DOI: http://dx.doi.org/10.1590/1678-992X-2018-0409

ISSN 1678-992X

SCIENTIA AGRICOLA

Research Article

Estimation of leaf nutrient concentration from hyperspectral reflectance in *Eucalyptus* using partial least squares regression

Luiz Felipe Ramalho de Oliveira*, Reynaldo Campos Santana

Universidade Federal dos Vales do Jequitinhonha e Mucuri/FCA – Depto. de Engenharia Florestal, Rod. MGT 367, km 583, 5000 – Alto da Jacuba – 39100-000 – Diamantina, MG – Brasil.

*Corresponding author <luizfelipe@florestal.eng.br>

Edited by: Mohammad Bagher Hassanpouraghdam

Received December 10, 2018

Accepted May 12, 2019

ABSTRACT: Leaf hyperspectral reflectance has been used to estimate nutrient concentrations in plants in narrow bands of the electromagnetic spectrum. The aim of this study was to estimate leaf nutrient concentrations using leaf hyperspectral reflectance and verify the variable selection methods using the partial least squares regression (PLSR). Two studies were carried out using stands with *Eucalyptus* clones. Study I was established in *Eucalyptus* stands with three clones, classifying leaves into five colour patterns using the Munsell chart for plant tissues. Immediately after leaf collection, leaf reflectance was read and the chemical analysis was performed. Study II was carried out in commercial clonal stands of *Eucalyptus* performing the same leaf sampling and chemical analysis as used in Study I. All leaf reflectance spectra were smoothed and three more pre-processing procedures were applied. In addition, three methods of PLSR were tested. The first derivative was more accurate for predicting nitrogen (R_{cv}^2 = 0.95), phosphorous (R_{cv}^2 = 0.93), and sulphur concentration (R_{cv}^2 = 0.85). The estimates for concentrations of calcium (R_{cv}^2 = 0.81), magnesium (R_{cv}^2 = 0.22), and potassium (R_{cv}^2 = 0.76) were more accurate using the logarithm transformation. Only the estimates for iron concentrations were performed with higher accuracy (R_{cv}^2 = 0.35) using the smoothed reflectance. The copper concentrations were more accurate (R_{cv}^2 = 0.78) using the logarithm transformation. Concentrations of boron (R_{cv}^2 = 0.68) and manganese (R_{cv}^2 = 0.79) were more accurate using the first derivative, while zinc (R_{cv}^2 = 0.31) concentration was most accurate using the second derivative.

Keywords: remote sensing, tree monitoring, modelling, variable selection

Introduction

Non-destructive methods for measuring nutrient concentration, involving different sensors, have been used in forest and agricultural nutritional management (Lu et al., 2018; Oliveira et al., 2017; Pandey et al., 2017). Among these sensors, hyperspectral reflectance stands out, as it allows obtaining information on the relationships between plants and electromagnetic energy in narrow bands of the electromagnetic spectrum. Didactically, the electromagnetic spectrum is divided into regions, according to wavelengths. The visible (VIS, 400 – 700 nm), near infrared (NIR, 700 – 1300 nm) and shortwave infrared (1300 – 2500 nm) regions of the spectrum are the regions that mostly interact with plants *in vivo*, since the leaf pigments in plants, especially chlorophylls, interact with the electromagnetic energy in the VIS region and the leaf structure interacts in the NIR region (Gates et al., 1965).

Not only do leaf pigments and structures relate to electromagnetic energy in the VIS-NIR regions, but leaf nutrients also do (Abdel-Rahman et al., 2017; Mahajan et al., 2014; Pimstein et al., 2011. Nutrients have several functions in plant metabolism in the areas of leaf structure and pigment synthesis, plant energy, and metabolism as well as the electron transport chain, among others (Marschner, 1995). Moreover, under deficiency of some nutrients, leaf chlorophyll concentrations are reduced and leaf reflectance is changed (Adams et al., 2000a, b; Mariotti et al., 1996). For this reason and due to other metabolic alterations, it is possible to detect when plants experience nutrient deficiencies through leaf and tissue chloroses and necrosis (Dell et al., 1996).

Because of the complexity of nutrient functions and the number of plant chemical compounds, estimating leaf nutrients by means of hyperspectral reflectance is not a simple procedure with difficulties with autocorrelation and collinearity (Blackburn, 2007; Wold et al., 2001). To overcome these problems, some modelling techniques have been used to predict leaf nutrient concentrations through hyperspectral reflectance, such as the partial least squares regression (PLSR) linked to variable selection methods (Mehmood et al., 2012). Therefore, this study aimed to estimate leaf nutrient concentrations using leaf hyperspectral reflectance and verify the variable selection methods using the PLSR.

Material and methods

Two studies (Study I and Study II) were carried out using *Eucalyptus* stands in the municipalities of Lassance and Três Marias, Minas Gerais, Brazil, about 800 m a.s.l. (Figure 1).

Study I was established in *Eucalyptus* stands which were 25 months old, established with 7.0 × 1.3 m tree spacing, with three clones (*E. urophylla x E. grandis*: GG680, *E. urophylla x E. grandis*: GG682 and hybrid of *E. urophylla* ST Blake: I144). In these stands, nine plots of 10 ha were allocated, three per clone. Leaves from the lower part of tree crowns were visually classified into five

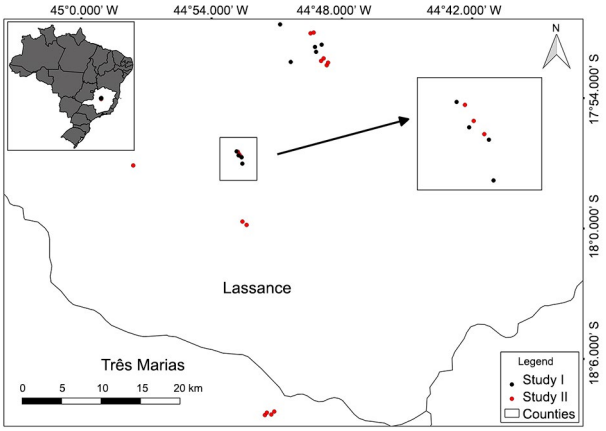

Figure 1 – Plots of Study I and Study II.

Table 1 – Leaf colour patterns sampled in Eucalyptus stands.

Colour name diagrams	Matte colour chips	Hue	Value	Chroma	Munsell colour code
Brilliant yellow green		7.5 GY	8	8	7.5 GY 8/8
Light yellow green		7.5 GY	8	4	7.5 GY 8/4
Brown		7.5 YR	4	2	7.5 YR 4/2
Yellow		2.5 Y	7	6	2.5 Y 7/6
Strong yellow		2.5 Y	8	10	2.5 Y 8/10

colour patterns using the Munsell chart for plant tissues (Gretag-Macbeth, New Winsor, NY, USA). The leaf colour patterns were defined by the clear expression of the biochemical cycling of nutrients (Saur et al., 2000) (Table 1).

