Wavelet-Based Functional Mixed Models

Jeffrey S. Morris

Department of Biostatistics

The University of Texas

MD Anderson Cancer Center

Rice University

2/27/2008

Linear Mixed Models:
General Linear Model

• **Suppose we have *N* data points y_i, *i=1, …, N* that we would like to explain by a set of *p* explanatory variables x_{ij}; *i=1, …, N; j=1, …, p*. The x_{ij} may be continuous variables, or discrete (dummy) variables. A General Linear Model is given by:**

$$y_i = \sum_{j=1}^{p} x_{ij} b_j + e_i$$

• **In many applications, we assume the errors are independent and identically distributed, so $e_i \sim N(0, \sigma_e^2)$**

Linear Mixed Models:
General Linear Model

- **This form is very general, with different specifications of x_{ij} corresponding to various classical statistical methods**
  - **Continuous: Linear regression**
  - **Dummy variables (2 groups): t-test**
  - **Dummy variables (>2 groups): ANOVA**
  - **Combination of continuous and dummy variables: Analysis of Covariance**

- **Note: b_j are partial effects; i.e. describe the effect of predictor x_{ij} after adjusting for all other effects in the model**

Linear Mixed Models:
General Linear Model

- **This model can be written in matrix form:**

$$y = Xb + e$$

$$e \sim N(0, \sigma_e^2 I_N)$$

Where y is an $N\times1$ response vector, X is an $N\times p$ design matrix, and b is a $p\times1$ vector of fixed effects

- **Typical goal: estimate and perform inference on fixed effect vector b**

- **The form of the design matrix, X, indicates the type of modeling being done**

Linear Mixed Models: Examples

- **Simple linear regression**

To perform a simple linear regression of *Y* on *x*, specify:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

- b_0=**Intercept**
- b_1=**slope**

- **Polynomial Regression:**

To perform a polynomial regression of Y on x, specify:

$$X = \begin{bmatrix} 1 & (x_1 - \bar{x}) & (x_1 - \bar{x})^2 & \cdots & (x_1 - \bar{x})^p \\ 1 & (x_2 - \bar{x}) & (x_2 - \bar{x})^2 & \cdots & (x_2 - \bar{x})^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_N - \bar{x}) & (x_N - \bar{x})^2 & \cdots & (x_N - \bar{x})^p \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

- b_0=**Intercept,** b_j=**polynomial level j coefficient**
- **Note: Centering x reduces collinearity**

Linear Mixed Models: Examples

- **Multiple linear regression**

To perform multiple regression of y on $\{x_1, x_2, \ldots, x_p\}$:

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & x_{2N} & \cdots & x_{pN} \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

- b_0 =Intercept

- $b_1, b_2, \ldots, b_p$ =partial slopes for factors $1, \ldots, p$

Linear Mixed Models: Examples

- **T-tests:** Define dummy variable (categorical factor)

- $x=1$ if group 0, -1 if group 1

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

- b_0=Intercept (overall mean of y)

- b_1=Main effect for group (=0 if no group effect)

  Note: F-test for H_0:b_1=0 is equivalent to t^2

- What if there are >2 groups we want to compare? 1-way ANOVA

Linear Mixed Models: Examples

- **ANOVA (Cell Means model): (g groups)**

- **Let $x_{ij}=1$ if subject i in group j, and <u>no intercept</u>**

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1g} \\ x_{21} & x_{22} & \cdots & x_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Ng} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_g \end{bmatrix}$$

- **b_j: overall mean for group j**

  $b_j=b_{j'}$ implies groups j and j' share common mean

- ***Contrasts* can be used to set up comparisons of interest**

Linear Mixed Models: Examples

- A *p×1* vector *c={c$_1$, c$_2$,…,c$_g$}* specifies a ***contrast***

$$L = \sum_{j=1}^{g} c_j b_j, \qquad \text{if } \sum_{j=1}^{g} c_j = 0$$

- **Specifies a linear combination of fixed effects that indicate a certain comparison of interest.**

- **E.g. Can be used with cell means model to test difference between groups *j* and *j'*: $c_j=1$, $c_{j'}=-1$**

- **Given set of *l* contrasts of interest, can construct *l×g* *contrast matrix C***

  Main Effect(ANOVA):

$$C = \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}$$

Linear Mixed Models: Examples

- **Analysis of Covariance (ANOCOVA)**

- **Cell Means model: (g groups, p covariates)**

- Let $x_{ij}=1$ if subject i in group j, $j=1, \ldots, g$
  Let $x_{ij}=$ covariate $(j-g+1)$ for subject i, $j=g+1, \ldots g+p$

$$
X = \begin{bmatrix}
x_{11} & \cdots & x_{1g} & x_{1(g+1)} & \cdots & x_{1(g+p)} \\
x_{21} & \cdots & x_{2g} & x_{2(g+1)} & \cdots & x_{2(g+p)} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
x_{N1} & \cdots & x_{Ng} & x_{N(g+1)} & \cdots & x_{N(g+p)}
\end{bmatrix}, \quad
b = \begin{bmatrix}
b_1 \\
\vdots \\
b_g \\
b_{g+1} \\
\vdots \\
b_{g+p}
\end{bmatrix}
$$

- b_j, $j=1, \ldots, g$: separate intercepts by group j

- b_j, $j=g+1, \ldots, g+p$: partial slopes for p covariates

Linear Mixed Models: Examples

- **What if we have multiple categorical factors we would like to simultaneously consider?**

- **2-way ANOVA: Includes main effects for factors 1 and 2, plus their 2-way interaction**

- **Interaction: The effect of factor 1 is different for different levels of factor 2 (and vice versa)**

- **Again, different parameterizations can be used**
  - **Overparameterized model: (messy!)**
  - **Main effect/interaction model**
  - **Cell means model, with contrasts**

- **Can also interact continuous factors with categorical factors e.g. growth curves with intercepts and slopes varying by group**

Linear Mixed Models: Examples

- **Cell Means model: (g=g_1×g_2 groups)**
- Let $x_{ij}=1$ if subject i in group j, and no intercept

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1g} \\ x_{21} & x_{22} & \cdots & x_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Ng} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_g \end{bmatrix}$$

- b_j: overall mean for group j

  $b_j=b_{j'}$ implies groups j and j' share common mean
- *Contrasts* can be used to set up comparisons of interest, e.g. main effect and interactions

Linear Mixed Models: Examples

Suppose factor 1 has 2 levels, and factor 2 has 3 levels, with $b=[b_{11}, b_{12}, b_{13}, b_{21}, b_{22}, b_{23}]'$

$$C_1 = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix}$$

$$C_{12} = \begin{bmatrix} 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix}$$

- *$L_1=C_1b$:* **Main effect for factor 1**
- *$L_2=C_2b$:* **Main effect for factor 1**
- *$L_{12}=C_{12}b$:* **2-way interaction effect**

Linear Mixed Models: Examples

- **General Linear Model:**

$$y = Xb + e$$

$$e \sim N(0, \sigma_e^2 I_N)$$

- **What if errors are not iid?**
  - **Subsampling** **(multiple randomly sampled observations per subject)**
  - **Longitudinal Data** **(multiple observations per subject over time, so observations are ordered)**

- **Solutions:**
  - **Assume parametric error structure on** $e \sim N(0,R)$
  - **Include** *random effect* **terms in the model.**

Linear Mixed Models: Examples

- **Simple Random Effects model:** Suppose we observe r repeated measurements of response y for each of n subjects (total of $N=rn$ obs.)

- Let $y=[y_{11}, y_{12}, \ldots, y_{1r}, y_{21}, \ldots, y_{2r}, \ldots, y_{nr}]$

- Let X be a design matrix for the fixed effects b

- Define an $N \times n$ random effects design matrix Z to have elements $z_{ij}=1$ if obs i from subject j, 0 otherwise.

- A random effects model can be written as:

$$y = Xb + Zu + e$$

$$u \sim N(0, \sigma_u^2 I_n)$$

$$e \sim N(0, \sigma_e^2 I_r)$$

Linear Mixed Models: Examples

$$y = Xb + Zu + e$$

$$u \sim N(0, \sigma_u^2 I_n)$$

$$e \sim N(0, \sigma_e^2 I_r)$$

- u_k = **random effect for subject** k**:** $E(y|$**subject** $k)=Xb+u_k$

- σ_u^2**: describes subject-to-subject variability**

- σ_e^2**: describes variability between observations within subject**

  - **If we integrate the random effects** u **out of the model, thus marginalizing over** u,

$$y = Xb + e^* \qquad e^* \sim N(0, \Sigma)$$

Where $\Sigma = \sigma_u^2 ZZ' + \sigma_e^2 I_N$ **has block** *compound symmetry* **structure**

Block Compound Symmetry

$$\Sigma = \begin{bmatrix} \begin{matrix} \sigma_{e_*}^2 & \rho\sigma_{e_*}^2 & \rho\sigma_{e_*}^2 \\ \rho\sigma_{e_*}^2 & \sigma_{e_*}^2 & \rho\sigma_{e_*}^2 \\ \rho\sigma_{e_*}^2 & \rho\sigma_{e_*}^2 & \sigma_{e_*}^2 \end{matrix} & 0 & \cdots & 0 \\ 0 & \begin{matrix} \sigma_{e_*}^2 & \rho\sigma_{e_*}^2 & \rho\sigma_{e_*}^2 \\ \rho\sigma_{e_*}^2 & \sigma_{e_*}^2 & \rho\sigma_{e_*}^2 \\ \rho\sigma_{e_*}^2 & \rho\sigma_{e_*}^2 & \sigma_{e_*}^2 \end{matrix} \vdots & & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \begin{matrix} \sigma_{e_*}^2 & \rho\sigma_{e_*}^2 & \rho\sigma_{e_*}^2 \\ \rho\sigma_{e_*}^2 & \sigma_{e_*}^2 & \rho\sigma_{e_*}^2 \\ \rho\sigma_{e_*}^2 & \rho\sigma_{e_*}^2 & \sigma_{e_*}^2 \end{matrix} \end{bmatrix} \left. \vphantom{\begin{matrix} 1 \\ 1 \end{matrix}} \right\} n \text{ blocks}$$

$r \times r$ blocks

- ***n* diagonal blocks of size *r×r*, each having compound symmetry**
- **Note: $\sigma_{e*}^2 = \sigma_e^2 + \sigma_u^2$ is the *total variability* and**

  $\rho = \sigma_u^2 / \sigma_e^2 + \sigma_u^2$ **is the *intraclass correlation***

Linear Mixed Models: Examples
General Linear Mixed Model

$$\underset{N\times 1}{\underbrace{y}} = \underset{N\times p}{\underbrace{X}}\,\overset{p\times 1}{\overbrace{b}} + \underset{N\times m}{\underbrace{Z}}\,\overset{m\times 1}{\overbrace{u}} + \underset{N\times 1}{\underbrace{e}}$$

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

- **Var(y)=Σ=ZPZ'+R**

$$\hat{b} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y = \text{generalized least squares estimator}$$

$$\hat{u} = PZ'\Sigma^{-1}(y - X\hat{b}) = \text{best linear unbiased predictor (BLUP)}$$

- **Most common applications use independent random effects and residual errors, $P=\sigma_u^2 I_m$ and $R=\sigma_e^2 I_N$**
- **Other possible covariance structures can also be used for P and R, as needed**

Linear Mixed Models: Examples
Fitting Linear Mixed Models

- **Standard software packages** are available for fitting linear mixed models
  - **PROC MIXED** (SAS)
  - **lme/nlme** (S+/R)
- **Based on EM or Newton-Raphson algorithms**
  - MLE for fixed effect functions (GLS)
  - REML for covariance parameters
- **Straightforward to perform frequentist inference on fixed effect functions, or linear combinations (contrasts) – F tests**
- **Trickier to test whether VC=0 but recent work in this area**
- **Can also be fit using Bayesian hierarchical model**
  - If standard improper priors (uninformative) are specified for fixed effects b and covariance parameters, it can be shown that the posterior mode Bayesian estimators are equivalent to MLE frequentist estimators, and the confidence/credible bands are similar

Functional Data Analysis

- **Functional Data Analysis (Ramsay&Silverman 1997)**
  - Methods for data where ideal units of observation are curves
  - Sometimes distinguish between methods for sparsely sampled functional data (few obs/curve, longitudinal) and functions sampled on fine grid (many obs/curve, functional)
  - Our main concern here is on functional data on fine grid

- **Challenge of Functional Data Analysis: 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

Functional Mixed Model (FMM)

- **Idea:** Relate *functional response* to set of scalar predictors through *functional coefficients*, while adjusting for possible *correlation between functions* induced by design.

