



ESPECIALIZAÇÃO EM INTELIGÊNCIA ARTIFICIAL

Fernando dos Santos Alves Fernandes

Previsão de Séries Temporais de Evapotranspiração de Referência com Modelos Multivariados nos Diferentes Climas do Brasil

Ouro Preto - MG
2025

Fernando dos Santos Alves Fernandes

**Previsão de Séries Temporais de
Evapotranspiração de Referência com Modelos
Multivariados nos Diferentes Climas do Brasil**

Trabalho de Conclusão do Curso apresentado ao curso de Especialização em Inteligência Artificial, ofertado pelo Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - *Campus* Ouro Preto, como parte dos requisitos para a obtenção do título de Especialista em Inteligência Artificial.

Orientador: Prof. Dr. Carlos Alberto Severiano Junior.

Ouro Preto - MG

2025

F363p

Fernandes, Fernando dos Santos Alves.

Previsão de séries temporais de evapotranspiração de referência com modelos multivariados nos diferentes climas do Brasil [manuscrito] / Fernando dos Santos Alves Fernandes . – 2025.
95 f. : il.

Orientador: Carlos Alberto Severiano Júnior.

Trabalho de Conclusão de Curso (especialização) – Instituto Federal de Minas Gerais. *Campus* Ouro Preto, 2025.

1. Evapotranspiração de referência. 2. Modelos de previsão multivariada. 3. Domínios climáticos do Brasil. I. Severiano Júnior, Carlos Alberto. II. Instituto Federal de Minas Gerais. *Campus* Ouro Preto. III. Título.

CDU: 004.8

Catálogo: Kelly Cristiane Santos Morais - CRB-6/3217



MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS
Campus Ouro Preto
Diretoria de Pesquisa, Inovação e Pós-Graduação
Coordenação do Curso de Pós-graduação em Inteligência Artificial
Rua Pandiá Calógeras, 898 - Bairro Bauxita - CEP 35400-000 - Ouro Preto - MG
- www.ifmg.edu.br

FOLHA DE APROVAÇÃO

FERNANDO DOS SANTOS ALVES FERNANDES

**PREVISÃO DE SÉRIES TEMPORAIS DE EVAPOTRANSPIRAÇÃO DE REFERÊNCIA
COM MODELOS MULTIVARIADOS NOS DIFERENTES CLIMAS DO BRASIL**

Trabalho de Conclusão de Curso apresentado ao curso de ESPECIALIZAÇÃO EM INTELIGÊNCIA ARTIFICIAL, ofertado pelo Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - *Campus* Ouro Preto, como parte dos requisitos para a obtenção do título de ESPECIALISTA EM INTELIGÊNCIA ARTIFICIAL.

Aprovado em 11 de fevereiro de 2025, pela Banca Examinadora:

Prof. Dr. Carlos Alberto Severiano Junior - IFMG *Campus* Sabará - Orientador

Prof. Ma. Patrícia Oliveira e Lucas - IFNMG *Campus* Salinas

Prof. Dr. Rodrigo César Pedrosa Silva (DECOM/UFOP)



Documento assinado eletronicamente por **Carlos Alberto Severiano Junior, Professor**, em 18/02/2025, às 14:01, conforme Decreto nº 10.543, de 13 de novembro de 2020.



A autenticidade do documento pode ser conferida no site <https://sei.ifmg.edu.br/consultadocs> informando o código verificador **2205541** e o código CRC **E39A1389**.

23213.000151/2025-24

2205541v1

Documento assinado digitalmente
gov.br PATRICIA DE OLIVEIRA ELUCAS
Data: 18/02/2025 14:23:59-0300
Verifique em <https://validar.iti.gov.br>

Documento assinado digitalmente
gov.br RODRIGO CESAR PEDROSA SILVA
Data: 18/02/2025 15:47:45-0300
Verifique em <https://validar.iti.gov.br>

Resumo

Para alcançar o equilíbrio entre o aumento previsto na demanda alimentar e a necessidade de expansão das áreas agrícolas irrigadas, num cenário de mudanças climáticas em que a água é um recurso cada vez mais escasso, é fundamental estimar com precisão a demanda hídrica das culturas, o que pode ser feito por meio da evapotranspiração de referência (*ET_o*). Em cenários de escassez de dados meteorológicos e em regiões com condições climáticas variadas, estimar essa grandeza torna-se uma tarefa ainda mais desafiadora. Neste contexto, este trabalho busca avaliar o desempenho de diferentes modelos de previsão multivariada da evapotranspiração de referência, utilizando dados meteorológicos, nos seis principais tipos climáticos do Brasil. Para tanto, foram selecionadas seis bases de dados representativas dos diferentes domínios climáticos e aplicados quatro modelos multivariados de previsão de séries temporais: um modelo *Vector Autorregressive* (VAR), o modelo *Random Forest* (RF), uma *Convolutional Neural Network* (CNN) e um modelo baseado na arquitetura *Transformer*. Métodos específicos de cada modelo foram utilizados para determinar as importâncias das variáveis meteorológicas na previsão da *ET_o* em cada um dos climas. Os resultados em termos de RMSE (*Root Mean Squared Error*) foram comparados em dois cenários: todas as variáveis meteorológicas como entrada; apenas as quatro variáveis de maior importância, no quadro geral, como entrada. Os resultados mostraram que o modelo clássico (VAR) e o modelo baseado em árvore de decisão (RF) tiveram melhor desempenho em todos os domínios climáticos, comparados aos modelos CNN e *Transformer*. Além disso, a utilização somente das quatro variáveis de maior importância teve efeito significativo apenas no comportamento da CNN, que apresentou médias de RSME menores em relação aos seus valores do primeiro cenário, em que todas as variáveis meteorológicas foram utilizadas.

Palavras-chave: evapotranspiração de referência, modelos de previsão multivariada, domínios climáticos do Brasil, cálculo das importâncias.

Abstract

To achieve a balance between the expected increase in food demand and the need to expand irrigated agricultural areas, in a climate change scenario where water is an increasingly scarce resource, it is essential to accurately estimate crop water demand, which can be done through reference evapotranspiration (ET_o). In scenarios with limited meteorological data and in regions with varied climatic conditions, estimating this measure becomes an even more challenging task. In this context, this study seeks to evaluate the performance of different multivariate forecasting models for reference evapotranspiration, using meteorological data, across the six main climate types in Brazil. To this end, six datasets representative of different climate domains were selected, and four multivariate time series forecasting models were implemented: a Vector Autoregressive (VAR) model, a Random Forest (RF) model, a model based on convolutional neural networks (CNN), and a model based on the Transformer architecture. Specific methods for each model were used to determine the importance of meteorological variables in forecasting ET_o for each climate. The results in terms of Root Mean Squared Error (RMSE) were compared in two scenarios: using all meteorological variables as input; and using only the four most important variables overall as input. The results showed that the classical model (VAR) and the decision-tree-based model (RF) outperformed the CNN and Transformer-based models across all climate domains. Furthermore, using only the four most important variables had a significant effect only on the performance of the CNN model, which showed lower average RMSE values compared to its results in the first scenario, where all meteorological variables were used.

Keywords: reference evapotranspiration, multivariate forecasting models, climate domains of Brazil, variable importance calculation.

Lista de Figuras

Figura 1 – Mapa climático do Brasil de acordo com a classificação de Köppen-Geiger. Adaptado de (ALVARES et al., 2013)	22
Figura 2 – Mapa das diferentes zonas climáticas agrupadas pela temperatura e umidade. Classificação por Climas Zonais. Adaptado de (IBGE, 1993)	23
Figura 3 – Mapa dos climas do Brasil, com seis classificações e principais correntes marítimas. Adaptado de (OLIVEIRA, 2020)	24
Figura 4 – Base de dados selecionada para o clima Equatorial (EQ).	42
Figura 5 – Base de dados selecionada para o clima Semiárido (SA).	43
Figura 6 – Base de dados selecionada para o clima Tropical Litorâneo (TL).	43
Figura 7 – Base de dados selecionada para o clima Tropical de Altitude (TA).	44
Figura 8 – Base de dados selecionada para o clima Tropical Típico (TT).	45
Figura 9 – Base de dados selecionada para o clima Subtropical (ST).	46
Figura 10 – Séries temporais no clima Equatorial (EQ).	47
Figura 11 – Séries temporais no clima Semiárido (SA).	48
Figura 12 – Séries temporais no clima Tropical Litorâneo (TL).	48
Figura 13 – Séries temporais no clima Tropical Típico (TT).	49
Figura 14 – Séries temporais no clima Tropical de Altitude (TA).	49
Figura 15 – Séries temporais no clima Subtropical (ST).	50
Figura 16 – Divisão das bases de dados em 80% para treinamento e 20% para teste.	53
Figura 17 – Divisão das bases de dados em conjuntos de treinamento e teste com 5 janelas deslizantes, utilizando TimeSeriesSplit, da biblioteca scikit-learn.	54
Figura 18 – Barplot do RMSE dos modelos RF, CNN e modelo baseado em <i>Transformer</i> e média do modelo VAR, em cada clima do Brasil, com todas as variáveis como entrada.	62

Figura 19 – Barplot do RMSE dos modelos RF, CNN e <i>Transformer</i> e média do modelo VAR, em cada clima do Brasil, apenas com as variáveis de maior importância.	64
Figura 20 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Equatorial (EQ).	75
Figura 21 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Semiárido (SA).	76
Figura 22 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Tropical Litorâneo (TL).	77
Figura 23 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Tropical Típico (TT).	78
Figura 24 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Tropical de Altitude (TA).	79
Figura 25 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Subtropical (ST).	80
Figura 26 – Importância das variáveis para o modelo RF no clima Equatorial (EQ).	81
Figura 27 – Importância das variáveis para o modelo RF no clima Semiárido (SA).	81
Figura 28 – Importância das variáveis para o modelo RF no clima Tropical Litorâneo (TL).	82
Figura 29 – Importância das variáveis para o modelo RF no clima Tropical Típico.	82
Figura 30 – Importância das variáveis para o modelo RF no clima Tropical de Altitude (TA).	83
Figura 31 – Importância das variáveis para o modelo RF no clima Subtropical (ST).	83
Figura 32 – Impacto das variáveis meteorológicas para o modelo CNN no clima Equatorial (EQ).	84
Figura 33 – Impacto das variáveis meteorológicas para o modelo CNN no clima Semiárido (SA).	85
Figura 34 – Impacto das variáveis meteorológicas para o modelo CNN no clima Tropical Litorâneo (TL).	86

Figura 35 – Impacto das variáveis meteorológicas para o modelo CNN no clima Tropical Típico (TT).	87
Figura 36 – Impacto das variáveis meteorológicas para o modelo CNN no clima Tropical de Altitude (TA).	88
Figura 37 – Impacto das variáveis meteorológicas para o modelo CNN no clima Subtropical (ST).	89
Figura 38 – Impacto das variáveis meteorológicas para o modelo <i>Transformer</i> no clima Equatorial (EQ).	90
Figura 39 – Impacto das variáveis meteorológicas para o modelo <i>Transformer</i> no clima Semiárido (SA).	91
Figura 40 – Impacto das variáveis meteorológicas para o modelo <i>Transformer</i> no clima Tropical Litorâneo (TL).	92
Figura 41 – Impacto das variáveis meteorológicas para o modelo <i>Transformer</i> no clima Tropical Típico (TT).	93
Figura 42 – Impacto das variáveis meteorológicas para o modelo <i>Transformer</i> no clima Tropical de Altitude (TA).	94
Figura 43 – Impacto das variáveis meteorológicas para o modelo <i>Transformer</i> no clima Subtropical (ST).	95

Lista de Tabelas

Tabela 1 – Descrição estatística das séries no clima Equatorial (EQ).	50
Tabela 2 – Descrição estatística das séries no clima Semiárido (SA).	50
Tabela 3 – Descrição estatística das séries no clima Tropical Litorâneo (TL).	51
Tabela 4 – Descrição estatística das séries no clima Tropical Típico (TT).	51
Tabela 5 – Descrição estatística das séries no clima Tropical de Altitude (TA).	51
Tabela 6 – Descrição estatística das séries no clima Subtropical (ST).	51
Tabela 7 – Hiperparâmetros definidos para o modelo VAR.	55
Tabela 8 – Hiperparâmetros definidos para o modelo RF.	55
Tabela 9 – Hiperparâmetros definidos para o modelo CNN.	55
Tabela 10 – Hiperparâmetros definidos para o modelo baseado em <i>Transformer</i>	55
Tabela 11 – Variáveis de maior importância para cada modelo nos diferentes climas	60
Tabela 12 – Valores e incertezas de RMSE, em mmd^{-1} , dos modelos VAR, RF, CNN e <i>Transformer-based</i> nos diferentes climas do Brasil.	62
Tabela 13 – Valores e incertezas de RMSE, em mmd^{-1} , dos modelos VAR, RF, CNN e <i>Transformer-based</i> nos diferentes climas do Brasil, utilizando apenas 4 variáveis.	63

Sumário

1	INTRODUÇÃO	13
1.1	Objetivos	15
1.2	Organização do Trabalho	15
2	REFERENCIAL TEÓRICO	17
2.1	Evapotranspiração de Referência	17
2.1.1	Radiação solar	19
2.1.2	Temperatura do ar	20
2.1.3	Umidade do ar	20
2.1.4	Velocidade do vento	20
2.2	Domínios climáticos do Brasil	21
2.3	Modelos Multivariados	26
2.3.1	Modelagem de Múltiplas Séries Temporais	26
2.3.2	Modelo Vetorial Autorregressivo	27
2.3.3	Modelo Floresta Aleatória	28
2.3.4	Redes Neurais Convolucionais para Séries Temporais	30
2.3.5	Modelos Baseados em <i>Transformer</i>	32
2.3.6	Importância das Variáveis de Entrada	33
2.4	Trabalhos Relacionados	36
3	MÉTODO DO TRABALHO	40
3.1	Seleção das Bases de Dados	40
3.2	Análise Descritiva das Séries Meteorológicas	47
3.3	Experimentos Realizados	52
4	RESULTADOS E DISCUSSÃO	56
4.1	Análise de Importância das Séries Meteorológicas	56
4.2	Análise dos Resultados	61
4.2.1	Previsões com Todas as Variáveis	61
4.2.2	Previsões Apenas com as Variáveis de Maior Importância	62

4.2.3	Uma Breve Discussão dos Resultados	64
5	CONSIDERAÇÕES FINAIS	66
	REFERÊNCIAS	68
	APÊNDICE A – FUNÇÕES IMPULSO-RESPOSTA PARA O VAR	75
	APÊNDICE B – <i>GINI IMPORTANCE</i> PARA O RF	81
	APÊNDICE C – VALORES SHAP PARA A CNN	84
	APÊNDICE D – VALORES SHAP PARA O <i>TRANS-</i> <i>FORMER</i>	90

1 Introdução

De acordo com a Agência Nacional de Águas e Saneamento Básico (ANA), o Brasil possui mais de 8,2 milhões de hectares de terras agrícolas irrigadas, com potencial para expansão para mais de 13 milhões de hectares até 2040 (Agência Nacional de Águas e Saneamento Básico, 2021). A irrigação é o principal consumidor de água doce no mundo, correspondendo a 70% do total disponível, segundo a Organização das Nações Unidas para Alimentação e Agricultura (FAO). No entanto, diante das mudanças climáticas e do aumento previsto na demanda alimentar (cerca de 60% até 2050, para atender às necessidades alimentares de uma população global crescente), a quantidade de água destinada à agricultura não pode exceder 10% de aumento, devido à escassez desse recurso (Food and Agriculture Organization of the United Nations, 2017).

Esse cenário destaca a importância de um manejo eficiente da irrigação. Os projetos de irrigação devem garantir que as culturas recebam a quantidade de água necessária para seu desenvolvimento, evitando tanto o desperdício quanto o *déficit* hídrico. Para alcançar esse equilíbrio, é fundamental estimar com precisão a demanda hídrica das culturas, o que pode ser feito por meio da evapotranspiração de referência (*ET_o*) (Serviço Nacional de Aprendizagem Rural, 2019).

A *ET_o* representa a evapotranspiração de uma cultura padrão, geralmente grama ou alfafa, cultivada sob condições ideais de solo e clima e é influenciada por diversas variáveis meteorológicas, como temperatura máxima e mínima, umidade relativa do ar, velocidade do vento, precipitação e radiação solar. A estimativa de evapotranspiração da planta de referência permite obter, a partir de coeficientes específicos, a demanda hídrica para as principais culturas agrícolas, como o milho, soja, hortaliças e frutas (ALLEN; PEREIRA; SMITH, 1998; LUCAS, 2019; SHALOO et al., 2024).

Modelos empíricos, como a equação de Penman-Monteith (FAO-56 PM) — considerado um dos métodos mais precisos — e o modelo Hargreaves-Samani (HS), devido à sua capacidade de lidar com um número limitado de variáveis meteorológicas, são amplamente utilizados para estimar a *ET_o* (ALLEN; PEREIRA;

SMITH, 1998; KRISHNA, 2019). Contudo, esses modelos enfrentam limitações em contextos de escassez de dados meteorológicos ou em regiões com condições climáticas variadas (FERREIRA et al., 2021). Diante disso, métodos mais avançados de aprendizado de máquina (*machine learning*) e aprendizado profundo (*deep learning*) estão sendo explorados como alternativas promissoras para melhorar a precisão dessas estimativas (LUCAS et al., 2018; LUCAS, 2019; MENESES et al., 2020; FERREIRA; CUNHA, 2020; FERREIRA; CUNHA; FILHO, 2022; GUO; CAO; LIN, 2024).

Ainda em relação à escassez de dados climáticos, em áreas com pouca ou nenhuma cobertura de estações meteorológicas, técnicas de interpolação têm sido adotadas para gerar estimativas meteorológicas. Esses métodos permitem que dados de estações disponíveis sejam interpolados, fornecendo informações climáticas para regiões desprovidas de medições diretas (XAVIER; KING; SCANLON, 2016; ABATZOGLOU; DOBROWSKI; HEGEWISCH, 2018). No Brasil, essa abordagem resultou em uma base de dados meteorológicos interpolados com resolução espacial de até $0,1^\circ \times 0,1^\circ$ (latitude por longitude), que inclui estimativas de *ET_o* e cobre todo o território nacional (XAVIER et al., 2022). Essa solução amplia a disponibilidade de dados para treinar e validar modelos de previsão.

Outro aspecto importante a se considerar na utilização de modelos de previsão de séries temporais para estimação da *ET_o* no Brasil está no fato de que o país apresenta uma significativa diversidade de domínios climáticos, abrangendo os tipos: equatorial, equatorial ameno, tropical, tropical ameno, subtropical, temperado, semiárido e árido (NOVAIS, 2023). Cada um desses climas apresenta características específicas que afetam as variáveis meteorológicas e, conseqüentemente, a *ET_o*.

Diante dessa complexidade, espera-se que as variáveis meteorológicas tenham diferentes graus de influência sobre a *ET_o* em cada uma dessas regiões. Portanto, é provável que diferentes modelos de previsão se comportem de maneiras distintas, dependendo das condições climáticas e da localização geográfica das bases de dados utilizadas (BASÍLIO, 2023; FLEISCHMANN et al., 2023).

1.1 Objetivos

O objetivo geral deste trabalho consistiu em avaliar o desempenho de diferentes modelos de previsão multivariada da evapotranspiração de referência (ET_o), utilizando dados meteorológicos (temperaturas máxima e mínima, umidade relativa do ar, velocidade do vento, precipitação e radiação solar), nos seis principais tipos climáticos do Brasil: equatorial, semiárido, tropical, tropical de altitude, tropical litorâneo e subtropical.

Para tanto, fez-se necessário atingir os seguintes objetivos específicos:

- Selecionar e organizar bases de dados meteorológicos relevantes para cada uma das regiões climáticas, garantindo a adequação e a consistência dos dados para uso nos modelos de previsão.
- Selecionar e implementar diferentes modelos multivariados, para a previsão da evapotranspiração de referência, considerando a diversidade de variáveis envolvidas.
- Aplicar ferramentas de avaliação das importâncias das variáveis meteorológicas na previsão da ET_o , em cada um dos modelos selecionados, viabilizando a realização de experimentos em dois cenários: um com todas as variáveis como preditores; outro apenas com as variáveis consideradas de maior importância para a previsão.
- Analisar os resultados para cada um dos climas, comparando o desempenho dos modelos com base em métricas estatísticas como o RMSE (*Root Mean Square Error*), a fim de identificar os modelos com melhores desempenhos nos dois cenários.

1.2 Organização do Trabalho

Este trabalho está dividido em cinco capítulos. O Capítulo 1 contextualiza o tema abordado neste trabalho e apresenta os objetivos da pesquisa. O Capítulo 2 aborda a fundamentação teórica, introduzindo os conceitos de evapotranspiração

e a aplicação de modelos multivariados de previsão climática. O Capítulo 3 detalha a metodologia utilizada, incluindo a seleção das bases de dados climáticos, a implementação dos modelos e as técnicas de validação. No Capítulo 4, são apresentados os resultados dos experimentos, comparando o desempenho dos modelos nos diferentes climas. Finalmente, o Capítulo 5 traz as considerações finais, destacando as conclusões da pesquisa e possíveis direções para estudos futuros.

2 Referencial Teórico

Neste capítulo, são apresentados os principais conceitos acerca do processo de estimação da evapotranspiração de referência, que pode ser utilizada para determinar a demanda hídrica de uma cultura. A relação entre as variáveis meteorológicas e sua correlação com a ET_o também são apresentadas. A seção ainda apresenta definições e características de cada um dos modelos de previsão multivariada selecionados para a realização deste trabalho.

2.1 Evapotranspiração de Referência

A evapotranspiração (ET) é o processo pelo qual a água da superfície terrestre passa para a atmosfera, do estado líquido para o estado de vapor. Esse processo é resultado da soma da transpiração das plantas com a evaporação direta da água, seja ela proveniente da superfície das plantas molhadas, do solo úmido ou da cobertura do solo. Em termos gerais, pode-se definir evapotranspiração como “a soma da quantidade de água perdida por uma superfície coberta por vegetação, por meio da evaporação direta da superfície do solo e da água perdida pelas plantas por transpiração” (FILHO et al., 2011; JENSEN; ALLEN, 2016).

A perda de água para a atmosfera depende de fatores energéticos e atmosféricos, como a radiação solar, a temperatura do ar, a velocidade do vento e a umidade relativa. A radiação solar e a temperatura do ar estão diretamente relacionadas ao aumento da evaporação e transpiração, pois fornecem a energia necessária para que a água passe para o estado de vapor. A velocidade do vento também influencia, removendo a camada de ar úmido próximo à superfície e aumentando a evaporação. Por outro lado, a umidade relativa atua de forma inversa, reduzindo a evaporação quanto maior for sua presença no ar (ALLEN et al., 2006; LUCAS, 2019).

Como a evapotranspiração varia dependendo da vegetação, das condições do solo e do clima, surgiu a necessidade de padronizar seu processo de estimação para permitir comparações consistentes entre diferentes regiões e climas. É nesse contexto que entra o conceito de evapotranspiração de referência. A padronização

da evapotranspiração por meio da ET_o é importante para o manejo eficiente da irrigação e para a previsão de demandas hídricas em diferentes condições climáticas, facilitando a comparabilidade e a aplicação em modelos de previsão (ALLEN; PEREIRA; SMITH, 1998).

A ET_o pode ser definida como a taxa de evapotranspiração que ocorre em uma hipotética cultura de referência (normalmente grama ou, em países frios, alfafa), com uma área extensa e com condições ideais de crescimento, sem restrição hídrica. Essa cultura de referência é caracterizada por uma vegetação em crescimento ativo, com altura em torno de 12 cm, albedo (fração de radiação solar refletida) igual a 0,23 e resistência ao transporte de vapor de água constante e igual a 70 sm^{-1} (segundos por metro) (ALLEN et al., 2006).

O principal método para a estimativa da ET_o , recomendado pela Organização das Nações Unidas para Alimentação e Agricultura, é o método FAO Penman-Monteith, definido pela equação 1:

$$ET_{FAO} = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta\gamma(1 + 0.34u_2)} \quad (1)$$

em que:

- ET_{FAO} : evapotranspiração de referência padrão, [$mm.dia^{-1}$];
- R_n : radiação de rede na superfície da cultura, [$MJ.m^{-2}.dia^{-1}$];
- G : densidade do fluxo de calor do solo, [$MJ.m^{-2}.dia^{-1}$];
- T : temperatura média diária do ar a 2 metros de altitude, [$^{\circ}C$];
- u_2 : velocidade do vento a 2 metros de altitude, [$m.s^{-1}$];
- e_s : pressão de saturação de vapor, [kPa];
- e_a : pressão atual de vapor, [kPa];
- $e_s - e_a$: *deficit* de pressão de saturação de vapor, [kPa];
- Δ : curva de inclinação da pressão de vapor, [$kPa.^{\circ}C^{-1}$];
- γ : constante psicométrica, [$kPa.^{\circ}C^{-1}$].

Esse método é o mais indicado por sua capacidade de determinar a ET_o de uma superfície de referência utilizando dados meteorológicos, com valores consistentes em todas as regiões e climas (ALLEN; PEREIRA; SMITH, 1998). Entretanto, requer vários parâmetros climatológicos e físicos, alguns medidos diretamente pelas estações meteorológicas, outros derivados de cálculos adicionais (ALLEN; PEREIRA; SMITH, 1998). Por esse motivo, uma série de outros métodos empíricos, baseados somente em dados de temperatura do ar (Thornthwaite, Blaney-Criddle, Camargo, Hargreaves-Samani e Thornthwaite modificado) ou apenas em dados de radiação solar (Markkink, Jensen-Haise e radiação solar), podem ser utilizados para a estimativa de ET_o na ausência de todos os dados meteorológicos (MOURA et al., 2013).

