Functional Regression

Jeffrey S. Morris

The University of Texas MD Anderson Cancer Center, Houston, Texas 77230;
email: jefmorris@mdanderson.org

Keywords

Abstract

Functional data analysis (FDA) involves the analysis of data whose ideal units of observation are functions defined on some continuous domain, and the observed data consist of a sample of functions taken from some population, sampled on a discrete grid. Ramsay & Silverman's (1997) textbook sparked the development of this field, which has accelerated in the past 10 years to become one of the fastest growing areas of statistics, fueled by the growing number of applications yielding this type of data. One unique characteristic of FDA is the need to combine information both across and within functions, which Ramsay and Silverman called replication and regularization, respectively. This article focuses on functional regression, the area of FDA that has received the most attention in applications and methodological development. First, there is an introduction to basis functions, key building blocks for regularization in functional regression methods, followed by an overview of functional regression methods, split into three types: (*a*) functional predictor regression (scalar-on-function), (*b*) functional response regression (function-on-scalar), and (*c*) function-on-function regression. For each, the role of replication and regularization is discussed and the methodological development described in a roughly chronological manner, at times deviating from the historical timeline to group together similar methods. The primary focus is on modeling and methodology, highlighting the modeling structures that have been developed and the various regularization approaches employed. The review concludes with a brief discussion describing potential areas of future development in this field.

1. INTRODUCTION

In recent decades, technological innovations have produced data that are increasingly complex, high dimensional, and structured. A large fraction of these data can be characterized as functional data, and has sparked the rapid development of a new area of statistics, functional data analysis (FDA). With functional data, the ideal units of observation are functions defined on some continuous domain, and the observed data consist of a sample of functions taken from some population, with each function sampled on a discrete grid. This grid could be fine or sparse, regular or irregular, common or varying across sampled functions. Most commonly, these functions are defined on a 1D Euclidean domain, but this paradigm also includes functions on higher dimensional domains such as 2D or 3D image data or time-space data, as well as functions observed on manifolds and other non-Euclidean domains. The genesis of this field was largely motivated by longitudinal data, growth curves, or time series for which t represents time, typically observed on sparse grids. Recently, technological innovations in various application areas have produced functions on other domains, including spatial domains, imaging domains, spectral domains, and even genomic locations, which have brought with them higher dimensional and more complex functional data. The key idea characterizing FDA is to think of each function as a single structured object rather than a collection of data points, which engenders a simplicity of thought that enables the building of models that can simultaneously handle complex structure both within functions and between functions.

The recognition of FDA as its own field was sparked in large part by the seminal textbook of Ramsay & Silverman (1997). This book introduced the philosophical thinking behind FDA, and described many aspects of FDA including registration, clustering, graphical displays, descriptive statistics, and various strategies for regression analysis. In their final chapter, they made a key observation that two factors are simultaneously at work in FDA: replication and regularization. Replication involves combining information across functions to draw upon their commonalities and make inferences on the populations from which they were sampled, whereas regularization involves the borrowing of strength within a function, exploiting the expected underlying structural relationships within a function to gain efficiency and interpretability.

This review focuses on functional regression, the area of FDA that has received the most attention in applications and methodological development. The goal of this review is to describe the historical development of work in this area, and to highlight the role of regularization and replication in the various modeling approaches that have been developed. First, there is an overview of basis functions, key building blocks for functional regression methods, followed by a description of the general strategies used for regularization in functional regression using various basis functions. Next, the methodological developments for each type of functional regression are described in turn: (*a*) functional predictor regression (scalar-on-function), (*b*) functional response regression (function-on-scalar), and (*c*) function-on-function regression. Each section begins with a general discussion of the role of replication and regularization in the given type of regression, and then steps through the methodological development in a historical manner, at times deviating from the historical timeline to mention closely related work or describe a cluster of work done by the same researchers. The primary focus is on modeling and methodology, not so much on theory or computations, highlighting the modeling structures that have been developed and the various regularization approaches employed. The end of each section contains a summary and a list of publicly available software for fitting some of the methods discussed.

While much work has focused on point estimation, some papers have highlighted the inferential capabilities of their methods, and in these cases inferential approaches are discussed, which include point-wise and joint intervals, variable selection, and hypothesis testing, based on either

asymptotic results, bootstrap, or Bayesian modeling. Most methods involve linear regression and independently sampled 1D Euclidean functions, but others are more advanced, handling nonlinear regression, correlated functions, and functions on higher dimensional domains, such as images. The review highlights these advances.

2. EXAMPLE DATA SETS

In this section, three functional data sets are introduced to highlight some different types of functional data and provide a concrete backdrop for the discussion of functional regression methods. For the first two, functions are defined on a 1D Euclidean domain, with the first [diffusion tensor imaging (DTI) data set] having relatively simple, smooth functions sampled on a moderately sized grid ($\approx$50–100 observations per curve), and the second (mass spectrometry data set) having complex, spiky functions sampled on a high-dimensional grid ($\approx$12,000 observations per curve). The third (glaucoma data set) is included to illustrate functional data on non-Euclidean domains, with functions defined on a 2D partial spherical domain. **Figure 1** contains representative plots of
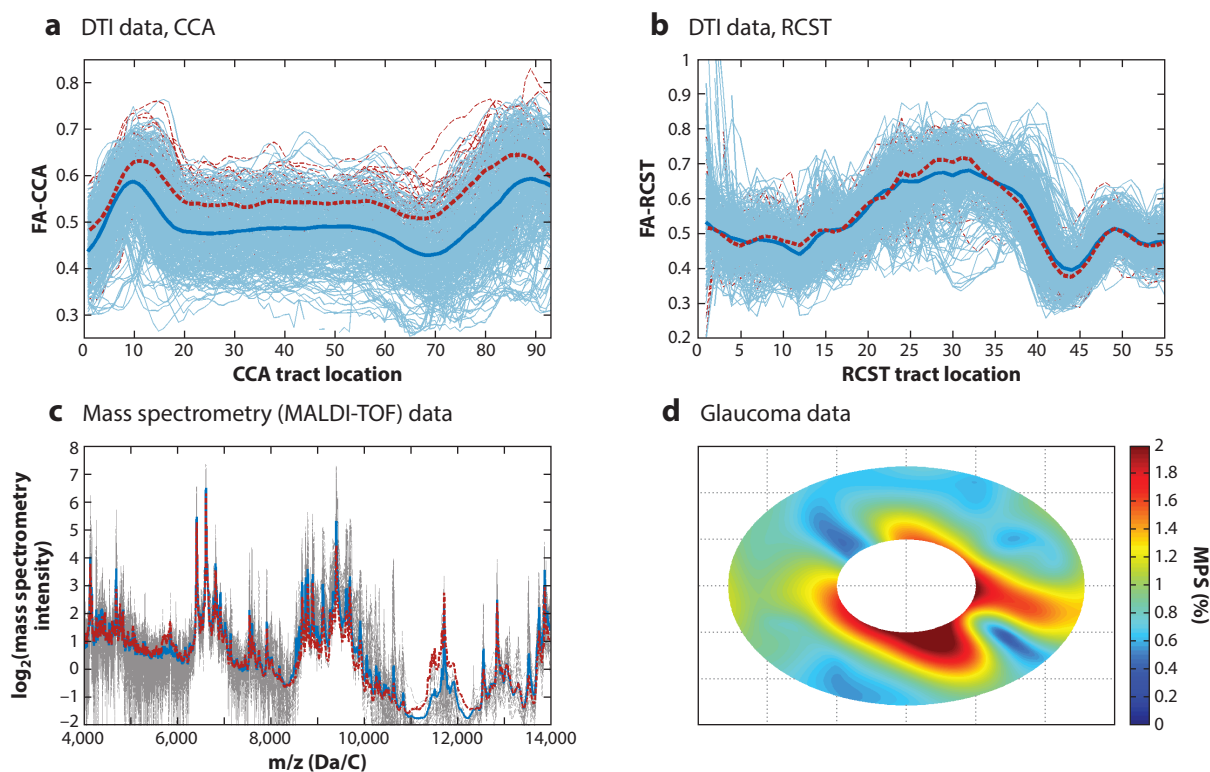
Figure 1

Example data sets. (*a*) Fractional anisotropy (FA) measurements from corpus collusum (CCA) and (*b*) right corticospinal tract (RCST) from the diffusion tensor imaging (DTI) data set. Curves from multiple sclerosis (MS) patients are plotted in thin blue, and from control patients in thin dashed red, with mean MS curve in thick blue and mean control curve in thick dashed red. (*c*) Log spectral intensities from the mass spectrometry data set, plotting part of spectrum from 4,000*D* to 14,000*D*. Gray lines are plotted spectra from 20 pancreatic cancer (*solid*) and 20 control (*dashed*) patients, with mean spectra for pancreatic cancer (*red*) and control (*blue*). (*d*) Polar azimuthal projection of partial spherical maximum principal strain (MPS) function for the left eye from one subject, age 66 years and Caucasian, in the glaucoma data set, under 45 mm Hg of interocular pressure.

these data sets. One can refer back to these example data in thinking about the different functional regression methods and associated modeling issues described in this review.

2.1. Diffusion Tensor Imaging Data (Simple, Smooth Functions, Moderate Grid)

This data set contains fractional anisotropy (FA) tract profiles from DTI data collected at Johns Hopkins University and the Kennedy-Krieger Institute, and is freely available as part of the *refund* R package. The data set involves 382 DTI scans from a total of 142 subjects, 100 with multiple sclerosis (MS) and 42 healthy controls. From each DTI scan, FA profiles are obtained along the corpus callosum (FA-CCA) and right corticospinal (FA-RCST) tracts, yielding functions $Y_i^{CCA}(t)$ and $Y_i^{RCST}(t), i = 1, \ldots, N = 382$. Additionally, for each subject the data include their MS status $X_i^{MS} = 1$ if MS, 0 if control, plus their Paced Auditory Serial Additional Test (PASAT) score X_i^{PASAT}, a continuous measurement of cognitive function relating to speed and flexibility in processing auditory information. As seen in **Figure 1***a,b*, these functions are relatively simple and smooth and sampled on a moderately sized grid of 93 (for CCA) or 55 (for RCST) positions.

This data set can be used to illustrate all three types of functional regression described in this review, and has been used in numerous papers in current literature (Greven et al. 2010, Goldsmith et al. 2011a, Staicu et al. 2011, Ivanescu et al. 2012, Scheipl et al. 2014). To predict MS status from the FA-CCA, one could perform functional predictor regression (Section 4) by regressing MS status X_i^{MS} on $Y_i^{CCA}(t)$ in Equation 2. Using functional response regression (Section 5), one could assess differences in the mean FA-CCA profiles between MS and control patients by regressing $Y_i^{CCA}(t)$ on X_i^{MS} in Equation 9, with functional coefficient $B(t)$ representing the difference in mean FA-CCA profile between MS and control subjects, and then either testing whether $B(t) = 0$ or using point-wise or joint intervals to characterize for which regions of t the groups differ. If multiple profiles per subject are modeled, then random effect functions should be added to the model to form a functional mixed effect model (FMM), such as Equation 10, to account for the correlation between curves from the same subject. After regressing $Y_i^{CCA}(t)$ on X_i^{MS}, one could predict future subjects' MS status from their FA-CCA profiles using functional discriminant analysis (Section 5) as an alternative to functional predictor regression. One could also regress $Y_i^{CCA}(t)$ on X_i^{PASAT} to estimate a functional slope $B^{PASAT}(t)$ that measures the linear association between FA-CCA at position t and PASAT score, or using semiparametric FMMs (Section 5), one could nonparametrically model a surface $f^{PASAT}(x, t)$ that measures the smooth effect of PASAT (x) on FA-CCA at position t in the curve. To relate the FA-CCA and FA-RCST curves to each other, one could use function-on-function regression (Section 6) to regress $Y_i^{CCA}(t)$ on $Y_i^{RCST}(s)$, using Equation 11 to estimate and perform inference on the resulting regression surface $B(s, t)$.

2.2. Mass Spectrometry Data (Complex, Spiky High-Dimensional Functions)

This data set contains mass spectra from a study to find serum proteomic markers for pancreatic cancer performed at the University of Texas MD Anderson Cancer Center. As described in Koomen et al. (2005), blood serum was taken from 139 pancreatic cancer patients and 117 healthy controls and run on a mass spectrometry instrument to obtain proteomic spectra $Y_i(t)$ for each sample $i = 1, \ldots, 256$. Each spectrum is a spiky function whose peaks correspond to proteins and peptides with molecular mass (actually mass per unit charge, m/z) roughly equal to t, and whose spectral intensity approximates its relative abundance. Each spectrum is observed on a high-dimensional grid of size $T = 12,096$ for a range of m/z values from 4,000D to 40,000D. As shown in **Figure 1***c*, these functions are spiky and irregular, and illustrate a type of complex, high-dimensional functional data that is not the focus of most functional regression development,

but is increasingly encountered in many applications. For data like these, flexibility, adaptiveness, and computational efficiency are important considerations for methodology development.

This data set can be used to illustrate both functional predictor regression and functional response regression, and has been used in studies by Morris et al. (2008), Zhu et al. (2011, 2012), and Morris (2012). Frequently, mass spectrometry data are analyzed by performing peak detection and quantification and then surveying the peaks, but this approach discards information in the spectra as no peak detection algorithm is perfect. Morris et al. (2008) and Morris (2012) demonstrated that modeling the entire spectrum as a function can yield more power for detecting differentially expressed proteins than peak detection–based approaches. Functional response regression can be performed to regress $Y_i(t)$ on X_i^{cancer} (= 1 if from cancer patient, = 0 if control) using Equation 9 and then assessing for which t is $B(t)$ different from 0, where $B(t)$ represents the difference between cancer and control mean spectra. After fitting this model, functional discriminant analysis can be done to classify future subjects based on spectra from their serum samples. Alternatively, one could predict cancer status using functional predictor regression to regress cancer status X_i^{cancer} on spectra $Y_i(t)$ using one of the methods based on Equation 2.

2.3. Glaucoma Data (Functions on a Partial Spherical Manifold)

This data set is from a glaucoma study investigating biomechanics of the scleral surface of the eye (Fazio et al. 2013). Using custom instrumentation, investigators were able to induce a fixed amount of interocular pressure and measure scleral displacement continuously around the eye. Their hypothesis is that individuals at higher risk of glaucoma, including older subjects and African Americans, should have lower displacement, especially closer to the optic nerve. To study this, the displacement for left and right eyes for 19 donors was measured for a series of nine interocular pressures (7 mm, 10 mm, 15 mm, 20 mm, 25 mm, 30 mm, 35 mm, 40 mm, and 45 mm Hg), yielding functions $Y_i(\theta, \phi), i = 1, \ldots, N = 19 \times 2 \times 9 = 342$ representing the maximum principal strain (MPS) at 120 circumferential locations $\phi \in (0°, 360°)$ and 120 meridional locations $\theta \in (9°, 24°)$, where $\theta = 0°$ corresponds to the optic nerve. These 2D functions are defined on a fine grid (size 14,400) on a partial spherical manifold. **Figure 1d** plots a polar azimuthal projection of one of the functions. For each eye, the subject's age X_i^{age} and race $X_i^{AA} = 1$ if African American, 0 otherwise, is also obtained.

Functional response regression of $Y_i(\theta, \phi)$ on X_i^{AA} yields functional coefficient $B^{AA}(\theta, \phi)$ that can be investigated to test whether MPS differs by race and, if so, in what regions of the eye, and including X_i^{age} yields a functional slope coefficient $B^{age}(\theta, \phi)$ to assess how MPS changes with aging. FMMs (Section 5) need to be used to account for correlation between eyes from the same subject and for longitudinal correlations across measurements for different pressure levels in the same eye. If not wanting to assume linearity in age, semiparametric FMMs can be used to estimate $f(age, \theta, \phi)$ that captures the effect of age on MPS as a smooth nonparametric effect varying across locations within the eye, indexed by (θ, ϕ). One such modeling effort for these data is presented in Lee et al. (2014). This example illustrates how functional regression approaches can simultaneously handle complex within- and across-function structure and be applied to functions on non-Euclidean domains.

3. REPLICATION, REGULARIZATION, AND BASIS FUNCTIONS

Inherently, any statistical method combines information across observations in some way to make inference on the populations from which they were sampled. Most basic methods involve either

replication, combining information across sampled units, or regularization, combining information within sampled units. FDA involves both.

