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 $\rightarrow$ mRNA $\rightarrow$ 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
- $g(t) = \text{intensity of spectrum at } m/z \text{ value } t$
- Intensity at peak (roughly) estimates the abundance of some protein with molecular weight of $t$ Daltons

7/7/2006

http://biostatistics.mdanderson.org/Morris
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

![Graphs showing activity levels over time for Child 1 and Child 2, with data from Day 1 and Day 2.](http://biostatistics.mdanderson.org/Morris)
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)
**Linear Mixed Models**

Linear Mixed Model (Laird and Ware, 1982):

\[
Y = X\beta + Zu + e
\]

- **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 closed interval $\mathcal{T}$.

\[
Y(t) = \underbrace{X \ B(t)}_{N \times p} + \underbrace{Z \ U(t)}_{N \times m} + \underbrace{E(t)}_{N \times m}
\]

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

- $P$ 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
- $X_{i3} = X_1 \times X_2$. $X_{i4} = 1$ for low laser intensity, -1 high.

- $B_0(t)$ = overall mean spectrum
- $B_1(t)$ = organ main effect function
- $B_2(t)$ = cell-line main effect
- $B_3(t)$ = org x cell-line int function
- $B_4(t)$ = laser intensity effect function

- $Z_{ik} = 1$ if spectrum $i$ is from mouse $k$ ($k=1, \ldots, 16$)

- $U_k(t)$ is random effect function for mouse $k$. 
Discrete Version of FMM

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

\[
Y_{N \times T} = X_{N \times p} B_{p \times T} + Z_{N \times m} U_{m \times T} + E_{N \times T}
\]

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

- $U$ and $E$ follow the Matrix Normal distn.
  - $U \sim MN(P, Q)$ implies $\text{Cov}\{U_{ij}, U_{ij'}\} = P_{ii'} Q_{jj'}$

- $P$ 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
- 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

\[ g(t) = \sum_{j,k} d_{jk} \psi_{jk}(t) \]

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

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

- **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:**

  \[ d = y W' \]

  \[ W' = T\text{-by-}T\] 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

Regularization by Local Linear Smoothing

\[ x \]
\[ y \]

\( \text{Spar}=0.20 \)
Adaptive Regularization

Regularization by Local Linear Smoothing

![Graph showing adaptive regularization with different spans](http://biostatistics.mdanderson.org/Morris)
Adaptive Regularization

Regularization by Local Linear Smoothing

Adaptive Regularization by Wavelet Shrinkage
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

\[
D = Y \times W'
\]

- 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

\( \mathbf{D} \): 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 \( k \)

\[
\begin{bmatrix}
\mathbf{D} & \mathbf{N} \times p \\
\mathbf{X} & \mathbf{B}^* \\
\mathbf{N} \times m & \mathbf{Z} \\
\mathbf{U}^* & \mathbf{E}^* \\
\mathbf{N} \times T & \mathbf{N} \times T
\end{bmatrix}
= \begin{bmatrix}
\mathbf{X} \mathbf{B}^* + \mathbf{Z} \mathbf{U}^* + \mathbf{E}^* \\
\mathbf{N} \times p \times T \\
\mathbf{p} \times m \times T \\
\mathbf{N} \times T
\end{bmatrix}
\]

- \( \mathbf{B}^* = \mathbf{BW}' \) & \( \mathbf{U}^* = \mathbf{UW}' \): Rows contain wavelet coefficients for the fixed and random effect functions,
- \( \mathbf{E}^* = \mathbf{EW}' \) is the matrix of wavelet-space residuals
- \( \mathbf{Q}^* = \mathbf{WQW}' \) and \( \mathbf{S}^* = \mathbf{WSW}' \) model the covariance structure between wavelet coefficients for a given function.
- \( \mathbf{P, Q}, \mathbf{R} \) and \( \mathbf{S}^* \) 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

- 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 $\beta_{ijk}^*$:

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

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

- Nonlinearly shrinks $B_{ijk}^*$ towards 0, leading to adaptively regularized estimates of $B_i(t)$.
- $\tau_{ij}$ & $\pi_{ij}$ 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 $\Omega$ 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 $i=1,...,p$, 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

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

Normalized Intensity vs. m/z (kDaltons)
Adaptive Regularization

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

- Able to preserve spikes in random effect functions, as well

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

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

Given posterior samples of all model quantities, we can perform 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 $\alpha$)

1. Compute pointwise posterior probabilities of effect size of interest being at least $\delta$
   \[ p_{il} = \Pr\{|B_j(t_l)| > \delta | Y\} \text{ for } l=1, \ldots, T \]

