Robust and Gaussian Spatial Functional Regression Models for Analysis of Event-Related Potentials

Hongxiao Zhu[a,*], Francesco Versace[b], Paul M. Cinciripini[b], Philip Rausch[c], Jeffrey S. Morris[d]

[a] *Department of Statistics, Virginia Tech, Blacksburg, VA, USA*
[b] *Department of Behavioral Science, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA*
[c] *Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany*
[d] *Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA*

Abstract

Event-related potentials (ERPs) summarize electrophysiological brain response to specific stimuli. They can be considered as correlated functions of time with both spatial correlation across electrodes and nested correlations within subjects. Commonly used analytical methods for ERPs often focus on pre-determined extracted components and/or ignore the correlation among electrodes or subjects, which can miss important insights, and tend to be sensitive to outlying subjects, time points or electrodes. Motivated by ERP data in a smoking cessation study, we introduce a Bayesian spatial functional regression framework that models the entire ERPs as spatially correlated functional responses and the stimulus types as covariates. This novel framework relies on mixed models to characterize the effects of stimuli while simultaneously accounting for the multilevel correlation structure. The spatial correlation among the ERP profiles is captured through basis-space Matérn assumptions that allow either separable or nonseparable spatial correlations over time. We induce both adaptive regularization over time and spatial smoothness across electrodes via a correlated normal-exponential-gamma (CNEG) prior on the fixed effect coefficient functions. Our proposed framework includes both Gaussian models as well as robust models using heavier-tailed distributions to make the regression automatically robust to outliers. We introduce predictive methods to select among Gaussian vs. robust models and models with separable vs. non-separable spatiotemporal correlation structures. Our proposed analysis produces global tests for stimuli effects across entire time (or time-frequency) and electrode domains, plus multiplicity-adjusted pointwise inference based on experimentwise error rate or false discovery rate to flag spatiotemporal (or spatio-temporal-frequency) regions that characterize stimuli differences, and can also produce inference for any prespecified waveform components. Our analysis of the smoking cessation ERP data set reveals numerous effects across different types of visual stimuli.

Keywords: Bayesian methods; Event-related potential; Functional data analysis; Functional mixed models; Functional regression; Correlated Normal-Exponential-Gamma.

*Correspondence to: Department of Statistics (MC0439), 250 Drillfield Drive, Virginia Tech, Blacksburg, VA 24061 USA
Email address: hongxiao@vt.edu (Hongxiao Zhu)

1. Introduction

Event-related potentials (ERPs) summarize electrophysiological brain responses to specific stimuli. They are generated by averaging electroencephalogram (EEG) segments recorded under repeated applications of a stimulus, with the averaging serving to reduce biological noise levels. ERPs represent temporal changes of electrical potential resulting from the firing of neurons in the brain, measured on a set of electrodes placed on the scalp. They have been widely used to assess brain cognition and information processing (Brandeis and Lehmann, 1986; Bressler, 2002). ERP studies produce for each electrode a waveform on a very fine temporal scale, which is sometimes represented using time-frequency representations such as spectrograms.

In cognitive neuroscience, psychophysiology, and related fields, analytical approaches on ERPs primarily focus on *ERP components*—waveforms with positive or negative voltage deflections (e.g., peaks or valleys). For example, the first peak with a negative voltage deflection occurring about 100 milliseconds (ms) after the onset of a stimulus is called the N100 (or N1) component, and the positive deflection peak occurring near 250–400 ms after the onset of a stimulus is called the P300 (or P3) component. These ERP components are often summarized by features such as the amplitude of the peak or the mean voltages in a time window. Based on these features, statistical analyses, such as analysis of variance (ANOVA) (Lamy et al., 2008; Lole et al., 2013), hypothesis testing (Cagy et al., 2006), regression (Itier et al., 2004; Vossen et al., 2011), classification (Venturini et al., 1992; Zhang et al., 2014), and clustering (Gonzalez-Rosa et al., 2011), are carried out to discover meaningful patterns.

While meaningful results have been found using this approach, limiting analyses to extracted components can be problematic and result in loss of information or false discoveries. First, any results in the data not contained in the pre-chosen components will be lost. Second, it is challenging to capture these components, as they do not occur at precisely the same time for each trial or subject, and so their estimation can attenuate the effect if the optimal location is not chosen or can lead to inflated type I error if locations are chosen to maximize the stimulus-induced signal (Kappenman and Luck, 2016). Third, this approach is typically used while modeling electrodes independently, while they are clearly correlated with each other, and as we show in our simulations failure to model this correlation can result in a loss of efficiency in estimation and inference. Fourth, these approaches often fail to produce global tests across all electrodes or time points, or account for the inherent multiple testing issue raised by performing inference across multiple components, time points and/or electrodes; such problems are exacerbated if multiple electrodes are analyzed and only those with the largest stimuli effects presented.

An alternative to this feature extraction approach is to analyze each electrode and time point (or time-frequency point) independently, which has been termed a mass univariate approach (MUA; Kiebel and Friston,

2004a). A notable work is the LIMO EEG package produced by Pernet et al. (2011) for two-level analysis. MUA is typically coupled with post-hoc smoothing of resulting t-statistics or p-values and adjustment for multiple testing via random field theory to control family-wise error rate (FWER). This approach can be effective, but by modeling electrodes, time points or time-frequency points independently, does not enable more global testing (Kiebel and Friston, 2004b) and can sacrifice efficiency relative to methods that account for these correlations.

Functional data analysis (FDA; Ramsay and Silverman, 1997) treats functions as objects, and accounts for correlation and regularity within functional objects using basis function representations and penalization, which can yield increased efficiency and greater inferential possibilities over methods that do not capture the intrafunctional correlation. Various FDA approaches have been introduced for the analysis of ERP data, typically modeling the temporal waveforms as the functional objects, and using a functional mixed model (FMM) to regress the ERP on the stimulus while adjusting for other factors. Kiebel and Friston (2004b) present hierarchical regression approaches that model the temporal waveforms using wavelet basis functions, using independent models per electrode, and yielding pointwise inference in the time or time-frequency domain. Wang et al. (2009) present a FMM for ERP data, including stimulus, electrode, stimulus $\times$ electrode as fixed effect functions along with subject-specific random effect functions and independent and identically distributed (iid) residual errors. They represent these functional effects through B-splines, and use functional ANOVA to perform global inference of whether the stimulus has any effect or not. Their approach does not, however, provide pointwise inference for individual time points or adjust for multiple testing, assumes iid residual errors, and has been applied to models with only a few selected electrodes. Davidson (2009) applies a Gaussian FMM to ERP data to each electrode separately using wavelet bases to represent the functions using the Bayesian method introduced in Morris and Carroll (2006), obtaining pointwise inference in the time domain that adjusts for multiple testing using false discovery rate (FDR). None of these methods model inter-electrode correlation, include both global and local inference with options for multiple testing adjustment by both FWER and FDR, or perform robust regression that can adjust for potential outlying time points, frequencies, or electrodes. Hasenstab et al. (2017) present methods to decompose the total variability of ERP data for a given scalp region into subject-specific and electrode-within-subject components, as well a component across scalp regions if multiple scalp regions are modeled, using multi-level functional principal components (fPC) to empirically estimate basis functions at each level. These methods provide an interesting approach for capturing the key structure of ERP data, but do not present regression models incorporating stimuli effects or perform inference to identify differences across experimental conditions.

In this paper, we present a Bayesian functional mixed model approach to model ERP data. This approach can account for nonstationary inter-electrode correlation, induces smoothness across electrodes in

the regression surfaces, is potentially robust to outlying curves or regions, and provides both global and pointwise inference to detect stimuli effects. To our knowledge, no other existing method for ERP data has all of these characteristics. Our framework treats the time or time-frequency waveforms as functional objects that are spatially correlated with nearby electrodes, and regresses these functions on any specified covariates with regression surfaces that are smooth in both time and space (i.e. across electrodes). Our simulations show that accounting for this correlation when present leads to greater power for detecting stimulus-induced effects. Our proposed framework utilizes either Gaussian models or robust models with heavier-tailed distributions when outliers are present, and can accommodate either separable or nonseparable inter-electrode spatial correlation parameterized by a Matérn structure. It yields fully Bayesian inference that can be used to perform a global test for stimulus effect across time or time-frequency and electrodes, and then localize any differences in the time or time-frequency and electrode domains, and if desired, can also test any prespecified waveform components that may be of interest, while adjusting for multiple testing using FWER or FDR criteria. The resulting continuous spatiotemporal effects help characterize EEG/ERP *microstates*—a sequence of quasi-stable spatial distributions (landscapes) connected by quick changes in landscapes (Lehmann et al., 2009; Milz et al., 2016). We present rigorous Bayesian model selection techniques to assess whether the Gaussian or robust model should be used, and whether the inter-electrode spatial correlation is needed and, if so, whether they should be separable or non-separable with time. The modeling framework we present can be considered to capture advantages of the existing modeling approaches—modeling the entire ERP data like MUA approaches, accounting for temporal correlation structure like the FDA methods, providing inference on prespecified time or time-frequency components like feature extraction approaches, while accounting for nonstationary inter-electrode correlation and achieve robustness to outliers.

While presented in the context of ERP data, the methods we introduce are general and can be applied to many other spatially correlated functional data sets, thus also contribute to the literature of functional regression. Functional regression has experienced rapid development in recent years (Morris, 2015). Comparing with existing methods, our proposed framework offers several unique features and advantages: (1) It simultaneously models fixed/random covariate effects and non-separable spatial correlation of the functions. In contrast, existing methods either only model complex spatiotemporal/multi-level correlation structures while not including the effects of covariates (Greven et al., 2010; Chen and Müller, 2012; Park and Staicu, 2015; Chen et al., 2017; Chen and Lynch, 2017; Hasenstab et al., 2017), or simply treat spatiotemporal information as covariates for fixed or random effects thus do not directly characterize correlations induced by spatial/temporal distances (Scheipl et al., 2015; Brockhaus et al., 2015; Scheipl et al., 2016). (2) It induces both adaptive regularization over time and spatial smoothness over electrodes in the functional regression coefficients, while most existing approaches either do not induce adaptive regularization (Staicu et al., 2010) or do not allow fixed effect functions to be spatially correlated (Morris and Carroll,

2006; Baladandayuthapani et al., 2008; Davidson, 2009; Steen, 2010; Zhou et al., 2010). (3) It provides an option to perform robust functional regression that is insensitive to outliers, while in existing methods, only a few consider robust regression (Zhu et al., 2011; Brockhaus et al., 2015; Scheipl et al., 2016). (4) Additionally, our Bayesian framework yields a rich set of inferential outputs including global or local tests for any transformation of model parameters, and adjusting for multiple testing using EWER or FDR criterion. It also includes model selection methods to determine Gaussian vs. robust models and models with separable vs. nonseparable spatiotemporal correlation structures.

2. Materials and Methods

2.1. The Smoking Cessation Study and the ERP Data

The ERP data studied in this paper were collected from a sub-study of a randomized clinical trial on smoking cessation (Cinciripini et al., 2013). This sub-study measures neurological responses to emotional cues in smokers under four types of visual stimuli—cigarette, pleasant, unpleasant and neutral. Investigators aim to test for systematic differences across the stimuli types and characterize any differences spatially (across scalp regions) and temporally. One hypothesis is that in nicotine-addicted individuals, cigarette-related cues will elicit ERPs comparable to those observed in the presence of the positive emotional stimuli.

