






DETERMINATION OF QUALITY ATTRIBUTES AND RIPENING STAGE USING VIS-NIR SPECTROSCOPY IN INTACT SERIGUELA AND UMBU FRUITS

Patricia de Araujo Souza¹ , Iara Jeanice Souza Ferreira²  & Daniel dos Santos Costa^{3*} 

1 - Federal University of São Francisco Valley, Petrolina, Pernambuco, Brazil

2 - State University of Campinas, College of Agricultural Engineering, Campinas, São Paulo, Brazil

3 - Federal University of São Francisco Valley, College of Agricultural and Environmental Engineering, Juazeiro, Bahia, Brazil

Keywords:

Spondias purpurea L
Spondias tuberosa
Total soluble solids
Firmness
Non-destructive methods

ABSTRACT

To produce seriguela and umbu on a large scale, it is important to detect the ripening stages and quality attributes of the fruits, to define the ideal harvest point. Thus, this study aimed to determine, in a non-destructive way, the quality attributes and ripening stages of intact seriguela and umbu fruits using Vis-NIR spectroscopy. A total of 150 seriguela fruits and 150 umbu fruits were used, at different ripening stages, and subjected to spectral analysis and reference laboratory testing to determine total soluble solids (TSS) and firmness. Spectral data were subjected to different pre-processing techniques. Regression and classification models were created through the statistical learning and machine learning methods. The models with the best performance for TSS were RF ($R^2_p = 0.94$) and PLSR ($R^2_p = 0.68$), and for firmness were PLSR ($R^2_p = 0.92$) and RF ($R^2_p = 0.58$), for seriguela and umbu, respectively. The model with the best performance in the classification was LDA, with a precision greater than 95% to discriminate the ripening stages of both fruits. Therefore, the Vis-NIR spectroscopy is a potential tool to determine the quality attributes and ripening stages, in a non-destructive way, of intact seriguela and umbu fruits.

Palavras-chave:

Spondias purpurea L
Spondias tuberosa
Sólidos solúveis totais
Firmeza
Métodos não destrutivos

DETERMINAÇÃO DE ATRIBUTOS DE QUALIDADE E ESTÁDIO DE MATURAÇÃO COM USO DA ESPECTROSCOPIA VIS-NIR EM FRUTOS INTACTOS DE SERIGUELA E UMBU

RESUMO

Para a produção de seriguela e umbu em larga escala é importante detectar os estádios de maturação e os atributos de qualidade dos frutos, a fim de definir o ponto ideal de colheita. Portanto, o objetivo do trabalho foi determinar, de forma não destrutiva, os atributos de qualidade e estádios de maturação de frutos intactos de seriguela e umbu com uso da espectroscopia Vis-NIR. Foram utilizados 150 frutos de seriguela e 150 frutos de umbu, em diferentes estádios de maturação, sendo esses submetidos às análises espectral e laboratorial de referência para determinação de sólidos solúveis totais (SST) e firmeza. Os dados espectrais foram submetidos a diferentes técnicas de pré-processamento. Foram desenvolvidos modelos de regressão e classificação a partir de métodos de aprendizagem estatística e de aprendizagem de máquina. Os modelos com melhor desempenho para SST foram RF ($R^2_p = 0,94$) e PLSR ($R^2_p = 0,68$), e para a firmeza foram o PLSR ($R^2_p = 0,92$) e RF ($R^2_p = 0,58$), para seriguela e umbu, respectivamente. O modelo de melhor desempenho na classificação foi o LDA com precisão maior que 95% para discriminar os estádios de amadurecimento de ambos os frutos. Portanto, a espectroscopia Vis-NIR é uma ferramenta potencial para determinar atributos de qualidade e estádios de maturação, de forma não destrutiva, em frutos intactos de seriguela e umbu.

INTRODUCTION

Brazil stands out in the production and export of tropical fruits, having a large diversity of fruit species (SOUZA *et al.*, 2016). However, there are exotic species that are little known, but have great market potential, such as the seriguela (*Spondias purpurea* L) and umbu (*Spondias tuberosa*) (SANCHES *et al.*, 2018). Seriguela and umbu fruits are mainly consumed *in natura*, but they can be used as raw material for processed products in agroindustry (SANTOS & OLIVEIRA, 2008).

One of the main difficulties regarding the growth and commercial development of the seriguela and umbu crops is the harvest and postharvest of the fruits (LIRA JÚNIOR *et al.*, 2010; BRENO *et al.*, 2012). These fruits are highly perishable during postharvest handling, being prone to softening and quickly reaching senescence (NERIS *et al.*, 2017; COSTA *et al.*, 2021). The harvest of seriguela fruits must be performed when they finish the physiological ripening stage and begin the maturing process (MALDONADO-ASTUDILLO *et al.*, 2014). On the other hand, the harvest of umbu fruits must be conducted when they are at the physiological ripening stage (LIMA *et al.*, 2018).

Therefore, it is extremely important to evaluate the fruits ripening parameters as indicators of the ideal harvest point, so that a greater use of production and a better acceptance of the product in the market can be achieved (LIMA & CASTRICINI, 2019). Nevertheless, these quality indices are normally determined by analytical methods that are destructive and time-consuming. Moreover, they require reagents, and cannot be carried out on many samples (NERIS *et al.*, 2017).

Hence, non-destructive techniques, such as Vis-NIR (visible near-infrared) spectroscopy, are an alternative for the rapid and non-destructive determination of quality attributes of fruits *in natura* (GABRIËLS *et al.*, 2020). Studies using Vis-NIR spectroscopy have shown satisfactory results in predicting quality attributes in numerous fresh tropical fruits, such as mangoes (SANTOS NETO *et al.*, 2017), oranges (CAYUELA & WEILAND, 2010; TORRES *et al.*, 2017), grapes (COSTA *et al.*, 2019), and acerola cherries (MORAES *et al.*, 2019). However, there are no specific studies using Vis-NIR spectroscopy to determine the quality

attributes and ripening stages of seriguela and umbu fruits.

As such, this study aimed to determine, in a non-destructive way, the quality attributes and ripening stages of intact seriguela and umbu fruits using Vis-NIR spectroscopy.

MATERIAL AND METHODS

Acquisition of fruits

A total of 150 seriguela fruits and 150 umbu fruits from plants located in the municipality of Juazeiro (Juazeiro, Bahia, Brazil: Latitude: 9° 26' 18" S; Longitude: 40° 30' 19" W), and under three different ripening stages were used: 50 fruits at the green ripening stage; 50 fruits at the intermediate ripening stage; and 50 fruits at the full ripening stage. The region is characterized by a BSh climate, according to the Köppen climate classification (ALVARES *et al.*, 2013), corresponding to an semiarid climate region. The division of fruits into three ripening stages was conducted based on the ripening stages suggested by Sampaio *et al.* (2008) and Narain *et al.* (1992) for seriguela and umbu, respectively.

Acquisition of reflectance spectra

After being harvested and separated into stages, the fruits were subjected to spectral analysis in a FieldSpec 3 spectrometer (Analytical Spectral Devices, Boulder, Colorado, USA), which has a spectral range from 350 to 2500 nm, resolution of 8 nm, reading speed of 100 ms, precision of ± 1 nm, InGaAs photodiode array, and 50 W quartz-tungsten-halogen light source. The reflectance spectra obtained were transformed into absorbance values. The region between ranges from 350 to 460 nm and from 2,060 to 2,500 nm was removed due to the excess of random noise.

Determination of postharvest quality attributes

The physicochemical quality attributes total soluble solids (TSS) and firmness were evaluated by analytical reference methods. TSS were determined using a digital refractometer (Invert Sugar Refractometer, HI 96804 Hanna), with values expressed as °Brix. Firmness was determined using a FR-5120 digital penetrometer (Rex Durometer), with values expressed as N (Newton).

