Research article

Journal of Human, Health and Halal Metrics; 2023, 4(2): 25-37
https://doi.org/10.30502/JHHHM.2023.424076.1077

Practical approaches in detection of food fraud with multivariate calibration

Somaye Vali Zade*

- *Halal Research Center of IRI, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran.*

This paper is open access under Creative Commons Attribution-NonCommercial 4.0 International license.

Submission: 6 November 2023 ***Revision***: 2 December 2023 ***Acceptance***: 3 December 2023

Abstract

Background and objective: In the realm of analytical chemistry, multivariate calibration involves creating mathematical models that connect diverse instrumental signals with analyte concentrations. This approach provides a mean to quantitatively analyze complex mixtures, particularly in multicomponent systems. To address food adulteration concerns, this paper explores the application of Raman spectroscopy and Partial Least Squares Regression (PLSR) using the MVC1 software. The main objective is to demonstrate the software's efficiency in quantifying the adulteration of hazelnut oil in extra virgin olive oil (EVOO).

Materials and methods: The analysis leverages the MVC1 software, a valuable tool for multivariate linear and nonlinear calibrations. One-leave-out cross-validation and the Durbin-Watson statistical test are employed to determine the optimal number of PLS factors and identify outliers. Statistical parameters including RMSEP, %REP, R^2, and explained variance are used to evaluate the calibration model's performance. Key figures of merit including sensitivity, analytical sensitivity, LOD, and LOQ, are computed to assess the analytical technique's precision and reliability.

Results and conclusion: The study effectively quantifies the percentage of adulteration in EVOO by hazelnut oil, a pressing concern in food authenticity and safety. The results demonstrate the MVC1 software's capability in establishing reliable calibration models. By achieving a balance between sensitivity and analytical sensitivity, the model accurately predicts analyte concentrations. It also sets robust detection and quantitation limits, ensuring precise analysis. This research showcases the practical application of advanced analytical techniques and software tools to address real-world problems, contributing to the authenticity and purity of food products in the market.

Keywords: Adulteration, Extra Virgin Olive Oil, Multivariate Calibration, MVC1, Partial Least Squares

1. Introduction

In the realm of analytical chemistry, multivariate calibration entails the development of mathematical models that establish connections between unselective multiple instrumental signals and analyte concentrations [1,2]. In contrast to univariate calibration, which focuses on single signals, the utilization of multivariate signals allows for the compensation of varying contributions from non-analytes within an unknown sample. This capability enables quantitative analysis even within inherently unselective multicomponent systems [3]. The success of multivariate calibration is evidenced by its extensive application, particularly in near-infrared (NIR) spectroscopy, dating back to the 1970s [4]. Over the last few decades, numerous other instrumental signals,

* Correspondence to: Somaye Vali Zade; Email: smy.valizade@gmail.com; Tel.: +98-21-88909033

including spectroscopic, electrochemical, chromatographic, and more, have been integrated into this dynamic realm of multivariate calibration.

The instrumentation required for the measurement of first-order data is quite straightforward; most spectroscopic, chromatographic, and voltametric equipment can readily provide this vectorial information. In contrast, the measurement of second- and third-order data can be achieved either through a single instrument or by combining multiple instruments. Advancements in analytical instrumentation have led to the proliferation of multiply hyphenated techniques, often referred to as hyperhyphenation or hypernation [5]. These techniques yield data of escalating complexity, presenting challenges from both experimental and theoretical perspectives.

Raman spectroscopy has emerged as a powerful analytical tool in the field of food science and technology. This non-destructive technique allows researchers to gain valuable insights into the composition, quality, and safety of food products. It is particularly effective in identifying and quantifying various compounds, including nutrients, contaminants, and additives, contributing to food authenticity and safety assessment. Raman spectroscopy's ability to provide rapid and precise data makes it invaluable in monitoring processes such as food production, storage, and quality control. Its impact on food science is evident in applications ranging from assessing food adulteration to understanding molecular changes during cooking and processing, ultimately ensuring the production of safe and high-quality food products.

First-order calibrations are powerful tools for analyzing mixtures, even in complex samples. In many commercial laboratory instruments, these methods are integrated into the device's software as standard procedures. Moreover, user-friendly software and computer coding for multivariate analysis are readily available. Despite these advantages, there is still a strong need for developing first-order multivariate calibration methods and gaining familiarity with the existing methods and software. Researchers can effectively address many quantitative analysis challenges using first-order calibration methods as their optimal solution.

To gain a deeper understanding of how to implement a first-order calibration method, this article aims to provide a practical insight into solving a problem related to food and adulteration using multivariate calibration. Efforts have been made to equip researchers in this field with a powerful and accessible software tool, enhancing their skills. Determining the percentage of adulteration in virgin olive oil serves as a chosen example, and the steps involved in a multivariate quantitative analysis, along with the performance metrics of the results, are thoroughly examined and discussed.

2. Materials and methods

2.1. Materials and sample preparation

This study analyzed 33 synthetic mixture samples of pure Extra Virgin Olive Oil and hazelnut oil. The edible oils were purchased from various suppliers. Pure extra virgin olive oil samples were blended with hazelnut oil adulterant to produce adulterated samples. The percentage of adulterations were from 3 to 90 percent of hazelnut oil.