EEG signals were recorded using a 129-electrode Geodesic Sensor Net (Geodesic EEG System 200; Electrical Geodesics Inc., Eugene, OR) during the presentation of pictures with pleasant, unpleasant, neutral, or cigarette-related content. Preprocessing of the EEG signals was then conducted; steps included high-pass and low-pass filtering, artifact removal, eye blink correction, as well as average re-referencing. More details can be found in Versace et al. (2011). The EEG signals were further segmented on the time interval $[-100, 800]$ ms with one measurement point for every 4 ms. The time zero indicates the onset of the picture. To increase signal-to-noise ratio, the signals for each subject were averaged together across the 24 replicate pictures for each stimulus type to produce ERP temporal waveforms. It would be possible to construct time-frequency representations from these data using spectrograms (Holan et al., 2010), multitapering (Maris and Oostenveld, 2007), or smooth localized complex exponential basis functions (SLEX; Ombao et al., 2002), but in this paper we focus on temporal waveforms. The preprocessing steps produce ERPs at $S = 129$ electrodes for each of the four stimulus types for each of the $M = 180$ subjects. The total number of ERP curves is $92,880$, and each curve contains measurements at $T = 225$ time points, resulting in a very large data set with > 20 million observations. In Figure 1(a), sample ERP waveforms from the first 10 subjects are plotted as grey lines for 16 selected electrodes, and the colored lines are the sample average for each of the four stimulus types calculated across all subjects. Figure 1(b) shows the layout of all 129 electrodes, partitioned into 11 cortical regions following Keil et al. (2002).

A special characteristic of ERP signals is the correlation induced by spatial locations of the electrodes. Figure 1(c) plots the correlation between pairs of electrodes (in the left central (R5) and the occipital

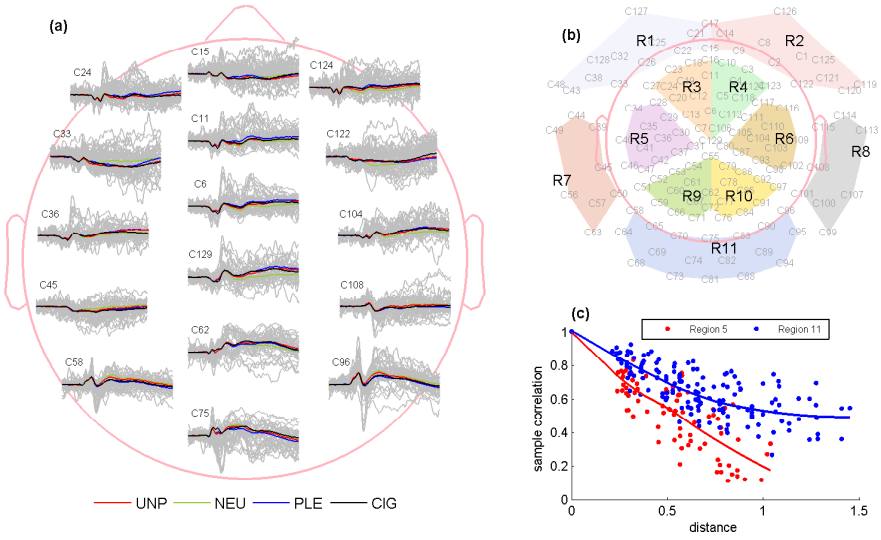

Figure 1: ERP plots: (a) ERP curves at 16 electrodes for 10 subjects. Colored curves are sample averages for the four stimuli. (b) Partition of the 129 sites into 11 regions: anterior frontal left/right (R1/R2), frontal left/right (R3/R4), central left/right(R5/R6), temporal left/right (R7/R8), parietal left/right (R9/R10) and occipital (R11). (c) Pairwise correlations between electrodes for region 5 (red) and 11 (blue). Each dot represents the Pearson correlation between one pair of electrodes calculated by pooling ERP measurement points across all subjects, stimulus types, and time grids. The lines are smoothed fits using local polynomial kernels.

(R11) cortical regions) as a function of the electrode distances. Figure 1(c) clearly demonstrates that the correlations decay with electrode distances. We aim to capture this spatial correlation structure in our modeling, which, as we will show by simulation, leads to greater sensitivity and specificity for detecting significant stimuli effects in location and time over methods ignoring this spatial correlation.

2.2. Functional Regression with Spatial Correlation

While the methods we introduce are general, here we present the proposed models in the context of ERP data reviewed in Section 2.1. This data set contains functional data with a complex inter-functional correlation structure that has both hierarchical and spatial elements. Suppose that there are M subjects, A stimulus types, and S electrodes. For each electrode, there are $\mathcal{L} = M \times A$ ERPs. Let $Y_{is}(t)$ represent the ith ERP for electrode s, where $i = 1, \ldots, \mathcal{L}$, $s = 1, \ldots, S$, and $t = t_1, \ldots, t_T$. Let $X_{ia} = 1$ if ERP i is from stimulus type a, and 0 otherwise; let $Z_{im} = 1$ if the ith ERP is from subject m, and 0 otherwise. The general functional response regression model we are interested in fitting is:

$$Y_{is}(t) \quad = \quad \sum_{a=1}^{A} X_{ia} B_{as}(t) + \sum_{m=1}^{M} Z_{im} U_m(t) + E_{is}(t), \quad t \in \mathcal{T}, \tag{1}$$

where $\mathcal{T}$ is a closed interval on the real line, $B_{as}(t)$ represents the effect of stimulus type a at electrode s, $U_m(t)$ is a mean-zero random effect function capturing the subject-level variability, and $E_{is}(t)$ is a mean-zero residual error function capturing the variability at the lowest level (i.e., electrode-level) of the hierarchy.

If modeling time-frequency representations instead of temporal waveforms, each functional quantity would simply be written as a function of both time and frequency. Our ultimate goal is to test for differences in $|B_{as}(t) - B_{a's}(t)|, a \neq a'$, determine which regions of the scalp location s and time t are significantly different, and if desired, assess any prespecified waveform components.

This model resembles other functional mixed models (FMMs) in the literature (Guo, 2002; Morris and Carroll, 2006; Zhu et al., 2011). However, in order to adequately capture the structure of ERPs, our model needs to regularize the fixed effect functions $\{B_{as}(t)\}$ over both time t and electrode s, plus account for spatial correlations in the residual errors $\{E_{is}(t)\}$ that may necessarily be nonstationary, i.e. vary over t. Also, ERP data frequently contain outliers, which can be outlying subjects, electrodes, or time points, and these outliers can strongly impact the functional regression results. Existing robust FMMs (Zhu et al., 2011) cannot accommodate any spatial interfunctional correlation in the fixed effect or the residual. Our proposed framework incorporates robust models that successfully accommodate these spatial correlations.

While model (1) is perhaps more intuitive, for the remainder of this paper we will work with a vectorized version of this model. By stacking the functions in model (1), we define $\boldsymbol{Y}(t) = (Y_{11}(t), \ldots, Y_{1S}(t), \ldots, Y_{\mathcal{L}1}(t), \ldots, Y_{\mathcal{L}S}(t))^T$, $\boldsymbol{B}(t) = (B_{11}(t), \ldots, B_{1S}(t), \ldots, B_{A1}(t), \ldots, B_{AS}(t))^T$, $\boldsymbol{U}(t) = (U_1(t), \ldots, U_M(t))^T$, and $\boldsymbol{E}(t) = (E_{11}(t), \ldots, E_{1S}(t), \ldots, E_{\mathcal{L}1}(t), \ldots, E_{\mathcal{L}S}(t))^T$. Model (1) can then be rewritten in vector form:

$$\mathbf{Y}(t) = \mathbf{X}\mathbf{B}(t) + \mathbf{Z}\mathbf{U}(t) + \mathbf{E}(t), \quad t \in \mathcal{T}. \tag{2}$$

Denote by $N = \mathcal{L}S$ the total number of ERPs measured and denote by $p = AS$ the total number of channel-specific fixed effects; then $\mathbf{X}$ in (2) is a $N \times p$ and $\mathbf{Z}$ is a $N \times M$ matrix, both containing only a single "1" in each row.

Basis-Transform Modeling Approach For efficient model fitting, we adopt a basis-transform modeling approach that involves representing the functions with a *lossless* or *near-lossless* basis representation, modeling in the dual space of basis coefficients, and then projecting the results back to the original data space for inference. Given a set of basis functions $\psi_k(t), k = 1, \ldots, K$, we use a truncated basis representation

$$Y_{is}(t) \quad = \quad \sum_{k=1}^{K} Y^*_{isk}\psi_k(t) \tag{3}$$

This transform is said to be *lossless* if $Y_{is}(t) \equiv \sum_k Y^*_{isk}\psi_k(t)$ for all observed t, so that the basis coefficients $\{Y^*_{isk}; k = 1, \ldots, K\}$ contain all information within the observed functional data $\{Y_{is}(t); t = t_1, \ldots, t_T\}$, which is the case for example with a wavelet transform. It is said to be *near-lossless* if

$$\left\| Y_{is}(t) - \sum_{k=1}^{K} Y^*_{isk}\psi_k(t) \right\| < \epsilon \quad \forall i = 1, \ldots, N \text{ and } s = 1, \ldots, S \tag{4}$$

for some small value ϵ and measure $\|\bullet\|$, which can be the case with a truncated wavelet representation

given enough basis functions. Near-losslessness may be sufficient for modeling as this condition assures that the chosen basis is sufficiently rich such that for practical purposes it can recapitulate the observed functional data, and visual inspection of the raw functions and basis transformation should reveal virtually no difference.

Any basis functions can be used, including commonly used splines, wavelets, Fourier bases, eigenfunctions or creatively constructed custom bases, and can be defined on multi-dimensional or non-Euclidean domains. If modeling time-frequency representations, 2D basis functions such as 2D wavelets (Martinez et al., 2013), 2D eigenfunctions (Chen and Jiang, 2017), or SLEX bases (Ombao et al., 2002) could be used. For the temporal ERP waveforms, in this paper we use wavelet bases, as has commonly been done in other papers in ERP literature (Kiebel and Friston, 2004b; Davidson, 2009).

Given a wavelet basis with mother wavelets $\{\psi_{jk}; j = 1, \ldots, J; k = 1, \ldots, K_j\}$ and father wavelets $\{\psi_{0k}; k = 1, \ldots, K_0\}$, we expand $Y(t)$, an element of $\mathbf{Y}(t)$, by $Y(t) = \sum_{j=0}^{J} \sum_{k=1}^{K_j} d_{jk}\psi_{jk}(t)$. Here, $\{d_{jk}\}$ are the wavelet coefficients that describe features of the ERP at scales indexed by j and locations indexed by k. For data on an equally spaced grid, this representation is *lossless* if all basis coefficients are retained, providing an exact representation of the original data. Model (2) can then be transferred to the dual space of wavelet coefficients:

$$\mathbf{D} = \mathbf{X}\mathbf{B}^* + \mathbf{Z}\mathbf{U}^* + \mathbf{E}^*, \tag{5}$$

where rows of $\mathbf{D}, \mathbf{B}^*, \mathbf{U}^*$ and $\mathbf{E}^*$ contain wavelet coefficients of entries in $\mathbf{Y}(t), \mathbf{B}(t), \mathbf{U}(t)$ and $\mathbf{E}(t)$ respectively, and columns are basis coefficients indexed by (j, k). We propose spatially correlated shrinkage priors for $\mathbf{B}^*$ in Section 2.2.2 that lead to adaptive regularization in t and spatial smoothness over s, and propose distributional assumptions for $\mathbf{E}^*$ and $\mathbf{U}^*$ for Gaussian models in Section 2.2.1 and robust models in Section 2.2.3 to accommodate the spatial correlation across electrodes and the correlation induced by the nested data structure.

2.2.1. Gaussian Functional Mixed Models with Spatial Correlation

We will capture the spatial correlation across electrodes through the residual term $\mathbf{E}(t)$. Suppose that the N functions in $\mathbf{E}(t)$ can be partitioned into $\mathcal{L}$ independent sets of correlated blocks, each of size S_l. For example, in the ERP data, $\mathbf{E}(t)$ contains $N = M \times A \times S$ elements. These elements can be partitioned into $\mathcal{L} = M \times A$ independent blocks, each corresponding to one subject-stimulus combination; and the size of each block is $S_l \equiv S$. One can order components in $\mathbf{E}(t)$ into the $\mathcal{L}$ blocks to obtain $\mathbf{E}(t) = (E_{11}(t), \ldots, E_{1S_1}(t), \ldots, E_{\mathcal{L}1}(t), \ldots, E_{\mathcal{L}S_\mathcal{L}}(t))^T$. We will model spatial functional correlation by assuming parametric covariance structures for the basis space residuals $\mathbf{E}^*$, which induces a flexible class of nonstationary correlations back in the data space.

