Modeling Children's Activity Levels using Wavelet-Based Functional Mixed Models

Jeffrey S. Morris

University of Texas M.D. Anderson Cancer Center, Houston, TX

Joint work with: **Louise Ryan, Cassandra Arroyo, Brent Coull**, and **Steve Gortmaker**

Overview

- Introduction
  - Planet Health Study/Accelerometer Data
  - Mixed Models/Wavelets
- Wavelet-based Functional Mixed Models
- Missing Data Methods for WFMM
- Analysis of Accelerometer Data from Planet Health Study
- Discussion/Conclusions

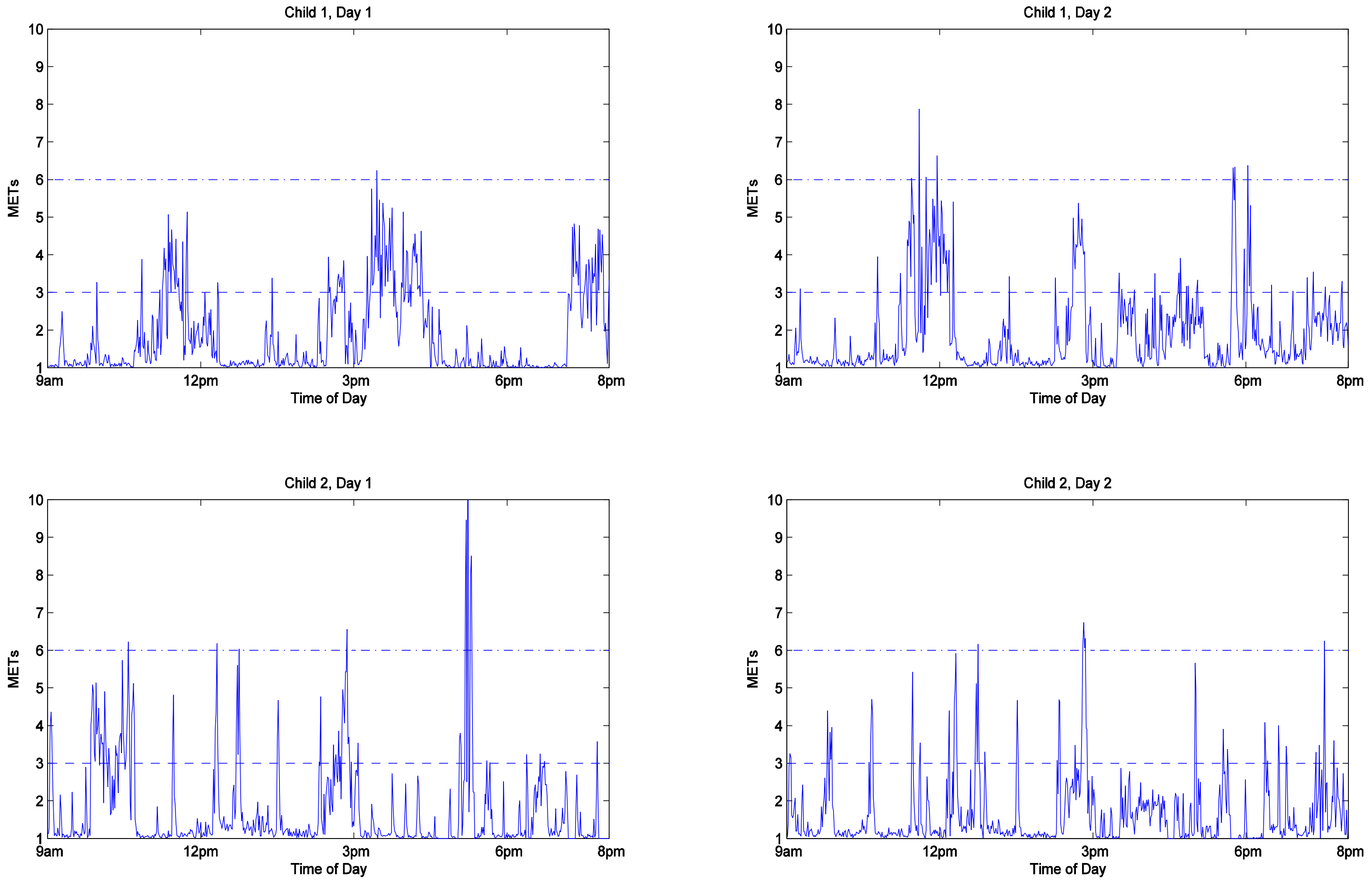
Planet Health Study

- **Planet Health:** school-based intervention study to reduce obesity in middle school youth by changing key physical activity and dietary risk factors.
- **Design:**
  - 10 schools paired up, randomized to intervention/control
  - 1295 children: nutritional, behavioral, and health-related outcomes measured at baseline (F95) and follow-up (S97)
- **Challenge:** How to measure physical activity levels?
  - Questionnaires
  - Accelerometers: objectively quantify activity levels
- **256 children:** monitored using TriTrac-R3D activity monitor for one or two 4-day sessions in S97

