57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

Compreendendo e Prevendo a Incidência de Queimadas na Amazônia: Uma Perspectiva de Séries Temporais

Rodrigo Gonçalves Novais, Igor Peres, Fernando Luiz Cyrino Oliveira, Rodrigo Caiado
Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ - Brasil - 22451-900

RESUMO

A floresta amazônica enfrenta ameaças crescentes de queimadas, impulsionadas por mudanças climáticas e intervenção humana. Este estudo aplica técnicas de séries temporais para prever a incidência de queimadas na região, utilizando dados da Coleção de Queimadas do MapBiomas (1985–2023). Foram desenvolvidos modelos estatísticos como SARIMA, simulações de Monte Carlo, bagging e séries temporais robustas. Estes modelos foram capazes de capturar corretamente os padrões sazonais do fenômeno estudado, apresentando bom desempenho preditivo. O SARIMA identificou padrões sazonais e autorregressivos; Monte Carlo introduziu estocasticidade para avaliação de riscos; e o bagging contribuiu para maior estabilidade das previsões. Modelos robustos apresentaram os menores erros no teste *out-of-sample*. Conclui-se que a integração de variáveis explicativas e técnicas avançadas é fundamental para aprimorar a previsão e apoiar políticas públicas de mitigação e planejamento.

PALAVRAS CHAVE: Amazônia, Queimadas, Previsão.

Tópicos: AS&DS, SIM

ABSTRACT

The Amazon rainforest faces increasing threats from wildfires, driven by climate change and human intervention. This study applies time series techniques to forecast fire incidence in the region using data from the MapBiomas Fire Collection (1985–2023). Statistical models such as SARIMA, Monte Carlo simulations, bagging, and robust time series methods were developed. These models successfully captured the seasonal patterns of the phenomenon, demonstrating good predictive performance. SARIMA identified seasonal and autoregressive structures; Monte Carlo introduced stochasticity for risk assessment; and bagging contributed to greater predictive stability. Robust models yielded the lowest out-of-sample errors. The study concludes that integrating explanatory variables and advanced techniques is essential to improve forecasting and support public mitigation and planning policies.

KEYWORDS: Amazon, Wildfires, Forecast.

Paper topics: AS&DS, SIM

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

1. Introdução

A Amazônia, bioma localizado na América do Sul e o maior do Brasil, é de suma importância para o mundo. Não só abriga milhões de espécies de plantas e animais, sendo um foco rico de biodiversidade, mas também presta serviços ambientais inestimáveis ao planeta. Como é uma densa floresta, armazena e sequestra grandes quantidades de CO2 em sua biomassa, o que contribui para a regulação da concentração do gás na atmosfera. Através da evapotranspiração realizada por sua flora, é responsável pela reciclagem da humidade e padrões de circulação atmosférica (Hadley e Walker), afetando não apenas a região como também sistemas climáticos em latitudes médias e altas (e.g., Europa e Ásia). Além disso, também é fundamental para regular o balanço hídrico da região, o controle de doenças, fornecimento de produtos e materiais e serviços culturais e estéticos. [Foley et al. 2007]

Apesar de seu inestimável valor, enfrenta ameaças persistentes decorrentes do desmatamento e da atividade de queimadas, impulsionadas pela intervenção humana e pelas mudanças climáticas. As queimadas na Amazônia acarretam consequências ecológicas e sociais severas, contribuindo para a perda de biodiversidade, emissões de carbono e impactos em comunidades indígenas e locais. Compreender a dinâmica da incidência de queimadas e prever com precisão sua ocorrência é essencial para a formulação de políticas públicas informadas, alocação de recursos e estratégias eficazes de mitigação [Brando et al. 2020; Lapola et al. 2023].

Mesmo com a significativa redução nas taxas de desmatamento na Amazônia brasileira nas últimas décadas, observou-se um aumento na incidência de incêndios florestais durante anos de seca extrema. Segundo [Aragão et al. 2018], embora tenha havido uma queda de 76% no desmatamento entre 2003 e 2015, o número de incêndios aumentou 36% durante a seca de 2015, atingindo áreas antes pouco afetadas. Esses incêndios, impulsionados por condições climáticas mais secas e pelo uso contínuo de fogo em atividades agropecuárias, demonstram um "desacoplamento" entre desmatamento e fogo, indicando que a redução do desmatamento por si só não é suficiente para conter as emissões de carbono e a degradação florestal associadas aos incêndios. [Alencar et al. 2022] destacam que o bioma amazônico foi o que mais teve seu ciclo natural de fogo afetado pelas ações humanas de mudança de uso do solo. Sua pesquisa ainda indica que, entre 1985 e 2020, o bioma concentrou 41,3% do fogo registrado no país, atrás apenas do bioma do cerrado, onde a vegetação é adaptada a regimes de queima periódicos, o que não ocorre na Amazônia.

Ainda segundo [Aragão et al. 2008], a seca aumenta significativamente o risco de incêndios florestais na Amazônia, pois reduz a umidade, intensifica o acúmulo de material combustível seco e facilita a propagação do fogo, mesmo em áreas com taxas menores de desmatamento. Uma importante variável correlacionada com este fenômeno, a variabilidade da temperatura da superfície dos oceanos Atlântico e Pacífico exerce forte influência sobre o regime de chuvas na Amazônia, modulando a ocorrência de secas severas por meio de conexões atmosféricas que alteram a circulação regional e a entrada de umidade na floresta [Santos et al. 2024].