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

$$U_k(t) \sim GP(0, Q)$$
$$E_i(t) \sim GP(0, S)$$

$$\underbrace{Y_i(t)}_{\substack{\text{response} \\ \text{functions}}} = \sum_{j=1}^{p} X_{ij} \underbrace{B_j(t)}_{\substack{\text{fixed effect} \\ \text{function}}} + \sum_{k=1}^{m} Z_{ik} \underbrace{U_k(t)}_{\substack{\text{random effect} \\ \text{functions}}} + \underbrace{E_i(t)}_{\substack{\text{residual error} \\ \text{functions}}}$$

- $B_j(t)$ summarizes partial effect of X_j on $Y(t)$

- $Q(t_1, t_2)$ and $S(t_1, t_2)$ are covariance surfaces on $\mathcal{T} \times \mathcal{T}$ describing the form of the function-function deviations

Discrete Version of FMM

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

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

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

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

- **Rows of B contain fixed effect functions on grid**

- **Q and S are within-curve covariance matrices $(T \times T)$ approximating surfaces on the grid**

  - **For irregular functional data, Q and S typically contain many nonstationarities, yet their dimension is too high to leave unstructured**

FDA: Functional Mixed Models

Growth curve example:

- **Goal:** Assess effect of gene A on tumor growth
- *N=20* mice, ½ from knockout and ½ from wild type groups
- **Xenograft model:** microscopic tumor implanted
- Tumor measurements at *T=6* time points *t={0,3,6,9,12,15}*
- $Y_i(t)$=log(tumor size) profile for animal *i*

$$Y_i(t) = \sum_{j=1}^{2} X_{ij} B_j(t) + E_i(t) \qquad E_i(t) \sim N\{0, S(t_1, t_2)\}$$

- **Cell Means** Model Used
  - $X_{i1}=1$ if animal *i* from knockout group, *0* otherwise
  - $X_{i2}=1$ if animal *i* from wild type group, *0* otherwise
- **Test:** $H_0: B_1(t) = B_2(t)$

FDA: Functional Mixed Models

- **Growth curves are (log) linear, so we can represent functions with parametric form (lines)**
  - **No intercepts (tumor size=0 at time=0 by design)**
  - $B_j(t) = b_j t;$ $\qquad\qquad E_i(t) = u_i t + e_i;$
  - $u_i \sim N(0, \sigma_u^2);$ $\qquad\qquad e_i \sim N(0, \sigma_e^2 I_6)$
  - $S = t't\, \sigma_u^2 + \sigma_e^2 I_6,$ $\qquad$ **e.g. if $\sigma_u^2 = 0.1$, $\sigma_e^2 = 1$:**

S

	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1.9	1.8	2.7	3.6	4.5
3	0	1.8	4.6	5.4	7.2	9
4	0	2.7	5.4	9.1	10.8	13.5
5	0	3.6	7.2	10.8	15.4	18
6	0	4.5	9	13.5	18	23.5

ρ_S

	1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1	0.61	0.65	0.67	0.67
3	0	0.61	1	0.83	0.86	0.87
4	0	0.65	0.83	1	0.91	0.92
5	0	0.67	0.86	0.91	1	0.95
6	0	0.67	0.87	0.92	0.95	1

Mixed model in matrix form:

Let $y=\{y_k, k=1, \ldots, NT\}$, $X_{k1}=1$ if obs k from knockout animal, $X_{k2}=1$ if obs k from wild type and $Z_{kl}=1$ if obs k from animal l

$$y = Xb + Zu + e \quad u \sim N\left(0, \sigma_u^2 I_N\right) \quad e \sim N\left(0, \sigma_e^2 I_{NT}\right)$$

$$X = \begin{bmatrix} X_{11}t_1 & X_{12}t_1 \\ X_{21}t_2 & X_{22}t_2 \\ \vdots & \vdots \\ X_{NT1}t_{NT} & X_{NT2}t_{NT} \end{bmatrix}$$

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

FDA: Functional Mixed Models

- **What if there is no acceptable parametric representation for the observed functions?**
- **Nonparametric representations** for functions: generalizations of nonparametric regression to multiple function case. Much recent work in this area.
- **Limitations of previous work:**
  - Most focus on smoothing methods like kernels and splines
  - Much intended for functional data on sparse, not fine, grids
  - Most models not fully general FMM (arbitrary X matrix, nested random effects, general representations for Q or S)
  - Many focus on estimation, not inference
- **Most of this work works with models that are special cases of the functional mixed model.**

Applications: SELDI Organ-by-Cell-Line

- **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**

Applications: SELDI Organ-by-Cell-Line

- **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**

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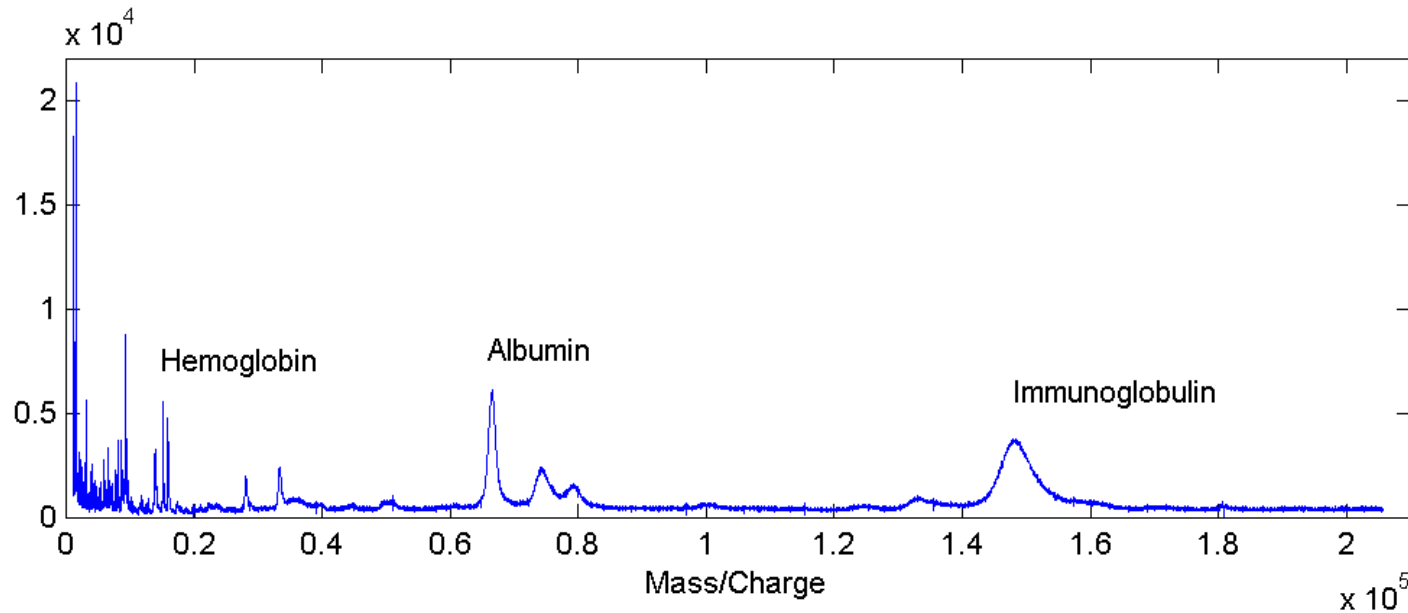
- **MALDI-TOF Spectrum: observed function**
- **$g(t)$ = intensity of spectrum at m/z value t**
- **Intensity at peak (roughly) estimates the abundance of some protein with molecular weight of t Daltons**

FDA: Functional Mixed Model

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.

Functional Mixed Model (FMM)

- **Idea:** Relate *functional response* to set of scalar predictors through *functional coefficients*, while adjusting for possible *correlation between functions* induced by design.

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

$$U_k(t) \sim GP(0, Q)$$
$$E_i(t) \sim GP(0, S)$$

$$\underbrace{Y_i(t)}_{\substack{\text{response} \\ \text{functions}}} = \sum_{j=1}^{p} X_{ij} \underbrace{B_j(t)}_{\substack{\text{fixed effect} \\ \text{function}}} + \sum_{k=1}^{m} Z_{ik} \underbrace{U_k(t)}_{\substack{\text{random effect} \\ \text{functions}}} + \underbrace{E_i(t)}_{\substack{\text{residual error} \\ \text{functions}}}$$

- $B_j(t)$ summarizes partial effect of X_j on $Y(t)$
- $Q(t_1, t_2)$ and $S(t_1, t_2)$ are covariance surfaces on $\mathcal{T} \times \mathcal{T}$ describing the form of the function-function deviations

Wavelet-Based Functional Mixed Models

- **We would like to fit functional mixed models to data like in our motivating examples (high dim., spiky)**

- **Need to find way to build flexibility into Q and S, yet be parsimonious enough to fit to large data sets.**

- **Need to find a way to accomplish adaptive regularization for $B_j(t)$ and $U_j(t)$**

- **Need to find way to compute this model that is computationally efficient enough to apply to extremely large data sets (100's-1000's of curves on grids of 10,000 or more)**

- **Want to be able to perform inference, not just estimation**

- **Key: Use of <u>wavelet space</u> instead of data space to do our modeling.**

Wavelet-Based Functional Mixed Models

- **Idea:** Represent functions by their wavelet expansion

- **Basis function approach:** $y(t) = \Sigma d_{jk}\psi_{jk}(t)$

- **Benefits of Using Wavelet Bases**

  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.

WFMM: 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.**

WFMM: 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.**

1. Project observed functions Y to wavelet space

- **After choice of (1) wavelet basis, (2) boundary correction method, and (3) number of decomposition levels J, 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 (rotate your coordinate axes according to wavelet bases).**

WFMM: Projecting to Wavelet Space

- **The matrix D that results contains the observed wavelet coefficients for our data set**

$$\underset{N \times T}{D} = \begin{bmatrix} \overbrace{c_{1,1}}^{K_1 \text{ coeff}} & \overbrace{d_{1,1}}^{K_1 \text{ coeff}} & \overbrace{d_{1,2}}^{K_2 \text{ coeff}} & \cdots & \overbrace{d_{1,J}}^{K_J \text{ coeff}} \\ c_{2,1} & d_{2,1} & d_{2,2} & \cdots & d_{2,J} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N,1} & d_{N,1} & d_{N,2} & \cdots & d_{N,J} \end{bmatrix}$$

- **With $T=2^J$, Haar wavelets, and full decomposition, we have $K_j=2^{j-1}$. Otherwise not quite true because of adjustments for $T \neq 2^J$ and boundary correction from other wavelet bases with wider support**

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.**

WFMM: Wavelet Space Model

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

Wavelet Representations

Y=DW

B=B*W

U=U*W

E=E*W

WFMM: Wavelet Space Model

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

<u>Wavelet Representations</u>

Y=DW

B=B*W

U=U*W

E=E*W

WFMM: Wavelet Space Model

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

Wavelet Representations

Y=DW

B=B*W

U=U*W

E=E*W

WW′=I

WFMM: Wavelet Space Model

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

<u>Wavelet Representations</u>

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*

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

$$U_k^* \sim MVN(0, Q^*)$$

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

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

- **$E^* = EW'$ is the matrix of wavelet-space residuals**

WFMM: Covariance Assumptions

- **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

- **Other alternatives possible (parent-sib correlation, block thresholding), but add more computational complexity.**

- **We have found independence to be very flexible: limits form of covariance matrices, but enough parameters and degrees of freedom to capture key nonstationary features of these matrices that are typical of our data**
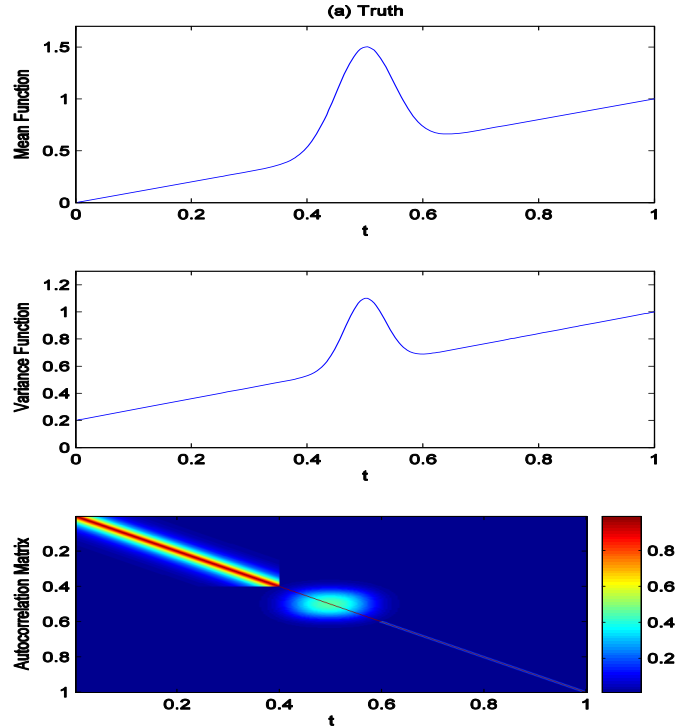
WFMM: Covariance Assumptions

- *Whitening*: autocorrelation of wavelet coefficients tends to die away rapidly across k; d_{jk} approximately uncorrelated (**Johnstone and Silverman 1997**)
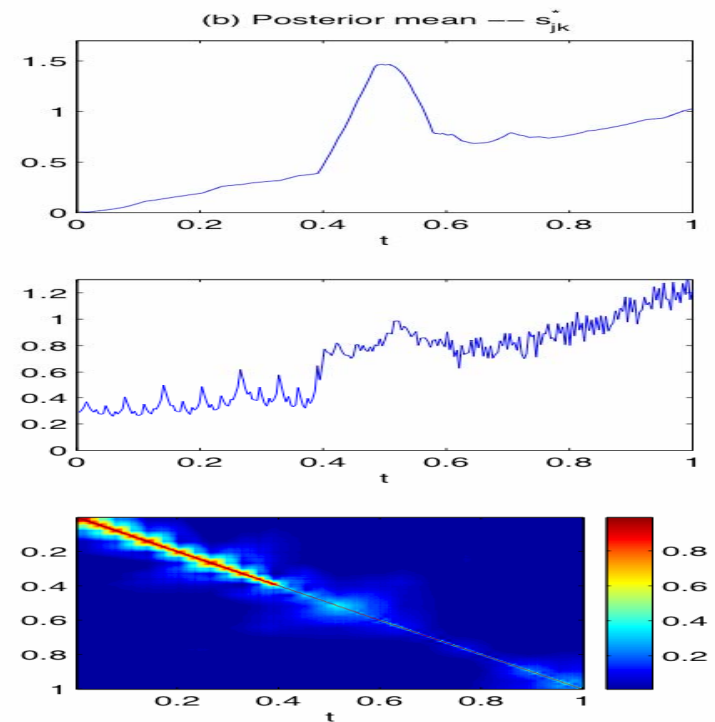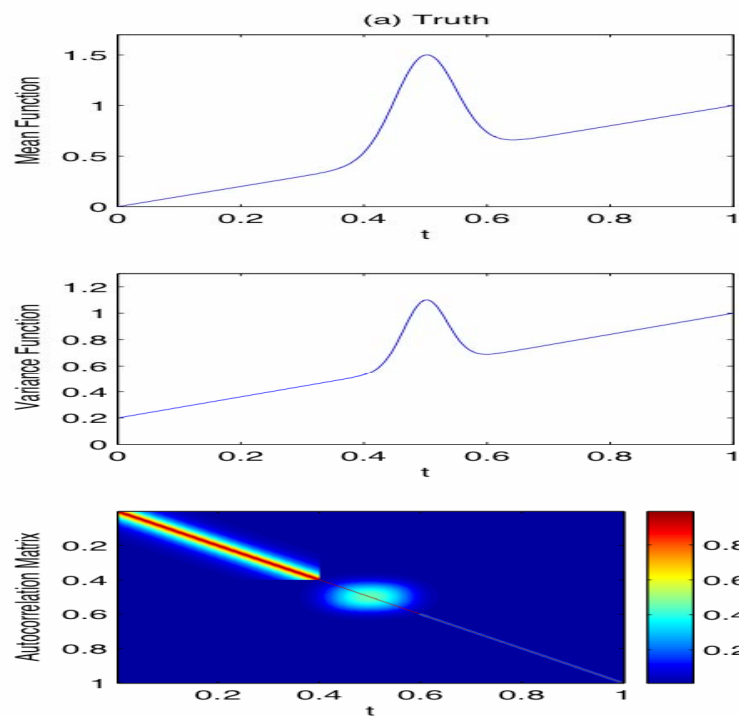
WFMM: Covariance Assumptions

