Ultra-Fast Approximate Inference Using Variational Functional Mixed Models

Shuning Huo
Department of Statistics, Virginia Tech
and
Jeffrey S Morris
Department of Biostatistics, Epidemiology and Informatics,
Department of Statistics, University of Pennsylvania
and
Hongxiao Zhu
Department of Statistics, Virginia Tech

July 25, 2022

Abstract

While Bayesian functional mixed models have been shown effective to model functional data with various complex structures, their application to extremely high-dimensional data is limited due to computational challenges involved in posterior sampling. We introduce a new computational framework that enables ultra-fast approximate inference for high-dimensional data in functional form. This framework adopts parsimonious basis to represent functional observations, which facilitates efficient compression and parallel computing in basis space. Instead of performing expensive Markov chain Monte Carlo sampling, we approximate the posterior distribution using variational Bayes and adopt a fast iterative algorithm to estimate parameters of the approximate distribution. Our approach facilitates a fast multiple testing procedure in basis space, which can be used to identify significant local regions that reflect differences across groups of samples. We perform two simulation studies to assess the performance of approximate inference, and demonstrate applications of the proposed approach by using a proteomic mass spectrometry dataset and a brain imaging dataset. Supplementary materials are available online.

Keywords: Approximate Bayesian Inference, Functional Data Analysis, Variational Bayes, Distributed Inference, Parallel Computing

1 Introduction

Modern high-throughput technologies enable the collection of a special type of high-dimensional data called *functional data*. The ideal observational units of functional data are curves or surfaces defined on some continuous domain and sampled on a discrete grid (Ferraty and Romain, 2018). Typical examples include longitudinal measurements, spectral curves, engineering signals, brain images, and many other digital measurements. While functional data often provide a rich source of information, they also pose extraordinary challenges to statistical methodology, mostly due to the high dimensionality, high volume, and complex data structures. It is not just desirable, but essential that analytical tools should be scalable to the increasing size or dimensionality and flexible to accommodate complex data structures. One motivating example is a brain imaging dataset which will be considered in our real data analysis. The data consist of 3-D brain images with about 10 million pixels per subject, making knowledge discovery challenging.

Extensive research work has been done in the field of *functional data analysis* to process and model functional data, among which the most studied area is functional data regression. Functional data regression focuses on characterizing the relationship between functional observations and other functional or non-functional variables. Notable work includes functional linear models (Cardot et al., 1999; Cuevas et al., 2002; Yao et al., 2005; Hall et al., 2007; Yuan and Cai, 2010), generalized functional linear models (James, 2002; Müller and Stadtmüller, 2005; Zhu and Cox, 2009), functional additive models (Fan et al., 2015), etc. Comprehensive reviews can be found in Ramsay and Silverman (2005), Morris (2015) and Wang et al. (2016).

Among existing models, functional mixed models (FMMs) provide a flexible functional response regression framework that can accommodate numerous complex structures induced by the experimental design. Compared with the frequentist counterpart (Guo, 2002), Bayesian functional mixed models (Morris and Carroll, 2006; Zhou et al., 2010) have achieved a great deal of success due to several advantages. First, inference through Markov chain Monte Carlo (MCMC) sampling is convenient; second, the variability from estimating nuisance parameters or any other parameters in the model is propagated through

the inference, which means that the final error bands account for all sources of variability. Indeed, Bayesian FMMs have been widely applied to different applications including the analysis of mass spectrometry data (Morris et al., 2008, 2011), accelerometer data (Morris et al., 2006), acoustic signals (Martinez et al., 2013), and medical images (Lancia et al., 2015; Zhu et al., 2018; Lee et al., 2019). Further extensions of Bayesian FMM have also been made to allow robust regression (Zhu et al., 2011, 2012) and function-on-function regression (Meyer et al., 2015), to incorporate spatial correlations (Zhang et al., 2016; Zhu et al., 2018) and multivariate functional responses (Zhu et al., 2017), to accommodate nonlinear effects for scalar predictors (Lee et al., 2019), and to perform quantile functional regression on distributions estimated from subject-specific data streams (Yang et al., 2020). These methods constitute a suite of Bayesian FMM-based toolbox (Morris, 2017) that cover a large scope of analysis involving high-dimensional, complex structured functional data. Its use of general basis functions in a *basis transform modeling approach* make it well suited for complex, high dimensional functional data with many observational points per function. By basis transformation, we can effectively perform inference in basis space and transform results back to data space for visualization and interpretation. This strategy brings numerous conveniences such as enabling parallel modeling and dimension reduction, and for suitably chosen basis, has the flexibility to capture global or local features of complex data.

Despite their effectiveness, Bayesian FMMs become computationally demanding for data with extraordinary high volume and dimensionality. The computational challenges primarily come from running Markov chain Monte Carlo (MCMC) sampling and the need to store a large number of posterior samples. Various strategies have been suggested to improve the computation scalability, for example, performing data compression in advance to reduce dimension (Morris et al., 2011) or calculating good MCMC initial values for MCMC (Zhu et al., 2012). Even with these strategies, the computation can still be a challenging task for large scale data as it still requires running MCMC until convergence. In this paper, we aim to develop an ultra-fast Bayesian FMM computational framework that is suitable for large scale functional data, avoiding expensive MCMC sampling and the hassle of storing posterior samples.

Since large scale functional data are a special type of Big Data, it is natural to hope that strategies for Big Data computation to be useful. For example, techniques such as divide-and-conquer (Wang et al., 2016) lead to embarrassingly parallel algorithms (Wikipedia contributors, 2019), and approximation is often effective to improve computation efficiency (Sarlos, 2006; Minka, 2001; Ruli et al., 2016). Despite the progress in Big Data computation, the fast computation of large-scale functional data has not received much attention.

In this paper, we propose a variational functional mixed model (VFMM) framework that takes advantages of several attractive computational strategies in Big Data computation. In particular, we adopt the divide-and-conquer strategy by first representing functional observations by parsimonious basis and then performing statistical inference for each component in basis space. The parsimonious basis representation enables efficient compression and embarrassingly parallel computation. It also facilitates an efficient multiple testing procedure in basis space which can be used to identify significant local regions in the original data domain. Instead of performing MCMC sampling, we rely on approximate inference (Sun, 2013). In particular, we approximate the posterior distribution by using variational Bayes, a method from machine learning that approximates posterior distributions through optimization (Blei et al., 2017). While this approximation sacrifices some of MCMC's accuracy, it provides large gains in terms of computational feasibility, especially in ultra-high dimensional settings. We design a fast iterative algorithm to estimate parameters of the approximated posterior distribution. Owing to its fast speed, our approach is ideal for obtaining quick initial estimate based on large scale data. If desired, it can be combined with full MCMC schemes to achieve more accurate Bayesian inference to the exact posterior distribution. We note that variational Bayes has already been applied in various regression setups (Faes et al., 2011; Ormerod and Wand, 2012; Ormerod et al., 2017; Luts et al., 2014; Luts and Wand, 2015; Hui et al., 2019; Zhang et al., 2019; Ray and Szabó, 2021) and other applications (Serra et al., 2019). In functional data analysis, it has also been successfully applied to registration (Earls and Hooker, 2017b,a) and scalar-on-function regressions including functional linear models (Goldsmith et al., 2011) and functional generalized additive models (McLean et al., 2014, 2017). To our knowledge, the current paper is the first

that deals with function-on-scalar regression with variational inference.

Relative to existing functional regression approaches, our proposed VFMM approach brings several advantages. (1) It enables distributed inference thus is scalable to large-scale functional data; (2) it avoids the hassle of running MCMC and storing posterior samples; (3) it facilitates the detection of significant local regions, and these regions can be visualized directly in the original data domain; (4) it is directly applicable to FMM with different basis choices as well as several of extensions of FMM, including function-on-function regression (Meyer et al., 2015), FMM for multivariate functional responses (Zhu et al., 2017), quantile functional regression (Yang et al., 2020), and semiparametric FMM that accommodates nonlinear covariate effects (Lee et al., 2019). Our results for the simulated and real data demonstrate the effectiveness of the proposed VFMM in estimating parameters, detecting interesting regions, and saving computation time and storage space.

The outline for the rest of this paper is as follows. In Section 2, we start by reviewing variational Bayes in Section 2.1, and introduce the proposed VFMMs in Section 2.2. An approach to detect significant regions by performing basis space tests will be discussed in Section 2.3. Estimation results and computational gains of VFMM are demonstrated by simulations in Section 3 using both curves and three-dimensional brain images. Two case studies are used to demonstrate the effectiveness of VFMM in Section 4. We provide a final discussion in Section 5.

2 Variational Functional Mixed Models

2.1 An Overview of Variational Bayes

While MCMC sampling provides a standard way to estimate parameters of FMM, its computation can be slow and its convergence can be difficult to diagnose especially in the context of large-scale problems. We aim to develop a computational framework for FMM that is faster and easier to scale-up to large datasets. The idea is to find a closed-form approximation to the posterior distribution by using variational Bayes. Here, we briefly review the basic idea of variational Bayes. A comprehensive review can be can be found

in Blei et al. (2017). Let θ denote all model parameters and y denote the observed data. Unlike MCMC sampling which provides a stochastic approximation of the exact posterior $p(\theta|y)$ using a set of samples, variational Bayes finds an analytical proxy $q_v(\theta)$ that is closest to $p(\theta|y)$. In particular, one estimates the parameters v of $q_v(\theta)$ in order to minimize the Kullback-Leibler divergence between $q_v(\theta)$ and $p(\theta|y)$,

$$\mathrm{KL}(q||p) = \int q_v(\theta) \log \frac{q_v(\theta)}{p(\theta|y)} d\theta.$$

Directly minimizing $\mathrm{KL}(q||p)$ is often difficult. Fortunately, one can decompose $\log p(y)$ as

$$\begin{aligned} \log p(y) &= \int q_v(\theta) \log \frac{q_v(\theta)}{p(\theta|y)} d\theta + \int q_v(\theta) \log \frac{p(y,\theta)}{q_v(\theta)} d\theta. \\ &= \mathrm{KL}(q||p) + E_{q_v} \left\{ \log \frac{p(\theta, y)}{q_v(\theta)} \right\}. \end{aligned} \tag{1}$$

