Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Kernel partial diagnostic robust potential to handle high-dimensional and irregular data space on near infrared spectral data

Divo Dharma Silalahi [a,*], Habshah Midi [b], Jayanthi Arasan [b], Mohd Shafie Mustafa [b], Jean-Pierre Caliman [a]

[a] *SMART Research Institute, PT. SMART TBK, Riau, Indonesia*
[b] *Institute of Mathematical Research, Universiti Putra Malaysia, Serdang, Malaysia*

ABSTRACT

In practice, the collected spectra are very often composes of complex overtone and many overlapping peaks which may lead to misinterpretation because of its significant nonlinear characteristics. Using linear solution might not be appropriate. In addition, with a high-dimension of dataset due to large number of observations and data points the classical multiple regressions will neglect to fit. These complexities commonly will impact to multicollinearity problem, furthermore the risk of contamination of multiple outliers and high leverage points also increases. To address these problems, a new method called Kernel Partial Diagnostic Robust Potential (KPDRGP) is introduced. The method allows the nonlinear solution which maps nonlinearly the original input $\mathbf{X}$ matrix into higher dimensional feature mapping with corresponds to the Reproducing Kernel Hilbert Spaces (RKHS). In dimensional reduction, the method replaces the dot products calculation of elements in the mapped data to a nonlinear function in the original input space. To prevent the contamination of the multiple outlier and high leverage points the robust procedure using Diagnostic Robust Generalized Potentials (DRGP) algorithm was used. The results verified that using the simulation and real data, the proposed KPDRGP method was superior to the methods in the class of non-kernel and some other robust methods with kernel solution.

1. Introduction

The Near Infrared Spectroscopy (NIRS) technology as secondary measurement has been widely used in varies agricultural process. It requires less (even no) sample preparation, non destructive, rapid, and chemical-free with its great potential for product quality control and real time analytical chemistry process. The NIRS uses a spectrometer to collect a continuous absorption spectrum. This spectrum associates with the chemical and physical composition of the sample material. Practically, NIRS applies the multivariate statistics method on the spectral dataset to build the fitted model. The model is built by associating the spectral dataset measured from the known sample with a set of reference data. In routine chemical analysis, this model is applied in the analytical process to predict the concentration of the unknown samples.

The NIR spectral data is presented as a high dimensional matrix. It contains wide wavelengths ranges (350–2500nm: 1 or 2nm interval) as predictor variables and takes a huge number of samples in line with the increasing volumes of routine laboratory analysis. Normally, each

spectrum NIR has more than thousands data points. Based on experience, the collected spectra also may often composed of complex overtone and many overlapping peaks that result in broad NIR absorption bands [1]. These are caused by the physical properties, chemical variation, and the instrumentation which lead to misinterpretation on the spectra information. This makes the spectral data processing becomes more complicated. To encounter this, a well assign of multivariate calibration based on statistical methods are highly considered.

In multivariate calibration, the Partial Least Squares Regression (PLS) method [2] is the most used. The PLS has benefit to tackle the presence of multicollinearity and high dimensional in the dataset. In fact, in the application the method is still less inferior due to its efficiency and interpretability. The complexity in the NIR spectral dataset is also an issue, where it may cause the risk of involving multiple outliers and high leverage points increases. This may decrease the effectiveness in the computational. As solution, then the robust procedure is needed to apply with the PLS method.

The PLS works by projecting the **X** matrix to a smaller set of uncorrelated variables called new components with corresponds to **y** vector. It is known as a class of linear methods with the basic linearity assumptions holds both in the relation between observations and for modeling in the latent variables [3]. However, in the real NIR spectral dataset the sets of predictor matrix **X** and response matrix **y** could have a nonlinear relationship. This is due to the scattering effect and noise problems in the dataset. To overcome this, the nonlinear solution is suggested as alternative to discover the nonlinear behavior in the input space [4]. The idea is the input data matrix is mapped into high dimensional nonlinear feature space through a nonlinear mapping function. Next, these mapped data are used in the post process of data analysis. Several studies with different dataset using the beneficial of applying various types of nonlinear high-dimensional space mapping method have been discussed [see [4, 5, 6, 7, 8, 9]]. This has proven its superiority as the most efficient method than such linear method [8, 10]. Considering the existing nonlinear methods, the kernel is the most used due to its flexibility and efficiency in the computational process [4, 6, 7, 11, 12, 13]. Instantaneously, the kernel solution presented here apply the famous Reproducing Kernel Hilbert Spaces (RKHS) in the mapping formulation to reach the nonlinear optimization procedure [7, 14] with universally result to consistent result [15]. It has been observed so far there is still less number of study considers its robustness particularly with improvement on the resistance diagnostics to the multiple outliers and high leverage points.

The outlier and/or high leverage data points are results of the experimental and/or non experimental error. The outlier is an unwanted observation in the **y** direction, oppositely the high leverage data point is an outlier in the **X** direction. These have potential to influence the fitting performance that lead to the misleading result. Using the non-robust statistical methods is not recommended, since the estimated parameter is sensitive to the small changes in the data distribution. Hence, seeking for the robust procedure that may not suffer from the influence is needed.

In 1973, Huber introduced the M-estimator [16] that minimizes the function of residual in the least square estimation. This estimator is well known as robust location parameter estimator which uses $\ell_1 - norm$ for measuring residual in the observation that is identified as outlier. The M-estimator fails to attain high breakdown point and has unbounded influence factor hence it is not resistant to the leverage points [17, 18]. To improve the low breakdown point, another robust method called S-estimator [17] was proposed by minimizing the scale of residual in M-estimator. In fact, the method still fails to identify the leverage points due to its unbounded influence factor [18]. In compliance to improve the resistance through bounded influence, a Generalized M-estimator (GM-estimator) [19] was suggested. The GM-estimator is still suffered in a loss of efficiency [20] because observations with good leverage point are still down-weighted. Another GM-estimator using Schweppe solution [21] was proposed by adjusting the leverage down-weighted with the size of residual. It was observed, the proposed GM has also less breakdown point (closer to 0), when the number of m predictor increases [22] with some outliers more likely happen. As development, the GM6 was introduced [23] and it has the bounded influence and high breakdown-point (with closely to 50%). The GM6 has benefit in a distinction between outliers and bad high leverage observations. It uses Least Trimmed Square (LTS) [24] as initial estimator than classical Ordinary Least Square (OLS). To fit the limitations of previous robust estimators, the MM-estimator [17] was introduced. The MM-estimator reaches the robustness by discovering the smallest possible dispersion of the residual in S-estimator and continues to find the resistant estimator through the efficiency of the M-estimator [20]. Some robust high breakdown-point estimator with a good efficiency such Least Trimmed Square (LTS) [24] and Least Median Squares (LMS) [25] usually are used as initial estimator in MM-estimator. However, the earlier robust methods effective only to detect the outlier and single high leverage points. In fact, they fail to detect multiple high leverage cases by reason of masking and/or swamping effects on the low leverage some (see [16, 22,

26, 27, 28]). As solution, the Diagnostic Robust Generalized Potential (DRGP) (see [29]) is assigned by initially classify the suspects high leverage points and real outlier using the Robust Mahalanobis Distance (RM_i^2) based on Minimum Volume Ellipsoid (MVE) [30, 31]. The DRGP calculates the generalized potentials criteria to each observations whether the suspected observations have potentially high leverage points or not.

The method presented here is the extension work of the previous study (in [7]) that introduces the robust algorithm procedure in the nonlinear Kernel-PLS. The new method is named as a Kernel Partial Diagnostic Robust Potential (KPDRGP) with the parameter estimators are highly resistance to the existing of multiple outlier and high leverage points in the dataset. The study provides a development and important contribution to tackle the challenges of scientific big data especially in vibrational spectroscopy technique. In the paper, some existing robust methods are also reviewed and integrated in the PLS term. Some discussions are provided to compare the methods through simulation using different datasets.

2. Some consideration methods

2.1. Partial least square regression

Partial Least Squares Regression (PLSR) [2] is the most used statistical method in the Chemometric analysis. Particularly for NIR spectral dataset, the PLSR has an important role to investigate the relationship between two different sets of predictor **X** and response **y**. Here, we limited the study only in the case of $n >> m$, where m represents the number of predictor variable and significantly greater or less than the number of sample size n. Related to the high dimensional problem in the NIR dataset the PLSR becomes a standard procedure to decrease the dimensional of predictor variables **X**. Technically, the predictor **X** is projected to a smaller set of uncorrelated variables called new components. The PLSR keeps a maximize covariance of the highly collinear original predictors in these new components and regress these to the single **y**. PLS has free distribution assumption means it does not matter whatever data distribution is, it is also opposed violations of independence, collinear, and small sample size that are known as main assumptions in regression.

Let consider a multiple regression model consists two different sets of predictor **X** and single response **y**, in matrix form these can be written as

$$\mathbf{y} = \mathbf{X}\,\mathbf{b}\, + \,\mathbf{e} \tag{1}$$

where **y**,**e** are $n \, x \, 1$ vector, **X** is $n \, x \, m$ matrix and **b** is $m \, x \, 1$ vector, the common solution of estimator **b** using OLS method is given by

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{2}$$

Earlier it considers the NIR dataset consisting large number of m predictors. So that there will infinite number of solution for estimator **b** as $\mathbf{X}^T\mathbf{X}$ is singular which not meet the usual trivial theorem on rank. As a special procedure to this case, it is needed to extract new components by maximizing a covariance criterion between predictor **X** and response **y** where linking the centered values of these two sets [32].

Using the Eq. (1) and initialize a starting score vector of **u** from single **y**, there exists an outer relation for predictor **X** which can be defined as

$$\mathbf{X} = \mathbf{V}\mathbf{P}^T\, + \,\mathbf{E} \tag{3}$$

where **P** is the matrix $m \, x \, l$ consisting loading vector $1 \, x \, m \{\mathbf{p}_g = (\mathbf{X}^T\,\mathbf{v}_g)/(\mathbf{v}_g{}^T\mathbf{v}_g)\}_{g=1}^l$ with $\mathbf{v}_g$ is the $n \, x \, 1$ column vector of scores $\mathbf{x}_j$ in **X** $\{\mathbf{v}_g = (\mathbf{X}\mathbf{w}_j)/(\mathbf{w}_j^T\mathbf{w}_j)\}_{g=1}^l$ involve $\mathbf{w}_j$ as a $m \, x \, 1$ vector of weight for **X** $\{\mathbf{w}_j = (\mathbf{X}^T\mathbf{u})/(\mathbf{u}^T\mathbf{u})\}_{j=1}^m$, **V** is a $n \, x \, l$ matrix of the $n \, x \, 1$ vector $\mathbf{v}_g$ and **E** is a $n \, x \, m$ matrix of residual in outer relation for predictor

