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

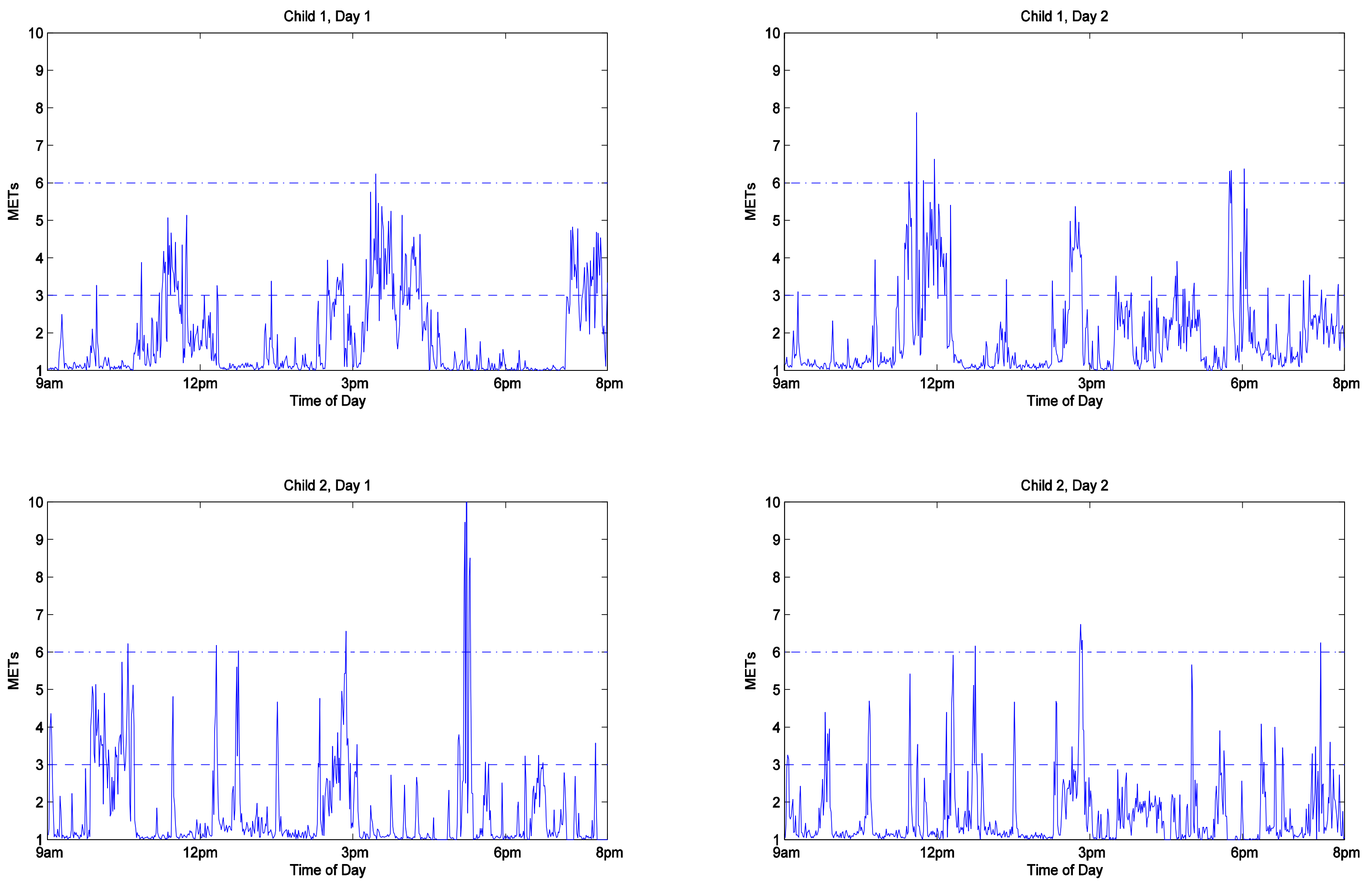
Functional Data

- **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



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):

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

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

- **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

Suppose we observe a sample of *N* curves, $Y_i(t)$, i=1, …, N, all defined on $\mathcal{T}$

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

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

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

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

Functional Mixed Model
(Discrete version)

Y= *N*-by-*T* matrix containing the observed spectra on sampling grid of size *T*

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

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

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

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, triceps calipers, 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

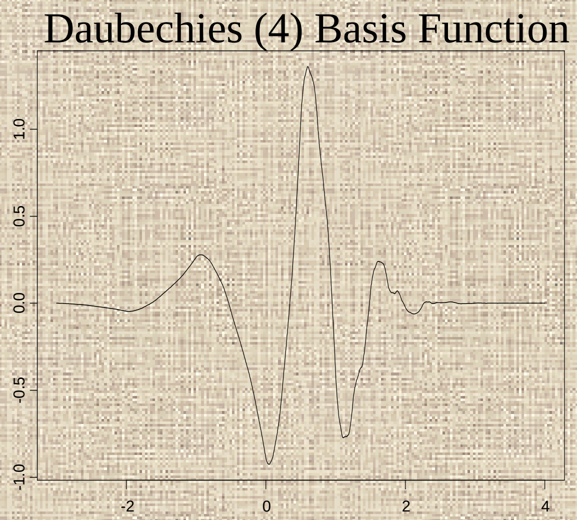
Introduction to Wavelets

- **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:** ***d* = y *W'***
  - ***W'* =*T*-by-*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**

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

- **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.**

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*

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

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

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

- ***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 β_{ijk}^*:

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

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

- **Nonlinearly shrinks β_{ijk}^* towards 0, leading to adaptively regularized estimates of $\beta_i(t)$.**
- **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

- **MCMC to obtain posterior samples of model quantities**
  - **Work with marginal likelihood; U* integrated out;**
- **Let Ω 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

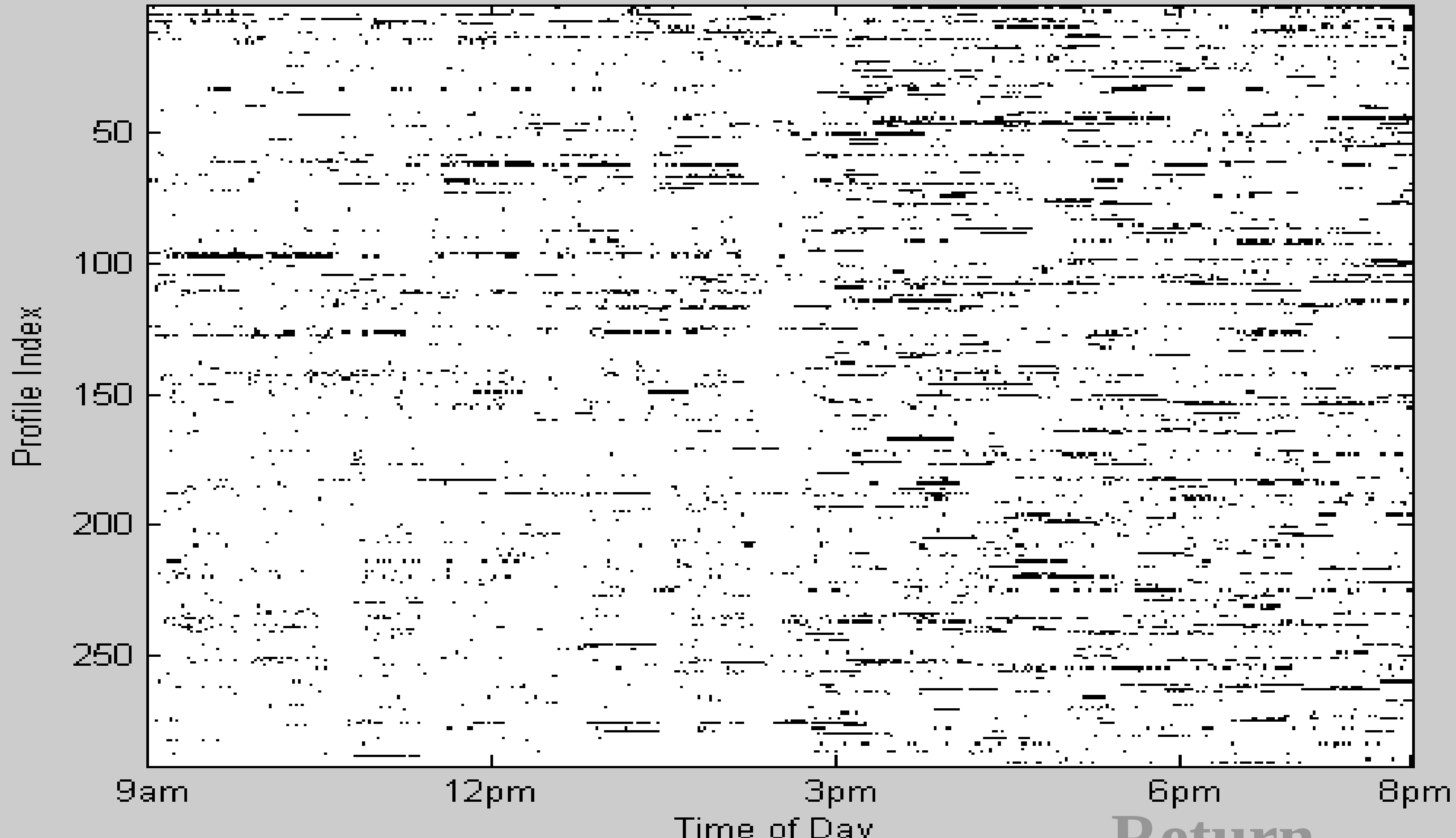
1. **Project observed functions Y into wavelet space.**
2. **Fit FMM in wavelet space**
   (Use MCMC to get posterior samples)
3. **Project wavelet-space estimates (posterior samples) back to data space.**

