

Using unmanned aerial vehicle and machine learning algorithm to monitor leaf nitrogen in coffee

Taya Cristo Parreiras¹ , Guilherme Henrique Expedito Lense¹ , Rodrigo Santos Moreira¹ ,
Derielsen Brandão Santana² , Ronaldo Luiz Mincato¹ 

¹Universidade Federal de Alfenas/UNIFAL, Alfenas, MG, Brasil

²Instituto Brasileiro de Geografia e Estatística/IBGE, Alfenas, MG, Brasil

Contact authors: tayacristo1@gmail.com, guilhermeelense@gmail.com, rodrigomagro@yahoo.com.br, derielsen@hotmail.com, ronaldo.mincato@unifal-mg.edu.br

Received in March 23, 2020 and approved in May 26, 2020

ABSTRACT

Nitrogen is an essential element for coffee production. However, when fertilization do not consider the spatial variability of the agricultural parameters, it can generate economic losses, and environmental impacts. Thus, the monitoring of nitrogen is essential to the fertilizing management, and remote sensing based on unmanned aerial vehicles imagery has been evaluated for this task. This work aimed to analyze the potential of vegetation indices of the visible range, obtained with such vehicles, to monitor the nitrogen content of coffee plants in southern Minas Gerais, Brazil. Therefore, we performed leaf analysis using the Kjeldahl method, and we processed the images to produce the vegetation indices using Geographic Information Systems and photogrammetry software. Moreover, the images were classified using the Color Index of Vegetation and the Maximum Likelihood Classifier. As estimator tool, we created Random Forest models of classification and regression. We also evaluated the Pearson correlation coefficient between the nitrogen and the vegetation indices, and we performed the analysis of variance and the Tukey-Kramer test to assess whether there is a significant difference between the averages of these indices in relation to nitrogen levels. However, the models were not able to predict the nitrogen. The regression model obtained a $R^2 = 0.01$. The classification model achieved an overall accuracy of 0.33 (33%), but it did not distinguish between the different levels of nitrogen. The correlation tests revealed that the vegetation indices are not correlated with the nitrogen, since the best index was the Green Leaf Index ($R = 0.21$). However, the image classification achieved a Kappa coefficient of 0.92, indicating that the tested index is efficient. Therefore, visible indices were not able to monitor the nitrogen in this case, but they should continue to be explored, since they could represent a less expensive alternative.

Keys words: Vegetation indices; RGB; machine learning; *Coffea arabica*.

1 INTRODUCTION

Coffee production can be impacted by several factors such as plagues, climate variations, planting system and density, terrain slope, soil quality, as well as the physiological characteristics of the plants (Ferraz et al., 2012). Therefore, management planning and the monitoring of the crops are key-factors to guarantee good productivity and to reduce socioeconomics and environmental impacts (Näsi et al., 2018).

Nitrogen is an essential element to crops, once it is part of the structure, the metabolism, and the osmoregulation of the plants, playing an essential role in photosynthesis (Taiz et al., 2017), therefore, is limiting factor to coffee production. On the other hand, when the N disponibility highly overcomes the plant's necessity, it can cause soil and atmosphere pollution, which can also lead to eutrophication of water bodies (Ballester et al., 2017).

Thus, precision agriculture (PA) has been gained space on the field, approaching science and technology to agricultural practices, helping to reduce costs, improve productivity by area, and minimize environmental impacts through the planning of actions as fertilization, irrigation, and disease control (Vega et al., 2015), and remote sensing is an important technology of the PA (Hunt Junior et al., 2018).

Therefore, researchers have been analyzing the potential vegetation indices (VI) obtained with visible sensors mounted

to unmanned aerial vehicles (UAV), along with digital image processing techniques, to monitor the spatial variability of the nitrogen content in different kinds of crops like barley (Escalante et al., 2019), grasses (Caturegli et al., 2019; Näsi et al., 2018), potatoes (Hunt Junior et al., 2018), wheat (Schirrmann et al., 2016), and maize (Zhang et al., 2020).

The UAV can help to overcome issues related to the leaf analysis, which may be expensive and time-consuming, since the mounted sensors have high resolution and the crops can be monitored during the entire season (Maes; Steppe, 2019). The VI are equations relating the spectral behavior of the surface vegetation, based on the plant responses in the different ranges of the electromagnetic spectrum, which is different according to the plant health (Caturegli et al., 2019; Ponzoni; Shimabukuro; Kuplich, 2012).

The VI relating the red, blue, and green reflectance have also been tested to detect the severity of infections, weed occurrence, to monitor growth vigor, and presented good perspectives to monitor nutrient status and yield prediction. Moreover, they represent a less expensive alternative when other bands, like Red-Edge and Near Infrared, are not available (Beniaich et al., 2019; Maes; Steppe, 2019).

To the data analysis, machine learning algorithms like Random Forest (RF), used in this study, and neural networks like Deep Learning Convolutional Neural Network (DCNN)

are gaining space and presenting promising results (Escalante et al., 2019; Osco et al., 2019; Näsi et al., 2018).

Although this approach is becoming more present, at the best of our knowledge, there are no published studies involving N content monitoring using UAV in coffee crops, which is an essential economic activity in Brazil, especially at southern Minas Gerais. Furthermore, researchers have been mainly applying multispectral VI, which require more expensive sensors, and makes the RGB-based imaging, a less expensive alternative.

