

# Multivariate functional response regression, with application to fluorescence spectroscopy in a cervical pre-cancer study <sup>☆</sup>

Hongxiao Zhu<sup>a,\*</sup>, Jeffrey S. Morris<sup>b</sup>, Fengrong Wei<sup>c</sup>, Dennis D. Cox<sup>d</sup>

<sup>a</sup>*Department of Statistics, Virginia Tech, Blacksburg, VA 24061*

<sup>b</sup>*The University of Texas MD Anderson Cancer Center, Houston, TX 77230*

<sup>c</sup>*Department of Mathematics, University of West Georgia, Carrollton, GA 30118*

<sup>d</sup>*Department of Statistics, Rice University, Houston, TX 77005*

---

## Abstract

Many scientific studies measure different types of high-dimensional signals or images from the same subject, producing multivariate functional data. These functional measurements carry different types of information about the scientific process, and a joint analysis that integrates information across them may provide new insights into the underlying mechanism for the phenomenon under study. Motivated by fluorescence spectroscopy data in a cervical pre-cancer study, a multivariate functional response regression model is proposed, which treats multivariate functional observations as responses and a common set of covariates as predictors. This novel modeling framework simultaneously accounts for correlations between functional variables and potential multi-level structures in data that are induced by experimental design. The model is fitted by performing a two-stage linear transformation—a basis expansion to each functional variable followed by principal component analysis for the concatenated basis coefficients. This transformation effectively reduces the intra- and inter-function correlations and facilitates fast and convenient calculation. A fully Bayesian

---

<sup>☆</sup>The code that demonstrates how to implement the proposed approach is available as an annex in the electronic version of this manuscript.

\*Correspondence to: Department of Statistics (MC0439), 250 Drillfield Drive, Virginia Tech, Blacksburg, VA 24061 USA. Telephone: 540-231-0400. Fax: 540-231-3863.

*Email addresses:* [hongxiao@vt.edu](mailto:hongxiao@vt.edu) (Hongxiao Zhu), [jefmorris@mdanderson.org](mailto:jefmorris@mdanderson.org) (Jeffrey S. Morris), [fwei@westga.edu](mailto:fwei@westga.edu) (Fengrong Wei), [dcox@rice.edu](mailto:dcox@rice.edu) (Dennis D. Cox)

approach is adopted to sample the model parameters in the transformed space, and posterior inference is performed after inverse-transforming the regression coefficients back to the original data domain. The proposed approach produces functional tests that flag local regions on the functional effects, while controlling the overall experiment-wise error rate or false discovery rate. It also enables functional discriminant analysis through posterior predictive calculation. Analysis of the fluorescence spectroscopy data reveals local regions with differential expressions across the pre-cancer and normal samples. These regions may serve as biomarkers for prognosis and disease assessment.

*Keywords:* Bayesian methods, functional data analysis, mixed models, multivariate functional regression, fluorescence spectroscopy, wavelets, principal component analysis

---

## 1. Introduction

Rapidly expanding modern technology enables the automatic collection of high-dimensional data in functional form. Examples include signals, images, and many emerging high-throughput digital measurements. The increased prevalence of such data promotes the development of *functional data analysis*. While 5 considerable efforts have been made to analyze functional data (Ramsay and Silverman, 1997), existing approaches often focus on modeling functional variables that have a common support and the same interpretation. Many applications, however, involve simultaneously collecting different functional variables. For 10 example, in economics, a firm’s *market value* and *trading volume* constitute two different financial profiles; in environmental studies, the air pollution level is often measured with the temperature and air pressure for a given location over a period of time; in neuroscience experiments, different types of brain images, such as electroencephalograms (EEGs), magnetic resonance images (MRIs), and dif- 15 fusion tensor images (DTIs), are often collected for the same subject, resulting in “multi-modal” imaging data (Uludag and Roebroek, 2014). In these situations, the multiple functional variables have different interpretations. They character-

ize different aspects of the same subject. Joint analysis that integrates these variables, also called *data fusion* (Castanedo, 2013), helps borrow information  
20 from different resources, which may provide new insights into the underlying mechanism for the phenomenon under study (e.g., how temperature and air pressure interact to influence the concentration of pollutants).

Despite the promise of borrowing strength across different resources, the joint analysis of multiple functional variables has not received much attention  
25 in the literature. There is extensive work on modeling functional data with hierarchical structures induced by crossed or nested experimental designs (Morris and Carroll, 2006; Greven et al., 2010; Zhu et al., 2011; Goldsmith et al., 2012; Scheipl et al., 2014; Goldsmith et al., 2015; Backenroth et al., 2016), spatial/temporal correlations (Giraldo et al., 2010; Delicado et al., 2010; Hörmann  
30 and Kokoszka, 2010; Aue et al., 2015), or both (Baladandayuthapani et al., 2008; Zhou et al., 2010; Staicu et al., 2010). However, these approaches are all based on the assumption that functional data are defined on a common domain and have the same or a similar interpretation. In the recent work of Zhu et al. (2016a) and Qiao et al. (2015), multiple functional variables are used in the  
35 functional graphical model setup, but these methods mainly focus on characterizing the conditional independence between functions using a graph. To our knowledge, there has been little effort to associate multivariate functional data with other covariates.

In the scalar-on-function regression in which scalar responses are regressed  
40 on functional predictors (Morris, 2015), multiple functional variables can be easily incorporated as predictors, regardless of whether they have the same interpretation (Zhu and Cox, 2009; Zhu et al., 2010; Fan et al., 2015). However, since the scalar-on-function regression treats functional variables as covariates, it models the conditional expectation of a scalar given the functional variables,  
45 and thus does not directly characterize the inter-function correlations or complicated multi-level structures between the functional variables. In this paper, we consider the joint analysis of multiple functional variables. We propose a multivariate functional response regression model that jointly regresses multivariate

functional data on a common set of predictors while simultaneously account-  
50 ing for potential multi-level structures in the data. In particular, we propose a  
flexible multivariate functional mixed model (MFMM) to incorporate intra- and  
inter-function correlations as well as correlation induced by hierarchical designs.

In the proposed MFMM, directly parameterizing intra- and inter-function  
correlations is challenging, because we do not have prior knowledge of the cor-  
55 relation structure, and existing nonparametric approaches, such as the methods  
of Yao (2007) and Yang et al. (2016), can only be used to estimate the intra-  
function correlation. To address this issue, we propose a convenient two-stage  
transformation approach through *lossless* or *near-lossless* basis representation.  
First, we represent each functional variable using a basis expansion. This trans-  
60 forms each functional variable to the dual space of basis coefficients. We then  
take the basis coefficients of all functional variables from one multivariate func-  
tional observation and concatenate them to form a long vector. Finally, we apply  
principal component analysis (PCA) by treating the entries in the concatenated  
vectors as variables and all observations as independent samples, which further  
65 reduces the correlations between the pooled basis coefficients. Assuming that  
the functional variables are Gaussian processes, the resulting principal compo-  
nent (PC) scores of the concatenated basis-coefficient sequences are approxi-  
mately independent of each other. This enables us to fit separate mixed models  
to each PC independently in the transformed space, using the PC scores from all  
70 observations as data. This results in fast and easily parallelizable computation.  
Since this two-stage transformation preserves the linear mixed model format and  
is invertible, Bayesian inferences can be performed in the original data space  
after inverse-transforming the posterior samples of the regression coefficients.  
This preserves the interpretability of the inferential results. Acknowledging its  
75 benefits, the basis-expansion-based transformation has been adopted in many  
functional data regressions; see Morris (2015) for a thorough review. Two-stage  
functional representations have also been used by James (2002), Reiss and Og-  
den (2007), and Reiss and Ogden (2010) in the context of scalar-on-function  
regression.

80 Compared to the existing methods, the proposed MFMM approach has the following major advantages. (1) It jointly models multiple functional variables, including those with different domains or different interpretations. (2) It accounts for intricate multi-level structures in the data. (3) The two-stage transformation avoids the hassle of directly parameterizing the intra- and inter-  
85 function correlations, leading to fast and easily parallelizable computation. (4) It yields interpretable Bayesian posterior inference in the original data space. (5) The posterior inference produces functional tests that adjust for the overall experiment-wise error rate (EER) or false discovery rate (FDR). (6) The Bayesian approach enables the classification/discrimination of new functional  
90 observations through posterior predictive calculation. Our analysis of fluorescence spectroscopy data reveals local regions with differential expressions across the pre-cancer and normal tissue samples. These regions may serve as biomarkers for prognosis and disease assessment.

The outline for the rest of this paper is as follows. In Section 2, we describe  
95 the motivating example—the fluorescence spectroscopy data in a cervical pre-cancer study. We describe the proposed MFMM in Section 3, demonstrate its performance using simulation studies in Section 4, and show fluorescence spectroscopy data analysis in Section 5. In Section 6, we describe a sensitivity study to evaluate whether our results are sensitive to different model choices  
100 and parameter setups. We provide a final discussion in Section 7. Details on the computational algorithm, parameter settings, and extra analytic results are available in the supplementary materials.

## 2. The fluorescence spectroscopy data in a cervical pre-cancer study

This work is motivated by the fluorescence spectroscopy data arising from  
105 a cervical pre-cancer study. Fluorescence spectroscopy is an optical technique that captures the spectra of fluorescent lights emitted by a given material (e.g., a tissue sample). It provides a non-invasive, low-cost alternative to existing approaches for the diagnosis and assessment of early stage cervical cancer (Ra-

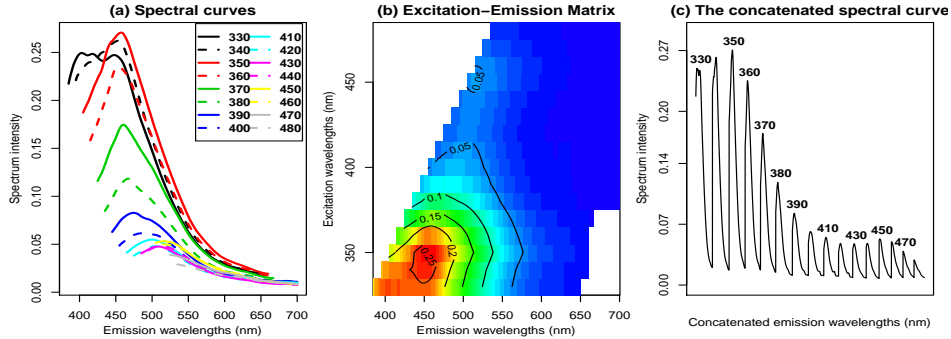


Figure 1: Plots of one fluorescence spectroscopy measurement. (a) Fluorescence spectral curves at different excitation wavelengths in nanometers (nms). (b) Image plot of the excitation-emission matrix (EEM). (c) Plot of a concatenated spectral curve, in which we have labeled the excitation wavelengths of some spectral curves at their peak intensities.

manujam et al., 1996). The data studied in this paper are collected from a  
 110 clinical study that used multiple fluorescence spectra to detect cervical abnormalities.

