Article

Journal of the Brazilian Chemical Society

https://dx.doi.org/10.21577/0103-5053.20240090

J. Braz. Chem. Soc. 2025, 36, 1, e-20240090 , 1-11

Assessing the Quality of Wheat Flour Blended with Cassava Starch Using a Handheld NIR Spectrophotometer and Chemometrics

Dâmaris N. C. Candeias,[a] Sara Regina R. C. de Barros,[b] Wellington S. Lyra,*,[c] David Douglas S. Fernandes[b] and Paulo Henrique G. D. Diniz*,[a]

[a]Programa de Pós-Graduação em Química Pura e Aplicada, Universidade Federal do Oeste da Bahia, 47810-04 Barreiras-BA, Brazil

[b]Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, 58429-500 Campina Grande-PB, Brazil

[c]Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, P.O. Box 5093, 58051-970 João Pessoa-PB, Brazil

This work developed an effective and eco-friendly analytical method to support quality control for Brazilian Bill proposal No. 5332/2009. The bill advocates the creation of the "Brazilian bread" by mandating the incorporation of cassava byproducts into wheat flour, with the aim of reducing Brazil's external dependency on wheat imports and promoting domestic cassava family farming. For this, a handheld near-infrared (NIR) spectrophotometer was utilized for on-site determination of cassava starch in commercial wheat flour. The quantitative analysis, which utilized NIR spectra pre-processed with offset correction (OFF) combined with linear baseline correction (LBC), alongside the successive projections algorithm for interval selection in partial least squares (iSPA-PLS), exhibited superior predictive capability, achieving a relative error of prediction below 8%. Furthermore, the classification process, which involved NIR spectra pre-processed with LBC and iSPA-PLS for discriminant analysis (iSPA-PLS-DA), yielded an impressive 94% overall accuracy, with only 3 misclassifications in the test set.

Keywords: quality control, wheat flour, near-infrared spectroscopy, variable selection, successive projections algorithm

Introduction

For centuries, bread has been a global cultural staple, typically made from wheat flour, water, and yeast, providing essential nutrients such as fiber, carbohydrates, and proteins in the human diet.[1,2] Bread typically uses wheat flour, a mix of enzymes, non-starch and starch polysaccharides, lipids, and gluten. Starch, a primary element, acts as a stabilizer, thickener, and binder in baked goods, replacing fat. Wheat flour's standout feature lies in its gluten proteins, enabling it to form a viscoelastic dough with gas retention ability, mixing tolerance, stretch resilience, and flexibility, crucial for high-quality wheat-based baking.[1,3]

Concerns about gluten safety in bread and gluten-containing foods have risen, leading to the promotion of gluten-free diets for better health. However, evidence disproves a link between gluten intake and appetite stimulation or weight gain. While most people can safely consume gluten-containing grains without experiencing health repercussions, some wheat peptides containing toxic epitopes may trigger celiac disease in genetically susceptible individuals.[4] In such a scenario, to avoid gluten intolerance, various ingredients like refined rice flour, cornmeal, corn, potato, and cassava starches have been alternatively used to prepare commercial gluten-free bread. Frequent ingestion of these items may increase health risks, such as nutrient deficiencies due to low protein, fiber, and micronutrient content, and chronic diseases due to their high glycemic impact. Furthermore, gluten-free flours lead to subpar bread characterized by rapid staling, poor texture, and unsatisfactory taste.[1]

Alternatively, the partial replacement of wheat flour by different cassava byproducts, including refined flour, shaving flour, and starch, has been already reported in the literature. Bread prepared by the addition of approximately

*e-mail: wellington@ccen.ufpb.br; paulodiniz.quimica@gmail.com, paulo.diniz@ufob.edu.br
Editor handled this article: Ivo M. Raimundo Jr. (Associate)

20% cassava flour to wheat flour did not cause adverse sensory effects.[5,6] In another study,[7] the sensory texture attributes were not significantly affected when the bread contained up to 30% of cassava byproduct mixed with wheat flour. More recently, Akintayo *et al*.[8] prepared bread from a wheat-cassava blend with added gluten, demonstrating similarities to wheat bread in terms of volume and firmness. Meanwhile, bread made from cassava flour with added gluten exhibited comparable crust color, texture, and overall acceptability to wheat bread.

In view of the above, the proposal of Bill No. 5332/2009 of the Brazilian Chamber of Deputies that creates the called "Brazilian bread" by compulsory addition of cassava byproducts to wheat flour purchased by the Brazilian Government may become viable.[9] It foresees the initial compulsory addition of 3% until reaching 10% after the 25th month of validity of the new law.[9] This bill aims to reduce the external dependency of Brazil on wheat imports, as domestic production remains insufficient to meet internal demands. Additionally, it seeks to promote increased cassava consumption, thereby supporting family farming. Despite nearly doubling wheat production between 2017 and 2022, Brazil experienced a decline from 10.8 million tons in 2022 to 8.1 million tons in 2023, coupled with an increase in imports from 4.5 million tons to 6.0 million tons.[10] This decrease aligns with uncertainties in the global wheat market due to the war between Russia and Ukraine, which accounted for nearly 30% of global wheat exports in 2021.[11] Additionally, the post-COVID-19 pandemic scenario, an energy crisis, shipping constraints, and recent climate-induced extreme events are other relevant factors that may lead to soaring wheat prices, reduced trade, and severe food insecurity. Therefore, this bill aligns with Brazil's Sustainable Development Goal No. 2, aiming to end hunger, achieve food security, improve nutrition, and promote sustainable agriculture.[11,12]

