Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Variable Selection in the Functional Linear Concurrent Model

Jeff Goldsmith[a‡], **Joseph E. Schwartz**[b,c]

We propose methods for variable selection in the context of modeling the association between a functional response and concurrently observed functional predictors. This data structure, and the need for such methods, is exemplified by our motivating example: a study in which blood pressure values are observed throughout the day, together with measurements of physical activity, location, posture, affect or mood, and other quantities that may influence blood pressure. We estimate the coefficients of the concurrent functional linear model using variational Bayes and jointly model residual correlation using functional principal components analysis. Latent binary indicators partition coefficient functions into included and excluded sets, incorporating variable selection into the estimation framework. The proposed methods are evaluated in simulations and real-data analyses, and are implemented in a publicly available R package with supporting interactive graphics for visualization. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: Ambulatory blood pressure; Functional data; Intensive longitudinal data; Spline smoothing; Variational Bayes; Wearable devices

1. Introduction

Portable, unobtrusive devices capable of monitoring blood pressure, physical activity, heart rate, location, and other quantities have made it possible to concurrently collect many data streams on study participants in parallel. In this and similar data settings, the functional linear concurrent model (FLCM) provides a useful framework for understanding the association between the functional response and functional covariates at a specific time. This model can be written

$$Y_i(t) = \beta_0(t) + \sum_{k=1}^{p} X_{ik}(t)\beta_k(t) + \delta_i(t), \tag{1}$$

and has been previously studied in the context of a relatively small number of predictors p [1, 2, 3]. However, the presence of a large number of data streams – a context that is becoming increasingly common – necessitates the incorporation of variable selection methods into the estimation of model (1). This problem has not yet been addressed, and the development of such methods is the main contribution of this manuscript.

Our approach begins with the expansion of coefficient functions $\beta_k(t)$ in model (1) in terms of a spline basis. Sparsity is induced by setting all spline coefficients for a given coefficient function to zero, at least approximately; to do so we use tools from Bayesian variable selection, in particular by specifying spike-and-slab priors for groups of spline coefficients. To model correlation in residual curves $\delta_i(t)$, we use a functional principal components expansion. We jointly estimate all model parameters using a computationally efficient variational Bayes algorithm and choose the tuning parameter, the nonzero "spike" prior variance, using cross validation. All methods are implemented in the publicly available `R` package `vbvs.concurrent`.

[a] *Department of Biostatistics, Columbia Mailman School of Public Health, Columbia University, USA*
[b] *Department of Medicine, Columbia University Medical Center, Columbia University, USA*
[c] *Department of Psychiatry and Behavioral Sciences, Stony Brook University, USA*
‡ *Email: helle@math.ku.dk*

As noted above, the estimation of parameters in the FLCM has been the subject of several previous papers, although none of these have focused on variable selection. Indeed, variable selection is the subject of a small but growing literature in functional data analysis. [4] developed variable selection tools for the linear scalar-on-function regression model; more recently, [5] and [6] have proposed methods for function-on-scalar regression models. Each of these made use of group penalties (e.g. group lasso or group MCP) for spline coefficients and, by recasting the functional regression model as a standard linear model, could apply off-the-shelf software for estimation and penalization. Because of the form of the least-squares estimate of the FLCM, it is not obvious that this strategy is applicable in the current setting. [6] also noted the importance of accounting for residual correlation when performing variable selection in a functional response model and used a "pre-whitening" approach to address this issue. In contrast, we jointly model the residual correlation structure and the regression coefficients in a Bayesian framework.

Variational Bayes algorithms have enjoyed some popularity in the functional data analysis literature, largely due to the computation efficiency of these approaches in comparison to Monte Carlo sampling and, in some cases, frequentist estimation methods. [7] and [8] developed variational algorithms for scalar-on-function regressions, and [9] used variational Bayes for function-on-scalar regression. [10] and [11] used variational approximations in a Bayesian approach to functional principal components analysis; this work is relevant in particular to our residual decomposition, although it did not address the incorporation of covariates in the mean structure. Recently, [12] used variational Bayes to combine a factor analysis related to Bayesian functional principal components analysis with the registration of features across curves.

Our method for variable selection is related to a recent EM-based approach to variable selection in the standard linear model [13]. In contrast to previous spike-and-slab normal mixture formulations, in which the spike distribution is considered to be a point mass at zero [14, 15], [13] posited a continuous spike prior. Doing so allowed the derivation of a closed-form EM algorithm and introduced the spike prior variance as a tuning parameter whose adjustment results in a sequence of progressively sparser models. Here, we adapt this basic approach to the context of variable selection in the FLCM and estimate parameters using variational Bayes rather than an EM algorithm.

The FLCM is a special case of function-on-function regression. Most generally, the linear function-on-function regression model allows response functions and predictor functions to be observed on different domains, and supposes a coefficient surface that relates the predictor functions to each point in the response domain via integration [16, 1]. Recent work for function-on-function regression has extended the basic model to allow for multiple functional predictors and scalar covariates, and has developed various estimation strategies [17, 18, 19]. In addition to the FLCM, another common special case of function-on-function regression is the historic functional regression model [20]. The historic model allows the response $Y_i(t)$ at time t to be influenced only by $X_i(s)$ for $s \leq t$; conceptually, this prevents future observations of the predictor functions to influence the response in the present. For more information regarding function-on-function regression and its special cases, see [21, Section 6].

Given this literature, our focus on the FLCM is driven by our motivating data, described in more detail in the following Section. Both responses and predictors are sparse and irregular across subjects, but are measured on the same time points within subjects. Thus, concurrent associations can be modeled but long term associations in the context of a historical model would be difficult to estimate and, indeed, are biologically unlikely. For more densely observed data, a historical model incorporating variable selection may be more appropriate.

Our presentation of methods in Section 3 will be general, but we first motivate our work by a discussion of data collected in the Masked Hypertension Study in Section 2. Simulations are conducted in Section 4, and an analysis of the motivating data is presented in Section 5. We end with a discussion in Section 6.

2. The Masked Hypertension Study

Elevated blood pressure, or hypertension, is associated with increased risk of several common diseases, most notably diabetes, myocardial infarction, and stroke. Hypertension diagnoses have historically been based solely on blood pressure measurements taken in the clinic setting, but a growing body of work demonstrates that out-of-clinic blood pressure provides a more complete assessment of cardiovascular health. Indeed, the term "masked hypertension" was coined to describe individuals with normal clinic blood pressure but elevated ambulatory or daytime blood pressure [22]; in the Masked Hypertension Study (888 subjects), 16% of participants with normal clinic blood pressure had masked hypertension [23].

Understanding daytime blood pressure level depends on frequent measurement as a participant engages in normal activities, and a 24-hour ambulatory blood pressure (ABP) monitoring device is used to this end. Briefly, participants in the Masked Hypertension Study wore a portable blood pressure cuff that measured blood pressure every 28 minutes over a 24-hour monitoring period; the relatively sparse measurement is attributable to the requirement that participants stop activities to ensure an accurate reading and the requirement, for this study, that they complete a 2-minute diary entry

immediately after each reading. Subject-level ABP measurements over the course of the day are the functional response in our FLCM.

Previous analyses of ambulatory blood pressure have mostly focused on average daytime (or nighttime) blood pressure [24, 25, 23]. However, individual ambulatory blood pressure readings are thought to depend on the specific context at the time of measurement, which changes throughout the day. Blood pressure may vary, for example, based on recent physical activity, mood, posture, location, activity type, and other factors. In the Masked Hypertension Study, these were assessed using accelerometers to quantify the intensity of activity and ecological momentary assessments, brief diary entries completed on a pre-programmed electronic diary (Palm Pilot Tungsten 3) immediately after each blood pressure reading, to quantify other variables. Physical activity and entries in the electronic diary over the course of the day are the 32 functional covariates in our proposed FLCM.

A subset of the data collected in the Masked Hypertension Study is shown in Figure 1. The left panel shows systolic blood pressure (SBP) measurements, the outcome in our application, for 40 participants taken between 10:00am and 10:00pm. Clear differences between participants exist, but there is also substantial within-subject variability over the course of the day. Remaining panels in Figure 1 show, for the same 40 participants, the measurements of frustration level, physical activity in the minute preceding ABP measurement, and a binary variable indicating whether the participant is working (according to the participant's self-reported current activity).

