PBPC
ISSN 2674-9432
Qualis A3 CAPES 2021-2024
DOI - Crossref
Latindex
Indexado no Google Acadêmico

Authentication and quantification of adulterants in high-quality honey from the Brazilian semiarid region based on NIR spectroscopy and chemometrics

*Franciel S. Costa[a], Gutenberg L. Silva[a], Heronides A. Dantas Filho[b], Marisa S. Gomes[a], Léia S. Carvalho[a], Marcos A. Sousa[a], Kelly G. F. Dantas[b], Tiago C. Menezes[*a]*

https://doi.org/10.36557/2674-9432.2026v5n1p2451-2471
Artigo recebido em 4 de Janeiro e publicado em 4 de Março de 2026

ARTIGO ORIGINAL

RESUMO

Este estudo propõe uma metodologia rápida, não destrutiva e sustentável baseada em espectroscopia no infravermelho próximo (NIR) combinada com quimiometria para autenticar méis multiflorais de alta qualidade produzidos na região semiárida do Piauí e quantificar adulterantes regionais. Amostras autênticas foram obtidas de cooperativas certificadas, enquanto misturas binárias foram preparadas em laboratório utilizando xaropes de açúcar e de milho e adulterantes ácidos (vinagre e limão), que são regionalmente empregados para alterar a acidez, o sabor e prolongar a vida útil. Técnicas de pré-processamento espectral (SNV, MSC e derivada de Savitzky–Golay) e diferentes modelos multivariados foram aplicados. Entre os modelos de classificação avaliados (SIMCA, DD-SIMCA e OCPLS), todos alcançaram sensibilidade máxima (1,00), sendo que o DD-SIMCA clássico apresentou o melhor desempenho geral ao atingir a maior especificidade e acurácia (0,95). Modelos de regressão por mínimos quadrados parciais (PLS) também foram desenvolvidos para prever a pureza e quantificar os xaropes adulterantes, apresentando alto poder preditivo (R^2 > 0,95) e baixo viés sistemático. O modelo para xarope de açúcar apresentou os menores erros de predição (RMSEP = 2,61%), indicando maior robustez. Os resultados confirmam a eficácia da espectroscopia NIR combinada com quimiometria como uma alternativa promissora para controle de qualidade, rastreabilidade e detecção de fraudes em méis do bioma Caatinga.

Palavras-chave: mel da região semiárida, espectroscopia NIR, classificação de uma classe, adulteração, regressão por mínimos quadrados parciais.

ABSTRACT

This study proposes a rapid, non-destructive, and sustainable methodology based on near-infrared (NIR) spectroscopy combined with chemometrics to authenticate high-quality multifloral honeys produced in the semiarid region of Piauí and to quantify regional adulterants. Authentic samples were obtained from certified cooperatives, while binary mixtures were prepared in the laboratory using sugar and corn syrups and acidic adulterants (vinegar and lemon), which are regionally employed to alter acidity, flavor, and extend shelf life. Spectral preprocessing techniques (SNV, MSC, and Savitzky–Golay derivative) and different multivariate models were applied. Among the classification models evaluated (SIMCA, DD-SIMCA and OCPLS), all achieved maximum sensitivity (1.00), with the classic DD-SIMCA providing the best overall performance by reaching the highest specificity and accuracy (0.95). Partial least squares (PLS) regression models were also developed to predict purity and quantify adulterant syrups, presenting high predictive power ($R^2 > 0.95$) and low systematic bias. The model for sugar syrup yielded the lowest prediction errors (RMSEP = 2.61%), indicating greater robustness. The results confirm the effectiveness of NIR spectroscopy combined with chemometrics as a promising alternative for quality control, traceability, and fraud detection in honeys from the Caatinga biome.

Keywords: honey from the semi-arid region, NIR spectroscopy, one-class classification, adulteration, partial least squares regression.

Instituição afiliada – a - Federal Institute of Education, Science, and Technology of Piauí, Campus Paulistana, Piauí, 64750-000, Brazil.
b - Institute of Exact and Natural Sciences, Federal University of Pará, 66075-110, Brazil.

Autor correspondente: *Tiago Corrêa Menezes - tiago.menezes@ifpi.edu.br*

1 INTRODUCTION

Honey is a natural and sweet substance produced by bees from the nectar of Flowers (Calle et al. 2023). Due to its nutritional value and bioactive properties, it has been extensively studied worldwide (Cartaxo et al. 2024). Honey is widely used as a sweetener in the food industry and as a functional ingredient in cosmetic formulations.[3] The most common variety, blossom honey, is produced by *Apis mellifera* bees from the nectar of multiple plant species, and its composition varies according to geographic origin, seasonal conditions, and local flora (Raypah et al. 2022).

In Brazil, commercial honey production is predominantly carried out using *Apis mellifera*, a stinging bee species responsible for large-scale yields from thousands of apiaries (Naves et al., 2023). The country has become one of the world's leading honey producers, benefiting from its climatic and botanical diversity, which favors the production of mono- and multifloral honeys with well-defined physicochemical properties and high nutritional value (Lianda; Castro, 2008; Trevisol et al., 2022). However, honey composition is influenced by multiple factors, including botanical origin, soil type, climate, processing, and storage conditions (Dranca et al., 2022; García-Seval et al., 2022; Silva et al., 2024; Silva et al., 2024).

