Web-based Supplementary Materials for Bayesian Function-on-Function Regression for Multilevel Functional Data by Meyer, M.J., Coull, B.A., Versace, F., Cinciripini, P., and Morris, J.S.

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.
**email:* mark.meyer@bucknell.edu

December, 2013

Web Appendix A: Additional Model Formulation Details

Comments regarding use of wavelets: Noting that the matrix formulation of the transformations involve an evaluation of wavelet basis functions on grids of T and V for the DWT on $\mathbf{Y}$ and $\mathbf{X}$ respectively, if T and V are powers of two, this decomposition will result in $T^* = T$ and $V^* = V$ wavelet coefficients, thus represents a lossless transform. Otherwise, padding is done according to some chosen boundary condition (e.g. periodic, reflection, and f with zeros), in which case T^* and V^* are not exactly equal to but are of the same order as T and V. We discuss choice of padding further in the simulation study included in the manuscript.

Wavelets tend to provide sparse representations for many functions, so one can achieve data compression with a near-lossless transform by eliminating wavelet coefficients that are negligible in magnitude for all curves. Wavelet thresholding has been widely used for compression and denoising of individual functions, and Morris et al. (2011) introduced a *joint compression* approach for the multiple function setting that finds a minimal subset of wavelet coefficients that jointly preserves $100\alpha\%$ of the total energy for *all* functions in a set. Let T^{W^*} and V^{W^*} represent the total number of coefficients left after such joint compression.

We can write these wavelet basis expansions in matrix form as $\mathbf{Y} = \mathbf{Y}^W \mathbf{W}_y$ and $\mathbf{X} = \mathbf{X}^W \mathbf{W}_x$, where $\mathbf{W}_y$ and $\mathbf{W}_x$ are $T^{W^*} \times T$ and $V^{W^*} \times V$ matrices, respectively, containing the retained wavelet basis functions evaluated on the T and V grids. Given orthogonal wavelets, we can also represent the DWT in matrix form as $\mathbf{Y}^W = \mathbf{Y}\mathbf{W}_y'$ and $\mathbf{X}^W = \mathbf{X}\mathbf{W}_x'$, or if non-orthogonal they can be represented $\mathbf{Y}^W = \mathbf{Y}\mathbf{W}_y^-$ and $\mathbf{X}^W = \mathbf{X}\mathbf{W}_x^-$. Thus, in the notation of Section 2.1, if we use wavelet transforms with joint compression for both $y(t)$ and $x(v)$, then we effectively define $\mathbf{\Xi} = W_y$ and $\mathbf{\Phi} = \mathbf{W}_x$, with $\mathbf{\Xi}^- = \mathbf{W}_y'$ and $\mathbf{\Phi}^- = \mathbf{W}_x'$, $\mathbf{Y}^* = \mathbf{Y}^W$ and $\mathbf{X}^* = \mathbf{X}^W$, and $T^* = T^{W^*}$ and $V^* = V^{W^*}$.

Comments regarding use of PCs and wPCs: In our model, calculations are linear in T^* but quadratic in V^*, so dimension reduction in $\mathbf{X}^*$ has especially important computational benefits. While joint wavelet compression can provide some dimension reduction, use of Principal Components Analysis (PCA) can provide additional dimension reduction while retaining orthogonality. In particular, consider performing the singular value decomposition of $\mathbf{X}^W = \mathbf{X}\mathbf{W}_x' = \mathbf{Q}\mathbf{\Sigma}\mathbf{P}'$. Noting that $\mathbf{X}^W$ is $N \times V^{W^*}$, we see that $\mathbf{Q}$, the matrix of left singular vectors, is $N \times V^{W^*}$ and both $\mathbf{\Sigma}$, the diagonal matrix of singular values, and $\mathbf{P}$, the matrix of right singular vectors, are $V^{W^*} \times V^{W^*}$. Supposing we keep $V^{svd} \lll V^{W^*}$ principal components, we can compute the wavelet-space PC scores $\mathbf{X}^* = \mathbf{X}^W \mathbf{P}_{svd}$, where $\mathbf{P}_{svd}$ is a $V^{W^*} \times V^{svd}$ matrix computed from the leading V^{svd} rows of $\mathbf{P}$. As long as enough wavelet coefficients and PCs are kept to account for an extremely high proportion of variability (e.g. $> 99\%$), this results in a near-lossless transform. This composite basis function strategy is equivalent to computing $\mathbf{X}^* = \mathbf{X}\mathbf{\Phi}^-$ with ***composite transform*** $\mathbf{\Phi}^- = \mathbf{W}_x'\mathbf{P}_{svd}$ and inverse transform $\mathbf{\Phi} = \mathbf{P}_{svd}'\mathbf{W}_x$ of dimension $V^* = V^{svd}$. Note that one could simply define $\mathbf{\Phi}$ to be the eigenvectors of a direct SVD on $\mathbf{X}$, but this composite wPC approach has advantages in that the joint compression in the wavelet space (1) reduces the dimensionality of $\mathbf{X}$ to speed up calculation of the SVD, (2) performs some denoising of the functions in $\mathbf{X}$ before calculation of the SVD, and (3) borrows strength locally within the function, thus accounting for the functional nature of the data.

PCs or wPCs can also be used for the response, if desired. In that case, the transform is done as above, and then the modeling strategies of 2.2 can be directly applied. Note that this allows flexible correlation structures for the random effect and residual error functions, with $\mathbf{\Sigma}_\epsilon = \mathbf{\Xi}'\mathbf{\Sigma}_\epsilon^*\mathbf{\Xi}$ and $\mathbf{\Sigma}_u = \mathbf{\Xi}'\mathbf{\Sigma}_u^*\mathbf{\Xi}$, where $\mathbf{\Xi}$ is the eigenvector matrix and $\mathbf{\Sigma}_\epsilon^*$ and $\mathbf{\Sigma}_u^*$ are diagonal matrix of variance components estimated separately for each PC. 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). In fact, Joliffe (1982) states that the original intention in PC regression was to perform variable selection over the PC dimensions, not just truncate to keep a small number explaining a major portion of the variability as is commonly done by many fPC practitioners. Thus, it may be desirable to use a variable selection approach as underlies the spike-and-slab prior rather than just appliying truncation.

Comments for use of Splines: We do not incoporate spline bases in this paper, since they will not tend to work well with functions with local features like our application and simulation data. However, in other settings, they may be desired and can be used in our method. The details are the topic of a more extensive paper on this topic, but here is a general outline of how it can be done: Placing a knot at each grid location on t, let B be a $T \times T$ design matrix of cubic B-splines. Assuming a 2nd derivative roughness penalty as in smoothing splines, the design matrix can be orthogonalized with respect to the $T \times T$ integrated 2nd derivative penalty matrix to yield $\mathbf{\Phi}$, a set of Demmler-Reinsch basis functions. From this, the unpenalized spline coefficients can be computed by $\mathbf{Y}^* = \mathbf{Y}\mathbf{\Phi}^-$. Assuming common variance components across the spline coefficients for the errors and random effect functions and a Gaussian prior (special case of spike-slab with $\pi_k = 1$) for the fixed effect functions in the basis space induces a smoothing spline for the fixed effects, random effects, and residual errors (using vague proper priors for the fixed linear coefficient in the Demmler-Reinsch basis). This provides an alternative Bayesian approach to fit spline-based functional mixed models that can be fit with our existing code.

