

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

VINICIUS LOVATEL ROCHA

**ANÁLISE E PREVISÃO DE INCÊNDIOS FLORESTAIS PARA
REGIÃO DE LAGES E XANXERÊ - SANTA CATARINA: UMA
ABORDAGEM EXPLORANDO MODELOS HÍBRIDOS DE PREVISÃO**

PATO BRANCO

2025

VINICIUS LOVATEL ROCHA   

**ANÁLISE E PREVISÃO DE INCÊNDIOS FLORESTAIS PARA
REGIÃO DE LAGES E XANXERÊ - SANTA CATARINA: UMA
ABORDAGEM EXPLORANDO MODELOS HÍBRIDOS DE PREVISÃO**

**Forest fire analysis and forecasting for the Lages and Xanxerê region
- Santa Catarina: an approach exploring hybrid forecasting models**

Dissertação apresentada como requisito para obtenção do título de Mestre em Engenharia de Produção e Sistemas pela Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador:

► Dr. Matheus Henrique Dal Molin Ribeiro   

PATO BRANCO

2025



Esta Dissertação está licenciada sob Creative Commons Atribuição–NãoComercial–SemDerivações 4.0 Internacional, que permite uso e distribuição em qualquer meio ou formato, desde que o trabalho original seja devidamente citado, o uso não seja comercial e nenhuma modificação ou adaptação seja feita.



**Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Pato Branco**



VINICIUS LOVATEL ROCHA

ANÁLISE E PREVISÃO DE INCÊNDIOS FLORESTAIS PARA REGIÃO DE LAGES E XANXERÊ - SANTA CATARINA: UMA ABORDAGEM EXPLORANDO MODELOS HÍBRIDOS DE PREVISÃO

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Engenharia De Produção E Sistemas da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Gestão Dos Sistemas Produtivos.

Data de aprovação: 29 de Agosto de 2025

Dr. Matheus Henrique Dal Molin Ribeiro, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Marco Antonio De Castro Barbosa, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Victor Henrique Alves Ribeiro, Doutorado - Pontifícia Universidade Católica do Paraná (Pucpr)

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 22/09/2025.

RESUMO

Anualmente, incêndios florestais impactam extensas áreas do estado de Santa Catarina, gerando prejuízos significativos ao meio ambiente, à economia e a patrimônios públicos e privados. Este estudo tem como objetivo prever a ocorrência de focos de incêndio nas regiões de Lages e Xanxerê, que concentram um elevado número de registros no estado. Com base nos dados operacionais fornecidos pelo Corpo de Bombeiros Militar de Santa Catarina (CBMSC), a pesquisa visa apoiar uma gestão mais eficiente dos recursos humanos e materiais, permitindo uma alocação otimizada do efetivo durante os períodos de maior risco. O estudo está estruturado em duas etapas. A primeira consiste na condução de uma revisão sistemática da literatura, com o propósito de identificar os principais métodos aplicados, os resultados obtidos, as limitações observadas e as lacunas de pesquisa no campo da previsão de incêndios florestais. Na segunda etapa, será realizado um estudo empírico voltado à análise e previsão dos focos de incêndio na região de Lages e Xanxerê, utilizando modelos univariados e multivariados, associados à técnica de decomposição de séries temporais denominada Variational Mode Decomposition (VMD). A avaliação dos modelos é realizada por meio de medidas de desempenho como a Raiz Quadrada do Erro Médio (RMSE), o Erro Absoluto Médio (MAE), o Erro Percentual Absoluto Médio Simétrico (sMAPE) e o Desvio Padrão (SD). A análise considera variáveis climáticas como precipitação, temperatura, umidade relativa, velocidade do vento e a ocorrência do fenômeno La Niña. Os modelos multivariados apresentaram desempenho superior quando aplicados sem o VMD. Em Lages, o melhor resultado ocorreu para horizonte $H = 2$ com dados transformados, sendo o Extreme Learning Machine (ELM) o mais eficaz (RMSE = 3,1084; MAE = 2,1924; sMAPE = 61,96%; DP = 2,8475). Já em Xanxerê, também para $H = 2$, o XGBoost sem transformação obteve maior precisão (RMSE = 4,4526; MAE = 2,9776; sMAPE = 73,43 %; SD = 4,5680). Por outro lado, a modelagem univariada combinando VMD e algoritmos de regressão se mostrou mais robusta para previsão mensal ($H = 1$). Em Lages, a melhor performance foi alcançada com dados transformados pela raiz quadrada, destacando-se o VMD-LASSO (RMSE = 1,2074; MAE = 0,9675; sMAPE = 58,90%; SD = 1,2401). Em Xanxerê, o modelo VMD-LASSO aplicado aos dados na forma original foi novamente o mais eficiente (RMSE = 1,5248; MAE = 1,3597; sMAPE = 67,79%; DP = 1,4683). Espera-se que este estudo sirva de ferramenta de apoio à tomada de decisão do CBMSC, contribuindo para o planejamento operacional e para a mobilização de efetivos em períodos críticos, fortalecendo as estratégias de prevenção e resposta a incêndios florestais.

Palavras-chave: Incêndios florestais, Previsão de Focos de Incêndio, Aprendizado de Máquina, Corpo de Bombeiros Militar de Santa Catarina.

ABSTRACT

Wildfires annually impact vast areas of the state of Santa Catarina, causing significant damage to the environment, economy, and both public and private property. This study aims to forecast wildfire occurrences in the regions of Lages and Xanxerê, which account for a high number of recorded incidents in the state. Based on operational data provided by the Military Fire Department of Santa Catarina (CBMSC), the research seeks to support more efficient management of human and material resources, enabling optimized personnel allocation during periods of increased risk. The study is structured in two stages. The first involves conducting a systematic literature review to identify the main forecasting methods applied, the results obtained, observed limitations, and research gaps in the field of wildfire prediction. The second stage comprises an empirical study focused on the analysis and forecasting of wildfire outbreaks in the Lages and Xanxerê regions, employing both univariate and multivariate models, combined with the time series decomposition technique known as Variational Mode Decomposition (VMD). Model performance is assessed using evaluation metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (sMAPE), and Standard Deviation (SD). The analysis considers climatic variables such as precipitation, temperature, relative humidity, wind speed, and the occurrence of the La Niña phenomenon. Multivariate models showed better performance when applied without VMD. In Lages, the best result was achieved for forecasting horizon $H = 2$ using transformed data, with the ELM model proving most effective (RMSE = 3.1084; MAE = 2.1924; sMAPE = 61.96%; SD = 2.8475). In Xanxerê, also for $H = 2$, the XGBoost model without transformation was the most accurate (RMSE = 4.4526; MAE = 2.9776; sMAPE = 73.43%; SD = 4.5680). On the other hand, the univariate modeling approach, combining VMD with regression algorithms, proved more robust for monthly forecasting ($H = 1$). In Lages, the best performance was obtained using square root-transformed data, with the VMD-LASSO model standing out (RMSE = 1.2074; MAE = 0.9675; sMAPE = 58.90%; SD = 1.2401). In Xanxerê, the best results were also obtained with original (non-transformed) data, again highlighting the VMD-LASSO model (RMSE = 1.5248; MAE = 1.3597; sMAPE = 67.79%; SD = 1.4683). This study is expected to serve as a decision-support tool for CBMSC, contributing to operational planning and personnel mobilization during critical periods, thereby strengthening wildfire prevention and response strategies.

Keywords: Forest Fires, Fire Outbreak Prediction, Machine Learning, Santa Catarina Military Fire Brigade .

LISTA DE ILUSTRAÇÕES

Figura 1 – Localização da cidade de Lages no contexto geográfico do Brasil e do estado de Santa Catarina.	13
Figura 2 – Localização da cidade de Xanxerê no contexto geográfico do Brasil e do estado de Santa Catarina.	14
Figura 3 – Estrutura do estudo envolvendo os artigos trabalhados.	20
Figura 4 – Processo de composição das palavras-chave com os respectivos operadores booleanos.	24
Figura 5 – Diagrama de seleção de estudos.	25
Figura 6 – Distribuição anual das publicações.	27
Figura 7 – Gráfico da distribuição das publicações por periódico.	28
Figura 8 – Análise de redes das palavras-chave.	29
Figura 9 – As 10 publicações mais relevantes.	31
Figura 10 – Série temporal de ocorrências mensais de incêndios florestais em Lages (SC).	42
Figura 11 – Teste de anomalias no conjunto de dados da série temporal de Lages (SC).	42
Figura 12 – Série temporal de ocorrências mensais de incêndios florestais em Xanxerê (SC).	42
Figura 13 – Teste de anomalias no conjunto de dados da série temporal de Xanxerê (SC).	43
Figura 14 – Funções de Autocorrelação e autocorrelação parcial da série temporal de Lages.	43
Figura 15 – Funções de Autocorrelação e autocorrelação parcial da série temporal de Xanxerê.	43
Figura 16 – Framework do estudo.	61
Figura 17 – Fluxo do procedimento metodológico proposto.	62
Figura 18 – Gráfico do modelo VMD-LASSO com dados em Sqrt para H=1.	66
Figura 19 – Gráfico do modelo VMD-LASSO com dados em Sqrt para H=2.	69
Figura 20 – Gráfico do modelo VMD-LASSO com dados em Sqrt para H=3.	71
Figura 21 – Gráfico do modelo VMD-LASSO com dados normais para H=1.	74
Figura 22 – Gráfico do modelo VMD-LASSO com dados normais para H=2.	77
Figura 23 – Gráfico do modelo VMD-LASSO com dados normais para H=3.	79
Figura 24 – Valores Shapley variáveis do modelo multivariado para cidade de Lages.	82
Figura 25 – Valores Shapley variáveis do modelo multivariado para cidade de Xanxerê.	83
Figura 26 – Desempenho comparativo dos melhores modelos para a cidade de Lages, considerando diferentes horizontes de previsão e abordagens (Univariada, Multivariada e híbrida com VMD).	84
Figura 27 – Desempenho comparativo dos melhores modelos para a cidade de Xanxerê, considerando diferentes horizontes de previsão e abordagens (Univariada, Multivariada e híbrida com VMD).	85
Figura 28 – Autorização dos coautores para a inclusão do artigo " <i>Wildfire Spots Analysis and Forecasting: Evaluation of Univariate and Multivariate Based on Variational Mode Decomposition Models</i> " na dissertação.	101

LISTA DE TABELAS

Tabela 1 – Critérios de inclusão dos estudos	25
Tabela 2 – Visão geral sistemática das principais questões de pesquisa.	27
Tabela 3 – Resumo de trabalhos relacionados a previsão de focos de incêndio sob diferentes perspectivas.	40
Tabela 4 – Desempenho de modelo único para previsão de um mês à frente.	64
Tabela 5 – Desempenho de modelos híbridos para previsão de um mês à frente.	65
Tabela 6 – Desempenho de modelo único para previsão de dois meses à frente.	67
Tabela 7 – Desempenho de modelos híbridos para previsão de dois meses à frente.	68
Tabela 8 – Desempenho de modelo único para previsão de três meses à frente.	70
Tabela 9 – Desempenho de modelos híbridos para previsão de três meses à frente.	71
Tabela 10 – Desempenho de modelo único para previsão de um mês à frente.	73
Tabela 11 – Desempenho de modelos híbridos para previsão de um mês à frente.	74
Tabela 12 – Desempenho de modelo único para previsão de dois meses à frente.	75
Tabela 13 – Desempenho de modelos híbridos para previsão de dois meses à frente.	76
Tabela 14 – Desempenho de modelo único para previsão de três meses à frente.	78
Tabela 15 – Desempenho de modelos híbridos para previsão de três meses à frente.	79

LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

SIGLAS

ACF	Função de Autocorrelação (do inglês, <i>Autocorrelation Function</i>)
AMO	Oscilação Multidecadal do Atlântico (do inglês, <i>Atlantic Multidecadal Oscillation</i>)
ANN	Rede Neural Artificial (do inglês, <i>Artificial Neural Network</i>)
ARIMA	Média Móvel Integrada AutoRegressiva (do inglês, <i>AutoRegressive Integrated Moving Average</i>)
ARIMAX	ARIMA com variáveis exógenas (do inglês, <i>ARIMA with exogenous variables</i>)
BPNN	Rede Neural com Retropropagação (do inglês, <i>Backpropagation Neural Network</i>)
BRT	Árvores de Regressão com Boosting (do inglês, <i>Boosted Regression Trees</i>)
CBMSC	Corpo de Bombeiros Militar de Santa Catarina
CNN	Rede Neural Convolucional (do inglês, <i>Convolutional Neural Network</i>)
CV	Validação Cruzada (do inglês, <i>Cross-Validation</i>)
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DL	Aprendizado Profundo (do inglês, <i>Deep Learning</i>)
ELM	Máquina de Aprendizado Extremo (do inglês, <i>Extreme Learning Machine</i>)
EMD	Decomposição em Modo Empírico (do inglês, <i>Empirical Mode Decomposition</i>)
EPAGRI	Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina
ETS	Erro-Tendência-Sazonalidade (do inglês, <i>Error-Trend-Seasonality</i>)
FDA	Avaliação de Perigo de Incêndio (do inglês, <i>Fire Danger Assessment</i>)
FSS	Índice de Habilidade Fracionária (do inglês, <i>Fraction Skill Score</i>)
FWI	Índice Meteorológico de Incêndio Florestal (do inglês, <i>Fire Weather Index</i>)
FR-CART	Risco de Incêndio – Árvores de Classificação e Regressão (do inglês, <i>Fire Risk - Classification and Regression Trees</i>)
FR-LR	Risco de Incêndio – Regressão Logística (do inglês, <i>Fire Risk - Logistic Regression</i>)
FR-MLP	Risco de Incêndio – Perceptron Multicamadas (do inglês, <i>Fire Risk - Multi-Layer Perceptron</i>)
FR-RF	Risco de Incêndio – Floresta Aleatória (do inglês, <i>Fire Risk - Random Forest</i>)
FR-SVM	Risco de Incêndio – Máquina de Vetores de Suporte (do inglês, <i>Fire Risk - Support Vector Machine</i>)
GBDT	Árvore de Decisão com Boosting por Gradiente (do inglês, <i>Gradient Boosting Decision Tree</i>)
GLM	Modelo Linear Generalizado (do inglês, <i>Generalized Linear Model</i>)
GLMNET	Modelo Linear Generalizado com Regularização Elastic Net (do inglês, <i>Generalized Linear Model via Elastic Net Regularization</i>)

GRU	Unidade Recorrente com Portões (do inglês, <i>Gated Recurrent Unit</i>)
GWO	Otimizador do Lobo Cinzento (do inglês, <i>Grey Wolf Optimizer</i>)
IA	Inteligência Artificial
IMFs	Funções de Modo Intrínseco (do inglês, <i>Intrinsic Mode Functions</i>)
INPE	Instituto Nacional de Pesquisas Espaciais
KPSS	Teste de Kwiatkowski-Phillips-Schmidt-Shin (do inglês, <i>Kwiatkowski-Phillips-Schmidt-Shin test</i>)
LASSO	Menor Operador Absoluto de Encolhimento e Seleção (do inglês, <i>Least Absolute Shrinkage and Selection Operator</i>)
LSTM	Memória de Longo e Curto Prazo (do inglês, <i>Long Short-Term Memory</i>)
LR	Regressão Logística (do inglês, <i>Logistic Regression</i>)
MAE	Erro Absoluto Médio (do inglês, <i>Mean Absolute Error</i>)
MA-FC	Previsão por Média Móvel (do inglês, <i>Moving Average Forecasting</i>)
MARS	Splines de Regressão Adaptativa Multivariada (do inglês, <i>Multivariate Adaptive Regression Splines</i>)
ML	Aprendizado de Máquina (do inglês, <i>Machine Learning</i>)
MLP	Perceptron Multicamadas (do inglês, <i>Multi-Layer Perceptron</i>)
MLR	Regressão Linear Múltipla (do inglês, <i>Multiple Linear Regression</i>)
NDMI	Índice de Umidade da Diferença Normalizada (do inglês, <i>Normalized Difference Moisture Index</i>)
NDVI	Índice de Vegetação da Diferença Normalizada (do inglês, <i>Normalized Difference Vegetation Index</i>)
OG	Objetivo Geral
ONI	Índice Oceânico de El Niño (do inglês, <i>Oceanic Niño Index</i>)
PACF	Função de Autocorrelação Parcial (do inglês, <i>Partial Autocorrelation Function</i>)
PB	Portfólio Bibliográfico
PS	<i>PostScript</i>
PSO	Otimização por Enxame de Partículas (do inglês, <i>Particle Swarm Optimization</i>)
QP	Questão de Pesquisa
RF	Floresta Aleatória (do inglês, <i>Random Forest</i>)
RFt	Floresta Aleatória para Séries Temporais (do inglês, <i>Random Forest for time series</i>)
RMSE	Raiz do Erro Quadrático Médio (do inglês, <i>Root Mean Squared Error</i>)
RSL	Revisão Sistemática de Literatura
SAVI	Índice de Vegetação Ajustado ao Solo (do inglês, <i>Soil-Adjusted Vegetation Index</i>)
SD	Desvio Padrão (do inglês, <i>Standard Deviation</i>)
sMAPE	Erro Percentual Absoluto Médio Simétrico (do inglês, <i>Symmetric Mean Absolute Percentage Error</i>)
STLF	Previsão usando decomposição Sazonal e Tendência por Loess (do inglês, <i>Seasonal and Trend decomposition using Loess Forecasting</i>)

SLFN	Rede Alimentada Direta de Camada Única (do inglês, <i>Single-Layer Feedforward Network</i>)
SVR	Máquina de Vetores de Suporte para Regressão (do inglês, <i>Support Vector Regression</i>)
SVM	Máquina de Vetores de Suporte (do inglês, <i>Support Vector Machine</i>)
TBATS	Modelo trigonométrico Box-Cox com erros ARMA, tendência e componentes sazonais (do inglês, <i>Trigonometric, Box-Cox, ARMA errors, Trend and Seasonal components</i>)
TST	Temperatura Superficial da Terra (do inglês, <i>Thermal Surface Temperature</i>)
TSGLM	Modelo Linear Generalizado para Séries Temporais (do inglês, <i>Time Series Generalized Linear Model</i>)
UTFPR	Universidade Tecnológica Federal do Paraná
VMD	Decomposição Variacional por Modos (do inglês, <i>Variational Mode Decomposition</i>)
XGBoost	<i>eXtreme Gradient Boosting</i>

SUMÁRIO

1	INTRODUÇÃO	11
1.1	PROBLEMA DE PESQUISA	11
1.2	OBJETIVOS DE PESQUISA	14
1.3	JUSTIFICATIVA E RELEVÂNCIA DA PESQUISA	16
1.4	ETAPAS DA PESQUISA	19
1.5	LISTA DE PUBLICAÇÕES	20
2	DESAFIOS NA PREVISÃO DE INCÊNDIOS FLORESTAIS - UMA REVISÃO DA LITERATURA	22
2.1	INTRODUÇÃO	22
2.2	METODOLOGIA	23
2.3	SELEÇÃO DO PORTFÓLIO BIBLIOGRÁFICO	24
2.3.1	Crítérios de Qualidade e Seleção dos Estudos	24
2.3.2	Análise Bibliométrica	26
2.3.3	Análise Sistemica	26
2.4	APRESENTAÇÃO DOS RESULTADOS	26
2.4.1	Quais são as variáveis empregadas na modelagem de previsão de focos de incêndio e o horizonte de previsão escolhido?	28
2.4.2	Quais os modelos de previsão empregados para modelagem dos dados, e as estratégias de treinamento utilizadas?	31
2.4.3	Quais são as principais limitações dos modelos estatísticos e de machine learning na previsão dos incêndios florestais?	32
2.5	CONCLUSÃO	33
3	PREVISÃO DE FOCOS DE INCÊNDIOS FLORESTAIS POR MEIO DE MODELOS HÍBRIDOS E DECOMPOSIÇÃO DE SÉRIES TEM- PORAIS: PERSPECTIVAS PARA DADOS NO SUL DO BRASIL.	35
3.1	INTRODUÇÃO	35
3.2	TRABALHOS RELACIONADOS	37
3.3	BASE DE DADOS	41
3.4	FUNDAMENTAÇÃO TEÓRICA	44
3.4.1	Média Móvel Integrada Autorregressiva	44
3.4.2	Modelo Autorregressivo Integrado de Médias Móveis com Variáveis Exógenas	45
3.4.3	Regressão por Vetores de Suporte	46
3.4.4	<i>Extreme Gradient Boosting</i>	47
3.4.5	Floresta Aleatória	48
3.4.6	Máquina de Aprendizado Extremo	49
3.4.7	<i>Regressão Lasso</i>	50
3.4.8	Decomposição de Modo Variacional	51
3.4.9	Interpretação de Variáveis de Entrada Utilizando Valores Shapley	53
3.4.10	Medidas de Desempenho	55
3.4.11	Validação Cruzada e Janela Deslizante	56
3.5	METODOLOGIA	57
3.5.1	Implementação dos Modelos e Ajuste de Hiperparâmetros	61
3.6	RESULTADOS	63

3.6.1	Resultados em Lages	63
3.6.1.1	Previsão de Incêndios Florestais para um Mês à Frente (H=1)	63
3.6.1.2	Previsão de Incêndios Florestais para Dois Meses à Frente (H=2)	66
3.6.1.3	Previsão de Incêndios Florestais para Três Meses à Frente (H=3)	69
3.6.2	Resultados em Xanxerê	72
3.6.2.1	Previsão de Incêndios Florestais para um Mês à Frente (H=1)	72
3.6.2.2	Previsão de Incêndios Florestais para Dois Meses à Frente (H=2)	75
3.6.2.3	Previsão de Incêndios Florestais para Três Meses à Frente (H=3)	77
3.6.3	Análise da Contribuição das Variáveis de Entrada para os Modelos de Previsão	78
3.6.4	Discussão sobre o Desempenho dos Modelos Univariados, Multivariados e Híbridos	81
3.7	CONCLUSÃO	86
4	DISCUSSÃO FINAL E IMPLICAÇÕES PRÁTICAS	87
	REFERÊNCIAS	89
	ANEXO A – AUTORIZAÇÃO CO-AUTORES PARA INCLUSÃO DE PESQUISA EM DISSERTAÇÃO	101

1 INTRODUÇÃO

Os incêndios florestais são um problema constante que afetam nosso planeta, impactando diversos aspectos como a economia, a saúde e a destruição de ecossistemas. Um grande número de incêndios é causado por humanos, embora outros fatores, como a seca, o vento, os raios e a topografia, tenham uma influência importante na ocorrência e na propagação do fogo (SAYAD *et al.*, 2019). A análise da influência dos fatores antropogênicos nos incêndios florestais tem sido realizada por muitos pesquisadores, e a importância desse tipo de pesquisa é demonstrada pelo fato de que a atividade humana é a principal causa dos incêndios no mundo (KOLANEK *et al.*, 2021). Nesse contexto, medidas para prevenir incêndios florestais tornam-se essenciais para prevenir sua propagação e minimizar a destruição ambiental.

A compreensão aprofundada dos mecanismos que desencadeiam e influenciam os incêndios florestais permite aprimorar significativamente a capacidade de prever sua ocorrência, intensidade e propagação. Isso não apenas possibilita uma resposta mais eficaz, como também contribui para o planejamento estratégico de longo prazo, visando à redução dos riscos e dos impactos provocados por esses eventos devastadores

Este estudo tem como objetivo explorar métodos de previsão de incêndios florestais, por meio da análise integrada de dados históricos de ocorrências e informações meteorológicas. Serão investigadas diferentes abordagens preditivas, incluindo o uso de tecnologias emergentes como a inteligência artificial e sistemas de aprendizado de máquina baseados em regressão. Busca-se, com isso, avançar na construção de sistemas de alerta precoce mais precisos e eficazes, capazes de apoiar autoridades e comunidades na gestão proativa e adaptativa dos incêndios florestais. A proposta reforça a importância de uma abordagem integrativa, que considere tanto os fatores climáticos quanto os padrões históricos de propagação do fogo.

1.1 PROBLEMA DE PESQUISA

Em escala global, estima-se que apenas 10% dos incêndios florestais possuem causas naturais, enquanto 90% são atribuídos à ação humana (CHAVES *et al.*, 2014). Embora esse dado remeta a mais de uma década, ele ainda se mostra condizente com a realidade atual, sendo frequentemente confirmado nas ocorrências atendidas, nas quais, em grande parte dos casos, uma queimada controlada acaba fugindo do controle. Isso evidencia a importância de compreen-

der a distinção entre queimadas e incêndios. Incêndios podem ser definidos como queimas não controladas, podendo ser naturais, acidentais ou criminosas (BERLINCK; BATISTA, 2020). Por sua vez, as queimadas estão relacionadas a práticas de manejo agrícola e florestal, ou seja, queimas controladas realizadas intencionalmente por ação humana (DIAS, 2009).

Os impactos decorrentes das queimadas são complexos e abrangem níveis ambientais, sociais e econômicos. Elas causam danos à biodiversidade e emitem grandes quantidades de dióxido de carbono (CO₂) para a atmosfera, contribuindo para as mudanças climáticas (BERLINCK; BATISTA, 2020). Os fatores físicos que influenciam significativamente um incêndio florestal pode ser agrupado em três categorias: topografia, vegetação e fatores meteorológicos (VIEGAS, 1998). Em termos gerais, esses fatores incluem condições de baixa precipitação, temperaturas extremas além do uso e ocupação do solo. Nos estudos sobre incêndios florestais, destacam-se dois pontos principais: a previsão antecipada e o monitoramento. A previsão antecipada envolve estimar a probabilidade de ocorrência de um incêndio florestal antes mesmo de sua ignição inicial. Isso é feito modelando a relação entre o risco de incêndio e fatores influentes, como condições meteorológicas e conteúdo de combustível (ABID, 2021). O monitoramento, por sua vez, abrange tanto os métodos tradicionais, que dependem da observação direta por humanos, quanto os sistemas automáticos de detecção à distância (ABID, 2021). O estudo de séries temporais de incêndios florestais está diretamente associado à previsão antecipada, permitindo a construção de cenários futuros que auxiliam o efetivo na compreensão do comportamento dos incêndios em determinados períodos, contribuindo para ações mais estratégicas e eficientes.

Diversos estudos têm investigado a previsão e mapeamento de incêndios florestais utilizando séries temporais e variáveis condicionantes, como fatores climáticos, topográficos, de vegetação e socioeconômicos (SHAO *et al.*, 2023; GAO *et al.*, 2023; SONG; WANG, 2020; TRUCCHIA *et al.*, 2022). Essas pesquisas destacam a importância de adaptar os modelos a características regionais, uma vez que os fatores influenciadores podem variar significativamente entre diferentes áreas. Além disso, a combinação de múltiplas variáveis e a aplicação de técnicas de aprendizado de máquina têm se mostrado promissoras para melhorar a precisão das previsões. No entanto, lacunas persistem, como a incorporação de dinâmicas não lineares, por exemplo, vento e umidade da vegetação e a integração de dados antropogênicos mais robustos — informações relacionadas às atividades humanas, como uso do solo, densidade populacional e proximidade de áreas urbanizadas —, aspectos que podem refinar ainda mais os modelos

preditivos.

Os incêndios florestais representam uma das principais áreas de atuação do Corpo de Bombeiros Militar de Santa Catarina (CBMSC). No âmbito do 5º Batalhão, sediado em Lages, a maior cidade de Santa Catarina em extensão territorial, com 2.654,1 km², foram registrados 477 incêndios florestais entre fevereiro de 2019 e dezembro de 2024. Esses eventos compõem uma série temporal de grande valor. Os dados estão disponíveis em um sistema interno de segurança pública: o software *Analytics*, desenvolvido pelo próprio CBMSC para uso exclusivo da corporação. Esse sistema está em operação desde 2019 e é responsável pela coleta das informações. Devido ao elevado número de ocorrências, torna-se crucial estudar mecanismos de previsão que possam quantificar os focos períodos específicos, visando uma gestão mais eficiente e assertiva das operações. Nesse contexto, o presente estudo buscará realizar previsões de curto prazo, com horizonte de 1, 2 e 3 meses à frente, possibilitando a adequação do efetivo operacional em períodos críticos. A escolha da cidade de Lages se justifica pelos altos índices de incêndios florestais, sendo um dos municípios com maior número de ocorrências no estado. Adicionalmente, será incluído o município de Xanxerê, permitindo avaliar o comportamento dos modelos preditivos em uma região distinta, o que amplia a robustez e a aplicabilidade dos resultados.

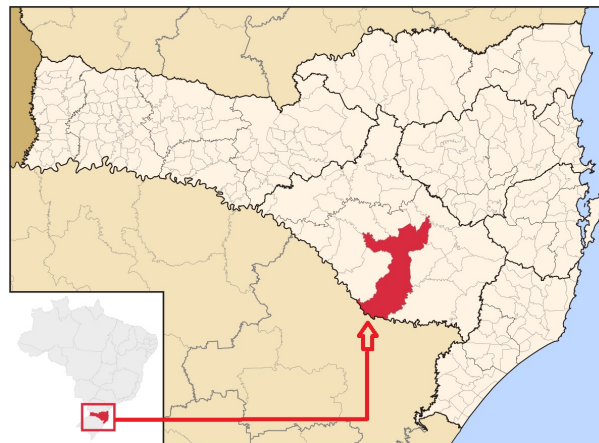


Figura 1 – Localização da cidade de Lages no contexto geográfico do Brasil e do estado de Santa Catarina.

Nessa perspectiva, é necessário adotar abordagens mais modernas e estratégias baseadas em dados para apoiar o trabalho dos bombeiros no enfrentamento dos incêndios florestais, bem como no planejamento logístico e na avaliação do manejo florestal diante de possíveis ameaças. A partir disso, apresenta-se a seguinte Questão de pesquisa (QP):

QP: Quais modelos estatísticos e de *Machine Learning* (ML) apresentam a maior



Figura 2 – Localização da cidade de Xanxerê no contexto geográfico do Brasil e do estado de Santa Catarina.

capacidade de previsão de incêndios florestais nas regiões de Lages e Xanxerê, em Santa Catarina?

O estudo em questão não apenas utiliza abordagens estatísticas tradicionais para a previsão de séries temporais, mas também explora o campo do aprendizado de máquina, um subcampo da Inteligência artificial (IA), como uma solução para previsão de séries temporais. A IA é uma ferramenta que pode identificar padrões em dados históricos e facilitar a tomada de decisões ao emular a inteligência humana (RIBEIRO, 2021). Com base nessas considerações e na análise dos dados, a aplicação de ML não apenas se destaca por sua eficácia na classificação, mas também emerge como uma abordagem promissora para realizar previsões de séries temporais.

Com base nas referências citadas, a integração de modelos de incêndios florestais com modelos de clima atmosférico tem contribuído significativamente para uma compreensão mais aprofundada da interação complexa entre o fogo e a atmosfera (SAYAD *et al.*, 2019). Estudos recentes conduzidos por pesquisadores como Aragão *et al.* (2018), Mota *et al.* (2019) e Pourghasemi *et al.* (2020) têm destacado a influência dinâmica de diversos fatores, como precipitação, temperatura, umidade do ar e uso da terra, na ocorrência de incêndios e queimadas. Portanto, é viável coletar dados observacionais ao longo do tempo para construir séries temporais, complementando-as com técnicas de aprendizado de máquina. Essa abordagem oferece a possibilidade de prever focos de incêndios florestais com maior precisão e antecedência.

1.2 OBJETIVOS DE PESQUISA

Esta pesquisa tem como objetivo geral:

- **OG:** Investigar, por meio de revisão sistemática da literatura (RSL) e experimentação empírica, as variáveis, modelos e estratégias frequentemente usadas para a previsão de incêndios florestais, com o objetivo de aprimorar a acurácia preditiva e subsidiar a tomada de decisão nas ações de prevenção e combate.

Para atingir esse objetivo geral, definem-se os seguintes objetivos específicos:

1. Realizar uma RSL com foco na identificação das principais variáveis, modelos preditivos e abordagens utilizadas na previsão de incêndios florestais, bem como suas limitações e potenciais melhorias. (Estudo 1).
2. Implementar e avaliar modelos de previsão baseados em séries temporais mensais para a estimativa de incêndios florestais, considerando diferentes horizontes de previsão (1, 2 e 3 meses à frente). (Estudo 2).
3. Integrar a técnica de Decomposição Variacional por Modos (VMD) ao processo de modelagem, aplicando-a em conjunto com algoritmos de regressão e aprendizado de máquina, tais como a Máquina de Vetores de Suporte para Regressão (SVR), o Impulsioneamento de Gradiente Extremo (XGBoost), a Floresta Aleatória (RF), a Máquina de Aprendizado Extremo (ELM) e o Operador de Seleção e Contração Absoluta Mínima (LASSO). (Estudo 2).
4. Realizar uma análise comparativa entre os modelos com e sem a aplicação do VMD, a fim de verificar o impacto dessa técnica na capacidade preditiva dos algoritmos considerados, testando tanto a abordagem univariada quanto a multivariada. (Estudo 2).
5. Avaliar a generalização dos modelos propostos em diferentes contextos geográficos, com foco nas regiões de Lages e Xanxerê (SC), visando analisar a robustez e a adaptabilidade das metodologias empregadas. (Estudo 2).
6. Aplicar uma análise interpretativa por meio dos *Shapley Values* (valores de Shapley), com o objetivo de quantificar a contribuição individual de cada variável preditora nas decisões dos modelos, possibilitando uma compreensão mais aprofundada da relevância de cada fator no comportamento dos focos de incêndio.

No Estudo 1, uma base teórica e prática é fornecida, com a identificação das variáveis e modelos mais utilizados, bem como de suas limitações. Essa abordagem permite compreender

quais fatores têm maior relevância na previsão de incêndios florestais e quais modelos demonstraram eficácia em outros contextos, podendo ser aplicados à região de Lages. Por sua vez, o Estudo 2 é voltado à avaliação e comparação de modelos específicos de regressão e aprendizado de máquina, juntamente com a decomposição por VMD, permitindo uma análise comparativa da capacidade preditiva desses modelos para a região de Lages. Além disso, dados de Xanxerê foram coletados, e será verificado se os algoritmos também são eficazes em outra região, ampliando a aplicabilidade e robustez dos modelos. Dessa forma, a questão de pesquisa proposta é diretamente abordada. Assim, os objetivos dos dois artigos se complementam, proporcionando uma análise abrangente que abrange desde a revisão das variáveis e modelos teóricos até a aplicação prática e a avaliação comparativa dos modelos na previsão de incêndios florestais para as regiões em questão.

1.3 JUSTIFICATIVA E RELEVÂNCIA DA PESQUISA

Um incêndio florestal é caracterizado por um fogo descontrolado que se espalha rapidamente e consome vários tipos de materiais inflamáveis presentes em uma floresta. Apesar de não ser muito apropriado, o termo “incêndio florestal” é, muitas vezes, generalizado para definir incêndios em outros tipos de vegetação, tais como capoeiras, campos e pradarias (SOARES; PAEZ, 1972). São ocorrências comuns em todo o mundo que podem destruir vastas áreas de floresta e matar inúmeros animais em poucas horas (GARCÊS; PIRES, 2023).