Specifically, we assume a separate Gaussian distribution per basis coefficient (column of $\mathbf{E}^*$), i.e., $\mathbf{E}^*_{jk} \sim$

$N(0, s_{jk}\mathbf{R}_{jk})$ independently across (j,k), where s_{jk} is a scale parameter with an inverse-gamma prior and $\mathbf{R}_{jk}$ is an $N \times N$ block-diagonal correlation matrix given by $I_{MA} \otimes \mathcal{R}_{jk}$ where I_{MA} is an identify matrix of size $M \times A$ and $\mathcal{R}_{jk}$ is an $S \times S$ correlation matrix determined by the correlation parameter $\boldsymbol{\rho}_{jk}$. With a slight abuse of notation, we denote by $\mathcal{R}_{jk}(s,s')$ the correlation between electrodes s and s'. By allowing the correlation parameter to vary over basis coefficients (j,k), this leads to a *nonseparable* correlation structure back in the data space, with $\text{corr}(E_{is}(t), E_{is'}(t')) = \sum_{j,k} \psi_{jk}(t)\mathcal{R}_{jk}(s,s')\psi_{jk}(t')$. In contrast, if one assumes that $\boldsymbol{\rho}_{jk} \equiv \boldsymbol{\rho}$ for all (j,k), we obtain a correlation structure with $\text{corr}(E_{is}(t), E_{is'}(t')) = \mathcal{R}(s,s')\sum_{j,k} \psi_{jk}(t)\psi_{jk}(t')$, which we refer to as a *separable* structure. Note that both types of correlation structures induce nonstationary processes in the data space as the spatial correlation varies with time t in both cases. There are numerous options for the correlation structure $\mathcal{R}_{jk}$ or $\mathcal{R}$ (Stein, 1999). Here, to induce spatial correlation across electrode locations on the scalp, we consider the Matérn structure—a common choice for point-referenced spatial data. In particular, we follow the parameterization of Baladandayuthapani et al. (2008) and Zhou et al. (2010), which assumes the following isotropic correlation structure:

$$\mathcal{R}_{jk}(s,s';\boldsymbol{\rho}_{jk}) = 2^{1-v_{jk}} \left(2\, d(s,s')\, v_{jk}^{1/2}/\alpha_{jk}\right)^{v_{jk}} K_{v_{jk}} \left(2\, d(s,s')\, v_{jk}^{1/2}/\alpha_{jk}\right) /\Gamma(v_{jk}), \quad d(s,s') > 0, \tag{6}$$

where $\boldsymbol{\rho}_{jk} = (\alpha_{jk}, v_{jk}) > 0$, $d(\cdot,\cdot)$ measures the distance (on the scalp surface) between two electrodes for ERPs, and $K_{v_{jk}}(\cdot)$ is the modified Bessel function of the second kind with order v_{jk}. The parameter α_{jk} controls the rate of decay when x increases, and v_{jk} controls the shape of the correlation function when x is small. Following Baladandayuthapani et al. (2008), we assume uniform priors for the elements of $\boldsymbol{\rho}_{jk}$, i.e., $\alpha_{jk} \sim \text{Unif}(0, C_\alpha)$, $v_{jk} \sim \text{Unif}(0, C_v)$ for constants C_α and C_v, and assume that α_{jk}, v_{jk} are mutually independent. The values of C_α and C_v are determined so that all combinations of (α, v) result in positive-definite correlation matrices given the electrode distances and the correlation structure in (6). Under this parameterization, $\boldsymbol{\rho}_{jk}$ or $\boldsymbol{\rho}$ can be updated through a Metropolis-Hastings step; see the supplementary materials for details.

Nested Correlation for ERPs from the Same Subject Besides the spatial correlation across electrodes, there is an additional layer of interfunctional correlation induced by the fact that we obtain separate ERPs for each subject from each stimuli. We accommodate this nested correlation through the random effect function of (2). Let $U_m(t)$ denote the mth entry of $\boldsymbol{U}(t)$. Assume that $U_m(t)$ is a Gaussian process with mean 0 and covariance kernel $Q(\cdot,\cdot)$ independently across m; then $\mathbf{U}^*_m$, the mth row of $\mathbf{U}^*$ in the dual space model (5), satisfies $\mathbf{U}^*_m \sim N(0, \mathbf{Q}^*)$ independently across m. Taking advantage of the whitening property of wavelet transforms, we make a simplified independence assumption between wavelet coefficients in $\mathbf{U}^*_m$ following Morris and Carroll (2006), which gives $\mathbf{Q}^* = \text{diag}(\{q^*_{jk}\})$, inducing nonstationary covariance assumptions in the original functional space with $\text{cov}\{U_m(t), U_m(t')\} = Q(t,t') = \sum_{j,k} \psi_{jk}(t)\mathbf{Q}^*_{jk}\psi_{jk}(t')$.

We use Gfmmc to represent Gaussian FMMs specified above, with $\text{Gfmmc}_{\boldsymbol{\rho}}$ representing a model with

separable correlation in the residual errors and Gfmmc$_{\boldsymbol{\rho}_{jk}}$ representing that with nonseparable correlation. While presented using the Matérn covariance, the Gfmmc models we introduce here can accommodate any interfunctional covariance structure in like manner. We will present regularization priors for $\mathbf{B}^*$ in Section 2.2.2, which will be incorporated in both Gfmmc and the robust models described in Section 2.2.3.

2.2.2. Spatially Correlated Shrinkage Priors for Fixed Effects

As noted in model (1), our approach allows stimuli effects to vary across both electrodes and time, and we expect our estimates to be regularized in both of these dimensions. We will accomplish both adaptive regularization over t and spatial smoothness across s using a correlated Normal-Exponential-Gamma (CNEG) prior for the basis space fixed effects. To our knowledge, this is the first use of such a correlated scale mixture prior to simultaneously smooth spatially-varying fixed effect functions.

More specifically, Let $\mathbf{B}^*_{jk}$ denote the (j,k)th column of $\mathbf{B}^*$. We assume that $\mathbf{B}^*_{jk} = \boldsymbol{\Gamma}\mathbf{b}^*_{jk}$, where $\boldsymbol{\Gamma}$ is a lower triangular matrix obtained from the Cholesky decomposition of a prior correlation matrix $\mathbf{R}_B$, i.e., $\mathbf{R}_B = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$. We assume that entries of $\mathbf{b}^*_{jk}$ are a priori independent, and each follows a Normal-Exponential-Gamma distribution with parameters a^B_{jk} and b^B_{jk}. We call the resulting prior for $\mathbf{B}^*_{jk}$ the CNEG prior following Griffin and Brown (2012), and write $\mathbf{B}^*_{jk} \sim \text{CNEG}(\boldsymbol{\Gamma}, a^B_{jk}, b^B_{jk})$. Technical details and discussions are available in Section 3 of supplementary materials. The CNEG prior encourages smoothness (spatial correlation) in each fixed effect $\mathbf{B}^*_{jk}$ across nearby electrodes. As a sparse prior in the wavelet space, it also induces adaptive regularization over t in data domain, i.e., it tends to retain large values of $B(t)$ with minimal attenuation while shrinking very small values of $B(t)$ towards zero to encourage sparsity (Morris and Carroll, 2006).

2.2.3. Robust Functional Mixed Models with Spatial Correlation

The Gaussian assumptions underlying the Gfmmc make the method described above sensitive to outliers, while it would be desirable for our method to be insensitive to outlying subjects, time points, or electrodes that can sometimes occur in practice. We now present *robust functional mixed models for correlated functional data* (Rfmmc). Denote the (j,k)th column of the wavelet domain model (5) by $\mathbf{d}_{jk} = \mathbf{X}\mathbf{B}^*_{jk} + \mathbf{Z}\mathbf{U}^*_{jk} + \mathbf{E}^*_{jk}$, and let $\mathbf{U}^*_{jk} = \{U^*_{mjk}\}_{m=1}^M$ and $\mathbf{E}^*_{jk} = \{E^*_{ijk}\}_{i=1}^N$. We use an CNEG prior for $\mathbf{B}^*_{jk}$ as above, and specify the random effect distribution using the scale mixtures of normals following Zhu et al. (2011): $U^*_{mjk} \sim N(0, \phi_{mjk})$, $\phi_{mjk} \sim \text{Exp}((\nu^U_{jk})^2/2)$, $(\nu^U_{jk})^2 \sim \text{Gamma}(a^U, b^U)$, where $\{\phi_{mjk}\}$ are mutually independent *scaling parameters* with exponential mixing distributions, and ν^U_{jk} are mutually independent *population scale parameters*. The above formulation is equivalent to setting double exponential (DE) distributions for random effects and residuals, which has the effect of accommodating heavier-tailed behavior (non-Gaussianity) of the data and downweighting the effect of outlying curves or regions.

To incorporate inter-electrode spatial correlation, we further assume that $\mathbf{E}^*_{jk}$ follows a scale-mixture-of-

normal setup with a block-diagonal correlation structure, i.e.,

$$\begin{aligned}\mathbf{E}^*_{jk} &\sim N(0, \mathbf{\Sigma}_{jk}), \quad \mathbf{\Sigma}_{jk} = \text{diag}\,\{\lambda_{ljk}; l = 1, \ldots, \mathcal{L}\} \otimes \mathcal{R}_{jk}, \\ \lambda_{ljk} &\sim \text{Exp}((\nu^E_{jk})^2/2), \quad (\nu^E_{jk})^2 \sim \text{Gamma}(a^E, b^E),\end{aligned}$$

where $\mathcal{R}_{jk}$ is the within-block correlation matrix and $\boldsymbol{\lambda}_{jk} = \{\lambda_{1jk}, \ldots, \lambda_{\mathcal{L}jk}\}$ contains independent scaling parameters. Under this setup, we can write the joint conditional density of $\boldsymbol{\lambda}_{jk}$ and $\boldsymbol{\rho}_{jk}$, as shown in Equation (1) in supplementary materials. Based on these results, we find that the conditional distribution of each λ_{ljk} is a generalized-inverse-Gaussian (GIG) distribution (Jørgensen, 1982).

The structure of $\mathcal{R}_{jk}$ can be parameterized following the same Matérn structure as in (6). Alternative correlation structures to the Matérn can be adopted without difficulty. A separable correlation structure is induced if one specifies $\boldsymbol{\rho}$ to be constant across all (j, k). When the $\boldsymbol{\rho}$ parameters depend on (j, k), the corresponding Rfmmc model is denoted by $\text{Rfmmc}_{\boldsymbol{\rho}_{jk}}$, and when $\boldsymbol{\rho}$ is common across (j, k), the model is denoted by $\text{Rfmmc}_{\boldsymbol{\rho}}$.

2.3. Posterior Analysis

We estimate parameters of the proposed models through posterior sampling using Markov chain Monte Carlo (MCMC) algorithms. Details are provided in the supplementary materials. Each posterior sample of $\mathbf{B}^*$ and $\mathbf{U}^*$ can be transformed back into the data space using the inverse wavelet transform, yielding posterior samples for $\mathbf{B}(t)$ and $\mathbf{U}(t)$ in the data space model (2) on a dense grid T. The posterior samples can also be computed for any function of the parameters, including the contrast effects between two stimuli and the averaged effect on a specific region, for example prespecified waveform components. Based on these samples, various inferential goals can be achieved.

2.3.1. Identify Significant Spatiotemporal Regions

A key inferential objective in the ERP data analysis is to identify spatial and temporal locations corresponding to electrophysiological effects that are different across different stimuli. This can be done by first calculating the contrast effects for a pair of stimuli. For example, denote by $B^{(g)}_{CIG,s}(t), B^{(g)}_{NEU,s}(t)$ the gth sample for the fixed effects at electrode s for the cigarette stimulus (CIG) and neutral stimulus (NEU) respectively. Then the contrast effect between CIG and NEU at electrode s can be calculated by $C^{(g)}_{CIG-NEU,s}(t) = B^{(g)}_{CIG,s}(t) - B^{(g)}_{NEU,s}(t)$. We can then identify the significant regions using $\{C^{(g)}_{CIG-NEU,s}(t)\}$. Most existing methods in the literature focus on the use of pointwise credible band for such questions, flagging any position t with a credible band that does not include zero. However, as emphasized in Crainiceanu et al. (2012), pointwise credible bands do not have joint coverage probabilities, and inference based on them does not adjust for family-wise/experimental-wise error rate (FWER/EWER) in the inherent multiple testing problem and thus is likely to result in high false discovery rates. Hence, we

propose two methods for flagging regions with global coverage properties: thresholding methods based on the simultaneous band scores (SimBaS) and the Bayesian false discovery rate (BFDR).