As $\log p(y)$ is a constant, minimizing $\mathrm{KL}(q||p)$ is equivalent to maximizing the second term in (1), which is referred to as the *evidence lower bound* (ELBO). Thus, inference with variational Bayes boils down to solving to a optimization problem. Often, $q_v(\theta)$ is restricted to take a simpler form than $p(\theta|y)$. One common restriction is through the mean-field assumption, i.e., assuming that θ can be partitioned into independent blocks, therefore the factorization $q_v(\theta) = \prod_i q_{v_i}(\theta_i)$ holds. Such simplification makes it possible to calculate $q_v(\theta)$ analytically. Additionally, more convenient calculations can be induced with assumptions about exponential family, i.e., assuming that (i) each conditional distribution $p(\theta_i \mid \theta_{(-i)}, y)$ belongs to the exponential family, and (ii) the approximate distribution $q_{v_i}(\theta_i)$ belongs to the same exponential family as $p(\theta_i \mid \theta_{(-i)}, y)$. In particular, write $p(\theta_i \mid \theta_{(-i)}, y)$ in canonical form $p(\theta_i \mid \theta_{(-i)}, y) = \exp\left\{\eta_i(\theta_{(-i)}, y)^T t(\theta_i) - A(\eta_i)\right\}$, where $\eta_i(\theta_{(-i)}, y)$ is the natural parameter, then $q_{v_i}(\theta_i)$ belongs to the same exponential family with natural parameter $\eta(v_i) = E_{q_{\setminus i}}\left\{\eta_i(\theta_{(-i)}, y)\right\}$, where $q_{\setminus i} = \prod_{j \neq i} q_{v_j}(\theta_j)$. Detailed derivations can be found in Appendix A of Blei (2006). The exponential family assumptions transfer the estimation of $q_{v_i}(\theta_i)$ to the estimation of natural parameters, making the calculations much more straightforward.

2.2 Functional Mixed Models with Variational Bayes

We consider the Bayesian Functional Mixed Model (FMM) framework of Morris and Carroll (2006) and demonstrate a variational Bayes approach for approximate inference. Let $\boldsymbol{Y}(t) = (Y_1(t), \ldots, Y_N(t))^T$ denote a vector of functional data responses. The FMM takes the form

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

where $\boldsymbol{B}(t) = (B_1(t), \ldots, B_p(t))^T$ is a vector of fixed effect coefficient functions associated with a $N \times p$ design matrix $\mathbf{X}$, $\boldsymbol{U}(t) = (U_1(t), \ldots, U_M(t))^T$ is a vector of random effect coefficient functions associated with a $N \times M$ design matrix $\mathbf{Z}$, and $\boldsymbol{E}(t) = (E_1(t), \ldots, E_N(t))^T$ is the vector of random errors. Morris and Carroll (2006) assumed that both $\boldsymbol{U}(t)$ and $\boldsymbol{E}(t)$ are mutually independent multivariate Gaussian processes. In particular, they used $\boldsymbol{E}(t) \sim \mathcal{N}(\mathbf{R}, S)$ to denote that $\boldsymbol{E}(t)$ is a multivariate mean-zero Gaussian process with a $N \times N$ between-function covariance matrix $\mathbf{R}$ and a within-function covariance surface $S(\cdot, \cdot)$, thus $\text{cov}\{E_i(t_1), E_{i'}(t_2)\} = \mathbf{R}_{i,i'}S(t_1, t_2)$, for $i, i' \in \{1, \ldots, N\}$ and $t_1, t_2 \in \mathcal{T}$. Similarly, they assumed that $\boldsymbol{U}(t) \sim \mathcal{N}(\mathbf{P}, Q)$.

Following a similar idea of Morris and Carroll (2006), we adopt a parsimonious basis representation to transform the FMM to basis space, which leads to divide-and-conquer computation. Consider the general FMM in (2), we assume that the responses $\{Y_i(t), i = 1, \ldots, N\}$ take values in $L^2(\mathcal{T})$, where $\mathcal{T}$ is a closed subset of $\mathbb{R}^d$, $d \geq 1$. Let $\{\phi_j\}_{j=1}^{\infty}$ denote a compactly supported, orthonormal basis of $L^2(\mathcal{T})$. We can expand $Y_i(t)$ by $Y_i(t) = \sum_{j=1}^{\infty} d_{ij}\phi_j(t)$ where $d_{ij} = \langle Y_i, \phi_j \rangle = \int_T Y_i(t)\phi_j(t)dt$. The coefficient sequence $(d_{i1}, d_{i2}, \ldots)$ lies in the space of square-summable sequences, denoted by $\ell^2 = \left\{ d_j : \sum_{j=1}^{\infty} d_j^2 < \infty \right\}$. Note that Morris and Carroll (2006) have used wavelet basis, in which case the basis functions $\{\phi_j\}$ have been indexed by (j, k) where j denotes the resolution level and k denotes location. In this paper, we use a single index j to index all basis functions for simplicity. Since $\mathbf{Y}(t)$ is written as the linear combination of $\mathbf{B}(t)$, $\mathbf{U}(t)$ and $\mathbf{E}(t)$, it is natural to assume that these unobserved functional components also take values in the same $L^2(\mathcal{T})$ space. With this assumption, all functional objects in FMM can be represented by a common basis. Specifically, denote $\boldsymbol{\Phi} = (\phi_1, \phi_2, \ldots)^T$, then all basis expansions can be represented by

linear operations: $\mathbf{Y} = \mathbf{D\Phi}$, $\mathbf{B} = \mathbf{B}^*\mathbf{\Phi}$, $\mathbf{U} = \mathbf{U}^*\mathbf{\Phi}$, and $\mathbf{E} = \mathbf{E}^*\mathbf{\Phi}$. Model (2) becomes $\mathbf{D\Phi} = \mathbf{XB}^*\mathbf{\Phi} + \mathbf{ZU}^*\mathbf{\Phi} + \mathbf{E}^*\mathbf{\Phi}$. Since $\mathbf{\Phi}$ preserves linear operation, the above model is equivalent to the dual space model

$$\mathbf{D} = \mathbf{XB}^* + \mathbf{ZU}^* + \mathbf{E}^*, \tag{3}$$

where rows of $\mathbf{D}, \mathbf{B}^*, \mathbf{U}^*$ and $\mathbf{E}^*$ contain wavelet coefficients of entries in $\mathbf{Y}(t), \mathbf{B}(t), \mathbf{U}(t)$ and $\mathbf{E}(t)$, respectively. This transforms FMM from the functional space $L^2(\mathcal{T})$ to the dual space ℓ^2.

In model (3), the random effects $\mathbf{U}^*$ and the error $\mathbf{E}^*$ are both zero-mean normal matrices, denoted by $\mathbf{U}^* \sim \mathrm{N}(\mathbf{P}, \mathbf{Q}^*)$, $\mathbf{E}^* \sim \mathrm{N}(\mathbf{R}, \mathbf{S}^*)$, where $\mathbf{Q}^*, \mathbf{S}^*$ denote covariances between columns of $\mathbf{U}^*$ and $\mathbf{E}^*$, respectively. For many choices of basis such as wavelets, correlations between basis coefficients are substantially reduced (Fan, 2003). This enables a simplified independence assumption for the covariance matrices $\mathbf{Q}^*$ and $\mathbf{S}^*$, i.e., $\mathbf{Q}^* = \mathrm{diag}(\{q_j^*\})$, $\mathbf{S}^* = \mathrm{diag}(\{s_j^*\})$. This further divides the dual space model (3) into many independent vector-response mixed effect models. We denote the jth model by

$$\mathbf{d}_j = \mathbf{Xb}_j^* + \mathbf{Zu}_j^* + \mathbf{e}_j^*, \tag{4}$$

where $\mathbf{d}_j$, $\mathbf{b}_j^*$, $\mathbf{u}_j^*$ and $\mathbf{e}_j^*$ denote the jth columns of $\mathbf{D}, \mathbf{B}^*, \mathbf{U}^*$ and $\mathbf{E}^*$ respectively. While the above divide-and-conquer strategy is suitable if using any orthonormal basis, we mainly focus on compactly supported orthonormal basis such as Haar wavelets, Daubechies Wavelets, and spherical wavelets. These bases have the ability to capture local features of functional data, enable parsimonious representation which allows further compression, and have discrete transformation operators that facilitate fast computation. They are generally applicable to curves, images, surfaces, etc. As we will explain in Section 2.3, using compactly supported orthonormal basis allows us to identify interesting local regions by performing multiple testing in basis space, avoiding the need of drawing posterior samples and inverse-transforming posterior samples back to the data domain.

Consider the jth model in (4), we denote $\mathbf{d}_j = (d_{1,j}, \ldots, d_{N,j})^T$, $\mathbf{b}_j^* = (b_{1,j}^*, \ldots, b_{p,j}^*)^T$, $\mathbf{u}_j^* = (u_{1,j}^*, \ldots, u_{m,j}^*)^T$, and $\mathbf{e}_j^* = (e_{1,j}^*, \ldots, e_{N,j}^*)^T$. In FMM, Morris and Carroll (2006) set a

"spike-slab" prior to each component of the fixed effect $\mathbf{b}_j^*$, i.e.,

$$b_{i,j}^* \sim \gamma_{i,j}^* N(0, \tau_{i,j}) + (1 - \gamma_{i,j}^*)\delta_0, \quad \gamma_{i,j}^* \sim \text{Bernoulli}(\pi_{i,j}), \tag{5}$$

where δ_0 is a point mass at 0. The spike-slab prior (Ishwaran and Rao, 2005) encourages sparsity in basis space, which induces adaptive regularization of the corresponding regression coefficient in function space (Morris, 2015). They further assumed $\mathbf{P}$, $\mathbf{R}$ to be identity matrices and set inverse-Gamma priors to the random effect and residual variances $\{q_j^*\}$, $\{s_j^*\}$. This leads to posterior inference via Markov chain Monte Carlo (MCMC) sampling.