Uma revisão de literatura recente [Jain et al. 2020] destaca que, do ponto de vista tecnológico, os principais desafios para implementar sistemas de previsão de incêndios incluem a complexidade dos processos físicos, limitação ao acesso de dados de qualidade e a dificuldade de generalização entre regiões. Nesse contexto, técnicas estatísticas e de aprendizado de máquina são preferidas ou usadas em complemento aos modelos baseados em processos naturais, por permitirem análises rápidas e adaptáveis a grandes volumes de dados e não requererem bases detalhadas, mesmo sem descrever explicitamente todos os processos físicos envolvidos na propagação de incêndios.

O objetivo deste trabalho é avaliar e comparar diferentes abordagens estatísticas e robustas para previsão da incidência de queimadas na Amazônia, utilizando dados históricos de cicatrizes de fogo do MapBiomas. A proposta consiste em aplicar, em um mesmo conjunto de dados, modelos amplamente usados na literatura (SARIMA), métodos de incorporação de incerteza (simulações de Monte Carlo), técnicas de agregação de previsões (bagging) e abordagens de otimização robusta

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

(séries temporais estáveis), analisando seus desempenhos in-sample e out-of-sample. Dessa forma, busca-se não apenas identificar a técnica com melhor desempenho preditivo, mas também discutir o potencial e as limitações dessas abordagens para subsidiar políticas de prevenção e mitigação de queimadas.

O restante deste trabalho está organizado da seguinte forma: a Seção 2 apresenta a revisão de literatura e descreve os fundamentos teóricos e metodológicos dos modelos adotados. A Seção 3 detalha os dados utilizados e o procedimento de modelagem. A Seção 4 apresenta os resultados obtidos para cada abordagem, enquanto a Seção 5 discute suas implicações, relacionando-as a estudos prévios e ao contexto das mudanças climáticas. Por fim, a Seção 6 sintetiza as principais conclusões e sugere direções para pesquisas futuras.

2. Marco teórico

Esta seção apresenta uma visão geral dos principais fundamentos teóricos subjacentes aos modelos empregados neste estudo. Os resultados da aplicação de cada método serão analisados detalhadamente na seção de resultados

2.1 Simulações de Monte Carlo e Latin Hypercube sampling

As simulações de Monte Carlo são uma ferramenta poderosa e versátil usada na análise de séries temporais para avaliar a robustez, a acurácia e o comportamento dos modelos sob diferentes cenários. A técnica envolve a geração de múltiplas amostras sintéticas de dados, introduzindo aleatoriedade nos resíduos do modelo. Essas simulações são particularmente efetivas ao trabalhar com processos estocásticos, como séries temporais, pois ajudam a quantificar a incerteza, prever valores futuros e testar o impacto de diferentes suposições sobre o desempenho do modelo [Rubinstein and Kroese 2017].

Após a especificação e predição do modelo, os resíduos (diferença entre valores reais e previstos) são gerados e ajustados a uma distribuição — geralmente aquela que apresenta o menor erro quadrático, embora outros métodos também possam ser aplicados, como os critérios de *Akaike Information Criterion* (AIC) ou *Bayesian Information Criterion* (BIC), ou testes como o de Kolmogorov-Smirnov. Em seguida, erros aleatórios podem ser extraídos dessa distribuição e somados de volta aos valores previstos, assumindo que a correlação temporal dos resíduos foi devidamente tratada durante a estimação do modelo, mimetizando a natureza estocástica dos dados [Rubinstein and Kroese 2017].

Utilizando os parâmetros estimados do modelo e os erros aleatórios gerados, séries temporais sintéticas são criadas iterativamente. Cada série sintética segue a dinâmica definida pelo modelo original, incluindo componentes autorregressivos, de média móvel e sazonais. As previsões são então geradas para cada série simulada, fornecendo uma gama de possíveis desfechos. Os resultados são agregados para produzir previsões probabilísticas, como intervalos de previsão [Rubinstein and Kroese 2017]. Neste trabalho, o conjunto das simulações, calculado como a mediana das simulações para cada tempo, foi obtido e suas métricas de ajuste foram comparadas aos demais modelos.

No que diz respeito ao procedimento de geração de erros, este estudo emprega a Latin Hypercube Sampling – LHS, um método estatístico utilizado para a amostragem eficiente do espaço de parâmetros em simulações numéricas e análises de incerteza. Diferentemente da amostragem aleatória simples, o LHS estratifica a distribuição de probabilidade acumulada de cada variável de entrada em intervalos iguais e garante que cada intervalo seja amostrado exatamente uma vez. Essa abordagem reduz a probabilidade de agrupamentos e assegura uma cobertura mais uniforme do espaço de parâmetros, levando a uma amostragem mais representativa e eficiente. No contexto de séries temporais e modelagem estocástica, o LHS é especialmente útil para reduzir a variância nas simulações de Monte Carlo ao gerar cenários diversos e bem distribuídos. Ao cobrir sistematicamente o espaço de parâmetros, pode fornecer estimativas mais precisas dos resultados do modelo com menos amostras, em comparação com métodos de amostragem aleatória tradicionais [Loh 1996].

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

2.2 Bagging

Bagging (abreviação de Bootstrap Aggregating) é um método de aprendizado em conjunto (ensemble) usado para melhorar a acurácia e estabilidade de modelos, reduzindo a variância e mitigando o sobreajuste (overfitting) [Bergmeir et al. 2016]. A ideia central do bagging é criar múltiplas versões de um conjunto de dados por meio de bootstrapping — amostragem com reposição — e treinar um modelo individual em cada amostra. As previsões desses modelos são então agregadas, tipicamente usando a média/mediana para problemas de regressão para produzir um resultado.

Neste trabalho, o bagging foi combinado com a decomposição STL (Seasonal-Trend decomposition using Loess) para lidar com as complexidades inerentes à série temporal de incidência de queimadas [Aragão et al. 2018; Marengo et al. 2012]. A decomposição STL foi aplicada primeiro para dividir a série em seus componentes principais: tendência, sazonalidade e resíduos.

