Biometrics 000, 000–000 DOI: 000
000 0000

Bayesian Conditional Functional Principal Components Analysis

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: 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: Functional data analysis; Peak alpha frequency; Autism spectrum disorder; Covariate-adjusted.

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

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 appealling 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 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, whree 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 2.1 lists prior choices and discusses the reasoning behind them, section 2.2 focuses on inference such as credible intervals for mean functions, Section 3 gives a thorough simulation study, Section 4 showcases the model on the motivating EEG case study, and Section 5 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 $\boldsymbol{x} = (x_1, \ldots, x_{d_1})^\top$ 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

$$y_i(t) = \mu(t, x_1, \ldots, x_{d_1}) + \sum_{j=1}^{k} \psi_j(t, x_1, \ldots, x_{d_1})\eta_{ij} + \epsilon_i(t) \tag{1}$$

where $\mu(t, x_1, \ldots, x_{d_1})$ is the conditional mean, $\psi_j(t, x_1, \ldots, x_{d_1})$ 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. MARGINAL MOMENTS HERE 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 decompose $\mu(\cdot)$ and

$\psi_j(\cdot)$ as linear combinations of spline bases. The complexity of the specific basis expansion will depend on the type of covariates and computational resources. For example, suppose $\boldsymbol{x} = (x_1)$ is a single continuous covariate. Then $\mu(t, x_1)$ can be specified by either $\mu(t, x_1) = x_1 \sum_{p=1}^{P} b_p(t)\beta_p$ or $\mu(t, x_1) = \sum_{p_1}^{P_1} \sum_{p_2}^{P_2} b_{p_1}(t) b_{p_2}(x)\beta_{p_1 p_2}$, where $b(\cdot)$ represents marginal bases, P, P_1, and P_2 are the dimensions of marginal bases and β represents coefficients for the mean structure. Either approach is valid but the tensor formulation is obviously more flexible while incurring a larger computational cost and a more complicated prior for regularization purposes. Continuing this example, $\psi_j(t, x_1)$ can be expanded in a similar manner with Λ_j denoting the set of latent basis coefficients.

This model has similarities to functional mixed effects models (Morris and Carroll, 2006; Guo, 2002). However, in our formulation we design priors to accomodate tensor expansions for greater flexibility. Random components are also drawn from a covariate-adjusted kernel as opposed to pre-specifying its shape. Estimating the distance metric yields important patterns of variability as is usual in functional principal component analyses. The probabilistic model is completed by specifying distributional on assumptions on η_{ij} and $\epsilon_i(t)$. We use η_{ij}

$$\mu(t|\boldsymbol{x_i}) = \sum_{d=1}^{d_1} \mu_d(t)\psi(x_{id}), \quad \phi_k(t|\boldsymbol{x_i}) = \sum_{d=1}^{D} \phi_{kd}(t) x_{id}$$

Furthermore, we expand $\mu_d(t)$ and $\phi_{kd}(t)$ in terms of a P-dimensional fixed spline basis. In this paper we use p-splines (Eilers and Marx, 1996), although other choices are available. Expanding $\mu_d(t)$ and $\phi_{kd}(t)$, we have

$$\mu_d(t) = \sum_{p=1}^{P} b_p(t)\theta_{pd}, \quad \phi_{kd}(t) = \sum_{p=1}^{P} b_p(t)\lambda_{kpd}$$

where $b_p(t)$, $p = 1, \ldots, P$, represent fixed basis splines. Let Θ be a $P \times D$ matrix with entry (p, d) equal to θ_{pd}, let Λ_k be a $P \times K$ matrix with entry (p, d) equal to λ_{kpd}, and let $b(t) = (b_1(t), \ldots, b_p(t))^\top$. Then Equation (1) can be rewritten as

$$y_i(t) = b(t)^\top \Theta \boldsymbol{x_i} + \sum_{k=1}^{K} b(t)^\top \Lambda_k \boldsymbol{x_i} \eta_{ik} + \epsilon_i(t)$$

Distributional assumptions must be imposed on η_{ik} and $\epsilon_i(t)$. In this paper, we simply use

