

## Bayesian Functional Covariance Regression

**John Shamschoian<sup>1</sup>, Damla Şentürk<sup>1</sup>, Shafali Jeste<sup>2</sup>, and Donatello Telesca<sup>1,\*</sup>**

<sup>1</sup>Department of Biostatistics, University of California, Los Angeles, California 90095, U.S.A.

<sup>2</sup> Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles,  
California 90095, U.S.A.

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

**SUMMARY:** Many electroencephalography (EEG) studies aim to compare cognitive function between and within diagnostic groups. In our motivating study, resting state EEG data is collected on in a sample of 59 children with autism spectrum disorder and 38 age-matched typically developing (TD) controls. Peak alpha frequency (PAF), the frequency of maximal power within the alpha range (6 - 14 Hz), is a biomarker related to cognitive development and is known to increase with age in TD children. In this article we model alpha spectral power, rather than just the peak location. Patterns of variability of alpha spectral power between children are obscured by factors such as age. In the present work we develop methodology to estimate covariate-adjusted dependency patterns of alpha band oscillations, allowing for valid group level inference.

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

## 1. Introduction

Autism spectrum disorder (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 appealing for diagnosis, prognosis, and intervention purposes (Jeste, Kirkham, Senturk, Hasenstab, Sugar, Kupelian, Baker, Sanders, Shimizu, Norona, et al., 2015). In this article’s motivating 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 alpha waves, which play a role in neural coordination and communication between distributed brain regions.

The study investigated peak alpha frequency (PAF), the frequency at which oscillations in the alpha rhythm [6-14 Hz] achieve maximal power and is known to shift from lower to higher frequencies as TD children age. The study found that children with ASD did not show increasing PAF with age. Furthermore, PAF was strongly correlated with non-verbal cognition. In this article we take a broader view and investigate the entire alpha spectrum as opposed to collapsing this information to a single point. EEG signals were recorded using a 128-channel sensor net at 500 Hz. After post-processing the raw EEG data, each child has 25 regions of interest and alpha spectral power captured from 6 to 14 Hz with .25 Hz increments. We propose to treat this data within a functional data framework, where each spectral power curve is considered one observation. See Wang et al. (2015) for a broad review on functional data analysis (FDA).

FDA is a mature body of literature designed to handle high dimensional data with smooth-

ness assumptions. Much attention has been devoted to estimating conditional means in a mixed model framework (Guo, 2002; Morris and Carroll, 2006; Montagna et al., 2012). Less studied is the problem of estimating conditional patterns of variability. Cardot (2007) developed a method to extract conditional patterns of variability for dense functional data and Jiang et al. (2010) extended this procedure to accommodate sparse functional data. These methods lie in the Frequentist framework and rely on bootstrap to perform uncertainty quantification. In addition, both methods do not scale appropriately for more than one covariate or group indicators. We propose a Bayesian covariate-adjusted FPCA model to estimate conditional patterns of variability of alpha spectral power, conditional on age and diagnostic status. Posterior sampling defines a straightforward mechanism for completing inference at the cost of specifying priors on all unknown parameters. The proposed method can accommodate group indicators or several variables due to some linearity assumptions.

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 depending 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. In contrast, Hoff and Niu (2012) extends to allow factor loading to depend on continuous or discrete covariates. However, this flexibility requires some linear assumptions, which is in spirit similar to linear regression. 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), and we will highlight the similarities and differences as we go along.

