

ORIGINAL ARTICLE

Estimation of sanity of a stand of *Pinus taeda* L. after the attack of *Sapajus nigritus* Kerr (1972) using vegetation index

Estimativa da sanidade de um povoamento de Pinus taeda L. após o ataque de Sapajus nigritus Kerr (1972) utilizando índices de vegetação

Carla Talita Pertille¹ , Marcos Benedito Schimalski¹ , Vilmar Picinatto Filho² ,
Verardo Liesenberg¹ , Edilson Batista de Oliveira³ , Fabiani das Dores Abati Miranda⁴ 

¹Universidade do Estado de Santa Catarina – UDESC. Lages, SC, Brasil

²Universidade Federal do Paraná – UFPR, Curitiba, PR, Brasil

³Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA, Colombo, PR, Brasil

⁴Universidade Tecnológica Federal do Paraná – UTFPR, Dois Vizinhos, PR, Brasil

How to cite: Pertille, C. T., Schimalski, M. B., Picinatto Filho, V., Liesenberg, V., de Oliveira, E. B., Miranda, F. D. A. (2020). Estimation of sanity of a stand of *Pinus taeda* L. after the attack of *Sapajus nigritus* Kerr (1972) using vegetation index. *Scientia Forestalis*, 48(126), e3323. <https://doi.org/10.18671/scifor.v48n126.03>

Abstract

This study aimed to identify regression and classification models using the most robust VI derived from Sentinel-2/MSI digital images for the detection of areas attacked by *Sapajus nigritus* in a *Pinus taeda* L. stand located in Bocaina do Sul, Santa Catarina State. In order to do this, three orbital images of the Sentinel-2/MSI constellation were used in pre-attack, attack and post-attack dates. A total of 46 field plots were allocated, in which tree individuals were classified according to the intensity of the year in: I: no damage, II: moderate, III: severe and VI: dead. Using Principal Component Analysis, vegetation index were selected as regressor variables for the development of regression models by Stepwise, Support Vector Machine (SVM) and Random Forest (RF) to estimate the percentage of attack classes per plot. The best model selection was based on regression model fit statistics. The most robust indexes were MCARI, NDI45 and NDVIRR, which estimated for the classes a percentage of attack of *S. nigritus* per plot with adjusted R² of 0.80, 0.46, 0.88 and 0.97; Syx of 13.49, 61.42, 15.11 and 8.69 (% 0.045 ha⁻¹) and RMSE of 1.95, 4.49, 2.11 and 0.32 (% 0.045 ha⁻¹). It was possible to estimate/identify areas attacked by this primate in the evaluated stand, highlighting the Stepwise model, followed by SVM and RF. However, it is recommended to test digital images of high/ultra-high spatial resolution to corroborate with such methodology.

Keywords: Vegetation indexes; Remote sensing; Red-edge.

Resumo

Essa pesquisa objetivou ajustar modelos de regressão e classificação utilizando os índices de vegetação (IV) mais robustos derivados a partir de imagens digitais de Sentinel-2/MSI para a detecção de áreas atacadas por *Sapajus nigritus*. Os dados de campo foram obtidos em um povoamento de *Pinus taeda* L. localizado em Bocaina do Sul, estado de Santa Catarina. Foram alocadas 46 parcelas, nas quais os indivíduos arbóreos foram classificados de acordo com a intensidade do ano em: I: sem dano, II: moderado, III: severo e IV: morto. Foram utilizadas três imagens orbitais da constelação Sentinel-2/MSI em datas coincidentes ao pré ataque, ataque e pós ataque. Por meio de Análise de Componentes Principais, foram selecionados os índices de vegetação como variáveis regressoras para o desenvolvimento de modelos de regressão por *Stepwise*, *Support Vector Machine* (SVM) e *Random Forest* (RF) para a estimativa do percentual das classes de ataque por parcela. A seleção de melhor modelo foi baseada em estatísticas de ajuste de modelos de

Financial support: Support Fund for the Maintenance and Development of Higher Education (FUMDES/SC), Santa Catarina Research Foundation (FAPESC), Brazilian National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES).

Conflict of interest: Nothing to declare.

Corresponding author: carlatpertille@gmail.com

Received: 25 March 2019.

Accepted: 01 July 2019.

Editor: Paulo Henrique Müller Silva



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

regressão. Os índices mais robustos foram o MCARI, NDI45 e NDVIRR, os quais estimaram para as classes um percentual de ataque de *S. nigritus* por parcela com R^2 ajustado de 0,80, 0,46, 0,88 e 0,97; Syx de 13,49, 61,42, 15,11 e 8,69 (% 0,045 ha⁻¹) e RMSE de 1,95, 4,49, 2,11 e 0,32 (% 0,045 ha⁻¹). Foi possível estimar/identificar áreas atacadas por esse primata no povoamento avaliado, destacando a modelagem Stepwise, seguida por SVM e RF. Entretanto, recomenda-se testar imagens digitais de alta/ultra alta resolução espacial para corroborar com tal metodologia.

Palavras-chave: Índices de vegetação; Sensoriamento remoto; Red-edge.

INTRODUCTION

The species *Sapajus nigritus* Kerr (1972), commonly known as capuchin monkey, is endemic to the Atlantic Forest, with geographic distribution from the state of Minas Gerais to the north of Rio Grande do Sul (Lynch Alfaro, 2012). Furthermore, according to Hendges et al. (2017), the capuchin monkey has significant adaptability to changes occurring in a given location and occupation of new areas, such as degraded and fragmented environments.

The exploitation of the Atlantic Forest directly interfered in the living conditions of this species, causing the monkey to seek alternative foods, such as the inclusion of the sap of *Pinus* sp. in their diet (Rocha, 2000). However, this consumption causes damages to forest stands of the genus *Pinus*. Because when removing the bark from the tree, the monkey-nail causes two types of damage: windowing (only a part of the bark of the trunk is removed) and girdling (occurs removal of the bark over the entire circumference of the trunk). These damages will compromise the growth of the tree and the attack of forest pests such as wood wasp (e.g., *Sirex noctilio* F.) or cause its death (Koehler & Firkowski, 1996; Mikich & Liebsch, 2009; Mikich & Liebsch, 2014).

Therefore, an alternative to monitor the impacts of this primate on forest plantations refers to the use of Remote Sensing (SR), a technology that enables the evaluation and monitoring of forest health in a remote way (Pause et al., 2016). For this, indicators of forest health can be used, such as visible evaluation of infestation levels, leaf characteristics such as defoliation and discoloration, trunk damage, number of dead or fallen trees, crown closure or leaf area index (LAI) (Lausch et al., 2017). The influence of stress factors should be associated with *in situ* observations (plot level) and the use of SR images (Trumbore et al., 2015) at different temporal and spatial scales (Lausch et al., 2017), as well as the use of Vegetation Indexes (VI).