Web Appendix B: MCMC Sampler

Working from Model (5) of the manuscript, the independence of the basis space allows us to split the model into T^* separate models for each basis coefficient in the y-space giving the model

$$\mathbf{y}_{(j,k)}^* = \mathbf{X}^*\beta_{(j,k)}^* + \mathbf{Z}\mathbf{u}_{(j,k)}^* + \mathbf{e}_{(j,k)}^*. \tag{1}$$

We now place priors on the coefficients by noting that $\beta_{(j,k)}^* = \left\{\beta_{(p,jk)}^*\right\}$ where p indexes the V^* retained principal components, $p = 1, \ldots, V^*$. We place spike-and-slab priors on the elements of $\beta_{(j,k)}^*$ for a given j, k, and p via

$$\beta_{(p,jk)}^* \sim \ \gamma_{(p,jk)}\mathcal{N}(0, \tau_{pj}) + (1 - \gamma_{p,jk})d_0, \quad \gamma_{(p,jk)} \sim \mathcal{B}(\pi_{pj})$$

where $\mathcal{B}$ denotes a Bernoulli distribution and d_0 represents a point-mass distribution at zero. Regularization parameters τ_{pj} and π_{pj} can be estimated using an Empirical Bayes-type approach as seen in Morris and Carroll (2006) and Malloy et al. (2010). Alternatively, priors may be placed as done in Zhu, Brown, and Morris (2011). For our model, we place an inverse gamma prior on the variances, τ_{pj}, of the Normal components of the mixture and a beta distribution on the mixture probabilities, π_{pj}, of the Bernoulli with respective hyper-parameters a_τ, b_τ, a_π, and b_π.

Following from Morris and Carroll (2006), we integrate out the random effects and work with marginalized likelihood. Morris and Carroll (2006) notes that this improves mixing over a naïve Gibbs sampler. The sampler alternates between sampling $\beta_{(j,k)}^*$ and the covariance parameters which we denote as $\mathbf{\Sigma}^*$. The random effects $\mathbf{u}_{(j,k)}^*$ are sampled when desired. The procedure iterates through the following steps:

Step 1: For each y-space coefficient indexed by (jk), we sample the fixed effect p from the distribution $f(\beta^*_{(p,jk)}|\mathbf{y}^*_{(j,k)}, \beta^*_{(-p),jk}, \mathbf{\Sigma}^*)$ where $\beta^*_{(-p),jk}$ is the set of all fixed-effect coefficients at j,k except the pth. Morris and Carroll (2006) demonstrate that $f(\cdot)$ is a mixture of a point-mass at zero and a normal with mean $\mu_{p,jk}$ and variance $\varepsilon_{p,jk}$. The mixture probability $\alpha_{p,jk}$ is given by

$$\alpha_{p,jk} = \Pr\left(\gamma_{p,jk} = 1|\mathbf{y}^*_{(j,k)}, \beta^*_{(-p),jk}, \mathbf{\Sigma}^*\right) = O_{p,jk}/\left(O_{p,jk}+1\right)$$

where

$$O_{p,jk} = \pi_{pj}/(1-\pi_{pj})\mathrm{BF}_{p,jk} \text{ and } \mathrm{BF}_{p,jk} = (1+\tau_{p,jk}/V_{p,jk})^{-1/2}\exp\left\{\frac{1}{2}\zeta^2_{p,jk}(1+V_{p,jk}/\tau_{p,jk})\right\}$$

and the forms of the mean and variance of the normal are

$$\mu_{p,jk} = \hat{\beta}^*_{(p,jk),\mathrm{MLE}}(1+V_{p,jk}/\tau_{p,jk})^{-1} \text{ and } \varepsilon_{p,jk} = V_{p,jk}(1+V_{p,jk}/\tau_{p,jk})^{-1}$$

where $\hat{\beta}^*_{(p,jk),\mathrm{MLE}}$ and $V_{p,jk}$ come from a maximum likelihood estimation of the function-on-function model.

Step 2: For each y-space coefficient indexed by (jk), we next sample the elements $\sigma^2_{U(j,k)}$ and $\sigma^2_{E(j,k)}$ of Σ^*_U and Σ^*_E respectively. For this we use a random-walk Metropolis-Hastings step with objective function

$$\begin{aligned}&f(\sigma^2_{U(j,k)}, \sigma_{E(j,k)}|\mathbf{y}^*_{(j,k)}, \beta^*_{(j,k)}) \propto \\ &|\Sigma_{jk}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}^*_{(j,k)} - \mathbf{X}^*\beta^*_{(j,k)})'\Sigma^{-1}_{jk}(\mathbf{y}^*_{(j,k)} - \mathbf{X}^*\beta^*_{(j,k)})\right\}f(\sigma^2_{U(j,k)}, \sigma_{E(j,k)}).\end{aligned}$$

where Σ_{jk} is the marginal variance of $\mathbf{y}^*_{(j,k)}$. For the proposal distribution, we use an independent Gaussian truncated at zero and centered at the previous values.

Step 3: Random effects $U^*_{(j,k)}$ for each (j,k) are sampled from their full conditional which is a Gaussian distribution with mean $\{\Psi^{-1}_{jk} + 1/\sigma^2_{U(j,k)}\}^{-1}\Psi^{-1}_{jk}\hat{\mathbf{u}}_{NS,jk}$ and variance $\{\Psi^{-1} + 1/\sigma^2_{U(j,k)}\}^{-1}$, where $\Psi_{jk} = \{\mathbf{Z}'(1/\sigma^2_{E(j,k)})\mathbf{Z}\}^{-1}$ and

$$\hat{\mathbf{u}}_{NS,jk} = \left\{\mathbf{Z}'(1/\sigma^2_{E(j,k)})\mathbf{Z}\right\}^{-1}\mathbf{Z}'\left(1/\sigma^2_{E(j,k)}\right)\left(\mathbf{y}^*_{(j,k)} - \mathbf{X}^*\beta^*_{(j,k)}\right)$$

Step 4: Finally, we update τ_{pj} and π_{pj} separately from $f(\tau_{pj}|\gamma_{(p,jk)}, \beta^*_{(j,k)}, a_\tau, b_\tau)$ and $f(\pi_{pj}|\gamma_{(p,jk)}, a_\pi, b_\pi)$. The form of these conditionals are an inverse-gamma and beta respectively. Hyperparameters a_τ, b_τ,a_π, and b_π are calculated using Empirical Bayes estimates which are described in Morris and Carroll (2006).

Notes: (1) Our approach can easily accommodate other shrinkage priors that might make sense for other basis functions, including Gaussians for spline bases, or other types of sparsity priors including Bayesian Lasso, Normal-Gamma, or Horseshoe Priors, which may have better sparsity and shrinkage properties under some settings; (2) These priors have connections to penalized likelihood methods, and their application in the basis space can induce smoothing or regularization across the coefficient surface $\boldsymbol{\beta}(v,t)$ in the data space; (3) The double-indexing inherent to multi-resolution bases like wavelets can be used for other bases, defining J clusters of basis coefficients $j = 1, \ldots, J$ containing K_j coefficients each, in order to allow clusters of coefficients to share common regularization parameters.

Additionally, code is available for running our model on one simulated dataset examined in the manuscript at `http://odin.mdacc.tmc.edu/~jmorris/FonF.zip`. The code relies on MATLAB functions of our design as well as the WFMM executable file developed by Morris and Carroll for the accompanying paper, Morris and Carroll (2006). Wavelet decompositions were done using the MATLAB function wavedec while principal component analysis was performed using the MATLAB function princomp.

