Wavelet-based Functional Mixed Models

Jeffrey S. Morris

University of Texas,

M.D. Anderson Cancer Center

Houston, TX, USA

joint work with **Raymond J. Carroll**

Functional Data Analysis

- **Functional data:**
  - Ideal units of observation: curves
  - Observed data: curves sampled on fine grid
- **Increasingly encountered in scientific research**
- **FDA (Ramsay & Silverman, 1997)**

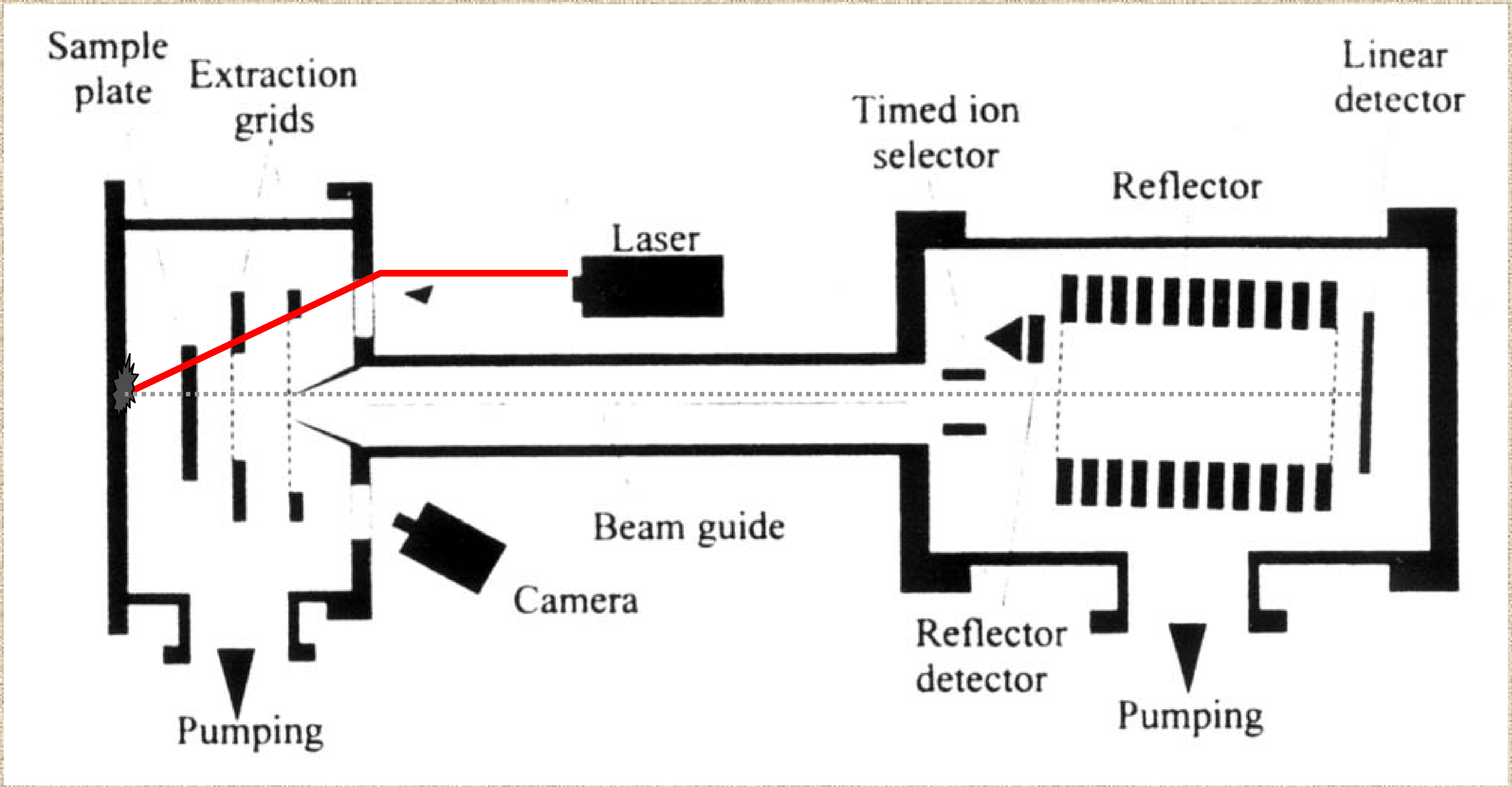
2 issues with functional data

1. **Regularization:** exploiting the assumed smoothness or regularity between measurements within a curve
2. **Replication:** combining information across N curves