Comparing the individual relationship of the spectral indices with the ground measured N, and evaluating them with Random Forest classification and regression models, our work aimed to assess the potential of RGB-based VI in monitoring N in a coffee-growing area, in Divisa Nova Municipality, Minas Gerais, Brazil's southeast.

2 MATERIAL AND METHODS

Study area and leaf sampling

The research was carried out in a 1 ha plot of coffee crops, located in Divisa Nova, southern Minas Gerais state, Brazil (21°20'S and 46°10'W, Datum SIRGAS 2000) (Figure 1). The Santo André farm has about 80 ha and produces coffee (Red Catuaí variety - *Coffea arabica*) in transition to organic type of production, with plants 0.65 m x 3.5 m apart.

According to Köppen classification, the climate is characterized by dry winters and tempered summers (Cwb)

(Alvares, 2013), with an average annual temperature between 20 and 22 °C, and a total annual rainfall in 2018 between 1800 and 2000 mm (Instituto Nacional de Meteorologia – INMET, 2020). The altitude of the area ranges from 820 to 900 m, with an average elevation of 860 m. The soil of the area is classified as Oxisol.

In the first stage of the study, we distributed 88 points of sampling, as shown in Figure 1, where 8 leaves were collected, 4 from each side of the plants. The healthy-looking leaves were removed from the plagiotropic branches, in the middle third of the coffee trees. This sampling was carried out on May 29th, 2019, between 7:30 am and 10:30 am. Moreover, to determine the N content, we conducted triplicates of the Kjeldahl method for leaf N quantification.

Along with the N amount expressed in kg kg^{-1} , obtained with the analysis, a parallel database was created to turn these measures into classes, where we classified them in three levels of nitrogen: deficient (if $N < 0.0288 \text{ kg kg}^{-1}$), appropriate ($N 0.0288 - 0.0322 \text{ kg kg}^{-1}$), and excessive ($N > 0.0322 \text{ kg kg}^{-1}$), as proposed by Guimarães et al. (1999). This new database was used to perform the Random Forest (RF) classification model, which will be discussed below.

On the same leaves that were analyzed, we used the chlorophyll meter SPAD 502, performing 2 readings per leaf, and registering the average for each point, that is, 16 readings per point. The SPAD values are given by the difference between the transmittance of red and infrared light that penetrated the leaf (Garza et al., 2020), and the objective of obtaining these values was to verify their relationship with the measured amounts of N.

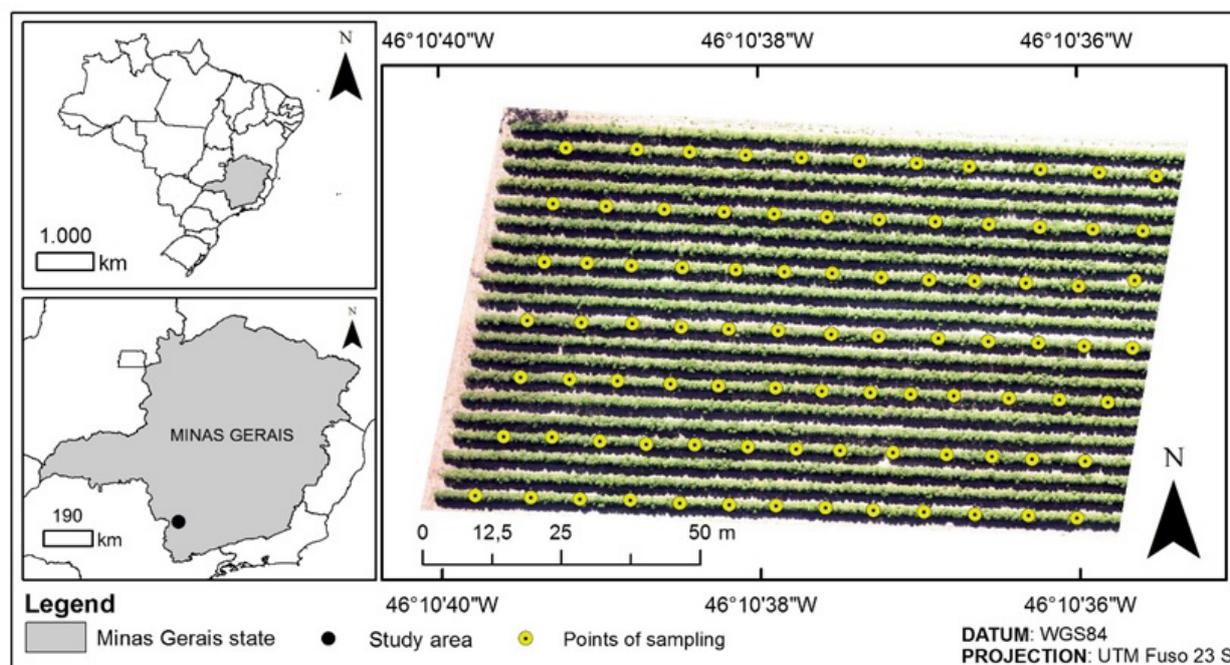


Figure 1: Study area, at the Santo André Farm, southern Minas Gerais, Brazil, and location of the points of sampling for leaf analysis and SPAD readings.