To expound, replication is the concept exploited in most commonly used statistical models. For example, given replicate sampled pairs of responses and predictors, regression models combine information across these pairs to discover the relationship between them and build a predictive model. Replication is also the underlying concept involved when accounting for correlation among the sampled observations, e.g., induced by multilevel designs, in which case information is combined together in a careful way depending on the relative variabilities and sample sizes at each stochastic level, for example, using mixed effect models or Bayesian hierarchical models. In FDA, we consider each function as the sampled unit, and replication involves how our models combine information across functions so we can make inferences on the populations from which they were drawn.

Regularization involves borrowing strength across observations within a function that are expected to be similar to each other, exploiting their expected regularity even though they are discretely observed. This is the key idea behind (single function) nonparametric regression or kriging in spatial data analysis, where estimation of a process at some location t is improved by borrowing strength from nearby observations within the function using some smoothing mechanism. For most FDA methods, regularization involves global smoothing, implying an assumption that similarity within the function is strictly related to distance between the observations. Global smoothing may not be sufficient to adequately capture the internal structure of some functions, especially more complex high-dimensional ones such as the mass spectrometry data in Section 2. For example, some functions are spatially heterogeneous, meaning some parts of the function are smoother (i.e., more similar across t) than others, and in other cases there are distant parts of the function that are correlated with each other. In these cases, more general forms of regularization are needed. Regularization is closely tied to the modeling of correlation for Gaussian data. It can be applied in the likelihood of the functional data to capture the within-function correlation, or in penalties or priors for functional coefficients in the regression model to improve their estimation. The key idea is that we believe that by exploiting this regularity, we can gain more efficient and interpretable estimators, better predictions, and at times calculations that are potentially faster and more stable.

Replication and regularization can be done separately using a two-step approach or together using a unified modeling approach. For example, in functional response regression when functions are all sampled on a common grid, a two-step approach would involve fitting separate regression models for each t and then smoothing the functional coefficients, or smoothing each individual function and then performing separate regression models on a grid of values for t. This approach is convenient, but may be inferior to unified modeling approaches that perform replication and regularization jointly because the borrowing of strength across functions can lead to more efficient regularization, and the regularization can more accurately inform how information is combined across functions (i.e., replication). This review highlights cases in which unified modeling is done and evidence of these benefits have been shown.

Basis functions are the building blocks of FDA and typically determine the mechanism by which regularization is done. Each basis function defines a linear combination among the locations within the function, which in effect induces a correlation among those functional regions sharing loadings of high magnitude, and thus establishing a specific framework for borrowing strength. Their use also allows finite numbers of basis coefficients to yield estimates and inference in an infinite-dimensional function space. Some of the most commonly used basis functions in FDA are splines, Fourier series, wavelets, and principal components (PCs), and each is suited for functions with certain characteristics. Splines are well suited to modeling smooth functions, and given their usual fitting approach work better when the dimension of the observational grid is not

too high. Fourier series transform the functions into the frequency domain and work well for functions with stationary periodic characteristics. Wavelets are multiresolution basis functions that decompose a signal in a dual frequency-time domain, and their local support make them ideal for irregular functions with spikes, discontinuities, and other nonstationary features. Both Fourier and wavelet bases possess fast algorithms for calculating basis coefficients and reconstructing the function from the basis coefficients when functions are regularly sampled on a grid of size T, with Fourier $O(T \log T)$ and wavelets $O(T)$, so they are well suited for functions sampled on high-dimensional, regular grids. PCs are basis functions empirically estimated from the data using eigendecompositions. They are well suited for functions with simple structure leading to rapidly decaying eigenvalues, and are able to capture distant, nonlocal correlations within the functions. I discuss some considerations and developments regarding the use of PC for FDA below. Kernels can also be thought of as basis functions that define a weighting over which local mean estimation or regression is done, and they are suitable for smooth functions when global bandwidths are used and can be used with spatially heterogeneous functions when varying bandwidths are used. Lin et al. (2004) showed the equivalence between certain kernel and spline models.

Regularization of functions is accomplished by applying one of three operations to its basis coefficients: (*a*) truncation, (*b*) roughness penalties (L2 penalization), or (*c*) sparsity (L1 penalization). For ordered bases, truncation involves the elimination of all bases after a certain point. This is most commonly utilized in principal component analysis (PCA), keeping only the first few PCs explaining the most variability, but can also be used with Fourier and wavelet bases ordered by increasing frequency if low-pass filters are applied, with orthogonal polynomials if polynomials above a certain order are removed, or with unpenalized regression splines for which the use of limited prespecified knot locations can be viewed as a form of basis truncation.

Roughness penalties penalize the integrated squared second derivative of the function, which involves some type of generalized L2 penalty on the basis coefficients. Given a single smoothing parameter, this leads to global smoothing and a ridge regression-like structure. Fitting can be done by penalized least squares (PenLS), a Bayesian model with Gaussian priors, or a linear mixed effect model (LMM), which is especially convenient when using Demmler-Reinsch basis functions (Demmler & Reinsch 1975) that are orthogonalized with respect to the roughness penalty. Smoothing splines involve natural cubic spline bases with knots at each observed data point while incorporating this roughness penalty, and penalized splines involve roughness penalties but use fewer knots. Other L2 penalties have been used in place of pure roughness penalties, with Marx & Eilers (1996) proposing an L2 penalty on the first difference of B-spline coefficients in their P-spline construction, and Ruppert et al. (2003) inducing a type of L2 penalization by assuming independent and identically distributed (iid) Gaussian distributions for truncated power series spline coefficients.

Sparsity in the basis coefficients is related to L1 penalization, which can be induced through Bayesian modeling using sparsity priors that involve stochastic search variable selection (SSVS) (George & McCulloch 1993) or nonlinear shrinkage [Bayesian least absolute shrinkage and selection operator (LASSO) (Park & Casella 2008), normal-gamma (Griffin & Brown 2010), horseshoe (Carvahlo et al. 2010), or generalized beta (Armagan et al. 2011)], or through PenLS using L1 sparsity penalties such as the LASSO (Tibshirani 1991) or one of its many variants. L1 penalization is a standard approach for wavelet regression, and it leads to a type of adaptive regularization that removes noise while preserving dominant local features, accommodating spatial heterogeneity (Donoho & Johnstone 1995). Sparsity is also the regularization mechanism of free-knot splines (Gervini 2006), and it has been used as an alternative to truncation in PC regression when one does not want to assume that the PC dimensions explaining the most variability are necessarily the most predictive (Joliffe 1982).

PC(A): principal component (analysis)

PenLS: penalized least squares

LMM: linear mixed effect model

SSVS: stochastic search variable selection

LASSO: least absolute shrinkage and selection operator

PCs play a key role in FDA and are worthy of some additional discussion. In PCA, the basis functions are empirically generated by estimating the within-function covariance surface, combining information across the replicated functions, and are thus an example of replication informing regularization. They have been a major workhorse in FDA, especially for simple smooth functions, in which case it is common for a few PCs to explain a high percentage of the total variability. Care must be taken, however, in using PCA for very complex, high-dimensional functional data for which the decay rate in the eigenvalues is slow, especially when the number of functions is small. In certain high-dimensional, low-sample-size settings, PCs have been shown to be inconsistent (Hall et al. 2005, Baik & Silverstein 2006, Jung & Marron 2009, Lee et al. 2010, Shen et al. 2013), as in these cases there is insufficient sample size to empirically estimate the complex covariance surface characterizing the functional process. These problems can be at least partially mitigated at times by using one of the functional principal component (fPC) approaches that incorporate regularization in the PC calculation and effectively reduce the dimensionality of the covariance surface, although it is currently not clear how to decide in a given situation whether these methods are sufficient to lead to well-estimated eigenfunctions.

Various approaches have been used to estimate regularized or sparse empirical basis functions and compute the corresponding scores using fPCA. Two-step approaches have been proposed, whereby one first smooths the individual curves and then computes PCs, or computes the PCs and then smooths the eigenfunctions (Ramsay & Silverman 2005, chapter 8). Besse & Ramsay (1986) smoothed sparse functional data to interpolate and then applied PCA. James (2002) first smoothed the raw curves using natural cubic splines and then computed the PCs on the spline coefficients. Johnstone & Lu (2009) talked about sparse PCA and described an approach that involves using sparsity in the wavelet space to obtain sparse PCA, and Mostacci et al. (2010), Røislien & Winje (2013), and Meyer et al. (2013) all used approaches involving wavelet denoising followed by PC calculation. Silverman (1996) introduced a unified approach for calculating PCs while penalizing roughness.

Other advances in PCA have allowed the calculation of PC scores in settings with functions observed on sparse and variable grids and for multilevel models with multiple sources of stochastic variability. In modeling repeated samples of functions within dose groups, Staniswallis & Lee (1998) used kernels to smooth the covariance surface from which the fPCs are estimated, using quadrature to estimate the fPC scores when functions are sampled on sparse, irregular grids that vary across functions. Yao et al. (2005a) developed a similar approach, principal analysis by conditional expectation (PACE), which borrows strength across sparse curves using a kernel-smoothed covariance matrix to estimate the eigenfunctions and computing the individual PC score as best linear unbiased predictors (BLUPs) of a LMM. Staniswallis & Lee (1998) performed separate fPC decompositions at two levels of a multilevel functional model, dose group and subject within dose group. In a more general context, Di et al. (2009) presented an approach, multilevel fPCA (ML-fPCA), that obtains separate fPC scores at each level of multilevel functional data based on smoothed covariance estimates constructed from method-of-moment estimates at the various levels of the multilevel models for densely sampled functions, and later generalized this to sparsely observed functions (Di et al. 2014). ML-fPCA was applied to estimate separate fPCs for random intercepts and slopes for longitudinally observed functional data by Greven et al. (2010), and Zipunnikov et al. (2011) described how to compute the fPCs in ML-fPCA models for extremely high-dimensional data, making use of the fact that $N \ll T$. Bayesian (Crainiceanu & Goldsmith 2010, Goldsmith et al. 2011b, McLean et al. 2013b) and bootstrap-based (Goldsmith et al. 2013) approaches have been developed to account for the variability in PC score estimation when fitting functional linear models (FLMs).

4. FUNCTIONAL PREDICTOR REGRESSION

Most work in functional predictor regression is based on a variant of the FLM, first introduced by Ramsay & Dalzell (1991) and first written in its commonly encountered form by Hastie & Mallows (1993):

$$Y_i = B_0 + \int X_i(t)B(t)dt + E_i, \qquad (1)$$

where $Y_i, i = 1, \ldots, N$ is a continuous response, $X_i(t)$ a functional predictor, $B(t)$ functional coefficient, B_0 intercept, and $E_i \sim N(0, \sigma^2)$ residual errors. For non-Gaussian responses, many have worked with the generalized FLM (GFLM), first introduced by Marx & Eilers (1999) for exponential family responses as

$$g\{E(Y_i)\} = B_0 + \int X_i(t)B(t)dt \qquad (2)$$

for some link function $g(\cdot)$.

Ramsay & Silverman (1997, section 10.4; 2005, section 15.3) discussed using truncated basis function expansions for $X_i(t) = \sum_{k=1}^{K_X} X_{ik}^* \phi_k(t)$ and $B(t) = \sum_{k=1}^{K_B} B_k^* \psi_k(t)$. If we let $\mathbf{X}_i^* = [X_{i1}^*, \ldots, X_{iK_X}^*]'$, $\mathbf{B}^* = [B_1^*, \ldots, B_{K_B}^*]'$, $\boldsymbol{\phi}(t) = [\phi_1(t), \ldots, \phi_{K_X}(t)]'$, and $\boldsymbol{\psi}(t) = [\psi_1(t), \ldots, \psi_{K_B}(t)]'$, then the functional regression terms can be written

$$\int X_i(t)B(t)dt = \mathbf{X}_i^* J_{\phi,\psi} \mathbf{B}^* \qquad (3)$$

$$= \mathbf{X}_i^{**} \mathbf{B}^*, \qquad (4)$$

where $\mathbf{X}_i^{**} = \mathbf{X}^* J_{\phi,\psi}$ and $J_{\phi,\psi} = \int \boldsymbol{\phi}(t)\boldsymbol{\psi}(t)'dt$, and thus the FLM simplifies into a linear model of dimension K_B. Notes: (*a*) If the same orthogonal basis functions are used for $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, then $J_{\phi,\psi} = I$ and $\mathbf{X}_i^{**} = \mathbf{X}_i^*$. (*b*) If all $X_i(t)$ are observed on the same regular grid t of size T, then one can write $J_{\phi,\psi} = \boldsymbol{\Phi}\boldsymbol{\Psi}'$, where $\boldsymbol{\Phi} = [\boldsymbol{\phi}(t_1), \ldots, \boldsymbol{\phi}(t_T)]$ and $\boldsymbol{\Psi} = [\boldsymbol{\psi}(t_1), \ldots, \boldsymbol{\psi}(t_T)]$.

As discussed in Section 3, regularization can be done in the basis space either through truncation, roughness penalties, or sparsity. For functional predictor regression, this regularization can be applied to the functional coefficients $B(t)$ and/or predictors $X_i(t)$. The regularization of $B(t)$ accomplishes several purposes: It (*a*) reduces collinearity in the regression fit, (*b*) increases interpretability of coefficient estimates, and (*c*) potentially increases estimation and prediction efficiency by making use of the functional nature of the data to borrow strength across t. Best choice of basis and regularization strategy is likely to vary across data sets, depending on the characteristics of the functions and the true functional coefficient. The use of basis functions and regularization across $X_i(t)$ has two potential purposes: to (*a*) reduce measurement error in predictors $X_i(t)$, which for Gaussian responses increases estimation efficiency and for nonlinear models involving non-Gaussian outcomes reduces measurement-error-induced bias (Carroll et al. 1995), and (*b*) accommodate functional predictors with sparse and/or irregular grids that vary across sampled functions.

Most existing methodological development in functional predictor regression has followed along this general strategy laid out by Ramsay & Silverman, with various choices of basis functions and regularization approaches. For each method we review, I point these out and relate the work to Model 3 above. Other model components have been introduced by others that lead to more flexible models beyond what is presented in Ramsay & Silverman (2005), including the accommodation of non-Gaussian or correlated responses, the inclusion of other nonfunctional fixed or random effect terms in the model, extension to multiple functional predictors, the accommodation of functions on irregular grids, extensions to nonlinear functional predictor

models, and the introduction of specific approaches for variable selection or inference. I highlight these developments, taking a historical context to describe the methods roughly in the order they were developed, grouped by type of basis functions used.

4.1. Principal Component Analysis for $X_i(t)$ and $B(t)$

The original work on FLM was done using multivariate analysis techniques for functions sampled on a common grid, with principal component regression (PCR) (Kendall 1957), partial least squares (PLS) (Wold 1966), or ridge regression used to reduce collinearity in the resulting high-dimensional multiple regression. In PCR, orthogonal empirically determined PC bases estimated from the decomposition of $X'X$ are used for both Φ and Ψ. According to Joliffe (1982), the original work on PCR (e.g., Kendall 1957) suggested a variable selection approach to regularization, supposing that the PCs explaining the largest amount of variability in $X_i(t)$ may not necessarily be the most important in predicting Y_i. However, because in many cases the first few PCs appear most important, over time many researchers have regularized by truncating at the first few PCs explaining a fixed proportion of total variability. Both strategies have been used in practice. Other methods extending PCR have been developed for function predictor regression.

Cardot et al. (1999) fit a FLM with PCR through truncation and discussed some theoretical results. Müller & Stadtmüller (2005) introduced functional PCR (fPCR) for GFLM, using smoothed fPC bases for both $X_i(t)$ and $B(t)$ with regularization done by PC truncation. The basis functions were estimated by decomposing a kernel-smoothed estimate of the covariance function after subtracting off a kernel-smoothed estimate of the overall mean, while removing white noise errors. When the functions are irregularly and sparsely sampled, the PC scores Y_i^* are computed using PACE (Yao et al. 2005a), which borrows strength across sparse curves to estimate the covariance matrix and estimates individual PC scores as BLUPs of a LMM.

In a multilevel functional data setting with multiple curves per subject, Crainiceanu et al. (2009) used ML-fPCA (Di et al. 2009) to estimate the multilevel PC scores, using a truncated set of subject-level scores as the predictors X_i^{**} in Model 4, effectively using subject-level PCs as bases for both ϕ and Ψ and regularizing by truncation. They used a restricted likelihood ratio test (RLRT) (Crainiceanu & Ruppert 2004) to select the truncation parameter, and also discussed a fully Bayesian Markov chain Monte Carlo (MCMC)-based version to obtain estimates and inference. Crainiceanu & Goldsmith (2010) developed a Bayesian modeling approach for PCR using WinBUGS, modeling the PC scores and eigenvalues stochastically (but still conditioning on estimated eigenvectors as fixed), which they showed leads to more accurate inference.

