Generalized Multilevel Functional-on-Scalar Regression and Principal Component Analysis

Jeff Goldsmith[1,*], Vadim Zipunnikov[2], and Jennifer Schrack[3]

[1]Department of Biostatistics, Mailman School of Public Health, Columbia University
[*]*jeff.goldsmith@columbia.edu*
[2]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University
[3]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University

February 21, 2014

Abstract

This manuscript considers regression models for generalized, multilevel functional responses: functions are *generalized* in that they follow an exponential family distribution and *multilevel* in that they are clustered within groups or subjects. This data structure is increasingly common across scientific domains and is exemplified by our motivating example, in which binary curves indicating physical activity or inactivity are observed for nearly six hundred subjects over five days. We use a generalized linear model to incorporate scalar covariates into the mean structure, and decompose subject-specific and subject-day-specific deviations using multilevel functional principal components analysis. Thus, functional fixed effects are estimated while accounting for within-function and within-subject correlations, and major directions of variability within and between subjects are identified. Fixed effect coefficient functions and principal component basis functions are estimated using penalized splines; model parameters are estimated in a Bayesian framework using `Stan`, a programming language that implements a Hamiltonian Monte Carlo sampler. Simulations designed to mimic the application indicate good estimation accuracy and inference with reasonable computation times for moderate datasets, in both cross-sectional and multilevel scenarios; code is publicly available. In the application we identify effects of age and BMI on the time-specific change in probability of being active over a twenty-four hour period; in addition, the principal components analysis identifies the patterns of activity that distinguish subjects and days within subjects.

Key Words: Penalized Splines, Generalized Functional Data, Bayesian Inference, Hamiltonian Monte Carlo, Accelerometry.

1 Introduction

1.1 Motivating data

Continuous monitoring of activity using accelerometers and other wearable devices promises to revolutionize the measurement of physical activity by providing objective, unbiased observation in unprecedented minute-by-minute detail over several days or weeks. Accelerometers generally measure activity through electrical signals that are a proxy measure for acceleration (Spierer et al., 2011; Trost et al., 2005; Ward et al., 2005). "Activity counts" are devised by summarizing the voltage signals across a monitoring period known as an epoch (a one-minute epoch is common), and can be dichotomized into "active" and "inactive" epochs to study sedentary behavior. Thus, these devices give rise to generalized multilevel functional observations: *generalized* because both activity counts and the derived binary "active" versus "inactive" outcomes do not follow a Gaussian distribution; *multilevel* because each subject has several days of data; and *functional* in that continuous 24-hour trajectories are considered the basic unit of observation.

Accelerometers have already been deployed to explore many pressing public health contexts. Examples include studies of the decline in physical activity associated with aging and frailty (Schrack et al., 2014), of the complex behavioral relationships between childhood asthma and physical activity (Rundle et al., 2009), and of the real-time surveillance and detection of symptomatic changes in congestive heart failure patients (Howell et al., 2010). Unfortunately, the analysis of accelerometer data typically reduces thousands of data points to a single summary, such as the total activity count over a 24-hour period, and few current methods utilize the richness of densely observed activity data. This immense data reduction leaves important scientific questions unaddressed. How are daily physical activity trajectories related to subject covariates, like age, gender, BMI, or socio-demographic status? To what degree do subjects differ from each other in their patterns of activity and inactivity, and to what degree do multiple days differ within one subject?

Our motivation in this manuscript is to identify covariate effects and characterize residual patterns of activity in accelerometer data collected from elderly subjects enrolled in the Baltimore Longitudinal Study on Aging (Schrack et al., 2014). BLSA is a study of normative human aging with healthy, functionally-independent participants. Once enrolled, participants are followed for life and undergo extensive testing every 1-4 years depending on age. The sub-sample we consider in this paper consists of 583 men and women who wore the Actiheart activity monitor for 5 days; we focus on binary "activity" and "inactivity" daily trajectories (see Figure 5 for example data from two subjects). The goals of this work are to describe and quantify the effects of age and BMI on the time-varying probability of being active over the course of a day, and to characterize the patterns of activity that differentiate subjects from each other and days within subjects. In addition to this motivating dataset, the proposed methods will be directly relevant to

existing and future accelerometer studies including the National Health and Nutrition Examination Survey (Troiano et al., 2008), the Women's Health Study (Shiroma et al., 2013), the Health ABC Study (Atkinson et al., 2007), and the Columbia Center for Children's Environmental Health birth cohort study.

1.2 Statistical framework

Speaking generally, we observe data $[Y_{ij}(t), \boldsymbol{x}_{ij}]$ for subjects $1 \leq i \leq I$, visits $1 \leq j \leq J_i$ and times $t \in [0, T]$, where $Y_{ij}(t)$ is a generalized response curve and $\boldsymbol{x}_{ij}$ is a length p vector of scalar covariates. For each time t, we assume $Y_{ij}(t)$ is a realization of random variable with an exponential family distribution. We introduce the generalized multilevel function-on-scalar regression model

$$\begin{aligned} \mathrm{E}[Y_{ij}(t)|b_i(t), v_{ij}(t)] &= \mu_{ij}(t) \\ g(\mu_{ij}(t)) &= \beta_0(t) + \sum_{k=1}^{p} x_{ij,k}\beta_k(t) + b_i(t) + v_{ij}(t) \end{aligned} \tag{1}$$

in which $g(\cdot)$ is a known link function, the $\beta_k(t)$ are fixed effect coefficient functions corresponding to the scalar covariates $\boldsymbol{x}$, $b_i(t)$ is a subject-specific random deviation from the fixed effect mean structure, and $v_{ij}(t)$ is a subject- and visit-specific random deviation from the subject-specific mean. As is detailed in later sections, we estimate fixed effect coefficients using a penalized spline expansion. The subject-level and subject-visit-level effects ($b_i(t)$ and $v_{ij}(t)$, respectively) are decomposed using a multilevel functional principal components analysis that separates within- and between-subject directions of variability, and principal component basis functions are estimated using penalized splines. All model parameters – including fixed effect spline coefficients, principal component spline coefficients, and principal components scores – are jointly estimated in a Bayesian analysis.

Elements of our analysis have antecedents in the statistical literature. Functional principal components analysis for cross-sectional continuous-valued curves has a long history in functional data analysis as a tool for dimension reduction and for identifying the major patterns that contribute to variation across curves; see Ramsay and Silverman (2005, §8.2) for an overview. Yao et al. (2005) describe a broadly used framework for FPCA in which the observed covariance matrix is calculated from discretely observed curves and decomposed, often following bivariate smoothing, to estimate principal component basis functions. Curve-specific scores (or loadings) are estimated using a mixed model framework. However, Goldsmith et al. (2013) noted that this standard FPCA method implicitly conditions on the estimated covariance and thus fails to account for uncertainty in estimated basis functions, meaning inference for individual curves can be poor. For multilevel functional data, Di et al. (2009) estimates both within- and between-subject covariances, and subsequently decomposes these into subject-level and visit-level principal component basis

functions; scores are again estimated in a mixed model framework. With some exceptions (for example, Cardot (2007) and Jiang and Wang (2010)), FPCA methods typically focus on the decomposition of curves around a common population mean rather incorporating covariates into the mean structure.

van der Linde (2008) develops a Bayesian approach to FPCA using low-dimensional spline expansions for basis functions and estimating parameters through a variational approximation to the full posterior; this work is based on the probabilistic and Bayesian (non-functional) PCA methods popularized in Tipping and Bishop (1999) and Bishop (1999). Bayesian PCA takes advantage of the connection between the mathematical formulation of traditional PCA and the Gaussian likelihood function. This connection does not encode the orthonormality constraints imposed by traditional PCA, and instead is more easily interpreted as a factor analysis. Some Bayesian methods have introduced orthonormality constraints (Šmídl and Quinn, 2007) at the expense of considerably more difficult inference and computation times, while others favor the latent factor analysis interpretation. We follow the latter approach. As with other factor analyses it is possible to rotate estimated components into their equivalent orthonormal space and thereby recover the appealing interpretation of traditional PCA.