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 4.1 focuses on inference such as credible intervals for mean functions, Section 5 gives a thorough simulation study, Section 6.2 showcases the model on the motivating EEG case study, and Section 7 concludes with a brief discussion. The sampling algorithm and simulation study information 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  $\mathbf{x} = (x_1, \dots, x_{d_1})^\top$  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

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

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

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

where  $\mu(t, \mathbf{x})$  is the conditional mean,  $\psi_j(t, \mathbf{x})$  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  $c(t, t', \mathbf{x})$  is

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

Specifying the form of  $\mu(\cdot)$  and  $\psi_j(\cdot)$  is a contentious topic and various approaches can be found in the literature 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). Each method has its own merit and we will compare our developments to some existing approaches in the context of covariance regression. Lending toward conceptually straight-forward prior specifications, we build  $\mu(\cdot)$  and  $\psi_j(\cdot)$  as linear combinations of spline bases. Borrowing notation from Scheipl et al. (2015),  $\mu(t, \mathbf{x})$  can be written as

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

where the set  $\cup\{\mathbf{x}_r\}_{r=1}^R = \mathbf{x}$ . This grouping framework leads to flexible specification of basis expansions. For example, when  $\mathbf{x}_r$  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  $\mathbf{x}_r = (x_{r1}, x_{r2})$  is a vector of covariates,  $f_r(t, \mathbf{x}_r)$  could be written as  $f_r(t, x_{r1}, x_{r2}) = f(t, x_{r1}, x_{r2}), x_{r1}f(t, x_{r2})$ , or  $x_{r1}x_{r2}f(t)$ . These terms are approximated by a set of basis functions with corresponding priors to encourage smooth effects.

The  $f_r(t, \mathbf{x}_r)$  terms can be represented by products of matrices and vectors. For example, let  $\sum_{j=1}^p \sum_{m=1}^{p_r} b_j(t) b_m^r(x_r) \beta_{rjm}$  be a tensor spline expansion for  $f(t, x_r)$  where  $\{b_j\}_{j=1}^p$  is a marginal basis expansion in  $t$  and  $\{b_m^r\}_{m=1}^{p_r}$  is a marginal basis expansion in  $x_r$ . Let  $\mathbf{b}(t) = (b_1(t), \dots, b_p(t))^\top$ ,  $\mathbf{b}^r(x_r) = (b_1^r(x_r), \dots, b_{p_r}^r(x_r))^\top$ , and arrange the coefficients into a  $p \times p_r$  matrix  $\beta_r$ . For the general case, we maintain that  $\mathbf{x}_r$  could be a vector with basis  $\mathbf{b}^r(\mathbf{x}_r) = (b_1^r(\mathbf{x}_r), \dots, b_{p_r}^r(\mathbf{x}_r))$ . Then

$$f_r(t, x_r) = \mathbf{b}(t)^\top \beta_r \mathbf{b}^r(x_r) \quad (6)$$

$$\mu(t, \mathbf{x}) = \mathbf{b}(t)^\top \beta \tilde{\mathbf{X}}(\mathbf{x}) \quad (7)$$

where  $\beta = (\beta_1 | \dots | \beta_R)$  and  $\tilde{\mathbf{X}}(\mathbf{x}) = (\mathbf{b}^1(\mathbf{x}_1) | \dots | \mathbf{b}^R(\mathbf{x}_R))^\top$ . Keeping track of dimensions,  $\beta$  is a  $p \times r(d_1)$  coefficient matrix and  $\tilde{\mathbf{X}}(\mathbf{x})$  is a  $r(d_1) \times 1$  vector, where  $r(d_1) = \sum_{r=1}^R p_r$ .

We place another additive model on  $\psi_j(t, \mathbf{x})$ . Namely,

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

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

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

$$\psi_j(t, \mathbf{x}) = \mathbf{b}(t)^\top \Lambda_j \tilde{\mathbf{X}}(\mathbf{x})$$

where  $\Lambda_{jr}$  is a  $p \times p_r$  loading matrix and  $\Lambda_j = (\Lambda_{j1} | \dots | \Lambda_{jR})$ . The additivity on  $\psi_j(t, \mathbf{x})$  implies that the covariance function 4 is

