

Realistic and simplified models of plant and leaf area indices for a seasonally dry tropical forest

Rodrigo de Queiroga Miranda^{a,*}, Rodolfo Luiz Bezerra Nóbrega^{b,c},
Magna Soelma Beserra de Moura^{b,d}, Srinivasan Raghavan^e, Josiclêda Domiciano Galvêncio^a

^a Laboratório de Sensoriamento Remoto e Geoprocessamento, Universidade Federal de Pernambuco, Recife, Pernambuco, 50670901, Brazil

^b Department of Geography and Environmental Science, University of Reading, RG6 6AH Reading, UK

^c Department of Life Sciences, Imperial College London, S15 7PY Ascot, UK

^d Brazilian Agricultural Research Corporation - Embrapa Tropical Semiárid, Petrolina, Pernambuco, 56302970, Brazil

^e Spatial Sciences Laboratory, Texas A&M University, College Station, TX 77845, USA

ARTICLE INFO

Keywords:

Caatinga
Landsat
Phenology
Semi-arid
Woody Area Index
Leaf Area Index

ABSTRACT

Leaf Area Index (LAI) models that consider all phenological stages have not been developed for the Caatinga, the largest seasonally dry tropical forest in South America. LAI models that are currently used show moderate to high covariance when compared to *in situ* data, but they often lack accuracy in the whole spectra of possible values and do not consider the impact that the stems and branches have over LAI estimates, which is of great influence in the Caatinga. In this study, we develop and assess PAI (Plant Area Index) and LAI models by using ground-based measurements and satellite (Landsat) data. The objective of this study was to create and test new empirical models using a multi-year and multi-source of reflectance data. The study was based on measurements of photosynthetic photon flux density (PPFD) from above and below the canopy during the periods of 2011–2012 and 2016–2018. Through iterative processing, we obtained more than a million candidate models for estimating PAI and LAI. To clean up the small discrepancies in the extremes of each interpolated series, we smoothed out the dataset by fitting a logarithmic equation with the PAI data and the inverse contribution of WAI (Wood Area Index) to PAI, that is the portion of PAI that is actually LAI (LAI_C). LAI_C can be calculated as follows: $LAI_C = 1 - (WAI/PAI)$. We subtracted the WAI values from the PAI to develop our *in situ* LAI dataset that was used for further analysis. Our *in situ* dataset was also used as a reference to compare our models with four other models used for the Caatinga, as well as the MODIS-derived LAI products (MCD15A3H/A2H). Our main findings were as follows: (i) Six models use NDVI (Normalized Difference Vegetation Index), SAVI (Soil-Adjusted Vegetation Index) and EVI (Enhanced Vegetation Index) as input, and performed well, with r^2 ranging from 0.77 to 0.79 (PAI) and 0.76 to 0.81 (LAI), and RMSE with a minimum of $0.41 \text{ m}^2 \text{ m}^{-2}$ (PAI) and $0.40 \text{ m}^2 \text{ m}^{-2}$ (LAI). The SAVI models showed values 20% and 32% (PAI), and 21% and 15% (LAI) smaller than those found for the models that use EVI and NDVI, respectively; (ii) the other models (ten) use only two bands, and in contrast to the first six models, these new models may abstract other physical processes and components, such as leaves etiolation and increasing protochlorophyll. The developed models used the near-infrared band, and they varied only in relation to the inclusion of the red, green, and blue bands. (iii) All previously published models and MODIS-LAI underperformed against our calibrated models. Our study was able to provide several PAI and LAI models that realistically represent the phenology of the Caatinga.

1. Introduction

The Leaf Area Index (LAI) is a widely adopted parameter in environmental sciences studies. It represents the one-sided area of leaves that covers a specific surface area (Fotis et al., 2018; Knotte et al., 2009; Mu et al., 2007; Rodriguez et al., 2009) and is one of the main parameters of both global and regional biosphere models (Arnold et al., 1998;

Bieger et al., 2017). LAI is used to scale up from leaf to vegetation photosynthesis and transpiration, energy balance of terrestrial surfaces, and many climatological and hydrological attributes such as atmospheric aerosols, water infiltration, and biogeochemical processes (Bonan, 1995).

There are two main approaches used to estimate LAI: (i) direct methods, in which the total leaf canopy is obtained by the summation of direct measurement of all individual leaf areas – this is usually a

* Corresponding author.

E-mail address: rodrigo.qmiranda@gmail.com (R.d.Q. Miranda).

<https://doi.org/10.1016/j.jag.2019.101992>

Received 5 August 2019; Received in revised form 16 October 2019; Accepted 18 October 2019

0303-2434/ © 2019 Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

destructive method because it requires the removal of all leaves and, therefore, is not viable at large scales; and (ii) indirect methods, which may require active or passive sensors to measure parameters that are highly correlated with LAI, such as light extinction coefficient (Jonckheere et al., 2004); or use the litterfall trap method, which is suitable for estimating LAI of deciduous plants (Almeida et al., 2019). Active sensors do not depend on solar radiation as they emit their own electromagnetic signals and capture those reflected, whereas passive sensors depend on solar radiation and are based on estimating the extent to which a given amount of leaf area will reduce radiation transmitted through a stratified arrangement of leaf elements within a canopy (Zheng and Moskal, 2009). This estimation can be determined using a radiative transfer model such as the PROSPECT and SAIL models (Jacquemoud et al., 2009; Jacquemoud and Baret, 1990; Knyazikhin et al., 1998; Verhoef, 1985, 1984) or abstracted by coefficients of an empirical model (Bastiaanssen, 1998; Galvncio et al., 2013; Machado, 2014).

Radiative transfer models are highly accurate, but require specific inputs, such as pigment concentration, cell diameter, and water content (Jacquemoud et al., 2009; Jacquemoud and Baret, 1990). These parameters can only be obtained with extensive fieldwork, while empirical models are purely statistical fast retrieval algorithms (Zheng and Moskal, 2009). To estimate LAI, the empirical models are mainly composed of regressions that relate LAI values to simple spectral responses and greenness indices (Almeida et al., 2019; Galvncio et al., 2013; Machado, 2014), such as the Soil Adjusted Vegetation Index (SAVI) and Normalized Difference Vegetation Index (NDVI) (Bastiaanssen, 1998; Galvncio et al., 2013). For LAI estimations at a regional scale, empirical models are generally reliable (Knote et al., 2009). However, in Brazil, more specifically in the seasonally dry tropical forest (SDTF) in the semi-arid region, known as the Caatinga, models that are currently used have not been developed using both intra- and inter-annual field measurements.

The Caatinga is the largest continuous SDTF in the Americas, with an open and mostly semi-arid landscape, as seen in many inter-plateau depressions (Ab'Saber, 1974; Silva et al., 2017). The Caatinga covers an area of approximately 900,000 km² (Silva et al., 2017), and exhibits at least 13 different physiognomies ranging from woodlands to sparsely distributed thorny shrubs (Silva et al., 2017). Its climate is characterized by high temperatures and low rainfall rates with high intra- and inter-annual variability both in space and time. The rainfall is normally concentrated over 2–4 months of the year, with the possibility of over 25% of the annual precipitation occurring in a single rainfall event (Miranda et al., 2018). The main landscape units that can be found in the Caatinga are canyons, ravines, mountains, sandy, and clayey plateaus (Leal et al., 2007). Their complex soil mosaics are commonly formed by four dominant soil orders (Latosols, Lithosols, Argisols, and Luvisols) (Menezes et al., 2012). The Caatinga holds over 3,150 species of 930 genera and 152 families of flowering plants (Silva et al., 2017). These plants have unique adaptations to endure conditions of spatiotemporally irregular water availability and extended droughts: approximately 85% of the Caatinga species lose all their leaves during the dry season (Leal et al., 2003; Silva et al., 2017). Thus, methods that attempt to measure the LAI by directly relating it to the intercepted radiation do not reflect only the area of the leaves, but also the surface of the woody area, which is mainly comprised of stems and branches (Cunha et al., 2019). The influence that stems and branches have over the LAI estimates can be addressed by computing the LAI as the difference between Plant Area Index (PAI) and the Woody Area Index (WAI) (Kalacska et al., 2005).

Current model estimates of LAI in the Caatinga show moderate to high covariance when compared to *in situ* data ($r^2 = 0.60\text{--}0.93$), but they might lack accuracy in the entire spectra of possible values because they are often developed using observations that do not cover complete intra-annual LAI variations due to phenology (e.g., Almeida et al., 2019; Galvncio et al., 2013; Machado, 2014). In addition, most of these models applied to the Caatinga have not entirely considered the influence of the

continuous variation of WAI, which is highly significant in the Caatinga as over 85% of plant's above-ground biomass is composed of stems and branches (Silva and Sampaio, 2008). The oversight of this uniqueness of SDTFs, such as the Caatinga, is likely to cause methodological drawbacks in estimating LAI. By not considering every all phenological stage, the intra-annual LAI changes may be reduced, and PAI can be wrongly addressed as LAI. As a consequence, models may provide unrealistic values for some periods of the year, especially in the dry season.