• **Simultaneous Band Scores (SimBaS).** The SimBaS are used to test whether a location of a contrast effect $C(s,t) = C_s(t)$ is significantly nonzero while controlling the EWER across $s = 1, \ldots, S$ and $t \in T$. To calculate SimBaS, we first generate simultaneous credible bands (SCBs) following Ruppert et al. (2003), i.e., $[\widehat{C}(s,t) - m_\alpha \, \widehat{\text{sd}}\{C(s,t)\}, \; \widehat{C}(s,t) + m_\alpha \, \widehat{\text{sd}}\{C(s,t)\}]$, where $\widehat{C}(s,t)$ is the sample mean, $\widehat{\text{sd}}\{C(s,t)\}$ is the sample standard deviation, and m_α is the $(1-\alpha)$ sample quantile of $\max_{s,t} \left\{ |C^{(g)}(s,t) - \widehat{C}(s,t)| / \widehat{\text{sd}}\{C(s,t)\} \right\}$, $g = 1, \ldots, H$. We then compute SimBaS by inverting the SCB procedure. Specifically, we calculate the SCB for a range of α values, and define the SimBaS at each (s,t) as the smallest α for which the $100(1-\alpha)\%$ SCB exclude zero at (s,t). This measure was first introduced in Meyer et al. (2015). Based on SimBaS, we can compute a global Bayesian p-value (GBPV) as $\min_{(s,t)}\{\text{SimBaS}(s,t)\}$, which can be used to test the global functional null hypothesis that $C(s,t) \equiv 0$. If GBPV$< \alpha$, we can conclude that there is some difference between stimuli types, and can subsequently localize these effects by flagging locations (s,t) as strongly significant if the corresponding SimBaS(s,t) is less than α.

• **Bayesian False Discovery Rate (BFDR).** At times, we are interested in identifying locations at which the magnitude of the contrast effect $C(s,t)$ is greater than some prespecified practical effect size δ. To do this, we first calculate the point-wise posterior probability $\widehat{p}(s,t) \approx \Pr(|C_s(t)| > \delta | \text{Data})$ from the posterior samples. The values $1 - \widehat{p}(s,t)$ can be interpreted as an estimate of the *local* FDR at location (s,t), if we consider a discovery to be a location where the effect is in fact greater than δ in magnitude. We then find a threshold ϕ_α for $\widehat{p}(s,t)$, for example corresponding to a prespecified expected FDR (averaged across all s and t) of α, and flag locations with $\widehat{p}(s,t) > \phi_\alpha$ as being significantly greater than δ. This strategy was introduced in the functional regression context by Morris et al. (2008). Further details are available in the supplementary materials.

Comparing the two methods, we see that the BFDR method uses the weaker FDR criterion but requires the pre-specification of a threshold δ, whereas the SimBaS analysis corresponds to FWER/EWER considerations but does not require specification of δ.

2.3.2. Model Selection via Posterior Predictive Likelihoods

We have proposed multiple spatial functional regression models, and it is natural to wonder for a given data set which model is ideal. We introduce a model selection approach using a training-validation strategy.

For our ERP data, we first randomly split the 180 subjects into a training set (containing 140 subjects) and a validation set (containing 40 subjects). We then fit various models to the training data and calculate the *posterior predictive likelihood* of the validation data using posterior samples obtained from the training procedure. Let $\boldsymbol{\theta}$ denote all model parameters. Let $\mathbf{D}^s, \mathbf{X}^s$ denote the data from a new subject in the validation set, and let $\mathcal{M}$ denote the model under consideration, then the posterior predic-

tive likelihood for the new subject can be approximated by Monte Carlo integration $f(\mathbf{D}^s|\mathcal{M},\mathbf{D},\mathbf{X},\mathbf{Z}) = \int f(\mathbf{D}^s|\mathbf{X}^s,\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathbf{D},\mathbf{X},\mathbf{Z},\mathcal{M})d\boldsymbol{\theta} \approx 1/H\sum_{g=1}^{H} f(\mathbf{D}^s|\mathbf{X}^s,\boldsymbol{\theta}^{(g)})$, where $\{\boldsymbol{\theta}^{(g)}, g = 1,\ldots,H\}$ are posterior samples of θ. Since larger posterior predictive likelihood indicates a better model fit to the validation data, to compare multiple models, it is sufficient to directly compare the log posterior predictive likelihood (LPPL). Notice that when computing the likelihood for new subjects, one needs to integrate out the random effects. We describe details in supplementary materials.

2.3.3. An Automated Workflow for Multiple-Inferential Tasks

Figure 2 presents a workflow that can serve as an automated pipeline for rigorously modeling this rich data. We first fit multiple models for each cortical region using the training set, then calculate LPPLs for the validation set and use them to select the best model for each region. The reasons for model fitting by cortical regions will be explained in Section 3.2. We then re-fit the best model to the full ERP data at each region and combine the posterior samples of the electrode-specific fixed effects from all regions. In order to present results continuously on the surface of the scalp, we interpolate fixed effects across all electrodes on the scalp and generate inferential summaries over a dense spatiotemporal domain. If time-frequency representations are modeled, these summaries will be over a dense grid on the 3D space-time-frequency domain. In case that inference on any desired prespecified waveform components (e.g. N100, P300, etc.) are desired, inferential summaries can be computed by selecting the corresponding peak locations or integrating over regions of t, which can be represented on the spatial scalp space.

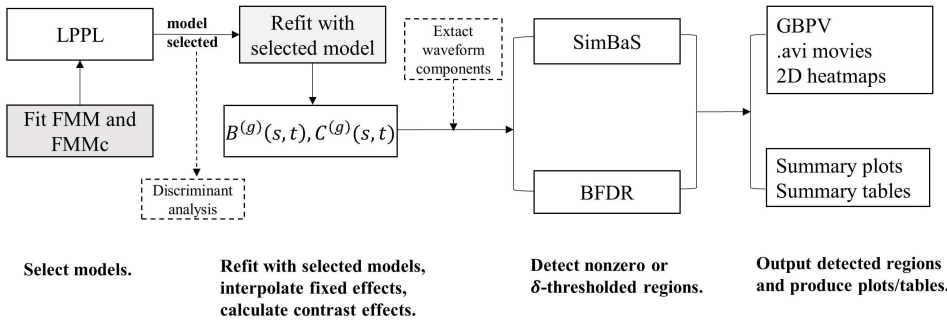

Figure 2: The suggested workflow for posterior inference in ERP data analysis. LPPL: log posterior predictive likelihood; SimBaS: Simultaneous Band Score; BFDR: Bayesian False Discovery Rate; GBPV: global Bayesian p-value.

3. Results

3.1. Simulation Study

We designed a simulation study to assess the performance of the proposed models. Data were simulated to resemble real ERP data. Our comparisons involve six models. Two are based on existing FMMs

that do not consider spatial correlations in either $\mathbf{B}(t)$ or $\mathbf{U}(t)$, including the Gaussian FMM (Gfmm) of Morris and Carroll (2006) and the Robust FMM (Rfmm) of Zhu et al. (2011). Four are spatial functional regression models proposed in this paper, including Gfmmc ($\text{Gfmmc}_{\boldsymbol{\rho}_{jk}}$, $\text{Gfmmc}_{\boldsymbol{\rho}}$) and Rfmmc ($\text{Rfmmc}_{\boldsymbol{\rho}_{jk}}$, $\text{Rfmmc}_{\boldsymbol{\rho}}$) models. For all models, we consider electrode-specific (i.e., spatially varying) fixed effects with a binary design matrix $\mathbf{X} = (\mathbf{X}_{11}, \ldots, \mathbf{X}_{1S}, \ldots, \mathbf{X}_{A1}, \ldots, \mathbf{X}_{AS})$, where S is the number of electrodes and A is the number of stimuli. For example, $(\mathbf{X}_{as})_i = 1$ indicates that the ith ERP curve belongs to the ath stimulus and the sth electrode.

To resemble the characteristics of real ERPs, we simulated data using the ERP curves from region R11 as the reference data. Specifically, we first fit the four proposed models ($\text{Gfmmc}_{\boldsymbol{\rho}_{jk}}$, $\text{Gfmmc}_{\boldsymbol{\rho}}$, $\text{Rfmmc}_{\boldsymbol{\rho}_{jk}}$ and $\text{Rfmmc}_{\boldsymbol{\rho}}$) to the ERPs from region R11. From the fitted models, we obtained the estimated values of $\mathbf{B}^*$ as well as the variance parameters for the random effect and residuals. We then treated these estimates as the true underlying parameters and simulated four data sets. The resulting data sets resemble real ERPs with different data distributions and spatial correlation structures. Each data set was generated based on one of the four models, which gives us the ground truth so that we can assess whether our model selection procedure can correctly select the true model, and evaluate the potential loss of efficiency if models are misspecified. In the supplementary materials, we plotted some simulated ERPs together with the true ERPs, which demonstrates that this simulation strategy has yielded ERPs with the functional characteristics of real ERP data. We denote the simulated data sets as $\text{G}_{\boldsymbol{\rho}_{jk}}$, $\text{G}_{\boldsymbol{\rho}}$, $\text{DE}_{\boldsymbol{\rho}_{jk}}$, $\text{DE}_{\boldsymbol{\rho}_{jk}}$, corresponding to the four proposed models. Here G indicates data with Gaussian random effect and residuals, DE indicates data with DE distributions for the random effect and residuals, and the subscripts $\boldsymbol{\rho}_{jk}$ or $\boldsymbol{\rho}$ specify the interfunctional correlation structures.

Each simulated data set contains 5760 ERP curves from $M = 80$ subjects, with each subject having 72 curves from $S = 18$ electrodes and $A = 4$ stimuli types. To reduce the computing time, we downsampled the time grid from 225 to 75 time points per curve. To assess the performance of the LPPL-based model selection procedure, another four validation sets were generated in the same way, with 20 subjects in each set. The above simulation was repeated five times, and results were evaluated using the following criteria.

Evaluation Criteria. We applied the six models to each simulated data set and calculated six summary statistics to evaluate the estimation performance. They included

$$\text{IMSE} = \frac{1}{AS}\sum_{a=1}^{A}\sum_{s=1}^{S}\frac{||\widehat{B}_{as}(t) - B_{as}(t)||^2}{||B_{as}(t)||^2},\ \text{IPVar} = \frac{1}{AS}\sum_{a=1}^{A}\sum_{s=1}^{S}\frac{\frac{1}{H}\sum_{g=1}^{H}||B_{as}^{(g)}(t) - \widehat{B}_{as}(t)||^2}{||B_{as}(t)||^2},$$

IWidth$= 1/(AS)\sum_{a=1}^{A}\sum_{s=1}^{S}||\widehat{w}_{B_{as}}(t)||^2/||B_{as}(t)||^2$, the coverage probability of the SCB for $\mathbf{B}(t)$ (CPrB_{95}), as well as the MSE $= \sum_{jk}(\widehat{\alpha}_{jk} - \alpha_{jk})^2/\sum_{jk}\alpha_{jk}^2$ and PVar $= \frac{1}{H}\sum_{g=1}^{H}\sum_{jk}(\alpha_{jk}^{(g)} - \hat{\alpha}_{jk})^2/\sum_{jk}\alpha_{jk}^2$ for α and v in the Matérn correlation. In the above formulae, the hat symbol denotes the posterior mean, $||\cdot||$ denotes

the L^2 norm, H denotes the number of posterior samples, and $\widehat{w}_{B_{as}}(t)$ denotes the width of the 95% pointwise credible band of $B_{as}(t)$. Here, IMSE and MSE summarize the deviation of the posterior mean about the truth; IPVar and PVar summarize the variability about the posterior mean.