X. Unlike in Principal Component Analysis (PCA), $\mathbf{v}_g$ is not simply as multiplication of the matrix $\mathbf{X}$ and loading matrix $\mathbf{P}$, a correspondence to $m \, x \, l$ weight matrix of $\mathbf{W}$ is considered. Following the similar procedure as like predictor $\mathbf{X}$, the outer relation for response $\mathbf{y}$ also can be given as

$$\mathbf{y} = \mathbf{u} \, \mathbf{q}^T + \mathbf{f} \tag{4}$$

In Eq. (4), variable $\mathbf{y}$ is in univariate term, $\mathbf{q}$ is the loading $l \, x \, 1$ vector$\{\mathbf{q}_g = (\mathbf{y}^T \mathbf{v}_g)/(\mathbf{v}_g^T \mathbf{v}_g)\}_{g=1}^l$, $\mathbf{u}$ is $n \, x \, l$ matrix of $\mathbf{y}$ block score, and $\mathbf{f}$ is a $n \, x \, 1$ vector of residual in $\mathbf{y}$. $\mathbf{u}$ is also called as linear inner relation between $\mathbf{X}$ block score and $\mathbf{y}$ block score which is calculated as $\{\mathbf{u} = \mathbf{b}_g \mathbf{v}_g$ with $\mathbf{b}_g = \mathbf{u}^T \mathbf{v}_g/(\mathbf{v}_g^T \mathbf{v}_g)\}_{g=1}^l$ or written as

$$\mathbf{u} = \mathbf{V} \, \mathbf{b}_{inner} + \mathbf{g} \tag{5}$$

where $\mathbf{b}_{inner}$ is a $l \, x \, 1$ vector of regression coefficient as solution using LS on the decomposition of vector $\mathbf{u}$, and $\mathbf{g}$ is $n \, x \, 1$ matrix of residual in the inner relation. Applying the normalization in $\mathbf{W}$, $\mathbf{q}$, $\mathbf{P}$ as the process for improving the inner relation and obtaining orthogonal $\mathbf{X}$ block scores, so that the mixed relation of PLSR model by integrating Eq. (4) and Eq. (5) can be formed as [2].

$$\begin{aligned} \mathbf{y} &= \mathbf{u} \, \mathbf{q}^T + \mathbf{f} \\ \mathbf{y} = \mathbf{V} \, \mathbf{b}_{inner} \, \mathbf{q}^T &+ \mathbf{g} \, \mathbf{q}^T + \mathbf{f} \\ \mathbf{y} = \mathbf{V} \, \mathbf{b}_{inner} \, \mathbf{q}^T &+ \breve{\mathbf{f}} \\ \mathbf{y} = \mathbf{V} \, \mathbf{a}^T &+ \breve{\mathbf{f}} \end{aligned} \tag{6}$$

the $l \, x \, 1$ vector coefficient $\mathbf{a}$ is $\mathbf{a}^T = \mathbf{b}_{inner} \, \mathbf{q}^T$ and $\breve{\mathbf{f}}$ denotes $n \, x \, 1$ vector of residual in mixed relation $\breve{\mathbf{f}} = \mathbf{g} \, \mathbf{q}^T + \mathbf{f}$. In Eq. (6) it holds $\mathbf{a} = \mathbf{V}^T \mathbf{y}$, without loss of generality that $\mathbf{X} = \mathbf{V}\mathbf{P}^T$ as in Eq. (3), then Eq. (6) can be redefined by multiplying the two sides with $m \, x \, l$ weight matrix of $\mathbf{W}$ that results

$$\begin{aligned} \mathbf{y} &= \mathbf{X} \, \mathbf{W}^* \, \mathbf{a} + \breve{\mathbf{f}} \\ \mathbf{y} &= \mathbf{X} \, \mathbf{W} \, (\mathbf{P}^T \mathbf{W})^{-1} \, \mathbf{a} + \breve{\mathbf{f}} \end{aligned} \tag{7}$$

with $\mathbf{V} = \mathbf{X} \, \mathbf{W}^*$ and $\mathbf{W}^* = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1}$. Let define $\mathbf{b}_{PLSR} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1}\mathbf{a}$ as matrix coefficient of mixed relation in PLSR, and then Eq. (7) is equivalent to

$$\mathbf{y} = \mathbf{X} \, \mathbf{b}_{PLSR} + \breve{\mathbf{f}} \tag{8}$$

$\breve{\mathbf{f}}$ need to be minimized. Applying the previous relation Eq. (3) and Eq. (4), it holds $\mathbf{W} = \mathbf{X}^T \mathbf{U}$, $\mathbf{P} = \mathbf{X}^T \mathbf{V}(\mathbf{V}^T \mathbf{V})^{-1}$ such that the estimator for parameter $\mathbf{b}_{PLSR}$ can be defined as

$$\widehat{\mathbf{b}}_{PLSR} = \mathbf{X}^T \mathbf{u} \, (\mathbf{V}^T \mathbf{X} \, \mathbf{X}^T \mathbf{u})^{-1} \, \mathbf{V}^T \mathbf{y}, \, \widehat{\mathbf{b}}_{PLSR} \in \Re^{mx1} \tag{9}$$

Eq. (5)

$\mathbf{b}_{PLSR}$ denotes the l dimensional vector of regression coefficient in PLSR. In the PLSR model, the Non Linear Iterative Partial Least Squares (NIPALS) algorithm is applied to computational.

2.2. Partial robust M-Regression

The Partial Robust M-Regression (PRM) is an alternative robust version to classical PLSR [33]. It uses a modified robust M-estimate [16] in PLSR by assigning additional reweighting function w_i as diagnostic tool [34]. The PRM succeed to identify the suspected outlier and leverage points both in $\mathbf{y}$ and $\mathbf{X}$ direction. To maximizing the covariance between the response and predictor variable as basic principle in PLSR, the PRM modifies the covariance function to obtain loadings and scores using a robust procedure.

Assigning proper weighted covariance

$$Cov_w(\mathbf{v}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} (w_i v_i y_i)$$

Then the loading vectors in $\mathbf{X}$ is determined in a sequential way through

$$\mathbf{p}_g = \arg\max_{p} \, Cov_w(\mathbf{X} \, \mathbf{p}, \mathbf{y})$$
$$\text{s.t} \, \| \, \mathbf{p} \, \| = 1$$

$Cov_w(\mathbf{p}_j, \mathbf{p}) = 0$ for all preciously computed $\mathbf{p}_j$

Continuing the Eq. (8), for $1 \le i \le n$, the OLS solution of $\mathbf{b}_{PLSR}$ for single $\mathbf{y}$ is defined as

$$\widehat{\mathbf{b}}_{PLSR} = \arg\min_{b} \left(\sum_{i=1}^{n} (y_i - \mathbf{v}_i \, \mathbf{b}_{PLSR})^2 \right) \tag{10}$$

The estimator is optimal if $E \, (\breve{f}) = 0$ and $Var \, (\breve{f}) = 1$ or where $\breve{f} \sim N \, (0, 1)$, otherwise fail to satisfy the normal assumption, the OLS losses its optimality, hence the robust estimator yields better.

The robust M-estimates reconstruct the squares in term of u then giving

$$\widehat{\mathbf{b}}_M = \arg\min_{b} \left(\sum_{i=1}^{n} \theta \, (y_i - \mathbf{v}_i \, \mathbf{b}_M) \right) \tag{11}$$

with $\theta \, (u) = u^2$, $\theta \, (y_i - \mathbf{v}_i \, \mathbf{b}_M) = (y_i - \mathbf{v}_i \, \mathbf{b}_M)^2$ as $\theta \, (u)$ is defined to be loss function which is symmetric and nondecreasing. Recall the $\breve{f}$ as residual $n \, x \, 1$ column vector $\{\breve{f}_i = y_i - \mathbf{v}_i \, \mathbf{b}_{PLSR}\}_{i=1}^n$ in the objective function of Eq. (11). Applying the partial derivative and reweighing technique [34], then the row weight in each observations is $w_i^r = \theta \, (\breve{f}_i)/\breve{f}_i^2$, taking $\theta \, (\breve{f}_i) = w_i^r \breve{f}_i^2$ thus equation in Eq. (11) can be rewritten as

$$\widehat{\mathbf{b}}_M = \arg\min_{b} \left(\sum_{i=1}^{n} w_i^r \breve{f}_i^2 \right) \tag{12}$$

The weight in Eq. (12) is considered less sensitive to the vertical outlier without pin-pointing the leverage points in $\mathbf{X}$ direction, follows Serneels [33] then another weight w_i^x is added to identify the suspected leverage points. The test criteria is $w_i^x \approx 0$ would be suspected as leverage point. The solution of partial robust M-regression hence can be formulized as

$$\widehat{\mathbf{b}}_{PRM} = \arg\min_{b} \left(\sum_{i=1}^{n} w_i^r \, w_i^x \breve{f}_i^2 \right) \tag{13}$$

where $w_i = w_i^r \, w_i^x$ is called to be the generalized weight. The Eq. (13) can be rewritten as

$$\widehat{\mathbf{b}}_{PRM} = \arg\min_{b} \left(\sum_{i=1}^{n} w_i \breve{f}_i^2 \right) \tag{14}$$

here, the weight w_i^r and w_i^x are calculated as

$$w_i^r = \rho \left(\frac{\breve{f}_i}{\widehat{\sigma}}, c \right) \tag{15}$$

with $\widehat{\sigma}$ uses robust $MAD \, (\breve{f}_1, \ldots, \breve{f}_n) = \underset{i}{median} \left| \breve{f}_i - \underset{j}{median} \, \breve{f}_j \right|$, $\rho \, (z, c)$ is a weight function [34] and constant c follows the Huber's function [16].

$$w_i^x = \rho \left(\frac{\| \, \mathbf{v}_i - med_{L1}(\mathbf{V}) \, \|}{median \, \| \, \mathbf{v}_i - med_{L1}(\mathbf{V}) \, \|}, c \right) \tag{16}$$

$\| \cdot \|$ is Euclidean norm, $med_{L1}(\mathbf{V})$ is a robust estimator of the center of the l dimensional score vectors, and $\mathbf{v}_i = (v_{i,1}, \ldots, v_{i,l})^T$ is the vector of component score matrix $\mathbf{V}$ that need to be estimated.

