Received: 5 July 2023 | Revised: 29 October 2023 | Accepted: 7 November 2023

DOI: 10.1002/cem.3527

RESEARCH ARTICLE

Journal of CHEMOMETRICS WILEY

Iterative re-weighted multilinear partial least squares modelling for robust predictive modelling

Puneet Mishra[1] | **Kristian Hovde Liland**[2]

[1]Food and Biobased Research, Wageningen University and Research, Wageningen, The Netherlands

[2]Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway

Correspondence
Puneet Mishra, Food and Biobased Research, Wageningen University and Research, Wageningen, The Netherlands.
Email: puneet.mishra@wur.nl

Abstract

Higher order data are commonly encountered in the domain of chemometrics, often generated by advanced analytical instruments. Due to the multilinear nature of the data, higher order data require different regression approaches compared with traditional two-way data for predictive modelling. The main aim is usually to extract the rich multilinear information, which is often lost if the data are simply analysed in unfolded form. A common algorithm for multilinear predictive modelling is N-way partial least squares (NPLS). However, a limitation of NPLS is that it inherently does not handle outlying observations, which can be detrimental to the model. This work presents a new robust multilinear predictive modelling approach based on iterative down-weighting of the outlier observations in both predictor and response space. A key benefit of the method is that it only requires a single extra parameter to tune. In this work, the algorithm is described, and the method is demonstrated on three real multilinear data sets. In all cases, the presented method outperformed the traditional NPLS modelling regarding the root mean squared error of prediction.

KEYWORDS

data fusion, multiblock, multivariate, robustness, spectroscopy

1 | INTRODUCTION

Higher order (multilinear) data are commonly encountered in the domain of chemometrics. For example, signals generated by high-end analytical instruments such as 3D front face fluorescence spectroscopy,[1] spatially and time-resolved optical spectroscopy[2] and mass spectrometry.[3] Sometimes, simple techniques such as optical spectroscopy (originally 2D) may form multilinear data structures (such as 3D), for example, by combining with temporal dimensions. In usual chemometrics terminology, any data having higher order than two-way data are termed as multilinear/multiway data, and the associated methods to analyse such data are termed as multilinear/multiway analysis methods.[4] Also, data having multilinear/multiway structure are denoted as tensor or sometimes multiway arrays.

In recent decades, several developments have been seen in relation to methods for analysing multilinear data.[5,6] Some particular methods were tensor decomposition methods such as Tucker[7] and its variants,[8] parallel factor analysis (PARAFAC)[9] and its faster variant,[10] tensor factorisation[11] and, especially for predictive modelling, the multilinear partial least squares (PLS) called NPLS.[4] One way of analysing multilinear data is to rearrange them into two-way form and apply traditional two-way modelling approaches such as principal component analysis (PCA) or PLS. However,

past works have shown that using two-way methods on rearranged data can lead to less transparency and models of lower predictive power when a multilinear structure provides a good approximation of the data.[6] However, the multiway methods have advantages with respect to predictions and interpretability.[12] In several studies, the prediction quality was maintained/improved, compared with two-way analysis. Still, the multilinear models enabled much simpler interpretation due to the lower number of free parameters[6] and decomposition along the original modes of the data.

NPLS is a direct extension of two-way PLS regression and can handle data of any higher order. NPLS is particularly useful when the aim is to do predictive modelling just like traditional PLS analysis. Due to the popularity of NPLS and wide application areas, several extensions of NPLS have also been proposed such as the Shifted Covariates REgression Analysis for Multi-way data method which allows handling data shifts in particular modes of multilinear data[13] and the N-way canonical PLS[14] which provides an efficient and super fast approach to modelling multiple responses by avoiding tensor deflation (the most time-consuming part). Note that in some cases, there has also been an interest in pre-decomposing the multilinear data with the use of techniques such as PARAFAC and then using traditional two-way analysis on the obtained scores.[1] However, this approach is similar to doing a principal component regression (PCR), which is mainly based on variance, while PLS regression modelling the covariance between the data and response is considered a better alternative for achieving a similar predictive performance model as PCR but with a fewer number of components.

Although several methods are now available to do multilinear PLS regression, there is one unexplored area of research for multilinear predictive modelling. The unexplored area is multilinear predictive modelling in the presence of outlying samples. Note that for tensor decomposition and exploration, there were methods proposed earlier such as robust Tucker[15] and PARAFAC[16] but none for the multilinear predictive modelling. Outlying samples are frequent in the area of analytical measurements from improper measurements with sensors or sample handling and due to instrumental or human errors during reference analysis.[17,18] Note that a common approach in chemometric modelling is to first do outliers removal and later use the cleaned data for modelling. However, in this work, we refer to handling outliers directly during the calibration process by either deselection or down-weighting of the samples. Unlike multilinear modelling approaches, for two-way modelling, there is a wide range of robust methods available to handle outlying samples during the calibration process.[19,20] Some methods are *robust SIMPLS*,[21] *iterative re-weighted PLS*,[22] *partial robust M (PRM) regression*[23,24] and *RoBoost PLS/PLS2*.[25,26] The methods handle outliers with varying criteria ranging from an estimation of robust distances from the model centre to repeated weighting of the samples to adjust the covariance estimation process. However, most of the multilinear predictive methods, such as NPLS and N-way canonical PLS, are simple extensions of the PLS approach to latent space modelling. Hence, a robust multilinear method can be achieved as the extension of two-way robust PLS methods just as demonstrated previously for robust features selection[18] and robust data fusion.[17] In this study, we demonstrate the development of an iterative re-weighted multilinear PLS (irNPLS) modelling method. However, the extension of any robust PLS method to its multilinear form should be feasible.