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)



Return

Incomplete Profiles

- **Lots of missing data** **(Missing Data)**

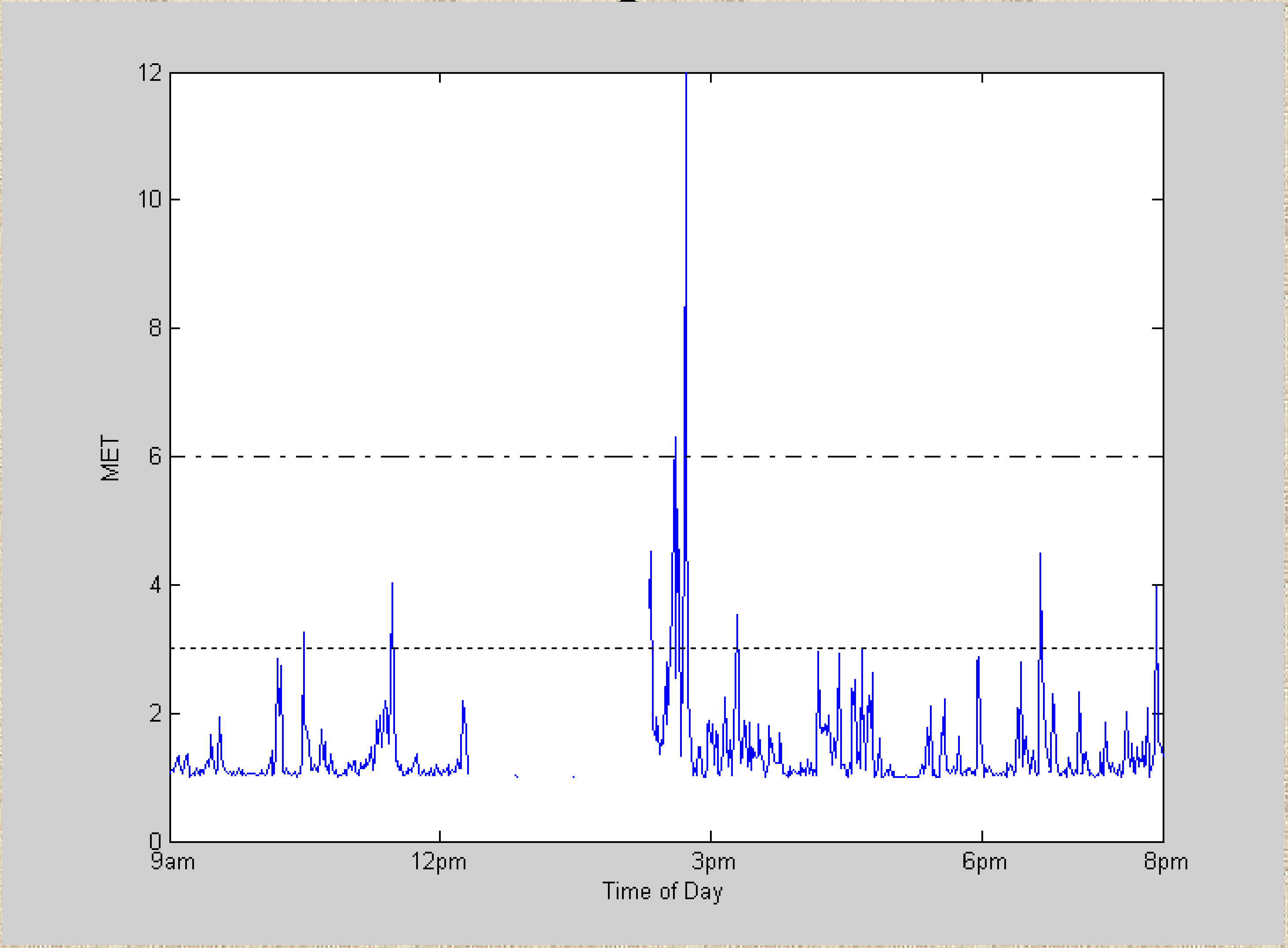
  Example of *incomplete profile*

- **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 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**

