Bayesian Function-on-Function Regression for Multilevel Functional Data

Mark J. Meyer[1,*], **Brent A. Coull**[2], **Francesco Versace**[3], **Paul Cinciripini**[3], and **Jeffrey S. Morris**[3]

[1]Department of Mathematics, Bucknell University, Lewisburg, Pennsylvania, U.S.A

[2]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, U.S.A

[3]The University of Texas M.D. Anderson Cancer Center, Houston, Texas, U.S.A

Summary

Medical and public health research increasingly involves the collection of complex and high dimensional data. In particular, functional data—where the unit of observation is a curve or set of curves that are finely sampled over a grid—is frequently obtained. Moreover, researchers often sample multiple curves per person resulting in repeated functional measures. A common question is how to analyze the relationship between two functional variables. We propose a general function-on-function regression model for repeatedly sampled functional data on a fine grid, presenting a simple model as well as a more extensive mixed model framework, and introducing various functional Bayesian inferential procedures that account for multiple testing. We examine these models via simulation and a data analysis with data from a study that used event-related potentials to examine how the brain processes various types of images.

Keywords

Basis Functions; Bayesian inference; Function-on-function regression; Functional data analysis; Functional mixed models; Functional Testing; Principal Components; Wavelet regression

1. Introduction

Medical and public health research increasingly involves the collection of complex and high dimensional data. In particular, functional data—where the unit of observation is a curve or set of curves that are finely sampled over a grid—is frequently obtained (Ramsay and Silverman, 2005). Moreover, researchers often sample multiple curves per subject which yields repeated functional measures. A common question is how to analyze the relationship between two functional variables. While the field of functional data analysis (FDA) has progressed considerably in recent years, gaps remain in the literature with regards to function-on-function regression where both the predictor and outcome are functional.

---

[*]mark.meyer@bucknell.edu.

Regression in FDA can be classified into three broad sub-classes: scalar-on-function, function-on-scalar, and function-on-function. Morris (2015) contains a thorough review on existing work on functional regression. Scalar-on-function regression, on which a large literature exists, involves a scalar outcome and fucntional predictor, with functional regression coefficients. See for instance Ramsay and Dalzell (1991), Cardot, Ferraty, and Sarda (1999), Reiss and Ogden (2007), Malloy et al. (2010), Goldsmith et al. (2011), McLean et al. (2012), Gertheiss, Maity, and Staicu (2013), and references therein. Function-on-scalar regression, also heavily investigated in the literature, involves regressing a functional predictor on to a set of scalar covariates, each of which has a functional regression coefficient. See for instance Brumback and Rice (1998), Morris and Carroll (2006), Reiss, Huang, and Mennes (2010), Staicu et al. (2011), Chen and Müller (2012), Goldsmith, Greven, and Crainiceanu (2013), and references therein.

In contrast, the literature addressing function-on-function regression, with functional outcome, functional predictor, and a coefficient surface, is rather sparse. Ramsay and Silverman (2005) do devote chapters to concurrent function-on-function regression and fully functional linear models for functional responses which assume iid residual errors, have no random effect functions, and no modeling of within-function correlation in the errors. Additionally, they focus on point estimation, not inference, and the authors further emphasize that Bayesian methods that model the variability in basis function selection would be welcome. Beyond that, much of the literature is dedicated to the historical functional linear model (HFLM), as described by Malfait and Ramsay (2003) and further examined by Harezlak et al. (2007) and Kim, entürk, and Li (2011). The primary assumption in an HFLM is that the association between curves is uni-directional. Function-on-function regression allowing for bi-directional associations—that is, with unconstrained regression coefficient surfaces—is explored by Yao, Müller, and Wang (2005), Müller and Yao (2008), and Wang (2014). Also, there are some recent technical reports on the topic from one research group, Ivanescu et al. (2012), Scheipl and Greven (2012), Scheipl, Staicu, and Greven (2014), that discuss a penalized spline approach, identifiability issues, and function-on-function regression in Functional Additive Mixed Models, respectively. Inferential procedures, if addressed, rely on 95% point-wise confidence intervals (PWCI) to determine significance without adjusting for multiple comparisons. These use penalized splines, which may not be the best basis choice in various settings, including the spiky functions in this paper. Much existing work also assumes iid residual curve-to-curve deviations, which is not realistic for most functional data settings.

To illustrate the function-on-function regression problem, we examine data from a smoking cessation trial conducted at the University of Texas M. D. Anderson Cancer Center (Cinciripini et al., 2013). Event Related Potentials (ERPs) were obtained at baseline during the presentation of a series of images depicting neutral, positive, negative, and cigarette-related contents. ERPs were collected using a 129 channel Geodesic Sensor Net. Finely sampled curves were produced over the course of 900 ms (100 ms prior to picture presentation and 800 ms after). Electrical potentials every 4ms were collected from 129 electrodes distributed on the surface of the scalp resulting in 225 measurements for each electrode. While many analyses are of interest for these data, in this paper we focus on

characterizing the time-varying relationship between ERP outputs from specific electrode pairs.

In this paper, we propose a general Bayesian function-on-function regression modeling framework that can accommodate this type of multilevel functional data which we fit using a Monte-Carlo Markov Chain (MCMC) procedure. The model is flexible enough to incorporate a variety of basis expansions including such common approaches as principal components, splines, and wavelet-based functional representations. Our approach not only allows for correlation between functions through random effect functions, but also allows heteroscedasticity and within-function correlation in the residual error functions, unlike existing methods assuming iid errors. While the approach can be applied generally for any number of functional predictors and arbitrary interactions with other discrete and continuous predictors, we present specific model formulations for both a single functional predictor of interest as well an interaction of a discrete factor with a functional predictor resulting in separate function-on-function regressions for each factor.

We propose three approaches for inference that account for multiplicity. First we extend the Bayesian False Discovery Rate (BFDR) procedure for functional regression implemented by Morris et al. (2008) to the function-on-function setting. Second, from joint credible bands as in Ruppert, Wand, and Carroll (2003), we generate two novel summaries: (1) Simultaneous Band Scores (SimBaS), a functional measure that summarizes for each position in the regression surface the smallest α for which the $100(1 - \alpha)\%$ joint credible bands exclude zero, and (2) Global Bayesian P-Values (GBPV), which can be interpreted as a type of Bayesian p-value corresponding to a global functional null hypothesis of no relationship. These summaries are of general interest and can be used in other functional regression settings. To our knowledge global inference procedures within function-on-function regression have not previously been considered in the literature.

Section 2 develops a simple version of our proposed function-on-function mixed model, presents a more general model, and describes our basis function modeling strategy. Section 3 details the BFDR, SimBaS, and GBPV inference procedures. In Section 4 we present the results of a simulation assessing model fit and the BFDR, SimBaS, and GBPV procedures. Section 5 presents the results obtained by applying the proposed methods to the ERP data, and Section 6 contains further discussion.

2. Function-on-Function Regression Model for Multi-Level Functional Data

Here we introduce the function-on-function model we will use to regress one function $y(t)$, $t \in \mathcal{T}$ on another $x(\upsilon)$, $\upsilon \in \nu$. First we consider a simple case with a single functional predictor and repeated measures of $\{y(t), x(\upsilon)\}$ pairs for each subject, and then in Section 2.3 we describe more complex models that can be handled by our approach.

Individual subjects are denoted as $i = 1, \ldots, n$. Let $c = 1, \ldots, C_i$ index repeated pairs of curves observed on subject i. Then for subject i, curve set c, we observe $x_{ic}(\upsilon)$ and $y_{ic}(t)$, $\{y_{ic}(t), x_{ic}(\upsilon) : t \in \mathcal{T}, \upsilon \in \nu\}$,

$$y_{ic}(t) = \alpha(t) + \int_{\upsilon \in \nu} x_{ic}(\upsilon)\beta(\upsilon,t)d\upsilon + U_i(t) + E_{ic}(t). \quad (1)$$

We assume observation-specific and subject-specific Gaussian process errors $E_{ic}(t) \sim \mathscr{GP}(0, \Sigma_E)$ and $U_i(t) \sim \mathscr{GP}(0, \Sigma_U)$. The integration over the entire support of υ allows an unconstrained exposure-response relationship, i.e. we do not assume the timing of an effect of x on y occurs in one direction or the other. That relationship is characterized by the surface $\beta(\upsilon, t)$.

In this paper, our focus is on functional data sampled on a common fine grid. Here, we consider a discretized version of Model (1). Let $y_{ic}(\cdot)$ be finely sampled on a grid $\mathbf{t} = [t_1 \cdots t_T]$ of length T. Similarly, $x_{ic}(\cdot)$ is observed on a grid $\mathbf{v} = [\upsilon_1 \cdots \upsilon_V]$ of length V. We can then define the row vectors $\mathbf{y}_{ic} = [y_{ic}(t_1) \cdots y_{ic}(t_T)]$ and $\mathbf{x}_{ic} = [x_{ic}(\upsilon_1) \cdots x_{ic}(\upsilon_V)]$ and express Model 1 in the discrete form

$$\mathbf{y}_{ic} = \mathbf{x}_{ic}\boldsymbol{\beta}\Delta_{\upsilon_j} + \mathbf{u}_i + \mathbf{e}_{ic}, \quad (2)$$

where $\mathbf{y}_{ic}$, $\mathbf{u}_i$, and $\mathbf{e}_{ic}$ are $1 \times T$, $\mathbf{x}_{ic}$ is $1 \times V$, and $\boldsymbol{\beta}$ is the $V \times T$ matrix of coefficients and $\Delta_{\upsilon_j} = \upsilon_j - \upsilon_{j-1}$. If we sample on an equally spaced grid, then Δ_{υ_j} is a constant equal to the distance between measurement occurrences. Also note that $\mathbf{e}_{ic} \sim \mathscr{N}(0, \Sigma_E)$ and $\mathbf{u}_i \sim \mathscr{N}(0, \Sigma_U)$. In practice, we center and scale both $y_{ic}(t)$ and $x_{ic}(\upsilon)$ and thus, without loss of generality, $\alpha(t)$ in Model (1) is zero and not needed in Model (2). If one of the two functions is not centered, than $\alpha(t)$ can be incorporated into $\boldsymbol{\beta}$ with a corresponding column of ones added to the design matrix.