To further assess the performance of BFDR and SimBaS in terms of flagging the regions with differential electrophysiological effects across stimuli, we defined two statistics—the thresholded false discovery rate (FDR_ϵ) and sensitivity (SEN_ξ). The FDR_ϵ is defined as the number of flagged locations with true value less than or equal to ϵ divided by the total number of flagged locations; the SEN_ξ is defined as the number of flagged locations with true value greater than ξ divided by the total number of locations with true value greater than ξ. These statistics are defined in order to evaluate the performance of the methods for flagging significant locations in the setting of absolutely continuous parameters. Besides FDR_ϵ and SEN_ξ, we defined the false negative rate (FNR_ξ) and specificity (SPEC_ϵ) in a similar fashion; details are available in the supplementary materials. Finally, we evaluated the model selection procedure by computing LPPL based on the validation data.

Simulation Results All six models were applied to each simulated data set. Intuitive visualizations of the estimated effects and the ground truth are provided as a scalp plot and a movie file in supplementary materials. The summary statistics were averaged across all five replications and listed in Table 1. Results from the "matched" model (the correct model) are highlighted using boldface. From Table 1, we see that Gfmm and Rfmm had larger IMSE and lower coverage rates than all the Gfmmc and Rfmmc models. This implies that when spatial correlation was present, ignoring such correlation results in larger estimation errors and less reliable inferential summaries. For the posterior variance, Gfmm and Rfmm had smaller IPVar and narrower IWidth than the Gfmmc and Rfmmc models, especially for data with Gaussian tails ($\text{G}_{\boldsymbol{\rho}_{jk}}$, $\text{G}_{\boldsymbol{\rho}}$). This pattern reflects the fact that treating correlated data as independent can cause overestimation of the effective sample size, which leads to underestimated posterior variances (Sainani, 2010). Comparing the four models that take into account spatial correlations, for data with DE tails ($\text{DE}_{\boldsymbol{\rho}_{jk}}$, $\text{DE}_{\boldsymbol{\rho}}$), the Rfmmc models achieved systematically lower IMSE, smaller IPVar and narrower IWidth than the Gfmmc models. For data with Gaussian tails, Rfmmc models still achieved IMSEs comparable to those of the Gfmmc models, and the results on IPVar, IWidth and CPrB_{95} are also comparable with the results from the Gfmmc models. These patterns indicate that for data with heavier (than Gaussian) tails, the robust models help reduce estimation error and improve estimation accuracy. If data have Gaussian tails, robust models do not trade off too much estimation or inferential performance relative to Gaussian models. These benefits of robust models have also been investigated by Zhu et al. (2011). The statistics for $\mathbf{U}(t)$ show similar patterns to those observed for $\mathbf{B}(t)$, and results are available in the supplementary materials.

We applied both BFDR ($\delta = 0.6$) and SimBaS on contract effects to detect spatiotemporal regions corresponding to differential electrophysiological effects across stimuli while controlling the overall FDR or

Table 1: Summary statistics of simulation study: integrated mean squared error (IMSE), integrated posterior variance (IPVar), integrated width of 95% credible interval (IWidth), and coverage probability of the 95% SCB (CPrB_{95}) of $\mathbf{B}(t)$; the averaged mean squared error (MSE) and the averaged posterior variance (PVar) of the Matérn parameters α and v; the $FDR_{.3}$ and $SEN_{1.25}$ calculated for regions flagged using BFDR ($\delta = 0.6$) and SimBaS approaches; the log posterior predictive likelihood (LPPL) of validation data sets; and the running time (based on 4000 MCMC iterations).

		$\mathbf{B}(t)$				α		v		BFDR ($\delta = .6$)		SimBaS		LPPL	Time
Data	Model	IMSE	IPVar	IWidth	CPrB_{95}	MSE	PVar	MSE	PVar	$\text{FDR}_{.3}$	$\text{SEN}_{1.25}$	$\text{FDR}_{.3}$	$\text{SEN}_{1.25}$	($\times 10^4$)	(hrs)
	Gfmm	0.465	0.045	0.675	0.751	–	–	–	–	0.021	0.871	0.019	0.779	-10.545	2.030
	Rfmm	0.539	0.043	0.646	0.706	–	–	–	–	0.054	0.805	0.066	0.677	-22.215	3.702
$\text{G}_{\boldsymbol{\rho}_{jk}}$	**Gfmmc$_{\boldsymbol{\rho}_{jk}}$**	**0.143**	**0.076**	**1.170**	**0.967**	**0.006**	**0.005**	**0.002**	**3.3e-4**	**0.012**	**0.965**	**0.035**	**0.812**	**-4.662**	**3.394**
	Gfmmc$_{\boldsymbol{\rho}}$	0.141	0.092	1.420	0.974	0.053	5.4e-5	0.050	4.0e-6	0.010	0.877	0.058	0.689	-4.713	2.123
	Rfmmc$_{\boldsymbol{\rho}_{jk}}$	0.151	0.073	1.120	0.965	0.110	0.011	0.004	3.2e-4	0.013	0.938	0.035	0.787	-5.220	7.191
	Rfmmc$_{\boldsymbol{\rho}}$	0.151	0.093	1.437	0.973	0.078	1.1e-4	0.051	3.9e-6	0.008	0.822	0.055	0.533	-5.265	5.561
	Gfmm	0.507	0.064	0.949	0.793	–	–	–	–	0.037	0.709	0.067	0.510	-10.681	2.357
	Rfmm	0.553	0.060	0.901	0.753	–	–	–	–	0.093	0.752	0.123	0.609	-22.295	3.681
	Gfmmc$_{\boldsymbol{\rho}_{jk}}$	0.140	0.093	1.424	0.982	0.005	0.005	0.002	3.2e-4	0.029	0.809	0.117	0.670	-4.733	3.445
$\text{G}_{\boldsymbol{\rho}}$	**Gfmmc$_{\boldsymbol{\rho}}$**	**0.142**	**0.092**	**1.421**	**0.981**	**1.9e-4**	**5.3e-5**	**2.6e-5**	**3.5e-6**	**0.027**	**0.808**	**0.096**	**0.665**	**-4.730**	**2.157**
	Rfmmc$_{\boldsymbol{\rho}_{jk}}$	0.139	0.093	1.431	0.983	0.140	0.012	0.003	3.1e-4	0.026	0.807	0.079	0.624	-5.244	7.225
	Rfmmc$_{\boldsymbol{\rho}}$	0.139	0.100	1.528	0.985	0.029	1.5e-4	5.4e-4	4.8e-6	0.024	0.805	0.066	0.590	-5.309	5.614
	Gfmm	0.581	0.071	1.069	0.794	–	–	–	–	0.036	0.957	0.012	0.646	-13.158	2.362
	Rfmm	0.566	0.045	0.677	0.707	–	–	–	–	0.004	1.000	0.037	0.967	-26.423	3.708
	Gfmmc$_{\boldsymbol{\rho}_{jk}}$	0.208	0.121	1.861	0.984	0.013	0.005	0.003	1.6e-4	0.020	0.981	0.002	0.785	-4.704	3.432
$\text{DE}_{\boldsymbol{\rho}_{jk}}$	Gfmmc$_{\boldsymbol{\rho}}$	0.221	0.175	2.692	0.993	0.053	3.6e-5	0.044	1.4e-6	0.002	0.794	0.025	0.539	-4.774	2.171
	Rfmmc$_{\boldsymbol{\rho}_{jk}}$	**0.112**	**0.063**	**0.969**	**0.981**	**0.053**	**0.005**	**0.001**	**1.6e-4**	**0.001**	**1.000**	**0.003**	**1.000**	**-3.451**	**7.206**
	Rfmmc$_{\boldsymbol{\rho}}$	0.123	0.092	1.406	0.989	0.074	7.7e-5	0.045	1.5e-6	0.000	1.000	0.005	0.954	-3.531	5.596
	Gfmm	0.685	0.117	1.761	0.853	–	–	–	–	0.129	0.710	0.010	0.485	-13.403	2.643
	Rfmm	0.703	0.077	1.147	0.770	–	–	–	–	0.039	0.944	0.010	0.697	-26.398	3.650
$\text{DE}_{\boldsymbol{\rho}}$	Gfmmc$_{\boldsymbol{\rho}_{jk}}$	0.265	0.183	2.822	0.982	0.009	0.005	0.002	1.6e-4	0.066	0.745	0.011	0.571	-4.850	3.391
	Gfmmc$_{\boldsymbol{\rho}}$	0.268	0.183	2.810	0.982	1.0e-4	8.7e-5	3.7e-5	2.5e-6	0.057	0.760	0.009	0.556	-4.847	2.173
	Rfmmc$_{\boldsymbol{\rho}_{jk}}$	0.107	0.088	1.348	0.985	0.057	0.006	0.002	1.5e-4	0.015	1.000	0.005	0.939	-3.508	7.237
	Rfmmc$_{\boldsymbol{\rho}}$	**0.107**	**0.092**	**1.421**	**0.987**	**0.019**	**9.7e-05**	**3.6e-4**	**2.3e-6**	**0.018**	**1.000**	**0.004**	**0.924**	**-3.584**	**5.507**

the FWER across all 18 electrodes and time points to be less than $\alpha = 0.05$. Results are assessed using the thresholded statistics $\text{FDR}_{.3}$ and $\text{SEN}_{1.25}$. These statistics are averaged across all six contrast effects and the five repeated simulations, and are listed in Table 1. From Table 1, we see that for data with heavier tails ($\text{DE}_{\boldsymbol{\rho}_{jk}}$ and $\text{DE}_{\boldsymbol{\rho}}$), the three robust models (Rfmm, $\text{Rfmmc}_{\boldsymbol{\rho}_{jk}}$, $\text{Rfmmc}_{\boldsymbol{\rho}}$) tend to show higher $\text{SEN}_{1.25}$ than their non-robust counterparts, and the two Rfmmc models always achieve higher $\text{SEN}_{1.25}$ than Rfmm. For data with Gaussian tails, the two Gfmmc models achieve higher $\text{SEN}_{1.25}$ than their Rfmmc counterparts. We also observe that the SimBaS approach gives systematically lower $\text{SEN}_{1.25}$ than the BFDR approach. This is not a surprise since FWER/EWER based approaches (e.g., SimBaS) are more conservative than FDR based approaches, hence tend to miss more discoveries. The results on $\text{FDR}_{.3}$ in Table 1 show that for data with heavier tails, the Rfmmc models tend to give lower FDRs than the other methods. For data with Gaussian tails, the Rfmmc models provide comparable, sometimes even lower, FDRs than their Gaussian counterparts. Additional statistics on FNR_{ξ} and SPEC_{ϵ} are available in the supplementary materials.

In Table 1, we also list the averaged LPPL. The results show that for all four simulated data sets, the correct models almost always achieved the maximum LPPL among the six models. An exception is the $\text{DE}_{\boldsymbol{\rho}}$ data, in which case although the data truly have separable correlation structure, the non-separable model $\text{Rfmmc}_{\boldsymbol{\rho}_{jk}}$ still gives a slightly higher LPPL. This suggests the robustness of non-separable models—we have little loss of efficiency when using the more flexible model even if data are generated from the simpler model. Therefore, it might be a reasonable strategy to use non-separable models by default. In addition to LPPL, in Table 1 we list the running time of each method, which shows that the robust methods cost roughly twice as much computational time than their Gaussian counterparts, and the models with non-separable correlation structures run slower than those with separable structures.