There is an extensive literature for function-on-scalar regression with real-valued response curves. Brumback and Rice (1998) and Guo (2002) use penalized splines to model both population-level effects and curve-level deviations – the former relied on the use of fixed effects for computational convenience and the latter utilized random effect models. Several approaches have been developed that focus on population fixed effects only, treating individual curves as errors around the covariate-dependent mean; (Ramsay and Silverman, 2005, §13.4) provides an introduction. Developments in Reiss and Huang (2010) and Scheipl et al. (2013) use penalized splines to model fixed effects in cross sectional and multilevel models, respectively, using cross validation or restricted maximum likelihood to select tuning parameters. A criticism of these approaches is that they make the assumption that functional errors are uncorrelated over the domain, which typically does not hold for functional data and can lead to poor inference for fixed effects. Wavelet-based Bayesian functional mixed models are presented in Morris and Carroll (2006) and Morris et al. (2011) with errors in the wavelet space assumed to be independent, an assumption heuristically justified by the "whitening" property of wavelet transformations. Bayesian penalized splines are used in Baladandayuthapani et al. (2007) assuming error curves are composed only of uncorrelated measurement error. Recently, Goldsmith and Kitago (2013) developed a Bayesian penalized spline approach for multilevel function-on-scalar regression that models potential residual correlations explicitly, and showed posterior credible intervals for fixed effects achieve nominal coverage in simulations.

In contrast to the rich literature for real-valued functional data, relatively little work exists for generalized functional responses. Hall et al. (2006) directly extend the real-valued FPCA method of Yao et al. (2005) to generalized data by positing a latent continuous process that, through a known link function, gives

rise to the observed generalized outcome. The mean and covariance are estimated using observed data, and the latent mean and basis functions are obtained by inverting the known link function. Inference is based on bootstrap confidence intervals, although coverage properties in simulations are not reported. Building on the developments in Schein et al. (2003) for non-functional Bayesian generalized PCA, van der Linde (2009) develops a variational Bayesian algorithm for generalized FPCA that uses low-dimensional spline representations for the mean and basis functions. The variational approach gives a computationally efficient approximation to the full posterior but neglects correlation among model parameters. In FPCA, the population mean and curve-specific deviations are highly correlated, meaning that a variational approach may give good point estimates but not yield accurate inference.

With respect to the preceding literature review, our methods are statistically novel in several important ways. We provide a unified framework for both generalized function-on-scalar regression and functional principal components analysis. From a regression standpoint, we explicitly model residual correlation to improve inference for population-level effects; at the same time, the FPCA framework describes major directions of variability. The use of fully Bayesian estimation and inference, rather than variational Bayes approximations, avoids unreasonable assumptions of posterior independence and provides joint inference that has been shown to have good numerical properties in simulations that mimic our motivating data. Finally, we introduce generalized multilevel functional data and develop methods for this scenario; all methods can be simplified appropriately for cross-sectional data.

The remainder of the paper is organized as follows. Section 2 presents the novel methodological contributions of the manuscript, and includes subsections on the model specification, computation, and rotating estimated components to induce orthonormality. Section 3 presents simulation studies designed to mimic the motivating data and explore the estimation accuracy and inferential properties of the proposed methods. Section 4 presents the real data analysis. We close with a discussion in Section 5.

2 Methods

2.1 Model

Notationally, for subjects $1 \leq i \leq I$ and visits $1 \leq j \leq J_i$ let $Y_{ij}(t)$ be a generalized response curve arising from an exponential family and $\boldsymbol{x}_{ij}$ be an accompanying length p vector of scalar covariates. Our interest is in fitting model (1), which contains population-level fixed effects $\beta_k(t)$, subject-level deviations $b_i(t)$ from the covariate-dependent mean, and subject-visit specific deviations $v_{ij}(t)$ from the subject-specific mean. Generalizing the multilevel FPCA approach (Di et al., 2009), we expand subject-specific (level 1)

and subject-visit-specific (level 2) effects in terms of populations basis functions and unique scores:

$$\begin{aligned} \mathrm{E}[Y_{ij}(t)|b_i(t), v_{ij}(t)] &= \mu_{ij}(t) \\ g(\mu_{ij}(t)) &= \beta_0(t) + \sum_{k=1}^{p} x_{ij,k}\beta_k(t) + b_i(t) + v_{ij}(t) \\ &\approx \beta_0(t) + \sum_{k=1}^{p} x_{ij,k}\beta_k(t) + \sum_{k=1}^{K^{(1)}} c_{ik}^{(1)}\psi_k^{(1)}(t) + \sum_{k=1}^{K^{(2)}} c_{ijk}^{(2)}\psi_k^{(2)}(t). \end{aligned} \tag{2}$$

The approximation in the third line stems from the use of truncated functional principal components expansions for subject-specific effects $b_i(t)$ and subject-visit-specific effects $v_{ij}(t)$, and is implicit in all FPCA methods. Level 1 and level 2 basis functions ($\psi_k^{(1)}(t)$ and $\psi_k^{(2)}(t)$, respectively) describe the major patterns that generate variation across subjects and across visits within subjects, and associated scores ($c_{ik}^{(1)}$ and $c_{ijk}^{(2)}$, respectively) indicate the subject- and subject-visit-specific contribution of each basis function.

In practice curves are observed on a finite grid of length D that, for notational simplicity, we assume is shared across subjects. For finite data, let $\boldsymbol{Y}$ be the $(\sum_i J_i) \times D$ matrix of row-stacked generalized functional response; $\boldsymbol{X}$ be the $(\sum_i J_i) \times (p+1)$ fixed effects design matrix constructed by row-stacking the $\boldsymbol{x}_{ij}$; $\boldsymbol{\beta}$ be the $(p+1) \times D$ matrix with rows containing $\beta_k(t)$ evaluated on the finite grid; $\boldsymbol{Z}$ be a $(\sum_i J_i) \times I$ random intercept design matrix for the subject-specific effects; $\boldsymbol{b}$ be the $I \times D$ matrix with rows containing $b_i(t)$ evaluated on the finite grid; and $\boldsymbol{v}$ be the $(\sum_i J_i) \times D$ matrix with rows containing $v_{ij}(t)$ evaluated on the finite grid. Fixed effects and FPCA basis functions at both levels are expressed using a spline expansion. Let $\boldsymbol{\Theta}$ denote a $D \times K_\Theta$ matrix of cubic B-spline basis functions evaluated over the finite grid on which functions are observed. Spline coefficients for the fixed effects $\beta_k(t)$, the level 1 FPCA basis functions $\psi^{(1)}(t)$, and the level 2 FPCA basis functions $\psi^{(2)}(t)$ are columns in the matrices $\boldsymbol{B}_X$, $\boldsymbol{B}_{\psi^{(1)}}$, and $\boldsymbol{B}_{\psi^{(2)}}$, respectively. Thus $\boldsymbol{\beta} = \boldsymbol{B}_X^T\boldsymbol{\Theta}^T$ and, letting $\boldsymbol{C}^{(1)}$ and $\boldsymbol{C}^{(2)}$ be the matrices created by row-stacking level 1 and level 2 scores for each subject and subject-visit, $\boldsymbol{b} = \boldsymbol{C}^{(1)}\boldsymbol{B}_{\psi^{(1)}}^T\boldsymbol{\Theta}^T$ and $\boldsymbol{v} = \boldsymbol{C}^{(2)}\boldsymbol{B}_{\psi^{(2)}}^T\boldsymbol{\Theta}^T$. Model (2) can now be re-expressed for finite data using

$$\begin{aligned} \mathrm{E}[\boldsymbol{Y}|\boldsymbol{b}, \boldsymbol{v}] &= \boldsymbol{\mu} \\ g(\boldsymbol{\mu}) &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{v} \\ &= \boldsymbol{X}\boldsymbol{B}_X^T\boldsymbol{\Theta}^T + \boldsymbol{Z}\boldsymbol{C}^{(1)}\boldsymbol{B}_{\psi^{(1)}}^T\boldsymbol{\Theta}^T + \boldsymbol{C}^{(2)}\boldsymbol{B}_{\psi^{(2)}}^T\boldsymbol{\Theta}^T. \end{aligned} \tag{3}$$