Os incêndios florestais representam um problema devido ao grande número de fatores que os influenciam. Na maioria dos países, a prevenção e mitigação desses incêndios exigem uma abordagem multi-institucional para tratar seus diversos aspectos e problemas específicos. Do ponto de vista científico, é necessária a cooperação de várias disciplinas na análise desse problema (VIEGAS, 1998). Algumas das ferramentas utilizadas para prever focos de incêndio incluem técnicas de análise multivariada de séries temporais, que permitem prever a ocorrência do fenômeno a partir de uma grande quantidade de dados sobre os valores de determinadas variáveis meteorológicas que influenciam a ocorrência de incêndios e o número de focos (VIGANÓ *et al.*, 2018). No Brasil, a detecção e o monitoramento operacional dos focos de queimadas e incêndios florestais detectados por satélites são realizados pelo Programa Queimadas, desenvolvido e administrado pelo Instituto Nacional de Pesquisas Espaciais (INPE) (JESUS *et al.*, 2020). Este monitoramento considera qualquer ponto quente na região delimitada, o que pode representar um desafio para a série temporal, já que, na região sul, é comum a prática de

queimadas controladas para limpeza de vegetação. Uma queimada pode, e muitas vezes ocorre, fugir do controle e se transformar em um incêndio, sendo então registrada como tal. Embora os dados de satélite sejam amplamente utilizados para o monitoramento de focos de calor, esses sistemas apresentam limitações reconhecidas na distinção entre queimadas controladas e incêndios florestais. Por isso, neste trabalho, optou-se por utilizar séries temporais de ocorrências atendidas por bombeiros, que refletem eventos com necessidade de intervenção humana, evitando superestimativas comuns em análises baseadas apenas em sensoriamento remoto.

A previsão de incêndios florestais é um desafio que pode trazer benefícios significativos para a sociedade e o meio ambiente. A mitigação de possíveis focos em determinados períodos pode auxiliar substancialmente as equipes de combate a incêndios, especialmente considerando o efetivo reduzido em alguns dias. A análise de séries temporais de incêndios, combinada com dados climáticos fornecidos pela Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (EPAGRI), fornecerá informações valiosas sobre as condições ideais para a ocorrência de possíveis focos de incêndio.

Sob o aspecto econômico, a previsão dos focos de incêndio permitirá uma alocação mais eficiente de recursos, como equipes de combate, equipamentos e logística, resultando em uma redução significativa dos custos operacionais (SILVA; PEREIRA, 2020). Além disso, previsões precisas poderão minimizar os impactos econômicos decorrentes de incêndios descontrolados, que afetam áreas produtivas, propriedades privadas e a infraestrutura local (COSTA; SANTOS, 2023). Ao antecipar os períodos de maior risco, será possível implementar ações preventivas, como campanhas de conscientização e alertas à população, evitando assim prejuízos consideráveis, especialmente para o setor agrícola e o turismo, duas importantes fontes de renda na região.

No aspecto ambiental, o controle eficaz dos incêndios florestais é fundamental para a preservação dos ecossistemas da região, que abriga uma biodiversidade rica e áreas de preservação ambiental (ALMEIDA, 2022). A degradação dessas áreas pode causar impactos irreversíveis na fauna, flora e nos recursos hídricos ((INPA), 2021). O estudo contribuirá para a mitigação dos danos ambientais, promovendo a preservação dos recursos naturais e o equilíbrio ecológico, fatores essenciais para garantir a sustentabilidade da região a longo prazo (ALMEIDA, 2022).

A pesquisa proposta será conduzida nos municípios de Lages e Xanxerê, localizados em Santa Catarina, regiões que historicamente apresentam elevados índices de incêndios flores-

tais. Lages, o maior município do estado em extensão territorial, concentra o maior número de ocorrências, enquanto Xanxerê também se destaca pelo volume significativo de atendimentos relacionados a queimadas, conforme registros do CBMSC. Esses dados, compostos por ocorrências efetivas, formam uma série temporal robusta, que será utilizada na primeira etapa do estudo. Na etapa seguinte, serão incorporadas variáveis climáticas de origem diária, incluindo precipitação total (mm), temperatura do ar (°C), umidade relativa (%) e velocidade do vento (m/s), visando enriquecer a análise preditiva dos focos de incêndio. Essas informações permitem identificar as condições favoráveis para a propagação de incêndios florestais. Além disso, a presença do fenômeno La Niña foi considerada como uma variável binária, devido a anomalias detectadas na série temporal associadas a esse fenômeno. Diferentemente de estudos anteriores, que em geral se concentram na criação de mapas de suscetibilidade, este estudo visa quantificar os possíveis focos de incêndio florestal para os meses subsequentes. Os dados coletados fornecem características climáticas específicas da região, obtidas por meio de um satélite de monitoramento. Em ambas as cidades, a criação de mapas de suscetibilidade teria um impacto limitado para os bombeiros, uma vez que sua atuação ocorre exclusivamente quando acionados. Dessa forma, a previsão de possíveis incêndios florestais pode tornar a gestão do efetivo mais eficiente, permitindo uma resposta mais rápida e eficaz aos incidentes.

O estudo é pioneiro na região e se destaca pela utilização de uma série temporal de ocorrências atendidas, evitando possíveis ruídos presentes em imagens de satélite. O uso de dados específicos e atualizados permite uma análise mais precisa das condições propícias para incêndios. As informações obtidas têm aplicabilidade prática, facilitando a alocação de recursos e a preparação estratégica, o que aumenta a eficácia na resposta às emergências. Com a gestão adequada dos combatentes em determinados períodos, o efetivo empregado no controle do fogo pode suprimir a propagação combatendo diretamente os possíveis focos. Isso ajuda a conter a ameaça, reduzindo a intensidade da destruição dos ecossistemas, propriedades e poluição ambiental. Além disso, o estudo estabelece um marco inicial que pode ser expandido e refinado por pesquisas futuras, contribuindo para um entendimento mais profundo dos incêndios florestais.

Além de sua relevância científica e prática, este estudo também se alinha aos Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos pela Organização das Nações Unidas (ONU). Em especial, relaciona-se ao ODS 3 (Saúde e Bem-Estar), na medida em que a previsão e prevenção de incêndios florestais contribuem para a redução da emissão de poluentes prejudiciais à saúde humana, e a ODS 13 (Ação Contra a Mudança Global do Clima), considerando

que o controle desses eventos auxilia na mitigação de seus efeitos sobre o aquecimento global e na preservação ambiental. Dessa forma, a pesquisa fortalece o compromisso com metas globais que buscam equilibrar desenvolvimento, saúde pública e proteção ambiental.

1.4 ETAPAS DA PESQUISA

Nessa seção abordaremos a metodologia empregada nos estudos propostos e de que forma eles se relacionam com a questão de pesquisa.

Para o primeiro artigo, realizamos uma RSL com uma abordagem qualitativa, buscando estudos sobre previsão de incêndios florestais nas bases *Scopus* e *Web of Science*, no período de 2019 a 2023. Esse processo resultou em um Portfólio Bibliográfico composto por 57 artigos, permitindo um aprofundamento no tema dos incêndios florestais. Foram analisadas as variáveis utilizadas, o horizonte de previsão, os modelos e as estratégias de treinamento, além das principais limitações dos métodos de regressão e aprendizado de máquina. Esse estudo proporcionou um embasamento sólido sobre as metodologias e os resultados anteriores, oferecendo direcionamento fundamental para a pesquisa proposta e a busca pelos resultados desejados.

No segundo estudo, adotamos uma estratégia univariada, utilizando apenas a série temporal dos focos de incêndio atendidos pelo Corpo de Bombeiros. A variável foi analisada em três formas: original, raiz quadrada e logarítmica, com o objetivo de verificar o impacto dessas transformações na performance dos modelos preditivos. Foram utilizadas metodologias como ARIMA, SVR, XGBoost, RF, ELM e LASSO, com a aplicação do método VMD. Essa abordagem teve como objetivo aprimorar a extração de características relevantes dos dados e, assim, aumentar a precisão das previsões. Na sequência, adotamos uma estratégia multivariada, com a inclusão de variáveis climáticas ao conjunto de dados, visando enriquecer o processo preditivo. Também foi introduzido o modelo ARIMAX, que permite incorporar variáveis exógenas à estrutura ARIMA tradicional, proporcionando uma modelagem mais completa e sensível às condições ambientais.

Para todas as abordagens, os algoritmos foram treinados, testados e avaliados com base em diferentes medidas de desempenho. Os resultados permitiram prever os possíveis focos de incêndio para os três meses subsequentes, além de identificar limitações e oportunidades para aprimoramentos futuros. A estratégia adotada proporcionou uma análise abrangente, contemplando desde a revisão das variáveis e modelos teóricos até a aplicação prática e a avaliação comparativa dos métodos empregados na previsão de incêndios florestais nas regiões estuda-

das. Além disso, os achados oferecem novas perspectivas complementares para investigações futuras.

Sob o aspecto macro, esses estudos oferecem uma visão abrangente sobre os incêndios florestais, destacando os métodos utilizados e os dados das séries temporais. A abordagem escolhida proporciona uma análise completa que vai desde a revisão das variáveis e modelos teóricos até a aplicação prática e a avaliação comparativa dos métodos de previsão. Isso não apenas aprimora a capacidade de previsão, mas também sugere novas possibilidades para futuros estudos, enriquecendo a compreensão e o manejo dos incêndios florestais na região proposta.

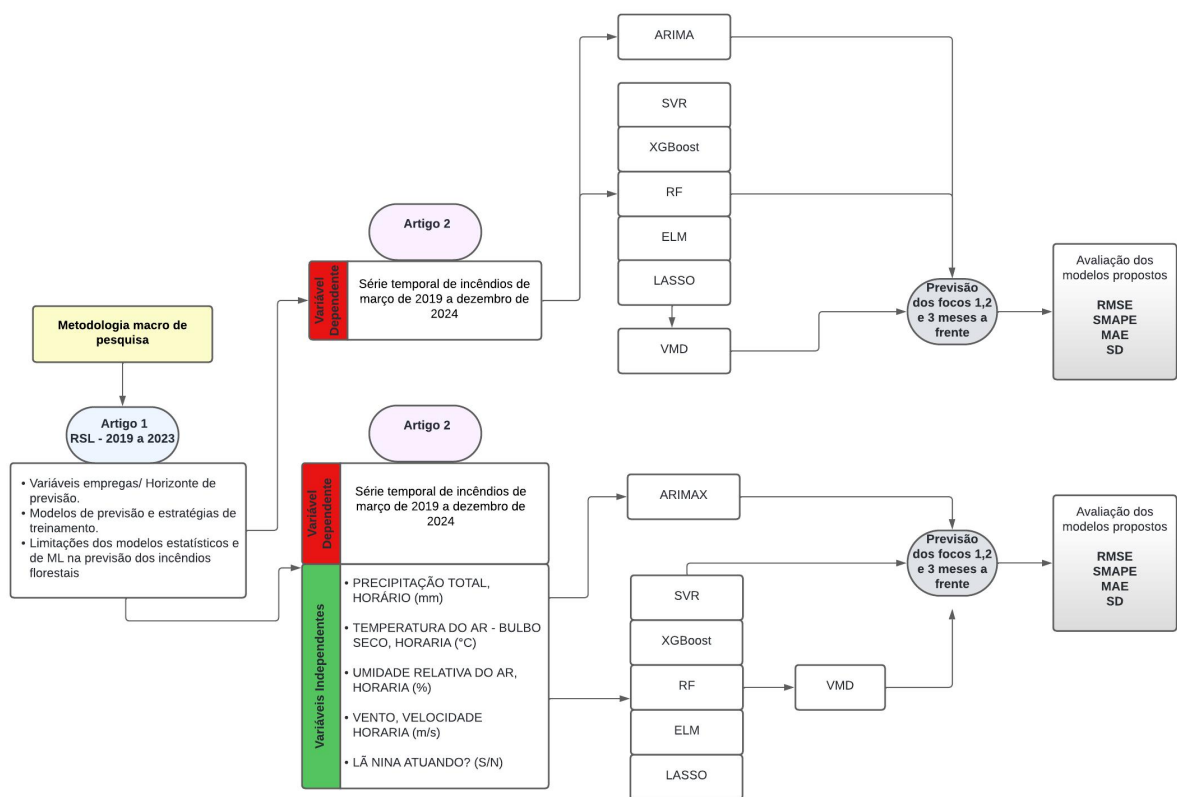


Figura 3 – Estrutura do estudo envolvendo os artigos trabalhados.

1.5 LISTA DE PUBLICAÇÕES

Dentre os estudos já publicados, destacam-se dois artigos apresentados no Encontro Nacional de Engenharia de Produção (ENGEPE), nos anos de 2023 e 2025, conforme descrito a seguir:

- - **Chioquetta, L. N., Deon, S., Rocha, V. L., & Lima, J. D. (2023).** Aplicação de técnicas de aprendizado de máquina para previsão da velocidade do vento. Apresentado no

(ENEGEP 2023), Fortaleza, CE, Brasil.

- - **Rocha, V. L., & Ribeiro, M. H. D. M.** (2024). Desafios na previsão de incêndios florestais: uma revisão da literatura. Apresentado no (ENEGEP 2024), Porto Alegre, RS, Brasil.
- - **Rocha, V. L., Ribeiro, M. H. D. M., Oliveira, G. A., & Chiu, A. S. F.** (2025). *Wildfire Spots Analysis and Forecasting: Contributing to Sustainability, Environmental Protection, and Economic Resilience. International Conference on Production Research (ICPR 2025)*, Chía, Colômbia.

ESTRUTURA DO ESTUDO

Este estudo está organizado da seguinte forma:

- **Capítulo 1:** Elementos pré-textuais: introdução, justificativa e contextualização.
- **Capítulo 2:** Desafios na previsão de incêndios florestais - Uma revisão da literatura.
- **Capítulo 3:** Desenvolvimento de modelos preditivos para focos de Incêndio na região de Lages, SC, combinando técnicas de decomposição e modelos de aprendizado de máquina.

2 DESAFIOS NA PREVISÃO DE INCÊNDIOS FLORESTAIS - UMA REVISÃO DA LITERATURA

O presente capítulo refere-se a um artigo apresentado no Encontro Nacional de Engenharia de Produção, realizado em Porto Alegre, Rio Grande do Sul, em 2024 (DOI: 10.14488/enegep2024_tn_wpg_413_2034_47421).

2.1 INTRODUÇÃO

Incêndios florestais são uma das principais preocupações ambientais, anualmente milhões de hectares são destruídos em todo o mundo, causando danos econômicos, ecológicos e perdas humanas (ABID, 2021). O clima e o tempo - incluindo temperatura, precipitação, vento e umidade atmosférica - são aspectos críticos da atividade de incêndios (FLANNIGAN *et al.*, 2016).

A ocorrência de incêndios florestais pode ser desencadeada tanto por fatores naturais, como a incidência de raios, quanto por ações humanas, como queimadas descontroladas, negligência na gestão de áreas de risco e outras atividades antrópicas. A região de Lages, no estado de Santa Catarina, registra os mais altos índices de incêndios florestais do estado. Portanto, dada sua significativa incidência, e considerando que é uma das áreas de atuação mais intensa dos bombeiros, a investigação de mecanismos de previsão torna-se um aspecto relevante tanto para gerir o pessoal envolvido no combate aos focos quanto para lidar com a propagação dos incêndios.

Nesse contexto, destaca-se o potencial do ML para a análise dos dados, incluindo séries temporais de incêndios e fatores climáticos, o que pode contribuir significativamente para a previsão de possíveis focos e fornecer insights valiosos para os bombeiros no combate eficaz aos incêndios florestais, bem como para o planejamento de serviços logísticos e a avaliação do manejo florestal diante de ameaças potenciais.

O objetivo deste estudo é realizar uma revisão da literatura científica, identificando possibilidades de pesquisa e então identificar o que está sendo desenvolvido na área de previsão de focos de incêndios florestais. Dessa forma, por meio da análise de artigos científicos relevantes, extraídos das bases de dados *Scopus* e *Web of science*, é possível identificar lacunas no conhecimento existente, visando alcançar uma compreensão abrangente dos métodos estatísticos e de ML na previsão de focos de incêndio. Para atingir esse objetivo, foram estabelecidos

objetivos específicos, que são: (I) Selecionar um portfólio bibliográfico compostos por artigos relacionados a previsão de focos de incêndio; (II) Realizar uma análise bibliométrica do (PB) e de suas referências, buscando mapear as redes de conhecimento e identificar lacunas na pesquisa, proporcionando uma visão abrangente do campo e orientando futuras investigações. (III) Realizaremos uma análise sistemática, abordando questões de pesquisa e examinando metodologias e indicadores.

Especificamente, revisaremos as variáveis utilizadas na previsão de incêndios florestais e os horizontes de previsão escolhidos em cada caso. Em seguida, examinaremos os modelos atualmente empregados na previsão de incêndios florestais, destacando as principais dificuldades encontradas e as considerações importantes na realização dessas previsões.

O restante deste artigo segue a seguinte estrutura: a Seção 2.2 descreve a metodologia do estudo. Na Seção 2.3 é descrita a seleção do portfólio bibliográfico, análises bibliométricas e sistêmicas. Na Seção 2.4, são apresentados e discutidos os resultados obtidos. A Seção 2.5 traz as conclusões finais e propõe direções para pesquisas futuras relacionadas ao tema abordado.

2.2 METODOLOGIA

A natureza da pesquisa se enquadra em exploratória e descritiva, pois explora e descreve o conhecimento sobre o tema em estudo, proporcionando uma maior compreensão.

Para realizar uma revisão bibliográfica abrangente e estruturada sobre o tema da previsão de incêndios florestais utilizando métodos de previsão como abordagens estatísticas clássicas e métodos relacionados a inteligência artificial, em como compreensão dos elementos básicos da modelagem e previsão de séries temporais, utilizou-se a abordagem PROKNOW-C, *Knowledge Development Process-Constructivist* (ENSSLIN *et al.*, 2013), como método de revisão de literatura. Esse método proporciona um levantamento sistemático de estudos relevantes, permitindo uma análise criteriosa dos trabalhos que abordam o assunto em questão. Dessa forma, foi possível identificar e selecionar adequadamente os artigos científicos que contribuem para a compreensão e o avanço nessa área de pesquisa. O método é composto por três etapas: seleção do portfólio; análise bibliométrica; e análise sistêmica (ENSSLIN *et al.*, 2013).

2.3 SELEÇÃO DO PORTFÓLIO BIBLIOGRÁFICO

Para a revisão de literatura sobre a previsão de incêndios florestais, foi elaborada uma estratégia de busca direcionada. Esta abordagem focou na utilização de bases de dados acadêmicas específicas e motores de busca especializados, como *Scopus* e *Web of Science*, para garantir a identificação de estudos relevantes e de alta qualidade dentro de um escopo definido. A estratégia de busca foi estruturada com uma combinação de palavras-chave e operadores booleanos (AND, OR), conforme ilustrado na Figura 4.

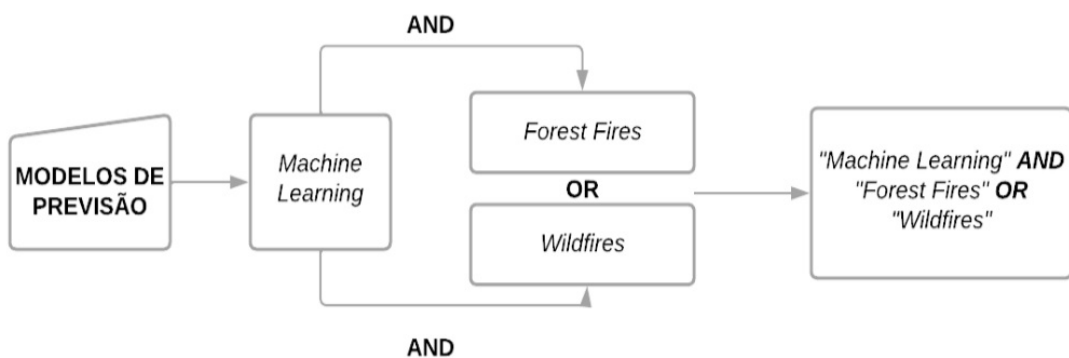


Figura 4 – Processo de composição das palavras-chave com os respectivos operadores booleanos.

Além disso, a busca foi realizada utilizando o Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), cobrindo o período de 2019 a 2023. A seleção dos artigos foi baseada em palavras-chave específicas, combinadas de diferentes maneiras para refinar os resultados da busca.

2.3.1 Critérios de Qualidade e Seleção dos Estudos

Para garantir a transparência e minimizar possíveis vieses, foram estabelecidos critérios de qualidade e inclusão de acordo com as diretrizes do método PROKNOW-C (ENSSLIN *et al.*, 2013). A seleção dos estudos seguiu critérios predefinidos, conforme descrito na Tabela 1, focando em trabalhos que aplicam técnicas de ML para prever incêndios florestais, utilizando séries temporais e variáveis relevantes para diferentes regiões.

Somente artigos revisados por pares, publicados em periódicos acadêmicos, anais de conferências e livros de editoras reconhecidas foram incluídos, com o inglês como idioma de publicação. A revisão abrange os avanços mais recentes da área nos últimos quatro anos. O processo de seleção dos estudos envolveu múltiplas etapas, incluindo triagem de títulos, resumos

Tabela 1 – Critérios de inclusão dos estudos

No.	Questão
1	O estudo está escrito em inglês
2	O estudo está disponível em texto completo
3	O estudo foi publicado em um periódico científico revisado por pares, conferência, livro ou capítulo de livro
4	O estudo está relacionado a <i>ML AND Forest Fires OR Wildfires</i>
5	O estudo foi publicado nos últimos quatro anos

e palavras-chave, seguida pela avaliação do texto completo. A seguir, são descritas as etapas percorridas.

Passo 1 : Realização da busca de estudos nas bases de dados selecionadas;

Passo 2 : Remoção dos estudos duplicados;

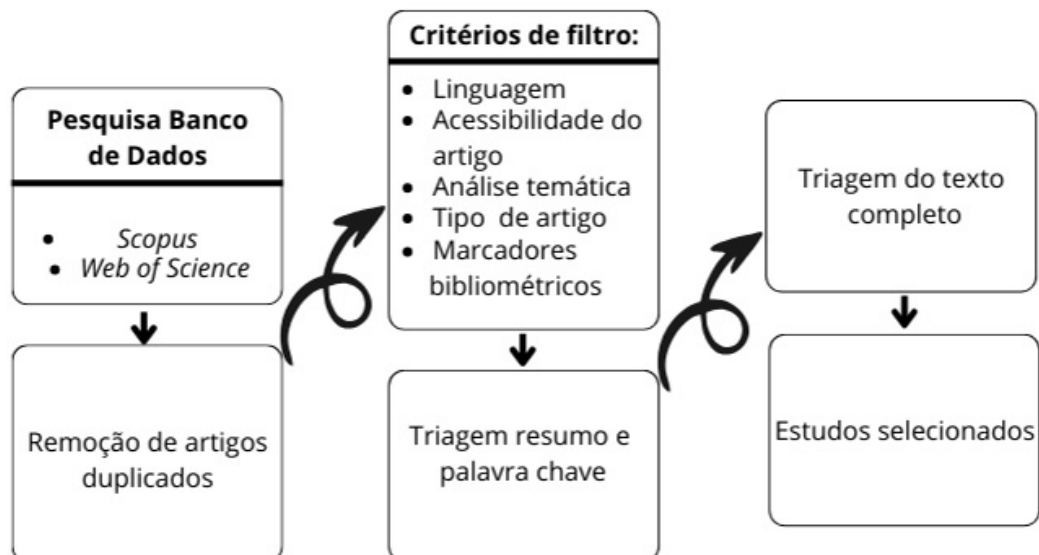
Passo 3 : Filtragem dos estudos conforme os critérios de qualidade estabelecidos;

Passo 4 : Triagem dos estudos com base nos resumos e nas palavras-chave;

Passo 5 : Avaliação dos estudos com base no texto completo;

Passo 6 : Inclusão dos estudos considerados relevantes.

A Figura 5 resume cada uma das etapas percorridas.

**Figura 5** – Diagrama de seleção de estudos.

2.3.2 Análise Bibliométrica

O método da análise bibliométrica não é baseado na análise de conteúdo das publicações, sendo o foco a quantidade de vezes em que os respectivos termos aparecem nas publicações ou a quantidade de publicações contendo os termos rastreados (YOSHIDA, 2010). A pesquisa em questão envolveu a realização de uma análise bibliométrica, que teve como foco a contagem de citações dos artigos relacionados ao objeto de estudo. Para tal, foram considerados diversos recursos, como o Google Acadêmico, periódicos científicos, autores renomados e artigos de destaque na área. Além disso, explorou-se o foco dos trabalhos e principais domínios, através do mapeamento de redes, utilizando o software VOSviewer como ferramenta de apoio.

2.3.3 Análise Sistêmica

A análise sistêmica, enquanto processo científico, permite examinar uma amostra representativa de artigos sobre um determinado tema de pesquisa a partir de uma visão teórica específica (GULARTE *et al.*, 2018). Seu objetivo é identificar lacunas e oportunidades de investigação, com base em uma afiliação teórica (LACERDA *et al.*, 2012). Nesta seção, pretende-se responder às perguntas de pesquisa formuladas neste estudo por meio de uma análise detalhada dos dados coletados. Ao realizar uma revisão abrangente dos dados pertinentes a essas questões de pesquisa (QPs), busca-se contribuir com o avanço do conhecimento existente e fornecer novos *insights* sobre a previsão de incêndios florestais. Com isso, foram definidas as seguintes QPs, conforme apresentado na Tabela 2:

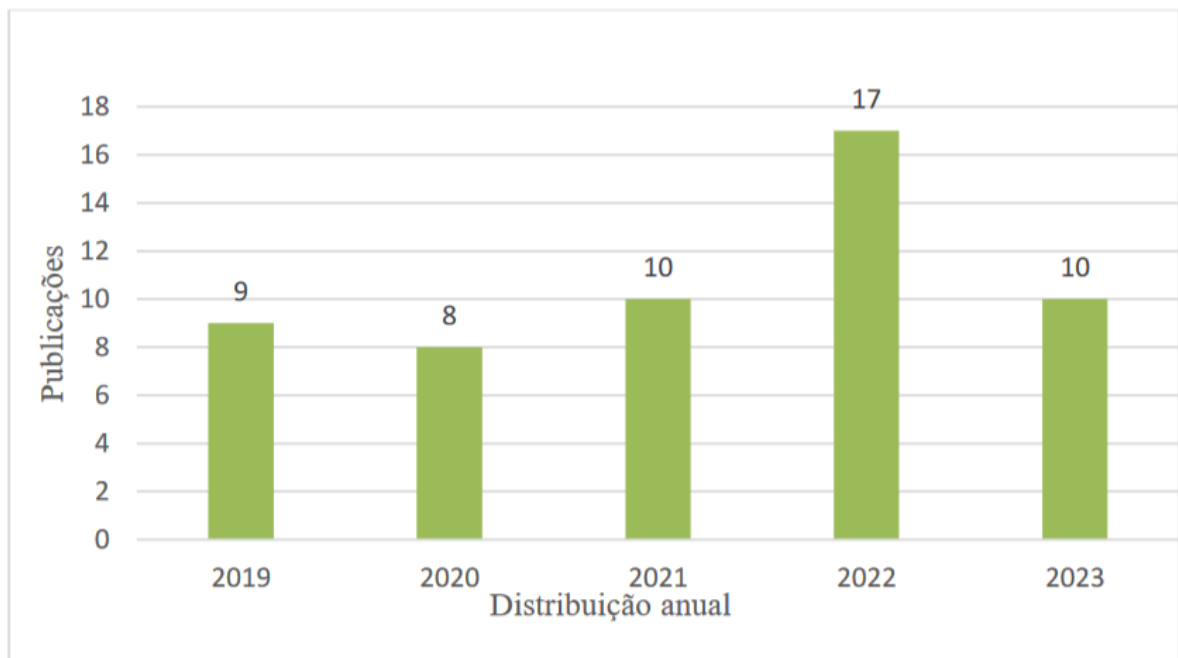
2.4 APRESENTAÇÃO DOS RESULTADOS

Nesta seção são apresentados os resultados da análise bibliométrica e sistêmica dos artigos que compõe o Portfólio Bibliográfico. Os artigos englobavam publicações de 2019 a 2023. De acordo com o levantamento, observou-se uma notável concentração de artigos nos anos de 2020 e 2022, representando 51% do total das publicações. Esse dado reflete o significativo interesse dos pesquisadores, especialmente devido à preocupante incidência de incêndios florestais nesse período. Em contrapartida, 34% das publicações ocorreram nos anos 2021 e 2023, enquanto 13% datam do ano de 2019.

Entre as 25 fontes distintas analisadas, os 57 trabalhos selecionados abrangem uma

Tabela 2 – Visão geral sistemática das principais questões de pesquisa.

No.	Questão	Motivação
QP1	Quais são as variáveis empregadas na modelagem de previsão de focos de incêndio e o horizonte de previsão escolhido?	Identificar as variáveis empregadas e o horizonte de previsão adotado permite garantir que todos os fatores relevantes estejam contemplados no modelo, possibilitando o ajuste da precisão das estimativas e favorecendo ações preventivas mais eficazes.
QP2	Quais os modelos de previsão empregados e o horizonte de previsão adotado permite garantir que todos os fatores relevantes estejam contemplados no modelo, possibilitando o ajuste da precisão das estimativas e favorecendo ações preventivas mais eficazes. os para modelagem dos dados, e as estratégias de treinamento utilizadas?	Entender os modelos de previsão empregados e as estratégias de treinamento é essencial para avaliar a eficácia do modelo e otimizar seu desempenho. Isso ajuda a escolher a abordagem mais adequada para os dados específicos e a garantir que o modelo seja treinado de forma eficiente para gerar previsões precisas.
QP3	Quais são as principais limitações dos modelos estatísticos e de ML na previsão dos incêndios florestais?	Conhecer as principais limitações dos modelos estatísticos e de ML na previsão de incêndios florestais possibilita a identificação de áreas de melhoria e o ajuste dos modelos para a obtenção de previsões mais precisas. Compreender essas limitações também contribui para a adaptação de estratégias, a seleção de técnicas adequadas e o desenvolvimento de soluções mais eficazes para enfrentar os desafios específicos da previsão de incêndios.

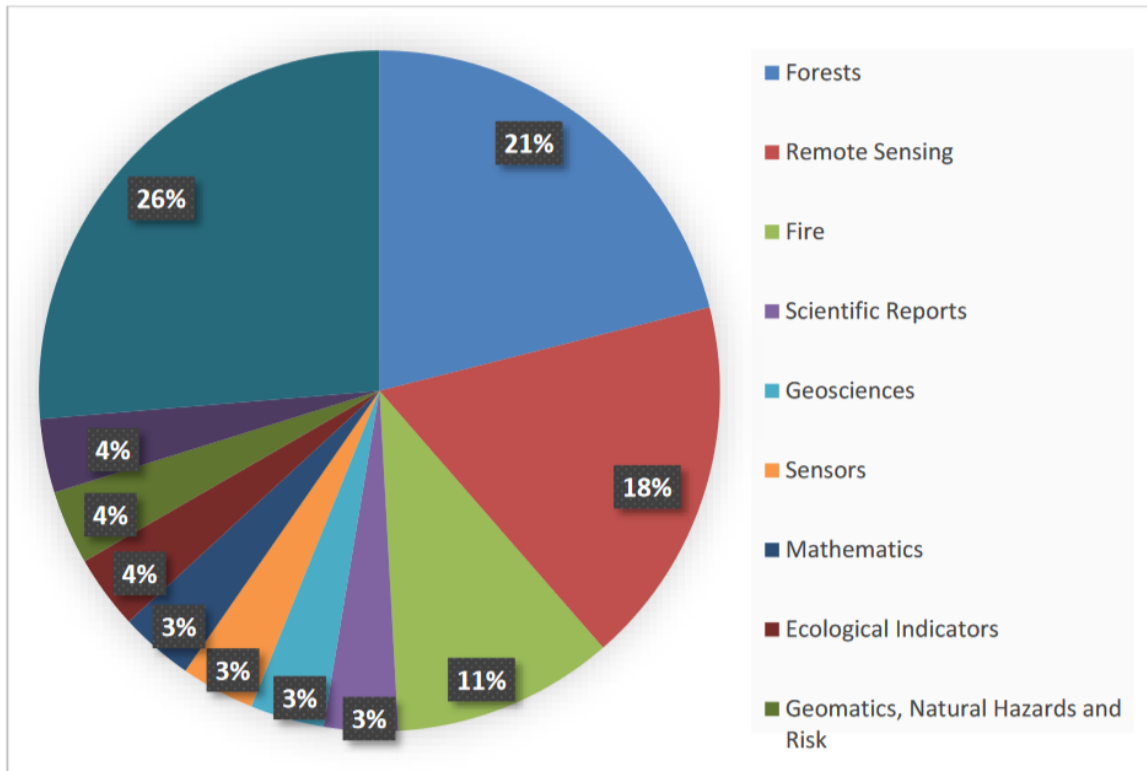


Fonte: Dados da pesquisa

Figura 6 – Distribuição anual das publicações.

ampla gama de áreas de pesquisa. Notavelmente, a revista *Forests* se destaca como a principal publicadora de estudos voltados para a previsão de incêndios florestais. A Figura 6 ilustra

graficamente essa distribuição dos artigos por periódico. Sua análise revela que as publicações mais frequentes estão vinculadas ao uso de ferramentas de aprendizado de máquina na previsão de incêndios e ao desenvolvimento de mapas de suscetibilidade. Em seguida, encontram-se os periódicos focados em aspectos relacionados ao risco de incêndio.



Fonte: Dados da pesquisa

Figura 7 – Gráfico da distribuição das publicações por periódico.

Analisando as palavras-chave na busca por artigos, o conjunto identificado no PB totalizou 2031 palavras coletas, destas, somente 62 se repetem por mais de cinco vezes nos artigos. Em sequência, para compreender a relação destes termos foi construída uma rede temática 8 com o auxílio do software de construção de redes bibliométricas VosViewer .

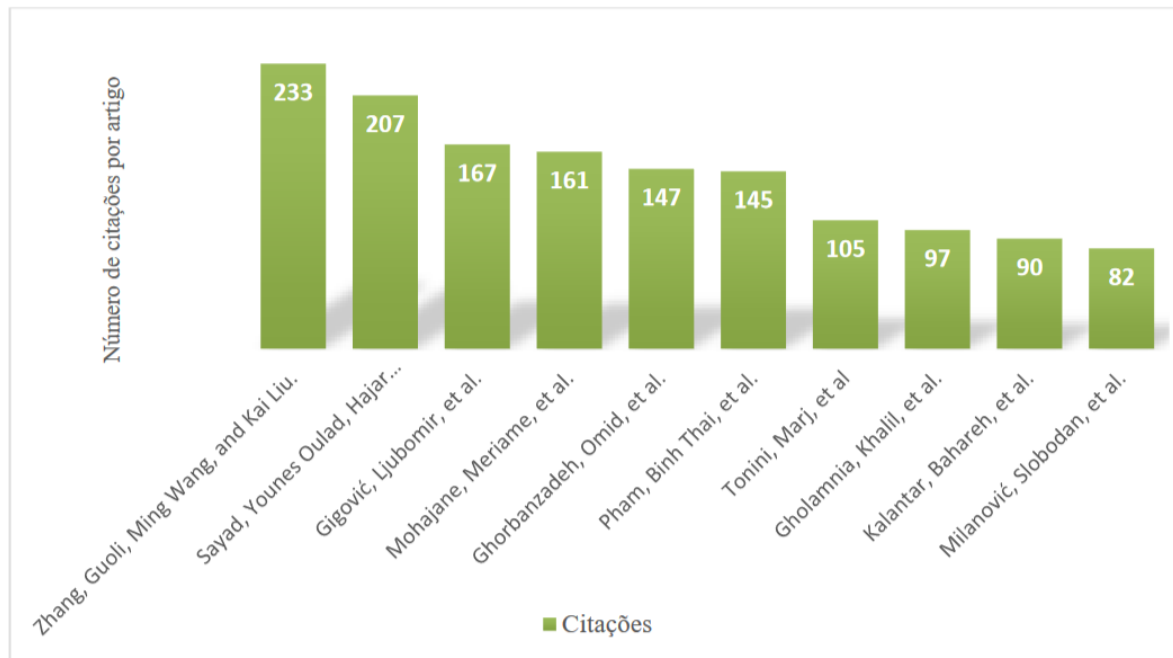
2.4.1 Quais são as variáveis empregadas na modelagem de previsão de focos de incêndio e o horizonte de previsão escolhido?

