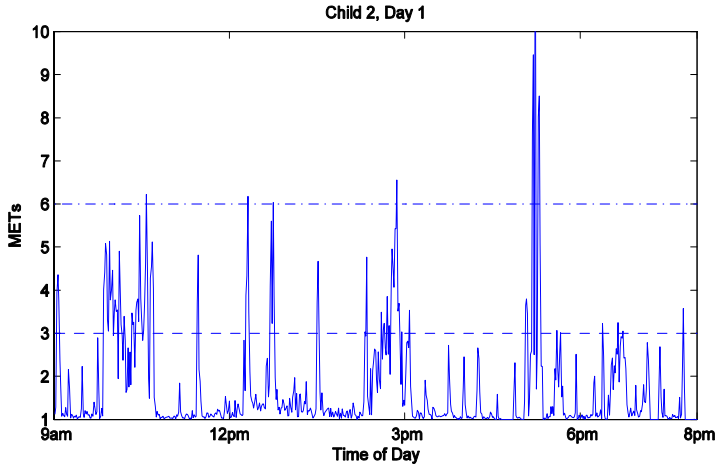
Example: Accelerometer Data

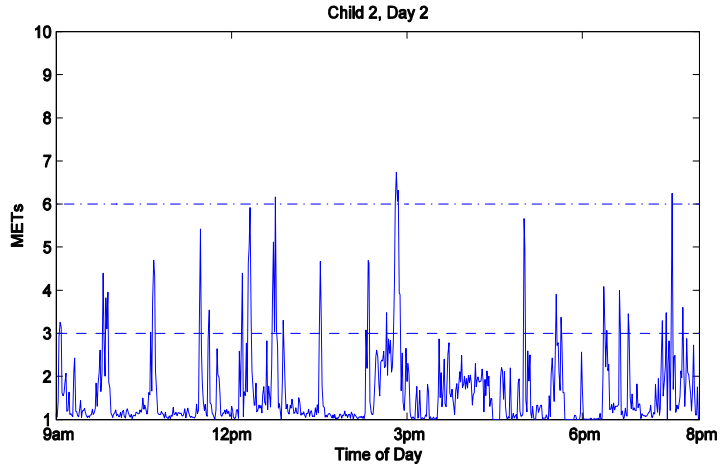
- **Accelerometers:** small motion sensors that digitally record minute-by-minute activity levels
  - Increasingly used in surveillance and intervention studies
- **TriTrac-R3D:** sensor worn on hip
  - Minute-by-minute record of motion in 3 planes
  - Condensed into single activity level measurement/minute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

Accelerometer Data





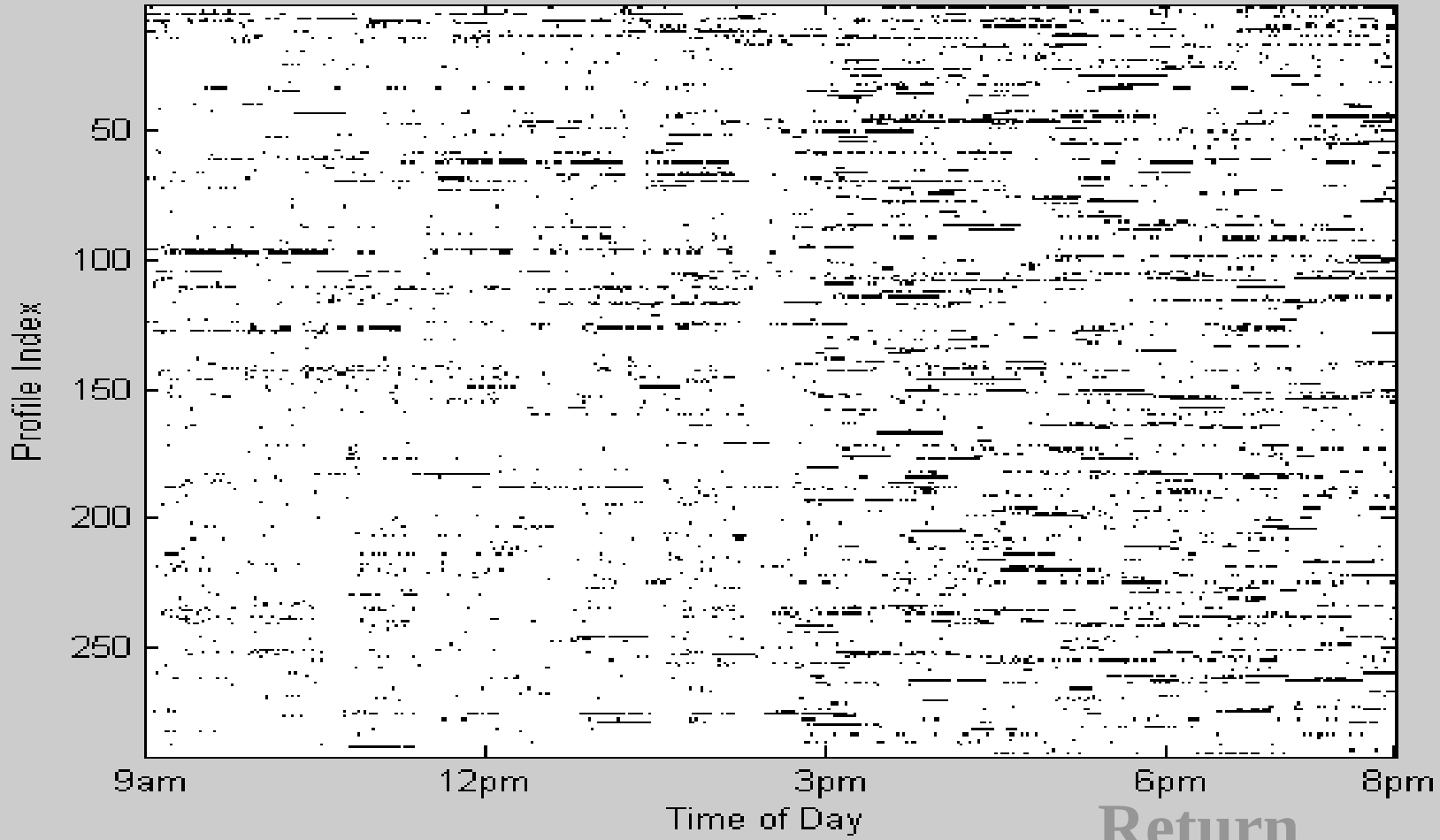




Accelerometer Data

- **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)



Return

Linear Mixed Models

Linear Mixed Model (Laird and Ware, 1982):

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

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

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

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

Functional Mixed Model (FMM)

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

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

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

- **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 $\text{Cov}\{U_i(t_1), U_j(t_2)\} = P_{ij} * Q(t_1,t_2)$**

- **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: SELDI Example

Let $Y_i(t)$ be the SELDI 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, …, 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*~MN(*P*,*Q*) implies Cov{U_{ij}, $U_{i'j'}$}=$P_{ii'}$*$Q_{jj'}$**
- ***P* and *R* are covariance matrices (*m* × *m* & *N* × *N*)**
- ***Q* and *S* are within-curve covariance matrices (*T* × *T*)**

Functional Mixed Models

- **Key feature of FMM: Does not require specification of parametric form for curves**
- **Kernels/fixed-knot splines may not work well for spatially heterogeneous or irregular functional data – inherent smoothness assumptions attenuate local features**
- **Wavelet Regression: nonparametric regression technique that better preserves local features present in the curves.**