- *Whitening*: autocorrelation of wavelet coefficients tends to die away rapidly across k; d_{jk} approximately uncorrelated (**Johnstone and Silverman 1997**)

WFMM: Covariance Assumptions

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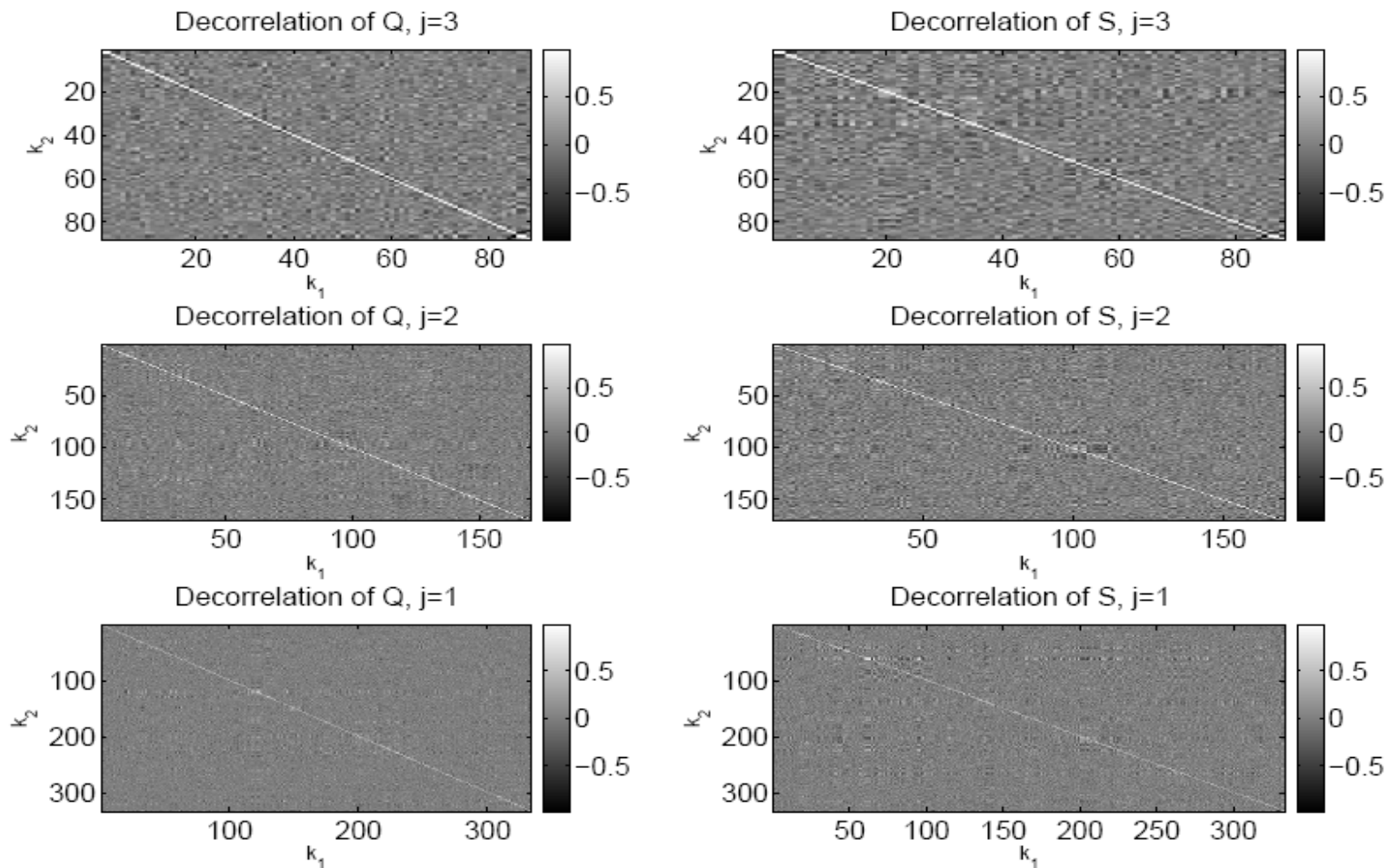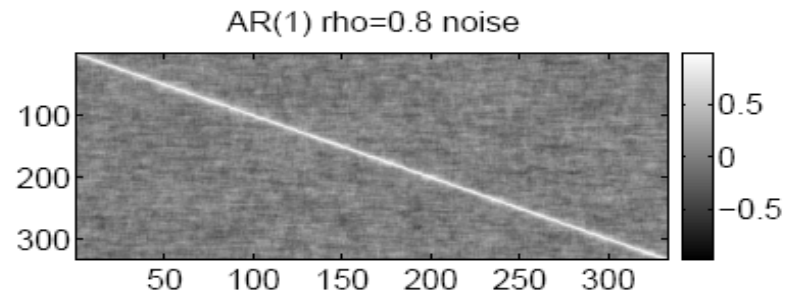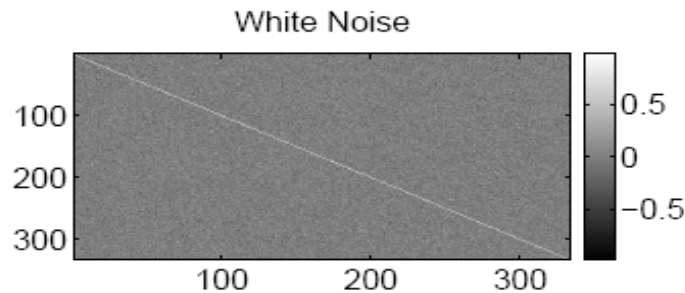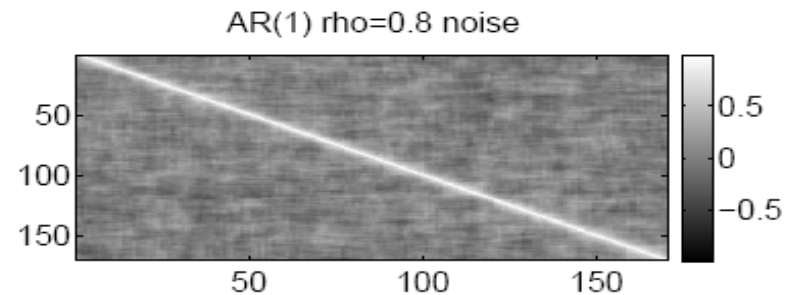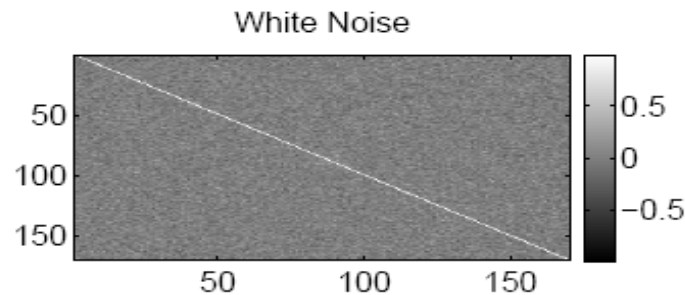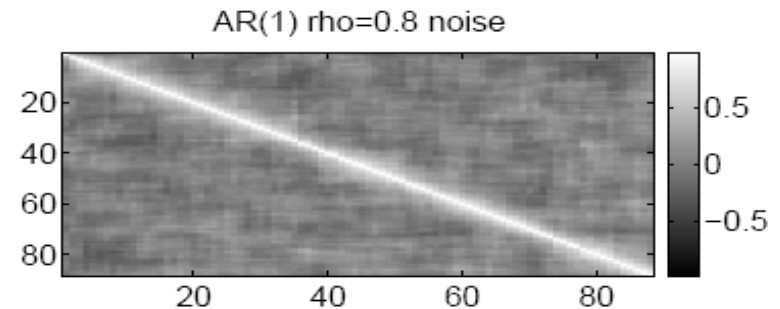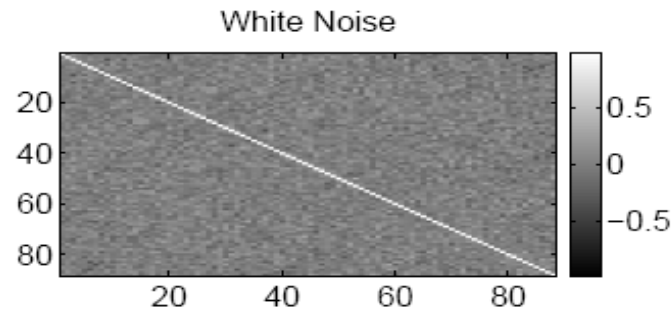
- **True mean**: line plus peak
- **True variance:** increasing in t, with extra var at peak
- **True autocorrelation:** Strong autocorrelation (0.9) at left, weak autocorrelation (0.1) right, extra at peak

WFMM: Covariance Assumptions

- **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**

WFMM: Covariance Assumptions
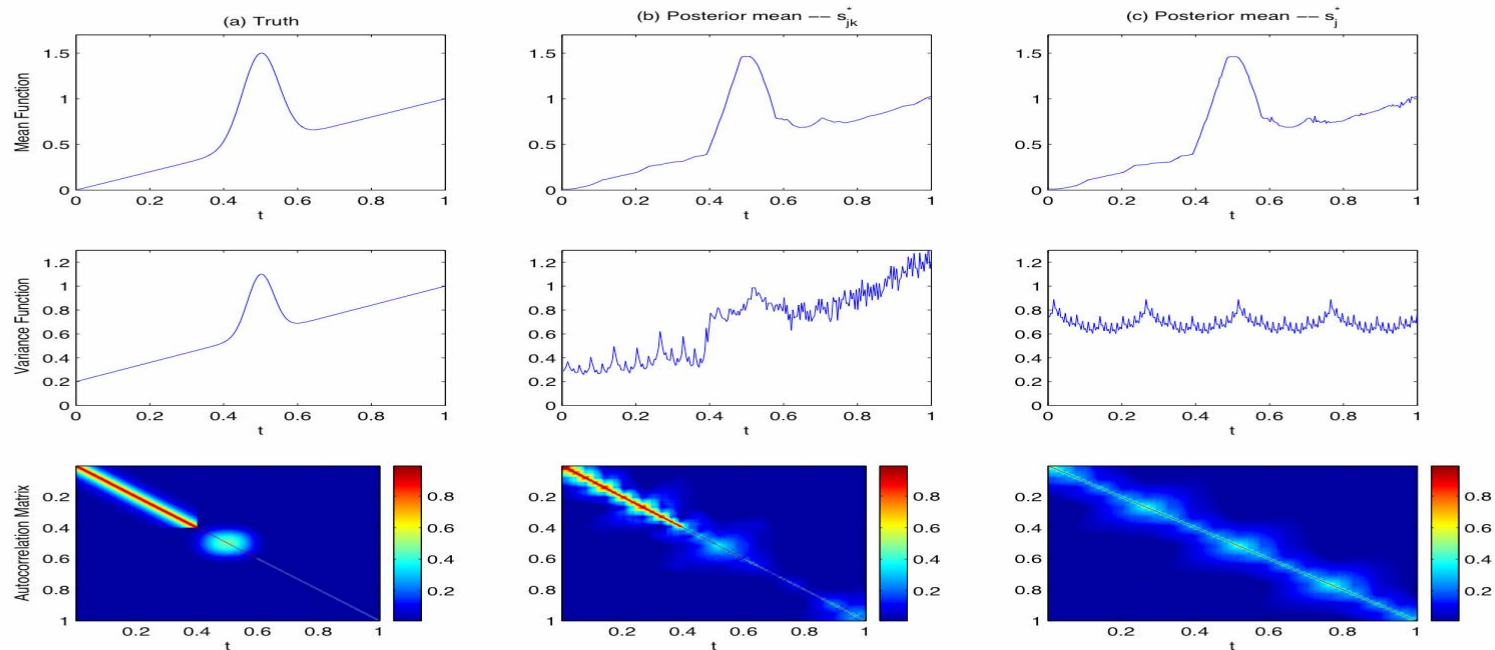
- **Most wavelet regression methods (Fan and Lin 1998, Morris, et al. 2003, Abramovich and Angelini 2007, Antoniadis and Sapatinas 2007) only index variances by scale j, but not location k.**
- **Not flexible enough to capture nonstatioinary covariance features**
- **Unnecessary restriction in multiple function case, since replicate functions allow estimation of separate VC for each (j,k)**

WFMM: Wavelet-Space Model

- **Key benefit of independence assumption: With Q,S diagonal, the columns of the WFMM are now unlinked, and can be modeled separately.**
- **A set of T regular linear mixed model instead of one HUGE functional mixed model.**
- **Key to <u>computational feasibility</u> (run time AND memory management)**

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

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

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

WFMM: Prior Assumptions

- **Another benefit of wavelet-space modeling is that it provides us a natural mechanism for inducing adaptive regularization for our estimates of the $B_a(t)$**

- **To induce nonlinear shrinkage, we assume a spike-slab prior on B_{ajk}^*:**

$$B_{ajk}^* = \gamma_{ajk}^* N(0, \tau_{aj}^2) + (1 - \gamma_{ajk}^*)\delta_0$$

$$\gamma_{ajk}^* = \text{Bernoulli}(\pi_{aj})$$

- Nonlinearly shrinks B_{ajk}^* towards 0, leading to ***adaptively regularized*** estimates of $B_a(t)$.