3. Proposed method

3.1. Partial robust MM-Regression

Extending the previous work in PRM [33], the improved robust MM-regression [17] was proposed to determine the final estimates in the PLSR. The estimator combines a high breakdown point (50%) of a class of LTS-estimators and better efficiency of modified M-estimators [33]. Using LTS-estimators [31] on PLSR as

$$\widehat{\mathbf{b}}_{LTS} = arg\ \min_b\ \widehat{\sigma}_{LTS}\ (\breve{f}_1, \breve{f}_2, \ldots, \breve{f}_n) \tag{17}$$

the solution in Eq. (17) is obtained as it finds the minimal possible dispersion of residual. Recall the procedures in the classical MM-estimators: 1) the M-estimators are used as initial estimates; 2) calculate the scale estimates $\widehat{\sigma}_{LTS}$ (LTS-estimators) [31] using the obtained residual $\breve{f}_i$ with objective function labeled as θ_0; 3) proceed the M-estimators with objective function θ_1 to calculate the final estimates. The improved on robust partial MM-regression is now the M-estimators replaced with the modified M-estimators. The aims are beside of having high breakdown point and more efficient estimator, the final estimator is resistant to the outliers and high leverage points. Following the Eq. (14), by applying the re-descending score function $\psi_1(u) = \frac{\partial\,\theta_1(u)}{\partial\,u}$ with $\left\{u = \frac{y_i - \mathbf{v}_i\widehat{\mathbf{b}}_M}{\widehat{\sigma}_{LTS}}\right\}$ and value of the scale estimates $\widehat{\sigma}_{LTS}$ the solution for partial robust modified MM-estimators is the slight modification on weight w_i^r which is

$$w_i^r = \rho\ \left(\frac{\breve{f}_i}{\widehat{\sigma}_{LTS}}, c\right) \tag{18}$$

with the additional resistant weight w_i^x as in Eq. (16). The objective function θ_1 does not need to be similar to the θ_0, but should satisfy the properties as below:

  i. θ is symmetric and continuously differentiable, with $\theta(0) = 0$
  ii. There exists $a > 0$ such that θ is rigidly increasing on $[0\ ,\ a]$ and constant on $[a\ ,\ \infty)$
  iii. $\theta_1(\breve{f}) \leq \theta_0(\breve{f})$ hence,

$$\sum_{i=1}^n \theta_1 \left(\frac{y_i - \mathbf{v}_i\,\widehat{\mathbf{b}}_M}{\widehat{\sigma}_{LTS}}\right) \leq \sum_{i=1}^n \theta_0 \left(\frac{y_i - \mathbf{v}_i\,\widehat{\mathbf{b}}_M}{\widehat{\sigma}_{LTS}}\right)$$

3.2. Partial robust GM6-Regression

The GM-estimator was introduced by Mallows [19] and Schweppe [21] with objective to improve the resistance of classical M-estimator to the suspected high leverage point through downweighing procedure. The general robust GM-estimator using equation in Eq. (1) is given as

$$\widehat{\mathbf{b}}_{GM} = arg\ \min_b\ \left(\sum_{i=1}^n w_i\,\theta\left(\frac{y_i - \mathbf{v}_i\,\widehat{\mathbf{b}}}{z_i\,\widehat{\sigma}}\right)\right) \tag{19}$$

where θ is a specified function which is absolutely continuous and non-decreasing on $[0\ ,\ \infty)$, w_i, z_i represent the weights and depend on the matrix of $\mathbf{V}$ that are updated iteratively from initial estimate using any robust method, and $\widehat{\sigma}$ is an estimate of robust scale. Taking the $\psi(\breve{f}) = \frac{\partial\,\theta(\breve{f})}{\partial\,\breve{f}}$, the GM-estimator in Eq. (19) then yields the solution as

$$\sum_{i=1}^n w_i\,\psi\ \left(\frac{y_i - \mathbf{v}_i\widehat{\mathbf{b}}}{z_i\,\widehat{\sigma}}\right)\ \mathbf{v}\ =\ 0 \tag{20}$$

the Mallows solution [19] provides only the weight w_i (ranged from 0 to 1), that is another weight $z_i = 1$. The weight w_i is a square root function of the diagonal elements of the hat matrix$\{\ w_i \propto (1 - h_{ii})^{1/2}\ \}$; $\mathbf{H} = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$ [21]. It reduces the involvement of observations with high influential points in the factor space. However, the estimator results in loss efficiency because the good leverage points are also downweighed. As improvement, the Schweppe [21] suggested the solution by adjusting the leverage weights function based on the size of residual $\breve{f}_i$ (in PLSR) with now the weight $z_i = w_i$; $\left\{w_i \propto \frac{\sqrt{1-h_{ii}}}{r_i}\,\psi\left(\frac{r_i}{\sqrt{1-h_{ii}}}\right)\right\}$; $\left\{r_i = \frac{\breve{f}_i}{\widehat{\sigma}}\right\}$ [35] and $\widehat{\sigma}$ as scale estimator in [36]. Although the strategy is better than Mallows, the Schweppe solution is problematic with low breakdown point [17] and it is not consistent when the residuals are asymmetric [37]. Moreover, it is suffered with the influence of multiple outliers in the factor space [36]. To overcome these limitation, the GM6 [23] with high breakdown point (closer to 50%) and bounded influence was introduced. The advantage of GM6 is it can classify whether the suspected leverages are good or bad. In the algorithm, the initial estimates b_0 uses high breakdown (BDP = 50%) LTS estimator [31] with the scale estimate is calculated using LMS estimator [31] which provides 0.95 efficiency$\{\ \widehat{\sigma} = 1.4826\ (1 + 5\,/(n - p))\ Median\ |r_i|\ \}$. The partial GM6 estimator can be defined as

$$\widehat{\mathbf{b}}_{GM6} = \widehat{\mathbf{b}}_0 + \left[\sum_{i=1}^n \mathbf{v}_i^T \mathbf{B}\ \mathbf{v}_i\right]^{-1} \times \sum_{i=1}^n \widehat{\sigma}\ \mathbf{W}\ \psi\left(\frac{r_i\,(\widehat{b}_0)}{\widehat{\sigma}\ w_i}\right)\ \mathbf{v}_i \tag{21}$$

it is known that the hat matrix h_{ii} fails to prevent the masking effects of multiple outliers in the factor space [38], the weighing strategy $\mathbf{W} = diag\ (w_i)$ using Robust Mahalanobis Distance (RM$_i^2$) [30] then is preferred

$$w_i \propto min\ \left(1, \frac{\chi_{0.95,p}^2}{\text{RM}_i^2}\right) \tag{22}$$

$$\text{RM}_i^2 = (\mathbf{v}_i - \mathbf{m}_v)^T \mathbf{C}_v^{-1} (\mathbf{v}_i - \mathbf{m}_v) \tag{23}$$

with $\mathbf{m}_v$ and $\mathbf{C}_v$ are robust location and shape estimates of the minimum volume ellipsoid (MVE) estimators. The value of MVE is calculated from the matrix of latent variables $\mathbf{V}$, $\mathbf{B} = diag\ \left(\psi'\left(\frac{r_i(\widehat{b}_0)}{\sigma\ w_i}\right)\right)$ as diagonal in the derivative of Huber's function ψ, and $\chi_{0.95,p}^2$ is the suitable chi-square distribution. In the GM6, the final estimates are calculated in the single step (Newton Raphson) rather than iteratively. Here, as alternative to S-estimator [36] the MM-estimator with better asymptotic efficiency is assigned as initial estimator.

3.3. Partial robust DRGP-Regression

As a cause of masking and/or swamping effect in the factor space, almost the existing robust methods only resistant to the influence of single outlier or high leverage point. With multiple high influential points in the factor space they are ineffective and fail to identify [22, 30, 39]. This may result to a poor fitting process, multicollinearity, and heteroscedastic [40]. Due to this problematic, Habshah [29] introduced the DRGP procedure in order to determine whether the observations potentially have multiple high leverage point or not. The procedure uses suitable robust cut-off point proposed by Imon [41] on the calculated generalized potentials p_{ii}^* to decide the suspected observations. Here, the robust DRGP is applied in the PLSR by assigning the new component variables $\mathbf{V}$ as input space instead of original data $\mathbf{X}$.

Recall the multiple regression model in Eq. (1), the easiest way to detect the influence point of outlier is based on their residual $\mathbf{e}$. Using the OLS method, the term of the true disturbance vector function of leverage and true error weight matrix $\mathbf{W}$ can be defined as

$$\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}} = \left(1 - \mathbf{V}\left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T\right)\mathbf{e} \tag{24}$$

where $\mathbf{W} = \mathbf{V}\left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T$ reflects joint effect of m predictors on the fitted values $\widehat{\mathbf{y}}$ and it is also known as hat matrix. The diagonal elements w_{ii} is the hat values in matrix $\mathbf{W}$ and it is considered as a measure of leverage values of $\mathbf{y}$ corresponds to $\widehat{\mathbf{y}}$.

The DRGP consists of two steps; the first step is to determine the suspected high leverage points using robust approach in Eq. (23), with robust cut-off criteria [39].

$$Median\left(\mathrm{RM}_i^2\right) + 3\,MAD\left(\mathrm{RM}_i^2\right) \tag{25}$$

this involves the calculation of median and Median Absolute Deviation (MAD) in the RM_i^2. The second step uses the diagnostic approach called generalized potential denoted as p_{ii}^* for all members in data set to confirm the suspicion. Let R as a set of "remaining" good observations and D as a set of "deleted" observations, then the R consists of $(n-d)$ observations after $d < (n-k)$ observations in D are deleted. In a case where the ith observations are deleted from the remaining set of R and further joins the deletion set in D. For any such i, the $w_{ii}^{-(D+i)}$ equals to $w_{ii}^{-(D)}/(1-w_{ii}^{-(D)})$ as shown below

$$w_{ii}^{-(D+i)} = \mathbf{v}_i^T\left(\mathbf{V}_R^T\mathbf{V}_R\right)^{-1}\mathbf{x}_i + \frac{\left(\mathbf{v}_i^T\left(\mathbf{V}_R^T\mathbf{V}_R\right)^{-1}\mathbf{v}_i\right)^2}{1 - \mathbf{v}_i^T\left(\mathbf{V}_R^T\mathbf{V}_R\right)^{-1}\mathbf{v}_i} = \frac{w_{ii}^{-(D)}}{\left(1 - w_{ii}^{-(D)}\right)} \tag{26}$$

for a case when the size of R is $(n-1)$ and $D = i$ the $w_{ii}^{-(D)}$ is equivalent as an natural extension of ith potentials p_{ii} in RM_i^2 [30] as $w_{ii}^{(-i)} = \mathbf{v}_i^T\left(\mathbf{V}_{(i)}^T\mathbf{V}_{(i)}\right)^{-1}\mathbf{v}_i = p_{ii}$, thus $w_{ii}^{-(D)}$ is the ith diagonal element in matrix $\mathbf{V}\left(\mathbf{V}_R^T\mathbf{V}_R\right)^{-1}\mathbf{V}^T$.

The generalized potentials criteria as in [41] is given as

$$p_{ii}^* = \begin{cases} \dfrac{w_{ii}^{-(D)}}{1 - w_{ii}^{-(D)}} & for\,i \in R \\[2ex] w_{ii}^{-(D)} & for\,i \in D \end{cases} \tag{27}$$

with the robust cut-off point consider both the dimension of the predictors and any account of the number of observations,

$$p_{ii}^* > Median\left(p_{ii}^*\right) + c\,MAD\left(p_{ii}^*\right) \tag{28}$$

here c is a constant value of 2 or 3 and $MAD\left(p_{ii}^*\right) = Median\left\{\left|p_{ii}^* - Median\left(p_{ii}^*\right)\right|\right\}/0.6475$. The criteria test, if the p_{ii}^* of suspicion is greater than the cut-off point in Eq. (28) the suspicion in the first step is true. While if it is not, then put back the observation into the dataset and recalculate the generalized potentials p_{ii}^* on the remaining subset.

3.4. Kernel - PLS

With the restriction of nonlinear relationship and improper spectra signature in the NIR dataset, the use of kernel function as solution is suggested. By relating the RKHS and feature space F in [7], each point in the original input vectors are mapped nonlinearly to a higher dimensional F. Recall the NIPALS algorithm in PLSR model, the kernel method replaces the dot products calculation of elements in the mapped data corresponding to a nonlinear function in the original input space. Consequently, the orthogonal weight vectors $\mathbf{w}$ and $\mathbf{a}$ as in PLSR cannot be estimated directly, hence the

NIPALS algorithm should be reconstructed into its nonlinear kernel-based variant [42].

