DETERMINATION OF SIGNIFICANT WAVELENGTHS FOR NITROGEN PREDICTION IN CITRUS LEAVES USING STEPWISE MULTIPLE LINEAR REGRESSION

Lucas Prado Osco[1], Érika Akemi Saito Moriya[2], Ana Paula Marques Ramos[1], José Marcato Júnior[3], Nilton Nobuhiro Imai[2], José Eduardo Creste[1]

[1] Universidade do Oeste Paulista – UNOESTE. 19067-175, Presidente Prudente, São Paulo, Brazil. E-mails: pradoosco@gmail.com, anaramos@unoeste.br, jcreste@unoeste.br; [2] Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP. 19060-900, Presidente Prudente, São Paulo, Brazil. E-mails: eakemisaito@gmail.com, nnimai@fct.unesp.br; [3]Universidade Federal de Mato Grosso do Sul. – UFMS. 79070-900, Campo Grande, Mato Grosso do Sul, Brazil. E-mail: jose.marcato@ufms.br;

ABSTRACT

Hyperspectral data measurements are a non-destructive and lesser time-consuming method for the determination of leaf nitrogen (N) concentration, which is highly needed for the appropriate management of citrus crops. We measured the spectrum of 32 different orange trees using a field spectroradiometer and calculate their reflectance and, later, absorbance levels. We proceeded to assemble these 32 spectra data, and determined the correlation between the N content in leaves and wavelengths measured values using the Stepwise Multiple Linear Regression (SMLR) method. As results, the blue (423 nm) and red (669 nm) wavelengths are better correlated with the N concentration on citrus leaves, with an R^2 equal to 0.9608 and 0.9663, respectively.

Key words — SMLR, hyperspectral, nutrient, spectrum, leaf.

1. INTRODUCTION

Nitrogen (N) in agricultural fields improves the crop yield, as it is highly correlated with foliar growing and chlorophyll content [1]. However, during applications, the non-absorbed portion brings concerns due to its leaching and volatilization phenomena, triggering environmental impacts [2]. To correctly manage their fertilization, farmers are requiring more precise and frequent monitoring of their crops. The traditional agronomic methods discourage a more frequent crop monitoring since they are based on plant tissue analyses, and so, labor-intensive, time-consuming and expensive [3]. For that, hyperspectral measurements are an alternative to predict nitrogen content in crops and orchards, as this method does not adopt the same procedures taken by those more traditional ones [4]. In citrus, a recent research predicted N concentration in leaves based on hyperspectral data [5].

To correlated the nitrogen content in orchards and the reflectance and absorbance of their leaves measured with spectroradiometers, several statistical methods have been adopted based on linear and non-linear prediction models [6; 7; 8]. The Stepwise Multiple Linear Regression (SMLR) is an improvement of the forward regression that re-exams every step of variables incorporated in the model during its previous steps [7]. This method is commonly used to calculate the relationship between leaf N concentration and the spectral wavelengths. Recent studies obtained good prediction levels in different agricultural crops with the application of the SMLR method [5; 6; 7; 8]. The method, however, receive critiques in the past [9], and since than its usage should be considered carefully. Therefore, this paper aims to investigate the usage of the SMLR method to identify significant wavelengths for the N prediction on citrus leaves.

2. MATERIAL AND METHODS

We conducted a field experiment during the end of March 2018, in a Valencia orange orchard in a farm located at Ubirajara, in São Paulo state, Brazil. We selected this period because it corresponds to the end of the summer and so the orange trees assume their vegetative stage. This farm performs soil and leaf analysis regularly, and the Valencia trees selected for this study where planted in 2013, with 700 trees ha^{-1}. The amount of N fertilization applied in soil was the same for all the trees, corresponding to 250 $kg.ha^{-1}$ of urea fertilizer. The historical data provided by the farm management shows different rates of nitrogen absorption by the plants, which motivate our study towards the area.

In field, we measured the radiance of different leaves of Valencia orange trees using an ASD FieldSpec HandHeld spectroradiometer. This equipment operates in the spectral range of 320 – 1075 nm, set on 512 channels with 1.6 nm spectral resolution and a Field of View (FOV) of 10°. We measured 10 spectral curves of 32 orange trees and used their averages values. The spectroradiometer was positioned at an angle of 45° and at an average height in relation to the orange tree. For each tree measurement, the radiance of a Lambertian reference surface (Spectralon® plate) was also measured under the same conditions of illumination to estimate the Hemispheric Conical Reflectance Factor (HCRF) using the methodology described by [10].