3.2. Application: Analysis of Smoking Cessation ERP Data

Recall that our goal in analyzing the ERP data is to characterize the differential neurological response of smokers across different visual stimuli spatially and temporally. While our proposed framework is suitable to include all electrodes, we choose to fit separate models for each of the 11 cortical regions for three reasons: (1) By using LPPL-based model selection, we observed that different models fit the data better for different regions. (2) The spatial correlation between electrodes appears to vary across scalp regions; see Figure 1(c) as well as Figure 13 in supplementary materials. Therefore, fitting separate models to each cortical region allows spatial covariance parameters, random effects, and residual distributions to vary across cortical regions, providing more flexibility. (3) Modeling brain signals by regions, as a *divide-and-conquer* approach, has also been adopted by other spatiotemporal modeling approaches such as Musgrove et al. (2016), who has shown that such strategy substantially improves computation efficiency while remaining insensitive to model misspecification and edge effects. Additionally, we have performed sensitivity analyses to demonstrate that our results are robust to different partitioning boundaries and parameter setups. In supplementary

Table 2: ERP data analysis: LPPLs on validation data. The values listed are on the scale of 10^4. The value with the highest LPPL in each region (row) is highlighted with boldface.

Region	Gfmm	Rfmm	Gfmmc$_{\boldsymbol{\rho}_{jk}}$	Gfmmc$_{\boldsymbol{\rho}}$	Rfmmc$_{\boldsymbol{\rho}_{jk}}$	Rfmmc$_{\boldsymbol{\rho}}$
Ant Frontal L (R1)	-17.62	-95.78	**-8.43**	-8.66	-10.10	-12.71
Ant Frontal R (R2)	-24.93	-81.79	-23.32	-21.32	**-11.14**	-13.53
Frontal L (R3)	-4.44	-88.68	**4.29**	4.09	3.34	2.53
Frontal R (R4)	-7.93	-87.99	-0.23	-0.38	**1.19**	0.52
Central L (R5)	-7.10	-107.32	5.37	5.22	**10.54**	8.63
Central R (R6)	-8.65	-109.98	1.89	1.68	**11.59**	9.70
Temporal L (R7)	-2.11	-59.17	**3.80**	3.72	1.13	-0.30
Temporal R (R8)	-1.13	-59.89	**4.94**	4.79	1.46	0.06
Parietal L (R9)	9.14	-101.17	**22.12**	21.99	19.03	18.28
Parietal R (R10)	14.24	-115.25	26.20	26.09	**26.33**	25.01
Occipital (R11)	9.72	-184.00	31.73	31.87	**41.04**	40.02

materials, we have also included a comparison between the region-by-region and global modeling approaches. The comparison demonstrates that for our data, similar results are obtained in either case, but the LPPL statistic suggests that the region-specific modeling fits the data better.

We first fit the six models used in the simulation to the training data, and assessed the best model separately for each of the 11 cortical regions. The run-time for training each model for each of the 11 regions is available in the supplementary materials. Results of model selection are listed in Table 2. Table 2 shows that for each region, the LPPLs based on Gfmm and Rfmm were systematically lower than those based on the Gfmmc and Rfmmc models, indicating that models taking spatial correlation into account provided better fits. Moreover, for each region, the maximum LPPL (marked in bold) was achieved either by Gfmmc$_{\boldsymbol{\rho}_{jk}}$ or by Rfmmc$_{\boldsymbol{\rho}_{jk}}$, which suggests that the non-separable correlation structure was more suitable for this data, indicating the spatial correlation varied temporally, and that for some cortical regions the robust model was preferred to the Gaussian model, suggesting the presence of some outliers.

After the best model was selected for each region, the selected model was used to fit the whole data set (with 180 subjects). The resulting posterior samples of the electrode-specific fixed effects were used for further analysis. To graphically present results continuously over the entire scalp region, not just at the electrodes, we interpolated posterior samples of the electrode-specific fixed effects pointwisely using a 2D interpolation onto a dense 67×67 geodesic grid (denoted by $\mathcal{D}$), and performed posterior inference based on the dense spatiotemporal grid $\mathcal{D} \times T$. We identified spatiotemporal regions that were significantly nonzero (or greater than δ in magnitude) for various contrast effects. For example, the contrast effect between "cigarette" and "neutral" was calculated by $C_{\text{cig-neu}}(s,t) = B_{\text{cig}}(s,t) - B_{\text{neu}}(s,t)$ pointwisely for each posterior sample. Since we have four stimuli, there are six pairs of contrast effects: cigarette vs. neutral (CIG-NEU), pleasant vs. neutral (PLE-NEU), unpleasant vs. neutral (UNP-NEU), cigarette vs. pleasant

(CIG-PLE), cigarette vs. unpleasant (CIG-UNP), and pleasant vs. unpleasant (PLE-UNP). Based on the posterior samples of the six contrast effects, we computed SimBaS and BFDR($\delta = 0.5$). We then calculated the GBPVs from the SimBaS for each contrast effect, and found that the GBPVs were less than 0.001 for all six contrast effects. This implies that for each pair of stimuli, there were at least some differences in their mean ERP effects. We flagged the spatiotemporal regions on the 3D domain $\mathcal{D} \times T$ using SimBaS (to detect nonzero regions) and BFDR (to detect regions with contrast effects greater than δ), using $\alpha = 0.05$ as the significance threshold.

Detailed results showing flagged regions over the entire (s, t) domain are displayed in .avi files, which mark flagged locations on a 2D scalp while stepping over time; see links to the files in the supplementary materials. Figures 3 and 4 summarize some of the key results in the figures based on SimBaS and BFDR respectively. The results for SimBaS are summarized and plotted in Figure 3, which contains integrated 2D-heatmaps for SimBaS values (row 1), integrated 2D-heatmaps for the mean contrast effect marked with flagged regions (SimBaS< 0.05) (row 2), as well as the scalp plots of SimBaS values calculated at two time intervals $[112, 160]$ ms (row 3) and $[232, 300]$ ms (row 4), using posterior samples averaged across time points within these intervals. The 2D-heatmaps in the first two rows demonstrate the results for all time (x-axis) and scalp locations (y-axis) while reordering the latter into blocks defined by the 11 cortical regions. The BFDR results are summarized in Figure 4, which demonstrates integrated 2D-heatmaps for the contrast effect marked with flagged regions (row 1), and scalp plots of local FDR values (i.e., $1 - \widehat{p}(s, t_i)$, where t_i is the ith time interval) calculated using posterior samples averaged across three time intervals: $[112, 160]$ ms, $[232, 300]$ ms, and $[440, 600]$ ms (respective rows 2-4).

Examining these integrated 2D-heatmaps or the corresponding .avi files, we see how the spatial distribution of the flagged regions evolves and changes over time. Six time intervals with evident patterns are highlighted in a table, and summary plots for SimBaS and BFDR results at each of these time intervals are produced. These results were presented in the supplementary materials (see Table 3 and Figures 5-10) together with a detailed description. Briefly speaking, no significant effects were detected before the image stimulus was shown ($[-100, 0]$ ms) and during the interval $[0, 100]$ ms. Between 112 ms and 160 ms, a time period known as the P1 region, we see a cigarette differential effect, whereby CIG was significantly different from NEU, PLE, and UNP in the parietal-occipital (R9-R11) region. From roughly 216 ms to 660 ms, we see various degrees of similarities between the response to the cigarette stimulus and that to the two emotional stimuli (PLE, UNP). To be more specific, from 216 ms to 232 ms, we observe similar response patterns for cigarette and pleasant stimuli; later at 232-300 ms, the response to the cigarette stimulus shows more similarity with the pleasant stimulus than the unpleasant stimulus; during the next period (300-440 ms), the cigarette stimulus evokes a pattern very similar to those evoked by both pleasant and unpleasant stimuli, in contrast with the neutral stimulus. Finally, from 660 ms-800 ms, we see significant differences between

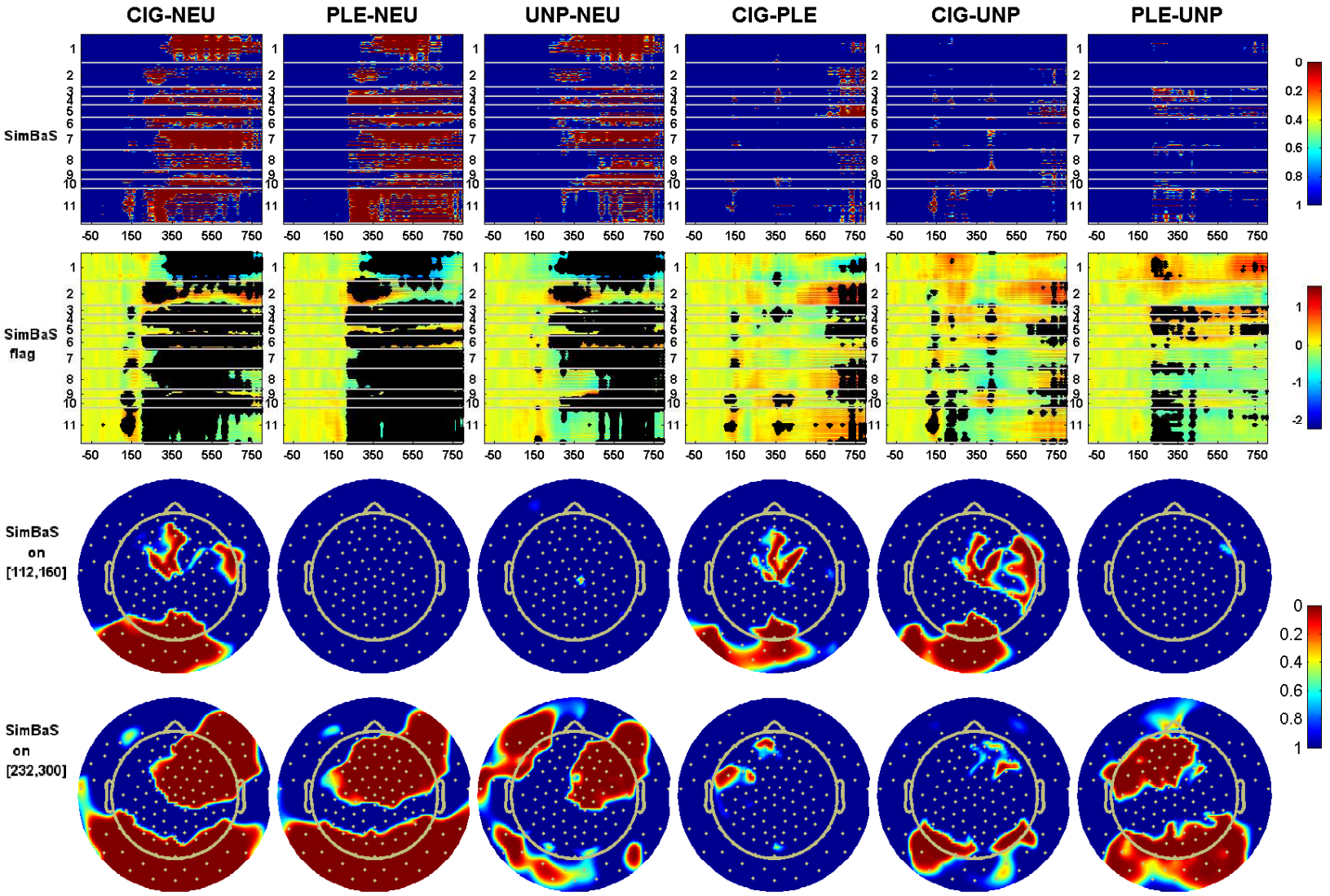

Figure 3: **Regions flagged by SimBaS.** Row 1: integrated heatmaps of the SimBaS plotted in 2D—the x-axis is time and the y-axis is vectorized spatial locations of the 2D scalp (indexed by region number). Row 2: integrated 2D heatmaps of means contrast effects (color maps) marked with SimBaS flagged regions (black dots). Row 3-4: scalp plots of new SimBaS values calculated at two time intervals ($[112, 140]$ ms and $[232, 300]$ ms) using posterior samples averaged across these intervals.