2.2. Raman measurements

The spectral gaining was conducted using Raman spectroscopy (Raman microscope model TDLG100, Teksan, Tehran, Iran). Each measurement was performed in the spectral range of 4200 to 200 cm-1. This device is equipped with a 785 nm YAG laser. Laser power, integration time, and average number were selected as 300 milliwatts, 3 s, and 5, respectively, to get optimum Raman peak intensity. The scans are done by placing the sample plate in the sample holder, where the light is focused onto the sample, and their average spectrums are measured after three scans. All the measurements were conducted at room temperature, and the collection time was 15 s. All the samples were scanned under same condition.

3. Results and discussion

3.1. PLS in action: Case studies illustrating the power of multivariate calibration

Calibration is the process of establishing a relationship between measured variables and one or more desired properties. Instrumental responses, such as spectra, typically consist of signals related to the concentration of analytes, and during the calibration process, a model is developed to determine their relationship with the concentration of the analyte(s). Univariable calibrations are usually not suitable for the analysis of complex mixtures due to the need for selectivity in the measured signal.

Multivariable calibration models are capable of establishing the relationship between measured signals, such as spectra in a region where there is even significant overlap between the responses of components, and the concentration of analyte(s). In this way, in the machine learning (ML) process, this multivariable relationship is discovered and used for predicting concentrations in unknown samples. Machine learning processes are performed using different computational algorithms. Here, we discuss the widely used machine learning algorithm known as "Partial Least Squares Regression" (PLSR).

Partial Least Squares Regression (PLSR) is a powerful machine learning tool used in multivariate calibration. It helps establish a relationship between complex spectroscopic data and the concentration of analytes. PLSR finds patterns in the data by capturing the most significant features in both the spectra and the concentration values. It does this by creating a set of latent variables, or factors, that best explain the variation in both datasets. PLSR is particularly valuable when dealing with highly overlapping signals, making it a reliable method for predicting analyte concentrations from intricate spectral information.

With the availability of reputable computational software and packages in the field of multivariate calibrations, machine learning users need not be experts in the development of computational and chemometric methods. Typically, software is designed in a manner that allows non-professional users to benefit from them in solving their problems. Nevertheless, an understanding of the principles and familiarity with performance metrics (figure of merits, FOM) and statistics related to analytical results are of great importance in assessing the effectiveness and validity of the outcomes. The free accessible software MVC1 is a highly valuable toolbox for multivariate linear and nonlinear calibrations, developed by the group of Professor Alejandro Olivieri [6,7]. This software provides unique reporting of performance metrics employed by the method and statistics related to result validity.

This research aims to provide a practical application of the statistics offered by the MVC1 software for a deeper understanding. The practical utility of this software is demonstrated in solving a real-world problem: the quantification of the percentage of adulteration of hazelnut oil in extra virgin olive oil using Raman spectroscopy and multivariate calibration method Partial Least Squares Regression (PLSR). The choice of this problem is twofold, considering both the significance of detecting adulteration in food products and as an illustrative case for an in-depth and descriptive examination of the application of the PLSR machine learning tool within the computational package MVC1. This study seeks to showcase how the practical functionalities of the software can be harnessed effectively. The main MVC1 screen has been shown in Figure 1, that is the interface between the users and the coded algorithm to do first order multivariate calibration.