Additionally, we collected approximately 4 leaves of each one 32 trees measured with the spectroradiometer. The leaves collected were the same measured, which were packed in appropriated bags to perform the laboratory analyze. The leaves were washed and dried in an oven at 75°C for 48 hours and they were, later, grounded. To quantify the chemical analysis, the Kjeldahl titration method [11] was applied, which consists of three different stages: 1) digestion, 2) distillation in N, and 3) titration with sulfuric acid (H_2SO_4). We determined the N concentration in leaves for each one of the 32 orange trees. Figura 1 displays two examples of spectra curves obtained by the sampled leaves.

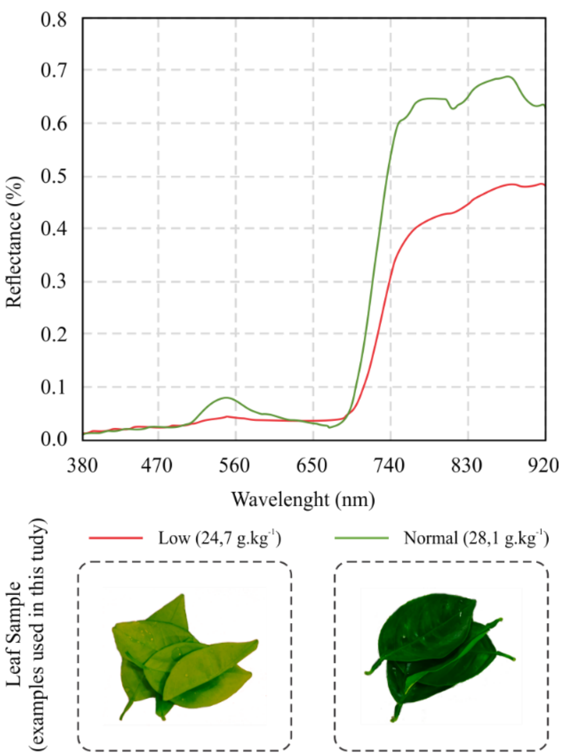

Figure 1. Wavelengths for leaves with different nitrogen

The values of leaf N concentration of each sample were assigned to their corresponded spectrum. To remove regions that displayed a low signal-to-noise ratio, we select the wavelengths between 380 and 920 nm. Following the Beer-Lambert law, which establishes that the concentration of an absorber is proportional to a sample absorbance, we converted the measured spectral reflectance to absorbance by applying the logarithm function of the inverse of the reflectance value [7]. This procedure was important to obtain an improved relationship between leaf N the and spectra data.

We perform three different combinations to cluster the data. First, we cluster the data based on Euclidian Distance method using N values of each sample, and then we calculate their spectral wavelengths means. This procedure resulted in seven different mean curves. Second, we adopted the Principal Components Analysis (PCA) which is used to convert correlated variables into newer linearly uncorrelated variables [1]. We applied the PCA using the wavelengths values measured by the spectroradiometer, and this resulted in four different curves. Third, we combine all the 32 curves in only three classes based on [12] standards, who measured the critical values of nitrogen for Valencia trees. These authors developed a non-linear regression between N leaf concentration and yield gained. Thus, we assumed the following N classification: low (< 24 $g.kg^{-1}$), medium (> 27 and < 29 $g.kg^{-1}$) and high (> 29 $g.kg^{-1}$). The medium class corresponds with the best possible yield gained, since, after this point, the curve from the regression tends to decline, and results in lesser yield values.

Using each one of the three combinations proposed, we applied the SMLR method, and calculated their correlation coefficient (r), coefficient of determination (R^2), standard deviation error (Std) and the root mean square difference (RMSD), between the leaf nitrogen concentration and the measured wavelengths means values, in order to evaluate the reliability of the SMLR method application. We also evaluate the raw data, i.e., all the 32 curves measured. Thus, we identified the spectra regions with high and low correlation with the N in leaves. For the analysis, were considered only the correlation coefficients (r) between 0.95 and -0.95.