Ao classificar os 57 artigos de revisão centrados na previsão de incêndios florestais, destaca-se que 10 deles foram mencionados em mais de 80 ocasiões, consolidando-se como os 10 trabalhos mais citados. Entre eles destaca-se o estudo de (ZHANG *et al.*, 2019) “*Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China*”,

posição entre populações vulneráveis e propriedades suscetíveis a incêndios florestais. O estudo conduzido por Kalantar *et al.* (2020) emprega uma abordagem que utiliza uma série temporal de incêndios, composta por 109 pontos quentes e mais 109 pontos não relacionados a incêndios, para fins de modelagem. Quanto às variáveis complementares, são empregadas quatro variáveis provenientes de dados topográficos, três variáveis de camadas de dados vetoriais e seis índices de imagem do satélite Landsat 8, incluindo *Land Surface Temperature* (LST), *Normalized Difference Moisture Index* (NDMI), *Normalized Difference Vegetation Index* (NDVI), *Soil Adjusted Vegetation Index* (SAVI), uso da terra e cobertura da terra. O objetivo é gerar mapas de suscetibilidade a incêndios florestais. Para isso, são utilizadas técnicas como *Multivariate Adaptive Regression Splines* (MARS), *Support Vector Machine* (SVM) e *Boosted Regression Trees* (BRT). Os dados são divididos em conjuntos de treinamento e teste, sendo que o conjunto de dados de treinamento passa por um processo de reamostragem utilizando *Cross-Validation* (CV), *bootstrap* e *optimism bootstrap*. (MOHAJANE *et al.*, 2021) conduziu um estudo para mapear a suscetibilidade de incêndios florestais no norte do Marrocos. Entre os modelos utilizados, incluíram-se *Frequency Ratio with Multilayer Perceptron* (FR-MLP), *Frequency Ratio with Logistic Regression* (FR-LR), *Frequency Ratio with Classification and Regression Trees* (FR-CART), *Frequency Ratio with Support Vector Machine* (FR-SVM) e *Frequency Ratio with Random Forest* (FR-RF). Vale ressaltar que o FR-RF se destacou por lidar melhor com as não linearidades entre as variáveis, além de demonstrar maior eficiência no processamento de grandes conjuntos de dados.

O artigo conduzido por Peng *et al.* (2023) utiliza séries temporais de pontos quentes de incêndio obtidos por satélite e integra dados meteorológicos, NDVI, tipos de vegetação, dados topográficos, e informações sobre redes de estradas e população como variáveis complementares. O objetivo do estudo é prever mapas de ocorrências de incêndios tanto a médio quanto a longo prazo. O estudo conduzido por Song e Wang (2020) visa criar um mapa global de suscetibilidade a incêndios. Para isso, utiliza uma série temporal de incêndios de 1997 a 2015 e integra dados de 12 parâmetros meteorológicos desde 1979, incluindo umidade específica, pressão de superfície, precipitação, cobertura do solo, tipo de vegetação, percentual de vegetação, cobertura de neve, evaporação potencial, fluxo de calor sensível e fluxo de calor do solo. Além disso, são considerados três índices climáticos oceânicos de longo prazo: o ONI, a AMO e a PDO. No estudo Kale *et al.* (2022), foram utilizadas séries temporais de imagens de satélite de incêndios ativos de 2003 a 2017. Além disso, foram consideradas variáveis independentes, como dias sem

chuva, temperatura média mensal máxima, tipo de vegetação e elevação. O objetivo foi prever focos de incêndio para os próximos 3 anos, utilizando a modelagem ARIMA.



Fonte: Dados da pesquisa

Figura 9 – As 10 publicações mais relevantes.

2.4.2 Quais os modelos de previsão empregados para modelagem dos dados, e as estratégias de treinamento utilizadas?

O estudo de Yousefi *et al.* (2020) utiliza séries temporais de incêndios florestais e fatores diversos (topográficos, geológicos, etc.) para modelagem. São empregados os métodos FDA, GLM e SVM, com 70% dos dados para treinamento e 30% para validação, visando criar mapas de risco de longo prazo e integrar múltiplos perigos com ferramentas de ML. Nos estudos abordados por Pérez-Porras *et al.* (2021), Shao *et al.* (2022) e Shao *et al.* (2023), os dados são divididos em 70% para treinamento e validação. Os métodos utilizados incluem RF, MLP, SVM e LR em (PÉREZ-PORRAS *et al.*, 2021); RF, SVM, MLP e GBDT em (SHAO *et al.*, 2022); e LR, RF, GBDT, SVM e Ensemble Model em (SHAO *et al.*, 2023). Esses métodos são aplicados na predição e classificação de mapas de incêndios florestais.

O artigo Trucchia *et al.* (2022) utiliza o algoritmo RF, com divisão dos dados em 75% para treinamento e 25% para validação. Além disso, os autores aplicam validação cruzada com 4 dobras (4-fold cross-validation), uma técnica que consiste em dividir os dados em quatro partes

iguais. Em cada rodada, três partes são usadas para treinar o modelo e a parte restante para validá-lo, repetindo esse processo quatro vezes para que todas as partes sejam utilizadas como validação. Essa abordagem permite uma avaliação mais robusta e generalizável do desempenho do modelo, reduzindo o risco de overfitting e viés decorrente da escolha de uma única divisão dos dados. O artigo Gao *et al.* (2023) também realiza validação cruzada, mas com 5 dobras, e divide os dados em 60% para treinamento e 40% para validação, empregando os algoritmos RF e BPNN. Por sua vez, o artigo Shmuel e Heifetz (2022) utiliza RF, XGBoost, MLP e LR, aplicando validação cruzada com 10 dobras e ajustando os hiperparâmetros. O estudo conduzido por RODRIGUES e TORRES (2020) investigou a intensidade do fogo, a altura das chamas e a taxa de propagação, utilizando GLMNET com Lasso e Elastic-Net) e RF. Os dados foram divididos em 60% para treinamento e 40% para validação.

Alguns estudos utilizam algoritmos de otimização para aprimorar suas análises. O estudo Fangrong *et al.* (2021), por exemplo, emprega GBDT otimizada pelo algoritmo CatBoost, dividindo os dados em 70% para treinamento e 30% para validação, e realizando validação cruzada com 10 dobras para otimização do CatBoost. Por outro lado, o artigo Nur *et al.* (2023) utiliza SVR com divisão dos dados em 70% para treinamento e 30% para validação, aplicando uma função aleatória. Além disso, o estudo emprega algoritmos de otimização, como GWO e PSO, para criar mapas de suscetibilidade para a cidade de Sydney, na Austrália.

2.4.3 Quais são as principais limitações dos modelos estatísticos e de machine learning na previsão dos incêndios florestais?

Em relação as limitações dos modelos encontrados nos estudos, primeiramente, destacam-se a disponibilidade dos dados. A previsão de incêndios requer uma ampla gama de dados provenientes de diversas fontes. Estudos como o de Janiec e Gadal (2020), encontraram dificuldades na coleta desses dados. Além disso, pesquisas como a de Shmuel e Heifetz (2022), não consideraram variáveis climáticas, resultando em uma previsão ligeiramente menos precisa. O estudo de Reyes-Bueno e Loján-Córdova (2022) também não fez o uso de variáveis climáticas na manipulação dos dados. Além disso, o estudo identificou problemas comuns nas imagens de satélite, como sombras ou máscaras causadas por nuvens durante a captura e a falta de diferenciação na intensidade ou tamanho dos incêndios ativos (ZHANG *et al.*, 2021). Esses fatores contribuem para o aumento dos ruídos nos dados.

A incerteza e a generalização dos modelos são evidentes. Estudos como os de Ber-

gonse *et al.* (2021) e Malandra *et al.* (2022) não integraram adequadamente as variáveis climáticas locais, resultando em uma modelagem generalizada que não considera as características climáticas específicas de cada região. Adicionalmente, o artigo Wang *et al.* (2023) destacou que diferentes tipos de vegetação, localizações geográficas e topografias afetam a capacidade de retenção de água do solo e dos combustíveis, influenciando a reação dos incêndios florestais ao clima e ao combustível. Portanto, é essencial estabelecer características específicas para as mudanças nos fatores climáticos e de combustível, e realizar avaliações de risco de incêndio florestal segmentando diferentes regiões geográficas. O estudo Tang *et al.* (2020) constatou que fatores como a acumulação de biomassa, o reflorestamento e índices espectrais associados a incêndios são fundamentais, pois áreas com maior material combustível tendem a permitir uma propagação mais rápida do fogo. Da mesma forma, o artigo Shmuel e Heifetz (2023) identificou uma limitação em seu estudo atual: enquanto ele avalia a taxa total de propagação do fogo, não prevê a direção exata para onde o incêndio se espalhará.

No trabalho de Pang *et al.* (2022), constatou-se que áreas com maior densidade populacional apresentam maior probabilidade de ocorrência de incêndios florestais provocados por ação humana. Dessa forma, é necessário selecionar métodos de seleção de características específicos para essas localidades. Conforme relatado por Cilli *et al.* (2022), é fundamental fornecer uma análise de explicabilidade para entender como as características contribuem para a classificação do mapa de suscetibilidade, no caso, a ocorrência de incêndios florestais. Devido à alta estocasticidade dos incêndios florestais, impulsionada por fatores externos ao modelo, como incêndios criminosos, alguns eventos continuarão sendo imprevisíveis.

2.5 CONCLUSÃO

Este capítulo apresentou os resultados da revisão sistemática da literatura (RSL) sobre a integração de abordagens para previsão de incêndios florestais. Três questões de pesquisa orientaram o estudo, permitindo uma análise estruturada das variáveis, modelos e limitações observadas na literatura. A primeira questão investigou as variáveis empregadas na modelagem de previsão de focos de incêndio e os horizontes de previsão adotados. A análise revelou que os estudos utilizam frequentemente imagens de satélite, séries temporais de focos de incêndio e variáveis climáticas, como temperatura, umidade e precipitação. Dados topográficos, características da vegetação e fatores humanos também são incorporados em alguns trabalhos. Observou-se grande variação nos horizontes de previsão, com predominância de períodos de

médio a longo prazo, enquanto horizontes curtos, essenciais para respostas rápidas das autoridades, são pouco explorados.

A segunda questão abordou os modelos de previsão e as estratégias de treinamento. Foram utilizados tanto métodos estatísticos, com boa interpretabilidade, quanto técnicas de aprendizado de máquina e híbridas, com maior capacidade de capturar padrões não lineares complexos. Estratégias de validação incluem divisões do tipo “split” e validação cruzada (k-fold), muitas vezes combinadas com algoritmos de otimização para aprimorar o desempenho dos modelos. A escolha da técnica depende da disponibilidade de dados, da complexidade do problema e da necessidade de generalização. A terceira questão analisou as limitações dos modelos. Destacou-se que a disponibilidade e qualidade dos dados são fatores críticos, e que grande parte dos estudos não integra adequadamente variáveis climáticas locais, resultando em generalizações excessivas. A complexidade do comportamento do fogo, a diversidade de vegetação, topografia e fatores humanos, além da estocasticidade dos incêndios, limitam a previsibilidade de eventos isolados. Também foram identificadas lacunas na modelagem de áreas específicas e na previsão em horizontes curtos, essenciais para respostas rápidas e eficazes no combate a incêndios florestais.

Com base nessa análise, a presente pesquisa busca preencher três lacunas importantes: (i) utilização de séries temporais de focos de incêndio provenientes do Corpo de Bombeiros, permitindo previsões regionais mais precisas; (ii) incorporação de variáveis climáticas e do fenômeno La Niña em modelos estatísticos, de aprendizado de máquina e híbridos, aumentando a capacidade de modelagem frente às condições ambientais variáveis; e (iii) avaliação de horizontes de previsão mais curtos, promovendo respostas mais ágeis para a gestão de incêndios florestais.

3 PREVISÃO DE FOCOS DE INCÊNDIOS FLORESTAIS POR MEIO DE MODELOS HÍBRIDOS E DECOMPOSIÇÃO DE SÉRIES TEMPORAIS: PERSPECTIVAS PARA DADOS NO SUL DO BRASIL.

O estudo foi submetido à revista *Applied Soft Computing*, um periódico internacional de referência na área de ciência de dados, com CiteScore 14.5 e Fator de Impacto 6.6, conforme os indicadores da Scopus e do *Journal Citation Reports* (JCR), respectivamente. Parte dos resultados deste trabalho foram previamente apresentada na *International Conference on Production Research 2025* (ICPR 28), contribuindo para o debate científico sobre métodos de previsão de incêndios e aplicações em sustentabilidade, proteção ambiental e resiliência econômica.

3.1 INTRODUÇÃO

Incêndios florestais podem causar danos irreparáveis ao meio ambiente, à fauna e à flora, além de propriedades humanas privadas. Esses incêndios podem ser causados tanto por ações naturais quanto humanas (SLAVIA *et al.*, 2019). Sua ocorrência pode ser desencadeada por causas naturais, como a incidência de raios, secas prolongadas, altas temperaturas e climas secos (SHARMA *et al.*, 2020), ou por ações humanas, como queimadas descontroladas, negligência no manejo de áreas de risco e outras atividades antropogênicas. No entanto, estima-se que apenas 10% dos incêndios ocorrem naturalmente, enquanto 90% são atribuídos à ação humana (CHAVES *et al.*, 2014). Nesse sentido, o estudo de mecanismos de previsibilidade é essencial para prevenir e controlar a propagação de focos de incêndio.

Nesse contexto, a análise de séries temporais de incêndios, combinada com a aplicação de modelos de aprendizado de máquina (ML), têm se destacado como uma abordagem fundamental para a previsão de séries temporais (STEFENON *et al.*, 2023b; KLAAR *et al.*, 2023b; STEFENON *et al.*, 2023). O objetivo principal da pesquisa é ajudar as equipes de combate a incêndios a responder de maneira eficaz aos incêndios florestais, ao mesmo tempo em que auxilia no planejamento de operações logísticas e na avaliação de estratégias de manejo florestal diante de ameaças iminentes.

Uma série temporal pode ser definida como uma sequência de observações feitas sequencialmente ao longo do tempo (da Silva *et al.*, 2024). Os dados são estruturados de forma que cada observação esteja vinculada a um momento preciso no tempo, permitindo assim a análise das variações de uma variável específica ou de um conjunto de variáveis ao longo do

tempo (RIBEIRO *et al.*, 2024a). Os registros são acumulados progressivamente ao longo de um período, o que possibilita a detecção de vários padrões, como autocorrelação, oscilações periódicas, flutuações sazonais, estabilidade, direções de tendência, variabilidade e características recorrentes (STEFENON *et al.*, 2024).

No contexto da previsão de incêndios florestais, duas áreas de pesquisa cruciais merecem atenção especial: (a) a previsão da ocorrência de incêndios (em termos de tempo e localização) e (b) a previsão da expansão dos incêndios florestais (a área queimada no futuro) (ARIF *et al.*, 2021). Nesse sentido, uma variedade de métodos de previsão é utilizada, que, quando combinados, possibilitam a criação de um modelo com maior precisão. Esses métodos fazem uso de dados de séries temporais de incêndios em combinação com variáveis meteorológicas. Também é possível incorporar índices de vegetação, como o NDVI (MICHAEL *et al.*, 2021), obtidos de imagens de satélite, para construir a série temporal de incêndios (GORYAEV *et al.*, 2021). Modelos de previsão podem ser baseados em algoritmos de aprendizado de máquina supervisionado (YAMASAKI *et al.*, 2024; STEFENON *et al.*, 2023a; MORENO *et al.*, 2024), regressão (SEMAN *et al.*, 2023) e técnicas de aprendizado profundo (KLAAR *et al.*, 2023a).

O objetivo deste trabalho é realizar previsões de incêndios florestais com o intuito de alcançar uma precisão confiável, contribuindo para uma preparação e resposta mais eficazes por parte das autoridades. Para isso, adota-se uma abordagem híbrida de previsão, combinando técnicas avançadas de decomposição de sinais (como o VMD), algoritmos de aprendizado de máquina (SVR, XGBoost, RF, ELM e LASSO) e métodos estatísticos (ARIMA e ARIMAX). O modelo proposto é aplicado para previsões de curto prazo, com horizontes de 1, 2 e 3 meses à frente, visando estimar a ocorrência de incêndios florestais nos municípios de Lages e Xanxerê, em Santa Catarina, Brasil. As previsões são baseadas em séries temporais mensais de ocorrências de incêndios registradas e atendidas pelo CBMSC, incorporando tanto abordagens univariadas quanto multivariadas. As variáveis utilizadas como preditoras englobam dados climáticos, como precipitação, temperatura, umidade relativa do ar e velocidade do vento, além de uma variável binária indicando a ocorrência do fenômeno La Niña. A acurácia dos modelos foi avaliada por meio de medidas como RMSE, SMAPE, MAE e SD. As análises foram conduzidas utilizando os dados em sua forma original, bem como com transformações logarítmica e de raiz quadrada para estabilização da variância e melhoria na performance dos modelos. As principais contribuições deste artigo são as seguintes:

- Incorporação da série temporal dos incidentes atendidos pelo CBMSC, o que pode resul-

tar em previsões substancialmente mais precisas. Isso desempenha um papel relevante na gestão dos tempos de resposta, permitindo uma alocação eficiente de recursos, o desenvolvimento de estratégias de combate mais eficazes e a promoção da conscientização pública. Além disso, o artigo oferece uma oportunidade para demonstrar como as previsões de incêndios florestais podem desempenhar um papel fundamental no apoio à conservação ambiental, especialmente na proteção de ecossistemas vulneráveis nos municípios de Lages e Xanxerê.

- A revisão da literatura relacionada revelou que os modelos de previsão enfrentam desafios significativos na previsão de incêndios florestais, tornando a seleção do modelo uma tarefa complexa. Para abordar essa questão, foram empregadas diversas abordagens de inteligência artificial na previsão de focos de incêndio. Foi realizada uma comparação entre diferentes algoritmos, incluindo SVR, XGBoost, RF, ELM e LASSO, com o objetivo de identificar o modelo mais eficaz para prever focos de incêndio nos horizontes observados. Além disso, foi utilizado o VMD para aprimorar os métodos estatísticos e de aprendizado de máquina, proporcionando uma análise mais robusta e precisa das séries temporais.
- Foram utilizadas variáveis climáticas — precipitação, temperatura, umidade relativa do ar, velocidade do vento e o fenômeno La Niña (representado por uma variável binária) — como preditoras nos modelos de previsão. Além disso, empregou-se a técnica *Shapley Values* para interpretar a importância relativa de cada variável na contribuição para as previsões, proporcionando maior transparência e compreensão sobre o comportamento dos modelos de aprendizado de máquina utilizados.

O restante deste artigo está estruturado da seguinte forma: A Seção 3.2 discute estudos relacionados no contexto da previsão de incêndios florestais. A Seção 3.4 analisa os conjuntos de dados utilizados e detalha o método empregado nesta pesquisa. A Seção 3.6 apresenta e discute os resultados obtidos. Por fim, a Seção 3.7 apresenta as conclusões finais e propõe direções para pesquisas futuras sobre o tema.

3.2 TRABALHOS RELACIONADOS

Nesta seção, são descritos estudos recentes relacionados à previsão de incêndios florestais, que utilizaram tanto modelos de aprendizado em conjunto quanto modelos individuais.

A previsão de modelos de séries temporais tem sido amplamente utilizada atualmente devido à sua capacidade de identificar a possibilidade de variação dos dados com base no treinamento em dados históricos (MEDEIROS *et al.*, 2022; FERNANDES *et al.*, 2022; BRANCO *et al.*, 2022).

Viganó *et al.* (2018) propuseram um estudo que combina os conceitos de MLR e ARIMA para criar um modelo matemático destinado à previsão do número de focos de incêndio. Este modelo incorporou dados de variáveis meteorológicas. As técnicas de análise multivariada aplicadas aos modelos MLR e ARIMA geraram resultados divergentes. O modelo MLR explicou 41% da variância, com um MAPE de 33,9%, demonstrando-se menos eficiente na previsão. Por outro lado, a modelagem de séries temporais usando a técnica ARIMA apresentou resultados satisfatórios, com uma explicação de 66,5% e MAPE de 31,6%, RMSE de 0,484 MAE de 0,381, representando melhorias significativas no desempenho geral do modelo ARIMA em comparação com o modelo MLR. É importante notar que as unidades de medida do RMSE e MAE são adimensionais devido ao processo de normalização. Para aprimorar ainda mais a técnica ARIMA, é imperativo incluir variáveis antropogênicas, que exercem uma influência significativa na ocorrência de incêndios florestais. Essa integração permitiria uma compreensão e previsão mais abrangentes dos eventos ao considerar tanto os fatores meteorológicos quanto os fatores associados à atividade humana.

Ferreira *et al.* (2020) propuseram uma abordagem que consiste em dividir a área de estudo em células de grade e extrair séries temporais de contagem de incêndios para ajustar os modelos de previsão. O principal objetivo foi estimar FSS, que representa a soma das contagens de incêndios detectadas em uma temporada. Vários modelos foram explorados neste estudo, incluindo ARIMA, ETS, Trigonométrico, TTBATS, Decomposição Sazonal-Tendência usando LOESS e STL, TSGLM, ANN e Prophet. Para treinar os modelos de previsão, foram utilizados os dados de MA-FC. Os resultados mostraram que todas as células apresentaram um MAE menor do que a severidade média da temporada de incêndios. Além disso, em 95% das células, os MAEs para a severidade prevista da temporada de incêndios (soma das MA-FCs previstas em uma temporada) foram menores do que os valores médios de severidade da temporada de incêndios, demonstrando a possibilidade de prever com precisão a severidade geral da temporada de incêndios. O modelo de previsão TBATS destacou-se como o mais eficaz.

Jamshed *et al.* (2022) propuseram um modelo baseado em LSTM e DL para a previsão precisa das áreas de desmatamento com base em dados históricos de incêndios florestais. Para o

modelo proposto, foi definida uma LSTM com 3 camadas. Usando uma única camada densa, o modelo obteve um RMSE de $4,142 \times 10^{-5}$, enquanto com 2 camadas densas, obteve um RMSE de 4.051×10^{-5} . Isso levou à seleção de 2 camadas densas, e além disso, foi feita uma correção nos dados ajustando o valor do dropout para 0.05. Essa diferença levou à escolha de duas camadas densas como a configuração mais apropriada.

O estudo conduzido por Slavia *et al.* (2019) envolveu a análise de pontos quentes em imagens de satélite como indicadores de incêndios florestais. Dados foram coletados mensalmente para criar uma série temporal de incêndios. O procedimento utilizado neste estudo foi o procedimento de Box-Jenkins para construir um modelo ARIMA. Na avaliação do modelo, obteve-se um MAE de 3,766891 nos resultados de teste, indicando que a precisão da previsão do modelo é de 96,23%. O estudo de Santos *et al.* (2023) incorporou uma abordagem que combina previsão de séries temporais e algoritmos de classificação. Esse método foi treinado em conjuntos de dados contendo informações meteorológicas e registros de pontos quentes. Eles usaram dados de pontos quentes capturados por imagens de satélite ao longo de um período de 20 anos, criando assim uma série temporal. Esta série temporal foi então aplicada a diferentes algoritmos de previsão, incluindo LSTM, GRU e CNN. Os pesquisadores aplicaram algoritmos de classificação aos dados e compararam os resultados obtidos com os índices de risco de incêndio florestal mais amplamente utilizados na literatura especializada específica para a região do Pantanal.

No estudo de Xie e Peng (2019), os dados foram obtidos de um repositório contendo 517 incêndios florestais, para os quais foram registrados 12 atributos diariamente no conjunto de dados. Isso incluía informações como tempo, localização espacial, quatro componentes do Sistema de Alerta de Incêndios Florestais (FWI), área total queimada e outros dados meteorológicos. No primeiro experimento, diferentes modelos de regressão foram utilizados para fazer previsões, e sob os critérios de RMSE e MAE, o modelo RFt, como técnica de aprendizado em conjunto, superou outras abordagens de regressão e foi mais provável de fornecer as melhores previsões da área queimada em comparação com o modelo GLM tradicional. No segundo experimento, foram utilizados modelos de classificação para prever grandes incêndios com base nas conclusões do primeiro experimento.

Ao considerar os trabalhos listados acima, fica claro que algumas oportunidades de pesquisa foram identificadas, especialmente em relação aos modelos RF e SVR, que são amplamente reconhecidos como referências em previsões, conforme evidenciado por estudos anterior-

res. Um resumo dos estudos mencionados sobre previsão de incêndios florestais é apresentado na Tabela 3, que destaca ainda mais essas oportunidades de pesquisa.

Tabela 3 – Resumo de trabalhos relacionados a previsão de focos de incêndio sob diferentes perspectivas.

Autores	Ano	Métodos	Intervalo De Amostragem	Horizonte De Previsão	Medidas
Viganó <i>et al.</i> (2018)	2018	MLR e ARIMA	Dados Horários	1,2,3,4 e 5	MAPE, RMSE e MAE
Slavia <i>et al.</i> (2019)	2019	ARIMA, ETS, TBATS, STLF, TSGLM e ANN	Dados Mensais	1,2,3,4,5,6	MAE
Xie e Peng (2019)	2019	RFd, RFt, GBMt, GLMt e DL	Dados Horários	10	RMSE e MAE
Ferreira <i>et al.</i> (2020)	2020	ARIMA, ETS, TBATS, STLF, TSGLM e ANN	Dados Horários	7	MAE
Singh <i>et al.</i> (2021)	2021	SVM	Dados dos focos	1	MSE, RMSE e MAE
Syarifudin <i>et al.</i> (2021)	2021	CNN	Dados diários e mensais	1 e 12	MAPE
Jamshed <i>et al.</i> (2022)	2022	LSTM	Dados Horários	150	RMSE
Santos <i>et al.</i> (2023)	2023	LSTM, GRU, e CNN	Dados Horários	1,2 e 3	MAE
Pinheiro e Rodrigues (2024a)	2024	ARIMA e ETS	Dados Mensais	12	RMSE, MASE, RMSSE e AvgRelRMSE
Song <i>et al.</i> (2025)	2025	SVM, RF, XGBoost, LightGBM e SHAP	Dados horários (ERA5)	Curto prazo (mensal, densidade de raios-alvo)	R^2 , MAE, MSE, SHAP e Kruskal-Wallis
Zhou <i>et al.</i> (2025)	2025	DNN com EOF, <i>encoder-decoder</i> LSTM, <i>self-attention</i> e FC layer	Dados Mensais	Previsão de 30 dias (mapas de incêndios urbanos)	RMSE e R^2
Partheepan <i>et al.</i> (2025)	2025	Sensoriamento remoto, índices e XGBoost	Dados mensais	Curto a médio prazo (não especificado)	Accuracy, MSE e R^2

Na pesquisa liderada por Singh *et al.* (2021), foram considerados a série temporal de incêndios, o FWI e vários parâmetros meteorológicos para prever incêndios florestais. O estudo utilizou dados da Índia e de Portugal para avaliar sua eficácia. Foram utilizados métodos convencionais de SVM e SVM paralela. Os resultados indicam que o modelo SVM paralelo obteve um RMSE de 63,45, enquanto o método convencional de SVM alcançou um RMSE de 63,5 com os dados de Portugal. Esses resultados sugerem que o modelo SVM paralelo pode representar uma alternativa eficiente para prever incêndios florestais, reduzindo o tempo de processamento e os requisitos de armazenamento.

Syarifudin *et al.* (2021), em sua pesquisa, propuseram um estudo sobre séries temporais de pontos quentes, registrados em ordem cronológica e que exibem um padrão repetitivo. Esses dados de pontos quentes foram combinados com 5 outros parâmetros: precipitação, duração da radiação solar, velocidade máxima do vento, velocidade média do vento e direção do vento. Foram feitas previsões para um horizonte de 12 meses, usando dados diários e mensais. O modelo escolhido para realizar essas previsões foi a CNN, e os resultados obtidos foram um MAPE de 1,0695 para as previsões diárias e 0,5814 para as previsões mensais. No entanto, é importante notar que este método não se mostrou tão eficaz em previsões diárias devido à presença de muitos dados ausentes em cada um dos parâmetros utilizados. As CNNs são aplicadas para várias aplicações, incluindo previsão (ZHA *et al.*, 2022), detecção de falhas e classificação (CORSO *et al.*, 2023) com base em aprendizado profundo (STEFENON *et al.*, 2023).

Adicionalmente, tanto na revisão da literatura quanto nos trabalhos correlatos, observamos que poucos estudos – ou mesmo nenhum – têm explorado o uso de modelos híbridos, que combinam técnicas de decomposição e algoritmos de aprendizado de máquina, para a previsão de focos de incêndio. Esse cenário ressalta a relevância e o ineditismo da abordagem proposta neste trabalho.

Inicialmente, investigamos métodos estatísticos aplicados isoladamente e, em seguida, analisamos a aplicação da Decomposição Modal Variacional (VMD) em conjunto com métodos estatísticos e de aprendizado de máquina na previsão de incêndios florestais. A escolha do VMD não se deve apenas à ausência de estudos que o apliquem a esse problema no Brasil, mas também às características do método: trata-se de uma técnica capaz de lidar com séries temporais não lineares e decompô-las em componentes mais estáveis, reduzindo ruídos e permitindo a extração de padrões ocultos. Assim, havia a expectativa de que sua integração a modelos estatísticos e de aprendizado de máquina resultasse em previsões mais precisas e robustas.

Essa lacuna destaca a importância e o potencial inovador da abordagem proposta, que busca preencher um espaço ainda pouco explorado no campo da previsão de incêndios florestais. Ao focar na redução do horizonte de previsão, nosso objetivo é aprimorar a precisão das estimativas, viabilizando respostas mais ágeis e eficazes às ocorrências de incêndios. A integração de técnicas avançadas de decomposição e aprendizado de máquina não apenas eleva a qualidade das análises preditivas, mas também oferece uma contribuição valiosa para a gestão de riscos ambientais no Brasil, fortalecendo estratégias de prevenção e mitigação.

3.3 BASE DE DADOS

Nesta seção, são descritos o conjunto de dados utilizado e os principais aspectos dos métodos empregados no estudo. A amostra considera os incidentes de incêndios florestais registrados nas cidades de Lages (SC) e Xanxerê (SC), Brasil. A escolha dessas localidades se justifica pelas elevadas taxas de ocorrência observadas nos últimos anos. A série temporal adotada abrange o período de fevereiro de 2019 a dezembro de 2024, com registros mensais, totalizando 477 incêndios florestais atendidos pelo CBMSC em Lages e 596 em Xanxerê. O número mais elevado em Xanxerê deve-se ao fato de o batalhão local prestar atendimento também a municípios vizinhos, enquanto em Lages a cobertura está concentrada majoritariamente no próprio território municipal. Os dados utilizados estão disponíveis publicamente no repositório GitHub (ROCHA, 2025).

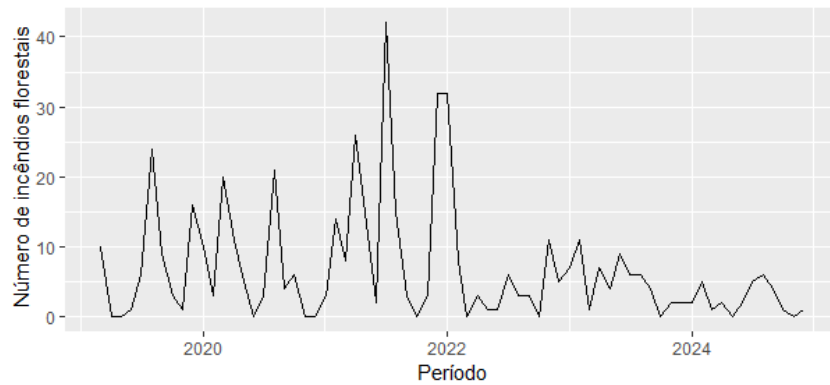


Figura 10 – Série temporal de ocorrências mensais de incêndios florestais em Lages (SC).

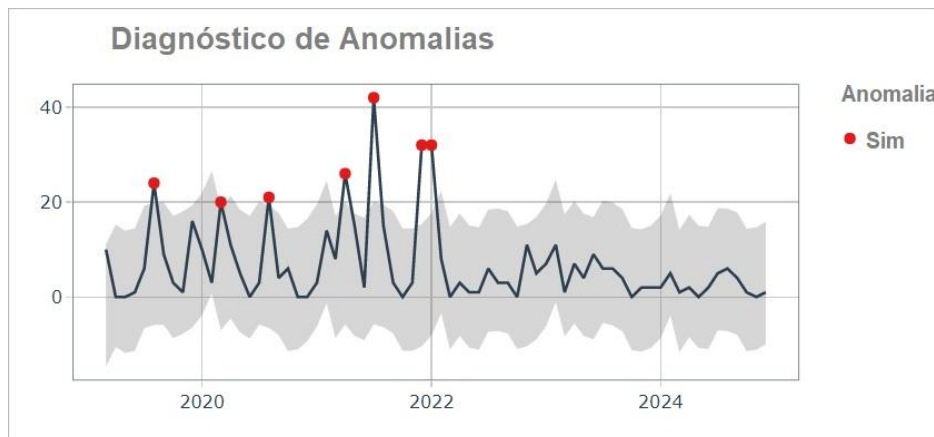


Figura 11 – Teste de anomalias no conjunto de dados da série temporal de Lages (SC).

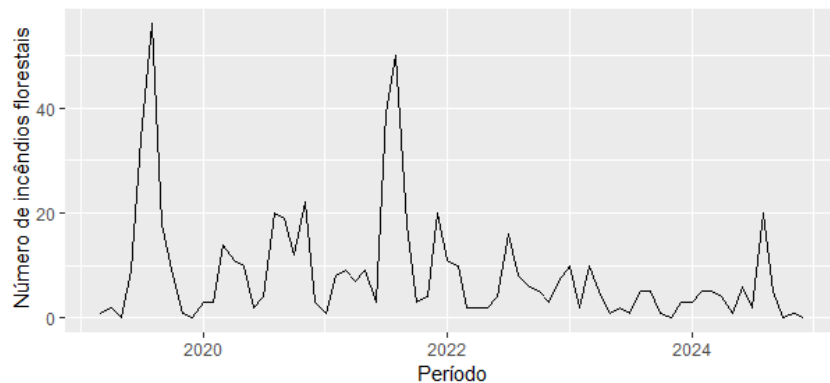


Figura 12 – Série temporal de ocorrências mensais de incêndios florestais em Xanxerê (SC).