Development of predictive models

Figure 1 summarizes the methods and procedures conducted in this study for the prediction of quality attributes and classification of the fruits ripening stages. After the fruits were harvested, they were subjected to acquisition of spectra and determination of physicochemical attributes. These data were used to develop predictive models.

Spectral data contain background and noise information, in addition to information regarding the samples. Hence, spectral data were subjected to different combinations of pre-processing for different purposes. To increase the signal/noise ratio, smoothing with a moving average filter (FMM) was applied to the spectral data in various segment sizes (7, 15, 25, and 35 points). To remove baseline displacement and inclination, the first and second Savitzky-Golay derivatives (GORRY, 1990; SAVITZKY & GOLAY, 1964) were applied to the smoothed data, with a second-order polynomial in various segment sizes (7,

15, 25, and 35 points). A multiplicative scatter correction (MSC) (ISAKSSON & NAES, 1988) and a standard normal variate (SNV) (BARNES *et al.*, 1989) were applied to the smoothed data and derivatives to remove the effect of light scattering. Finally, an orthogonal signal correction (OSC) was applied to the smoothed spectral data to remove part of the variation that has low correlation with the variable of interest (WOLD *et al.*, 1998).

Principal Component Analysis (PCA) was applied to the data to select samples with higher explained variance, which composed the calibration set, while the other samples composed the prediction set. The seriguela and umbu datasets were divided into 2/3 and 1/3 ratios, that is, 2/3 of the data were destined for calibration and cross-validation, and 1/3 for prediction.

Regression models were built for the quality attributes (TSS and firmness), and classification models were built for the ripening stages (green, intermediate, and full ripe). For this purpose,

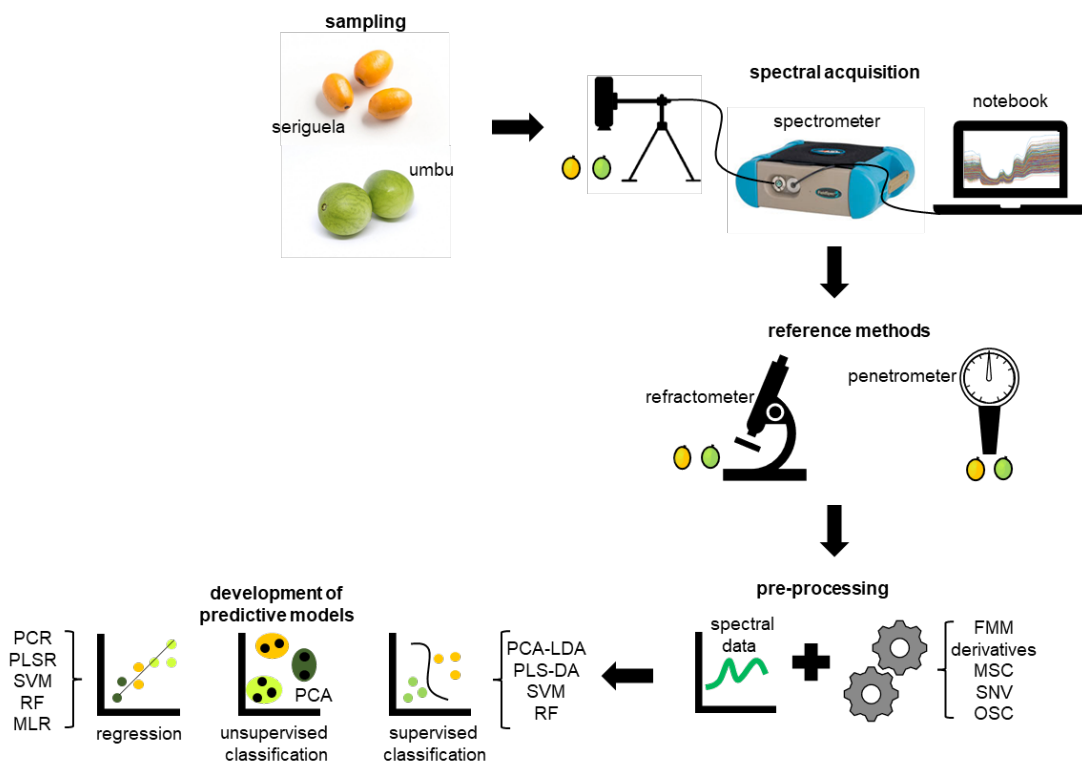


Figure 1. Experimental steps to build prediction and classification models. PCR: principal component regression; PLSR: partial least squares regression; SVM: support vector machine; RF: random forest; MLR: multiple linear regression; PCA: principal component analysis; PCA-LDA: principal component analysis - linear discriminant analysis; PLS-DA: partial least squares discriminant analysis; FMM: moving average filter; MSC: multiplicative scatter correction; SNV: standard normal variate; OSC: orthogonal signal correction

original or pre-processed spectral data were used as independent variables (X), whereas the quality attributes or ripening stages of the fruits were used as dependent variables (Y) (WOLD *et al.*, 1984).

Statistical learning methods were used to develop the regression models, namely the Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR), while the machine learning methods used were the Support Vector Machine (SVM) and Random Forest (RF). The NIPALS (Non-linear Iterative Partial Least-Squares) algorithm (GELADI & KOWALSKI, 1986) consisting of 100 iterations was used to build the PCR and PLSR models. The Sequential Minimal Optimization (SMO) algorithm was used to build the SVM models, and random trees consisting of 100 iterations were used to build the RF models.

To solve data matrix collinearity problems and reduce the number of input spectral variables to enable the development of Multiple Linear Regression (MLR) models, a variable selection (wavelengths) was performed from the graphic analysis of the spectral variable weights of the attributes TSS and firmness in the PLSR and PCR models. These selected spectral variables were used as input to the MLR models. After the MLR models construction, the selected variables were subjected to an analysis of variance (ANOVA), and the models were reconstructed with the significant variables (p -value ≤ 0.05). In addition, from the SVM and RF models developed with the original and pre-processed spectra, a variable selection was conducted with the Correlation-based Feature Selection (Cfs) filter and subsequent reconstruction of these models with the selected variables.

The regression model performance was evaluated using different statistical parameters, such as: Pearson's coefficient of determination (R^2); root mean square error of calibration (RMSEC); standard error of calibration (SEC); root mean square error of cross-validation (RMSECV); standard error of cross-validation (SECV); root mean square error of prediction (RMSEP); standard error of prediction (SEP); and bias. Two parameters were used to choose the best models: Pearson's coefficient of determination (R^2_{cv}), and standard error, both from cross-validation (SECV)

(GÓMEZ *et al.*, 2006). The statistical parameters are defined by Equations (1) to (8):

$$R^2(\text{Pearson}) = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})^2}{(n-1)\sigma_r\sigma_p} \quad (1)$$

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$SEC = \sqrt{\frac{n RMSEC^2}{n-1}} \quad (3)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{i*})^2}{n}} \quad (4)$$

$$SECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i - bias)^2}{n-1}} \quad (5)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{m}} \quad (6)$$

$$SEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i - bias)^2}{m-1}} \quad (7)$$

$$bias = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{m} \quad (8)$$

Where:

\hat{y}_i = value estimated by the calibration model;

\hat{y}_{i*} = value estimated by the model in the cross-validation stage;

y_i = reference value;

\bar{y} = average of reference values;

$\bar{\hat{y}}$ = average of predicted values;

n = number of samples at the calibration or validation stages;

m = number of predicted samples;

σ_r = standard deviation of reference values;

σ_p = standard deviation of predicted values; and

S_r^2 = variation of reference values.

