Bayesian Functional Covariance Regression

John Shamshoian[1], Damla Şentürk[1], Shafali Jeste[2], and Donatello Telesca[1],*

[1]Department of Biostatistics, University of California, Los Angeles, California 90095, U.S.A.

[2] Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles,

California 90095, U.S.A.

**email:* donatello.telesca@ucla.edu

SUMMARY:

Function on scalar regression relates continuous outcomes to scalar predictors in expectation and finds many uses in applied domains such as medicine, public health, biomechanics, and environmental science. Equally important as the mean of the conditional mean function is the covariance operator, which has has relatively underdeveloped regression methods with some notable exceptions. In this article, we develop a Bayesian functional covariance regression model to provide full posterior inference to address this disparity. We also develop novel low-dimensional summaries which graphically indicate the degree of covariate-dependent heteroschedasticity. We apply our model to two motivating case studies to make novel insights. The first case study evaluates a functional biomarker of neural development in children with autism spectrum disorder and the second case study explores the relationship between sleep patterns, age, and hypertension.

KEY WORDS: Autism spectrum disorder; Covariance regression; Functional data analysis; Peak alpha frequency; Power spectral density; Sleep heart health study.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Functional data analysis (FDA) is a broad collection of theory and methods designed to analyze conceptually infinite-dimensional data with smoothness assumptions. Functional principal components analysis (FPCA) (Wang et al., 2016) is ubiquitous in FDA, with applications ranging from biomedicine, gene expression data, environmental science, and many other scientific disciplines. Although the FPCA literature is growing rapidly to accommodate data arising from complex designs (Baladandayuthapani et al., 2008; Staicu et al., 2010; Greven et al., 2010; Zipunnikov et al., 2011; Park and Staicu, 2015; Scheffler et al., 2020), only a handful of methods allow incorporating exogenous covariate information in the covariance estimation or resulting FPCA. Cardot (2007) developed a nonparametric technique based on kernel smoothers to smooth covariance surfaces over a covariate in dense functional data settings. Jiang et al. (2010) extended the kernel smoothing technique to accommodate sparse functional data by employing conditional expectation Yao et al. (2005). Xiao et al. (2015) use sandwich smoothing (Xiao et al., 2013) to extract age-adjusted patterns of variation in a repeated measures circadian rhythm study. Scheffler et al. (2020) extend the concept of weak-separability to incorporate exogenous covariate information in a multidimensional functional setting.

In this paper we develop a Bayesian procedure to adjust covariances based on exogenous covariate information. We focus on the independent functional response case for simplicity although extensions to more general dependency structures are possible. Our proposed method has three main contributions to the existing literature of covariate-adjusted covariance modelling:

- **Straightforward inference.** The Bayesian paradigm offers a straightforward mechanism for uncertainty quantification. The alternative methods listed above must rely on the bootstrap. The literature surrounding bootstrapping in the presence of functional data is vague

on handling many tuning parameters. Although there are some theoretical guarantees of the bootstrap in functional data settings (Cuevas et al., 2006), the literature is underdeveloped for models incorporating covariate information into the covariance function. In this paper we focus on uncertainty quantification for covariate-adjusted mean functions, covariate-adjusted covariance functions, and subject-specific latent trajectories.

- **Model-based.** Existing methods are based on highly flexible nonparametric techniques. However, users employing these methods must stratify their inferences by discrete covariates such as sex. In contrast, the proposed method accommodates discrete covariates in a single analysis due to the additive structure, retaining as much information as possible in the data without creating stratified analyses.

- **Low-dimensional summaries.** We develop a low dimensional summary which aids in describing the amount of heterogeneity attributed to a particular covariate. It's not clear how to quantify this information using existing methods.

The proposed method is closely related to the notion of regularized covariance estimation. As an early reference for regularized covariance estimation, Flury (1984) developed a method to estimate a common set of principal components across k groups. This concept was generalized by Franks and Hoff (2019), who use partial pooling to estimate a set of principal components across k groups. Fox and Dunson (2015) developed a Bayesian nonparametric method for estimating a time-varying covariance matrix through factor matrix products, where the loading of the factor matrix depends on predictors. However it's unclear how to extend this method in the context of independent functional observations or include discrete covariates such as group indicators. Moreover, this particular method requires high dimensional Gibbs updates, which limits its computational feasibility. In contrast, the multivariate covariance regression model of Hoff and Niu (2012) can incorporate continuous as well as discrete covariates and only requires low dimensional Gibbs updates. However, this flexibility

is at the cost of some linear assumptions, making the method not as flexible as Fox and Dunson (2015) in some aspects. See Li et al. (2014) and Quintero and Lesaffre (2017) for extensions of this model to the multivariate multilevel case. The model presented in Section 2 can be seen as a functional extension of Hoff and Niu (2012) with some added flexibility, and we will highlight the similarities and differences as we go along.

We apply the proposed method to two case studies. The first case study involves electroencephalography (EEG) brain signals recorded on children with autism spectrum disorder (ASD). ASD is a complex neurodevelopmental disorder that affects about 1 in 54 children. ASD is characterized by difficulty in communication, restricted repetitive behaviors, and stereotypical behavior. Low functioning children may have limited behavioral repertoire, necessitating specialized assessment methods. Electroencephalography (EEG) provides a direct measure of postsynaptic brain activity and does not rely on behavioral output from young children with ASD, making EEG based biomarkers appealling for diagnosis, prognosis, and intervention purposes (Jeste, Kirkham, Senturk, Hasenstab, Sugar, Kupelian, Baker, Sanders, Shimizu, Norona, et al., 2015). In this study 59 heterogenous children with ASD and 38 age matched typically developing (TD) children had resting-state EEG signals recorded (Dickinson, DiStefano, Senturk, and Jeste, 2018). This study focused on oscillations in the alpha rhythm, which play a role in neural coordination and communication between distributed brain regions. We describe how random patterns of variation differ between children with ASD and their TD peers, providing novel insights into their neurodevelopmental differences.

The second case study analyzes EEG data collected as part of an in-home polysomnography for the Sleep Heart Health Study (SHHS). This large study was designed to identify factors for sleep-disordered breathing, such as age, blood pressure, or sleep patterns. In this article we simply quantify sleep as a functional response and explore how heteroschedasticity depends

on age and hypertension status. Specifically, we quantify how random variations of delta

spectral power change by age and hypertension status.

The rest of this paper is organized as follows: Section 2 gives the generating model for

functional data, Section 3 lists prior choices and discusses the reasoning behind them, Section

4 briefly discusses computations involved for posterior calculations, Section 5 gives a thorough

simulation study assessing errors and coverage properties, Section 6 showcases the model on

the two motivating case studies, and Section 7 concludes with a brief discussion. The sampling

algorithm and additional simulation details are given in the supplement.

2. Model

In this section we present the model associating patterns of variability and time-stable

covariates. Let $y_i(t)$ denote the outcome for subject i at point $t \in \mathcal{T}$ for some real compact

interval $\mathcal{T}$. Let $\boldsymbol{x} = (x_1, \ldots, x_{d_1})^\top$ denote a d_1-dimensional time-stable covariate for subject

i, with the dependence on i removed for ease of presentation. The k-dimensional data-

generating model is

$$y_i(t) = \mu(t, \boldsymbol{x}) + r_i(t, \boldsymbol{x}) + \epsilon_i(t) \tag{1}$$

$$r_i(t, \boldsymbol{x}) = \sum_{j=1}^{k} \psi_j(t, \boldsymbol{x}) \eta_{ij} \tag{2}$$

$$\eta_{ij} \sim N(0, 1), \quad \epsilon_i(t) \sim N(0, \varphi^2) \tag{3}$$

where $\mu(t, \boldsymbol{x})$ is the conditional mean, $\psi_j(t, \boldsymbol{x})$ form conditional latent functional bases,

$\eta_{ij} \sim N(0, 1)$ are subject-specific scores, and $\epsilon_i(t) \sim N(0, \varphi^2)$ represents measurement error.

