

# Restricted Likelihood Ratio Tests for Linearity in Scalar-on-Function Regression

Mathew W. McLean\*    Giles Hooker<sup>‡</sup>    David Ruppert <sup>‡</sup>

October 23, 2013

## Abstract

We propose a procedure for testing the linearity of a scalar-on-function regression relationship. To do so, we use the functional generalized additive model (FGAM), a recently developed extension of the functional linear model. For a functional covariate  $X(t)$ , the FGAM models the mean response as the integral with respect to  $t$  of  $F\{X(t), t\}$  where  $F(\cdot, \cdot)$  is an unknown bivariate function. The FGAM can be viewed as the natural functional extension of generalized additive models. We show how the functional linear model can be represented as a simple mixed model nested within the FGAM. Using this representation, we then consider restricted likelihood ratio tests for zero variance components in mixed models to test the null hypothesis that the functional linear model holds. The methods are general and can also be applied to testing for interactions in a multivariate additive model or for testing for no effect in the functional linear model. The performance of the proposed tests is assessed on simulated data and in an application to measuring diesel truck emissions, where strong evidence of nonlinearities in the relationship between the functional predictor and the response are found.

**Keywords:** Functional data analysis, Functional regression, Generalized additive model, P-spline, Restricted likelihood ratio test, P-spline analysis of variance, Tensor product smooth

---

\*Research Assistant Professor, Institute for Applied Mathematics and Computational Science, Texas A&M University, College Station, TX, 77843 (E-mail: mmclean@stat.tamu.edu)

<sup>‡</sup>Associate Professor, Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, 14853, USA (E-mail: giles.hooker@cornell.edu)

<sup>‡</sup>Andrew Schultz Jr. Professor of Engineering and Professor of Statistical Science, School of Operations Research and Information Engineering and Department of Statistical Science, Cornell University, 1170 Comstock Hall, Ithaca, NY 14853, USA (E-mail: dr24@cornell.edu)

# 1 Introduction

The purpose of this paper is to propose and compare formal tests for when a scalar-on-function regression is well-modelled by the functional linear model (FLM). The FLM is by far the most popular model for scalar-on-function regression with a number of applied and theoretical developments available in the literature; an overview is provided by Ramsay and Silverman (2005). We suppose one has data  $\{(X_i(t), Y_i) : t \in \mathcal{T}\}$  for  $i = 1, \dots, N$ , where  $X_i$  is a real-valued, continuous, square-integrable, random function on the compact interval  $\mathcal{T}$  and  $Y_i$  is a scalar. We make the common assumption that the predictor,  $X(\cdot)$ , is observed at a dense grid of points. The FLM assumes the mean effect of  $X$  on  $Y$  is linear for each fixed  $t$  as follows

$$Y_i = \beta_0 + \int_{\mathcal{T}} \beta(t) X_i(t) dt + \epsilon_i, \quad (1)$$

where  $\beta(\cdot)$  is the functional coefficient with  $\beta(t)$  describing the effect on the response of the functional predictor at time  $t$  and  $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ .

McLean et al. (2013a) proposed an extension of the FLM called the functional generalized additive model (FGAM), which assumes the following form for the response

$$Y_i = \beta_0 + \int_{\mathcal{T}} F\{X_i(t), t\} dt + \epsilon_i, \quad (2)$$

where  $F(\cdot, \cdot)$  is an unknown bivariate function that is assumed to be smooth. The FLM can be seen as a special case of FGAM where  $F(x, t) = \beta(t)x$ . An intuitive way to view (2), is to consider it as the limit as the number of predictors (i.e. the number of observation times,  $t_j$ ) goes to infinity in a (multivariate) additive model with component functions  $f_j\{X_i(t_j)\} \Delta t_j := F\{X_i(t_j), t_j\}$ . Each additive component can be seen as a term in the Riemann sum approximation to the integral in (2) (see, McLean et al. 2013a). Some asymptotic theory for different estimators of (2) is available in Wang and Ruppert (2013) and Müller et al. (2013). An extension to the case of sparsely observed functional predictors is also available (McLean et al. 2013b).

We briefly review some other nonlinear scalar-on-function regression models that have been proposed in the literature. One of the earliest such proposals, and perhaps the most general, is the nonparametric kernel estimator of Ferraty and Vieu (2006a). Several authors have considered additive models that use linear functionals of the predictor curves as covariates, e.g.

$$E(Y_i | X_i) = \beta_0 + f\{\langle \beta(t) X_i(t) \rangle\} = \beta_0 + f\left\{\int \beta(t) X_i(t) dt\right\},$$

for unknown  $\beta_0$ ,  $f(\cdot)$ , and  $\beta(t)$ . Two such examples are Müller and Yao (2008) and James and Silverman (2005). The former approach regresses on a finite number of functional principal components scores and the latter approach searches for linear functionals using projection pursuit. For ease of interpretation, FGAM may be preferred to these models as it incorporates the functional predictor directly in the conditional mean; a model that is additive in the principal component scores is not additive in  $X(t)$  itself, and vice versa. While these models and FGAM each have the FLM as a special case, each is capable of capturing true relationships for the conditional mean response that the others cannot, and they can thus be seen as complimentary. Single-index type models similar to James and Silverman (2005) are considered in Ait-Saïdi et al. (2008), Chen et al. (2011), and Febrero-Bande et al. (2013). Two other noteworthy extensions of the FLM are Guillas and Lai (2010), which examines the case when  $X$  is a bivariate function so that  $E(Y_i | X_i) = \beta_0 + \int \int \beta(s, t) X(s, t) ds dt$  and Li et al. (2010), which allows for interaction between a scalar and functional covariate through a single index.

This paper will focus on testing

$$H_0 : E(Y_i | X_i) = \theta_0 + \int_{\mathcal{T}} \beta(t) X_i(t) dt \quad (\text{FLM})$$

vs.

$$H_1 : E(Y_i | X_i) = \theta_0 + \int_{\mathcal{T}} F(X_i(t), t) dt \quad (\text{FGAM}).$$

To accomplish this, we will formulate FGAM as a mixed model using a representation proposed for penalized splines in Wood et al. (2013), obtaining a representation of FGAM as a mixed model with three pairwise independent vectors of random effects. Mixed model representations for penalized splines are popular for their conceptual simplicity and because they can be estimated using software and algorithms available for mixed models (see, e.g., Ruppert et al. 2003). This parameterization makes the nesting of the FLM within FGAM explicit, and reduces our problem to one of testing for zero variance components in a mixed model. To conduct the test for zero variance components, we use restricted likelihood ratio tests (RLRTs) for which the exact finite-sample distribution is known for the one variance component case (Crainiceanu and Ruppert 2004).

Though the papers mentioned above all propose nonlinear models for scalar-on-function regression, we are only aware of one other work (García-Portugués et al. 2013) that considers a formal test of the FLM being true under the null hypothesis. To date almost all hypothesis tests proposed for scalar-on-function regression are concerned

with testing for no effect in the FLM (i.e.  $H_0 : \beta \equiv 0$  in (1)). Examples include Cardot et al. (2003), Swihart et al. (2013), Gabrys et al. (2010), Kong et al. (2013), and Zhang and Chen (2007). Swihart et al. (2013) is notable for taking a similar approach to the one we consider here; they use a penalized spline-mixed model for the FLM and use an RLRT for a zero variance component to test their desired hypothesis. Other hypothesis tests available for scalar-on-function regression include a test for equality of two coefficient functions estimated from independent data sets and a test for no quadratic effect in the functional quadratic model (Horváth and Kokoszka 2012, Ch. 10 and 12, resp.).

García-Portugués et al. (2013) proposes a Cramér-von Mises statistic for testing the null hypothesis that the FLM is the true model and uses a wild bootstrap to approximate the null distribution of the statistic. The method estimates the coefficient function in the FLM using a finite number of basis functions without penalization and has an unspecified alternative hypothesis. In contrast, we use *penalized* splines which allow for greater flexibility and specify an alternative hypothesis that the FGAM is the true model. We will compare our RLRT with the test used by these authors in our numerical studies.

The remainder of the paper proceeds as follows. In Section 2 we discuss mixed model representations for FGAM. Section 3 presents our proposed restricted likelihood ratio tests for linearity of FGAM. Section 4 provides a simulation study of our proposed approaches, and Section 5 concludes with an application of our methods to some motor vehicle emissions data.

