TECHNOMETRICS
2021, VOL. 63, NO. 3, 396–409
https://doi.org/10.1080/00401706.2020.1802350

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS
Check for updates

Robust Function-on-Function Regression

Harjit Hullait[a], David S. Leslie[b], Nicos G. Pavlidis[c], and Steve King[d]

[a]STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, UK; [b]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK; [c]Department of Management Science, Lancaster University, Lancaster, UK; [d]Rolls Royce plc, London, UK

ABSTRACT

Functional linear regression is a widely used approach to model functional responses with respect to functional inputs. However, classical functional linear regression models can be severely affected by outliers. We therefore introduce a Fisher-consistent robust functional linear regression model that is able to effectively fit data in the presence of outliers. The model is built using robust functional principal component and least squares regression estimators. The performance of the functional linear regression model depends on the number of principal components used. We therefore introduce a consistent robust model selection procedure to choose the number of principal components. Our robust functional linear regression model can be used alongside an outlier detection procedure to effectively identify abnormal functional responses. A simulation study shows our method is able to effectively capture the regression behavior in the presence of outliers, and is able to find the outliers with high accuracy. We demonstrate the usefulness of our method on jet engine sensor data. We identify outliers that would not be found if the functional responses were modeled independently of the functional input, or using nonrobust methods.

ARTICLE HISTORY
Received June 2019
Accepted July 2020

KEYWORDS
Outlier detection; Robust functional data analysis; Robust model selection

1. Introduction

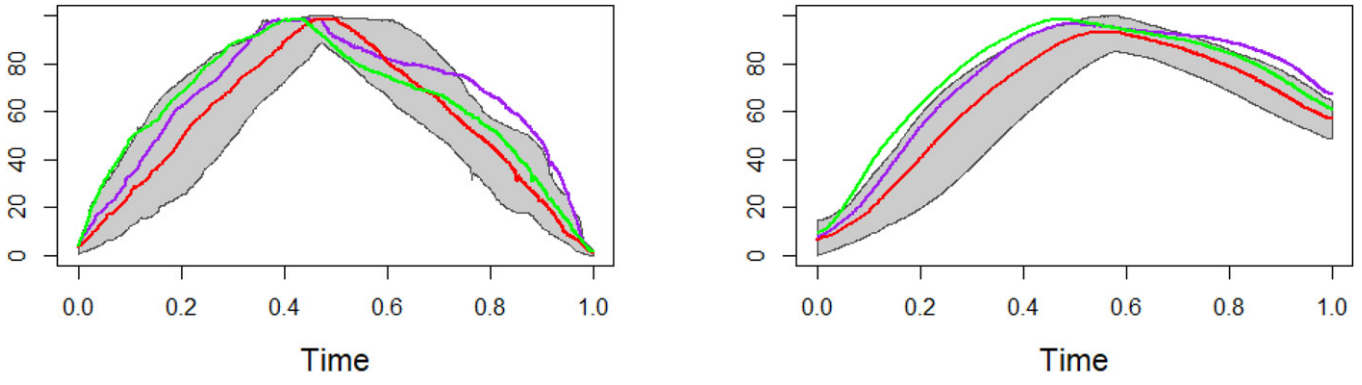
Functional linear regression (FLR) in the function-on-function case (Ramsay and Dalzell 1991) is a widely used technique for modeling functional responses with respect to functional inputs. However, classical FLR models can be severely affected by outliers as we will demonstrate via a simulation study in Section 6. We therefore develop a robust FLR (RFLR) model, which is able to effectively fit the data in the presence of outliers. The model is built using the robust functional principal component (FPC) model by Bali et al. (2011) and the multivariate least trimmed squares (MLTS) estimator by Agulló, Croux, and Van Aelst (2008).

Our study of FLR is motivated by the need to identify unusual jet engine behavior using sensor data captured during a *pass-off* test (the last test prior to the deployment of a jet engine). Jet engines are highly complex machines consisting of tens of thousands of parts, and although the physical behavior of individual components can be modeled by systems of partial or ordinary differential equations, such a description is far from adequate to model the entire engine (Houstis et al. 2002). During a pass-off test an engineer that controls the speed of the engine performs a number of maneuvers which can be defined as different accelerations and decelerations starting and ending at a set idle speed. Although each maneuver is characterized by a distinct engine speed profile, its execution differs each time it is performed. Throughout the duration of the test sensor measurements for numerous engine parameters, including speed, pressure, temperature, and vibration at different parts on the engine, are recorded at very high frequency. The volume of data generated in each test is very large and only a small proportion of this is manually inspected to identify potential engine issues. Currently there exists no engine simulation model to predict the evolution of these parameters during the maneuvers performed by the controller. Instead engineers use their judgment to determine whether the sensor measurements deviate sufficiently from their expectations to indicate a potential engine issue.

An automated approach to identifying pass-off test segments that require closer inspection will enable a larger proportion of the pass-off data to be processed, and can thus improve the early identification of engine issues. However, to achieve this goal it is first necessary to develop accurate models of the expected behavior of a healthy engine during different maneuvers. The available data consist of sensor measurements captured during pass-off tests. For each individual engine, we do not know whether it had issues or not, but we do know that the dataset contains tests for engines that had issues.

In this work, we develop RFLR to model engine temperature parameters during the vibration survey (VS), a key pass-off test maneuver during which the engine is gradually accelerated to a certain speed and then slowly decelerated. Both engine speed and temperature are continuous quantities that are recorded at very high frequencies. Therefore, the central assumption underlying functional data analysis (FDA), namely that the observed time series constitute discrete measurements from a smooth functional process is realistic in our application. We

CONTACT Harjit Hullait h.hullait@lancaster.ac.uk STOR-i Centre for Doctoral Training, Lancaster University, Lancaster LA1 4YW, UK.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TECH.
© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



(a) Turbine Pressure Ratio (b) Turbine Gas Temperature

Figure 1. Envelope plots of turbine pressure ratio (TPR) and turbine gas temperature (TGT) time series in gray. Sample 33 (green) is identified as an outlier when the relationship between TPR and TGT is not accounted for, but is not associated with an engine issue. Samples 106 (purple) and 77 (red) are not identified as outliers by direct outlier detection methods, but are linked with engine faults. These are correctly identified using RFLR.

use FLR due to the documented ability of this methodology to capture complex dependency structures (Morris 2015). Ramsay and Silverman (2005) have illustrated that the FLR model simultaneously uses information across functions to identify their commonalities (called *replication*), and information about the expected underlying structural relationships within each function (called *regularization*). However, the flexibility to use information from "future" time points implies that the models developed do not truly approximate the underlying mechanical process.

For our analysis, we will use 199 VS datasets, which include information about the turbine pressure ratio (TPR) (the human controlled parameter that measures engine speed) and various temperature features including the turbine gas temperature (TGT). Figure 1 provides envelope plots of TPR and TGT for all the VS maneuvers. To anonymize the data, we have transformed the time index onto the interval $[0, 1]$ and the sensor measurements to the range $[0, 100]$. The VS maneuvers are performed by a human controller, which causes variability in TPR as depicted in Figure 1(a). This naturally affects the TGT curves and can mask unusual behavior arising from an engine issue. In fact outlier detection methods applied directly on the TGT curves (as shown in Figure 1(b)) fail to identify meaningful outliers, and instead identify exclusively curves produced by unusual TPR profiles. The green TGT curve in Figure 1(b) is such an example. Direct outlier detection methods identify this curve as an outlier but there are no issues with this engine. This curve only appears to be an outlier due to the relatively uncommon TPR schedule during this maneuver (see Figure 1(a)).

We therefore require a method of detecting outliers in the presence of the controller induced variability. In particular, we expect that the relationship between engine speed (TPR) and engine temperature for different VS maneuvers should be the same irrespective of the way the maneuver is performed. For example given a certain engine acceleration we would expect a certain temperature response. If, however, the response differs from expectation this could be indicative of an engine issue. In Section 6, we show that classical FLR can be severely affected when the dataset contains outliers, which in our case correspond to data from VS maneuvers of engines with issues. The RFLR model we propose aims to overcome this limitation. Applied on the data in Figure 1, our model identifies the TGT curves with purple and red color as outliers. Notice that both curves do not appear to diverge from the general pattern in the envelope plot in Figure 1(b), and are only identified as outliers when the functional input from the corresponding TPR curves is taken into account. Both VS maneuvers correspond to engines for which issues were identified during the test.

The article is organized as follows. In Section 2, we outline the classical FLR model. In Section 3, we will outline robust FDA techniques to obtain an RFLR model. We also introduce a robust model selection procedure. In Section 4, we prove consistency results for the RFLR model and the robust model selection procedure. In Section 5, we describe an outlier detection method, which acts on the residuals of the RFLR model. In Section 6, we perform a simulation study to illustrate the model fitting and outlier detection capabilities of the robust model. In Section 7, we apply the robust model on the engine data and highlight unusual observations that cannot be detected by using outlier detection methods directly on the temperature curves. The article ends with concluding remarks in Section 8.

2. Classical Functional Data Analysis

In this section, we give a brief summary of the FDA tools that we will later apply in our model. In the following sections, we will use the vector space $L^2(I)$ which is the Hilbert space of square integrable functions on the compact interval I with the inner product $\langle f, g\rangle = \int_I f(t)g(t)dt$ for functions $f, g \in L^2(I)$.

We will define $X(t), Y(t)$ to be univariate stochastic processes defined on I, with mean functions $\mu^X(t)$ and $\mu^Y(t)$, and covariance functions $C_X(s,t) = \text{cov}\{X(s), X(t)\}$ and $C_Y(s,t) = \text{cov}\{Y(s), Y(t)\}$ for all $s, t \in I$. We shall define $\boldsymbol{x}(t) = [x_1(t), \ldots, x_n(t)]$ and $\boldsymbol{y}(t) = [y_1(t), \ldots, y_n(t)]$ to be n independent and identically distributed realizations of $X(t)$ and $Y(t)$, respectively.

In practice, we observe $x_i(t)$ and $y_i(t)$ at discrete time points. We shall assume for simplicity of exposition that observations are made at equally spaced time points $t_1, \ldots, t_T$. We will outline functional linear regression and functional principal component analysis (FPCA) with respect to the underlying functions

$x(t), y(t)$. In Section 2.3, we need to use the discretely observed data to define a suitable model selection criterion.

2.1. Functional Linear Regression

In this section, we will introduce the classical FLR model (Ramsay and Dalzell 1991). In FLR, we model the relationship between predictor $x_i(t)$ and response $y_i(t)$ as

$$y_i(t) = \alpha(t) + \int_I x_i(s)\beta(s,t)ds + \epsilon_i(t), \tag{1}$$

where $\alpha(t)$ is the intercept function, $\beta(s,t)$ is the regression function, and $\epsilon_i(t)$ is the error process. For a fixed t, we can think of $\beta(s,t)$ as the relative weight placed on $x_i(s)$ to predict $y_i(t)$. As in Chiou, Yang, and Chen (2016), we will assume the mean functions $\mu^X(t) = 0$ and $\mu^Y(t) = 0$ which thereby means $\alpha(t) = 0$. This is a reasonable assumption as in practice we can calculate the mean functions $\mu^X(t)$ and $\mu^Y(t)$ efficiently for dense data and then preprocess the data by subtracting $\mu^X(t)$ and $\mu^Y(t)$ from the observed curves.