In this study, we aimed to create and test new empirical models using a multi-year and multi-source set of reflectance data. We rely on the premise that by providing multiple reflectance data combinations as input and that by accounting for the WAI component of the PAI we will be able to provide models that are more accurate and better adjusted to the Caatinga. Our objectives were to evaluate the efficiency of new LAI models derived from Landsat reflectance using fitted regressions and field measurements from a typical Caatinga formation area in Brazil, and to test new empirical approaches using previously published models currently used for the Caatinga.

1.1. Study area

Data were collected in an area of shrubby hyperxerophytic Caatinga forest area (Fig. 1) (Kiill, 2017), located at the Embrapa Tropical Semi-arid Research Station in the state of Pernambuco, Brazil (9°2'33"S, 40°19'16"W; at 350 m a.s.l.). The vegetation in this area consists of shrubs, trees, herbaceous plants, and Cactaceae. The canopy average height is 4.5 m. The plant phenological stages in the Caatinga are usually four: foliar development, maturity, senescence, and dormancy (Rankine et al., 2017), and this cycle follows the rainfall patterns closely (Silva et al., 2017). Most species in the Caatinga are deciduous, and respond quickly to slight changes in soil water availability, breaking the dormancy of wood growth; and that allows the plants to sprout most of their leaves in only a few days in the beginning of the rainy season (Machado et al., 1997). The dominant plant species (approximately 90% of the total relative dominance) in our study area were *Commiphora leptophloeos*, *Schinopsis brasiliensis*, *Mimosa tenuiflora*, *Cenostigma microphyllum*, *Sapium glandulosum*, *Cnidoscylus quercifolius*, *Handroanthus spongiosus*, *Manihot pseudoglaziovii*, *Croton conduplicatus*, and *Jatropha mollissima* (Kiill, 2017). Although the Cactaceae (*Pilosocereus gounellei* and *Pilosocereus pachycladus*) have a fairly constant vegetative phenology throughout the year, these plants have a relative dominance of less than 5% and an insignificant production of leaves; therefore they were not considered in our LAI estimates. The climate is dry semi-arid (Alvares et al., 2013), with the rainy season between January and April and an average annual temperature of 26°C. Although the average historical annual rainfall is approximately 500 mm, the average rainfall was less than 300 mm during our study period, which is the most severe drought in this region's recorded history. These conditions were particularly interesting for our study, allowing a precise assessment of the WAI influence on the total PAI.

2. Methodology

2.1. Field measurements

LAI was derived from field measurements of photosynthetic photon flux density (PPFD) taken from above and below the canopy using two different non-destructive methods. The measurements were conducted throughout the year in order to cover all plant phenological stages and covered five years: 2011–2012 and 2016–2018. The first method measured PPFD using three quantum sensors (one LI-190SA sensor to measure the above-canopy PPFD, and two LI-191 sensors for the below-canopy data) installed in a 16-m meteorological tower in the study area. All sensors were connected to a data acquisition system (CR1000, Campbell Scientific Inc.), which was programmed to compute averages of 30-s measurements taken at 30-min intervals from January 2011 to December 2012. In order to maximize the quality of our measurements,

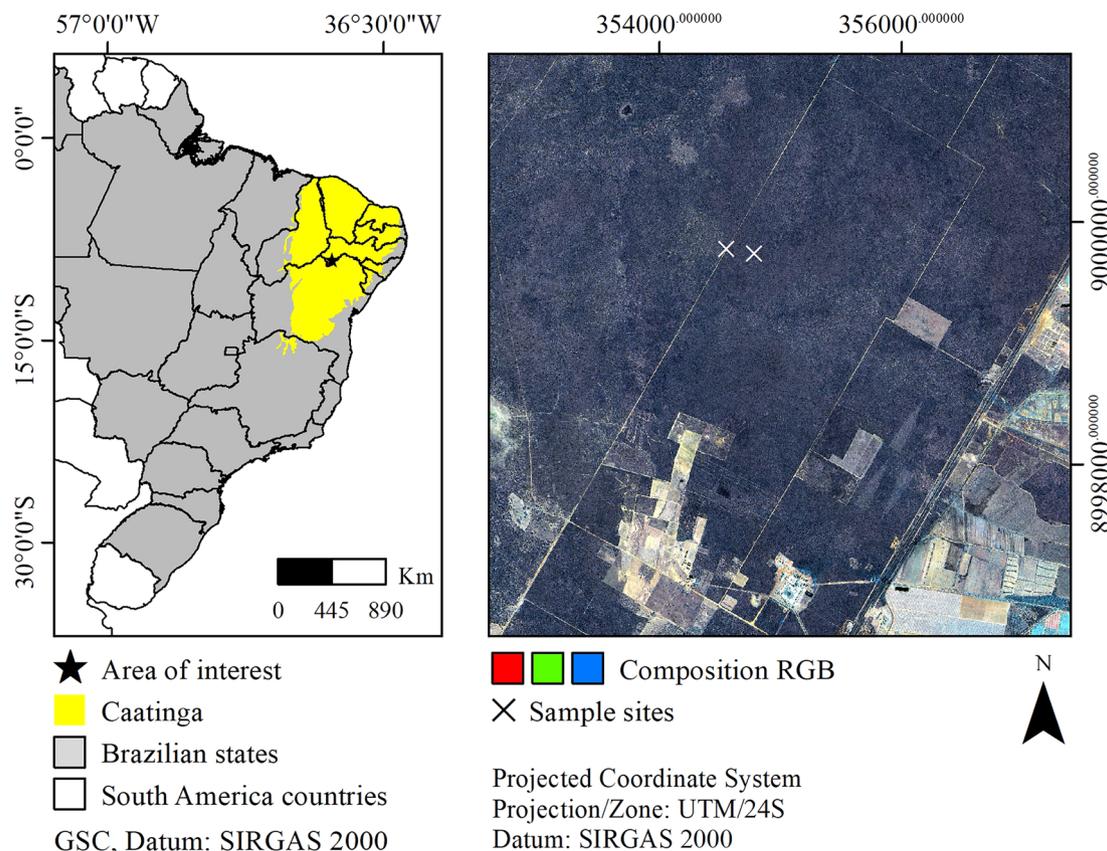


Fig. 1. Location of the seasonally dry tropical forest experimental area at the Embrapa Semiárid Research Station in the state of Pernambuco (Brazil).

we filtered all data, considering only the average of the measurements between 10 a.m. and 2 p.m. each day (GMT -3), when the zenith angle is close to zero. The second measurement approach was applied on a weekly basis (68.97% of the entire dataset) from January 2016 to November 2018, with the following exceptions: 19.54% (≥ 8 days of interval between measurements - DBM), 8.05% (≥ 14 DBM) and 3.44% (≥ 21 DBM). The dataset consisted of LAI estimates based on the transmission of light through the canopy at various angles by using an AccuPAR ceptometer (AccuPAR® LP-80, Decagon Devices). The AccuPAR has a linear ceptometer with 80 sensors, capable of measuring PPFd at the photosynthetically active radiation (PAR) range (400–700 nm wavelength) from 0 to $2500 \mu\text{mol m}^{-2}$. The above-canopy PPFd and solar zenith angle measurements were obtained in a nearby (about 10 m away) clear area, and the below-canopy PPFd was acquired by holding the AccuPAR beneath the canopy at approximately 0.4 m above ground. The dataset from this approach was linearly interpolated to produce the daily data required to match the satellite overpass times. We used the data collected to predict scattered and transmitted PPFd, as well as to predict light extinction, as proposed by Norman (1979).