O procedimento de bagging adaptado seguiu os seguintes passos, conforme adaptado de [Bergmeir et al. 2016] e reproduzido na figura 1:

1. Decomposição: A série foi decomposta usando STL para extrair o componente de resíduos, removendo a influência de tendências de longo prazo e ciclos sazonais.
2. Bootstrap dos Resíduos: A amostragem bootstrap foi aplicada especificamente aos resíduos. Através dessa amostragem, múltiplas variações da série de resíduos foram geradas, capturando a incerteza e a variabilidade potencial.
3. Reconstrução: Para cada amostra bootstrap dos resíduos, a série foi reconstruída somando novamente os componentes de tendência e sazonalidade. Isso produziu múltiplas realizações da série original, cada uma com características de ruído ligeiramente diferentes.
4. Estimação do Modelo: Modelos como o SARIMA ou outros métodos de previsão de séries temporais foram ajustados a cada série reconstruída. A diversidade nas séries reconstruídas permitiu ao conjunto de modelos capturar uma gama mais ampla de padrões potenciais.
5. Agregação: As previsões de todos os modelos foram agregadas para produzir a previsão final. Neste estudo, foi usada a mediana de todas as previsões.

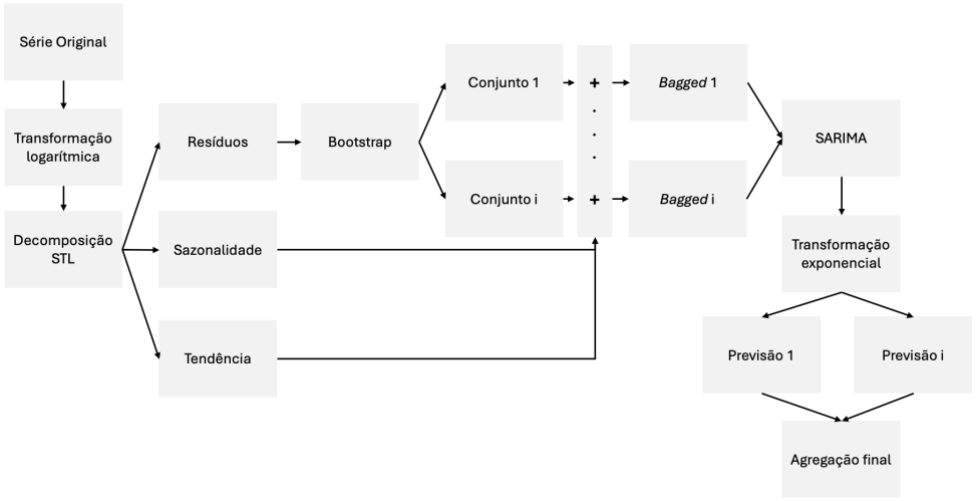


Figura 1 - Fluxograma da solução implementada

2.3 Séries temporais estáveis

[Bertsimas and Paskov 2020] introduzem o conceito de modelos de séries temporais estáveis (*Stable Time Series*), projetados para manter robustez frente a mudanças de regime, ao

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

mesmo tempo em que apresentam bom desempenho durante períodos estacionários. Uma inovação fundamental desta metodologia está na sua integração com os princípios de aprendizado em conjunto (*ensemble learning*). Em vez de treinar múltiplos modelos independentes, múltiplos conjuntos de treinamento são gerados e o modelo é treinado simultaneamente em todos eles.

A metodologia começa com o arcabouço de Séries Temporais Estáveis (STS), que constrói modelos avaliados em diversos conjuntos sintéticos de treinamento. Esses conjuntos são criados para simular diversos cenários potenciais, incluindo períodos com mudanças de regime. O arcabouço de otimização é expresso matematicamente como um problema de minimização–maximização, cujo objetivo é minimizar os erros quadráticos ponderados no pior caso entre os cenários [Bertsimas and Paskov, 2020]. Sua formulação matemática é dada pela equação 1:

$$\min_{\xi,\gamma} \max_{z\in Z} \sum_{t=p+1}^{n} z_t\left(x_t - \gamma - \xi_1 x_{t-1} - \cdots - \xi_p x_{t-p}\right)^2 \quad (1)$$

onde x_t representa a variável de interesse no tempo t, $\xi_1, \ldots, \xi_p$ representam os parâmetros de um modelo autorregressivo a serem estimados para x_t, e γ é sua constante. O termo z_t representa os pesos dos cenários, e $\mathcal{Z}$ é o conjunto de todos os cenários possíveis considerados na otimização.

O conjunto de incerteza $\mathcal{Z}_1$ é definido como o conjunto de todos os subconjuntos de tamanho k, definido na equação 2:

$$\mathcal{Z}1 = \left\{ z: \sum_{i=1}^{n} z_i = k,\ z_i \in \{0,1\},\ i \in [n] \right\} \quad (2)$$

Na solução ótima, cada z_i assume valor 0 ou 1, com a interpretação de que, se $z_i = 1$, a observação (x_i, y_i) é alocada ao conjunto de treinamento; caso contrário, é atribuída ao conjunto de validação. O parâmetro k determina a proporção dos dados que permanecerá no conjunto de treino. Por exemplo, ao definir $k = 0{,}7n$, obtém-se a divisão típica 70/30 e, ao definir $k = 0{,}5n$, obtém-se a divisão 50/50 etc. Valores menores de k resultam em modelos mais conservadores. o parâmetro k não define um corte fixo no início ou no fim da série, mas sim a proporção de observações incluídas no conjunto de treinamento em cada cenário adversarial, que é escolhido de forma a maximizar o erro e, assim, testar o modelo sob as condições mais desafiadoras possíveis.

