

## Near infrared spectroscopy sugar quantification in intact orange

Karla Rodrigues Borba<sup>1</sup>, Karla Siebert Sapelli<sup>2</sup>, Poliana Cristina Spricigo<sup>3</sup> & Marcos David Ferreira<sup>4</sup>

### SUMMARY

One of the critical points in the orange processing industry is the fruit quality evaluations by destructive methods. Therefore, non-destructive analysis can represent an advance in the system. The reflectance spectroscopy technique of near infrared (NIR) was employed with spectrophotometer model Spectrum 100N associated with chemometric method regression of Partial Least Squares (PLS) to quantify the total sugars concentration in intact oranges (cv. Valencia). Reference values for total concentration sugars were obtained by high-performance liquid chromatography with chromatograph equipped Refractive Index detector (RI). Spectral data were pre-processed by Normal Standard Variation (SNV), centering on the middle and 1st derivative with 21 points. Chemometric data analysis was done using the software Pirouette<sup>®</sup> for the calibration models construction with regression method of Partial Least Squares (PLS). After this step, the prediction was performed. The model presented Pearson correlation coefficients (r) prediction of 0.88 and Root Mean Square Error of Prediction (RMSEP) of 14.12 g L<sup>-1</sup>. Preliminary results show spectroscopy technique as an alternative non-destructive analysis for oranges quality determination, showing a strong correlation and low prediction error value.

**Index terms:** infrared, chemometric, PLS, total sugar.

### Espectroscopia de infravermelho próximo na quantificação de açúcar em laranja intacta

### RESUMO

Um dos pontos críticos da indústria de processamento de laranja é a avaliação da qualidade dos frutos por métodos destrutivos. Portanto, a análise não destrutiva pode representar um avanço no sistema. A técnica de espectroscopia de reflectância no infravermelho próximo (NIR) foi empregada com o espectrofotômetro modelo Spectrum 100N associado à regressão do método quimiométrico dos mínimos quadrados parciais (PLS) para quantificar a concentração de açúcares totais em laranjas intactas (cv Valencia). Os valores de referência para a concentração total de açúcares foram obtidos por cromatografia líquida de alto desempenho com detector de índice de refração (RI) equipado com cromatografia. Os dados espectrais foram pré-processados pela Variação Padrão Normal (SNV), centrando-se no meio e 1ª derivada com 21 pontos. A análise de dados quimiométricos foi feita

<sup>1</sup> Department of Food and Nutrition, School of Pharmaceutical Sciences, Universidade Estadual Paulista – UNESP, Araraquara, SP, Brazil

<sup>2</sup> Universidade Estadual do Centro-Oeste – UNICENTRO, Guarapuava, PR, Brazil

<sup>3</sup> Department of Plant Science, Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ, Universidade de São Paulo – USP, Piracicaba, SP, Brazil

<sup>4</sup> Empresa Brasileira de Pesquisa Agropecuária – Embrapa, São Carlos, SP, Brazil

**Corresponding author:** Karla Rodrigues Borba, Department of Food and Nutrition, School of Pharmaceutical Sciences, Universidade Estadual Paulista – UNESP, Rod. Araraquara-Jaú, Km 1, CEP 14801-902, Araraquara, SP, Brazil. E-mail: [borbakr@gmail.com](mailto:borbakr@gmail.com)

usando o software Pirouette® para a construção dos modelos de calibração com o método de regressão de mínimos quadrados parciais (PLS). Após este passo, a previsão foi realizada. O modelo apresentou a predição de coeficientes de correlação de Pearson ( $r$ ) de 0,88 e o erro quadrático médio da Predição (RMSEP) de 14,12 g L<sup>-1</sup>. Os resultados preliminares mostram a técnica de espectroscopia como uma análise alternativa não destrutiva para determinação de qualidade de laranjas, mostrando uma forte correlação e baixo valor de erro de predição.