Using equations (1, 2, 3) the conditional covariance function $c(t, t', \boldsymbol{x})$ is

$$c(t, t', \boldsymbol{x}) = \sum_{j=1}^{k} \psi_j(t, \boldsymbol{x}) \psi_j(t', \boldsymbol{x}) \tag{4}$$

We acknowledge that various approaches exist for specifying the form of $\mu(\cdot)$ and $\psi_j(\cdot)$

including local polynomial smoothers (Fan and Gijbels, 1996), kernel smoothers (Ferraty and

Vieu, 2006), Gaussian process methods (Yang et al., 2016; Fox and Dunson, 2015), and spline procedures (Ramsay, 2004). Lending toward conceptually straight-forward prior assumptions, we build $\mu(\cdot)$ and $\psi_j(\cdot)$ as linear combinations of spline bases. Borrowing notation from Scheipl et al. (2015), $\mu(t, \boldsymbol{x})$ can be written as

$$\mu(t, \boldsymbol{x}) = \sum_{r=1}^{R} f_r(t, \boldsymbol{x}_r) \tag{5}$$

where the set $\cup\{\boldsymbol{x}_r\}_{r=1}^{R} = \boldsymbol{x}$. This grouping framework leads to flexible specification of basis expansions. For example, when $\boldsymbol{x}_r$ is a single scalar covariate $f_r(t, x_r)$ could be a functional linear effect $x_r f(t)$ or a smooth effect $f(t, x_r)$. If $\boldsymbol{x}_r = (x_{r_1}, x_{r_2})$ is a vector of covariates, $f_r(t, \boldsymbol{x}_r)$ could be written as $f(t, x_{r_1}, x_{r_2})$, $x_{r_1} f(t, x_{r_2})$, or $x_{r_1} x_{r_2} f(t)$. These terms are approximated by a set of basis functions with corresponding priors to encourage smooth effects. For the general case, we maintain that $\boldsymbol{x}_r$ could be a vector with basis $\boldsymbol{b}^r(\boldsymbol{x}_r) = (b_1^r(\boldsymbol{x}_r), \ldots, b_{p_r}^r(\boldsymbol{x}_r))$. Then

$$f_r(t, x_r) = \boldsymbol{b}(t)^\top \beta_r \, \boldsymbol{b}^r(\boldsymbol{x}_r) \tag{6}$$

$$\mu(t, \boldsymbol{x}) = \boldsymbol{b}(t)^\top \beta \boldsymbol{b}^x(\boldsymbol{x}_i) \tag{7}$$

where $\beta = (\beta_1 \,|\, \cdots \,|\, \beta_R)$ and $\tilde{\boldsymbol{X}}(\boldsymbol{x}) = (\boldsymbol{b}^1(\boldsymbol{x}_r) \,|\, \cdots \,|\, \boldsymbol{b}^R(\boldsymbol{x}_r))^\top$. Keeping track of dimensions, β is a $p \times r(d_1)$ coefficient matrix and $\boldsymbol{b}^x(\boldsymbol{x}_i)$ is a $r(d_1) \times 1$ vector, where $r(d_1) = \sum_{r=1}^{R} p_r$. We place another additive model on $\psi_j(t, \boldsymbol{x})$. Namely,

$$\psi_j(t, \boldsymbol{x}) = \sum_{r=1}^{R} l_{jr}(t, \boldsymbol{x}_r) \tag{8}$$

where l_{jr} are smooth effects in terms of $\boldsymbol{x}_r$ for the jth conditional latent function basis function. Similar to $f_r(t, \boldsymbol{x}_r)$ and $\mu(t, \boldsymbol{x})$, $l_{jr}(t, \boldsymbol{x}_r)$ and $\psi_j(t, \boldsymbol{x})$ can also be written as products of matrices and vectors:

$$l_{jr}(t, x_r) = \boldsymbol{b}(t)^\top \Lambda_{jr} \, \boldsymbol{b}^r(\boldsymbol{x}_r)$$

$$\psi_j(t, \boldsymbol{x}) = \boldsymbol{b}(t)^\top \Lambda_j \boldsymbol{b}^x(\boldsymbol{x}_i)$$

where Λ_{jr} is a $p \times p_r$ loading matrix and $\Lambda_j = (\Lambda_{j1} \,|\, \cdots \,|\, \Lambda_{jR})$. The additivity on $\psi_j(t, \boldsymbol{x})$

implies that the covariance function 4 is

$$c(t, t', \boldsymbol{x}) = \sum_{j=1}^{k} \left(\sum_{r=1}^{R} \sum_{r'=1}^{R} l_{jr}(t, \boldsymbol{x}_r) l_{jr'}(t', \boldsymbol{x}_{r'}) \right)$$

This convolution structure makes it difficult to define low dimensional summaries of covariate influence on the covariance function. Instead, we propose a low dimensional summary which quantifies the impact of a covariate on the l_{jr} functions directly. Let

$$g_r(t, \boldsymbol{x}_r) = \sum_{j=1}^{k} l_{jr}(t, \boldsymbol{x}_r)^2 \tag{9}$$

summarize the effect of $\boldsymbol{x}_r$ across ψ_j, $j = 1, \ldots, k$. If the impact of $\boldsymbol{x}_r$ on ψ_j, $j = 1, \ldots, k$ is negligable, then $g_r(t, \boldsymbol{x}_r)$ will be near zero. Consequently if $g_r(t, \boldsymbol{x}_r)$ is near zero, the $c(t, t', \boldsymbol{x})$ will not be sensitive to changes in $\boldsymbol{x}_r$. To the best of our knowledge, this is the first attempt at quantifying the impact of covariates on a covariance function which is only possible through the additivity assumption on $\psi_j(t, \boldsymbol{x})$ from equation 8.

Unlike previous work on functional covariance regression (Cardot, 2007; Jiang et al., 2010), equations (1, 2, 3) specify a generative model for functional covariance regression. Complete with priors detailed in Section 3, posterior inference is evaluated through Markov-Chain Monte Carlo (MCMC). This is important because empirical methods require resampling to perform inference. However, resampling techniques for functional data come with pitfalls. Any fitted model will have smoothing bias (to regularize rapidly varying functions). A parametric bootstrap would generate data from a biased model, and subsequently cause even more bias by smoothing once again for each bootstrap replicate. Nonparametric bootstrapping causes undersmoothing due to the presence of repeated functions.

Outside of the FDA literature, Hoff and Niu (2012) developed a similar model for covariance regression with multivariate data. However, as Fox and Dunson (2015) note, their mapping from predictors to covariance assumes a parametric form, thus limiting the model's expressivity. To overcome this parametric limitation, the authors develop a factor matrix process estimate a time-varying covariance matrix characterizing influenza incidence across the

United States. This approach is flexible but each gibbs sample iteration requires a cholesky decomposition of an $n \times n$ matrix where n is the number of subjects. The basis transform approach taken here would only require a cholesky decomposition of a $p \cdot r(d_1) \times p \cdot r(d_1)$ matrix for each iteration. Therefore the basis transform approach is likely to scale better for large data sets such as the SHHS, provided $r(d_1)$ is not too large.

