

COMUNICADO  
TÉCNICO

97

Boa Vista, RR  
Dezembro, 2022

**Embrapa**

# Efeito da Extensão da Área de Estudo sobre a Distribuição Potencial de Espécies: Um Estudo com Modelos Maxent para *Raoiella indica* Hirst (Acari: Tenuipalpidae) no Brasil

George Amaro  
Elisangela Gomes Fidelis  
Ricardo Siqueira da Silva  
Cesar Augusto Marchioro

# Efeito da Extensão da Área de Estudo sobre a Distribuição Potencial de Espécies: Um Estudo com Modelos Maxent para *Raoiella indica* Hirst (Acari: Tenuipalpidae) no Brasil<sup>1</sup>

<sup>1</sup> George Amaro, Pesquisador da Embrapa Roraima, MSc. Economia, Boa Vista, RR, Brasil. E-mail: george.amaro@embrapa.br. Elisangela Gomes Fidelis, Pesquisadora da Embrapa Recursos Genéticos e Biotecnologia, DSc. Entomologia, Brasília, DF, Brasil. E-mail: elisangela.fidelis@embrapa.br. Ricardo Siqueira da Silva, Professor e Pesquisador da Universidade Federal dos Vales do Jequitinhonha e Mucuri, DSc. Fitotecnia, Diamantina, MG, Brasil. E-mail: ricardo.siqueira@ufvjm.edu.br. Cesar Augusto Marchioro, Professor e Pesquisador da Universidade Federal de Santa Catarina, DSc. Zoologia, Curitiba, SC, Brasil. E-mail: cesmarc@gmail.com.

O interesse e necessidade em descrever, compreender e fazer previsões sobre a distribuição geográfica e ambiental de espécies existe há muito tempo (Wallace, 1860; Grinnell, 1917), de forma que diversos métodos foram desenvolvidos e vêm sendo utilizados em várias aplicações (Soberón et al., 2017). Em geral, são utilizados modelos que buscam quantificar as relações entre locais de ocorrência conhecidos de uma determinada espécie (apenas observações de presença ou de presença e ausência) e variáveis ambientais nesses locais (Elith, 2017), possibilitando que alguma inferência seja feita sobre a ocorrência dessa espécie em outras regiões onde condições ambientais similares possam ocorrer no espaço ou no tempo. Desde o final dos anos 1990, duas terminologias principais têm sido utilizadas para se referir aos métodos de modelagem que buscam estabelecer

associações ambientais à ocorrência de espécies e as relações dessas associações com suas distribuições geográficas: “modelos de distribuição de espécies” e “modelos de nicho ecológico” (Peterson; Soberón, 2012).

Atualmente, Maxent está entre os métodos mais utilizados para a modelagem de distribuição de espécies e tem demonstrado bom desempenho em comparação com outros métodos (Elith et al., 2006; Heikkinen et al., 2012; Venette 2017). Maxent é um algoritmo de aprendizado de máquina que foi desenvolvido especificamente para aproximar uma distribuição desconhecida para cenários onde apenas os locais de presença de uma espécie são conhecidos (Phillips et al., 2006; Elith et al., 2011; Merow et al., 2013). Basicamente, o algoritmo contrasta ambientes de pontos de presença com

aqueles amostrados como *background* para determinar quais combinações de variáveis melhor predizem ocorrências. O *background* inclui toda a área de estudo de onde são extraídas informações que caracterizam o ambiente acessível para a espécie ocorrer (Sillero; Barbosa, 2021). Os dados de fundo (referenciados neste texto como *background*, conforme definição de Phillips et al., 2009) não são tentativas de “adivinhar” os locais de ausência, mas sim uma forma de caracterizar os ambientes na região de estudo. Nesse sentido, o contexto ambiental é o mesmo, independentemente de onde a espécie foi encontrada. Os dados de *background* estabelecem o domínio ambiental do estudo, enquanto os dados de presença devem estabelecer sob quais condições uma espécie tem maior probabilidade de estar presente do que a média (Hijmans; Elith, 2021).

A extensão da área de estudo é uma questão importante na adequação do habitat e na modelagem da distribuição de espécies, uma vez que diz respeito à representação geográfica total ou parcial da área ocupada pela espécie de interesse, exatamente por ser onde os pontos de *background* são amostrados. Dependendo do objetivo do estudo, pode ser muito importante capturar toda a extensão da presença de uma espécie para verificar se seu nicho completo é representado. Contudo, em contextos em que existam registros de ocorrência em áreas invadidas, a inclusão desses registros permanece discutível, com no argumento comum de que essas presenças se aproximam melhor do nicho

fundamental de uma espécie e do potencial de invasão futura, enquanto os registros de presença da área nativa melhor refletem o nicho realizado (Elith, 2017). Estudos anteriores concluíram que modelos de distribuição para espécies invasoras se beneficiam da utilização de dados da área invadida (Mau-Crimmins et al., 2006; Broennimann; Guisan, 2008; Beaumont et al., 2009), uma vez que isso possibilita identificar melhor as condições em que a espécie pode se estabelecer e prosperar. No entanto, resultados contrários também foram encontrados (Vaclavik; Meentemeyer, 2012; Barbet-Massin et al., 2018).

Modelos baseados no Maxent utilizam apenas dados de pontos de presença e de *background*. Métodos que se baseiam em dados de presença-*background* simplesmente distinguem entre habitats adequados e menos adequados, considerando a extensão da área de estudo analisada, incluindo os pontos de presença e a delimitação da área de estudo, e não entre habitats ocupados e desocupados, como fazem os métodos de presença-ausência (Sillero, 2011).