Example: Mass Spectrometry Proteomics

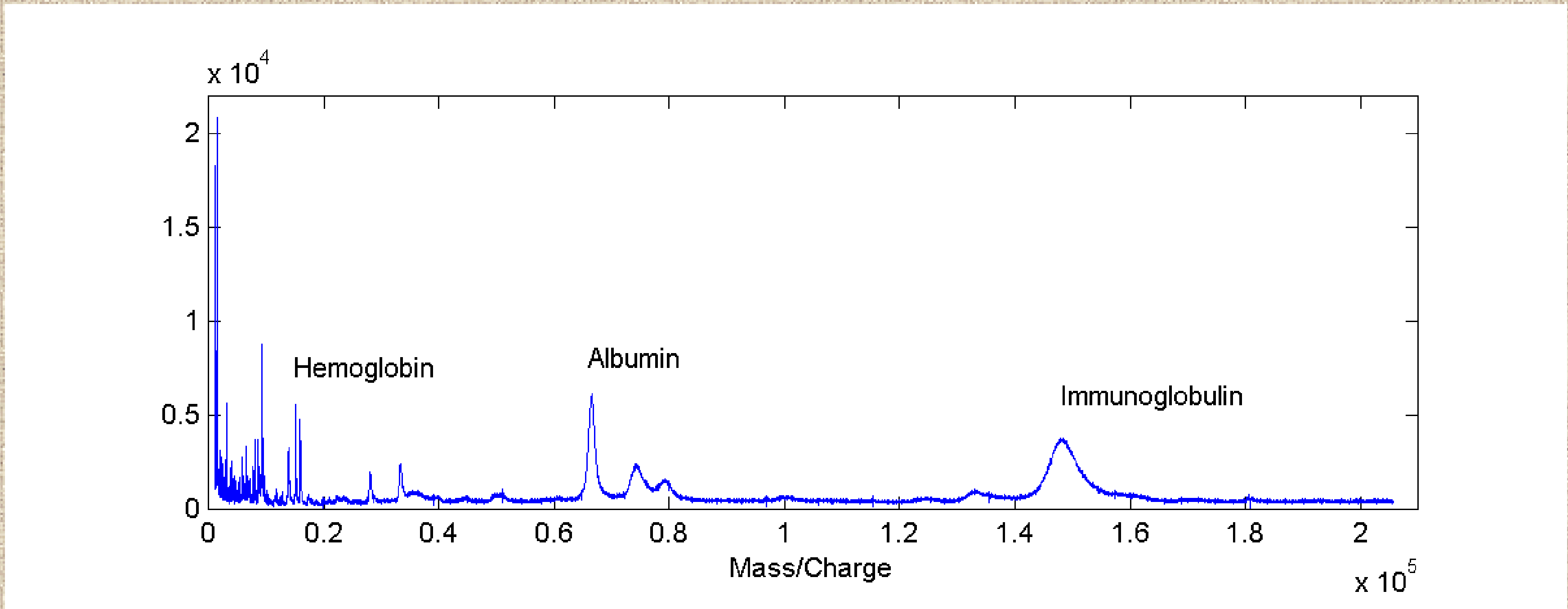
- **Central dogma**: DNA → mRNA → protein
- **Microarrays**: measure expression levels of 10,000s of genes in sample (amount of mRNA)
- **Proteomics:** look at proteins in sample.
  - Gaining increased attention in research
    - Proteins more biologically relevant than mRNA
    - Can use readily available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

MALDI-TOF Schematic



Vestal and Juhasz. *J. Am. Soc. Mass Spectrom*. **1998**, *9*, 892.

Sample MALDI-TOF Spectrum



- **MALDI-TOF: Spectrum is “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 nude 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 from each mouse – placed on 2 MALDI plates
- Samples run at 2 different laser intensities (low/ high)
- Total of 32 spectra (observed functions), 2 per mouse
- Sampled on equally-spaced grid of roughly 24,000
  - Downsampled to grid of size 2000

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

Linear Mixed Models

Linear Mixed Model (Laird and Ware, 1982):

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

$$u \sim N(0, \overbrace{P}^{m\times m})$$
$$e \sim N(0, \underbrace{R}_{N\times N})$$

- **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, …, *N*, on a compact set $\mathcal{T}$

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

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

- **DEFN**: $U(t)$~$MGP(P,Q)$ **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 $\mathcal{T}\times\mathcal{T}$**

Discrete Version of FMM

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

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

$$\begin{aligned} U &\sim MN(P,Q) \\ E &\sim MN(R,S) \end{aligned}$$

- ***U* and *E* follow the Matrix Normal distn.**
  - $U \sim MN(P,Q)$ 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
- Most existing literature for nonparametrically modeling functional data is based on kernels or splines (Guo 2002).
- Kernels/fixed-knot splines may not work well for spatially heterogeneous data
- Wavelet Regression: nonparametric regression technique that better preserves local features present in the curves.

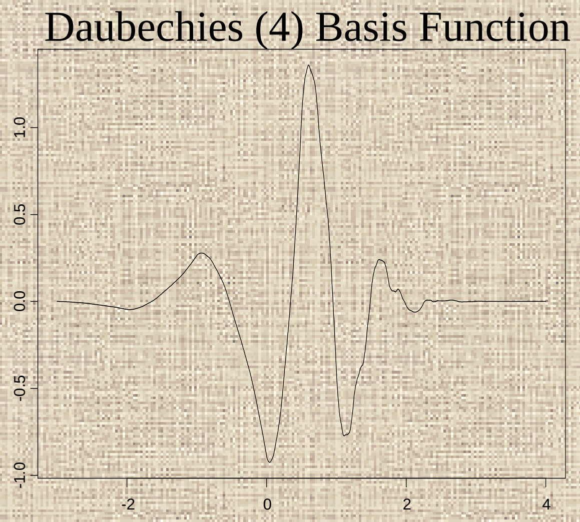
Introduction to Wavelets

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

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

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

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



- **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:** $\mathbf{d = y\, W'}$
  - ***W'* =*T*-by-*T* orthogonal projection matrix**
- **Inverse DWT (IDWT):** $\mathbf{y = d\, W}$

Wavelet Regression

- **Wavelet Regression:**
  - **Row vector *y*: response on equally-spaced grid *t* (length *T*)**

$$\boxed{\mathbf{y} = g(\mathbf{t}) + \varepsilon} \qquad \varepsilon \sim N(0, \sigma^2 I_T)$$

1. Project data into wavelet space using DWT.

d=y ***W'* where *W'* is the orthogonal DWT matrix**

Wavelet Regression

- **Wavelet Regression:**
  - **Row vector *y*: response on equally-spaced grid *t* (length *T*)**

$$\boxed{\mathbf{y}W' = g(\mathbf{t})W' + \varepsilon W'} \qquad \varepsilon \sim N(0, \sigma^2 I_T)$$

