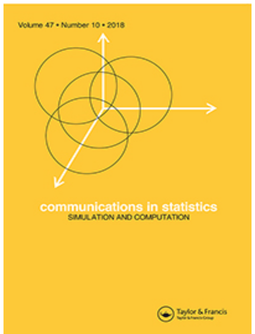

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Robust sparse functional regression model

Jasdeep Pannu & Nedret Billor

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2020.1767784

Taylor & Francis
Taylor & Francis Group

Check for updates

REVIEW ARTICLE

Robust sparse functional regression model

Jasdeep Pannu[a] and Nedret Billor[b]

[a]Mathematics and Statistics, California State University Sacramento, Sacramento, California, USA;
[b]Mathematics and Statistics, Auburn University, Auburn, Alabama, USA

ABSTRACT
The presence of outliers, in general, affects the performance of the conventional statistical methods which require the homogeneity of observations. In this study, we consider variable selection problem in a functional regression model when a functional dataset contains outliers. We propose a functional adaptive group LASSO variable selection method based on the weighted least absolute deviation which takes into account the effect of outliers in both x and y directions for a functional regression model with a scalar response and multiple functional predictors. Further, we demonstrate, through simulated and real datasets, that the proposed methods perform well.

ARTICLE HISTORY
Received 9 October 2018
Accepted 5 May 2020

KEYWORDS
Functional outliers; Functional regression model; LASSO; Least absolute deviation; Robustness; Variable selection

MATHEMATICS SUBJECT CLASSIFICATION
62 Statistics

1. Introduction

Recently, variable selection has become one of the most important steps in a regression model due to the rapid rise in the production of complex and large datasets. As developments in functional data analysis have also gained momentum with the rapid data collection techniques due to the advancement in computer technology, variable selection has even become more important for a functional regression model due to the inherent high dimensionality of the parameter functions. Therefore, it is inevitable to handle the high dimensionality problem by using variable selection to get simple and interpretable models with high prediction accuracy. There have been proposed several variable selection methods based on penalized least squares idea, such as least absolute shrinkage and selection operator (LASSO) (Tibshirani 1996; Tibshirani et al. 2005; Zou and Hastie 2005; Yuan and Lin 2006) for a functional linear regression model. However, these conventional variable selection methods that use the quadratic risk function is sensitive to outliers in a functional dataset, thus would give us incorrect model with low prediction accuracy. In this study, we consider the problem of variable selection for functional regression model in the presence of outliers. Recently, a few robust methods have been proposed to handle outliers to obtain more reliable models in functional regression framework (Arribas-Gil and Romo 2013; Hubert, Rousseeuw, and Segaert 2015; Aneiros and Vieu 2016; Berrendero, Cuevas, and Torrecilla 2016; Huang and Sun 2016; Matsui 2017; Dai and Genton 2019; Picheny, Servien, and Villa-Vialaneix 2019). Recently,

CONTACT Jasdeep Pannu j.pannu@csus.edu Mathematics and Statistics, California State University Sacramento, Shasta Hall, Sacramento, CA 95819-6088, USA.

Pannu and Billor (2017) have proposed a robust functional variable selection based on group LASSO and least absolute deviation method (*LAD-gLASSO*) in the presence of outliers in y direction. They have shown that the *LAD-gLASSO* performs better than the Gertheiss, Maity, and Staicu (2013) classical variable selection method, functional *groupLASSO* in the presence of outliers in response. Although this method works well in the presence of some type of outliers it has several limitations. The first limitation is that since the amount of contribution for each functional predictor on the response may vary penalizing all regression coefficient functions with the same penalty term is not correct. The second limitation is that *LAD-gLASSO* method takes into account only the outliers in the response, Y, not the outliers in functional predictor, $X(t)$. The functional outliers in a functional predictor as well as in response may cause the most serious issues for the estimators of the model parameters and the predicted model. Therefore, we propose a method that handles these limitations. For the first limitation, we propose an adaptive version (*LAD-agLASSO*) of *LAD-gLASSO* to obtain more accurate model when there are outliers in response, Y. To improve the model resulted from the *LAD-gLASSO* method, we suggest to use varying penalties for each regression coefficient function. To overcome the second limitation of functional *LAD-gLASSO* of being sensitive to outliers in functional predictor, $X(t)$, we propose a weighted version of functional *LAD-agLASSO* called functional *Weighted LAD-adaptive groupLASSO* (*WLAD-agLASSO*) that takes into account the effect of outliers in both functional outliers, $X(t)$ and response, Y, for functional regression model. Weighted *LAD* regression estimation has been proposed by Ellis and Morgenthaler (1992), Hubert and Rousseeuw (1997), and Giloni, Simonoff, and Sengupta (2006) to deal with outliers in predictors for ordinary multiple regression model. Arslan (2012) has proposed *Weighted LAD-LASSO* (*WLAD-LASSO*) as a robust variable selection method to handle the issue of outliers in response and explanatory variables for ordinary multiple regression model. But, to our knowledge, no such method exists for functional regression model. Our proposed method functional *WLAD-agLASSO* is not only resistant to outliers in the response variable, but also minimizes the effect of outliers in functional predictors (leverage curves), by introducing weights which are dependent on the functional predictors only. These weights are introduced to downweight the leverage curves and thus reducing their effect on the estimation process.

This article is organized as follows. In Sec. 2, we provide the methodologies for functional *LAD-agLASSO* and functional *WLAD-agLASSO*, respectively. Simulation studies are presented extensively to assess the performances of our proposed methods in Sec. 3. In Sec. 4, a real data application is considered. Finally, a summary and conclusion are given in Sec. 5.

2. Methodology

We consider a functional linear model with the scalar response and p functional predictors:

$$Y_i = \alpha + \Sigma_{j=1}^{p} \int_{\mathcal{T}_I} X_{ij}(t)\beta_j(t)dt + \epsilon_i, \quad i = 1, \ldots, N, \tag{1}$$

where the random error terms ϵ_i are assumed to be independent normally distributed with mean 0 and variance σ^2. The functional predictors $X(t)$ are assumed to have mean function equal to zero. α is a scalar parameter, $\beta_j(t)$ is a parameter function for $j = 1, \ldots, p,$ and $\mathcal{T}_I \subset \Re \to \Re$ is the domain of integration.

We apply the same method described in Gertheiss, Maity, and Staicu (2013) to the model in (1), to overcome the inherent infinite dimensionality problem and reformulate it as an ordinary multiple regression model by approximating the parameter functions, $\beta_j(t)$ with l number of pre-basis, $\phi(t),$ as

$$\beta_j(t) = \Sigma_{b=1}^{l} c_{jb}\phi_{jb}(t), \tag{2}$$

and approximating the integral in (1) by Riemann sum with

$$\int X_{ij}(t)\beta_j(t)dt \approx \Sigma_m X_{ij}(t_m)\beta_j(t_m). \tag{3}$$

Then the model in (1) can be written as:

$$Y_i = \alpha + \Sigma_{j=1}^{p} \mathbf{\Phi_{ij}^T c_j} + \epsilon_i, \tag{4}$$

where $\mathbf{c_j} = (c_{j1}, \ldots, c_{jl})^T, \mathbf{\Phi_{ij}} = (\Phi_{ij1}, \ldots, \Phi_{ijl})^T, \Phi_{ijb} = \delta_j \Sigma_m X_{ij}(t_{jm})\phi_{jb}(t_{jm})$ and $\delta_j = t_{jm} - t_{j,m-1}$ for $i = 1, \ldots, N$ and $j = 1, \ldots, p$. In the new model in (4), $\mathbf{\Phi_{ij}}$ are known and α and $\mathbf{c_j}$'s are the unknown regression coefficients that need to be estimated.

As discussed in Pannu and Billor (2017), one of the main assumptions in functional regression model is that data should be homogeneous, which is free of outliers which is almost never true in real life. Therefore, it is desirable to develop robust statistical methods that are resistant to outliers in functional data. Pannu and Billor (2017) proposed a robust functional variable selection method, functional *LAD-groupLASSO* (*LAD-gLASSO*) by using the least absolute deviation and the same penalty function as suggested in Gertheiss, Maity, and Staicu (2013), in the presence of outliers.

$$\Sigma_{i=1}^{n} |Y_i - \alpha - \Sigma_{j=1}^{p} \mathbf{\Phi_{ij}}^T \mathbf{c_j}| + P_{\lambda,\varphi}(\beta_j), \tag{5}$$

where $P_{\lambda,\varphi}(\beta_j)$ is the penalty function as introduced by Meier, Van de Geer, and Bühlmann (2009) and used by Gertheiss, Maity, and Staicu (2013) for functional variable selection. Specifically,

$$P_{\lambda,\varphi}(\beta_j) = \lambda(||\beta_j||_2^2 + \varphi||\beta_j''||_2^2)^{1/2}, \tag{6}$$

where $||.||_2^2 = \int (.)^2 dt$ is the L^2 norm and β_j'' is the second derivative of β_j. An adequate combination of λ (shrinkage parameter) and φ (smoothing parameter) has to be chosen to avoid over shrinkage of coefficients and very wiggly hard to interpret solutions. *K*-fold (5 or 10) cross-validation is one of the ways to select this combination in which the prediction error of the model given by the sum of squared errors $\Sigma_i(Y_i - \widehat{Y}_i)^2$ is minimized.