PCA was applied as an exploratory data analysis method to verify the possibility of clustering samples in relation to the ripening stages. The predictor variables were specific regions of the spectrum, namely the visible and near-infrared region (350-1,000 nm), first short-wave near-

infrared region (1,001-1,830 nm), and second short-wave near-infrared region (1,831-2,500 nm). The classification models were developed based on the formation of clusters.

The Principal Component Analysis - Linear Discriminant Analysis (PCA-LDA), Partial Least Squares Discriminant Analysis (PLS-DA), Support Vector Machine (SVM), and Random Forest (RF) methods were used to develop the classification models. The performance of the classification models was evaluated using the confusion matrix for three ripening stages (green, intermediate, and full ripe). From the confusion matrix, precision, or global accuracy (GAC), sensitivity (SEN), selectivity (SEL), and false-positive rate (FPR) values were calculated using Equations (9) to (12):

$$GAC = \frac{TP+TN}{TP+FN+FP+TN} 100 \quad (9)$$

$$SEL = \frac{TN}{FP+TN} 100 \quad (10)$$

$$SEN = \frac{TP}{TP+FN} 100 \quad (11)$$

$$FPR = \frac{FP}{FP+TN} 100 \quad (12)$$

Where TP, TN, FP, and FN correspond to the true-positive, true-negative, false-positive, and false-negative values, respectively.

The construction of the regression models through the PLSR and PCR methods, as well as the construction of the classification models through the LDA and PLS-DA methods, was carried out using the Unscrambler X 10.4 software (CAMO ASA, Oslo, Norway). The construction of the

regression and classification models through the SVM and RF methods was performed using the Weka 3.8.4 software (University of Waikato, New Zealand).

RESULTS AND DISCUSSION

Descriptive statistics

Table 1 indicates the descriptive statistics of the quality attributes for the calibration and prediction datasets in seriguela and umbu fruits. The coefficients of variation (CV) of the quality attributes were high, except for the TSS of umbu fruits, which showed CV of 11.90% and 10.15% for calibration and prediction, respectively. CV are important variability indicators of the dataset. The high variation in data enables the development of robust predictive models.

Figure 2 shows the behavior of the attributes TSS and firmness of seriguela and umbu fruits throughout the maturing process. The TSS content increases during the maturing process of both fruits. This increase in TSS content can be attributed to the hydrolysis of polysaccharides, such as starch, and their transformation into simple sugars (SAMPAIO *et al.*, 2008). Nevertheless, seriguela fruits had a sharp increase for TSS after the intermediate stage, whereas umbu fruits showed a small increase. This behavior occurs due to the increase in respiration rate during the maturing process of seriguela fruits, while in umbu fruits this does not occur (SAMPAIO *et al.*, 2008; MENEZES *et al.*, 2017; SILVA *et al.*, 2020). The increased respiration causes the events of increase in TSS and reduction between the acid-sugar balance (MONTALVO-GONZÁLEZ *et al.*, 2011).

Table 1. Descriptive statistics of the quality attributes TSS and firmness for the calibration and prediction sets in seriguela and umbu fruits

Quality attributes	Calibration					Prediction				
	Mean	Maximum	Minimum	SD	CV (%)	Mean	Maximum	Minimum	SD	CV (%)
Seriguela										
TSS (°Brix)	11.69	24.60	4.50	5.78	49.44	11.76	21.60	4.50	5.31	45.14
Firmness (N)	35.33	79.35	2.15	20.54	58.14	33.26	76.00	2.75	20.28	60.95
Umbu										
TSS (°Brix)	11.49	14.30	7.80	1.37	11.90	11.55	13.70	8.30	1.17	10.15
Firmness (N)	16.94	60.85	0.35	14.69	86.70	18.44	56.20	0.50	13.10	71.07

TSS = Total Soluble Solids; SD = standard deviation; CV = coefficient of variation

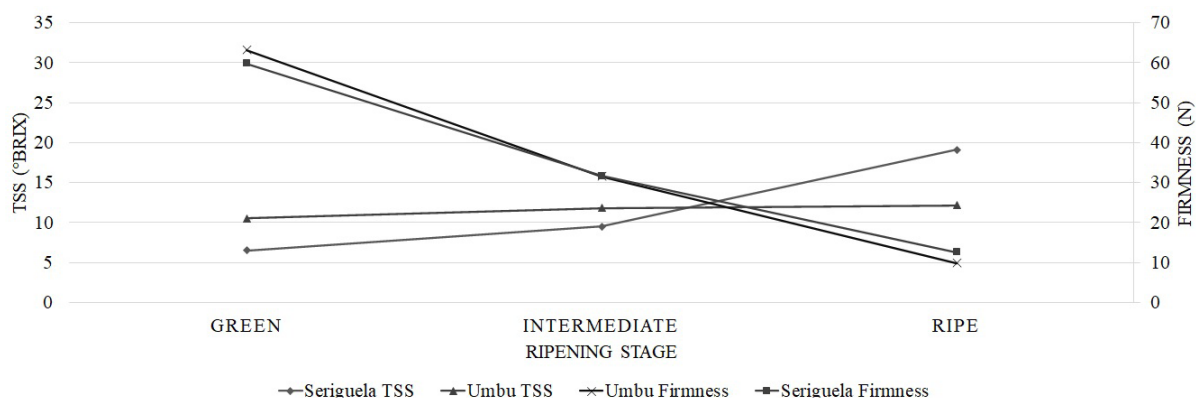


Figure 2. Behavior of TSS and firmness in seriguela and umbu fruits throughout the maturing process

During the same process, it is also possible to observe a loss of firmness. That occurs due to structural and biochemical alterations in pectin, hemicelluloses, and celluloses which constitute the cell wall. These cellular changes include the solubilization and depolymerization of cell wall polysaccharides, and alterations in their arrangement patterns, causing fruit softening (SAMPAIO *et al.*, 2008).

Exploratory analysis of the spectrum

The absorbance spectra for the three ripening stages of seriguela are indicated in Figure 3A. A distinct behavior in the visible range from 544 to 629 nm in the spectral curves of the three ripening stages was observed. This characteristic is due to the metabolism of chlorophylls and carotenoids (MALDONADO-ASTUDILLO *et al.*, 2014).

Seriguela fruits exhibit a sequence of changes in the pericarp color during the maturing process, consisting of three phases: phase I from dark green to light green, phase II from light green to orange-yellow, and phase III from orange-yellow to purple-red. During these ripening stages, there is a reduction in chlorophyll content and a continuous increase in carotenoids content. Therefore, the loss of green color in fruits is due to the chlorophyll structure degradation, and the appearance of orange and red colors is due to the biosynthesis of carotenoids (SAMPAIO *et al.*, 2008).

An absorption peak was reported in the region from 461 to 500 nm, which corresponds to the spectral absorption region of carotenoids

(BRENN & BERARDO, 2004), mainly in ripe seriguela fruits. Moreover, Figure 3A shows that, in the wavelength range from 629 to 713 nm, the spectral behavior at the green stage was like the intermediate stage, differing from the spectral behavior at the full ripe stage. This occurs because fruits at the intermediate ripening stage still have chlorophyll content, similarly to the green stage in this spectral region. This, in turn, does not occur in fruits that are in the final maturing process, in which there was a complete degradation of chlorophyll (SAMPAIO *et al.*, 2008).

Figure 3B indicates the absorbance spectra for the three ripening stages of umbu. An absorption peak was observed in the region around 647 nm for the three ripening stages, and there were no differences in the shape of the spectral curves. This behavior occurs due to characteristics of the umbu maturing process, in which three phases can be identified: phase I, characterized by a hard texture and green color, phase II, characterized by a firm texture and light green color, and phase III, characterized by a soft texture and yellowish light green color (NARAIN *et al.*, 1992), in a way that a complete degradation of chlorophyll does not occur in ripe fruits.