There are several VI in the literature, however, the Normalized Difference Vegetation Index (NDVI) is one of the most used to diagnose plant health (Rullan-Silva et al., 2013). In other studies, such as that of Townsend et al. (2012), short-wave infrared (SWIR) based indexes were superior to NDVI. Indexes based on red-edge have demonstrated satisfactory performance in the study by Adelabu et al. (2014). Therefore, Rullan-Silva et al. (2013) recommend testing different VI for each specific case.

A potential source of data for studies in this area refers to the Sentinel-2 satellite mission, produced by the European Space Agency (ESA). This constellation of satellites presents several advantages for use in studies aimed at the diagnosis of forest health, such as: free data with spatial resolution comparable to commercial satellites, spectral bands of red border specific for vegetation studies and temporal resolution of 5 days (European Space Agency, 2010). As a consequence, the use of images of this sensor for studies related to the health of forest stands, monitoring the stage and extent of attack and/or infestation levels is important.

In this scenario, this research aimed to adjust regression models using the most robust VI derived from Sentinel-2 multispectral images for the detection of areas attacked by *Sapajus nigritus* in a *Pinus taeda* L. stand located in Bocaina do Sul, state of Santa Catarina.

MATERIAL AND METHODS

Study area

The research was conducted at the Cerro Rico farm (Figure 1) located in Bocaina do Sul, in the state of Santa Catarina, with average coordinates UTM (Universal Transverse Mercator)

E: 602272 meters and N: 6936448 meters and with an area of 1535.57 hectares. The municipality presents a Cfb (humid subtropical mesothermic climate), according to the classification of Köppen, average temperature of 16.5° C, relative humidity of 79.3% and precipitation well distributed throughout the year, with annual average of 1500 mm (Alvares et al., 2013).

The area has plantations of *Pinus taeda* L., *Eucalyptus* spp. and native forest remnants. The effective plantation area of *Pinus taeda* L. was planted in 2002 and 2003, totaling 236.21 ha, with spacing of 3 x 2.5 meters. There was no silvicultural intervention (slabs) and the harvest system used was the full tree.

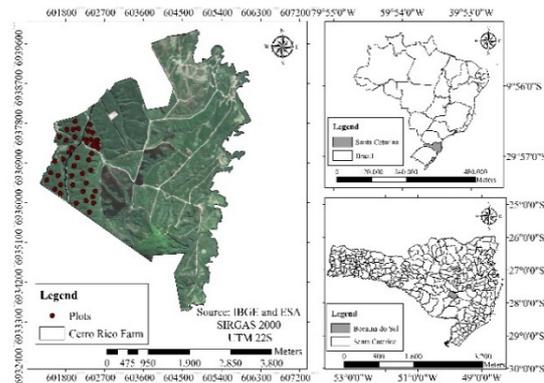


Figure 1. Location of the stand of *Pinus taeda* L. in Bocaina do Sul, Santa Catarina State, Brazil.

Data collect

The sampling procedure used was simple random sampling, representing the same probability of selection of plots. Thus, 46 circular plots of variable radius of 5 to 10 meters were installed in the area to meet sample adequacy based on Neto et al. (2002).

The size and shape of the plots were defined according to the *S. nigritus* attack pattern, which affects isolated and random trees in the stands. Therefore, circular plots of variable radius of 5 to 10 meters were used to allow the inclusion of attacked and non-attacked trees. Of the 46 plots, three had a radius of 5 meters and 43 had a radius of 10 meters. It is important to note that there was no overlapping of plots.

In all plots, the geographic position was obtained using a GPS receiver (Global Positioning System) of the Garmin Etrex Legend ® model, with data georeferenced in the WGS - 84 coordinate system (World Geodetic System 1984). The method of geodetic positioning used was the absolute one.

In order to evaluate the damage caused by *S. nigritus* in *Pinus taeda* L. trees, four qualitative damage classes were proposed: I: without damage; II: moderate; III: severe and VI: 100%. The evaluation was performed in a visual way, obeying the following criteria: trees without damage (they did not present any type of damage); moderate (presented damage), severe (presented the two types of damage: windowing and girdling or only girdling) and dead (upper part of the dry cup). Thus, the attack percentage of each class was obtained by the ratio of the number of trees of the respective class of damage by the total number of trees in the plot.

Remote sensing data

Three images of the satellite constellation Sentinel-2/MSI (Multispectral Instrument) were used in data corresponding to three times: pre-attack (image 05/05/2017); attack (11/11/2017) and post attack (04/18/2018) with the purpose of performing the temporal analysis of the attacks of this primate in the area. The criterion for establishing the set intervals of the images used corresponded to the symptoms of *S. nigritus* attack on *Pinus taeda* L. All images were purchased at the Copernicus Open Access Hub, with the bands and characteristics described in European Space Agency (2010), in orbit 221 and point 79.

The images were obtained as data at level 1C, i.e., data with reflectance at the top of the atmosphere (TOA). In the ENVI version 5.3 (Environment for Visualizing Images) (EXELIS, 2018), the images were processed to level 2A, with atmospheric reflectance: Bottom of the Atmosphere (BOA). The atmospheric correction was performed using the FLAASH (Fast Line-of-Sight Atmospheric Analysis of Hypercubes) algorithm. Then, all image bands were resampled to the 10-meter spatial resolution using the Sentinel Application Platform (SNAP) (European Space Agency, 2018).

With the images properly processed, the vegetation index described in Table 1 were derived. These VI were chosen according to the spectral ranges that they address and the reports found in the literature on forest health.

Table 1. Vegetation indexes calculated for the study area.

