



Journal of the American Statistical Association

Publication details, including instructions for authors and subscription information:

<http://amstat.tandfonline.com/loi/uasa20>

Using Wavelet-Based Functional Mixed Models to Characterize Population Heterogeneity in Accelerometer Profiles

Jeffrey S. Morris, Cassandra Arroyo, Brent A. Coull, Louise M. Ryan, Richard Herrick and Steven L. Gortmaker

Jeffrey S. Morris is Associate Professor and Richard Herrick is Software Engineer, Department of Biostatistics & Applied Mathematics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030. Cassandra Arroyo is Research Instructor, Social Epidemiology Research Division, Morehouse School of Medicine, Atlanta, GA 30310. Brent A. Coull is Associate Professor, Department of Biostatistics, Louise M. Ryan is Professor, Department of Biostatistics, and Steven L. Gortmaker is Professor, Faculty of Public Health, Harvard School of Public Health, Boston, MA 02115. Jeffrey Morris, Louise Ryan, and Brent Coull were supported by grants from the National Cancer Institute (CA-107304 to Morris, CA-48061 to Ryan, and ES012044 to Coull). Steven Gortmaker was supported by National Institutes of Child Health and Human Development (grant HD-30780) and the Centers for Disease Control and Prevention (Prevention Research Centers grants U48/CCU115807 and U48/DP00064-00S1). The authors thank the associate editor and two reviewers whose comments led to a substantially improved manuscript.

Available online: 01 Jan 2012

To cite this article: Jeffrey S. Morris, Cassandra Arroyo, Brent A. Coull, Louise M. Ryan, Richard Herrick and Steven L. Gortmaker (2006): Using Wavelet-Based Functional Mixed Models to Characterize Population Heterogeneity in Accelerometer Profiles, *Journal of the American Statistical Association*, 101:476, 1352-1364

To link to this article: <http://dx.doi.org/10.1198/016214506000000465>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://amstat.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Using Wavelet-Based Functional Mixed Models to Characterize Population Heterogeneity in Accelerometer Profiles: A Case Study

Jeffrey S. MORRIS, Cassandra ARROYO, Brent A. COULL, Louise M. RYAN, Richard HERRICK, and Steven L. GORTMAKER

We present a case study illustrating the challenges of analyzing accelerometer data taken from a sample of children participating in an intervention study designed to increase physical activity. An accelerometer is a small device worn on the hip that records the minute-by-minute activity levels throughout the day for each day it is worn. The resulting data are irregular functions characterized by many peaks representing short bursts of intense activity. We model these data using the wavelet-based functional mixed model. This approach incorporates multiple fixed-effects and random-effects functions of arbitrary form, the estimates of which are adaptively regularized using wavelet shrinkage. The method yields posterior samples for all functional quantities of the model, which can be used to perform various types of Bayesian inference and prediction. In our case study, a high proportion of the daily activity profiles are incomplete (i.e., have some portion of the profile missing), and thus cannot be modeled directly using the previously described method. We present a new method for stochastically imputing the missing data that allows us to incorporate these incomplete profiles in our analysis. Our approach borrows strength from both the observed measurements within the incomplete profiles and from other profiles, from the same child as well as from other children with similar covariate levels, while appropriately propagating the uncertainty of the imputation throughout all subsequent inference. We apply this method to our case study, revealing some interesting insights into children's activity patterns. We point out some strengths and limitations of using this approach to analyze accelerometer data.

KEY WORDS: Accelerometer; Bayesian inference; Functional data analysis; Missing data; Mixed models; Wavelet; Wavelet regression.

1. INTRODUCTION

Physical activity during childhood and adolescence has an impact on a number of different aspects of subsequent health and well-being. Studies have suggested that increased physical activity during the childhood and adolescent years can (1) help build and maintain healthy bones, muscles, and joints; (2) help control weight, build lean muscles, and reduce fat; (3) prevent or delay the development of hypertension and help reduce blood pressure in some adolescents with hypertension; (4) reduce feelings of depression and anxiety; (5) positively influence mental health and feelings of independence; and (6) positively influence the development of self-esteem (Centers for Disease Control 2000, 2001; Dietz 1998; Pearce 1999; Strauss, Rodzilshy, Burack, and Colin 2001).

The systematic study of physical activity is complicated by the technical challenges of measurement. One common approach is to use self-report questionnaires, but these are known to correlate poorly with actual activity levels, often overestimating them (Matthews and Freedson 1995; Epstein, Paluch, Coleman, Vito, and Anderson 1996; Coleman, Saelens, Wiedrich-Smith, Finn, and Epstein 1997). Recently, more objective approaches have been used, including computerized accelerometers, small motion sensors that can digitally record

minute-by-minute activity levels. Their small size and automatic measurement capabilities make them well suited for objectively quantifying activity levels in large populations over long periods (Westerterp 1999). They are being increasingly used to monitor activity levels in large surveillance studies and intervention trials in children as well as adults (Rowlands, Eston, and Ingledeu 1999; Gortmaker 2002; Going et al. 2003; Talbot, Gaines, Huynh, and Metter 2003; Algase, Beattie, Leitsch, and Beel-Bates 2003; Abbott and Davies 2004). In this article we consider accelerometer data obtained from Planet Health (Gortmaker et al. 1999), a Boston-area study of a school-based intervention designed to reduce obesity in middle school youth by changing key physical activity and dietary risk factors.

The Planet Health study used the TriTrac-R3D activity monitor (Hemokentics Inc., Madison, WI), which is worn in a small pouch on the hip and provides minute-by-minute acceleration counts computed from motion sensors in three-dimensional space. The resulting daily profiles are irregular functional data characterized by many peaks representing short bursts of intense activity. Figure 1 presents two daily profiles from 9 AM to 8 PM from two different children in the Planet Health study. Other accelerometers yield similar data. In studies involving accelerometers, it is standard practice to obtain four to eight daily profiles for each subject in the study. These profiles can be analyzed directly or used to calibrate and validate questionnaire-based measures of physical activity. In practice, many of the profiles are incomplete, having periods for which the activity levels are missing (e.g., if the accelerometer is removed).

Approaches currently used to analyze these data are based on a simple summary, such as the average daily activity level (Talbot et al. 2003), 30-minute averages (Going et al. 2003; Cradock et al. 2004), or the proportion of time spent above specific cutoff levels corresponding to sedentary, moderately vigorous, and vigorous activity (Rowlands et al. 1999; Goldfield,

Jeffrey S. Morris is Associate Professor (E-mail: jeffmo@mdanderson.org) and Richard Herrick is Software Engineer, Department of Biostatistics & Applied Mathematics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030. Cassandra Arroyo is Research Instructor, Social Epidemiology Research Division, Morehouse School of Medicine, Atlanta, GA 30310. Brent A. Coull is Associate Professor, Department of Biostatistics, Louise M. Ryan is Professor, Department of Biostatistics, and Steven L. Gortmaker is Professor, Faculty of Public Health, Harvard School of Public Health, Boston, MA 02115. Jeffrey Morris, Louise Ryan, and Brent Coull were supported by grants from the National Cancer Institute (CA-107304 to Morris, CA-48061 to Ryan, and ES012044 to Coull). Steven Gortmaker was supported by National Institutes of Child Health and Human Development (grant HD-30780) and the Centers for Disease Control and Prevention (Prevention Research Centers grants U48/CCU115807 and U48/DP00064-00S1). The authors thank the associate editor and two reviewers whose comments led to a substantially improved manuscript.

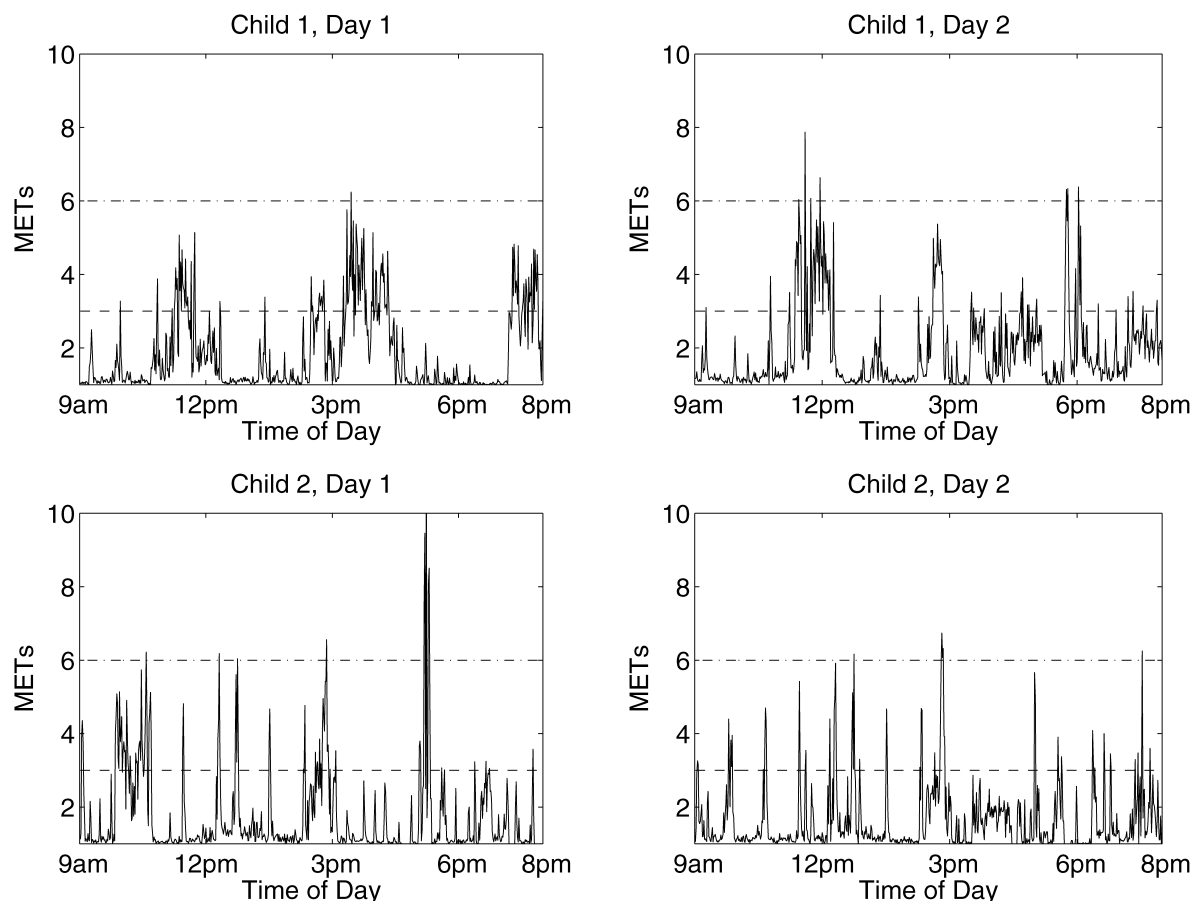


Figure 1. Sample Profiles From 2 Days for 2 Different Children, From 9 AM to 8 PM. Activity levels > 6 METs are considered vigorous, whereas activity levels between 3 and 6 METs are moderate.

