

1.03.99 – Ciência da Computação.

**DETERMINAÇÃO DA INFLUÊNCIA DE VARIÁVEIS METEOROLÓGICAS E AEROSSÓIS NA DISTRIBUIÇÃO DE ESPÉCIES NA BACIA AMAZÔNICA**Renato O. Miyaji<sup>1\*</sup>, Pedro L. P. Corrêa<sup>2</sup>

1. Estudante da Escola Politécnica da Universidade de São Paulo (POLI-USP)
2. Professor da POLI-USP - Departamento de Eng. Computação e Sist. Digitais

**Resumo**

Localizada na região central da Floresta Amazônica, Manaus (AM) é considerada como um laboratório ideal para o estudo dos efeitos da ação antrópica no clima e nos ecossistemas terrestres em uma floresta tropical. Dessa forma, entre 2014 e 2015, foi realizado o projeto GoAmazon, que coletou dados atmosféricos na região. A partir dessas informações, foi possível a realização de análises. No entanto, nota-se a necessidade da aplicação de técnicas para análise dos dados do projeto, de modo a gerar novas informações. Nesse contexto, este projeto buscou utilizar métodos de interpolação espacial para agregar novos conhecimentos sobre pluma de Manaus. Esses foram utilizados em conjunto com os dados sobre a fauna local, de modo a aplicar uma modelagem de distribuição de espécies e determinar a amplitude dos nichos ecológicos das espécies, podendo, dessa forma identificar que as concentrações de isopreno e ozônio possuem forte influência na ocorrência da espécie *Coragyps atratus*.

**Palavras-chave:** Interpolação Espacial; GoAmazon; Nicho Ecológico.

**Apoio financeiro:** CNPq

**Trabalho selecionado para a JNIC:** Universidade de São Paulo (USP).

**Introdução**

Para promover um desenvolvimento sustentável, a Organização das Nações Unidas (ONU) criou 17 objetivos para serem cumpridos até 2030. Dentre esses, está “13 – Ação contra a mudança global do clima” (IBGE, 2021). Atualmente, a Floresta Amazônica está em pauta por conta de sua importância climática e por concentrar uma parcela significativa da biodiversidade mundial. Assim, estudos e análises desenvolvidos a respeito dela são de grande importância para o cumprimento desse objetivo.

Nesse contexto, entre 2014 e 2015, foi realizado o projeto GoAmazon 2014/15. Desenvolvido na região entre Manaus (AM) e Manacapuru (AM), a campanha possuía como objetivo principal entender e quantificar as interações complexas que se desenvolvem entre a vegetação, a atmosfera e a produção de aerossóis. Além de sua conexão com as nuvens, os aerossóis e a precipitação na área de estudo, tanto sob condições naturais, quanto sob efeitos da atividade antrópica. Assim, buscou-se aumentar a disponibilidade de dados climáticos a respeito da região (MARTIN et al., 2016).

A coleta foi feita por meio de nove estações de pesquisa situadas nas proximidades de Manaus e de voos realizados por aeronaves que realizaram voos em baixa altitude. Foram realizados dois períodos de operação intensa (IOPs) durante a estação seca e a chuvosa.

Nos repositórios do Portal da Biodiversidade do Instituto Chico Mendes (ICMBIO, 2021) e do *Global Biodiversity Information Facility* (GBIF, 2021) estão disponíveis dados referentes à biodiversidade da região amazônica. Dessa maneira, utilizando os dados climáticos e de aerossóis em conjunto com esses, pode-se desenvolver um experimento de modelagem de distribuição de espécies, de modo a analisar a distribuição de espécies existentes na área de interesse, com base nas condições ambientais locais.

O objetivo deste projeto foi a aplicação de métodos de interpolação espacial utilizando os dados do repositório do projeto GoAmazon 2014/15 para a geração de novos conhecimentos. A partir desses, buscou-se utilizar a modelagem de distribuição de espécies para determinar as amplitudes dos nichos ecológicos das espécies observadas na região de Manaus e Manacapuru (AM).

