The development of green analytical methods to monitor adulteration in honey by UV-visible spectroscopy and chemometrics models

Omar Elhamdaoui[1]*, *Aimen* El Orche[2], *Houda* Bouchafra[3], *Miloud* El Karbane[1], *Amine* Cheikh[3], and *Mustapha* Bouatia[1]

[1]Laboratory of Analytical Chemistry & Bromatology, Team of Formulation and Quality Control of Health Products Faculty of Medicine and Pharmacy, Mohammed V University Rabat, Morocco
[2]Laboratory of Chemical Processes and Applied Materials University of Sultan Moulay Slimane Beni-Mellal, Morocco
[3]Faculty of Pharmacy, Abulcasis University Rabat, Morocco

Abstract. The development of green and environmentally friendly analytical methods for agri-food products is an essential element to be treated by green analytical chemistry. In this study, UV-Visible spectroscopy, combined with a mathematical and statistical or chemometrics algorithm, has been developed to monitor honey quality. Partial Least Squares Regression (PLS-R) and Support Vector Machine Learning Regression (SVM-R) showed an adequate quantification of the percentage of impurity. The use of these models demonstrates a high ability to predict the quality of honey. R-square's high value shows this ability, and the low value of root mean square error of calibration and cross-validation (RMSECV, RMSEC). The results indicate that UV-Visible spectroscopy allied with the Chemometrics algorithms can provide a quick, non-destructive, green, and reliable method to control the quality and predict honey's adulteration level.

1 Introduction

Honey is one of the oldest food products most adulterated after olive oil and milk [1]. Sugars and water are the key chemical constituents of honey (typically 80% carbohydrate and 17% water). In contrast, proteins, pigments, aromas and flavors, free amino acids, vitamins, biologically active compounds, and numerous volatile compounds constitute the minor components [2]. According to the Codex definition, Alimentarius honey is supposed to be a pure product, and consequently, the addition or removal of the other substance is prohibited [3]. Honey adulteration is a challenging problem in the world. Honey can be adulterated in various ways. The addition of sugar syrup is the usual primary method of adulterating honey, one of the most common contaminants due to its similar composition to honey and low price [4].

*Corresponding author: omar.elhamdaaoui@gmail.com

Currently, techniques are applied to authenticate honey include high-performance liquid chromatography, gas chromatography, C-isotope approach, and nuclear magnetic resonance [5,6]. Although some of these methods have demonstrated to be useful for the detection of contaminants in honey, these techniques present disadvantages like an expensive, destructive, time consuming, and they require a skilled operator; consequently, these methods cannot be efficiently executed in routine food laboratories where screening techniques for quick analysis are more suitable.

This study aimed to evaluate the ability of partial least squares regression and artificial intelligence, particularly support vector machine coupled to UV-Visible spectroscopy data to develop quick and suitable methods for the optimization and discrimination of the pure and adulterated honey and predict the adulteration level of honey.

2 Method

A total of forty-five samples of honey have been prepared and adulterated using sugar syrup and analyzed by UV-Visible spectroscopy. The adulteration samples used for learning ranged between 0.5% and 30%. The following equation calculated all levels of adulteration investigated:

% Adulteration= (mass of adulterant in the sample) ×100/ (total mass of sample)

Linear regression is one of the most essential and well-understood machine learning and algorithms in statistics. As such, in the statistics domain, linear regression has been developed and is being studied as a model for understanding the relationship between one independent (or predictor) variable and one continuous dependent (or outcome) variable.