Some general definitions associating with the theory in RHKS by Aronszajn (see in [43]) are provided as below:

Definition 1. (*Inner product*). *Let H be a vector space over field $\Re$. A function $\langle\cdot,\cdot\rangle_H{:}H \times H \to \Re$ is said to be an inner product on H must satisfy the conditions*

   i. Bilinearity: $\langle\alpha f + \beta g, k\rangle_H = \alpha\langle f, k\rangle_H + \beta\langle g, k\rangle_H \forall f,\,g,\,k \in H,\;\forall\,\alpha,\,\beta \in \Re$
   ii. Symmetric: $\langle f, g\rangle_H = \langle g, f\rangle_H \forall f, g \in H$
   iii. Positive definiteness: $\langle f, f\rangle_H \geq 0, \forall f \in H$ and $\langle f, f\rangle_H = 0$ if and only if $f = 0$.

Definition 2. (*Normed space*) *is a linear (vector) space H over field $\Re$ on which a norm on H is defined and for $f, g \in H$ and $\forall\,\alpha \in \Re$ satisfies the properties*

   i. Positive homogeneity: $\|\alpha f\| = |\alpha|\,\|f\|$
   ii. Triangle inequality: $\|f + g\| \leq \|f\| + \|g\|$
   iii. Positive definiteness: $\|f\| \geq 0$ and $\|f\| = 0$ if and only if $f = 0$

Definition 3. (*Convergent*). *Let H be an inner product space, a sequence $\{f_n\}_{n=1}^{\infty}$ is said to be Convergent if for $\forall\,\varepsilon > 0$ there exists an element $f \in H$, such that $\|f - f_n\| < \varepsilon$ as $n \to \infty$.*

Definition 4. (*Cauchy sequence*). *Let H be an inner product space, a sequence $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence in H if $\forall\,\varepsilon > 0$ there exists a positive integer N, such that $\|f_n - f_m\| < \varepsilon$ whenever $n, m \geq N$.*

Any convergent sequence is also a Cauchy sequence.

Definition 5. (*Complete*). *A linear (vector) space H is said to be complete if the Cauchy sequence is convergent.*

Definition 6. (*Hilbert Space*). *A normed space is said to be a Hilbert space on which the norm is induced by an complete inner product space $\langle f, g\rangle$ as the relation $\|f\|_H{:} = \sqrt{\langle f, f\rangle_H}$.*

Definition 7. (*Kernel*). *Consider $\mathbf{X}$ be an arbitrary non-empty set. A function $K{:}\mathbf{X} \times \mathbf{X} \to \Re$ is said to be a kernel if $\Re$-Hilbert space and a map $\phi{:}\mathbf{X} \to H$ are exist, such that $\forall\,\mathbf{x}_i, \mathbf{x}_i^T \in \mathbf{X}$,*

$$K\left(\mathbf{x}_i, \mathbf{x}_i^T\right){:} = \left\langle\phi(\mathbf{x}_i), \phi(\mathbf{x}_i)^T\right\rangle_H \tag{29}$$

Definition 8. (Reproducing Kernel Hilbert Space). Let H be a Hilbert space of function on non-empty set $\mathbf{X}$ whose $\langle f, g\rangle$ as complete inner product space and $\|f\|_H{:} = \sqrt{\langle f, f\rangle_H}$ as the normed space in H, for f, $g \in H$. Reproducing kernel of H is said to be a function of $K{:}\mathbf{X} \times \mathbf{X} \to \Re$ if satisfies the properties

   i. $K_{\mathbf{x}}(\cdot) = K(\mathbf{x},\;\cdot) \in H$ for any $\mathbf{x} \in \mathbf{X}$, where
   ii. Reproducing property: $f(x) = \langle f(\cdot), K(x,\;\cdot)\rangle_H$, for any $f \in H$ and $\mathbf{x} \in \mathbf{X}$.

Applying the function $K_{\mathbf{x}}$ at $\mathbf{y}$ to the properties (ii) in (Def.8) giving

$$f(\mathbf{x}) = \langle f(\mathbf{y}), K(\mathbf{x}, \mathbf{y})\rangle_H, \text{ for } \forall f \in H,\text{ and}$$

$$K_{\mathbf{x}}(\mathbf{y}) = \langle K_{\mathbf{x}}, K_{\mathbf{y}}\rangle, \text{ for } \mathbf{x}, \mathbf{y} \in \mathbf{X}$$

with the symmetric function of two variables in K satisfying the Mercer's theorem [44].

Theorem 1. *Let H be a Hilbert space of function over non-empty set $\mathbf{X}$. The following properties are equivalent*

1. H has a reproducing kernel
2. For any $\mathbf{x} \in \mathbf{X}$, the function $\phi_{\mathbf{x}}:H \to \Re$ defined by $\phi_{\mathbf{x}}(f) = f(\mathbf{x})$ is continuous.

Following Mercer's theorem, each positive definite kernel $K(\mathbf{x},\mathbf{y})$ on a compact domain $\mathbf{X} \times \mathbf{X}$ may be written as

$$K(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{s} \lambda_i \phi_i(\mathbf{x}) \, \phi_i(\mathbf{y}), \; S \leq \infty \tag{30}$$

where λ_i is the sequence of eigenvalues with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_S$ of $K(\mathbf{x},\mathbf{y})$ and $\{\phi_i\}_{i=1}^{s}$ is the infinite sequence of eigenfunctions. Moreover, Eq. (30) can be reformulated as

$$K(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{s} \sqrt{\lambda_i}\phi_i(\mathbf{x}) \, \sqrt{\lambda_i} \, \phi_i(\mathbf{y}) = (\boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{y})) = (\, \boldsymbol{\phi}(\mathbf{x})^T \cdot \boldsymbol{\phi}(\mathbf{y})\,) \tag{31}$$

It is obvious to see that any kernel $K(\mathbf{x},\mathbf{y})$ in Eq. (31) also corresponds to a canonical (Euclidean) dot product in a possibly higher dimensional feature space F, with the mapping function

$$\begin{aligned}\boldsymbol{\phi}:\mathbf{X} &\to F \\ \mathbf{x} &\to \left(\sqrt{\lambda_1}\,\phi_1(\mathbf{x}), \sqrt{\lambda_2}\,\phi_2(\mathbf{x}), \ldots, \sqrt{\lambda_s}\,\phi_s(\mathbf{x})\right)\end{aligned} \tag{32}$$

with $\{\,\{\sqrt{\lambda_i}\,\phi_i(\mathbf{x})\}_{i=1}^{s}, \mathbf{x} \in \mathbf{X}\}$ denotes as feature mapping.

Based on these properties, the description of the nonlinear kernel is discussed. Assuming a nonlinear transformation function of the input variables $\mathbf{x}_i \in \mathbf{X} \subset \Re^m$ into the feature space F.

$$\boldsymbol{\phi}:\mathbf{x}_i \in \mathbf{X} \subset \Re^m \to \phi(\mathbf{x}_i) \in F \tag{33}$$

with $\boldsymbol{\phi}(\cdot)$ is a nonlinear mapping function in the input space of the original input vectors to the feature space F. The mapping $\boldsymbol{\phi}$ replaces $\mathbf{x}_i$ with the $\phi(\mathbf{x}_i)$ and result to the high-dimension which can even be infinite feature space F. Once the component score matrix $\mathbf{V}$ in linear PLSR is obtained, then a nonlinear PLSR is computed as the original input data space. Define $\boldsymbol{\phi}$ as an $n \times s$ matrix of mapped space data with the ith row is the vector $\phi(\mathbf{x}_i)$ in s dimensional feature space F. Here the use of nonlinear kernel function is preferred instead of an explicit nonlinear mapping. The Eq. (31) giving the deflation

$$\mathbf{K} = \boldsymbol{\phi}\,\boldsymbol{\phi}^T \tag{34}$$

with $\mathbf{K}$ represents as $n \times n$ kernel Gram matrix of the cross dot products among all mapped input data points $\{\phi(\mathbf{x}_i)\}_{i=1}^{n}$. Using the normalized component scores $\mathbf{v}_g$, the deflation of $\mathbf{K}$ and $\mathbf{y}$ are formulized as

$$\begin{aligned}\boldsymbol{\phi}_{g+1}\,\boldsymbol{\phi}_{g+1}^T &\leftarrow \left(\boldsymbol{\phi}_g - \mathbf{v}_g \mathbf{v}_g^T \boldsymbol{\phi}_g\right)\left(\boldsymbol{\phi}_g - \mathbf{v}_g \mathbf{v}_g^T \boldsymbol{\phi}_g\right)^T \\ \mathbf{K}_{g+1} &\leftarrow \left(\mathbf{I} - \mathbf{v}_g \mathbf{v}_g^T\right)\mathbf{K}_g \left(\mathbf{I} - \mathbf{v}_g \mathbf{v}_g^T\right)\end{aligned} \tag{35}$$

and

$$\mathbf{y}_{g+1} \leftarrow \mathbf{y}_g - \mathbf{v}_g \mathbf{v}_g^T \mathbf{y}_g = \mathbf{y}_g\left(\mathbf{I} - \mathbf{v}_g \mathbf{v}_g^T\right) \tag{36}$$

the coefficient matrix of the KPLS regression model in F then can be written as

$$\widehat{\mathbf{b}}_{KPLS} = \boldsymbol{\phi}^T \mathbf{u}\left(\mathbf{V}^T \mathbf{K}^T \mathbf{U}\right)^{-1}\mathbf{V}^T\mathbf{y}, \quad \mathbf{b}_{KPLS} \in \Re^{mxp} \tag{37}$$

and the final prediction of component concentration in the regression model is given as

$$\begin{aligned}\widehat{\mathbf{y}} &= \boldsymbol{\phi}\,\boldsymbol{\phi}^T\mathbf{u}\left(\mathbf{V}^T\boldsymbol{\phi}\boldsymbol{\phi}^T\,\mathbf{u}\right)^{-1}\mathbf{V}^T\mathbf{y} \\ \widehat{\mathbf{y}}_v &= \mathbf{K}_v\mathbf{u}\left(\mathbf{V}^T\mathbf{K}\,\mathbf{u}\right)^{-1}\mathbf{V}^T\mathbf{y}\end{aligned} \tag{38}$$

where $\widehat{\mathbf{y}}$ and $\widehat{\mathbf{y}}_v$ are the prediction of calibration set and validation set, respectively. $\mathbf{K}_v$ is the $n_v x\, n$ kernel matrix of validation set with elements are composed of $\mathbf{K}_{ij} = \mathbf{K}(\mathbf{x}_i,\mathbf{x}_j)$, which are input vectors of calibration set $\{\,\mathbf{x}_i\}_{i=n+1}^{n+n_v}$ and validation set $\{\,\mathbf{x}_j\}_{j=1}^{n}$. As suggested by Wu [45, 46], centering the kernel both in calibration and validation are very important to provide the bias term to be zero. Centralization on the mapped data in F simply can be calculated as

$$\begin{aligned}\mathbf{K} &= \left(\mathbf{I} - \frac{1}{n}1_n 1_n^T\right)\mathbf{K}\left(\mathbf{I} - \frac{1}{n}1_n 1_n^T\right) \\ \mathbf{K}_v &= \left(\mathbf{K}_v - \frac{1}{n}1_{n_v}1_n^T\mathbf{K}_v\right)\left(\mathbf{I} - \frac{1}{n}1_n 1_n^T\right)\end{aligned} \tag{39}$$

with $\mathbf{I}$ represents as n-dimensional identity matrix, and $1_n, 1_{n_v}$ denotes the vectors whose elements equal to 1, with lengths n and n_v, respectively.