Image acquisition, processing, and vegetation indices generation

The UAV used in this study was a multirotor Phantom 4 Advanced (DJI, Shenzhen, China) that has a visible (RGB) sensor 1" CMOS of 20 megapixels. The flights were planned using the Pix4D Capture software (Pix4D Inc., San Francisco, USA), where the area coverage grid was created, and the flight parameters (altitude, overlap, NADIR field of view, etc.) were determined. The uncalibrated images were taken on the same day as the leaf sampling (May 29th, 2019), at noon, with a clear sky and no clouds. The parameters of the flight are shown in Table 1.

Table 1: Date and parameters of the flight.

Flight date	Altitude	Overlap/ Sidelap	Average GSD* cm px ⁻¹	Flight Time
May 29th, 2019	50 m	80/80 (%)	1.27	0.12 h

*GSD: Ground Sample Distance by pixel.

Firstly, the acquired images were processed in the Pix4D Mapper 4.4.12 software (Pix4D Inc., San Francisco, USA), where the 171 photos were combined to produce the georeferenced orthomosaic and the digital surface model (DSM) of the entire study area. Using the DSM, we generate 100 points well distributed over the area to extract values of elevation, ignoring the values regarding the plants. Then, using these values, we interpolated them with the IDW tool (Environmental Systems Research Institute - Inc. – ESRI, 2017) to create the Digital Elevation Model (DEM) of the area.

Then, the bands of the orthomosaic were separated using the Raster Calculator tool of the QGIS 3.8.3 (QGIS Development Team, 2020), which generated three spectral maps: Red, Green, and Blue. The separated bands were normalized, also using the Raster Calculator, to determine the chromatic levels, as proposed by Arroyo, Guijarro and Pajares (2016) with Equation 1.

$$\frac{\text{Red}}{\text{Red}+\text{Green}+\text{Blue}}, \frac{\text{Green}}{\text{Red}+\text{Green}+\text{Blue}} \text{ and } \frac{\text{Blue}}{\text{Red}+\text{Green}+\text{Blue}} \quad (1)$$

Image classification

To remove as much background interference as possible, isolating the spectral response of coffee trees, we performed a supervised classification using the Color Index of Vegetation (CIVE) (Kataoka et al., 2003), given by Equation 2. This index is indicated to generate a contrast between plants and soil, and has a good performance to estimate the fraction of vegetation cover (Beniaichi et al., 2019).

$$\text{CIVE} = 0.441 * \text{Red} - 0.881 * \text{Green} + 0.385 * \text{Blue} + 18.78745 \quad (2)$$

Therefore, using the Image Classification tool, from the Spatial Analyst extension of the ArcMap 10.5 (ESRI, 2017), we collected 75 training samples from the soil and the coffee plants to perform the classification. The algorithm chosen for the process was the Maximum Likelihood Classifier (MLC), which uses the training samples to delimit homogeneous regions, pixel by pixel, to calculate the mean and the covariance matrix, estimating the probability of determined pixel belongs to a given class (Cechim Júnior; Johann; Antunes, 2017).

Moreover, to verify the accuracy of this classification, 100 random points were generated in the study area to verify the agreement index between the real and the classified image, given by the Kappa coefficient (k) through Equation 3 (Cohen, 1960), measured using the “kappa2” and “kappam.fleiss” functions, from the “irr” package of the R software (Gamer et al., 2019; R Core Team, 2019).

$$k = P_o - P_e / (1 - P_e) \quad (3)$$

In which: P_o = proportion of agreement between classified and real; P_e = proportion of units in which concordance is expected by the chance.

Vegetation indices generation

The classification of the previous step, which resulted in a binary raster image, was converted into a polygon file, from which only those polygons referring to the coffee lines were selected. Thus, the vegetation indices could be calculated and generated only within the areas of interest, that is, those occupied by the coffee trees.

The vegetation indices (VI) selected for the study were the following: Excess Green (ExG), Green Leaf Index (GLI), Visible Atmospheric Resistance Index (VARI), Normalized Green-Red Difference Index (NGRDI) and Color Index of Vegetation (CIVE). The VI were generated executing their corresponding equations (Table 2), also using the Raster Calculator tool, with the normalized bands.

Table 2: Vegetation Indices (VI), equations and references.

VI	Equation	Reference
GLI	$\frac{2 * \text{Green} - \text{Red} - \text{Blue}}{2 * \text{Green} + \text{Red} + \text{Blue}}$	Louhaichi et al. (2001)
ExG	$2 * \text{Green} - \text{Red} - \text{Blue}$	Woebbecke et al. (1995)
NGRDI	$\frac{\text{Green} - \text{Red}}{\text{Green} + \text{Red}}$	Tucker (1979)
VARI	$\frac{\text{Green} - \text{Red}}{\text{Green} + \text{Red} - \text{Blue}}$	Gitelson et al. (2002)
CIVE	$0.441 * \text{Red} - 0.881 * \text{Green} + 0.385 * \text{Blue} + 18.78745$	Kataoka et al. (2003)

Where: GLI (Green Leaf Index), ExG (Excess Green), NGRDI (Normalized Difference Green Red Index), VARI (Visible Atmospherically Resistant Index), and Color Index of Vegetation (CIVE).