Nas Figuras 11 e 13, podemos observar que a área cinza representa o comportamento adequado da série temporal, indicando que as ocorrências de incêndios estão dentro dos níveis médios históricos. Por outro lado, os pontos vermelhos indicam um comportamento atípico, sugerindo períodos com alta ocorrência. Essas observações podem ser explicadas pelos fenômenos meteorológicos que ocorreram nos anos de 2021 e 2022. Em 2021, o estado de Santa Catarina foi afetado pelo fenômeno *La Niña*, resultando em calor intenso e seca severa, o que levou a um aumento significativo nas ocorrências de incêndios. No ano seguinte, 2022, o fenô-

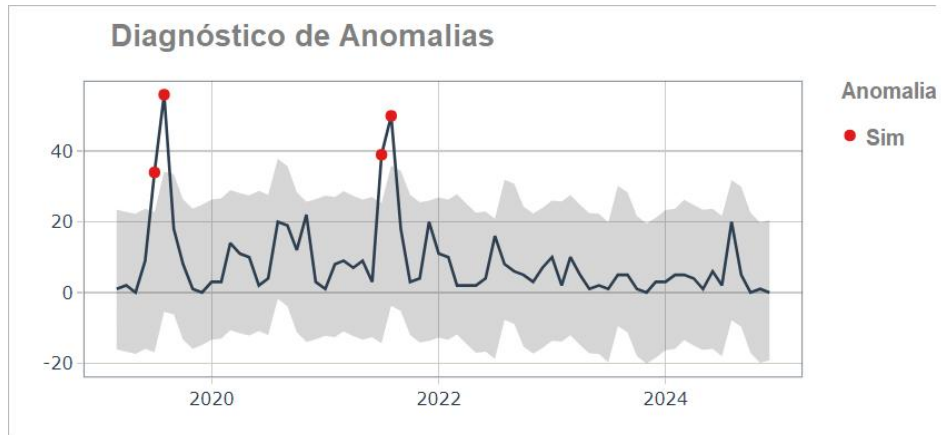


Figura 13 – Teste de anomalias no conjunto de dados da série temporal de Xanxerê (SC).

meno começou a perder intensidade, e a temporada de chuvas voltou ao normal, contribuindo para a redução das ocorrências de incêndios.

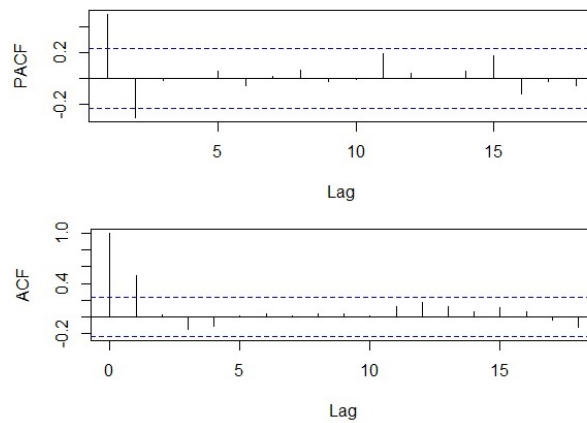


Figura 14 – Funções de Autocorrelação e autocorrelação parcial da série temporal de Lages.

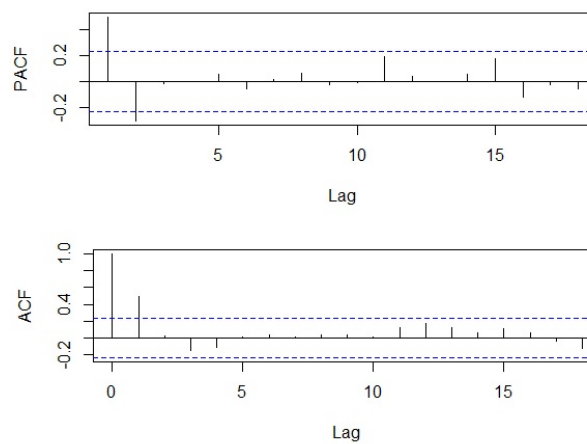


Figura 15 – Funções de Autocorrelação e autocorrelação parcial da série temporal de Xanxerê.

Nas figuras 14 e 15, o ACF apresenta uma queda rápida para valores próximos de zero, com oscilações entre valores positivos e negativos cujas magnitudes diminuem progressiva-

mente ao longo dos defasagens. Esse comportamento sugere que as séries temporais analisadas possuem características de estacionariedade. O teste KPSS foi aplicado para avaliar a estacionariedade das séries temporais. A estatística do teste foi de 0,1321 para Lages e 0,0505 para Xanxerê. Considerando os valores críticos de 0,119 (10%), 0,146 (5%), 0,176 (2,5%) e 0,216 (1%), conclui-se que a série de Xanxerê é estacionária com forte evidência, uma vez que o valor da estatística está abaixo de todos os níveis de significância. Para Lages, embora o valor ultrapasse o limiar de 10%, ele permanece abaixo dos níveis de 5% e mais estritos, indicando que a série pode ser considerada estacionária sob critérios mais conservadores.

3.4 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, serão detalhados os modelos utilizados para a previsão, a técnica de decomposição de sinal aplicada aos dados e as métricas selecionadas para avaliação do desempenho dos modelos. A análise incluirá uma descrição aprofundada dos métodos de previsão, destacando suas abordagens e configurações específicas. A técnica de decomposição de sinal será discutida em termos de sua contribuição para a extração de padrões relevantes, enquanto as métricas de desempenho serão explicadas para garantir uma compreensão clara dos critérios de avaliação e comparação entre os modelos.

3.4.1 Média Móvel Integrada Autorregressiva

O modelo ARIMA é uma abordagem estatística amplamente utilizada para previsão de séries temporais, com base em valores históricos. Ele combina três componentes: autorregressivo (AR), de diferenciação (I — *Integrated*) e de média móvel (MA), permitindo capturar tanto dependências temporais quanto tendências nos dados (BOX *et al.*, 2015; HYNDMAN; ATHANASOPOULOS, 2018).

A formulação geral do modelo ARIMA(p, d, q) é dada por:

$$\Phi(B)(1 - B)^d y_t = \Theta(B)a_t. \quad (1)$$

Onde

- y_t é o valor observado da série no tempo t ;
- a_t é um ruído branco (erro aleatório com média zero e variância constante);

- B é o operador de defasagem, tal que $By_t = y_{t-1}$;
- $(1 - B)^d$ representa a diferenciação de ordem d , responsável por tornar a série estacionária;
- $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o polinômio da parte autorregressiva (AR), de ordem p ;
- $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ é o polinômio da parte de média móvel (MA), de ordem q .

Neste estudo, foi utilizada a função `auto.arima` do pacote `forecast`, a qual seleciona automaticamente os parâmetros p , d e q com base em critérios de informação, como AIC, AICc e BIC.

3.4.2 Modelo Autorregressivo Integrado de Médias Móveis com Variáveis Exógenas

O modelo ARIMAX, conforme proposto por Bierens (1987), representa uma extensão do modelo ARIMA tradicional, permitindo a inclusão de variáveis exógenas (X) no processo de modelagem. Dado um processo de série temporal com $(k + 1)$ variáveis, no qual y_t e os k componentes de x_t são variáveis aleatórias com valores reais, o modelo ARIMAX assume uma estrutura capaz de capturar tanto a dependência temporal da variável endógena quanto a influência das variáveis externas.

Matematicamente, o modelo pode ser representado da seguinte forma:

$$\Phi(L)y_t = \Theta(L)\varepsilon_t + \Gamma(L)x_t. \quad (2)$$

Em que L é o operador defasador, tal que $L^s y_t = y_{t-s}$, ε_t representa o erro aleatório (ruído branco), e $\Phi(L)$, $\Theta(L)$ e $\Gamma(L)$ são polinômios nas defasagens de ordens p , q e r , respectivamente. O polinômio $\Phi(L)$ descreve a parte autorregressiva (AR), $\Theta(L)$ a parte de média móvel (MA), e $\Gamma(L)$ modela os efeitos das variáveis exógenas ao longo do tempo. A identificação do modelo requer, inicialmente, a verificação da estacionariedade da variável endógena, representada pela série temporal dos focos de incêndio, e das variáveis exógenas — temperatura, pluviosidade, umidade relativa do ar, velocidade do vento e o fenômeno La Niña (variável binária), bem como a definição apropriada das ordens p , q e r .

O modelo ARIMAX foi ajustado com a função `auto.arima()` do pacote `forecast`, a qual seleciona automaticamente os hiperparâmetros do modelo, incluindo as ordens p , d e q , bem como os componentes sazonais P , D , Q e m , com base nos critérios de informação AIC, AICc e BIC. As variáveis exógenas (`xreg`) incluíram defasagens das séries meteorológicas (temperatura, umidade, vento e precipitação), defasagens da própria variável resposta (focos de incêndio) e a variável binária associada ao fenômeno *La Niña*. A estimação dos parâmetros foi realizada por máxima verossimilhança, conforme a implementação padrão da função.

3.4.3 Regressão por Vetores de Suporte

O SVR é um algoritmo de aprendizado de máquina supervisionado usado principalmente para problemas de regressão. Esse método incorpora uma região de insensibilidade ao erro, conhecida como ϵ -tube, ao redor da função de previsão, permitindo uma maior tolerância a pequenos desvios nos dados (AWAD; KHANNA, 2015). Assim, é apresentado como um problema de otimização ao definir uma função de perda convexa insensível ao ϵ a ser minimizada, buscando o tubo mais plano que inclui a maior parte dos casos de treinamento.

A equação da função de regressão é dada por (SMOLA; SCHÖLKOPF, 2004):

$$f(x) = \langle w, x \rangle + b \quad (3)$$

Em que $f(x)$ representa a função de regressão, x são os dados de entrada, w é o vetor de pesos, $\langle w, x \rangle$ denota o produto escalar entre w e x , e b é o termo de interceptação.

Essa formulação foi originalmente proposta por Drucker *et al.* (1996), sendo posteriormente estendida por diversos autores. O problema de otimização correspondente pode ser expresso como:

$$\min_{w, b, \xi_i, \xi_i^*} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right\} \quad (4)$$

Sujeito às restrições:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (5)$$

Nessa formulação:

- $C > 0$ é um parâmetro que controla o equilíbrio entre a complexidade do modelo e a penalização dos erros;
- ϵ define a margem de tolerância no qual erros não são penalizados;
- ξ_i e ξ_i^* são variáveis de folga que permitem erros maiores que ϵ .

Essa estrutura permite ao SVR lidar com dados ruidosos e complexos, sendo apropriado para problemas de previsão com comportamento não linear.

3.4.4 *Extreme Gradient Boosting*

O XGBoost é uma técnica avançada baseada no princípio do *gradient boosting*, amplamente utilizada para tarefas de regressão e classificação em conjuntos de dados estruturados. Proposto por Chen e Guestrin (2016), o XGBoost incorpora mecanismos eficientes para lidar com grandes volumes de dados e evitar sobreajuste, como regularização explícita, paralelismo e tratamento de dados esparsos.

O XGBoost adota uma estratégia aditiva: a cada iteração k , uma nova árvore f_k é adicionada ao modelo para melhorar as previsões. O modelo final é composto por K árvores de decisão, conforme a seguinte equação:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F}, \quad (6)$$

onde:

- \hat{y}_i é a predição do modelo para a amostra i ;
- x_i é o vetor de características da amostra;
- f_k representa a k -ésima árvore de decisão pertencente ao espaço de funções \mathcal{F} , composto por árvores com estrutura e pontuação nos nós folha.

A função objetivo regularizada que o XGBoost busca minimizar é composta por dois termos: a função de perda e a regularização da complexidade do modelo:

$$\mathcal{L}(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (7)$$

onde: $l(y_i, \hat{y}_i)$ é uma função de perda convexa e diferenciável (por exemplo, erro quadrático médio) e $\Omega(f_k)$ é o termo de regularização dado por:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2, \quad (8)$$

em que:

- T representa o número de folhas da árvore f ;
- ω é o vetor de pontuações (valores) atribuídos às folhas;
- γ penaliza o número de folhas para evitar árvores muito complexas;
- λ regula a magnitude dos valores atribuídos às folhas.

Sheridan *et al.* (2016) destacam que, além da regularização, o XGBoost implementa outras estratégias eficazes para mitigar o sobreajuste e melhorar a performance:

- fShrinkage (redução da taxa de aprendizado): após cada iteração, os novos pesos são escalados por um fator $\eta \in (0, 1)$, permitindo um aprendizado mais gradual e robusto;
- Subamostragem de colunas: para cada árvore, apenas um subconjunto aleatório de variáveis é utilizado, o que reduz a correlação entre as árvores e acelera o treinamento;
- Tratamento eficiente de dados esparsos: com uma abordagem chamada *sparsity-aware split finding*, o algoritmo treina eficientemente mesmo em bases com grande quantidade de valores ausentes ou zeros, comuns em aplicações de Relações Quantitativas entre Estrutura e Atividade (QSAR).

Essas características fazem do XGBoost uma das ferramentas mais poderosas para previsão em dados estruturados, sendo amplamente utilizado em competições e aplicações científicas.

3.4.5 Floresta Aleatória

A Floresta Aleatória é um método de aprendizado supervisionado baseado na construção de um conjunto de árvores de decisão. Originalmente proposto por Breiman Breiman (2001), o modelo pode ser utilizado tanto para classificação quanto para regressão. No contexto

de regressão, o algoritmo busca prever uma variável de resposta contínua por meio da média das previsões geradas por diversas árvores independentes.

Formalmente, a floresta consiste em uma coleção de árvores de decisão $\{(x, \Theta_k), k = 1, \dots, K\}$, onde cada Θ_k representa um vetor de variáveis aleatórias que determinam a construção da k -ésima árvore. Essas variáveis são independentes e identicamente distribuídas. Para cada entrada x , obtém-se uma previsão $h_k(x)$ da árvore k , e a saída final da floresta é a média das previsões:

$$\hat{y}(x) = \frac{1}{K} \sum_{k=1}^K h_k(x). \quad (3)$$

Nesta equação:

- $\hat{y}(x)$ é o valor previsto para a entrada x ;
- $h_k(x)$ representa a saída da k -ésima árvore de regressão;
- K é o número total de árvores na floresta.

A aleatoriedade do modelo é introduzida em dois momentos: (i) na seleção de subconjuntos de dados por amostragem com reposição (bootstrap), e (ii) na escolha aleatória de subconjuntos de variáveis candidatas a cada divisão dos nós da árvore. Essa diversidade entre as árvores reduz a variância do modelo final e aumenta sua capacidade de generalização (BREIMAN, 2001; LIU *et al.*, 2012).

3.4.6 Máquina de Aprendizado Extremo

A Extreme Learning Machine (ELM) é um algoritmo de treinamento para redes neurais do tipo Perceptron de Múltiplas Camadas com uma única camada oculta, conhecidas em inglês como *Single-Layer Feedforward Networks* (SLFN), que são redes neurais feedforward com uma única camada intermediária de neurônios ocultos. A ELM converge muito mais rapidamente do que os métodos tradicionais e apresenta desempenho promissor (HUANG *et al.*, 2006). Nesta abordagem, os parâmetros dos nós ocultos são definidos de maneira aleatória, enquanto os pesos de saída são calculados de forma analítica.

A ELM é aplicada a SLFNs. Essas redes, com L nós ocultos, são representadas pelas

seguintes equações descritas por (HUANG *et al.*, 2006):

$$f_L(x) = \sum_{i=1}^L G_i(x, a_i, b_i) \cdot \beta_i, \quad a_i \in \mathbb{R}^d, \beta_i \in \mathbb{R}. \quad (4)$$

onde:

- $G_i(\cdot)$ é a função de ativação do i -ésimo nó oculto.
- a_i é o vetor de pesos (funções aditivas) ou centro (funções RBF) que conecta a camada de entrada ao i -ésimo nó oculto.
- b_i é o viés (funções aditivas) ou parâmetro de escala (funções de base radial, *Radial Basis Function* - RBF). No caso RBF, $b_i > 0$.
- β_i é o peso da saída.

Quando usamos uma função de ativação aditiva g , a função G_i é definida como:

$$G_i(x, a_i, b_i) = g(a_i \cdot x + b_i). \quad (5)$$

Quando utilizamos funções de ativação com base radial (RBF), a função G_i assume a forma:

$$G_i(x, a_i, b_i) = g(b_i \|x - a_i\|^2). \quad (6)$$

Em (6), $b_i > 0$ é um escalar que controla a largura da função radial (ex.: $g(z) = e^{-z}$ para Gaussianas).

3.4.7 Regressão Lasso

O Lasso, proposto por Tibshirani (1996), é um método de regressão que realiza seleção de variáveis e regularização via penalização ℓ_1 . Dados N pares de observações (\mathbf{x}_i, y_i) com preditores padronizados:

$$\sum_{i=1}^N x_{ij} = 0, \quad \frac{1}{N} \sum_{i=1}^N x_{ij}^2 = 1 \quad \text{para } j = 1, \dots, p, \quad (8)$$

o problema de otimização do Lasso é definido como:

$$\min_{(\beta_0, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}} \left[\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \lambda \|\boldsymbol{\beta}\|_{\ell_1} \right], \quad \lambda \geq 0, \quad (9)$$

onde $\|\beta\|_{\ell_1} = \sum_{j=1}^p |\beta_j|$ é a norma ℓ_1 dos coeficientes.

Extensão para Elastic Net:

O *Elastic Net* Zou e Hastie (2005) generaliza o Lasso adicionando uma penalização ℓ_2 :

$$P_\alpha(\beta) = (1 - \alpha) \frac{1}{2} \|\beta\|_{\ell_2}^2 + \alpha \|\beta\|_{\ell_1}, \quad \alpha \in [0, 1]. \quad (10)$$

Casos especiais:

- $\alpha = 1$: Reduz-se ao Lasso
- $\alpha = 0$: Equivale à regressão Ridge

3.4.8 Decomposição de Modo Variacional

A ideia central do, VMD é construir um modelo variacional e resolver o problema variacional para obter múltiplos componentes de modo (ZHAO *et al.*, 2024). Cada componente de modo possui uma largura de banda limitada pela frequência central. A VMD supera várias limitações de métodos tradicionais de decomposição de sinais. Ela é mais robusta ao ruído e a questões de amostragem, pois incorpora um termo de fidelidade aos dados, ajustável com base no nível de ruído (CUI *et al.*, 2024). Além disso, a VMD evita o processo iterativo de filtragem presente na Decomposição Empírica de Modos (EMD, do inglês *Empirical Mode Decomposition*), que pode ser suscetível às condições iniciais e resultar em mistura de modos.

Conforme abordado por Ribeiro *et al.* (2024b), por meio do pré-processamento com VMD, a série temporal original é decomposta em um número pré-definido k de funções IMFs, ou funções de modo $u_k(t)$, descritas por:

$$u_k(t) = A_k(t) \cos(\phi_k(t)). \quad (9)$$

Onde $\phi_k(t)$ é uma função não-decrescente ao longo do tempo t (ou seja, $\phi_k'(t) \geq 0$), a envoltória $A_k(t)$ é não-negativa, e a frequência $\omega_k(t) = \phi_k'(t)$, que é instantânea, varia mais lentamente do que a fase $\phi_k(t)$. Nesse sentido, cada modo é comprimido ao redor do centro (ω_k), que é definido durante todo o processo de decomposição. As etapas essenciais para calcular a largura de banda dos modos são, conforme definido por (RIBEIRO *et al.*, 2024b):

1. Para cada modo u_k , calcula-se a transformada de Hilbert para obter um espectro de frequência unilateral do sinal original;

2. Ao reorganizar a frequência central aproximada, o espectro de cada modo é modulado para sua respectiva banda-base;
3. Com o sinal modulado, a largura de banda é estimada por meio de suavização Gaussiana. Assim, trata-se de um problema de variação restrita, apresentado a seguir:

$$\min_{(u_k, \omega_k)} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (10)$$

$$\text{sujeito a: } \sum_k u_k(t) = f.$$

Ao aplicar o VMD, conseguimos extrair informações relevantes que podem ser utilizadas pelos modelos de previsão. A separação dos dados em sub-séries facilita a identificação de padrões e tendências específicas, tornando as previsões mais precisas. Isso é especialmente importante em cenários complexos, como a previsão de incêndios florestais, onde variáveis climáticas e condições ambientais podem ter um impacto significativo.

A remoção de ruído em dados de séries temporais envolve a eliminação de perturbações indesejadas do sinal, enquanto se retém a informação útil subjacente. De acordo com Stefenon *et al.* (2025), a Transformada Wavelet decompõe uma série temporal em diferentes componentes de frequência. Ela também fornece informações sobre a localização temporal dessas frequências, o que pode ser útil para a remoção de ruído. Isso pode aprimorar o modelo com previsões mais precisas, superando os modelos padrões. Além disso, ao contrário da Transformada Wavelet, que utiliza escalas fixas de frequência, o VMD adota uma abordagem adaptativa para extrair múltiplos modos de oscilação, permitindo uma análise mais precisa do sinal e facilitando a distinção entre ruído e informação relevante. Isso torna o VMD uma ferramenta eficaz para melhorar a precisão de modelos preditivos, especialmente em dados complexos e não estacionários.

Quando combinados, eles reconstroem o sinal original, preservando uma largura de banda limitada em cada componente (RIBEIRO *et al.*, 2023). Entre as principais vantagens do VMD, destacam-se seu menor custo computacional, em termos de tempo e memória, e sua robustez contra o problema de mistura de modos (RIBEIRO *et al.*, 2023).

A técnica VMD foi aplicada em dois cenários distintos. No caso univariado, considerou-se apenas a série alvo (*FIRE*), adotando-se $K = 5$, de modo a extrair cinco modos decompostos. Esse valor foi definido com base em estudos prévios da literatura e em testes preliminares que garantiram uma boa separação dos componentes. No caso multivariado, que

incluiu variáveis exógenas (temperatura, umidade, vento, precipitação e a variável binária La Niña), o número de modos foi definido como $K = 3$, buscando evitar redundância entre os modos e preservar a interpretabilidade dos sinais.

Em ambos os cenários, os demais hiperparâmetros do VMD foram mantidos fixos em valores amplamente utilizados na literatura: $\alpha = 2000$ (penalização do ruído, conferindo suavidade aos modos), $\tau = 0$ (tolerância dual), inicialização aleatória ($init=1$), ausência de modo DC ($DC=FALSE$) e critério de convergência de 10^{-6} . Os modos resultantes foram então utilizados como entradas nos modelos de previsão, em conjunto com as defasagens da série original e das variáveis exógenas, conforme o caso.

3.4.9 Interpretação de Variáveis de Entrada Utilizando Valores Shapley

A Inteligência Artificial Explicável (XAI, do inglês *Explainable Artificial Intelligence*) é uma área da Inteligência Artificial que busca tornar os modelos mais transparentes e compreensíveis, por meio de técnicas e métodos desenvolvidos para esse propósito. Além de fornecer suporte às decisões geradas pelo modelo, a XAI permite identificar a importância e a contribuição de cada variável nos resultados previstos (ZUEGE *et al.*, 2025). Isso é particularmente relevante na utilização de algoritmos de caixa-preta, como SVR, XGBoost e redes neurais, que, embora altamente eficazes em termos preditivos, carecem de interpretabilidade intrínseca.

A aplicação de técnicas de Explicação em Inteligência Artificial (XAI) torna-se essencial em cenários onde decisões automatizadas impactam diretamente a sociedade, como na gestão de riscos ambientais e na previsão de incêndios florestais. Estudos recentes demonstram o uso de SHAP para interpretar modelos de suscetibilidade ao fogo na Austrália, destacando os fatores meteorológicos e topográficos que mais influenciam o risco (ABDOLLAHI; PRADHAN, 2023). Outro trabalho realizado na China combina XGBoost e SHAP para identificar variáveis-chave em áreas montanhosas vulneráveis, permitindo a construção de mapas de risco explicáveis (LIU *et al.*, 2024). Essa explicabilidade fortalece a confiança dos usuários e tomadores de decisão, além de viabilizar auditorias mais eficazes na identificação de vieses e inconsistências.

Dentre as principais técnicas de XAI, destaca-se o SHAP, que se fundamenta na Teoria dos Jogos Cooperativos, mais especificamente nos valores de Shapley. Esse método avalia a contribuição marginal de cada variável para a predição, considerando todas as possíveis combinações de variáveis. Assim, o SHAP fornece uma interpretação tanto global quanto local do

modelo, atribuindo importância às variáveis com base no impacto causal que exercem nas previsões. Além disso, ele é capaz de identificar apenas as características que realmente causam o comportamento anômalo em conjuntos de instâncias de ataque ou anomalia (ROSHAN; ZAFAR, 2021). Contudo, é importante destacar que a aplicação do SHAP pode demandar elevado custo computacional, especialmente em modelos complexos e bases de dados extensas.

Pode-se formalizar o cálculo dos valores SHAP, que são definidos como os valores de Shapley aplicados à função de expectativa condicional do modelo. A partir desse conceito, o SHAP quantifica a contribuição marginal de cada variável para uma predição específica, considerando todas as combinações possíveis de variáveis presentes e ausentes (LUNDBERG; LEE, 2017).

Conforme definido por Lundberg e Lee (2017), matematicamente, o valor SHAP para uma variável i é dado por:

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (11)$$

onde:

- N representa o conjunto total de variáveis de entrada;
- S é qualquer subconjunto de N que não inclui a variável i ;
- $f_S(x_S)$ corresponde à expectativa da predição do modelo, condicionada às variáveis no conjunto S ;
- $f_{S \cup \{i\}}(x_{S \cup \{i\}})$ representa a expectativa da predição quando a variável i é adicionada ao conjunto S .

Na prática, calcular $f_S(x_S)$ requer que o modelo consiga lidar com entradas contendo variáveis ausentes. Como isso não é possível na maioria dos modelos, essa expectativa é estimada a partir da média condicional dos resultados do modelo, dada a presença de S . Este conceito é expresso pela seguinte relação:

$$f(h_x(z')) = \mathbb{E}[f(z)|z_S] \quad (12)$$

Esse valor representa a expectativa da saída do modelo, condicionado às variáveis presentes no conjunto S , enquanto as variáveis complementares \bar{S} são tratadas como desconhecidas.

Quando se assume independência entre as variáveis, a expressão pode ser aproximada como:

$$\mathbb{E}[f(z)|z_S] \approx \mathbb{E}_{z_{\bar{S}}}[f(z)] \quad (13)$$

Além disso, sob a hipótese de que o modelo seja linear, é possível simplificar ainda mais o cálculo:

$$\mathbb{E}[f(z)|z_S] \approx f([z_S, \mathbb{E}[z_{\bar{S}}]]) \quad (14)$$

Essas aproximações são particularmente úteis para reduzir o custo computacional do cálculo dos valores SHAP, principalmente em modelos complexos com grandes quantidades de variáveis. Nesse trabalho será analisado a importância das variáveis climáticas, considerando uma abordagem multivariada, o que permitirá interpretar de forma mais abrangente as interações entre as variáveis no processo de predição.

3.4.10 Medidas de Desempenho

Para uma amostra de n observações reais $y = (y_1, y_2, \dots, y_n)$ e suas respectivas previsões $\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$, os resultados dos modelos foram avaliados por meio dos indicadores RMSE, sMAPE, MAE e SD. Essas medidas foram calculadas com base na comparação entre os valores previstos e observados de focos de incêndio, sendo utilizadas para quantificar o erro e a variabilidade das previsões.

Com isso, torna-se possível comparar o desempenho entre os diferentes modelos analisados, a fim de identificar aquele com maior precisão preditiva. Vale ressaltar que as unidades de medida do RMSE, MAE e SD correspondem à unidade da variável de resposta (número de focos de incêndio), enquanto o sMAPE é expresso em porcentagem, permitindo uma avaliação relativa dos erros ao longo do tempo.

A equação para o RMSE é definida da seguinte forma (WILLMOTT, 1981):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (15)$$

O sMAPE (HYNDMAN; KOEHLER, 2006) é expressa pela seguinte equação:

$$\text{sMAPE} = \frac{1}{n} \sum_{t=1}^N \frac{|F_t - A_t|}{(|A_t| + |F_t|)^2} \times 100\%. \quad (16)$$

A medida sMAPE é adotada devido à presença de resultados zero na série temporal, o que torna impossível calcular o MAPE.

O MAE mede a média das diferenças absolutas entre os valores reais e previstos, refletindo o erro médio sem considerar o sinal. A equação é definida da seguinte forma (HODSON, 2022):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (17)$$

Ao calcular o desvio padrão de uma amostra, utiliza-se esse valor como uma estimativa da variabilidade da população da qual a amostra foi retirada, descrevendo o grau de dispersão das medições (ALTMAN; BLAND, 2005).

A fórmula a seguir pode ser utilizada para calcular o desvio padrão de uma amostra:

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n - 1}}. \quad (18)$$

onde:

- Σ : Símbolo que representa a “soma”.
- e_i : O i -ésimo erro de previsão.
- \bar{e} : O erro médio.
- n : O tamanho da amostra.

3.4.11 Validação Cruzada e Janela Deslizante

A validação cruzada aplicada a séries temporais, também conhecida como *rolling window*, é uma abordagem essencial para avaliar o desempenho de modelos preditivos quando há dependência temporal entre as observações, conforme descrito por (HYNDMAN; ATHANASSOPOULOS, 2013). Essa técnica simula uma situação de previsão em tempo real, garantindo que, a cada iteração, o modelo seja treinado com dados passados e testado com dados futuros, respeitando a ordem cronológica e evitando o vazamento de informações.

Esse procedimento é particularmente recomendado para problemas em que a estrutura temporal afeta diretamente o comportamento da variável de interesse, como previsão de consumo energético, séries financeiras e ocorrência de incêndios florestais. Segundo Bergmeir *et al.* (2018), o uso de métodos tradicionais de validação cruzada com partições aleatórias, como o k-fold convencional, pode comprometer a validade da avaliação em séries temporais, uma vez que desconsidera a autocorrelação entre as observações.

Existem duas estratégias principais de validação cruzada para esse tipo de dado: a janela deslizante (*rolling window*) e a janela expansível (*expanding window*). Na abordagem de janela deslizante, o tamanho da amostra de treinamento permanece fixo e é deslocado ao longo do tempo, substituindo os dados mais antigos por novos a cada repetição. Já na abordagem de janela expansível, novos dados são adicionados ao conjunto de treinamento em cada iteração, aumentando gradualmente o volume de informações disponíveis para o ajuste do modelo. A escolha entre uma ou outra depende da natureza do processo gerador dos dados e da necessidade de adaptação a padrões temporais mais recentes (CERQUEIRA *et al.*, 2020).

3.5 METODOLOGIA

Esta seção apresenta a descrição detalhada das etapas realizadas no estudo, desde a análise exploratória dos dados até a aplicação dos modelos preditivos e a agregação final das previsões. São ressaltadas as técnicas empregadas, as abordagens adotadas e as justificativas para as escolhas metodológicas realizadas.

Etapa 1 - Análise exploratória dos dados.

O estudo é dividido em duas classes de modelos: aqueles que utilizam a decomposição por VMD e aqueles que não utilizam a decomposição. Para a análise dos dados, foram considerados três tipos de séries temporais: os dados originais e os dados transformados por meio de duas técnicas distintas. A Transformação 1 refere-se à aplicação do logaritmo natural ajustado, $\log(y + 1)$ (PINHEIRO; RODRIGUES, 2024b), enquanto a Transformação 2 corresponde à aplicação da raiz quadrada, $\sqrt{y + 0,5}$ (PINHEIRO; RODRIGUES, 2024b), sendo que y representa os valores da série temporal. Essas transformações foram escolhidas para lidar com a assimetria dos dados e a presença de valores baixos ou nulos, com o objetivo de estabilizar a variância e melhorar a performance dos modelos de previsão.

Transformação 1, por meio da aplicação do logaritmo natural, visa reduzir a assimetria da distribuição dos dados e estabilizar sua variância, tornando padrões e tendências mais discerníveis. Transformação 2, baseada na raiz quadrada, tem como objetivo principal suavizar flutuações extremas e minimizar o efeito de valores atípicos muito baixos ou próximos de zero. Ambas as técnicas são essenciais para preparar adequadamente séries temporais com distribuição não normal e heterogeneidade na variância, garantindo maior robustez e precisão em modelos preditivos ao mitigar distorções estatísticas e melhorar a generalização dos resultados. Após a aplicação dessas transformações, é necessário reverter o processo para obter as previsões na escala original da série temporal.

Etapa 2 - Decomposição da Série Temporal Utilizando VMD e Análise das Componentes.

O VMD pré-processa a série bruta em um número finito e pré-definido k de funções de modo. Esses modos capturam diferentes faixas de frequência dos dados, abrangendo tanto componentes de baixa quanto de alta frequência. Para os dados univariados — compostos exclusivamente pela série temporal de focos de incêndio — a VMD foi aplicada definindo-se $k = 5$ componentes principais. Já ao incorporar as variáveis climáticas exógenas, testaram-se diferentes valores de k , sendo $k = 3$ o parâmetro considerado ótimo, conforme indicado pela análise da entropia de Shannon. Em ambos os casos, a VMD gerou IMFs que capturam faixas de frequência distintas dos dados, facilitando a separação de padrões de curto e longo prazo para posterior modelagem. Posteriormente, essas componentes foram utilizadas como entradas para diferentes modelos preditivos: VMD-SVR, VMD-XGBoost, VMD-RF, VMD-ELM e VMD-LASSO.

Etapa 3 - Aplicação dos modelos de previsão para cada uma das componentes.

Os modos gerados pelo VMD foram normalizados para estabilizar as escalas entre os componentes. Em seguida, os dados foram divididos em conjuntos de treinamento (70%) e teste (30%) para avaliação dos modelos. Foi trabalhado com uma janela inicial de 90%, isto é, o modelo foi treinado com 44 dados e efetuou a previsão para um dado, calculando as medidas de desempenho. Posteriormente, o modelo foi treinado com 45 dados e as medidas de desempenho foram recalculadas, repetindo o processo até que todos os dados disponíveis fossem utilizados. Após a decomposição da série temporal nas suas componentes, as previsões foram geradas para cada um dos modos utilizando os modelos preditivos previamente definidos. Para cada componente, os modelos foram treinados

com os dados de treinamento e avaliados nos dados de teste, com o cálculo das medidas de desempenho a cada iteração. As previsões foram analisadas para horizontes de 1, 2 e 3 meses à frente. Observa-se que os melhores desempenhos foram alcançados com defasagens de 6 períodos para Lages e de 5 períodos para Xanxerê.

Etapa 4 - Cálculo das previsões de cada componente e agregação final.

Com as previsões obtidas para cada um dos modos, o processo de agregação foi realizado para recompor a previsão original da série temporal. A agregação consiste na soma das previsões individuais de cada componente, resultando em uma previsão global que reflete a combinação das diferentes faixas de frequência representadas pelas IMFs. Esse método permite capturar as dinâmicas complexas e as variações temporais da série original, utilizando as informações extraídas de cada componente de forma integrada.

A fórmula de agregação utilizada é a seguinte:

$$\hat{y}_t = \sum_{i=1}^k \text{IMF}_{i,t} \quad (19)$$

onde: \hat{y}_t é a previsão final para o instante t , $\text{IMF}_{i,t}$ é a previsão para a i -ésima componente (IMF), e k é o número total de componentes dos Modos obtidas pelo VMD. Dessa forma, a previsão final é uma combinação ponderada das previsões de cada componente, proporcionando uma estimativa mais robusta e precisa da série temporal original.

