Dealing with Incomplete Profiles in Wavelet-Based Functional Mixed Models

Jeffrey S. Morris

UT MD Anderson Cancer Center

Houston, Texas

Joint work with Louise Ryan, Steve Gortmaker, Brent Coull, Cassandra Arroyo, and Dick Herrick
Outline

• Introduction:
  – Functional Data
  – Example: Accelerometers
• Functional Mixed Models
• Wavelet-based Functional Mixed Models
• Posterior Predictive Distribution-Based Multiple Imputation Scheme
• Application/Results
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
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

Chili 1, Day 1

Chili 2, Day 1

Chili 1, Day 2

Chili 2, Day 2
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)

- **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):

\[
Y_{N \times 1} = X_{N \times p} \beta_{p \times 1} + Z_{N \times m} u_{m \times 1} + e_{N \times 1}
\]

- **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.

\[
u \sim N(0, D_{m \times m})
\]

\[
e \sim N(0, R_{N \times N})
\]
Functional Mixed Model

Suppose we observe a sample of $N$ curves, $Y_i(t)$, $i=1, \ldots, N$, all defined on $T$.

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

- $B_j(t)$ = fixed effect functions
- $U_k(t)$ = random effect functions
- $E_i(t)$ = residual error processes
- $Q$ and $S$ are covariance surfaces on $T \times T$
  - $S(t_1,t_2) = \text{Cov}\{E_i(t_1), E_i(t_2)\}$: 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 Model
(Discrete version)

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

\[
\begin{align*}
Y & = X \begin{bmatrix} B_1 \\ \vdots \\ B_p \end{bmatrix} + Z \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix} + E \\
& = N \times T \times p + N \times m + N \times T
\end{align*}
\]

\[
U_i \sim MVN(0, Q) \quad E_i \sim MVN(0, S)
\]

• \( B_{ij} \) 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)
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, triceps calipers, 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**
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**: \( d = y W' \)
  - \( W' = T \times T \) orthogonal projection matrix

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

\[
D = Y W' \\
N \times T \quad N \times T \quad T \times T
\]

- Projects the observed curves into the space spanned by the wavelet bases.
Wavelet-Based FMM: General Approach

1. Project observed functions $Y$ into wavelet space.

2. Fit FMM in wavelet space (Use MCMC to get posterior samples)

3. Project wavelet-space estimates (posterior samples) back to data space.
Wavelet Space FMM

**D** : empirical wavelet coefficients for observed curves
Row *i* contains wavelet coefficients for observed curve *i*
Each column **double-indexed** by wavelet scale *j* and location *k*

\[
\begin{align*}
D \left( N \times T \right) &= \left( X \left( N \times p \right) B^* \left( p \times T \right) \right) + \left( Z \left( N \times m \right) U^* \left( m \times T \right) \right) + \left( E^* \left( N \times T \right) \right) \\
U^* &\sim MVN(0, Q^*) \\
E^* &\sim MVN(0, S^*)
\end{align*}
\]

- **B** = **BW’** & **U** = **UW’** : Rows contain wavelet coefficients for the fixed and random effect functions,
- **E** = **EW’** is the matrix of wavelet-space residuals
- **Q** = **WQW’** and **S** = **WSW’** model the covariance structure between wavelet coefficients for a given function.
- **Q** and **S** are typically too large to estimate in an unstructured fashion: special structure assumed.
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. $Q^*$ and $S^*$).

### MCMC Steps

1. **Sample from $f(B^*|D,\Omega)$:**
   Mixture of normals and point masses at 0 for each $i,j,k$.

2. **Sample from $f(\Omega|D,B^*)$:**
   Metropolis-Hastings steps for each $j,k$

3. If desired, **sample from $f(U^*|D,B^*,\Omega)$:**
   Multivariate normals
Wavelet-Based FMM: General Approach

1. **Project** observed functions $Y$ **into wavelet space.**
2. **Fit FMM** in wavelet space (Use MCMC to get posterior samples)
3. **Project** wavelet-space estimates (posterior samples) **back to data space.**
Wavelet-Based FMM

3. **Project** wavelet-space estimates (posterior samples) **back to data space.**

- Apply IDWT to posterior samples of $B^*$ to get posterior samples of fixed effect functions $B_j(t)$ for $i=1,\ldots, p$, on grid $t$.

  - $B=B^*W$

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

- Can be used for Bayesian inference/prediction
Heatmap of Missingness
(Black=missing)
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.
   • 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.
   • Borrow information from nearby times in incomplete profiles.
   • Makes predictions for missing regions “connected” with observed.

4. Supplement WFMM with step to stochastically impute values for missing data.
   • Inference appropriately accounts for uncertainty in imputation
Incomplete Profile
Predictive Distribution

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

\[ \Sigma_i(t_1, t_2) = COV \{ Y_i(t_1), Y_i(t_2) \mid Y^C \} \]
Imputation distribution

\[ \mu_{i}^{M,O} = \mu_{i}^{M} + \sum_{i}^{M,O} \left( \sum_{i}^{O,O} \right)^{-1} \left( Y_{i}^{O} - \mu_{i}^{O} \right) \]

\[ \sum_{i}^{M,O} = \sum_{i}^{M, M} - \sum_{i}^{M, O} \left( \sum_{i}^{O,O} \right)^{-1} \sum_{i}^{O, M} \]
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

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

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

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

\[ M_i^* = M_i W' \]

\[ V_i^* = WV_i W' \]
Selected Results: School Effects

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

- **WFMM 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

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

Let $Y_{ij}(t)$ be accelerometer profile on day $j$ from child $i$

$$Y_{ij}(t) = B_0(t) + \sum_{k=1}^{p_1} X_{ik} B_{k}^{\text{child}}(t) + \sum_{k=1}^{p_2} X_{ijk} B_{k}^{\text{day}}(t) + U_i(t) + E_{ij}(t)$$

- $X_{ik} =$ child-level covariates (school, race, gender, BMI, % body fat, avg hrs of TV/week)
- $X_{ijk} =$ day-level covariates (day-of-week, DST)
- $B_0(t) =$ overall mean profile
- $B_k^{\text{child}}(t) =$ effect functions for child-level covariates
- $B_k^{\text{day}}(t) =$ effect functions for day-level covariates
- $U_i(t) =$ Random effect function for child $i$
Results: Bouts

- Can compute posterior predictive probabilities of bouts for children
Results: Bouts

- Can compute posterior predictive probabilities of bouts for children
- Model-based predictive probabilities not far from empirically-estimated probabilities
- May want heavier tails
Some Results

- **School #5:**
  - Spikes every 48min (changing classes)
  - 3 lunch periods
  - School out at 2:15pm

- **BMI effect positive**
  - Artifact of preprocessing?

- **Daylight Savings Time**
  - More active after DST
  - Especially 2-3pm, 5-7pm

- **Avg hrs TV/week**
  - More TV=less active
  - Especially 3-5pm, 7-8pm
  - More active at lunch
Results
Adaptive Regularization

Regularization by Local Linear Smoothing

Adaptive Regularization by Wavelet Shrinkage
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
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

\[ d_{jk} = X \beta_{jk}^* + Z u_{jk}^* + e_{jk}^* \]

\[ u_{jk}^* \sim N(0, q_{jk}^*) \]
\[ e_{jk}^* \sim N(0, s_{jk}^*) \]
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?
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 data – inherent smoothness assumptions attenuate local features
- **Wavelet Regression**: nonparametric regression technique that better preserves local features present in the curves.
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)