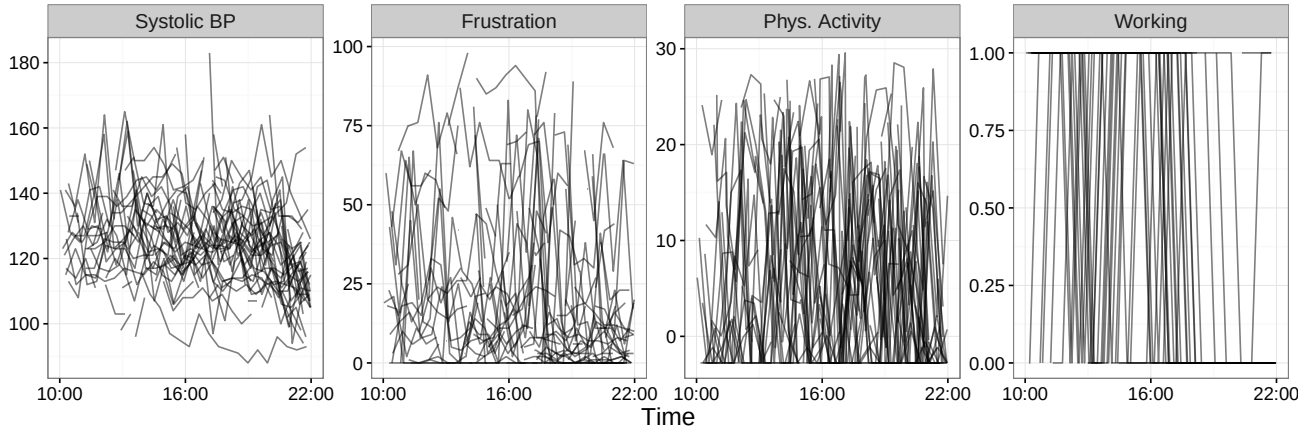

Figure 1. Observed ambulatory systolic blood pressure data for 40 participants between 10:00am and 10:00pm.

The potential for time-specific dependence of blood pressure on recent activity and other contextual variables was one factor in the design of the Masked Hypertension Study and motivates our use of the FLCM. The number of these variables requires selection methods to prevent overfitting and provide accurate predictions. It is also hoped that the results of our variable selection analysis can be used to limit the length of ecological momentary assessments (i.e. the number of questions asked following each blood pressure reading) in future studies to reduce the burden on participants.

3. Methods

Our goal is to fit the functional linear concurrent model while inducing sparsity in the estimated coefficient functions. As part of our strategy, we model residuals using a functional principal components expansion. That is, we expand the residuals $\delta_i(t)$ in (1) in terms of a shared basis to obtain

$$Y_i(t) = \sum_{k=1}^{p} X_{ik}(t)\beta_k(t) + \sum_{k=1}^{K_\Psi} c_{ik}\psi_k(t) + \epsilon_i(t). \tag{2}$$

In (2), the $\psi_k(t)$ are the functional principal component basis functions (FPCs) and the c_{ik} are the random subject-specific loadings; together, these account for correlation within residual curves. The $\epsilon_i(t)$ is thus a mean-zero uncorrelated error curve, which we will assume to have a constant variance σ^2_ϵ. The intercept $\beta_0(t)$ is omitted from (2) because both predictors $X_{ik}(t)$ and responses $Y_i(t)$ are centered prior to estimation (see Section 3.3 for details).

Model (2) is conceptual in that predictor and response curves are observed on a finite grid which is, in our case, sparse and irregular across subjects. Nonetheless, it is useful for framing our approach and emphasizing the functional nature underlying the observed data. In Section 3.1 we detail the model specification for observed data, while in Sections 3.2 and 3.3 we describe our fitting algorithm and practical considerations, respectively.

3.1. Model specification for observed data

We specify our model with sparse and irregular data in mind, although the following applies equally to curves measured on a dense grid shared across subjects. For subject i, $1 \leq i \leq I$, assume that the response and predictor curves are observed at time points $\boldsymbol{t}_i = \{t_{i1}, \ldots, t_{iJ_i}\}$. Let functions evaluated at $\boldsymbol{t}_i$ denote $J_i \times 1$ vectors of those functions on the observed time points (e.g. $Y_i(\boldsymbol{t}_i) = [Y_i(t_{i1}), \ldots, Y_i(t_{iJ_i})]^T$, $X_{ij}(\boldsymbol{t}_i) = [X_{ij}(t_{i1}), \ldots, X_{ij}(t_{iJ_i})]^T$, and $\beta_k(\boldsymbol{t}_i) = [\beta_k(t_{i1}), \ldots, \beta_k(t_{iJ_i})]^T$).

Coefficient functions $\beta_k(t)$ in (2) are expanded in terms of a fixed spline basis $\boldsymbol{\Theta}(t)$ made up of K_Θ functions $\theta_1(t), \ldots, \theta_{K_\Theta}(t)$. Let $\boldsymbol{\Theta}(\boldsymbol{t}_i)$ be the $J_i \times K_\Theta$ spline evaluation matrix on the grid $\boldsymbol{t}_i$; then $\beta_k(\boldsymbol{t}_i) = \boldsymbol{\Theta}(\boldsymbol{t}_i)\boldsymbol{b}_k^\beta$ where $\boldsymbol{b}_k^\beta$ is the vector of spline coefficients for the kth coefficient function. For the fixed effects in (2) evaluated on $\boldsymbol{t}_i$, substitution yields

$$\begin{aligned}
\sum_{k=1}^{p} X_{ik}(\boldsymbol{t}_i) \cdot \beta_k(\boldsymbol{t}_i) &= \sum_{k=1}^{p} X_{ik}(\boldsymbol{t}_i) \cdot [\boldsymbol{\Theta}(\boldsymbol{t}_i)\boldsymbol{b}_k^\beta] \\
&= \sum_{k=1}^{p} \left[(X_{ik}(\boldsymbol{t}_i) \otimes \mathbf{1}_{K_\Theta}) \cdot \boldsymbol{\Theta}(\boldsymbol{t}_i) \right] \boldsymbol{b}_k^\beta \\
&\equiv \sum_{k=1}^{p} \boldsymbol{X}_{ik}^* \boldsymbol{b}_k^\beta
\end{aligned}$$

where "$\cdot$" is the element-wise product, "$\otimes$" is the Kronecker product, and $\mathbf{1}_K$ is a length K column vector with each entry equal to 1. Defining $\boldsymbol{X}_i^* = \left[\boldsymbol{X}_{i1}^* | \ldots | \boldsymbol{X}_{ip}^*\right]$ and $\boldsymbol{b}^\beta = \left[\left(\boldsymbol{b}_1^\beta\right)^T | \ldots | \left(\boldsymbol{b}_{K_\Theta}^\beta\right)^T \right]^T$, we have that

$$\sum_{k=1}^{p} X_{ik}(\boldsymbol{t}_i) \cdot \beta_k(\boldsymbol{t}_i) = \boldsymbol{X}_i^* \boldsymbol{b}^\beta. \tag{3}$$

Lastly, we note that $\boldsymbol{\beta}(\boldsymbol{t}_i)$, the $J_i \times p$ matrix of coefficient functions evaluated on the grid $\boldsymbol{t}_i$, is given by $\boldsymbol{\Theta}(\boldsymbol{t}_i)\boldsymbol{B}^\beta$ where $\boldsymbol{B}^\beta$ is the matrix of spline coefficients with kth column equal to $\boldsymbol{b}_k^\beta$.

Variable selection is induced through our prior specification on the $\boldsymbol{b}_k^\beta$. For each k, our spike-and-slab prior is

$$\boldsymbol{b}_k^\beta \sim \mathrm{N}\left[0, (1-\gamma_k)v_0\sigma_\epsilon^2 \boldsymbol{I}_{K_\Theta} + \gamma_k v_1 \sigma_\epsilon^2 \boldsymbol{I}_{K_\Theta}\right]$$

where γ_k is the latent binary inclusion indicator, v_0 and v_1 control the spike and slab variances, respectively, and $\boldsymbol{I}_K$ is a $K \times K$ identity matrix. Each γ_k is assigned a Bernoulli prior with probability p_γ; in turn, p_γ is assigned a Beta$[.5, .5]$ prior. Choices for v_0 and v_1 are discussed in Section 3.3.

