Biometrics 000, 000–000 DOI: 000
000 0000

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: 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.

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 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 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 $\boldsymbol{x} = (x_1, \ldots, x_{d_1})^\top$ 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

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

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), $\mu(t, \boldsymbol{x})$ can be written as

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

where the set $\cup\{\boldsymbol{x}_r\}_{r=1}^R = \mathbf{x}$. 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 $f_r(t, x_{r_1}, x_{r_2}) = f(t, x_{r_1}, x_{r_2}), x_{r_1} f(t, x_{r_2})$, 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.

The $f_r(t, \boldsymbol{x}_r)$ 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 $\boldsymbol{b}(t) = (b_1(t), \ldots, b_p(t))^\top$, $\boldsymbol{b}^r(x_r) = (b_1^r(x_r), \ldots, b_{p_r}^r(x_r))^\top$, and arrange the coefficients into a $p \times p_r$ matrix β_r. 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

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

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

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

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

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

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

where Λ_{jr} is a $p \times p_r$ loading matrix and $\Lambda_j = \big(\Lambda_{j1} \,|\, \cdots \,|\, \Lambda_{jR}\big)$. 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

$$g_r(t, \boldsymbol{x}_r) = \sum_{j=1}^{k} l_{jr}(t, \boldsymbol{x}_r)^2 \;/\; \sum_{r=1}^{R} \sum_{j=1}^{k} l_{jr}(t, \boldsymbol{x})^2$$

summarize the effect of $\boldsymbol{x}_r$ across ψ_j, $j = 1, \ldots, k$. Clearly $g_r(t, \boldsymbol{x}_r) \in [0, 1]$ and $\sum_{r=1}^{R} g_r(t, \boldsymbol{x}_r) = 1$. 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 possible through the additivity assumption on $\psi_j(t, \boldsymbol{x})$ in 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 equation **??**, 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

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

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

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

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

This prior is improper but provided that proper priors are set for 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_{rj} 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 $x \sim Gamma(a, b)$, then $\mathbb{E}[x] = a/b$). In addition we use ideas from the Gamma Multiplicative Process Prior (GMPP) (Bhattacharya and Dunson, Bhattacharya and Dunson; Montagna et al., 2012) to assign priors to τ_{rj}^* and ϕ_{rj}. The prior for τ_{rj}^* is

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

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

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

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

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. 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 $t_1 < \ldots < t_N$ 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 6.

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

6. Case Study

7. Discussion

You can experiment with fancier tables than Table **??**.

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

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

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.

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,**.

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

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`.