As the only naturally produced sweetener suitable for direct consumption, honey is often subject to economically motivated adulteration through the addition of cheaper sweeteners (Se et al., 2019; Soares et al., 2017). Common adulterants include high-fructose corn syrup (HFCS), corn syrup (COSS), inverted sugar syrup (ISS), and cane sugar syrup (CASS), whose use depends on regional cost and availability (Biswas; Chaudhari, 2024; Guler et al., 2014; Ruiz-Matute et al., 2007).

Globally, the choice of adulterant varies: wheat and rice syrups are common in Turkey and France; maltose and rice syrups are widely used in China due to their detection difficulty; date syrup is predominant in Iran; and in India, jaggery syrup, a dark unrefined sugar from palm tree sap, is frequently used (Tosun, 2013; Li et al., 2017; Al Eid, 2006; Mishra et al., 2010). In Brazil, major adulterants include glucose, fructose, sucrose, and cane molasses, particularly from C4 plants such as corn and sugarcane, whose isotopic signatures can be detected via carbon isotope ratio (13C/12C) analysis (Jaafar et al., 2020).

Honeys produced in Brazil's semi-arid northeast, especially in the state of Piauí, have unique compositional profiles shaped by the Caatinga biome. The arid-adapted flora imparts a rich content of phenolic compounds and flavonoids, along with a distinct mineral composition and high antioxidant potential (Silva et al., 2024). Despite their value, these honeys are often adulterated with low-cost syrups, compromising their authenticity and market potential.

While well-established analytical techniques such as gas chromatography (GC) and liquid chromatography (LC) are effective for fraud detection, they require extensive sample preparation, the use of solvents and reagents, and entail high operational costs (Calle et al., 2023; Sotiropoulou et al., 2021; Martinello et al., 2022). There is, therefore, a growing need for alternative methods that are faster, more affordable, and environmentally sustainable. Near-infrared (NIR) spectroscopy, particularly when combined with chemometrics, presents a promising solution by enabling rapid analyses with little or no sample preparation, at relatively low cost and with reduced environmental impact (Menezes et al., 2025; Jaafar et al., 2020).

Recent studies by Bodor et al. (2023), Raypah et al. (2022) and Calle et al. (2023) have demonstrated the successful application of NIR spectroscopy and multivariate models for detecting and quantifying sugar adulteration in honey, yielding high correlation coefficients (R^2) and low prediction errors (RMSEP). These findings confirm the potential of NIR-chemometric approaches as robust tools for quality control and fraud detection.

Given the high diversity of honeys from Brazil's semi-arid region, this study aims to (i) authenticate high-quality multifloral honeys produced in the semiarid region of Piauí, Brazil, using near-infrared (NIR) spectroscopy combined with one-class chemometric models, and (ii) quantify the levels of adulterants commonly found in the region, including sugar-based syrups (corn syrup and homemade sugar syrup) and acidic adulterants (vinegar and lemon juice). The inclusion of the latter, although not containing sugars, is justified by their regional use to alter the physicochemical parameters of honey, such as acidity and flavor, with the intent of masking defects or extending shelf life. The proposed approach therefore combines classification models and multivariate regression to provide a rapid, non-destructive, and sustainable tool for quality control and traceability of honeys from the Caatinga biome.

2 METHOD

Samples

Multifloral honey samples, certified as high-quality products from the semi-arid region of Piauí, Brazil, were obtained from cooperatives located in the municipalities of Picos, Jacobina, and Paulistana. 41 authentic samples were collected from five production batches harvested between 2022 and 2024.

The high-quality honey supplied by the cooperatives is produced by *Apis mellifera*, a stinger-bearing Africanized honeybee species. The authenticity of the samples was certified by the cooperatives, which also ensured full traceability from the beehives to the honey extraction process.

Sample collection was conducted at the cooperatives' industrial facilities following standardized protocols. To avoid compositional changes, the samples were stored in light-protected plastic vials and maintained at 25 °C until adulterated mixture preparation and spectral acquisition. Homemade sugar syrup (100 g of homemade sugar to 50 mL of water) and citrus lemon juice were prepared in the laboratory of the Federal Institute of Piauí (IFPI), Paulistana campus, whereas corn syrup and acetic acid vinegar were purchased from local supermarkets. These low-cost products, commonly associated with honey adulteration in the region, were used as adulterants to prepare the mixtures described below.

Preparation of adulterated mixtures

The sample set consisted of 328 binary mixtures, which were prepared in the laboratory by intentionally adding different proportions of homemade sugar syrups, lemon citrus juice, commercial vinegar, and corn syrup (adulterants) to samples of organic honey produced between 2022 and 2024. All adulterated samples were prepared to have a final weight of 10 grams. In addition to the binary mixtures, the sample set also included 41 authentic samples from the 5 batches of organic honey acquired, as well as pure samples of the adulterants (mentioned above), as detailed in Table 1.

The levels of adulterants in the binary mixtures ranged from 1% to 30% (w/w) in 1% (w/w) intervals and from 30% to 95% (w/w) in 5% (w/w) intervals

Table 1.

Authentic honey samples and adulterated mixtures prepared.

