

Combining multiple models to predict the geographical distribution of the Baru tree (*Dipteryx alata* Vogel) in the Brazilian Cerrado

Nabout, JC.^{a,b*}, Soares, TN.^b, Diniz-Filho, JAF.^c, De Marco Júnior, P.^{a,b},
Telles, MPC.^b, Naves, RV.^c and Chaves, LJ.^{a,c}

^aCiências Ambientais, Universidade Federal de Goiás – UFG, Pró-reitoria de Pesquisa e Pós-graduação – PRPPG, CP 131, CEP 74001-970, Goiânia, GO, Brazil

^bDepartamento de Biologia Geral, Instituto de Ciências Biológicas - ICB, Universidade Federal de Goiás – UFG, Campus Samambaia (Campus II), CP 131, CEP 74001-970, Goiânia, GO, Brazil

^cEscola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás – UFG, CEP 74001-970, Goiânia, GO, Brazil

*e-mail: naboutjc@hotmail.com

Received March 11, 2009 – Accepted May 14, 2009 – Distributed November 30, 2010

(With 4 figures)

Abstract

The Brazilian Cerrado is a biome of great biodiversity, but detailed information about the diversity and distribution of species in this region is still insufficient for both testing ecological hypotheses and for conservation purposes. Among native plants in the Cerrado, *Dipteryx alata* Vogel (commonly known as the “Baru” tree), has a high potential for exploitation. The aims of this paper were to predict the potential spatial distribution of *D. alata* in the Brazilian Cerrado utilising five different niche modelling techniques. These techniques usually provide distinct results, so it may be difficult to choose amongst them. To adjust for this uncertainty, we employ an ensemble forecasting approach to predict the spatial distribution of the Baru tree. We accumulated a total of 448 occurrence points and modelled the subsequent predicted occurrences using seven climatic variables. Five different presence-only ecological niche modelling techniques (GARP, Maxent, BIOCLIM, Mahalanobis Distance and Euclidean Distance) were used and the performance of these models was compared using Receiver Operating Characteristics (ROC) and the Area Under the Curve (AUC). All models presented AUC values higher than 0.68, and GARP presented the highest AUC value, whereas Euclidean Distance presented the lowest. The ensemble forecasting approach suggested a high suitability for the occurrence of the Baru tree in the Central-Western region of the Brazilian Cerrado. Our study demonstrated that modelling species distribution using ensemble forecasting can be an important computational tool for better establishing sampling strategies and for improving our biodiversity knowledge to better identify priority areas for conservation. For the Baru tree, we recommend priority actions for conservation in the central region of the Cerrado Biome.

Keywords: conservation, niche modelling, ensemble forecast, Brazilian Cerrado.

Combinando múltiplos modelos para prever a distribuição geográfica do Barueiro (*Dipteryx alata* Vogel) no Cerrado brasileiro

Resumo

O Cerrado brasileiro é considerado um bioma de grande biodiversidade, no entanto, informações detalhadas sobre a diversidade e distribuição das espécies ainda é insuficiente para testar hipóteses ecológicas e para propostas de conservação biológica. Dentre as plantas do cerrado com amplo potencial de exploração, destaca-se *Dipteryx alata* Vogel, (vulgarmente conhecida como Barueiro). O objetivo deste trabalho foi prever a distribuição geográfica potencial de *D. alata* no Cerrado brasileiro, utilizando cinco diferentes técnicas de modelagem de nicho. Essas técnicas usualmente apresentam distintos resultados, dificultando a escolha de um modelo. Para corrigir essa incerteza, usamos a estratégia de combinar os modelos gerados (*ensemble forecasting*) para prever a distribuição espacial potencial do Baru. Foi acumulado um total de 448 pontos de ocorrência para as espécies e modeladas, usando sete variáveis climáticas. Foram utilizadas cinco diferentes técnicas de modelagem de nicho usando-se dados de presença (GARP, Maxent, BIOCLIM, distância de Mahalanobis e Distância Euclidiana) e a performance dos modelos foi comparada usando-se a curva ROC (*Receiver Operating Characteristics*) e AUC (*Area Under the Curve*). Todos os modelos apresentaram valores de AUC superiores a 0,68, além disso, o GARP apresentou maior valor de AUC enquanto que a distância Euclidiana apresentou menor valor. A combinação dos cinco modelos sugere que as regiões de elevada adequabilidade de habitat para essa espécie encontram-se na região centro-oeste do bioma Cerrado. Por fim, o presente estudo demonstrou que a modelagem da distribuição geográfica da espécie, usando-se a “combinação de modelos” é uma importante ferramenta computacional para auxiliar nas estratégias de amostragem das espécies e ainda identificar áreas prioritárias para conservação. Para o Barueiro, são recomendáveis ações prioritárias para conservação na região central do Bioma cerrado (i.e. vale do Araguaia).

Palavras-chave: conservação, modelagem de nicho, conjunto de previsões, cerrado.

1. Introduction

Most tropical plants species are still poorly characterised with respect to their geographical distribution, a problem referred to as “Wallace shortfall” (Whittaker et al., 2005; Bini et al., 2006). Geographical distribution modelling has become an important tool in filling this gap and in furthering ecology, conservation and evolution studies. Specific applications include the study of the expansion of invasive species (Carroll et al., 2001; Peterson et al., 2007; Zhu et al., 2007; Reino et al., 2009), the effects of climatic changes on biodiversity (Heikkinen et al., 2006) and the establishment of conservation plans for threatened species (Solano and Feria, 2007; Marage et al., 2008).