Out of several robust techniques for robust two-way modelling, we decided to develop the iterative re-weighing approach because this approach is not based on the principle of outlier removal but on the concept of down-weighting the samples based on their degree of outlyingness. Outlier removal may result in a severe loss of information as long as the outliers still contain some valuable information, and thus, intelligent robust methods adapt the weights according to the outlyingness or inconsistency of the observations.[19] Down-weighting has particular importance when the model needs to be calibrated on a low number of calibration samples as the samples are not removed from the modelling.[19] Furthermore, outlier removal can be an extreme case of down-weighting where some extreme samples obtain almost zero weights in the re-weighting; however, this is totally data driven and the user does not interfere in the outlier detection or down-weighting. Recently, several new chemometric methods have emerged based on the concept of iterative re-weighting such as *RoBoost PLS/PLS2*,[25,26] *iterative weighted covariates selection*[18] and *iterative weighted multiblock modelling*.[17]

The aim of this study was to develop a new robust multilinear predictive method. The method is a combination of a robust two-way modelling technique called iterative re-weighted PLS (irPLS) and multilinear PLS modelling called NPLS. The method is the first robust multilinear method to do predictive modelling for higher order data in the presence of outlying observations. The algorithm and test of the method on real data sets are demonstrated below.

2 | MATERIALS AND METHOD

In the following, the linear algebra underlying the algorithm is presented, followed by a discussion about important steps in the algorithm. Later, the results of the method by testing it on real data sets are presented. All matrices and tensors are denoted with bold uppercase typeface such as $\mathbf{X}$ and $\underline{\mathbf{X}}$, all vectors are denoted with bold lowercase typeface such as $\mathbf{y}$, and all scalars are denoted with italic typeface such as a. Tensor multiplications follow the definitions explained in Liland et al,[14] but collapsing dimensions are assumed implicitly as either the sample dimension or all feature/variable dimensions, that is, the tensor-matrix product of dimensions $(n \times a \times b) \times (a \times b)$ collapses to a vector of dimension $n \times 1$.

2.1 | Algorithm

Define $\mathbf{y}$ $(n \times 1)$ as the response vector, $\underline{\mathbf{X}}$ as a tensor with n samples in the first mode, and let A be the desired number of components to be extracted. Let $\mathbf{D}$ be the initial sample weights. Note that $\mathbf{D}$ is a diagonal matrix initially having $1/n$ as the weight for all samples for each component. Both the predictor and the response are assumed to be median centred (less influence of outliers than mean centred).[17,18] Let α be the tuning parameter defining the aggressiveness in weighting down outliers. This will be further discussed after the algorithm.

Algorithm for iterative weighted multilinear partial least squares modelling

$for\ a = 1 : A$	- loop over A components to be extracted
$\quad while\ crit > 10^{-5}$	- loop for sample re-weighting
$\quad\quad \mathbf{V} = \underline{\mathbf{X}}^t \mathbf{D} \mathbf{y}$	- candidate loading weights
$\quad\quad switch\ (d = dim(\underline{\mathbf{V}}))$	- processing alternatives according to the $\underline{\mathbf{V}}$-dimensionality
$\quad\quad\quad d = 1 : \mathbf{v}^{[1]} = \mathbf{v}/\lVert\mathbf{v}\rVert$	- $\underline{\mathbf{V}} = \mathbf{v}$ (vector): only normalise
$\quad\quad\quad d = 2 : \mathbf{v}^{[1]}, \mathbf{v}^{[2]} \Leftarrow svd(\mathbf{V})$	- $\underline{\mathbf{V}} = \mathbf{V}$ (matrix): use its dominant pair of left-/right singular vectors
$\quad\quad\quad d \geq 3 : \mathbf{v}^{[1]}, ..., \mathbf{v}^{[d]} \Leftarrow parafac(\underline{\mathbf{V}})$	- $\underline{\mathbf{V}}$ (tensor): use its one-component PARAFAC model vectors
$\quad\quad end$	
$\quad\quad \underline{\mathbf{V}}^O = outer(\mathbf{v}^{[1]}, ..., \mathbf{v}^{[d]}) \rightarrow \underline{\mathbf{V}}_A^O$	- form loading weight tensor from vector outer product and accumulate
$\quad\quad \mathbf{t}_a = \frac{\underline{\mathbf{X}}\mathbf{V}^O}{\lVert\underline{\mathbf{X}}\mathbf{V}^O\rVert}$	- estimate normalised score vector
$\quad\quad \mathbf{q}_a = \mathbf{y}^t \mathbf{D} \mathbf{t}_a$	- temporary regression coefficients
$\quad\quad \mathbf{r} = \mathbf{y} - \mathbf{t}_a \mathbf{q}_a^t$	- estimate residuals
$\quad\quad \mathbf{r} = \mathbf{r}/\sqrt{1\text{-diag}(\mathbf{t}_a \mathbf{t}_a^t)}$	- compute adjusted residuals
$\quad\quad \mathbf{u} = \frac{\mathbf{r} \times 0.6745}{\alpha \times \text{ MAD}(\mathbf{r})}$	- standardise the residuals with median absolute deviation (MAD)
$\quad\quad for\ i = 1 : n$	- loop over samples
$\quad\quad\quad if\ \lvert\mathbf{u}_i\rvert > 1$	- limit large residuals
$\quad\quad\quad\quad \mathbf{r}_i \Leftarrow 0$	(implemented as element-wise replacement)
$\quad\quad\quad else$	
$\quad\quad\quad\quad \mathbf{r}_i \Leftarrow (1 - \mathbf{u}_i{}^2)^2$	- bisquare function based weight estimation
$\quad\quad\quad end$	
$\quad\quad end$	
$\quad\quad crit = \sum(\lvert\mathbf{r}\rvert - \lvert\text{diag}(\mathbf{D})\rvert)$	- update criterion for loop
$\quad\quad \mathbf{D} \Leftarrow \mathbf{I}_n \odot \mathbf{r}$	- update weight matrix with $\mathbf{r}$ on the diagonal, otherwise zeros
$\quad end$	- end of re-weighting loop
$\quad \mathbf{y} \Leftarrow \mathbf{y} - \mathbf{t}_a \mathbf{q}_a^t$	- robust $\mathbf{y}$ deflation
$\quad \mathbf{P}_a = \underline{\mathbf{X}}^t \mathbf{D} \mathbf{t}_a$	- storing robust $\mathbf{X}$ loadings
$\quad \mathbf{W}_a = \mathbf{V}^O$	- storing robust loading weights
$\quad \underline{\mathbf{X}} \Leftarrow \underline{\mathbf{X}} - \mathbf{t}_a \mathbf{P}_a^t$	- robust $\underline{\mathbf{X}}$ deflation
end	- end of component loop
$\mathbf{R} = \mathbf{W}(\mathbf{P}^t\mathbf{W})^{-1}$	- projections for score prediction
$\mathbf{B} = \text{cumsum}(\mathbf{R}\mathbf{q}^t)$	- regression coefficients
$\mathbf{B}_0 = \bar{\mathbf{Y}} - \bar{\underline{\mathbf{X}}}\mathbf{B}$	- median compensation

