# REGION-REFERENCED SPECTRAL POWER DYNAMICS OF EEG SIGNALS: A HIERARCHICAL MODELING APPROACH

BY QIAN LI[1,*], JOHN SHAMSHOIAN[1], DAMLA ŞENTÜRK[1], CATHERINE SUGAR[1], SHAFALI JESTE[2], CHARLOTTE DISTEFANO[2] AND DONATELLO TELESCA[1,†]

[1]*Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles,*
[*]*qianl@ucla.edu;* [†]*dtelesca@ucla.edu*

[2]*Department of Psychiatry and Department of Psychiatry and Biobehavioral Sciences, UCLA David Geffen School of Medicine, University of California, Los Angeles*

Functional brain imaging through electroencephalography (EEG) relies upon the analysis and interpretation of high-dimensional, spatially organized time series. We propose to represent time-localized frequency domain characterizations of EEG data as region-referenced functional data. This representation is coupled with a hierarchical regression modeling approach to multivariate functional observations. Within this familiar setting we discuss how several prior models relate to structural assumptions about multivariate covariance operators. An overarching modeling framework, based on infinite factorial decompositions, is finally proposed to balance flexibility and efficiency in estimation. The motivating application stems from a study of implicit auditory learning, in which typically developing (TD) children, and children with autism spectrum disorder (ASD) were exposed to a continuous speech stream. Using the proposed model, we examine differential band power dynamics as brain function is interrogated throughout the duration of a computer-controlled experiment. Our work offers a novel look at previous findings in psychiatry and provides further insights into the understanding of ASD. Our approach to inference is fully Bayesian and implemented in a highly optimized Rcpp package.

**1. Introduction.** The human brain and its functional relation to biobehavioral processes like motor coordination, memory formation and perception as well as pathological conditions like Parkinson's disease, epilepsy and autism have been a subject of intense scientific scrutiny (Broyd et al. (2009), Murias et al. (2007), Uhlhaas and Singer (2010)). An important and highly prevalent imaging paradigm aims to study macroscopic neural oscillations projected onto the scalp in the form of electrophysiological signals and record them by means of electroencephalography (EEG).

Currently, a typical multi-channel EEG imaging study is carried out with a geodesic EEG-net, composed of up to 256 electrodes. After precise placement on the scalp, electrodes collect electrophysiological signals at high time resolution through event-related potentials (ERP) or event-related oscillations (ERO). In this setting, endogenous or exogenous events can result in frequency-specific changes to ongoing EEG oscillations. The spectral features of the resulting time series are often used to quantify such changes. Specifically, in a time-frequency analysis the spatiotemporal dynamics of frequency-specific power are examined, as they relate to sensory, motor and/or cognitive processes (Gou, Choudhury and Benasich (2011), Mills, Coffey-Corina and Neville (1993), Scheffler et al. (2020)).

Our work is motivated by a study of language acquisition in young children with autism spectrum disorder (ASD), a developmental condition that affects an individual's communication and social interactions (Lord et al. (2000)). It is thought that typically developing (TD)

infants, as young as 6 months old, start to parse continuous streams of speech to actively discern word patterns (Kuhl (2004)). Infants diagnosed with ASD, tend to feature late development of linguistic skills (Eigsti et al. (2011)). Because neither verbal instructions nor behavioral evaluations can be performed on toddlers, EEG platforms have been recognized as an effective and noninvasive functional brain imaging tool.

In an idealized setting, for each study unit we aim to characterize variability in the relative log-power of a specific frequency band, recorded at a scalp location $s$, at time $t$. This aim is, however, made more complex by several estimation- and data-related challenges. Raw EEG recordings, in fact, suffer from exogenous contamination and poor spatial resolution as well as nonstationarity. To mitigate these issues, several preprocessing and regularization strategies have been proposed in the literature (Scheffler et al. (2020)). In this article we build on our previous work and couple the characterization of EEG spectral dynamics as region-referenced functional data with a Bayesian model for multivariate functional observations. Figure 1 illustrates the fundamental nature of the data structure considered in our motivating case study by depicting specific band power trajectories through time in trial and across three of 11 brain regions in a study of language acquisition. Given this data structure, we are interested in region-specific, time-varying differences in band power dynamics between diagnostic groups. Crucially, valid inference depends on the correct specification of large multivariate covariance functions.
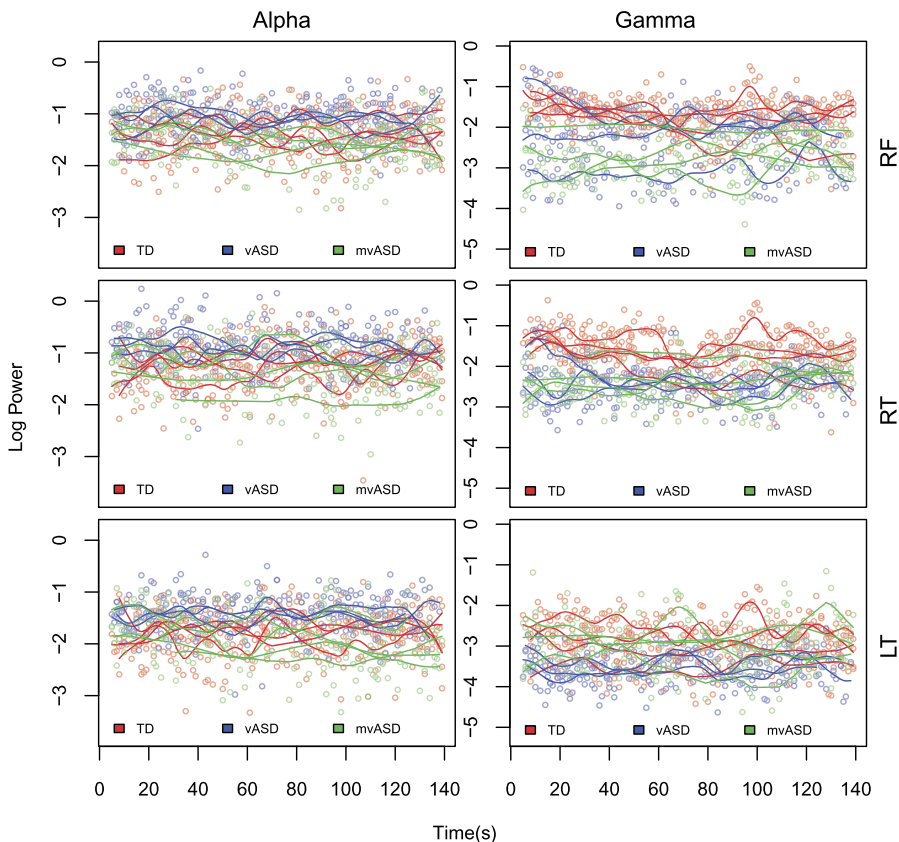


FIG. 1. Power dynamics in language acquisition. *Log-power dynamics both raw and smoothed, in three of 11 anatomical regions: left temporal (LT), right temporal (RT) and right frontal (RF). For each region we plot the principal component power for two frequency bands (Alpha, Gamma) during time on trial. We select three subjects from each of the three diagnostic groups: typically developing (TD), verbal ASD (vASD) and minimally verbal ASD (mvASD).*

Many approaches to functional data analysis rely on functional principal components analysis as a fundamental tool for advanced modeling (James, Hastie and Sugar (2000), Yao, Müller and Wang (2005)). Methods for the analysis of multivariate functional data are more sporadic but tend to follow similar strategies with specific attention paid to data scales (Chiou, Chen and Yang (2014)) or potential heterogeneity in the functional evaluation domain (Happ and Greven (2018)). The Bayesian literature on the subject is more sparse, with notable exceptions to be found in Baladandayuthapani et al. (2008), who consider nested spatially correlated functional data under assumptions of separability, and the more general work of Zhou et al. (2010), who consider similar data structures under more flexible covariance models.

The use of typical spatial assumptions is, however, problematic in the analysis of region-referenced EEG data. Namely, Euclidean distances between electrodes may not accurately characterize functional dependency of within-subject, relative log-power readings. Therefore, instead of relying on prespecified covariance structures (e.g., conditionally autoregressive models) to describe within-subject correlation, we propose to estimate them through functional factor analysis.