Each measurement consists of multiple spectral curves measured on the same cervical tissue site. The measurement was taken in the following way: an excitation light at a certain fixed excitation wavelength is produced to illuminate  
 115 the cervical tissue. During illumination, the excitation light is absorbed by various endogenous fluorescent molecules in the tissue, resulting in the emission of fluorescent light. The emitted fluorescent light is then captured by an optical detector that produces a spectrum as a smooth curve. The excitation light can be varied at a sequence of excitation wavelengths, producing multiple  
 120 spectral curves for each measurement. In Figure 1 (a), we demonstrate one measurement. It contains 16 spectral curves measured at excitation wavelengths ranging from 330 nm to 480 nm, with increments of 10 nm, where each spectral curve contains fluorescence intensities recorded on a range of emission wavelengths between 385 nm and 700 nm. Here, the fluorescence intensity is  
 125 in arbitrary units. It has been normalized through dividing by the excitation light energy measurement for each excitation wavelength (Lee et al., 2005) and thus is comparable across different excitation wavelengths. We often use a color

band to represent the intensities of each spectral curve and stack all 16 spectra according to their excitation wavelengths, as shown in Figure 1 (b). We call such data the excitation-emission matrices (EEMs). To display multiple EEMs in one plot, we often concatenate spectral curves to form a one-dimensional vector. A plot is shown in Figure 1 (c), in which the excitation wavelengths of the spectra are labeled at their peak intensities.

The data presented in this paper were collected using the same instrument (called FastEEM3) in the same clinic (British Columbia Cancer Agency, Vancouver, CA). They were preprocessed following a six-step procedure, the details of which were described by Marín et al. (2006). The processed data contain 534 EEM measurements, among which 143 are from pre-cancer samples, and 391 are from normal samples. Here, pre-cancer refers to tissue sites that were diagnosed as cervical intraepithelial neoplasia (CIN) II or worse, and normal refers to sites that were diagnosed as CIN I or better. Both are based on pathological diagnosis of cervical tissue biopsies. All EEMs were measured from sites with colposcopic tissue type “squamous”, and from pre-menopausal patients.

As shown in Figure 1, an important characteristic of the EEM data is that, the supports of the 16 spectral curves are different—the left and right boundaries are cropped differently at different excitations. Such cropping is common practice in EEM data preprocessing, with the purpose of excluding unwanted noise (Chang et al., 2002). Besides different supports, the 16 spectral curves are likely to be correlated due to their natural ordering caused by excitation wavelengths. These data constitute an example of multivariate functional data with functional variables supported on different domains.

Analytical approaches that extract features by treating curves as vectors (Yamal et al., 2012; Chang et al., 2013) are often considered suboptimal because they fail to take into account the “functional” nature of the curves. Zhu and Cox (2009) and Zhu et al. (2010) adopted scalar-on-function regressions to select “important” functional predictors while simultaneously performing classification. Although scalar-on-function regressions are effective for many purposes, they do not directly model the correlation between curves and are not

flexible enough to incorporate complex multi-level data structures. To address  
160 these issues, we introduce a general functional-response regression framework for  
the joint analysis of multivariate functional data. This new framework allows us  
to model the intra- and inter-function correlations, to estimate the effects of var-  
ious covariates, and to incorporate potential hierarchical structures in the data.  
The application to EEM data helps identify systematic differences between pre-  
165 cancer and normal samples, and enables the prediction of disease status for new  
observations.

### 3. Methods, calculations, and posterior analysis

In this section, we establish the notion of multivariate functional objects  
and introduce the proposed regression framework. To clarify the notations, we  
170 use underlined letters (e.g.,  $\underline{Y}$ ,  $\underline{\mu}$ ) to denote a multivariate functional object,  
and use boldfaced letters (e.g.,  $\mathbf{Y}$ ,  $\mathbf{B}$ ) to denote vectors of multivariate func-  
tional objects. Regular sequences, vectors, or matrices are also highlighted using  
boldface (e.g.,  $\mathbf{d}$ ,  $\mathbf{t}$ ,  $\mathbf{X}$ ).

#### 3.1. The multivariate functional mixed models

We define a multivariate functional object following a setup similar to that  
175 of Zhu et al. (2016a). Let  $\underline{Y} = (Y^1, \dots, Y^p)$  denote a collection of random  
functions, with the  $j$ th component  $Y^j$  taking values in  $L^2(T_j)$ , and  $T_j$  be-  
ing a closed subset of  $\mathbb{R}^{d_j}$ ,  $d_j \geq 1$ . The domain of  $\underline{Y}$  can then be denoted  
by  $T = \bigsqcup_{j=1}^p T_j$ , where  $\bigsqcup$  denotes the disjoint union defined by  $\bigsqcup_{j=1}^p T_j =$   
180  $\bigcup_{j=1}^p \{(t, j) : t \in T_j\}$ . We use the notations  $\mathbf{t}_j$  and  $\mathbf{t}$  to denote the sets of finite  
discrete grids of  $T_j$  and  $T$  on which the functional data are measured. For each  
 $j$ , let  $\{\phi_{jk}\}_{k=1}^\infty$  denote an orthonormal basis of  $L^2(T_j)$ . We define an extended  
basis function  $\underline{\psi}_{jk} = (0, \dots, 0, \phi_{jk}, 0, \dots, 0)$ , which is a  $1 \times p$  vector of functions  
with the  $j$ th component  $\phi_{jk}$ , and 0 functions elsewhere. Then the collection  
185  $\{\underline{\psi}_{jk}, j = 1, \dots, p, k = 1, \dots, \infty\}$  forms an orthonormal basis of  $L^2(T)$ . Let  
 $(L^2(T), \mathcal{B}(L^2(T)), P)$  be a probability space, where  $\mathcal{B}(L^2(T))$  is the Borel  $\sigma$ -



algebra on  $L^2(T)$ . Then the probability law (distribution) of  $\underline{Y}$  is induced by the probability measure  $P$ .

We further define  $\underline{Y}$  as a multivariate Gaussian process as follows. Let  $\underline{\mu}_0 = (\mu_0^1, \dots, \mu_0^p)$  be an element in  $L^2(T)$ . Denote by  $\mathcal{K} = \{k_{ij} : T_i \times T_j \rightarrow \mathbb{R}\}$  a collection of covariance kernels such that  $\text{cov}\{Y^i(s), Y^j(t)\} = k_{ij}(s, t)$ ,  $s \in T_i, t \in T_j$ . We assume that  $\mathcal{K}$  is positive semidefinite and trace class. Here, positive semidefinite means that

$$\sum_{i,j=1}^p \sum_{k,l=1}^{\infty} d_{ik}d_{jl} \int_{T_j} \int_{T_i} k_{ij}(s, t)\phi_{ik}(s)\phi_{jl}(t)dsdt \geq 0$$

for any square-summable sequence  $\{d_{ik}, i = 1, \dots, p, k = 1, \dots, \infty\}$ , and trace class means that

$$\sum_{j=1}^p \sum_{l=1}^{\infty} \int_{T_j} \int_{T_i} k_{jl}(s, t)\phi_{jl}(s)\phi_{jl}(t)dsdt < \infty.$$

Then  $\underline{\mu}_0$  and  $\mathcal{K}$  uniquely determine a Gaussian process on  $L^2(T)$  (Prato, 2006), which we call a multivariate Gaussian process, and write  $\underline{Y} \sim \text{MGP}(\underline{\mu}_0, \mathcal{K})$ .

We now describe a mixed model framework treating  $\underline{Y}$  as the response. Let  $\{\underline{Y}_i, i = 1, \dots, N\}$  represent  $N$  multivariate functional responses. Let  $\{X_{ia}, \dots, X_{iA}\}$  be  $A$  covariates for the  $i$ th measurement. Furthermore, suppose that the  $N$  measurements form  $M$  batches (i.e., groups of measurements) so that the batches may be systematically different from each other. Such a batch structure may be caused by having multiple measurements (or samples) per subject (or unit). Let  $Z_{im} = 1$  if the  $i$ th measurement belongs to the  $m$ th batch, and 0 otherwise. Then a *multivariate functional mixed model* is written as

$$\underline{Y}_i = \sum_{a=1}^A X_{ia}\underline{B}_a + \sum_{m=1}^M Z_{im}\underline{U}_m + \underline{E}_i, \quad i = 1, \dots, N, \quad (1)$$

where  $\underline{B}_a = (B_a^1, \dots, B_a^p)$  is the multivariate functional fixed effect corresponding to the  $a$ th covariate,  $\underline{U}_m = (U_m^1, \dots, U_m^p)$  is the multivariate functional random effect corresponding to the  $m$ th batch, and  $\underline{E}_i = (E_i^1, \dots, E_i^p)$  is the residual that is assumed to be independent across the index  $i$ . The term

$\sum_{m=1}^M Z_{im} \underline{U}_m$  captures the group structure in the data and can be omitted  
 205 if such a structure is not present. The random effects  $\{\underline{U}_m\}$  are assumed to be  
 independent across index  $m$ , and independent of the residuals  $\{\underline{E}_i\}$ . Addition-  
 ally, all items in  $\{\underline{U}_m\}$  and  $\{\underline{E}_i\}$  are assumed to have zero means.

We further stack the responses and the regression coefficients in model (1)  
 to form vectors of multivariate functional objects. Model (1) is re-written as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Z}\mathbf{U} + \mathbf{E}, \quad (2)$$

210 where  $\mathbf{Y} = (\underline{Y}_1, \dots, \underline{Y}_N)^T$ ,  $\mathbf{B} = (\underline{B}_1, \dots, \underline{B}_A)^T$ ,  $\mathbf{U} = (\underline{U}_1, \dots, \underline{U}_M)^T$ , and  
 $\mathbf{E} = (\underline{E}_1, \dots, \underline{E}_N)^T$ . Here,  $\mathbf{X}$  is a  $N \times A$  design matrix for the fixed effect, and  
 $\mathbf{Z}$  is a  $N \times M$  binary design matrix for the random effect. Model (2) constitutes a  
 general multivariate functional mixed model framework, abbreviated as MFMM.  
 Our goal is to estimate  $\mathbf{B}$  and  $\mathbf{U}$  and characterize their uncertainties.