2.2. Plant Area Index (PAI) partitioning

In our study, we defined PAI as the sum of WAI and LAI (Magalhães et al., 2018), and the WAI as the contribution of woody material such as stems, branches, and trunks to the light interception of PAI. In order to carry out this partition of our data, we first took the minimum LAI (LAI_{MIN}) value of each year as the WAI, which was verified by visual evaluation of hemispheric photos from a phenological monitoring database (Fig. 2); then we fixed this value from the day of the LAI_{MIN} to the first subsequent day with rainfall over 2.5 mm. Based on field observations, we assumed that low-precipitation ($\leq 2.5 \text{ mm d}^{-1}$) events did not cause any significant phenological change in the ecosystem. The

WAI was assumed to change between sequential dry seasons gradually; we gap-filled the WAI dataset with a linear interpolation between the fixed-value periods of each year. To avoid small discrepancies in the extremes of each interpolated series, we smoothed the dataset by fitting a logarithmic equation (Eq. 1) with the PAI data and the inverse contribution of WAI to PAI, which is the percentage of PAI that is actually LAI (here called LAI_C). LAI_C was calculated as follows: $\text{LAI}_C = 1 - (\text{WAI}/\text{PAI})$. The WAI values were subtracted from the PAI to develop our *in situ* LAI dataset, which is used for further analysis.

$$\text{WAI} = \{1 - [\ln(\text{PAI}) \times 0.5]\} \times \text{PAI} \quad (1)$$

2.3. Landsat data processing

We selected the Landsat Surface Reflectance Level-2 products for the entire study period (total of 110 candidate images). These products are designed to provide atmospherically and geometrically corrected reflectance data with 30-m resolution for every 16 days. These data are generated using the auxiliary climate data from MODIS (e.g., water vapor, ozone, geopotential height, and aerosol optical thickness) and two different algorithms: 1) the Second Simulation of a Satellite Signal in the Solar Spectrum (6S) algorithm to the data derived from Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images; and 2) a unique radiative transfer model to the Landsat 8 Operational Land Imager (OLI) data. The data were extracted from two sample sites (Fig. 1), and all clear pixels were filtered using the respective Quality Band (QA band) of each product ($\text{L5-7} = 66$, and $\text{L8} = 322$), resulting in a 70-record dataset. The dataset was then submitted to an iterative model-fitting approach to create new PAI and LAI models. The Landsat Collection 2 Level-2 products include reflectance values derived from three sensors (TM/Landsat 5, ETM+/Landsat 7, and OLI/Landsat 8) with 30-m spatial resolution. The different bands were matched to create an equivalent dataset of reflectance across all

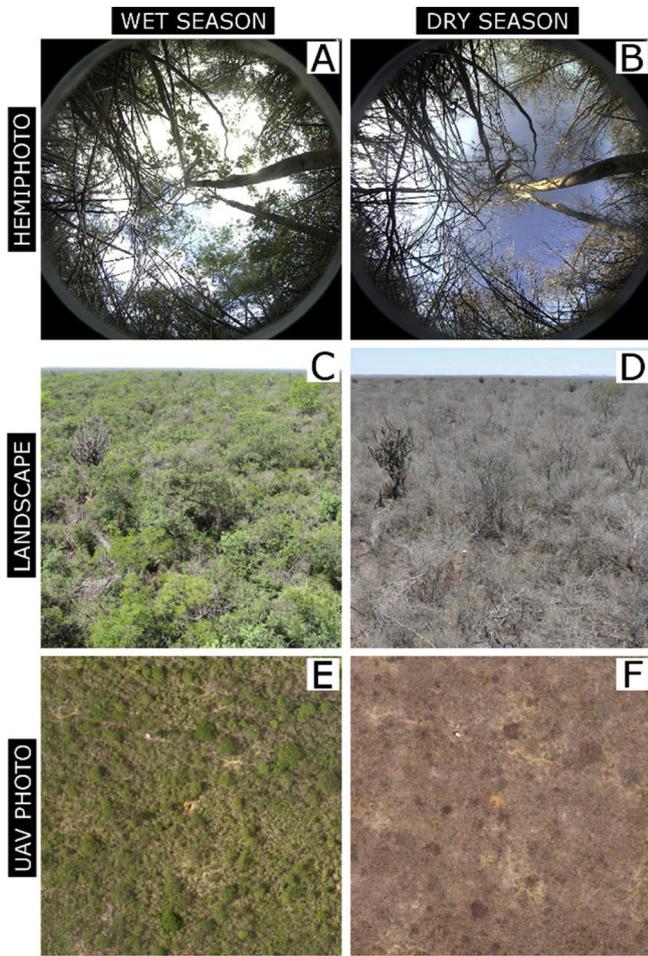


Fig. 2. Contrast in the Caatinga between its wet (A, C and E) and dry (B, D and F) conditions. A–B are hemispheric photos taken from below the vegetation in 12/18/2018 and 9/27/2018 respectively; C–D are landscape photos taken horizontally in 2/5/2016 and 10/20/2017 at the height of 14 m; E–F are orthophotos taken by drone (unmanned aerial vehicle) at 80 m height in 02/16/2018 and 10/20/2017, respectively.

Table 1

Equivalence table of the bands of the sensors TM/Landsat 5, ETM+ /Landsat 7 and OLI/Landsat 8.

OLI/Landsat 8 (nm)	ETM+ /Landsat 7 and TM/Landsat 5 (nm)	Equivalent bands for this study (nm)
–	$\rho_1^{ETM+/TM} = 450\text{--}520$	$\rho_1 = [\rho_2^{OLI}, \rho_1^{ETM+/TM}]$
$\rho_2^{OLI} = 452\text{--}512$	$\rho_2^{ETM+/TM} = 520\text{--}600$	$\rho_2 = [\rho_3^{OLI}, \rho_2^{ETM+/TM}]$
$\rho_3^{OLI} = 533\text{--}590$	$\rho_3^{ETM+/TM} = 630\text{--}690$	$\rho_3 = [\rho_4^{OLI}, \rho_3^{ETM+/TM}]$
$\rho_4^{OLI} = 636\text{--}673$	$\rho_4^{ETM+/TM} = 770\text{--}900$	$\rho_4 = [\rho_5^{OLI}, \rho_4^{ETM+/TM}]$
$\rho_5^{OLI} = 851\text{--}879$	$\rho_5^{ETM+/TM} = 1,550\text{--}1,750$	$\rho_5 = [\rho_6^{OLI}, \rho_5^{ETM+/TM}]$
$\rho_6^{OLI} = 1,566\text{--}1,651$	–	–
$\rho_7^{OLI} = 2,107\text{--}2,294$	$\rho_7^{ETM+/TM} = 2,090\text{--}2,350$	$\rho_7 = [\rho_7^{OLI}, \rho_7^{ETM+/TM}]$

sensors (Table 1). These products are freely available through the LSDS Science Research and Development (LSRD) database of the U.S. Geological Survey (<https://espa.cr.usgs.gov/>).

2.4. Model calibrations

We developed PAI and LAI models based on the combinations of bands (ρ_1 to ρ_7); vegetation indices (*NDVI*, *SAVI* and *EVI*; Eqs. 2 to 4); transformation functions, i.e., x , $1/x$, $\ln(x)$, $\log_{10}(x)$, \sqrt{x} , x^2 , e^x ; and basic

mathematical operations. These models were obtained by using an exhaustive training iteration process ($> 10^6$ iterations) that selected the best results based on the highest coefficient of determination (r^2) with the lowest Root Mean Square Error (RMSE). We used the Percent Bias (PBIAS) and the concordance correlation coefficient (ρ_c) as auxiliary performance indices. In our regression analysis, we used linear, logarithmic, exponential, and power functions to fit the observed data. We obtained *NDVI*, *SAVI*, and Enhanced Vegetation Index (*EVI*) using Eqs. 2 to 4, where $C1$ (6) and $C2$ (7.5) are the coefficients of the aerosol resistance, G (2.5) is a gain factor, and L is the soil effect constant, according to Rouse et al. (1974) and Huete (1988). Our L for the *EVI* and *SAVI* were set using a sensitivity analysis, varying the factor L from -1 to 1 with intervals of 0.01. The best L value occurred when simulated data achieved the highest r^2 with the lowest RMSE. The L values found were 0.07 (for PAI models) and 0.37 (for the LAI models) for equations using *SAVI*; and 1 for both PAI and LAI models for equations using *EVI*. The number of models evaluated can be calculated using Eq. 5, where nc is the number of parameters entered into the model. All independent data were previously tested with the Variance Inflation Factor ($VIF = 1/(1 - r^2)$) to avoid any significant multicollinearity. We considered data to be independent when $VIF < 10$. All processing was performed using an interpreter Python 2.7.15 with only basic modules installed (freely available at <https://github.com/razeayres/correlator>).

$$NDVI = \left(\frac{\rho_4 - \rho_3}{\rho_4 + \rho_3} \right) \tag{2}$$

$$SAVI = \frac{(1 + L) \times (\rho_4 - \rho_3)}{L + \rho_4 + \rho_3} \tag{3}$$

$$EVI = G \times \frac{\rho_4 - \rho_3}{\rho_4 + C1 \times \rho_3 - C2 \times \rho_1 + L} \tag{4}$$

$$f(nc) = C_{\begin{matrix} \rho_1 & \rho_4 & NDVI \\ \rho_2 & \rho_5 & SAVI \\ \rho_3 & \rho_7 & EVI \end{matrix}}^{nc} \times C_{\begin{matrix} x & 1/x & \ln(x) \\ \log_{10}(x) & \sqrt{x} & x^2 \\ e^x & & \end{matrix}}^{nc} \times \begin{cases} 1, & nc = 1 \\ C_{\begin{matrix} + \\ \div \end{matrix}}^{(nc-1)} + (nc-2), & nc \geq 2 \end{cases} \tag{5}$$