To enable efficient variational Bayes computation, we slightly modify the priors in (5) and random effect/residual distributions of FMM by setting

$$\begin{aligned} b_{i,j}^* \sim \gamma_{i,j}^* N(0, q_j^* \tau_{i,j}) + (1 - \gamma_{i,j}^*)\delta_0, \quad \gamma_{i,j}^* \sim \text{Bernoulli}(\pi_j), \\ q_j^* \sim IG(a_j, b_j), \quad \mathbf{u}_j^* \sim N(0, q_j^* \mathbf{I}), \quad \mathbf{e}_j^* \sim N(0, q_j^* \zeta_j \mathbf{I}). \end{aligned}$$

Here, we have factored out the random effect variance q_j^* from the prior variance of fixed effect $b_{i,j}^*$ and the residual variance. This new parameterization allows for convenient update of the approximate distribution of q_j^*; details can be found in the supplementary materials. Based on the above model setup, the joint posterior distribution of $\{b_{i,j}^*\}, \{\gamma_{i,j}^*\}$ and q_j^* can be written as

$$\begin{aligned} & p(\{b_{i,j}^*\}, \{\gamma_{i,j}^*\}, q_j^* \mid \mathbf{d}_j, \zeta_j, \tau_{i,j}, \pi_j) \\ \propto \;& p(\mathbf{d}_j \mid \{b_{i,j}^*\}, q_j^*, \zeta_j)\; p(q_j^*) \prod_{i=1}^{p} p(b_{i,j}^* \mid \gamma_{i,j}^*, \tau_{i,j})\, p(\gamma_{i,j}^* \mid \pi_j) \end{aligned}$$

We treat $\pi_j, \tau_{i,j}, \zeta_j$, and (a_j, b_j) as hyperparameters and make mean-field assumptions for the approximate distributions of $\{b_{i,j}^*\}$,$\{\gamma_{i,j}^*\}$, and q_j^*. This enables an efficient variational EM algorithm. In particular, we assume that the approximate distribution can be factored as follows:

$$q(\{b_{i,j}^*\}, \{\gamma_{i,j}^*\}, q_j^*) = q(q_j^*) \prod_{i=1}^{p} q(b_{i,j}^* \mid \gamma_{i,j}^*) q(\gamma_{i,j}^*). \tag{6}$$

As the conditional posterior of $\{b_{i,j}^*\}$,$\{\gamma_{i,j}^*\}$, and q_j^* all fall in the exponential family, we follow Blei (2006) by assuming that each factor in the above approximate distribution also

falls in the same exponential family. Therefore, in the E-step, the estimation of the approximate distribution boils down to the estimation of the natural parameters in exponential family. This facilitates fast calculation of the approximate distributions. Specifically, we get $q(\gamma^*_{i,j} = 1) = \widetilde{\pi}_{i,j}$, $q(b^*_{i,j} \mid \gamma^*_{i,j} = 1)$ is the density of $\mathcal{N}(\widetilde{\mu}_{i,j},\ \widetilde{\sigma}^2_{i,j})$, and $q(q^*_j)$ is the density of $IG(\widetilde{a}_j,\ \widetilde{b}_j)$, where $\widetilde{\pi}_{i,j}, (\widetilde{\mu}_{i,j},\ \widetilde{\sigma}^2_{i,j})$ and $(\widetilde{a}_j,\ \widetilde{b}_j)$ are variational parameters whose formulae are derived in supplementary materials.

In the M-step, conditional on the estimation of the approximate distributions, we update values of the hyperparameters $\pi_j, \tau_{i,j}$ and ζ_j by directly maximizing the ELBO. Specifically, we are able to analytically solve the values of π_j and $\tau_{i,j}$ by setting the first derivative of ELBO to zero. The value of ζ_j, however, needs to be searched by using an optimization algorithm. In general, the calculation of ELBO involves computing the determinant and inverse of an N by N covariance matrice, which can be slow. However, as shown in our derivation in supplementary materials, if the design matrix $\mathbf{Z}$ is blockwise, we can speed up the calculation of ELBO substantially. The values of the hyperparameters (a_j, b_j) are determined by matching the mean of the inverse-Gamma prior with the initial estimate of q^*_j while setting the prior variance to be fairly large (e.g., 10^3).

We list steps of the VFMM algorithm in Algorithm 1. To ensure fast convergence, we adopt Henderson's mixed model equations (Searle et al., 1992, pages 275-286) to initialize parameters. More technical details about Algorithm 1 and the initialization algorithm are available supplementary materials. While options for more parameter settings and detailed tuning are available, the only required inputs are the observed data $\mathbf{Y}$ and the design matrices $\mathbf{X}$, $\mathbf{Z}$. As mentioned earlier, the FMM (and VFMM) framework is based on a simplified independence assumption in basis space, which divides the basis space model into many independent vector-response mixed effect models. This strategy make it possible to enjoy the computational benefit of parallelization while still accommodate autocorrelation within the function. For this reason, in VFMM, calculation are performed independently across the index j, and Algorithm 1 can either be performed by using vector-wise calculation or be distributed to multi-core computational units. Vector-wise calculation is suitable for small to median scale calculations, for example, the basis space model with dimension less

than $O(10^4)$. For higher dimensions, one may split the basis space model (3) across the index j (i.e., across the columns) into a few sub-models and perform vector-wise calculation in parallel. This is indeed what we did in the real data application on 3-D brain images.

Algorithm 1: The VFMM algorithm. The formulae (S1)-(S8) for calculating $\widetilde{\pi}_{i,j}, (\widetilde{\mu}_{i,j}, \widetilde{\sigma}^2_{i,j})$, $(\widetilde{a}_j, \widetilde{b}_j)$ and $ELBO$ are in section 1 of supplementary materials.

Initialize all parameters;
while $ELBO^{(t)} - ELBO^{(t-1)} > tolerance$ **do**
  for *all* j **do**
    Update $q(\gamma^*_{i,j} = 1) = \text{Bernoulli}(\widetilde{\pi}_{i,j})$ based on log odds in (S1);
    Update $q(b^*_{i,j}|\gamma^*_{i,j} = 1) = \text{N}(\widetilde{\mu}_{i,j}, \widetilde{\sigma}^2_{i,j})$ following (S2)-(S3); $q(b^*_{i,j}|\gamma^*_{i,j} = 0) = \delta_0$;
    Update $q(q^*_j) = IG(\widetilde{a}_j, \widetilde{b}_j)$ following (S4)-(S5);
    Update $ELBO^{(t)}$ following (S6)-(S8);
    Update $\pi_j = \sum_{i=1}^{p} \widetilde{\pi}_{i,j}/p$ and $\tau_{i,j} = \widetilde{a}_j(\widetilde{\sigma}^2_{i,j} + \widetilde{\mu}^2_{i,j})/\widetilde{b}_j$;
    Update ζ_j by hill-climbing heuristic research;
    Update $ELBO^{(t)}$ again following (S6)-(S8).
  end
end

2.3 Region Detection via Basis Space Testing

In addition to estimating unknown parameters, another key inferential objective is to identify local regions that reflect significant differences across groups of samples. In Bayesian FMM, since posterior samples are available, detection of local regions can be achieved by controlling family-wise error rate or Bayesian expected false discovery rate across a grid of $\mathcal{T}$ in data domain as done by Morris et al. (2008) and Meyer et al. (2015). In the VFMM framework, since we are targeting at avoiding posterior sampling in order to improve computational efficiency, we propose a new basis-space testing strategy to detect local regions.

Let $C(t)$ denote a contrast effect of interest. For example, $C(t) = B_1(t) - B_2(t)$ represents the contrast effect between groups 1 and 2 if $B_1(t)$ and $B_2(t)$ are the respective

group means. We focus on detecting regions on which $|C(t)| > \delta$ for a prespecified threshold δ. To achieve this, we assume that $C(t)$ can be represented by the basis expansion $C(t) = \sum_{j=1}^{\infty} c_j \phi_j(t)$. Instead of performing multiple testing on a grid of $\mathcal{T}$, we aim to obtain a region identifier $\widetilde{C}(t) = \sum_{j=1}^{\infty} c_j 1_{\{|c_j|>\epsilon\}} \phi_j(t)$, where ϵ is a small positive threshold. We see that $\widetilde{C}(t) \to C(t)$, as $\epsilon \to 0$ for all t. The component $c_j 1_{\{|c_j|>\epsilon\}}$ represents significantly nonzero component. When using a compactly supported basis, we expect a subset of $\{c_j\}$ to be zero (i.e., $\{H_{0j} : c_j = 0\}$ holds for some j) if $C(t)$ contains zero regions. Therefore, we will identify nonzero components in basis space by performing a sequence of basis space testing $\{H_{0j} : |c_j| > \epsilon$ vs. $H_{aj} : |c_j| \leq \epsilon \ \ ; j = 1, \ldots, K\}$ while controlling the Bayesian false discovery rate across all j. Here, we have assumed that there exists a finite truncation at K such that all H_{0j} will not be rejected beyond K. This is a reasonable assumption as for most smooth functions, higher K corresponds to high frequency components which are usually predominantly noise. After all, numerical calculation has to be performed in a finite dimensional manner even though the theoretical representation is infinite. The parameter K can be determined by performing a truncation to the basis expansion by controlling the percentage of total energy retained. We propose a testing procedure which consists of the following steps:

1. For each component $j = 1, \ldots, K$ in basis space, estimate a probability discovery function $p_\epsilon(j) = \text{pr}\{|c_j| > \epsilon \,|\, \text{Data}\}$ by $\widehat{p}_\epsilon(j)$ based on the approximate distributions $q(b^*_{i,j}|\gamma^*_{i,j})$. Specifically, since $q(b^*_{i,j}|\gamma^*_{i,j} = 1)$ is normal, the approximate distribution of c_j can be fully derived, based on which $\widehat{p}_\epsilon(j)$ can be calculated.

2. Sort $\{\widehat{p}_\epsilon(j)\}$ in descending order to obtain the order statistics $\{\widehat{p}_{\epsilon,(l)}, l = 1, \ldots, K\}$.

3. Set $\phi_\alpha = \widehat{p}_{\epsilon,(s)}$, where $s = \max\{l^* : (l^*)^{-1} \sum_{l=1}^{l^*} \{1 - \widehat{p}_{\epsilon,(l)}\} \leq \alpha\}$.

4. Suppose $J \subset \{1 : K\}$ are the indices that are significant. We reconstruct $\widetilde{C}(t)$ by $\widetilde{C}^\dagger(t) = \sum_{j \in J} c_j \phi_j(t)$ and flag regions on which $|\widetilde{C}^\dagger(t)| > \delta$.

