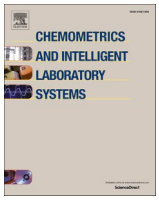
Chemometrics and Intelligent Laboratory Systems 240 (2023) 104883

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



# J-Score: A new joint parameter for PLSR model performance evaluation of spectroscopic data

Jokin Ezenarro [a], Daniel Schorn-García [a], Laura Aceña [a], Montserrat Mestres [a], Olga Busto [a], Ricard Boqué [b,*]

[a] Universitat Rovira i Virgili. Instrumental Sensometry (iSens), Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo 1, Tarragona, 43007, Spain
[b] Universitat Rovira i Virgili. Chemometrics, Qualimetrics and Nanosensors Group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo 1, Tarragona, 43007, Spain

ARTICLE INFO

*Keywords:*
Partial least squares regression
Latent variables
Root mean square error
Validation
Vibrational spectroscopy
Preprocessing

ABSTRACT

Since its beginnings, many parameters have been proposed to evaluate the goodness of Partial Least Squares Regression (PLSR) models and thus help chemometricians to choose the most appropriate one. This article proposes a new performance evaluation parameter for regression models based on spectroscopic data, the J-Score, which combines some of the most commonly used model evaluation parameters (Ratio of Performance to Deviation, Calibration and Validation Root Mean Square Errors and Regression Vector) into a single indicator. The J-Score can help non-experienced analysts select both the adequate number of Latent Variables (LVs) and the best preprocessing technique for their dataset in an automated way. The performance of the J-Score has been compared to other evaluation methods with different datasets, demonstrating that it can be used for different types of samples and spectroscopic data; that it is stable and objective, and offers an easy way to select the optimal number of LVs.

## 1. Introduction

Partial Least Squares Regression (PLSR) is one of the most important algorithms in multivariate data analysis for its simplicity, versatility and applicability, as it can fit multiple response variables in a single model and makes interpretation of results more intuitive. This method is based on finding new dimensions (Latent Variables, LVs) from the original data by maximizing the covariance of the **X** and **Y** blocks (predictors and predicted values), reducing the dimensionality and using the new variables to build a regression equation [1]. Thus, it is especially useful when the predictors are highly collinear or when the number of predictors (independent variables) is higher than the number of observations and ordinary least-squares regression fails. This is why the PLSR algorithm is widely used in spectroscopic measurements, which are the basis of relatively low-cost and fast analysis methods and constitute one of the broadest research areas of analytical chemistry [2].

Spectroscopic data being the source of a large amount of information implies the need to apply a deep data analysis, for which multivariate data analysis techniques, such as the mentioned PLSR, are needed. However, data analysis needs to be supervised by an expert analyst with enough experience to build regression models and ensure good performance. Thus, the analyst, after making the data appropriate for the modelling process by removing outliers and preprocessing the spectra, must select the optimum model dimensionality (also known as model rank), which is the number of LVs that constitute the multivariate regression model [3]. Then, to decide if the regression model is adequate, the analyst must choose the most suitable quality assessment parameters among the existing ones, taking into account that some of them account for the same properties of the model while some others are complementary [4]. For example, typical properties of the model are: the shape of the curves of Root Mean Squared Error (RMSE) of Calibration ($RMSE_{Cal}$), Cross Validation ($RMSE_{CV}$) or Prediction ($RMSE_{P}$) vs. the number of Latent Variables (LV); the values of Explained Variance (EV) by each LV; the shape of the loadings and the Regression Vector (RV); or the plot of Q residuals vs Hotelling $T^2$ values. However, it should be noted that the adequacy of these values and plots is usually decided based on the opinion of the analyst, which comes from experience and training, but sometimes may not be objective enough.

* Corresponding author.
*E-mail address:* ricard.boque@urv.cat (R. Boqué).

https://doi.org/10.1016/j.chemolab.2023.104883
Received 31 January 2023; Received in revised form 1 June 2023; Accepted 5 June 2023
Available online 7 June 2023

Therefore, based on all the mentioned statistical parameters and plots the PLS algorithm can offer, several model performance evaluation parameters have been proposed, such as PRESS (Predicted Residual Error Sum of Squares) [5], Average Variance Extracted (AVE) [6], Composite Reliability ($\rho_c$) [6] or Ratio of Prediction to Deviation (RPD) [7]. These parameters only rely on one of the properties of the PLSR model, such as the EV or RMSE. This last parameter is the most commonly used but it must be used carefully as it could provide over-optimistic results if the regression is overfitted to the calibration data and the model cannot be generalized or if the validation dataset is not representative enough [8,9].

In this work, a novel tool for evaluating the overall performance PLSR models for spectroscopic data quantitatively is proposed, by considering several model properties and describing its suitability in a global way. A tool that allows non-experienced analysts choose the optimal number of LVs or the adequate data preprocessing. This has been done by converting some of the aforementioned parameters into numerical indexes and then combining them into a general score that can be used for the assessment of the performance of a model or for the comparison between models in a fast and objective way.

## 2. Material and methods

### 2.1. J-Score

For the calculation of a global score, which reflects the general quality and future performance of a regression model, more than one parameter has been considered, covering all the properties of the model that an analyst would use for its evaluation in a direct or indirect way.

#### 2.1.1. Inverse Ratio of Performance to deviation

RMSE is the most used quality indicator of a regression model, particularly applied to prediction and cross-validation sets ($RMSE_P$ and $RMSE_{CV}$); however, the magnitude of these values depends on the order of magnitude of the data; so, if models built using different datasets need to be compared, the RMSE values must be normalized. Ratio of Performance to Deviation (RPD) is a commonly used parameter to solve this problem. It is calculated by dividing the standard deviation of the reference data ($s_Y$) with the RMSE, providing values that can go from zero to infinity [10]. Instead, the inverse of RPD, as shown in Equation (1), goes from 0 to 1; 0 meaning a small error and 1 meaning an error equal to $s_Y$.