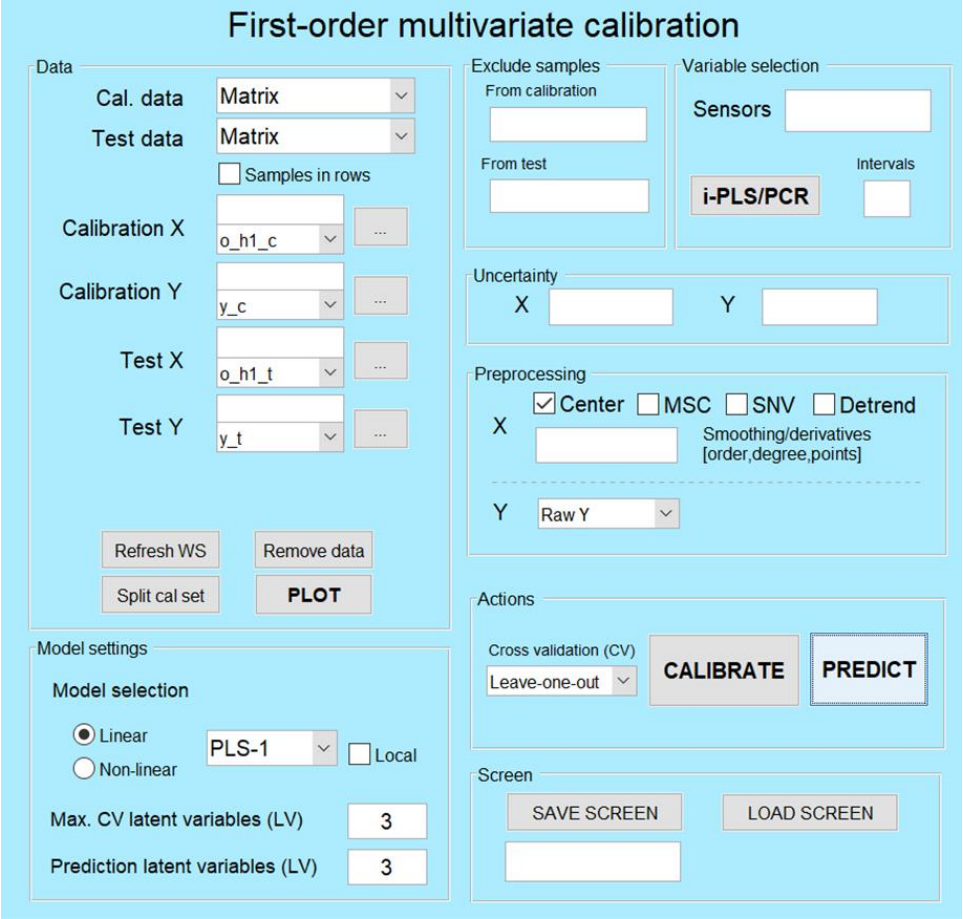

Figure 1- Main MVC1 screen for first order multivariate calibration

3.2. Quantifying hazelnut oil adulteration in extra virgin olive oil with Raman spectroscopy and machine learning

Determining the percentage of adulteration of hazelnut oil in extra virgin olive oil through Raman spectroscopy and machine learning methods holds significant importance in the realm of food authenticity and safety. The adulteration of extra virgin olive oil with lower-cost alternatives is a prevalent issue in the food industry. Detecting such adulteration is vital not only for ensuring the integrity and quality of olive oil but also for safeguarding consumer health. Raman spectroscopy, with its capability to provide molecular fingerprinting, offers a powerful tool for identifying subtle variations in oil composition. When coupled with machine learning techniques like Partial Least Squares Regression (PLSR), it becomes possible to quantify adulteration levels accurately. This research not only serves as a practical example of applying advanced analytical methods to a pressing real-world problem but also contributes to the broader efforts to maintain the authenticity and purity of food products in the market.

To train the PLS machine learning model effectively, a representative set of samples, referred to as the calibration or training set, is required. This set should consist of extra virgin olive oil samples with varying levels of hazelnut oil adulteration, ranging from 0% to 60%. In Figure 2a, spectra of 20 samples with different degrees of olive oil adulteration are presented. Each sample's Raman spectrum comprises 731 recorded Raman shift values. Additionally, 13 independent samples with percentage of adulteration in the

calibration set's range are shown in Figure 2b. These independent samples are reserved for the final testing of the model's performance in quantifying the adulteration percentage in samples not involved in the model's training process.

Prior to entering the calibration model process, data can undergo preprocessing to prepare it for machine learning, making the relevant information regarding adulteration percentage more accessible to the model. These preprocessing steps can sometimes lead to simpler calibration models by eliminating or reducing the impact of signal components with weaker connections to the desired response (in this case, the percentage of impurity adulteration). Various methods are employed for this purpose. In Figures 2c and 2d, Raman spectra of the calibration and test sets after applying mean centering preprocessing on the measured signals in each Raman shift value are depicted.

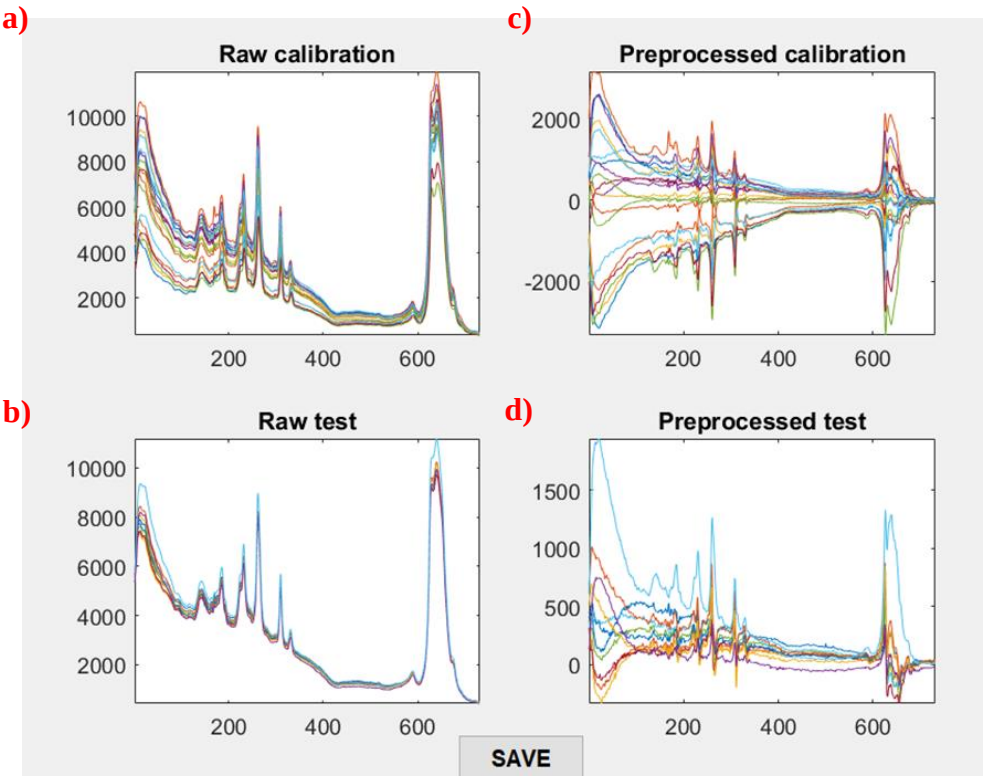

Figure 2- Raman spectra of adulterated extra virgin olive oil samples with hazelnut oil; a) raw calibration set, b) raw test set, c) mean centered calibration set, and d) mean centered test set.