Regarding the literature, several analytical techniques have been explored to authenticate the geographical origin, variety, farming type (conventional or organic), and discriminate wheat grain and flour from other cereals/adulterants. These encompass the use of isotopic, multi-elemental, genetic, omics, spectroscopic, and image analysis combined with qualitative and quantitative chemometric methods.[13] In this context, near-infrared (NIR) spectroscopy, including hyperspectral imaging (HSI), stands out in analyzing wheat flour and wheat-based products. It is user-friendly, cost-effective, time-saving, and environmentally friendly due to its non-destructive and chemical-free nature.[3,14,15] However, few studies involving the addition of cassava by-products to wheat flour have been reported elsewhere using hyperspectral images,[16] handheld microNIR spectrometer,[17] and benchtop spectrophotometers.[18] In such a scenario, the primary constraint initially stemmed from the necessary laboratory setup for spectral measurements. Therefore, the miniaturization of NIR spectrometers gains prominence because it expands compact and portable technology for on-site and in-field measurements and can extend their use to unconventional user environments in line with the principles of "true green analytical chemistry". These devices embody characteristics such as compactness, affordability, robustness, user-friendliness, portability, and ergonomic design. They are favored by users across various fields for their efficiency. Their compact size facilitates on-the-go measurements, while affordability enhances accessibility. Enhanced robustness ensures consistent performance, while user-friendly interfaces simplify operation. Their portability seamlessly integrates into diverse workflows, and ergonomic designs prioritize user comfort, rendering them indispensable for efficient, cost-effective analyses tailored to specific requirements. As a result, this facilitates enhanced precision, efficiency, and speed in conducting food integrity checks, thereby ensuring the quality, safety, and authenticity of the supply chain, as demonstrated in recent reviews.[19-23]

This work aims to develop an analytical method using a handheld NIR spectrophotometer and chemometrics for the *in situ* determination of cassava starch added to commercial bread-making wheat flour. The goal is to support quality control for the Brazilian Bill proposal No. 5332/2009. Traditional full spectrum quantitative partial least squares (PLS) and qualitative partial least squares discriminant analysis (PLS-DA) models will be compared with variable selection-based models using the successive projections algorithm (SPA). The determination of cassava starch content involves selecting individual variables in multiple linear regression (SPA-MLR) and wavelength intervals in PLS (iSPA-PLS). Discrimination between pure and cassava starch-containing wheat flour is conducted by selecting individual variables in linear discriminant analysis (SPA-LDA) and interval selection in PLS-DA (iSPA-PLS-DA). A comprehensive description of these SPA-based methods is provided by Vallese *et al*.[24]

Experimental

Samples and data acquisition

This study involves a total of 45 pure wheat flour samples and 80 mixtures of cassava starch and wheat flour at eight levels: 3, 5, 7, 10, 15, 20, 25, and 30 mg kg^{-1}. This sample collection was obtained by blending randomly four batches of commercial wheat flour and two batches of cassava starch acquired in supermarkets located in João Pessoa city, Paraíba State, Brazil. Representative and homogeneous portions of the bulk materials were efficiently obtained by a quartering process followed by a sieving step. Lastly, they were sealed in plastic bags and placed in a vacuum desiccator at room temperature until analysis.[18]

NIR spectra from the studied samples were recorded in triplicate using a portable DLP NIRscan Nano Evaluation Module (Texas Instruments®, Dallas, Texas, USA) in the range from 900 to 1700 nm at a room temperature of 25 ± 1 °C. The diffuse reflectance measurements were performed using a pattern pixel width of 2 nm, a digital resolution of 228 wavelength points, by integrating 32 scans with a Hadamard transform, and using a polytetrafluorethylene powder (Sigma-Aldrich, St. Louis, MO, USA) as a blank. Only the average spectrum from all replicates of each sample was used to construct the chemometric models.

Chemometric procedure

Initially, six different pre-processing methods were applied to overcome the drawbacks caused by systematic variations on the baseline: (*i*) offset correction (OFF), (*ii*) linear baseline correction (LBC), (*iii*) OFF + LBC, (*iv*) multiplicative scatter correction (MSC), (*v*) standard normal variate transformation (SNV), and (*vi*) Savitzky-Golay first derivative with second-order polynomial and 13-points window (SGD).

To determine the cassava starch content in wheat flour, 56 calibration and 24 prediction samples were chosen using the SPXY algorithm. Full spectrum PLS, SPA-MLR, and iSPA-PLS (partitioned into 5, 10, 15, and 20 intervals) were then assessed. All models underwent validation via leave-one-out cross-validation, gauged by the lowest root mean square error of cross-validation (RMSECV) and the highest correlation coefficient (r_{cv}). Prediction samples were solely employed for final data evaluation and comparing multivariate calibration models. The predictive ability of the models was appraised based on root mean square error of prediction (RMSEP), correlation coefficient (r_{pred}), relative error of prediction (REP), and the ratio performance to deviation (RPD). Additionally, an ordinary least squares (OLS) fitting was estimated, and the estimated intercept (a) and slope (b) were compared (with ideal values of 0 and 1) through the elliptical joint confidence region (EJCR) test to appraise model accuracy.[25]

For qualitative analysis, an exploratory analysis of the data was initially performed using principal component analysis (PCA). Then, the Kennard-Stone (KS) uniform sampling algorithm was used to partition the dataset into 30 training and 15 test pure wheat flour samples, in addition to 56 training and 24 test cassava starch-blended wheat flour samples.[21] The performance of the PLS-DA, SPA-LDA, and iSPA-PLS-DA models was compared in terms of sensitivity, specificity, efficiency, accuracy, and Matthew's correlation coefficient (MCC):[26]

$$\text{Sensitivity}\ (\%) = \left(\frac{\text{TP}}{\text{TP} + \text{FN}}\right) \times 100 \tag{1}$$

$$\text{Specificity}\ (\%) = \left(\frac{\text{TN}}{\text{TN} + \text{FP}}\right) \times 100 \tag{2}$$

$$\text{Efficiency}\ (\%) = \sqrt{\frac{\text{TP} \times \text{TN}}{(\text{TP} + \text{FN})(\text{TN} + \text{FP})}} \times 100 \tag{3}$$

$$\text{Accuracy}\ (\%) = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}\right) \times 100 \tag{4}$$

$$\text{Matthew's correlation coefficient} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{5}$$

where TP: true positive, TN: true negative, FP: false positive, and FN: false negative. Sensitivity and specificity were calculated, considering the pure wheat flour samples as the target class.

All chemometric procedures for data pre-processing, data splitting, and multivariate quantitative and qualitative modeling were implemented using Matlab 2018b (Natick, MA, USA, Mathworks).[27]

Results and Discussion

Quantification of the cassava starch content in wheat flour

Before the construction of the chemometric models, NIR spectra of pure wheat flour samples (burgundy lines) and wheat flour blended with cassava starch (green lines) were pre-processed with offset correction (Figure 1a), linear baseline correction (Figure 1b), offset correction coupled with linear baseline correction (Figure 1c), multiplicative scatter correction (Figure 1d), standard normal variate transformation (Figure 1e), and Savitzky-Golay first derivative with second-order polynomial and 13-points window (Figure 1f). As can be seen, all pre-processing techniques were efficient to remove systematic variations on the baseline contained in the NIR spectra. In the following, they were used to construct quantitative PLS, SPA-MLR, and iSPA-PLS models for determining the cassava starch content added into wheat flour. Table 1 presents the results of the quantitative analysis obtained by the different multivariate calibration models, demonstrating the numeric advantage of using these different pre-processed NIR spectra.