2.5. Models verification

To verify the accuracy of all models in this study, we first assessed the applicability of parametric statistics to all data with the Shapiro–Wilk (for normality) and Brown–Forsythe (for homoscedasticity) tests (Zar, 1996), and then we conducted a comparison between the remotely sensed data and the estimates from the field observations using the Monte Carlo cross-validation technique (Xu and Liang, 2001), considering 91 different sampling sizes varying from 5 to 95% of the total data at 1% intervals. Each sample was evaluated by its r^2 and computed as the mean of 50 random repetitions. The methods of cross-validation are widely adopted, and they were used to check whether models tend to over-adjust to the *in situ* dataset distribution (Hawkins, 2004). This over-adjustment would mean that excellent results would be obtained only in calibration (Shao, 1993), while during verification, the accuracy of the model would drastically drop. This approach allows for a good calibration (Shao, 1993). In addition, we used the models proposed by Bastiaanssen (1998) (Eq. 6), Galvncio et al. (2013) (Eq. 7), Machado (2014) (Eq. 8), and Almeida et al. (2019) (Eq. 9), and derived from MODIS data (MCD15A3H/A2H) to produce independent data required for comparing with our field observations and for testing our models. Except for the model developed by Bastiaanssen (2018), these other models were specifically developed for the Caatinga. However, the model of Bastiaanssen (1998) has been widely used to estimate LAI in this region (e.g., Bezerra et al., 2014;

Oliveira et al., 2015; Santos et al., 2017).

$$LAI = -\frac{\ln\left[\frac{(0.69 - SAVI)^{0.59}}{0.59}\right]}{0.91} \tag{6}$$

$$LAI = e^{1.426 + \frac{-0.542}{NDVI}} \tag{7}$$

$$LAI = 0.102 \times e^{5.341 \times NDVI} \tag{8}$$

$$LAI = 9.555 \times EVI - 1.324 \tag{9}$$

For Eqs. 6 to 9, we used the same Landsat dataset produced for the models calibrations; for the MODIS MCD15A2H/A3H products, we used all images for the entire study period (total of 830 candidate images). These products are designed to provide data with a spatial resolution of 500 m every four days (MCD15A3H) or every eight days (MCD15A2H). They are based on a complex algorithm that uses both the daily surface reflectance values of the MODIS sensor on one or both of the Terra and Aqua satellites and the data from a radiative transfer model, which are stored in a two-dimensional lookup table (Yang et al., 2006). These reflectance data are already corrected for atmospheric interferences such as atmospheric gases and aerosols, and they are freely available through the Earth Explorer online tool of the U.S. Geological Survey (<https://earthexplorer.usgs.gov/>). For all products, scale corrections were performed using the Geospatial Data Abstraction Library and clear land dry forest pixels were filtered using the Quality Band (QA band, value 0).

3. Results and discussion

Six of our selected models use *NDVI*, *SAVI* and *EVI* as input (Eqs. 15 to 17, and 23 to 25 in Table 2). These models exhibited r^2 values ranging from 0.77 to 0.79 for PAI and 0.78 to 0.81 for LAI, and RMSE with a minimum of $0.41 \text{ m}^2 \text{ m}^{-2}$ for PAI and $0.40 \text{ m}^2 \text{ m}^{-2}$ for LAI. The *SAVI* models (Eqs. 15 and 23) showed RMSE values smaller than the ones found for the models that use *EVI* and *NDVI*. We ascribe the better accuracy with the *SAVI* models over the other vegetation indices to the fact that *SAVI* takes into consideration the effects of soil background, while not showing high variability as *EVI* does for sparsely vegetated areas, which in turn produces infrared reflectance at low levels due to

dry soil background (Lu et al., 2015). In addition, *SAVI* better reflects the surface roughness, which affects momentum, heat, and water vapor fluxes (Bastiaanssen, 1998), and varies according to the phenological stages of the Caatinga (Teixeira et al., 2008). These models are useful because they allow easy retrieval of the PAI or LAI from remote sensing data. For example, many *NDVI* products, using a large variety of sensor data, are freely available, and they can be used to acquire physical information for large forest areas.

Our models presented a better performance when fitted linearly rather than in any other non-linear form (Table 2). This is the opposite of what was shown by some *NDVI*-LAI relationship models (Liu et al., 2012; Tavakoli et al., 2014). Liu et al. (2012) conducted an experiment in the Ningxia Hui Autonomous District, in Northwest China, and they found saturation of *NDVI* at high LAI values. Tavakoli et al. (2014), in 16 plots of winter wheat (*Triticum aestivum* L., cv. Cubus) in an experimental station located in Marquardt in Germany, found the best *NDVI*-LAI relation when fitting data logarithmically. In fact, this saturation of LAI as function of *NDVI* is commonly expressed by a logarithmic relationship. However, *NDVI* values tend to be poorly associated with those from ground observations in SDTFs (Guzmán et al., 2019). In our study, the vegetation indices did not exhibit saturation related to the LAI of the Caatinga vegetation, which resulted in a linear covariance as reflect in Eqs. 23–25. Magalhães et al. (2018) showed that a linear model simulates better LAI in a SDTF by arguing that *NDVI* can only saturate in vegetation types with LAI above $5 \text{ m}^2 \text{ m}^{-2}$. That supports our findings since this threshold is above the values we used to develop LAI models for the Caatinga.

In this study, absolute non-saturated simulated LAI values varied from 0 to $4.56 \text{ m}^2 \text{ m}^{-2}$ (0.61 to $5.23 \text{ m}^2 \text{ m}^{-2}$ for PAI values), while Bastiaanssen (1998) exhibited values from 0 to ca. $4.45 \text{ m}^2 \text{ m}^{-2}$ (considering only the non-saturated values), Galvêncio et al. (2013) from 0.63 to $1.98 \text{ m}^2 \text{ m}^{-2}$, Machado (2014) from 0.25 to $3.7 \text{ m}^2 \text{ m}^{-2}$, and Almeida et al. (2019) from ca. 0 up to $4.26 \text{ m}^2 \text{ m}^{-2}$ in average. All of these previously models do not consider the temporal variations due to the phenological stages of the Caatinga on a continuous multi-year basis, thus the range of possible simulated values is smaller when compared to our models. Bastiaanssen (1998) derived LAI using different equations for only seven types of land use cover types (cotton,

Table 2
Calibration of PAI and LAI models created through an iterative process using Landsat reflectance data.

	Model		r^2 ¹	RMSE ²	ρ_c	PBIAS ²
PAI	Eq. 10	$y = 10.1 \times (\rho_4 - \sqrt{\rho_3}) + 3.1$	0.79	0.41	0.88	0.33
	Eq. 11	$y = -13.2 \times (\sqrt{\rho_2} - \rho_4) + 3.1$	0.77	0.44	0.87	1.84
	Eq. 12	$y = -13.5 \times \left(\frac{\log_{10}(\rho_4)}{\ln(\rho_3)}\right) + 6.1$	0.77	0.43	0.87	-1.84
	Eq. 13	$y = -20.3 \times (\rho_3 - \rho_4^2) + 3$	0.77	0.43	0.87	-0.83
	Eq. 14	$y = -3.2 \times (\ln(\rho_3) \times \sqrt{\rho_4}) - 1.4$	0.79	0.41	0.88	-0.22
	Eq. 15 ³	$y = 3.5 \times (e^{SAVI}) - 2.7$	0.79	0.41	0.88	1.10
	Eq. 16	$y = 4.8 \times (e^{EVI}) - 3.7$	0.77	0.45	0.86	3.72
	Eq. 17	$y = 5 \times (NDVI^2) + 1.3$	0.79	0.43	0.89	1.04
	LAI	Eq. 18	$y = \left(\frac{\rho_4^2}{\rho_1}\right) - 0.1$	0.79	0.41	0.88
Eq. 19		$y = -9.7 \times \left(\frac{\log_{10}(\rho_3)}{\left(\frac{1}{\rho_4}\right)}\right) - 1.2$	0.78	0.42	0.88	-4.84
Eq. 20		$y = 11.2 \times (\sqrt{\rho_4} - e^{\rho_3}) + 8.3$	0.76	0.44	0.86	7.15
Eq. 21		$y = 12.2 \times (\sqrt{\rho_4} - \sqrt{\rho_2}) - 1.2$	0.76	0.44	0.86	-0.73
Eq. 22		$y = 19.6 \times (\rho_4^2 - e^{\rho_3}) + 21.4$	0.78	0.42	0.87	-3.01
Eq. 23 ³		$y = 11 \times (SAVI^2) + 0.2$	0.81	0.40	0.89	0.04
Eq. 24		$y = 6.5 \times (EVI) - 0.4$	0.78	0.42	0.88	-5.71
Eq. 25		$y = 4.9 \times (NDVI^2) + 0.1$	0.80	0.41	0.89	4.39

¹ Significant at $p = 0.05$.