For each colour pattern, 30 leaves were collected in each of the nine plots. A leaf was collected from each tree in a random zigzag walk. The set of 30 leaves constituted a composite sample, totalling 45 composite samples (3 clones × 3 plots × 5 leaf colour patterns). This sampling was performed to obtain a wide variation in leaf nutrient concentrations.

Immediately collection, leaf reflectance (400-900 nm) was read in the abaxial part of each leaf, 10 mm from the lower border, on the left side of the leaves using a CI-710 mini-spectrometer (CID Bio- Science - Camas, Washington, USA). Leaf reflectance was analysed using SpectraSnap! (Software version 1.1.3.150, CID Bio-Science) with 300 milliseconds of integration time, a boxcar with 10 points and two scans for averaging. From the set of 30 smoothed spectra in each sample, the average value for leaf reflectance was obtained for each of the 45 composite samples.

Subsequently, the 30 leaves of each composite sample were placed in paper bags and oven-dried with forced air circulation at 65 ºC. After drying, the 45-composite samples were digested in nitro-perchloric solution and concentrations of calcium (Ca), magnesium (Mg), sulphur (S), zinc (Zn), iron (Fe) and manganese (Mn) were determined by spectrophotometry. The phosphorus (P) concentration was determined by colorimetry, the potassium (K) concentration by flame photometry and the total nitrogen (N) concentration by the Kjeldahl method following sulphuric digestion.

Study II was carried out in commercial stands of the hybrid of *Eucalyptus urophylla* ST Blake (I144 clone), which were 9, 12, 15, and 25 months old. Sixteen plots of 10 ha were allocated, four plots per age (Figure 1). Leaves from 25 trees in these plots were sampled. They were taken from tall trees in the upper canopy of the stands (80 % percentile). For each tree, four completely expanded leaves, without physical damages, were collected at the four cardinal points near in the middle of the *Eucalyptus* crowns. The 100-leaf set comprised a composite sample to determine nutrients and measure leaf reflectance. The procedures for leaf collecting and chemical analysis were the same as in Study I.

The results obtained in Studies I and II were grouped into a single database with 45-leaf nutrient concentrations and reflectance samples from Study I and 16-leaf nutrient concentrations and reflectance samples from Study II. All leaf reflectance was smoothed by the Savitzky-Golay algorithm using polynomial 2 (SR, Savitzky and Golay, 1964). In addition, three more pre-processing procedures were used comparatively: leaf reflectance logarithmic transformation (LT, Eq. 1), first (FD) and second derivative (SD).

$$\ln\left(\rho\lambda_j\right) = \log\left(\frac{1}{\rho\lambda_j}\right) \quad (1)$$

where: $\rho\lambda_j$ - leaf reflectance in the j wavelength.

The pre-processed reflectance values for each composite sample (n = 61) were set in PLSR as a set of predictors variables and a single wavelength represented a variable (λ = 2884). The leaf nutrient concentrations, obtained by the chemical analysis, were set as dependent variables (n = 61) in PLSR. For each nutrient, a PLSR was adjusted through the SIMPLS algorithm on all predictors and dependent variables after pre-processing (Eq. 2 and 3).

$$y_i = \beta_0 + \sum_{k=1}^{r} \beta_k T_{ik} + e_i \quad (i = 1, \ldots, n) \quad (2)$$

$$T_{ik} = \sum_{j=1}^{m} c_{kj}\rho_{ij} \quad (k = 1, \ldots, r) \quad (3)$$

where: y_i - nutrient concentration in the i sample; T_{ik} - latent variable k in the i sample; ρ_{ij} - leaf reflectance of sample i in wavelength j; m - total number of wavelengths, 2884; n - total number of samples; e_i - error; β_k - regression coefficients of the k latent variable; r - number of latent variables; c_{kj} - coefficients of the k latent variable in wavelength j.

For each nutrient, three methods of PLSR were tested for each reflectance pre-processing procedure. The method called PLS1 was a fit for the model using the hyperspectral reflectance in all wavelengths. Moreover, the methods wrapper and filter of variable selection were used to select the wavelengths (inputs) for the PLSR models. The iterative predictor weighting PLS (IPW) was performed using 100 iterations and selecting the wavelengths by means of the regression coefficients. In each iteration, the predictor importance was computed after PLS modelling and this importance was used to re-scale the reflectance in each single wavelength and eliminate the least important wavelengths with values below 0.1 before subsequent model re-fitting (Forina et al., 1999).

Another variable selection method used was the variable importance in PLSR projection (VIP) (Wold et al., 1993). This method accumulated the reflectance importance in each wavelength *j* that was reflected by the weight (*w*) from each component and selected the wavelength with VIP > 1 (Eq. 4).

$$VIP_j = \sqrt{p \sum_{a=1}^{A} \left[SS_a \left(W_{aj} / \|W_a\| \right)^2 \right] \Big/ \sum_{a=1}^{A} \left(SS_a \right)} \quad (4)$$

where: SS_a - sum of squares explained by the ath component; $\left(W_{aj} / \|W_a\|\right)^2$ - importance of the jth wavelength.

The number of latent variables (LV) was based on the minimum predicted residual error sum of squares (PRESS) to avoid over fitting of the model. During the process of training and modelling of validation, several models were created to predict concentration of each nutrient. The model that presented less root mean squares error ($RMSE_{train}$ and $RMSE_{cv}$, Eq. 5), as well as the highest coefficient of determination in the training (R_{train}^2) and validation (R_{cv}^2), was selected. The leave-one-out cross-validation (LOOCV) was used, as it provides lower bias and fewer errors when it is used with a limited number of samples (Cawley and Talbot, 2003).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Yi - yi)^2}{n}} \quad (5)$$

where: *Yi* is the nutrient concentration of sample *i* estimated by the equation; *yi* is the nutrient observed in the laboratory of sample *i*; *n* is the total number of samples. *RMSE*: g kg^{-1} for N, P, K, Ca, Mg and S; mg kg^{-1} for B, Zn, Mn, Fe and Cu.

All statistical procedures were carried out using the software R Core Team (2017) version 3.4.0, platform support R Studio version 1.0.143. A method process simplification can be visualised in the workflow in Figure 2.

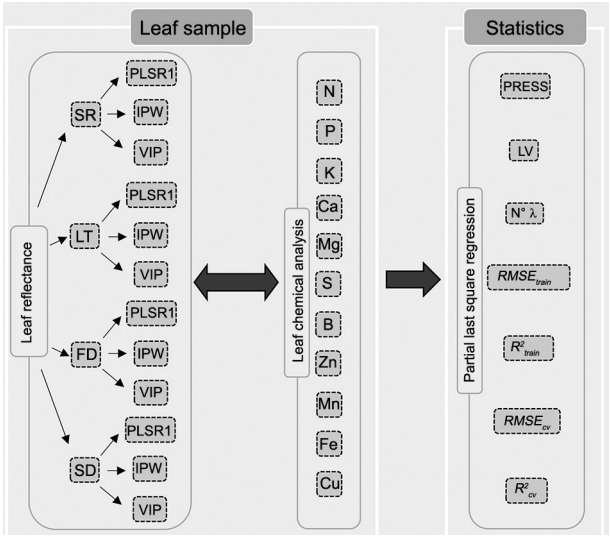

Figure 2 – Workflow of partial least square regression nutrient estimates.