To complete the datasets used to run the Random Forest (RF) models, at the same georeferenced points of sampling, we generated circular buffers with radius = 0.7 m. Then, using the Zonal Statistics as Table tool (ESRI, 2017), the average values of each VI were extracted in the areas corresponding to these buffers. Therefore, the regression model dataset contained the amount of N (kg kg^{-1}) plus the VI values, while the classification model dataset, was composed by the levels of N (deficient, appropriate, and excessive) plus the VI data.

Variance and mean analysis, and Random Forest models

Initially, we generated a correlation matrix to evaluate the relationship between the VI and the measured N (kg kg^{-1}) through the Pearson Correlation Coefficient (PCC). We also performed the analysis of variance (ANOVA) using R, and then, using the package DTK (Lau, 2015), in R environment, we executed the Tukey-Kramer test ($p < 0.05$), since the samples were unbalanced, relating the VI data to the three levels of N.

The Random Forest is a machine learning algorithm, used to make predictions through regression or classification, and that can handle a high number of variables without overfitting. Its implementation is relatively simple and fast, being currently, one of the most accurate available techniques for this task (Biau, 2012).

The RF works with the analysis of decision trees, creating a “forest” of trees that are combined by the bagging method, with the premise that this random combination will generate better models with more accurate predictions (Osco et al., 2019; Biau, 2012).

In this work, this algorithm was used to generate a model of regression, to estimate the N content, and another of classification, to estimate the N levels, using the values of VI as explanatory variables. Both models were executed using the randomForest package in the software R (Liaw; Wiener, 2002; R Core Team, 2019).

The datasets of the models were separated into training subsets, with 80% of the data, and test subsets, with 20% of the data. Also, it was determined the creation of 500 trees, with 2 variables at each node. This last parameter was defined using the tuneRF function, of the same package.

To evaluate the regression model, we use the coefficient of determination (R^2) and the root mean of square error (MSE) between the predicted and the measured data in the test subset. To evaluate the classification model, we used the Kappa coefficient (k), and the overall accuracy, which is a measure that discounts the Out of bag (OOB) error rate, and estimates the classification error rate in the test subset as the trees

are added to the forest. However, for both models, as the trees are added, the reduction of the errors is limited, avoiding overfitting.

Moreover, the Random Forest also assesses the relevance of variables by generating a score of importance, given by the increase of the Mean Squared Error (%IncMSE) for the regression model, and by the Mean Decrease Accuracy for the classification model. Both scores reflect the OOB error rate assessment, evaluating the impact of the exchange of each variable for the increase of the errors in the predictions, that is, determining how important is each variable to the precision of the forecasts.

3 RESULTS AND DISCUSSION

Leaf analysis and Digital Elevation Model

The basic statistics of the N and SPAD values measured are shown in Table 3. The results indicate that, in general, the N content is appropriated, according to the level proposed by Guimarães et al. (1999).

However, considering these levels, only 53% of the leaf samples were appropriate (0.0288 to $0.0322 \text{ kg kg}^{-1}$), while 23.5% of the analyzes showed deficiency ($N < 0.0288 \text{ kg kg}^{-1}$), and another 23.5%, exhibited an excess ($N > 0.0322 \text{ kg kg}^{-1}$). Using the DEM (Figure 2), it was possible to understand how the nitrogen varies in the studied area.

We performed a Pearson correlation test with the N amounts (kg kg^{-1}) and the respective values of altitude, with all sampled points, to evaluated if there is a variation of the N levels according to the altitude. The result presented $\rho = -0.48$ (significant at $p < 0.01$), and therefore, although it is not a perfect negative correlation, it shows a significant tendency of accumulation of the nutrient in the lower parts of the area, which can be explained by the transportation of nutrients mainly due to the action of rainfall and the surface runoff. The variation of the N content in relation to the terrain elevation is shown in the Table 4.

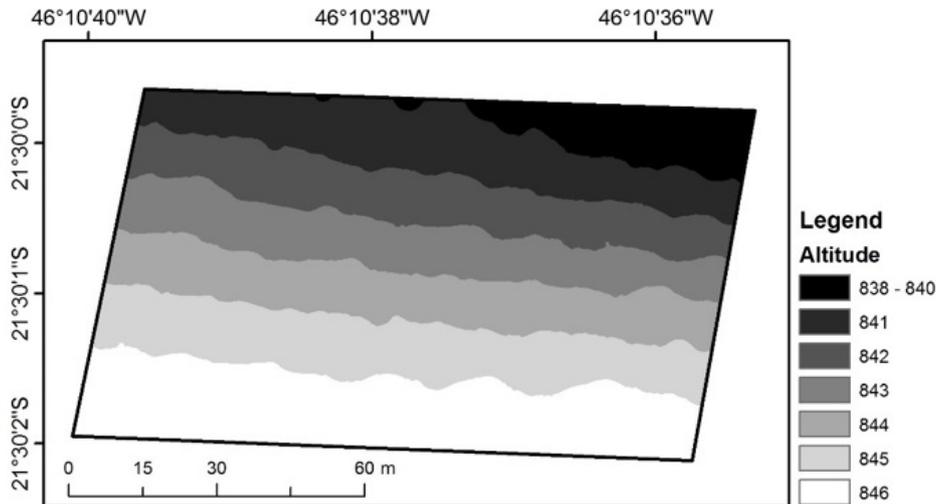
The correlation analysis between N content and SPAD values, otherwise, indicated low relationship, $\rho = 0.16$ between these parameters, which was not expected, since the SPAD readings usually are well related with plant nutrition (Reis et al., 2009; Pôrto et al., 2011).