² RMSE is in $\text{m}^2 \text{ m}^{-2}$, and PBIAS is showed as percentage.

³ L values in the *SAVI* calculations were 0.07 (for the PAI) and 0.37 (for the LAI).

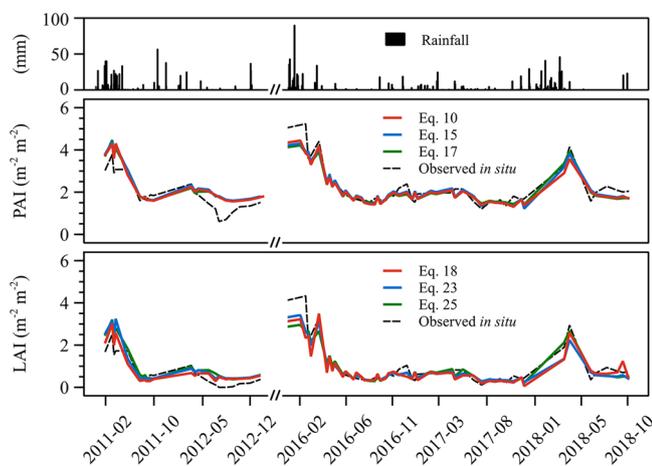


Fig. 3. Comparison of temporal variation of PAI (in the middle) and LAI (bottom) observed *in situ* to the best simulated models created through an iterative process using Landsat reflectance data.

maize, soy, wheat, fruit trees, vegetables, and native forests), none of which were similar to the dry forest in our study area. The study of Galvncio et al. (2013) was based on a comparison of data obtained using an AccuPAR analyzer with indices created from spectroradiometry from a single day of measurements. The model proposed by Machado (2014) was developed in a Caatinga area of the National Park of Catimbau using only one Landsat 5 TM image, combined with 54 field-derived LAI measurements acquired three times over 20 days using simultaneous averages of diffuse light interception at five different zenith angles using sensors with fisheye lens. Almeida et al. (2019) created LAI models using the litterfall trap method in the Caatinga, and collected data for three representative species. Although the LAI models of Almeida et al. (2019) show high correlation to their field measurements, they were not able to consider the entire growth cycle in their analysis due to limitations of the litterfall fall method such as the discrete distribution of the measurements over time, and misestimation of the foliar development because of the appearance of new leaves between assessments.

The best-performing new models that use different band combinations were Eqs. 10–14, and 18–22 (Table 2). These equations may represent other physical processes and components, such as leaf etiolation and increasing protochlorophyll, which is reported to influence the blue band of the visible spectrum (ρ_1 in Eq. 17) (Gates et al., 1965). Medeiros et al. (2019) suggested the near-infrared (NIR) band may be a good indicator of leaf radiation reflectance patterns among different species, which reflect variations in leaf size, form, and type, and even plant habit. Our models used the NIR band, and they varied only to the inclusion of the red, green, and blue bands. The amount of energy reflected or absorbed in these bands varies according to the physicochemical and biophysical properties of the target (Edwards et al., 2013). All bodies reflect or emit electromagnetic radiation at different wavelengths and in different ways, and the result is a reflectance curve or spectral signature (Schmugge et al., 2002). This set of unique interactions restricts the bands that distinguish certain characteristics of a target and allows various parameters quantification (e.g., pigment concentration and plant structure complexity) (Blackburn, 2007; Dawson et al., 1998; Schmugge et al., 2002). Usually, vegetation reflects about half of the incident radiant flux in the NIR band (Zhao et al., 2007); therefore, this is a band very sensitive to biomass and LAI. Leaves predominantly absorb energy at the blue–red spectrum and reflect the energy in the green and NIR bands because of the interaction with chlorophyll, carotenoids, and the mesophyll itself (Gates et al., 1965). Thus, the green and NIR bands are considered bands of high reflectance (Fan et al., 2018). In comparison to the green band, the NIR has a relatively higher multiple reflectance through within-canopy

layers, which reduces the canopy light extinction coefficient (Zheng and Moskal, 2009).

We consider Eqs. 10, 15, and 17 to be optimal solutions for the estimation of PAI, and Eqs. 18, 23, and 25 for the estimation of LAI in the Caatinga (Table 2). Although studies have highlighted the dubious quality of data acquired by remote sensing in the blue band because of wavelength-dependent atmospheric interference (e.g., Carter et al., 2009; Motohka et al., 2009), Eq. 18 has performed very well with $r^2 = 0.79$ and $RMSE = 0.41 \text{ m}^2 \text{ m}^{-2}$, with values comparable to Eq. 23 (which does not use the blue band) with $r^2 = 0.81$ and $RMSE = 0.40 \text{ m}^2 \text{ m}^{-2}$. The greatest contribution of Eq. 18 is its natural proximity to a 1:1 relation to *in situ* measurements ($\rho_c = 0.88$, $PBIAS = -0.01$), which provides greater ability to simulate values near to zero. Although our models require observations in the NIR band, many images have NIR sensors and, if well calibrated, they allow for LAI to be estimated based on reflectance from spectral mixture or coarse resolution composites. These images include those captured by phenological cameras, unmanned aerial vehicles, and high-resolution monitoring satellites (e.g., QuickBird and IKONOS).

The accuracy of our best models can be visualized when plotting their estimates alongside observed PAI and LAI data (Fig. 3). Our best performance models were able to emulate the variance of LAI in our study period (Table 2). Eqs. 10, 15, and 23 are biased towards overestimation ($PBIAS = 0.33, 1.10$ and 0.04 respectively) and Eq. 18 presented a small underestimation bias ($PBIAS = -0.01$). In general, based on our findings, models developed with independent bands of a sensor produced low-magnitude values near-optimal zero, indicating accurate model simulation, while the models created using vegetation indices exhibited moderate bias.

In the cross-validation analyses, Eq. 10 produced maximum values ($r_{max} = 0.81$) and minimum values ($r_{min} = 0.68$) similar to Eq. 18 ($r_{max} = 0.83$ and $r_{min} = 0.59$). In comparison, the *NDVI*, *SAVI* and *EVI* models yielded higher maximum and minimum values for both PAI ($r_{max} = 0.83, 0.82$, and 0.80 , respectively; $r_{min} = 0.73$ for all models) and LAI ($r_{max} = 0.84, 0.85$, and 0.83 ; $r_{min} = 0.71, 0.71$, and 0.69 , respectively). The other models presented r values ranging from 0.64 to 0.84 (for PAI), and 0.6 to 0.83 (for LAI). The 0.01 standard deviation of r was the same for all models, indicating that they are reliable. When varying the amount of data taken for cross-validation from 5 to 95%, the r values tend to show minimal variations (Figs. 4 and 5). However, in our verification we did not observe any statistically significant pattern in correlation owing to the removal of data from the calibration of the models, which confirms that these models are highly robust to estimate LAI.

The approaches proposed by Bastiaanssen (1998) (Eq. 6), Galvncio et al. (2013) (Eq. 7), Machado (2014) (Eq. 8), and Almeida et al. (2019) (Eq. 9), as well as the MCD15A3H/A2H data underperformed compared to our models, when compared to our *in situ* PAI and LAI data (Table 3), even though correlations were significant ($p < 0.05$). Eqs. 6 and 7 performed the best in terms of accuracy, while Eqs. 8 and 9 presented the highest covariances. Although the MODIS products are supposed to reproduce the LAI seasonality well, we observed that they do not respond well for the Caatinga during the dry season; the lowest values were around $0.5 \text{ m}^2 \text{ m}^{-2}$ when real LAI values were practically zero. We tested the fitness of MODIS products against our *in situ* PAI and LAI datasets. The lower correlation with the MCD15A3H product in comparison to the MCD15A2H product indicates that the proportion of high-quality data for the dry forest area is lower for periods of composition of four days than for the eight-day version (Table 3). These periods of composition are created from the highest value observed *in situ*; thus, the greater the number of values available for the determination of LAI of each pixel, the higher the probability of it being an accurate value. This is because data obtained by satellites are influenced by a number of different atmospheric factors such as water vapor, cloud cover, and aerosols, thus any given bit of satellite data may not yield accurate results (Yang et al., 2006). Our results indicate that

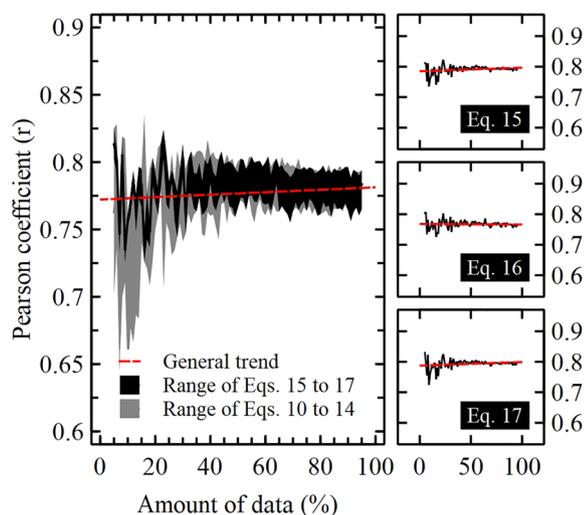


Fig. 4. Cross-validation of the PAI models created through an iterative process using Landsat reflectance data. Detailed validations of Eqs. 15–17 are on the right side.