A extensão geográfica ou ambiental sobre a qual esses pontos são amostrados afetam potencialmente o desempenho do modelo (VanDerWal et al., 2009; Barbet-Massin et al., 2018). Há argumentos sobre ser vantajoso restringir a amostragem de *background* com base no raciocínio ecológico, ou de uma forma que explique o viés de amostragem, em vez de simplesmente amostrar todo o ambiente de fundo, possivelmente em

uma área geográfica irrealisticamente grande (Phillips et al., 2009). Isso deve ser considerado com cautela ao se tratar de espécies invasoras e adicionar considerações ecológicas ou estatísticas prévias ao amostrar os pontos de *background* pode ser importante e ter um efeito significativo nos modelos derivados das observações de presença, uma vez que o uso de uma extensão grande, mas irrealista, apresentou efeito potencialmente adverso no modelo e nas previsões resultantes (VanDerWal et al., 2009; Elith et al., 2010). Melhores resultados podem ser obtidos ao restringir as amostras de *background* para áreas próximas às presenças (Phillips et al., 2009). Restringir a amostragem a áreas onde a espécie não foi observada para evitar observações conflitantes de presença e ausência irá provavelmente aumentar o viés inicial (Wisz; Guisan, 2009) e pode resultar em previsões exageradas dos locais ocupados por uma determinada espécie (Hanberry et al., 2012). Assim, esse trabalho tem como objetivo demonstrar o efeito da extensão da área de estudo sobre a modelagem de distribuição de espécies com o algoritmo Maxent, utilizando dados reais da espécie *Raoiella indica* Hirst (Acari: Tenuipalpidae).

Foi utilizado o ambiente R, versão 4.2.1 (R Core Team, 2022); o pacote *geobr* versão 1.6.5999 (Pereira; Goncalves, 2022), para dados espaciais do Brasil; o pacote *terra* versão 1.5-21 (Hijmans, 2022), para análises e transformação de dados espaciais; o pacote *flexsdm* 1.3.0 (Velazco et al., 2022), para todos os procedimentos de modelagem; o pacote

*pROC* versão 1.18.0 (Robin et al., 2011), para fazer o gráfico e estimativas relacionadas à curva ROC; e, o pacote *tmap* (Tennekes, 2018) para elaboração dos mapas de probabilidades e binário.

A partir das observações de ocorrência de *R. indica* utilizadas por Amaro et al. (2021)<sup>2</sup>, que compreendem pontos de presença em áreas nativas e invadidas de *R. indica*, seguiu-se o seguinte procedimento de modelagem, com todos os parâmetros dos modelos mantidos constantes, variando-se apenas a extensão da área de calibração:

- limpeza dos dados de ocorrência, utilizando-se o pacote *CoordinateCleaner* 2.0-20 (Zizka et al., 2019), com os testes “zeros”, “seas”, “equal”, “institutions”, “duplicates”, “centroids”, “gbif”, “validity”, “capitals”, resultando 203 observações, das 220 originais;
- redução do viés de amostragem, pela aplicação de um filtro ambiental (Varela et al., 2014) sobre as ocorrências, através da função *occfilt\_env*, com o uso de cinco classes, resultando 150 observações finais, utilizadas para a calibração do modelo;
- a área de calibração (estudo) foi definida pela função *calib\_area*, através de um *buffer* ao redor dos pontos de

<sup>2</sup> Incluindo dados de pesquisas de campo realizados em Roraima (equipe da pesquisadora da Embrapa, Dra. Elisangela Gomes Fidelis) e em São Paulo e Paraná (equipe do professor, Dr. Gilberto José de Moraes, da ESALQ/USP).

presença (ocorrências), com raio variado (200, 400, 800 e 1200 km), pelo uso de um polígono convexo mínimo (MCP) e pela definição de um retângulo envolvente, cobrindo toda a área definida pelos pontos, compreendendo seis áreas diferentes;

- para controlar a autocorrelação espacial potencial entre os dados de calibração e teste do modelo e melhorar sua transferibilidade, foi utilizado o método *spatial block cross-validation* (Roberts et al., 2017; Santini et al., 2021) para particionamento dos dados, através da função *part\_sblock*, com quatro partições, mantendo os valores *default* para as outras opções, resultando em 47, 63, 28, 12 pontos de presença em cada uma das partições criadas;
- amostragem de pontos de *background* (o entorno ambiental) na área de calibração, foi feita de acordo com as partições criadas, tendo sido estabelecido pontos aleatórios (*method = random*, da função *sample\_background*) proporcionais às ocorrências, sendo definidos, dessa forma, 3.133, 4.200, 1.867 e 800 pontos para cada partição (10.000 no total, padrão do Maxent);
- as 19 variáveis bioclimáticas da base de dados WorldClim 2.1 (Fick; Hijmans, 2017) em resolução es-

pacial de 2,5 minutos (~ 5km), selecionadas como variáveis abióticas, foram corrigidas com relação à colinearidade através do fator de inflação de variância (*variance inflation factor – VIF*; Marquardt, 1970), sendo que as nove variáveis remanescentes (*Bio2, Bio3, Bio8, Bio9, Bio13, Bio14, Bio15, Bio18 e Bio19*) foram utilizadas para modelagem;

- o Maxent possui dois parâmetros principais que podem ser ajustados: a) *features*; e b) multiplicador de regularização (RM). Uma *feature* corresponde a uma transformação matemática das diferentes covariáveis utilizadas no modelo para permitir que relações complexas sejam modeladas (Elith et al., 2010). O multiplicador de regularização é um parâmetro que adiciona novas restrições, ou seja, é uma penalidade imposta ao modelo. O principal objetivo é evitar a complexidade excessiva e/ou *overfitting*, controlando a intensidade das *features* usadas para construir o modelo (Elith et al., 2010; Shcheglovitova; Anderson, 2013). Nesse contexto, foi utilizada a função *tune\_max* para calibrar 50 modelos Maxent, resultantes da combinação de RMs de 0,5 até 5 com incremento de 0,5 e FCs “L” (linear), “Q” (quadrática), “H” (*hinge*), “LQ” e “LQH”, com saída no formato *cloglog*. Pode-se omitir *fea-*

tures de produto porque as curvas de resposta (marginais) para cada variável preditora definem completamente o modelo e são mais fáceis de interpretar do que aquelas que dependem dos valores de outras variáveis. Um limiar que maximiza a soma de sensibilidade e especificidade (*max\_sens\_spec*) foi utilizado para obter o mapa binário da adequação ambiental e a métrica para selecionar a melhor combinação de parâmetros foi TSS (Allouche et al., 2006). Os modelos Maxent foram ajustados através de um processo de Poison não homogêneo (Phillips et al., 2017);

- a curva ROC foi calculada, juntamente com a AUC parcial (expressa em termos percentuais), para os modelos resultantes, considerando limites de 10%, para construção de um gráfico;
- o modelo selecionado foi projetado para o Brasil com o uso da função *sdm\_predict*, mantendo as projeções até o limite do limiar utilizado e saída no formato *cloglog*, para obtenção do mapa com as probabilidades e do mapa binário (a partir da aplicação do limiar definido), sendo a área calculada pelo mapa binário, com o uso da função *expanse* do pacote *terra*.

