VALIDATION OF ECOLOGICAL NICHE MODELS CAN BE IMPROVED BY GEO-OBJECT ORIENTED ANALYSIS

Carlos Antonio da Silva Junior [1], Mendelson Lima [1], José Francisco de Oliveira-Júnior [2], Paulo Eduardo Teodoro [3], Pedro V. Eisenlohr [1], Leandro José da Silva [1], Reginaldo Carvalho dos Santos [1] and Bruna Martins de Lima [1]

[1]State University of Mato Grosso, Cáceres-MT, carlosjr@unemat.br; [2]Federal University of Alagoas, Maceió-AL, jose.junior@icat.ufal.br; e [3]Federal University of Mato Grosso do Sul, Chapadão do Sul-MS, eduteodoro@hotmail.com

ABSTRACT

The goal of our study is providing a new validation concept into ENMs. This could be done by applying remote sensing (RS) techniques, which enables mapping of large areas and provide detailed information on land use. In order to validate the ENM models using the GEOBIA technique we used the species *Bertholletia excelsa*. The models were built on the ‘biomod2’ package. The images were obtained from orbital sensor Operational Land Imager (OLI) on board the Landsat-8 satellite. 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, using 25 sites accessible for validation, which added 45 new records of this species. GEOBIA has been shown to be a tool with a high potential to validate ENMs, as well as in the extraction of arboreal species from medium resolution spatial images.

Key words **—** *Bertholletia excelsa*, vegetation index, Amazonia.

RESUMO

O objetivo do nosso estudo é fornecer um novo conceito de validação para os ENMs. Isso pode ser feito aplicando técnicas de sensoriamento remoto (SR), que permitem o mapeamento de grandes áreas e fornecem informações detalhadas sobre o uso da terra. Para validar os modelos ENM usando a técnica GEOBIA, utilizamos a espécie Bertholletia excelsa. Os modelos foram construídos no pacote "biomod2". As imagens foram obtidas do sensor orbital Operational Land Imager (OLI) a bordo do satélite Landsat-8. Calculamos os índices de vegetação (EVI, SAVI, LAI e NDVI) e os aplicamos à técnica GEOBIA. Um total de 693 locais possíveis de B. excelsa foram detectados, usando 25 locais acessíveis para validação, que adicionaram 45 novos registros dessa espécie. O GEOBIA tem se mostrado uma ferramenta com alto potencial para validar ENMs, bem como na extração de espécies arbóreas a partir de imagens espaciais de média resolução.

Palavras-chave **—** *Bertholletia excelsa*, índice de vegetação, Amazônia.

1. INTRODUCTION

A plethora of species will probably be extinct even during this century [1]. Knowledge about the species that still exist is important for the elaboration of conservation strategies [2]. These approaches is the 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 their environmental suitability [3].

A fundamental issue in ENM involves the validation of the models, which are not yet totally solved [4]. Every model must be validated; those which fail are simply discarded [5]. Thus, there is a need for highly accurate methods that allow selecting areas of easy access to obtain occurrence records that could validate our models in loco. Our study provides new avenues on validation of ENMs by proposing a new approach in which one can find new occurrence records of modeled species within the suitable areas indicated by projection maps.

Such new avenues can be provided by remote sensing (RS) techniques, which are the least cost solutions that contribute to the mapping of land use and occupation of forest, agricultural and urban areas due to their ability to provide detailed information on the coverage of the soil in extensive areas [6]. In addition, they present a synoptic coverage and high temporal resolution capacity, which allows monitoring the phenology of regions with vegetal cover [6,7].

We expected a positive answer, because each species has a specific spectral signature, and species that stand out the canopy are likely detected by remote sensing (RS) in multi-resolution. If this is true, it will be possible to improve the occurrence records of a tree species based on RS, because, if RS has proved to be useful to validate ENMs, we could assume RS as a powerful tool to detect the traits extracted by the target species. We tested RS as an ecological niche model validation tool with nut tree (*Bertholletia excelsa*, Lecythidaceae), a tree protected by the Brazilian Government (Law Decree Nº 5.975/2006), which has great importance as a non-timber product (Peres et al. 2003) and is considered vulnerable by the International Union for Conservation of Nature [8].

ISBN: 978-85-17-00097-3

2. MATERIAL AND METHODS

The studied area was the state of Mato Grosso in the municipality of Alta Floresta, Midwest region of Brazil, located between the geographical coordinates of latitude 09º 34' to 56º 14'S and longitude 10º 21' to 56º 13'W, with an approximate area of 8,947,069 km2 (Figure 1). The altitude of the area varies where 52% of the territory is above 279 m while only 3% is below 300 m. The predominant type of climate were characterized according to Köppen-Geiger's classification: "Cwi" (rainy tropical). In addition, it was sought to know the use and occupation of the land by means of classified database of the region [9,10].