Now let N be the total number of observed response curves. Stacking the row vectors by subject, $\mathbf{Y}$ and $\mathbf{X}$ represent the $N \times T$ and $N \times V$ matrices of observed curves. Further, $\mathbf{Z}$ is the $N \times n$ random effects design matrix. Our discretized model for all subjects is then

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \mathbf{E}, \quad (3)$$

where $\boldsymbol{\beta}$ is as defined in Model (2), $\mathbf{U}$ is the $n \times T$ matrix of subject specific random effect functions on the grid, and $\mathbf{E}$ is the $N \times T$ matrix of model errors, interpretable as residual curve-to-curve deviations. Assuming an equally spaced grid, we omit Δ_{υ_j} from Model (3).

As is typical in functional regression (Morris, 2015), we will represent $y(t)$ and $x(\upsilon)$ using basis representations prior to model fitting. The basis transform modeling approach we use allows us fast calculations while inducing regularization in the coefficient surface $\beta(\upsilon, t)$ through basis-space prior distributions, and can be used with various choices of lossless or near-lossless basis choices, including wavelets, splines, PCs, and Fourier series. We will begin by describing the modeling approach in terms of general basis functions, and then we will present the rest of the modeling details using specific basis functions chosen for our simulation and data analysis, with the Web Appendix A mentioning some adaptations for other bases.

2.1 General Basis Transform Modeling Approach

Here we describe our general basis function transform approach for fitting the function-on-function regression models, which involves projecting both the functional responses and predictors into a chosen basis space, fitting the model in the basis space, and then transforming the results back to the original function space for interpretation and inference.

Let $y_{ic}(t)=\sum_{j=1}^{T^*} y_{icj}^* \xi_j(t)+r_{ic}^{\xi}(t)$ and $x_{ic}(v)=\sum_{j=1}^{V^*} x_{icj}^* \phi_j(v)+r_{ic}^{\phi}(v)$ be some chosen truncated basis expansions for the functional responses and predictors, respectively, and suppose this expansion results in a lossless (

$$r_{ic}^{\xi}(t)=0$$

and

$$r_{ic}^{\phi}(v)=0$$

$\forall$ i, c, and observed t), or virtually lossless (

$$r_{ic}^{\xi}(t)<\varepsilon$$

and

$$r_{ic}^{\phi}(v)<\varepsilon$$

for some small ε) transform so the $r(\cdot)$ can effectively be ignored. Potential choices with these properties include wavelets, Fourier bases, splines (given sufficient knots), PCs, or independent components. Let Ξ be a matrix of size $T^* \times T$ containing the basis functions on discrete grid $\mathbf{t}$ with element (j, t) given by $\xi_j(t)$, and likewise let Φ be a $V^* \times V$ matrix containing the basis functions for $x(v)$ on the grid $\mathbf{v}$. Considering the discretely sampled functions in matrix form, we can write the basis expansion as $\mathbf{Y} = \mathbf{Y}^*\Xi$ and $\mathbf{X} = \mathbf{X}^*\Phi$, with $\mathbf{Y}^*$ and $\mathbf{X}^*$ being $N \times T^*$ and $N \times V^*$ matrices, respectively, containing the basis coefficients for the observed functions. Here we assume that Φ and Ξ are of full row rank, possibly but not necessarily orthogonal, so rank$(\Phi) = V^*$, rank$(\Xi) = T^*$ and $\Phi\Phi'$ and $\Xi\Xi'$ are invertible matrices of size $V^* \times V^*$ and $T^* \times T^*$.

Replacing each functional quantity in Model (3) with its basis expansion, we have

$$\mathbf{Y}^*\Xi=\mathbf{X}^*\Phi\Phi'\beta^*\Xi+Z\mathbf{U}^*\Xi+\mathbf{E}^*\Xi, \quad (4)$$

where β^* is $V^* \times T^*$, $\mathbf{U}^*$ is $n \times T^*$, and $\mathbf{E}^*$ is $N \times T^*$, representing quantities of Model (3) in the transformed basis space. When Φ is orthogonal so that $\Phi\Phi' = \mathbf{I}_{V^*}$, if we multiply each side of (4) by $\Xi^- = \Xi'(\Xi\Xi')^{-1}$, then we arrive at the *basis space model*

$$Y^{*} = X^{*}\beta^{*} + ZU^{*} + E^{*}. \quad (5)$$

When Φ is not orthogonal, we instead replace β^* in Model (5) with $\beta^{\dagger} = \Phi\Phi'\beta^*$. Thus, we can fit this basis space model after first transforming the functional responses and predictors to their respective basis spaces, $Y^* = Y\Xi^-$ and $X^* = X\Phi^-$, with $\Phi^- = \Phi'(\Phi\Phi')^{-1}$, and then after fitting the model, transform back to the original function space to obtain estimates and inference for $\beta = \Phi'\beta^*\Xi$ when Φ is orthogonal, $\beta = \Phi^-\beta^{\dagger}\Xi$ otherwise. Note that for some choices of basis functions, fast transform algorithms can be used in lieu of matrix multiplication to compute the basis functions or transform back to the original space, e.g., discrete wavelet transform (DWT) for wavelets, discrete Fourier transform (DFT) for Fourier bases, and fast algorithms for computing independent components (Hyvarinen et al., 2001).

We take a Bayesian approach to fit Model (5), using an Markov Chain Monte Carlo (MCMC) procedure to sample from the posterior distributions using appropriate prior distributions for each model parameter. Different choices of basis functions corresponds to different choices of Ξ and Φ. For example, for wavelets Ξ and Φ are inverse discrete wavelet transform (IDWT) matrices, for principal components they are the eigenvectors, possibly rescaled by the eigenvalues, for Fourier series they are the Inverse Discrete Fourier Transform (IDFT) matrices, and for splines they can be constructed based on B-splines or orthogonalized B-spline design matrices. Note that the same basis transform does not need to be used for both $y(t)$ and $x(\upsilon)$. In this paper, we use wavelet bases to represent the functional form of $y(t)$, and for $x(\upsilon)$, we use a composite strategy involving wavelets followed by principal components or wavelet Principal Components (wPC), which is similar to strategies used by Johnstone and Lu (2009) and Røislien and Winje (2012).

2.2 Model Formulation

Here, we present our modeling details using wavelets for $y(t)$ and wPC for $x(\upsilon)$. First, we transform the functions to the wavelet space by applying the $O(T)$ DWT to each row of Y and X, which can be represented as

$$\mathbf{y}_{ic} \xrightarrow{\text{DWT}} \mathbf{y}_{ic}^{*} = \left\{ y_{ic,jk}^{*} \right\}$$

and

$$\mathbf{x}_{ic} \xrightarrow{\text{DWT}} \mathbf{x}_{ic}^{*} = \left\{ x_{ic,s\ell}^{*} \right\}$$

. Wavelets are multi-resolution bases that are double-indexed by scale and location. The scales are $j = 1, \ldots, J^y$ and $s = 1, \ldots, S^x$ and locations

$$k = 1, \ldots, K_j^y$$

and

$$\ell = 1, \ldots, L_s^x$$

for **Y** and **X**, respectively. The dimension of

$$\mathbf{y}_{ic}^*$$

is $1 \times T^*$ where

$$T^* = \sum_{j=1}^{J^y} k_j^y$$

. Similarly,

$$\mathbf{x}_{ic}^*$$

has dimensions $1 \times V^*$ where

$$V^* = \sum_{s=1}^{S^x} \ell_s^x$$

. After performing a DWT on each row of **X**, we then compute the wavelet-space PC scores by applying a singular value decomposition to obtain the matrix of right singular vectors. All coefficients can be kept for a lossless transform or a large number kept for a near-lossless transform.

Both transformations can be represented in matrix form. Thus performing the DWT on **Y** is equivalent to taking **Ξ** to be the matrix of wavelet basis functions evaluated on the T grid. And performing wPC on **X** is equivalent to letting **Φ** be the composite transform consisting of the product of the matrix of wavelet basis functions and the reduced matrix of right singular values. Web Appendix A has further discussion on the details of wPC.