3. Prior Distributions

In this section we place priors on all unknown quantities of interest. We begin by placing prior on $\mu(t, \boldsymbol{x})$. As we have seen in equations 6 and 7, this amounts to placing a prior on each β_r submatrix. The rows of β_r are associated with a $p \times p$ penalty matrix K, and the columns of β_r are associated with a $p_r \times p_r$ penalty matrix K_r. These penalties are designed to encourage smoothness and can target magnitude penalization, squared derivative shrinkage, or local changes in β_r through a differencing penalty. In this paper we penalize the second order difference of β_r coefficients in both directions, but other penalties could be used as well. A prior for β_r respecting the tensor structure is constructed as follows (Wood, 2017). Let $\tilde{K} = I_{p_r \times p_r} \otimes K$ and $\tilde{K}_r = K_r \otimes I_{p \times p}$. The prior for the vectorized form of β_r is

$$\text{vec}(\beta_r) \,|\, \tau_{1xr}, \tau_{1tr} \sim$$

$$\exp\{-0.5\text{vec}(\beta_r)^\top (\tau_{1xr}\tilde{K}_r + \tau_{1tr}\tilde{K})\text{vec}(\beta_r)\}$$

where τ_{1xr}, τ_{1tr} are smoothing parameters. If $p_r = 1$ then β_r is a $p \times 1$ vector and $\tilde{K}_r = 0$. In this case the prior simplifies to

$$\beta_r \,|\tau_{1tr} \sim \exp\{-0.5\tau_{1tr}\beta_r^\top K\beta_r\}$$

This prior is improper but provided that proper priors are set for τ_{1tr}, τ_{1xr}, the posterior of β_r will be proper (Lang and Brezger, 2004). Priors for $\psi_j(t, \boldsymbol{x})$. However, $\psi_j(t, \boldsymbol{x})$ for larger j should also contribute less to the fit than earlier terms. Therefore the prior for $\psi_j(t, \boldsymbol{x})$

should encourage smoothing and shrinkage aspects. Let Λ_{rj} be the analogous component to β_r. Re-using the same penalty matrices as above, the prior for Λ_{rj} is

$$\text{vec}(\Lambda_{rj}) \mid \tau_{2jxr}, \tau_{2jtr}, \tau_{rj}^*, \phi_{rj} \sim$$

$$\exp\{-0.5\text{vec}(\Lambda_{rj})^\top (\tau_{2jxr}\tilde{K}_r + \tau_{2jtr}\tilde{K} + \tau_{rj}^*\phi_{rj})\text{vec}(\Lambda_{rj})\}$$

where τ_{2jxr}, τ_{2jtr} are smoothing parameters and ϕ_{rj}, τ_{rj}^* are shrinkage parameters. Here ϕ_{rj} is a diagonal matrix with dimension $p \cdot p_r \times p \cdot p_r$. If $p_r = 1$ (so that Λ_{rj} is a column vector), then the prior becomes

$$\Lambda_{rj} \mid \tau_{2jtr}, \tau_{rj}^*, \phi_{rj} \sim$$

$$\exp\{-0.5\Lambda_{rj}^\top (\tau_{2jtr}\tilde{K} + \phi_{rj}\tau_{rj}^*)\Lambda_{rj}\}$$

similar to the case when β_r is a column vector. In summary, these priors for $\mu(t, \boldsymbol{x})$ and $\psi_j(t, \boldsymbol{x})$ are design to reflect our assumptions of smoothness in the functional data outcomes. All smoothing parameters are given a gamma prior distribution,

$$\tau_{1xr}, \tau_{1tr} \sim \text{Gamma}(a_\tau, b_\tau)$$

$$\tau_{2jxr}, \tau_{2jtr} \sim \text{Gamma}(a_\tau, b_\tau)$$

where we use the 'rate' parameterization of the Gamma distribution (i.e., if $x \sim \text{Gamma}(a, b)$, then $\mathbb{E}[x] = a/b$). In addition we borrow the Gamma Multiplicative Process Prior (GMPP) from Bhattacharya and Dunson (2011); Montagna et al. (2012) to assign priors to τ_{rj}^* and ϕ_{rj}. Let ϕ_{irj} denote the ith diagonal element of ϕ_{rj}.

$$\phi_{irj} \sim \text{Gamma}(a_\phi, b_\phi) \tag{10}$$

$$\tau_{rj}^* = \prod_{l=1}^{r} \delta_{lr} \tag{11}$$

$$\delta_{r1} \sim \text{Gamma}(a_{r0}, 1) \tag{12}$$

$$\delta_{rl} \sim \text{Gamma}(a_{r1}, 1), \; l > 1 \tag{13}$$

$$a_{r0}, a_{r1} \sim \text{Gamma}(2, 1) \tag{14}$$

so that the τ_{rj} are stochastically increasing in j. This shrinkage is data-adaptive so that later entries of Λ_{rj} may or may not be shrunk toward zero depending on the model fit. In particular, specifying a conservative choice of k (number of latent functional factors) should not change results too much compared to setting k to some "optimal" choice.

4. Posterior inference

We discuss posterior inference for subject-specific latent trajectories, mean functions, and principal directions of variation. In this paper we use the posterior mean as a point estimate and symmetric pointwise credible intervals for uncertainty quantification. Let $\hat{\beta}_m$, $\hat{\Lambda}_{jm}$, and $\hat{\eta}_{ijm}$ be the mth posterior draw of β, Λ_j, and η_{ij}. The point estimate for the ith subject-specific latent trajectory, $\tilde{y}_i(t)$, is $\frac{1}{M} \sum_{m=1}^{M} \{\boldsymbol{b}(t)^\top \hat{\beta}_m \boldsymbol{b}^x(\boldsymbol{x}_i) + \sum_{j=1}^{k} \boldsymbol{b}(t)^\top \hat{\Lambda}_{jm} \boldsymbol{b}^x(\boldsymbol{x}_i)\}$. Similarly, the point estimate for the covariate-adjusted mean, $\mu(t, \boldsymbol{x})$, is $\frac{1}{M} \sum_{m=1}^{M} \boldsymbol{b}(t) \hat{\beta}_m \boldsymbol{b}^x(\boldsymbol{x}_i)$. Let $C_m(\boldsymbol{x})$ be the matrix given by evaluating the mth posterior draw of $c(t, t', \boldsymbol{x})$ on a subset of $\mathcal{T} \times \mathcal{T}$. Performing a spectral decomposition on $C_m(\boldsymbol{x})$ yields the mth draw of ordered orthonormal posterior principal directions of variation (eigenfunctions) $\tilde{\psi}_{jm}(t, \boldsymbol{x})$, $j = 1, \ldots, k$. However, this method can be computational expensive due to the spectral decomposition of arbitrarily high-dimensional matrix $C_m(\boldsymbol{x})$ due to the nature of functional data. To alleviate this burden, we follow Aguilera and Aguilera-Morillo (2013) and only perform a spectral decomposition on a low-dimensional matrix, $\sum_{j=1}^{k} \hat{\Lambda}_{jm} \boldsymbol{b}^x(\boldsymbol{x}_i) \boldsymbol{b}^x(\boldsymbol{x}_i)^\top \hat{\Lambda}_{jm}^\top$, so that we never actually form $C_m(\boldsymbol{x})$ in our implementation. Since $\tilde{\psi}_{jm}(t, \boldsymbol{x})$ is only identifiable up to sign change, each $\psi_{jm}(t, \boldsymbol{x})$ may be potentially multiplied by -1 to orient all eigenfunctions correctly. More details on extracting and orienting posterior eigenfunctions are provided in Web Appendix D. Once all posterior eigenfunctions are oriented correctly, a point-estimate is obtained by simply taking the posterior pointwise mean.