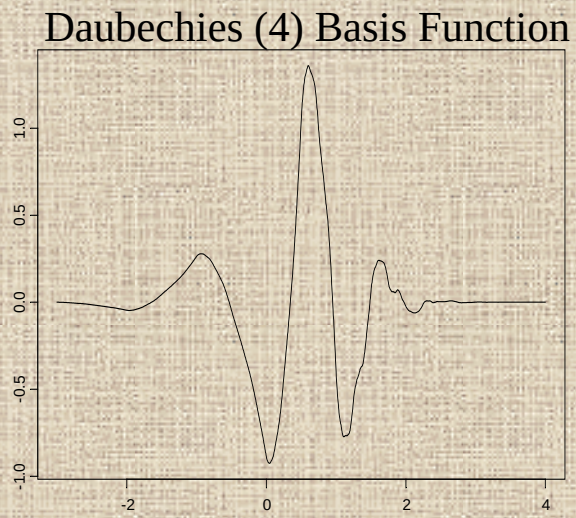
Introduction to Wavelets

- **Wavelets: families of orthonormal basis functions**

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

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

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

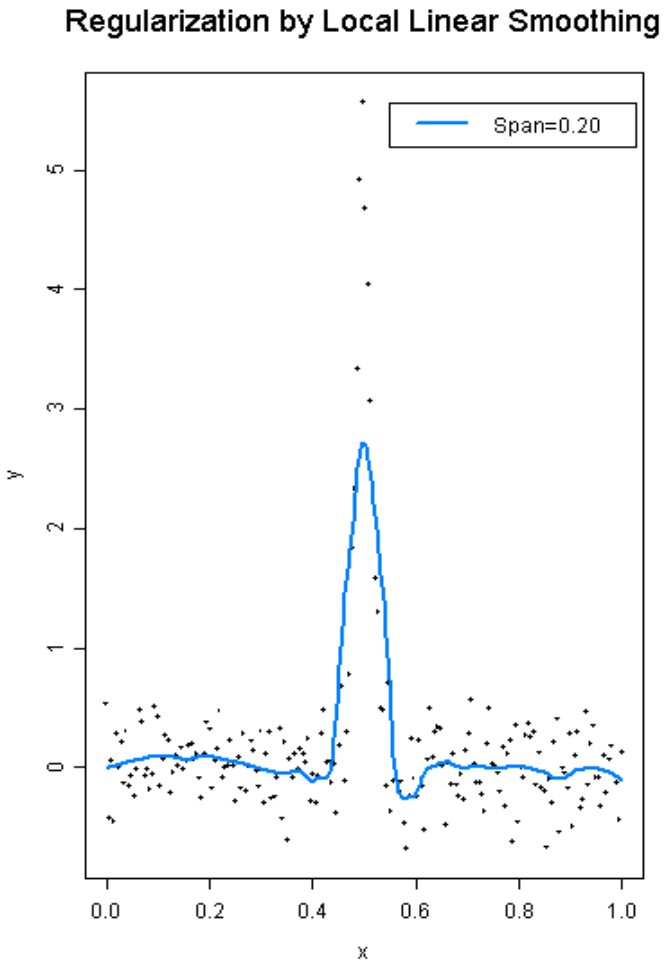


- **Discrete Wavelet Transform (DWT): fast algorithm {O(*T*)} for obtaining *T* empirical wavelet coefficients for curves sampled on equally-spaced grid of length *T*.**
- **Linear Representation: $d = y\,W'$**
  - ***W'* =*T*-by-*T* orthogonal projection matrix**
- **Inverse DWT (IDWT): $y = d\,W$**

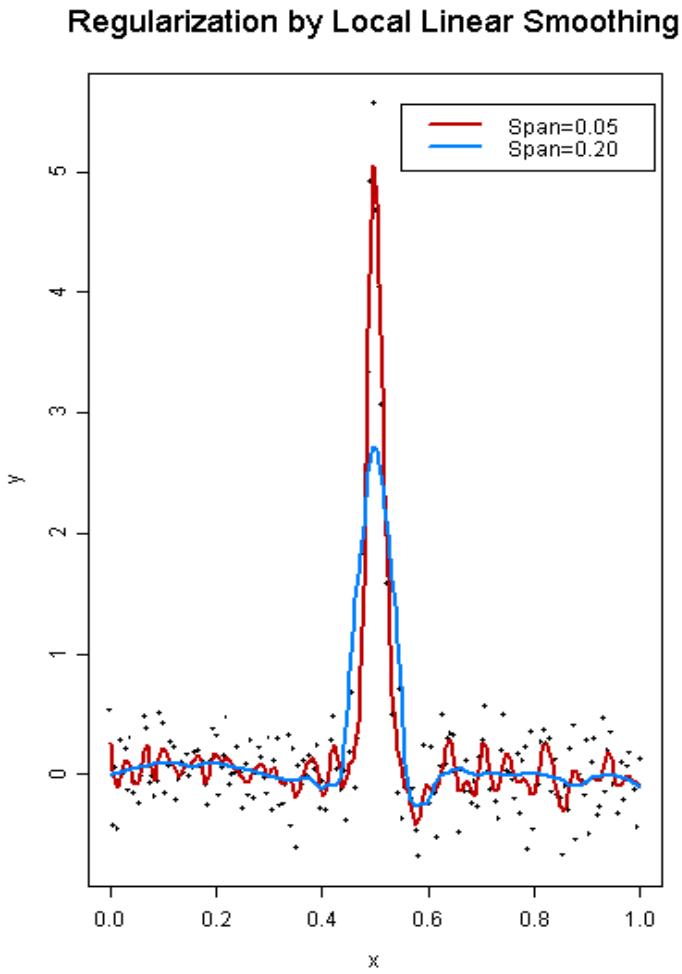
Wavelet Regression

- **Useful properties of wavelets:**
  - **Whitening property**
  - **Compact support**
  - **Parsimonious representation**
- **Wavelet Regression – 3 step process**
  1. **Project data into wavelet space**
  2. **Threshold/shrink coefficients**
  3. **Project back to data space**
- **Yields *adaptively regularized* nonparametric estimates**

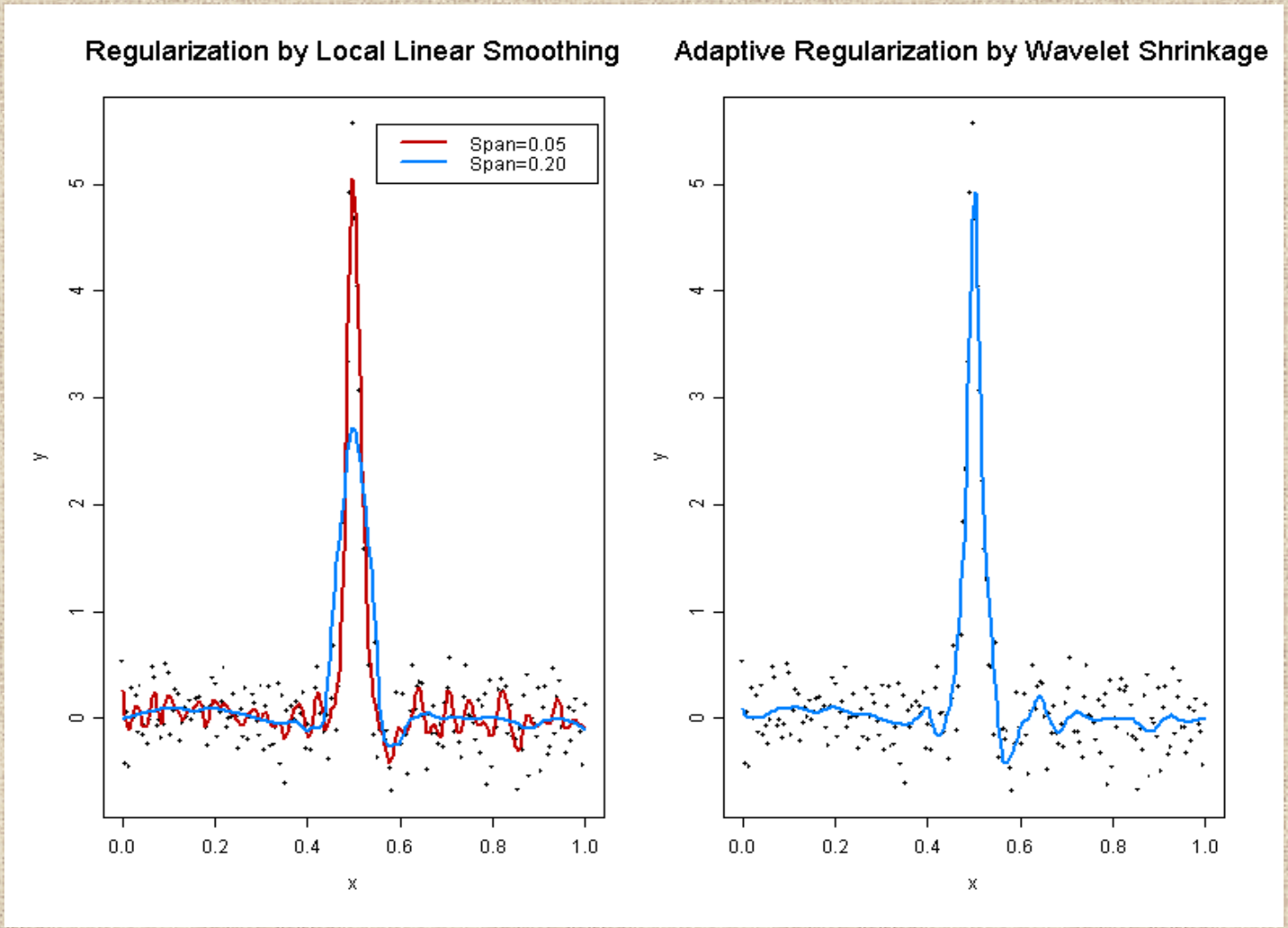
Adaptive Regularization



Adaptive Regularization



Adaptive Regularization



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-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

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*

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

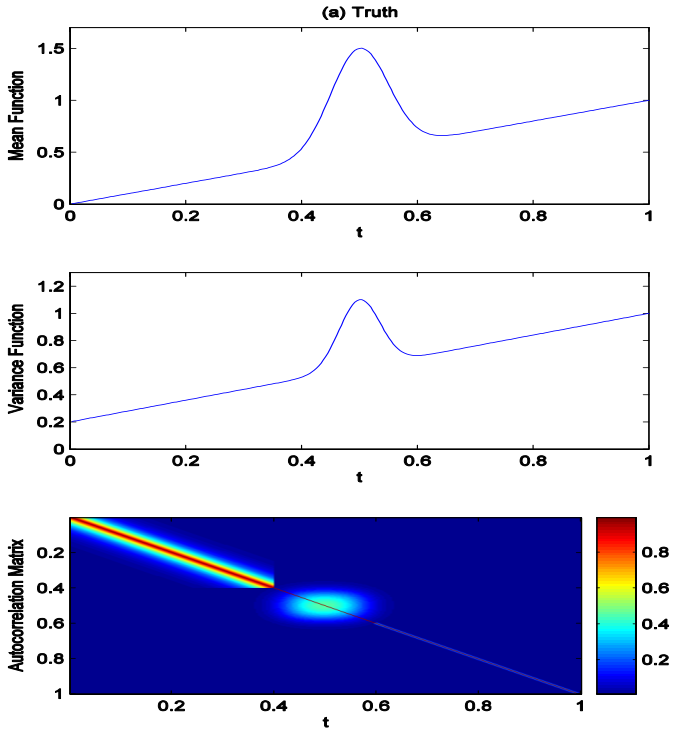
$$U^* \sim MN(P, Q^*)$$
$$E^* \sim MN(R, S^*)$$