**Termos de indexação:** infravermelho, quimiométrica, PLS, açúcar total.

## INTRODUCTION

Brazil is the world's largest orange grower, responsible for about 25% of the global production (CitrusBR 2011; FAO, 2015). Most of the orange production is directed for juice extraction, where quality plays an essential role in consumer's acceptability. Sugars are greatly important to orange quality, denoting 85% of the soluble solids, a commonly parameter used for classification purposes during citrus marketing (Neves, 2010). Sugars and organic acids balance comprised gives flavors to fruits, that may be either pleasant or unpleasant to the consumer (Spoto et al., 2006).

Performed by destructive techniques, conventional quality control requires time, reagents, team training, and waste generation. In addition, these methods limit analyzes to only a representative batch of the total samples, since it is a destructive. The use of near-infrared spectroscopy in the quality analysis of fruits and vegetables is an alternative technique. The major advantages of near-infrared (NIR) spectroscopy is related to a short time results – only a few seconds – as well as non-destructive character (Nicolai et al., 2007).

Several studies have reported the use of NIR infrared to quantify sugar (reducing, total and SSC) in fruits. Oliveira et al. (2014) quantified sugars (sucrose, glucose and fructose) in passion fruit samples, comparing the NIR and MIR techniques. Jamshidi et al. (2014) developed SSC prediction models in oranges, as well as Cayuela & Weiland (2010) and Gómez et al. (2006). This study aim to evaluate NIR technique and the chemometric method partial least squares (PLS) regression in an effort to develop models for determining the total sugars in Valência oranges.

## MATERIALS AND METHODS

### Plant material

Oranges (*Citrus sinensis*) of Valência cultivar were harvested in 2015 at a commercial orchard in the state of São Paulo, Brazil. In the laboratory, fruits were sanitized,

selected and stored at 12 °C ± 1.5 (R.H. = 90% ± 5). From 100 oranges harvested, 70 fruits were used to the calibration with crossvalidation and 30 were used to external validation.

### Characterization of samples

Fresh weight, equatorial diameter, soluble solids content, and pH of oranges samples were analyzed. Fruits were individually weighed in a Marte AS 2000C electronic scale. The values were expressed in grams. The cross-sectional diameter (i.e. the equatorial diameter) was measured with a Mitutoyo digital caliper.

The soluble solids content (SSC) was quantified in triplicate by inserting 2 mL of the processed samples in a refractometer Atago RX-5000cx. The results were expressed as %SSC (AOAC, 1992). The measurement of pH was performed by immersing the QX 1500 QUALSTRON equipment electrode in 20-30 mL orange juice extracted.

### Near-infrared (NIR) spectroscopy

The NIR spectra of oranges were analyzed through reflectance in a commercial infrared spectrophotometer, model Spectrum 100N (Perkin-Elmer Corp., Norwalk, CT). Spectra were acquired between 4000-10000 cm<sup>-1</sup>, with a resolution of 32 cm<sup>-1</sup>, and using 32 scans per sample to improve the signal-to-noise ratio. The analysis was performed in triplicate by the contact of equatorial region of the whole fruit peel to the equipment crystal.

### Reference method for Total Sugars

Total sugars were quantified by HPLC (High performance liquid chromatography) through a methodology adapted from Kelebek et al. (2009), using a Varian liquid chromatograph equipped with a HPX 87H column and a refractive index (RI) detector. The column was maintained at room temperature (25 °C). Sulfuric acid (H<sub>2</sub>SO<sub>4</sub> 0.5 mM) was

used as the mobile phase with a flow rate of 0.6 mL min<sup>-1</sup>. Twenty µL of each sample were injected in the system.

Juice samples of each fruit were extracted and stored at -28 °C. Samples were thawed, diluted to 1:10 in ultrapure water (Milli-Q®) and filtered through a disposable filter made up of hydrophilic Teflon (porosity of 0.45 µm) prior to the analyses. Standards comprised 99%+ pure sucrose, 99.5%+ pure D-glucose, and 99%+ pure D-fructose (Sigma Life Science). The total sugars value was obtained by adding the values of sucrose, fructose and glucose.