$$\frac{1}{RPD} = \frac{RMSE}{s_Y} \quad \text{(Eq. 1)}$$

#### 2.1.2. Calibration to validation error ratio

The difference between $RMSE_{Cal}$ and $RMSE_{CV}$ (or $RMSE_P$) is bigger when the model is overfitted, this is, when the model is too specific for the samples used in the calibration and cannot predict new samples correctly. Therefore, when another set of samples is projected onto the model, the model is unable to provide reliable predictions due to its specificity as, apart from explaining the variability related to the sample properties, it is also explaining unrelated noise [3]. So, one minus this ratio (Equation (2)) shows how much overfitting the regression model has; 0 being the same error in both sample sets and 1 meaning the calibration error is zero or that the cross-validation error is infinitely bigger.

$$\text{Error Ratio} = 1 - \frac{RMSE_{Cal}}{RMSE_{CV}} \quad \text{(Eq. 2)}$$

#### 2.1.3. Regression Vector Noise Index

The noise of the Regression Vector (RV) of a PLSR model can be defined as the variability that is not related to the chemical and/or physical properties of the samples that affect the spectra used to build the model. This is reflected in the RV of the model when the model is explaining the variance related to the noise instead of the information of interest. It is easy to see when working with continuous vectors like those obtained from spectra because noise causes the RV to lose its continuity. This noise is an indicator of the level of overfitting of the model and therefore, of its lack of suitability. In the present work, the quantification of this noise is performed by applying a Gaussian smoothing to the regression vector, this is, approximating each point of the vector following the normally distributed weighting of the surrounding points. Therefore, the number of points to be used for the smoothing process (Smoothing Window, SW) must be decided based on the properties of the data. In this study, as spectra of very different characteristics have been used, a new algorithm has been implemented to quantitatively decide the SW, based on the method proposed by Lin H. et al. [11].

For each dataset, an average spectrum is calculated and then the first derivative (gap derivative) is applied to the average spectrum; then, the positions of the zeros (indexes of the variables with value zero) are obtained, which are the maxima and minima of the original spectra as shown in Fig. 1. Finally, the mean distance between correlative zeros is calculated (Eq. (3)).

$$SW = \frac{\sum_{i=1}^{k} x_{i+1} - x_i}{k} \quad \text{(Eq. 3)}$$

Where **x** is the vector containing the indexes of the variables where the average of the spectra in the dataset has a maximum or minimum point. Obtained by indexing the zeros of the derived average spectrum of the dataset.

The Regression Vector Noise Index ($NI_{RV}$) is obtained by first calculating the residuals between the original vector and the smoothed one (Fig. 2), which is an idealization of how the vector should really be if there was no noise [12]. Then, all the residuals are summed up and, finally, the value is relativized by dividing by the summatory of the absolute coefficients of the original RV, as described in Equation (4).

$$NI_{RV} = \frac{\sum |RV - RV_{smoothed}|}{\sum |RV|} = \frac{\sum |\text{Noise Vector}|}{\sum |RV|} \quad \text{(Eq. 4)}$$

#### 2.1.4. Determination coefficient difference to unity

The determination coefficient between the reference values of the studied property and the values predicted using the regression model is calculated by dividing the square of the covariance between predicted (usually in cross-validation) and reference values by the product of the variances of predicted and reference values, as shown in Equation (5) [13]. This parameter also goes from 0 to 1.

$$R^2 = \frac{s_{XY}^2}{s_{XX} s_{YY}} \quad \text{(Eq. 5)}$$

In this case, as the optimal value of the index is 1 instead of 0, the

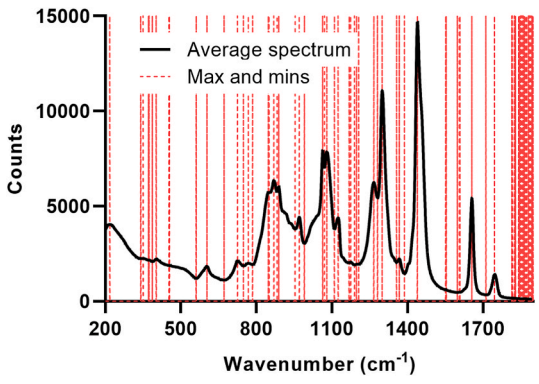



**Fig. 1.** Example of the average spectrum of a dataset and its maximum and minimum points.

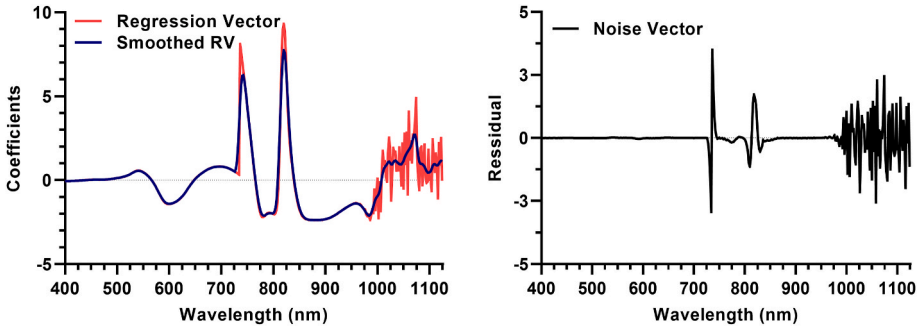



**Fig. 2.** a) An example of a PLSR model's regression vector and the same vector after applying Gaussian smoothing. b) The difference between the original RV and the smoothed RV.

difference of the determination coefficient to unity (1-$R^2$) is used as an index to show how good the model predictions are by comparing them with the reference values of the predicted parameter, values that are considered to be the real ones.

#### 2.1.5. Residual variance

Residual Variance (ResVar) is the ratio of the variance in **Y** that is not captured by the model, this is, the variance that the LVs used to build the PLSR model do not explain, relative to the variance of the calibration data. When this index is close to 0 means that the model is explaining almost the whole variance of the reference data (**Y** matrix). The residual variance does not account for overfitting, which means that the model could also be including the noise of the calibration data. This implies that samples that were not used in the calibration process may not be predicted correctly.