## 2 Mixed Models Representations for FGAM

### 2.1 Notation and Basic Mixed Model Parameterization

In this section we introduce notation and briefly discuss a simple mixed model representation for FGAM in order to motivate the alternative mixed model representation that will be the focus of the paper. The linear mixed model we present here uses the parameterization of the bivariate surface in (2) that was used by McLean et al. (2013a), which we now describe. For both the  $x$  and  $t$  axes, we define cubic B-spline bases which we denote by  $\{B_j^X(x) : j = 1, \dots, K_x\}$  and  $\{B_k^T(t) : k = 1, \dots, K_t\}$ , respectively. For simplicity, the knots for the bases are chosen to be equally-spaced distances apart. The surface in (2) becomes

$$F(x, t) = \sum_{j=1}^{K_x} \sum_{k=1}^{K_t} B_j^X(x) B_k^T(t) \theta_{jk},$$

where the  $\theta_{jk}$ 's are unknown tensor-product B-spline coefficients.

Let  $\mathbb{B}_x$  denote the  $NJ \times K_x$  matrix of the  $x$ -axis B-splines evaluated at  $\text{vec}(\mathbf{X})$ , where  $\mathbf{X}$  is the  $N \times J$  matrix of observed functional predictor values ( $N$  curves measured  $J$  times each). Similarly, define  $\mathbb{B}_t$  to be the  $NJ \times K_t$  matrix of  $t$ -axis B-splines evaluated at  $\text{vec}(\mathbf{T})$ , where  $\mathbf{T}$  is the  $N \times J$  matrix of observation times for the functional predictor. In the densely-observed functional predictor setting assumed for this work, each  $X(t)$  will be pre-smoothed, and the statistician is free to choose  $\mathbf{T}$  and  $J$  as he/she chooses. We use a second-order penalty for both the  $x$  and  $t$  axes,  $\int \{\partial_{xx} F(x, t)\}^2 dx$  and  $\int \{\partial_{tt} F(x, t)\}^2 dt$ , respectively. To form the tensor product surface, we use the box product, also known as the row-wise Kronecker product, which is defined for two matrices  $\mathbb{A}_1$  and  $\mathbb{A}_2$  of dimension  $n \times m_1$  and  $n \times m_2$ , respectively as  $\mathbb{A}_1 \square \mathbb{A}_2 := \mathbb{A}_1 \otimes \mathbf{1}_{m_1}^T \odot \mathbf{1}_{m_2}^T \otimes \mathbb{A}_2$ , where  $\otimes$  represents the Kronecker product and  $\odot$  denotes element-wise matrix multiplication. We must approximate the integral in (2) using numerical quadrature. For this we define an  $N \times NJ$  matrix  $\mathbb{L}$  of quadrature weights. For example, if we simply use the midpoint rule,  $\mathbb{L} = J^{-1}(\mathbb{I}_N \otimes \mathbf{1}_J^T)$ . The penalized log-likelihood for FGAM is then

$$-\frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{1}\beta_0 - \mathbb{L}\mathbb{B}_x \square \mathbb{B}_t \boldsymbol{\theta}\|^2 - \boldsymbol{\theta}^T \mathbb{S}(\lambda_x, \lambda_t) \boldsymbol{\theta} \quad (3)$$

where  $\mathbf{y} = (y_1 \dots, y_N)^T$ ,  $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1K_t}, \theta_{21}, \dots, \theta_{K_x K_t})^T$ , and  $\mathbb{S}(\lambda_x, \lambda_t)$  is the penalty matrix for the tensor product smooth (e.g., Wood 2006; Marx and Eilers 2005).

McLean et al. (2013a) used generalized cross validation (GCV) to select the smoothing parameters  $\lambda_x$  and  $\lambda_t$ , which is equivalent to placing an improper Gaussian prior on the spline coefficients. To fit FGAM using mixed modelling software, we require a proper Gaussian prior for the random effects. This can be achieved by separating the model into a component in the null space of the penalty term for FGAM and a component in the orthogonal complement of the null space, which we refer to as the range space for consistency with Wood et al. (2013). One easy way to do this is using an eigendecomposition of the penalty matrix.

We decompose  $\mathbb{S}(\lambda_x, \lambda_t)$  as  $\mathbb{S}(\lambda_x, \lambda_t) = \mathbb{U}\mathbb{D}\mathbb{U}^T$ , where  $\mathbb{U}$  is orthogonal and  $\mathbb{D}$  is diagonal. Owing to the second-order penalties,  $\mathbb{S}(\lambda_x, \lambda_t)$  will have four zero-eigenvalues. We split  $\mathbb{U}$  into  $\mathbb{U} = [\mathbb{U}_p : \mathbb{U}_n]$ , where  $\mathbb{U}_n$  contains the columns of  $\mathbb{U}$  corresponding to the zero eigenvalues of the penalty matrix and  $\mathbb{U}_p$  consists of the remaining columns. We write  $\mathbb{D}_+$  to denote the diagonal submatrix of  $\mathbb{D}$  containing only the positive eigenvalues of  $\mathbb{S}(\lambda_x, \lambda_t)$ . The columns of  $\mathbb{U}_p$  form a basis for the

range space of the FGAM penalty. FGAM can then be reparametrized as follows

$$\begin{pmatrix} \mathbf{b} \\ \boldsymbol{\beta} \end{pmatrix} := \mathbf{U}^T \boldsymbol{\theta}, \quad \mathbb{X} := \mathbb{L}\mathbb{B}_x \square \mathbb{B}_t \mathbf{U}_n, \quad \mathbb{Z} := \mathbb{L}\mathbb{B}_x \square \mathbb{B}_t \mathbf{U}_p,$$

and we then have the mixed model

$$\begin{aligned} \mathbf{y} &= \mathbf{1}\beta_0 + \mathbb{X}\boldsymbol{\beta} + \mathbb{Z}\mathbf{b} + \boldsymbol{\epsilon}, \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbb{I}_N), \quad \mathbf{b} \sim N(\mathbf{0}, [\mathbf{U}_p^T \mathbb{S}(\lambda_x, \lambda_t) \mathbf{U}_p]^{-1}). \end{aligned} \quad (4)$$

We can further reparametrize (4) in terms of variance components  $\sigma_x^2 := \sigma^2/\lambda_x$  and  $\sigma_t^2 = \sigma^2/\lambda_t$  to coincide with the mixed model literature.

This model can be fit via maximum likelihood (ML) or restricted maximum likelihood (REML) using some mixed modelling software such as the `nlme` package in R (Pinheiro et al. 2013), but not by other popular software such as SAS (SAS Institute Inc. 2008) or R package `lme4` (Bates et al. 2013) because of the complicated covariance structure for the random effect. There is some evidence to suggest that REML should be preferred to methods that minimize prediction error, such as GCV or Akaike's information criteria (AIC) and its siblings for estimating smoothing parameters in penalized spline models because it avoids their occasional tendencies to badly undersmooth and offers slightly better RMSE performance in practise (Reiss and Ogden 2009). However, for the purposes of this work, we need a model with a simpler covariance structure that makes the nesting of the FLM within FGAM explicit. This will be the subject of the next section.

## 2.2 Penalized Spline ANOVA Parameterization

In this section we show how FGAM may be parameterized using the tensor product spline basis construction of Wood et al. (2013) which parallels smoothing spline ANOVA (for e.g., Wang 2011), the main difference being the use of low-rank spline bases. The full model will be projected onto a tensor sum of orthogonal subspaces, with each component in the new construction either unpenalized or with its own unique penalty that is interpretable in terms of the original model component functions. This is convenient because it will lead to a mixed model representation where each random effect has a diagonal covariance matrix independent of the other effects.

For this construction, we begin with eigendecompositions of marginal penalty matrices,  $\mathbb{P}_x$  and  $\mathbb{P}_t$ . For a second-order penalty,  $\mathbb{P}_x$  has rank  $K_x - 2$  and  $(n, m)$ -entry

$$(P_x)_{m,n} = \int \partial_{xx} \{B_m^{(X)}(x)\} \partial_{xx} \{B_n^{(X)}(x)\} dx; \quad m, n = 1, \dots, K_x.$$

We express  $\mathbb{P}_x$  as  $\mathbb{P}_x = \mathbb{U}_x \mathbb{D}_x \mathbb{U}_x^T$ , where  $\mathbb{U}_x$  is orthogonal and  $\mathbb{D}_x$  is diagonal with two zeros on the diagonal. As before, we let  $\mathbb{U}_{n,x}$  be the columns of  $\mathbb{U}_x$  corresponding to the zero eigenvalues and  $\mathbb{U}_{p,x}$  be the remaining columns and let  $\mathbb{D}_{+,x}$  be the diagonal matrix containing all positive eigenvalues of  $\mathbb{P}_x$ . We then form  $\mathbb{Z}_x = \mathbb{B}_x \mathbb{U}_{p,x} \mathbb{D}_{+,x}^{-1/2}$  and  $\mathbb{X}_x = \mathbb{B}_x \mathbb{U}_{n,x}$ . The matrix  $\mathbb{Z}_x$  forms a basis for the random effects of a univariate smooth in  $x$  (i.e., a basis for the range space of the marginal penalty for  $x$ ) and  $\mathbb{X}_x$  forms a basis for the fixed effects of the smooth. The marginal penalty matrix will become the identity matrix of appropriate dimension except with its last two diagonal entries equal to zero (see Wood et al. 2013). We form  $\mathbb{X}_t$  and  $\mathbb{Z}_t$  analogously using  $\mathbb{B}_t$  and a corresponding penalty matrix  $\mathbb{P}_t$ .