$\eta_{ik} \sim N(0,1)$ and $\epsilon_i(t) \sim N(0, \sigma^2)$. Of course, more complex choices are warranted by the application at hand. The model structure and assumptions yields the covariate-dependent mean

$$\mathbb{E}(y_i(t)|\boldsymbol{x_i}) = b(t)^\top \Theta \boldsymbol{x_i}$$

and covariate-dependent covariance

$$\text{cov}(y_i(t), y_i(t')|\boldsymbol{x_i}) = \sum_{k=1}^{K} b(t)^\top \Lambda_k \boldsymbol{x_i} \boldsymbol{x_i}^\top \Lambda_k^\top b(t') \tag{2}$$

This model can be interpreted as a functional extension to Hoff and Niu (2012) via incorporating basis spline expansions. In contrast to Hoff and Niu (2012), we avoid Wishart priors on the residual error process to increase model identifiability and to reflect our assumptions of the presence of measurement error. This model also shares similarities with functional mixed effects models (Morris and Carroll, 2006; Guo, 2002) in the time-varying linear structure of the mean. However, Equation (1) models the subject to subject variability in a different manner. Instead of using a random intercept or random slope, this model uses a covariate-dependent functional probabilistic principal component analysis decomposition. This decomposition is nonparametric and is suitable when covariance estimation is one of the goals of analysis.

In contrast to covariance estimation via kernel smoothers (Cardot, 2007; Jiang et al., 2010), this model uses a quadratic time-varying covariance structure as seen from Equation (2). Although the aforementioned approaches are extremely flexible, they can face serious practical difficulties when the covariate dimension is not miniscule (Montagna et al., 2012). The linear structure approach relieves much of the computational burden. At the same time, the linear structure allows for group covariates, which is essential for the case study presented in this paper.

2.1 *Prior Distributions*

The Bayesian paradigm requires one to specify prior distributions for all unknown quantities. We use smoothing priors in the mean and covariance parameters to prevent overfitting. Let Θ_d and Λ_{kd} denote the dth column of Θ and Λ_k respectively. We assign a Gaussian Markov Random Field prior to Θ_d and Λ_{kd} such that

$$\Theta_d|\tau_{0d} \propto \exp\left(-\frac{\tau_{0d}}{2}\Theta_d^\top K_0 \Theta_d\right), \quad \Lambda_{kd}|\tau_{1d} \propto \exp\left(-\frac{\tau_{1d}}{2}\Lambda_{kd}^\top K_0 \Lambda_{kd}\right)$$

where K_0 is the first-order difference matrix with $rank(K_0) = P-1$. Since K_0 does not have full rank, these priors are improper. In addition, τ_{0d} and τ_{1d} are assigned Gamma priors, so that $\tau_{0d}, \tau_{1d} \sim \text{Gamma}(a_\tau, b_\tau)$ with $\mathbb{E}(\tau_{0d}) = a_\tau/b_\tau$. We set $a_\tau = 1$ and $b_\tau = .0005$ which has been recommended by Lang and Brezger (2004). We place a conventional Inverse-Gamma prior on σ^2, so that $\sigma^{-2} \sim \text{Gamma}(a_\sigma, b_\sigma)$. We set $a_\sigma, b_\sigma = .0001$ to achieve a diffuse yet proper prior on σ^{-2}.

2.2 *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 $t_1 < \ldots < t_N$ 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 α^* be the $(1-\alpha)$ sample quantile of $\max_{1\leqslant j\leqslant N} |f(t_j) - \mathbb{E}\{f(t_j)\}|/\text{SD}\{f(t_j)\}$. Then $\mathbb{E}\{f(t_j)\} \pm \alpha^* \text{SD}\{f(t_j)\}, 1 \leqslant j \leqslant N$ 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 4.

3. 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 journl; use `table*` for tables that need to span two columns. For figures, use of `figure` and `figure*` is analogous.

[Table 1 about here.]

4. Case Study

5. Discussion

You can experiment with fancier tables than Table 1.

We can get bold symbols using `\bmath`, for example, $\boldsymbol{\alpha}_i$.

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

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

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.

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.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science* pages 89–102.

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.

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.

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.

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,**.

Wang, J.-L., Chiou, J.-M., and Mueller, H.-G. (2015). Review of functional data analysis. *arXiv preprint arXiv:1507.05135* .

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 2.2, 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`.

Table 1
This is a simple table.

Estimator	β_1	β_2	β_3
MLE	10.18	−3.26	0.13
OLS	9.92	−3.19	0.11
WLS	9.88	−3.33	0.12