Foram calculados os parâmetros do modelo e suas métricas de avaliação. As métricas utilizadas foram: IMAE (*Inverse Mean Absolute Error*), TPR (*True Positive Rate*), TNR (*True Negative Rate*), Sorensen, Jaccard, FPB (*F Presence-Background*), OR (*Omission Rate*), TSS (*True Skill Statistics*), área sob a curva (AUC), e o índice de Boyce contínuo. O IMAE é calculado como  $1 - \text{Mean Absolute Error}$  para ser consistente com as outras métricas, onde quanto maior o valor de uma determinada métrica de desempenho, maior a precisão do modelo. A TPR ou sensibilidade é proporção de presenças corretamente previstas entre todas as presenças (Fielding; Bell, 1997). Em contrapartida, a TNR, ou especificidade, mede a proporção de ausências previstas corretamente entre todas as ausências (Fielding; Bell, 1997). Sorensen e Jaccard são os índices de similaridade de Sørensen (Sorensen, 1948; Leroy et al., 2018) e Jaccard (Jaccard, 1912; Leroy et al., 2018), respectivamente, largamente utilizados em aplicações voltadas à ecologia. FPB corresponde à média harmônica ponderada entre precisão e sensibilidade (Daskalaki et al., 2006), considerando pontos de presença e *background*. OR, ou taxa de omissão, indica a porcentagem de localidades de teste que cai em áreas não previstas como adequados para a espécie e deve ser baixa para um bom desempenho do modelo (Phillips et al., 2006; Fielding; Bell, 1997). TSS é definida como a média da taxa de sucesso de previsão para sites presentes

e ausentes (Allouche et al., 2006). A área sob a curva (AUC) ROC (Fielding; Bell, 1997) expressa uma métrica independente de *threshold* e, a despeito de diversas críticas, continua sendo utilizada. Finalmente, o índice de Boyce contínuo, que mede a concordância entre os gradientes previstos de adequação do habitat e a distribuição de pontos de teste retidos (Hirzel et al., 2006). Essa métrica requer apenas dados de presença e é independente de um *threshold*. Um valor mais alto indica que as previsões do modelo são consistentes com a distribuição observada de presenças nas regiões de teste.

A Tabela 1 apresenta os resultados obtidos, com os parâmetros do modelo selecionado para cada extensão e suas métricas (médias e desvios padrões) para avaliação de sua capacidade de discriminação (capacidade de classificar corretamente presenças e ausências) e comparações, fornecidos pelo pacote *flexsdm*. A avaliação dos modelos é imprescindível para detectar erros potenciais ou baixa capacidade preditiva (Fielding; Bell, 1997). Ressalta-se que não é objetivo deste trabalho identificar um melhor modelo, mas apenas considerar os efeitos possíveis da escolha da extensão da área de estudo nos resultados que podem ser obtidos.

**Tabela 1.** Resultados dos modelos e métricas (valores médios e desvios padrões) utilizados para avaliação de sua capacidade de discriminação, calculados pelo pacote *flexsdm* para diferentes tamanhos de *buffer*, polígono convexo mínimo (MCP) e retângulo envolvente (RET), considerando a extensão total da área de estudo.

Resultados	200 km	400 km	800 km	1.200 km	MCP	RET
RM	2,5	3,5	0,5	0,5	1,0	1,0
Features	q	h	lq	lq	h	h
Valor do <i>Threshold</i> <sup>A</sup>	0,690703	0,687512	0,506775	0,479131	0,501255	0,525651
Área (km <sup>2</sup> ) <sup>B</sup>	5492933	12954210	28696074	45292591	48048774	63768341
Área Adequada (km <sup>2</sup> ) <sup>C</sup>	1827308	2515338	3449437	3487684	2928716	2852826
% Área Adequada <sup>D</sup>	33,27	19,42	12,02	7,70	6,10	4,47
Background % <sup>E</sup>	14,79	6,23	2,79	1,75	1,71	12,67
Métricas	200 km	400 km	800 km	1.200 km	MCP	RET
IMAE (dp)	0,08363	0,03996	0,02618	0,03148	0,02594	0,01359
TPR (média)	0,79177	0,77527	0,80412	0,82721	0,76427	0,76228
TPR (dp)	0,18247	0,12063	0,10910	0,07880	0,05271	0,01755
TNR (média)	0,64525	0,68148	0,72681	0,75340	0,78860	0,79712
TNR (dp)	0,25593	0,17824	0,15366	0,15500	0,13786	0,14162
Sorensen (média)	0,07620	0,07634	0,09206	0,10689	0,13563	0,13952

Métricas	200 km	400 km	800 km	1.200 km	MCP	RET
Sorensen (dp)	0,02792	0,02461	0,03300	0,04164	0,09768	0,08855
Jaccard (média)	0,03977	0,03981	0,04849	0,05684	0,07504	0,07684
Jaccard (dp)	0,01500	0,01325	0,01807	0,02304	0,05846	0,05174
FPB (média)	0,07954	0,07962	0,09698	0,11368	0,15009	0,15368
FPB (dp)	0,02999	0,02650	0,03613	0,04609	0,11691	0,10347
OR (média)	0,20823	0,22473	0,19588	0,17279	0,23573	0,23772
OR (dp)	0,18247	0,12063	0,10910	0,07880	0,05271	0,01755
TSS (média)	0,43703	0,45675	0,53093	0,58061	0,55287	0,55941
TSS (dp)	0,14600	0,11318	0,13460	0,14216	0,13622	0,14694
AUC (média)	0,70872	0,71874	0,76574	0,79539	0,79945	0,79337
AUC (dp)	0,08617	0,07037	0,07752	0,06945	0,05127	0,05565
Boyce (média)	0,84563	0,81091	0,83890	0,80365	0,81696	0,80409
Boyce (dp)	0,11442	0,26607	0,17418	0,24251	0,22391	0,19829

<sup>A</sup> máximo (sensibilidade + especificidade).