Modelling the geographic distribution of a species is usually done by the spatial projection of functions describing the ecological niche of that species. These processes relate the known occurrences of a species and relevant environmental data to produce predicted geographic occurrences (Guisan and Zimmermann, 2000). There are several techniques that can be employed for niche modelling, such as Generalised Linear Models (McCullagh and Nelder, 1989), Generalised Additive Models (Hastie and Tibshirani, 1990), Principal Component Analysis (Robertson et al., 2003), Artificial Neural Networks (Mastrorillo et al., 1997), Maximum Entropy (Phillips et al., 2006) and others (see Elith et al., 2006; Reino et al., 2009). Many recent papers discuss these statistical techniques, how to evaluate the models they generate, and how to detect which model is more robust or accurate (Segurado and Araújo, 2004; Elith et al., 2006). However, Araújo and New, 2007, suggested that using ensemble forecasts – combining different models and parameters and using techniques to explore the resultant variety of potential distributions of a species – can be more informative than interpreting each method independently (see Marmion et al., 2009 for a recent statistical evaluation of ensembling techniques).

The Brazilian Cerrado is a region that has been highlighted as one of the 25 large biogeographical regions that requires priority conservation approaches, known as biodiversity “hotspots” (Myers et al., 2000). Nevertheless, information on the diversity and distribution of individual species within this biome is still insufficient for both understanding the ecological processes underlying diversity patterns and for establishing conservation policies. Among native plants in the Cerrado, *Dipteryx alata* Vogel (commonly known as the “Baru” tree) has a high potential for exploitation. This native fruit plant of the family Fabaceae is widely distributed but endemic to the Cerrado biome. It is found primarily in dense savanna formations and in dry forests (Ribeiro et al., 2004; Sano et al., 2004). This species has many potential human uses including food, medicinal applications and as wood for building (Almeida et al., 1998). Moreover, it is considered a key species for sustaining the natural fauna in the Cerrado as its fruit feeds birds, mammals (bats), rodents and monkeys, especially during the dry season when other food sources are scarce (Sano et al., 2004).

The goals of this paper were to predict the potential spatial distribution of *Dipteryx alata* in the Brazilian Cerrado

biome using five different presence-only ecological niche modelling techniques: Genetic Algorithm for Rule-set Prediction - GARP, Maximum Entropy - Maxent, Bioclimatic envelope - BIOCLIM, Mahalanobis Distance and Euclidean Distance, and to compare their performances using ROC curve verification. We employed ensemble forecasting to produce a map using all of the models simultaneously as the results of this method have been proposed to generate a better and more conservative overall solution (Araújo and New, 2007). We also evaluated the similarity of the maps generated by the five different techniques using cluster analysis and produced a unique prediction of species occurrence based on an ensemble map.

2. Material and Methods

2.1. Data

The geographical coordinates available for the locations of *Dipteryx alata* were obtained from the literature (e.g., Soares et al., 2008), field sampling and online databases at the Centro de Referência em Informação Ambiental (CRIA: <http://www.cria.org.br/>). We accumulated a total of 448 geographically unique occurrence points, representing the most complete database for the geographic distribution of this species (Figure 1). The Baru tree is endemic to the Cerrado biome (Ribeiro et al., 2004), which occupies approximately 2.5 million km² in the Central Region of Brazil. Moreover, this area includes a high heterogeneity of vegetation types, soil, geology, geomorphology and climate (Silva et al., 2006).

Some climatic variables are strongly correlated. To account for this collinearity, we obtained values for the environmental variables at the occurrence points of the Baru tree and performed a Principal Component Analysis (PCA) on this matrix (Legendre, P. and Legendre, L., 1998). In



Figure 1. Location of 448 occurrence data-points for *Dipteryx alata* used in the analyses.

the PCA, the first two components explained 72.23% of the total variability. A subset of climatic variables were then selected based on their loadings in these two principal components, allowing to define a set of six climatic variables for niche modelling that were not strongly correlated across the distribution of the Baru tree: mean annual temperature, temperature seasonality (coefficient of variation), mean temperature in the wettest quarter, mean temperature in the driest quarter, precipitation in the coldest quarter, precipitation in the warmest quarter, as well as one topographic variable (altitude) derived from the WORDCLIM database (<http://www.worldclim.org/>). All of the selected variables were converted to a grid resolution of 0.0417 degrees (nearly 4 km).

2.2. Species distribution modeling methods

We used five different techniques to model the potential geographic distribution of the Baru tree, based on presence-only data (i.e., only occurrences are known, absences are unknown). The specific models tested included Bioclimatic Envelope (BIOCLIM), Euclidean Distance, Mahalanobis Distance, GARP and Maxent. All of these techniques are frequently used to model species distributions, and especially Maxent and GARP have been used in many recent studies (see Peterson et al., 2007). All of the models were projected onto the contour map of the Brazilian Cerrado (Almeida et al., 1998). We analysed only the continuous regions of the Cerrado found in central Brazil and did not consider isolated patches that are found in the Northern and Northwestern parts of the Amazon. A brief description of each modelling technique and the related key references are provided below.

Bioclimatic Envelope (BIOCLIM) – This technique is based on the production of envelopes that contain presence points in a multidimensional environmental space. For this model we used DIVA-GIS software (Hijmans et al., 2002). The “Bioclim” option was used for identifying the suitability of areas for the Baru tree. Cells for which one or more climate variables were outside of the 0-100 percentile envelope obtained a code of “0”. The cells within the 5-95 percentile were coded as “3”; those external to this range but within the 2.5-97.5 percentile were coded as “2”; and the cells external to this range, but within the 0-100 percentile for all climate variables, were coded as “1”. The program then generated categories of species suitability, varying from “not suitable” to “most favourable”.