FPC basis functions $\psi_k(t)$ in (2), like coefficient functions $\beta_k(t)$, are expanded in terms of a fixed spline basis; for notational convenience we use the same basis $\boldsymbol{\Theta}(t)$. We let $\psi_k(\boldsymbol{t}_i) = \boldsymbol{\Theta}(\boldsymbol{t}_i)\boldsymbol{b}_k^\psi$ where $\boldsymbol{b}_k^\psi$ is the vector of spline coefficients for the kth FPC basis function. Substituting this expansion for the FPC basis functions in (2) yields

$$\begin{aligned}
\sum_{k=1}^{K_\Psi} c_{ik}\psi_k(\boldsymbol{t}_i) &= \sum_{k=1}^{K_\Psi} c_{ik}[\boldsymbol{\Theta}(\boldsymbol{t}_i)\boldsymbol{b}_k^\psi] \\
&= [\boldsymbol{c}_i^T \otimes \boldsymbol{\Theta}(\boldsymbol{t}_i)]\boldsymbol{b}^\psi
\end{aligned} \tag{4}$$

where $\boldsymbol{c}_i = [c_{i1}, \ldots, c_{iK_\Psi}]^T$ and $\boldsymbol{b}^\psi = \left[(\boldsymbol{b}_1^\psi)^T | \ldots | (\boldsymbol{b}_{K_\Psi}^\psi)^T\right]^T$. In contrast to the situation for fixed effects, in which the predictor functions are known and spline coefficients must be estimated, here both $\boldsymbol{c}_i$ and $\boldsymbol{b}^\psi$ are unknown. A useful expression equivalent to that in (4) emphasizing that the $\boldsymbol{c}_i$ must be estimated is $[\boldsymbol{\Theta}(\boldsymbol{t}_i)\boldsymbol{B}^\psi]\boldsymbol{c}_i$, where $\boldsymbol{B}^\psi$ is the matrix of spline coefficients with kth column equal to $\boldsymbol{b}_k^\psi$.

The prior specification for parameters in the residual FPC decomposition draws on probabilistic and Bayesian PCA for non-functional data [26, 27]. In particular, scores $\boldsymbol{c}_i$ are given independent standard Normal priors $\boldsymbol{c}_i \sim \mathrm{N}\left[0, \boldsymbol{I}_{K_\Psi}\right]$ and, for each k, spline coefficients are given Normal priors $\boldsymbol{b}_k^\psi \sim \mathrm{N}\left[0, \sigma_{\psi_k}^2 \boldsymbol{I}_{K_\Theta}\right]$. Variances $\sigma_{\psi_k}^2$ and σ_ϵ^2 are modeled using

uninformative inverse gamma priors. Thus, our complete model specification is

$$\begin{aligned}
Y_i(\boldsymbol{t}_i) &\sim \text{N}\left[\boldsymbol{X}_i^*\boldsymbol{b}^\beta + [\boldsymbol{c}_i^T \otimes \boldsymbol{\Theta}(\boldsymbol{t}_i)]\boldsymbol{b}^\psi, \sigma_\epsilon^2 \boldsymbol{I}_{J_i}\right] \text{ for subjects } 1 \le i \le I; \\
\boldsymbol{b}_k^\beta &\sim \text{N}\left[0, (1-\gamma_k)v_0\sigma_\epsilon^2 \boldsymbol{I}_{K_\Theta} + \gamma_k v_1 \sigma_\epsilon^2 \boldsymbol{I}_{K_\Theta}\right] \text{ for } 1 \le k \le p \\
\gamma_k &\sim \text{Bernoulli}[p_\gamma] \text{ for } 1 \le k \le p \\
p_\gamma &\sim \text{Beta}[.5, .5] \\
\boldsymbol{b}_k^\psi &\sim \text{N}\left[0, \sigma_{\psi_k}^2 \boldsymbol{I}_{K_\Theta}\right] \text{ for } 1 \le k \le K_\Psi \\
\boldsymbol{c}_i &\sim \text{N}\left[0, \boldsymbol{I}_{K_\Psi}\right] \text{ for subjects } 1 \le i \le I \\
\sigma_{\psi_k}^2 &\sim \text{IG}\left[.5, .5\right] \text{ for } k = 1 \ldots K_\Psi \\
\sigma_\epsilon^2 &\sim \text{IG}\left[.5, .5\right]. \qquad (5)
\end{aligned}$$

3.2. *Variational algorithm*

To estimate the parameters in (5) we implement a variational Bayes algorithm as an alternative to sampler-based estimation. This approach is motivated by three factors: i the general accuracy of variational algorithms, at least for posterior modes, to estimate model parameters; ii the computational efficiency of variational algorithms; and iii the emphasis on prediction accuracy rather than on posterior inference, making accurate estimation sufficient for many cases. We also argue that traditional Markov chain Monte Carlo implemented in a Gibbs sampler, for example, would perform poorly for the proposed model: the mean structure and FPC decomposition of the residual curves in the proposed model are highly correlated, which is expected to lead to poor mixing. Some possible directions for future work to address these concerns are noted in Section 6.

For a detailed introduction to variational Bayes, see [28] and [29, Chapter 10]; here be briefly review the general methodology. Variational Bayes methods seek an approximation $q(\boldsymbol{\phi})$ to the full posterior $p(\boldsymbol{\phi}|\boldsymbol{y})$ for parameter vector $\boldsymbol{\phi}$ and data $\boldsymbol{y}$. To this end, q is restricted to a class of functions that are more tractable than the full posterior distribution. In our algorithm, we assume that for some partition $\{\phi_1, \cdots, \phi_L\}$ of $\boldsymbol{\phi}$ the posterior distribution factors such that $q(\boldsymbol{\phi}) = \prod_{l=1}^L q_l(\phi_l)$, and each q_l is a parametric density function. Within this class, we wish to choose the element q^* that minimizes the Kullback-Leibler distance from $p(\boldsymbol{\phi}|\boldsymbol{y})$. It can be shown that the optimal component densities q_l^* are given by

$$q_l^*(\phi_l) \quad \propto \quad \exp\left\{E_{\phi_{-l}} \log\left[p(\boldsymbol{y}, \boldsymbol{\phi})\right]\right\} \propto \exp\left\{E_{\phi_{-l}} \log\left[p(\phi_l|\text{rest})\right]\right\}$$

where rest $\equiv \{\boldsymbol{y}, \phi_1, \ldots, \phi_{l-1}, \phi_{l+1}, \ldots, \phi_L\}$ is the collection of the observed data and all parameters other than ϕ_l, and $p(\phi_l|\text{rest})$ is the full conditional distribution of ϕ_l.

The parameters of the q_l^* are updated iteratively and deterministically using expressions involving the current estimates of the remaining parameters and the data. This suggests an algorithm of the following form:

1. Initialize parameter estimates;
2. Update the mean and variance of $\boldsymbol{b}^\beta$;
3. For $1 \le k \le p$, update the mean of γ_k; update p_γ;
4. Update the mean and variance of $\boldsymbol{b}^\psi$;
5. For $1 \le i \le I$, update the mean and variance of $\boldsymbol{c}_i$;
6. For $1 \le k \le p$, update $\sigma_{\psi_k}^2$; update σ_ϵ^2;
7. Repeat steps 2-6 until convergence

The complete variational algorithm is given in the Supplementary Materials. We initialize parameters so that $\boldsymbol{b}^\beta = 0$, $\gamma_k = 1$ for all k, $\boldsymbol{b}^\psi = 0$, and c_{ik} is drawn from a standard Normal for all i and k. Variational algorithms are often monitored for convergence through a lower bound on the divergence between the true posterior and the variational approximation; we have found that monitoring convergence through estimated coefficients is also reasonable, as is using a fixed (relatively large) number of iterations.

3.3. *Practical concerns*

3.3.1. Centering and scaling Similar to non-functional variable selection settings, we recommend that predictor functions are centered to have mean zero and scaled to have variance one over their domain. Additionally, we center the response function so that the intercept can be omitted from (2).

From observed predictors $\tilde{X}_{ik}(t)$ we construct normalized predictors $X_{ik}(t) = \frac{\tilde{X}_{ik}(t) - \mu_k(t)}{\sqrt{\text{Var}_k(t)}}$ where $\mu_k(t)$ and $\text{Var}_k(t)$ are the mean and variance curves. Estimates of the mean and variance curves can be obtained from a two-stage procedure.