3. Materiais e métodos

O conjunto de dados utilizado nesta análise foca nas cicatrizes de fogo na região da Amazônia Legal [MapBiomas Brasil 2024], extraído da Coleção de Queimadas do MapBiomas. Se trata de uma série temporal mensal, cobrindo o período de 1985 a 2023, capturando as áreas queimadas em toda a Amazônia Legal. Cada ponto da série representa a área total afetada pelo fogo, medida em hectares, e é derivado de imagens de satélite processadas por técnicas avançadas de sensoriamento remoto. Dados omissos foram substituídos pela média dos mesmos meses em anos adjacentes, tendo em vista a preservação da sazonalidade da variável. A Figura 1 exibe o boxplot mensal dos dados, o qual evidencia uma forte componente sazonal: agosto, setembro e outubro concentram os maiores valores de área queimada na região.

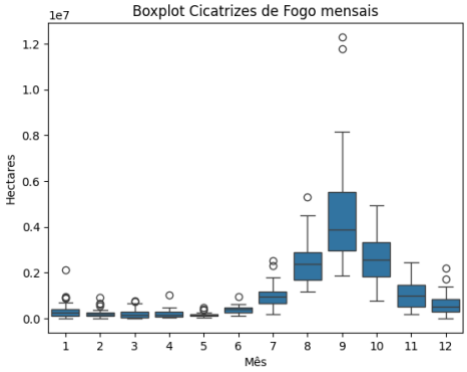


Figura 1 - Boxplot mensal das cicatrizes de queimadas na Amazônia Legal

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

A Figura 2 apresenta a série temporal completa, que destaca os padrões sazonais da série. Notadamente, os valores máximos exibem variabilidade significativa ao longo dos anos, onde alguns, como 2007 e 2010, se sobressaem por sua atividade de fogo substancialmente maior em comparação à média. Segundo [Silva et al. 2023], que também estudou as tendências espaciotemporais do fogo na região, entre 2004-2005 e 2009-2010 a Amazônia sofreu com secas intensas, catalisadas pelas altas temperaturas do oceano Atlântico Norte.

No entanto, as causas climáticas não foram as únicas associadas ao aumento da incidência de incêndios no bioma. Nos últimos sete anos, quatro desses eventos foram registrados (2017, 2019, 2020 e 2022), coincidindo com o alerta de [De Area Leão Pereira et al. 2019], que destacaram que o corte de mais da metade do orçamento do Ministério do Meio Ambiente já em 2017 (mais de 100 milhões de dólares à época) poderia resultar em uma redução da fiscalização e consequente afrouxamento na aplicação das leis ambientais. Ademais, [Silveira et al. 2020] verificaram que as anomalias de fogo registradas na Amazônia brasileira em 2019 foram atribuídas principalmente à combinação de fatores socioeconômicos e institucionais, mais do que a condições climáticas atípicas.

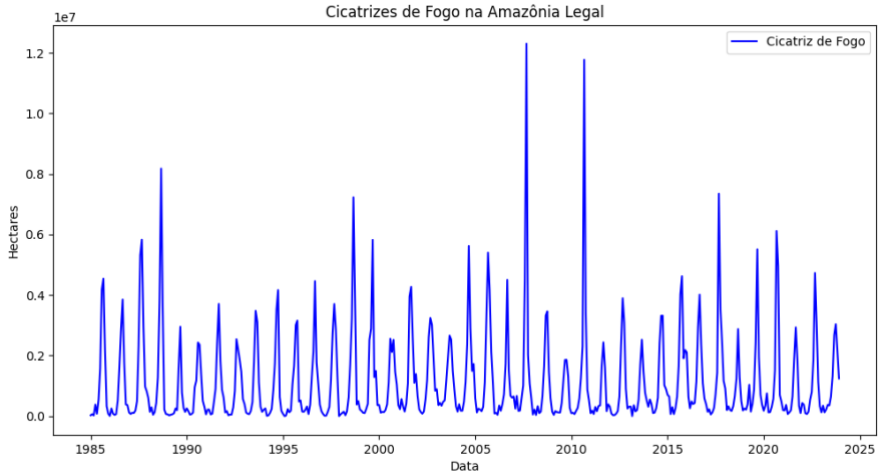


Figura 2 - Série temporal de cicatrizes de queimadas

A metodologia proposta combina modelos estatísticos tradicionais e técnicas robustas para previsão de queimadas na Amazônia, conforme detalhado na Seção 2. O ponto de partida é o modelo SARIMA, utilizado para capturar padrões sazonais e autorregressivos a partir da série histórica. Para lidar com incertezas e mudanças estruturais, incorporamos simulações de Monte Carlo, geradas via Latin Hypercube Sampling, bagging baseado em decomposição STL e reamostragem bootstrap, e o arcabouço de séries temporais estáveis, que otimiza o desempenho sob diferentes cenários de regime. Essa integração busca desenvolver previsões mais robustas.

4.0 Resultados

4.1 Estimação do modelo SARIMA

Inicialmente, a metodologia Box-Jenkins foi aplicada para estimar um modelo SARIMA que tratasse corretamente a autocorrelação serial da variável, ao mesmo tempo em que fosse parcimonioso. A transformação logarítmica foi utilizada tanto para tratar a variância excessiva quanto para restringir os valores previstos ao conjunto de números reais positivos, uma vez que não há razão para uma cicatriz de queimada assumir valor negativo. O modelo foi treinado utilizando o logaritmo da base de dados de 1985 a 2022, enquanto 2023 foi reservado como conjunto de teste para avaliar a acurácia do modelo. Essa escolha permitiu manter uma janela de teste suficientemente curta para que o coeficiente de média móvel estimado (com 12 períodos) ainda fosse efetivo. O modelo resultante foi um $SARIMA(1,0,1)(0,1,1)[12]$, cujas métricas de acurácia *Root Mean Squared Error* (RMSE) e *Mean Absolute Deviation* (MAD) estão descritas na Tabela 1. Ademais, o modelo apresentou p-valor de 0,90 no teste de Ljung–Box, indicando que a autocorrelação serial foi adequadamente tratada.