Kalakanis, Ernst, and Epstein 2000; Westerterp 2002; Abbott and Davies 2004). Although these summaries are a reasonable starting point, they have serious limitations, mostly because they do not make full use of the rich information contained in the functional data. The proportions and daily averages do not account for time-of-day variability, and the proportions do not even use the actual activity levels. The arbitrary binning inherent to taking 30-minute averages results in attenuation of the signal, and conclusions may be sensitive to the choice of endpoints for the bins. A second important limitation of these approaches is that they may not effectively handle the missing data in the incomplete profiles, which is a serious concern considering the large degree of missingness that can occur in these data.

Methods that model the functional profiles in their entirety have the potential to extract more information from the data compared with methods based on arbitrarily chosen summary measures. For example, these methods are able to identify at what time of day children are most active, detect covariate effects that vary throughout different times of the day, and allow the variability in activity levels to differ across time of day. Functional data analysis (FDA) (Ramsay and Silverman 1997) is a general term for approaches that consider the functional profiles as single entities rather than simply collections of individual data points. A key challenge with these approaches is that they must deal both with regularization (i.e., borrowing strength across measurements within a profile) and replication

(i.e., combining information across multiple profiles) to make inferences on the population from which they came.

The irregularity of accelerometer profiles makes modeling them in this manner especially challenging. They clearly cannot be represented by simple parametric structures, and the most commonly used nonparametric methods involve smoothing by kernels or splines, which are not well suited for modeling functional data with many local features like peaks. Because there are typically multiple daily profiles per child, the possibility of correlation between profiles must be accounted for. In addition, many factors can potentially affect the profiles, including day of week, time of year, school, and child-level characteristics, including sex, age, and obesity status. The effects of these factors may vary by time of day. A suitable modeling approach should be able to simultaneously account for functional effects of multiple covariates like these, some of which are of interest to investigators and others that are potential confounders.

Morris and Carroll (2006) introduced a new method for analyzing functional data that accommodates these concerns. This method is based on the functional mixed model, a generalization of the linear mixed model (Laird and Ware 1982) to functional data. This method is appropriate for modeling irregular profiles with many peaks, because regularization is accomplished using adaptive wavelet shrinkage. The functional mixed model allows multiple fixed-effects functions of arbitrary form, which can simultaneously represent time-varying effects for multiple covariates that are discrete or continuous in

nature. It also allows multiple random-effects functions of arbitrary form, which can model the correlation between profiles from the same child. The reasonably flexible assumptions made on the covariance matrices for the random effects and residual errors accommodate a broad range of nonstationary covariance structures. The method's output consists of posterior samples for all quantities in the model, which can be used to perform Bayesian inference.

Whereas the method of Morris and Carroll (2006) cannot handle incomplete profiles, it is based on a unified modeling approach, making it possible to develop rigorous methods for imputing the missing data. In this article we introduce new missing-data methods for the wavelet-based functional mixed model (WFMM). Briefly, our approach is to stochastically sample the missing regions of the incomplete profiles from their approximate posterior predictive distributions, as estimated from a preliminary model fit using only the complete daily profiles. This approach borrows strength from both the observed measurements within the incomplete profile and other profiles and from the same child as well as from other children with similar covariate levels, and appropriately propagates the uncertainty due to the imputation throughout any subsequent inference.

We apply these methods to analyze the TriTrac-R3D data from the Planet Health study to explore patterns in the children's activity level profiles, to identify factors related to activity levels, and to characterize the relative contributions of the day-to-day and child-to-child sources of variability to aid the design of future studies. We also assess the model fit and discuss the strengths and limitations of this approach for analyzing accelerometer data.

The remainder of the article is organized as follows. In Section 2 we describe the Planet Health study and explore the characteristics of the TriTrac-R3D accelerometer data from that study. In Section 3 we briefly overview wavelets and the WFMM, and in Section 4 we introduce new methods for imputing missing data for incomplete profiles in the WFMM setting. In Section 5 we present the results of our case study analysis, and in Section 6 we discuss these results and assess the strengths and limitations of this approach for analyzing accelerometer data. We provide the derivation of the posterior predictive distributions used to impute the missing data in the Appendix.

2. ACCELEROMETER DATA FROM THE PLANET HEALTH STUDY

Childhood obesity is a major health problem in the United States. Planet Health was a school-based intervention designed to reduce obesity in middle school youth by changing key physical activity and dietary risk factors. The Planet Health study (Gortmaker et al. 1999) involved 10 Boston-area middle schools, which were paired up and randomized to either receive the intervention or serve as control schools. For each of the 1,295 children involved in the study, various nutritional, behavioral, and health-related outcomes were measured at baseline (Fall 1995) and follow-up (Spring 1997). A subsample of 256 children was randomly selected to participate in a substudy focusing on the technical issue of objectively measuring physical activity using accelerometers. The activity levels of each child in the substudy were monitored using the TriTrac-R3D

activity monitor for one or two 4-day sessions between February and May 1997.

The TriTrac-R3D activity monitor is a pocket-sized motion sensor that, when worn on the hip in a nylon pouch, measures motion in three planes (horizontal, vertical, and medio-lateral). The device stores data over several days, after which the data can be downloaded to a computer to provide a minute-by-minute record of movement in each dimension. The counts in the three planes may be combined into a single activity level for each minute by computing the vector magnitude. The vector magnitudes may then be converted into estimated energy expenditure per minute due to activity (EE). The TriTrac-R3D uses a proprietary formula involving the subject's weight to compute EE; Gortmaker (2002) calculated this formula, and then converted the EEs into METs (metabolic units) using $MET = (EE + BMR)/BMR$, where BMR is the child's basal metabolic rate, estimated based on sex, age, and weight using the World Health Organization equations (*Recommended Daily Allowances* 2000). The minimum activity level is 1 MET, which indicates the child is stationary; activities yielding levels between 3 and 6 METs (e.g., walking) are considered moderately intense, and activities yielding levels of more than 6 METs (e.g., running) are considered vigorous (Rowlands, Thomas, Eston, and Topping 2004).

We focused on the weekday accelerometer profiles from children in the five control schools that did not receive the intervention. This dataset consisted of 550 daily profiles from 112 children. Each daily profile contained 1,440 measurements, the minute-by-minute activity levels for a single day, measured in METs. If no movement was recorded for 30 or more consecutive minutes, then we assumed that the child removed the monitor, and the corresponding measurements were considered missing. Figure 2 contains a heat map of the 550 daily profiles. A heat map is a graphical device useful for representing high-dimensional datasets. Each row corresponds to a single profile, and each column is 1 minute in the day. The color of the pixel (i, j) indicates the activity level of profile i at time t_j , with higher activity levels represented by lighter colors. For this graph, we coded missing measurements as 0 (black) and censored moderate and vigorous activity levels (>3.0 METs, white) to improve the contrast of the figure. The white dashed horizontal lines delineate the schools, which we refer to as schools A, B, C, D, and E. Note the imbalance in the number of profiles obtained from each school.

There were a large number of missing measurements in the data. Supplementary material available on the first author's website (http://biostatistics.mdanderson.org/Morris/papers_files/TriTrac_supplement.html) contains a plot of the proportion of daily profiles with nonmissing measurements as a function of the time of the day, along with a histogram of the proportion of minutes missing from 9 AM to 8 PM for the 550 profiles. We focused our analysis on the data from 9 AM to 8 PM, because most of the profiles were missing outside this region. Only 95 of the 550 profiles were complete from 9 AM to 8 PM, and 187 profiles had more than 80% of this region missing. We focused on the 292 profiles that were at least 50% complete from 9 AM to 8 PM, which were from 106 children. All of these profiles were obtained between February 10, 1997 and May 28, 1997. We also considered the criteria of 80% complete or 20% complete and found similar results.

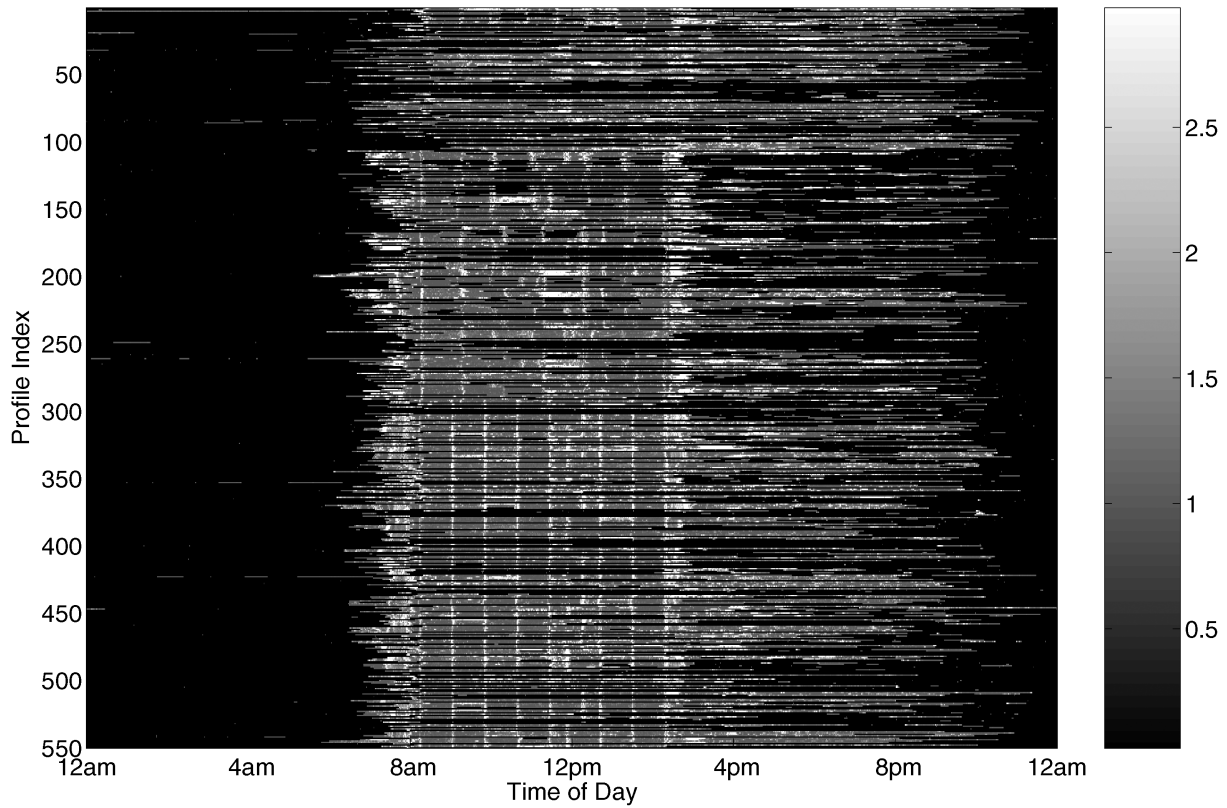


Figure 2. Heat Map of All Weekday Accelerometer Profiles. The rows are the daily profiles, ordered by school and child within school. The columns are the minutes of the day, from 12:01 AM through 12:00 midnight. The black regions of the plot indicate regions of the profiles where the activity level is missing. Because the minimum activity level is 1 MET, all nonmissing activity levels are some shade of gray, not black. The white regions indicate bouts, or regions where the activity profile is > 3.0 METs. The dotted white lines mark the divisions between profiles from different schools.