A estratégia recursiva, também conhecida como estratégia iterativa, foi empregada para gerar previsões de curto prazo, para 1, 2 e 3 meses à frente, conforme descrito no experimento. Inicialmente, um único modelo é ajustado para realizar a previsão de um passo à frente. Em seguida, o valor previsto é utilizado como entrada para o mesmo modelo, visando a previsão do próximo passo à frente. Nesse contexto, a estrutura de treinamento adotada neste artigo é apresentada da seguinte forma:

$$y(t+1, k) = f \{y(t, k), \dots, y(t - n_y + 1, k)\} + \epsilon \quad (20)$$

onde: f é a função associada ao modelo adotado na fase de treinamento, $y(t)$ são os dados de entrada (valores passados da série temporal), n_y é o número de defasagens (onde $n_y = 6$ e $n_y = 5$, devido aos melhores resultados obtidos), h é o número de passos à frente (1, 2 ou 3), e k refere-se ao componente k do VMD. Para nosso caso, foi definido $k = 5$

- apenas a série temporal e $k = 3$ série temporal + variáveis climáticas . Além disso, ϵ é o erro aleatório, que deve ser independente e identicamente distribuído. Para prever até 3 meses a frente, as seguintes estruturas são consideradas:

$$y(t+h, k) = \begin{cases} f[y(t), y(t-1), \dots, y(t-n_y+h)] & \text{se } h = 1 \\ f[\hat{y}(t), y(t-1), \dots, y(t-n_y+h)] & \text{se } h = 2 \\ f[\hat{y}(t), \hat{y}(t-1), \dots, y(t-n_y+h)] & \text{se } h = 3. \end{cases} \quad (21)$$

Etapa 5 - Aplicação das Transformações Inversas.

Após a aplicação dessas transformações, é necessário reverter o processo para obter as previsões na escala original da série temporal. Para isso, aplica-se a transformação inversa correspondente a cada predição \hat{y}_t : (a) se foi utilizada a raiz quadrada, a reversão é feita por $\hat{y}_t^* = (\hat{y}_t)^2$; (b) se foi aplicada a transformação logarítmica, utiliza-se $\hat{y}_t^* = \exp(\hat{y}_t) - 1$.

Etapa 6 - Avaliação Preditiva e Visualização.

Nesta etapa, o desempenho dos modelos preditivos é avaliado por meio do RMSE, MAE, SMAPE e SD dos erros residuais. Essas medidas permitem quantificar a acurácia e a estabilidade das previsões. Além disso, são geradas visualizações para facilitar a análise qualitativa dos resultados, incluindo gráficos que comparam a série observada com os valores previstos transformados de volta à escala original (\hat{Y}^*), bem como a evolução dos erros ao longo do tempo.

Etapa 7 - Análise de Interpretabilidade com Valores de Shapley.

Após a seleção do melhor modelo multivariado para cada localidade analisada, os valores de SHAP foram calculados com o objetivo de interpretar as previsões realizadas. Para isso, foram gerados gráficos de resumo (SHAP summary plots), que apresentam visualmente a importância relativa das variáveis envolvidas no processo preditivo. Ao final, o modelo retorna o vetor de previsões $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$, correspondente aos valores estimados para o horizonte definido.

A Figura 16 ilustra o framework metodológico proposto neste estudo, o qual abrange as duas abordagens avaliadas: com e sem a aplicação da decomposição por VMD. O diagrama evidencia como o VMD é incorporado ao processo de modelagem, permitindo a decomposição

da série temporal em componentes mais informativos e distintos em termos de frequência. Essa etapa visa aprimorar a representação dos dados, favorecendo a captura de padrões relevantes e, conseqüentemente, melhorando o desempenho preditivo dos modelos na tarefa de previsão de focos de incêndios florestais. Complementarmente, a Figura 17 apresenta, de forma estruturada, o algoritmo proposto, destacando a sequência das etapas realizadas, que abrangem o pré-processamento dos dados, a decomposição por VMD, a aplicação dos modelos de previsão e, por fim, a agregação dos resultados das componentes.

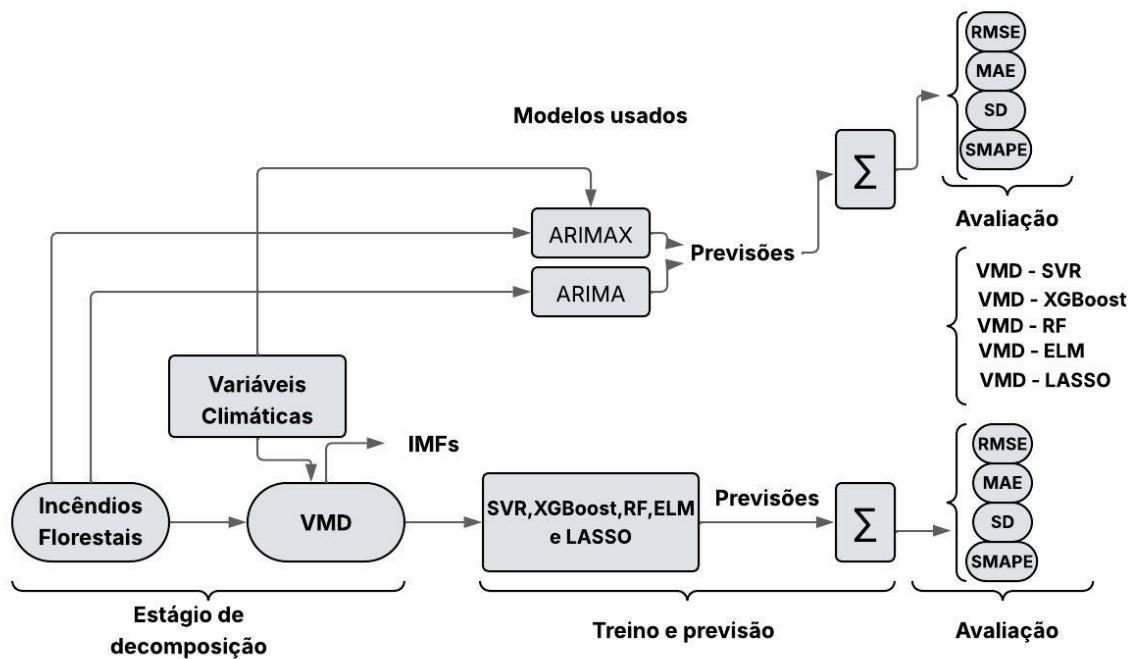


Figura 16 – Framework do estudo.

3.5.1 Implementação dos Modelos e Ajuste de Hiperparâmetros

Nesta seção, são detalhados os modelos utilizados para previsão, a técnica de decomposição de sinal aplicada aos dados e as medidas de performance selecionadas para avaliação do desempenho. Os experimentos foram conduzidos na linguagem R (versão 4.3.1), utilizando o ambiente RStudio e pacotes específicos para cada abordagem. Os modelos de aprendizado de máquina foram implementados com o pacote `caret`, utilizando validação cruzada k -fold ($k = 5$) e busca em grade (`tuneLength=5`) para ajuste dos hiperparâmetros, enquanto as variáveis foram previamente centralizadas e escaladas (`center` e `scale`) para garantir comparabilidade entre os preditores. O conjunto de modelos incluiu SVR com kernel radial, XG-

Algorithm 1 Procedimento Metodológico com Decomposição VMD

```

1: Entrada: Série temporal  $Y = \{y_1, y_2, \dots, y_n\}$ 
2: Saída: Previsões agregadas  $\hat{Y}$ 
3: Etapa 1: Análise Exploratória e Transformação dos Dados
4: for cada série  $y_t \in Y$  do
5:   Aplicar Transformação 1:  $y'_t = \log(y_t + 1)$ 
6:   Aplicar Transformação 2:  $y''_t = \sqrt{y_t + 0.5}$ 
7: end for
8: Etapa 2: Decomposição com VMD
9: for cada série transformada  $y_t$  do
10:   Definir  $k$ : número de componentes
11:   Aplicar VMD  $\Rightarrow$  obter modos  $IMF_1, IMF_2, \dots, IMF_k$ 
12: end for
13: Etapa 3: Aplicação dos Modelos de Previsão
14: for cada modo  $IMF_k$  do
15:   Normalizar  $IMF_k$ 
16:   Dividir dados: Treinamento (70%) e Teste (30%)
17:      $\triangleright$  As entradas dos modelos podem ser:
18:     (a) apenas lags da variável original, ou
19:     (b) lags + variáveis exógenas (climáticas)
20:   for cada modelo preditivo (SVR, XGBoost, RF, ELM, LASSO) do
21:     Treinar modelo em  $IMF_k$ 
22:     Avaliar modelo (Rolling Window com  $nlags = 5$  ou  $6$ )
23:     Obter previsão  $\hat{IMF}_{k,t+1}$ 
24:   end for
25: end for
26: Etapa 4: Agregação das Previsões
27: for cada instante  $t$  do
28:   Calcular  $\hat{y}_t = \sum_{k=1}^K \hat{IMF}_{k,t}$ 
29: end for
30: Etapa 5: Aplicação das Transformações Inversas
31: for cada previsão  $\hat{y}_t$  do
32:   Aplicar a transformação inversa correspondente:
33:   (a) Se foi usada raiz quadrada:  $\hat{y}_t = (\hat{y}_t^*)^2 - 0.5$ .
34:   (b) Se foi usada log:  $\hat{y}_t = \exp(\hat{y}_t^*) - 1$ .
35: end for
36: Etapa 6: Avaliação Preditiva e Visualização
37: Calcular medidas de desempenho: RMSE, MAE, sMAPE, SD
38: Gerar gráficos:
39:   - Série observada vs previsão  $\hat{Y}^*$ 
40:   - Série de erros ao longo do tempo
41: Etapa 7: Análise de Interpretabilidade com Shapley Values
42: for Cada modelo considerado do
43:   Calcular valores de Shapley para entrada de predição
44:   Gerar gráfico de importância das variáveis (SHAP summary plot)
45: end for
46: Retornar:  $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$ 

```

Figura 17 – Fluxo do procedimento metodológico proposto.

Boost, Random Forest, ELM e LASSO. Cada modelo teve seus hiperparâmetros ajustados de acordo com a abordagem própria: SVR explorou o parâmetro de penalização C e a largura do kernel, XGBoost ajustou número de árvores, profundidade máxima, taxa de aprendizado e regularização, Random Forest avaliou número de árvores e variáveis por divisão, ELM definiu número de neurônios e função de ativação, e LASSO selecionou a regularização λ dentro de uma grade predefinida com $\alpha = 1$. Os valores finais dos hiperparâmetros foram obtidos a partir dos modelos treinados utilizando a função `bestTune` do pacote `caret`, correspondendo ao conjunto que apresentou melhor desempenho nas métricas de validação.

3.6 RESULTADOS

Nesta seção, são apresentados os resultados dos quatro experimentos, organizados por cidade. As análises consideram, para cada horizonte de previsão, duas configurações principais: (i) abordagem univariada, utilizando apenas a série temporal dos focos de incêndio, tanto na forma original quanto combinada com a decomposição VMD; e (ii) abordagem multivariada, que incorpora variáveis climáticas como preditores adicionais, também avaliadas com e sem o pré-processamento por VMD. Dessa forma, são comparados os desempenhos dos modelos em quatro cenários: univariado sem VMD, univariado com VMD, multivariado sem VMD e multivariado com VMD.

3.6.1 Resultados em Lages

Nas subseções seguintes, os resultados das medidas de desempenho para 1, 2 e 3 meses à frente são analisados considerando as abordagens propostas.

3.6.1.1 Previsão de Incêndios Florestais para um Mês à Frente (H=1)

As Tabelas 4 e 5 sintetizam os resultados obtidos pelos modelos avaliados na previsão de focos de incêndio com horizonte de 1 mês à frente, considerando distintas configurações de entrada e transformações nos dados.

Na Tabela 4, são apresentados os desempenhos dos modelos aplicados sem o pré-processamento por decomposição VMD. Para a configuração baseada exclusivamente na série temporal (Univariado), o modelo ARIMA, com transformação logarítmica $\log(y+1)$, apresen-

tou a melhor performance, com RMSE de 2,0605 e MAE de 1,8195, demonstrando adequada capacidade preditiva ao considerar apenas a estrutura histórica da série. Entre os métodos de aprendizado de máquina, destaca-se o SVR, também com transformação logarítmica, que obteve desempenho competitivo (RMSE = 2,7065). Além dessas métricas, é importante destacar que o modelo ARIMA apresentou um dos menores valores de Desvio Padrão (SD = 0,8457) entre os métodos testados, evidenciando maior consistência e estabilidade nas previsões realizadas.

Com a incorporação de variáveis climáticas (Multivariado), o desempenho dos modelos apresentou variações expressivas. Modelos tradicionais, como o ARIMAX, não evidenciaram ganhos significativos, enquanto o ELM, utilizando a mesma transformação logarítmica, apresentou a menor MAE (1,7046) e sMAPE de 76,38%, indicando que a introdução de variáveis exógenas demanda modelos capazes de capturar interações mais complexas entre os fatores climáticos e a série temporal. No entanto, o ELM apresentou um SD de 1,1594, valor mais elevado em relação ao ARIMA, sugerindo maior variabilidade entre as previsões e, conseqüentemente, menor estabilidade preditiva.

Tabela 4 – Desempenho de modelo único para previsão de um mês à frente.

Entrada	Transformação	Critério	SVR	XGBoost	RF	ELM	LASSO	ARIMA/ARIMAX
Univariado	Sem Transf.	RMSE	3,2506	5,2044	4,7773	3,4535	5,0266	4,9266
		MAE	2,7971	4,1810	4,2337	2,5765	4,6675	4,4980
		sMAPE	85,1143	107,1687	99,2418	118,4081	103,3154	126,0169
		SD	2,3458	4,0839	2,7826	3,4814	2,5081	2,3208
	Log(y+1)	RMSE	2,7065	3,7916	3,1188	9,3915	2,8990	2,0605
		MAE	2,3522	3,1567	2,8797	7,1938	2,4589	1,8195
		sMAPE	58,9033	60,8259	65,2524	95,7568	59,8286	101,2545
		SD	2,2701	2,9511	2,6100	7,3710	2,3151	1,1166
	Sqrt(y+0.5)	RMSE	2,8400	5,0432	3,7603	3,0780	3,5596	3,2522
		MAE	2,4295	4,3642	3,3960	2,7106	3,1493	3,0307
		sMAPE	81,5145	104,7289	92,7226	84,0614	88,8379	148,7102
		SD	2,2652	4,0954	2,6743	2,4315	2,2873	2,3586
Multivariado	Sem Transf.	RMSE	5,2975	3,0948	12,2358	10,4901	5,0274	8,5337
		MAE	4,7212	2,6411	11,5228	9,3207	4,5215	7,4375
		sMAPE	99,1088	142,7877	132,1098	120,1576	99,2745	86,0851
		SD	2,6411	16,6602	4,2173	5,4249	2,2522	6,1496
	Log(y+1)	RMSE	5,5317	5,0012	4,1861	2,2934	2,8380	2,8304
		MAE	4,7581	3,9657	3,7190	1,7046	2,5204	2,2419
		sMAPE	99,3315	85,0895	90,7257	76,3767	78,9977	108,1683
		SD	3,1306	3,5752	2,4163	2,3097	2,3503	3,2123
	Sqrt(y+0.5)	RMSE	5,8086	9,9857	6,6929	3,8056	3,5510	3,2818
		MAE	5,2212	8,1753	6,1963	3,2403	3,1759	2,6124
		sMAPE	103,2015	114,5592	109,3992	94,4047	86,2778	142,7888
		SD	2,8770	5,9059	2,8848	3,7325	2,3098	3,1984

A Tabela 5 reporta os resultados obtidos mediante a aplicação do pré-processamento por decomposição VMD. De modo geral, a utilização do VMD promoveu melhorias substanciais na performance dos modelos, especialmente quando considerada apenas a série temporal (Univariado + VMD). Nesse cenário, o modelo VMD-LASSO com transformação log(y+1) al-

cançou o melhor desempenho, com RMSE de 1,2074, MAE de 0,9675 e sMAPE de 58,90%, indicando um ajuste mais acurado e estável à escala original dos focos de incêndio. Ademais, o VMD-LASSO apresentou o menor SD = 0,4917) dentre todos os modelos analisados, reforçando sua elevada estabilidade e confiabilidade para aplicações em cenários de previsão mensal.

A comparação entre os melhores modelos com e sem VMD evidencia uma redução expressiva nas métricas de erro: a aplicação do VMD-LASSO com transformação $\log(y+1)$ resultou em uma diminuição de aproximadamente 41,4% no RMSE e 46,8% no MAE em relação ao melhor modelo sem decomposição (ARIMA com $\log(y+1)$). Além disso, o sMAPE apresentou uma redução de cerca de 21,5 pontos percentuais, reforçando a superioridade do pré-processamento via VMD na melhoria da acurácia preditiva. Essa superioridade também se evidencia pela redução do SD, que passou de 0,8457 (ARIMA) para 0,4917 (VMD-LASSO), indicando que, além de mais preciso, o modelo pré-processado é também mais consistente.

Por outro lado, quando as variáveis climáticas foram integradas ao modelo juntamente com a decomposição VMD (Multivariado + VMD), observou-se uma degradação no desempenho preditivo em comparação ao uso exclusivo da série temporal com VMD. Esse comportamento sugere que a introdução de variáveis exógenas pode ter acrescentado ruído ou complexidade adicional não suficientemente capturada pelos modelos utilizados. Tal efeito também se reflete no aumento do SD, com valores superiores a 0,9000, indicando maior dispersão e incerteza nas previsões.

Tabela 5 – Desempenho de modelos híbridos para previsão de um mês à frente.

Modelo de Entrada	Transformação	Critério	VMD-SVR	VMD-XGBoost	VMD-RF	VMD-ELM	VMD-LASSO
Univariado+ VMD	Sem Transformação	RMSE	1,5919	2,9071	3,1454	1,7334	1,5248
		MAE	1,4004	2,6174	2,8159	1,5008	1,3596
		sMAPE	68,0135	88,6548	87,9530	69,3955	67,7893
		SD	1,4715	2,3164	2,2434	1,5341	1,4683
	Log(y+1)	RMSE	1,2094	2,2951	2,2799	1,5132	1,2074
		MAE	0,9760	1,8733	1,8712	1,3221	0,9675
		sMAPE	59,1241	77,6037	77,5433	67,6093	58,8999
		SD	1,2424	1,9182	1,8635	1,4240	1,2401
	Sqrt(y+0.5)	RMSE	1,1488	2,3313	2,8031	1,2334	1,1377
		MAE	0,9381	2,0522	2,3183	1,0187	0,9353
		sMAPE	59,1924	81,6217	83,6387	61,1639	59,0311
		SD	1,1497	1,7959	1,8028	1,1873	1,1504
Multivariado+ VMD	Sem Transformação	RMSE	7,5408	9,0245	7,5790	7,7225	7,5627
		MAE	6,2574	7,2609	6,3808	6,4523	6,3119
		sMAPE	109,6111	131,7786	111,8876	110,9032	109,8808
		SD	5,6097	7,7550	5,5802	5,6583	5,5735
	Log(y+1)	RMSE	4,9505	5,2354	5,0484	5,3153	5,0559
		MAE	3,8779	4,0914	3,9398	4,1801	3,9745
		sMAPE	91,4644	93,5736	91,9044	93,3710	92,2550
		SD	4,4970	4,6974	4,5796	4,7331	4,5624
	Sqrt(y+0.5)	RMSE	5,7275	5,9910	5,9131	5,9648	5,8240
		MAE	4,6796	4,8384	4,7757	4,9405	4,7706
		sMAPE	94,7044	93,4773	93,9233	97,1606	95,2021
		SD	4,7206	4,8955	4,8543	4,7945	4,7578

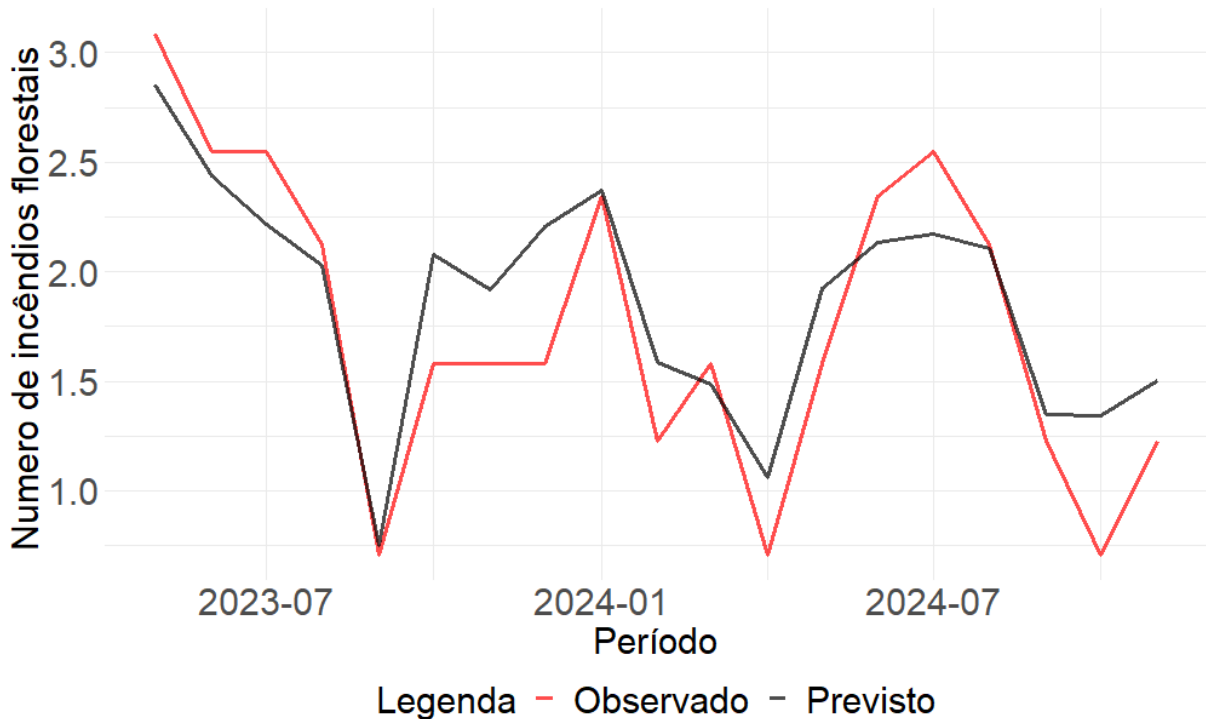


Figura 18 – Gráfico do modelo VMD-LASSO com dados em Sqrt para H=1.

Em síntese, os resultados indicam que, para a série temporal isolada, a combinação do pré-processamento por VMD com modelos lineares regulares (como LASSO e SVR), associada à transformação logarítmica dos dados, proporciona a configuração mais robusta, equilibrando erro e estabilidade na previsão mensal de focos de incêndio. Nessa perspectiva, o modelo VMD-LASSO 18 com $\log(y+1)$ destaca-se como a abordagem mais promissora para o horizonte H=1, não apenas pelos menores valores de erro, mas também pela menor variabilidade observada, conforme evidenciado pelo SD.

3.6.1.2 Previsão de Incêndios Florestais para Dois Meses à Frente (H=2)

As Tabelas 6 e 7 sintetizam os resultados obtidos pelos modelos de previsão de focos de incêndio para o horizonte de 2 meses, considerando distintas configurações de entrada e diferentes transformações de dados.

Na configuração baseada exclusivamente na série temporal (Univariado), o modelo SVR com transformação logarítmica $\log(y+1)$ destacou-se, apresentando os menores valores de RMSE (2,7786) e MAE (2,4362), superando inclusive métodos tradicionais como o ARIMA. Esses resultados demonstram a capacidade dos modelos de aprendizado de máquina em capturar padrões temporais complexos, especialmente quando associados a transformações adequadas.

Por outro lado, o modelo ARIMA com transformação de raiz quadrada $\sqrt{y + 0,5}$ apresentou a menor variabilidade (SD = 0,7501), indicando maior estabilidade nas previsões, embora tenha registrado um sMAPE elevado (141,5173).

Quando incorporadas variáveis climáticas (Multivariado), o modelo ELM, sem qualquer transformação, revelou-se o mais eficiente, atingindo um RMSE de 3,1084 e o menor MAE absoluto (2,1924) entre todos os cenários analisados. Além disso, apresentou um sMAPE (61,9637) inferior aos de outros modelos, como o LASSO (100,0505) e o XGBoost (104,8838), sugerindo que a arquitetura ELM é particularmente apta a integrar variáveis exógenas sem a necessidade de transformações prévias.

Tabela 6 – Desempenho de modelo único para previsão de dois meses à frente.

Entrada	Transformação	Critério	SVR	XGBoost	RF	ELM	LASSO	ARIMA/ARIMAX
Univariado	Sem Transf.	RMSE	3,3462	5,0822	4,5036	3,5184	5,4612	6,1150
		MAE	2,8722	4,0895	3,9806	2,6588	5,0160	5,8872
		sMAPE	85,6894	106,7063	97,6987	120,8878	105,0689	149,3552
		SD	2,4220	3,9940	2,6640	3,5730	2,7821	1,8114
	Log(y+1)	RMSE	2,7786	3,7726	3,0705	12,9885	3,0784	2,8723
		MAE	2,4362	3,0515	2,8564	8,9171	2,6319	2,6797
		sMAPE	60,1017	60,0011	65,2800	99,2336	62,1248	128,4170
		SD	2,3281	3,0157	2,5799	107,5712	2,4367	1,1327
	Sqrt(y+0.5)	RMSE	2,9158	5,1050	3,6259	3,4998	3,8646	3,9619
		MAE	2,5153	4,2096	3,2860	3,0243	3,3850	3,9023
		sMAPE	82,5752	100,9711	92,3673	87,0867	90,6178	141,5173
		SD	2,3299	4,3467	2,6161	2,6958	2,4862	0,7501
Multivariado	Sem Transf.	RMSE	5,7025	8,6425	8,7283	3,1084	5,8572	3,8131
		MAE	5,1134	6,9457	8,2695	2,1924	5,2988	2,3507
		sMAPE	98,3209	104,8838	117,7895	61,9637	100,0505	68,6395
		SD	2,6197	5,6824	2,8652	2,8475	2,5606	3,8546
	Log(y+1)	RMSE	5,9455	5,1711	3,2138	3,0635	2,9852	3,4173
		MAE	4,8107	3,9782	2,8730	2,3289	2,6352	2,4589
		sMAPE	90,3311	85,3023	78,9892	69,8721	76,0717	64,9639
		SD	3,8130	3,7816	2,5178	3,1090	2,5617	3,0359
	Sqrt(y+0.5)	RMSE	5,8731	10,3572	4,5665	3,1102	3,8579	3,5469
		MAE	5,1879	9,0275	3,9579	2,5759	3,4173	2,3409
		sMAPE	97,7034	116,5445	87,6640	87,8365	84,0658	61,3771
		SD	2,8244	5,6314	2,6797	3,1816	2,5784	3,2706

A aplicação da decomposição VMD à série temporal pura (Univariado + VMD) resultou em ganhos expressivos de acurácia. O modelo VMD-LASSO com transformação $\sqrt{y + 0,5}$ alcançou desempenho , com um RMSE de 1,2115, representando uma redução de 56,4% em relação ao melhor modelo sem VMD, e um MAE de 0,9836, o menor erro absoluto registrado. Além disso, apresentou um SD de 1,2169, indicando elevada estabilidade preditiva. Esse modelo também apresentou um sMAPE competitivo (59,8280), reforçando que a combinação entre decomposição de sinais, regularização e transformações matemáticas potencializa a captura de padrões subjacentes à série.

A combinação das variáveis climáticas com a decomposição VMD (Multivariado +

VMD) apresentou desafios consideráveis. Embora o modelo VMD-RF com $\log(y+1)$ tenha alcançado o melhor RMSE (5,0502) dentro dessa configuração, seu desempenho foi inferior ao obtido pelo ELM sem VMD (RMSE: 3,1084). Destaca-se também o modelo VMD-ELM com $\sqrt{y+0,5}$, cujo sMAPE foi de 95,8187, valor que confirma a tendência de degradação das métricas nesse cenário. O aumento médio de 62% no RMSE e a elevação da variabilidade ($SD > 4,6$) indicam que a introdução de variáveis exógenas acrescenta ruído que não é adequadamente mitigado pela decomposição VMD.

A análise comparativa permite estabelecer que o modelo VMD-LASSO com $\sqrt{y+0,5}$ mostrou-se superior para séries temporais puras, com reduções de 56,4% no RMSE e de 59,6% no MAE em relação ao SVR sem VMD. Nos cenários que incluem variáveis climáticas, as abordagens sem decomposição, notadamente o ELM sem transformação, mantêm uma vantagem expressiva, com um RMSE 38,4% menor que o do melhor modelo com VMD. Observa-se ainda que transformações logarítmicas beneficiam modelos que incorporam variáveis climáticas, enquanto transformações de raiz quadrada são mais adequadas em cenários que utilizam a decomposição VMD.

Tabela 7 – Desempenho de modelos híbridos para previsão de dois meses à frente.

MODELO DE ENTRADA	TRANSFORMAÇÃO	CRITÉRIO	VMD-SVR	VMD-XGBoost	VMD-RF	VMD-ELM	VMD-LASSO
Univariado + VMD	Sem Transformação	RMSE	1,6831	2,8542	3,1011	2,0838	1,5902
		MAE	1,4665	2,5551	2,7697	1,8265	1,4013
		sMAPE	68,2984	87,5697	87,3675	73,3792	67,6006
		SD	1,5276	2,2980	2,2234	1,7548	1,5090
	Log(y+1)	RMSE	1,2669	2,4743	2,3207	2,3549	1,2593
		MAE	1,0273	1,9848	1,9153	1,9007	1,0192
		sMAPE	60,0891	79,3414	78,1419	74,4228	59,9587
		SD	1,2991	2,0900	1,9152	2,0677	1,2934
	Sqrt(y+0.5)	RMSE	1,2443	2,4045	2,9216	1,6653	1,2115
		MAE	1,0037	2,0875	2,3899	1,2976	0,9836
		sMAPE	60,2476	82,4707	84,1559	64,7197	59,8280
		SD	1,2312	1,8331	1,9473	1,5359	1,2169
Multivariado + VMD	Sem Transformação	RMSE	7,3295	8,8104	7,3941	9,0003	7,4993
		MAE	5,8951	6,9076	6,2629	7,6636	6,0265
		sMAPE	102,6427	112,7964	110,9191	129,2060	101,9041
		SD	5,3034	7,7566	5,4282	7,4999	5,4678
	Log(y+1)	RMSE	5,6254	5,3144	5,0502	7,9627	5,8694
		MAE	4,2195	4,1555	3,8766	5,6153	4,4148
		sMAPE	90,1874	93,1470	89,9303	105,9968	93,0330
		SD	5,0874	4,8140	4,6016	7,2262	5,3056
	Sqrt(y+0.5)	RMSE	5,8041	6,0487	5,9461	6,7392	5,8785
		MAE	4,5621	4,8797	4,7728	5,2591	4,6336
		sMAPE	89,8350	94,0693	93,1952	95,8187	90,5984
		SD	4,6726	4,9116	4,8850	5,4879	4,6792

Dessa forma, para o horizonte de previsão $H = 2$ meses, a decomposição VMD consolida-se como uma ferramenta essencial apenas para configurações baseadas exclusivamente na série temporal, caso em que o modelo VMD-LASSO19 com $\sqrt{y+0,5}$ oferece precisão e estabilidade inigualáveis. Por outro lado, em cenários que incluem variáveis climáticas, abordagens diretas, como o ELM sem transformação, continuam sendo mais eficazes, suge-

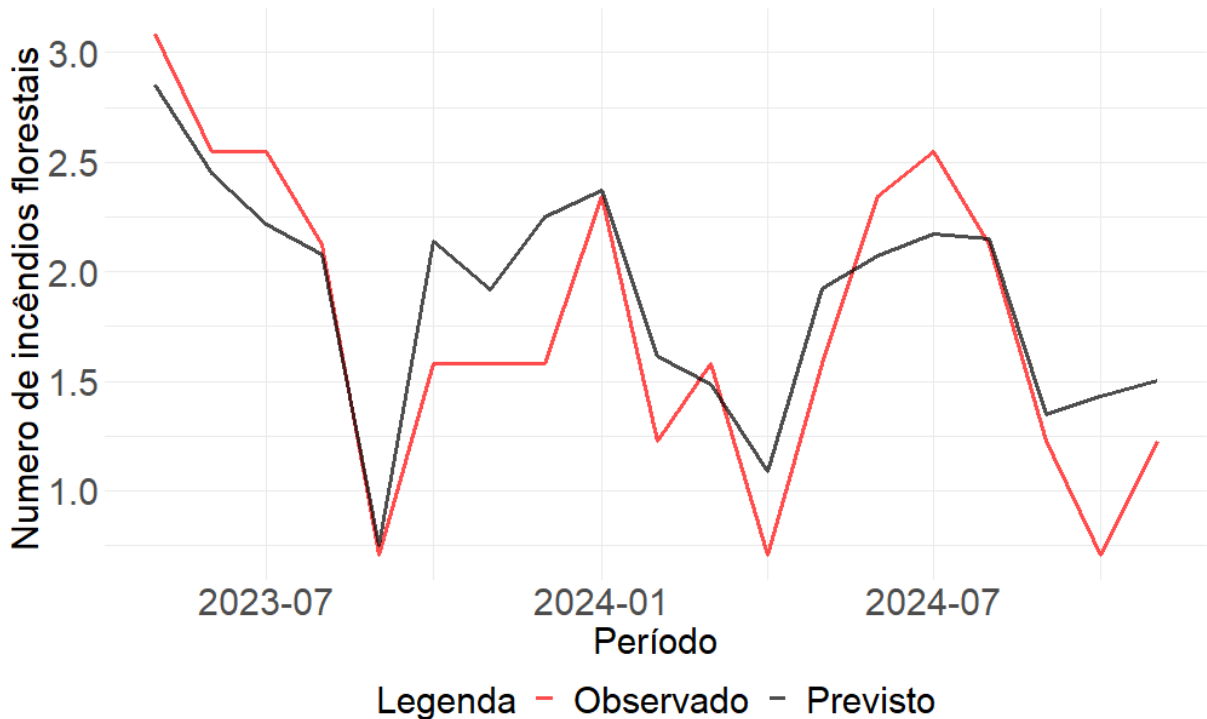


Figura 19 – Gráfico do modelo VMD-LASSO com dados em Sqrt para H=2.

rindo que a complexidade adicional introduzida pelo VMD não se traduz em ganhos preditivos quando há a presença de fatores exógenos.

3.6.1.3 Previsão de Incêndios Florestais para Três Meses à Frente (H=3)

As Tabelas 8 e 9 sintetizam os desempenhos dos modelos aplicados à previsão de focos de incêndio com horizonte de 3 meses, levando em conta distintas configurações de entrada e técnicas de transformação dos dados.

Na configuração baseada exclusivamente na série temporal, o modelo SVR com transformação logarítmica ($\log(y+1)$) destacou-se ao apresentar o menor RMSE (2,7824) e MAE (2,4731), evidenciando robustez na captura de padrões temporais de médio prazo. A aplicação da transformação logarítmica demonstrou ser eficaz na normalização da distribuição dos dados, contribuindo para a redução dos erros em comparação aos dados não transformados. Entretanto, o modelo ELM, sem qualquer transformação, apresentou o menor MAE absoluto (2,5684), embora tenha registrado um sMAPE elevado (121,3065), indicando uma maior sensibilidade a valores extremos.