- τ_{aj}^2 & π_{aj} are **regularization parameters**

WFMM: Prior Assumptions

- **How to choose smoothing parameters? Not intuitive.**
  - **Antoniadis, Sapatinas, and Silverman (1998)** suggest a parameterization in which π_j decrease exponentially in j, and relate them theoretically to the underlying Besov space
  - **Clyde and George (2000)** estimated these from data using local MLE, in *empirical Bayes* fashion. We follow.

- **Involves iterative (EM) algorithm to estimate the proportion of "large" wavelet coefficients at each level j (π_{ij}), and their associated variances (τ_{ij}), with lower bound on τ_{ij} to avoid excessive linear shrinkage**

- **Similar in principal to idea of estimating smoothing parameters from data in nonparametric regression**

WFMM: Prior Assumptions

- **To complete the specification of the Bayesian model, we need priors on the variance components.**

- **When $P=R=I$, we only need to worry about q_{jk} and s_{jk}**
  - **We use *diffuse* conjugate Inverse Gamma priors**
  - **Center on starting value estimates (MLE's), and set scale parameter such that prior has 1/1000 information of data**
  - **These empirical Bayes priors are by default automatically calculated by the code, which is important since it would not be feasible to set these by hand since there are $(H+1)T$ variance components.**

WFMM: 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

WFMM: Model Fitting

- **Updating fixed effects, B*:** Gibbs step (conjugate spike-slab mixture) for each *a=1, …, p, j=1, …, J, k=1, …, Kj*

$$B^*_{ajk} = \alpha_{ijk} N\left\{\hat{B}^*_{ajk,MLE} LS_{aj}, \hat{V}_{ajk} LS_{aj}\right\} + (1 - \alpha_{ijk}) I_0$$

$$\hat{B}_{ajk,MLE} = \{X_a{}'\Sigma_{jk}^{-1}X_a\}^{-1} X_a{}'\Sigma_{jk}^{-1}(d_{jk} - X_{(-a)}\hat{B}_{(-a)jk,MLE})$$

$$V_{ajk} = \mathrm{Var}(\hat{B}^*_{ajk,MLE}) = (X_a{}'\Sigma_{jk}^{-1}X_a)^{-1} \qquad \Sigma_{jk} = ZPZ'q_{jk} + Rs_{jk}$$

$$LS_{aj} = \underbrace{\left(\frac{T^2_{aj}}{T^2_{aj}+1}\right)}_{\textbf{Linear Shrinkage}}$$

$$\alpha_{ijk} = \underbrace{\mathrm{Pr}\left\{\gamma_{ajk} = 1 \mid \mathbf{d}_{ajk}\right\}}_{\text{Nonlinear Shrinkage}}$$

$$T^2_{aj} = \tau^2_{aj} / V_{ajk}$$

$$\alpha_{ajk} = \Pr\left(\gamma_{ajk} = 1 \mid \mathbf{d}_{ajk}\right) = \frac{O_{ajk}}{O_{ajk} + 1}, \qquad O_{ajk} = \text{Posterior Odds}$$
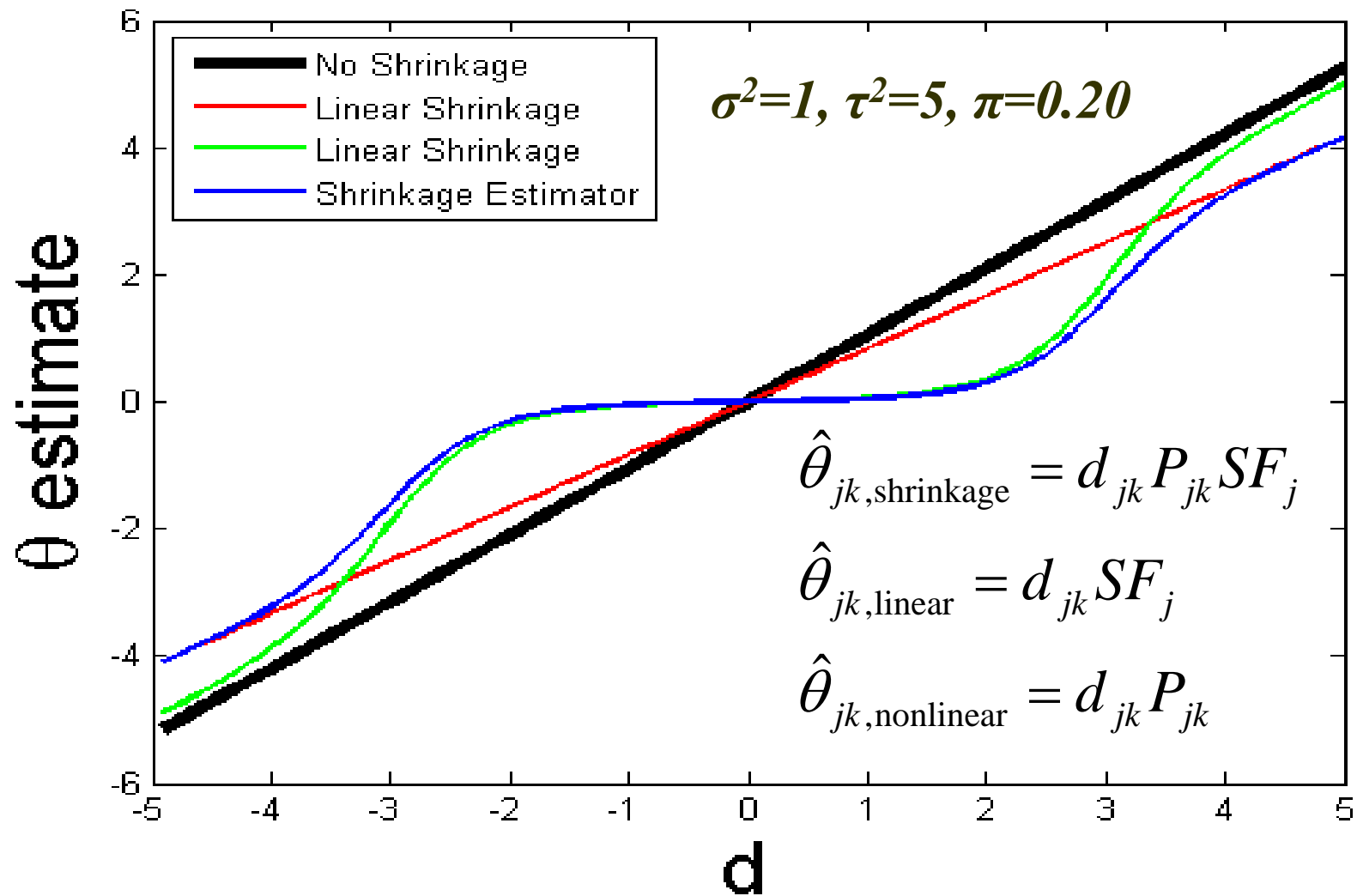
$$\underbrace{O_{ajk}}_{\textbf{Posterior Odds}} = \underbrace{\left(\frac{\pi_{aj}}{1 - \pi_{aj}}\right)}_{\textbf{Prior Odds}} \underbrace{\left(1 + T_{aj}^2\right)^{-1} \exp\left\{\frac{\zeta_{ajk}^2}{2}\left(\frac{T_{aj}^2}{T_{aj}^2 + 1}\right)\right\}}_{\textbf{Bayes Factor}}$$
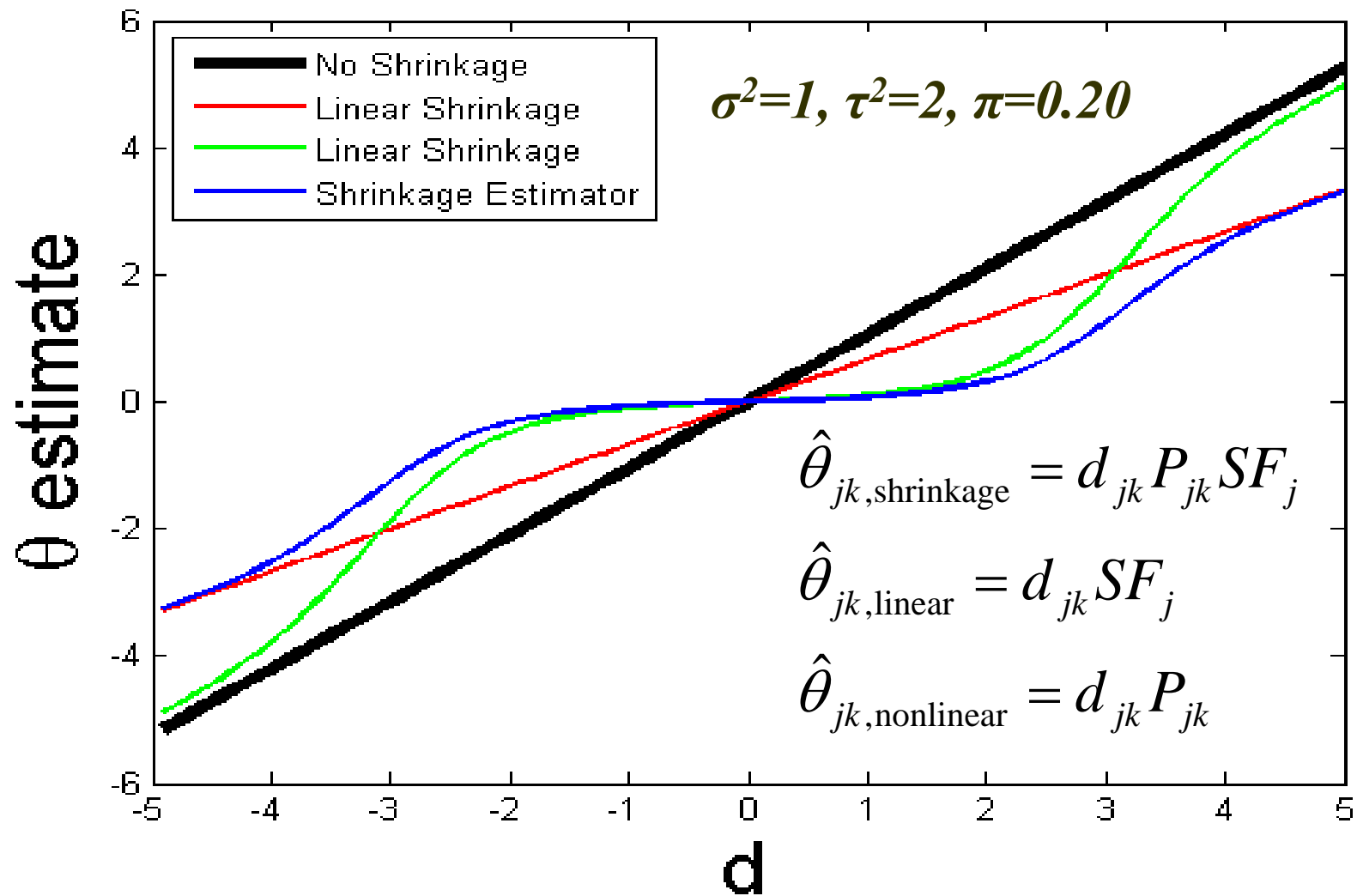
$$\zeta_{ajk} = \hat{B}_{ajk,MLE}^* / \sqrt{V_{ajk}}$$
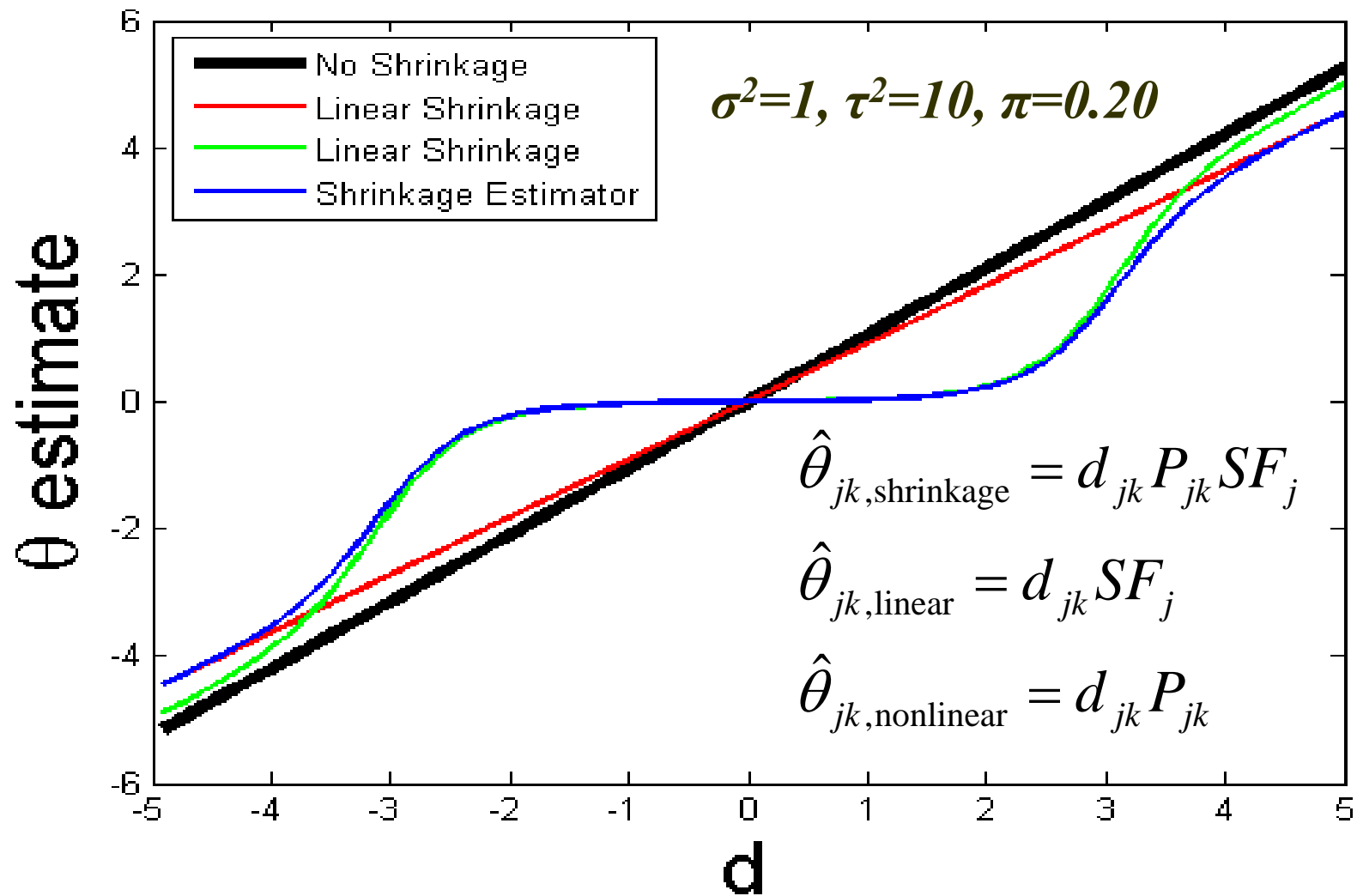
Wavelets: Wavelet Regression

