Int. J. Nonlinear Anal. Appl. 12 (2021) No. 1, 1037-1055
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.23023.2458

A New Robust Algorithm for Penalized Regression Splines Based on Mode-Estimation

Ahmed Eldeeb[a,b,*], Sabreen Desoky[b], Mohamed Ahmed[b]

[a] Department of Business Administration, college of Business, King Khaled University, KSA.
[b] Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Alexandria University, Egypt.

(Communicated by Madjid Eshaghi Gordji)

Abstract

The main purpose of present article is proposed an effective method for robust fitting penalized regression splines models. According to such a context a comparative analysis with two common robust techniques, M-type estimator, S-type estimator, and non-robust least squares (LS) for penalized regression splines (PRS) has been implemented. Because the penalized regression splines are recently a common approach to smoothing noisy data for its simplicity, efficiency, and significantly reducing disturbance of outliers and its flexibility in monitoring nonlinear data trends. In many cases, it is difficult to determine the most suitable form and a way of designing a data is needed when faced with many smoothing problems. The executing aspects of fitting precision and robustness of the four estimators have a thorough evaluation of their performance on R codes. A comparative analysis demonstrates that the proposed method can resist the noise effect in both simulated and real data examples compared to other robust estimators with different combinations of contamination. These findings are used as guidance for finding a specific method to pulsing smoothing noisy data.
AMS Classification: 62F35, 62-08, 62-04

Keywords: M-estimator; S-estimator; modal regression; penalized regression splines; Smoothing.

1. Introduction

The penalized technique of spline smoothing was already seen as a popular nonparametric approach for the noisy smoothing data in the design domain Wegman and Wright (1983), Wood (2000) and Kim, and Gu (2004). Over the last two decades because of its convenient and flexible

*Corresponding author
Email addresses: asedky@kku.edu.sa (Ahmed Eldeeb), sabreenfraag@yahoo.com (Sabreen Desoky), drmoh.alyahmed@gmail.com (Mohamed Ahmed)

Received: October 2020 *Revised:* December 2020

fitting inner knots and parameter smoothing. The concept of integrating the regression splines with slightly fewer inner knots relative to the sample size, and the penalized parameter introduction can at least be traced back to O'Sullivan (1986) who used the B-spline base in estimation.
In the early stages, Kelly and Rice (1990) approximated a smoothing curve spline with hybrid splines fitted with the inner knots equal to the sample data and the penalty value parameter, which was normally determined by the generalized cross-validating criterion. In addition, Marx and Eilers (1998) suggested flexible criteria to choose an optimal penalty parameter to deal with the coefficients of the B-splines. Wand (1999) has also derived a quick and simple approach to find the optimum penalty parameter for the regression of the penalized spline. In addition, Ruppert et al. (2003) examined spatially different penalties and choosing the inner knots layout.
A penalized minimization problem can be derived in the closed-form formula. Such penalties appear to be the least, but regression splines cannot be disturbance resistant from outliers. An easy and straightforward idea is to change squared residuals by a slower loss function close to the estimator for M-type regression Wold et al. (2001). Achieving atypical observation alleviate the outliers effects. Early investigation of the technique for M-type smoothing. The work on cubic splines for regression was introduced by Cox (1983) and by Lee and Oh (2007). Another effective model estimate is the Sestimation by Eilers and Marx (1996), with its high breakdown and extremely robust Tharmaratnam et al (2010). The regression curve for the penalized spline is defined by minimizing the squared residuals, subject to a bound at the spline coefficient scale. Therefore, a closed-form form formula can be extracted from the penalized minimization problem. Obviously, those penalized least square regression splines cannot be resistant to the perturbation of the outliers. Likewise, the clear and simple concept of the M-type estimator is to replace the squared with a loss function, to alleviate the outlier effects. Early research started on the M-type smoothing technique by Huber (1981) who worked on cubic regression splines. The S-type estimation which has a high breakdown point and minimizes the scale of residuals in an extremely robust way, can be another robust model Rousseeuw and Yohai (1984). Computing the S-type regression estimates have developed a fast algorithm in Salibian-Barrera and Yohai (2006). The results of comparison in showed that the S-estimator is more capable of resisting outlier disturbance compared to the M-type estimator, but with smaller efficiency Wang et al. (2014). However, until 15 years ago, little work was done to put the penalized spline regression into the robust estimate category. Oh et al. (2004) proposed a robust smoothing method introduced to the time analysis of variable started by simply enforcing on the generalized cross-validation (GCV) criterion the favorite loss function of Huber Tharmaratnam et al. (2010). An iterative procedure for conniving the M-type estimator of penalized regression splines via the introduction of empirical pseudo data was proposed by Lee and Oh (2007) and Finger (2013) subsequently verified that such an M-type estimator compromised the efficiency of different estimation methods in insurance applications among the very robust estimators and the very effective least square type estimators for penalized spline regressions. Meanwhile Tharmaratnam et al. (2010) have also submitted S-type estimator to substitute a least square estimate with an efficient estimator for penalized regression splines which is equivalent to a weighted penalized least square regression and which is well-conforming even in the case of highly contaminated samples. This article introduces the new robust method of smoothing penalized regression splines accordingly and carries out a comparative study with M-type estimator and the S-type estimating technique for penalized regression splines. Section 3 involves the proposed method with its algorithm. Comparative studies are conducted with both simulated data and a real weather ballon data set. Concluding remarks are included in Section 5 .

2. Robust Penalized Regression Splines

The idea of the penalized regression splines begins with the following model:

$$y_i \sim N\left(f\left(x_i\right), \sigma_{\varepsilon}^2\right), i = 1, \ldots, n \tag{2.1}$$

where $\mathrm{f}\left(x_i\right)$ is supposed to be a smooth but unknown regression function which needs to be estimated based on the sample observations $\left(x_i, y_i\right), i = 1, \ldots, n$ and σ_{ε}^2 denotes the constant variance of the random deviation error between the response variable and the regression function $\mathrm{f}(x)$. It's a flexible definition in terms of penalized spline smoothing since the different basis functions will refer to the various penalized spline smoothing functions. The truncated polynomial bases are a common collection, but other choices can be explored similarly,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \sum_{k=1}^{K} \beta_{p+k}\left(x_i - \xi_k\right)_+^p + \varepsilon_i \tag{2.2}$$

where this regression splines are a linear combination of $(\mathrm{k} + \mathrm{p} + 1)$ of functions

$$1, x, \ldots, x^p, \left(x - \xi_1\right)_+^p, \left(x - \xi_2\right)_+^p, \ldots, \left(x - \xi_K\right)_+^p \tag{2.3}$$

where

$$\left(x_i - \xi_k\right)_+^p = \begin{cases} \left(x_i - \xi_k\right)^p & \text{if } x_i > \xi_k \\ 0 & \text{Otherwise} \end{cases} \tag{2.4}$$

which is called truncated-power basis function, $\left(x_i - \xi_k\right)_+^p = \max\left(x_i - \xi_k, 0\right)$ and $\beta = \left(\beta_0, \beta_{1\ldots}, \beta_{p+k}\right) T$ denotes a vector of regression coefficients, $\xi_j, \ldots, \xi_K$ is the specified inner knots, and p indicates the exponential order for truncated power basis. The vector of parameters' estimator can be written in matrices form as

$$\hat{\beta} = \left(X^T X + \lambda D\right)^{-1} X^T y \tag{2.5}$$

We can rewrite the regression splines model in general form using base function B-spline:

$$y_i = \sum_{j=1}^{K+p+1} \beta_j B_j\left(x_i; p\right) + \varepsilon_i \tag{2.6}$$

where $B_j\left(x_i; p\right)$ is B -spline number j in p degree.