As shown in Pannu and Billor (2017), functional *LAD-gLASSO* outperforms the classical group LASSO method, functional *gLASSO* proposed by Gertheiss, Maity, and Staicu (2013) in the presence of outliers in the response, but has two limitations as mentioned in Sec. 1. First limitation is that it applies same amount of shrinkage to all of the regression coefficients and hence is not consistent in terms of model selection (Fan and Li 2001). Efficiency can also suffer due to one shrinkage parameter

(Zou 2006). As a result, we introduce an adaptive tuning parameter which assigns a different tuning parameter for each coefficient, allowing the shrinkage to vary from coefficient to coefficient in Sec. 2.1. To address the second limitation of this method being highly sensitive to outliers in the functional predictors we introduce a weighted version of *LAD* in Sec. 2.2.

2.1. Functional LAD-adaptive groupLASSO

The functional *LAD-gLASSO* method proposed by Pannu and Billor (2017) imposes same penalty on all the coefficient functions. We consider a penalty that is adaptive in nature to allow for different shrinkage and smoothness for the different covariates as introduced by Zou (2006) and used by Gertheiss, Maity, and Staicu (2013). We call this method functional *LAD-adaptive groupLASSO* (*LAD-agLASSO*) which produces the estimators minimizing the objective function

$$\Sigma_{i=1}^{n}|Y_i - \alpha - \Sigma_{j=1}^{p}\mathbf{\Phi_{ij}}^T\mathbf{c_j}| + P_{\lambda,\varphi}(\beta_j), \tag{7}$$

for the regression parameters in a functional regression model given in (1).

Here $P_{\lambda,\varphi}(\beta_j)$ is now defined as,

$$P_{\lambda,\varphi}(\beta_j) = \lambda(\kappa_j||\beta_j||_2^2 + \nu_j\varphi||\beta_j''||_2^2)^{1/2}, \tag{8}$$

where $\|.\|_2^2 = \int(.)^2dt$ is the L^2 norm, β_j'' is the second derivative of β_j, κ_j, and ν_j are the data-adaptive weights. The choice of weights κ_j and ν_j is meant to reflect some subjectivity about the true parameter functions. We use initial parameters estimates (based on smoothing solely) to choose these weights. Let $\ddot{\beta}_j{}'$ s, be the initial estimates of the coefficient functions β_j' s, using for example, quantile regression implemented in the R package *quantreg* (Koenker 2013). Then, the adaptive weights can be defined as $\kappa_j = 1/\parallel\ddot{\beta}_j\parallel$ Z$\kappa_j = 1/\parallel\beta_j^{\cdot\cdot}\parallel$ and $\nu_j = 1/\parallel\ddot{\beta}_j{}''\parallel$. The penalty function $P_{\lambda,\varphi}(\beta_j)$ in (8), can be decomposed using Cholesky decomposition as discussed by Gertheiss, Maity, and Staicu (2013):

$$P_{\lambda,\varphi}(\beta_j) = \lambda(\mathbf{c_j}^T(L_{\varphi,j}L_{\varphi,j}^T)\mathbf{c_j})^{1/2}, \tag{9}$$

where $L_{\varphi,j}$ is nonsingular lower triangular matrix. Now using (9) the model in (7) reduces to the following:

$$\Sigma_{i=1}^{n}|Y_i - \alpha - \tilde{\Phi}_{ij}^T\tilde{\mathbf{c}}_\mathbf{j}| + \lambda\Sigma_{j=1}^{p}||\tilde{\mathbf{c}}_\mathbf{j}||, \tag{10}$$

where $\tilde{\mathbf{c}}_\mathbf{j} = L_{\varphi,j}^T\mathbf{c_j}$ and $\tilde{\Phi}_{ij} = L_{\varphi,j}^{-1}\Phi_{ij}$. Now $\hat{\alpha}$ and $\widehat{\mathbf{c_j}}$‘s are the minimizers of (10) and the coefficient function $\beta(t)$ is estimated by $\hat{\beta}_j(t) = \Sigma_{b=1}^{l}\phi_{jb}(t)\widehat{c_{jb}}$ for $j = 1, ..., p$.

Although functional *LAD-agLASSO* aims to improve the model selection compared to functional *LAD-gLASSO*, being based on simple *LAD* gives better results only when outliers are present in the response variable. In the presence of outliers in functional predictors, it is not robust. In the next section, we propose an approach called functional *Weighted LAD-adaptive groupLASSO* (*WLAD-agLASSO*) to handle this issue.

2.2. Functional weighted LAD-adaptive groupLASSO

In this section, we propose functional *Weighted LAD-adaptive groupLASSO* (*WLAD-agLASSO*) to minimize the effect of outliers present not only in the y response but also in the functional predictors on the model selection. Reconsider the objective function for functional *LAD-agLASSO* in (10) and introduce weights w_i to the function which are determined by a robust measure of predictors and are chosen to downweight the leverage points, that is, outliers in the functional predictors. We call this new objective function as functional *WLAD-agLASSO*:

$$\Sigma_{i=1}^{n} w_i |Y_i - \alpha - \tilde{\Phi}_{ij}^T \tilde{\mathbf{c}}_{\mathbf{j}}| + \lambda \Sigma_{j=1}^{p} ||\tilde{\mathbf{c}}_{\mathbf{j}}||. \tag{11}$$

The weights w_i in (11) are obtained using the robust distances of the functional predictors so that the outlying observations in the functional predictors will have large distances and the corresponding weights will be small. Therefore, it is expected that the resulting regression estimator will be robust against the outliers in the response variable and leverage curves. The weights are computed using the weight definition given in Hubert and Rousseeuw (1997). Specifically, the algorithm to find the weights for a multivariate dataset is as following:

1. Calculate the robust location and scatter estimates, $\tilde{\mu}$ and $\tilde{\Sigma}$ for the location vector and the scatter matrix of the data $\boldsymbol{x}_1, \boldsymbol{x}_2,$ ta $\boldsymbol{x}_n \in \Re^p$. One can use high breakdown point location and scatter estimators such as *MCD* (minimum covariance determinant). The idea behind *MCD* is to find observations whose empirical covariance matrix has the smallest determinant, yielding a pure subset of observations from which to compute standards estimates of location and covariance. The *MCD* estimator has been introduced by Rousseeuw (1984). The implementation in R package *rrcov* (Todorov 2009) uses the fast *MCD* algorithm of Rousseeuw and Driessen (1999) to approximate the *MCD* estimator.
2. Compute the robust distances: $RD(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \tilde{\mu})^T \tilde{\Sigma}^{-1} (\boldsymbol{x}_i - \tilde{\mu})$.
3. Calculate the weights $w_i = min\left\{1, \frac{p}{RD(\boldsymbol{x}_i)}\right\}$ for $i = 1, ..., n$.

We will adapt this algorithm to the model in (4). Next, we will conduct an extensive simulation study and apply the proposed methods to real datasets in Sec. 4.

3. Simulation study

In this section, we assess the performance of the proposed methods through simulation study. We consider performing simulation study in two parts. The first part uses toy example and the second part focuses on 50 simulation runs. The evaluation of the methods is done through some metrics. In both parts of the simulation study, we consider the following three models:

- **Model (1):** The presence of outliers only in the scalar response Y.
- **Model (2):** The presence of outliers in the scalar response Y and the functional predictors $X(t)$.
- **Model (3):** Clean functional data, that is no outliers exist in the functional data.

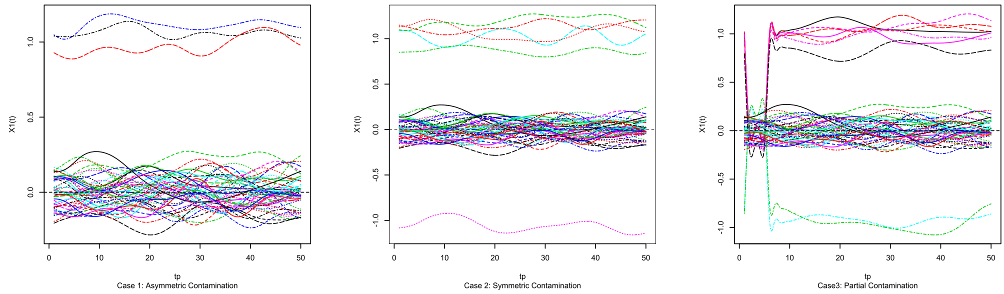

Figure 1. The contaminated $X_1(t)$ curves for contamination cases 1–3 ($q = 15\%$).

The data are generated and contaminated the same way as discussed in Pannu and Billor (2017).

Generating Functional Predictors $X(t)$:

The functional covariates $X(t)$ are generated similarly as in Tutz and Gertheiss (2010) from:

$$X_{ij}(t) = [\sigma(t)]^{-1}\Sigma_{r=1}^{5}(a_{ijr}\, sin\,(\pi t(5 - a_{ijr})/150) - m_{ijr}), \tag{12}$$

where $i = 1, \dots, n, j = 1, \dots, p, a_{ijr} \sim U(0, 5)$, $m_{ijr} \sim U(0, 2 * \pi)$ and $\sigma(t)$ is defined so that $var[X_{ij}(t)] = 0.01$.