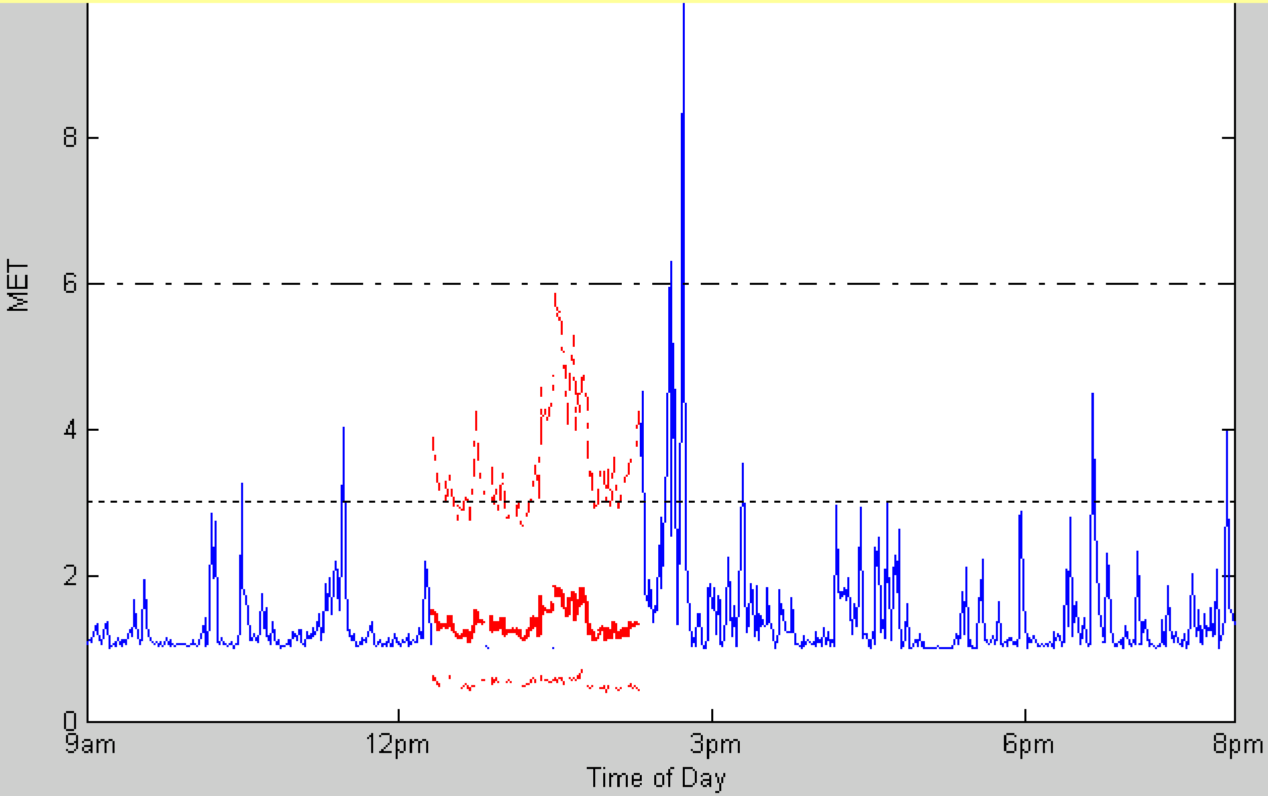
Incomplete Profile



Predictive Distribution

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

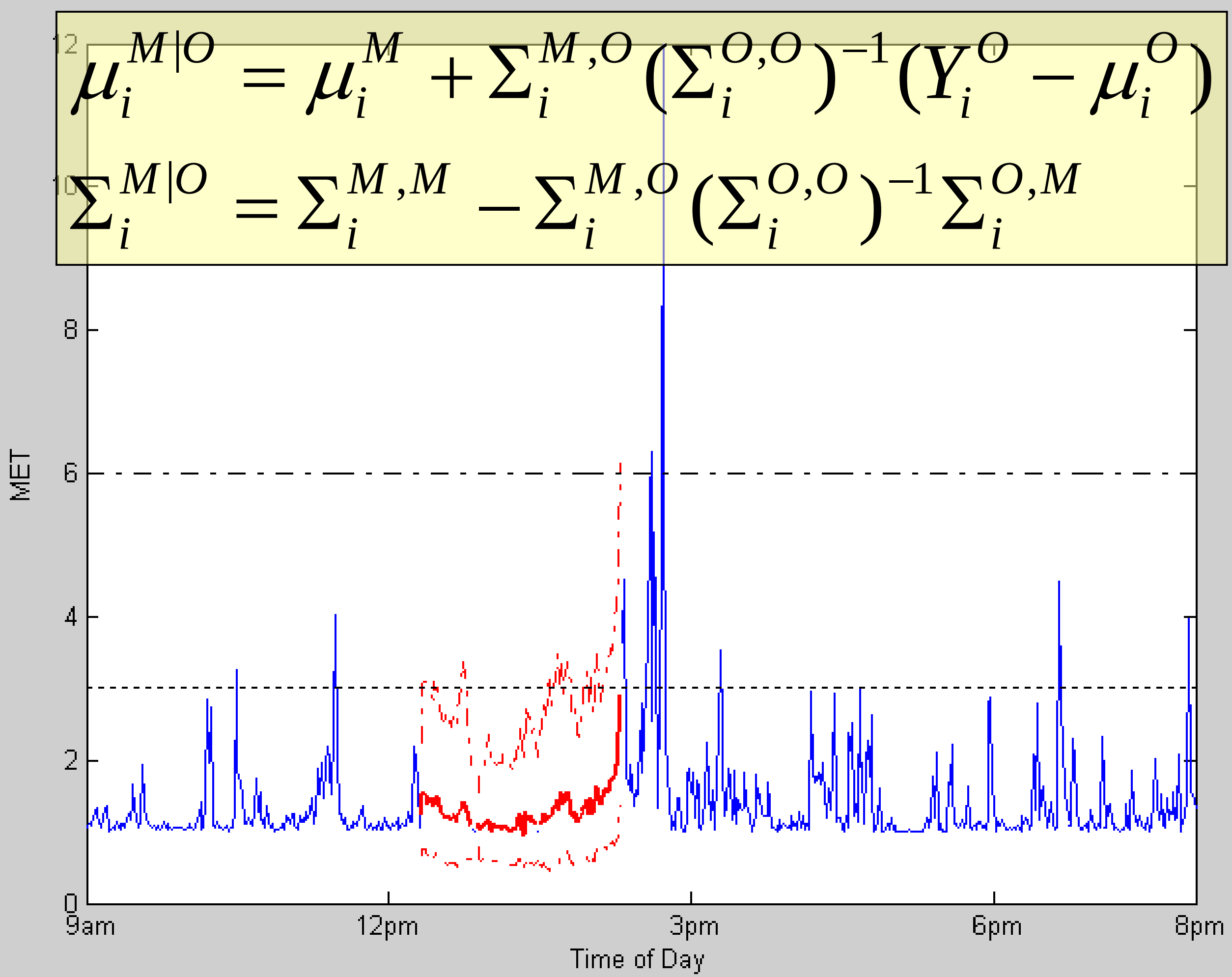
$$\Sigma_i(t_1, t_2) = COV\ \{Y_i(t_1), Y_i(t_2) \mid Y^C\}$$