In particular, this article introduces a probabilistic framework for multivariate functional data regression which hinges on two ideas: (1) the constructive definition of Gaussian processes through basis expansions and (2) the modeling of covariance operators through functional factor analysis (Montagna et al. (2012)). While neither approach is new, we highlight the application of these concepts to functional brain imaging though EEG. In the process we discuss important consequences associated with specific implementation choices as they relate to transparent assumptions on the structure of multivariate covariance functions. Our proposal seeks to develop a general framework for automatic shrinkage and regularization through infinite factor models (Bhattacharya and Dunson (2011)). The resulting methodology is suitable for nonparametric estimation with any projection methods which is amenable for use in a broad range of applications.

The article is organized as follows. In Section 2 we outline a conceptual framework for region-referenced band powers dynamics. A hierarchical functional model for region-referenced functional data is introduced in Section 3. We assess the performance of the proposed model on engineered datasets in Section 4 and apply our work to a neurocognitive study on auditory implicit learning in Section 5. We conclude with a critical discussion of our work in Section 6.

**2. Time-dependent, region-referenced band-power.**   EEG signals, measured through high-resolution geodesic networks of electrodes, allow for highly localized interrogation of cortical functions throughout the duration of a computer-controlled experiment. In large studies the high-dimensional nature of the resulting time series poses potential modeling and computational challenges.

Due to spatial proximity, neighboring electrodes collect signals likely to be highly multicollinear. In this context the analysis of EEG data is often preceded by a dimension reduction exercise, aimed at discarding redundant information and aid interpretation through the definition of region-level summaries. Specifically, electrodes are partitioned according to anatomical regions on the scalp. Subsequently, region-level summaries are defined by averaging power spectra or by selecting the power spectrum of a single electrode within the region. Alternatively, given that multicollinearity of signals defines similarity in their spectral features, spectral PCA (Brillinger (1981)) has been proposed to pool information within anatomical regions with minimal loss of information. Applications to EEG data have been recently proposed by Ombao and Ho (2006) and Scheffler et al. (2020). Our approach follows closely the development in Scheffler et al. (2020), who perform spectral PCA on nonoverlapping EEG segments. Our target summaries, however, focus on specific frequency bands, as opposed to the entire power spectrum.

We outline and define basic concepts used in our work and defer details to Section 2 of the Supplementary Material (Li et al. (2020a)). Let $X_{ij}(s, t)$ denote the raw EEG recording, conceptualized as a locally stationary, zero-mean $q_j$-dimensional random process observed on subject $i$, $i = 1, \ldots, n$, within anatomical region $j$, $j = 1, \ldots, p$, for segment $t$, $t = 1, \ldots, m$, at a sampling rate $U$ across discretized time $s$, ($s = 0, \pm1, \ldots, \pm U/2$), assuming a 1s analysis window. For each observation $X_{ij}(s, t)$, FFT is performed to obtain Fourier coefficients $d_{ij}(\omega, t) = U^{-1/2} \sum_s X_{ij}(s, t) \exp\{-i2\pi\omega s\}$, at frequency $\omega$. The raw periodogram matrix $I_{ij}(\omega, t) = d_{ij}(\omega, t)\overline{d_{ij}}(\omega, t)'$ is computed, where $\overline{d_{ij}}(\omega, t)'$ is the transpose of the complex conjugate of $d_{ij}(\omega, t)$. Following Ombao and Ho (2006), we define kernel smoothed spectral density matrices $\tilde{I}_{ij}(\omega, t)$, by smoothing $I_{ij}(\omega, t)$ across $\omega$, using a Daniell kernel with bandwidths selected through generalized cross validation (GCV). Crucially, smoothing bandwidth are held fixed within region, resulting in $\tilde{I}_{ij}(\omega, t)$ being Hermitian and nonnegative definite.

A one-dimensional, region-level summary may be obtained by defining the *principal power* $\lambda_{ij}^*(\omega, t) := c_{ij}^{-1}(t) \max_{|z|=1} z' \tilde{I}_{ij}(\omega, t) z$, as the normalized leading eigenvalue of $\tilde{I}_{ij}(\omega, t)$, with $c_{ij}(t) = \int \lambda_{ij}^*(\omega, t) \, d\omega$. Within region the principal power summarizes common frequency-level variation across electrodes along the direction of the leading eigenvector. In contrast to Scheffler et al. (2020), who consider region-referenced, longitudinal-functional representations, we adopt a simplified view and focus on specific frequency bands. Specifically, for a frequency range $(\omega_a, \omega_b)$ this article emphasizes modeling the *time-varying principal band power* $\gamma_{ij}(t) := \int_{\omega_a}^{\omega_b} \lambda_{ij}^*(\omega, t) \, d\omega$.

In adults, typical frequency bands and their spectral boundaries are *delta* ($<4$ Hz), *theta* (4–8 Hz), *alpha* (8–15 Hz), *beta* (15–32 Hz) and *gamma* (32–50 Hz). This view is motivated by traditional approaches in neurocognitive science which differentiate functionally distinct frequency bands as they are thought to target distinct neurocognitive and biobehavioral processes (Saby and Marshall (2012)).

## 3. Hierarchical models for region-referenced functional data.

3.1. *Data projections.* Given a frequency band of interest, let $Y_{ij}(t) = \log\{\gamma_{ij}(t)\}$ be the logarithm of the principal band power for subject $i$, anatomical region $j$, evaluated at time $t$. In practice, for each subject only a finite number of observations are collected discretely at $t_i = \{t_{i1}, t_{i2}, \ldots, t_{im_i}\}$. However, for ease of notation and without loss of generality, we assume $t \in \mathcal{T} = [a, b]$. Collecting all region-level log band-power measurements into a $p$-dimensional vector $Y_i(t) = \{Y_{i1}(t), \ldots, Y_{ip}(t)\}'$, we characterize subject-level observations as possibly contaminated multivariate functional data. Specifically, let $f_i(t) = \{f_{i1}(t), \ldots, f_{ip}(t)\}'$ be a set of $p$-dimensional random functions with each $f_{ij}(t)$, $j = 1, \ldots, p$ in $\mathcal{L}_2(\mathcal{T})$, a Hilbert space of square integrable functions with respect to the Lebesque measure on $\mathcal{T}$. We assume

$$(3.1) \qquad Y_i(t) = f_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim N_p(0, \Sigma_\epsilon),$$

with residual covariance $\Sigma_\epsilon = \text{diag}(\sigma_{\epsilon1}^2, \ldots, \sigma_{\epsilon p}^2)$. Given $f_i(t)$ at time $t$, $Y_i(t)$ is assumed to arise from independent realizations of a heteroscedastic multivariate Gaussian distribution.

We characterize the random functions $f_i(t)$ through finite-dimensional projections onto $q$ cubic B-spline bases $b(t) = \{b_1(t), \ldots, b_q(t)\}'$. Specifically, let $\Theta_i = \{\theta_{ijk}\} \in \mathbb{R}^{p \times q}$ be a matrix of random basis coefficients; we express 3.1 as follows:

$$(3.2) \qquad Y_i(t) = \Theta_i b(t)' + \epsilon_i(t).$$

The representation above is readily available for hierarchical modeling. Given priors on $\Theta_i$ and $\Sigma_\epsilon$, standard posterior inference and computation applies to the seemingly challenging problem of modeling multivariate functional data. Crucially, if the elements of $\Theta_i$ are

assumed Gaussian, given $\boldsymbol{b}(t)$, the simple construction in 3.2 implies that $\boldsymbol{f}_i(t)$ follows a $p$-dimensional Gaussian process, with second-order properties fully determined by covariance assumptions on the random basis coefficients.

In the following sections we develop an overarching structure for prior specification, including possible dependence on covariate information and discuss modeling strategies and their implications for flexibility and efficiency in estimation.