Contamination of Y:

To create outliers in response Y, we use the standard normal distribution, the t-distribution with 2 degrees of freedom, and the t-distribution with 7 degrees of freedom for the errors ϵ.

Contamination of $X(t)$: The contamination of functional predictors $X(t)$ is done as described by Fraiman and Muniz (2001). The following three cases of contamination are considered:

- **Case (1):** Asymmetric contamination $Z_j(t) = X_j(t) + cM$ where c is 1 with probability q and 0 with probability 1-q and $q = \{0\%; 5\%; 10\%; 15\%; 20\%\}$; M is the contamination constant size equal to 10 and $X_j(t)$ is as defined in (12).
- **Case (2):** Symmetric contamination $Z_j(t) = X_j(t) + c\sigma M$ where $X_j(t)$, c and M are as defined before and σ is a sequence of random variables independent of c that takes the values 1 and -1 with probability 0.5.
- **Case (3):** Partial contamination $Z_j(t) = X_j(t) + c\sigma M$ if $t > T$ and $Z_j(t) = X_j(t)$ if $t < T$, where $T \sim U[0, 10]$.

The effects of these different types of contamination on $X_1(t)$ at 15% level are shown in Figure 1.

We consider several contamination levels (0%, 15%, 25%, and 40%) of the data. We present the results for 0% and 15% contamination levels. In addition, we provide results for 25% contamination level of Model (1). We also provide information based on 40% contamination levels and comment on empirical breakdown point based on this simulation study.

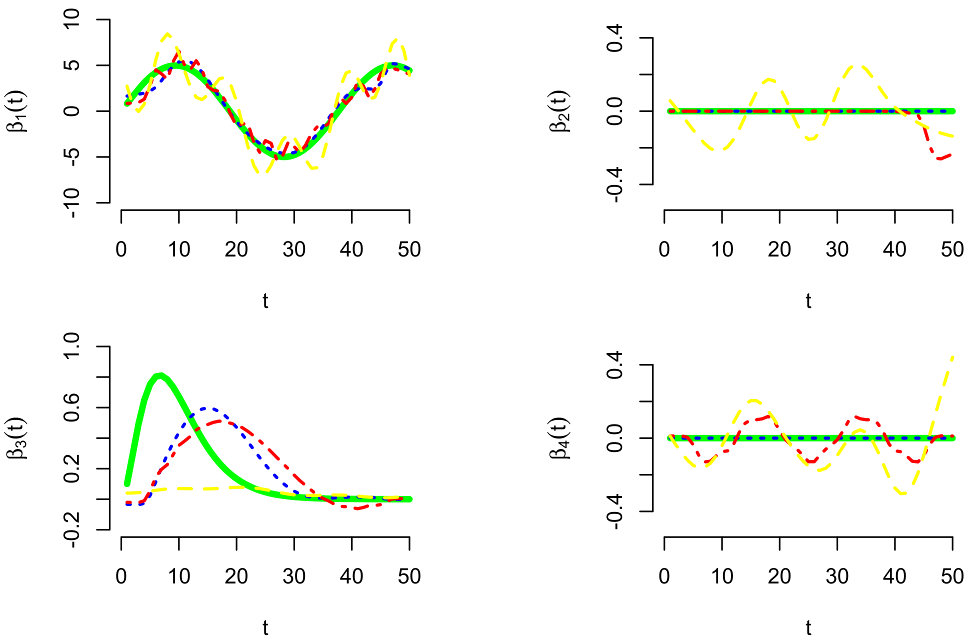

Figure 2. Fitting results of true beta functions (green) using robust functional *LAD-agLASSO* (blue) ($\lambda = 10$, $\varphi = 10$), robust functional *LAD-gLASSO* (red) ($\lambda = 10$, $\varphi = 10^2$), and classical functional *agLASSO* (yellow) ($\lambda = 10^2$, $\varphi = 10^3$) at 15% contamination of Y for Model (1).

Simulation Study Part I

The results for three model scenarios Model (1), Model (2), and Model (3) are as following.

Model (1): The presence of outliers only in the scalar response *Y*.

We consider 50 replications of each of the four functional covariates $X(t)$ observed at 50 equidistant points in (0, 50) and the contamination level of Y is 15%.

The true model is given as

$$Y_i = \alpha + \int_0^{50} \beta_1(t) X_{i1}(t) dt + \int_0^{50} \beta_3(t) X_{i3}(t) dt + \epsilon_i, \tag{13}$$

where $i = 1, \ldots, 50$. In order to create outliers in response Y, the errors are generated from the standard normal distribution, the t-distribution with two degrees of freedom, and the t-distribution with 7 degrees of freedom. The contamination is done for 15% of Y. The model is set up where the response is related only to $X_1(t)$ and $X_3(t)$. Functional predictors $X_2(t)$ and $X_4(t)$ are essentially zero functions and should be excluded from the model by the applied method. The coefficient function $\beta_1(t)$ and $\beta_3(t)$ are sine and gamma function, respectively as shown in Figure 2.

Results for Model (1): We compare our proposed method robust functional *LAD-agLASSO* with robust functional *LAD-gLASSO* (Pannu and Billor 2017) and classical functional *agLASSO* (Gertheiss, Maity, and Staicu 2013). We use function *rq.fit.lasso ()* from the *R* packages *quantreg* (Koenker 2013) to execute robust functional *LAD-agLASSO*. With the green curves representing the true beta functions, we observe that the robust functional *LAD-agLASSO* (blue) performs better than both robust functional

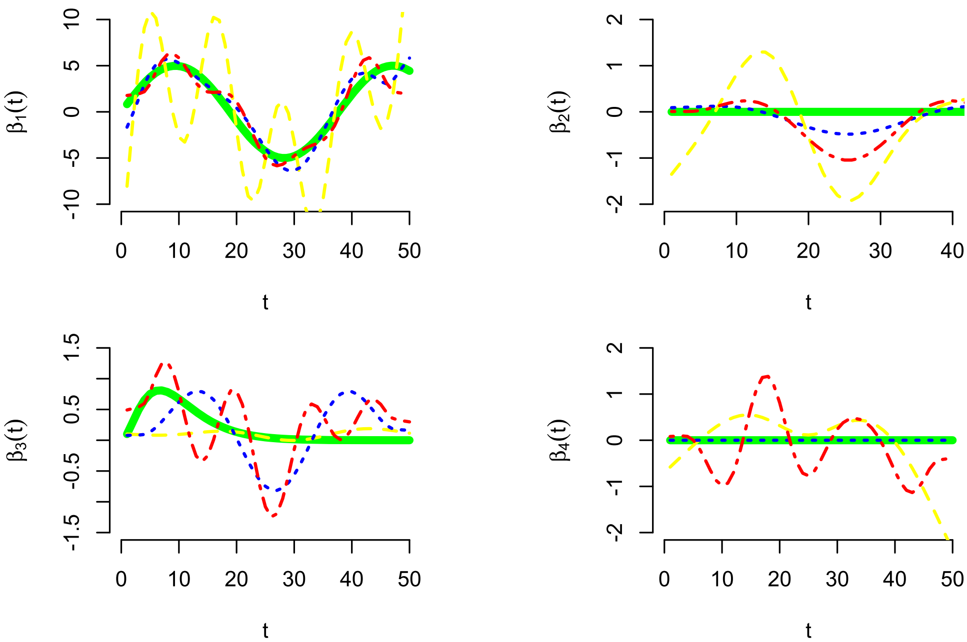

Figure 3. Fitting results of true beta functions (green) using robust functional *LAD-agLASSO* (blue) ($\lambda = 10^2, \varphi = 10$), robust functional *LAD-gLASSO* (red) ($\lambda = 10^2$, $\varphi = 10^3$), and classical functional *agLASSO* (yellow) ($\lambda = 10^2$, $\varphi = 10^4$) at 25% contamination of Y for Model (1).

LAD-gLASSO (red) and classical functional *agLASSO* (black) (Figure 2). Robust functional *LAD-gLASSO* is giving one false positive by not excluding $X_4(t)$ from the model and classical functional *agLASSO* performs worse among all in terms of estimation and selection of coefficients.

Furthermore, Figure 3 shows the comparison of these methods at 25% contamination level of Y. We can see that our proposed method robust functional *LAD-agLASSO* still works better than the other compared methods. Additionally, we perform the comparison at 40% contamination level of Y and find that robust functional *LAD-agLASSO* breaks down empirically.

Model (2): The presence of outliers in the scalar response Y and the functional predictors $X(t)$.

For this model, we consider 100 curves for each of four functional covariates $X(t)$ at 50 equidistant points in (0, 50) and the contamination level is 15% for both the response Y and the functional predictors $X(t)$.