VI	Formula	Reference
Anthocyanin Reflectance Index	$(1/B3) - (1/B5)$	Gitelson et al. (2001)
Anthocyanin Reflectance Index	$(B8/B3) - (B8/B5)$	Gitelson et al. (2001)
Chlorophyll Red-Edge	$(B5/B8) - 1$	Gitelson et al. (2003a)
Red and red-edge modified chlorophyll index	$(B8) / (1 * B4 + (1 - a) * B5) - 1$	Xie et al. (2018)
Carotenoid Reflectance Index 1	$(1/B2) - (1/B3)$	Gitelson et al. (2002)
Carotenoid Reflectance Index 1	$(1/B2) - (1/B5)$	Gitelson et al. (2002)
Enhanced Vegetation Index	$2.5 * (B8 - B4) / (B8 + 6 * B4 + 7.5 * B2 + 1)$	Jiang et al. (2008)
Enhanced Vegetation Index 2	$2.5 * (B8 - B4) / (B8 + 2.4 * B4 + 1)$	Jiang et al. (2008)
Green Normalized Difference Index	$(B8 - B3) / (B8 + B3)$	Gitelson et al. (1996)
Inverted Red-Edge Chlorophyll Index	$(B7 - B4) * B6/B5$	Clevers et al. (2000)
Modified Chlorophyll Absorption in Reflectance Index	$1 - 0.2 * (B5 - B3) / (B5 - B4)$	Daughtry (2000)
Modified Soil Adjusted Vegetation Index	$\frac{(B8 - B4)}{(B8 + B4 + L)}(1 + L)$	Qi et al. (1994)
Second Modified Soil Adjusted Vegetation Index	$(B8 + 1) - 0.5 * \text{sqrt}((2 * B8 - 1)^2 + 8 * B4)$	Qi et al. (1994)
MERIS Terrestrial Chlorophyll Index	$(B6 - B5) / (B5 - B4)$	Dash e Curran (2007)
Normalized Difference Infrared Index	$(B8 - B11) / (B8 + B11)$	Hardisky et al. (1983)
Normalized Difference Vegetation Index	$(B8 - B4) / (B8 + B4)$	Rouse et al. (1974)
Red Edge NDVI	$(B8 - B6) / (B8 + B6)$	Fernández-Manso et al. (2016)
Red and red-edge normalized difference vegetation index	$(B8 - (1 * B4 + (1 - a) * B5)) / (B8 + (1 * B4 + (1 - a) * B5))$	Xie et al. (2018)
Plant Senescence Reflectance Index Near Infrared	$(B4 - B2) / (B6)$	Merzlyak et al. (1999)
Pigment Specific Simple Ratio	$(B8/B4)$	Blackburn (1998)
Soil Adjusted Vegetation Index	$1.5 * (B8 - B4) / (B8 + B4 + 0.5)$	Huete (1988)
Sentinel-2 Red-Edge Position	$705 + 35 * (0.5 * (B7 + B4) - B5) / (B6 - B5)$	Guyot & Baret (1988)

Note: B2: blue (490 nm); B3: green (560 nm); B4: red (665 nm); B5: red-edge 1 (705 nm); B6: red-edge 2 (740 nm); B7: red-edge 3 (783 nm); B8: NIR (842 nm); B8A: red-edge 4 (865 nm); B11: SWIR 1 (1610 nm); B12: SWIR 2 (2190 nm); L and a: adjustment factors (considered as 0.5).

The delimitation of the plots in the images was made from the coordinates of the central point of the plot. This point served to delineate a circumference with a radius of 12 m, since the creation of plots with larger radius could decrease the GPS positioning error. This is due to the error in the GNSS (Global Navigation Satellite System) positioning method. The plots were delimited with this radius to include a significant number of trees attacked by *S. nigritus*. For the existence of this error, an area of influence of 12 meters of radius was constructed to generate the area of the parcels in the images. This procedure was performed in a Geographic Information System (GIS) environment (Environmental Systems Research Institute, 2018) using the buffer tool. Then, the mean value of each VI per plot was obtained for each of the three images used.

Selection of variables

The regressor (VI) variables were selected by the Principal Components Analysis (PCA). This analysis is used when the objective is to reduce the number of variables without causing loss of information and quality (Andrade et al., 2007) and also to reduce the variance of the data. For each class of damage, VIs with an importance percentage equal to or greater than 85% were considered as predictor variables in the construction of regression models to estimate the percentage of attack per plot (% 0.045 ha⁻¹).

Methods for adjusting regression models

The Stepwise method is a regression technique that determines which variables have the greatest influence on the dependent variables (output set). This technique can reduce the number of variables that will constitute the regression model and is used in exploratory studies and to describe the relationship between the variables (Abbad & Torres, 2002). In this study, the criterion used was AIC, since it is the statistical criterion that presents a measure of discrepancy between the lines of the true model and the approximate model.

Support Vector Machine (SVM) is a powerful and highly flexible modeling class. Initially, it was developed for classification models, but was transformed into robust regression models to reduce the effect of discrepant data on regression equations. The characteristics of the SVM for regression were: radial and polynomial kernel, degree ranging from 1 to 3 and cost of 0.5, 1, 2 to 4, according to Hawrylo et al. (2018).

The RF methodology uses a versatile set of decision trees, and performs well for estimating linear and nonlinear models by balancing bias and variance (Breiman, 2001). For the use of this technique, two parameters must be defined: the number of input variables randomly chosen in each division (mtry) and the number of trees (ntree). From tests that determined the low variability in the data, the mtry was 2 and the ntree was 3 and 6. These parameters were used following the guidance of Hawrylo et al. (2018). The adjusted models are described in Table 2.

Table 2. Models adjusted by the techniques Stepwise, Support Vector Machine and Random Forest in the estimation of attack of *S. nigritus* per plot (% 0.045 ha⁻¹) in a *Pinus taeda* L stand.

Model	Method	Class	Formula
1	Stepwise	Without damage	$A = \beta_0 + \beta_1 * IV + \beta_2 * \ln IV + \beta_3 * 1/IV^2 + \beta_4 * 1/IV^2^2 + \beta_5 * 1/IV^4$
2	SVM	Without damage	$A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * IV^2^2 + \beta_5 * IV^2^3 + \beta_6 * IV^2^4 + \beta_7 * IV^5$
3	RF	Without damage	$A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * IV^2^2 + \beta_4 * 1/IV + \beta_5 * EXP IV^2^4$ $A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * IV^2 + \beta_4 * IV^3 + \beta_5 * IV^4 +$
1	Stepwise	Moderate	$\beta_6 * IV^5 + \beta_7 * IV^2^2 + \beta_8 * IV^2^3 + \beta_9 * IV^2^4 + \beta_{10} * IV^2^5 + \beta_{11} * \ln IV^2 + \beta_{12} * EXP IV + \beta_{13} * EXP IV^2 + \beta_{17} * 1/IV^3$

Table 2. Continued...