Web Appendix C: Additional Simulation Details

Equations for the four simulation scenarios discussed in the manuscript are found below:

$$\frac{7}{500}\frac{1}{\sqrt{(2\pi)0.003}}\exp\left[-\frac{1}{(2)(0.003)}\left(\frac{t}{225}-\frac{v}{225}\right)^2\right] \text{ (ridge),} \tag{2}$$

$$\frac{437}{10000}\frac{1}{\sqrt{(2\pi)(0.03)}}\exp\left[-\frac{1}{(2)(0.03)}\left(\frac{t}{225}-\frac{v}{225}-0.5\right)^2\right] \text{ (lagged),} \tag{3}$$

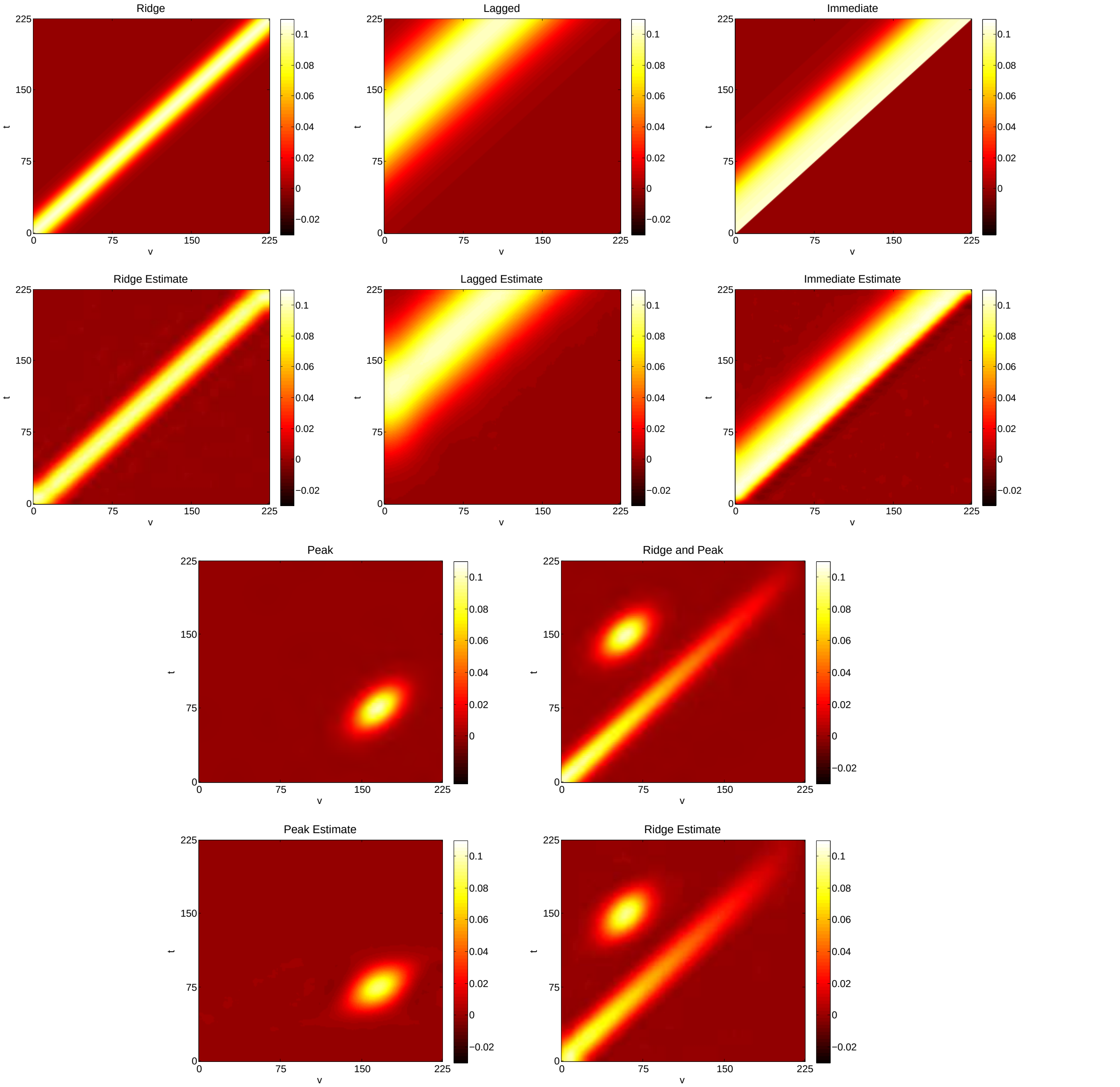
$$\frac{1029}{10000}\left[1+\frac{1}{1+\exp\left(\frac{0.25-\frac{t}{225}+\frac{v}{225}}{0.05}\right)}\right] \text{ (immediate),} \tag{4}$$

$$\frac{1225}{10}(2\pi)^{-1}|\Sigma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\left[\begin{array}{c} v \\ t \end{array}\right]-\left[\begin{array}{c} 150 \\ 60 \end{array}\right]\right)'\Sigma^{-1}\left(\left[\begin{array}{c} v \\ t \end{array}\right]-\left[\begin{array}{c} 150 \\ 60 \end{array}\right]\right)\right] \text{ (peak)} \tag{5}$$

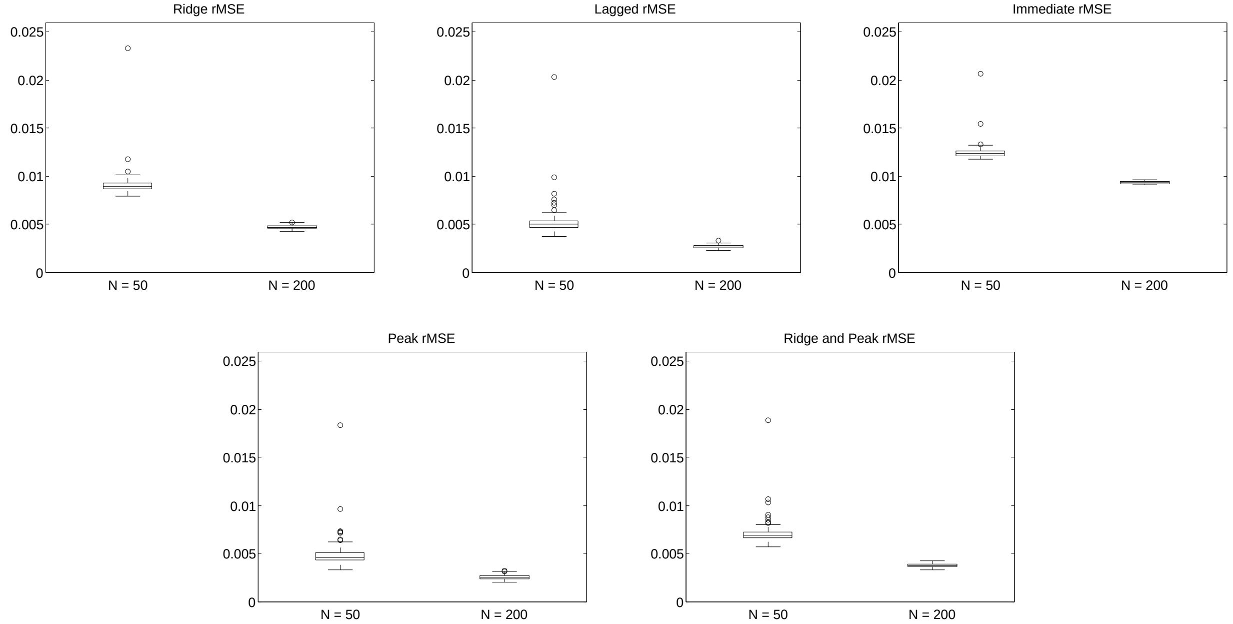
where

$$\Sigma = \left[\begin{array}{cc} 15 & 0.5(15)(15) \\ 0.5(15)(15) & 15 \end{array}\right].$$