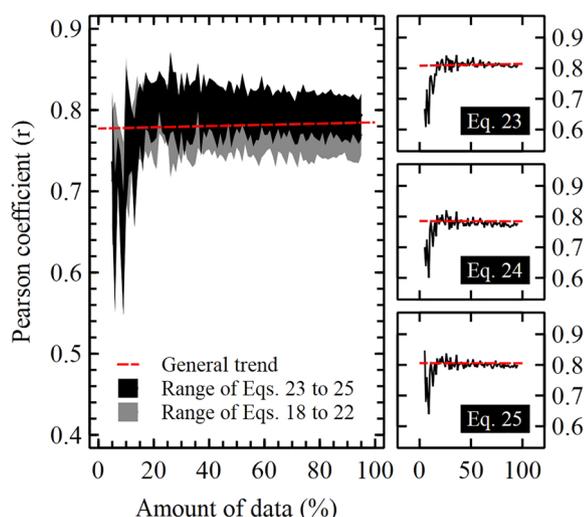


Fig. 5. Cross-validation of the LAI models created through an iterative process using Landsat reflectance data. Detailed cross-validations of Eqs. 23–25 are on the right side.

studies that rely on LAI from MODIS products for vegetation assessment in SDTFs, such as the Caatinga, are likely to incorporate bias due to unrealistically high LAI values during the dry season, when plants lose most of their leaves and the vegetation consist predominantly of non-photosynthetic biomass (Leal et al., 2003; Silva et al., 2017). The reduction of photosynthesis has direct consequences for the evapotranspiration and gross primary productivity, affecting the carbon storage and CO₂ exchange capacity (Morais et al., 2017; Nagler et al., 2003).

The LAI in this study, as seen in all models for the Caatinga, can be defined as effective LAI, which is the portion of LAI that effectively intercepts the light, not directly considering grouped foliage. This grouping of leaves can be quantified by a vegetation dispersion parameter Ω (clumping index) (Nilson, 1971), which often can be determined by a random distribution (Chen and Black, 1992). The “true” LAI is not easy to achieve; it requires intensive fieldwork and systematic sampling, using all possible allometric relationships (Frazer et al., 1997; Weiss et al., 2004). Since the approaches of estimating PAI and LAI used in our study are based on the light extinction, the WAI values as a result of the difference between PAI and LAI are likely to be underestimated

Table 3
Comparison of the previously published models, and the MCD15A3H/A2H products with *in situ* data.

Reference	Parameter	r ² ¹	RMSE ²	ρ_c	PBIAS ²
Bastiaanssen (1998)	PAI	0.73	1.54	0.35	-69.44
	LAI	0.75	0.51	0.84	-26.53
Galvıncio et al. (2013)	PAI	0.73	1.32	0.33	-57.44
	LAI	0.72	0.52	0.75	2.33
Machado (2014)	PAI	0.76	1.28	0.61	-40.48
	LAI	0.78	1.01	0.72	43.12
Almeida et al. (2019)	PAI	0.77	1.66	0.40	-73.74
	LAI	0.77	0.65	0.81	-36.80
MCD15A3H	PAI	0.66	1.39	0.29	-60.08
	LAI	0.65	0.57	0.71	-4.00
MCD15A2H	PAI	0.77	1.26	0.38	-55.07
	LAI	0.78	0.46	0.82	8.04

Summary of selected models					
Eq. 10	PAI	0.79	0.41	0.88	0.33
Eq. 15 ³	PAI	0.79	0.41	0.88	1.10
Eq. 17	PAI	0.79	0.41	0.89	1.04
Eq. 18	LAI	0.79	0.41	0.88	-0.01
Eq. 23 ³	LAI	0.81	0.40	0.89	0.04
Eq. 25	LAI	0.80	0.41	0.89	4.39

¹ Significant at p = 0.05.
² RMSE is in m² m⁻², and PBIAS is showed as percentage.
³ L values in the SAVI calculations were 0.07 (for the PAI) and 0.37 (for the LAI).

when LAI is high (Nackaerts et al., 2000; Stenberg, 1996). This is attributed to the fact that when LAI values are very high, the leaves cover the woody area and reduce the role of light interception of the branches and stems (Chen et al., 1997), which in turn leaves the PAI and LAI values very similar (e.g., Feb 2012 in Fig. 3).

Our models are easy-to-use PAI and LAI predictors that can be applied to estimate these indices for the Caatinga. The models also can be used to simulate other Caatinga types (such as in transitional areas), but since they rely on calibration coefficients, minor adjustments might be required to approximate minimum and maximum LAI. Regional applicability can be considered as moderate-high, because the phytophysiology dominated by shrubs is the main and most abundant in the Caatinga (Silva et al., 2017). However, at a regional scale, our models may be used as backup models in a physical approach that does not require calibration to achieve maximum generalization. Further improvements may include (i) pooling coefficients adjusted for other areas of Caatinga with different levels of degradation, which could be similar to what was made by Bastiaanssen (1998) when developing LAI models; (ii) the adjustment of these equations using field data from other types of Caatinga vegetation, where some plants, such as Cactaceae and Bromeliaceae, may have a more significant presence, and the soil exposure may be different; (iii) the removal of the influence of non-photosynthetic plant material, such as flowers, fruits and petioles, on LAI measurements; and (iv) approximation of LAI to more realistic values, developing and introducing a new Ω to more efficiently account for leaf dispersion directly in the models, instead of abstracting it in regression coefficients. This could solve systematic problems, such as misestimation of LAI at a given phenological stage.

4. Conclusions

Our study developed and assessed several PAI and LAI models to be realistically representative for the phenology of a typical Caatinga ecosystem. Given the high frequency of our *in situ* measurements (mostly measured on a daily or weekly basis), all Caatinga phenological stages were covered and reproduced in our models. The joint usage of ground and satellite data presented an efficient way to assess both PAI and LAI models. The results included parameterizations with the visible

and infrared spectral bands, which allowed the use of many currently available datasets to estimate LAI.

The models produced results with high accuracy (up to $r^2 = 0.81$ and $\text{RMSE} = 0.41 \text{ m}^2 \text{ m}^{-2}$). The significant improvement of our models over the others used for the Caatinga is due to the consideration of WAI, which previously had not been considered in calibrations for the Caatinga, and the temporal variations of LAI, which allowed us to create more generalist models that can be used during different phenological stages of the Caatinga vegetation.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Acknowledgements

We thank CAPES (Brazilian Coordination for the Improvement of Higher Level Personnel) for funding this study through the project PVE A103/20132015, FACEPE (*Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco*) for funding this through the projects FACEPE APQ 0646-9.25/16 and FACEPE APQ 0062-1.07/15 (Caatinga-FLUX), and CNPq (National Council for Scientific and Technological Development of Brazil) for funding this through the project < GN2 > 402834/2016-0 < GN2 > - CNPq Universal 01/2016. This work is also supported by the UK/Brazil Nordeste project, funded jointly through the UK Natural Environment Research Council (NE/N012526/1 ICL and NE/N012488/1 UoR) and FAPESP (The São Paulo Research Foundation) (FAPESP 2015/50488-5).