2.2 | Comments on the algorithm

NPLS is a direct extension of two-way PLS modelling. The key difference in multilinear PLS, compared with two-way PLS modelling, lies in how the high-dimensional covariance matrix/tensor is handled during covariance estimation. When estimating covariance for two-way data in the PLS modelling step, it results in a 1D vector. However, for three-way data, it becomes a matrix, and for four-way data, it becomes a 3D tensor, and so on. In summary, the dimension of the estimated covariance for higher order data is equivalent to the number of variables in each mode. To manage such high-dimensional covariance matrices/tensors, the main approach is to perform singular value decomposition (SVD) when the covariance is a matrix or PARAFAC when the covariance is a tensor. It is important to note that SVD and PARAFAC are used to extract one factor at a time, with the assumption that the single factor captures the dominant variation in the covariance. The resulting loadings from SVD and PARAFAC are then utilised for score estimation.

In the proposed method, it is hypothesised that the presence of outlying observations can contaminate the estimation of the covariances. Hence, the SVD and PARAFAC of the covariance matrix/tensor will carry over this contamination during the estimation of the scores. To deal with this, the key idea is to down-weight any potential outliers in the data set such that the estimated covariance is minimally contaminated by the noise of the outliers. The proposed algorithm deploys an iterative re-weighting step which iteratively estimates the sample weights using joint information of the residuals and leverages estimated using the scores. By using the joint information from residual and leverages estimated using scores, the method handles outliers in both the predictor and response. The iterative weighting idea is inspired by chemometric methods proposed three decades back[22,27] and some recent methods.[17,18,25,26]

In the older proposed iterative re-weighting PLS approaches,[22,27] the main aim was to down-weight the outliers based on information only about the $\mathbf{y}$ residuals. Only using $\mathbf{y}$ residuals means that the earlier methods were only able to deal with outliers present in the response, which in most of the cases is sufficient as more errors are made during reference wet chemistry analysis than analytical measurements with instruments. However, in many cases, outliers can also be present in predictors requiring a robust additional criterion to deal with outliers in both predictor and response. Some recent methods[25,26] proposed using additional estimates apart from $\mathbf{y}$ residuals such as scores and $\mathbf{X}$ residuals to estimate sample weights. Later, methods estimated a single weight per sample by taking the element-wise product of the weights estimated by different criteria. However, a main challenge with the approaches using multiple different criteria is that the number of parameters to optimise increased proportionally. Such a large number of parameters can be difficult to optimise when a low number of samples are available in the calibration set. In our earlier works, we have proposed a new approach to jointly use the information from scores and $\mathbf{y}$ residuals without increasing the number of parameters to optimise. The approach is based on the estimation of the adjusted residuals (Equation 1) using the leverages estimated using the estimated scores[17,18]:

$$r_{i,adj} = \frac{r_i}{\sqrt{1-h_i}}, \tag{1}$$

where r_i are the ordinary least squares (OLS) residuals and h_i are the least squares fit leverage values estimated using the scores. The leverage values h_i are bounded between 0 and 1 ($0 \leq h_i \leq 1$). A leverage value h_i close to 1 indicates that $\mathbf{x}_i$ is deemed an extreme value (or outlier relative to other samples in the calibration set). Leverages adjust the residuals by increasing the weight of high-leverage data points that have a large effect on the least squares fit, that is, exaggerating the degree of outlyingness that the subsequent weighting is based on. Adjusted residuals are then standardised as in Equation (2).

$$u_i = \frac{r_{i,adj}}{\alpha s} = \frac{r_i}{\alpha s\sqrt{1-h_i}} = \frac{t_i}{\alpha}, \tag{2}$$

where α is a tuning constant and s is an estimate of the standard deviation of the error term given by $s = \mathrm{MAD}/0.6745$. MAD is the median absolute deviation of the adjusted residuals from their median. The constant 0.6745 makes the estimate unbiased for the normal distribution. Note that the residuals are later replaced by the bisquare function $\mathbf{r}_i \Leftarrow (1-\mathbf{u}_i^2)^2$. The tuning constant α defines the aggressiveness towards down-weighting outliers. For example, when $\alpha \to \infty, u_i \leftarrow 0$, then all samples will be given equal weights ($r_i \leftarrow 1$) and the algorithm will converge to NPLS. As the $\alpha \to 0$, the method will become highly aggressive and down-weighting inliers. It is important to tune α using validation approaches like earlier studies.[17,18,26] However, note that compared with other earlier methods,[25,26] our proposed

approach only requires optimisation of one additional parameter α, apart from the total number of latent variables, to deal with outliers in both predictor and response. The single extra parameter can be easily optimised either using a validation set or cross-validation. Note that Equation (2) is similar to estimating studentised residuals t_i. In this way, estimation of u_i is similar to estimation of t_i/α.