Mean centering preprocessing, from a computational standpoint, is straightforward. The Raman spectroscopic data matrix essentially contains the Raman spectra of each sample arranged in rows of the data matrix. In this matrix, each column represents the variations in the Raman signal at a specific wavelength for all samples. To apply mean centering preprocessing, the mean of the Raman signals in each column is subtracted from every signal in that column. This process can eliminate the influence of constant values at each wavelength, such as background signals, and sometimes result in the creation of simpler and more interpretable calibration models. As it can be shown in the preprocessed data in Figures 2c and 2d, negative values appear.

The MVC1 software, initially developed for first-order multivariate calibration, made its debut in the public domain as a MATLAB graphical user interface (GUI) and was first documented in a research paper

[6]. This release marked an enhancement of an earlier version, which was originally a Visual Basic program. The most recent MATLAB-based iteration of MVC1 is readily accessible for users and can be freely obtained at http://www.iquir-conicet.gov.ar/descargas_/mvc1.zip. Furthermore, an independent, compiled version of the software is also available, eliminating the need for MATLAB installation on the user's computer. You can access this stand-alone version through the following link: https://www.dropbox.com/sh/nruf3lp0ge1gbww/AAAj6r97UBMIhgQmukRGYFPKa?dl1⁄40.

The MVC1 software was utilized, and the necessary steps for performing PLSR calibration were executed.

3.3. One-Leave-Out cross validation

The determination of the number of PLS factors plays a crucial role in defining the complexity of the training model, and it is essential to find the optimal model. To achieve this, a one-leave-out cross-validation procedure is employed. In this cross-validation process, the number of PLS factors is systematically adjusted by iteratively excluding one training sample at a time and using the remaining samples for constructing the latent factors and regression. Subsequently, the predicted concentrations are compared with the actual concentrations for each calibration sample, and the predicted error sum of squares (PRESS = $\Sigma(C_{actual} - C_{predicted})^2$) is computed. Within the MVC1 software, there is a dedicated option for specifying the maximum number of factors to be evaluated.

Following the completion of the cross-validation procedure, the ideal number of factors for prediction purposes is determined based on the Haaland and Thomas [8] criterion, which relies on the F significant test. Once the one-leave-out cross-validation algorithm concludes, several statistical parameters are computed as a function of the number of retained factors. These parameters include the Prediction Residual Error Sum of Squares (PRESS), the Root Mean Square Error of Cross-Validation (RMSECV), the Relative Error of Cross-Validation (RECV%, calculated as RMSECV divided by the average concentration of the mean parameter of interest in the training samples), and the correlation coefficient between actual and predicted concentrations (R^2). Additionally, a built-in feature for outlier detection is available to facilitate this crucial task.

Figure 3 illustrates the results of applying MVC1 to determine the appropriate number of factors (latent vectors) in the PLS model based on one leave-out cross-validation algorithm. In this dataset, at this stage five factors have created a model that minimizes the PRESS value. The software also provides a logarithmic plot to visualize changes in PRESS, particularly when the scales of variations are significantly different. This graphical representation aids in understanding the model's performance.

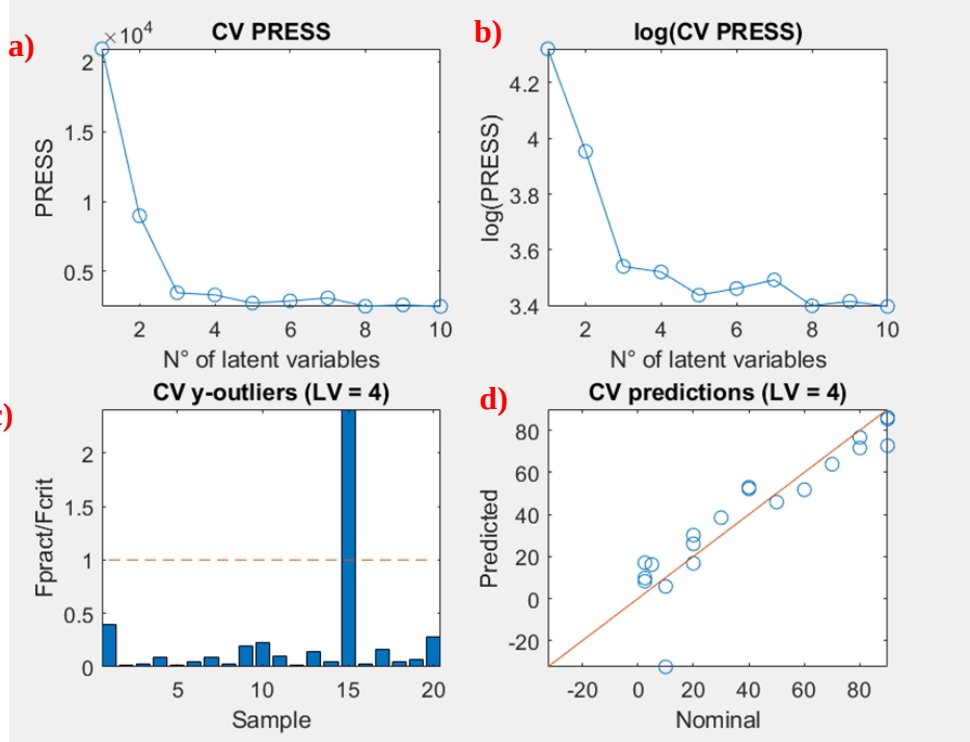

Figure 3- MVC1 graphic for CV results, a), b) PRESS and log(PRESS) as a function of number of latent vectors, c) graphical representation of F-test for outlier detection in calibration set, and d) predicted vs. nominal concentration in CV.

