

# APLICAÇÃO DE DADOS DE OCORRÊNCIA DE INVENTÁRIO FLORESTAL NA MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIES FLORESTAIS MADEIREIRAS NO LESTE DO ACRE, BRASIL

Symone Maria de Melo Figueiredo<sup>(1)</sup> e Evandro Orfanó Figueiredo<sup>(2)</sup>

<sup>1</sup>Universidade Federal do Acre, Centro de Ciências Biológicas e da Natureza. E-mail: symone.figueiredo@ufac.br; <sup>2</sup>Empresa Brasileira de Pesquisa Agropecuária - Embrapa, Centro de Pesquisa Agroflorestal do Acre -CPAF – Acre . E-mail: evandro.figueiredo@embrapa.br.

## RESUMO

Em razão da pouca disponibilidade de dados de ocorrência de espécies florestais na Amazônia, o objetivo desse trabalho foi avaliar se a utilização de inventários florestais pode melhorar a estimativa de probabilidade de ocorrência e identificar os limites da distribuição potencial de espécies florestais de interesse madeireiro na região leste do estado do Acre. As variáveis ambientais preditoras foram: altitude, declividade, índice de vegetação por diferença normalizada (NDVI) e distância vertical à drenagem mais próxima, com aplicação do algoritmo de máxima entropia. Os modelos de distribuição de espécies obtiveram valores de AUC entre 0,76 a 0,92 usando dados de inventários florestais, superior ao desempenho dos modelos usando os dados da rede *speciesLink* (0,50 a 0,64). A altitude e o NDVI foram as variáveis mais importantes na modelagem para todas as espécies.

**Palavras-chave** — Maxent, Modelflora, Amazônia.

## ABSTRACT

*Due to the limited availability of data for the occurrence of forest species in the Amazon, the aim of this study was to evaluate the use of forest inventories can improve the estimation of probability of occurrence, identify the boundaries of the potential distribution and habitat preference of a group species of timber forest species of interest in the eastern region of the state of Acre. Environmental predictor variables were: altitude, slope, vegetation index (NDVI) and vertical distance to nearest drainage with application of maximum entropy algorithm. The species distribution models achieved AUC values between 0,76 to 0,92 using data from forest inventories, higher performance models using data from speciesLink network (0,50 to 0,64). The altitude and NDVI were the most important variables in the model for all species.*

**Key words** — Maxent, Modelflora, Amazon.

## 1. INTRODUÇÃO

Nos últimos anos, o uso e a aplicação dos modelos de distribuição de espécies tem aumentado resultando em implicações fundamentais para o estudo da biodiversidade e para a tomada de decisão em conservação ([1]; [2]) e auxiliando na compreensão dos requerimentos ecológicos das espécies [3].

Os modelos, em geral, relacionam a distribuição dos pontos de ocorrência da espécie em locais conhecidos com um conjunto multivariado de informações sobre o ambiente desses locais, por meio de ajustes de funções para prever em que locais no espaço geográfico são prováveis sua ocorrência ([4]; [5]).

Apesar das potencialidades do uso e aplicação da modelagem de distribuição de espécies para a conservação da biodiversidade, planejamento, desenvolvimento e implementação de políticas públicas reguladoras da exploração dos recursos naturais, existem problemas associados ao uso dessa técnica ([6]; [7]; [8]). A escassez e a baixa precisão espacial dos dados de ocorrência das espécies estão entre os principais entraves para o planejamento de conservação nos trópicos [6].

Assim, uma das principais questões nos estudos que envolvem a modelagem preditiva da distribuição de espécies é como obter informações sobre os registros de ocorrência destas espécies. No caso das espécies florestais, uma alternativa é a utilização dos dados provenientes dos levantamentos para a realização dos planos de manejo florestal, reconhecidos como uma fonte de dados para estudos de diversidade e distribuição espacial de espécies em sistemas florestais amazônicos que apresentam uma precisão espacial maior do que as informações provenientes de outras bases. Desta forma, o objetivo desse trabalho foi avaliar o efeito da utilização de dados provenientes de inventários florestais na estimativa de probabilidade de ocorrência de um grupo de espécies florestais de interesse madeireiro, em uma escala regional, na região leste do estado do Acre.

## 2. MATERIAIS E MÉTODOS

Os estudos foram desenvolvidos em 18 (dezoito) áreas de planos de manejo florestal (licenciados pelo órgão ambiental estadual) e a projeção dos mapas de distribuição potencial de

espécies compreendeu uma região delimitada pela Bacia hidrográfica do Rio Purus, no Acre, com extensão territorial de 80.004 km<sup>2</sup>.