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

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

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

summarize the effect of  $\mathbf{x}_r$  across  $\psi_j$ ,  $j = 1, \dots, k$ . If the impact of  $\mathbf{x}_r$  on  $\psi_j$ ,  $j = 1, \dots, k$  is negligible, then  $g_r(t, \mathbf{x}_r)$  will be near zero. Consequently if  $g_r(t, \mathbf{x}_r)$  is near zero, the  $c(t, t', \mathbf{x})$  will not be sensitive to changes in  $\mathbf{x}_r$ . 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  $\psi_j(t, \mathbf{x})$  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  $\mu(t, \mathbf{x})$ . As we have seen in equations 6 and 7, this amounts to placing a prior on each  $\beta_r$  submatrix. The rows of  $\beta_r$  are associated with a  $p \times p$  penalty matrix  $K$ , and the columns of  $\beta_r$  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  $\beta_r$  through a differencing penalty. In this paper we penalize the second order difference of  $\beta_r$  coefficients in both directions, but other penalties could be used as well. A prior for  $\beta_r$  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  $\beta_r$  is

$$\begin{aligned} \text{vec}(\beta_r) \mid \tau_{1xr}, \tau_{1tr} &\sim \\ \exp\{-0.5\text{vec}(\beta_r)^\top (\tau_{1xr}\tilde{K}_r + \tau_{1tr}\tilde{K})\text{vec}(\beta_r)\} \end{aligned}$$

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

$$\beta_r \mid \tau_{1tr} \sim \exp\{-0.5\tau_{1tr}\beta_r^\top \tilde{K}\beta_r\}$$

This prior is improper but provided that proper priors are set for  $\tau_{1tr}, \tau_{1xr}$ , the posterior of  $\beta_r$  will be proper (Lang and Brezger, 2004). Priors for  $\psi_j(t, \mathbf{x})$ . However,  $\psi_j(t, \mathbf{x})$  for larger  $j$  should also contribute less to the fit than earlier terms. Therefore the prior for  $\psi_j(t, \mathbf{x})$  should encourage smoothing and shrinkage aspects. Let  $\Lambda_{rj}$  be the analogous component to  $\beta_r$ . Re-using the same penalty matrices as above, the prior for  $\Lambda_{rj}$  is

$$\begin{aligned} \text{vec}(\Lambda_{rj}) \mid \tau_{2jxr}, \tau_{2jtr}, \tau_{rj}^*, \phi_{rj} &\sim \\ \exp\{-0.5\text{vec}(\Lambda_{rj})^\top (\tau_{2jxr}\tilde{K}_r + \tau_{2jtr}\tilde{K} + \tau_{rj}^*\phi_{rj})\text{vec}(\Lambda_{rj})\} \end{aligned}$$

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

$$\begin{aligned} \Lambda_{rj} \mid \tau_{2jtr}, \tau_{rj}^*, \phi_{rj} &\sim \\ \exp\{-0.5\Lambda_{rj}^\top (\tau_{2jtr}\tilde{K} + \phi_{rj}\tau_{rj}^*)\Lambda_{rj}\} \end{aligned}$$

similar to the case when  $\beta_{rj}$  is a column vector. In summary, these priors for  $\mu(t, \mathbf{x})$  and  $\psi_j(t, \mathbf{x})$  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 use ideas from the Gamma Multiplicative Process Prior (GMPP) (Bhattacharya and Dunson, 2011; Montagna et al., 2012) to assign priors to  $\tau_{rj}^*$  and  $\phi_{rj}$ . The prior for  $\tau_{rj}^*$  is

$$\begin{aligned}\tau_{rj}^* &= \prod_{l=1}^r \delta_{lr} \\ \delta_{r1} &\sim \text{Gamma}(a_{r0}, 1) \\ \delta_{rl} &\sim \text{Gamma}(a_{r1}, 1), \quad l > 1 \\ a_{r0}, a_{r1} &\sim \text{Gamma}(2, 1)\end{aligned}$$

so that the  $\tau_{rj}$  are stochastically increasing in  $j$ . This shrinkage is data-adaptive so that later entries of  $\Lambda_{rj}$  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. Simulations

### 4.1 Markov-Chain Monte-Carlo and Posterior Distributions

Analytic posterior distributions are intractable, so we rely on Markov-Chain Monte-Carlo techniques to draw samples from all relevant posterior distributions. Since all full conditionals of blocks of parameters are available in closed form, a simple Gibbs sampler updates each parameter block sequentially. See Web Appendix A for all block parameter updating steps.