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

Tabela 1 – Medidas de aderência SARIMA

	In-sample	*Out-of-sample*
RMSE	603,919	614,241
MAD	346,125	430,714

Este modelo é considerado como a baseline deste estudo. Visualmente, o modelo estimado conseguiu capturar as características sazonais e autorregressivas da variável, replicando o comportamento *out-of-sample*, como demonstrado na Figura 3.

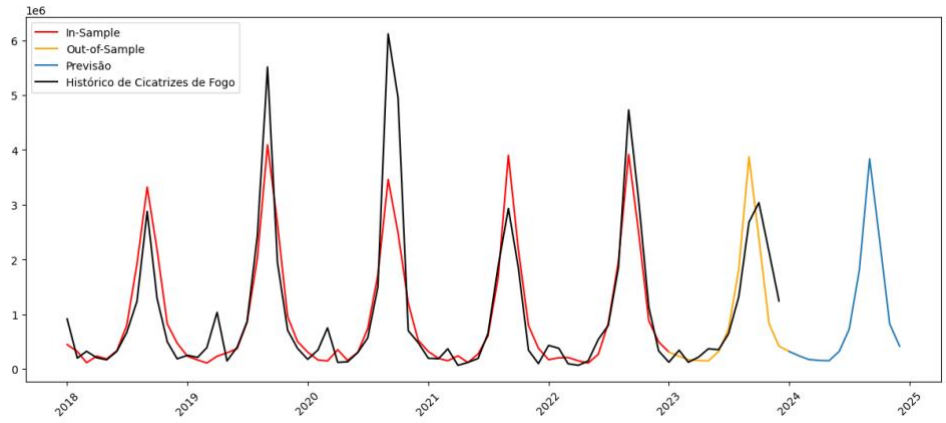


Figura 3 - Previsões do SARIMA vs cicatrizes de queimadas (linha do tempo encurtada para preservar detalhes visuais)

4.2 Simulações de Monte Carlo

Para calcular as simulações de Monte Carlo, um termo de erro aleatório foi somado às previsões do modelo SARIMA estimado. Esse procedimento foi repetido 1000 vezes, representando o número total de simulações, utilizando o método de LHS (Latin Hypercube Sampling). O erro aleatório seguiu a mesma distribuição ajustada aos resíduos do modelo, uma distribuição Weibull Dupla (Double Weibull), que apresentou o menor erro quadrático entre todas as distribuições testadas, conforme mostrado na Tabela 2.

Além do erro quadrático, a Tabela 2 apresenta outras métricas de ajuste que permitem avaliar diferentes aspectos da qualidade da distribuição escolhida. Os valores de AIC e BIC que equilibram qualidade do ajuste e complexidade de parametrização da distribuição, penalizando as que possuem número mais parâmetros estimados. Como os valores foram próximos, priorizamos a escolha da distribuição pelo menor erro quadrático. Já as colunas Est. KS e Valor-p KS correspondem, respectivamente, à estatística e ao valor-p do teste de Kolmogorov–Smirnov, utilizado para avaliar se a distribuição ajustada aos resíduos difere significativamente dos dados observados. Uma estatística KS próxima de zero e valores-p elevados sugerem que não há evidências para rejeitar a hipótese nula de que os resíduos seguem a distribuição especificada, reforçando a adequação do ajuste. Neste estudo, todas as distribuições passaram neste teste e, portanto, poderiam representar corretamente a distribuição do erro.

Como a autocorrelação serial foi corretamente tratada, assumiu-se que os erros são independentes e identicamente distribuídos (i.i.d). O conjunto das simulações apresentou métricas de acurácia piores, tanto *in* quanto *out-of-sample* conforme a Tabela 3.

Tabela 2 – Ajuste das distribuições aos resíduos

Distribuição	Erro quadrático	AIC	BIC	Est. KS	Valor-P KS
Double Weibull	0.35	878.23	890.52	0.03	0.70
Laplace	0.36	862.10	870.29	0.03	0.70
LogLaplace	0.36	866.22	878.51	0.03	0.67
Double Gamma	0.36	875.21	887.50	0.04	0.32
Generalized Normal	0.36	858.05	870.34	0.03	0.68

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

Tabela 3 – Métricas de aderência do conjunto (ensamble) de simulações de Monte Carlo

	In-sample	*Out-of-sample*
RMSE	582,205	656,713
MAD	358,949	448,616

4.3 Simulações de Bagging

Para calcular as simulações com bagging, a série logaritmizada foi transformada em seus componentes de tendência, sazonalidade e resíduos via decomposição STL (*Seasonal-Trend decomposition using Loess*). Os resíduos foram então processados por um procedimento de moving block bootstrap, com tamanho de bloco igual a 24 observações. Isso foi feito 1000 vezes, representando o número total de simulações. Os resíduos reamostrados foram somados aos componentes originais de tendência e sazonalidade, gerando 1000 novas séries temporais. Cada uma dessas séries foi usada para treinar um modelo com a mesma especificação do modelo SARIMA original, mas permitindo que os coeficientes variassem.