Model	Method	Class	Formula
2	SVM	Moderate	$A = \beta_0 + \beta_1 * IV + \beta_2 * 1 / IV^2 + \beta_3 * 1 / IV^2 + \beta_4 * IV^{2^2} + \beta_5 * IV^4 + \beta_6 * \frac{1}{IV^5} + \beta_7 * \ln IV^2 + \beta_8 * EXP IV^2^3 +$
3	RF	Moderate	$A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * IV^2^4 + \beta_4 * IV^2^5 + \beta_5 * \ln IV^2^2 + \beta_6 * EXP IV + \beta_7 * EXP IV^2 + \beta_8 * 1 / IV^3$
1	Stepwise	Severe	$A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * IV^2 + \beta_4 * \ln IV^2 + \beta_5 * IV^3 + \beta_6 * EXP IV^4 + \beta_7 * 1 / IV^5 + \beta_8 * 1 / IV^2^2 + \beta_9 * EXP IV^2^3 + \beta_{10} * \ln IV$
2	SVM	Severe	$A = \beta_0 + \beta_1 * IV + \beta_2 * 1 / IV^2 + \beta_3 * 1 / IV^2 + \beta_4 * 1 / IV^3 + \beta_5 * 1 / IV^4 + \beta_6 * 1 / IV^5 + \beta_7 * 1 / IV^2^2 + \beta_8 * 1 / IV^2^3 + \beta_9 * \ln IV^2 + \beta_{10} * EXP IV^2$
3	RF	Severe	$A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * 1 / IV * 1 / IV^2 + \beta_4 * IV^3 + \beta_5 * IV^4 + \beta_6 * 1 / IV^5 + \beta_7 * EXP 1 / IV^2^2 + \beta_8 * EXP IV^2^3 + \beta_9 * \ln 1 / IV^2 + \beta_{10} * EXP 1 / IV^2$
1	Stepwise	Dead	$A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * IV^3 + \beta_4 * IV^4 + \beta_5 * IV^5 + \beta_6 * IV^2^2 + \beta_7 * IV^2^3 + \beta_8 * IV^2^4 + \beta_9 * IV^2^5 + \beta_{10} * \ln IV + \beta_{11} * \ln IV^2 + \beta_{12} * EXP IV^2 + \beta_{13} * EXP IV^3$
2	SVM	Dead	$A = \beta_0 + \beta_1 * IV + \beta_2 * IV^2 + \beta_3 * EXP IV^5 + \beta_4 * \sqrt{IV^5} + \beta_5 * \frac{1}{IV^2} + \beta_6 * \frac{1}{IV^2^3} + \beta_7 * \frac{1}{IV^3} * 1 / IV^2 + \beta_8 * \frac{1}{IV^4} * 1 / IV$
3	RF	Dead	$A = \beta_0 + \beta_1 * IV + \beta_2 * EXP IV^3 + \beta_3 * EXP IV^2^5 + \beta_4 * 1 / IV^2 + \beta_5 * 1 / IV^3^5 + \beta_6 * IV * IV^2 + \beta_7 * IV * IV^3 + \beta_8 * IV^2 * IV^3$

Note: A: attack per plot (% 0.045 ha⁻¹); βi: parameters to be estimated; ln: natural logarithm based on the constant e (2,71828182845904); EXP: natural exponential function.

The criteria for choosing the best model were the following: higher adjusted coefficient of determination (R² adjusted) (Equation 1), lower standard error values of the estimate (Syx%) (Equation 2 and 3), as suggested by Nicoletti et al. (2016), Akaike Information Criterion (AIC) (Equation 4), Bayesian Information Criterion (BIC) (Equation 5) and Root Mean Squared Error (RMSE) (Equation 6). The statistical factor pointed out by Schneider et al. (2009), test F with the level of significance of 5% of probability, was also considered.

$$R^2_{aj} = 1 - \left\{ (1 - R^2) * \left(\frac{n-1}{n-p} \right) \right\} \tag{1}$$

$$Syx = \sqrt{\frac{\sum (y - \hat{y}_i)^2}{n-p}} \tag{2}$$

$$Syx\% = \frac{Syx}{\hat{Y}} * 100 \tag{3}$$

$$AIC = n * \ln(SQ_{res}) - n * \ln(n) + 2p \tag{4}$$

$$BIC = -2 \log(L_p) + [(p+1)+1] \log(n) \tag{5}$$

$$RMSE = \sqrt{\frac{\sum (y - \hat{y}_i)^2}{n}} \tag{6}$$

Note: R^2_{aj} : adjusted coefficient of determination; n : number of observations; p : number of parameters of the equation; S_{yx} : standard error of estimate (% 0.045 ha⁻¹); y : attack observed ((% 0.045 ha⁻¹); y_i : attack estimated(% 0.045 ha⁻¹); S_{yx} (%):standard error of the percentage estimate (%); \hat{y} : mean of the observed values ((% 0.045 ha⁻¹); p : number of model parameters; SQ_{res} : Sum of Squares of the residues obtained by ANOVA; L_p : maximum likelihood function of the model; RMSE: Root Mean Square Error (% 0.045 ha⁻¹).

Statistical analyses

For the validation of the best model, 20% of the data were separated, in which the chi-square test (χ^2) was applied at a significance level of 95% probability. The Chi-Square test was proposed by Pearson (1990) and is expressed by the Greek letter χ^2 . In this method, the hypothesis of existence or not of differences between observed and predicted values is verified. For each class of damage assessed, the statistical analyses involved the normality test by the Kolmogorov-Smirnov test at 5% probability. The design was a completely randomized design with four treatments (percentage of attack observed in the field, percentage of attack estimated by the model of the Stepwise technique, percentage of attack estimated by the model of the SVM technique and percentage of attack estimated by the RF model). Analyses were performed in software R version 3.4.1. (R Core Team, 2018).

RESULTS

From the field inventory data it was found that the percentage attack values per plot differed for each class analyzed. In all plots we can see a higher number of trees not attacked than attacked, as can be seen in the statistics of the class without damage (Figure 2). This pattern indicates that, in the evaluated stand, a greater number of non-attacked trees were found; however, the attacked trees represented large losses of wood volume due to the damages caused by severe and moderate damages.