To obtain our tensor product construction, we form all pairwise box products of elements of  $\{\mathbb{X}_x, \mathbb{Z}_x\}$  with elements of  $\{\mathbb{X}_t, \mathbb{Z}_t\}$ . Our design matrix for the tensor product smooth becomes

$$\mathcal{M} = [\mathbb{X}_x \square \mathbb{X}_t : \mathbb{X}_x \square \mathbb{Z}_t : \mathbb{Z}_x \square \mathbb{X}_t : \mathbb{Z}_x \square \mathbb{Z}_t].$$

The term  $\mathbb{X}_x \square \mathbb{X}_t$  corresponds to the unpenalized, fixed effects part of the smooth, and the three other terms are bases for the random effects with each component having a separate ridge penalty. Let  $\mathbf{x} = \text{vec}(\mathbf{X})$  and  $\mathbf{t} = \text{vec}(\mathbf{T})$ ; we can re-parameterize the null space bases as  $\mathbb{X}_x = [\mathbf{1} : \mathbf{x}]$ ,  $\mathbb{X}_t = [\mathbf{1} : \mathbf{t}]$ , and  $\mathbb{X}_x \square \mathbb{X}_t = [\mathbf{1} : \mathbf{x} : \mathbf{t} : \mathbf{x} \odot \mathbf{t}]$ . The function  $F(x, t)$  is decomposed into an unpenalized, parametric part and three separate nonparametric parts each subject to different, non-overlapping penalties

$$\begin{aligned} \text{term} : \quad & \underbrace{F(x, t)}_{\text{penalty}} = \underbrace{\beta_0 + \beta_1 x + \beta_2 t + \beta_3 x t}_{\text{unpenalized}} \\ & \lambda_t \int (\partial_{tt} F)^2 + \lambda_x \int (\partial_{xx} F)^2 \\ + \quad & \underbrace{f_1(t) + x f_2(t)}_{\lambda_1 \int [(\partial_{tt} f_1)^2 + (\partial_{tt} f_2)^2]} + \underbrace{g_1(x) + t g_2(x)}_{\lambda_2 \int [(\partial_{xx} g_1)^2 + (\partial_{xx} g_2)^2]} + \underbrace{h(x, t)}_{\lambda_3 \int (\partial_{xxtt} h)^2}. \end{aligned} \quad (5)$$

If the terms  $\beta_2 t$  and  $f_1(t)$  in (5) are integrated w.r.t.  $t$ , they become confounded with the intercept,  $\beta_0$ . Hence, for identifiability  $\mathbf{t}$  and  $\mathbb{Z}_t (= \mathbf{1} \square \mathbb{Z}_t)$  are dropped from  $\mathcal{M}$ . This construction is summarized in Table 1. An alternative approach for ensuring identifiability is discussed in McLean et al. (2013a). If we define

$$\mathbb{X} = [\mathbf{1} : \mathbf{x} : \mathbf{x} \otimes \mathbf{t}], \quad \mathbb{Z}_1 = \mathbf{x} \square \mathbb{Z}_t, \quad \mathbb{Z}_2 = \mathbb{Z}_x \square \mathbb{X}_t, \quad \mathbb{Z}_3 = \mathbb{Z}_x \square \mathbb{Z}_t,$$

and let  $\mathcal{L}(\mathbf{F})$  denote the  $N$ -vector with  $i$ th entry

Table 1: Description of penalized components of the tensor production construction (5) with  $f_1(t)$  removed for identifiability

Term	Basis for functions of the form	Penalty
$\mathbb{X} = [\mathbf{1} : \mathbf{x} : \mathbf{x} \otimes \mathbf{t}]$	$\beta_0 + \beta_1 x + \beta_3 xt$	unpenalized
$\mathbb{Z}_1 = \mathbf{x} \square \mathbb{Z}_t$	$xf_2(t)$	$\int (\partial_{tt} f_1)^2 + (\partial_{tt} f_2)^2$
$\mathbb{Z}_2 = \mathbb{Z}_x \square \mathbb{X}_t$	$g_1(x) + tg_2(x)$	$\int (\partial_{xx} g_1)^2 + (\partial_{xx} g_2)^2$
$\mathbb{Z}_3 = \mathbb{Z}_x \square \mathbb{Z}_t$	$h(x, t)$ w/o above terms	$\int (\partial_{xxtt} h)^2$

$\int_{\mathcal{T}} F\{X_i(t), t\} dt$ , then we can write FGAM in mixed model form as

$$\mathbf{y} = \beta_0 + \mathcal{L}(\mathbf{F}) \approx \mathbb{L}\mathbb{X}\boldsymbol{\beta} + \sum_{j=1}^3 \mathbb{L}\mathbb{Z}_j \mathbf{b}_j + \boldsymbol{\epsilon}; \quad (6)$$

$$\mathbf{b}_j \sim N(\mathbf{0}, \sigma_j^2 \mathbb{I}_{q_j}), \quad j = 1, 2, 3;$$

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_e^2 \mathbb{I}_N),$$

with the dimension of the vectors of random effects being  $q_1 = K_t - 2$ ,  $q_2 = 2(K_x - 2)$ , and  $q_3 = (K_x - 2)(K_t - 2)$  for  $\mathbf{b}_1$ ,  $\mathbf{b}_2$ , and  $\mathbf{b}_3$ , respectively.

We now have a mixed model where each vector of random effects,  $\mathbf{b}_j$ , has a simple, diagonal covariance structure. This allows us to use `lme4`, which is optimized to work with sparse covariance matrices, or e.g., PROC MIXED in SAS. This is not possible for the mixed model (4). More importantly for the goal of this paper, it is explicitly clear how the FLM is nested in FGAM in this parameterization: Referencing Table 1, the integrand for the FLM,  $\beta(t)X(t)$ , can be seen to correspond to the first penalized term,  $xf_2(t)$ , and the unpenalized terms. Thus, in the mixed model (6), the FLM is represented by the fixed effects and the random effect  $\mathbf{b}_1$ . Therefore, a goodness of fit test for the FLM is equivalent to testing whether  $\sigma_2 = \sigma_3 = 0$ . We will refer to (6) as the FGAMM (for Functional Generalized Additive Mixed Model) later in our numerical experiments when we compare this parametrization with the one used in McLean et al. (2013a).



### 3 Tests for Linearity of Scalar-on-Function Regression

#### 3.1 Approximate Restricted Likelihood Ratio Tests

The restricted likelihood (Patterson and Thompson 1971) for model (6) is

$$\begin{aligned} \ell_R(\mathbf{y}) = & -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}_J| \\ & - \frac{(\mathbf{Y} - \mathbb{L}\mathbb{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_J^{-1} (\mathbf{Y} - \mathbb{L}\mathbb{X}\boldsymbol{\beta})}{2\sigma^2} - \frac{1}{2} |\mathbb{X}^T \mathbb{L}^T \boldsymbol{\Sigma}_J^{-1} \mathbb{L}\mathbb{X}|, \end{aligned}$$

where  $\boldsymbol{\Sigma}_J := \text{cov}(\mathbf{Y}) = \sigma^2 \mathbb{I}_N + \sum_{j=1}^3 \sigma_j^2 \mathbb{L} \mathbb{Z}_j \mathbb{Z}_j^T \mathbb{L}^T$ . We wish to test  $H_0 : \sigma_2 = \sigma_3 = 0$  using a restricted likelihood ratio test (RLRT). For the one variance component linear model  $\mathbf{y} = \mathbb{X}\boldsymbol{\beta} + \mathbb{Z}\mathbf{b}$  with  $\mathbf{b} \sim N(\mathbf{0}, \sigma_1^2 \mathbb{I}_{q_1})$ , we know the *exact* finite-sample distribution for the RLRT statistic (Crainiceanu and Ruppert 2004):