The above method of determining the Bayesian expected FDR threshold ϕ_α is sketched in a diagram in Figure 1, where the solid decreasing line denotes the ordered $\widehat{p}_\epsilon(j)$. The areas marked by A, B, C, D are the estimated proportions for true positives, false positives,

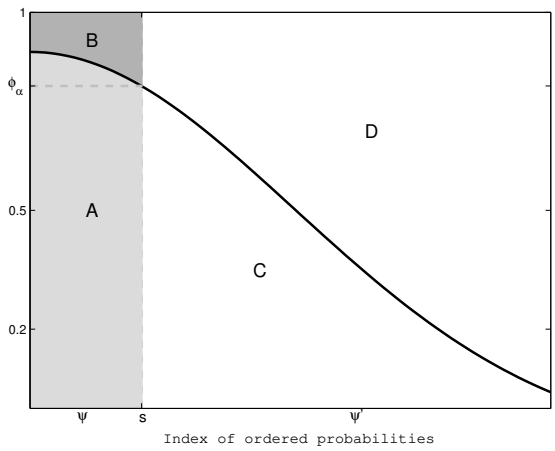

Figure 1: A diagram for determining the Bayesian FDR threshold ϕ_α.

false negatives and true negatives respectively. The threshold ϕ_α is indeed determined by constraining $B/(A+B) \leq \alpha$. The set of locations $\psi = \{l : \widehat{p}_{\epsilon,(l)} > \phi_\alpha\}$ corresponds to the set of *discoveries*. The threshold ϕ_α is a cut-point on the estimated posterior probabilities that correspond to an expected Bayesian FDR of α. In addition to the Bayesian expected FDR calculated by $B/(A+B)$, one can further calculate the corresponding Bayesian expected false negative rate (FNR) by $C/(C+D)$, sensitivity (SEN) by $A/(A+C)$, and specificity (SPEC) by $D/(B+D)$. The Bayesian expected statistics yield estimates of the statistics FDR, FNR, SEN, SPEC *in basis space* without knowing the true underlying function $\mathbf{B}(t)$ and its basis space counterpart $\mathbf{B}^*$. In simulations when we know the true function $\mathbf{B}(t)$ and $\mathbf{B}^*$, we can also compute the true FDR, FNR, SEN, and SPEC statistics; details are provided in Section 3.2. We call statistics computed using the true $\mathbf{B}(t)$ the "realized" quantities. We note that these statistics are calculated in basis space which does not require posterior samples; similar statistics can also be calculated in the data domain if we draw posterior samples from the approximate distribution, inverse-transform the samples back to the data domain, and follow the Bayesian FDR control procedure used in Morris et al. (2008).

3 Simulation Study

3.1 Simulation Setup

We designed a simulation study to assess the computational benefit of VFMM and evaluated the potential loss of accuracy when using approximate inference in VFMM. To mimic realistic inter- and intra-function correlations, we generated simulated data by resembling two real datasets described in Section 4—the cancer organ-by-cell-line proteomics data and the Tensor-based morphometry (TBM) brain imaging data. The first dataset provides an example of functional data observed on a one-dimensional domain, which we refer to as the 1-D case; and the second dataset represents functional data measured on a three-dimensional domain, which we refer to as the 3-D case. More details of the reference datasets are available in Section 4.

In the 1-D case, we simulated data by first fitting Baysian FMM to the reference dataset, from where we obtained posterior means of several key parameters in basis space, including the fixed effects $\{\mathbf{b}_j^*\}$, the random effect variances $\{q_j^*\}$, and the residual variances $\{s_j^*\}$. We then simulated data treating these estimated parameters as the underlying truth. The simulation involves generating random effects and residuals from multivariate normal distributions, and generating the responses $\{\mathbf{d}_j\}$ in basis space following model (4). Functional responses were finally obtained by applying inverse wavelet transform to $\{\mathbf{d}_j\}$. A total of 128 functional responses were simulated, four functions from each of the 32 "animals", with each function sampled on an equally spaced grid of 512.

In the 3-D case, we generated brain images on a $128 \times 128 \times 128$ grid following the reference data. We mimicked the cell mean design in the reference data by assuming that there are four groups, 25 subjects in each group. Therefore, the design matrix $\mathbf{X}$ is a 100 by 4 binary matrix, in which a one in the (i, j)th position indicates that the ith sample belongs to the jth group. Under this design, the fixed effect $\mathbf{B}(t)$ contains the four group means. We set the value of $\mathbf{B}(t)$ by adding different artificial shapes to different local regions of a template image. The artificial shapes include a cube with staircase intensity, a ball with highest intensity in center, a ball with linear intensity change, and a diamond

shape. The template image was chosen to be the minimal deformation template of the TBM data, created based on Magnetic Resonance (MR) scans of 40 randomly selected normal subjects. To simulate the random effect, we assume that the 100 brain images come form 25 independent batches, four in each batch, so the design matrix for the random effect is $\mathbf{I} \otimes \mathbf{1}$, where $\mathbf{I}$ is a 25×25 identity matrix and $\mathbf{1}$ is a 4×1 vector of ones. We simulated the 25 random batch effect functions using Karhunen–Loéve expansion with the first three components, where eigenvalues were fixed and eigen-functions were generated from a zero-mean random Gaussian process with the squared exponential kernel.

3.2 Evaluation Criteria

We applied VFMM to the two simulated datasets and compared its performance with that of the Bayesian FMM. Several summary statistics are defined to evaluate the performance in estimation and in identifying significant local regions. Specifically, we defined two summary statistics to assess the estimation performance of $\mathbf{B}^*$, including the averaged mean square error (AMSE) and the averaged posterior variability (APVar). Let $\mathbf{B}_a^*$ denote the ath row of $\mathbf{B}^*$; $a = 1, \ldots, p$. We define AMSE by $\text{AMSE} = p^{-1} \sum_{a=1}^{p} ||\widehat{\mathbf{B}}_a^* - \mathbf{B}_a^*||^2 / ||\mathbf{B}_a^*||^2$, where $\widehat{\mathbf{B}}_a^*$ denotes the posterior mean of $\mathbf{B}_a^*$. AMSE summarizes the variability of posterior mean relative to the truth. Furthermore, we defined APVar by $\text{APVar} = (pG)^{-1} \sum_{a=1}^{p} \sum_{g=1}^{G} ||\mathbf{B}_a^{(g)} - \widehat{\mathbf{B}}_a^*||^2 / ||\mathbf{B}_a^*||^2$, where $\{\mathbf{B}_a^{(g)}, g = 1, \ldots, G\}$ denotes the posterior samples of $\widehat{\mathbf{B}}_a^*$. APVar summarizes the posterior variability relative to the posterior mean. To assess the estimation performance of random effects, we defined two statistics for the random effect variances $\{q_j^*\}$: $\text{MSE} = \sum_j (\widehat{q}_j^* - q_j^*)^2 / \sum_j (q_j^*)^2$ and $\text{PVar} = 1/G \sum_{g=1}^{G} \{\sum_j (q_j^{*(g)} - \widehat{q}_j^*)^2 / \sum_j (q_j^*)^2\}$ where $\widehat{q}_j^*$ denotes the posterior mean of q_j^*. Note that for general inference, VFMM does not require any posterior sampling. In this simulation, in order to create a fair comparison with FMM, we drew posterior samples from the approximate distribution and used the samples to calculate the above statistics.

To evaluate the performance of region detection, we also calculated the true FDR, FNR, SEN and SPEC statistics in data domain by assuming that $\mathbf{B}(t)$ is known. These statistics, also called "realized" statistics as noted in Section 2.3, measure the proportion of regions

that are correctly or incorrectly identified under different considerations. For example, if we use the criterion $|\widetilde{C}^{\dagger}(t)| > \delta$ to flag locations on a grid of $\mathcal{T}$, FDR is calculated through dividing the number of grid points that are falsely detected by the total number of grid points detected; SEN is obtained by calculating the proportion of correctly detected grid points among all grid points that truly satisfies $|C(t)| > \delta$; FNR is calculated by finding that, among those non-detected grid points, which proportion truly satisfies $|C(t)| > \delta$; and SPEC is obtained by calculating that, among all grid points that truly satisfies $|C(t)| \leq \delta$, which proportion is correctly non-detected.

3.3 Results

We repeated each of the 1-D and 3-D simulations 50 times to mitigate the Monte Carlo variability in the data generation process. In each simulation, we applied both VFMM and Bayesian FMM, and compared their performance on estimation and region detection. When fitting both VFMM and FMM, we adopted wavelet transformation by using Daubechies wavelets with six resolution levels. In the 3-D simulation case, wavelet compression was performed before applying both models; the truncation parameter was chosen so that the remaining components retain at least 80% of the total energy (sum of squared wavelet coefficients), which reduced the dimension from around 10.8 million to $49,152$ after compression. When applying Bayesian FMM, we ran 5000 MCMC iterations and treated the first 3000 iterations as the burnin period. In Table 1, we reported the mean and standard deviation (in parentheses) of each summary statistic, calculated across the 50 repetitions. Note that for the 3-D case, the MSE statistic for $\{q_j^*\}$ is not available since the true values of $\{q_j^*\}$ are unknown, owing to the fact that the random effect $\mathbf{U}(t)$ was simulated directly in data domain.

When performing region detection, in the 1-D case, we focused on detecting regions with absolute values greater than $\delta = \log_2(1.5)$ on the cell line effect, organ effect, organ-cell-line interaction effect and the mean effect; in the 3-D case, we focused on detecting regions with pairwise contrast effect greater than $\delta = 5$. In the 1-D case, to better compare VFMM with FMM, we flagged regions directly in the data domain by directly controlling

Table 1: Simulation results on estimation and region detection for 1-D and 3-D cases.

		Estimation				Time
		$\mathbf{B}^*$		$\{q_j^*\}$		
Data	Model	AMSE	APVar	MSE	PVar	(hrs)
1-D	FMM	.0109 (.0016)	.0108 (7.1e-04)	.0010 (8.1e-05)	.0981 (.0202)	.4700 (.0024)
	VFMM	.0107 (.0017)	.0069 (3.6e-04)	.0012 (6.8e-05)	.0032 (6.1e-04)	.0226 (.0014)
3-D	FMM	.1020 (.0056)	.0014 (6.4e-03)	—	.0960 (.0340)	7.530 (.0058)
	VFMM	.1440 (.0053)	.0002 (4.8e-03)	—	.0056 (7.4e-03)	.3500 (.0002)
		Region Detection				
		FDR	FNR	SEN	SPEC	
1-D	FMM	.066 (.022)	.076 (.016)	.864 (.033)	.964 (.014)	
	VFMM	.086 (.025)	.064 (.016)	.889 (.032)	.951 (.017)	
3-D	FMM	.028 (0.027)	.065 (0.015)	.918 (0.045)	.993 (0.011)	
	VFMM	.032 (0.029)	.067 (0.016)	.921 (0.041)	.992 (0.013)	

Bayesian expected FDR on a grid of $\mathcal{T}$ following the approach of Morris et al. (2008); "realized" statistics were calculated by comparing flagged regions with the ground truth in data domain. Note that this approach requires using posterior samples of the fixed effects for both FMM and VFMM. In the 3-D case, since storing posterior samples and performing inverse wavelet transforms of them are both computationally expensive, we performed basis-space testing described in Section 2.3. This testing procedure does not require generating posterior samples in the VFMM case. In this testing procedure, we set thresholds $\epsilon = 0.07$ and $\delta = 5$. We evaluated the performance of region detection in data domain by using "realized" statistics. For both 1-D and 3-D simulation, we controlled the overall Bayesian expected FDR across all contrast effects to be less than the significance level $\alpha = 0.05$. In addition to estimation and region detection, Table 1 also listed computation time in hours.