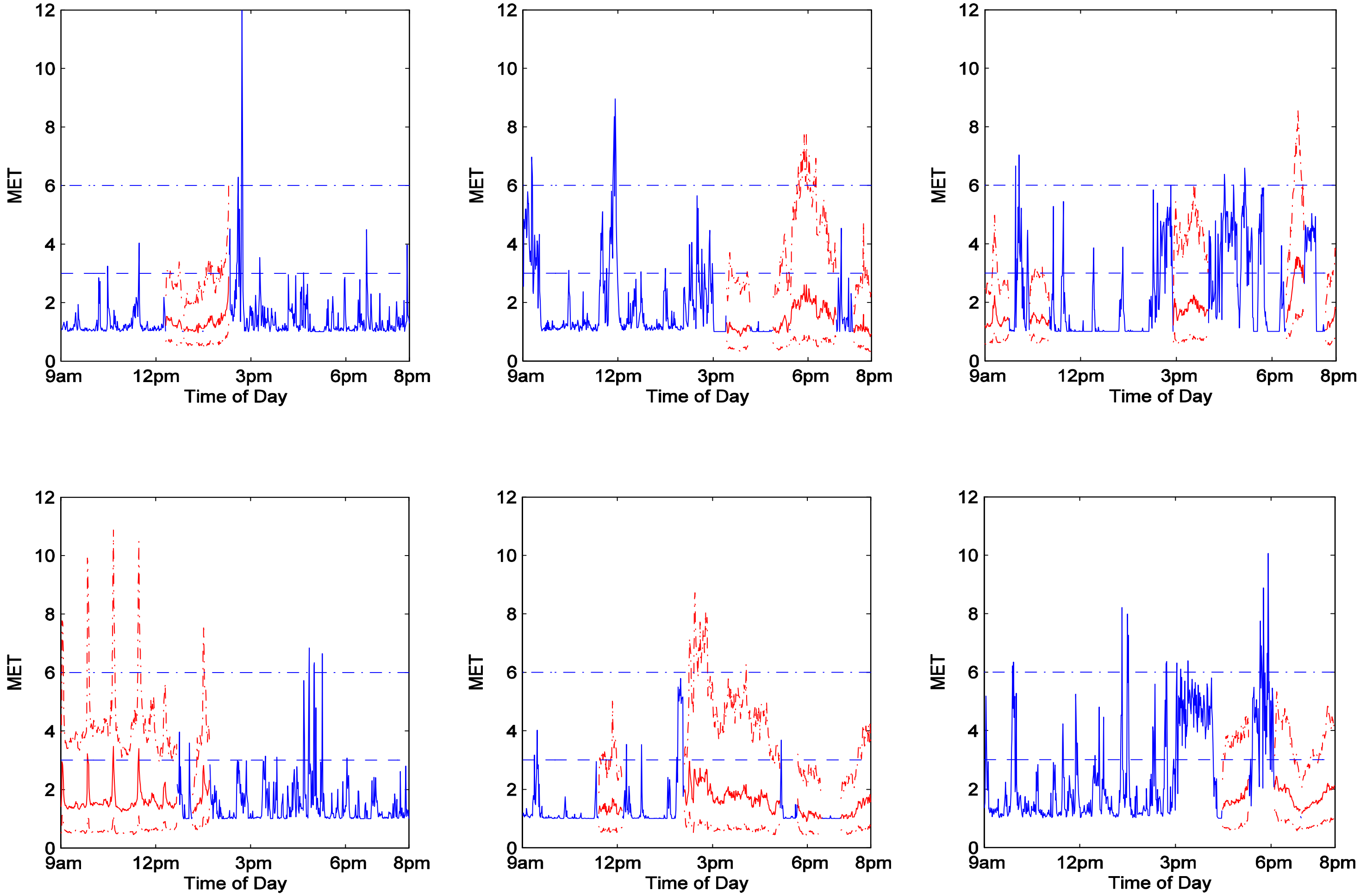
Imputation distribution

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

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



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

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

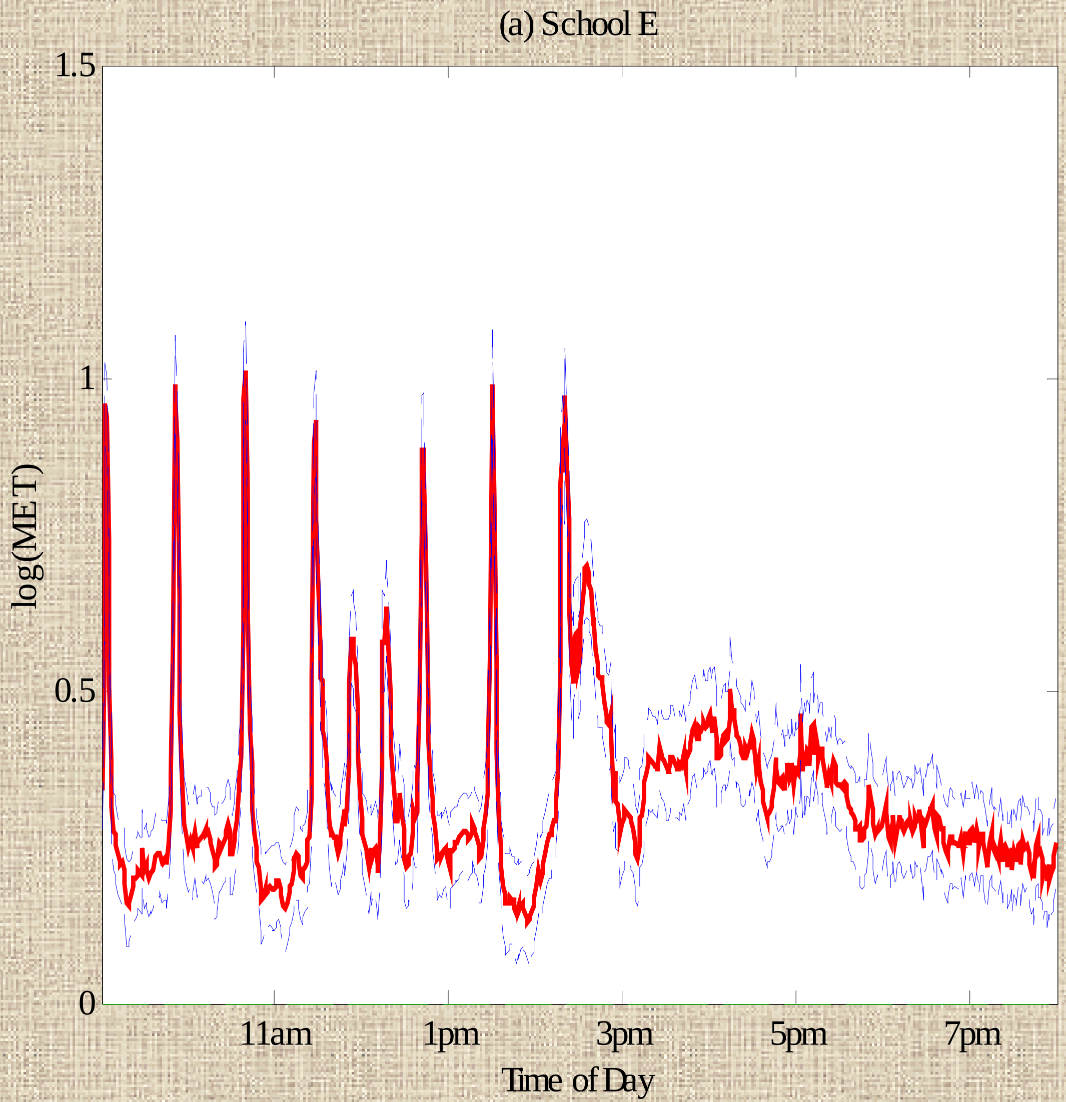
$$V_i(t_1, t_2) = \begin{cases} 0 \text{ if either } t_1 \text{ or } t_2 \text{ obs.} \\ \Sigma_i^{M|O}(t_1, t_2) \text{ otherwise} \end{cases}$$

$$M_i^* = M_i W'$$

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

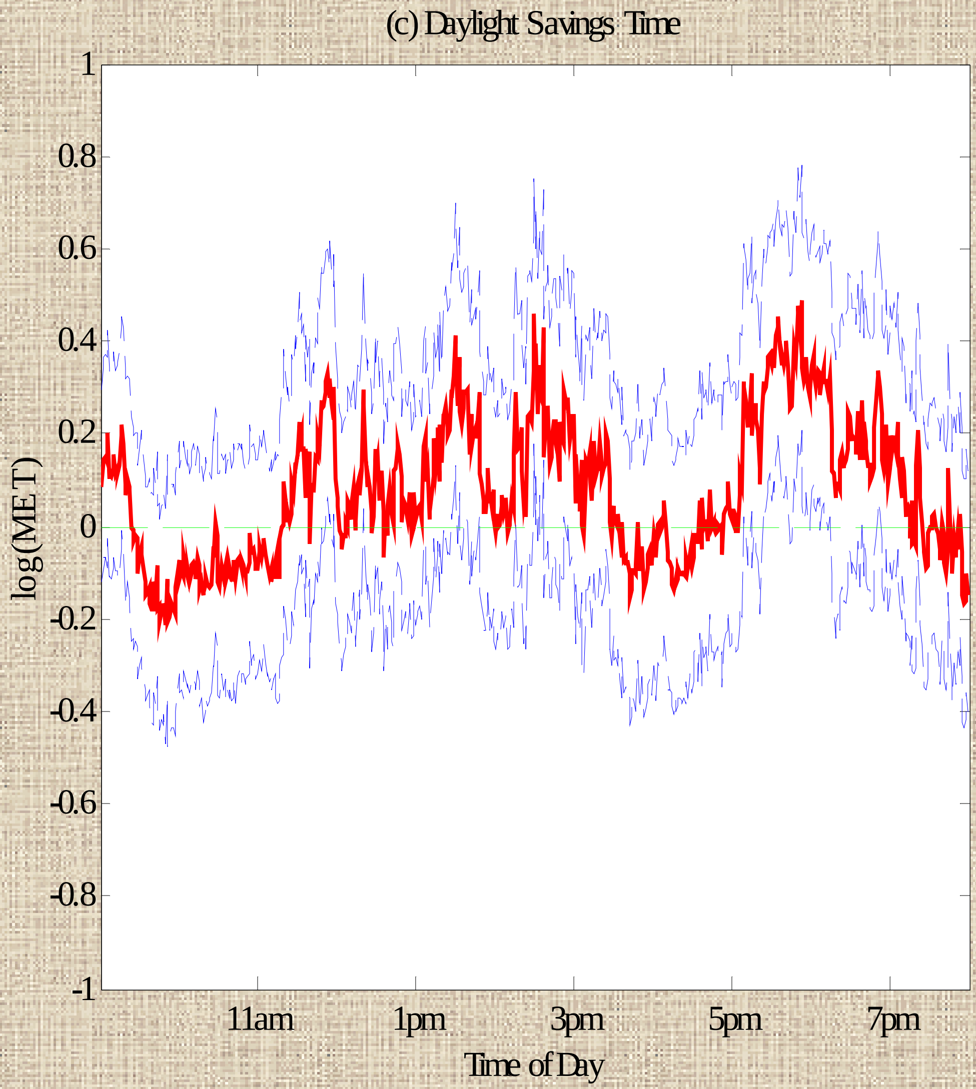
- Add step to MCMC whereby "missing" wavelet coefficients $D_{ijk} \sim N(M^*_{ijk}, V^*_{ijk})$