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



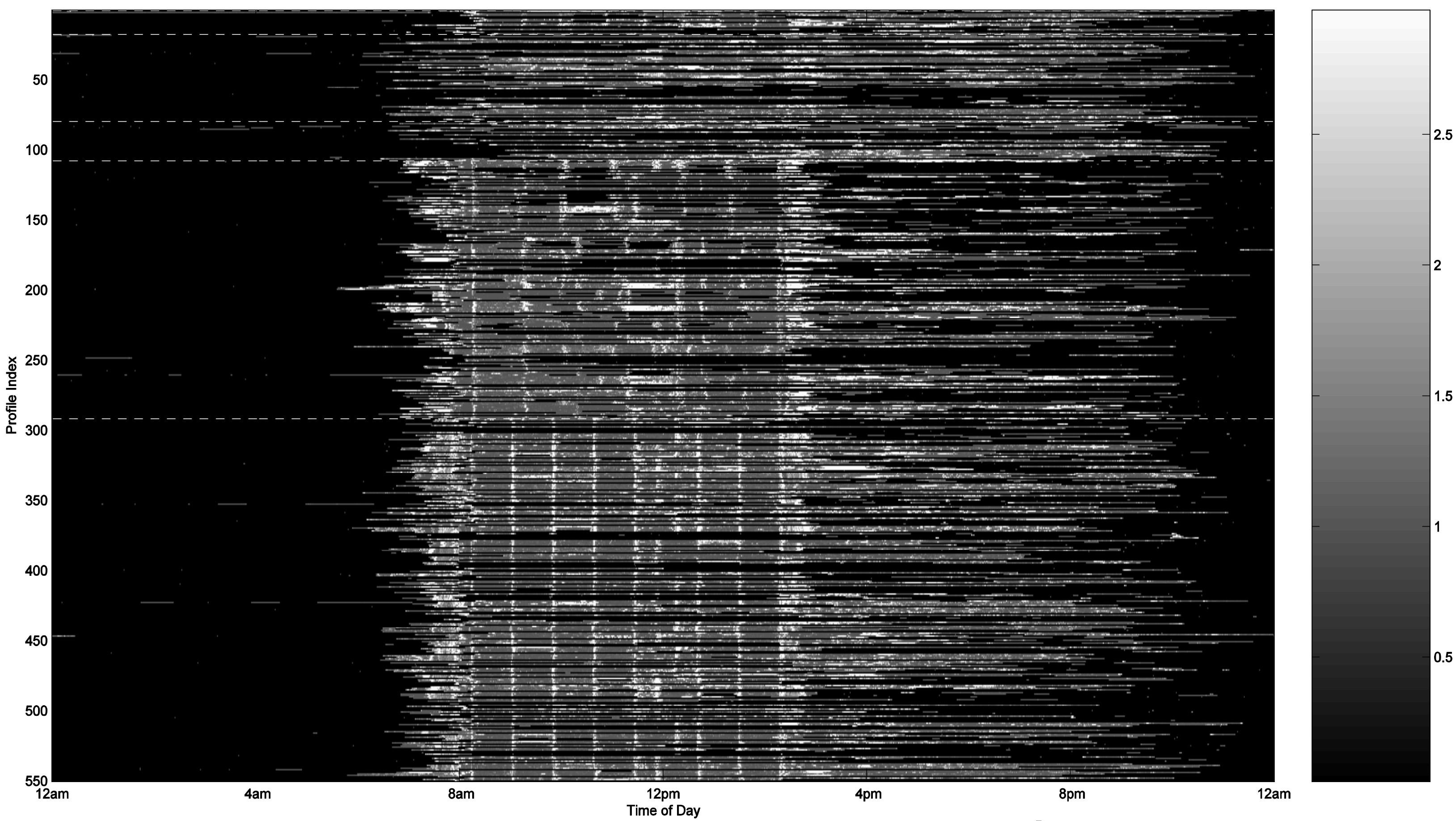
Analyzing Accelerometer Data

- Major challenge: How to analyze these data?
- Standard analysis strategies involve simple, arbitrary summary measures:
  - Average daily activity levels
  - Proportion of time with activity levels corresponding to sedentary, moderately vigorous, and vigorous activities
  - 30-minute averages
- Limitations of these approaches:
  - Do not make full use of information in functional data
  - Cannot effectively deal with incomplete profiles
- Methods that model the entire functions may be more effective in extracting information from these data

Accelerometer Data

- Here, we focus only on weekday profiles from children at control schools
  - 550 profiles from 112 children, 1440 meas./profile
  - Enormous amount of data; can be displayed using Heatmap
- Lots of *missing data*
  - Use profiles >50% complete from 9am-8pm
    - 292 profiles from 106 children

Heatmap



Accelerometer Data

Accelerometer Data

- **Other measurements:**
  - **Child-level covariates**: school, race, gender, age, weight, height, BMI, % body fat, avg hrs of TV/week
  - **Day-level covariates:** day of week, calendar date
- **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.

Statistical Challenges

- In order to model these data as functions, we need a method that can:

1. Model functions of **arbitrary form**
2. Jointly model ***functional effects*** of multiple covariates
3. Account for **correlation between profiles** from same child
4. Incorporate information from **incomplete profiles**

Linear Mixed Models

Linear Mixed Model (Laird and Ware, 1982):

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

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

- **Fixed effects part, $X\beta$, accommodate a broad class of mean structures, including main effects, interactions, and linear coefficients.**
- **Random effects part, Zu, provide a convenient mechanism for modeling correlation among the N observations.**

Functional Mixed Model

Morris and Carroll (2004) (Discrete version)

Y= *N*-by-*T* matrix containing the observed curves all sampled on a common equally-spaced grid of length *T*, *t*.

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

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

- **B_{ij} is the effect of covariate i at location t_j**
- **U and E follow the Matrix Normal distribution**
  - **U~MN(P,Q) implies Cov{U_{ij}, $U_{i'j'}$}=$P_{ii'}$*$Q_{jj'}$**
- **P and R are between-curve covariance matrices**
- **Q and S are within-curve covariance matrices ($T \times T$)**

Example: Model

Let Y be 292×660 matrix containing *292* accelerometer profiles for each minute from 9am-8pm.

$$\boxed{Y = XB + ZU + E}$$

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

Functional Mixed Models

- **Key feature of FMM**: Does not require specification of parametric form for curves
- Most existing literature for nonparametrically modeling functional data is based on kernels or splines (Guo 2002).
- Kernels/fixed-knot splines may not work well for spatially heterogeneous data
- **Wavelet Regression**: nonparametric regression technique that better preserves local features present in the curves.

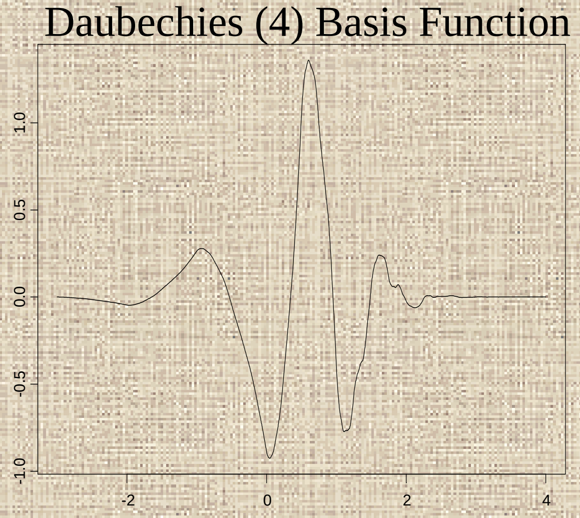
Introduction to Wavelets

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

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

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

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



- **Discrete Wavelet Transform (DWT): fast algorithm {O(*T*)} for obtaining *T* empirical wavelet coefficients for curves sampled on equally-spaced grid of length *T*.**
- **Linear Representation:** $\mathbf{d} = \mathbf{y}\, \boldsymbol{W}'$
  - ***W'* =*T*-by-*T* orthogonal projection matrix**
- **Inverse DWT (IDWT):** $\boldsymbol{y} = \boldsymbol{d}\, \boldsymbol{W}$

