

Combinação de Modelos Baseados em Transformers para Previsão de Dados Eólicos

Douglas Querino dos Anjos

Pontifícia Universidade Católica do Paraná – PUCPR
R. Imac. Conceição, 1155 - Prado Velho, Curitiba – PR
douglas.anjos@pucpr.edu.br

Gilberto Reynoso-Meza

Pontifícia Universidade Católica do Paraná, PPGEPS
R. Imac. Conceição, 1155 - Prado Velho, Curitiba – PR
g.reynosomeza@pucpr.br

Ramon Gomes da Silva

Pontifícia Universidade Católica do Paraná, PPGEPS
R. Imac. Conceição, 1155 - Prado Velho, Curitiba – PR
ramon.gsilva@pucpr.br

RESUMO

Este trabalho investiga o uso de modelos combinados baseados em Transformers para previsão de séries temporais, avaliando as arquiteturas Informer, Autoformer, Transformer e CALF. Os modelos foram treinados com dados eólicos em duas configurações, com e sem *denoising*, e avaliados em diferentes horizontes de previsão. Duas estratégias de combinação foram exploradas: uma baseada em Otimização Multiobjetivo (MOO), utilizando NSGA-III para balancear MAPE e RMSE, e outra por meio da técnica multicritério SAW, que seleciona dinamicamente o melhor modelo a cada instante com base em janelas móveis. Os resultados mostraram que o modelo SAW obteve desempenho superior ao MOO, beneficiando-se da adaptação ao comportamento dos dados, enquanto o MOO demonstrou superioridade aos modelos base. Além disso, o *denoising* não teve impacto claro na acurácia, mas reduziu significativamente o tempo de treinamento. Conclui-se que abordagens combinadas podem oferecer previsões mais robustas que modelos individuais, especialmente em cenários com variabilidade temporal acentuada.

PALAVRAS-CHAVE. Otimização Multiobjetivo, Decisão Multicritério, Previsão de Séries Temporais.

Combination of Transformer-Based Models for Wind Data Forecasting

ABSTRACT

This study explores the use of combined Transformer-based models for time series forecasting, evaluating the Informer, Autoformer, Transformer, and CALF architectures. models were trained on wind data under two configurations, with and without denoising, and evaluated across different forecasting horizons. Two combination strategies were developed: a Multi-Objective Optimization (MOO) approach using NSGA-III to balance MAPE and RMSE, and another one through the SAW multicriteria technique, which dynamically selects the most suitable model at each time step based on sliding windows. Results show that the SAW model outperformed the MOO-based approach by adapting to the evolving behavior of the data, while the MOO model consistently provided stable performance compared to the based model. Moreover, denoising did not have a clear impact on accuracy but significantly reduced training time. It is concluded that combined approaches can offer more robust forecasts than individual models, especially in scenarios with pronounced temporal variability.

KEYWORDS: Multi-objective Optimization, Multicriteria Decision-making, Time Series Forecasting.

1. Introdução:

Nas últimas décadas, a energia eólica tem se tornado uma das principais alternativas renováveis na matriz energética mundial. Estimativas apontavam que, entre 2020 e 2024, mais de 355 GW de capacidade eólica seriam adicionados globalmente, com o Brasil desempenhando um papel relevante nesse cenário. Até 2021, o país já ocupava a 7ª posição mundial, com mais de 8.300 turbinas distribuídas em 695 parques eólicos [Silva et al. 2022]. No entanto, apesar de sua expansão, a geração eólica enfrenta um desafio fundamental: a sua natureza intermitente, causada pela variabilidade do vento, que compromete a previsibilidade da oferta energética. Diante disso, a previsão da velocidade do vento torna-se estratégica para garantir o equilíbrio do sistema elétrico, o planejamento do despacho de energia e a gestão eficiente de recursos [Silva et al. 2022].

A utilização de modelos combinados aliada a técnicas de decomposição tem se mostrado eficaz na previsão de séries temporais de energia eólica. Abordagens que integram o método VMD (*Variational Mode Decomposition*) para decompor a série em componentes modais, associadas a modelos como o ELM (*Extreme Learning Machine*) resultaram em reduções expressivas de erro, com destaque para melhorias de até 34,76% no RMSE. Esses métodos permitem capturar melhor a variabilidade dos dados e aprimorar a acurácia preditiva, mesmo em contextos com alta complexidade temporal [Ribeiro et al. 2024].