Partial Least Squares regression (PLS) is a chemometric approach that is widely used to establish a quantitative correlation between two matrices of data X and Y. X data matrix, containing generally the spectral data of a calibration set, and another matrix, Y, containing quantitative values(Fig. 1) [7,8].
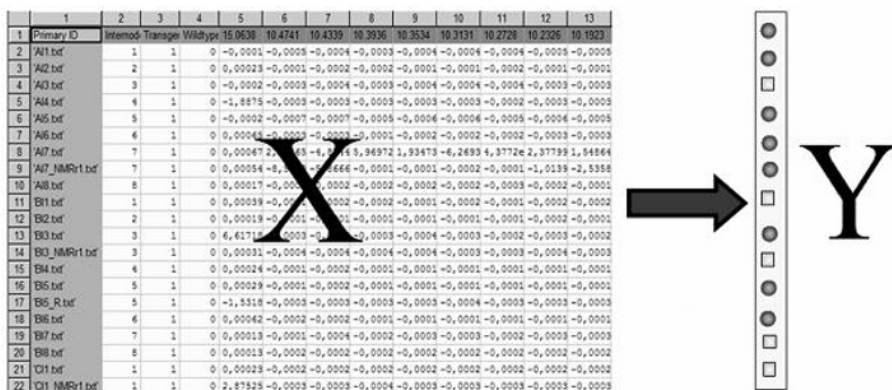
Fig 1. the information presented in the X matrix presents the numerical data and the Y matrix contains the numerical values of the predicted response [9].

The Support vector machines (SVM) approach is one of the supervised learning algorithms. It uses a mechanism called "kernels", which essentially calculates the distance between two observations. The SVM approach provides a decision limit that maximizes the distance between the most similar observations of different categories. [10].

3 Results and discussion
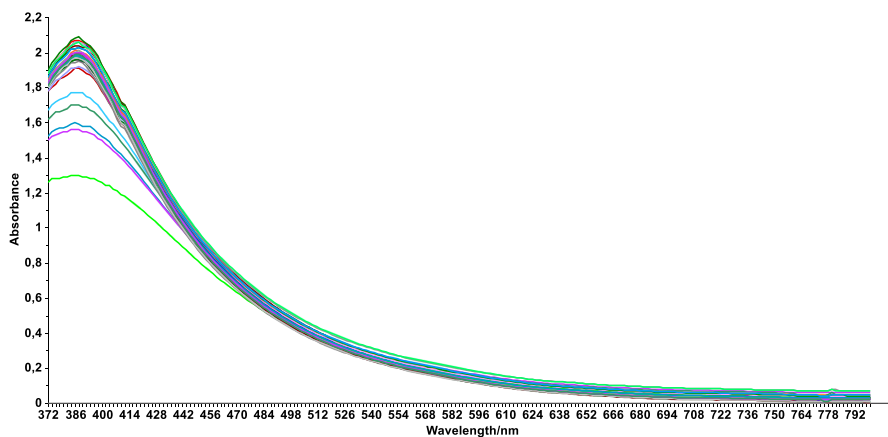
Fig. 2. UV-Visible spectra of adulterated and non-adulterated honey

The visual observation of the spectrum (Fig. 2) does not allow us to notice or to detect the difference between the adulterated and non-adulterated honey; for this reason, we use chemometric tools based on the statistical treatment of the measurement results to draw more information and to quantify the percentage of impurity.
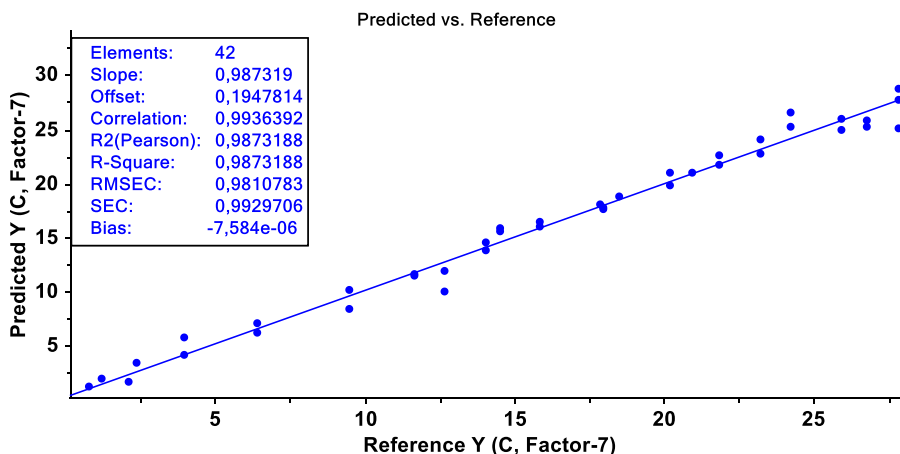
Fig. 3. Reference value vs. predicted value (PLS-R)