**Metodologia**

A interpolação espacial foi desenvolvida utilizando a linguagem *Python*, através da aplicação *web Jupyter Notebook*. Para ela, utilizou-se os dados coletados pela aeronave G-1 e disponibilizados no portal do *Atmospheric Radiation Measurement* (ARM, 2021). Separou-se os dados para voo. Então, aplicou-se o filtro: foram consideradas apenas as medições realizadas abaixo da altitude de 1500 m, de modo a garantir homogeneidade. Posteriormente, criou-se a malha de interpolação. Com a malha pronta, realizou-se a interpolação para cada voo. Utilizou-se três diferentes métodos: uma interpolação linear baricêntrica, através de *splines* e pelo algoritmo de *nearest neighbor*. A partir dos valores interpolados, calculou-se o método de sumarização. Para todas as variáveis, utilizou-se a média. Para a temperatura, ainda se aplicou o máximo e o mínimo. A metodologia aplicada é apresentada na Figura 1.

Com o objetivo de se validar os resultados obtidos, aplicou-se uma validação cruzada. Por meio do método de *k-fold*, com  $k=3$ , foi possível avaliar funções de custo, como o erro médio absoluto (*Mean Absolute Error*), o erro médio absoluto percentual (*Mean Absolute Percentage Error*) e a raiz quadrada do erro médio (*Root-Mean-Square Error*). Através dessas pôde-se avaliar a acurácia da interpolação. Além disso, também se considerou a suavidade das curvas e a fidelidade aos dados originais.



Figura 1 – Metodologia de interpolação espacial

O experimento de modelagem de distribuição de espécies foi conduzido por meio da linguagem R. Os dados referentes à existência de espécies foram tratados. Filtrou-se apenas os dados dentro da mesma área obtida pelo processo de interpolação espacial e para as mesmas datas. Obteve-se 734 espécies distintas. Filtrou-se apenas para espécies que possuíssem no mínimo 17 ocorrências distintas. Dessa forma, foi possível criar um *dataset* de dados bioclimáticos.

A metodologia utilizada é a proposta por Pinaya & Corrêa (2014). Construiu-se a hipótese inicial de que a variação da concentração de aerossóis e poluentes, devido à ação antrópica, influencia a distribuição de espécies na região de Manaus e Manacapuru (AM). Na etapa de modelagem, selecionou-se as variáveis predictoras. A quantidade de variáveis deveria ser reduzida, de modo a evitar que o modelo incorporasse padrões aleatórios. Para isso, utilizou-se o coeficiente de Pearson, eliminando as variáveis de alta correlação entre si. Também foi aplicada a técnica de Análise de Componentes Principais (PCA). Em seguida, selecionou-se e executou-se o modelo de Máxima Entropia, por geralmente apresentar os melhores resultados em relação aos demais para bases de dados pequenas e médias (PHILLIPS et. al., 2005). Para testar o desempenho preditivo do modelo utilizou-se a *Area Under the Receiver Operating characteristic Curve* (AUC-ROC). Por fim, validou-se a hipótese científica. A metodologia do experimento é apresentada na Figura 2.

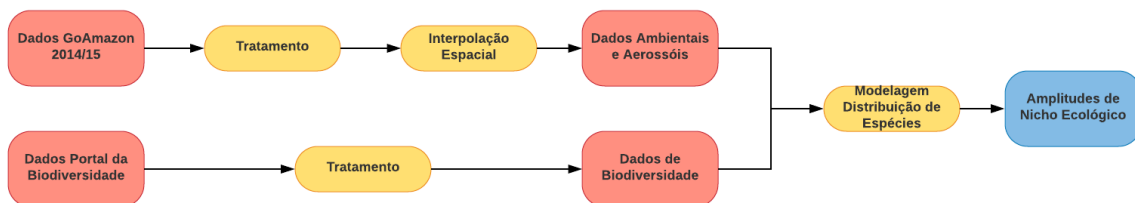


Figura 2 – Metodologia do experimento de modelagem de distribuição de espécies

## Resultados e Discussão

Na Tabela 1, pode-se observar o MAE e o RMSE, calculados para a variável concentração de monóxido de carbono na estação seca. Observa-se que o método de interpolação por *splines* apresenta os maiores erros médios. Já entre os demais métodos, o linear apresenta consistentemente erros menores em comparação com o *nearest neighbor*. Para todas as variáveis, nota-se o mesmo comportamento: o método de interpolação linear apresenta um MAPE médio menor, de 9,29 %, em comparação com os demais: 10,09 % para o método por *splines* e de 12,75 % para o método *nearest neighbor*.