As observed in Table 1, all PLS models achieved good results, regardless of the pre-processing used. They obtained RMSECV and RMSEP values consistently well below the first level of cassava starch addition of 3 mg kg^{-1}, while the cross-validation and prediction correlation coefficients were higher than 0.970. In this context, the OFF + LBC/PLS model showed good predictive capability, achieving an RMSEP of 1.458 mg kg^{-1}, r_{cv} of 0.984, RPD of 5.511, and REP of 8.461%, using 10 latent variables (Figure 2a). On the other hand, when individual variable selection is performed by SPA-MLR, the predictive capability of all generated models is not satisfactory, with RPD values below 3 and/or REP values above 16%. An exception occurs for the OFF/SPA-MLR model, which, although obtaining slightly higher RMSECV and RMSEP values than 3 mg kg^{-1}, achieved the lowest REP value in this study (7.058%), along with appropriate r_{pred} of 0.957 and RPD of 3.204, selecting only 7 individual variables (Figure 2d). For iSPA-PLS modeling, all models generally obtained similar results in terms of RMSE, r, and latent variables, regardless of the pre-processing used. Nevertheless, the selection of 12 out of 20 intervals by SPA improved the results obtained by PLS when OFF+LBC was chosen as pre-processing (Figure 2g). In this case, LBC/iSPA-PLS reduced RMSEP to 1.376 mg kg^{-1}, REP to 7.987%, and the number of latent variables included in the model to 8, while increasing r_{pred} to 0.986 and RPD to 5.838. Therefore, the OFF+LBC/iSPA-PLS model can be defined as the best approach for quantifying cassava starch in commercial wheat flour.

For comparison purposes, it is possible to observe that the individual variables selected by OFF/SPA-MLR (Figure 2d) and the intervals selected by OFF + LBC/iSPA-PLS (Figure 2g) are in the same regions of the main typical

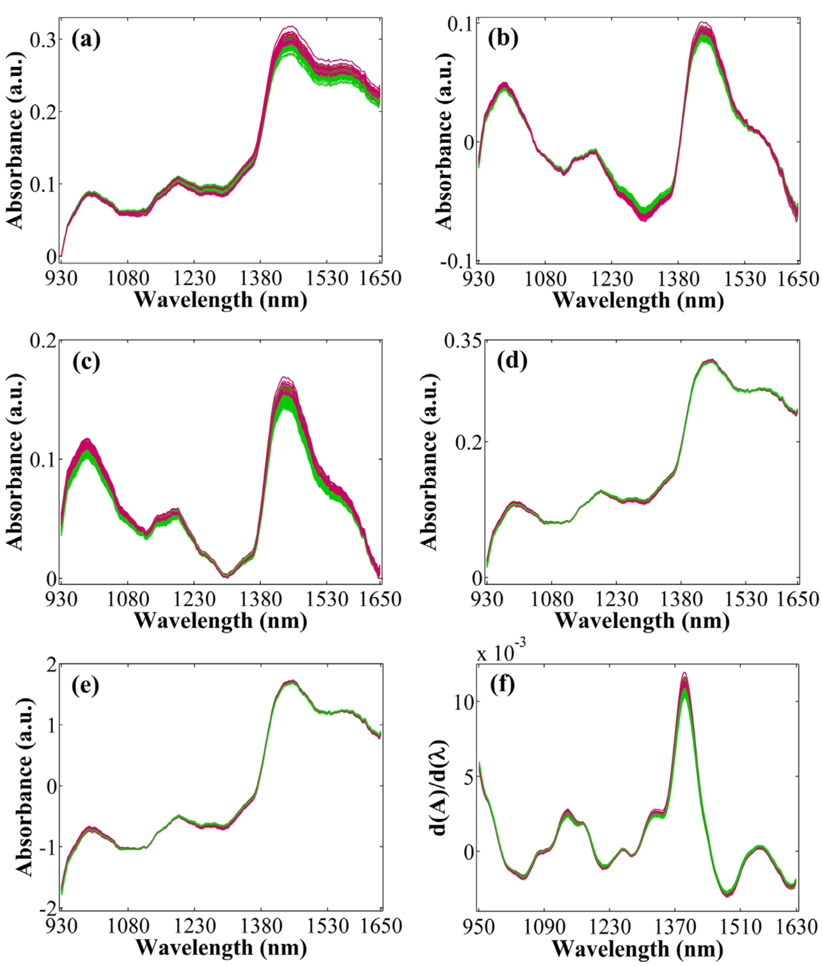

Figure 1. NIR spectra of pure wheat flour samples (burgundy lines) and wheat flour blended with cassava starch (green lines) pre-processed with offset correction (a), linear baseline correction (b), offset correction coupled with linear baseline correction (c), multiplicative scatter correction (d), standard normal variate transformation (e), and Savitzsky-Golay first derivative with second-order polynomial and 13-points window (f).

Table 1. Best results obtained by the different multivariate calibration models for the determination of the cassava starch content in wheat flour using different pre-processed NIR spectra