### 215 3.2. Model reparameterization with a two-stage transformation

Since  $\underline{Y}_i$  has been written as the linear combination of  $\{\underline{B}_a\}$ ,  $\{\underline{U}_m\}$  and  $\underline{E}_i$  in  
 model (1), it is natural to assume that the unobserved multivariate functional  
 objects  $\{\underline{B}_a\}$ ,  $\{\underline{U}_m\}$  and  $\{\underline{E}_i\}$  also take values in the same  $L^2(T)$  space as  
 $\underline{Y}_i$ . With this assumption, all multivariate functional objects in the MFMM  
 220 can be represented by basis expansions using a common orthonormal basis.  
 Taking  $\underline{Y}_i$  as an example, let  $\{\phi_{jk}\}_{k=1}^\infty$  denote an orthonormal basis of  $L^2(T_j)$ ,  
 then  $Y_i^j(t) = \sum_{k=1}^\infty d_{ijk} \phi_{jk}(t)$  where  $d_{ijk} = \langle Y_i^j, \phi_{jk} \rangle = \int_{T_j} Y_i^j(t) \phi_{jk}(t) dt$ . The  
 coefficient sequence  $\mathbf{d}_i^j = (d_{ij1}, d_{ij2}, \dots)$  lies in the space of square-summable  
 sequences, denoted by  $\ell_j^2 = \{d_{jk} : \sum_{k=1}^\infty d_{jk}^2 < \infty\}$ . Let  $\mathbf{d}_i = (\mathbf{d}_i^1, \dots, \mathbf{d}_i^p)$ , then  
 225  $\mathbf{d}_i$  is a concatenated vector of infinite sequences and  $\mathbf{d}_i$  lies in  $\ell^2 = \prod_{j=1}^p \ell_j^2$ .  
 Since  $\ell_j^2$  and  $L^2(T_j)$  are isometrically isomorphic for each  $j$ , once an orthonormal  
 basis of  $L^2(T)$  has been chosen, there is a one-to-one mapping between  $L^2(T)$   
 and  $\ell^2$ . This enables us to transform the MFMM from  $L^2(T)$  to  $\ell^2$ .

To be more specific, denote by  $\{\mathbf{d}_i\}$ ,  $\{\mathbf{b}_a\}$ ,  $\{\mathbf{u}_m\}$ ,  $\{\mathbf{e}_i\}$  the  $\ell^2$  space counter-  
 parts of  $\{\underline{Y}_i\}$ ,  $\{\underline{B}_a\}$ ,  $\{\underline{U}_m\}$ , and  $\{\underline{E}_i\}$ , respectively. We can stack the sequences

in  $\ell^2$  to form two-dimensional (2-d) arrays. For example, denote

$$\mathbf{D} = \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_N \end{pmatrix} = \begin{pmatrix} \mathbf{d}_1^1 & \mathbf{d}_1^2 & \cdots & \mathbf{d}_1^p \\ \mathbf{d}_2^1 & \mathbf{d}_2^2 & \cdots & \mathbf{d}_2^p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{d}_N^1 & \mathbf{d}_N^2 & \cdots & \mathbf{d}_N^p \end{pmatrix},$$

then  $\mathbf{D}$  is a 2-d array with  $N$  rows and  $p$  column-blocks, and each block (e.g.,  $\mathbf{d}_1^2$ ) is an infinite sequence. We similarly denote  $\mathbf{B}^* = (\mathbf{b}_1, \dots, \mathbf{b}_A)^T$ ,  $\mathbf{U}^* = (\mathbf{u}_1, \dots, \mathbf{u}_A)^T$ , and  $\mathbf{E}^* = (\mathbf{e}_1, \dots, \mathbf{e}_N)^T$ . The basis expansions in  $L^2(T)$  can then be represented through linear operations:  $\mathbf{Y} = \mathbf{D}\Psi$ ,  $\mathbf{B} = \mathbf{B}^*\Psi$ ,  $\mathbf{U} = \mathbf{U}^*\Psi$ ,  $\mathbf{E} = \mathbf{E}^*\Psi$ , where  $\Psi$  is a 2-d array (with an infinite number of rows and  $p$  columns) consisting of basis functions, i.e.,

$$\Psi = \begin{pmatrix} \psi^1 & & & \\ & \psi^2 & & \\ & & \ddots & \\ & & & \psi^p \end{pmatrix},$$

and  $\psi^j = (\phi_{j1}, \phi_{j2}, \dots)^T$  is a column vector of basis functions in  $L^2(T_j)$  for  $j = 1, \dots, p$ . Since  $\Psi$  preserves the linear operation, model (2) is equivalent to the following model in the  $\ell^2$  space:

$$\mathbf{D} = \mathbf{X}\mathbf{B}^* + \mathbf{Z}\mathbf{U}^* + \mathbf{E}^*, \quad (3)$$

Note that  $\mathbf{D}$ ,  $\mathbf{B}^*$ ,  $\mathbf{U}^*$ , and  $\mathbf{E}^*$  are 2-d arrays with the same number of rows as their  $L^2(T)$ -space counterparts and an infinite number of columns. Since  $\ell^2$  is the dual space of  $L^2(T)$ , we call model (3) the dual-space model.

The dual-space model in (3) brings several advantages for model fitting. First, many basis expansions, such as Fourier expansion and wavelets, have the “whitening” property, meaning that, compared with the correlation between function values at two adjacent grid points, the correlation between the basis coefficients is substantially reduced. As an illustration, Figure 2 (b)-(c) shows the empirical correlations before and after wavelet transformation based on simulated EEM data. This property allows us to assume simplified correlation

structures (e.g., diagonal) for the intra-function correlation in the dual-space model. In the dual space, the intra-function correlation refers to the correlation between the basis coefficients of the same function. Second, for smooth functional data, the coefficient sequence in  $\ell^2$  is often sparse (e.g., quickly decays to zero), thus various compression techniques can be applied to zero-out the small coefficients or truncate the tails of the coefficient sequence. Third, the one-to-one mapping between  $l^2$  and  $L^2(T)$  allows us to freely transform posterior samples of the regression coefficient functions between the two spaces, so that posterior inferences can be performed directly in the data domain.

Under model (3), we need to parameterize the distributions for the 2-d arrays:  $\mathbf{D}$ ,  $\mathbf{B}^*$ ,  $\mathbf{U}^*$ , and  $\mathbf{E}^*$ . Since the original data are assumed to be multivariate Gaussian processes, these 2-d arrays are discrete Gaussian processes. If we assume a separable correlation structure, the correlations of a 2-d array can be characterized by the Kronecker product of the between-row correlation and between-column correlation. Even with the whitening property of the basis expansion, characterizing the correlation of such 2-d arrays is still challenging, since we usually do not have a priori knowledge about the inter-functional correlation. We therefore propose a second stage transformation—an orthogonal linear transformation (a rotation) that transforms model (3) to a new coordinate system determined by the PC directions of  $\mathbf{D}$ .

Specifically, let  $\mathbf{V}$  denote a 2-d array (with infinite number of rows and columns), for which the columns are the eigenvectors of  $\mathbf{D}^T \mathbf{D}$ . We obtain the PC scores by projecting rows of  $\mathbf{D}$  on the eigenvectors, i.e.,  $\tilde{\mathbf{D}} = \mathbf{D}\mathbf{V}$ . Each row of  $\tilde{\mathbf{D}}$  consists of PC scores of the corresponding row in  $\mathbf{D}$ . By applying the same rotation/projection to  $\mathbf{B}^*$ ,  $\mathbf{U}^*$ ,  $\mathbf{E}^*$ , we obtain  $\tilde{\mathbf{B}} = \mathbf{B}^* \mathbf{V}$ ,  $\tilde{\mathbf{U}} = \mathbf{U}^* \mathbf{V}$ , and  $\tilde{\mathbf{E}} = \mathbf{E}^* \mathbf{V}$ . Since the rotation operator  $\mathbf{V}$  is again a linear operator, model (3) is equivalent to

$$\tilde{\mathbf{D}} = \mathbf{X}\tilde{\mathbf{B}} + \mathbf{Z}\tilde{\mathbf{U}} + \tilde{\mathbf{E}}. \quad (4)$$

In the above model,  $\tilde{\mathbf{D}} = (\tilde{d}_{ik}), i = 1, \dots, N, k = 1, 2, \dots$ , where  $\tilde{d}_{ik}$  denotes the  $k$ th (uncentered) PC score for the row  $i$  of  $\mathbf{D}$ . Note that  $\tilde{\mathbf{D}}, \tilde{\mathbf{B}}, \tilde{\mathbf{U}}, \tilde{\mathbf{E}}$  have

the same dimensions as their counterparts in model (3). Since  $\mathbf{V}$  is orthogonal ( $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$ ), the second-stage transformation is also invertible. Due to the Gaussian assumption of  $\{\underline{Y}_i\}$ , the PC scores of  $\mathbf{D}$  (i.e., columns of  $\tilde{\mathbf{D}}$ ) are approximately independent of each other. Therefore we can make a reasonable  
275 assumption that the columns of  $\tilde{\mathbf{D}}, \tilde{\mathbf{B}}, \tilde{\mathbf{U}}, \tilde{\mathbf{E}}$  in model (4) are independent of each other. This facilitates a much simpler prior setup and calculation, as elaborated in Section 3.3.