Credible intervals are computed in the following manner. Let $\mu_f(t)$ and $\sigma_f(t)$ be the posterior mean and standard deviation of some functional $f(\cdot)$ evaluated at t. A $(1-\alpha) \cdot 100$ credible

interval for $f(t)$ is $[\hat{f}(t) - z^*_{1-\alpha/2}\sigma_f(t), \hat{f}(t) + z^*_{1-\alpha/2}\sigma_f(t)]$ where $z^*_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of

the standard normal distribution. Pointwise intervals have nominal average coverage when

the model is correctly calibrated. However, pointwise intervals are anti-conservative when

performing inference for an entire function $f(\cdot)$. In this case we recommend computing

simultaneous credbile bands. Our implementation includes this option and we give more

details in Web Appendix B.

5. Simulations

We assess the performance of the proposed method in terms of mean, covariance, and

latent response point estimation and coverage. We generate data under two scenarios. The

first scenario has heteroscedasticity depending on a covariate x scaled within the unit

interval, and the second scenario imposes no such relationship between heteroscedasticity

and covariates. Both scenarios include a covariate-dependent mean surface. We fit data

generated in both scenarios by (1) the proposed method, adjusting for covariate-dependent

heteroscedasticity and (2) foregoing this adjustment and assuming the covariate only impacts

the mean structure. We generate data in both scenarios by simulating from the model. We

generate 300 datasets per scenario and true model parameters are kept fixed over each

dataset. We include two sample sizes $N = 100$ and $N = 400$ to numerically verify mean and

covariance point estimation convergence.

We begin by using a p-spline of dimension 10 to expand the functional arguement t and

a p-spline of dimension 5 (not including an intercept) to expand the covariate argument x.

According to notation from section 2, $\boldsymbol{b}(t) = (b_1(t), \ldots, b_{10}(t))^\top$, $\boldsymbol{b}^2(x) = (b_1^2(t), \ldots, b_5^2(t))^\top$,

and $\mu(t,x) = f_1(t) + f_2(t,x) = \boldsymbol{b}(t)\,\beta_1 + \boldsymbol{b}(t)\,\beta_2\,\boldsymbol{b}^2(x)$. The model parameters β_1 and $\mathrm{vec}(\beta_2)$

have prior mean zero and precision $\Omega_1 + \epsilon \cdot I_{10\times10}$ and $\Omega_2 + \epsilon \cdot I_{50\times50}$ respectively. The nugget

term $\epsilon \cdot I$ is used to simulate from a full-rank multivariate normal and in our experiments

we set $\epsilon = .1$. To ensure smoothness of the resulting response functions $y_i(t)$, we set $\Omega_1 = \tau_{1t1}K$ and $\Omega_2 = \tau_{1x2}(K_2 \otimes I_{10\times10}) + \tau_{1t2}(I_{6\times6} \otimes K)$, where K and K_2 are second order discrete penalty matrices associated with p-splines. We use $k = 4$ latent functional factors for the random component $r_i(t,x)$, so $r_i(t,x) = \sum_{j=1}^4 \psi_j(t,x)$. Under data-generating scenario 1, we set $\psi_j(t,x) = l_{j1}(t) + l_{j2}(t,x)$. Under data-generating scenario 2, we set $\psi_j(t,x) = l_{j1}(t)$. Expanding out the terms, $l_{j1}(t) = \boldsymbol{b}(t)^\top \Lambda_{j1}$ and $l_{j2}(t,x) = \boldsymbol{b}(t)\Lambda_{j2}\boldsymbol{b}^2(x)$. The model parameters Λ_{j1} and Λ_{j2} prior mean zero and prior precision Γ_{j1} and Γ_{j2} respectively. We set $\Gamma_{j1} = \tau_{2t1}K + \tau_{1j}^*$ and $\Gamma_{j2} = \tau_{2x2}(K_2 \otimes I_{10\times10}) + \tau_{2t2}(I_{6\times6} \otimes K) + \tau_{2j}^*$. Smoothing and shrinkage parameters are $\tau_{1t1} = \tau_{1t2} = \tau_{2t1} = \tau_{2t2} = 10$, $\tau_{1x2} = \tau_{2x2} = 100$, $\tau_{1j}^* = 2^j$, and $\tau_{2j}^* = 4^j$. Measurement error is $\varphi^2 = .20^2$ and functional data $y_i(t)$ is generated according to equation 1. The functional argument t is evaluated at 100 equally spaced locations within the unit interval and x_i is equal to $(i-1)/(N-1)$.

It will be convenient to divide data-generating scenarios and fitting procedures into four separate cases. Case 1 has data generated from scenario 1, fit with a model that adjusts for heteroscedasticity due to the covariate. Case 2 has data generated from scenario 2, fit with a model that does not adjust for heteroscedasticity due to the covariate. Case 3 has data generated from scenario 2, fit with a model that adjusts for heteroscedasticity due to the covariate. Case 4 has data generated from scenario 2, fit with a model that does not adjust for heteroscedasticity due to the covariate. We fit each dataset with with $k = 6$ latent functional factors and expand the functional argument into a p-spline basis of dimension 10. We expand the covariate argument into a p-spline basis of dimension 7. In doing so, we simply enlarge the model space and test its ability to regularize unnecessary latent dimensions and basis expansions. We place uniform priors on all smoothing parameters and a Gamma$(0.0001, 0.0001)$ prior on the inverse measurement error, φ^{-2}. We fit each model with 50,000 samples, discarding 25,000 as burn-in and keeping every 5th observation to save

computer memory. We keep track of coverage, average credible interval width, and relative

integrated squared error (RISE). The coverage performance assess a model's calibration and

goodness of fit, whereas RISE assesses a model's performance in terms of point estimation.

Let $t_1, \ldots, t_{100}$ be 100 uniformly placed points along the domain of the functional argument,

$x_1^*, \ldots, x_{10}^*$ be uniformly placed points along the interval $[0.1, 0.9]$, and $\tilde{y}_i(t)$ denote the latent

response of $y_i(t)$ uncontaminated with random error. Let $\mathbb{1}(\cdot)$ be the indicator function

that returns 1 when the argument is true and 0 otherwise. The coverage of $\mu(t, x)$ is

$\frac{1}{1000} \sum_{i=1}^{10} \sum_{m=1}^{100} \mathbb{1}\{\mu(t_m, x_i^*) \in \widehat{I}_{mi}\}$, the coverage of $c(t, t', x)$ is $\frac{1}{50500} \sum_{i=1}^{10} \sum_{m=1}^{100} \sum_{m'=1}^{m} \mathbb{1}\{c(t_m, t_{m'}, x_i^*) \in$

$\widehat{I}_{mm'i}\}$, and the coverage of $\tilde{y}_i(t)$ is $\frac{1}{N \times 100} \sum_{i=1}^{N} \sum_{m=1}^{100} \mathbb{1}\{\tilde{y}_i(t_m) \in \widehat{I}_{y,mi}\}$, where $\widehat{I}_{mi}$ is the nom-

inal 95% posterior interval about $\mu(t_m, x_i)$ and $\widehat{I}_{mm'i}$ is the nominal 95% posterior interval

about $c(t_m, t_{m'}, x_i)$, and $\widehat{I}_{y,mi}$ is the nominal 95% posterior interval about $\tilde{y}_i(t)$. The RISE of

$\mu(t, x)$ is $100 \cdot \frac{\sum_{i=1}^{10} \int_{\mathcal{T}} \{\widehat{\mu}(t, x_i^*) - \mu(t, x_i^*)\}^2 \, dt}{\sum_{i=1}^{10} \int_{\mathcal{T}} \mu(t, x_i^*)^2 \, dt}$, the RISE of $c(t, t', x)$ is $100 \cdot \frac{\sum_{i=1}^{10} \int_{\mathcal{T}} \int_{\mathcal{T}} \{\widehat{c}(t, t', x_i^*) - c(t, t', x_i^*)\}^2 \, dt \, dt'}{\int_{\mathcal{T}} \int_{\mathcal{T}} \sum_{i=1}^{10} c(t, t', x_i^*)^2 \, dt \, dt'}$,

and the RISE of $\tilde{y}_i(t)$ is $100 \cdot \frac{\sum_{i=1}^{N} \int_{\mathcal{T}} \{\widehat{y}_i(t) - \tilde{y}_i(t)\}^2 \, dt}{\sum_{i=1}^{N} \int_{\mathcal{T}} \tilde{y}_i(t)^2 \, dt}$. In practice we use the trapezoidal rule to

evaluate all integrals. We also compute average credible band width for $\mu(t, x)$, $c(t, t', x)$ and

$\tilde{y}_i(t)$. All coverage, RISE, and interval width results are reported in Table 1.