	Parameter						
	RMSECV / (mg kg^{-1})	r_{CV}	RMSEP / (mg kg^{-1})	r_{Pred}	RPD	REP / %	LV
Offset correction (OFF)							
PLS	1.341	0.989	1.614	0.983	5.440	10.110	9
SPA-MLR (7)[a]	3.454	0.984	4.631	0.957	3.204	7.058	–
iSPA-PLS (15/8)[a]	2.381	0.963	1.473	0.989	5.960	9.228	7
Linear baseline correction (LBC)							
PLS	1.525	0.986	1.506	0.987	5.470	9.076	8
SPA-MLR (8)[a]	2.614	0.957	2.807	0.940	2.934	16.921	–
iSPA-PLS (5/4)[a]	1.634	0.984	1.362	0.989	6.046	8.211	8
Offset correction coupled with linear baseline correction (OFF + LBC)							
PLS	1.439	0.987	1.458	0.984	5.511	8.461	10
SPA-MLR (11)[a]	2.266	0.968	4.429	0.837	1.814	25.706	–
iSPA-PLS (20/12)[a]	2.009	0.975	1.376	0.986	5.838	7.987	8
Multiplicative scatter correction (MSC)							
PLS	2.107	0.970	1.805	0.981	5.054	11.665	7
SPA-MLR (7)[a]	3.295	0.927	3.871	0.906	2.357	25.017	–
iSPA-PLS (10/8)[a]	1.973	0.974	1.805	0.982	5.054	11.666	7
Standard normal variate (SNV)							
PLS	2.107	0.970	1.805	0.981	5.054	11.665	7
SPA-MLR (7)[a]	3.308	0.926	3.873	0.906	2.356	25.030	–
iSPA-PLS (5/4)[a]	2.268	0.966	1.772	0.982	5.150	11.449	7
Savitzky-Golay derivative (SGD)							
PLS	1.686	0.981	2.009	0.978	4.429	11.958	7
SPA-MLR (10)[a]	2.045	0.972	2.894	0.955	3.074	17.227	–
iSPA-PLS (20/18)[a]	1.641	0.982	1.889	0.981	4.711	11.242	7

[a]Number of selected intervals or variables by SPA (e.g., SPA-MLR (10) means 10 individual variables selected by the SPA algorithm, or iSPA-PLS (10/6) means 5 intervals selected from 10 by the SPA algorithm). RMSECV: root mean square error of cross-validation; r_{cv}: correlation coefficient for cross-validation; RMSEP: root mean square error of prediction; r_{pred}: correlation coefficient for prediction; RPD: ratio performance to deviation for prediction; LV: latent variables used in the model; REP: relative error of prediction; PLS: partial least squares; SPA-MLR: successive projections algorithm for variable selection in multiple linear regression; iSPA-PLS: successive projections algorithm for interval selection in partial least squares.

vibrational absorption bands assigned to the O–H, N–H, and C–H relative to the primary structural components of organic compounds present in wheat flour and cassava starch, such as water, lipids, and proteins. In both cases, most of the selected variables/intervals occur at N–H and O–H absorption bands because cassava starch generally contains 17% protein (approximate) and 0.1% amylose (dry solids), while wheat flour contains 28% protein and 0.4% amylose, respectively.[14,18] Additionally, Figures 2b, 2e, and 2h illustrate the good distribution of the calibration (green circles) and prediction (orange squares) samples nearby the bisectrix line in the predicted *versus* actual plots for the quantification of cassava starch in commercial wheat flour using OFF + LBC/PLS, OFF/SPA-MLR, and OFF + LBC/iSPA-PLS, respectively. Undoubtedly, the OFF + LBC/iSPA-PLS model provides the most accurate result, as confirmed by the smaller size of the joint confidence ellipse containing the ideal theoretical point in Figure 2i, compared to the others obtained for the OFF + LBC/PLS (Figure 2c) and OFF/SPA-MLR (Figure 2f) models. The weights obtained from both the OFF + LBC/PLS and OFF + LBC/iSPA-PLS models (as shown in Figure S1 of the Supplementary Information (SI) section) confirmed that the enhanced performance of the iSPA-PLS model stems from integrating analytical information into the initial eight weights. In contrast, the PLS model exhibited increased noise in the ninth and tenth weights.

Regarding the literature, Su and Sun[16] used hyperspectral images (900-1700 nm) for quantitative detection of Irish organic wheat flour blended with cassava flour in the