Similar absorption peaks can be observed in Figure 3A and 3B in the near-infrared region for the seriguela and umbu fruits. There were absorption peaks around 980 nm, which corresponds to the water absorption spectral range, and around 1,200 nm and 1,450 nm, corresponding to the lengthening of C – H bonds in CH₂ of carbohydrates due to the second harmonic (MOGHIMI *et al.*, 2010).

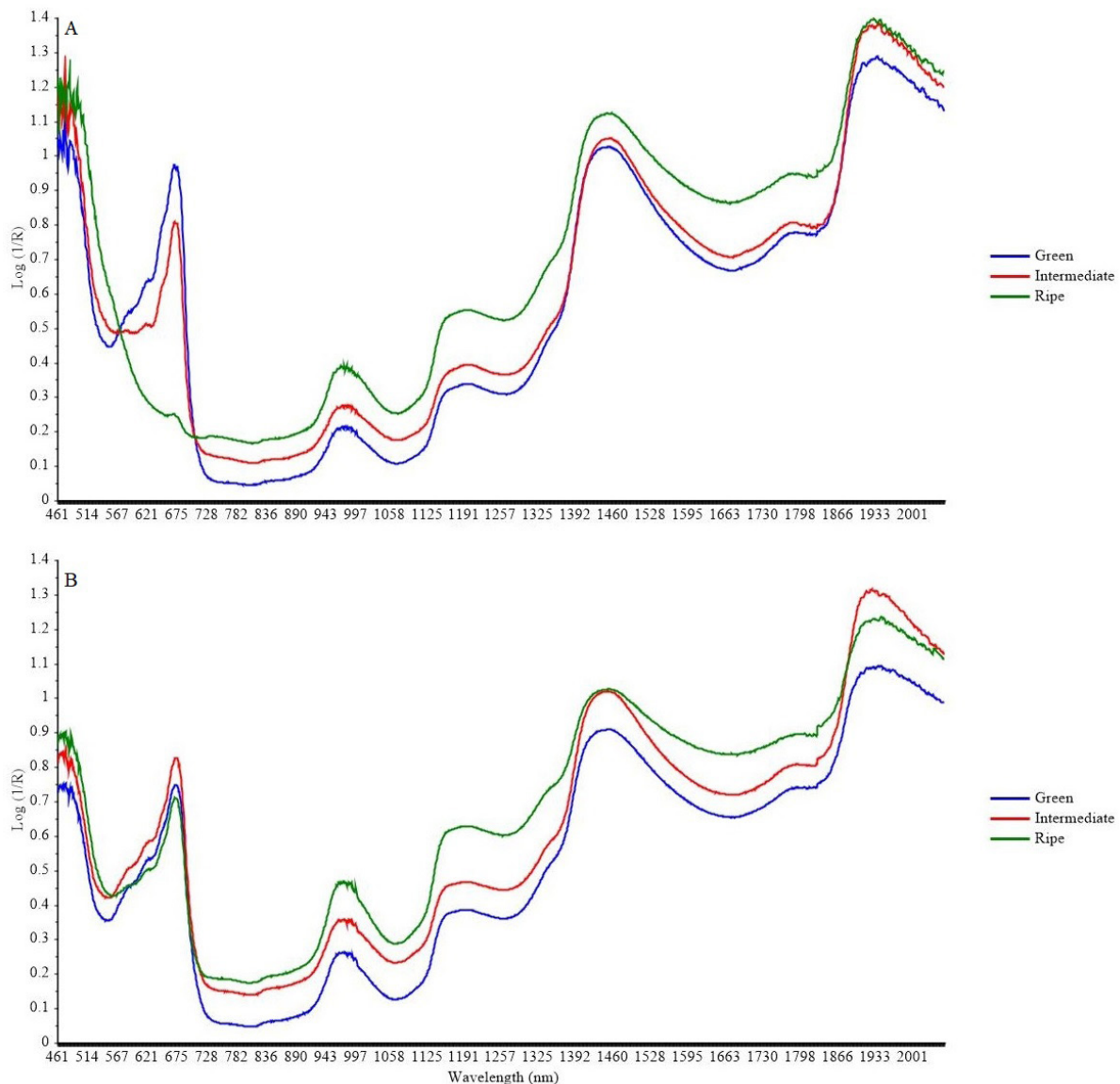


Figure 3. Raw mean absorbance spectra of (A) seriguelas; and (B) umbus for each evaluated ripening stage (green, intermediate, and full ripe)

Predictive models of the quality attributes

Full-spectrum regression models and Cfs of variables

For seriguela fruits, TSS were determined using the RF model associated with OSC (35 MAF) and Cfs, with R^2 of 0.96 and 0.94 for cross-validation and prediction, respectively. Firmness was determined using the PLSR model associated with OSC (35 MAF), with R^2 of 0.94 and 0.92 for cross-validation and prediction, respectively. SECV of 1.22 °Brix and 4.86 N, and SEP of 1.34 °Brix and 5.78 N were obtained for TSS and firmness, respectively.

For umbu fruits, TSS were determined using the PLSR model associated with OSC (35 MAF), with R^2 of 0.67 and 0.68 for cross-validation and prediction, respectively. Firmness was determined using the RF model associated with OSC (35 MAF), with R^2 of 0.74 and 0.58 for cross-validation and prediction, respectively. SECV of 0.80 °Brix and 7.47 N, and SEP of 0.68 °Brix and 8.57 N were obtained for TSS and firmness, respectively (Table 2).

The association of regressions with the orthogonal signal correction (OSC) provided a satisfactory performance of the mathematical models, both in calibration and prediction, reducing the modeling complexity. This behavior occurred

Table 2. Statistical learning and machine learning predictive models for TSS and firmness of intact seriguela and umbu fruits using the full spectrum (461-2059 nm) and variable selection

Quality attributes	Model	Pre-processing	Factor	Calibration			Cross-validation			Prediction		
				R ² c	RMSEC	SEC	R ² cv	RMSECV	SECV	R ² p	RMSEP	SEP
Seriguela												
TSS (°Brix)	PLSR	OSC (35 MAF)	1	0.96	1.13	1.14	0.96	1.16	1.16	0.92	1.52	1.54
	PCR	OSC (35 MAF)	1	0.96	1.13	1.14	0.96	1.16	1.16	0.92	1.52	1.53
	SVM	Raw + Cfs	*	0.95	1.26	1.26	0.96	1.29	1.29	0.93	1.41	1.42
	RFR	OSC (35 MAF) + Cfs	*	0.99	0.48	0.48	0.96	1.22	1.22	0.94	1.38	1.34
Firmness (N)	PLSR	OSC (35 MAF)	1	0.95	4.74	4.77	0.94	4.83	4.86	0.92	5.90	5.78
	PCR	OSC (35 MAF)	1	0.95	4.74	4.77	0.94	4.83	4.86	0.92	5.90	5.78
	SVM	raw + Cfs	*	0.92	5.98	6.01	0.91	6.08	6.07	0.92	6.17	5.89
	RFR	OSC (35 MAF) + Cfs	*	0.99	1.88	1.89	0.94	5.07	5.08	0.92	5.77	5.71
Umbu												
TSS (°Brix)	PLSR	OSC (35 MAF)	7	0.83	0.56	0.57	0.67	0.79	0.80	0.68	2.24	0.68
	PCR	OSC (35 MAF)	1	0.48	0.98	0.98	0.46	1.00	1.01	0.44	1.89	0.88
	SVM	OSC (35 MAF)	*	0.87	0.49	0.49	0.57	0.91	0.91	0.58	0.91	0.87
	RFR	OSC (35 MF)	*	0.94	0.38	0.38	0.43	1.04	1.05	0.44	0.89	0.89
Firmness (N)	PLSR	OSC (35 MAF)	1	0.72	7.80	7.84	0.70	7.97	8.01	0.59	24.63	8.49
	PCR	OSC (35 MAF)	1	0.72	7.80	7.84	0.70	7.97	8.01	0.59	24.63	8.49
	SVM	Raw	*	0.79	6.85	6.88	0.62	9.03	9.07	0.59	8.42	8.49
	RFR	OSC (35 MAF)	*	0.96	2.98	2.99	0.74	7.44	7.47	0.58	8.49	8.57

