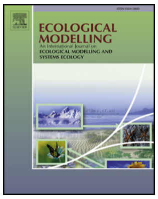
Ecological Modelling 380 (2018) 22–30

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



Improving the validation of ecological niche models with remote sensing analysis

Leandro José-Silva[a], Reginaldo Carvalho dos Santos[a], Bruna Martins de Lima[a], Mendelson Lima[a], José Francisco de Oliveira-Júnior[b], Paulo Eduardo Teodoro[c,*], Pedro V. Eisenlohr[a], Carlos Antonio da Silva Junior[a]

[a] State University of Mato Grosso, Alta Floresta, Mato Grosso, Brazil
[b] Institute of Atmospheric Sciences, Federal University of Alagoas, Alagoas, Maceió, Brazil
[c] Federal University of Mato Grosso do Sul, Chapadão do Sul, Mato Grosso do Sul, Brazil

ARTICLE INFO

Keywords:
Bertholletia excelsa
Vegetation index
Amazon
GEOBIA
Operational land imager

ABSTRACT

Ecological Niche models (ENMs) are tools that allow us to approximate the area of suitability for a species, thereby allowing elaboration of conservation strategies. The validation of these models in situ is not always possible due to costly access remote areas where conserved species are often found. The goal of our study was to provide a new validation concept for ENMs by applying remote sensing (SR) techniques, such as Geographic Object-Based Image Analysis (GEOBIA), which enables mapping of large areas and provides detailed information on land use. To assess the GEOBIA validation technique, we selected the species *Bertholletia excelsa* (Brazil nut), a tree that has great importance as a non-timber forest product and is considered vulnerable by the International Union for Conservation of Nature (IUCN). Models were built on the 'biomod2' package, and evaluation was conducted using the area under the receiver operating characteristic curve (AUC) and True Skill Statistics (TSS) metrics. Images were obtained from the orbital Operational Land Imager (OLI) on board the Landsat-8 satellite and the thematic maps were evaluated using Kappa and Overall Accuracy Statistics. We calculated vegetation indices (EVI, SAVI, LAI, and NDVI) and applied them to the GEOBIA technique. A total of 693 possible sites of *B. excelsa* were detected. Of these, 25 accessible sites were used for validation, and 45 new records of *B. excelsa* were added in the study area. GEOBIA was demonstrated to have high potential for validating ENMs, as well as in the extraction of arboreal species from medium-resolution spatial images.

1. Introduction

A plethora of species will probably become extinct by the end of this century (Loreau et al., 2006; Tilman et al., 2017). Knowledge about species that still exist is important for the elaboration of conservation strategies (Willis et al., 2007; Aarts et al., 2012), and there are several approaches in the literature that can be applied to explain and predict the distribution of endangered species (Kerr, 1997; Myers et al., 2000; Pearson et al., 2002; Peterson, 2003; Baillie et al., 2004; Maciel et al., 2016). One of these approaches is Ecological Niche Modeling (ENM) (sometimes referred to as Species Distribution Modeling or Habitat Suitability Modeling), which is a tool that by can predict geographic species distribution based on environmental suitability (e.g., Pearson, 2007; Siqueira et al., 2009; Franklin, 2010). The ENM is based on the ecological niche theory and uses occurrence points along with the ecological requirements present at these points (generally environmental variables) as an approximation of the multidimensional niche (Elith et al., 2006). Subsequently, these ecological requirements are extrapolated to areas of potential distribution (Franklin, 2010).