References

- Ab'Saber, A.N., 1974. O domínio morfoclimático Semi-Árido das caatingas brasileiras. *Geomorfologia* 1–39.
- Almeida, C.L., Carvalho, T.R.Ade, Araújo, J.C., 2019. Leaf area index of Caatinga biome and its relationship with hydrological and spectral variables. *Agric. For. Meteorol.* 279, 107705. <https://doi.org/10.1016/j.agrformet.2019.107705>.
- Alvares, C.A., Stape, J.L., Sentelhas, P.C., Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. *Meteorol. Zeitschrift* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large-area hydrologic modeling and assessment: part I. Model development. *J. Am. Water Resour. Assoc.* 34, 73–89. <https://doi.org/10.1111/j.1752-1688.1998.tb05961.x>.
- Bastiaanssen, W., 1998. *Remote Sensing in Water Resources Management: The State of the Art*, ix. ed. International Water Management Institute (IWMI), Colombo, Sri Lanka.
- Bezerra, J.M., Moura, G.B.D.A., Silva, B.B., Lopes, P.M.O., Silva, Ê.F.F., 2014. Parâmetros biofísicos obtidos por sensoriamento remoto em região semiárida do estado do Rio Grande do Norte, Brasil. *Rev. Bras. Eng. Agrícola e Ambient* 18, 73–84. <https://doi.org/10.1590/S1415-43662014000100010>.
- Bieger, K., Arnold, J.G., Rathjens, H., White, M.J., Bosch, D.D., Allen, P.M., Volk, M., Srinivasan, R., 2017. Introduction to SWAT+, a completely restructured version of the soil and water assessment tool. *JAWRA J. Am. Water Resour. Assoc.* 53, 115–130. <https://doi.org/10.1111/1752-1688.12482>.
- Blackburn, G.A., 2007. Hyperspectral remote sensing of plant pigments. *J. Exp. Bot.* 58, 855–867. <https://doi.org/10.1093/jxb/erl123>.
- Bonan, G.B., 1995. Land-Atmosphere interactions for climate system models: coupling biophysical, biogeochemical, and ecosystem dynamical processes. *Remote Sens. Environ.* 51, 57–73. [https://doi.org/10.1016/0034-4257\(94\)00065-U](https://doi.org/10.1016/0034-4257(94)00065-U).
- Carter, G., Lucas, K., Blossom, G., Lassiter, C., Holiday, D., Mooneyhan, D., Fastring, D., Holcombe, T., Griffith, J., 2009. Remote sensing and mapping of Tamarisk along the Colorado River, USA: a comparative use of summer-acquired Hyperion, Thematic Mapper and QuickBird data. *Remote Sens.* 1, 318–329. <https://doi.org/10.3390/rs1030318>.
- Chen, J.M., Black, T.A., 1992. Foliage area and architecture of plant canopies from sunfleck size distributions. *Agric. For. Meteorol.* 60, 249–266. [https://doi.org/10.1016/0168-1923\(92\)90040-B](https://doi.org/10.1016/0168-1923(92)90040-B).
- Chen, J.M., Rich, P.M., Gower, S.T., Norman, J.M., Plummer, S., 1997. Leaf area index of boreal forests: theory, techniques, and measurements. *J. Geophys. Res. Atmos.* 102, 29429–29443. <https://doi.org/10.1029/97JD01107>.
- Cunha, J., Nóbrega, R.L.B., Rufino, I., Erasmi, S., Galvão, C., Valente, F., 2019. Surface albedo as a proxy for land-cover clearing in seasonally dry forests: evidence from the Brazilian Caatinga. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2019.111250>.
- Dawson, T.P., Curran, P.J., Plummer, S.E., 1998. LIBERTY—modeling the effects of leaf biochemical concentration on reflectance spectra. *Remote Sens. Environ.* 65, 50–60. [https://doi.org/10.1016/S0034-4257\(98\)00007-8](https://doi.org/10.1016/S0034-4257(98)00007-8).
- Edwards, B.L., Namikas, S.L., D'Sa, E.J., 2013. Simple infrared techniques for measuring beach surface moisture. *Earth Surf. Process. Landforms* 38, 192–197. <https://doi.org/10.1002/esp.3319>.
- Fan, X., Kawamura, K., Guo, W., Xuan, T.D., Lim, J., Yuba, N., Kurokawa, Y., Obitsu, T., Lv, R., Tsumiyama, Y., Yasuda, T., Wang, Z., 2018. A simple visible and near-infrared (V-NIR) camera system for monitoring the leaf area index and growth stage of Italian ryegrass. *Comput. Electron. Agric.* 144, 314–323. <https://doi.org/10.1016/j.compag.2017.11.025>.
- Fotis, A.T., Morin, T.H., Fahey, R.T., Hardiman, B.S., Bohrer, G., Curtis, P.S., 2018. Forest structure in space and time: biotic and abiotic determinants of canopy complexity and their effects on net primary productivity. *Agric. For. Meteorol.* 250–251, 181–191. <https://doi.org/10.1016/j.agrformet.2017.12.251>.
- Frazier, G.W., Lertzman, K.P., Trofymow, J.A., 1997. A Method for Estimating Canopy Openness, Effective Leaf Area Index, and Photosynthetically Active Photon Flux Density Using Hemispherical Photography and Computerized Image Analysis Techniques. *Pacific Forestry Centre, Victoria, B. C.*
- Galvêncio, J.D., Moura, M.S.B., Silva, T.G., Silva, B.B., Naue, C.R., 2013. LAI improved to dry forest in Semiarid of the Brazil. *Int. J. Remote Sens. Appl.* 3, 193. <https://doi.org/10.14355/ijrsa.2013.0304.04>.
- Gates, D.M., Keegan, H.J., Schleter, J.C., Weidner, V.R., 1965. Spectral properties of plants. *Appl. Opt.* 4, 11. <https://doi.org/10.1364/AO.4.000011>.
- Guzmán, J.A., Sanchez-Azofeifa, G.A., Espírito-Santo, M.M., 2019. MODIS and PROBA-V NDVI products differ when compared with observations from phenological towers at four tropical dry forests in the Americas. *Remote Sens. (Basel)* 11, 2316. <https://doi.org/10.3390/rs11192316>.
- Hawkins, D.M., 2004. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* 44, 1–12. <https://doi.org/10.1021/ci0342472>.
- Huete, A., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).
- Jacquemoud, S., Baret, F., 1990. PROSPECT: a model of leaf optical properties spectra. *Remote Sens. Environ.* 34, 75–91. [https://doi.org/10.1016/0034-4257\(90\)90100-Z](https://doi.org/10.1016/0034-4257(90)90100-Z).
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+SAIL models: a review of use for vegetation characterization. *Remote Sens. Environ.* 113, S56–S66. <https://doi.org/10.1016/j.rse.2008.01.026>.
- Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F., 2004. Review of methods for in situ leaf area index determination part I. Theories, sensors and hemispherical photography. *Agric. For. Meteorol.* 121, 19–35. <https://doi.org/10.1016/j.agrformet.2003.08.027>.
- Kalaeska, M., Calvo-Alvarado, J.C., Sanchez-Azofeifa, G.A., 2005. Calibration and assessment of seasonal changes in leaf area index of a tropical dry forest in different stages of succession. *Tree Physiol.* 25, 733–744. <https://doi.org/10.1093/treephys/25.6.733>.
- Kiill, L.H.P., 2017. *Caracterização da vegetação da reserva legal da Embrapa Semiárido. Doc. 281, Série Documentos.*
- Knote, C., Bonafe, G., Di Giuseppe, F., 2009. Leaf area index specification for use in mesoscale weather prediction systems. *Mon. Weather Rev.* 137, 3535–3550. <https://doi.org/10.1175/2009MWR2891.1>.
- Knyazikhin, Y., Martonchik, J.V., Diner, D.J., atmosphere-corrected M. data, Myneni, R.B., Verstraete, M., Pinty, B., Gobron, N., 1998. Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data. *J. Geophys. Res. Atmos.* 103, 32239–32256. <https://doi.org/10.1029/98JD02461>.
- Leal, I.R., Tabarelli, M., Silva, J.M.C., 2003. *Ecologia e conservação da Caatinga. Editora Universitária da UFPE, Recife.*
- Leal, I.R., Wirth, R., Tabarelli, M., 2007. Seed dispersal by ants in the semi-arid Caatinga of North-East Brazil. *Ann. Bot.* 99, 885–894. <https://doi.org/10.1093/aob/mcm017>.
- Liu, F., Qin, Q., Zhan, Z., 2012. A novel dynamic stretching solution to eliminate saturation effect in NDVI and its application in drought monitoring. *Chinese Geogr. Sci.* 22, 683–694. <https://doi.org/10.1007/s11769-012-0574-5>.
- Lu, L., Kuenzer, C., Wang, C., Guo, H., Li, Q., 2015. Evaluation of three MODIS-derived Vegetation Index time series for dryland vegetation dynamics monitoring. *Remote Sens.* 7, 7597–7614. <https://doi.org/10.3390/rs70607597>.
- Machado, C.C.C., 2014. *Alterações na superfície do Parque Nacional do Catimbau (PE-Brasil): consolidação dos aspectos biofísicos na definição dos indicadores ambientais do bioma Caatinga. Universidade Federal de Pernambuco.*
- Machado, I.C.S., Barros, L.M., Sampaio, E.V.S.B., 1997. Phenology of caatinga species at Serra Talhada, PE, Northeastern Brazil. *Biotropica* 29, 57–68. <https://doi.org/10.1111/j.1744-7429.1997.tb00006.x>.
- Magalhães, S.F., Calvo-Rodríguez, S., Do Espírito Santo, M.M., Sánchez Azofeifa, G.A., 2018. Determining the K coefficient to leaf area index estimations in a tropical dry forest. *Int. J. Biometeorol.* 62, 1187–1197. <https://doi.org/10.1007/s00484-018-1522-6>.
- Medeiros, E.Se S., Machado, C.C.C., Galvêncio, J.D., Moura, M.S.B., Araujo, H.F.P., 2019. Predicting plant species richness with satellite images in the largest dry forest nucleus in South America. *J. Arid Environ.* 166, 43–50. <https://doi.org/10.1016/j.jaridenv.2019.03.001>.
- Menezes, R.S.C., Sampaio, E.V.S.B., Giongo, V., Pérez-Marin, A.M., 2012. Biogeochemical cycling in terrestrial ecosystems of the Caatinga Biome. *Braz. J. Biol.* 72, 643–653.
- Miranda, R.D.Q., Galvêncio, J.D., Moraes, Y.C.B., Moura, M.S.B., Jones, C.A., Srinivasan, R., 2018. Dry forest deforestation dynamics in Brazil's Pontal basin. *Rev. Caatinga* 31, 385–395. <https://doi.org/10.1590/1983-21252018v31n215rc>.
- Moraes, Y.C.B., Araújo, Mdo S.B., Moura, M.S.B., Galvêncio, J.D., Miranda, R.Q., 2017. Análise do sequestro de carbono em áreas de Caatinga do semiárido pernambucano. *Rev. Bras. Meteorol.* 32, 585–599. <https://doi.org/10.1590/0102-7786324007>.
- Motokha, T., Nasahara, K.N., Miyata, A., Mano, M., Tsuchida, S., 2009. Evaluation of