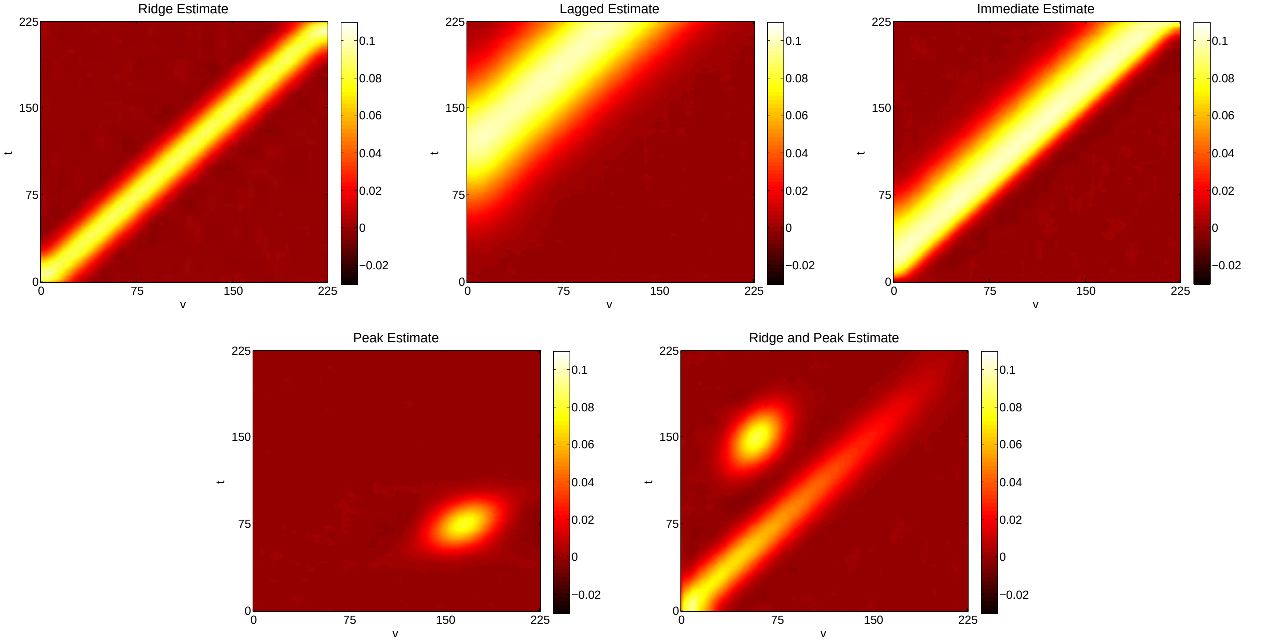
Graphical representations of each of these can be found both in the manuscript as well as in Web Figure 1.



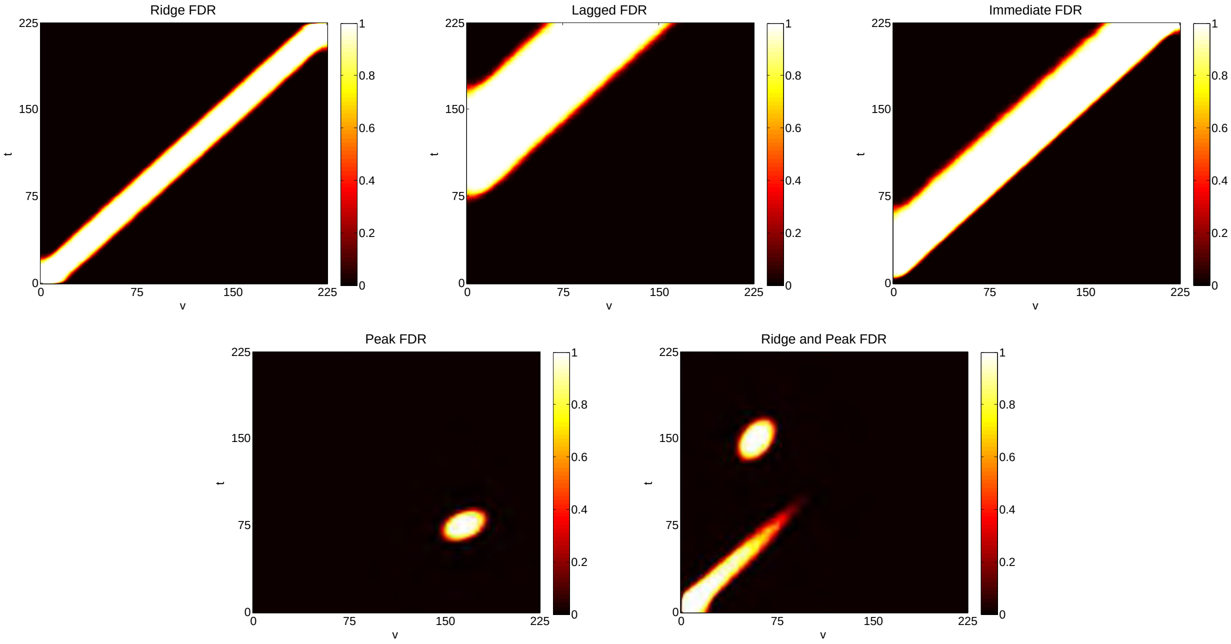
Web Figure 1: Heat maps of the true surfaces for simulation study are above heat maps of the estimated surfaces for each simulated scenario based on a sample size of $n = 100$ with two measure per subject, $C_i = 2\ \forall\ i$, for a total of $N = 200$ observations. Each surface is the average of the posterior estimate for the true surface based on 200 simulated datasets.



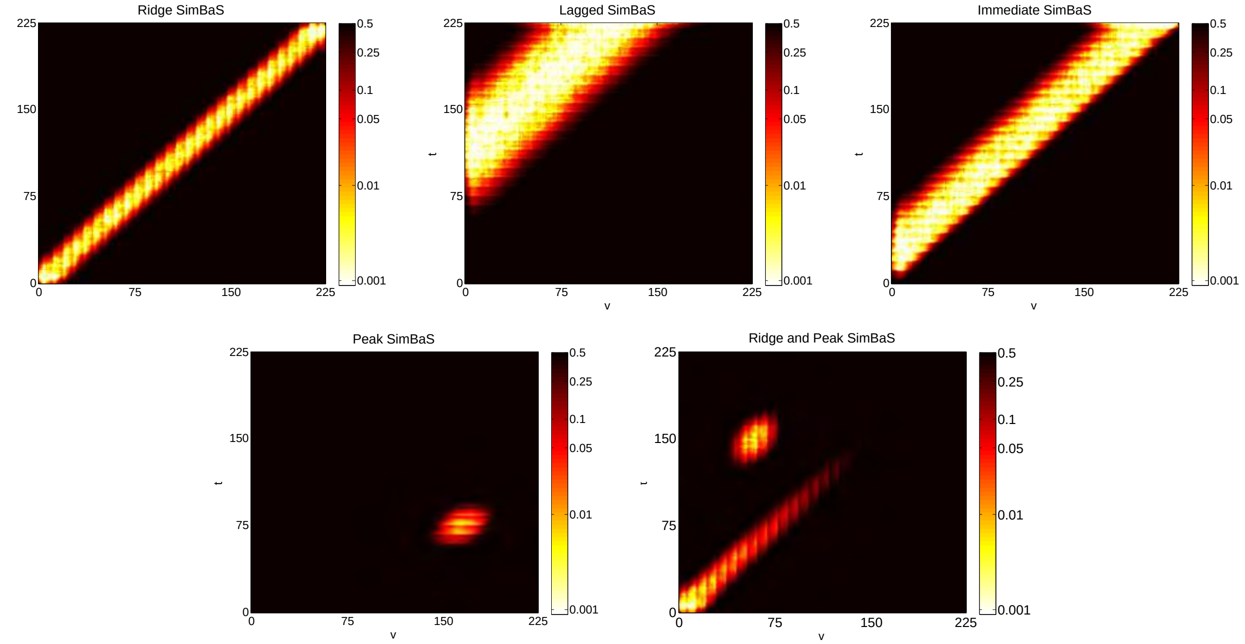
Web Figure 2: Boxplots comparing the root-Mean Square Error amongst varying sample sizes by simulation scenario. Not surprisingly, as sample size increases, rMSE decreases. Note that the graphics are listed by total number of observations with $N = 50$ on the left and $N = 200$ on the right.