Results

The SR presented higher standard deviation in the VIS region, especially at wavelengths between 550 and 700 nm (Figure 3A). Similarly, the LT increased the standard deviation mainly at the wavelength near the blue region (400 – 500 nm) (Figure. 3B). In the FD pre-processing, the point of maximum inflection of the curve in the red edge (IPP, ~700 nm) was the region with the highest standard deviations (Figure 3C). As observed in FD, the wavelengths close to 700 and 730 nm represented the region with the highest standard deviation in SD (Figure 3D).

Table 2 shows the descriptive statistics for nutrient concentrations of *Eucalyptus* leaves. Large standard deviations are observed in practically all nutrient concentrations, as well as the large range between the nutrient concentration minimum and maximum. On average, nutrient concentrations were classified as optimal for the crop, except for N, P and Fe, which were below optimal.

None of the models used to predict macronutrients showed better accuracy when using the SR and SD pre-procedures (Table 3). The FD was more accurate to predict N and P concentrations using six LV and PLS1. The same pre-processing was used to predict S concentration using 16 LV and select the wavelengths with the VIP method. The estimates of Ca, Mg, and K concentrations were more accurate using the LT. The Ca concentration estimate was more accurate using the 12 LV and the IPW methods. On the other hand, estimates of Mg and K concentration were more accurate using 11 LV while selecting wavelengths using the VIP method.

Among the models developed for micronutrient concentrations, only the estimates for Fe concentrations were performed with high accuracy using the SR preprocess and the VIP method (Table 4). Estimates of Cu concentrations were more accurate using the LT pre-process, with PLS1 and 12 LV. Estimates of B and Mn concentrations were more accurate using the FD with seven LV and selecting the wavelengths using the VIP method. Estimates of Zn concentration were more accurate using the SD with just two LV and the same variable selection method, similar to estimates for B and Mn concentrations.

Table 2 – Descriptive statistics of leaf nutrient concentrations in Eucalyptus.

Statistics	N	P	K	Ca	Mg	S	B	Zn	Mn	Fe	Cu
	g kg^{-1}						mg kg^{-1}				
$\bar{x}$	10.20	0.62	6.73	10.11	1.92	1.43	94.89	22.44	1689.89	66.59	9.26
s.d.	5.94	0.36	2.66	4.14	0.39	0.26	40.56	5.67	905.52	31.33	1.35
Min	2.36	0.18	1.95	3.45	1.03	0.93	25.30	7.83	288.97	24.28	7.60
Max	18.95	1.24	12.46	19.71	3.00	2.02	171.40	35.83	4066.00	182.52	13.10

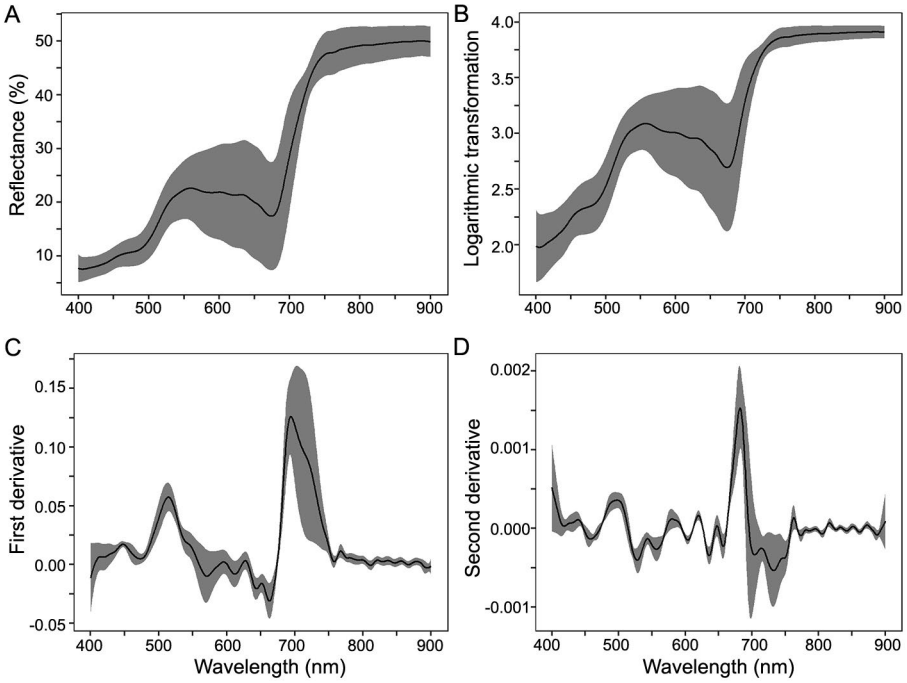

Figure 3 – Average and standard deviation of pre-processing leaf reflectance.

Estimates of N concentration were more accurate than all estimates for nutrient concentrations (Table 3), resulting in an accurate model prediction (Figure 4). Moreover, concentrations of P, Ca, S, Mn, and Cu were estimated with high accuracy (Table 3) and resulted in an excellent model prediction (Figure 4). Although accuracy for K and B estimates were not as high as for the nutrient concentrations described previously, estimates for K and B were still fairly accurate (Table 3) with a few errors (Figure 4). Estimates for Mg, Zn, and Fe concentrations were the least accurate and included a large number of errors than for nutrients mentioned previously.

Regarding PLSR coefficients, the absolute values indicate the contribution of each wavelength in the predictive models. In other words, the higher the absolute coefficient value, the greater the influence of wavelength on the model. In the N concentration model, the coefficients for wavelengths at 740 and 780 nm presented more absolute values (Figure 5). Moreover, the pattern of coefficient values observed in the N modelling is similar to that found for the P concentration modelling. However, the high absolute coefficient value was around 680 nm for the P concentration modelling.

For the K concentration modelling, the VIP method selected wavelengths in the regions for blue, green, red, and a small part of NIR. In this model, the coefficient assigned to 400 nm had a greater absolute value (Figure 5). Since the Ca concentration modelling used the IPW variable selection method, wavelengths in all hyperspectral reflectance regions were used in the modelling. Nevertheless, higher absolute coefficient values were assigned to wavelengths in the red edge and NIR regions.

The Mg concentration was estimated using wavelengths in the blue, red, red edge and NIR regions, with higher absolute values in the blue region and around 820 nm (Figure 5). The S concentration was predicted using almost all regions included in this study, with a gap between 706 and 870 nm and higher absolute coefficient values at 705 and 880 nm. Similarly, the B concentration was estimated using all regions of the electromagnetic spectrum with a higher absolute coefficient value around 790 nm.

Table 3 – Partial least squares (PLS) statistics for macronutrients.