3.5. Kernel - PRM

The Kernel-PRM (simply called as KPRM) develops the concept of PRM to incorporate the nonlinear feature in the original input factor dataset. In the prior study [47], the KPRM uses kernel function to transform the nonlinear relationship in the original input space into a linear PLS through high-dimensional feature space. The method provides generalized weight w_i scheme as robust procedure to the nonlinear kernel to remove the outlier and downgrade the influence of bad high leverage point in the dataset. A modified PRM as proposed in the KPRM is the input matrix $\mathbf{X}$ is subsequently replaced as the outer product $\boldsymbol{\phi}\,\boldsymbol{\phi}^T$ of the $n\, x\, n$ kernel Gram matrix $\mathbf{K}$. A main concern in the KPRM procedure is based on the weighing scheme in $\mathbf{K}$. Let define $\breve{\mathbf{K}}$ as the weighted matrix of $\mathbf{K}$, so that

$$\breve{\mathbf{K}} = (\boldsymbol{\Omega}\phi)\,(\boldsymbol{\Omega}\phi)^T = \boldsymbol{\Omega}\mathbf{K}\boldsymbol{\Omega} \tag{40}$$

where $\boldsymbol{\Omega}$ is said to be the diagonal weight matrix, with the ith elements in the diagonal matrix equals to the generalized weight in the PRM.

3.6. Kernel Partial Diagnostic Robust Potential

Another alternative of robust method in the class of nonlinear kernel called Kernel Partial Diagnostic Robust Potential (KPDRGP) is proposed. The KPDRGP follows the principles of the Partial Robust DRGP-Regression that uses the suitable robust cut-off point on the calculated generalized potentials p_{ii}^* to confirm the suspected high leverage. Instead of applying the explicit nonlinear mapping of $\mathbf{X}$ to the feature space F, the kernel function $\boldsymbol{\phi}\,\boldsymbol{\phi}^T$ as the cross dot products among all mapped input data points $\{\phi(\mathbf{x}_i)\}_{i=1}^{n}$ is used. Here, a Gaussian kernel function is applied due to its superiority to handle nonlinear dataset [48].

$$\mathbf{K}_{i,j} = \mathbf{K}\left(\,\mathbf{x}_i,\mathbf{x}_j\right) = exp\left(-\frac{\|\,\mathbf{x}_i - \mathbf{x}_j\,\|^2}{\sigma^2}\right) \tag{41}$$

where σ^2 is the non-robust variance of original training data $\mathbf{X}$ and $\|\cdot\|$ is Euclidean norm of two matrices.

Let define $\breve{\mathrm{R}}\,\mathrm{M}_i^2$ as robust Mahalanobis distance that employ the ith elements kernel Gram matrix $\mathbf{K}$ in the input space, such that

$$\breve{\mathrm{R}}\,\mathrm{M}_i^2 = (\breve{\mathbf{v}}_i - \breve{\mathbf{m}}_v)^T\mathbf{C}_v^{-1}(\breve{\mathbf{v}}_i - \breve{\mathbf{m}}_v) \tag{42}$$

where $\breve{\mathbf{m}}_v$ and $\breve{\mathbf{C}}_v$ are robust location and shape estimates of the minimum volume ellipsoid (MVE) estimators. These robust estimator are calculated from the matrix of new components $\breve{\mathbf{V}}$ with kernel Gram matrix $\mathbf{K}$ in the input space. To justify the suspicion high leverage points it is important

to determine how $w_{ii}^{-(D)}$ is adjusted. Define new $w_{ii}^{-(D)}$ as the ith diagonal element in matrix $\breve{\mathbf{v}}$ $(\breve{\mathbf{v}}_R^T \breve{\mathbf{v}}_R)^{-1} \breve{\mathbf{v}}^T$ and result to the new generalized potential $\breve{p}_{ii}^*$. This value then is used as diagnostic tool to confirm the suspicion through cut-off point in Eq. (28).

$$\breve{p}_{ii}^* = \begin{cases} \dfrac{\breve{w}_{ii}^{-(D)}}{1 - \breve{w}_{ii}^{-(D)}} \, for\, i \in R \\[2em] \breve{w}_{ii}^{-(D)} \quad for\, i \in D \end{cases} \tag{43}$$

3.7. Algorithm Kernel Partial Diagnostic Robust Potential

Let define $\mathbf{K} = \boldsymbol{\phi}(\mathbf{X}) \, \boldsymbol{\phi}(\mathbf{X})^T$ be the kernel Gram matrix of the cross dot products among all mapped input data points $\{ \phi(\mathbf{x}_i) \}_{i=1}^n$ in feature space F. In addition, $\mathbf{K}$ must be centered, and $\mathbf{y}$ need to be normalized to satisfy the normal assumption.

1. Apply a nonlinear mapping function $\{ \boldsymbol{\phi} : \mathbf{x}_i \in \mathbf{X} \subset \Re^m \to \phi(\mathbf{x}_i) \in F \}$ in the input space of the original input vectors to the feature space F. The mapping $\boldsymbol{\phi}$ replaces $\mathbf{x}_i$ with the $\phi(\mathbf{x}_i)$ and result to the high-dimensional feature space F.

2. Replace the outer product $\phi(\mathbf{x}_i) \, \phi(\mathbf{x}_i)^T$ by the kernel matrix $\mathbf{K}$. Apply an identity matrix $\mathbf{I}$ as the initial weight $\boldsymbol{\Omega}$ on the kernel matrix to get the weighted $\breve{\mathbf{K}}_g$ and output vector $\breve{\mathbf{y}}_g$, $\breve{\mathbf{y}}_g = \boldsymbol{\Omega} \, \mathbf{y}_g$.

3. Initialize the weighted output vector $\breve{\mathbf{y}}$ as a column weighted latent variable $\breve{\mathbf{u}}_1 = \breve{\mathbf{y}}_g$.

4. Regress $\breve{\mathbf{K}}_g$ on $\breve{\mathbf{u}}_1$ to obtain the weight $\breve{\mathbf{w}}_g = (\breve{\mathbf{K}}_g^T \breve{\mathbf{u}}_1) / (\breve{\mathbf{u}}_1^T \breve{\mathbf{u}}_1)$, then normalize $\breve{\mathbf{w}}_g$ and rename it as $\breve{\mathbf{v}}_g = \breve{\mathbf{w}}_g / \|\breve{\mathbf{w}}_g\|$. $\breve{\mathbf{v}}_g$ is the weighted of new component variables with kernel Gram matrix $\breve{\mathbf{K}}_g$ in the input space.

5. Calculate the weights in $\breve{\mathbf{y}}_g$ by $\breve{\mathbf{c}}_g = (\breve{\mathbf{y}}_g^T \breve{\mathbf{v}}_g) / \breve{\mathbf{v}}_g^T \breve{\mathbf{v}}_g$ and project the $\breve{\mathbf{y}}_g$ on the $\breve{\mathbf{c}}_g$ to calculate the new component variable in $\breve{\mathbf{y}}_g$ using $\breve{\mathbf{u}}_g^* = \breve{\mathbf{y}}_g \, \breve{\mathbf{c}}_g$ and normalize $\breve{\mathbf{u}}_g^*$.

6. Determine the $\breve{\mathbf{y}}_g$ and $\breve{\mathbf{v}}_g$ improvement using $\breve{\mathbf{u}}_\Delta = \breve{\mathbf{u}}_g^* - \breve{\mathbf{u}}_g$, if $\breve{\mathbf{u}}_\Delta > $ tol, go to step 1 to 5 until convergence using $\breve{\mathbf{u}}_g^*$. While if $\breve{\mathbf{u}}_\Delta < $ tol, the first latent PLS is found and proceed to the step 7. The tol refer to tolerance limit for convergence which is set to 10^{-4}.

7. Deflate the $\breve{\mathbf{K}}_g$ and $\breve{\mathbf{y}}_g$ using $\breve{\mathbf{K}}_{g+1} = (1 - \breve{\mathbf{v}}_g \breve{\mathbf{v}}_g^T) \, \breve{\mathbf{K}}_g (1 - \breve{\mathbf{v}}_g \breve{\mathbf{v}}_g^T)$, $\breve{\mathbf{y}}_{g+1} = \breve{\mathbf{y}}_g (1 - \breve{\mathbf{v}}_g \breve{\mathbf{v}}_g^T)$

8. Continue to step 4 until all l PLS components are determined.

9. Here, the new calculated $\breve{\mathbf{V}}$ as nonlinear PLSR are used as the new input space in the robustness procedures.

10. For each i point on $(\breve{\mathbf{v}}_i, \breve{\mathbf{m}}_i)$ pair, calculate the $\breve{\mathbf{R}} \, \mathbf{M}_i^2$

11. Use the cut-off point criteria $Median \, (\breve{\mathbf{R}} \, \mathbf{M}_i^2) + 3MAD \, (\breve{\mathbf{R}} \, \mathbf{M}_i^2)$ to determine the suspected outlier and high leverage points in each i point of $\breve{\mathbf{R}} \, \mathbf{M}_i^2$.

12. Calculate the generalized potential values for each observation. If the $\breve{p}_{ii}^*$ of suspicion is less than the cut-off point then put back the observation into the dataset and then recalculate the generalized potentials $\breve{p}_{ii}^*$ on the remaining subset R.

13. The observations that are identified as real outlier and high leverage points are removed from the remaining subset R.

14. Update the weight Ω by assigning the weight score as 0 to the deleted observation, while the rest observations in the subset R are weighted as 1.

15. Continue to step 2 to recalculate the new weighted $\breve{\mathbf{K}}_g$ and output vector $\breve{\mathbf{y}}_g$, re-perform the rest steps until all PLS estimates are obtained.

4. Results and discussion

To examine the superiority, all the proposed methods: PLS, PRM, PRMM, PRGM6, PRDRGP, KPLS, KPRM, and KPDRGP were evaluated in order using the simulation data and real data. The simulation data uses the $sin(x)$ function with added noise $e \sim N(0, 0.05)$, while the real data uses NIR spectral dataset of oil palm (*Elaeis guineensis*) fresh fruit mesocarp (see Figure 1). This fruit samples are a part of breeding trial experiment in Palapa Estate, PT. Ivomas Tunggal, Riau Province, Indonesia. There are two response variables observed from the fruit mesocarp sample: percent of oil to dry mesocarp (%ODM) and percent of oil to wet mesocarp (%OWM). The spectral data were collected (in contact) using the QualitySpec Trek from Analytical Spectral Devices (ASD Inc., Boulder, Co., USA) on 12 different sampling positions in the fruit bunch (without destructive). These data are good enough to evaluate the non linear effect of the proposed methods. The comprehensive algorithms about the proposed methods were written into R-language using Ri386 software (http://r-project.org) with 3.4.2 version for windows.