From Table 1, we see that for the estimation of $\mathbf{B}^*$, in the 1-D case, VFMM resulted in

similar AMSE (.0107 vs. .0109) and lower APVar (.0069 vs. .0108) than FMM; and in the 3-D case, VFMM resulted in higher AMSE (0.1440 vs. 0.0990) and lower APVar (.0002 vs. .0014) than FMM. The pattern on estimating $\{q_j^*\}$ is similar with that of $\mathbf{B}^*$. The lower APVar statistic of VFMM reflects the effects of using mean-field assumption in the posterior approximation—assuming independence across different components of the joint posterior usually results in underestimates of posterior variance, a well-known problem of variational inference (Giordano et al., 2015; Blei et al., 2017). For region detection, we see that for the 1-D case, VFMM resulted in higher false discovery rate (0.086 vs. 0.066), lower false negative rate (0.064 vs. 0.076), with slightly higher sensitivity (0.889 vs. 0.864) and lower specificity (0.951 vs. 0.964). These results indicate that, with narrower credible bands, VFMM tends to flag more locations, leading to higher false discoveries and sensitivity. Statistics for the 3-D case show similar patterns as in the 1-D case.

Regarding computation time, Table 1 demonstrates a clear advantage of VFMM relative to FMM. The 1-D simulations are performed on a mac laptop with 2.2 GHz Intel Core i7 CPU and 16 GB of RAM. All 3-D simulations are performed on a CPU cluster equipped with 190 nodes, 32 cores per node, and each core consists of two Intel Broadwell E5-2683v4 @ 2.1GHz processors with 128 GB of RAM. For the FMM, the computation time required for running 5000 MCMC iterations increases from 0.47 hours in the 1-D case to 7.53 hours in the 3-D case. The VFMM, on the other hand, requires only 0.0226 hours (1.356 minutes) for the 1-D cases and 0.35 hours for the 3-D case. In addition to shorter computation time, the storage space required for VFMM is substantially reduced as it does not require drawing posterior samples. Specifically, the storage space required for FMM to save 2000 posterior samples increases from 257 megabytes in the 1-D case to 4.94 gigabytes in the 3-D case, whereas VFMM only requires 1.38 megabyte in the 1-D case and 3.64 megabyte in the 3-D case. These computation benefits make VFMM attractive for large-scale data.

In addition to summary statistics shown in Table 1, in Figures 2 and 3, we also plot the estimation and region detection results for selected contrast effects. Figure 2 shows the results of the cell line effect $C(t) = (B_1(t) - B_2(t) + B_3(t) - B_4(t))/2$. Green and blue lines mark the posterior means of VFMM and FMM respectively. The yellow line shows the true

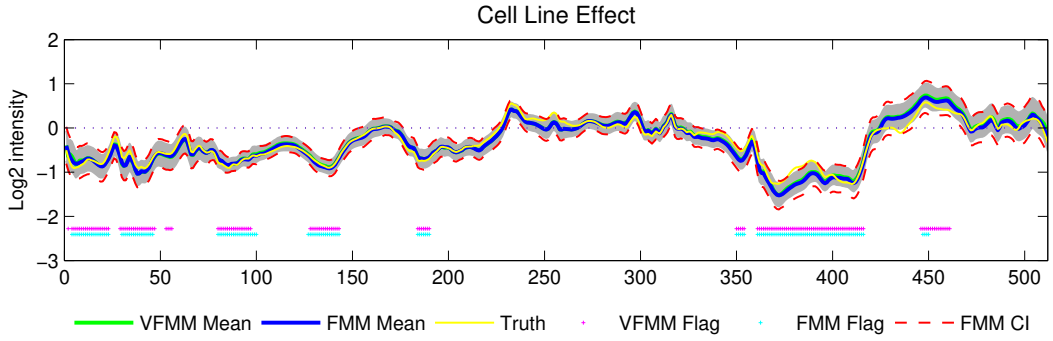

Figure 2: The 1-D simulation case: estimation and region detection results for the simulated cell line effect.

value of $C(t)$. The 95% percent credible bands were shown by shaded gray area for VFMM and by red dash lines for FMM. Here, the 95% percent credible bands were calculated by finding the $(0.025, 0.975)$ percentiles pointwisely on a grid of $\mathcal{T}$ based on posterior samples of $C(t)$. Detected regions are flagged by magenta and cyan dots at the bottom of the plot. From Figure 2, we see that while VFMM and FMM resulted in very close mean estimates, VFMM produced slightly narrower credible bands and more flagged locations than FMM, a result that is consistent with the higher FDR and sensitivity observed in Table 1. This indicates that VFMM is more aggressive than FMM on region detection, a potential issue of mean-field variational inference. More plots for the 1-D simulated case are available in the supplementary materials.

Figure 3 shows the flagged regions by VFMM and FMM for the contrast effect $B_1(t) - B_2(t)$ in the 3-D simulated case, along with the truth. For demonstration convenience, only one a 2-D slice of the 3-D image is shown. White areas are regions that are not flagged. Colors represent estimated values. For this contrast effect, we expect two significant local regions to be flagged—one corresponds to a cube with staircase intensity, the other corresponds to a ball with the highest intensity in the center. From Figure 3, we see that VFMM and FMM perform similarly in identifying the two regions. Both VFMM and FMM estimated the ball shape very well and identified the staircase pattern with less accuracy on the boundary. More plots for the 3-D simulated case are available in the supplementary

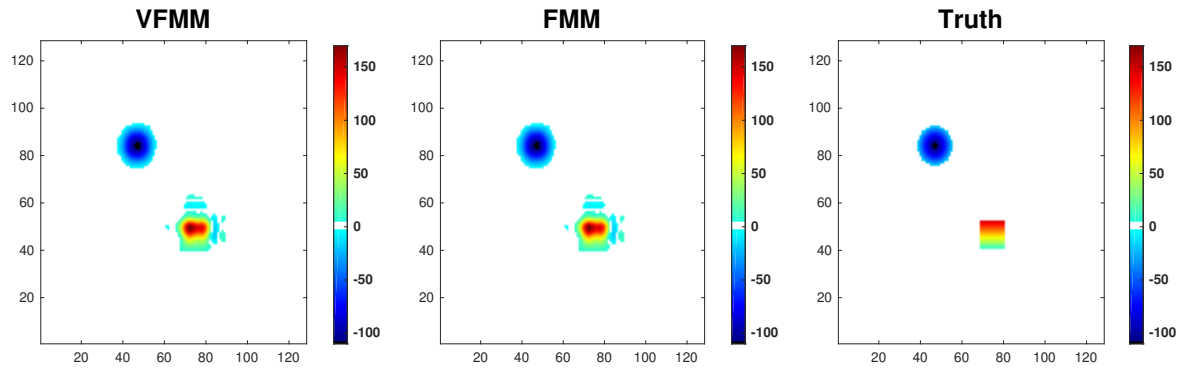

Figure 3: The 3-D simulation case: region detection results for the contrast effects $\mathbf{B}_1(t) - \mathbf{B}_2(t)$ along with the truth. Only one 2-D slice of the the 3-D image is plotted. White areas are regions that are not flagged. Colors represent estimated values. Left, middle and right figures correspond to results of VFMM, FMM and the truth respectively.

materials.

4 Real Data Applications

We demonstrate the applications of VFMM by analyzing two real datasets, a 1-D cancer proteomics dataset and a 3-D brain imaging dataset. Performance of VFMM is compared with Bayesian FMM.

4.1 Analysis of the 1-D Cancer Proteomics Data

In this analysis, we study the proteomic spectra collected from a MALDI-TOF mass spectrometer during a cancer cell line study. During this study, a tumor from one of two cancer cell lines was implanted into either the brains or lungs of 16 nude mice. The cell lines were A375P, a human melanoma cancer cell line with low metastatic potential, and PC3MM2, a highly metastatic human prostate cancer cell line. The goal was to find blood serum proteins that are differentially expressed between organ implant sites, implanted cell line types, or the organ-by-cell line interaction. Data were collected by drawing a blood serum sample from each animal and run it through a MALDI-TOF mass

spectrometer. This produces a proteomic spectrum $y(t)$ that is a function supported on a 1-D domain. The proteomic spectrum consists of many peaks; a peak at a location t corresponds to a protein/peptide in the sample with molecular mass of t Daltons. The spectral intensity $y(t)$ provides a rough estimate of the corresponding protein abundance. During the experiment, we obtained two spectra for each mouse, one using a low laser intensity and one using a high laser intensity. Here, we consider the part of the spectrum between $t = 2,000$ and $t = 14,000$ Daltons, a range that includes $T = 7,985$ points per spectrum. Preprocessing steps, include background correction, normalization of the mass spectra, and a $\log_2$ transformation of the intensities were performed before applying VFMM and FMM. More details of preprocessing can be found in Morris et al. (2005).