Foram selecionadas nove espécies florestais de interesse madeireiro para a realização da modelagem preditiva de distribuição de espécie: *Amburana acreana* (Ducke) A.C.Sm.; *Ceiba pentandra* (L.) Gaertn.; *Apuleia leiocarpa* (Vogel) J.F.Macbr.; *Aspidosperma parvifolium* A.DC.; *Aspidosperma macrocarpon* Mart.; *Astronium lecointei* Ducke.; *Castilla ullei* Warb.; *Clarisia racemosa* Ruiz e Pav; *Hura crepitans* L. Foi realizada coleta de material botânico para herborização e confirmação do nome científico das espécies. As exsicatas foram depositadas no herbário UFACPZ do Parque Zoobotânico da Universidade Federal do Acre (UFAC).

Foram utilizados dois grupos de dados de ocorrência das espécies selecionadas cujos registros de coordenadas geográficas estão localizados na área de estudo. No primeiro grupo de ocorrências, os dados são oriundos de coleções (herbários e museus) disponíveis na rede *speciesLink* e são comumente utilizados nos estudos de modelagem. No segundo grupo, foram usados os dados de inventários florestais realizados com técnicas de precisão em áreas com Plano de Manejo Florestal Sustentável (PMFS). Os registros de ocorrência das espécies florestais de interesse madeireiro são provenientes de inventários florestais realizados mediante o uso das técnicas recomendadas pelo Modelo Digital de Exploração Florestal (Modeflora), que perfazem um total de 25.860 hectares de áreas manejadas.

O Modeflora é um conjunto de procedimentos voltados para o manejo de precisão em florestas tropicais desenvolvido pela Embrapa Acre e Embrapa Floresta [9]. Esse banco de dados foi disponibilizado pela Embrapa Acre e pelo Instituto de Meio Ambiente do Acre (IMAC).

As variáveis preditoras dos modelos foram: altitude, declividade, exposição do terreno, distância vertical à drenagem mais próxima (HAND) e índice de vegetação por diferença normalizada (NDVI). Foi realizada uma análise de correlação entre pares de variáveis a fim de eliminar aquelas com autocorrelação espacial, com objetivo de evitar redundância de informações e reduzir a chance de sobreajustamento dos modelos.

As variáveis topográficas (altitude, declividade, exposição ou orientação do terreno) foram geradas em ambiente SIG, a partir de dados do Shuttle Radar Topographic Mission (SRTM) com espaçamento (resolução) de 3 arco-segundos (~90 m). A distância vertical à drenagem mais próxima está relacionada indiretamente com a profundidade do lençol freático, que por sua vez, indica a disponibilidade da água do solo. Assim, o desnível ou a distância vertical de cada ponto em relação à drenagem mais próxima foi obtido com o algoritmo HAND (Height Above the Nearest Drainage) usando como dado de entrada o SRTM [10].

Para o cálculo do NDVI, foi utilizado o mosaico das imagens multiespectrais do satélite Landsat TM (Thematic

Mapper), bandas 3 (0,63 – 0,69 μm) e 4 (0,76 – 0,90 μm), correspondente a área de estudo, reamostrado para resolução de 90 m, compatível com as demais variáveis. As imagens foram obtidas do banco de imagens do Instituto Nacional de Pesquisas Espaciais (INPE), georreferenciadas com base na imagem geocover e são do período seco do ano de 2011 (antes da exploração das áreas manejadas) em que se verifica pouca cobertura de nuvens. Os valores do NDVI indicam a quantidade de vegetação verde presente no pixel [11].

Nos modelos foram incluídas apenas as variáveis ambientais com coeficiente de correlação de Pearson menor que 0,7 para evitar variáveis altamente correlacionadas. Para estimar quais são as variáveis mais importantes no modelo, foi usado o teste de Jackknife [12].

Na modelagem preditiva de distribuição das espécies foram utilizados dois algoritmos em função do número de registros de presença disponíveis, isso porque os dados de ocorrência das espécies florestais disponíveis na rede *speciesLink* são em número reduzido. Assim, nos casos do número de presença ser inferior a dez registros, optou-se por uma modelagem exploratória usando o algoritmo de distância ambiental euclidiana disponível no software *openModeller* [13]. Nos demais casos, ou seja, com dados oriundos das coleções de herbários com o mínimo de 10 registros de ocorrência e com os dados oriundos dos inventários florestais, foi utilizado o algoritmo de máxima entropia por meio do software Maxent [4].