As previsões dos modelos treinados foram agrupadas e sua mediana calculada, resultando no conjunto de simulações bagging. As métricas de acurácia *in-sample* foram muito semelhantes ao modelo de referência, mas *out-of-sample* houve melhora significativa, como mostra a Tabela 4:

Tabela 4 – Medidas de aderência do bagging

	In-sample	*Out-of-sample*
RMSE	617,623	596,173
MAD	353,732	397,058

4.4 Séries Temporais Estáveis

Por fim, o modelo de referência também foi comparado a uma abordagem proveniente da literatura de otimização robusta: o modelo de Séries Temporais Estáveis proposto por [Bertsimas and Paskov 2020]. Como especificado anteriormente na seção de metodologia, esse modelo consiste em um problema de min-max. O problema de maximização interno seleciona amostras para aumentar o erro, enquanto o problema de minimização externo busca minimizar a soma dos erros quadráticos. O parâmetro k é um hiperparâmetro que deve ser ajustado de acordo com as necessidades do usuário e representa a proporção de dados usada no conjunto de treinamento do modelo. Neste trabalho, foram testados os valores $k = 0.7$, $k = 0.5$ e $k = 0.3$, onde $k = 1$ representaria a utilização de todas as observações históricas.

Na prática, espera-se que o desempenho *in-sample* seja pior do que no modelo SARIMA original, já que este utiliza todas as observações históricas possíveis para otimizar os erros quadráticos. Contudo, espera-se que o desempenho *out-of-sample* seja melhor, já que os coeficientes estimados tendem a ser mais conservadores e menos sensíveis às mudanças nos padrões históricos. A Tabela 5 mostra que essas expectativas são praticamente confirmadas, com um desempenho *out-of-sample* ainda melhor que o proporcionado pelo bagging.

Tabela 5 – Medidas de aderência do modelo Séries Temporais Estáveis

	k = .7		k = .5		k = .3	
	In-sample	Out-sample	In-sample	Out-sample	In-sample	Out-sample
RMSE	687,177	579,415	698,947	579,183	769,827	620,229
MAD	378,154	393,561	380,104	387,999	389,770	360,249

5. Discussão

Os resultados empíricos obtidos evidenciam diferenças entre as abordagens testadas. O modelo SARIMA de referência apresentou bom desempenho na captura da sazonalidade e da estrutura autorregressiva, com RMSE e MAD estáveis entre os conjuntos *in-sample* e *out-of-sample* (Tabela 1), reforçando sua adequação como *baseline*. As simulações de Monte Carlo, embora úteis para quantificar a incerteza, resultaram em erros ligeiramente superiores, possivelmente devido à propagação adicional de variabilidade associada ao termo estocástico (Tabela 3). Por outro lado, o bagging proporcionou melhora relevante no RMSE fora da amostra, confirmando que a agregação de previsões oriundas de séries reamostradas pode reduzir o impacto

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

de flutuações aleatórias e capturar melhor padrões recorrentes (Tabela 4). Finalmente, o modelo de Séries Temporais Estáveis superou todas as abordagens no desempenho *out-of-sample* para diversos valores de k, indicando maior robustez (Tabela 5).

As técnicas robustas de modelagem aplicadas neste estudo — simulações de Monte Carlo, aprendizado em conjunto via bagging e séries temporais estáveis — fornecem importantes avanços metodológicas para enfrentar mudanças de padrões dentro do contexto histórico. No entanto, todas as metodologias testadas apresentaram níveis de erro relativamente elevados. Um possível caminho para mitigar esta limitação é recomendado por Reichstein et al. 2019, que defendem que modelos híbridos, combinando modelos baseados em processos com técnicas de aprendizado de máquina. Nesse arcabouço, o conhecimento físico incorporado fornece consistência mecanicista e garante interpretabilidade ao modelo, enquanto o aprendizado de máquina explora padrões latentes em dados observacionais, melhorando a capacidade preditiva e reduzindo vieses.

Modelos híbridos são geralmente preferidos quando comparados aos puramente determinísticos, uma vez que estes enfrentam limitações inerentes em cenários práticos. Tais limitações incluem a qualidade insuficiente e a baixa disponibilidade de dados ambientais necessários para parametrizar com precisão esses processos físicos, além da alta demanda computacional associada às simulações mecanicistas detalhadas [Jain et al. 2020]. Como resultado, os campos de ciência e gestão de incêndios florestais dependem fortemente de modelos empíricos e estatísticos para representar processos em escalas meso, sinótica, estratégica e global. A eficácia desses modelos depende da sua capacidade de representar as interações tipicamente complexas entre variáveis relevantes, bem como da qualidade e acessibilidade dos dados subjacentes.

Uma aplicação prática dessa modelagem híbrida é observada em Brando et al. 2020, que acoplaram o modelo de carbono e vegetação a um módulo probabilístico de fogo, calibrado com dados observacionais, para simular de forma integrada o efeito combinado de clima, uso da terra e dinâmica da vegetação sobre a ocorrência de incêndios. De forma semelhante, Le Page et al. 2017 também utilizaram um modelo de natureza híbrida, para explorar as sinergias entre mudanças climáticas e expansão agrícola na Amazônia, incorporando dados de satélite para calibrar a dinâmica de fogo de sub-bosque. Embora não tenham empregado aprendizado de máquina, esses trabalhos ilustram como a combinação de modelos físicos e estatísticos pode aumentar a robustez das projeções e apoiar políticas de mitigação mais eficazes — um caminho promissor para o avanço da modelagem de queimadas na Amazônia.

Ademais, modelos estatísticos como o SARIMA, utilizado neste estudo, têm demonstrado boa capacidade de capturar padrões sazonais de queimadas. No entanto, sua capacidade preditiva pode ser significativamente ampliada pela incorporação de índices climáticos exógenos, como o Índice Oceânico de El Niño (ONI) e a Oscilação Multidecadal do Atlântico (AMO). A relevância desses índices é bem estabelecida na literatura: por exemplo, Chen et al., 2011 demonstraram associações claras entre anomalias na temperatura da superfície do mar (SSTs) e temporadas severas de queimadas na América do Sul. Da mesma forma, Dos Reis et al., 2021 destacaram a forte ligação entre eventos extremos de seca relacionados a anomalias do ENSO e ocorrências de queimadas na Amazônia central, reforçando ainda mais a necessidade de incluir tais variáveis climáticas. Além disso, achados recentes de Zanchettin and Rubino, 2024 indicam que tendências de aquecimento acelerado podem atenuar significativamente a variabilidade multidecadal do Atlântico, fazendo com que a dinâmica de temperatura do oceano se mantenha mais quente por períodos prolongados. Esse cenário climático emergente representa um desafio crucial para modelos que se baseiam em relações climáticas históricas, pois essas relações podem se alterar no futuro próximo.