Least Squares for Penalized Regression Splines (LSPRS)

Given a group of sample observations $\left(x_1, y_1\right), \ldots, \left(x_n, y_n\right)$, an increasingly popular way for obtaining the estimation of $\mathrm{f}(\cdot)$ is via transforming it into the category of a least squares problem, where we need to seek out the member from the class $\mathrm{f}(x; \beta)$ which minimizes the sum of squared residuals Ruppert et al. (2003)

$$\sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \sum_{k=1}^{K} \beta_{p+k}\left(x_i - \xi_k\right)_+^p\right)\right)^2 + \lambda \sum_{k=1}^{K} \beta_k^2 \tag{2.7}$$

By using the Lagrange multiplier, the coefficients of the penalized least squares regression splines $\hat{\hat{\beta}}_{LS}$ are equivalent to mini

$$n\sum_{\mathrm{i}=1}\left(\mathrm{y_i}-\mathrm{f}\left(\mathrm{x_i};\beta\right)\right)^2+\lambda\mathrm{k}\sum_{\mathrm{j}=1}\beta^2_{\mathrm{p+j}} \tag{2.8}$$

For any parameter smoothing (penalty) $\lambda \geq 0$ with regard to β, violating the biases of the fitted spline reverse curve with the variance, and with diagonal matrix D that only penalizes the spline coefficient. The direct calculations can be made to the closed-formula minimizer for objective function (5) Ruppert et al. (2003). Therefore, the corresponding estimate vector $\hat{\mathrm{y}}$ is expressed by the penalized least square regression splines as

$$\hat{y} = X\hat{\beta} = X\left(X^TX+\lambda D\right)^{-1}X^Ty \tag{2.9}$$

$$\hat{y} = S_\lambda y \tag{2.10}$$

Where $S_\lambda = X\left(X^TX+\lambda D\right)^{-1}X^T$ is called a smoothing matrix. In the case of normal least squares S_λ is typically called smoothing matrix which corresponds to a hat matrix.

Choosing the Smoothing Parameter $\boldsymbol{\lambda}$

Using a simple trial-and-error technique, we can either manually choose the value of λ, or we can implement an automated process. In the PRS model, the smoothing parameter plays a key role, as the value of this parameter affects the shape of the corresponding data set curve, so it is important to use a suitable value for that parameter.
Figure 1 shows the difference between the effective curve using a relatively large parameter value λ (right shape) and using a relatively small parameter value (left shape).

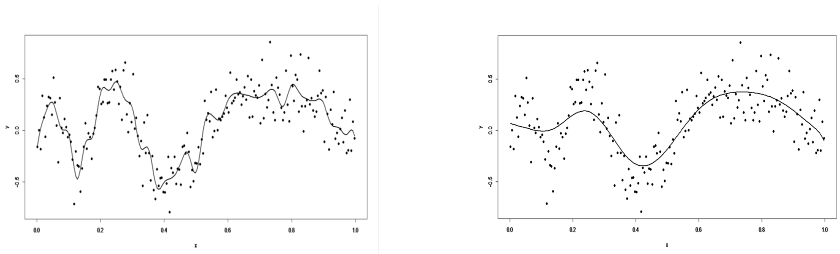


Figure 1: The right shape is the curve that is reconciled with a relatively large parameter value, and the left-hand curved shape that is reconciled with a relatively small parameter value.

S_λ, which can be regarded as a linear conversion matrix by which the vector of estimates y is linear as $\hat{\mathrm{y}} = \mathrm{Ly}$. An appropriate curve is generally a matter of concern, not an approximation at each point, so the overall error across values of the variable X is usually calculated. One simple alternative which avoids relying on the values of variable x enabled is the mean sum of Squared errors (MSSE)

$$MSSE = \sum_{i=1}^{n}\left\{\hat{f}\left(x_i\right)-f\left(x_i\right)\right\}^2 \tag{2.11}$$

Therefore, MSSE can be written as Ruppert et al. (2009)

$$\text{MSSE}(\hat{f}) = \sum_{i=1}^{n} \left\{ E\hat{f}\left(x_i\right) - f\left(x_i\right) \right\}^2 + \text{Var}\left\{ \hat{f}\left(x_i\right) \right\} \tag{2.12}$$

$$= \text{Bias}^2\{\hat{f}(x)\} + \text{Var}\{\hat{f}(x)\} \tag{2.13}$$

In the case of the PRS model, the equation is as follows:

$$\text{MSSE}(\hat{f}) = \left\| \left(S_\lambda - I\right) f \right\|^2 + \sigma_\varepsilon^2 \, \text{tr}\left(S_\lambda S_\lambda^T\right) \tag{2.14}$$

Where the first term refers to the square of bias and the second term refers to the sum of the variations. Draper and Smith (1998) clarified that estimates of the PRS model take the form of Ridge regression estimates, which are in turn, biased input regression capacities used in reducing estimates change Krivobokova (2006). This can be said as a continuous function, in a single parameter, the smoothing range of the estimates within the PRS model is ready to change Craven and Wahba (1978). The higher the value of this parameter the more effective the curve is, the more distraction has been and vice versa. It is easy to measure and does not require any terminology for the distribution of model errors using the crossvalidation norm (CV) Hurvich et al. (1998). It is one of the parameters used to pick a smoothing parameter in the PRS. The practical sample is divided into n parts and the model is then reconciled to the measurement of the remaining square residuals by using a (n-1) part and the remaining square residuals will be measured. The average residual CV is estimated as follows:

$$CV = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{f}^{-i}\left(x_i\right) \right)^2 \tag{2.15}$$

Where $\hat{f}^{-i}\left(x_i\right)$ refers to the model that is fitted using the sample available after excluding observation number i. Calculating CV requires to set $\mathrm{Q} = \mathrm{n}$ to fit a n number of times model, which is mathematically difficult, but the relationship is achieved for many linear functions which depend on error squares as a loss function. For considering out the penalty parameter λ in (5), GCV technique is utilized hereafter, which is computed by leave-one-out of the residual from sum of squares to avoid overfitting in the regression splines Wilcox (2011). CV criterion is given by

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{f}^{-i}\left(x_i, \lambda\right) \right)^2 \tag{2.16}$$

The augmented penalty parameter λ can be determined by minimizing CV (λ) over $\lambda \geq 0$. For the linear smoothing matrix, $H\lambda = \mathrm{X}\left(\mathrm{X}^T\mathrm{X} + \lambda D\right)^{-1}\mathrm{X}^T$, and it can be verified as

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ y_i - \hat{f}^{-i}\left(x_i\right) \right\}^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{y_i - \hat{f}\left(x_i\right)}{1 - h_{ii}} \right]^2 \tag{2.17}$$

Where h_{ii} is the ith number of the main diagonal in estimation matrix Ruppert et al. (2003). GCV criterion is proposed simply by replacing the diagonal elements with their average $\left(\frac{\text{tr}(S_\lambda)}{n}\right)$ in PRS model Craven and Wahba (1978).

$$GCV_\lambda = n\|y - \hat{f}(X)\|^2 / (n - \text{trace}(H(\lambda)))^2 \tag{2.18}$$

M -Type Estimator for Penalized Regression Splines LS may greatly suffer from various disturbances mainly when the data is contaminated with outliers. A penalized regression estimator can be constructed easily by replacing the squared residual loss function with the following M-type criterion to alleviate effects of outliers Lee and Oh (2007):

$$\hat{f}_{\text{robust}}(x) = \operatorname{argmin}_{\beta}\left\{\sum_{i=1}^{n} \rho_c\left(y_i - \beta^T \mathrm{x}_i\right) + \lambda \beta^T \mathrm{D} \beta\right\} \tag{2.19}$$

where ρ is even, nondecreasing in $[0, +\infty)$ and $\rho(0) = 0$, for which a common choice is taken of the Huber loss function Tharmaratnam et al. (2010) with cutoff $c > 0$,

$$\rho_c(x) = \begin{cases} x^2 & |x| \leq c \\ 2c|x| - c^2 & |x| > c \end{cases} \tag{2.20}$$