First, one estimates $\hat{\mu}_k(t)$ by smoothing observed predictors $\tilde{X}_{ik}(t)$ over t; second, one estimates $\widehat{\text{Var}}_k(t)$ by smoothing $\left[\tilde{X}_{ik}(t) - \hat{\mu}_k(t)\right]^2$. Several smoothing approaches could be used in these steps, including popular spline-based approaches. However, due in part to the sparsity and irregularity of the grid on which our motivating data are observed, spline-based smoothing results in negative estimates of $\widehat{\text{Var}}_k(t)$ for some values of k and t. In our application we instead use K nearest neighbors to estimate the mean and variance. Specifically, for time t we define the neighborhood $S(t)$ as the set $\{t_{ij} : i \in 1, \ldots, I, j \in 1, \ldots, J_i\}$ of K observed grid points nearest to t in Euclidean distance. Based on this neighborhood we construct estimates

$$\hat{\mu}_k(t) = \frac{1}{K} \sum_{t_{ij} \in S(t)} \tilde{X}_{ik}(t_{ij}) \text{ and } \widehat{\text{Var}}_k(t) = \frac{1}{K} \sum_{t_{ij} \in S(t)} \left[\tilde{X}_{ik}(t_{ij}) - \hat{\mu}_k(t)\right]^2.$$

In these sums, the index i in the predictor $\tilde{X}_{ik}$ corresponds to the subject i in the argument t_{ij}. This normalization approach is also used to obtain centered and scaled responses $Y_i(t)$ from observed responses $\tilde{Y}_i(t)$.

An alternative is to center at the subject level by, for example, defining a subject-specific neighborhood $S_i(t) = \{t_{ij} : j \in 1, \ldots, J_i\}$. Doing so would model the effect of changes in predictors relative to each subject's mean and variance. How to best implement this for sparse data is not obvious, and our efforts in this direction for our data did not yield results meaningfully different from those using the above approach.

3.3.2. Tuning parameters Our variational algorithm estimates coefficient functions for fixed values of v_0 and v_1, which control the spike and slab variances. We fix v_1 to be relatively large to avoid unnecessary penalization on selected coefficients. Allowing v_0 to have a small but positive value, rather than setting the spike variance to zero, helps to absorb negligible nonzero coefficient functions into the spike distribution. We treat v_0 as a tuning parameter and choose its value via five-fold cross validation in which subjects, rather than observations within subjects, are partitioned into training and validation sets.

The values of K_Θ and K_Ψ, the number of spline basis functions in our expansions and the number of FPCs in our decomposition of the residual curves, are also fixed outside of our variational algorithm. Because we don't impose smoothness through penalization, the value of K_Θ acts as an implicit tuning parameter. The effect of the choice of K_Θ should be carefully evaluated in applications, potentially through the use of cross validation. For K_Ψ, a useful strategy in practice is to initially use a large value to allow the examination of a scree plot, and then to select a smaller value that nonetheless explains a large proportion of the residual variance. Alternatively, it may be possible to use an information criterion to inform the choice of both K_Θ and K_Ψ, as proposed in [30] for a model involving a FPC decomposition with parameters estimated by an EM algorithm.

3.3.3. Rotation of FPCs The Bayesian FPC analysis, like similar approaches, does not explicitly introduce orthogonality in the estimated basis functions. Although the estimation approach is valid even without these constraints, the interpretation of the FPC analysis is more straightforward when FPCs are orthogonal. For this reason, we suggest that estimated basis functions be rotated into an equivalent orthonormal space at the convergence of the variational algorithm.

3.3.4. Implementation Lastly, we note that all methods are implemented in the `R` package `vbvs.concurrent`, installable from GitHub using the following code:

```
install.packages("devtools")
devtools::install_github("jeff-goldsmith/vbvs.concurrent")
```

This package contains code for fitting the FLCM with the proposed method, as well as an estimation algorithm that uses variational Bayes without variable selection and an OLS-based approach. Helper functions for implementing five-fold cross validation to select v_0, extracting coefficients, and making predictions are also included. Interactive graphics for these FLCM methods can be produced using the `refund.shiny` package [31, 32]; this package can be installed from GitHub or CRAN.

4. Simulations

We now conduct simulations to explore the properties of the methods developed in Section 3. For reference, we also compare to a variational Bayes algorithm analogous to our proposal except that it omits the variable selection step; that is, we fit the FLCM using only a "slab" prior on regression coefficients.

We generate response curves according to the model

$$Y_i(t) = \sum_{k=1}^{50} X_{ik}(t)\beta_k(t) + \sum_{k=1}^{2} c_{ik}\psi_k(t) + \epsilon_i(t). \tag{6}$$

with $t \in [0,1]$. We set $\beta_1(t) = \sin(2\pi t)$, $\beta_2(t) = \cos(2\pi t)$, $\beta_3(t) = 1$, and $\beta_k(t) = 0$ for $k = 4, \ldots 50$. We additionally fix $\psi_1(t) = \sin(2\pi t)$ and $\psi_2(t) = \cos(2\pi t)$. Principal component scores are generated using $c_{i1} \sim \mathrm{N}[0, 9]$ and $c_{i2} \sim \mathrm{N}[0, 1]$, and uncorrelated errors $\epsilon_i(t)$ are generated from a standard Normal distribution. The number of observations J_i for subject i is drawn uniformly from $\{10, \ldots, 15\}$ and the grid $\boldsymbol{t}_i$ on which the response $Y_i(t)$ and the predictors $X_{ik}(t)$ are observed is drawn from a Unif$[0,1]$. Observed covariate functions $X_{ik}(\boldsymbol{t}_i)$ are generated independently from a standard Normal distribution. Under this design, the signal-to-noise ratio varies over t from a minimum of 0.2 to a maximum of 1.

Two hundred and fifty datasets are generated according to the above design for each $I \in \{25, 50, 100, 200\}$. For each dataset, we fit the FLCM using variational Bayes with and without variable selection (abbreviated VBVS and VB, respectively). Throughout this Section, we fix $K_\Theta = 5$ and $K_\Psi = 3$. The latter choice is made to demonstrate performance under modest misspecification of the FPC model for residual curves; additional results $K_\Psi = 0$ are described briefly in this Section and presented more fully in an Appendix. The slab variance v_1 is set to 100, and for VBVS the spike variance v_0 is chosen by five-fold cross validation. Code implementing our simulations is available in the Supplementary Material.

The VBVS and VS methods are first evaluated for their estimation accuracy. The top rows Figure 2 shows estimated coefficients across all 250 replicates using the sample size $I = 50$, separately for VBVS and VB. We present estimates of nonzero coefficient functions $\beta_1(t)$, $\beta_2(t)$, and $\beta_3(t)$, and zero coefficient function $\beta_4(t)$ as a representative of results for $\beta_k(t)$ for $k = 4, \ldots 50$. The bottom row of Figure 2 shows boxplots of the integrated squared error ISE $= \int_0^1 [\beta_k(t) - \hat{\beta}_k(t)]^2\, dt$ for the same coefficient functions and for each sample size. Boxplots are shown on the log scale.

The results in Figure 2 indicate that VBVS outperforms VB in terms of estimation accuracy for nonzero coefficient functions and, especially, for zero coefficient functions. The difference between approaches decreases as sample size increases, which is expected, but persists even when $I = 200$. For $I = 50$, some large outlying ISEs are observed for the VBVS method. These correspond to over-penalized estimates, readily identifiable in the top panel, that occur when the coefficient function is not selected as by the VBVS algorithm. Such outliers occur more frequently for $\beta_1(t)$ than other coefficients because scores c_{i1} for the first FPC, which has the same form as $\beta_1(t)$, have relatively high variance. Nonetheless, the estimated coefficient functions appear largely unbiased for the true functions shown in red, and the frequency of outlying ISEs drops for larger sample sizes.

Figure 3 includes panels showing, for each sample size, the proportion of replicates in which coefficient functions were selected by VBVS; the overall model size for VBVS; and the computation time for VBVS and VB. True positive rates are roughly .4 for a sample size of 25, but increase quickly as sample size grows; meanwhile, true negative rates begin around .6 for the smallest sample size and also increase quickly. The impact of these results on model size are apparent in the middle panel of Figure 3. When $I = 25$ the model size distribution is bimodal, with no or all variables being included with some frequency, but model size quickly converges to a narrow distribution around the true value.

Computation times, shown in the right panel of Figure 3, are generally low but include some outliers and increase roughly quadratically with sample size. Computation times are shown for single model fits, and are comparable comparing the VBVS and VB methods. This does not, however, reflect the increased burden for VBVS required by cross validation to choose tuning parameters. The extent of this burden depends on the nature of the cross validation procedure used (e.g. on the size of the grid considered for v_0). Given the generally low computation times for a single model fit, the computational burden for VBVS even with cross validation is manageable.