Wavelets: Wavelet Regression

$\sigma^2=1,\ \tau^2=5,\ \pi=0.20$

$$\hat{\theta}_{jk,\text{shrinkage}} = d_{jk}P_{jk}SF_j$$

$$\hat{\theta}_{jk,\text{linear}} = d_{jk}SF_j$$

$$\hat{\theta}_{jk,\text{nonlinear}} = d_{jk}P_{jk}$$

Wavelets: Wavelet Regression

$\sigma^2 = 1,\ \tau^2 = 2,\ \pi = 0.20$

Legend:
- **No Shrinkage** (black)
- Linear Shrinkage (red)
- Linear Shrinkage (green)
- Shrinkage Estimator (blue)

$$\hat{\theta}_{jk,\text{shrinkage}} = d_{jk}P_{jk}SF_j$$

$$\hat{\theta}_{jk,\text{linear}} = d_{jk}SF_j$$

$$\hat{\theta}_{jk,\text{nonlinear}} = d_{jk}P_{jk}$$

WFMM: Model Fitting

- **Updating of variance components Ω:**

- **Could use Gibbs step for q_{jk} and s_{jk}: $f(\Omega|D,B^*,U^*)$**

    – **Problem:** As $q_{jk} \to 0$, $U^*_{jk} \to 0$, get stuck in local mode ("black hole phenomenon")

- **Solution:** *Marginalize* **model by integrating out U^*, then update $f(\Omega|D,B^*)$, then $f(U^*|D,B^*,\Omega)$**

    – Equivalent to updating (Ω,U^*) simultaneously, since $f(U^*,\Omega|D,B^*)$ $=f(\Omega|D,B^*) \times f(U^*|D,B^*,\Omega)$

    – **Another advantage:** If inference on random effects not needed, then there is no need to update them!

    – **Disadvantage:** Can't update q_{jk} and s_{jk} with simple Gibbs steps, since $f(\Omega|D,B^*)$ is not a known distributional form ; instead must use Metropolis-Hastings step

WFMM: Model Fitting

- **Metropolis-Hastings** for q_{jk}, s_{jk}
  - Use *random-walk proposal*: Proposed value for new MCMC iteration $q_{jk}^{(1)}$ is centered at value at previous iteration $q_{jk}^{(0)}$, as $N(q_{jk}^{(0)}, propvar_{jk})$
- **Key is appropriate choice of *propvar$_{jk}$*; if too big, sampler won't move, if too small, will not explore posterior (get stuck in local modes)**
  - Typically found by trial-and-error; not feasible here!
- **We estimate *propvar$_{jk}$* as *1.5*Var(q$_{jk,MLE}$)*, which should be slightly more than variance of the true complete conditional distribution; works well**

WFMM: Model Fitting

- **Updating of random effects, U*: Gibbs step, Normal**

$$f(u_{jk}^* \mid d_{jk}, B_{jk}^*, \Omega) = Normal(M_{jk}, V_{jk})$$

$$V_{jk} = \left\{ \Psi_{jk}^{-1} + (q_{jk}^* \times P)^{-1} \right\}^{-1}$$

$$M_{jk} = \underbrace{V_{jk} \Psi_{jk}^{-1}}_{\text{Linear Shrinkage}} \hat{u}_{jk,MLE}$$

$$\Psi_{jk} = \left\{ Z'(s_{jk}^* \times R)^{-1} Z \right\}^{-1}$$

$$\hat{u}_{jk,MLE} = \Psi_{jk} Z'(s_{jk}^* \times R)^{-1} (d_{jk} - XB_{jk}^*)$$

- **When *P*=*R*=*I*, this step is very quick and trivial**

WFMM: 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.**

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.**

WFMM: Back to Data Space

- **For each fixed effect function *a=1, …, p*, and MCMC sample *g=1, …, G*, construct:**

$$B_a^{*(g)} = \left[B_{a11}^{*(g)} \quad B_{a12}^{*(g)} \quad \cdots \quad B_{aJK_J}^{*(g)} \right]$$

- **Then apply the pyramid-based IDWT to this vector to obtain a posterior sample of $B_a^{(g)}(t)$ on the grid t, since $B_a = B_a^* W'$**

- **In the same way, posterior samples of *U(t)* can be obtained by applying the IDWT to $U^{*(g)}$, if desired.**

- **You can also compute posterior samples for any contrast involving the fixed effect functions, *CB***

WFMM: Back to Data Space

- **We can also obtain posterior samples of Q or S from the wavelet space versions $Q^{*(g)}=diag\{q_{jk}^{(g)}\}$ and $S^{*(g)}=diag\{s_{jk}^{(g)}\}$ using the results of Vannucci and Corradi (1999), based on the fact that $Q=WQ^*W'$**

- **Can be done efficiently, in $O(T^2)$, by applying the 2-d IDWT to $Q^{*(g)}$ and $S^{*(g)}$**

- **This gives us estimates of the within-curve covariance surfaces, The diagonal of these give us estimates of the *variance function***

WFMM: Inference

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

1. Pointwise posterior **credible intervals** for functional effects or contrasts
2. **Posterior probabilities** for pointwise hypotheses of interest
3. **Bayes Factors** for functional testing
4. Can account for multiple testing in identifying significant regions of curves by controlling the **expected Bayesian FDR**
5. Can compute **posterior predictive distributions**, which can be used for model-checking or classification.

WFMM: Inference

- ## Pointwise Inference:

- **Given posterior samples of $B_a(t)$, can construct posterior mean (shrinkage) estimates for $B_a(t)$, or any contrast of fixed effect functions CB**

- **Can compute quantiles of posterior distribution for $B_a(t)$ for each t, which can be used for pointwise credible intervals. $Pr\{Q_{j,\alpha/2}(t) < B_a(t) < Q_{j,1-\alpha/2}(t)|D\} = 1-\alpha$**

- **Can compute probability of having a certain minimum effect size: $p_a(t) = Pr\{|B_a(t)| > \delta|D\}$, e.g. log: $\delta=1 \Leftrightarrow$ 2-fold**
  - **Takes both Statistical and Clinical significance into account**

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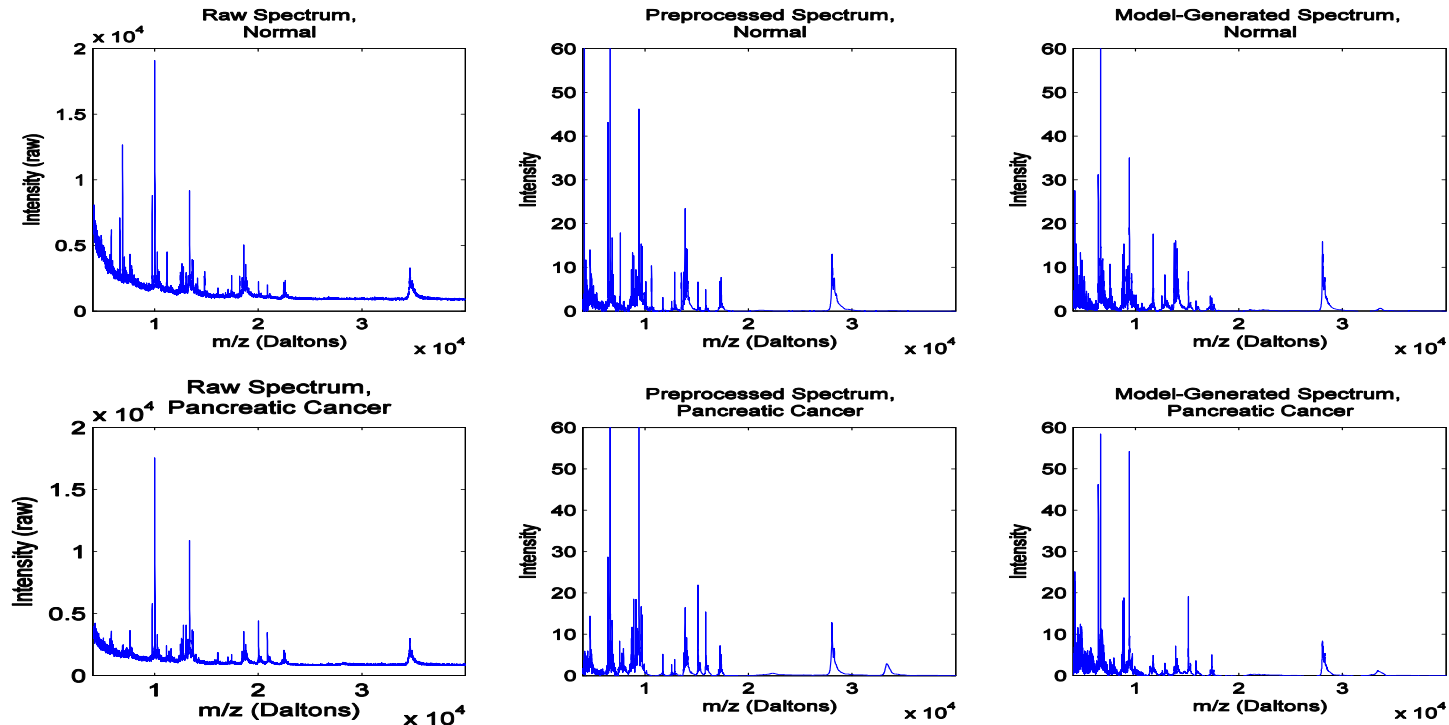
- **Draws of spectra from posterior predictive distribution yield data that looks like real MALDI data (3rd column), indicating reasonable model fit.**

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

(a) School E
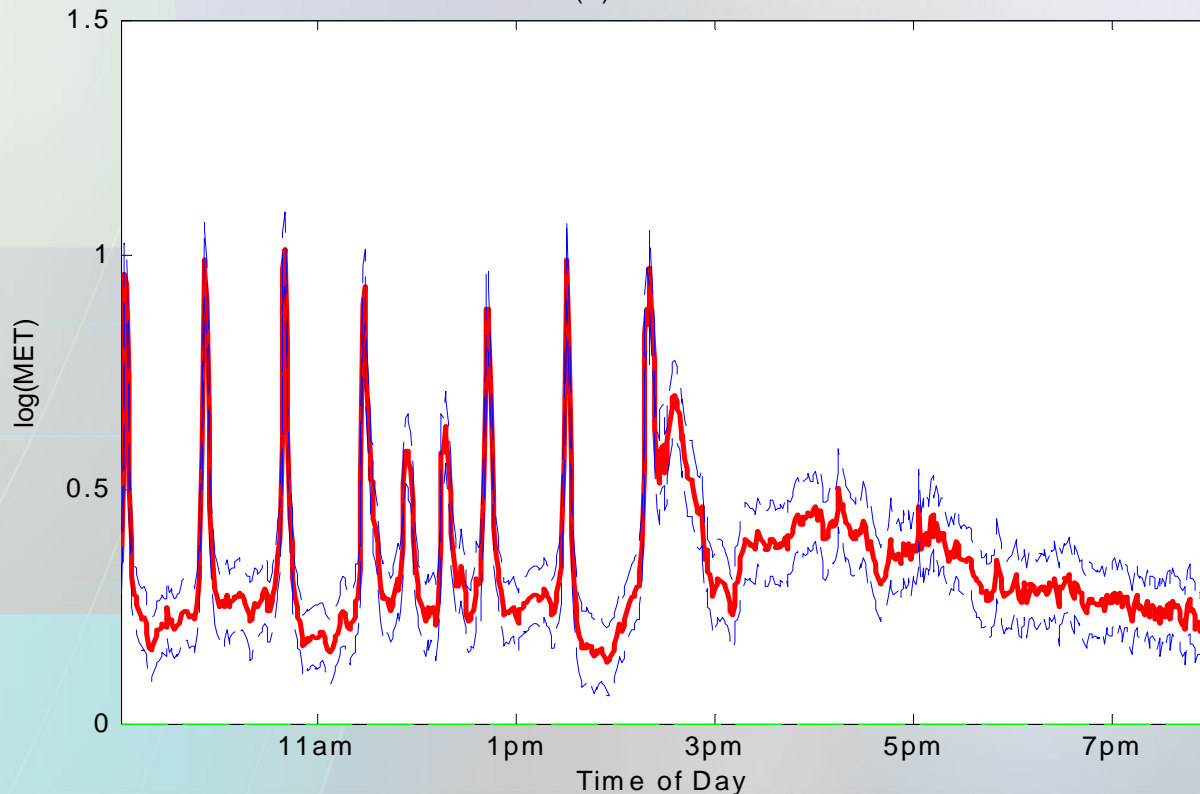
- **Able to preserve dominant spikes in mean curves, if present**