Thus, after transforming the data, recall our basis space model (5) is given by $\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{Z}\mathbf{U}^* + \mathbf{E}^*$. Consistent with previous work (Morris and Carroll (2006), Morris et al. (2008), Zhu, Brown, and Morris (2011), among others), we assume independence in the wavelet space. That is, for the subject specific version of Model (5), $\mathbf{y}_{ic}^* = \mathbf{x}_{ic}^*\boldsymbol{\beta}^* + \mathbf{u}_i^* + \mathbf{e}_{ic}^*$, we assume $\mathbf{e}_{ic}^* \sim \mathcal{N}(0, \boldsymbol{\Sigma}_e^*)$ where $\boldsymbol{\Sigma}_e^*$ is a diagonal matrix with elements varying by j, k, $\boldsymbol{\Sigma}_e^* = \mathrm{diag}\left\{\sigma_{e(j,k)}^2\right\}$, and equivalently $\mathbf{u}_i^* \sim \mathcal{N}(0, \boldsymbol{\Sigma}_u^*)$ where $\boldsymbol{\Sigma}_u^* = \mathrm{diag}\left\{\sigma_{u(j,k)}^2\right\}$. The induced within-function covariances in the data space are given by $\boldsymbol{\Sigma}_e = \boldsymbol{\Xi}'\boldsymbol{\Sigma}_e^*\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}_u = \boldsymbol{\Xi}'\boldsymbol{\Sigma}_u^*\boldsymbol{\Xi}$, which with wavelets accommodates a broad class of covariances allowing heteroscedasticity and differing degrees of autocorrelation, and thus different degrees of borrowing of strength, in different regions of the function (Morris and Carroll, 2006).

The basis space independence assumption allows us to split Model (5) into a series of T^* separate models for each basis coefficient in the y-space, double-indexed by (j, k), giving

$\mathbf{y}^*_{(j,k)} = \mathbf{X}^* \beta^*_{(j,k)} + \mathbf{Z}\mathbf{u}^*_{(j,k)} + \mathbf{e}^*_{(j,k)}$, where $\mathbf{y}^*_{(j,k)}$ and $\mathbf{e}^*_{(j,k)}$ are $N \times 1$, $\beta^*_{(j,k)}$ is $V^* \times 1$, and $\mathbf{u}^*_{(j,k)}$ is $n \times 1$. $\mathbf{X}^*$ and $\mathbf{Z}$ are as previously defined. This separability allows computational scalability to extremely large T, as calculations are linear in T^*, sparse near-lossless basis functions frequently yield $T^* \ll T$, and when cluster computing resources are available, allows parallel computing across (j, k). For prior specification, we assume vague proper priors on the variance components and a spike-and-slab prior similar to that found in Morris and Carroll (2006), Malloy et al. (2010), and others. Note that the spike-and-slab prior for selecting among PCs has been used in other PC regression contexts (Joliffe, 1982; Aston et al., 2010; Yang et al., 2013). Posterior samples are generated for β^* and projected back into the data-space using $\beta = \Phi^- \beta^* \Xi$. These posterior samples are used to perform Bayesian inference on β, as detailed in Section 3. The MCMC algorithm used for estimation can be found in the Web Appendix. Code for running the above model is also available online (http://odin.mdacc.tmc.edu/~jmorris/FonF.zip). DWTs and wPC were performed using the MATLAB functions wavedec and princomp which are available in the wavelet and statistics toolboxes (MATLAB, 2013a).

2.3 More Complex Function-on-Function Mixed Models

Model (1) is a special case of a general function-on-function mixed model that incorporates arbitrary scalar covariates $\{\mathscr{X}_a, a = 1, \ldots, p_s\}$, functional covariates $\{X_a(\upsilon_a), a = 1, \ldots, p_f\}$, scalar-by-function interactions, and multiple levels of random effect covariates $\{Z_l^h, h = 1, \ldots, H; l = 1, \ldots, L_h\}$. In principle, our approach can also accommodate function-by-function interactions, but we omit that here. The general model can be written

$$
\begin{aligned}
y_i(t) = &\sum_{a=1}^{p_s} \mathscr{X}_{ia} B_a(t) + \sum_{a=1}^{p_f} \int_{\upsilon_a \in \nu_a} X_{ia}(\upsilon_a) \beta_a(\upsilon_a, t) d\upsilon_a \\
&+ \sum_{a_s=1}^{p_{s_I}} \sum_{a_f=1}^{p_{f_I}} \int_{\upsilon_{a_f} \in \nu_{a_f}} \mathscr{X}_{ia_s} X_{ia_f}(\upsilon_{a_f}) \beta_{a_s a_f}(\upsilon_{a_f}, t) d\upsilon_{a_f} + \sum_{h=1}^{H} \sum_{l=1}^{L_h} Z_{il}^h U_l^h(t) + E_i(t),
\end{aligned}
\tag{6}
$$

where $B_a(t)$ are functional coefficients for scalar predictors, $\beta_a(\upsilon_a, t)$ are function-on-function coefficient surfaces for functional predictors, $\beta_{a_s a_f}(\upsilon_{a_f}, t)$ coefficient surfaces for the interaction of scalar covariate a_s and functional predictor a_f, and the random effects $U_l^h(t) \sim \mathscr{GP}(0, \Sigma_U^h)$. The multiple levels of random effects allow the model to handle various types of multi-level models needed to accommodate many complex designs commonly encountered in practice. Our code is capable of fitting this general model.

For the ERP data considered in Section 5, we include a discrete factor *image type* both as a main effect as well as effect modifier for the functional predictor, which allows different functional intercepts and function-on-function regression surfaces for each image type. See Model (9) in Section 5 for specification. Inference can then be performed on any number of desired statistics resulting from the model.

2.4 Model Assessment

To assess our models we examine a functional R^2 as well as the proportion of total variance contributed by the random effects. For the functional R^2 measures, we propose the use of an

average functional R^2 which Harezlak et al. (2007) implemented. Given the predicted values $\hat{y}_{ic}(t)$, we formulate R^2 as

$$R_{ave}^2 = \frac{1}{T}\int_0^T R^2(t)dt \quad\text{where}\quad R^2(t) = 1 - \frac{\sum_{c=0}^C \sum_{i=1}^n (y_{ic}(t) - \hat{y}_{ic}(t))^2}{\sum_{c=0}^C \sum_{i=1}^n (y_{ic}(t))^2}.$$

Since we sample on a grid, the integral above can be approximated as

$R_{ave}^2 \approx \frac{1}{T}\sum_0^T R^2(t)\Delta(t)$ where (t) is the distance in time between measurements. For assessing the contribution of the random effects we use the ratio $\mathrm{tr}(\Sigma_U)/\mathrm{tr}(\Sigma_U + \Sigma_E)$ where tr denotes the trace.

3. Posterior Functional Inference

Previous work in the function-on-function setting has focused solely on estimation or inference based on the construction of point-wise confidence intervals over the surface considering intervals that don't contain zero as significant (Scheipl, Staicu, and Greven, 2014). However, such an approach does not account for the inherent multiple testing problem from testing multiple locations within the coefficient surface. When applied to Bayesian credible intervals, we refer to this as the point-wise credible interval (PWCI) procedure. This unadjusted approach may lead to coefficients spuriously designated as significant. Thus we propose two posterior functional inference procedures aimed at identifying significant regions of a surface while controlling overall α, either using false discovery rate or experiment wise error rate, plus a Bayesian global test for testing whether the regression surface is identically zero.

First, we extend to the function-on-function setting previous Bayesian False Discovery Rate (BFDR) approaches that were originally proposed by Müller, Parmigiani, and Rice (2006), and then extended to functional regression by Morris et al. (2008) and Malloy et al. (2010). The BFDR is reliant upon the selection of δ-fold intensity change. Ideally this value is biologically motivated, however such a value may not necessarily exist or may be difficult to determine. Therefore, we also consider joint credible bands similar to those considered by Ruppert, Wand, and Carroll (2003) and invert them to construct what we call Simultaneous Band Scores (SimBaS) that are based on experiment wise error rate.

Suppose we have M MCMC samples. Let $\beta^{(m)}(\upsilon, t)$ be one posterior sample of the estimated surface for sample m, $m = 1, \ldots, M$. Then for a specific υ, $\upsilon = 1, \ldots, V$, and t, $t = 1, \ldots, T$, with δ a pre-determined intensity change of practical interest, consider

$$P_{BFDR}(\upsilon, t) = Pr\{|\beta(\upsilon, t)| \le \delta | y\} \approx \frac{1}{M}\sum_{m=1}^M 1\{|\beta^{(m)}(\upsilon, t)| \le \delta\}.$$

Note that standardized predictors should be used so δ has a consistent scale across $\beta(\upsilon, t)$. To correct for the discrete nature of the MCMC we replace any $P_{BFDR}(\upsilon, t) = 0$ with the

quantity $(2M)^{-1}$. As discussed by Morris et al. (2008), this quantity can be interpreted as a local FDR for a true discovery defined by an effect size of at least δ in magnitude.

For a pre-specified global FDR-bound α, we select the set of points (locations) satisfying $\psi = \{(\upsilon, t) : P_{BFDR}(\upsilon, t) \quad \nu_\alpha\}$. To obtain ν_α, we sort $\{P_{BFDR}(\upsilon, t), \upsilon = 1, \ldots, V, t = 1, \ldots, T\}$ in ascending order across all sets of locations. This gives us the set $\{P_{(r)}, r = 1, \ldots, R\}$, where $R = VT$ or the ordered set of probabilities calculated above. We then define

$\lambda = \max \left[r^* : \frac{1}{r^*} \sum_{r=1}^{r^*} P_{(r)} \leq \alpha \right]$. The cutoff for selecting significant coefficients is then $\nu_\alpha = P_{(\lambda)}$.