Os modelos baseados em Transformers vêm ganhando destaque na área de previsão de séries temporais, especialmente por sua habilidade em identificar e explorar dependências de longo prazo nos dados. Apesar disso, esses modelos enfrentam dificuldades ao lidar com sequências extensas, principalmente devido ao alto custo computacional e às exigências de memória. Para contornar essas limitações, surgiram novas abordagens, como o Informer, que aprimoram o mecanismo de autoatenção (*self-attention*), tornando-o mais eficiente e escalável [Zhou et al. 2021].

A otimização multiobjetivo (*Multiobjective Optimization* - MOO) tem se consolidado como uma ferramenta poderosa para lidar com problemas complexos que envolvem múltiplos critérios de desempenho, especialmente quando esses critérios entram em conflito. Diferente das abordagens tradicionais que buscam uma única solução ótima, a MOO visa encontrar um conjunto de soluções de compromisso — conhecido como fronteira de Pareto — no qual cada alternativa representa um equilíbrio distinto entre os objetivos considerados [Meza et al. 2014]. Uma das principais vantagens da MOO é a capacidade de identificar e explorar soluções não dominadas, oferecendo ao tomador de decisão um leque de possibilidades alinhadas às suas preferências e restrições específicas, além disso, a MOO fornece uma compreensão mais clara das interações entre objetivos conflitantes, o que contribui para decisões mais estratégicas e fundamentadas [Ribeiro e Meza, 2019].

Diante da crescente complexidade dos sistemas de apoio à decisão, a integração entre a otimização multiobjetivo e os métodos de decisão multicritério (*Multi-Criteria Decision Making* - MCDM) tem se mostrado uma abordagem promissora para lidar com problemas reais que envolvem múltiplas dimensões de avaliação. Enquanto a MOO oferece um conjunto de soluções ótimas de compromisso, conhecidas como soluções de Pareto, a MCDM fornece estruturas formais para auxiliar o decisor na escolha da alternativa mais adequada, considerando preferências, julgamentos qualitativos e critérios muitas vezes conflitantes [Almeida 2013].

[Anjos e Meza, 2024] propuseram uma abordagem híbrida combinando Otimização Multiobjetivo (MOO) com as previsões dos modelos Informer, Transformer e Autoformer. Com base nessa proposta, este estudo amplia a metodologia por meio da incorporação do modelo CALF e da aplicação de critérios de decisão multicritério (MCDM) para a seleção dinâmica do previsor mais adequado.

A estrutura deste artigo está organizada em 5 seções. Na Seção 2, são apresentados os principais conceitos e ferramentas teóricas que fundamentam o estudo. A Seção 3 descreve os dados utilizados e a metodologia aplicada. Os resultados obtidos e a respectiva discussão são

detalhados na Seção 4. A Seção 5 traz as conclusões do trabalho. Pôr fim, apresentam-se os agradecimentos, seguidos das referências utilizadas.

2. Referencial Teórico:

Esse capítulo é destinado a apresentar um referencial teórico dos diversos algoritmos utilizados nesse artigo.

2.1 Série Temporal:

Uma série temporal é composta por um conjunto de observações x_t , onde cada valor é registrado em um instante específico no tempo t . Quando essas observações são coletadas em instantes pertencentes a um conjunto discreto T_0 , como ocorre quando as medições são feitas em intervalos fixos de tempo (por exemplo, a cada hora ou a cada dia), temos uma série temporal em tempo discreto, que é o foco principal de diversas abordagens analíticas. Por outro lado, quando as observações são realizadas continuamente ao longo de um intervalo de tempo, como em $T_0 = [0,1]$, a série é classificada como uma série temporal em tempo contínuo [Brockwell e Davis 2016].