4.1. Simulation data

The simulation data uses the $sin(x)$ function to generate the nonlinear behavior dataset. Earlier it was known that some existing multivariate calibration only capable to capture the linear datasets, otherwise it fails to capture the dataset structure with highly nonlinear. In this section, 41 samples data were randomly generated in the range of $[0, 10]$ as training set and were calculated with correspond to the sine function in Eq. (44). To evaluate the sensitivity of the proposed method, 7 random outliers were manually added in the training set. The outliers were marked orderly followed their sample number which are 10, 12, 18, 19, 28, 33, and 36.

$$f(x) = sin(-3/4x) \tag{44}$$

To test the consistency in the training model, 101 samples was also uniformly generated using the range of $[0, 10]$ as testing set. The performance of the methods both using non-kernel and kernel methods are presented in Figure 2 and Figure 3.

As seen in Figure 3, by applying the kernel transformation on the methods the prediction results is really promising. The methods succeed to capture the structure of highly nonlinear dataset. While without kernel transformation (see Figure 2), all the methods fails to fit the model. The proposed KPDRGP method shows a better fitting line compared with the KPLS and KPRM. With the contamination of outliers in the training dataset, the KPDRGP is less sensitive and succeeds to decrease the influence. It is interested to evaluate the robustness of the proposed DRGP procedure with and without kernel transformation in PLSR through their final weight values (see Figure 4).

With the influence of nonlinear effect in the dataset, the linear PRDRGP algorithm fails to identify the true outliers in the dataset (see Figure 4). While by transforming the data into a high dimensional feature space using kernel Hilbert Space method, it succeeds to capture the true outliers in the dataset. The KPDRGP method decides the sample number: 10, 12, 18, 32, 33, 34, and 36 were the true outliers. However it is still failed to screen the remaining 2 outliers (sample number: 19 and 34), this might be caused by the extraction process in the data transformation. In Figure 3, the KPDRGP also succeed to prevent the over-lower prediction in the testing data. The summarization of the prediction results both using training and testing dataset can be seen in Table 1. Here, the statistical measures uses Desirability Indices [49] such as Root Mean
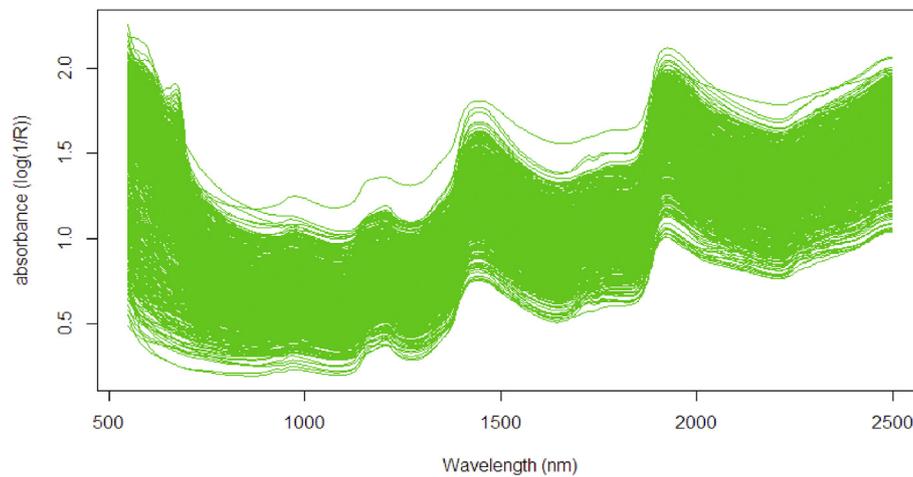
Figure 1. First run of averaged spectra on oil palm fresh fruit mesocarp.
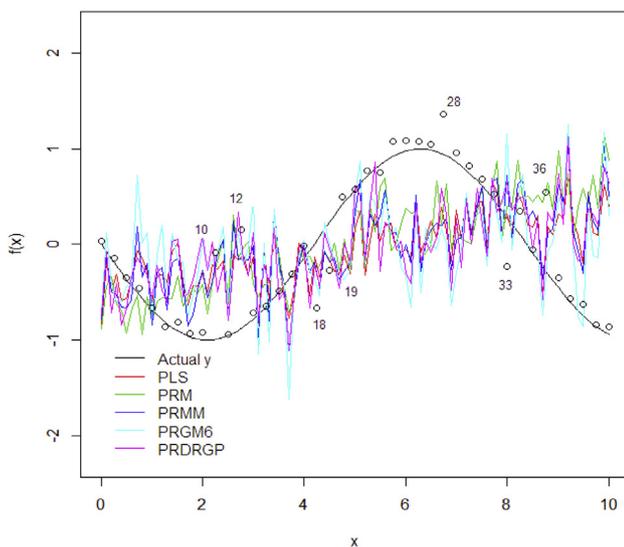
Figure 2. Predictions on testing data using training model of non-kernel methods.
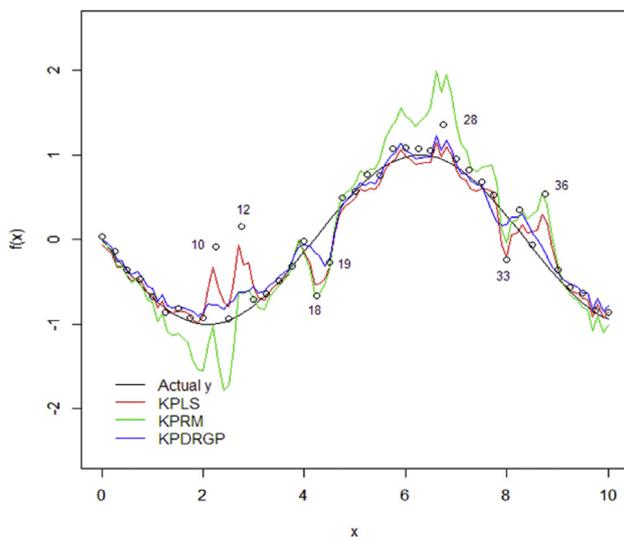
Figure 3. Predictions on testing data using training model of kernel methods.

Squared Error (RMSE), Coefficient of Determination (R^2), Ratio of Prediction to Deviation (RPD), Bias, and Standard Error (SE).

In Table 1, by preventing the effects of outlier in the fitting process, the KPDRGP produces the lowest prediction error (SE and RMSE) both in training and testing data set are about 0.150 and 0.144, respectively. The method shows the highest R^2 both in training and testing dataset which are 0.980 and 0.965, respectively. This R^2 value represents the variability for a response variable that can be explained by predictor variables. The method also shows better reliability value (RPD) with less bias than the other methods. In the class of kernel, the KPLS method is inferior due to contamination from outliers in the dataset. By downgrading the effect of multiple outliers in the dataset, the KPDRGP is superior to KPRM.

4.2. Oil to dry mesocarp data

The real NIR spectra dataset on fresh fruit mesocarp was used to examine the robustness of the kernel solution. The dataset contains 960 observations and 489 wavelengths data points as predictor variables. In this section, the %ODM was used as response variable. As seen in Figure 1, the higher spectral absorbance corresponds to a higher %ODM in fruit mesocarp. Figure 5 shows the frequency distribution of the % ODM values used as training set. To build a good training model, the training set should contain all possible values. Earlier, it was considered if the dataset has high dimensional problem as result of huge number of observation and predictor variables. This condition may risk to the multicollinearity that lead to the heteroscedastic problem in the residual. To encounter this, the PLSR method is preferred to summarize the covariance between predictor variable and the response variable by downscaling the original variables **X** into less number of new PLS components **V**.

In this study, the maximum of PLS components used in the model were limited to 20 and 30 components. This strategy was assigned to evaluate the efficiency of PLSR model using less and large number of PLS components. The assumption is the less number of PLS components used in the training model the most efficient is. As seen in Table 2 it is observed that in the class of non-kernel method, all the proposed methods (PLS, PRM, PRMM, PRGM6, PRDRGP) are not better than the class of kernel partial method (KPLS, KPRM, KPDRGP). The KPDRGP method produces the lowest prediction error (SE and RMSE) even using less and higher number of PLS which are about 0.466 and 0.204, respectively. The reliability of the training model based on its RPD is also greater than 3; means that the training model constructed arestrong enough to prove the reliability for routine analysis. With contamination of outliers in the dataset, the KPLS fails to prevent the influence hence result to a higher SE with lower R^2 compared to the KPRM and KPDRGP.
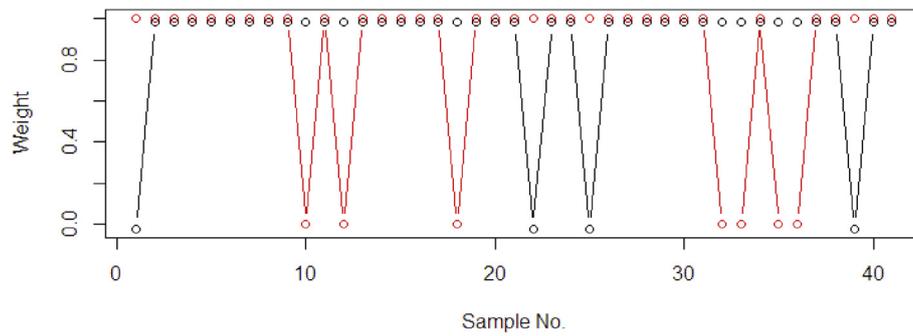
Figure 4. The final sample weight values using DRGP algorithm: PRDRGP (black) and KPDRGP (red).

Table 1. Statistical measures on prediction results using sine function simulation data.

Dataset	Desirability Index	PLS	PRM	PRMM	PRGM6	PRDRGP	KPLS	KPRM	KPDRGP
Training	RMSE	0.631	0.690	0.645	0.662	0.654	0.302	0.265	0.101
	R^2	0.201	0.109	0.171	0.141	0.156	0.920	0.958	0.980
	RPD	1.117	1.021	1.092	1.064	1.077	1.087	2.656	5.724
	Bias	-0.106	0.003	-0.062	-0.081	-0.084	0.002	0.026	0.001
	SE	0.639	0.698	0.653	0.671	0.662	0.320	0.269	0.150
Testing	RMSE	0.647	0.704	0.686	0.768	0.713	0.235	0.679	0.143
	R^2	0.133	0.107	0.104	0.029	0.104	0.854	0.891	0.965
	RPD	1.065	0.979	1.004	0.897	0.966	2.928	1.015	4.819
	Bias	0.032	0.089	0.049	0.028	0.049	0.024	0.114	0.054
	SE	0.650	0.707	0.690	0.772	0.717	0.236	0.682	0.144
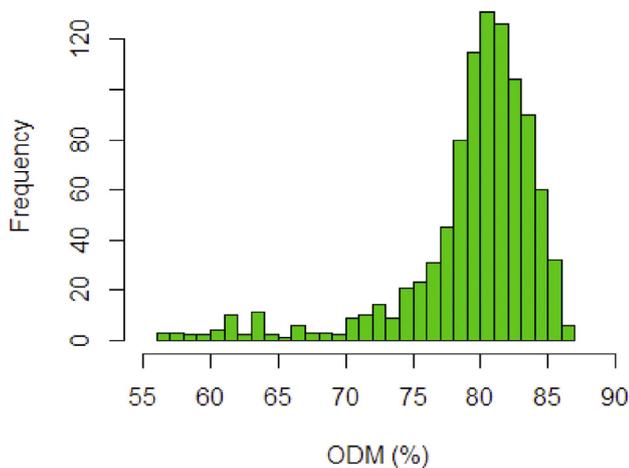
Figure 5. Frequency distribution histogram of %ODM values.