Alternatively, in the spirit of Ruppert, Wand, and Carroll (2003), consider constructing joint credible bands. A $100(1 - \alpha)\%$ credible band of $\beta(\upsilon, t)$ must satisfy

$$Pr\{L(\upsilon, t) \leq \beta(\upsilon, t) \leq U(\upsilon, t) \forall \upsilon \in \nu, t \in \mathscr{T}\} \geq 1 - \alpha, \quad (7)$$

where $L(\upsilon, t)$ and $U(\upsilon, t)$ are the lower and upper bounds respectively. It follows from Ruppert, Wand, and Carroll (2003) that an interval satisfying (7) is

$$I_\alpha(\upsilon, t) = \hat{\beta}(\upsilon, t) \pm q_{(1-\alpha)} \left[\text{St.}\hat{\text{D}}\text{ev} \left\{ \hat{\beta}(\upsilon, t) \right\} \right],$$

where $\hat{\beta}(\upsilon, t)$ and $\text{St.}\hat{\text{D}}\text{ev} \left\{ \hat{\beta}(\upsilon, t) \right\}$ are the mean and standard deviation for a given (υ, t) taken over all M MCMC samples, adjusted for autocorrelation as described in Web Appendix G. The variable $q_{(1-\alpha)}$ is the $(1 - \alpha)$ quantile taken over M of the quantity

$$Z^{(m)} = \max_{\upsilon \in \nu, t \in \mathscr{T}} \left| \frac{\beta^{(m)}(\upsilon, t) - \hat{\beta}(\upsilon, t)}{\text{St.}\hat{\text{D}}\text{ev} \left\{ \hat{\beta}(\upsilon, t) \right\}} \right|.$$

These joint bands control for multiple testing in a strong experiment-wise fashion while also not requiring a pre-specified δ-fold intensity change as in the BFDR.

Now consider inverting these joint bands, constructing $I_\alpha(\upsilon, t)$ for multiple levels of α and determining for each (υ, t) the minimum α at which each interval excludes zero, denoted $P_{SimBaS}(\upsilon, t) = \min \{\alpha : 0 \notin I_\alpha(\upsilon, t)\}$, which can be directly computed by

$$P_{SimBaS}(\upsilon, t) = \frac{1}{M} \sum_{m=1}^{M} \left\{ \left| \frac{\hat{\beta}(\upsilon, t)}{\text{St.}\hat{\text{D}}\text{ev} \left\{ \hat{\beta}(\upsilon, t) \right\}} \right| \leq Z^{(m)} \right\}. \quad (8)$$

We call these probabilities Simultaneous Band Scores or SimBaS. Similar to the BFDR and PWCI, we can select a specific α and identify (υ, t) for which $P_{SimBaS}(\upsilon, t) < \alpha$ as significant, which is equivalent to checking if the joint credible intervals cover zero at a specific α-level. We can also compute global Bayesian p-values (GBPV), $P_{GBPV} = \min_{\upsilon,t}\{P_{SimBaS}(\upsilon, t)\}$, a measure for testing the global null hypothesis that $\beta(\upsilon, t) = 0 \; \forall \; \upsilon \in \nu, t \in \mathscr{T}$, when desired.

The BFDR, SimBaS, and GBPV can be computed for individual surfaces $\beta(\upsilon, t)$ or any transformation or contrast defined across surfaces. For example, in the two surface setting, interest focuses on applying the procedure to both $\beta_g(\upsilon, t)$, $g = 0, 1$, as well as the difference surface, $D(\upsilon, t) = \beta_1(\upsilon, t) - \beta_0(\upsilon, t)$. This allows us to detect differences between surfaces corresponding to two different experimental conditions or groups and determine where those differences occur.

4. Simulation

We generate data in two phases. First, we draw $\mathbf{x}_{ic}$, $\mathbf{u}_i$, and $\mathbf{e}_{ic}$. Second, we generate $\mathbf{y}_{ic}$ using $\mathbf{y}_{ic} = \mathbf{x}_{ic}\beta + \mathbf{u}_i + \mathbf{e}_{ic}$, where β is one of four true surfaces of association. To generate predictor curves, random effects, and model errors, we use Gaussian Processes with auto-regressive 1 [AR(1)] covariance structures. Estimates for parameters of the covariance of $\mathbf{x}_{ic}$ come from estimating autoregressive parameters from the output of one electrode from our ERP data. We assign pairs of curves to each subject to induce repeated measures and consider three different sample sizes: $n = 25$, 50, and 100. Repeated measures brings the total number of observations up to $N = 50$, 100, and 200 respectively. We select parameters for the covariances of $\mathbf{u}_i$ and $\mathbf{e}_{ic}$ as $\sigma_u^2 = 0.05$ and $\rho_u = 0.75$ and $\sigma_e^2 = 0.1$ and $\rho_e = 0.5$ respectively. Additionally, the maximum value of any location on the true surface does not exceed 0.13. Prior to constructing $\mathbf{y}_{ic}$, we center and scale $\mathbf{x}_{ic}$ across i, c across all curves so that the variance at each time point is 1.

We select true surfaces to mimic biologically plausible time varying associations. The first and third rows of Figure 1 contains the heat maps of each surface. Each surface represents a different type of association, equations for which can be found in Web Appendix C. The ridge surface represents a relationship where the strongest association between $x(\upsilon)$ and $y(t)$ occurs along the line $\upsilon = t$. In other words, changes in $y(t)$ are associated with concurrent changes in $x(\upsilon)$. The lagged surface suggests a relationship where changes in $x(\upsilon)$ at a given time are associated with later changes in $y(t)$, but the strongest effect is delayed. The relationship between $x(\upsilon)$ and $y(t)$ in the immediate surface is similar to that in the lagged, however the strongest effect occurs immediately before dying off. The peak scenario demonstrates a setting where changes in $y(t)$ at a given time are associated with later changes in $x(\upsilon)$ and the association is characterized by a single peak. Finally, we consider one slightly more complicated scenario which involves a ridge effect that attenuates as both t and υ increase along with a single peak above the attenuated ridge.

For each surface, we generate 200 simulated data sets and draw 2000 posterior samples, discarding 1000 of them. We use Daubechies wavelets with four vanishing moments and three levels of decomposition. In preliminary simulations, zero-padding reduced edge effects better than symmetric-half point padding. Thus we implement zero-padding for all models. Motivated by the ERP data structure, we set the total number of time points in both time domains to be 225. For the wPC decomposition, we keep components accounting for 99.0% of the variability in $\mathbf{X}^*$. A single posterior estimate with near average rMSE for each surface can be found in Figure 1. Results from all three sample sizes were similar, thus we only present simulations for $n = 25$, $N = 50$ here. Results for $n = 100$, $N = 200$ can be found in the Web Appendix. For each dataset we also calculate root Mean Square Error (rMSE).

We also examine the performance of the BFDR, SimBaS, and GBPV procedures in simulation using a global α of 0.05. For the BFDR, we use a δ-intensity change of 0.05 which is roughly half the max signal from each surface. For comparison, we also generate unadjusted PWCIs. To evaluate the three procedures, we calculate false discovery rate, sensitivity, experiment-wise error rate (EWER), and type I error. Define false discovery rate, FDR_ε, as the number of selected locations (υ, t) with true value ε divided by the total number of selected locations. Next define the sensitivity, SEN_Y, as the number of selected locations (υ, t) with true magnitude $> Y$ divided by the total number of locations with true magnitude $> Y$. $EWER_\varepsilon$ is calculated as the proportion of simulated datasets with at least one falsely discovered location, i.e. a selected location with true value ε. Type I error is calculated using a null simulation with true surface $\beta(\upsilon, t) = 0 \; \forall \; \upsilon \in \nu, t \in \mathcal{T}$ and determining the proportion of simulated datasets with at least one location identified as significant.

Figure 1 allows for direct comparison of each estimated surface to the truth. For all surfaces, we see the model performed quite well, effectively reconstructing all the true surfaces. Estimation improves as sample size increases. Not surprisingly, rMSE decreases as sample size increases though even the smallest sample size produced small rMSEs. Heat maps containing the averaged set of identified coefficients for the BFDR and the average SimBa scores across datasets can be found in the Web Appendix. Both procedures correctly identified regions of elevated association in all four surfaces.

Table 1 displays both the average false discovery rate, FDR_ε, and the average sensitivity, SEN_Y, for each scenario using $\varepsilon = 0.01, 0.05$ and $Y = 0.05, 0.075$. For each procedure, we use $\alpha = 0.05$ to select the set of selected locations. We can see that the BFDR and SimBaS procedures performs similarly well by both measures, though BFDR does better for a higher ε and Y. While the PWCI has very good sensitivity, it comes at the cost of an inated false discovery rate. $EWER_\varepsilon$ is calculated using $\varepsilon = 0.01$. Additionally, SimBaS controls experiment-wise type I error quite well at 0.05. While BFDR has a slightly low type I error of 0.04, PWCI has a very high value of 0.645. To assess P_{GBPV} we determine the percent of datasets under each scenario with $P_{GBPV} < 0.05$ which was all datasets in each scenario.

These simulation results suggest our method performs well both in estimation and in inference. Even at the smallest sample size we considered, the model effectively reproduces the true surface. Both the BFDR and SimBaS capture the regions of strongest association without spuriously selecting too many non-significant coefficients. They also control well for type I error. Further, BFDR and SimBaS outperform the PWCI while maintaining reasonable sensitivity. Increasing sample size improves these facets of the model. Our method using wavelets and wPCs better captured the local features of these simulations than the spline-based methods of Scheipl, Staicu, and Greven (2014), as shown in Web Appendix F.