2.1 Transformer:

Estudos recentes sobre a aplicação de modelos Transformer em séries temporais destacam suas contribuições inovadoras e as adaptações realizadas para lidar com dados sequenciais. Esses modelos vêm sendo amplamente empregados em tarefas como previsão, detecção de anomalias e classificação de séries temporais. Entre suas principais vantagens, destaca-se a capacidade de identificar dependências de longo prazo, o que é especialmente relevante em séries com forte presença de sazonalidade e tendência. Originalmente concebido para o domínio do processamento de linguagem natural, o Transformer passou por adaptações específicas para o contexto temporal, incluindo ajustes nos mecanismos de atenção e na codificação posicional. Em relação a métodos clássicos como ARIMA (*Autoregressive Integrated Moving Average*) ou redes recorrentes (*Recurrent Neural Network* - RNNs), os Transformers se sobressaem em previsões de longo horizonte, sobretudo pela sua habilidade de processar informações de múltiplas etapas temporais em paralelo e captar relações complexas entre os dados ao longo do tempo [Wen, et al. 2023].

2.2 Autoformer:

O modelo Transformer convencional é reconhecido por sua eficiência em capturar relações de longo prazo em sequências temporais, graças ao seu mecanismo de atenção. No entanto, sua principal limitação reside na alta demanda computacional, decorrente da complexidade quadrática do cálculo de atenção em relação ao tamanho da sequência. Para contornar esse problema, o Autoformer introduz uma abordagem inovadora baseada na decomposição das séries temporais e em um mecanismo de autocorrelação, o que permite extrair padrões periódicos com maior precisão. Essa estrutura facilita a identificação de subséries recorrentes, promovendo o agrupamento de informações semelhantes e contribuindo para previsões mais eficazes e menos custosas computacionalmente. Estudos experimentais evidenciam que o Autoformer supera diversos modelos anteriores ao alcançar menores valores de erro quadrático médio (*Mean Squared Error* - MSE) e erro absoluto médio (*Mean Absolute Error* - MAE) em diferentes tipos de dados, incluindo séries de consumo energético, tráfego urbano e variáveis meteorológicas [Wu et al. 2022].

Além disso, o Autoformer mostra um desempenho superior ao lidar com séries temporais de alta complexidade, onde há forte presença de padrões não lineares e interações de longo prazo. Enquanto outros modelos Transformers recorrem a estratégias de atenção esparsa para reduzir a carga computacional, o Autoformer se diferencia por combinar desempenho robusto com alta precisão, tornando-se uma escolha promissora em aplicações desafiadoras como planejamento energético, modelagem de tráfego e previsão do tempo [Wu, et al. 2022].

2.3 Informer:

O Informer surge como uma solução eficaz para superar as principais limitações do Transformer tradicional, especialmente em contextos envolvendo séries temporais extensas. Entre os desafios enfrentados pelo modelo original está o elevado custo computacional e de memória, decorrente da complexidade quadrática do mecanismo de autoatenção (*self-attention*). Para resolver esse problema, o Informer introduz o mecanismo de autoatenção esparsa, conhecido como *ProbSparse Attention*, que reduz a complexidade para $O(L \log L)$, onde L representa o tamanho da sequência. Essa abordagem possibilita ganhos expressivos de eficiência tanto em tempo quanto em consumo de memória. Outra inovação importante trazida pelo Informer é a adoção de um decodificador generativo, capaz de realizar previsões de longo prazo em uma única etapa de inferência. Com isso, elimina-se a necessidade de decodificação passo a passo, o que reduz significativamente o acúmulo de erros e torna o processo de previsão mais estável e confiável [Zhou et al. 2021].

2.4 CALF:

Os Modelos de Linguagem de Grande Escala (*Large Language Model* - LLMs) têm ganhado destaque em diversas aplicações, mas enfrentam desafios ao lidar com séries temporais devido à diferença entre dados textuais e numéricos. Para contornar essa limitação, o modelo CALF (*Cross-modal Alignment for Language Models on Forecasting*) propõe uma arquitetura com dois ramos: um textual, baseado em LLMs pré-treinados, e outro temporal, voltado à extração de padrões diretamente das séries temporais. O alinhamento entre essas duas modalidades é realizado por meio de atenção cruzada e projeções via PCA (*Principal Component Analysis*), uma técnica de redução de dimensionalidade que transforma os dados em um novo sistema de coordenadas, preservando as direções de maior variância. Isso permite mapear os dados temporais para o espaço semântico dos *embeddings* de linguagem de forma mais compacta e informativa [Liu et al. 2025].