Os fatores meteorológicos que determinam a evapotranspiração são os parâmetros climáticos associados à energia necessária para a vaporização e remoção do vapor de água da superfície de evaporação. Dentre esses parâmetros climáticos, os principais são descritos a seguir.

2.1.1 Radiação solar

O processo de evapotranspiração é determinado pela quantidade de energia disponível para transformar a água líquida em vapor, sendo a radiação solar a principal fonte dessa energia. A radiação solar que atinge a superfície terrestre varia conforme a localização geográfica, a altitude e a época do ano. Devido à inclinação da Terra e à sua posição em relação ao Sol, essa radiação muda de acordo com a latitude e as estações. Além disso, fatores atmosféricos como a turbidez e a cobertura de nuvens afetam a quantidade de radiação solar que efetivamente chega ao solo, refletindo ou absorvendo parte dela (ALLEN; PEREIRA; SMITH, 1998; ALLEN et al., 2006).

Ao analisar o impacto da radiação solar sobre a evapotranspiração de referência, é importante destacar que nem toda a energia recebida é utilizada no processo de evaporação. Parte dessa radiação é direcionada para o fluxo de calor sensível, que aquece o ar, e para o fluxo de calor latente, responsável pela evaporação da água. Outra parte aquece a superfície do solo. Portanto, na relação entre radiação solar e evapotranspiração os fatores atmosféricos e de superfície

influenciam diretamente o balanço de energia e, conseqüentemente, a quantidade de água evaporada (ALLEN; PEREIRA; SMITH, 1998; ALLEN et al., 2006).

2.1.2 Temperatura do ar

A radiação solar absorvida pela atmosfera e o calor emitido pela Terra aumentam a temperatura do ar. O calor sensível do ar circundante transfere energia para a cultura e influencia a taxa de evapotranspiração. Em dias ensolarados e quentes, a perda de água por evapotranspiração é maior do que em tempo nublado e frio (ALLEN; PEREIRA; SMITH, 1998; ALLEN et al., 2006).

2.1.3 Umidade do ar

Enquanto o fornecimento de energia solar e o calor do ar circundante são os principais motores da evaporação da água, o fator determinante para a remoção do vapor d'água é a diferença na pressão de vapor entre a superfície evapotranspirante e o ar. A pressão de vapor representa a quantidade de vapor d'água presente no ar. Em regiões áridas, quentes e secas, campos bem irrigados consomem grandes quantidades de água devido à abundância de energia disponível e à baixa umidade relativa do ar, o que aumenta a capacidade da atmosfera de absorver vapor d'água, intensificando a evapotranspiração. Em contraste, em regiões tropicais úmidas, apesar da alta energia solar, a elevada umidade relativa do ar diminui a demanda por evapotranspiração, uma vez que o ar já está próximo da saturação, limitando a quantidade de vapor d'água adicional que ele pode armazenar. Portanto, a taxa de evapotranspiração em ambientes tropicais úmidos tende a ser menor do que em regiões áridas (ALLEN; PEREIRA; SMITH, 1998; ALLEN et al., 2006).

2.1.4 Velocidade do vento

A evapotranspiração depende em grande parte da turbulência gerada pelo vento, que substitui o ar saturado com vapor d'água por ar mais seco, restaurando o gradiente de pressão de vapor entre a superfície evapotranspirante e o ar circundante. Esse gradiente é o principal mecanismo que impulsiona a remoção do vapor d'água, e o vento intensifica esse processo ao remover continuamente o ar saturado. Em

condições de vento forte, a evapotranspiração é mais rápida, enquanto ventos fracos diminuem a eficiência, pois o ar saturado permanece sobre a superfície por mais tempo (ALLEN; PEREIRA; SMITH, 1998; ALLEN et al., 2006).

A influência do vento varia conforme o clima. Em regiões áridas e quentes, onde a umidade do ar é baixa e há abundância de energia, o vento facilita o transporte de grandes quantidades de vapor d'água, aumentando significativamente a taxa de evapotranspiração. Já em climas úmidos, a alta umidade relativa limita a capacidade do ar de absorver vapor adicional, o que faz com que o vento tenha um efeito menor. Nesses ambientes, o vento apenas substitui o ar levemente saturado, resultando em uma menor variação na evapotranspiração (ALLEN; PEREIRA; SMITH, 1998; ALLEN et al., 2006).

2.2 Domínios climáticos do Brasil

O sistema de classificação climática mais amplamente utilizado em geografia, climatologia e ecologia é o de Köppen-Geiger, também conhecido simplesmente como classificação de Köppen. Essa metodologia baseia-se na relação entre clima e vegetação, partindo do princípio de que os diferentes tipos de vegetação são condicionados pelo tipo climático. A classificação considera a sazonalidade e os valores médios anuais e mensais de temperatura e precipitação para definir os tipos climáticos (ARNFIELD, 2011).

O sistema de Köppen utiliza um código composto por letras maiúsculas e minúsculas, onde cada combinação define um tipo e subtipo climático. A primeira letra classifica os climas em cinco grandes grupos: *A* (tropical), *B* (seco), *C* (temperado), *D* (continental) e *E* (polar). A segunda letra indica o regime de precipitação sazonal (os meses mais secos e a quantidade de chuva), enquanto a terceira letra define o nível de calor. Por exemplo, o código “*Af*” se refere a um clima equatorial, caracterizado por chuvas abundantes, com precipitação mínima de 60 mm em todos os meses do ano. Já o código “*BSh*” identifica climas semiáridos quentes, enquanto “*Cfa*” descreve climas temperados sem estação seca, com temperaturas médias mensais superiores a 22 °C nos meses mais quentes e invernos com temperaturas que variam entre -3 °C e 18 °C (ARNFIELD, 2011).

Segundo a classificação de Köppen-Geiger, o território brasileiro pode ser dividido em três tipos climáticos principais: o clima tropical (*A*), o clima semiárido (*B*) e o clima subtropical úmido (*C*). Esses tipos apresentam variações de subtipos dependendo da região, como o clima equatorial (*Af*) na Amazônia, o clima tropical típico (*Aw*) em áreas do cerrado, e o clima subtropical (*Cfa*) predominante no sul do país (Figura 1). Essas classificações refletem as grandes diferenças regionais de temperatura e precipitação, resultando em distintos padrões de vegetação e uso da terra (ALVARES et al., 2013).

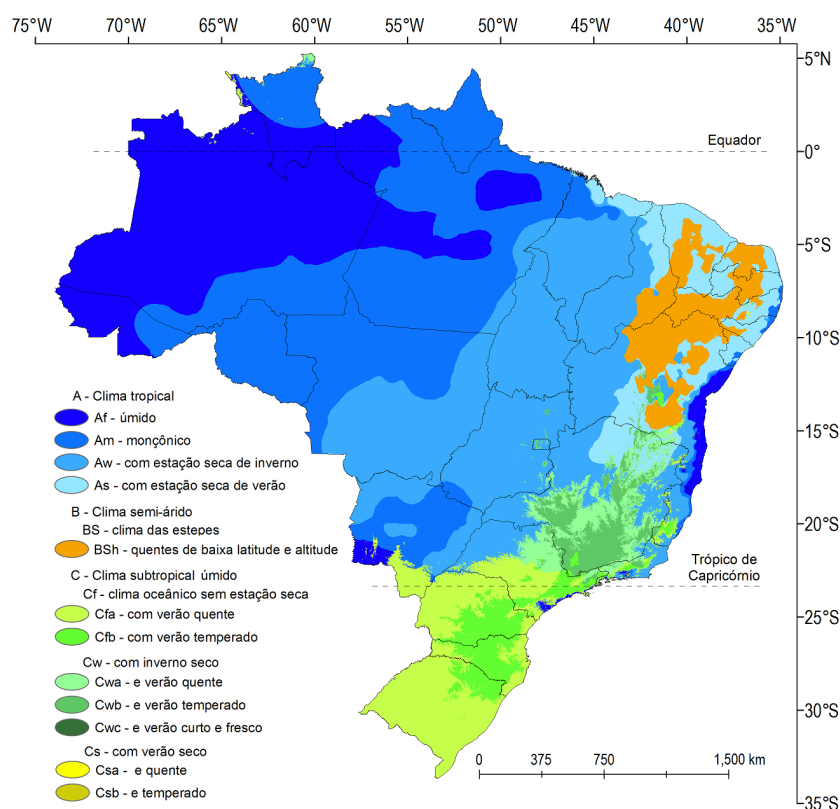


Figura 1 – Mapa climático do Brasil de acordo com a classificação de Köppen-Geiger. Adaptado de (ALVARES et al., 2013)

Os climas do Brasil também podem ser classificados de acordo com as zonas climáticas ou zonas térmicas e, portanto, podem ser agrupados com base nas características de temperatura e umidade. Neste caso, os tipos climáticos podem ter apenas cinco (Figura 2: Equatorial, Tropical Zona Equatorial, Tropical Nordeste

Oriental, Tropical Brasil Central e Temperado) ou seis classificações (Figura 3: Equatorial (EQ), Semiárido (SA), Subtropical (ST), Tropical de Altitude (TA), Tropical Litorâneo (TL) e Tropical Típico (TT)) (GOUVEIA, s.d.; IBGE, 1993).

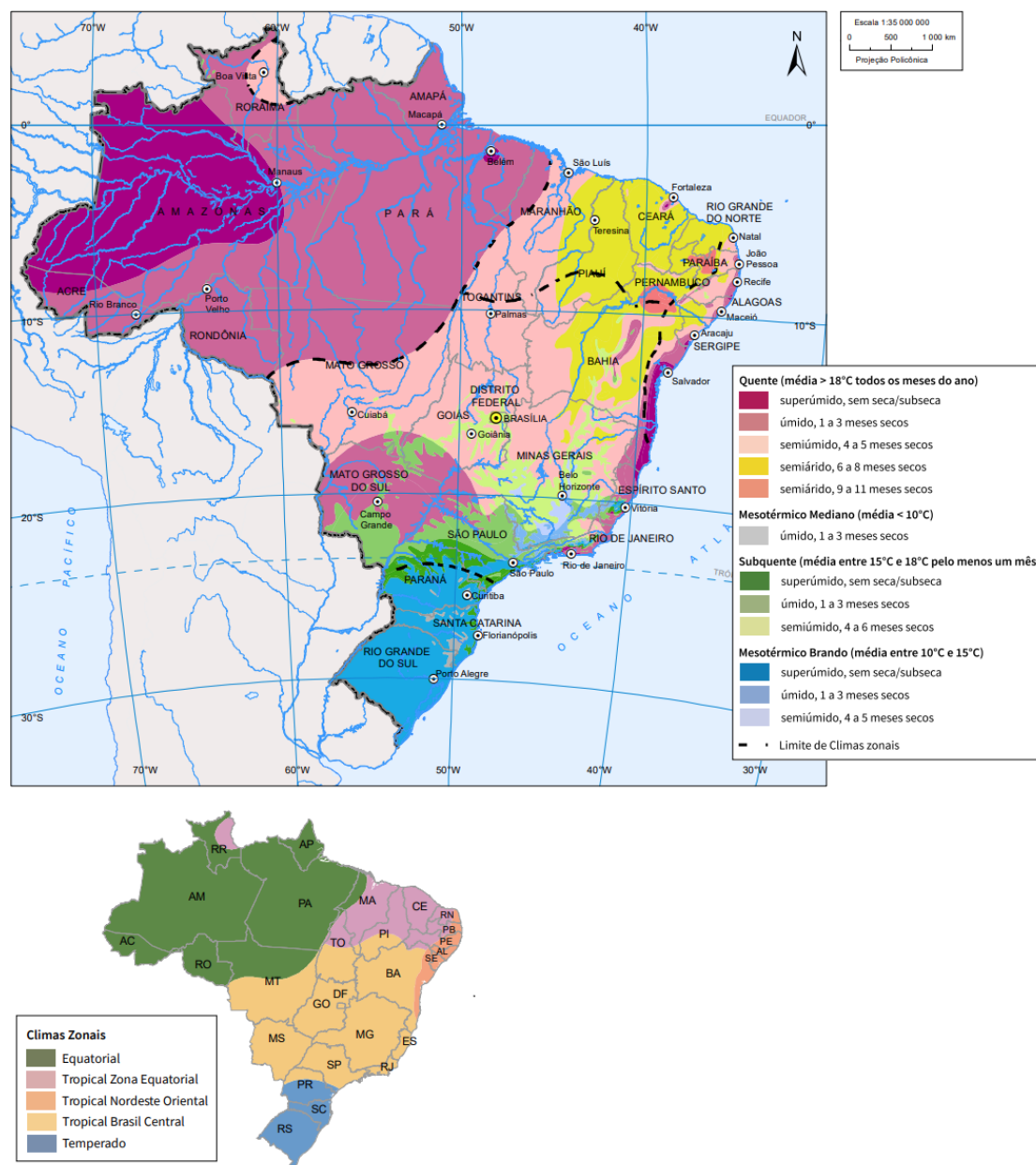


Figura 2 – Mapa das diferentes zonas climáticas agrupadas pela temperatura e umidade. Classificação por Climas Zonais. Adaptado de (IBGE, 1993)



Figura 3 – Mapa dos climas do Brasil, com seis classificações e principais correntes marítimas. Adaptado de (OLIVEIRA, 2020)

O clima **Equatorial (EQ)**, predominante na região Norte e em parte do Centro-Oeste do Brasil, é caracterizado por temperaturas elevadas durante o ano todo e altos índices de precipitação, que podem ultrapassar os 2500 mm anuais. As temperaturas médias ultrapassam os 25°C e, apesar do calor constante, eventos de friagem ocorrem no inverno em decorrência de frentes frias vindas do pólo Sul que resultam em quedas de temperatura que podem atingir os 10°C. A frequência das chuvas na região não impede a ocorrência de uma estação seca em Manaus, no estado do Amazonas (GOUVEIA, s.d.; OLIVEIRA, 2020).

Seguindo para o centro do Brasil, encontramos o clima **Tropical Típico (TT)**, também chamado de continental ou semi-úmido, que abrange a região central do país e predomina no Centro-Oeste. Esse clima é marcado por duas estações bem definidas: um verão quente e chuvoso, e um inverno seco, com temperaturas ligeiramente mais amenas. As médias anuais de temperatura ficam acima de 18°C, com uma amplitude térmica em torno de 7°C. A precipitação anual varia entre 1000 e 1800 mm, com chuvas concentradas nos meses de verão, mas em razão de massas de ar continentais, o Pantanal matogrossense pode sofrer de longas estiagens (GOUVEIA, s.d.; OLIVEIRA, 2020).

Já o clima tropical **Semiárido (SA)**, típico da região Nordeste, é caracterizado por altas temperaturas médias anuais, em torno de 27°C, com baixa amplitude térmica (cerca de 5°C) e chuvas irregulares que não ultrapassam 800 mm/ano. Nessa região, as precipitações são escassas e mal distribuídas ao longo do ano, sendo uma das áreas mais secas do país. A barreira contra as massas de ar oceânicas formada pelo planalto do Borborema, associada à célula de alta pressão existente na região, que dissipa as massas de ar úmidas e à menor temperatura das águas do Sul do Equador que provocam uma redução das chuvas nessa área, estão entre as causas da semiaridez presente nesse domínio climático (GOUVEIA, s.d.; OLIVEIRA, 2020).

Nas regiões mais elevadas do Sudeste e em outras regiões de planaltos e serras, o clima **Tropical de Altitude (TA)** predomina. Esse clima é influenciado pelas maiores altitudes, resultando em temperaturas mais amenas, com médias anuais abaixo de 18°C, e amplitude térmica entre 7°C e 9°C. O regime de chuvas é semelhante ao do clima tropical, mas podem ocorrer as chamadas chuvas de relevo. E a ocorrência de frentes frias no inverno pode levar à formação de geadas,

especialmente nas áreas de maior altitude (GOUVEIA, s.d.; OLIVEIRA, 2020).

Ao longo da costa brasileira, o clima **Tropical Litorâneo (TL)**, também conhecido como tropical úmido, se estende do Rio Grande do Norte até o Rio de Janeiro. Esse clima é fortemente influenciado pela proximidade do oceano, o que garante temperaturas médias anuais entre 18°C e 26°C e altos índices de precipitação, em torno de 1500 mm/ano, podendo alcançar 4000 mm em trechos da serra do Mar no litoral paulista. A distribuição das chuvas varia conforme a região: no Nordeste, as precipitações são mais frequentes no outono e inverno, enquanto no Sudeste, o verão é a estação mais chuvosa (GOUVEIA, s.d.; OLIVEIRA, 2020).

Finalmente, na região Sul do país, abaixo do Trópico de Capricórnio, o clima **Subtropical (ST)** prevalece. Com suas estações do ano bem demarcadas, os verões são quentes e os invernos podem ser rigorosos, com a ocorrência de geadas e, em algumas áreas de maior altitude, até mesmo neve. As temperaturas no inverno frequentemente caem abaixo de 0°C, mas as médias anuais ficam por volta dos 18°C. O regime de chuvas é bem distribuído ao longo do ano, com totais pluviométricos superiores a 1250 mm anuais, o que diferencia esse clima dos demais no Brasil (GOUVEIA, s.d.; OLIVEIRA, 2020).

2.3 Modelos Multivariados

Nesta seção, são apresentados os modelos multivariados selecionados para avaliar a previsão da evapotranspiração de referência em diferentes climas do Brasil. A escolha dos modelos baseou-se em sua capacidade de capturar padrões complexos em dados temporais e em sua aplicabilidade comprovada em estudos similares, dentre os trabalhos relacionados apresentados na Seção 2.4. Os quatro modelos escolhidos representam uma diversidade de abordagens, desde métodos tradicionais até técnicas modernas baseadas em aprendizado profundo como *Transformers*, de modo a comparar sua eficácia em diferentes condições climáticas.

2.3.1 Modelagem de Múltiplas Séries Temporais

Na previsão univariada, cada série temporal é analisada individualmente, sendo modelada como uma combinação de suas próprias defasagens (também

chamadas de atrasos ou *lags*) e, opcionalmente, de variáveis exógenas. Por outro lado, a modelagem multivariada envolve a construção de um modelo preditivo unificado que considera todas as séries temporais simultaneamente (RODRIGO; ORTIZ, 2022). Essa abordagem busca capturar os padrões principais que regem as séries, reduzindo o impacto de possíveis ruídos individuais, podendo gerar generalizações mais robustas entre as séries, embora, em alguns casos, possa sacrificar algum conhecimento específico de cada uma (RODRIGO; ORTIZ, 2023).

2.3.2 Modelo Vetorial Autorregressivo

O modelo vetorial autorregressivo ou *vector autoregressive* (VAR) é uma técnica clássica amplamente utilizada para a previsão de séries temporais multivariadas, especialmente quando as variáveis têm dependências temporais entre si. Ao contrário de modelos univariados, como o ARIMA, o VAR consegue capturar as interações dinâmicas entre múltiplas variáveis ao longo do tempo (LÜTKEPOHL, 2005). Essa capacidade o torna adequado para problemas onde múltiplas variáveis influenciam-se mutuamente, como ocorre com variáveis meteorológicas (SHAHIN; ALI; ALI, 2014; ABDALLAH et al., 2020).

O VAR assume que cada variável na série temporal pode ser explicada por uma combinação linear das suas próprias defasagens e das defasagens das outras variáveis (LÜTKEPOHL, 2005). Assim, para k variáveis e p defasagens, cada variável é modelada como:

$$y_{k,t} = A_{k,0} + A_{k,1}y_{k,t-1} + A_{k,2}y_{k,t-2} + \dots + A_{k,p}y_{k,t-p} + \epsilon_{k,t} \quad (2)$$

Onde:

- y_t é um vetor contendo os valores de todas as variáveis no tempo t ;
- $A_0, A_1, A_2, \dots, A_p$ são matrizes de coeficientes que capturam a influência de cada defasagem nas variáveis atuais;
- ϵ_t é um vetor de termos de erro (resíduos), assumidos como ruído branco multivariado.

Um modelo VAR com defasagem de uma unidade de tempo ($p = 1$) e com 2 variáveis ($k = 2$), por exemplo, seria representado por duas equações: $y_{1,t} = A_{1,0} + A_{1,1} * y_{1,t-1} + A_{1,2} * y_{2,t-1} + \epsilon_{1,t}$ e $y_{2,t} = A_{2,0} + A_{2,1} * y_{1,t-1} + A_{2,2} * y_{2,t-1} + \epsilon_{2,t}$.

Em relação ao aprendizado, essa etapa busca estimar as matrizes de coeficientes $A_0, A_1, A_2, \dots, A_p$ por meio de métodos de regressão linear multivariada como OLS (*Ordinary Least Squares*), minimizando a soma dos erros quadráticos. Durante o treinamento, o modelo ajusta esses coeficientes para capturar as relações entre as variáveis ao longo do tempo (LütKEPOHL, 2005; HANCK et al., 2024).

Uma das características importantes do VAR é que ele assume estacionariedade, ou seja, que as relações entre as variáveis permaneçam estáveis ao longo do tempo. Quando as variáveis não são estacionárias, é necessário aplicar diferenciações para estabilizar a média. Para a previsão, o modelo usa valores observados e previstos em defasagens anteriores. Por exemplo, ao prever o valor do tempo $t + 1$, o VAR se baseia nos valores de $t, t - 1, \dots, t - p$. A natureza recursiva do modelo permite que ele faça previsões de longo prazo, ainda que a precisão tenda a diminuir à medida que o horizonte de previsão aumenta (LütKEPOHL, 2005; HANCK et al., 2024).

No contexto deste trabalho, o VAR foi utilizado para prever a *ETo* com base em variáveis climáticas como temperatura e umidade relativa do ar, radiação solar e velocidade do vento. A capacidade do modelo de capturar a influência simultânea dessas variáveis é particularmente relevante, dado que essas condições meteorológicas se influenciam mutuamente e, juntas, determinam a evapotranspiração. Embora o VAR seja um modelo linear, sua simplicidade é uma vantagem por fornecer previsões interpretáveis e servir de *benchmark* para avaliar o desempenho de modelos mais complexos, como os baseados em redes neurais e *Transformers*.

2.3.3 Modelo Floresta Aleatória

O modelo floresta aleatória ou *random forest* (RF) é um algoritmo de aprendizado de máquina baseado em árvores de decisão, muito utilizado para problemas de classificação e regressão. Ele é uma técnica de *ensemble*, ou seja, combina múltiplas árvores de decisão para produzir previsões mais robustas e precisas. A força do modelo vem de sua capacidade de reduzir o risco de *overfitting*

(que corresponde a um ajuste excessivo), ao usar um grande número de árvores, cada uma treinada com subconjuntos aleatórios dos dados e das variáveis preditoras (*features*) (LOUPPE, 2014).

O modelo floresta aleatória constrói várias árvores de decisão de forma independente e calcula a média (ou o voto majoritário, em casos de classificação) das previsões de todas as árvores para obter o resultado final. Cada árvore é treinada com uma amostra aleatória do conjunto de dados (técnica conhecida como *bootstrap*) e, em cada divisão da árvore, considera-se uma seleção aleatória das variáveis preditoras. Isso aumenta a diversidade entre as árvores, melhorando a capacidade de generalização do modelo (LOUPPE, 2014).

Matematicamente, de forma simplificada, a previsão para uma variável-alvo (*target*) y é dada por:

$$\hat{y} = \frac{1}{N} \sum_{t=1}^N f_t(x) \quad (3)$$

Onde:

- N é o número total de árvores, e
- $f_t(x)$ é a previsão da t -ésima árvore para a amostra x .

Durante o treinamento, o algoritmo RF constrói cada árvore até atingir uma profundidade máxima ou até que as folhas contenham um número mínimo de amostras. O uso de dados amostrados por *bootstrap* e a seleção de variáveis reduzem a correlação entre árvores, melhorando a robustez do modelo. Além disso, como o RF não faz suposições sobre a distribuição dos dados, ele é capaz de capturar tanto relações lineares quanto não lineares entre as variáveis (LOUPPE, 2014). Isso o torna particularmente útil para dados complexos e heterogêneos, como séries temporais climáticas.

Embora o RF não seja originalmente um modelo de séries temporais, ele pode ser adaptado para previsões temporais multivariadas ao utilizar as defasagens das variáveis como novas *features* (BROWNLEE, 2020). Para cada instante t , o modelo utiliza valores passados das variáveis de entrada, como temperatura, umidade e radiação solar, para prever a *ETo* no tempo $t + 1$ (FERREIRA; CUNHA,

2020; FERREIRA; CUNHA; FILHO, 2022; SHALOO et al., 2024). Essa abordagem permite capturar padrões históricos nas variáveis climáticas e prever seus valores futuros de forma precisa, mesmo quando as relações entre as variáveis são complexas e não lineares.

No contexto deste trabalho, o modelo RF foi utilizado para prever a *ETo* com base em variáveis meteorológicas. A capacidade do modelo de capturar interações complexas entre as variáveis é uma vantagem importante, uma vez que a *ETo* é afetada por múltiplas variáveis interdependentes. Além disso, como o modelo é resistente a *overfitting* e funciona bem com um grande número de variáveis, ele serve como uma boa alternativa para a previsão de séries temporais meteorológicas.

2.3.4 Redes Neurais Convolucionais para Séries Temporais

As redes neurais convolucionais para séries temporais ou *temporal convolutional neural networks* (TCNs) são uma variação das redes neurais convolucionais, especialmente adaptadas para lidar com dados sequenciais e temporais. Diferentemente das CNNs tradicionais, que são amplamente usadas em processamento de imagens, as TCNs operam ao longo da dimensão do tempo para identificar padrões temporais complexos em dados multivariados (BAI; KOLTER; KOLTUN, 2018; GONÇALVES, 2021). Elas são utilizadas em várias aplicações, como previsão financeira, meteorologia e outras séries temporais multivariadas, devido à sua capacidade de modelar dependências de curto e longo prazo.