Euclidean Distance – This technique is structured on the existence of an optimum ecological point, defined by the generation of a centroid for all occurrence points in the total ecological space. The distance between this “optimum” and the observed values for each cell in the environmental grid within the geographical study area is inverse to the suitability of the environment at that site. Smaller distances indicate more similarity among regions, and a higher probability of a species being present. The Euclidean Distance produces a circular envelope around the optimum within the ecological space.

Mahalanobis Distance (Typicality probabilities) – This method is similar to Euclidean Distance, but the Mahalanobis distance produces an ellipsoidal envelope around the optimum ecological space by taking into account the covariances among environmental variables. The distance of a given locality to the centroid in the environmental space is given by (Equation 1):

$$D^2 = (x - m)^T C^{-1} (x - m) \quad (1)$$

where D^2 is the Mahalanobis distance, x is the vector of environmental measures at a location, m is the vector of the mean environmental measures for all known instances of the Baru tree, and C^{-1} is the inverse of the variance/covariance matrix among environmental measures for all known occurrences of the species.

Maxent (Maximum Entropy Method) – This recent technique provides better results than other programs, especially for small datasets (Pearson et al., 2006; Elith et al., 2006). The Maxent technique produces predictions and inferences from incomplete information (see Phillips et al., 2006), and is based on the principle of estimating a “target” probability distribution by finding the probability distribution of maximum entropy (i.e., the most spread out or the closest to uniform distribution), subjected to a set of constraints that represent the incomplete information about the “target” distribution. The parameters used in the Maxent model in this paper were 1000 iterations and a convergence threshold value equal to 10^{-5} (the program default), as well as removing all duplicates.

GARP (Genetic Algorithm for Rule Set Production) – This is a widely used algorithm based on artificial intelligence that works by combining rule sets for the purpose of generating the most accurate prediction in the considered region (Stockwell and Noble, 1992). The rules represent a group of possible multivariate relationships between the species occurrence points and environmental variables. These include bioclimatic rules, atomic rules and logistic regression (Stockwell and Peters, 1999). We adopted the following optimization parameters: 200 runs, a 0.001 convergence limit and a maximum of 2000 iterations. To generate the best subset output we indicated 0% for the omission threshold and 10% of the commission threshold. We evaluated 20 models using this subset. These models were then overlaid producing a number of models from the best subset at a given place or region. In both GARP and Maxent, absence data is need for some of the analyses (GARP) or for model testing (Maxent). In these cases, both software packages simulate pseudo-absence data as required (see below).

2.3. Evaluation and comparison of models

Model evaluation focused on their predictive performance and included the determination of a minimum threshold of quantitative output for the potential presence of the species. The model sensitivity was defined as the proportion of true presences in relation to total presences predicted by the model. Model specificity was defined as the proportion of true absences in relation to absences predicted by the

model. In this manner, a Receiver Operating Characteristics (ROC) curve was obtained by plotting the sensitivity against $1 - \text{specificity}$ for the varying probability threshold values, generating a threshold independent model evaluation method (Manel et al., 2001). Moreover, the area under the ROC curve (AUC) has been used extensively in species distribution modelling, characterising the performance of the model, at all possible thresholds, based on a single value that could be used as an objective approach to comparing different models (Elith et al., 2006; Phillips et al., 2006). The AUC ranges from 0 to 1, where 1 indicates high performance while values < 0.5 indicate low performance (Luoto et al., 2005; Elith et al., 2006). Despite recent criticisms (e.g., Lobo et al., 2008), AUC can still be useful when comparing models from a single species in a similar geographical domain.

To use this analytical machinery without the existence of a dataset that includes true absences, Phillips et al. (2006) generated a sample of 10000 pseudo-absence points as a training sample and to estimate the ROC curve and the AUC values for the Maxent procedure. We repeated the same process with the other model predictions to allow for a proper comparison of these methods.

For the analysis of similarity among the models, quantitative measures from each model output were used. These usually included continuous measures of the species occurrence, e.g., suitability in Maxent and Bioclim, distances in Mahalanobis and Euclidian Distance, and the number of models in GARP. These quantitative measures were truncated (corresponding to the point on the ROC curve that has the shortest distance to the top-left corner) and the minimum threshold was used to generate presence-absence data (Cantor et al., 1999). Values smaller than the threshold were considered zero, whereas values higher than threshold were considered one. A map of the number of models predicting the occurrence of the Baru tree for each location can then be obtained, providing a better idea of the regions for which predictions are independent of the model characteristics (an unweighted average – see Marmion et al., 2009). Finally, we used Simple Matching coefficients to produce a pairwise similarity matrix between models. This similarity matrix was clustered using an UPGMA algorithm (Legendre, P. and Legendre, L., 1998), allowing a visual evaluation of the similarity among the five different models.

3. Results

The potential geographical distribution of the Baru tree generated by all of the models suggested that the incidence of this species was higher in the Central-Western region of the Brazilian Cerrado (Figure 2). Among the predictions generated, the GARP model showed the highest occupational area, with approximately 63.99% of the Brazilian Cerrado biome occupied by this species. On the other hand, the model obtained using the Maxent method presented the smallest occupational area, of only 23.15% of the biome (Table 1).

All of the models tested presented AUC values higher than 0.68 (Table 1), with the highest value obtained for the GARP model and the lowest for Euclidean Distance model. Despite the great variation in the AUC values among all models, the analysis of similarity suggested the formation of two different groups of predictions: i) those formed only by GARP and ii) those formed by Mahalanobis Distance, Bioclim, Euclidean Distance and Maxent (within which Maxent and Bioclim were the most similar) (Figure 3).