1. **Project data into wavelet space using DWT.**

   d=y ***W'* where *W'* is the orthogonal DWT matrix**

Wavelet Regression

- **Wavelet Regression:**
  - **Row vector *y*: response on equally-spaced grid *t* (length *T*)**

$$\boxed{\mathbf{d} = g(\mathbf{t})W' + \varepsilon W'} \qquad \varepsilon \sim N(0, \sigma^2 I_T)$$

1. **Project data into wavelet space using DWT.**

   d=y ***W'* where *W'* is the orthogonal DWT matrix**

Wavelet Regression

- **Wavelet Regression:**
  - **Row vector *y*: response on equally-spaced grid *t* (length *T*)**

$$\boxed{\mathbf{d} = \boldsymbol{\theta} + \varepsilon W'} \qquad \varepsilon \sim N(0, \sigma^2 I_T)$$

1. **Project data into wavelet space using DWT.**

   d=y ***W'* where *W'* is the orthogonal DWT matrix**

Wavelet Regression

- **Wavelet Regression:**
  - **Row vector *y*: response on equally-spaced grid *t* (length *T*)**

$$\boxed{\mathbf{d} = \boldsymbol{\theta} + \varepsilon^*} \qquad \varepsilon^* \sim N(0, \sigma^2 I_T)$$

1. **Project data into wavelet space using DWT.**

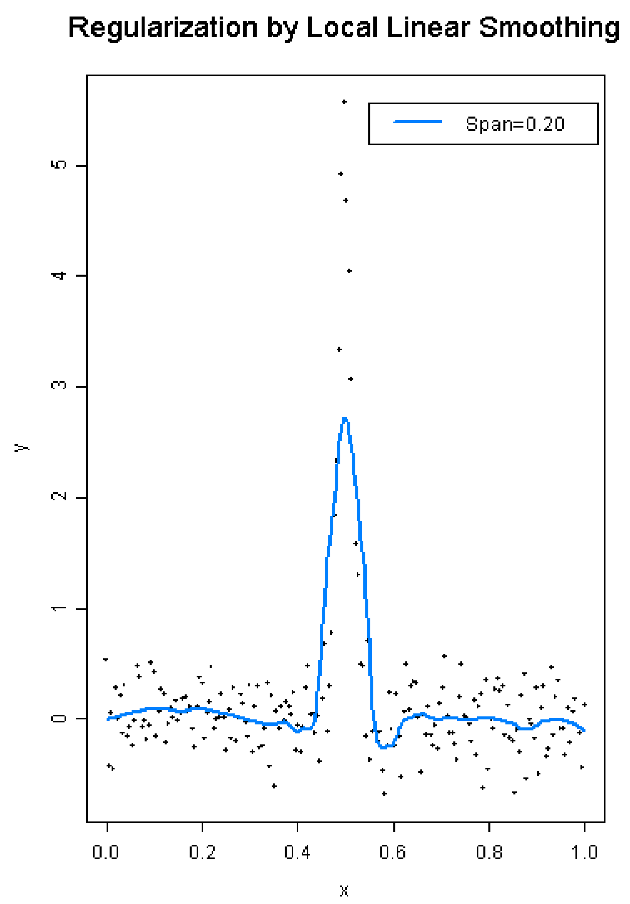
   d=y ***W'*** **where *W'* is the orthogonal DWT matrix**

2. **Estimate θ by thresholding (or placing mixture prior on θ)**
3. **Project back to data space using IDWT**