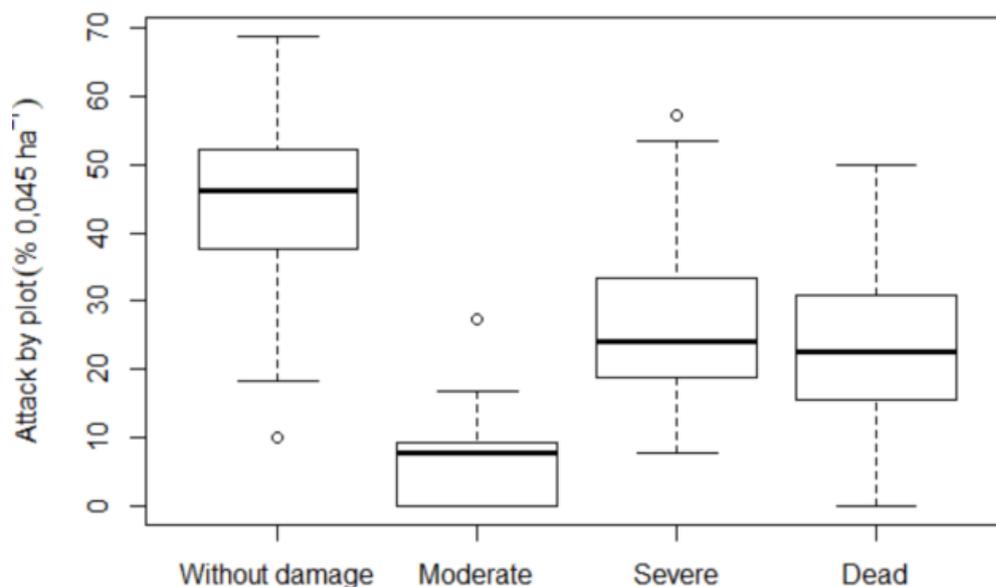


Figure 2. Descriptive statistics of the variable percentage of attack by plot in a stand of *Pinus taeda* L. attacked by *S. nigritus*.

The selection of the most representative indexes by class by the PCA (Figure 3) to be used in the construction of the regression models revealed that for each analyzed image, of the 24 index derived, only three index stood out (Table 3).

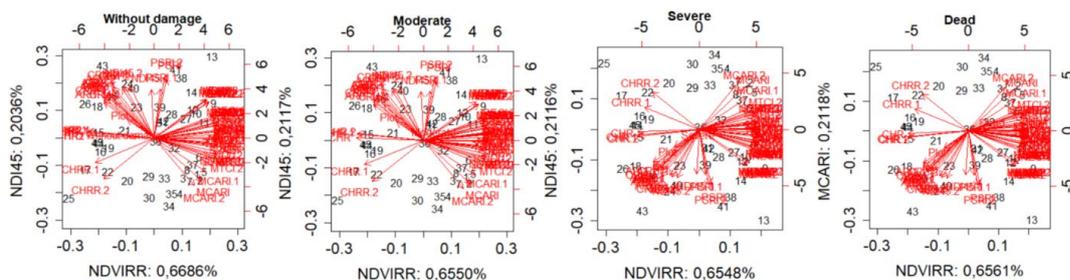


Figure 3. Selection of regression variables using PCA for a settlement of *Pinus taeda* L. attacked by *S. nigritus*.

Table 3. Selection of regression variables using PCA for a *Pinus taeda* L. stand attacked by *S. nigritus* in Bocaina do Sul - SC.

Class	VI	ProportionofVariance
Without damage	NDVIRR e NDI45	0.8722
Moderate	NDVIRR e NDI45	0.8667
Severe	NDVIRR e NDI45	0.8664
Dead	NDVIRR e MCARI	0.8667

Note: VI: Vegetation Index; MCARI: Modified Chlorophyll Absorption in Reflectance Index; NDI45: Normalized Difference Index 45 e NDVIRR: Red and Red-edge Normalized Difference Vegetation Index.

All adjusted regression models estimated the percentage of attack class per plot with moderate adjustment metrics (Table 4). The adjusted R^2 variation was 0.4274 to 0.9766, while error percentages ranged from 0.32 to 13.08 (% 0.045 ha⁻¹).

Table 4. Statistical analysis of the regression models tested for the Sentinel-2 sensor to estimate the percentage of capuchin monkey attack class per plot (% 0.045 ha⁻¹) for a *Pinus taeda* L. stand in Bocaina do Sul.

Model	Method	VI	Without damage						
			R ² aj	Syx	Syx (%)	F	AIC	BIC	RMSE
1	Stepwise	NDVIRR e NDI45	0.8042	5.85	13.49	5.63	216.31	268.57	1.95
2	SVM	NDVIRR e NDI45	0.9	13.19	30.38	-	-	-	13.08
3	RF	NDVIRR e NDI45	0.8697	12.89	29.92	-	-	-	8.87
Model	Method	VI	Moderate						
			R ² aj	Syx	Syx (%)	F	AIC	BIC	RMSE
1	Stepwise	NDVIRR e NDI45	0.4603	0.95	62.56	1.96	197.53	249.78	1.49
2	SVM	NDVIRR e NDI45	0.4274	4.43	61.42	-	-	-	4.89
3	RF	NDVIRR e NDI45	0.4508	5.91	81.87	-	-	-	6.51
Model	Method	VI	Severe						
			R ² aj	Syx	Syx (%)	F	AIC	BIC	RMSE
1	Stepwise	NDVIRR e NDI45	0.8815	4.15	15.11	8.89	170.63	226.05	2.11
2	SVM	NDVIRR e NDI45	0.9047	12.07	43.95	-	-	-	10.72
3	RF	NDVIRR e NDI45	0.8548	11.75	42.78	-	-	-	6.49
Model	Method	VI	Dead						
			R ² aj	Syx	Syx (%)	F	AIC	BIC	RMSE
1	Stepwise	NDVIRR e MCARI	0.9766	1.96	8.69	43.95	93.86	150.87	0.32
2	SVM	NDVIRR e MCARI	0.69	11.38	50.27	-	-	-	11.67
3	RF	NDVIRR e MCARI	0.9629	13.21	58.37	-	-	-	12.98