Sample type	Number of samples
Authentic honey	41
Prepared mixtures	
Binary (BH:SS)	129
Binary (BH:CS)	162
Binary (BH:V)	18
Binary (BH:LJ)	19
Total	369

Note. BH: Bee honey; SS: Sugar syrup; CS: Corn syrup; V: Vinegar; LJ: Lemon juice

NIRS data acquisition

The absorbance spectra were acquired in transmittance mode using a high-resolution Fourier Transform NIR spectrometer (ABB Bomem, model MB3600) equipped with a deuterated triglycine sulfate (NIR-DTGS) detector. This instrument was operated in the wavelength range from 12,800 to 4,000 cm^{-1} (780 nm to 2,500 nm), with a spectral resolution of 4 cm^{-1}. A quartz cuvette with an optical path of 1 mm was used to make the transmittance measurements. All samples were analyzed in triplicate, and each spectrum was acquired as an average of 16 scans. Background spectra were measured with the 1 mm quartz cuvette empty and clean in the sample compartment. The ambient temperature was kept close to 23 ºC and relative humidity between 50% and 60%.

Chemometric procedures

Spectral preprocessing techniques, including standard normal variate (SNV), multiplicative scatter correction (MSC), and Savitzky–Golay smoothing with a 19-point window and first derivative, were evaluated to correct a moderate scatter effect

observed in the spectra. Mean centering was applied prior to all modeling procedures. The performance of the tested preprocessing techniques, as well as the behavior of the high-quality honey samples, were assessed using Principal Component Analysis (PCA) models.

As detailed below, the methodology initially involved a one-class classification approach to identify authentically high-quality honey samples. Subsequently, PLS regression models were constructed to predict honey purity and the content of homemade sugar syrup and corn syrup used as adulterants. After several exploratory analyses of the spectral data, the most informative spectral regions for model construction were determined to be between 4,130 cm^{-1} and 8,000 cm^{-1}. This range was selected due to the presence of the most relevant absorption bands associated with the principal chemical variations of honey constituents and their adulterants, significantly contributing to the performance and robustness of the chemometric models developed.

Classification models

One-class classification techniques were applied to categorize samples from five batches of high-quality honey originating from the Brazilian semi-arid region, with or without the presence of adulterants in their composition. This method involves defining a boundary around the target class, allowing a sample to be classified based on whether it falls within or outside this boundary (Xu et al., 2014). This approach is particularly relevant given the broad diversity and complexity in the composition of honey characteristics of the semi-arid northeastern region.

Three one-class classifiers were employed, traditional Soft Independent Modeling of Class Analogies (SIMCA), Data Driven SIMCA (DDSIMCA) (Zontov et al., 2017) and One-class partial least squares (Xu et al., 2013; Pomerantsev, 2008).

While traditional SIMCA relies on fixed thresholds or confidence intervals to establish acceptance areas, its variant DD-SIMCA considers the distributions of scores distances and orthogonal distances are dependent on the data itself. Consequently, in DD-SIMCA method performs a statistical analysis on score distances and orthogonal distances to construct the acceptance areas, according to a specified value of the type I error (Hourant et al., 2000; Zontov et al., 2017). Outliers have a considerable impact

on the estimation of the mean and covariance matrix, thereby influencing the classical PCA decomposition. Robust statistical techniques can be employed to calculate the parameters of the chi-square distribution for score distances and orthogonal distances (Hourant et al., 2000). This enables more effective treatment of extreme objects and outliers within the analysis.

In one-class partial least squares (OCPLS) (Pomerantsev, 2008), a variant of partial least squares (PLS) regression, a latent variable approach is utilized to establish fundamental relationships between a matrix X (containing samples arranged along its rows and features along its columns) and a response vector (containing as many rows as the number of samples and all its values set to 1). A cross-validation step is usually carried out to find the significant number of latent variables. OCPLS employs two metrics to assign a given sample to a target class. The Absolute Centered Residuals (ACR) of predicted response measures the magnitude of the deviation of each sample from the model's prediction. The ACR of a given sample belonging to the training set is calculated as the difference between observed and predicted responses subtracted by the mean of training errors. The Hotelling's T^2 statistic assesses the variability of the target class with basis on the score distances of each sample to the center of the class in the space of the latent variables. Given a confidence level, thresholds are defined based on the distributions of Hotelling's T^2 score distances and ACR values for training objects (estimated by cross validation).

In the training step, the target class for the classification models included authentic honey samples. In the test step, samples belonging to the target class were evaluated, as well as samples adulterated with homemade sugar syrups, corn syrup, and samples with acidity adulterated using commercial vinegar and lemon juice.

The performance of the models can be assessed by the analysis of contingency tables, which provide results of false negative (FN), false positive (FP), true negative (TN), and true positive (TP).

Sensitivity, specificity and accuracy are common figures of merit used to evaluate classification and authentication models. The first is the ability of the model to correctly classify positive samples as positive and is represented by Eq. (1).

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

where: TP = number of positive samples classified as positive; FN = number of positive samples classified as negative.

Specificity is the ability of the model to correctly classify negative samples as negative and is mathematically represented by Eq. (2).

$$Specificity = \frac{TN}{FP+TN} \quad (2)$$

where: TN = number of negative samples classified as negative; FP = number of negative samples classified as positive.