5. Application

5.1 Description of ERP Data Set

To illustrate the performance of our proposed methods, we analyzed data from the Department of Behavioral Sciences at the University of Texas M. D. Anderson Cancer Center. As part of a smoking cessation trial, researchers obtained Event Related Potentials (ERPs) at baseline for subjects viewing a series of images of different types, including neutral, emotional (positive and negative), and cigarette-related.

EEG was continuously recorded during image presentation and collected using a 129-channel Geodesic Sensor Net and amplified with AC-coupled high-input impedance (200 MΩ) amplifier (Geodesic EEG System 250; Electrical Geodesics, Inc., Eugene, OR) referenced to the Cz electrode. The time series were preprocessed as described in Versace et al. (2010a), with 0.1Hz high pass and 100Hz low pass filters, blink-corrected using spatial filtering, transformed to average reference, segmented into 900ms segments from 100ms before each image shown to 800ms after, obvious artifacts removed, and ERPs obtained by averaging across images for each image type per subject/electrode. After this processing, for each subject, we are left with functions of length 225 for each image type for all 129 electrodes.

Example curves recorded from 180 participants at electrode Cz (#129, in the middle of the crown of the head) during presentation of cigarette-related and neutral images can be seen in Figure 2 with the average over curves included in red. Curves under the other image-types are similar in appearance. The irregularity and localized spikiness of the raw curves motivates our use of wavelets and wPC in our modeling approach (Figure 2).

While many analyses are of interest for these data, in this paper we aim to characterize the time-varying relationship between ERP output from pairs of electrodes, focusing on two pairs in particular. The first pair is 55 and 129. Electrode 129, as previously mentioned, is positioned at the top of the head and electrode 55 is located directly behind it. We expect these two adjacent electrodes to be positively associated along the diagonal, $t = \upsilon$, axis. The second pair is 75 and 11. Electrode 75 is an occipital electrode located at the back of the head while electrode 11 is at the front. Output from these two electrodes is expected to exhibit a negative correlation and thus we anticipate a negative association along the diagonal axis. For each pair of electrodes, we jointly model the association between the electrodes under both the neutral and cigarette image conditions resulting in a multilevel data structure. Thus, for each model, subjects have four curves resulting from measurements from two electrodes while viewing two different image types. We would like to assess the inter-electrode relationships, determine whether these relationships vary over the time the image is shown, and differ by image type. Note that our goal in this analysis is not a broad-scale analysis of all pairwise associations, for which other approaches may be more suitable (e.g. Montagna et al. (2012)), but rather perform function-on-function regression analyses of pre-specified pairs.

5.2 Analysis

We fit two models to the data. In general, the model is given by

$$y_{icg}(t)=1(g=0)\left[\alpha_0(t)+\int_{\upsilon\in\nu}x_{ic0}(\upsilon)\beta_0(\upsilon,t)d\upsilon\right]+1(g=1)\left[\alpha_1(t)+\int_{\upsilon\in\nu}x_{ic1}(\upsilon)\beta_1(\upsilon,t)d\upsilon\right]+U_i(t)+E_{ic}(t), \quad (9)$$

where g denotes group membership, 0 for neutral, 1 for cigarette. For the first model we used Electrode 129 as the outcome function and Electrode 55 as the predictor function. For the second we used Electrode 11 as the outcome and Electrode 75 as the predictor. In both models, inference was drawn on both image-specific surfaces, β_0 and β_1, as well as the difference surface $D(\upsilon, t) = \beta_1(\upsilon, t) - \beta_0(\upsilon, t)$. As in the simulation study, we used Daubechies wavelets with four vanishing moments, three levels of decomposition, and zero-padding. Prior to decomposition, we standardized both outcome and predictor functions by time. After DWT, the dimensions of the transformed functional outcomes from Electrodes 129 and 11 were both 360×245. After wPC and preserving 99% of the total variability in the signals, the dimensions of the transformed functional predictors were 360×72 for Electrode 55 and 360×62 for Electrode 75. We obtained 1000 posterior samples from the MCMC after a burn-in 1000.

MCMC convergence was assessed using Geweke's Convergence Diagnostic (Geweke, 1992) and evaluation of first order auto-correlation coefficients for each coefficient on the surface. Both diagnostics suggest chain convergence. Details of the diagnostics along with histograms of Metropolis-Hastings acceptance rates can be found in the Web Appendix. For the model regressing Electrode 129 on to 55, the 95% interval of acceptance rates across all variance components was (0.431, 0.777) with a mean of 0.588. Acceptance rates for the model regressing Electrode 11 on to 75 were similar with 95% interval (0.405, 0.787) and mean 0.56.

We considered inference for both models using all three procedures. For the BFDR procedure, we selected a global α of 0.05 when implementing it on the difference surfaces. We choose a somewhat strict intensity change of $\delta = 0.05$ to focus on large differences between the surfaces. Choice of δ will potentially affect results, thus additional choices of δ can be found in the Web Appendix. We also implemented BFDR on the image-specific surfaces in both models. There the α-level was reduced to 0.025 for each surface, however the intensity change, δ, remained at 0.05 so to only select relatively large associations. For comparability, the same intensity change was used for both models. For the PWCI, we also used $\alpha = 0.05$.

Figure 3 contains posterior means of all three surfaces for both models. Examination of the posterior estimates of the difference surfaces found in the first column of Figure 3 suggest little to no systematic difference between image type in both models. When we look at the image-specific surfaces in the model using electrodes 129 and 55 (top row, second and third column, Figure 3), we see an elevated ridge of association along the $t = \upsilon$ diagonal, which is the relationship we anticipated between these two adjacent electrodes. Note that this relationship is strongest in the first 300 ms in the ERP or 200 ms post picture presentation (image presentation occurred at $t = \upsilon = 0$), corresponding to the initial response to viewing the image. Transitioning to the image-specific surfaces of the model using electrodes 11 and 75 (bottom row, second and third column, Figure 3), we see a valley of negative association

along the $t = \upsilon$ diagonal that also begins to die out around 200 ms to 300 ms past presentation. Once again, this is consistent with the expected relationship between these two electrodes.

Figure 4 contains results from the BFDR procedure on the difference surface for both models. Each heat map plots the posterior probabilities P_{BFDR}. We see that for both models, most locations have a low probability of being greater than δ. In fact, plotting ψ, we see no regions identified as significant (see the Web Appendix), suggesting there is little evidence that the correlation across the two electrodes differs across image types. The second and third columns of Figure 4 show the application of the BFDR to the image-specific surfaces in each model, and again the heat maps plot the posterior probabilities P_{BFDR}. We see that the probabilities along the ridge are quite large suggesting that ridge of positive association is significant up until almost 300 ms past image presentation. However the negative association along the ridge we saw in the second model has lower probabilities along the ridge at the $\delta = 0.05$ cut-off. The regions identified as significant by the procedure as well as additional choices of δ are presented in heat maps in Web Appendix.

For the SimBaS procedure, we plot heat maps of the logged SimBa Scores in Figure 5. We see that for both difference surfaces, the SimBa scores are all relatively large (at least 0.5 or more), and the global Bayesian p-value for both is $P^D_{GBPV} = 0.5$, suggesting there is not enough evidence to conclude differences in the coefficient surfaces between image types. The second and third columns of Figure 5 show the SimBaS procedure applied to the image-specific surfaces. These heat maps are also plotted on the log-scale so to distinguish variations in small SimBa scores. For both models, we see evidence of a non-zero coefficient surface for each image type (

$$P^0_{GBPV} = P^1_{GBPV} = 0.001$$

for the model using Electrodes 129 and 55,

$$P^0_{GBPV} = 0.001$$

and

$$P^1_{GBPV} = 0.005$$

for the model using Electrodes 11 and 75). Additionally, the SimBaS procedure detects the ridge of positive association in first model but only finds some of the negative associations in second.

Heat maps of significant locations using PWCI can be found in the Web Appendix. Not surprisingly, the PWCI is more sensitive to minor variations in the surface where there appears to be no systematic association. While both BFDR and SimBaS found no significant locations in the difference surfaces, the PWCI discovers a number of regions and also finds

a number of significant locations in the image-specific surfaces that are off the $t = \upsilon$ axis while suggesting the association lingers longer. Given the results in the simulation studies, we interpret these results cautiously, as they may likely be spurious, and feel more confident in the multiplicity-adjusted inference from the BFDR and SimBaS procedures.

We calculated functional R^2 as described in Section 2.4. The model regressing Electrode 129 on to 55 had $R^2_{ave} = 0.698$ while the model regressing Electrode 11 on to 75 had $R^2_{ave} = 0.455$. We also examined the proportion of the variance attributable to the random effects. For the model regressing Electrode 129 on to 55, the random effects contributed to 35.6% to the total variability and residuals 64.4%, while for Electrode 11 regressed on 75 random effects contributed to 43.5% to the total variability and residuals 56.5%.