Web Figure 3: Each figure contains a heat map displaying the average estimated surface from the simulation for the smallest sample size $N = 50$, $n = 25$ taken over all 200 simulated datasets.

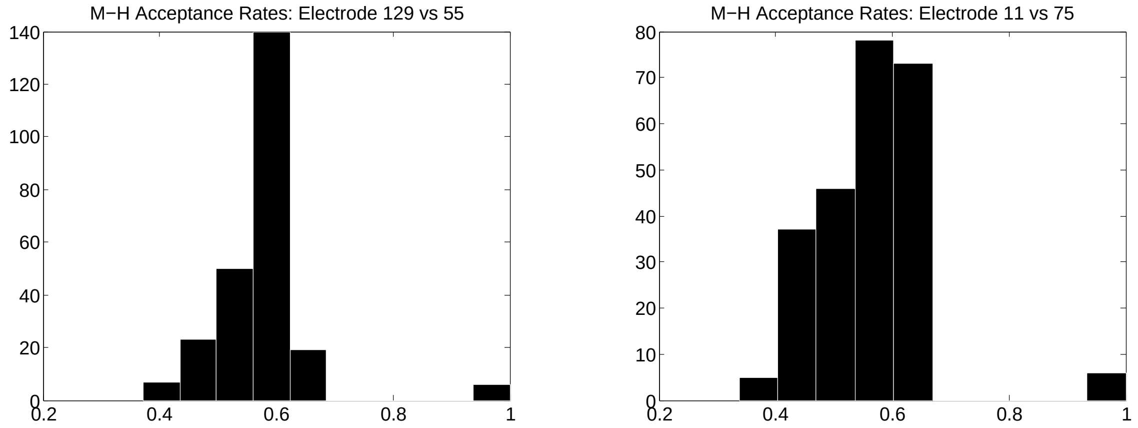


Web Figure 4: Each figure contains a heat map displaying the FDR acceptance region averaged over 200 simulated data sets. The white regions indicate locations (v, t) flagged in every or almost every data set. Black locations (v, t) were not flagged in any or almost any data sets. At the edge of each flagged region, locations that were flagged only occasionally can be seen in varying shades of red and orange. These figures were based on the smallest sample size $N = 50$, $n = 25$.

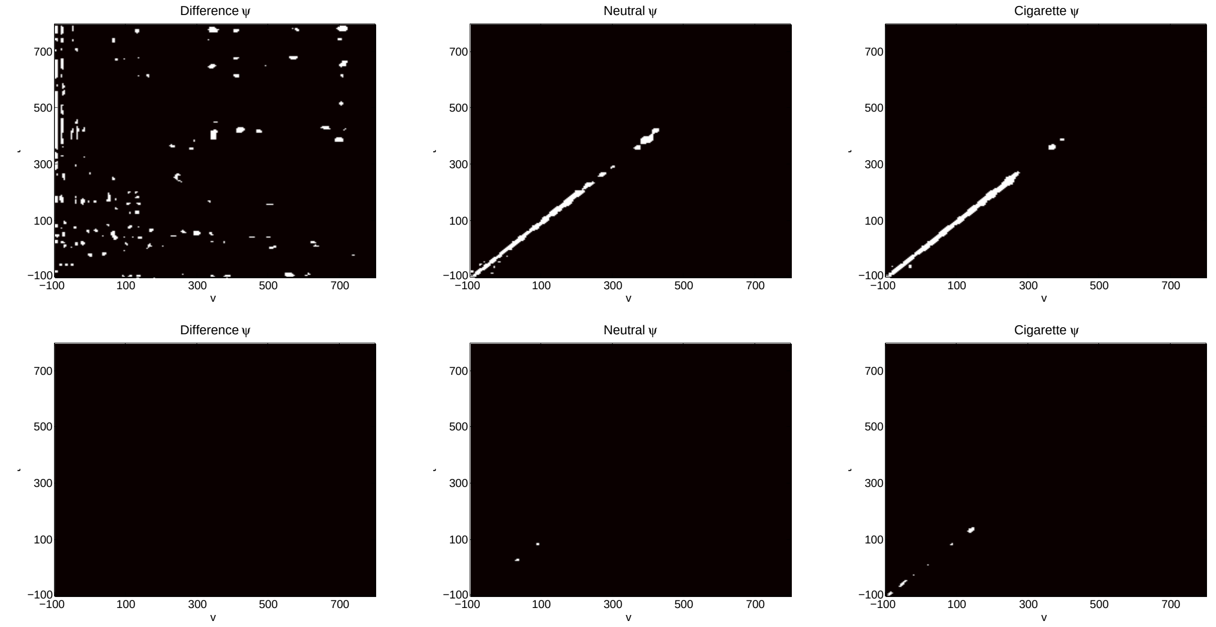


Web Figure 5: Each figure contains a heat map displaying the SimBa Scores. All plots are on the log-scale, however for interpretability, the color scale has been exponentiated. For consistency of interpretation, the color scale was reversed so that the darker the red the more significant the coefficient and the darker the blue, the less significant. The scale is set to exhibit the variation in the SimBa Scores over the region of elevated significance. As such, the max of the scale does not accurately reflect the SimBa Scores where the true surface lacks association. Each plot represents the average SimBaS for each coefficient over the 200 simulated data sets. These figures were based on the smallest sample size $N = 50$, $n = 25$.