Wavelet Regression

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

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

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

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

Wavelet Regression

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

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

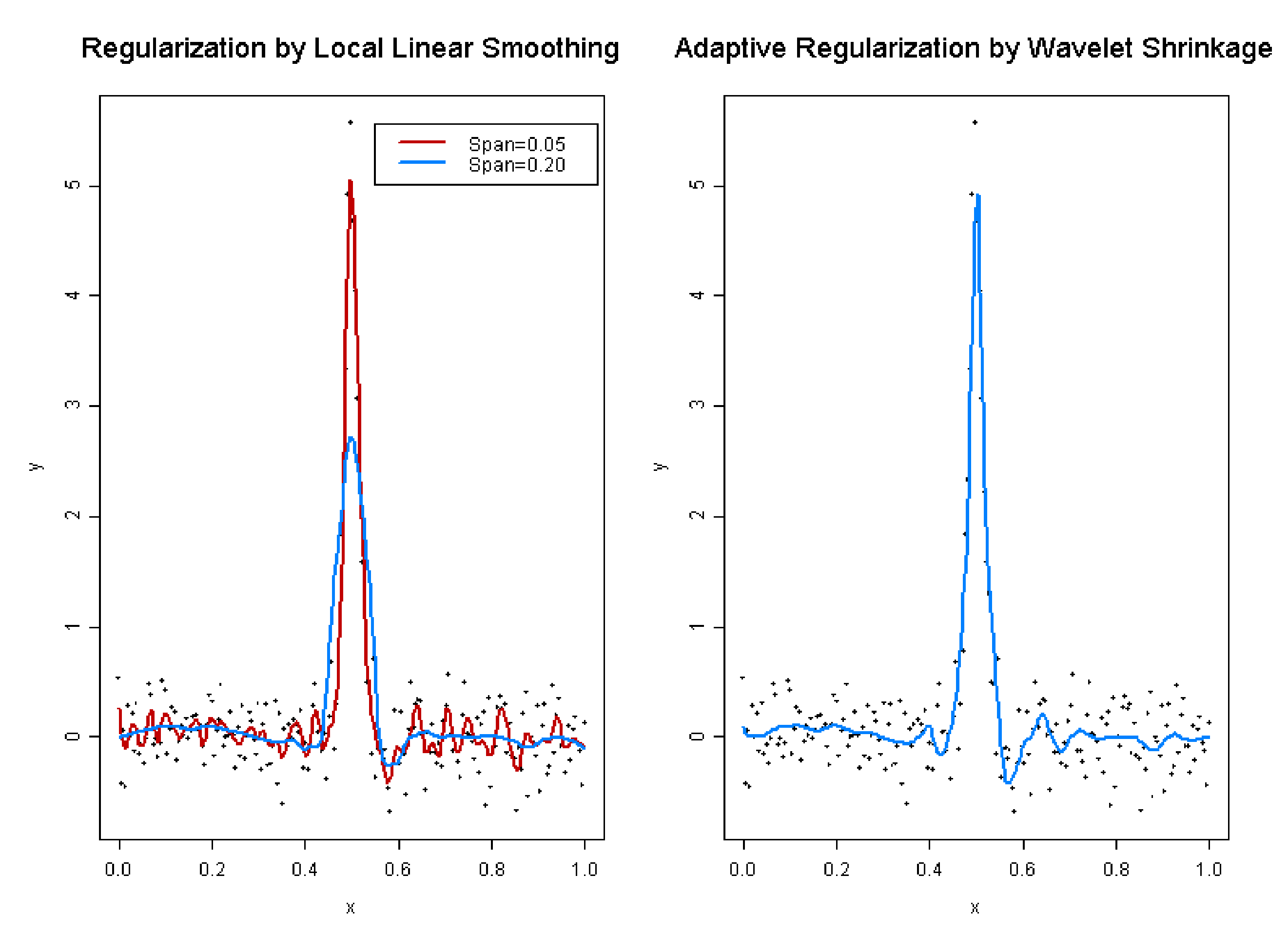
1. **Project data into wavelet space using DWT.**
   d=y ***W'*** **where *W'* is the orthogonal DWT matrix**

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

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

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

Adaptive Regularization



Wavelet-based FMM:

- Morris, et al. (2003) introduced new Bayesian method generalizing wavelet regression to hierarchical functional framework
  - Bayesian inference on mean curve and random effect functions for nested functional data
- Morris and Carroll (2004) extended this work to functional mixed model
  - Arbitrary functional fixed and random effects
  - Wavelet-based functional mixed models

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**

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

- **Projects the observed curves into the space spanned by the wavelet bases.**

Wavelet-based FMM:

General Approach

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

Projecting FMM to Wavelet Space

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

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

Projecting FMM to Wavelet Space

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

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

Projecting FMM to Wavelet Space

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

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

Projecting FMM to Wavelet Space

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

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

Projecting FMM to Wavelet Space

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

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

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

Prior Assumptions

Mixture prior on B^*_{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 $B_i(t)$.**
- **τ_{ij} & π_{ij} are regularization parameters**
  - **Can be estimated from the data using empirical Bayes**
  - **Extend Clyde&George (1999) to functional mixed model**

Model Fitting

- Let Ω be a vector containing ALL covariance parameters (i.e. for *P*, *Q**, *R*, and *S**).

MCMC Steps

1. **Sample from $f(B^*|D,\Omega)$:**

   Mixture of normals and point masses at 0 for each *i,j,k*.

2. **Sample from $f(\Omega|D,B^*)$:**

   Metropolis-Hastings steps for each *j,k*

3. **If desired, sample from $f(U^*|D,B^*,\Omega)$:**

   Multivariate normals

Wavelet-based FMM:

General Approach

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

Wavelet-based FMM

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

- **Apply IDWT to posterior samples of *B** to get posterior samples of *B***
  - **B=B*W**
- **Also for *U*, *P*, *Q*, *R*, and *S***
- **Can be used for Bayesian inference/prediction**
  - **Posterior probabilities**
  - **Bayes Factors**
  - **Posterior predictive probabilities**