When the target of inference is a function  $f$ , as opposed to a single point, we adopt methodology from Crainiceanu et al. (2007) to form simultaneous credible bands. Suppose the domain of  $f$  is  $[t_1, t_N]$  and let  $t_1 < \dots < t_N$  be a fine grid of points on this interval. Let  $\mathbb{E}\{f(t_j)\}$  and  $\text{SD}\{f(t_j)\}$  be the pointwise posterior mean and standard deviation of  $f(t_j)$  respectively. Let  $\alpha^*$  be the  $(1 - \alpha)$  sample quantile of  $\max_{1 \leq j \leq N} |f(t_j) - \mathbb{E}\{f(t_j)\}| / \text{SD}\{f(t_j)\}$ . Then  $\mathbb{E}\{f(t_j)\} \pm \alpha^* \text{SD}\{f(t_j)\}$ ,  $1 \leq j \leq N$  constitute  $(1 - \alpha)$  simultaneous credible intervals.

This simultaneous credible band will be used to evaluate uncertainty in the mean and aspects of the covariance in Section 6.2.

## 5. Operative Characteristics

Please see the file `biomsample.tex` for fancy examples of making tables. Here is a very simple one. Use `table` for tables that are narrow enough to fit in one column of the typeset journal; use `table*` for tables that need to span two columns. For figures, use of `figure` and `figure*` is analogous.

## 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 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 baseline SHHS cohort included 52.9% women and 47.1% men. The race distribution was 81% Caucasians, 9.% Native-Americans, 8.0% African-Americans, 1.3% Asians, and 0.03% ‘other’ race category. The minimum age of enrollment was 40 years and there was no upper limit. The mean age was 62.9 (SD 11.0) years old.

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 ( $\geq 75 \mu\text{V}$ ) 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 of increased odds of incident hypertension in both men and women, independent of confounders such as indices of 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 estimation to quantify sleep EEG.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

Of interest is the C3/A2 EEG channel sampled at 125 Hz. Sleep scoring is usually a visually-based counting process, where every 30-second epoch of sleep is assigned a discrete stage

## 7. Discussion

You can experiment with fancier tables than Table ??.

We can get bold symbols using `\bm{math}`, for example,  $\alpha_i$ .

Put your final comments here.

## ACKNOWLEDGEMENTS

The authors thank Professor A. Sen for some helpful suggestions, Dr C. R. Rangarajan for a critical reading of the original version of the paper, and an anonymous referee for very useful comments that improved the presentation of the paper.

## REFERENCES

- Akerstedt, T., Hume, K., Minors, D., and Waterhouse, J. (1997). Good sleep—its timing and physiological sleep characteristics. *Journal of sleep research* **6**, 221.
- 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.
- Cardot, H. (2007). Conditional functional principal components analysis. *Scandinavian journal of statistics* **34**, 317–335.
- 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.
- Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially adaptive bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* **16**, 265–288.
- 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.
- Guo, W. (2002). Functional mixed effects models. *Biometrics* **58**, 121–128.
- 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.
- Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 179–199.
- Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology* **72**, 184–187.
- 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.
- 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.
- 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 Mueller, H.-G. (2015). Review of functional data analysis.

*arXiv preprint arXiv:1507.05135* .

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

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.

#### SUPPORTING INFORMATION

Web Appendix A, referenced in Section 4.1, is available with this paper at the Biometrics website on Wiley Online Library.

#### APPENDIX

*Title of appendix*

Put your short appendix here. Remember, longer appendices are possible when presented as Supplementary Web Material. Please review and follow the journal policy for this material, available under Instructions for Authors at <http://www.biometrics.tibs.org>.