A máxima entropia (Maxent) é um método para realizar previsões ou inferências a partir de informações incompletas, estima a probabilidade de adequação do habitat onde é possível a ocorrência da espécie, encontrando a distribuição de probabilidade de máxima entropia (mais próxima do uniforme) sujeita a um conjunto de restrições, no qual os valores esperados em cada variável ambiental devem corresponder à suas médias observadas nas amostras ([4]; [5]; [14]).

Na aplicação do algoritmo Maxent, para cada espécie foi realizada a modelagem com dez ensaios ou repetições (número de execuções do modelo) e, em seguida, calculada a média dos resultados de todos os modelos gerados por espécie. Para isso foi feita a reamostragem dos dados utilizando a técnica subsample [15], gerando aleatoriamente dez sub-amostras do conjunto de dados, sem reposição. Para a execução do modelo, em cada sub-amostra os dados foram divididos aleatoriamente, 70% para amostras de treino e 30% para amostras de teste externo (validação dos modelos). Nos modelos gerados apenas com os dados da rede *speciesLink*, as amostras de teste externo foram os dados de ocorrência coletada em campo por ocasião dos inventários florestais.

Para comparar a área estimada nos modelos com aquelas observadas no mesmo ponto pelas amostras de validação foi utilizado o gráfico do receptor-operador (ROC-plot) onde a área sob a curva (Area Under the Curve – AUC) é a medida de desempenho. Este método é bastante utilizado porque é uma medida global de desempenho independente de limites de corte [16]. A variação dos valores de AUC entre os

modelos obtidos com os dados provenientes dos inventários florestais e dos modelos obtidos usando os dados do *speciesLink* foi comparado utilizando o teste de Mann-Whitney.

A confecção do mapa da diferença entre os modelos foi resultado da subtração dos mapas binários individuais gerados a partir dos mapas de probabilidade convertidos para mapas com valores de 0 (ausência) ou 1 (presença). O parâmetro utilizado para a conversão foi definido com base no menor valor de probabilidade aplicado ao conjunto de dados de treinamento do algoritmo (*Minimum training presence*) denominado limiar ou limite de corte.

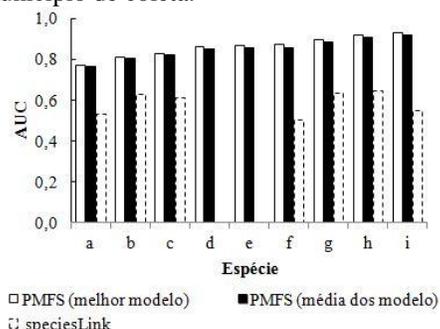
### 3. RESULTADOS

Em termos gerais, a quantidade de registros de ocorrência das espécies florestais oriundas das coleções disponíveis na rede *speciesLink* é bastante limitada quando comparada com os dados dos inventários florestais. Enquanto foram utilizados 36.065 registros de ocorrência dos inventários florestais, com a rede *speciesLink* foram apenas 74, para as nove espécies.

Em consequência, nem todas as espécies foram modeladas com o uso dos dados disponíveis na rede, devido ao número insuficiente de presenças ou erros de coordenadas.

Em outros casos, foi necessária a aplicação de um algoritmo mais simples (distância ambiental euclidiana), limitando o uso de modelos de distribuição de espécies a uma abordagem exploratória dos dados.

Os modelos gerados usando os dados de ocorrência oriundos das coleções de herbários disponíveis na rede *speciesLink* resultaram em AUC entre 0,50 a 0,64 ( $\pm 0,007$ ) usando o algoritmo de máxima entropia e distância euclidiana (Figura 1). O algoritmo de distância ambiental foi aplicado para *Astronium lecointei* e *Aspidosperma macrocarpon*, devido o número limitado de registros de ocorrência, incluindo dados cujas coordenadas geográficas correspondia à sede do município de coleta.