We observe at least 95% coverage for both $\mu(t, x)$ and $\tilde{y}_i(t)$ under all four cases. Adjusting

for covariate-dependent heteroscedasticity does not seem to affect the coverage rates for

$\mu(t, x)$ and $\tilde{y}_i(t)$. Coverage rates for $c(t, t', x)$ are slightly below the nominal 95% in case 1,

indicating that inference on $c(t, t', x)$ may be anti-conservative under covariate-dependent

heteroschedasticity. As expected, coverage rates are low for $c(t, t', x)$ under case 2, where the

data-generating truth has covariate-dependent heteroscedasticity but the model does not

account for this. The coverage rate gets smaller from $N = 100$ to $N = 400$ samples due to

tighter credible bands. We also see that adjusting for covariate-dependent heteroscedasticity

does not alter the width of credible intervals by a substantial amount when the data-

generating truth has no covariate-dependent heteroscedasticity.

We notice slightly greater RISE for $\mu(t, x)$ and $\tilde{y}_i(t)$ in case 2 compared to case 1, indicating that adjusting for covariate-depending heteroscedasticity can lead to efficiency gains in estimating the mean structure, even with only $N = 100$ samples and only modest covariate-dependent heteroscedasticity. Estimation error for $\mu(t, x)$ and $\tilde{y}_i(t)$ is about the same under cases 3 and 4, indicating not much evidence of efficiency loss when the proposed adjustment is not necessary. However, the covariance estimation error is larger for case 3 than case 4, which indicates that the covariate-dependent heteroscedasticity adjustment can lead to slightly larger estimation error in the covariance structure when this adjustment is not necessary.

Overall, the simulation shows that the proposed method can lead to gains in estimation and coverage (compared to a homoscedastic model) when the data has covariate-dependent heteroscedasticity. Cases 3 and 4 show that estimation error for the mean structure is robust to choice of model under a homoscedastic data-generating mechanism. However, making this adjustment under a homoschedastic data-generating mechanism leads to slightly larger estimation error in the covariance structure. We recommend exploring values of $g(t, x)$ to decide whether or not to adjust the variance by covariates.

[Table 1 about here.]

6. Case Studies

6.1 *Application to ASD study*

Peak alpha frequency (PAF), the frequency at which oscillations in the alpha range demonstrate maximal power, shows well-characterised increases with chronological age during childhood in typically developing children (Somsen et al., 1997; Dustman et al., 1999; Stroganova et al., 1999; Chiang et al., 2011; Cragg et al., 2011; Miskovic et al., 2015). PAF has been shown to index neural development in TD children (Valdés-Hernández et al., 2010; Segalowitz et al., 2010; Rodríguez-Martínez et al., 2017). However, a recent study by

Dickinson et al. (2018) found that children with ASD did not show the typical increase in PAF with age. In the present study, we investigate the relationship between diagnostic status, age, and alpha spectral density. Treating alpha spectral density as a functional response avoids complicated PAF identification procedures Dickinson et al. (2018) and retains more information as opposed to collapsing the entire alpha spectral band to a single point (Scheffler et al., 2019, 2020). In this study we wish to characterize functional mean and covariance dependence of alpha spectral density in terms of age and diagnostic status.

In the Dickinson et al. (2018) study, resting-state EEG was collected on 39 TD children and 58 children with ASD aged 2 to 12 years old. EEG signals were sampled at 500 Hz for two minutes and interpolated to a 10-20 25 channel montage (Jasper, 1958; Perrin et al., 1989). Alpha spectral density $\Omega \in [6 \text{ Hz}, 14 \text{ Hz}]$ were obtained for each electrode. To facilitate comparisons across regions and subjects, alpha spectral density was normalized to unit area. See Scheffler et al. (2019) for more details on data acquisition and pre-processing. The resulting data structure is region-referenced functional data, which induces correlated responses within subjects. Unfortunately, the proposed method can not accommodate correlated functional data. Instead, for demonstration purposes, we only examine one region at a time. We also acknowledge that Scheffler et al. (2020) already successfully applied a covariate-adjusted hybrid principal components analysis, which can accommodate correlated functional responses within subject. While the goals of Scheffler et al. (2020) are similar to the ones posed in this article, we note that this method is (1) based on a Bayesian method, which yields stochastic regularization and straight-forward inference and (2) the low dimensional summary in eq. 9 provides some insight into how the variability of normalized alpha spectral density as a function of age and diagnostic status.

We treat normalized alpha spectral density from the T8 region as the functional response and adjust the mean and covariance functions by diagnostic status, age, and a diagnostic

status by age interaction. This interaction is scientifically motivated because as previously mentioned, PAF increases with age for TD children but not so for children with ASD. For the frequency dimension, we use a p-spline basis with 12 degrees of freedom using a second order differencing penalty. We expand age via p-splines with 6 degrees of freedom, again with a second order difference penalty. We run a markov chain monte carlo algorithm for 200,000 iterations. To save on memory purposes, we only save every 20 iterations. Of the leftover 10,000 iterations, we discard the first 5,000 as burn-in and keep the subsequent 5,000 for inference.

Figure 7 plots normalized alpha spectral density over several ages of child and diagnostic group with associated 95% pointwise credible bands. The means curves drift apart as the cross-sectional age increases, which supports the notion that differences in alpha spectral band dynamics between ASD and TD children at similar ages can successfully predict diagnostic status Scheffler et al. (2019).

Furthermore, we explore age and group adjusted patterns of variability. Figure 7 displays the leading age and group adjusted eigenfunction. The general shape for a particular age is a unimodal curve peaked around 8-10 Hz. Notice that the peak of this curve tends to shift from lower to higher frequencies as age increases for the TD group, whereas the peak frequency is relatively constant over age for the ASD group. This finding was previously discovered by Scheffler et al. (2020) who estimate covariate-adjusted eigenfunctions in a similar to the proposed method but also account for the regional dimension. On the other hand, posterior summaries $g(t, \boldsymbol{x})$ in Figure 7 indicate that most variability is due to group as opposed age.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

6.2 *Application to Sleep Heart Health Study*

The Sleep Heart Health Study (SHHS) was a prospective cohort study designed to investigate obstructive sleep apnea (OSA) and other sleep-disordered breathing (SDB) as risk factors for the development of cardiovascular disease (Quan et al., 1997). Parent cohort studies and recruitment targets for these cohorts are the following: Atherosclerosis Risk in Communities Study (1,750 participants), Cardiovascular Health Study (1,350 participants), Framingham Heart Study (1,000 participants), Strong Heart Study (600 participants), New York Hypertension Cohorts (1,000 participants), and Tucson Epidemiologic Study of Airways Obstructive Diseases and the Health and Environment Study (900 participants). An unattended in-home polysomnography was completed by participants between November 1, 1995 and January 31, 1998. Between January 2001 and June 2003, a second polysomnogram was obtained in 3295 of the participants.

