International Statistical Review

International Statistical Review (2017), 85, 2, 228–249 doi:10.1111/insr.12163

Methods for Scalar-on-Function Regression

Philip T. Reiss[1,2,3], Jeff Goldsmith[4], Han Lin Shang[5] and R. Todd Ogden[4,6]

[1]*Department of Child and Adolescent Psychiatry, New York University School of Medicine, New York, NY, USA*
E-mail: phil.reiss@nyumc.org
[2] *Department of Population Health, New York University School of Medicine, New York, NY, USA*
[3]*Department of Statistics, University of Haifa, Haifa, Israel*
[4]*Department of Biostatistics, Columbia University Mailman School of Public Health, New York, NY, USA*
[5]*Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, Australia*
[6]*New York State Psychiatric Institute, New York, NY, USA*

Summary

Recent years have seen an explosion of activity in the field of functional data analysis (FDA), in which curves, spectra, images and so on are considered as basic functional data units. A central problem in FDA is how to fit regression models with scalar responses and functional data points as predictors. We review some of the main approaches to this problem, categorising the basic model types as linear, non-linear and non-parametric. We discuss publicly available software packages and illustrate some of the procedures by application to a functional magnetic resonance imaging data set.

Key words: Functional additive model; functional generalised linear model; functional linear model; functional polynomial regression; functional single-index model; non-parametric functional regression.

1 Introduction

Regression with functional data is perhaps the most thoroughly researched topic within the broader literature on functional data analysis (FDA). It is common (e.g. Ramsay and Silverman, 2005; Reiss *et al.*, 2010) to classify functional regression models into three categories according to the role played by the functional data in each model: scalar responses and functional predictors ('scalar-on-function' regression); functional responses and scalar predictors ('function-on-scalar' regression); and functional responses and functional predictors ('function-on-function' regression). This article focuses on the first case and reviews linear, non-linear, and non-parametric approaches to scalar-on-function regression. Domains in which scalar-on-function regression (hereafter, SoFR) has been applied include chemometrics (Goutis, 1998; Marx & Eilers, 1999; Ferraty *et al.*, 2010), cardiology (Ratcliffe *et al.*, 2002), brain science (Reiss & Ogden, 2010; Goldsmith *et al.*, 2011; Huang *et al.*, 2013), climate science (Ferraty *et*

al., 2005; Baíllo & Grané, 2009) and many others. We refer the reader to Morris (2015) for a recent review of functional regression in general and to Wang *et al.* (2015) for a broad overview of FDA.

In cataloguing the many variants of SoFR, we have attempted to cast a wide net. A major contribution of this review is our attempt not merely to describe many approaches in what has become a vast literature, but to distill a coherent organisation of these methods. To keep the scope somewhat manageable, we do not attempt to survey the functional *classification* literature. We acknowledge, however, that classification and regression are quite closely related–especially insofar as functional logistic regression, a special case of the functional generalised linear models considered in Section 5.3, can be viewed as a classification method. Our emphasis is more methodological than theoretical, but for brevity, we omit a number of important methodological issues such as confidence bands, goodness-of-fit diagnostics, outlier detection and robustness.

In the 'vanilla' data setting (Sections 2–4), we consider an independent, identically distributed (iid) sample of random pairs $(\mathcal{X}_i, y_i)$, $i = 1, \ldots, n$, where y_i is a real-valued scalar outcome and the predictor $\mathcal{X}_i$ belongs to a space $\mathcal{F}$ of real-valued functions on a finite interval $\mathcal{I} \subset \mathbb{R}$. The most common choice for $\mathcal{F}$ seems to be the Hilbert space $L^2(\mathcal{I})$ with the usual inner product $\langle \mathcal{X}_1, \mathcal{X}_2 \rangle = \int_{\mathcal{I}} \mathcal{X}_1(t)\mathcal{X}_2(t)dt$. Cuevas (2014) discusses more general spaces in which the functional data may 'live'.

In the general practice of FDA, functions $\mathcal{X}_1(t), \ldots, \mathcal{X}_n(t)$ are observed on a set of discrete grid points that can be sparse or dense, regular or irregular and possibly subject to measurement errors. Several 'preprocessing' steps are typically taken before modelling the data. Aside from smoothing the functional data, in some cases, it is appropriate to apply registration or feature alignment, or if the grid points differ across observations, to interpolate to a dense common grid. Measurement error is expected to be low in some (e.g. chemometric) applications, but when it is not, it can have important effects on the regression relation. Some methods (e.g. James, 2002) account explicitly for such error. Here, in order to keep the focus on the various regression models, we shall mostly assume functional data observed on a common dense grid with negligible error.

The functional linear model (FLM) is a natural extension of multiple linear regression to allow for functional predictors. Many techniques have been developed to fit this model, and we review these in Section 2. Non-linear extensions of this basic approach are presented in Section 3. In Section 4, we discuss non-parametric approaches to SoFR, which are based on distances among the predictor functions. For simplicity of exposition, these three sections consider only the most basic and most common scenario: a single functional predictor and one real-valued scalar response. Generalisations and extensions, including the inclusion of scalar covariates, multiple functional predictors, generalised response values and repeated observations, are reviewed in Section 5. Section 6 presents some ideas on how to choose among the many methods. Available software for SoFR is described in Section 7, and an application to brain imaging data appears in Section 8. Some concluding discussion is provided in Section 9.

2 Linear Scalar-on-Function Regression

The scalar-response functional linear model can be expressed as

$$y_i = \alpha + \int_{\mathcal{I}} \mathcal{X}_i(t)\beta(t)dt + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\beta(\cdot)$ is the coefficient function and errors ε_i are iid with mean zero and constant variance σ^2.

The coefficient function $\beta(\cdot)$ has a natural interpretation: locations t with largest $|\beta(t)|$ are most influential to the response. In order to enforce some regularity in the estimate, a common general approach to fitting model (1) is to expand $\beta(\cdot)$ (and possibly the functional predictors as well) in terms of a set of basis functions. Basis functions can be categorised as either (i) a priori fixed bases, most often splines or wavelets or (ii) data-driven bases, most often derived by functional principal component analysis (FPCA) or functional partial least squares (FPLS). The next two subsections discuss these two broad alternatives.

2.1 *Regularised* a priori *Basis Functions*

One general class of methods for fitting Eqn 1 restricts the coefficient function $\beta(\cdot)$ to the span of an a priori set of basis functions while imposing a penalty or a prior to prevent overfitting. Assume $\beta(t) = \boldsymbol{b}(t)^T\boldsymbol{\gamma}$, where $\boldsymbol{b}(t) = [b_1(t), \ldots, b_K(t)]^T$ is a set of basis functions and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)^T \in \mathbb{R}^K$. We estimate α and $\boldsymbol{\gamma}$ by finding

$$(\hat{\alpha}, \hat{\boldsymbol{\gamma}}) = \underset{\alpha,\gamma}{\arg\min} \left[\sum_{i=1}^{n} \left\{ y_i - \alpha - \int_{\mathcal{I}} \mathcal{X}_i(t)(\boldsymbol{b}(t)^T\boldsymbol{\gamma})dt \right\}^2 + \lambda P(\boldsymbol{\gamma}) \right] \tag{2}$$

for some $\lambda > 0$ and some penalty function $P(\cdot)$. The estimate of the coefficient function is thus $\hat{\beta}(t) = \boldsymbol{b}(t)^T\hat{\boldsymbol{\gamma}}$.

In spline approaches (e.g. Hastie and Mallows, 1993; Marx and Eilers, 1999; Cardot *et al.*, 2003), the penalty is generally a quadratic form $\boldsymbol{\gamma}^T\boldsymbol{L}\boldsymbol{\gamma}$, which measures the roughness of $\beta(t) = \boldsymbol{b}(t)^T\boldsymbol{\gamma}$, and hence (2) is a generalised ridge regression problem. When $\boldsymbol{L} = \left[\int_{\mathcal{I}} b_i^{(q)}(t) b_j^{(q)}(t)dt\right]_{1\le i,j\le K}$, as in Ramsay and Silverman (2005), the quadratic form equals $\int_{\mathcal{I}}[\beta^{(q)}(t)]^2dt$; cubic B-splines with a second derivative penalty ($q = 2$) are particularly popular. Alternatively, the P-spline formulation of Marx and Eilers (1999) takes $\boldsymbol{L} = \boldsymbol{D}^T\boldsymbol{D}$, where $\boldsymbol{D}$ is a differencing matrix. Higher values of λ enforce greater smoothness in the coefficient function. Standard methods for automatic selection of λ include restricted maximum likelihood and generalised cross-validation (Craven & Wahba, 1979; Ruppert *et al.*, 2003; Reiss & Ogden, 2009; Wood, 2011). While B-splines are the basis functions most often combined with roughness penalties, other bases are possible. For example, Fourier bases may be employed when the functions are periodic. Marx and Eilers (1999) discuss smoothing of the curves $\mathcal{X}_i(t)$ when evaluating the integral in (2) in practice.

While splines and roughness penalties are a natural choice when the coefficient function is expected to be smooth, in some applications, $\beta(\cdot)$ may be irregular, with features such as spikes or discontinuities. Wavelet bases (e.g. Ogden, 1997), which provide sparse representations for irregular functions, have received some attention in recent years. In the framework of (2), Zhao *et al.* (2012) propose the wavelet-domain lasso, which combines wavelet basis functions $\boldsymbol{b}(\cdot)$ with the ℓ_1 penalty $P(\boldsymbol{\gamma}) = \sum_{k=1}^{K}|\gamma_k|$ in (2). Other sparsity penalties for wavelet-domain SoFR are considered by Zhao *et al.* (2015) and Reiss *et al.* (2015). Not all sparse approaches rely on wavelet bases; see, for example, James *et al.* (2009) and Lee and Park (2011).