## Data processing

The chemometric analysis were performed using the Origin 8.1 (OriginLab, Northampton, MA, USA) and Pirouette v.4.5 (Infometrix, Inc. Bothell – WA) softwares. NIR spectra were centered on the mean value and were pre-processed using Standard Normal Variation (SNV) + first derivative of Savitzky-Golay with window twenty one points, and Normalize (0-1) + Multiplicative signal correction (MSC).

For applying the partial least squares (PLS) regression, were build matrices in the NIR analytical signals (spectra) were used to set up the X matrix that comprised the independent variables. The Y matrix was built with data obtained by chemical reference tests related to the dependent variables. PLS application to the X and Y matrices aimed at correlating spectra and fruit quality data. To establish the PLS models for each quality parameter, the total data set was randomly algorithm divided into two groups. One of these was used to build the PLS model for calibration + internal validation with the internal cross validation test leave-one-out, comprising 70% of the data, whereas the other group was used to build the PLS model for external validation and comprised 30% of the data.

The models evaluation was taken from the Pearson correlation coefficient (r) of the calibration (r Cal), Validation (r Val) and prediction (r), and using the root-mean-square errors of calibration (RMSEC), validation (RMSEV), and prediction (RMSEP).

## RESULTS AND DISCUSSION

### Characterization of samples

Descriptive statistics of the physicochemical characterization parameters of the samples are presented in Table 1. The mean and standard derivation of fresh

**Table 1.** Results of the characterization of orange samples

	Mean ± sd	Interval
<b>Mass (g)</b>	163.05 ± 41.09	93.72-232.97
<b>SSC (%)</b>	8.58 ± 1.113	6.28-11.48
<b>pH</b>	3.10 ± 0.0.28	2.67-3.77

mass values were 163.05 g and 41.0 g, respectively, in agreement with Magwaza et al. (2013), who evaluated Valência oranges quality using NIR spectroscopy. The SSC determines, indirectly, the sweetness of the juice (Kader, 2002). The amplitude of variation of the SSC parameter found in this study is equivalent to that verified by Jamshidi et al. (2014), where values of SSC in the range of 7.9 to 11% of SSC were described. Regarding pH values, values below 3.5 represent acid oranges and those with pH above 3.5 are considered less acidic (Ting, 1969).