$$\begin{aligned} \text{RLRT} &= 2 \sup_{H_1} \ell_R(\mathbf{y}) - 2 \sup_{H_0} \ell_R(\mathbf{y}) \\ &= \sup_{\lambda} \left[ (N - q_0 - 1) \log\{1 + U_n(\lambda)\} - \sum_{k=1}^{q_1} \log(1 + \lambda\mu_k) \right]; \end{aligned} \quad (7)$$

where  $\lambda = \sigma_1^2/\sigma^2$  and  $U(\lambda) = N(\lambda)/D(\lambda)$  with

$$\begin{aligned} N(\lambda) &= \sum_{k=1}^{q_1} \frac{\lambda\mu_k}{1 + \lambda\mu_k} w_k^2, \\ D(\lambda) &= \sum_{k=1}^{q_1} \log(1 + \lambda\mu_k) + \sum_{k=q_1+1}^{N-q_0} w_k^2; \end{aligned}$$

for  $w_k \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ ;  $k = 1, \dots, N - q_0 - 1$ . The  $\mu_k$  are the eigenvalues of the matrix  $\mathbb{Z}^T (\mathbb{I}_N - \mathbb{X}(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T) \mathbb{Z}$ . This distribution may be simulated from very quickly. The eigendecomposition of the  $q_1 \times q_1$  matrix need only be computed once, and then all that is required to obtain a draw from the RLRT distribution is simulation of  $q_1$   $\chi_1^2$  random variables and one  $\chi_{N-q_0-q_1-1}^2$  random variable.

While the theory is fully developed for the one variance component case, extensions to tests for models with multiple variance components (which we will require for FGAM) have proven much harder, and this is still an open problem. An approach that has proven to work well empirically is that of Greven et al. (2008), which used ideas from pseudo-likelihood estimation and relied on the assumption that the restricted likelihood ratio tests (RLRT) for their variance components of interest could be accurately approximated

by an RLRT that assumes the nuisance random effects are known. Another possible approach has been proposed by Wang and Chen (2012), which developed F-tests for penalized spline models estimated in the mixed model framework. The Wald-type test developed by Wood (2013) is not considered because it was designed for testing whether a component of a smooth is identically zero, not for testing whether the component is in the null space of its penalty. We choose to work with the approach of Greven et al. (2008) because it has been shown through extensive simulation studies by Scheipl et al. (2008) to work well and because the method is available in an R package by the same authors. The simulations in Wang and Chen (2012) also confirmed the effectiveness of the Greven et al. (2008) approach, with their F-tests only offering minor improvements in the case where a nuisance variance component is very close to zero. We refer to the Greven et al. (2008) method as the pseudo-RLRT. A more computationally intensive approach, would be to approximate the null distribution of the RLRT statistic using a parametric bootstrap as in Pinheiro and Bates (2000).

An additional open question not addressed by the above papers is how to test for multiple variance components being simultaneously zero under the null hypothesis. This is the situation that we are faced with for FGAM and the subsequent sections will propose some ideas for how to deal with this problem.

### 3.1.1 Test $\sigma_2$ and $\sigma_3$ Separately

The first approximation we consider is to conduct two separate RLRTs for the two non-FLM variance components. In the first test, the random effect  $\mathbf{b}_3$  is fixed as the zero vector under both the null and alternative hypotheses,  $H_0 : \sigma_2 = 0, \sigma_3 = 0$  and  $H_1 : \sigma_2 > 0, \sigma_3 = 0$ . In the second test,  $\mathbf{b}_2 \equiv \mathbf{0}$  and  $\sigma_3$  is tested for equality with zero. Both tests still involve one nuisance variance component and so we use the approximate RLRT from Greven et al. (2008) for each test and then apply a Bonferroni correction to account for multiple testing. We refer to the approach as method “Bonferroni” in later sections.

### 3.1.2 Test assuming $\sigma_2 = \sigma_3$

The second approximation we consider is to assume that the amount of smoothness for each non-FLM component of the decomposition in (5) are equal, i.e. that  $\sigma_2 = \sigma_3$ . Using this assumption, we are again reduced to testing one variance component for equality to zero in the presence of one nuisance variance component, for which we use the method of Greven et al. (2008). However, unlike the previous section, we only need to conduct one test. We refer to the approach as method “EqualVC” in later sections.

### 3.2 A Test For No Effect In the Functional Linear Model

Before assessing whether an FLM or FGAM provides a better fit to the data, one will want to determine whether the functional predictor has any effect on the response at all. This is quite simple to test in our framework. By simply dropping the random effects  $\mathbf{b}_2$  and  $\mathbf{b}_3$ , we can test for no effect by considering  $H_0 : \beta_2 = \beta_3 = 0, \sigma_1 = 0$  versus  $H_1 : \beta_2 \neq 0$  or  $\beta_3 \neq 0$  or  $\sigma_1 > 0$  (FLM is true). The exact distribution of the LRT statistic for this test is known due to Crainiceanu and Ruppert (2004). Note that a restricted likelihood ratio test is inappropriate here because the fixed effects are different under the two hypotheses. One can also use either an LRT or RLRT to test  $H_0 : \sigma_1 = 0$  vs.  $H_1 : \sigma_1 > 0$  which is a test that the effect of  $X(t)$  is linear in  $t$ ;  $Y_i = \beta_0 + \mathbf{L}^T\{\mathbf{x}_i \odot (\beta_1 + \beta_2 \mathbf{t})\}$ . If one instead uses a first order penalty for  $x$  and  $t$ , then a test for no effect is equivalent to testing  $\sigma_1 = 0$ . This proposal is similar to one recently considered in Swihart et al. (2013) for the penalized functional regression model of Goldsmith et al. (2011). Those authors first perform a functional principal components analysis to estimate the predictor trajectories and then estimate the coefficient function in the FLM using penalized splines with a first-order difference penalty and different mixed model representation than the one considered here. It is also possible to test for a quadratic effect of the form  $\int \zeta(t)X^2(t)dt$  if one uses a third order penalty for the marginal basis for  $x$ .

## 4 Simulation Study

In this section we study the performance of our proposed tests for linearity of FGAM on simulated data for two different setups. First, in Section 4.1, we generate the response variable using a convex combination of an FLM and an FGAM in the functional predictor. This is done to assess the size of the departure from linearity that our tests can detect in a way that is interpretable in terms of the original models. In Section 4.2, we assess empirical Type I error rates and power for our tests by generating the response from the mixed model in Section 2.2 for several different values of the variance components and compare performance with tests that know the value of the nuisance parameters.

We also consider the recently proposed method of García-Portugués et al. (2013) mentioned in the introduction. This is the only work besides ours we are aware of that focuses on a null hypothesis of the FLM being true. Their method is implemented in the R package `fdm.usc` (Febrero-Bande and Oviedo de la Fuente 2012). To perform their test, their software first fits an FLM. In our simulation studies, specifying more than four basis functions for the functional coefficient produced singularities in the

model matrix that caused their software to fail. We speculate that this is due to the lack of regularization in the method. We therefore only report results for the four basis function case for this method, which we label “GPGMFB”.

For the methods that involve RLRTs, computations are done in R (R Core Team 2012) using the package `RLRsim` (Scheipl et al. 2008). The package requires fitted model objects for the model under both hypotheses, as well as a fit to the data with nuisance variance components equal to zero. These fits are obtained using the package `lme4` (Bates et al. 2013).

#### 4.1 True Model as Convex Combination of FLM and FGAM

Here, we fit each model to 500 simulated data sets. The true functional covariates are given by  $X(t) = \sum_{j=1}^4 \xi_j \phi_j(t)$ , with  $\xi_j \sim N(0, 8j^{-2})$  and  $\{\phi_1(t), \dots, \phi_4(t)\} = \{\sin(\pi t), \cos(\pi t), \sin(2\pi t), \cos(2\pi t)\}$ . Each functional predictor was observed at 30 equally-space points. To generate the response, we take a convex combination of a bivariate function linear in  $x$  and one nonlinear in  $x$ ,

$$F_1(x, t) = 2x \sin(\pi t), \text{ and } F_2(x, t) = 10 \cos\left(-\frac{x}{8} + \frac{t}{4} - 5\right),$$

with  $t = [0, 1]$ . The response is given by

$$Y_i = \int_0^1 [\phi F_1\{X_i(t), t\} + (1 - \phi) F_2\{X_i(t), t\}] dt + \epsilon_i,$$

with  $\epsilon_i \sim N(0, 1)$  and  $0 \leq \phi \leq 1$ . The constants in  $F_1$  and  $F_2$  were chosen so that each surface contributed roughly equally to the signal for each generated data set prior to multiplication by  $\phi$ . Notice that when  $\phi = 1$ , the null hypothesis that the true model is an FLM is true. Both true surfaces along with some generated functional predictors are shown in Figure 1.