6. Discussion

Functional data analysis is an expanding field requiring more work to fill in gaps in the literature and build upon the general knowledge of the field. Previous work on function-on-function regression is limited. Here we present a general approach to function-on-function regression modeling which benefits from several attributes. First, our approach can use any basis function for $y(t)$ and $x(\upsilon)$ allowing us to handle functions of various types, including those with spiky and smooth features, can be used with functions on higher dimensional domains like images, and allows us to parsimoniously model correlated residuals rather than assuming iid errors. Second, we get fully Bayesian inferences on all model quantities including point-wise credible intervals, posterior probabilities interpretable as Bayesian FDRs, joint credible intervals, and SimBaS that provide global and experiment-wise inferential quantities. Third, our inference procedures correctly identify regions of elevated association without falsely identifying too many non-significant coefficients. Fourth, our method resides within the functional mixed model (FMM) framework as put-forth by Morris and Carroll (2006) that handles correlation between functions and random effects through random effect function distribution, and thus accounting for the various sources of variability in multi-level models. Using the mixed model representation of penalized splines, this also allows the relaxation of the linearity assumption between $Y(t)$ and $X(\upsilon)$, although details are beyond the scope of this paper. Finally, the FMM framework also allows any combination of continuous and discrete scalar predictors, functional predictors, and their interactions, allowing function-on-function regression to be done in a much broader modeling context.

We demonstrated by simulation that our model performs well for realistic sample sizes and forms of functional association with fits improving as sample size increases. Simulations also show the BFDR and SimBaS procedures have better false discovery and type I error rates than the PWCI with comparable sensitivity. Our approaches for global inference and multiple-testing adjustment for Bayesian inference using BFDR, SimBaS, and GBPV are of general interest and can be used in other functional regression settings.

Our application displays the ability of the model to estimate the forms of the relationship of ERP output between different electrodes on the scalp. With the neighboring electrodes, a positive association was expected and seen along the diagonal axis $t = \upsilon$ while a negative

association was expected and seen between electrodes on opposite sides of the scalp. Further, both our inference procedures were able to detect these associations as significant, even the one based on experiment wise error rate.

In summary, the function-on-function mixed model with basis-space modeling comprises a flexible approach to the function-on-function regression setting. The method performed well in both simulation and application. Further studies are needed to explore the model's performance in more complex settings, including non-functional components beyond a factor variable, incorporating multiple functional predictors, and various types of random effect correlation structures. Additionally, further examination of data reduction techniques could improve the modeling prowess of the method.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Aston JA, Choiu J, Evans J. Linguistic pitch analysis using functional principal component mixed effect models. Journal of the Royal Statistical Society, Series C. 2010; 59:297–317.

Brumback BA, Rice JA. Smoothing spline models for the analysis of nested and crossed samples of curves. Journal of the American Statistical Association. 1998; 93:398–408.

Cardot H, Ferraty F, Sarda P. Functional linear model. Statistics & Probability Letters. 1999; 45:11–22.

Chen K, Müller HG. Modeling Repeated Functional Observations. Journal of the American Statistical Soceity. 2012; 107:1599–1609.

Cinciripini PM, Robinson JD, Karam-Hage M, Minnix JA, Lam CY, Versace F, Brown VL, Engelmann JM, Wetter DW. The effects of varenicline and bupropion-SR plus intensive counseling on prolonged abstinence from smoking, depression, negative affect, and other symptoms of nicotine withdrawal. JAMA Psychiatry. 2013; 70:522–533. [PubMed: 23536105]

Gertheiss J, Maity A, Staicu A-M. Variable Selection in Generalized Functional Linear Models. Stat. 2013; 2:86–101. [PubMed: 25132690]

Geweke, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernado, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. Bayesian Statistics. Vol. 4. Oxford, UK: Clarendon Press; 1992.

Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized Functional Regression. Journal of Computational and Graphical Statistics. 2011; 20:830–851. [PubMed: 22368438]

Goldsmith J, Greven S, Crainiceanu CM. Corrected Confidence Bands for Functional Data Using Principal Components. Biometrics. 2013; 69:41–51. [PubMed: 23003003]

Harezlak J, Coull BA, Laird NM, Magari SR, Christiani DC. Penalized solutions to functional regression problems. Computational Statistics & Data Analysis. 2007; 51:4911–4925. [PubMed: 18552972]

Hyvarinen, A.; Karhunen, J.; Oja, E. Independent Component Analysis. Vol. 26. Wiley-Interscience; 2001.

Ivanescu AE, Staicu A-M, Greven S, Scheipl F, Greven S. Penalized function-on-function regression. Johns Hopkins University, Dept. of Biostatistics Working Papers. 2012 Working Paper **254**, http://biostats.bepress.com/jhubiostat/paper254.

Johnstone IM, Lu AY. On Consistency and Sparsity for Principal Components Analysis in High Dimensions. Journal of the American Statistical Society. 2009; 104:682–693.

Joliffe IT. A note on the use of principal components in regression. Journal of the Royal Statistical Society, Series C. 1982; 31(3):300–303.

Kim K, entürk D, Li R. Recent history functional linear models for sparse longitudinal data. Journal of Statistical Planning and Inference. 2011; 141:1554–1566. [PubMed: 21691421]

Malfait N, Ramsay JO. The historical functional linear model. The Canadian Journal of Statistics. 2003; 31:115–128.

MATLAB and Wavelet and Statistics Toolboxes Release. Natick, MA, USA: The MathWorks, Inc.; 2013a.

Malloy EJ, Morris JS, Adar SD, Suh H, Gold DR, Coull BA. Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. Biostatistics. 2010; 11:432–452. [PubMed: 20156988]

McLean MW, Hooker G, Staicu A-M, Scheipl F, Ruppert D. Functional Generalized Additive Models. Journal of Computational and Graphical Statistics, to appear. 2012

Montagna S, Tokdar ST, Neelon B, Dunson DB. Bayesian latent factor regression for functional and longitudinal data. Biometrics. 2012; 68:1064–1073. [PubMed: 23005895]

Morris JS. Functional regression. Annual Reviews of Statistics and its Application to appear. 2015

Morris JS, Brown PJ, Herrick RC, Baggerly KA, Coombes KR. Bayesian Analysis of Mass Spectrometry Proteomic Data Using Wavelet-Based Functional Mixed Models. Biometrics. 2008; 64:479–489. [PubMed: 17888041]

Morris JS, Carroll RJ. Wavelet-based functional mixed models. Journal of the Royal Statistical Society, Series B. 2006; 68:179–199.

Müller P, Parmigiani G, Rice K. FDR and Bayesian multiple comparison rules. Johns Hopkins University, Dept. of Biostatistics Working Papers. 2006 Working Paper **115**, http://biostats.bepress.com/jhubiostat/paper115/.

Müller HG, Yao F. Functional Additive Models. Journal of the American Statistical Society. 2008; 103:1543–1544.

Ramsay JO, Dalzell CJ. Some tools for functional data analysis. Journal of the Royal Statistical Society: Series B. 1991; 53:539–561.

Ramsay, JO.; Silverman, BW. Functional Data Analysis. 2nd ed.. Springer; 2005.

Reiss PT, Huang L, Maarten M. Fast Function-on-Scalar Regression with Penalized Basis Expansions. The International Journal of Biostatistics. 2010; 6

Reiss PT, Ogden RT. Functional principal components regression and functional partial least squares. Journal of the American Statistical Association. 2007; 102:984–996.

Røislien J, Winje B. Feature extraction across individual time series observations with spikes using wavelet principal component analysis. Statistics in Medicine. 2012; 32:3660–3669. [PubMed: 23553851]

Ruppert, D.; Wand, MP.; Carroll, RJ. Semiparametric Regression. Cambridge University Press; 2003.

Scheipl F, Greven S. Identifiability in penalized function-on-function regression models. University of Munich Technical Report. 2012; 125

Scheipl F, Staicu A-M, Greven S. Functional Additive Mixed Models. Journal of Computational and Graphical Statistics, to appear. 2014 Available at *arXiv:1207.5947v5*.

Staicu A-M, Crainiceanu CM, Reich DS, Ruppert D. Modeling Functional Data with Spatially Heterogeneous Shape Characteristics. Biometrics. 2011; 68:331–343. [PubMed: 22050118]

Versace F, Minnix JA, Robinson JD, Lam CY, Brown VL, Cinciripini PM. Brain reactivity to emotional, neutral, and cigarette-related stimuli in smokers. Addiction Biology. 2010a; 16:296–307. [PubMed: 21182573]

Versace F, Robinson JD, Lam CY, Minnix JA, Brown VL, Carter BL, Wetter DW, Cinciripini PM. Cigarette cues capture smokers attention: Evidence from event-related potentials. Psychophysiology. 2010b; 47:435–441. [PubMed: 20070579]

Wang W. Linear Mixed Function-on-Function Regression Models. Biometrics. 2014

Yang WH, Wikle CK, Holan SH, Wildhaber ML. Ecological prediction with nonlinear multivariate time-frequency functional data models. Journal of Agricultural, Biological, and Environmental Statistics. 2013; 18(3):450–474.

Yao F, Müller HG, Wang JL. Functional Linear Regression Analysis for Longitudinal Data. The Annals of Statistics. 2005a; 33:2873–2903.