#### 2.1.6. J-Score formulation

The five indexes obtained can be joined by averaging and, in this way a global score that goes from 0 to 1 is obtained, Joint-Score or J-Score (described in Equation (6)), where the value indicates the quality of the model: the lower the value the better the model. For choosing the optimal number of Latent Variables, a curve of the J-Score for the model with different LVs can be calculated, where the minimum of the curve shows the ideal number of LVs to be used in the regression model.

$$J-Score=\left(\frac{RMSE_{CV}}{s_Y}+\left(1-\frac{RMSE_{Cal}}{RMSE_{CV}}\right)+NI_{RV}+\left(1-0R^2_{CV}\right)+ResVar_Y\right)\Big/5 \quad \text{(Eq. 6)}$$

The time the algorithm takes to calculate the J-Score for only one model is insignificant, but if the J-Score is to be implemented in any kind of algorithm for automatic evaluation of PLSR models, like comparing many models with different data preprocessing and number of LVs, the computation time must be optimized, even if Equation (6) could be used with the mentioned purpose (after studying its behavior). Although, when studying the weights of the individual indices of the original equation (Equation (6)), as explained in the results section (3.1), it was observed that getting rid of the determination coefficient and the residual variance does not change the shape of the J-Score curve and reduces the computation time significantly. So, an optimized J-Score has been proposed, calculated as,

$$\mathrm{J-Score}=\left(\frac{\mathrm{RMSE_{CV}}}{\mathrm{s_Y}}+1-\frac{\mathrm{RMSE_{Cal}}}{\mathrm{RMSE_{CV}}}+\mathrm{NI_{RV}}\right)\Big/3 \quad \text{(Eq. 7)}$$

where the optimal PLSR model is defined as the model with the lowest J-Score.

In the present work, the $RMSE_{CV}$ has been calculated using 10-fold random subsets CV as it is one of the most widely used and robust random CVs and the one used by original authors in most of the models of the example datasets [14]. An important point to emphasize is that this term can be substituted in Equation (7) by the RMSE obtained using any other way of cross validation if the arrangement of the data requires it or even by the $RMSE_P$ if there is an adequate test set for validation.

### 2.2. Datasets

For studying the behavior of the J-Score in different scenarios, four spectroscopic datasets were selected, three of them published in articles that use and describe PLSR models for them, and another one built from the spectra published in another article. To get the highest possible variability, studies that use different spectroscopic techniques and different sample types were chosen, in order to assess the universality and applicability of the method.

#### 2.2.1. Dataset 1 – UV–Vis spectra of nucleotides

Four datasets with different theoretical LVs were simulated from the Ultraviolet–Visible (UV–Vis) spectra of four pure nucleotides: adenylic acid, cytidylic acid, guanylic acid and uridylic acid [15]. The first dataset is formed by generating random concentrations of adenylic acid, then multiplying them by its pure spectrum and finally adding 5% of random noise. The second one is a mixture of random concentrations of adenylic and cytidylic acids multiplied by their pure spectra and 5% of noise, and so on until all of them are included. In this way, four datasets of 100 samples with theoretically 1, 2, 3 and 4 LVs were generated. Only mean centering (MC) was used as preprocessing.

#### 2.2.2. Dataset 2 – Raman spectra of meat

The samples of this dataset were 16 pork carcasses taken from the daily production stock of a slaughterhouse with the goal of predicting their fat content by Raman spectroscopy. In total, 134 samples were acquired: each one was divided in two, one half was analyzed by Raman spectroscopy and the other half was used for reference analysis of Iodine Value (IV), which is a parameter used to express fat content [16]. IVs were calculated by applying the American Oil Chemists' Society recommended practice 1c-8522 'Calculated Iodine Value', with the extensions by Pettersson [17] which is based on the individual concentrations of fatty acids determined by gas chromatography. The spectral data were preprocessed using Standard Normal Variate (SNV) and MC.

#### 2.2.3. Dataset 3 – MIR spectra of wine fermentation

This dataset was obtained by measuring the middle-infrared spectra of a wine fermentation process with the goal of monitoring the process. In total, 36 microvinifications were monitored: 14 in normal fermentation conditions and 22 intentionally contaminated fermentations with different concentrations of lactic acid bacteria. Attenuated Total Reflectance-Middle Infrared (ATR-MIR) measurements were collected during alcoholic and malolactic fermentations and density and pH were daily measured using a portable densimeter and a pH-meter, respectively, for their prediction using PLSR. The spectra were preprocessed by using Savitzky-Golay (S-G) 2$^{nd}$ order polynomial smoothing with a 15-

point-window, SNV normalization and MC. For pH prediction two spectral regions were selected and joined [18].

#### 2.2.4. Dataset 4 – NIR spectra of forages

The dataset consists of Near Infrared (NIR) reflectance spectra of 305 forage samples recorded under the same conditions by a specialized laboratory for predicting their humidity. Moisture was measured by gravimetry, drying the samples in the oven at 103–105 °C for 4 h. The dataset was delivered to six participants, all of whom had previous knowledge and experience on multivariate calibration, to determine how different the results from several laboratories were when they used their preferred multivariate calibration method and software [19]. The different preprocessings used in the original paper were replicated.

## 3. Results and discussion

### 3.1. J-Score optimization

The effect of each index in the J-Score was separately studied as shown in Fig. 3. One of the first facts observed is that the inverse RPD, 1-$R^2$ and Residual Variance values (Fig. 3a, d and 3e) are directly proportional, as it can be seen in Fig. 3f, as they describe similar properties of the PLSR model. Instead, the inverse RPD, the Error Ratio and the Noise Index are considerably independent between them as they describe other properties of the model. Even more, the calculation of the determination coefficient between predicted and reference values and the residual variance of the **Y** block consumes around 90% of the computation time of a model's J-Score due to the properties of the PLS algorithm used (plsregress from Machine Learning Toolbox in Matlab). In addition, to study the relevance of the indices considered in Equation (6), Principal Component Analyses (PCA) were performed on the indices obtained for the PLSR models of several example datasets. Matrices were built with the five indices in columns and number of LVs in rows for each model. From this it was concluded that as 2 PCs explained around 98% of the variance and they included information from all three indices considered in Equation (7) (as it can be seen for an example dataset in Fig. S6 a and b), these PCs could be used to join the indices by summing up their scores. The result, shown in Fig. S7, offered curves similar to those of the RMSE, with no additional information so it was discarded as a more robust way of joining the indices.