Additional simulation results included in an appendix show analogous results for $K_\Psi = 0$. Briefly, for this choice VBVS also outperforms VB in the ways described above. However, the results for VBVS with $K_\Psi = 0$ are generally worse than for VBVS with $K_\Psi = 3$, which suggests that accounting for residual correlation when it exists improves the accuracy of variable selection and estimation.

5. Application results

In this section we undertake analyses of the Masked Hypertension Study, introduced in Section 2. For the present analysis, we have restricted the dataset to the 563 participants with at least 10 ambulatory blood pressure measurements taken approximately every half-hour over the period 10:00am to 10:00pm in a single day. While both systolic and diastolic measurements were taken, we focus here on systolic BP as the outcome of interest. Additional variables were collected using ecological momentary assessments and accelerometers. In particular, we have measures of mood (excited; frustrated; happy; tired; angry; anxious), alcohol and caffeine consumption, activity type (relaxing; working;

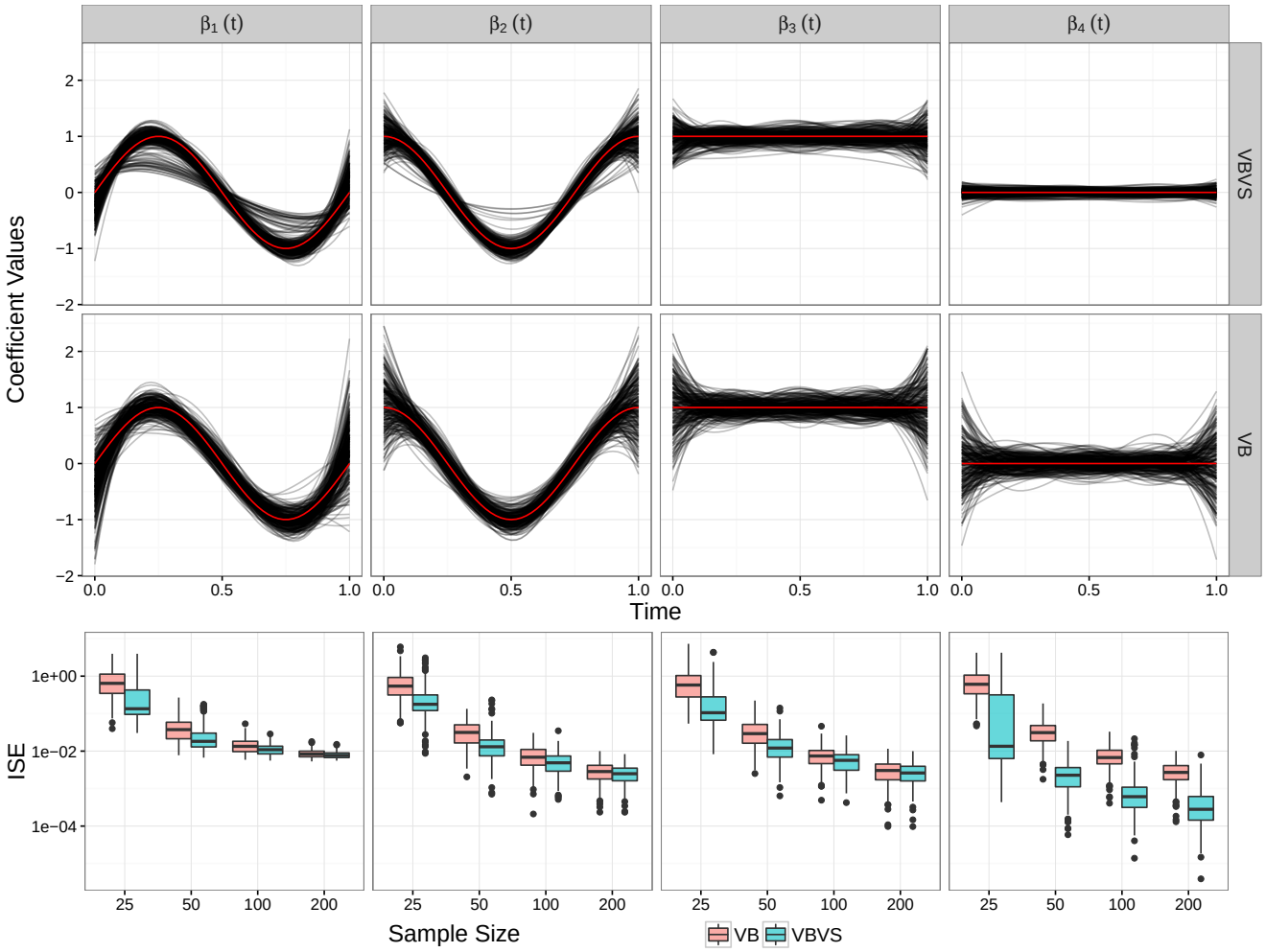

Figure 2. The top rows show estimated coefficient functions in black using the VBVS and VB method with $I = 50$; true coefficient functions are overlaid in red. Results for $\beta_k(t)$ with $k \in \{1, 2, 3, 4\}$ appear in separate columns. The bottom rows show boxplots of integrated squared errors for the same coefficient functions, separately for VBVS and VB and for each sample size included in the simulation.

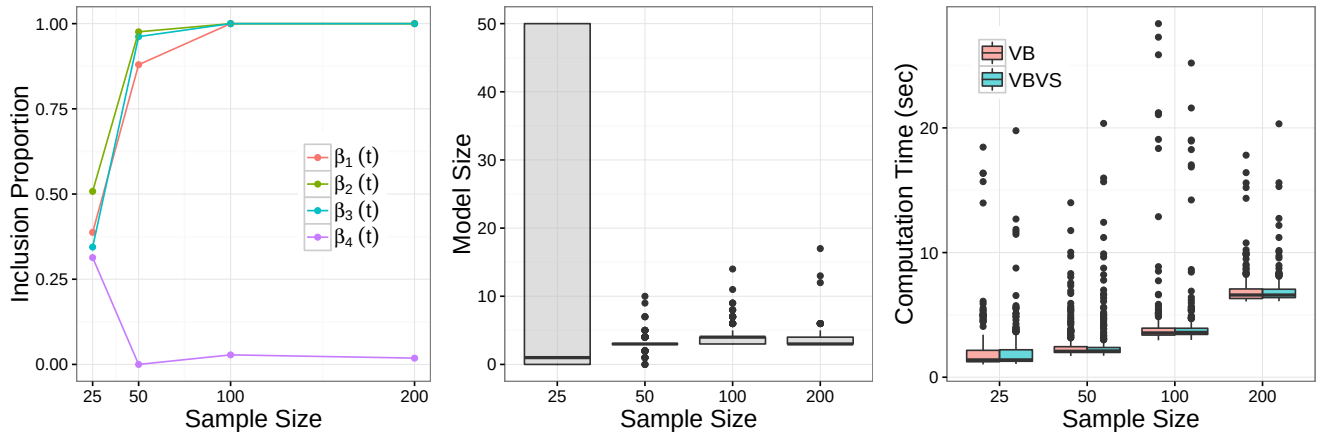

Figure 3. The left panel shows the proportion of replicates in which coefficient functions were selected by the VBVS approach. The middle panel shows the total model size for the VBVS approach. The right panel shows the computation time, in seconds, for a single model fit of the VBVS and VB approaches.

doing chores; commuting; exercising; having a meal; none of these), current exertion level, current pain rating, posture (sitting / reclining; moving; standing), location (at work; at home; other), and physical activity in the first, second, third, fourth, and fifth minute preceding the blood pressure measurement. Because completion of the ecological momentary

assessment requires the participant to be awake, and because the accelerometer was removed during sleep, we restrict our analysis to observations between 10:00am and 10:00pm. In addition to the time-varying covariate functions, we include baseline systolic and diastolic blood pressures (the mean of nine readings taken by a trained technician over three visits to the clinic), as well as age, BMI, and sex; these are included as covariate functions that are constant over time. In total, there are 32 predictor functions in our analysis.