We applied both VFMM and FMM to the preprocessed dataset by adopting the same basis transformations and design matrices. In particular, we applied a discrete wavelet transform to each spectrum by using the Daubechies wavelets with eight vanishing moments, periodic boundary extension mode, and nine resolution levels. We set the fix effect design matrix $\mathbf{X}$ by using the cell mean model for the factorial design with an additional column for the laser intensity effect, so that $\mathbf{X}$ is a 32×5 matrix. Columns one to four indicates four treatment groups: brain-A375P, brain-PC3MM2, lung-A375P, lung-PC3MM2, respectively, while column five indicated whether the observations were from high (coded as 1) or low (coded as -1) laser intensity. The random effect design matrix Z was a 32 $\times$16 binary matrix with $Z_{ib} = 1$ indicating that spectrum i came from the bth animal. Based on estimation of the fixed effects, we detected nonzero regions on three contrast effects: the organ effect $C_1(t) = (B_1(t) + B_2(t) - B_3(t) - B_4(t))/2$, the cell-line effect $C_2(t) = (B_1(t) - B_2(t) + B_3(t) - B_4(t))/2$, and the organ-by-cell line interaction effect $C_3(t) = (B_1(t) - B_2(t) - B_3(t) + B_4(t))/2$. Results of region detection were compared between VFMM and FMM.

To detect significant regions on contrast effects, for both models, we applied the Bayesian expected FDR control with $\delta = 1$ on the measurement grid in data domain following Morris et al. (2008). Note that this requires posterior samples for the fixed effects; for VFMM we have sampled from the estimated posterior distributions to get posterior samples. For

both models, we detected regions by controlling the overall Bayesian expected FDR across all contrast effects. The significant level for both approaches was set to be $\alpha = 0.05$. In Figure 4, we compared significant regions flagged by FMM and VFMM on the cell line effect (top pane) together with a zoom-in plot on the [2950, 3070] kilodaltons region (bottom pane). The flagged regions were marked on the mean cell line effect obtained by VFMM. We use different colors to denote whether the locations were flagged by VFMM only, FMM only, or both VFMM and FMM. From Figure 4, we observe that for the cell line effect, a total of 9 regions were flagged by both models. For the common flagged regions, VFMM usually flagged wider intervals than FMM. Besides regions that were flagged by both approaches, there were three extra regions with more than 10 contiguous grid points that were flagged by VFMM but not by FMM; these regions include intervals $[2587, 2602]$, $[3809, 3837]$, $[3856, 3876]$ kilodaltons. There is one contiguous region flagged by FMM, $[2214, 2300]$ kilodaltons, which VFMM flagged as $[2203, 2279]$ and $[2286, 2302]$. Similar patterns were observed for the organ effect and organ-by-cell-line interaction effect. For the organ effect, 16 regions were flagged by both models and two extra regions were flagged by VFMM but not by FMM. For the organ-by-cell-line interaction effect, seven regions were flagged by both models and four regions were flagged by VFMM but not by FMM. Again, we shall interpret the additional flagging cautiously due to the possible higher FDR associated with VFMM. Figures on region detection for the organ effect and organ-by-cell-line interaction effect are available in supplementary materials.

In addition to the plot in Figure 4, we also calculated the Bayesian expected SEN, FNR and SPEC following the approach described in Section 2.3. For the 1-D cancer proteomics data case, since posterior samples were obtained for both FMM and VFMM, we have calculated these statistics in data domain. For the 3-D brain imaging data case, the statistics were calculated in the compressed wavelet domain. Results are listed in Table 2. The statistics reported in Table 2 are based on controlling the Bayesian expected FDR across all wavelet components and all three contrast effects in basis space. The significant level for the BFDR control was chosen to be $\alpha = 0.05$. From Table 2, we see that for the 1-D cancer proteomic data, VFMM achieves higher sensitivity (0.720 vs. 0.612), comparable

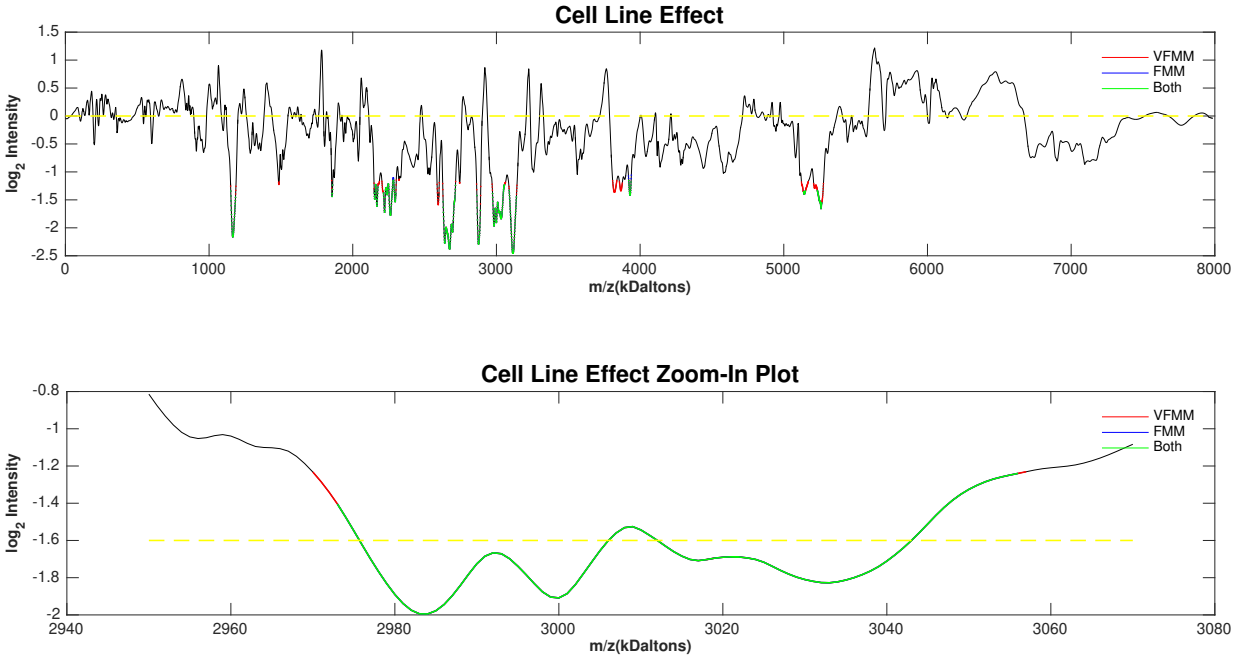

Figure 4: Cancer proteomics data analysis. Top pane: significant nonzero regions flagged by VFMM and FMM on the cell line effect; bottom pane: a zoom-in plot on the [2950, 3070] kilodaltons region. Regions were flagged on the mean estimate obtained by VFMM. Red, blue, and green colors denote locations flagged by VFMM only, FMM only, or both VFMM and FMM, respectively.

specificity (0.980 vs. 0.981), and lower FNR (0.129 vs. 0.187) than FMM. Table 2 also reported computation time. For the 1-D data, FMM takes around 6.2 hours to perform 4000 MCMC iterations, whereas it only takes 0.55 hours for VFMM to converge. The regions we discovered by using both VFMM and FMM, together with the Bayesian expected statistics, provide statistical evidence about which proteins/peptides are differential expressed across organs, cell-lines, and their interactions. These proteins/peptides may serve as potential biomarkers for the assessment, diagnosis, and treatment of cancer.

4.2 Analysis of the 3-D Brain Imaging Data

In this analysis, we consider the 3-D brain imaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI is a multisite study that aims to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of

Alzheimer's disease (AD). We will analyze the preprocessed 3-D Tensor-based morphometry (TBM) data shared by the Laboratory of Neuro Imaging at University of California, Los Angeles School of Medicine. This dataset can be downloaded from the ADNI website `http://adni.loni.usc.edu/`. TBM is an image analysis technique that measures brain structural differences relative to a common anatomical template Frackowiak et al. (2003). To generate TBM images, a minimal deformation template (MDT) was first created based on Magnetic Resonance (MR) scans of 40 randomly selected normal subjects. All other brain MR images were aligned to the MDT using a nonlinear inverse-consistent elastic intensity-based registration algorithm Stein et al. (2010). For each brain image, a Jacobian matrix field was derived based on the gradients of the deformation field that warped the brain image to the MDT. Volumetric tissue differences were then assessed at each voxel by calculating the determinant of the Jacobian matrix. The determinant value encodes local volume excess or deficit relative to the MDT. The pre-processing TBM data consist of 816 subjects, among which 228 were healthy elderly controls (118 Male, 110 Female), 396 were diagnosed with mild cognitive impairment (MCI; 255 Male, 141 Female), and 192 were diagnosed as Alzheimer's disease (AD; 101 Male, 91 Female). Each TBM image was measured on a common $220\times220\times220$ grid in 3-D.

By analyzing the TBM data, we aim to estimate the contrast effects between groups (i.e., the differences between group means) and detect local brain regions with systematic volumetric expansion or compression across patient groups with different diagnostic status or genders. We adopt the cell mean design by setting $\mathbf{X}$ to be a 816×6 binary matrix, with 1's in every row indicating the diagnosis stage and the subject's gender. In particular, columns 1-6 of $\mathbf{X}$ correspond to subgroups Normal-Male, Normal-Female, MCI-Male, MCI-Female, AD-Male, AD-Female respectively. Under this design, the contrast effect between AD and Normal can be calculated by pre-multiplying $(-1/2, -1/2, 0, 0, 1/2, 1/2)$ to the fixed effects $\mathbf{B}$ (or $\mathbf{B}^*$) and the gender effects can be calculated by pre-multiplying $(1/3, -1/3, 1/3, -1/3, 1/3, -1/3)$. Since there is only one image per subject, no random effect was modeled. We applied VFMM and FMM by adopting the same basis transformations and design matrix. In particular, we applied a 3-D discrete wavelet transform

Table 2: Real data application results: Bayesian expected sensitivity, false negative rate and specificity for region detection, calculated in wavelet domain.

		Region Detection			Time
Data	Model	FNR	SEN	SPEC	(hrs)
1-D	FMM	0.187	0.612	0.981	6.2
	VFMM	0.129	0.720	0.980	0.55
3-D	FMM	0.030	0.980	0.928	13.9
	VFMM	0.028	0.981	0.929	1.19

to each image by using the Daubechies wavelets with four vanishing moments, periodic boundary extension mode and four resolution levels. To further reduce the dimension, in wavelet domain we applied an efficient wavelet compression algorithm, which reduces the dimension from $10,657,241$ to $22,096$ while retaining 96% of the total energy.