The effect of removing the Regression Vector Noise Index from Equation (7) was also studied, as it is the index that makes the J-Score only applicable to continuous data and has a similar tendency to the Calibration to Validation error ratio. Although in some cases the

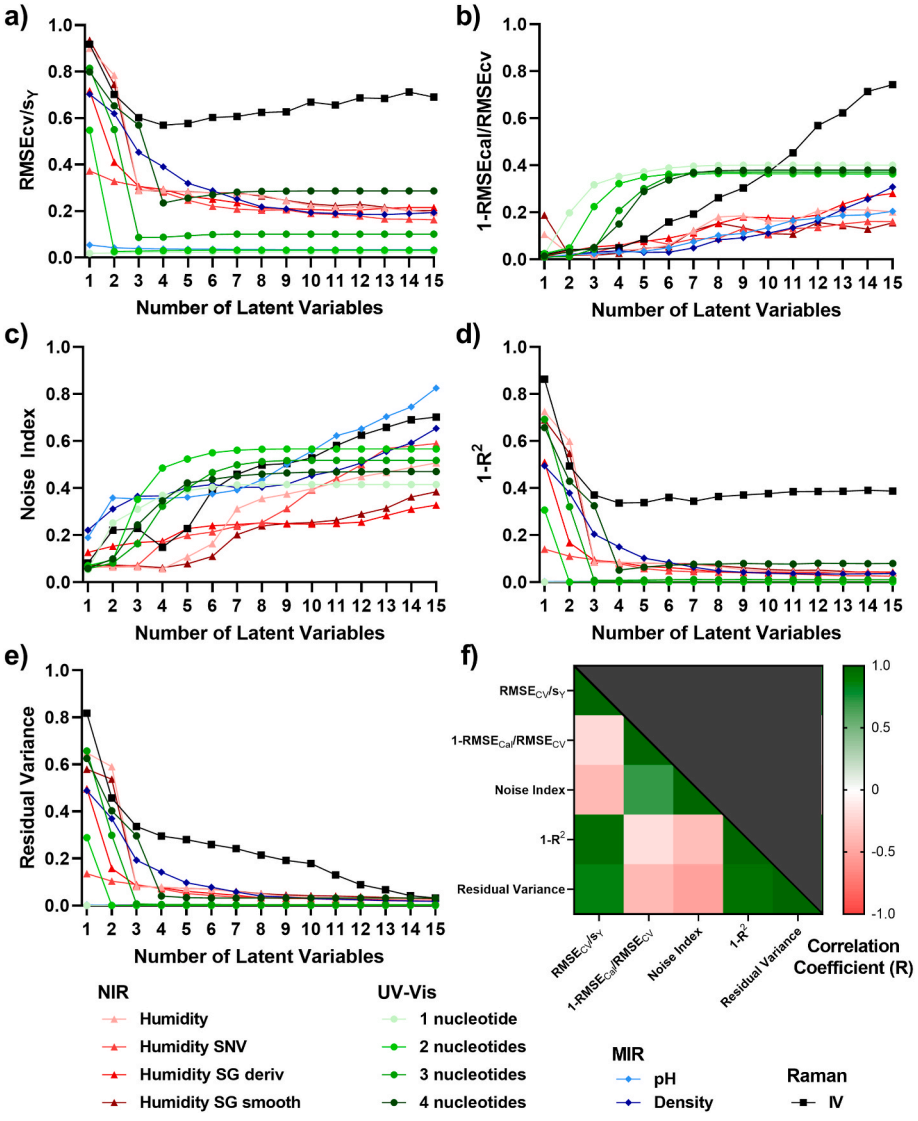



**Fig. 3.** The curves for each individual factor considered in the J-Score for all the models studied in this paper. a) Inverse RPD b) Calibration to Validation Error Ratio c) RV Noise Index d) Determination coefficient difference to unity e) Residual Variance f) Correlation coefficients between the terms.

obtained curves have the same minimum point (UV–Vis and Raman datasets), results show that including this term in the selection of the optimal dimensionality turns out in more parsimonious models when the minimum point is not so clear (Fig. S6 c, as an example). Therefore, the formulation of a score to be used for non-continuous data should be studied in more detail in a future work.

Considering this, it was decided that using the optimized formula for the J-Score (Equation (7)), which does not include the last two terms but does include the RV Noise Index, would be more appropriate because it is faster and offers more robust results, since the choice of the optimal number of LVs is the same for Equation (6) and Equation (7) in all cases (as shown in Fig. S8).

In Fig. 3a–e it can also be observed that all the considered indices (Fig. 3a, b and 3c) not only can theoretically go from 0 to 1 but in practical cases they cover that range and none of the terms has a bigger weight in the J-Score. This being a fact, a same weight average is proposed in Equation (7), as it does not bring in a bias due to the analyst's experience and it is the simplest way to join the indices. It could be possible to adjust the weights of the terms being analyzed, such as increasing the weight of the inverse RPD due to the frequent use of the $RMSE_{CV}$ and its correlation with other factors. However, this approach may introduce an 'expertise' bias and the introduced weights would have to be studied as they could change the obtained conclusions.

### 3.2. Performance of the J-Score in the example datasets

#### 3.2.1. Example 1 – UV–Vis spectra of nucleotides (dataset 1)

In order to study the adequacy of the J-Score to select the best number of LVs, PLSR models were fitted for the datasets simulated using different mixtures of pure spectra of four nucleotides: with one, two, three and four components plus a noise component in all cases. This allows comparing the number of LVs selected by the J-Score with a reference value that has not been selected by an analyst; therefore, a value that is not biased nor subjective. As these cases are theoretical it is easy to see where the slope of the RMSE curves make a clear change, indicating the number of LVs that should be selected (Fig. 4). This is also the point where the $RMSE_{Cal}$ and $RMSE_{CV}$ curves start to diverge. The J-Score curves show a clear minimum where the RMSE curves have a change in slope, making it easier to objectively decide the number of LVs to be selected for a PLSR model.

These models also show that the way of calculating the J-Score, and particularly the Noise Index, is adequate for UV–Vis spectroscopy, where the absorption bands are usually wide and smooth. In fact, the conclusions obtained from the J-Score curves coincide with the theoretical number of LVs to use and are consistent with those obtained from the RMSE curves. Although these are simple scenarios where the number of LVs to be selected is easy to decide, the use of the J-Score has shown to be more robust and objective.

#### 3.2.2. Example 2 – Raman spectra of meat (dataset 2)

In this section, the performance of the J-Score when dealing with Raman spectra is studied. A specific characteristic of this type of spectroscopy is that the collected spectra show sharper and closer peaks than in other spectrophotometric techniques. The Noise Index is based on a smoothing algorithm, explained in section 2.1.3, and adjusting the SW based on the resolution of the spectra makes the smoothed RV adequate to subtract the noise even in sharp spectra like Raman, as it can be seen in Fig. 5.