### Total sugar reference analysis

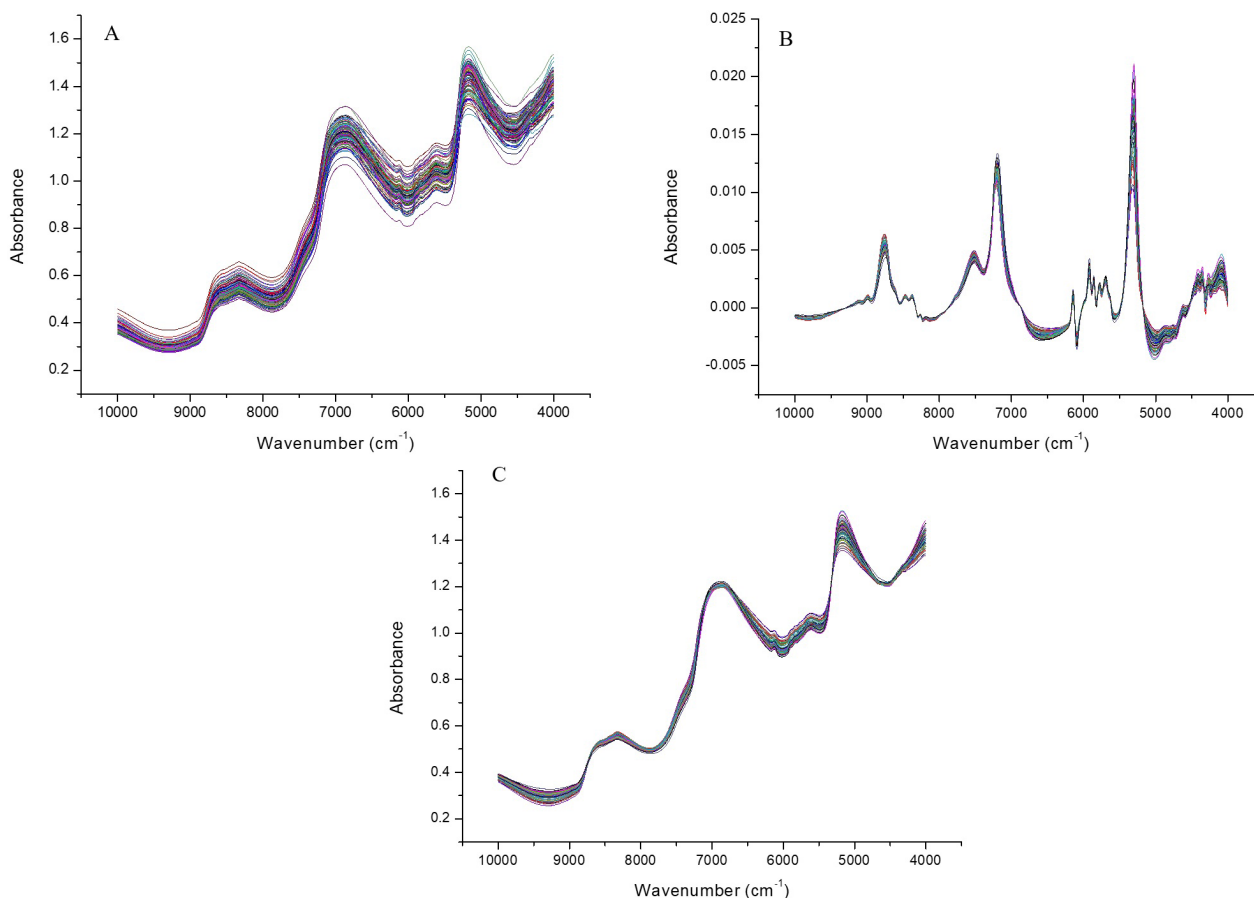
The average, standard deviation and data range value for total sugars analyzed in the calibration + internal validation and prediction sets are shown in Table 2. The results revealed a wide variation range of total sugar value from the samples used in calibration sets with cross-validation and external validation. Such range is essential for models to present applicability, since the regression must be calibrated with representative samples (Nicolai et al., 2014). The averages of total sugars in the calibration + internal validation and external validation sets did not present significant differences (p>0.05). Therefore, these sets presented equivalent averages and value ranges.

### NIR spectra

Due to the light scattering effects caused by the samples lack of homogeneity some transformations and pre-processing in the spectra before starting the construction of the PLS models were required. The unprocessed NIR reflectance spectra is shown in Figure 1A. On the other hand NIR reflectance spectra submitted to mean center + SNV + first derivative with 21 window points in Figure 1B, and mean center + normalize (0-1) + MSC (Multiplicative signal correction) in Figure 1C. Data processing corrected the

**Table 2.** Statistical analysis of the calibration + internal validation and external validation sets: average, standard deviation, and range of total sugars (g L<sup>-1</sup>)

	N = Sample universe	Range	Mean	SD
Calibration + Internal Validation	70	77.98-182.17	114.84	26.4
External Validation	30	72.30-182.17	116.67	28.18

**Figure 1.** (A) Original NIR spectra, without pre-processing and transformations (B) NIR spectra submitted to SNV + first derivation (21); (C) NIR spectra submitted to normalize + MSC.

issues related to spectra baseline variation and light scattering and also highlighted the previously overlapped peaks.

The PLS models comprising NIR spectra were built with two different variable sets: NIRfull, where all the spectral variables were used (10000cm<sup>-1</sup>-4000cm<sup>-1</sup>); and NIRselected, from which the variables within the intervals 10000 – 9600 cm<sup>-1</sup> (characteristic RNH<sub>2</sub> absorption region) and 4400-4000 cm<sup>-1</sup> (C-H and N-H combination region) were excluded (Yang & Irudayaraj, 2002; Lin et al., 2009). Two sharp peaks were observed in the NIR signals of both variable sets. The first peak

occurred at 7000 cm<sup>-1</sup> while the second one occurred at 5000 cm<sup>-1</sup>, indicating regions of OH asymmetric stretching as well as bending combinations (Osborne et al., 1993).