Holan et al. (2010) developed Bayesian methods to classify subjects based on a functional predictor that is a nonstationary time series. They computed the spectrogram, a time-varying Fourier spectrum, to transform the time series to the time-frequency dual space, and then constructed a logistic GFLM with the spectrogram image as a predictor, using PC for the basis functions and SSVS for regularization. Randolph et al. (2012) introduced a method for PCR for which regularization is done by a structured L2 penalty that incorporates presumed structure directly into the estimation process through a linear penalty operator, leading to a weighted ridge regression in the PC space.

4.2. Splines for $X_i(t)$ and/or $B(t)$

Hastie & Mallows (1993) pointed out that standard multivariate approaches such as PCR do not take the ordering inherent to the functional data into account and mentioned the growing importance of functional data modeling. They introduced the FLM (Model 1), and used penalized

splines for $B(t)$ while assuming all functions were sampled on a common grid t. Others have also incorporated spline bases for regularization in the FLM or GFLM settings.

Marx & Eilers (1999) introduced penalized signal regression (PSR) for the GFLM that uses B-splines for $B(t)$ that are penalized by the P-spline first difference L2 penalty introduced in Eilers & Marx (1996). Designed for functional predictors on a common grid t, their approach does not transform X ($\Phi = I_T$) but uses the B-spline design matrix for Ψ, resulting in $X^{**} = X\Psi'$ in Model 4. They subsequently extended this work in various ways, adding multiple linear and additive scalar fixed effects and linear random effects to the model [generalized linear additive smooth structures (GLASS)] (Eilers & Marx 2002), accommodating multidimensional signals such as images [multidimensional penalized signal regression (MPSR)] (Marx & Eilers 2005), and allowing prespecified weights for adaptive smoothing [space-varying penalized signal regression (SPSR)] (Li & Marx 2008).

James (2002) fit a GFLM using natural cubic spline bases to represent the predictor functions $X_i(t)$ with regularization done by the truncation inherent in the knot selection. To account for measurement error and accommodate varying sampling grids across functions, they specified a measurement error model $X_i = X_i^* \Phi_i + \epsilon_i$, where X_i and Φ_i are vectors with the functional predictors and natural cubic spline basis functions evaluated on the observational grid t_i, and X_i^* a vector containing the spline coefficients for a common set of knots.

Zhang et al. (2007) presented a spline-based FLM with periodic spline bases used for both Φ and Ψ, and regularization by roughness penalties. Like James (2002) and Müller & Stadtmüller (2005), their approach accommodated irregularly sampled functions and adjusted for measurement error in $X_i(t)$. Crambes et al. (2009) presented a smoothing spline–based FLM, and proved some theoretical results; Maronna & Yohai (2013) presented a robust version of this model that is insensitive to outliers. Yuan & Cai (2010) introduced a general reproducing kernel Hilbert space approach to functional linear regression regularized by roughness penalties, implemented using smoothing spline–like kernels.

PSR: penalized signal regression

GLASS: generalized linear additive smooth structures

MPSR: multidimensional penalized signal regression

SPSR: space-varying penalized signal regression

DWT: discrete wavelet transform

4.3. Wavelets for $X_i(t)$ and $B(t)$

Although not explicitly describing their model as functional regression, Brown et al. (2001) introduced a Bayesian, wavelet-based approach for fitting FLMs to data on a common equally spaced grid t. They used orthogonal wavelet bases to represent both $X_i(t)$ and $B(t)$, and used SSVS to select a subset of important wavelet coefficients. Their approach effectively used the orthonormal wavelet transform matrix for both Φ and Ψ, but coefficients were actually calculated using the O(T) pyramid-based discrete wavelet transform (DWT) algorithm, allowing the approach to scale up to functional data on very large grids. As mentioned above, the use of wavelets and sparsity leads to adaptive regularization, making it well-suited for modeling spatially heterogeneous functions with local features such as spikes. Subsequent work involving wavelet representations for $X_i(t)$ and $B(t)$ includes that of Wang et al. (2007), Malloy et al. (2010), and Zhao et al. (2012). Wang et al. (2007) predicted a dichotomous response using a Bayesian probit regression model, first denoising the predictor function $X_i(t)$ using a Bayesian LASSO to induce L1 penalization on the wavelet coefficients, and then performing SSVS on these coefficients to adaptively regularize $B(t)$. Malloy et al. (2010) fit lag time models using wavelets with regularization done by SSVS and included scalar fixed and random effect predictors to model correlated responses from a nested design. Zhao et al. (2012) modeled using wavelets and regularized using the LASSO as a sparsity prior.

4.4. General Bases for $X_i(t)$ and/or $B(t)$

Ratliffe et al. (2002a) presented methods for the FLM using common basis transform Φ for $X_i(t)$ and $B(t)$, with $X_i^{**} = X_i^* \Phi \Phi'$ and regularization done by truncation. They used Fourier basis coefficients, but presented their method for any general basis, extending to a logistic FLM presented in Ratliffe et al. (2002b). Ogden et al. (2002) transformed 2D image data into 1D variance functions, to which they fit a FLM using Fourier bases for both $X_i(t)$ and $B(t)$, regularizing by roughness penalty and fitting by PenLS.

Zhu & Cox (2009) developed a group-LASSO approach for performing variable selection across multiple potential functional predictors in a GFLM. Given regularly sampled functional predictors, they assumed a common truncated orthonormal basis transform for both $X_i(t)$ and $B(t)$. Zhu & Vannucci (2010) and Lee & Park (2011) utilized sparsity in the basis coefficients for regularization. Zhu & Vannucci (2010) modeled dichotomous outcomes using a Bayesian probit model, using general orthogonal basis functions for both $X_i(t)$ and $B(t)$, truncating at a large number and then performing SSVS to select among them. Lee & Park (2011) introduced a FLM for Gaussian responses that utilizes a common general basis transform for both $X_i(t)$ and $B(t)$ and regularized via one of a number of sparsity penalties on the coefficients, including LASSO, adaptive LASSO, or smoothly clipped absolute deviation (SCAD). Fan & James (2014) also used the group-LASSO to perform variable selection across multiple functional predictors, using a common orthogonal basis transform for $X_i(t)$ and $B(t)$, in their case for Gaussian responses.

James et al. (2009) introduced a method [functional linear regression that's interpretable (FLiRTI)] that encourages sparsity in the $B(t)$ for interpretability, accomplished via L1 penalization to shrink the coefficients toward sparsity in a prespecified derivative. They used piecewise constant basis functions for $B(t)$, but mentioned their approach can be used for any high-dimensional basis expansion, including splines, Fourier, and wavelets. Goldsmith et al. (2014) introduced a Bayesian method to predict a Gaussian response from an image predictor, using a combination of Ising prior and Markov random field prior to encourage significant regions of the coefficient surface to cluster and borrow strength from each other.

4.5. Combinations of Principal Components and Splines for $X_i(t)$ and $B(t)$

A number of papers have used PCs and splines together to perform regularization. One approach has been to represent the functional predictor $X_i(t)$ using splines, and then perform a PC decomposition of the spline coefficients. After fitting unpenalized cubic splines to the functional predictors, James (2002) performed a PC decomposition on the matrix of spline coefficients $X^* = [X_1^*, X_2^*, \ldots, X_N^*]$, and used this basis to further transform the $X(t)$. In the form of Model 2, they effectively used natural spline bases for Φ and eigenvectors of the spline bases for Ψ, and fit their model using the expectation maximization (EM) algorithm and weighted least squares (WLS), with the spline coefficient eigenvalues as the weights.

Reiss & Ogden (2007) used a similar strategy in a method that combines the ideas of PCR and PSR (Marx & Eilers 1999). For Gaussian outcomes, their model effectively transforms $X_i(t)$ using a $T \times K_x$ B-spline design matrix Φ, and then uses a PC decomposition on $\mathbf{X}\Phi$ to estimate a truncated set of eigenfunctions Ψ, with $X^{**} = X \Phi \Psi$ in Model 4. Regularization is done via roughness penalties (or P-spline penalties) applied to either the fPC (fPCR$_C$) or the regression coefficients (fPCR$_R$). They presented a similar strategy, but replaced PC bases with PLS bases (fPLS$_C$, fPLS$_R$), empirical bases that maximize correlation between $X_i(t)$ and Y_i instead of simply explaining variability in $X_i(t)$. Reiss & Ogden (2010) extended the fPCR$_R$ approach to handle exponential family responses and 2D image predictors. Radial basis B-splines were used

to ensure radial symmetry in the images. They also presented inferential methods, including bootstrap-based simultaneous confidence bands and using RLRT (Crainiceanu & Ruppert 2004) to test the functional coefficient alternative versus a constant null.

An alternative approach is to transform the functional predictors $X_i(t)$ using PCA while parameterizing the coefficient function $B(t)$ using splines. Cardot et al. (2003) introduced a two-step method that performs PCR and then uses B-splines with a roughness penalty to smooth the resulting $B(t)$. Crainiceanu & Goldsmith (2010) presented a Bayesian GFLM with PCs for $X_i(t)$ and B-splines for $B(t)$, regularized using the random walk prior described by Lang & Brezger (2004) on the spline coefficients, and modeling the PC scores and eigenvalues stochastically. Goldsmith et al. (2011b) developed a variational Bayes approximation for this model, and found this was sufficient for estimation but not necessarily for inference. Goldsmith et al. (2011a) described a frequentist variant of this method, called penalized functional regression (PFR), using PCs for Φ and truncated power series for Ψ, and regularizing by L2 penalties on the truncated power series spline coefficients as presented in Ruppert et al. (2003). Their approach can handle sparsely observed, irregular, or multilevel functional data using PACE or ML-fPCA to obtain PC scores to use in the GFLM as predictors, and also includes nonfunctional linear predictors. Instead of truncation they suggest keeping a large number of PCs, making the primary purpose of their PC decomposition the handling of irregularly sampled functions rather than dimension reduction. Note that when used with regularly sampled data and keeping all PCs, their approach is like PSR (Marx & Eilers 1999), but using truncated power series and iid Gaussian L2 penalties in place of the B-splines with first difference penalties. Swihart et al. (2013) presented functional hypothesis testing procedures based on RLRT for comparing functional versus constant effects and selecting among multiple functional predictors in this setting.

Goldsmith et al. (2012) extended this approach to handle data with repeated measurements in both the response and functional predictor by adding scalar random effects to the GFLM. Their specific model includes iid random effects (e.g., to model nested data), but clearly could be applied using other more complex random effects structures as well. Their approach does not account for the correlation among replicate functions in the calculation of the PC scores, but Gertheiss et al. (2013) presented an alternative approach that does. Using an extension of ML-fPCA (Di et al. 2009) for longitudinal functional data (LFPCA) (Greven et al. 2010), they computed separate eigenvectors for random intercepts and slopes, estimated a shared PC score between them using a regression approach, and then carried this PC score forward into the GFLM.

4.6. Nonlinear Functional Predictor Models

The previously described methods assumed linear functional effects. A number of methods for nonlinear functional predictor regression have been proposed.

Li & Marx (2008) extended the PSR method of Marx & Eilers (1999) to include general additive polynomial terms such as $\int \{X_i(t)\}^2 B_2(t) dt$. Yao & Müller (2010) allowed full quadratic functional predictors,

$$Y_i = B_0 + \int X_i(t) B_1(t) dt + \int\int X_i(t) X_i(s) B_2(t, s) dt ds + E_i. \qquad (5)$$

Note that this yields greater flexibility than the PSR method described by Li & Marx (2008), which only allows the diagonal cross product in the second term. They used PC decompositions to estimate empirical basis functions for $X_i(t)$, and used these as basis functions for $X_i(t)$, $B_1(t)$, and $B_2(t, s)$, with regularization by truncation. Their approach generalizes to full polynomials of any order. Yang et al. (2013) extended the spectrogram-based method of Holan et al. (2010) to

this penalized quadratic regression setting, using a Bayesian modeling approach with SSVS across the PC dimensions for regularization.

In the first nonparametric extension of the FLM, James & Silverman (2005) introduced a method called functional adaptive model estimation (FAME) that extends projection pursuit regression to functional predictors and exponential family outcomes. The model is given by

$$g\{E(Y_i)\} = B_0 + \sum_{k=1}^{K} f_k \left\{ \int X_i(t) B_k(t) \right\} dt, \tag{6}$$

where f_k is a smooth function of unspecified form. They used natural cubic splines to represent $X_i(t)$, $B_k(t)$, and $f_k(\cdot)$, and regularized using roughness penalties. Their model was extended to included multiple functional predictors.

Various researchers have incorporated nonlinearities in functional predictor regression using a special case of projection pursuit, single-index models, involving $f(\int X_i(t) B(t) dt)$ for some smooth $f(\cdot)$. Li et al. (2010) used single-index models to model interactions of scalar and functional predictors, using complete orthogonal basis representations for the $X_i(t)$. Marx et al. (2011) used single-index models to extend the PSR of Marx & Eilers (1999) in the setting of 2D functional predictors, using tensor B-splines for the predictor surface and 1D B-splines for the index function, regularizing using P-spline penalties. Fan & James (2014) allowed separate additive single-index terms for p functional predictors, using a common orthogonal basis for $X_{ij}(t)$ and $B_j(t)$ and another basis for $g_j(\cdot)$, $j = 1, \ldots, p$, using truncation for penalization and group-LASSO to perform variable selection across the functional predictors.

Müller & Yao (2008) introduced functional additive models (FAMs), which extend the fPC regression approach of Yao et al. (2005b) to nonlinear models that involve additive nonparametric functions of the PC scores, using the model

$$Y_i = B_0 + \sum_{k=1}^{K_x} f_k(Y_{ik}^*) + E_i, \tag{7}$$

where Y_{ik}^* are the PC scores for a truncated set of K_x PCs, and f_k is a general function smoothed using kernel approaches. Zhu et al. (2014) introduced a variant of this approach that did not perform truncation over PCs, but instead used COSSO regularization, a form of L1 penalization, based on an assumption that the function lives in a particular reproducing kernel Hilbert space. This allows dimensions explaining less total variability in $X_i(t)$ but that are more important for predicting Y_i to be included in the regression.

McLean et al. (2012) presented a functional generalized additive model (FGAM) that extends generalized additive models (GAMs) (Hastie & Tibshirani 1986) to the generalized functional predictor regression framework for noise-free functions observed on a common grid t:

$$g\{E(Y_i)\} = B_0 + \int f\{X_i(t), t\} dt + E_i(t), \tag{8}$$

where g is a general link function for an exponential family and $f(x, t)$ is a smooth functional additive regression surface. They parameterized this 2D surface using tensor B-splines, regularized by P-spline-type L2 penalization, and fit the model by penalized iteratively WLS with generalized cross validation (GCV) to estimate the smoothing parameters. McLean et al. (2013) introduced a Bayesian FGAM for sparsely observed functions on irregular grids. Effectively extending the model presented by Crainiceanu & Goldsmith (2010), they handled the measurement error and unequally spaced $X_i(t)$ using PC decompositions, updating the PC scores within the MCMC conditional on the eigenfunctions, eigenvalues, and mean curves. They used tensor product B-splines with random walk penalties on the coefficients as in Lang & Brezger (2004).

4.7. Summary of Functional Predictor Regression Models

There has been a great deal of work in functional predictor regression, especially in the past 10 years. Although some methods include nonlinear effects of some sort, most methods work with a version of the Gaussian model or GFLM, Models 1 or 2, respectively. The key differences among the methods are their choice of basis to represent the predictors $X_i(t)$ and/or the functional coefficient $B(t)$, with the most common choices being PCs, splines, wavelets, or some combination thereof, and their approach for regularization. No single basis function approach is likely to be superior in all settings, as each makes sense for particular types of data, as discussed in Section 3. New methods for choosing among various potential basis functions and assessing which regularization approaches work best would be a welcome addition to the literature, and would help users assess which approach to use given a particular data set.