Além disso, o modelo introduz funções de perda específicas — *Feature Regularization Loss* e *Output Consistency Loss* — para garantir que as representações geradas pelos dois ramos permaneçam coerentes e semanticamente alinhadas. Com o uso de técnicas como LoRA (*Low-Rank Adaptation*) para um treinamento eficiente, o CALF se mostra eficaz em tarefas de previsão de séries temporais de curto e longo prazo, superando modelos existentes tanto em precisão quanto em eficiência, inclusive em cenários com poucos dados disponíveis [Liu et al. 2025].

2.5 EMD e Remoção de ruído (*Denoising*):

Na análise de métodos de pré-processamento voltados à previsão de séries temporais, técnicas como a decomposição empírica de sinais (*Empirical Mode Decomposition* - EMD) e estratégias de remoção de ruído com base em decomposição, têm se destacado por seu impacto na melhoria da acurácia das previsões. A EMD, em particular, é amplamente valorizada por sua flexibilidade em lidar com sinais não estacionários, ao segmentar séries temporais em componentes chamadas de funções de modo intrínseco (*Intrinsic Mode Functions* - IMFs). Essas funções possibilitam uma exploração mais precisa das variações locais de frequência em dados como a velocidade do vento, contribuindo para uma análise mais refinada dos padrões subjacentes [Liu e Chen 2019].

Um ponto importante no uso da EMD está relacionado à primeira função de modo intrínseco (IMF1), que geralmente contém componentes de alta frequência. Essa componente pode incorporar tanto ruídos quanto informações relevantes, mas sua alta volatilidade frequentemente compromete a capacidade preditiva dos modelos. Assim, é comum que o IMF1 seja eliminado do conjunto de dados durante o processo de decomposição, por ser considerado uma fonte de ruído. Essa exclusão contribui para suavizar a série temporal, tornando-a mais regular e, conseqüentemente, mais adequada à modelagem preditiva, com ganhos expressivos em termos de precisão [Liu e Chen 2019].

2.6 Multi-Objective Optimization Desig (MOOD):

O Design de otimização multiobjetivo é composto por três etapas principais. A primeira consiste na formulação do Problema Multiobjetivo (*Multi-Objective Problem* – MOP). Em seguida, aplica-se a Otimização Multiobjetivo propriamente dita, com o objetivo de gerar um conjunto de soluções de compromisso, conhecido como fronteira de Pareto. Por fim, realiza-se a seleção da alternativa mais adequada utilizando métodos de Decisão Multicritério (MCDM), com base nas preferências do decisor. [Ribeiro e Meza 2019].

Formulação do Problema Multiobjetivo (MOP): Essa etapa inicial consiste em modelar corretamente o problema a ser resolvido, expressando os diferentes objetivos a otimizar através de um vetor objetivo conforme a Equação 1, onde m representa o número de metas consideradas.

$$J(\theta) = [J_1(\theta), J_2(\theta), \dots, J_m(\theta)] \quad (1)$$

As variáveis de decisão são reunidas em um vetor decisão conforme Equação 2, com n sendo a quantidade total de variáveis, sujeitas a restrições de desigualdade ou igualdade.

$$\theta = [\theta_1, \theta_2, \dots, \theta_n] \quad (2)$$

As Equações 3 e 4 representam as restrições de desigualdade e igualdade respectivamente, além de limites inferiores e superiores definidos pela Equação 5.

Restrição de desigualdade:

$$g(\theta) \leq 0 \quad (3)$$

Restrição de igualdade:

$$h(\theta) = 0 \quad (4)$$

Limites do espaço de decisão:

$$\underline{\theta}_i \leq \theta_i \leq \overline{\theta}_i \quad (5)$$

A modelagem correta permite a obtenção de soluções de compromisso conhecidas como não dominadas, que compõem a chamada fronteira de Pareto — um conjunto em que nenhuma solução pode ser melhorada em um critério sem comprometer outro [Ribeiro e Meza 2019].