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

3. Identify cutpoint $\varphi_\alpha$ on posterior probabilities that controls expected Bayesian FDR to be $\leq \alpha$
   \[ \varphi_\alpha = p_{i(\lambda)}, \text{ where } \lambda = \max \left[ l^* : \sum_{l=1}^{l^*} \{1 - p_{i(l)}\} \leq \alpha \right] \]

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

- Using $\alpha=0.05$, $\delta=1$ (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

7/7/2006  http://biostatistics.mdanderson.org/Morris
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 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

Time of Day

MET

0 2 4 6 8 10 12

9am 12pm 3pm 6pm 8pm

7/7/2006
Predictive Distribution

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

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

\[
\mu_i^{M|O} = \mu_i^M + \sum_i^{M,O} (\sum_i^{O,O})^{-1} (Y_i^O - \mu_i^O) \\
\sum_i^{M|O} = \sum_i^{M,M} - \sum_i^{M,O} (\sum_i^{O,O})^{-1} \sum_i^{O,M}
\]
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

\[
M_i(t) = \begin{cases}
Y_i(t) & \text{if } t \text{ observed} \\
\mu_i(t) & \text{otherwise}
\end{cases}
\]

\[
V_i(t_1, t_2) = \begin{cases}
0 & \text{if either } t_1 \text{ or } t_2 \text{ obs.} \\
\Sigma_{i}^{M|O}(t_1, t_2) & \text{otherwise}
\end{cases}
\]

- Add step to MCMC whereby “missing” wavelet coefficients \( D_{ijk} \sim N(M_{ijk}^*, V_{ijk}^*) \)

\[
M_i^* = M_i W'
\]

\[
V_i^* = WV_i W'
\]
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


• Computer code/papers on web at http://biostatistics.mdanderson.org/Morris/papers.html
Projecting FMM to Wavelet Space

\[
\begin{align*}
Y_{N \times T} & = \underbrace{X_{N \times p}}_{p \times T} B_{p \times T} + \underbrace{Z_{N \times m}}_{m \times T} U_{m \times T} + \underbrace{E_{N \times T}}_{N \times T}
\end{align*}
\]

\[
U_i \sim MVN(0, Q)
\]

\[
E_i \sim MVN(0, S)
\]
Projecting FMM to Wavelet Space

\[
\begin{align*}
Y_{N \times T} &\mathrel{\leftarrow} W'_{T \times T} = X_{N \times p} B_{p \times T} + Z_{N \times m} U_{m \times T} + E_{N \times T} \\
U_i &\sim MVN(0, Q) \\
E_i &\sim MVN(0, S)
\end{align*}
\]
Projecting FMM to Wavelet Space

\[ Y_{N \times T} = X_{p \times T} B_{T \times T} W'_{N \times T} + Z_{m \times T} U_{T \times T} W'_{N \times T} + E_{T \times T} W'_{N \times T} \]

\[ U_i \sim MVN(0, Q) \]

\[ E_i \sim MVN(0, S) \]
Projecting FMM to Wavelet Space

\[
\begin{align*}
Y &\quad \text{\(N \times T\)} \\
W' &\quad \text{\(T \times T\)} \\
= &\quad \text{\(X \quad N \times p\)} \\
&\quad \text{\(B \quad T \times T\)} \\
&\quad \text{\(W' \quad N \times m\)} \\
&\quad \text{\(Z \quad T \times T\)} \\
&\quad \text{\(U \quad m \times T\)} \\
&\quad \text{\(E \quad N \times T\)} \\
\end{align*}
\]

\[
\begin{align*}
U_i W' &\sim \text{MVN (0, } W Q W' \text{)} \\
E_i W' &\sim \text{MVN (0, } W S W' \text{)}
\end{align*}
\]
Projecting FMM to Wavelet Space

\[
\begin{align*}
\mathbf{D} & = \begin{bmatrix} \mathbf{X} & \mathbf{B}^* & \mathbf{Z} & \mathbf{U}^* & \mathbf{E}^* \end{bmatrix} \\
& = \begin{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{B}^* \end{bmatrix} & \begin{bmatrix} \mathbf{Z} & \mathbf{U}^* \end{bmatrix} & \mathbf{E}^* \end{bmatrix} \\
& = \mathbf{X} \mathbf{B}^* + \mathbf{Z} \mathbf{U}^* + \mathbf{E}^* \\
& = \begin{bmatrix} \mathbf{X} \mathbf{B}^* \mathbf{Z} & \mathbf{E}^* \end{bmatrix} \\
\end{align*}
\]

\[
\begin{align*}
\mathbf{U}_{i}^* & \sim \text{MVN}(0, \mathbf{Q}^*) \\
\mathbf{E}_{i}^* & \sim \text{MVN}(0, \mathbf{S}^*) \\
\end{align*}
\]