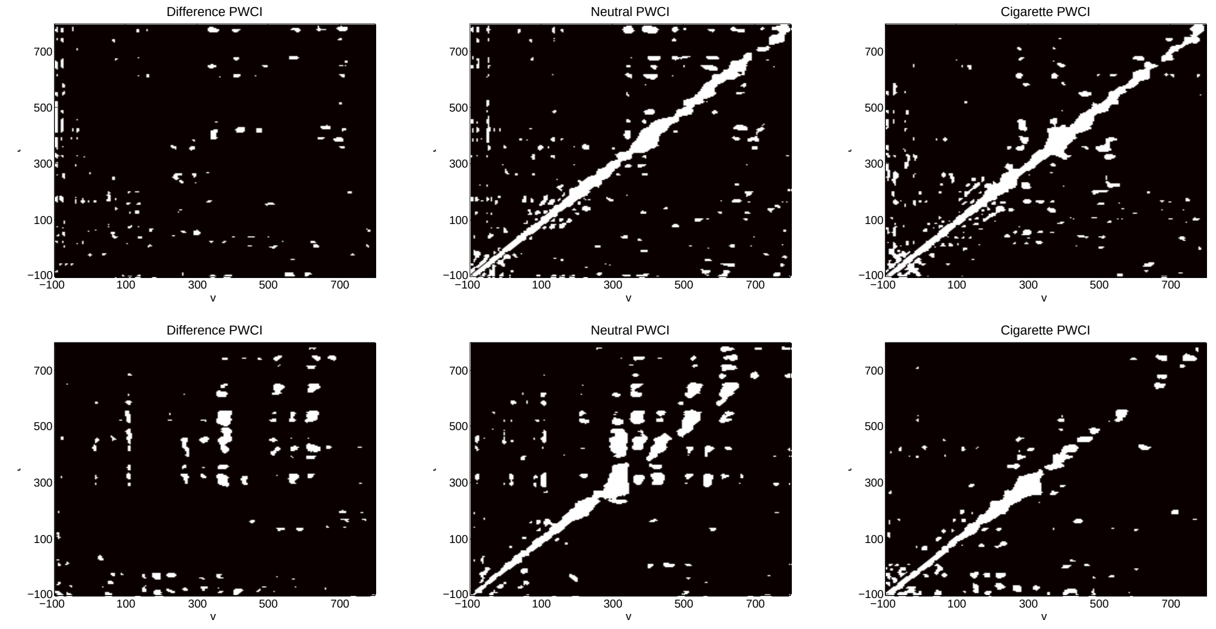
Web Appendix D: Additional Application Results



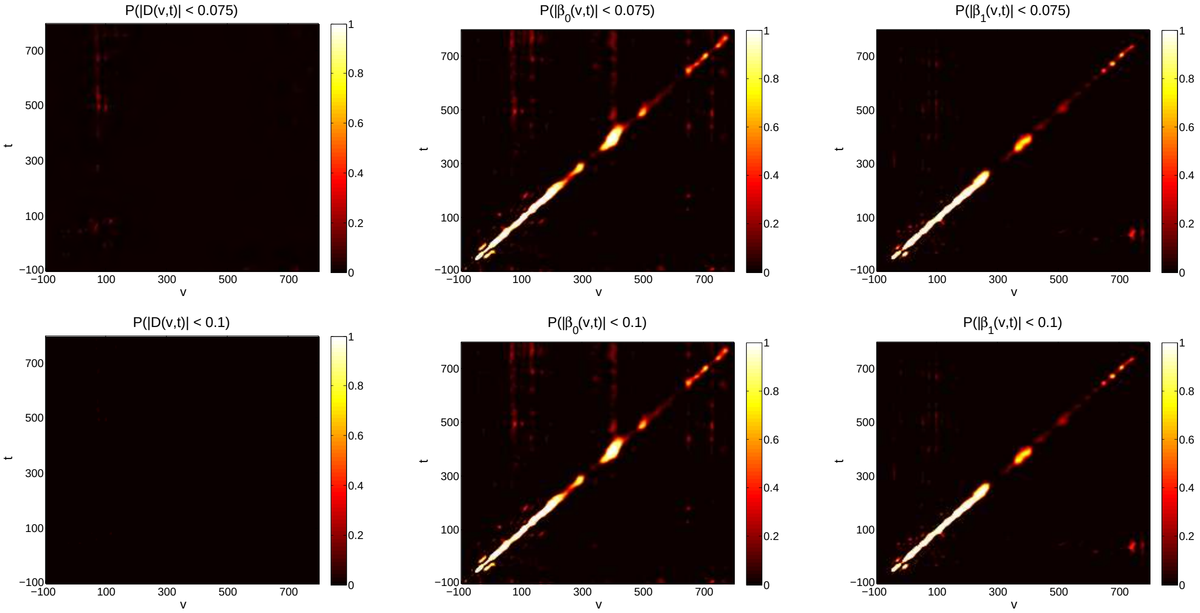
Web Figure 6: Histograms of acceptance rates for the Metropolis-Hastings step of the sampler which samples the variance components of the model. The histogram on left is for the model regressing Electrode 129 on to Electrode 55. The histogram on the right is for the model regressing Electrode 11 on to Electrode 75.



Web Figure 7: Heat maps of the set of flagged locations (v,t), ψ, from the BFDR. Significant locations appear in white, non-significant locations are in black. For the difference surface, $\alpha = 0.05$ while the image-specific surfaces use $\alpha = 0.025$. The δ-intensity change is 0.05 for all surfaces. The top row contains results from the model using electrodes 129 and 55. The bottom row contains results from the model using electrodes 75 and 11.



Web Figure 8: Heat maps of the set of flagged locations (v,t) using PWCI. Locations in white were flagged as significant by the procedure while locations in black are not-significant. The top row contains results from the model using electrodes 129 and 55. The bottom row contains results from the model using electrodes 75 and 11.



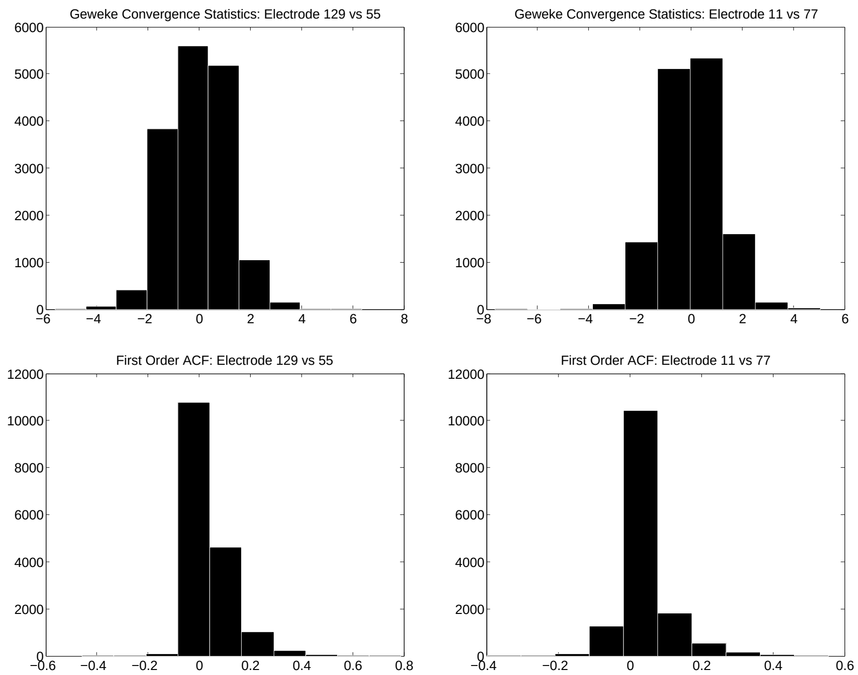
Web Figure 9: Heat maps of posterior probability that a location (v, t) is larger in absolute value than the δ intensity change. The top row shows a $\delta = 0.075$ while the bottom row shows a $\delta = 0.1$. In the manuscript, we show results for $\delta = 0.05$. These heat maps are for the model regressing electrode 129 on to 55.