Selected Results: School Effects



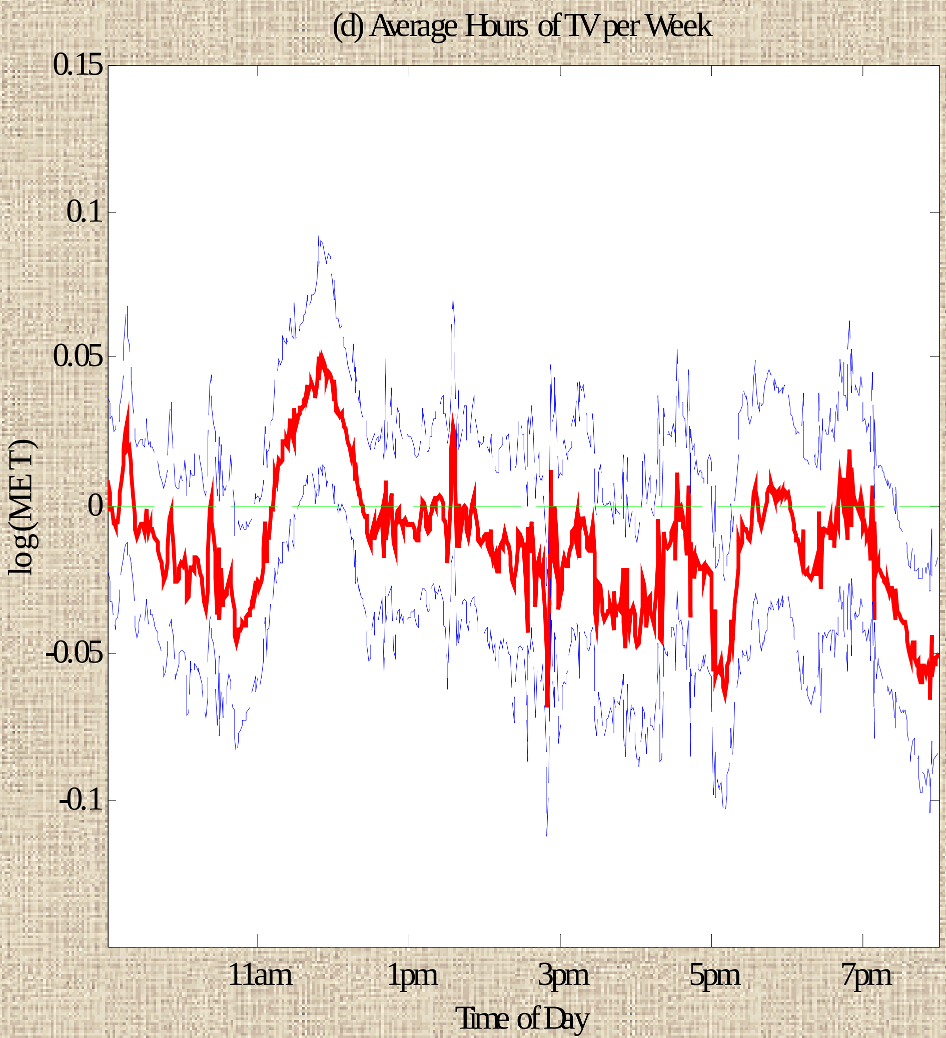
- **School schedules evident in effects**
  - Spikes every 48min (changing classes)
  - 3 lunch periods
  - School out at 2:15pm
- **Not so evident in individual curves**

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

- **<u>Work presented here is from 2 papers</u>**

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*, to appear.**

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

Accelerometer Example

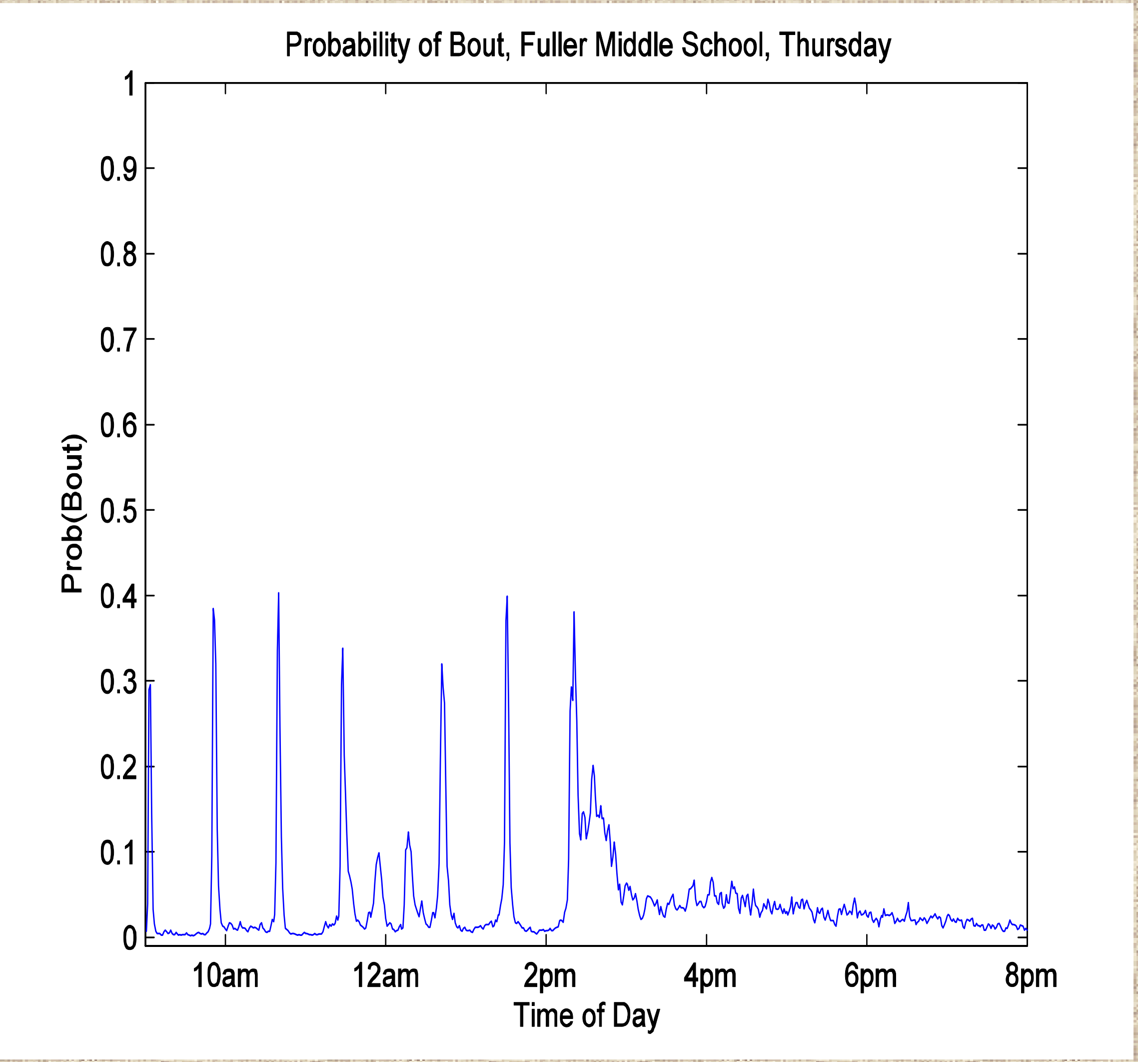
Let $Y_{ij}(t)$ be accelerometer profile on day j from child i

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

- **X_{ik}= child-level covariates (school, race, gender, BMI, % body fat, avg hrs of TV/week)**