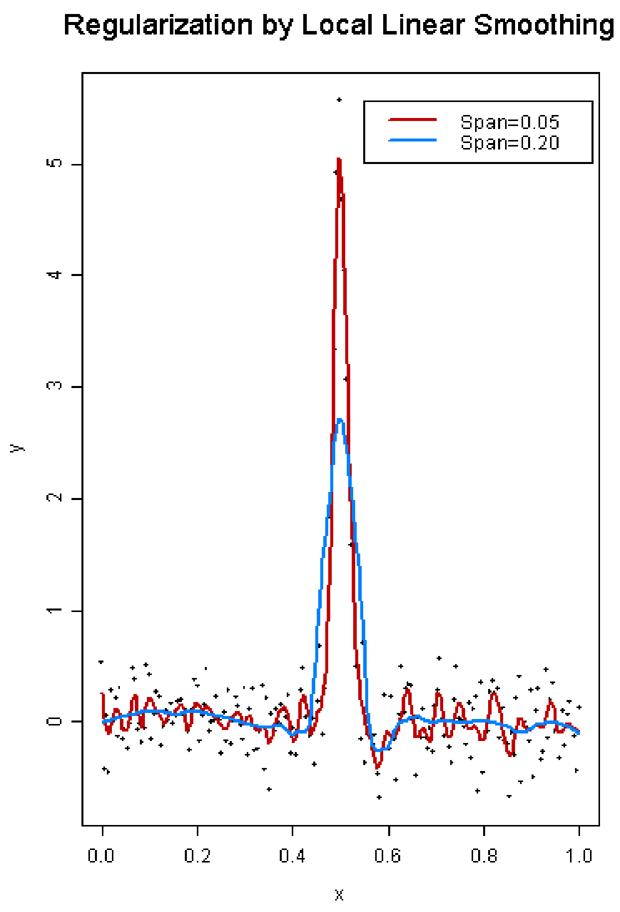
$$\boxed{\hat{g}(t) = \hat{\theta} W}$$

- **Yields *adaptive regularized* nonparametric estimate of *g*(t).**

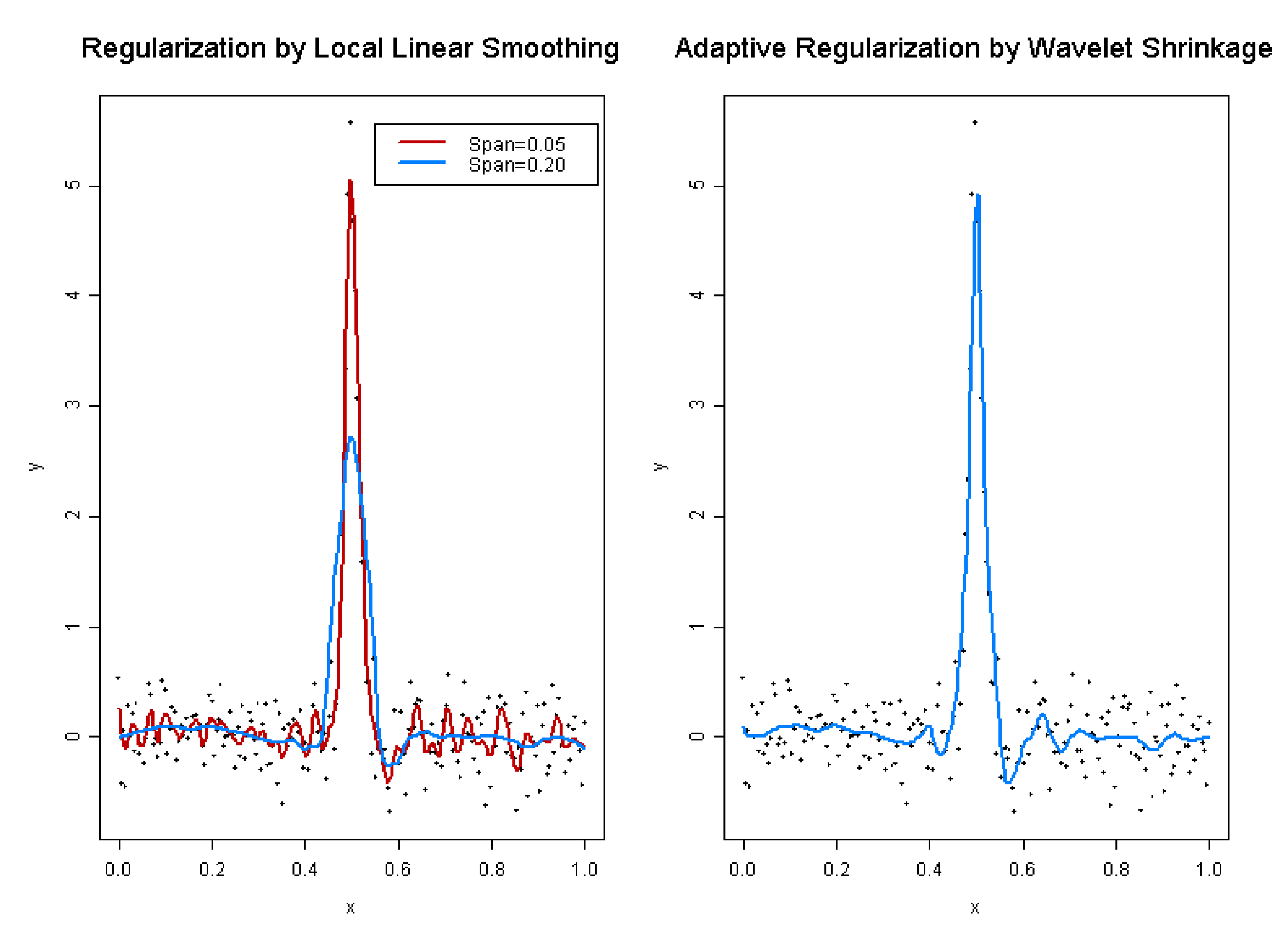
Adaptive Regularization



Adaptive Regularization



Adaptive Regularization



Wavelet-based Hierarchical Functional Models

- Most existing wavelet regression methods for single function case
- Morris, Vannucci, Brown, and Carroll (2003)
  - Bayesian wavelet-based method for estimating mean function for functional data from nested design.
  - Extended wavelet regression to hierarchical functional context.
- Goal: Develop Bayesian wavelet-based methodology for functional mixed model setting.

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**

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

- **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**.

Projecting FMM to Wavelet Space

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

$$\boxed{\begin{aligned} U &\sim MN(P,Q) \\ E &\sim MN(R,S) \end{aligned}}$$

Projecting FMM to Wavelet Space

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

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

Projecting FMM to Wavelet Space

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

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

Projecting FMM to Wavelet Space

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

$$U\mathbf{W'} \sim MN(P, \mathbf{W}Q\mathbf{W'})$$
$$E\mathbf{W'} \sim MN(R, \mathbf{W}S\mathbf{W'})$$

Projecting FMM to Wavelet Space

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

$$\mathbf{U}^* \sim MN(P, \mathbf{Q}^*)$$

$$\mathbf{E}^* \sim MN(R, \mathbf{S}^*)$$

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,**
- ***E*=EW'*** **is the matrix of residuals**
- ***Q*=WQW'*** **and** ***S*=WSW'*** **model the covariance structure between wavelet coefficients for function.**

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.**
  - **Assume wavelet coefficients within given random effect function or residual error process are independent**
  - **Heuristically justified by whitening property of DWT**
  - **Common assumption in wavelet regression**
  - **Is parsimonious in wavelet space (*T* parameters), yet leads to flexible class of covariance structures in data space**

Wavelet Space Model

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

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

Model Each Column Separately

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

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

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

Single Wavelet Coefficient Model

- **Independence assumption allows us to fit wavelet-space model one column at a time.**
  - **i.e., we have a series of T (scalar) mixed models, with the only shared parameters being the between-curve covariance parameters in P and R.**
- **In principal, we could fit this model using standard mixed models software.**
- **However, fitting this model without additional mechanism for regularization would result in rough, noisy estimates of the p fixed effects functions $B_i(t)$**

Prior Assumptions

Mixture prior on β_{ijk}^*:

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

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

- **Nonlinearly shrinks β_{ijk}^* towards 0, leading to adaptively regularized estimates of $\beta_i(t)$.**
- **τ_{ij} & π_{ij} are regularization parameters**
  - **Can be estimated from the data using empirical Bayes**
  - **Extend Clyde&George (1999) to functional mixed model**

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. for P, Q^*, R, 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 $B_i(t)$ for $i=1,\ldots, p$, on grid t.**
  - **B=B*W**
- **Posterior samples of $U(t)$, P, Q, R, and S are also available, if desired**
- **Can be used for Bayesian inference/prediction**

  Posterior probabilities, Bayes Factors, posterior predictive probabilities easy to compute.