Once the SW for the Noise Index is decided, each one of the indexes that make up the J-Score can be calculated. As shown in Fig. 6a, the cross-validation RMSE usually has a downward trend that can either stabilize around an asymptote, have a small valley and rise again, or keep going down slowly. In either case it is not obvious the number of LVs that should be selected for the PLSR model, so some sort of criteria such as a minimum relative change on the RMSE when adding another LV is usually stablished to decide it [20,21]. Secondly, the ratio between $RMSE_{Cal}$ and $RMSE_{CV}$ is calculated. Fig. 6b shows a growing curve with the prediction error difference between calibration and validation sample predictions, which indicates model overfitting. However, this

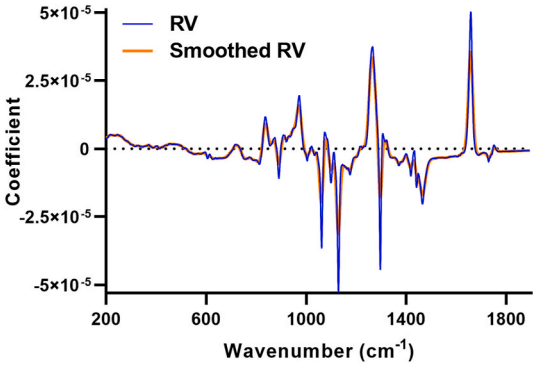



**Fig. 5.** Regression vector and smoothed regression vector of the PLSR model (three LVs) for prediction of Iodine Value in meat using Raman spectroscopy.

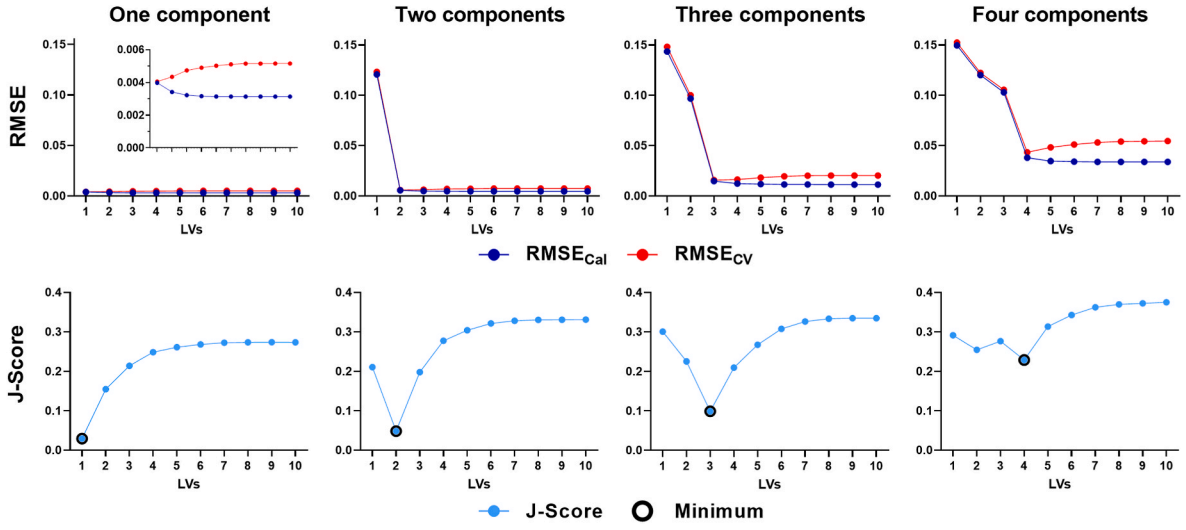



**Fig. 4.** Upper row shows the $RMSE_{Cal}$ and $RMSE_{CV}$ curves of the PLSR models built for the datasets with 1, 2, 3 and 4 theoretical components, respectively. The lower row shows the J-Score curves for the same models.

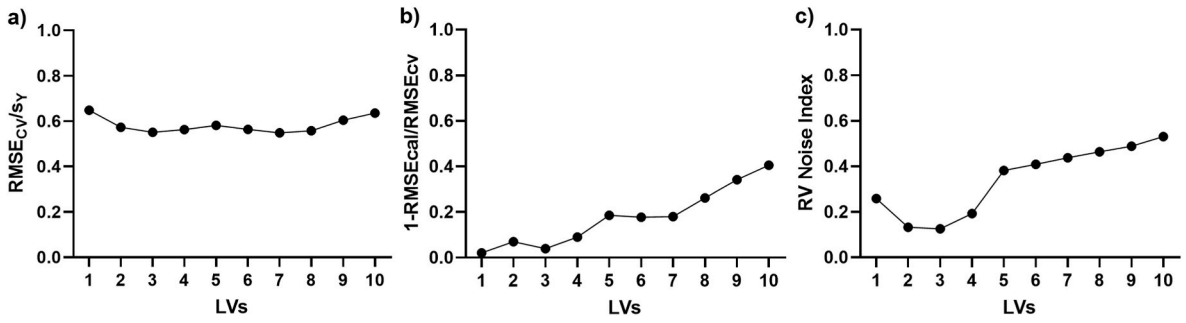



**Fig. 6.** Curves of each of the indices of the J-Score for PLSR models with different number of LVs: a) Normalized RMSE (inverse RPD) of cross-validation. b) Ratio of calibration RMSE and cross-validation RMSE. c) Noise index of regression vector.