For each child, we had a number of covariate measurements, including school, race/ethnicity, sex, age, weight, height, body mass index (BMI), triceps skinfold (measured by skinfold calipers applied to the triceps), and average number of hours spent watching television per day. We also had a record of the day of week and the calendar date on which each profile was obtained.

We were interested in assessing how activity levels tended to vary throughout the day, across schools, across different days of the week, over time from early to late spring, and across various child-level covariates. We also were interested in assessing the relative variability from day to day and child to child, to provide design recommendations for future studies. We accomplished these goals by applying the WFMM to these data, after adapting the procedure to accommodate incomplete profiles. We assessed how well this model fit the data, and specifically checked how closely it predicted the frequency of bouts of moderate and vigorous activity for groups of children as a function of time of day.

3. WAVELETS AND FUNCTIONAL MIXED MODELS

3.1 Brief Review of Wavelets

Wavelets are basis functions that can be used to represent other functions, often very parsimoniously. A basic introduction to wavelets has been given by Vidakovic (1999). A wavelet

series approximation to a continuous function $y(t)$ is given by

$$y(t) = \sum_k c_{J,k} \phi_{J,k}(t) + \sum_{j=1}^J \sum_k d_{j,k} \psi_{j,k}(t), \quad (1)$$

where J is the number of scales and k ranges from 1 to K_j , the number of coefficients at scale j . The functions $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ are wavelet basis functions that provide a location-scale decomposition of the observed function. They are dilations and translations of a father and mother wavelet function, $\phi(t)$ and $\psi(t)$, with $\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k)$ and $\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)$. The coefficients $c_{J,k}, d_{J,k}, \dots, d_{1,k}$ are the *wavelet coefficients*. The $c_{J,k}$, called the *smooth coefficients*, represent smooth behavior of the function at coarse scale J ; the $d_{j,k}$, called *detail coefficients*, represent deviations of the function at scale j , where smaller j correspond to finer scales. The wavelet coefficients at scale j essentially correspond to differences of averages of 2^{j-1} time units, spaced 2^j units apart. In addition, by examining the phase properties of the wavelet bases, we can associate each wavelet coefficient on each scale with a specific set of time points.

Theoretically, each coefficient may be obtained by taking the inner product of the function and the corresponding wavelet basis function, although in practice more efficient approaches are used. If the function is sampled on an equally spaced grid of length T , then the coefficients may be computed using a pyramid-based algorithm called the *discrete wavelet transform*.

(DWT) in just $O(T)$ operations. Applying the DWT to a row vector of observations \mathbf{y} produces a row vector of wavelet coefficients, $\mathbf{d} = (c_{J,1}, \dots, c_{J,K_J}, d_{J,1}, \dots, d_{1,K_1})$. This transformation is a linear projection, so it may also be represented by matrix multiplication, $\mathbf{d} = \mathbf{y}W'$, with W' being the DWT projection matrix. Similarly, the inverse discrete wavelet transform (IDWT) may be used to project wavelet coefficients back into the data space and can also be represented by matrix multiplication by the IDWT projection matrix W , the transpose of the DWT projection matrix. We use the method implemented in the Matlab Wavelet Toolbox (Misiti, Misiti, Oppenheim, and Poggi 2000) to compute the DWT, although other implementations can just as well be used. Supplementary material (http://biostatistics.mdanderson.org/Morris/papers_files/TriTrac_supplement.html) describes how this implementation deals with data for which T is not a power of 2 and discusses some of its properties.

Wavelets can be used to perform nonparametric regression using the following three-step procedure. First, noisy data \mathbf{y} are projected into the wavelet domain using the DWT, yielding empirical wavelet coefficients \mathbf{d} . The coefficients are then thresholded by setting to 0 any coefficients smaller in magnitude than a specified threshold and/or nonlinearly shrunk toward 0 using one of a number of possible frequentist or Bayesian approaches. These result in estimates of the true wavelet coefficients, which would be the wavelet coefficients for the true function if there were no noise. Finally, these estimates are projected back to the original data domain using the IDWT, yielding a denoised nonparametric estimate of the true function. Because most signals may be represented by a small number of wavelet coefficients, yet white noise is distributed equally among all wavelet coefficients, this procedure yields denoised function estimates that tend to retain dominant local features of the function. In this article we refer to this property as *adaptive regularization*, because the function is regularized (i.e., denoised or smoothed) in a way that adapts to the characteristics of the function. This property makes the procedure useful for modeling functions with many local features like peaks. References on wavelet regression have been given by Vidakovic (1999, chaps. 6 and 8), Donoho and Johnstone (1995), Chipman, Kolaczyk, and McCulloch (1997), Vidakovic (1998), Abramovich, Sapatinas, and Silverman (1998), Clyde, Parmigiani, and Vidakovic (1998), and Clyde and George (2000).

Most work in wavelet regression to date has been limited to the single-function case, with a few exceptions. For example, Brown, Fearn, and Vannucci (2001) performed Bayesian variable selection on wavelet coefficients of functions serving as predictors for a scalar response. Chang and Vidakovic (2002) proposed a method to obtain a regularized estimate for the mean function over a sample of curves. Morris, Vannucci, Brown, and Carroll (2003) developed a wavelet-based method for analyzing hierarchical functional data, in which the functions are sampled from a nested hierarchical design. This method produces adaptively regularized estimates and inference for the overall mean function and individual profiles. Morris and Carroll (2006) generalized this work to the functional mixed models setting.

3.2 Functional Mixed Models

Suppose that we observe n functional profiles, $Y_i(t)$, $i = 1, \dots, n$, all defined on the compact set $\mathcal{T} \in \mathbb{R}^1$. A functional mixed model for these profiles is given by

$$Y_i(t) = \sum_{j=1}^p X_{ij}B_j(t) + \sum_{k=1}^m Z_{ik}U_k(t) + E_i(t), \quad (2)$$

where X_{ij} are covariates, $B_j(t)$ are functional (time-varying) fixed effects, Z_{ik} are elements of the design matrix for functional random effects $U_k(t)$, and $E_i(t)$ are residual error processes. Here we assume that $U_k(t)$ are independent and identically distributed (iid) mean-0 Gaussian processes with covariance surface $Q(t_1, t_2)$, and $E_i(t)$ are iid mean-0 Gaussian processes with covariance surface $S(t_1, t_2)$, with $U_k(t)$ and $E_i(t)$ independent. Note that Q indicates the covariance across the random-effect functions $k = 1, \dots, m$, and S indicates the covariance across the residual error processes for the n curves, after conditioning on the fixed and random effects. Note that the iid assumptions on the individual random-effect functions and residual error processes makes (2) a special case of the more general functional mixed model presented by Morris and Carroll (2006). Because the response, fixed-effects, random-effects, and residual error processes are all functional, heuristically this model can be viewed as fitting separate mixed models at each time point, but with an extra layer added for regularization, that is, to borrow strength across observations within a function.

Suppose that all observed profiles are sampled on the same equally spaced grid \mathbf{t} of length T . Let Y be the $n \times T$ matrix containing the observed profiles on the grid, with each row containing one observed profile on the grid \mathbf{t} . A discrete, matrix-based version of this mixed model can be written as

$$Y = XB + ZU + E. \quad (3)$$

The matrix X is an $n \times p$ design matrix containing (nonfunctional) covariates; B is a $p \times T$ matrix with rows containing the corresponding *fixed-effect functions* on the grid \mathbf{t} . The quantity B_{ij} denotes the effect of the covariate in column i of X on the response at time t_j . The matrix U is an $m \times T$ matrix with rows containing *random-effect functions* on the grid \mathbf{t} , and Z is the corresponding $n \times m$ design matrix. Each row of the $n \times T$ matrix E contains the residual error process for the corresponding observed profile. We assume that the rows of U are iid $MVN(\mathbf{0}, Q)$ and the rows of E are iid $MVN(\mathbf{0}, S)$, independent of U , with Q and S being $T \times T$ covariance matrices that are discrete approximations to the covariances surfaces in (2) on the grid.

This model is very flexible and can be used to represent a wide range of functional data. The fixed-effect functions may be group mean functions, interaction functions, or functional linear effects for continuous covariates, depending on the structure of the design matrix. The random-effect functions provide a convenient mechanism for modeling between-function correlation, for example, when multiple profiles are obtained from the same individual. The model places no restrictions on the form of the fixed-effect or random-effect functions. Because the forms of the covariance matrices Q and S are also left unspecified, it is typically not feasible to fit this model without first making further assumptions.

3.3 Wavelet-Based Functional Mixed Models

Morris and Carroll (2006) did not fit model (4) directly, but instead projected the observed profiles into the wavelet space, then worked with the wavelet-space version of the model. This allowed the modeling to be done in a more parsimonious and computationally efficient manner and provided a convenient mechanism for adaptively regularizing the fixed-effect functions.

The projection is accomplished by applying the DWT to each row of Y , yielding a matrix of wavelet coefficients $D = YW'$, where W' is the DWT projection matrix. Row i of D contains the wavelet coefficients for profile i , with the columns corresponding to individual wavelet coefficients and double-indexed by scale j and location k . It is easy to show that the wavelet-space version of model (3) is

$$D = XB^* + ZU^* + E^*, \quad (4)$$

where each row of $B^* = BW'$ contains the wavelet coefficients corresponding to one of the fixed-effect functions, each row of $U^* = UW'$ contains the wavelet coefficients for a random-effect function, and $E^* = EW'$ contains the wavelet-space residuals. The rows of U^* and E^* remain independent mean-0 Gaussians, but with covariance matrices $Q^* = WQW'$ and $S^* = WSW'$.