Example: Model

Let $Y(t)$ be the N=32 MALDI spectra, preprocessed.

$$\boxed{Y(t) = XB(t) + ZU(t) + E(t)}$$

- $X=\{1, X_{organ}, X_{cell\text{-}line}, X_{int}, X_{laser}\}$ **, where**

  X_{organ}**=1 for lung, -1 brain.** $X_{cell\text{-}line}$**=1 for A375P, -1 for PC3MM2**

  $X_{int} = X_{organ} * X_{cell\text{-}line}$ X_{laser}**=1 for low laser intensity, -1 high.**

- $B(t)=\{B_0(t), B_1(t), B_2(t), B_3(t), B_4(t)\}$**, where**

  $B_0(t)$ = 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**

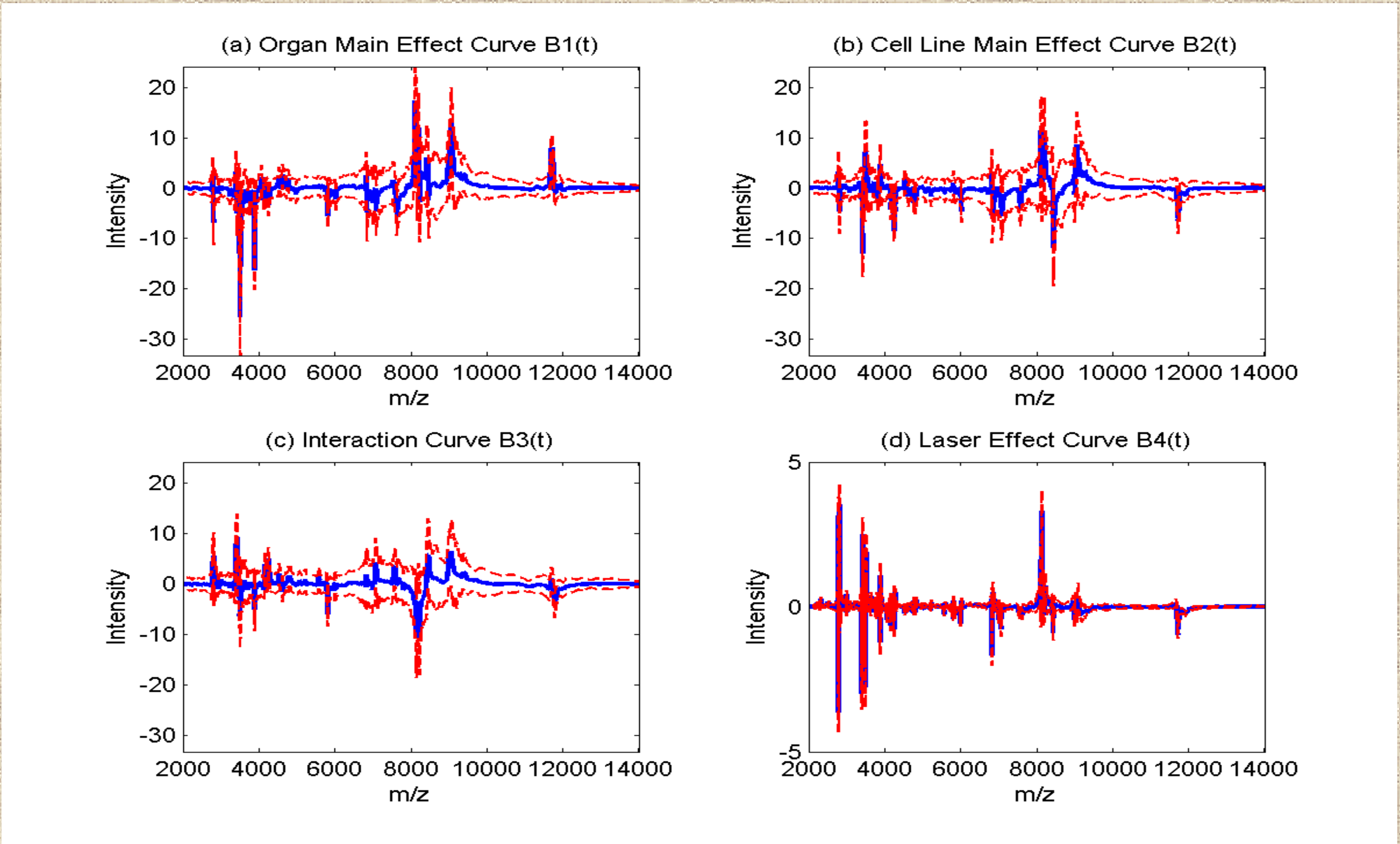
  $B_4(t)$ = laser intensity effect **function**

- $Z=1_2 \otimes I_{16}$**, and** $U(t)=\{U_1(t), \ldots, U_{16}(t)\}$ **are** mouse random effect **functions.**