We consider two sample sizes,  $N = 100, 500$ . We used ten basis functions for each axis when fitting the FGAM. Results for other numbers of basis functions were similar, unless the number of basis functions is made too small. This is because low-dimensional random effects are difficult for mixed model software to estimate. In `lme4`, having low-dimensional random effects results in an increased number of zero estimates for the variance components corresponding to the low-dimensional random effects. The results for this simulation study are summarized in Figure 2 where for each of our proposed tests we plot the proportion of the 500 simulations where the null hypothesis is rejected by  $(1 - \phi)$ . We use  $1 - \phi$  so that zero on the x-axis corresponds to the null model (FLM) being true. For brevity, we report results for significance level,  $\alpha = 0.05$  only.

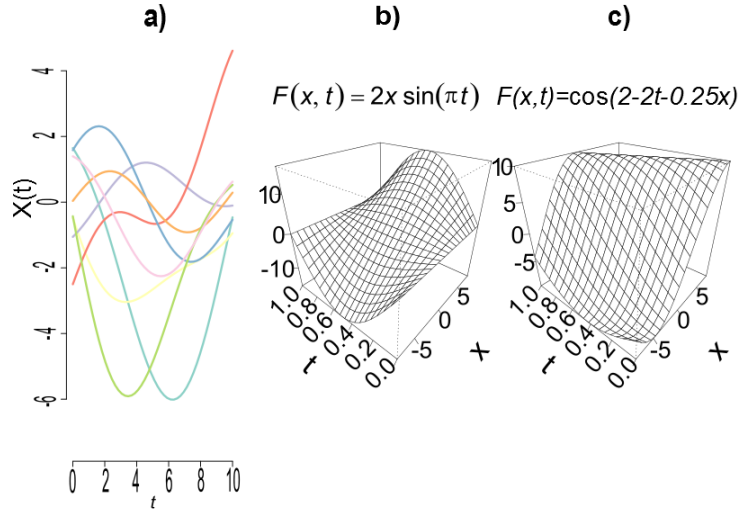


Figure 1: a) Sample of 8 simulated predictor curves,  $X(t)$ . b) True surface,  $F_1(x, t)$ , corresponding to the FLM. c) True surface  $F_2(x, t)$ , corresponding to FGAM

The EqualVC method is able to achieve a type I error rate fairly close to the nominal level and also has the highest power of any of the methods. Method Bonferroni is seen to be conservative, as expected. GPGMFB has lower power and an observed type I error rate slightly further from the nominal level than method EqualVC, but is less conservative than Bonferroni.

## 4.2 True Model as Mixed Model From Section 2.2

We now change how the response is generated so that it comes from the mixed model (6). We will explore the trade-off between  $\sigma_2^2$  and  $\sigma_3^2$  values to assess power and to explore whether the assumption of EqualVC that  $\sigma_2 = \sigma_3$  can be problematic. The functional predictors are generated in the same manner as the previous section. For a given simulated sample of  $N$  curves, the response is formed as follows. First, the Section 2.2 parameterization is used to form the bases  $\mathbb{X}$ ,  $\mathbb{Z}_1$ ,  $\mathbb{Z}_2$ , and  $\mathbb{Z}_3$ . Next, the random effect vector for the nuisance variance component corresponding to the FLM term (see Table 1) in the construction is drawn as  $\mathbf{b}_1 \sim N(\mathbf{0}, 4\mathbb{I}_{q_1})$  and the two random effects vectors corresponding to non-FLM terms are drawn as  $\mathbf{b}_j \sim N(\mathbf{0}, \sigma_j^2 \mathbb{I}_{q_j})$ ;  $j = 2, 3$ .

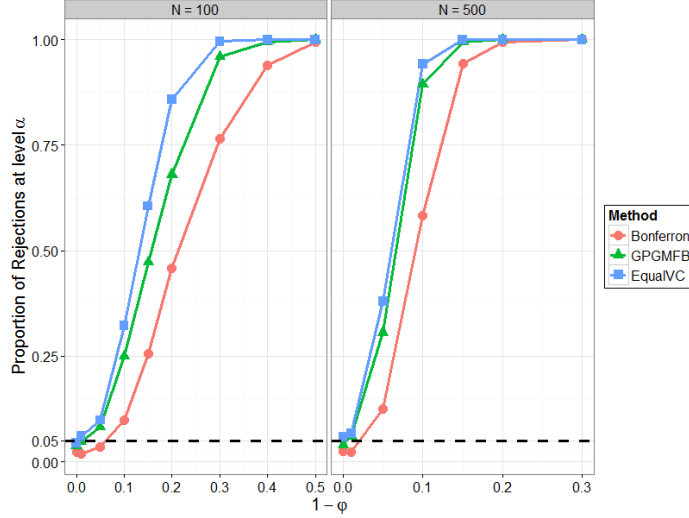


Figure 2: Proportion of rejected null hypotheses for each testing method under consideration in Section 4.1

The response is then given by

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sum_{j=1}^3 \mathbb{Z}\mathbf{b}_j + \boldsymbol{\epsilon},$$

where  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbb{I}_N)$  and  $\boldsymbol{\beta} = (1, 0.01, 0.01)^T$ .

In this section we again consider Bonferroni and EqualVC, and to better assess the performance of method Bonferroni, we consider a “quasi-oracle” test that knows the true value of the nuisance random effects vector for each simulation. In more detail, the pseudo-residuals used as inputs to the pseudo-RLRTs for this method are  $\mathbf{Y} - \mathbb{Z}_1\mathbf{b}_1$ , for the true  $\mathbf{b}_1$  instead of its prediction using REML. The method still tests  $\sigma_2$  and  $\sigma_3$  separately and uses a Bonferroni correction, but with no nuisance variance component and only one variance component to test at a time, we are in exactly the framework of Crainiceanu and Ruppert (2004), where the distribution of the test statistic is known and easily simulated from. We label this method “KnownSig1”. Note also that changing the number of basis functions with this data generation scheme changes the dimension of the random effects in the true model. The values of  $\sigma_j^2$  for  $j = 2, 3$ ; considered are  $\sigma_j^2 = (0, 0.04, 0.1, 0.25, 0.5, 0.75)$  for  $N = 100$  and  $\sigma_j^2 = (0, 0.004, 0.04, 0.14, 0.2, 0.3)$  when  $N = 500$ .

As in the previous section, we generate 500 data sets for each simulation setting,

use ten basis functions for each axis when fitting FGAM, and report the proportion of times each method rejects the null hypothesis that the FLM is the true model. The empirical power of the proposed tests for significance level  $\alpha = 0.05$  is plotted in Figure 3. Each panel corresponds to a different value of  $\sigma_2^2$ ; for simplicity we report only three values of  $\sigma_2^2$  for each value of  $N$ . The results for the other values of  $\sigma_2^2$  are similar.

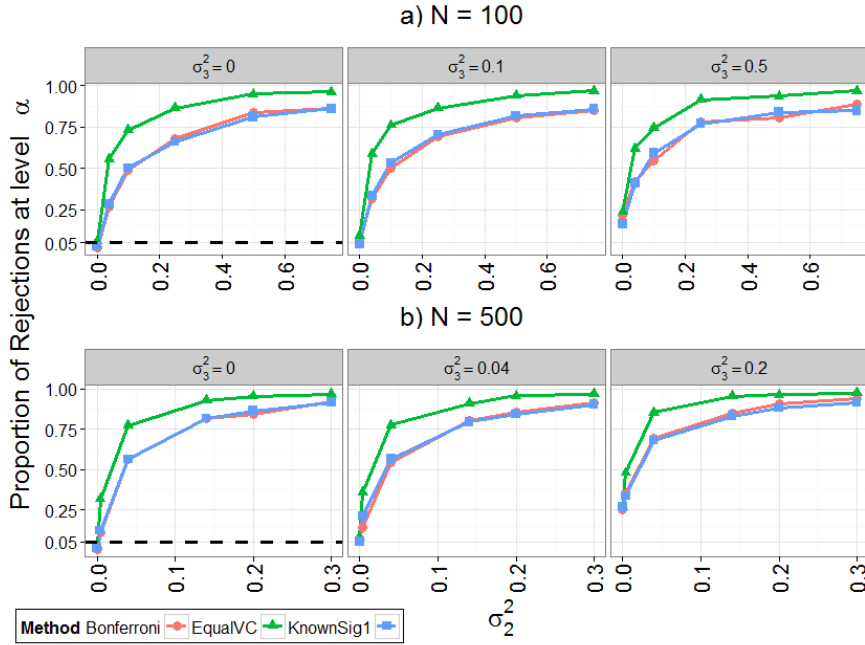


Figure 3: Proportion of rejected null hypotheses at  $\alpha = 0.05$  over 500 simulations for a grid of several values of  $\sigma_2^2$  and  $\sigma_3^2$

The method EqualVC is again the most powerful for this study. We see similar levels of disparity between EqualVC and Bonferroni as in the simulation study of the previous section. It is promising that method EqualVC outperforms the method that knows the nuisance random effect. We can also see that knowing the nuisance random effect has little effect on the power of the test as Bonferroni performs nearly identically to KnownSig1.