- ***B*=BW'*** **&** ***U*=UW'*** **: Rows contain wavelet coefficients for the fixed and random effect functions,**
- ***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

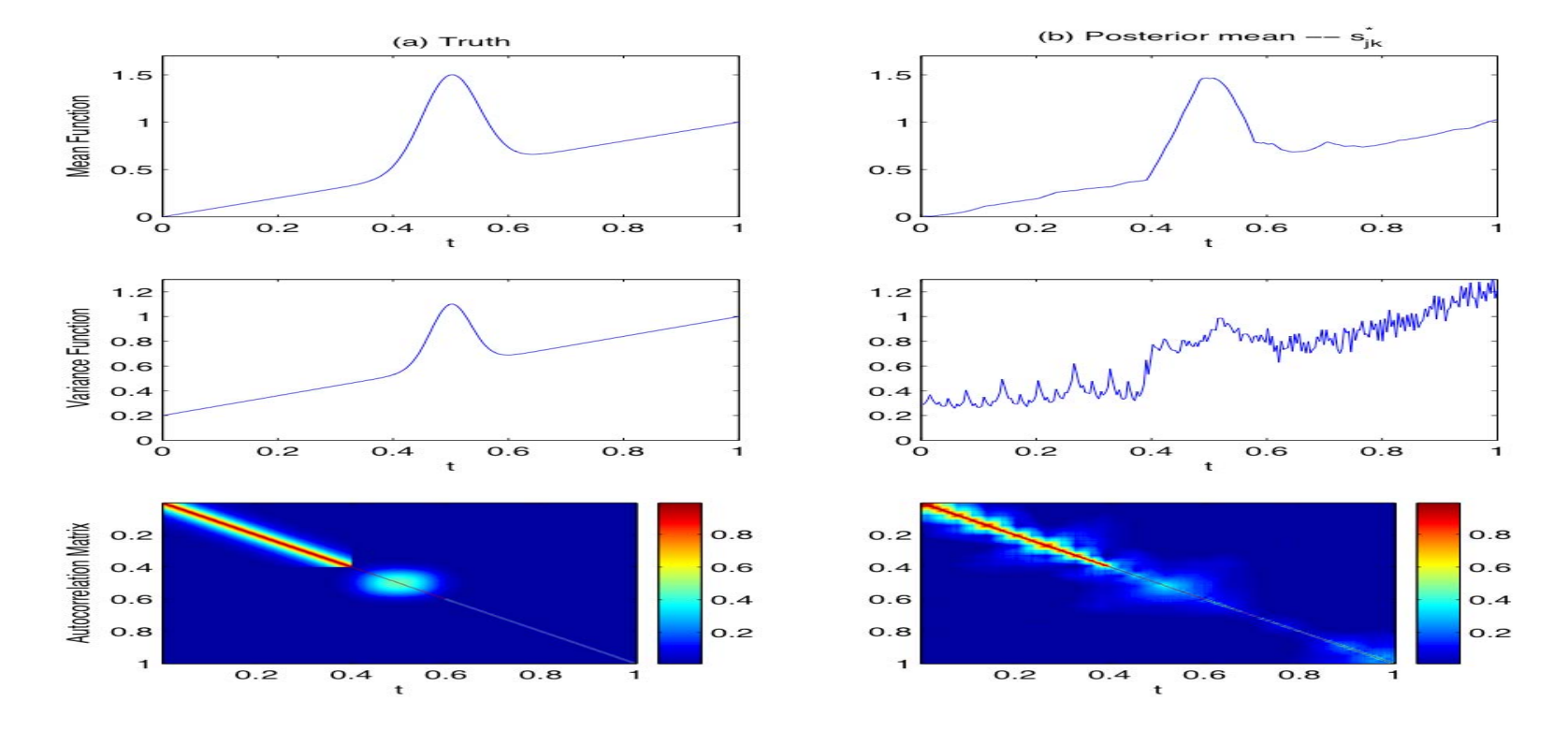
- **We choose parametric structures for P and R to model the covariance structure between the curves.**
  - **Based on the experimental design**
  - **As in linear mixed models.**
- **We assume the between-wavelet covariance matrices Q^* and S^* are diagonal ($Q^*=\text{diag}\{q_{jk}\}$, $S^*=\text{diag}\{s_{jk}\}$).**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

Simulation: Covariance Structure



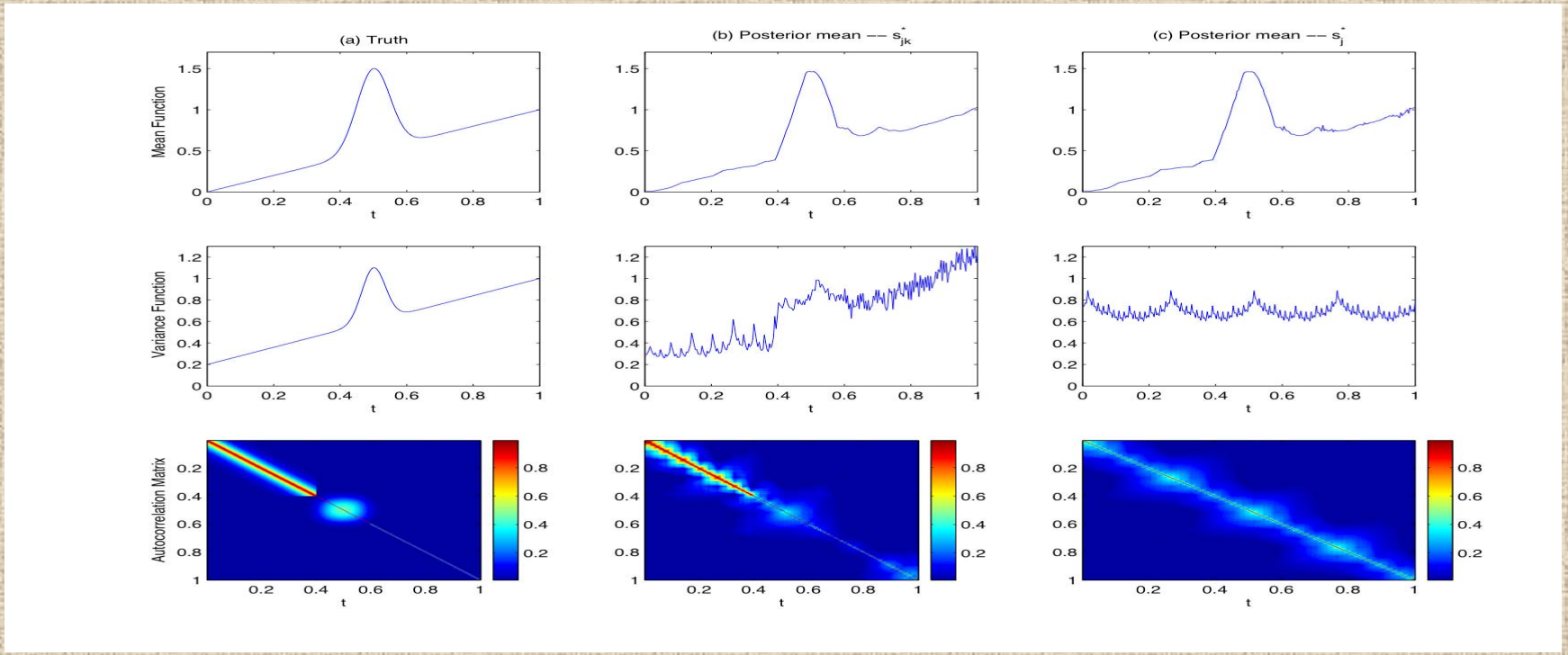
- **True mean: line plus peak**
- **True variance: increasing in t, with extra var at peak**
- **True autocorrelation: Strong autocorrelation (0.9) at left, weak autocorrelation (0.1) right, extra at peak**

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) 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 VC by (*j*,*k*)**

Prior Assumptions

Mixture prior on β_{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**
  - **Can be elicited, or estimated from the data using empirical Bayes approach (extend Clyde&George 1999 to FMM)**