The American Academy of Sleep recognizes four sleep stages: stage N1 (light sleep), stage N2 (relaxation), stage N3 (slow-wave sleep), and stage R (rapid-eye movement). Importantly, N3 sleep, or slow-wave sleep (SWS) consists of high amplitude ($\geqslant 75\ \mu$V) and low-frequency $(0.5 - 4.0)$ delta waves. SWS is considered to be most restorative sleep stage and to be associated with sleep quality, sleep maintenance, and functions toward memory consolidation (Bonnet, 1987; Akerstedt et al., 1997; Walker, 2009). Mander et al. (2017) reports that advancing into the fifth decade of age comes with (1) reduced SWS time and (2) increased time spent in lighter sleep stages (N1, N2). In addition, Javaheri et al. (2018) reports that lower levels of percent SWS sleep are associated with increased odds of incident hypertension in both men and women, independent of confounders such as sleep apnea, age, and sex. Most clinical studies (including the papers referenced above) quantify sleep by classifying time-varying electrical phenoma into discrete sleep stages. Subsequently, amount of SWS is represented by a single number, which is commonly percentage of sleep time in stage N3.

However, this approach comes with several limitations (Crainiceanu et al., 2009) including low intraclass correlation coefficient, no biological basis, and loss of temporal information. In this paper follow Crainiceanu et al. (2009); Di et al. (2009) and use power spectral density analysis to quantify sleep EEG. Our present goal is to characterize age-related changes in sleep between hypertension and non-hypertension groups.

We use the discrete fast Fourier transform to analyze the sleep EEG data in the frequency domain. The entire night of sleep is broken into 30-second adjacent sleep epochs to account for temporal effects. These 30-second windows are processed using Welch's method of 50% overlapping windows (4 second windows with 2 seconds overlap), where the intervals are windowed using a tapered Tukey window. The epoch-level power spectral density estimate is the average of the windowed power spectra. We use a band-pass filter to attenuate signals less than .3 Hz and greater than 35 Hz with a .5 Hz transition width. We mask epochs where artifacts detected using the method described in Buckelmüller et al. (2006). We also mask epochs that are statistical outliers using (2, 2) Hjorth parameters Hjorth (1970). We compute delta spectral power by summing power spectra from 0.5 Hz to 4.0 Hz in .25 Hz increments on an epoch-by-epoch basis. To facilitate comparisons across participants, we compute epoch-level relative delta power spectral density (RDPSD) by dividing this power density by a summation of spectral density from .5 Hz to 35 Hz in .25 Hz increments. All filtering and power spectral density computations were performed in Luna (Purcell, 2020), which is an open-source software package for manipulating and analyzing polysomnographic readings.

We restrict our attention to EEG data collected on the first visit. We also filter participants to only include those who scored at least four hours of artifact-free signal. We analyze the first two hours of sleep, resulting in a dense grid of length 240 epochs for each participant. The 10th and 90th percentiles of age are 47 and 77 years-old, respectively. There are 2177

participants with hypertension and 3081 participants without hypertension. We use a p-spline

of dimension 24 for the marginal basis for epoch and a p-spline of dimension 8 for the marginal

basis for age. Each basis is associated with a second order differencing penalty matrix. We

also include a hypertension dummy variable and a hypertension-age interaction. Exploratory

analysis from Castro et al. (1986); Carroll et al. (2020) shows 9 principal components are

needed to explain 99% of variability (unadjusted for age and hypertension group). We fit our

model with $k = 12$ latent factors since unneeded latent factors will be set close to zero due

to the prior shrinkage process in equations 11-14. We ran the MCMC algorithm described in

the supplement for 200,000 iterations with a burn-in of 100,000, saving every 20th iteration

for memory requirements.

Figure 7 displays mean RDPSD with associated 95% point-wise credible intervals for ages

50, 60, 70, and years old for both hyertension and non-hypertension groups. Relative power

spectral density decreases in N3 (around epoch 60) with age for both groups. The seemingly

little decrease of SWS from midlife to late life is supported by Van Cauter et al. (2000), who

found that percent SWS predominantly decreases from early adulthood to midlife, with no

further decrease into late life. Their study included 149 healthy men aged 16 to 83 years.

Our methods support their finding after considering a much richer metric than percent SWS

in the context of a much larger observational study. There is considerable overlap in the

credible intervals between the hypertension and non-hypertension RDPSD, which does not

support the hypothesis that hypertension is associated with increased levels of RDPSD.

[Figure 4 about here.]

We conducted an eigen-analysis on the model-based covariance surfaces. The estimated lead-

ing eigenfunction for a 50 year old with no hypertension accounts for 31% of total variability

and is positive over the entire epoch grid, suggesting that this component represents an

overall RDPSD size construct. This eigenfunction, nor subsequent eigenfunctions vary much

over age or group. The intercept effect is of $g(t, \boldsymbol{x}_r)$ is very large compared to age, group, and interaction effects, which means that between-subject heteroscedasticity does not depend much on these covariates. RDPSD is a heteroscedastic-stable metric when accounting for age and hypertension.

[Figure 5 about here.]

[Figure 6 about here.]

7. Discussion

We developed a probabilistic model suitable for functional data with covariate-dependent heteroscedasticity. The proposed apprach is computationally feasible for moderately large datasets such as the sleep study, which has $N = 5258$ subjects, 240 timepoints, and 16 basis functions for the covariate dimension. This feasibility is owed to tractible lower dimensional gibbs sampling updates. We also engineer low dimensional summaries of covariate-dependent heteroschedasticity, which can be used in practice to graphically aid in deciding whether to use a simpler model.

REFERENCES

Aguilera, A. and Aguilera-Morillo, M. (2013). Penalized pca approaches for b-spline expansions of smooth functional data. *Applied Mathematics and Computation* **219,** 7805 – 7819.

Akerstedt, T., Hume, K., Minors, D., and Waterhouse, J. (1997). Good sleep–its timing and physiological sleep characteristics. *Journal of sleep research* **6,** 221.

Baladandayuthapani, V., Mallick, B. K., Young Hong, M., Lupton, J. R., Turner, N. D., and

Carroll, R. J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* **64,** 64–73.

Bhattacharya, A. and Dunson, D. B. (2011). Sparse bayesian infinite factor models. *Biometrika* pages 291–306.

Bonnet, M. H. (1987). Sleep restoration as a function of periodic awakening, movement, or electroencephalographic change. *Sleep* **10,** 364–373.

Buckelmüller, J., Landolt, H.-P., Stassen, H., and Achermann, P. (2006). Trait-like individual differences in the human sleep electroencephalogram. *Neuroscience* **138,** 351–356.

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

Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Mueller, H.-G., and Wang, J.-L. (2020). *fdapace: Functional Data Analysis and Empirical Dynamics.* R package version 0.5.5.

Castro, P. E., Lawton, W. H., and Sylvestre, E. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28,** 329–337.

Chiang, A., Rennie, C., Robinson, P., Van Albada, S., and Kerr, C. (2011). Age trends and sex differences of alpha rhythms including split alpha peaks. *Clinical Neurophysiology* **122,** 1505–1517.

Cragg, L., Kovacevic, N., McIntosh, A. R., Poulsen, C., Martinu, K., Leonard, G., and Paus, T. (2011). Maturation of eeg power spectra in early adolescence: a longitudinal study. *Developmental science* **14,** 935–943.

Crainiceanu, C. M., Caffo, B. S., Di, C.-Z., and Punjabi, N. M. (2009). Nonparametric signal extraction and measurement error in the analysis of electroencephalographic activity during sleep. *Journal of the American Statistical Association* **104,** 541–555.

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating

functions with functional data. *Computational statistics & data analysis* **51,** 1063–1074.

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *The annals of applied statistics* **3,** 458.