**Figura 1. Testes de desempenho dos modelos de distribuição de espécie de acordo com a origem dos dados de ocorrência, no leste do Acre. a: *Clarisia racemosa*; b: *Castilla ullei*; c: *Apuleia leiocarpa*; d: *Aspidosperma parvifolium*; e: *Ceiba pentandra*; f: *Astronium lecointei*; g: *Hura crepitans*; h: *Amburana acreana*; i: *Aspidosperma macrocarpon*.**

Usando os inventários florestais como fonte de dados de ocorrência, foram gerados 90 modelos usando o algoritmo de

máxima entropia (10 modelos por espécie). O valor de AUC variou de 0,76 a 0,92 ( $\pm 0,007$ ) e 0,81 a 0,93 ( $\pm 0,006$ ) para a média e o melhor modelo, respectivamente (Figura 1).

Os valores de AUC dos modelos gerados com os dados dos inventários florestais são significativamente superiores do que os modelos gerados com os dados do *speciesLink* (Teste Mann-Whitney, p-value=0,001).

A diferença no desempenho dos modelos de acordo com a origem do banco de dados influenciou na extensão da área predita da bacia.

### 4. DISCUSSÃO

Os modelos gerados com os dados do *speciesLink* foram pouco informativos e não mostraram um bom desempenho, devido à baixa disponibilidade de registros de ocorrência (Figura 1). Como consequência, a extensão da área predita foi menor em relação aos modelos gerados com os dados de inventário florestal, havendo mais discordância do que concordância entre eles, usando o mesmo algoritmo e as mesmas variáveis ambientais.

Todos os modelos gerados com dados de inventário florestal tiveram desempenho superior aos modelos produzidos com os dados de herbários, justificado pela diferença entre o número de registros de ocorrência e sua distribuição no espaço geográfico que refletem na predição dos modelos, pois os resultados da modelagem dependem da combinação entre as variáveis ambientais associadas com a presença observada da espécie (restrita nos dados de herbários) no subespaço do nicho realizado [17].

Para melhorar a predição dos modelos de distribuição das espécies usando registros de herbários, é necessária a incorporação de novos dados e que estes sejam mais bem distribuídos no tempo e no espaço [18].

Para trabalhos em escala local, os dados de ocorrência de espécies disponíveis das coleções de herbários é um fator limitante pela escassez no número de registros e pela qualidade associada às coordenadas geográficas desses pontos, sobretudo na região amazônica considerando que, em geral, os municípios tem grande dimensão territorial. Esses problemas são relatados por outros pesquisadores que sugerem um número mínimo entre 40 a 70 observações ([19]; [6]), abaixo do qual diminui o desempenho dos modelos. A utilização de outras fontes de dados de registros de presença, como os inventários florestais com técnicas de precisão, possibilita aumentar o número de observações da espécie de interesse e melhorar a predição dos modelos.

O uso da variável NDVI como preditora foi problemático no caso das espécies cujas coordenadas dos registros disponíveis na rede *speciesLink* estavam concentradas na sede municipal (baixos valores de NDVI), como consequência a distribuição potencial foi maior em áreas desmatadas, levando a uma interpretação errônea dos requisitos ambientais da espécie. Portanto, o uso do NDVI deve ser feito com cautela, principalmente quando os registros de ocorrência não tem precisão de coordenadas

geográficas e estão concentrados em áreas historicamente desmatadas.

Por outro lado, nos mapas de probabilidade existem regiões ao norte e a nordeste da bacia que não foram preditas como áreas de adequabilidade ambiental para as espécies, que coincidem com as áreas já desmatadas no leste do Acre. Isso aconteceu devido ao uso do NDVI como uma das variáveis preditoras dos modelos, que expressa o vigor da vegetação fotossinteticamente ativa [11], portanto, o valor de NDVI é menor em áreas desmatadas.

Além disso, a concentração dos dados de ocorrência utilizados em áreas florestadas de planos de manejo contribuiu para que as áreas desmatadas fossem preditas como não adequadas ambientalmente porque não há habitat remanescente (floresta) para a espécie. O uso dessa variável pode substituir a necessidade de aplicação de uma máscara para excluir áreas desmatadas da paisagem (background) conforme recomendado por [14].

## 5. CONCLUSÕES

A utilização de dados provenientes de inventários florestais de planos de manejo realizados com técnicas de precisão tem grande potencial em estudos de biodiversidade, conservação e gestão florestal, com o uso de técnicas de modelagem preditiva de distribuição de espécie.

A altitude e o NDVI foram as variáveis preditoras mais importantes na modelagem para todas as espécies, auxiliando na compreensão das condições ambientais adequadas para a possível ocorrência das espécies ao longo da bacia.