3. RESULTS

Table 1 shows the correlation analysis for each grouping considered. The cluster based on the raw data resulted in the worst results. The clustering based on the Euclidian Distance method, although reduced the number of spectra curves, it did not present a good relation to the amount of N measured. When adopting the PCA method, we found out a high correlation in the blue region (380 – 450 nm). The best result was obtained when we used the Nitrogen Classification method, were our data was separated into the three different nitrogen classes (low, medium and high). For this case, we found out spectral regions highly correlated, from the visible blue (402 – 495 nm) and red regions (639 – 688 nm) to the near infrared region (718 – 920 nm). Figure 1 shows the variation of the correlation coefficient for the best method, i.e., the grouping based on Nitrogen Classification.

Table 1. Results of the correlation coefficients and the SMLR analysis for the different grouping methods.

Grouping	Measurements	Correlation (r)	Band Regions	SMLR id.	R^2	Std error	RMSD
Raw Data	32	-0.041 / 0.050	none	----	----	----	----
Euclidian Distance	7	-0.422 / -0.257	none	----	----	----	----
Principal Components Analysis (PCA)	4	-0.996 / -0.562	380 – 450 nm	393 nm*	0.9044	0.0229 / 0.0629	0.0363
Nitrogen Classification[1]	3	-0,998 / -0.867	402 – 495 nm 639 – 688 nm 718 – 920 nm	423 nm** 669 nm**	0.9608 0.9663	0.0085 / 0.0284 0.0077 / 0.0167	0.0089 0.0063

[1] classification based upon the [12] standards; * significant at 5% using the t test; ** significant at 1% using the t test. (min / max) (r > ± 0.95) (a / b)

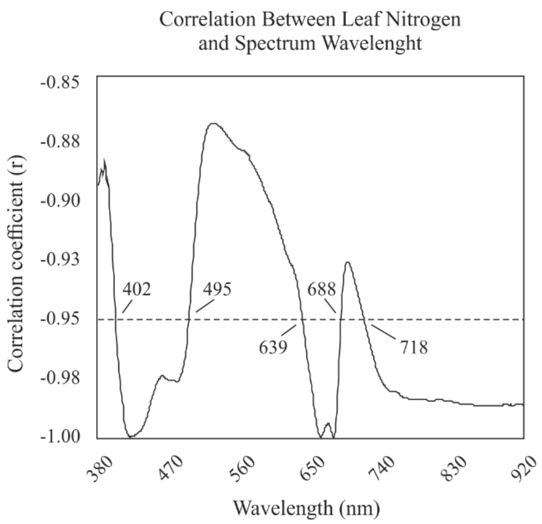

Figure 1. Correlation between absorbance values at each wavelength and the citrus leaf nitrogen concentration

Figure 1 demonstrates that there is a strong negative correlation between the N content from the measured leaves with the absorption levels registered for some spectra regions. This implies that to blue, red and near-infrared wavelength the higher the leaf N, the lesser is the absorption rate. The region between 495 nm and 639 nm corresponds to green wavelengths mostly. Even though N is known to affect the greenness of vegetation, our results show little difference among leaves samples with higher nitrogen (> 29 $g.kg^{-1}$) and with medium nitrogen concentration (< 27 and < 29 $g.kg^{-1}$). Moreover, the region between 688 and 718 nm also did not express a strong correlation as the others (688 – 718 nm) wavelengths (Figure 1).

From analysis based on SMLR method, we found out that the 393 nm wavelength had a good linear relationship with the N content for the PCA group ($R^2 = 0.90$; p-value = 0.0482), and both 423 nm and 669 nm wavelengths maintain the strongest linear relation for the Nitrogen Classification group ($R^2 = 0.96$; p-value < 0.0001). The clustering method based on the [12] standards of N in leaves also performed better results in relation to the standard deviation error and RMSE, showing lower values than those obtained from the PCA method. Therefore, the approach to identify the wavelengths using the Nitrogen Classification grouping was the best and the results are shown in Figure 2.

The Figure 2 reveals that there is a negative relationship between nitrogen content and the absorption percentage. Our results show that, for blue, red and near-infrared regions, it was possible to predict N content in citrus leaves using these wavelengths, although the blue and red regions presented results with higher significance. This implies that is possible to measure leaf N content on citrus using the absorbance measured in these regions.