More flexible, albeit potentially more complex, models can be built by replacing the penalty with an explicit prior structure in a fully Bayesian framework. Spline approaches of this type are developed by Crainiceanu and Goldsmith (2010) and Goldsmith *et al.* (2011); wavelet approaches based on Bayesian variable selection include those of Brown *et al.* (2001) and Malloy *et al.* (2010).

2.2 Regression on Data-Driven Basis Functions

Alternatively, the coefficient function in (1) can be estimated using a data-driven basis. The most common choice is the eigenbasis associated with the covariance function $\Gamma(t,s) = \mathrm{Cov}[\mathcal{X}_i(t), \mathcal{X}_i(s)]$, that is, the orthonormal set of functions $\phi_1(t), \phi_2(t), \ldots$ such that for each j and all $t \in \mathcal{T}$, $\int_{\mathcal{T}} \Gamma(t,s)\phi_j(s)ds = \lambda_j \phi_j(t)$ for eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Expressing each of the functional predictors by its truncated Karhunen–Loève expansion

$$\mathcal{X}_i(t) = \sum_{a=1}^{A} x_{ia}\phi_a(t) \tag{3}$$

and the coefficient function $\beta(t) = \sum_{a'=1}^{A} \beta_{a'}\phi_{a'}(t)$ using the same basis, the integral in (1) becomes

$$\sum_{a=1}^{A}\sum_{a'=1}^{A} x_{ia}\beta_{a'} \int_{\mathcal{T}} \phi_a(t)\phi_{a'}(t)dt = \sum_{a=1}^{A} x_{ia}\beta_a,$$

reducing (1) to the ordinary multiple regression model

$$y_i = \alpha + \sum_{a=1}^{A} x_{ia}\beta_a + \varepsilon_i \tag{4}$$

(Cardot *et al.*, 1999). In practice, the eigenfunctions are estimated by FPCA (Rice and Silverman, 1991; Silverman, 1996; Yao *et al.*, 2005) and treated as fixed in subsequent analysis.

The number A of retained components acts as a tuning parameter that controls the shape and smoothness of $\beta(\cdot)$. Ways to choose A as a part of the regression analysis include explained variability, bootstrapping (Hall & Vial, 2006), information criteria (Yao *et al.*, 2005; Li *et al.*, 2013) and cross-validation (Hosseini-Nasab, 2013).

Data-driven bases other than functional principal components can also be utilised, such as FPLS (Preda and Saporta, 2005; Escabias *et al.*, 2007; Reiss and Ogden, 2007; Aguilera *et al.*, 2010; Delaigle and Hall, 2012) and functional sliced inverse regression (Ferré & Yao, 2003; 2005).

2.3 Hybrid Approaches

A number of papers combine the data-driven basis and a priori basis approaches of the previous two subsections. For example, Amato *et al.* (2006) employ the so-called sufficient dimension reduction methods for component selection but implement them in the wavelet domain. The functional principal component regression method referred to as $\mathrm{FPCR_R}$ in Reiss and Ogden (2007) restricts the coefficient function to the span of leading functional principal components, but fits model (1) by penalised splines (see also Horváth and Kokoszka, 2012; Araki *et al.*, 2013). Goldsmith *et al.* (2011) use FPCA to pre-process predictor functions and penalised splines to model the coefficient function. The regularisation strategy of Randolph *et al.* (2012) incorporates information from both the predictors and a linear penalty operator.

2.4 Functional Polynomial Regression

We conclude the discussion of FLMs with functional polynomial models. These are linear in the coefficients but not in the predictors, in contrast to the models of the next two sections, which are not linear in either sense.

The functional quadratic regression model of Yao and Muller (2010) can be expressed as

$$y_i = \alpha + \int_{\mathcal{I}} \mathcal{X}_i(t)\beta(t)dt + \int_{\mathcal{I}}\int_{\mathcal{I}} \mathcal{X}_i(s)\mathcal{X}_i(t)\gamma(s,t)ds\ dt + \varepsilon_i. \tag{5}$$

Here, we have both linear and quadratic coefficient functions, $\beta(t)$ and $\gamma(s,t)$; when the latter is zero, (5) reduces to the FLM (1). By expressing elements of (5) in terms of functional principal components, responses can be regressed on the principal component scores. Adding higher-order interaction terms results in more general functional polynomial regression models.

3 Non-Linear Scalar-on-Function Regression

In many applications, the assumption of a linear relationship between $\mathcal{X}$ and y is too restrictive to describe the data. In this section, we review several models that relax the linearity assumption; Section 4 will describe models (usually termed 'non-parametric') that are even more flexible than those described here.

3.1 Single-Index Model

We begin by presenting the functional version of the single-index model (Stoker, 1986):

$$y_i = h\left(\int_{\mathcal{I}} \mathcal{X}_i(t)\beta(t)dt\right) + \varepsilon_i, \qquad i = 1,\ldots,n, \tag{6}$$

which extends the FLM by allowing the function $h(\cdot)$ to be any smooth function defined on the real line. Fitting this model requires estimation of both the coefficient function $\beta(\cdot)$ and the unspecified function $h(\cdot)$ and is typically accomplished in an iterative way. For given $\beta(\cdot)$, $h(\cdot)$ can be estimated using splines, kernels or any technique for estimating a smooth function; for given $h(\cdot)$, $\beta(\cdot)$ can be estimated in a similar fashion; and the process is iterated until convergence. Several of the methods described in Section 2 for estimating an FLM have been combined with a spline method for estimating $h(\cdot)$ in (6) (e.g. Eilers *et al.*, 2009). Alternatively, a kernel estimator can be used for $h(\cdot)$ (e.g. Ait-Saïdi *et al.*, 2008; Ferraty *et al.*, 2011).

3.2 Multiple-Index Model

A natural extension of model (6) is to allow multiple linear functionals of the predictor via the multiple-index model

$$y_i = \sum_{j=1}^{J} h_j\left(\int_{\mathcal{I}} \mathcal{X}_i(t)\beta_j(t)dt\right) + \varepsilon_i. \tag{7}$$

Models of this kind, which extend projection pursuit regression to the functional predictor case, are developed by James and Silverman (2005), Chen *et al.* (2011) and Ferraty *et al.* (2013).

Setting $\beta_j(t) = \phi_j(t)$, the j-th FPC basis function reduces (7) to $y_i = \sum_{j=1}^{J} h_j(x_{ij}) + \varepsilon_i$, where we use x_{ij} to denote FPC scores as in (3). Muller and Yao (2008) refer to this as a 'functional additive model', generalising the FLM of Section 2.2, which reduced to the multiple regression model (4) with respect to the FPC scores. An extension of this method, incorporating a sparsity-inducing penalty on the additive components, is proposed by Zhu *et al.* (2014).

3.3 Continuously Additive Model

Müller *et al.* (2013) and McLean *et al.* (2014) propose the model

$$y_i = \alpha + \int_{\mathcal{I}} f[\mathcal{X}_i(s), s]ds + \varepsilon_i, \tag{8}$$

where $f(\cdot,\cdot)$ is a smooth bivariate function that can be estimated by penalised tensor product B-splines. As an aid to interpretation, note that if $s_\ell = s_0 + \ell\Delta s$, $\ell = 1, \ldots, L$, and $\mathcal{I} = [s_0, s_L]$, then for large L, (8) implies

$$E(y_i|\mathcal{X}_i) \approx \sum_{\ell=1}^{L} f[\mathcal{X}_i(s_\ell), s_\ell]\Delta s = \sum_{\ell=1}^{L} g_\ell[\mathcal{X}_i(s_\ell)],$$

where $g_\ell(x) = f(x, s_\ell)\Delta s$. The expression at the right shows that (8) is the limit (as $L \to \infty$) of an additive model–or in the generalised linear extension considered by McLean *et al.* (2014), of a generalised additive model (Hastie & Tibshirani, 1990; Wood, 2006). Hence, McLean *et al.* (2014) employ the term 'functional generalised additive model'.

4 Non-Parametric Scalar-on-Function Regression

The monograph of Ferraty and Vieu (2006) has popularised a *non-parametric* paradigm for SoFR, in which the model for the conditional mean of y is not only non-linear but essentially unspecified, that is

$$E(y|\mathcal{X}) = m(\mathcal{X}) \tag{9}$$

for some operator $m : \mathcal{F} \longrightarrow \mathbb{R}$. Note that mathematically, the FLM can also be formulated as an operator, but as one that is linear–whereas Ferraty and Vieu (2006) focus primarily on non-linear operators m. (For further discussion of the terms *non-linear* and *non-parametric* SoFR, see Section 9.)

Approaches to estimating m extend traditional non-parametric regression methods (Härdle *et al.*, 2013) from the case in which the $\mathcal{X}_i$ are scalars or vectors to the case of function-valued $\mathcal{X}_i$. For example, the most popular non-parametric approach, which we consider next, generalises the Nadaraya–Watson (NW) smoother (Nadaraya, 1964; Watson, 1964) to functional predictors.

4.1 Functional Nadaraya–Watson Estimator

The functional NW estimator of the conditional mean $m(X) = \mathrm{E}(y|\mathcal{X} = X)$ is

$$\widehat{m}(X) = \frac{\sum_{i=1}^{n} K[d(X, \mathcal{X}_i)/h]y_i}{\sum_{i=1}^{n} K[d(X, \mathcal{X}_i)/h]}, \tag{10}$$

where $K(\cdot)$ is a kernel function, which we define as a function supported and decreasing on $[0, \infty)$; $h > 0$ is a bandwidth; and $d(\cdot,\cdot)$ is a semi-metric. Here, we define a semi-metric on $\mathcal{F}$ as a function $d : \mathcal{F} \times \mathcal{F} \longrightarrow [0, \infty)$ that is symmetric and satisfies the triangle inequality, but $d(f_1, f_2) = 0$ does not imply $f_1 = f_2$. (Such a function is often called a 'pseudo-metric'; our terminology has the advantage of implying that a semi-*norm* $\|\cdot\|$ on $\mathcal{F}$ induces a semi-metric $d(f_1, f_2) = \|f_1 - f_2\|$.) Smaller values of $d(X, \mathcal{X}_i)$ imply larger $K[d(X, \mathcal{X}_i)/h]$ and thus larger weight assigned to y_i.