The above function of Huber is apparently a parabola near zero. A default choice of the tuning constant $c = k\hat{\sigma}, \mathrm{k} = 1.345$ aims for a 95% asymptotic efficiency with respect to the standard normal distribution. Denote the derivative of ρ_c is ψ_c. When a set of fitted residuals

$$r_i = y_i - f\left(x_i; \beta\right), i = 1, \ldots, n \tag{2.21}$$

The robust M-scale estimator for the estimation of the standard deviation σ_{ε} of these residuals can be used as follows Wilcox (2011), based on the penalized spline regression:

$$\sum_{i=1}^{n} \psi_c\left(y_i - f\left(x_i; \beta\right) \hat{\sigma}_{\varepsilon}\right) = 0 \tag{2.22}$$

A conventional option for a measure of scale with a high breakdown point can be the value of ω determined by

$$\hat{\sigma}^{(j+1)} = 1.4826 \times MAD\left(\hat{\varepsilon}_i\right) \tag{2.23}$$

Where MAD is the median absolute deviation statistic, and MAD is actually the sample median of the n values, with its finite sample breakdown point approximately being 0.5 and MAD in place normally needs to be rescaled to a more familiar context to estimate after all, particularly when residues are sampled from normal distribution Wilcox (2011). Especially,

$$MADN = MAD/Z_{0.75} \approx 1.4836MAD \tag{2.24}$$

By utilizing the aforementioned spline basis functions Lee and Oh (2007), M-type penalized spline estimator using the standardized residuals can be computed by obtaining the estimate of $\hat{f}^{(j+1)}(x)$ as follows

$$\hat{f}^{(j+1)}(x) = X\left(X^{\tau} X + \lambda D\right)^{-1} X^T Z^{(j+1)} \tag{2.25}$$

Under such circumstances, an M -type iterative algorithm for calculating the penalized regression splines is proposed in Wilcox (2011) by introducing the empirical pseudo data with iterative algorithm where the penalized least square estimator converges asymptotically. Regarding the M-type estimator for the penalized regression splines, the optimal value of smoothing parameter λ is obtained by GCV.

S-Estimation for Penalized Regression Splines

The low breakdown point of M-type estimator was explained by Yao and Li (2014). The robustness characteristics of the aforesaid M-type estimator for the penalized regression splines are not fully satisfactory. Estimation for penalized regression splines, the core concept of which is using flexible Stype estimator to replace the conventional least squares estimator, is another robust approach to smoothing noisy results. The coefficient vector of the penalized S -regression splines $\hat{\beta}_s$ is defined as follows:

$$\hat{\beta}_s = \operatorname{argmin}_\beta \left(n\hat{\sigma}_n^2(\beta) + \lambda\beta^T D\beta\right) \tag{2.26}$$

Where $\hat{\sigma}_n^2(\beta)$ is robust M-scale estimator and for each vector $\beta, \hat{\sigma}_n^2(\beta)$ Satisfies the equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - x_i^T\beta}{\hat{\sigma}_n(\beta)}\right) = E_\Phi[\rho(Z)] \tag{2.27}$$

To make the estimate consistent if the distribution is normal, where Φ refers to the standard normal distribution. Considering this, substitute the estimation of $\hat{\sigma}_\varepsilon(\beta)$ by MADN in (22) with the addition of the absolute value of residual median; namely,

$$\hat{\sigma}_s = MADN + \operatorname{Median}(r(\beta)) \tag{2.28}$$

The penalized S-estimator for the regression splines model is computed as $\hat{\mathrm{m}}_s = X\hat{\beta}_s$, with its coefficients vector $\hat{\beta}_s$ satisfying both (26) and (27), simultaneously. The formula below shows that such S-estimators are actually equal to a weighted penalized least squares regression Wang et al. (2014) And, thus, a further simple deformation arrives at the coefficient vector $\widehat{\beta}_s$ with an iteration form

$$\hat{\beta}_s = \left\{X^\tau W\left(\hat{\beta}_s\right)X + \frac{\lambda}{\tau\left(\hat{\beta}_S\right)}D\right\}^{-1} X^T W\left(\hat{\beta}_S\right)y \tag{2.29}$$

$$W(\beta) = \operatorname{diag}\left(W_i(\beta)\right) \in \mathbb{R}^{n\times n} \tag{2.30}$$

Where

$$W_i(\beta) = \rho'\left(\tilde{r}_i(\beta)\right)/\tilde{r}_i(\beta) \tag{2.31}$$

$$\tilde{r}_i(\beta) = \frac{\left(y_i - x_i^T\beta\right)}{\hat{\sigma}_n(\beta)} \tag{2.32}$$

And $\tau\left(\hat{\beta}_S\right) = n\hat{\sigma}_n^2(\beta)/\left[(y - X\beta)^T W(\beta)(y - X\beta)\right]$ This shows that the penalty parameter λ is then determined based on the next Regularized GCV criterion

$$RGCV_\lambda = \frac{n_w\left\|W(\hat{\beta})^{\frac{1}{2}}(y - X\hat{\beta})\right\|^2}{\left(n_w - \operatorname{trace}\left(H_S(\lambda)\right)\right)^2} \tag{2.33}$$

where

$$H_S(\lambda) = \tilde{X}\left(\tilde{X}^T\tilde{X} + \left(\lambda/\tau\left(\hat{\beta}_S\right)\right)D\right)^{-1}\tilde{X}^T \tag{2.34}$$

$$= W\left(\hat{\beta}_S\right)^{\frac{1}{2}} X\left(X^\tau W\left(\hat{\beta}_S\right)X + \left(\lambda/\tau\left(\hat{\beta}_S\right)\right)D\right)^{-1} X^T W\left(\hat{\beta}_S\right)^{\frac{1}{2}} \tag{2.35}$$

and n_w denotes the number of nonzero elements in weight matrix $W\left(\hat{\beta}_s\right)$. Therefore, the iterative $\mathrm{S}-$ type estimator algorithm for penalized regression splines based on the previous in-depth discussions had proposed in Finger (2013), which the penalized least squares estimator converges asymptotically.

3. The Proposed Method

The median and mode have significantly better results of robustness when outliers are present. In addition, since the modal regression focuses on the relationship between most of the data points, a more meaningful forecast of points and a higher probability of prediction coverage than the mean regression when the error distribution is skewed Zhao et al. (2014). Yao and Li (2014) suggested to estimate the modal regression parameter β for linear regression models called MODLR that assumes $f(y \mid x)$ is a linear function of variable x, Mode $(y \mid x) = y_i = x_i^T\beta$, by maximizing

$$Q(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}\phi_h\left(y_i - x_i^T\beta\right) \tag{3.1}$$

Where $\varepsilon = y - x^T\beta$ and $\phi_h(t) = h^{-1}\phi\left(\frac{t}{h}\right)$, $\phi_h(t)$ is a kemel density function, and h is a symmetric bandwidth. We shall assume, for the remainder of the article, that β is the usual normal computational density.

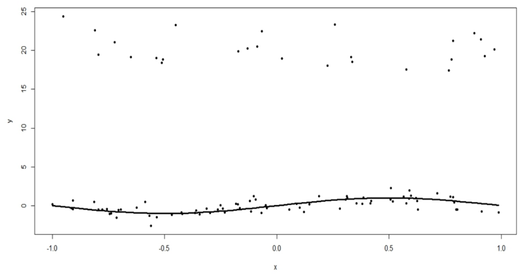
The conditional density of $f(\varepsilon \mid \mathrm{x})$ refers to error distribution, Yao, and Li (2014) showed that an error distribution based on x. However, if $f(\varepsilon \mid \mathrm{x})$ is skewed, the coefficients β of modal regression and Conventional norm regression coefficients will vary. The modal regression may also be a linear function of x, but it is not in Conventional norm regression. They proposed an EM algorithm to minimize a kernel-based objective function for the estimation of modal regression coefficients and suggested a way of generating asymmetrical intervals of predictions which can be better coverable than symmetric intervals if distributions are extremely skewed.