FLR in the function-on-function case can be modeled parametrically (Yao, Müller, and Wang 2005; Chiou, Yang, and Chen 2016) or nonparametrically (Ferraty, Keilegom, and Vieu 2012; Ivanescu et al. 2015; Scheipl, Staicu, and Greven 2015). We use a parametric approach which models the regression matrix in terms of predefined basis functions. We obtain the approximations $\hat{x}_i(t)$ and $\hat{y}_i(t)$ in terms of (M, K) pre-chosen orthonormal basis functions $\phi_m^X(t), \phi_k^Y(t)$, respectively:

$$\hat{x}_i(t) = \sum_{m=1}^{M} z_{im}\phi_m^X(t) \text{ and } \hat{y}_i(t) = \sum_{k=1}^{K} w_{ik}\phi_k^Y(t),$$

where $z_{im}, w_{ik} \in \mathbb{R}$. We define $\boldsymbol{\phi}^{\boldsymbol{X}}(t) = [\phi_1^X(t), \ldots, \phi_M^X(t)]$, $\boldsymbol{\phi}^{\boldsymbol{Y}}(s) = [\phi_1^Y(s), \ldots, \phi_K^Y(s)]$, $\boldsymbol{z}_i = [z_{i1}, \ldots, z_{iM}]$, and $\boldsymbol{w}_i = [w_{i1}, \ldots, w_{iK}]$. We then model the regression surface using a double basis expansion (Ramsay and Silverman 2005):

$$\beta(s,t) = \sum_{m=1}^{M}\sum_{k=1}^{K} b_{mk}\phi_m^X(s)\phi_k^Y(t) = \boldsymbol{\phi}^{\boldsymbol{X}}(s)^T\boldsymbol{B}\boldsymbol{\phi}^{\boldsymbol{Y}}(t), \tag{2}$$

for an $M \times K$ regression matrix $\boldsymbol{B}$. Using the above in Equation (1); expressing $\epsilon_i(t) = \boldsymbol{q}_i^T\boldsymbol{\phi}^{\boldsymbol{Y}}(t)$, and using the orthonormality of $\boldsymbol{\phi}^{\boldsymbol{X}}(t)$ and $\boldsymbol{\phi}^{\boldsymbol{Y}}(t)$, gives

$$\begin{aligned}
\hat{y}_i(t) &= \int_I \hat{x}_i(s)\beta(s,t)ds + \epsilon_i(t),\\
\hat{y}_i(t) &= \int_I [\boldsymbol{z}_i^T\boldsymbol{\phi}^{\boldsymbol{X}}(s)][\boldsymbol{\phi}^{\boldsymbol{X}}(s)^T\boldsymbol{B}\boldsymbol{\phi}^{\boldsymbol{Y}}(t)]ds + \epsilon_i(t),\\
\hat{y}_i(t) &= \boldsymbol{z}_i^T\left\{\int_I \boldsymbol{\phi}^{\boldsymbol{X}}(s)\boldsymbol{\phi}^{\boldsymbol{X}}(s)^T ds\right\}\boldsymbol{B}\boldsymbol{\phi}^{\boldsymbol{Y}}(t) + \epsilon_i(t),\\
\hat{y}_i(t) &= \boldsymbol{z}_i^T\boldsymbol{B}\boldsymbol{\phi}^{\boldsymbol{Y}}(t) + \epsilon_i(t),\\
\boldsymbol{w}_i^T\boldsymbol{\phi}^{\boldsymbol{Y}}(t) &= \boldsymbol{z}_i^T\boldsymbol{B}\boldsymbol{\phi}^{\boldsymbol{Y}}(t) + \boldsymbol{q}_i^T\boldsymbol{\phi}^{\boldsymbol{Y}}(t),\\
\boldsymbol{w}_i^T\int_I \boldsymbol{\phi}^{\boldsymbol{Y}}(t)\boldsymbol{\phi}^{\boldsymbol{Y}}(t)^T dt &= \boldsymbol{z}_i^T\boldsymbol{B}\int_I \boldsymbol{\phi}^{\boldsymbol{Y}}(t)\boldsymbol{\phi}^{\boldsymbol{Y}}(t)^T dt + \boldsymbol{q}_i^T\\
&\quad\times\int_I \boldsymbol{\phi}^{\boldsymbol{Y}}(t)\boldsymbol{\phi}^{\boldsymbol{Y}}(t)^T dt,\\
\boldsymbol{w}_i &= \boldsymbol{z}_i^T\boldsymbol{B} + \boldsymbol{q}_i^T.
\end{aligned} \tag{3}$$

This parameterization of the residual function is also used by Chiou, Yang, and Chen (2016). We can then estimate $\boldsymbol{B}$ using standard multivariate regression methods typically assuming Gaussian $\boldsymbol{q}_i$.

2.2. Functional Principal Component Analysis

In this section, we describe FPCA, which we will use to estimate the basis functions $\boldsymbol{\phi}^{\boldsymbol{X}}(t)$ and $\boldsymbol{\phi}^{\boldsymbol{Y}}(t)$ for $x_i(t)$ and $y_i(t)$, respectively. These basis functions give effective, low-dimensional representations and will be used in the FLR model described in Section 2.1.

FPCA is a method of finding dominant modes of variance for functional data. These dominant modes of variance are called the FPCs. FPCA is also used as a dimensionality reduction tool, as a set of observed curves can be effectively approximated by a linear combination of a small set of FPCs.

Let $X(t)$ be a mean-centered stochastic process with covariance function $C_X(s,t)$, with eigenfunctions $\phi_m^X(t)$ and eigenvalues λ_m^X with $m \in \mathbb{N}$, where the eigenfunctions are ordered in decreasing order of the corresponding eigenvalues. The Karhunen–Loéve theorem (Shang 2014) states

$$X(t) = \sum_{m=1}^{\infty} \xi_m^X\phi_m^X(t),$$

where each ξ_m^X is a random variable. Let $x_i(t)$, $i = 1, \ldots, n$ be n independent realizations of $X(t)$. We can write $x_i(t) = \sum_{m=1}^{\infty} z_{im}\phi_m^X(t)$ where the principal component scores, z_{im}, constitute realizations of ξ_m^X. Since the eigenfunctions are orthonormal the scores can be defined as $z_{im} = \int_I x_i(t)\phi_m^X(t)dt$.

In practice, a large proportion of the variability in $x_i(t)$ can be captured by using a finite number M, of the first eigenfunctions. We can define the M-truncation as

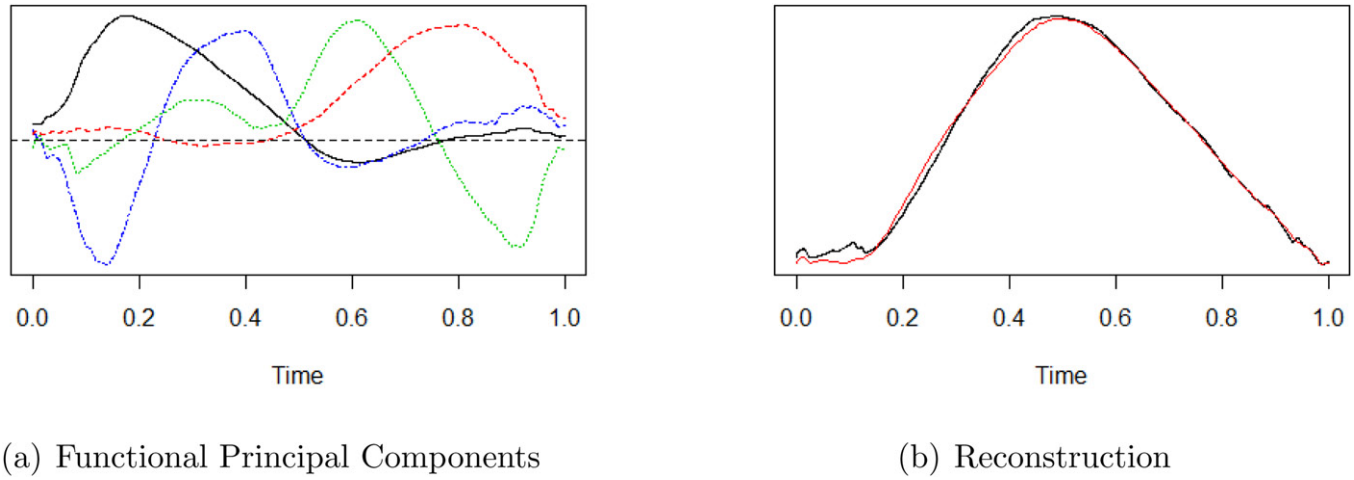
$$\hat{x}_i^M(t) = \sum_{m=1}^{M} z_{im}\phi_m^X(t). \tag{4}$$

This minimizes the reconstruction error, subject to the constraint that M basis functions are used

$$\frac{1}{n}\sum_{i=1}^{n}\|x_i - \hat{x}_i^M\|^2 = \frac{1}{n}\sum_{i=1}^{n}\int_I [x_i(t) - \hat{x}_i^M(t)]^2 dt. \tag{5}$$

To choose M, we will use an information criterion discussed in the next subsection. An analogous procedure is used to determine the K eigenfunctions $\phi_k^Y(t)$ for $y(t)$.

In Figure 2(a), we depict the first four eigenfunctions for the TGT curves shown in Figure 1(b). Note that the eigenfunctions do not resemble the shape of the TGT curves because, as described above, they are estimated after the data is preprocessed by subtracting the mean function, $\alpha(t)$ in Equation (1). In Figure 2(b), we depict the reconstruction of a single TGT curve using the first four eigenfunctions (four-truncation). The reconstruction is formed by taking the mean function plus a weighted sum of the eigenfunctions. We can see that using only four eigenfunctions produces an accurate approximation of the shape of this TGT curve except at the start where the discrepancy is noticeable.



(a) Functional Principal Components (b) Reconstruction

Figure 2. Plot of the first four eigenfunctions for turbine gas temperature (TGT) data and a plot of a TGT curve (black) and its reconstruction using the first four eigenfunctions (red).

2.3. Bayesian Information Criterion for FLR

In the FLR model described in Section 2.1, we need to choose terms M and K. Typically M and K are chosen independently (Yao, Müller, and Wang 2005), however, the estimation of $\beta(s,t)$ also depends on M and K and this should be incorporated into the estimation of these terms. In this section, we formulate a Bayesian information criterion (BIC) to determine the basis size M and K, similarly to Matsui (2020).

A component of the BIC is the log-likelihood, often expressed as a squared error term. It is tempting to use the squared error resulting from Equation (3). However, the objective is to fit the data $\boldsymbol{y}_i$ so we should use a likelihood of this data instead of a squared error term of basis coefficients.