It turns out that it is much easier to detect  $\sigma_2$  being non-zero when  $\sigma_3$  is small or zero than vice versa; for example, for  $N = 100$  EqualVC rejects  $H_0$  in 93.6% of the simulations when  $(\sigma_2^2, \sigma_3^2) = (0.5, 0)$ , but in only 23.4% of simulations when  $(\sigma_2^2, \sigma_3^2) = (0, 0.5)$ . Note that the relation between the dimension of  $\mathbf{b}_2$  and  $\mathbf{b}_3$  in

this setup with  $K_x = K_t$  and second-order penalties is  $q_3 = (q_2/2)^2$ . We also see that method EqualVC does not seem to lose its advantage in power over the other methods when one variance component is non-zero while the other is zero.

The type I error rates varied little as either the sample size or number of basis functions changed. Averaging over the two samples sizes, for  $\alpha = 0.05$ , the empirical type I error rate was 0.013 for Bonferroni, 0.048 for EqualVC, and 0.019 for KnownSig1. Given how close its rate is to the nominal level and its strong power performance compared to the other methods in both simulation sections, we recommend using the EqualVC method which assumes a priori  $\sigma_2 = \sigma_3$  and then conducts a single pseudo-RLRT using the Greven et al. (2008) approach.

To further assess the adequacy of the EqualVC method, we also considered using the parametric bootstrap as it is implemented in `lme4` (Bates et al. 2013) for a subset of the  $(\sigma_2^2, \sigma_3^2)$  pairs. We found that using the bootstrap produced similar results to the EqualVC method, with only negligible differences between the two methods in both empirical Type I error and power. This is similar to the results found in Scheipl et al. (2008). We take this as further support that the EqualVC approach is sound, and prefer using it to the bootstrap due to its significantly reduced computational overhead.

## 5 Analysis of Emissions Data

In this section we apply our proposed procedures to study the quantities of various pollutants in truck exhaust emissions. The data come from chassis dynamometer emissions readings from the Coordinating Research Council E55/59 emissions inventory program (Clark et al. 2007). The goal of the study was to assess particulate matter emissions in heavy-duty trucks in California. Vehicles were tested in a lab setting designed to mimic everyday driving conditions. Particulate matter was captured using 70 mm filters on the dilute exhaust. Each vehicle in the study was repeatedly driven in four different driving cycles for an extend period. For example, one driving cycle might be cruising at highway speeds and the next stop-and-go city driving. For our application, we consider a simplified problem, attempting to model the emissions of one truck from the study. We attempt to predict the logarithm of particulate matter every twenty seconds using the truck's recorded speed and/or acceleration in the immediately preceding forty seconds. This resulted in 157 samples. We chose to sample particulate matter every 20 seconds to ensure there was no temporal dependence between response samples. Figure 4 plots both the original speed data and estimated accelerations for all blocks in the data grouped according to driving cycle.

In the subsection that follows, we analyze the fit of the FLM to these data using



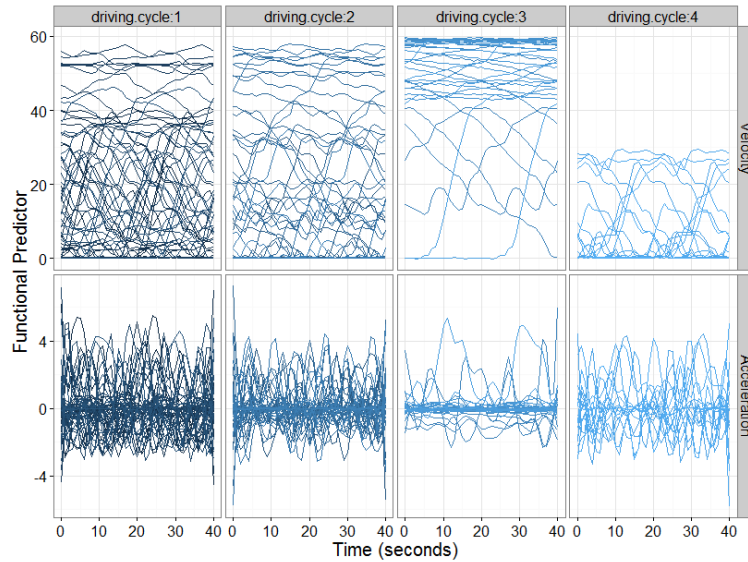


Figure 4: Speed and acceleration trajectories in forty seconds blocks grouped according to speed pattern (driving cycle)

our proposed tests for linearity. After that, we compare FLM and FGAM out-of-sample predictive performance for this data set and also compare predictive performance of the parameterization discussed in Section 2.2 with the parameterization used in McLean et al. (2013a).

## 5.1 FLM Fit Assessment

Leaving the finer details of our fitting procedures to the next section, we now discuss some diagnostics for assessing the fit of both the FLM and the FGAM including use of our proposed testing procedures. We consider predicting particulate matter using vehicle acceleration as the functional predictor and also include a categorical covariate for the driving cycle. Some residual plots for an FLM fit to the entire data set using tuning parameters that had been chosen to optimize performance for the next section are given in Figure 5. The top row of plots shows the residuals grouped according to the driving cycle covariate and the bottom shows the residuals plotted against the predicted value and also a normal Q-Q plot of the residuals. We can see evidence of a nonlinear association between the residuals and the response and also that the variance of the residuals is not constant across the driving cycle factor. The Q-Q plot

indicates non-normality of the residuals. To assess this more formally, we consider

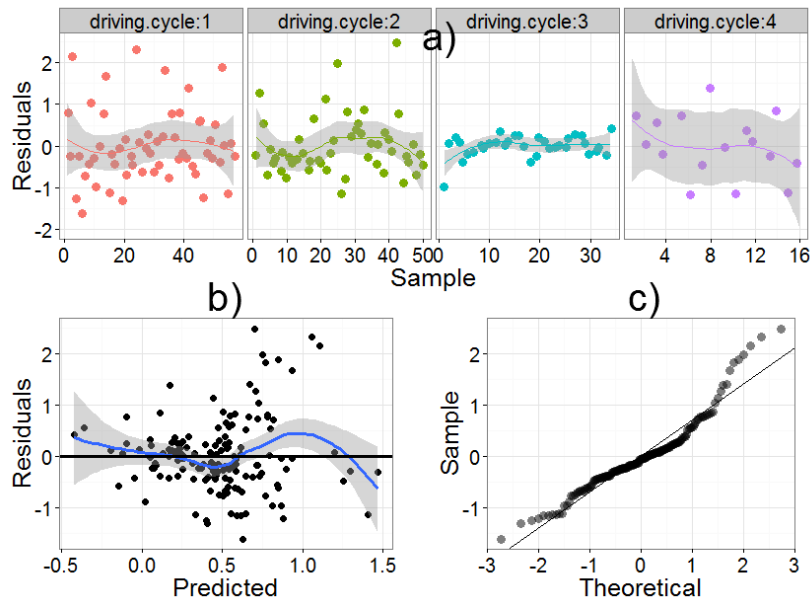


Figure 5: Diagnostic plots for an FLM fit with truck acceleration as the functional predictor: a) plots the residuals grouped by driving cycle, b) plots residuals vs. predicted value, and c) is the normal Q-Q plot

the proposed tests of Section 3. Using the EqualVC method from Section 3.1.2, we obtain a p-value  $\approx 0$ . This very strongly suggests the FLM is not adequate here. We also obtain a p-value that is zero to machine precision using the Bonferroni method from Section 3.1.1. The results remain overwhelming regardless of the number of basis functions used.

The residual plots for an FGAM fit to the data using the basis construction from this chapter can be seen in Figure 6. We can see that the magnitude of the residuals has gone down and that all three plots seem to be less in violation of the model assumptions than the FLM fit. The variance of the residuals also appears to be more constant across driving cycles.

Figure 7 shows contours of the estimated surface obtained by using all 157 samples and the acceleration curves as predictors. The surface was estimated using  $1m\epsilon 4$  (Bates et al. 2013). Also plotted are the individual components of the basis construction of Section 2.2; the unpenalized component, along with the three penalized components (see Table 1). The marginal bases for the  $x$  and  $t$  axes were both of dimension eight.