3.2. *Hierarchical priors and dependence on covariates.* A typical study of functional brain imaging aims to relate subject-level, time-stable covariate information $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{id})'$, with a region-referenced, dynamic outcome $\boldsymbol{Y}_i(t)$. We focus on the conditional expectation $E\{\boldsymbol{Y}_i(t) \mid \boldsymbol{w}_i\} = E(\boldsymbol{\Theta}_i \mid \boldsymbol{w}_i)\boldsymbol{b}(t)'$ as the principal measure of association between covariates and brain function outcomes. Let $M(\boldsymbol{w}_i) := E(\boldsymbol{\Theta}_i \mid \boldsymbol{w}_i)$; a familiar modeling framework for $\boldsymbol{\Theta}_i$ relies on the following matrix-mixed effects construction:

$$(3.3) \qquad \boldsymbol{\Theta}_i = M(\boldsymbol{w}_i) + \boldsymbol{Z}_i, \qquad \text{vec}(\boldsymbol{Z}_i) \sim N_{pq}(0, \boldsymbol{\Sigma}_z),$$

where $\boldsymbol{Z}_i \in \mathbb{R}^{p \times q}$ captures subject specific regional and functional variation. Assuming $\boldsymbol{w}_i$ is organized as a regression design vector, let $\boldsymbol{\Psi}_\ell \in \mathbb{R}^{p \times q}$, $\ell = 1, \ldots, d$, be regression coefficients; a matrix linear model defines $M(\boldsymbol{w}_i) = \sum_{\ell=1}^{d} \boldsymbol{\Psi}_\ell w_{i\ell}$. In this setting the regression functions, $\boldsymbol{\mu}_\ell(t) := \boldsymbol{\Psi}_\ell \boldsymbol{b}(t)'$, are $p$-dimensional varying coefficients to be interpreted in relation to the design structure encoded in $\boldsymbol{w}_i$ (Zhu, Li and Kong (2012)).

Given a basis projection, the second-order behavior indexing functional and region-level dependence for the the multivariate Gaussian process $\boldsymbol{f}_i(t)$ is fully determined by the covariance matrix $\boldsymbol{\Sigma}_z$. While finite-dimensional, this matrix is likely high-dimensional, as it spans both the number of regions $p$ and a potentially large number of basis functions $q$. We discuss three modeling approaches, aimed at regularization through shrinkage and simplifying structural assumptions, namely: a naïve-Bayes (NB) prior, a separable sandwich (SS) prior and a regularized nonseparable (NS) prior.

*Naïve-Bayes (NB) prior.* This construction exploits two simplifying assumptions: (1) the random coefficients matrix $\boldsymbol{Z}_i$ is assumed to follow a Matrix Gaussian distribution, and (2) covariance along time is parametrized through Bayesian P-splines smoothing penalties (Lang and Brezger (2004)). Let $\boldsymbol{\Omega}_0 \in \mathbb{R}^{q \times q}$ be a deterministic penalty matrix and $\boldsymbol{S} \in \mathbb{R}^{p \times p}$ be a covariance matrix indexing dependence between anatomical regions. The *NB* prior assumes

$$(3.4) \qquad \boldsymbol{Z}_i \sim \mathcal{MN}(0, \boldsymbol{S}, \boldsymbol{\Omega}_0^{-1}), \qquad \boldsymbol{S}^{-1} \sim W(\nu, \boldsymbol{S}_0^{-1}/\nu).$$

The prior above structures $\boldsymbol{\Sigma}_z = \boldsymbol{S} \otimes \boldsymbol{\Omega}_0^{-1}$, implying separability of covariation between regions and time points. The matrix $\boldsymbol{S}$ serves both the purpose of modeling dependence between brain regions through its of-diagonal elements and the purpose of establishing region-level adaptive smoothing through its diagonal elements as they multiply $\boldsymbol{\Omega}_0^{-1}$. More details about this matrix and the choice of hyperparameters are discussed in Web Appendix C. While greatly simplified, when compared to a completely unstructured $\boldsymbol{\Sigma}_z$, this construction is potentially problematic, as it hinges on a relatively rigid parametrization for the time-covariance while enforcing only mild Wishart regularization of $\boldsymbol{S}$.

*Separable sandwich (SS) prior.* A more balanced approach to the prior in 3.4 retains the assumption of separability but implements adaptive regularized estimation of both the time and region covariance. In particular, we extend the approach of Montagna et al. (2012) and propose a two-way Bayesian latent factor model for $\boldsymbol{Z}_i$. Let $\boldsymbol{\Upsilon} = \{\upsilon_{jr}\} \in \mathbb{R}^{p \times k_1}$ and $\boldsymbol{\Gamma} = \{\gamma_{vs}\} \in \mathbb{R}^{q \times k_2}$ be two loading matrices. Also, let $\boldsymbol{H}_i = \{h_{rs}\} \in \mathbb{R}^{k_1 \times k_2}$ be a random matrix with $h_{rs} \sim_{iid} N(0, 1)$, $r = 1, \ldots, k_1$, $s = 1, \ldots, k_2$; we write

$$(3.5) \qquad \boldsymbol{Z}_i = \boldsymbol{\Upsilon} \boldsymbol{H}_i \boldsymbol{\Gamma}' + \boldsymbol{R}_i \quad \text{with } \boldsymbol{R}_i \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{\Sigma}_p, \boldsymbol{\Sigma}_q),$$

where $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_q$, both are diagonal with anisotropic components. This construction, implies

$$\boldsymbol{\Sigma}_z = \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}' \otimes \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Sigma}_p \otimes \boldsymbol{\Sigma}_q.$$

Typically, as in latent-factor models, a truncation $k_1 < p$ and $k_2 < q$ is selected to define a low-rank representation of the region and time covariances. Rather than selecting the number of latent factors, we consider the multiplicative shrinkage prior of Bhattacharya and Dunson (2011), so that:

$$\upsilon_{jr} \mid \phi_{jr}, \tau_r \sim \mathrm{N}(0, \phi_{jr}^{-1}\tau_r^{-1}), \qquad \phi_{jr} \sim \mathrm{Ga}(\nu_1/2, \nu_1/2), \qquad \tau_r = \prod_{u=1}^{r} \delta_u,$$

$$\delta_1 \sim \mathrm{Ga}(a_{11}, 1), \qquad \delta_u \sim \mathrm{Ga}(a_{12}, 1) \quad \text{when } u > 1;$$

$$\gamma_{vs} \mid \rho_{vs}, \kappa_s \sim \mathrm{N}(0, \rho_{vs}^{-1}\kappa_s^{-1}), \qquad \rho_{vs} \sim \mathrm{Ga}(\nu_2/2, \nu_2/2), \qquad \kappa_s = \prod_{\ell=1}^{s} \pi_v,$$

$$\pi_1 \sim \mathrm{Ga}(a_{21}, 1), \qquad \pi_v \sim \mathrm{Ga}(a_{22}, 1) \quad \text{when } \ell > 1.$$

When $a_{11} > 1$, $a_{12} > 1$, this prior defines stochastically increasing precisions as more columns are added to $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Gamma}$. Specific details about this shrinkage strategy and choice of hyperparameters are reported in the Supplementary Material (Section 3) (Li et al. (2020a)). In what follows, we build on latent factor representations and forego the assumption of separability altogether.

*Nonseparable (NS) prior.* As noted in Cressie and Huang (1999), the class of separable models ((3.4), (3.5)) is severely limited since it cannot capture region-time interaction. The cross-covariance functions between the time series at any region has the same shape, regardless of region location. Separable structures are often chosen for convenience rather than for their ability to fit the data well. These limitations make the separability approach difficult to justify in the setting of functional neuroimaging. A simple approach, which foregoes this assumption while still defining a regularized representation of $\boldsymbol{\Sigma}_z$, involves dealing with $\mathrm{vec}(\boldsymbol{Z}_i)$ directly. Specifically, let $\boldsymbol{\Xi} = \{\xi_{cs}\} \in \mathbb{R}^{pq \times k}$, be a loading matrix and $\boldsymbol{\eta}_i \sim N_k(0, I_k)$; we write

$$(3.6) \qquad \mathrm{vec}(\boldsymbol{Z}_i) = \boldsymbol{\Xi}\boldsymbol{\eta}_i + \boldsymbol{r}_i \quad \text{with } \boldsymbol{r}_i \sim N_{pq}(0, \boldsymbol{\Sigma}_r),$$