- optical satellite remote sensing for rice paddy phenology in monsoon Asia using a continuous in situ dataset. *Int. J. Remote Sens.* 30, 4343–4357. <https://doi.org/10.1080/01431160802549369>.
- Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.* 111, 519–536. <https://doi.org/10.1016/j.rse.2007.04.015>.
- Nackaerts, K., Coppin, P., Muys, B., Hermy, M., 2000. Sampling methodology for LAI measurements with LAI-2000 in small forest stands. *Agric. For. Meteorol.* 101, 247–250. [https://doi.org/10.1016/S0168-1923\(00\)00090-3](https://doi.org/10.1016/S0168-1923(00)00090-3).
- Nagler, P.L., Inoue, Y., Glenn, E., Russ, A., Daughtry, C.S., 2003. Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes. *Remote Sens. Environ.* 87, 310–325. <https://doi.org/10.1016/j.rse.2003.06.001>.
- Nilson, T., 1971. A theoretical analysis of the frequency of gaps in plant stands. *Agric. Meteorol.* 8, 25–38. [https://doi.org/10.1016/0002-1571\(71\)90092-6](https://doi.org/10.1016/0002-1571(71)90092-6).
- Norman, J.M., 1979. Modelling the complete crop canopy. In: Barfield, B.J., Gerber, J. (Eds.), *Modification of the Aerial Environment*. American Society of Agricultural Engineers, St. Joseph, USA, pp. 249–277.
- Oliveira, L.M.Mde, Montenegro, S.M.G.L., Silva, B.B., Moura, A.E.S.S., 2015. Balanço de radiação por sensoriamento remoto em bacia hidrográfica da Zona da Mata nordestina. *Rev. Bras. Meteorol.* 30, 16–28. <https://doi.org/10.1590/0102-778620130652>.
- Rankine, C., Sánchez-Azofeifa, G.A., Guzmán, J.A., Espirito-Santo, M.M., Sharp, I., 2017. Comparing MODIS and near-surface vegetation indexes for monitoring tropical dry forest phenology along a successional gradient using optical phenology towers. *Environ. Res. Lett.* 12, 105007. <https://doi.org/10.1088/1748-9326/aa838c>.
- Rodriguez, R., Real, P., Espinosa, M., Perry, D.A., 2009. A process-based model to evaluate site quality for *Eucalyptus nitens* in the Bio-Bio Region of Chile. *Forestry* 82, 149–162. <https://doi.org/10.1093/forestry/cpn045>.
- Rouse, J.J.W., Haas, R.R.H., Schell, J.J.A.J., Deering, D.W., Harlan, J.C., 1974. *Monitoring the Vernal Advancement and Retrogradation (greenwave Effect) of Natural Vegetation, NASA/GSFC Type III Final Report*. NASA Goddard Space Flight Centre, Greenbelt, USA.
- Santos, C.A.G., Silva, R.M., Silva, A.M., Brasil Neto, R.M., 2017. Estimation of evapotranspiration for different land covers in a Brazilian semi-arid region: a case study of the Brígida River basin, Brazil. *J. South Am. Earth Sci.* 74, 54–66. <https://doi.org/10.1016/j.jsames.2017.01.002>.
- Schmugge, T.J., Kustas, W.P., Ritchie, J.C., Jackson, T.J., Rango, A., 2002. Remote sensing in hydrology. *Adv. Water Resour.* 25, 1367–1385. [https://doi.org/10.1016/S0309-1708\(02\)00065-9](https://doi.org/10.1016/S0309-1708(02)00065-9).
- Shao, J., 1993. Linear model selection by Cross-Validation. *J. Am. Stat. Assoc.* 88, 486. <https://doi.org/10.2307/2290328>.
- Silva, G.C., Sampaio, E.V.D.S.B., 2008. Biomassas de partes aéreas em plantas da caatinga. *Rev. Árvore* 32, 567–575. <https://doi.org/10.1590/S0100-67622008000300017>.
- Silva, J.M.C., Leal, I.R., Tabarelli, M., 2017. *Caatinga: The Largest Tropical Dry Forest Region in South America*. Springer International Publishing, Cham, Switzerland. <https://doi.org/10.1007/978-3-319-68339-3>.
- Stenberg, P., 1996. Correcting LAI-2000 estimates for the clumping of needles in shoots of conifers. *Agric. For. Meteorol.* 79, 1–8. [https://doi.org/10.1016/0168-1923\(95\)02274-0](https://doi.org/10.1016/0168-1923(95)02274-0).
- Tavakoli, H., Mohtasebi, S.S., Alimardani, R., Gebbers, R., 2014. Evaluation of different sensing approaches concerning to nondestructive estimation of leaf area index (LAI) for winter wheat. *Int. J. Smart Sens. Intell. Syst.* 7, 337–359. <https://doi.org/10.21307/ijssis-2017-659>.
- Teixeira, A.H., de, C., Bastiaanssen, W.G.M., Ahmad, M.D., Moura, M.S.B., Bos, M.G., 2008. Analysis of energy fluxes and vegetation-atmosphere parameters in irrigated and natural ecosystems of semi-arid Brazil. *J. Hydrol.* 362, 110–127. <https://doi.org/10.1016/j.jhydrol.2008.08.011>.
- Verhoef, W., 1985. Earth observation modeling based on layer scattering matrices. *Remote Sens. Environ.* 17, 165–178. [https://doi.org/10.1016/0034-4257\(85\)90072-0](https://doi.org/10.1016/0034-4257(85)90072-0).
- Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sens. Environ.* 16, 125–141. [https://doi.org/10.1016/0034-4257\(84\)90057-9](https://doi.org/10.1016/0034-4257(84)90057-9).
- Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., Coppin, P., 2004. Review of methods for in situ leaf area index (LAI) determination Part II. Estimation of LAI, errors and sampling. *Agric. For. Meteorol.* 121, 37–53. <https://doi.org/10.1016/j.agrformet.2003.08.001>.
- Xu, Q.-S., Liang, Y.-Z., 2001. Monte Carlo cross validation. *Chemometr. Intell. Lab. Syst.* 56, 1–11. [https://doi.org/10.1016/S0169-7439\(00\)00122-2](https://doi.org/10.1016/S0169-7439(00)00122-2).
- Yang, W., Tan, B., Huang, Dong, Rautiainen, M., Shabanov, N.V., Wang, Y., Privette, J.L., Huemmrich, K.F., Fensholt, R., Sandholt, I., Weiss, M., Ahl, D.E., Gower, S.T., Nemani, R.R., Knyazikhin, Y., Myneni, R.B., 2006. MODIS leaf area index products: from validation to algorithm improvement. *IEEE Trans. Geosci. Remote Sens.* 44, 1885–1898. <https://doi.org/10.1109/TGRS.2006.871215>.
- Zar, J.H., 1996. *Biostatistical Analysis*, 3rd ed. Prentice Hall, Upper Saddle River.
- Zhao, D., Huang, L., Li, J., Qi, J., 2007. A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy. *ISPRS J. Photogramm. Remote Sens.* 62, 25–33. <https://doi.org/10.1016/j.isprsjprs.2007.01.003>.
- Zheng, G., Moskal, L.M., 2009. Retrieving Leaf Area Index (LAI) using remote sensing: theories, methods and sensors. *Sensors* 9, 2719–2745. <https://doi.org/10.3390/s90402719>.