Ideally, the bandwidth h should strike a good balance between the squared bias of $\widehat{m}(X)$ (which increases with h) and its variance (which decreases as h increases) (Ferraty *et al.*, 2007). Rachdi and Vieu (2007) consider a functional cross-validation method for bandwidth selection and prove its asymptotic optimality. Shang (2013, 2014a,2014b, 2015) and Zhang *et al.* (2014) propose a Bayesian method for simultaneously selecting the bandwidth and the unknown error density and show that it attains greater estimation accuracy than functional cross-validation.

Observe that if the fixed bandwidth h in (10) is replaced by $h_k(X)$, the k-th-smallest of the distances $d(X, \mathcal{X}_i)$ $(i = 1, \ldots, n)$, then we instead have a functional version of (weighted) k-nearest neighbours regression (Burba *et al.*, 2009).

4.2 Choice of Semi-Metric

The performance of the functional NW estimator can depend crucially on the chosen semi-metric (Geenens, 2011). Optimal selection of the semi-metric is discussed by Ferraty and Vieu (2006, Chapters 3 and 13) and is addressed using marginal likelihood by Shang (2015).

For smooth functional data, it may be appropriate to use the derivative-based semi-metric

$$d_q^{\text{deriv}}(\mathcal{X}_1, \mathcal{X}_2) = \sqrt{\int_{\mathcal{I}} \left[\mathcal{X}_1^{(q)}(t) - \mathcal{X}_2^{(q)}(t)\right]^2 dt},$$

where $\mathcal{X}^{(q)}$ is the qth-order derivative of $\mathcal{X}$. In practice, $q = 2$ is a popular choice (e.g. Goutis, 1998; Ferraty and Vieu, 2002,). The use of B-spline approximation for each curve allows straightforward computation of the derivatives.

For non-smooth functional data, it may be preferable to adopt a semi-metric based on FPCA truncated at A components,

$$d_A^{\text{FPCA}}(\mathcal{X}_1, \mathcal{X}_2) = \sqrt{\sum_{a=1}^{A} \left[\int_{\mathcal{I}} \{\mathcal{X}_1(t) - \mathcal{X}_2(t)\}\phi_a(t)dt\right]^2} = \sqrt{\sum_{a=1}^{A} (x_{1a} - x_{2a})^2},$$

where the last expression uses truncated expansions defined as in (3). A semi-metric based on FPLS components (Preda & Saporta, 2005; Reiss & Ogden, 2007) can be defined analogously. Chung *et al.* (2014) introduced a semi-metric based on thresholded wavelet coefficients of the functional data objects.

Given a set of possible semi-metrics with no a priori preference for any particular option, one can select the one that minimises a prediction error criterion such as a cross-validation score; more generally, one can adopt an ensemble predictor (see Section 6.2.2 and Fuchs *et al.*, 2015).

4.3 Functional Local Linear Estimator

The functional NW estimator (10) can also be written as

$$\widehat{m}(X) = \hat{\alpha} \equiv \min_{\alpha \in \mathbb{R}} \sum_{i=1}^{n} K[d(X, \mathcal{X}_i)/h](y_i - \alpha)^2.$$

As an alternative to this 'local constant' estimator, Baillo and Grane (2009) consider a functional analogue of local polynomial smoothing (Fan & Gijbels, 1996), specifically a local (functional) linear approximation

$$m(\mathcal{X}_i) \approx \alpha + \int_{\mathcal{I}} \beta(t)[\mathcal{X}_i(t) - X(t)]dt$$

for $\mathcal{X}_i$ near X, where $\alpha \in \mathbb{R}$ and $\beta \in L^2(\mathcal{I})$. This motivates the minimization problem

$$(\hat{\alpha}, \hat{\beta}) = \min_{\alpha,\beta} \sum_{i=1}^{n} K[d(\mathcal{X}_i, X)/h] \left[y_i - \alpha - \int_{\mathcal{I}} \beta(t)[\mathcal{X}_i(t) - X(t)]dt \right]^2,$$

whose solution yields the functional local linear estimate $\widehat{m}(X) = \hat{\alpha}$. Barrientos-Marin *et al.* (2010) propose a compromise between the NW (local constant) and local linear estimators, while Boj *et al.* (2010) offer a formulation based on more general distances.

4.4 A Reproducing Kernel Hilbert Space Approach

A rather different non-parametric method (Preda, 2007) is based on the notion of a positive semidefinite kernel $k(\mathcal{X}_1, \mathcal{X}_2)$ (Wahba, 1990; Schölkopf & Smola, 2002), which can roughly be thought of as defining a similarity between functions $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{F}$; the Gaussian kernel $\exp[-\int_{\mathcal{I}}\{\mathcal{X}_1(t) - \mathcal{X}_2(t)\}^2 dt/(2\sigma^2)]$ for some $\sigma > 0$ is an example. Briefly, any such $k(\cdot,\cdot)$ (not to be confused with the univariate kernel function $K(\cdot)$ of (10)) defines a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ of maps $m : \mathcal{F} \longrightarrow \mathbb{R}$, equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$. Preda (2007) considers more general loss functions, but for squared error loss, his proposed estimate of m in (9) is

$$\widehat{m}_\lambda = \operatorname*{arg\,min}_{m \in \mathcal{H}_k} \left[\sum_{i=1}^{n} [y_i - m(\mathcal{X}_i)]^2 + \lambda \|m\|^2_{\mathcal{H}_k} \right], \tag{11}$$

where λ is a non-negative regularisation parameter, as in criterion (2) for the FLM. A key RKHS result, the representer theorem (Kimeldorf & Wahba, 1971; Schölkopf *et al.*, 2001), leads to a much-simplified minimization in terms of the Gram matrix $\boldsymbol{K} = [k(\mathcal{X}_i, \mathcal{X}_i)]_{1 \le i,j \le n}$: (11) can be written as $\widehat{m}_\lambda(\cdot) = \sum_{i=1}^{n} \gamma_i k(\mathcal{X}_i, \cdot)$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^T$ minimises

$$\|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\gamma}\|^2 + \lambda \boldsymbol{\gamma}^T \boldsymbol{K} \boldsymbol{\gamma}$$

with $\boldsymbol{y} = (y_1, \ldots, y_n)^T$.

Note that this non-parametric formulation is distinct from the RKHS approach of Cai and Yuan (2012) to the functional linear model (1)–in which the coefficient function $\beta(\cdot)$, rather than the (generally non-linear) map $m : \mathcal{F} \longrightarrow \mathbb{R}$, is viewed as an element of an RKHS–as well as from the RKHS method of Zhu *et al.* (2014).

Expression (11) is a functional analogue of the criterion minimised in smoothing splines (Wahba, 1990), much as (10) generalises the NW smoother to functional predictors. We believe that, just as the RKHS-based smoothing spline paradigm has spawned a very flexible array of tools for non-parametric and semi-parametric regression (Ruppert *et al.*, 2003; Wood, 2006; Gu, 2013), there is great potential for building upon the regularised RKHS approach to non-parametric SoFR and thereby, perhaps, connecting FDA with machine learning.

As a further link between the (reproducing) kernel approach of Preda (2007) and other non-parametric approaches (such as the NW estimator) that are based on (semi-metric) distances among functions, we note that there is a well-known duality between kernels and distances (e.g. Faraway, 2012, p. 410). Because both kernels and distances can be defined for more general data types than functional data, the non-parametric FDA paradigm is readily extensible to 'object-oriented' data analysis (Marron & Alonso, 2014).

5 Generalisations and Extensions

Although many of the references cited in the previous three sections have considered more general scenarios, our presentation thus far has considered only the simplest situation: a single functional predictor and a real-valued response. However, situations that arise in practice often require various extensions, including scalar covariates, multiple functional predictors and models appropriate for responses that arise from a general exponential family distribution. In this section, we describe some of these generalisations and extensions.

5.1 Including Scalar Covariates

Including scalar covariates in the FLM (1) is fairly straightforward. For methods based on penalisation, covariates can be included by simply applying the penalty only to the spline basis coefficients. When using a data-driven basis, as noted earlier, the FLM reduces to the ordinary multiple regression model (4), so adding scalar covariates is even more routine. With non-linear and non-parametric strategies, however, incorporating scalar covariates can be somewhat more challenging. For example, Aneiros-Pérez and Vieu (2006, 2008) and Aneiros *et al.* (2011) considered semi-functional partial linear models of the form

$$y_i = \boldsymbol{z}_i^T \boldsymbol{\alpha} + m(\mathcal{X}_i) + \varepsilon_i, \tag{12}$$

which include linear effects of scalar covariates $\boldsymbol{z}_i$, estimated using weighted least squares, and effects of functional predictors $\mathcal{X}_i(\cdot)$, estimated non-parametrically via NW weights.

5.2 Multiple Functional Predictors

A number of recent papers have considered the situation in which the i-th observation includes multiple functional predictors $\mathcal{X}_{i1}, \ldots, \mathcal{X}_{iR}$, possibly with different domains $\mathcal{I}_1, \ldots, \mathcal{I}_R$. The FLM (1) extends naturally to the multiple functional regression model

$$y_i = \alpha + \sum_{r=1}^{R} \int_{\mathcal{I}_r} \mathcal{X}_{ir}(t)\beta_r(t)dt + \varepsilon_i.$$

Penalised or fully Bayesian approaches to selecting among candidate functional predictors have been proposed by Zhu *et al.* (2010), Gertheiss *et al.* (2013) and Lian (2013).