with $\boldsymbol{\Sigma}_r = \mathrm{diag}(\sigma_{r1}^2, \ldots, \sigma_{rpq}^2)$, implying $\boldsymbol{\Sigma}_z = \boldsymbol{\Xi}\boldsymbol{\Xi}' + \boldsymbol{\Sigma}_r$. This formulation leads to a probabilistic version of the multivariate FPCA of Chiou, Chen and Yang (2014), with normalization handled through explicit assumptions of heteroscedasticity through $\boldsymbol{\Sigma}_r$. Echoing the approach used in the sandwich prior, the model is completed with a shrinkage prior on the loading matrix $\boldsymbol{\Xi}$, so that:

$$\boldsymbol{\xi}_{(s)}^{(c)} \mid \boldsymbol{\Omega}_s, \tau_c \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_s^{-1}\tau_c^{-1}), \qquad \boldsymbol{\Omega}_s \sim \mathrm{W}(\nu, \boldsymbol{\Omega}_0/\nu), \qquad \tau_c = \prod_{l=1}^{c} \delta_l,$$

$$\delta_1 \sim \mathrm{Ga}(a_1, 1), \qquad \delta_l \sim \mathrm{Ga}(a_2, 1) \quad \text{when } l > 1,$$

where $\boldsymbol{\xi}_{(s)}^{(c)}$ indicates the $s$th $p$-dimensional block, $s \in \{1, \ldots, q\}$, of the $c$th column of $\boldsymbol{\Xi}$, $c = 1, \ldots, k$.

It is easy to show that the vectorized model in 3.6 encompasses both the *separable sandwich* prior in 3.5 and the *naïve-Bayes* prior in 3.4. We examine the finite sample properties of these modeling strategies under several simulation scenarios in the Section 4.

The three priors listed above span a broad range of model flexibility. We recognize, however, that several other nuanced configurations have been considered in the literature, particularly, when sampling schedules allow for design-driven structural assumptions. For example,

in the context of longitudinal functional data, methodological simplifications to multivariate functional data include both functional random effects models (Greven et al. (2010)) and simplified multivariate functional principal components representations (Park and Staicu (2015)). Weakly separable covariance structures have been characterized in Chen and Lynch (2017), bridging the complexity of SS and NS models, by expressing a multiway covariance as the linear combination of rank-one, strongly-separable terms. This idea has found several applications and extensions in the context of longitudinal functional data literature (Lynch and Chen (2018), Scheffler et al. (2020), Shamshoian et al. (2019)). While a comprehensive comparison is out of scope for this work, in the context of our case study and through simulations, we aim to assess how a general NS model can avoid losses in efficiency through adaptive regularization.

3.3. *Posterior inference.* Posterior inference is based on Markov chain Monte Carlo simulations from the target measure. All prior models discussed in Section 3.2 are amenable to simple Gibbs sampling implementations. A highly-optimized R package for data manipulation and inference is freely available from `github` at https://github.com/Qian-Li.

The MCMC transition schedule implemented in our package is optimized to marginalize out subject by region coefficients whenever possible, therefore limiting the degree of autocorrelation between posterior draws. As is the case with GP-type regression, in applying our method beyond EEG data some care is needed, as the scalability of simple posterior sampling schemes could be an issue in high dimensions (large $n$ or $p$). In these cases, specialized approximation strategies are, however, applicable, as described in Nishimura and Suchard (2018).

We note that the loading matrix $\mathbf{\Xi}$ in (3.6) as well as $\mathbf{\Upsilon}$ and $\mathbf{\Gamma}$ in (3.5) are not likelihood identified due to invariance to orthogonal rotations. Crucially, however, $\mathbf{\Xi}\mathbf{\Xi}'$, $\mathbf{\Upsilon}\mathbf{\Upsilon}'$ and $\mathbf{\Gamma}\mathbf{\Gamma}'$ are all uniquely identified, leading to likelihood-identifiability of $\mathbf{\Sigma}_z$. Given posterior samples from the model parameters, Monte Carlo estimates of all quantities of interest, including simultaneous credible bands are obtained in a relatively straightforward fashion (Baladandayuthapani, Mallick and Carroll (2005)). Detailed calculations, including full conditional distributions are reported in the Supplementary Material (Li et al. (2020a)).

**4. Experiments on engineered data.** To evaluate the finite sample performance of the hierarchical model and priors described in Section 3, we carried out an extensive Monte Carlo study considering data generated under several dependence structure (separable, nonseparable), dependence strength, sample size and signal-to-noise ratio scenarios. The signal-to-noise ratio is defined by $\text{SNR} = \frac{1}{n}\frac{1}{p}\sum_{i=1}^{n}\sum_{j=1}^{p}\int_{\mathcal{T}} f_{ij}(t)^2 \, dt/\sigma_\epsilon^2$. Intuitively, SNR in this context can be intended as an inflation quotient for the error-variance $\sigma_\epsilon^2$. The general goal of these experiments aims to assess how well the mean $M(\mathbf{w}_i)$ and covariance $\mathbf{\Sigma}_z$ are recovered under different priors and simulation truths. Further details regarding the data generation scheme are available in Web Appendix (Section 4).

Let $\hat{\boldsymbol{\mu}}_\ell(t)$ and $\widehat{\mathbf{\Sigma}}_z$ denote posterior mean estimates for their parametric counterparts (Section 3.2). To quantify the quality of estimates, we consider relative squared errors, defined as follows:

- *Mean*: Average relative squared error across regions and covariates, s.t.

$$\text{RSE}(M) = \frac{1}{dp}\sum_{\ell=1}^{d}\left[\frac{\int_t\{\hat{\boldsymbol{\mu}}_\ell(t) - \boldsymbol{\mu}_\ell(t)\}^2 \, dt}{\int_t \boldsymbol{\mu}_\ell(t)^2 \, dt}\right]' \mathbf{1}_p;$$

TABLE 1
*Simulation study. Mean and covariance recovery under naïve-Bayes (NB), separable-sandwich (SS), and nonseparable (NS) priors. Relative errors are reported under separable and nonseparable simulation truths, sample size escalation ($n = 10, 20, 30$) and two signal-to-noise ratio scenarios (SNR $= 0.2, 1.0$)*

| | $n = 10$ | | $n = 20$ | | $n = 50$ | |
|---|---|---|---|---|---|---|
| | Mean | Cov. | Mean | Cov. | Mean | Cov. |
| *Separable* | (SNR $= 0.2$) | | | | | |
| NB | 0.0346 | 0.3255 | 0.0248 | 0.3150 | 0.0162 | 0.3120 |
| SS | 0.0384 | 0.3192 | 0.0272 | 0.3004 | 0.0174 | 0.2806 |
| NS | 0.0386 | 0.3380 | 0.0272 | 0.3117 | 0.0174 | 0.2900 |
| | (SNR $= 1.0$) | | | | | |
| NB | 0.0277 | 0.5146 | 0.0195 | 0.5126 | 0.0124 | 0.5180 |
| SS | 0.0272 | 0.1987 | 0.0192 | 0.1817 | 0.0122 | 0.1677 |
| NS | 0.0272 | 0.2595 | 0.0192 | 0.2294 | 0.0122 | 0.1915 |
| *Nonseparable* | (SNR $= 0.2$) | | | | | |
| NB | 0.0382 | 0.4217 | 0.0276 | 0.4189 | 0.0183 | 0.4226 |
| SS | 0.0413 | 0.3633 | 0.0289 | 0.3547 | 0.0185 | 0.3470 |
| NS | 0.0415 | 0.3427 | 0.0290 | 0.3188 | 0.0185 | 0.2892 |
| | (SNR $= 1.0$) | | | | | |
| NB | 0.0339 | 0.6601 | 0.0236 | 0.6110 | 0.0153 | 0.6164 |
| SS | 0.0340 | 0.3412 | 0.0235 | 0.3357 | 0.0152 | 0.3321 |
| NS | 0.0338 | 0.2721 | 0.0234 | 0.2433 | 0.0151 | 0.2101 |

- *Covariance*: Relative squared Frobenius error, st.

$$\mathrm{RSE}(\Sigma_z) = \frac{\|\Sigma_z - \widehat{\Sigma}_z\|_F^2}{\|\Sigma_z\|_F^2}.$$

In relation to these metrics, we observe the following behavior:

*Sample size and SNR.* For each of 100 datasets, we simulate $n = 10, 20, 50$ individuals in each of three diagnostic groups, with functional observations collected over six anatomical regions. Each data-set is considered under two signal-to-noise ratio scenarios, SNR $= 0.2, 1.0$. We assume observations are defined on a common time-grid $\in [0, 1]$ with random missing patterns, where between *zero* and 80% of the time-points are discarded to mimic observed data. Results are summarized in Table 1.