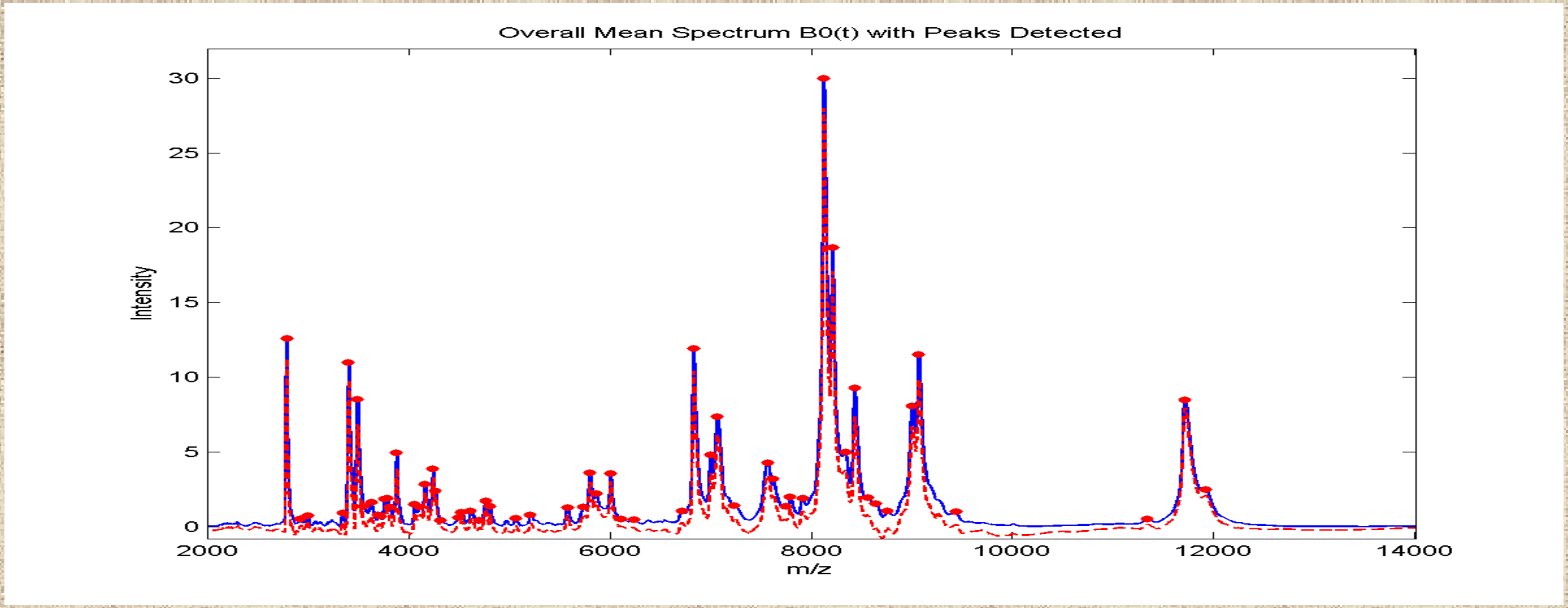
Example: Model Fitting

- **Daubechies 8 wavelet basis, *J*=11 levels**
- **Empirical Bayes procedure used to estimate regularization parameters π_{ij} and τ_{ij} from data.**
- **Burn-in 1000; 20,000 MCMC samples; thin=10**
- **Took 7hr 53min on Win2000 P-IV 2.8GHz 2GB RAM**
  - **That is Matlab code; C++ code takes ~2 hours.**
- **Trace plots indicated good convergence properties**
- **Metropolis Hastings acceptance probabilities good:**
  - **Range of (0.04, 0.53)**
  - **(10^{th},50^{th},90^{th}) percentiles of (0.20, 0.29, 0.50)**

Example: Results



Example: Peak Detection



- We first did "**peak detection**"
  - **Local maxima** in posterior mean (denoised) estimate of $B_0(t)$ with High posterior probability of nonzero mean; $\Pr\{B_0(t)>0|Y\}\geq 0.95$
- Using this criterion, we found 58 peaks
- We restrict inference to values of t at peaks

Example: Flagged Peaks

Detecting 'significant' peaks: (assoc. w/ organ,cl,int)

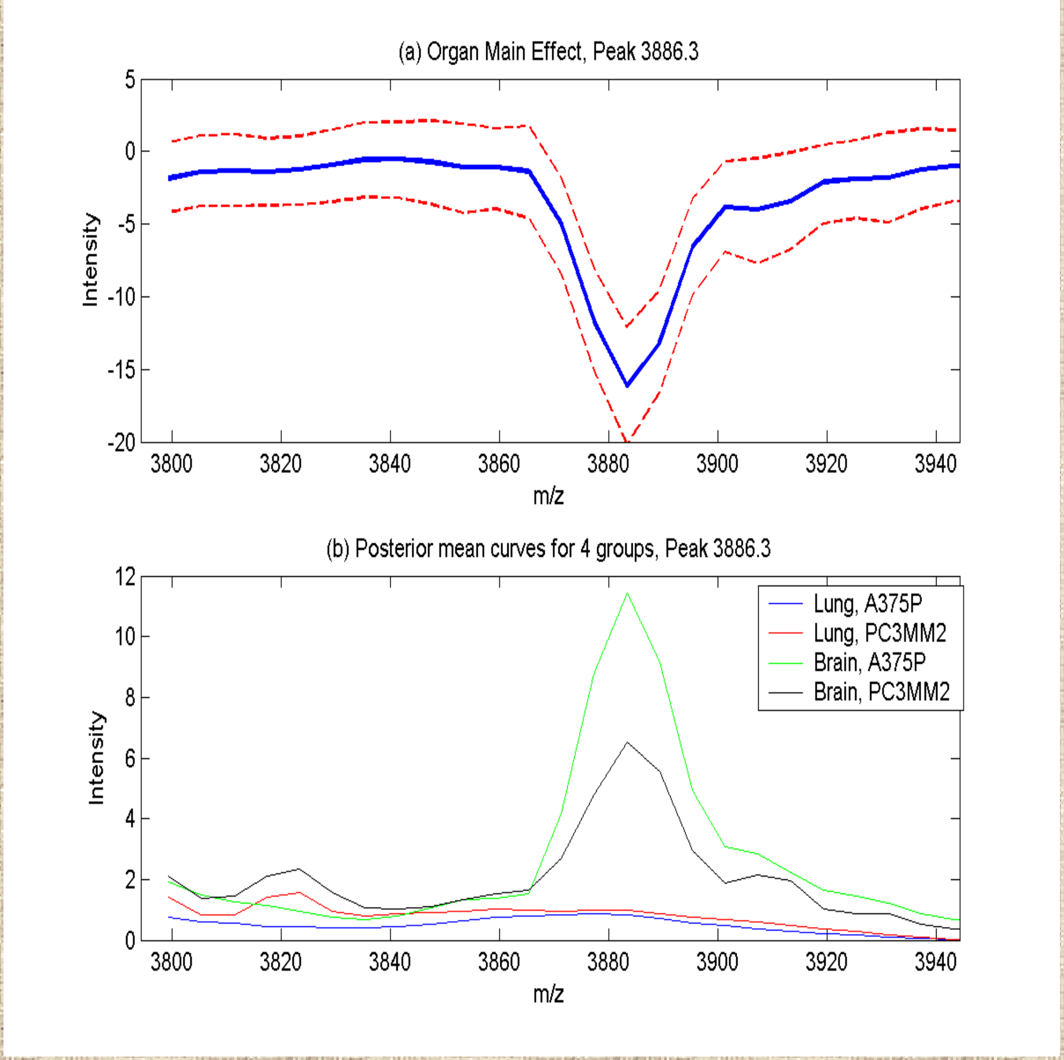
For each t at a peak, compute $p_i(t)=\min[\Pr\{B_i(t)>0\},\Pr\{B_i(t)<0\}]$