PLS	N			P			K			Ca			Mg			S		
	PLS1[a]	IPW	VIP	**PLS1**[a]	IPW	VIP	PLS1	IPW	**VIP**[a]	PLS1	**IPW**[a]	VIP	PLS1	IPW	**VIP**[a]	PLS1	IPW	**VIP**[a]
Smoothed reflectance																		
N° LV	9	9	8	5	5	7	12	12	11	12	12	15	3	2	3	5	13	25
N° λ	*All*	155	1369	*All*	103	1178	*All*	366	1100	*All*	43	788	*All*	43	794	*All*	307	1524
$RMSE_{train}$	1.05	1.05	1.08	0.09	0.09	0.08	0.93	1.60	1.09	1.24	1.36	1.29	0.38	0.38	0.38	0.11	0.06	0.03
$RMSE_{cv}$	1.34	1.35	1.33	0.10	0.10	0.10	1.46	1.74	1.45	2.02	1.94	2.05	0.42	0.40	0.41	0.12	0.11	0.13
R^2_{train}	0.97	0.97	0.97	0.94	0.94	0.95	0.88	0.63	0.83	0.91	0.89	0.90	0.05	0.03	0.04	0.84	0.94	0.98
R^2_{cv}	0.95	0.95	0.95	0.92	0.93	0.93	0.70	0.56	0.70	0.76	0.78	0.75	0.03	0.01	0.01	0.78	0.81	0.76
Logarithmic transformation																		
N° LV	7	7	8	5	5	8	11	13	**11**	12	**12**	13	11	11	**11**	9	12	2
N° λ	*All*	57	1390	*All*	56	1346	*All*	181	**1063**	*All*	**756**	1096	*All*	113	**863**	*All*	123	1596
$RMSE_{train}$	1.10	1.10	1.00	0.09	0.09	0.07	0.90	0.81	**0.94**	1.10	**1.09**	1.24	0.25	0.26	**0.25**	0.09	0.06	0.11
$RMSE_{cv}$	1.30	1.30	1.27	0.10	0.10	0.10	1.38	1.48	**1.30**	1.81	**1.80**	1.91	0.39	0.40	**0.34**	0.12	0.10	0.12
R^2_{train}	0.97	0.97	0.97	0.94	0.94	0.96	0.88	0.91	**0.87**	0.93	**0.93**	0.91	0.57	0.54	**0.58**	0.89	0.95	0.81
R^2_{cv}	0.95	0.95	0.95	0.93	0.93	0.93	0.73	0.68	**0.76**	0.81	**0.81**	0.78	0.01	0.01	**0.22**	0.80	0.84	0.78
First derivative																		
N° LV	**6**	6	4	**6**	6	4	4	9	5	7	6	8	3	5	4	8	5	**16**
N° λ	***All***	93	1196	***All***	51	1270	*All*	50	1439	*All*	88	1377	*All*	61	869	*All*	67	**1218**
$RMSE_{train}$	**0.90**	0.98	1.07	**0.06**	0.07	0.08	1.35	1.06	1.24	1.24	1.42	1.32	0.35	0.28	0.32	0.06	0.10	**0.04**
$RMSE_{cv}$	**1.26**	1.34	1.25	**0.09**	0.10	0.09	1.67	1.52	1.63	1.85	1.94	1.90	0.42	0.41	0.40	0.10	0.11	**0.10**
R^2_{train}	**0.98**	0.97	0.97	**0.97**	0.96	0.96	0.74	0.84	0.78	0.91	0.88	0.90	0.18	0.48	0.31	0.95	0.87	**0.97**
R^2_{cv}	**0.95**	0.95	0.96	**0.93**	0.93	0.93	0.60	0.67	0.62	0.80	0.78	0.79	0.03	0.02	0.01	0.85	0.81	**0.85**
Second derivative																		
N° LV	4	4	4	4	4	3	3	2	7	3	3	4	3	3	3	10	7	9
N° λ	*All*	83	1200	*All*	54	1182	*All*	24	1192	*All*	66	1334	*All*	111	993	*All*	38	1101
$RMSE_{train}$	1.00	1.07	1.09	0.07	0.08	0.09	1.16	1.37	0.96	1.38	1.50	1.60	0.30	0.30	0.30	0.05	0.08	0.06
$RMSE_{cv}$	1.37	1.33	1.34	0.10	0.10	0.10	1.53	1.56	1.48	1.78	1.91	1.98	0.44	0.41	0.40	0.11	0.11	0.11
R^2_{train}	0.97	0.97	0.97	0.96	0.95	0.94	0.81	0.73	0.87	0.89	0.87	0.95	0.39	0.40	0.41	0.96	0.90	0.95
R^2_{cv}	0.95	0.95	0.95	0.92	0.92	0.93	0.66	0.65	0.69	0.81	0.78	0.77	0.08	0.02	0.01	0.81	0.82	0.81

[a]In bold = Reflectance pre-processing and variable selection method chosen for nutrient estimates. N = Nitrogen; P = Phosphorous; K = Potassium; Ca = Calcium; Mg = Magnesium; S = Sulphur; PLS1 = Hyperspectral reflectance in all wavelengths; IPW = Iterative predictor weighting partial least square; VIP = Variable importance in partial least squares regression projection; N° LV = Number of latent variables; N° λ = Number of wavelengths = $RMSE_{train}$ = Root mean square error of training; $RMSE_{cv}$ = Root mean square error of cross validation = R^2_{train} = Coefficient of determination of training; R^2_{cv} = Coefficient of determination of cross validation.

In the Zn concentration estimates, the higher absolute values were assigned to wavelengths for the blue and NIR regions, while the wavelengths for the blue, red, red edge and NIR regions were used to predict Mn concentration with higher absolute coefficient values at 680 nm and around the NIR region. The wavelengths in the blue region also had higher absolute coefficient values in the estimates for Fe and Cu concentrations. More specifically, in Cu concentration estimates, 520 nm was the wavelength with the highest absolute coefficient value.

Discussion

The pre-processing caused changes in different regions of the spectrum (Figure 3), which is important for modelling leaf nutrient concentration with large standard deviations (Table 2). A higher standard deviation for leaf reflectance indicates that these regions may be more favourable for extracting differences in independent samples and undertaking robust modelling. In the SR and LT pre-processes, the highest standard deviations at 550 and 700 nm and below 550 nm are comparable to those in several plant species (Asner et al., 2011; Dechant et al., 2017). Likewise, regions of leaf reflectance related to physiological traits, like the wavelengths at around 700 nm (Horler et al., 1983), presented higher standard deviations in FD and SD (Figures 3 C, D).

In this study, the number of LV used to estimate nutrient concentrations ranged between 2 and 16 (Table 3, 4). Pandey et al. (2017) found similar results when predicting nutrient concentrations using hyperspectral images in maize and soybean. In the PLSR modelling for hyperspectral data and biochemical traits in plants, it is usual to use more than 10 LV (Nguyen et al., 2006; Ramoelo et al., 2013). However, high LV values may undergo overfitting, which could compromise the model predictive power (Wold et al., 2001). On the other hand, in some studies, biochemical traits were predicted with less than 10 LV, for example in *Cirtrus sinensis* (L.) Osbeck cv Tarocco, *Picea rubens* Sarg., *Abies balsamea* (L.) Mill. and *Beta vulgaris* L. (Swiss chard) (Abdel-Rahman

Table 4 – Partial least squares (PLS) statistics for micronutrients.