One can also consider two types of functional interaction terms. An interaction between a scalar and a functional predictor (e.g. McKeague and Qian, 2014) is formally similar to another functional predictor, whereas an interaction between two functional predictors (e.g. Yang *et al.*, 2013) resembles a functional quadratic term as in (5).

The 'functional additive regression' model of Fan *et al.* (2015) extends the functional single-index model (6) to the case of multiple predictors.

In the non-parametric FDA literature, Febrero-Bande and Gonzalez-Manteiga (2013) consider the model

$$y_i = \eta\left[\sum_{r=1}^{R} m_r(\mathcal{X}_{ir})\right] + \varepsilon_i, \tag{13}$$

where η can be a known function, or estimated non-parametrically; the $m_r(\cdot)$'s are non-linear partial functions of $\mathcal{X}_{ir}(t)$. Model (13), like model (8), is referred to by the authors as

a 'functional generalised additive model'. Lian (2011) studied the 'functional partial linear regression' model

$$y_i = \int_{\mathcal{I}} \mathcal{X}_{i1}(t)\beta(t)dt + m(\mathcal{X}_{i2}) + \varepsilon_i, \tag{14}$$

which combines linear and non-parametric functional terms. Note that both predictors in (14) are functional, whereas the linear terms in the 'semi-functional' model (12) are scalars.

5.3 Responses with Exponential Family Distributions

In all models considered to this point, the response variable has been a continuous, real-valued scalar. In many practical applications, the response is discrete, such as a binary outcome indicating the presence or absence of a disease. Many of the aforementioned methods have been generalised to allow responses with exponential-family distributions, including both linear (Marx & Eilers, 1999; James, 2002; Müller & Stadtmüller, 2005; Reiss & Ogden, 2010; Goldsmith *et al.*, 2011; Aguilera-Morillo *et al.*, 2013) and non-linear (James & Silverman, 2005; McLean *et al.*, 2014) models. For a single functional predictor and no scalar covariates, the functional generalised linear model can be written as $g(\mu_i) = \alpha + \int_{\mathcal{I}} \mathcal{X}_i(t)\beta(t)dt$, where $\mu_i = E[y_i|\mathcal{X}_i(\cdot)]$ and g is a known link function. Estimation for this model is analogous to the methods described in Section 2: using a data-driven basis recasts the functional model as a standard generalised linear model, and regularised basis expansion methods can be implemented using penalisation methods for GLMs. Scalar covariates and multiple functional predictors can be incorporated as in Sections 5.1 and 5.2.

5.4 Multilevel and Longitudinal Scalar-on-Function Regression

In recent years, it has become more common to collect repeated functional observations from each subject in a sample. In these situations, the data are $\left[\mathcal{X}_{ij}(\cdot), y_{ij}\right]$ for observations $j = 1, \ldots, J_i$ within each of subjects $i = 1, \ldots, n$. A relevant extension of the FLM is

$$y_{ij} = \alpha + b_i + \int_{\mathcal{I}} \mathcal{X}_{ij}(t)\beta(t)dt + \varepsilon_{ij}$$

with random effects $b_i \sim N\left(0, \sigma_b^2\right)$ used to model subject-specific effects. Goldsmith *et al.* (2012) directly extend the spline-based estimation strategy described in Section 2.1 for this model, and Gertheiss *et al.* (2013) use longitudinal FPCA (Greven *et al.*, 2010) to construct a data-driven basis for coefficient functions. Crainiceanu *et al.* (2009) consider the related setting in which the functional predictor is repeatedly observed but the response is not and estimate the coefficient function using a data-driven basis derived using multilevel FPCA (Di *et al.*, 2009).

5.5 Multidimensional Functional Predictors

While most methodological development of SoFR has focused on one-dimensional functional predictors as in (1), a growing number of authors have considered two-dimensional or three-dimensional signals or images as predictors in regression models. While this extension is relatively straightforward conceptually (replacing the single integral in (1) with a double or triple integral), it can involve significant technical challenges, including higher 'natural dimensionality', the need for additional tuning parameter values and correspondingly greater computational requirements.

Within the linear SoFR model framework, Marx and Eilers (2005) extend their penalised spline regression to handle higher dimensional signals by expressing the signals in terms of a tensor B-spline basis and applying both a 'row' and a 'column' penalty. Reiss and Ogden (2010) consider two-dimensional brain images as predictors, expressing them in terms of their eigenimages and enforcing smoothness via radially symmetric penalisation. Holan *et al.* (2010) and Holan *et al.* (2012) reduce the dimensionality of the problem by projecting the images on their (2D) principal components. Guillas and Lai (2010) apply penalised bivariate spline methods and consider two-dimensional functions on irregular regions. Zhou *et al.* (2013) and Zhou and Li (2014) propose methods that exploit the matrix or tensor structure of image predictors; Huang *et al.* (2013) and Goldsmith *et al.* (2014) develop Bayesian regression approaches for three-dimensional images; and Wang *et al.* (2014) and Reiss *et al.* (2015) describe wavelet-based methods.

5.6 Other Extensions

Our stated goal in this paper has been to describe and classify major areas of research in SoFR, and we acknowledge that any attempt to list all possible variants of SoFR would be futile. In this subsection, we very briefly mention a few models that do not fit neatly into the major paradigms discussed in Sections 2 through 4 or their direct extensions in Sections 5.1 through 5.5, knowing this list is incomplete.

5.6.1 Other non-independent, identically distributed settings

This review has focused on iid data pairs $(\mathcal{X}_i, y_i)$, with the exception of Section 5.4. Other departures from the iid assumption have received some attention in the SoFR literature. For example, Delaigle *et al.* (2009) considered heteroscedastic error variance, while Ferraty *et al.* (2005) studied α-mixing data pairs.

5.6.2 Mixture regression

A data set may be divided into latent classes, such that each class has a different regression relationship of the form (1). This is the model considered by Yao *et al.* (2011), who represent the predictors in terms of their functional principal components and apply a multivariate mixture regression model fitting technique. Ciarleglio & Ogden (2016) consider sparse mixture regression in the wavelet domain.

5.6.3 Point impact models

In some situations, it may be expected that only one point, or several points, along the function will be relevant to predicting the outcome. The model (1) could be adapted to reflect this by replacing the coefficient function β by a Dirac delta function at some point θ. This is the 'point impact' model considered by Lindquist and McKeague (2009) and McKeague and Sen (2010), who consider various methods for selecting one or more of these points. Ferraty *et al.* (2010) consider the same situation but within a non-parametric setting.

5.6.4 Derivatives in scalar-on-function regression

Derivatives have been incorporated in SoFR investigations in two completely different senses. The first is ordinary (first or higher-order) derivatives of functional observations with respect to the argument t. Some types of functional data may reflect vertical or linear shifts that are irrelevant to predicting y (Goutis, 1998). A key example is near-infrared (NIR) spectroscopy curves, which are used by analytical chemists to predict the contents of a sample: for

example, a wheat sample's spectrum may play the role of a functional predictor, with protein content as the response. It is sometimes helpful to remove such shifts by taking first or second differences (approximate derivatives) of the curves as a preprocessing step. For the same reason, derivative semi-metrics (see earlier discussion, Section 4.2) can be more useful than the L^2 metric for non-parametric SoFR with spectroscopy curves (Ferraty & Vieu, 2002).

A second type of derivative, studied by Hall *et al.* (2009), is the functional (Gâteaux) derivative of the operator m in the non-parametric model (9). Roughly speaking, for a given $X \in \mathcal{F}$, the functional derivative is a linear operator m_X such that for a small increment $\Delta X \in \mathcal{F}$, we have $m(X + \Delta X) \approx m(X) + m_X(\Delta X)$. In the linear case, the functional derivative is given by the coefficient or slope function, that is, $m_X(g) = \int_{\mathcal{I}} \beta(t)g(t)dt$ for all X and all g. In the non-parametric case, functional derivatives allow one to estimate functional gradients, which are in effect locally varying slopes. Muller and Yao (2010) simplify the study of functional derivatives and gradients by imposing the additive model framework of Muller and Yao (2008) (see earlier discussion, Section 3.2).

5.6.5 Conditional quantiles and mode

Up to now, we have been concerned with modelling the mean of y (or a transformation thereof, in the GLM case), conditional on functional predictors. Cardot *et al.* (2005) propose instead to estimate a given quantile of y, conditional on $\mathcal{X}$, by minimising a penalised criterion that is similar to (2), but with the squared error loss in the first term replaced by the 'check function' used in ordinary quantile regression (Koenker & Bassett, 1978). Chen and Muller (2012), on the other hand, estimate the entire conditional distribution of y by fitting functional binary GLMs with $I(y \leq y_0)$ (where $I(\cdot)$ denotes an indicator) as response, for a range of values of y_0. Quantiles can then be inferred by inverting the conditional distribution function. Ferraty *et al.* (2005) also estimate the entire conditional distribution, but adopt a non-parametric estimator of a weighted-average form reminiscent of (10). That paper and a number of subsequent ones have studied applications in the geosciences, such as an extreme value analysis of ozone concentration (Quintela-del-Río & Francisco-Fernández, 2011). The mode of the non-parametrically estimated conditional distribution serves as an estimate of the conditional mode of y, whose convergence rate is derived by Ferraty *et al.* (2005).

6 Which Method to Choose?

Sections 2 through 5 have presented a perhaps overwhelming variety of models and methods for SoFR. How can a user decide which is the appropriate method in a given situation? We offer here a few suggestions, which may be divided into a priori and data-driven considerations.