3.4. Outlier detection

In multivariate calibration for prediction, an outlier refers to a data point that significantly deviates from the expected or normal behavior of the dataset. Outliers can be caused by errors, noise, or irregularities in the data. Outlier detection is the process of identifying and flagging these data points to prevent them from adversely affecting the accuracy and reliability of the calibration model. Detecting and handling outliers is crucial because they can distort the relationship between the measured variables and the property of interest, potentially leading to incorrect predictions. Effective outlier detection techniques help ensure that the multivariate calibration model provides robust and accurate results.

The MVC1 software conducts an outlier detection test for both calibration and prediction samples. This test is based on assessing the significance of F statistics. It involves calculating the variance of the residuals derived from the reconstruction of the spectral data of a test sample using a PLS model, and comparing it to the variance of residuals from all calibration samples. More precisely, the experimental F_{exp} value is computed using the following equation:

$$F_{exp} = \frac{I \sum_{j=1}^{J} e_j^2}{\sum_{j=1}^{J} \sum_{i=1}^{I} e_{ij}^2}$$

In this equation, I represents the number of calibration samples, e_j is the element of the residual vector e = x - x_A for the specific test sample at wavelength j, and e_{ij} corresponds to the spectral residues at wavelength j for calibration sample i. F_{exp} is then compared to the critical F_{crit} value with I and I * J degrees of freedom to determine significance.

In case the calculated F value is smaller than the critical F value, it implies that the null hypothesis is accepted, and the test sample is not considered an outlier. The MVC1 software simplifies outlier detection by reporting the ratio of the calculated F value to the critical F value for each sample (when the test sample did not participate in the calibration model

creation). If this ratio is greater than 1, it indicates the presence of an outlier. As shown in Figure 3c, during the cross-validation phase, sample number 15, as indicated in the bar plot, is identified as an outlier. By performing this analysis, the MVC1 software advises users to remove any potential outlier sample from the calibration set before proceeding with the calibration process.

During the calibration phase, when creating a calibration model for the mixture of extra virgin olive oil and hazelnut oil, using four PLS factors, sample 15 was identified as an outlier. Consequently, the optimization of the number of model factors was carried out after excluding the outlier sample. Ultimately, as depicted in Figure 4, a five-factor calibration model is constructed without any outliers in this dataset. The prediction results for the test set can be observed in Figure 5.

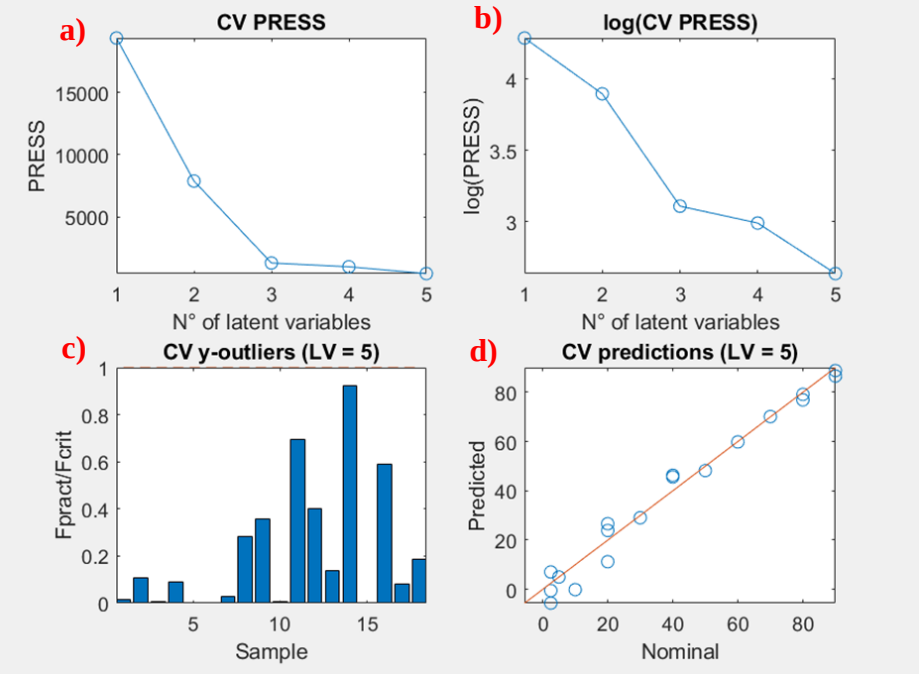

Figure 4- MVC1 graphic for CV results after removing the sample 15 from calibration set; a) and b) PRESS and log(PRESS) as a function of number of latent vectors, c) graphical representation of F-test for outlier detection in calibration set, and d) predicted vs. nominal concentration in CV