Based on group means obtained from VFMM and FMM in wavelet domain, we calculated pair-wise contrast effects between the AD (N=192), MCI (N=396), and normal (N=228) groups as well as the contrast effect between male and female groups. To identify local regions on these contrast effects, we performed basis-space testing in the compressed wavelet domain by following the procedure proposed in Section 2.3. For both VFMM and FMM models, we set $\epsilon = 0.5$ and controlled the overall Bayesian expected FDR across all compressed wavelet components and all four contrast effects at the significant level $\alpha = 0.05$. Significant local regions were flagged in data domain by using threshold $\delta = 20$. In Figure 5, we demonstrate regions flagged by VFMM for each of the four contrast effects by using sliced 2D plots. The flagged regions (colored by red or blue) were plotted on top of the MDT background image (the gray scale image). For each contrast effect, we showed the flagged regions via three views: the axial, sagittal and coronal views, sliced in the middle of the 3D brain along three directions. Similar 2D plots for FMM are available in supplementary materials. In addition to the 2D plots, we also produced interactive plots in three views for each contrast effect, which allows users to visualize the flagged regions

for all slices by using scroll bars.

Flagged regions on contrast effects reveal local volumetric tissue change for one group relative to another group. From Figure 5 and the interactive plots, we observe that, on the AD-Normal contrast effect, there is a profound positive contrast effect in the lateral ventricle region, which indicates cerebrospinal fluid (CSF) inflation in AD patients. In addition, positive contrast effects are also seen in the circular sulcus of the insula bilaterally, suggesting brain volume expansions in these regions. Furthermore, we observe negative contrast effects in the temporal and parietal regions and the hippocampus, which suggests brain atrophy in these regions in AD patients. The contrast effects for AD-MCI and MCI-Normal show similar patterns but with lower values and smaller regions. This indicates graduate volumetric tissue changes from Normal to MCI and from MCI to AD.

On the Male-Female contrast effect, we observe positive contrast effects in the lateral ventricle region, which indicates more CSF inflation for males relative to females. Additionally, we also observe positive contrast effects on the top portion of the frontal region. This suggests higher brain volume (i.e., less tissue loss) for males in this region. Finally, we observe negative contrast effects in the pariental and temporal regions, indicating lower brain volume (i.e., more brain atrophy) for males in these regions. Results of FMM are similar to VFMM; the 2D and interactive plots are shown in supplementary materials. In addition to plots, in the bottom section of Table 2, we also listed Bayesian expected SEN, FNR and SPEC, calculated just as in the 1-D case described in section 4.1. These results demonstrate that VFMM achieves Bayesian expected statistics very close to FMM. Regarding computation, we have split the wavelet components to three blocks and performed posterior calculation in parallel for both VFMM and FMM. It took 13.9 hours for FMM to finish $10,000$ MCMC iterations with a burnin period of 5000 iterations, whereas it only took 71 minutes for VFMM to converge. Additionally, FMM took 5.1 gigabytes to store 5000 posterior samples and the posterior results for VFMM only took 6.9 megabytes of storage space.

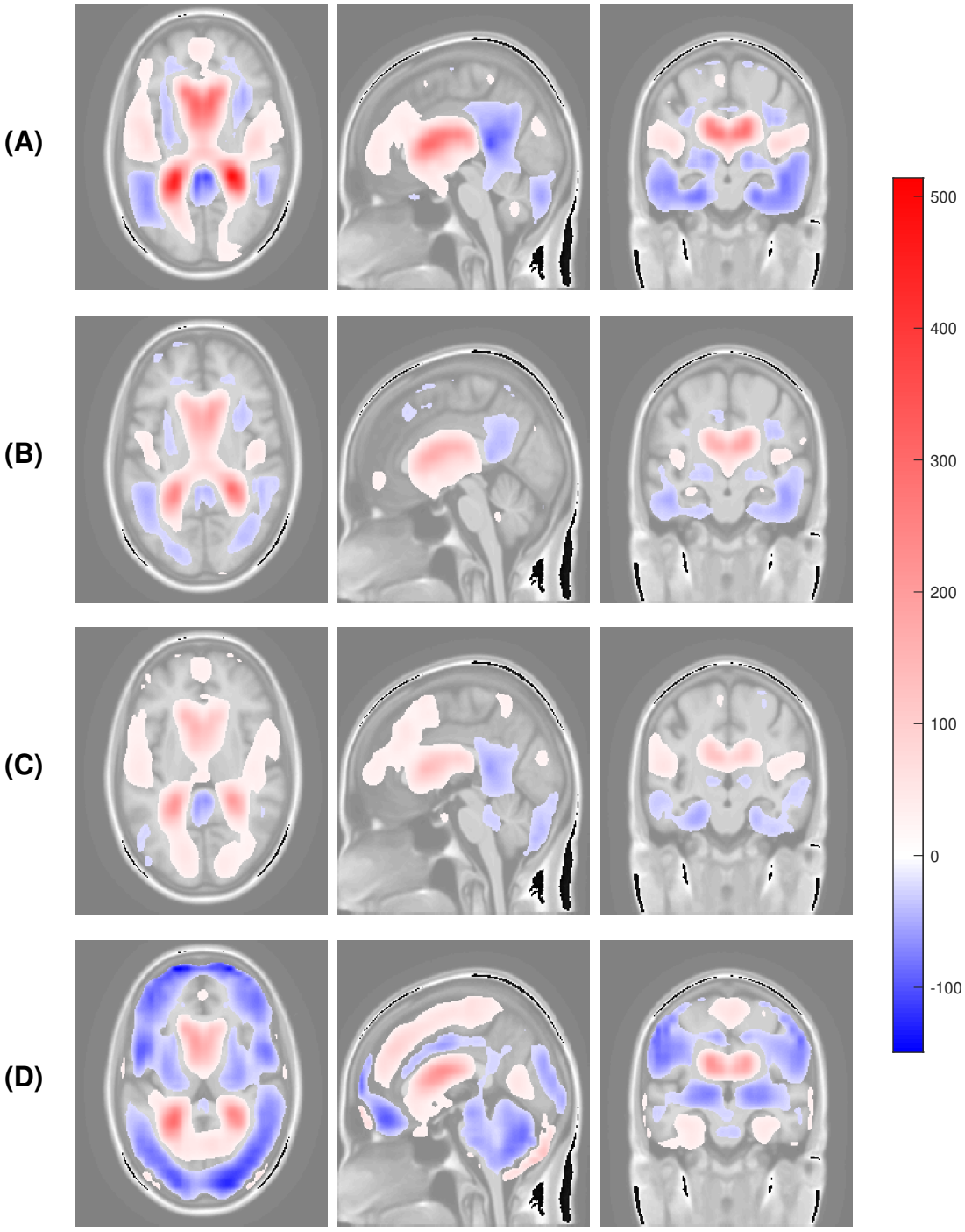

Figure 5: TBM brain imaging data analysis: plots of regions detected for four contrast effects: (A) AD-Normal; (B) AD-MCI; (C) MCI-Normal; (D) Male-Female. Each row illustrates three 2D images according to three views—the axial (sliced at $z = 110$), sagittal (sliced at $x = 110$), and coronal (sliced at $y = 110$) views, from left to right.

5 Discussion

To improve computational efficiency of Bayesian FMM, we have proposed VFMM, a new computational approach that enables ultra-fast approximate inference. This novel framework approximates the posterior distribution in wavelet domain by using variational Bayes, avoiding expensive MCMC sampling and the need to store posterior samples. Our approach leads to an automated, efficient algorithm suitable for a large family of high-dimensional functional data. In order to detect interesting local regions, we proposed a fast basis-space testing approach that does not require posterior samples. Our applications to the 1-D cancer proteomics data and 3-D brain imaging data demonstrate comparable Bayesian expected sensitivity and specificity relative to Bayesian FMM.

A key contribution of VFMM is its computational advantage. As shown by our simulation and real data analysis, VFMM can accomplish Bayesian estimation in minutes whereas MCMC-based Bayesian FMM may require days to finish a few thousand MCMC iterations. Additionally, VFMM only takes very little storage space to store the estimated model parameters, whereas MCMC-based FMM may need gigabytes of storage to save posterior samples. Finally, the algorithm of VFMM can be easily implemented in parallel on manycore CPUs or multicore GPUs, making it suitable for extremely high-dimensional computational tasks.

We have focused on wavelet basis in the proposed VFMM due to its nice whitening properties and several appealing characteristics, such as the multi-resolution representation, the choice of compact supported wavelets, and the ability to perform efficient compression. However, as emphasized by several works (Morris et al., 2011; Meyer et al., 2015; Zhang et al., 2016), the Bayesian FMM framework is not limited to wavelet bases; it can be used with other lossless or near lossless basis functions that have whitening (decorrelation) property. While other basis functions can be generally applied to VFMM just as in FMM, compactly supported bases are desirable in order to detect local regions by using our proposed basis-space testing approach. Examples include a family of wavelets (e.g., Daubechies, spherical wavelets) and orthogonal splines (Mason et al., 1993).

We have assumed Gaussian process distributions for the random effect and residual

functions. As we have highlighted in Section 1, the current framework can be directly applied to FMM under different basis choices as well as several extensions of FMM without any major modifications. For other extensions, such as the robust FMM with heavier tailed distributions (Zhu et al., 2011) and FMM with spatial-temporal correlations in residuals (Zhang et al., 2016; Zhu et al., 2018), modifications are needed in order to estimate parameters induced by assuming heavier-tailed distribution or between-function correlation. Extension to the former case should be straightforward as the conditional posteriors are still in the exponential family. Extension to the latter case may involve applying search algorithms to numerically optimize the ELBO, which can be computationally more challenging.

Both FMM and VFMM assume spike-slab priors for fixed effects in basis space. We could easily use other sparsity-inducing priors such as Bayesian Lasso (Park and Casella, 2008), scale mixtures of normals (Griffin and Brown, 2005), horse shoe (Carvalho et al., 2010), and non-local priors (Johnson and Rossell, 2010). In our experience, these choices produce similar results in terms of regularization. Nevertheless, both models could be extended by including other alternative priors which is left to future work. Notable works on variational Bayes approaches to variable selection include Ormerod et al. (2017), Serra et al. (2019), and Ray and Szabó (2021), etc.

The VFMM is based on variational Bayes. A well known issue of variational Bayes is the approximation error caused by the mean-field assumption. Our simulations demonstrate that VFMM tends to provide narrower credible intervals in fixed and random effects, which often leads to higher sensitivity and sometimes higher false discovery rate or lower specificity in region detection. However, following our observation in simulations, we conjecture that the effect of reduced specificity should not be a big concern if the number of true positive components in basis space is small relative to the total number of components, which is usually true if one adopts sparse basis such as wavelets. After all, given its computational benefits, the proposed VFMM provides an ideal option for producing fast initial results. If desired, full Bayesian analysis can still be performed by using outputs of VFMM as initial values.