<sup>B</sup> Extensão da área utilizada para calibrar o modelo (área de estudo).

<sup>C</sup> Área da distribuição geográfica potencial, calculada com base no mapa binário (presença/ausência).

<sup>D</sup> Percentual da área de distribuição geográfica potencial relativo à extensão total da área de estudo.

<sup>E</sup> Percentual da área de estudo ocupado pelos pontos de *background*.

Cores mais escuras representam melhores métricas.

Ficou claro que o aumento na área de estudo inflacionou a maioria das métricas utilizadas, conforme se pode observar pela Tabela 1, exceto para TPR e OR, cujo padrão aponta para comportamentos diferentes entre áreas delimitadas por *buffer* e por polígonos. Esse resultado parece corroborar o argumento de buscar um critério ecológico para delimitar adequadamente a extensão da área de estudo e sugere que o método utilizado para delimitação também apresente alguma influência, o que deverá ser melhor investigado.

Os resultados de modelos dessa natureza podem ser viesados se a área de estudo incluir regiões que são

adequadas para a espécie, mas que não são ocupadas por ela, devido a: (1) falta de dispersão (ou falta de permanência das populações); (2) interações bióticas; e/ou (3) mudanças ambientais causadas pelo homem (Anderson, [s.d.]). Por essas razões, levando em consideração as informações disponíveis, deve-se escolher uma região de estudo (ambiente geográfico), que seja o mais consistente possível com a suposição de que nenhum desses três fatores seja responsável pela espécie habitar um subconjunto ambientalmente inadequado (Anderson, [s.d.]). Com base nisso, embora os modelos resultantes do uso de MCP e do retângulo envolvente tenham boas métricas, poderia ser argumentado que sua utilização,

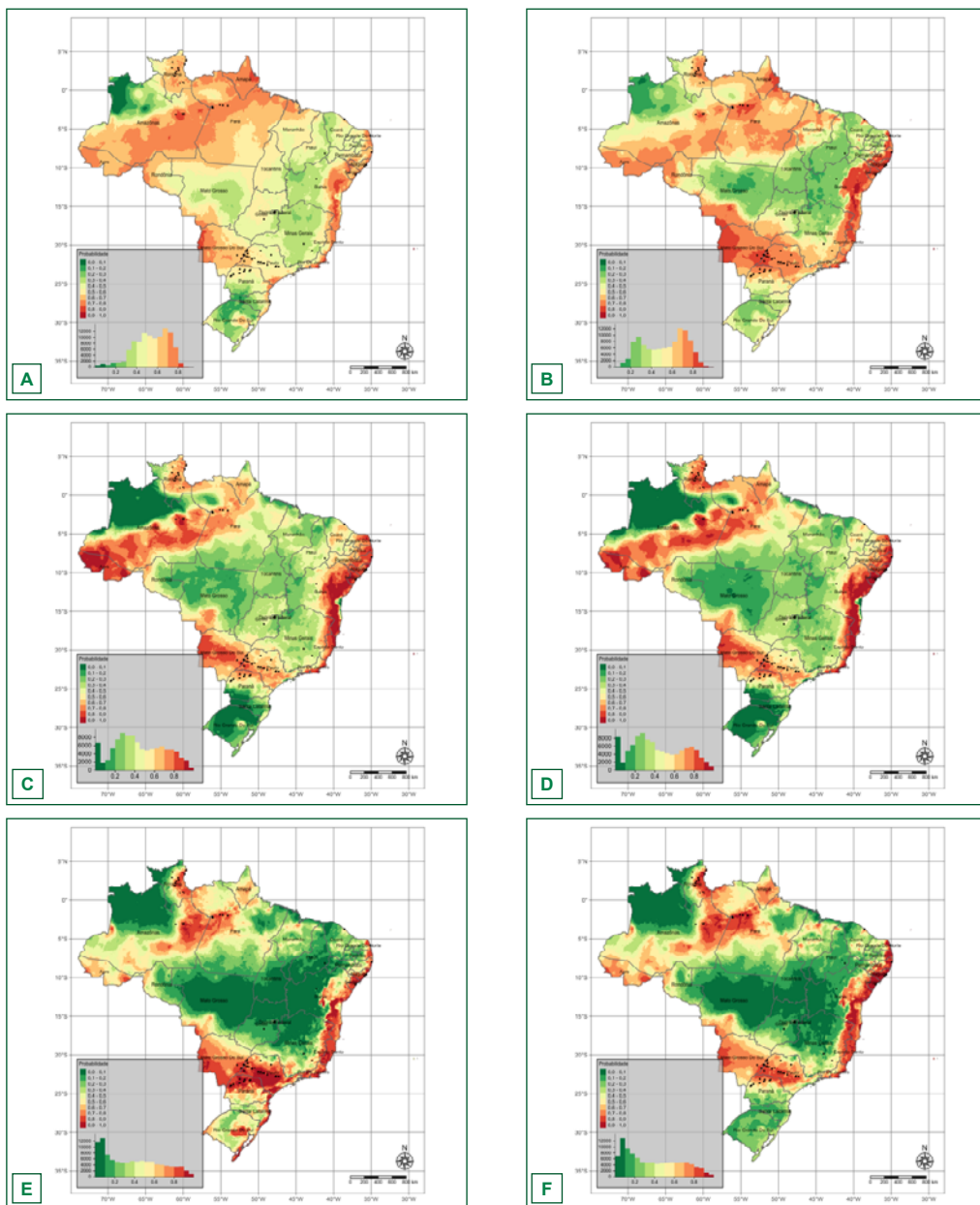


neste caso, seria inadequada, considerando a grande extensão da África sem ocorrências, apesar de estar na mesma faixa climática onde *R. indica* tenha sido registrada. Contudo, não foi possível identificar que pesquisas tenham sido feitas nessa região sem registrar a presença de *R. indica* e nem o motivo de, eventualmente, não haverem sido feitas tais pesquisas. Por outro lado, deixar de considerar essa região, talvez represente perda de informação relevante para a construção do modelo, especialmente se for considerada a definição de *background* e seu efeito em modelos Maxent.