Com a incorporação de variáveis climáticas à série temporal, o modelo ELM com a transformação $\log(y+1)$ emergiu como o mais preciso, alcançando o menor RMSE (3,1044)

e um sMAPE competitivo (74,5389). Destaca-se ainda o modelo LASSO com $\log(y+1)$, que obteve o melhor MAE (2,5590) e o menor sMAPE (70,5759) nessa categoria, evidenciando que a combinação entre regularização e transformação logarítmica potencializa a integração das variáveis exógenas. Por outro lado, o modelo ARIMAX com a transformação $\sqrt{y+0,5}$ apresentou a menor variabilidade ($SD = 2,6331$), embora tenha registrado desempenho inferior nas demais medidas de desempenho.

Tabela 8 – Desempenho de modelo único para previsão de três meses à frente.

Entrada	Transformação	Critério	SVR	XGBoost	RF	ELM	LASSO	ARIMA/ARIMAX
Univariado	Sem Transf.	RMSE	3,3945	4,3812	4,4526	3,3626	5,5077	6,0876
		MAE	2,9247	3,4897	3,9519	2,5684	5,1397	5,8340
		sMAPE	86,3201	92,8960	98,1953	121,3065	106,6284	148,5923
		SD	2,4148	3,4838	2,7419	3,3728	2,5738	1,9044
	Log(y+1)	RMSE	2,7824	4,1326	3,0396	9,1483	3,0677	3,1252
		MAE	2,4731	3,4542	2,8192	6,5518	2,6627	2,9023
		sMAPE	60,6517	66,3192	64,7762	91,6483	62,8265	128,3368
		SD	2,2873	3,2453	2,5960	7,7294	2,4009	1,2698
	Sqrt(y+0,5)	RMSE	2,9385	5,2300	3,6263	3,5491	3,8896	4,4822
		MAE	2,5599	4,3556	3,3149	3,1362	3,4226	4,3491
		sMAPE	83,3302	99,7417	93,7133	89,0955	91,3848	141,4900
		SD	2,3004	4,2809	2,6909	2,5074	2,4166	1,1875
Multivariado	Sem Transf.	RMSE	4,4344	20,1804	8,2310	4,3882	5,5142	4,6403
		MAE	3,9274	16,9893	7,4831	3,8988	4,9905	3,9336
		sMAPE	83,7500	131,4529	107,2244	83,4699	93,2225	95,8630
		SD	3,3755	11,3511	3,5171	2,9249	3,3438	2,6965
	Log(y+1)	RMSE	4,4698	5,9626	3,7611	3,1044	3,1805	4,2499
		MAE	3,7482	4,6330	3,1104	2,6085	2,5590	3,5497
		sMAPE	79,8668	84,6355	75,6645	74,5389	70,5759	92,7270
		SD	3,5011	4,9517	3,4070	3,1738	3,1848	2,7389
	Sqrt(y+0,5)	RMSE	4,3141	5,8919	4,7213	6,7578	3,6862	4,6595
		MAE	3,6942	4,8104	3,9024	5,5738	3,3195	3,9916
		sMAPE	80,6212	126,3337	81,4918	106,1843	79,6962	97,8428
		SD	3,4606	6,0297	3,7038	6,0749	3,1663	2,6331

A aplicação do pré-processamento por decomposição VMD na série temporal pura resultou em ganhos expressivos de desempenho. O modelo VMD-LASSO com a transformação $\sqrt{y+0,5}$ apresentou desempenho interessante, reduzindo o RMSE para 1.3069 (uma diminuição de 53,0% em relação ao melhor modelo sem VMD) e alcançando o menor MAE absoluto (1,0370), além de uma variabilidade muito baixa ($SD = 1,3114$), indicando alta estabilidade. O sMAPE obtido por essa configuração também foi competitivo (60,5776), reforçando a eficácia da sinergia entre a decomposição de modos, a regularização e as transformações não-lineares.

Entretanto, a integração de variáveis climáticas juntamente com o VMD apresentou desafios consideráveis. A combinação dessas variáveis com a decomposição (Multivariado + VMD) resultou em uma degradação do desempenho. O modelo VMD-RF com a transformação $\log(y+1)$ apresentou o melhor desempenho dentro dessa categoria, com RMSE de 4,1050 e MAE de 3,3530, mas ainda assim inferior ao modelo ELM sem VMD (RMSE de 3,1044).

Observa-se que, de forma geral, o sMAPE aumentou em média 17,4 pontos percentuais em relação aos modelos sem VMD, e a variabilidade (SD superior a 3,3) foi consistentemente mais elevada, indicando maior instabilidade nas previsões.

Tabela 9 – Desempenho de modelos híbridos para previsão de três meses à frente.

MODELO DE ENTRADA	TRANSFORMAÇÃO	CRITÉRIO	VMD-SVR	VMD-XGBoost	VMD-RF	VMD-ELM	VMD-LASSO
Univariado + VMD	Sem Transformação	RMSE	1,7930	2,8796	3,1212	2,5136	1,6817
		MAE	1,5606	2,5837	2,7846	2,1372	1,4967
		sMAPE	68,6449	87,9600	87,3108	75,3496	68,5364
		SD	1,6060	2,3092	2,2445	1,9703	1,5927
	Log(y+1)	RMSE	1,3402	2,4157	2,2737	2,4279	1,3456
		MAE	1,0506	1,9089	1,8485	2,1215	1,0517
		sMAPE	59,9765	76,9599	76,7180	78,5449	59,8951
		SD	1,3736	2,0037	1,8467	1,8895	1,3817
	Sqrt(y+0,5)	RMSE	1,3416	2,3490	2,7939	1,8445	1,3069
		MAE	1,0518	1,9428	2,2960	1,3943	1,0370
		sMAPE	61,0472	78,7611	83,0961	66,1882	60,5776
		SD	1,3163	1,7190	1,8252	1,6319	1,3114
Multivariado + VMD	Sem Transformação	RMSE	6,9984	8,0199	6,9881	13,6294	7,3902
		MAE	5,7824	6,9485	5,9450	10,9074	6,0445
		sMAPE	108,2520	121,9203	107,1662	133,3914	108,2326
		SD	4,8496	6,5300	4,7373	13,2882	5,2923
	Log(y+1)	RMSE	6,0377	4,2148	4,1050	23,0975	6,7068
		MAE	4,3142	3,7285	3,3530	10,6113	4,7074
		sMAPE	87,7457	93,1852	87,9200	113,2589	95,6152
		SD	5,0040	3,3129	3,4940	21,6831	5,6610
	Sqrt(y+0,5)	RMSE	5,7389	5,5325	5,1089	9,0888	5,6689
		MAE	4,7009	4,9197	4,4518	6,2311	4,7690
		sMAPE	95,9495	98,9630	95,4339	100,3713	97,0587
		SD	4,0254	3,8695	3,7906	7,4134	3,8474

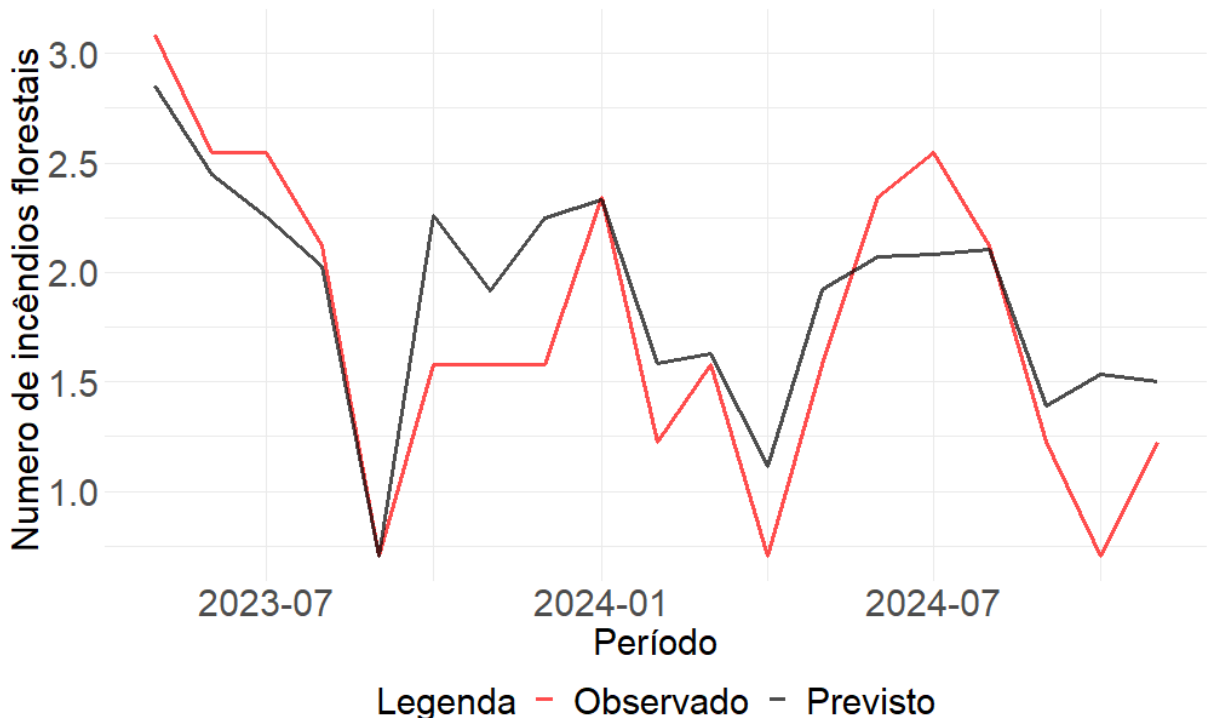


Figura 20 – Gráfico do modelo VMD-LASSO com dados em Sqrt para H=3.

Sendo assim, verifica-se que a decomposição VMD mostrou-se altamente benéfica para séries temporais puras, como demonstrado pelo modelo VMD-LASSO com $\sqrt{y+0,5}$, que reduziu o RMSE em 53,0% e o MAE em 58,1% em relação ao SVR sem VMD, além de

diminuir a variabilidade (SD) em 42,6%, confirmando uma maior estabilidade. Por outro lado, a aplicação do VMD em conjunto com variáveis climáticas revelou limitações importantes, com modelos que apresentaram, em média, RMSE 32,2% superior aos sem VMD, e com o VMD-RF mostrando um desempenho negativo em comparação ao ELM sem VMD, com um aumento de 32,2% no RMSE. Em relação às transformações aplicadas, observa-se que a transformação $\log(y+1)$ foi a mais adequada para modelos que incorporam variáveis climáticas, como no caso do ELM sem VMD, enquanto a transformação $\sqrt{y+0,5}$ mostrou-se ideal para a combinação de VMD com séries temporais puras, como no VMD-LASSO.

Diante disso, conclui-se que, para o horizonte de previsão de 3 meses, a decomposição VMD é altamente eficaz quando aplicada exclusivamente às séries temporais puras, proporcionando ganhos substanciais em precisão e estabilidade, conforme evidenciado pelo desempenho superior do modelo VMD-LASSO 20 com $\sqrt{y+0,5}$. Em contrapartida, em cenários que incluem variáveis climáticas, abordagens diretas, como o ELM com $\log(y+1)$, mantêm uma vantagem crítica, sugerindo que a complexidade adicional introduzida pela decomposição não compensa, podendo inclusive intensificar o ruído quando fatores exógenos são incorporados. A degradação consistente do sMAPE (+17.4 pontos percentuais) e do SD (+24,1%) em modelos que combinam VMD e variáveis climáticas reforça a necessidade de explorar estratégias alternativas para a integração eficiente de dados climáticos.

3.6.2 Resultados em Xanxerê

Nas subseções seguintes, os resultados das medidas de desempenho para 1, 2 e 3 meses à frente são analisados considerando as abordagens propostas.

3.6.2.1 Previsão de Incêndios Florestais para um Mês à Frente (H=1)

As Tabelas 10 e 11 apresentam os resultados dos modelos de previsão de focos de incêndio para a região de Xanxerê, com horizonte de um mês, considerando abordagens univariadas e multivariadas, com e sem pré-processamento por VMD.

Sem o uso do VMD (Tabela 10), o modelo ARIMA com transformação logarítmica $\log(y+1)$ destacou-se na abordagem univariada, alcançando o menor RMSE (3,6019) e MAE (2,9197), com previsões estáveis (SD = 3,0809). Esse desempenho representou uma redução de 46.8% no RMSE em relação ao segundo melhor modelo, o XGBoost com $\sqrt{y+0,5}$. Na

abordagem multivariada, o RF com $\log(y+1)$ obteve os melhores resultados, com RMSE de 5,0700, MAE de 4,4498 e o menor sMAPE entre os multivariados (85,5676). Embora o ELM sem transformação tenha registrado o menor MAE absoluto (3,9015), apresentou maior RMSE (6,2672) e sMAPE elevado (105,0349).

Tabela 10 – Desempenho de modelo único para previsão de um mês à frente.

Entrada	Transformação	Critério	SVR	XGBoost	RF	ELM	LASSO	ARIMA/ARIMAX
Univariado	Sem Transf.	RMSE	8,5484	5,8637	7,8992	11,2104	7,1774	6,7827
		MAE	8,2622	4,7136	7,3846	9,7653	6,4340	4,2154
		sMAPE	126,0695	101,6382	119,8024	132,0961	115,6620	178,6449
		SD	4,6799	5,1897	5,0212	7,0949	5,2293	6,5524
	Log(y+1)	RMSE	8,4145	6,3257	5,7938	8,5188	6,1333	3,6019
		MAE	7,3020	4,9576	4,9372	7,3839	5,1520	2,9197
		sMAPE	97,5812	78,2233	83,2673	99,8601	82,9872	168,8973
		SD	6,0960	5,7996	4,8317	6,6359	5,2107	3,0809
	Sqrt(y+0,5)	RMSE	8,6808	5,2313	6,3187	9,7272	6,4071	6,0021
		MAE	8,0211	3,8027	5,5633	6,9072	5,3102	4,1228
		sMAPE	123,1073	90,6256	108,9790	118,7764	106,1790	176,8013
		SD	5,5298	4,9859	5,0120	9,1788	5,4108	5,5258
Multivariado	Sem Transf.	RMSE	5,2127	10,7329	7,6323	6,2672	9,5629	11,6134
		MAE	4,1572	6,6542	7,1295	3,9015	8,2860	8,1294
		sMAPE	100,2106	93,5049	110,6707	105,0349	124,9658	79,0099
		SD	5,1463	10,0228	4,8384	6,3906	8,6566	8,0412
	Log(y+1)	RMSE	8,8198	6,9044	5,0700	15,2217	9,2337	5,1365
		MAE	7,4544	5,8755	4,4498	14,1129	7,9232	4,2359
		sMAPE	107,6525	90,8631	85,5676	131,5069	111,5198	106,4646
		SD	7,2203	4,7982	4,1247	7,1459	6,9882	4,4168
	Sqrt(y+0,5)	RMSE	5,5590	5,1217	5,2794	12,7220	6,4582	5,5421
		MAE	3,9144	4,4078	4,5944	12,0905	5,2407	4,5399
		sMAPE	92,7138	95,9909	94,3397	130,0290	108,4461	110,5359
		SD	5,5738	4,6117	4,5294	4,0562	5,8479	4,8318

Com a aplicação do VMD (Tabela 11), o modelo VMD-LASSO 21, sem transformações, apresentou desempenho superior na abordagem univariada, com RMSE de 1,5248 e MAE de 1,3597, representando uma redução de 57,7% no RMSE em relação ao melhor modelo sem VMD. Além disso, apresentou a menor variabilidade ($SD = 1,4683$). Entretanto, transformações como $\log(y+1)$ e $\sqrt{y+0,5}$ degradaram seu desempenho, indicando que a série original é mais adequada para o VMD.

Na abordagem multivariada com VMD, o melhor resultado foi obtido pelo VMD-RF com $\log(y+1)$, mas com desempenho inferior aos modelos multivariados sem VMD. O RMSE aumentou 40,6% em relação ao RF sem VMD, e todas as configurações apresentaram sMAPE superiores a 91,67, além de maior variabilidade ($SD > 6,0$), indicando perda de precisão e estabilidade.

Em suma, o VMD demonstrou ser altamente vantajoso para séries temporais puras, reduzindo o erro em mais de 57% e garantindo máxima estabilidade, superando inclusive modelos tradicionais como o ARIMA. No entanto, mostrou-se contraproducente em abordagens multi-

Tabela 11 – Desempenho de modelos híbridos para previsão de um mês à frente.

MODELO DE ENTRADA	TRANSFORMAÇÃO	CRITÉRIO	VMD-SVR	VMD-XGBoost	VMD-RF	VMD-ELM	VMD-LASSO
Univariado + VMD	Sem Transformação	RMSE	1,5919	2,9071	3,1455	1,7334	1,5248
		MAE	1,4004	2,6174	2,8159	1,5008	1,3597
		sMAPE	68,0135	88,6548	87,9531	69,3955	67,7893
		SD	1,4715	2,3164	2,2435	1,5341	1,4683
	Log(y+1)	RMSE	3,2046	3,3382	3,5751	2,8744	3,2032
		MAE	1,6463	1,9817	2,2631	1,8540	1,6209
		sMAPE	68,3325	71,0297	75,5729	72,5334	67,9009
		SD	3,2772	3,4141	3,6030	2,9048	3,2685
	Sqrt(y+0,5)	RMSE	2,9435	3,6514	3,5981	2,9039	2,9211
		MAE	1,8003	2,6780	2,7220	1,8339	1,7713
		sMAPE	73,4091	81,2408	82,5749	74,3674	73,1516
		SD	3,0194	3,5143	3,3618	2,9628	3,0000
Multivariado + VMD	Sem Transformação	RMSE	9,0105	9,2548	9,2957	9,2060	9,1321
		MAE	8,0797	8,1858	8,3039	8,2278	8,1404
		sMAPE	114,4985	114,8223	115,9850	115,2395	113,9238
		SD	5,7948	5,9512	5,8841	5,8815	5,8607
	Log(y+1)	RMSE	7,1439	7,3024	7,1319	7,2068	7,2543
		MAE	5,5660	5,7633	5,5294	5,6260	5,6542
		sMAPE	93,2957	94,3841	91,6716	93,4833	93,5633
		SD	6,2229	6,2950	6,1824	6,2665	6,2902
	Sqrt(y+0,5)	RMSE	7,8588	8,1239	7,9316	7,9716	7,8853
		MAE	6,5753	6,7398	6,6543	6,6567	6,6089
		sMAPE	101,3418	102,0604	102,4872	102,0029	100,9379
		SD	6,0588	6,2438	6,0284	6,1287	6,0406

variadas, elevando em média 40% o RMSE e 25% o MAE, possivelmente devido à introdução de ruídos que os algoritmos não conseguiram mitigar.

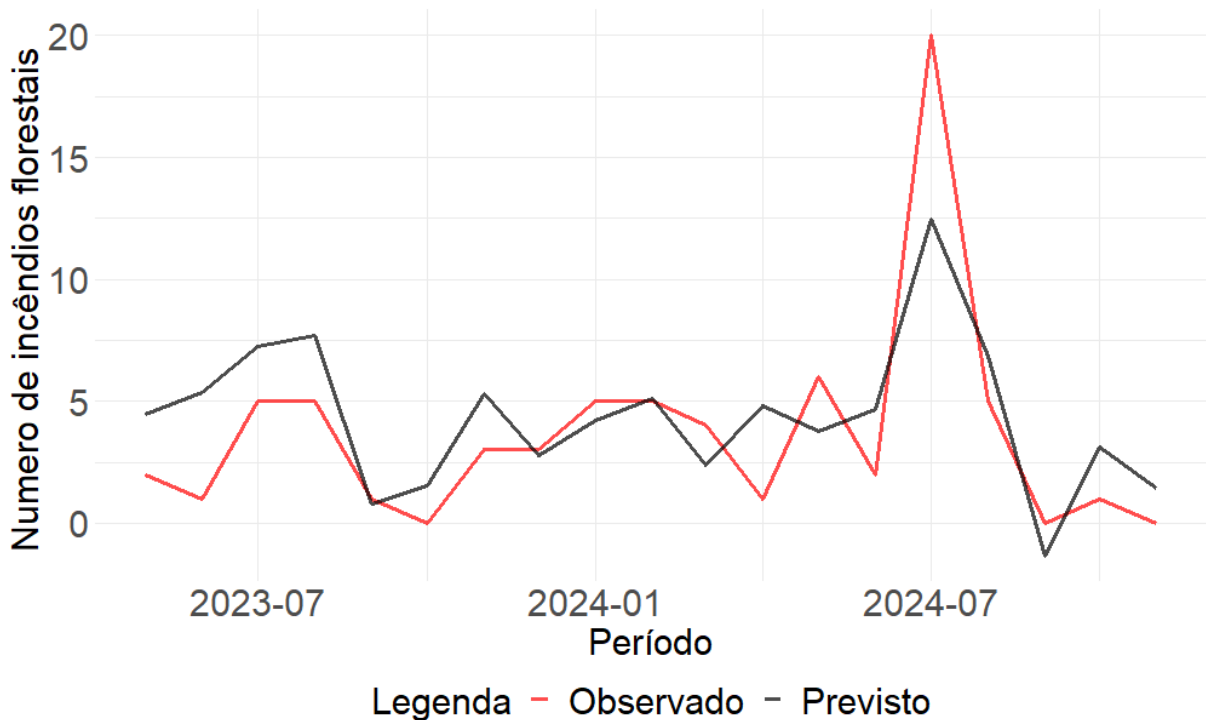


Figura 21 – Gráfico do modelo VMD-LASSO com dados normais para H=1.

3.6.2.2 Previsão de Incêndios Florestais para Dois Meses à Frente (H=2)

As Tabelas 12 e 13 apresentam os resultados dos modelos de previsão de focos de incêndio para a região de Xanxerê no horizonte de 2 meses, considerando abordagens univariadas e multivariadas, com e sem o pré-processamento VMD.

Sem a utilização do VMD, na abordagem univariada, o modelo XGBoost com transformação de raiz quadrada $\sqrt{y + 0,5}$ destacou-se por apresentar o menor MAE (3,6969) e um RMSE competitivo (5,3466). Esse modelo também apresentou um bom equilíbrio entre precisão e robustez, com sMAPE de 88,8822 e uma variabilidade moderada ($SD = 5,1548$), embora superior à de modelos como o ARIMA. Ainda nesta abordagem, o modelo ARIMA com $\log(y+1)$ obteve o menor RMSE (4,8709), mas apresentou MAE mais elevado (3,9343) e um sMAPE significativamente alto (135,6365), sugerindo maior sensibilidade a *outliers*.

Na abordagem multivariada, que considera a série temporal juntamente com variáveis climáticas, o modelo mais preciso foi o XGBoost sem transformação, registrando o menor RMSE (4,4526), um MAE excepcionalmente baixo (2,9776) e um sMAPE de 73,4312, indicando ótimo desempenho para dados multivariados. O modelo LASSO com $\log(y+1)$ também apresentou desempenho relevante, com RMSE de 5,1955 e MAE de 4,1124, embora inferior ao XGBoost. O RF com $\log(y+1)$ obteve um sMAPE competitivo (91,9623), mas com um RMSE superior (5,4644), indicando menor precisão absoluta.

Tabela 12 – Desempenho de modelo único para previsão de dois meses à frente.

Entrada	Transformação	Critério	SVR	XGBoost	RF	ELM	LASSO	ARIMA/ARIMAX
Univariado	Sem Transf.	RMSE	7,4121	5,7401	7,3370	6,1717	7,7781	7,3742
		MAE	7,2196	4,2295	6,7371	5,2954	7,1047	5,4583
		sMAPE	121,5396	96,9637	115,4884	108,2161	117,9783	145,6765
		SD	4,1871	5,4178	4,9962	5,3890	5,4026	5,8084
	Log(y+1)	RMSE	7,3324	5,9633	5,5264	9,2625	6,4943	4,8709
		MAE	6,6668	4,6908	4,6541	8,0721	5,6286	3,9343
		sMAPE	96,9412	77,8649	80,9863	103,3728	87,7965	135,6365
		SD	5,1719	5,5198	4,6998	7,0302	5,3154	3,5931
	Sqrt(y+0,5)	RMSE	7,5746	5,3466	6,1946	10,4805	6,6025	6,1193
		MAE	7,1828	3,6969	5,4444	8,0908	5,7532	4,7036
		sMAPE	121,2340	88,8822	108,0911	121,3953	108,9215	176,0389
		SD	4,7863	5,1548	4,9470	9,1358	5,3236	4,6921
Multivariado	Sem Transf.	RMSE	6,5057	4,4526	8,7366	6,0426	7,5145	9,2642
		MAE	5,8942	2,9776	8,3603	4,1933	6,2715	7,1875
		sMAPE	104,5074	73,4312	115,3429	195,5049	126,2348	89,4565
		SD	5,2256	4,5682	4,4775	4,4692	7,4089	9,2663
	Log(y+1)	RMSE	7,4112	10,2109	5,4644	7,4428	5,1955	7,4296
		MAE	6,3744	8,3194	4,6316	5,6323	4,1124	5,8564
		sMAPE	103,8635	107,4363	91,9623	101,3656	87,9354	111,8536
		SD	5,9402	7,3794	4,8536	7,2604	5,0127	7,6096
	Sqrt(y+0,5)	RMSE	6,6734	7,4663	6,1602	9,6541	5,7996	7,5211
		MAE	5,9516	5,4768	5,4675	8,3250	4,8217	5,6377
		sMAPE	105,0178	95,5539	98,9577	108,8597	96,1109	117,1792
		SD	5,4472	7,2505	5,0123	6,7100	5,3437	7,2863

Ao se aplicar o pré-processamento VMD, observou-se um impacto expressivo na abordagem univariada. O modelo VMD-LASSO, sem qualquer transformação, alcançou desempenho significativamente superior, com RMSE de 1,5902 (redução de 64,5% em relação ao melhor modelo sem VMD), MAE de 1,4014, sMAPE de 67,6006 e a menor variabilidade entre todos os modelos ($SD = 1,5090$). No entanto, a aplicação de transformações como $\log(y+1)$ e $\sqrt{y+0,5}$ resultou em aumento superior a 80% nos erros, indicando que os dados originais são os mais adequados para o VMD nesta configuração.

Por outro lado, a aplicação do VMD em modelos multivariados não trouxe benefícios. O melhor desempenho foi observado com o modelo VMD-RF com $\log(y+1)$, que ainda assim apresentou RMSE de 7,0373 e MAE de 5,4553. Esses valores foram significativamente superiores aos obtidos sem o VMD, como, por exemplo, o RMSE 58% maior do VMD-RF em comparação com o XGBoost sem VMD. Além disso, todas as configurações multivariadas com VMD apresentaram sMAPE acima de 91,0 e alta variabilidade (SD superior a 6,0), sugerindo instabilidade nas previsões e perda de precisão percentual.

Tabela 13 – Desempenho de modelos híbridos para previsão de dois meses à frente.

MODELO DE ENTRADA	TRANSFORMAÇÃO	CRITÉRIO	VMD-SVR	VMD-XGBoost	VMD-RF	VMD-ELM	VMD-LASSO
Univariado + VMD	Sem Transformação	RMSE	1,6831	2,8543	3,1011	2,0838	1,5902
		MAE	1,4665	2,5552	2,7697	1,8265	1,4014
		sMAPE	68,2985	87,5698	87,3675	73,3792	67,6006
		SD	1,5277	2,2980	2,2235	1,7548	1,5090
	Log(y+1)	RMSE	3,2228	3,3727	3,5970	3,1672	3,2099
		MAE	1,7098	2,0419	2,2976	2,1810	1,6404
		sMAPE	69,4531	72,5062	76,0673	77,4354	68,0296
		SD	3,3015	3,4419	3,6188	3,1360	3,2780
	Sqrt(y+0,5)	RMSE	2,9848	3,6375	3,6405	3,0075	2,9357
		MAE	1,8691	2,6235	2,7458	1,9362	1,7936
		sMAPE	74,3836	80,0891	82,5119	75,8467	73,2526
		SD	3,0564	3,5196	3,3985	3,0559	3,0140
Multivariado + VMD	Sem Transformação	RMSE	8,9126	9,2775	9,3744	9,0923	9,0078
		MAE	8,0429	8,1811	8,3692	8,2612	8,0730
		sMAPE	110,7986	114,0194	116,1609	116,6452	112,6573
		SD	5,7153	5,9828	5,9476	5,7596	5,8316
	Log(y+1)	RMSE	7,1810	7,3487	7,0373	8,3144	7,3420
		MAE	5,6093	5,8780	5,4553	6,5824	5,7537
		sMAPE	92,7655	96,1852	91,0064	100,7601	93,6806
		SD	6,2453	6,2681	6,1020	7,0992	6,3332
	Sqrt(y+0,5)	RMSE	7,6558	8,1017	7,8227	9,1826	7,9531
		MAE	6,4564	6,8612	6,5988	7,2901	6,6592
		sMAPE	101,7526	104,3023	102,6412	105,6752	99,6921
		SD	5,8973	6,1021	5,9171	7,3621	6,1007

A análise comparativa evidencia que o VMD é altamente eficaz quando aplicado a séries univariadas, com destaque para o modelo VMD-LASSO 22 sem transformação, que superou modelos tradicionais como ARIMA e XGBoost, alcançando uma redução superior a 64% nos erros e a menor variabilidade registrada. No entanto, sua aplicação em contextos multivariados mostrou-se ineficaz, resultando em aumento médio de 58% no RMSE e 83% no MAE. A combinação entre o VMD e variáveis exógenas parece introduzir ruídos que não são mitiga-

dos pelo processo de decomposição. Com relação às transformações, conclui-se que os dados originais são mais adequados para a abordagem univariada com VMD, enquanto no cenário multivariado sem VMD, o uso de dados originais para o XGBoost e a transformação $\log(y+1)$ para o RF proporcionam os melhores desempenhos.

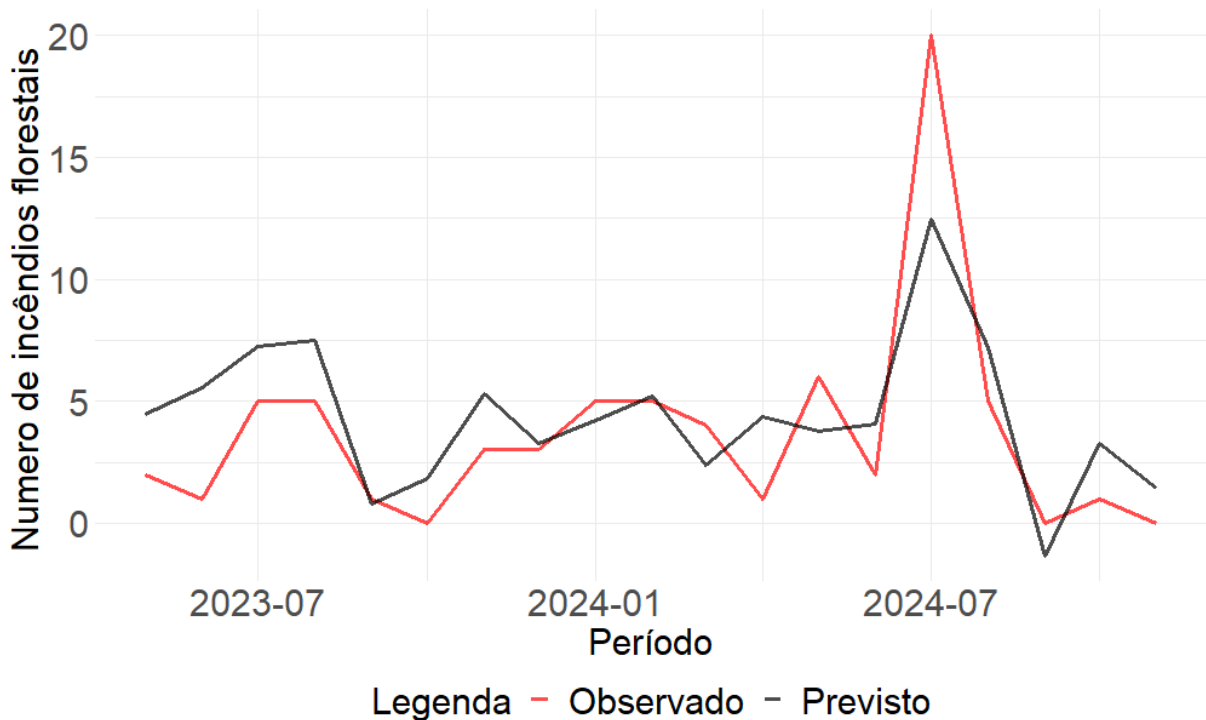


Figura 22 – Gráfico do modelo VMD-LASSO com dados normais para H=2.

3.6.2.3 Previsão de Incêndios Florestais para Três Meses à Frente (H=3)

As Tabelas 14 e 15 resumem o desempenho dos modelos de previsão de focos de incêndio em Xanxerê para um horizonte de três meses, comparando abordagens univariadas e multivariadas, com e sem VMD.

Sem VMD, na análise univariada o XGBoost com transformação de raiz quadrada $\sqrt{y+0,5}$ obteve o menor RMSE (5,2272) e MAE (3,8293), equilíbrio sólido entre precisão (sMAPE = 90,16) e estabilidade (SD = 4,9513). Em segundo lugar, o ARIMA com $\log(y+1)$ apresentou RMSE de 5,5650, mas MAE elevado (5,3144) e sMAPE alto (157,49), indicando maior sensibilidade a picos atípicos. Na configuração multivariada sem VMD, o ELM com $\sqrt{y+0,5}$ foi o mais eficaz, com RMSE de 4,6695, MAE de 3,0878 e sMAPE de 79,62, superando XGBoost $\log(y+1)$, RMSE = 5,7386; MAE = 4,7596) e ARIMAX (RMSE = 5,1365; MAE = 4,2359).

Tabela 14 – Desempenho de modelo único para previsão de três meses à frente.

Entrada	Transformação	Critério	SVR	XGBoost	RF	ELM	LASSO	ARIMA/ARIMAX
Univariado	Sem Transf.	RMSE	8,1148	5,7501	7,4034	6,7606	8,9616	8,8340
		MAE	7,5814	4,5574	6,9013	5,3463	8,3544	8,1063
		sMAPE	121,5875	100,8754	117,3786	104,0217	121,8214	172,3540
		SD	5,3463	5,1995	5,0419	6,0468	4,9363	3,8464
	Log(y+1)	RMSE	7,4020	7,6413	5,8978	9,2068	7,1125	5,5650
		MAE	6,7477	6,3888	4,9040	7,5866	6,4773	5,3144
		sMAPE	96,8081	88,3610	81,9631	96,4157	93,4305	157,4861
		SD	5,0412	6,4413	5,0231	6,2873	4,9472	1,8087
	Sqrt(y+0,5)	RMSE	8,1572	5,2272	6,3450	9,4173	7,6859	7,1318
		MAE	7,5661	3,8293	5,4820	7,8492	6,9927	6,6587
		sMAPE	121,0987	90,1606	106,8138	127,0741	114,5418	167,5820
		SD	5,4312	4,9513	5,1561	6,8531	5,0396	2,7979
Multivariado	Sem Transf.	RMSE	5,9743	7,5819	10,4428	17,9596	8,0440	11,6134
		MAE	5,4616	6,8560	9,6847	17,2677	7,2221	8,1294
		sMAPE	95,9729	102,2028	116,5094	140,3300	105,8269	79,0099
		SD	4,6888	5,0894	4,6820	5,0844	5,8623	4,4905
	Log(y+1)	RMSE	7,8480	5,7386	6,0080	5,5826	6,5247	5,1365
		MAE	6,9144	4,7596	5,4186	3,6190	5,8295	4,2359
		sMAPE	103,5342	83,3961	94,1057	122,3529	97,9169	106,4646
		SD	6,0581	5,1852	4,5458	4,8474	5,2732	5,0864
	Sqrt(y+0,5)	RMSE	6,5642	6,5996	6,9097	4,6695	7,4342	5,5421
		MAE	5,8670	6,0876	6,4241	3,0878	6,6812	4,5399
		sMAPE	98,1951	98,8483	101,9129	79,6184	104,1303	110,5359
		SD	5,2251	4,5855	4,6370	4,5179	5,7113	4,2397

Ao aplicar o VMD, a versão univariada do LASSO sem transformação dominou, reduzindo o RMSE a 1,6818 (-67,8% em relação ao melhor sem VMD), com MAE = 1,4968, sMAPE = 68,54 e SD = 1,5927. Transformações logarítmica e de raiz quadrada, nesse caso, ampliaram os erros em mais de 100%, o que reforça a adequação dos dados originais para VMD. Já na abordagem multivariada com VMD, mesmo o melhor modelo (VMD-RF com $\log(y+1)$, RMSE = 7,2352; MAE = 5,7163) ficou abaixo dos equivalentes sem VMD, apresentando RMSE 55% maior que o ELM puro. Além disso, todas as configurações multivariadas com VMD exibiram sMAPE acima de 94,7, e o VMD-ELM chegou a RMSE = 18,1869, evidenciando incompatibilidade entre VMD e variáveis exógenas.