The true model is assumed to be as

$$Y_i = \alpha + \int_0^{50} \beta_1(t)X_{i1}(t)dt + \int_0^{50} \beta_3(t)X_{i3}(t)dt + \epsilon_i, \tag{14}$$

where $i = 1, \ldots 100$ and $\epsilon_i \sim N(0,4)$. The shapes of parameter functions $\beta_1(t)$ through $\beta_4(t)$ are shown in Figure 4 by green curves. The model is set up where the response is related only to $X_1(t)$ and $X_3(t)$. The functional predictors $X(t)$ are contaminated as discussed above using three cases of contamination that is, Case 1 (Asymmetric Contamination), Case 2 (Symmetric Contamination) and Case 3 (Partial Contamination). We use function

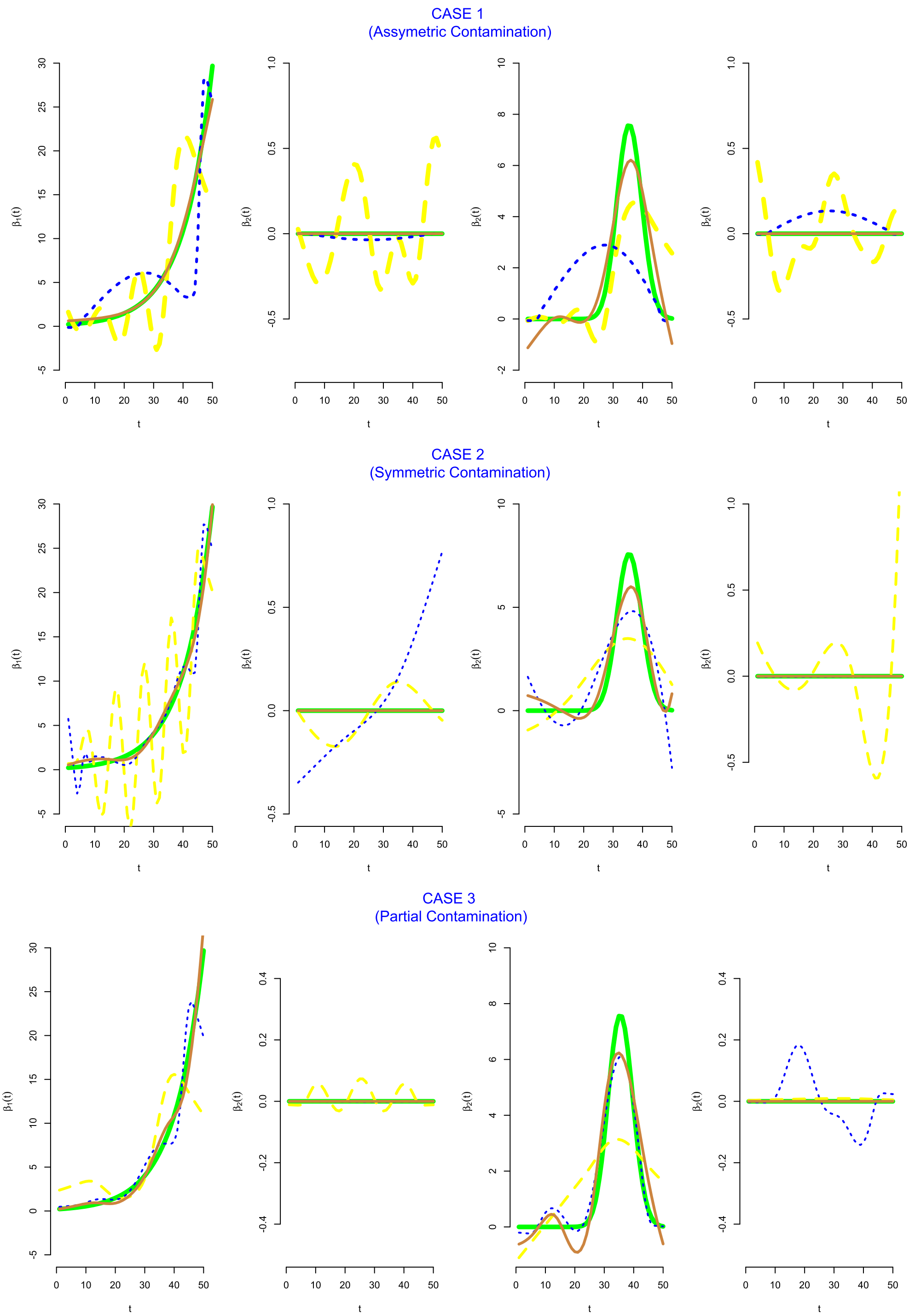

Figure 4. Fitting results of true beta functions (green) using robust functional *WLAD-agLASSO* (tan) ($\lambda = 10$, $\varphi = 10^2$), robust functional *LAD-agLASSO* (blue) ($\lambda = 10^2$, $\varphi = 10^3$), and classical functional *agLASSO* (yellow) ($\lambda = 10^3$, $\varphi = 10^3$) at 15% contamination of both Y and $X(t)$ for Model (2).

CovMcd () from the *R* package *rrcov* (Todorov 2009) to execute functional *WLAD-agLASSO* since this method requires the robust distances to calculate the weights, w_i, as described in Sec. 2.2.

Results for Model (2): To compare the classical functional *agLASSO* (black) (Gertheiss, Maity, and Staicu 2013) method with robust functional *LAD-agLASSO* (blue) and robust functional *WLAD-agLASSO* (tan) methods, we computed the estimated beta functions for every method and cases (1) through (3) and displayed them in Figure 4. In this figure, we clearly observe that the classical functional *agLASSO* (black) (Gertheiss, Maity, and Staicu 2013) method performs the worst than the robust methods. Further, robust functional *WLAD-agLASSO* (tan) based estimated beta functions (which mainly work the best in the presence of outliers in functional predictors as well as the response variable) are estimated much more accurately than the robust functional *LAD-agLASSO* (blue) based beta functions. This indicates that introducing weights which are functions of robust distances to handle outliers in functional predictors improves the estimation and selection of the $\beta(t)$ functions. Moreover, we investigated the performance of robust functional *WLAD-agLASSO* at 25% and 40% contamination levels of both response Y and the functional predictors $X(t)$. We observe that robust functional *WLAD-agLASSO* still performs better than the classical functional *agLASSO* (Gertheiss, Maity, and Staicu 2013) and the robust functional *LAD-agLASSO* at 25% contamination level but breaks down empirically at 40% contamination level.

Model (3): Clean functional data.

We generate 100 replications for each of $X_1(t), X_2(t)$ and $X_3(t)$ at 50 equidistant time points in (0, 50).

The true model is assumed to be

$$Y_i = \alpha + \int_0^{50} \beta_1(t) X_{i1}(t)dt + \int_0^{50} \beta_3(t) X_{i3}(t)dt + \epsilon_i, \tag{15}$$

where $i = 1, \ldots, 100$ and $\epsilon_i \sim$ N(0,1). The parameter functions $\beta_1(t)$ and $\beta_3(t)$ have an exponential function shape and square root function shape, respectively as represented by green curves in Figure 5. The parameter function $\beta_2(t)$ is essentially zero. The model is set up where the response is related only to $X_1(t)$ and $X_3(t)$.

Results for Model (3): We compare the methods we propose in this paper with the classical functional *agLASSO* (Gertheiss, Maity, and Staicu 2013). Figure 5 shows that all methods, classical functional *agLASSO* (black), robust functional *LAD-agLASSO* (blue), and robust functional *WLAD-agLASSO* (tan), perform equally well in terms of estimation and selection of coefficients in the absence of outliers. This indicates that the newly proposed robust methods are consistent when data are clean.

Simulation Study Part II

In the second part of this study, we focus on the assessment of the performance of the proposed methods in terms of sparseness accuracy and prediction performance for Models (1) through (3). Therefore, we conducted extensive simulations in a variety of settings. The data are generated and contaminated the same way as described above. We consider 10 functional predictors and generate 100 sample curves for each functional predictor $X(t)$ observed at 100 equidistant time points. The shapes of the

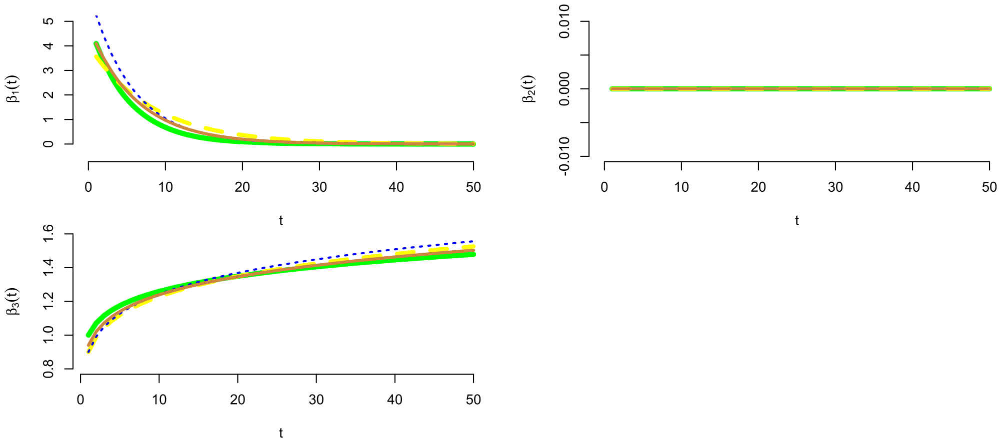

Figure 5. Fitting results of true beta functions (green) using robust functional *WLAD-agLASSO* (tan) ($\lambda = 10$, $\varphi = 10$), robust functional *LAD-agLASSO* (blue) ($\lambda = 10$, $\varphi = 10^2$), and classical functional *agLASSO* (yellow) ($\lambda = 10$, $\varphi = 10^2$) at 0% contamination for Model (3).