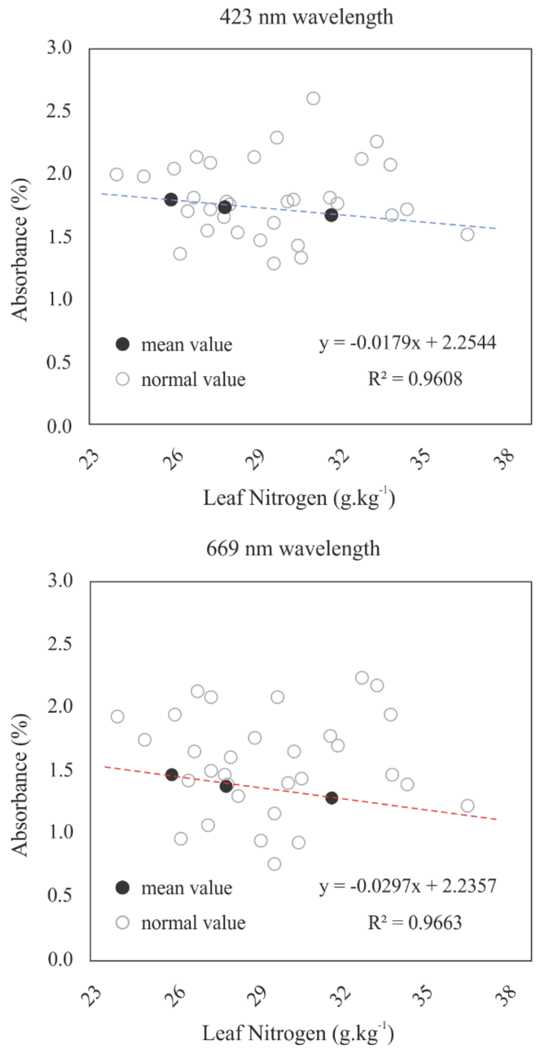

Figure 2. Regression for the best wavelength SMLR analysis.

4. DISCUSSION

Our method presented interesting results, which resemble the results obtained by [7] with measurements performed in controlled laboratory conditions. One of their methods also suggests that the blue and red regions are highly correlated with the chlorophyll content, which is being known as an indicator of N concentration [13]. The main idea behind it is that the N accumulation on leaves ingresses the amount of absorption on the visible light region, and reduces the amount of absorption at the red-edge and near infrared regions [6]. Our results indicate the same implications for the red-edge and near infrared wavelengths. However, for the visible region, we observed an opposite relationship between the absorption rate and the leaf N concentration.

We found that the higher the nitrogen content in leaves, the lesser was the absorption rate by both blue and red regions, a condition that was also observed with the red-edge and the near infrared wavelengths. When [7] analyzed the N content in citrus leaves, they indicated an optimal production threshold between 24 and 27 $g.kg^{-1}$. Our reduced number of sampling used to analyze wavelengths response may be an indicator of that difference. Still, another reason for that may be that we based our classification model in an agronomical study [12] with different thresholds. This study indicates that, by a non-linear regression analysis, the ideal leaf nitrogen concentration intervals for the Valencia orange trees should be between 27 and 29 $g.kg^{-1}$ in order to obtain the maximum yield production. From the 32 samples collected, the majority were above the 29 $g.kg^{-1}$ threshold. This excessive content of N in leaves may lead to unbalanced conditions and interfere with their physiology [3], and clearly in their reflectance and absorption rates, which is implied in our results.

Related to the PCA method, the reduced number of samples did present an expected absorption rate on the blue region, but the mean curves created were not useful for fitting the N concentration to the variable of input. This was also observed by [6] who also applied the PCA method, however in a different statistical model than the SMLR. Nonetheless, the application of the SMLR resulted in similar values of N prediction when compared to values found to other cultures [8]. Even so, the SMLR method cannot be view as a definitive solution to the analysis of spectral data, since it showed different results and has received some critics on its performance [9]. Other methods as the partial least-squares regression (PLSR) and back propagation neural networks (BPNN) already resulted in better prediction models than the SMLR for other crops [1], which opens up new possibilities for the spectrum data evaluation. Likewise, the idea of using different combinations or implementing spectral indices in the analysis [14] could improve the results stated here, as we intend to apply in further investigations.

5. CONCLUSIONS

We conclude that the SMLR method for the nitrogen prediction in citrus leaves is appropriated, but only when adopted the mean values of all the collected wavelengths, segregating the N content into different classes (low, medium and high). We identified that blue (423 nm) and red (669 nm) spectral regions are highly correlated with N concentration, and it may be estimated throughout a linear regression. These finds corroborate with previous analysis conducted by other researches. However, our results contradict the expected absorption rates, establishing a negative relationship between the total of radiance absorbed and the values of N found on the leaves samples. We suggest applying other statistical methods besides the SMLR to understand the differences found here, and verify the performance of different methods.