We have a set of models $J = \{(M,K)\,|\,M = 1,\ldots,M_{\max}, K = 1,\ldots,K_{\max}\}$, where $M_{\max}$ and $K_{\max}$ are preset maximum number of FPCs that will be considered in the model. Let vector $\vec{\boldsymbol{y}}_i$ be the values of $y_i(t)$ evaluated at discrete time points: $\vec{\boldsymbol{y}}_i = [y_i(t_1),\ldots,y_i(t_T)]$. Let $\boldsymbol{z}_i^{(M)}$ be the first M principal scores of $x_i(t)$ with respect to the FPCs $\boldsymbol{\phi}^{\boldsymbol{X}}(t)$ and let $\boldsymbol{\phi}^{(\boldsymbol{K})}$ be the matrix with (k,i) entry $\phi_k^Y(t_i)$. We assume there exists a true model (M_0,K_0) with associated $M_0 \times K_0$ matrix $\boldsymbol{B}^{\boldsymbol{M_0,K_0}}$ such that

$$\vec{\boldsymbol{y}}_i = (\boldsymbol{z}^{(\boldsymbol{M_0})}{}_i)^T \boldsymbol{B}^{\boldsymbol{M_0,K_0}} \boldsymbol{\phi}^{(\boldsymbol{K_0})} + \boldsymbol{\epsilon}_i, \tag{6}$$

where the error $\boldsymbol{\epsilon}_i = [\epsilon_i(t_1),\ldots,\epsilon_i(t_T)]$ is assumed for simplicity to be sampled from $N(0,v^2\boldsymbol{I_T})$, where $\boldsymbol{I_T}$ is the identity matrix of size T. The assumption of independent observations in Equation (8) is not tenable in the general case. However, in our application, this simplification is realistic because in accord with the practice followed by engineers, we subsample the high frequency sensor data at every 200 time points.

For Model (M,K), we define the parameters $\boldsymbol{\theta}^{\boldsymbol{M,K}} = (\boldsymbol{B}^{\boldsymbol{M,K}}, v^{M,K})$ and the prediction $\hat{\boldsymbol{y}}^{\boldsymbol{M,K}}{}_i = (\boldsymbol{z}^{(\boldsymbol{M})}{}_i)^T \boldsymbol{B}^{\boldsymbol{M,K}} \boldsymbol{\phi}^{(\boldsymbol{K})}$. We want to identify this true model (M_0,K_0), which we can use to obtain consistent estimates of $\boldsymbol{\theta}^{\boldsymbol{M_0,K_0}}$. For each Model (M,K), we define the likelihood for sample i as

$$f(\vec{\boldsymbol{y}}_i|\boldsymbol{\theta}^{\boldsymbol{M,K}}) = \frac{1}{(2\pi)^{\frac{T}{2}}(v^{M,K})^T} \exp \left\{ -\frac{[\vec{\boldsymbol{y}}_i - \hat{\boldsymbol{y}}^{\boldsymbol{M,K}}{}_i]^T[\vec{\boldsymbol{y}}_i - \hat{\boldsymbol{y}}^{\boldsymbol{M,K}}{}_i]}{2(v^{M,K})^2} \right\}, \tag{7}$$

and the log-likelihood $l(\boldsymbol{\theta}^{\boldsymbol{M,K}}) = \sum_{i=1}^n \log(f(\vec{\boldsymbol{y}}_i|\boldsymbol{\theta}^{\boldsymbol{M,K}}))$. As in Eilers and Marx (1996)

$$\mathrm{BIC}_n(M,K) = -2l(\hat{\boldsymbol{\theta}}^{\boldsymbol{M,K}}) + \omega(M,K)\log(n), \tag{8}$$

where $\hat{\boldsymbol{\theta}}^{\boldsymbol{M,K}}$ is the maximum likelihood estimator and the penalty $\omega(M,K) = MK + 1$, in which MK is the number of free parameters in the model and the 1 comes from v. We will denote $(M^*,K^*)_n = \arg\min_{(M,K)\in J} \mathrm{BIC}_n(M,K)$, which depends on the sample size n.

To summarize, we estimate the FPCs for X and Y and solve the FLR model for different models (M,K). We then choose model $(M^*,K^*)_n$ that minimizes the BIC criterion. The robust equivalent of this procedure is given in Algorithm 1.

3. Robust Functional Linear Regression

In Section 2, we have defined the FLR model and have outlined the use of FPCA bases to estimate parameters of the model. In this section, we will introduce robust versions of the FDA techniques outlined in Section 2. This will allow us to fit a model that captures the expected relationship of the predictor and response functions even in the presence of outliers. We shall also propose a robust BIC procedure for model selection.

We will replace classical FPCA with robust FPCA estimates by Bali et al. (2011) which ensure that outliers do not unduly affect the FPCA estimates. Note that FPCA minimizes the residual error given in (5). To obtain robust FPCA estimates, Bali et al. (2011) minimized a robust scale estimator, using a projection pursuit approach, which iteratively performs a weighted least squares till the estimators stabilize.

Analogous to (4), the robust FPCs $\tilde{\phi}_m^X(t)$, with $m = 1,\ldots,M$, and $\tilde{\phi}_k^Y(t)$ with $k = 1,\ldots,K$, are orthonormal functions such that

$$\tilde{x}_i(t) = \sum_{m=1}^{M} \tilde{z}_{im}\tilde{\phi}_m^X(t), \quad \tilde{y}_i(t) = \sum_{k=1}^{K} \tilde{w}_{ik}\tilde{\phi}_k^Y(t), \tag{9}$$

are good approximations for $x_i(t)$ and $y_i(t)$. We define $\tilde{y}_i(t) = \tilde{\boldsymbol{w}}_i^T\tilde{\boldsymbol{\phi}}^{\boldsymbol{Y}}(t)$ and assume as in (3) that $\boldsymbol{\epsilon}_i = \tilde{\boldsymbol{q}}_i^T\tilde{\boldsymbol{\phi}}^{\boldsymbol{Y}}(s)$. We can now write the robust counterpart of (3) as

$$\tilde{\boldsymbol{w}}_i^T = \tilde{\boldsymbol{z}}_i^T\tilde{\boldsymbol{B}} + \tilde{\boldsymbol{q}}_i^T. \tag{10}$$

To obtain a robust estimate of the regression matrix $\tilde{\boldsymbol{B}}$, we will use the MLTS estimator by Agulló, Croux, and Van Aelst (2008), to mitigate the affect of outliers with respect to the regression relationship.

Let $[\cdot] : \mathbb{R} \to \mathbb{Z}$ denote the function that rounds its input to the nearest integer. Then, for $\alpha \in [0, 1]$ we define the set $\mathcal{S} = \{S \subset \{1, \ldots, n\}, |S| = [\alpha n]\}$. The objective of MLTS is to find a subset S such that

$$S = \arg\min_{S \in \mathcal{S}} \sum_{i \in S} \|\tilde{\boldsymbol{w}}_i^T - \tilde{\boldsymbol{z}}_i^T \tilde{\boldsymbol{B}}\|^2.$$

This is robust as outliers will not be in the subset by definition so shall not affect the model estimation. According to the engineers' assessment between 1% and 5% of the data correspond to engines with issues. To ensure that no data from such engines are included in the calculation of the robust estimators we used a subset of size $[0.8n]$.

3.1. Robust Bayesian Information Criterion for FLR

The BIC model selection method is known to be non-robust (Machado 1993). In particular, outliers can significantly affect the log-likelihood estimation. We therefore outline a robust BIC (RBIC) model, which, similar to MLTS, maximizes over a subset of samples S. RBIC can therefore give good model selection performance in the presence of outliers.

Denote as $\tilde{\boldsymbol{\theta}}^{\boldsymbol{M,K}} = (\tilde{\boldsymbol{B}}^{\boldsymbol{M,K}}, \tilde{\nu}^{M,K})$ the robust estimated parameters for model (M, K), where $\tilde{\boldsymbol{B}}^{\boldsymbol{M,K}}$ is the robust estimator of the regression matrix, obtained by solving Equation (10); and $\tilde{\nu}^{M,K}$ is a robust trimmed estimator of the variance in Equation (7). The robust prediction of $\boldsymbol{y}_i$ by model (M, K) is given by $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}^{\boldsymbol{M,K}} = (\tilde{\boldsymbol{z}}^{(\boldsymbol{M})}{}_i)^T \tilde{\boldsymbol{B}}^{\boldsymbol{M,K}} \tilde{\boldsymbol{\Phi}}^{(\boldsymbol{K})}$, where $\tilde{\boldsymbol{\Phi}}^{(\boldsymbol{K})} = [\tilde{\boldsymbol{\phi}}^{\boldsymbol{Y}}{}_1, \ldots, \tilde{\boldsymbol{\phi}}^{\boldsymbol{Y}}{}_K]$ are the first K robust FPCs used in Equation (9), and estimated through the method of Bali et al. (2011). We define the trimmed log-likelihood for model (M, K) and set S as

$$\tilde{l}(\tilde{\boldsymbol{\theta}}^{\boldsymbol{M,K}}, S) = \sum_{i \in S} \left(\frac{[\vec{\boldsymbol{y}}_i - \tilde{\boldsymbol{y}}^{\boldsymbol{M,K}}{}_i]^T [\vec{\boldsymbol{y}}_i - \tilde{\boldsymbol{y}}^{\boldsymbol{M,K}}{}_i]}{(\tilde{\nu}^{M,K})^2} \right) + rT \log(2\pi) + 2rT \log(\tilde{\nu}^{M,K}). \quad (11)$$

We will define $S^{M,K} = \arg\min_{S \in \mathcal{S}} \tilde{l}(\tilde{\boldsymbol{\theta}}^{\boldsymbol{M,K}}, S)$, where $\mathcal{S} = \{S \subset \{1, \ldots, n\}, |S| = [\alpha n]\}$ where $\alpha = 0.8$. Then

$$\begin{aligned} \text{RBIC}_n(M, K) &= -2 \min_{S \in \mathcal{S}} \tilde{l}(\tilde{\boldsymbol{\theta}}^{\boldsymbol{M,K}}, S) + \omega(M, K) \log([\alpha n]), \\ &= -2\, \tilde{l}(\tilde{\boldsymbol{\theta}}^{\boldsymbol{M,K}}, S^{M,K}) + \omega(M, K) \log([\alpha n]). \quad (12) \end{aligned}$$

We will denote $(\tilde{M}, \tilde{K})_n = \arg\min_{(M,K) \in J} \text{RBIC}_n(M, K)$, and assume that there is a unique minimizer, as is standard in the BIC literature (Schwarz 1978).

In Algorithm 1, we outline the calculation of the RFLR model, which incorporates the RBIC procedure. In the algorithm, we estimate the model for different values of (M, K) and choose the model with the minimum RBIC value. We consider $M = 1, \ldots, M_{\max}$ and $l = 1, \ldots, K_{\max}$ where $M_{\max}, K_{\max}$ are chosen to ensure that 99.99% of the variance in the raw data is captured.