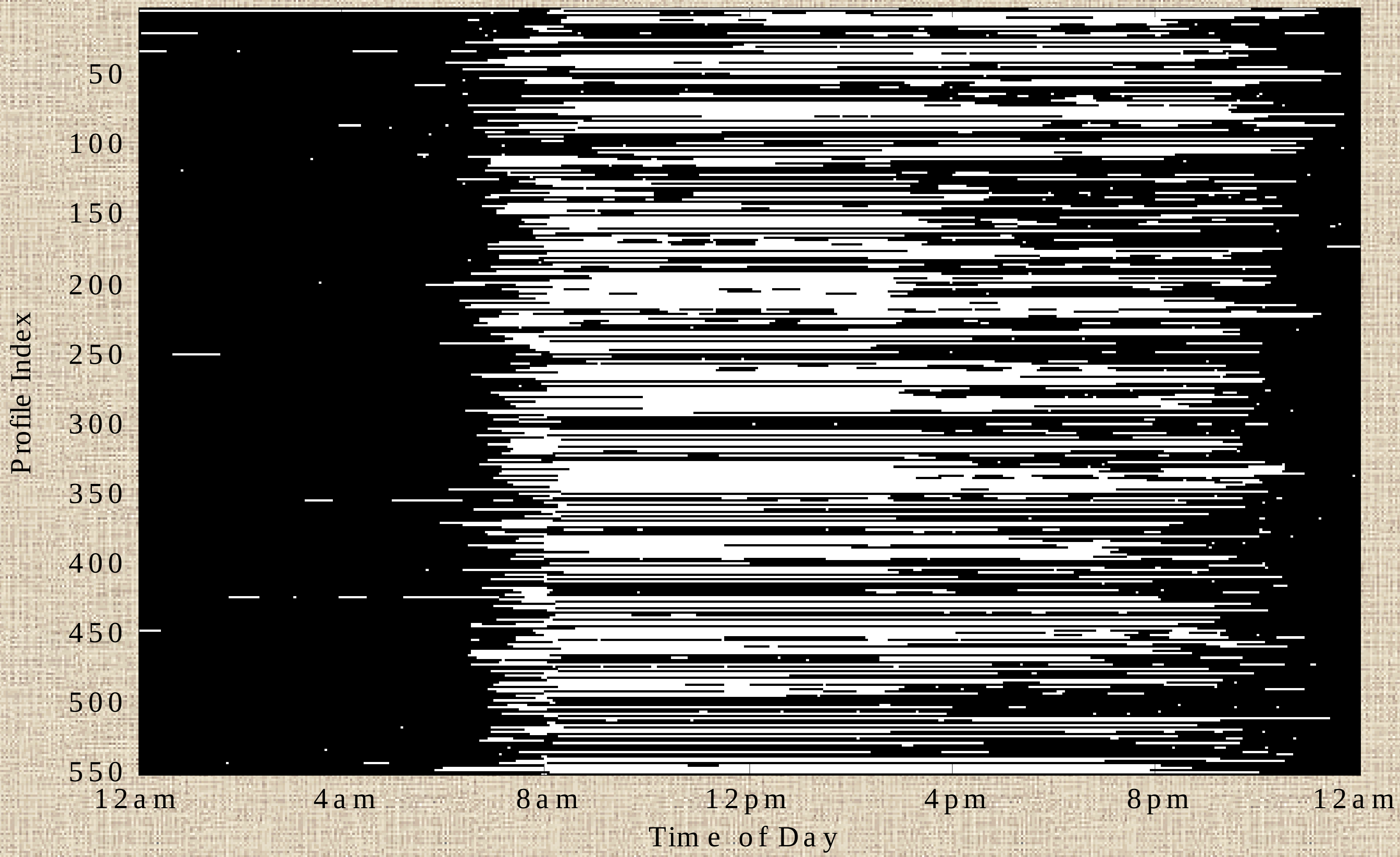
Incomplete Profiles

- **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* ?**

Missing Data

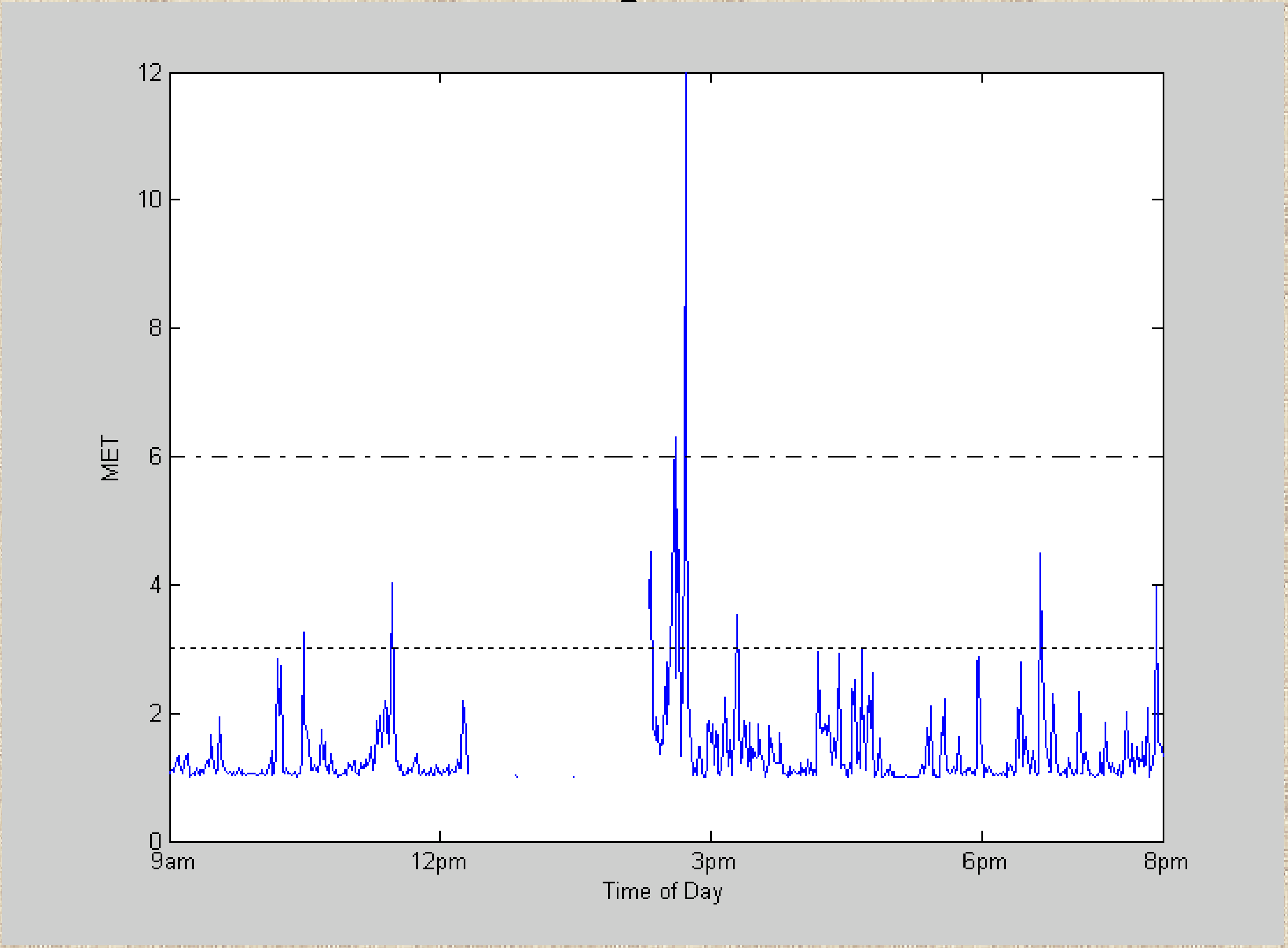


Accelerometer Data

Approach: Incomplete Profiles

1. First fit model to *complete profiles*, get estimates and standard errors for model parameters.
2. Use these to estimate *predictive distributions* for the 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