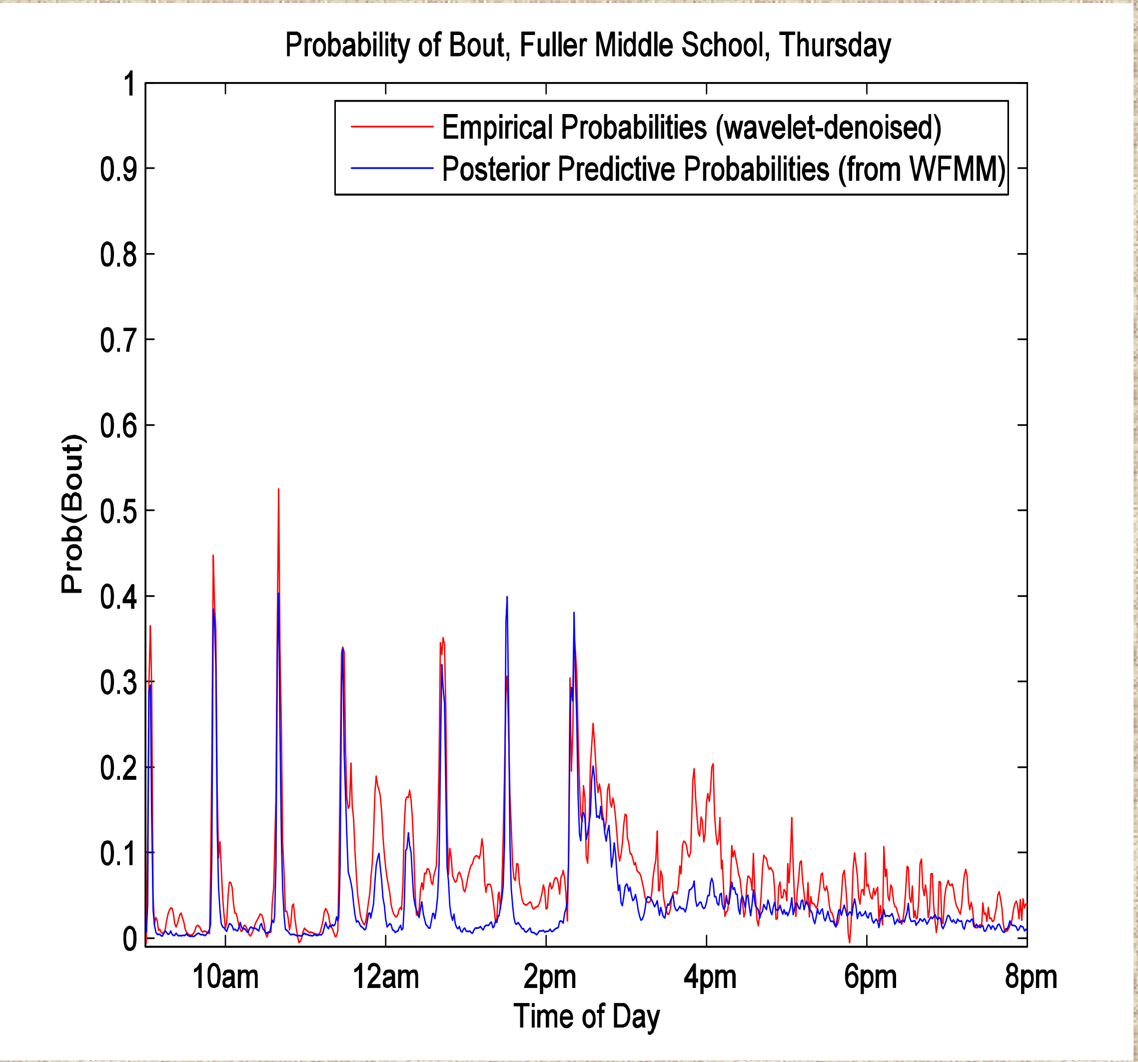
  X_{ijk}=day-level covariates (day-of-week, DST)
- **$B_0(t)$ = overall mean profile**
  $B_k^{child}(t)$ = effect functions for child-level covariates
  $B_k^{day}(t)$ = effect functions for day-level covariates
- **$U_i(t)$ = Random effect function for child i**

Results: Bouts



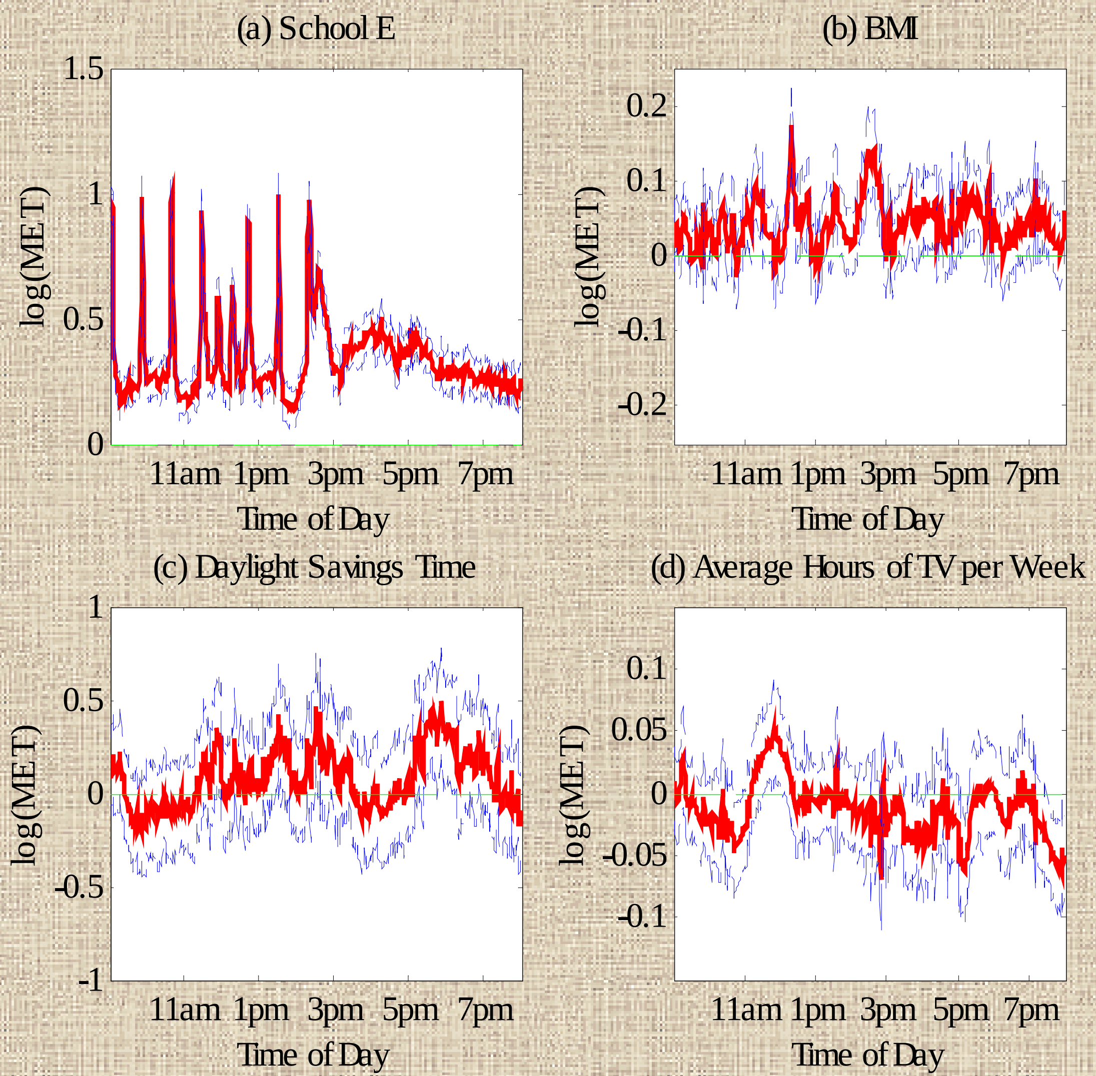
- **Can compute posterior predictive probabilities of bouts for children**