Notationally, model (3) is formulated in a similar fashion as the continuous-valued cross sectional function-on-scalar regression models described in Ramsay and Silverman (2005, §13.4) and the continuous-valued multilevel function-on-scalar regression models described in Goldsmith and Kitago (2013). Inference for model (3) focuses on spline coefficients in the matrices $\boldsymbol{B}_X$, $\boldsymbol{B}_{\psi^{(1)}}$, and $\boldsymbol{B}_{\psi^{(2)}}$, and on the principal com-

ponent score matrices $\boldsymbol{C}^{(1)}$ and $\boldsymbol{C}^{(2)}$.

To ensure flexibility we use a rich B-spline basis by taking K_Θ large, but impose smoothness on resulting coefficient function estimates through the prior specification. In particular, we assume the following priors for the columns of $\boldsymbol{B}_X$, $\boldsymbol{B}_{\psi^{(1)}}$, and $\boldsymbol{B}_{\psi^{(2)}}$:

$$\begin{aligned}
\boldsymbol{B}_{X_k} &\sim \mathrm{N}\left[0, \sigma^2_{X_k} P^{-1}\right], \text{ for } 0 \leq k \leq p \\
\boldsymbol{B}_{\psi^{(1)}_k} &\sim \mathrm{N}\left[0, \sigma^2_{\psi^{(1)}_k} P^{-1}\right], \text{ for } 1 \leq k \leq K^{(1)} \\
\boldsymbol{B}_{\psi^{(2)}_k} &\sim \mathrm{N}\left[0, \sigma^2_{\psi^{(2)}_k} P^{-1}\right], \text{ for } 1 \leq k \leq K^{(2)}.
\end{aligned} \tag{4}$$

In (4), P is a pre-specified penalty matrix that enforces smoothness through the connection between Bayesian priors and quadratic penalization (Ruppert et al., 2003; Crainiceanu et al., 2005). We use $P = \alpha P_0 + (1-\alpha)P_2$ where P_0 and P_2 are zeroth- and second-order derivative penalty matrices (Eilers and Marx, 1996). Taking $0 < \alpha \leq 1$ balances the universal shrinkage encoded in P_0 with the smoothness constraint of P_2, while ensuring P is positive definite and priors are proper. In our simulations and real data analyses we set $\alpha = .1$ to predominantly enforce smoothness rather than shrinkage as is common in FDA; sensitivity analyses have indicated robustness to the choice of α in this analysis.

To complete the model specification, scores vectors are assigned independent standard Normal priors $\boldsymbol{c}_i^{(1)} \sim \mathrm{N}\left[0, I_{K^{(1)}}\right]$ and $\boldsymbol{c}_i^{(2)} \sim \mathrm{N}\left[0, I_{K^{(2)}}\right]$, a choice motivated by the factor analysis interpretation of the model as discussed in the introduction and in Tipping and Bishop (1999). Values for $K^{(1)}$ and $K^{(2)}$ are fixed constants chosen large enough to model major directions of uncertainty, keeping in mind that smoothness and shrinkage constraints help to control the effective dimension of the estimated basis. Finally, variance components $\sigma^2_{X_k}$, $\sigma^2_{\psi^{(1)}_k}$ and $\sigma^2_{\psi^{(2)}_k}$ are assigned IG$[0.01, 0.01]$.

2.2 Computation Using Stan

The model in Section 2.1 is implemented in `Stan` (Stan Development Team, 2013; Hoffman and Gelman, 2011), using an `R` interface (R Development Core Team, 2009) for data entry and for summarizing posterior samples. `Stan` is an open-source, general purpose programming language for Bayesian analysis that, at the user interface level, has similarities with `BUGS` (Lunn et al., 2009) or `JAGS` (Plummer, 2003). Samples are generated using Hamiltonian Monte Carlo, an MCMC algorithm that avoids random walk behavior by using the gradient of the log-posterior (Neal, 2011). In comparison with earlier MCMC algorithms such as the Gibbs sampler (Geman and Geman, 1984), Hamiltonian Monte Carlo offers fast convergence and parameter space exploration when posteriors are highly correlated, such as in the case of the fixed, subject-specific, and subject-day-specific effects in model (2). Code for both model (2) and for an analogous

cross-sectional model described in Section 3 is publicly available on the first author's website.

Computation time is a concern in all Bayesian approaches, especially for high-dimensional data such as those we consider. Here, computation times were reasonable for the moderate datasets considered in the simulations – taking several minutes for cross-sectional datasets consisting of up to 100 curves measured on grids of length 100, and taking at most a few hours for multilevel datasets with up to 100 subjects and 4 curves per subject. Real data analyses were more computationally expensive due to the higher dimensionality and increased complexity, and took several days. Details for computation time are provided in Sections 3 and 4.

2.3 Rotation

As noted in the introduction, we describe our methodology as a "functional principal components analysis" despite omitting orthonormality constraints on the estimated basis functions. FPCA has an appealing and well-established interpretation as estimating the major directions of variation within and between subjects, and we obtain that interpretation here through a singular value decomposition of the estimated basis functions. In this subsection we formalize that rotation omitting notation for level 1 and level 2 basis functions: both are obtained using the same steps.

FPCA is typically posed as an expansion $b_i(t) \approx \sum_{k=1}^{K} c_{ik}^* \psi_k^*(t)$, with the $\psi_k^*(t)$ orthonormal basis functions and scores c_{ik}^* uncorrelated zero mean random variables with non-increasing variances λ_k. Basis functions and variances are estimated using a truncated Karhunen-Loève decomposition of the covariance matrix $\text{Var}(b_i(t))$. Within each iteration of the sampler we estimate $\boldsymbol{C}\boldsymbol{B}_\psi^T\boldsymbol{\Theta}^T = \boldsymbol{C}\boldsymbol{\Psi}$ where $\boldsymbol{\Psi}$ are basis functions evaluated on a finite grid, and we wish to obtain an equivalent $\boldsymbol{C}^*\boldsymbol{\Psi}^*$ for which $\boldsymbol{\Psi}^*$ is an orthonormal basis. To do so, we use the singular value decomposition of the $\boldsymbol{\Psi}$ estimated without orthonormality constraints $\boldsymbol{\Psi} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}$ with $\boldsymbol{U}, \boldsymbol{V}$ unitary matrices and $\boldsymbol{D}$ diagonal. Making a substitution, we have $\boldsymbol{C}\boldsymbol{\Psi} = \boldsymbol{C}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}$ and define $\boldsymbol{C}^* := \boldsymbol{C}\boldsymbol{U}\boldsymbol{D}$ and $\boldsymbol{\Psi}^* := \boldsymbol{V}$. Moreover, the prior assumption that $\text{Var}(\boldsymbol{c}_i) = I$ implies that for each row $\boldsymbol{c}_i^*$ of $\boldsymbol{C}^*$, $\text{Var}(\boldsymbol{c}_i^*) = \text{Var}(\boldsymbol{c}_i\boldsymbol{U}\boldsymbol{D}) = \boldsymbol{D}\boldsymbol{U}^T\text{Var}(\boldsymbol{c}_i)\boldsymbol{U}\boldsymbol{D} = \boldsymbol{D}\boldsymbol{U}^T\boldsymbol{I}\boldsymbol{U}\boldsymbol{D} = \boldsymbol{D}^2$. Thus estimates of the score variance components λ_k are provided by squaring the diagonal entries of $\boldsymbol{D}$. This rotation can be conducted within each iteration of the sampler and, accounting for potential sign changes in the basis functions, provides a posterior distribution of orthonormal basis functions.