Data: Let $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ be mean-corrected time series of length T for $i = 1, \ldots, n$.
1. Estimate $\{\tilde{\phi}_1^X(t), \ldots, \tilde{\phi}_{M_{\max}}^X(t)\}$, $\{\tilde{\phi}_1^Y(t), \ldots, \tilde{\phi}_{K_{\max}}^Y(t)\}$ (Bali et al. 2011).
for $M = 1, \ldots, M_{\max}$ **do**
  for $K = 1, \ldots, K_{\max}$ **do**
    Estimate the regression matrix $\boldsymbol{B}^{\boldsymbol{M,K}}$ using MLTS (Agulló, Croux, and Van Aelst 2008).
    Obtain the $\text{RBIC}_n(M, K)$ value using (12)
  end for
end for
2. Select model $(\tilde{M}, \tilde{K})_n$.
return Regression matrix $\tilde{\boldsymbol{B}}$ from model $(\tilde{M}, \tilde{K})_n$ and $\{\tilde{\phi}_1^X(t), \ldots, \tilde{\phi}_{\tilde{M}}^X(t)\}$, $\{\tilde{\phi}_1^Y(t), \ldots, \tilde{\phi}_{\tilde{K}}^Y(t)\}$.

Algorithm 1: RFLR procedure

4. Consistency Results

In Section 3, we proposed a RFLR model for the function-on-function problem. In this section, we outline our asymptotic consistency results. The proofs are contained in the supplementary materials. Asymptotic consistency is a minimum requirement for a good model. As Silverman (1996) argued asymptotic consistency results should be seen as large-sample approximation results, which can aid our intuition about a method and are practically relevant when reasonably large amounts of data are available. Silverman (1996) also argued that such results are particularly important for FDA because in certain contexts "obvious" procedures are not consistent and may therefore produce misleading insights.

Definition 1. Let $X_1, X_2, \ldots, X_n$ be a sequence of real-valued random variables. An estimator $T_n := T(X_1, X_2, \ldots, X_n)$ of a parameter θ is said to be (asymptotically) *consistent* if for all $\varepsilon > 0$

$$\lim_{n \to \infty} P(|T_n - \theta| > \varepsilon) = 0.$$

Definition 2. Let $X_1, X_2, \ldots, X_n$ be a sequence of real-valued random variables with an associated cumulative distribution function F_θ, which depends on an unknown parameter θ. Let the estimator $T_n := T(F_n)$ of a parameter θ, be a function of the empirical distribution function F_n. We say this estimator is *Fisher-consistent* for the parameter θ if

$$T(F_\theta) = \theta.$$

Remark 1. Fisher consistency is equivalent to (asymptotic) consistency if the empirical distribution function F_n converges pointwise to the true distribution function F_θ. This can be shown to be the case for iid real multivariate random variables using the Glivenko–Cantelli theorem (Pollard 2012).

Lemma 4.1. Assume C1–C6 (given in the supplementary material) hold then the robust regression function $\tilde{\beta}(s, t)$ is Fisher-consistent.

Corollary 4.1. If $\{x_1(t), y_1(t)\}, \ldots, \{x_n(t), y_n(t)\}$ are iid samples with cumulative distribution function (F_X, F_Y). Then, assuming

C1-C6 (given in the supplementary material) hold, $\tilde{\beta}(s,t)$ is consistent.

Note that $x_i(t)$ and $y_i(t)$ are defined on a finite number of eigenfunctions, so are defined by finite score vectors. Therefore, Corollary 4.1 follows from Lemma 4.1 and Remark 1, which states almost sure convergence of the empirical distribution for iid multivariate random variables. In this case, Fisher-consistency is equivalent to consistency.

Theorem 4.1. Given Assumptions 1 and 2 (given in the supplementary material) hold, and there exists a true model (M_0, K_0), then $(\tilde{M}, \tilde{K})_n$ is a consistent estimator of (M_0, K_0).

5. Outlier Detection

There is a rich literature on outlier detection methods for functional data. This includes methods based on outlyingness measures such as Arribas-Gil and Romo (2014) and Dai and Genton (2018), and methods relying on functional depth such as the thresholding approach by Febrero-Bande, Galeano, and Gonzãlez-Manteiga (2008) and the functional boxplot by Sun and Genton (2011). These methods are typically applied directly to functional representations of the response time-series, which in our case would be those of the engine temperature parameters. As the discussion around Figure 1 in the introduction indicated outliers identified through this approach can be meaningless, because the dependency between engine speed and temperature is ignored. Our approach relies on the premise that once this dependency is accounted for through an RFLR model the residual curve $r_i(t) = y_i(t) - \tilde{y}_i(t)$ of an engine that has issues will differ from other residual curves, and thus constitute an outlier. We therefore apply the outlier detection method of Febrero-Bande, Galeano, and Gonzãlez-Manteiga (2008) to residual curves from RFLR. We next outline the notion of depth and the motivation for selecting this approach.

To order univariate data one can directly use order statistics, however, ordering multivariate data is not trivial. This motivates depth functions, defined as $D : \mathbb{R}^d \to \mathbb{R}$, that map a multivariate point to a univariate depth value (Tukey 1975). A key property of a depth function is that a point x_1 that is closer to the "center" of a data cloud than a point x_2 will be assigned a larger depth value. Points sufficiently far from the center can be considered as outliers, and can be identified by their low depth values. The notion of depth has been extended to functional data (Nieto-Reyes and Battey 2016) and is referred to as functional depth. The functional depth based outlier detection method of Febrero-Bande, Galeano, and Gonzãlez-Manteiga (2008) allows us to detect not only residual curves with large integrated squared error, but also *shape outliers* (Hubert, Rousseeuw, and Segaert 2015). The latter are curves that do not appear unusual if viewed at each time point but are abnormal across the entire trajectory. Shape outliers can be missed if we use methods that rely solely on integrated square error.

There are a number of depth functions for functional data that can be chosen. We will use the h-modal depth (Cuevas, Febrero, and Fraiman 2007) to rank samples r_i, as it works well in practice. For a given kernel G_h (typically Gaussian with bandwidth h), the h-modal depth of r_i with respect to $\boldsymbol{r} = (r_1, \ldots, r_n)$ is given by

$$D(r_i|\boldsymbol{r}, h) = \frac{1}{n}\sum_{j=1}^{n} G\left(\frac{\|r_i - r_j\|}{h}\right). \tag{13}$$

The h-modal depth has two useful properties. First, it uses a distance metric therefore the depth values will be proportional to the distance from the "center." Second, it does not assume a unique center and therefore it is appropriate when there are more than one "mode" in the data cloud.

In the algorithm, we need to choose the bandwidth h and a threshold C to identify outliers. The bandwidth h is taken to be the 15th percentile of the empirical distribution of $\{\|r_i - r_j\|, i, j = 1, \ldots, n\}$ (Febrero-Bande, Galeano, and Gonzãlez-Manteiga 2008). The threshold C is chosen such that $P(D(r_i|\boldsymbol{r}, h) \leq C) = \delta$, where δ is a pre-chosen percentile. To estimate the threshold C they use a bootstrapping approach, which estimates a value of C for different random sets of samples and then aggregates these estimates. We describe the outlier detection algorithm in Algorithm 2.

Data: Centered curves $\{x_i(t), y_i(t)\}$ for $i = 1, \ldots, n$ and percentile δ.
1. Use Algorithm 1 to obtain $\tilde{\phi}_k^Y(t)$, $\tilde{z}_m$, and $\tilde{\boldsymbol{B}}$.
2. Calculate residual curves $r_i(t)$.
3. Estimate bandwidth h.
4. For each $r_i(t)$ calculate $D(r_i|\boldsymbol{r}, h)$.
5. Estimate C for given percentile δ.
6. If $D(r_i|\boldsymbol{r}, h) < C$ sample i is an outlier.

Algorithm 2: Outlier detection using RFLR.

6. Simulation Study

In this section, we will provide a simulation study to investigate the finite sample properties of RBIC and RFLR in comparison to BIC and classical FLR (CFLR). Code to reproduce the experimental results reported in this section can be found at the GitHub repository (*https://github.com/hullait/RobustFLR*). In the simulation study, we will generate data using a FLR process and corrupt a certain number of samples, which will be the outliers. The outliers have been designed to be undetectable, if the response curves are considered independently of the predictor curves. Therefore, standard functional data outlier detection algorithms such as those discussed in Section 5 will perform poorly.

The main motivation for the RFLR model is to obtain good model fitting in the presence of outliers. In this simulation study, we compare the fitting error (FE) given in (14), for the non-outlier samples using the robust model, which uses RFLR and RBIC with the classical approach using CFLR and BIC. We define the indicator variable $u_i = 1$ if sample i is an outlier and zero otherwise. Letting $\hat{y}_i(t)$ be the estimation of $y_i(t)$ and given that proportion a of the samples have been contaminated then FE is given by

$$\text{FE} = \frac{1}{(1-a)n}\sum_{i=1}^{n}(1-u_i)||y_i - \hat{y}_i||^2. \tag{14}$$

Next we compare the outlier detection capabilities of robust and classical approaches using the receiver operating characteristic (ROC) curve to determine the sensitivity/specificity trade-off for different thresholds. If we have perfect outlier detection for all thresholds then the area under the curve (AUC) of the ROC curve would be one. We can therefore use the AUC value as a measure of outlier detection accuracy regardless of threshold.

FPCA is performed by taking the principal components of a cubic B-spline representation of each of the predictor and response curves (Ramsay and Silverman 2005). The robust FPCA approach outlined in Section 3 is performed using the CR algorithm proposed by Croux and Ruiz-Gazen (1996) on the same B-spline coefficients. The MLTS estimator is calculated using the heuristic given by Agulló, Croux, and Van Aelst (2008) using different trimming proportions $(1-\alpha)$ for $\alpha \in [0,1]$.

6.1. Scenarios

We will generate samples $x(t)$ using a FPCA based model with mean function $\mu_X(t) = -10(t-0.5)^2 + 2$ for $t \in [0,1]$ and eigenfunctions:

$$\phi_1^X = \sqrt{2}\sin(\pi t),\ \ \phi_2^X = \sqrt{2}\sin(7\pi t),\ \ \phi_3^X = \sqrt{2}\cos(7\pi t).$$

The principal scores are sampled from Gaussian distributions with mean 0 and variances 40, 10, and 1 for the eigenfunctions, respectively. Note that we do not create any outliers in the FPCA decompositions of the predictor curves. We generate 400 predictor curves $x_1(t),\ldots,x_{400}(t)$, which are observed at $T = 500$ equidistant points in the interval $[0,1]$.

The samples $y(t)$ will have eigenfunctions:

$$\phi_1^Y = \sqrt{2}\sin(12\pi t),\ \ \phi_2^Y = \sqrt{2}\sin(5\pi t),\ \ \phi_3^Y = \sqrt{2}\cos(2\pi t),$$

and mean function $\mu_Y(t) = 60\exp(-(t-1)^2)$. We will generate $\beta(s,t) = \boldsymbol{\phi}^{\boldsymbol{X}}(s)^T\boldsymbol{B}\boldsymbol{\phi}^{\boldsymbol{Y}}(t)$ where $\boldsymbol{B}$ will have random entries between $[-3,3]$. We generate non-outlier curves:

$$y_i(t) = \mu_Y(t) + \int_I \beta(s,t)(x_i(s) - \mu_X(s))ds + \epsilon_i(t),$$

where the residual function $\epsilon_i(t) = \boldsymbol{q}_i^T\boldsymbol{\phi}^{\boldsymbol{Y}}(t) + d_i$ where $\boldsymbol{q}_i^T$ and d_i are sampled iid from $N(0,0.1)$. We will consider three cases when the proportion of outliers are $a = 0.1, 0.2$ and 0.3.