Model Fitting

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

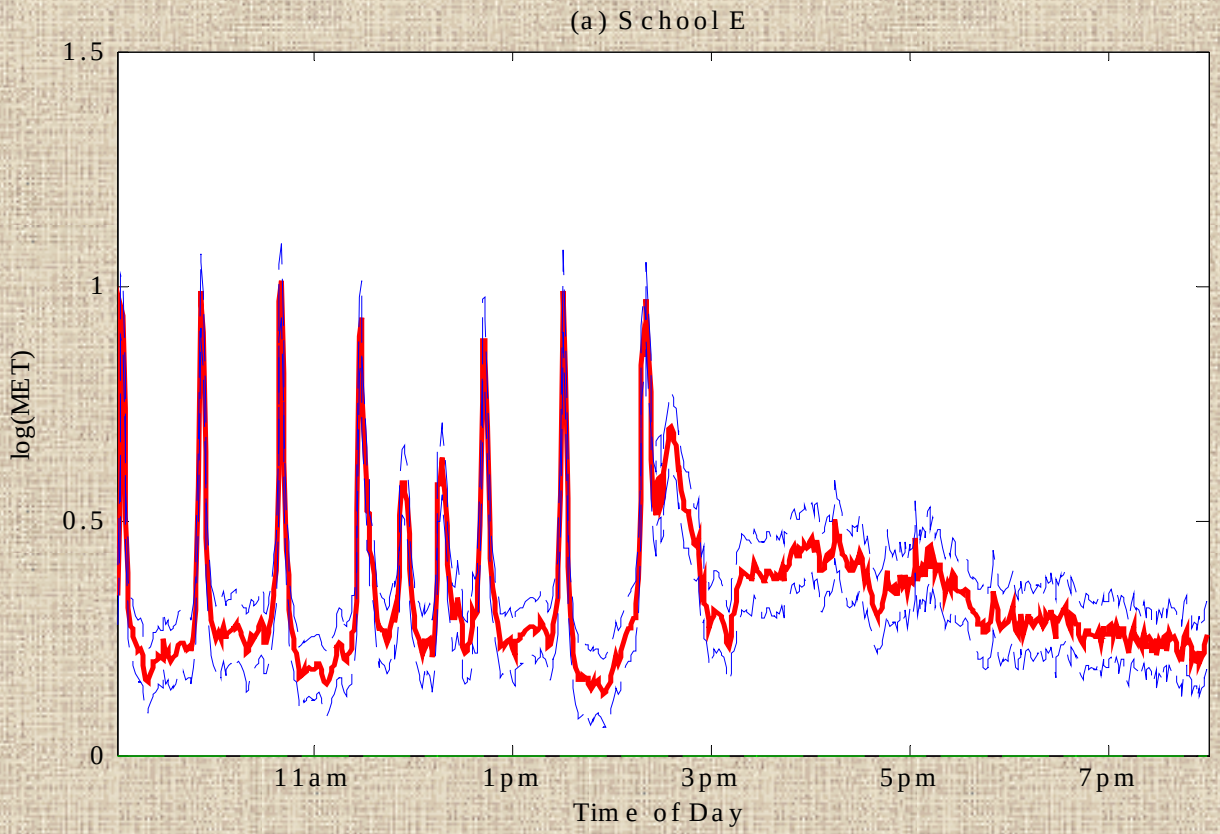
Wavelet-Based FMM

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

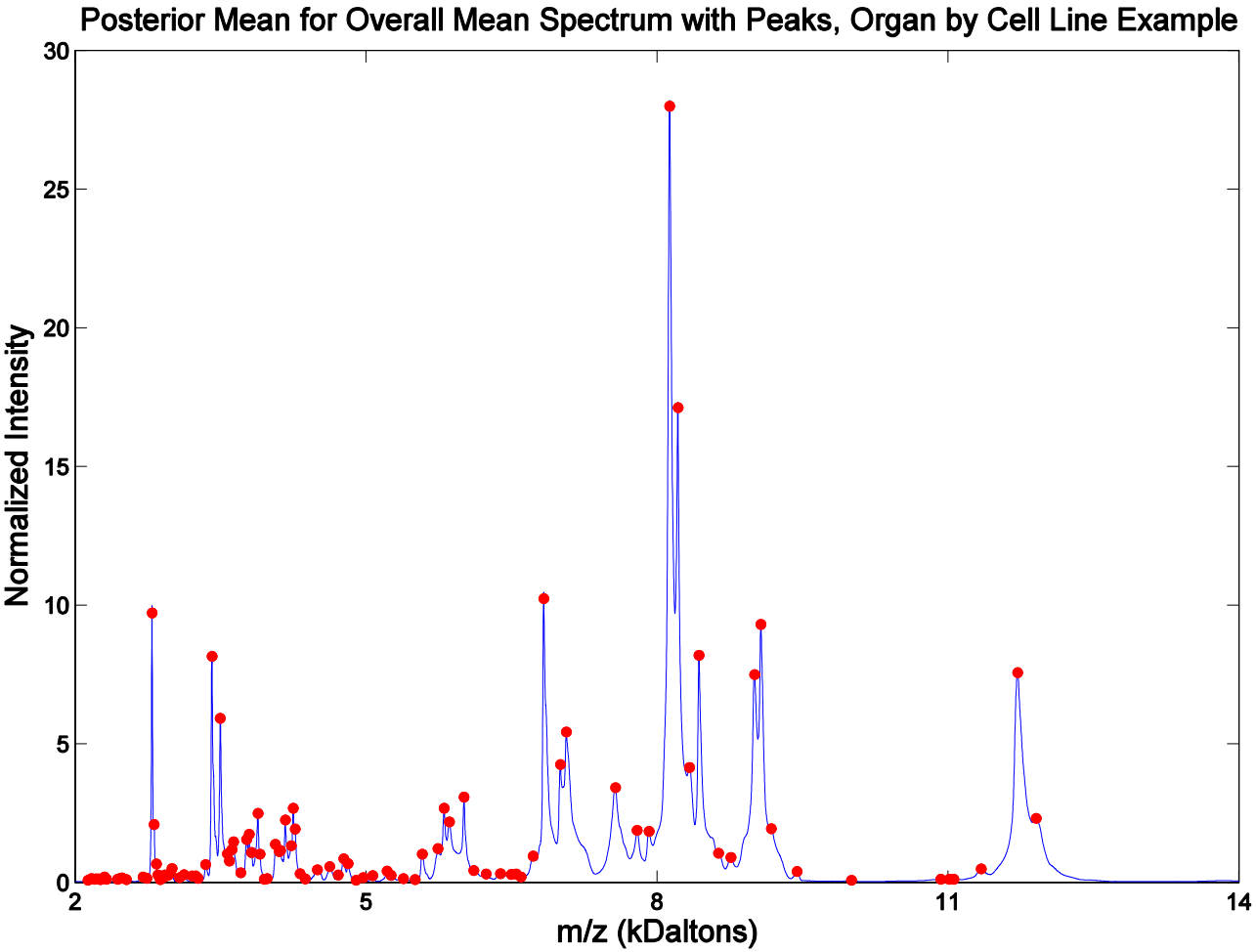
Adaptive Regularization

- **Posterior samples/estimates of fixed effect functions $B_i(t)$ *adaptively regularized* from shrinkage prior**



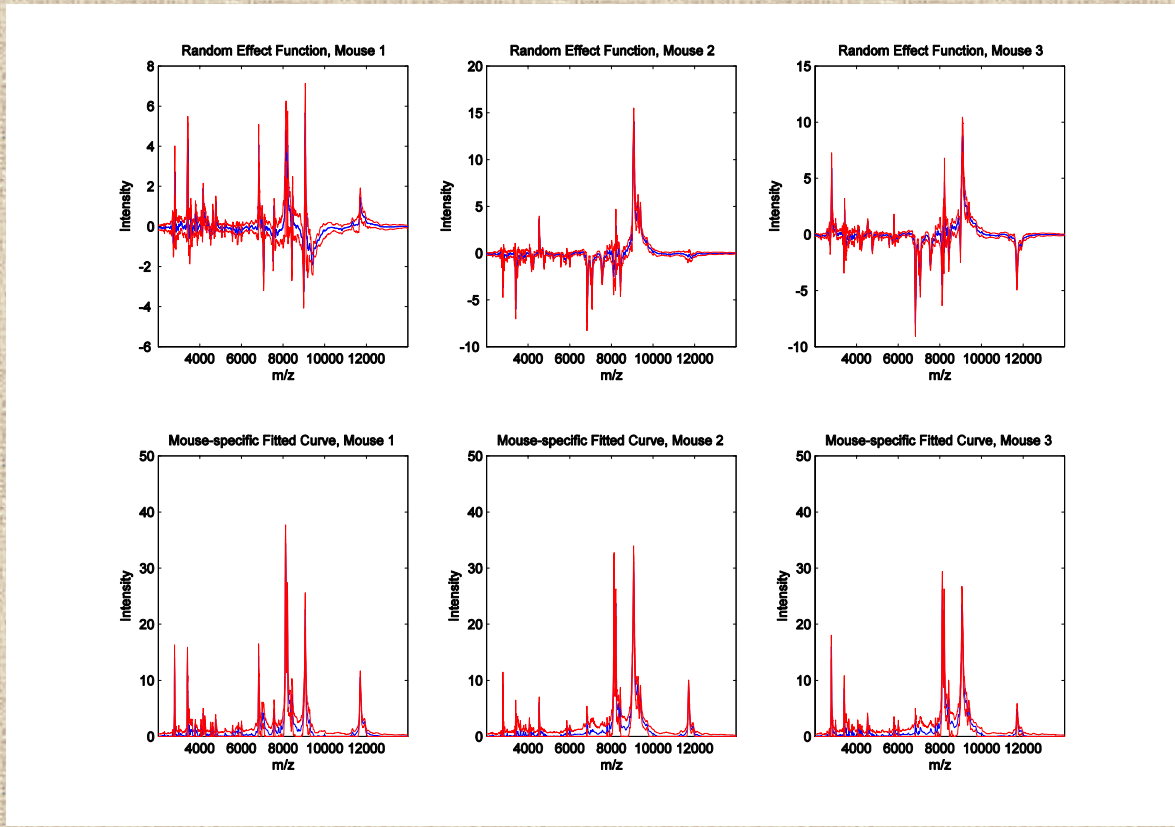
- **Able to preserve dominant spikes in mean curves, if present**