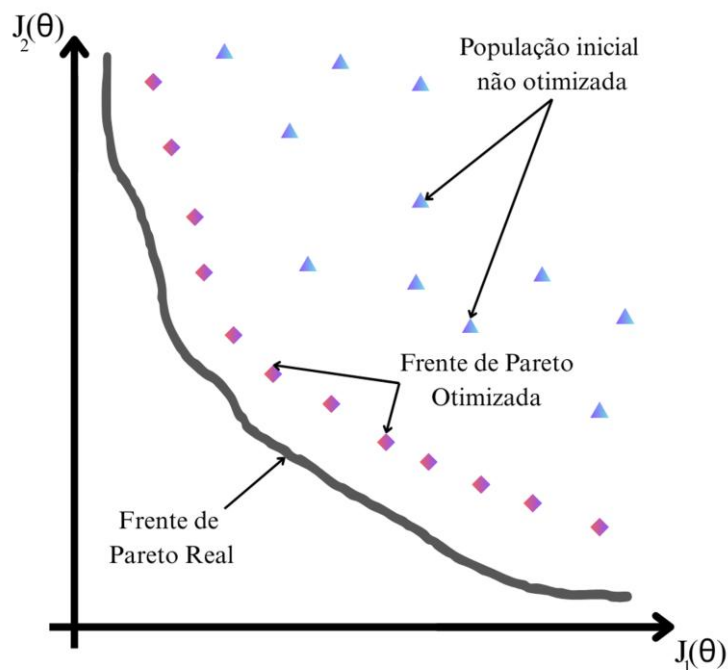
Aplicação da Otimização Multiobjetivo (MOO): Nessa fase, são aplicados algoritmos de busca para encontrar as soluções que compõem a fronteira de Pareto. Entre os métodos mais utilizados destacam-se o NSGA-II (*Non-dominated Sorting Genetic Algorithm II*), o MOEA/D (*Multiobjective Evolutionary Algorithm based on Decomposition*) e o spMODE-II que é uma variante da evolução diferencial para múltiplos objetivos. Esses algoritmos são projetados para gerar um conjunto representativo de soluções, possibilitando ao tomador de decisão visualizar claramente os compromissos entre os objetivos [Ribeiro e Meza 2019]. Em especial, o NSGA-III (*Non-dominated Sorting Genetic Algorithm III*) se destaca para problemas que envolvem muitos objetivos simultaneamente. O NSGA-III utiliza pontos de referência distribuídos uniformemente em um hiperplano, o que assegura uma boa dispersão das soluções no espaço de busca [Deb e Jain 2014].

Apoio à Decisão Multicritério (MCDM): Após obter a fronteira de soluções, o foco se desloca para a escolha da alternativa mais compatível com os interesses e restrições do decisor [Ribeiro e Meza 2019]. Técnicas de MCDM, como o uso de índices de preferência, gráficos com coordenadas paralelas e métodos de decisão multicritério baseados em atributos MADM, auxiliam na análise comparativa entre as opções disponíveis. Um exemplo tradicional de método MADM é o SAW (*Simple Additive Weighting*), amplamente utilizado devido à sua simplicidade e à transparência dos resultados. O método SAW calcula uma soma ponderada dos desempenhos de cada alternativa, considerando a importância relativa de cada critério para a decisão final [Taherdoost 2023].

A Figura 1 exemplifica a dinâmica da otimização multiobjetivo, partindo de uma população inicial não otimizada que, após o processo de busca, se aproxima de uma representação discreta da fronteira de Pareto. Ressalta-se que, embora a fronteira otimizada possa atingir alta qualidade, ela nunca será idêntica à fronteira de Pareto ideal. Além disso, quanto maior a proximidade da solução em relação à fronteira real, maior será o custo computacional associado.

Na Figura 1, os triângulos azuis representam a população inicial aleatória, que ainda não passou pelo processo de otimização. Os losangos vermelhos indicam as soluções otimizadas que compõem a fronteira de Pareto aproximada, obtida a partir da convergência da população inicial. Por fim, a curva em cinza ilustra a fronteira de Pareto ideal, a qual serve como referência teórica de desempenho máximo. Observa-se que os losangos vermelhos tendem a se concentrar nas proximidades da curva cinza, representando a otimização realizada pelo algoritmo ao aproximar as soluções da fronteira de Pareto ideal, ainda que sem alcançá-la de forma exata.

Figura 1 – Processo de otimização da frente de pareto.



Fonte: Adaptada [Meza et al. 2014].

3 Dados e Metodologia:

Este capítulo descreve os dados utilizados e a metodologia adotada para o desenvolvimento e a avaliação dos modelos de previsão.