A TCN é composta por camadas convolucionais unidimensionais (1D) que aplicam filtros ao longo da dimensão temporal (BAI; KOLTER; KOLTUN, 2018). Dois conceitos fundamentais garantem seu desempenho:

- **Convolução causal:** Garante que o valor previsto para um momento t não dependa de informações futuras ($t + 1, t + 2, \dots$). Isso é essencial para que as previsões respeitem a sequência temporal dos dados;
- **Convolução dilatada:** Amplia o campo de visão do modelo sem aumentar a profundidade da rede. Ao aplicar um espaçamento entre os dados analisados, a convolução dilatada permite capturar padrões temporais mais longos sem precisar empilhar muitas camadas convolucionais.

A arquitetura da TCN também pode incorporar conexões residuais, inspiradas nas redes *residual net* (ResNet), para facilitar a propagação de gradientes. Essas conexões criam atalhos que permitem que a entrada de uma camada seja diretamente somada à sua saída, proporcionando caminhos adicionais para a propagação de informações. Essa abordagem não apenas ajuda a evitar o desaparecimento de gradientes durante o treinamento, mas também melhora a estabilidade do aprendizado e a eficiência da otimização (ZENG, 2019).

O treinamento de uma TCN envolve o ajuste dos filtros convolucionais para minimizar uma função de perda, como o RMSE. Ao contrário das arquiteturas recorrentes, a TCN possui um caminho de retropropagação diferente da direção temporal da sequência, evitando a utilização do método de *backpropagation through time* (BPTT), que é uma grande limitação das *recurrent neural networks* (RNNs) (BAI; KOLTER; KOLTUN, 2018).

Em séries multivariadas, como as usadas na previsão de *ETo*, cada variável (temperatura do ar, umidade relativa do ar, radiação solar e velocidade do vento) é tratada como um canal separado na entrada da rede. Isso permite que a TCN aprenda não apenas padrões individuais de cada variável, mas também interações entre elas (GONÇALVES, 2021).

Uma vez treinada, a TCN utiliza os dados históricos para prever os valores futuros das variáveis-alvo. Graças às convoluções causais e dilatadas, o modelo consegue fazer previsões de curto, médio e longo prazo com eficiência computacional. Além disso, a estrutura convolucional permite processar dados em paralelo, acelerando o tempo de inferência em comparação com modelos recorrentes, como *long short-term memory* (LSTM) e *gated recurrent unit* (GRU) (BAI; KOLTER; KOLTUN, 2018).

Neste trabalho, a TCN é usada para prever a evapotranspiração de referência com base em variáveis climáticas, como temperatura e umidade relativa do ar, radiação solar, precipitação e velocidade do vento. A capacidade da TCN de capturar padrões complexos e interações entre as variáveis é interessante para esse contexto, dado que a *ETo* depende simultaneamente de múltiplos fatores climáticos.

2.3.5 Modelos Baseados em *Transformer*

Os *Transformers*, propostos pela primeira vez em 2017 (VASWANI et al., 2023), são uma classe de modelos originalmente introduzida para tarefas de processamento de linguagem natural (NLP), mas que têm sido cada vez mais utilizados em problemas de previsão de séries temporais multivariadas, incluindo a previsão de *ETo* (BELLIDO-JIMÉNEZ et al., 2022; AHMADI et al., 2023).

A principal inovação dos *Transformers* é o uso do **mecanismo de atenção** (*attention mechanism*), que permite ao modelo atribuir diferentes pesos para diferentes momentos ou variáveis na entrada, dependendo da sua relevância para a previsão atual. Diferentemente de modelos recorrentes como LSTM e GRU, os *Transformers* não dependem de uma propagação sequencial de estados ao longo do tempo, o que melhora a eficiência computacional e permite paralelização durante o treinamento. Além disso, os *Transformers* dispensam completamente as recorrências e as convoluções (VASWANI et al., 2023).

De acordo com os mesmos autores, cada *Transformer* é composto por camadas dos seguintes tipos:

- **Atenção Multi-Cabeças** (*Multi-Head Attention*): O mecanismo de atenção multi-cabeça nos *Transformers* permite que diferentes “cabeças” atendam a diferentes aspectos dos dados de entrada. Cada cabeça pode focar em diferentes padrões ou interações entre as variáveis. Por exemplo, em uma previsão de *ETo*, uma cabeça poderia se concentrar na relação entre a radiação solar e a temperatura, enquanto outra poderia aprender as interações entre umidade e vento. Isso permite ao modelo capturar interações complexas e dependências entre variáveis de maneira mais eficaz do que modelos sequenciais como RNNs.
- **Feedforward denso**: Após a aplicação do mecanismo de atenção, onde o modelo aprende quais partes da sequência de entrada são mais relevantes em cada momento, as camadas densas são aplicadas a cada passo de tempo para processar e combinar essas informações, capturando padrões locais não lineares.

- **Normalização e *Residual Connections*:** Como o modelo trabalha com múltiplas variáveis meteorológicas ao longo do tempo, a normalização ajuda a estabilizar a aprendizagem das representações complexas dessas variáveis. O uso de conexões residuais em *Transformers* tem papel similar ao seu uso numa CNN: ajudar o modelo a aprender representações mais profundas sem perder a estabilidade.

O treinamento dos *Transformers* para previsão de séries temporais envolve minimizar uma função de perda (como RMSE ou MAE) por meio de **otimização por descida de gradiente**, utilizando algoritmos como o Adam. Como os dados são sequenciais, um pré-processamento é necessário para organizar as séries temporais em janelas deslizantes. Além disso, como a arquitetura original dos *Transformers* não possui uma noção explícita de tempo, é comum a adição de **codificações posicionais** (*positional encodings*), que indicam a ordem temporal dos dados (WU et al., 2020).

Os *Transformers* exigem uma grande quantidade de dados para treinar de forma eficaz, mas podem capturar dependências tanto de curto (LIM et al., 2020; ZERVEAS et al., 2020) quanto de longo (WU et al., 2020; ZHOU et al., 2021) prazos, algo importante no contexto de previsão da evapotranspiração de referência.

Neste trabalho, um modelo baseado em *Transformers* é aplicado para prever a evapotranspiração de referência com base em variáveis meteorológicas interdependentes. Sua capacidade de processar múltiplas variáveis simultaneamente, e de capturar padrões complexos nas interações entre elas, é particularmente relevante em previsões multivariadas de *ET_o*.

2.3.6 Importância das Variáveis de Entrada

As variáveis de entrada, também conhecidas como recursos (*features*) ou preditores, desempenham papel importante na precisão dos modelos de previsão de séries temporais, portanto, explorar sua importância contribui para a melhora no desempenho e a eficiência dos modelos.

As variáveis podem ser classificadas em três principais categorias:

- **Numéricas:** recursos que representam dados quantitativos, expressos como valores numéricos (inteiros ou decimais), como temperatura ($^{\circ}\text{C}$), peso (kg) e idade (anos).
- **Catagóricas:** recursos que representam dados qualitativos, significando a categoria à qual um ponto de dados pertence, como cor do cabelo (loiro, moreno, preto) e satisfação do cliente (satisfeito, neutro, insatisfeito).
- **Ordinais:** recursos que são um subtipo dos recursos catagóricos, possuindo uma ordem ou classificação inerente, como classificações de filmes (1 estrela, 2 estrelas, etc.) e experiência de atendimento ao cliente (ruim, média, excelente).

Entender a importância das variáveis de entrada oferece desempenho aprimorado ao modelo, uma vez que, ao identificarem-se as variáveis mais influentes, pode-se priorizá-las durante o treinamento do modelo, levando a previsões mais precisas. Além disso, consegue-se tempos de treinamento mais rápidos, já que focar nas variáveis mais relevantes agiliza o processo de treinamento, economizando tempo e recursos computacionais. Outra vantagem desse entendimento encontra-se na diminuição do risco de *overfitting*, que pode ocorrer quando um modelo memoriza os dados de treinamento em vez de aprender padrões gerais. Ao focar nas variáveis mais importantes, evita-se que o modelo se torne excessivamente dependente de pontos de dados específicos (COLLEGE, 2024).

Várias técnicas podem ser empregadas para calcular a importância das variáveis de entrada. Modelos interpretáveis como o *Random Forest*, disponibilizam medidas de importância das variáveis, sendo as mais largamente utilizadas a *Impurity Importance* e a *Permutation Importance*. A *Impurity Importance*, também conhecida como média de diminuição de impureza ou *mean decrease impurity* (MDI), comumente está associada ao índice de Gini, que é o critério de divisão dos nós nas árvores de classificação e regressão, por isso é frequentemente chamada de *Gini Importance* ou importância de Gini (LOUPPE, 2015; NEMBRINI; KÖNIG; WRIGHT, 2018).

A importância geral de uma variável de entrada em uma árvore de decisão, ou a *Gini Importance* pode ser computada da seguinte maneira: passa-se por todas as divisões para as quais o recurso foi usado e mede-se o quanto ele reduziu a variância

ou o índice de Gini em comparação ao nó pai. A soma de todas as importâncias é dimensionada para 100% (ou 1.0). Isso significa que cada importância pode ser interpretada como uma parcela da importância geral do modelo (LOUPPE, 2015; MOLNAR, 2024).

A *Permutation Importance* mede o aumento no erro de previsão do modelo após permutarmos os valores das variáveis de entrada, o que quebra a relação entre o variável e o resultado verdadeiro. Uma variável preditora é “importante” se, ao embaralharem-se seus valores aumenta-se o erro do modelo, porque neste caso, o modelo “confiou” na variável para a previsão. Uma variável de entrada é “sem importância” se, ao embaralharem-se seus valores, mantém-se o erro do modelo inalterado, uma vez que, neste caso, o modelo ignorou a variável durante o processo de previsão (MOLNAR, 2024).

No modelo VAR, as *impulse response functions* (IRF) podem ser usadas para medir o impacto dinâmico de um choque em uma variável sobre ela mesma e sobre as demais variáveis do sistema ao longo do tempo. Em outras palavras, as IRF’s mostram como um sistema VAR responde a uma perturbação inesperada em uma variável e as respostas ao impulso traçam a resposta dos valores atuais e futuros de cada uma das variáveis a um aumento de uma unidade no valor atual de um dos erros do VAR, assumindo que esse erro retorna a zero em períodos subsequentes e que todos os outros erros são iguais a zero. Esse mecanismo contribui para o entendimento de quais variáveis influenciam mais fortemente as outras no sistema. (LÜTKEPOHL, 2005).

Com modelos “caixa-preta”, como as redes neurais, também é possível utilizar o método *Permutation Feature Importance*, no entanto, outra abordagem popular utilizada para medir a importância das variáveis baseia-se na teoria dos jogos. SHAP (*SHapley Additive exPlanations*) de (LUNDBERG; ERION; LEE, 2019) é um método para explicar previsões individuais baseado em *Shapley values*. O objetivo do SHAP é explicar a previsão de uma instância x calculando a contribuição de cada variável de entrada para essa previsão.

O método calcula valores de Shapley da teoria dos jogos cooperativos, os valores das variáveis de uma instância agem como jogadores em uma coalizão e os valores de Shapley dizem como distribuir de forma justa o “pagamento” entre as *features*. Um jogador pode ser um valor de *feature* individual, por exemplo, para

dados tabulares ou pode ser um grupo de valores de *features*. Por exemplo, para explicar uma imagem, os pixels podem ser agrupados em superpixels e a previsão distribuída entre eles. Uma inovação que o SHAP traz é que a explicação do valor de Shapley é representada como um método de atribuição de característica aditiva, cujo modelo é linear (MOLNAR, 2024).

Portanto, para cada um dos modelos de previsão multivariada, pode-se utilizar um ou mais dos métodos mencionados para medir a importância das variáveis meteorológicas e melhorar a precisão e desempenho dos modelos na previsão de *ET_o*. Neste trabalho, foram utilizados o *Impulse Response Function* para o modelo VAR, o *Gini Importance* para o modelo *Random Forest*, e o SHAP para os modelos baseados em CNN e *Transformers*.

2.4 Trabalhos Relacionados

A previsão da evapotranspiração de referência é um tema amplamente estudado, com uma variedade de abordagens exploradas para melhorar a precisão das previsões em diferentes contextos climáticos. Entre as técnicas mais promissoras para esse fim, destacam-se as de inteligência artificial e aprendizado de máquina, incluindo redes neurais, máquinas de vetores de suporte (SVMs), sistemas fuzzy; modelos baseados em árvores de decisão, como o *Random Forest*, o Gradient Boosting Machine (GBM) e o XGBoost; e métodos de aprendizado profundo, como redes convolucionais (CNN), redes recorrentes como LSTM e *Transformers*. Essas técnicas têm mostrado resultados variados em estudos que utilizam diferentes configurações de variáveis meteorológicas e contextos ambientais.

A influência de variáveis climáticas na evapotranspiração foi avaliada em áreas de cerrado e cana-de-açúcar. Usando análise de componentes principais (PCA) e técnicas de agrupamento, os autores identificaram que a evapotranspiração é fortemente impactada por fatores como disponibilidade de energia e condições atmosféricas e da vegetação. As variáveis incluíram temperatura, umidade relativa, pressão atmosférica, velocidade do vento e irradiância solar, com três componentes principais explicando entre 75% e 82% da variância dos dados (RUHOFF et al., 2009).

No campo das comparações entre métodos empíricos de estimativa de ET_o , o desempenho do método de Penman-Monteith, na ausência de dados climáticos essenciais como a radiação solar, em comparação com o método de Hargreaves-Samani, que depende apenas de dados de temperatura, foi avaliado em um estudo com 20 estações meteorológicas de Minas Gerais, e o método de Penman-Monteith apresentou melhor desempenho, mesmo sem a totalidade das variáveis (ALENCAR; SEDIYAMA; MANTOVANI, 2015).

Redes neurais convolucionais temporais para prever a ET_o em Sete Lagoas, Minas Gerais, foram exploradas usando dados históricos de 1993 a 2015. Modelos *ensemble* de CNN superaram modelos individuais e um modelo Naïve, especialmente para horizontes curtos (LUCAS, 2019). Em outra direção, modelos de autorregressão vetorial foram aplicados para previsão de curto prazo de parâmetros climáticos, mostrando 97% de precisão em variáveis como temperatura mínima e umidade relativa máxima (ABDALLAH et al., 2020).

Outras pesquisas abordaram o uso de aprendizado profundo e aprendizado de máquina em contextos com dados limitados ou incompletos. Por exemplo, o uso dos modelos ANFIS, MARS, *Random Forest* e XGBoost para estimar ET_o usando dados diários e horários mostraram que esses modelos superaram as equações convencionais de Penman-Monteith e Hargreaves-Samani em condições de dados incompletos (FERREIRA; CUNHA, 2020). Em continuidade a este trabalho, a generalização de modelos como ANN e MARS em regiões semiáridas foi investigada, concluindo que a calibração local é fundamental tanto para modelos de aprendizado de máquina quanto para equações empíricas (FERREIRA et al., 2021).

O modelo XGBoost também foi utilizado para a previsão de ET_o em 3 localidades da Região das Matas de Minas, em Minas Gerais, avaliando o modelo em dois cenários: utilizando todas as variáveis meteorológicas; e utilizando somente dados de temperatura. O autor viu potencial no modelo, mas notou a ocorrência de superestimativas nos valores de ET_o nos dois cenários (ROSA, 2022).

Alguns estudos exploraram abordagens híbridas e *fuzzy* para previsão de ET_o . Sistemas *fuzzy* e árvores de decisão foram combinadas para previsão multivariada, com aplicações que incluíram o regime pluvial do rio Xingu, evidenciando a interpretabilidade dos modelos baseados em regras em relação a métodos como redes neurais (SILVA, 2021). No Brasil, séries temporais nebulosas ou *fuzzy time*

series (FTS) foram utilizadas para previsões de *ETo*, mostrando que o modelo FTS multivariado ponderado superou abordagens tradicionais em previsões pontuais, embora tenha tido desempenho inferior em previsões intervalares (LUCAS et al., 2022).

Em outro contexto geográfico, arquiteturas híbridas de *Transformers*-CNN e *Transformers*-LSTM foram implementadas para prever *ETo* diária na Andaluzia, Espanha, mostrando que, apesar de a precisão dos modelos diminuir para horizontes maiores que três dias, eles superaram significativamente modelos tradicionais, como RF, SVM e ELM, em condições variadas de aridez e proximidade do mar (BELLIDO-JIMÉNEZ et al., 2022). Na Austrália, propôs-se o modelo MVMD-BRT, que combinou decomposição de modo variacional com árvores de regressão, superando modelos comparáveis como RF e CFNN na previsão diária de *ETo* (ALI et al., 2023).

Estudos comparativos entre abordagens também são comuns. Modelos baseados em árvores de decisão (*Decision Tree*, RF, e XGBoost) foram comparados com os resultados obtidos com modelos CNN, avaliando previsões de *ETo* em diferentes regiões do Brasil (BASÍLIO, 2023). Em outro trabalho, seis modelos híbridos de aprendizado de máquina foram avaliados para estimativa de *ETo* na China, e o modelo DNN-RF teve o melhor desempenho nas métricas CC, MAE, RMSE, RAE e RRSE (ELBELTAGI et al., 2023).

Estudos asiáticos têm investigado diferentes variáveis para melhorar as previsões de *ETo*. Na China, o GBM foi identificado como o modelo mais eficaz ao incorporar variáveis meteorológicas para estimativas de *ETo* com base na equação de Penman-Monteith (GADDIKERI et al., 2023). Em outro estudo asiático, o modelo VAR foi utilizado para investigar a relação causal entre variáveis meteorológicas e *ETo* em Penang, Malásia, mostrando a importância do período de dados históricos para a precisão (HOU et al., 2023).

Com base na revisão desses estudos, percebe-se que diversas abordagens têm sido exploradas para a previsão da evapotranspiração de referência em diferentes contextos climáticos, especialmente no uso de técnicas de aprendizado de máquina e modelos híbridos. Contudo, a maioria dos trabalhos foca em contextos geográficos específicos e em horizontes temporais variados, com resultados que muitas vezes não são diretamente comparáveis.

Neste cenário, o presente trabalho propõe uma avaliação do desempenho de modelos de previsão multivariada de ET_o nos diferentes climas brasileiros, uma lacuna importante considerando a diversidade climática do país. Espera-se que os resultados obtidos contribuam tanto para o avanço no entendimento do comportamento da ET_o em regiões com diferentes condições climáticas, quanto para a aplicação prática na gestão de recursos hídricos e na agricultura de precisão.

3 Método do trabalho

Neste capítulo, são apresentados os critérios de seleção das bases de dados meteorológicos, detalhes de implementação dos modelos de previsão multivariada e critérios de avaliação dos resultados. Também encontram-se descritos detalhes dos experimentos realizados.

As etapas para o desenvolvimento deste estudo consistiram em:

1. Caracterizar os climas do Brasil, de acordo com as classificações apresentadas na Seção 2.2;
2. Selecionar as bases de dados representativas de cada clima, cujos critérios são descritos na próxima seção;
3. Selecionar os modelos multivariados para a previsão da evapotranspiração de referência, conforme descrito na Seção 2.3;
4. Definir diferentes cenários para experimentação, com base na análise de importância das variáveis, cujas ferramentas utilizadas foram apresentadas na Seção 2.3.6;
5. Realizar experimentos nos cenários definidos, avaliando os modelos com base nos erros de previsão;
6. E realizar uma análise e discussão dos resultados.

3.1 Seleção das Bases de Dados

Como mencionado no Capítulo 1, a escassez de dados meteorológicos em várias regiões do Brasil levou à necessidade de utilizar técnicas de interpolação para gerar estimativas climáticas onde não é possível realizar medições diretas. Neste trabalho, as bases de dados selecionadas foram extraídas de um repositório de dados cujo estudo utilizou interpolação espacial para gerar estimativas climáticas em regiões com baixa densidade de estações meteorológicas (XAVIER et al., 2022).

O objetivo deste trabalho foi avaliar modelos de previsão multivariada em diferentes climas brasileiros. Para isso, foram selecionadas seis bases de dados meteorológicos representando as principais regiões climáticas do Brasil, incluindo o Cerrado, a Amazônia, o Semiárido e as regiões Sul e Sudeste, que possuem condições climáticas distintas. As bases foram escolhidas de modo a capturar a variabilidade climática do Brasil, considerando a diversidade de temperatura, umidade e precipitação, essenciais para a avaliação dos modelos. A confiabilidade dessas estimativas climáticas, geradas por meio de interpolação, foi validada, garantindo que os dados são adequados para o uso nos experimentos aqui propostos (XAVIER et al., 2022).

As séries históricas selecionadas compreendem o período entre 2000 e 2019, constituindo 20 anos de dados diários, referentes às seguintes variáveis meteorológicas:

- R_s : radiação solar, [$MJ.m^{-2}.dia^{-1}$];
- u_2 : velocidade do vento a 2 metros de altitude, [$m.s^{-1}$];
- T_{max} : temperatura máxima diária do ar, [$^{\circ}C$];
- T_{min} : temperatura mínima diária do ar, [$^{\circ}C$];
- RH : umidade relativa do ar, [%];
- pr : precipitação, [$mm.dia^{-1}$];
- ET_o : evapotranspiração de referência, [$mm.dia^{-1}$];

As bases de dados meteorológicos foram selecionadas levando em consideração a aproximação geográfica em relação a cidades-referência representantes dos diferentes domínios climáticos do Brasil: Manaus/AM, Quixeramobim/CE, Recife/PB, Cuiabá/MT, Belo Horizonte/MG e Alegrete/RS (OLIVEIRA, 2020).

Manaus, localizada no estado do Amazonas, a -3,09 graus de latitude e -60,06 graus de longitude, foi tomada como referência para o clima Equatorial devido à sua localização central na floresta amazônica e à disponibilidade de dados meteorológicos para a região. A base de dados selecionada para esse clima tem coordenadas -2,15 de latitude e -59,85 de longitude (Figura 4).

Pontos para ETo's de Interesse - Amazonas

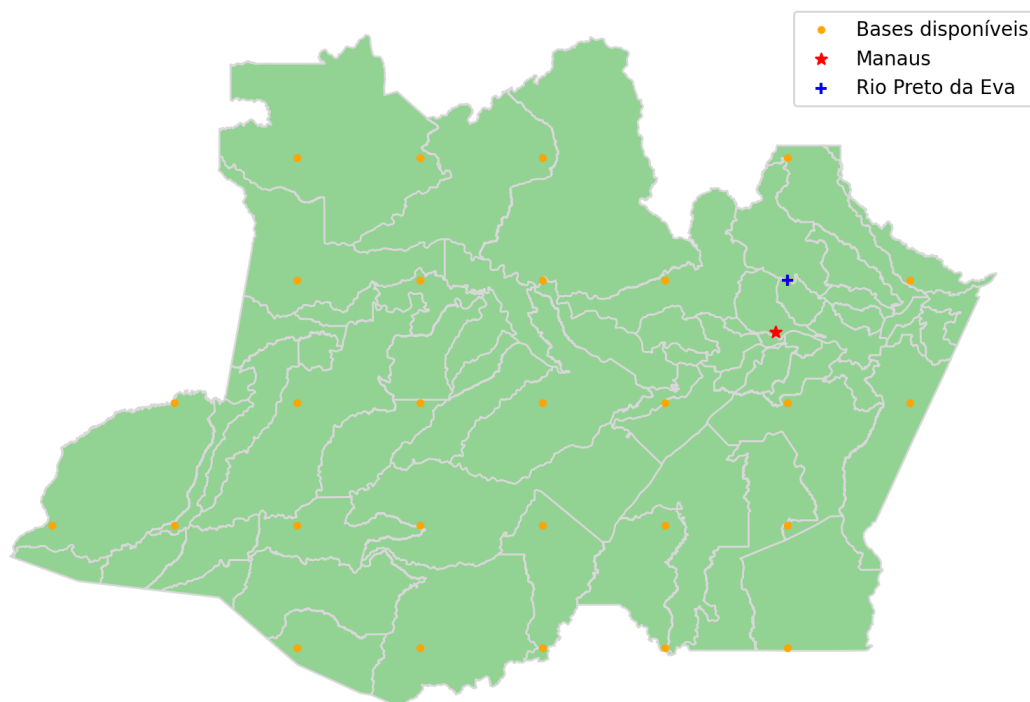


Figura 4 – Base de dados selecionada para o clima Equatorial (EQ).

Quixeramobim, no estado do Ceará, com coordenadas de -5,2 graus de latitude e -39,3 graus de longitude, foi escolhida para representar o clima Semiárido, sendo uma cidade situada no interior do Nordeste, área caracterizada pela baixa pluviosidade e alta variabilidade climática. A base mais próxima geograficamente tem coordenadas -4,35 de latitude e -40,05 graus de longitude (Figura 5).

A cidade de Recife, em Pernambuco, localizada a -8,06 graus de latitude e -34,87 graus de longitude, foi escolhida para representar o clima Tropical Litorâneo, caracterizado por sua posição litorânea e pela influência de ventos marítimos. A base selecionada para representar o clima Tropical Litorâneo está a -8,75 graus de latitude e -35,65 graus de longitude (Figura 6).