3 Simulations

We demonstrate the performance of our method using a simulation in which generated data mimic the motivating application. In the following subsections we consider both cross-sectional and multilevel scenarios; our focus is on assessing the estimation accuracy and inferential properties of the proposed methods. All

code for the following simulations is publicly available.

3.1 Cross-sectional simulations

We generate binary response curves $Y_i(t)$ on an equally spaced grid of length 100 according to the model

$$\begin{aligned} \mathrm{E}[Y_i(t)|b_i(t)] &= \mu_i(t) \\ g(\mu_i(t)) &= \beta_0(t) + x_{i1}\beta_1(t) + \sum_{k=1}^{2} c_{ik}^{(1)}\psi_k^{(1)}(t), \end{aligned} \tag{5}$$

and use the logit link $g(\cdot)$. Model (5) is a simplification of model (2) for the cross-sectional case with one scalar covariate. We let $t \in [0, 1]$ represent a 24-hour period as in the motivating accelerometer study, and in the following use descriptions motivated by this context. The intercept is $\beta_0(t) = -1.5 - \sin(2t\pi) - \cos(2t\pi)$, which roughly mimics a circadian rhythm over one day. The fixed effect $\beta_1(t) = \frac{1}{20}\phi(\frac{t-.6}{.15^2})$, where $\phi(\cdot)$ is the standard Normal density function, affects the probability of activity in the afternoon but not in the late evening or early morning, and we generate scalar predictors using $x_{i1} \sim N(0, 25)$. The orthogonal basis functions are chosen to be $\psi_1^{(1)}(t) \propto -1.5 - \sin(2t\pi) - \cos(2t\pi)$ and $\psi_2^{(1)}(t) \propto -\sin(4t\pi)$, and are normalized to 1. The first basis function is amplifies or diminishes the circadian rhythm found in $\beta_0(t)$, broadly giving higher or lower overall activity patterns, while the second affects activity probabilities in the early and later afternoon. Subject-level PC scores are generated using variance components are set to $\lambda_1 = 3$ and $\lambda_2 = 1.5$.

One hundred datasets are constructed according to the preceding model for all combinations of sample size $I \in \{50, 100\}$ and number of estimated principal components $\hat{K} \in \{2, 5\}$, giving a total of four possible simulation designs. For $\hat{K} = 5$ the number of estimated PC basis functions is larger than the number of true basis functions, which is held at $K = 2$ throughout. For each dataset, we estimate model parameters using a simplification of the methods described in Section 2 for cross sectional data. Estimation and inference is based on posterior means and quantiles of 5000 iterations from the sampler, after discarding the first 2000 as burn-in; visual inspection and diagnostics for the one simulated dataset indicate that these levels are sufficient for convergence to and exploration of the posterior distribution. We quantify estimation accuracy for fixed effects using the integrated mean squared error IMSE $= \int_0^1 (\beta_k(t) - \hat{\beta}_k(t))^2\, dt$ and for the latent subject probability trajectories using the average integrated mean squared error AIMSE $= \frac{1}{I}\sum_{i=1}^{I}\int_0^1 (\mu_i(t) - \hat{\mu}_i(t))^2\, dt$. Inference is evaluated using average pointwise coverage of 95% posterior credible intervals.

Figure 1 illustrates the simulation design and results for a single dataset with $I = 50$ and $\hat{K} = 5$. Simulated latent probability curves $\mu_i(t)$ are shown in the left panel, and demonstrate the structure of

activity trajectories as well as their variability across subjects. The true and estimated fixed effects $\beta_0(t)$ and $\beta_1(t)$ are plotted in the middle panels, along with 95% credible intervals. Finally, the right panel shows the observed binary data $Y_i(t)$ for one subject in blue, as well as the latent probability curve $\mu_i(t)$ and a sample from the posterior distribution of $\mu_i(t)$.

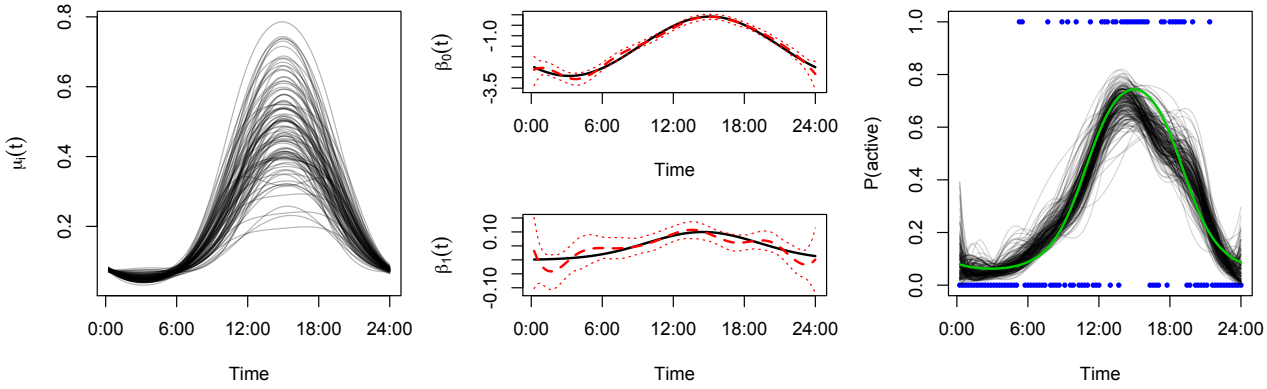

Figure 1: Illustration of data and results for cross-sectional simulations. The left panel show simulated probability curves $\mu_i(t)$ for all subjects $i \in 1, \ldots, I$. The middle panels show fixed effects $\beta_0(t)$ and $\beta_1(t)$ (top and bottom, respectively) in black, with estimates in red and 95% pointwise credible intervals in dashed red lines. The right panel shows observed binary responses $Y_1(t)$ for subject $i = 1$ in blue, the corresponding probability curve $\mu_1(t)$ in green, and a sample from the posterior of $\mu_1(t)$ in black.

Table 1 provides the mean (across simulated datasets) IMSE for fixed effects and AIMSE for latent probability trajectories, as well as mean pointwise coverage and computation time. As one would expect, increasing sample size improves estimation accuracy for fixed effects; the estimation of fixed effects is not noticeably affected by changing the number of estimated principal components. Estimation accuracy for latent subject effects is only moderately improved by increasing the sample size since subject-level information is not increased, and any improvement would necessarily depend on improved estimation of fixed effects and PC basis functions. Coverage for fixed effects and latent subject trajectories is near nominal levels in all cases, although intervals are somewhat conservative for $\beta_1(t)$ and anti-conservative for $\mu_i(t)$ when $\hat{K} = 2$. Computation times increase both with sample size I and the number of estimated PC basis functions $\hat{K}$, and range from a few minutes to roughly half an hour. For reference, the IMSEs appearing in Figure 1 are $\text{IMSE}(\hat{\beta}_0(t)) = 0.012$ and $\text{IMSE}(\hat{\beta}_1(t)) = 0.0034$.