PLS	B			Zn			Mn			Fe			Cu		
	PLS1	IPW	**VIP**[a]	PLS1	IPW	**VIP**[a]	PLS1	IPW	**VIP**[a]	PLS1	IPW	**VIP**[a]	**PLS1**[a]	IPW	VIP
Smoothed reflectance															
N° LV	10	11	9	3	11	3	14	13	21	12	3	**11**	14	11	5
N° λ	*All*	62	1571	*All*	94	524	*All*	58	821	*All*	140	**672**	*All*	80	821
$RMSE_{train}$	18.97	19.87	21.96	4.93	3.42	4.56	225.3	246.7	184.1	16.64	27.10	**17.84**	0.38	0.51	0.66
$RMSE_{cv}$	27.75	26.23	27.88	5.30	5.33	4.87	451.7	455.0	448.2	28.56	29.72	**25.00**	0.75	0.75	0.76
R^2_{train}	0.78	0.76	0.70	0.23	0.63	0.34	0.94	0.92	0.96	0.71	0.24	**0.67**	0.92	0.86	0.75
R^2_{cv}	0.52	0.57	0.52	0.11	0.10	0.25	0.75	0.74	0.75	0.16	0.09	**0.35**	0.68	0.68	0.68
Logarithmic transformation															
N° LV	9	9	9	5	5	6	12	12	12	11	10	9	**12**	12	11
N° λ	*All*	59	1621	*All*	415	525	*All*	256	1219	*All*	511	554	***All***	68	1219
$RMSE_{train}$	20.55	20.61	21.24	4.57	4.56	4.41	250.9	261.5	315.4	17.29	20.06	19.25	**0.39**	0.40	0.47
$RMSE_{cv}$	27.22	27.56	28.17	5.24	5.24	5.04	432.9	448.8	493.7	27.25	29.12	25.22	**0.63**	0.65	0.65
R^2_{train}	0.74	0.74	0.72	0.34	0.23	0.39	0.92	0.92	0.88	0.69	0.58	0.62	**0.92**	0.91	0.88
R^2_{cv}	0.54	0.53	0.51	0.13	0.13	0.20	0.77	0.75	0.70	0.23	0.12	0.34	**0.78**	0.77	0.76
First derivative															
N° LV	6	6	**7**	5	5	5	9	9	**7**	2	4	4	6	6	10
N° λ	*All*	83	**1277**	*All*	88	730	*All*	411	**1112**	*All*	44	1377	*All*	237	1252
$RMSE_{train}$	17.67	18.10	**16.74**	3.71	3.77	3.84	237.1	240.7	**277.5**	27.37	25.51	26.05	0.52	0.53	0.43
$RMSE_{cv}$	24.41	24.70	**22.66**	5.02	5.02	4.77	421.4	428.7	**407.5**	30.92	31.69	31.43	0.73	0.74	0.70
R^2_{train}	0.81	0.80	**0.83**	0.56	0.55	0.53	0.93	0.93	**0.90**	0.22	0.32	0.30	0.85	0.85	0.90
R^2_{cv}	0.63	0.62	**0.68**	0.20	0.20	0.28	0.78	0.77	**0.79**	0.01	0.01	0.01	0.70	0.69	0.72
Second derivative															
N° LV	4	4	3	3	4	**2**	7	7	7	6	20	5	8	7	6
N° λ	*All*	61	1309	*All*	141	**628**	*All*	57	734	*All*	30	683	*All*	18	1245
$RMSE_{train}$	17.64	17.73	22.08	3.86	3.68	**4.04**	240.2	341.7	335.4	17.09	18.45	18.44	0.36	0.64	0.46
$RMSE_{cv}$	23.77	23.66	25.35	5.00	5.06	**4.67**	419.8	523.0	524.1	29.83	30.62	25.77	0.70	0.83	0.61
R^2_{train}	0.81	0.81	0.70	0.53	0.57	**0.48**	0.93	0.86	0.86	0.70	0.65	0.65	0.93	0.77	0.88
R^2_{cv}	0.65	0.65	0.60	0.21	0.19	**0.31**	0.78	0.66	0.66	0.07	0.03	0.31	0.73	0.61	0.78

[a]In bold = Reflectance pre-processing and variable selection method chosen for nutrient estimates. B = Boron; Zn = Zinc; Mn = Manganese; Fe = Iron; Cu = Copper; PLS1 = Hyperspectral reflectance in all wavelengths; IPW = Iterative predictor weighting partial least square; VIP = Variable importance in partial least squares regression projection; N° LV = Number of latent variables; N° λ = Number of wavelengths = $RMSE_{train}$ = Root mean square error of training; $RMSE_{cv}$ = Root mean square error of cross validation = R^2_{train}: Coefficient of determination of training; R^2_{cv} = Coefficient of determination of cross validation.

et al., 2014; 2017; Menesatti et al., 2010). Therefore, the LOOCV was performed to evaluate the model predictive capacity (Tables 3, 4).

In general, the variable selection methods increased the PLSR models (Tables 3, 4). In estimates for K, Mg, S, B, Zn, Mn, and Fe concentrations, greater accuracy was obtained with the VIP method. The IPW method was only used with the highest accuracy for estimates of Ca concentration. In general, variable selection methods increase accuracy of hyperspectral modelling and the use of whole wavelengths has a negative influence on the predictive PLSR capacity (Abdel-Rahman et al., 2017; Filzmoser et al., 2012). However, estimates of N, P, and Cu concentrations were more accurate using the wavelengths for the entire reflectance spectrum (400 – 900 nm).

It is known that the relationship between N, P, and Cu concentrations and the reflectance in the studied region of reflectance spectrum is very strong (Pimstein et al., 2011; Mahajan et al., 2014; Oliveira et al., 2017). For this reason, even using the whole reflectance spectrum, the model accuracy for these nutrients was high (Figure 4). Nevertheless, the variable selection methods increase accuracy for estimates of N and P concentrations in oilseed rape (*Brassica napus* L.) (Zhang et al., 2013).

Estimates of N concentration had the greatest accuracy, followed by estimates of P concentration (Figure 4). In general, the concentrations of these nutrients are the most accurate in spectral data estimates (Asner et al., 2011; Pandey et al., 2017; Ramoelo et al., 2013). Both nutrients are highly related to photosynthetic traits in plants (Dechant et al., 2017). However, the same result may not be observed in all plant species (Abdel-Rahman et al., 2017; Menesatti et al., 2010). In addition, Asner et al. (2011), using spectroscopy to determine nutrients in humid tropical forest canopies, were only successful in estimating N and P concentrations with less accuracy.

The higher absolute coefficient values related to wavelengths in the red and red edge regions explained the strong relationship between these wavelengths and N and P concentrations (Figure 5). The red and red edge regions are widely used for estimates of leaf nutrient for several forest and agricultural crops (Mutanga et al., 2005; Oliveira et al., 2017; Schlemmer et al., 2013; Yu et al., 2014).