All three priors recover the mean structure equally well under both SNR scenarios. Independently of the generating dependence structure, average relative squared errors improve as more subjects are included for analysis. More meaningful differences are observed when we consider recovery of the covariance structure. When data are generated under separable covariance, *SS* priors exhibit the best performance, with improved accuracy for larger sample sizes. Some loss in efficiency is observed when considering *NS* priors. However, even with only 10 subjects the loss in efficiency is estimated to be only about 30%, reducing to about 4% when sample size escalates to $n = 50$. When data are generated from nonseparably covarying processes, *NS* priors exhibit the highest efficiency and are the only model improving meaningfully in accuracy as sample size increases. In all simulation settings, *NB* priors seem to perform poorly in relation to other alternatives.

In summary, our findings agree with well-known results in multivariate and longitudinal data analysis (Wakefield (2013)). Consistent recovery of the mean function seems to be relatively insensitive to the covariance model. Inference and uncertainty quantification about the mean, however, requires correct recovery of the dependence structure encoded in $\Sigma_z$. While

meaningful gains in efficiency can be achieved if separability assumptions are warranted in applications, the encompassing regularized prior (*NS*) leads to a generally more flexible and efficient modeling framework without the need for stringent structural assumptions.

Extended results, including sensitivity to model specification are reported in the Supplementary Material (Li et al. (2020a)). While theoretical large sample results are not examined in the manuscript, all simulation results can be reproduced and extended with our R package and supplementary documentation.

## 5. EEG and language acquisition in TD and ASD infants.

5.1. *Study background.* We consider a functional brain imaging study carried out by our collaborators in the Jeste laboratory at UCLA. The study aims to characterize differential functional features associated with language acquisition in TD and ASD children. EEG data were recorded for 144s using an 128 electrode HydroCel Geodesic Sensor Net for nine TD, 32 ASD children ranging between 4 and 12 years of age. The EEG data is divided into nonoverlapping segments of 1.024 seconds, producing a maximum of 140 observable segments for each subject at each electrode. Details about data preprocessing are deferred to Web Appendix A. Individual sensors were partitioned between 11 anatomical regions made up of four to seven electrodes; left and right for the temporal region (LT and RT) and left, right and middle for the frontal, central and posterior regions (LF, RF, MF, LC, RC, MC, LP, RP and MP, respectively). Region-level power dynamics were estimated as outlined in Section 2. Figure 1 illustrates a sample of smoothed power trajectories within the *alpha* and *gamma* frequency bands for both TD and ASD children. ASD children were further classified as verbal ASD (vASD—14 children) and minimally verbal ASD (mvASD—19 children) in relation to their verbal developmental quotients (vDQ).

The broad scientific goal of this study aims to understand the functional underpinnings associated with language acquisition through the human ability to parse otherwise continuous speech streams into meaningful words. In order to replicate this natural phenomenon, children were exposed to continuous synthetic speech constructed through a collection of 12 phonemes. By defining phoneme triplets (e.g., *pa-bi-ku*, *da-ro-pi*) as deterministic pseudowords and exposing children to random continuous permutations of pseudowords, study subjects were given the opportunity for implicit statistical learning of word segmentation. During this process we seek to detect differential neurophysiological response across anatomical regions and oscillation frequencies.

5.2. *Group mean trajectory analysis.* Our analysis focuses on time-varying, region-referenced differences among diagnostic groups (TD, vASD, mvASD) in relation to two frequency bands, namely, *alpha* (8–15 Hz) and *gamma* (32–50 Hz). Specifically, alpha waves are thought to play an active role in network coordination and communication (Fink and Benedek (2014)), while gamma waves are thought to index conscious perception and tend to correlate with implicit learning processes (Gruber and Müller (2002)).

Our analyses are based on the nonseparable model in (3.6), with 10 latent factors, and data projected on 12 B-splines basis functions. Main findings across four of 11 regions are summarized graphically in Figure 2 for alpha waves and in Figure 3 for gamma waves. In both models the varying coefficients, representing the mean structure in 3.3, take as input a simple factorial covariate indexing diagnostic group membership and yield group-level means across anatomical regions. For each time-varying mean we also represent simultaneous credible bands by probability shading to include a gradient of 0.2, 0.6 and 0.9 posterior coverage. A brief discussion about sensitivity to the prior model, number of latent factors and projection methods is deferred to Section 6.
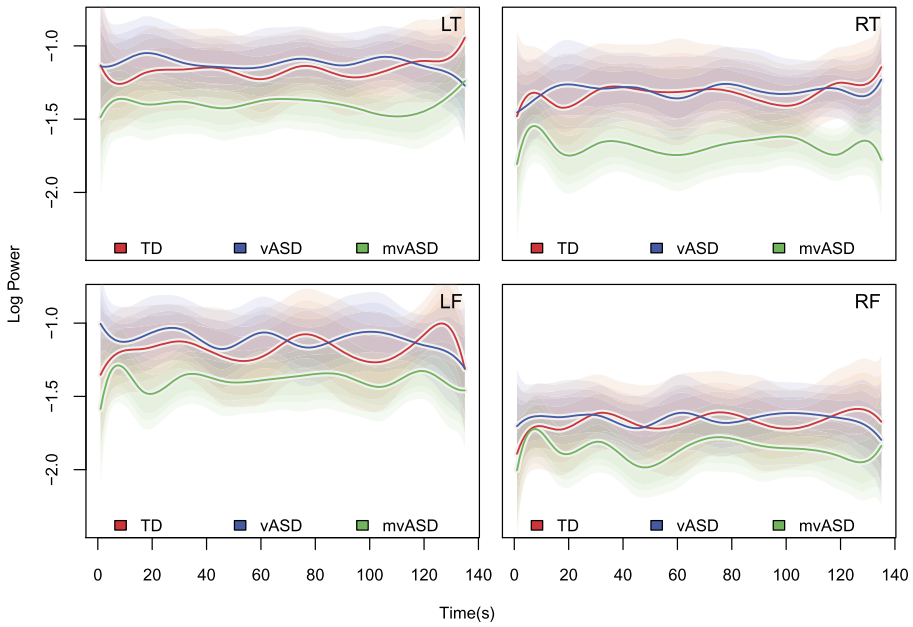
FIG. 2. Alpha power dynamics. *Average log power dynamics stratified by diagnostic group*: *Typically developing* (*TD*), *verbal ASD* (*vASD*) *and minimally verbal ASD* (*mvASD*). *Four of* 11 *anatomical regions are represented, namely*: *Left frontal* (*LF*), *right frontal* (*RF*), *left temporal* (*LT*) *and right temporal* (*RT*). *In each region we display the posterior mean, accompanied by simultaneous credible bands, shaded to include* 0.2, 0.6 *and* 0.8 *of all posterior samples.*
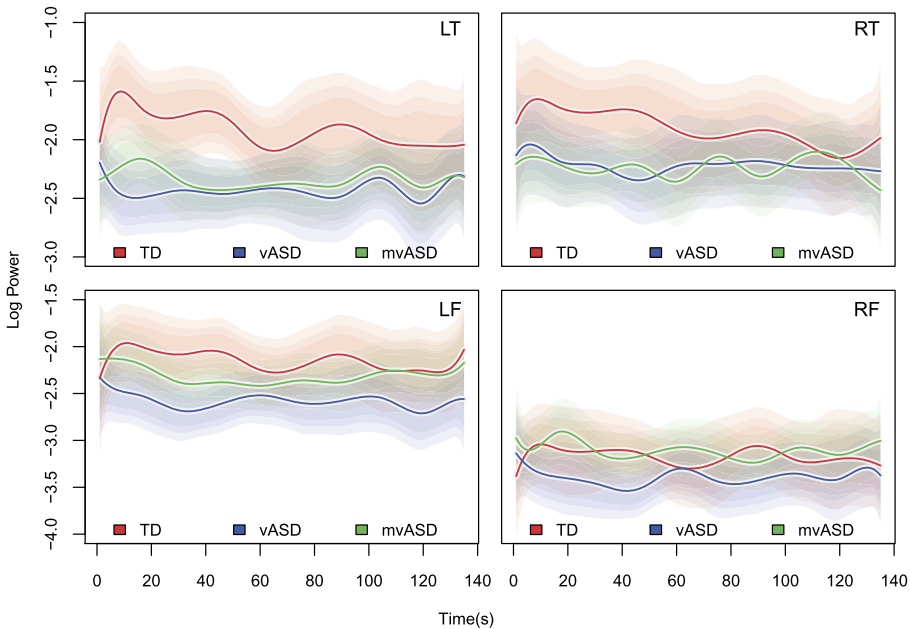