The analysis of the sum of these models (the unweighted ensemble forecast, Figure 4) produced a wide-ranging model in which 72.25% of the Cerrado is occupied by the Baru tree. Furthermore, the ensemble distribution indicated a higher model overlap (i.e., more accumulated evidence for the occurrence of the species, or more environmental suitability) for the Baru tree in Goiás State in the Central-Western region of the Cerrado. On the other hand, in extreme Northern region of the Brazilian Cerrado, no model suggested the occurrence of the Baru tree.

4. Discussion

4.1. Ensemble forecasting and distribution modeling

Tools for modelling ecological niches have been widely used to predict the geographic distribution of species (Filipe et al., 2004; Pearson et al., 2007) and can be used to define areas where environmental conditions are appropriate for protection of a species whose habitat is threatened (Araújo et al., 2004; Chefaoui et al., 2005; Martinez et al., 2006). Knowledge of the potential distribution areas of the Baru tree is of extreme importance, even more so when the highly diverse nature of the Cerrado biome is considered (see Myers et al., 2000).

The efficiency of these various niche models has been frequently tested using the AUC, (see Peterson et al., 2007; Lobo et al., 2008). In this work, the GARP and Maxent models provided better results than the other models, according to their AUC values. Despite the high frequency of papers using GARP, recent comparative work has suggested that the Maxent models provide better results (Phillips et al., 2004; Elith et al., 2006; Phillips et al., 2006) and that they are more computationally efficient. However, Nabout et al., (2009) recorded higher AUC values for GARP when compared to Maxent when modelling the potential distributions of American fiddler crabs. Moreover, other papers support the importance of employing different modelling techniques. For Farber and Kadmon (2003), the Mahalanobis Distance method produced more accurate species distribution models than the Bioclimatic Envelope technique. More recently, Tsoar et al. (2007) modelled the distribution of 42 species of land snails, nesting birds, and insectivorous bats using six different modelling techniques and reported that GARP and Mahalanobis Distance obtained the highest accuracy. Johnson and Gillingham (2005) observed different spatial distribution models and showed that models produced by Mahalanobis Distance and GARP techniques were not similar, demonstrating the

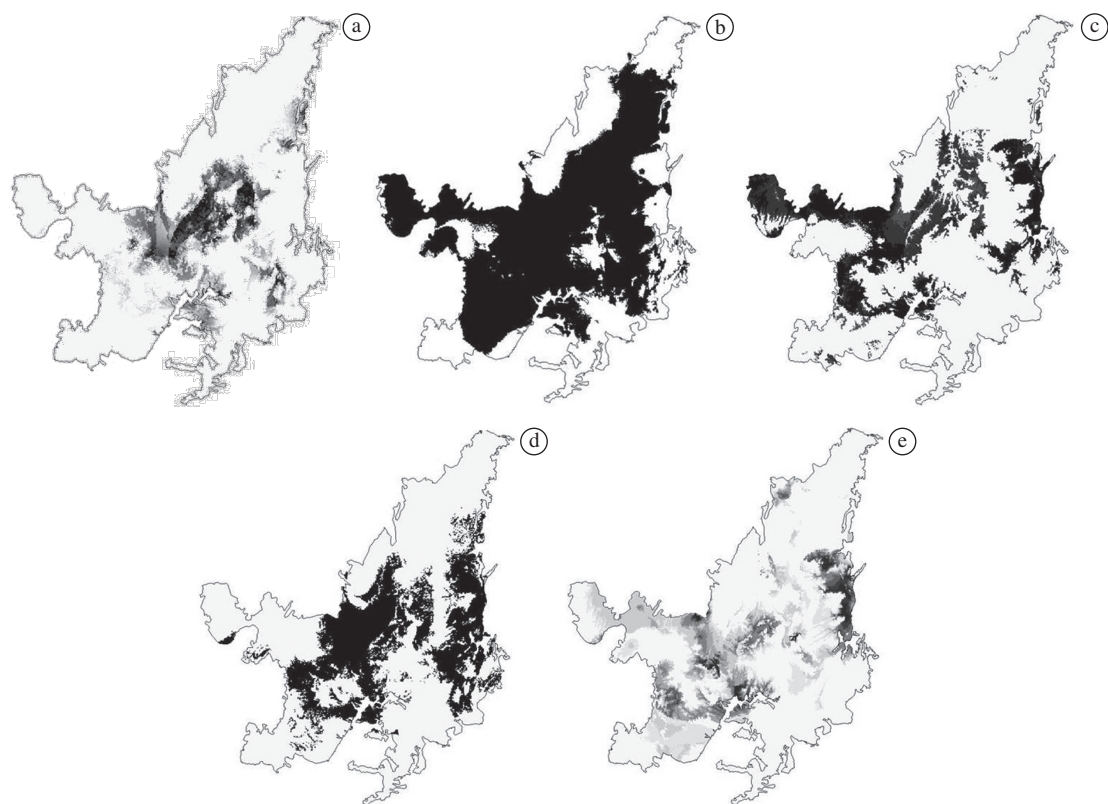


Figure 2. Spatial distribution modelling of Baru tree in Brazilian savanna (“Cerrado”): a) Maxent model; b) GARP Model; c) Mahalanobis distance; d) Euclidean distance and e) Bioclimatic envelope. Increasing darkness of shading indicates higher environmental suitability.

Table 1. Evaluation of five models based on AUC values, and predicted area of each model (in %) in relation to Brazilian Cerrado.