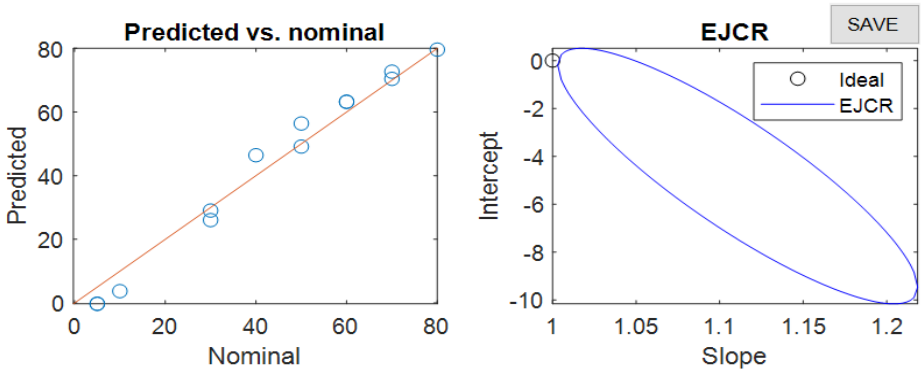

Figure 5- MVC1 output for prediction the test set; a) predicted vs nominal concentration (red line represents the ideal calibration curve), and b) elliptical confidence region obtained by EJCR test for assessing accuracy and precision of the model

3.5. Statistic parameters

The reported statistical parameters for calibration models, play a vital role in assessing the performance and reliability of these models. These parameters are crucial in gauging the effectiveness of calibration models, ensuring the quality of analytical results, and enabling informed decision-making in various applications, including quality control and product testing.

MVC1 software provides a comprehensive list of statistical parameters for evaluating the constructed calibration model (Figure 6). These parameters are essential tools for users to thoroughly assess the model's performance. They include metrics that gauge the quality of predictions and the model's ability to handle complex data. In the following, a concise explanation of these key parameters will be provided.

Statistics

RMSEP	4.1336
REP%	10.4427
R2	0.9935
Cal. X residuals	19.8108
Expl. var. X	100.0000
Expl. var. y	99.9565
Durbin-Watson DW	1.7711
Durbin-Watson p	0.6240

Figure 6- MVC1 statistics output optimized PLS model for prediction the percentage of adulteration in extra virgin olive oil

3.6. Root Mean Square Error in Prediction (RMSEP)

RMSEP is a statistical metric used to assess the performance of a calibration model in predicting the concentration or values of an analyte in unknown samples. RMSEP quantifies the accuracy of these predictions by measuring the average magnitude of the differences between the predicted values and the actual observed values. A lower RMSEP indicates a more accurate model, while a higher RMSEP suggests less accuracy in the predictions. RMSEP is a valuable parameter for evaluating the predictive power of a calibration model in various applications. RMSEP calculated according to:

$$RMSEP = \sqrt{\frac{\sum_{n=1}^{N_{val}} \left(y_{norm,n} - y_{pred,n}\right)^2}{N_{val}}}$$

where N_{val} is the number of validation samples, $y_{nom,n}$ is the nominal concentration of analyte n in the validation samples, and $y_{pred,n}$ is the predicted concentration in the same samples.

3.7. Relative Error of Prediction (%REP)

%REP is calculated as the ratio of the Root Mean Square Error in Prediction (RMSEP) to the mean concentration of the analyte within the calibration dataset.

$$REP = 100\frac{RMSE}{\langle y \rangle}$$

In simpler terms, it evaluates how well the calibration model's predictions match the true values of the analyte. A lower %REP indicates that the model's predictions are closer to the actual concentrations, signifying better predictive performance. Conversely, a higher %REP implies a larger discrepancy between predictions and true values, suggesting reduced predictive accuracy.

This parameter is indispensable in analytical chemistry because it helps researchers and analysts assess the reliability and precision of calibration models, ensuring that they produce trustworthy and accurate predictions, ultimately leading to more robust and effective analytical methods. By understanding and monitoring %REP, one can fine-tune calibration models and optimize their performance for various applications.

3.8. R-Squared (correlation coefficient) R^2

R^2 is a statistical measure that represents the proportion of the variance in the dependent variable (concentration) that can be explained or accounted for by the independent variables (e.g., spectral data). R^2, ranges from 0 to 1, where:

- R^2 = 0 means that none of the variance in the dependent variable is explained by the independent

variables, indicating a poor model fit.
- R^2 = 1 means that all of the variance in the dependent variable is explained by the independent variables, indicating a perfect model fit.
In calibration modeling, a higher R^2 value generally indicates a better fit of the model to the data, suggesting that the independent variables (e.g., spectral data) can explain a significant portion of the variance in the dependent variable (e.g., chemical concentrations). It's an essential parameter for assessing the quality and reliability of calibration models, as it quantifies how well the model captures the relationship between variables.