### PLS models of sugar contents

Table 3 shows de PLS models, to calibration, results of different pre-processing methods and spectral ranges. All PLS models have been validated for the external validation stage, because they presented high correlation

**Table 3.** Statistical outcomes of the PLS calibration models built up with the NIR analytical signals for Total sugars (g L<sup>-1</sup>)

Pre-Processing	Calibration + internal validation						
	Wave range (cm <sup>-1</sup> )	N	Factor	RMSEV	r <sub>val</sub>	RMSEC	r <sub>cal</sub>
SNV + First derivation	NIRfull	70	10	13.5	0.85	7.52	0.96
	NIR selected (9600 cm <sup>-1</sup> -4400cm <sup>-1</sup> )	70	10	12.94	0.89	8.3	0.96
Normalize (0-1) + MSC	NIRfull	70	10	13.96	0.85	10.37	0.93
	NIR selected (9600 cm <sup>-1</sup> -4400cm <sup>-1</sup> )	70	10	13.84	0.87	11.47	0.93

RMSEV and RMSEC = root-mean-square errors of validation and calibration, respectively; r<sub>val</sub> and r<sub>cal</sub> = Pearson's correlation coefficients of validation and calibration, respectively; N = sample universe; SNV = Standard Normal Variation; MSC = Multiplicative signal correction.

**Table 4.** Statistical outcomes of the external validation to NIR analytical signals for total sugars (g L<sup>-1</sup>)

Pre-Processing	External Validation					
	Wave range (cm <sup>-1</sup> )	N	Factor	RMSEP	r	RPD
SNV + First derivation	NIR full	30	10	15.96	0.87	1.98
	NIR selected (9600 cm <sup>-1</sup> -4400cm <sup>-1</sup> )	30	9	14.12	0.88	1.95
Normalize (0-1) + MSC	NIR full	30	10	18.24	0.82	1.74
	NIR selected (9600 cm <sup>-1</sup> -4400cm <sup>-1</sup> )	30	9	14.98	0.87	1.97

RMSEP = root-mean-square errors of external validation, r = Pearson's correlation coefficients of external validation; N = sample universe; SNV = Standard Normal Variation; MSC = Multiplicative signal correction; RPD = residual predictive deviation.

( $r > 0.70$ ) and low RMSEV. The optimal number of factors was ten, to all range sets. Comparing the results in table 1, it shows that the optimal PLS model is obtained in the range NIR selected (9600 cm<sup>-1</sup>-4400 cm<sup>-1</sup>).