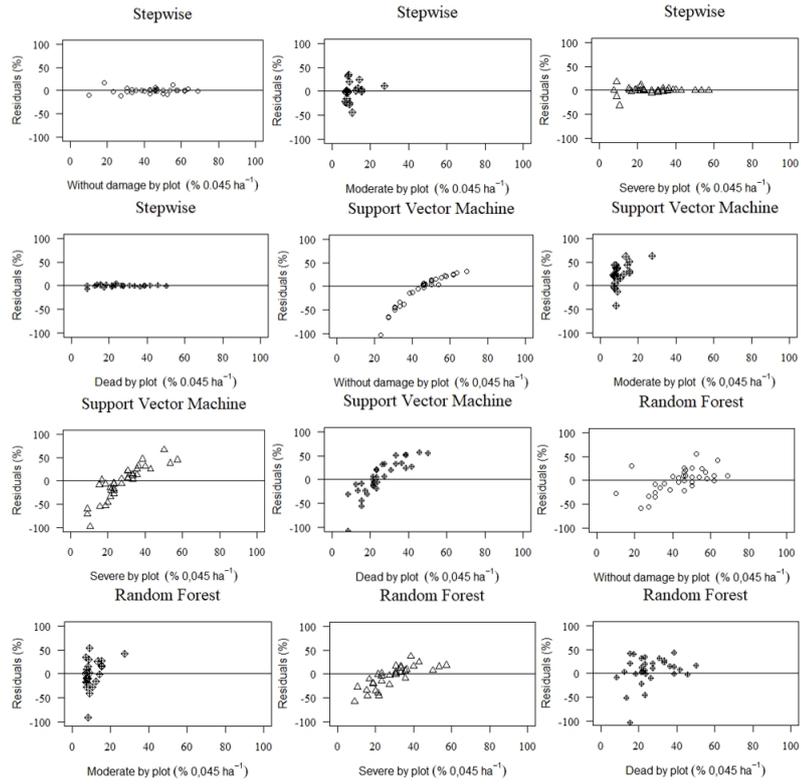


Figure 4. Graphical dispersion of the residuals of the models adjusted for the classes of damage by the tested technique in estimating the percentage of attack of *S. nigritus* per plot (% 0.045 ha⁻¹) in a *Pinus taeda* L. stand in Bocaina do Sul.

The regression methods tested indicated that the Stepwise model was superior to the others, a fact attributed to the criterion of addition or removal of variables to the model. On the other hand, the models developed by machine learning revealed instability in the construction of the models, by the criterion of separation of training samples in the case of SVM and RF decision forests. Due to the superior performance of the Stepwise model, it was possible to notice in Figure 4 that there was a more concentrated distribution of residues around the regression line. For the moderate and severe class, the residuals were high for all the techniques tested, due to the great variability of the data and the smaller number of trees belonging to this class (Figure 5).

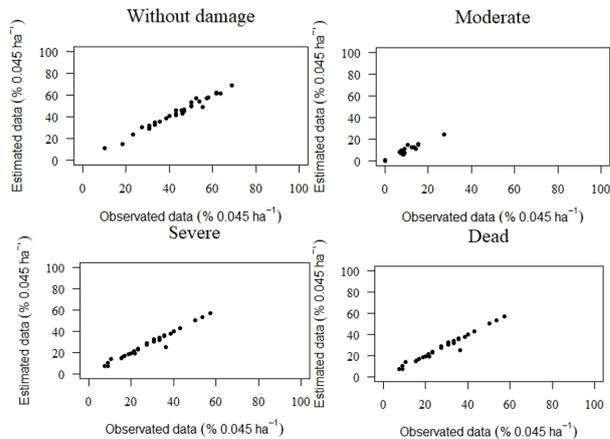


Figure 5. Graphical dispersion of the estimated and observed values for the damage classes by modeling in the estimation of the percentage of attack of *S. nigritus* per plot (% 0.045 ha⁻¹) in a *Pinus taeda* L. stand in Bocaina do Sul.

The validation of the best model by damage class indicated that there are no differences between the attack values estimated by said model and the attack data for each class observed in the field. Statistical analyses corroborate these results, revealing that there was no significant difference between the percentage of attack observed by the inventory and the data remotely located for each attack class.

DISCUSSION

The VI selected by the PCA (MCARI, NDI45 and NDVIRR) express the spectral behavior of the vegetation in the spectral ranges of each index. Stressed plants such as given by seizures, diseases, temperatures and nutrition induce the loss of chlorophyll, drastically interfering with the absorption of photosynthetically active radiation. The result of this process is the increase of the reflectance in the visible band, especially the bands in the region of red (665 nm) and blue (560 nm) (Carter & Knapp, 2001). This fact is related to the spectral sensitivity of each index as well as to the variance of the derived values. Thus, these index were distinguished by the smaller variance of the data and greater representativeness with the attack classes.

The red-edge bands have greater sensitivity to variations in chlorophyll content than visible bands to detect changes in chlorophyll values induced by stress (Eitel et al., 2010) or *S. nigritus* attack. This also causes a decrease in reflectance in the infrared range, which interacts with the healthy leaf structure and is spread through the air cell interfaces of the spongy mesophyll (Jensen, 2000).

Factors such as age, density, type of underbrush vegetation, soil type and forest management practices may also contribute to strengthen the relation of vegetation index to *S. nigritus* attack, as well as to consolidate the knowledge of the pattern of attack of this primate in forest stands. Monitoring of forest health can also be associated with climate, degree of disturbance, forest-type phenology and stress-causing agent (Stone & Coops, 2004).

In addition, another factor that interfered with the ratio of index to damage per plot was the amount of vegetation coverage within the Sentinel-2 pixels (spatial resolution of 10 m) and the degree of homogeneity within the pixel found in the forest stand under study, as pointed out by Zarco-Tejada et al. (2018). The identification of the plots with the positioning of GNSS signal receivers influenced the relationship between the VI and the categories of damage. The errors associated with the use of GPS in forests involve delays in the ionosphere and troposphere (decrease of intensity of the GPS signal when crossing the atmosphere), multi-lamination (GPS reflection by objects or structures before reaching the receiver), receiver clock (ie, the receiver clock accuracy is lower than the GPS satellites' atomic clocks) and orbit errors (incorrect information about satellite position) (Monico et al., 2014).

The regression models for each attack class developed by traditional modeling (Stepwise) and by the learning machines (SVM and RF) demonstrated the moderate performance of all techniques for all images. The addition of regressor variables by the Stepwise technique (Forward and Backward) resulted in linear models with highest value of adjusted R^2 and lower error percentages. The nonlinear models calibrated by SVM had adjustment metrics higher than the Stepwise technique with a smaller number of regressor variables. And the RF modeling resulted in non-linear models and less stable than the others.

However, the model developed by Stepwise had the lowest percentages of error, AIC, BIC, RMSE and better waste distribution. The moderate attack class obtained fit the statistics with the highest error percentages for all techniques tested. This fact can be explained by the recognition of this type of damage in the images and later in the indexes used in the models. The models adjusted to estimate the percentage of severe attack per plot revealed a good performance of all modeling techniques. However, the models developed by Stepwise were superior to the others. And the estimates of the last class of attack, referring to dead trees, were significant, especially the Stepwise, RF and SVM technique.

The models found by Stepwise estimate a variable of interest with the smallest model fit metrics. According to Piñeiro et al. (2008), the means are extremely significant, since the R^2 indicates the linear variations of the data volume and are explained by the variation of the

predicted values, being an indicator that represents the main forms of performance of the model. The RMSE, another calculated indicator, can be understood as a proxy of the mean scope (divided by $n-1$) of values not related to the regression relation. This estimative variable is the mean root error of the observed values in relation to the regression of the observed values versus the root mean square prediction and error of the values related to the regression rate of the predicted versus observed values.