Belo Horizonte (Latitude: -19,92, Longitude: -43,94), em Minas Gerais, foi usada como referência para o clima Tropical de Altitude, devido à sua localização elevada e às características climáticas distintas da região Sudeste. A base de dados localizada em Pará de Minas (Latitude: -19,75, Longitude: -44,45) é a mais próxima

Pontos para ETo's de Interesse - Ceará

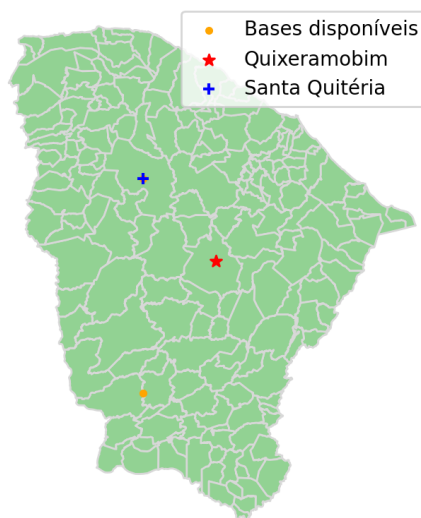


Figura 5 – Base de dados selecionada para o clima Semiárido (SA).

Pontos para ETo's de Interesse - Pernambuco

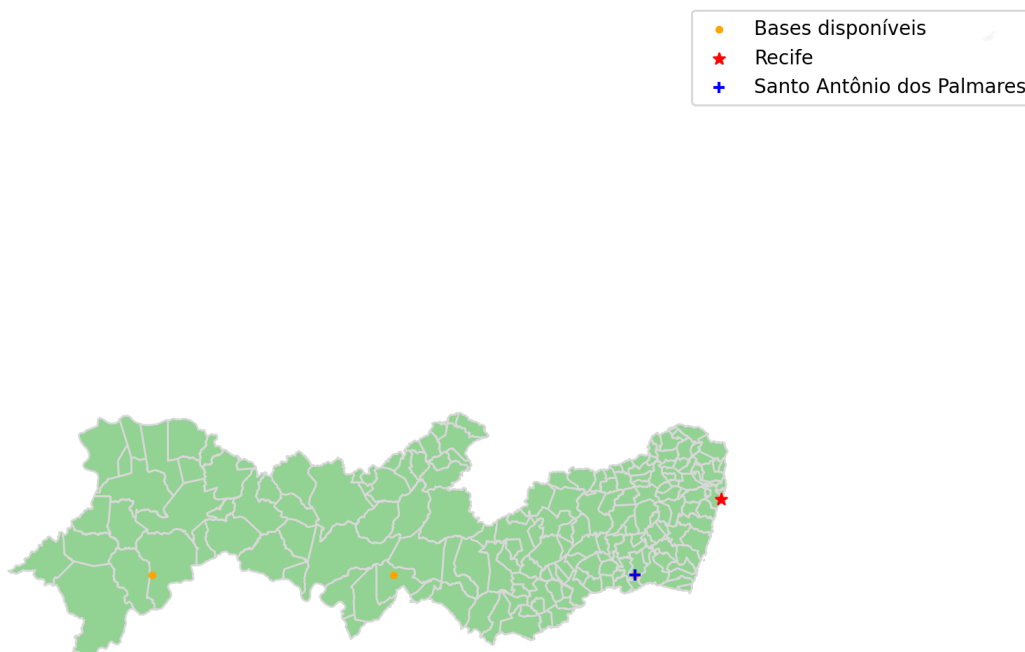


Figura 6 – Base de dados selecionada para o clima Tropical Litorâneo (TL).

e, por isso, foi selecionada (Figura 7).

Pontos para ETo's de Interesse - Minas Gerais

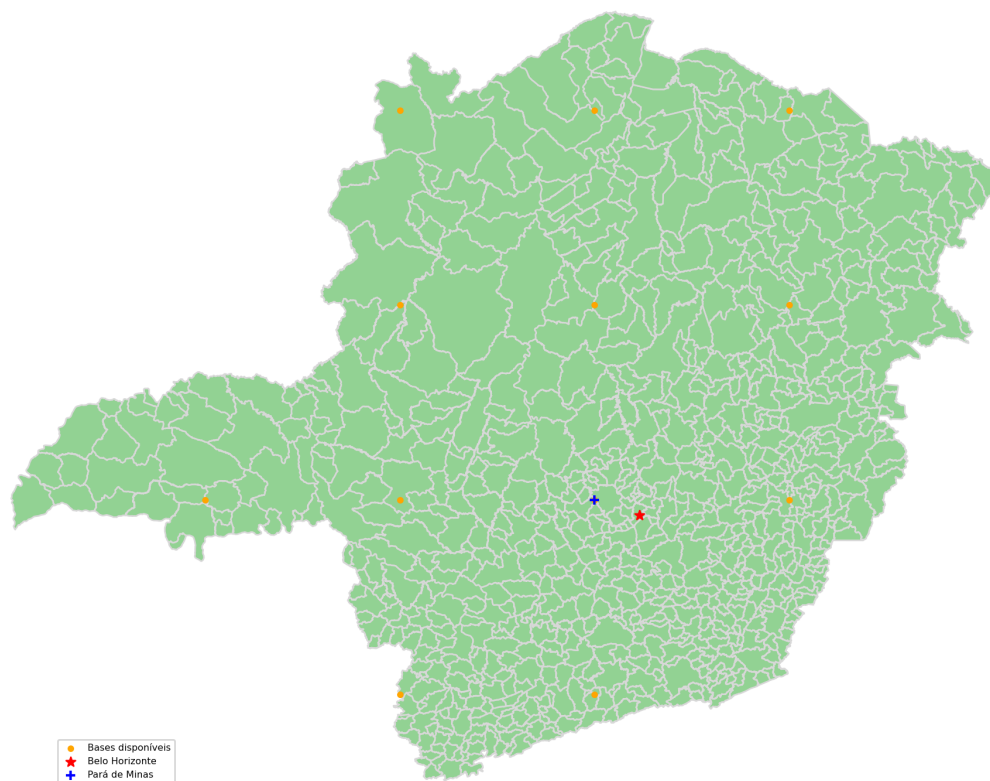


Figura 7 – Base de dados selecionada para o clima Tropical de Altitude (TA).

Para representar o clima Tropical, Cuiabá (Latitude: -15,57, Longitude: -56,08), no Mato Grosso, com estações secas bem definidas e altas temperaturas, foi escolhida como cidade-referência. A base de dados mais próxima fica em Rio da Casca, a -15,35 de latitude e -55,45 de longitude (Figura 8).

Por fim, Alegrete, no Rio Grande do Sul, cujas coordenadas são -29,8 graus de latitude e -55,8 graus de longitude, foi escolhida para o clima Subtropical, por estar no extremo sul do Brasil, com variações de temperatura significativas entre as estações. A base de dados mais próxima tem coordenadas (Latitude: -30,75, Longitude: -55,45) dentro da região do município de Sant'Ana do Livramento (Figura 9).



Figura 8 – Base de dados selecionada para o clima Tropical Típico (TT).

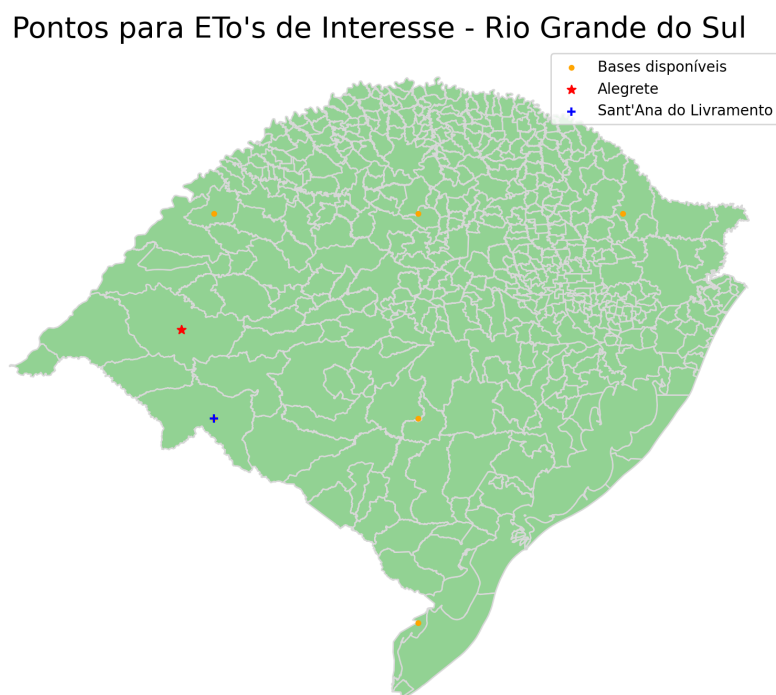


Figura 9 – Base de dados selecionada para o clima Subtropical (ST).

3.2 Análise Descritiva das Séries Meteorológicas

Nas Figuras 10, 11, 12, 13, 14 e 15, apresentam-se as séries temporais das variáveis meteorológicas selecionadas para representar os seis climas brasileiros. As médias de precipitação, temperatura, radiação e umidade relativa, caracterizam bem os diferentes domínios climáticos, de acordo com as classificações apresentadas na Seção 2.2.

Analisando as Tabelas 1, 2, 3, 4, 5 e 6, têm-se: uma média de 2500 *mm/ano* de precipitação, temperatura média acima dos 25°C e alta umidade relativa (acima de 80%) no clima Equatorial; o menor volume médio anual de precipitação, a maior média de radiação solar e de temperatura máxima no clima Semiárido; maior amplitude térmica no clima Tropical de Altitude e maior média de velocidade do vento no clima Tropical Litorâneo, comparadas, respectivamente, com a amplitude térmica e com a média de velocidade do vento do clima Tropical Típico; e as mais baixas temperaturas (entre 0°C e 18°C, com mínima de -3,5°C) e menores índices de radiação solar no clima Subtropical;

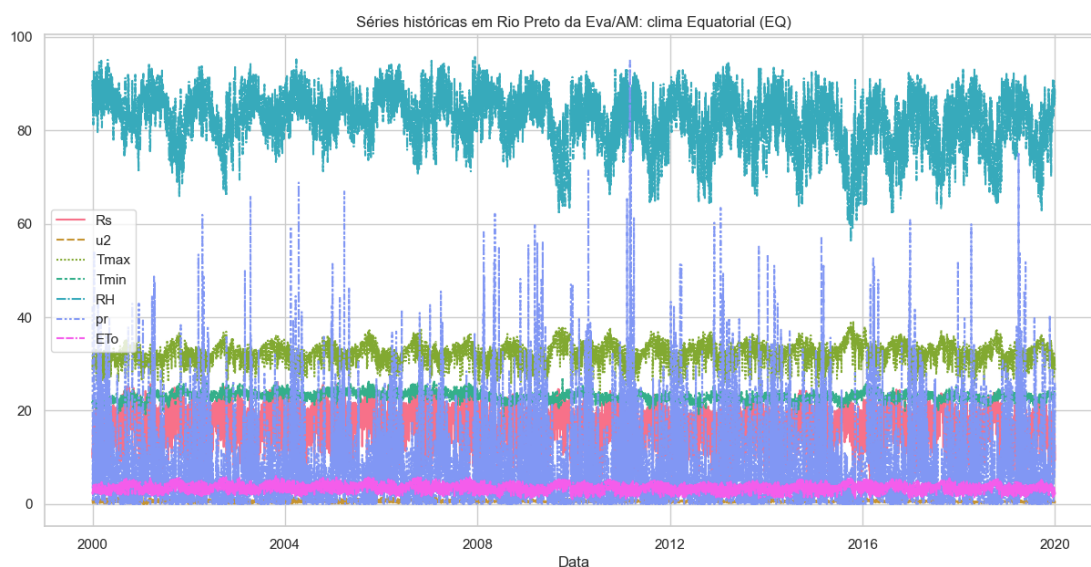


Figura 10 – Séries temporais no clima Equatorial (EQ).

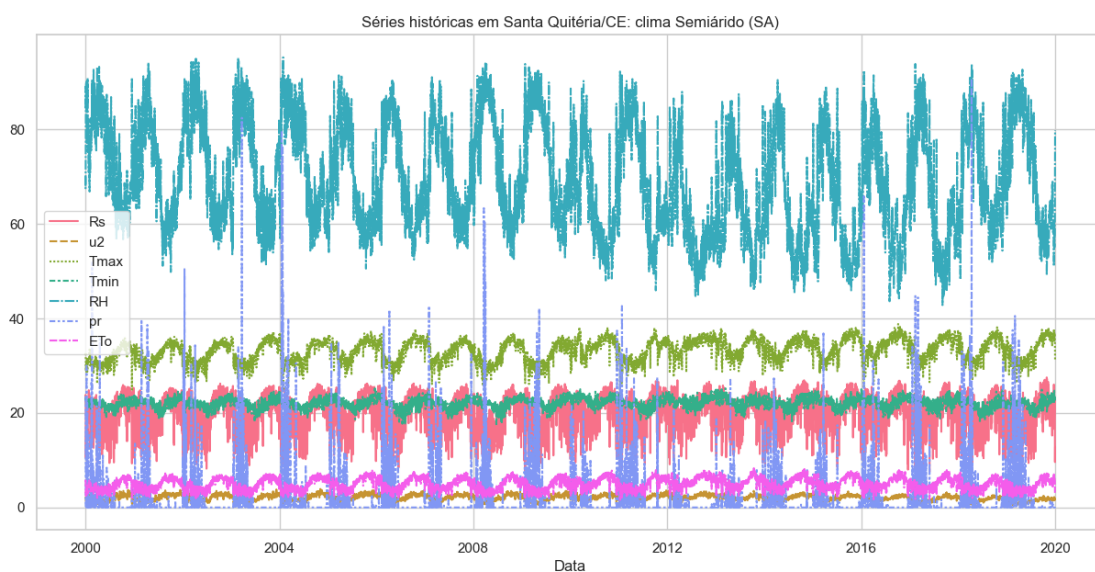


Figura 11 – Séries temporais no clima Semiárido (SA).

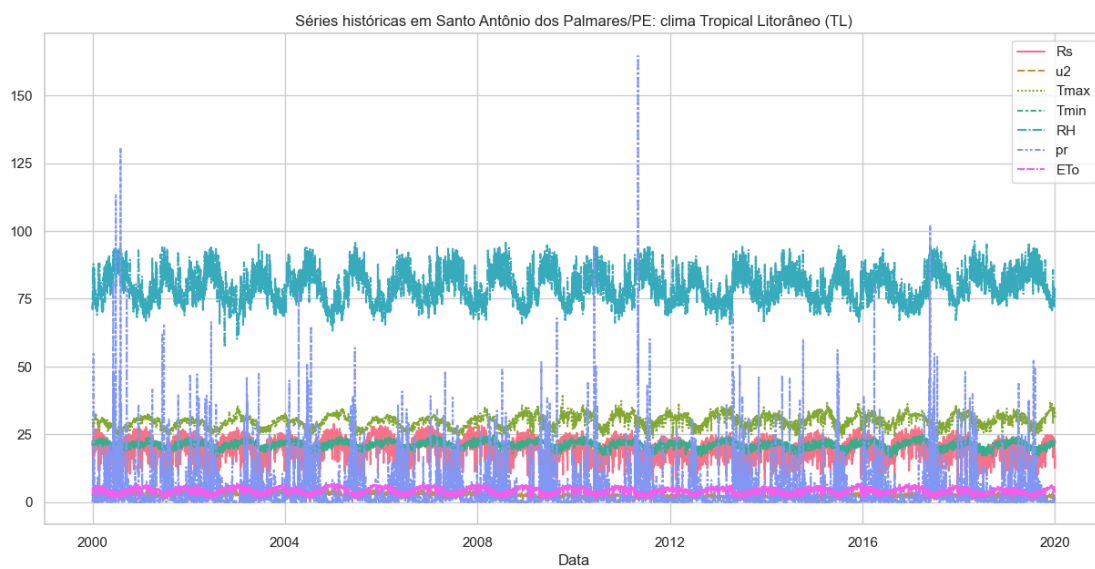


Figura 12 – Séries temporais no clima Tropical Litorâneo (TL).

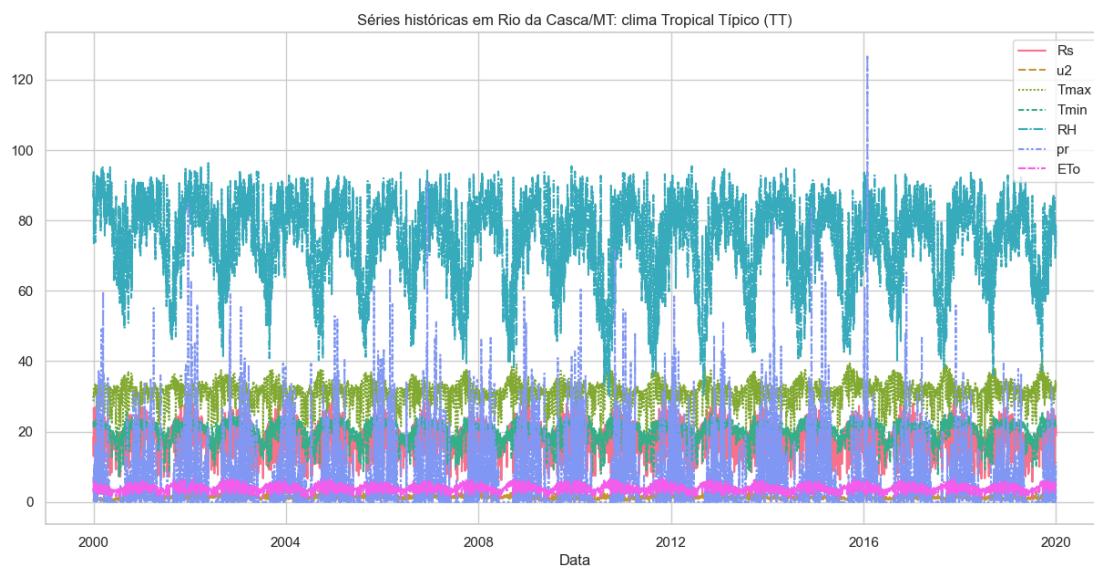


Figura 13 – Séries temporais no clima Tropical Típico (TT).

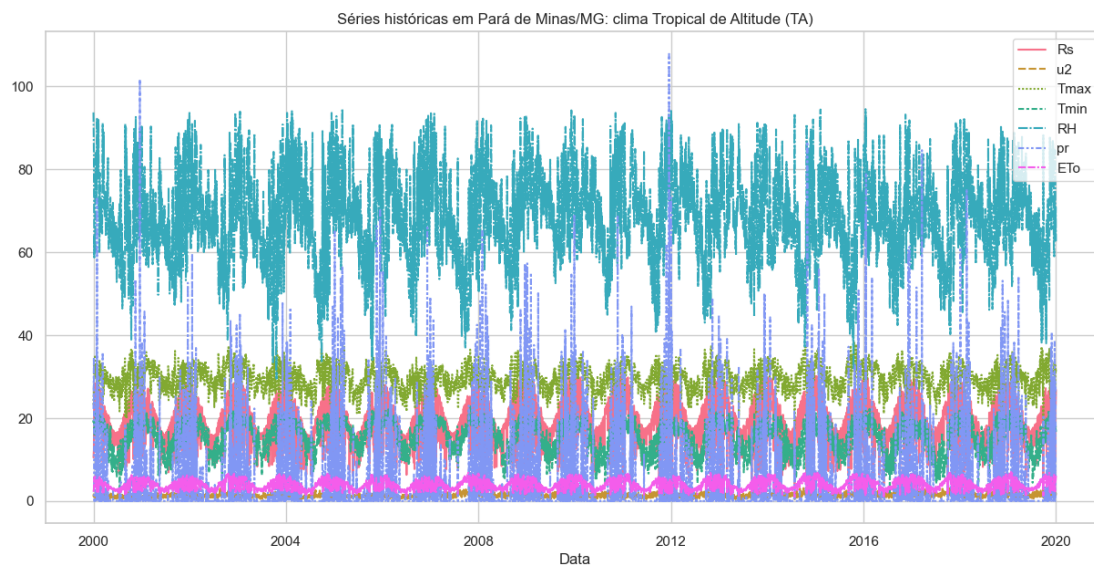


Figura 14 – Séries temporais no clima Tropical de Altitude (TA).

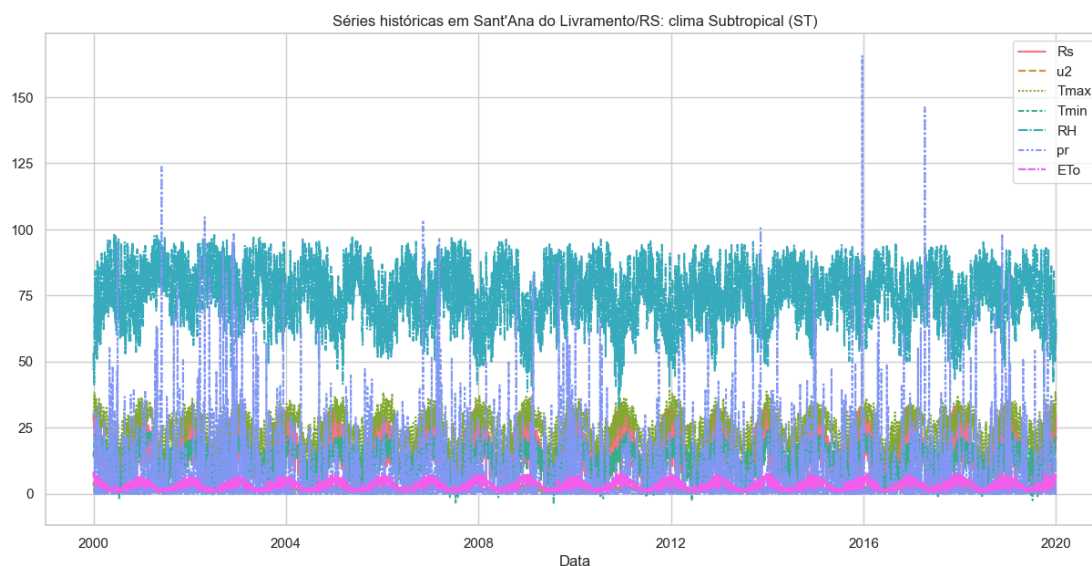


Figura 15 – Séries temporais no clima Subtropical (ST).

Tabela 1 – Descrição estatística das séries no clima Equatorial (EQ).

	Média ± Desvio Padrão	Mínimo	Máximo
Rs [$MJ.m^{-2}.dia^{-1}$]	16.391255 ± 4.391255	3.0	26.1
u2 [$m.s^{-1}$]	0.610732 ± 0.222563	0.0	2.3
Tmax [$^{\circ}C$]	32.266215 ± 2.205483	22.6	39.1
Tmin [$^{\circ}C$]	23.003956 ± 0.952058	18.8	27.1
RH [%]	81.889788 ± 6.048656	56.1	95.7
pr [$mm.dia^{-1}$]	7.017372 ± 9.068709	0.0	95.1
ETo [$mm.dia^{-1}$]	3.538508 ± 0.897718	1.0	6.0

Tabela 2 – Descrição estatística das séries no clima Semiárido (SA).

	Média ± Desvio Padrão	Mínimo	Máximo
Rs [$MJ.m^{-2}.dia^{-1}$]	19.8087 ± 3.7783	7.0	27.5
u2 [$m.s^{-1}$]	1.9643 ± 0.6158	0.5	3.9
Tmax [$^{\circ}C$]	33.2240 ± 2.3120	24.2	38.8
Tmin [$^{\circ}C$]	22.0630 ± 1.0808	17.5	25.2
RH [%]	69.0780 ± 11.4813	42.8	95.2
pr [$mm.dia^{-1}$]	1.6408 ± 5.3445	0.0	90.7
ETo [$mm.dia^{-1}$]	4.9825 ± 1.2738	1.6	8.3

Tabela 3 – Descrição estatística das séries no clima Tropical Litorâneo (TL).

	Média \pm Desvio Padrão	Mínimo	Máximo
Rs [$MJ.m^{-2}.dia^{-1}$]	19.0644 \pm 4.5862	4.1	28.9
u2 [$m.s^{-1}$]	2.3683 \pm 0.7268	0.8	5.4
Tmax [$^{\circ}C$]	29.9075 \pm 2.2818	22.2	39.7
Tmin [$^{\circ}C$]	21.1921 \pm 1.2908	16.5	25.0
RH [%]	79.3385 \pm 5.9280	57.2	96.3
pr [$mm.dia^{-1}$]	4.0552 \pm 8.2106	0.0	164.6
ETo [$mm.dia^{-1}$]	4.2168 \pm 1.1162	1.1	7.2

Tabela 4 – Descrição estatística das séries no clima Tropical Típico (TT).

	Média \pm Desvio Padrão	Mínimo	Máximo
Rs [$MJ.m^{-2}.dia^{-1}$]	17.7279 \pm 4.0295	5.7	29.5
u2 [$m.s^{-1}$]	1.2774 \pm 0.4480	0.1	4.2
Tmax [$^{\circ}C$]	31.2871 \pm 3.0548	16.1	39.9
Tmin [$^{\circ}C$]	20.2662 \pm 2.6980	6.9	25.8
RH [%]	73.9019 \pm 12.2403	30.0	96.3
pr [$mm.dia^{-1}$]	4.9982 \pm 8.7870	0.0	127.0
ETo [$mm.dia^{-1}$]	3.9064 \pm 0.9611	1.2	6.8

Tabela 5 – Descrição estatística das séries no clima Tropical de Altitude (TA).

	Média \pm Desvio Padrão	Mínimo	Máximo
Rs [$MJ.m^{-2}.dia^{-1}$]	18.4022 \pm 4.9933	3.9	31.0
u2 [$m.s^{-1}$]	1.2709 \pm 0.4266	0.2	3.4
Tmax [$^{\circ}C$]	29.5976 \pm 2.9037	17.5	38.7
Tmin [$^{\circ}C$]	15.8885 \pm 3.7374	1.6	22.8
RH [%]	67.9168 \pm 10.7055	29.5	94.6
pr [$mm.dia^{-1}$]	3.5602 \pm 9.0291	0.0	107.8
ETo [$mm.dia^{-1}$]	3.8956 \pm 1.1660	1.0	7.0

Tabela 6 – Descrição estatística das séries no clima Subtropical (ST).