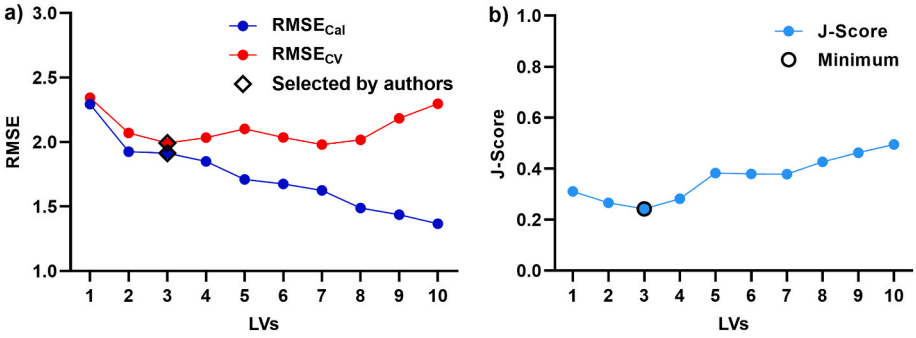



**Fig. 7.** a) $RMSE_{Cal}$ and $RMSE_{CV}$ curves of the PLSR models for Iodine Value prediction, as used in the reference paper to decide the number of LVs (apart from $RMSE_P$) [16]. b) J-Score curve for the same models.

curve does not always have a critical change of slope nor a valley, so it is difficult to decide the optimal number of LVs using this parameter. Thirdly, the NI curve is calculated using the procedure explained in section 2.1.3. (Fig. 6c). This index is the one that best shows when a model is overfitted, as the RV includes a notable amount of noise when an unnecessary LV is added, so the curve shows an evident change of slope. Finally, as explained in section 2.1.6, when the three indices are joint the J-Score is obtained.

In the example at hand, the original authors selected the number of LVs based on the curves of $RMSE_{Cal}$, $RMSE_{CV}$ (shown in Fig. 7a) and

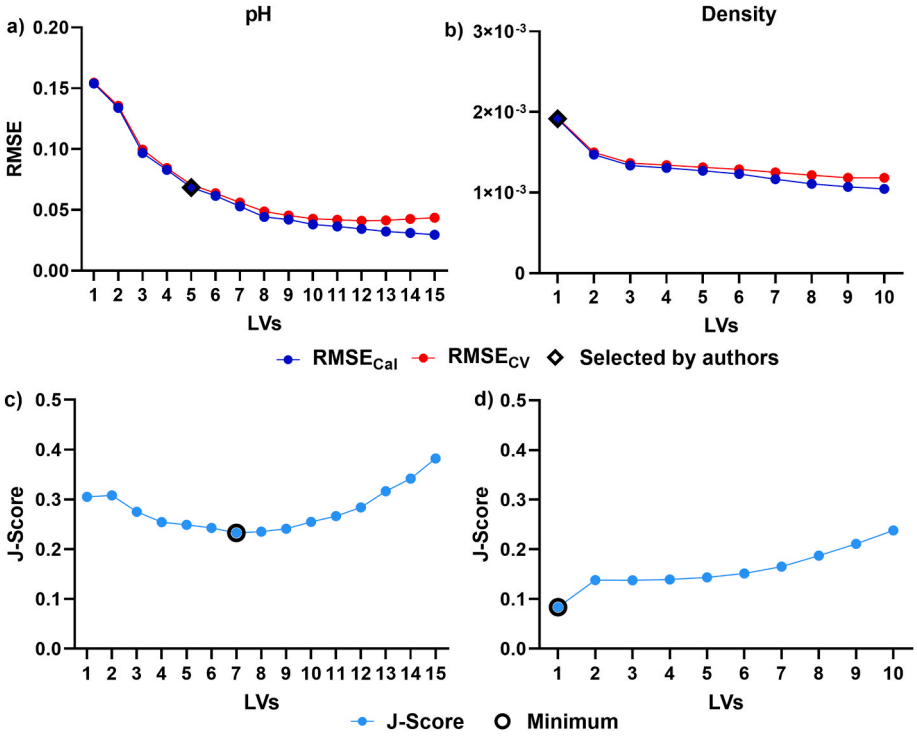



**Fig. 8.** a) and b) $RMSE_{Cal}$ and $RMSE_{CV}$ curves of the PLSR models for prediction of pH and density in wine with MIR spectra, respectively. c) and d) J-Scores of the same models.

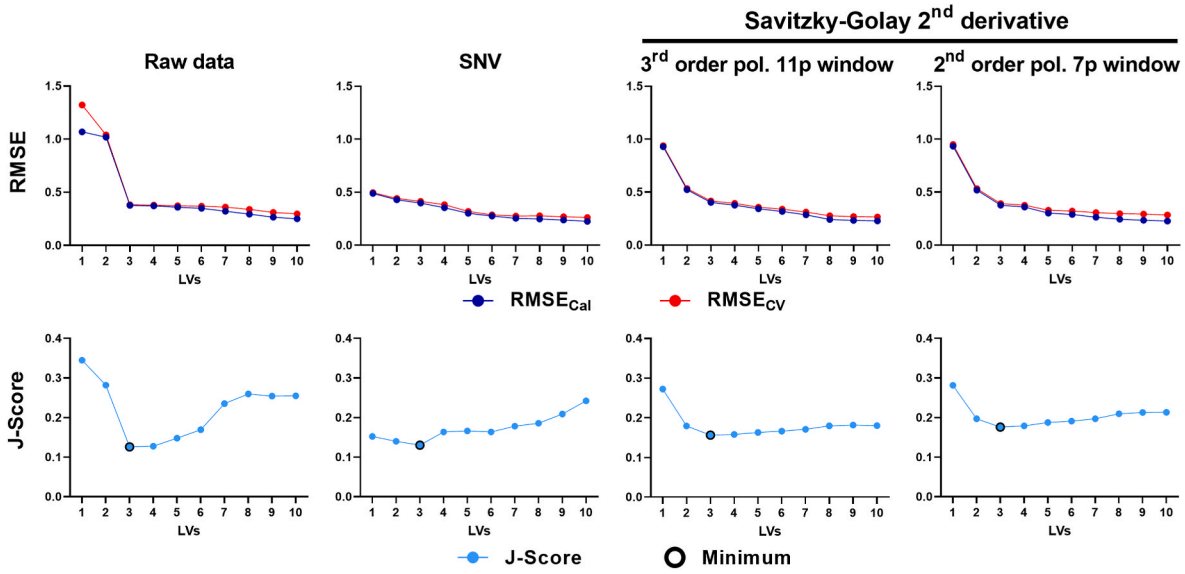



**Fig. 9.** Upper row shows the $RMSE_{Cal}$ and $RMSE_{CV}$ curves of the moisture PLSR models with different data preprocessings used in Ruisánchez et al. The lower row shows the J-Score curves for the same models.