### 3.3. Model setup, prior specification, and posterior inference

We extract the  $k$ th column of  $\tilde{\mathbf{D}}, \tilde{\mathbf{B}}, \tilde{\mathbf{U}}, \tilde{\mathbf{E}}$ , respectively, and write model (4)  
280 as

$$\tilde{\mathbf{D}}_k = \mathbf{X}\tilde{\mathbf{B}}_k + \mathbf{Z}\tilde{\mathbf{U}}_k + \tilde{\mathbf{E}}_k, \quad k = 1, 2, \dots \quad (5)$$

where  $\tilde{\mathbf{D}}_k$  and  $\tilde{\mathbf{E}}_k$  are  $N \times 1$  vectors,  $\tilde{\mathbf{B}}_k$  is of size  $A \times 1$ , and  $\tilde{\mathbf{U}}_k$  is  $M \times 1$ . Denoting the  $a$ th entry of  $\tilde{\mathbf{B}}_k$  by  $\tilde{B}_{ak}$ , we set the random effect and residual distributions as well as the priors for fixed effects as

$$\begin{aligned} \tilde{\mathbf{E}}_k &\sim N(0, s_k \mathbf{I}_N), \quad \tilde{\mathbf{U}}_k \sim N(0, q_k \mathbf{I}_M), \quad \tilde{B}_{ak} \sim \gamma_{ak} N(0, \zeta_{ak}) + (1 - \gamma_{ak}) \delta_0, \\ s_k &\sim IG(a_k^s, b_k^s), \quad q_k \sim IG(a_k^q, b_k^q), \quad \zeta_{ak} \sim IG(a_k^\zeta, b_k^\zeta), \quad \gamma_{ak} \sim \text{Bern}(\pi_{ak}), \end{aligned}$$

and set  $\pi_{ak} \sim \text{Beta}(a_k^\pi, b_k^\pi)$ . These distributions are assumed to be independent  
285 across the column index  $k$ . The prior for  $\tilde{B}_{ak}$  takes the ‘‘spike-and-normal-slab’’ form, with the spike part being  $\delta_0$ , the point mass at zero. This prior enables *adaptive shrinkage* for the fixed effect, i.e., smaller components of  $\tilde{\mathbf{B}}$  are encouraged toward zero while large components are retained, resulting in a sparse estimate of  $\tilde{\mathbf{B}}$ . Under the above model setup, posterior sampling can be  
290 performed using a Markov chain Monte Carlo (MCMC) algorithm. Details of the MCMC algorithm and methods for estimating the initial values and setting prior parameters are described in the supplementary materials.

One practical issue involved in model fitting and calculation is how to deal with the infinite-dimensional arrays in model (3) and model (4). While theoretically  
295 infinite-dimensional, the dimensions of the 2-d arrays  $\mathbf{D}$  and  $\tilde{\mathbf{D}}$  in practical

calculations are constrained by the number of measurement points and sample size. For example, the number of bases one can use to expand a curve is often smaller than or equal to the number of measurement points of the curve, and the number of positive eigenvalues one can obtain in the PCA of  $\mathbf{D}$  is less than the smaller dimension of  $\mathbf{D}$ . In most applications, the functional data are smooth and densely measured, so that the number of measurement points is usually much higher than the sample size. In these situations, we often choose to compress the basis coefficients so that the close-to-zero values are replaced by zeros, or truncate the coefficients sequence when the basis coefficients decay monotonically. Compression and truncation have the effects of smoothing (removing high frequency noise) and reducing the dimension of functional data. They are often preferred for high-dimensional problems. In this paper, we recommend keeping the two-stage transformation “lossless” or “near-lossless”, meaning that we only perform moderate truncation or compression so that nearly all essential information is retained after the transformation. This can be achieved by controlling the fraction of information retained (FIR). For example, during the basis expansion at stage one, we can truncate the sequence  $\mathbf{d}_i^j$  at  $\tau_j$  so that the ratio  $R_{ij} = \sum_{k=1}^{\tau_j} d_{ijk}^2 / \sum_k d_{ijk}^2$  is greater than or equal to a pre-specified threshold (e.g., 0.999) for all  $i = 1, \dots, N$  and  $j = 1, \dots, p$ . In the PCA at stage two, we simply truncate the PC scores at  $\tau$  so that  $R_{pc} = \sum_{k=1}^{\tau} \hat{\lambda}_k / \sum_k \hat{\lambda}_k$  is greater than or equal to a pre-specified threshold (e.g.,  $1-1\mathbf{e}-6$ ). Here,  $\hat{\lambda}_k$  denotes the  $k$ th estimated eigenvalue of  $\mathbf{D}^T \mathbf{D}$ .

### 3.4. Posterior inference

The MCMC algorithm for the proposed model yields posterior samples of the model parameters, from which inferential summaries can be constructed. The posterior samples for  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{U}}$  can be transformed back into the data space using the two-stage inverse-transform:  $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{V}^T\mathbf{\Psi}$  and  $\mathbf{U} = \tilde{\mathbf{U}}\mathbf{V}^T\mathbf{\Psi}$ , yielding posterior samples for  $\mathbf{B}$  and  $\mathbf{U}$  in the data space model (2). The posterior samples can also be computed for any function of the parameters, such as the contrast effects between two groups. For example, in the fluorescence

spectroscopy data analysis, a key inferential objective is to identify locations on the EEM measurements that are differentially expressed across the normal samples and pre-cancer samples. Let  $\mathbf{X}$  denote a  $N \times 2$  matrix with the  $i$ th row being  $(1, 0)$  if measurement  $i$  is from the pre-cancer group and  $(0, 1)$  if from the normal group. Then the corresponding coefficient  $\mathbf{B} = (\underline{B}_1, \underline{B}_2)^T$  contains the mean effect for the pre-cancer group ( $\underline{B}_1$ ) and the normal group ( $\underline{B}_2$ ), respectively. Based on posterior samples of  $\mathbf{B}$ , the contrast effect between the two groups can be calculated by  $\underline{C} = \underline{B}_2 - \underline{B}_1$  for each posterior sample. We can then identify significant nonzero regions on  $\underline{C}$ . These regions reflect significant differences between the two groups.

**Flagging Regions.** Most existing methods rely on pointwise credible bands to identify significant local regions. They first construct the 95% credible interval on each grid separately, then flag a position (a grid point) as significantly nonzero if the credible band does not include zero in that position. However, since pointwise credible bands do not adjust for joint coverage probabilities, inference based on these approaches fails to adjust for the family-wise (experimental-wise) error rate in the inherent multiple testing problem, which may lead to high false discovery rates (Crainiceanu et al., 2012). In this paper, we flag regions using two approaches with global coverage properties: a thresholding method based on the simultaneous band scores (SimBaS) and a Bayesian false discovery rate (BFDR) method.

1. *Simultaneous band scores (SimBaS).* To detect regions that are significantly nonzero on a contrast effect  $\underline{C} = (C^1, \dots, C^p)$ , we first generate simultaneous credible bands (SCBs) for the multivariate functional object  $\underline{C}$  following Ruppert et al. (2003). The SCB takes the form  $[\widehat{\underline{C}} - m_\alpha \widehat{\text{sd}}\{\underline{C}\}, \widehat{\underline{C}} + m_\alpha \widehat{\text{sd}}\{\underline{C}\}]$ , where  $\widehat{\underline{C}}$  is the sample mean,  $\widehat{\text{sd}}\{\underline{C}\}$  is the sample standard deviation, and  $m_\alpha$  is the  $(1 - \alpha)$  sample quantile of  $\max_{\mathbf{t}} \left\{ |\underline{C}^{(g)} - \widehat{\underline{C}}| / \widehat{\text{sd}}\{\underline{C}\} \right\}$ ,  $g = 1, \dots, H$ . Here,  $g$  is the index for (and  $H$  is the total number of) posterior samples, and  $\mathbf{t}$  denotes the discrete grid of the domain  $T$  on which the posterior samples were evaluated. We then calculate the SCB for a range of  $\alpha$  values, and define the SimBaS at each grid point  $t \in \mathbf{t}$  as the smallest  $\alpha$  for which the  $100(1 - \alpha)\%$

SCB excludes zero at that point. This measure was introduced by Meyer et al. (2015). Given the joint coverage properties of the SimBaS, we can compute a global Bayesian p-value (GBPv) as  $\min_t \{\text{SimBaS}(t)\}$ , which can be used to test the global functional null hypothesis that  $H_0 : \underline{C} \equiv 0$  vs.  $H_a : C^j \neq 0$  for some  $j$ . If  $\text{GBPv} < \alpha$ , we conclude that there is some difference between the two groups, and can subsequently localize these effects by flagging locations as strongly significant if the corresponding  $\text{SimBaS}(t)$  is less than  $\alpha$ .

2. *Bayesian false discovery rate (BFDR)*. At times, we are interested in identifying locations at which the contrast effect  $\underline{C}$  is greater than some pre-specified practical effect size  $\delta$ . To do this, we first calculate the pointwise posterior probability  $\hat{p}_j(t) \approx \Pr(|C^j(t)| > \delta \mid \text{Data})$  from the posterior samples for  $j = 1, \dots, p$ . The values  $\{1 - \hat{p}_j(t)\}$  can be interpreted as an estimate of the *local FDR* at location  $t$  for functional variable  $j$ , if we consider a discovery to be a location where the effect is in fact greater than  $\delta$  in magnitude. We can then find a threshold  $\phi_\alpha$  for  $\{\hat{p}_j(t)\}$ , for example corresponding to a pre-specified expected FDR (averaged across all locations and all  $j$ ) of  $\alpha$ , and flag locations with  $\hat{p}_j(t) > \phi_\alpha$  as being significantly greater than  $\delta$ . Similar strategies have been adopted by Newton et al. (2004) and Morris et al. (2008).

Comparing the two methods, we see that the BFDR method uses the weaker FDR criterion but requires the pre-specification of a threshold  $\delta$ , whereas the SimBaS analysis corresponds to EER considerations but does not require specification of an effect size of interest,  $\delta$ .

**Discriminant Analysis.** Besides flagging significant local regions, the proposed MFMM framework also facilitates discriminant analysis through posterior predictive calculation. We use the symbol  $\dagger$  to denote data from a new observation. Specifically, let  $\tilde{\mathbf{D}}^\dagger$  denote a row vector containing the PC scores for the new observation, calculated using the eigenvectors estimated in the training set. Let  $c^\dagger \in \{1, \dots, S\}$  denote the class label that corresponds to the design vector  $\mathbf{X}^\dagger$ . For example, in the EEM case, we can set  $c^\dagger = 1$  if the design vector for the new observation is  $\mathbf{X}^\dagger = (1, 0)$ , which corresponds to the pre-cancer class, and set  $c^\dagger = 2$  if  $\mathbf{X}^\dagger = (0, 1)$ , which corresponds to the normal class. In this



case,  $S = A = 2$ , where  $A$  is the number of covariates in model (1). In general,  $S$  can be less than  $A$ , meaning that not all covariates need be relevant to the classes.