Método	MAE	RMSE
<i>Linear</i>	0,0152	0,0300
<i>Spline</i>	0,0209	0,0478
<i>Nearest Neighbor</i>	0,0165	0,0332

Tabela 1 – MAE e RMSE obtidos para interpolação de CO (ppm)

Na Figura 3, pode-se observar a interpolação obtida para a concentração de monóxido de carbono em um dos voos voo 18. Nela, pode-se notar que a interpolação por *spline* apresenta a maior suavidade das curvas. Em seguida, os resultados por interpolação linear possuem a segunda maior suavidade. Já a interpolação por *nearest neighbor* apresenta pouca suavidade e limiares bruscos entre as regiões.

Considerando os resultados das comparações realizadas, optou-se por utilizar o método de interpolação espacial linear, por conta dos menores erros e suavidade maior nas curvas.

Para a Modelagem de Distribuição de Espécies, selecionou-se a espécie que apresentava a maior quantidade de ocorrências: a *Coragyps atratus*, o urubu de cabeça preta. Essa apresentava 54 pontos de ocorrência distintos e pouca dispersão espacial, fatores que contribuem para um melhor desempenho dos modelos de distribuição de espécies (HERNANDEZ et al., 2006).

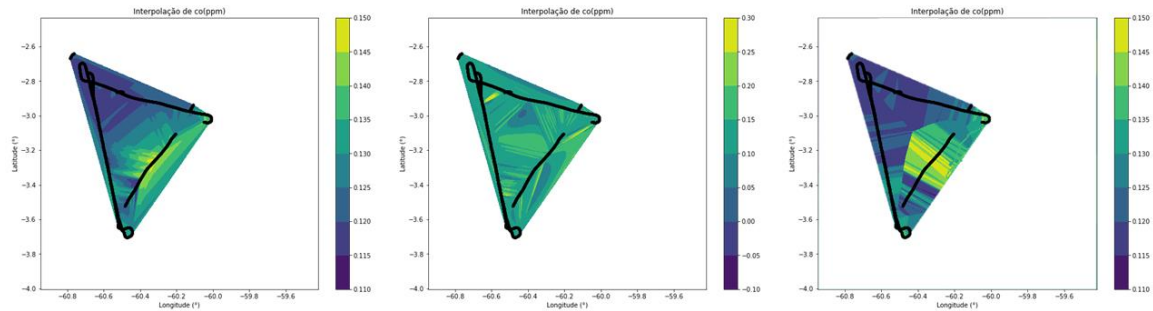


Figura 3 – Comparação da suavidade de curvas para método linear (esquerdo), *spline* (centro) e *nearest neighbor* (direita)

Na etapa de Pré-Análise de Dados, calculou-se a matriz de correlação para o *dataset*, optando por retirar uma das variáveis dos pares que eram muito correlacionados entre si – com módulo superior a 80 % (MATEO et al., 2013). Além disso, uma análise de componentes principais (PCA) também foi realizada, avaliando a variância explicada, observando que os nove primeiros componentes principais representavam aproximadamente 95 % da variância explicada total. Dessa forma, praticamente não haveria perda ao se reduzir a dimensão do conjunto de dados para nove, retirando as variáveis altamente correlacionadas entre si.

Assim, a partir do *dataset* com nove atributos e uma variável-resposta correspondente à ocorrência de *Coragyps atratus*, foi possível aplicar o Modelo de Máxima Entropia. Obteve-se um modelo com desempenho superior ao classificador base (aleatório) com ROC-AUC de 73 %, como visto na Figura 4. Desse modo, o modelo desenvolvido possui elevada capacidade preditora.