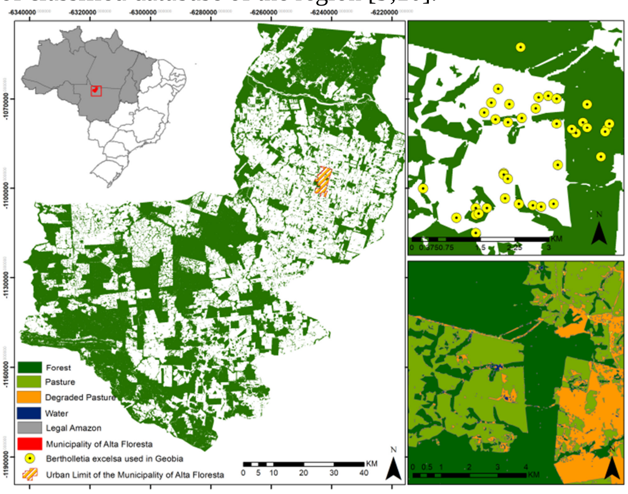

Figure 1. Record of Bertholletia excelsa used to train the algorithm in the municipality of Alta Floresta.

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

The ENMs were built, evaluated and projected using the 'biomod2' package [11]. 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 in 70% for training and 30% for testing with 10 runs in each algorithm, which allowed obtaining a robust estimate of the performance of each model. 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.

To map all areas of *Bertholletia excelsa* images of the Operational Land Imager (OLI) orbital sensor on board the Landsat-8 satellite were used. Scenes were obtained from the United States Geological Survey (USGS - Earth Explorer) database for the year 2017 of Julian days 167 and 174, respectively. With the images processed, thematic maps of pasture areas were generated, with OLI monotemporal image with vegetation index. The accuracy of thematic maps with the specialization of pasture areas was evaluated by Kappa (κ) and Overall Accuracy (OA) metrics. To determine Kappa values and Overall Accuracy, 170 randomly distributed sample points (Figure 1) were used, based on the knowledge of the area and points collected by the GPS device (Garmin eTrex Venture Hc).

Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI) and Leaf Area Index (LAI) images from the OLI sensor were used. They corresponded to path/row 227/67 and 228/68, collection 1 level 1 [12] from the Landsat-8 satellite.

The sequence phases necessary for classifying the Bertholletia excelsa areas are represented in Fig. 1. This kind of analysis uses GEOBIA (GEOgraphic-Object-Based Image Analysis), data mining and the integration of both when applied to OLI sensor. The multiresolution segmentation was performed using eCognition 8.0, in which the developed objects (polygons) were exposed to heterogeneity decision, which can be adjusted by selecting the scale parameter, the spectral band weights, the shape and compactness factors. The adjustment of a scale parameter may influence the size of the segments developed.

3. RESULTS

The average ENMs indicate climatic suitability for the Amazonian Domain, being in agreement with the known distribution area of the species, provided by Flora do Brasil 2020 under construction (2017) (Figure 2). The ENMs generated by the algorithms presented AUC mean values between 0.88 and 0.97. The TSS mean values presented results from 0.757 to 0.862. Thus, both metrics indicated good or excellent predictive power, with high accuracy. In addition, there was low standard deviations, which suggests a high degree of concordance among individual models (Table 1). Thus, the ENMs of *Bertholletia excelsa* present a better explanation than expected by chance.

Table 1. Mean values of AUC and TSS and standard deviation (SD) of 100 models built for each of the nine selected algorithms.

Algorithms	AUC ± SD	TSS ± SD
GLM	0.972 ± 0.005	0.862 ± 0.027
GAM	0.943 ± 0.029	0.838 ± 0.060
ANN	0.957 ± 0.025	0.840 ± 0.061
MARS	0.939 ± 0.045	0.829 ± 0.064
FDA	0.933 ± 0.024	0.778 ± 0.041
GBM	0.942 ± 0.064	0.852 ± 0.090
CTA	0.883 ± 0.042	0.757 ± 0.091
RF	0.949 ± 0.063	0.861 ± 0.120
MAXENT	0.942 ± 0.012	0.809 ± 0.042