3.2 Multilevel simulations

For the multilevel case, we generate binary response curves $Y_{ij}(t)$ according to the model

$$\begin{aligned} \mathrm{E}[Y_{ij}(t)|b_i(t), v_{ij}(t)] &= \mu_{ij}(t) \\ g(\mu_{ij}(t)) &= \beta_0(t) + x_{i1}\beta_1(t) + \sum_{k=1}^{2} c_{ik}^{(1)}\psi_k^{(1)}(t) + \sum_{k=1}^{2} c_{ijk}^{(2)}\psi_k^{(2)}(t) \end{aligned} \tag{6}$$

	IMSE			Coverage			Comp. Time
	$\beta_0(t)$	$\beta_1(t)$	$\mu_i(t)$	$\beta_0(t)$	$\beta_1(t)$	$\mu_i(t)$	(in sec)
$I = 50; \hat{K} = 2$	0.026	0.006	0.066	0.936	0.969	0.919	389
$\hat{K} = 5$	0.024	0.005	0.064	0.943	0.983	0.965	950
$I = 100; \hat{K} = 2$	0.012	0.003	0.050	0.944	0.971	0.924	771
$\hat{K} = 5$	0.011	0.003	0.050	0.943	0.978	0.959	1805

Table 1: Cross-sectional results averaged across the 100 simulated datasets. Integrated mean squared errors are defined as IMSE $= \int_0^1 (\beta_p(t) - \hat{\beta}_p(t))^2\, dt$ for fixed effects and AIMSE $= \frac{1}{I}\sum_{i=1}^{I} \int_0^1 (\mu_i(t) - \hat{\mu}_i(t))^2\, dt$ for probability curves. Coverage is averaged over 95% pointwise credible intervals. Computation time is reported in seconds.

again assuming a logit link function. The fixed effects $\beta_0(t)$ and $\beta_1(t)$, scalar covariate x_{i1}, level 1 basis functions $\psi_1^{(1)}(t)$ and $\psi_2^{(1)}(t)$, and level 1 variances $\lambda_1^{(1)}$ and $\lambda_2^{(1)}$ are as in Section 3.1. To this, we add level 2 basis functions $\psi_1^{(2)}(t) \propto -1.5 - \sin(2t\pi) - \cos(2t\pi)$ and $\psi_2^{(2)}(t) \propto -\cos(4t\pi)$, again normalized to 1. Level 2 variance components are $\lambda_1^{(2)} = 3$ and $\lambda_2^{(2)} = 1.5$. In this setting, the level 1 basis functions describe subject-level directions of variability and the level 2 basis functions describe subject-day-specific directions of variability. Setting $\psi_1^{(1)}(t) = \psi_1^{(2)}(t)$ has the interpretation that the major pattern distinguishing subjects as distinguishes days within a subject, but increases the difficulty of the estimation problem.

For all simulations we let J_i, the number of days observed per subject, be 4. Once again, one hundred datasets are constructed for all combinations of sample size $I \in \{50, 100\}$ and number of estimated principal components $\hat{K}^{(1)} = \hat{K}^{(2)} \in \{2, 5\}$, giving a total of four possible simulation designs. Model parameters are estimated using the methodology described in Section 2; we use chains of length 5000, discarding the first 2000 as burn-in.

Table 2 provides the mean (across 100 simulated datasets) IMSE for fixed effects and AIMSE for latent probability trajectories, as well as mean pointwise coverage and computation time. These results reiterate the major points found in the cross sectional simulations. In particular, estimation accuracy for fixed effects improves as sample size increases; estimation of subject effects also improves as sample size increases, although to a lesser extent than for fixed effects; in all cases, coverage for fixed effects and latent probability trajectories in near nominal levels; and the coverage of intervals for the latent subject-specific trajectories $\mu_i(t)$ and latent subject-day-specific trajectories $\mu_{ij}(t)$ increases as $\hat{K}$ increases. In this setting, increasing $\hat{K}$ does not affect estimation accuracy for $\beta_1(t)$ but may negatively affect accuracy for $\beta_0(t)$ due to the flexibility in the model or because $\psi_1^{(1)}(t) = \psi_1^{(2)}(t)$; meanwhile, increasing $\hat{K}$ may improve coverage for both fixed effects. For the multilevel simulations, computation times are larger but not prohibitive, and generally take between one and four hours.

	IMSE				Coverage				Comp. Time
	$\beta_0(t)$	$\beta_1(t)$	$\mu_i(t)$	$\mu_{ij}(t)$	$\beta_0(t)$	$\beta_1(t)$	$\mu_i(t)$	$\mu_{ij}(t)$	(in sec)
$I = 50; \hat{K^{(1)}} = \hat{K^{(1)}} = 2$	0.0073	0.00027	0.038	0.059	0.941	0.923	0.920	0.934	2893
$\hat{K^{(1)}} = 5$	0.0084	0.00026	0.039	0.060	0.946	0.946	0.956	0.968	5938
$I = 100; \hat{K^{(1)}} = \hat{K^{(2)}} = 2$	0.0035	0.00014	0.030	0.050	0.945	0.938	0.934	0.946	6943
$\hat{K^{(2)}} = 5$	0.0040	0.00014	0.030	0.051	0.942	0.944	0.959	0.969	12290

Table 2: Multilevel simulation results averaged across 100 datasets. Integrated mean squared errors are defined as IMSE $= \int_0^1 (\beta_p(t) - \hat{\beta}_p(t))^2\, dt$ for fixed effects and AIMSE $= \frac{1}{I}\sum_{i=1}^{I} \int_0^1 (\mu_i(t) - \hat{\mu}_i(t))^2\, dt$ for probability curves. Coverage is averaged over 95% pointwise credible intervals. Computation time is reported in seconds.

4 Application

We now apply methods of Section 2 to the motivating data. For 583 subjects, we observe age, BMI, and minute-by-minute activity count trajectories for 5 days. For our analysis these activity counts are dichotomized into "active" and "inactive" by thresholding the observed activity counts at 10. Similar results are obtained from other thresholds between 0 and 25; this range is fairly conservative for defining activity in order to allow for low-intensity activity commonly observed in elderly subjects. To reduce the computational burden of the analysis, data are thinned to one data point for every 10 minutes, giving 144 observations per subject per day. Our model considers age and BMI, centered at 60 and 25 respectively, as potential predictors of activity and we set $K^{(1)} = K^{(2)} = 8$. The dimension of the B-spline basis $\mathbf{\Theta}$ is $K_\Theta = 10$, which suffices to estimate the smooth effects observed in this application. We fit model (3) using 5000 iterations of the sampler, discarding 2000 as burn-in; total computation time was 10 days.