4.8. Available Software for Functional Predictor Regression

A number of the methods discussed in this section have publicly available software. The methods from Ramsay & Silverman's (2005) textbook are contained in the R package *fda* and are also available in Matlab, as described in an accompanying computational textbook (Ramsay et al. 2009). The R package *fpca* implements the sparse PCA approaches of James et al. (2000), and R and Matlab code for James & Hastie (2001) and FLiRTI (James et al. 2009) are available at **http://www-bcf.usc.edu/~gareth/research/Research.html**. The PACE software (Yao et al. 2005a) and its applications can be downloaded at **http://www.stat.ucdavis.edu/PACE/**, and the software from Müller & Stadtmüller (2005) are available at **http://anson.ucdavis.edu/~mueller/data/spqr.html**. The software for a number of methods can be found in the *refund* R package, which started with *fpcr* (Reiss & Ogden 2007, 2010) and later was appended with the packages *pfr* (Goldsmith et al. 2011a), *lpfr* (Goldsmith et al. 2012), *peer* (Randolph et al. 2012), *lpeer* (Kundu et al. 2012), *wnet* (Zhao et al. 2012), and *fgam* (McLean et al. 2013). The software for the Bayesian methods of Goldsmith et al. (2010) can be found at **http://www.biostat.jhsph.edu/~ccrainic/software.html**.

5. FUNCTIONAL RESPONSE REGRESSION

Functional response regression involves the regression of functional responses on a set of scalar predictors. Given a sample of functional responses $Y_i(t_j), i = 1, \ldots, N; j = 1, \ldots, T_i$ and scalar predictors $X_{ia}, a = 1, \ldots, p$, a general linear functional response regression model is given by

$$Y_i(t_j) = \sum_{a=1}^{p} X_{ia} B_a(t_j) + E_i(t_j), \tag{9}$$

where functional coefficient $B_a(t)$ represents the partial effect of predictor X_a on the response at position t. The goal of functional response regression is often estimation of $B_a(t)$ and then either testing whether $B_a(t) = 0$ or assessing for which t is it true that $B_a(t) \neq 0$. The $E_i(t)$ are the curve-to-curve residual error deviations, frequently assumed to be iid mean zero Gaussians with covariance $S(t_1, t_2)$, whose structure describes the within-function covariance. Sometimes the curve-to-curve deviations are split into the sum of a curve-level random effect function $U_i(t_j)$ and white noise residual errors $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, with the primary benefit being the ability to obtain denoised estimates of the individual curves as $\hat{Y}_i(t) = \sum_a X_{ia} \hat{B}_a(t) + \hat{U}_i(t)$. If these random effect functions are represented by a truncated basis representation $U_i(t) = \sum_{k=1}^{K_u} U_{ik}^* \phi_k(t)$, and it is assumed that $\mathbf{U}_i^* \sim \mathcal{N}(0, S^*)$ where $\mathbf{U}_i^* = (U_{i1}^*, \ldots, U_{iK_u}^*)'$ and S^* is a $K_u \times K_u$ covariance matrix

for the basis coefficients, then this implies that $S(t_1, t_2) = \boldsymbol{\phi}(t_1)S^*\boldsymbol{\phi}(t_2)' + \sigma^2 I(t_1 = t_2)$, where $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_{K_u}(t))$. Thus, the form of the within-function covariance is determined by choice of basis and basis space covariance matrix S^*.

The concepts of replication and regularization enter into functional response regression modeling in various ways. Replication is involved in the regression and in accounting for potential between-function correlation, which is discussed below in Sections 5.5–5.7. Regularization of the functional coefficients $B_a(t)$ increases interpretability, potentially increases estimation and precision accuracy, and allows interpolation to values of t between sampled grid points. Regularization of the curve-level random effect functions $U_i(t)$ can have similar benefits. As with functional predictor models, this regularization is typically done using either truncation, roughness penalties, or sparsity in some chosen basis space. Further, the inclusion of some within-function correlation structure in $S(t_1, t_2)$ induces regularization, as it defines a manner by which strength is borrowed within the function in estimation and inference. This affects estimation of $B_a(t)$ when taken into account using weighted least squares for which the within-function covariance assumptions determine the weights (Hart & Wehrly 1986, Lin et al. 2004, Wu & Liang 2004, Morris & Carroll 2006, Krafty et al. 2008, Reiss et al. 2010, Staicu et al. 2010), as is typical for proper regression analyses of correlated data. This borrowing of strength is related to the idea of seemingly unrelated regression (Zellner 1962). Further, inference on $B_a(t)$ is affected by the within-function covariance assumptions, especially joint inference such as confidence bands (Lee & Morris 2014).

Much of the existing work on functional response regression can be related back to Model 9. Methods differ in the scope of their models, their mode of regularization of $B_a(t)$, and the assumptions made on the within-function covariance $S(t_1, t_2)$. We highlight these aspects of each method, focusing on the modeling, but sometimes including brief discussions of inferential capabilities of the models if they are emphasized. Some methods work for sparse, variable grids between functions, whereas others require regular, fine, and/or common grids across functions. Some approaches are able to model correlations between functions, and for these methods we highlight their capabilities and manner of capturing this correlation. Most methods in the literature are for 1D functions on Euclidean domains and involve linear modeling (in X), but some methods handle 2D functions like images and functions on non-Euclidean manifolds or allow nonlinearities (in X), and these are also highlighted. Most methods are designed for functions sampled on small-to-moderate-sized grids, although some have been designed to scale up to enormously sized functional data, e.g., the mass spectrometry data set in Section 2.

5.1. Growth Curves

Early work focused on estimation of the mean function $\mu(t)$ $(X_{ia} = 1)$ from an iid sample of curves, which has been called growth curve analysis in the classic setting where t represents time. Rao (1958) talked about the use of PCs to represent $\phi(t)$ in the curve-to-curve deviations, in which case the form of S is determined by the estimated PCs. Laird & Ware (1982) discussed LMMs, which can directly be used to fit growth curves when linear parametric forms are used for the mean function and curve-to-curve deviations, frequently through orthogonal polynomials. Hart & Wehrly (1986) nonparametrically estimated the mean of a sample of curves using kernel smoothing with parametric assumptions for S, e.g., an AR(1). They found that the optimal bandwidth could be smaller or larger depending on the autocorrelation, demonstrating that the within-function covariance assumptions affect the regularization of the functional mean estimation. Rice & Silverman (1991) estimated the mean of a sample of curves nonparametrically using splines with roughness penalties, with S parameterized using a fPC decomposition involving a roughness penalty in PC calculation, with regularization by truncation.

Barry (1995) introduced a Bayesian model with Gaussian processes for both the overall mean curves and curve-to-curve deviations, which is like a smoothing spline with separate roughness penalties for the mean curves and deviations. Shi et al. (1996) represented the mean curve with B-splines regularized by truncation inherent in the knot selection, and S parameterized by a PC decomposition of the B-spline coefficients plus white noise. Zhang et al. (2000) proposed a semiparametric model that adds parametric fixed effects to a model with a nonparametric mean function, using periodic spline bases for the mean curve, regularizing by roughness penalty. They represented S parametrically using an Ornstein-Uhlenbeck process, and performed global testing for the difference of two mean curves.

James et al. (2000) represented the mean curve using B-splines with regularization by truncation inherent in the knot location, represented the curve-to-curve deviations with B-splines, modeled the basis space covariance matrix S^* as unstructured or through PC decompositions, and included white noise errors. Martinez et al. (2010) implemented a Bayesian version of this approach using roughness penalties for the mean curve and curve-to-curve deviations, and using reversible jump MCMC (RJMCMC) to perform Bayesian model averaging over the number of PCs in the model. Rice & Wu (2001) also used regression splines with fixed knots to represent the overall mean function and curve-to-curve deviations, both regularized by the truncation inherent in knot placement, and the covariance S^* left unstructured.

Wu & Zhang (2002) used local polynomials to represent the mean function and curve-level random effect functions, with independent but heteroscedastic residual errors. Yao et al. (2005a) kernel-smoothed the overall mean function and then represented S using fPC obtained by performing a PC decomposition of the kernel-smoothed covariance matrix after subtracting off white noise errors. For functions sampled on a common grid, Gervini (2006) represented curve-level random effect functions using a PC decomposition, with the fPCs and overall mean function both adaptively regularized using free-knot splines.

Bigelow & Dunson (2007) presented a Bayesian approach that fit the overall mean and curve-level random effect functions using truncated linear splines; regularizing using sparsity priors at knot locations, with the between-spline-coefficient covariance S^* assumed diagonal but heteroscedastic; and using RJMCMC for model fitting. Thompson & Rosen (2008) also presented a Bayesian approach in which they characterized the overall mean and curve-level random effect functions using B-splines and regularized using sparsity priors applied to large set of candidate knots, with S^* left unstructured with an inverse Wishart prior. Fox & Dunson (2012) presented a Bayesian approach that expressed the mean function and curve-level random effects using multiresolution Gaussian processes, adaptively regularized through SSVS across potential multiresolution partitions. S was constrained to be the sum of multiresolution piecewise, weighted exponential kernels plus white noise. Storlie et al. (2013) presented a Bayesian approach that expressed the mean and curve-specific functions as Gaussian processes using a covariance kernel based on Bernoulli polynomials that has connections to smoothing splines, with this kernel determining the structure of S up to four scalar variance components.

Ogden & Green (2010) projected the observed curves into the wavelet space to estimate mean functions for curves sampled on a fine, common grid. The mean function was regularized by inducing sparsity in the wavelet space through multiplicity-adjusted hypothesis tests for each wavelet coefficient, and, as described by Morris & Carroll (2006), the wavelet-space covariance S^* was diagonal but heteroscedastic across wavelet coefficients, which leads to nonstationary features in S, allowing heteroscedasticity and varying degrees of smoothness across t. Similarly, Bunea et al. (2011) also modeled functional data on a fine, common grid and used a flexible sparse basis like wavelets to represent the mean function, regularizing via sparsity achieved by hard thresholding rules. Staicu et al. (2012) introduced methods for estimating the mean curve for functions with

skewed distributions, using a parametric distribution with shape and scale parameters varying over *t*, using *t*-copulas to account for within-function correlation.

5.2. Joint Bands and Multiple Testing

Various researchers have proposed joint confidence/credible bands with global functional coverage for the mean functions in growth curve models. Li & Hsing (2010) and Degras (2011) dealt with densely sampled functions and used local linear kernel approaches, and Bunea et al. (2011) used orthonormal basis functions. Ma et al. (2012) dealt with sparsely sampled functions and used piecewise constant splines. Storlie et al. (2013) presented approaches using B-splines and Gaussian processes. Using an approach described by Ruppert et al. (2003), joint credible bands can be constructed from MCMC samples of any Bayesian functional model or from bootstrap samples based on a Gaussian approximation. Crainiceanu et al. (2012) used this approach to produce bootstrap-based joint bands for the difference of two mean functions, considering various parametric and nonparametric bootstrap strategies and regularizing point-wise estimates of the mean functions through penalized splines. This approach has also been used for joint inference on functional regression coefficients in the general Bayesian FMM framework (Meyer et al. 2015, Lee et al. 2014, Lee & Morris 2014, Zhang et al. 2015). Joint bands can be inverted to yield functional inference (e.g., see simultaneous band scores, *SimBaS*, in Meyer et al. 2015) that controls global type I error rate. Additionally, for Bayesian methods posterior probabilities can be computed that have connections to local false discovery rate (FDR) (e.g., see Morris et al. 2008), and can be used to flag regions of functional regression coefficients while controlling FDRs.

5.3. Functional Analysis of Variance

Some papers have generalized growth curves to multiple populations $[X_{ia} = \delta_a(i)]$, which can be viewed as functional regression with discrete predictors, or functional analysis of variance (FANOVA). Barry (1996) extended previous work to test for differences in mean curves, using random intercepts to induce compound symmetry structure in *S*. For growth curves sampled within each of a number of doses, Staniswallis & Lee (1998) estimated mean functions across a set of doses using a complete, orthonormal basis expansion in *t*, represented as the inner product of the basis functions and separate nonparametric smooth functions of dose for each basis function. They represented *S* using fPCs, with an eigendecomposition done on kernel-smoothed estimates of the covariance matrices, plus white noise errors, and used Bonferroni-corrected ANOVA tests to perform inference. Spitzner et al. (2003) constructed FANOVA tests using PCs or Fourier bases to represent 2D functional responses, and Abramovich & Angelini (2006) constructed FANOVA tests based on wavelet representations of the functions.

5.4. General Functional Response Regression

Many of the methods for functional growth curves and FANOVA described above could, in principle, be generalized to the functional response regression setting, regressing the function on a general design matrix **X** as in Model 9, with some additional software development required. Some researchers have developed methods with the general Model 9 specifically in mind. Faraway (1997) described a two-step approach in which they smooth or interpolate the observed curves, and then perform point-wise regressions for each *t*, ignoring the within-function correlation, and then perform inference using a bootstrap or multivariate techniques based on an eigendecomposition

of the empirically estimated curve-to-curve deviations. Ramsay & Silverman (1997) first proposed separate point-wise regressions for each t with no regularization, and second discussed regularizing the functional regression coefficients through roughness penalties and basis function modeling. Their focus was estimation, not inference, and they assumed white noise errors and no curve-level random effects, so they effectively used $S(t_1, t_2) = \sigma_e^2 I(t_1 = t_2)$. Wu & Chiang (2000) fit a general linear functional response regression model, assuming iid errors and no random effect functions, and used kernel smoothing to regularize the functional coefficients, and Chiang et al. (2001) proposed smoothing spline regularization for the same model.

Reiss et al. (2010) studied Ramsay and Silverman's approach to general functional response regression for curves sampled on a common, fine grid, with basis representations (splines) for the observed curves and functional coefficients and regularization done by roughness penalties, while ignoring within-function correlation. They called this method ordinary least squares (OLS), and compared it with an approach where the within-function covariance was estimated in the basis space in an unstructured fashion and accounted for in estimation by WLS. They also introduced automatic smoothing parameter estimation methods and inferential techniques for this model. Shi et al. (2007) specified a general functional response regression model using Gaussian processes to capture the within-function covariance structure S. They used a multistep approach to fit the model, representing each observed function using a K_y-dimensional B-spline decomposition, performing a regression for each B-spline coefficient, and then fitting a Gaussian process to the empirical curve-to-curve deviations. Krafty et al. (2008) worked with Model 9 using PC decompositions to capture the within-function covariance. Functional coefficients were modeled by smoothing splines with roughness penalties and estimated by iteratively reweighted least squares (IRLS), with inference done by point-wise confidence intervals. They mentioned that smoothing spline estimates of functional coefficients from correlated data are consistent but inefficient if the within-function correlation is ignored (Lin et al. 2004).

5.5. Modeling Correlated Functional Data and Functional Mixed Effects Models

All of the aforementioned methods assumed an independent sample of functions and cannot capture between-function correlation induced by the experimental design. This correlation is present, for example, in multilevel functional models for which multiple functions are observed for each subject or within clusters, such as the DTI data set in Section 2, in longitudinally observed functional data where functions are observed serially in time or some other variable like pressure in the glaucoma data set of Section 2, or in spatially correlated functional data where functions are observed for each unit on a spatial grid or lattice. It is important to take the induced correlation between functions into account in estimation, to ensure observed functions are properly weighted in the regression through WLS, and in inference, to ensure that standard errors, point-wise bands, or joint bands properly account for all sources of variability in the data.

One way to capture this correlation is to specify a between-function covariance structure R in the curve-to-curve deviations in Model 9 that is separable from the within-function covariance S, for example, assuming the total covariance is given by a tensor product $R \otimes S$, implying COV $\{E_i(t_1), E_{i'}(t_2)\} = R_{ii'} S(t_1, t_2)$. This strategy has been used in many time-space models in the spatial modeling literature, and it is assumed in the context of spatially correlated functional data by the separable version of the functional conditional autoregressive (CAR) model introduced by Zhang et al. (2015), which assumes a block CAR structure for R.

Alternatively, correlation between functions can be induced by adding random effect functions to the general linear functional response regression model (Model 9) to form FMMs. For example,

consider the following FMM that was introduced in this form by Morris & Carroll (2006):

$$Y_i(t_j) = \sum_{a=1}^{p} X_{ia} B_a(t_j) + \sum_{b=1}^{H} \sum_{l=1}^{M_b} Z_{ibl} U_{bl}(t_j) + E_i(t_j), \tag{10}$$

where H is the number of levels of random effects, Z_{ibl} are random effect covariates at level b with corresponding random effect functions $U_{bl}(t)$. Suppose the $U_{bl}(t)$ are iid (over l) mean zero Gaussians with covariance $Q_b(t_1, t_2)$ representing the within-function covariance structure at random effect level b. Although conditionally independent, these random effect functions can induce correlation between the functions through the structure of their design matrices.