The fitting performance between measured and predicted values using PRDRGP and KPDRGP methods on %ODM dataset are shown in Figure 6. It highlights how well the kernel solution on the proposed KPDRGP (Figure 6(c) and 6(d)) can improve the prediction result using training model (see Table 2). This can be inspected through how close the predicted values are to the fitted regression line using without kernel (Figure 6(a)) and with kernel (6(c)) solution applied in the PRDRGP model. In the study, the PRDRGP fits the data adequately (with $R^2 > 95\%$) compared to the others class of robust PLS such as PRM, PRMM, and PRGM6. When the kernel solution is applied, the accuracy in the KPDRGP increases. The KPDRGP (Figure (6(d)) yields the lowest prediction residual which is closed to 0 than PRDRGP (Figure 6(b)).

4.3. Oil to wet mesocarp data

Using similar NIR spectral dataset of oil palm fresh fruit mesocarp as predictor variable, in this section the %OWM (see Figure 7) was assigned as response variable. As seen in Figure 7, the distribution of the response variable covers all possible value of oil content in the wet mesocarp. The maximum number of PLS

Table 2. Statistical measures on prediction results using %ODM dataset.

PLS	Desirability Index	PLS	PRM	PRMM	PRGM6	PRDRGP	KPLS	KPRM	KPDRGP
$l = 20$	RMSE	3.305	3.344	3.474	3.601	3.343	0.617	0.561	0.466
	R^2	0.556	0.578	0.562	0.548	0.570	0.925	0.988	0.993
	RPD	1.551	1.533	1.475	1.423	1.519	8.299	9.073	11.588
	Bias	0.240	0.298	0.426	0.498	0.313	0.011	-0.021	-0.004
	SE	3.307	3.346	3.476	3.603	3.325	0.618	0.561	0.466
$l = 30$	RMSE	2.934	2.992	3.166	2.652	2.969	0.273	0.242	0.203
	R^2	0.632	0.660	0.631	0.633	0.665	0.937	0.997	0.998
	RPD	1.747	1.713	1.619	1.932	1.726	18.804	19.911	25.579
	Bias	0.060	0.133	0.385	0.196	0.142	0.014	0.018	0.001
	SE	2.996	2.993	3.168	2.954	2.971	0.273	0.242	0.204

Fitted Values

Residual

(a)

(b)

(c)

(d)
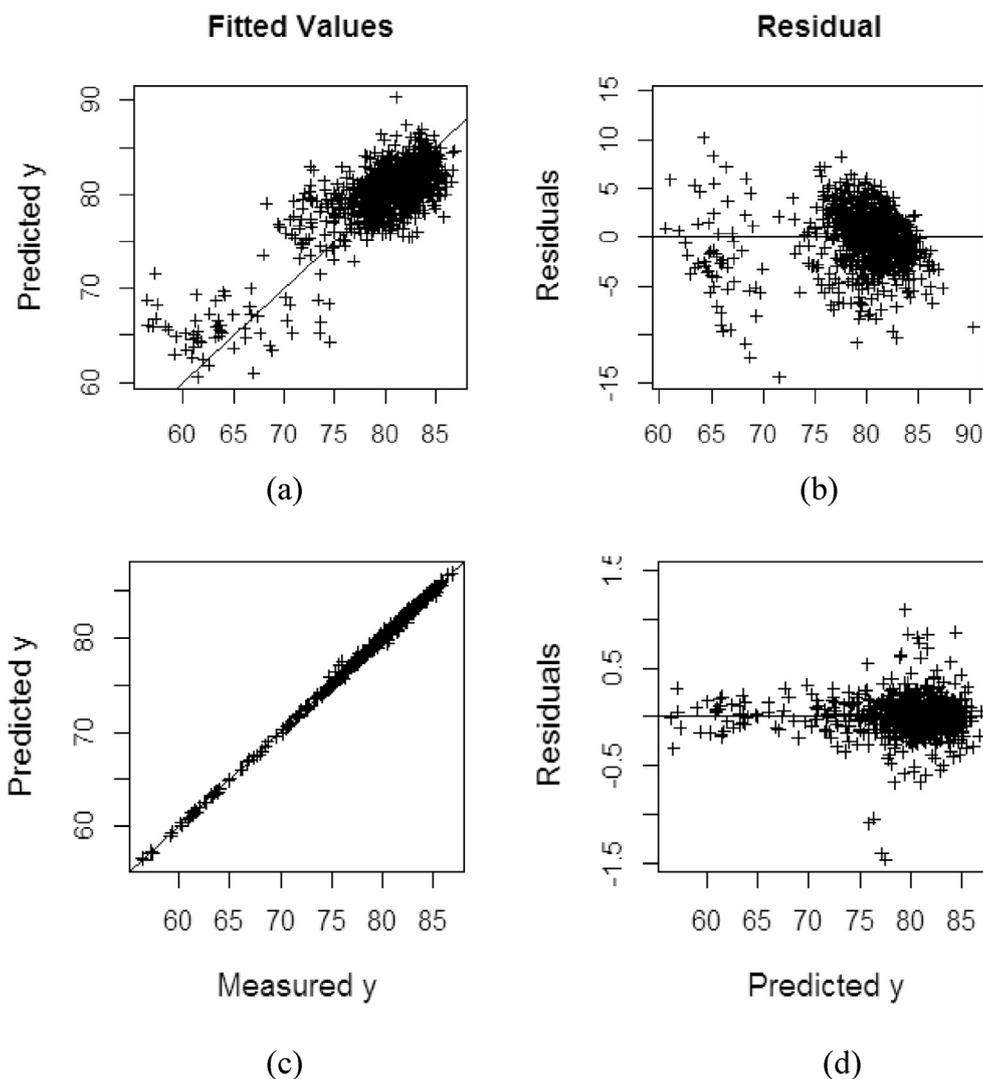
Figure 6. Fitting and residual performance on %ODM dataset: (a) measured vs predicted values on PRDRGP; (b) predicted vs residual on PRDRGP; (c) measured vs predicted values on KPDRGP; (d) predicted vs residual on KPDRGP.
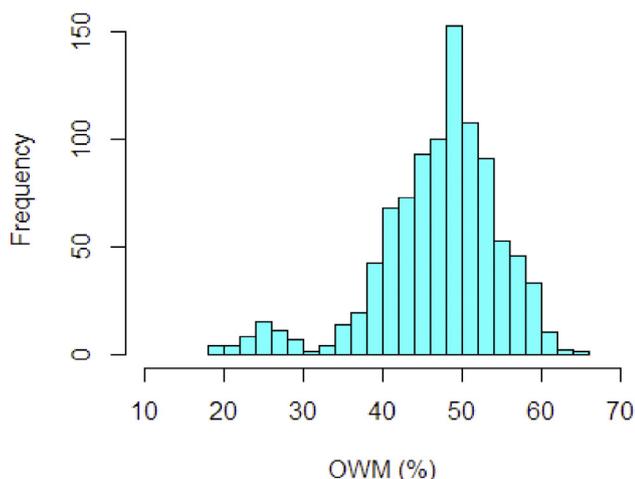
Figure 7. Frequency distribution histogram of %OWM values.

components was set to 20 and 30 variables. In Table 3, the result shows that the class of non-kernel partial methods which are PRDRGP, PRM, and PRGM6 generally superior to PLS and PRMM. The methods yield a higher R^2 with lower prediction error (SE and RMSE). According to their reliability index using RPD value, all the non-kernel methods fail to meet (RPD <3). The proposed KPDRGP method still provides the highest fitted regression line (R^2>95%) with lowest prediction error even using less and higher predictors. The proposed robust KPDRGP is also more reliable (with RPD>3) than the other methods. With the possible contamination of outliers in the dataset, the non-robust KPLS fails to prevent the effect hence results to low accuracy compared to the KPRM and KPDRGP.

Without kernel transformation applied in the input variable, all the class of non-kernel methods fail to fit the training model. This also lead to produce high bias in the predicted residual (see Table 2) compared to their kernel version. The proposed robust PRDRGP method also does not fit the data adequately (see Figure 8(a) and 8(b)) and still produce high bias (see Table 2). While, by integrating kernel transformation in the model, the method (KPDGP) improved the accuracy and reduce the bias (see Figure 8(c) and 8(d)). This is to confirm that using different dataset, in the class of robust with and without kernel method, the proposed robust PLSR using DRGP procedure is still superior.

Table 3. Statistical measures on prediction results using %OWM dataset.