## 6. REFERÊNCIAS

- [1] MOTA-VARGAS, C.; ROJAS-SOTO, O.R. The importance of defining the geographic distribution of species for conservation: The case of the Bearded Wood-Partridge. *Journal for Nature Conservation*, 20: 10-17. 2012.
- [2] SHCHEGLOVITOVA, M.; ANDERSON, R.P. Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecological Modelling*, 269: 9-17. 2013.
- [3] MILLAR, C.S.; BLOUIN-DEMERS, G. Habitat suitability modelling for species at risk is sensitive to algorithm and scale: A case study of Blanding's turtle, *Emydoidea blandingii*, in Ontario, Canada. *Journal for Nature Conservation*, 20: 18-29. 2012.
- [4] PHILLIPS, A.T.; ANDERSON, R.P.; SCHAPIRE, R.E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190: 2331-2359. 2006.
- [5] PEARSON, A.T.; PAPES, M.; EATON, M. Transferability and model evaluation in ecological from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34(1): 102-117. 2007.
- [6] CAYUELA, L.; GOLICHER, D.J.; NEWTON, A.C.; KOLB, M.; ALBURQUERQUE, F.S.; ARETS, E.J. et al. Species distribution modeling in the tropics: problems, potentialities, and the

role of biological data for effective species conservation. *Tropical Conservation Science*, 2(3): 319-352. 2009.

- [7] ROCCHINI, D.; HORTAL, J.; LENGYEL, S.; LOBO, J.M.; JIMÉNEZ-VALVERDE, A.; RICOTTA, C.; BACARO, G.; CHIARUCCI, A. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, 35: 211-226. 2011.
- [8] KAMINO, L.H.; STEHMANN, J.R.; AMARAL, S.; DE MARCO, P.; RANGEL, T.F.; DE SIQUEIRA, M.F.; Hortal, J. Challenges and perspectives for species distribution modelling in the neotropics. *Biology letters*, 8(3): 324-326. 2012.
- [9] FIGUEIREDO, E.O.; BRAZ, E.M.; D'OLIVEIRA, M.V.N. Manejo de Precisão em Florestas Tropicais: Modelo Digital de Exploração Florestal. Rio Branco: Embrapa Acre, 183 p. 2007.
- [10] RENNÓ, C.D.; NOBRE, A.D.; CUARTAS, L.A.; SOARES, J.V.; HODNETT, M.G.; TOMASELLA, J.; WATERLOO, M.J. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazônia. *Remote Sensing of Environment*, 112: 3469-3481. 2008.
- [11] PRATES-CLARK, C.D.C.; SAATCHI, S. S.; AGOSTI, D. Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data. *Ecological Modelling*, 211(3): 309-323. 2008.
- [12] YANG, X.Q.; KUSHWAHA, S. P. S.; SARAN, S.; XU, J.; ROY, P.S. Maxent modeling for predicting the potential distribution of medicinal plant. *Ecological Engineering*, 51: 83-87. 2013.
- [13] MUÑOZ, M.E.S.; GIOVANNI, R.; SIQUEIRA, M.F.; SUTTON, T.; BREWER, P.; PEREIRA, R.S.; CANHOS, D.A.L.; CANHOS, V.P. openModeller: a generic approach to species potential distribution modeling. *Geoinformatica* 1: 111-135. 2011.
- [14] ELITH, J.; PHILLIPS, S.J.; HASTIE, T.; DUDIK, M.; CHEE, Y.E.; YATES, C.J. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17 (1): 43-57. 2011.
- [15] GIANNINI, T.C.; SIQUEIRA, M. F.; ACOSTA, A. L.; BARRETO, F. C.C.; SARAIVA, A.M.; ALVES-DOS-SANTOS, I. Desafios atuais da modelagem preditiva de distribuição de espécies. *Rodriguésia*, 63(3): 733-749. 2012.
- [16] FIELDING, A. H.; BELL, J.F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1): 38-49. 1997.
- [17] SOBERÓN, J.; PETERSON, A.T. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics*, 2: 1-10. 2005.
- [18] FEELEY, K. J.; SILMAN, M. R. Keep collecting: accurate species distribution modelling requires more collections than previously thought. *Diversity and Distributions*, 17(6): 1132-1140. 2011.
- [19] DRAKE, J.M.; RANDIN, C.; GUIBAN, A. Modeling ecological niches with support vector machines. *Journal of Applied Ecology*, 43: 424-432. 2006.