To illustrate, let $\mathbf{Y}(t) = (Y_1(t), \ldots Y_N(t))'$, $\mathbf{B}(t) = (B_1(t), \ldots, B_p(t))'$, and Z_b be an $N \times M_b$ matrix with element (i, l) given by Z_{ibl}. For a given t, the marginal $N \times N$ covariance matrix for $\{\mathbf{Y}(t)|\mathbf{B}(t)\}$ is given by $\sum_b \mathbf{Z}_b \mathbf{Z}'_b Q_b(t, t) + S(t, t) I_N$, which introduces between-function correlation through the random effect design matrices $\mathbf{Z}_b$, with each component $\mathbf{Z}_b \mathbf{Z}'_b Q_b(t, t)$ capturing a different source of between-function variability. For example, by specifying a design matrix with $Z_{ibl} = 1$ only if curve i is from cluster l, one induces a compound symmetry correlation structure among all curves within the cluster because of the shared cluster-specific random effect function $U_{bl}(t)$. For longitudinally observed functional data, if d_i represents the longitudinal time point at which curve $Y_i(t)$ is observed, then by defining $Z_{ib'l} = d_i$ if from subject l, 0 otherwise, and with $U_{b'l}$ representing the random slope for subject l, this structure will capture serial correlation across the longitudinally observed curves. Notes: (*a*) These between-function correlations cannot be induced by curve-level random effects, i.e., if all random effect functions are indexed by curve, $U_{ibl}(t)$, which is an important distinction between Model 10 and the type of FMM introduced in Guo (2002). (*b*) When the structure of Q_b and S allow heteroscedasticity across t [i.e., $Q_b(t, t)$ and $S(t, t)$ vary over t], then the level of between-function correlation is allowed to vary over the functional index t. (*c*) When Q_b and S allow within-function correlation (i.e., are not diagonal), then the $N \times N$ marginal cross-covariance of $\mathbf{Y}(t_1)$ and $\mathbf{Y}(t_2)$ is given by $\sum_b \mathbf{Z}_b \mathbf{Z}'_b Q_b(t_1, t_2) + S(t_1, t_2) I_N$.

Additional between-function correlation can be induced in FMMs by specifying a separate between-function covariance P_b across the M_b random effect functions at level b, and then using tensor products $[P_b \otimes Q_b(t_1, t_2)$, Morris & Carroll 2006] or row tensor products $[P_b \odot Q_b(t_1, t_2) = \{P_b \otimes I(t_1 = t_2) + I_N \otimes Q_b(t_1, t_2)\}$, Scheipl et al. 2014] for the total covariance of $U_{bl}(t_1)$ and $U_{bl'}(t_2)$.

5.6. Different Types of Random Effects in Functional Response Regression Models

To clarify an important issue in this context, random effect model components are used in various ways in functional regression, and not all of them capture between-function correlation. First, in single-function spline smoothing, random effects on the spline coefficients are used as a mechanism of penalization to induce regularization within the function, which could be called random effects for penalization. These are called new style random effects by Hodges (2013, chapter 13), and are most familiar to spline smoothers. Second, in functional response regression the curve-level random effect functions defined in Model 9 capture within-function correlation in the curve-to-curve deviations, but again the purpose is strictly regularization, as these curve-level random effects, if independent, cannot induce correlations among the functions. Third, random effect functions can be defined for units of observations such as clusters or subjects for which multiple curves are observed. These could be called multilevel random effect functions, and serve multiple purposes. They capture multiple sources of between-function variability in the data and induce between-function correlation, so are involved in replication, determining the manner in which

information is combined across functions in performing functional response regression. They are also involved in regularization because within-function covariance assumptions for each level of random effect functions define how strength is borrowed within a function, which impacts estimation and inference of the functional coefficients. These are the type of random effect functions featured in the FMM in Model 10.

In classic LMMs, random effects are integrated out and fixed effects are estimated using marginal likelihood, leading to a WLS-like fitting procedure that reweights the observations according to the covariance structure induced by the random effects. As a result, inference on the fixed effects integrates over the populations sampled at the random effect levels, allowing inference to be made on the corresponding reference population. For some methods involving FMMs, the random effect functions are not integrated out of the model but instead conditioned upon in an additive multistep fitting procedure. In these cases, there is no integration over these sources of variability in estimation and inference of the fixed effect functions. Thus, these conditional random effect functions behave much like fixed effect functions, providing smoothed estimates for specific individuals but not allowing inference on the fixed functional regression coefficients to cover the reference population from which they were drawn.

5.7. Papers Involving Functional Mixed Effect Models

Brumback & Rice (1998) introduced the first published approach for modeling multilevel functional data, a smoothing-spline based method for nested or crossed curves. Their model contains mean functions at the group level, plus deviation curves at the subject and curve-within-subject levels, all represented by smoothing splines using the Demmler-Reinsch parameterization plus white noise errors. They split the random functions at each level (group, subject, curve) into the sum of fixed intercepts and slopes and random iid Demmler-Reinsch spline coefficients. The between-function correlation at the different levels is captured by these random spline coefficients, but the form of the subject-level (Q) and cycle-level (S) within-function covariance matrices is restricted, fully determined by the spline basis and roughness penalty up to a single variance component, the smoothing parameter.

Guo (2002) first introduced the term functional mixed effect model for a model that included arbitrary numbers of fixed effect functions and q curve-level random effect functions, all modeled as smoothing splines, and iid errors. Because this model includes only curve-level random effect functions, it cannot capture between-function correlation or accommodate multilevel models like the FMM presented above in Model 10. Splitting the random effect functions into linear and spline components, they modeled intercepts and slopes as random effects, and so their within-function covariance S is determined by the sum of a parametric part induced by the covariance of the random intercept and slope, a spline part whose form is fixed by the spline basis and penalty up to the scalar smoothing parameters, plus white noise. They fit their model using either PROC MIXED or a faster Kalman filter–based approach, and described point-wise confidence bands and functional testing based on RLRTs. Qin & Guo (2006) extended this approach to induce periodic assumptions on the fixed and random effect smoothing splines.

Morris et al. (2003) presented a Bayesian, wavelet-based method for nested functional data with three hierarchical levels, group, subject within group, and curve within subject, with fixed mean functions for each group and random effect functions at the subject and curve levels. They assumed diagonal covariance matrices for the wavelet coefficients of the random effect functions at the subject and curve levels, Q^* and S^*, respectively. The variance components were indexed by wavelet scale but not location, constraining the form of the within-function covariances Q and S to stationary-type structures. They presented a fully Bayesian model for estimation and inference.

Fixed effect functions were regularized using SSVS priors on the wavelet coefficients, and their estimation accounts for the between-function correlation induced by the random effect functions in a WLS-like fashion.

Morris & Carroll (2006) extended this approach to the general FMM (Model 10) with general design matrix X, multiple levels of random effect functions, plus curve-to-curve deviations, in a method intended for functional data sampled on a common fine grid. Using a basis transform modeling approach, they computed $T^* \approx T$ wavelet coefficients for each observed function, thus projecting the data into the wavelet space, and fit a version of the FMM in the wavelet space, with diagonal but heteroscedastic assumptions for Q_b^* and S^* across wavelet coefficients. Unlike Morris et al. (2003), who only let variance components vary across wavelet scale, in this work they allowed each of the T^* wavelet coefficients to have its own variance component. While limiting the form of the within-function covariances Q_b and S relative to unstructured covariances (which would not be feasible to fit in most settings), this structure is flexible enough to capture important nonstationarities such as heteroscedasticity and different degrees of within-function autocorrelation across t. This greatly speeds calculations, making the procedure $O(T^*)$ and allowing it to scale up to perform fully Bayesian analysis for enormously sized functional data sets such as mass spectrometry (Morris et al. 2008), quantitative image data (Morris et al. 2011, Fazio et al. 2013), or whole-genome analyses (Lee & Morris, 2014).

Fixed effect functions are adaptively regularized through sparsity priors in the wavelet space, and random effect functions are adaptively regularized through the wavelet-space Gaussian distributions, with coefficient-specific variances that behave like scale mixtures and thus have adaptive shrinkage properties (Morris & Carroll 2006). The random effect function and curve-to-curve deviation covariance structures involve respective tensor products of between- and within-function covariances $P_b \otimes Q_b$ and $R \otimes S$, respectively, which allows highly structured between-function covariances, e.g., for spatially correlated functional data (Zhang et al. 2015). Even when conditionally independent random effect functions are used ($P_b = I_{M_b}$, $R = I_N$), the multiple levels of functional random effects with general design matrices enable great flexibility in capturing between-function correlation, for example, modeling nested structures through random functional intercepts or modeling longitudinally observed functions through random functional linear coefficients that are parametric in time but nonparametric in the functional index (Lee et al. 2014). Further, by adding random effect levels with design matrices consisting of Demmler-Reinsch spline bases constructed for some continuous predictor X_i, the model can also incorporate nonparametric functional additive terms $f(X_i, t)$ in place of functional linear terms $X_i B(t)$ (Lee et al. 2014). The fully Bayesian modeling approach integrates over all stochastic levels, yielding WLS-like fixed effect function estimates and inference that account for the various sources of between- and within-function variability captured by the model. Joint credible bands can be computed from which functional hypotheses can be tested (Meyer et al. 2015), and point-wise posterior probabilities, interpretable as local FDR, can be computed to flag regions of the fixed effect curves as significant while controlling global FDR (Morris et al. 2008).

Morris et al. (2006) extended this method to deal with partially missing functions, i.e., with missing grid points for some of the functions, and Morris et al. (2011) extended this framework to 2D or 3D image data. While the method was introduced using wavelet bases, subsequent work (Morris et al. 2011, Meyer et al. 2015, Lee et al. 2014, Zhang et al. 2015) has described how the method can be applied using other basis functions such as Fourier bases, PC, and splines while using the same basis transform modeling strategy using lossless or near-lossless basis transforms. The forms of the within-function covariances at each level are determined by the choice of basis functions up to a set of T^* coefficient-specific variance components. Regularization is done by either truncation of bases using a joint basis truncation strategy introduced in Morris et al. (2011)

and/or through roughness or sparsity penalties in the basis space, both of which can be induced as special cases of a mixture prior consisting of a spike at 0 and a Gaussian scale mixture. Zhu et al. (2011) developed a robust version of this Bayesian FMM that uses heavier-tailed distributions for the likelihood and random effect distributions to produce functional coefficient estimates and inference that are insensitive to outliers. As discussed below, this framework can also be used to perform general function-on-function regression (Meyer et al. 2015) and analyze spatially correlated functional data (Zhang et al. 2015).

Antoniadis & Sapatinas (2007) presented a FMM with general fixed effect functions, curve-level random effect functions, and iid residual errors, using wavelet bases to represent the functions. Unlike Morris & Carroll (2006), who focus primarily on estimation and point-wise inference on the fixed effects, their main focus is on functional inference. Their random effect curves are assumed iid, with within-function covariance matrix S completely determined by the assumed Besov space up to a single scalar variance component. This, along with the iid error assumptions, limits its flexibility, but allows a simple functional hypothesis test for the significance of the random effects. They do not apply any regularization to the fixed effect functions.

Berhane & Molitor (2008) introduced a complex, multilevel FMM for growth curves of individuals nested within clusters. Their model included fixed effect functions for overall intercepts and covariate effects, plus functional random effect intercepts and slopes for covariate effects at both the cluster and random effect levels, with iid errors. The fixed and random effect functions were parameterized using natural cubic B-splines, with regularization done by truncation inherent in the knot selection. They used a two-step method to fit this model.

Aston et al. (2010) presented a PC-based FMM for functions sampled on common grids. After subtracting a point-wise (unsmoothed) mean function, they estimated the eigenfunctions based on the empirical covariance matrix, and then computed the PC scores for each curve, truncating at T^* coefficients, effectively projecting the data into the space spanned by the overall PCs. They next fit independent LMMs for each PC score, regularizing through sparsity by performing stepwise variable selection across the PC dimensions for each fixed effect function. Their model included a single level of random effect functions at the curve level.

Chen & Wang (2011) presented a spline-based FMM for sparse functional data with general functional fixed effect structures, curve-level random effects, and residual errors with a parametric covariance [e.g., AR(1)] that is heteroscedastic across t. They represented the functional coefficients, random effects, and the log of residual variances (varying over t) using spline bases with roughness penalties for regularization. They nonparametrically estimated the random effect functions and an unstructured covariance for the random effect spline coefficients S^* using an EM algorithm and accounted for the within-function covariance by updating the fixed effects using a WLS-like approach. With only curve-level random effects, their model cannot accommodate between-function correlation.

Greven et al. (2012) presented a model for longitudinally observed functional data $Y_{ij}(d_{ij}, t)$, with d representing the longitudinal index and t the functional index. Their model includes an overall mean function $\nu(d_{ij}, t)$ and three levels of random effect functions: random slope and random intercept functions at the subject level and a curve-level random effect function. Their primary focus is simply estimation of the ML-fPC at these three levels and their accompanying scores, which they do after subtracting off a 2D-kernel smoothed estimate of the mean function that does not take the between- or within-function covariances into account. Goldsmith & Kitago (2015) introduced a Bayesian method for general functional response regression for nested data. Their model has general fixed effect functions, represented by B-splines regularized by a roughness penalty, plus subject-level random effect functions and curve-to-curve residual error deviations. Their subject-level random effect functions are represented by penalized splines, with

their effective within-function covariance structure Q determined up to a scalar constant by the spline and penalty choice. The within-function covariance S for the curve-to-curve deviations is left unstructured and given an inverse Wishart prior. Estimation is done using variational Bayes.

5.8. Spatially Correlated Functional Data

Some papers have dealt specifically with functional response regression in settings where the observed functions are correlated according to some spatial grid. Within a nested multilevel functional model with correlated functions within subjects and subjects within groups, Baladandayuthapani et al. (2008) introduced a Bayesian method using truncated power series spline bases to represent the functions at each level of the hierarchy and capturing the correlation between curves using a Matern structure on the spline coefficients. Their within-function covariances at each hierarchical level Q_b consisted of the sum of random polynomial terms plus a spline part whose structure is determined by the choice of basis and penalty up to a scalar smoothing parameter. Staicu et al. (2010) presented an alternative approach to this model, using ML-fPC bases instead of splines at each level of the hierarchy and polynomials for the group mean functions, and estimating the spatial variogram using method-of-moments. The ML-fPCs yield flexibility in the within-function covariance assumptions at each level, and their use of WLS to update the group mean functions integrates over the different levels of variability in the model. Their fitting was done using a multistep approach that involves sequential estimation of the variogram, ML-fPC bases, group mean functions, and the residual error variances, yielding point estimates of each quantity. Zhou et al. (2010) also used ML-fPC at each level of the hierarchy, but instead of parametric mean functions, used polynomial splines to nonparametrically represent the mean functions and ML-fPC, with regularization by roughness penalties. They captured the between-function spatial correlation at the curve level by specifying a stationary covariance structure among the curve-level ML-fPC scores. Unlike Baladandayuthapani et al. (2008) and Staicu et al. (2010), with these assumptions their model allows nonseparable spatial correlations when projecting the correlation from the fPC space back to the function space. In the context of areal functional data, Zhang et al. (2015) introduced a functional CAR model. The separable version of their model was previously mentioned above, and they also introduced a more flexible nonseparable version that projects the functions to a chosen basis space and fits CAR models in the basis space with separate CAR parameters for each of T^* basis coefficients. This construction leads to a CAR model between the functions that allows the spatial correlation to vary across t, yet borrows strength across t in estimating the correlation as determined by the chosen basis functions. Embedded within the general FMM framework in Model 10, this allows functionally varying spatial correlation to be accounted for in general functional response regression.

5.9. Semiparametric Functional Mixed Effect Models

Most of the existing functional response regression methods involve fixed effect functions that are nonparametric in t, but linear in x. There is recent work extending the idea of GAMs (Hastie & Tibshirani 1986) to functional response regression, leading to a new class of semiparametric models for functional response regression that include terms that are either parametric or nonparametric in x. In the last chapter of Wood (2006), additive mixed models (AMMs) are described as an extension of Laird & Ware (1982) with generalized additive fixed effects and parametric random effects. If parametric linear forms are assumed in t for the fixed and random effects, then AMMs can be used to fit parametric functional AMMs to estimate terms such as $f(x, t)$, representing an effect of X on $Y(t)$ that is nonparametric in x but parametric in t. The models are fit using *mgcv* in

R, which utilizes *lme* to fit the LMMs, and it is mentioned that computational problems can arise when fitting very complex models.