Results: Bouts



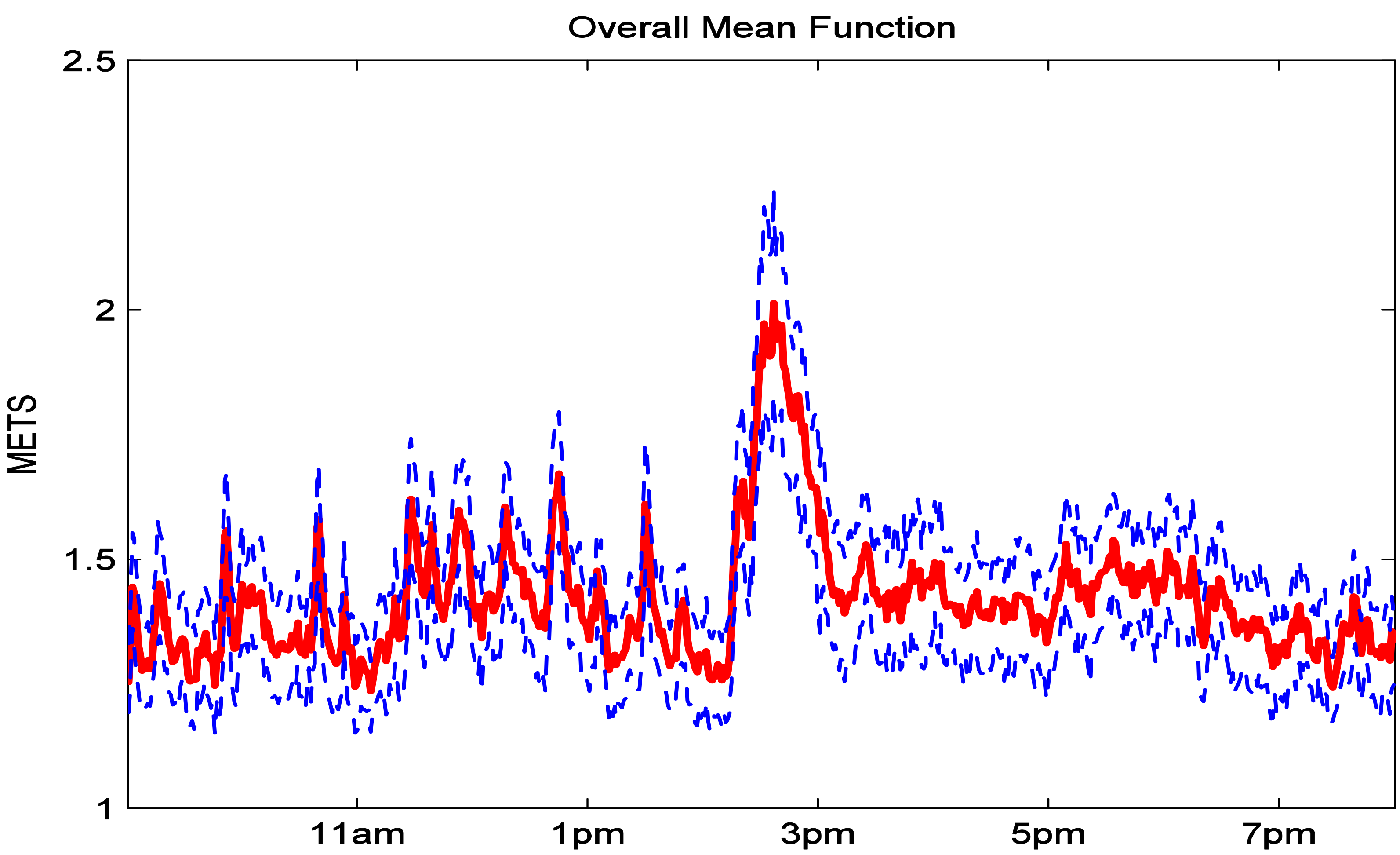
- **Can compute posterior predictive probabilities of bouts for children**
- **Model-based predictive probabilites not far from empirically-estimated probabilities**
- **May want heavier tails**

Some Results

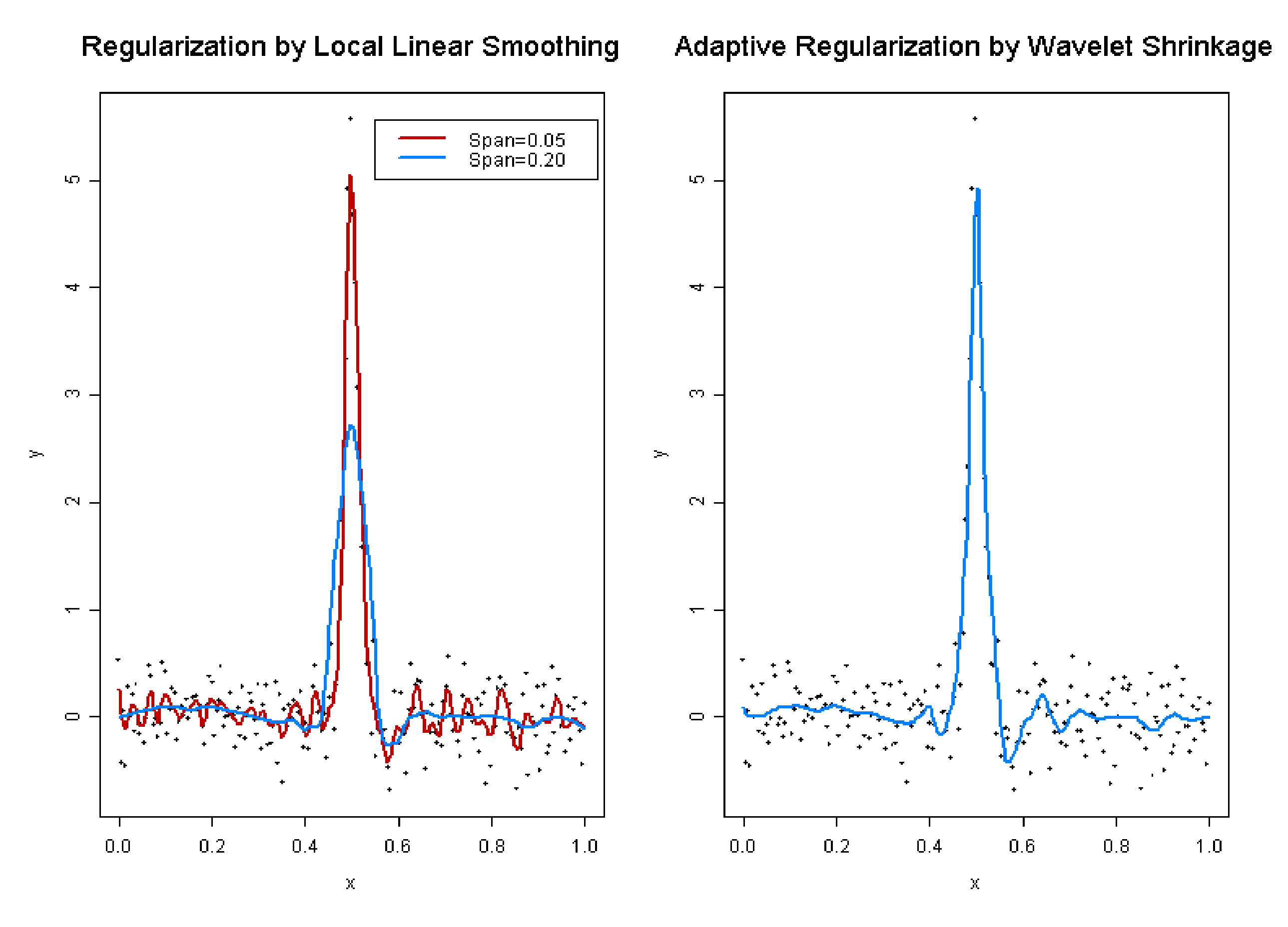


- **School #5:**
  - **Spikes every 48min (changing classes)**
  - **3 lunch periods**
  - **School out at 2:15pm**
- **BMI effect positive**
  - **Artifact of preprocessing?**
- **Daylight Savings Time**
  - **More active after DST**
  - **Especially 2-3pm, 5-7pm**
- **Avg hrs TV/week**
  - **More TV=less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