A fundamental issue in ENM involves model validation, which is still not well-developed (Peterson et al., 2008; Jiménez Valverde et al., 2011). Every model must be validated; those which fail validation are simply discarded (Monte et al., 1996; Morrison et al., 1998; Ottaviani et al., 2004). There are several techniques for model validation, but a rather widespread idea in ENM is that the best validation is realized in situ (Siqueira et al., 2009). However, this type of validation is not always possible due to the cost of expeditions or even difficulty in accessing areas where the species of interest is expected to find suitable conditions (Hertzog et al., 2014). Thus, there is a need for highly accurate methods for selecting areas of easy access to obtain in situ occurrence records that can be used to validate models. Our study proposes a new approach for validating ENMs in which one can find new

---

[*] Corresponding author at: Rodovia MS 306, km. 305, Postal Box 112, 79560000, Chapadão do Sul, MS, Brazil.
E-mail address: paulo.teodoro@ufms.br (P.E. Teodoro).

https://doi.org/10.1016/j.ecolmodel.2018.04.013
Received 1 March 2018; Received in revised form 21 April 2018; Accepted 23 April 2018

occurrence records for modeled species within suitable areas as indicated by projection maps.

Such new avenues can be provided by remote sensing (RS) techniques, which are the lowest-cost solutions for mapping land use, particularly the distribution of forest, agricultural, and urban areas, due to their ability to provide detailed information on the soil coverage over extensive areas (Eberhardt et al., 2016; Bellón et al., 2017). In addition, they provide synoptic coverage and high temporal resolution capacity, which allows for monitoring the phenology of regions with vegetal cover (Bellón et al., 2017; Yang et al., 2017). Specifically, Geographic Object-Based Image Analysis (GEOBIA) is a robust method for mapping treetops and differentiating them from the surrounding landscape, thereby accommodating different types of data and levels of quality variables (O'Neil-Dunne et al., 2014; Silva Junior et al., 2018), which could be applied to facilitate and improve the validation of ENMs.

We sought to reduce the area projected by the models in order to facilitate discovery of target species occurrences in the field, with specific collection sites indicated based on the use of GEOBIA within the ENM scenario. Thus, we aimed to assess whether ENMs can be validated by means of GEOBIA. We expected a positive answer, because each species has a specific spectral signature, and species that stand out of the canopy are likely to be detected by multi-resolution GEOBIA. If this is true, it would be possible to improve the occurrence records of a tree species using GEOBIA. If GEOBIA is able to validate ENMs, then we can assume that GEOBIA is able to detect the extracted traits of the target species. We tested GEOBIA as an ecological niche model validation tool with the tree *Bertholletia excelsa* (Brazil nut, from the Lecythidaceae family), a tree protected by the Brazilian Government (Law Decree Nº 5.975/2006) that has great importance as a non-timber product (Peres et al., 2003) and is considered vulnerable by the International Union for Conservation of Nature (IUCN, 2017).

2. Material and methods

2.1. Study area

The study area was located in the state of Mato Grosso in the municipality of Alta Floresta, in the Midwest region of Brazil at latitude 09° 34′ to 56° 14′S and longitude 10° 21′ to 56° 13′W, encompassing an approximate area of 8,947,069 km^2 (Fig. 1). The altitude of the region varies substantially, with 52% of the territory above 279 m altitude, and only 3% below 300 m. The predominant climate was characterized according to Köppen-Geiger's classification as Cwi (rainy tropical) (Álvares et al., 2013). In addition, the land use and cover of the region were determined using a classified database (Silva Junior and Lima, 2018).

2.2. Occurrence records and environmental predictors

We obtained the occurrence records of *B. excelsa* from the SpeciesLink (http://www.splink.org.br) and GBIF (http://www.gbif.org) databases. We only included records that had an image or that were input by experts, and we eliminated inconsistent data, such as unreliable coordinates (which lacked georeferenced information) or repeated coordinates.

The bioclimatic variables obtained were solar radiation, wind speed, water vapor pressure, twelve variables related to temperature, and another nine related to precipitation, totaling 22 environmental layers extracted from the WorldClim database (Hijmans et al., 2005). Since this set of variables presents a high degree of collinearity, and also to avoid subjectivity in their selection, we retained the first seven axes determined by Principal Component Analysis (PCA), thus ensuring ± 95% of the data variation was captured. We used the 'rasterPCA' function of the *RStoolbox* package in R 3.4.2 (Leutner and Horning, 2016; R Core Development Team, 2017).