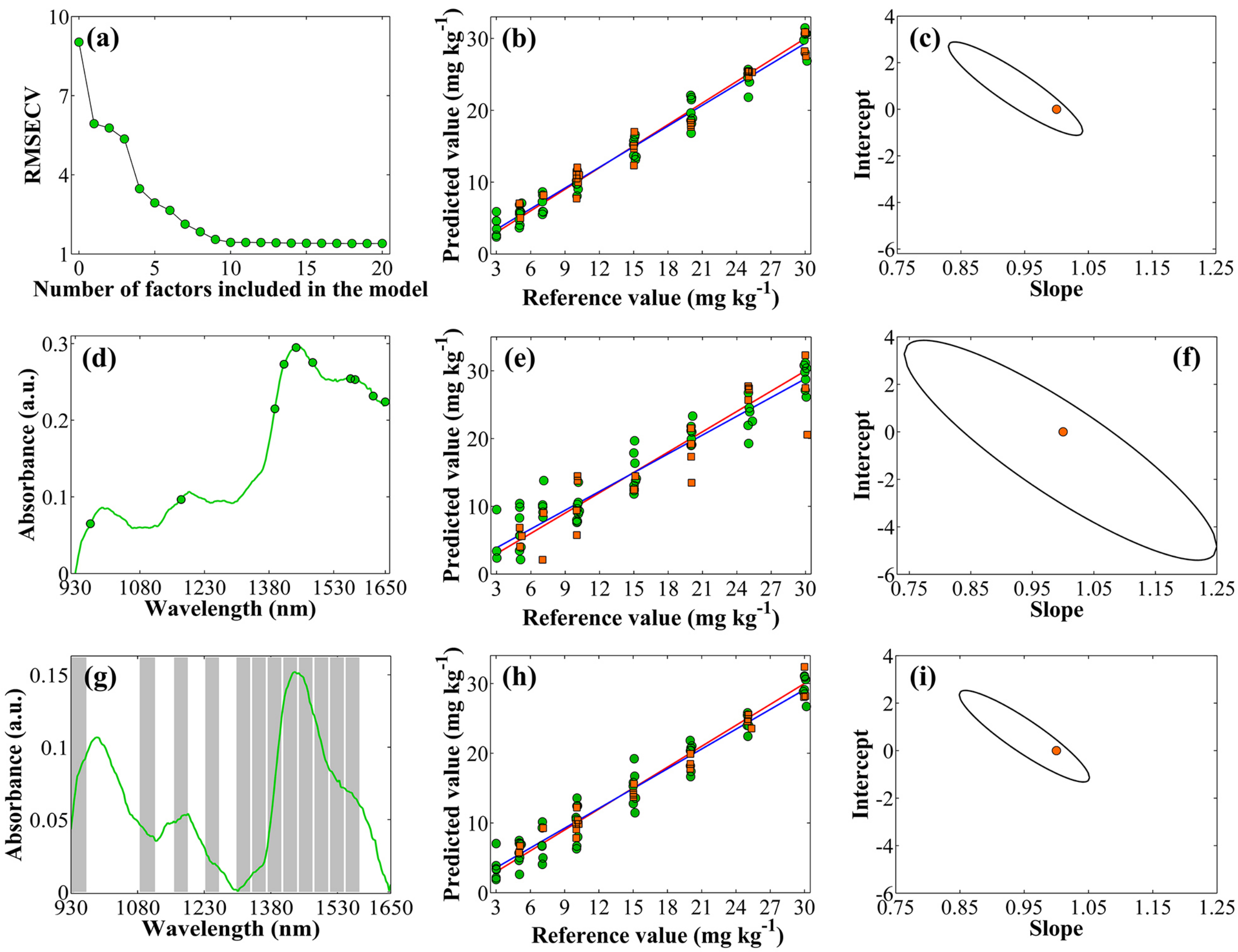

Figure 2. Best results of the determination of the cassava starch content in commercial wheat flour using the NIR spectra pre-processed with offset correction coupled with linear baseline correction (OFF + LBC), OFF, and OFF + LBC for the construction of the PLS, SPA-MLR, and iSPA-PLS models, respectively. The selected number of latent variables in PLS, individual variables in SPA-MLR, and intervals in iSPA-PLS are shown in (a,d,g), respectively, along with their respective predicted versus reference plots in (b,e,h) containing the calibration (green circles) and prediction (orange squares) samples, and confidence ellipses of the prediction models including the theoretical ideal points in (c,f,i).

range of 3-75% (m/m). For this, partial least squares (PLS) coupled with a first-derivative and mean centering iteration algorithm and variable selection based on model regression coefficients achieved a determination coefficient of prediction (R^2_{pred}) of 0.986 and a root mean square error of prediction (RMSEP) of 0.026. Other important figures of merit such as RPD and REP were not provided. On the other hand, Almeida and co-workers[18] employed a benchtop NIR spectrometer to quantify the cassava starch content in commercial wheat flour in the range of 3 to 30 mg kg^{-1}. The best result was attained by PLS using spectra preprocessed with the first derivative Savitzky-Golay smoothing with a second-order polynomial and 21-point window size, reaching a r_{pred} of 0.995, RMSEP of 1.004 mg kg^{-1}, RPD_{pred} of 9.682, and REP of 8.665%. Therefore, the results of the proposed work stand out for having a lower REP value (7.987%), in addition to using a cheaper, easy-to-use, and portable instrument.

Discrimination between pure and cassava starch-containing wheat flour

To make an initial exploratory analysis of the data, Figure 3 shows the score plots obtained by PCA using the NIR spectra were pre-processed with OFF (Figure 3a), LBC (Figure 3b), OFF + LBC (Figure 3c), MSC (Figure 1d), SNV (Figure 1e), and SGD (Figure 1f). As can be seen, there is a good trend in separating among the studied samples for all pre-processing used, besides existing some overlapping between pure and cassava starch-containing wheat flour. The observed patterns arise from variances in the chemical composition of wheat flour and cassava starch, particularly concerning the absorption bands of O–H, N–H, and C–H in relation to water, lipids, proteins, and amylose contents.[14,18] This correlation is evidenced in the loading plots depicted in Figure S2 (SI section). To solve this, discriminant

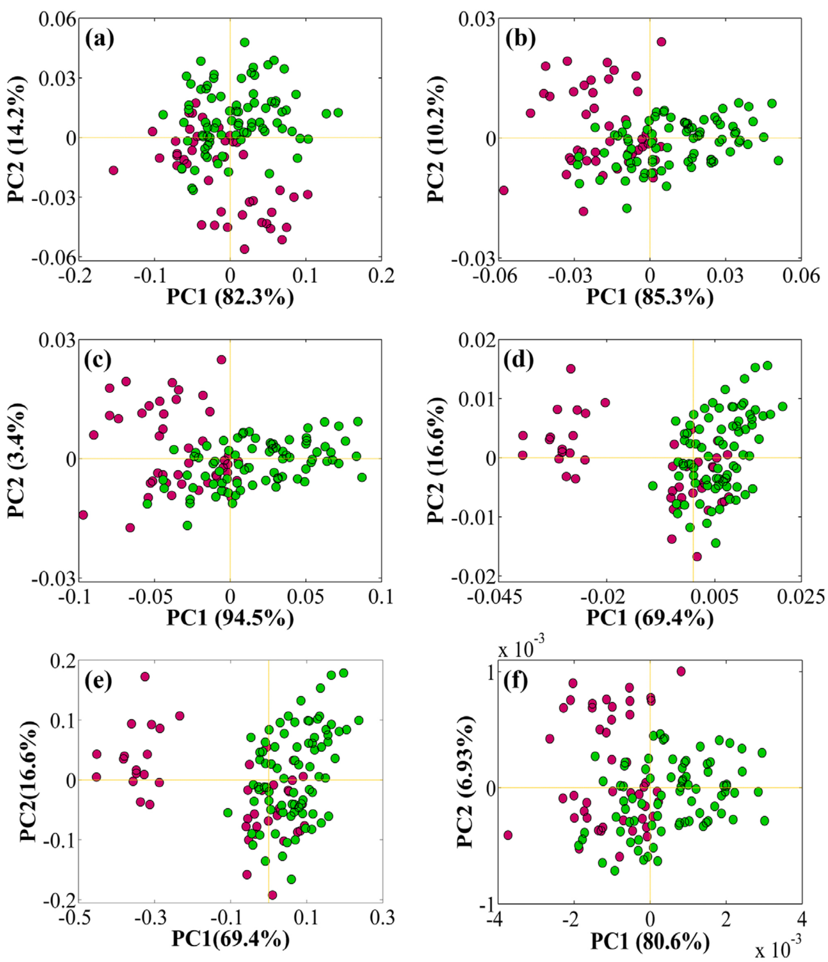

Figure 3. PCA score plots obtained using the NIR spectra of pure wheat flour samples (burgundy circles) and wheat flour blended with cassava starch (green circles) pre-processed with offset correction (a), linear baseline correction (b), offset correction coupled with linear baseline correction (c), multiplicative scatter correction (d), standard normal variate transformation (e), and Savitzky-Golay first derivative with second-order polynomial and 13-points window (f).