	AUC	Area occupied of Cerrado (%)
Maxent	0.82	23.15
GARP	0.92	63.99
Mahalanobis distance	0.77	30.84
Euclidean distance	0.68	29.79
Bioclim	0.73	36.00

influence of model choice on the outcome and development of conservation areas. Sérgio et al. (2007) demonstrated that Maxent was better than GARP and ENFA (Ecological Niche Factor Analysis) for four species of Bryophytes. In summary, this short (and certainly not exhaustive) list demonstrates a great idiosyncrasy, where the best model varies according to the species and the region studied. This methodological uncertainty justifies the development of ensemble forecasting techniques for species distribution modelling (Araújo and New, 2007; see Marmion et al., 2009 for a recent evaluation).

In our analyses, the ensemble forecast map showed smaller uncertainty in the predicted distribution of the Baru tree in the central-western region of the Brazilian

Cerrado and not in the peripheries. Many ecological and physical conditions can influence the distribution, abundance and occurrence probability of trees. In many cases, higher abundance and occurrence probability occurs at the centre of a species’ geographical range and declines toward the peripheries (called the “abundant centre” distribution) (Brown, 1984; Gaston and Blackburn, 2000). The bioclimatic conditions of this biome edge do not favour the incidence of this species (see also Diniz-Filho et al., 2009). The Brazilian Cerrado is also surrounded by other biomes with different environmental characteristics, such as the Pantanal, the Amazonian basin and the Caatinga. Moreover, many studies have used the “abundant centre” distribution as the basis for hypotheses about ecological and evolutionary processes (see revisions from Sagarin and Gaines, 2002), mainly because central and peripheral populations respond in different ways to ecological, genetic and evolutionary processes.

The “ecological niche” is a fundamental principle used in producing models that are designed to predict the occurrence of a species (Elith et al., 2006; Stockwell, 2006). These models reflect the potential niche; however, the absence of data relating to dispersal limitation, species interactions (i.e., interspecific competition) or historical factors (Araújo and Guisan, 2006) can generate overestimated geographical distributions. Based on this, Soberón (2007)

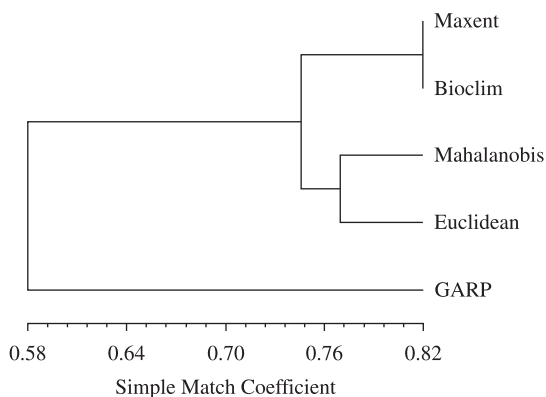


Figure 3. UPGMA dendrogram demonstrated the relationship among models of potential spatial distribution of Baru tree in Brazilian savanna.

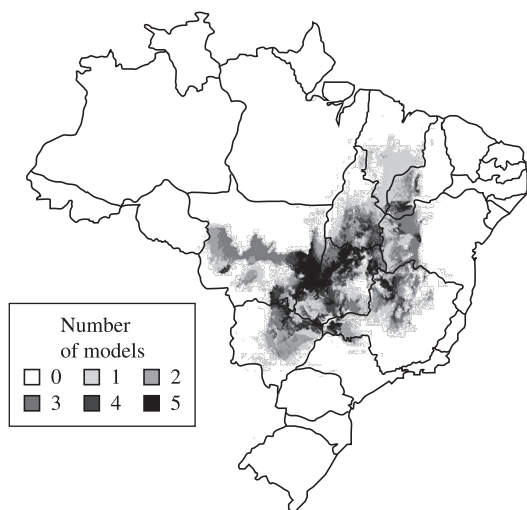


Figure 4. Combination of the five models (ensemble forecasting of maps from Figure 2) of Baru tree in Brazilian savanna.

suggested separating niches into two principal classes: i) the Grinnellian niche, which considers the environmental conditions at which a species can live, defined fundamentally by non-interactive variables (*scenopoetic*), and ii) the Eltonian niche, obtained by considering the biotic interactions and resource-consumer dynamics. The distribution models adopted in this study are based on the Grinnellian niche concept and can produce wider geographic ranges than are observed in nature. The Eltonian niche, although clearly important, is much more difficult to measure and model. Nevertheless, it is important to highlight that including biological interactions into species distribution modelling is an important research frontier (Kearney, 2006; Morin et al., 2007; Araújo and Luoto, 2007).

4.2. Implications for ecology and conservation

Populations of the Baru tree can form large clusters in any continuous area within the Brazilian Cerrado. In a

survey of woody species within this biome, the Baru tree was found in 86 of 316 sampled regions (Ribeiro et al., 2004), demonstrating the wide incidence of this species. This range of incidence can be associated with conditions of higher soil fertility (Almeida et al., 1998). The Baru tree is hermaphroditic and is pollinated by insects (Thum and Costa, 1999). The dispersal of its fruits can occur by gravity (short distance dispersal) and by animals, mainly by bats that are long distance dispersers (Ribeiro et al., 2004; Sano et al., 2004).

Our ensemble forecasting model produced more correspondence between the expected distribution of this species and the observed sampled distribution than when each method was viewed separately, in which case wider distributions were generated. Our models can be utilised to suggest sampling areas for further study of the botanical characteristics and the genetics of the Baru tree. More importantly, they can be used to indicate conservation regions – an urgent issue considering the current rapid increase of human land use in the Brazilian Cerrado.

