Wavelet-Based Functional Mixed Models

Jeffrey S. Morris

UT MD Anderson Cancer Center
Houston, Texas

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: Accelerometer Data

- **Accelerometers**: small motion sensors that digitally record minute-by-minute activity levels
  - Increasingly used in surveillance and intervention studies

- **TriTrac-R3D**: pocket-sized sensor worn on hip
  - Minute-by-minute record of motion in 3 planes
  - Condensed into single activity level measurement/minute, on either acceleration or “METs” scale (multiple of resting metabolic rate)

- **METs for different activities** (Rowlands, et al. 2004)
  - 1 MET = minimum activity level
  - 3-6 MET = moderately intense activity
  - >6 MET = vigorous activity
Accelerometer Data
Accelerometer Data

- **Planet Health Study** (Harvard University):
  - Intervention study investigating activity levels of middle school children in Boston area schools
  - Children’s activity levels objectively monitored using TriTrac-R3D activity monitor for one or two 4-day sessions
  - **Data considered**: 292 daily profiles/103 children/5 schools

- **Goals**:
  1. Assess how activity levels vary throughout day, across schools, across different days of the week, over time from early to late Spring, and across various child-level covariates.
  2. Assess relative variability in activity levels from day-to-day and child-to-child, in order to guide future study design.
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.
Suppose we observe a sample of $N$ curves, $Y_i(t)$, $i=1, \ldots, N$, all defined on $\mathcal{T}$

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

- $B_j(t)$ = fixed effect functions
- $U_k(t)$ = random effect functions
- $E_i(t)$ = residual error processes
- $Q$ and $S$ are covariance surfaces on $\mathcal{T} \times \mathcal{T}$
  - $S(t_1,t_2) = \text{Cov}\{E_i(t_1),E_i(t_2)\}$: describes within-curve covariance structure of residual curve-to-curve deviations

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

$E_i(t) \sim GP(0, S)$
Functional Mixed Models

- **Key feature of FMM**: Does not require specification of parametric form for curves
- Similar model used by Guo (2002 *Biometrics*)
  - Different treatment of random effect functions
    (Not as flexible in modeling curve-to-curve variability)
  - Represented functions using smoothing splines
- Kernels/fixed-knot splines may not work well for spatially heterogeneous 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 \( \mathcal{O}(T) \) for obtaining \( T \) empirical wavelet coefficients for curves sampled on equally-spaced grid of length \( T \).

- **Linear Representation**: \( d = y W' \)
  - \( W' \) = \( T \)-by-\( T \) orthogonal projection matrix

- **Inverse DWT (IDWT)**: \( y = d W \)
Wavelet Regression

- **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 of function
- Morris, et al. (2003) extended to hierarchical functional model (Bayesian)
Adaptive Regularization

Regularization by Local Linear Smoothing

Adaptive Regularization by Wavelet Shrinkage

8/4/2005 ENAR 2005 Austin, TX
Functional Mixed Model
(Discrete version)

\( Y = \text{N-by-T matrix containing the observed spectra on sampling grid of size } T \)

\[
\begin{align*}
Y_{N \times T} &= \begin{bmatrix} X_{p \times T} & B_{N \times p} & Z_{m \times T} & U_{m \times T} & E_{N \times T} \end{bmatrix} \\
U_i &\sim MVN(0, Q) \\
E_i &\sim MVN(0, S)
\end{align*}
\]

- \( B_{ij} \) is the effect of covariate \( i \) at location \( t_j \)
- \( Q \) and \( S \) are covariance matrices (\( T \times T \))
- Note: Some structure must be assumed on form of \( Q \) and \( S \) (discussed later)
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 **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

   $$D = Y W'$$

   - 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.
Projecting FMM to Wavelet Space

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

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

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

\[
Y_{N \times T} = WX^{T}_{T \times T} + B_{N \times p}^{p \times T} + Z_{N \times m}^{m \times T} + E_{N \times T}
\]