PLS-DA, SPA-LDA, and iSPA-PLS-DA models were then applied, as detailed in Table 2.

Observing Table 2, the PLS-DA models using OFF, MSC, and SNV achieved 100% correct classifications in the training set. However, they are clearly overfitted due to the high number of latent variables included in the models (between 9 and 16). Applying the principle of parsimony, when NIR spectra pre-processed with SGD were used to construct a PLS-DA model with 6 latent variables (Figure 4a), sensitivities were 96.7 and 93.3%, specificities were 92.9 and 87.5%, efficiencies were 94.7 and 90.4%, accuracies were 94.2 and 89.7%, and MCC values were 0.88 and 0.79 in the training and test sets, respectively. The same results were obtained for the training set using the SGD/SPA-LDA model, selecting only 10 individual variables (Figure 4c). However, its predictive ability was lower, achieving 86.7% sensitivity, 83.3% specificity, 85.0% efficiency, 84.6% accuracy, and 0.69 MCC. All other SPA-LDA models using different pre-processing techniques exhibited very poor results, as evidenced by low MCC values, at most 0.66 for the training set and 0.50 for the test set. It is important to emphasize that MCC serves as a crucial metric in binary classification, especially when handling an imbalanced distribution of samples among classes. A value near +1 indicates a highly satisfactory sample assignment for these two classes. Conversely, an MCC value close to −1 implies complete misrecognition of all samples, while zero might suggest random sample assignment.[22]

For iSPA-PLS-DA, models using OFF, OFF + LBC, MSC, and SNV also showed overfitting. Excluding these models, the highest MCC values among all studied models were obtained using LBC/iSPA-PLS-DA, reaching 0.92 and 0.84 for the training and test sets, respectively, employing only 5 latent variables and selecting only 4 out of 15 intervals (Figure 4e). This indicates that only 1 and 3 pure wheat flour samples were misclassified in the training and test sets, respectively, while only 2 wheat flour samples containing cassava starch were incorrectly identified as pure in the training set. On the other hand, all wheat flour

Table 2. Best results obtained by different classification models in discriminating wheat flour (as containing or not containing cassava) starch using different pre-processed NIR spectra

Pre-processing technique	PLS-DA				SPA-LDA				*i*SPA-PLS-DA			
	Training		Test		Training		Test		Training		Test	
Offset correction (OFF)												
	30	–	11	4	21	9	8	7	30	–	11	4
	–	56	2	22	4	52	2	22	–	56	1	23
Variables in the model[a]	9				6				9 (5/3)[b]			
Sensitivity / %	100		73.3		70.0		53.3		100		73.3	
Specificity / %	100		92.9		92.9		91.7		100		95.8	
Efficiency / %	100		82.1		80.6		69.9		100		83.8	
Accuracy / %	100		84.6		84.9		76.9		100		87.2	
MCC	1.00		0.67		0.66		0.50		1.00		0.73	
Linear baseline correction (LBC)												
	28	2	13	2	22	8	11	4	29	1	12	3
	4	52	2	22	8	48	7	17	2	54	–	24
Variables in the model[a]	6				4				5 (15/4)[b]			
Sensitivity / %	93.3		86.7		73.3		73.3		96.7		80.0	
Specificity / %	92.9		91.7		85.7		70.8		96.4		100.0	
Efficiency / %	93.1		89.1		79.3		72.1		96.5		89.4	
Accuracy / %	93.0		89.7		81.4		72.0		96.5		92.3	
MCC	0.85		0.78		0.66		0.50		0.92		0.84	
Offset correction coupled with linear baseline correction (OFF + LBC)												
	30	–	11	4	29	1	11	4	30	–	12	3
	2	54	3	21	8	48	5	19	1	55	2	22
Variables in the model[a]	8				7				16 (5/ 3)[b]			
Sensitivity / %	100		73.3		96.7		73.3		100		80.0	
Specificity / %	96.4		87.5		85.7		79.2		98.2		91.7	
Efficiency / %	98.2		80.1		91.0		76.2		99.1		85.6	
Accuracy / %	97.7		82.0		89.5		76.9		98.8		87.2	
MCC	0.95		0.62		0.59		0.43		0.97		0.73	
Multiplicative scatter correction (MSC)												
	30	–	10	5	22	8	8	7	30	–	11	4
	–	56	1	23	5	51	5	19	–	56	2	22
Variables in the model[a]	15				3				10 (5/3)[b]			
Sensitivity / %	100		66.7		73.3		53.3		100		73.3	
Specificity / %	100		95.8		91.1		79.2		100		91.7	
Efficiency / %	100		79.9		81.7		65.0		100		82.0	
Accuracy / %	100		84.6		84.9		69.2		100		84.6	
MCC	1.00		0.68		0.66		0.34		1.00		0.67	
Standard normal variate (SNV)												
	30	–	11	4	22	8	8	7	30	–	11	4
	–	56	3	21	5	51	5	19	–	56	3	21
Variables in the model[a]	16				3				9 (20/7)[b]			
Sensitivity / %	100		73.3		73.3		53.3		100		73.3	
Specificity / %	100		87.5		91.1		79.2		100		87.5	
Efficiency / %	100		80.1		81.7		65.0		100		80.1	
Accuracy / %	100		82.0		84.9		69.2		100		82.0	
MCC	1.00		0.62		0.66		0.34		1.00		0.62	

Table 2. Best results obtained by different classification models in discriminating wheat flour (as containing or not containing cassava) starch using different pre-processed NIR spectra (cont.)

	Savitzky-Golay derivative (SGD)											
	29	1	14	1	29	1	13	2	29	1	13	2
	4	52	3	21	4	52	4	20	3	53	4	20
Variables in the model[a]	6				10				6 (5/4)[b]			
Sensitivity / %	96.7		93.3		96.7		86.7		96.7		86.7	
Specificity / %	92.9		87.5		92.9		83.3		94.6		83.3	
Efficiency / %	94.7		90.4		94.7		85.0		95.6		85.0	
Accuracy / %	94.2		89.7		94.2		84.6		95.3		84.6	
MCC	0.88		0.79		0.88		0.69		0.90		0.69	

[a]The number of latent variables in PLS-DA-based models or selected variables in SPA-LDA models; [b]the number of selected intervals used for the iSPA-PLS-DA model construction is indicated in parenthesis (e.g., iSPA-PLS-DA (15/4) means 4 intervals selected from 15 by the SPA algorithm). PLS-DA: partial least squares discriminant analysis; SPA-LDA: successive projections algorithm for variable selection in linear discriminant analysis; iSPA-PLS-DA: successive projections algorithm for interval selection in partial least squares discriminant analysis; MCC: Matthew's correlation coefficient.