Accuracy refers to the percentage of both true positives and true negatives predicted with respect to the total number of samples and is represented by Eq. (3).

$$Accuracy = \frac{TP+TN}{N_T} \quad (3)$$

Where N_T is the total number of samples (TP + TN + FP + FN).

PLS regression models

PLS models were constructed to predict the content of adulterant syrups (homemade sugar and corn syrup) and the purity of honey in authentic and intentionally adulterated samples. The optimal number of latent variables was determined using the venetian blinds method for cross-validation of the PLS models. The SPXY (sample set partitioning based on joint x–y distances) algorithm (Galvão et al., 2005) was employed with previously pre-processed spectra to select the cross-validation and external validation sets for the PLS models, at proportions of 70% and 30%, respectively. Hotelling's T^2 and Q Residuals statistics were used for outlier detection.

The predictive capability of the developed PLS models was assessed using the analytical performance parameters RMSECV (Root Mean Square Error of Cross-validation), RMSEP (Root Mean Square Error of Prediction), determination coefficient (R^2), and bias. The known proportions of adulterant content and authentic honey in the prepared mixtures were used as reference values to build the PLS models.

Software

All calculations involving data pre-processing, selection of samples by SPXY (Galvão et al., 2005) were made by in-house routines implemented with MATLAB™ (MathWorks, Inc., Natick, MA). SIMCA and PLS regression models were developed in the PLS_Toolbox 8.2 interface (Eigenvector Research, Inc.) working under Matlab. Data Driven SIMCA models were built using the DDSIMCA guide user interface described by Zontov et al. (2017). One-class PLS classifier was employed using Matlab routines as described by Xu et al. (2014).

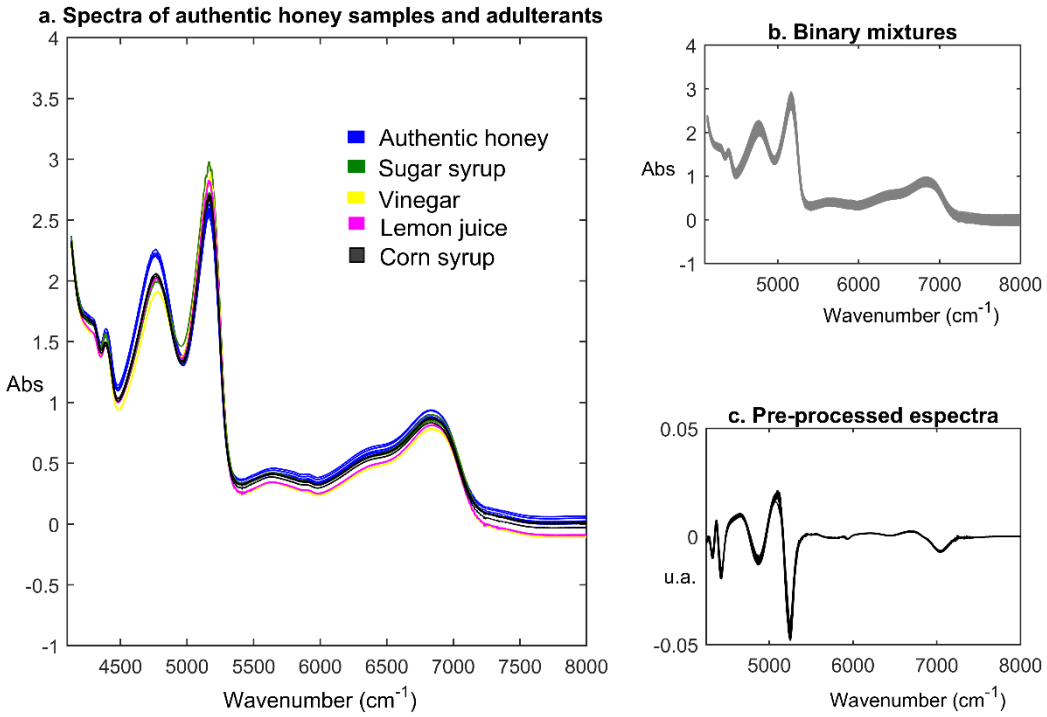
3 RESULTS AND DISCUSSION

Spectral data

In Figure 1, the NIR spectra of honey samples are presented. Two intense absorption bands are prominent at 5,165 cm^{-1} and 4,760 cm^{-1}, associated, respectively, with the first O–H harmonic, the first overtone of O–H stretching, the combination of O–H stretching and bending, the first overtone of O–H bending, and the C–O stretching band. In both the raw spectra (Figure 1a) and the preprocessed spectra (Figure 1c), a pronounced peak is observed in the 5,160 cm^{-1} region, which, according to Bodor et al. (2023), is related to water content. This band exhibited higher absorbance values in sugar syrup samples, followed by other adulterants, which can be attributed to the higher water content in these matrices (Biswas; Chaudhari, 2024). Conversely, the band at 4,760 cm^{-1} showed greater absorption in authentic honey samples, possibly associated with the characteristic sugars of these honeys. Despite these specific differences, the overall spectral profiles of authentic and non-authentic honey samples are remarkably similar, highlighting the need to employ multivariate approaches that simultaneously consider all wavelengths.

Figure 1

Raw NIR spectra of authentic honey samples and their adulterants (a), binary mixtures containing honey and adulterant (b), and preprocessed spectra (c).