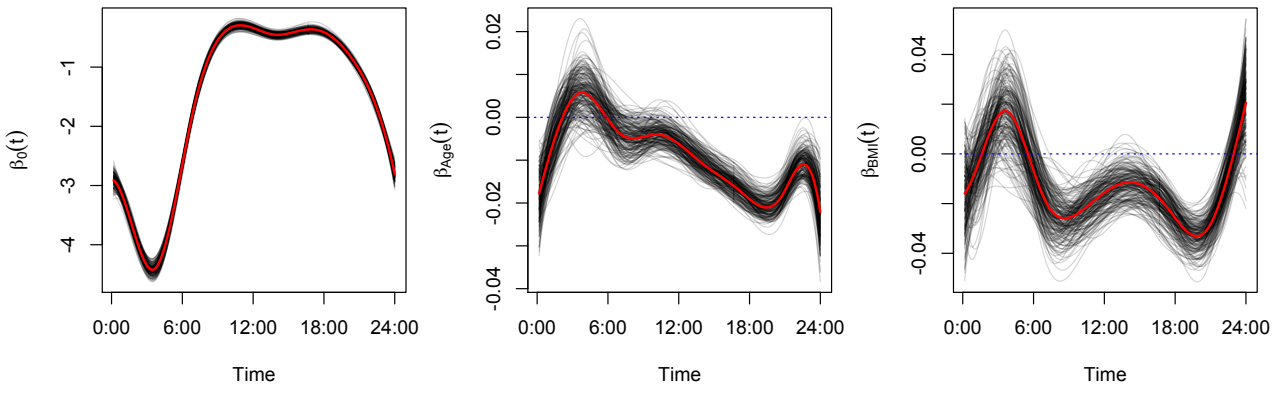

Figure 2: Estimated fixed effects (red) from the real data analysis with samples from the posterior (black). The left panel shows the intercept $\beta_0(t)$; the middle panel shows the age effect $\beta_{Age}(t)$; the right panel shows the BMI effect $\beta_{BMI}(t)$.

Figures 2 and 3 provide the estimated fixed effect coefficients. In Figure 2 we show the estimated effect in red and a posterior sample in black. The intercept $\beta_0(t)$ gives the log odds of activity for 60 year old

subject with a BMI of 25, and has an expected circadian rhythm shape. Coefficient functions $\beta_{age}(t)$ and $\beta_{BMI}(t)$ have a log odds ratio interpretation; for example, $\beta_{age}(t)$ is the change in the log odds of activity for each one year increase in age, keeping BMI fixed, over a 24-hour time course. From the posterior distribution, it seems that both age and BMI have significant negative effects on the probability of being active during daytime hours. The effect of age is most pronounced in the late afternoon, perhaps as a result of increased fatigue in older subjects, while BMI is most significant in the mid-morning and mid-afternoon. Figure 3 demonstrates these effects by plotting the fitted probability of being active over a 24-hour period for several age and BMI levels.

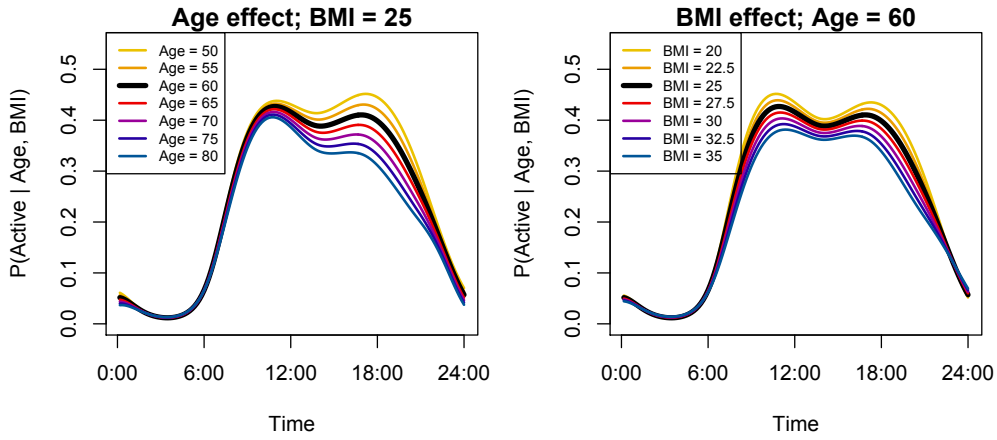

Figure 3: The left panel shows the effect on the probability of being active of varying age while keeping BMI fixed. The right panel shows the effect of varying BMI while keeping age fixed. In both panels, the subject- and subject-day-specific effects are set to zero.

In addition to fixed effects, we estimate level 1 and level 2 principal component basis functions $\psi^{(1)}(t)$ and $\psi^{(2)}(t)$, which have been rotated to induce orthonormality as described in 2.3. These functions model the subject- and subject-day-specific residual dependency in the 24-hour trajectories unaccounted for in the covariate-dependent mean. The top row of Figure 4 shows the directions of variation explained by the first two level 1 basis functions by plotting $g\left(\beta_0(t) \pm \sqrt{\lambda_k^{(1)}}\psi_k^{(1)}(t)\right)$, $k = 1, 2$; the third panel shows the scree plot for the level 1 decomposition. The major directions that distinguish subjects are a general shift in the probability of being active and a contrast in the probability of being active in the daytime and non-daytime hours. Similar plots are shown for the level 2 decomposition in the second row of Figure 4. Although these figures show the basis functions using the probability of activity, the percent variance explained is calculated and the orthonormality property enforced in the log odds of activity scale. The proportion of residual (after removing fixed effects) variance explained by subject level effects, given by $\frac{\sum_{k=1}^{K^{(1)}} \lambda_k^{(1)}}{\sum_{k=1}^{K^{(1)}} \lambda_k^{(1)} + \sum_{k=1}^{K^{(2)}} \lambda_k^{(2)}}$ is 0.46 in this application, indicating moderate stability within subjects over multiple days.

Finally, we compare fitted values and observed data in Figure 5. The top row contains plots for a 85