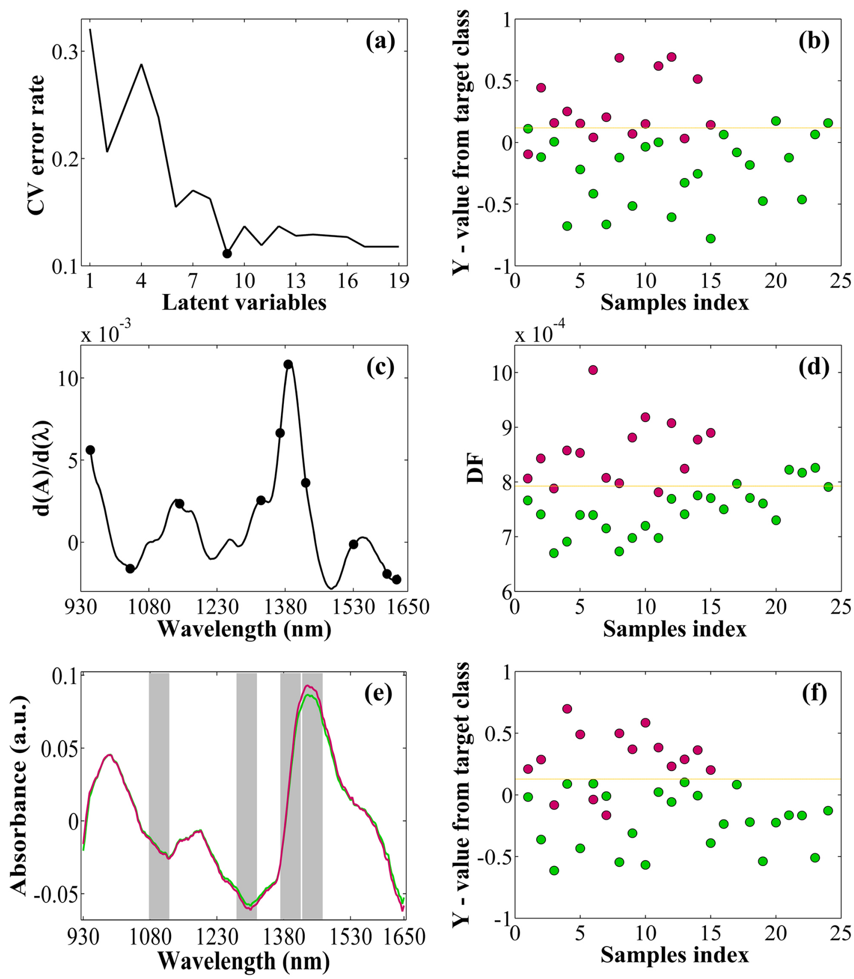

Figure 4. Best results of the classification between pure wheat flour samples (burgundy circles) and wheat flour blended with cassava starch (green circles) using the NIR spectra pre-processed with Savitzsky-Golay first derivative with second-order polynomial and 13-points window (SGD), SGD, and linear baseline correction for the construction of the PLS-DA, SPA-LDA, and iSPA-PLS-DA models, respectively. The selected number of latent variables in PLS-DA, individual variables in SPA-LDA, and intervals in iSPA-PLS-DA are shown in (a,c,e), respectively, along with their respective response plots in (b,d,f). The orange horizontal line representing the interclass boundary.

samples containing cassava starch were correctly classified in the test set. In other words, this implies sensitivities were 96.7 and 80.0%, specificities were 96.4 and 100%, efficiencies were 96.5 and 89.4%, accuracies were 96.5

and 92.3%, and MCC values were 0.92 and 0.84 in the training and test sets, respectively. For illustration and comparison, Figures 4b, 4d, and 4f show the classification results obtained by the SGD/PLS-DA, SGD/SPA-LDA, and LBC/iSPA-PLS-DA models with the projections of the pure (burgundy circles) and cassava starch-containing (green circles) wheat flour, with the orange horizontal line representing the interclass boundary.

Regarding the literature, Tao *et al.*[17] employed a handheld microNIR spectrometer (1150-2150 nm) to detect cassava flour adulteration in wheat flour at five levels of 5, 10, 20, 30 and 40% using PCA-LDA and PLS-DA. The two-class discriminant models attained predictive accuracies exceeding 95.00% in distinguishing between pure and adulterated wheat flour. In the case of six-class discriminant models, only wheat flours without cassava flour achieved 100% accuracy. Samples with additions from 5 to 40% cassava flour displayed accuracies ranging from 56.25 to 100%. Almeida and co-workers[18] used a benchtop NIR spectrometer (10000 to 4000 cm^{-1}) authenticate wheat flour for bread-making against additions of the cassava starch at eight different levels (3 to 30% (m m^{-1})) using data-driven soft independent modelling of class analogy (DD-SIMCA) as a one-class classifier. The best result was achieved using 7 principal components, reaching 95.5% sensitivity in the training and 100% sensitivity and specificity in the test set. Since the cassava starch content is higher (5 to 40%) in the study developed by Tao *et al.*,[17] the chemometric approach involving class modeling rather than discriminant analysis in the work of Almeida and co-workers[18] is different, it is not suitable to make any comparison with the proposed work. However, it is worth emphasizing the advantage of using a compact device for this purpose.

Conclusions

This study highlights the potential of a portable NIR spectrometer for analyzing commercial bread-making wheat flour blended with cassava starch, offering both quantitative and qualitative insights. The use of interval selection through SPA significantly enhanced the model performance of both types of analysis. Exploring spectral information in an automatic and unbiased manner, particularly in –OH and –NH absorption bands related to water, protein, and amylose, proved crucial for success. The robust results demonstrated effective cassava starch quantification and wheat flour classification (as containing or not containing cassava starch), following the principles of Green Analytical Chemistry. This contributes to Sustainable Development Goal 2 and aligns with the objectives of the Brazilian Bill proposal No. 5332/2009, aiming to reduce wheat imports and promote domestic production and consumption of cassava, thereby supporting family farming. The proposed method ensures non-destructive quality control for the raw material used in producing the future "Brazilian bread," providing a reliable and environmentally conscious analytical solution.