As shown in Fig. 3, the differences between the predicted and reference values are given by the equation $f = Y_{Estimated} - Y_{Predicted}$. However, the differences decrease with the increase of latent variables. This figure illustrates a strong correlation between actual and predicted values. The use of the PLS-R model for quantifying adulteration demonstrates high-efficiency
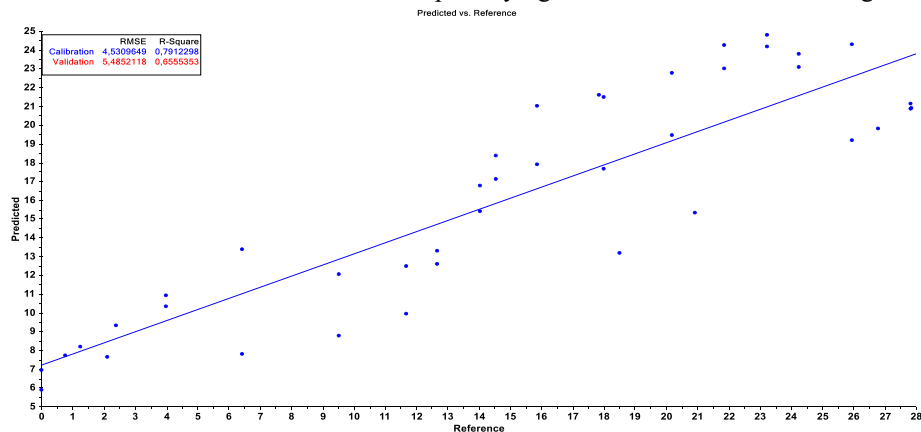
capabilities represented by a high value of the R^2, greater than 98% for both calibration and cross-validation, and lower values of RMSEC and RMSECV of 0.98 and 5.06, respectively.

Fig. 4. Reference value vs. predicted value (SVM-R)

The obtained results of the SVM approach showed relatively good results expressed by the R^2 of 0.78 which considering acceptable, and the root means a square error of calibration and validation shown in Table 1. The processing of spectral data by specific algorithms (SNV, smoothing, and baseline correction), which aim to remove undesired noise, improve the SVMR and PLS-R mathematical models' reliability.

Table 1. Table summarizing the SVM regression model's statistical parameters that evaluate the model's quality developed.

Statistical Parameters	Calibration	Cross-validation
R-square	0.79	0.65
RMSE	4.53	5.48

The results obtained by this study show the ability to combine the UV-Visible spectral method with mathematical and statistical regression methods in quantifying the adulteration of honey by sugars. These results also show that the PLS method gives better results than the SVM regression method.

4 Conclusion

In this work, we proposed an exploratory study of UV-Visible spectroscopy's applicability to predict the adulteration level of honey with sugar syrup. Mathematical and statistical tools

were found appropriate for applying a feasible experimental technique to monitor honey quality. The benefit of this method is non-destructive, rapid, and doesn't require environmentally harmful reagents.

Reference

1. J. C. Moore, J. Spink, and M. Lipp, J. Food Sci. (2012)
2. D. W. Ball, J. Chem. Educ. (2007)
3. European Commission, Off. J. Eur. Communities (2002)
4. J. D. Kelly, C. Petisco, and G. Downey, J. Agric. Food Chem. (2006)
5. J. W. White and K. Winters, J. Assoc. Off. Anal. Chem. (1989)
6. S. Wang, Q. Guo, L. Wang, L. Lin, H. Shi, H. Cao, and B. Cao, Food Chem. (2015)
7. S. Wold, M. Sjostrom, and L. Eriksson, Chemom. Intell. Lab. Syst. (2001)
8. A. El Orche, M. Bouatia, and M. Mbarki, J. Anal. Methods Chem. (2020)
9. J. Trygg and T. Lundstedt, in *Handb. Metabonomics Metabolomics* (2007)
10. K. M. Mendez, S. N. Reinke, and D. I. Broadhurst, Metabolomics (2019)