Scheipl et al. (2014) described a generalization of AMMs to the functional response setting that allows $f(x, t)$ to be nonparametric in x and t. They introduced an elegant notation for specifying design matrices for an impressively general array of fixed effect function structures, including intercepts and linear or smooth nonparametric effects for scalar predictors with constant or functional (in t) effects, plus linear or smooth nonparametric effects for functional predictors. Their notation subsumes the fixed effect structures used in current literature. Their construction uses row tensor products of between- and within-function design matrices, each with their own roughness penalty that consists of a scalar smoothing parameter and specified penalty matrix. It is primarily designed to be used with penalized splines for the smooth effects and functional coefficients, each with their own separate roughness penalties for regularization. Their construction assumes iid errors but also allows multiple levels of random effect functions, whose covariance structure is a row tensor product (see definition above) and is completely determined by the choice of basis functions and specified between- and within-function penalties (serving as covariance matrices) except for two scalar variance components that are the between- and within-function smoothing parameters. Given N functions each sampled on a grid of size T, they simplify the entire model down to a large AMM or LMM of size NT that is fit using *mgcv* from Wood (2011), from which point-wise intervals and RLRT tests can be computed. The use of a single large model limits the size of data sets that it can feasibly accommodate, so as they mention it does not scale up well to functional data sets with enormous T, e.g., the mass spectrometry data set in Section 2, but works well for sparsely sampled functions and can accommodate unequal sampling grids between functions.

As described above, Lee et al. (2014) showed how the Bayesian FMM of Morris & Carroll (2006) and the additional methodological publications that followed can accommodate additive coefficients $f(x, t)$ that are nonparametric in x and t by adding a level of random effect functions with corresponding design matrix defined based on Demmler-Reinsch spline bases on x, with the variance component regulating the smoothing in x allowed to vary by basis coefficient $k = 1, \ldots, T^*$. By allowing the smoothing parameters to vary across the response basis coefficients, their approach allows the degree of smoothness in x to vary and borrow strength over t. Their Bayesian modeling approach allows full joint or point-wise inference to be done on the coefficient surface or linear functionals, e.g., derivatives. This method is designed for functions sampled on a common fine grid.

5.10. Spectral Domain Functional Data Analysis of Time Series

A number of papers apply FDA approaches to time series using spectral domain modeling, which effectively involves projecting the data to the Fourier domain and modeling there. Diggle & Wassel (2007) used Fourier domain modeling involving the log-spectrum for modeling of dense growth curves. Guo (2003) performed smoothing-spline ANOVA on log periodograms, with separate periodograms for partitions of time to allow locally stationary processes. They used tensor product spline bases to model the coefficients with roughness penalties for regularization, and accommodated general fixed effect specifications. Qin et al. (2009) applied a state-space modeling approach to scale this approach up to large data sets, and applied it to EEG data. Freyermuth et al. (2010) estimated growth curves in the spectral domain by representing the overall mean and curve-specific deviations of log-spectra using tree-structured wavelets, with adaptive regularization done by truncating to a limited number of wavelet coefficients, keeping the same set of wavelet coefficients for each curve and the overall mean. Krafty et al. (2011) fit a FMM in the spectral domain for stationary time series, with general fixed effect functions and handling two-level nested

designs in the random effects, with n_j time series from each of N subjects, and random effect spectra for each of the N subjects. They have q random effects per subject, each with their own covariance. The covariance was estimated by first estimating all of the subject-level random effect functions as smoothing splines, with a common smoothing parameter across subjects, and then estimating the covariance nonparametrically from these smooth estimates. Inference was done by point-wise confidence bands. Their approach can be extended to locally stationary time series by partitioning time and using the tensor product approach of Guo (2003). Martinez et al. (2013) applied the Bayesian wavelet-based image FMM of Morris et al. (2011) to spectrogram images summarizing time-varying spectral content, allowing functional response regression of nonstationary time series on multiple predictors of various types, while accounting for potential correlation among the time series induced by the experimental design through random effect functions.

5.11. Functional Discriminant Analysis Using Functional Response Regression

Although functional response regression models treat the function $Y_i(t)$ as the response and any class indicator X as the predictor, after fitting functional response regression models one can perform functional discriminant analysis to classify subjects and curves based on their functional data, in which case the functional response regression model serves to characterize a parametric joint density for the functions within each group. This has been done in various modeling contexts. Do & Kirk (1999) estimated the mean curve through penalized splines with roughness penalties, and used Fourier domain-smoothed PCs to model the within-function covariance matrices for each group. Hall et al. (2001) performed a PC decomposition using the combined data set including all groups, and constructed discriminant functions on the PC scores. James & Hastie (2001) used spline representations for the mean functions, and PCs for the errors to construct their discriminant function. Li & Yu (2008) introduced a method combining linear discriminant analysis and support vector machines. Bayesian approaches for functional discriminant analysis have the added benefit of being based on predictive distributions, integrating over the uncertainty of the model parameters rather than simply conditioning on estimates of them. Zhu et al. (2012) introduced Bayesian methods to perform functional discriminant analysis in the FMM framework of Morris & Carroll (2006). This allows the model to adjust for other covariates affecting the functions and enables classification based on a set of correlated functions. Methods were described for both the Gaussian and robust (Zhu et al. 2011) versions of the FMM, and for use with 1D functions or higher dimensional functions like images, using wavelets or any other chosen basis modeled using a basis transform modeling approach. Stingo et al. (2012) presented a Bayesian, wavelet-based method for functional discriminant analysis. They modeled in the wavelet space, using an Inverse Wishart prior on the covariance for each group to allow more general covariance structures than approaches assuming independence in the wavelet space.

5.12. Summary of Functional Response Regression Models

The literature and software for functional response regression is much less developed than that for functional predictor regression. A vast majority of the published papers simply involve estimation of a mean of a sample of functions, which has been called growth curve analysis in classic literature. These methods differ in their approach for smoothing the mean function, with different choices of basis functions or regularization approaches, and also in their approach for modeling the correlation over t in the curve-to-curve deviations, if they even model such correlation. Of the methods for general functional response regression, most do not handle correlation between curves, so are only suitable for independently sampled functions. Again, methods differ in choice

of basis functions and assumptions made on the curve-to-curve deviations. Methods that assume independent and identically distributed residual errors and no curve-to-curve deviations to capture the within-function correlation are not realistic for the FDA setting because the nature of the functional data suggests there should be correlation across t and often heteroscedasticity in the curve-to-curve deviations. These may yield reasonable point estimates but the inference may be questionable because of the incorrect independence and homoscedasticity assumptions. Many of the spline-based functional response regression models accommodate within-function correlation in the curve-to-curve deviations, but through the choice of basis function and penalty make assumptions on the within-function covariance that fixes its structure up to a constant, the smoothing parameter. These types of simplifying assumptions are necessary in single-function spline smoothing, but not in the FDA setting because the replicate functions allow the estimation of more rich structures in the covariance of the curve-to-curve deviations. Other methods accommodate more flexibility in these matrices, either modeling them with PC decompositions, unstructured covariance matrices, or parsimonious but still somewhat flexible assumptions induced by assuming independence in a chosen basis space. The covariance assumptions in the curve-to-curve deviations may be important, as they can affect the inference on the functional regression coefficients, and also strongly come into play when performing functional discriminant analysis.

There is a small but growing literature on methods that can model correlated functions in the functional response regression setting. This literature includes models for spatially correlated functional data, longitudinally observed functional data, and general FMMs that can include various levels and types of random effect functions that can induce correlation both between and within functions and take these correlations into account in constructing point estimates and inference for the functional coefficients. Most of the methods are intended for functional data consisting of smooth, simple functions sampled on a coarse grid, and they make modeling assumptions that are not flexible enough and/or computationally scalable to settings with complex functional data sampled on fine grids, for example, spiky functions such as mass spectrometry data (Morris et al. 2008) or quantitative image data such as functional magnetic resonance imaging (Morris et al. 2011). These complex, high-dimensional data often have richly structured within-curve covariance matrices, yet in many of these settings their complexity, along with the fact that $T \gg N$, raises questions about whether PCs can be well-enough estimated to be reliably used.

5.13. Available Software for Functional Response Regression

Only a few of the discussed functional response regression methods have available software. Stand-alone executable software for the Bayesian FMM methods introduced by Morris & Carroll (2006) is available at **https://biostatistics.mdanderson.org/SoftwareDownload/ SingleSoftware.aspx?Software_Id=70** for Windows, Linux, or OS X. With properly defined design matrices $\mathbf{X}$ and $\mathbf{Z}_b$, this stand-alone software can be directly used to implement the MCMC in the methods presented by Morris et al. (2006) and Morris et al. (2008), for 2D or 3D image responses presented by Morris et al. (2011), function-on-function regression presented by Meyer et al. (2013), and the semiparametric FMMs and longitudinally observed functional modeling approaches presented by Lee et al. (2014). The current version uses wavelet bases, although one can fit Bayesian FMMs using other basis functions if they compute the transforms and inverse transforms outside the package and input the data matrix in the basis space. Broken down into preprocessing, MCMC, and postprocessing modules, the executable is parallelizable, allowing it to be used with enormous data sets. As described by Herrick & Morris (2006), this method is at worst linear in the number of observations per function T, and even faster when joint compression (Morris et al. 2011) is used to perform truncation on the basis coefficients. The

method automatically produces a set of posterior summary statistics (including means, standard errors, point-wise and joint-credible bands, and point-wise posterior probabilities of prespecified effect sizes) for each fixed effect function and prespecified contrast of fixed-effect functions for each t or specified linear combinations across the t, plus saves the posterior samples so any other Bayesian estimation, inference, prediction, or diagnostics can be computed.

The smoothing spline FMM of Guo (2002) can be applied using a SAS macro at **http://www.jstatsoft.org/v43/c01**. The method for spatially correlated hierarchical functional data of Baladandayuthapani et al. (2008) has software available at **http://works.bepress.com/veera/5**. The fast function-on-scalar regression methods of Reiss et al. (2010) is implemented in the R package *fosr* in the *refund* package, along with an implementation of Fan & Zhang's (2000) two-step approach in the *fos2r* package.

Wood (2006) has implemented his GAM approaches using *mgcv*, which includes a package *gam* to fit GAMs, which uses *lme* in R to fit the LMMs. The R package *pffr* in *refund* contains a wrapper that sets up the models of Scheipl et al. (2014) in a way so they can all be fit using Wood's *mgcv* package in R.

6. FUNCTION-ON-FUNCTION REGRESSION

There has been comparatively little work on function-on-function regression done to date. Ramsay & Dalzell (1991) introduced the FLM, which is a function-on-function regression model with unconstrained surface coefficient $B(s, t)$:

$$Y_i(t) = B_0(t) + \int X_i(s) B(s, t) ds + E_i(t). \tag{11}$$

This setting combines the issues present in functional predictor and functional response regression settings, including regularization of the predictor function for denoising, reducing collinearity, and dealing with variable grids across functions; regularization of the coefficient surface in both dimensions in order to obtain more interpretable and efficient estimators; and regularization in terms of modeling the within-function correlation in the residual errors and taking this into account in estimation through WLS-type approaches and in estimation. Additionally, when the response curves are correlated, random effect functions or correlated errors may be necessary to take between-function correlation into account.

Ramsay & Dalzell (1991) fit Model 11 using piecewise Fourier bases for $B(s, t)$, whereas Besse & Cardot (1996) developed spline-based approaches. Ramsay & Silverman (1997) presented this model and discussed using basis functions $\Phi(s)$ and $\Psi(t)$ for $X_i(s)$ and $Y_i(t)$, respectively, in which case the estimator can be written $B(s, t) = \Phi(s)'\mathbf{B}\Psi(t)$, where $\mathbf{B}$ is a $K_x \times K_y$ matrix containing the coefficient surface in the basis space. They discussed various approaches to regularization, including no regularization, truncation of X and/or Y, and bivariate roughness penalties. All of these methods assumed iid residual errors, no random effect functions, and no modeling of within-function correlation in the errors.

Yao et al. (2005b) presented fPC-based methods for this model where $Y(t)$ and $X(s)$ are modeled using fPC decompositions with iid measurement errors. The PC scores for X and Y were computed by PACE (Yao et al. 2005a), and estimation of the functional coefficient surface $B(s, t)$ amounts to a series of independent simple linear models regressing estimated response PC scores on predictor PC scores. Wu & Müller (2011) used WLS to account for the correlation within the functions in estimating the regression coefficients, and Chen et al. (2011) adapted this approach to perform the regression of the sparse, noisy observations $Y_i(t_{ij})$ on the predictor function $X_i(s)$ rather than the estimated PC scores of the denoised estimator.

Ivanescu et al. (2012) extended the PFR of Goldsmith et al. (2011a) to the function-on-function setting, using PC expansions of the functional predictors while keeping many PCs, and regularizing the functional coefficient using P-splines. Their model includes scalar fixed effects, potentially multiple function-on-function regression terms, iid residual errors, and no random effect functions. Their approach accommodates sparse or densely sampled functions, with or without measurement error. Scheipl & Greven (2012) discussed identifiability issues in the context of this model.

Work has been done on two constrained versions of Model 11. One is the concurrent model (Ramsay & Silverman 2005), in which both X and Y are defined on the same domain t and the prediction of $Y_i(t)$ only depends on concurrently observed $X_i(t)$, i.e., $B(s, t) = B(t)$. This is a special case of the varying coefficient model (Hastie & Tibshirani 1993). Wu et al. (1998) described kernel-based methods with k functional predictors and iid errors, and the functional coefficients smoothed by kernel estimation, and Hoover et al. (1998) described smoothing spline and locally weighted polynomial approaches. Fan & Zhang (2000) described a two-step approach, first performing point-wise regressions and then smoothing the resulting functional coefficients. Liang et al. (2003) included measurement errors in the predictor functions using random effect functions, using unstructured covariance. The functional regression coefficients and random effect functions were both represented by B-spline bases, with penalization done by the truncation inherent in knot selection, and estimated using a WLS-like procedure. Wu & Liang (2004) discussed a similar approach that includes random intercepts and slopes for each subject and independent but heteroscedastic errors, and used local polynomial regression to regularize the functional coefficients. Huang et al. (2004) represented the functional coefficients using B-splines with regularization by the truncation inherent in knot selection, estimating the within-function covariance matrix as unstructured in the B-spline basis space. They accounted for the within-function correlation in inference, but used a least-squares procedure for estimation that ignores this correlation. Like James et al. (2000), Zhou et al. (2008) represented the mean functions for two longitudinal responses using B-splines regularized by roughness penalties and the respective random effect functions by PC of B-spline coefficients, and then computed the functional surface by regressing the PC scores on each other.

A second constrained model that has been studied is the historical functional linear model (hFLM) (Malfait & Ramsay 2003), for which $B(s, t)$ is lower triangular, allowing only $X_i(s)$ with $s \leq t$ to inform prediction of $Y_i(t)$. Malfait & Ramsay (2003) represented $B(s, t)$ using custom basis functions $\phi(s, t)$ that are tent-like piecewise linear functions on some grid, constructed to ensure the $s \leq t$ constraint, and fit a model with iid errors. Harezlak et al. (2007) considered a variety of regularization techniques for linear B-spline basis functions, including basis truncation, roughness penalties, and sparsity penalties (LASSO). Kim et al. (2011) presented a method using fPC decompositions for both Y and X and assuming iid errors. The functional coefficient is written out as a product of varying coefficent in t and predetermined basis function such as Fourier, splines, or PCs in s, which after integration of s out of the model within the domain $(0, t)$, leads to a varying coefficient model in t. Eigendecompositions of the marginal variances of $X(s)$ and $Y(t)$ and the cross-covariance of $(X(s), Y(t))$ also come into play in their modeling.

Most work to date has assumed iid errors and no random effect functions, and so it does not take within-function or between-function correlations into account in estimation or inference, or incorporate many of the modeling structures developed for functional response regression, with two exceptions. The functional AMM of Scheipl et al. (2014) discussed in Section 5.9 also accommodates functional predictors, allowing linear regression and two types of additive regression: additive in the PCs as presented by Müller & Yao (2008) or additive in the sense of Hastie & Tibshirani (1986), effectively generalizing McLean et al. (2012). Their approach requires discretizing the predictor

coefficient on a grid of *s*. Additionally, Meyer et al. (2013) extended the Bayesian FMM methods of Morris & Carroll (2006) and the additional methodological publications that followed to the function-on-function regression setting, using (potentially different) basis transformations on both the functional responses and predictors and basis truncation, roughness penalties, or sparsity used for regularization. This model is embedded in the general FMM setting, allowing multiple levels of fixed effect functions for scalar and functional predictors and multiple levels of random effect functions, fit in a way that both within- and between-function variability are accounted for in estimation and inference. Point-wise and joint functional inference is done on the coefficient surfaces.