Exploratory Analysis

Principal Component Analysis (PCA) was applied to investigate the similarity and dissimilarity relationships between authentic honey samples and their adulterants. The score plot based on the first two principal components (PC1 and PC2) is shown in figure 2a, encompassing authentic samples from five honey batches and various adulterants. The samples were clustered according to the type of adulteration and authenticity, indicating clear distinctions among them. Notably, corn syrup exhibited a distribution close to that of the authentic honey samples, which may hinder its detection and quantification as an adulterant.

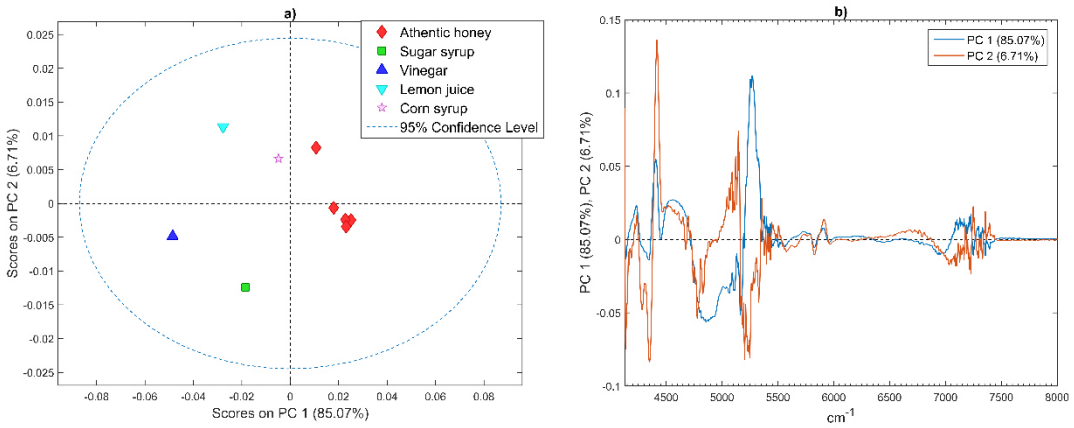
The first two principal components explained 91.78% of the total variance in the spectral data, capturing most of the relevant information. The loading plots (Figure 2b) revealed the wavelengths with the highest contribution to the discrimination between honey and adulterants, particularly at 4,300 cm^{-1}, 4,800 cm^{-1}, 5,250 cm^{-1}, and 5,400

cm^{-1}. According to Calle et al. (2023), these peaks correspond to overtone regions of C-H stretching, O-H stretching (from water), and N-H stretching, which are associated with key honey components such as carbohydrates, water, proteins, and amino acids (Biswas; Chaudhari, 2024).

The findings demonstrate that honey adulteration with common substances such as homemade sugar syrup, acetic vinegar, and corn syrup can be effectively identified through clear group separations.

Figure 2.

PCA scores and loadings of the preprocessed spectra for five authentic honey samples (one from each batch) and their adulterants.



Classification of Honey Samples

Classification models—SIMCA, DD-SIMCA and OCPLS—were employed for honey authentication, aiming to discriminate between authentic and non-authentic samples. For the development of one-class models, a total of 319 samples were used, comprising 41 authentic and 278 adulterated samples. The adulteration levels ranged from 10% to 100%, a range selected based on limitations observed in detecting adulterations at lower levels, as well as the low practical occurrence of fraud involving adulterant concentrations below 10%.

Sample splitting for training and test sets was performed exclusively with target-class samples (authentic honeys) using the Kennard–Stone algorithm, with a 70:30 ratio (n = 28 for training and n = 13 for testing). Acceptance thresholds were defined based on the training set and subsequently used to predict sample authenticity in the testing phase.

The SIMCA model was constructed using principal component analysis (PCA) as a preliminary step to identify outliers via Hotelling's T^2 and Q-residual statistics. The most effective discrimination criterion was the square root of the sum of squared standardized T^2 and Q values,

which yielded a sensitivity of 1.00, specificity of 0.92, and overall accuracy of 0.92. One sample (from the oldest batch) was identified as an outlier and excluded from the training set, likely due to elevated HMF levels, as previously reported by Kesić et al. (2014).

The classic version of the DD-SIMCA model, developed using six principal components and without outlier removal, outperformed SIMCA, achieving a sensitivity of 1.00, specificity of 0.95, and overall accuracy of 0.95. The acceptance region was defined based on the chi-squared distribution at a significance level of 1% ($\alpha = 0.01$). The absence of outliers indicated greater intraclass homogeneity, which contributed to more accurate modeling of authentic sample variability.

The OCPLS model, constructed with four latent variables, also achieved maximum sensitivity (1.00), although with slightly lower specificity (0.91) and overall accuracy (0.91). These results suggest a greater tendency for false positives, a common trade-off in models with high sensitivity. No outliers were detected, indicating robustness in modeling the target class. Classification criteria were based on comparison of ACR values and score distances with thresholds defined during model training, as proposed by Xu et al. (2014).

Table 2.