	Média \pm Desvio Padrão	Mínimo	Máximo
Rs [$MJ.m^{-2}.dia^{-1}$]	16.6600 \pm 8.0904	1.1	33.8
u2 [$m.s^{-1}$]	1.8228 \pm 0.6895	0.3	5.6
Tmax [$^{\circ}C$]	23.9273 \pm 6.1288	6.2	39.5
Tmin [$^{\circ}C$]	13.2129 \pm 5.2539	-3.5	25.8
RH [%]	74.7658 \pm 10.6155	33.1	98.4
pr [$mm.dia^{-1}$]	4.4609 \pm 11.0853	0.0	165.6
ETo [$mm.dia^{-1}$]	3.2298 \pm 1.7440	0.5	8.4

3.3 Experimentos Realizados

Uma vez selecionadas as bases de dados representativas de cada clima do Brasil (Seção 3.1), os modelos de previsão de séries temporais foram implementados e os experimentos realizados. Para cada modelo escolhido (VAR, RF, CNN e modelo baseado em *Transformer*), foram desenvolvidos *scripts* em Python (versão 3.11.3), devido à sua versatilidade e à ampla disponibilidade de bibliotecas especializadas em aprendizado de máquina e aprendizado profundo.

Alguns modelos, conforme descrito na Seção 2.3, necessitaram de adaptações para lidar com séries temporais multivariadas, garantindo que tanto as variáveis quanto suas defasagens fossem utilizadas como *features* de entrada. Neste estudo, optou-se por um atraso de 4 dias para todas as variáveis, valor definido com base nos experimentos realizados em (LUCAS et al., 2018). Essa escolha foi confirmada experimentalmente no modelo VAR, através da análise dos critérios de informação Akaike (AIC), Bayesiano (BIC) e Hannan-Quinn (HQIC), cujos valores foram reduzidos (HOU et al., 2023).

Para garantir a reprodutibilidade dos experimentos, os hiperparâmetros dos modelos, apresentados nas Tabelas 7, 8, 9 e 10, foram configurados da seguinte forma:

- Random Forest (RF): Nenhuma otimização inicial foi realizada; utilizou-se os valores padrão da biblioteca scikit-learn, versão 1.6¹, com $n_estimators = 100$.
- CNN: Configuração baseada no modelo CNN1 de (LUCAS et al., 2018).
- Transformer: Valores padrão da biblioteca PyTorch, versão 2.6², com exceção dos parâmetros $input_dim$, $output_dim$, num_epochs e $batch_size$, definidos experimentalmente.

Para garantir a integridade da ordem temporal dos dados, os modelos foram treinados com as seguintes premissas:

¹ RandomForestRegressor: <https://scikit-learn.org/.../sklearn.ensemble.RandomForestRegressor.html>

² torch.nn.Transformer: <https://pytorch.org/docs/stable/.../torch.nn.Transformer>

- O VAR lida naturalmente com múltiplas variáveis e valores defasados.
- Na CNN, as séries e suas defasagens foram explicitamente passados como entrada.
- Para os modelos VAR e CNN, as bases de dados foram divididas em 80% dos registros para o treinamento e 20% para a fase de teste (Figura 16).
- Os modelos RF e *Transformer* necessitaram do uso de janelas deslizantes para treinamento com dados sequenciais. Foram definidas 5 janelas deslizantes, garantindo que, em cada validação, os conjuntos de treinamento fossem aumentados progressivamente, mantendo cerca de 16% dos dados para teste (Figura 17).

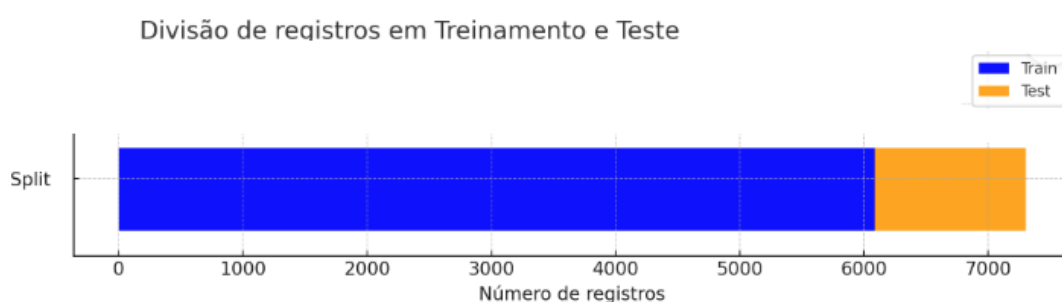


Figura 16 – Divisão das bases de dados em 80% para treinamento e 20% para teste.

A adoção de métodos diferentes para a divisão das bases de dados em conjuntos de treinamento e de teste, ora em um percentual para cada fase, ora pelo método das janelas deslizantes, se deu em razão da ordem em que os modelos foram implementados. Nos primeiros modelos implementados (VAR e CNN), utilizou-se a divisão 80/20 pela facilidade de implementação. Durante o processo de implementação dos modelos RF e *Transformer*, percebeu-se que o método mais recomendado era o baseado na divisão em janelas deslizantes. Portanto, recomenda-se a padronização, com a adoção do método da divisão por janelas deslizantes para todos os modelos.

Os experimentos realizados neste estudo consistiram em 30 execuções de cada modelo utilizando todas as bases de dados representativas dos seis climas

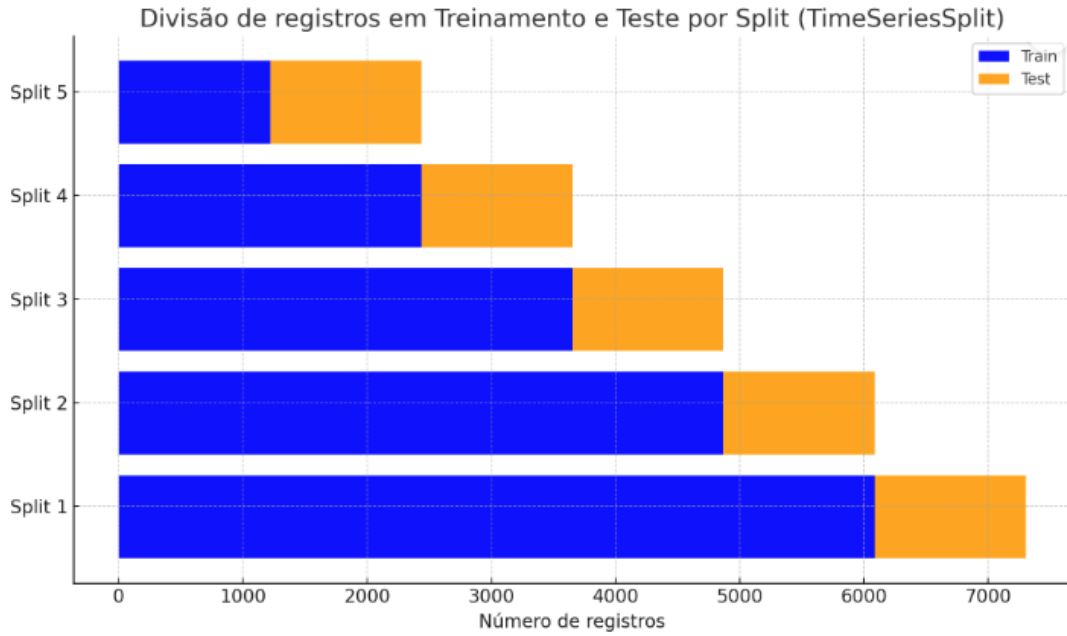


Figura 17 – Divisão das bases de dados em conjuntos de treinamento e teste com 5 janelas deslizantes, utilizando TimeSeriesSplit, da biblioteca scikit-learn.

brasileiros. A métrica de avaliação utilizada foi a Raiz do Erro Médio Quadrático ou *Root Mean Squared Error* (RMSE), representada pela Equação 4.

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}} \quad (4)$$

Onde:

- N é o número de amostras,
- y_i é o valor observado para a amostra i , e
- \hat{y}_i é o valor previsto pelo modelo para a amostra i .

Inicialmente, todas as variáveis (Rs , $u2$, $Tmax$, $Tmin$, RH , pr , ETo) e suas defasagens foram utilizadas como entrada. Posteriormente, após uma análise da importância das variáveis, os experimentos foram repetidos considerando apenas as quatro séries mais relevantes e o histórico da ETo .

Todas as bases de dados e *scripts* Python encontram-se disponíveis em <https://github.com/especializacao-ia-ifmg-op/tcc>.

Hiperparâmetro	Valor
max_lags : número máximo de defasagens (ordem)	4

Tabela 7 – Hiperparâmetros definidos para o modelo VAR.

Hiperparâmetro	Valor
n_estimators : número de árvores aleatórias	100

Tabela 8 – Hiperparâmetros definidos para o modelo RF.

Hiperparâmetro	Valor
n_var : número de variáveis de entrada	7
n_lags : número de defasagens	4
num_conv : número de camadas de convolução	1
filters : número de filtros da camada de convolução	32
pool_size : janela máxima de agrupamento <i>pool</i>	3
kernel_size : tamanho da janela de convolução	11

Tabela 9 – Hiperparâmetros definidos para o modelo CNN.

Hiperparâmetro	Valor
input_dim : número de variáveis de entrada, incluindo as defasagens	24
nhead : número de camadas de atenção	8
num_layers : número de sub-camadas do <i>Encoder</i>	6
output_dim : número de variáveis-alvo (<i>ETo</i>)	1
learning_rate : taxa de aprendizado	0.001
num_epochs : número de iterações para aprendizagem	100
batch_size : número de amostras para aprendizagem	32

Tabela 10 – Hiperparâmetros definidos para o modelo baseado em *Transformer*.

4 Resultados e Discussão

Nesta seção, são apresentados os principais resultados dos experimentos realizados. Com o intuito de evidenciar a relação entre as séries e os respectivos domínios climáticos, também são apresentados os gráficos das séries temporais utilizadas e sua descrição estatística correspondente. Realiza-se ainda, uma análise da importância das variáveis de entrada em cada um dos modelos implementados, utilizando as ferramentas apropriadas em cada caso. Por fim, são apresentadas as representações gráficas dos comportamentos dos modelos, individualmente e comparativamente, em cada clima brasileiro.

4.1 Análise de Importância das Séries Meteorológicas

Visando melhor compreensão dos resultados das previsões realizadas, buscou-se determinar a importância das séries meteorológicas utilizadas como entradas para os modelos multivariados. Cada um dos modelos dispõe de diferentes ferramentas para avaliar o impacto de cada variável de entrada (ou preditor) na variável de saída (variável alvo).

Inicialmente, avaliou-se, para cada clima, a existência de correlação entre todas as variáveis, através dos mapas de correlação das variáveis. Observou-se que, em todos os domínios climáticos, a ET_0 tem uma forte correlação com a radiação solar (R_s), com a temperatura máxima do ar (T_{max}) e a umidade relativa do ar (RH). Com as duas primeiras variáveis, a correlação é positiva, ou seja, um aumento em R_s ou T_{max} produz um aumento em ET_0 , enquanto a correlação entre ET_0 e a umidade relativa é negativa, indicando que uma diminuição em RH produz uma diminuição em ET_0 , e vice-versa.

No clima Semiárido, a velocidade do vento apresenta uma forte correlação positiva, enquanto no clima Tropical Litorâneo, a precipitação tem significativa correlação negativa com a ET_0 . No clima Subtropical, a temperatura mínima (T_{min}) é mais uma variável que se correlaciona positivamente com a ET_0 .

Constatadas as correlações entre as variáveis meteorológicas, foram utilizadas ferramentas, próprias de cada tipo de modelo, na tentativa de determinar quais as variáveis, e suas defasagens, apresentaram maior impacto no resultado das previsões de ETo . Para o modelo VAR, foi utilizado o método *Impulse Response Function* (IRF), para o modelo *Random Forest*, utilizou-se o *Gini Importance*, e para os modelos baseados em redes neurais (CNN e *Transformer*), utilizou-se o SHAP.

Os resultados da aplicação de IRF às bases de dados dos diferentes climas com o modelo VAR, considerando um período de 10 dias à frente do momento do impulso, são apresentados nos gráficos de impulso-resposta no Apêndice A. Em cada gráfico, vê-se a resposta da variável ETo a um impulso em cada uma das variáveis de entrada.

No clima Equatorial, por exemplo, um impulso positivo na radiação solar (Rs) gera um impulso inicial positivo na ETo , que rapidamente retoma seu equilíbrio, enquanto a umidade relativa (RH) gera uma resposta negativa em ETo , cujo impacto demora a se dissipar. Outras variáveis, como a velocidade do vento ($u2$), a temperatura máxima ($Tmax$) e a temperatura mínima ($Tmin$), geram uma resposta inicial positiva que, gradativamente, atinge um impacto negativo e aos poucos permite que a ETo volte ao seu estado de equilíbrio, enquanto o gráfico da precipitação (pr) sugere que esta variável não exerce qualquer impacto significativo na ETo , uma vez que a resposta inicial ao impulso é igual a zero.

De maneira geral, a ETo teve o mesmo comportamento em relação à radiação solar em todos os climas: a resposta inicial a um impulso em Rs é positiva, mas a ETo retorna rapidamente ao seu estado de equilíbrio. A resposta da ETo a impulsos na umidade relativa também é similar em todos os domínios climáticos: inicialmente, a resposta é negativa, mas gradualmente, a ETo se aproxima do equilíbrio, mas mantém uma resposta ligeiramente negativa durante todo o tempo.

A precipitação não apresentou impacto na resposta da ETo , na maioria dos climas, indicando que não influencia significativamente a previsão da ETo . Curiosamente, no clima Tropical de Altitude, a ETo apresentou uma resposta positiva a um impulso em pr , que se manteve persistente ao longo dos períodos. No clima Tropical Típico, a resposta inicial ao impulso em pr foi nulo, mas a ETo teve uma influência levemente positiva, que se manteve ao longo do tempo.

Em todos os seis climas, a ETo teve o mesmo comportamento em relação ao

impulso em $u2$ e $Tmax$: ora teve uma resposta inicial positiva, que gradualmente alcançou o equilíbrio, ora essa resposta inicial positiva se manteve ao longo dos períodos da análise. Já, a temperatura mínima, na maioria das vezes gerou uma resposta semelhante à gerada por $u2$ e $Tmax$ (climas EQ, TL, TT, TA e ST), mas em SA, a resposta inicial da ETo foi positiva, passando a ser levemente negativa ao longo do tempo.

Analisando os gráficos de importância das variáveis (Apêndice B), obtidos através do método *Gini Importance* do modelo *Random Forest*, observa-se que, em todos os climas, a variável de maior importância é a radiação solar (Rs), com exceção do clima Semiárido (SA), em que a temperatura máxima do ar ($Tmax$) aparece como a variável de maior importância e Rs como a segunda mais importante.

Nos climas Equatorial, Tropical Típico e Subtropical, $Tmax$ apresenta-se como a segunda variável mais importante, seguida pela ETo defasada de um dia ou pela velocidade do vento ($u2$). No clima Tropical Litorâneo, quem aparece como segunda variável mais importante é a umidade relativa do ar (RH), enquanto no clima Tropical de Altitude, a ETo , defasada de um dia, é que assume essa posição, seguida pela temperatura mínima ($Tmin$).

De maneira geral, verifica-se que as séries de maior importância para o modelo *Random Forest* foram: a radiação solar (Rs), a temperatura máxima ($Tmax$), a umidade relativa do ar (RH) e a velocidade do vento ($u2$), além da própria ETo .

Essa mesma ideia de importância das variáveis foi aplicada ao modelo CNN, por meio do método SHAP, descrito na Seção 2.3.6. Os valores SHAP representados graficamente no Apêndice C, indicam o impacto de cada variável meteorológica (ou de seu valor defasado) no resultado da previsão obtida pelo modelo. As linhas verticais, de cima para baixo, indicam a ordem decrescente de importância das variáveis, ou seja, o nível de impacto das variáveis no resultado da previsão. As amostras de valor mais alto apresentam coloração mais avermelhada, enquanto as de mais baixo valor apresentam coloração azulada. Os pontos à direita do eixo central indicam um impacto positivo, enquanto os mais à esquerda indicam um impacto negativo no resultado da previsão da variável alvo.

Para 50 pontos amostrados, o gráfico do clima Equatorial indica que as variáveis de maior impacto foram, considerando-se apenas as 7 primeiras: a umidade

relativa (RH), defasada de três dias; a temperatura mínima ($Tmin$), defasada de três dias; a temperatura máxima ($Tmax$), defasada de quatro dias; a própria ETo ; a velocidade do vento ($u2$), defasada de um dia; $Tmin$ e $Tmax$. Portanto, verifica-se que RH_lag3 e $u2_lag1$ apresentaram impacto negativo, de baixa magnitude. Quanto a $Tmin_lag3$, ETo , $Tmin$ e $Tmax$, todas apresentaram impacto positivo, a primeira com alta magnitude e as demais com baixa magnitude.

No clima Semiárido, as variáveis de maior impacto foram a umidade relativa, as temperaturas mínima e máxima e a velocidade do vento, com defasagens de 3 ou 4 dias: RH_lag3 , $Tmin_lag3$, $Tmin_lag4$, RH , $u2_lag3$, $Tmax$ e $Tmax_lag4$. As variáveis que apresentaram maior magnitude foram RH (negativamente) e $Tmin$ (positivamente), ambas com defasagem de três dias. Os climas Tropical Litorâneo e Tropical Típico também tiveram a umidade relativa, defasada em 3 dias, como a variável de maior impacto. Entretanto, curiosamente, RH_lag3 teve impacto positivo sobre a magnitude da saída no clima Tropical Típico.

Nos climas Tropical de Altitude e Subtropical, a variável de maior impacto previsão foi a temperatura mínima, defasada em três dias ($Tmin_lag3$), seguida, respectivamente, pela umidade relativa, defasada em 3 dias (RH_lag3) e pela velocidade do vento, defasada em 4 dias ($u2_lag4$).

De forma geral, o método SHAP indicou que o modelo CNN considerou as seguintes variáveis como as de maior impacto no valor do resultado obtido nas previsões de ETo : umidade relativa (RH), temperaturas mínima ($Tmin$) e máxima ($Tmax$), velocidade do vento ($u2$) e radiação solar (Rs), com defasagens de 1 a 4 dias ($_lag1$ a $_lag4$).

Com o modelo baseado em *Transformer*, foram considerados apenas 25 pontos de amostra para o cálculo dos valores SHAP. Os resultados para os seis climas brasileiros estão representados nos gráficos do Apêndice D.

No clima Equatorial, as variáveis de maior impacto na previsão de ETo foram: a umidade relativa, defasada em 1 e 2 dias (RH_lag_1 , RH_lag_2); a radiação solar, defasada de 1 dia (Rs_lag_1); a temperatura mínima, defasada em 1, 4 e 2 dias ($Tmin_lag_1$, $Tmin_lag_4$, $Tmin_lag_2$); a temperatura máxima, com defasagem de 2 dias ($Tmax_lag_2$); e a velocidade do vento, defasada em 1 dia ($u2_lag_1$).

No Semiárido, a precipitação aparece como uma das variáveis mais im-

portantes, com defasagens de 4, 3 e 1 dia (pr_lag_4 , pr_lag_3 , pr_lag_1). A radiação solar também se apresenta como uma variável de impacto importante na previsão de ETo , com seus valores defasados em 1, 3 e 2 dias (Rs_lag_1 , Rs_lag_3 , Rs_lag_2). A temperatura máxima, defasada em 3 dias ($Tmax_lag_3$) é outra variável que gera impacto importante na saída do modelo.

A radiação solar aparece como uma variável de impacto nos climas Tropical Litorâneo, Tropical Típico, Tropical de Altitude e, principalmente, no clima Subtropical, em que aparece representada por seus valores defasados de 1 a 4 dias (Rs_lag_1 , Rs_lag_2 , Rs_lag_3 , Rs_lag_4). Outra variável presente nos resultados de todos os modelos é a umidade relativa, defasadas entre 1 a 4 dias, com destaque para os climas Tropical Litorâneo e Tropical de Altitude. A velocidade do vento também se mostrou uma variável de impacto significativo, principalmente, nos climas Tropicais (TL, TT e TA). No clima Subtropical, além da radiação solar, a temperatura mínima, com defasagens de 1, 4 e 3 dias ($Tmin_lag_1$, $Tmin_lag_4$, $Tmin_lag_3$), tem impacto significativo na previsão, segundos os resultados.

Portanto, o método SHAP para o modelo baseado em *Transformer* apresentou como variáveis de maior importância para a previsão de ETo : a radiação solar (Rs), a umidade relativa (RH), as temperaturas mínima ($Tmin$) e máxima ($Tmax$) e a velocidade do vento ($u2$), para todos os climas. Além disso, a precipitação (pr) se mostrou importante para os climas tropicais (TL, TT e TA).

Com base na análise das representações gráficas das importâncias das variáveis para cada modelo e clima, os resultados foram sintetizados na Tabela 11. É importante destacar que a variável Rs não foi representada nessa tabela devido à sua correlação consistentemente positiva com a ETo . Além disso, a variável ETo , sendo a variável alvo, não está incluída entre as variáveis explicativas.

Tabela 11 – Variáveis de maior importância para cada modelo nos diferentes climas

Modelo	EQ	SA	TL	TT	TA	ST
VAR	Tmax, Tmin, RH	u2, Tmax, Tmin, RH	u2, Tmax, Tmin, RH	u2, RH, pr	u2, Tmin, RH, pr	Tmin, RH, pr
RF	Tmax, u2, Tmin, RH	Tmax, u2, RH	RH, Tmax, u2	Tmax, u2, RH, Tmin	Tmin, u2, Tmax, RH	Tmax, u2, RH, Tmin
CNN	RH, Tmin, Tmax, u2	RH, Tmin, u2, Tmax	RH, Tmax, u2, Tmin	RH, Tmax, Tmin, pr, u2	Tmin, RH, pr	Tmin, u2, RH
Transformer	RH, Tmin, Tmax, u2	pr, RH, Tmax, u2	Tmin, Tmax, RH, u2	u2, RH, pr, Tmax	RH, u2, pr	Tmin, RH, u2

4.2 Análise dos Resultados

Nesta seção são apresentados e discutidos os resultados da previsão multivariada de séries temporais diárias da evapotranspiração de referência (ET_o). Os experimentos foram realizados conforme descrito na seção 3.3 e os resultados correspondem ao RMSE das previsões obtidas pelos modelos em dois cenários: 1) utilizando todas as sete variáveis meteorológicas; 2) utilizando apenas as quatro variáveis de maior impacto na ET_o , segundo a análise de importâncias realizada.

4.2.1 Previsões com Todas as Variáveis

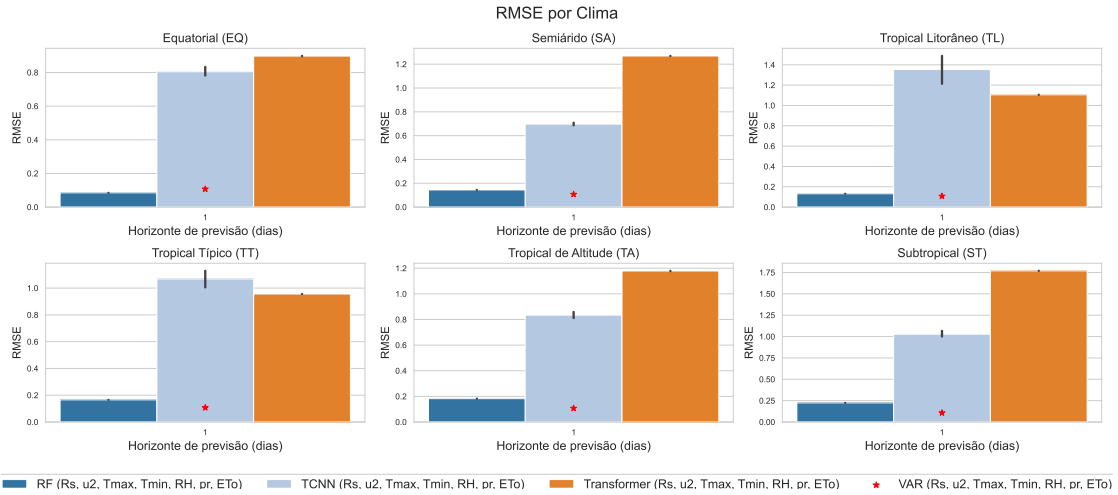
A Tabela 12 apresenta a média do RMSE e o desvio padrão de 30 execuções dos modelos VAR, RF, CNN e modelo baseado em *Transformer*, das previsões feitas um horizonte de previsão igual a 1 dia a frente, utilizando todas as variáveis meteorológicas como entrada. Esses resultados também podem ser visualizados graficamente na Figura 18. Percebe-se, pelos valores das médias, que o modelo VAR apresentou o melhor desempenho em relação aos demais modelos, em quase todos os climas brasileiros, com exceção do clima Equatorial, em que o RF apresentou melhor desempenho.

Os modelos de aprendizado profundo apresentaram médias maiores que o modelo estatístico e o modelo baseado em árvore. Comparando-se apenas esses modelos mais complexos, o CNN apresentou médias melhores em todos os climas do Brasil, com exceção dos climas Tropical Litorâneo e Tropical Típico, em que o modelo baseado em *Transformer* foi ligeiramente melhor.

Analisando por clima os gráficos *boxplot*, vê-se que o modelo VAR apresentou comportamento melhor nos climas Semiárido e Tropical Litorâneo, enquanto o modelo RF teve melhor desempenho no clima Equatorial. O modelo CNN apresentou médias de RMSE menores nos climas Equatorial e Semiárido, maior média e variabilidade no clima Tropical Litorâneo, enquanto o modelo baseado em *Transformer* foi melhor nos climas Equatorial e Tropical Típico.