6.1 A Priori *Considerations*

Some authors have questioned whether the flexibility of the general non-parametric model (9) is worth the price paid in terms of convergence rates. The small-ball probability $\phi_X(u) = \Pr(\|\mathcal{X} - X\| < u)$ typically converges to 0 exponentially fast as $u \to 0$, and consequently, the functional NW estimator (10) converges at a rate that is logarithmic, as opposed to a power of $1/n$ (Hall *et al.*, 2009). Geenens (2011) succinctly interprets the problem as a function-space 'curse of dimensionality', but shows that a well-chosen semi-metric d increases the concentration of the functional predictors and thus allows for more favourable convergence of the NW estimator to m.

Aside from asymptotic properties, choice of a method may be guided by the kind of interpretation desired, which may in turn depend on the application. As noted in Section 2, the FLM

offers a coefficient function, which has an intuitive interpretation. Non-linear and especially non-parametric model results are less interpretable in this sense. But for applications in which one is interested only in accurate prediction, this advantage of the FLM is immaterial.

Regarding the FLM, we noted in Section 2.1 that smoothness of the data is an important factor when choosing among a priori basis types, such as splines versus wavelets. Among data-driven basis approaches to the FLM (Section 2.2), those that rely on a more parsimonious set of components may sometimes be preferred, again on grounds of interpretability. In this regard, FPLS has an advantage over FPCR, as emphasised by Delaigle and Hall (2012). On the other hand, if one is interested only in the coefficient function, not in contributions of the different components, then the relative simplicity of FPCR is an advantage over FPLS. A key advantage of FPCR over spline or wavelet methods is that it is more readily applied when the functional predictors are sampled not densely but sparsely and/or irregularly (longitudinal data).

We tend to view a wide variety of methods as effective in at least some settings, but one method for which we have limited enthusiasm is selecting a 'best subset' among a large set of FPCs, as opposed to regressing on the leading FPCs. Whereas leading FPCs offer an optimal approximation in the sense of Eckart and Young (1936), if one is not selecting the leading FPCs, then it is not clear why the FPC basis is the right one to use at all. Another basis that is by construction relevant for explaining y, such as FPLS components, may be more appropriate. This view finds support in the empirical results of Febrero-Bande *et al.* (2015).

Other considerations in choosing among methods include conceptual complexity, the number of tuning parameters that must be selected, sensitivity to how preprocessing is performed and computational efficiency. In practice, a primary factor for many users is the availability of software that is user-friendly and has the flexibility to handle data sets that may be more complex than the iid setup of the Introduction. Available software is reviewed in Section 7, and the real-data example presented in Section 8 highlights the need for software that is flexible.

6.2 Data-Driven Choice

Relying on one's *a priori* preference is unlikely to be the best strategy for choosing how to perform SoFR. Next, we discuss two ways to let the data help determine the best approach.

6.2.1 Hypothesis testing

Hypothesis testing is an appropriate paradigm when one wishes to compare a simpler versus a more complex SoFR model, with strong evidence required in order to reject the former in favour of the latter. By now, there are enough articles on tests of SoFR-related hypotheses to justify a separate review paper; here, we offer just a few remarks. Many papers have proposed tests for the FLM (1). Some methods test the null hypothesis $\beta(t) \equiv 0$ in (1) versus the alternative that $\beta(t) \neq 0$ for some $t \in \mathcal{I}$ (e.g. Cardot *et al.*, 2003; Lei, 2014). The (restricted) likelihood ratio test of Swihart *et al.* (2014) can be applied either to that zero-effect null or, alternatively, to the null hypothesis that $\beta(t)$ is a constant–a hypothesis that, if true, allows one to regress on the across-the-function average rather than resorting to functional regression. McLean *et al.* (2015) treat the linear model (1) as the null, to be tested versus the additive model (8), while García-Portugués *et al.* (2014) consider testing against a more general alternative. Horvath and Kokoszka (2012) and Zhang *et al.* (2014) investigate hypothesis testing procedures to choose the polynomial order in functional polynomial models such as the quadratic model (5). Delsol *et al.* (2011) propose a test for the null hypothesis that m in the non-parametric model (9) belongs to a given family of operators.

6.2.2 Ensemble predictors

In most practical cases, there are not just two plausible models–a null and an alternative–but many reasonable options for performing SoFR, and it is impossible to know in advance which model and estimation strategy will work best for a given data set. Because of this, Goldsmith and Scheipl (2014) extended the idea of model stacking (Wolpert, 1992) (or superlearning, van der Laan *et al.*, 2007) to SoFR. A large collection of estimators are applied to the data set of interest and are evaluated for prediction accuracy using cross-validation. These estimators are combined into an 'ensemble' predictor based on their individual performance. These authors found that multiple approaches yielded dramatically different relative performance across several example data sets–underlining the value of trying a variety of approaches to SoFR when possible.

7 Software for Scalar-on-Function Regression

Several packages for R (R Core Team, 2015) implement SoFR. The function `fRegress` in the `fda` package (Ramsay *et al.*, 2009) fits linear models in which either the response and/or the predictor is functional. The `fda.usc` package (Febrero-Bande & Oviedo de la Fuente, 2012) implements an extensive range of parametric and non-parametric functional regression methods, including those studied by Ramsay and Silverman (2005), Ferraty and Vieu (2006) and Febrero-Bande and Gonzalez-Manteiga (2013). The `refund` package (Huang *et al.*, 2015) implements penalised functional regression, including several variants of the FLM and the additive model (8), and allows for multiple functional predictors and scalar covariates, as well as generalised linear models. Optimal smoothness selection relies on the `mgcv` package of Wood (2006). The `mgcv` package itself is one of the most flexible and user-friendly packages for SoFR, allowing for the whole GLASS structure of Eilers and Marx (2002) plus random effects; see section 5.2 of Wood (2011), which incidentally was the second paper to use the term 'scalar-on-function regression'. In `mgcv`, functional predictor terms are treated as just one instance of 'linear functional' terms (cf. the 'general spline problem' of Wahba, 1990). The `refund.wave` package (Huo *et al.*, 2014), a spinoff of `refund`, implements scalar-on-function and scalar-on-image regression in the wavelet domain.

In MATLAB (MathWorks, Natick, MA, USA), the `PACE` package implements a wide variety of methods by Müller, Wang, Yao and co-authors, including a versatile collection of functional principal component-based regression models for dense and sparsely sampled functional data. From a Bayesian viewpoint, Crainiceanu and Goldsmith (2010) have developed tools for functional generalised linear models using WinBUGS (Lunn *et al.*, 2000).

8 A Brain Imaging Example

Our work has made us aware of the expanding opportunities to apply SoFR in cutting-edge biomedical research, a trend that we expect will accelerate in the coming years. For such applications to be feasible, one must typically move beyond the basic model (1) and incorporate scalar covariates and/or multiple functional predictors, allow for non-iid data and consider hypothesis tests for the coefficient function(s). The following example illustrates how FLMs can incorporate some of these features (thanks to flexible software implementation, as advocated in Section 6.1) and can shed light on an interesting scientific question. For comparisons of linear and non-linear and non-parametric approaches applied to several data sets, we refer the reader to Goldsmith and Scheipl (2014).

Lindquist (2012) analysed functional magnetic resonance imaging (fMRI) measures of response to hot (painful) and warm (non-painful) stimuli applied to the left volar forearm in 20 volunteers. Here, we apply SoFR to examine whether the blood oxygen level dependent (BOLD) response measured by fMRI in the lateral cerebellum predicts the pain intensity, as rated after each stimulus on a scale from 100 to 550 (with higher values indicating more pain). Each trial consisted of thermal stimulation for 18 s; then a 14-s interval in which a fixation cross was presented on a screen; then the words 'How painful?' appeared, and after another 14-s interval, the participant rated the pain intensity. The BOLD signal was recorded at 23 2-s intervals. There were $J_i = 39$–48 such trials per volunteer, and 940 in total. As shown in Figure 1(a), the lateral cerebellum BOLD signal tends to be higher in the hot trials, but only during the fixation cross interval.

We fitted the model

$$y_{ij} = \alpha_i + \gamma I_{ij}^{\text{hot}} + \int_{\mathcal{I}} \mathcal{X}_{ij}(t)\beta(t)dt + \varepsilon_{ij}, \quad i = 1, \ldots, n, j = 1, \ldots, J_i, \tag{15}$$

in which y_{ij} is the log pain score for the i-th participant's j-th trial, I_{ij}^{hot} is an indicator for a hot stimulus, the α_i's are iid normally distributed random intercepts, $\mathcal{I}$ is the time interval of the trial and the ε_{ij}'s are iid normally distributed errors with mean zero.

Unsurprisingly, the expected difference in log pain score between hot and warm trials is found to be hugely significant ($\hat{\gamma} = 0.599$, $p < 2 \cdot 10^{-16}$). The coefficient function is also very significantly non-zero, based on the modified Wald test of Wood (2013) or the likelihood ratio test of Swihart *et al.* (2014).