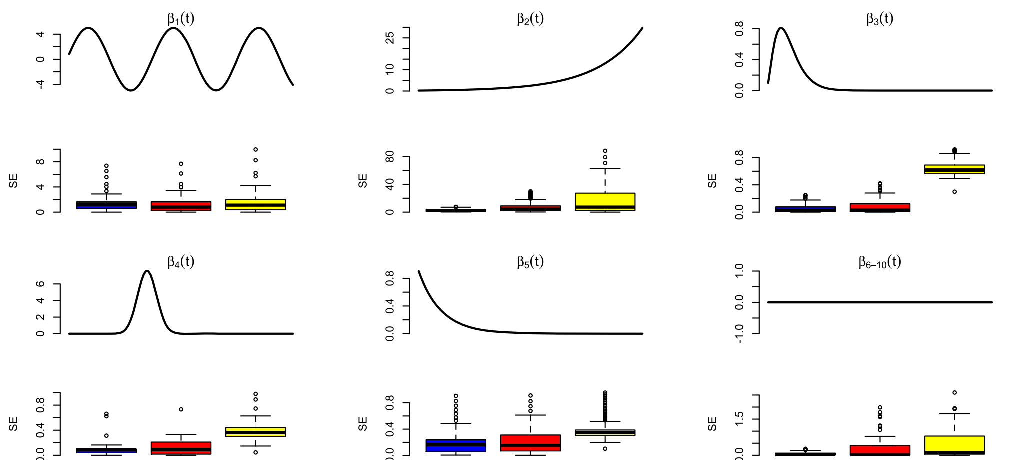

Figure 6. Comparison of SE of robust functional *LAD-agLASSO* (blue), robust functional *LAD-gLASSO* (red), and classical functional *agLASSO* (yellow), and at 15% contamination for Model (1).

coefficient functions are depicted in Figure 6. The true model depends on $\beta_1(t)$-$\beta_5(t)$ and the $\beta_6(t)$-$\beta_{10}(t)$ are essentially 0. We employ *MAE* (mean absolute error) of prediction and *SE* (squared error) metrics.

In order to evaluate the performance of the proposed methods, we employ two metrics.

To evaluate the predictive performance of the proposed methods, *MAE* of prediction which is defined as

$$MAE = \frac{1}{n}\sum_{i} Y_i - \widehat{Y}_i|, \tag{16}$$

is used. In this measure, Y_i and $\widehat{Y}_i$ are the true and predicted values, respectively. n is the number of observations.

For the sparseness accuracy, the SE defined as

$$SE = \int (\hat{\beta}_j(t) - \beta_j(t))^2 dt, \tag{17}$$

is used and $\hat{\beta}_j(t)$ and $\beta_j(t)$ are the estimated and true coefficient functions, respectively.

Figures 6–8 correspond to boxplots of *SE* for Model (1), Model (2), and Model (3), respectively. Figures 9–11 show the boxplots of *MAE* for Model (1), Model (2), and Model (3), respectively. Figures 6 and 9 show that our proposed methodology robust functional *LAD-agLASSO* performs better compared to robust functional *LAD-gLASSO* (Pannu and Billor 2017) and classical functional *agLASSO* (Gertheiss, Maity, and Staicu 2013) when there are outliers in the *Y* direction, but Figures 7 and 10 reveal that it does not perform well when we have outliers in both the response and functional predictors. Instead, our second proposed method robust functional *WLAD-agLASSO* works better in that scenario. Lastly, it can be seen in Figures 8 and 11 that all methods perform equally well in the absence of outliers. Furthermore, we examine our proposed methods robust functional *LAD-agLASSO* and functional *WLAD-agLASSO* at 25% and 40% contamination levels for Model (1) and Model (2) settings and we observed that these methods break down at 40% contamination level which these results are not included in this paper.

Furthermore, Table 1 shows the proportions of 100 simulation runs with the respective functional predictor being selected and average model size using functional *LAD-agLASSO*, functional *LAD-gLASSO* and classical functional *agLASSO* for Model (1) where we have outliers only in the response variable. We see in Table 1 that the true predictors $X_1(t) - X_5(t)$ are selected most frequently and predictors $X_6(t) - X_{10}(t)$ which are irrelevant to the true model are less frequently selected by the functional *LAD-agLASSO* compared to functional *LAD-gLASSO* and classical functional *gLASSO*. To summarize, the percentage of false positives and false negatives reduces when functional *LAD-agLASSO* is used.

4. Real data application

4.1. Sugar data

First, we consider a sugar spectra data described by Munck et al. (1998). The dataset consists of 268 samples of dissolved sugar measured spectrofluorometrically every 8 hours for about 3 months during a campaign at a sugar plant in Scandinavia. The emission spectra range from 275 to 560 nm measured at 571 wavelengths (in 0.5 intervals) at 7 excitation wavelengths (230, 240, 255, 290, 305, 325, and 340 nm). As discussed in Gertheiss, Maity, and Staicu (2013), the objective of this experiment is to study how the fluorescence spectra relate to the ash content (which measures the amount of inorganic impurities in the refined sugar, Bro 1999). This dataset is a good candidate to compare our proposed methods since it has outliers (red curves) in the functional predictors (seven excitation wavelengths) and the scalar response (Ash content) as shown in Figures 12 and 13, respectively. The outliers in functional predictors are detected using the depth function (Febrero, Galeano, and González-Manteiga 2008) in the *R* package *fda.usc* (Febrero-Bande and Oviedo de la Fuente 2012). Figure 14

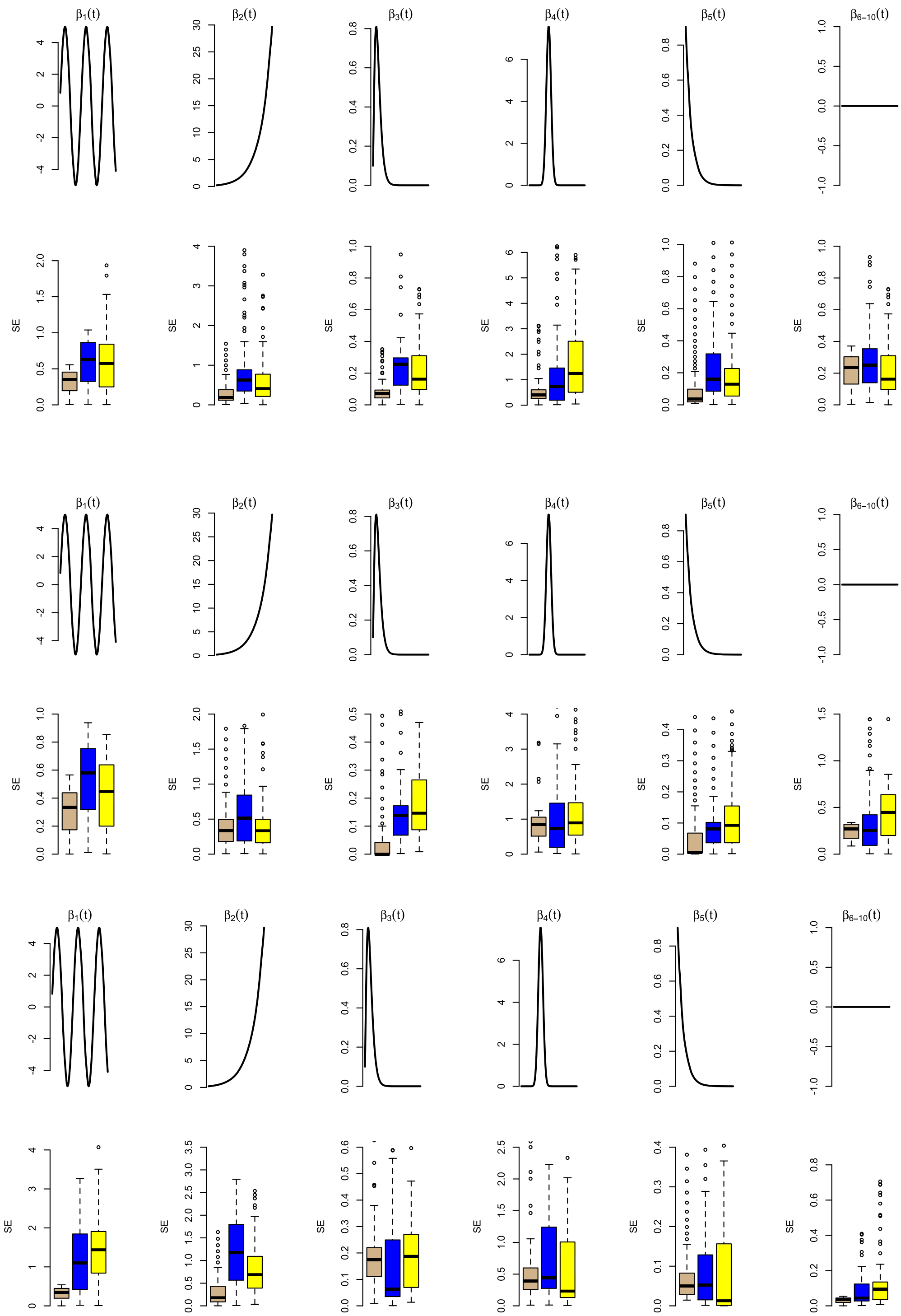

Figure 7. Comparison of SE of robust functional *WLAD-agLASSO* (tan), robust functional *LAD-agLASSO* (blue), and classical functional *agLASSO* (yellow) at 15% Asymmetric (Case1), Symmetric (Case 2), and Partial (Case 3) contamination, respectively for Model (2).