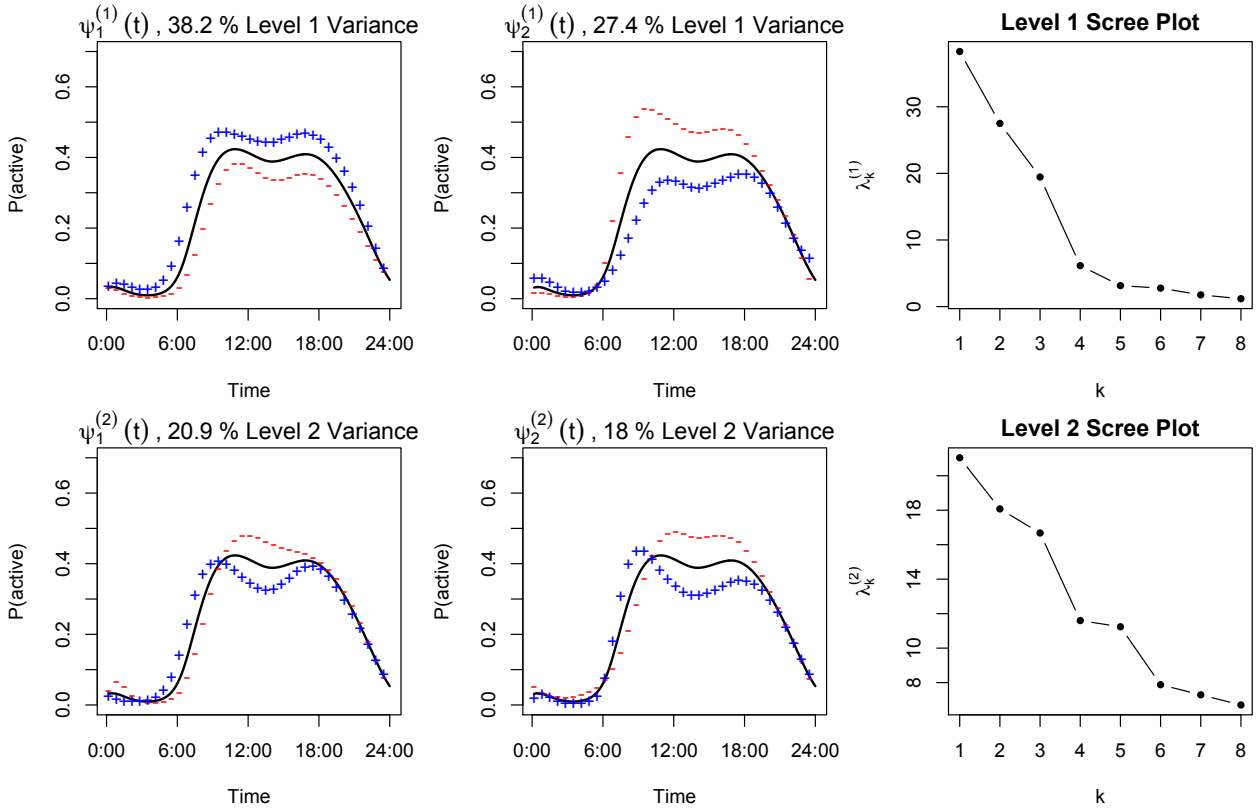

Figure 4: Estimated MFPCA basis functions and scree plots for subject-level and subject-day-level effects (top and bottom row, respectively). Basis functions are illustrated by plotting $g\left(\beta_0(t) \pm \sqrt{\lambda_k^{(L)}}\psi_k^{(L)}(t)\right)$ for basis functions $k \in \{1, 2\}$ and levels $L \in \{1, 2\}$.

year old subject with a BMI of 26.5. The left and middle panels show observed data for two different days as blue dots and a moving average of the observed data as a green trajectories. Subject-day-specific estimates, combining fixed effects with level 1 and level 2 FPC effects, are overlayed: the posterior mean $\hat{Y}_{ij}(t)$ is shown as a red curve and a posterior sample is shown in black. The right panel shows the moving average trajectory for each of the five observed days as separate green curves. Subject-specific estimates, combining fixed effects with only level 1 FPC effects, are again overlayed with the posterior mean $\hat{Y}_i(t)$ in red and a posterior sample in black. Data for a second subject, aged 51 years with a BMI of 23.8, is shown in the bottom row of Figure 5. Our method accurately captures both large scale patterns and detailed phenomena, giving accurate estimates of the probability of being active over a 24-hour period using relatively few principal components and scores.

5 Concluding remarks

The generalized multilevel function-on-scalar regression and principal components analysis techniques developed in this manuscript are necessary tools in modern functional data analysis and are required by

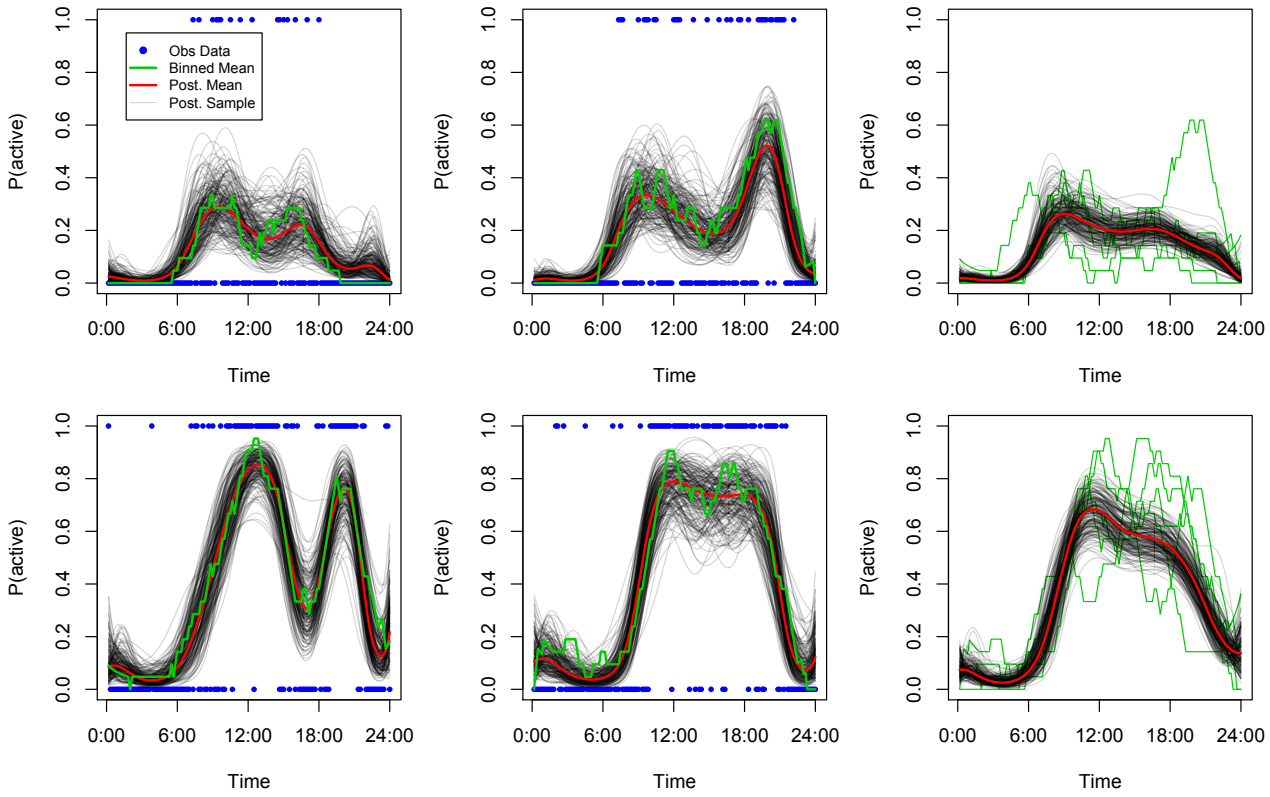

Figure 5: Fitted values for two subjects, separately by row. In each row, the left and middle panels show observed binary values $Y_{ij}(t)$ as blue dots (separate days are shown in each panel). A moving average of the observed data is shown in green. Estimates of subject-day-specific probability trajectories $\hat{\mu}_{ij}(t)$ are shown in red, and a sample from the posterior of $\mu_{ij}(t)$ is shown in black. In the right panel of each row, moving averages for each observed day of the subject are shown in green, estimates of subject-specific probability trajectories $\hat{\mu}_i(t)$ are shown in red, and a sample from the posterior of $\mu_i(t)$ is shown in black.

our application. From a methodological perspective, this work has two major motivations that have often been neglected in functional data analysis. For the problem of function-on-scalar regression, some effort is needed to account for residual correlation within functions to develop reasonable inferential procedures. Meanwhile, in functional principal components analyses, it is common to condition (implicitly or explicitly) on the estimated mean and basis functions when predicting latent subject-specific trajectories and constructing related confidence/credible intervals. Both of these issues are made more difficult in the context of generalized and multilevel functional data. Our approach has been to jointly model all parameters of interest in a Bayesian context, and in doing so we have attempted to develop a unified framework for both function-on-scalar regression and functional principal components analysis.

In the motivating real-data analysis, we quantify and confirm a scientifically plausible expectation: that the probability of activity decreases as individuals age and as BMI increases, and these effects are dynamic over the course of the day. Moreover, we identify the major patterns of activity that distinguish subjects from each other and that distinguish days within subjects. By focusing on a binary activity variable we

address a concern that is distinct from the intensity of activity, instead examining changes in sedentary behavior in response to changes in covariates. Of course, an analysis of the changes in activity intensity is also warranted, as is the consideration of other potentially important covariates the allowing for non-linear effects.