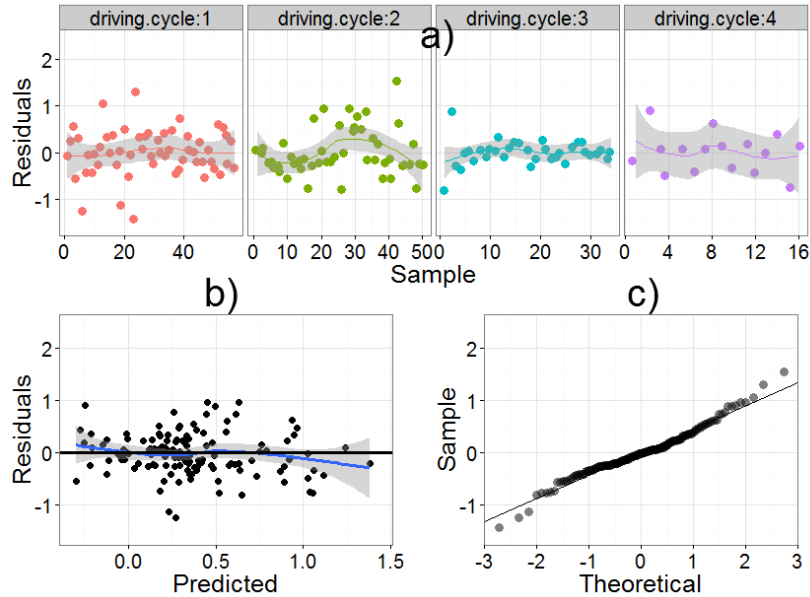


Figure 6: Diagnostic plots for an FGAMM fit with truck acceleration as the functional predictor: a) plots the residuals grouped by driving cycle, b) plots residuals vs. predicted value, and c) is the normal Q-Q plot

Interestingly, the variance component for the FLM portion of the fit was estimated to be very close to zero in this case.

The FLM fails only marginally better if the truck speeds are used as the functional predictor. We omit the diagnostic measures, but the out-of-sample prediction performance using either covariate or both is examined in the next section.

## 5.2 Out-of-Sample Prediction of Particulate Matter

As further confirmation that the FGAM provides a better fit to this data than the FLM, we considered out-of-sample prediction of the log-particulate matter. We also compare the two different basis constructions for the tensor product surface in our model: the McLean et al. (2013a) construction (FGAM) and the construction from the current work (FGAMM). We also fit the fully nonparametric kernel estimator of Ferraty and Vieu (2006a). We considered fitting models with a single functional predictor using either the truck speeds or the accelerations, as well as models including both of these functional predictors at once. Since multiple functional predictors or scalar covariates

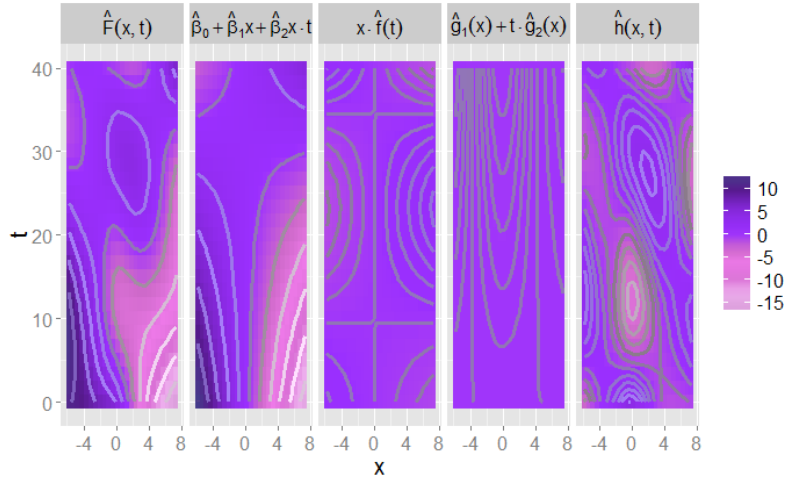


Figure 7: Contours of estimated surface,  $\hat{F}(x, t)$ , and components of FGAMM fit from using acceleration trajectories as predictors. The second panel is the parametric component of the fit, third is the component parametric in  $x$  and nonparametric in  $t$ , fourth vice versa, and finally  $\hat{f}(x, t)$  is nonparametric in both and subject to a fourth order penalty

do not seem to be implemented for the model in Ferraty and Vieu (2006a), we could not consider the model with both functional predictors for that method. It is also for this reason that we do not include a categorical predictor for the driving cycle for any of the models in the reported results. Perhaps surprisingly, including the categorical predictor for the FLM and FGAM methods had little effect on out-of-sample predictive performance.

For FLM, FGAM, and FGAMM, smoothing parameters were chosen using REML. The nonparametric kernel estimator was fit using code from the authors (Ferraty and Vieu 2006b), and includes automatic bandwidth selection. Several different basis dimensions were considered for both the FLM and FGAMs. Results varied little as the number of basis functions changed for each of the methods, so for compactness we only report the values that produced the lowest root-mean-square error (RMSE) averaged over the different predictors for each method. For the FLM this was ten basis functions for the functional coefficient, for FGAM this was six basis functions for both axes, and for FGAMM this was ten basis functions for each axis for the one functional predictor models and seven basis functions for the two predictor model. Both the FGAM and the FLM can be fit in R using the `refund` package (Crainiceanu et al. 2013). The underlying estimation is handled by the R recommended package `mgcv` (Wood 2011).

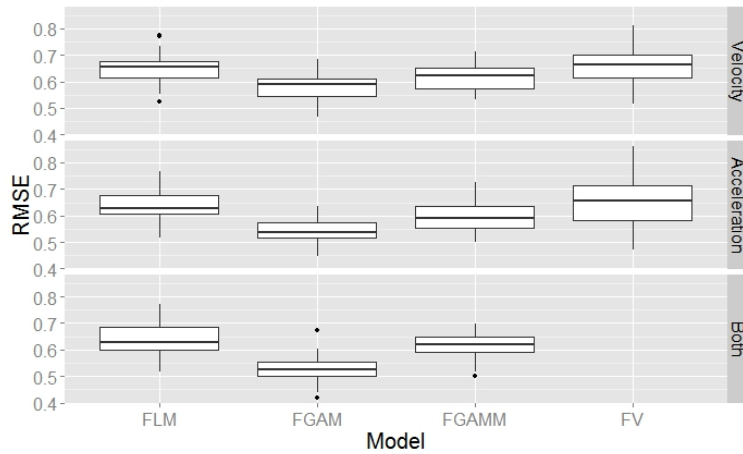


Figure 8: Boxplots for prediction error for FLM, FGAM, FGAMM, and FV using truck speed, acceleration, or both at once as predictors

For FGAMM, the variance components are estimated by the package `lme4` (Bates et al. 2013). The data was randomly divided so that 105 samples ( $\approx$  two thirds of the data) were using for training the models and the remaining samples were used for testing. Boxplots of the RMSEs for predicting the test set samples over 25 different random partitions into training-test sets is displayed in Figure 8.

We see that both FGAM and FGAMM had lower mean RMSE for the speed, acceleration, and two functional predictor models. While FGAMM performed similarly to FGAM when truck speed was the functional predictor, the FGAM basis construction provided better out-of-sample predictions for the other two models. Both parameterizations for the FGAM gave superior prediction results to the Ferraty and Vieu model as well. The lowest mean RMSE was achieved by the FGAM that included both functional predictors and used the more standard tensor product construction. A method for continuously predicting emissions over time for the data from this study using a functional response model is considered by Asencio et al. (2013). A non-functional data approach on an expanded data set from the original study authors can be found in Clark et al. (2010).

## 6 Conclusion

This work has addressed an important question, which to this point has few answers in the literature: when is a scalar-on-function regression problem not well-modelled by the functional linear model? How often is the relationship between the response and functional predictor truly nonlinear? Using an alternative mixed model representation for the recently proposed functional generalized additive model, we were able to develop simple tests for assessing linearity of an FGAM fit to functional data. Through two simulation studies we were able to find an approach that gave type I error rates quite close to the nominal level and also had high power. In an application to measuring the amount of pollutants emitted by heavy-duty trucks in various driving conditions, we presented strong evidence that particulate matter could not be adequately predicted from truck speed or acceleration using a functional linear model, whereas the (nonlinear) FGAM provided a much better fit to the data.