Adaptive Regularization



Adaptive Regularization

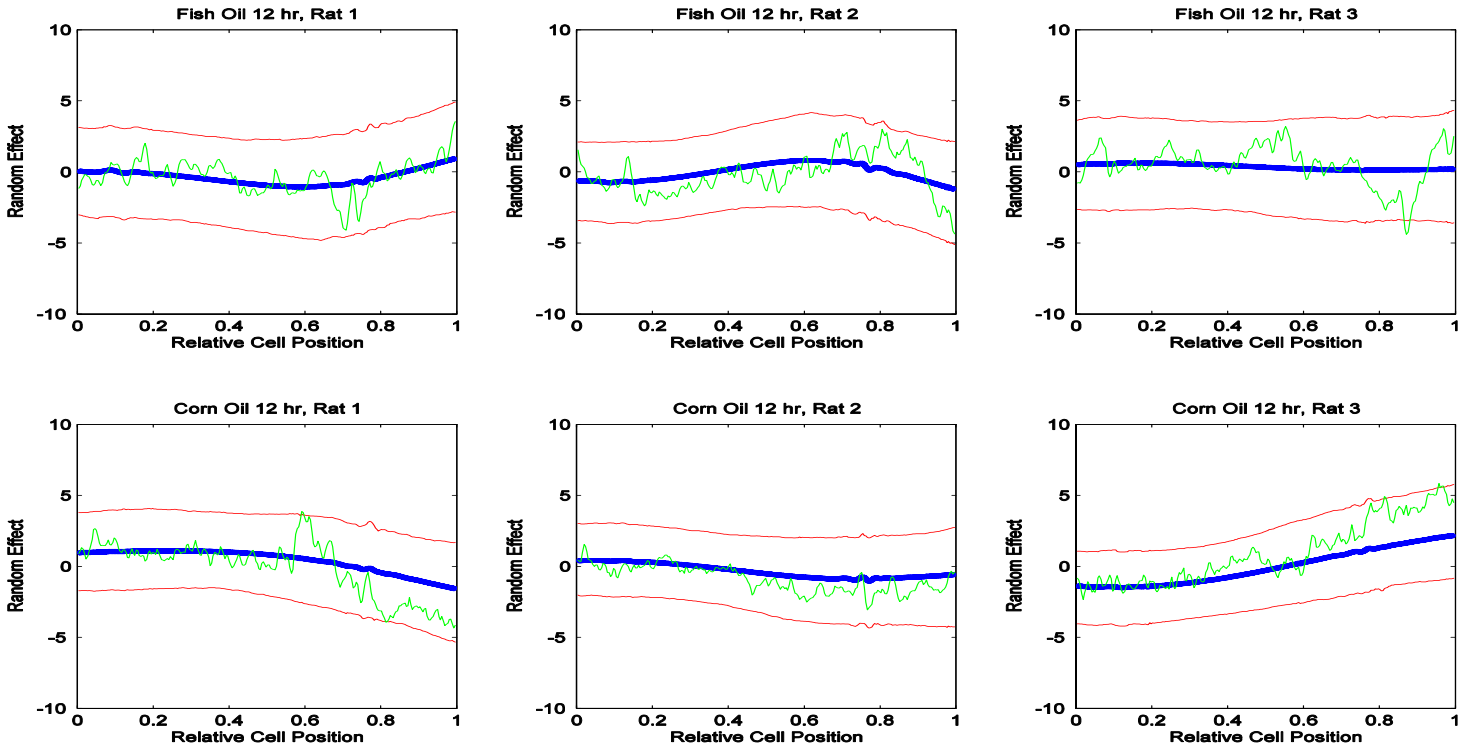
- **Posterior samples/estimates of random effect functions $U_i(t)$ are also *adaptively regularized* from Gaussian prior, since each wavelet coefficient has its own random effect & residual variance**



- **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 any desired Bayesian inference or prediction:

1. **Pointwise posterior credible intervals for funct. 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

   $p_{il}=\Pr\{|B_j(t_l)|>\delta|\mathbf{Y}\}$ for $l=1, \ldots, T$

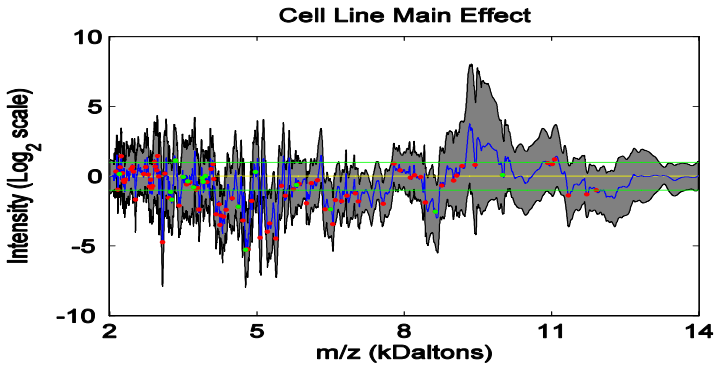
2. Sort peaks in descending order of p_{il} $\{p_{i(l)}, l=1, \ldots, T\}$

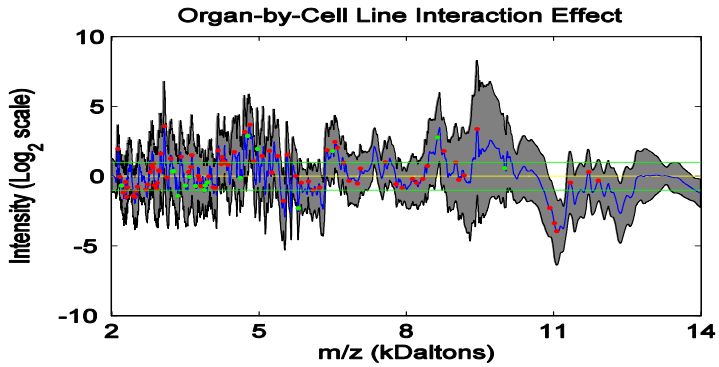
3. Identify cutpoint φ_α on posterior probabilities that controls expected Bayesian FDR to be $\leq \alpha$

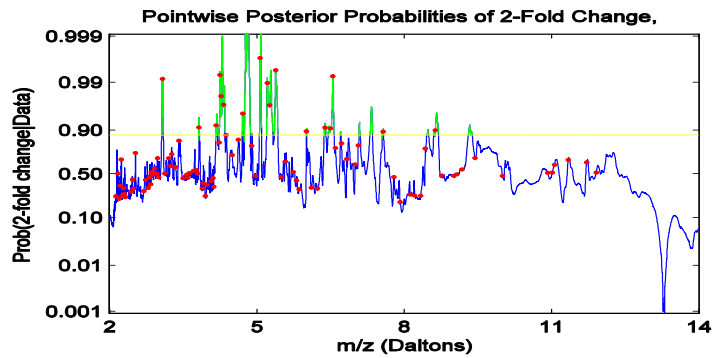
   $\varphi_\alpha=p_{i(\lambda)}$, where $\lambda = \max\left[l^* : \sum_{l=1}^{l^*} \{1-p_{i(l)}\} \leq \alpha \right]$

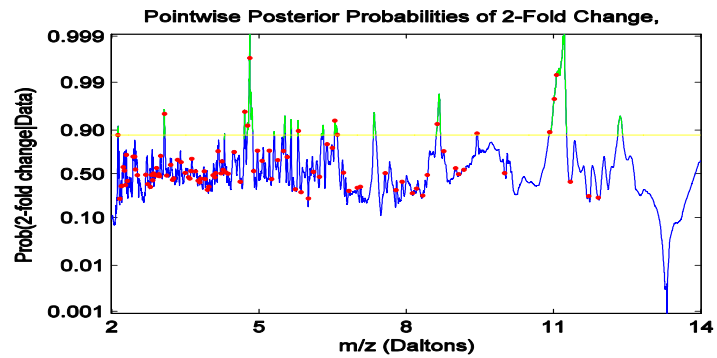
4. Flag the set of locations $\{t_l : p_{il} \leq \varphi_\alpha\}$ as significant (According to model, expect only α to be false pos.)