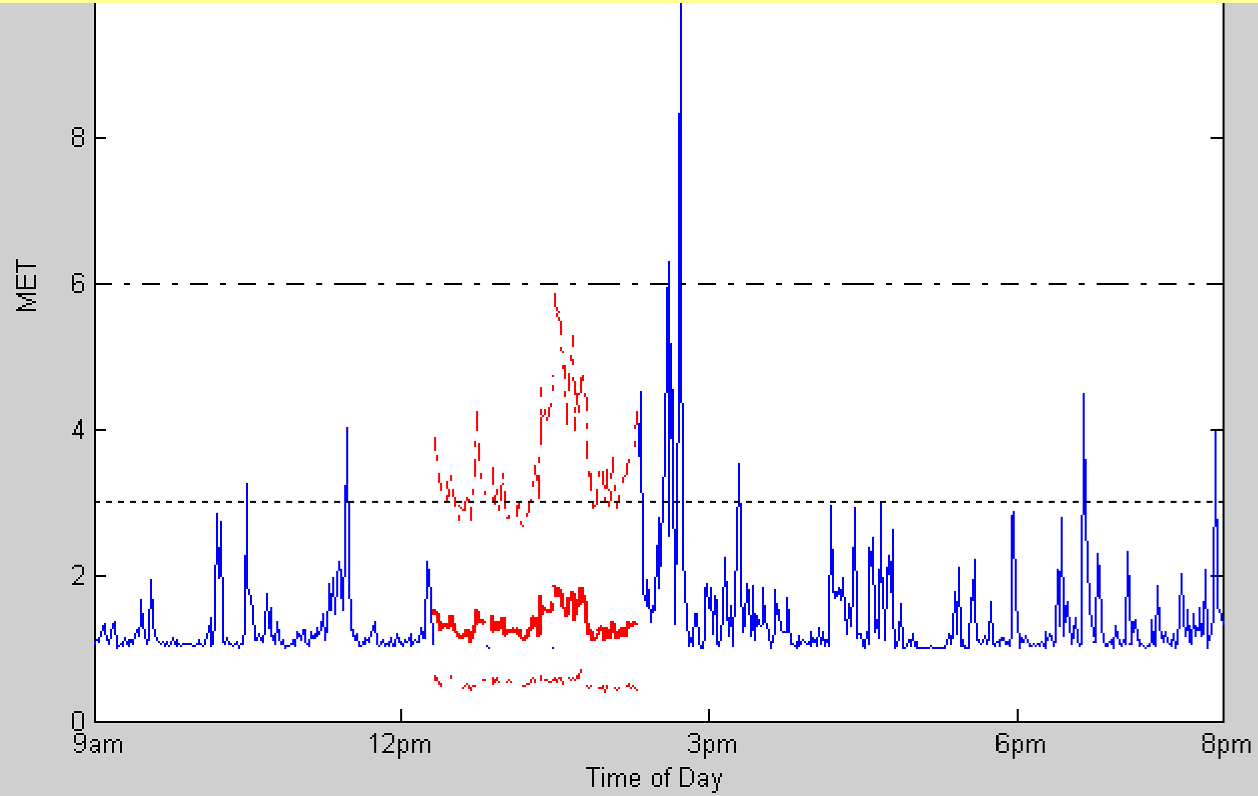


Return

Predictive Distribution

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

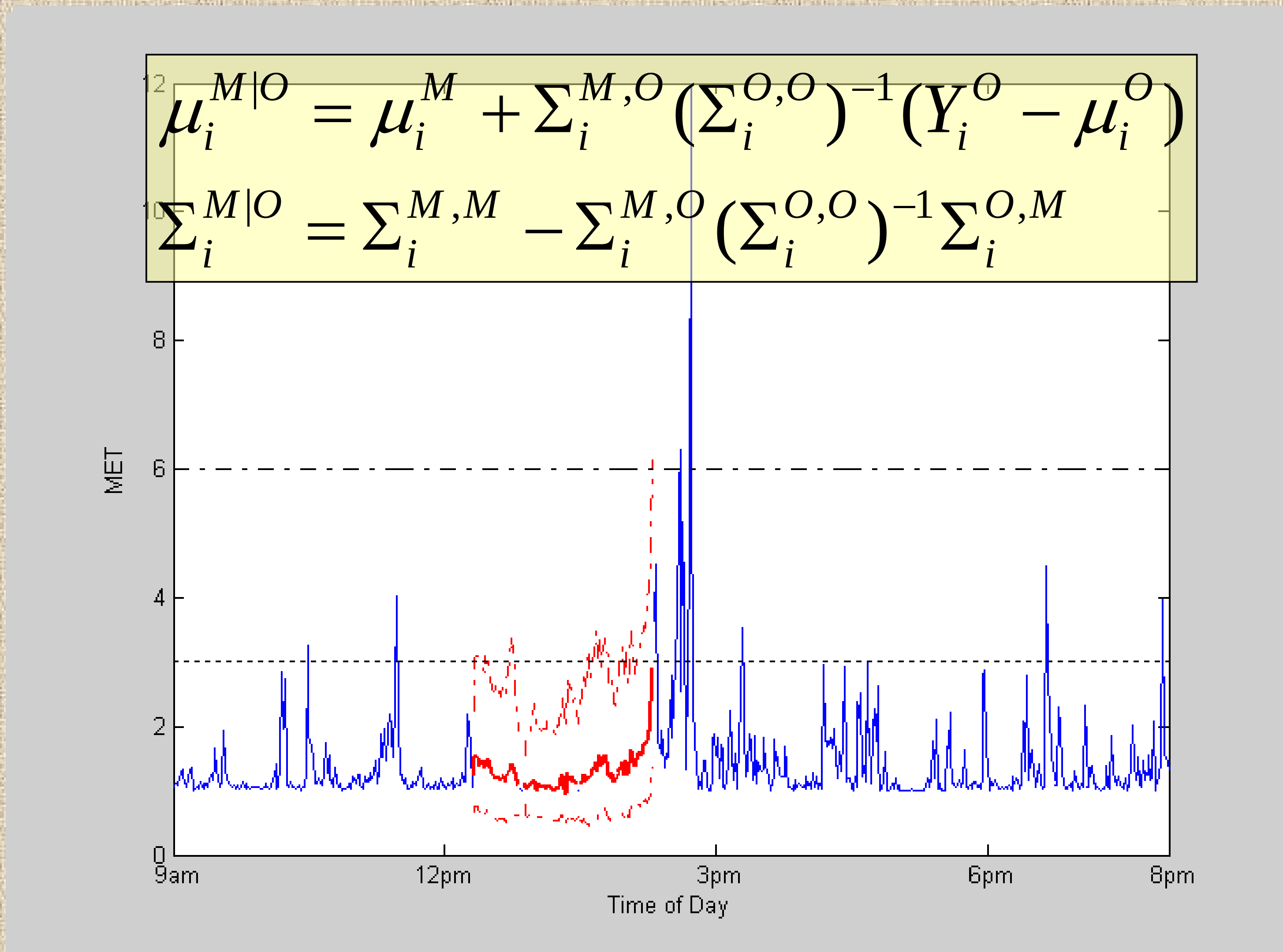
$$\Sigma_i(t_1, t_2) = COV\ \{Y_i(t_1), Y_i(t_2) \mid Y^C\}$$