R²c = Coefficient of determination of Calibration; R²cv = Coefficient of determination of Cross-validation; R²p = Coefficient of determination of Prediction; RMSEC = Root Mean Square Error of Calibration; SEC = Standard Error of Calibration; RMSECV = Root Mean Square Error of Cross-validation; SECV = Standard Error of Cross-validation; RMSEP = Root Mean Square Error of Prediction; SEP = Standard Error of Prediction

because the OSC removed spectral signals that are not related to the quality attributes (RAMBO *et al.*, 2013; HEMRATTRAKUN *et al.*, 2021).

The statistical learning models (PLSR and PCR) performed similarly to the machine learning models (SVM and RF). However, the PLSR and PCR models have a lower computational complexity and cost than the SVM and RF models, while the predictive modeling is more efficient in this case (WOLD *et al.*, 1984). Thus, predictive models for determining the quality attributes of intact seriguela and umbu fruits using Vis-NIR spectroscopy in the spectral range from 461 to 2,059 nm are robust.

MLR models with variable selection

From the graphic analysis of spectral variable weights, provided by the PLSR model, a probable association between specific wavelengths and quality attributes was observed (Figure 4). From

the loading curves of the independent variables and ANOVA, with a *p*-value of less than 5%, the following wavelengths were selected: 471, 520, 1400, 1423, 1426, 1456, 1465, 1660 nm for TSS in seriguela; 977, 1009, 1016, 1110, 1133, 1189, 1329, 1522, 1543, 1679, 1809, 1846 nm for firmness in seriguela; 645, 778, 956, 963, 1043, 1198, 1315, 1472, 1594, 1733, 1788, 1816, 1849, 1884, 1954, 1989, 1990, 2025 nm for TSS in umbu; and 470, 504, 644, 971, 980, 1143, 1525, 1652, 1667, 1750 e 1953 nm for firmness in umbu.

For TSS, in seriguela and umbu fruits, the wavelengths in the visible region were related to photosynthetic pigments (MA *et al.*, 2021), and in the near-infrared region, the spectral signatures were associated with the lengthening of O – H bonds of water, organic acids, and sugars due to the third OH overtone (GODDU & DELKER, 1960; MOGHIMI *et al.*, 2010).

For firmness, in seriguela and umbu fruits, the

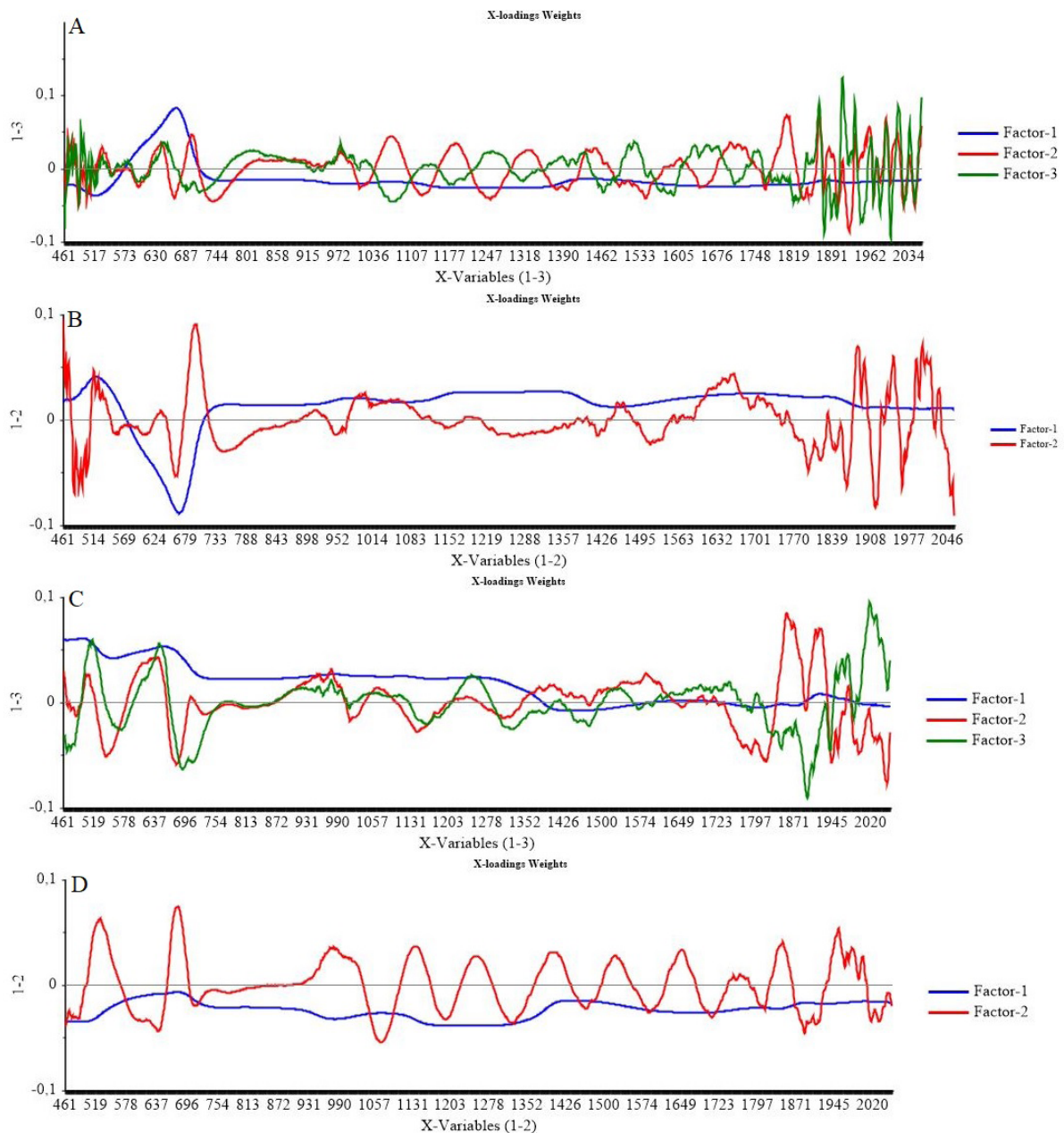


Figure 4. Loading plots concerning spectral variables of seriguela fruits for (A) TSS; (B) firmness; and umbu fruits for (C) TSS; and (D) firmness

wavelengths in the near-infrared are related to the lengthening of C – H covalent bonds of the CH₂ group due to the first and second harmonics (FU *et al.*, 2008). Fruit firmness is a cohesive force measurement between the pectin, which are substances present in the middle lamellas, and exert a cementing function between the cell walls. As maturing occurs, these compounds are degraded, causing fruit softening (ENGELSEN *et al.*, 1998). Confirming these results, Bizzani *et al.* (2017) reported that pectin is responsive to energy emitted in some spectral regions in the range from

1,000 to 2,500 nm (near-infrared region).