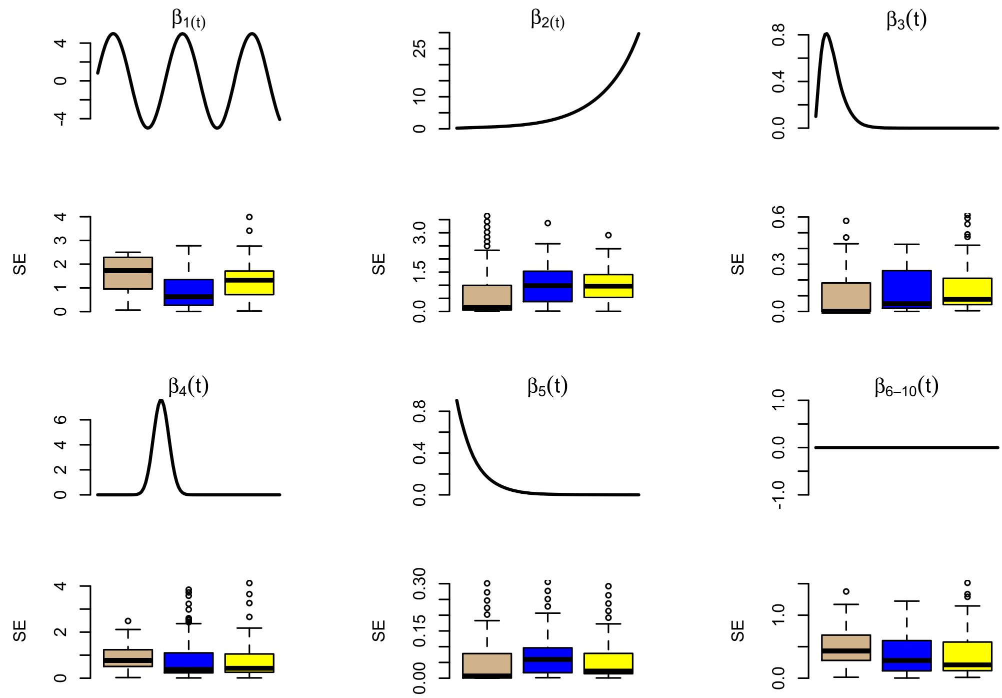

Figure 8. Comparison of SE of robust functional *WLAD-agLASSO* (tan), robust functional *LAD-agLASSO* (blue), and classical functional *agLASSO* (yellow) for Model (3).

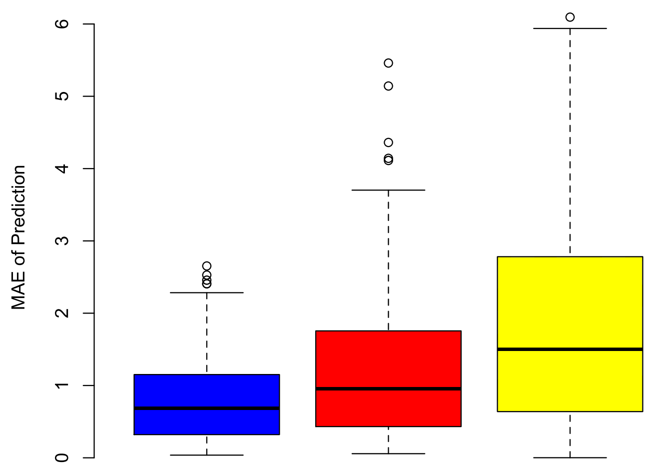

Figure 9. Comparison of MAE of robust functional *LAD-agLASSO* (blue), robust functional *LAD-gLASSO* (red), and classical functional *agLASSO* (yellow) at 15% contamination for Model (1).

shows the comparison of functional *WLAD-agLASSO* (tan) and functional *LAD-agLASSO* (blue). We can see that both methods exclude excitation wavelengths 255 nm and 340 nm, but functional *WLAD-agLASSO* (tan) also excludes excitation wavelengths 230 nm and 305 nm from the model. The analysis of the same data done using classical

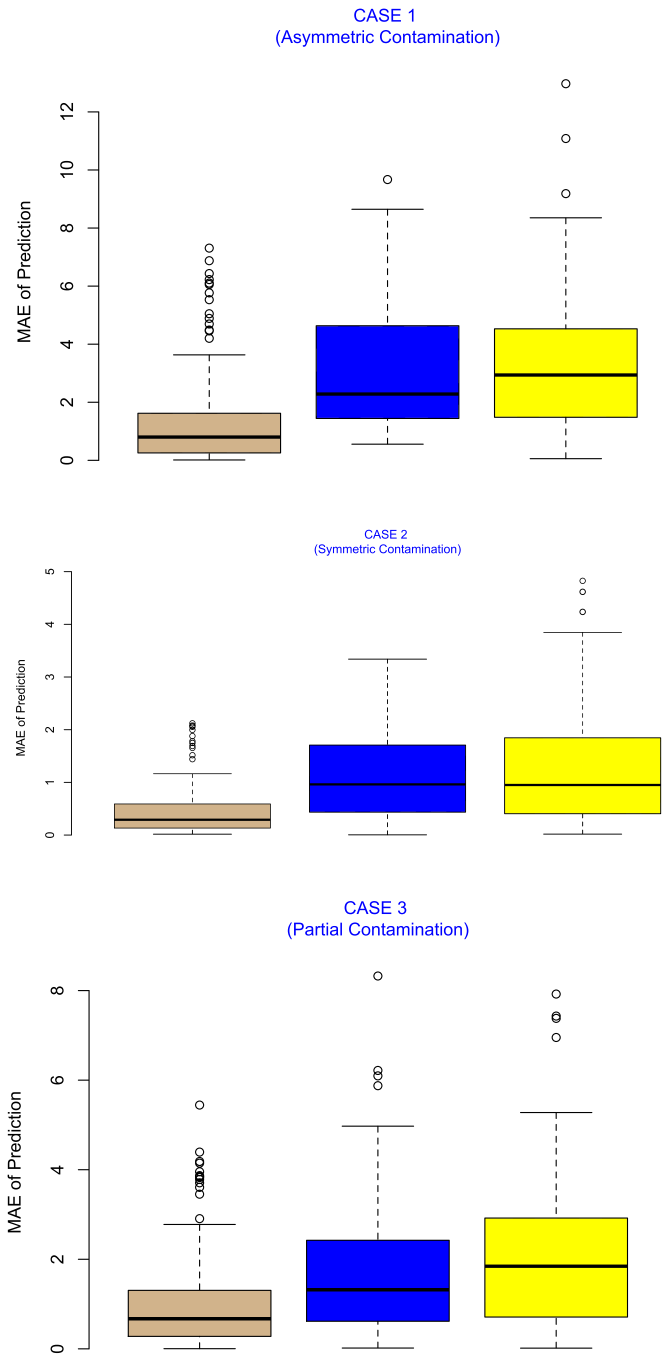

Figure 10. Comparison of MAE of robust functional *WLAD-agLASSO* (tan), robust functional *LAD-agLASSO* (blue), and classical functional *agLASSO* (yellow) at 15% contamination for Model (2).

functional *agLASSO* by Gertheiss, Maity, and Staicu (2013) gave different results in which excitation wavelength 230 nm, 255 nm, and 305 nm were excluded from the model.

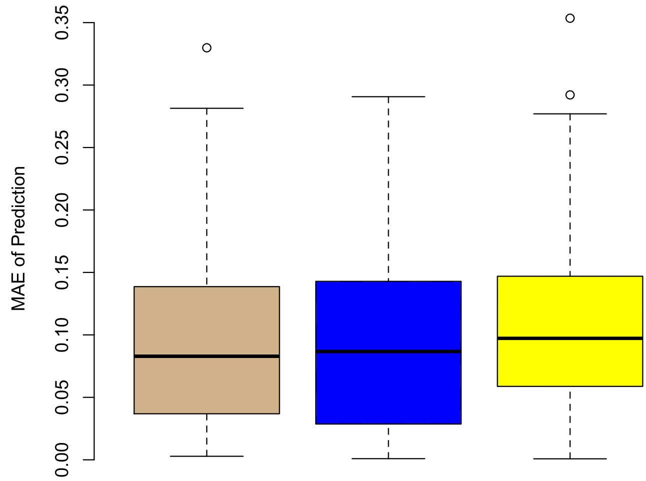

Figure 11. Comparison of MAE of robust functional *WLAD-agLASSO* (tan), robust functional *LAD-agLASSO* (blue), and classical functional *agLASSO* (yellow) for Model (3).