Since Zhao et al. (2014) B-spline base functions were used in the modal regression, which requires determining the appropriate number and location of nodes and because the penalized regression splines models overcome the problem of the number and location of the appropriate nodes by using a large number of nodes using the smoothing parameter. Therefore, our method proposes to use modal regression with penalized regression splines by Maximizing the objective function:

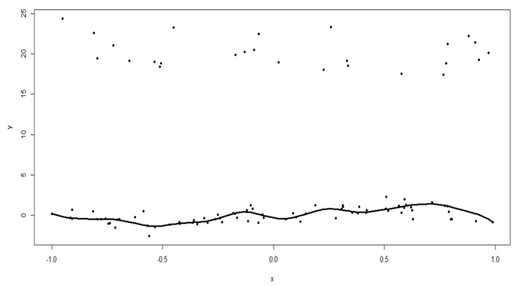
$$Q_h(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}\phi_h\left(y_i - x_i^T\beta\right) - \lambda\sum_{k=1}^{K}\beta_{k+p}^2 \tag{3.2}$$

The values of the two parameters h and λ control the shape of the curve that can be obtained when estimating the model, and a figure 2 can be viewed to illustrate this. Where the degree of smoothing parameter λ controls the smoothing degree of the curve, the greater the value, the greater the degree to which the curve is smoothed (see upper-left and middle-right shapes), the opposite is true (see middle left and lower right shapes). While the parameter h controls the degree of curved robustness, using a relatively small value for that parameter makes the resulting estimate more robust and unaffected by the presence of outliers (see upper left and middle left shapes). When a relatively large value is used for parameter h, it affects estimates in the presence of outliers (see middle right and lower right shapes). Therefore, it is necessary to select appropriate values to the parameters λ and h to obtain a well-fitted curve for the data. This is what we have tried to do in the current article.

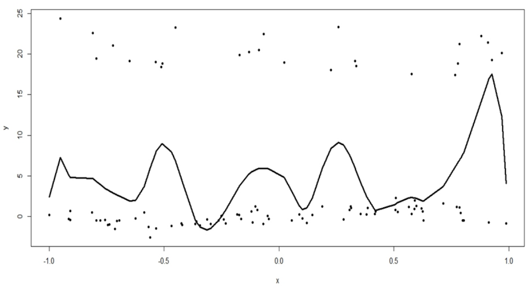
(a) Original function



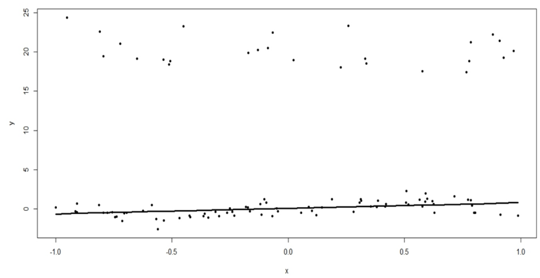
(b) small h and small λ



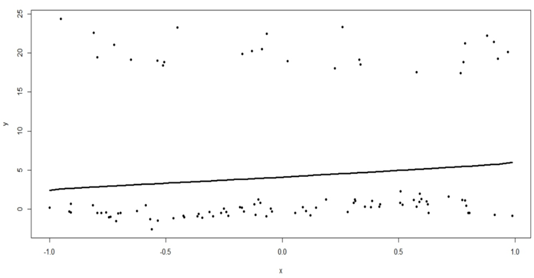
(c) large h and small λ



(d) small h and large λ



(e) large h and large λ



(f) Choosing h and λ using our proposed method

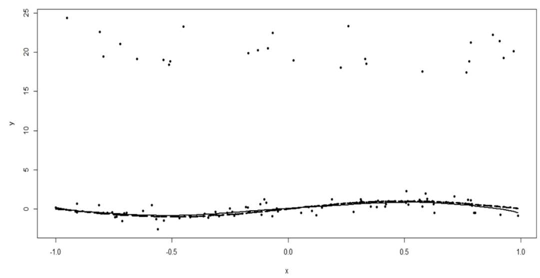


Figure 2: The upper-left shape is a scatter plot of a set of data that contains approximately 25% of outliers attached to the actual function curve (solid line). Other shapes represent Mode penalized regression splines with (Solid lines) by using different values for h and λ. The dashed line curve in the lower-right shape is the curve of the real data function

To solve equation (2.29) Yao et al. (2012) extended the modal expectation-maximization (MEM) proposed by Lee and Oh (2007). Similar to an expectation-maximization (EM) algorithm, the MEM algorithm also consists of two steps: E-step and M-step and requires one to iterate the E-step and the M-step until the algorithm converges.
The corresponding penalty parameter λ is determined using GCV criterion for the LS fitting and M type estimation methods, the regularized GCV criterion for the S -type estimation method and

in the proposed method choosing the appropriate value from selected values nominated using the robust generalized cross-validation (RGCV) to choose the optimum value. As for the bandwidth parameter h, it will be based on Yao et al (2012) to select the optimal value from a set of the following parameter:

$$h_l = 0.5\hat{\sigma} \times 1.02^{(l-1)}, l = 1, 2, \ldots.., L \tag{3.3}$$

The optimal value for h is determined to correspond to the lowest value of a scale $\mathrm{r(h)}$

$$r(h) = \frac{\widehat{G}(h)\hat{F}(h)^{-2}}{\hat{\sigma}^2} \tag{3.4}$$

Where

$$\hat{F}(h) = \frac{1}{n}\sum_{i=1}^{n} \phi_h''\left(\hat{\varepsilon}_i\right) \tag{3.5}$$

$$\hat{G}(h) = \frac{1}{n}\sum_{i=1}^{n} \left\{\phi_h'\left(\hat{\varepsilon}_i\right)\right\}^2 \tag{3.6}$$

Algorithm:

The following steps explain our algorithm:
1- Put $\ell = 1$ and repeat the following steps until L number of iterations is reached.
2- Specify the number of i values nominated for the smoothing parameter (which will be referred to as the symbols $\lambda_1, \lambda_2, \ldots\ldots, \lambda_1$)
3- Iterate the following steps for each parameter's candidate value for the smoothing parameter λ_i.
3-a) Obtain R number of initial estimates for the parameters vector indicated by the β -symbol $\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \ldots\ldots, \hat{\beta}_R^{(0)}$
3-b) Calculate the value of the objective function shown in (37) for each vector $\hat{\beta}_r^{(0)}$, and then determine the initial value of smoothing parameter h, using the Silverman rule Weglarczyk (2018):

$$h_0 = 1.06 \times \hat{\sigma} \times n^{-\frac{1}{5}} \tag{3.7}$$

3-c) Choose the vector of estimates corresponding to the largest value of the objective function $\hat{\beta}^{(0)}$, use this vector to calculate the corresponding residuals and standard deviation and then specify the value for the parameter h_l.

$$\mathrm{h}_1 = 0.5\widehat{\sigma} \times 1.02^{(1-1)}, \mathrm{l} = 1, 2, \ldots\ldots, \mathrm{L} \tag{3.8}$$

3-d) Select the following weights for the observations

$$\pi\left(j \mid \hat{\beta}^{(0)}\right) = \frac{\phi_{h_l}\left(y_j - \boldsymbol{x}_j^T\hat{\beta}^{(0)}\right)}{\sum_{i=1}^{n}\phi_{h_l}\left(y_i - x_i^T\hat{\beta}^{(0)}\right)}, j = 1, 2, \ldots n \tag{3.9}$$