Results: SELDI Example



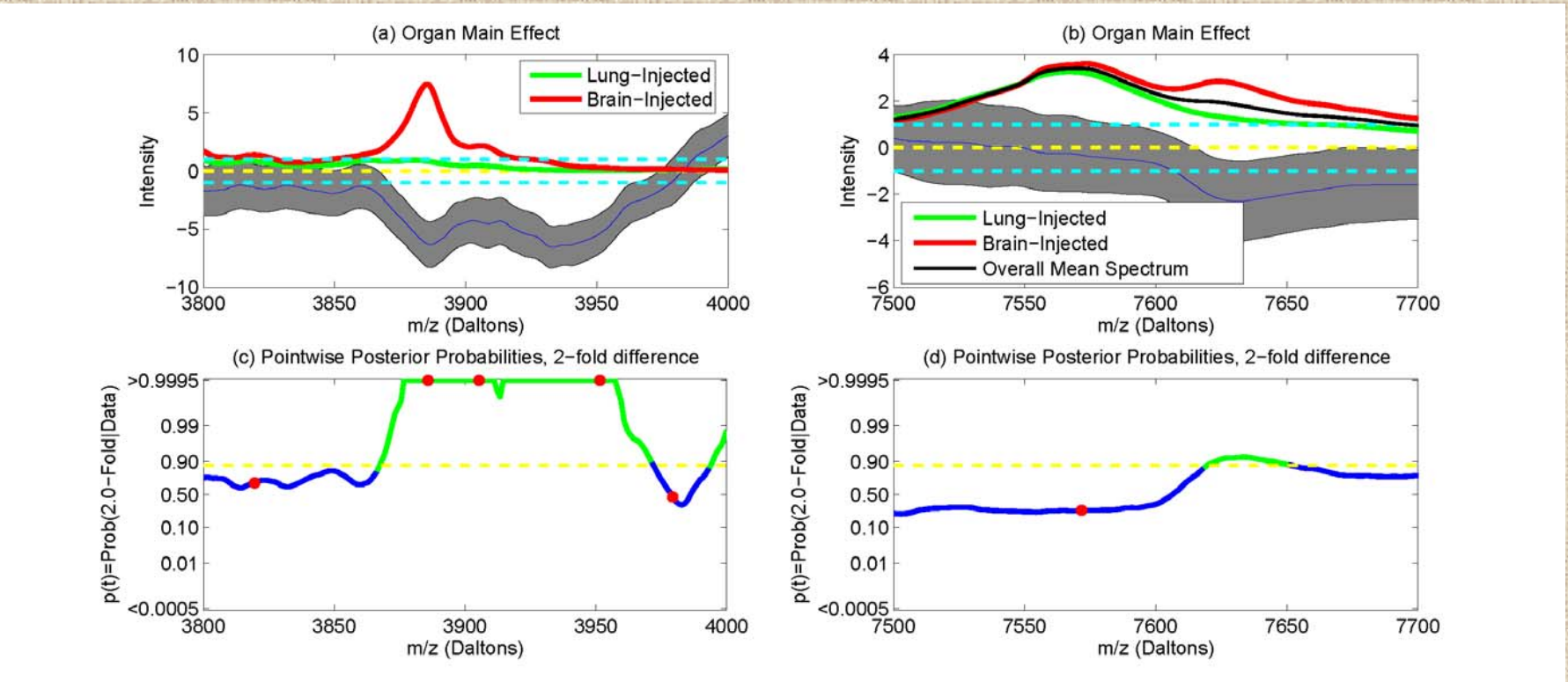






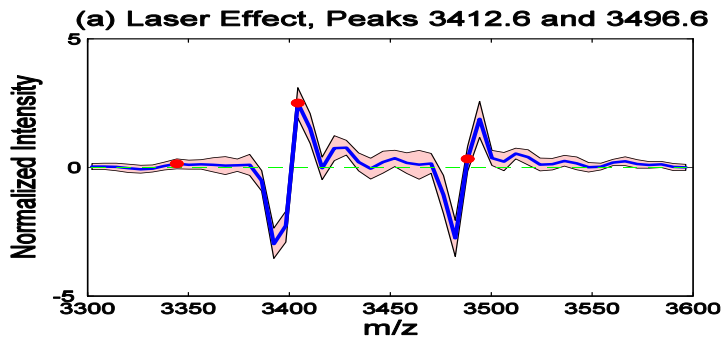
- **Using α=0.05, δ=1 (2-fold expression on $\log_2$ scale), we flag a number of spectral regions.**

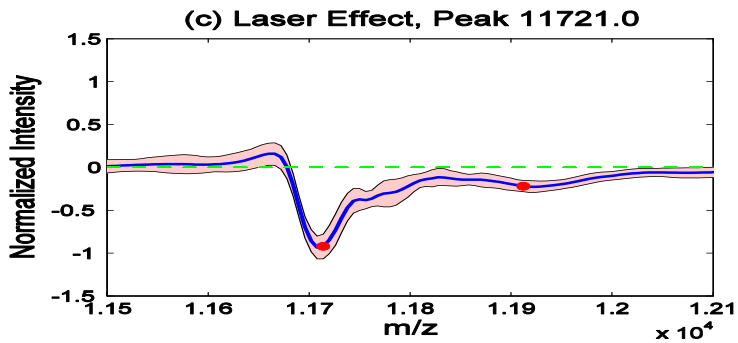
Results: SELDI Example

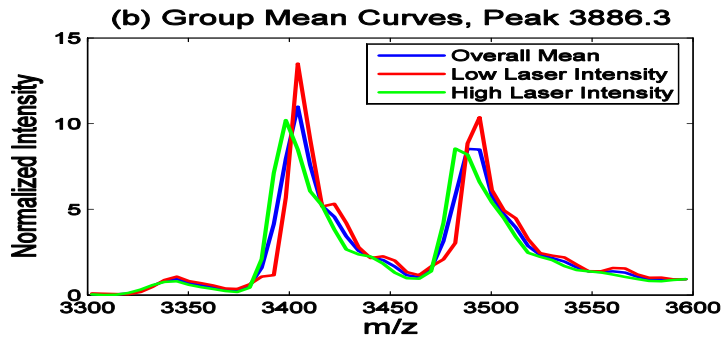


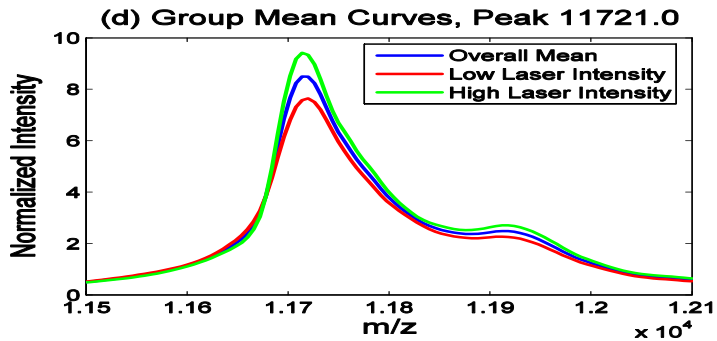
- **3900 D (CGRP-II): dilates blood vessels in brain**
- **7620 D (nerogranin): active in synaptic modeling in brain (Not detected as peak)**

Results: SELDI Example



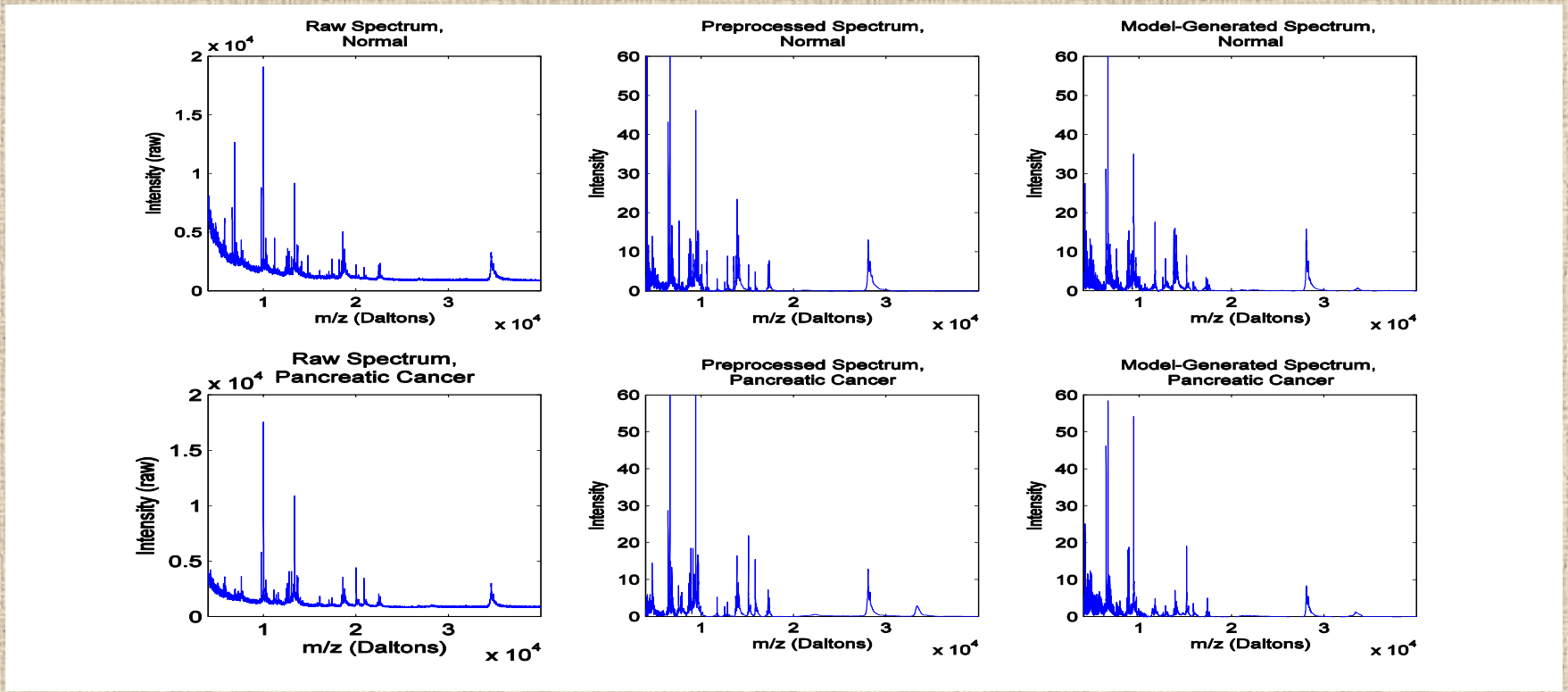






- **Inclusion of nonparametric functional laser intensity effect is able to adjust for systematic differences in the *x* and *y* axes between laser intensity scans**