Dickinson, A., DiStefano, C., Senturk, D., and Jeste, S. S. (2018). Peak alpha frequency is a neural marker of cognitive function across the autism spectrum. *European Journal of Neuroscience* **47,** 643–651.

Dustman, R. E., Shearer, D. E., and Emmerson, R. Y. (1999). Life-span changes in eeg spectral amplitude, amplitude variability and mean frequency. *Clinical neurophysiology* **110,** 1399–1409.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice.* Springer Science & Business Media.

Flury, B. N. (1984). Common principal components in k groups. *Journal of the American Statistical Association* **79,** 892–898.

Fox, E. B. and Dunson, D. B. (2015). Bayesian nonparametric covariance regression. *The Journal of Machine Learning Research* **16,** 2501–2542.

Franks, A. M. and Hoff, P. (2019). Shared subspace models for multi-group covariance estimation. *Journal of Machine Learning Research* **20,** 1–37.

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic journal of statistics* **4,** 1022–1054.

Hjorth, B. (1970). Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology* **29,** 306–310.

Hoff, P. D. and Niu, X. (2012). A covariance regression model. *Statistica Sinica* pages 729–753.

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* **10,** 370–375.

Javaheri, S., Zhao, Y. Y., Punjabi, N. M., Quan, S. F., Gottlieb, D. J., and Redline, S. (2018). Slow-wave sleep is associated with incident hypertension: the sleep heart health study. *Sleep* **41,** zsx179.

Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., Baker, E., Sanders, A. J., Shimizu, C., Norona, A., et al. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with asd. *Developmental science* **18,** 90–105.

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

Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of computational and graphical statistics* **13,** 183–212.

Li, B., Bruyneel, L., and Lesaffre, E. (2014). A multivariate multilevel gaussian model with a mixed effects structure in the mean and covariance part. *Statistics in medicine* **33,** 1877–1899.

Mander, B. A., Winer, J. R., and Walker, M. P. (2017). Sleep and human aging. *Neuron* **94,** 19–36.

Miskovic, V., Ma, X., Chou, C.-A., Fan, M., Owens, M., Sayama, H., and Gibb, B. E. (2015). Developmental changes in spontaneous electrocortical activity and network organization from early to late childhood. *Neuroimage* **118,** 237–247.

Montagna, S., Tokdar, S. T., Neelon, B., and Dunson, D. B. (2012). Bayesian latent factor regression for functional and longitudinal data. *Biometrics* **68,** 1064–1073.

Park, S. Y. and Staicu, A.-M. (2015). Longitudinal functional data analysis. *Stat* **4,** 212–226.

Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. (1989). Spherical splines for scalp po-

tential and current density mapping. *Electroencephalography and clinical neurophysiology* **72,** 184–187.

Purcell, S. (2020). *luna: Analysis of Sleep Signal Data.* https://github.com/remnrem/luna, http://zzz.bwh.harvard.edu/luna.

Quan, S. F., Howard, B. V., Iber, C., Kiley, J. P., Nieto, F. J., O'Connor, G. T., Rapoport, D. M., Redline, S., Robbins, J., Samet, J. M., et al. (1997). The sleep heart health study: design, rationale, and methods. *Sleep* **20,** 1077–1085.

Quintero, A. and Lesaffre, E. (2017). Multilevel covariance regression with correlated random effects in the mean and variance structure. *Biometrical Journal* **59,** 1047–1066.

Ramsay, J. O. (2004). Functional data analysis. *Encyclopedia of Statistical Sciences* **4,**.

Rodríguez-Martínez, E., Ruiz-Martínez, F., Paulino, C. B., and Gómez, C. M. (2017). Frequency shift in topography of spontaneous brain rhythms from childhood to adulthood. *Cognitive neurodynamics* **11,** 23–33.

Scheffler, A. W., Dickinson, A., DiStefano, C., Jeste, S., and Şentürk, D. (2020). Covariate-adjusted hybrid principal components analysis. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 391–404. Springer.

Scheffler, A. W., Telesca, D., Sugar, C. A., Jeste, S., Dickinson, A., DiStefano, C., and Şentürk, D. (2019). Covariate-adjusted region-referenced generalized functional linear model for eeg data. *Statistics in Medicine* **38,** 5587–5602.

Scheipl, F., Staicu, A.-M., and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics* **24,** 477–501.

Segalowitz, S. J., Santesso, D. L., and Jetha, M. K. (2010). Electrophysiological changes during adolescence: a review. *Brain and cognition* **72,** 86–100.

Somsen, R. J., van't Klooster, B. J., van der Molen, M. W., van Leeuwen, H. M., and Licht,

R. (1997). Growth spurts in brain maturation during middle childhood as indexed by eeg power spectra. *Biological psychology* **44,** 187–209.

Staicu, A.-M., Crainiceanu, C. M., and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **11,** 177–194.

Stroganova, T., Orekhova, E., and Posikera, I. (1999). Eeg alpha rhythm in infants. *Clinical Neurophysiology* **110,** 997–1012.

Valdés-Hernández, P. A., Ojeda-González, A., Martínez-Montes, E., Lage-Castellanos, A., Virués-Alba, T., Valdés-Urrutia, L., and Valdes-Sosa, P. A. (2010). White matter architecture rather than cortical surface area correlates with the eeg alpha rhythm. *Neuroimage* **49,** 2328–2339.

Van Cauter, E., Leproult, R., and Plat, L. (2000). Age-related changes in slow wave sleep and rem sleep and relationship with growth hormone and cortisol levels in healthy men. *Jama* **284,** 861–868.

Walker, M. P. (2009). The role of slow wave sleep in memory processing. *Journal of Clinical Sleep Medicine* **5,** S20–S26.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3,** 257–295.

Wood, S. N. (2017). *Generalized additive models: an introduction with R.* CRC press.

Xiao, L., Huang, L., Schrack, J. A., Ferrucci, L., Zipunnikov, V., and Crainiceanu, C. M. (2015). Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics* **16,** 352–367.

Xiao, L., Li, Y., and Ruppert, D. (2013). Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology* pages 577–599.

Yang, J., Zhu, H., Choi, T., Cox, D. D., et al. (2016). Smoothing and mean–covariance

estimation of functional data with a bayesian hierarchical model. *Bayesian Analysis* **11,** 649–670.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American statistical association* **100,** 577–590.