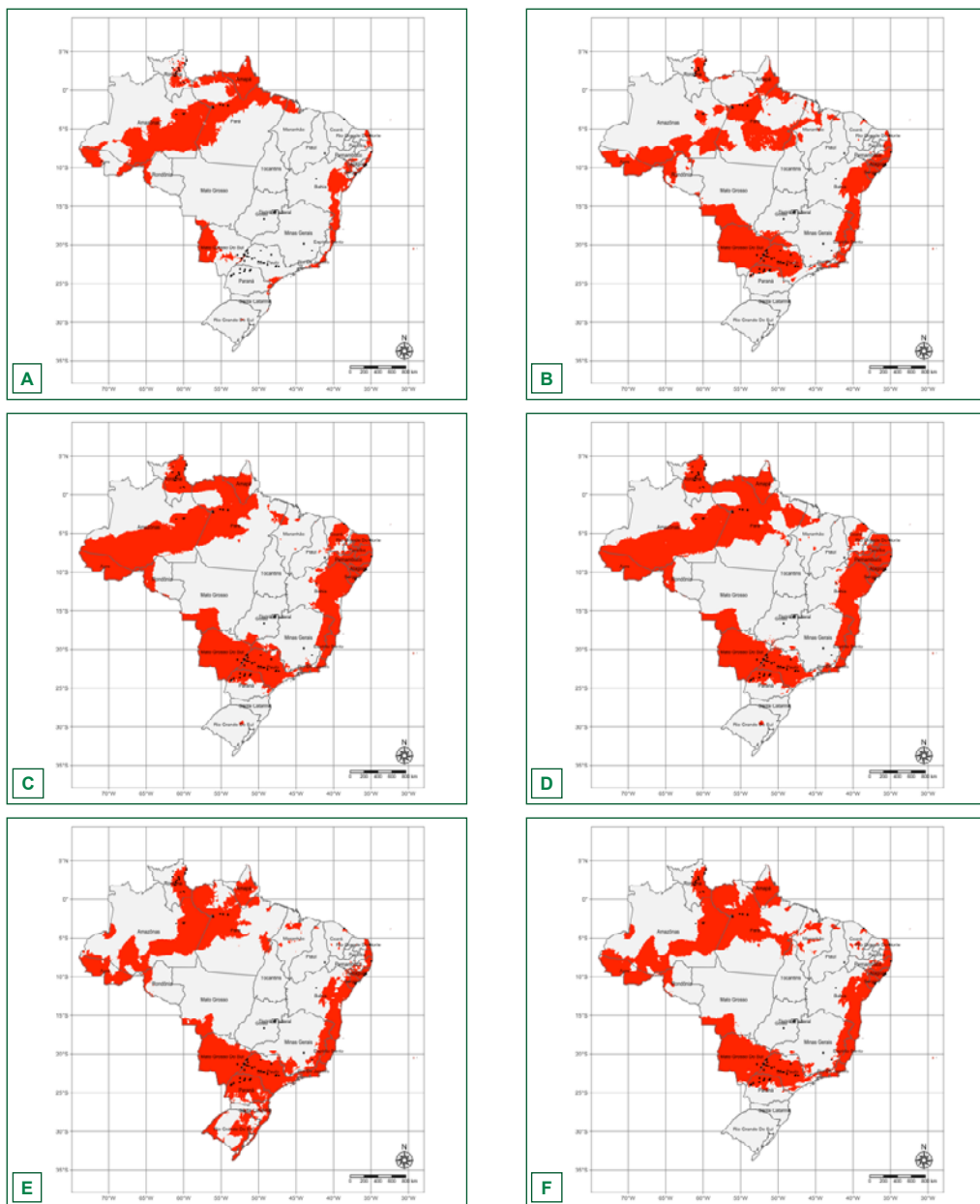
Considerando ainda que *R. indica* é uma espécie invasora, classificada como praga quarentenária presente no Brasil, o objetivo de um modelo de sua distribuição potencial seria identificar, juntamente com avaliações relativas à importância econômica de culturas hospedeiras, áreas em que represente risco mais acentuado e, dessa forma, orientar políticas públicas de controle fitossanitário. Sendo assim, o custo de falsos negativos seria potencialmente maior do que o custo de falsos positivos

para aquelas localidades em que as culturas hospedeiras tenham relevância econômica.

Pode-se observar pelas diferenças apresentadas nas Figura 1 e 2 o efeito da variação da escolha da área de estudo no resultado projetado do modelo, influenciando diretamente a identificação de áreas geográficas adequadas para a espécie (sua distribuição potencial). Previsões com áreas muito pequenas, comparativamente a outras, podem representar *overfitting* do modelo e resultarão em previsões falso-negativas de registros de localidade observados. Este é frequentemente o caso quando a amostragem da área é insuficiente para capturar toda a gama de condições ambientais ocupadas por uma espécie. Áreas muito grandes, por outro lado, resultam na identificação de locais que não estão ocupados pela espécie (previsões falso-positivas), que podem ser causadas por localidades que estejam fora da sua distribuição real, ou indicar restrições de dispersão devido a algum tipo de barreira ou, ainda, interações bióticas como competição e predação (Raxworthy et al., 2007).