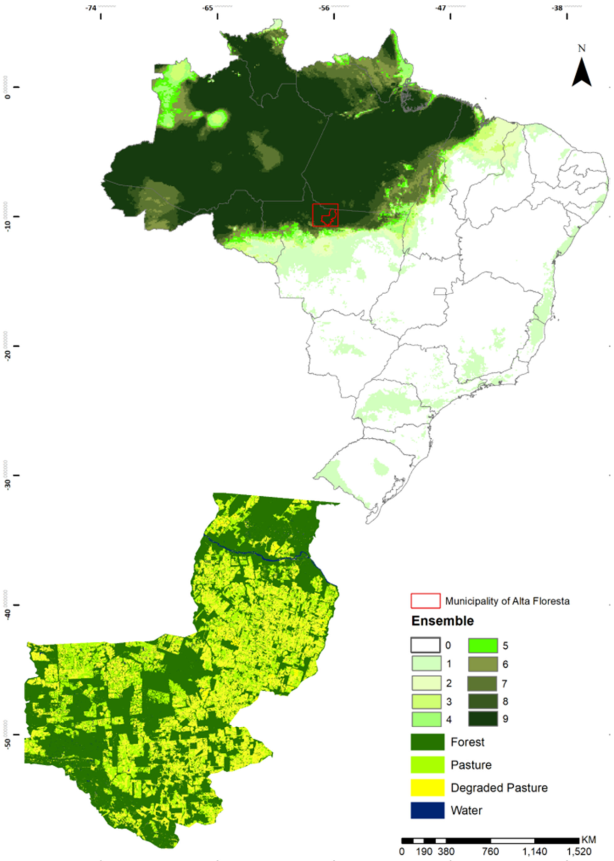

Figure 2. On the right, the average consensus of the distribution of *Bertholletia excelsa* to the algorithms. Dark color indicates suitable areas and white color indicates no existence of suitable areas of occurrence. On the left border, the municipality of Alta Floresta – MT.

Vegetation indices indicate that, in the study area, mean values for NDVI were high (0.873 ± 0.0587), showing sites with high plant biomass, such as the forests in the southeast (SE) and northwest (NW) of the municipality. The SAVI and EVI values were not elevated, maintaining the mean of 0.530 and 0.560 respectively, with smaller deviations from the mean (± 0.0335 and ± 0.0418), indicating that the sites present an intermediate forest canopy [12]. The lower values of SAVI and EVI also show that anthropic activities, such as pastures, are more pronounced in the central part of the municipality. This same result is evidenced by the LAI mean (4.396 ± 0.472), which indicates the heterogeneity of the sites differentiating forests from areas modified by anthropic activities [13].

Aiming at analyzing the quality of the classifier, the Kappa index value 0.71 and Overall Accuracy was 0.85, which indicates the robustness of the classifier [14]. From the 170 terrestrial truth points selected, 100 identified the occurrence of *B. excelsa* and 70 other uses, obtaining high levels of accuracy (Overall Accuracy 93.99% and Kappa index 0.87). This result showed the reliability of the classifier in the use of medium resolution images.

Within the municipality of Alta Floresta, MT, a region of high environmental suitability predicted by the ENMs (Figure 2), such tool allowed to detect 693 sites indicative of possible occurrence of *B. excelsa*. We simulate a situation in which it iswould be not possible to validate all 693 sites at the field; in such a situation, medium spatial resolution images would be used to select 25 accessible sites. Sampling of these 25 sites allowed the addition of 45 new occurrence records for *B. excelsa* species (Figure 3).

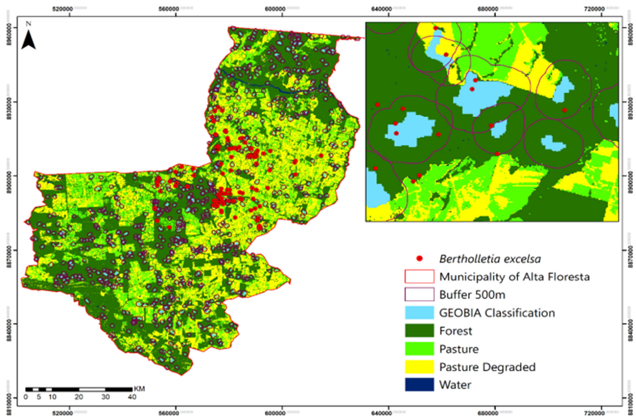

Figure 3. Core area of occurrence of *Bertholletia excelsa* and the validated sites in the municipality of Alta Floresta.

4. DISCUSSION

We were successful in providing a new framework to validate ENMs by using remote sensing. We found great potential to extract tree species in a multi-resolution OLI image (30 x 30m), a methodological proposal that is ideal for scholars and researchers who need a starting point in the validation of ENMs. We confirmed both hypothesis and demonstrated high potential of using GEOBIA in the context of these models. First, we found support of GEOBIA to validate the ENMs, because this image analysis tool was able to hit the target of *B. excelsa* predicted occurrence in the landscape. Second, GEOBIA has guided the obtainment of new occurrence records of *B. excelsa* in a high environmental suitability area predicted by projection maps.

Through the GEOBIA analysis, it was possible to map the landscapes covering the entire geographical study area. This tool makes it possible to map the soil use and occupation, classify and characterize forests, and its accuracy is close to 90%; however, this value depends on the resolution of the input image [15].