Table 1. Proportions of runs with the respective functional predictor being selected and average model size.

	X_1 (t)	X_2 (t)	X_3 (t)	X_4 (t)	X_5 (t)	X_6 (t)	X_7 (t)	X_8 (t)	X_9 (t)	X_10 (t)	Average model size
Functional LAD-agLASSO	1	0.98	1	1	0.88	0.24	0.15	0.17	0.05	0.10	5.57
Functional LAD-gLASSO	1	0.92	0.90	0.79	0.65	0.32	1	0.48	1	0.65	7.71
Classical Functional agLASSO	1	0.88	0.60	0.80	0.85	1	0.36	1	1	0.54	8.03

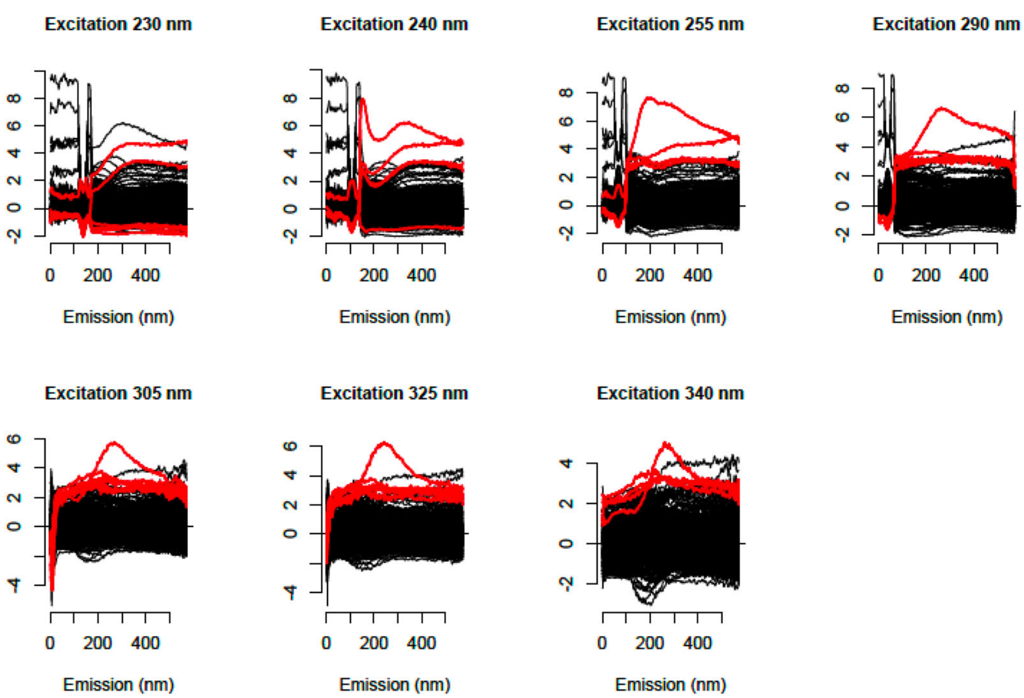

Figure 12. Red curves showing outliers in seven Excitation Wavelengths: 230, 240, 255, 290, 305, 325, and 340 nm, respectively.

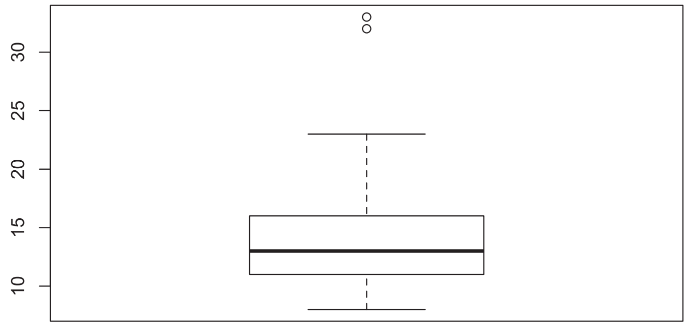

Figure 13. Box plot showing outliers in the Ash Content.

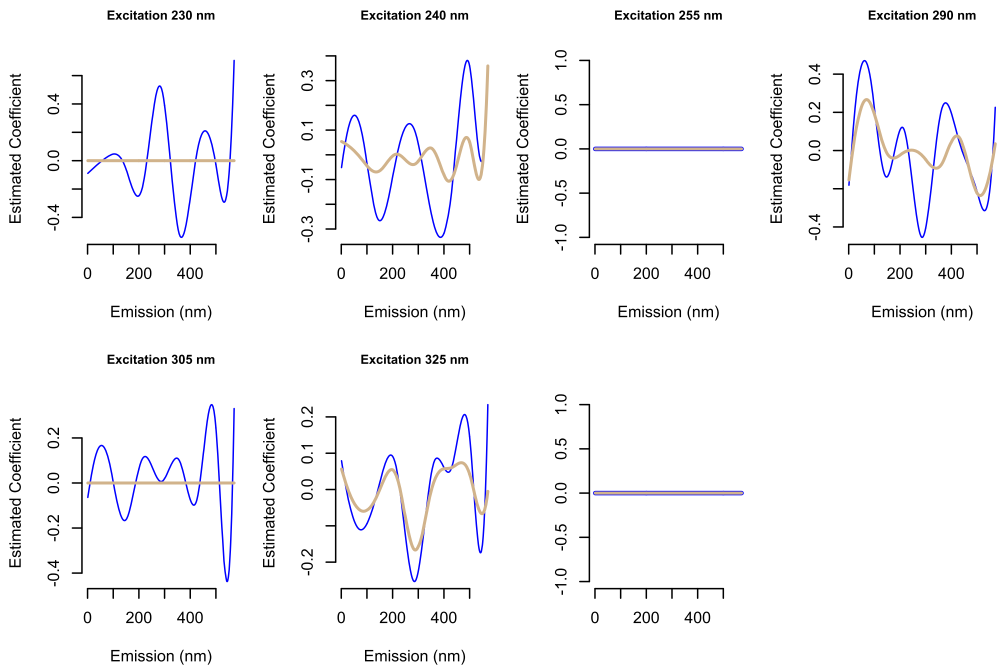

Figure 14. Estimated Coefficients for Sugar data using functional *WLAD-agLASSO* (tan) and functional *LAD-agLASSO* (blue).

4.2. Weather data

Next, we reconsider the weather dataset used by Matsui and Konishi (2011). The dataset is collected over time for 12 months at 79 stations in Japan from 1971 to 2000. As discussed by Pannu and Billor (2017) this dataset has outliers in both response (PRECIPITATION) and functional variables (TEMPERATURE, PRESSURE, HUMIDITY, and DAYLIGHT). From our simulation study in Sec. 3, we observed that functional *WLAD-agLASSO* works best when there are outliers in both response and predictors compared to the other proposed method. So we apply functional *WLAD-agLASSO* to the weather data. The fitting results are shown in Figure 15. The variables PRESSURE and DAYLIGHT are excluded from the model. The results indicate that

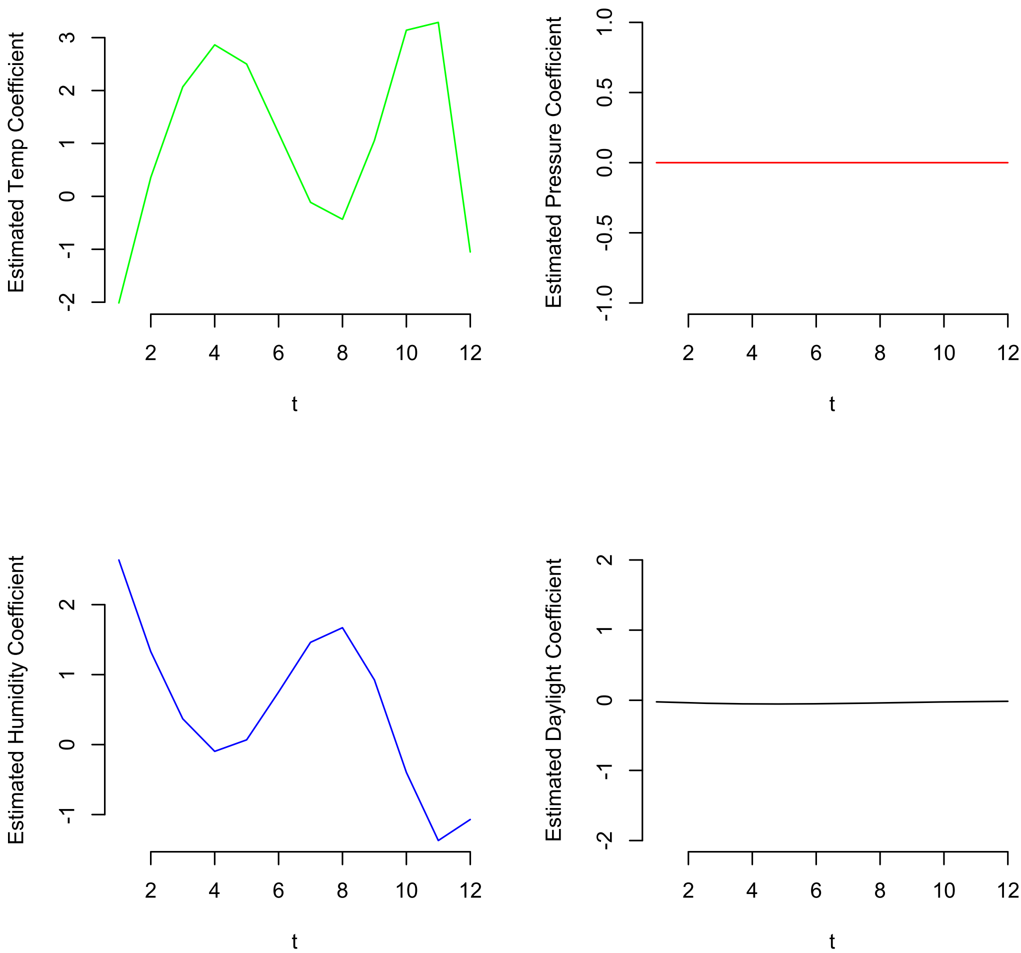

Figure 15. Estimated variable coefficients for weather data using functional *WLAD-agLASSO*.

there is no significant relationship between these variables and the PRECIPITATION. The remaining variables, TEMPERATURE and HUMIDITY, may relate to the precipitation. A different model was given by functional *LAD-gLASSO* in Pannu and Billor (2017) for the same dataset. As a result, the functional *WLAD-agLASSO* is a better method when the data have outliers in both functional predictors and *Y* variable.