FIG. 3. Gamma power dynamics. *Average log power dynamics stratified by diagnostic group*: *typically developing* (*TD*), *verbal ASD* (*vASD*) *and minimally verbal ASD* (*mvASD*). *Four of* 11 *anatomical regions are represented, namely*: *left frontal* (*LF*), *right frontal* (*RF*), *left temporal* (*LT*) *and right temporal* (*RT*). *In each region we display the posterior mean, accompanied by simultaneous credible bands, shaded to include* 0.2, 0.6 *and* 0.8 *of all posterior samples.*

Starting with a group-by-region analysis of the alpha frequency band in Figure 2, we note that, for all three groups and across all 11 brain regions, there is no discernible time-varying pattern in the average relative alpha frequency which is maintained relatively flat across time-in-trial. The TD and vASD groups are, essentially, indistinguishable with minor differences likely due to sampling variability. Crucially, meaningfully lower levels in the average relative alpha frequency are observed for the mvASD group. While we fall short of considering formal notions of statistical significance, we highlight potentially meaningful findings in the RT region, where 90% simultaneous bands fail to overlap for some of the time-on-trial. More importantly, the lower level in relative alpha frequency is consistent across most brain regions, suggesting a potential role of alpha waves in distinguishing brain functional characteristics in the minimally verbal ASD group. Finally, within group, for all diagnostic groups we confirm a well-known phenomenon, known as left-right alpha asymmetry, which is significant over time at frontal regions, diminished at temporal regions and negligible at central and posterior regions.

Replicating the same analysis for gamma frequencies in Figure 2, we note that, while alpha waves seem to isolate the mvASD group, gamma waves tend to separate TD children from the vASD and mvASD groups. More precisely, the TD group exhibits significantly higher average gamma band power than vASD and mvASD, especially at the early stages of the experiment. This pattern is most evident at temporal regions, both left and right (LT, RT). Furthermore, the average gamma band power for TD children exhibits a sharp increase at the beginning of time-on-trial, followed by a decreasing trend through the end of the experiment. This finding suggests that, differently from the ASD groups, TD children consciously perceive the beginning of the speech stream as a stimulus. Compatible findings by Gruber and Müller (2002) report a decreased response when stimuli recur, which they projected to be linked to a "neural savings" mechanism within a cell assembly representing an object, that is, a word in our case. As observed for alpha band power, we note significant frontal asymmetry throughout the study and across diagnostic groups. While previous studies, for example, Rojas and Wilson (2014), pointed out left-dominant asymmetry as a discriminative feature between TD and ASD cohorts, our findings do not replicate this observation within the context of a language acquisition experiment.

We compared the NB, SS and NS prior models, by computing their expected log pointwise predictive density leave-one-out cross-validation (elpd-loo)(Vehtari, Gelman and Gabry (2017), Watanabe (2013), Vehtari, Gelman and Gabry (2015)). In the context of our case study, let $p\{Y_i(t) \mid Y_{-i}(t)\}$ be the leave-one-out predictive density for subject $i$ ($i = 1, \ldots, n$). The elpd-loo is computed as follows:

$$(5.1) \qquad \text{elpd}_{\text{loo}} = \sum_{i=1}^{n} \log p(Y_i(t)|Y_{-i}(t)).$$

Treating gamma band power as the response, NB has $\text{elpd}_{\text{loo}} = -11{,}586$ (SE $= 947$), SS with six latent factors for the time and spatial domains has $\text{elpd}_{\text{loo}} = -10{,}981$ (SE $= 911$) and NS with 10 latent factors has $\text{elpd}_{\text{loo}} = -10{,}951$ (SE $= 910$). Treating alpha power as the response, NB has $\text{elpd}_{\text{loo}} = -14{,}698$ (SE $= 1227$), SS has $\text{elpd}_{\text{loo}} = -14{,}211$ (SE $= 1209$) and NS has $\text{elpd}_{\text{loo}} = -14{,}260$ (SE $= 1214$). For both the gamma and alpha power bands, SS and NS tend have a relatively similar performance, while NB seems to underperform in both settings. Our analyses are therefore based on the NS prior, as it is more general and does not seem to induce any loss in predictive power for this case study.

To assess goodness of fit, we compute 90% pointwise and simultaneous (grouping by subject and region) posterior predictive credible bands treating gamma power as the response (Gelman, Meng and Stern (1996)). Simultaneous bands are obtained as in Crainiceanu et al. (2007). Figure 4 displays four random, subject-region gamma power dynamics accompanied
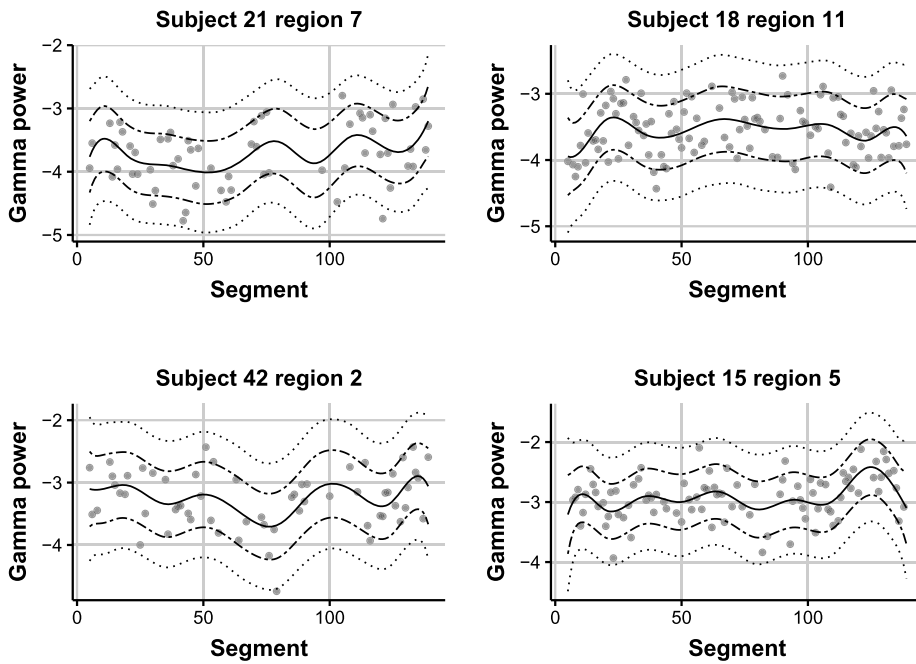
FIG. 4. Gamma power goodness of fit. *Posterior mean trajectory (solid line) for four random subject region responses. We report pointwise (dot-dash lines) and simultaneous (dotted lines) 90% posterior predictive credible bands. Top left: subject 21 (mvASD), region 7. Top right: subject 18 (ASD), region 11. Bottom left: subject 42 (mvASD), region 2. Bottom right: subject 15 (ASD), region 5.*

with maximum a posteriori fit and 90% pointwise posterior prediction bands. The figure illustrates adequate fit and predictive coverage for all four subjects. Over all subjects, the interquartile range (IQR) of predictive coverage rate is (78.4%, 88.5%) and (97.1% to 100%) using pointwise prediction bands and simultaneous prediction bands, respectively. Similar IQR coverage rates are obtained when considering alpha power dynamics, indicating that the model is reasonably well calibrated (Dawid (1985)).