Scientific discussion on how to validate ecological models has been debated since 1960s [16]. The traditional methods that are used for validation of areas with potential climatic suitability [17] have several limitations. In general, areas of climate suitability are extensive, often even larger than entire biogeographical regions, which makes it difficult to choose a site to an in situ validation. This is because ENMs with such large suitable areas have not been able to precisely answer where is the most appropriate place to invest time and resource for model validation.

5. CONCLUSIONS

We thus emphasize the main novel modeling aspect brought by this study: the facility of validating ENMs by remote sensing techniques, avoiding financial and time expenses when searching for occurrence records in large areas of environmental suitability of the species.

6. REFERENCES

[1] Tilman, D.; Clark, M.; Williams, D.R.; Kimmel, K.; Polasky, S.; and Packer, C., "Future threats to biodiversity and pathways to their prevention", *Nature*, v.546, n.1, p.73-81, 2017.

[2] Aarts, G.; Fieberg, J.; and Matthiopoulos, J., "Comparative interpretation of count, presence–absence and point methods for species distribution models", *Methods in Ecology and Evolution*, v.3, n.1, p.177–187, 2012.

[3] Franklin, J., "*Mapping species distributions: spatial inference and prediction*", Cambridge University Press. 2010.

[4] Jiménez Valverde, A.; Barve, N.; Lira-Noriega, A.; Maher, S.P.; Nakazawa, Y.; Papeş, M.; Soberón, J.; Sukumaran, J.; and Peterson, A.T., "Dominant climate influences on North American bird distributions", *Global Ecology and Biogeography*, v.20, n.1, p.114-118, 2011.

[5] Ottaviani, D.; Lasinio, G.J.; and Boitani, L, "Two statistical methods to validate habitat suitability models using presence-only data", *Ecological Modelling*, v.179, n.4, p.417-443, 2004.

[6] Bellón, B.; Bégué, A.; Seen, D.L.; Almeida, C.P.; and Simões, M., "A Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series", *Remote Sensing*, v.6, n.6, p.600, 2017.

[7] Yang, C.; Wu, G.; Ding, K.; Shi, T.; Li, Q.; and Wang, J., "Improving Land Use/Land Cover Classification by Integrating Pixel Unmixing and Decision Tree Methods", *Remote Sensing*, v.9, n.12, p.1222-1234, 2017.

[8] IUCN - Americas Regional Workshop (Conservation & Sustainable Management of Trees, Costa Rica, November 1996). 1998. *Bertholletia excelsa*. The IUCN Red List of Threatened Species 1998: e.T32986A9741363. http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32986A9741363.en. Download on 09 October 2017.

[9] Silva Junior, C.A.; and Lima, M., "Soy Moratorium in Mato Grosso: Deforestation undermines the agreement", *Land Use Policy*, v.71, n.1, p.540-542, 2018.

[10] José-Silva, L.; Santos, R.C.; Lima, B.M.; Lima, M.; Oliveira-Júnior, J.F.; Teodoro, P.E.; Eisenlohr, P.V.; and Silva Junior, C.A., "Improving the validation of ecological niche models with remote sensing analysis", *Ecological Modelling*, v.380, n.1, p.22-30, 2018.

[11] Thuiller, W.; Georges, D.; Engler, R.; Breiner, F.; Georges, M.D.; and Thuiller, C.W., "Package 'biomod2'", 2016.

[12] Huete, A.R.; Liu, H.Q.; Batchily, K.; and Van Leeuwen, W., "A comparison of vegetation indices over a global set of TM images for EOS-MODIS", *Remote Sensing of Environment*, v.59, n.3, p.440-451, 1997.

[13] Aragão, L.E.O.; Shimabukuro, Y.E.; Santo, F.D.E.; and Williams, M., "Landscape pattern and spatial variability of leaf area index in Eastern Amazonia", *Forest Ecology and Management* v.211, n.3, p.240-256, 2005.

[14] Pax-Lenney, M.; and Woodcock, C.E., "The effect of spatial resolution on the ability to monitor the status of agricultural lands", *Remote Sensing Environment*, v.61, n.2, p.210- 220, 1997.

[15] O'Neil-Dunne, J.; MacFaden, S.; and Royar, A., "A versatile, production-oriented approach to high-resolution tree-canopy mapping in urban and suburban landscapes using GEOBIA and data fusion", *Remote Sensing*, v.6, n.12, p.12837-12865, 2014.

[16] Rykiel Jr, E.J., "Testing ecological models: the meaning of validation", *Ecological Modelling*, v.90, n.3, p.229-244, 1996.

[17] Guisan, A.; Thuiller, W.; and Zimmermann, N.E. "*Habitat Suitability and Distribution Models: With Applications in R*", Cambridge University Press. 2017.