3.9. Explained variance

Explained variance in the context of calibration modeling refers to the portion of the total variance in both the independent variable (X, usually spectral data) and the dependent variable (Y, often analyte concentrations) that is accounted for by the calibration model. In other words, it quantifies how well the model captures the relationships between the measured spectral information (X) and the actual concentrations of analytes (Y).
For the independent variable X, explained variance assesses how much of the spectral variability is effectively utilized by the model to predict the dependent variable Y, such as analyte concentrations. Higher explained variance in X suggests that the model efficiently extracts relevant spectral information for prediction.
Similarly, for the dependent variable Y, explained variance measures the proportion of the variance in analyte concentrations that is accurately predicted by the calibration model. A higher explained variance in Y indicates that the model can effectively estimate the analyte concentrations. In calibration, maximizing the explained variance in both X and Y is a key objective to ensure the model's accuracy and reliability in predicting analyte concentrations based on spectral data. The explained variance in the context of calibration modeling can be calculated using the following equation:

Explained Variance (EV) = 1 – [(**Residual Variance** (RV)) / (**Total Variance** (TV))]
Where:
- EV represents the explained variance.
- RV represents the residual variance, which is the variance of the differences between the observed values and the predicted values (the errors).
- TV represents the total variance, which is the variance of the observed values.
The explained variance (EV) is a value between 0 and 1, where a higher value indicates that the model can explain a larger proportion of the total variance in the data, implying a better fit of the model to the data.

3.10. Durbin–Watson statistical test

Various tests have been proposed in the scientific literature to explore whether there are non-linear relationships between multiple variables and concentrations [9]. One straightforward approach is to begin by constructing a PLS-1 model using the ideal number of latent variables. This model predicts the concentration of the substance of interest in a group of validation samples. If the system behaves non-linearly, the prediction errors will display noticeable correlations when organized based on increasing predicted concentration values.
To accurately identify these correlations and avoid any misleading visual cues, the Durbin-Watson statistical test, introduced by Durbin and Watson in 1950, is applied. The test calculates the Durbin-Watson indicator (DW), which is defined as:

$$DW = \frac{\sum_{n=1}^{N_{val}-1}(r_{n+1} - r_n)^2}{\sum_{n=1}^{N} {r_n}^2}$$

Here, r_n represents the prediction residue at the nth position. A high DW value suggests uncorrelated residuals, as it results from a significant number of differences between positive and negative values. In contrast, correlated residuals lead to a low DW value, as they generate series of both positive and negative values with relatively few differences between consecutive residues. To make a statistical judgment about these DW values, there is an associated probability 'p'. When 'p' is less than 0.05, the null hypothesis (indicating uncorrelated residuals) is

rejected, signaling the presence of correlations among residuals, and vice versa.

After the calibration model is constructed and its validation based on statistical parameters is completed, the model can be used to predict an independent set (the test set) to assess its predictive capabilities. MVC1 software provides useful graphical metrics in the test set, as shown in Figure 5, for predicting the test set of hazelnut oil adulteration in extra virgin olive oil. In this figure, two plots illustrate the predicted concentration values for hazelnut oil impurity in 13 test samples relative to their actual concentrations (left plot). Another unique feature in the MVC1 software is the drawing of EJCR plot (right plot in Figure 5), which requires further explanation.

When dealing with a broad range of analyte concentrations in test samples where constant variance cannot be assumed, it is advisable to employ linear regression by plotting predicted values against nominal analyte concentrations. The assessment of these results should not merely revolve around determining if the ideal conditions of a unit slope and zero intercept fall individually within their respective confidence intervals around the means. Instead, a more robust approach is the elliptical joint confidence region (EJCR) test, which entails the creation of an EJCR for both the slope and intercept within the linear plot mentioned earlier. This test involves scrutinizing whether the ideal point (with slope of 1 and intercept of 0) lies within the boundaries of the ellipse, thus providing a more comprehensive evaluation of the linear regression model's performance. The EJCR plot in Figure 5 shows relatively a good prediction model of adulteration modeling in extra virgin olive oil.

In the realm of analytical chemistry, figures of merit emerge as crucial quantitative parameters, serving as indispensable tools for characterizing and benchmarking the performance of analytical techniques. These metrics are meticulously tailored to specific analytical methodologies, enabling scientists and researchers to evaluate the precision, accuracy, and reliability of various approaches. Figures of merit play a pivotal role in guiding the selection of the most appropriate analytical methods for specific applications, ultimately contributing to the quality and robustness of analytical data. Moreover, they facilitate the objective comparison of different techniques, ensuring that the chosen methodology aligns with the analytical goals of the study. In this field, figures of merit are instrumental in driving advancements in analytical chemistry, leading to more accurate and effective analytical processes and outcomes.

The MVC1 toolbox performs the calculation of figures of merit for PLS regression, as illustrated in Figure 7. These figures of merit serve as a comprehensive performance evaluation for quantifying adulteration in extra virgin olive oil with hazelnut oil. To enhance clarity, concise explanations for each of these terms will be provided.

AFOMs

SEN	63.3433
Anal. SEN	3.1974
LODmin	3.6321
LODmax	8.5579
LOQmin	10.8964
LOQmax	25.6736

Figure 7- Calculated Analytical Figures of Merit for PLSR modeling of Raman spectra vs Hazelnut adulteration in extra virgin olive oil

In the realm of analytical chemistry, sensitivity often takes the form of an inverse relationship with the length of the regression coefficients vector. This relationship signifies the method's ability to detect and respond to subtle changes in the input signals, typically representing concentration or specific measurement values.