A discrimination analysis involves predicting  $c^\dagger$  using posterior samples obtained from the training procedure. In particular, we predict by calculating the posterior predictive odds (PPO) of  $c^\dagger = r$  versus  $c^\dagger = 1$ , i.e.,

$$\text{PPO}(r) = \frac{f(\tilde{\mathbf{D}}^\dagger | c^\dagger = r, \tilde{\mathbf{D}}, \mathbf{X}, \mathbf{Z})}{f(\tilde{\mathbf{D}}^\dagger | c^\dagger = 1, \tilde{\mathbf{D}}, \mathbf{X}, \mathbf{Z})} \cdot \frac{\Pr(c^\dagger = r)}{\Pr(c^\dagger = 1)}, \quad r \in \{2, \dots, S\}, \quad (6)$$

where  $\Pr(c^\dagger = r)$  and  $\Pr(c^\dagger = 1)$  are the pre-specified prior probabilities for the class designation, and  $f(\tilde{\mathbf{D}}^\dagger | c^\dagger, \tilde{\mathbf{D}}, \mathbf{X}, \mathbf{Z})$  is the posterior predictive density for the new observation, calculated by  $\int f(\tilde{\mathbf{D}}^\dagger | c^\dagger, \Theta) f(\Theta | \tilde{\mathbf{D}}, \mathbf{X}, \mathbf{Z}) d\Theta$  using Monte Carlo integration. Here,  $\Theta$  denote all model parameters and  $f(\Theta | \tilde{\mathbf{D}}, \mathbf{X}, \mathbf{Z})$  is the posterior density of the parameters in the training procedure. When calculating the predictive density  $f(\tilde{\mathbf{D}}^\dagger | c^\dagger, \Theta)$ , we need to integrate out the random effect term  $\mathbf{Z}^\dagger \mathbf{U}^\dagger$  since the random effect for the new observation cannot be estimated from the training procedure. This integration can be done analytically using the properties of Gaussian random variables. The PPO is generally applicable to the binary or multiple-class case. With the odds computed using equation (6), the posterior predictive probabilities for class designations can be determined straightforwardly using

$$\Pr(c^\dagger = r | \tilde{\mathbf{D}}^\dagger, \mathbf{X}^\dagger, \tilde{\mathbf{D}}, \mathbf{X}, \mathbf{Z}) = \begin{cases} \frac{\text{PPO}(r)}{1 + \sum_{v=2}^S \text{PPO}(v)}, & \text{if } r = 2, \dots, S. \\ \frac{1}{1 + \sum_{v=2}^S \text{PPO}(v)}, & \text{if } r = 1. \end{cases}$$

#### 4. Simulation study

##### 4.1. Simulation 1. Assess estimation accuracy and discrimination performance

We designed a simulation study to demonstrate the performance of the proposed MFMM approach. In particular, we want to assess the benefit of jointly modeling multivariate functional data versus independently modeling each functional variable. To capture realistic inter- and intra-function correlations, we

400 based our simulation upon the real EEM data described in Section 2, i.e., treating the EEM data as the reference data. We generated simulated data by first fitting MFMM to the reference EEM data using model (2). We set the design matrix  $\mathbf{X}$  by following a cell mean design, i.e., the  $i$ th row of  $\mathbf{X}$  equals  $(1, 0)$  if the  $i$ th EEM measurement is from a pre-cancer sample and  $(0, 1)$  if from the  
 405 normal sample. Correspondingly, the elements in  $\mathbf{B} = (\underline{B}_1, \underline{B}_2)^T$  denote the mean of the pre-cancer group and normal group, respectively. Since the real EEM data do not contain multiple batches, the  $\mathbf{ZU}$  term is omitted in the reference run.

From the reference run, we obtained the estimated values for  $\tilde{\mathbf{B}}$  and  $\{s_k\}$   
 410 in the transformed domain. We then simulated data, treating the estimated  $\tilde{\mathbf{B}}$  as the true fixed effect. We simulated the random effect  $\tilde{\mathbf{U}}$  and residuals  $\tilde{\mathbf{E}}$  based on the normal distributions described in Section 3.3, using pre-specified variance components  $\{q_k, s_k\}$ . Here, we specify  $q_k = r_1 \hat{s}_k$  and  $s_k = r_2 \hat{s}_k$ , where  $\hat{s}_k$  is the posterior mean estimated from the reference run and  $r_1, r_2 \in (0, 1)$  are  
 415 tuning parameters that control the signal to noise ratio (SNR). Specifically, we set  $r_1 = 0.075$ ,  $r_2 = 0.025$  so that the resulting SNR, defined by the ratio (in absolute value) between  $\mathbf{X}\tilde{\mathbf{B}}$  and  $\mathbf{ZU} + \mathbf{E}$ , component-wise, has a mean value around 5. In this simulation, we set the design matrix  $\mathbf{X}$  to be identical to that of the reference EEM data, and set  $\mathbf{Z}$  so that the samples are randomly  
 420 assigned to  $M = 5$  batches. In order to test the predictive performance, we simulated one training set (with 534 EEM observations) and one test set (260 EEM observations) independently using the same parameter setup. We repeated the above simulation procedure 20 times and summarized the results by averaging the statistics calculated from each simulation.

425 The simulated data and the effect of two-stage transformation are demonstrated in Figure 2, where panel (a) shows simulated samples for the two groups (10 samples each) as concatenated spectral curves. Unlike in Figure 1 (c), here, we use the x-axis to label the excitation wavelengths of the spectral curves. Such relabeling helps distinguish which segment of the concatenated curve corresponds to which spectrum. We performed a two-stage transformation, a wavelet  
 430

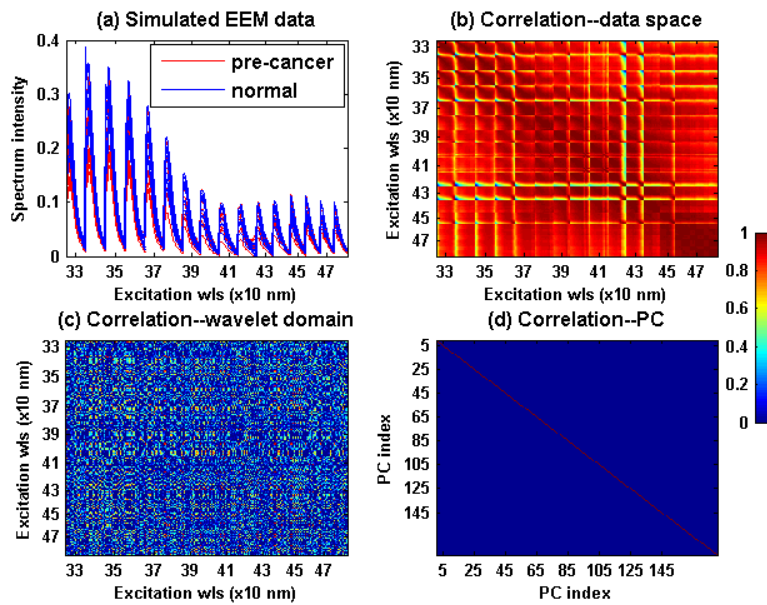


Figure 2: Simulated data: (a) Simulated samples plotted as concatenated spectral curves for the pre-cancer group (red) and the normal group (blue); (b) image plot of the sample correlation of the concatenated spectral curves; (c) sample correlation in wavelet domain; (d) sample correlation of the PC scores after the two-stage transformation. Here, the x-axis of (a) and both axes of (b)-(c) have been relabeled by the excitation wavelengths (wls) of the spectral curves.

transformation followed by a PCA on the concatenated wavelet coefficients, and plotted the sample correlation before and after each stage of the transformation in Figure 2 (b)-(d). Panel (b) shows the sample correlation of the concatenated spectral curves before the two-stage transformation. It demonstrates high intra- and inter-function correlations (with average correlation around 0.87). After the first-stage wavelet transformation, the sample correlation of the concatenated wavelet coefficients is reduced substantially (with average value around 0.004), as shown in Figure 2 (c). Yet high correlation blocks are still observed for certain wavelet coefficients. In Figure 2 (b)-(c), we again relabeled the x-axis and y-axis using the excitation wavelengths to help distinguish which correlation block corresponds to which spectral curves. After performing PCA of the concatenated wavelet coefficients, the PC scores become nearly independent (with average correlation around  $5e-16$ ), as shown in Figure 2 (d). Furthermore, since the two-stage transformation is invertible, we are able to reconstruct the original data from the PC scores, which gives the maximum reconstruction error  $4.4e-4$ .

**Evaluation criteria.** We applied the proposed MFMM model to the simulated dataset. To demonstrate the benefits of joint modeling, we also applied a functional mixed model (FMM) with the same model setup (i.e., the same two-stage transformation, parameter setup, and prior setup) to each of the functional variables independently. We collected the posterior samples from both scenarios (MFMM and FMM), and calculated six summary statistics to evaluate the estimation performance. The statistics included

$$\text{IMSE} = \frac{1}{Ap} \sum_{a=1}^A \sum_{j=1}^p \frac{\|\widehat{B}_{aj}(t) - B_{aj}(t)\|^2}{\|B_{aj}(t)\|^2}, \quad \text{IPVar} = \frac{1}{Ap} \sum_{a=1}^A \sum_{j=1}^p \frac{\frac{1}{H} \sum_{g=1}^H \|B_{aj}^{(g)}(t) - \widehat{B}_{aj}(t)\|^2}{\|B_{aj}(t)\|^2},$$

$\text{IWidth} = 1/(Ap) \sum_{a=1}^A \sum_{j=1}^p \|\widehat{w}_{B_{aj}}(t)\|^2 / \|B_{aj}(t)\|^2$ , the mean coverage probability of the 95% SCB for each functional variable in  $\mathbf{B}$  (denoted by  $\text{CPr}_{\text{scb}}$ ), as well as the coverage probability of the 95% aggregated SCB for  $\mathbf{B}$  (denoted by  $\text{CPr}_{\text{agg}}$ ), which adjusts for the EER of all 16 functional variables in  $\mathbf{B}$  simultaneously. In the above formulae,  $\|\cdot\|$  denotes the  $L^2$  norm,  $H$  is the number of posterior samples,  $\widehat{w}_{B_{aj}}(t)$  is the width of the 95% pointwise credible band of  $B_{aj}(t)$  (the  $j$ th component of  $\underline{B}_a$ ) estimated on the grid  $T_j$ . The IMSE sum-

Table 1: Summary statistics of simulation 1: integrated mean squared error (IMSE), integrated posterior variance (IPVar), integrated width of 95% credible interval (IWidth), coverage probability of the 95% SCB ( $CPr_{scb}$ ) and the 95% aggregated SCB ( $CPr_{agg}$ ) for  $\mathbf{B}$  and  $\mathbf{U}$ , averaged across 20 repeated simulations; the area under the ROC curve (AUC) when predicting class labels of the test set, and extended AUC calculated when constraining 1-specificity < 10% ( $AUC_{0.1}$ ).