Table 3 shows the performance parameters of the MLR models based on the spectral signatures selected for the quality attributes. Regarding seriguela fruits, both for TSS and firmness, the MLR models reached R² above 0.94 and 0.91 for cross-validation and prediction, respectively. SECV of 1.11 °Brix and 4.89 N, and SEP of 1.52 °Brix and 6.04 N were obtained for TSS and firmness, respectively. For umbu fruits, the MLR models achieved performance for TSS and firmness with R² above 0.63 in cross-validation, and above 0.55

in prediction. SECV of 0.84 °Brix and 7.62 N, and SEP of 0.82 °Brix and 8.52 N were obtained for TSS and firmness, respectively.

These results were satisfactory and are like those found by Li *et al.* (2013), who developed PLSR prediction models with R² between 0.70 and 0.87 to quantify TSS and firmness in pears, and Costa *et al.* (2019), who developed MLR models with R² greater than 0.90 to quantify TSS in grapes, both using Vis-NIR spectroscopy.

Variable selection was important to make the modeling process less complex, and increase the prediction performance using more specific and relevant wavelengths for the characterization of physicochemical parameters (LUO *et al.*, 2012).

Classification models of ripening stages

Principal component analysis (PCA)

The pretreatment that provided the best results for PCA was the OSC (35 MAF), explaining 100% of spectral variability in the first principal component for the seriguela and umbu fruits

(Figure 5). The existence of an overlap is observed between the intermediate ripe fruits, and the green and full ripe fruits. That occurs due to a gradual change in pericarp color of seriguelas during the ripening process. Fruits at the intermediate ripening stage are in a physiological phase between green and ripe, and their biochemical and cellular constituents are in a transition state (VLAIC *et al.*, 2018), which makes it difficult for samples at the intermediate ripening stage to be clustered into a well-defined class.

PCA in the visible region (Figure 6A) reveals the same explained variance as PCA for the full spectrum in the first two principal components. For the ranges of near-infrared 1 and 2 (Figure 6B and 6C), in turn, there was a reduction in the explained spectral variation in comparison with PCA for the full spectrum. This result reinforces the importance of the visible spectral region for discriminating the ripening stages of seriguelas. For umbu fruits, PCA in the near-infrared region 1 showed the highest explained variance (Figure 6E), when compared

Table 3. Multiple linear regression models for the quality attributes TSS and firmness of seriguela and umbu fruits using selected spectral signatures

Quality attributes	Model	Pre-processing	Calibration			Cross-validation			Prediction		
			R ² c	RMSEC	SEC	R ² cv	RMSECV	SECV	R ² p	RMSEP	SEP
Seriguela											
TSS (°Brix)	MLR	OSC (35 MAF)	0.97	1.05	1.00	0.96	1.11	1.11	0.92	1.51	1.52
Firmness (N)	MLR	OSC (35 MAF)	0.96	4.58	4.29	0.94	4.89	4.89	0.91	6.09	6.04
Umbu											
TSS (°Brix)	MLR	OSC (35 MAF)	0.75	0.75	0.68	0.63	0.84	0.84	0.55	2.10	0.82
Firmness (N)	MLR	OSC (35 MAF)	0.78	7.24	6.82	0.73	7.62	7.62	0.58	32.59	8.52

R² = Coefficient of determination of Calibration; R²_{cv} = Coefficient of determination of Cross-validation; R²_p = Coefficient of determination of Prediction; RMSEC = Root Mean Square Error of Calibration; SEC = Standard Error of Calibration; RMSECV = Root Mean Square Error of Cross-validation; SECV = Standard Error of Cross-validation; RMSEP = Root Mean Square Error of Prediction; SEP = Standard Error of Prediction

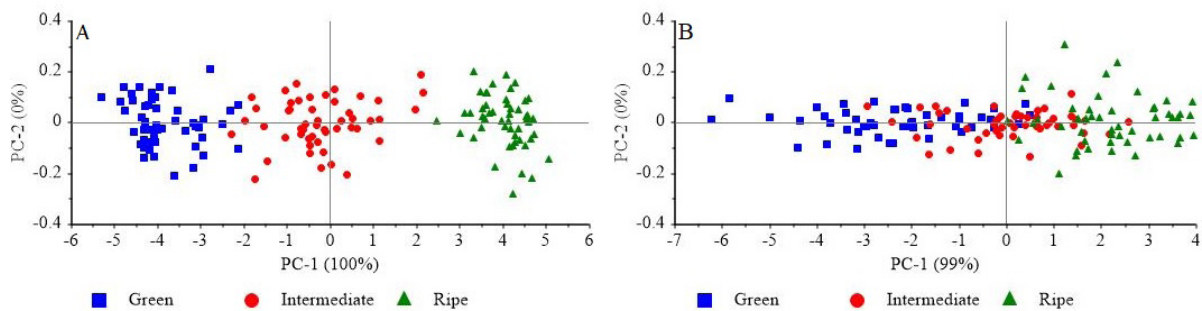


Figure 5. Principal component analysis of the full spectrum for the ripening stages in (A) seriguela and (B) umbu

to other spectral regions (Figure 6D and 6F). This result is due to biochemical reactions throughout the ripening stages of the fruits.

Supervised classification

Table 4 shows the best supervised classification

models for the discrimination of ripening stages in intact seriguela and umbu fruits. The LDA model provided the best discrimination of the three ripening stages with a precision of 97.78% and 95.31%, and false-positive rate of 1.67% and 3.32%, for seriguela and umbu, respectively.

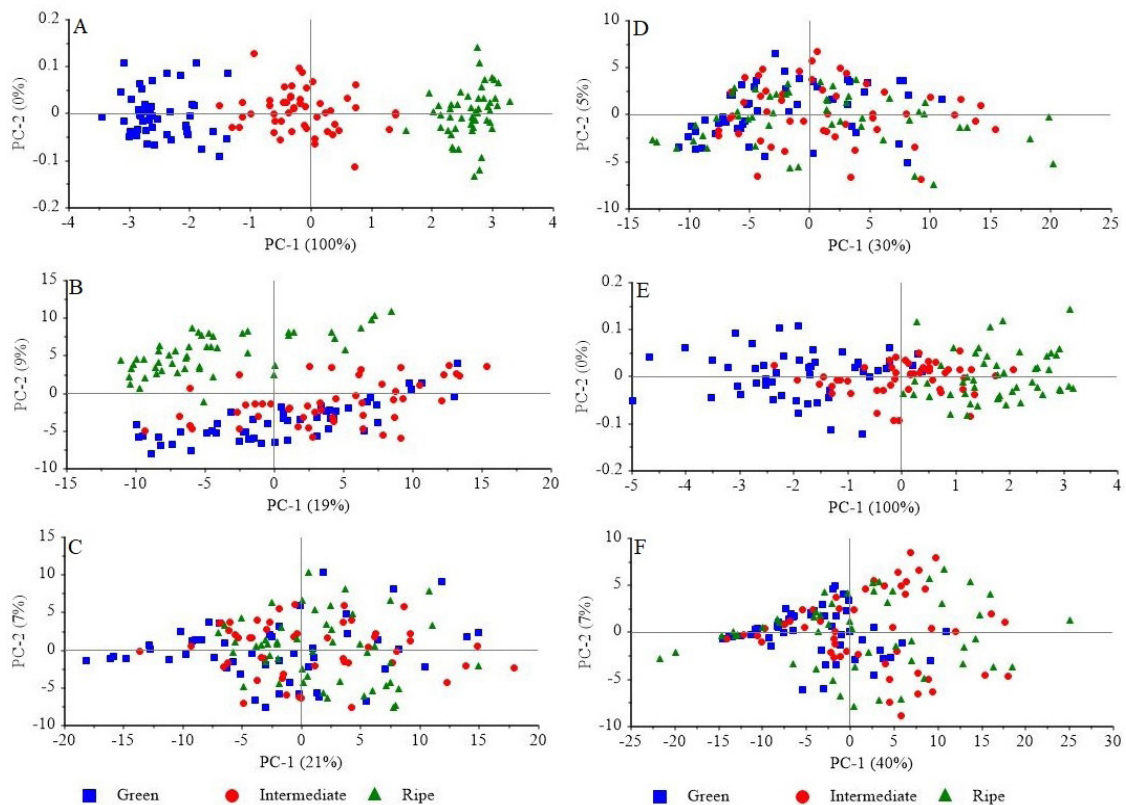


Figure 6. Principal component analysis for the ripening stages of seriguelas in the (A) visible (350-1,000 nm); (B) near-infrared 1 (1,001-1,830 nm); (C) near-infrared 2 (1,831-2,500); and umbu in the (D) visible (350-1,000 nm); (E) near-infrared 1 (1,001-1,830 nm); and (F) near-infrared 2 (1,831-2,500)