Imputation distribution

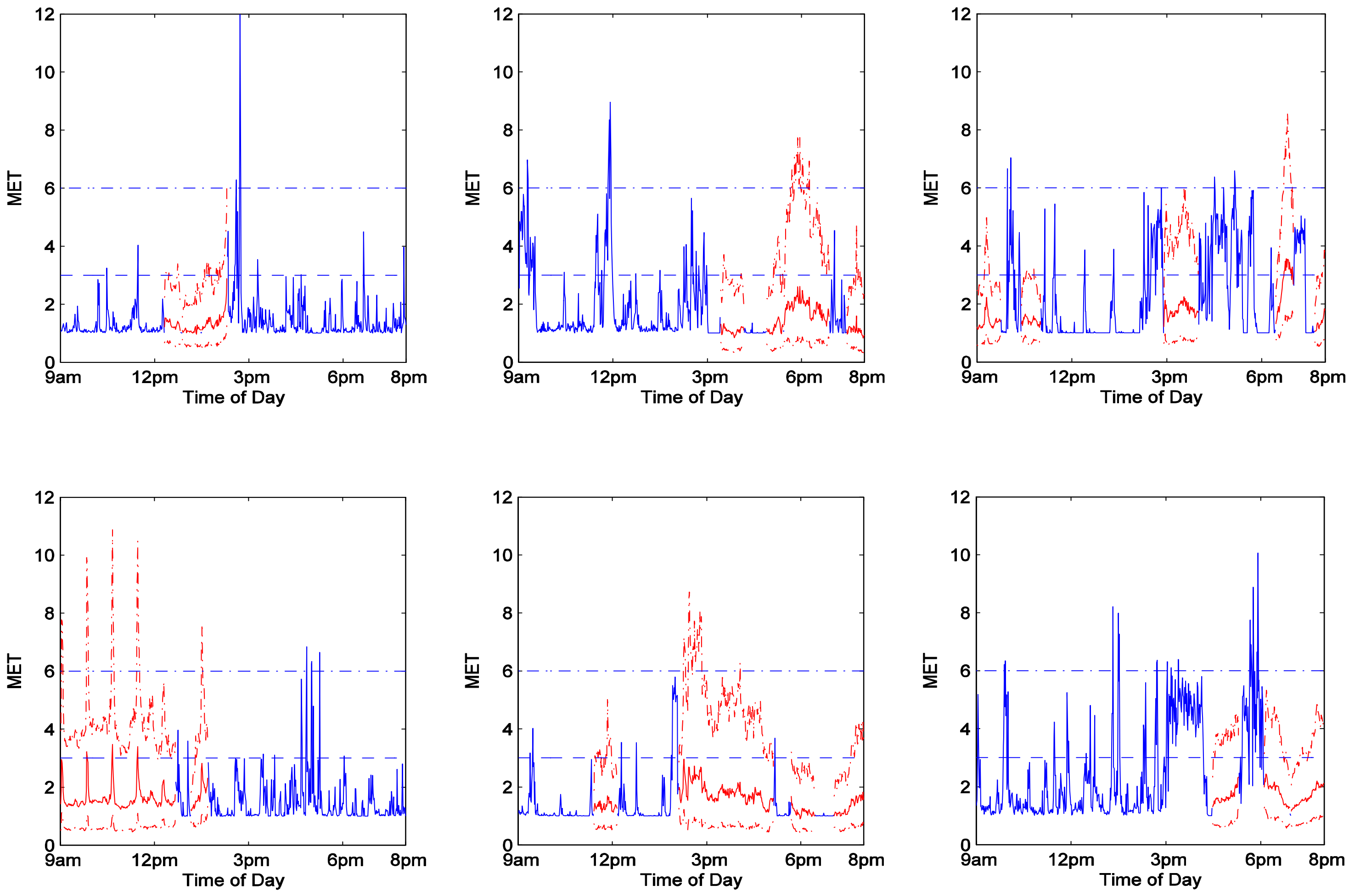
$$\mu_i^{M|O} = \mu_i^{M} + \Sigma_i^{M,O}(\Sigma_i^{O,O})^{-1}(Y_i^{O} - \mu_i^{O})$$

$$\Sigma_i^{M|O} = \Sigma_i^{M,M} - \Sigma_i^{M,O}(\Sigma_i^{O,O})^{-1}\Sigma_i^{O,M}$$