We apply the methods described in Section 3 to these motivating data, and have two main objectives. First, in Section 5.1, we focus on the ability of the proposed methods to improve prediction accuracy, for various sample sizes, in comparison to fitting the FLCM without variable selection. Second, in Section 5.2, we perform an analysis of the complete dataset and interpret the results. Throughout, we set $v_1 = 100$ and use five-fold cross validation to choose v_0; the choice of K_Θ and K_ψ is discussed separately in Sections 5.1 and 5.2.

5.1. *Predictive performance*

To examine the effect of variable selection on prediction accuracy in the FCLM in the Masked Hypertension Study, we randomly sample participants without replacement from the complete data to construct non-overlapping training and validation sets. We then fit model 1 to the training data using the methods described in Section 3; we also fit a variational approach without variable selection for comparison. Fitted values $\hat{Y}_i(t_{ij})$ for subjects i in the validation set are computed from the regression coefficients estimated from the training data and compared to observed values. We record which variables are selected, the estimated coefficients $\hat{\beta}_k(t)$, and the average prediction mean squared error $APMSE = \frac{1}{\sum_i J_i} \sum_i \sum_{j \in J_i} \left(Y_i(t_{ij}) - \hat{Y}_i(t_{ij}) \right)^2$ for subjects i in the validation set.

Training sets are chosen to include $I \in \{25, 50, 100, 200\}$ subjects. Validation sets include 350 subjects for each training sample size so that the APMSE is based on the same number of subjects in all cases. The use of relatively small training sets in this Section is intended to emphasize the importance of variable selection on prediction accuracy when analyzing real data, and augments the simulations in Section 4. Considering a sequence of training set sizes facilitates a discussion of fitting the FLCM with and without variable selection as sample size grows. An application to the complete data follows in Section 5.2.

The above process, beginning with the training and validation sampling and ending with the computation of the APMSE for the validation set, is repeated 250 times for each training set size. For this analysis we set $K_\Theta = 5$ and $K_\psi = 2$, although similar results are obtained for other choices of these values.

Figure 4 shows the frequency with which each possible variable is selected across the 250 replications when the training set consists of 100 subjects. As expected, clinical SBP was selected as relevant in almost all replications. Surprisingly, however, other variables are never or only rarely selected. This may suggest that the variation observed in Figure 1 is either measurement error or biological variation not directly attributable to measured covariates; it may also suggest the effects of concurrently observed covariates are too small to be included in our predictive model. Additional discussion of this result appears in Section 6.

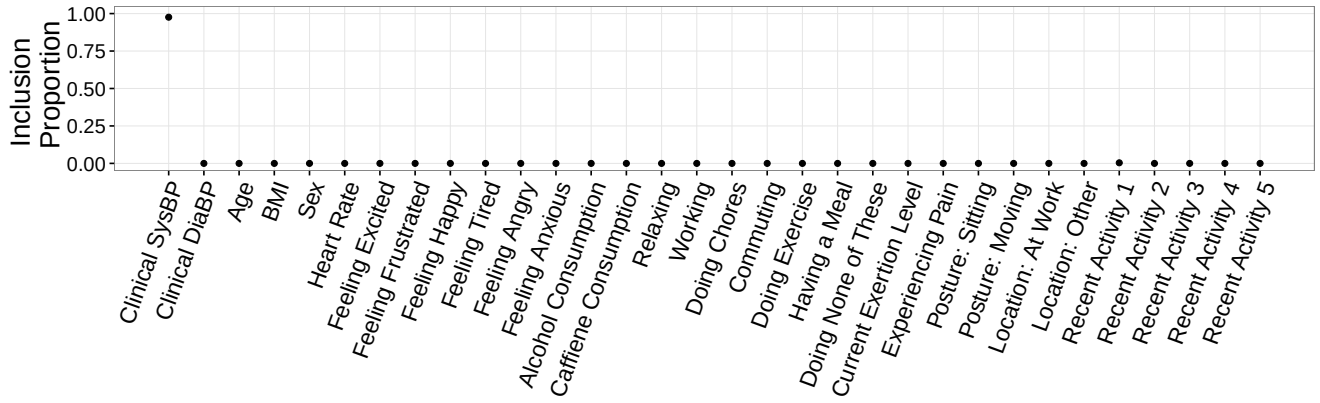

Figure 4. Proportion of the 250 replicates with training set size 100 in which potential covariates were selected by the proposed method.

Figure 5 shows the APMSE for each replication separately for each training set size. In addition to the variational Bayes algorithms with and without variable selection (again abbreviated VBVS and VB, respectively), we include two other models for comparison. First, we include an "intercept only" model to confirm that the FLCM improves over using the

in-sample mean for these data. Second, based on the results in Figure 4, the "Clinical SBP Only" model fits the FLCM with only clinical SBP as a predictor; because there is a single predictor, this model is fit with no variable selection.

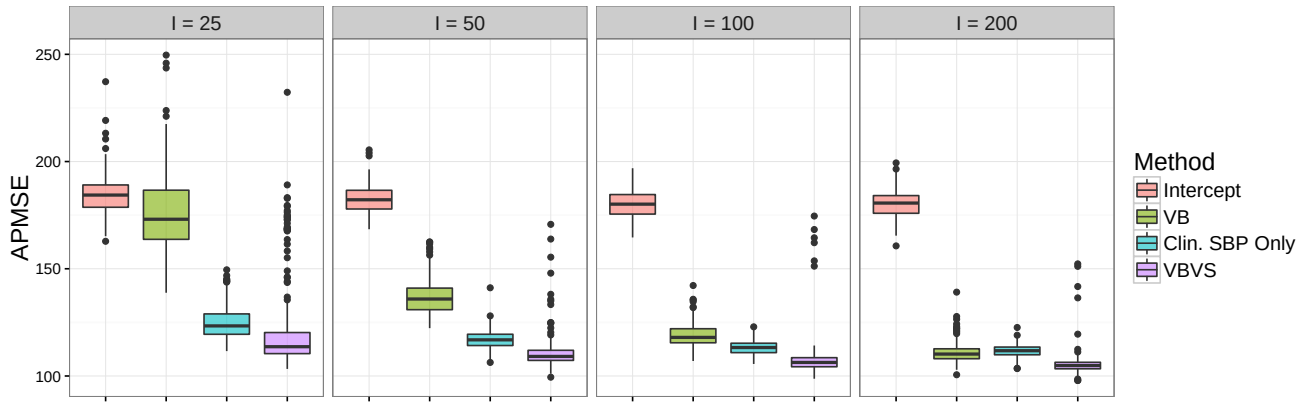

Figure 5. Boxplots of the average prediction mean squared error (APMSE) for each of four approaches: intercept only, variational Bayes without variable selection (VB), variational Bayes with variable selection (VBVS), and a model with only clinical SBP as a predictor. Panels show results for varying training set sizes I.

The VBVS approach outperforms the VB approach in terms of APMSE for all training set sizes, emphasizing the detrimental effect of including unimportant covariates on prediction accuracy. Comparing APMSE across panels indicates that the VBVS approach has reasonable performance even for small samples sizes. Perhaps surprisingly, the VBVS approach modestly but consistently outperforms the Clinical SBP Only model even though, in most cases, the VBVS method only selects clinical SBP as a predictor. We speculate that this improvement stems from the heavy penalization applied to non-selected covariates in the VBVS approach, in place of their complete omission from the Clinical SBP Only model. Lastly, the outlying APMSE values for the VBVS method are generally the result of cases in which no variables are selected, and occur less frequently for larger sample sizes.

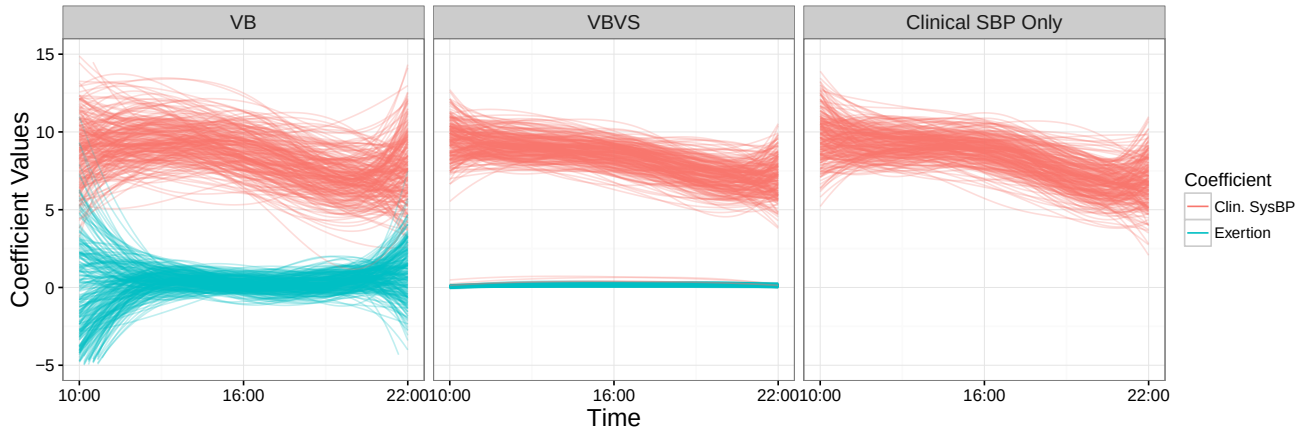

Figure 6. Estimated coefficient functions for clinical SBP and current exertion level for VB, VBVS, and Clinical SBP Only models.