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

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

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

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

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

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

\[
U_i \cdot W' \sim MVN(0, WQW')
\]

\[
E_i \cdot W' \sim MVN(0, WSW')
\]
Projecting FMM to Wavelet Space

\[ D_{N \times T} = \underbrace{X_{p \times T} B_{m}^{*} + Z_{m \times T} U_{N \times T}^{*}}_{N \times p} + \underbrace{E_{N \times T}^{*}}_{N \times m} \]

\[ U_{i}^{*} \sim MVN(0, Q^{*}) \]

\[ E_{i}^{*} \sim MVN(0, S^{*}) \]
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.
Model Each Column Separately

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

\[ u_{jk} \sim N(0, q_{jk}^{*}) \]

\[ e_{jk} \sim N(0, s_{jk}^{*}) \]
Prior Assumptions

Mixture prior on $\beta_{ijk}^*$:

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

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

- Nonlinearly shrinks $\beta_{ijk}^*$ towards 0, leading to adaptively regularized estimates of $\beta_i(t)$.
- $\tau_{ij}$ & $\pi_{ij}$ 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 \( \Omega \) 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 $B_j(t)$ 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.
- Can be used for Bayesian inference/prediction
Model

Let $Y$ be $292 \times 660$ matrix containing 292 accelerometer profiles for each minute from 9am-8pm.

$$Y = XB + ZU + E$$

- **X** = $292 \times 14$ matrix of covariates
  - School effects (5), gender, % body fat, BMI, day-of-week (4), daylight savings time, avg tv hrs/wk
- **B** = $14 \times 660$ matrix of fixed effects functions
  - $B_{ij}$ is effect of covariate $i$ at time $t_j$
- **Z** = $292 \times 106$ matrix indicating child for each profile
- **U** = $106 \times 660$ matrix of random effect functions (1/child)
- **E** = $292 \times 660$ matrix of residual errors
Results

Overall Mean Function

METS

11am 1pm 3pm 5pm 7pm
Selected Results: School Effects

- School schedules evident in effects
  - Spikes every 48 min (changing classes)
  - 3 lunch periods
  - School out at 2:15 pm
- Not so evident in individual curves
Selected Results: BMI Effect

- BMI Coded as continuous factor (mean-centered)

- BMI effect positive ($p<0.0005$)
  - Higher BMI, more active
  - Preprocessing artifact?

- Should raw activity levels be monitored instead of METs?
Selected Results: **DST Effect**

- DST – April 6th
- More active after DST (overall 8%, $p=0.062$)
- Especially strong:
  - As school is letting out (2:15-3:00, 25%, $p=0.03$)
  - In early evening (5:30-7:00, 30%, $p=0.01$)
- Note: Sunset was
  - 5:10-6:15 before DST
  - 7:15-8:10 after DST
Selected Results: TV hours/wk

- TVhrs coded as continuous factor (standardized)
- TVhrs effect negative (-1.3% per sd, $p=0.03$)
  - More TV, less active
  - 3:00-5:30, -2.6%, $p=0.02$
  - 7:00-8:00, -3.6%, $p=0.008$
- Positive effect over lunch
  - +2.7%, $p=0.03$
  - More TV, on average more active over lunch
Results: Covariance Analysis

- Variability: 91% day-to-day, 9% child-to-child
  - Important to have many days per child
- Study variability as function of $t$
  - Child-to-child variability: school day > after school
  - Day-to-day variability: after school > school day
  - Relative day-to-day variability after school: 95%-99%
- Equivalent designs:
  - 108 children, 4 days/child
  - 72 children, 8 days/child
  - 54 children, 16 days/child
- Less children, more days, save $$?
Discussion

• Introduced unified modeling approach for FDA
  – Can accommodate very irregular functions
• 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 (not discussed here)

• Method **generalizes to higher dimensional functions**, e.g. image data, space/time (fixed domain) data.
Acknowledgements

• Work presented here is from 2 papers (both currently under review)