Zhu H, Brown PJ, Morris JS. Robust, adaptive functional regression in functional mixed model framework. Journal of the American Statistical Association. 2011; 106:1167–1179. [PubMed: 22308015]

Zhu H, Brown PJ, Morris JS. Robust classification of functional and quantitative image data using functional mixed models. Biometrics. 2012; 68(4):1260–1268. [PubMed: 22670567]
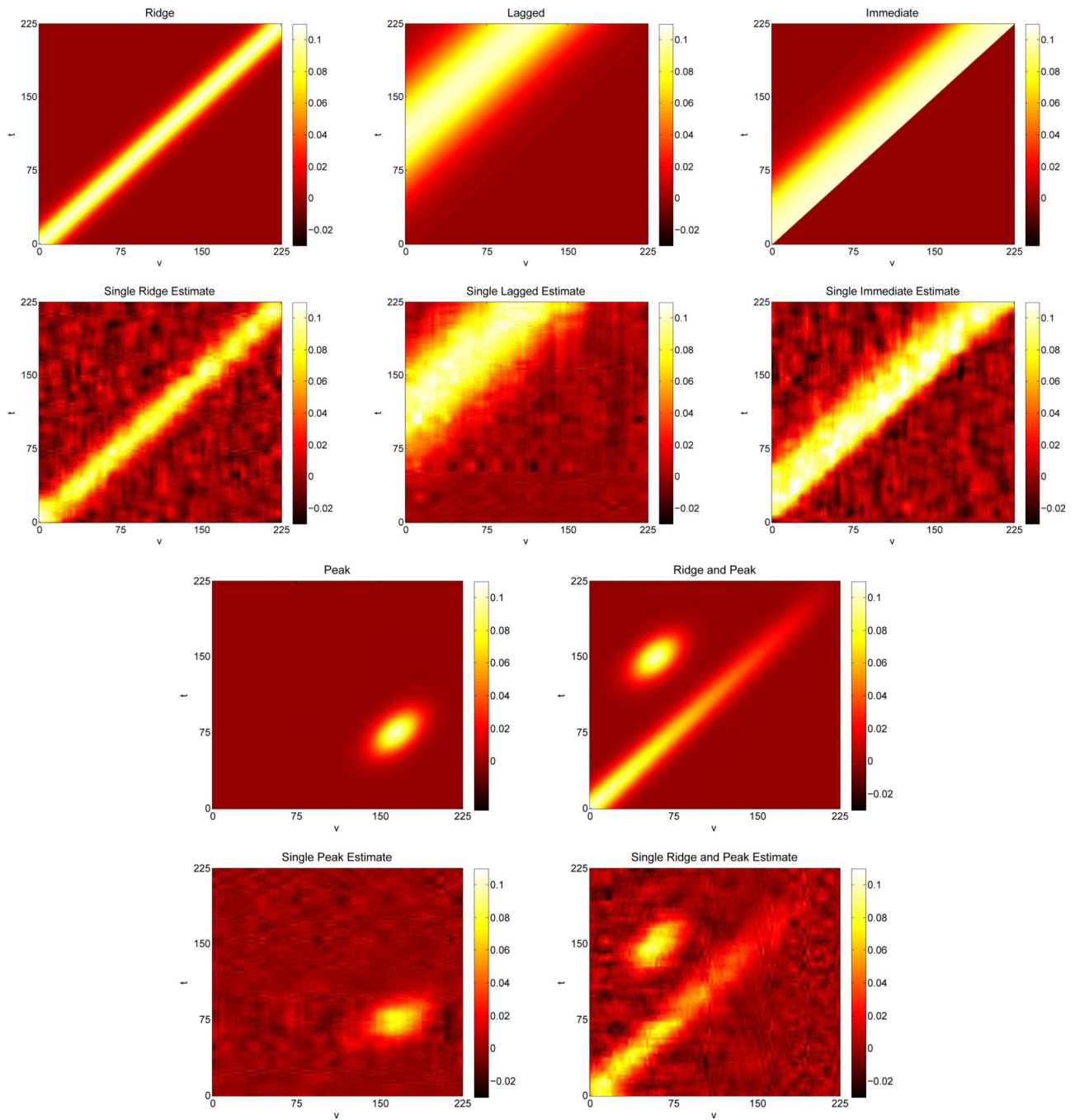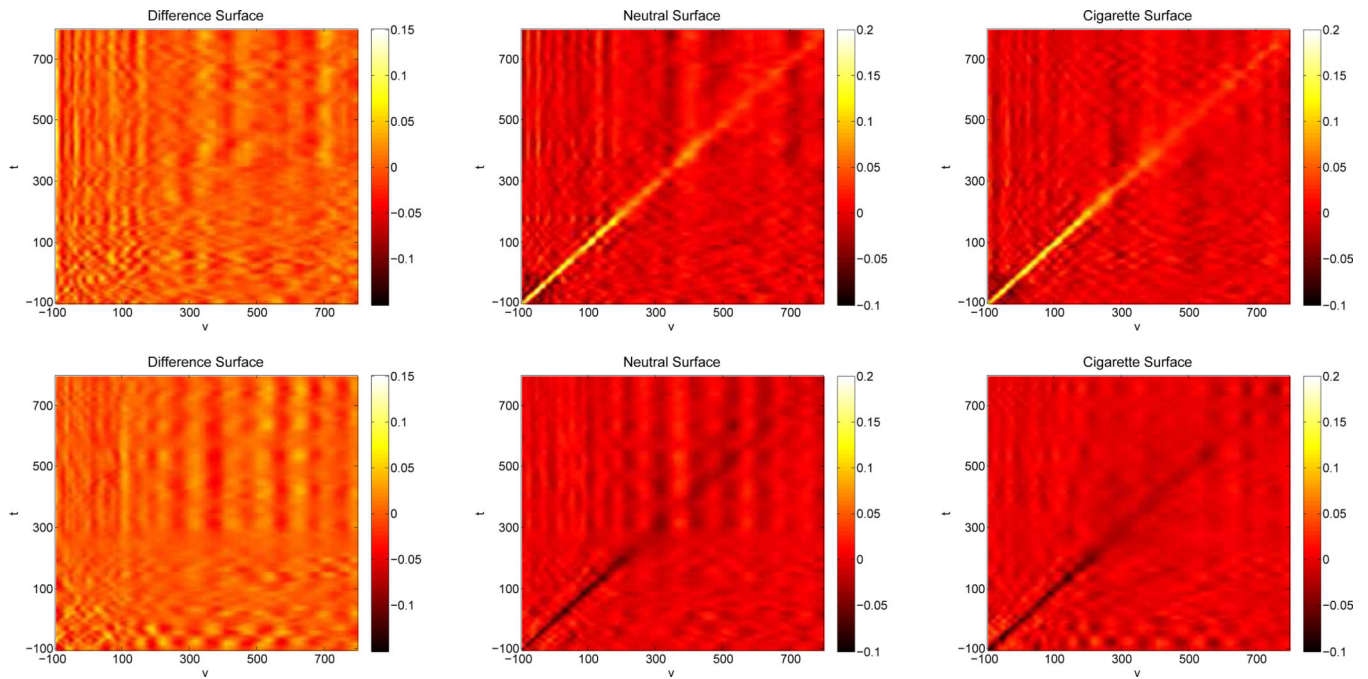
Figure 1.
Heat maps of the true surfaces for simulation study are above heat maps of a single estimated surfaces for each simulated scenario based on a sample size of $n = 25$ with two measure per subject, $C_i = 2 \,\forall\, i$, for a total of $N = 50$ observations. Each surface has near average rMSE and is based on a total of 2000 MCMC chains with the first 1000 discarded.
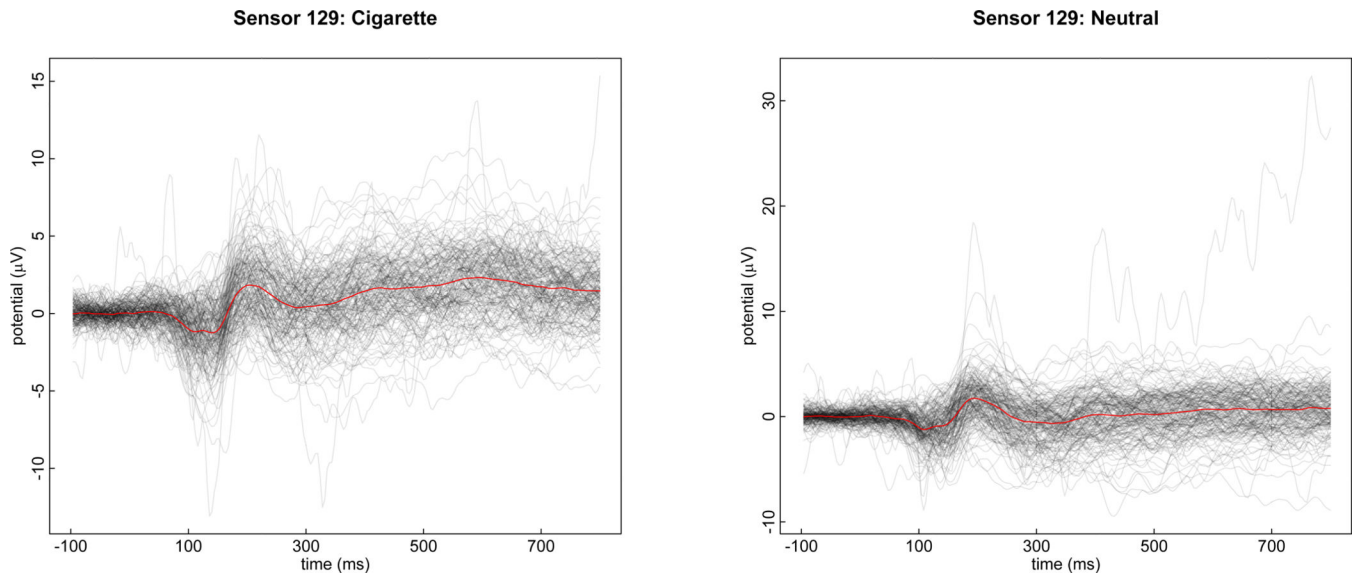
Figure 2.
On the left, raw profile curves are plotted in gray with the mean in red from electrode 129 under the cigarette image condition. On the right, are raw curves and the mean from electrode 129 under the neutral image condition.
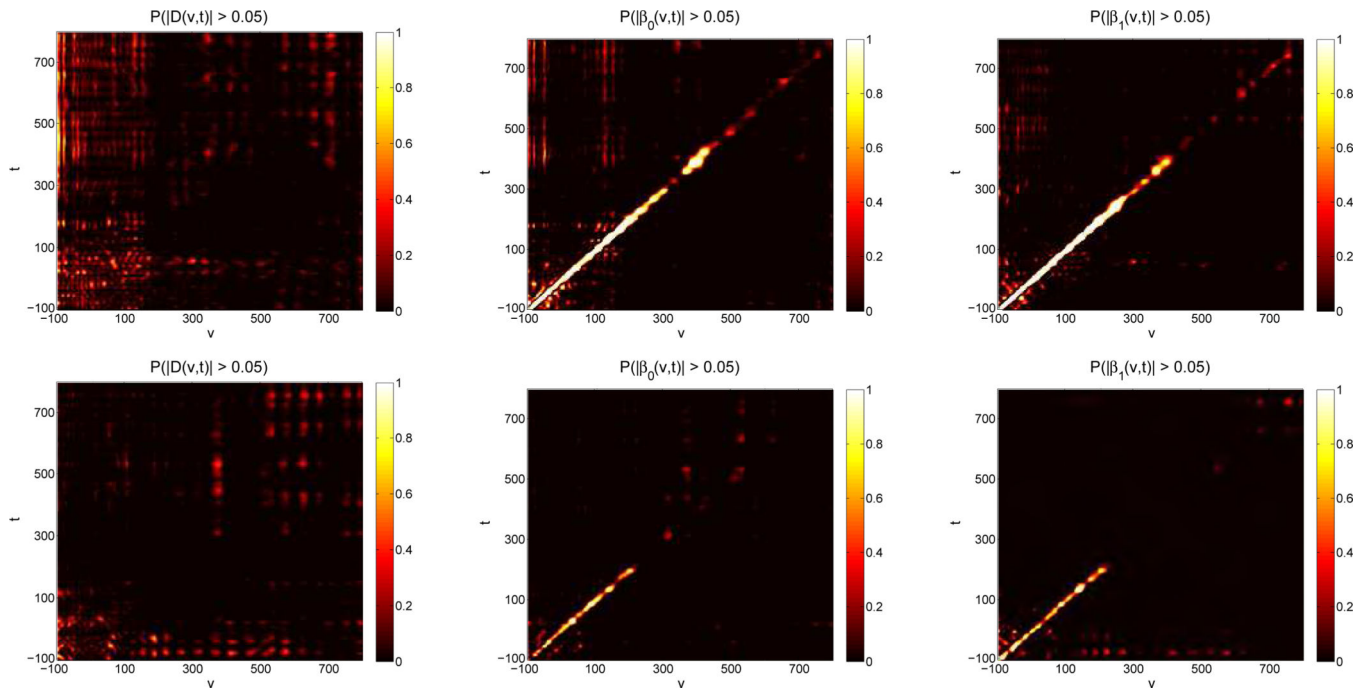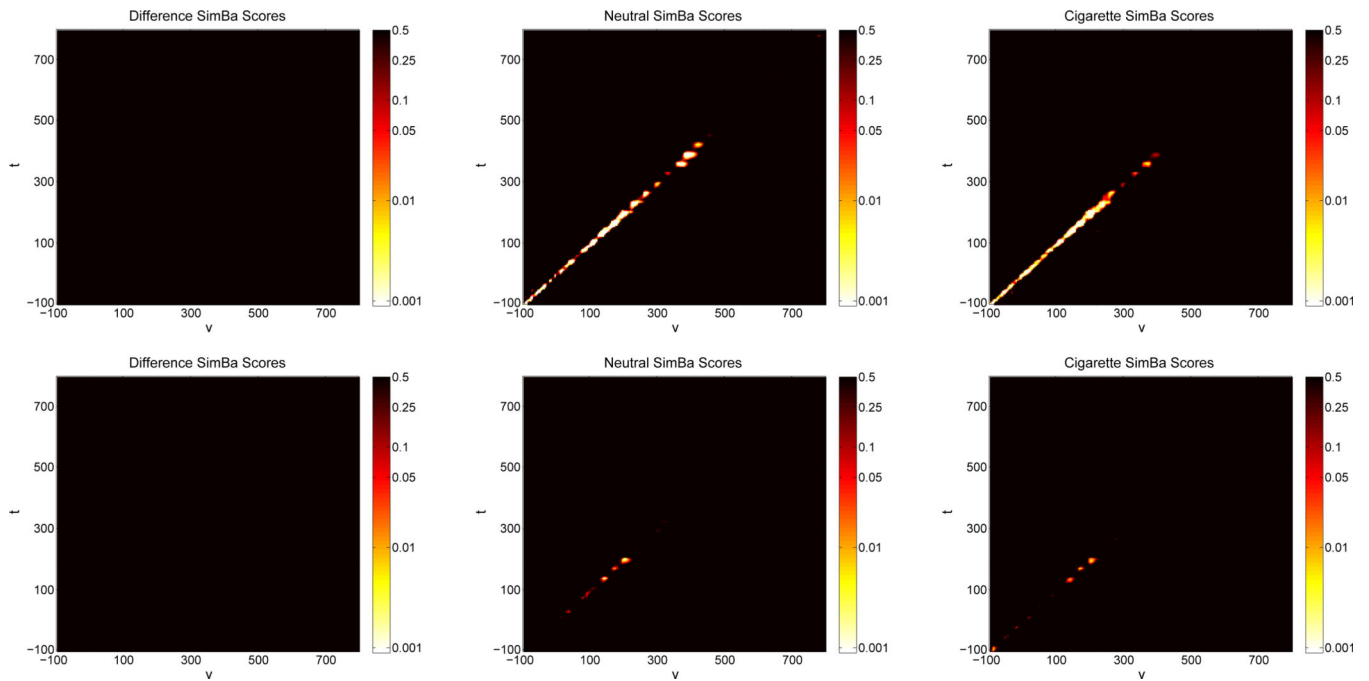
Figure 3.
The top row contains surface estimates for the association between electrodes 129 and 55. Posterior surfaces comparing electrodes 11 to 75 are in the second row. The estimated posterior surface of the difference between cigarette and neutral is found in the first column. Group specific surface estimates are in the second and third columns, Neutral and Cigarette respectively. ERP output from electrode 129 is the response and the output from electrode 55 is the predictor for the first model and electrode 75 is the predictor of electrode 11 in the second model.