Thus, the diagonal elements W is determined by these diagonal weights. 3-e) Calculate the value of the robust GCV criterion

$$RGCV_\lambda = \frac{n\left\|W^{1/2}\left(y - X\hat{\beta}^{(0)}\right)\right\|^2}{(n - \mathrm{trace}(H(\lambda)))^2} \tag{3.10}$$

Where

$$H(\lambda) = W^{\frac{1}{2}} X \left(X^T W X + \lambda D\right)^{-1} X^T W^{\frac{1}{2}} \tag{3.11}$$

4- Choose a value for λ corresponding to the smallest value of a criterion $RGCV_\lambda$, put $\mathrm{k} = 0$ and repeat the next three steps
(I) Step E : The weights are calculated in this step using

$$\pi\left(j \mid \beta^{(k)}\right) = \frac{\phi_{h_l}\left(y_j - \mathrm{x}_j^T \beta^{(k)}\right)}{\sum_{i=1}^{n} \phi_{h_l}\left(y_i - x_i^T \beta^{(k)}\right)} \propto \phi_{h_l}\left(y_j - \mathrm{x}_j^T \beta^{(k)}\right), \quad j = 1, \ldots\ldots n \tag{3.12}$$

(II) Step M : In this step, we update the parameters estimator vector using weighted ridge regression

$$\beta^{(\mathrm{k}+1)} = \left(\mathrm{X}^{\mathrm{T}} \mathrm{W}_{\mathrm{k}} \mathrm{X} + \lambda \mathrm{D}\right)^{-1} \mathrm{X}^{\mathrm{T}} \mathrm{W}_{\mathrm{k}} \mathrm{y} \tag{3.13}$$

(III) If $\left\|\widehat{\beta}^{(\mathrm{k})} - \hat{\beta}^{(\mathrm{k}+1)}\right\| < \epsilon \left\|\hat{\beta}^{(\mathrm{k})}\right\|$ or $k =$ Itermax, terminate and output the convergent coefficient vector and calculate $\mathrm{r}\left(\mathrm{h_i}\right)$ modify l to $\mathrm{l} + 1$, if 1 is reached to L go to step 5 else go to Step 1.
5- Choosing the corresponding h_i value for the smallest $r\left(h_i\right)$ to be the best value for the robustness parameter h_b, specifying the corresponding smoothing parameter value λ_b and determining the vector of the final predicted values $\hat{y}_b$

$$\hat{y}_b = X\left(X^T W_{h_b} X + \lambda_b D\right)^{-1} X^T W_{h_b} y \tag{3.14}$$

4. Performance Evaluation

This section is devoted to the comparative experiments upon both simulated data and one real weather balloon data set for the foregoing explored penalized spline smoothing methods, including the non-robust LS presented by Mallows (1995) and two robust smoothing approaches, the M -type estimator by Lee and Oh (2007) and S - estimation by Tharmaratnam et al. (2010) and the proposed method as described in Algorithm for penalized regression splines.

Configuration and Criteria:

It can obviously be said that for all three robust methods an exponential order $\mathrm{p} = 3$ for truncated power base will be applied and the knot locations will be chosen by Ruppert et al. $(2003), \kappa_k = ((k+1)/(\mathrm{k}+2))^{\mathrm{th}}$ sample quantile of the Unique(x), $k = 1, \ldots, K$; together with a simple choice of the knot number $K = \min(35, (1/4)\times$ Number of Unique(x)), where $x = \left(x_1, x_2, \ldots, x_n\right)^T$. the corresponding penalty parameter λ is determined according to GCV criterion for the LS fitting and M - estimation methods and the RGCV criterion for the S -estimation method and for the proposed method. Meanwhile, the termination tolerance $\varepsilon = 10^{-6}$ and the maximum iteration numbers 100 are allocated randomly to the exploration of the ultimate optimum. In the meantime, ACCER 8700 (CPU: Intel Core i7-1070 @ $8\mathrm{GHz}$, R-4.0.3.tar.gz built on Windows 10 Ultimate, are the hardware and software environment for the experiments below (64 Bit).

Simulated Data.

To test the method proposed in this article, we have created a function that exhibits various kinds of overall pattems blurred by different examples of noise. In Lee and Oh (2007) and Tharmaratnam et al. (2010) the simulated analysis is presented the same way Wang et al. (2014).

The settings for the simulation analysis are made for the different sample dimensions in design variables $\mathrm{x}_1 \ldots, \mathrm{x}_\mathrm{n}$ from the uniform distribution at interval $[-1,1]$. For all settings, these values are maintained to reduce the uncertainty of the simulation. $\mathrm{N} = 25, 100$, and 250 sample sizes were used. We used the following functions for the mean structure which describe a number of functions, which represent a variety of shapes:
$f_1(x) = \sin(\pi x), f_2(x) = \sin\left(2\pi(1-x)^2\right), f_3(x) = x + x^2 + x^3 + x^4$ and $f_4(x) = -20 + e^{3x}$. Function $f_2(x)$ is the same one used by Lee and Oh (2007) to facilitate a comparison with the results presented therefor, the error distribution we used five possibilities, ordered according to the heaviness of their tails: (i) uniform distribution $(-1,1)$, (ii) normal distribution $\mathrm{N}(0, 0.72)$, (iii) logistic distribution $(0,1)$, (iv) slash distribution, defined as $\mathrm{N}(0,1)/$ uniform $(0,1)$, and (v) Cauchy distribution $(0,1)$. Both cauchy and slash distributions are heavy tailed. We compare four penalized regression splines estimation methods in this simulation study: (A) the non-robust method for penalized regression splines estimation as in equation (9), using LS method, (B) penalized M-type regression splines estimators as studied by Lee and Oh (2007), (C) penalized Stype regression splines estimators as studied by Tharmaratnam et al. (2010) and (D) The proposed method in this article, using penalized modal regression splines estimators, and employing algorithm as described in Section 3 . For all four methods, we use truncated cubic splines $(\mathrm{p} = 3)$ with $\mathrm{K} = 6, 25$, or $35\mathrm{knots}$ (corresponding to sample sizes $25, 100$, and 250), spread equally according to the quantiles of the data. We have tried with different choices of K as well (results not shown) and found similar results. The penalty parameter λ is chosen by minimizing the GCV criterion for the LS estimation method. Robust cross-validation (RCV) is used for the M-type regression splines estimation method as proposed by Lee and Oh (2007). RGCV defined in (Section 2) is used for the S-type estimation method Tharmaratnam et al. (2010) and for the proposed method, we set the tolerance level in the three robust methods to 10^{-6} and the maximum number of iterations was set to 500 . To examine the robustness of the approaches against outliers, we randomly generated different percentages of outliers $(5\%, 10\%, 20\%$ and $30\%)$ for each of the two cases, first to get scattered outliers in normal distribution with mean 20 and standard deviation 20, and second to get concentrated outliers in normal distribution with mean 20 and standard deviation 2. The results of the estimation of the penalized spline regression model using the S estimator and the proposed mode estimator are based on the quality of the initial values of robustness's parameter h. Only 50 values have been used for the parameter h in the proposed algorithm to reduce the time needed to perform the current simulation study, the 50 values were specified using the equation of Yahoo et al. (2012):

$$h = 0.5\hat{\sigma} \times 1.02^j, j = 0, \ldots, 49 \tag{4.1}$$

Simulation Results

The goodness of fit of the estimated model is measured by calculating the median of mean squared errors (MSE) and median absolute deviation of mean squared errors. Denoting $\hat{f}_j\left(x_i\right)$ the estimated 4^+ value of $f\left(x_i\right)$ for simulation run j $(j = 1 \ldots . J = 1000)$, mean squared error (MSE) is defined by

$$MSE_j = \frac{1}{n}\sum_{i=1}^{n}\left(f\left(x_i\right) - \hat{f}_j\left(x_i\right)\right)^2, j = 1, 2, \ldots.., J \tag{4.2}$$