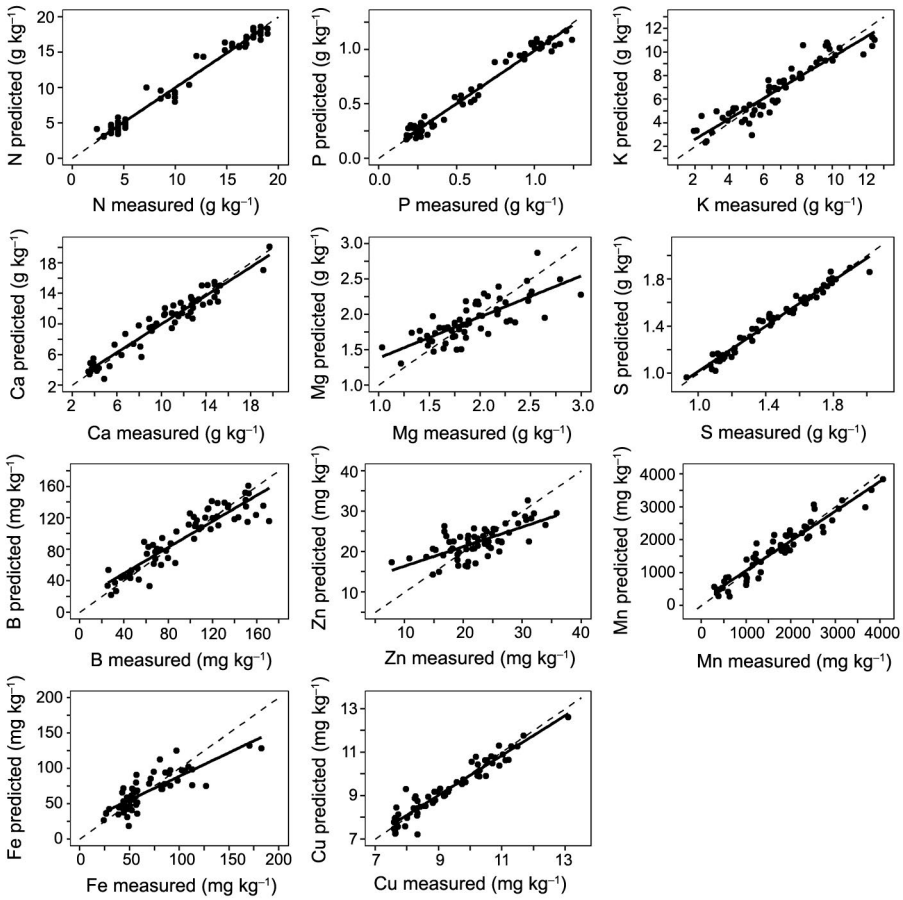

Figure 4 – Relationship between leaf nutrient concentrations measured and predicted.

Regarding estimates of K concentration, several wavelengths in the VIS and NIR regions were selected (Figure 5). These regions are related to changes in K concentration in leaves (Guo et al., 2017; Zhai et al., 2013). In the VIS region, the blue, green, red and red edge regions were selected to estimate K concentration using the VIP method (Figure 5). These same regions were used to classify canola (*Brassica napus* L.), plants with K deficiency using a hyperspectral camera (Severtson et al., 2016).

The red edge and NIR presented higher absolute coefficient values for estimates of Ca concentration (Figure 5). This is probably due to the relationship between the structural components and the energy in these regions. One of the functions of Ca in higher plants is cell wall synthesis and cell membrane integrity; thus, Ca has a structural function (Marschner, 1995). Moreover, leaf reflectance in NIR regions demonstrates a strong relationship with the structural components of plants (Gates et al., 1965).

Higher absolute coefficient values in the blue region and near 820 nm were obtained for estimates of Mg concentration (Figure 5). Mg is the central atom in the chlorophyll molecule ring and indirectly, Mg concentration influences leaf reflectance in the VIS region changing chlorophyll concentration in leaves (Marschner, 1995). Leaf chlorophyll has absorption peaks in the electromagnetic spectrum in the blue and red regions (Lichtenthaler, 1987). Most likely, higher absolute coefficient values are in wavelengths in the blue region.

Concentration of S was estimated using almost all regions in the electromagnetic spectrum with higher absolute coefficient values closer to 700 and 880 nm (Figure 5). Wavelengths in these regions have been used in the management of S in wheat (*Triticum* aestivum L.) and rice (*Oryza sativa* L.) (Mahajan et al., 2014; 2017). Using the VIP method, wavelengths in all regions of the spectrum were selected for estimates of B concentration with higher absolute coefficient values at approximately 790 nm (Figure 5). Similar to Ca, B plays a structural role in plants (Marschner, 1995) and has a strong relationship with the NIR region, as explained previously. In general, there have been problems with hyperspectral sensors to estimate B concentrations, namely in soybean (*Glicine max* L., variedade Thorne) and maize (*Zea mays*, B73 inbred) crops (Pandey et al., 2017). In this study, although accuracy of B concentration estimates was not the highest, leaf concentration was estimated with R^2_{cv} = 0.68 and $RMSE_{cv}$ = 22.66 mg kg^{-1} (Table 4 and Figure 4).

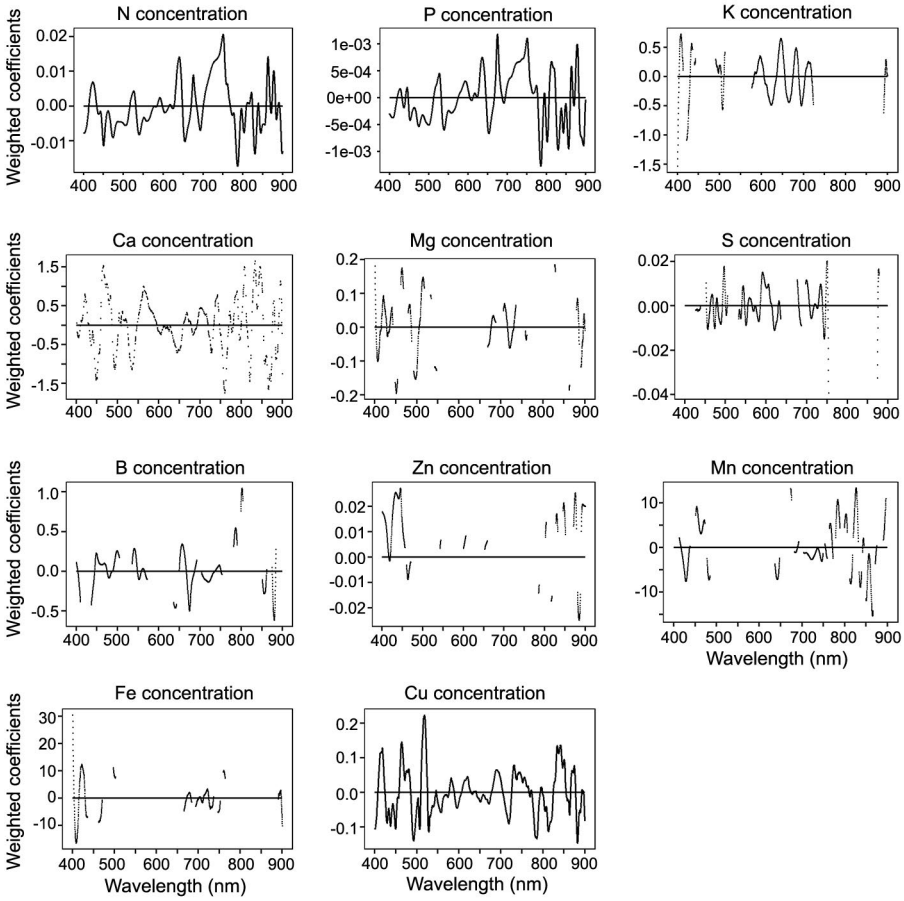

Figure 5 – Coefficient weights for wavelengths for selected partial least squares regression (PLSR) models.