Motivated by the whitening property of the wavelet transform, many wavelet regression methods in the single-function setting assume that the wavelet coefficients for a given function are mutually independent. Morris and Carroll (2006) effectively made this assumption in the WFMM by constraining Q^* and S^* to be diagonal. Allowing the variance components to differ across both wavelet scale j and location k yields $Q^* = \text{diag}(q_{jk})$ and $S^* = \text{diag}(s_{jk})$. This assumption reduces the dimensionality of Q and S from $T(T+1)/2$ to T , while still accommodating a reasonably wide range of nonstationary within-profile covariance structures for both the random-effects and residual error processes. For example, it allows heteroscedasticity and differing degrees of smoothness at different regions of the curves. Figure 1 of Morris and Carroll (2006) illustrates this point. By using transforms from wavelet packet tables (see Percival and Walden 2000, chap. 6), it may be possible to accommodate an even broader class of covariance matrices.

Morris and Carroll (2006) used a Markov chain Monte Carlo (MCMC) scheme to generate posterior samples for quantities of model (4). They used vague proper priors for the variance components and independent mixture priors for the elements of B^* . Specifically, the prior for B_{ijk}^* , the wavelet coefficient at scale j and location k for fixed-effect function i , was $B_{ijk}^* = \gamma_{ijk} \text{normal}(0, \tau_{ij}) + (1 - \gamma_{ijk})\delta_0$, with $\gamma_{ijk} \sim \text{Bernoulli}(\pi_{ij})$ and δ_0 being a point mass at 0. This prior is commonly used in Bayesian implementations of wavelet regression (see, e.g., Clyde et al. 1998; Abramovich et al. 1998). Using this mixture prior causes the posterior mean estimates of the B_{ijk}^* to be nonlinearly shrunken toward 0, resulting in adaptively regularized estimates of the fixed-effect functions. The parameters τ_{ij} and π_{ij} are *regularization parameters* that determine the relative trade-off of variance and bias in the nonparametric estimation. They may be either prespecified or estimated from the data using an empirical Bayes method (see Morris and Carroll 2006 for details).

Posterior samples for each fixed-effect function, $\{B_i^{(g)}, g = 1, \dots, G\}$, on the grid \mathbf{t} are obtained by applying the IDWT to the posterior samples of the corresponding complete set of wavelet coefficients $B_i^{*(g)} = [B_{i11}^{*(g)}, \dots, B_{iJK_j}^{*(g)}]$, and similarly for the random-effect functions U_i . If desired, posterior samples for the covariance matrices Q and S may also be computed using matrix multiplication $Q^{(g)} = WQ^{*(g)}W'$ and $S^{(g)} = WS^{*(g)}W'$. Because $Q^{*(g)}$ and $S^{*(g)}$ are diagonal, this may be accomplished in an equivalent but more efficient manner by applying the two-dimensional version of the IDWT (2d-IDWT) to $Q^{*(g)}$ and $S^{*(g)}$ (Vannucci and Corradi 1999). These posterior samples of the quantities in model (3) may subsequently be used to perform any desired Bayesian inference.

4. HANDLING INCOMPLETE PROFILES IN WAVELET-BASED FUNCTIONAL MIXED MODELS

The foregoing method cannot be used to analyze data for which some profiles are incomplete (i.e., not fully observed on the grid \mathbf{t}), because the first step of the method, the DWT, cannot be applied to incomplete profiles. Under these circumstances, one alternative is to simply perform a “complete-case” analysis that uses only the complete profiles; however, this is inefficient, because it ignores information contained in the incomplete profiles. This problem is especially severe when a high proportion of observed profiles are incomplete, which is true for the TriTrac accelerometer data from the Planet Health study.

In this section we introduce two missing-data approaches that can be used to incorporate information from the incomplete profiles in the WFMM when at least some complete profiles are available. Both methods involve imputation of “missing-wavelet coefficients” from predictive distributions at each iteration of the MCMC. Missing-wavelet coefficients are wavelet coefficients for an individual profile for which the support of the corresponding basis function intersect missing regions of the profile. In the first method, the imputation distributions are defined based on the posterior means of model parameters obtained from an MCMC applied to the complete-case data, whereas in the second method, posterior predictive means and variances are used. These approaches are appropriate when the data are missing completely at random (MCAR) (Little and Rubin 2002). Craddock et al. (2004) described a validation study in which they essentially concluded that the missingness in the accelerometer data from the Planet Health study was MCAR.

Suppose that we have an incomplete profile, $Y_i(t)$, observed on the set of times \mathbf{t}^O but missing on the set of times \mathbf{t}^M , with $\mathbf{t}^O \cup \mathbf{t}^M = \mathbf{t}$, an equally spaced grid on the region \mathcal{T} . From (3), the model for \mathbf{Y}_i , a row vector containing the full profile on the grid \mathbf{t} , is given by $\mathbf{Y}_i = X_i B + Z_i U + \mathbf{E}_i$, where X_i is $1 \times p$, B is $p \times T$, Z_i is $1 \times m$, U is $m \times T$, and \mathbf{E}_i is $1 \times T$. The rows of U are independently distributed as $\text{normal}\{0, Q\}$, and $\mathbf{E}_i \sim \text{normal}\{0, S\}$. Let Ω be the vector containing the covariance parameters determining the matrices Q and S . The joint distribution of the profile on the grid, conditional on the fixed effects, covariance parameters, and random effects, is $\{\mathbf{Y}_i | B, \Omega, U\} \sim \text{normal}(\mu_i, \Sigma_i)$, with

$$\mu_i = X_i B + Z_i U \quad (5)$$

and

$$\Sigma_i = S. \quad (6)$$

Partition $\mathbf{Y}_i = (\mathbf{Y}_i^O, \mathbf{Y}_i^M)$, with $\mathbf{Y}_i^O = \{Y_i(t) : t \in \mathbf{t}^O\}$ and $\mathbf{Y}_i^M = \{Y_i(t) : t \in \mathbf{t}^M\}$ being vectors of length T_i^O and T_i^M containing the observed and missing regions of the profile. Similarly, partition

$$\boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^O, \boldsymbol{\mu}_i^M) \quad \text{and} \quad \Sigma_i = \begin{bmatrix} \Sigma_i^{O,O} & \Sigma_i^{O,M} \\ \Sigma_i^{M,O} & \Sigma_i^{M,M} \end{bmatrix}.$$

By standard conditional normal calculations, we find that

$$\{\mathbf{Y}_i^M | \mathbf{Y}_i^O, B, \Omega, U\} \sim \text{normal}(\boldsymbol{\mu}_i^{M|O}, \Sigma_i^{M|O}), \quad (7)$$

with $\boldsymbol{\mu}_i^{M|O} = \boldsymbol{\mu}_i^M + \Sigma_i^{M,O}(\Sigma_i^{O,O})^{-1}(\mathbf{Y}_i^O - \boldsymbol{\mu}_i^O)$ and $\Sigma_i^{M|O} = \Sigma_i^{M,M} - \Sigma_i^{M,O}(\Sigma_i^{O,O})^{-1}\Sigma_i^{O,M}$. We call this conditional distribution the *imputation distribution*. This distribution represents a linear regression of the missing data on the observed parts of the profile.

In the Bayesian paradigm, a natural way to deal with the missing data \mathbf{Y}_i^M is to treat it as an unknown parameter vector to be updated at each iteration of the MCMC from its complete conditional distribution, which is the imputation distribution (7). Although conceptually straightforward, this fully Bayesian approach is computationally infeasible here. It requires numerous applications of the DWT and IDWT and extra matrix inversions and multiplications at each iteration of the MCMC, because the missingness is in the data space and the MCMC is applied to the wavelet space version of the model.

A less computationally intensive alternative is to first perform a “complete-case analysis” to obtain posterior samples of the model parameters, then use these to characterize the imputation distribution (7). This distribution may then be projected into the wavelet space to yield a distribution from which random draws of the “missing-wavelet coefficients” can be taken at each iteration of the MCMC. This approach does not require any DWTs, IDWTs, or extra matrix inversions or multiplications inside the MCMC.

The following specific steps are necessary to implement this approach, using the posterior mean estimates from the complete case run to characterize the mean and variance in (7):

1. Fit the wavelet-based functional mixed model to the set of complete profiles Y^C . Compute \widehat{B} , \widehat{U} , and \widehat{S} , the posterior mean estimates of the fixed-effect functions, random-effect functions, and residual covariance matrix using the procedure described in Section 3.3.

2. For each incomplete profile $Y_i(t)$:

- a. Compute $\widehat{\boldsymbol{\mu}}_i^{M|O}$ and $\widehat{\Sigma}_i^{M|O}$ in (7), using the posterior mean estimates of the model quantities in place of $\boldsymbol{\mu}_i$ and Σ_i .

- b. Form the mean imputed profile $M_i(t)$ on the grid \mathbf{t} , \mathbf{M}_i , by setting $M_i(t) = Y_i(t)$ for $t \in \mathbf{t}^O$ and $M_i(t) = \widehat{\boldsymbol{\mu}}_i^{M|O}$ for $t \in \mathbf{t}^M$.

- c. Form the covariance matrix for the imputed profile $V_i(t_1, t_2)$ on $\mathbf{t} \times \mathbf{t}$, by setting $V_i(t_1, t_2) = \widehat{\Sigma}_i^{M|O}(t_1, t_2)$ if $t_1, t_2 \in \mathbf{t}^M$ and $V_i(t_1, t_2) = 0$ if $t_1 \in \mathbf{t}^O$ or $t_2 \in \mathbf{t}^O$.

- d. Apply the DWT to the mean imputed profile, yielding the corresponding vector of wavelet coefficients, $\mathbf{M}_i^* = \mathbf{M}_i W'$, double-indexed by wavelet scale j and location k .

- e. Apply the two-dimensional DWT to the covariance matrix for the imputed profile to compute the corresponding covariance matrix of the imputed wavelet coefficients, $V_i^* = W V_i W'$. Because of our independence assumption between the wavelet coefficients within a given function, we restrict attention to the diagonal elements of this matrix, $\mathbf{v}_i^* = \text{diag}(V_i^*)$. Each element of the vector \mathbf{v}_i^* , v_{ijk}^* , contains the imputation variance for a single wavelet coefficient, double-indexed by scale j and location k .