In particular, Figure 1(b) shows that $\hat{\beta}(t)$ is clearly positive for the same time points that evince a warm/hot discrepancy in Figure 1(a). One could venture a 'functional collinearity' explanation for this, in view of the high correlation (0.28) between the indicator variable I_{ij}^{hot} and the functional predictor term $\int_{\mathcal{I}} \mathcal{X}_{ij}(t)\hat{\beta}(t)dt$. In other words, one might suspect that the highly positive values of $\hat{\beta}(t)$ around the 20-s to 28-s range are just an artefact of the higher

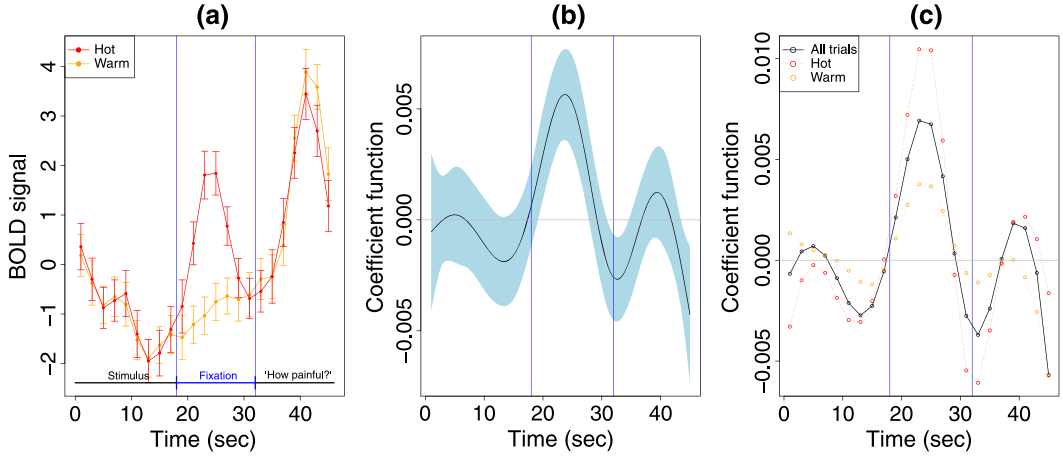

Figure 1. *(a) Overall mean lateral cerebellum blood oxygen level dependent (BOLD) signal at each of the 23 time points, with approximate 95% confidence intervals, for warm-stimulus and hot-stimulus trials. (b) Estimated coefficient function, with approximate pointwise 95% confidence intervals, for model (15). (c) Coefficient function estimate for the full data set (as in (b)) compared with those obtained with only the hot or only the warm trials. [Colour figure can be viewed at wileyonlinelibrary.com]*

BOLD signal in that range for hot versus warm trials. But that seems incorrect, because $\hat{\beta}$ looks very similar even when we fit separate models with only the warm or only the hot trials (Figure 1(c)). A more cogent explanation is that brain activity detected by the BOLD signal partially mediates the painful effect of the hot stimulus, a causal interpretation developed in depth by Lindquist (2012).

The SoFR implementations of Wood (2011) in the R package `mgcv` and of Goldsmith *et al.* (2011) in the `refund` package–which were used to create Figure 1(b) and (c), respectively–make it routine to include scalar covariates and random effects, and to test the significance of the coefficient function, as we have done here. It is also straightforward to test for a 'scalar-by-function' interaction between stimulus type and BOLD signal (found to be non-significant); or to include multiple BOLD signal predictors. Besides the lateral cerebellum, the data set includes such functional predictors for 20 other pain-relevant regions. Several of these have significant effects on pain, even adjusting for the lateral cerebellum signal, but the associated increments in explained variance are quite small.

9 Discussion

This paper has emphasised methodological and practical aspects of some widely used SoFR models. Just a few brief words will have to suffice regarding asymptotic issues. For functional (generalised) linear models (see Cardot and Sarda, 2011, for a survey), convergence rates have been studied both for estimation of the coefficient function (e.g. Müller and Stadtmüller, 2005; Hall and Horowitz, 2007; Dou *et al.*, 2012) and for prediction of the response (e.g. Cai and Hall, 2006; Crambes *et al.*, 2009); see Hsing & Eubank (2015) for a succinct treatment, and Reimherr (2015) for a recent contribution on the key role played by eigenvalues of the covariance operator. A few recent papers (e.g. Wang *et al.*, 2014; Brunel and Roche, 2015) have derived non-asymptotic error bounds for FLMs. In non-parametric FDA, only prediction error is considered, with small-ball probability playing a central role. Ferraty and Vieu (2011) review both pointwise and uniform convergence results. A detailed analysis of the functional NW estimator's asymptotic properties was recently provided by Geenens (2015).

Regarding the term *non-parametric* in the FDA context, Ferraty *et al.* (2005) explain: 'We use the terminology *functional non-parametric*, where the word *functional* refers to the infinite dimensionality of the data and where the word *non-parametric* refers to the infinite dimensionality of the model'. But some would argue that, because the coefficient function in (1) lies in an infinite-dimensional space, this nomenclature makes even the functional linear model 'non-parametric'. While a fully satisfying definition may be elusive, we find it most helpful to think of non-parametric SoFR as an analogue of ordinary non-parametric regression, that is, as extending the non-parametric model (9) to the case of functional predictors.

Our use of the term *non-linear* in Section 3 can likewise be questioned, because that term applies equally well to non-parametric SoFR. But we could find no better term to encompass models that are not linear but that impose more structure than the general model (9). Note that in non-functional statistics as well, *non-linear* usually refers to models that have some parametric structure, as opposed to leaving the mean completely unspecified.

Other problems of nomenclature have arisen as methods for SoFR have proliferated. We have seen that 'functional (generalised) additive models' may be additive with respect to FPC scores (Müller & Yao, 2008; Zhu *et al.*, 2014); points along a functional predictor domain (Müller *et al.*, 2013; McLean *et al.*, 2014); or multiple functional predictors with either non-linear (Fan *et al.*, 2015) or non-parametric (Febrero-Bande & González-Manteiga, 2013) effects. Likewise, authors have considered every combination of scalar linear (SL), functional linear (FL) and

functional non-parametric (FNP) terms, so that 'partial linear' models may refer to any two of these three: SL+FL (Shin, 2009), SL+FNP (Aneiros-Pérez & Vieu, 2008) or FL+FNP (Lian, 2011). Finally, some of the non-linear approaches of Section 3 are sometimes termed 'non-parametric' because they incorporate general smooth link functions. One of our aims here has been to reduce terminological confusion.

We hope that, by distilling some key ideas and approaches from what is now a sprawling literature, we have provided readers with useful guidance for implementing scalar-on-function regression in the growing number of domains in which it can be applied.

Acknowledgments

We thank the Co-Editor-in-Chief, Prof. Marc Hallin, and the Associate Editor and referees, whose feedback enabled us to improve the manuscript significantly. We also thank Prof. Martin Lindquist for providing the pain data, whose collection was supported by the U.S. National Institutes of Health through grants R01MH076136-06 and R01DA035484-01, and Pei-Shien Wu for assistance with the bibliography. Philip Reiss' work was supported in part by grant 1R01MH095836-01A1 from the National Institute of Mental Health. Jeff Goldsmith's work was supported in part by grants R01HL123407 from the National Heart, Lung, and Blood Institute and R21EB018917 from the National Institute of Biomedical Imaging and Bioengineering. Han Lin Shang's work was supported in part by a Research School Grant from the ANU College of Business and Economics. Todd Ogden's work was supported in part by grant 5R01MH099003 from the National Institute of Mental Health.

References

Aguilera, A.M., Escabias, M., Preda, C. & Saporta, G. (2010). Using basis expansions for estimating functional PLS regression: applications with chemometric data. *Chemometr. Intell. Lab.*, **104**(2), 289–305.

Aguilera-Morillo, M.C., Aguilera, A.M., Escabias, M. & Valderrama, M.J. (2013). Penalized spline approaches for functional logit regression. *Test*, **22**(2), 251–277.

Ait-Saïdi, A., Ferraty, F., Kassa, R. & Vieu, P. (2008). Cross-validated estimations in the single-functional index model. *Statistics*, **42**(6), 475–494.

Amato, U., Antoniadis, A. & De Feis, I. (2006). Dimension reduction in functional regression with applications. *Comput. Stat. Data Anal.*, **50**(9), 2422–2446.

Aneiros, G., Cao, R., Vilar-Fernández, J. M. & Muñoz-San-Roque, A. (2011). Functional prediction for the residual demand in electricity spot markets. In *Recent Advances in Functional Data Analysis and Related Topics*, Ed. F. Ferraty, pp. 9–15. Berlin: Springer-Verlag.

Aneiros-Pérez, G. & Vieu, P. (2006). Semi-functional partial linear regression. *Stat Probabil. Lett.*, **76**(11), 1102–1110.

Aneiros-Pérez, G. & Vieu, P. (2008). Nonparametric time series prediction: a semi-functional partial linear modeling. *J. Multivar. Anal.*, **99**(5), 834–857.

Araki, Y., Kawaguchi, A. & Yamashita, F. (2013). Regularized logistic discrimination with basis expansions for the early detection of Alzheimer's disease based on three-dimensional MRI data. *Adv. Data Anal. Classif.*, **7**(1), 109–119.

Baíllo, A. & Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivar. Anal.*, **100**(1), 102–111.

Barrientos-Marin, J., Ferraty, F. & Vieu, P. (2010). Locally modelled regression and functional data. *J. Nonparametr. Stat.*, **22**(5), 617–632.

Boj, E., Delicado, P. & Fortiana, J. (2010). Distance-based local linear regression for functional predictors. *Comput. Stat. Data Anal.*, **54**(2), 429–437.

Brown, P. J., Fearn, T. & Vannucci, M. (2001). Bayesian wavelet regression on curves with application to a spectroscopic calibration problem. *JASA*, **96**, 398–408.

Brunel, E. & Roche, A. (2015). Penalized contrast estimation in functional linear models with circular data. *Statistics*, **49**(6), 1298–1321.