A partir do modelo treinado foi feita a previsão para toda a área de dados disponível, de modo a se obter mapas de distribuição potencial, que indicam a probabilidade de ocorrência da espécie de interesse em função das coordenadas de latitude e longitude. Na Figura 4, esse pode ser visto para a espécie *Coragyps atratus*. Do mapa de distribuição potencial gerado, percebe-se que as regiões com as maiores probabilidades de ocorrência concentram-se na porção leste dos mapas, nas proximidades da cidade de Manaus (AM).

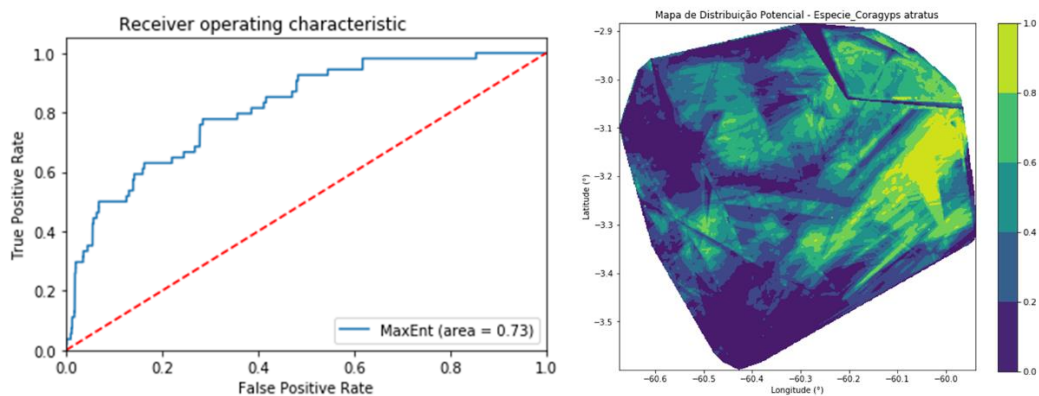


Figura 4 – Curva ROC (esquerda) e mapa de distribuição potencial (direita) para *Coragyps atratus*

Ademais, foi possível determinar as curvas de resposta, que indicam a dependência existente entre a probabilidade de ocorrência das espécies em função do valor de cada uma das variáveis preditoras, que pode ser vista na Figura 5. Assim, foi possível analisar a hipótese científica inicial, notando que as variáveis com maior influência positiva na probabilidade de ocorrência foram: a concentração de óxidos de nitrogênio, as temperaturas mínima e máxima, a fração volumétrica de água e a concentração de acetona. Especificamente para essa última, os efeitos possuem módulo inferior. Já as variáveis concentração de isopreno, ozônio, metano e monóxido de carbono apresentaram efeitos negativos. Nota-se que as condições ideais para as quais a probabilidade de ocorrência é máxima são semelhantes às que ocorrem nas proximidades de Manaus (AM), onde a maior parte das ocorrências reais da espécie são observadas.

Dessa forma, foi possível comprovar a hipótese científica inicial de que variáveis meteorológicas e de aerossóis possuem efeitos na probabilidade de ocorrência da espécie analisada. Em especial, nota-se que a concentração do poluente isopreno apresentou elevada influência negativa na espécie, sendo a probabilidade de ocorrência praticamente nula quando a concentração supera 3 ppb. Da mesma forma, a concentração de ozônio mostrou possuir um efeito negativo, com a probabilidade de ocorrência baixa quando a concentração é superior a 40 ppb. Esses resultados preliminares a respeito da influência de aerossóis e variáveis meteorológicas na

distribuição de espécies de pássaros na região Amazônica mostraram-se praticamente inéditos na literatura, sendo necessária a realização de análises mais aprofundadas de cunho fisiológico para aprimorar a compreensão a respeito dessas interações observadas.