3. Fit the WFMM to the full dataset, including the complete and incomplete profiles. At each iteration of the MCMC, add a step whereby for each incomplete profile indexed by i , we draw each wavelet coefficient d_{ijk} ($j = 1, \dots, J$, $k = 1, \dots, K_j$) from its wavelet-space imputation distribution, which is normal(M_{ijk}^* , v_{ijk}^*). For any “observed” wavelet coefficients for which the corresponding wavelet basis function, $\psi_{jk}(t)$, is completely contained within \mathbf{t}^O , the imputation variance v_{ijk}^* will be 0 and the imputation mean M_{ijk}^* will be d_{ijk} , the wavelet coefficient that we would have obtained by applying the DWT to the profile \mathbf{Y}_i if it were completely observed.

This procedure assumes that for each incomplete profile \mathbf{Y}_i , we have at least one complete profile from that child, because it conditions on estimates of the corresponding child-level random-effect function in U from the complete-case analysis. The procedure is easily adapted for profiles for which this is not true by substituting $\boldsymbol{\mu}_i = X_i B$ and $\Sigma_i = Q + S$ for (5) and (6). Recall that in our example Q represents the child-to-child covariance and S represents the day-to-day covariance. Including of the child-to-child variability tends to inflate the imputation variances.

This approach has several desirable qualities. It is easy to implement and makes full use of the functional data, borrowing strength both within and between profiles while performing the imputation. The borrowing of strength from within the profile is accomplished through the conditional distribution in (7), and the borrowing of strength between profiles occurs from the use of parameter estimates from the complete-case analysis to characterize the distributions in (5) and (6). Because multiple draws are taken from the imputation distribution, the estimated variability of the imputation is propagated throughout any subsequent inference.

But one weakness of this approach is that by conditioning on the parameter estimates from the complete profile model fit, it fails to propagate the uncertainty of the parameter estimation into the imputation variances. This problem can be overcome by instead using posterior predictive distributions, which integrate over this uncertainty; that is, replace the conditional mean and covariance in (5) and (6) in the imputation distribution in (7) by the posterior predictive mean and covariance, $\mu_i^{\text{pred}} = E(\mathbf{Y}_i | Y^C)$ and $\Sigma_i^{\text{pred}} = \text{var}(\mathbf{Y}_i | Y^C)$. Given a set of G MCMC samples from the complete profile run, these may be estimated by

$$\widehat{\mu}_i^{\text{pred}} = G^{-1} \sum_{g=1}^G \mu_i^{(g)} = \widehat{\mu}_i \quad (8)$$

and

$$\widehat{\Sigma}_i^{\text{pred}} = \widehat{S} + (G - T)^{-1} \sum_{g=1}^G \{\mu_i^{(g)} - \widehat{\mu}_i^{\text{pred}}\} \{\mu_i^{(g)} - \widehat{\mu}_i^{\text{pred}}\}', \quad (9)$$

where $\mu_i^{(g)} = X_i B^{(g)} + Z_i U^{(g)}$; $B^{(g)}$ and $U^{(g)}$ are the posterior samples of B and U from iteration g of the complete-case MCMC, and \hat{S} the posterior mean estimator of the within-function residual covariance matrix S . Details of the derivations of (8) and (9) are given in the Appendix. These distributions appropriately propagate the uncertainty in parameter estimation throughout the imputation distributions. They are not fully efficient, however, because they integrate over the posterior distribution of the parameters conditional on only the complete-case data, whereas the fully Bayesian approach would integrate over the posterior conditional on all available data.

Note that because of the mixture prior placed on B^* , the predictive distribution $f(\mathbf{Y}_i | Y^C)$ is not multivariate normal. Thus, by representing this distribution by its first two moments and using the normal distribution to perform our imputation, we are using an approximation. This keeps the procedure computationally feasible, because sampling from the actual posterior predictive distribution would require multiple DWT and IDWT calculations within the MCMC, as was the case for the fully Bayesian approach. Our use of the posterior predictive means and variances improves on the conditional approach because it propagates the uncertainty due to parameter estimation in the correct manner, as indicated by the second term in (9).

5. ANALYSIS OF TRITRAC DATA FROM PLANET HEALTH

5.1 Complete-Case Analysis

We first modeled the set of complete profiles to obtain posterior predictive distributions to characterize the imputation distributions for the missing parts of the incomplete profiles. There were a total of 95 complete daily profiles (i.e., no missing observations between 9 AM and 8 PM) from 61 children. Let Y^C be the 95×660 response matrix, each row of which contained the log accelerometer profile (in log METs) for one day from one child, sampled on the time grid $\mathbf{t} = \{t_m, m = 1, \dots, 660\}$, consisting of every minute from 9 AM to 8 PM. We log-transformed the data to stabilize the variance and make the data more symmetric to accommodate the Gaussian error assumptions underlying the functional mixed model.

We modeled these data using the functional mixed model (3), $Y^C = XB + ZU + E$. The matrix X was a 95×13 design matrix containing child-level and day-level covariates that correspond to the 13×660 matrix B , with rows that were the fixed-effect functions on the grid \mathbf{t} . Entry B_{ij} described the effect of covariate i on the activity levels at time t_j . The matrix Z was a 95×61 design matrix with corresponding 61×660 matrix U , for which each row contained the random-effect function for a single child on the grid \mathbf{t} . The 95×660 matrix E contained the residual errors on the same grid. We assumed that the rows of U and E were mutually independent mean-0 Gaussians with 660×660 covariance matrices Q and S .

The X matrix that we used included columns corresponding to an overall functional “intercept,” a sex effect (1 if boy, -1 if girl), an effect for higher tricep skinfolds measurements (1 if >20 , -1 if ≤ 20), a BMI effect, an effect for the average number of hours per day watching television, day-of-week effects, a seasonal effect (1 if after the beginning of daylight savings time [DST], 0 if before DST), and school effects. All effects were functional, meaning that they were allowed to vary

as a function of time of day. Both the BMI and TV hours covariates were treated as continuous, with the BMI covariate mean-centered and the TV hours covariate standardized. School effects were included only for schools B, D, and E, because there were only 1 complete profile from school A and 4 complete profiles from school C, out of the total of 11 and 15 that were at least 50% complete from these 2 schools. We determined this was not sufficient replication for estimating the corresponding school fixed effects.

Written out in scalar form, the model for profile i at time t_j , $Y_i^C(t_j)$, was $Y_i^C(t_j) = B_0(t_j) + \sum_{k=1}^{12} X_{ik} B_k(t_j) + \sum_{k=1}^{61} Z_{ik} \times U_k(t_j) + E_i(t_j)$, where $U_k(t_j)$ were mean-0 Gaussians with $\text{cov}\{U_k(t_j), U_{k'}(t_{j'})\} = Q_{jj'}$ if $k = k'$, 0 otherwise; $E_i(t_j)$ were mean-0 Gaussians with $\text{cov}\{E_i(t_j), E_{i'}(t_{j'})\} = S_{jj'}$ if $i = i'$, 0 otherwise; and $\text{cov}\{U_k(t_j), E_i(t_{j'})\} = 0$ for all i, j, j', k .

We fit the WFMM described in Section 3.3 and obtained posterior samples for all parameters in this model. We chose a Daubechies wavelet with four vanishing moments (Daubechies 1992) and performed the wavelet decomposition to $J = 8$ levels. We used maximum likelihood estimates as starting values for the MCMC for all model parameters. We estimated the regularization parameters using the empirical Bayes approach described by Morris and Carroll (2006), with $\pi_{ij} \approx 1$ and τ_{ij} chosen very large for the smooth coefficients and the two coarsest levels of wavelet coefficients to make the degree of shrinkage at these levels negligible, and used vague proper priors for the variance components. After a burn-in of 1,000, we obtained 20,000 MCMC samples of the model parameters, and kept every tenth one. This took roughly 7 hours to run on our Xeon 3.2-GHz Windows 2000 PC with 2 GB of RAM, and the posterior samples occupied roughly 300 MB. Trace plots and Metropolis–Hastings acceptance probabilities indicated that the MCMC had good convergence properties.

5.2 Imputation Distributions for Incomplete Profiles

We used these posterior samples to estimate the posterior predictive means and variances for each incomplete profile, $\hat{\mu}_i^{\text{pred}}$ and $\hat{\Sigma}_i^{\text{pred}}$, in (8) and (9). As described in Section 4, we used these estimates as the basis for the imputation distributions for the missing data.

Figure 3 contains six incomplete profiles, with their missing regions replaced by the mean and 95% pointwise bounds of their imputation distributions. Note how the imputation distributions borrow strength from nearby observations and tend to “connect” the imputed data with nearby observed data. They also allow the magnitude of the nearby observed data to play a role in the imputation. Moreover, the imputation variances tend to be smaller for the values immediately adjacent to observed values and in regions of the curves with less variability in the population.

Using the DWT, we projected these imputation distributions into the wavelet space. At each iteration of the MCMC applied to the full dataset, we took random draws from this distribution to fill in the “missing-wavelet coefficients.”

5.3 Model for Analysis of the Full Dataset

We used the same underlying model for the full data analysis as for the complete-case analysis, but removed the functional intercept term and allowed each school to have its own