PLS	Desirability Index	PLS	PRM	PRMM	PRGM6	PRDRGP	KPLS	KPRM	KPDRGP
$l = 20$	RMSE	4.471	4.516	4.535	4.487	4.517	0.923	0.768	0.686
	R^2	0.644	0.658	0.655	0.662	0.658	0.956	0.990	0.993
	RPD	1.726	1.709	1.702	1.720	1.709	8.363	9.972	11.646
	Bias	-0.029	0.151	0.004	0.002	0.007	-0.026	0.031	0.003
	SE	4.473	4.519	4.537	4.489	4.519	0.923	0.768	0.686
$l = 30$	RMSE	4.085	4.126	4.249	4.180	4.130	0.404	0.374	0.279
	R^2	0.710	0.714	0.697	0.712	0.714	0.967	0.998	0.999
	RPD	1.890	1.870	1.816	2.097	1.869	19.098	20.946	28.252
	Bias	0.075	0.048	0.011	-0.030	0.097	0.003	-0.010	0.002
	SE	4.087	4.129	4.251	4.082	4.133	0.404	0.374	0.279
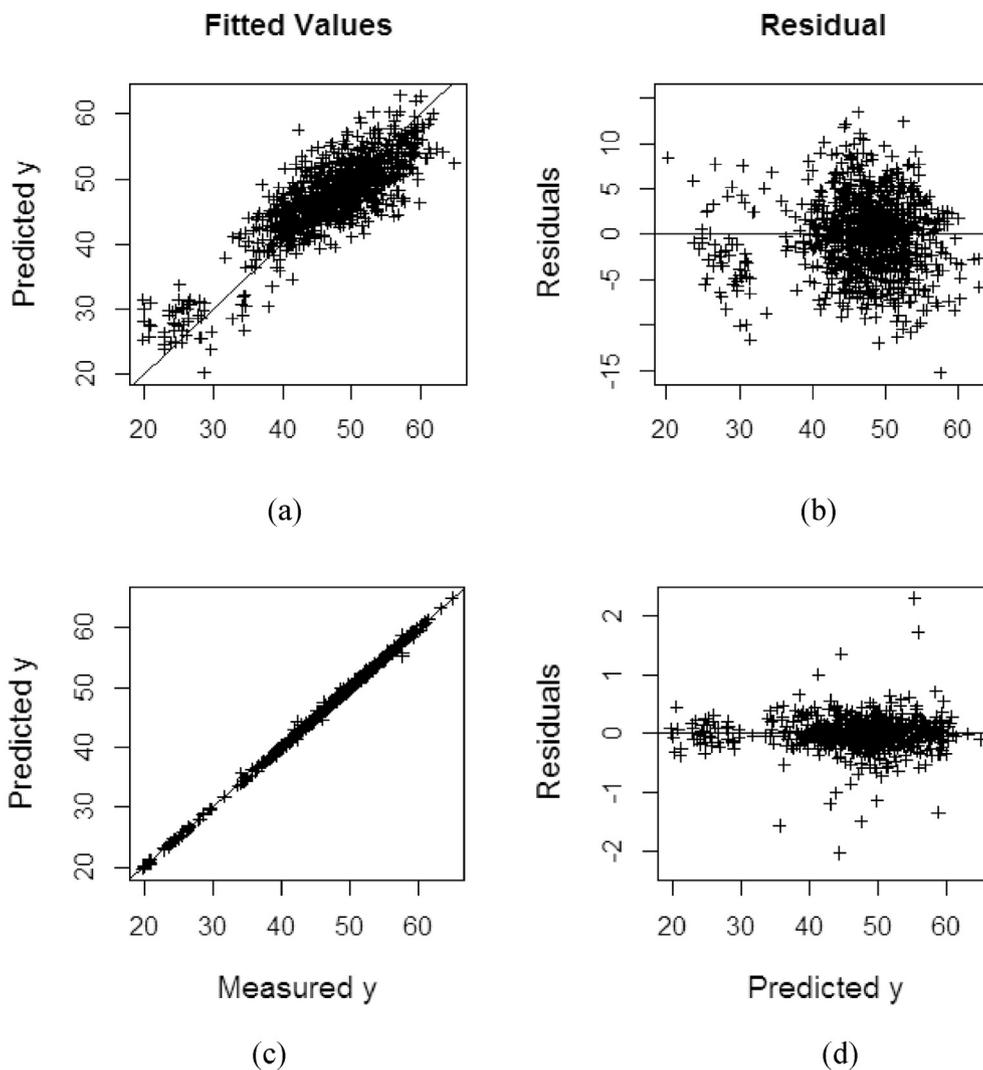
Figure 8. Fitting and residual performance on %OWM dataset: (a) measured vs predicted values on PRDRGP; (b) predicted vs residual on PRDRGP; (c) measured vs predicted values on KPDRGP; (d) predicted vs residual on KPDRGP.

5. Conclusions

The development of the robust DRGP algorithm in the PLSR method with kernel solution in the input space to handle the high dimensional problem and irregular data space has been discussed. Based on the study using simulation and real NIR spectral dataset, the proposed KPDRGP method was superior compared to the other upgraded methods using some considered robust methods. The KPDRGP succeed to capture the highly nonlinear relationship between the predictor variables against response variables through high-dimensional feature mapping. The reliability of the method also was evaluated; the result also shows the high promising using the method for routine prediction. With the contamination of the multiple outliers and high leverage points in the dataset, KPDRGP was not suffered otherwise the method prevents the influence during the fitting process.

Declarations

References

[1] M. Manley, Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials, Chem. Soc. Rev. 43 (2014) 8200–8214.
[2] H. Wold, in: P.R. Krishnaiah (Ed.), Multivariate Analysis, third ed., Academic Press, New York, 1973.
[3] R. Rosipal, Nonlinear Partial Least Squares: an Overview, N Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques, 2010, pp. 169–189.
[4] X. Ma, Y. Zhang, H. Cao, S. Zhang, Y. Zhou, Nonlinear regression with high-dimensional space mapping for blood component spectral quantitative analysis, J. Spectr. 2018 (2018) 1–8.
[5] Y. Liu, X. Sun, J. Zhou, H. Zhang, C. Yang, Linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis–NIR diffuse reflectance spectroscopy, Mathematical and Computer Modeling 51 (2010) 1438–1443.
[6] K.P. Bennett, M.J. Embrechts, An optimization perspective on kernel partial least squares regression, Nato Science Series Sub Series III Computer and Systems Sciences 190 (2003) 227–250.
[7] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in reproducing kernel hilbert space, J. Mach. Learn. Res. 2 (2001) 97–123.
[8] H. Yang, P.R. Griffiths, J.D. Tate, Comparison of partial least squares regression and multi-layer neural networks for quantification of nonlinear systems and application to gas phase Fourier transform infrared spectra, Anal. Chim. Acta 489 (2003) 125–136.
[9] V. Centner, J. Verdú-Andrés, B. Walczak, D. Jouan-Rimbaud, F. Despagne, L. Pasti, et al., Comparison of multivariate calibration techniques applied to experimental NIR data sets, Appl. Spectrosc. 54 (2000) 608–623.
[10] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction, Chemometr. Intell. Lab. Syst. 88 (2007) 183–188.
[11] V. Sindhwani, M.H. Quang, A.C. Lozano, Scalable matrix-valued kernel learning for high-dimensional nonlinear multivariate regression and granger causality, in: Proceedings of the Conference on Uncertainty in Articial Intelligence, UAI), 2013.
[12] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1998) 1299–1319.
[13] S. Rännar, F. Lindgren, P. Geladi, S. Wold, A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: theory and algorithm, J. Chemometr. 8 (1994) 111–125.
[14] C. Preda, Regression models for functional data by reproducing kernel Hilbert spaces methods, J. Stat. Plann. Inference 137 (2007) 829–840.
[15] G. Blanchard, N. Krämer, Kernel partial least squares is universally consistent, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2018, pp. 57–64.
[16] P.J. Huber, Robust regression: asymptotics, conjectures and Monte Carlo, Ann. Stat. 1 (1973) 799–821.
[17] V.J. Yohai, High breakdown-point and high efficiency robust estimates for regression, Ann. Stat. 15 (1987) 642–656.
[18] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, Robust statistics: the approach based on influence functions, Technometrics 29 (1987) 307–341.
[19] C.L. Mallows, On Some Topics in Robustness, Technical Memorandum, Bell Telephone Laboratories, 1975.
[20] C. Yu, W. Yao, Robust linear regression: a review and comparison, Commun. Stat. Simulat. Comput. 46 (2017) 6261–6282.
[21] E. Handschin, F. Schweppe, J. Kohlas, A.A.F.A. Fiechter, Bad data analysis for power system state estimation, IEEE Trans. Power Apparatus Syst. 94 (1975) 329–337.
[22] R. Maronna, O. Bustos, V. Yohai, Bias-and efficiency-robustness of general M-estimators for regression with random carriers, in: Smoothing Techniques for Curve Estimation, Springer, 1979, pp. 91–116.
[23] C.W. Coakley, T.P. Hettamansperger, A bounded influence, high breakdown, efficient regression estimator, J. Am. Stat. Assoc. 88 (1993) 872–880.
[24] P.J. Rousseeuw, Regression techniques with high breakdown point 12, The Institute of Mathematical Statistics Bulletin, 1983, p. 115.
[25] A.F. Siegel, Robust regression using repeated medians, Biometrika 69 (1982) 242–244.
[26] D.C. Hoaglin, R.E. Welsch, The hat matrix in regression and ANOVA, Am. Statistician 32 (1978) 17–22.
[27] D. Peña, F.J. Prieto, Multivariate outlier detection and robust covariance matrix estimation, Technometrics 43 (2001) 286–310.
[28] P.F. Velleman, R.E. Welsch, Efficient computing of regression diagnostics, Am. Statistician 35 (1981) 234–242.
[29] M. Habshah, M.R. Norazan, A.H.M.R. Imon, The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression, J. Appl. Stat. 36 (2009) 507–520.
[30] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, 1987, p. 589.
[31] P.J. Rousseeuw, V. Yohai, Regression by Means of S-Estimators, Robust and Nonlinear Time Series Analysis, Springer, 1984, pp. 256–272.
[32] T. Naes, H. Martens, Multivariate Calibration, John Wiley & Sons, 1992.
[33] S. Serneels, C. Croux, P. Filzmoser, P.J. Van Espen, Partial robust M-regression, Chemometr. Intell. Lab. Syst. 79 (2005) 55–64.
[34] D.J. Cummins, C.W. Andrews, Iteratively reweighted partial least squares: a performance analysis by Monte Carlo simulation, J. Chemometr. 9 (1995) 489–507.
[35] A.D. Chave, D.J. Thomson, A bounded influence regression estimator based on the statistics of the hat matrix, J. Roy. Stat. Soc.: Series C (Applied Statistics) 52 (2003) 307–322.
[36] A. Bagheri, H. Midi, M. Ganjali, S. Eftekhari, A comparison of various influential points diagnostic methods and robust regression approaches: reanalysis of interstitial lung disease data, Appl. Math. Sci. 28 (2010) 1367–1386.
[37] R.J. Carroll, A.H. Welsh, A note on asymmetry and robustness in linear regression, Am. Statistician 42 (1988) 285–287.
[38] Z. Du, D.P. Wiens, Jackknifing, weighting, diagnostics and variance estimation in generalized M-estimation, Stat. Probab. Lett. 46 (2000) 287–299.
[39] A.H.M.R. Imon, Identifying multiple influential observations in linear regression, J. Appl. Stat. 32 (2005) 929–946.
[40] A.H.M.R. Imon, M. Khan, A Solution to the problem of multicollinearity caused by the presence of multiple high leverage points, Int. J. Stat. Sci 2 (2003) 37–50.
[41] A.H.M.R. Imon, Identifying multiple high leverage points in linear regression, Journal Of Statistical Studies 3 (2002) 207–218.
[42] H. Wold, Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach, J. Appl. Probab. 12 (1975) 117–142.
[43] N. Aronszajn, Theory of reproducing kernels, Trans. Am. Math. Soc. 68 (1950) 337–404.
[44] H. Sun, Mercer theorem for RKHS on noncompact sets, J. Complex 21 (2005) 337–349.
[45] W. Wu, D. L Massart, S. De Jong, Kernel-PCA algorithms for wide data. Part II: fast cross-validation and application in classification of NIR data, Chemometr. Intell. Lab. Syst. 37 (1997) 271–280.
[46] W. Wu, D.L. Massart, S. De Jong, Kernel-PCA algorithms for wide data. Part I: theory and algorithms, Chemometr. Intell. Lab. Syst. 36 (1997) 165–172.
[47] R.D. Jia, Z.Z. Mao, Y.Q. Chang, S.N. Zhang, Kernel partial robust M-regression as a flexible robust nonlinear modeling technique, Chemometr. Intell. Lab. Syst. 100 (2010) 91–98.
[48] T. Zhang, An introduction to support vector machines and other kernel-based learning methods, AI Mag. 22 (2001) 103.
[49] H. Trautmann, The Desirability index as an Instrument for Multivariate Process Control, Dortmund University, Germany, 2004.