$RMSE_P$ [16]. These curves show two valleys with no clear optimal point, so a criterion based on a change of slope and overall performance of the models was used by the authors to select three LVs. However, using the J-Score curve (Fig. 7b) a minimum point is shown in the model with three LVs, which is consistent with the results of the original study.

### 3.2.3. *Example 3 – MIR spectra of wine fermentation (dataset 3)*

In this section, examples of PLSR models where the number of LVs is hard to decide are studied. In the original study, the prediction models of the pH value of wine samples using MIR spectroscopy show a smooth RMSE curve with no critical change of slope nor a minimum (Fig. 8a). In this case, the authors decided to use five LVs based on the RMSE curve together with the explained variance [18]. On the contrary, the J-Score curve for the same models (Fig. 8c) shows a minimum point at seven LVs. Although the performance of the two models is similar, the decision based on the J-Score may differ from the initial one. This highlights the difficulty for the analyst when evaluating the models and interpreting the descriptive statistics with the aim of avoiding overfitting.

When predicting the density of the samples using the same spectra, the original authors decided that only one LV was needed, but as it can be seen in Fig. 8b that decision cannot be made based on the RMSE curve but based on the already low RMSE value of the model with one LV. Instead, in the J-Score curve shown in Fig. 8d, the model with one LV is a clear minimum showing the optimal model, agreeing with the expert eye of the authors.

### 3.2.4. *Example 4 – NIR spectra of forages (dataset 4)*

In this section, the example dataset is used to show that the J-Score can be used not only for comparing models with different number of LVs, but for comparing models built using data with different preprocessing (just MC; SNV and MC or S-G and MC) and both variations at the same time, in order to select the optimal model between the considered ones. As it can be seen in Fig. 9, the different data preprocessings change the shape of both RMSE curves in a considerable way, making harder or easier to choose the adequate number of LVs in each case. On the contrary, the J-Score curves show a minimum in all cases. In all four preprocessing methods, three LVs were selected as optimal, showing that in this case the J-Score is consistent in deciding the dimensionality of the

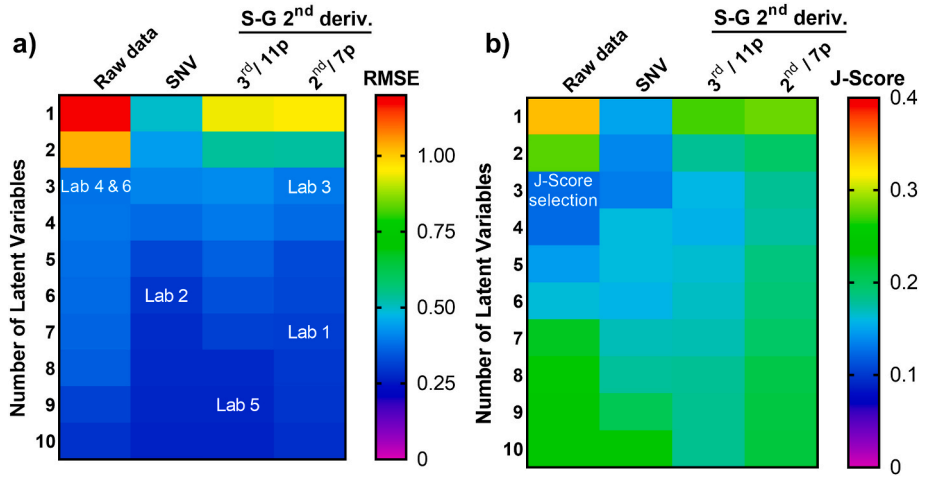



**Fig. 10.** a) RMSECV values of PLSR models for moisture in forage using different spectral preprocessing (for Savitzky-Golay, the polynomial order and the number of points of the moving window are stated) and different number of LVs. Labels showing the model selected by each laboratory in Ruisánchez et al. b) J-Scores of the same models and a label showing the optimal one.

moisture information contained in the spectra, regardless of the preprocessing, even if this cannot be generalized and does not have to happen in every case.

When both, data preprocessing and number of LVs, are compared together, the heat maps shown in Fig. 10 are obtained, which offer a global view of all the models considered. In the case of $RMSE_{CV}$ (Fig. 10a) it is hard to decide the optimal point in the map. In fact, all the experts (represented as laboratories in Fig. 10) selected different models except two of them that agreed on that the model with no preprocessing and three LVs was the optimal one. On the other side, the same map built using the J-Score (Fig. 10b) clearly shows a valley with three-four models where the minimum is also the model with no preprocessing and three LVs, supporting the choice of two out of six laboratories.

Furthermore, in Fig. 10b it can be seen that the models where the spectra were preprocessed with derivatives have a higher J-Score than the other ones. This happens because, although they have a similar $RMSE_{CV}$, the regression vectors of the models with a derivative are noisier. The derivative amplifies the noise of the spectra as much as the information of interest, making the models less robust. In addition, the rough derivative is noisier than the S-G derivative, as the last one smooths the data. Therefore, the J-Score quantifies the fact that when two models have the same prediction error but one of them has been more preprocessed it will be less robust, so the model with less preprocessing should be selected. A fact that cannot be seen using only $RMSE_{CV}$ in Fig. 10a.

The different laboratories selecting very different models illustrates the fact that each analyst has their own model evaluating system with different weights for the different possible evaluation parameters. This is based on knowledge and experience in the field, and the lack of it may end up in choosing a suboptimal model. Instead, the J-Score offers an objetive and global way of comparing the obtained models, which concordates with the criteria of some of the experts.

## 4. Conclusions

A novel indicator for choosing the optimal number of latent variables of a PLS-R models has been proposed. This parameter has proven to be objective, as it is based on the combination of numerical properties of the model; robust, because it has a good reproducibility, meaning that it is stable when calculated repeatedly; universal, as it can be applied to many data sources that offer continuous vectors; and fast, as it can be used for the evaluation of many models in a short period of time in an automated way. The J-Score is an evaluation tool that can be used by non-experienced analysts for choosing the optimal number of Latent Variables, as well as for choosing the best preprocessing techniques to use in a fast and automated way. Furthermore, it has shown to be adequate for all types of spectroscopic data tried (UV–Vis, NIR, MIR and Raman).