Posterior Mean for Overall Mean Spectrum with Peaks, Organ by Cell Line Example
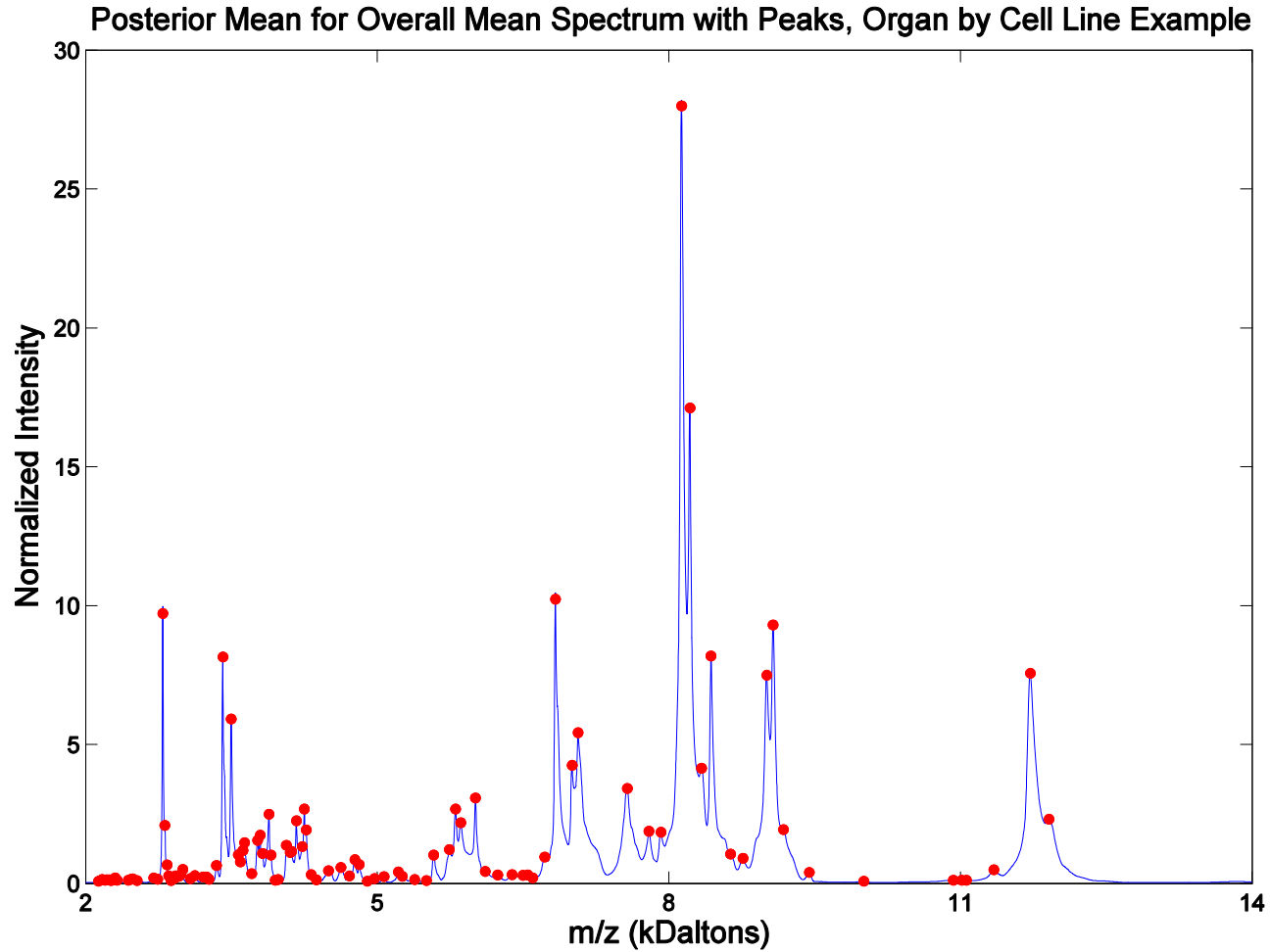
WFMM: Properties

- **Posterior samples/estimates of random effect functions $U_j(t)$ also appear to be *adaptively regularized***
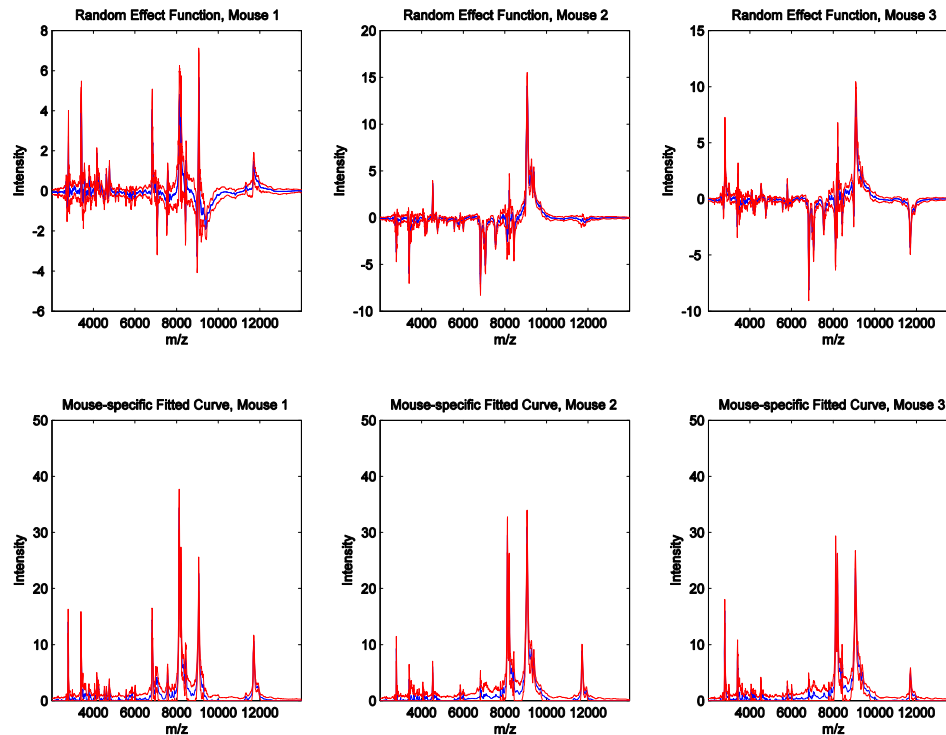
- **Able to preserve spikes in random effect functions, as well**

- **Important for estimation of random effect functions AND for valid inference on fixed effect functions.**

WFMM: Properties

- **How can the random effects be adaptively regularized when they have a Gaussian prior (with linear shrink)?**

  - $U^*_{jk} \sim N(0, q_{jk})$

- **However**, note that each wavelet coefficient has its own variance component q_{jk}, which determines the amount of shrinkage from $u^*_{jk,MLE}$ towards **0** (smaller q relative to s, more shrinkage towards **0**)

  - Since (j,k) corresponding to strong signals in the random effect functions will tend to be larger, q_{jk} will also tend to be larger, leading to less shrinkage.

  - Unimportant (j,k) will tend to have small q_{jk}, and thus lead to more shrinkage

  - Although linear for each wavelet coefficient, the shrinkage is nonlinear when viewed for the entire set of wavelet coefficients together.

- **This same dynamic allows very smooth random effect estimates when the data supports this notion.**

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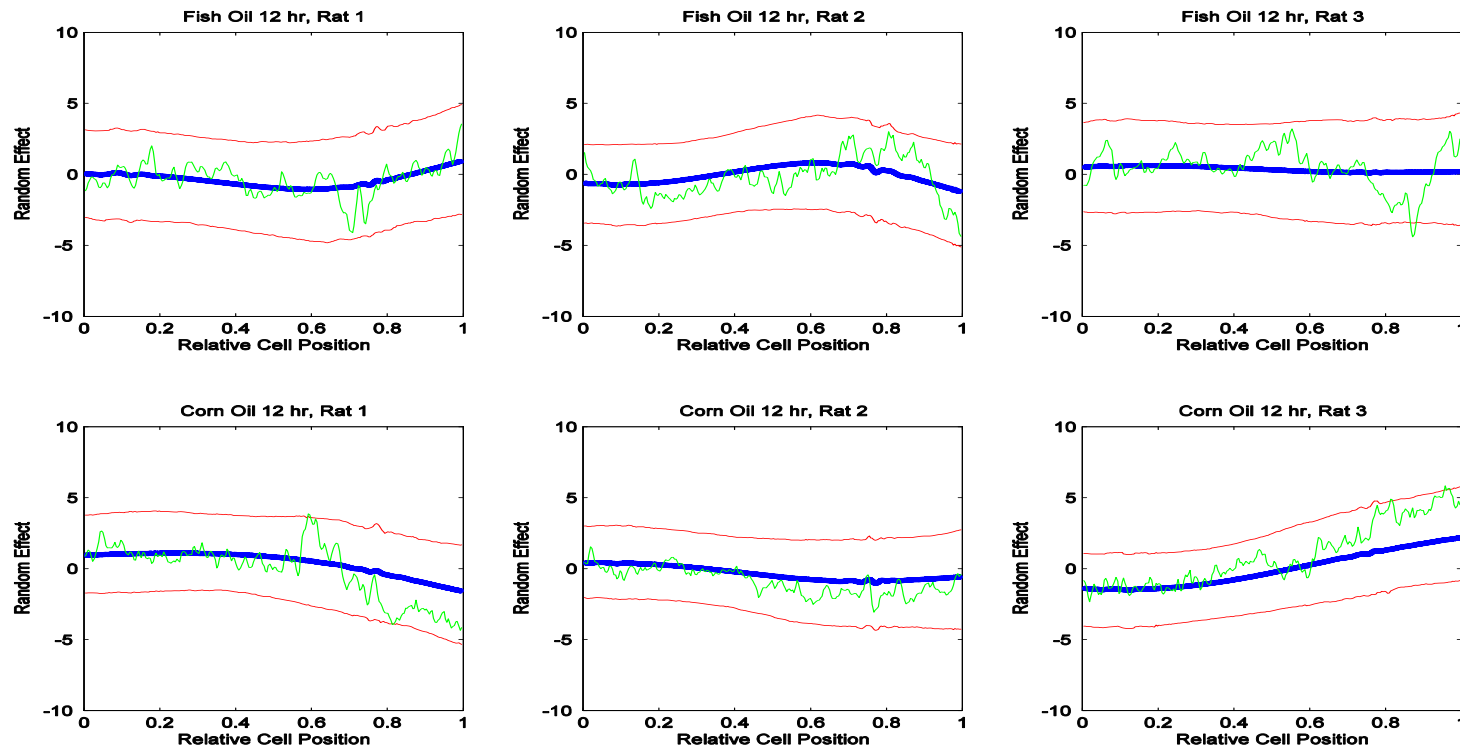
WFMM: Inference

- **Multiple testing** problem is inherent in pointwise inference: T positions, T tests

- **One Solution**: perform **functional test** of $H_0 : B_a(t) = 0$
  - Popular approach for some researchers (FANOVA)
  - Not as interesting to me: key is for what t we have $B_a(t)$ differ from zero, and by how much.
  - We can do functional testing in WFMM using **Bayes Factors**

- **Another Solution**: Apply some sort of multiple testing adjustment to the inference. Here we will look at applying the *False Discovery Rate* (**FDR**) ideas popular in microarrays to this pointwise functional setting.

- **Adjusting for Multiple testing**
- **Classic approach: Bonferroni**
  T tests, use α/T as sig. level
  - **Controls *experiment-wise error rate***
    Pr(at least one false positive)$< \alpha$
  - **Too conservative for some fields (e.g. microarrays), so alternative criterion devised**
- ***False Discovery Rate* (FDR):**
  Benjamini and Hochberg (1995)
  Controls proportion of "discoveries" that are false to be no more than α

$Truth$

	H_0	H_1
H_0	A	B
H_1	C	D

Decision

$$FDR = \frac{C}{C+D}$$

WFMM: Inference

- **Many procedures have been developed for controlling FDR**
  - **Benjamini and Hochberg (1995), Yekutieli and Benjamin (1999), Benjamini and Liu (1999), Storey (2002), Storey (2003), Genovese and Wasserman (2002), Ishwaran and Rao (2003), Pounds and Morris (2003), Efron (2004), Newton (2004), Pounds and Cheng (2004)**

- **Most methods follow one of three approaches:**
  1. **Fit mixture model on tests stats (known distn under H_0)**
  2. **Fit mixture model on p-values $\{U(0,1)$ under $H_0\}$**
  3. **Fit Bayesian mixture model, with point mass prior on H_0**

- **Methods can either set limit on *global FDR* (α), or on *local FDR* for each statistic *i* (Storey 2003)**
  - q_i=min $Pr(H_0$ true for gene i|Gene i in $RR)$
  - **Note:** p-value is Pr(Gene i in RR|H_0 true for statistic i)