The Bayesian procedure we develop was shown in realistic simulations to have good estimation and inferential properties. Not surprisingly, computation time can be a serious concern particularly as sample sizes, grid densities, and the number of estimated principal component basis functions grow. Future work focusing on variational Bayes or other approximations could address these concerns and, we suspect, would result in good estimation; however, the decrease in computational burden may be accompanied by poorer inferential performance due to the assumptions needed for such an approximation. Balancing these will depend on the particular data scenario, and both will be important.

6 Acknowledgments

We thank Luigi Ferrucci, Principal Investigator of the Baltimore Longitudinal Study on Aging, for encouraging the use of the BLSA accelerometer data that motivated this work and for his scientific insight and guidance.

References

Atkinson, H. H., Rosano, C., Simonsick, E. M., Williamson, J. D., Davis, C., Ambrosius, W. T., Rapp, S. R., Cesari, M., Newman, A. B., Harris, T. B., Rubin, S. M., Yaffe, K., Satterfield, S., and Kritchevsky, S. B. "Cognitive function, gait speed decline, and comorbidities: the health, aging and body composition study." The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 62:844–850 (2007).

Baladandayuthapani, V., Mallick, B., Young Hong, M., Lupton, J., Turner, N., and Carroll, R. J. "Bayesian Hierarchical Spatially Correlated Functional Data Analysis with Application to Colon Carcinogenesis." Biometrics, 64:64–73 (2007).

Bishop, C. M. "Bayesian PCA." Advances in neural information processing systems, 382–388 (1999).

Brumback, B. and Rice, J. "Smoothing spline models for the analysis of nested and crossed samples of curves." Journal of the American Statistical Association, 93:961–976 (1998).

Cardot, H. "Conditional functional principal components analysis." Scandinavian journal of statistics, 34:317–335 (2007).

Crainiceanu, C. M., Ruppert, D., and Wand, M. P. "Bayesian analysis for penalized spline regression using WinBUGS." Journal of statistical software, 14:165–185 (2005).

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. "Multilevel Functional Principal Component Analysis." Annals of Applied Statistics, 4:458–488 (2009).

Eilers, P. H. C. and Marx, B. D. "Flexible smoothing with B-splines and penalties." Statistical Science, 11:89–121 (1996).

Geman, S. and Geman, D. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." Pattern Analysis and Machine Intelligence, IEEE Transactions on, (6):721–741 (1984).

Goldsmith, J., Greven, S., and Crainiceanu, C. M. "Corrected Confidence Bands for Functional Data using Principal Components." Biometrics, 69:41–51 (2013).

Goldsmith, J. and Kitago, T. "Assessing Systematic Effects of Stroke on Motor Control using Hierarchical Function-on-Scalar Regression." Technical Report (2013).

Guo, W. "Functional mixed effects models." Biometrics, 58:121–128 (2002).

Hall, P., Müller, H.-G., and Wang, J. L. "Properties of Principal Component Methods for Functional and Longitudinal Data Analysis." Annals of Statistics, 34:1493–1517 (2006).

Hoffman, M. D. and Gelman, A. "The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo." arXiv preprint arXiv:1111.4246 (2011).

Howell, J., Strong, M., Weisenberg, J., Kakade, A., Gao, Q., Cuddihy, P., Delisle, S., Kachnowski, S., and Maurer, M. "Maximum Daily 6 Minutes of Activity: An Index of Functional Capacity Derived from Actigraphy and Its Application to Older Adults with Heart Failure." Journal of the American Geriatric Society, 58:931–936 (2010).

Jiang, C.-R. and Wang, J.-L. "Covariate adjusted functional principal components analysis for longitudinal data." The Annals of Statistics, 38:1194–1226 (2010).

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. "The BUGS project: Evolution, critique and future directions (with discussion)." Statistics in Medicine, 28:3049–3082 (2009).

Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P., and Gutstein, H. "Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data." Annals of Applied Statistics, 5:894–923 (2011).

Morris, J. S. and Carroll, R. J. "Wavelet-based functional mixed models." Journal of the Royal Statistical Society: Series B, 68:179–199 (2006).

Neal, R. "MCMC Using Hamiltonian Dynamics." Handbook of Markov Chain Monte Carlo, Chapter 5, 113–162 (2011).

Plummer, M. "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling." In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March, 20–22 (2003).

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2009). ISBN 3-900051-07-0.
URL http://www.R-project.org

Ramsay, J. O. and Silverman, B. W. Functional Data Analysis. New York: Springer (2005).

Reiss, P. T. and Huang, L. "Fast Function-on-Scalar Regression with Penalized Basis Expansions." International Journal of Biostatistics, 6:Article 28 (2010).

Rundle, A., Goldstein, I. F., Mellins, R. B., Ashby-Thompson, M., Hoepner, L., and Jacobson, J. S. "Physical activity and asthma symptoms among New York City Head Start children." Journal of Asthma, 46:803–809 (2009).

Ruppert, D., Wand, M. P., and Carroll, R. J. Semiparametric Regression. Cambridge: Cambridge University Press (2003).

Schein, A. I., Saul, L. H., and Ungar, A. "A generalised linear model for principal component analysis of binary data." In Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics (2003).

Scheipl, F., Staicu, A.-M., and Greven, S. "Additive Mixed Models for Correlated Functional Data." Under Review (2013).

Schrack, J. A., Zipunnikov, V., Goldsmith, J., Bai, J., Simonshick, E. M., Crainiceanu, C. M., and Ferrucci, L. "Assessing the "Physical Cliff": Detailed Quantification of Aging and Physical Activity." Journal of Gerontology: Medical Sciences (2014).

Shiroma, E. J., Freedson, P. S., Trost, S. G., and Lee, I. M. "Patterns of accelerometer-assessed sedentary behavior in older women." Journal of the American Medical Association, 310:2562–2563 (2013).

Spierer, D. K., Hagins, M., Rundle, A., and E, P. "A comparison of energy expenditure estimates from the Actiheart and Actical physical activity monitors during low intensity activities, walking, and jogging." European Journal of Applied Physiology, 111:659–667 (2011).

Stan Development Team. Stan Modeling Language User's Guide and Reference Manual, Version 1.3 (2013). URL http://mc-stan.org/

Tipping, M. E. and Bishop, C. "Probabilistic Principal Component Analysis." Journal of the Royal Statistical Society: Series B, 61:611–622 (1999).

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., and McDowell, M. "Physical activity in the United States measured by accelerometer." Medicine & Science in Sports & Exercise, 40:181–188 (2008).

Trost, S. G., McIver, K. L., and Pate, R. R. "Conducting accelerometer-based activity assessments in field-based research." Medicine and Science in Sports and Exercise, 37:S531–543 (2005).

van der Linde, A. "Variational Bayesian Functional PCA." Computational Statistics and Data Analysis, 53:517–533 (2008).

—. "A Bayesian latent variable approach to functional principal components analysis with binary and count." Advances in Statistical Analysis, 93:307–333 (2009).

Šmídl, V. and Quinn, A. "On Bayesian principal component analysis." Computational Statistics & Data Analysis, 51:4101–4123 (2007).

Ward, D. S., Evenson, K. R., Vaughn, A., Rodgers, A. B., and Troiano, R. P. "Accelerometer use in physical activity: best practices and research recommendations." Medicine and science in sports and exercise, 37:S582–588 (2005).

Yao, F., Müller, H., and Wang, J. "Functional data analysis for sparse longitudinal data." Journal of the American Statistical Association, 100(470):577–590 (2005).