Supplementary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

Dâmaris N. C. Candeias gratefully acknowledges Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Brazil). David Douglas S. Fernandes is thankful to Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ, Brazil). Paulo Henrique G. D. Diniz also thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) (grant No. 409246/2023-9).

Author Contributions

Dâmaris N. C. Candeias was responsible for methodology, investigation, formal analysis, writing original draft; Sara Regina R. C. de Barros for methodology, investigation, visualization, writing original draft; Wellington S. Lyra for methodology, investigation, visualization, writing original draft; David Douglas S. Fernandes for software, formal analysis, validation, visualization, data curation, writing original draft; Paulo Henrique G. D. Diniz for conceptualization, data curation, supervision, funding acquisition, project administration, writing original draft, review and editing.

References

1. Vinod, B. R.; Asrey, R.; Rudra, S. G.; Urhe, S. B.; Mishra, S.; *Food Chem. Adv.* **2023**, *3*, 100473. [Crossref]
2. Dahiya, S.; Bajaj, B. K.; Kumar, A.; Tiwari, S. K.; Singh, B.; *Process Biochem*. **2020**, *99*, 290. [Crossref]
3. Badaró, A. T.; Tavares, J. P. H.; Blasco, J.; Aleixos-Borrás, N.; Barbin, D. F.; *Food Control* **2022**, *140*, 109115. [Crossref]
4. Brouns, F.; Geisslitz, S.; Shewry, P. R.; *J. Cereal Sci.* **2022**, *105*, 103447. [Crossref]
5. Begum, R., Rakshit, S. K., Rahman, S. M. M.; *Int. J. Food Prop.* **2011**, *14*, 185. [Crossref]
6. Eddy, N. O.; Udofia, P. G.; Eyo, D.; *Afr. J. Biotechnol.* **2007**, *6*, 2415. [Crossref]
7. Jensen, S.; Skibsted, L. H.; Kidmose, U.; Thybo, A. K.; *LWT* **2015**, *60*, 292. [Crossref]
8. Akintayo, O. A.; Oyeyinka, S. A.; Aziz, A. O.; Olawuyi, I. F.;

Kayode, R. M. O.; Karim, O. R.; *J. Food Sci.* **2020**, *85*, 2310. [Crossref]

9. Câmara dos Deputados, Brasil, Projeto de Lei No. 5332/2009, *Cria o "Pão Brasileiro", a Ser Produzido com Farinha de Trigo Adicionada de Farinha de Mandioca Refinada, de Farinha de Raspa de Mandioca ou de Fécula de Mandioca, Adquiridos Pelo Poder Público, e Estabelece Regime Tributário Especial Para a Farinha de Trigo Misturada, e dá Outras Providências*; Brasília-DF, 2009. [Link] accessed in May 2024
10. Companhia Nacional de Abastecimento (CONAB); *Acompanhamento da Safra Brasileira Grãos-Safra 2023/24*, 4º Levantamento; CONAB: Brasília, DF, Brazil, 2024. [Link] accessed in May 2024
11. Lin, F.; Li, X.; Jia, N.; Feng, F.; Huang, H.; Huang, J.; Fan, S.; Ciais, P.; Song, X.-P.; *Global Food Sec*. **2023**, *36*, 100661. [Crossref]
12. Behnassi, M., El Haiba, M.; *Nat. Hum. Behav.* **2022**, *6*, 754. [Crossref]
13. Liu, H.-Y.; Wadood, S. A.; Xia, Y.; Liu, Y.; Guo, H.; Guo, B.-L.; Gan, R.-Y.; *Crit. Rev. Food Sci. Nutr.* **2023**, *63*, 33. [Crossref]
14. Zhang, S.; Liu, S.; Shen, L.; Chen, S.; He, L.; Liu, A.; *Curr. Res. Food Sci.* **2022**, *5*, 1305. [Crossref]
15. Du, Z.; Tian, W.; Tilley, M.; Wang, D.; Zhang, G.; Li, Y.; *Compr. Rev. Food Sci. Food Saf.* **2022**, *21*, 2956. [Crossref]
16. Su, W.-H.; Sun, D.-W.; *J. Food Eng.* **2017**, *200*, 59. [Crossref]
17. Tao, F.; Liu, L.; Kucha, C.; Ngadi, M.; *Biosyst. Eng.* **2021**, *203*, 34. [Crossref]
18. Duarte, E. S. A.; Almeida, V. E.; Costa, G. B.; Araújo, M. C. U.; Véras, G.; Diniz, P. H. G. D.; Fernandes, D. D. S.; *Food Chem.* **2022**, *368*, 130843. [Crossref]
19. Giussani, B.; Gorla, G.; Riu, J.; *Crit. Rev. Anal. Chem.* **2024**, *54*, 11. [Crossref]
20. Gopal, J.; Muthu, M.; *Trends Anal. Chem.* **2024**, *171*, 117504. [Crossref]
21. Yan, H.; Neves, M. D. G.; Noda, I.; Guedes, G. M.; Ferreira, A. C. S.; Pfeifer, F.; Chen, X.; Siesler, H. W.; *Chemosensors* **2023**, *11*, 272. [Crossref]
22. Gullifa, G.; Barone, L.; Papa, E.; Giuffrida, A.; Materazzi, S.; Risoluti, R.; *Front. Chem.* **2023**, *11*, 1214825. [Crossref]
23. Beć, K. B.; Grabska, J.; Huck, C. W.; *Foods* **2022**, *11*, 1465. [Crossref]
24. Vallese, F. D.; Paoloni, S. G.; Springer, V.; Fernandes, D. D. S.; Diniz, P. H. G. D.; Pistonesi, M. F.; *J. Food Comp. Anal.* **2024**, *126*, 105925. [Crossref]
25. Gomes, A. A.; Azcarate, S. M.; Diniz, P. H. G. D.; Fernandes, D. D. S.; Veras, G.; *Food Chem.* **2022**, *370*, 131072. [Crossref]
26. Oliveri, P.; Downey, G.; *Trends Anal. Chem.* **2012**, *35*, 74. [Crossref]
27. *MATLAB*, version R2018b, Natick, MA, USA: MathWorks, 2018.

Submitted: January 17, 2024
Published online: May 29, 2024