Tabela 12 – Valores e incertezas de RMSE, em mmd^{-1} , dos modelos VAR, RF, CNN e *Transformer-based* nos diferentes climas do Brasil.

Modelo	EQ	SA	TL	TT	TA	ST
VAR	$0.1530 \pm 2.82e-17$	$0.1077 \pm 1.41e-17$	$0.1073 \pm 2.82e-17$	0.1481 ± 0.0	0.1589 ± 0.0	$0.1338 \pm 2.82e-17$
RF	$0.0849 \pm 1.72e-4$	$0.1444 \pm 6.45e-4$	$0.1313 \pm 4.89e-4$	$0.1656 \pm 5.82e-4$	$0.1823 \pm 5.60e-4$	$0.2237 \pm 5.35e-4$
CNN	0.8047 ± 0.0733	0.6969 ± 0.0344	1.3535 ± 0.3796	1.0682 ± 0.1703	0.8340 ± 0.0691	1.0276 ± 0.0964
Transformer	0.8973 ± 0.0036	1.2691 ± 0.0017	1.1050 ± 0.0017	0.9559 ± 0.0017	1.1784 ± 0.0036	1.7688 ± 0.0016

Figura 18 – Barplot do RMSE dos modelos RF, CNN e modelo baseado em *Transformer* e média do modelo VAR, em cada clima do Brasil, com todas as variáveis como entrada.

4.2.2 Previsões Apenas com as Variáveis de Maior Importância

De acordo com a Tabela 13, no cenário em que as séries de temperatura mínima e de precipitação são suprimidas como entradas para os modelos, os resultados em termos de RMSE mantiveram os modelos VAR e *Random Forest* como os de melhor desempenho, comparados aos modelos CNN e modelo baseado em *Transformer*.

Entre os modelos de aprendizado profundo, neste cenário o modelo CNN foi melhor em todos os climas, comparado ao modelo baseado em *Transformer*, que no cenário anterior apresentou médias de RMSE menores nos climas TL e TT.

Analisando os modelos individualmente em todos os climas, percebe-se que o comportamento do modelo VAR foi similar nos dois cenários: com todas as variáveis

Tabela 13 – Valores e incertezas de RMSE, em mmd^{-1} , dos modelos VAR, RF, CNN e *Transformer-based* nos diferentes climas do Brasil, utilizando apenas 4 variáveis.

Modelo	EQ	SA	TL	TT	TA	ST
VAR	$0.1527 \pm 2.823e-17$	$0.1073 \pm 1.412e-17$	0.1074 ± 0.0	$0.1480 \pm 5.643e-17$	$0.1586 \pm 2.823e-17$	$0.1331 \pm 2.823e-17$
RF	$0.0879 \pm 2e-4$	$0.1462 \pm 5e-4$	$0.1322 \pm 4e-4$	$0.1756 \pm 5e-4$	$0.1941 \pm 4e-4$	$0.2212 \pm 5e-4$
CNN	0.8217 ± 0.0745	0.6999 ± 0.0468	0.6913 ± 0.0749	0.7669 ± 0.0577	0.7491 ± 0.0183	0.9852 ± 0.0529
Transformer	0.8983 ± 0.0049	1.2677 ± 0.0011	1.1063 ± 0.0021	0.9561 ± 0.0012	1.1723 ± 0.0021	1.7682 ± 0.0015

e com a supressão das séries de temperatura mínima e de precipitação. Isso reflete a análise das funções impulso-resposta, em que se verifica que um impulso na precipitação não gera qualquer resposta inicial na *ETo* e que o comportamento da *ETo* em relação à *Tmin* foi muito similar à resposta em relação à *Tmax*, que não foi suprimida no segundo cenário.

O mesmo acontece com o modelo RF. Neste caso, o comportamento pode ser explicado pelos resultados dos cálculos das importâncias obtidos pelo método *Gini Importance*, que indicam que as variáveis suprimidas (*Tmin* e *pr*) influenciaram muito pouco na previsão da *ETo*, em todos os climas. Os gráficos podem indicar ainda que na etapa de seleção aleatória dos preditores, as árvores que incluíram *Tmin* e *pr* tiveram menor peso no resultado final da previsão pelo modelo.

Em relação ao modelo CNN, observa-se que a supressão das variáveis *Tmin* e *pr* afetou o seu comportamento em todos os domínios climáticos. Sobretudo nos climas Tropical Litorâneo e Tropical Típico, o modelo obteve melhores resultados em termos de média de RMSE. No clima Tropical de Altitude, a média foi levemente maior, mas com exceção dos climas Equatorial e Semiárido, em que os erros de previsão foram maiores do que no cenário com todos os preditores, neste segundo cenário os resultados tiveram menores desvios-padrão..

Por outro lado, quanto ao modelo baseado em *Transformer*, a não utilização da temperatura mínima e da precipitação como variáveis preditoras manteve o comportamento do modelo em todos os climas, como médias muito semelhantes, sugerindo que a supressão das variáveis de precipitação e temperatura mínima não promoveu qualquer ajuste ao modelo.

Graficamente, os valores de RMSE da tabela acima estão representados na Figura 19. Visualmente, percebe-se como os modelos mais simples (VAR e RF) tiveram melhor desempenho na previsão multivariada da evapotranspiração, em

relação aos modelos CNN e *Transformer*.

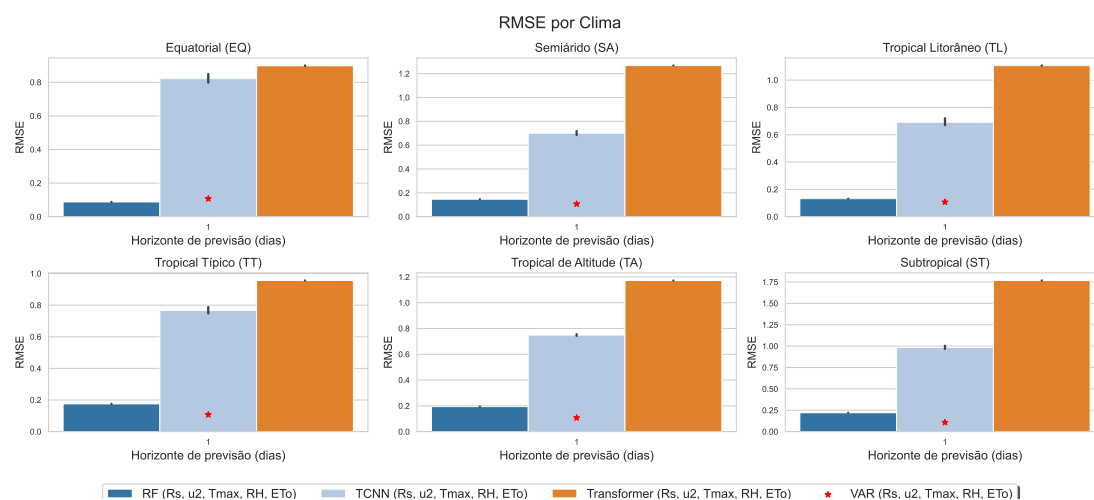


Figura 19 – Barplot do RMSE dos modelos RF, CNN e *Transformer* e média do modelo VAR, em cada clima do Brasil, apenas com as variáveis de maior importância.

4.2.3 Uma Breve Discussão dos Resultados

Os resultados deste trabalho indicaram os modelos VAR e *Random Forest* como os de maior consistência em todos os climas brasileiros avaliados, superando os modelos de aprendizado profundo. Esse comportamento está parcialmente alinhado com a literatura, mas também apresenta divergências.

Por exemplo, a precisão do modelo VAR na previsão de parâmetros climáticos é confirmada por estudo que destaca seu potencial em contextos multivariados e sob condições climáticas diversas (ABDALLAH et al., 2020). O desempenho consistente do RF também está de acordo com os achados de estudo, que evidencia sua eficácia em capturar relações não lineares e lidar com variáveis heterogêneas, superando métodos empíricos como Penman-Monteith e Hargreaves-Samani (FERREIRA; CUNHA, 2020).

Além disso, embora as CNNs sejam viáveis e competitivas na previsão univariada de ET_o , modelos tradicionais como ARIMA e WFTS continuam se destacando, especialmente em horizontes de previsão mais curtos (LUCAS, 2019). Neste estudo,

com modelos multivariados, a CNN demonstrou ser sensível à seleção de variáveis, com desempenho melhorado quando apenas as variáveis mais importantes foram utilizadas como entrada. Este comportamento é condizente com os resultados de estudos que destacam a capacidade das convoluções dilatadas de capturar padrões temporais relevantes com entradas otimizadas (BAI; KOLTER; KOLTUN, 2018; LUCAS, 2019).

Por outro lado, os modelos baseados em *Transformer* não superaram os modelos tradicionais neste estudo, contrastando com trabalho em que *Transformers* combinados com CNNs e LSTMs tiveram melhor desempenho em previsões de *ETo* na Espanha (BELLIDO-JIMÉNEZ et al., 2022). Os resultados também contrariam a hipótese de que modelos multivariados de aprendizado profundo, como a TCN e o *Transformer* (originalmente proposto para tarefas de processamento de linguagem natural) teriam vantagem sobre modelos estatísticos na previsão de variáveis meteorológicas, como a *ETo* (AHMADI et al., 2023). Isso pode estar relacionado à variabilidade climática brasileira e à dependência desses modelos de grandes volumes de dados de alta qualidade, como destacado por (ZERVEAS et al., 2020).

Quanto à limitação no volume de dados, é importante destacar que as bases de dados utilizadas neste estudo foram limitadas ao período de 2001 a 2020. Uma expansão para incluir todos os dados históricos do repositório de dados de origem (XAVIER et al., 2022) poderia capturar padrões climáticos mais amplos e melhorar a representação de regiões sub-representadas, como o Semiárido.

5 Considerações Finais

O objetivo deste trabalho foi avaliar o desempenho de diferentes modelos multivariados na previsão da evapotranspiração de referência (ET_o) em seis climas brasileiros, utilizando dados meteorológicos como radiação solar, umidade relativa, temperatura máxima e mínima, velocidade do vento e precipitação.

As bases de dados, selecionadas com base na proximidade geográfica com cidades-referência em cada clima, mostraram-se representativas das características climáticas regionais definidas segundo a classificação por zonas térmicas.

Os modelos VAR, RF, CNN e um modelo baseado em *Transformer* foram selecionados por sua diversidade de abordagens, abrangendo desde métodos tradicionais estatísticos até técnicas modernas de aprendizado profundo. Esses modelos foram avaliados quanto à sua eficácia em prever a evapotranspiração de referência em diferentes condições climáticas brasileiras, demonstrando a capacidade de adaptação e aplicação de cada abordagem em cenários distintos.

A análise de importância das variáveis meteorológicas evidenciou que cada modelo atribui diferentes níveis de influência a essas variáveis nas regiões analisadas, proporcionando um grau de explicabilidade para estudos de ET_o . Com base nessas observações, foram realizados experimentos em dois cenários: utilizando todas as variáveis e, posteriormente, apenas as quatro variáveis de maior relevância.

Modelos mais simples, como VAR e RF, superaram os modelos de aprendizado profundo (CNN e *Transformer*) na maioria dos climas, reforçando a eficácia de abordagens menos complexas para previsões de ET_o em contextos multivariados. O desempenho variou entre os climas, sugerindo que a adequação dos modelos está ligada às condições climáticas e características específicas das bases de dados.

Esses resultados podem contribuir para estratégias mais eficientes no manejo da irrigação. Por exemplo, em regiões onde a radiação solar é a variável dominante, como nos climas equatorial e tropical, técnicas que considerem apenas essa variável podem otimizar o uso da água. No Semiárido, onde a temperatura máxima e a velocidade do vento se destacam, o seu monitoramento pode ser utilizado para ajustar práticas agrícolas e melhorar a eficiência hídrica.

Futuras pesquisas poderiam explorar a integração de modelos híbridos, combinando a capacidade interpretativa dos modelos mais simples com as capacidades dos modelos de aprendizado profundo. Também seria interessante investigar arquiteturas personalizadas de TCN e *Transformers*, próprias para séries temporais e otimizadas para dados climáticos, e avaliar a influência de bases de dados ampliadas. Além disso, ajustes nos hiperparâmetros dos modelos e a utilização de métodos de seleção de variáveis podem ajudar a melhorar a eficiência das previsões e sua aplicação prática no contexto da irrigação.

Referências

ABATZOGLOU, J. T.; DOBROWSKI, S. Z.; HEGEWISCH, K. C. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, v. 5, p. 170191, 2018. Disponível em: <<https://doi.org/10.1038/sdata.2017.191>>. Acesso em: 03 out. 2024.

ABDALLAH, W.; ABDALLAH, N.; MARION, J.-M.; OUEIDAT, M.; CHAUVET, P. *A vector autoregressive methodology for short-term weather forecasting: tests for Lebanon*. 2020. 1555 p. Disponível em: <<https://doi.org/10.1007/s42452-020-03292-y>>.

Agência Nacional de Águas e Saneamento Básico. *Atlas irrigação: uso da água na agricultura irrigada*. Brasília: [s.n.], 2021. Disponível em: <<https://acesse.one/mycN3>>. Acesso em: 23 set. 2024.

AHMADI, A.; DACCACHE, A.; SADEGH, M.; SNYDER, R. L. Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency. *Computers and Electronics in Agriculture*, v. 215, p. 108424, 2023. ISSN 0168-1699. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0168169923008128>>.

ALENCAR, L. P. D.; SEDIYAMA, G. C.; MANTOVANI, E. C. Estimativa da evapotranspiração de referência (eto padrão fao), para minas gerais, na ausência de alguns dados climáticos. *Engenharia Agrícola*, Associação Brasileira de Engenharia Agrícola, v. 35, n. 1, p. 39–50, Jan 2015. ISSN 0100-6916. Disponível em: <<https://doi.org/10.1590/1809-4430-Eng.Agric.v35n1p39-50/2015>>.

ALI, M.; JAMEI, M.; PRASAD, R.; KARBASI, M.; XIANG, Y.; CAI, B.; ABDULLA, S.; Ahsan Farooque, A.; LABBAN, A. H. New achievements on daily reference evapotranspiration forecasting: Potential assessment of multivariate signal decomposition schemes. *Ecological Indicators*, v. 155, p. 111030, 2023. ISSN 1470-160X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1470160X2301172X>>.

ALLEN, R.; PEREIRA, L.; RAES, D.; SMITH, M. Evapotranspiración del cultivo. guías para la determinación de los requerimientos de agua de los cultivos. *Estudio FAO Riego Y Drenaje No 56*, 01 2006.

ALLEN, R.; PEREIRA, L.; SMITH, M. *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56*. [S.l.]: Food and Agriculture Organization of the United States (FAO), 1998. v. 56. 333 p.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. de M.; SPAROVEK, G. Koppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, Schweizerbart Science Publishers, Stuttgart, Germany, v. 22, n. 6, p. 711–728, 12 2013. Disponível em: <<https://doi.org/10.1127/0941-2948/2013/0507>>. Acesso em: 07 out. 2024.

ARNFIELD, A. J. *Köppen climate classification*. 2011. Disponível em: <<https://www.britannica.com/science/Koppen-climate-classification>>. Acesso em: 07 out. 2024.

BAI, S.; KOLTER, J. Z.; KOLTUN, V. *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling*. 2018. Disponível em: <<https://arxiv.org/abs/1803.01271>>.

BASÍLIO, R. da S. *Previsão de séries temporais de evapotranspiração de referência e desenvolvimento de estação meteorológica*. Monografia (Graduação em Engenharia de Controle e Automação) — Universidade Federal de Ouro Preto - Campus Morro do Cruzeiro, Ouro Preto, 2023.

BELLIDO-JIMÉNEZ, J. A.; ESTÉVEZ, J.; VANSCHOREN, J.; GARCÍA-MARÍN, A. P. Agroml: An open-source repository to forecast reference evapotranspiration in different geo-climatic conditions using machine learning and transformer-based models. *Agronomy*, v. 12, n. 3, 2022. ISSN 2073-4395. Disponível em: <<https://www.mdpi.com/2073-4395/12/3/656>>.

BROWNLEE, J. *Random Forest for Time Series Forecasting*. 2020. Acesso em: 10 out. 2024. Disponível em: <<https://machinelearningmastery.com/random-forest-for-time-series-forecasting/>>.

COLLEGE, V. R. S. E. *Feature Importance with Random Forests*. 2024. Acesso em: 05 nov. 2024. Disponível em: <<https://www.geeksforgeeks.org/feature-importance-with-random-forests/>>.

ELBELTAGI, A.; SRIVASTAVA, A.; LI, P.; JIANG, J.; JINSONG, D.; RAJPUT, J.; KHADKE, L.; AWAD, A. Forecasting actual evapotranspiration without climate data based on stacked integration of dnn and meta-heuristic models across China from 1958 to 2021. *Journal of Environmental Management*, v. 345, p. 118697, 2023. ISSN 0301-4797. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0301479723014858>>.

FERREIRA, L.; CUNHA, F.; SILVA, G. da; CAMPOS, F.; DIAS, S. H.; SANTOS, J. Generalizability of machine learning models and empirical equations for the estimation of reference evapotranspiration from temperature in a semiarid region.

Anais da Academia Brasileira de Ciências, v. 23, p. e20200304, 03 2021. Disponível em: <<https://doi.org/10.1590/0001-3765202120200304>>. Acesso em: 03 out. 2024.

FERREIRA, L. B.; CUNHA, F. F. da. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. *Agricultural Water Management*, v. 234, p. 106113, 2020. ISSN 0378-3774. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378377419322383>>. Acesso em: 03 out. 2024.

FERREIRA, L. B.; CUNHA, F. F. da; FILHO, E. I. F. Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across brazil. *Agricultural Water Management*, v. 259, p. 107281, 2022. ISSN 0378-3774. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378377421005588>>. Acesso em: 07 out. 2024.

FILHO, M. A. C.; PEREIRA, F. A.; ANGELOCCI, L. R.; COELHO, E. F.; OLIVEIRA, G. X. S. O processo de evapotranspiração. In: SOUZA, V. F.; MARQUELLI, W. A.; COELHO, E. F.; FILHO, M. A. C. (Ed.). *Irrigação e fertirrigação em fruteiras e hortaliças*. Brasília, DF: Embrapa Informação Tecnológica, 2011. p. 91–113. Disponível em: <<https://encurtador.com.br/167ke>>. Acesso em: 03 out. 2024.

FLEISCHMANN, A. S.; LAIPELT, L.; PAPA, F.; PAIVA, R. C. D. d.; ANDRADE, B. C. de; COLLISCHONN, W.; BIUDES, M. S.; KAYSER, R.; PRIGENT, C.; COSIO, E.; MACHADO, N. G.; RUHOFF, A. Patterns and drivers of evapotranspiration in south american wetlands. *Nature Communications*, v. 14, n. 1, p. 6656, Oct 2023. ISSN 2041-1723. Disponível em: <<https://doi.org/10.1038/s41467-023-42467-0>>.

Food and Agriculture Organization of the United Nations. *Water for Sustainable Food and Agriculture Organization of the United Nations. A report produced for the G20 Presidency of Germany. Technical report*. Roma: Food and Agriculture Organization of the United States (FAO), 2017.

GADDIKERI, V.; JATAV, M. S.; SIDDHARAM; ASHA, K. R.; AISWARYA, L.; PREETI; NAGESWAR, B. Predictive modeling and comparative analysis of reference evapotranspiration with machine learning algorithms. *International Journal of Environment and Climate Change*, v. 13, n. 11, p. 1623–1634, Oct. 2023. Disponível em: <<https://journalijecc.com/index.php/IJECC/article/view/3317>>.

GONÇALVES, R. *Understanding Random Forest, From Theory to Practice*. Tese (Doutorado) — Department of Electrical Engineering and Computer Science,

Faculty of Engineering, University of Porto, Porto, Portugal, 2021. Disponível em: <<https://repositorio-aberto.up.pt/bitstream/10216/135947/2/492480.pdf>>. Acesso em: 14 out. 2024.

GOUVEIA, R. *Climas do Brasil*. s.d. Disponível em: <<https://www.todamateria.com.br/climas-do-brasil/>>. Acesso em: 07 out. 2024.

GUO, P.; CAO, J.; LIN, J. Establishment of a reference evapotranspiration forecasting model based on machine learning. *Agronomy*, v. 14, n. 5, 2024. ISSN 2073-4395. Disponível em: <<https://www.mdpi.com/2073-4395/14/5/939>>.

HANCK, C.; ARNOLD, M.; GERBER, A.; SCHMELZER, M. *Introduction to Econometrics with R*. 2024. <<https://www.econometrics-with-r.org/16.1-vector-autoregressions.html>>. Disponível em: <<https://www.econometrics-with-r.org/ITER.pdf>>. Acesso em: 07 out. 2024.

HOU, P. S.; FADZIL, L. M.; MANICKAM, S.; AL-SHAREEDA, M. A. Vector autoregression model-based forecasting of reference evapotranspiration in malaysia. *Sustainability*, v. 15, n. 4, 2023. ISSN 2071-1050. Disponível em: <<https://www.mdpi.com/2071-1050/15/4/3675>>.

IBGE. *Atlas geográfico escolar*. 9. ed. Rio de Janeiro: IBGE, Centro de Documentação e Disseminação de Informações, 1993. 240 p. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv102069.pdf>>. Acesso em: 07 out 2024.

JENSEN, M. E.; ALLEN, R. G. *Evaporation, Evapotranspiration and Irrigation Water Requirements - ASCE Manuals and Reports on Engineering Practice No. 70*. 2. ed. Reston, Virginia: American Society of Civil Engineers, 2016. 528 p. Disponível em: <<https://doi.org/10.1061/9780784414057>>. Acesso em: 05 out 2024.

KRISHNA, P. R. A. Evapotranspiration and agriculture-a review. *Agricultural Reviews*, v. 40, p. 1–11, 2019. Disponível em: <<https://api.semanticscholar.org/CorpusID:134377557>>.

LIM, B.; ARIK, S. O.; LOEFF, N.; PFISTER, T. *Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting*. 2020. Disponível em: <<https://arxiv.org/abs/1912.09363>>.

LOUPPE, G. *Understanding Random Forest, From Theory to Practice*. Tese (Doutorado) — Department of Electrical Engineering & Computer Science, Faculty of Applied Sciences, University of Liège, Liège, 2014. Disponível em: <<https://arxiv.org/abs/1407.7502>>. Acesso em: 14 out. 2024.

- LOUPPE, G. *Understanding Random Forests: From Theory to Practice*. 2015. Disponível em: <<https://arxiv.org/abs/1407.7502>>.
- LUCAS, P.; MAIA, R.; VICENTE, M.; LEITE, C. Using artificial neural network to estimate reference evapotranspiration. *Global Science and Technology*, v. 11, p. 229–240, 12 2018.
- LUCAS, P. d. O. e. *Previsão de séries temporais de evapotranspiração de referência com redes neurais convolucionais [recurso eletrônico] / Patrícia de Oliveira e Lucas*. 100 p. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Escola de Engenharia, 2019.
- LUCAS, P. d. O. e.; ALVES, M. A.; SILVA, P.; GUIMARÃES, F. Reference evapotranspiration prediction for precision agriculture using fuzzy time series. In: *Anais do Simpósio Brasileiro de Pesquisa Operacional*. Campinas: Galoá, 2022. Disponível em: <<https://proceedings.science/sbpo/sbpo-2022/trabalhos/reference-evapotranspiration-prediction-for-precision-agriculture-using-fuzzy-ti?lang=pt-br>>. Acesso em: 23 out. 2024.
- LUNDBERG, S. M.; ERION, G. G.; LEE, S.-I. *Consistent Individualized Feature Attribution for Tree Ensembles*. 2019. Disponível em: <<https://arxiv.org/abs/1802.03888>>.
- LütKEPOHL, H. *New Introduction to Multiple Time Series Analysis*. Heidelberg, Berlin: Springer, 2005.
- MENESES, K. C. d.; APARECIDO, L. E. D. O.; MENESES, K. C. d.; FARIAS, M. F. d. Estimating potential evapotranspiration in maranhão state using artificial neural networks. *Revista Brasileira de Meteorologia*, Sociedade Brasileira de Meteorologia, v. 35, n. 4, p. 675–682, Oct 2020. ISSN 0102-7786. Disponível em: <<https://doi.org/10.1590/0102-77863540072>>.
- MOLNAR, C. *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*. 2024. Acesso em: 05 nov. 2024. Disponível em: <<https://christophm.github.io/interpretable-ml-book/>>.
- MOURA, A. R. C.; MONTENEGRO, S. M. G. L.; ANTONINO, A. C. D.; AZEVEDO, J. R. G. d.; SILVA, B. B. d.; OLIVEIRA, L. M. M. d. Evapotranspiração de referência baseada em métodos empíricos em bacia experimental no estado de pernambuco - brasil. *Revista Brasileira de Meteorologia*, Sociedade Brasileira de Meteorologia, v. 28, n. 2, p. 181–191, Jun 2013. ISSN 0102-7786. Disponível em: <<https://doi.org/10.1590/S0102-77862013000200007>>.