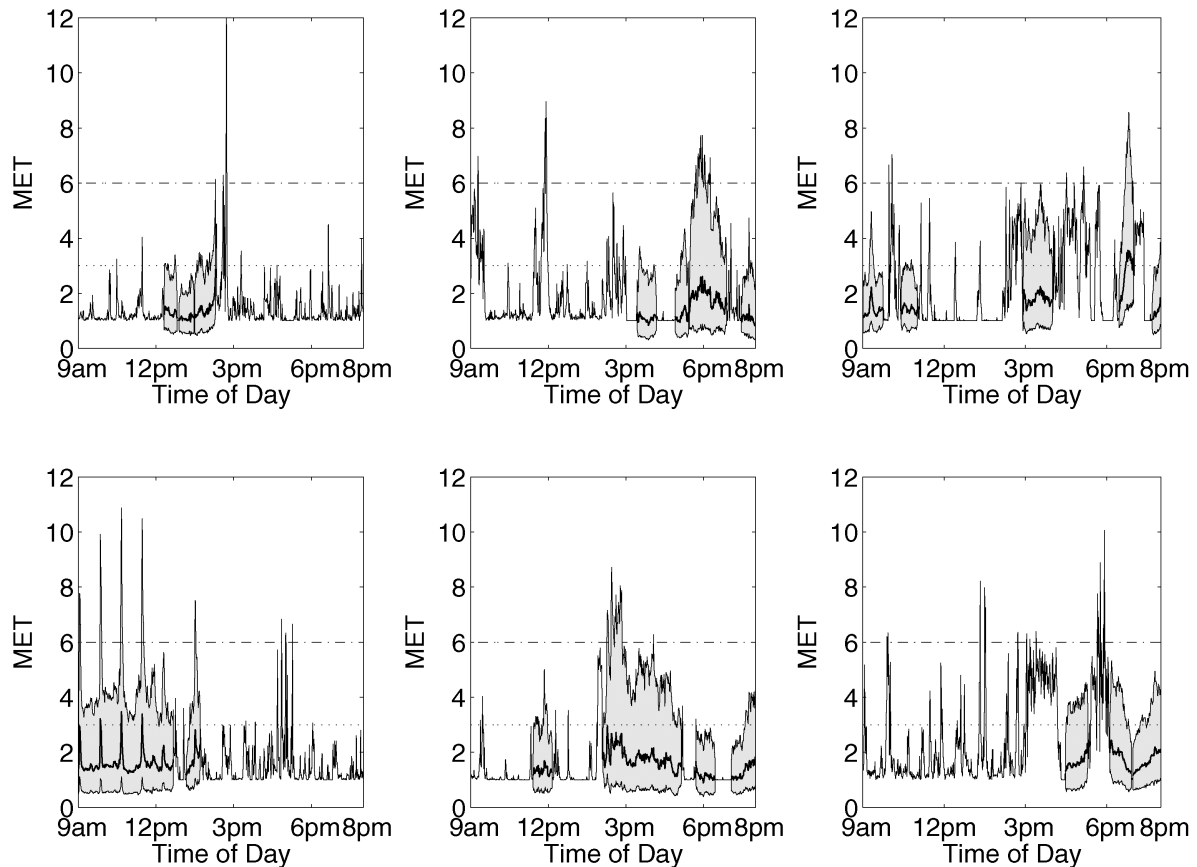


Figure 3. Imputation Distributions. Some incomplete profiles, with mean and 95% pointwise bounds from the joint imputation distribution substituted for the missing regions, as indicated by the shaded regions. These are based on the normal approximation to the posterior predictive distribution, estimated from posterior samples obtained from a MCMC scheme applied to the complete-case data.

fixed-effect function, because the full dataset contained enough profiles to reliably estimate them all. We obtained posterior samples for all parameters in this model using MCMC, with the missing-wavelet coefficients sampled from their imputation distributions at each iteration, as described in Section 4. We used the same method settings as used in the complete-case analysis. We ran 4 parallel chains, each with a burn-in of 1,000 and consisting of 2,500 MCMC samples, and kept every 5. This took roughly 40 hours to run, and the posterior samples occupied roughly 300 MB. For 87% of the variance components, the Metropolis–Hastings acceptance probabilities were between .25 and .50. Trace plots indicated good convergence properties.

5.4 Results

5.4.1 Analysis of Fixed-Effect Functions. To consider the marginal time-of-day effect on activity levels, we computed the overall mean curve by equally weighting each of the schools' fixed-effect functions and conditioning on the mean values of all other covariates. Figure 4(a) contains the posterior mean and 90% posterior pointwise credible bounds for the overall mean curve in the log METs scale. Throughout the school day, the curve is characterized by numerous spikes indicating periods of coordinated activity, such as class switches and lunch periods for the different schools. There is a large spike between 2:15 PM and 3:00 PM corresponding to the end of the school day. The mean activity profiles are less spiky after school, in-

dicating less coordinated activity across days and children. The mean activity levels remained relatively constant from 3:00 PM to 6:30 PM, then decreased later in the evening.

For each fixed-effect function, we plotted the posterior mean curve and computed pointwise 90% posterior bounds. Two of these are plotted in Figures 4(b) and 4(c), and other curves are available by request from the first author. Recall that the effects are additive on the log scale, so they are multiplicative on the MET scale. Exponentiating the values in these plots yields the multiplicative effects, which subsequently can be converted into percent increases or decreases in the MET scale.

To further summarize the results, we computed the average percent increases/decreases corresponding to each fixed-effect function, aggregating within the following time intervals, chosen based on the structure evident in the overall mean curve: (1) all day, 9 AM–8 PM; (2) morning, 9 AM–11:30 AM; (3) lunch, 11:30 AM–12:30 PM; (4) afternoon, 12:30 PM–2:15 PM; (5) going home, 2:15 PM–3:00 PM; (6) after school, 3:00 PM–5:30 PM; (7) early evening, 5:30 PM–7:00 PM; and (8) late evening, 7:00 PM–8:00 PM. The 90% credible intervals for these quantities are reported in square brackets. The fixed-effect curves for the different schools were characterized by numerous peaks present throughout the school day. These were most evident for school E; see Figure 4(b). This school provided the most daily profiles in this dataset (146 out of 292) and appeared to have the most regular schedule of the schools in the study; see Figure 2. The regular pattern of peaks reveals

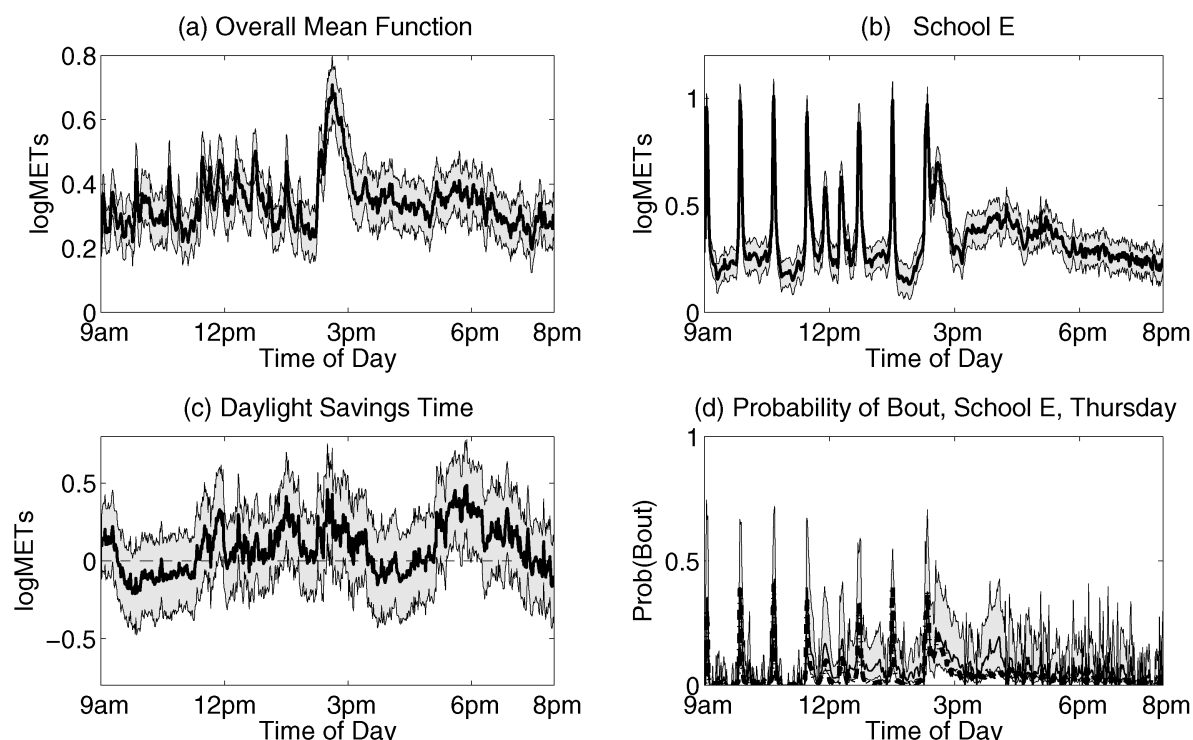


Figure 4. Results. Posterior mean curves and 90% pointwise posterior credible intervals for the fixed-effects curves for (a) overall mean curve, (b) school E, and (c) daylight savings time, plus (d) posterior predictive probabilities of bouts for a child from school E on Thursdays. The curve in (a) is obtained by averaging equally over the schools and conditioning on the mean value in the dataset for all other covariates. The curve in (b) can be interpreted as the mean log profile on a Monday before daylight savings time for a child from school E with average BMI who is average with respect to the average number of hours of TV watched per day. The curve in (c) is added to the mean log profile for any day after daylight savings time (April 6). The solid line in curve (d) indicates the wavelet-denoised empirical probabilities of bouts for each time point averaging over the 42 accelerometer profiles obtained on Thursdays from children enrolled at school E (—), along with 95% pointwise confidence bands computed using the method described by Bruce and Gao (1996). The dashed-dotted line indicates the posterior predictive probability of bouts from the functional mixed model for a child enrolled at school E on Thursday, for whom the levels of all other covariates are at their mean levels.

the school's class schedule. The large peaks in the morning and afternoon are exactly 48 minutes apart and indicate class switches. The three peaks around noon are 24 minutes apart and likely indicate different lunch periods. From the large extended spike in the afternoon, we surmise that school let out around 2:15 PM. Although regular peaks were evident in the school's main-effect curve, they were not obvious in any individual curve because of the high levels of day-to-day and child-to-child variability. This is shown in the bottom two panels of Figure 1, which present profiles from a child at this school.

We found that children tended to be less active on Tuesdays (3.3% decrease [−6.2%, −.2%]), Thursdays (2.9% decrease [−5.5%, −.2%]), and Fridays (3.6% decrease [−6.8%, −.4%]), compared with Mondays, aggregating over the entire day. These effects were especially strong during the after-school period (3:00–5:30; Tuesday, −7.7% [−12.6%, −2.2%]; Thursday, −7.8% [−12.5%, −3.1%]; Friday, −5.8% [−11.4%, −.1%]). A similar but less pronounced trend was present on Wednesdays (overall, 1.9% decrease [−5.0%, 1.4%]; after school, 5.1% decrease [−10.5%, .9%]).

In 1997 daylight savings time (DST) in Boston started on April 6. We found that activity levels were generally higher after DST than before DST (overall, 8.2% increase [−.5%, 18.1%]); see Figure 4(c). This effect was especially strong as school was letting out (2:15–3:00; 25% increase

[4.2%, 53.4%]) and in the early evening (5:30–7:00; 29.5% increase, [5.6%, 59.2%]), a time period during which the sun was still out after DST, but had already set before DST.

Boys tended to be slightly more active than girls on average, most notably over lunch (4.8% increase [.5%, 9.1%]).

5.4.2 Variance Component Analysis. We summarized the relative day-to-day variability by $\rho = \text{Tr}(S) / \{\text{Tr}(Q) + \text{Tr}(S)\}$, where $\text{Tr}(Q)$ and $\text{Tr}(S)$ are the traces of the child-to-child and day-to-day covariance matrices Q and S . The posterior mean of ρ was that 90.9%, indicating that 90.9% of the variability in the log accelerometer profiles was day-to-day and 9.1% of it was child-to-child. The 90% posterior credible interval for ρ was [88.8%, 92.8%].