It is also important to emphasize that it is recommended to test the reliability of the model with data from other sensors that have spectral bands of red, NIR, red-edge and SWIR, as recommended by Eitel et al. (2011). In addition, the importance of a satellite platform with red-edge bandwidth information for the development of the detection and monitoring system for defoliation of insects and other pests was highlighted by Adelabu et al. (2014), as well as for the classification of health level for forest management.

Some limitations of this research should were the size and geo-referencing of the field plots, the GPS receivers used, the spatial resolution of the images (10 meters), characteristics of the area such as soil type, trees and topographical conditions. Therefore, it is recommended that in future work that such methodology be tested in high to ultra high spatial resolution sensors in other forest areas affected by *S. nigritus* at different times and with different vegetation index.

CONCLUSION

For the data set used in this research, the most robust index for identifying areas attacked by *S. nigritus* per plot in a *Pinus taeda* L. stand were MCARI (Modified Chlorophyll Absorption in Reflectance Index), NDVIRR (Red and Red -Edge Modified Chlorophyll Index) and NDI45 (Normalized Difference Index 45). The regression models used to estimate the attack classes by plot ($\% 0.045 \text{ ha}^{-1}$) had significant adjustment metrics for the Stepwise technique.

It was possible to estimate/identify areas attacked by this primate in the evaluated stand; however, it is recommended to test images obtained by high and ultra high spatial resolution digital images to strengthen the methodology presented here.

REFERENCES

- Adelabu, S., Mutanga, O., & Adam, E. (2014). Evaluating the impact of red-edge band from Rapid eye image for classifying insect defoliation levels. *ISPRS Journal of Photogrammetry and Remote Sensing*, 95, 34-41. <http://dx.doi.org/10.1016/j.isprsjprs.2014.05.013>.
- Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22(6), 711-728.
- Abbad, G., & Torres, C. V. (2002). Regressão múltipla stepwise e hierárquica em Psicologia Organizacional: aplicações, problemas e soluções. *Estudos de Psicologia*, 7(spe), 19-29. <http://dx.doi.org/10.1590/S1413-294X2002000300004>.
- Andrade, E. M., Araújo, L. F. P., Rosa, M. F., Gomes, R. B., & Lobato, F. A. O. (2007). Fatores determinantes da qualidade das águas superficiais na bacia do Alto Acaraú, Ceará, Brasil. *Ciência Rural*, 37(6), 1791-1797. <http://dx.doi.org/10.1590/S0103-84782007000600045>.
- Blackburn, G. A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing*, 19(4), 657-675. <http://dx.doi.org/10.1080/014311698215919>.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <http://dx.doi.org/10.1023/A:1010933404324>.
- Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4), 677-684. PMID:11302854. <http://dx.doi.org/10.2307/2657068>.
- Clevers, J. G. P. W., De Jong, S. M., Epema, G. F., Addink, E. A., & Box, P. O. (2000, July 11-13). Meris and the Red-Edge Index. *2nd EARSeL workshop* (pp. 1-16). Enschede: EARSeL.
- Dash, J., & Curran, P. J. (2007). Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Advances in Space Research*, 39(1), 100-104. <http://dx.doi.org/10.1016/j.asr.2006.02.034>.

- Daughtry, C. (2000). Estimating corn leaf Chlorophyll Concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2), 229-239. [http://dx.doi.org/10.1016/S0034-4257\(00\)00113-9](http://dx.doi.org/10.1016/S0034-4257(00)00113-9).
- Eitel, J. U. H., Keefe, R. F., Long, D. S., Davis, A. S., & Vierling, L. A. (2010). Active ground optical remote sensing for improved monitoring of seedling stress in nurseries. *Sensors (Basel)*, 10(4), 2843-2850. PMID:22319275. <http://dx.doi.org/10.3390/s100402843>.
- Eitel, J. U. H., Vierling, L. A., Litvak, M. E., Long, D. S., Schulthess, U., Ager, A. A., Krofcheck, D. J., & Stoscheck, L. (2011). Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sensing of Environment*, 115, 3640-3646.
- European Space Agency – ESA (2010). *Introducing Sentinel-2*. Retrieved in 2017, June 06, from http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel2/Introducing_Sentinel-2.
- European Space Agency – ESA. (2018). *SNAP and Sentinel Toolboxes*. Retrieved in 2018, August 06, from <http://step.esa.int/main/download/>
- Environmental Systems Research Institute – ESRI. (2018). *ArcGIS Professional GIS for the desktop, version 10.4.1*. Retrieved in 2018, Maio 20, from <https://support.esri.com/en/Products/Desktop/arcgis-desktop/arcmap/10-4-1>
- EXELIS (2018). *Visual Information Solutions*. Retrieved in 2018, May 24, from <https://www.harrisgeospatial.com/SoftwareTechnology/ENVI.aspx>
- Fernández-Manso, A., Fernández-Manso, O., & Quintano, C. (2016). SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation*, 50, 170-175. <http://dx.doi.org/10.1016/j.jag.2016.03.005>.
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3), 289-298. [http://dx.doi.org/10.1016/S0034-4257\(96\)00072-7](http://dx.doi.org/10.1016/S0034-4257(96)00072-7).
- Gitelson, A. A., Merzlyak, M. N., Chivkunova, O. B., Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 74(1), 38-45. PMID:11460535. [http://dx.doi.org/10.1562/0031-8655\(2001\)074<0038:OPANEO>2.0.CO;2](http://dx.doi.org/10.1562/0031-8655(2001)074<0038:OPANEO>2.0.CO;2).
- Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, 75(3), 272-281. PMID:11950093. [http://dx.doi.org/10.1562/0031-8655\(2002\)075<0272:ACCIPL>2.0.CO;2](http://dx.doi.org/10.1562/0031-8655(2002)075<0272:ACCIPL>2.0.CO;2).
- Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., Leavitt, B. (2003a). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*, 30(5), 52-1-52-3.
- Guyot, G., & Baret, F. (1988, January 18-22). Utilisation de la Haute Resolution Spectrale pour Suivre l'état des Couverts Vegetaux. In *4th International Colloquium "Spectral Signatures of Objects in Remote Sensing"* (Publication SP, 287, pp. 279-286). Aussois, Paris: ESA.
- Hardisky, M., Klemas, V., & Smart, R. (1983). The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of spartina alterniflora canopies. *Photogrammetric Engineering and Remote Sensing*, 49, 77-83.
- Hawrylo, P., Bednarz, B., Wężyk, P., & Szostak, M. (2018). Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2. *European Journal of Remote Sensing*, 51(1), 194-205. <http://dx.doi.org/10.1080/22797254.2017.1417745>.
- Hendges, C. D., Melo, G. L., Gonçalves, A. S., Cerezer, F. O., & Cáceres, N. C. (2017). Landscape attributes as drivers of the geographical variation in density of *Sapajus nigritus* Kerr, 1792, a primate endemic on the Atlantic Forest. *Acta Oecologica*, 84, 57-63. <http://dx.doi.org/10.1016/j.actao.2017.08.007>.
- Huete, A. (1988). A soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, 25(3), 295-309. [http://dx.doi.org/10.1016/0034-4257\(88\)90106-X](http://dx.doi.org/10.1016/0034-4257(88)90106-X).
- Jensen, J. R. (2000). *Remote sensing of the environment - An earth resource perspective* (544 p.). Minnesota: Pearson Prentice Hall.
- Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10), 3833-3845. <http://dx.doi.org/10.1016/j.rse.2008.06.006>.
- Koehler, A., & Firkowski, C. (1996). Descascamento de pinuspormacaco-prego (*Cebusapella*). *Floresta*, 24(1-2), 61-75.