7. DISCUSSION

There has been a great deal of work on functional regression, especially in the past decade, but clearly there is still much work to be done. Most of the work to date focuses on 1D continuous-valued functions observed on Euclidean domains, and most methods are designed for smooth functional data on sparse grids as commonly encountered in longitudinal data settings. The scope of FDA is much broader than this, with many applications yielding 2D or 3D functions, such as quantitative image data, functions on non-Euclidean manifolds (e.g., spheres, closed surfaces), discrete- or matrix-values functions, and complex functions with numerous features and sampled on a high-dimensional, fine grid.

A few of the papers highlighted here deal with image data or 2D/3D functions. Given the emerging importance of these types of data in various applications and the often-times naïve analyses commonly used in practice for them, further development in this area is crucial. There are some preliminary papers involving functional data on manifold domains and a few others dealing with discrete-valued functions, but not much, and so this is another emerging area in the coming years. Methods that involve general basis function approaches can potentially be directly applied to manifold data in cases where basis functions honoring the inherent constraints can be constructed. Very few methods have been designed for complex, high-dimensional functional data, which are encountered in a number of emerging applications generating automated measurements on grids of time and/or space. The Bayesian FMM methods of Morris & Carroll (2006) and the additional methodological publications that followed have been specifically developed with this setting in mind, and have been used for very high-dimensional genomic (Morris et al. 2008, Lee & Morris 2014) and imaging data (Morris et al. 2011, Fazio et al. 2013), in which the functions are complex with many local features and are observed on grids of size thousands to millions. Most other functional regression methods with available software are at least quadratic in the number of observations per function *T*, so cannot scale up to these enormous sizes. More functional regression methods need to be developed with these types of complex functions and scalability to enormous functions in mind.

There are many alternative basis functions and regularization approaches used in functional regression analyses. One open problem is to determine when each type of basis is suitable and preferable. PCs are a workhorse for many FDA methods, but more study needs to be done to determine in what settings they should be used. With complex, high-dimensional functions that have many features and a relatively small sample size, the empirical covariance matrices upon which the PCs are computed are poorly estimated, and it is not clear how many of the eigenfunctions are well estimated, or what their use implies when they are not. There are preliminary theoretical results on high-dimension, low sample size asymptotics (see Section 3), but the practical implications of these results need to be explored. In settings where PCs are not well-estimated, what other bases should be used? Alternative scalable approaches that can provide some degree of flexibility in representing the covariances but are estimable in such settings would be helpful. Sparse graphical

models may be useful in this context. Another welcome addition to the literature would be statistically principled approaches to select among a set of candidate bases. The literature would benefit from a better understanding of regularization and the development of more methods to adaptively regularize data to adapt to local features. More incorporation of modern sparsity penalties and priors into the functional regression modeling will likely lead to some of these developments.

Statistical principles suggest that regularization in the various ways discussed in this review should lead to more efficient estimation and more accurate inference. It would be helpful to have careful studies of the importance of regularization of the functional coefficients for estimation and inference in these various models. How much of a difference does it make? Is it more important for estimation or inference? Also, it would be helpful to determine in what settings it is more important to account for within-function correlation in estimation of the functional coefficients, how accurately one must represent or estimate the correlation in order to derive a benefit, and for what types of inferential or predictive tasks it is most important. There are similar relevant questions in terms of accounting for between-function correlation when it is present that seem to be poorly understood by most researchers, and need to be addressed. These issues are challenging to study theoretically given the vastly different types of functional data encountered in practice (i.e., not all are simple smooth functions), the difficulty to determine the most appropriate asymptotic (e.g., based on number of curves, positions within curve, or numbers of clusters of curves), and the difficulty of assessing whether asymptotic results are relevant to a particular applied setting.

Many published methods have focused primarily on point estimation. Although there are approaches in the literature for computing point-wise bands, joint bands, hypothesis tests, and model selection statistics for functional regression, there is still more work that needs to be done in these areas, and care needs to be taken so that the inferential methods take into account all of the various within- and between-function sources of variability that characterize these rich data.

SUMMARY POINTS

1. Functional data analysis (FDA) is a rapidly developing field in statistics whose growth is fueled by growing numbers of applications yielding this type of structured data. Functional regression, comprised by regression methods involving functional responses and/or predictors, is the area of functional data analysis receiving the most attention in applications and methodology development.

2. The crux of FDA is the need to combine information both across and within functions, which Ramsay & Silverman (1997) call replication and regularization, respectively. These are the key concepts underlying functional regression methods.

3. Basis functions are key building blocks of functional regression methods, and they determine how the model borrows strength across observations within a curve in performing the regression.

4. Regularization is accomplished by applying one of three operations to the chosen basis coefficients: truncation, roughness penalties, or sparsity penalties.

5. Most functional predictor regression models work with the same underlying functional linear model (FLM). The key distinctions among the methods are which basis functions are used to represent the functional predictor and/or regression coefficient, and which operation is used to perform regularization.

6. Many papers modeling functional data as responses focus on estimation of an overall mean function, although more recently a number of methods deal with a general functional response regression model. These methods primarily differ in terms of their choice of basis functions and approach to regularization of the functional coefficients, as well as in their approach for modeling intrafunctional correlation in the residuals or random effects.

7. Most functional response regression work assumes independently sampled functions, although there is a small but growing literature on methods that can model interfunctional correlation through functional mixed effect models (FMMs) or models for longitudinally or spatially sampled functional data.

8. There has been comparatively little work on function-on-function regression methods, although there is a growing literature that includes concurrent, historical, and unconstrained function-on-function models.

FUTURE ISSUES

1. Most functional regression methods are designed for simple, smooth sparsely sampled functions on Euclidean domains. More work is needed for methods that are flexible and computationally efficient enough to handle complex functions on high-dimensional grids and multidimensional and/or non-Euclidean domains.

2. Although a number of existing papers deal with inferential issues, the field needs further development of methods to perform inference in the functional regression context, especially while taking the multiple testing issues into account.

3. Most existing functional regression methods are intended for independently sampled functions. More development is needed for methods that can account for interfunctional correlation induced by multilevel designs and longitudinally or spatially sampled functions.

4. There is little work to date on functional regression methods for discrete-valued functional data, so this is another area requiring further development.

5. Basis functions are ubiquitous in functional regression, but there is very little in the literature about how to determine which set of basis functions is most suitable for a given application. The literature is in need of methods and strategies to select basis functions and provide guidance to determine the settings in which each type of basis works well.

6. Statistical principles suggest that models that capture the within-function and between-function correlation should yield more efficient estimators and more accurate inference in the functional regression context, but it is not clear how much of a difference it makes or how accurately the true covariance structure must be captured to realize such benefits. It would be helpful to have careful studies and theoretical investigations of the benefit of accounting for these sources of variability, the consequences when it is not properly taken into account, and an idea of how accurately one must represent or estimate the true correlation structure in order to derive a benefit.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

Abramovich F, Angelini C. 2006. Testing in mixed-effects FANOVA models. *J. Stat. Plan. Inf.* 136:4326–48

Antoniadis A, Sapatinas T. 2007. Estimation and inference in functional mixed-effect models. *Comp. Stat. Data Anal.* 51:4793–813

Armagan A, Dunson D, Clyde M. 2011. Generalized beta mixtures of Gaussians. In *Advances in Neural Information Processing Systems*, Vol. 24, ed. J Shawe-Taylor, R Zemel, P Bartlett, F Pereira, K Weinberger, pp. 523–31. La Jolla, CA: Neural Inf. Process. Syst. Found. **http://papers.nips.cc/paper/4439-generalized-beta-mixtures-of-gaussians**

Aston JA, Chiou J, Evans J. 2010. Linguistic pitch analysis using functional principal component mixed effect models. *J. R. Stat. Soc. C* 59:297–317

Baik J, Silverstein JW. 2006. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivar. Anal.* 97:1382–408

Baladandayuthapani V, Mallick BK, Hong MY, Lupton JR, Turner ND, Carroll RJ. 2008. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* 64:321–22

Barry D. 1995. A Bayesian model for growth curve analysis. *Biometrics* 51(2):639–55

Barry D. 1996. An empirical Bayes approach to growth curve analysis. *Statistician* 45:3–19

Berhane K, Molitor NT. 2008. A Bayesian approach to functional-based multilevel modeling of longitudinal data: applications to environmental epidemiology. *Biostatistics* 4:686–99

Besse PC, Cardot H. 1996. Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Can. J. Stat.* 24:467–87

Bigelow JL, Dunson DB. 2007. Bayesian adaptive regression splines for hierarchical data. *Biometrics* 63:724–32

Brown PJ, Fearn T, Vannucci M. 2001. Bayesian wavelet regression on curves with application to a spectroscopic calibration problem. *J. Am. Stat. Assoc.* 96:398–408

Brumback BA, Rice JA. 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Stat. Assoc.* 93:961–76

Bunea F, Ivanescu AE, Wegkamp MH. 2011. Adaptive inference for the mean of a Gaussian process in functional data. *J. R. Stat. Soc. B* 73(4):531–58

Cardot H, Ferraty F, Mas A, Sarda P. 2003. Testing hypotheses in the functional linear model. *Scand. J. Stat.* 30:241–55

Cardot H, Ferraty F, Sarda P. 1999. Functional linear model. *Stat. Prob. Lett.* 45:11–22

Carroll RJ, Ruppert D, Stefanski LA. 1995. *Measurement Error in Nonlinear Models: A Modern Perspective*. New York: Springer-Verlag

Carvahlo CM, Polson NG, Scott JG. 2010. The horseshoe estimator for sparse signals. *Biometrika* 97(2):465–80

Chen H, Wang Y. 2011. A penalized spline approach to functional mixed effect model analysis. *Biometrics* 67:861–70

Crainiceanu CM, Goldsmith AJ. 2010. Bayesian functional data analysis using WinBUGS. *J. Stat. Softw.* 32(11):i11

Crainiceanu CM, Ruppert D. 2004. Likelihood ratio tests in linear mixed models with one variance component. *J. R. Stat. Soc. B* 66(1):165–85

Crainiceanu CM, Staicu AM, Di CZ. 2009. Generalized multilevel functional regression. *J. Am. Stat. Assoc.* 104:1550–61

Crainiceanu CM, Staicu AM, Ray S, Punjabi N. 2012. Bootstrap-based inference on the difference in the means of two correlated functional responses. *Stat. Med.* 31:3223–40

Crambes C, Kneip A, Sarda P. 2009. Smoothing spline estimators for functional linear regression. *Ann. Stat.* 37(1):35–72

Degras D. 2011. Simultaneous confidence bands for nonparametric regression with functional data. *Stat. Sin.* 21:1735–65

Demmler A, Reinsch C. 1975. Oscillation matrices with spline smoothing. *Numer. Math.* 24:375–82

Di CZ, Crainiceanu CM, Caffo BM, Punjabi NM. 2009. Multilevel functional principal component analysis. *Ann. Appl. Stat.* 3:458–88

Di CZ, Crainiceanu CM, Jank WS. 2014. Multilevel sparse functional principal component analysis. *Stat* 3(1):126–43

Diggle PJ, Wasel A. 1997. Spectral analysis of replicated biomedical time series. *J. R. Stat. Soc. C* 46(1):31–71

Do KA, Kirk K. 1999. Discriminant analysis of event-related potential curves using smoothed principal components. *Biometrics* 55:174–81

Donoho DL, Johnstone IM. 1995. Minimax estimation via wavelet shrinkage. *Ann. Stat.* 26(3):879–921

Eilers PHC, Marx BD. 1986. Flexible smoothing with B-splines and penalties. *Stat. Sci.* 11(2):89–121

Eilers PHC, Marx BD. 2002. Generalized linear additive smooth structures. *J. Comput. Graph. Stat.* 11:758–83

Fan J, Zhang JT. 2000. Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B* 62:303–22

Fan Y, James GM, Radchenko P. 2014. *Functional additive regression*. Tech. Rep. Marshall Sch. Bus., Univ. South. Calif., Los Angeles, CA. **http://www–bcf.usc.edu/~gareth/research/FAR.pdf**

Faraway JJ. 1997. Regression analysis for a functional response. *Technometrics* 39:254–61

Fazio MA, Grytz R, Morris JS, Bruno L, Gardiner S, et al. 2013. Age-related changes in human peripapillary scleral stiffness. *Biomech. Model. Mechanobiol.* 13(3):551–63

Fox EB, Dunson DB. 2012. Multiresolution Gaussian processes. *Adv. Neural Inf. Process. Syst.* 25:746–54

Freyermuth JM, Ombao H, von Sachs R. 2010. Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *J. Am. Stat. Assoc.* 105(490):634–46

George EI, McCulloch RE. 1993. Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* 88:881–89

Gertheiss J, Goldsmith J, Crainiceanu C, Greven S. 2013. Longitudinal scalar-on-function regression with application to tractography data. *Biostatistics* 14(3):447–61

Gervini D. 2006. Free-knot spline smoothing for functional data. *J. R. Stat. Soc. Ser. B* 68(4):671–88

Goldsmith J, Bobb J, Crainiceanu C, Caffo B, Reich D. 2011a. Penalized functional regression. *J. Comput. Gr. Stat.* 20:830–51

Goldsmith J, Crainiceanu CM, Caffo B, Reich D. 2012. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc. C* 61(3):453–69

Goldsmith J, Huang L, Crainiceanu CM. 2014. Smooth scalar-on-image regression via spatial Bayesian variable selection. *J. Comput. Gr. Stat.* 23(1):46–64

Goldsmith J, Kitago K. 2013. *Assessing systematic effects of stroke on motor control using hierarchical scalar-on-function regression*. Tech. Rep., Columbia Univ., New York, NY. **http://jeffgoldsmith.com/Downloads/BayesPFSR_Paper.pdf**

Goldsmith J, Wand MP, Crainiceanu C. 2011b. Functional regression via variational Bayes. *Electron. J. Stat.* 5:571–602

Greven S, Crainiceanu C, Caffo B, Reich D. 2010. Longitudinal functional principal components analysis. *Electron. J. Stat.* 4:1022–54

Griffin JE, Brown PJ. 2010. Inference with normal-gamma prior distributions in regression problems. *Bayesian Anal.* 5:171–88

Guo W. 2002. Functional mixed effect models. *Biometrics* 58:121–28

Guo W. 2003. Smoothing spline ANOVA for time-dependent spectral analysis. *J. Am. Stat. Assoc.* 98(463):643–52

Hall P, Marron JS, Neeman A. 2005. Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B* 67:427–44

Hall P, Poskitt DS, Presnell B. 2001. A functional data-analytic approach to signal discrimination. *Technometrics* 43:1–9

Harezlak J, Coull BA, Laird NM, Magari SR, Christiani DC. 2007. Penalized solutions to functional regression problems. *Comput. Stat. Stat. Anal.* 51(10):4911–25

Hart JD, Wehrly TE. 1986. Kernel regression estimation using repeated measurements data. *J. Am. Stat. Assoc.* 81:1080–88

Hastie T, Mallows C. 1993. [A statistical view of some chemometrics regression tools]: Discussion. *Technology* 35(2):140–43

Hastie T, Tibshirani R. 1986. Generalized additive models. *Stat. Sci.* 1(3):297–318

Hastie T, Tibshirani R. 1993. Varying coefficient models. *J. R. Stat. Soc. Ser. B* 55:757–96

Herrick RC, Morris JS. 2006. Wavelet-based functional mixed model analysis: computational considerations. *Joint Stat. Meet. 2006 Proc.*, *ASA Sect. Stat. Comput.*, *Seattle*, pp. 2051–53. Alexandria, VA: Am. Stat. Assoc.