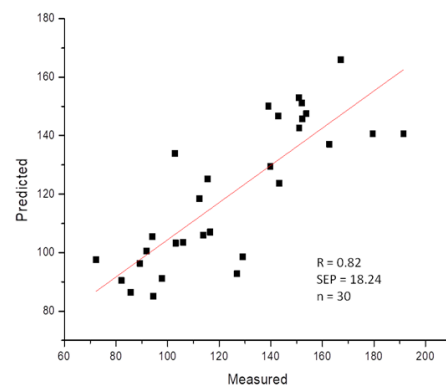
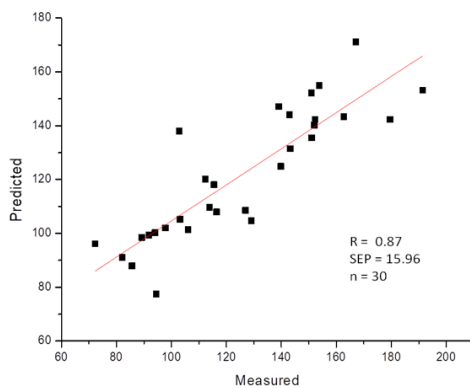
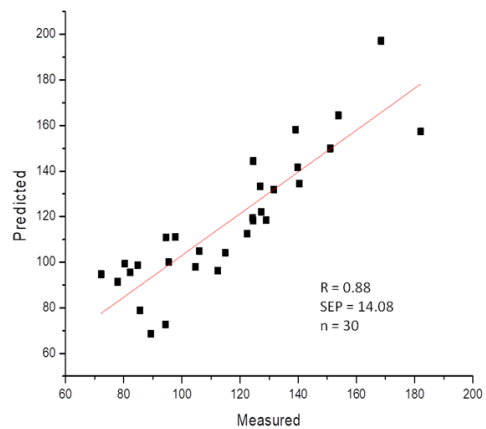
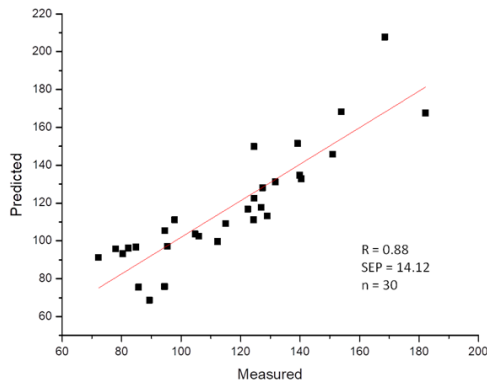
In the external validation the PLS models developed with the NIRselected range also presented results with higher correlation coefficient and less RMSEP (Table 4). Regarding the different types of pre-processing, the application of SNV + First derivation generated the best model, with  $r = 0.88$  and RMSEP = 14.12. To illustrate the importance of models the comparison between the predicted, based on NIR PLS models, total sugars values and the values obtained with the reference method are shown in Figure 2 and 3.

The RPD (residual predictive deviation) is the ratio of the standard deviation for a specific reference population and RMSEP of the prediction set (Mariani et al., 2014). According to Nicolai et al. (2007) a RPD between 1.50-2.00, indicates a model discriminative and low from high

values in the response variable. These models resulted in RPD within the range of 1.5 and 2, therefore they have the power to discriminate between high and low values.

In a similar study, Jamshidi et al. (2014) applied the NIR technique for sugar prediction (SSC) in orange samples, where the correlation value found was  $r = 0.93$ . Sánchez et al. (2012), also analyzing intact oranges by means of NIR obtained models with correlation values between 0.84 and 0.78. NIR spectroscopy has been used for the development of sugars prediction models in other fruits cultivars. For example, in the sucrose, glucose and fructose quantification in passion fruit (Oliveira et al., 2014), in intact mangoes (Delwiche et al., 2008), for SSC in intact plums (Costa & Lima, 2013), and apricot (Oliveira et al., 2014).

This preliminary result demonstrated NIR as a potential technique to determine total sugar content and classify intact oranges.



**Figure 2.** Reference (measured) and predicted values of Total sugars ( $\text{g L}^{-1}$ ) in intact orange fruit Valencia by NIR. PLS models external validation SNV + First derivation correction. (A) NIRfull sets and (B) NIRselected sets.

**Figure 3.** Reference (measured) and predicted values of Total sugars ( $\text{g L}^{-1}$ ) in intact orange fruit Valencia by NIR. PLS models external validation normalize + MSC correction. (A) NIRfull sets and (B) NIRselected sets.

## ACKNOWLEDGEMENTS

We would like to thank the Brazilian agency FAPESP (project grant 13/23479-0) and CAPES for their financial support to this work.

## REFERENCES

AOAC - Association of Analytical Chemists International (1992). Official Methods of analysis of the Association of Analytical Chemists International. 14th ed. Washington: AOAC.

Cayuela JA & Weiland C (2010) Intact orange quality prediction with two portable NIR spectrometers. *Postharvest Biology and Technology* 58(2): 113-120.

CitrusBR - Associação Nacional dos Exportadores de Sucos Cítricos 2011. A indústria brasileira de suco de laranja. São Paulo: CitrusBR. p. 37.

Costa RC & Lima KMG (2013) Prediction of parameters (Soluble Solid and pH) in Intact Plum using NIR spectroscopy and wavelength selection. *Journal of the Brazilian Chemical Society* 24(4): 1351-1356.