A frequentist test for goodness of fit was applied using the framework developed by Yuan and Johnson (2012). Although standardized residual means are consistently close to zero, the p-values for different quantile thresholding scenarios suggest some evidence of lack-of-fit with the residual variance model. A detailed analysis at two levels of the model hierarchy is reported in the Supplementary Material (Li et al. (2020a)). We note that, while these findings are unlikely to influence inference on population summaries, some care would be needed in predictive settings, where a scale mixture extension of the residual model would likely account for the observed overdispersion with respect to Gaussian sampling.

5.3. *Effects of age and verbal-DQ.* Rather than considering a coarse classification of subjects into TD, vASD and mvASD, we examine how alpha and gamma band power trajectories change as a function of age and verbal DQ. Due to the relatively small-sized sample, the distribution of subjects' demographics among three groups is somewhat unbalanced. To be specific, the TD cohort is significantly older ($94.7 \pm 28.8$, in months) with hider vDQ ($120.6 \pm 11.4$), the v-ASD cohort are younger ($67.1 \pm 58$) with medium vDQ ($89.3 \pm 22.7$) and mv-ASD has a wide age range ($85.6 \pm 24.0$), however, significantly lower by vDQ ($23.6 \pm 10.9$). Our analysis is therefore not perfect but still warranted under the assumption of generalized additive effects. In particular, the mean structure in 3.3 is represented as the additive combination of time-varying coefficients, including an intercept function, a co-
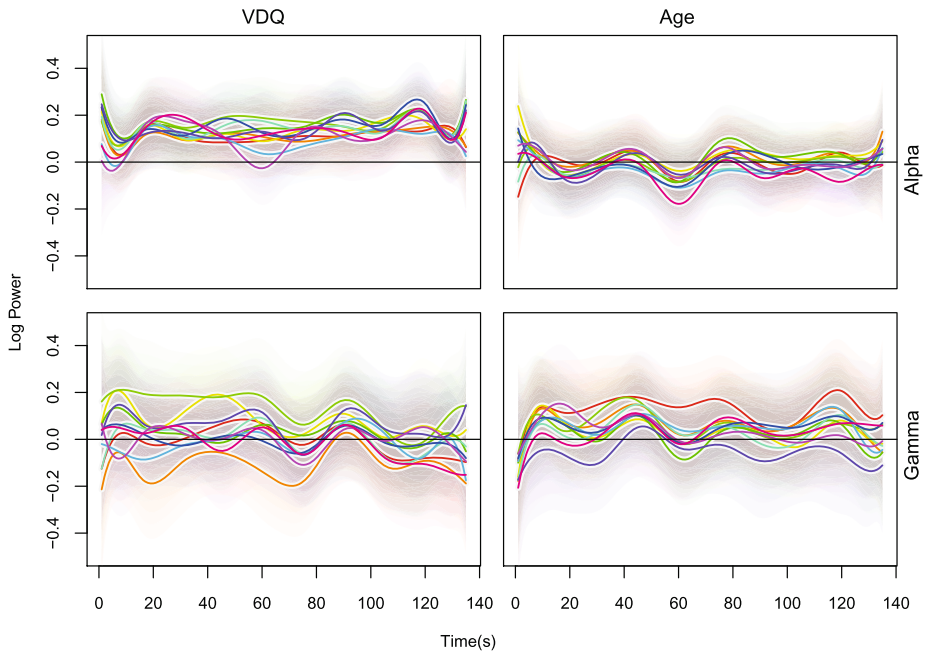
FIG. 5. *Varying coefficients for age and vDQ. Varying coefficients associated with age and verbal DQ across all anatomical regions for the gamma and alpha frequency bands. Each varying coefficient is accompanied by simultaneous credible bands*, *shaded to include* 0.2, 0.6 *and* 0.9 *of all posterior samples.*

efficient function for the main effect of age and a coefficient function for the main effect of vDQ.

Our results are summarized graphically in Figure 5. We display the varying coefficients for age and vDQ for both the alpha and gamma band power, across all 11 brain regions. For each curve we include probability shading for simultaneous 0.2, 0.6 and 0.9 credible bands. The effects of age and vDQ on alpha band power is consistent across all brain regions. While age does not seem to be associated with the outcome, higher vDQ levels are consistently and significantly associated with higher alpha band power levels. This finding is consistent with the group-level analysis in Section 5.2, where mvASD children were found to exhibit significantly lower alpha levels, when compared to TD and vASD children. On the other hand, the relationship between vDQ and gamma band power is highly heterogeneous across brain regions, with vDQ positively associated with higher gamma band power in temporal regions (LT and RT) and negatively associated with the outcome in the right frontal regions (RF). The effect of age on gamma power is more consisted across brain regions, with older children exhibiting higher gamma in some regions and significantly so in LF.

**6. Discussion.** We introduced a hierarchical modeling framework for the analysis of region-referenced functional data in the context of functional brain imaging through EEG. Our work hinges on two classical ideas, namely, the constructive definition of Gaussian processes through basis functions and a flexible representation of regularized dependence within and across anatomical regions through latent factor expansions. Our proposal is implemented within a high-performance computation R package, supporting three prior models, which include both separable and nonseparable covariance structures for region-referenced functional observations. We showed that the proposed approach has satisfactory operating characteristics under extensive numerical experiments as well as appealing inferential properties due to straightforward handling of posterior functionals.

The application of our method relies on projections into the space spanned by basis functions. This feature is engineered into our proposal in order to ensure expert knowledge can be included in the analysis by not limiting the applicability of our method to B-spline projections but, potentially, including functional spaces spanned, for example, by wavelets or periodic functions which may be more appropriate in some applications of multivariate functional data analysis. Specific modeling choices, like the number and placement of spline knots, can be based on straightforward calculation of information criteria (Gelman, Hwang and Vehtari (2014)). Similar considerations apply to the choice of prior model, including restrictions on the structure of large covariance oparators. Our implementation depends on choosing the number of latent factors encoding regularized estimation of covariance operators. From our simulation experiments and data analysis, we conclude that regularization through product Gamma priors yields results which are robust to this specific choice. In practice, the number of latent factors included in the analysis can be compared to the number of principal component functions included in a standard FPCA analysis. A crucial difference in our modeling approach is that we interpret regularized estimation through continuous penalization rather than discrete truncation to a specific number of principal components. Therefore, we recommend relative overparametrization by selecting a larger number of latent factors in default analyses or by performing formal Bayesian model selection for the number of latent factors (Vehtari, Gelman and Gabry (2017)).

Our case study, considering region-referenced EEG data, shows how complex data structures may be analyzed within the familiar hierarchical modeling framework, coupled with varying coefficient models. Our work focused on inference for the mean structure of region-referenced functions. However, important strides can be made in a formal characterization of mean and covariance dependence on predictors. This is notably relevant in the field of functional brain imaging, where large levels of subject heterogeneity are often observed. We note that, for the covariance structure, simple group comparisons are indeed possible within the proposed framework by simply running separate analyses. Several questions involving regional and functional correlation are, therefore, easily answered from a straightforward examination of posterior summaries. A promising, perhaps more principled approach would, for example, include a formal representation of covariance heterogeneity through differential subspace structures (Franks and Hoff (2019)).

Finally, our work discusses statistical significance only informally through pairwise comparisons of simultaneous credible bands within anatomical regions. Within the context of multivariate functional data analysis, more work on the meaning and construction of uncertainty bounds is, however, needed with formal considerations of multiplicity within structured dependence settings.

A user-friendly R implementation, with examples, is available in the Supplementary Materials (Li et al. (2020b)) and online as an Rcpp package at https://github.com/Qian-Li/HFM.

## SUPPLEMENTARY MATERIAL

**Supplement to "Region-referenced spectral power dynamics of EEG signals: A hierarchical modeling approach"** (DOI: 10.1214/20-AOAS1374SUPPA; .pdf). Additional analytical details, extended simulation studies and analyses.

**Source code for "region-referenced spectral power dynamics of EEG signals: A hierarchical modeling approach"** (DOI: 10.1214/20-AOAS1374SUPPB; .zip). R package implementation for the methods described in the paper.