In Scenario 1, outliers will be generated by replacing $\boldsymbol{B}$ with $\boldsymbol{B}_1 = \boldsymbol{B}+\boldsymbol{R}$ where $\boldsymbol{R}$ has random entries sampled from $N(0,0.5)$ giving $\beta_1(s,t) = \boldsymbol{\phi}^{\boldsymbol{X}}(s)^T\boldsymbol{B}_1\boldsymbol{\phi}^{\boldsymbol{Y}}(t)$. Outliers $y_i'(t)$ are given by

$$y_i'(t) = \mu_Y(t) + \int_I \beta_1(s,t)(x_i(s) - \mu_X(s))ds + \epsilon_i(t).$$

In Scenario 2, we generate outliers by adding a random B-spline function $p(t)$ defined on an interval of length $1/10$. Letting $\beta_2(s,t) = \boldsymbol{\phi}^{\boldsymbol{X}}(s)^T\boldsymbol{B}_2[\boldsymbol{\phi}^{\boldsymbol{Y}}(t), p(t)]$, for 3×4 matrix $\boldsymbol{B}_2 = [\boldsymbol{B}, \boldsymbol{l}]$ for $\boldsymbol{l} \sim N(2,1)$, then the outliers $y_i''(t)$ are given by

$$y_i''(t) = \mu_Y(t) + \int_I \beta_2(s,t)(x_i(s) - \mu_X(s))ds + \epsilon_i(t).$$

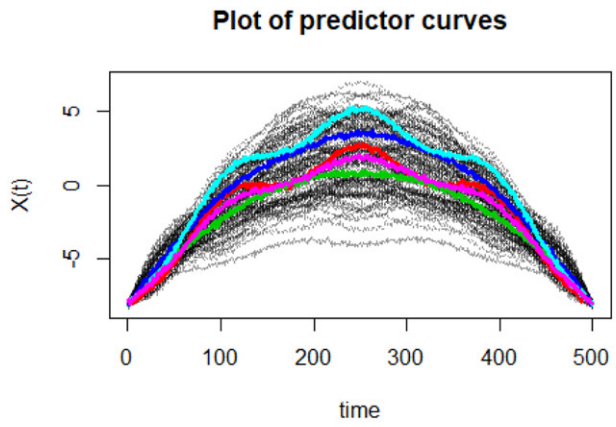
Note that the outliers in Scenario 1 affect the regression function across the entire interval whereas the outliers in Scenario 2 only affect a small interval of the curves.

In Figure 3, we have a plot of the predictor curves $x_i(t)$ and response curves $y_i(t)$ with outliers from Scenario 1 and Scenario 2. The figure shows the outliers are masked by the variability in the curves and therefore cannot by identified using standard outlier detection algorithms. To make the outliers clearer, we have plotted the residuals of the response curves using the true regression function and mean functions. In the bottom row of Figure 3, we can see that the outliers in Scenario 2 are localized to a fixed interval whereas in Scenario 1 the outliers affect the response curve at all time points.

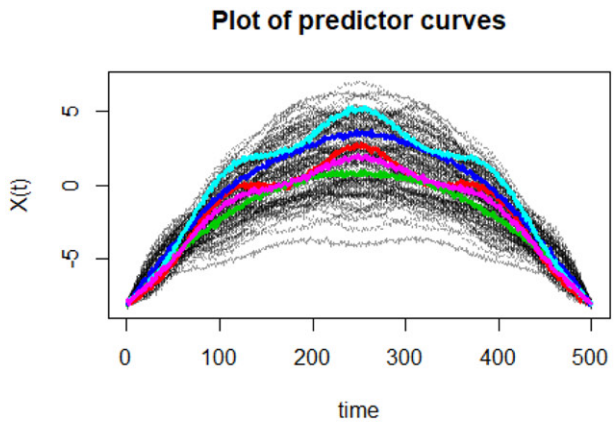
The RFLR model depends on the proportion of trimming α. To investigate the effect of the trimming we will consider trimming proportions $\alpha = 0.1, 0.2$, and 0.3. We shall also investigate the performance using BIC and RBIC with fixed trimmed sample size of $r = [0.8n]$.

We sample 400 predictor and response curve datasets and generate classical and robust models to calculate the average FE (14). In Tables 1 and 2, we present the results for Scenarios 1 and 2, respectively. The CFLR model gives a smaller FE value in the case of no-outliers $a = 0$, however, the robust model still gives good model fits. If we compare the FE using BIC and RBIC, we can see that BIC gives better model choices when $a = 0$. This is due to BIC using all the data and in particular using samples in the tails of the distribution. In the presence of outliers, the robust model outperforms the classical model, and as expected the difference in FE increases as the number of outliers increases. We should also note that RBIC is giving better model choices than BIC when outliers are present. Next, we can see using trimming proportion $\alpha = 0.1$ we obtain significantly large FE values when $a = 0.3$. However, the FE values for $\alpha = 0.2$ and 0.3 are very similar in the case of $a = 0.3$. The outliers generated can have different sizes, therefore in the $\alpha = 0.2$ robust model only small outliers are present, which only affect the model fitting slightly .

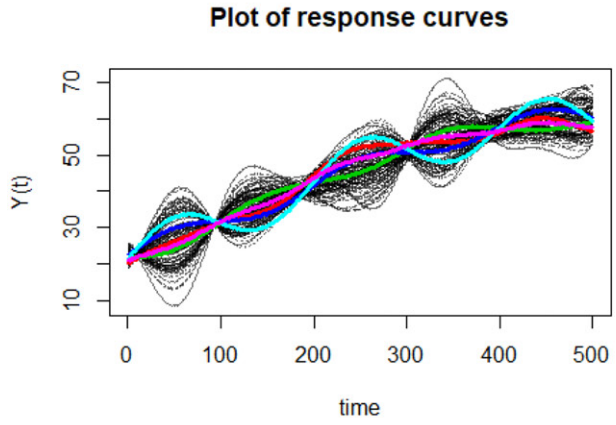
In Figure 4, we have two ROC curves generated for one of the repetitions in Scenarios 1 and 2 in which we have contaminated 20% of the samples. In both scenarios, the robust model outperforms the classical model. We also deploy the approach of Febrero-Bande, Galeano, and Gonzãlez-Manteiga (2008) to the response curves, disregarding the predictor curves (henceforth called the Direct approach). The ROC curves show that the robust and classical models are more effective than the Direct method in identifying the outliers in Scenarios 1 and 2. By only using the specificity and sensitivity for a fixed threshold a lot of information is being lost, therefore, a better comparison would be the area under the curve (AUC). Using the AUC metric, we can understand the model outlier detection capabilities overall, in particular how well are the outliers separated from the other samples. We have taken the average AUC values over the 100 iterations performed for Scenario 1, which are shown in Table 3. We have considered the average AUC values for trimming levels $\alpha = 0.1, 0.2$, and 0.3. The robust models give larger AUC values than the classical model. However, the different trimming levels do not seem to have a significant effect on the AUC values. In Scenario 2, we have the results in Table 4. The same patterns appear as in Scenario 1 except the AUC values are notably smaller. This is to be expected given the outliers in Scenario 2 are defined on a small time interval.



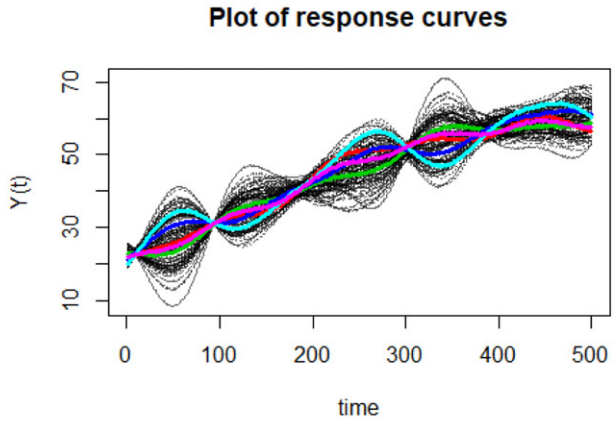
(a) $x_i(t)$



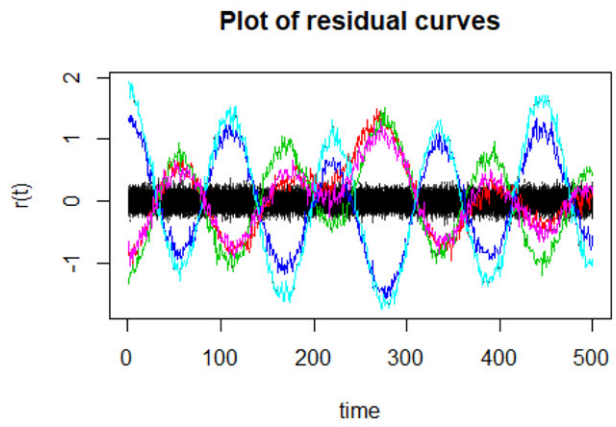
(b) $x_i(t)$



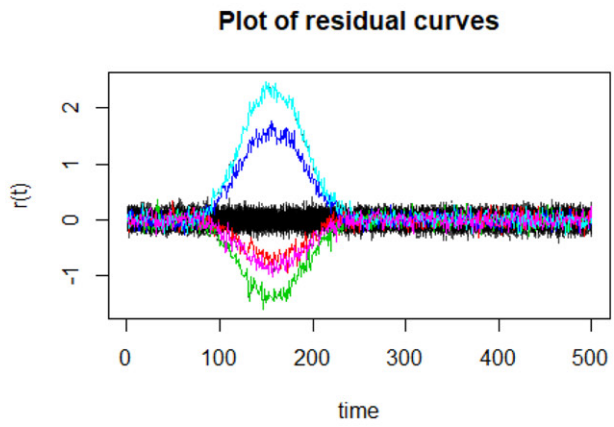
(c) $y_i^{(1)}(t)$



(d) $y_i^{(2)}(t)$



(e) $r_i^{(1)}(t)$



(f) $r_i^{(2)}(t)$

Figure 3. *Left:* Plots of the predictor curves $x_i(t)$, response curves $y_i^{(1)}(t)$ and residuals curves $r_i^{(1)}(t)$ for Scenario 1. *Right:* Plots of the predictor curves $x_i(t)$, response curves $y_i^{(2)}(t)$ and residuals curves $r_i^{(2)}(t)$ for Scenario 2. The residual curves are generated using the true regression function and mean functions. In each scenario there are 5 outliers each in a distinctive color. The predictors curves $x_i(t)$ are identical for both scenarios, and the response curves look very similar due to mean and functional components masking the outliers. However, the residuals are clearly distinctive.