The sample weights are modified using the bisquare function. Bisquare-based weighting is commonly implemented in iterative re-weighting methods[17,18,22,25,26,28] due to its simplicity in implementation and efficient operation. Readers are free to choose from many weighting functions as presented in Cummins and Andrews.[22] Note that the input to any new weighting function will be standardised adjusted residuals as estimated in Equation (2). In the presented algorithm, Y-residuals and leverage information from the scores at every latent variable extraction step were used for estimating sample weights. Some earlier studies have also utilised spectral residuals (X-residuals).[26] The current algorithm can also be modified to include X-residual information. For example, during the estimation of the leverages, an average of in-space and out-of-space leverage can be computed. However, in the current study, incorporating X-residual information did not result in any improvement in the predictive performance of the models.

Note that in the literature, methods exist for decomposing higher dimensional arrays such as robust Tucker[15] and PARAFAC.[16] However, in this study, the aim was predictive modelling based on covariance maximisation; hence, robust versions of Tucker and PARAFAC were not used. Although to further refine the method, the current PARAFAC step of decomposing covariance tensor can be replaced with robust PARAFAC.

2.3 | Data sets for method demonstration

This study uses three different real multilinear data sets openly available from https://ucphchemometrics.com/. All three data sets were related to excitation–emission fluorescence and associated wet chemistry analysis performed on the same samples. In the following, each data set is described. The irNPLS model performance was compared with an NPLS model. In the case of the Enzyme data set, there was already data partitioned into calibration and test sets. For the yoghurt and sugar data sets, one out of three samples were selected for the test set and the remaining samples were used for model cross-validation and calibration. Model parameters were optimised using five-fold cross-validation. The number of latent variables and the α parameter search space were both in the interval of Integers 1–20. Cross-validation was performed on the grid of all 400 pairs. Note that the same algorithm was used for both irNPLS and NPLS. Only for NPLS, the α parameter was set to infinity and median centring was replaced by mean centring to give equal weights to all samples. The minimum of the cross-validation error curve (surface in case of irNPLS) was used to select the optimal number of latent variables and α parameter. The model trained with optimal parameters was tested on the test set. The MATLAB code for the algorithm is available at https://github.com/puneetmishra2/IRNPLS.

2.3.1 | Enzyme activity during fermentation

Data were first published in Mortensen and Bro[29] and related to the monitoring of fed-batch cultivation in a bioreactor. Sampling was performed (100 mL) from the bioreactor four to seven times during a batch run. In total, 25 batches were monitored. At-line multi-channel fluorescence measurements were performed on homogenised culture broth (10 mL). The samples were later analysed for protease degrading *N*,*N*-dimethyl-casein. The coloured complex developed by the reaction of the enzyme with 2,4,6-trinitrobenzene sulfonic was determined spectroscopically. Data were generated with 15 excitation filters ranging from 270 to 550 nm with a spectral resolution of 20 nm and 15 emission filters ranging from 310 to 590 nm also with a spectral resolution of 20 nm. In the reported literature there was a mention of a total of 283 samples,[29] while the downloaded data set had a total of 338 samples in the calibration set and 53 samples in the test set. In this work, we compared the NPLS model with the irNPLS model from scratch; hence, we used all the available samples in the data set. A drawback of this is that the results cannot be directly compared with the results earlier achieved in Mortensen and Bro.[29]

2.3.2 | Sugar ash content

Data were first published in Bro[1] and consisted of 268 excitation–emission fluorescence measurements of sugar samples. The emission spectra from 275 to 560 nm were measured in 0.5 nm intervals (571 wavelengths) at seven excitation

wavelengths (230, 240, 255, 290, 305, 325 and 340 nm). Samples were measured during 3 months of operation in late autumn from a sugar plant in Scandinavia. As a reference parameter and measure of inorganic impurities in the refined sugar, ash content was measured. Ash content was determined by conductivity.[1] There were a total of 178 samples in the calibration set and 90 samples in the test set.

2.3.3 | Parma ham age prediction

Data were first published in Møller et al[30] and consisted of excitation–emission fluorescence measurements on Parma ham samples. The aim was to predict the age of Parma ham with fluorescence measurements. Data were generated with 15 excitation filters ranging from 270 to 550 nm with a spectral resolution of 20 nm and 15 emission filters ranging from 310 to 590 nm also with a spectral resolution of 20 nm. In total, 67 samples of Parma ham were measured. There were a total of 44 samples in the calibration set and 23 samples in the test set.