2.3. Model building

The ENMs were built, evaluated, and projected using the 'biomod2' package (Thuiller et al., 2016). The algorithms used were *Classification Tree Analysis* (CTA), *Flexible Discriminant Analysis* (FDA), *Multiple Adaptive Regression Splines* (MARS), *Artificial Neural Networks* (ANN), *Generalized Boosting Model* (GBM), *Random Forest* (RF), *Maximum Entropy* (MAXENT), *Generalized Additive Models* (GAM), *Generalized Linear Models* (GLM), and *Surface Range Envelope* (SRE). The models were calibrated in a random sample of the data, with 70% used for training and 30% for testing. Each algorithm was used for 10 runs, which allowed a robust estimate of the performance to be obtained for each model (Pearce and Ferrier, 2000; Franklin, 2009). We used 10 sets of points as pseudo-absences, all of them outside the bioclimatic envelope of the species, totaling 100 models (10 sets of pseudo-absence x 10 runs) for each algorithm.

From the continuous projection maps, we prepared binary maps (i.e., maps that indicate suitability or non-suitability areas for *B. excelsa*) based on the threshold that maximizes the sum of the sensitivity plus the specificity (Liu et al., 2013). This threshold equals finding the point on the Receiver Operating Characteristic (ROC) curve whose tangent slope is equal to 1 (Liu et al., 2005). Use of this threshold produces higher sensitivity in most cases and higher TSS in many cases than is achieved by the other methods (Liu et al., 2013). In preparation of the consensus models, we obtained the mean of the individual models, thus providing more reliable predictions with lower levels of uncertainty (Araújo and New, 2007).

2.4. Model evaluation

Two methods were used to evaluate model performance. The first was the area under the ROC curve (AUC; Elith et al., 2006). The AUC values were classified as Excellent at 0.90–1.00, Good at 0.80–0.90, Average at 0.70–0.80, Poor at 0.60–0.70, or Failure at < 0.60 (Swets, 1988; Araujo et al., 2005). Thus, AUC = 1 would be expected if the entire area of the model was under the curve or, in other words, in a model with 100% correct predictions. In turn, failure models would occupy a limited area ($< 60\%$) under the curve. The second method used the True Skill Statistics (TSS), which has values between -1 and + 1. Values less than 0.4 are indicative of low predictive power, 0.4–0.8 of good predictive power, and 0.8 to 1 of excellent predictive power; values equal to or less than zero indicate performance no better than explained by chance (Allouche et al., 2006; Zhang et al., 2015). The algorithm SRE failed ($TSS < 0.4$) and was thereby discarded.

2.5. Collection of orbital images

To map all areas of *B. excelsa*, images from the Operational Land Imager (OLI) orbital sensor on board the Landsat-8 satellite were used. Landsat-8 continues the Landsat program, while providing advances to sensor capability that include new bands in spectral regions that respond to aerosols and cirrus, two thermal bands, better signal-to-noise sensor performance, and better radiometric resolution (Table 1). Scenes 227/67 and 228/67 were obtained from the United States Geological Survey Earth Explorer (USGS) database for Julian days 167 and 174, respectively, in year 2017 (GLCF, 2004). The OLI monotemporal vegetation index rasters were processed into thematic maps of pasture areas, whose accuracy was evaluated using the Kappa (κ) and Overall Accuracy (OA) Statistics (details on κ are in Landis and Koch, 1977). To determine the κ and OA values, 170 randomly distributed sample points were used (Fig. 1) based on prior knowledge of the area, and their locations were determined by GPS (Garmin eTrex Venture Hc).

The free availability of the Landsat series data provides opportunities for the analysis of terrestrial change at multiple time scales (Silva Junior et al., 2014). By means of the radiometric calibration process in ENVI 5.1, all bands of both scenes were transformed from digital