Table 1 presents summary values of the MSE (median and median absolute deviation) for the four estimation methods for the normal error distribution and with mean function $f_1(x)$. In all sample sizes, the median MSE of the proposed method is smaller than that of the other three methods for samples with more than 5% of outliers except in sample size of $n = 100$ with 5% of outliers method of penalized M-type regression splines estimation works better. For LS and penalized M-type regression splines estimators, MSE is clearly increasing with the percentage of outliers increasing. For penalized S-type regression splines estimation, the MSE values tend to be quite stable, the MSE of the proposed method tend to be quite stable with small values. As expected, the goodness of fit as measured by the MSE values improves for larger sample sizes with all percentages of contamination. Table 1 clearly shows that the penalized least squares method may already break down with only 5% of outliers. For penalized M-type regression splines estimation, the breakdown arrives earlier, showing the need for taking the scale into consideration in the fitting method and working with a bounded ρ -function, a clearer increase (breakdown) is observed for the penalized S-type regression splines estimation method when the presence of outliers. For the proposed method of penalized mode-regression splines estimation, the MSE values are relatively small even with 30% of scattered outliers for all sample sizes a clearer increase (breakdown) is observed for the proposed method when the presence of outliers reaches 30% of the sample size. The median of absolute deviations is still at the same level as before, the LS-estimator's MSE grow very rapidly. Similarly, the penalized M-type regression splines estimator's MSE grow rapidly after 10% of outliers. These results are confirming that the penalized M-type regression splines estimation method works better with less than 10% of outliers, while the penalized S-type regression splines estimation method increases but is much less than the penalized M, while the proposed method works well for all considered percentages of outliers.

Figure 3: TABLE 1 Median and median absolute deviation (between parentheses) of MSE for data generated with mean structure $f_1(x)$, error terms from $N(0, 0.72)$ distribution, and for different sample sizes. from $N\left(20, 2^2\right)$

ϵ	$f_2(x)$				$f_3(x)$				$f_4(x)$			
	LS	M	S	Mode	LS	M	S	Mode	LS	M	S	Mode
0%	0.08	0.17	0.33	0.29	0.02	0.03	0.32	0.15	0.05	0.06	0.32	0.17
	(0.03)	(0.07)	(0.14)	(0.13)	(0.02)	(0.02)	(0.14)	(0.08)	(0.02)	(0.03)	(0.14)	(0.08)
5%	1.87	0.19	0.35	0.09	1.41	0.04	0.35	0.04	5.73	0.07	0.35	0.06
	(1.2)	(0.08)	(0.16)	(0.03)	(1.13)	(0.03)	(0.16)	(0.03)	(4.08)	(0.03)	(0.15)	(0.03)
10%	5.42	0.25	0.36	0.09	4.79	0.06	0.35	0.04	18.21	0.09	0.36	0.07
	(3.05)	(0.11)	(0.16)	(0.03)	(2.86)	(0.04)	(0.16)	(0.03)	(10.21)	(0.05)	(0.16)	(0.03)
20%	17.47	0.51	0.41	0.14	16.44	0.18	0.4	0.04	59.31	0.25	0.37	0.07
	(6.91)	(0.27)	(0.24)	(0.08)	(6.72)	(0.12)	(0.25)	(0.03)	(23.54)	(0.15)	(0.2)	(0.05)
30%	38.33	1.74	0.4	0.36	36.09	0.76	0.29	0.04	130.94	1.04	0.28	0.16
	(11.64)	(1.28)	(0.32)	(0.08)	(11.17)	(0.58)	(0.28)	(0.03)	(40.19)	(0.85)	(0.2)	(0.13)

We have further checked our proposed method with the penalized S-type regression splines estimation method and the penalized M-type regression splines estimation method. We generated errors ε_i from a normal distribution and included different percentages of outliers for sample size $\mathrm{n} = 100$. For each of these settings we computed the MSE over 1000 simulation runs; the results are presented in Table 2 . The proposed method gives the smallest median MSE values for all data generated from functions goniometric $f_2(x)$, the polynomial $f_3(x)$, and the exponential $f_4(x)$ mean functions if there are 5% and more of outliers. The penalized S has shown no relative superiority over the penalized M unless with 30% contamination and the penalized M-type regression splines estimation method works better than the penalized S for the cases with 5% and 10% of outliers in all generation functions $f_2(x)$, $f_3(x)$, and $f_4(x)$.

Figure 4: TABLE 2 Median and median absolute deviation (between parentheses) of MSE for data generated from functions $f_2(x)$, $f_3(x)$, and $f_4(x)$ with error terms from $N(0, 0.72)$ for sample size $n = 100$ with different percentages ε of outliers generated from $N\left(20, 2^2\right)$

ε	$f_2(x)$				$f_3(x)$				$f_4(x)$			
	LS	M	S	Mode	LS	M	S	Mode	LS	M	S	Mode
0%	0.08	0.17	0.33	0.29	0.02	0.03	0.32	0.15	0.05	0.06	0.32	0.17
	(0.03)	(0.07)	(0.14)	(0.13)	(0.02)	(0.02)	(0.14)	(0.08)	(0.02)	(0.03)	(0.14)	(0.08)
5%	1.87	0.19	0.35	0.09	1.41	0.04	0.35	0.04	5.73	0.07	0.35	0.06
	(1.2)	(0.08)	(0.16)	(0.03)	(1.13)	(0.03)	(0.16)	(0.03)	(4.08)	(0.03)	(0.15)	(0.03)
10%	5.42	0.25	0.36	0.09	4.79	0.06	0.35	0.04	18.21	0.09	0.36	0.07
	(3.05)	(0.11)	(0.16)	(0.03)	(2.86)	(0.04)	(0.16)	(0.03)	(10.21)	(0.05)	(0.16)	(0.03)
20%	17.47	0.51	0.41	0.14	16.44	0.18	0.4	0.04	59.31	0.25	0.37	0.07
	(6.91)	(0.27)	(0.24)	(0.08)	(6.72)	(0.12)	(0.25)	(0.03)	(23.54)	(0.15)	(0.2)	(0.05)
30%	38.33	1.74	0.4	0.36	36.09	0.76	0.29	0.04	130.94	1.04	0.28	0.16
	(11.64)	(1.28)	(0.32)	(0.08)	(11.17)	(0.58)	(0.28)	(0.03)	(40.19)	(0.85)	(0.2)	(0.13)

Next, we measure the effects of the various error distributions on the performance of the estimates. The results are shown in Table 3 for sample size $\mathrm{n} = 100$ and true mean function $f_1(x)$. The proposed method in uniform distribution gives the smallest median MSE values in all percentages of outliers, only in one case is 5% contamination the results of the proposed method and the penalized M-type regression splines estimation were similar. The penalized S-type regression splines estimation showed a relative superiority over the penalized M only at 30% contamination. The proposed method in logistic and slash distribution gives the smallest median MSE values with data contamination from 10% to the 30%. While the penalized M-type regression splines estimation works better for the samples with 5%. Also, the proposed method in caushy distribution gives the smallest median MSE values with data contamination from 20% to the 30%, while in 10% contamination the results of the proposed method and the penalized M-type regression splines estimation were similar, and the penalized M gives the smallest median MSE values with data contamination 5%.