3 | RESULTS

A summary of NPLS and irNPLS analyses for the fermentation data is shown in Figure 1. Cross-validation analysis results for NPLS are shown in (Figure 1A) and for irNPLS in (Figure 1B). For NPLS, cross-validation errors are presented as a function of the number of latent variables, while for irNPLS, errors are presented as a function of both the latent variables and the α parameter. For irNPLS (Figure 1B), for the majority of α parameters, the cross-validation error for irNPLS was lower than NPLS. For example, the irNPLS achieved a similar cross-validation error as NPLS but with a lower number of latent variables. For NPLS, the minimum was found at 16 latent variables (Figure 1A). For

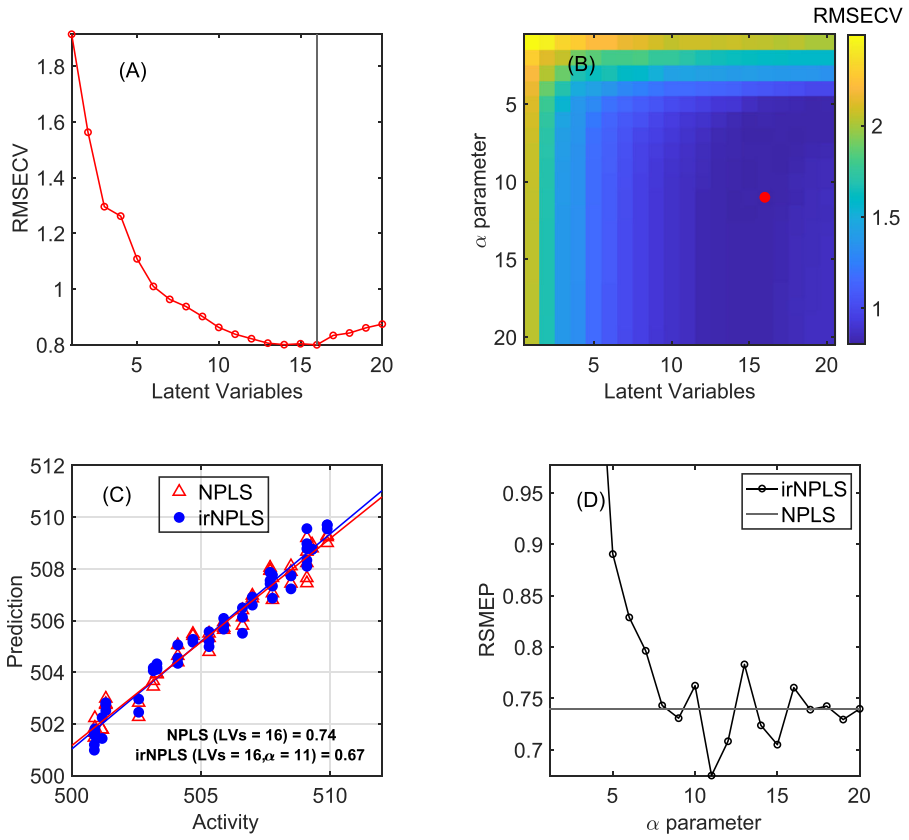

FIGURE 1 Fermentation data analysis results: (A) cross-validation trends for NPLS with a vertical line indicating the minimum, (B) cross-validation trends for irNPLS with a red dot indicating the minimum, (C) prediction plot with $y = x$ line in black and (D) posterior analysis to compare irNPLS and NPLS as a function of the α parameter. LVs, latent variables.

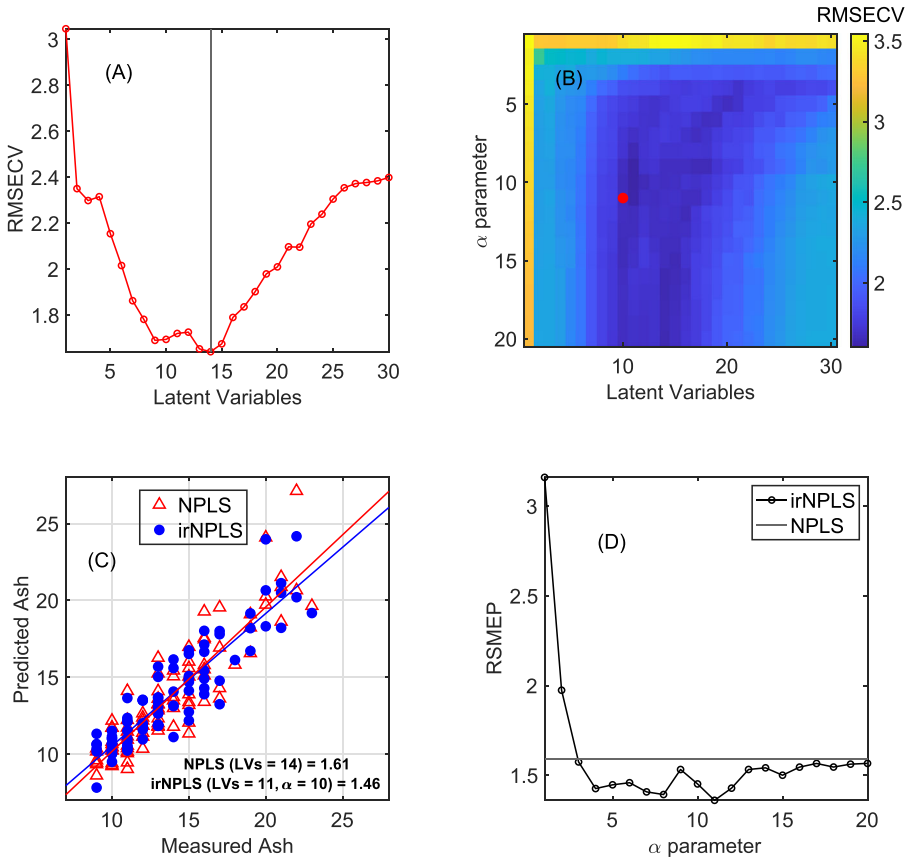

FIGURE 2 Sugar data analysis results: (A) cross-validation trends for NPLS with a vertical line indicating the minimum, (B) cross-validation trends for irNPLS with a red dot indicating the minimum, (C) prediction plot with $y = x$ line in black and (D) posterior analysis to compare irNPLS and NPLS as a function of the α parameter. LVs, latent variables.

irNPLS, the minimum was noted at 16 latent variables and α value of 11. The model created with optimal parameters was tested on the test set, and the results are shown in Figure 1C. As can be noted, compared with NPLS, the irNPLS decreased the root mean squared error of prediction (RMSEP) from 0.74 to 0.67. The irNPLS achieved lower error compared with NPLS using the same number of latent variables. The error decrease has directly to do with samples re-weighting as the posterior analysis of the data with the same number of latent variables (16 for NPLS and irNPLS) but different α parameter demonstrated that irNPLS achieved lower errors (Figure 1D) when $\alpha > 10$.