Histogram

Global: $\alpha=C/(C+D)$

Local: $q(pv)=c/(c+d)$
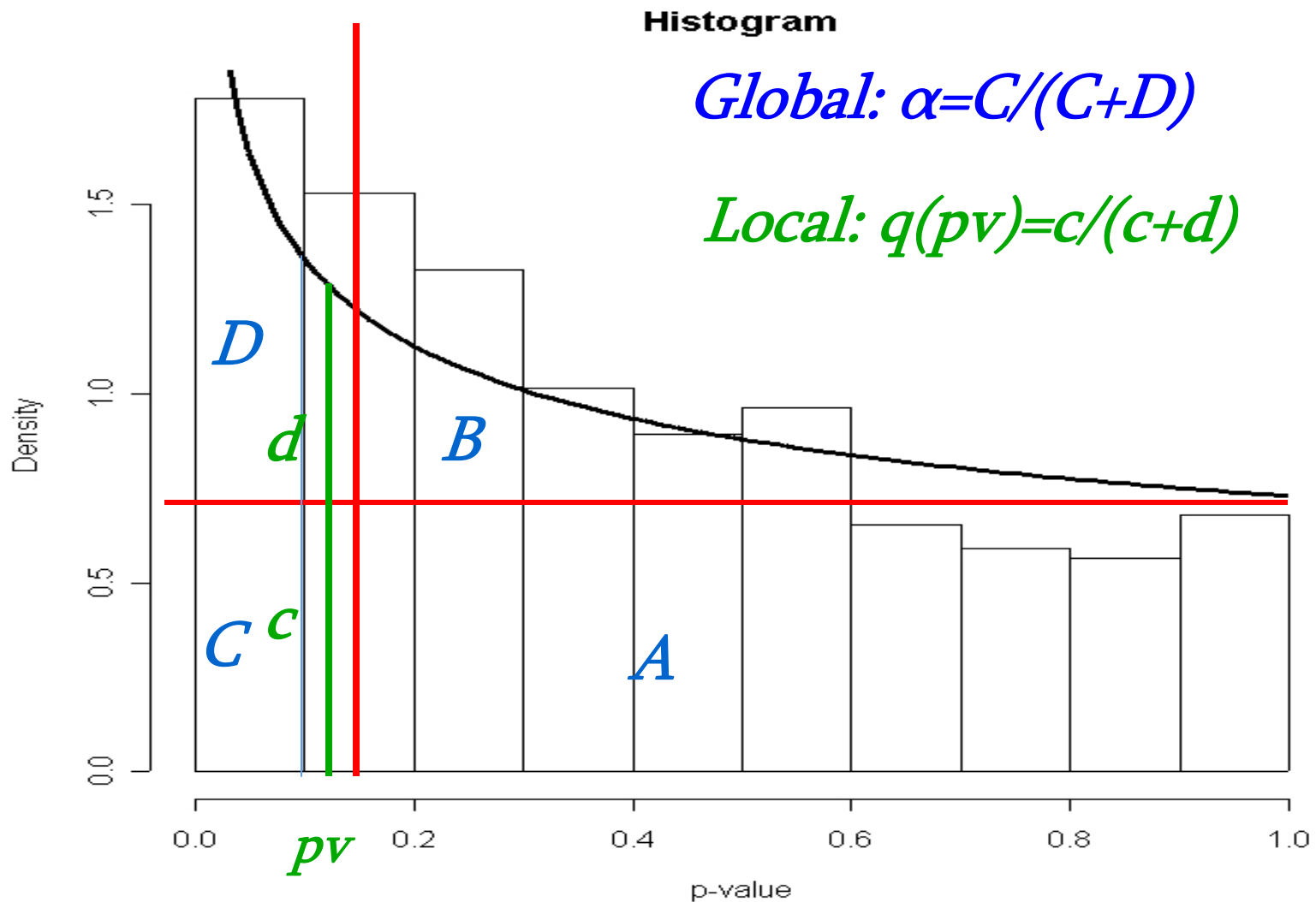
WFMM: Inference

- **<u>Bayesian FDR</u> from posterior probabilities**
  - Posterior probs of effect sizes: $p_i = Pr\{|B_i| < \delta | Y\}$, $i = 1, \ldots, T$
- If we define a "discovery" to be effect size at least δ (H_{0i}: $|B_i| < \delta$), then p_i is a *Bayesian local FDR*
- To control the *global expected Bayesian FDR*
  1. Sort in ascending order of p_i $\{p_{(i)}, i = 1, \ldots, T\}$
  2. Identify cutpoint φ_α on posterior probabilities that controls expected Bayesian FDR to be $\leq \alpha$
     $\varphi_\alpha = p_{(\lambda)}$, where
     $$\lambda = \max\left[i^* : \sum_{i=1}^{i^*} p_i \leq i^* \alpha \right]$$
  3. Flag the set of statistics $\{i: p_i \leq \varphi_\alpha\}$ as significant (According to model, we expect only α to be false pos.)

WFMM: Inference

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.**

WFMM: Inference

FDR for pointwise functional inference:

- **Substituting Lebesgue measure for counting measures, we can apply FDR ideas to pointwise inference:**

- **From MCMC samples, we have the quantities:**
  $p_{al}=\text{Pr}\{|B_a(t_l)|>\delta|\mathbf{Y}\}$ **for** $l=1, \ldots, T$

- **Using** $H_{0t}:|B_a(t)|<\delta$, p_{al} **is a** *Bayesian local FDR* **for fixed effect function** *a* **at position** t_l **within the curve.**

- **As described above, we can find a cutpoint** φ_α **on the** p_{al} **that controls the global average Bayesian FDR<α**

$$\alpha > \frac{|\,t\,|: t \text{ flagged falsely}\,|\,B_a(t)\,|< \delta}{|\,t\,|: t \text{ flagged as "significant"}}$$

Applications: SELDI Organ-by-Cell-Line

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

$$\log_2\{Y_i(t)\} = \sum_{j=1}^{5} X_{ij}B_j(t) + \sum_{k=1}^{16} Z_{ik}U_k(t) + E_i(t)$$

- $X_{i1}=1$ if lung/A375P, 0 ow $X_{i2}=1$ for brain/**A375P**, 0 ow
  $X_{i3}= 1$ if lung/PCMM2 $X_{i4}=1$ for brain/PCMM2
  $X_{i5}= 1$ if high laser intensity, -1 if low

- $B_j(t)$ = **overall mean** spectrum for treatment group $j=1,2,3,4$

  $B_5(t)$ = **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.
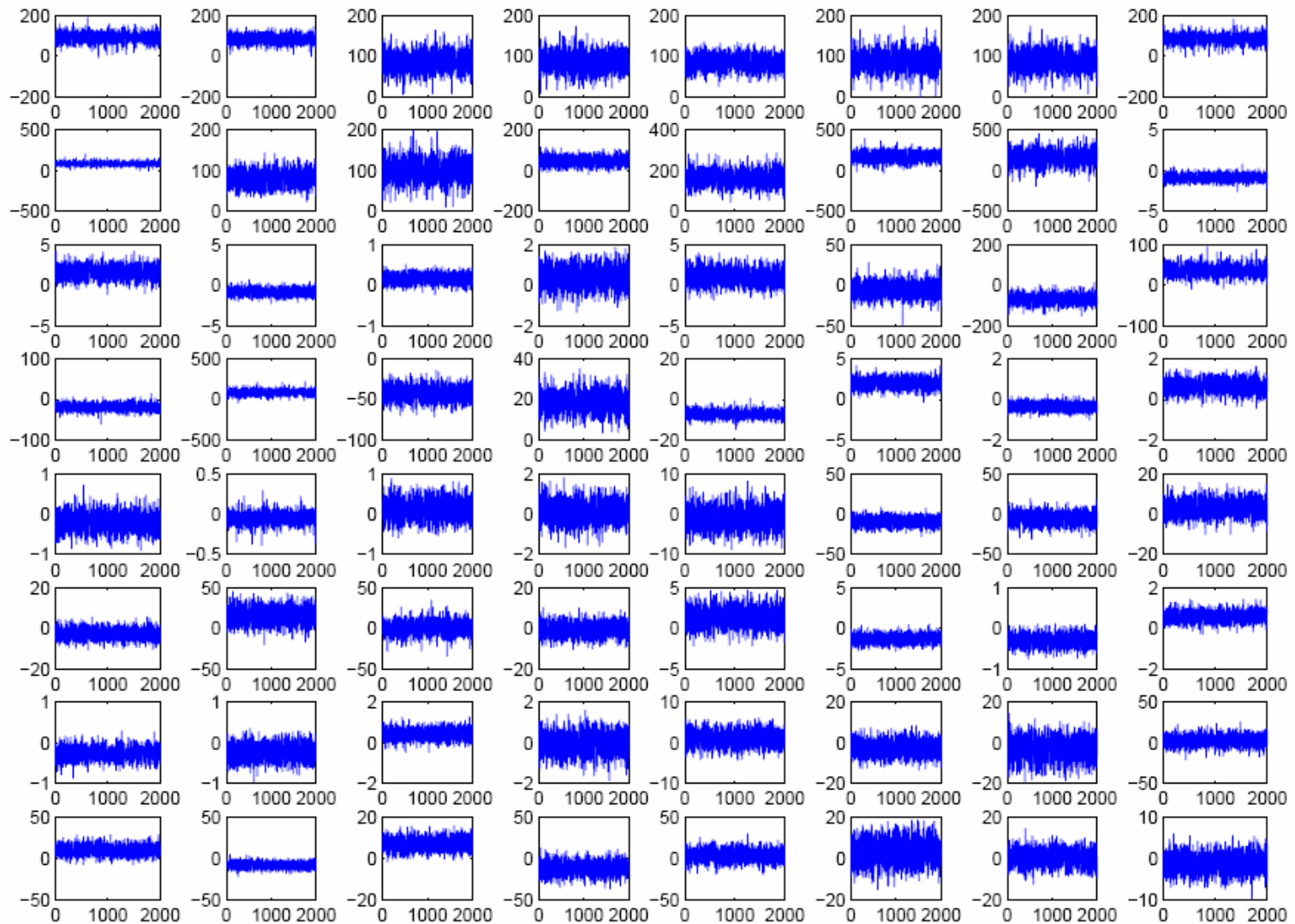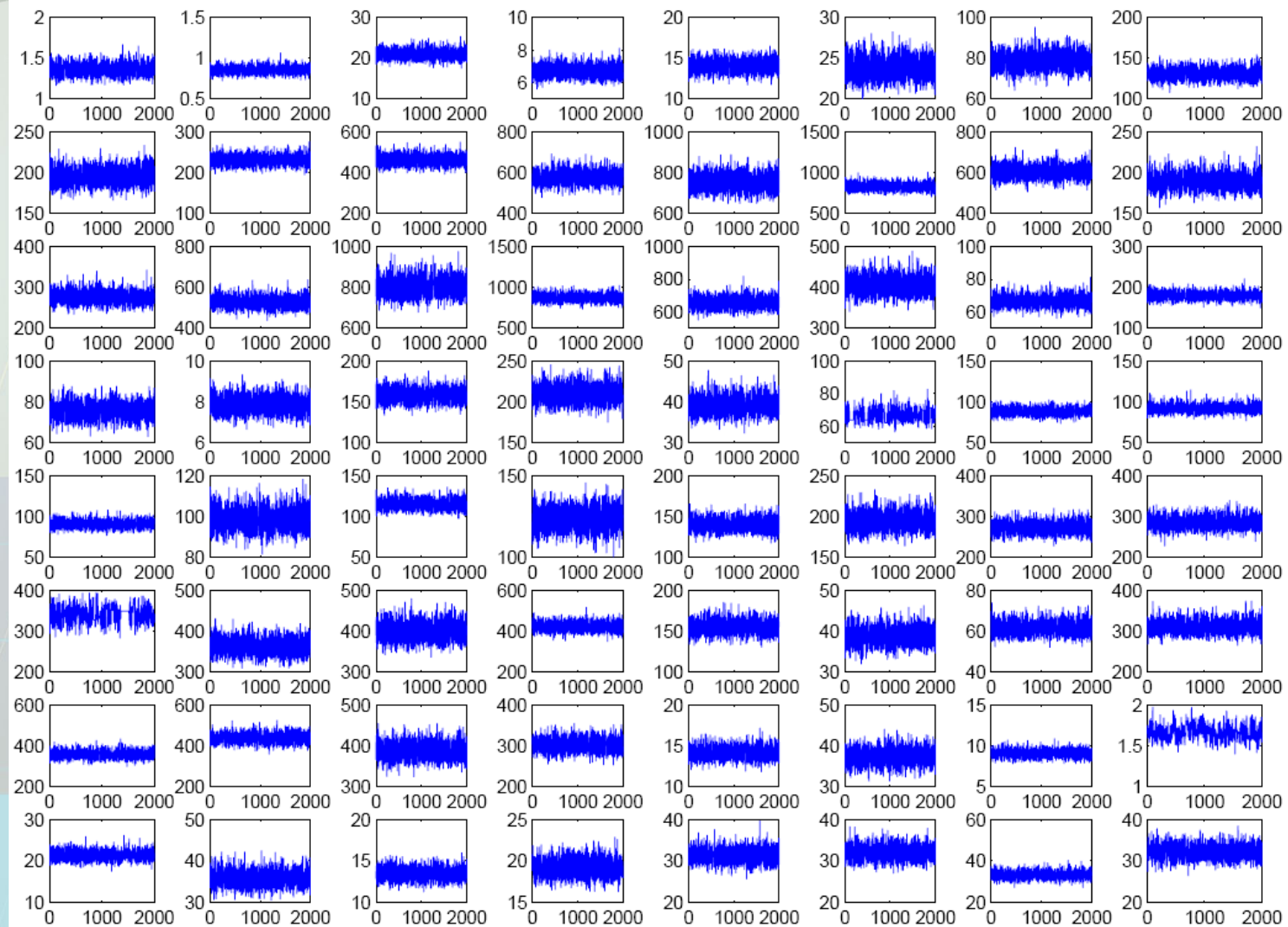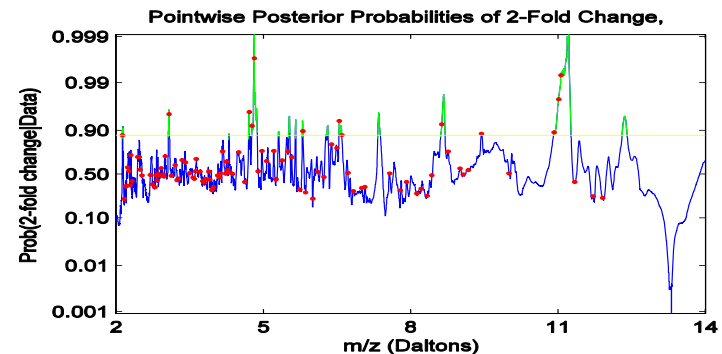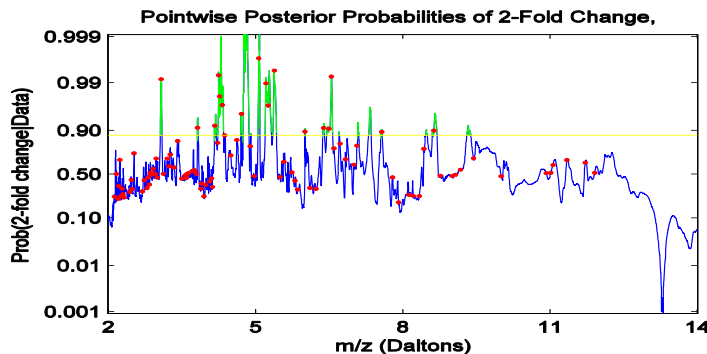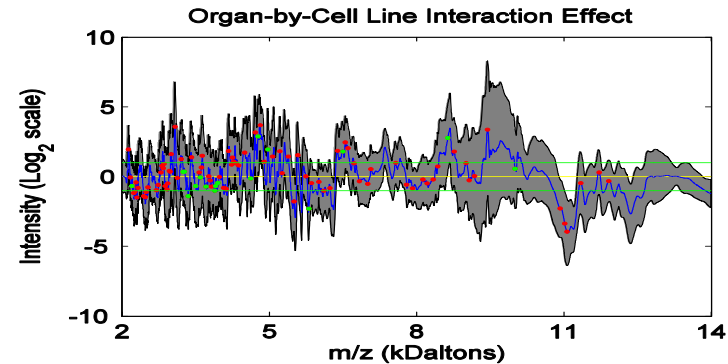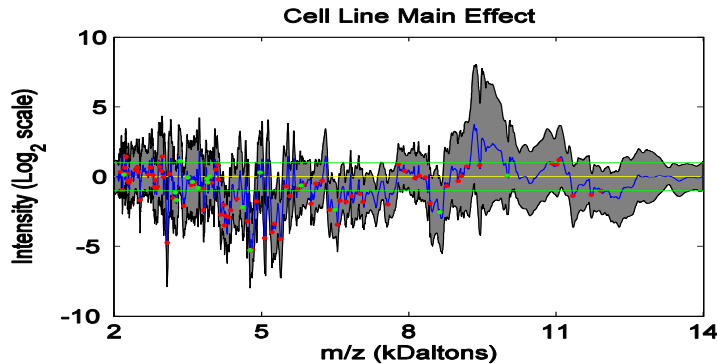
- **Using $\alpha=0.05$, $\delta=1$ (2-fold expression on $\log_2$ scale), we flag a number of spectral regions.**

Applications: SELDI Organ-by-Cell-Line

Table 1: Selected flagged regions from organ by cell line example. Location of selected region (in Daltons per coulomb) is given, along with which effect was deemed significant, estimated maximum fold change difference within the region, and a description of the effect. These effects comprise all those with $p_l > 0.9995$.