Parameters of analytical performance for the SIMCA, DD-SIMCA and OCPLS classification models developed

	SIMCA		*DD-SIMCA (classic)*		*OCPLS*	
	Total	Outliers	Total	Outliers	Total	Outliers
Nº of training samples	28	1	28	0	28	0
	Target	Non-target	Target	Non-target	Target	Non-target
Nº of test samples	13	278	13	278	13	278
Nº of PCs/LVs	5		6		4	
Sensitivity	1.00		1.00		1.00	
Specitivity	0.92		0.95		0.91	
Acuracy	0.92		0.95		0.91	

PLS regression models

Partial Least Squares (PLS) regression models were developed to predict honey purity, corn syrup content, and homemade sugar syrup content. The samples used for model development covered a wide range of purity levels, from 0% to 100%, with the

total number of samples for calibration (70%) and prediction (30%) varying according to the variable under study (Table 3).

The honey purity model was constructed using 269 samples for calibration and 116 for prediction, encompassing the full purity range (0–100%). This model included 44 authentic honey samples and samples adulterated with homemade sugar syrup (132) and corn syrup (167), in addition to samples adulterated with commercial acetic vinegar (18), homemade lemon juice (19), and pure adulterant samples. The model yielded coefficients of determination (R^2) of 0.95 for calibration and 0.97 for prediction. The root mean square error of cross-validation (RMSECV) was 6.86, while the root mean square error of prediction (RMSEP) was 4.92. The prediction bias was close to zero (–1.54), indicating no significant systematic trend. Despite the large number of samples, this model exhibited higher absolute errors compared to the others, suggesting moderate yet satisfactory precision.

For the corn syrup content, 136 samples were used for calibration and 59 for prediction, covering the same concentration range (0–100%). The model showed high coefficients of determination (R^2 = 0.98 for both calibration and prediction), with an RMSECV of 3.77 and RMSEP of 3.35. The prediction bias was also negligible (0.13), demonstrating high precision and the absence of systematic error, indicating suitable robustness for predicting this parameter.

The model for homemade sugar syrup content employed 110 samples for calibration and 47 for prediction, equally distributed across the 0–100% range. This model exhibited the best performance among the three, with very high coefficients of determination (R^2 = 0.99 for both calibration and prediction), and the lowest RMSECV (2.86) and RMSEP (2.61) values, along with an almost null prediction bias (–0.20). Although based on a smaller sample size, the model demonstrated high robustness and reliability for predicting sugar syrup content in the analyzed samples.

Table 3.

Parameters of analytical performance of PLS models

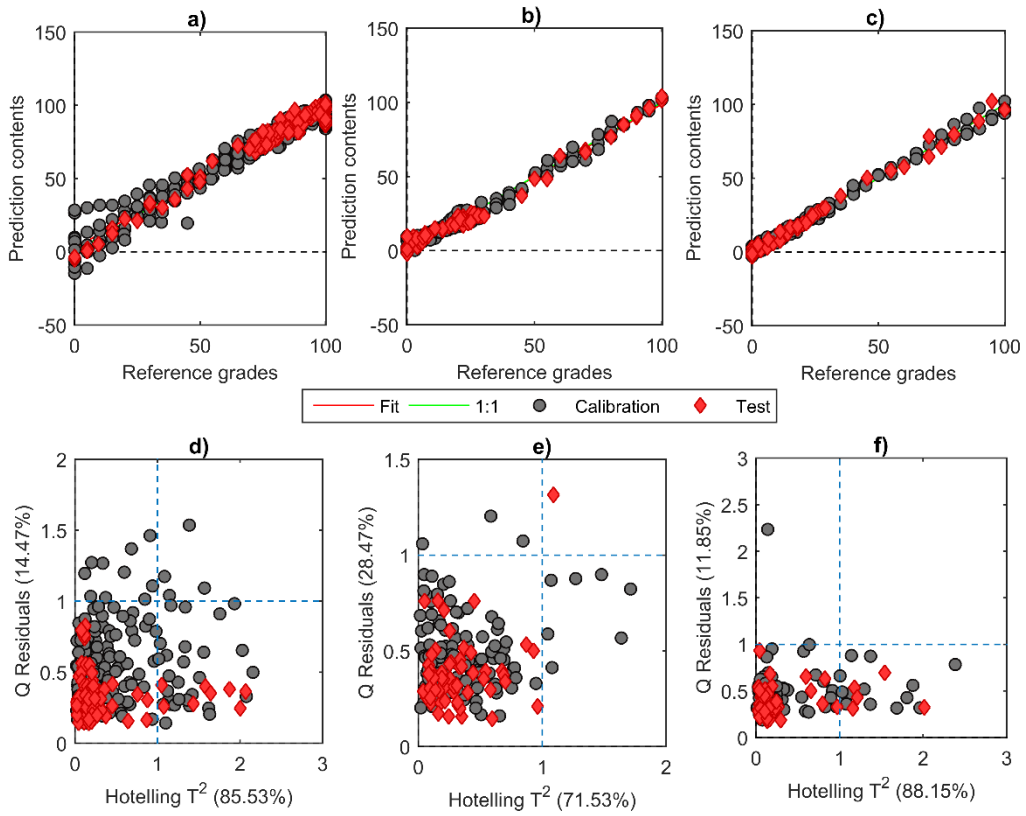
Model	N. of samples		Purity	VL	R^2		RMSECV	RMSEP	bias	
	Cal.	Pred.	range (%)		Cal.	Pred.			Cal.	Pred.