For the sugar data set as well, the irNPLS model achieved a lower RMSEP of 1.46 compared with 1.61 achieved with NPLS (Figure 2C). The cross-validation analysis (Figure 2A,B) also showed that the irNPLS model achieved lower cross-validation errors with a lower number of latent variables compared with the NPLS model (red trend). The total number of latent variables for NPLS was 14 and for irPLS was 11 with α parameter as 10 at the optimal RMSECV. The posterior analysis of the data with the same number of latent variables (14 for NPLS and irNPLS) but different α parameter values demonstrated that irNPLS achieved lower errors in general (Figure 2D).

For the Parma ham data set, the RMSEP for irNPLS was also lower than NPLS (Figure 3C). The NPLS model was based on seven latent variables, and the irNPLS model was based on seven latent variables with α parameter of 20. In general, the cross-validation analysis for the majority of α values attained lower errors than the NPLS cross-validation errors (Figure 3A,B). Especially, after seven latent variables, the cross-validation error of NPLS analysis increased while for irNPLS remained stable. This can be an indication that irNPLS is less prone to over-fitting as it focuses on learning refined information with the sample re-weighting. A similar trend of cross-validation errors was noted for all the earlier data sets as well. The posterior analysis of the data with the same number of latent variables (seven for NPLS and irNPLS) but different α parameter values demonstrated that irNPLS achieved lower errors with higher α values (Figure 3D).

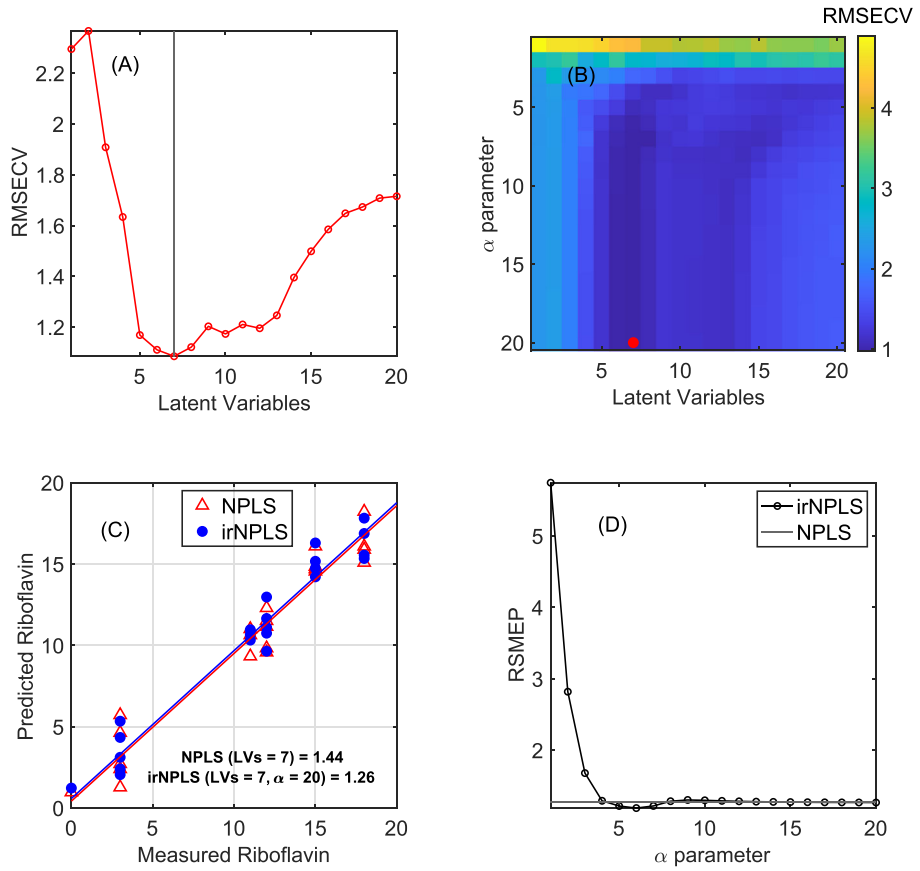

FIGURE 3 Parma ham data analysis results: (A) cross-validation trends for NPLS with a vertical line indicating the minimum, (B) cross-validation trends for irNPLS with a red dot indicating the minimum, (C) prediction plot with $y = x$ line in black and (D) posterior analysis to compare irNPLS and NPLS as a function of the α parameter. LVs, latent variables.

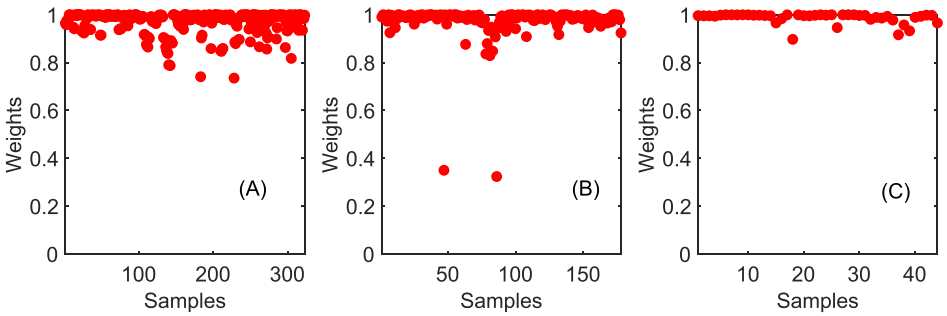

FIGURE 4 Sample weights obtained for different data set modelling scenarios: (A) fermentation data, (B) sugar data and (C) Parma ham data.