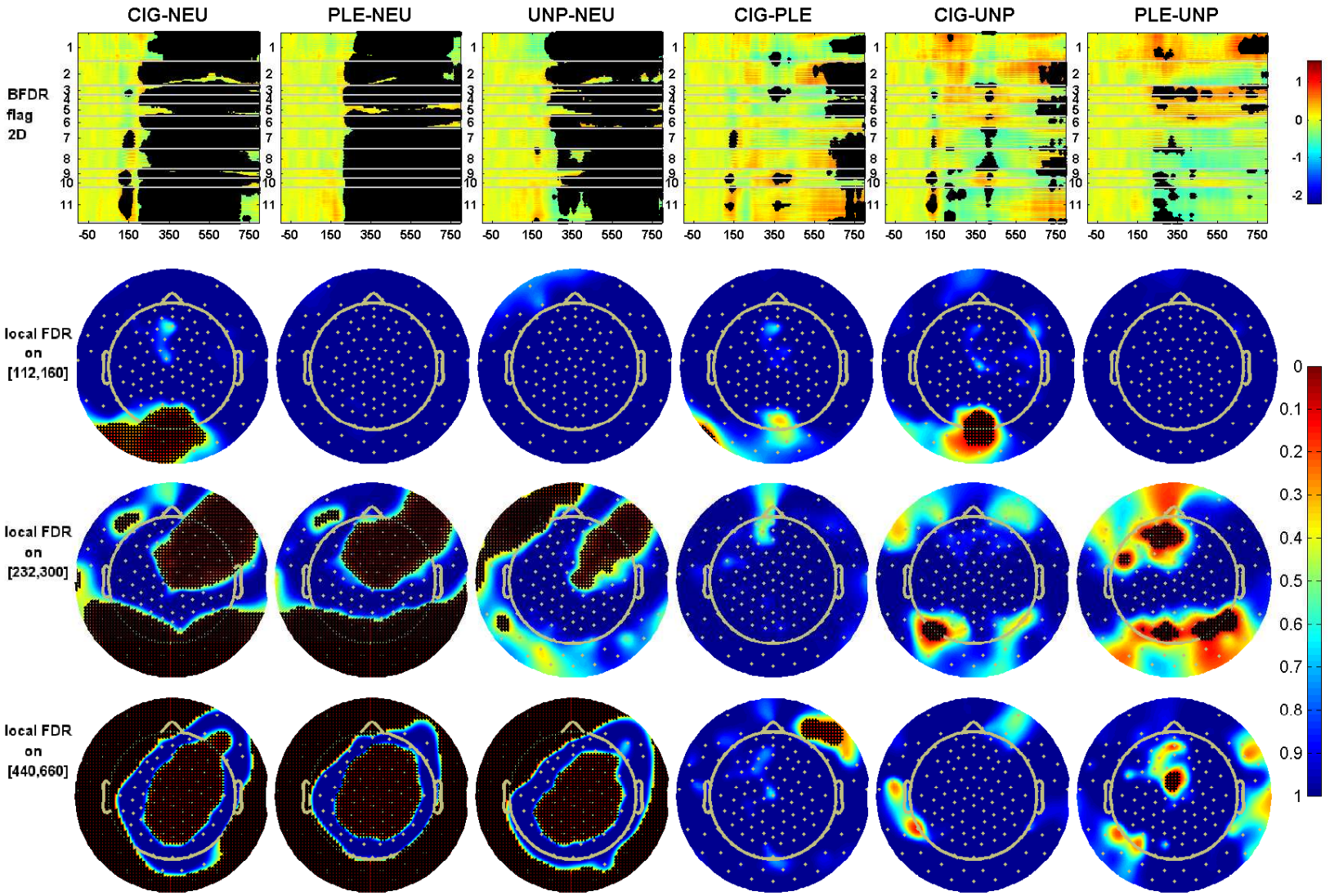

Figure 4: **Regions flagged by BFDR.** Row 1: integrated 2D heatmaps of mean contrast effects marked with BFDR (δ=0.5) flagged regions (black dots)—the x-axis is time and the y-axis is vectorized spatial locations of the 2D scalp (indexed by region number). Row 2-4: scalp plots of local FDR at three time intervals ($[112, 140]$ ms, $[232, 300]$ ms, and $[440, 660]$ ms), marked with BFDR flagged regions. Here the local FDR and the BFDR flagging results were re-calculated based on posterior samples averaged across the time intervals.

the response to all pairs of stimuli. These effects could indicate important neurological signals in smokers that are indirect measurements of their cravings. These signals can potentially be exploited in predicting smoking cessation success or providing longitudinal assessments of cessation drug efficacy.

Sensitivity Analysis The results presented above rely on several modeling choices, including model fitting by scalp regions, determination of the prior correlation parameter for $\mathbf{B}^*_{jk}$ based on preliminary estimates $\widehat{\mathbf{B}}^*_{jk}$, and selection of models using cross-validation. To assess the sensitivity of the outputs to these modeling choices, we repeated several analyses by refitting the $\text{Gfmm}_{\rho_{jk}}$ using a different cortical partition, different spatial hyperpriors, and a different cross-validation. Results are in the supplementary materials. These analyses show that our results are not sensitive to different cortical partition boundaries and different choices of spatial prior parameters for $\mathbf{B}^*_{jk}$; and different cross-validations lead to similar model selection pattern, with slight differences on choosing between $\text{Gfmmc}_{\rho_{jk}}$ and $\text{Rfmmc}_{\rho_{jk}}$ in four regions.

4. Discussion

To compare the effects of different stimuli on the ERP curves in smokers, we have proposed functional response regression models for correlated functional data. These methods flexibly capture the complex data structure yet yield intuitive and natural inferential summaries. Our application to the ERP data demonstrates patterns of differential electrophysiological effects across stimuli, and characterizes similarities and differences in the effects evoked by cigarette and emotional stimuli in contrast to the neutral stimuli. Our approach provides full Bayesian inference over the entire ERP to localize the key stimuli effects on the scalp and over time, which enables us to detect effects that may have been missed had analyses been limited to prespecified waveform components, and by incorporating spatial inter-electrode correlation and robustness to outliers, may have resulted in greater power to detect stimuli effects according to the results of our simulation study.

We have analyzed an ERP data set in a smoking session study. The same data set has been analyzed by Versace et al. (2011) by using a standard ERP analysis approach. In their analysis, they first applied a temporal principal component analysis (PCA) to the ERPs, from where they identified six temporal regions of interest by using the peak locations of the loading factors of PCA. The mean voltages were then calculated by averaging across time windows centered at these temporal locations. Based on the mean voltages, a randomization test was performed to identify significant differences between the emotional/cigarette stimuli and the neutral stimulus. Versace et al. (2011)'s analysis demonstrated similar neurological responses in the presence of cigarette and emotional cues for two of the temporal regions, the 452–508 ms and the 212–316 ms time windows. It also showed that the cigarette-related pictures enhanced the amplitude of the P1 component (136-144 ms) above the levels measured in the emotional and neutral conditions. These findings are consistent with our findings described in Section 3.2 and the supplementary materials. Our analysis,

however, provides more detailed findings in terms of when and where the significant differences present between any pair of stimuli, as demonstrated by the .avi files, Figures 5–10, and Table 3 in supplementary materials. This is the key advantage of modeling the entire ERP data set without using reductionistic feature extraction.

While we have focused on modeling the stimulus effects for a group of individuals using averaged EEGs (ERPs), the proposed framework can also be used to model EEG data from multiple trials on a single individual. It can be further used to model EEGs at both the individual and group level simultaneously. This can be done in two different ways. (i) The first way is to model data from both levels all together, adding subject- and trial-specific random effect functions. Our modeling framework allows multiple levels of random effects, enabling great flexibility for capturing different sources of variability. While in principal this could be done with our existing software, for large studies like this one the sample sizes would be enormous, which would add considerably to the computational complexity. (ii) An alternative strategy would be to use a two-step approach, first modeling each individual's data independently with first-level MCMC to estimate the ERPs per subject, and then taking these as the data in a second-stage group-level MCMC to estimate the stimuli effects. This approach allows us to propagate the uncertainty of the first level model to the second, and the computation is easily parallelizable. This approach has been used in a different context by Morris et al. (2006) to deal with missing functional data.

We used the Matérn family to model the interfunctional spatial correlations. Depending on the nature of the correlation, other parametric families such as the continuous-time AR(1) structure can be easily incorporated (Louis, 1988; Simpson et al., 2014). For functional data indexed by points on a lattice, one could also assume local correlation patterns. For example, Zhang et al. (2015) used conditional autoregressive (CAR) assumptions to model local correlations between functions on a lattice, which can also be easily incorporated into our framework.

While we have focused on wavelets, our dual space models can be used with many other bases including splines and principal components. The choice of basis should be based on the characteristics of the functional data (Morris, 2015). Our analyses here modeled the temporal ERP waveforms, but our framework and software can also model the time-frequency representations of the ERPs, with the only required change being the specification of appropriate basis functions for that 2D space. Besides modeling electrode data measured on the scalp surface, our modeling framework can also be used to model reconstructed brain source signals that could be inferred from the EEG data, e.g., using the *surface Laplacian* technique (Hjorth, 1975; Kayser and Tenke, 2015; Carvalhaes and de Barros, 2015). Linking our approach to the source signal identification in a joint framework would be a very interesting problem, but beyond the intended scope of this paper.

One potential limitation of our proposed approach is the computation time for Bayesian inference. In

supplementary materials, we listed the computation time for running each of the six models for the 11 scalp regions, and also performed a run-time analysis to evaluate how the proposed framework scales with various data setups. While our algorithms can be run concurrently for all six models for each scalp region, it still takes $O(10)$ hours to train the models and calculate the LPPLs. While relatively long compared to simpler analytical approaches, this computing time is not inordinate, given the extensive time to conduct studies yielding these rich data. It is our view that this extra computing time is a good trade-off given the ability of our model to capture information anywhere in space-time and to account for the complex spatiotemporal correlation structures. One can further reduce the computation cost in two ways: by using near-lossless basis via wavelet compression (Morris et al., 2011), or by replacing the MCMC-based posterior sampling by approximation approaches such as variational Bayesian inference (Blei and Jordan, 2006). Based on our experience, we expect that the use of a near-lossless basis retaining $> 99.5\%$ total energy for each ERP would result in a speed-up of 5-20 fold with very little loss of information, and the use of variational inference usually reduce the computation time to the scale of minutes (with a sacrifice of narrower confidence bands).

While our models have numerous complex features that capture various types of spatiotemporal correlation while inducing robustness to outliers, the model specification and running of software is relatively straightforward, so accessible to a broad class of researchers. Algorithms are developed in Matlab and C, and compiled using Matlab compiler (MATLAB Compiler). The complied code and demo scripts are shared through the link: `http://www.apps.stat.vt.edu/zhu/other/FMMC_v0_compiled_May7_2018.zip`. We are also working on integrating these algorithms with an R package (R Core Team, 2017), which will generalize a preliminary R package developed by Rausch et al. (2013).

Supplementary Materials

The supplementary materials are enclosed with this submission.

Acknowledgments

Hongxiao Zhu was supported by Institute for Critical Technology and Applied Science, Virginia Tech (ICTAS-JFC 175139) and National Science Foundation (NSF-DMS 1611901). Jeffrey S. Morris was supported by National Science Foundation (NSF-DBI 1550088), National Cancer Institute (R01-CA178744, P30-CA016672), and National Institute of Drug Abuse (R01-DA017073).

References

Baladandayuthapani, V., Mallick, B.K., Hong, M.Y., Lupton, J.R., Turner, N.D., Carroll, R.J., 2008. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. Biometrics 64, 64–73.

Blei, D.M., Jordan, M.I., 2006. Variational inference for dirichlet process mixtures. Bayesian Anal. 1, 121–143. URL: https://doi.org/10.1214/06-BA104, doi:10.1214/06-BA104.

Brandeis, D., Lehmann, D., 1986. Event-related potentials of the brain and cognitive processes: approaches and applications. Neuropsychologia 24, 151–168.

Bressler, S.L., 2002. Event-related potentials, in: Arbib, M. (Ed.), The Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge MA, pp. 412–415.

Brockhaus, S., Scheipl, F., Hothorn, T., Greven, S., 2015. The functional linear array model. Statistical Modelling 15, 279–300.

Cagy, M., Infantosi, A.F.C., Franca, A.J., Lemle, M., 2006. Statistical analysis of event-related potential elicited by verb-complement merge in brazilian portuguese. Braz. J. Med. Biol. Res. 39, 1465–1474.