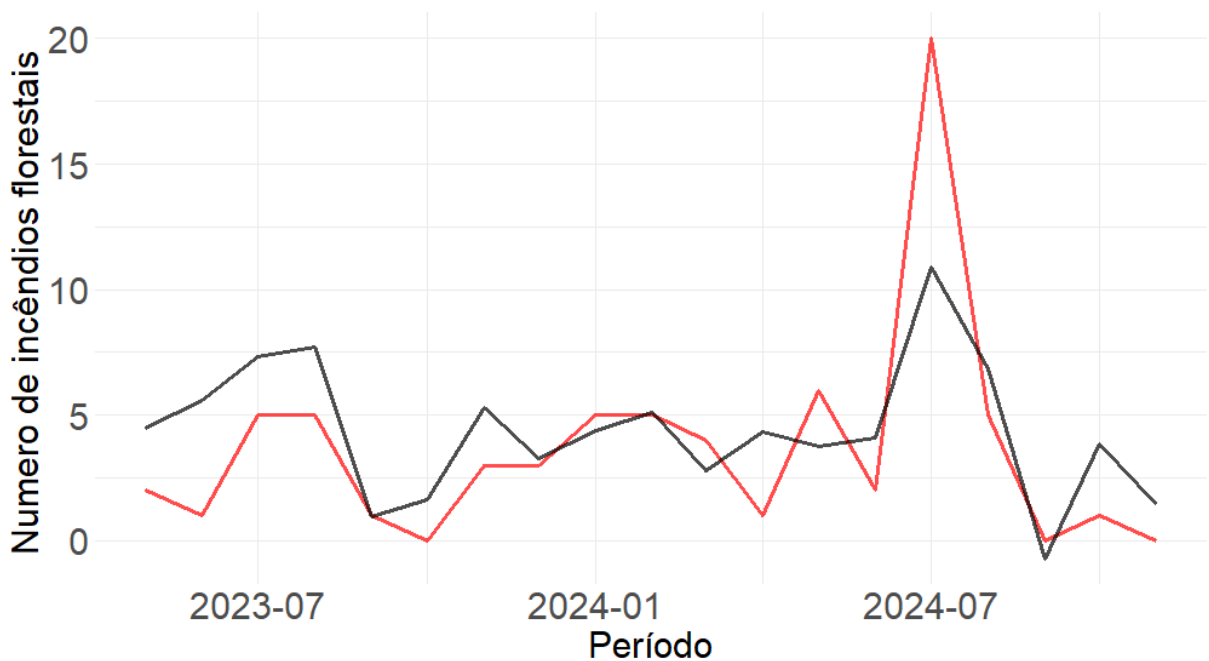
Sendo assim, para séries univariadas de três meses, o VMD-LASSO 23 sem transformação é o mais recomendado, graças à drástica redução de erro e alta estabilidade. Para modelos multivariados, o ELM com $\sqrt{y + 0,5}$ sem VMD mantém a melhor performance. O uso de VMD em cenários multivariados deve ser evitado, pois tende a degradar os resultados e pode provocar falhas extremas.

3.6.3 Análise da Contribuição das Variáveis de Entrada para os Modelos de Previsão

A figura 25 dos valores SHAP permite avaliar a contribuição das variáveis no modelo SVR para previsão de focos de incêndio em Xanxerê-SC. As variáveis analisadas incluem defa-

Tabela 15 – Desempenho de modelos híbridos para previsão de três meses à frente.

MODELO DE ENTRADA	TRANSFORMAÇÃO	CRITÉRIO	VMD-SVR	VMD-XGBoost	VMD-RF	VMD-ELM	VMD-LASSO
Univariado + VMD	Sem Transformação	RMSE	1,7931	2,8797	3,1212	2,5136	1,6818
		MAE	1,5607	2,5837	2,7846	2,1372	1,4968
		sMAPE	68,6450	87,9600	87,3109	75,3496	68,5364
		SD	1,6061	2,3092	2,2445	1,9704	1,5927
	Log(y+1)	RMSE	3,5204	3,7365	3,7635	3,1335	3,5453
		MAE	1,7434	2,1501	2,3550	2,2358	1,7070
		sMAPE	69,3673	72,6793	76,5879	77,8163	68,4598
		SD	3,6070	3,8241	3,7988	3,0433	3,6192
	Sqrt(y+0,5)	RMSE	3,3466	3,7222	3,7354	3,1029	3,3402
		MAE	1,9672	2,7041	2,8105	2,0949	1,9046
		sMAPE	74,7457	81,3461	83,2242	77,7912	73,7324
		SD	3,4283	3,6027	3,5012	3,0938	3,4308
Multivariado + VMD	Sem Transformação	RMSE	8,6538	8,4613	8,7610	8,5189	9,0950
		MAE	7,6250	7,3433	7,7151	7,2671	7,9030
		sMAPE	116,3036	107,7745	111,3659	112,0327	118,7383
		SD	6,2694	5,7517	5,8828	7,1021	6,8683
	Log(y+1)	RMSE	9,5606	7,6484	7,2352	18,1869	10,1099
		MAE	7,0511	5,9928	5,7163	10,2128	7,3053
		sMAPE	104,2861	95,9493	94,7393	114,4691	103,6595
		SD	8,4616	6,5594	6,2078	17,0103	8,8656
	Sqrt(y+0,5)	RMSE	8,2265	8,0231	7,8411	11,0189	8,8139
		MAE	6,5433	6,7061	6,5202	7,6830	7,1044
		sMAPE	102,8916	102,6595	101,4031	108,2454	103,9815
		SD	6,9460	6,0793	6,0657	9,8489	7,0699



Legenda - Observado - Previsto

Figura 23 – Gráfico do modelo VMD-LASSO com dados normais para H=3.

sagens meteorológicas — temperatura ($LAG_{1,1}$), umidade ($LAG_{1,2}$), vento ($LAG_{1,3}$) e precipitação ($LAG_{1,4}$) —, histórico de focos (LAG_1 a LAG_5) e o fenômeno La Niña (binário). Valores SHAP positivos indicam aumento no risco, enquanto negativos o reduzem.

A temperatura no mês anterior ($LAG_{1,1}$) apresentou o maior impacto positivo (SHAP $\approx +0,4$), indicando que temperaturas mais elevadas contribuem significativamente para o aumento do risco de incêndios, em razão do maior ressecamento da vegetação. De forma se-

melhante, o número de focos registrados três meses antes (LAG_3) também exerceu influência relevante, o que pode estar relacionado à presença de padrões sazonais e à reincidência de queimadas em ciclos trimestrais, possivelmente vinculados ao tempo necessário para a regeneração da vegetação.

Por outro lado, a precipitação ($LAG_{1.4}$) apresentou SHAP nulo, indicando ausência de influência direta na previsão. Isso sugere que seu efeito pode ser indireto ou não linear. Observa-se também que variáveis de curto prazo, como a temperatura do mês anterior, e de médio prazo, como os focos no terceiro mês anterior, são mais relevantes, enquanto focos mais recentes (LAG_1) não figuram entre os principais preditores, indicando que o risco depende de tendências acumuladas e não apenas de eventos imediatos. O fenômeno La Niña não apresentou impacto relevante no modelo, possivelmente por seu efeito ser mediado por variáveis como temperatura ou por apresentar influência regional reduzida.

Diante desses resultados, destaco que a temperatura ($LAG_{1.1}$) e os focos históricos do terceiro mês anterior (LAG_3) são variáveis-chave para monitoramento e devem compor sistemas de alerta precoce. Variáveis com SHAP próximo de zero, como a precipitação, podem ser reavaliadas ou removidas para simplificar o modelo. Operacionalmente, períodos de calor exigem ações preventivas imediatas, e a ocorrência de incêndios demanda vigilância intensificada no terceiro mês subsequente.

Os resultados indicam que o risco de incêndio na região é influenciado principalmente pela inércia climática e por padrões de recorrência temporal, reforçando a importância de modelos que considerem defasagens estratégicas. Além disso, embora vento e umidade não tenham apresentado impacto relevante neste cenário, sua influência pode se manifestar em outras situações ou configurações.

A figura 24 dos valores SHAP permite avaliar a contribuição das variáveis no modelo SVR para previsão de focos de incêndio em Lages-SC. Diferentemente do que foi observado em Xanxerê, o comportamento das variáveis em Lages revela padrões opostos, com predominância de efeitos redutores sobre o risco de incêndios.

A análise indica que a umidade no mês anterior ($LAG_{1.2}$) é o fator de maior impacto negativo no risco, com um valor SHAP de aproximadamente -1.0. Esse resultado sugere que aumentos na umidade reduzem de forma significativa a probabilidade de ocorrência de incêndios, o que é consistente com o entendimento físico do fenômeno, já que altos níveis de umidade limitam a combustão da vegetação. Cada unidade adicional na umidade resulta na diminuição

de uma unidade na escala de risco do modelo.

Outro fator relevante foi o histórico de focos no mês anterior (LAG_1), que apresentou um valor SHAP em torno de -0.5, indicando que a ocorrência recente de incêndios está associada à redução do risco atual. Esse comportamento pode ser explicado pela depleção de combustível após os incêndios, tornando a vegetação local temporariamente menos suscetível, além da possibilidade de intensificação das ações de controle e fiscalização logo após eventos recentes.

A temperatura no mês anterior ($LAG_{1.1}$) também apresentou impacto negativo, com valor SHAP de aproximadamente -0.5, o que representa um padrão contraintuitivo, uma vez que, na maioria dos contextos, temperaturas elevadas tendem a aumentar o risco de incêndio. Esse comportamento anômalo pode estar relacionado às características climáticas específicas de Lages, onde picos de temperatura frequentemente estão associados a chuvas convectivas, típicas de calor intenso, ou ainda à interação entre calor e alta umidade, onde o calor úmido não exerce o mesmo efeito de secagem que o calor seco.

De forma geral, os padrões observados mostram que a umidade histórica atua como o principal inibidor de incêndios na região, sendo o fator com maior peso na redução do risco. A ocorrência de incêndios no mês anterior também contribui para a redução do risco, sugerindo que a vegetação não se regenera como combustível no intervalo de até um mês, deslocando o pico de risco para períodos posteriores. A relação inversa entre temperatura e risco reforça a hipótese de que características climáticas locais, como a associação entre calor e precipitação, desempenham papel determinante na dinâmica dos incêndios em Lages.

O contraste com os resultados obtidos em Xanxerê é evidente, uma vez que, enquanto lá a temperatura elevada é fator de aumento de risco, em Lages tanto a temperatura quanto a umidade atuam como fatores de supressão. Além disso, assim como no outro modelo, a variável La Niña novamente não apresentou impacto significativo, com valor SHAP próximo de zero, indicando ausência de influência direta nas previsões para este contexto.

3.6.4 Discussão sobre o Desempenho dos Modelos Univariados, Multivariados e Híbridos

A análise comparativa dos modelos de previsão de focos de incêndio em Lages e Xanxerê revela padrões críticos que transcendem as particularidades regionais. Para séries temporais puras, os modelos híbridos que combinam decomposição VMD com técnicas de aprendizado de máquina (especialmente VMD-LASSO) demonstraram superioridade em ambos os municípios.

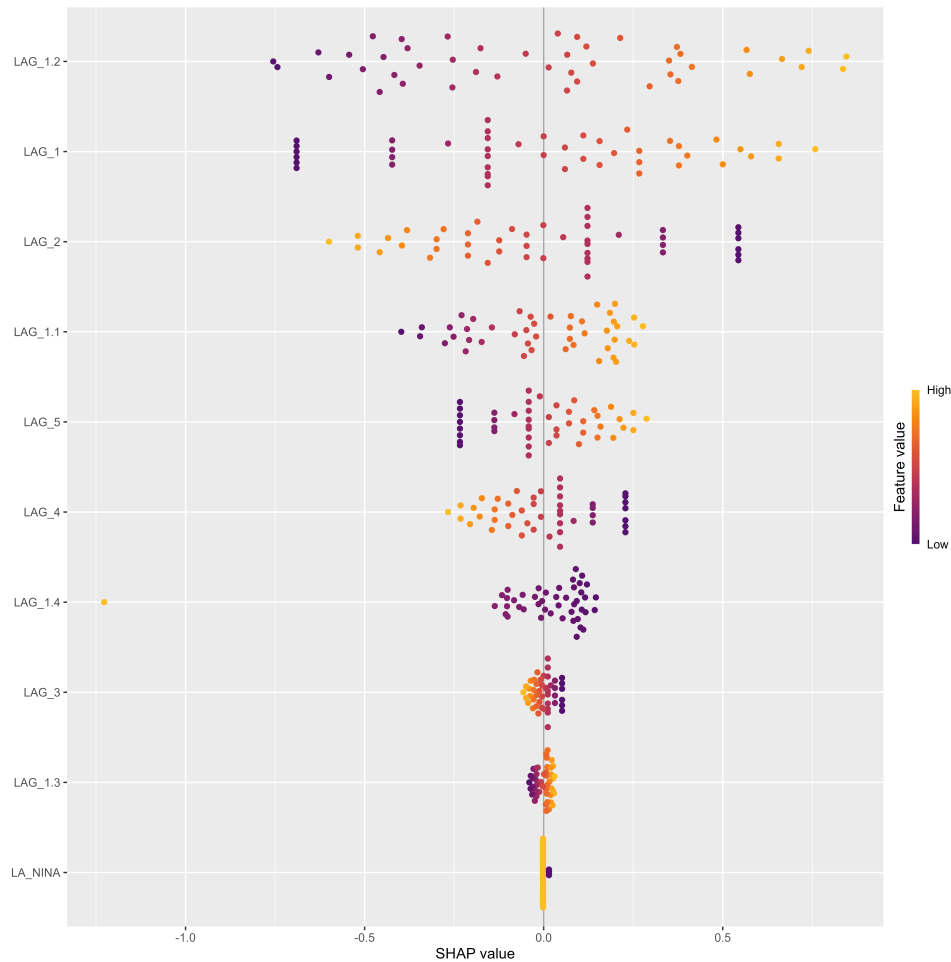


Figura 24 – Valores Shapley variáveis do modelo multivariado para cidade de Lages.

Em Lages, essa abordagem reduziu o RMSE em mais de 40% nos três horizontes de previsão (H=1,2,3), enquanto em Xanxerê as reduções foram ainda mais expressivas, alcançando 67.8% no horizonte de três meses. Além da precisão, o VMD-LASSO garantiu estabilidade operacional, com SD consistentemente abaixo de 1.6 – até três vezes menor que modelos convencionais como ARIMA ou XGBoost sem pré-processamento. Contudo, observou-se um contraste nas transformações de dados: enquanto em Lages as transformações logarítmica ou de raiz quadrada otimizaram os resultados, em Xanxerê o uso de dados brutos foi essencial, pois transformações aumentaram os erros em mais de 80%, indicando que as características intrínsecas das séries locais demandam estratégias distintas.

Em contrapartida, a incorporação de variáveis climáticas expôs limitações estruturais dos modelos híbridos. Em ambas as cidades, a aplicação do VMD em abordagens multivariadas degradou sistematicamente o desempenho: o RMSE aumentou mais de 32% em Lages e 40% em Xanxerê, acompanhado de instabilidade crítica ($SD > 6.0$ em Xanxerê-H=1). Casos extremos, como o VMD-ELM em Xanxerê (H=3) com RMSE 289% superior ao modelo sem

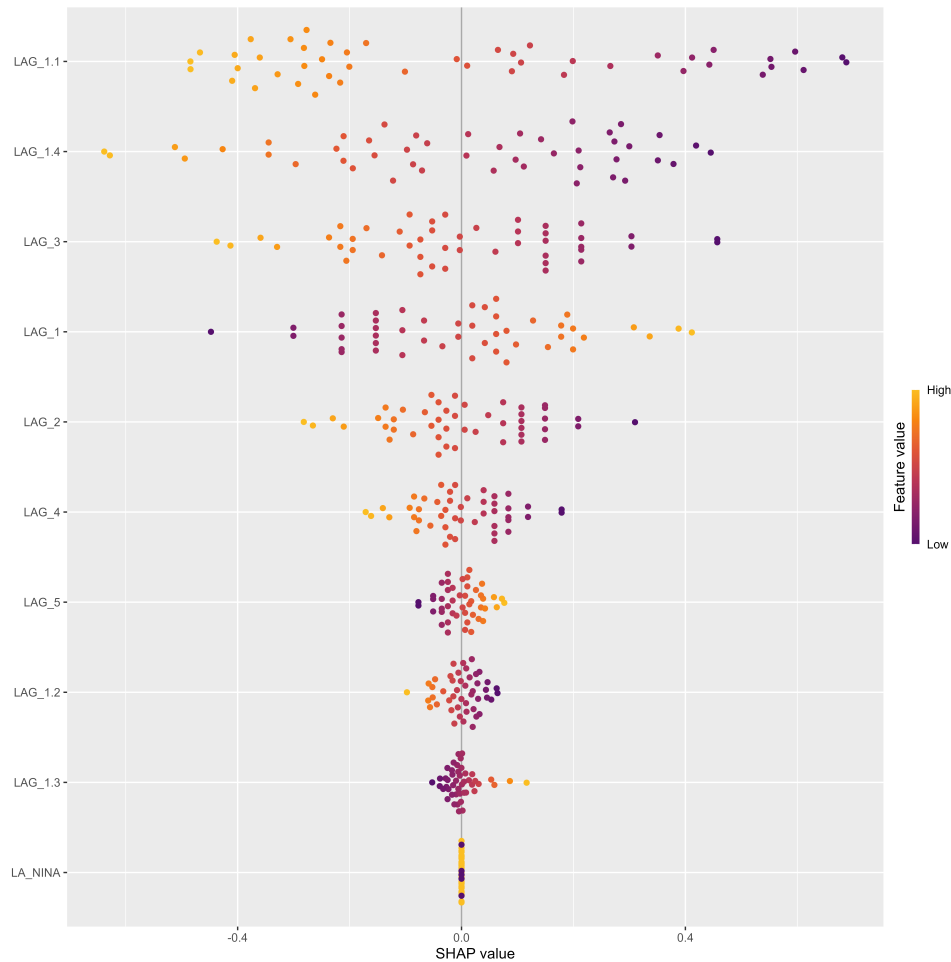


Figura 25 – Valores Shapley variáveis do modelo multivariado para cidade de Xanxerê.

decomposição, evidenciam que o VMD dessincroniza os sinais climáticos da série temporal, introduzindo ruído não mitigável. Essa falha torna modelos diretos – como ELM, XGBoost e RF sem VMD – alternativas indiscutivelmente superiores para cenários multivariados. Em Lages, o ELM com transformação logarítmica destacou-se, enquanto em Xanxerê o XGBoost sem transformação ($H=2$) e o ELM com raiz quadrada ($H=3$) foram mais eficazes, reforçando que interações clima-fogo exigem não apenas modelos não-lineares, mas também tratamento diferenciado de dados conforme o contexto regional. Tal comportamento observado na presente análise encontra respaldo na literatura, que evidencia que a eficácia do VMD está diretamente condicionada à natureza dos dados e ao alinhamento entre os sinais decompostos e as variáveis exógenas.

Nesse contexto, estudos recentes reforçam a efetividade do VMD na modelagem de séries temporais. No artigo de Lin *et al.* (2023), por exemplo, observou-se que o uso do VMD permitiu decompor a sequência de velocidade do vento — sujeita a diversas perturbações — em múltiplas subsequências que incluem tanto termos de perturbação quanto de tendência. As

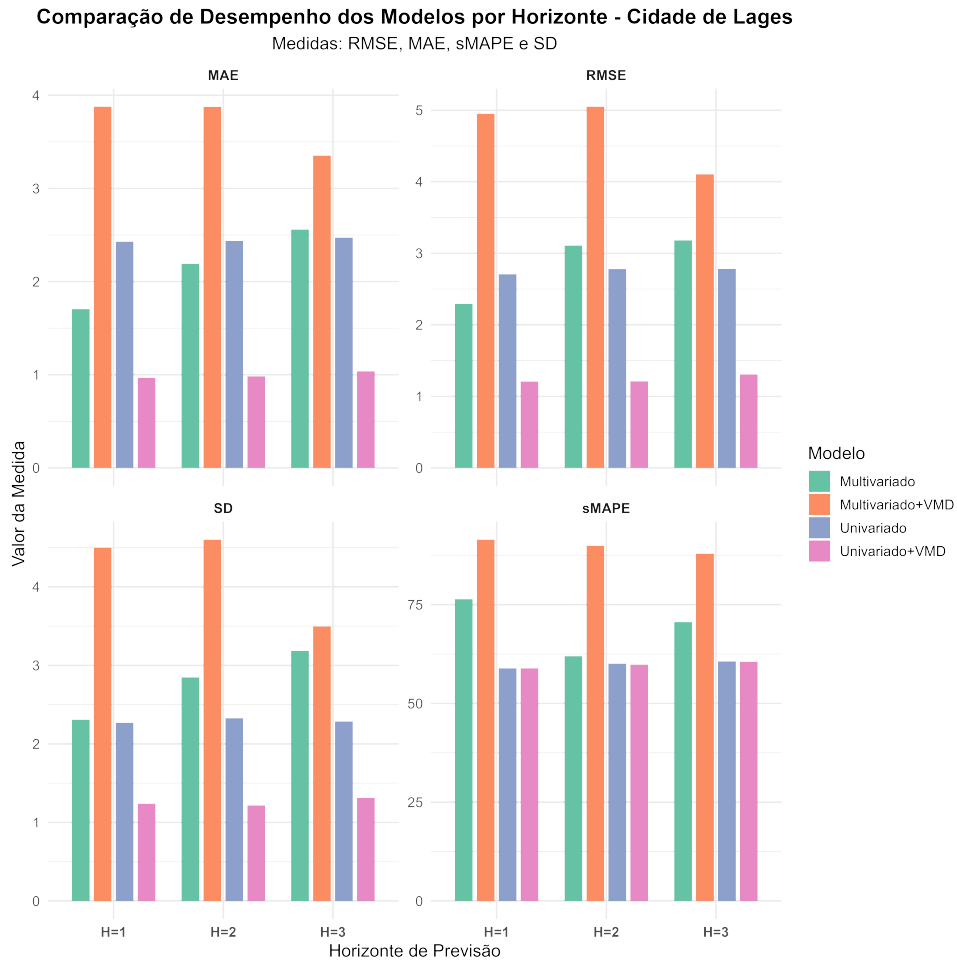


Figura 26 – Desempenho comparativo dos melhores modelos para a cidade de Lages, considerando diferentes horizontes de previsão e abordagens (Univariada, Multivariada e híbrida com VMD).

diferenças de amplitude e frequência entre essas subsequências são atribuídas às escalas espaço-temporais dos diferentes processos físico-químicos envolvidos. Esses achados evidenciam que o VMD apresenta desempenho superior quando combinado a outros algoritmos. Além disso, foi constatado que, embora o EMD não exija processo de seleção de parâmetros, sua decomposição oferece desempenho inferior ao VMD, especialmente quando este é ajustado com parâmetros otimizados.

Corroborando essa perspectiva, o estudo de Chen *et al.* (2023) explora a combinação do VMD com redes neurais LSTM, destacando-se como uma técnica eficaz para aprimorar a precisão de modelos preditivos em séries temporais — sobretudo na presença de dados ruidosos e componentes residuais complexos. Essa abordagem oferece uma solução robusta para lidar com flutuações e ruídos nos dados, sendo particularmente útil em domínios como monitoramento de sinais GNSS, análise ambiental e outras aplicações que envolvem séries temporais não lineares e de difícil modelagem. Quando comparado ao LSTM puro, o modelo DVMD-

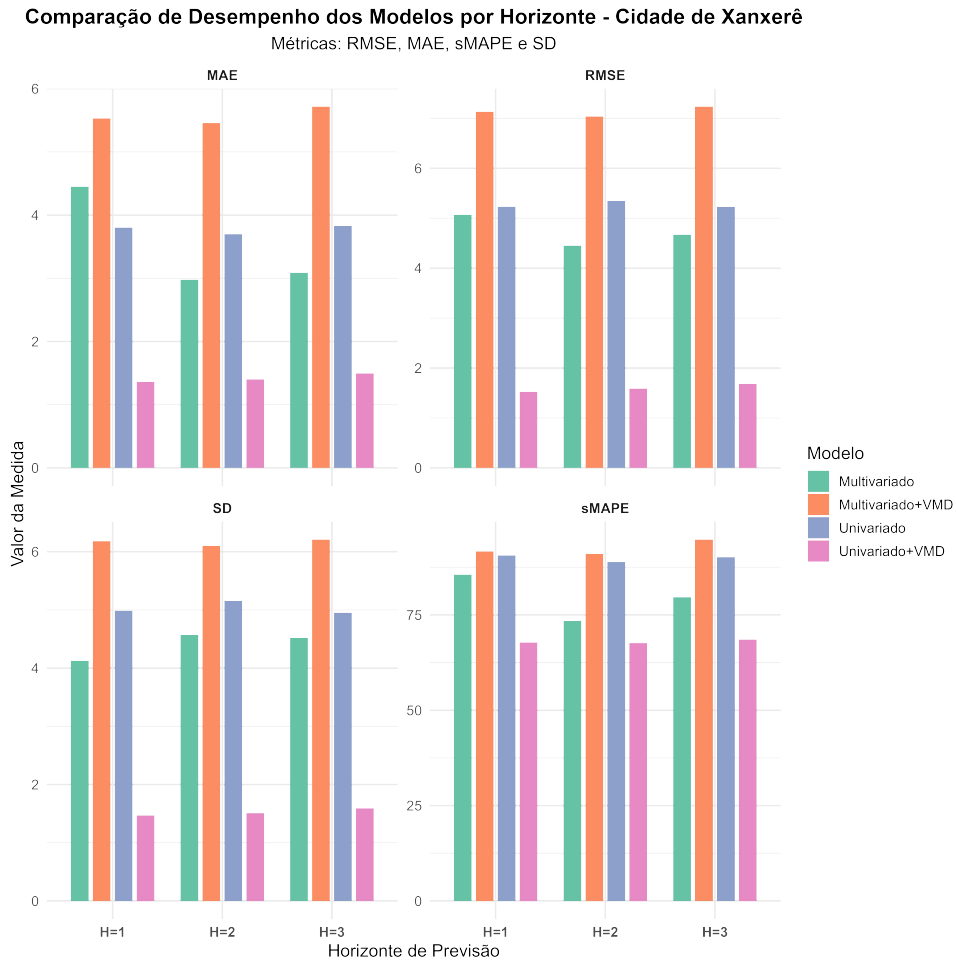


Figura 27 – Desempenho comparativo dos melhores modelos para a cidade de Xanxerê, considerando diferentes horizontes de previsão e abordagens (Univariada, Multivariada e híbrida com VMD).

LSTM apresentou um aumento de 36,50% na precisão da previsão do modelo de ruído ótimo e uma melhoria de 33,02% na previsão da velocidade. Esses resultados atestam a superioridade do DVMD-LSTM, que não apenas aprimora a capacidade preditiva, mas também incorpora de forma mais eficiente as características de ruído presentes nos dados. Assim, a integração entre VMD e LSTM se consolida como uma estratégia poderosa para obter previsões mais precisas em cenários desafiadores.

Dessa forma, torna-se evidente que o VMD, quando integrado a modelos preditivos, oferece avanços significativos em relação aos métodos tradicionais aplicados isoladamente. A sua capacidade de decompor séries temporais em modos específicos possibilita uma modelagem mais precisa, robusta e adaptável a diferentes cenários e tipos de dados. Esses aprimoramentos consolidam o VMD como uma ferramenta essencial e cada vez mais relevante para aplicações de previsão de séries temporais, sobretudo quando combinado a técnicas modernas de aprendizado de máquina.

3.7 CONCLUSÃO

A análise comparativa realizada neste estudo demonstra que os modelos híbridos baseados em decomposição VMD, particularmente o VMD-LASSO, alcançaram desempenho excepcional na previsão univariada de focos de incêndio em Lages e Xanxerê, reduzindo o RMSE em até 67,8% e mantendo estabilidade operacional notável com SD consistentemente abaixo de 1.6. Contudo, a incorporação de variáveis climáticas - temperatura, precipitação, umidade, vento e La Niña - revelou desafios significativos. A principal dificuldade reside na dessincronização dos sinais climáticos durante o processo de decomposição VMD, que sistematicamente degradou o desempenho dos modelos híbridos, aumentando o erro em mais de 40% e gerando instabilidade crítica em ambos os municípios. Essa limitação foi agravada pela sensibilidade regional às transformações de dados, onde estratégias benéficas em Lages tornaram-se prejudiciais em Xanxerê, aumentando os erros em mais de 80%.

Os resultados obtidos oferecem ferramentas valiosas para o CBMSC na gestão de efetivos. Para planejamento estratégico com até três meses de antecedência, o modelo VMD-LASSO aplicado a séries históricas puras permite alocação preventiva de bombeiros em períodos críticos. Embora os resultados dos modelos multivariados não tenham atingido o desempenho esperado em termos de erro e estabilidade, abordagens diretas — como ELM e XGBoost sem decomposição —, quando ajustadas às especificidades regionais, demonstraram elevado potencial para geração de alertas precoces baseados nas sinergias entre variáveis climáticas. Combinações críticas, como baixos índices de umidade, ventos intensos e padrões associados ao fenômeno La Niña, são eficazmente capturadas por esses modelos. Essa capacidade preditiva, ainda que com limitações, permite aprimorar o planejamento operacional, otimizando o dimensionamento de efetivos sazonais e fortalecendo as estratégias de prevenção e resposta a incêndios florestais.

Para avanços futuros, torna-se premente explorar técnicas de decomposição alternativas como EEMD ou Wavelets que preservem a sincronia entre variáveis exógenas, desenvolver arquiteturas híbridas com redes neurais recursivas capazes de alinhar componentes decompostos a dados climáticos, e implementar protocolos automáticos de seleção de transformações baseados em características locais. A harmonização entre decomposição e variáveis multivariadas permanece o principal desafio aberto nesta frente de pesquisa, cuja superação poderá unificar as vantagens atualmente segmentadas entre abordagens univariadas e multivariadas, ampliando ainda mais a precisão e aplicabilidade operacional desses modelos na proteção ambiental.

4 DISCUSSÃO FINAL E IMPLICAÇÕES PRÁTICAS

Esta pesquisa buscou responder à pergunta central sobre quais variáveis, modelos e estratégias são mais eficazes para a previsão de incêndios florestais e de que forma podem subsidiar a tomada de decisão no planejamento operacional. O trabalho contribuiu para preencher lacunas importantes identificadas na literatura, incluindo: a utilização de séries temporais de focos de incêndio provenientes do Corpo de Bombeiros, permitindo previsões regionais mais precisas; a incorporação de variáveis climáticas e do fenômeno La Niña em modelos estatísticos, de aprendizado de máquina e híbridos, aumentando a capacidade de modelagem frente às condições ambientais variáveis; e a avaliação de horizontes de previsão mais curtos, promovendo respostas mais ágeis para a gestão de incêndios florestais. Para que os estudos preditivos sejam eficazes, é fundamental que as bases de dados estejam sempre atualizadas, garantindo que os modelos reflitam as condições mais recentes e reais.

Os objetivos específicos da pesquisa foram alcançados ao identificar variáveis e abordagens preditivas relevantes, implementar modelos de séries temporais mensais e integrar técnicas de decomposição como a VMD aos algoritmos de aprendizado de máquina. A análise comparativa entre modelos univariados e multivariados evidenciou que, embora a incorporação de variáveis climáticas e regionais seja importante para capturar padrões complexos, a série histórica pura univariada continua sendo a abordagem mais robusta e confiável para previsão prática. A aplicação dos *Shapley Values* permitiu compreender a importância relativa das variáveis preditoras, fornecendo subsídios claros para a priorização de fatores críticos na gestão de incêndios.

Do ponto de vista prático, os resultados fornecem ferramentas valiosas para o Corpo de Bombeiros Militar de Santa Catarina. A previsão antecipada possibilita planejar a alocação do efetivo, incluindo os bombeiros comunitários, de forma mais eficiente e direcionada, permitindo uma gestão adequada e estratégica, em que a escala de serviço diária é organizada com base em períodos de maior risco, ao invés de depender de critérios aleatórios. A ideia de gerenciar o efetivo considerando os possíveis dias com maior risco de incêndios florestais já foi apresentada ao comando imediato, demonstrando a aplicabilidade dos resultados. Mesmo com limitações em cenários multivariados, modelos de aprendizado de máquina sem decomposição mostraram potencial para alertas precoces, identificando combinações críticas de condições climáticas que aumentam a probabilidade de incêndio. Dessa forma, os resultados demonstram que é possível

transformar dados e análises preditivas em ações concretas de gestão, fortalecendo o planejamento operacional e a proteção ambiental.

Em síntese, a pesquisa atingiu seu objetivo geral ao propor estratégias preditivas que combinam séries temporais históricas, variáveis climáticas e técnicas híbridas de modelagem. Embora haja espaço para avanços na harmonização de decomposição e dados multivariados, os achados reforçam que modelos univariados ajustados regionalmente oferecem a base mais sólida para apoiar decisões operacionais e aprimorar a capacidade de prevenção e combate a incêndios florestais.