Table 1. Average fitting errors (FE) for 100 replications for Scenario 1, using classic FPCA and robust FPCA with different amount of trimming in the MLTS estimator and using models selected by BIC and RBIC.

	Trim	Model	$a = 0$	$a = 0.1$	$a = 0.2$	$a = 0.3$
Classic	$\alpha = 0.0$	BIC	**5.326**	18.441	48.771	101.320
Robust	$\alpha = 0.1$	BIC	8.283	14.166	21.118	33.907
	$\alpha = 0.1$	RBIC	9.285	**9.179**	10.674	28.393
	$\alpha = 0.2$	BIC	8.288	14.178	15.750	16.623
	$\alpha = 0.2$	RBIC	9.292	9.207	**9.535**	13.436
	$\alpha = 0.3$	BIC	8.294	14.199	15.815	16.518
	$\alpha = 0.3$	RBIC	9.301	9.214	9.544	**12.334**

NOTE: The bold values are the minimum values in each row.

Table 2. Average fitting errors (FE) for 100 replications for Scenario 2, using classic FPCA and robust FPCA with different amount of trimming in the MLTS estimator and using models selected by BIC and RBIC.

	Trim	Model	$a = 0$	$a = 0.1$	$a = 0.2$	$a = 0.3$
Classic	$\alpha = 0.0$	BIC	**5.326**	17.252	48.906	85.063
Robust	$\alpha = 0.1$	BIC	8.283	15.242	21.524	28.758
	$\alpha = 0.1$	RBIC	9.285	**9.074**	9.919	18.546
	$\alpha = 0.2$	BIC	8.288	16.745	20.652	21.928
	$\alpha = 0.2$	RBIC	9.292	9.191	**8.997**	13.628
	$\alpha = 0.3$	BIC	8.294	16.808	20.695	21.750
	$\alpha = 0.3$	RBIC	9.301	9.233	9.018	**11.439**

NOTE: The bold values are the minimum values in each row.

Table 3. Average AUC values over 100 replications for Scenario 1, using proportion of outliers $a = 0.1, 0.2$, and 0.3.

		$a = 0.1$	$a = 0.2$	$a = 0.3$
Direct	–	0.532	0.538	0.550
Classic	$\alpha = 0.0$	0.960	0.898	0.797
Robust	$\alpha = 0.1$	0.995	0.991	0.953
	$\alpha = 0.2$	**0.996**	**0.996**	0.987
	$\alpha = 0.3$	**0.996**	**0.996**	**0.990**

NOTE: Using Direct compared to classic FPCA with BIC, and using robust FPCA with RBIC and trimming levels $\alpha = 0.1, 0.2$, and 0.3. The bold values are the minimum values in each row.

Table 4. Average AUC values over 100 replications for Scenario 2, using proportion of outliers $a = 0.1, 0.2$, and 0.3.

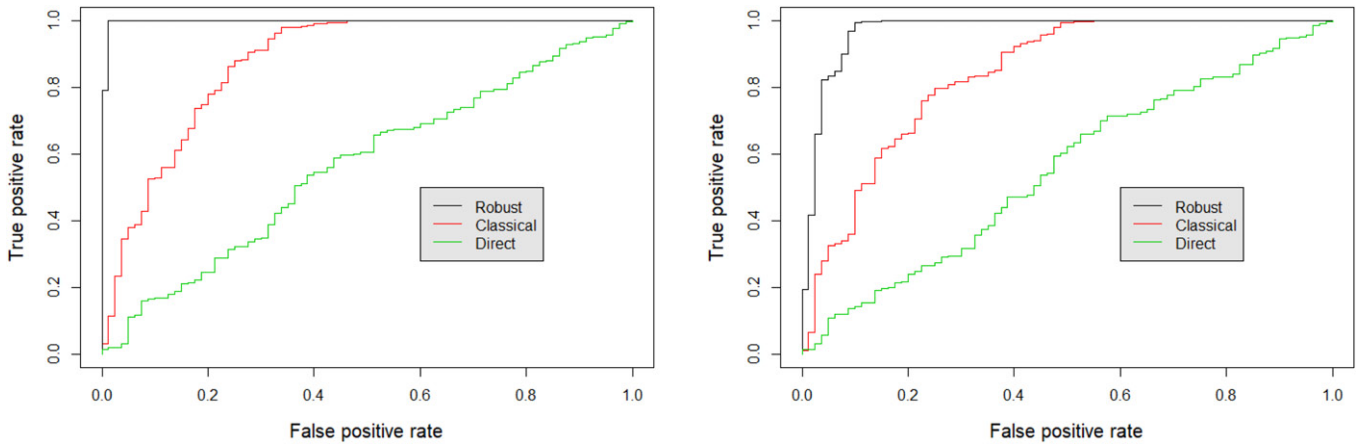
		$a = 0.1$	$a = 0.2$	$a = 0.3$
Direct	–	0.512	0.548	0.554
Classic	$\alpha = 0.0$	0.922	0.838	0.734
Robust	$\alpha = 0.1$	**0.985**	0.964	0.932
	$\alpha = 0.2$	0.980	**0.980**	0.966
	$\alpha = 0.3$	0.980	**0.980**	**0.968**

NOTE: Using Direct compared to classic FPCA with BIC, and using robust FPCA with RBIC and trimming levels $\alpha = 0.1, 0.2$, and 0.3. The bold values are the minimum values in each row.

7. Jet Engine data

The jet engine dataset contains sensor measurements taken during 199 VS maneuvers. These maneuvers have been taken from multiple engine tests. To minimize the variability in the data we have ensured that the tests were performed on the same engine type and at the same test facility as environmental factors can affect engine parameters during the test. The VS maneuver has a distinctive engine speed profile with a slow acceleration followed by a slow deceleration, as shown in Figure 1(a). The TPR is the engineer-controlled measure of engine thrust. As shown in Figure 1 engine temperature parameters like the TGT follow a broadly similar pattern to TPR, during the acceleration phase. As expected, the temperature parameters reach their maximum value with some delay compared to TPR and although they decrease during the deceleration phase they are considerably higher than their initial values at the end of the maneuver. This highlights the trajectory-dependent behavior that we seek to model.

We do not have labels for whether any of the individual engines have outliers but we do have log books from the engine test, from which we can obtain insights into the VS maneuvers which our method flags as outliers. There are a number of temperature features measured within an engine including the TGT, discussed previously. In addition, we have four other temperature readings T25 (the temperature reading at sensor station 25), T30 (the temperature reading at sensor station 30),



(a) Scenario 1 (b) Scenario 2

Figure 4. ROC curve for one instance of Scenarios 1 and 2 with the proportion of outlier $a = 0.2$ and proportion trimmed $\alpha = 0.2$.

TCAR (temperature of casing at rear), and TCAF (temperature of casing at front). All the temperature features are shown in Figure 7. The TCAR is particularly interesting as it has two distinct curve behaviors. The VS maneuvers time series are of similar length. To standardize we have fitted a B-spline basis of 400 basis functions (with equally spaced knots) to each to ensure the time series are well approximated. We have experimented with different number of basis functions in the range between 200 and 500 and our results were not affected by this choice. A lower dimensional fit can be achieved with a smarter placement of knots, but we avoided this option to minimize the degree of manual tuning required. We have taken 1000 equally spaced points on the B-spline representations to be our inputs $x_i(t)$ and $y_i(t)$. Note that if all time series are of equal length we could use standard PCA. However, as shown by Ramsay and Silverman (2005) this can lead to noisy estimates of the principal components, which can negatively affect the subsequent model fitting.

We first estimate an RFLR model for each temperature parameter, through the procedure outlined in Algorithm 1. As specified in this algorithm, the number of eigenfunctions to represent the predictor and response curves, M and K, respectively, is determined through RBIC. In all cases M and K range between three and five.

We can then apply the outlier detection algorithm described in Algorithm 2. We compare these outliers with those detected on the temperature curves directly and using CFLR and BIC in Algorithm 2. We can look at the residuals curves to verify whether the outliers appear to be abnormal. In particular, we want to show that using functional regression we are able to determine outliers that would otherwise be missed by investigating the temperature curves directly.

Using the depth based outlier detection (Direct) (Febrero-Bande, Galeano, and González-Manteiga 2008) directly on the temperature curves (with a default threshold of $\delta = 0.01$), we obtain the outliers in Table 5. We can see that the outliers in the TPR are the same as the outliers in the temperature features. This suggests the outliers being identified are arising from the controller induced variability. We therefore need to model the dependency between the control feature (TPR) and the temperature features.

We applied the outlier detection algorithm given in Algorithm 2 using CFLR and BIC with threshold $\delta = 0.01$. The outliers identified are given in Table 5. The residuals curves are shown in Figures 5 and 6, with the outliers colored in blue. It is not clear from this plot that the outliers are truly different from the other data.

Lastly we applied Algorithm 2 using RFLR and RBIC with threshold $\delta = 0.01$. The outlier samples are given in Table 5 for each temperature feature. In Figures 5 and 6, we have the residual curves using RFLR. We can see that the RFLR model fits the majority of the temperature curves well. The outliers that are picked up clearly look abnormal, with significant deviations from the general behavior. The RFLR model is therefore able to identify interesting behavior, which may otherwise have been undetected. Engineers have informed us that Sample 24 comes from an engine in which they detected damaged hardware. All the other outliers in the RFLR column of Table 5 were also noted to come from engines that displayed odd behavior during the pass-off test. This is not the case for the outliers reported in the CFLR column.

Table 5. Outliers detected for temperature features (Temp) using outlier detection on the temperature features directly (Direct), and the outliers found using CFLR and RFLR.

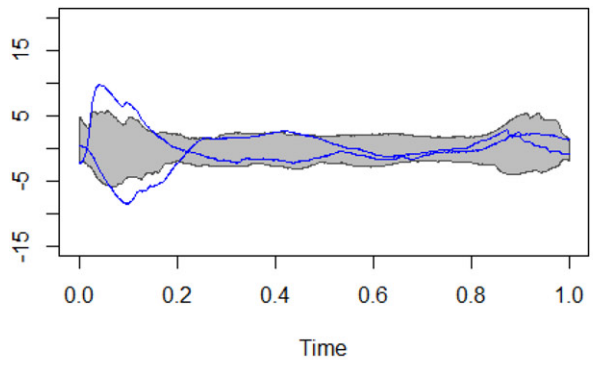
Temp	Direct	CFLR	RFLR
TPR	33, 106, 167	–	–
T25	33, 106, 167	24, 182	24, 70, 106
T30	33, 106, 167	24, 182, 192	24, 44, 70, 106, 196
TGT	33, 106, 167	119, 153	44, 70, 106, 117
TCAR	33, 106	36, 91, 106	70, 106
TCAF	33, 167	65, 167, 170, 171	24, 70, 106

In Figure 7, we have a plot of the temperature parameters with the outliers identified using the curves directly in green, those using the RFLR model in red and those detected by both in purple. We can see that the outliers from the RFLR model do not necessarily appear as abnormal if we look at the temperature curves directly. Sample 106 is identified as an outlier by multiple temperature features and also when the depth based outlier detection is used on the temperature curves directly. Comparing the outliers identified using a classical approach, we can see Sample 24 is identified as an outlier multiple times using the classical and robust approaches. However, most of the outliers from the classical approaches differ with the outliers identified using the robust approach. We can also see that the outliers using the RFLR are significantly more distinctive than the outliers using CFLR.