- Lausch, A., Erasmí, S., King, D. J., Magdon, P., & Heurich, M. (2017). Understanding forest health with remote sensing-Part II-A review of approaches and data models. *Remote Sensing*, 9(2), 129-133. <http://dx.doi.org/10.3390/rs9020129>.
- Lynch Alfaro, J. W. (2012). How different are robust and gracile Capuchin Monkeys? An argument for the use of *Sapajus* and *Cebus*. *American Journal of Primatology*, 74(4), 273-286. PMID:22328205. <http://dx.doi.org/10.1002/ajp.22007>.
- Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, 106(1), 135-141. <http://dx.doi.org/10.1034/j.1399-3054.1999.106119.x>.
- Mikich, S. B., & Liebsch, D. (2009). *O Macaco-prego e os Plantios de Pinus spp.* (Comunicado Técnico, 234, pp. 1-5). Colombo, PR: Embrapa Florestas.
- Mikich, S. B., & Liebsch, D. (2014). Damage to forest plantations by tufted capuchins (*Sapajus nigritus*): too many monkeys or not enough fruits? *Forest Ecology and Management*, 314, 9-16. <http://dx.doi.org/10.1016/j.foreco.2013.11.026>.
- Monico, J. F. G., Rofatto, V. F., & Mendonça, M. A. M. D. (2014). Introdução e aplicação da diluição da precisão das ambiguidades GNSS - ADOP. *Boletim de Ciências Geodésicas*, 20(2), 279-299.
- Neto, B. D. B., Scarminio, I. S., & Bruns, R. E. (2002). *Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria* (2. ed., 405 p.). São Paulo: Editora Unicamp.
- Nicoletti, M. F., Souza, K., Silvestre, R., França, M. C., & Rolim, F. A. (2016). Relação Hipsométrica para *Pinus taeda* L. em diferentes fases do ciclo de corte. *Floresta e Ambiente*, 23(1), 80-89.
- Pause, M., Schweitzer, C., Rosenthal, M., Keuck, V., Bumberger, J., Dietrich, P., Heurich, M., Jung, A., & Lausch, A. (2016). In situ/Remote sensing integration to assess forest health – A review. *Remote Sensing*, 8(6), 471.
- Pearson, K. (1990). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 5(50), 157-175.
- Piñeiro, G., Perelman, S., Guerschman, J. P., & Paruelo, J. M. (2008). How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216(3-4) 316-322.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119-126. [http://dx.doi.org/10.1016/0034-4257\(94\)90134-1](http://dx.doi.org/10.1016/0034-4257(94)90134-1).
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved in 2018, April 24, from <https://www.Rproject.org/>
- Rocha, V. J. (2000). Macaco-Prego, como controlar essa nova praga florestal? *Floresta*, 30(1/2), 95-99.
- Rouse, J. W., Haas, R. H., & Schell, J. A. (1974). Monitoring the Vernal Advancement of Retrogradation (Green Wave Effect) of Natural Vegetation, Texas: Remote Sensing Center, Texas A&M University College Station.
- Rullan-Silva, C. D., Olthoff, A. E., Delgado De La Mata, J. A., & Pajares-Alonso, J. A. (2013). Remote monitoring of forest insect defoliation. A review. *Forest Systems*, 22(3), 377. <http://dx.doi.org/10.5424/fs/2013223-04417>.
- Schneider, P. R., Schneider, P. S. P., & Souza, C. A. M. (2009). *Análise de regressão aplicada à Engenharia Florestal*. 2. ed. Santa Maria: Facos.
- Stone, C., & Coops, N. C. (2004). Assessment and monitoring of damage from insects in Australian eucalypt forests and commercial plantations. *Australian Journal of Entomology*, 43(3), 283-292. <http://dx.doi.org/10.1111/j.1326-6756.2004.00432.x>.
- Townsend, P. A., Singh, A., Foster, J. R., Rehberg, N. J., Kingdon, C. C., Eshleman, K. N., & Seagle, S. W. (2012). A general Landsat model to predict canopy defoliation in broadleaf deciduous forests. *Remote Sensing of Environment*, 119, 255-265. <http://dx.doi.org/10.1016/j.rse.2011.12.023>.
- Trumbore, S., Brando, P., & Hartmann, H. (2015). Forest health and global change. *Science*, 349(6250), 814-818. PMID:26293952. <http://dx.doi.org/10.1126/science.aac6759>.
- Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., Casa, R., Pignatti, S., Laneve, G., Pascucci, S., Dong, Y., & Ye, H. (2018). Vegetation indices combining the red and red-edge spectral information for leaf area index retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(11), 1482-1493. <http://dx.doi.org/10.1109/JSTARS.2018.2813281>.
- Zarco-Tejada, P. L., Hornero, A., Hernández-Clemente, R., & Beck, P. S. A. (2018). Understanding the temporal dimension of the red-edge spectral region for forest decline detection using high-

resolution hyperspectral and Sentinel-2A imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 137, 134-148. PMID:29551855. <http://dx.doi.org/10.1016/j.isprsjprs.2018.01.017>.

Author's contributions: CTP: conceptualization, writing, methodology, data processing, formal analysis. MBS: acquisition of resources, supervision; writing, formal analysis. VCP, VL, EBO and FADM: writing and formal analysis.