3.1 Dados:

Os dados utilizados foram fornecidos pela Universidade Tecnológica Federal do Paraná (UTFPR) que foram gerados em um estudo voltado para a construção de um mapa de potencial eólico na Região Metropolitana de Curitiba (RMC). Em particular, foram utilizados os dados coletados pela torre anemométrica instalada na Lapa, uma das cinco torres posicionadas pela UTFPR. Esses dados refletem características de um ambiente urbano e semiurbano [Vieira 2020].

As torres foram posicionadas em locais estratégicos, considerando a topografia e as condições geográficas, para fornecer dados precisos e confiáveis. A Torre da Lapa foi equipada com sensores instalados a 30 metros de altura para medir a velocidade e a direção do vento, além de outros parâmetros meteorológicos como pressão, temperatura e umidade relativa [Vieira 2020].

3.2 Metodologia:

Nesta pesquisa, foram utilizados quatro modelos de previsão baseados em arquiteturas Transformer: Informer, Autoformer, Transformer e CALF. Cada modelo foi treinado com janelas deslizantes de 3 dias para a parte de teste (432 amostras, com intervalo de 10 minutos entre elas), considerando três diferentes horizontes de previsão: 10, 25 e 50 passos à frente. Todos os experimentos foram realizados com e sem aplicação de técnicas de *denoising* por EMD.

Para cada configuração experimental (modelos, horizontes e *denoising*), foram executadas 100 rodadas de treino e previsão, deslocando-se a janela deslizante de 3 dias em uma hora a cada nova rodada. O critério de parada para o treinamento consistiu em atingir uma função de perda (loss) inferior a 0,3 ou completar no máximo 500 epochs, garantindo a convergência dos modelos ou evitando sobreajuste em casos de não convergência.

Além dos modelos individuais, foi desenvolvido um modelo combinado por meio de MOO, utilizando o algoritmo NSGA-III. A combinação foi realizada por média ponderada das previsões dos quatro modelos conforme a Equação 6, com os pesos otimizados visando a minimização das métricas de erro percentual absoluto médio (*Mean Absolute Percentage Error* – MAPE) e erro quadrático médio (*Root Mean Squared Error* – RMSE).

$$Y_{\{Modelo\ Combinado\ MOO\}} = \frac{\omega_1 * Y_I + \omega_2 * Y_A + \omega_3 * Y_T + \omega_4 * Y_C}{(\omega_1 + \omega_2 + \omega_3 + \omega_4)} \quad (6)$$

Na Equação 6, Y_I é o vetor de previsão gerado pelo modelo Informer, Y_A corresponde ao vetor de previsão do Autoformer, Y_T ao vetor de previsão do Transformer e Y_C ao vetor de previsão do modelo CALF.

Os valores ω_1 , ω_2 , ω_3 e ω_4 são os pesos atribuídos às previsões dos quatro modelos. Os limites inferiores e superiores impostos para os pesos na otimização foram definidos na Equação 7 como:

$$0 \leq \omega_1, \omega_2, \omega_3, \omega_4 \leq 1 \quad (7)$$

O problema MOO foi definido como:

Vetor de Decisão: O vetor de decisão do problema θ_p é composto pelos pesos atribuídos a cada modelo, ou seja:

$$\theta_p = [\omega_1, \omega_2, \omega_3, \omega_4] \quad (8)$$

Vetor de Funções Objetivo:

$$J(\theta_p) = [MAPE(\theta_p), RMSE(\theta_p)] \quad (9)$$

Após a obtenção da frente de Pareto no instante t_{-1} , e a posterior seleção do ponto mais adequado por meio de *pseudo weights*, foi realizada uma nova previsão para o tempo t_0 utilizando os pesos definidos. A técnica de *pseudo weights* consiste em atribuir um peso relativo para cada função objetivo, calculado com base na distância normalizada de cada solução em relação ao pior desempenho observado naquela dimensão. Essa medida reflete, portanto, o quanto uma solução se distancia de um cenário indesejado, funcionando como um critério de preferência implícito [Blank 2020].