Model	$\mathbf{B}$					Prediction
	IMSE	IPVar	IWidth	$CPr_{scb}$	$CPr_{agg}$	AUC
MFMM	0.016	0.014	0.197	0.950	0.978	0.966
FMM	0.031	0.013	0.187	0.928	0.969	0.951
Model	$\mathbf{U}$					$AUC_{0.1}$
	IMSE	IPVar	IWidth	$CPr_{scb}$	$CPr_{agg}$	
MFMM	3.021	3.020	42.472	0.951	0.978	0.874
FMM	20.199	4.820	66.466	0.930	0.970	0.789

marizes the deviation of the posterior mean about the truth, and the IPVar summarizes the variability about the posterior mean.

**Simulation results.** We applied MFMM and FMM to each of the 20 simulated data sets. In each simulation, we ran 5000 MCMC iterations and treated the first 3000 iterations as the burn-in period. We monitored the convergence of the MCMC samples; see Section 7 and the supplementary materials for details. We obtained posterior samples for  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{U}}$  in the transformed PC domain and inverse-transformed them to the data domain. During the fittings of both the MFMM and FMM, we performed discrete wavelet decomposition in stage one with no compression, and performed a near lossless PCA in stage two so that the truncation parameter  $\tau$  (or  $\{\tau_j\}$  in the FMM case) is chosen to retain at least 99.99% of the total variation. Based on the posterior samples, we calculated the summary statistics for  $\mathbf{B}$  and  $\mathbf{U}$ . We averaged these statistics across all 20 repetitions and listed them in Table 1.

From Table 1, we see that MFMM resulted in systematically lower IMSE and higher coverage probabilities ( $CPr_{scb}$  and  $CPr_{agg}$ ) for  $\mathbf{B}$  and  $\mathbf{U}$  than the FMM approach. This implies that, when functional variables are correlated, modeling multivariate functional data jointly using MFMM may lead to smaller estimation

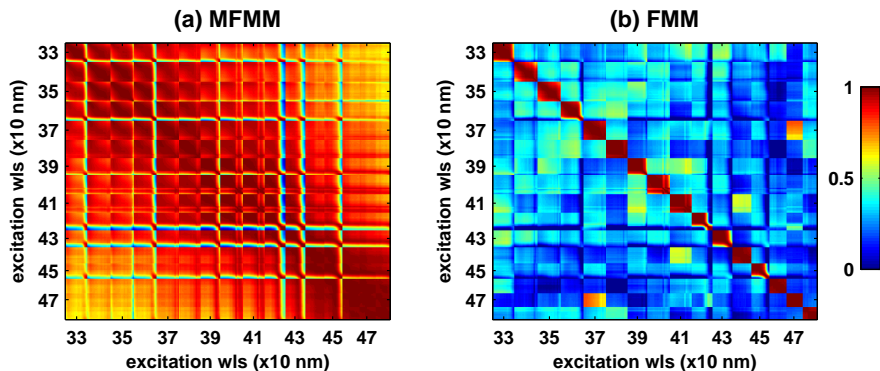


Figure 3: Simulation 1: posterior estimation of the intra- and inter-function correlation for the group mean  $\underline{B}_1$  using MFMM (a) and FMM (b). Here, the axes have been relabeled by the excitation wavelengths (wls) of the spectral curves similarly to Figure 2 (b)-(c).

errors for the fixed and random effect, as well as more reliable confidence bands.

480 For the posterior variance, MFMM provides comparable IPVar and IWidth for  $\mathbf{B}$  but considerably smaller IPVar and IWidth for  $\mathbf{U}$ . This suggests that, compared to the FMM approach, the MFMM approach has evidently improved the estimation accuracy of the random effect  $\mathbf{U}$ .

In addition to the summary statistics, in Figure 3 we demonstrate the posterior 485 estimation of the intra- and inter-function correlation for the group mean  $\underline{B}_1$  using MFMM and FMM. In Figure 3, similarly to Figure 2 (b)-(c), we relabeled the axes to indicate which block corresponds to which spectral curve using the excitation wavelengths of each spectral curve. Figure 3 shows that joint modeling using MFMM resulted in high intra- and inter-function correlations 490 for the estimated fixed effect  $\underline{B}_1$  (with average correlation 0.81), whereas the independent modeling approach using FMM only retained high intra-function correlation, and resulted in a generally lower inter-functional correlation (average value 0.26).

Besides estimating parameters, we also applied discriminant analysis to predict 495 the class labels (pre-cancer versus normal tissue) of the test data. To summarize the prediction results, we used two statistics: the area under the receiver operating characteristic (ROC) curve, denoted by AUC, and a modified  $AUC_{0.1}$

statistic. To calculate the AUC, we first generated the ROC curve by plotting 1-Specificity (x-axis) versus Sensitivity (y-axis) for all possible thresholds applied  
 500 to the predictive probabilities for the test data. The AUC is the area under the ROC curve computed using numerical integration. AUC takes values in  $[0, 1]$ , with higher values implying better classification performance. The modified  $AUC_{0.1}$  statistic focuses on the lower left portion of the ROC curves. It describes the classification performance while controlling the false positive rate  
 505 to be less than 10% (which is often preferred in medical practice). The  $AUC_{0.1}$  is re-scaled (multiplied by 10) so that it also takes values in  $[0, 1]$ . We averaged the two summary statistics across the 20 repeated simulations; results are listed in the last column in Table 1. From Table 1, we see that the MFMM approach resulted in slightly higher AUC and considerably higher  $AUC_{0.1}$  than the FMM  
 510 approach. This indicates that joint modeling using MFMM has the potential to improve the discrimination power over that obtained from independent modeling using FMM.

#### 4.2. Simulation 2. Assess the power of the global test for contrast effects

Motivated by the fact that a multivariate test may be more powerful when  
 515 the variables to be tested are correlated, we assessed whether MFMM results in higher power for hypothesis testing. We considered the same design as in simulation 1, and aimed to assess the power of an omnibus test for the contrast effect  $\underline{C}$ , i.e.,  $H_0 : \underline{C} \equiv 0$  versus  $H_1 : C^j(t) \neq 0$  for at least one  $j$ . Here,  $\underline{C} = \underline{B}_2 - \underline{B}_1$ . According to the definition, the power of a hypothesis test is the  
 520 probability of rejecting  $H_0$  when  $H_1$  holds. Therefore, to compare the power, we simulated data under  $H_1$ . To facilitate comparable power (power that is not always 1 in all cases), we simulated data by controlling the true contrast effect based upon simulation 1. Specifically, we adopted the same  $\underline{B}_1$  estimated from the reference run, but let  $\underline{B}_2 = \underline{B}_1 + r\underline{C}$ , where  $\underline{C}$  is the mean contrast effect  
 525 estimated in the reference run. We set  $r = 0.15$  so that the true contrast effect was fairly small, thus there as a higher chance of accepting  $H_0$ . This setup would lead to power with values lower than 1.

We simulated 1000 datasets following the same procedure as in simulation 1. For each simulated dataset, we applied both MFMM and FMM, and collected the posterior samples of  $\underline{C}$ . We then adopted the SimBaS approach on the grid  $\mathbf{t}$  of  $\underline{C}$  while controlling the EER to be less than 0.05. Based on SimBaS, we calculated the GBPV and rejected  $H_0$  when  $\text{GBPV} < 0.05$ . We finally calculated the power as the frequency of successfully rejecting  $H_0$  across the 1000 repetitions. Results show that the SimBaS-based omnibus test gives power of 0.85 for the MFMM approach and 0.69 for the FMM approach. This suggests that MFMM can improve the power of testing the global contrast effect.

## 5. Analysis of fluorescence spectroscopy data

We applied MFMM to the fluorescence spectroscopy data using a design similar to that of the reference run in simulation 1. In particular, the design matrix  $\mathbf{X}$  is a  $534 \times 2$  binary matrix following the cell mean design, and the corresponding regression coefficient  $\mathbf{B} = (\underline{B}_1, \underline{B}_2)^T$  contains the group means for the pre-cancer and normal samples. The random effect term  $\mathbf{ZU}$  is omitted due to the absence of a batch/group effect. Our goal for this analysis is two-fold: to characterize systematic differences between pre-cancer and normal samples, and to predict disease status using cross-validation.