Estimates of Zn concentration had the lowest accuracy with higher absolute coefficient values related to wavelengths in the blue and NIR region (Table 4 and Figure 5). In the plant, Zn is as metal ion; thus, higher precision in its estimate using leaf hyperspectral reflectance was expected (Pandey et al., 2017). However, estimates of Zn concentration using leaf reflectance are not frequently reported and the weak relationship may be a reason for this. For instance, Zn deficiency was not detected by reflectance in soybean plants under Zn omission in the growth solution (Adams et al., 2000a, b).

Higher absolute coefficient values were attributed to wavelengths in the blue, red, red edge and NIR regions to estimate Mn concentration (Figure 5), while higher absolute coefficient values were related to wavelengths in the blue region for estimates of Fe concentration. Estimates of Cu concentration had higher absolute coefficient values in wavelengths at around 520 nm (Figure 5). Zhang et al. (2017) also selected a wavelength in the green region to estimate Cu concentration in vegetation stress. Nevertheless, as described for Zn, the relationship between leaf reflectance and micronutrients is little explained in the literature (Adams et al., 2000a; Pandey et al., 2017; Zhang et al., 2017).

Only in estimates of B and K concentration, wavelengths around 425 – 435 nm and 460 – 468 nm were not used, respectively (Figure 5). Since 91 % of macro- and micronutrients models used leaf reflectance in these wavelengths, it can be an indicative of the importance of these wavelengths in leaf nutrient modelling of concentrations. On the other hand, only estimates of N, P, and Cu concentrations used leaf reflectance in the wavelengths 517 – 532 nm, 763 – 778 nm, 790 – 798 nm, 805 – 819 nm, 832 – 841 nm, and 866 – 890 nm. As previously discussed, these wavelengths are avoided to estimate most nutrient concentrations in *Eucalyptus* clones.

In this study, we used more than 2,000 predictors in 61 samples of leaf reflectance and nutrient concentrations. Although there are more predictors than dependent variables, PLSR is a robust method than can overcome this problem and deal with small sets of samples and large numbers of estimated parameters associated to models for hyperspectral data (Abdel-Rahman et al., 2017; Ramoelo et al., 2013). Moreover, more advanced analytical and sensing methods were previously used to estimate macro and micronutrient concentrations in several plant species (Pullanagari et al., 2016; Abdel-Rahman et al., 2017; Pandey et al., 2017; Wang et al.,

2017). However, the use of PLSR in leaf hyperspectral reflectance presented a rapid and accurate method to estimate leaf macro- and micronutrient concentrations in *Eucalyptus* clones.

Conclusion

Leaf hyperspectral reflectance was capable to estimate leaf nutrient concentration in *Eucalyptus* clones using partial least squares regression and variable selection methods. In general, the variable selection methods increased accuracy of nutrient concentration estimates. Although the number of studies on relationships between nutrient concentrations and leaf reflectance is increasing, the biochemical meaning behind their models is still poorly described.

The NIR region evaluated in this work was up to 900 nm, the complete NIR and SWIR regions (until 2500 nm) may have wavelengths that could be useful to estimate accurately leaf nutrient concentrations in *Eucalyptus* clones. The authors recommend that the results be up-scaled to another time and space points to evaluate the operational use of this approach.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Universidade Federal dos Vales do Jequitinhonha e Mucuri -UFVJM, Conselho Nacional de Desenvolvimento Científico e Tecnológico (SWE scholarship holder), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), and Gerdau Florestal SA.

Authors' Contributions

Conceptualization: Oliveira, L.F.R.; Santana, R.C. Data acquisition: Oliveira L. F.R. Data analysis: Oliveira, L.F.R. Design of methodology: Oliveira L.F.R. Writing and editing: Oliveira, L.F.R.; Santana, R.C.

References

Abdel-Rahman, E.M.; Mutanga, O.; Odindi, J.; Adam, E.; Odindo, A.; Ismail, R. 2014. A comparison of partial least squares (PLS) and sparse PLS regressions for predicting yield of Swiss chard grown under different irrigation water sources using hyperspectral data. Computers and Electronics in Agriculture 106: 11–19.

Abdel-Rahman, E.M.; Mutanga, O.; Odindi, J.; Adam, E.; Odindo, A.; Ismail, R. 2017. Estimating Swiss chard foliar macro- and micronutrient concentrations under different irrigation water sources using ground-based hyperspectral data and four partial least squares (PLS)-based (PLS1, PLS2, SPLS1, SPLS2) regression algorithms. Computers and Electronics in Agriculture 132: 21–31.

Adams, M.L.; Norvell, W.A.; Philpot, W.D.; Peverly, J.H. 2000a. Spectral detection of micronutrient deficiency in 'Bragg' soybean. Agronomy Journal 92: 261–268.

Adams, M.L.; Norvell, W.A.; Philpot, W.D.; Peverly, J.H. 2000b. Toward the discrimination of manganese, zinc, copper, and iron deficiency in 'Bragg' soybean using spectral detection methods. Agronomy Journal 92: 268–274.

Asner, G.P.; Martin, R.E.; Knapp, D.E.; Tupayachi, R.; Anderson, C.; Carranza, L.; Martinez, P.; Houcheime, M.; Sinca, F.; Weiss, P. 2011. Spectroscopy of canopy chemicals in humid tropical forests. Remote Sensing of Environment 115: 3587–3598.

Blackburn, G.A. 2007. Hyperspectral remote sensing of plant pigments. Journal of Experimental Botany 58: 855–867.

Cawley, G.C.; Talbot, N.L.C. 2003. Efficient leave-one-out cross-validation of Kernel Fischer discriminant classifiers. Pattern Recognition 36: 2585–2592.

Dechant, B.; Cuntz, M.; Vohland, M.; Schulz, E.; Doktor, D. 2017. Estimation of photosynthesis traits from leaf reflectance spectra: correlation to nitrogen concentration as the dominant mechanism. Remote Sensing of Environment 196: 279–92.

Dell, B. 1996. Diagnosis of nutrient deficiencies in eucalypts. p. 417-440. In: Attiwill, P.M.; Adams, M.A., eds. Nutrition of eucalypts. CSIRO, Camberra, Australia.