To evaluate how these sources of variability varied over t , we considered the child-to-child and day-to-day variances as functions of t , $V_q(t) = Q(t, t)$ and $V_s(t) = S(t, t)$ (plots available in supplementary material at http://biostatistics.mdanderson.org/Morris/papers_files/TriTrac_supplement.html). We found that child-to-child variability was greater during the school day than after school, whereas day-to-day variability was greater after school than during the school day. The relative day-to-day variability as a function of time of day, $\rho(t) = V_s(t) / \{V_q(t) + V_s(t)\}$, revealed that the percent of day-to-day variability increased from 85–90% during the school day up to 95–99% after 3 PM.

5.4.3 Assessing Bouts. The literature on accelerometer data includes much work on studying “bouts,” or short bursts of intense activity. Here we considered bouts to be periods during which the child was engaging in moderate or vigorous activity, which corresponded to activity levels of >3.0 METs (Rowlands et al. 2004). In the literature, frequently performed data analyses include estimating the probabilities of bouts for different groups of individuals. In our WFMM approach, we did not directly model these probabilities, but the unified Bayesian framework underlying the method made it easy to compute these probabilities as a function of time of day for any group of individuals using the posterior samples output from the MCMC.

Given a child with covariate levels \mathbf{X}^* , the posterior predictive probability that the child experiences one bout at time t in a hypothetical future daily activity profile $\mathbf{Y}^*(t)$ is given by $\phi_{\text{WFMM}}(t) = \Pr\{\mathbf{Y}^*(t) > 3.0 | Y, \mathbf{X}^*\} = \int I\{\mathbf{Y}^*(t) > 3.0\} \times f\{\mathbf{Y}^*(t) | \theta, \mathbf{X}^*\} f(\theta | Y) d\theta$, where $\theta = \{B, Q, S\}$ are the fixed-effect and covariance parameters in the model and Y are the profiles used to fit the model. This expression can be estimated from the G posterior samples $\theta^{(g)} = \{B^{(g)}, Q^{(g)}, S^{(g)}\}$, $g = 1, \dots, G$, obtained from the MCMC, by $\hat{\phi}_{\text{WFMM}}(t) = G^{-1} \sum_{g=1}^G \Phi\{[3.0 - \mu^{(g)}(t)]/\sigma^{(g)}(t)\}$, where $\mu^{(g)}(t) = \mathbf{X}^* \times B^{(g)}(t)$, $\sigma^{(g)}(t) = \sqrt{Q^{(g)}(t, t) + S^{(g)}(t, t)}$, and $\Phi(x)$ is the standard normal cumulative distribution function evaluated at x .

For example, we computed the probability of bouts as a function of the time of the day for Thursdays before DST for children from school E. This curve is given by the dashed line in Figure 4(d). To check how well our model captured these bouts, we also computed an empirical estimate of the probability of bouts for each time point averaging over the $n = 42$ accelerometer profiles obtained on Thursdays before DST from children enrolled at school E. The raw empirical estimator was $\hat{\phi}_{\text{EMP}}(t) = n^{-1} \sum_{i=1}^n I(Y_i(t) > 3.0)$. The solid line in Figure 4(d) is a wavelet-denoised version of this curve, denoised using *Sureshrink* (Donoho and Johnstone 1995) with soft thresholding. Although we did not directly model the probability of bouts, our WFMM appeared to do a reasonable job capturing these features of the data.

6. DISCUSSION

We first discuss the specific results of our analysis, and then assess the fitness of the WFMM for analyzing accelerometer data. We were not surprised by the large DST effect that we observed in the data. Considering the weather in Boston, children naturally tended to be more active after April 6 than before April 6. In 1997, the sun set in Boston at times ranging from 5:10 PM to 6:15 PM from February 10 through April 5, then from 7:16 PM through 8:11 PM from April 6 through May 28. This may explain the very large spike in the DST effect between 5 PM and 7 PM.

In our relative variability analysis, we found that after adjusting for the covariates, the day-to-day variability dominated the child-to-child variability, especially during the after-school hours. In other words, the difference between the average log profiles of a very active child and a very inactive child with the same covariate levels is small compared with the day-to-day variability in activity for a given child, especially after school. This suggests that it is important to get many days per child

to accurately quantify each child's typical level of activity. Although it is already typical in these studies to monitor each child for 4–8 days, it may be a better design to obtain even more days per child, especially if this does not result in much increase in cost for the study.

Although it does not directly model the bouts, the WFMM yields estimates of the probability of bouts at different times of the day for children with different levels of the covariates. It appears that this model does a reasonable job of capturing these bouts, indicating that it is at least somewhat effective in modeling some of the tail behavior in the distributions. However, there is some evidence of attenuation in our model-based estimates of these probabilities, especially during the after-school period, in which there is less coordinated activity and more day-to-day variability. These may be related to our underlying assumptions of normality and equal covariances across different groups of children. It is possible in our framework to allow different covariance parameters across different groups of children and to relax the normality assumptions using scale mixtures of normals. These adaptations would accommodate heavier tails and may do an even better job capturing these bouts.

The WFMM is a powerful tool for analyzing accelerometer data. It allows one to consider the joint functional effects of multiple covariates and to model correlations between profiles obtained from the same individual using random-effect functions. The functions are allowed to be of arbitrary form, and the within-function covariances are allowed to be nonstationary. The fixed-effect functions are adaptively regularized using the principles of wavelet shrinkage. Less adaptive methods using kernels or splines would result in more attenuation of dominant local features in the fixed-effects curves, possibly resulting in reduced power for inference.

It is also possible to introduce other random-effect functions to account for covariance due to other clustering factors, such as school or neighborhood. In the present analysis we chose to model the schools using fixed effects because there were so few of them, but given more schools, we could use random effects. It would be a good idea to use more schools in these studies, because the school-to-school variability appears to be large, and it would be interesting and important to study a wider range of schools with different socioeconomic makeups.

Our model is linear in that we assume that the effects are linear in the covariates. It would be interesting to consider generalizing or testing this assumption. But this framework is still very flexible because the linear coefficients are allowed to be time-varying, and their functional form over time is left arbitrary. A much less flexible model, for example, would be to allow the overall mean function to be arbitrary but to force the fixed effects to be constant over time. It would be interesting to develop formal testing procedures for testing whether it is necessary to allow the coefficient to be time-varying for a given covariate; this is a topic for future investigation.

In addition, we note that this model is very flexible in its representation of the covariances of the profiles from day to day, S , and the covariances of the random-effect profiles from child to child, Q . Our assumption of independence in the wavelet space reduces the dimensionality of these covariances from $T(T+1)/2$ to T , yet accommodates a reasonably wide range of nonstationary covariance structures. Different variance components are estimated for each wavelet coefficient, which allows

the day-to-day and child-to-child variability to differ across both scale and location. This accommodates various types of nonstationarity, allowing the day-to-day and child-to-child variances across profiles to vary over time, and also allowing the level of smoothness in the random-effect functions and residual error processes to vary over time.

The missing-data methods introduced in this article allow the WFMM to be applied to data with incomplete profiles, which are commonly encountered in accelerometer data. In our case study, these methods allowed us to increase the number of profiles in the analysis by a factor of three. The resulting gain in precision is evident if one compares the posterior bounds from the full-data and complete-case analyses (not shown).

The WFMM method has some weaknesses and limitations. It is based on linear models and so assumes additive errors (in our case study, additive on the log scale). Other types of variability may not be adequately picked up by these models. The normality assumption is somewhat restrictive, although, as mentioned previously, this could be relaxed using scale mixtures. As is true with nonfunctional mixed models, some models can be unstable to fit, especially when there is near collinearity in the design matrices or small effective sample sizes for some of the variance components. These problems are exacerbated in the functional context, because we are effectively fitting T simultaneous scalar mixed models. It would also be useful to have a global functional test for more formally assessing the significance of the fixed-effect functions; this is an area that we will investigate in the near future. Our method is computationally and memory intensive; the analyses performed for this article took a total of about 50 computer hours. But this is not unreasonable considering the time that it takes to collect the data. The method is also highly parallelizable, so it could be done in much less time whenever parallel computing resources are available.

Although computationally intensive, the method is straightforward to implement. The minimal information one must specify to use our scripts are the Y , X , and Z matrices; the wavelet basis; the number of levels of decomposition; the number of MCMC samples; and the burn-in. By default, maximum likelihood starting values, empirical Bayes regularization parameters, Fisher's information-based proposal variances for the Metropolis–Hastings, and vague priors on the variance components are automatically computed.

Although straightforward to implement, this method is considerably complex, combining wavelets with mixed models and wrapping it all up inside a MCMC. This begs the question of what all this complexity buys, of what this method can do that simpler approaches cannot. This functional data modeling approach can be used to perform the same types of standard analyses found in the existing literature, including analyses of average daily activity levels, 30-minute averages, and probabilities of bouts for different groups of individuals, but these can be done more efficiently, because we can effectively incorporate incomplete profiles into the analysis. More important, the functional approach opens new possibilities in terms of analyses that can be done and types of information that can be extracted from these rich data, for example, allowing us to perform inference on time-varying effects. Given the posterior samples from our model fit, we can perform nearly any type of inference that

we desire, functional or pointwise, on fixed-effect functions, random-effect functions, or covariance functions. The Bayesian approach propagates the uncertainty from the different sources of variability in the model and the different parameters estimated throughout any subsequent inference. The WFMM, supplemented with the missing-data methods introduced in this article, comprises a promising tool for extracting information from accelerometer data in activity-level studies.

APPENDIX: DERIVATION OF POSTERIOR PREDICTIVE MEANS AND COVARIANCES

Suppose that Y^C is an $N^C \times T$ matrix whose rows contain N^C complete profiles, observed on an equally spaced grid \mathbf{t} of length T . Assume that this matrix is modeled using the functional mixed model (3) with parameters $\Theta = \{B, U, Q, S\}$. Let \mathbf{Y}_i be another profile following the same model that is independent of each profile in Y^C conditional on Θ . Here we show that, given a set of G posterior samples of $(\Theta|Y^C)$, we can estimate the posterior predictive mean, $\mu_i^{\text{pred}} = E(\mathbf{Y}_i|Y^C)$, and covariance matrix, $\Sigma_i^{\text{pred}} = \text{var}(\mathbf{Y}_i|Y^C)$, by the expressions in (8) and (9).