We applied MFMM to the full dataset that contains 143 pre-cancer samples and 391 normal samples. In the model fitting, we adopted a two-stage transformation—a lossless discrete wavelet transform (with no wavelet compression) followed by PCA of the concatenated wavelet coefficients. We truncated the resulting PC scores by constraining that  $\text{FIR} \geq 1-1\text{e-}6$ . We ran 5000 MCMC iterations and treated the first 3000 iterations as the burn-in period. The convergence of the MCMC samples has been monitored, and more details can be found in Section 7 and the supplementary materials. The running time for 5000 MCMC iterations was around 8 minutes for the full data set. Based upon the posterior samples of  $\underline{B}$ , we calculated the posterior samples of the contrast effect between normal samples and pre-cancer samples by  $\underline{C}^{(g)} = \underline{B}_2^{(g)} - \underline{B}_1^{(g)}$  for



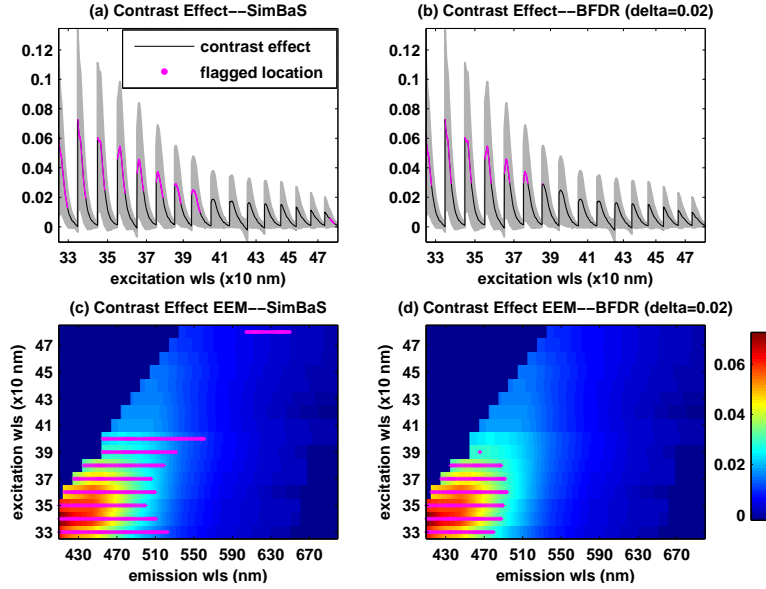


Figure 4: The posterior mean of the contrast effect  $\underline{C}$  plotted as a concatenated curve, flagged by SimBaS (a) and BFDR (b), and the posterior mean of  $\underline{C}$  plotted as an EEM image, flagged by SimBaS (c) and BFDR (d). In (a) and (b), the shaded gray regions are the 95% SCB calculated by adjusting EER across all 16 curves.

$$g = 1, \dots, H.$$

We applied the SimBaS approach to the posterior samples of  $\underline{C}$  to identify significant nonzero regions, and applied the BFDR ( $\delta = 0.02$ ) approach to identify regions that are greater than 0.02, using  $\alpha = 0.05$  as the significance threshold in both cases. From SimBaS, we found the GBPV to be 0.004. This implies that overall, there were some differences between the normal and pre-cancer samples. We demonstrate the posterior mean of  $\underline{C}$  and the flagged regions in Figure 4 using both concatenated curve plots and EEM plots. From Figure 4 (a) and (b), we see that the contrast effect  $\underline{C}$  is mostly positive, indicating that the normal samples tend to have higher intensity than the pre-cancer samples. Overall, the SimBaS and the BFDR approach identified similar regions: both flagged the region around excitations 330 – 380 nms and emissions 410 – 490 nms. Since the SimBaS is targeted to detect nonzero regions, whereas BFDR is targeted to identify regions with  $|\underline{C}| > 0.02$ , the SimBaS flagged more significant

locations than the BFDR. For example, SimBaS flagged regions of curves at excitations 390, 400 and 480 nms; whereas these regions were found to be non-significant by the BFDR.

To assess the performance of MFMM on discriminating disease status, we re-fit the data using a 4-fold cross-validation. Specifically, we randomly split the full dataset into four blocks while reserving the proportion of disease and normal samples within each subset. We then iterated the training–predicting procedure by first training MFMM using three of the four blocks and then predicting the class labels for the remaining block. The resulting posterior predictive probabilities for disease status were treated as prediction scores, with higher values indicating higher chances of belonging to the disease class. Based on these scores, we plotted the empirical ROC curve (Figure 5). The corresponding AUC for this discrimination analysis is 0.65. If we make decision by choosing a threshold to maximize the sum of sensitivity and specificity, we achieve sensitivity of 0.69 and specificity of 0.59, with a misclassification rate of 38%. This discrimination result is close to but not as high as the result reported by Zhu and Cox (2009). A sensible explanation is that this study is based on data collected from a fixed tissue type and from patients with a fixed menopausal status; therefore, the variables of tissue type and menopause statuses were not used as predictors in classification.

The reasons for excluding samples from another tissue type (tissue type columnar) and from patients with other menopausal statuses (peri- and post-menopause) are as follows: (1) No samples were available for columnar tissue for the pre-cancer group, and only three samples were available for patients with per- and post-menopausal status in the pre-cancer group. With too few samples, we can hardly provide a reliable estimate of the tissue or menopausal effects using MFMM. (2) Data from columnar tissue and patients with peri- or post-menopausal status were heavily unbalanced across the normal group and the pre-cancer group. Therefore, the tissue type and menopausal status may confound the disease status in the training procedure, causing spurious classification results. For example, if all samples with columnar tissue type

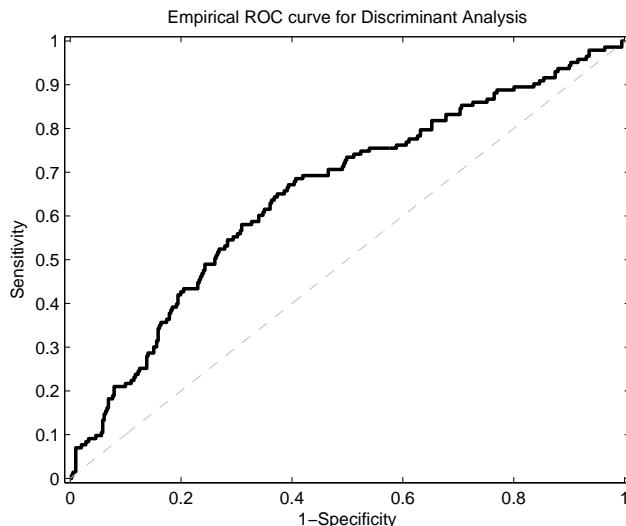


Figure 5: Plot of empirical ROC curve based on 4-fold cross validation.

are from the normal group, purely classifying samples with columnar tissue as normal would provide a better-than-random-guess result. On the other hand, we want to use only disease-related information from the EEM data, rather than information on the unbalanced design, to classify samples. Nevertheless, when the sample size is sufficiently large and samples are balanced across the two classes, including other covariates may improve the overall classification performance.

In summary, our analysis of the fluorescence spectroscopy data revealed a major local region around excitations 330 – 380 nms and emissions 420 – 480 nms on the EEM that reflects differences between the normal and pre-cancer samples. In particular, normal samples tend to have significantly higher intensity than the pre-cancer samples in this region. If desired, we can adjust the  $\delta$  parameter in the BFDR analysis to identify locations with other levels of differences. The flagged regions may serve as biomarkers for future disease assessment and diagnosis.

## 6. Sensitivity analysis

It is of interest to study whether the results presented in Section 5 are sensitive to different model choices or parameter settings. For example, whether using a basis other than wavelets or different prior parameters would affect the posterior estimates and region detection outputs. To answer these questions, we repeated several analyses under different model choices using the real EEM data. These analyses include: (i) **Using cosine basis in the first-stage transformation.** We re-fit MFMM by replacing the wavelets by cosine basis expansion. The cosine basis takes the form  $\{\phi_1(t) = 0, \phi_k = \sqrt{2} \cos((k-1)\pi t), k \geq 2\}$ . (ii) **Using a smaller variance when setting the Inverse-Gamma and Beta priors.** In our original analysis in Section 5, we set parameters for the inverse-Gamma and Beta priors by letting the mode of these distributions be equal to their initial values and letting the variance be a pre-specified large value (i.e.,  $10^3$  for the inverse-Gamma and 0.06 for the Beta priors). In this sensitivity analysis, we adjusted the prior parameters by reducing the prior variance for inverse-Gamma to 10 and reducing the prior variance for Beta to 0.04. (iii) **Using a larger variance when setting the inverse-Gamma and Beta priors.** We repeated analysis (ii) and increased the prior variance for inverse-Gamma to  $10^6$  and increased the prior variance for Beta to 0.08.

The results of the sensitivity analysis are plotted similarly to Figure 4 (a)-(b) for each of the cases and are demonstrated in the supplementary materials. These analyses show that the posterior mean, the 95% SCB, and the regions flagged by the BFDR ( $\delta = 0.02$ ) approach are nearly identical across all analyses. The flagged locations using the SimBaS approach also demonstrate similar patterns, with slight variations appear at a few low-intensity regions on excitation curves at 380-390 and 460-480 nms. This is not surprising because SimBaS aims to flag regions on  $\underline{C}$  that are significantly nonzero, and locations with close-to-zero intensities on a continuous curve are hard to determine. In summary, these sensitivity analyses indicate that our posterior estimation and region detection results are not sensitive to different basis choices and different

parameter settings for the inverse-Gamma and Beta priors.

## 7. Discussion

To jointly model multiple functional variables, we have proposed a MFMM  
650 approach that regresses multivariate functional data on a set of scalar covariates  
while flexibly accounting for potential multi-level data structures. We fit the  
model after performing a convenient two-stage transformation. We provided  
a fully Bayesian approach to obtain the posterior samples in the transformed  
domain and performed posterior inference in the original data space after inverse  
655 transforming the posterior samples. Our approach produces functional tests  
to flag local regions of the EEMs while accounting for multiple testing using  
the FDR or EER criterion. It also facilitates a discrimination analysis of new  
observations.

The proposed Bayesian approach requires running an MCMC algorithm to  
660 obtain posterior samples. As other MCMC-based analyses, the convergence  
of the MCMC samples needs to be monitored and tested. We monitored the  
behavior of the posterior samples by checking the trace plots, the autocorre-  
lation plots, and calculating the effective sample size of the retained samples.  
We tested the convergence of the chains by calculating Geweke's Z-statistics  
665 (Geweke, 1992) based on the samples after burnin. These diagnostics were per-  
formed for both the simulation and the real data analysis. Summary plots and  
more detailed discussions are available in the supplementary materials.

The fluorescence spectroscopy EEM data represent multivariate functional  
data with different supports. Besides the EEM data, the proposed MFMM  
670 can be applied to other multivariate functional data with potentially different  
dimensions and/or interpretations, such as various types of neuroimaging, ge-  
nomic, proteomic and engineering data. For example, we can use MFMM to  
jointly model 1-d EEG signals and 3-d MRIs. When two functional variables  
are of different dimensions, one needs to apply different basis representations  
675 at the first stage of the transformation. For example, 1-d, 2-d, or 3-d wavelets

can be used to represent different functional variables. Other bases, such as splines, Fourier series, or principal components, also have higher dimensional counterparts, and thus can be used just as well. At the second stage, while we have proposed PCA for the purpose of de-correlation and dimension reduction,  
680 one may consider other transformations with similar characteristics.

We have proposed a Bayesian approach for inference. It is possible to apply frequentist approaches to estimate parameters in model (4) by using, for example, the restricted maximum likelihood (REML) and the best linear unbiased estimation prediction (BLUP) methods as described in Searle et al. (1992).  
685 However, the point estimates and their uncertainties are hard to be interpreted in the transformed space. Therefore, if using frequentist approaches, one may need to perform bootstrapping and characterize the uncertainty of the estimated parameters after inverse-transforming the bootstrapped estimates back to the original data domain.