Purity content of honey	269	116	100 – 0	3	0.95	0.97	6.86	4.92	-0.02	-1.54
Corn syrup content	136	59	100 – 0	3	0.98	0.98	3.77	3.35	0.07	0.13
Sugar syrup content	110	47	100 – 0	3	0.99	0.99	2.86	2.61	0.01	-0.20

Figure 3 presents the plots of reference values versus predicted values (a–c) and Hotelling's T^2 scores versus Q residuals (d–f) for the honey purity, corn syrup content, and homemade sugar syrup content models, respectively. All three models demonstrate excellent correlation between reference and predicted values, with data points closely aligned along the 1:1 line, indicating high predictive capability. A slight dispersion is observed in plot (a), particularly for calibration samples with lower purity levels, which is consistent with the higher RMSEP observed for the honey purity model. In contrast, the corn syrup (b) and sugar syrup (c) content models show tighter distributions, especially the latter, confirming their superior precision and robustness.

The Hotelling's T^2 versus Q residuals projections (d–f) further support the quality of the models by showing a low incidence of samples beyond the critical limits (blue dashed lines). The honey purity model (d) shows approximately 14.47% of samples exceeding the Q residual threshold, indicating slightly higher variability. The corn syrup content model (e) displayed the highest percentage of outliers (28.47%), which may be attributed to greater compositional variability among the samples or to specific systematic errors. On the other hand, the sugar syrup content model (f) exhibited only 11.85% of samples outside the limits, the lowest among the three, confirming it's superior robustness and statistical stability in the face of sample variability.

Figure 3.
Reference versus predicted values and Hotelling's T² scores versus Q residuals for the models of purity (a), corn syrup content (b), and homemade sugar syrup content (c) in honey samples from the semi-arid region of Piauí.



4 CONCLUSION

All models achieved maximum sensitivity (1.00), demonstrating a strong ability to correctly identify authentic samples. The classic DD-SIMCA exhibited the best overall performance, with specificity and accuracy of 0.95, outperforming SIMCA and OCPLS, the latter showing a greater tendency toward false positives.

The partial least squares (PLS) calibration models showed excellent data fit ($R^2 > 0.95$), high predictive capacity, and no relevant systematic bias, confirming their reliability for quantitative applications. The sugar syrup quantification model stood out as the most robust, presenting the lowest prediction error (RMSEP = 2.61%), followed by the corn syrup model, which also demonstrated satisfactory performance. The honey purity model, although statistically adequate, displayed higher prediction error and lower robustness, likely due to the inclusion of multiple adulterants, such as vinegar and

lemon juice, which increased dataset variability while broadening the model's applicability by enabling the detection of a wider range of contaminants.

Overall, these findings reinforce the potential of NIR spectroscopy combined with chemometrics as a rapid, reliable, and cost-effective approach for honey authentication, supporting quality control, traceability, and food safety monitoring.