We start by showing (8). It is easy to show, using alternating conditional expectation arguments, that if X and Y are independent, conditional on Z , then $E(X|Y) = E_{Z|Y}\{E(X|Z)\}$. Because \mathbf{Y}_i and Y^C are independent conditional on Θ , this means that $\mu_i^{\text{pred}} = E_{\Theta|Y^C}\{E(\mathbf{Y}_i|\Theta)\} = E_{\Theta|Y^C}\{X_i B + Z_i U\}$. Estimating this expectation using the MCMC samples of Θ , we get $\hat{\mu}_i^{\text{pred}} = \hat{\mu}_i$.

To show (9), we first note that alternating conditional expectation arguments can be used to show that if X and Y are independent, conditional on Z , then $\text{var}(X|Y) = E_{Z|Y}\{\text{var}(X|Z)\} + \text{var}_{Z|Y}\{E(X|Z)\}$. This means that we can write $\Sigma_i^{\text{pred}} = \text{var}(\mathbf{Y}_i|Y^C) = E_{\Theta|Y^C}\{\text{var}(\mathbf{Y}_i|\Theta)\} + \text{var}_{\Theta|Y^C}\{E(\mathbf{Y}_i|\Theta)\}$.

Because $\text{var}(\mathbf{Y}_i|\Theta) = S$, the first term can be estimated by \hat{S} , the posterior mean estimator of S computed from the MCMC samples, which is the first term in (9). Next we can write $\text{var}_{\Theta|Y^C}\{E(\mathbf{Y}_i|\Theta)\} = E_{\Theta|Y^C}\{[E(\mathbf{Y}_i|\Theta) - E_{\Theta|Y^C}\{E(\mathbf{Y}_i|\Theta)\}]' [E(\mathbf{Y}_i|\Theta) - E_{\Theta|Y^C}\{E(\mathbf{Y}_i|\Theta)\}]\} = E_{\Theta|Y^C}\{[X_i B + Z_i U - \mu_i^{\text{pred}}]' [X_i B + Z_i U - \mu_i^{\text{pred}}]\}$. This expression can be estimated using the MCMC samples by $(G - T)^{-1} \times \sum_{g=1}^G \{\mu_i^{(g)} - \hat{\mu}_i^{\text{pred}}\}' \{\mu_i^{(g)} - \hat{\mu}_i^{\text{pred}}\}$, which is the second part of (9).

[Received January 2005. Revised October 2005.]

REFERENCES

- Abbott, A. P., and Davies, P. S. (2004), "Habitual Physical Activity and Physical Activity Intensity: Their Relation to Body Composition in 5.0–10.5-Year-Old Children," *European Journal of Clinical Nutrition*, 58, 285–291.
- Abramovich, F., Sapatinas, T., and Silverman, B. W. (1998), "Wavelet Thresholding via a Bayesian Approach," *Journal of the Royal Statistical Society, Ser. B*, 60, 725–749.
- Alga, D. L., Beattie, E., Leitsch, S. A., and Beel-Bates, C. A. (2003), "Biomechanical Activity Devices to Index Wandering Behavior in Dementia," *American Journal of Alzheimer's Disease and Other Dementias*, 18, 85–92.
- Brown, P. J., Fearn, T., and Vannucci, M. (2001), "Bayesian Wavelet Regression on Curves With Application to a Spectroscopic Calibration Problem," *Journal of the American Statistical Association*, 96, 398–408.
- Bruce, A. G., and Gao, H. Y. (1996), "Understanding WaveShrink: Variance and Bias Estimation," *Biometrika*, 83, 727–745.
- Centers for Disease Control (2000), "Promoting Better Health for Young People Through Physical Activity and Sports: A Report to the President From the Secretary of Health and Human Services and the Secretary of Education," technical report, Centers for Disease Control and Prevention.
- (2001), "Physical Activity and Good Nutrition, Essential Elements to Prevent Chronic Disease and Obesity, at a Glance 2001," technical report, Centers for Disease Control and Prevention.
- Chang, W., and Vidakovic, B. (2002), "Wavelet Estimation of a Base-Line Signal From Repeated Noisy Measurements by Vertical Block Shrinkage," *Computational Statistics and Data Analysis*, 40, 317–328.

- Chipman, H. A., Kolaczyk, E. D., and McCulloch, R. E. (1997), "Adaptive Bayesian Wavelet Shrinkage," *Journal of the American Statistical Association*, 92, 1413–1421.
- Clyde, M., and George, E. I. (2000), "Flexible Empirical Bayes Estimation for Wavelets," *Journal of the Royal Statistical Society, Ser. B*, 60, 681–698.
- Clyde, M., Parmigiani, G., and Vidakovic, B. (1998), "Multiple Shrinkage and Subset Selection in Wavelets," *Biometrika*, 85, 391–401.
- Coleman, K. J., Saelens, B. E., Wiedrich-Smith, M. D., Finn, J. D., and Epstein, L. H. (1997), "Relationships Between TriTrac-R3D Vectors, Heart Rate, and Self-Report in Obese Children," *Medicine and Science in Sports and Exercise*, 29, 1535–1542.
- Cradock, A. L., Weicha, J. L., Peterson, K. E., Sobol, A. M., Colditz, G. A., and Gortmaker, S. L. (2004), "Youth Recall and TriTrac Accelerometer Estimates of Physical Activity Levels," *Medicine and Science in Sports and Exercise*, 36, 525–532.
- Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: Society for Industrial and Applied Mathematics.
- Dietz, W. H. (1998), "Health Consequences of Obesity in Youth: Childhood Predictors of Adult Disease," *Pediatrics*, 101, 518–558.
- Donoho, D. L., and Johnstone, I. M. (1995), "Adapting to Unknown Smoothness via Wavelet Shrinkage," *Journal of the American Statistical Association*, 90, 1200–1224.
- Epstein, L. H., Paluch, R. A., Coleman, K. J., Vito, D., and Anderson, K. (1996), "Determinants of Physical Activity in Obese Children Assessed by Accelerometer and Self-Report," *Medicine and Science in Sports and Exercise*, 28, 1157–1164.
- Going, S., Thompson, J., Cano, S., Stewart, D., Stone, E., Harnack, L., Hastings, C., Norman, J., and Corbin, C. (2003), "The Effects of the Pathways Obesity Prevention Program on Physical Activity in American Indian Children," *Preventive Medicine*, 37, S62–S69.
- Goldfield, G. S., Kalakanis, L. E., Ernst, M. M., and Epstein, L. H. (2000), "Open-Loop Feedback to Increase Physical Activity in Obese Children," *International Journal of Obesity and Related Metabolic Disorders*, 24, 888–892.
- Gortmaker, S. L. (2002), "The Role of Physical Activity Environment in Obesity Among Children and Adolescents in the Industrialized World," in *Obesity in Childhood and Adolescents*, eds. C. Chen and W. H. Dietz, Philadelphia: Vevey/Lippencott Williams & Wilkins, pp. 177–194.
- Gortmaker, S. L., Peterson, K., Wiecha, J., Sobol, A., Dixit, S., Fox, M., and Laird, N. (1999), "Reducing Obesity via a School-Based Interdisciplinary Intervention Among Youth: Planet Health," *Archives of Pediatric and Adolescent Medicine*, 153, 409–418.
- Johnstone, I. M., and Silverman, B. W. (1997), "Wavelet Threshold Estimators for Data With Correlated Noise," *Journal of the Royal Statistical Society, Ser. B*, 59, 319–351.
- Laird, N., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.
- Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis With Missing Data*, Hoboken, NJ: Wiley.
- Matthews, C. E., and Freedson, P. S. (1995), "Field Trial of a Three-Dimensional Activity Monitor: Comparison With Self-Report," *Medicine and Science in Sports and Exercise*, 27, 1071–1078.
- Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. M. (2000), *Wavelet Toolbox for Use With Matlab: User's Guide*, Natick, MA: Mathworks, Inc.
- Morris, J. S., and Carroll, R. J. (2006), "Wavelet-Based Functional Mixed Models," *Journal of the Royal Statistical Society, Ser. B*, 68, 179–199.
- Morris, J. S., Vannucci, M., Brown, P. J., and Carroll, R. J. (2003), "Wavelet-Based Non-Parametric Modeling of Hierarchical Functions in Colon Carcinogenesis," *Journal of the American Statistical Association*, 98, 573–597.
- Pearce, K. D. (1999), "Race, Ethnicity, and Physical Activity," *Leisure Today JOPERD*, 70, 25–28.
- Percival, D. B., and Walden, A. T. (2000), *Wavelet Methods for Time Series Analysis*, New York: Cambridge.
- Ramsay, J. O., and Silverman, B. W. (1997), *Functional Data Analysis*, New York: Springer-Verlag.
- (2000), *Recommended Dietary Allowances* (10th ed.), Washington, DC: National Academy Press.
- Rowlands, A. V., Eston, R. G., and Ingledew, D. K. (1999), "Relationship Between Activity Levels, Aerobic Fitness, and Body Fat in 8- to 10-Year-Old Children," *Journal of Applied Physiology*, 86, 1428–1435.
- Rowlands, A. V., Thomas, P. W., Eston, R. G., and Topping, R. (2004), "Validation of the RT3 Triaxial Accelerometer for the Assessment of Physical Activity," *Medicine and Science in Sports and Exercise*, 36, 518–524.
- Strauss, R. S., Rodzilshy, D., Burack, G., and Colin, M. (2001), "Psychosocial Correlates of Physical Activity in Healthy Children," *Archives of Pediatrics and Adolescent Medicine*, 155, 897–902.
- Talbot, L. A., Gaines, J. M., Huynh, T. N., and Metter, E. J. (2003), "A Home-Based Pedometer-Driven Walking Program to Increase Physical Activity in Older Adults With Osteoarthritis of the Knee: A Preliminary Study," *Journal of the American Geriatric Society*, 51, 387–392.
- Vannucci, M., and Corradi, F. (1999), "Covariance Structure of Wavelet Coefficients: Theory and Models in a Bayesian Perspective," *Journal of the Royal Statistical Society, Ser. B*, 61, 971–986.
- Vidakovic, B. (1998), "Nonlinear Wavelet Shrinkage With Bayes Rules and Bayes Factors," *Journal of the American Statistical Association*, 93, 173–179.
- (1999), *Statistical Modeling by Wavelets*, New York: Wiley.
- Westerterp, K. R. (1999), "Physical Activity Assessment With Accelerometers," *International Journal of Obesity and Related Metabolic Disorders*, 23 (Suppl 3), S45–S49.
- (2002), "Pattern and Intensity of Physical Activity," *Nature*, 410, 539.