Carvalhaes, C., de Barros, J.A., 2015. The surface laplacian technique in eeg: Theory and methods. International Journal of Psychophysiology 97, 174 – 188. doi:10.1016/j.ijpsycho.2015.04.023. on the benefits of using surface Laplacian (current source density) methodology in electrophysiology.

Chen, K., Delicado, P., Müller, H.G., 2017. Modelling function-valued stochastic processes, with applications to fertility dynamics. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 177–196. doi:10.1111/rssb.12160.

Chen, K., Lynch, B., 2017. Weak Separablility for Two-way Functional Data: Concept and Test. ArXiv e-prints arXiv:1703.10210.

Chen, K., Müller, H.G., 2012. Modeling repeated functional observations. J. Am. Stat. Assoc. 107, 1599–1609.

Chen, L.H., Jiang, C.R., 2017. Multi-dimensional functional principal component analysis. Statistics and Computing 27, 1181–1192.

Cinciripini, P.M., Robinson, J.D., Karam-Hage, M., Minnix, J.A., Lam, C., Versace, F., Brown, V.L., Engelmann, J.M., Wetter, D.W., 2013. Effects of vareniclineand bupropion sustained-release use plus intensive smoking cessation counseling on prolonged abstinence from smoking and on depression, negative affect, and other symptoms of nicotine withdrawal. JAMA Psychiatry 70, 522–533.

Crainiceanu, C.M., Staicu, A.M., Ray, S., Punjabi, N., 2012. Bootstrap-based inference on the difference in the means of two correlated functional processes. Stat. Med. 31, 3223–3240.

Davidson, D., 2009. Functional Mixed-Effect models for electrophysiological responses. Neurophysiology 41, 71–79.

Gonzalez-Rosa, J.J., Vazquez-Marrufo, M., Vaquero, E., Duque, P., Borges, M., Gomez-Gonzalez, C.M., Izquierdo, G., 2011. Cluster analysis of behavioural and event-related potentials during a contingent negative variation paradigm in remitting-relapsing and benign forms of multiple sclerosis. BMC Neurology 11, 64.

Greven, S., Crainiceanu, C., Caffo, B., Reich, D., 2010. Longitudinal functional principal component analysis. Electron. J. Stat. 4, 1022–1054.

Griffin, J.E., Brown, P.J., 2012. Structuring shrinkage: some correlated priors for regression. Biometrika 99, 481–487.

Guo, W., 2002. Functional mixed effects models. Biometrics 58, 121–128.

Hasenstab, K., Scheffler, A., Telesca, D., Sugar, C.A., Jeste, S., DiStefano, C., Şentürk, D., 2017. A multi-dimensional functional principal components analysis of eeg data. Biometrics 73, 999–1009. doi:10.1111/biom.12635.

Hjorth, B., 1975. An on-line transformation of eeg scalp potentials into orthogonal source derivations. Electroencephalography and Clinical Neurophysiology 39, 526–530.

Holan, S., Wikle, C., Sullivan-Beckers, L., Cocroft, R., 2010. Modeling complex phenotypes: Generalized linear models using spectrogram predictors of animal communication signals. Biometrics 66, 914–24.

Itier, R.J., Taylor, M.J., Lobaugh, N.J., 2004. Spatiotemporal analysis of event-related potentials to upright, inverted, and contrast-reversed faces: Effects on encoding and recognition. Psychophysiology 41, 643–653.

Jørgensen, B., 1982. Statistical Properties of the Generalized Inverse Gaussian Distribution. Lecture Notes in Statistics, Springer-Verlag New York, New York, U.S.A.

Kappenman, E.S., Luck, S.J., 2016. Best practices for event-related potential research in clinical populations. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 1, 110–115. URL: http://doi.org/10.1016/j.bpsc.2015.11.007.

Kayser, J., Tenke, C.E., 2015. On the benefits of using surface laplacian (current source density) methodology in electrophysiology. International Journal of Psychophysiology 97, 171–173. doi:https://doi.org/10.1016/j.ijpsycho.2015.06.001. on the benefits of using surface Laplacian (current source density) methodology in electrophysiology.

Keil, A., Bradley, M.M., Hauk, O., Rockstroh, B., Elbert, T., Lang, P.J., 2002. Large-scale neural correlates of affective picture processing. Psychophysiology 39, 641–649.

673 Kiebel, S.J., Friston, K.J., 2004a. Statistical parametric mapping for event-related potentials: I. generic
674 considerations. NeuroImage 22, 492 – 502. doi:doi.org/10.1016/j.neuroimage.2004.02.012.

675 Kiebel, S.J., Friston, K.J., 2004b. Statistical parametric mapping for event-related potentials (ii): a hierar-
676 chical temporal model. NeuroImage 22, 503 – 520. doi:10.1016/j.neuroimage.2004.02.013.

677 Lamy, D., Salti, M., Bar-Haim, Y., 2008. Neural correlates of subjective awareness and unconscious process-
678 ing: An erp study. J. Cognitive Neurosci. 21, 1435–1446.

679 Lehmann, D., Pascual-Marqui, R.D., Michel, C., 2009. Eeg microstates. Scholarpedia 4, 7632.

680 Lole, L., Gonsalvez, C.J., Barry, R.J., De Blasio, F.M., 2013. Can event-related potentials serve as neural
681 markers for wins, losses, and near-wins in a gambling task? a principal components analysis. International
682 Journal of Psychophysiology 89, 390–398.

683 Louis, T.A., 1988. General methods for analysing repeated measures. Statistics in Medicine 7, 29–45.

684 Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of eeg- and meg-data. Journal of Neuro-
685 science Methods 164, 177 – 190. doi:https://doi.org/10.1016/j.jneumeth.2007.03.024.

686 Martinez, J.G., Bohn, K.M., Carroll, R.J., Morris, J.S., 2013. A study of mexican free-tailed bat chirp
687 syllables: Bayesian functional mixed models for nonstationary acoustic time series. Journal of the American
688 Statistical Association 108, 514–526. doi:10.1080/01621459.2013.793118.

689 MATLAB Compiler, 2012b. Matlab. The MathWorks, Natick, MA, USA.

690 Meyer, M.J., Coull, B.A., Versace, F., Cinciripini, P., Morris, J.S., 2015. Bayesian function-on-function
691 regression for multilevel functional data. Biometrics 71, 563–574. doi:10.1111/biom.1299.

692 Milz, P., Faber, P., Lehmann, D., Koenig, T., Kochi, K., Pascual-Marqui, R., 2016. The functional
693 significance of eeg microstatesassociations with modalities of thinking. NeuroImage 125, 643 – 656.
694 doi:https://doi.org/10.1016/j.neuroimage.2015.08.023.

695 Morris, J.S., 2015. Functional Regression. Annu. Rev. Stat. Appl. 2, 321–359.

696 Morris, J.S., Arroyo, C., Coull, B.A., Louise, M.R., Herrick, R., Gortmaker, S., 2006. Using wavelet-based
697 functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study.
698 J. Am. Statist. Ass. 101, 1352–1364.

699 Morris, J.S., Baladandayuthapani, V., Herrick, R.C., Sanna, P., Gutstein, H., 2011. Automated analysis of
700 quantitative image data using isomorphic functional mixed models, with application to proteomics data.
701 Ann. Appl. Stat. 5, 894–923.

Morris, J.S., Brown, P.J., Herrick, R.C., Baggerly, K.A., Coombes, K.R., 2008. Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. Biometrics 64, 479–489.

Morris, J.S., Carroll, R.J., 2006. Wavelet-based functional mixed models. J. Royal Statist. Soc. Ser. B 68, 179–199.

Musgrove, D.R., Hughes, J., Eberly, L.E., 2016. Fast, fully bayesian spatialtemporal inference. Biostatistics 17, 291–303.

Ombao, H., Raz, J., von Sachs, R., Guo, W., 2002. The slex model of a non-stationary random process. Annals of the Institute of Statistical Mathematics 54, 171–200. URL: https://doi.org/10.1023/A:1016130108440, doi:10.1023/A:1016130108440.

Park, S.Y., Staicu, A.M., 2015. Longitudinal functional data analysis. Stat 4, 212–226. doi:10.1002/sta4.89. sta4.89.

Pernet, C.R., Chauveau, N., Gaspar, C., Rousselet, G.A., 2011. Limo eeg: A toolbox for hierarchical linear modeling of electroencephalographic data. Computational Intelligence and Neuroscience 2011, 1–1. doi:10.1155/2011/831409.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ramsay, J.O., Silverman, B.W., 1997. Functional Data Analysis. Springer-Verlag, New York.

Rausch, P., Morris, J.S., Sommer, W., Krifka, M., 2013. When you are thrown a curve: Two r packages for swerving with wavelet-based functional mixed models. Linguistic Evidence Conference.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, UK.

Sainani, K., 2010. The importance of accounting for correlated observations. PM&R 2, 858–861.

Scheipl, F., Gertheiss, J., Greven, S., 2016. Generalized functional additive mixed models. Electron. J. Statist. 10, 1455–1492. doi:10.1214/16-EJS1145.

Scheipl, F., Staicu, A.M., Greven, S., 2015. Functional additive mixed models. J. Comp. Graph. Stat. 24, 477–501.

Simpson, S.L., Edwards, L.J., Styner, M.A., Muller, K.E., 2014. Kronecker product linear exponent ar(1) correlation structures for multivariate repeated measures. PLOS ONE 9, 1–10. doi:10.1371/journal.pone.0088864.

731 Staicu, A., Crainiceanu, C.M., Carroll, R.J., 2010. Fast methods for spatially correlated multilevel functional
732 data. Biostatistics 11, 177–194.

733 Steen, J., 2010. An analysis of ERP data by wavelet-based functional mixed effect modeling. Master's thesis.
734 Ghent University. Belgium.

735 Stein, M.L., 1999. Interpolation of spatial data. Springer Series in Statistics, Springer-Verlag, New York.
736 Some theory for Kriging.

737 Venturini, R., Lytton, W.W., Sejnowski, T.J., 1992. Neural network analysis of event related potentials and
738 electroencephalogram predicts vigilance, in: Moody, J., Hanson, S., Lippmann, R. (Eds.), Advances in
739 Neural Information Processing Systems 4. Morgan Kaufmann, San Mateo, California, pp. 651–658.

740 Versace, F., Minnix, J.A., Robinson, J.D., Lam, C.Y., Brown, V.L., Cinciripini, P.M., 2011. Brain reactivity
741 to emotional, neutral and cigarette-related stimuli in smokers. Addiction Biology 16, 296–307.

742 Vossen, H., Breukelen, G.V., Hermens, H., Van Os, J., Lousberg, R., 2011. More potential in statistical
743 analyses of event-related potentials: a mixed regression approach. Int. J. Methods Psychiatr. Res. 20,
744 e56–e68.

745 Wang, X., Yang, Q., Fan, Z., Sun, C.K., Yue, G.H., 2009. Assessing time-dependent association between
746 scalp eeg and muscle activation: A functional random-effects model approach. Journal of Neuroscience
747 Methods 177, 232–240. doi:10.1016/j.jneumeth.2008.09.030.

748 Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K., Majewski, T., Czerniak, B.A., Morris, J.S.,
749 2015. Functional CAR models for large spatially correlated functional datasets. J. Am. Statist. Ass. 111,
750 772–786.

751 Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., Cichocki, A., 2014. Aggregation of sparse linear discrimi-
752 nant analyses for event-related potential classification in brain-computer interface. Int. J. Neur. Syst. 24,
753 1450003.

754 Zhou, L., Huang, J.Z., Martinez, J.G., Maity, A., Baladandayuthapani, V., Carroll, R.J., 2010. Reduced
755 rank mixed effects models for spatially correlated hierarchical functional data. J. Am. Statist. Ass. 105,
756 390–400.

757 Zhu, H., Brown, P.J., Morris, J.S., 2011. Robust, adaptive functional regression in functional mixed model
758 framework. J. Am. Statist. Ass. 495, 1167–1179.