Image classification

The supervised classification of the image that was made with the CIVE to separate the coffee lines from the background (soil, shade, and invasive plants), presented an almost perfect accuracy, since the Kappa coefficient of the classification was 0.92, with significance statistic of $p < 0.01$. The final classified image is shown in Figure 3.

Table 3: Basic statistics of the leaf analysis of the nitrogen content and the SPAD values.

Parameter	Minimum	Maximum	Average	Standard Deviation
N	0.024 kg kg ⁻¹	0.0357 kg kg ⁻¹	0.0302 kg kg ⁻¹	0.0024 kg kg ⁻¹
SPAD values	50.3	70.3	59.95	3.53

**Figure 2:** Digital Elevation Model (DEM) of the study area.**Table 4:** Variation of the N content as the terrain elevation changes.

Elevation (meters)	Average N content (kg kg ⁻¹) ± Standard Deviation
838 – 840 m	0.0329 ± 0.00016
840.1 - 841 m	0.0313 ± 0.00022
841.1 - 842 m	0.0297 ± 0.00019
842.2 - 843 m	0.0319 ± 0.00018
843.1 - 844 m	0.0285 ± 0.00021
844.1 - 845 m	0.0288 ± 0.00018
845.1 - 846 m	0.0294 ± 0.00018

Vegetation Indices

The correlation analysis showed that the VI are not related to the N content, since the best correlation obtained was with the Green Leaf Index (GLI), $\rho = 0.21$ ($p < 0.05$). However, most VI are highly related to each other. The Color Index of Vegetation (CIVE) is a VI that highlight the red band (absorption band on plants), that is why it presented inversely proportional correlation with the N and the other VI, especially with the GLI (Table 5).

Considering the levels of N (deficient, appropriate and excessive), the analysis of variance (ANOVA) showed that, between the VI, only the CIVE presented significant variation, and the SPAD values also did not varied significantly within the N levels (Table 6).

The analysis of means, made with the Tukey-Kramer test for unbalanced samples revealed that this difference exhibited in the CIVE occurred between the excessive - appropriate classes (Table 7).

Random Forest models

Both the regression model, built with the N data in kg kg⁻¹, and the classification model, made with the N data classified in levels, did not obtain significant results in the predictions. The results of the models are shown in Table 8.

In the classification model, the appropriate class reached an overall accuracy of 73%. However, the unbalancing of the samples can help to explain this outcome. There were 44 samples of this class, while there were only 19 for the excessive, and 21 for the deficient. The confusion matrix of the training sample shows that the model estimated 67% of the samples as “appropriated”, which although representative of this level on the field, did not mean sensibility to distinguish among the different N levels. This outcome can also be explained by the inexistence of significant differences within the vegetation indices, as shown previously with the Tukey-Kramer test.

Regarding the importance of each variable, the CIVE was the most important for the regression model, while for the classification model, the VARI was the most relevant, followed by the CIVE (Table 9).

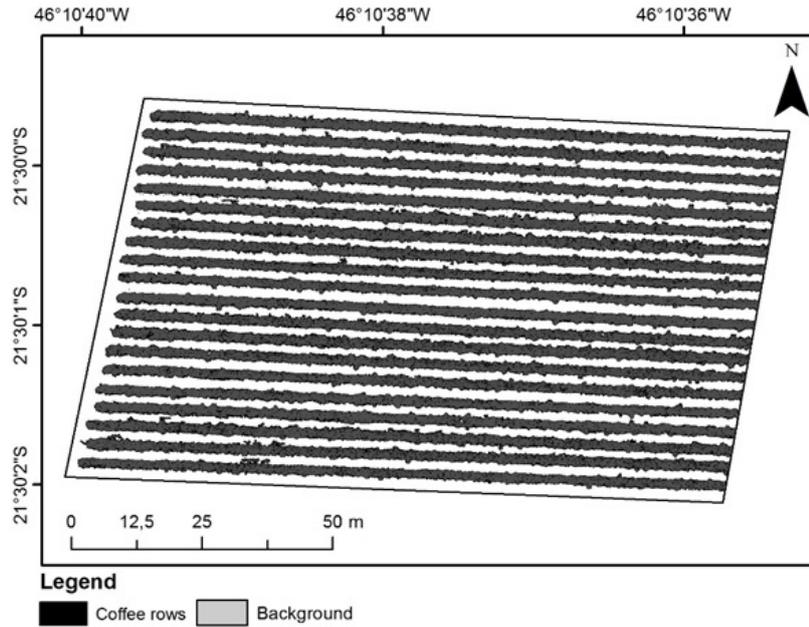


Figure 3: Classified image indicating the coffee rows.