Table 4. Performance of the classification models for the discrimination of ripening stages of seriguela and umbu using the full spectrum (461 to 2,059 nm) and variable selection

Model	Pre-processing	Precision	Sensitivity	Selectivity	FPR
Seriguela					
LDA	OSC (35 MAF)	97.78	96.67	98.33	1.67
PLS-DA	35 MAF	94.82	92.66	96.05	3.95
SVM	OSC (35 MAF) + Cfs	96.00	93.75	97.06	2.20
RF	OSC (35 MAF) + Cfs	94.00	93.33	94.29	3.40
Umbu					
LDA	SNV (25 MAF)	95.31	93.53	96.68	3.32
PLS-DA	SNV (25 MAF)	86.01	76.17	88.23	11.77
SVM	Raw	78.00	78.95	92.31	10.10
RF	Raw	66.00	63.64	86.36	18.40

FPR = false-positive rate

The classifiers performances are like those reported by Costa *et al.* (2019), who while studying Vis-NIR spectroscopy in red wine grapes at different ripening stages, found a precision of 93.15% for the PLS-DA model. The discrimination of ripening stages of seriguelas is related to colorimetric alterations in the peel, due to the degradation of chlorophyll and concentration of pigment substances. In this sense, the visible spectral region is associated with the transition of electrons, which can be used to determine the color of the samples (XIAO *et al.*, 2018). For umbu fruits, a reduction in the performance of the classification models is observed. This occurs because the maturing process does not cause an expressive peel color change, maintaining a less intense green color in the fruits at the full ripening stage, in relation to fruits at the green ripening stage (NARAIN *et al.*, 1992).

Thus, the classification models allowed to discriminate the ripening stages of seriguela and umbu fruits using Vis-NIR spectroscopy, with the visible spectral region having a greater contribution.

CONCLUSION

- The predictive regression models with the best performance for seriguela fruits were the RF and PLSR for TSS and firmness, respectively ($R^2 > 92\%$), and for umbu fruits were the PLSR and RF for TSS and firmness, respectively ($R^2 > 0.68\%$).
- The construction of supervised classification models allowed to discriminate the three different ripening stages for the seriguela and umbu fruits, in which the LDA was the classification model that showed the most satisfactory result, with a precision above 95%.
- Therefore, the Vis-NIR spectroscopy is a potential tool to determine the quality attributes and ripening stages, in a non-destructive way, of intact seriguela and umbu fruits.

AUTHORSHIP CONTRIBUTION STATEMENT

SOUZA, P.A.: Formal Analysis, Methodology,

Project administration, Supervision, Writing – review & editing; **FERREIRA, I.J.S.:** Formal Analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing; **COSTA, D.S.:** Conceptualization, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing.

DECLARATION OF INTERESTS

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

REFERENCES

- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorol. Z.**, v. 22, p. 711-728, 2013.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772-777, 1989.
- BIZZANI, M.; FLORES, D. W. M.; COLNAGO, L. A.; FERREIRA, M. D. Non-invasive spectroscopic methods to estimate orange firmness, peel thickness, and total pectin content. **Microchemical Journal**, v. 133, p. 168-174, 2017.
- BRENNA, O. V.; BERARDO, N. Application of near-infrared reflectance spectroscopy (NIRS) to the evaluation of carotenoids content in maize. **Journal of Agricultural and Food Chemistry**, v. 52, n. 18, p. 5577-5582, 2004.
- BRENO, P.; CARVALHO FILHO, C. D.; MATTA, V. M.; MENEZES, J. S.; LIMA, P. C.; PINTO, C. O.; CONCEIÇÃO, L. E. M. G. Produção e caracterização físico-química de fermentado de umbu. **Ciência Rural**, v. 42, n. 9, p.1688-1693, 2012.

- CAYUELA, J. A.; WEILAND, C. Intact orange quality prediction with two portable NIR spectrometers. **Postharvest Biology and Technology**, v. 58, n. 2, p. 113-120, 2010.
- COSTA, B. L.; SOUZA, P. A.; SANTOS, E. R. M.; SENA NETO, B. G.; SANTOS, S. C. L.; LUCAS, G. K. S.; SILVA, R. J.; ROCHA, J. P. M.; CARBEIRO, L. C. Qualidade pós-colheita dos frutos do Umbuzeiro (*Spondias tuberosa*) submetidos ao recobrimento com Fécula de Mandioca e PVC. **Society and Development**, v. 10, n.1, p. 1-12, 2021.
- COSTA, D. S.; MESA, N. F. O.; FREIRE, M. S.; RAMOS, R. P.; MEDEROS, B. J. T. Development of predictive models for quality and maturation stage attributes of wine grapes using Vis-NIR reflectance spectroscopy. **Postharvest Biology and Technology**, v. 150, p. 166-178, 2019.
- ENGELSEN, S. B.; MIKKELSEN, E.; MUNCK, L. New approaches to rapid spectroscopic evaluation of properties in pectic polymers. **Progress in Colloid and Polymer Science**, v. 108, p. 166-174, 1998.
- FU, X.; YING, Y.; ZHOU, Y.; XIE, L.; XU, H. Application of NIR spectroscopy for firmness evaluation of peaches. **Journal of Zhejiang University SCIENCE B**, v. 9, n. 7, p. 552-557, 2008.
- GABRIËLS, S. H. E. J.; MISHRA, P.; MENSINK, M. G. J.; SPOELSTRA, P.; WOLTERING, E. J. Non-destructive measurement of internal browning in mangoes using visible and near-infrared spectroscopy supported by artificial neural network analysis. **Postharvest Biology and Technology**, v. 166, p. 1-9, 2020.
- GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986.
- GÓMEZ, A. H.; HE, Y.; PEREIRA, A. G. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. **Journal of Food Engineering**, v. 77, n. 2, p. 313-319, 2006.
- GODDU, R. F.; DELKER, D. A. Spectra-structure correlations for near-infrared region. **Analytical Chemistry**, v. 32, n. 1, p. 140-141, 1960.
- GORRY, P. A. General least-square smoothing and differentiation by the convolution (Savitzky-Golay) method. **Analytical Chemistry**, v. 62, n. 6, p. 570-573, 1990.
- HEMRATTRAKUN, P.; NAKANO, K.; BOONYAKIAT, D.; OHASHI, S.; MANIWARA, P.; THEANJUMPOL, P.; SEEHANAM, P. Comparison of reflectance and interactance modes of visible and near-infrared spectroscopy for predicting persimmon fruit quality. **Food Analytical Methods**, v. 14, p. 117-126, 2021.
- ISAKSSON, T.; NAES, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. **Applied Spectroscopy**, v. 42, n. 7, p. 1273-1284, 1998.
- LI, J.; HUANG, W.; ZHAO, C.; ZHANG, B. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. **J. Food Eng.**, v. 116, p. 324-332, 2013.
- LIMA, M. A. C.; SILVA, S. M.; OLIVEIRA, V. R. Umu - *Spondias tuberosa*. In: RODRIGUES, S.; SILVA, E. O.; BRITO, E. S. (ed.). **Exotic fruits reference guide**. London: Academic Press, 2018.
- LIMA, M. A. C.; CASTRICINI, A. Qualidade e pós-colheita do umbu. **Informe Agropecuário**, v. 40, n. 307, p. 80-90, 2019.
- LIRA JÚNIOR, J. S.; BEZERRA, J. E. F.; LEDERMAN, I. E.; MOURA, ROBERTO J. M. Produção e características físico-químicas de clones de ciriguela na Zona da Mata Norte de Pernambuco. **Revista Brasileira de Ciências Agrárias**, v. 5, n. 1, p. 43-48, 2010.
- LUO, X.; TAKAHASHI, T.; KYO, K.; ZHANG, S. Wavelength selection in vis/NIR spectra for detection of bruises on apples by ROC analysis. **J. Food Eng.**, v. 109, p. 457-466, 2012.