- NEMBRINI, S.; KÖNIG, I.; WRIGHT, M. The revival of the gini importance? *Bioinformatics (Oxford, England)*, v. 34, 05 2018.
- NOVAIS, G. T. *Climas do Brasil: classificação climática e aplicações*. 1. ed. Porto Alegre, RS: Editora Totalbooks Eireli, 2023. Acesso em: 05 out 2024.
- OLIVEIRA, M. *Climas do Brasil*. 2020. Disponível em: <<https://www.tudogeo.com.br/2020/05/07/climas-do-brasil/>>. Acesso em: 07 out. 2024.
- RODRIGO, J. A.; ORTIZ, J. E. *Global Forecasting Models: Modeling Multiples Time Series with Machine Learning — cienciadedatos.net*. 2022. <<https://cienciadedatos.net/documentos/py44-multi-series-forecasting-skforecast.html>>. [Accessed 05-12-2024].
- _____. *Global Forecasting Models: Comparative Analysis of Single and Multi-Series Forecasting Modeling — cienciadedatos.net*. 2023. <<https://cienciadedatos.net/documentos/py53-global-forecasting-models>>. [Accessed 05-12-2024].
- ROSA, D. R. Q. *Aplicação de machine learning para predição da evapotranspiração de referência (ET_o)*. Monografia (Especialização em Inteligência Artificial e Big Data) — Departamento de Ciências de Computação e Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, São Paulo, 2022.
- RUHOFF, A.; BRAZIL, C.; COLLISCHONN, W.; UVO, C.; ROCHA, H.; MACHADO, O.; CABRAL, R. Análise multivariada do processo de evapotranspiração em Áreas de cerrado e cana-de-açúcar. *Revista Brasileira de Recursos Hídricos*, v. 14, 01 2009.
- Serviço Nacional de Aprendizagem Rural. *Irrigação: gestão e manejo*. / *Serviço Nacional de Aprendizagem Rural. Coleção Senar, 250*. Brasília: [s.n.], 2019. 84 p. Disponível em: <<https://ead.senar.org.br/senar-play/cartilhas/irrigacao-gestao-e-manejo>>.
- SHAHIN, M. A.; ALI, M. A.; ALI, A. B. M. S. Vector autoregression (var) modeling and forecasting of temperature, humidity, and cloud coverage. In: _____. *Computational Intelligence Techniques in Earth and Environmental Sciences*. Dordrecht: Springer Netherlands, 2014. p. 29–51. ISBN 978-94-017-8642-3. Disponível em: <https://doi.org/10.1007/978-94-017-8642-3_2>.
- SHALOO; KUMAR, B.; BISHT, H.; RAJPUT, J.; MISHRA, A. K.; TM, K. K.; BRAHMANAND, P. S. Reference evapotranspiration prediction using machine learning models: An empirical study from minimal climate

- data. *Agronomy Journal*, v. 116, n. 3, p. 956–972, 2024. Disponível em: <<https://access.onlinelibrary.wiley.com/doi/abs/10.1002/agj2.21504>>.
- SILVA, R. R. C. *Previsão de Séries Temporais Multivariadas Utilizando Modelo Híbrido Interpretável com Árvores de Decisão Fuzzy*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal Minas Gerais, Belo Horizonte, 2021.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. *Attention Is All You Need*. 2023. Disponível em: <<https://arxiv.org/abs/1706.03762>>.
- WU, N.; GREEN, B.; BEN, X.; O'BANION, S. *Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case*. 2020. Disponível em: <<https://arxiv.org/abs/2001.08317>>.
- XAVIER, A. C.; KING, C. W.; SCANLON, B. R. Daily gridded meteorological variables in brazil (1980–2013). *International Journal of Climatology*, v. 36, n. 6, p. 2644–2659, 2016. Disponível em: <<https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.4518>>.
- XAVIER, A. C.; SCANLON, B. R.; KING, C. W.; ALVES, A. I. New improved brazilian daily weather gridded data (1961–2020). *International Journal of Climatology*, v. 42, n. 16, p. 8390–8404, 2022. Disponível em: <<https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7731>>.
- ZENG, B. *Towards Understanding Residual Neural Networks*. Tese (Master's thesis) — Massachusetts Institute of Technology, 2019. Disponível em: <<https://dspace.mit.edu/bitstream/handle/1721.1/123067/1127292128-MIT.pdf>>.
- ZERVEAS, G.; JAYARAMAN, S.; PATEL, D.; BHAMIDIPATY, A.; EICKHOFF, C. *A Transformer-based Framework for Multivariate Time Series Representation Learning*. 2020. Disponível em: <<https://arxiv.org/abs/2010.02803>>.
- ZHOU, H.; ZHANG, S.; PENG, J.; ZHANG, S.; LI, J.; XIONG, H.; ZHANG, W. *Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting*. 2021. Disponível em: <<https://arxiv.org/abs/2012.07436>>.

APÊNDICE A – Funções Impulso-Resposta para o VAR

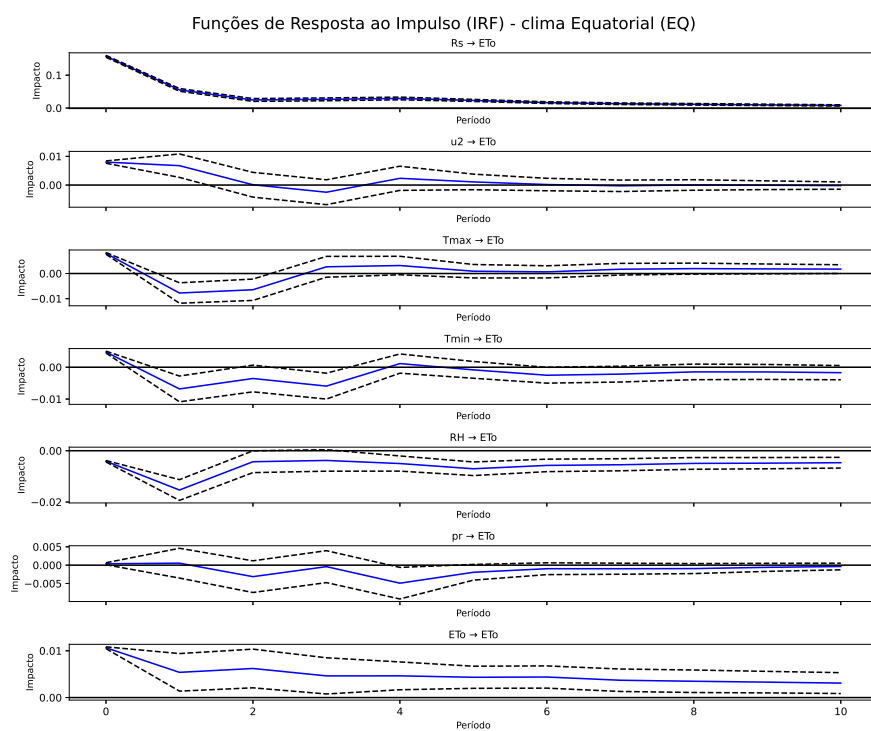


Figura 20 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Equatorial (EQ).

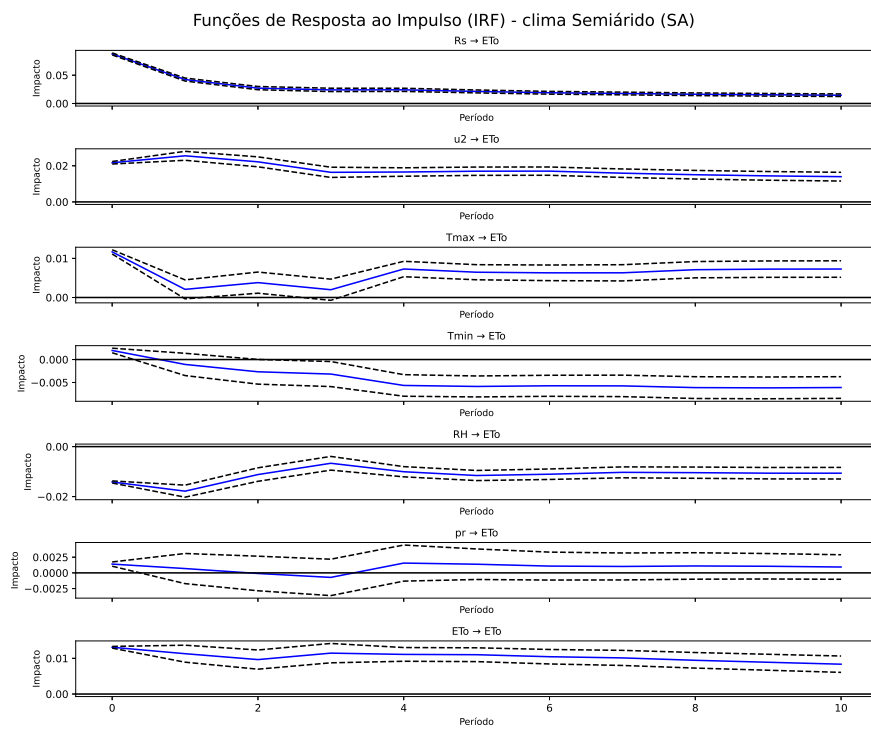


Figura 21 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Semiárido (SA).

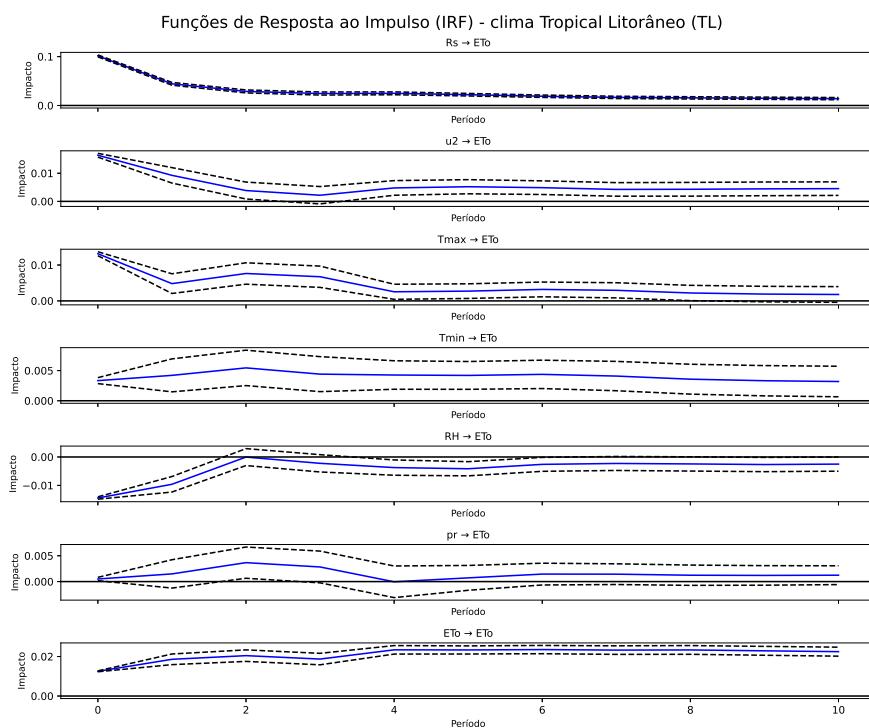


Figura 22 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Tropical Litorâneo (TL).

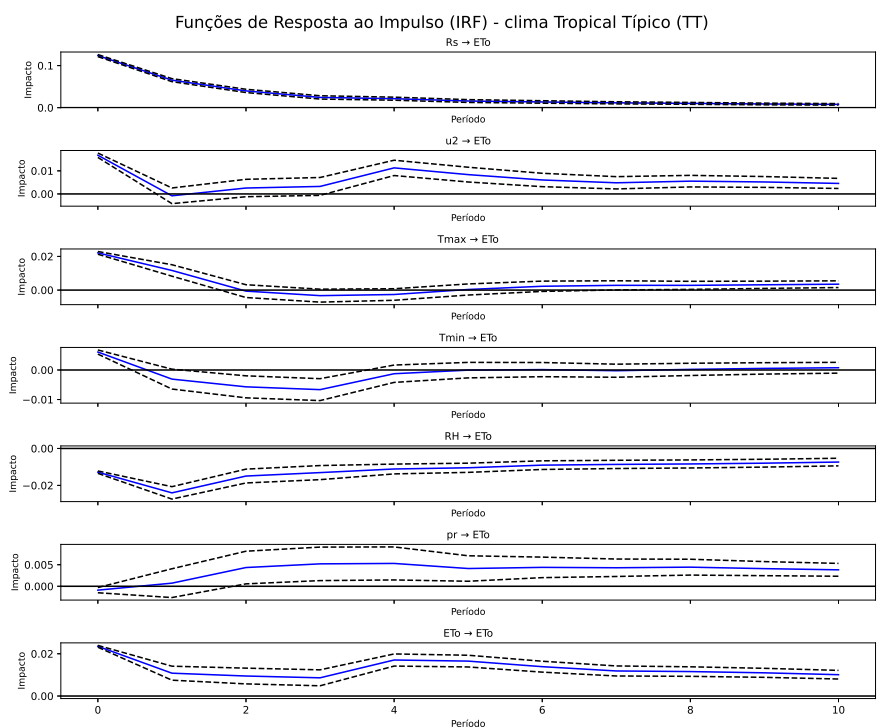


Figura 23 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Tropical Típico (TT).

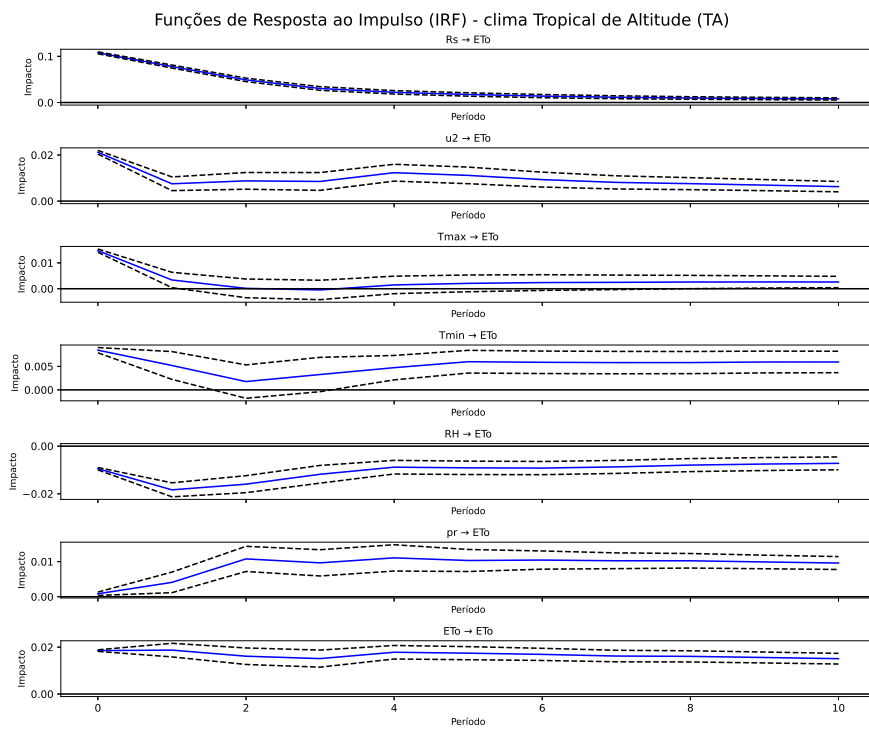


Figura 24 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Tropical de Altitude (TA).

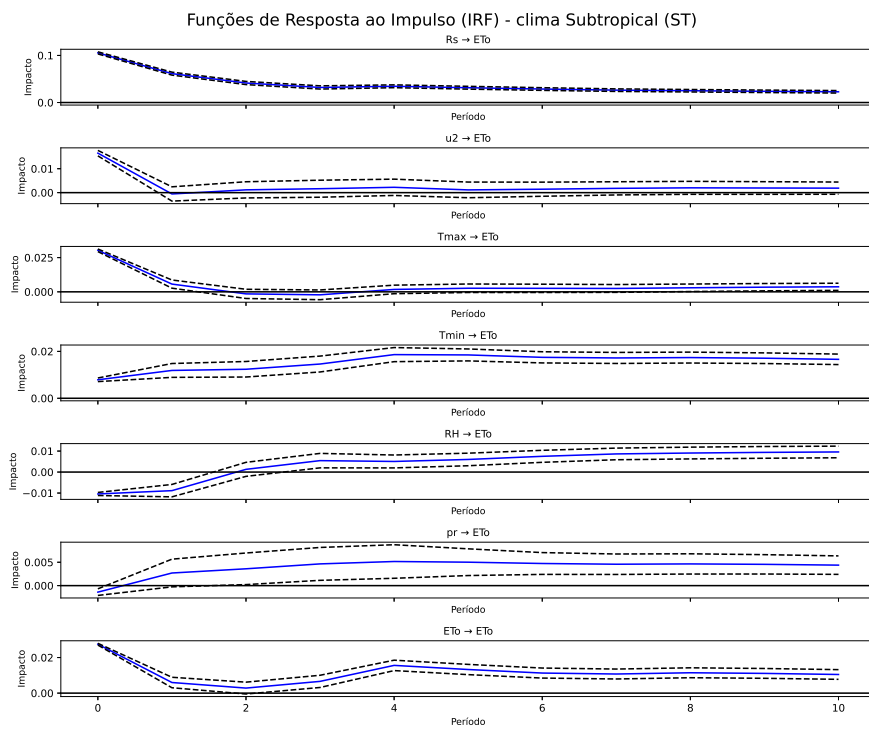


Figura 25 – Funções de Resposta ao Impulso (IRF) para o modelo VAR no clima Subtropical (ST).

APÊNDICE B – *Gini* *Importance* para o RF

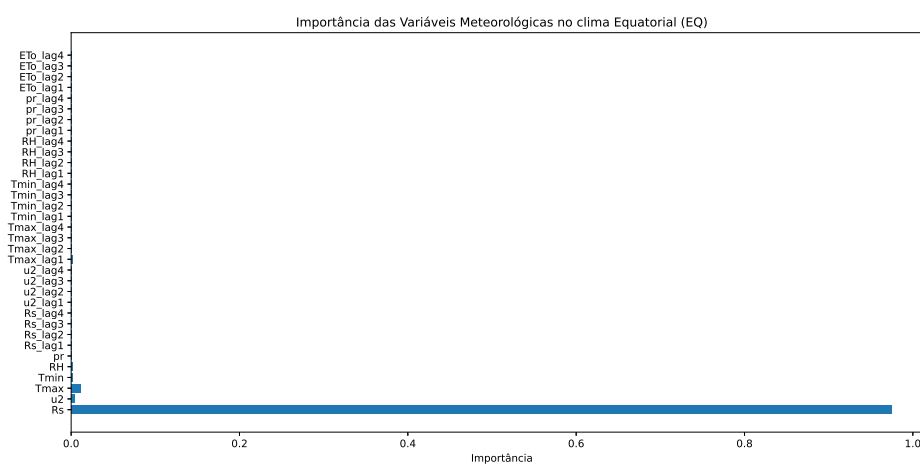


Figura 26 – Importância das variáveis para o modelo RF no clima Equatorial (EQ).

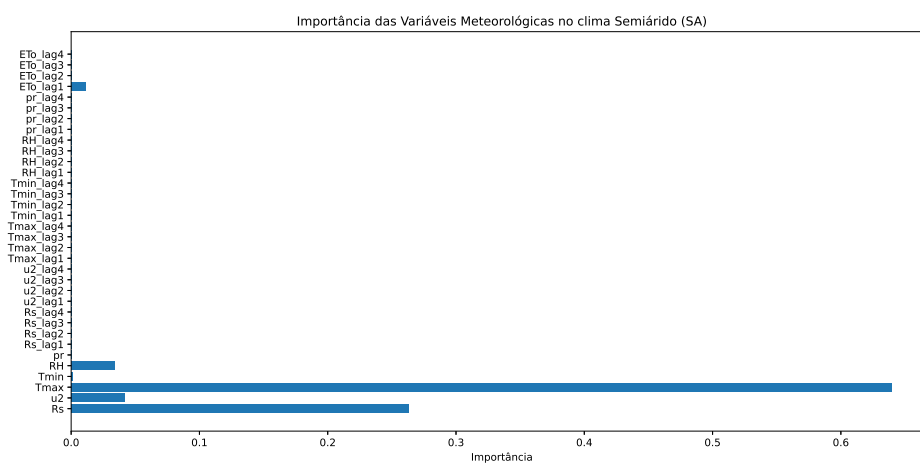


Figura 27 – Importância das variáveis para o modelo RF no clima Semiárido (SA).

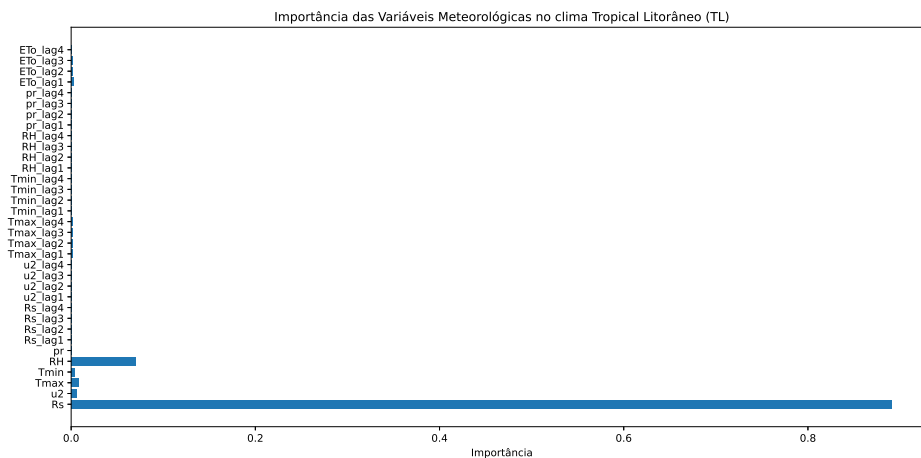


Figura 28 – Importância das variáveis para o modelo RF no clima Tropical Litorâneo (TL).

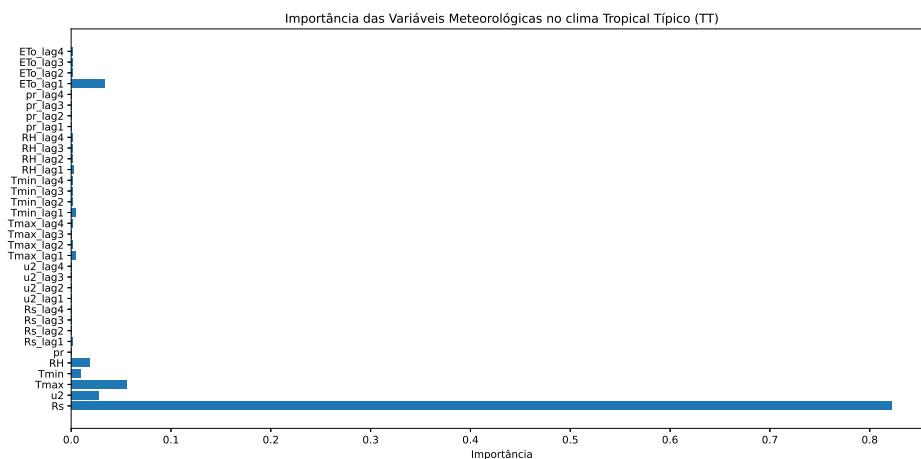


Figura 29 – Importância das variáveis para o modelo RF no clima Tropical Típico.

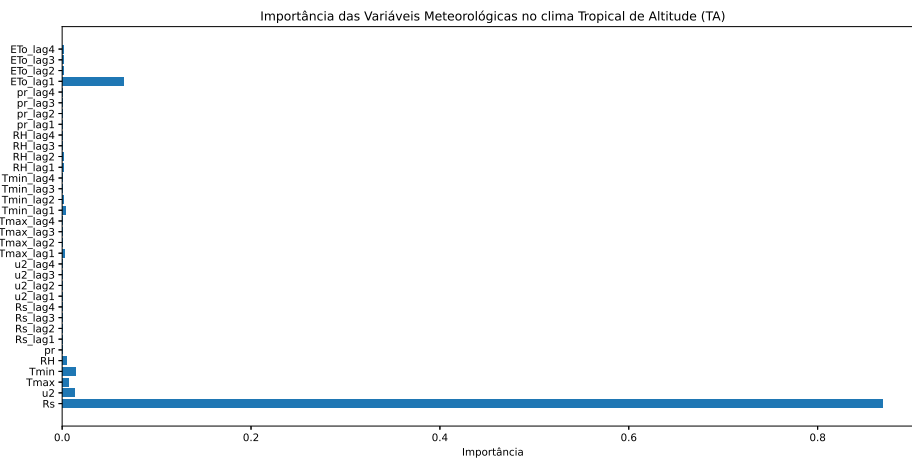


Figura 30 – Importância das variáveis para o modelo RF no clima Tropical de Altitude (TA).

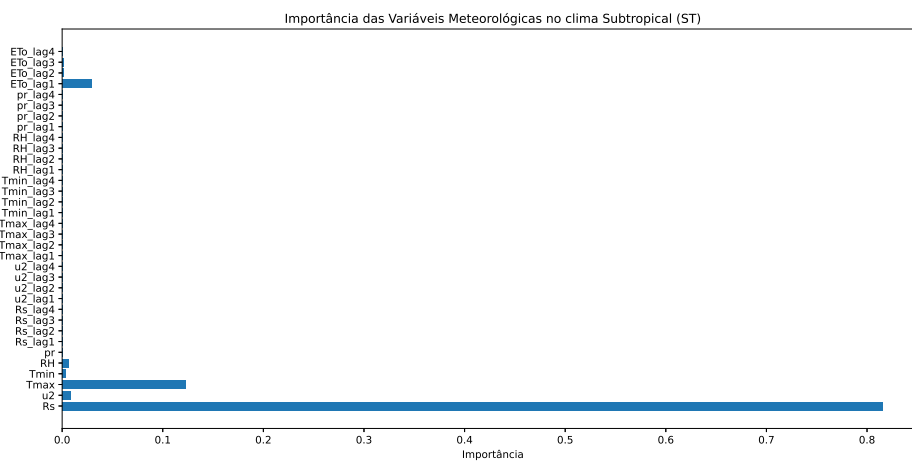


Figura 31 – Importância das variáveis para o modelo RF no clima Subtropical (ST).