Next, we investigate the robustness of the methods against two types of outliers, simulated cases with scattered outliers or with concentrated cloud of outliers. We generated data from scattered outliers, simulated using normal distribution with mean 20 and standard deviation 20 and concentrated cloud, simulated using normal distribution with mean 20 and standard deviation 2., and included different percentages of outliers for sample size $\mathrm{n} = 100$. For each of these settings we computed the MSE over 1000 simulation runs; the results are presented in Table 4. Regardless

Figure 5: TABLE 3 Median and median absolute deviation (between parentheses) of MSE for data generated with mean structure $f_1(x)$, error terms from different distributions, and for sample sizes $n = 100$ with different percentages ε of outliers generated from $N(20, 2^2)$

ϵ		0%		5%		10%		20%		30%	
uniform	LS	0.02	(0.01)	1.52	(1.22)	5.17	(2.77)	18.13	(6.94)	38.1	(12.38)
	M	0.02	(0.01)	0.03	(0.02)	0.05	(0.04)	0.16	(0.1)	0.71	(0.58)
	S	0.3	(0.08)	0.31	(0.09)	0.32	(0.1)	0.32	(0.12)	0.28	(0.18)
	Mode	0.21	(0.06)	0.03	(0.02)	0.03	(0.02)	0.03	(0.02)	0.04	(0.04)
logistic	LS	0.16	(0.1)	1.58	(1.25)	5.29	(3)	17.97	(6.84)	38.19	(12.58)
	M	0.16	(0.09)	0.2	(0.11)	0.31	(0.2)	1.08	(0.7)	4.62	(3.79)
	S	1.86	(0.85)	1.99	(0.96)	2.37	(1.51)	4.09	(4.58)	23.2	(32.83)
	Mode	0.35	(0.2)	0.23	(0.15)	0.23	(0.15)	0.2	(0.12)	0.22	(0.14)
slash	LS	6.92	(8.78)	8.23	(10.02)	11.87	(11.86)	25.17	(18.91)	45.74	(24.21)
	M	0.28	(0.17)	0.4	(0.27)	0.63	(0.47)	2.72	(2.15)	15.27	(14.2)
	S	3.56	(2.55)	4.09	(3.22)	5.19	(5.34)	18.66	(24.91)	58.58	(69.86)
	Mode	0.32	(0.2)	0.36	(0.25)	0.4	(0.33)	0.52	(0.51)	1	(1.24)
cauchy	LS	4.24	(5.42)	5.62	(6.26)	9.7	(9.22)	23.59	(15.26)	41.84	(20.63)
	M	0.17	(0.1)	0.23	(0.14)	0.36	(0.26)	1.57	(1.22)	7.89	(7.45)
	S	1.67	(1.2)	1.93	(1.75)	2.48	(2.64)	4.33	(6.2)	27.33	(40.17)
	Mode	0.24	(0.17)	0.26	(0.19)	0.31	(0.24)	0.34	(0.29)	0.46	(0.48)

of the type of contamination whether it is scattered or concentrated. The proposed method has the smallest median MSE values with data contamination from 10% to the 30%. While the M-type estimation works better for the samples with 5%. The order of M-type and S-type estimation magnitude varied according to the type of contamination at 30%, with the order of M-type being the third in the case of concentrated contamination and the second in scattered, while the S-type was second in concentrated contamination and third in the case of scattered contamination. Table

Figure 6: TABLE 4 Median and median absolute deviation (between parentheses) of MSE for data generated from Scattered outliers and Concentrated cloud for sample size $n = 100$.

	Concentrated cloud				Scattered outliers			
ϵ	LS	M	S	Mode	LS	M	S	Mode
0%	0.03	0.03	0.32	0.16	0.03	0.03	0.32	0.16
	(0.02)	(0.02)	(0.13)	(0.07)	(0.02)	(0.02)	(0.13)	(0.07)
5%	1.56	0.04	0.36	0.04	1.74	0.03	0.35	0.04
	(1.22)	(0.02)	(0.17)	(0.03)	(1.79)	(0.02)	(0.16)	(0.03)
10%	5.17	0.06	0.35	0.04	5.79	0.05	0.37	0.05
	(3.09)	(0.04)	(0.16)	(0.03)	(4.66)	(0.03)	(0.17)	(0.03)
20%	17.43	0.2	0.39	0.04	19.34	0.12	0.43	0.08
	(6.96)	(0.13)	(0.24)	(0.03)	(11.43)	(0.08)	(0.27)	(0.07)
30%	38.35	0.83	0.3	0.06	42.66	0.37	0.42	0.25
	(11.85)	(0.67)	(0.27)	(0.05)	(19.71)	(0.27)	(0.39)	(0.13)

5 summarizes the results of the simulation study of the effect of increasing the number of initial values and the number of smoothing parameter values on the performance of the four estimation methods $(\mathrm{LS}, \mathrm{M}, \mathrm{S},$ and Mode). 350 ”Run” returns have been performed It is clear that the order

Figure 7: Table 5: Median and median absolute deviation (between parentheses) of MSE for data generated with from the estimation methods $(\mathrm{LS}, \mathrm{M}, \mathrm{S},$ and Mode) in the cases of small and large number of initial values and candidate values for a smoothing parameter.

ϵ	A small number of initial and candidate values for a smoothing parameter				A large number of initial and candidate values for a smoothing parameter			
	LS	M	S	Mode	LS	M	S	Mode
0%	0.03	0.03	0.32	0.16	0.03	0.03	0.46	0.16
	(0.02)	(0.02)	(0.14)	(0.08)	(0.02)	(0.02)	(0.32)	(0.08)
5%	1.56	0.04	0.36	0.04	1.57	0.04	0.46	0.03
	(1.25)	(0.03)	(0.17)	(0.03)	(1.25)	(0.03)	(0.29)	(0.02)
10%	5.20	0.06	0.36	0.04	5.21	0.06	0.45	0.03
	(3.05)	(0.04)	(0.16)	(0.03)	(3.06)	(0.04)	(0.27)	(0.02)
20%	17.14	0.19	0.41	0.04	17.22	0.19	0.45	0.04
	(6.66)	(0.13)	(0.29)	(0.03)	(6.63)	(0.13)	(0.30)	(0.02)
30%	38.27	0.83	0.32	0.05	38.27	0.83	0.37	0.04
	(12.77)	(0.67)	(0.28)	(0.05)	(12.67)	(0.67)	(0.29)	(0.02)

of the four methods in terms of performance quality did not vary between small and large number of initial values and the number of smoothing parameter values except in two cases, corresponding to the contamination ratios $(5\%$ and $30\%)$. If a small number of values were used at 5% contamination, the proposed method and M-type estimator had the same performance. It is also clear from a table 5 that both estimates have the same median value and the same absolute deviation value as the median, while the mode estimate rank becomes 1 and the m-rated rank equals 2 if the number of values increases, which means that there is an improvement in the estimated performance of the mode in that case. As for the 30% contamination rate, it can be noted that the two different estimates are S and M where the two values had the same order if the number of values was small, and the order of 3 and 2 , if the number of values was increased. We can summarize some general results of the simulation study to use as a practical guidance for finding a specific method to pulsing smoothing noisy data:

1. the best estimate of the four comparable estimates in the case of data contamination is 10% to 30%, i.e., if data is moderately or significantly contaminated, the expression is expressed.
2. M is the best estimate in case of small data contamination (5%).
3. LS is best estimated if data is not polluted (0% contamination).
4. The proposed method is generally best estimated in the simulation study, where it was with the lowest ranking of the general average.

Balloon data

In this section, we have used the balloon dataset from the R software's library ftnonpar. The data are radiation measurements from the sun, taken from a flight of a weather balloon. Due to the rotation of the balloon, or for some other reasons, outliers were introduced because the measuring device was occasionally blocked from the sun. The response variable Y is a radiation measurement, and the explanatory variable x is the index of the measurement. The sample size equals 4984. Displayed in Figure 8 are regression estimates obtained by LS method, M-type penalized regression splines estimation, S-type penalized regression splines estimation and our proposed method of Mode penalized regression splines estimation. For the time $0.3 - 0.9$, the non-robust curve estimate of LS suffers from the presence of the outliers. That is, the estimated curve was pulled downwardly away from the majority of the observations, which is clearly visible around the value of $x = 0.8$. On the other hand, this phenomenon does not affect all robust methods, including our proposed method.