Figure 4.
Heat maps containing the posterior probabilities from the BFDR procedure using a δ intensity change of 0.05. Locations (υ, t) in white have a high probability of being greater than δ and thus likely to be included in ψ, the set of locations identified as significant. Black locations (υ, t) have a low probability of being greater than δ and are thus less likely to be identified as significant. The top row contains results from the model using electrodes 129 and 55 while the second row contains results from the model using electrodes 75 and 11.

Figure 5.
Heat maps containing the SimBa scores for each surface of both models. The top row contains results from the model using electrodes 129 and 55 while the second row contains results from the model using electrodes 75 and 11. Scores are plotted on the log-scale with the color axis on the exponential scale. White regions represent coefficients with low SimBa scores, black regions represent coefficients with high SimBa scores.

Table 1

FDR, sensitivity, experiment-wise error rate (EWER), and type I error values by inference procedure. The BFDR use a δ intensity change of 0.05. To determine assessment values for SimBaS, a cutoff of $\alpha = 0.05$ was used. Likewise, the PWCI used 95% point-wise credible intervals to determine significant locations.

Measure	Surface	BFDR	SimBaS	PWCI
$FDR_{0.01}$	Lagged	0.06%	0.08%	5.80%
	Peak	0.48%	0.75%	22.9%
	Ridge	0.12%	0.19%	20.5%
	Immediate	2.25%	2.80%	20.9%
	Ridge + Peak	0.20%	0.35%	17.4%
$FDR_{0.05}$	Lagged	5.74%	13.9%	44.7%
	Peak	4.01%	20.4%	73.5%
	Ridge	9.75%	15.6%	53.3%
	Immediate	5.74%	7.58%	38.1%
	Ridge + Peak	5.45%	22.7%	73.4%
$SEN_{0.05}$	Lagged	98.1%	96.2%	99.9%
	Peak	64.9%	73.4%	99.9%
	Ridge	96.8%	93.4%	99.9%
	Immediate	97.9%	93.8%	99.9%
	Ridge + Peak	63.6%	76.3%	99.9%
$SEN_{0.075}$	Lagged	99.9%	99.3%	100%
	Peak	94.4%	88.2%	99.9%
	Ridge	99.8%	97.6%	99.9%
	Immediate	99.9%	96.2%	99.9%
	Ridge + Peak	95.7%	93.3%	99.9%
$EWER_{0.01}$	Lagged	7.00%	16.5%	100%
	Peak	4.50%	10.5%	100%
	Ridge	9.50%	49.0%	100%
	Immediate	100%	100%	100%
	Ridge + Peak	6.00%	29.0%	100%
Type I Error	Null	4.00%	5.00%	64.5%