Table 5: Correlation matrix between the measured N x VI, and between VI x VI.

	N (kg kg ⁻¹)	GLI	NGRDI	VARI	ExG	CIVE
N (kg kg ⁻¹)	1					
GLI	0.21*	1				
NGDRI	0.18	0.79*	1			
VARI	0.15	0.69*	0.98*	1		
ExG	0.17	0.64*	0.97*	0.98*	1	
CIVE	-0.20	-0.98*	-0.69*	-0.57*	-0.51*	1

Where: N (nitrogen); GLI (Green Leaf Index); NGRDI (Normalized Green Red Difference Index); VARI (Visible Atmospherically Resistant Index); ExG (Excess Green); CIVE (Color Index of Vegetation).

**statistical significance at $p < 0.05$.

Table 6: Analysis of variance between the vegetation indices and the SPAD values in relation to the nitrogen levels.

Variable	F value	Probability (>F)
CIVE	3.404	0.03
ExG	0.169	0.84
GLI	1.968	0.15
VARI	0.113	0.89
NGRDI	0.304	0.74
SPAD	0.202	0.82

Where: GLI (Green Leaf Index); NGRDI (Normalized Green Red Difference Index); VARI (Visible Atmospherically Resistant Index); ExG (Excess Green); CIVE (Color Index of Vegetation).

Table 7: Average values of the vegetation indices and the SPAD values according to the levels of nitrogen analyzed with the Tukey-Kamer test ($p < 0.05$).

Variable	Average	Variable	Average
CIVE		GLI	
Appropriate	18.770 a	Excessive	0.186 a
Deficient	18.768 ab	Deficient	0.176 a
Excessive	18.757 b	Appropriate	0.175 a
EXG		NGRDI	
Excessive	0.726 a	Excessive	0.0965 a
Appropriate	0.723 a	Appropriate	0.0927 a
Deficient	0.721 a	Deficient	0.0910 a
VARI		SPAD	
Excessive	0.1422 a	Excessive	68.18 a
Appropriate	0.1409 a	Appropriate	59.69 a
Deficient	0.1371 a	Deficient	59.50 a

Where: GLI (Green Leaf Index); NGRDI (Normalized Green Red Difference Index); VARI (Visible Atmospherically Resistant Index); ExG (Excess Green); CIVE (Color Index of Vegetation). Means followed by different letters differ ($p < 0.05$) by Tukey-Kramer test.

Table 8: Performance of the Random Forest models.

Model	Overall accuracy	Kappa Coefficient	R ²	MSE (kg kg ⁻¹)
Regression	-	-	-0.06	0.00636
Classification	0.33	-0.02	-	-

Where: R² (coefficient of determination); MSR (mean squared error).

Table 9: Contribution of each vegetation index for the Random Forest models of regression and classification, where higher values represents higher relevance.

Vegetation Index	Regression Model	Classification Model
	Increase of mean squared error	Mean Decrease Accuracy
Color Index of Vegetation (CIVE)	8.73 %	2.48
Green Leaf Index (GLI)	6.54 %	-4.44
Normalized Green-Red Difference Index (NGRDI)	1.50 %	-1.34
Excess Green (ExG)	1.33 %	-2.29
Visible Atmospherically Resistant Index (VARI)	1.21 %	3.19

The main objective of finding a model able to explain and predict N-variability is to assist farmers in the identification of zones with deficiency or excess, with fastness e precision, thus preventing economic losses and environmental negative impacts caused by unnecessary fertilization. However, our approach of using the Random Forest models could not monitor and predicted the N content or level, since the vegetation indices used in this study did not present significant variation that it would allow them to distinguish between the different nutrition status over the field.

Zhang et al. (2020), Hunt Junior et al. (2018) and Schirrmann et al. (2016) also found weak results monitoring the N variability with RGB imagery in maize, potatoes, and winter wheat, respectively. Garza et al. (2020), did not find significant correlation with the visible VI Triangular Greenness Index (TGI) and the amount of N in citrus trees. However, Zhang et al. (2020) reported that the Excess Green (ExG) may have the potential for this task, and Schirrmann et al. (2016) found positive relationships with other biophysical parameters of the crop, like fresh and dry biomass, and Leaf Area Index.

Regarding the leaf area, Hunt Junior et al. (2018) suggest that the vegetation cover index (CVR) can be an alternative in the monitoring of the N content with RGB images, due to the high spatial resolution, and due to the relationship between N content and plant development. Garza et al. (2020) also affirm that the leaf area can be affected by the nutritional status of the plants, so monitoring it can lead to the detection of stresses like diseases and lack of nutrients, and Näsı et al. (2018) found high correlations between the amounts of N and biomass in barley crops, indicating the potential of the approach.

Beniaich et al. (2019) estimated the cover vegetation index of millet and jack beans using RGB images and found good results with the CIVE and the ExG. In our study, the CIVE presented an excellent performance in the leaf cover classification. Moreover, Cunha, Sirqueira Neto and Hurtado (2019) obtained good results estimating the volume of coffee crops using true color images from UAV. Therefore, the monitor of the growth, within different rates of N fertilizing with this VI could be an alternative to be explore.