Dessa forma, as mudanças climáticas representam desafios significativos para modelos de incêndio que não consideram cenários de aquecimento futuro. A frequência e intensidade de eventos de fogo induzidos por secas estão aumentando, alterando os regimes históricos do ecossistema amazônico. Por exemplo, [Aragão et al. 2018] demonstraram que as queimadas associadas à seca compensam substancialmente a redução nas emissões de carbono obtida pela queda nas taxas de desmatamento. Da mesma maneira, a crescente frequência e severidade de secas

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

plurianuais [Chen et al. 2025] ameaçam a estabilidade de estruturas preditivas construídas exclusivamente com base em regimes de fogo históricos. Esse novo cenário reforça a urgência de incorporar dinâmicas climáticas projetadas nos esforços de modelagem de queimadas, especialmente considerando pontos de inflexão potenciais, como a "savanização" da Amazônia [Nobre et al. 2016]. Nesse contexto, a realização de simulações sob diferentes hipóteses para a incidência de queimadas, considerando projeções alternativas para ONI, AMO e outras variáveis explicativas, como intensidade de seca, constitui um caminho promissor para antecipar riscos e apoiar políticas de mitigação mais adaptativas.

As estratégias de validação são fundamentais para o desenvolvimento e aplicação robusta de modelos. A literatura recente oferece bons exemplos de metodologias de validação diversificadas. Brando et al., 2020 validaram seu modelo acoplado por meio de extensas comparações com cicatrizes de queimadas observadas e focos de ignição detectados (R^2: 0,83–0,88; métricas de RMSE). Le Page et al., 2017 validaram seu modelo utilizando áreas queimadas observadas, verificadas por imagens de satélite. Silvestrini et al. 2011 empregaram um modelo probabilístico validado por meio de hot pixels detectados por satélites da *National Oceanic and Atmospheric Administration* (NOAA), obtendo sobreposição espacial superior a 50% em raio de 11 km e alta acurácia temporal. Fonseca et al., 2019 calibraram seu modelo MaxEnt com base em conjuntos históricos de dados de queimadas, garantindo consistência e confiabilidade. Neste estudo, nosso modelo foi validado por meio da acurácia de previsões *out-of-sample*, com foco especial em RMSE e MAD para avaliar o desempenho preditivo.

Apesar dos avanços obtidos, permanecem limitações importantes. O modelo proposto oferece previsões em nível de bioma, o que restringe sua utilidade para ações locais de prevenção e resposta, geralmente baseadas em avaliações de risco mais detalhadas espacialmente. Além disso, não incorpora projeções de mudanças climáticas futuras, como adotado por alguns estudos recentes, o que limita sua aplicabilidade em planejamento de longo prazo. Pesquisas futuras devem priorizar a integração de cenários climáticos atualizados do IPCC e de modelos regionais, aliados a estruturas híbridas que combinem processos físicos, estatística e aprendizado de máquina. A inclusão de variáveis socioeconômicas e geográficas também pode ampliar substancialmente a capacidade preditiva e a relevância prática de modelos de previsão de queimadas na Amazônia.

6. Conclusão

Os resultados deste estudo destacam as complexidades envolvidas na previsão da incidência de queimadas na Amazônia e a necessidade de empregar diferentes técnicas de modelagem para capturar a dinâmica desse fenômeno. Embora o modelo SARIMA tenha servido como baseline confiável para lidar com sazonalidade e autorregressão, métodos avançados como simulações de Monte Carlo, bagging e modelos de séries temporais estáveis ofereceram alternativas para lidar com incertezas e mudanças estruturais nos dados.

Em termos de performance preditiva, os modelos bagging e de séries temporais estáveis apresentaram os menores valores de RMSE e MAD *out-of-sample*. O modelo desenvolvido foi capaz de capturar corretamente os padrões sazonais do fenômeno estudado, apresentando um bom nível de desempenho preditivo. No entanto, o alto erro de previsão geral evidencia os desafios inerentes à modelagem de um sistema tão volátil e multifacetado.

Esses resultados sugerem que pesquisas futuras devem focar na integração de variáveis explicativas adicionais, como fatores climáticos (intensidade da seca, temperaturas oceânicas, cenários de aquecimento global) e socioeconômicos (fiscalização ambiental, mudanças legislativas, desenvolvimento social), para melhorar o desempenho dos modelos. Além disso, a exploração mais aprofundada de conjuntos de incerteza como $\mathcal{Z}_2$, conforme proposto no arcabouço de séries temporais estáveis, pode oferecer um caminho para alcançar um maior equilíbrio entre estabilidade e acurácia. Outras técnicas também podem ser consideradas, como suavização exponencial ETS (Error, Trend, Seasonality) e abordagens de aprendizado de máquina que utilizem métodos em conjunto para capturar relações não lineares e interações complexas nos dados, aprimorando ainda mais a acurácia e robustez preditiva.

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

Ao refinar as metodologias e incorporar fatores externos, esta linha de pesquisa pode contribuir para ferramentas de previsão mais confiáveis, auxiliando tomadores de decisão na gestão dos impactos ambientais e socioeconômicos causados por queimadas na Amazônia.