Flag any peak for which $p_i(t)$ is very small.

Using this criterion, we flagged 9 peaks as interesting

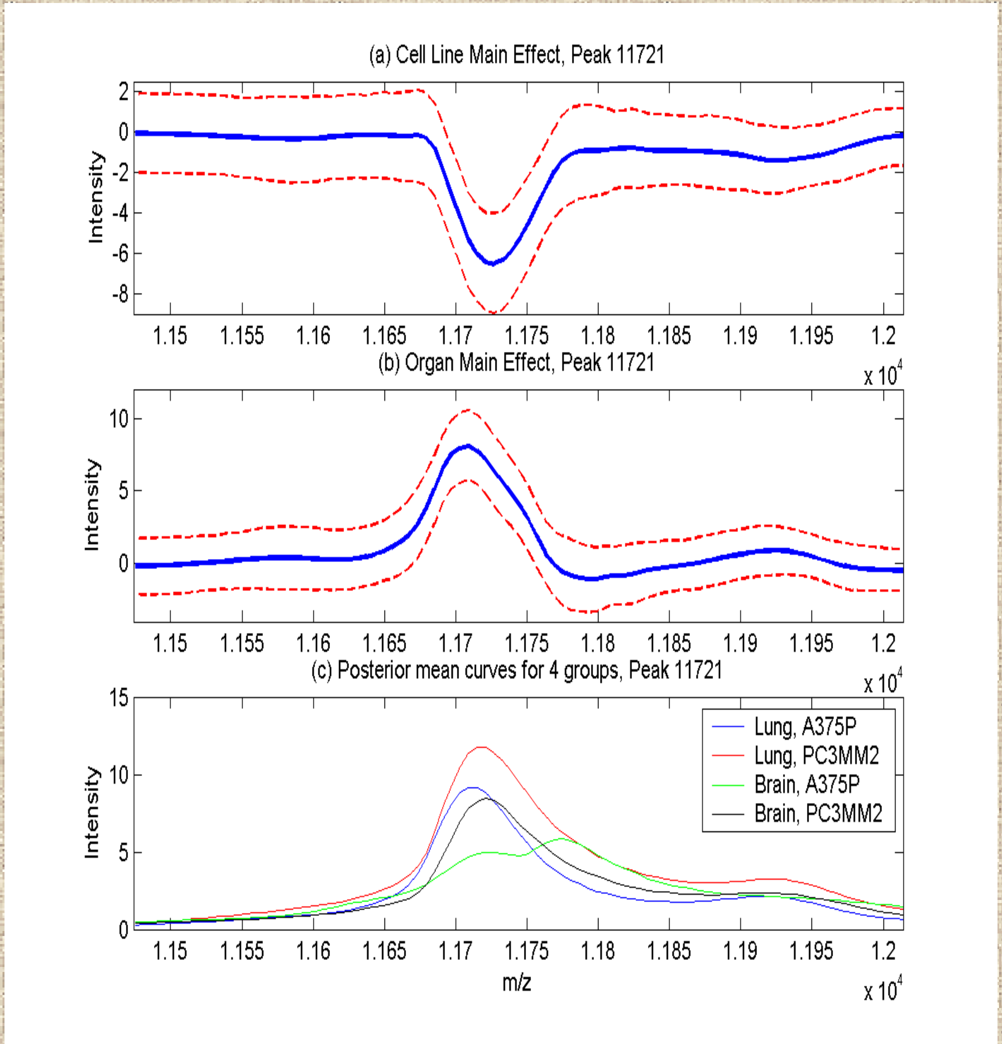
m/z	Effect	*p*	Comment
3412.6	int.	<0.0005	PC3MM2>A375P for brain-injected only
3496.6	organ	<0.0005	Only expressed in brain-injected mice
3886.3	organ	<0.0005	Only expressed in brain-injected mice
4168.2	int.	0.0005	PC3MM2>A375P in brain-injected only
4252.1	int.	<0.0005	PC3MM2>A375P in brain-injected only
4270.1	cell line	<0.0005	PC3MM2>A375P
5805.3	int.	<0.0005	brain>lung only for mice given A375P cell-line
6015.2	cell line	<0.0005	PC3MM2>A375P
11721	cell line	<0.0005	PC3MM2>A375P
11721	organ	<0.0005	lung>brain