REFERENCES

1. Al Eid, S. M.; *Int. J. Food Sci. Nutr.* **2006**, *57*, 83. https://doi.org/10.1080/09637480600658286.
2. Biswas, A.; Chaudhari, S. R. Exploring the role of NIR spectroscopy in quantifying and verifying honey authenticity: A review. *Food Chemistry*, v. 445, p. 138712, **2024**. DOI: https://doi.org/10.1016/j.foodchem.2024.138712.
3. Bodor, Z.; Majadi, M.; Benedek, C.; Zaukuu, J.-L. Z.; Veresné Bálint, M.; Csajbókné Csobod, É.; Kovacs, Z.; *Chemosensors* **2023**, *11*, 89. https://doi.org/10.3390/chemosensors11020089.
4. Calle, J. L. P.; Punta-Sánchez, I.; González-de-Peredo, A. V.; Ruiz-Rodriguez, A.; Ferreiro-González, M.; Palma, M.; *Foods* **2023**, *12*, 2491. https://doi.org/10.3390/foods12132491.
5. Cartaxo, H. B.; Silva, D. E.; Almeida, L. M.; *Rev. Bras. Filos. Hist.* **2024**, *14*, 4611. https://doi.org/10.18378/rbfh.v13i4.11158.
6. Dranca, F.; Ropciuc, S.; Pauliuc, D.; Oroian, M.; *LWT–Food Sci. Technol.* **2022**, 113910. https://doi.org/10.1016/j.lwt.2022.113910.
7. Galvão, R. K. H.; Araujo, M. C. U.; Jose, G. E.; Pontes, M. J. C.; Silva, E. C.; Saldanha, T. C. B.; *Talanta* **2005**, *67*, 736. https://doi.org/10.1016/j.talanta.2005.03.025.
8. García-Seval, V.; Mártinez-Alfaro, C.; Saurina, J.; Nuñes, O.; Sentellas, S.; *Foods* **2022**, *11*, 2345. https://doi.org/10.3390/foods11152345.
9. Guler, A.; Kocaokutgen, H.; Garipoglu, A. V.; Onder, H.; Ekinci, D.; Biyik, S.; *Food Chem.* **2014**, *155*, 155. https://doi.org/10.1016/j.foodchem.2014.01.033.
10. Hourant, P.; Baeten, V.; Morales, M. T.; Meurens, M.; Aparicio, R.; *Appl. Spectrosc.* **2000**, *54*, 628. https://doi.org/10.1366/0003702001950733.
11. Jaafar, M. B.; Othman, M. B.; Yaacob, M.; Talip, B. A.; Ilyas, M. A.; Ngajikin, N. H.; Fauzi, N. A. M.; *Int. J. Integr. Eng.* **2020**, *12*, 125. https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/5696.
12. Kesić A.; Crnkić A.; Hodžić Z.; Ibrišimović N.; Šestan A. Effects of Botanical Origin and Ageing on HMF Content in Bee Honey. *J. Sci. Res. Rep.* **2014**, *3* (8), 1057-1066. 10.9734/JSRR/2014/8095.
13. Lianda, R. L. P.; Castro, R. N.; *Quim. Nova* **2008**, *31*, 1472. https://doi.org/10.1590/S0100-40422008000600036.
14. Li, S.; Zhang, X.; Shan, Y.; Su, D.; Ma, Q.; Wen, R.; Li, J.; *Food Chem.* **2017**, *218*, 231. https://doi.org/10.1016/j.foodchem.2016.08.105.
15. Martinello, M.; Stella, R.; Baggio, A.; Biancotto, G.; Mutinelli, F.; *Metabolites* **2022**, *12*, 985. https://doi.org/10.3390/metabo12100985.
16. Menezes, T. C.; Castro, G. A. B.; Fernandes, H. A.; Gutjahr, K. E.; Dantas Filho, H. A.; Silva, N. C.; Dantas, K. G. F.; *J. Food Compos. Anal.* **2024**, *122*, 106988. https://doi.org/10.1016/j.jfca.2024.106988.
17. Mishra, S.; Kamboj, U.; Kaur, H.; Kapur, P.; *Int. J. Food Sci. Nutr.* **2010**, *61*, 306. https://doi.org/10.3109/09637480903476415.
18. Naves, K.; Menezes Filho, A. C. P.; Souza, A. O.; Taques, A. S.; Alves, I.; Ventura, M. V. A.; *Braz. J. Sci.* **2023**, *2*, 19. https://doi.org/10.14295/bjs.v2i5.366
19. Pomerantsev, A. L.; *J. Chemom.* **2008**, *22*, 601. https://doi.org/10.1002/cem.1147.
20. Raypah, M. E.; Omar, A. F.; Muncan, J.; Zulkurnain, M.; Najib, A. R. A.; *Molecules* **2022**, *27*, 2324. https://doi.org/10.3390/molecules27072324.
21. Ruiz-Matute, A. I.; Soria, A. C.; Martinez-Castro, I.; Sanz, M. L.; *J. Agric. Food Chem.* **2007**, *55*, 7264. https://doi.org/10.1021/jf070559j.
22. Se, K. W.; Wahab, R. A.; Yaacob, S. N. S.; Ghoshal, S. K.; *J. Food Compos. Anal.* **2019**, *80*, 16. https://doi.org/10.1016/j.jfca.2019.04.001.
23. Silva, D. J. S.; Soares, A. K. O.; Campos, A. L. S.; Bendini, J. N.; Moreira-Araújo, R. S. R.; *Obs. Econ. Latinoam.* **2024**, *22*, e5043. https://doi.org/10.55905/oelv22n6-031.

24. Silva, S. M.; Marcolin, L. C.; Borba, V. S.; Arias, J. L. O.; Kupski, L.; Barbosa, S. C.; Primel, E. G.; *Braz. J. Food Technol.* **2024**, *27*, e2023132. https://doi.org/10.1590/1981-6723.13223.
25. Soares, S.; Amaral, J. S.; Oliveira, M. B. P. P.; Mafra, I.; *Compr. Rev. Food Sci. Food Saf.* **2017**, *16*, 1072. https://doi.org/10.1111/1541-4337.12278.
26. Sotiropoulou, N. S.; Xagoraris, M.; Revelou, P. K.; Kaparakou, E.; Kanakis, C.; Pappas, C.; Tarantillis, P.; *Foods* **2021**, *10*, 1671. https://doi.org/10.3390/foods10071671.
27. Tosun, M.; *Food Chem.* **2013**, *138*, 1629. https://doi.org/10.1016/j.foodchem.2012.11.068.
28. Trevisol, G.; Bueno, M. P.; Oliveira, J. P. L.; Macedo, K. G.; *Rev. Gest. Secretariado* **2022**, *13*, 352. https://doi.org/10.7769/gesec.v13i3.1321.
29. Xu, L.; Goodarzi, M.; Shi, W.; Cai, C.; Jiang, J.; *Chemom. Intell. Lab. Syst.* **2014**, *139*, 75. https://doi.org/10.1016/j.chemolab.2014.09.005.
30. Xu, L.; Si-Min, Y.; Cai, C. B.; Yu, X. P.; *Chemom. Intell. Lab. Syst.* **2013**, *126*, 49. https://doi.org/10.1016/j.chemolab.2013.04.008.
31. Zontov, Y. V.; Rodionova, O. Y.; Kucheryavskiy, S. V.; Pomerantsev, A. L.; *Chemom. Intell. Lab. Syst.* **2017**, *167*, 23. https://doi.org/10.1016/j.chemolab.2017.05.010