Results: SELDI Example



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

Incomplete Profiles

- **Lots of missing data** **(Missing Data)**

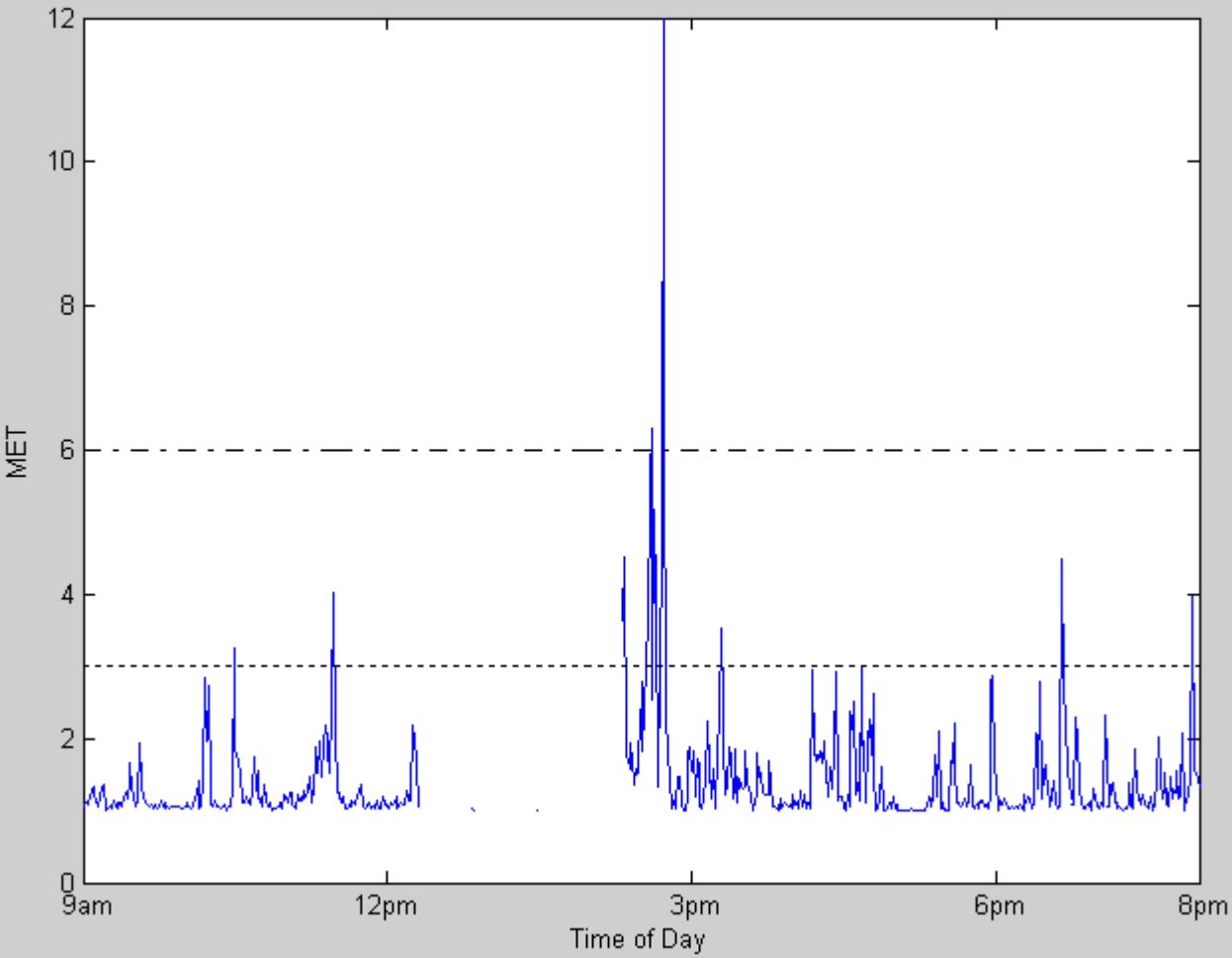
  Example of *incomplete profile*

- **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 (fig)
   - Borrow information about what the curves in these regions look like.
   - Account for child-specific and day-specific covariates.
3. Regress missing data on the observed data to obtain *imputation distribution* for missing regions (fig)
   - 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