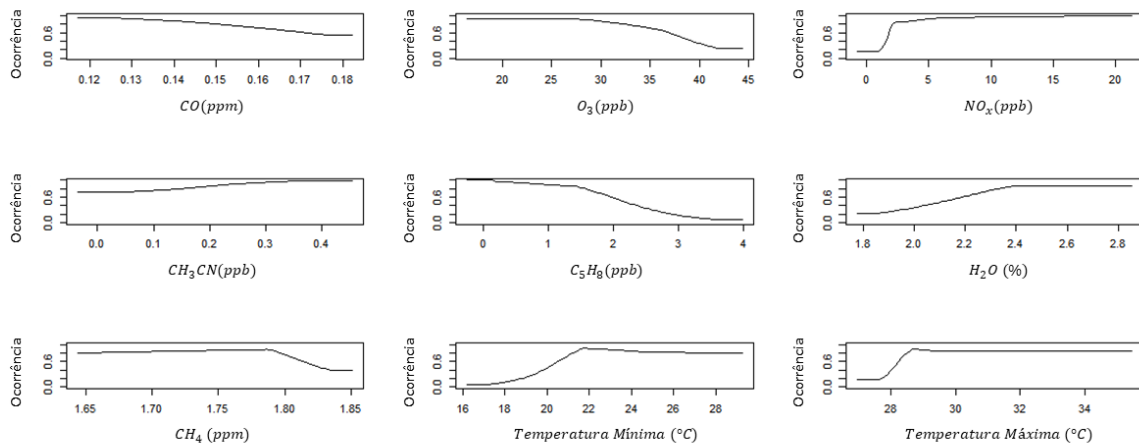


Figura 5 – Curvas de resposta para *Coragyps atratus*

## Conclusões

Com base nos resultados obtidos, nota-se que o objetivo do trabalho foi cumprido. Em relação à interpolação espacial, foi possível coletar e tratar os dados referentes a 10 variáveis ambientais e de aerossóis. Então, desenvolveu-se uma metodologia para a aplicação da interpolação espacial para os dados obtidos e testou-se três diferentes técnicas, selecionando a melhor delas.

Dessa forma, optou-se pelo método linear, obtendo superfícies de interpolação para cada variável, sendo essas utilizadas em conjunto com os dados de ocorrência de espécies, para a construção de um *dataset* que foi utilizado para a aplicação da modelagem de distribuição de espécies, após a etapa de seleção de variáveis preditoras.

Foi selecionada a espécie *Coragyps atratus* para ser analisada. Um modelo foi ajustado, avaliando-se seu desempenho. Além disso, foi possível gerar mapas de distribuição potencial, nos quais foi realizada a previsão da ocorrência das espécies sobre toda a área de dados disponível. Por fim, foram gerados os gráficos de resposta. Dessa forma, pôde-se avaliar a hipótese científica, concluindo-se que as concentrações dos poluentes isopreno e ozônio possuem elevada influência negativa na ocorrência da espécie analisada – resultados praticamente inéditos na literatura.

## Referências bibliográficas

- ARM. Atmospheric Radiation Measurement User Facility. Disponível em: <https://www.arm.gov/>. Acesso em 13 Agosto 2021.
- PINAYA, J.L.D. CORRÊA, P.L.P. Metodologia para Definição das Atividades do Processo de Modelagem de Distribuição de Espécies. XXXIV Congresso da Sociedade Brasileira de Computação. 2014.
- GBIF. Global Biodiversity Information Facility. Disponível em: <https://www.gbif.org/>. Acesso em 13 Agosto 2021.
- HERNANDEZ, P. A., GRAHAM, C. H., MASTER, L. L., ALBERT, D. L. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, v. 29, n. 5, p. 773–785, 2006.
- ICMBIO. Portal da Biodiversidade do Instituto Chico Mendes de Conservação da Biodiversidade. Disponível em: <https://portaldabiodiversidade.icmbio.gov.br/portal/>. Acesso em 13 Agosto 2021.
- MATEO, R. G. VANDERPOORTEN, A. MUÑOZ, J. LAENEN, B. DÉSAMORÉ, A. Modelling species distributions from heterogeneous data for the biogeographic regionalization of the European bryophyte flora. *Plos One*, v. 8, n. 2, 2013.
- MARTIN, S. ET AL. Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon 2014/5). *Atmospheric Chemistry and Physics*, 2016.
- PHILLIPS, S. J. et al. Maximum entropy modeling of species geographic distribution. *Ecological Modelling*, v. 190 (2006), p. 231-259. 2005.