Finally, we have adopted mean-field variational inference. Alternatively, stochastic variational inference (Hoffman et al., 2013) has been proposed which offers better scalability for data with large sample size. However, in VFMM, we are often dealing high-dimensional functional data with relatively small/moderate sample size; for example, our brain imaging data contain images with $O(10^6)$ measurement points with sample size 816. This is the typical situation in most medical and genomics applications since collecting large samples is usually expensive. Thus, in these cases, the subsampling strategy of stochastic variational inference is not needed. Nevertheless, it remains a promising solution if the sample size becomes large.

Supplementary Materials

The supplementary materials contain details of the VFMM algorithm, Henderson's mixed model equations, ELBO under a block design, and additional simulation and real data results. Demonstration code written in Matlab is available.

Acknowledgements

HZ was supported by National Science Foundation of the United States grant 1611901 and 1762577. JSM was supported by NIH grants R01-CA244845, R01-CA178744, and UL1-TR001878. The authors report there are no competing interests to declare.

References

Blei, D. M. (2006). Variational Inference for Dirichlet Process. *Bayesian Analysis 1*, 121–144.

Blei, D. M., A. Kucukelbir, and D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association 112*, 859–877.

Cardot, H., F. Ferraty, and P. Sarda (1999). Functional linear model. *Statistics & Probability Letters 45*(1), 11–22.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010, 04). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Cuevas, A., M. Febrero, and R. Fraiman (2002). Linear functional regression: the case of fixed design and functional response. *Canadian Journal of Statistics 30*(2), 285–300.

Earls, C. and G. Hooker (2017a). Combining functional data registration and factor analysis. *Journal of Computational and Graphical Statistics 26*(2), 296–305.

Earls, C. and G. Hooker (2017b). Variational Bayes for Functional Data Registration, Smoothing, and Prediction. *Bayesian Analysis 12*(2), 557 – 582.

Faes, C., J. T. Ormerod, and M. P. Wand (2011). Variational bayesian inference for parametric and nonparametric regression with missing data. *Journal of the American Statistical Association 106*(495), 959–971.

Fan, Y. (2003). On the approximate decorrelation property of the discrete wavelet transform for fractionally differenced processes. *IEEE Transactions on Information Theory 49*(2), 516–521.

Fan, Y., G. M. James, and P. Radchenko (2015, 10). Functional additive regression. *The Annals of Statistics 43*(5), 2296–2325.

Ferraty, F. and Y. Romain (Eds.) (2018, aug). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.

Frackowiak, R., K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny (2003). *Human Brain Function* (2nd ed.). Academic Press.

Giordano, R. J., T. Broderick, and M. I. Jordan (2015). Linear response methods for accurate covariance estimates from mean field variational bayes. In *Neural Information Processing Systems*.

Goldsmith, J., W. MP., and C. Crainiceanu (2011). Functional regression via variational bayes. *Electron. J. Statist. 5*, 572–602.

Griffin, J. E. and P. J. Brown (2005, May). Alternative prior distributions for variable selection with very many more variables than observations. Technical report, University of Warwick, Centre for Research in Statistical Methodology.

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

Hall, P., J. L. Horowitz, et al. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Hoffman, M. D., D. M. Blei, C. Wang, and J. Paisley (2013). Stochastic variational inference. *Journal of Machine Learning Research 14*(4), 1303–1347.

Hui, F. K. C., C. You, H. L. Shang, and S. Müller (2019). Semiparametric regression using variational approximations. *Journal of the American Statistical Association 114*(528), 1765–1777.

Ishwaran, H. and J. S. Rao (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics 33*(2), 730 – 773.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B 64*(3), 411–432.

Johnson, V. E. and D. Rossell (2010). On the use of non-local prior densities in bayesian hypothesis tests. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 72*(2), 143–170.

Lancia, L., P. Rausch, and J. S. Morris (2015). Automated quantitative analysis of ultrasound tongue contours via wavelet-based functional mixed models. *Journal of the Acoustical Society of America 137*, EL178–EL183.

Lee, W., M. F. Miranda, P. Rausch, V. Baladandayuthapani, M. Fazio, J. C. Downs, and J. S. Morris (2019). Bayesian semiparametric functional mixed models for serially correlated functional data, with application to glaucoma data. *Journal of the American Statistical Association 114*(526), 495–513.

Luts, J., T. Broderick, and M. P. Wand (2014). Real-time semiparametric regression. *Journal of Computational and Graphical Statistics 23*(3), 589–615.

Luts, J. and M. P. Wand (2015). Variational Inference for Count Response Semiparametric Regression. *Bayesian Analysis 10*(4), 991 – 1023.

Martinez, J. G., K. M. Bohn, R. J. Carroll, and J. S. Morris (2013). A study of mexican free-tailed bat chirp syllables: Bayesian functional mixed models for nonstationary acoustic time series. *Journal of the American Statistical Association 108*(502), 514–526.

Mason, J. C., G. Rodriguez, and S. Seatzu (1993). Orthogonal splines based on B-splines — with applications to least squares, smoothing and regularisation problems. *Numer Algor 5*, 25–40.

McLean, M. W., G. Hooker, A.-M. Staicu, F. Scheipl, and D. Ruppert (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics 23*(1), 249–269.

McLean, M. W., F. Scheipl, G. Hooker, S. Greven, and D. Ruppert (2017). Bayesian functional generalized additive models with sparsely observed covariates.

Meyer, M. J., B. A. Coull, F. Versace, P. Cinciripini, and J. S. Morris (2015). Bayesian function-on-function regression for multilevel functional data. *Biometrics 71*(3), 563–74.

Minka, T. P. (2001). Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, San Francisco, CA, USA, pp. 362–369. Morgan Kaufmann Publishers Inc.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application 2*, 321–359.

Morris, J. S. (2017). Comparison and contrast of two general functional regression modeling frameworks. *Statistical Modeling 17*(1-2), 59–85.

Morris, J. S., C. Arroyo, B. A. Coull, M. R. Louise, R. Herrick, and S. Gortmaker (2006). Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study. *Journal of the American Statistical Association 101*, 1352–1364.

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

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

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

Morris, J. S., K. R. Coombes, J. Koomen, K. A. Baggerly, and R. Kobayashi (2005). Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics 21*, 1764–1775.

Müller, H.-G. and U. Stadtmüller (2005, 04). Generalized functional linear models. *Ann. Statist. 33*(2), 774–805.

Ormerod, J. T. and M. P. Wand (2012). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics 21*(1), 2–17.

Ormerod, J. T., C. You, and S. Müller (2017). A variational Bayes approach to variable selection. *Electronic Journal of Statistics 11*(2), 3549 – 3594.

Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis, Section Edition.* New York: Springer.

Ray, K. and B. Szabó (2021). Variational bayes for high-dimensional linear regression with sparse priors. *Journal of the American Statistical Association 0*(0), 1–12.

Ruli, E., N. Sartori, and L. Ventura (2016). Improved laplace approximation for marginal likelihoods. *Electron. J. Statist. 10*(2), 3986–4009.

Sarlos, T. (2006, Oct). Improved approximation algorithms for large matrices via random projections. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pp. 143–152.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. New York: John Wiley & Sons.

Serra, J. G., J. Mateos, R. Molina, and A. K. Katsaggelos (2019). Variational em method for blur estimation using the spike-and-slab image prior. *Digital Signal Processing 88*, 116–129.

Stein, J., X. Hua, S. Lee, A. Ho, A. Leow, A. Toga, A. Saykin, L. Shen, T. Foroud, N. Pankratz, M. Huentelman, D. Craig, J. Gerber, A. Allen, J. Corneveaux, B. Dechairo, S. Potkin, M. Weiner, and P. Thompson (2010). Voxelwise genome-wide association study (vgwas). *NeuroImage 53*, 1160–1174.

Sun, S. (2013). A review of deterministic approximate inference techniques for Bayesian machine learning. *Neural Computing and Applications 23*(7-8), 2039.

Wang, C., M.-H. Chen, E. Schifano, J. Wu, and J. Yan (2016). Statistical methods and computing for big data. *Stat Interface 9*(4), 399–414.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*(1), 257–295.

Wikipedia contributors (2019). Embarrassingly parallel — Wikipedia, the free encyclopedia. [Online; accessed 17-October-2019].

Yang, H., V. Baladandayuthapani, A. U. K. Rao, and J. S. Morris (2020). Quantile Function on Scalar Regression Analysis for Distributional Data. *Journal of the American Statistical Association 115*(529), 90–106.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics 33*(6), 2873–2903.

Yuan, M. and T. T. Cai (2010, 12). A reproducing kernel hilbert space approach to functional linear regression. *Ann. Statist. 38*(6), 3412–3444.

Zhang, C.-X., S. Xu, and J.-S. Zhang (2019). A novel variational bayesian method for variable selection in logistic regression models. *Computational Statistics & Data Analysis 133*, 1–19.

Zhang, L., V. Baladandayuthapani, H. Zhu, K. A. Baggerly, T. Majewski, B. A. Czerniak, and J. S. Morris (2016). Functional car models for large spatially correlated functional datasets. *Journal of the American Statistical Association 111*(514), 772–786. PMID: 28018013.

Zhou, L., J. Z. Huang, J. G. Martinez, A. Maity, V. Baladandayuthapani, and R. J. Carroll (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association 105*(489), 390–400.

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

Zhu, H., P. J. Brown, and J. S. Morris (2012). Robust classification of functional and quantitative image data using functional mixed models. *Biometrics 68*(4), 1260–1268.

Zhu, H. and D. D. Cox (2009). A functional generalized linear model with curve selection in cervical pre-cancer diagnosis using fluorescence spectroscopy. In *Optimality: The Third Erich L. Lehmann Symposium*, Volume 57 of *Lecture Notes Monograph Series*, pp. 173–189. Institute of Mathematical Statistics, Beachwood, OH, USA.

Zhu, H., J. S. Morris, F. Wei, and D. D. Cox (2017). Multivariate functional response regression, with application to fluorescence spectroscopy in a cervical pre-cancer study. *Computational Statistics and Data Analysis 111*, 88–101.

Zhu, H., F. Versaceb, P. M. Cinciripini, P. Rausch, and J. S. Morris (2018). Robust and gaussian spatial functional regression models for analysis of event-related potentials. *NeuroImage 181*, 501–512.