Web Appendix E: Convergence Diagnostics

We assess convergence of surface coefficients in the wavelet-space in two ways. First, we examine the Geweke Convergence diagnostic as described in Geweke (1992) which generates standard Gaussian Z-scores to test the equality of the first 10% and last 50% of the chain. Second, we determine the first order autocorrelation coefficient for each chain. Histograms for Geweke Z-scores and first order autocorrelation coefficients can be found in Web Figure 10. Diagnostics were performed on both Application models. In addition to graphical displays, we generated numerical summaries which can be found in Web Table 1. The mean values of the Geweke Z-scores are near zero for both models (0.001 and 0.014 respectively). Further, the middle 95% of Geweke Z-scores for both models fall in the intervals $(-2.118, 2.151)$ and $(-2.049, 2.11)$ respectively. This suggests that most Z-scores are within roughly 2 standard deviations of zero. Additionally, the mean first order autocorrelation coefficients are also near zero, 0.042 and 0.031 respectively. The distribution of first order coefficients is tightly packed about zero with the middle 95% of coefficients for both models falling in the intervals $(-0.046, 0.252)$ and $(-0.059, 0.233)$ respectively. Graphically, we see in Web Figure 10 the distributions of Geweke Z-scores and first order autocorrelation coefficients are roughly symmetric about zero. This, combined with our numerical summaries, suggests that convergence of coefficients has, on the whole, been achieved.

Web Table 1: Summary statistics for Geweke convergence diagnostic and first order autocorrelation coefficient. Statistics consist of the mean as well as 2.5 percentile and 97.5 percentile. Diagnostics are broken down by application model.

	Geweke			1st Order ACF		
Model	Mean	2.5%-ile	97.5%-ile	Mean	2.5%-ile	97.5%-ile
Electrode 129 vs 55	0.001	−2.118	2.151	0.042	−0.046	0.252
Electrode 11 vs 77	0.014	−2.049	2.114	0.031	−0.059	0.233



Web Figure 10: Histograms of convergence diagnostics. The top row contains the Geweke Z-scores calculated as in Geweke (1992). The bottom row contains first order autocorrelation coefficients. The first column corresponds to diagnostics for the model assessing Electrodes 129 and 55 while the second column corresponds to diagnostics for the model examining Electrodes 11 and 77.

Web Appendix F: Comparison to Scheipl, Staicu, and Greven (2014)

Here we briefly compare the approach used by Scheipl, Staicu, and Greven (2014) for function-on-function regression to the approach we develop in the manuscript. A complete comparison is not entirely appropriate given the constraints of their modeling procedure. For example, their model does not allow for non-iid errors. Additionally, we were unable to use their code to implement functional random effects, only scalar random effects. Nevertheless, we used their approach to fit simplified versions of our models for both application and simulation–only one simulated dataset is used for each scenario so for bet comparison, refer to Web Figure 3 which shows the results of our model applied to the same simulated datasets. We fit the model using the defaults provided in the code from Scheipl, Staicu, and Greven (2014).

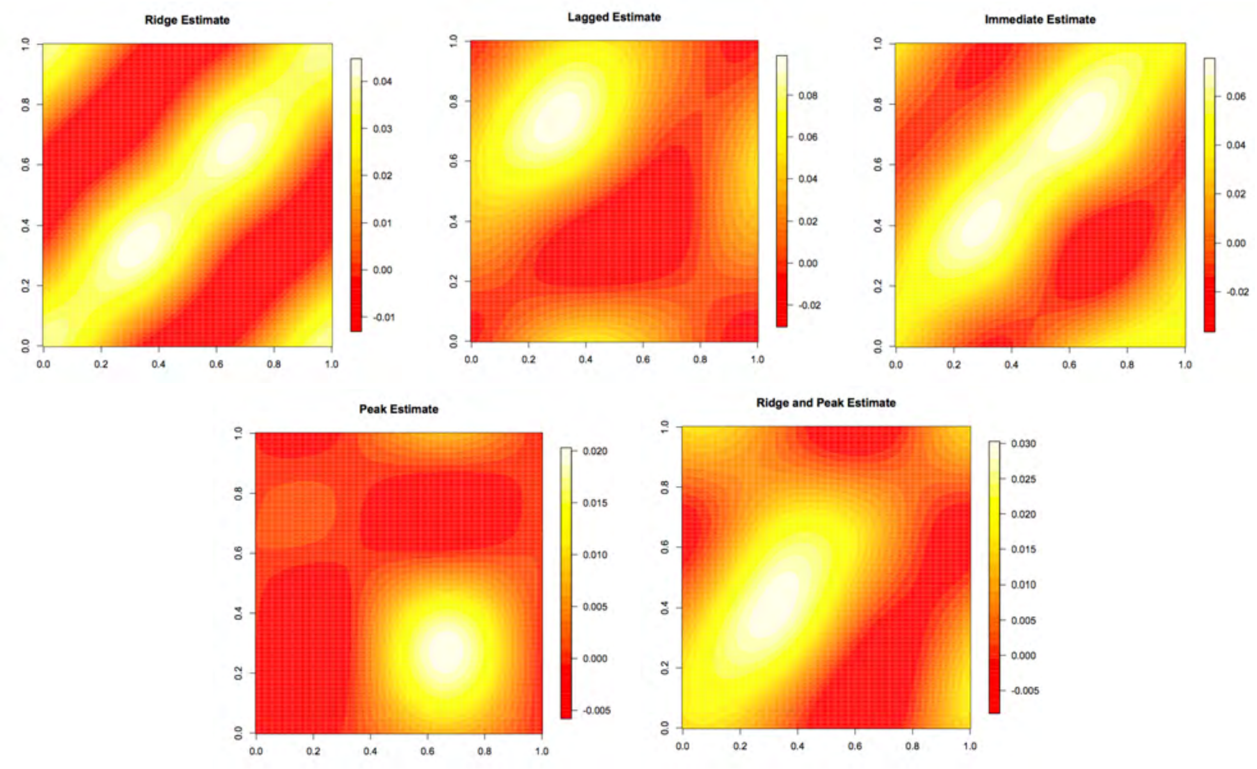
To compare in simulation, we fit the model

$$y_{ic}(t) = \alpha(t) + \int_{v \in \mathcal{V}} x_{ic}(v)\beta(v,t)dv + U_i + E_{ict}.$$

And to compare in application, we fit a simplified version of equation (9):