The panels of Figure 6 show the estimated coefficient functions for clinical SBP and current exertion level for the VB, VBVS, and Clinical SBP Only model fits across the 250 replications with $I = 100$. Recall that predictor functions are centered and scaled, meaning that coefficient functions represent the change in systolic ABP (mmHg) for a one standard deviation increase in the predictor. Estimated coefficient functions for the effect of SBP are more variable for VB than for VBVS, illustrating the added noise in estimates when unimportant predictors are included; meanwhile, the estimates for the VBVS and Clinical SBP Only models are nearly indistinguishable. Estimates of the coefficient for exertion are near but not exactly equal to zero for the VBVS approach in each replication, in contrast to the noisy estimates of the VB method. Across replications, the coefficient for clinical SBP is positive over the course of the day; it declines very gradually until the late afternoon, at which time the coefficient function decreases more rapidly. This dip may reflect a weakening of the association between clinic SBP, which is taken in the daytime, and ambulatory SBP readings in the evening.

5.2. *Full data analysis*

Our study of prediction accuracy in Section 5.1 explored the importance of variable selection in the context of the FLCM for a range of training set sizes. We now present results from the analysis of our subset ($I = 563$) of the Masked Hypertension Study data using the proposed VBVS approach; once again, we compare to the VB approach to provide context.

The values of K_Θ and K_ψ act as tuning parameters, and we select them using the techniques described in Section 3.3.2. To choose K_ψ, we first set $K_\Theta = 12$ and fit the FLCM using VBVS with $K_\psi = 10$. Results of this model fit indicate that the first four functional principal components suffice to explain 99% of the residual variance, and we set $K_\psi = 4$ for the remainder of our analyses. We then choose $K_\Theta \in \{6, 9, 12\}$ and v_0 using cross validation, and find that $K_\Theta = 6$ slightly outperforms the larger options.

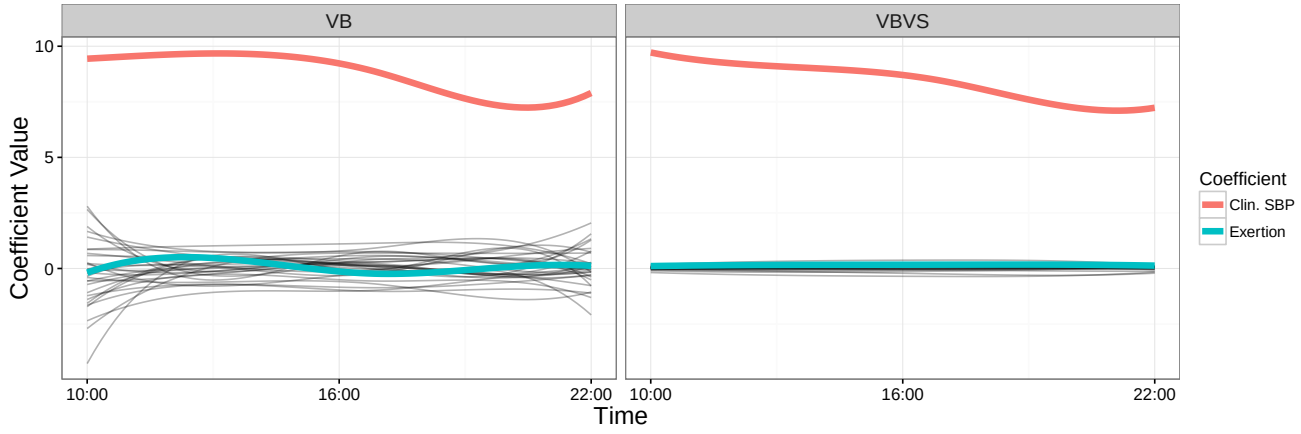

Figure 7. Coefficient functions for each covariate, estimated using variational Bayes without variable selection ("VB", left panel) and with variable selection ("VBVS", right panel), in the full data analysis. Coefficients for clinic SBP and Exertion are colored to correspond to Figure 5; other coefficients are shown in black.

Coefficient functions estimated using VB and VBVS are shown in the left and right panels of Figure 7. In both panels, we highlight the coefficient functions for clinical SBP and for current exertion level to aid in comparing these to estimates shown in Figure 6. In the VBVS analysis, only clinical SBP is selected as an important predictor; the estimate and its interpretation is similar to that in Figure 6. Remaining coefficient functions are shrunk toward zero over their full domain. There are two clear groups of estimated coefficient functions in the VB analysis: most are clustered near zero, while the coefficient for clinical SBP alone is far from zero. The comparison of these results with those for VBVS indicates the effect of the "spike" prior on the magnitude and variability of coefficient functions for non-selected covariates.

Our proposed approach jointly models the coefficient functions using a spike-and-slab prior and residual correlation using an FPC decomposition. While our emphasis in this manuscript is on the former, the results of the FPC analysis are useful in understanding the structure of the variability that is unexplained by covariate functions. To illustrate these results for the Masked Hypertension Study, Figure 8 shows $\pm\sqrt{\lambda_k}\psi_k(t)$ for $k = 1$ and 2 in the left and right panels, respectively. The first FPC is largely a mean shift, indicating that some participants have uniformly higher or lower ambulatory SBP than expected based on their covariates. Those who are higher show a masked hypertension effect relative to their expected value, meaning that ambulatory BP is higher than expected based on clinic BP, while those who are lower show a relative white-coat effect in which the clinic setting induces higher BP than is observed in ambulatory settings. The second FPC highlights distinct time-of-day effects: participants with high scores for FPC 2 have greater than expected ambulatory SBP in the morning and afternoon and lower than expected SBP in the evening. Thus the second FPC may differentiate those who have more work stress than home stress and vice versa. The first and second FPCs explain 85% and 12% percent of the smooth residual variance, respectively.

6. Concluding remarks

We have developed methods for variable selection in the functional linear concurrent model; although we are motivated by a specific application, the methods are general and broadly useful. There are several key contributions to the functional data analysis literature: the use of Bayesian variable selection methods, in contrast to the currently-used group lasso or similar approaches; jointly modeling the regression coefficients with sparsity and the residual covariance structure;

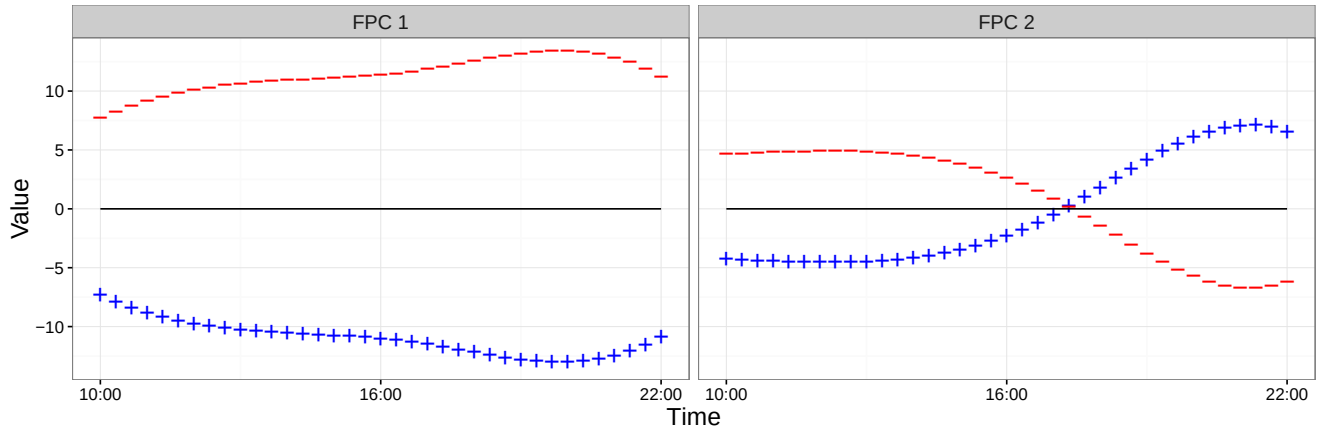

Figure 8. Illustration of the FPC basis functions. The panels are $\pm\sqrt{\lambda_k}\psi_k(t)$ for $k = 1$ (left) and $k = 2$ (right).

support for both densely- and sparsely-sampled functional responses; and the availability of robust, user-friendly code in an `R` package.