Burba, F., Ferraty, F. & Vieu, P. (2009). k-nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.*, **21**(4), 453–469.
Cai, T. T. & Hall, P. (2006). Prediction in functional linear regression. *Ann. Stat.*, **34**(5), 2159–2179.
Cai, T. T. & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *JASA*, **107**, 1201–1216.
Cardot, H., Crambes, C. & Sarda, P. (2005). Quantile regression when the covariates are functions. *J. Nonparametr. Stat.*, **17**(7), 841–856.
Cardot, H., Ferraty, F., Mas, A. & Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.*, **30**(1), 241–255.
Cardot, H., Ferraty, F. & Sarda, P. (1999). Functional linear model. *Stat Probabil. Lett.*, **45**(1), 11–22.
Cardot, H., Ferraty, F. & Sarda, P. (2003). Spline estimators for the functional linear model. *Stat. Sin.*, **13**, 571–591.
Cardot, H. & Sarda, P. (2011). Functional linear regression. In *The Oxford Handbook of Functional Data Analysis*, Eds. F. Ferraty & Y. Romain, pp. 21–46. New York: Oxford University Press.
Chen, D., Hall, P. & Müller, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Stat.*, **39**(3), 1720–1747.
Chen, K. & Müller, H.-G. (2012). Conditional quantile analysis when covariates are functions, with application to growth data. *J. R. Stat. Soc. Series B Stat. Methodol.*, **74**(1), 67–89.
Chung, C., Chen, Y. & Ogden, R. T. (2014). Functional data classification: a wavelet approach. *Comput. Stat.*, **14**, 1497–1513.
Ciarleglio, A. & Ogden, R. T. (2016). Wavelet-based scalar-on-function finite mixture regression models. *Comput. Stat. Data Anal.*, **93**, 86–96.
Crainiceanu, C. M. & Goldsmith, J. (2010). Bayesian functional data analysis using WinBUGS. *J. Stat. Softw.*, **32**, 1–33.
Crainiceanu, C. M., Staicu, A. M. & Di, C.-Z. (2009). Generalized multilevel functional regression. *JASA*, **104**, 1550–1561.
Crambes, C., Kneip, A. & Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Stat.*, **37**(1), 35–72.
Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**(4), 317–403.
Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Stat. Plan. Inference*, **147**, 1–23.
Delaigle, A. & Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *Ann. Stat.*, **40**(1), 322–352.
Delaigle, A., Hall, P. & Apanasovich, T. V. (2009). Weighted least squares methods for prediction in the functional data linear model. *Electron. J. Stat.*, **3**, 865–885.
Delsol, L., Ferraty, F. & Vieu, P. (2011). Structural test in regression on functional variables. *J. Multivar. Anal.*, **102**(3), 422–447.
Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. & Punjabi, N. M. (2009). Multilevel functional principal component analysis. *Ann. Appl. Stat.*, **3**(1), 458–488.
Dou, W.W., Pollard, D. & Zhou, H.H. (2012). Estimation in functional regression for general exponential families. *Ann. Stat.*, **40**(5), 2421–2451.
Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**(3), 211–218.
Eilers, P.H.C., Li, B. & Marx, B.D. (2009). Multivariate calibration with single-index signal regression. *Chemometr. Intell. Lab.*, **96**(2), 196–202.
Eilers, P. H. C. & Marx, B. D. (2002). Generalized linear additive smooth structures. *J. Comp. Graph. Stat.*, **11**(4), 758–783.
Escabias, M., Aguilera, A.M. & Valderrama, M.J. (2007). Functional PLS logit regression model. *Comput. Stat. Data Anal.*, **51**(10), 4891–4902.
Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London: Chapman & Hall.
Fan, Y., James, G. & Radchenko, P. (2015). Functional additive regression. *Ann. Stat.*, **43**(5), 2296–2325.
Faraway, J.J. (2012). Backscoring in principal coordinates analysis. *J. Comp. Graph. Stat.*, **21**(2), 394–412.
Febrero-Bande, M., Galeano, P. & González-Manteiga, W. (2015). Functional principal component regression and functional partial least-squares regression: an overview and a comparative study. *Int. Stat. Rev.*, DOI 10.1111/insr.12116, (to appear in print).
Febrero-Bande, M. & González-Manteiga, W. (2013). Generalized additive models for functional data. *Test*, **22**(2), 278–292.
Febrero-Bande, M. & Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the R package fda.usc. *J. Stat. Softw.*, **51**(4), 1–28.

Ferraty, F., Goia, A., Salinelli, E. & Vieu, P. (2013). Functional projection pursuit regression. *Test*, **22**(2), 293–320.
Ferraty, F., Hall, P. & Vieu, P. (2010). Most-predictive design points for functional data predictors. *Biometrika*, **97**(4), 807–824.
Ferraty, F., Laksaci, A. & Vieu, P. (2005). Functional time series prediction via conditional mode estimation. *C. R. Acad. Sci., Series I*, **340**(5), 389–392.
Ferraty, F., Mas, A. & Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust N Z J Stat.*, **49**(3), 267–286.
Ferraty, F., Park, J. & Vieu, P. (2011). Estimation of a functional single index model. In *Recent Advances in Functional Data Analysis and Related Topics*, Ed. F. Ferraty, pp. 111–116. Berlin: Springer.
Ferraty, F., Rabhi, A. & Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon. *Sankhyā: The Indian Journal of Statistics*, **67**(2), 378–398.
Ferraty, F. & Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Stat.*, **17**(4), 545–564.
Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* New York: Springer.
Ferraty, F., Vieu, P. & 2009. Additive prediction and boosting for functional data. *Comput. Stat. Data Anal.*, **53**(4), 1400–1413.
Ferraty, F. & Vieu, P. (2011). Kernel regression estimation for functional data. In *The Oxford Handbook of Functional Data Analysis*, Eds. F. Ferraty & Y. Romain, pp. 72–129. New York: Oxford University Press.
Ferré, L. & Yao, A. F. (2003). Functional sliced inverse regression analysis. *Stat.*, **37**(6), 475–488.
Ferré, L. & Yao, A.-F. (2005). Smoothed functional inverse regression. *Stat. Sin.*, **15**, 665–683.
Fuchs, K., Gertheiss, J. & Tutz, G. (2015). Nearest neighbor ensembles for functional data with interpretable feature selection. *Chemometr. Intell. Lab.*, **146**, 186–197.
García-Portugués, E., González-Manteiga, W. & Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar responses. *J. Comp. Graph. Stat.*, **23**(3), 761–778.
Geenens, G. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.*, **5**, 30–43.
Geenens, G. (2015). Moments, errors, asymptotic normality and large deviation principle in nonparametric functional regression. *Stat. Probabil. Lett.*, **107**, 369–377.
Gertheiss, J., Goldsmith, J., Crainiceanu, C. & Greven, S. (2013). Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics*, **14**(3), 447–461.
Gertheiss, J., Maity, A. & Staicu, A.-M. (2013). Variable selection in generalized functional linear models. *Stat.*, **2**(1), 86–101.
Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B. & Reich, D. (2011). Penalized functional regression. *J. Comp. Graph. Stat.*, **20**(4), 830–851.
Goldsmith, J., Crainiceanu, C. M., Caffo, B. & Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc. Series C Stat. Methodol.*, **61**(3), 453–469.
Goldsmith, J., Huang, L. & Crainiceanu, C. M. (2014). Smooth scalar-on-image regression via spatial Bayesian variable selection. *J. Comp. Graph. Stat.*, **23**(1), 46–64.
Goldsmith, J. & Scheipl, F. (2014). Estimator selection and combination in scalar-on-function regression. *Comput. Stat. Data Anal.*, **70**, 362–372.
Goldsmith, J., Wand, M. P. & Crainiceanu, C. M. (2011). Functional regression via variational Bayes. *Electron. J. Stat.*, **5**, 572–602.
Goutis, C. (1998). Second-derivative functional regression with applications to near infra-red spectroscopy. *J. R. Stat. Soc. Series B Stat. Methodol.*, **60**(1), 103–114.
Greven, S., Crainiceanu, C. M., Caffo, B. & Reich, D. (2010). Longitudinal functional principal component analysis. *Electron. J. Stat.*, **4**, 1022–1054.
Gu, C. (2013). *Smoothing Spline ANOVA Models*, 2nd ed. New York: Springer.
Guillas, S. & Lai, M.-J. (2010). Bivariate splines for spatial functional regression models. *J. Nonparametr. Stat.*, **22**(4), 477–497.
Hall, P. & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Stat.*, **35**(1), 70–91.
Hall, P., Müller, H.-G. & Yao, F. (2009). Estimation of functional derivatives. *Ann. Stat.*, **37**(6A), 3307–3329.
Hall, P. & Vial, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Series B Stat. Methodol.*, **68**(4), 689–705.
Härdle, W., Müller, M., Sperlich, S. & Werwatz, A. (2013). *Nonparametric and Semiparametric Models*, 2nd ed. New York: Springer.
Hastie, T. & Mallows, C. (1993). Discussion of a statistical view of some chemometrics regression tools by by I.E. Frank and J.H. Friedman. *Technometrics*, **35**(2), 140–143.