Return

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

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

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

$$M_i^* = M_i W'$$

$$V_i^* = W V_i W'$$

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

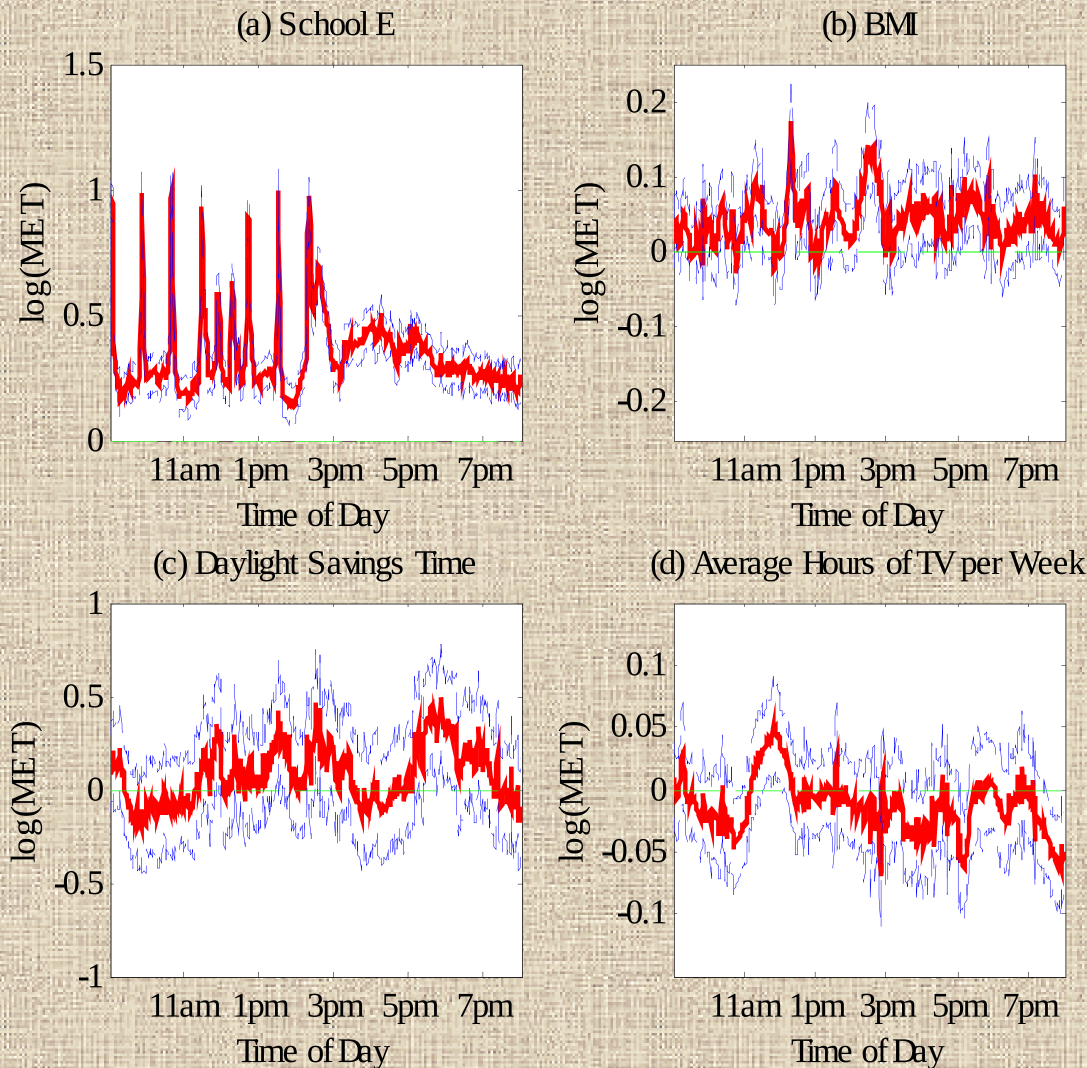
Model

Let Y be 292×660 matrix containing *292* accelerometer profiles for each minute from 9am-8pm.