The Brazilian Cerrado has been under pressure from agriculture (mainly soybean) and pastureland since the 1960's (Ratter et al., 1997), and recently by the advance of sugarcane cultivation (Sawyer, 2008). In general, the Central-Western and Central-Southern regions of this biome have been encroached by agricultural advancement and pastureland development (Rangel et al., 2007). These regions have also been transformed into sugarcane culture for the production of biofuels, which have had significant impacts on the ecology of the Cerrado. It is important to note that, according to our modelling, these regions present the highest suitability for the Baru tree. Several environmental impacts in these regions due to habitat fragmentation and losses are expected to affect these populations by disturbing dispersal and gene flow, influencing population density and genetic diversity and occasionally causing local extinctions. In addition to the local and regional effects of agricultural activities, global change is the other leading factor that can impact the diversity and distribution of this species (Sawyer, 2008; Tundisi and Matsumura-Tundisi, 2008). In light of this, it is very important to better design the conservation strategies for the Baru tree and other endemic plants within the Cerrado, especially at broad geographical scales.

The “Workshop on Priority Areas for Conservation in the Cerrado and Pantanal Region” identified 87 areas for conservation (see Brasil, 1999; Cavalcanti and Joly, 2002). The ensemble model for the Baru tree predicted its occurrence in many of these areas (mainly within the Araguaia valley and the Rio das Mortes wetland). However, in practice, the Federal and State reserve systems in the Cerrado region are insufficient both in extent and representation - Only 1.6% of this area is under protection (Cavalcanti and Joly, 2002). Despite the small number of protected regions (reserves), the Baru tree has been maintained locally by small landowners and communities (located mainly in the interior of the Goiás State), which derive their livelihood from commercialised almond production. Adopting better

conservation plans for this biome will be important to the preservation of ecological and evolutionary processes that are maintaining Baru tree populations.

Finally, our study demonstrated that ecological niche modelling can be an important computational tool, and that using an ensemble forecasting approach produced a more robust and more theoretically consistent model. Ensemble forecasting, beyond producing a more coherent model when compared with the observed species occurrences, revealed that the Baru tree presumably exhibits higher environmental suitability in Central-Western region of the Cerrado. Since this is the region with the most intense human occupation, we recommend priority action for successful Baru tree conservation, especially in areas such as the Araguaia valley. Moreover, we encourage others scientists to use these models for defining sampling strategies for different types of ecological, botanical and genetic studies. We also recommend that a similar strategy of producing ensemble forecasting models be adopted for other plants within the Cerrado Biome.

Acknowledgements — We are grateful to the Secretaria de Ciência e Tecnologia do Estado de Goiás (SECTEC-GO) for financial support for the Project, as part of the PRONEX program for establishing conservation priorities in the Brazilian Cerrado. JCN and TNS received a doctoral scholarship from CAPES and CNPq, respectively. JAFDF, PDM, MPCT and LJC are supported by productivity grants from CNPq.

References

- ALMEIDA, SP., PROENÇA, CEB., SANO, SM. and RIBEIRO, JF., 1998. *Cerrado: espécies vegetais úteis*. Planaltina: Embrapa-CPAC. 464 p.
- ARAÚJO, MB. and GUIBAN, A., 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, vol. 33, no. 10, p. 1677-1688.
- ARAÚJO, MB. and LUOTO, M., 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, vol. 16, no. 6, p. 743-753.
- ARAÚJO, MB. and NEW, M., 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, vol. 22, no. 1, p. 42-47.
- ARAÚJO, MB., CABEZA, M., THUILLER, W., HANNAH, L. and WILLIAMS, PH., 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, vol. 10, no. 9, p. 1618-1626.
- BINI, LM., DINIZ-FILHO, JAF., RANGEL, TFLVB., BASTOS, RP. and PINTO, MP., 2006. Challenging Wallacean and Linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distribution*, vol. 12, no. 5, p. 475-482.
- BRASIL., 1999. *Ações prioritárias para a conservação da biodiversidade no Cerrado e Pantanal*. Brasília, DF: Ministério do Meio Ambiente. FUNATURA, Conservation international, Fundação Biodiversitas, Universidade de Brasília.
- BROWN, JH., 1984. On the relationship between abundance and distribution of species. *The American Naturalist*, vol. 124, no. 2, p. 255-279.
- CANTOR, SB., SUN, CC., TORTOLERO-LUNA, G., RICHARDS-KORTUM, R. and FOLLEN, M., 1999. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology*, vol. 52, no. 9, p. 885-892.
- CARROLL, C., NOSS, RF. and PAQUET, PC., 2001. Carnivores as focal species for conservation planning in the Rocky Mountain region. *Ecological Applications*, vol. 11, no. 4, p. 961-980.
- CAVALCANTI, R. and JOLY, C. 2002. The conservation of the cerrados. In OLIVEIRA, PS. and MARQUIS, RJ. (Ed.). *The cerrado of Brazil*. New York: Columbia University Press. p. 351-367.
- CHEFAOUI, RM., HORTAL, J. and LOBO, JM., 2005. Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. *Biological Conservation*, vol. 122, no. 2, p. 327-338.
- DINIZ-FILHO, JAF., NABOUT, JC., BINI, LM., SOARES, TN., TELLES, MPC., DE MARCO, P. and COLLEVATTI, RG., 2009. Niche modelling and landscape genetics of Caryocar brasiliensis (Pequi tree: Caryocaraceae) in Brazilian Cerrado: an integrative approach for evaluating central-peripheral population patterns. *Tree Genetics & Genomes*, vol. 5, no. 4, p. 617-627. DOI: 10.1007/s11295-009-0214-0.
- ELITH, J., GRAHAM, CH., ANDERSON, RP., DUDIK, M., FERRIER, S., GUIBAN, A., HIJMANS, RJ., HUETTSMANN, F., LEATHWICK, JR., LEHMANN, A., LI, J., LOHMANN, LG., LOISELLE, BA., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, JM., PETERSON, AT., PHILLIPS, SJ., RICHARDSON, K., SCACHETTI-PEREIRA, R., SCHAPIRE, RE., SOBERON, J., WILLIAMS, S., WISZ, MS. and ZIMMERMANN, NE., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, vol. 29, no. 2, p. 129-151.
- FARBER, O. and KADMON, R., 2003. Assessment of alternative approaches for bioclimatic modelling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, vol. 160, no. 1-2, p. 115-130.
- FILIFE, AF., MARQUES, TA., SEABRA, S., TIAGO, P., RIBEIRO, F., MOREIRA DA COSTA, L., COWX, IG. and COLLARES-PEREIRA, MJ., 2004. Selection of priority areas for fish conservation in Guadiana River Basin, Iberian Peninsula. *Conservation Biology*, vol. 18, no. 1, p. 189-200.
- GASTON, KJ. and BLACKBURN, TM., 2000. *Pattern and process in macroecology*. London: Blackwell, 377p.
- GUIBAN, A. and ZIMMERMANN, NE., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, vol. 135, no. 1-2, p. 147-186.
- HASTIE, T. and TIBSHIRANI, R., 1990. *Generalised Additive Models*. London: Chapman and Hall. 413p.
- HEIKKINEN, RK., LUOTO, M., ARAÚJO, MB., VIRKKALA, R., THUILLER, W. and SYKES, MT., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, vol. 30, no. 6, p. 751-777.
- HIJMANS, RJ., GUARINO, L. and ROJAS, E., 2002. *DIVA-GIS. A geographic information system for the analysis of biodiversity data*. Lima, Peru: International Potato Center. [Manual, version 2]