All three data sets had some samples which should not be given equal weights, as it can lead to detrimental model performance as noted in the earlier analysis. To demonstrate that irNPLS was actually able to handle different outlying samples, sample weights for all three data sets for the first latent variable are shown in Figure 4. As can be noted, the fermentation data (Figure 4A) had minimal outliers but many samples attained lower than 1 weight and a couple had weights from around 0.8 and down. For the sugar data set (Figure 4B), there were two extreme outlier achieving 0.4 weight while several other outliers attained weights between 0.9 and 0. For the Parma ham data set (Figure 4C), there were some samples that achieved weights lower than 1.

In the earlier analyses, the RMSEP was primarily highlighted as a comparison measure. However, to gain more detailed insights into the improvements achieved with the addition of the iterative re-weighting step, further detailed

statistics were estimated and plotted in Figure 5. The 'r' parameter was estimated as the slope of regression line of predicted versus true values. The 'bias' was estimated as average difference between true and predicted values. The standard error of prediction (SEP) was estimated as square root of the difference between squared RMSEP and squared bias. For the enzymatic fermentation data (Figure 5A), it can be noted that the reduction in RMSEP was both due to a reduction in bias and SEP. The 'r' values were similar for both NPLS and irNPLS. In fact, the 'r' values were very similar for all three data sets. For the sugar data (Figure 5B), the reduction in RMSEP was mainly due to SEP, as the bias was similar (or close) for both NPLS and irNPLS. For the Parma ham data (Figure 5C), the reduction in RMSEP was both due to a reduction in bias and SEP.

4 | DISCUSSION AND FUTURE WORK

We have presented a novel approach to multilinear PLS modelling which intelligently handles outlying samples during the calibration processes. This in practical terms means that the user does not need to perform additional outlier removal analysis as the method automatically does that during modelling, thus saving additional computing and analysis time. Furthermore, unlike the traditional outlier detection approaches where the user often ends up removing the outlying samples, this method only down-weights the samples based on their level of outlyingness. Outlier removal is a special case of this method, and it happens when the samples are extreme and assigned zero weights; however, such cases are purely data driven. Furthermore, the method uses information from both the residuals and the scores to estimate adjusted residuals for sample weight estimation. Hence, the method is sufficient to handle both predictor and response outliers. The testing of the method on three real data set scenarios showed that the method achieved lower root mean squared errors of prediction compared with the NPLS algorithm. The proposed algorithm can also be considered a new generic NPLS algorithm because NPLS is a special case of the presented algorithm. The NPLS model can easily be obtained by setting the α parameter to infinity. Doing this will assign equal weights to all samples leading to NPLS. The current algorithm is also suitable for doing traditional PLS and irPLS analysis, as PLS is a special case when loading weights are simply vectors. We foresee wide applications of the method, especially not only for cases where users suspect the presence of outliers but also for general PLS modelling as the method is automatic in terms of dealing with outliers.

The proposed algorithm can be considered as the backbone for the development of further new multilinear methods. For example, the same algorithm can be adapted to build a multi-response algorithm which can be used to do joint prediction of multiple responses or classification modelling. The modification to deal with multiple responses can be achieved by implementing the multilinear canonical PLS approach.[14] The extension of multilinear canonical PLS is simple and can be achieved by replacing the following steps for loading weight estimation. At first, candidate loading weights are estimated as $\underline{\mathbf{W}} = \underline{\mathbf{X}}^t\mathbf{Y}$. Then, candidate scores are estimated as $\mathbf{Z} = \underline{\mathbf{X}}\underline{\mathbf{W}}$. Finally, weighted canonical analysis is performed, the dominant left canonical weights are retained $\mathbf{c} \Leftarrow canoncorr(\mathbf{Z}, \mathbf{Y}, \mathbf{D})$ and the loading weights are computed as $\underline{\mathbf{V}} = \underline{\mathbf{W}}\mathbf{c}$ before being subjected to normalisation/SVD/PARAFAC. Furthermore, the residuals from multiple responses can be jointly handled as a product as done in earlier works.[18,26] The same algorithm can also be adapted to do higher order robust feature selection[18] for multilinear data with methods such as N-way covariates selection.[31] The feature selection step will involve binarising the robust loading weight vector using a criterion such as maximum covariance. Developing further new robust methods inspired by the presented algorithm is a direction of future work.

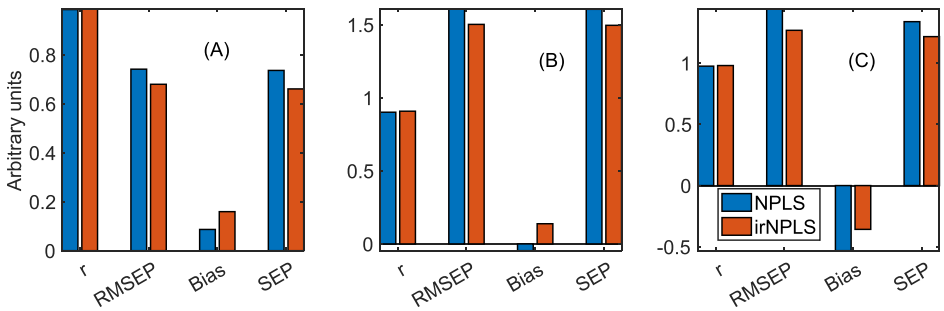

FIGURE 5 Detailed prediction statistics: (A) fermentation data, (B) sugar data and (C) Parma ham data.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/cem.3527.