6. REFERENCES

[1] Stellacci, A. M.; Castrignanò, A. and Troccoli, A. "Selecting optimal hyperspectral bands to discriminate nitrogen status in durum wheat: a comparison of statistical approaches," *Environmental Monitoring and Assessment*, 188-199 (pp.), 2016.

[2] Cilia, C.; Panigada C.; Rossini, M.; Meroni, M.; Busetto, L.; Amaducci, S.; Boschetti, M.; Picchi, V. and Colombo, R. "Nitrogen Status Assessment for Variable Rate Fertilization in Maize through Hyperspectral Imagery," *Remote Sen.*, 6 (v.), 6549-6565 (pp.), 2014.

[3] Huerta, R. F. M.; Gonzalez, R. G. G.; Medina, L. M. C.; Pacheco, I.; Olivarez, J. P. and Velazquez, R. V. "A Review of Methods for Sensing the Nitrogen Status in Plants: Advantages, Disadvantages and Recent Advances," *Sensors*, 13 (v.), 10823-10843 (pp.), 2013.

[4] Wang, J.; Shen, C.; Liu, N.; Jin, X.; Fan, X.; Dong, C. and Xu, Y. "Non-Destructive Evaluation of the Leaf Nitrogen Concentration by In-Field Visible/Near-Infrared Spectroscopy in Pear Orchards," *Sensors*, 17 (v.), 3 (n.), 538-553 (pp.), 2017.

[5] Yanli, L.; Qiang, L.; Shaolan, H.; Shilai, Y.; Xuefeng, L.; Rangjin, X.; Youngqiang, Z. and Lie, D. "Prediction of nitrogen and phosphorus contents in citrus leaves based on hyperspectral imaging", *International Journal of Agricultural and Biological Engineering*, 8 (v.), 80-88 (pp.), 2015.

[6] Connell, J. L. O.; Byrd, K. B. and Kelly, M. "Remotely-Sensed Indicators of N-Related Biomass Allocation in Schoenoplectus acutus," *Plosone*, 9 (v.), 3 (n.), 2014.

[7] Min, M. and Lee, W. "Determination of significant wavelengths and prediction of nitrogen content for citrus," *American Society of Agricultural Engineers*, 48 (v.), 2 (n.), 455-461 (pp.), 2005.

[8] Miphokasap, P. and Wannasiri, W. "Estimations of Nitrogen Concentration in Sugarcane Using Hyperspectral Imagery," *Sustainability*, 10 (v.), 1266-1282 (pp.), 2018.

[9] Grossman, Y. L.; Ustin, S. L.; Jacquemoud, S.; Sanderson; E. W.; Schmuck, G. and Verdebou, J. "Critique of Stepwise Multiple Linear Regression for the Extraction of Leaf Biochemistry Information from Leaf Reflectance Data," *Remote Sensing of Environment*, 56 (v.), 182-193 (pp.), 1996.

[10] Anderson, K.; Rossini, M.; Labrador, J. P.; Balzarolo, M.; Arthur, A.; Fava, F.; Julitta, T. and Vescovo, L. "Inter-comparison of hemispherical conical reflectance factors (HCRF) measured with four fibre-based spectrometers," *Remote sensing and sensors*, 21 (v.), 1 (n.), 605-617 (pp.), 2013.

[11] Labconco, C. A, "*Guide to Kjeldahl Nitrogen Determination Methods and Apparatus*," Lab. Corp.: Houston, USA, 10 (p.), 1998.

[12] Quaggio, J. A.; Cantarella, H. and Raij, B. van. "Phosphorus and potassium soil test and nitrogen leaf analysis as a base for citrus fertilization," *Nutrient Cycling in Agroecosystems*, 52 (v.), 67-74 (pp.), 1998.

[13] Atzberger, C.; Guerif, M.; Baret, F. and Werner, W. "Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat," *Comp. and Elec. Agr.*, 73 (v.), 165-173 (pp.), 2010.

[14] Muharam, F. M.; Mass, S. J.; Bronson, K. F. and Delahunty, T. "Estimating Cotton Nitrogen Nutrition Status Using Leaf Greenness and Ground Cover Information," *Remote Sensing*, 7 (v.), 7007-7028 (pp.), 2015.