Hastie, T. & Tibshirani, R. (1990). *Generalized Additive Models.* Boca Raton, FL: Chapman & Hall/CRC.
Holan, S., Yang, W., Matteson, D. & Wikle, C. (2012). An approach for identifying and predicting economic recessions in real-time using time-frequency functional models. (with discussion). *Appl. Stoch. Model Bus.*, **28**, 485–489.
Holan, S. H., Wikle, C. K., Sullivan-Beckers, L. E. & Cocroft, R. B. (2010). Modeling complex phenotypes: generalized linear models using spectrogram predictors of animal communication signals. *Biometrics*, **66**, 914–924.
Horváth, L. & Kokoszka, P. (2012). *Inference for Functional Data with Applications.* New York: Springer.
Hsing, T. & Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* West Sussex, UK: John Wiley & Sons.
Hosseini-Nasab, M. (2013). Cross-validation approximation in functional linear regression. *JSCS*, **83**(8), 1429–1439.
Huang, L., Goldsmith, J., Reiss, P. T., Reich, D. S. & Crainiceanu, C. M. (2013). Bayesian scalar-on-image regression with application to association between intracranial DTI and cognitive outcomes. *NeuroImage*, **83**, 210–223.
Huang, L., Scheipl, F., Goldsmith, J., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C. & Reiss, P. (2015). refund: regression with functional data. R package version 0.1-13.
Huo, L., Reiss, P. & Zhao, Y. (2014). refund.wave: wavelet-domain regression with functional data. R package version 0.1.
James, G. M. (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Series B Stat. Methodol.*, **64**(3), 411–432.
James, G. M. & Silverman, B. (2005). Functional adaptive model estimation. *JASA*, **100**, 565–576.
James, G. M., Wang, J. & Zhu, J. (2009). Functional linear regression that's interpretable. *Ann. Stat.*, **37**(5A), 2083–2108.
Kimeldorf, G. & Wahba, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.*, **33**(1), 82–95.
Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**(1), 33–50.
Lee, E. R. & Park, B. U. (2011). Sparse estimation in functional linear regression. *J. Multivar. Anal.*, **105**(1), 1–17.
Lei, J. (2014). Adaptive global testing for functional linear models. *JASA*, **109**, 624–634.
Li, Y., Wang, N. & Carroll, R. J. (2013). Selecting the number of principal components in functional data. *JASA*, **108**, 1284–1294.
Lian, H. (2011). Functional partial linear model. *J. Nonparametr. Stat.*, **23**(1), 115–128.
Lian, H. (2013). Shrinkage estimation and selection for multiple functional regression. *Stat. Sin.*, **23**, 51–74.
Lindquist, M. & McKeague, I. (2009). Logistic regression with Brownian-like predictors. *JASA*, **104**, 1575–1585.
Lindquist, M. A. (2012). Functional causal mediation analysis with an application to brain connectivity. *JASA*, **107**, 1297–1309.
Lunn, D. J., Thomas, A., Best, N. & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.*, **10**(4), 325–337.
Malloy, E., Morris, J., Adar, S., Suh, H., Gold, D. & Coull, B. (2010). Wavelet-based functional linear mixed models: an application to measurement error-corrected distributed lag models. *Biostatistics*, **11**(3), 432–452.
Marron, J. S. & Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical J.*, **56**(5), 732–753.
Marx, B. D. & Eilers, P. H. C. (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, **41**(1), 1–13.
Marx, B. D. & Eilers, P. H. C. (2005). Multidimensional penalized signal regression. *Technometrics*, **47**(1), 12–22.
McKeague, I. W. & Qian, M. (2014). Estimation of treatment policies based on functional predictors. *Stat. Sin.*, **24**(3), 1461–1485.
McKeague, I. W. & Sen, B. (2010). Fractals with point impact in functional linear regression. *Ann. Stat.*, **38**, 2559–2586.
McLean, M. W., Hooker, G. & Ruppert, D. (2015). Restricted likelihood ratio tests for linearity in scalar-on-function regression. *Stat. Comput.*, **25**(5), 997–1008.
McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F. & Ruppert, D. (2014). Functional generalized additive models. *J. Comp. Graph. Stat.*, **23**(1), 249–269.
Morris, J. S. (2015). Functional regression. *Annu. Rev. Stat. Appl.*, **2**, 321–359.
Müller, H.-G. & Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Stat.*, **33**(2), 774–805.
Müller, H.-G., Wu, Y. & Yao, F. (2013). Continuously additive models for nonlinear functional regression. *Biometrika*, **100**(3), 607–622.
Müller, H.-G. & Yao, F. (2008). Functional additive models. *JASA*, **103**, 1534–1544.
Müller, H.-G. & Yao, F. (2010). Additive modelling of functional gradients. *Biometrika*, **97**(4), 791–805.
Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.*, **9**(1), 141–142.
Ogden, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis.* Boston: Birkhauser.
Preda, C. (2007). Regression models for functional data by reproducing kernel Hilbert spaces methods. *J. Stat. Plan. Inference*, **137**(3), 829–840.
Preda, C. & Saporta, G. (2005). PLS regression on a stochastic process. *Comput. Stat. Data Anal.*, **48**(1), 149–158.

Quintela-del-Río, A. & Francisco-Fernández, M. (2011). Nonparametric functional data estimation applied to ozone data: prediction and extreme value analysis. *Chemosphere*, **82**(6), 800–808.
R Core Team. (2015). *R: a language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing.
Rachdi, M. & Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Stat. Plan. Inference*, **137**(9), 2784–2801.
Ramsay, J. O., Hooker, G. & Graves, S. (2009). *Functional Data Analysis with R and MATLAB.* New York: Springer.
Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer.
Randolph, T. W., Harezlak, J. & Feng, Z. (2012). Structured penalties for functional linear models–partially empirical eigenvectors for regression. *Electron. J. Stat.*, **6**, 323–353.
Ratcliffe, S. J., Leader, L. R. & Heller, G. Z. (2002). Functional data analysis with application to periodically stimulated fetal heart rate data: I. Functional regression. *Stat. Med.*, **21**(8), 1103–1114.
Reimherr, M. (2015). Functional regression with repeated eigenvalues. *Stat. Probabil. Lett.*, **107**, 62–70.
Reiss, P.T., Huang, L. & Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *Int. J. Biostat.*, **6**(1). Article 28.
Reiss, P. T., Huo, L., Zhao, Y., Kelly, C. & Ogden, R. T. (2015). Wavelet-domain regression and predictive inference in psychiatric neuroimaging. *Ann. Appl. Stat.*, **9**(2), 1076–1101.
Reiss, P. T. & Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *JASA*, **102**, 984–996.
Reiss, P. T. & Ogden, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *J. R. Stat. Soc. Series B Stat. Methodol.*, **71**(2), 505–523.
Reiss, P. T. & Ogden, R. T. (2010). Functional generalized linear models with images as predictors. *Biometrics*, **66**(1), 61–69.
Rice, J. & Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Series B Stat. Methodol.*, **53**, 233–243.
Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric Regression.* Cambridge: Cambridge University Press.
Schölkopf, B., Herbrich, R. & Smola, A. J. (2001). A generalized representer theorem, pp. 416–426.
Schölkopf, B. & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* Cambridge, Massachusetts: MIT Press.
Shang, H. L. (2013). Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density. *Comput. Stat. Data Anal.*, **67**, 185–198.
Shang, H. L. (2014). Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density. *J. Nonparametr. Stat.*, **26**, 599–615.
Shang, H. L. (2014). Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density. *Comput. Stat.*, **29**, 829–848.
Shang, H. L. (2015). A Bayesian approach for determining the optimal semi-metric and bandwidth in scalar-on-function quantile regression with unknown error density and dependent functional data. *J. Multivar. Anal.*, DOI 10.1016/j.jmva.2015.06.015, (to appear in print).
Shin, H. (2009). Partial functional linear regression. *J. Stat. Plan. Inference*, **139**(10), 3405–3418.
Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Stat.*, **24**(1), 1–24.
Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica*, **54**(6), 1461–1481.
Swihart, B. J., Goldsmith, J. & Crainiceanu, C. M. (2014). Restricted likelihood ratio tests for functional effects in the functional linear model. *Technometrics*, **56**, 483–493.
van der Laan, M. J., Polley, E. C. & Hubbard, A. E. (2007). Super learner. *Stat. Appl. Genet. Molec. Biol.*, **6**(1), Article 25.
Wahba, G. (1990). *Spline Models for Observational Data.* Philadelphia: Society for Industrial and Applied Mathematics.
Wang, J.-L., Chiou, J.-M. & Müller, H.-G. (2015). Review of functional data analysis. arXiv preprint arXiv:1507.05135.
Wang, X., Nan, B., Zhu, J., Koeppe, R. & the Alzheimer's Disease Neuroimaging Initiative. (2014). Regularized 3D functional regression for brain image data via Haar wavelets. *Ann. Appl. Stat.*, **8**(2), 1045–1064.
Watson, G. S. (1964). Smooth regression analysis. *Sankhya A*, **26**, 359–372.
Wolpert, D. H. (1992). Stacked generalization. *Neural Netw.*, **5**, 241–259.
Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R.* London: Chapman & Hall.
Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Series B Stat. Methodol.*, **73**(1), 3–36.

Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, **100**(1), 221–228.
Yang, W.-H., Wikle, C. K., Holan, S. H. & Wildhaber, M. L. (2013). Ecological prediction with nonlinear multivariate time-frequency functional data models. *J. Agr. Biol. Envir. St.*, **18**(3), 450–474.
Yao, F., Fu, Y. & Lee, T. C. M. (2011). Functional mixture regression. *Biostatistics*, **12**(2), 341–353.
Yao, F., Müller, H. & Wang, J. (2005). Functional data analysis for sparse longitudinal data. *JASA*, **100**, 577–590.
Yao, F. & Müller, H.-G. (2010). Functional quadratic regression. *Biometrika*, **97**(1), 49–64.
Zhang, T., Zhang, Q. & Wang, Q. (2014). Model detection for functional polynomial regression. *Comput. Stat. Data Anal.*, **70**, 183–197.
Zhang, X., King, M. L. & Shang, H. L. (2014). A sampling algorithm for bandwidth estimation in a nonparametric regression model with a flexible error density. *Comput. Stat. Data Anal.*, **78**, 218–234.
Zhao, Y., Chen, H. & Ogden, R. T. (2015). Wavelet-based weighted LASSO and screening approaches in functional linear regression. *J. Comp. Graph. Stat.*, **24**(3), 655–675.
Zhao, Y., Ogden, R. T. & Reiss, P. T. (2012). Wavelet-based LASSO in functional linear regression. *J. Comp. Graph. Stat.*, **21**(3), 600–617.
Zhou, H. & Li, L. (2014). Regularized matrix regression. *J. R. Stat. Soc. Series B Stat. Methodol.*, **76**(2), 463–483.
Zhou, H., Li, L. & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *JASA*, **108**, 540–552.
Zhu, H., Vannucci, M. & Cox, D. D. (2010). A Bayesian hierarchical model for classification with selection of functional predictors. *Biometrics*, **66**(2), 463–473.
Zhu, H., Yao, F. & Zhang, H. H. (2014). Structured functional additive regression in reproducing kernel Hilbert spaces. *J. R. Stat. Soc. Series B Stat. Methodol.*, **76**(3), 581–603.