DATA AVAILABILITY STATEMENT

All data used in this work are publicly available.

ORCID

Puneet Mishra https://orcid.org/0000-0001-8895-798X
Kristian Hovde Liland https://orcid.org/0000-0001-6468-9423

REFERENCES

1. Bro R. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemom Int Lab Syst*. 1999;46(2):133-147.
2. Igne B, Talwar S, Feng H, Drennen JK, Anderson CA. Near-infrared spatially resolved spectroscopy for tablet quality determination. *J Pharmaceut Sci*. 2015;104(12):4074-4081.
3. Johnsen LG, Skou PB, Khakimov B, Bro R. Gas chromatography–mass spectrometry data processing made easy. *J Chromatogr A*. 2017;1503:57-64.
4. Bro R. Multiway calibration. Multilinear PLS. *J Chemom*. 1996;10(1):47-61.
5. Coppi R. An introduction to multiway data and their analysis. *Comput Stat Data Anal*. 1994;18(1):3-13.
6. Andersson CA, Bro R. The N-way Toolbox for MATLAB. *Chemom Int Lab Syst*. 2000;52(1):1-4.
7. Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. *J Math Phys*. 1927;6(1–4):164-189.
8. Kolda TG, Bader BW. Tensor decompositions and applications. *SIAM Rev*. 2009;51(3):455-500.
9. Bro R. PARAFAC. Tutorial and applications. *Chemom Int Lab Syst*. 1997;38(2):149-171.
10. Yu H, Augustijn D, Bro R. Accelerating PARAFAC2 algorithms for non-negative complex tensor decomposition. *Chemom Int Lab Syst*. 2021;214:104312.
11. Takeuchi K, Tomioka R, Ishiguro K, Kimura A, Sawada H. Non-negative multiple tensor factorization. In: 2013 IEEE 13th International Conference on Data Mining. IEEE; 2013:1199-1204.
12. Bro R, Heimdal H. Enzymatic browning of vegetables. Calibration and analysis of variance by multiway methods. *Chemom Int Lab Syst*. 1996;34(1):85-102.
13. Marini F, Bro R. SCREAM: a novel method for multi-way regression problems with shifts and shape changes in one mode. *Chemom Int Lab Syst*. 2013;129:64-75.
14. Liland KH, Indahl UG, Skogholt J, Mishra P. The canonical partial least squares approach to analysing multiway datasets—N-CPLS. *J Chemom*. 2022;36(7):e3432.
15. Pravdova V, Estienne F, Walczak B, Massart DL. A robust version of the Tucker3 model. *Chemom Int Lab Syst*. 2001;59(1-2):75-88.
16. Engelen S, Frosch S, Jørgensen BM. A fully robust parafac method for analyzing fluorescence data. *J Chemom: A J Chemom Soc*. 2009;23(3):124-131.
17. Mishra P, Liland KH. An algorithm for robust multiblock partial least squares predictive modelling. *J Chemom*. 2023;2023:e3480.
18. Mishra P, Liland KH. Iterative re-weighted covariates selection for robust feature selection modelling in the presence of outliers (irCovSel). *J Chemom*. 2023;37(2):e3458.
19. Filzmoser P, Todorov V. Review of robust multivariate statistical methods in high dimension. *Anal Chim Acta*. 2011;705(1-2):2-14.
20. Møller SF, von Frese Frese J, Bro R. Robust methods for multivariate data analysis. *J Chemom: A J Chemom Soc*. 2005;19(10):549-563.
21. Hubert M, Branden KV. Robust methods for partial least squares regression. *J Chemom: A J Chemom Soc*. 2003;17(10):537-549.
22. Cummins DJ, Andrews CW. Iteratively reweighted partial least squares: a performance analysis by Monte Carlo simulation. *J Chemom*. 1995;9(6):489-507.
23. Serneels S, Croux C, Filzmoser P, Van Espen PJ. Partial robust M-regression. *Chemom Int Lab Syst*. 2005;79(1-2):55-64.
24. Hoffmann I, Serneels S, Filzmoser P, Croux C. Sparse partial robust M regression. *Chemom Int Lab Syst*. 2015;149:50-59.
25. Metz M, Abdelghafour F, Roger J-M, Lesnoff M. A novel robust PLS regression method inspired from boosting principles: RoBoost-PLSR. *Anal Chim Acta*. 2021;1179:338823.
26. Metz M, Ryckewaert M, Mas-Garcia S, et al. RoBoost-PLS2-R: an extension of RoBoost-PLSR method for multi-response. *Chemom Int Lab Syst*. 2022;222:104498.
27. Wakelinc IN, Macfie HJH. A robust PLS procedure. *J Chemom*. 1992;6(4):189-198.
28. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J Chemom: A J Chemom Soc*. 1998;12(5):301-321.
29. Mortensen PP, Bro R. Real-time monitoring and chemical profiling of a cultivation process. *Chemom Int Lab Syst*. 2006;84(1–2):106-113.
30. Møller JKS, Parolari G, Gabba L, Christensen J, Skibsted LH. Monitoring chemical changes of dry-cured Parma ham during processing by surface autofluorescence spectroscopy. *J Agri Food Chem*. 2003;51(5):1224-1230.

31. Mishra P, Liland KH, Indahl UG. Swiss knife covariates selection: a unified algorithm for covariates selection in single block, multi-block, multiway, multiway multiblock cases including multiple responses. *J Chemom*. 2022;36(10):e3441.

How to cite this article: Mishra P, Liland KH. Iterative re-weighted multilinear partial least squares modelling for robust predictive modelling. *Journal of Chemometrics*. 2024;38(1):e3527. doi:10.1002/cem.3527