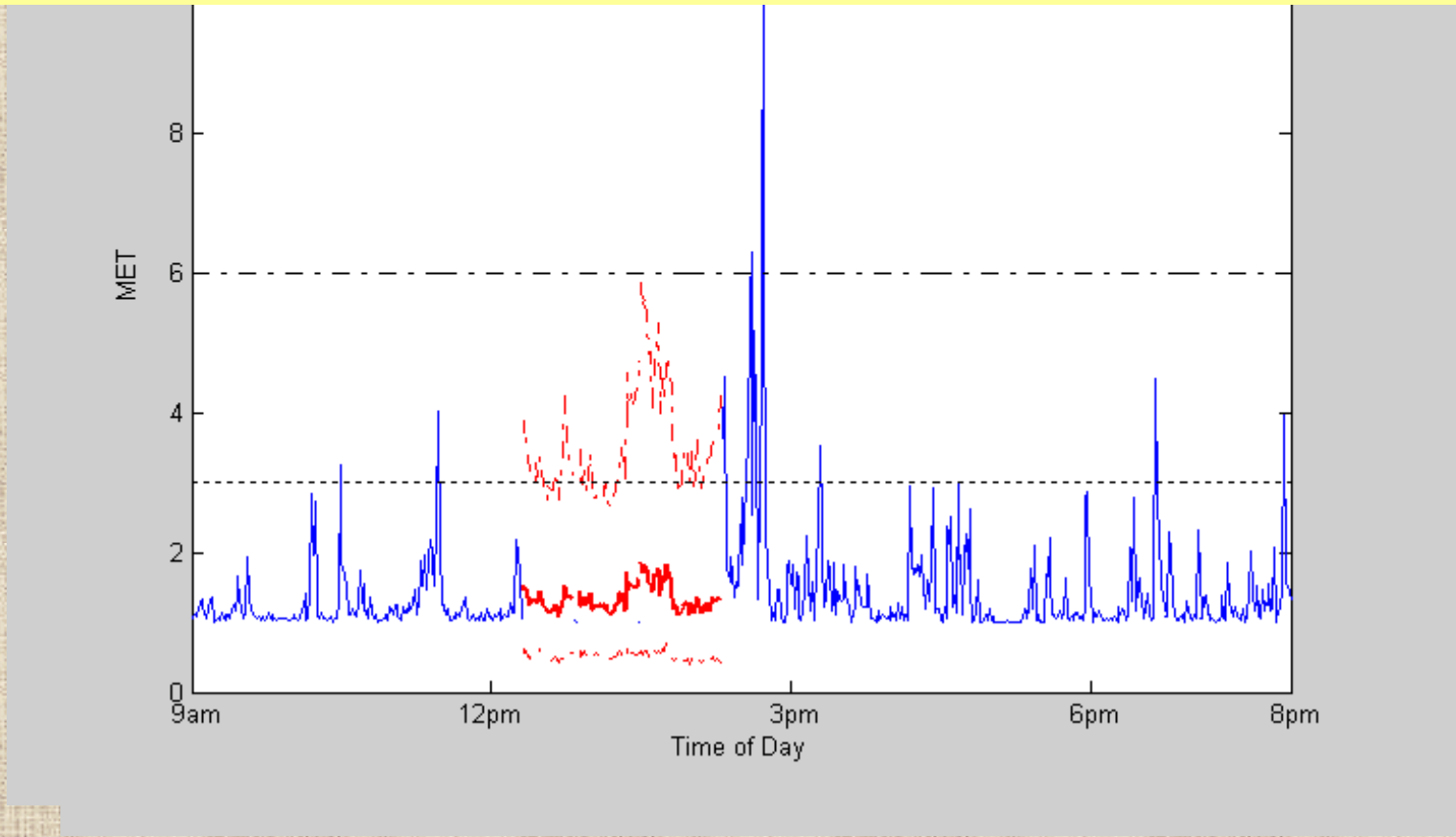
Incomplete Profile



Predictive Distribution

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

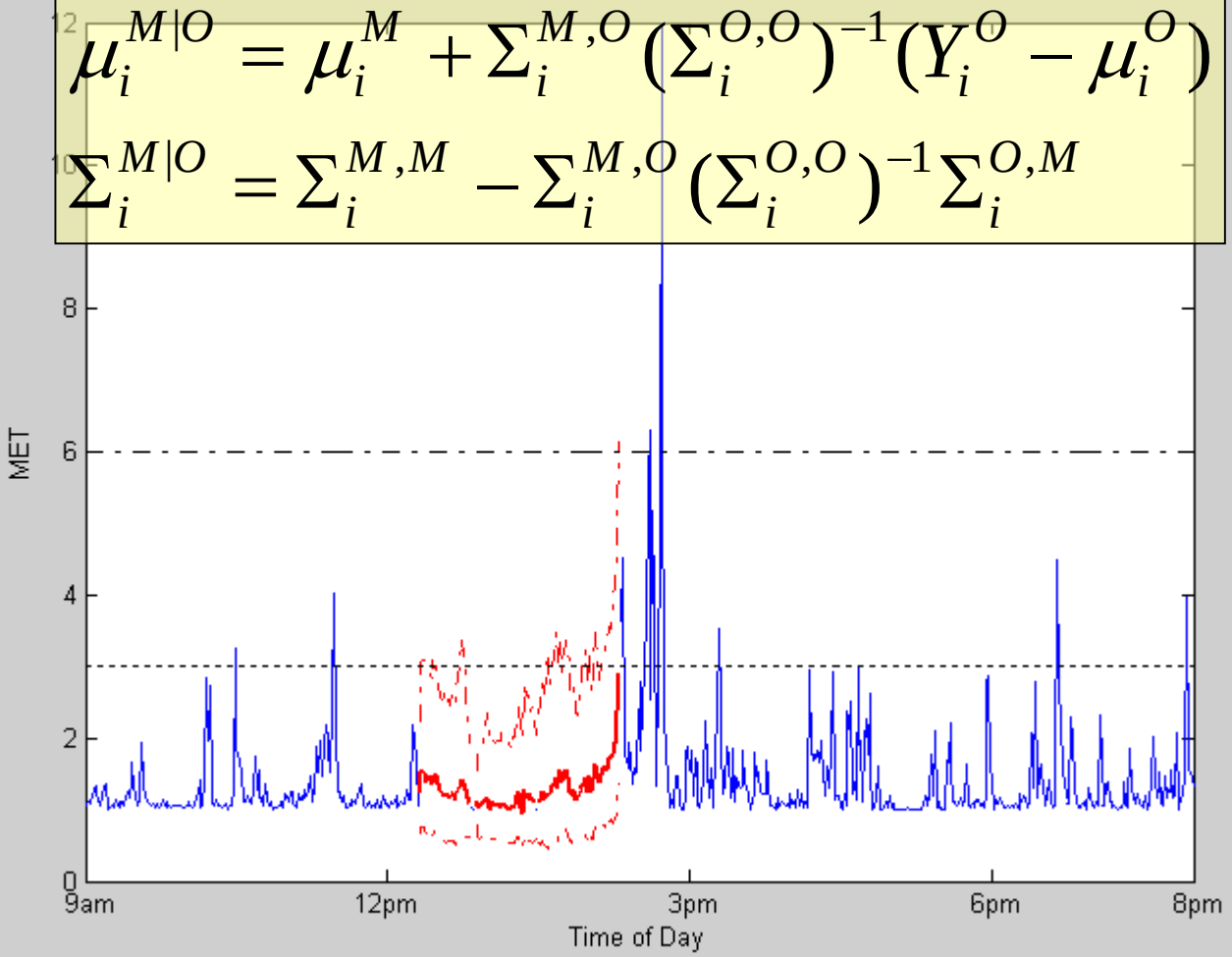
$$\Sigma_i(t_1, t_2) = COV\ \{Y_i(t_1), Y_i(t_2) \mid Y^C\}$$



Imputation distribution

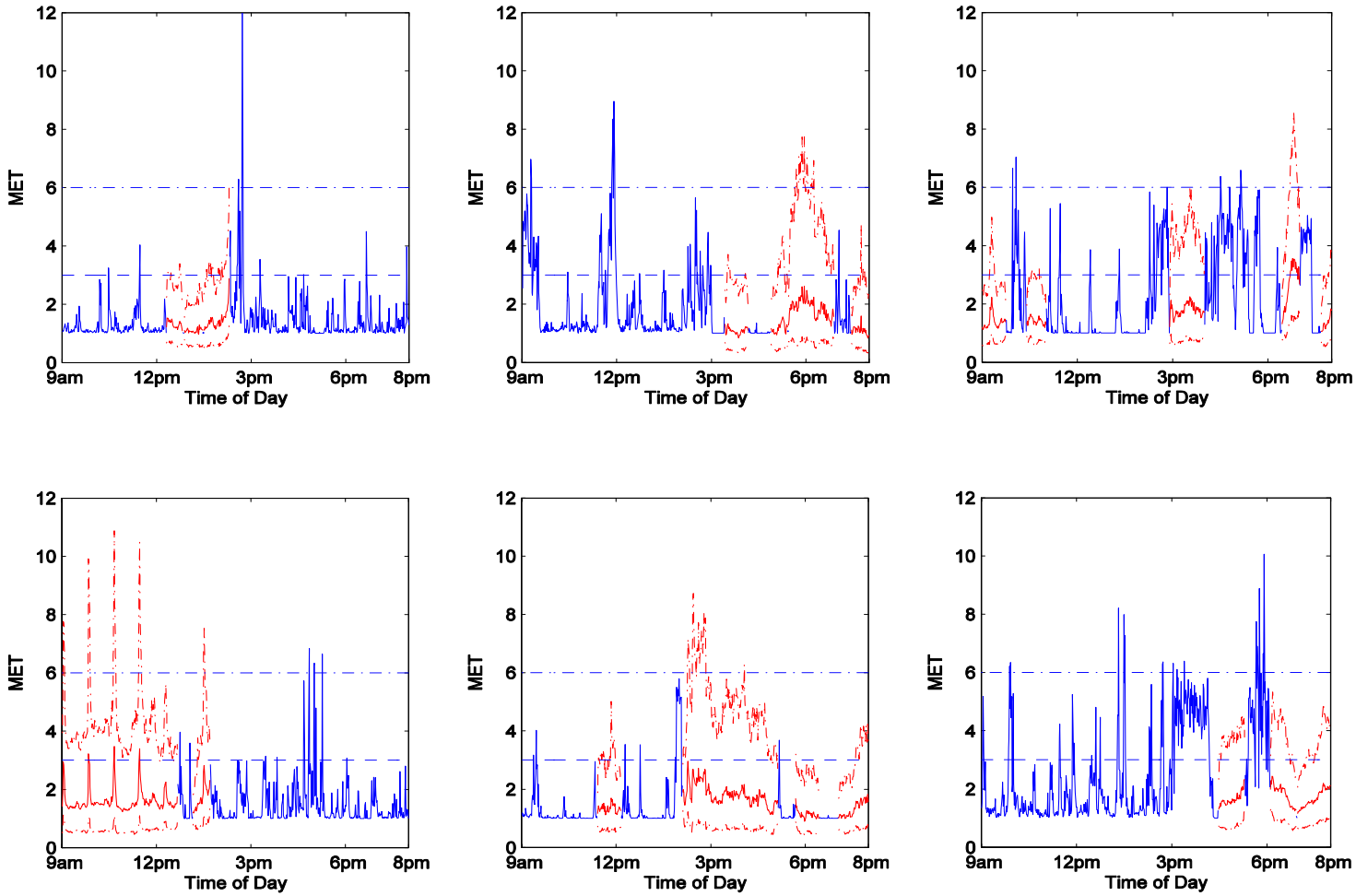
$$\mu_i^{M|O} = \mu_i^M + \Sigma_i^{M,O}(\Sigma_i^{O,O})^{-1}(Y_i^O - \mu_i^O)$$

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



Return

Incomplete Profiles



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

- **Introduced 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 generalizes to higher dimensional functions, e.g. image data, space/time (fixed domain) data**
- **The Gaussian assumptions can be robustified**

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*, to appear.
3. "*Bayesian Analysis of Mass Spectrometry Proteomics Data using Wavelet Based Functional Mixed Models*" (2006) Jeffrey S. Morris, Philip J. Brown, Richard Herrick, Keith A. Baggerly, and Kevin R. Coombes, submitted, available on MDACC's bepress site.

- **Computer code/papers on web at** `http://biostatistics.mdanderson.org/Morris/papers.html`

Projecting FMM to Wavelet Space

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

$$U_i \sim MVN(0, Q)$$

$$E_i \sim MVN(0, S)$$

Projecting FMM to Wavelet Space

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

$$U_i \sim MVN(0, Q)$$

$$E_i \sim MVN(0, S)$$

Projecting FMM to Wavelet Space

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

$$U_i \sim MVN(0, Q)$$

$$E_i \sim MVN(0, S)$$

Projecting FMM to Wavelet Space

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

$$U_i\mathbf{W}' \sim MVN(0, \mathbf{W}Q\mathbf{W}')$$

$$E_i\mathbf{W}' \sim MVN(0, \mathbf{W}S\mathbf{W}')$$

Projecting FMM to Wavelet Space

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

$$\mathbf{U}_i^{*} \sim MVN(0, \mathbf{Q}^{*})$$

$$\mathbf{E}_i^{*} \sim MVN(0, \mathbf{S}^{*})$$