Additionally, we generate 50 bootstrap samples from the weather data. For each bootstrap sample, functional regression modeling is performed using functional *WLAD-agLASSO*, functional *LAD-agLASSO* and functional *LAD-gLASSO*. We examine how many times each variable is selected. The results are shown in Table 2. The table shows that functional *WLAD-agLASSO* gives us the smallest model size and functional *LAD-gLASSO* gives the highest model size among three methods. Also, the mean TEMP is selected most frequently among the four variables, followed by HUMIDITY by functional *WLAD-agLASSO*. This reveals significant relationships of these variables to the precipitation. On the other hand, the average PRESSURE and DAYLIGHT are less frequently selected by functional *WLAD-agLASSO*. From the results, there seems to be less of a significant relationship between these variables and the precipitation.

Table 2. Proportions of runs with the respective functional predictor being selected and average model size.

	TEMP	PRESSURE	HUMIDITY	DAYLIGHT	Avg. Model Size
Functional *WLAD-agLASSO*	1	0.36	0.98	0.40	2.74
Functional *LAD-agLASSO*	1	0.90	0.98	0.92	3.80
Functional *LAD-gLASSO*	1	0.94	0.98	0.96	3.88

5. Summary and discussion

In this paper, we proposed two robust variable selection procedures, functional *LAD-agLASSO* and functional *WLAD-agLASSO* for functional linear regression models in the presence of outliers. These methods overcome the limitations of the method robust functional *LAD-gLASSO* proposed by Pannu and Billor (2017). The first limitation functional *LAD-gLASSO* is that it penalizes all regression coefficient functions with the same penalty term which is not correct since the amount of contribution for each functional predictor on the response may vary. The second limitation is that functional *LAD-gLASSO* method takes into account only the outliers in the response, Y, variable not the outliers in functional predictor, $X(t)$. But the functional outliers in functional predictors may cause the most serious issues for the estimators of the model parameters and the predicted model. We found that our first proposed method functional *LAD-agLASSO* overcomes the first limitation of functional *LAD-gLASSO* by providing more accurate model when there are outliers only in the response variable. Our second proposed method, functional *WLAD-agLASSO*, is robust to outliers in both the functional predictors and the response, Y and hence overcomes the second limitation of functional *LAD-gLASSO*. We presented extensive simulation studies at 15% contamination level and observed that our proposed methods perform better than other methods. In addition, the examination of our proposed methods at 25% and 40% contamination levels reveals that our proposed methods still perform better than other methods at 25% contamination level, but break down empirically at contamination level of 40%. Furthermore, we did not explore the theoretical properties of our proposed methods in this paper, but we wish to do so in future.

References

Aneiros, G., and P. Vieu. 2016. Sparse nonparametric model for regression with functional covariate. *Journal of Nonparametric Statistics* 28 (4):839–21. doi:10.1080/10485252.2016.1234050.

Arribas-Gil, A., and J. Romo. 2013. Shape outlier detection and visualization for functional data: The outliergram. arXiv:1306.1718.

Arslan, O. 2012. Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis* 56 (6):1952–65. doi:10.1016/j.csda.2011.11.022.

Berrendero, J., A. Cuevas, and J. Torrecilla. 2016. Variable selection in functional data classification: A maxima-hunting proposal. *Statistica Sinica* 26:619–38. doi:10.5705/ss.202014.0014.

Bro, R. 1999. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometrics and Intelligent Laboratory Systems* 46 (2):133–47. doi:10.1016/S0169-7439(98)00181-6.

Dai, W., and M. Genton. 2019. Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis* 131:50–65. doi:10.1016/j.csda.2018.03.017.

Ellis, S., and S. Morgenthaler. 1992. Leverage and breakdown in L1 regression. *Journal of the American Statistical Association* 87 (417):143–8. doi:10.2307/2290462.

Fan, J., and R. Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96 (456):1348–60. doi:10.1198/016214501753382273.

Febrero, M., P. Galeano, and W. González-Manteiga. 2008. Outlier detection in functional data by depth measures with application to identify abnormal *NOx* levels. *Environmetrics* 19 (4): 331–45. doi:10.1002/env.878.

Febrero-Bande, M., and M. Oviedo de la Fuente. 2012. Statistical computing in functional data analysis: The R Package fda.usc. *Journal of Statistical Software* 51 (4):1–28. http://www.jstatsoft.org/v51/i04/. doi:10.18637/jss.v051.i04.

Fraiman, R., and G. Muniz. 2001. Trimmed means for functional data. *Test* 10 (2):419–40. doi:10.1007/BF02595706.

Gertheiss, J., A. Maity, and A. M. Staicu. 2013. Variable selection in generalized functional linear models. *Stat* 2 (1):86–101. doi:10.1002/sta4.20.

Giloni, A., J. Simonoff, and B. Sengupta. 2006. Robust weighted LAD regression. *Computational Statistics & Data Analysis* 50 (11):3124–40. doi:10.1016/j.csda.2005.06.005.

Huang, H., and Y. Sun. 2016. Total variation depth for functional data. arXiv:1611.04913.

Hubert, M., and P. Rousseeuw. 1997. Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference* 57 (1):153–63. doi:10.1016/S0378-3758(96)00041-9.

Hubert, M., P. Rousseeuw, and P. Segaert. 2015. Multivariate functional outlier detection. *Statistical Methods & Applications* 24 (2):177–202. doi:10.1007/s10260-015-0297-8.

Koenker, R. 2013. Quantreg: Quantile regression. R Package Version 5.05. http://CRAN.Rproject.org/package=quantreg

Matsui, H. 2017. Selection of variables and decision boundaries for functional data via bi-level selection. arXiv:1702.02010 [stat.ME].

Matsui, H., and S. Konishi. 2011. Variable selection for functional regression models via the L1 regularization. *Computational Statistics & Data Analysis* 55 (12):3304–10. doi:10.1016/j.csda.2011.06.016.

Meier, L., S. Van de Geer, and P. Bühlmann. 2009. High-dimensional additive modeling. *The Annals of Statistics* 37 (6B):3779–821. doi:10.1214/09-AOS692.

Munck, L., L. Nørgaard, S. B. Engelsen, R. Bro, and C. A. Andersson. 1998. Chemometrics in food science a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometrics and Intelligent Laboratory Systems* 44 (1–2):31–60. doi:10.1016/S0169-7439(98)00074-4.

Pannu, J., and N. Billor. 2017. Robust group-Lasso for functional regression model. *Communications in Statistics - Simulation and Computation* 46(5):3356–74. doi:10.1080/03610918.2015.1096375.

Picheny, V., R. Servien, and N. Villa-Vialaneix. 2019. Interpretable sparse SIR for functional data. *Statistics and Computing* 29 (2):255–67. doi:10.1007/s11222-018-9806-6.

Rousseeuw, P. J. 1984. Least median of squares regression. *Journal of the American Statistical Association* 79 (388):871. doi:10.1080/01621459.1984.10477105.

Rousseeuw, P. J., and K. V. Driessen. 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41 (3):212–23. doi:10.1080/00401706.1999.10485670.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1):267–88. doi:10.1111/j.2517-6161.1996.tb02080.x.

Tibshirani, R., S. Michael, R. Saharon, Z. Ji, and K. Keith. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67 (1): 91–108.

Todorov, V. 2009. rrcov: Scalable robust estimators with high breakdown point. R Package Version 1.4-3. https://CRAN.R-project.org/package=rrcov

Tutz, G., and J. Gertheiss. 2010. Feature extraction in signal regression: A boosting technique for functional data. *Journal of Computational and Graphical Statistics* 19 (1):154–74. doi:10.1198/jcgs.2009.07176.

Yuan, M., and Y. Lin. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (1):49–67. doi:10.1111/j.1467-9868.2005.00532.x.

Zou, H. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101 (476):1418–29. doi:10.1198/016214506000000735.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2):301–20. doi:10.1111/j.1467-9868.2005.00503.x.