**Figura 1.** Mapas da distribuição geográfica potencial de *Raoiella indica* no Brasil, apresentando a probabilidade relativa em intervalos de 0,1 para buffer de 200 (a), 400 (b), 800 (c) e 1.200km (d), MCP (e) e retângulo envolvente (f), considerando os dados de presença, variáveis predictoras e parâmetros utilizados para modelagem.



**Figura 2.** Mapas de presença (vermelho) / ausência de *Raziella indica* no Brasil, a partir da aplicação do threshold definido, para buffer de 200 (a), 400 (b), 800 (c) e 1.200km (d), MCP (e) e retângulo envolvente (f), considerando os dados de presença, variáveis preditoras e parâmetros utilizados para modelagem.

O modelo com *buffer* de 200 km, ainda que apresente uma complexidade inferior aos demais modelos que utilizaram o mesmo princípio para delimitar a extensão, foi o segundo mais penalizado, com  $RM = 2.5$ . Muitos registros de presença, especialmente nos estados de São Paulo e Paraná ficaram fora da área considerada como adequada para a espécie, ou seja, é um modelo que apresenta erros de omissão facilmente identificáveis.

De forma geral, os modelos MCP e RET apresentaram métricas bastante semelhantes (Tabela 1) e, embora a área total da extensão utilizada para o modelo RET tenha sido 33% maior do que área de extensão do modelo MCP, o seu resultado, em termos de área de distribuição potencial de *R. indica* prevista (Figura 2), foi 3% menor. O mapa binário do modelo MCP incluiu áreas nos estados de Santa Catarina e Rio Grande do Sul, que foram classificadas como de menor probabilidade de ocorrência (Figura 1) pelo modelo RET e, pela aplicação do corte promovido pelo limiar (*threshold*) estabelecido, não foram incluídas por este último.

Pode-se perceber também que, à medida que aumenta a área de extensão, a distribuição geográfica potencial se torna mais pontual (menos inclusiva, de forma geral). Dessa forma, pontos de presença em áreas de menor probabilidade não pertencem à área delimitada pelo limiar resultante, especialmente considerando que os modelos com *buffer* de 200 e 400 km possuem os valores mais altos de *threshold*.

A extensão da área utilizada durante o processo de modelagem tem influências abrangentes no resultado do modelo e, especialmente, se for muito limitada, a importância de alguns fatores na delimitação da distribuição pode ser subestimada (Barve et al., 2011). Aumentar a extensão inclui frequentemente ausências que estão mais distantes ambientalmente das presenças, de modo que os valores da AUC aumentam, assim como outras medidas de discriminação (Lobo et al., 2008), conforme pode ser observado na Tabela 1.

Fica evidente que processos de dispersão e habitat devem ser considerados para que se possa delimitar adequadamente a área de extensão para modelagem da dispersão de espécies. Contudo, nem sempre essas informações estão disponíveis e, no caso especial de espécies invasoras, as características de dispersão da espécie não dependem apenas de sua capacidade, mas, principalmente, de outros agentes que atuam como vetores. A utilização de um processo de “modelagem exploratória”, como feito neste trabalho, pode auxiliar a entender os detalhes subjacentes à dispersão atual e potencial de espécies, o que pode ser determinante para a escolha adequada de parâmetros de calibragem dos modelos, especialmente considerando os riscos econômicos envolvidos, decorrentes da presença de espécies invasoras.

Preferencialmente, a extensão da área de estudo e, portanto, dos dados

de ocorrência, deve cobrir a área considerada acessível à espécie de interesse durante o período de tempo considerado, e os dados de fundo (*background*) devem ser limitados a essa mesma extensão (Barve et al., 2011). No entanto, se as informações de ocorrência não cobrirem toda a extensão acessível a uma espécie, a extensão geográfica desses dados ainda deve definir a extensão para o desenvolvimento do modelo. Se os dados de *background* forem extraídos de áreas fora da faixa de amostragem do local de ocorrência, os resultados serão alterados e as métricas de avaliação infladas.

Finalmente, deve-se entender que o resultado de modelos como Maxent referem-se à adequação relativa do ambiente, ou seja, um local é mais adequado que outro, para a espécie de interesse, considerando as variáveis e parâmetros utilizados durante a modelagem, mas os valores não refletem a ocupação.

## Agradecimentos

À Empresa Brasileira de Pesquisa Agropecuária – Embrapa, cujo suporte financeiro, através do projeto “Avaliação da Distribuição Geográfica e Riscos Econômicos Potenciais da Mosca-Oriental-das-Frutas (*Bactrocera dorsalis*) e da Mosca-da-Carambola (*Bactrocera carambolae*) no Brasil.” (10.20.03.056.00.00), desenvolvido na Embrapa Roraima, possibilitou a realização deste trabalho.

Aos revisores anônimos que ofereceram valiosas contribuições para melhoria do texto. E a toda equipe de suporte da Embrapa Roraima, pelo apoio recebido durante a realização das atividades necessárias, especialmente à bibliotecária da Embrapa Roraima, Jeana Garcia Beltrão Macieira, pela normalização bibliográfica.