$$\boxed{Y = XB + ZU + E}$$

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

Some Results

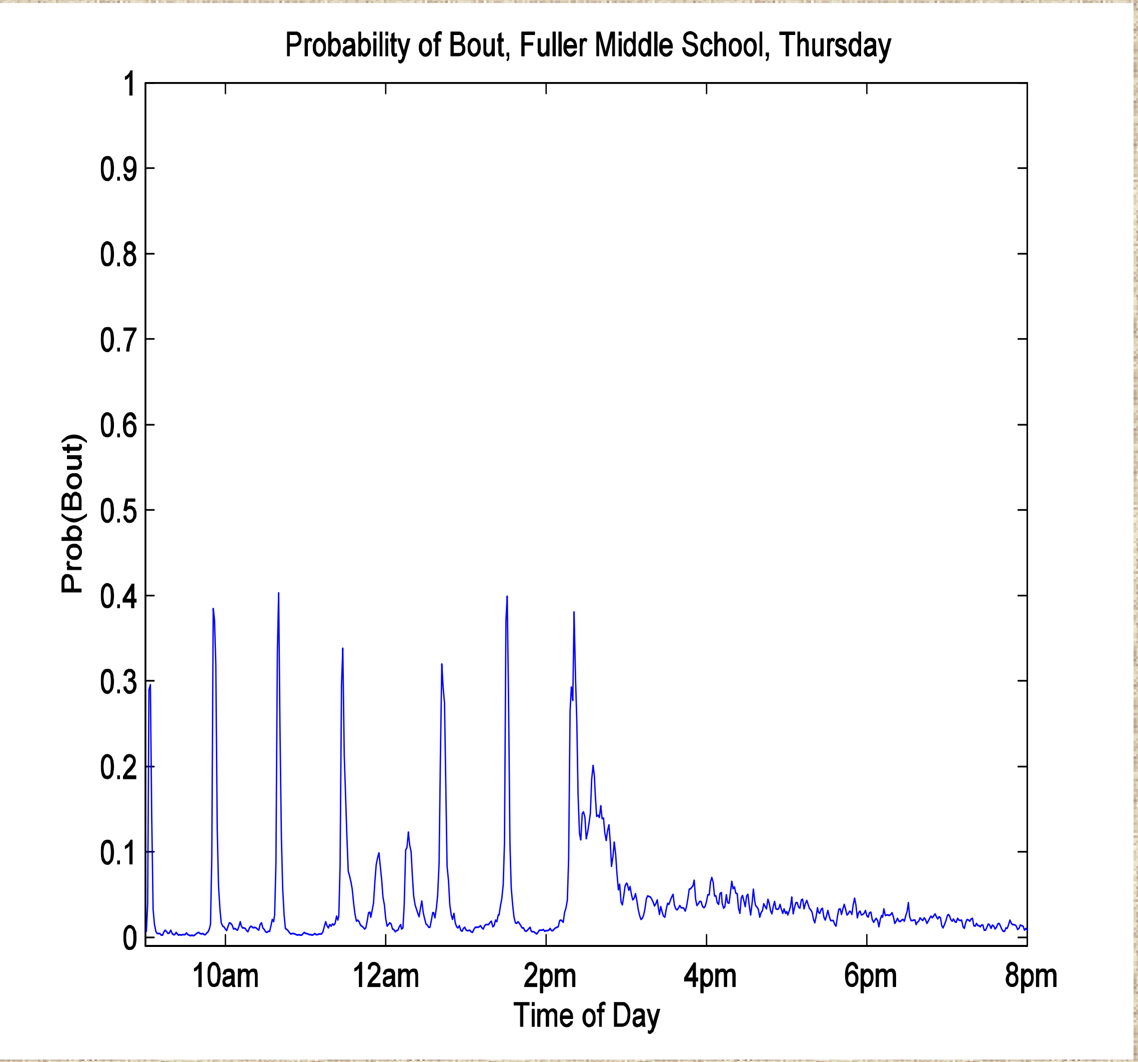


- **School #5:**
  - Spikes every 48min (changing classes)
  - 3 lunch periods
  - School out at 2:15pm
- **BMI effect positive**
  - Artifact of preprocessing?
- **Daylight Savings Time**
  - More active after DST
  - Especially 2-3pm, 5-7pm
- **Avg hrs TV/week**
  - More TV=less active
  - Especially 3-5pm, 7-8pm
  - More active at lunch

Results: Covariance Analysis

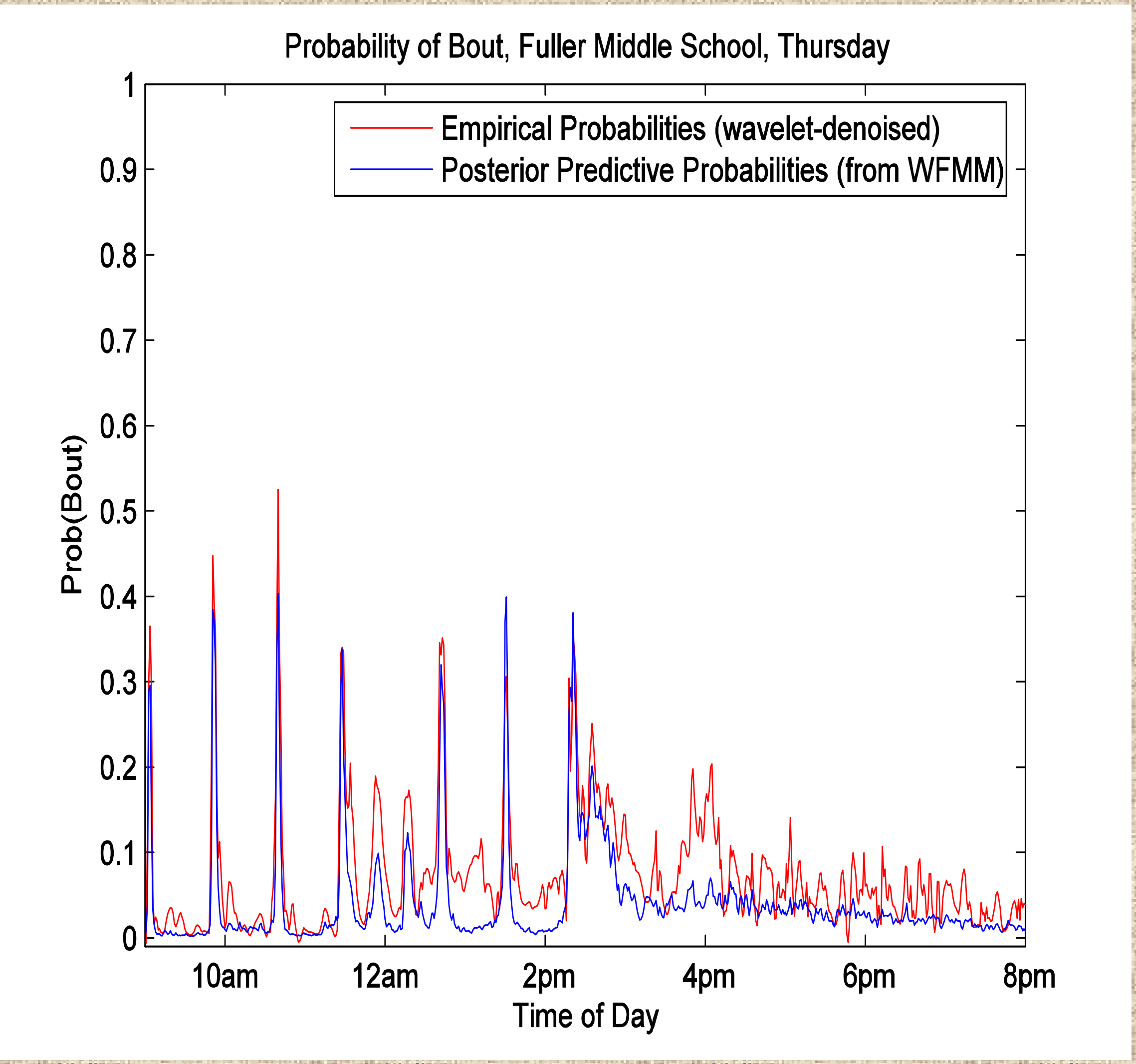
- Variability: 91% day-to-day, 9% child-to-child
  - Important to have many days per child
- Study variability as function of *t*
  - Child-to-child variability: school day > after school
  - Day-to-day variability: after school > school day
  - Relative day-to-day variability after school: 95%-99%
-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

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 probabilites not far from empirically-estimated probabilities
- May want heavier tails

Discussion

- **WFMM: unified modeling approach for FDA**
  - **Extract information from samples of accelerometer profiles**
- **Method based on mixed models; is FLEXIBLE**
  - **Accommodates a wide range of experimental designs**
  - **Addresses large number of research questions**
- **Posterior samples allow Bayesian inference and prediction**
  - **Posterior credible intervals; pointwise or joint**
  - **Predictive distributions for future sampled curves**
  - **Predictive probabilities for group membership of new curves**
  - **Bayesian functional inference can be done via Bayes Factors**
- **Since a unified modeling approach is used, all sources of variability in the model propagated throughout inference.**

Discussion

- **Since functions adaptively regularized using wavelet shrinkage, the method is appropriate for spatially heterogeneous functional data, like accelerometer data.**
- **Approach is Bayesian. The only informative priors to elicit are regularization parameters, which can be estimated from data using empirical Bayes.**
- **Method complex, but code straightforward to implement.**
- **Method generalizes to higher dimensional functions, e.g. image data, space/time (fixed domain) data.**
- **We used wavelet bases, but approach can be generalized to other orthogonal basis functions.**
- **Unified modeling approach makes it possible to develop rigorous methods for incorporating of observed functions with missing data**
  - **Makes it possible to relax equally-spaced grid assumption**

Acknowledgements

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