Zipunnikov, V., Caffo, B., Yousem, D. M., Davatzikos, C., Schwartz, B. S., and Crainiceanu, C. (2011). Multilevel functional principal component analysis for high-dimensional data. *Journal of Computational and Graphical Statistics* **20,** 852–873.
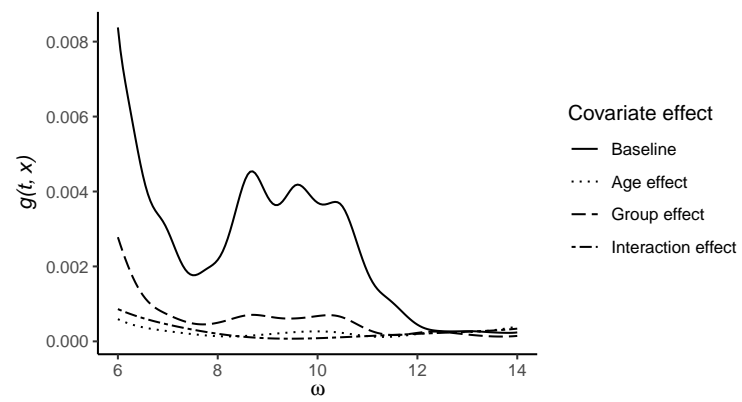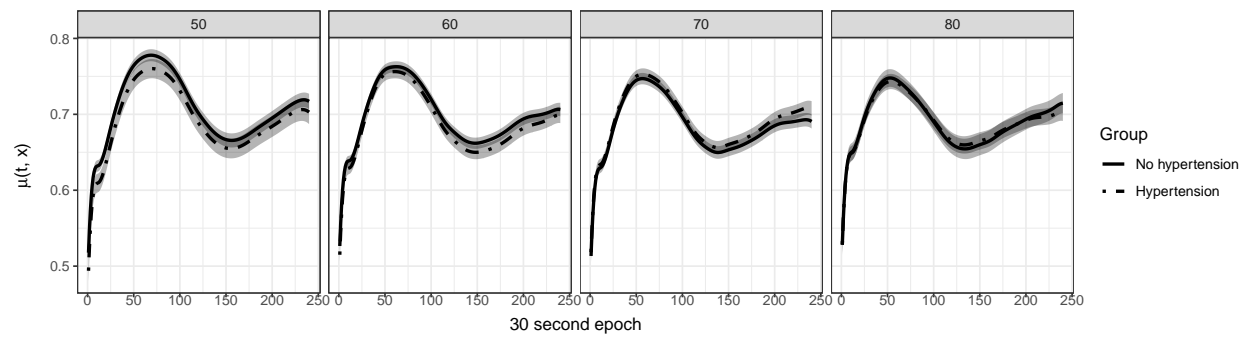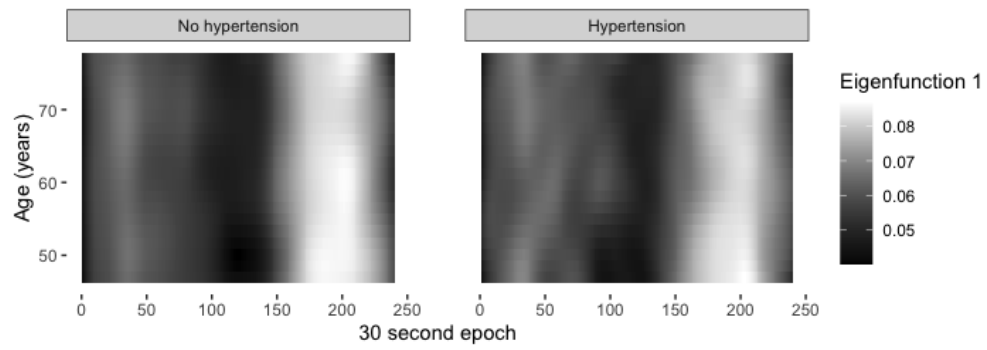
Appendix

Title of appendix
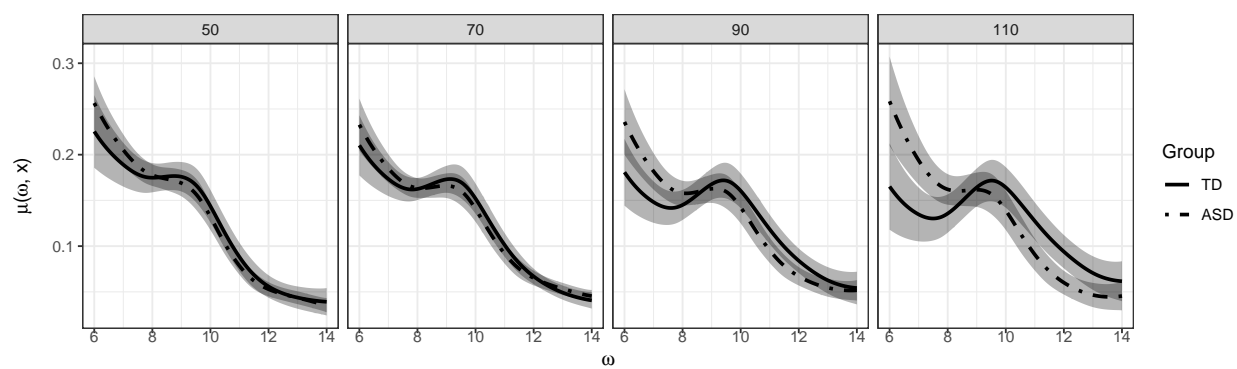
Figure 1. Posterior mean alpha spectral power for ASD and TD groups at age 50, 70, 90, and 110 months.
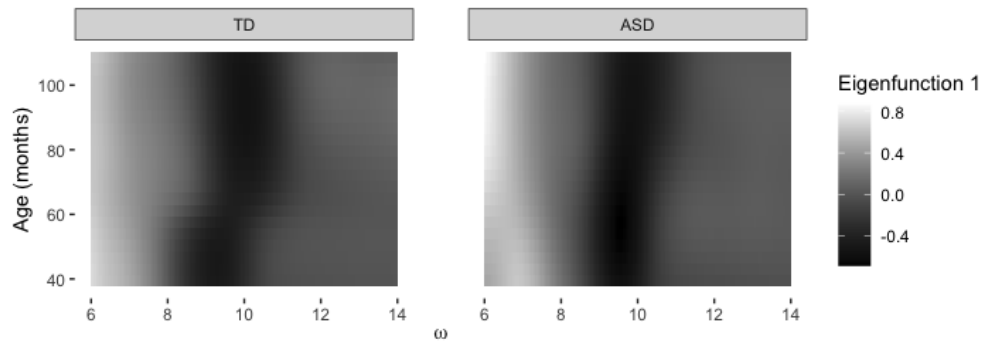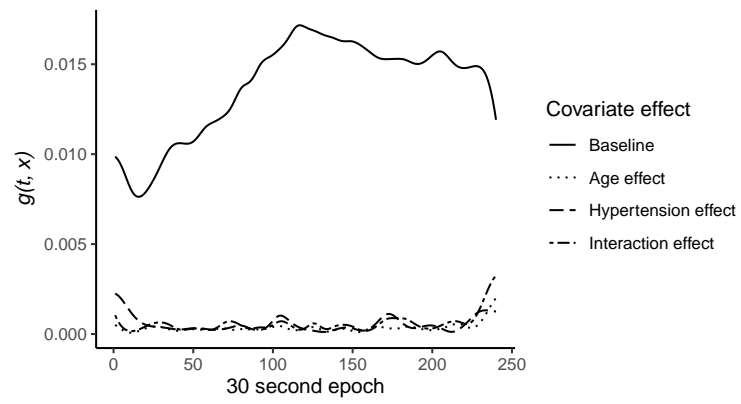
Figure 2. Posterior mean of first leading covariate-adjusted eigensurfaces.

Table 1

Simulation results over 300 data sets for each sample size and case combination. Coverage, error (RISE), and interval width are averaged over all data sets. Refer to Section 5 for details on how coverage, error, and interval width are calculated.

Sample size	Quantity	Case	Coverage	Error	Interval width
$N = 100$	$\mu(t, x)$	1	.98	.84	.30
		2	.98	.91	.32
		3	.98	.78	.30
		4	.98	.79	.30
	$c(t, t', x)$	1	.89	7.00	.14
		2	.72	9.76	.11
		3	.98	4.31	.13
		4	.97	3.29	.11
	$y_i(t)$	1	.97	.25	.17
		2	.97	.29	.19
		3	.98	.21	.17
		4	.98	.20	.16
$N = 400$	$\mu(t, x)$	1	.99	.19	.18
		2	.98	.21	.19
		3	.99	.18	.18
		4	.98	.18	.18
	$c(t, t', x)$	1	.88	3.10	.09
		2	.46	7.23	.06
		3	.98	1.01	.06
		4	.97	.84	.05
	$y_i(t)$	1	.97	.13	.16
		2	.97	.18	.18
		3	.98	.12	.16
		4	.98	.12	.16