Region	Effect type	max FC	Comment
3866.3-3971.3	organ	1/93.9	Only in brain-injected mice
3658.3-3739.0	organ	1/118.5	Only in brain-injected mice
9902.6-10044.0	organ	46.1	Only in lung-injected mice
4762.2-4874.8	interaction	1/13.7	PC3MM2>A375P, especially brain
4748.2-4868.3	cell-line	1/39.7	PC3MM2>A375P
3743.4-3565.3	organ	1/35.0	Brain>Lung
4952.6-5008.2	organ	1/32.8	Brain>Lung
4519.9-4697.5	organ	27.5	Lung>Brain
5051.3-5093.3	cell-line	1/23.5	PC3MM2>A375P
3993.4-4061.3	organ	21.0	Lung>Brain (on upslope of peak)
10912-11269	interaction	1/16.4	Brain>Lung for A375P only

Applications: SELDI Organ-by-Cell-Line
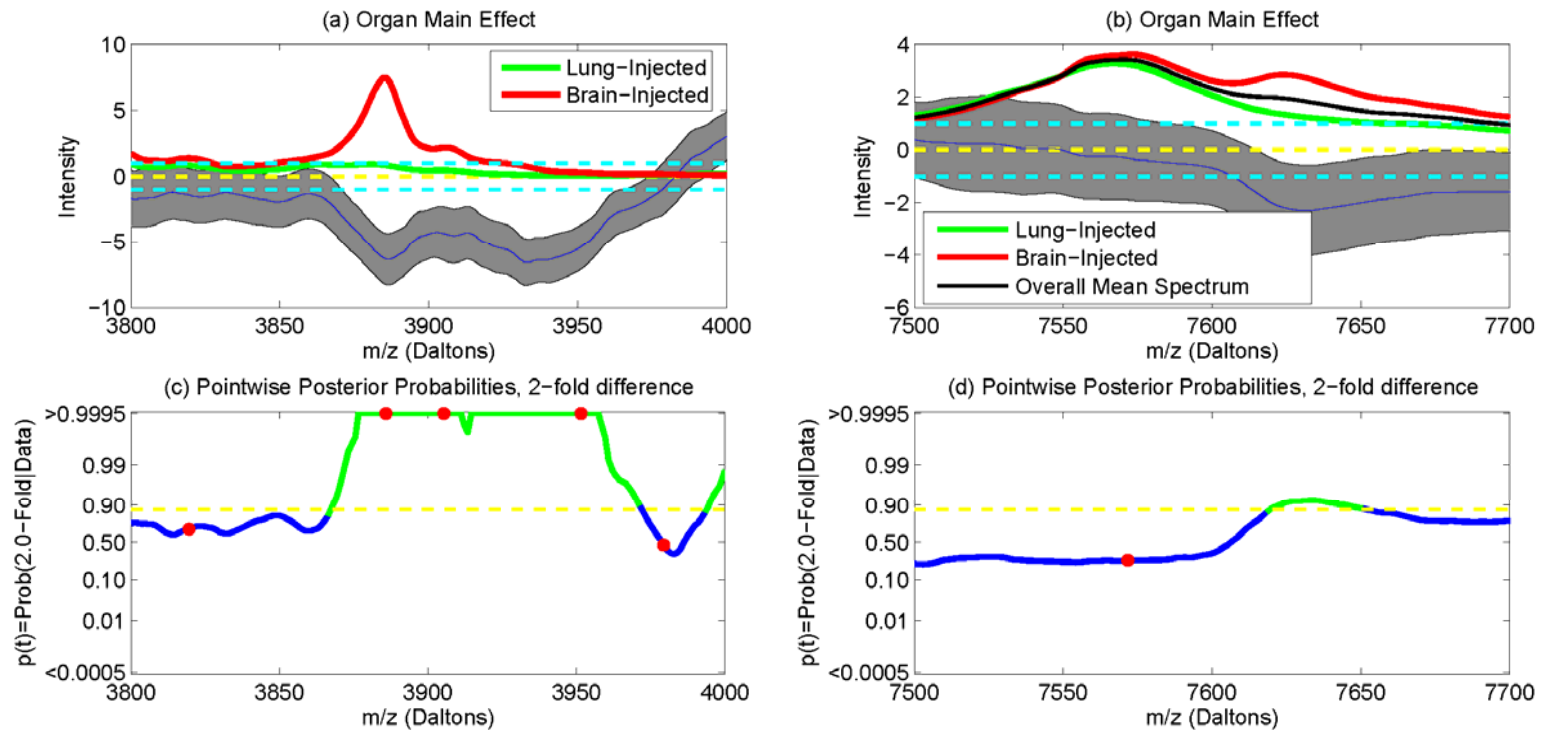

- **3900 D (~100-fold) (CGRP-II): dilates blood vessels in brain**
- **7620 D (~5-fold) (neurogranin): active in synaptic modeling in brain (Not detected as peak)**
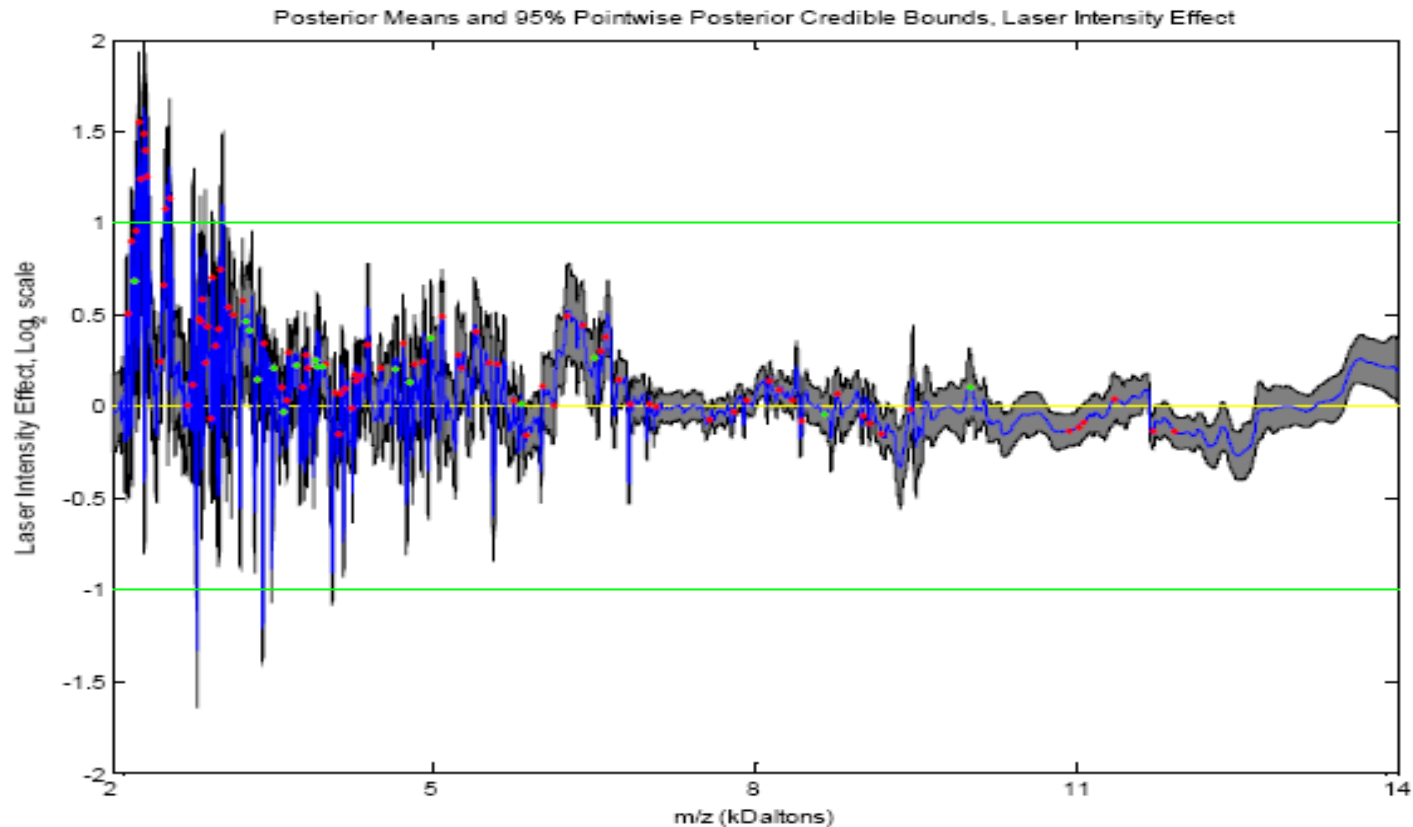
Supplementary Figure 7: *Laser Intensity Effect, Organ-by-Cell-Line Example.* Plot of the posterior mean of the laser intensity effect function (blue line), along with 95% pointwise posterior bounds (gray shaded regions). The green lines mark 2-fold expression differences. The green/red dots indicate detected peaks that were/were not flagged as significant for either the organ, cell-line, or interaction main effects.
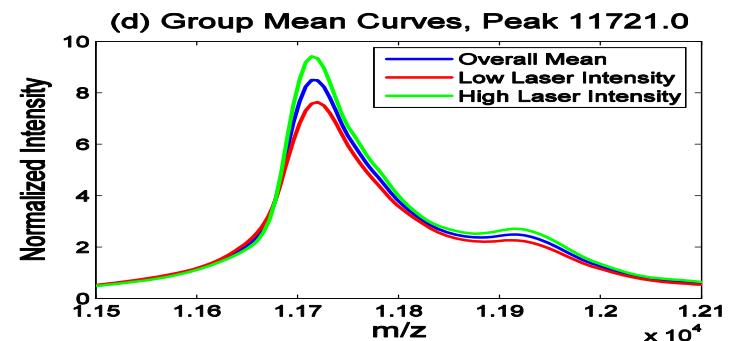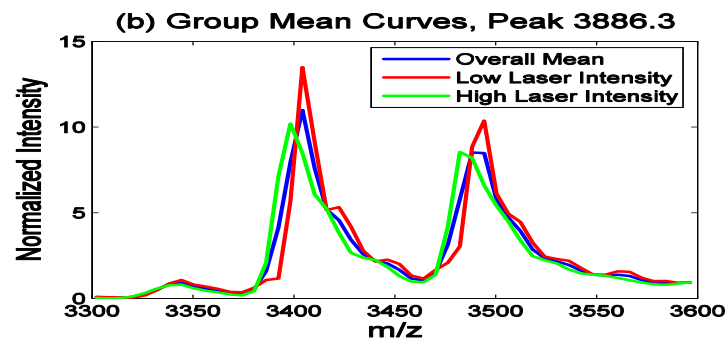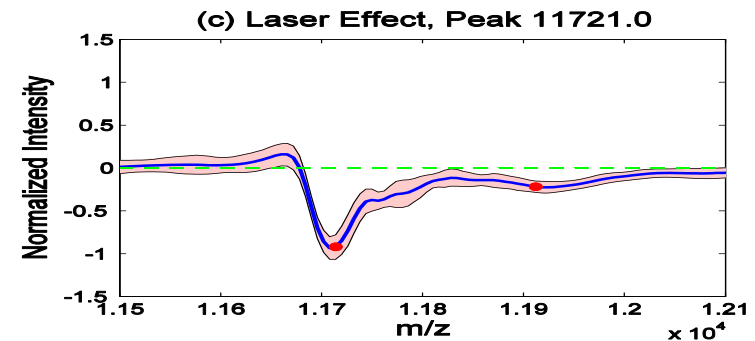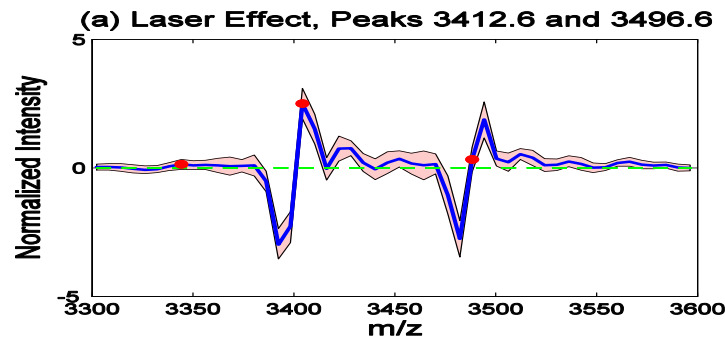
- **Inclusion of nonparametric functional laser intensity effect is able to adjust for systematic differences in the *x* and *y* axes between laser intensity scans**

Conclusion

- **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.

Conclusion

- **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 has been 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*, **online.**

- **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