## Funding

Grant PID2019-104269RR-C33 funded by MCI/AEI/10.13039/501100011033. Grant URV Martí i Franqués – Banco Santander (2021PMF-BS-12). This publication has been possible with the support of the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (2020 FISDU 00221; Schorn-Garcia, D.).

## CRediT authorship contribution statement

**Jokin Ezenarro:** Methodology, Software, Formal analysis, Investigation, Writing – Original Draft; **Daniel Schorn-García:** Validation, Resources, Writing – Original Draft; **Laura Aceña:** Project administration, Writing – Review & Editing; **Montserrat Mestres:** Writing – Review & Editing; **Olga Busto:** Conceptualization, Writing – Review & Editing; **Ricard Boqué:** Investigation, Conceptualization, Writing – Review & Editing, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

The authors would like to thank Zscheile et al.; Lyndgaard et al.; Cavaglia et al. and Ruisánchez et al. for making their data sets available for this research. The MATLAB version of the J-Score calculation function is available upon request to isens@urv.cat.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2023.104883.

## References

[1] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 186 (1986).
[2] Y. Ozaki, Near-infrared spectroscopy—its versatility in analytical chemistry, Anal. Sci. 28 (2012) 545–563, https://doi.org/10.2116/ANALSCI.28.545, 6. 28 (2012).
[3] N.M. Faber, R. Rajkó, How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative, Anal. Chim. Acta 595 (2007) 98–106, https://doi.org/10.1016/j.aca.2007.05.030.
[4] R.D. Clark, Boosted leave-many-out cross-validation: the effect of training and test set diversity on PLS statistics, J. Comput. Aided Mol. Des. 17 (2003) 265–275.
[5] M. Stone, Cross-validatory choice and assessment of statistical predictions, J. Roy. Stat. Soc. B 36 (1974) 111–133, https://doi.org/10.1111/J.2517-6161.1974.TB00994.X.
[6] W.W. Chin, How to write up and report PLS Analyses, in: Handbook of Partial Least Squares, Springer, Berlin, Heidelberg, 2010, pp. 655–690, https://doi.org/10.1007/978-3-540-32827-8_29.
[7] P. Williams, K. Norris, Variables affecting near-infrared reflectance spectroscopy analysis, in: P. Williams, K. Norris (Eds.), Near-Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists Inc., Saint Paul, MN, 2002, pp. 143–167.
[8] M.W. Liemohn, A.D. Shane, A.R. Azari, A.K. Petersen, B.M. Swiger, A. Mukhopadhyay, RMSE is not enough: guidelines to robust data-model comparisons for magnetospheric physics, J. Atmos. Sol. Terr. Phys. 218 (2021), 105624, https://doi.org/10.1016/J.JASTP.2021.105624.
[9] N. Zhao, Z.S. Wu, Q. Zhang, X.Y. Shi, Q. Ma, Y.J. Qiao, Optimization of parameter selection for partial least squares model development, Sci. Rep. 5 (2015) 11647, https://doi.org/10.1038/srep11647.
[10] V. Bellon-Maurel, E. Fernandez-Ahumada, B. Palagos, J.M. Roger, A. McBratney, Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy, TrAC, Trends Anal. Chem. 29 (2010) 1073–1081, https://doi.org/10.1016/j.trac.2010.05.006.
[11] H.C. Lin, L.L. Wang, S.N. Yang, Automatic determination of the spread parameter in Gaussian smoothing, Pattern Recogn. Lett. 17 (1996) 1247–1252, https://doi.org/10.1016/0167-8655(96)00082-7.
[12] C. Liu, S.X. Yang, X. Li, L. Xu, L. Deng. Noise Level Penalizing Robust Gaussian Process Regression for NIR Spectroscopy Quantitative Analysis, 201, Chemometrics and Intelligent Laboratory Systems, 2020, https://doi.org/10.1016/j.chemolab.2020.104014.
[13] A.G. Asuero, A. Sayago, A.G. González, The correlation coefficient: an overview, Crit. Rev. Anal. Chem. 36 (2006) 41–59, https://doi.org/10.1080/10408340500526766.
[14] R. Bro, K. Kjeldahl, A.K. Smilde, H.A.L. Kiers, Cross-validation of component models: a critical look at current methods, Anal. Bioanal. Chem. 390 (2008) 1241–1251, https://doi.org/10.1007/s00216-007-1790-1.
[15] F.P. Zscheile, H.C. Murray, G.A. Baker, R.G. Peddicord, Instability of linear systems derived from spectrophotometric analysis of multicomponent systems, Anal. Chem. 34 (1962) 1776–1780, https://doi.org/10.1021/ac60193a036.
[16] L.B. Lyndgaard, K.M. Sørensen, F. van den Berg, S.B. Engelsen, Depth profiling of porcine adipose tissue by Raman spectroscopy, J. Raman Spectrosc. 43 (2012) 482–489, https://doi.org/10.1002/jrs.3067.
[17] S. Pétursson, Clarification and expansion of formulas in AOCS recommended practice Cd 1c-85 for the calculation of iodine value from FA composition, J. Am. Oil Chem. Soc. 79 (2002) 737–738, https://doi.org/10.1007/s11746-002-0551-1.

[18] J. Cavaglia, D. Schorn-García, B. Giussani, J. Ferré, O. Busto, L. Aceña, M. Mestres, R. Boqué, ATR-MIR spectroscopy and multivariate analysis in alcoholic fermentation monitoring and lactic acid bacteria spoilage detection, Food Control 109 (2020) 106947, https://doi.org/10.1016/j.foodcont.2019.106947.

[19] I. Ruisánchez, F.X. Rius, S. Maspoch, J. Coello, T. Azzouz, R. Tauler, L. Sarabia, M. C. Ortiz, J.A. Fernández, D. Massart, A. Puigdomènech F, C. García, Preliminary results of an interlaboratory study of chemometric software and methods on NIR data. Predicting the content of crude protein and water in forages, Chemometr. Intell. Lab. Syst. 63 (2002) 93–105, https://doi.org/10.1016/S0169-7439(02)00039-4.

[20] D.W. Osten, Selection of optimal rregression models via cross-validation, J. Chemom. 2 (1988) 39–48, https://doi.org/10.1002/cem.1180020106.

[21] D.M. Haaland, E.v. Thomas, Partial least-squares methods for spectral Analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information, Anal. Chem. 60 (1988) 1193–1202.