Example: Results



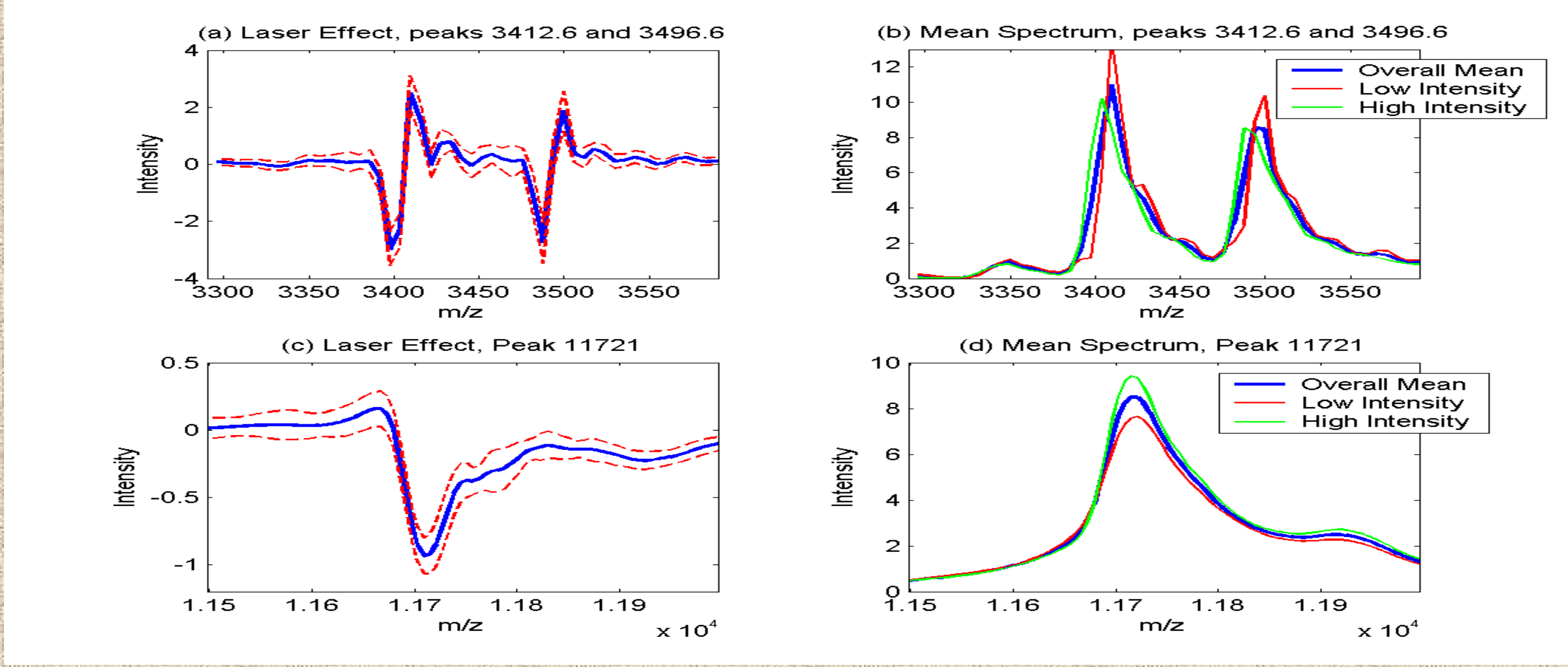
- **Specific to brain-injected mice**
- **May be CGRP-II (3882.34 Dal), peptide in mouse proteome that dilates blood vessels in brain**
- **Host response to tumor implanted in brain?**

Example: Results



- **Higher in mice injected with metastatic (PC3-MM2) cell line**
- **May be MTS1 (11721.43 Dalt), metastatic cell protein in mouse proteome.**
- **Also higher in lung-injected mice than brain-injected mice**

Example: Results



- **Laser intensity effect adjusts for:**
  - **Shifts in intensities**
  - **Offsets in m/z scale**
- **Important proof of principle that "linear" functional term can be used to adjust for functional effects of nuisance factors**

Discussion

- **Introduced unified modeling approach for FDA**
  - **Applied here to MALDI-TOF, but method is general.**
- **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 group membership 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

- **Since functions adaptively regularized using wavelet shrinkage, the method is appropriate for spatially heterogeneous functional data.**
- **Approach is Bayesian. The only informative priors to elicit are regularization parameters, which can be estimated from data using empirical Bayes.**
- **Method complex, but code straightforward to implement.**
- **Method generalizes to higher dimensional functions, e.g. image data, space/time (fixed domain) data.**
- **We used wavelet bases, but approach can be generalized to other orthogonal basis functions.**
- **Major challenges in developing unified statistical modeling approach for replicated functional data, but worth the effort.**

Acknowledgements

- **Statistical Collaborators on work related to WFMM: Raymond J. Carroll, Louise Ryan, Phil Brown, Marina Vannucci, Kevin Coombes, Keith Baggerly, Naisyin Wang, Brent Coull, and Cassandra Arroyo.**
- **Other Scientific Collaborators: Steve Gortmaker, Joanne Lupton, Josh Fidler, Lihong Wang, Victor Prieto, Stan Hamilton, Ryuji Kobayashi, David Hawke, John Kooman, Nancy Turner, and Meeyoung Hong.**
- **Thanks to Dick Herrick for assistance in optimizing the code for the method, and for converting the Matlab code to C++.**