Delwiche SR, Mekwatanakarn W & Wang CY (2008) Soluble solids and simple sugars measurement in intact mango using near infrared spectroscopy. *HortTechnology* 3(9): 410-416.

FAO - Food and Agriculture Organization (2015) Faostat Agriculture Data. Rome: FAO. Available from: <<http://faostat.fao.org>>. Accessed: 10 jan. 2015.

Gómez AH, He Y & Pereira AG (2006) Non-destructive measurement of acidity, soluble solids and firmness of

- Satsuma mandarin using Vis/NIR-spectroscopy techniques. *Journal of Food Engineering* 77: 313-319.
- Jamshidi B, Mohajerani E & Ghassemian H (2014) Prediction of soluble solids in oranges using visible/near-infrared spectroscopy: effect of peel. *International Journal of Food Properties* 17: 1460-1468.
- Kader AA (2002) Postharvest technology of horticultural crops. Oakland: Agriculture and Natural Resources, University of California.
- Kelebek H, Selli S, Canbas A & Cabaroglu T (2009) HPLC determination of organic acids, sugars, phenolic compositions and antioxidant capacity of orange juice and orange wine made from a Turkish cv. Kozan. *Microchemical Journal*, 91: 187-192.
- Lin M, Rasco AB, Cavinato AG & Al-Holy M (2009) Infrared (IR) Spectroscopy - Near - Infrared Spectroscopy and Mid-Infrared Spectroscopy. In: Sun DW (Ed). *Infrared spectroscopy for food quality analysis and control*. New York: Academic Press, p. 119-143.
- Magwaza LS, Opara UL, Terry LA, Landahl S, Cronje PJR, Nieuwoudt HH, Hanssens A, Saeys W & Nicolai BM (2013) Evaluation of Fourier transform-NIR spectroscopy for integrated external and internal quality assessment of Valencia oranges. *Journal of Food Composition and Analysis* 31: 144-154.
- Mariani NC, Costa RC, Lima KM, Nardini V, Cunha Júnior LC1 & Teixeira GH (2014) Predicting soluble solid content in intact jaboticaba [*Myrciaria jaboticaba* (Vell.) O. Berg] fruit using near-infrared spectroscopy and chemometrics. *Food Chemistry* 159: 458-462.
- Neves, M. F. (Coord) (2010) O retrato da citricultura brasileira. São Paulo: CitrusBr. Available from: <[http://www.citrusbr.com/download/biblioteca/Apresentacao\\_Marcos\\_Fava\\_evento\\_valor.pdf](http://www.citrusbr.com/download/biblioteca/Apresentacao_Marcos_Fava_evento_valor.pdf)>. Accessed: 10 jan. 2015.
- Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI & Lammertyn J (2007) Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology* 46: 99-118.
- Nicolai BM, Defraeye T, Ketelaere B, Herremans E, Hertog MLATM, Saeys W, Torricelli A, Vandendriessche T & Verboven P (2014) Nondestructive measurement of fruit and vegetable quality. *Annual Review of Food Science and Technology* 5: 285-312.
- Oliveira GA, Castilhos FR, Renard CMGC & Bureau S (2014) Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Research International* 60: 154-162.
- Osborne B, Fearn T & Hindle P (1993) *Practical NIR spectroscopy with applications in food and beverage analysis second*. Harlow UK: Addison-Wesley Longman Ltd.
- Sánchez M-T, Haba M-J, Serrano I & Pérez-Marín D (2012) Application of NIRS for nondestructive measurement of quality parameters in intact oranges during on-tree ripening and at harvest. *Food Analytical Methods* 6(3): 826-837.
- Spoto MFH, Regiano-D'Arce MAB & Oetterer M (2006). *Fundamentos de ciência e tecnologia de alimentos*. Barueri: Manole.
- Ting SV (1969) Distribution of soluble components and quality factors in the edible portion of citrus fruits. *Journal of Science Horticulture* 32: 515-519.
- Yang H & Irudayaraj J (2002) Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques. *The Journal of Pharmacy and Pharmacology* 54(9): 1247-1255.

---

*Received: May 01, 2017*

*Accepted: November 11, 2017*