APÊNDICE C – Valores SHAP para a CNN

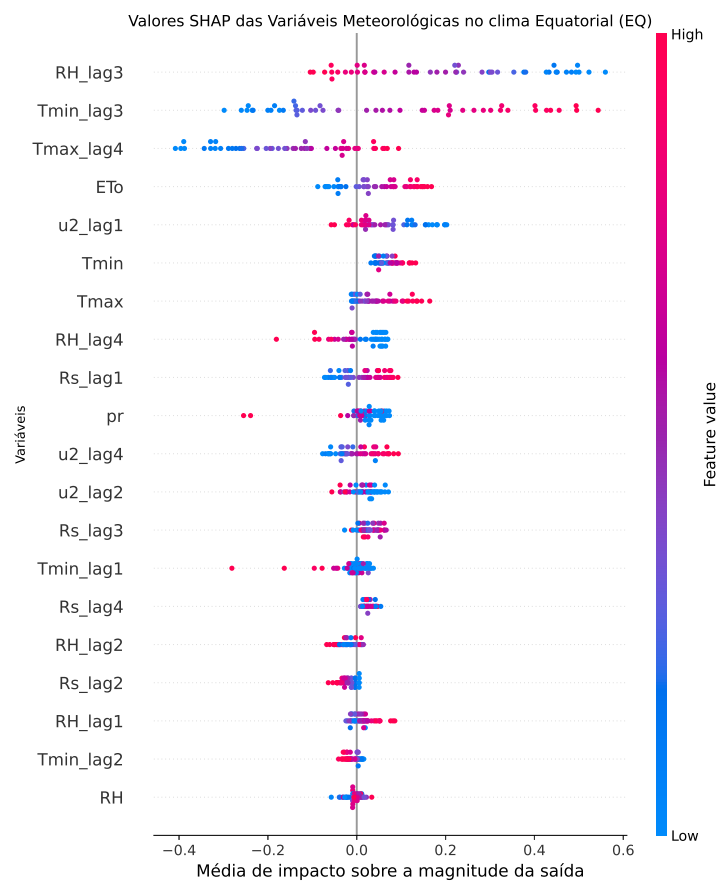


Figura 32 – Impacto das variáveis meteorológicas para o modelo CNN no clima Equatorial (EQ).

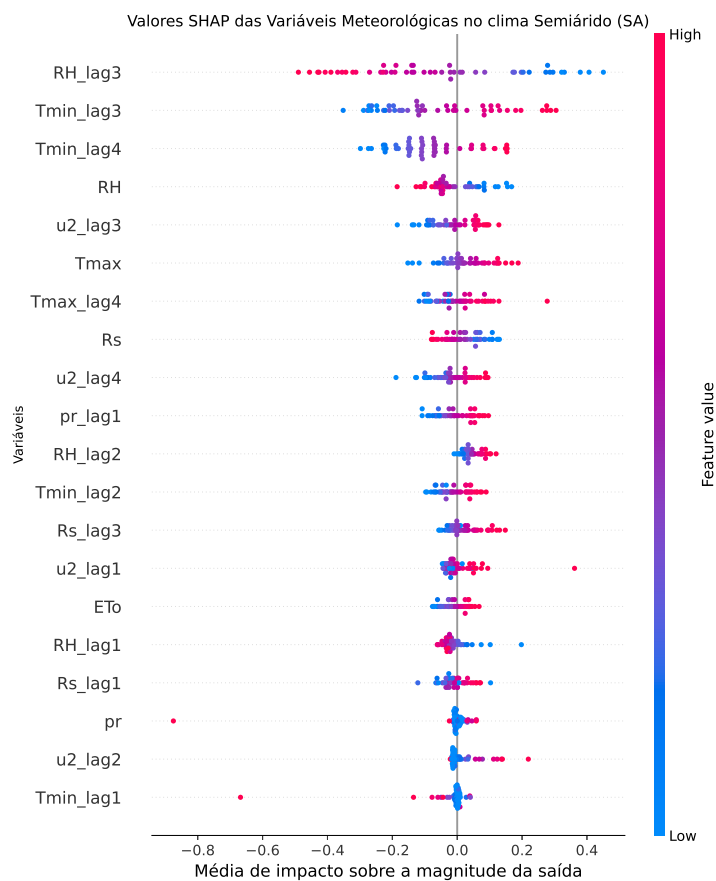


Figura 33 – Impacto das variáveis meteorológicas para o modelo CNN no clima Semiárido (SA).

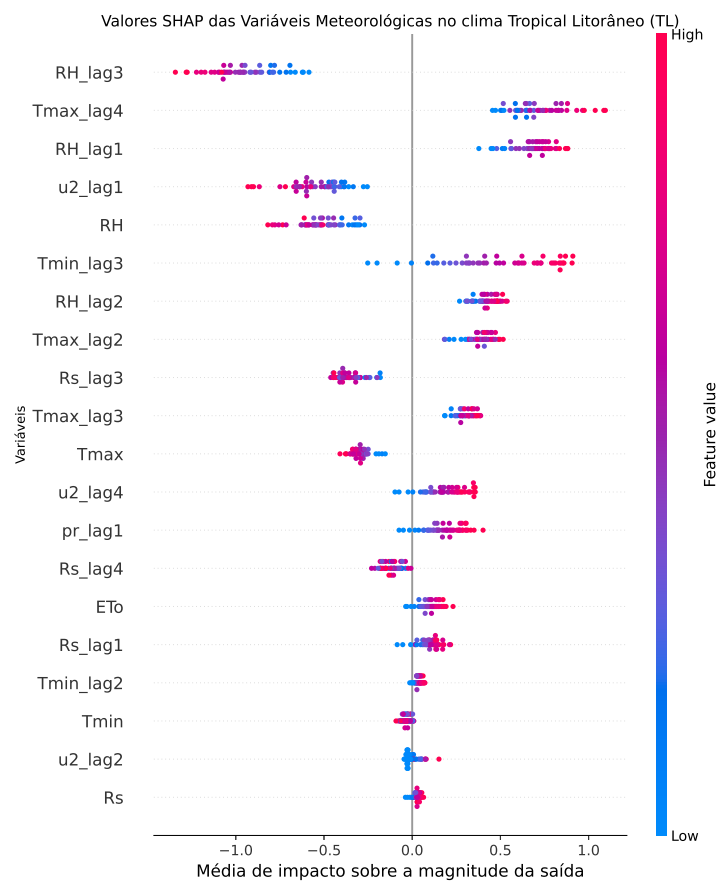


Figura 34 – Impacto das variáveis meteorológicas para o modelo CNN no clima Tropical Litorâneo (TL).

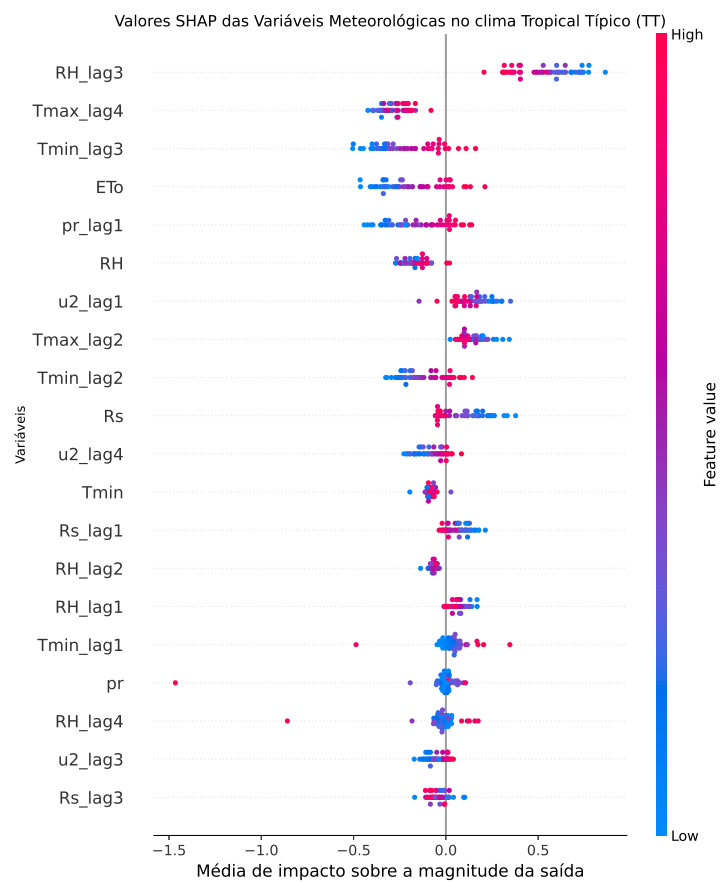


Figura 35 – Impacto das variáveis meteorológicas para o modelo CNN no clima Tropical Típico (TT).

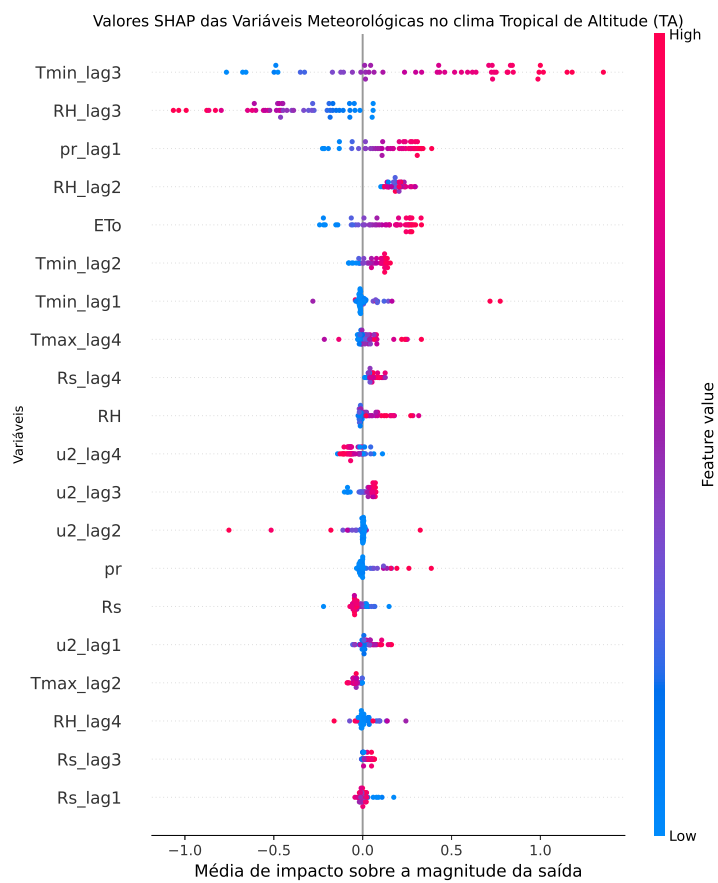


Figura 36 – Impacto das variáveis meteorológicas para o modelo CNN no clima Tropical de Altitude (TA).

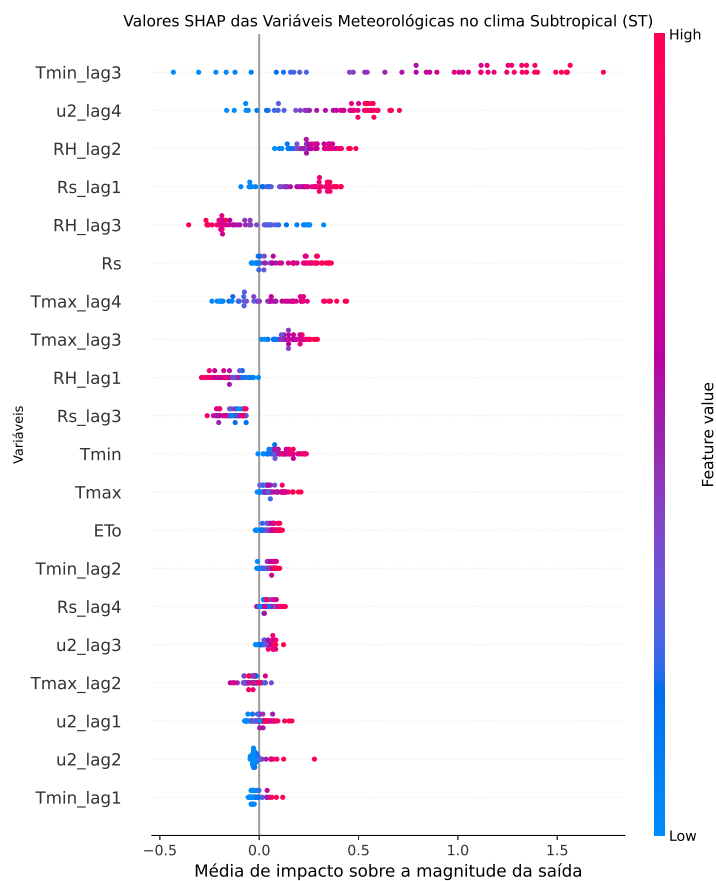


Figura 37 – Impacto das variáveis meteorológicas para o modelo CNN no clima Subtropical (ST).

APÊNDICE D – Valores SHAP para o *Transformer*

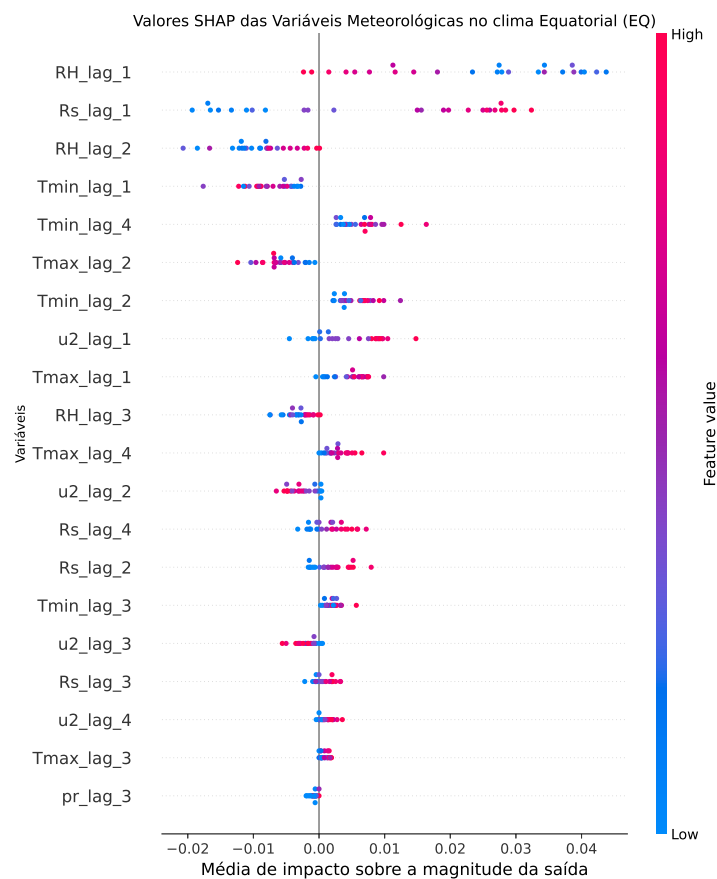


Figura 38 – Impacto das variáveis meteorológicas para o modelo *Transformer* no clima Equatorial (EQ).

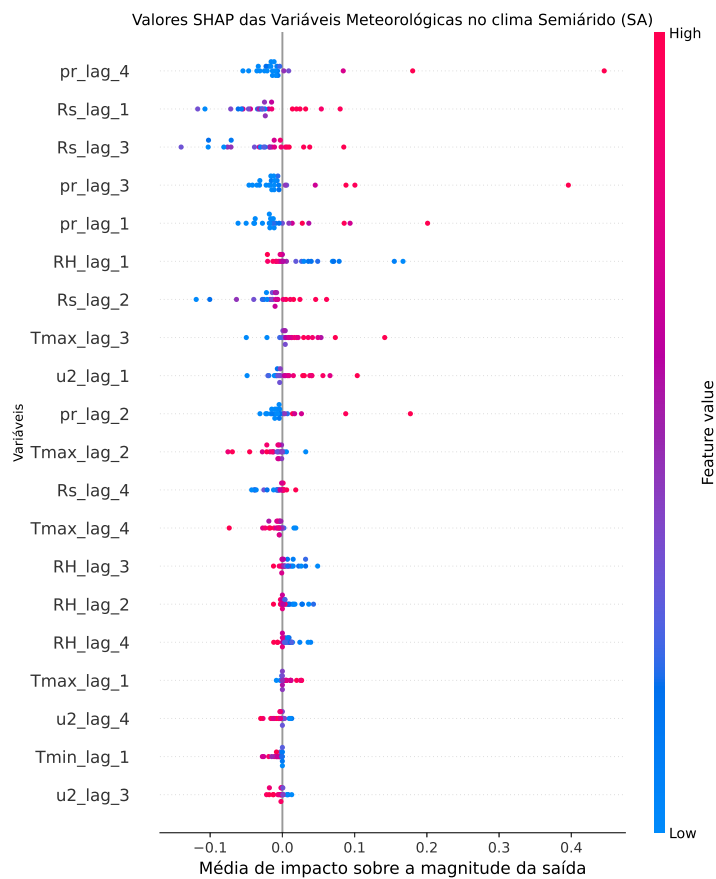


Figura 39 – Impacto das variáveis meteorológicas para o modelo *Transformer* no clima Semiárido (SA).

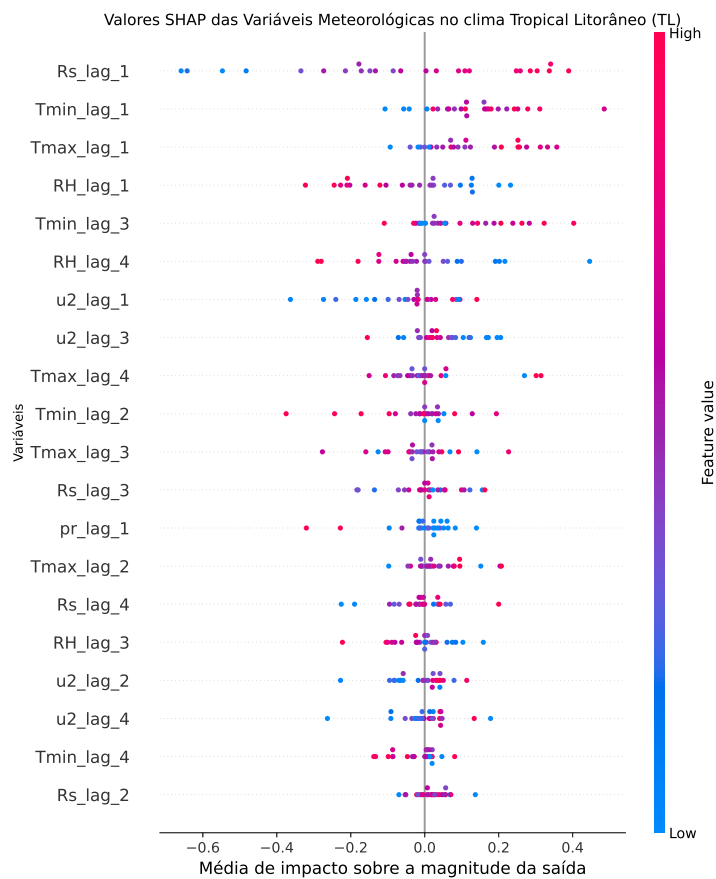


Figura 40 – Impacto das variáveis meteorológicas para o modelo *Transformer* no clima Tropical Litorâneo (TL).

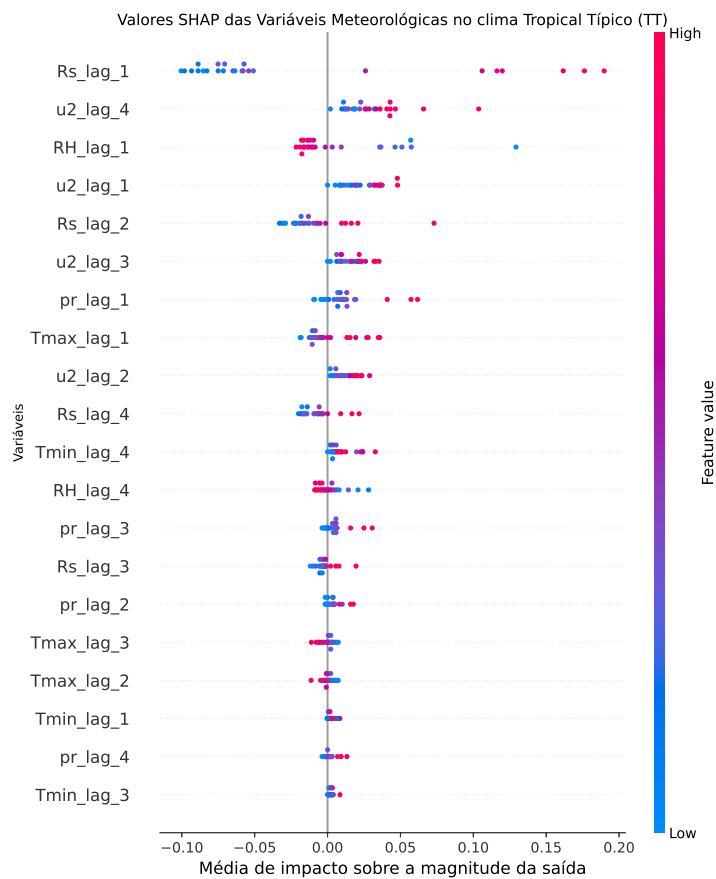


Figura 41 – Impacto das variáveis meteorológicas para o modelo *Transformer* no clima Tropical Típico (TT).

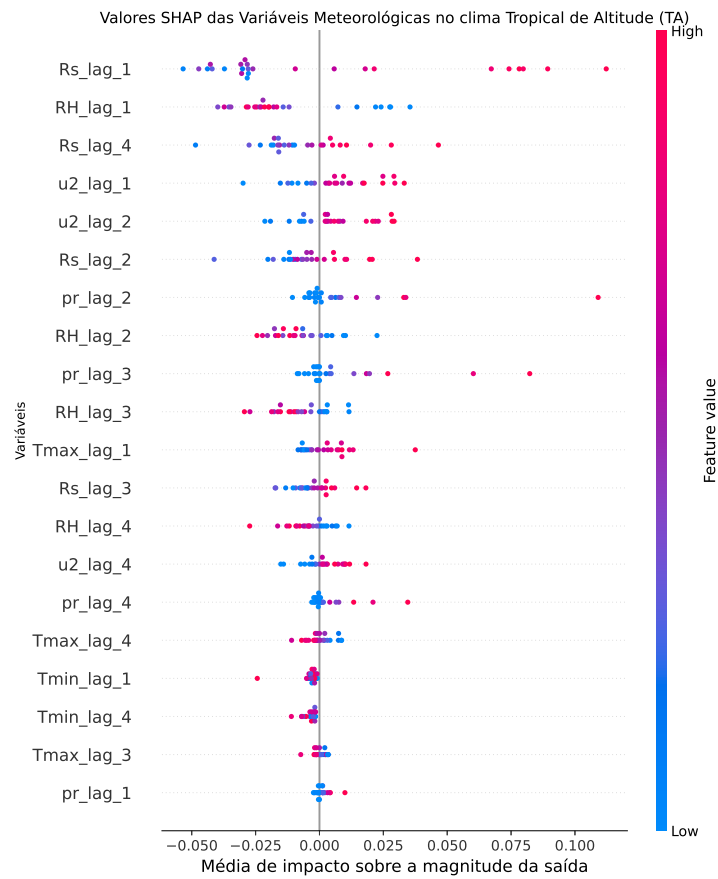


Figura 42 – Impacto das variáveis meteorológicas para o modelo *Transformer* no clima Tropical de Altitude (TA).

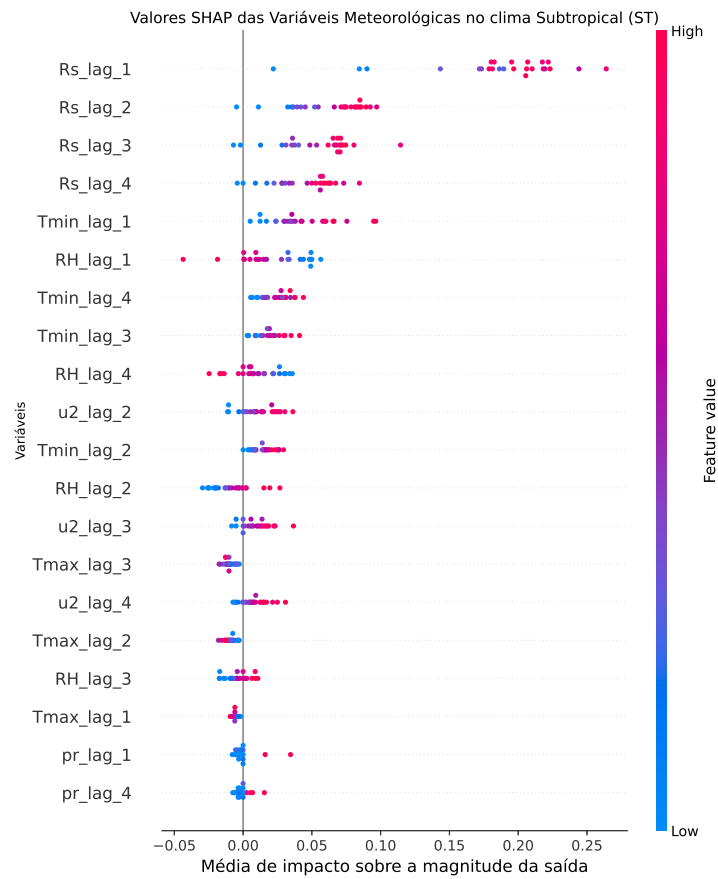


Figura 43 – Impacto das variáveis meteorológicas para o modelo *Transformer* no clima Subtropical (ST).