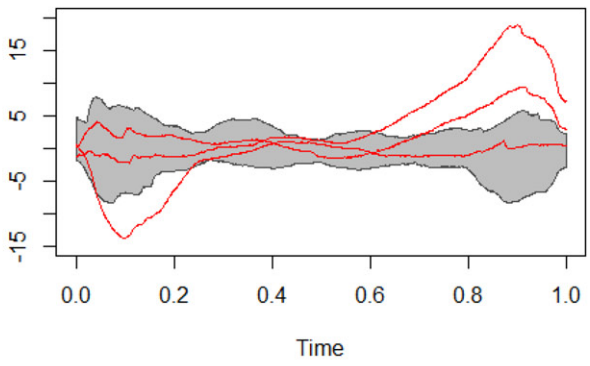
The RFLR model produces a regression function $\tilde{\beta}(s,t)$, which models the relationship between the engine speed and engine temperature. In Figure 8, we depict two indicative regression surfaces for the temperature parameters T25 and TGT. Note that the number of eigenfunctions used to represent the response and the predictor curves, K and M, respectively, affects the estimated $\tilde{\beta}(s,t)$. For both regression functions in Figure 8 large in absolute value entries appear on or close to the diagonal, which corresponds to $s = t$. However, not all the weight is on the diagonal, and in particular it is commonly the case that $\tilde{\beta}(s,t)$ is nonzero for $s > t$. This suggests that using information from the current and past time-points only ($s \leqslant t$) can substantially limit the accuracy of an FLR model. FLR models that are restricted to not use information from future time-points are called *historical FLR* models (Malfait and Ramsay 2003). We developed such models and indeed found that they fit the data less well. A further limitation of this approach is that no robust historical FLR models exist and developing one is challenging. One major difficultly is that the basis functions required for the historical FLR model are not necessarily orthogonal.

8. Conclusion

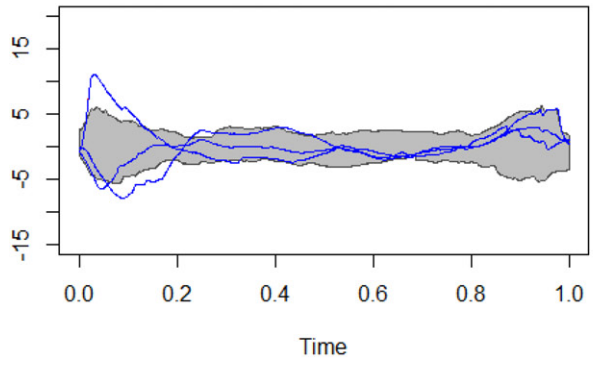
There exist a number of functional regression models for functional inputs and responses, however, these methods are not robust to outliers. We have introduced an RFLR model that is able to produce good model fits in the presence of outliers. Alongside the RFLR model, we have also introduced a robust model selection procedure and proven the consistency of the RFLR and model selection procedure. Using a simulation study



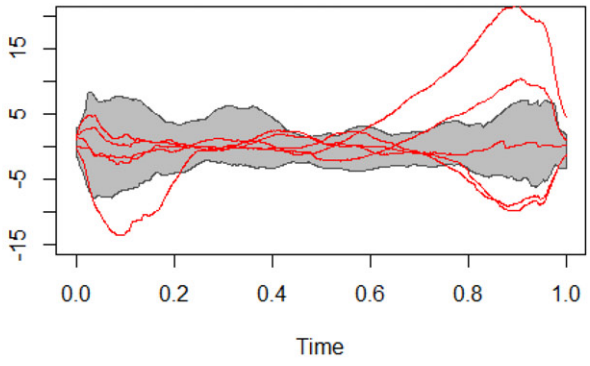
(a) Residuals of T25 through CFLR



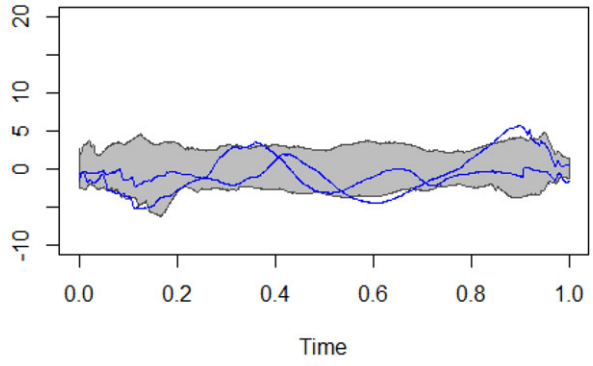
(b) Residuals of T25 through RFLR



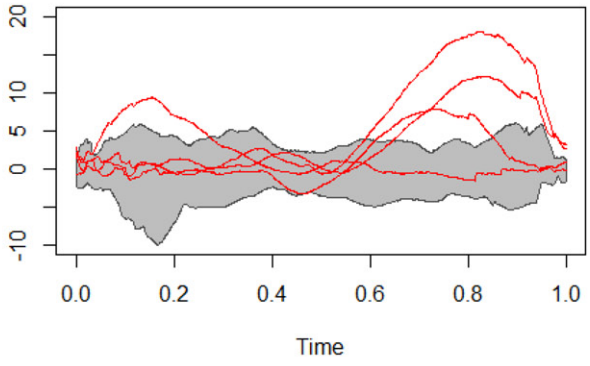
(c) Residuals of T30 through CFLR



(d) Residuals of T30 through RFLR

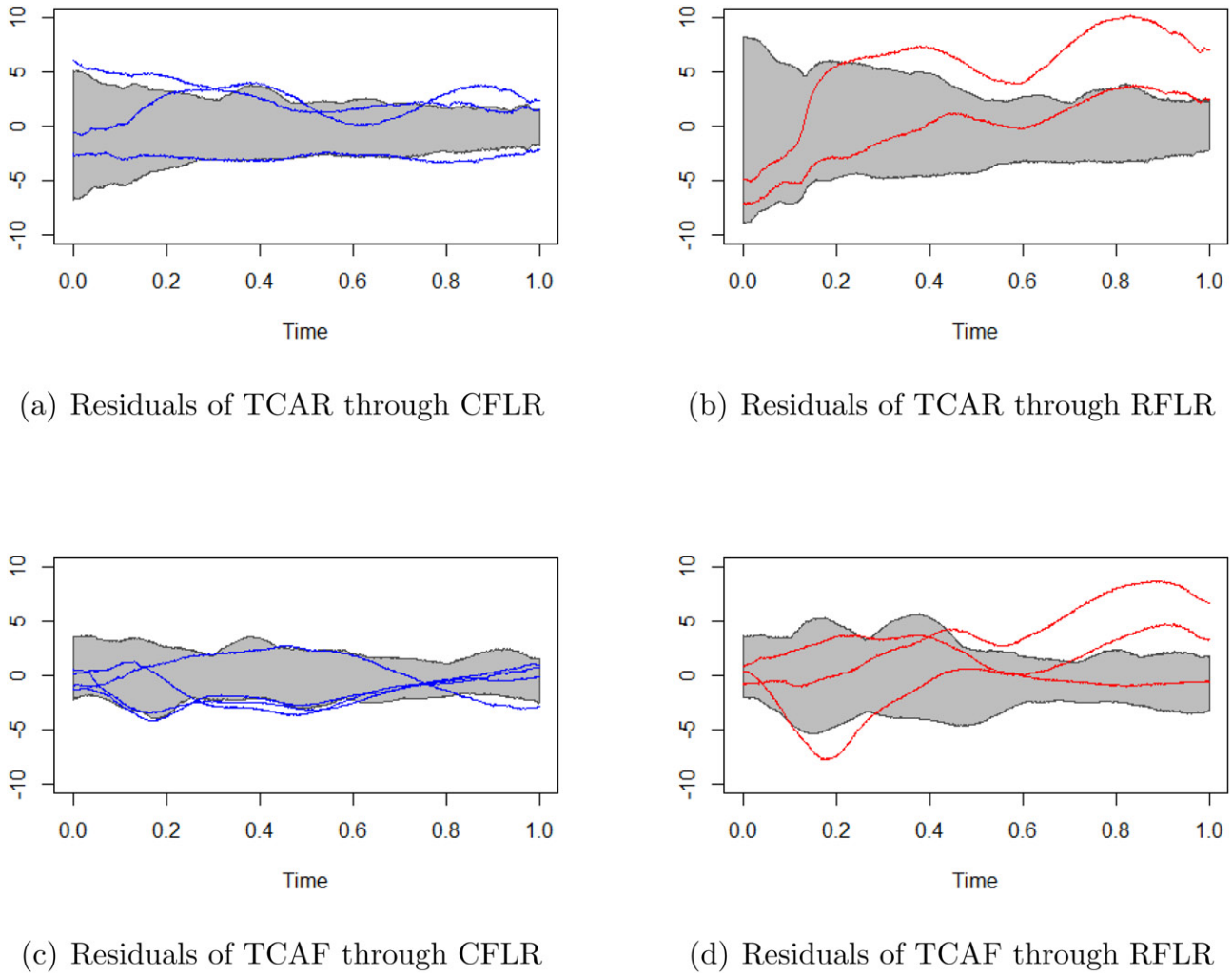


(e) Residuals of TGT through of CFLR



(f) Residuals of TGT through RFLR

Figure 5. Plots of the residual curves for temperature at stations 25 and 30, T25 and T30, respectively, and turbine gas temperature (TGT). Nonoutliers are in the gray region while outliers are depicted in blue and red for the CFLR and RFLR models, respectively. Figures on the left column are obtained through CFLR while those on the right through RFLR.



(a) Residuals of TCAR through CFLR

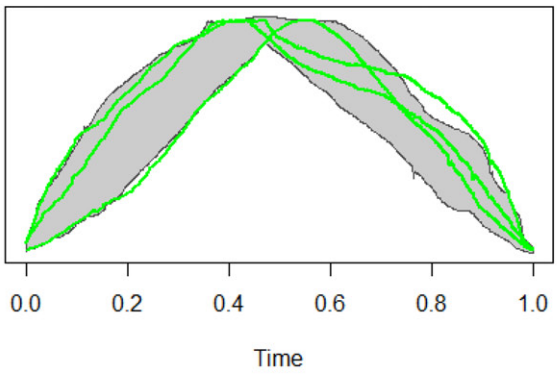
(b) Residuals of TCAR through RFLR

(c) Residuals of TCAF through CFLR

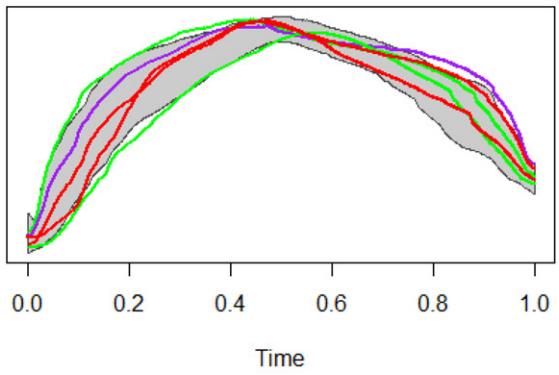
(d) Residuals of TCAF through RFLR

Figure 6. Plots of residuals curves for temperature of casing at rear and at Front, TCAR and TCAF, respectively. Nonoutliers are in the gray region and outliers are depicted in blue and red for the CFLR and RFLR models, respectively. Figures on the left column are obtained through CFLR while those on the right through RFLR.