We have assumed Gaussian distributions for the random effects and residuals in the proposed MFMM. Extensions that incorporate heavier tailed distributions can be performed following the work of Zhu et al. (2011). Furthermore, in addition to the multi-level structure, we can incorporate between-function spatial or temporal correlations in the residual terms and/or the fixed effect  
690 coefficient functions, following strategies similar to Zhang et al. (2014) and Zhu et al. (2016b).

Algorithms used in this paper are coded using Matlab and C. In the supplementary material, we have provided Matlab code that calls a C executable for implementing the proposed approach. We demonstrate how to use this code  
700 through an example that runs simulation 1. In case of higher dimensional functions, especially when the number of measurement points is more than  $O(10^4)$ , the scalability of our algorithm can be improved substantially by using both basis coefficient compression and PC truncation. Depending on the characteristics of the functional data and the chosen basis, speed improvements of 20-fold  
705 or more are possible with moderate compression. Finally, for extremely high dimensional datasets such as neuroimaging and genomic data, the independence

assumption across columns of model (4) enables us to perform Bayesian inference in parallel, using either a graphics processing unit (Sanders and Kandrot, 2011) or other parallel programming techniques.

## 710 **Acknowledgments**

The authors thank Lee Ann Chastain for her assistance in editing. Hongxiao Zhu was supported by Institute for Critical Technology and Applied Science, Virginia Tech (ICTAS-JFC 175139) and National Science Foundation (NSF-DMS 1611901). Jeffrey S. Morris was supported by National Cancer Institute (R01-CA107304, R01-CA178744, P30-CA016672) and National Institute 715 of Drug Abuse (R01-DA017073). Dennis D. Cox was supported by National Cancer Institute (PO1-CA082710).

## **References**

- Aue, A., Dubart Norinho, D., Hörmann, S., 2015. On the prediction of stationary functional time series. *Journal of the American Statistical Association* 720 110, 378–392.
- Backenroth, D., Goldsmith, J., Kitago, T., Krakauer, J., 2016. Modeling heterogeneity in motor learning using heteroskedastic functional principal components. abstract, JSM 2016 .
- 725 Baladandayuthapani, V., Mallick, B.K., Hong, M.Y., Lupton, J.R., Turner, N.D., Carroll, R.J., 2008. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* 64, 64–73.
- Castanedo, F., 2013. A review of data fusion techniques. *The Scientific World Journal* 2013.
- 730 Chang, C.Y., Chang, C.C., Hsiao, T., 2013. Fluorescence intrinsic characterization of excitation-emission matrix using multi-dimensional ensemble empirical mode decomposition. *Int. J. Mol. Sci.* 14, 22436–22448.

- Chang, S.K., Follen, M., Malpica, A., Utzinger, U., Staerkel, G., Cox, D., Atkinson, E.N., MacAulay, C., Richards-Kortum, R., 2002. Optimal excitation  
735 wavelengths for discriminantion of cervical neoplasia. *IEEE Transactions on Biomedical Engineering* 49, 1102–1110.
- Crainiceanu, C.M., Staicu, A.M., Ray, S., Punjabi, N., 2012. Bootstrap-based inference on the difference in the means of two correlated functional processes. *Statistics in Medicine* 31, 3223–3240.
- 740 Delicado, P., Giraldo, R., Comas, C., Mateu, J., 2010. Statistics for spatial functional data: some recent contributions. *Environmetrics* 21, 224 – 239.
- Fan, Y., James, G.M., Radchenko, P., 2015. Functional additive regression. *Ann. Statist.* 43, 2296–2325.
- Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the  
745 calculation of posterior moments. *Bayesian Statistics* 4 .
- Giraldo, R., Delicado, P., Mateu, J., 2010. Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics* 15, 66–82.
- Goldsmith, J., Crainiceanu, C.M., Caffo, B., Reich, D., 2012. Longitudinal  
750 penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc. Ser. C Appl. Stat.* 61, 453–469.
- Goldsmith, J., Zipunnikov, V., Schrack, J., 2015. Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics* 71, 344–353.
- 755 Greven, S., Crainiceanu, C., Caffo, B., Reich, D., 2010. Longitudinal functional principal component analysis. *Electronic Journal of Statistics* 4, 1022–1054.
- Hörmann, S., Kokoszka, P., 2010. Weakly dependent functional data. *ArXiv e-prints* [arXiv:1010.0792](https://arxiv.org/abs/1010.0792).



- James, G.M., 2002. Generalized linear models with functional predictors. *Journal of the Royal Statistical Society, Series B* 64, 411–432.
- 760
- Lee, J.S., Shuhatovich, O., Price, R., Pikkula, B., Follen, M., McKinnon, N., MacAulay, C., Knight, B., Richards-Kortum, R., Cox, D.D., 2005. Design and preliminary analysis of a study to assess intra-device and inter-device variability of fluorescence spectroscopy instruments for detecting cervical neoplasia. *Gynecologic Oncology* 99, S98 – S111.
- 765
- Marín, N.M., MacKinnon, N., MacAulay, C., Chang, S.K., Atkinson, E.N., Cox, D., Serachitopol, D., Pikkula, B., Follen, M., Richards-Kortum, R., 2006. Calibration standards for multicenter clinical trials of fluorescence spectroscopy for in vivo diagnosis. *Journal of Biomedical Optics* 11, 014010.
- 770
- Meyer, M.J., Coull, B.A., Versace, F., Cinciripini, P., Morris, J.S., 2015. Bayesian function-on-function regression for multilevel functional data. *Biometrics* doi:10.1111/biom.1299.
- Morris, J.S., 2015. Functional Regression. *Annual Review of Statistics and Its Application* 2, 321–359.
- 775
- Morris, J.S., Brown, P.J., Herrick, R.C., Baggerly, K.A., Coombes, K.R., 2008. Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. *Biometrics* 64, 479–489.
- Morris, J.S., Carroll, R.J., 2006. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* 68, 179–199.
- 780
- Newton, M.A., Noueiry, A., Sarkar, D., Ahlquist, P., 2004. Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* 5, 155–176.
- Prato, G.D., 2006. *An Introduction to Infinite-Dimensional Analysis*. Springer, New York.

- 785 Qiao, X., James, G., Lv, J., 2015. Functional graphical models. Technical Report. University of Southern California.
- Ramanujam, N., Mitchell, M., Mahadevan, A., Thomsen, S., Malpica, A., Wright, T., Atkinson, N., Richards-Kortum, R., 1996. Spectroscopic diagnosis of cervical intraepithelial neoplasia(cin) in vivo using laser induced fluorescence spectra at multiple excitation wavelengths. *Lasers Surg. Med.* 19, 790 63–67.
- Ramsay, J.O., Silverman, B.W., 1997. *Functional Data Analysis*. Springer-Verlag, New York.
- Reiss, P.T., Ogden, R.T., 2007. Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association* 102, 795
- Reiss, P.T., Ogden, R.T., 2010. Functional generalized linear models with images as predictors. *Biometrics* 1, 61–69.
- Ruppert, D., Wand, M.P., Carroll, R.J., 2003. *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, UK.
- Sanders, J., Kandrot, E., 2011. *CUDA by Example, An Introduction to General-Purpose GPU Programming*. Addison-Wesley, Boston, MA.
- Scheipl, F., Staicu, A.M., Greven, S., 2014. Functional additive mixed models. 805 *Journal of Computational and Graphical Statistics* 24, 477–501.
- Searle, S.R., Casella, G., McCulloch, C.E. (Eds.), 1992. *Variance Components*. John Wiley & Sons, New York.
- Staicu, A., Crainiceanu, C.M., Carroll, R.J., 2010. Fast methods for spatially correlated multilevel functional data. *Biostatistics* 11, 177–194.

- 810 Uludag, K., Roebroek, A., 2014. General overview on the merits of multimodal  
neuroimaging data fusion. *NeuroImage* 102, Part 1, 3 – 10. *Multimodal Data  
Fusion*.
- Yamal, J.M., Zewdie, G.A., Cox, D.D., Atkinson, N.E., Cantor, S.B., MacAulay,  
C. and Follen, M., 2012. Accuracy of optical spectroscopy for the detection  
815 of cervical intraepithelial neoplasia without colposcopic tissue information;  
a step toward automation for low resource settings. *Journal of Biomedical  
Optics* 17.
- Yang, J., Zhu, H., Choi, T., Cox, D.D., 2016. Smoothing and meancovariance  
estimation of functional data with a bayesian hierarchical model. *Bayesian*  
820 *Anal.* 3, 649–670.
- Yao, F., 2007. Asymptotic distributions of nonparametric regression estimators  
for longitudinal or functional data. *Journal of Multivariate Analysis* 98, 40–56.
- Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., Cichocki, A., 2014. Ag-  
gregation of sparse linear discriminant analyses for event-related potential  
825 classification in brain-computer interface. *Int. J. Neur. Syst.* 24.
- Zhou, L., Huang, J.Z., Martinez, J.G., Maity, A., Baladandayuthapani, V.,  
Carroll, R.J., 2010. Reduced rank mixed effects models for spatially correlated  
hierarchical functional data. *Journal of the American Statistical Association*  
105, 390–400.
- 830 Zhu, H., Brown, P.J., Morris, J.S., 2011. Robust, adaptive functional regression  
in functional mixed model framework. *J. Am. Statist. Ass.* 495, 1167–1179.
- Zhu, H., Cox, D.D., 2009. A functional generalized linear model with curve  
selection in cervical pre-cancer diagnosis using fluorescence spectroscopy, in:  
Rojo, J. (Ed.), *Optimality: The Third Erich L. Lehmann Symposium*, Insti-  
835 tute of Mathematical Statistics, Beachwood, Ohio, USA. pp. 173–189.
- Zhu, H., Strawn, N., Dunson, D.B., 2016a. Bayesian graphical models for mul-  
tivariate functional data. *Journal of Machine Learning Research* 17, 1–27.

- Zhu, H., Vannucci, M., Cox, D.D., 2010. A Bayesian hierarchical model for classification with selection of functional predictors. *Biometrics* 66, 463–473.
- <sup>840</sup> Zhu, H., Versace, F., Cinciripini, P.M., Morris, J.S., 2016b. Robust spatial functional regression models for analysis of event-related potentials. Manuscript