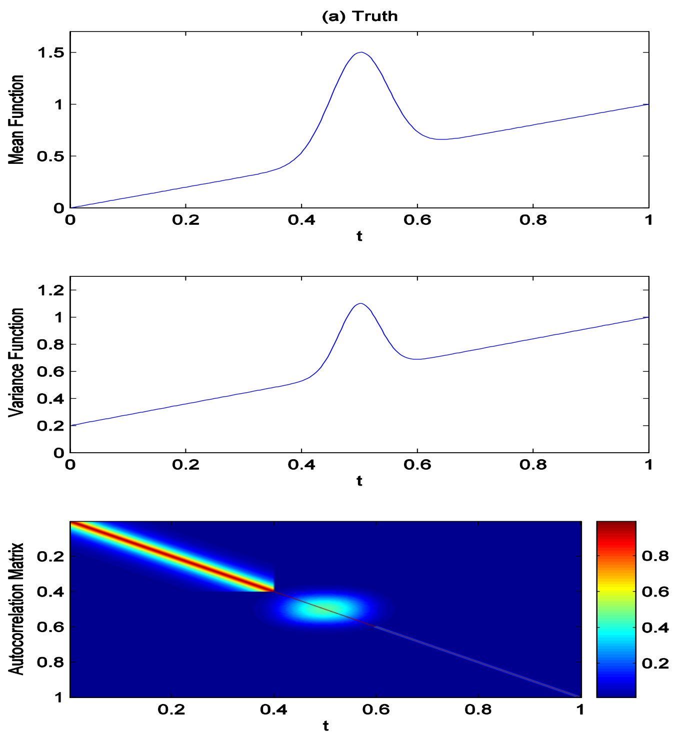
Results



Adaptive Regularization

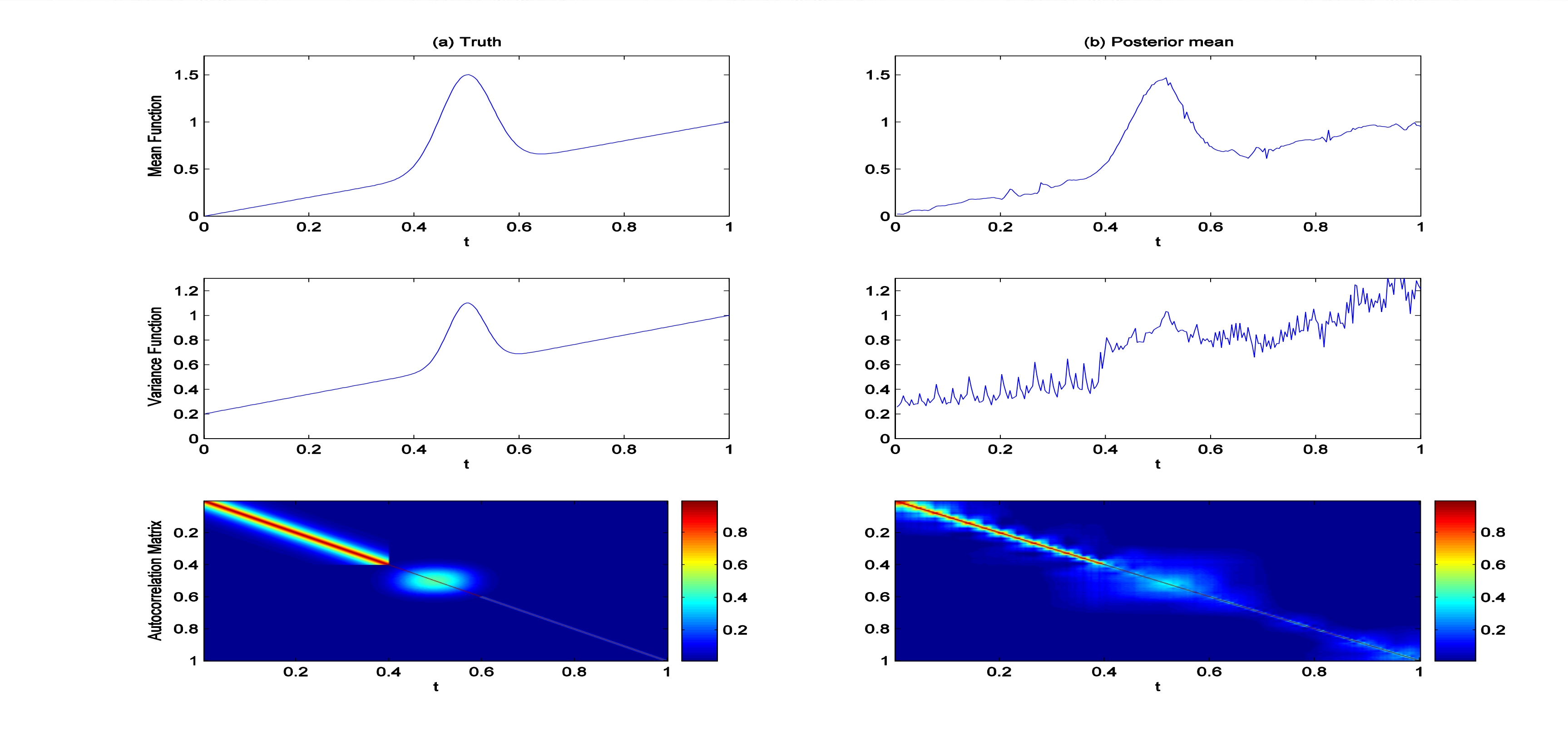


Simulation: Covariance Structure



- **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

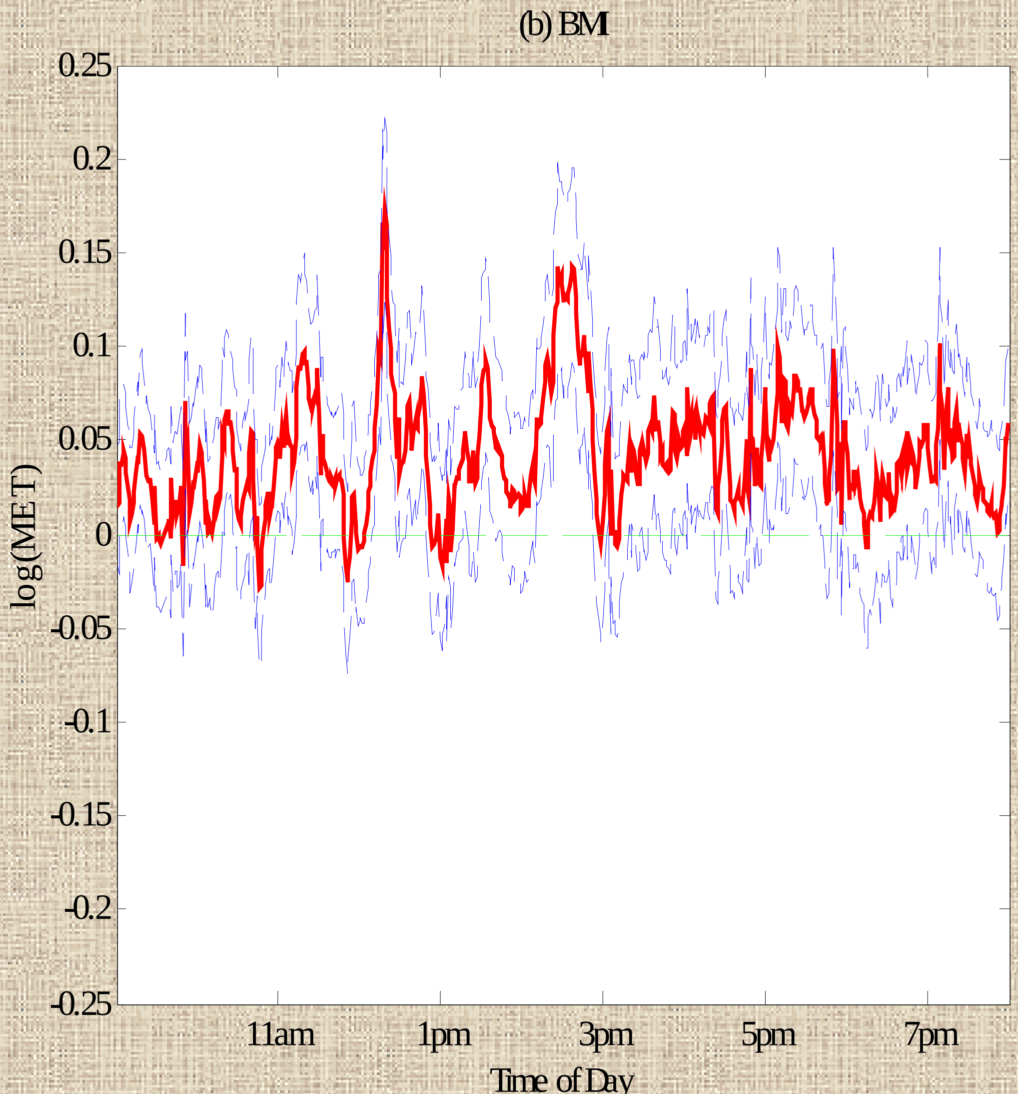
Model Each Column Separately

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

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

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

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* (plot) nonparametric estimates of function**
- **Morris, et al. (2003) extended to hierarchical functional model (Bayesian)**
- **Morris and Carroll (2006) extended to general functional mixed model framework (wavelet-based functional mixed model)**