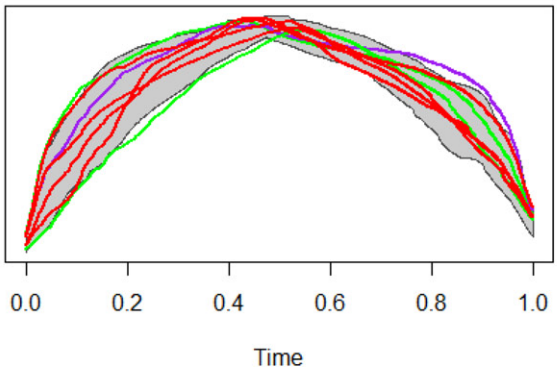
we have shown the need for a robust approach to obtain good models in the presence of outliers. The RFLR model is also effective in identifying global and localized outliers. Finally using jet engine sensor data as a motivating application for RFLR, we have identified unusual temperature behavior. In particular, the outliers identified in the jet engine sensor data would not have been detected if we modeled the response variables independently of human controlled driving variable.



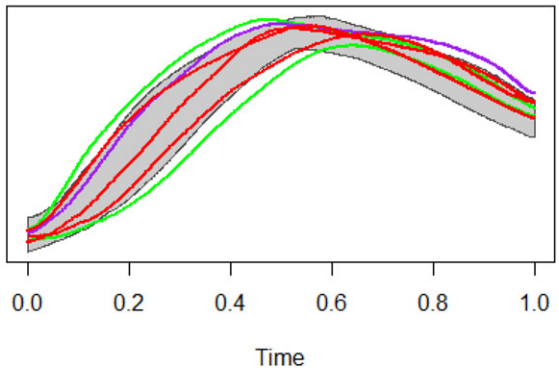
(a) Turbine Pressure Ratio



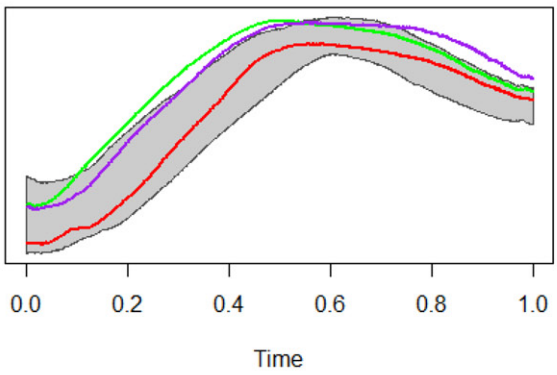
(b) Temperature at station 25



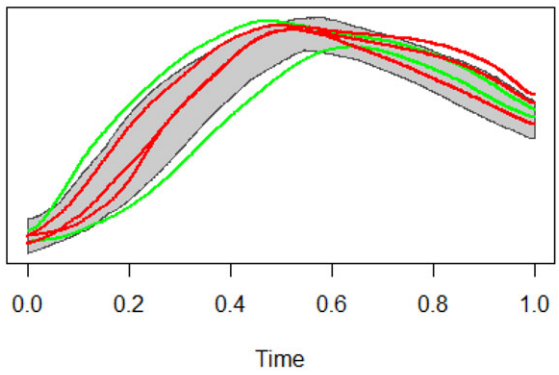
(c) Temperature at station 30



(d) Turbine Gas Temperature

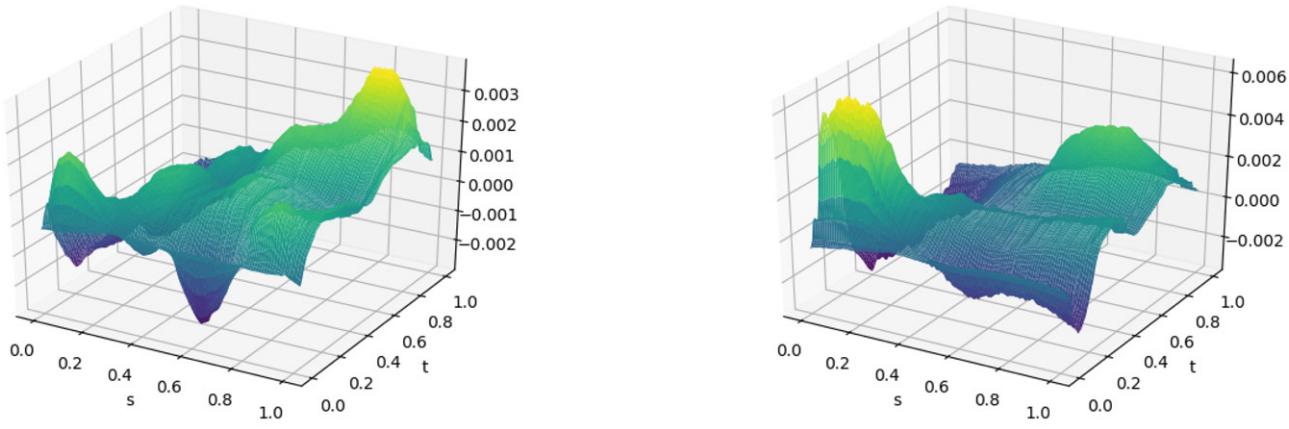


(e) Temperature of Casing at Rear



(f) Temperature of Casing At Front

Figure 7. Plots of temperature time series with non-outliers in gray region and outliers using robust FLR in red; those using the curves directly in green and those for both in purple.



(a) T25 (b) TGT

Figure 8. Plot of the regression functions $\tilde{\beta}(s,t)$ for T25 and TGT parameters.

Supplementary Materials

Supplementary material: A PDF containing proofs of the results in Section 4.

RobFLR: R code used for simulation study available at github/hullait/RobFLR

Funding

This work was supported by EPSRC grant number EP/L015692/1 (STOR-i). The authors also acknowledge Rolls Royce plc (RR) for financial support.

References

Agulló, J., Croux, C., and Van Aelst, S. (2008), "The Multivariate Least-Trimmed Squares Estimator," *Journal of Multivariate Analysis*, 99, 311–338. [396,400,402]

Arribas-Gil, A., and Romo, J. (2014), "Shape Outlier Detection and Visualization for Functional Data: The Outliergram," *Biostatistics*, 15, 603–619. [401]

Bali, J. L., Boente, G., Tyler, D. E., and Wang, J.-L. (2011), "Robust Functional Principal Components: A Projection-Pursuit Approach," *The Annals of Statistics*, 39, 2852–2882. [396,399,400]

Chiou, J.-M., Yang, Y.-F., and Chen, Y.-T. (2016), "Multivariate Functional Linear Regression and Prediction," *Journal of Multivariate Analysis*, 146, 301–312. [398]

Croux, C., and Ruiz-Gazen, A. (1996), "A Fast Algorithm for Robust Principal Components Based on Projection Pursuit," in *COMPSTAT*, ed. A. Prat, Physica-Verlag HD, pp. 211–216. [402]

Cuevas, A., Febrero, M., and Fraiman, R. (2007), "Robust Estimation and Classification for Functional Data via Projection-Based Depth Notions," *Computational Statistics*, 22, 481–496. [401]

Dai, W., and Genton, M. G. (2018), "Multivariate Functional Data Visualization and Outlier Detection," *Journal of Computational and Graphical Statistics*, 27, 923–934. [401]

Eilers, P. H. C., and Marx, B. D. (1996), "Flexible Smoothing With B-Splines and Penalties," *Statistical Science*, 11, 89–121. [399]

Febrero-Bande, M., Galeano, P., and Gonzãlez-Manteiga, W. (2008), "Outlier Detection in Functional Data by Depth Measures, With Application to Identify Abnormal NO_x Levels," *Environmetrics*, 19, 331–345. [401,402,405]

Ferraty, F., Keilegom, I. V., and Vieu, P. (2012), "Regression When Both Response and Predictor Are Functions," *Journal of Multivariate Analysis*, 109, 10–28. [398]

Houstis, E., Catlin, A., Tsompanopoulou, P., Gottfried, D., Balakrishnan, G., Su, K., and Rice, J. (2002), "GasTurbnLab: A Multidisciplinary Problem Solving Environment for Gas Turbine Engine Design on a Network of Nonhomogeneous Machines," *Journal of Computational and Applied Mathematics*, 149, 83–100. [396]

Hubert, M., Rousseeuw, P. J., and Segaert, P. (2015), "Multivariate Functional Outlier Detection," *Statistical Methods & Applications*, 24, 177–202. [401]

Ivanescu, A. E., Staicu, A.-M., Scheipl, F., and Greven, S. (2015), "Penalized Function-on-Function Regression," *Computational Statistics*, 30, 539–568. [398]

Machado, J. A. F. (1993), "Robust Model Selection and M-Estimation," *Econometric Theory*, 9, 478–493. [400]

Malfait, N., and Ramsay, J. O. (2003), "The Historical Functional Linear Model," *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 31, 115–128. [405]

Matsui, H. (2020), "Quadratic Regression for Functional Response Models," *Econometrics and Statistics*, 13, 125–136. [399]

Morris, J. S. (2015), "Functional Regression," *Annual Review of Statistics and Its Application*, 2, 321–359. [397]

Nieto-Reyes, A., and Battey, H. (2016, 02), "A Topologically Valid Definition of Depth for Functional Data," *Statistical Science*, 31, 61–79. [401]

Pollard, D. (2012), *Convergence of Stochastic Processes*, New York: Springer. [400]

Ramsay, J. O., and Dalzell, C. J. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society*, Series B, 53, 539–572. [396,398]

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, Springer Series in Statistics, New York: Springer. [397,398,402,405]

Scheipl, F., Staicu, A.-M., and Greven, S. (2015), "Functional Additive Mixed Models," *Journal of Computational and Graphical Statistics*, 24, 477–501. [398]

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [400]

Shang, H. L. (2014), "A Survey of Functional Principal Component Analysis," *AStA Advances in Statistical Analysis*, 98, 121–142. [398]

Silverman, B. W. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," *The Annals of Statistics*, 24, 1–24. [400]

Sun, Y., and Genton, M. G. (2011), "Functional Boxplots," *Journal of Computational and Graphical Statistics*, 20, 316–334. [401]

Tukey, J. W. (1975), "Mathematics and the Picturing of Data," in *International Congress of Mathematicians* (Vol. 2), ed. R. D. James, pp. 523–532. [401]

Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903. [398,399]