A critical aspect to understand about sensitivity is that its units are represented as (signal × concentration^{-1}). This implies that the parameter's value depends on the type of signal being measured. Consequently,

sensitivity remains intricately tied to the specific attributes of the signals in question. The challenge arises when one attempts to directly compare different analytical techniques. These techniques often rely on entirely distinct types of signals, and their sensitivity values can vary significantly due to the signal-related units. Consequently, using sensitivity as a universal benchmark for comparing analytical techniques with diverse signal bases may prove inefficient. Thus, while sensitivity remains a valuable metric within its respective context, a more versatile yardstick is often required when evaluating techniques based on widely different types of signals.

Analytical sensitivity, denoted as γ, is a useful parameter that helps us understand sensitivity in analytical methods. It's valuable because it considers the specific type of signal used in calibration models. This means that sensitivity can vary between methods based on different signals, making direct comparisons challenging. γ, on the other hand, is a better choice for comparisons because it's calculated in a way that makes it more consistent and easier to understand.

Limit of detection (LOD) serves as a critical parameter in analytical chemistry and can be computed using a formula that takes into account both type I and type II errors. Specifically, it is estimated as:

$$LOD = 3.3\ s_0$$

Here, s_0 represents the standard error in the concentration of an analyte in a sample that is free from the analyte of interest. The value "3.3" corresponds to the probabilities for type I (α) and type II (β) errors, set at 0.05. However, it's important to note that the LOD is influenced by the concentrations of other components within a given sample. Additionally, to calculate LOD, it is necessary to estimate standard error in the predicted analyte concentration for a blank sample (s_0).

In the context of multivariate analysis, the blank samples can be variable, leading to a range of blank leverages. Consequently, this results in a spectrum of detection limits, ranging from a minimum LOD (LOD_{min}) to a maximum LOD (LOD_{max}). This approach acknowledges the variability in the blank samples and provides a range that of LOD values, which is particularly useful for analytical assessments in situations where the blank samples can differ significantly [10].

Limit of quantitation (LOQ), a crucial parameter in analytical chemistry, is the analyte concentration at which the relative prediction error is limited to a maximum of 10%. Much like the limit of detection (LOD), LOQ can be calculated based on the same principles and assumptions. Using the following equation:

$$LOQ = 10\ s_0$$

Here, s_0 still represents the standard error in the concentration of the analyte in a sample that is free from the analyte of interest. This equation allows us to determine the point at which analytical measurements become sufficiently precise for quantifying the analyte with a relative error of no more than 10%. Understanding both the LOD and LOQ is vital in analytical chemistry, as they establish the lower limits for reliable detection and quantification of analytes, ensuring the accuracy and precision of analytical results.

4. Conclusion

This contribution has demonstrated the practical application of the MVC1 software in quantifying hazelnut oil adulteration in extra virgin olive oil using Raman spectroscopy and PLS regression. The study emphasized the importance of figures of merit in analytical chemistry, discussed the concept of analytical sensitivity (γ), and introduced LOD and LOQ as crucial parameters. The MVC1 software proved to be a valuable tool for constructing a robust PLSR calibration model. It simplified key steps like outlier detection and determination of optimal factors. The study showcased the significance of sensitivity and analytical sensitivity for understanding signal-response relationships. By combining software tools, advanced analytical techniques, and a thorough understanding of figures of merit, this research addresses the critical issue of food authenticity and safety. It

highlights the role of performance metrics in evaluating analytical methods and contributes to the advancement of analytical chemistry for ensuring the quality of food products.

References

1. Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J. Straight line regression and calibration. In: Handbook of chemometrics and qualimetrics. Elsevier. First edition. Amsterdam. 1997; Part A, 171: 231.

2. Martens H, Naes T. Multivariate calibration. John Wiley & Sons. 1992.

3. Bro R. Multivariate calibration: what is in chemometrics for the analytical chemist? Analytica Chimica Acta. 2003; 500(1-2): 185-194. https://doi.org/10.1016/S0003-2670(03)00681-0

4. Martens H, Martens M. Multivariate analysis of quality: an introduction. John Wiley & Sons. 2001.

5. Wilson ID, Brinkman UAT. Hype and hypernation: multiple hyphenation of column liquid chromatography and spectroscopy. TrAC Trends in Analytical Chemistry. 2007; 26(9): 847-854. https://doi.org/10.1016/j.trac.2007.07.007

6. Olivieri AC, Goicoechea HC, Inon FA. MVC1: an integrated MatLab toolbox for first-order multivariate calibration. Chemometrics and Intelligent Laboratory Systems. 2004; 73(2):189-197. https://doi.org/10.1016/j.chemolab.2004.03.004

7. Chiappini FA, Goicoechea HC, Olivieri AC. MVC1_GUI: A MATLAB graphical user interface for first-order multivariate calibration. An upgrade including artificial neural networks modelling. Chemometrics and Intelligent Laboratory Systems. 2020; 206: 104162. https://doi.org/10.1016/j.chemolab.2020.104162

8. Haaland DM, Thomas EV. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Analytical Chemistry. 1988; 60(11): 1193-1202. https://doi.org/10.1021/ac00162a020

9. Centner V, De Noord O, Massart D. Detection of nonlinearity in multivariate calibration. Analytica Chimica Acta. 1998; 376(2): 153-168. https://doi.org/10.1016/S0003-2670(98)00543-1

10. Allegrini F, Olivieri AC. IUPAC-consistent approach to the limit of detection in partial least-squares calibration. Analytical Chemistry. 2014; 86(15): 7858-7866. https://doi.org/10.1021/ac501786u