# REFERENCES

BALADANDAYUTHAPANI, V., MALLICK, B. K. and CARROLL, R. J. (2005). Spatially adaptive Bayesian penalized regression splines (P-splines). *J. Comput. Graph. Statist.* **14** 378–394. MR2160820 https://doi.org/10.1198/106186005X47345

BALADANDAYUTHAPANI, V., MALLICK, B. K., HONG, M. Y., 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, 321–322. MR2422820 https://doi.org/10.1111/j.1541-0420.2007.00846.x

BHATTACHARYA, A. and DUNSON, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika* **98** 291–306. MR2806429 https://doi.org/10.1093/biomet/asr013

BRILLINGER, D. R. (1981). *Time Series*: *Data Analysis and Theory*, 2nd ed. *Holden-Day Series in Time Series Analysis*. Holden-Day, Inc., Oakland, CA. MR0595684

BROYD, S. J., DEMANUELE, C., DEBENER, S., HELPS, S. K., JAMES, C. J. and SONUGA-BARKE, E. J. (2009). Default-mode brain dysfunction in mental disorders: A systematic review. *Neurosci. Biobehav. Rev.* **33** 279–296.

CHEN, K. and LYNCH, B. (2017). Weak separablility for two-way functional data: Concept and test. arXiv preprint arXiv:1703.10210.

CHIOU, J.-M., CHEN, Y.-T. and YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statist. Sinica* **24** 1571–1596. MR3308652

CRAINICEANU, C. M., RUPPERT, D., CARROLL, R. J., JOSHI, A. and GOODNER, B. (2007). Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *J. Comput. Graph. Statist.* **16** 265–288. MR2370943 https://doi.org/10.1198/106186007X208768

CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494 https://doi.org/10.2307/2669946

DAWID, A. P. (1985). Calibration-based empirical probability. *Ann. Statist.* **13** 1251–1285. With discussion by Mark J. Schervish and a reply by the author. MR0811493 https://doi.org/10.1214/aos/1176349736

EIGSTI, I.-M., DE MARCHENA, A. B., SCHUH, J. M. and KELLEY, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders* **5** 681–691.

FINK, A. and BENEDEK, M. (2014). EEG alpha power and creative ideation. *Neurosci. Biobehav. Rev.* **44** 111–123.

FRANKS, A. M. and HOFF, P. (2019). Shared subspace models for multi-group covariance estimation. *J. Mach. Learn. Res.* **20** Paper No. 171, 37. MR4048982

GELMAN, A., HWANG, J. and VEHTARI, A. (2014). Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24** 997–1016. MR3253850 https://doi.org/10.1007/s11222-013-9416-2

GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. With comments and a rejoinder by the authors. MR1422404

GOU, Z., CHOUDHURY, N. and BENASICH, A. A. (2011). Resting frontal gamma power at 16, 24 and 36 months predicts individual differences in language and cognition at 4 and 5 years. *Behav. Brain Res.* **220** 263–270.

GREVEN, S., CRAINICEANU, C., CAFFO, B. and REICH, D. (2010). Longitudinal functional principal component analysis. *Electron. J. Stat.* **4** 1022–1054. MR2727452 https://doi.org/10.1214/10-EJS575

GRUBER, T. and MÜLLER, M. M. (2002). Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human EEG. *Cogn. Brain Res.* **13** 377–392.

HAPP, C. and GREVEN, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Amer. Statist. Assoc.* **113** 649–659. MR3832216 https://doi.org/10.1080/01621459.2016.1273115

JAMES, G. M., HASTIE, T. J. and SUGAR, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602. MR1789811 https://doi.org/10.1093/biomet/87.3.587

KUHL, P. K. (2004). Early language acquisition: Cracking the speech code. *Nat. Rev., Neurosci.* **5** 831–843.

LANG, S. and BREZGER, A. (2004). Bayesian P-splines. *J. Comput. Graph. Statist.* **13** 183–212. MR2044877 https://doi.org/10.1198/1061860043010

LI, Q., SHAMSHOIAN, J., ŞENTÜRK, D., SUGAR, C., JESTE, S., DISTEFANO, C. and TELESCA, D. (2020a). Supplement to "Region-referenced spectral power dynamics of EEG signals: A hierarchical modeling approach." https://doi.org/10.1214/20-AOAS1374SUPPA

LI, Q., SHAMSHOIAN, J., ŞENTÜRK, D., SUGAR, C., JESTE, S., DISTEFANO, C. and TELESCA, D. (2020b). Source code for: "Region-referenced spectral power dynamics of EEG signals: A hierarchical modeling approach" https://doi.org/10.1214/20-AOAS1374SUPPB

LORD, C., RISI, S., LAMBRECHT, L., COOK, E. H., LEVENTHAL, B. L., DILAVORE, P. C., PICKLES, A. and RUTTER, M. (2000). The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30** 205–223.

LYNCH, B. and CHEN, K. (2018). A test of weak separability for multi-way functional data, with application to brain connectivity studies. *Biometrika* **105** 815–831. MR3877867 https://doi.org/10.1093/biomet/asy048

MILLS, D. L., COFFEY-CORINA, S. A. and NEVILLE, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *J. Cogn. Neurosci.* **5** 317–334.

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. MR3040013 https://doi.org/10.1111/j.1541-0420.2012.01788.x

MURIAS, M., WEBB, S. J., GREENSON, J. and DAWSON, G. (2007). Resting state cortical connectivity reflected in EEG coherence in individuals with autism. *Biological Psychiatry* **62** 270–273.

NISHIMURA, A. and SUCHARD, M. A. (2018). Prior-preconditioned conjugate gradient method for accelerated Gibbs sampling in "large n & large p" sparse Bayesian regression. Technical Report, arXiv preprint arXiv:1810.12437.

OMBAO, H. and HO, M. R. (2006). Time-dependent frequency domain principal components analysis of multichannel non-stationary signals. *Comput. Statist. Data Anal.* **50** 2339–2360. MR2225573 https://doi.org/10.1016/j.csda.2004.12.011

PARK, S. Y. and STAICU, A.-M. (2015). Longitudinal functional data analysis. *Stat* **4** 212–226. MR3405402 https://doi.org/10.1002/sta4.89

ROJAS, D. C. and WILSON, L. B. (2014). $\gamma$-Band abnormalities as markers of autism spectrum disorders. *Biomark. Med.* **8** 353–368.

SABY, J. N. and MARSHALL, P. J. (2012). The utitlty of EEG band power analysis in the study of infancy and early childhood. *Dev. Neuropsychol.* **37** 253–273.

SCHEFFLER, A., TELESCA, D., LI, Q., SUGAR, C. A., DISTEFANO, C., JESTE, S. and ŞENTÜRK, D. (2020). Hybrid principal components analysis for region-referenced longitudinal functional EEG data. *Biostatistics* **21** 139–157. MR4043850 https://doi.org/10.1093/biostatistics/kxy034

SHAMSHOIAN, J., SENTURK, D., JESTE, S. and TELESCA, D. (2019). Bayesian analysis of multidimensional functional data.

UHLHAAS, P. J. and SINGER, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nat. Rev., Neurosci.* **11** 100–113.

VEHTARI, A., GELMAN, A. and GABRY, J. (2015). Pareto smoothed importance sampling. arXiv preprint arXiv:1507.02646.

VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

WAKEFIELD, J. (2013). *Bayesian and Frequentist Regression Methods*. *Springer Series in Statistics*. Springer, New York. MR3025225 https://doi.org/10.1007/978-1-4419-0925-1

WATANABE, S. (2013). A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **14** 867–897. MR3049492

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

YUAN, Y. and JOHNSON, V. E. (2012). Goodness-of-fit diagnostics for Bayesian hierarchical models. *Biometrics* **68** 156–164. MR2909864 https://doi.org/10.1111/j.1541-0420.2011.01668.x

ZHOU, L., HUANG, J. Z., MARTINEZ, J. G., MAITY, A., BALADANDAYUTHAPANI, V. and CARROLL, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *J. Amer. Statist. Assoc.* **105** 390–400. MR2757206 https://doi.org/10.1198/jasa.2010.tm08737

ZHU, H., LI, R. and KONG, L. (2012). Multivariate varying coefficient model for functional responses. *Ann. Statist.* **40** 2634–2666. MR3097615 https://doi.org/10.1214/12-AOS1045