In our application, it is perhaps surprising that so few of the collected variables were selected as predictors of systolic blood pressure: these variables were recorded specifically because such associations were anticipated. There are several implications of this finding that are worth considering. First, we acknowledge that measurement of relevant variables may be imperfect. As with all surveys, the ecological momentary assessment relies on self-report; for variables like happiness or anger, there may be individual-level differences in the ratings of these emotions. Multiple parallel data streams also requires careful alignment of recording devices; it is possible that the accelerometer and ambulatory blood pressure monitor were mis-aligned for some participants, although steps were taken to prevent this. Second, we note that the subject-level patterns uncovered by the FPCA decomposition of the residuals indicates subject-level effects that are unexplained by baseline variables or covariate functions, and there remains substantial unexplained variation in ambulatory blood pressure readings. Whether this is natural biological variation, measurement error, or the result of other concurrent covariates is unclear. Lastly, we emphasize that although only clinical SBP was selected as a predictor of ambulatory SBP, the proposed approach was necessary to make this finding. We also acknowledge that it is possible that some EMA variables are statistically significantly associated with ambulatory BP, but the effects are sufficiently small to be omitted in our prediction-focused analysis.

Several future directions for statistical work are possible. Adaptations of our methods for other functional regression settings, such as function-on-scalar and scalar-on-function, are possible. In function-on-scalar regression specifically, our joint modeling approach may have important benefits over the "pre-whitening" approach of [6]: extensions to multilevel functional responses appears more straightforward in the Bayesian setting. Including an explicit penalty on the smoothness of coefficient functions for selected covariates would be similar to estimation strategies without variable selection; however, how to include both smoothness and sparsity constraints in our current framework is not clear. The development of a sampling-based Bayesian approach is also reasonable, although it should be undertaken with some care. In [33], the authors pose a function-on-scalar regression with an FPCA decomposition of residual curves and use a Hamiltonian Monte Carlo sampler [34], implemented in `Stan` [35]; a similar implementation may be appropriate here. Alternative approaches to modeling residual correlation can also be considered; in [9] a Wishart prior models this correlation, and Gibbs sampling worked well. Whether a similar approach is suitable for higher-dimensional curves and sparse data is unclear. Lastly, a mixture of the concurrent functional model and the historic functional model [20] could be relevant for including the effect of recent physical activity on current blood pressure, although based on our findings for the Masked Hypertension Study such methodological developments may not be necessary.

7. Acknowledgments

The first author's research was supported in part by Award R01HL123407 from the National Heart, Lung, and Blood Institute, by Award R21EB018917 from the National Institute of Biomedical Imaging and Bioengineering, and by Award R01NS097423-01 from the National Institute of Neurological Disorders and Stroke. The second author's research, including the collection of all data for the Masked Hypertension Study, was supported by Award P01HL047540 from the National Heart, Lung, and Blood Institute.

References

1. Ramsay JO, Silverman BW. *Functional Data Analysis*. New York: Springer, 2005.
2. Fan J, Zhang W. Statistical methods with varying coefficient models. *Statistics and its Interface* 2008; **1**:179.
3. Şentürk D, Nguyen DV. Varying coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica* 2011; **21**:1831–1856.
4. Gertheiss J, Goldsmith J, Crainiceanu C, Greven S. Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics* 2013; **14**:447–461.
5. Barber RF, Reimherr M, Schill T. The function-on-scalar lasso with applications to longitudinal gwas. *Under Review* 2016; .
6. Chen Y, Goldsmith J, Ogden T. Variable selection in function-on-scalar regression. *Stat* 2016; .
7. Goldsmith J, Wand MP, Crainiceanu CM. Functional regression via variational Bayes. *Electronic Journal of Statistics* 2011; **5**:572–602.
8. McLean MW, Scheipl F, Hooker G, Greven S, Ruppert D. Bayesian functional generalized additive models for sparsely observed covariates. *Under Review* 2013; .
9. Goldsmith J, Kitago T. Assessing systematic effects of stroke on motor control using hierarchical function-on-scalar regression. *Journal of the Royal Statistical Society: Series C* 2016; **65**:215–236.
10. van der Linde A. Variational Bayesian Functional PCA. *Computational Statistics and Data Analysis* 2008; **53**:517–533.
11. van der Linde A. A Bayesian latent variable approach to functional principal components analysis with binary and count. *Advances in Statistical Analysis* 2009; **93**:307–333.
12. Earls C, Hooker G. Combining functional data registration and factor analysis. *Journal of Computational and Graphical Statistics* 2016; .
13. Ročková V, George EI. EMVS: The EM approach to bayesian variable selection. *Journal of the American Statistical Association* 2014; **109**:828–846.
14. George EI, McCulloch RE. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 1993; **88**:881–889.
15. George EI, McCulloch RE. Approaches for Bayesian variable selection. *Statistica Sinica* 1997; **7**:339–373.
16. Ramsay JO, Dalzell CJ. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B* 1991; **53**:539–572.
17. Brockhaus S, Scheipl F, Hothorn T, Greven S. The functional linear array model. *Statistical Modelling* 2015; **15**:279–300.
18. Ivanescu AE, Staicu AM, Scheipl F, Greven S. Penalized function-on-function regression. *Computational Statistics* 2015; **30**:539–568.
19. Scheipl F, Staicu AM, Greven S. Functional additive mixed models. *Journal of Computational and Graphical Statistics* 2015; **24**:477–501.
20. Malfait N, Ramsay JO. The historical functional linear model. *Canadian Journal of Statistics* 2003; **31**:115–128.
21. Morris JS. Functional regression analysis. *Annual Review of Statistics and Its Application* 2015; **2**(1).
22. Pickering TG, Davidson K, Gerin W, Schwartz JE. Masked hypertension. *Hypertension* 2002; **40**:795–796.
23. Schwartz JE, Burg MM, Shimbo D, Broderick JE, Stone AA, Ishikawa J, Sloan R, Yurgel T, Grossman S, Pickering TG. Clinic blood pressure underestimates ambulatory blood pressure in untreated employer-based us population: Results from the masked hypertension study. *Circulation* In press; .
24. Shimbo D, Newman JD, Schwartz JE. Masked hypertension and prehypertension: diagnostic overlap and interrelationships with left ventricular mass: the Masked Hypertension Study. *American journal of hypertension* 2012; **25**:664–671.
25. Abdalla M, Goldsmith J, Muntner P, Diaz KM, Reynolds K, Schwartz JE, Shimbo D. Is isolated nocturnal hypertension a reproducible phenotype? *American journal of hypertension* 2015; :hpv058.
26. Tipping ME, Bishop C. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B* 1999; **61**:611–622.
27. Bishop CM. Bayesian PCA. *Advances in Neural Information Processing Systems* 1999; :382–388.
28. Ormerod J, Wand MP. Explaining variational approximations. *The American Statistician* 2010; **64**:140–153.
29. Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
30. Huang H, Li Y, Guan Y. Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *Journal of the American Statistical Association* 2014; **109**:1412–1424.
31. Wrobel J, Goldsmith J. *refund.shiny: Interactive plotting for functional data analyses* 2015. R package version 0.2.
32. Wrobel J, Park SY, Staicu AM, Goldsmith J. Interactive graphics for functional data analyses. *Stat* 2016; **5**:88–101.
33. Goldsmith J, Zipunnikov V, Schrack J. Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics* 2015; **71**:344–353.
34. Hoffman MD, Gelman A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *arXiv preprint arXiv:1111.4246* 2011; .
35. Stan Development Team. *Stan Modeling Language User's Guide and Reference Manual, Version 2.10.0* 2016. URL http://mc-stan.org/.