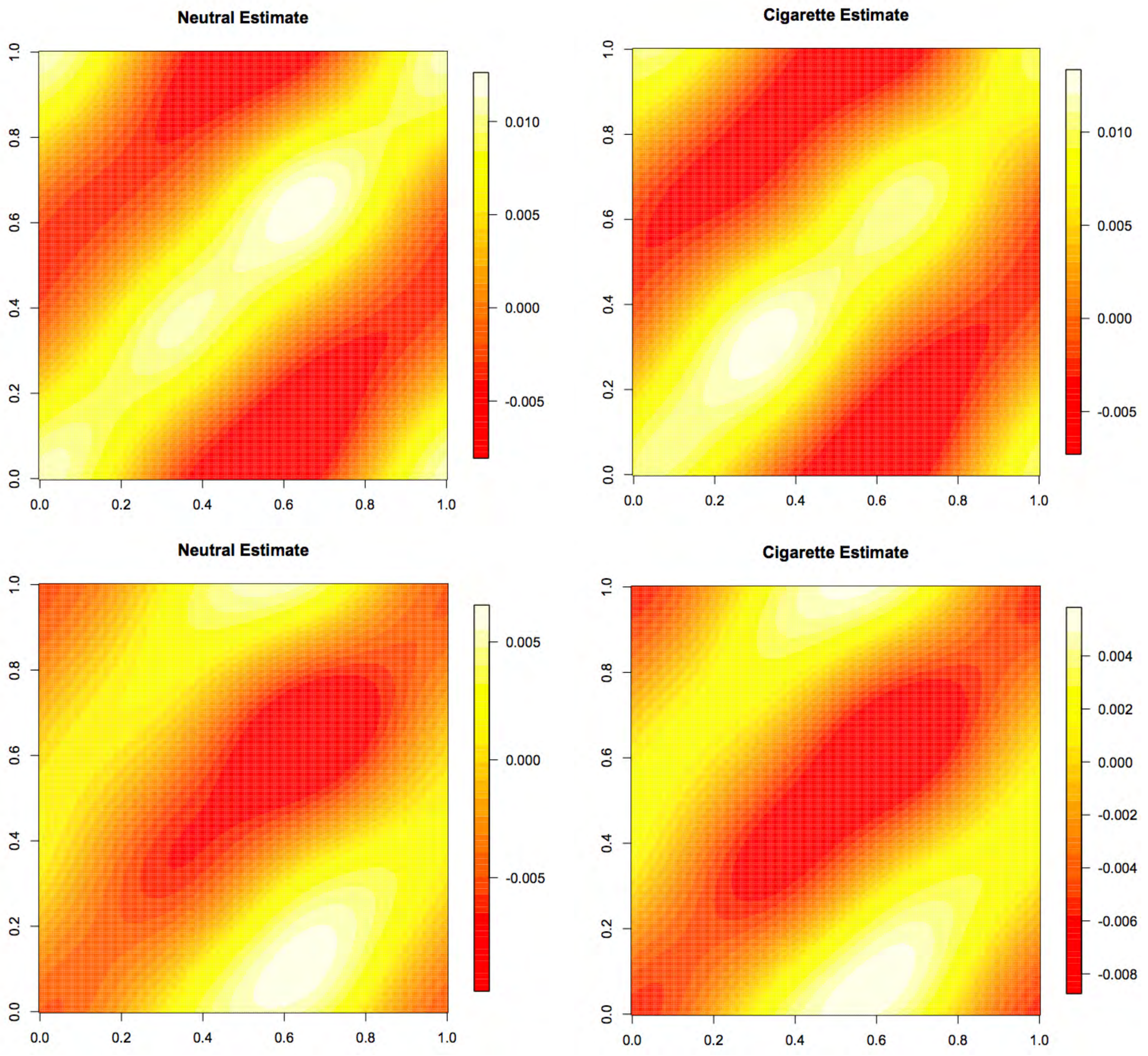
$$\begin{aligned} y_{icg}(t) = 1(g=0) & \left[\alpha_0(t) + \int_{v \in \mathcal{V}} x_{ic0}(v)\beta_0(v,t)dv\right] \\ &+ 1(g=1)\left[\alpha_1(t) + \int_{v \in \mathcal{V}} x_{ic1}(v)\beta_1(v,t)dv\right] + U_i + E_{ict}. \end{aligned}$$

Where in both preceding models U_i is a scalar random intercept, $U_i \sim \mathcal{N}(0, \sigma_U)$, and E_{ict} is assumed iid over measurements of t, $E_{ict} \sim \mathcal{N}(0, \sigma_E)$. Heat maps of estimated surfaces for simulated datasets can be found below in Web Figure 11. Using the true surfaces found in Web Figure 1 as reference, we see that for each estimated surface below, the effect size is consistently under estimated except for perhaps the Lagged surface. The estimated surfaces struggle to capture the regions of elevated association, tending to over estimate their breadth. Further, detailed features tend to get lost. For instance, the "cliff" of the Immediate effect is not detected and the ridge and peak are indistinguishable for that scenario. Also, their approach has issues estimating regions that are truly zero. This is particularly evident around the edges of each surface where larger positive effects are estimated despite being truly zero.



Web Figure 11: Each figure contains a heat map displaying a single estimated surface from the simulation for the smallest sample size $N = 50$, $n = 25$. Estimation is performed using the method examined by Scheipl, Staicu, and Greven (2014).

Similar issues arise when examining estimates from the model for the application data. Heat maps of the cigarette and neutral surfaces are in Web Figure 12 (note that no difference surface is included as their model does not facilitate inference on that surface). For both models, regressing first electrode 129 on to electrode 55 and then regressing electrode 11 on to electrode 75, the detail of the association is lost as the estimated surfaces over estimate the breadth of the region of elevated association along the ridge. Further, away from the ridge, estimates head in the opposite direction of the the association before shifting back at the corners. These edge effects were seen on the simulated data and appear to exist also in application.



Web Figure 12: Each figure contains a heat map displaying estimated surfaces from the application using the approach examined by Scheipl, Staicu, and Greven (2014). The top row corresponds to the model regressing electrode 129 on to electrode 55. The bottom corresponds to the model regressing electrode 11 on to electrode 75.

Web Appendix G: Adjustment for Autocorrelation in posterior variance estimates

MCMC generates correlated observations from the posterior distribution. Given a serial sample of correlated posterior samples $X_i, i = 1, ..., N$, with $\text{corr}(X_i, X_{i+k}) = \rho_k$, then from Anderson (1971), p.448 (eqn 51), we know that

$$E(s^2) = \sigma^2 A(N, \rho_k) \tag{6}$$

$$A(N, \rho_k) = 1 - \frac{2}{N-1} \sum_{k=1}^{N-1} (1 - k/N) * \rho_k. \tag{7}$$

where $A(N, \rho_k)$ is the bias factor, which is bounded above by 1. Thus, the sample mean of autocorrelated posterior samples is biased downwards for estimating the posterior variance, which affects any inference based on these posterior variances, e.g. the calculation of joint credible bands by the method of Ruppert, et al. (2003) or the SimBaS in this paper.

If the data are AR1(ρ), then $\rho_k = \rho^k$. Based on this function, we see that the bias in the sample variance increases for large ρ and for small N. Following is a table of the bias factor $A(N, \rho)$ for AR1 samples:

						ρ					
N	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	0.999
100	0.993	0.984	0.971	0.951	0.922	0.877	0.800	0.636	0.328	*	*
1,000	0.999	0.998	0.997	0.995	0.992	0.987	0.980	0.962	0.924	0.746	0.011
2,000	1.000	0.999	0.999	0.998	0.996	0.994	0.990	0.981	0.962	0.868	0.301
5,000	1.000	1.000	0.999	0.999	0.998	0.998	0.996	0.992	0.985	0.946	0.626
10,000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.996	0.992	0.973	0.793
100,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.978
1,000,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998

Thus, we see that if we have 2000 posterior samples, the bias becomes non-negligible roughly when $\rho > 0.9$, when we have 5000 posterior samples, $\rho > 0.95$. Thus, if the autocorrelation is not too big or we run enough samples, this bias is neglible. Otherwise, however, we need to do an adjustment for this bias or our intervals will be biased. Note that thinning is commonly used to reduce the autocorrelation. The formula (7) can be used to do the adjustment when necessary.

References

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley.

Aston, J. A., Choiu, J., and Evans, J. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society, Series C* **59**, 297-317.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (ed JM Bernado, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

Joliffe I. T. (1982) A note on the use of principal components in regression. *Journal of the Royal Statistical Society, Series C* **31(3),** 300-303.

Malloy, E. J., Morris, J. S., Adar, S. D., Suh, H., Gold, D. R., and Coull, B. A. (2010). Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. *Biostatistics* **11,** 432–452.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68,** 179–199.

Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P., and Gutstein, H. (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *The Annals of Applied Statistics* **5,** 894–923.

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

Yang, W. H., Wikle, C. K., Holan, S. H., and Wildhaber, M. L. (2013). Ecological prediction with nonlinear multivariate time-frequency functional data models. *Journal of Agricultural, Biological, and Environmental Statistics* **18(3),** 450-474.

Zhu H., Brown, P. J., and Morris, J. S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association* **106,** 1167–1179.