Referências

Alencar, A. A. C., Arruda, V. L. S., Silva, W. V. Da, et al. (jan 2022). Long-Term Landsat-Based Monthly Burned Area Dataset for the Brazilian Biomes Using Deep Learning. *Remote Sensing*, v. 14, n. 11, p. 2510.

Aragão, L. E. O. C., Anderson, L. O., Fonseca, M. G., et al. (13 feb 2018). 21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions. *Nature Communications*, v. 9, n. 1, p. 536.

Aragão, L. E. O. C., Malhi, Y., Barbier, N., et al. (27 may 2008). Interactions between rainfall, deforestation and fires during recent years in the Brazilian Amazonia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 363, n. 1498, p. 1779–1785.

Bergmeir, C., Hyndman, R. J. and Benítez, J. M. (apr 2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, v. 32, n. 2, p. 303–312.

Bertsimas, D. and Paskov, I. (nov 2020). Time Series that are Robust to Regime Changes.

Brando, P. M., Soares-Filho, B., Rodrigues, L., et al. (10 jan 2020). The gathering firestorm in southern Amazonia. *Science Advances*, v. 6, n. 2, p. eaay1632.

Chen, L., Brun, P., Buri, P., et al. (17 jan 2025). Global increase in the occurrence and impact of multiyear droughts. *Science*, v. 387, n. 6731, p. 278–284.

Chen, Y., Randerson, J. T., Morton, D. C., et al. (11 nov 2011). Forecasting Fire Season Severity in South America Using Sea Surface Temperature Anomalies. *Science*, v. 334, n. 6057, p. 787–791.

De Area Leão Pereira, E. J., Silveira Ferreira, P. J., De Santana Ribeiro, L. C., Sabadini Carvalho, T. and De Barros Pereira, H. B. (1 oct 2019). Policy in Brazil (2016–2019) threaten conservation of the Amazon rainforest. *Environmental Science & Policy*, v. 100, p. 8–12.

Dos Reis, M., Graça, P. M. L. D. A., Yanai, A. M., Ramos, C. J. P. and Fearnside, P. M. (jun 2021). Forest fires and deforestation in the central Amazon: Effects of landscape and climate on spatial and temporal dynamics. *Journal of Environmental Management*, v. 288, p. 112310.

Foley, J. A., Asner, G. P., Costa, M. H., et al. (2007). Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. *Frontiers in Ecology and the Environment*, v. 5, n. 1, p. 25–32.

Fonseca, M. G., Alves, L. M., Aguiar, A. P. D., et al. (2019). Effects of climate and land-use change scenarios on fire probability during the 21st century in the Brazilian Amazon. *Global Change Biology*, v. 25, n. 9, p. 2931–2946.

Jain, P., Coogan, S. C. P., Subramanian, S. G., et al. (dec 2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, v. 28, n. 4, p. 478–505.

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430

Lapola, D. M., Pinho, P., Barlow, J., et al. (27 jan 2023). The drivers and impacts of Amazon forest degradation. *Science*, v. 379, n. 6630, p. eabp8622.

Le Page, Y., Morton, D., Hartin, C., et al. (20 dec 2017). Synergy between land use and climate change increases future fire risk in Amazon forests. *Earth System Dynamics*, v. 8, n. 4, p. 1237–1246.

Loh, W.-L. (1996). ON LATIN HYPERCUBE SAMPLING. *The Annals of Statistics*, v. 24, n. 5, p. 2058–2080.

MapBiomas Brasil (2 dec 2024). . https://brasil.mapbiomas.org/metodo-mapbiomas-fogo/, [accessed on Dec 2].

Marengo, J. A., Tomasella, J., Soares, W. R., Alves, L. M. and Nobre, C. A. (1 jan 2012). Extreme climatic events in the Amazon basin. *Theoretical and Applied Climatology*, v. 107, n. 1, p. 73–85.

Nobre, C. A., Sampaio, G., Borma, L. S., et al. (27 sep 2016). Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proceedings of the National Academy of Sciences*, v. 113, n. 39, p. 10759–10768.

Reichstein, M., Camps-Valls, G., Stevens, B., et al. (feb 2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, v. 566, n. 7743, p. 195–204.

Rubinstein, R. Y. and Kroese, D. P. (2017). *Simulation and the Monte Carlo method*. Third edition ed. Hoboken, New Jersey: Wiley.

Santos, C. A. G., Santos, D. C. Dos, Brasil Neto, R. M., et al. (1 sep 2024). Analyzing the impact of ocean-atmosphere teleconnections on rainfall variability in the Brazilian Legal Amazon via the Rainfall Anomaly Index (RAI). *Atmospheric Research*, v. 307, p. 107483.

Silva, R. M. D., Lopes, A. G. and Santos, C. A. G. (jan 2023). Deforestation and fires in the Brazilian Amazon from 2001 to 2020: Impacts on rainfall variability and land surface temperature. *Journal of Environmental Management*, v. 326, p. 116664.

Silveira, M. V. F., Petri, C. A., Broggio, I. S., et al. (14 dec 2020). Drivers of Fire Anomalies in the Brazilian Amazon: Lessons Learned from the 2019 Fire Crisis. *Land*, v. 9, n. 12, p. 516.

Silvestrini, R. A., Soares-Filho, B. S., Nepstad, D., et al. (2011). Simulating fire regimes in the Amazon in response to climate change and deforestation. *Ecological Applications*, v. 21, n. 5, p. 1573–1590.

Zanchettin, D. and Rubino, A. (29 oct 2024). Accelerated North Atlantic surface warming reshapes the Atlantic Multidecadal Variability. *Communications Earth & Environment*, v. 5, n. 1, p. 639.

DOI: 10.59254/sbpo-2025-212430
http://live-proceedings.pantheonsite.io/p/212430