## Referências

- ALLOUCHE, O.; TSOAR, A.; KADMON, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). **Journal of Applied Ecology**, v. 43, n. 6, p. 1223–1232, dez., 2006.
- AMARO, G.; FIDELIS, E. G.; SILVA, R. S. da; MEDEIROS, C. M. de. Current and potential geographic distribution of red palm mite (*Raoiella indica* Hirst) in Brazil. **Ecological Informatics**, v. 65, p. 101396, nov. 2021.
- ANDERSON, R. P. El Modelado De Nichos Y Distribuciones: No Es Simplemente “Clic, Clic, Clic”. **Biogeografía**, v. 8, p. 4-27, 2015. Disponível em: < [https://www.researchgate.net/publication/287194385\\_Anderson\\_R\\_P\\_2015\\_El\\_modelado\\_de\\_nichos\\_y\\_distribuciones\\_no\\_es\\_simplemente\\_clic\\_clic\\_clic\\_Modeling\\_niches\\_and\\_distributions\\_it%27s\\_not\\_just\\_click\\_click\\_click\\_La\\_modelisation\\_de\\_niche\\_et\\_de\\_distributions](https://www.researchgate.net/publication/287194385_Anderson_R_P_2015_El_modelado_de_nichos_y_distribuciones_no_es_simplemente_clic_clic_clic_Modeling_niches_and_distributions_it%27s_not_just_click_click_click_La_modelisation_de_niche_et_de_distributions)>. Acesso em: 27 jul. 2022.
- BARBET-MASSIN, M.; ROME, Q.; VILLEMANT, C.; COURCHAMP, F. Can species distribution models really predict the expansion of invasive species? **PLoS ONE**, n. 13, e0193085, 2018.
- BARVE, N.; JIMÉNEZ-VALVERDE, A.; LIRANORIEGA, A.; MAHER, S. P.; PETERSON, A. T.; SOBERÓN, J.; VILLALOBOS, F. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. **Ecological Modelling**, v. 222, n. 11, p. 1810–1819, jun. 2011.
- BEAUMONT, L. J.; GALLAGHER, R. V.; THUILLER, W.; DOWNEY, P. O.; LEISHMAN, M. R.; HUGHES, L. Different climatic envelopes among invasive populations may lead to underestimations of current

- and future biological invasions. **Divers. Distrib.**, n. 15, p. 409–420, 2009.
- BROENNIMANN, O.; GUIGAN, A. Predicting current and future biological invasions: both native and invaded ranges matter. **Biol. Lett.**, n. 4, p. 585–589, 2008.
- DASKALAKI, S.; KOPANAS, I.; AVOURIS, N. Evaluation of classifiers for an uneven class distribution problem. **Applied Artificial Intelligence**, n. 20, p. 381–417, 2006.
- ELITH, J. Predicting distributions of invasive species. In: ROBINSON, J.; WALSHE, T.; BURGMAN, J.; NUNN, M. (Ed.). **Invasive species: risk assessment and management**. Cambridge, UK: Cambridge University Press, 2017.
- ELITH, J.; GRAHAM, C. H.; ANDERSON, R. P.; DUDÍK, M.; FERRIER, S.; GUIGAN, A. R.; HIJMANS, J.; HUETTMANN, F. J.; LEATHWICK, R.; LEHMANN, A. Novel methods improve prediction of species' distributions from occurrence data. **Ecography**, n. 29, p. 129–151, 2006.
- ELITH, J.; KEARNEY, M.; PHILLIPS, S. The art of modelling range-shifting species. **Methods in ecology and evolution**, v. 1, n. 4, p. 330–342, 2010.
- ELITH, J.; PHILLIPS, S. J.; HASTIE, T.; DUDÍK, M.; CHEE, Y. E.; YATES, C. J. A statistical explanation of MaxEnt for ecologists. **Divers. Distrib.**, n. 17, p. 43–57, 2011.
- FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International journal of climatology**, v. 37, n. 12, p. 4302–4315, 2017.
- FIELDING, A. H.; BELL, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. **Environ. Conserv.**, n. 24, p. 38–49, 1997.
- GRINNELL, J. Field tests of theories concerning distributional control. **American Naturalist**, n. 51, p. 115–128, 1917. <http://dx.doi.org/10.1086/279591>.
- HANBERRY, B. B.; HE, H. S.; PALIK, B. J. Pseudoabsence generation strategies for species distribution models. **Plos One**, n. 7, 2012.
- HEIKKINEN, R. K.; MARMION, M.; LUOTO, M. Does the interpolation accuracy of species distribution models come at the expense of transferability? **Ecography**, n. 35, p. 276–288, 2012.
- HIJMANS, R. J.; ELITH, J. **Species Distribution Models**. 2021. Disponível em <[https://rspatial.org/raster/sdm/raster\\_SDM.pdf](https://rspatial.org/raster/sdm/raster_SDM.pdf)>. Acesso em: 12 jul. 2022.
- HIRZEL, A. H.; LE LAY, G.; HELFER, V.; RANDIN, C.; GUIGAN, A. Evaluating the ability of habitat suitability models to predict species presences. **Ecological Modelling**, n. 199, p. 142–152, 2006. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>.
- JACCARD, P. The distribution of the flora in the alpine zone. 1. **New phytologist**, v. 11, n. 2, p. 37–50, 1912.
- LEROY, B.; DELSOL, R.; HUGUENY, B.; MEYNARD, C. N.; BARHOUMI, C.; BARBET-MASSIN, M.; BELLARD, C. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. **Journal of Biogeography**, v. 45, n. 9, p. 1994–2002, set. 2018.
- LOBO, J. M.; JIMÉNEZ-VALVERDE, A.; REAL, R. AUC: a misleading measure of the performance of predictive distribution models. **Global Ecology and Biogeography**, n. 17, p. 145–151, 2008.
- MARQUARIDT, D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. **Technometrics**, n. 12, p. 591–612, 1970.
- MAU-CRIMMINS, T. M.; SCHUSSMAN, H. R.; GEIGER, E. L. Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. **Ecol. Modell.**, n. 193, p. 736–746, 2006.
- MEROW, C.; SMITH, M. J.; SILANDER, J. A. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. **Ecography**, n. 36, p. 1058–1069, 2013.
- PEREIRA, R. H. M.; GONCALVES, C. N. **Geobr: Download Official Spatial Data Sets of Brazil**. GitHub repository, 2019. Disponível em: <<https://github.com/ipeaGIT/geobr>>. Acesso em: 27 jul. 2022.
- PETERSON, A. T.; SOBERÓN, J. Species distribution modeling and ecological niche