Hodges JS. 2013. *Richly Parameterized Linear Models: Additive, Time Series, and Spatial Models Using Random Effects*. New York: Chapman & Hall

Holan SH, Wikle CK, Sullivan-Beckers LE, Cocroft RB. 2010. Modeling complex phenotypes: generalized linear models using spectrogram predictors of animal communication signals. *Biometrics* 66(3):914–24

Hoover DR, Rice JA, Wu CO, Yang LP. 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85:809–22

Huang JZ, Wu CO, Zhou L. 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Stat. Sin.* 14:763–88

Ivanescu AE, Staicu AM, Greven S, Scheipl F, Crainiceanu CM. 2012. Penalized function-on-function regression. *Tech. Rep.* In press. doi: 10.1007/s00180-014-0548-4

James GM. 2002. Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B* 64:411–32

James GM, Hastie TJ. 2001. Functional linear discriminant analysis for irregularly sampled curves. *J. R. Stat. Soc. Ser. B* 63:533–50

James GM, Hastie TJ, Sugar CA. 2000. Principal component models for sparse functional data. *Biometrika* 87(3):587–602

James GM, Silverman BW. 2005. Functional adaptive model estimation. *J. Am. Stat. Assoc.* 100:565–76

James GM, Wang J, Zhu J. 2009. Functional linear regression that's interpretable. *Ann. Stat.* 37(5A):2083–108

Johnstone IM, Lu AY. 2009. On consistency and sparsity for principal component analysis in high dimensions. *J. Am. Stat. Assoc.* 104(486):683–93

Joliffe IT. 1982. A note on the use of principal components in regression. *J. R. Stat. Soc. Ser. C* 31(3):300–3

Jung S, Marron J. 2009. PCA consistency in high dimension, low sample size context. *Ann. Stat.* 37:4104–30

Kendall MG. 1957. *A Course in Multivariate Analysis*. London: Griffin

Kim K, Senturk D, Li R. 2011. Recent history functional linear models for sparse longitudinal data. *J. Stat. Plan. Inference* 141(4):1554–66

Koomen JM, Shih LN, Coombes KR, Li D, Xiao LC, et al. 2005. Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. *Clin. Cancer Res.* 11(3):1110–18

Krafty RT, Gimotty PA, Holtz D, Coukos G, Guo W. 2008. Varying coefficient model with unknown within-subject covariance for analysis of tumor growth curves. *Biometrics* 64(4):1023–31

Krafty RT, Hall M, Guo W. 2011. Functional mixed effects spectral analysis. *Biometrika* 98(3):583–98

Kundu MG, Harezlak J, Randolph TW. 2012. Longitudinal functional models with structured penalties. arXiv:1211.4763 [stat.AP]

Laird NM, Ware JH. 1982. Random-effects models for longitudinal data. *Biometrics* 38(4):963–74

Lang S, Bretzger A. 2004. Bayesian P-splines. *J. Comp. Graph. Stat.* 13:183–212

Lee ER, Park BU. 2012. Sparse estimation in functional linear regression. *J. Multivar. Anal.* 105:1–17

Lee S, Zou F, Wright FA. 2010. Convergence and prediction of principal component scores in high-dimensional settings. *Ann. Stat.* 38:3605–29

Lee W, Baladandayuthapani V, Fazio M, Downs C, Morris JS. 2014. *Semiparametric functional mixed models for longitudinally observed functional data, with application to glaucoma data*. Tech. Rep., Univ. Tex. MD Anderson Cancer Cent., Houston, Tex.

Lee W, Morris JS. 2014. *Identification of differentially expressed methylated loci using wavelet-based functional mixed models*. Tech. Rep., Univ. Tex. MD Anderson Cancer Cent., Houston, Tex.

Li B, Marx BD. 2008. Sharpening P-spline signal regression. *Stat. Model.* 8:367–83

Li B, Yu Q. 2008. Classification of functional data: a segmentation approach. *Comp. Stat. Data. Anal.* 52:4790–800

Li Y, Hsing T. 2010. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Stat.* 38(6):3321–51

Li Y, Wang N, Carroll RJ. 2010. Generalized functional linear models with semiparametric single-index interactions. *J. Am. Stat. Assoc.* 105(490):621–33

Liang H, Wu H, Carroll RJ. 2003. The relationship between virology and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics* 4(2):297–312

Lin X, Wang N, Welsh AH, Carroll RJ. 2004. Equivalent kernels of smoothing splines in nonparametric regression for longitudinal/clustered data. *Biometrika* 91(1):177–93

Ma Y, Yang L, Carroll RJ. 2012. A simultaneous confidence band for sparse longitudinal regression. *Stat. Sin.* 22:95–122

Malfait N, Ramsay JO. 2003. The historical functional linear model. *Can. J. Stat.* 31:115–28

Malloy EJ, Morris JS, Adar SD, Suh HH, Gold DR, Coull BA. 2010. Wavelet-based functional linear mixed models: an application to measurement error–corrected distributed lag models. *Biostatistics* 11(3):432–52

Maronna RA, Yohai VJ. 2013. Robust functional linear regression based on splines. *Comput. Stat. Data Anal.* 65:46–55

Martinez JG, Fiang F, Zhou L, Carroll RJ. 2010. Longitudinal functional principal component modeling via stochastic approximation Monte Carlo. *Can. J. Stat.* 38(2):256–70

Martinez JG, Bohn KM, Carroll RJ, Morris JS. 2013. A study of Mexican free-tailed bat syllables: Bayesian functional mixed modeling of nonstationary acoustic time series. *J. Am. Stat. Assoc.* 108(502):514–26

Marx BD, Eilers PHC. 1999. Generalized linear regression on sampled signals and curves: a *P*-spline approach. *Technometrics* 41(1):1–13

Marx BD, Eilers PHC. 2002. Multivariate calibration stability: a comparison of methods. *J. Chemom.* 16:129–40

Marx BD, Eilers PHC. 2005. Multidimensional penalized signal regression. *Technometrics* 47:13–22

Marx BD, Eilers PHC, Li B. 2011. Multidimensional single-index signal regression. *Chemom. Intell. Lab. Syst.* 109:120–30

McLean MW, Hooker G, Staicu AM, Scheipl F, Ruppert D. 2012. Functional generalized additive models. *J. Comput. Gr. Stat.* 23:249–69

McLean MW, Scheipl F, Hooker G, Greven S, Ruppert D. 2013. *Bayesian functional generalized additive models with sparsely observed covariates*. arXiv:1305.3585 [stat.ME]

Meyer MJ, Coull BA, Versace F, Cinciripini P, Morris JS. 2015. Bayesian function-on-function regression for multi-level functional data. *Biometrics* In press

Morris JS. 2012. Statistical methods for proteomic biomarker discovery using feature extraction or functional data analysis approaches. *Stat. Interface* 5(1):117–36

Morris JS, Arroyo C, Coull BA, Ryan LM, Herrick RC, Gortmaker SL. 2006. Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study. *J. Am. Stat. Assoc.* 101(476):1352–64

Morris JS, Baladandayuthapani V, Herrick RC, Sanna PP, Gutstein H. 2011. Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *Ann. Appl. Stat.* 5:894–923

Morris JS, Brown PJ, Herrick RC, Baggerly KA, Coombes KR. 2008. Bayesian analysis of mass spectrometry data using wavelet-based functional mixed models. *Biometrics* 12:479–89

Morris JS, Carroll RJ. 2006. Wavelet-based functional mixed models. *J. R. Stat. Soc. B* 68(2):179–99

Morris JS, Vannucci M, Brown PJ, Carroll, RJ. 2003. Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *J. Am. Stat. Assoc.* 98:573–83

Mostacci E, Truntzner C, Cardot H, Ducoroy P. 2010. Multivariate denoising methods combining wavelets and principal component analysis for mass spectrometry data. *Proteomics* 10:2564–72

Müller HG, Stadtmüller U. 2005. Generalized functional linear models. *Ann. Stat.* 33:774–805

Ogden RT, Greene E. 2010. Wavelet modeling of functional random effects with application to human vision data. *J. Stat. Plann. Inference* 140:3797–808

Ogden RT, Miller CE, Takezawa K, Ninomiya S. 2002. Functional regression in crop lodging assessment with digital images. *J. Agric. Biol. Environ. Stat.* 7(3):389–402

Park T, Casella G. 2008. The Bayesian lasso. *J. Am. Stat. Assoc.* 103:672–80

Qin L, Guo W. 2006. Functional mixed-effects model for periodic data. *Biostatistics* 7:225–34

Qin L, Guo W, Litt B. 2009. A time-frequency functional model for locally stationary time series. *J. Comput. Graph. Stat.* 18(3):675–93

Ramsay JO, Dalzell CJ. 1991. Some tools for functional data analysis. *J. R. Stat. Soc. Ser. B* 53:539–72

Ramsay JO, Hooker G, Graves S. 2009. *Functional Data Analysis with R and MATLAB*. New York: Springer-Verlag

Ramsay JO, Silverman BW. 1997. *Functional Data Analysis*. New York: Springer-Verlag. 1st ed.

Ramsay JO, Silverman BW. 2005. *Functional Data Analysis*. New York: Springer-Verlag. 2nd ed.

Randolph TW, Harezlak J, Feng Z. 2012. Structured penalties for functional linear models—partially empirical eigenvectors for regression. *Elec. J. Stat.* 6:323–53

Rao CR. 1958. Some statistical methods for comparison of growth curves. *Biometrics* 14(1):1–17

Ratliffe SJ, Heller GZ, Leader LR. 2002a. Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. *Stat. Med.* 21(8):1103–14

Ratliffe SJ, Heller GZ, Leader LR. 2002b. Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression. *Stat. Med.* 21(8):1115–27

Reiss PT, Huang L, Mennes M. 2010. Fast function-on-scalar regression with penalized basis expansions. *Int. J. Biostat.* 6(1):28

Reiss PT, Ogden RT. 2007. Functional principal component regression and functional partial least squares. *J. Am. Stat. Assoc.* 102(479):984–96

Reiss PT, Ogden RT. 2010. Functional generalized linear models with images as predictors. *Biometrics* 66(1):61–69

Rice JA, Silverman BW. 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. B* 53(1):233–43

Rice JA, Wu CO. 2001. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 57:253–59

Røislien J, Winje B. 2013. Feature extraction across individual time series observations with spikes using wavelet principal component analysis. *Stat. Med.* 32(21):3660–69

Ruppert D, Wand MP, Carroll RJ. 2003. *Semiparametric Regression*. New York: Cambridge Univ. Press

Scarpa B, Dunson DB. 2009. Bayesian hierarchical functional data analysis via contaminated information priors. *Biometrics* 65:772–80

Scheipl F, Greven S. 2012. *Identifiability in penalized function-on-function regression models*. Tech. Rep. No. 125, Dept. Stat., Univ. Munich, Ger.

Scheipl F, Staicu AM, Greven S. 2014. Functional additive mixed models. *J. Comput. Gr. Stat.* In press

Shen D, Shen H, Marron JS. 2013. Consistency of sparse PCA in high dimension, low sample size contexts. *J. Multivar. Anal.* 115:317–33

Shi M, Weiss RE, Taylor JMG. 1996. An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *J. R. Stat. Soc. Ser. C* 45:151–63

Silverman BW. 1996. Smoothed principal component analysis by choice of norm. *Ann. Stat.* 24:1–24

Spitzner DJ, Marron JS, Essick GK. 2003. Mixed model functional ANOVA. *J. Am. Stat. Assoc.* 98:263–72

Staicu AM, Crainiceanu CM, Carroll RJ. 2010. Fast methods for spatially correlated multilevel functional data. *Biostatistics* 11(2):177–94

Staicu AM, Crainiceanu CM, Reich DS, Ruppert D. 2012. Modeling functional data with spatially heterogeneous shape characteristics. *Biometrics* 68:331–43

Staniswallis JG, Lee JJ. 1998. Nonparametric regression analysis of longitudinal data. *J. Am. Stat. Assoc.* 93:1403–18

Stingo F, Vannucci M, Downey G. 2012. Bayesian wavelet-based curve classification via discriminant analysis with Markov tree priors. *Stat. Sin.* 22:465–88

Storlie CB, Fugate ML, Higdon DM, Huzurbazar AV, Francois EG, McHugh DC. 2013. Methods for characterizing and comparing populations of shock wave curves. *Technometrics* 55(4):436–49

Swihart BJ, Goldsmith J, Crainiceanu CM. 2013. *Restricted likelihood ratio tests for functional effects in the functional linear model*. Work. Pap. 247, Dept. Biostat., Johns Hopkins Univ., Baltimore, MD

Thompson WK, Rosen O. 2008. A Bayesian model for sparse functional data. *Biometrics* 64(1):54–63

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58:267–88

Wang XF, Ray S, Mallick BK. 2007. Bayesian curve classification using wavelets. *J. Am. Stat. Assoc.* 102(479):962–73

Wang XF, Yang Q, Fan Z, Sun CK, Yue GH. 2009. Assessing time-dependent association between scalp EEG and muscle activation: a functional random-effects model approach. *J. Neurosci. Methods* 177:232–40

Wold H. 1966. Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis*, ed. PR Krishnaiah, pp. 383–487. New York: Acad. Press

Wood SN. 2006. *Generalized Additive Models: An Introduction with R*. New York: Chapman & Hall

Wood SN. 2011. Fast stable restricted maximum likelihood and margin likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. B* 73(1):3–36

Wu CO, Chiang CT. 2000. Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Stat. Sin.* 10:433–56

Wu CO, Chiang CT, Hoover DR. 1998. Asymptotic confidence regions for kernel smoothing of the varying coefficients model. *J. Am. Stat. Assoc.* 93:1388–402

Wu H, Liang H. 2004. Backfitting random varying-coefficients models with time-dependent smoothing covariates. *Scand. J. Stat.* 31:3–19

Wu H, Zhang JT. 2002. Local polynomical mixed-effects models for longitudinal data. *J. Am. Stat. Assoc.* 97:883–97

Wu S, Müller HG. 2011. Response-adaptive regression for longitudinal data. *Biometrics* 67:852–60

Yang WH, Wikle CK, Holan SH, Wildhaber ML. 2013. Ecological prediction with nonlinear multivariate time-frequency functional data models. *J. Agric. Biol. Environ. Stat.* 18(3):450–74

Yao F, Müller HG. 2010. Functional quadratic regression. *Biometrika* 97(1):49–64

Yao F, Müller HG, Wang JL. 2005a. Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* 100:577–90

Yao F, Müller HG, Wang JL. 2005b. Functional linear regression for longitudinal data. *Ann. Stat.* 33:2873–903

Yuan M, Cai TT. 2010. A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Stat.* 28(6):3412–44

Zellner A. 1962. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *J. Am. Stat. Assoc.* 57:348–86

Zhang D, Lin X, Sowers MF. 2000. Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics* 56:31–39

Zhang D, Lin X, Sowers MF. 2007. Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics* 63:351–62

Zhang L, Baladandayuthapani V, Zhu H, Baggerly KA, Majewski TA, et al. 2015. Functional CAR models for large spatially correlated functional datasets. *J. Am. Stat. Assoc.* In press

Zhao Y, Ogden RT, Reiss PT. 2012. Wavelet-based LASSO in functional linear regression. *J. Comput. Gr. Stat.* 21(3):600–17

Zhou L, Huang JZ, Carroll RJ. 2008. Joint modeling of paired sparse functional data using principal components. *Biometrika* 95(3):601–19

Zhou L, Huang JZ, Martinez JG, Maity A, Baladandayuthapani V, Carroll RJ. 2010. Reduced rank mixed effects models for spatially correlated hierarchical functional data. *J. Am. Stat. Assoc.* 105:390–400

Zhu H, Brown PJ, Morris JS. 2011. Robust, adaptive functional regression in functional mixed model framework. *J. Am. Stat. Assoc.* 106(495):1167–79

Zhu H, Brown PJ, Morris JS. 2012. Robust classification of functional and quantitative image data using functional mixed models. *Biometrics* 68:1260–68

Zhu H, Cox DD. 2009. A functional generalized linear model with curve selection in cervical pre-cancer diagnosis using fluorescence spectroscopy. *Optimality: Third Erich L. Lehmann Symp.* 57:173–89

Zhu H, Vannucci M, Cox DD. 2010. A Bayesian hierarchical model for classification with selection of functional predictors. *Biometrics* 66:463–73

Zhu H, Yao F, Zhang HH. 2014. Structured functional additive regression in reproducing kernel Hilbert spaces. *J. R. Stat. Soc. B* 76(3):581–603. doi: 10.1111/rssb.12036

Zipunnikov V, Caffo B, Yousem D, Davitzikos C, Schwartz BS, Crainiceanu C. 2011. Multilevel functional principal component analysis for high-dimensional data. *J. Comput. Gr. Stat.* 20:852–73