- JOHNSON, CJ. and GILLINGHAM, MP., 2005. An evaluation of mapped species distribution models used for conservation planning. *Environmental Conservation*, vol. 32, no. 2, p. 1-12.
- KEARNEY, M., 2006. Habitat, environment and niche: What are we modelling? *Oikos*, vol. 115, no. 1, p. 186-191.
- LEGENDRE, P. and LEGENDRE, L., 1998. *Numerical ecology*. Amsterdam: Elsevier. 853p.
- LOBO, JM., JIMÉNEZ-VALVERDE, A. and REAL, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, vol. 17, no. 2, p. 145-151.
- LUOTO, M., POYRY, J., HEIKKINEN, RK. and SAARINEN, K., 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography*, vol. 14, no. 6, p. 575-584.
- MANEL, S., WILLIAMS, HC. and ORMEROD, SJ., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, vol. 38, no. 5, p. 921-931.
- MARAGE, D., GARRAUD, L. and RAMEAU, JC., 2008. The influence of management history on spatial prediction of *Eryngium spinalba*, an endangered endemic species. *Applied Vegetation Science*, vol. 11, no. 1, p. 139-148.
- MARMION, M., PARVIANIEN, M., LUOTO, M., HEIKKINEN, RK. and THULLER, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, vol. 15, no. 1, p. 59-69.
- MARTINEZ, I., CARRENO, F., ESCUDERO, A. and RUBIO, A., 2006. Are threatened lichen species well-protected in Spain? Effectiveness of a protected areas network. *Biological Conservation*, vol. 133, no. 4, p. 500-511.
- MASTRORILLO, S., LEK, S., DAUBA, F. and BELAUD, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology*, vol. 38, no. 2, p. 237-246.
- MCCULLAGH, P. and NELDER, JA., 1989. *Generalized Linear Models*. 2 ed. London: Chapman and Hall, 540p.
- MORIN, X., AUGSPURGER, C. and CHUINE, I., 2007. Process-based modelling of tree species' distributions. What limits temperate tree species' range boundaries? *Ecology*, vol. 88, no. 9, p. 2280-2291.
- MYERS, N., MITTERMEIER, RA., MITTERMEIER, CG., FONSECA, GAB. and KENTS, J., 2000. Biodiversity hotspots for conservation priorities. *Nature*, vol. 403, no. 6772, p. 853-858.
- NABOUT, JC., MARCO JÚNIOR, PD., BINI, LM. and DINIZ-FILHO, JAF., 2009. Distribuição geográfica potencial de espécies americanas do caranguejo "violinista" (gênero *Uca*) (crustacea, decapoda) com base em modelagem de nicho ecológico. *Iheringia, série zoologia*, vol. 99, no. 1, p. 92-98.
- PEARSON, RG., RAXWORTHY, CJ., NAKAMURA, M. and PETERSON, AT., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, vol. 34, no. 1, p. 102-117.
- PEARSON, RG., THULLER, W., ARAÚJO, MB., MARTINEZ-MEYER, E., BROTONS, L., MCCLEAN, C., MILES, L., SEGURADO, P., DAWSON, TP. and LEES, DC., 2006. Model based uncertainty in species range prediction. *Journal of Biogeography*, vol. 33, no. 10, p. 1704-1711.
- PETERSON, A., WILLIAMS, R. and CHEN, G., 2007. Modeled global invasive potential of Asian gypsy moths, *Lymantria dispar*. *Entomologia Experimentalis et Applicata*, vol. 125, no. 1, p. 39-44.
- PHILLIPS, SJ., ANDERSON, RP. and SCHAPIRE, RE., 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, vol. 190, no. 3-4, p. 231-259.
- PHILLIPS, SJ., DUDÍK, M. and SCHAPIRE, RE., 2004. A maximum entropy approach to species distribution modeling. In *Proceedings of 21 International Conference on Machine Learning*. New York: ACM Press. p. 655-662.
- RANGEL, TFLVB., BINI, LM., DINIZ-FILHO, JAF., PINTO, MP., CARVALHO, P. and BASTOS RP., 2007. Human development and biodiversity conservation in Brazilian Cerrado. *Applied Geography*, vol. 27, no. 1, p. 14-27.
- RATTER, JA., RIBEIRO, JF. and BRIDGEWATER, S., 1997. The Brazilian cerrado vegetation and threats to its biodiversity. *Annals of Botany*, vol. 80, no. 3, p. 223-230.
- REINO, L., MOYA-LARAÑO, J. and HEITOR, AC., 2009. Using survival regression to study patterns of expansion of invasive species: will the common waxbill expand with global warming? *Ecography*, vol. 32, no. 2, p. 237-246. DOI: 10.1111/j.1600-0587.2008.05354.x
- RIBEIRO, JF., SANO, SM. and BRITO, MA., 2004. *Baru (Dipteryx alata Vog.)*. Jaboticabal: Funep. p. 41.
- ROBERTSON, MP., PETER, CI., VILLET, MH. and RIPLEY, BS., 2003. Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, vol. 164, no. 2-3, p. 153-167.
- SAGARIN, RD. and GAINES, SD., 2002. The 'abundant centre' distributions: to what extent is it a biogeographical rule? *Ecology Letters*, vol. 5, no. 1, p. 137-147.
- SANO, SM., RIBEIRO, JF. and DE BRITO, MA., 2004. *Baru: biologia e uso*. Planaltina, DF: Embrapa cerrados, p. 52.
- SAWYER, D., 2008. Climate change, biofuels and eco-social impacts in the Brazilian Amazon and Cerrado. *Philosophical Transactions of the Royal Society B*, vol. 363, no. 1498, p. 1747-1752.
- SEGURADO, P. and ARAÚJO, MB., 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography*, vol. 31, no. 10, p. 1555-1568.
- SÉRGIO, C., FIGUEIRA, R., DRAPER, D., MENEZES, R. and SOUSA, AJ., 2007. Modelling bryophyte distribution based on ecological information for extent of occurrence assessment. *Biological Conservation*, vol. 135, no. 3, p. 341-351.
- SILVA, JF, FARIÑAS, MR., FELFILI, JM. and KLINK, CA., 2006. Spatial heterogeneity, land use and conservation in the Cerrado region of Brazil. *Journal of Biogeography*, vol. 33, no. 3, p. 536-548.
- SOARES, TN., CHAVES, LJ., TELLES, MPC., RESENDE, LV. and DINIZ-FILHO, JAF., 2008. Landscape conservation genetics of *Dipteryx alata* (baru tree: Fabaceae) from Cerrado region of central Brazil. *Genetica (The Hague)*, vol. 132, no. 1, p. 9-19.