In previous studies, researchers have been establishing plots with different treatments to create a contrast over the study area. Our study, on the other hand, was executed in an area without the control of the fertilizing, which was done homogeneously by the farm administration. Moreover, in the monitoring of N-variability of different crops, the correlations are usually higher in the vegetative phase (Schirrmann et al., 2016; Vega et al., 2015; Zhang et al., 2020), which was not the case of this survey, done after the harvest.

The authors that used RGB-imagery report variable results, and the methodology seems to play an important role in these outcomes. Escalante et al. (2019) achieved very good results using different deep convolutional neural networks to monitor three different rates of N with barley, and their model were especially accurate to identify zones of deficiency, with a performance higher than 95%. Näsı et al. (2018), also monitoring the N variability with barley, managed to improve their results combining RGB spectral data with 3D features.

Therefore, exploring the wide range of machine learning algorithms is a way of deeply examine the potential of the RGB-imagery to this task, as well as incorporate 3D models in the analysis, which can be done with the same photogrammetry software, although it requires a much powerful computer.

Caturegli et al. (2019) bring an interesting approach that involves the conversion of the RGB images into values of Hue, Saturation, and Brightness (HSB) to generate the Dark Green Colour Index (DGCI), which led them to promising results, with high correlations of this index with the N content and even with the NIR-based VI Normalized Difference Vegetation Index (NDVI).

Therefore, although the use of RGB-imagery can lead to very poor results, it all depends on continuously exploring different approaches of data acquiring, processing, and analysis. Base on the results reported by some authors, the visible range should not be discarded from the task of monitoring the nutritional status of crops.

Regardless of the fact that we did not establish different fertilizing treatments, the leaf analysis revealed the occurrence of sites with inadequate N content, with a noticeable spatial distribution, like higher amounts in lower areas. This reasserts

the importance of the N-monitoring so producers can create management zones, perform the applications at variable rates, using a faster, relatively less expensive, and more appropriate strategy to reach the goals.

Despite the results, RGB imagery is a less expensive alternative that should continue to be tested, once it could come to represent a changing point to the planning of the fertilization. If this approach reaches success, it can also beneficiate small farmers, helping to reduce the costs of the production, to improve productivity, and to minimize the environmental impacts related to the fertilizers.

4 CONCLUSIONS

The Random Forest models were not able to explain and predict the nitrogen variability, since the regression model achieved an R^2 of -0.06, while the classification model presented an overall accuracy of 33%, with a Kappa coefficient of -0.02.

The analysis of variance pointed to a significant variance only in the Color Index of Vegetation, and the Tukey-Kramer test showed that this difference occurred between the appropriate and the excessive levels. Therefore, this index could be implemented in further studies.

The analysis of correlation shown that the Excess Green had almost perfect positive correlation with the Normalized Green-Red Difference Index, $\rho = 0.97$ ($p < 0.05$), and with the Visible Atmospherically Resistant Index, where $\rho = 0.98$ ($p < 0.05$). Moreover, the Color Index of Vegetation, that highlights the red band, showed an almost perfect negative correlation with the Green Leaf Index, that in turn, highlights the green band, where $\rho = 0.98$ ($p < 0.05$).

The Green Leaf Index was the only vegetation index that presented a significant correlation with the N content, although low, where $\rho = 0.21$ ($p < 0.05$).

The method of image classification done using the Color Index of Vegetation and the Maximum Likelihood Classifier algorithm presented very good performance, achieving a Kappa coefficient of 0.92, an almost perfect accordance between real and classified image.

There is a tendency of occurrence of higher nitrogen amount in lower areas of the field, since the Pearson Correlation Coefficient was significant as $\rho = -0.48$ ($p < 0.01$).

5 ACKNOWLEDGMENTS

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Finance Code 001. The authors thank the Minas Gerais State Research Support Foundation (FAPEMIG) and the National Council for Scientific and Technological Development (CNPq) for the grants

awarded. The authors also thank the Legado Cafés that conceded the study area, and the Experimental Nutrition Laboratory of the Nutrition School of the Federal University of Alfenas for conceding space, orientation and equipment for leaf analysis.

6 REFERENCES

- ALVARES, C. A. et al. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22(6):711-728, 2013.
- ARROYO, J.; GUIJARRO, M.; PAJARES, G. An instance-based learning approach for thresholding in crop images under different outdoor conditions. *Computers and Electronics in Agriculture*, 127:669-679, 2016.
- BALLESTER, C. et al. Assessment of In-Season Cotton Nitrogen Status and Lint Yield Prediction from Unmanned Aerial System Imagery. *Remote Sensing*, 9(11):1149, 2017.
- BENIAICHI, A. et al. Determination of vegetation cover index under different soil management systems of cover plants by using an unmanned aerial vehicle with an onboard digital photographic camera. *Semina: Ciências Agrárias*, 40(1):49-66, 2019.
- BIAU, G. Analysis of a Random Forests model. *Journal of Machine Learning Research*, 13:1063-1095, 2012.
- CATUREGLI, L. et al. Normalized Difference Vegetation Index versus Dark Green Colour Index to estimate nitrogen status on bermudagrass hybrid and tall fescue. *International Journal of Remote Sensing*, p. 1-16, 2019.
- CECHIM JÚNIOR, C.; JOHANN, J. A.; ANTUNES, J. F. G. Mapping of sugarcane crop area in the Paraná State using Landsat/TM/OLI and IRS/LISS-3 images. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 21(6):427-432, 2017.
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37-46, 1960.
- CUNHA, J. P. A. R.; SIRQUEIRA NETO, M. A.; HURTADO, S. M. C. Estimating vegetation volume of coffee crops using images from unmanned aerial vehicles. *Engenharia Agrícola*, 39:41-47, 2019.
- ESCALANTE, H. J. et al. Barley yield and fertilization analysis from UAV imagery: A deep learning approach. *International Journal of Remote Sensing*, 40(7):2493-2516, 2019.

- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE - INC. – ESRI. **ARCGIS Professional GIS for the desktop version 10.5**. Redlands, California, 2017. Available in: <<https://desktop.arcgis.com/en/quick-start-guides/10.5/arcgis-desktop-quick-start-guide.htm/>>. Access in: January 10, 2019.
- FERRAZ, G. A. S. et al. Agricultura de precisão no estudo de atributos químicos do solo e da produtividade de lavoura cafeeira. **Coffee Science**, 7(1):56-67, 2012.
- GAMER, M. et al. **irr**: various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. 2019 Available in: <<https://cran.r-project.org/web/packages/irr/irr.pdf>>. Access in: April 30, 2020.
- GARZA, B. N. et al. Quantifying Citrus Tree Health Using True Color UAV Images. **Remote Sensing**, 12(1):170, 2020.
- HUNT JUNIOR, E. R. et al. Monitoring nitrogen status of potatoes using small unmanned aerial vehicles. **Precision Agriculture**, 19:314-333, 2018.
- GITELSON, A. A. et al. Novel algorithms for remote estimation of vegetation fraction. **Remote Sensing of Environment**, 80:76-87, 2002.
- GUIMARÃES, P. T. G. et al. Cafeeiro. In: RIBEIRO, A. C.; GUIMARÃES, P. T. G.; ALVARES, V. H. **Recomendações para o uso de corretivos e fertilizantes em Minas Gerais: 5ª aproximação**. Viçosa: UFV, 1999. p.289-302.
- INSTITUTO NACIONAL DE METEOROLOGIA – INMET. **Estações Convencionais**. 2020. Available in: <<http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesConvencionais>>. Access in: January 27, 2020.
- KATAOKA, T. et al. Crop growth estimation system using machine vision. **Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics**, 2(1):1079-1083, 2003.
- LAU, M. K. **DTK**: Dunnett-Tukey-Kramer Pairwise Multiple Comparison Test Adjusted for Unequal Variances and Unequal Sample Sizes. R package version 3.5. 2015. Available in: <<https://CRAN.R-project.org/package=DTK>>. Access in: April 30, 2020.
- LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R News**, 2(3):18-22, 2002.
- LOUHAICHI, M. et al. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. **Geocarto International**, 16(1):65-70, 2001.
- MAES, W. H.; STEPPE, K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. **Trends in Plant Science**, 24(2):152-164, 2019.
- NÄSI, R. et al. Estimating Biomass and Nitrogen Amount of Barley and Grass Using UAV and Aircraft Based Spectral and Photogrammetric 3D Features. **Remote Sensing**, 10(7):1082, 2018.
- OSCO, L. P. et al. Predicting Canopy Nitrogen Content in Citrus-Trees Using Random Forest Algorithm Associated to Spectral Vegetation Indices from UAV-Imagery. **Remote Sensing**, 11(24):2925, 2019.
- PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento Remoto da Vegetação**. 2. ed. São Paulo: Oficina de Textos, 2012. 176p.
- PÔRTO, M. L. et al. Índice SPAD para o diagnóstico do estado de nitrogênio na cultura da abobrinha. **Horticultura Brasileira**, 29(3):311-315, 2011.
- QGIS Development Team, 2020. **QGIS Geographic Information System**. Open Source Geospatial Foundation Project. Available in: http://www.qgis.org/wiki/Developers_Manual>. Access in March 30, 2020.
- R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, 2019.
- REIS, A. R. et al. Photosynthesis, Chlorophylls, and SPAD Readings in Coffee Leaves in Relation to Nitrogen Supply. **Communications in Soil Science and Plant Analysis**, 40:1512-1528, 2009.
- SCHIRRMANN, M. et al. Monitoring Agronomic Parameters of Winter Wheat Crops with Low-Cost UAV Imagery. **Remote Sensing**, 8(9):706, 2016.
- TAIZ, L. et al. **Fisiologia e Desenvolvimento Vegetal**. 6. ed. Porto Alegre: Artmed, 2017. 858p.
- TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, 8:127-150, 1979.
- VEGA, F. A. et al. Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. **Biosystems Engineering**, 132:19-27, 2015.
- WOEBBECKE, D. M. et al. Color indices for weed identification under various soil, residue, and lighting conditions. **Transactions of the ASAE**, 38(1):259-269, 1995.
- ZHANG, M. et al. Estimation of maize yield and effects of variable-rate nitrogen application using UAV-based RGB imagery. **Biosystems Engineering**, 189:24-35, 2020.