Of interest for future work is developing a test where the null model is FGAM to assess if an even more general model, such as the kernel estimator of Ferraty and Vieu (2006a) is necessary. This is more complicated because the model under the alternative is not a linear mixed model and because there is no explicit nesting of the model under the null hypothesis within the alternative model. It is also of interest to consider hypothesis testing for models with multiple functional predictors, assessing if adding a second functional covariate significantly improves model fit or perhaps assessing whether there is a significant interaction between two functional predictors. For example, for the emissions data, including a covariate for speed in a model already containing acceleration seems to offer little improvement in fit; we would like to formally assess this using similar testing procedures to ones developed in this work. Assessing goodness of fit using Bayes factors for a Bayesian version of FGAM (McLean et al. 2013b) is the subject of ongoing work.

## Acknowledgements

Much of this work was completed while the first author was a PhD student at Cornell University. He thanks NSERC for support while completing his studies. The authors wish to thank Ray Carroll for helpful comments on an early version of the manuscript and Oliver Gao for providing the emissions data.

## References

- Ait-Saïdi, A., F. Ferraty, R. Kassa, and P. Vieu (2008). "Cross-validated estimations in the single-functional index model". In: *Statistics* 42.6, pp. 475–494.
- Asencio, M., G. Hooker, and H. O. Gao (2013). "Functional Convolution Models". In: *Statistical Modelling*. In press.
- Bates, D., M. Maechler, and B. Bolker (2013). *lme4: Linear mixed-effects models using Eigen and syntax, R package version 1.0-4*.
- Cardot, H., F. Ferraty, A. Mas, and P. Sarda (2003). "Testing hypotheses in the functional linear model". In: *Scandinavian Journal of Statistics* 30.1, pp. 241–255.
- Chen, D., P. Hall, and H. G. Müller (2011). "Single and multiple index functional regression models with nonparametric link". In: *The Annals of Statistics* 39.3, pp. 1720–1747.
- Clark, N. N., M. Gautam, W. S. Wayne, G. W. Lyons, and G. J. Thompson (2007). *Heavy-duty vehicle chassis dynamometer testing for emissions inventory, air quality modeling, source apportionment, and air toxics emissions inventory*. Tech. rep. CRC Rep. No. E55/59. Alpharetta, GA: Coordinating Research Council, Inc. (CRC).
- Clark, N. N., K. A. Vora, L. Wang, M. Gautam, W. S. Wayne, and G. J. Thompson (2010). "Expressing cycles and their emissions on the basis of properties and results from other cycles". In: *Environmental science & technology* 44.15, pp. 5986–5992.
- Crainiceanu, C. M., P. T. Reiss, J. Goldsmith, L. Huang, L. Huo, and F. Scheipl (2013). *refund: Regression with Functional Data, R package version 0.1-7*.
- Crainiceanu, C. M. and D. Ruppert (2004). "Likelihood ratio tests in linear mixed models with one variance component". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66.1, pp. 165–185.
- Febrero-Bande, M., P. Galeano, and W. González-Manteiga (2013). "Generalized additive models for functional data". In: *TEST* 22.2, pp. 278–292.
- Febrero-Bande, M. and M. Oviedo de la Fuente (2012). "Statistical Computing in Functional Data Analysis: The R Package *fda.usc*". In: *Journal of Statistical Software* 51.4, pp. 1–28.
- Ferraty, F. and P. Vieu (2006a). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Verlag.
- (2006b). *npfda NonParametric Functional Data Analysis*. URL: <http://www.math.univ-toulouse.fr/staph/npfda/> (visited on 2013-10-08).
- Gabrys, R., L. Horváth, and P. Kokoszka (2010). "Tests for error correlation in the functional linear model". In: *Journal of the American Statistical Association* 105.491, pp. 1113–1125.

- García-Portugués, E, W. González-Manteiga, and M. Febrero-Bande (2013). "A goodness-of-fit test for the functional linear model with scalar response". In: *Journal of Computational and Graphical Statistics*, to appear.
- Goldsmith, J., J. Bobb, C. M. Crainiceanu, B. Caffo, and D. Reich (2011). "Penalized Functional Regression". In: *Journal of Computational and Graphical Statistics* 20.4, pp. 830–851.
- Greven, S., C. M. Crainiceanu, H. Küchenhoff, and A. Peters (2008). "Restricted likelihood ratio testing for zero variance components in linear mixed models". In: *Journal of Computational and Graphical Statistics* 17.4, pp. 870–891.
- Guillas, S. and M. J. Lai (2010). "Bivariate splines for spatial functional regression models". In: *Journal of Nonparametric Statistics* 22.4, pp. 477–497.
- Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*. Springer.
- James, G. M. and B. W. Silverman (2005). "Functional Adaptive Model Estimation." In: *Journal of the American Statistical Association* 100.470, pp. 565–577.
- Kong, D., A. M. Staicu, and A. Maity (2013). *Classical Testing in Functional Linear Models*. URL: [http://www4.stat.ncsu.edu/~staicu/papers/ClassicalTest\\_FLM\\_KSM.pdf](http://www4.stat.ncsu.edu/~staicu/papers/ClassicalTest_FLM_KSM.pdf) (visited on 2013-10-06). Submitted.
- Li, Y., N. Wang, and R. J. Carroll (2010). "Generalized functional linear models with semiparametric single-index interactions". In: *Journal of the American Statistical Association* 105.490.
- Marx, B. D. and P. H. C. Eilers (2005). "Multidimensional penalized signal regression". In: *Technometrics* 47.1, pp. 13–22.
- McLean, M. W., G. Hooker, A. M. Staicu, F. Scheipl, and D. Ruppert (2013a). "Functional Generalized Additive Models". In: *Journal of Computational and Graphical Statistics*. doi: 10.1080/10618600.2012.729985.
- McLean, M. W., F. Scheipl, G. Hooker, S. Greven, and D. Ruppert (2013b). *Bayesian Functional Generalized Additive Models with Sparsely Observed Covariates*. arXiv: 1305.3585 [stat.ME]. (Visited on 2013-10-06).
- Müller, H. G., Y. Wu, and F. Yao (2013). "Continuously additive models for nonlinear functional regression". In: *Biometrika* 100.3. doi: 10.1093/biomet/ast004.
- Müller, H. G. and F. Yao (2008). "Functional additive models". In: *Journal of the American Statistical Association* 103.484, pp. 1534–1544. ISSN: 0162-1459.
- Patterson, H. D. and R. Thompson (1971). "Recovery of inter-block information when block sizes are unequal". In: *Biometrika* 58.3, pp. 545–554.
- Pinheiro, J. C. and D. M. Bates (2000). *Linear mixed-effects models: basic concepts and examples*. Springer.



- Pinheiro, J. C., D. M. Bates, S. DebRoy, Deepayan Sarkar, and R Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-111.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN: 3-900051-07-0.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. Second. Springer.
- Reiss, P. T. and R. T. Ogden (2009). "Smoothing parameter selection for a class of semiparametric linear models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 505–523.
- Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Cambridge University Press.
- SAS Institute Inc. (2008). *SAS/STAT®9.2 User's Guide*. Cary, NC: SAS Institute Inc.
- Scheipl, F., S. Greven, and H. Küchenhoff (2008). "Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models". In: *Computational Statistics & Data Analysis* 52.7, pp. 3283–3299.
- Swihart, B. J., J. Goldsmith, and C. M. Crainiceanu (2013). *Restricted likelihood ratio tests for functional effects in the functional linear model*. Tech. rep. 247. Johns Hopkins University, Dept. of Biostatistics.
- Wang, X. and D. Ruppert (2013). *Optimal Prediction in an Additive Functional Model*. arXiv: 1301.4954 [math.ST]. (Visited on 2013-10-06).
- Wang, Y. (2011). *Smoothing Splines: Methods and Applications*. CRC Press.
- Wang, Y. and H. Chen (2012). "On testing an unspecified function through a linear mixed effects model with multiple variance components". In: *Biometrics* 68.4, pp. 1113–1125. doi: 10.1111/j.1541-0420.2012.01790.x.
- Wood, S. N. (2006). "Low-Rank Scale-Invariant Tensor Product Smoother for Generalized Additive Mixed Models". In: *Biometrics* 62.4, pp. 1025–1036.
- (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.1, pp. 3–36.
- (2013). "On p-values for smooth components of an extended generalized additive model". In: *Biometrika* 100.1, pp. 221–228.
- Wood, S. N., F. Scheipl, and J. J. Faraway (2013). "Straightforward intermediate rank tensor product smoothing in mixed models". In: *Statistics and Computing* 23.3, pp. 341–360.
- Zhang, J.-T. and J. Chen (2007). "Statistical inferences for functional data". In: *The Annals of Statistics* 35.3, pp. 1052–1079.