- MA, T.; XIA, Y.; INAGAKI, T.; SATORU TSUCHIKAWA, S. Rapid and nondestructive evaluation of soluble solids content (SSC) and firmness in apple using Vis-NIR spatially resolved spectroscopy. **Postharvest Biology and Technology**, v. 173, p. 1-12, 2021.
- MALDONADO-ASTUDILLO, Y. I.; ALIA-TEJACAL, I.; NÚÑEZ-COLÍN, C. A.; JIMÉNEZ-HERNÁNDEZ, J.; PELAYO-ZALDÍVAR, C.; LÓPEZ-MARTÍNEZ, V.; ANDRADE-RODRÍGUEZ, M.; BAUTISTA-BANOS, S.; VALLE-GUADARRAMA, S. Postharvest physiology and technology of *Spondias purpurea* L. and *S. mombin* L. **Scientia Horticulturae**, v. 174, p. 193-206, 2014.
- MENEZES, P. H. S.; SOUZA, A. A.; SILVA, E. S.; MEDEIROS, R. D.; BARBOSA, N. C.; SORIA, D. G. Influência do estágio de maturação na qualidade físico-química de frutos de umbu (*Spondias tuberosa*). **Scientia Agropecuaria**, v. 8, n. 1, p. 73-78, 2017.
- MOGHIMI, A.; AGHKHANI, M. H.; SAZGARNIA, A.; SARMA, M. Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. **Biosystems Engineering**, v. 106, n. 3, p. 295-302, 2010.
- MORAES, F. P. M.; COSTA, R. C.; MORAIS, C. L. M.; MEDEIROS, F. G. M.; FERNANDES, T. R. N.; HOSKIN, R. T.; LIMA, K. M. G. Estimation of ascorbic acid in intact acerola (*Malpighia emarginata* DC) fruit by NIRS and Chemometric Analysis. **Horticulturae**, v. 5, n. 12, p. 1-10, 2019.
- MONTALVO-GONZÁLEZ, E.; GARCÍA, H. S.; OCA, M. M.; TOVAR-GÓMEZ, B. Effect of light on Mexican plum stored under different storage conditions. **Journal of Food**, v. 9, n. 1, 2011.
- NARAIN, N.; BORA, P. S.; HOLSCHUH, H. J.; VASCONCELOS, M. A. S. Variation in physical and chemical composition during maturation of umbu (*Spondias tuberosa*) fruits. **Food Chemistry**, v. 44, n. 4, p. 255-259, 1992.
- NERIS, T. S.; LOSS, R. A.; GUEDES, S. F. Caracterização físico-química da seriguela (*Spondias purpurea* L.) colheitadas no município de Barra do Bugres/MT em diferentes estádios de maturação. **Sustenere**, v. 7, n. 1, p. 1-10, 2017.
- RAMBO, M. K. D.; AMORIM, E. P.; FERREIRA, M. M. C. Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues. **Analytica Chimica Acta**, v. 775, p. 41-49, 2013.
- SAMPAIO, S. A.; BORA, P. S.; HOLSCHUH, H. J. Postharvest respiration and maturation of some lesser-known exotic fruits from Brazil: ciriguela (*Spondias purpurea* L.). **Ceres**, v. 55, n. 2, p. 141-145, 2008.
- SANCHES, A. G.; SILVA, M. B.; MOREIRA, E. G. S.; SANTOS, E. X. Atraso na maturação e qualidade pós-colheita de seriguela exposta a radiação ultravioleta-C. **Nativa**, v. 6, n. 3, p. 225-232, 2018.
- SANTOS NETO, J. P. S.; ASSIS, M. W. D.; CASAGRANDE, I. P.; CUNHA JÚNIOR, L. C.; TEIXEIRA, G. H. A. Determination of 'Palmer' mango maturity indices using portable near infrared (VIS-NIR) spectrometer. **Postharvest Biology and Technology**, v. 130, p. 75-80, 2017.
- SANTOS, C. A. F.; OLIVEIRA, V. R. Inter-relações genéticas entre espécies do gênero *Spondias* com base em marcadores AFLP. **Revista Brasileira de Fruticultura**, v. 30, n. 3, p. 731-735, 2008.
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627-1639, 1964.
- SILVA, L. R. I.; PEREIRA, T. S.; SILVA, V. M. A.; SILVA, G. M.; RAMOS LUIZ, M.; SANTOS, N. C.; ALMEIDA, R. L. J.; RIBEIRO, V. H. A.; MUNIZ, C. E. S.; SILVA EDUARDO, R.; SILVA, R. de A. Propriedades termofísicas da polpa de seriguela em diferentes estádios de maturação. **Research, Society and Development**, v. 9, n. 8, p. 1-15, 2020.

SOUZA, S. F.; SILVA, J. L. M.; GUEDES, J. P. M.; LIMA, J. R. F. Competitividade e parcela de mercado das exportações brasileiras de manga: uma análise do modelo *Constant Market Share*. **Revista Econômica do Nordeste**, Fortaleza, v. 47, n. 1, p. 39-48, 2016.

TORRES, I.; PÉREZ-MARÍN, D.; HABA, M. J.; SANCHEZ, M. T. Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors. **Biosystems Engineering**, v. 153, p. 140-148, 2017.

VLAIC, R. A.; MURESAN, V.; MURESAN, A. E.; MURESAN, C. C.; PAUCEAN, A.; MITRE, V.; CHIS, S. M.; MUSTE, S. The changes of polyphenols, flavonoids, anthocyanins and chlorophyll content in plum peels during growth phases: from fructification to ripening. **Notulae**

Botanicae Horti Agrobotanici Cluj-Napoca, v. 46, n. 1, p. 148-155, 2018.

WOLD, S.; RUHE, A.; WOLD, H. The collinearity problem in linear regression. The partial least square (PLS) approach to generalized inverses. **SIAM Journal on Scientific Computing**, v. 5, n. 3, p. 735-743, 1984.

WOLD, S.; ANTTI, H.; LINDGREN, F.; ÖHMAN, J. Orthogonal signal correction of near-infrared spectra. **Chemometrics and Intelligent Laboratory Systems**, v. 44, n. 1, p. 175-185, 1998.

XIAO, H.; LI, A.; LI, M.; SUN, Y.; TU, K.; WANG, S.; PAN, L. Quality assessment and discrimination of intact white and red grapes from *Vitis vinifera* L. at five ripening stages by visible and near-infrared spectroscopy. **Scientia Horticulturae**, v. 233, p. 99-107, 2018.