- SOBERÓN, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, vol. 10, no. 12, p. 1115-1123.
- SOLANO, E. and FERIA, TP., 2007. Ecological niche modelling and geographic distribution of the genus *Polianthes* L. (Agavaceae) in Mexico: using niche modelling to improve assessments of risk status. *Biodiversity Conservation*, vol. 16, no. 16, p. 1885-1900.
- STOCKWELL, DRB., 2006. Improving ecological niche models by data mining large environmental datasets for surrogate models. *Ecological Modelling*, vol. 192, no. 1-2, p. 188-196.
- STOCKWELL, DRB. and NOBLE, IR., 1992. Induction of Sets of Rules from Animal Distribution Data - A Robust and Informative Method of Data-Analysis. *Mathematics and Computers in Simulation*, vol. 33, no. 1-6, p. 385-390.
- STOCKWELL, DRB. and PETERS, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, vol. 13, no. 2, p. 143-158.
- THUM, AB. and COSTA, EC., 1999. Entomofauna visitante das inflorescências de *Arecastrum romanzoffiana* (cham.) Glassm. (palmae). *Revista da Faculdade de Zootecnia, Veterinária e Agronomia de Uruguaiana*, vol. 5/6, no. 1, p. 43-47.
- TSOAR, A., ALLOUCHE, O., STEINITZ, O., ROTEM, D. and KADMON, R., 2007. A comparative evaluation of presence-only methods for modelling species distributions. *Diversity and Distribution*, vol. 13, no. 4, p. 397-405.
- TUNDISI, JD. and MATSUMURA-TUNDISI, T., 2008. Biodiversity in the neotropics: ecological, economic and social values. *Brazilian Journal of Biology*, vol. 68, no. 4, p. 913-915.
- WHITTAKER, RJ., ARAÚJO, MB., PAUL, J., LADLE, RJ., WATSON, JEM. and WILLIS, KJ., 2005. Conservation biogeography: assessment and prospect. *Diversity and Distribution*, vol. 11, no. 1, p. 3-23.
- ZHU, L., SUN, OJ., SANG, E., LI, Z. and MA, K., 2007. Predicting the spatial distribution of an invasive plant species (*Eupatorium adenophorum*) in China. *Landscape Ecology*, vol. 22, no. 8, p. 1143-1154.