- modeling: getting the concepts right. **Nat. Conservação**, n. 10, p. 102–107, 2012.
- PHILLIPS, S. J.; ANDERSON, R. P.; DUDÍK, M.; SHAPIRE, R. E.; BLAIR, M. E. Opening the black box: An open-source release of Maxent. **Ecography**, n. 40, p. 1–7, 2017. <https://doi.org/10.1111/ecog.03049>.
- PHILLIPS, S. J.; ANDERSON, R. P.; SCHAPIRE, R. E. Maximum entropy modeling of species geographic distributions. **Ecol. Modell.**, n. 190, p. 231–259, 2006.
- PHILLIPS, S. J.; DUDÍK, M.; ELITH, J.; GRAHAM, C. H.; LEHMANN, A.; LEATHWICK, J.; FERRIER, S. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. **Ecological Applications**, n. 19, p. 181–197, 2009.
- R CORE TEAM. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2022. <https://www.R-project.org/>.
- RAXWORTHY, C. J.; INGRAM, C. M.; RABIBISOA, N.; PEARSON, R. G. Applications of Ecological Niche Modeling for Species Delimitation: A Review and Empirical Evaluation Using Day Geckos (*Phelsuma*) from Madagascar. **Systematic Biology**, v. 56, n. 6, p. 907–923, dez. 2007.
- ROBERTS, D. R.; BAHN, V.; CIUTI, S.; BOYCE, M. S.; ELITH, J.; GUILLERA-ARROITA, G.; WARTON, D. I. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. **Ecography**, n. 40, p. 913–929, 2017. <https://doi.org/10.1111/ecog.02881>.
- ROBIN, X.; TURCK, N.; HAINARD, A.; TIBERTI, N.; LISACEK, F.; SANCHEZ, J. C.; MÜLLER, M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. **BMC Bioinformatics**, v. 12, p. 77, 2011.
- SANTINI, L.; BENÍTEZ-LÓPEZ, A.; MAIORANO, L.; ČENGIĆ, M.; HUIJBREGTS, M. A. J. Assessing the reliability of species distribution projections in climate change research. **Diversity and Distributions**, v. 27, n. 6, p. 1035–1050, 2021.
- SHCHEGLOVITOVA, M.; ANDERSON, R. P. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. **Ecological Modelling**, v. 269, p. 9–17, 2013.
- SILLERO, N. What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. **Ecological Modelling**, v. 222, n. 8, p. 1343–1346, abr. 2011.
- SILLERO, N.; BARBOSA, M. Common mistakes in ecological niche models. **International Journal of Geographical Information Science**, v. 35, p. 213–226, 2021.
- SOBERÓN, J.; OSORIO-OLIVERA, L.; PETERSON, T. Conceptual differences between ecological niche modeling and species distribution modeling. **Revista Mexicana de Biodiversidad**, n. 88, p. 437–441, 2017.
- SORENSEN, T. A. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. **Biol. Skar.**, v. 5, p. 1–34, 1948.
- TENNEKES, M. tmap: Thematic Maps in R. **Journal of Statistical Software**, v. 84, n. 6, p. 1–39, 2018.
- VACLAVIK, T.; MEENTEMEYER, R. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. **Divers. Distrib.**, n.18, p. 73–83, 2012.
- VANDERWAL, J.; SHOO, L. P.; GRAHAM, C.; WILLIAM, S. E. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? **Ecological Modelling**, v. 220, p. 589–594, 2009.
- VARELA, S.; ANDERSON, R. P.; GARCÍA-VALDÉS, R.; FERNÁNDEZ-GONZÁLEZ, F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. **Ecography**, p. 1084–1091, jan. 2014.
- VELAZCO, S. J. E.; ROSE, M. B.; ANDRADE, A.; MINOLI, I.; FRANKLIN, J. flexsdm: An R package for supporting a comprehensive and flexible species distribution modelling workflow. **Methods in Ecology and Evolution**, v. 13, p. 2041–210X.13874, 2022.
- VENETTE, R. C. Climate analyses to assess risks from invasive forest insects: simple matching to advanced models. **Curr. For. Reports**, n. 3, p. 255–268, 2017.

WALLACE A. R. On the zoological geography of the Malay Archipelago. **Proceedings of the Linnean Society of London**, n. 4, p. 172-184, 1860. <http://dx.doi.org/10.1111/j.1096-3642.1860.tb00090.x>.

WISZ, M. S.; GUISAN, A. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. **BMC Ecology**, n. 9, p. 8, 2009.

ZIZKA, A.; SILVESTRO, D.; ANDERMANN, T.; AZEVEDO, J.; RITTER, C.D.; EDLER, D.; FAROOQ, H.; HERDEAN, A.; ARIZA, M.; SCHARN, R.; SVANTESSON, S.; WENGSTRÖM, N.; ZIZKA, V.; ANTONELLI, A. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. **Methods in Ecology and Evolution**, n. 10, p. 7, 2019.



Exemplares desta edição  
podem ser adquiridos na:

**Embrapa Roraima**

Rodovia BR 174, Km 8 - Distrito Industrial  
Caixa Postal 133 - CEP. 69.301-970  
Boa Vista | RR  
Fone/Fax: (95) 4009-7100  
Fax: +55 (95) 4009-7102  
www.embrapa.br  
www.embrapa.br/fale-conosco/sac

**1ª edição**



**Comitê de Publicações da Unidade**

Presidente

*Edmilson Evangelista Da Silva*

Secretário

*Daniel Augusto Schurt*

Membros

*Karine Dias Batista, Cássia Ângela Pedrozo,  
Newton de Lucena Costa, Carolina Volkmer de  
Castilho, George Correa Amaro, Oscar Jose  
Smiderle, Sandro Loris Aquino Pereira*

Normalização Bibliográfica

*Jeana Garcia Beltrão Macieira*

Revisão editorial

*Jeana Garcia Beltrão Macieira*

Revisão de texto

*Ilda Maria Sobral de Almeida*

Projeto gráfico da coleção

*Carlos Eduardo Felice Barbeiro*

Editoração eletrônica

*Phábrica de Produções:*

*Alecsander Coelho, Daniela Bissigui, Érsio Ribeiro e Paulo Ciola*

Foto da capa

*George Amaro*