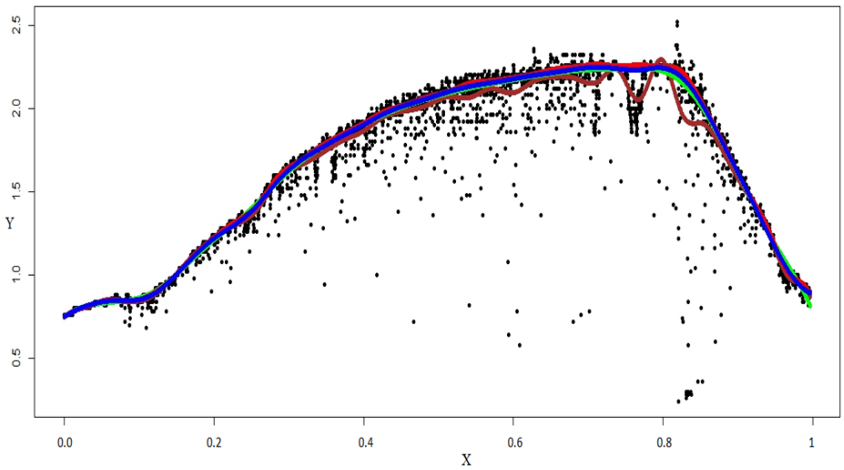


Figure 8: Fitted values for the balloon data. LS -type regression splines method (brown), penalized S-type regression spline method (red), penalized M-type regression splines method (green) and mode penalized regression(blue)

5. Concluding Remarks

In this article an effective method is proposed for robust fitting penalized regression splines models. Generally, smoothing methods may be influenced by outliers. The proposed method is easy to implement. Mode penalized regression splines estimators superior to least squares penalized regression splines, M-type penalized regression splines estimators and S-type penalized regression splines estimators in moderate or high contamination percentage. The procedure performs very well in many scenarios of simulated data and real data. The penalized M-type regression splines estimation works better for the cases with a small percentage of contamination while penalized S-type regression splines estimation works well for higher percentages of contamination, but mode penalized regression splines estimation works better for all percentages of contamination. In the absence of outliers, the efficiency of the proposed method is not very high, that is the price to pay for a high robustness. The asymptotic properties of mode penalized regression have not yet been studied and are a topic of our further research. We expect that consistency, and asymptotic normality still hold, under appropriate regularity conditions. These results would be useful, for example to construct confidence bands for the curves.

Supplemental Materials R-code:

We make the R code available with functions implementing the mode penalized regression s-plines estimators used in this article through the journal's supplementary materials. The file also in-cludes the code that we used for a comparison with penalized least squares penalized M-estimation and S-estimation methods.

Acknowledgement: The corresponding author would like to express their gratitude to King Khalid University, Saudi Arabia for providing administrative and technical support.

References

[1] E. J. Wegman, , and I. W. Wright, . *Splines in statistics*. Journal of the American Statistical Association, 78(382), (1983)351-365.

[2] S. N. Wood, . *Modelling and smoothing parameter estimation with multiple quadratic penalties*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(2),(2000)413-428.

[3] Y. J. Kim, , and C. Gu, . *Smoothing spline Gaussian regression: more scalable computation via efficient approx-imation*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(2),(2004)337-356.

[4] F. O'Sullivan, . *A statistical perspective on ill-posed inverse problems*. Statistical science,(1986) 502-518.

[5] C. Kelly, and J. Rice, . *Monotone smoothing with application to dose-response curves and the assessment of synergism*. Biometrics,(1990) 1071-1085.

[6] B. D. Marx, and P. H. Eilers, . *Direct generalized additive modeling with penalized likelihood*. Computational Statistics and Data Analysis, 28(2),(1998) 193-209.

[7] M. P. Wand, . *Miscellanea. On the optimal amount of smoothing in penalized spline regression*. Biometrika, 86(4), (1999)936-940.

[8] D. Ruppert, M. P. Wand, and R. J. Carroll, . *Semiparametric regression (No. 12)*. Cambridge university press(2003).

[9] S.Wold, M. Sjstrm, and L. Eriksson, . *PLS-regression: a basic tool of chemometrics*. Chemometrics and intelli-gent laboratory systems, 58(2), (2001)109-130.

[10] K. Tharmaratnam, G. Claeskens, C. Croux, and M. Salibin-Barrera, . *S-estimation for penalized regression splines*. Journal of Computational and Graphical Statistics, 19(3),(2010) 609-625.

[11] D. D. Cox, . *Asymptotics for M-type smoothing splines*. The Annals of Statistics,(1983) 530-551.

[12] T. C. Lee, and H. S. Oh, . *Robust penalized regression splines fitting with application to additive mixed modeling*. Computational Statistics, 22(1),(2007) 159-171.

[13] P. H. Eilers, and B. D. Marx, . *Flexible smoothing with B-splines and penalties*. Statistical science,(1996) 89-102.

[14] B. Wang, W. Shi, and Z. Miao, . *Comparative analysis for robust penalized spline smoothing methods*. Mathe-matical problems in engineering(2014).

[15] P. J. Huber, . *Robust Statistics*. New York: John Wiley and Sons(1981).

[16] P. Rousseeuw, and V. Yohai, . *Robust regression by means of S-type estimators*. In Robust and nonlinear time series analysis (pp. 256-272). Springer, New York, NY(1984).

[17] M. Salibian-Barrera, and V. J. Yohai, . *A fast algorithm for S-type regression estimates*. Journal of computational and Graphical Statistics, 15(2),(2006) 414-427.

[18] H. S. Oh, D. Nychka, T. Brown, and P. Charbonneau, . *Period analysis of variable stars by robust smoothing*. Journal of the Royal Statistical Society: Series C (Applied Statistics), 53(1),(2004) 15-30.

[19] R. Finger, . *Investigating the performance of different estimation techniques for crop yield data analysis in crop insurance applications*. Agricultural Economics, 44(2),(2013) 217-230.

[20] C. L. Mallows, . *More comments on Cp*. Technometrics, 37(4),(1995) 362-372.

[21] D. Ruppert, M. P. Wand, and R. J. Carroll, . *Semiparametric regression during 20032007*. Electronic journal of statistics, 3,(2009) 1193.

[22] N. R. Draper, and H. Smith, . *Applied regression analysis (Vol. 326)*. John Wiley and Sons(1998).

[23] T. Krivobokova, . *Theoretical and practical aspects of penalized spline smoothing*(2006).

[24] P. Craven, and G. Wahba, . *Smoothing noisy data with spline functions*. Numerische mathematik, 31(4),(1978) 377-403.

[25] C. M. Hurvich, J. S. Simonoff, and C. L. Tsai, . *Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60(2),(1998) 271-293.

[26] R. R. Wilcox, . *Introduction to robust estimation and hypothesis testing*. Academic press(2011).

[27] W. Zhao, R. Zhang, J. Liu, and Y. Lv, . *Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression*. Annals of the Institute of Statistical Mathematics, 66(1),(2014) 165-191.

[28] W. Yao, and L. Li, . *A new regression model: modal linear regression*. Scandinavian Journal of Statistics, 41(3),(2014) 656-671.

[29] B. W. Silverman, . *Density estimation for statistics and data analysis (Vol. 26)*. CRC press(1986).

[30] S. Wglarczyk, . Kernel density estimation and its application. In ITM Web of Conferences (Vol. 23). EDP Sciences(2018).

[31] E. Cantoni, and E. Ronchetti, . *Resistant selection of the smoothing parameter for smoothing splines*. Statistics and Computing, 11(2),(2001) 141-146.

[32] W. Yao, B. G.Lindsay, and R. Li, . *Local modal regression*. Journal of nonparametric statistics, 24(3),(2012) 647-663.