REFERÊNCIAS

- ABDOLLAHI, Arnick; PRADHAN, Biswajeet. Explainable artificial intelligence (xai) for interpreting the contributing factors feed into the wildfire susceptibility prediction model. **Science of the Total Environment**, Elsevier, v. 879, p. 163004, 2023.
- ABID, Faroudja. A survey of machine learning algorithms based forest fires prediction and detection systems. **Fire technology**, Springer, v. 57, n. 2, p. 559–590, 2021.
- ALMEIDA, J. P. **Ecosistemas e Biodiversidade: Conservação de Zonas Ripárias**. São Paulo: Editora Sustentabilidade, 2022. ISBN 978-85-1234-567-8.
- ALTMAN, Douglas G; BLAND, J Martin. Standard deviations and standard errors. **Bmj**, British Medical Journal Publishing Group, v. 331, n. 7521, p. 903, 2005.
- ARAGÃO, Luiz EOC; ANDERSON, Liana O; FONSECA, Marisa G; ROSAN, Thais M; VEDOVATO, Laura B; WAGNER, Fabien H; SILVA, Camila VJ; JUNIOR, Celso HL Silva; ARAI, Egidio; AGUIAR, Ana P *et al.* 21st century drought-related fires counteract the decline of amazon deforestation carbon emissions. **Nature communications**, Nature Publishing Group UK London, v. 9, n. 1, p. 536, 2018.
- ARIF, M; ALGHAMDI, KK; SAHEL, SA; ALOSAIMI, SO; ALSAHAFT, ME; ALHARTHI, MA; ARIF, M. Role of machine learning algorithms in forest fire management: A literature review. **J. Robot. Autom**, v. 5, p. 212–226, 2021.
- AWAD, Mariette; KHANNA, Rahul. **Efficient learning machines: theories, concepts, and applications for engineers and system designers**. [S.l.]: Springer nature, 2015.
- BERGMEIR, Christoph; HYNDMAN, Rob J; KOO, Bonsoo. A note on the validity of cross-validation for evaluating autoregressive time series prediction. **Computational Statistics & Data Analysis**, Elsevier, v. 120, p. 70–83, 2018.
- BERGONSE, Raffaello; OLIVEIRA, Sandra; GONÇALVES, Ana; NUNES, Sílvia; DACAMARA, Carlos; ZÊZERE, José Luis. Predicting burnt areas during the summer season in portugal by combining wildfire susceptibility and spring meteorological conditions. **Geomatics, Natural Hazards and Risk**, Taylor & Francis, v. 12, n. 1, p. 1039–1057, 2021.
- BERLINCK, Christian N; BATISTA, Eugênia KL. Good fire, bad fire: It depends on who burns. **Flora**, Elsevier, v. 268, p. 151610, 2020.

BIERENS, Herman J. Armax model specification testing, with an application to unemployment in the netherlands. **Journal of Econometrics**, Elsevier, v. 35, n. 1, p. 161–190, 1987.

BOX, George EP; JENKINS, Gwilym M; REINSEL, Gregory C; LJUNG, Greta M. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.

BRANCO, Nathielle Waldrigues; CAVALCA, Mariana Santos Matos; STEFENON, Stefano Frizzo; LEITHARDT, Valderi Reis Quietinho. Wavelet LSTM for fault forecasting in electrical power grids. **Sensors**, v. 22, n. 21, p. 8323, 2022.

BREIMAN, Leo. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.

CERQUEIRA, Vitor; TORGO, Luis; MOZETIČ, Igor. Evaluating time series forecasting models: An empirical study on performance estimation methods. **Machine Learning**, Springer, v. 109, n. 11, p. 1997–2028, 2020.

CHAVES, Michel Eustáquio Dantas; MATAVELI, Guilherme Augusto Verola; JUSTINO, Rodrigo Cesário. Uso da modelagem estatística para monitoramento da vegetação no parque nacional da serra da canastra, minas gerais. **Caderno de Geografia**, Pontifícia Universidade Católica de Minas Gerais, v. 24, n. 1, p. 120–132, 2014.

CHEN, Hongkang; LU, Tieding; HUANG, Jiahui; HE, Xiaoxing; YU, Kegen; SUN, Xiwen; MA, Xiaping; HUANG, Zhengkai. An improved vmd-lstm model for time-varying gnss time series prediction with temporally correlated noise. **Remote Sensing**, MDPI, v. 15, n. 14, p. 3694, 2023.

CHEN, Tianqi; GUESTRIN, Carlos. Xgboost: A scalable tree boosting system. *In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794.

CILLI, Roberto; ELIA, Mario; D'ESTE, Marina; GIANNICO, Vincenzo; AMOROSO, Nicola; LOMBARDI, Angela; PANTALEO, Ester; MONACO, Alfonso; SANESI, Giovanni; TANGARO, Sabina *et al.* Explainable artificial intelligence (xai) detects wildfire occurrence in the mediterranean countries of southern europe. **Scientific reports**, Nature Publishing Group UK London, v. 12, n. 1, p. 16349, 2022.

CORSO, Marcelo Piccolotto; STEFENON, Stefano Frizzo; SINGH, Gurmail; MATSUO, Marcos Vinicius; PEREZ, Fábio Luis; LEITHARDT, Valderi Reis Quietinho. Evaluation of visible contamination on power grid insulators using convolutional neural networks. **Electrical Engineering**, v. 105, p. 3881–3894, 2023.

COSTA, F. G.; SANTOS, M. T. Impactos econômicos da degradação do solo no agronegócio. *In: RIBEIRO, A. B. (Ed.). Desafios do Século XXI: Agricultura e Sustentabilidade. [S.l.]: EDUFF, 2023. p. 127–150.*

CUI, Zhesen; LI, Tian; DING, Zhe; WU, Jinran *et al.* Probabilistic oil price forecasting with a variational mode decomposition-gated recurrent unit model incorporating pinball loss. **Data Science and Management**, Elsevier, 2024.

da Silva, Evandro Cardozo; FINARDI, Erlon Cristian; STEFENON, Stefano Frizzo. Enhancing hydroelectric inflow prediction in the Brazilian power system: A comparative analysis of machine learning models and hyperparameter optimization for decision support. **Electric Power Systems Research**, v. 230, p. 110275, 2024.

DIAS, Genebaldo Freire. Queimadas e incêndios florestais: cenários e desafios: subsídios para a educação ambiental. **Brasília: IBAMA**, v. 2, 2009.

DRUCKER, Harris; BURGESS, Christopher J; KAUFMAN, Linda; SMOLA, Alex; VAPNIK, Vladimir. Support vector regression machines. **Advances in neural information processing systems**, v. 9, 1996.

ENSSLIN, Sandra Rolim; ENSSLIN, Leonardo; BACK, Felipe; LACERDA, Rogério Tadeu de Oliveira. Improved decision aiding in human resource management: a case using constructivist multi-criteria decision aiding. **International Journal of Productivity and Performance Management**, Emerald Group Publishing Limited, v. 62, n. 7, p. 735–757, 2013.

FANGRONG, Zhou; HAO, Pan; ZHENYU, Gao; XUYONG, Huang; GUOCHAO, Qian; YU, Zhu *et al.* Fire prediction based on catboost algorithm. **Mathematical Problems in Engineering**, 2021.

FERNANDES, Filipe; STEFENON, Stéfano Frizzo; SEMAN, Laio Oriel; NIED, Ademir; FERREIRA, Fernanda Cristina Silva; SUBTIL, Maria Cristina Mazzetti; KLAAR, Anne Carolina Rodrigues; LEITHARDT, Valderi Reis Quietinho. Long short-term memory stacking model to predict the number of cases and deaths caused by COVID-19. **Journal of Intelligent & Fuzzy Systems**, IOS Press, v. 6, n. 42, p. 6221–6234, 2022.

FERREIRA, Leonardo N; VEGA-OLIVEROS, Didier A; ZHAO, Liang; CARDOSO, Manoel F; MACAU, Elbert EN. Global fire season severity analysis and forecasting. **Computers & geosciences**, Elsevier, v. 134, p. 104339, 2020.

FLANNIGAN, MD; WOTTON, BM; MARSHALL, GA; GROOT, WJ De; JOHNSTON, J; JURKO, N; CANTIN, AS. Fuel moisture sensitivity to temperature and precipitation: climate change implications. **Climatic Change**, Springer, v. 134, p. 59–71, 2016.

GAO, Chao; LIN, Honglei; HU, Haiqing. Forest-fire-risk prediction based on random forest and backpropagation neural network of heihe area in heilongjiang province, china. **Forests**, MDPI, v. 14, n. 2, p. 170, 2023.

GARCÊS, Andreia; PIRES, Isabel. The hell of wildfires: The impact on wildlife and its conservation and the role of the veterinarian. **Conservation**, MDPI, v. 3, n. 1, p. 96–108, 2023.

GHOLAMNIA, Khalil; NACHAPPA, Thimmaiah Gudiyangada; GHORBANZADEH, Omid; BLASCHKE, Thomas. Comparisons of diverse machine learning approaches for wildfire susceptibility mapping. **Symmetry**, MDPI, v. 12, n. 4, p. 604, 2020.

GHORBANZADEH, Omid; BLASCHKE, Thomas; GHOLAMNIA, Khalil; ARYAL, Jagannath. Forest fire susceptibility and risk mapping using social/infrastructural vulnerability and environmental variables. **Fire**, MDPI, v. 2, n. 3, p. 50, 2019.

GORYAEV, VM; BASANGOVA, EO; GOLDBVARG, TB; BEMBITOV, DB; DJACHNAEVA, EN; LIDZHI-GARYAEV, VV. Forecasting steppe fires using remote sensing data of time series. *In*: IOP PUBLISHING. **IOP Conference Series: Materials Science and Engineering**. [S.l.], 2021. v. 1047, p. 012092.

GULARTE, Luis Carlos Pais; BORTOLUZZI, Sandro César; LIMA, José Donizetti de; PINTO, Michele Aparecida Nepomuceno; GOFFI, Andréia dos Santos. Análise bibliométrica e sistêmica da literatura sobre viabilidade econômica-financeira, gestão e sustentabilidade da reciclagem de resíduos da construção civil. **Exacta**, Universidade Nove de Julho, v. 16, n. 4, p. 45–58, 2018.

HODSON, Timothy O. Root mean square error (rmse) or mean absolute error (mae): When to use them or not. **Geoscientific Model Development Discussions**, Göttingen, Germany, v. 2022, p. 1–10, 2022.

HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. Extreme learning machine: theory and applications. **Neurocomputing**, Elsevier, v. 70, n. 1-3, p. 489–501, 2006.

HYNDMAN, Rob J; ATHANASOPOULOS, G. 8.1 stationarity and differencing. **Forecasting: Principles and practices, Melbourne, Australia, OTexts**, 2013.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. [S.l.]: OTexts, 2018. Available at <https://otexts.com/fpp2/>.

HYNDMAN, Rob J; KOEHLER, Anne B. Another look at measures of forecast accuracy. **International journal of forecasting**, Elsevier, v. 22, n. 4, p. 679–688, 2006.

(INPA), Instituto Nacional de Pesquisas Ambientais. **Relatório Técnico: Efeitos de Incêndios Florestais em Recursos Hídricos**. [S.l.], 2021. Disponível em: <https://www.inpa.gov.br/relatorios/TR-2021-089>.

JAMSHED, Muhammad Ali; THEODOROU, Charalambos; KALSOOM, Tahera; ANJUM, Nadeem; ABBASI, Qammer H; UR-REHMAN, Masood. Intelligent computing based forecasting of deforestation using fire alerts: A deep learning approach. **Physical Communication**, Elsevier, v. 55, p. 101941, 2022.

JANIEC, Piotr; GADAL, Sébastien. A comparison of two machine learning classification methods for remote sensing predictive modeling of the forest fire in the north-eastern siberia. **Remote Sensing**, MDPI, v. 12, n. 24, p. 4157, 2020.

JESUS, Janisson Batista de; ROSA, Cristiano Niederauer da; BARRETO, Íkaro Daniel de Carvalho; FERNANDES, Milton Marques. Análise da incidência temporal, espacial e de tendência de fogo nos biomas e unidades de conservação do brasil. **Ciência Florestal**, SciELO Brasil, v. 30, p. 176–191, 2020.

KALANTAR, Bahareh; UEDA, Naonori; IDREES, Mohammed O; JANIZADEH, Saeid; AHMADI, Kouros; SHABANI, Farzin. Forest fire susceptibility prediction based on machine learning models with resampling algorithms on remote sensing data. **Remote Sensing**, MDPI, v. 12, n. 22, p. 3682, 2020.

KALE, Manish P; MISHRA, Asima; PARDESHI, Satish; GHOSH, Suddhasheel; PAI, DS; ROY, Parth Sarathi. Forecasting wildfires in major forest types of india. **Frontiers in Forests and Global Change**, Frontiers Media SA, v. 5, p. 882685, 2022.

KLAAR, Anne Carolina Rodrigues; STEFENON, Stefano Frizzo; SEMAN, Laio Oriel; MARIANI, Viviana Cocco; COELHO, Leandro Santos. Optimized EWT-Seq2Seq-LSTM with attention mechanism to insulators fault prediction. **Sensors**, MDPI, v. 23, n. 6, p. 3202, 2023.

KLAAR, Anne Carolina Rodrigues; STEFENON, Stefano Frizzo; SEMAN, Laio Oriel; MARIANI, Viviana Cocco; COELHO, Leandro Santos. Structure optimization of ensemble learning methods and seasonal decomposition approaches to energy price forecasting in Latin America: A case study about Mexico. **Energies**, MDPI, v. 16, n. 7, p. 3184, 2023.

KOLANEK, Aleksandra; SZYMANOWSKI, Mariusz; RACZYK, Andrzej. Human activity affects forest fires: The impact of anthropogenic factors on the density of forest fires in poland. **Forests**, MDPI, v. 12, n. 6, p. 728, 2021.

LACERDA, Rogério Tadeu de Oliveira; ENSSLIN, Leonardo; ENSSLIN, Sandra Rolim. Uma análise bibliométrica da literatura sobre estratégia e avaliação de desempenho. **Gestão & Produção**, SciELO Brasil, v. 19, p. 59–78, 2012.

LIN, Chuanying; LI, Xingdong; SHI, Tiefeng; SHENG, Junhao; SUN, Shufa; WANG, Yangwei; LI, Dandan. Forecasting of wind speed under wind-fire coupling scenarios by combining hs-vmd and am-lstm. **Ecological Informatics**, Elsevier, v. 77, p. 102270, 2023.

LIU, Jia; WANG, Yukuan; LU, Yafeng; ZHAO, Pengguo; WANG, Shunjiu; SUN, Yu; LUO, Yu. Application of remote sensing and explainable artificial intelligence (xai) for wildfire occurrence mapping in the mountainous region of southwest china. **Remote Sensing**, MDPI, v. 16, n. 19, p. 3602, 2024.

LIU, Yanli; WANG, Yourong; ZHANG, Jian. New machine learning algorithm: Random forest. *In*: SPRINGER. **Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings 3. [S.l.]**, 2012. p. 246–252.

LUNDBERG, Scott M; LEE, Su-In. A unified approach to interpreting model predictions. **Advances in neural information processing systems**, v. 30, 2017.

MALANDRA, Francesco; VITALI, Alessandro; MORRESI, Donato; GARBARINO, Matteo; FOSTER, Daniel E; STEPHENS, Scott L; URBINATI, Carlo. Burn severity drivers in italian large wildfires. **Fire**, MDPI, v. 5, n. 6, p. 180, 2022.

MEDEIROS, A.; SARTORI, A.; STEFENON, S. F.; MEYER, L. H.; NIED, A. Comparison of artificial intelligence techniques to failure prediction in contaminated insulators based on leakage current. **Journal of Intelligent & Fuzzy Systems**, v. 42, n. 4, p. 3285–3298, 2022.

MICHAEL, Yaron; HELMAN, David; GLICKMAN, Oren; GABAY, David; BRENNER, Steve; LENSKY, Itamar M. Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. **Science of The Total Environment**, Elsevier, v. 764, p. 142844, 2021.

MOHAJANE, Meriame; COSTACHE, Romulus; KARIMI, Firoozeh; PHAM, Quoc Bao; ESSAHLAOU, Ali; NGUYEN, Hoang; LANEVE, Giovanni; OUDIJA, Fatiha. Application of remote sensing and machine learning algorithms for forest fire mapping in a mediterranean area. **Ecological Indicators**, Elsevier, v. 129, p. 107869, 2021.

MORENO, Sinvaldo Rodrigues; SEMAN, Laio Oriel; STEFENON, Stefano Frizzo; COELHO, Leandro dos Santos; MARIANI, Viviana Cocco. Enhancing wind speed forecasting through synergy of machine learning, singular spectral analysis, and variational mode decomposition. **Energy**, v. 292, p. 130493, 2024.

MOTA, Pedro Henrique Santos; ROCHA, Samuel José Silva Soares da; CASTRO, Nero Lemos Martins de; MARCATTI, Gustavo Eduardo; FRANÇA, Luciano Cavalcante de Jesus; SCHETTINI, Bruno Leão Said; VILLANOVA, Paulo Henrique; SANTOS, Hugo Thaner dos;

SANTOS, Alexandre Rosa dos. Forest fire hazard zoning in mato grosso state, brazil. **Land use policy**, Elsevier, v. 88, p. 104206, 2019.

NUR, Arip Syaripudin; KIM, Yong Je; LEE, Joon Ho; LEE, Chang-Wook. Spatial prediction of wildfire susceptibility using hybrid machine learning models based on support vector regression in sydney, australia. **Remote Sensing**, MDPI, v. 15, n. 3, p. 760, 2023.

PANG, Yongqi; LI, Yudong; FENG, Zhongke; FENG, Zemin; ZHAO, Ziyu; CHEN, Shilin; ZHANG, Hanyue. Forest fire occurrence prediction in china based on machine learning methods. **Remote Sensing**, MDPI, v. 14, n. 21, p. 5546, 2022.

PARTHEEPAN, Shouthiri; SANATI, Farzad; HASSAN, Jahan. Modelling bushfire severity and predicting future trends in australia using remote sensing and machine learning. **Environmental Modelling & Software**, Elsevier, v. 188, p. 106377, 2025.

PENG, Wanyu; WEI, Yugui; CHEN, Guangsheng; LU, Guofan; YE, Qing; DING, Rungping; HU, Peng; CHENG, Zhenyu. Analysis of wildfire danger level using logistic regression model in sichuan province, china. **Forests**, MDPI, v. 14, n. 12, p. 2352, 2023.

PÉREZ-PORRAS, Fernando-Juan; TRIVIÑO-TARRADAS, Paula; CIMA-RODRÍGUEZ, Carmen; LARRIVA, Jose-Emilio Meroño-de; GARCÍA-FERRER, Alfonso; MESAS-CARRASCO, Francisco-Javier. Machine learning methods and synthetic data generation to predict large wildfires. **Sensors**, MDPI, v. 21, n. 11, p. 3694, 2021.

PINHEIRO, Ana Caroline; RODRIGUES, Paulo Canas. Hierarchical time series forecasting of fire spots in brazil: A comprehensive approach. **Stats**, v. 7, n. 3, p. 647–670, 2024. ISSN 2571-905X. Disponível em: <https://www.mdpi.com/2571-905X/7/3/39>.

PINHEIRO, Ana Caroline; RODRIGUES, Paulo Canas. Hierarchical time series forecasting of fire spots in brazil: A comprehensive approach. **Stats**, MDPI, v. 7, n. 3, p. 647–670, 2024.

POURGHASEMI, Hamid Reza; GAYEN, Amiya; LASAPONARA, Rosa; TIEFENBACHER, John P. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. **Environmental research**, Elsevier, v. 184, p. 109321, 2020.

REYES-BUENO, Fabián; LOJÁN-CÓRDOVA, Julia. Assessment of three machine learning techniques with open-access geographic data for forest fire susceptibility monitoring—evidence from southern ecuador. **Forests**, MDPI, v. 13, n. 3, p. 474, 2022.

RIBEIRO, Matheus Henrique Dal Molin. **Time series forecasting based on ensemble learning methods applied to agribusiness, epidemiology, energy demand, and renewable energy**. 2021. Tese (Doutorado) — Pontifícia Universidade Católica do Paraná, 2021.

RIBEIRO, Matheus Henrique Dal Molin; SILVA, Ramon Gomes da; MORENO, Sinvaldo Rodrigues; CANTON, Cristiane; LARCHER, José Henrique Kleinübing; STEFENON, Stefano Frizzo; MARIANI, Viviana Cocco; COELHO, Leandro dos Santos. Variational mode decomposition and bagging extreme learning machine with multi-objective optimization for wind power forecasting. **Applied Intelligence**, v. 54, p. 3119–3134, 2024.

RIBEIRO, Matheus Henrique Dal Molin; SILVA, Ramon Gomes da; MORENO, Sinvaldo Rodrigues; CANTON, Cristiane; LARCHER, José Henrique Kleinübing; STEFENON, Stefano Frizzo; MARIANI, Viviana Cocco; COELHO, Leandro dos Santos. Variational mode decomposition and bagging extreme learning machine with multi-objective optimization for wind power forecasting. **Applied Intelligence**, Springer, v. 54, n. 4, p. 3119–3134, 2024.

RIBEIRO, Matheus Henrique Dal Molin; SILVA, Ramon Gomes da; RIBEIRO, Gabriel Trierweiler; MARIANI, Viviana Cocco; COELHO, Leandro dos Santos. Cooperative ensemble learning model improves electric short-term load forecasting. **Chaos, Solitons & Fractals**, Elsevier, v. 166, p. 112982, 2023.

ROCHA, Vinicius Lovatel. **Forest fires data in Xanxerê and Lages – Santa Catarina**. 2025. <https://github.com/ViniLova/FOREST-FIRES-DATA-IN-XANXER-AND-LAGES---SANTA-CATARINA>. Acesso em: 11 ago. 2025.

RODRIGUES, Vinícius Barros; TORRES, Fillpe Tamiozzo Pereira. Fire behavior prediction using machine learning algorithms. **Brazilian Journal of Biometrics**, v. 38, n. 3, p. 343–352, 2020.

ROSHAN, Khushnaseeb; ZAFAR, Aasim. Utilizing xai technique to improve autoencoder based model for computer network anomaly detection with shapley additive explanation (shap). **arXiv preprint arXiv:2112.08442**, 2021.

SANTOS, Bruna Zamith; SORIANO, Balbina Maria Araujo; NARCISO, Marcelo Gonçalves; SILVA, Diego Furtado; CERRI, Ricardo. A new time series framework for forest fire risk forecasting and classification. *In: IEEE. 2023 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2023. p. 1–8.

SAYAD, Younes Oulad; MOUSANNIF, Hajar; MOATASSIME, Hassan Al. Predictive modeling of wildfires: A new dataset and machine learning approach. **Fire safety journal**, Elsevier, v. 104, p. 130–146, 2019.

SEMAN, Laio Oriel; STEFENON, Stefano Frizzo; MARIANI, Viviana Cocco; COELHO, Leandro Santos. Ensemble learning methods using the Hodrick–Prescott filter for fault forecasting in insulators of the electrical power grids. **International Journal of Electrical Power & Energy Systems**, v. 152, p. 109269, 2023. ISSN 0142-0615.

SHAO, Yakui; FENG, Zhongke; CAO, Meng; WANG, Wenbiao; SUN, Linhao; YANG, Xuanhan; MA, Tiantian; GUO, Zanquan; FAHAD, Shahzad; LIU, Xiaohan *et al.* An ensemble model for forest fire occurrence mapping in china. **Forests**, MDPI, v. 14, n. 4, p. 704, 2023.

SHAO, Yakui; FENG, Zhongke; SUN, Linhao; YANG, Xuanhan; LI, Yudong; XU, Bo; CHEN, Yuan. Mapping china's forest fire risks with machine learning. **Forests**, MDPI, v. 13, n. 6, p. 856, 2022.

SHARMA, Richa; RANI, Shalli; MEMON, Imran. A smart approach for fire prediction under uncertain conditions using machine learning. **Multimedia Tools and Applications**, Springer, v. 79, n. 37, p. 28155–28168, 2020.

SHERIDAN, Robert P; WANG, Wei Min; LIAW, Andy; MA, Junshui; GIFFORD, Eric M. Extreme gradient boosting as a method for quantitative structure–activity relationships. **Journal of chemical information and modeling**, ACS Publications, v. 56, n. 12, p. 2353–2360, 2016.

SHMUEL, Assaf; HEIFETZ, Eyal. Global wildfire susceptibility mapping based on machine learning models. **Forests**, MDPI, v. 13, n. 7, p. 1050, 2022.

SHMUEL, Assaf; HEIFETZ, Eyal. A machine-learning approach to predicting daily wildfire expansion rate. **Fire**, MDPI, v. 6, n. 8, p. 319, 2023.

SILVA, L. O.; PEREIRA, R. M. Políticas de mitigação de impactos ambientais em unidades de conservação. **Revista Brasileira de Gestão Ambiental**, v. 14, n. 3, p. 45–58, 2020.

SINGH, Kajol R; NEETHU, KP; MADHUREKAA, K; HARITA, A; MOHAN, Pushpa. Parallel svm model for forest fire prediction. **Soft Computing Letters**, Elsevier, v. 3, p. 100014, 2021.

SLAVIA, Athaya Putri; SUTOYO, Edi; WITARSYAH, Deden. Hotspots forecasting using autoregressive integrated moving average (arima) for detecting forest fires. *In: IEEE. 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS). [S.l.]*, 2019. p. 92–97.

SMOLA, Alex J; SCHÖLKOPF, Bernhard. A tutorial on support vector regression. **Statistics and computing**, Springer, v. 14, p. 199–222, 2004.

SOARES, Ronaldo Viana; PAEZ, Gilberto. Correlações entre alguns fatores meteorológicos e ocorrências de incêndios florestais na região centro paranaense. **Floresta**, v. 4, n. 2, 1972.

SONG, Sida; ZHOU, Xiao; YUAN, Shangbo; CHENG, Pengle; LIU, Xiaodong. Interpretable artificial intelligence models for predicting lightning prone to inducing forest fires. **Journal of Atmospheric and Solar-Terrestrial Physics**, Elsevier, v. 267, p. 106408, 2025.

SONG, Yongjia; WANG, Yuhang. Global wildfire outlook forecast with neural networks. **Remote Sensing**, MDPI, v. 12, n. 14, p. 2246, 2020.

STEFENON, Stefano Frizzo; SEMAN, Laio Oriel; AQUINO, Luiza Scapinello; COELHO, Leandro dos Santos. Wavelet-Seq2Seq-LSTM with attention for time series forecasting of level of dams in hydroelectric power plants. **Energy**, Elsevier, v. 274, p. 127350, 2023.

STEFENON, Stefano Frizzo; SEMAN, Laio Oriel; da Silva, Luiza Scapinello Aquino; MARIANI, Viviana Cocco; COELHO, Leandro dos Santos. Hypertuned temporal fusion transformer for multi-horizon time series forecasting of dam level in hydroelectric power plants. **International Journal of Electrical Power & Energy Systems**, v. 157, p. 109876, 2024.

STEFENON, Stefano Frizzo; SEMAN, Laio Oriel; MARIANI, Viviana Cocco; COELHO, Leandro Santos. Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices. **Energies**, v. 16, n. 3, p. 1371, 2023.

STEFENON, Stefano Frizzo; SEMAN, Laio Oriel; NETO, Nemesio Fava Sopelsa; MEYER, Luiz Henrique; MARIANI, Viviana Cocco; COELHO, Leandro dos Santos. Group method of data handling using Christiano-Fitzgerald random walk filter for insulator fault prediction. **Sensors**, v. 23, n. 13, p. 6118, 2023.

STEFENON, Stefano F; SEMAN, Laio O; YAMAGUCHI, Cristina K; COELHO, Leandro Dos S; MARIANI, Viviana C; MATOS-CARVALHO, João P; LEITHARDT, Valderi RQ. Neural hierarchical interpolation time series for reservoir level multi-horizon forecasting in hydroelectric power plants. **IEEE Access**, IEEE, 2025.

STEFENON, Stefano Frizzo; SINGH, Gurmail; SOUZA, Bruno José; FREIRE, Roberto Zanetti; YOW, Kin-Choong. Optimized hybrid YOLOu-Quasi-ProtoPNet for insulators classification. **IET Generation, Transmission & Distribution**, v. 17, n. 15, p. 3501–3511, 2023.

SYARIFUDIN, Mohammad Anang; NOVITASARI, Dian Candra Rini; MARPAUNG, Faridawaty; WAHYUDI, Noor; HAPSARI, Dian Puspita; SUPRIYATI, Endang; FARIDA, Yuniar; AMIN, Faris Muslihul; NUGRAHENI, RR Diah; NARISWARI, Rinda *et al.* Hotspot prediction using 1d convolutional neural network. **Procedia Computer Science**, Elsevier, v. 179, p. 845–853, 2021.

TANG, Xianzhe; MACHIMURA, Takashi; LI, Jiufeng; LIU, Wei; HONG, Haoyuan. A novel optimized repeatedly random undersampling for selecting negative samples: A case study in

an svm-based forest fire susceptibility assessment. **Journal of Environmental Management**, Elsevier, v. 271, p. 111014, 2020.

TIBSHIRANI, Robert. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, Oxford University Press, v. 58, n. 1, p. 267–288, 1996.

TRUCCHIA, Andrea; MESCHI, Giorgio; FIORUCCI, Paolo; GOLLINI, Andrea; NEGRO, Dario. Defining wildfire susceptibility maps in Italy for understanding seasonal wildfire regimes at the national level. **Fire**, MDPI, v. 5, n. 1, p. 30, 2022.

VIEGAS, Domingos Xavier. Forest fire propagation. **Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 356, n. 1748, p. 2907–2928, 1998.

VIGANÓ, Hevelyne Henn da Gama; SOUZA, Celso Correia de; NETO, José Francisco Reis; CRISTALDO, Marcia Ferreira; JESUS, Leandro de. Prediction and modeling of forest fires in the pantanal. **Revista Brasileira de Meteorologia**, Revista Brasileira de Meteorologia, v. 33, n. 2, p. 306–316, 2018.

WANG, Zili; HE, Binbin; CHEN, Rui; FAN, Chunquan. Improving wildfire danger assessment using time series features of weather and fuel in the Great Xing'an Mountain region, China. **Forests**, MDPI, v. 14, n. 5, p. 986, 2023.

WILLMOTT, Cort J. On the validation of models. **Physical Geography**, Taylor & Francis, v. 2, n. 2, p. 184–194, 1981.

XIE, Ying; PENG, Minggang. Forest fire forecasting using ensemble learning approaches. **Neural Computing and Applications**, Springer, v. 31, n. 9, p. 4541–4550, 2019.

YAMASAKI, Marcos; FREIRE, Roberto Zanetti; SEMAN, Laio Oriel; STEFENON, Stefano Frizzo; MARIANI, Viviana Cocco; dos Santos Coelho, Leandro. Optimized hybrid ensemble learning approaches applied to very short-term load forecasting. **International Journal of Electrical Power & Energy Systems**, v. 155, p. 109579, 2024.

YOSHIDA, Nelson Daishiro. Análise bibliométrica: um estudo aplicado à previsão tecnológica. **Future Studies Research Journal: Trends and Strategies**, v. 2, n. 1, p. 52–84, 2010.

YOUSEFI, Saleh; POURGHASEMI, Hamid Reza; EMAMI, Sayed Naeim; POUYAN, Soheila; ESKANDARI, Saeedeh; TIEFENBACHER, John P. A machine learning framework for multi-hazards modeling and mapping in a mountainous area. **Scientific Reports**, Nature Publishing Group UK London, v. 10, n. 1, p. 12144, 2020.

ZHA, Wenshu; LIU, Yuping; WAN, Yujin; LUO, Ruilan; LI, Daolun; YANG, Shan; XU, Yanmei. Forecasting monthly gas field production based on the cnn-lstm model. **Energy**, Elsevier, v. 260, p. 124889, 2022.

ZHANG, Guoli; WANG, Ming; LIU, Kai. Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china. **International Journal of Disaster Risk Science**, Springer, v. 10, n. 3, p. 386–403, 2019.

ZHANG, Zhen; WANG, Leilei; XUE, Naiting; DU, Zhiheng. Spatiotemporal analysis of active fires in the arctic region during 2001–2019 and a fire risk assessment model. **Fire**, MDPI, v. 4, n. 3, p. 57, 2021.

ZHAO, Yonghui; PENG, Xunhui; TU, Teng; LI, Zhen; YAN, Peiyu; LI, Chao. Woa-vmd-scinet: Hybrid model for accurate prediction of ultra-short-term photovoltaic generation power considering seasonal variations. **Energy Reports**, Elsevier, v. 12, p. 3470–3487, 2024.

ZHOU, Yang; LIN, Peihui; WANG, Naiyu. A deep neural network approach for regional-scale 30-day accumulated urban fire occurrence forecast. **Fire Safety Journal**, Elsevier, v. 152, p. 104331, 2025.

ZOU, Hui; HASTIE, Trevor. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, Oxford University Press, v. 67, n. 2, p. 301–320, 2005.

ZUEGE, Cesar Vinicius; STEFENON, Stefano Frizzo; YAMAGUCHI, Cristina Keiko; MARIANI, Viviana Cocco; GONZALEZ, Gabriel Villarrubia; COELHO, Leandro dos Santos. Wind speed forecasting approach using conformal prediction and feature importance selection. **International Journal of Electrical Power & Energy Systems**, Elsevier, v. 168, p. 110700, 2025.

ANEXO A – AUTORIZAÇÃO CO-AUTORES PARA INCLUSÃO DE PESQUISA EM DISSERTAÇÃO

E-mail de Universidade Tecnológica Federal do Paraná - Autorização pa... <https://mail.google.com/mail/u/1/?ik=0449cd70da&view=pt&search=a...>



VINICIUS LOVATEL ROCHA <viniciuslovatel@alunos.utfpr.edu.br>

Autorização para o artigo compor a dissertação.

3 mensagens

VINICIUS LOVATEL ROCHA <viniciuslovatel@alunos.utfpr.edu.br> 1 de outubro de 2025 às 10:52
Para: gilson@utfpr.edu.br, ramon.gsilva@pucpr.br

Bom dia,

Gostaria de pedir a autorização dos senhores para que o artigo "Wildfire Spots Analysis and Forecasting: Evaluation of Univariate and Multivariate Based on Variational Mode Decomposition Models" submetido ao periódico Applied Soft Computing possa compor a dissertação "ANÁLISE E PREVISÃO DE INCÊNDIOS FLORESTAIS PARA REGIÃO DE LAGES E XANXERÊ - SANTA CATARINA: UMA ABORDAGEM EXPLORANDO MODELOS HÍBRIDOS DE PREVISÃO".

Atenciosamente,

Vinicius Lovatel Rocha

gilson@utfpr.edu.br <gilson@utfpr.edu.br> 1 de outubro de 2025 às 10:53
Para: VINICIUS LOVATEL ROCHA <viniciuslovatel@alunos.utfpr.edu.br>, "ramon.gsilva@pucpr.br"
<ramon.gsilva@pucpr.br>

De pleno acordo,

Gilson Adamczuk Oliveira

De: VINICIUS LOVATEL ROCHA <viniciuslovatel@alunos.utfpr.edu.br>
Enviado: quarta-feira, 1 de outubro de 2025 10:52
Para: gilson@utfpr.edu.br <gilson@utfpr.edu.br>; ramon.gsilva@pucpr.br <ramon.gsilva@pucpr.br>
Assunto: Autorização para o artigo compor a dissertação.

[Texto das mensagens anteriores oculto]

Ramon Gomes da Silva <ramon.gsilva@pucpr.br> 1 de outubro de 2025 às 10:56
Para: VINICIUS LOVATEL ROCHA <viniciuslovatel@alunos.utfpr.edu.br>, "gilson@utfpr.edu.br" <gilson@utfpr.edu.br>

De acordo, Vinicius.

Ramon

From: VINICIUS LOVATEL ROCHA <viniciuslovatel@alunos.utfpr.edu.br>
Sent: Wednesday, October 1, 2025 10:52:13 AM
To: gilson@utfpr.edu.br <gilson@utfpr.edu.br>; Ramon Gomes da Silva <ramon.gsilva@pucpr.br>
Subject: Autorização para o artigo compor a dissertação.

Geralmente, você não recebe emails de viniciuslovatel@alunos.utfpr.edu.br. [Saiba por que isso é importante](#)

[Texto das mensagens anteriores oculto]

Figura 28 – Autorização dos coautores para a inclusão do artigo "Wildfire Spots Analysis and Forecasting: Evaluation of Univariate and Multivariate Based on Variational Mode Decomposition Models" na dissertação.