Filzmoser, P.; Gschwandtner, M.; Todorov, V. 2012. Review of sparse methods in regression and classification with application to chemometrics. Journal of Chemometrics 26: 42–51.

Forina, M.; Casolino, C.; Pizarro Millan, C. 1999. Iterative predictor weighting (IPW) PLS: a technique for the elimination of useless predictors in regression problems. Journal of Chemometrics 13: 165–184.

Gates, D.M.; Keegan, H.J.; Schleter, J.C.; Weidner, V.R. 1965. Spectral properties of plants. Applied Optics 4: 11–20.

Guo, X.; Zhu, X.; Li, C.; Wei, Y.; Yu, X.; Zhao, G.; Sun, H. 2017. Hyperspectral inversion of potassium concentration in apple leaves based on vegetation index. Agricultural Sciences 8: 825–836.

Horler, D.N.; Dockray, M.; Barber, J. 1983. The red edge of plant leaf reflectance. International Journal of Remote Sensing 4: 273–288.

Lu, S.; Lu, F.; You, W.; Wang, Z.; Liu, Y.; Omasa, K. 2018. A robust vegetation index for remotely assessing chlorophyll content of dorsiventral leaves across several species in different sensors. Plant Methods 14: 1-15.

Lichtenthaler, H.K. 1987. Chlorophylls and carotenoids: the pigments of photosynthetic biomembranes. p. 350-382. In: Colowick, S.P.; Kaplan, N.O., eds. Methods in enzymology. Academic Press, San Diego, CA, USA.

Mahajan, G.R.; Sahoo, R.N.; Pandey, R.N.; Gupta, V.K.; Kumar, D. 2014. Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (*Triticum aestivum* L.). Precision Agriculture 15: 499–522.

Mahajan, G.R.; Pandey, R.N.; Sahoo, R.N.; Gupta, V.K.; Datta, S.C.; Kumar, D. 2017. Monitoring nitrogen, phosphorus and sulphur in hybrid rice (*Oryza sativa* L.) using hyperspectral remote sensing. Precision Agriculture 18: 736–761.

Marschner, H. 1995. Mineral Nutrition of Higher Plants. Academic Press, London, UK.

Mariotti, M.; Ercoli, L.; Masoni, A. 1996. Spectral properties of iron-deficient corn and sunflower leaves. Remote Sensing of Environment 58: 282–288.

Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. 2012. A review of variable selection methods in Partial Least Squares Regression. Chemometrics and Intelligent Laboratory Systems 118: 62–69.

Menesatti, P.; Antonucci, F.; Pallottino, F.; Roccuzzo, G.; Allegra, M.; Stagno, F.; Intrigliolo, F. 2010. Estimation of plant nutritional status by VIS-NIR spectrophotometric analysis on orange leaves [Citrus sinensis (L) Osbeck cv Tarocco]. Biosystems Engineering 105: 448–453.

Mutanga, O.; Skidmore, A.K.; Kumar, L.; Ferwerda, J. 2005. Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. International Journal of Remote Sensing 26: 1093–1108.

Nguyen, H.T.; Kim, J.H.; Nguyen, A.T.; Nguyen, L.T.; Shin, J.C.; Lee, B-W. 2006. Using canopy reflectance and partial least squares regression to calculate within-field statistical variation in crop growth and nitrogen status of rice. Precision Agriculture 7: 249–264.

Oliveira, L.F.R.; Oliveira, M.L.R.; Gomes, F.S.; Santana, R.C. 2017. Estimating foliar nitrogen in *Eucalyptus* using vegetation indexes. Scientia Agricola 74: 142–147.

Pandey, P.; Ge, Y.; Stoerger, V.; Schnable, J.C. 2017. High throughput *in vivo* analysis of plant leaf chemical properties using hyperspectral imaging. Frontiers in Plant Science 8: 1–12.

Pimstein, A.; Karnieli, A.; Surinder, K.B.; Bonfil, D.J. 2011. Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. Field Crops Research 121: 125–135.

Pullanagari, R.R.; Kereszturi, G.; Yule, I.J. 2016. Mapping of macro and micro nutrients of mixed pastures using airborne AisaFENIX hypespectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing 117: 1–10.

Ramoelo, A.; Skidmore, A.K.; Chon M.A.; Mathieu, R.; Heitkönig, I.M.A.; Dudeni-Tlhone, N.; Schlerf, M.; Prins, H.H.T. 2013. Non-linear partial least squares regression increases the estimation accuracy of grass nitrogen and phosphorus using *in situ* hyperspectral and environmental data. ISPRS Journal of Photogrammetry and Remote Sensing 82: 27–70.

Savitzky, A.; Golay, M.J.E. 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36: 1627–1639.

Saur, E.; Nambiar, E.K.S.; Fife, D.N. 2000. Foliar nutrient retranslocation in *Eucalyptus globulus*. Tree Physiology 20: 1105–1112.

Schlemmer, M.; Gitelson, A.; Schepers, J.; Ferguson, R.; Peng, Y.; Shanahan, J.; Rundquist, D. 2013. Remote estimation of nitrogen and chlorophyll concentrations in maize at leaf and canopy levels. International Journal of Applied Earth Observation and Geoinformation 25: 47–54.

Severtson, D.; Callow, N.; Flower, K.; Neuhaus, A.; Olejnik, M.; Nansen, C. 2016. Unmanned aerial vehicle canopy reflectance data detects potassium deficiency and green peach aphid susceptibility in canola. Precision Agriculture 17: 659–677.

Yu, K-Q.; Zhao, Y-R.; Shao, Y-N.; Liu, F.; He, Y. 2014. Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant. PLoS ONE 9: e0134071.

Wang, L.; Zhou, X.; Zhu, X.; Guo, W. 2017. Estimation of leaf nitrogen concentration in wheat using the MK-SVR algorithm and satellite remote sensing data. Computers and Electronics in Agriculture 140: 327–337.

Wold, S.; Johansson, E.; Cocchi, M. 1993. PLS: Partial least squares projections to latent structures. p. 523-550. In: Kubinyi, H., ed. 3D QSAR in drug design. ESCOM Science, Leiden, Netherlands.

Wold, S.; Sjöström, M.; Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58: 109–130.

Zhai, Y.; Cui, L.; Zhou, X.; Gao, Y.; Fei, T.; Gao, W. 2013. Estimating of nitrogen, phosphorus, and potassium concentrations in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-squares regression and support vector machine regression methods. International Journal of Remote Sensing 34: 2502–2518.

Zhang, X.; Liu, F.; He, Y.; Gong, X. 2013. Detecting macronutrients concentration and distribution in oilseed rape leaves based on hyperspectral imaging. Biosystems Engineering 115: 56-65.

Zhang, C.; Ren, H.; Qin, Q.; Ersoy, O.K. 2017. A new narrow band vegetation index for characterizing the degree of vegetation stress due to copper: the copper stress vegetation index (CSVI). Remote Sensing Letters 8: 576–585.