Logo após, os erros MAPE e RMSE dos quatro modelos e do modelo combinado, foram avaliadas com base em pontuações de 1 a 5 atribuídas separadamente para MAPE e RMSE, onde 1 representa o menor erro e 5 o maior, permitindo a comparação entre os modelos individuais e o modelo combinado de forma padronizada.

As pontuações individuais de RMSE e MAPE foram então somadas, resultando em uma pontuação agregada por modelo. Essa pontuação final serviu como critério para ranquear os

modelos, permitindo identificar, de forma robusta e quantitativa, qual deles apresentou o melhor desempenho geral ao longo dos experimentos.

Adicionalmente, foi implementado um segundo modelo combinado baseado em MCDM, especificamente utilizando o método SAW. Nesse caso, os erros (MAPE e RMSE) produzidos pelos quatro modelos (Informer, Transformer, Autoformer e CALF) foram normalizados e avaliados a partir da soma dos valores em janelas móveis com tamanhos de 1, 3 e 7 passos anteriores. O modelo de janela 1 foi chamado MM1P, de janela 3 de MM3P e o de janela 7 de MM7P. Cada uma dessas janelas representou uma abordagem distinta de decisão. A seleção do modelo para a próxima previsão era feita com base no menor valor de soma dos erros normalizados dentro da janela. O modelo escolhido era então tratado como uma quinta alternativa, permitindo sua avaliação comparativa frente aos demais. A mesma lógica de pontuação (1 a 5 por métrica) e pontuação agregada por modelo foi utilizada também para este novo modelo, possibilitando uma análise do potencial da abordagem baseada em decisão multicritério frente aos modelos utilizados.

A Equação 10 define o vetor T_{MR} , composto pela soma dos valores normalizados de MAPE e RMSE do modelo Transformer, calculados ao longo da janela de análise. O elemento t_{MR0} corresponde à primeira posição do vetor, representando a soma dos valores normalizados de MAPE e RMSE no instante inicial da janela.

$$T_{MR} = t_{MR0} + t_{MR1} + \dots + t_{MRn} \quad (10)$$

De forma análoga, o vetor I_{MR} na Equação 11 representa os valores equivalentes para o modelo Informer. O vetor A_{MR} na Equação 12 está para o modelo Autoformer. O vetor C_{MR} na Equação 13 está para o modelo CALF.

$$I_{MR} = i_{MR0} + i_{MR1} + \dots + i_{MRn} \quad (11)$$

$$A_{MR} = a_{MR0} + a_{MR1} + \dots + a_{MRn} \quad (12)$$

$$C_{MR} = c_{MR0} + c_{MR1} + \dots + c_{MRn} \quad (13)$$

A Tabela 1 apresenta um exemplo de seleção de modelos com base em decisão multicritério. Para cada instante de tempo, os menores valores entre os vetores T_{MR} , I_{MR} , A_{MR} e C_{MR} estão destacados em vermelho, de forma a evidenciar a escolha preferencial em cada caso.

Tabela 1 – Exemplo de Seleção de Modelos por Decisão Multicritério com Destaque dos Melhores Desempenhos

<i>Tempo</i>	T_{MR}	I_{MR}	A_{MR}	C_{MR}	<i>Modelo Escolhido</i>
t_0	t_{MR0}	i_{MR0}	a_{MR0}	c_{MR0}	<i>Transformer</i>
t_1	t_{MR1}	i_{MR1}	a_{MR1}	c_{MR1}	<i>Transformer</i>
t_2	t_{MR2}	i_{MR2}	a_{MR2}	c_{MR2}	<i>Informer</i>
t_3	t_{MR3}	i_{MR3}	a_{MR3}	c_{MR3}	<i>CALF</i>
t_4	t_{MR4}	i_{MR4}	a_{MR4}	c_{MR4}	<i>Autoformer</i>
...	...	...	...	...	...
t_n	t_{MRn}	i_{MRn}	a_{MRn}	c_{MRn}	<i>Informer</i>

Fonte: Autor (2025)

4. Resultados e Discussões:

Em uma avaliação comparativa entre os modelos Transformer, Informer, Autoformer e CALF, observou-se um comportamento instável no desempenho individual de cada arquitetura ao longo das diferentes rodadas de execuções. De modo geral, cada modelo apresentou momentos de destaque, assumindo a liderança no ranking de desempenho em determinadas janelas; no entanto, essa superioridade se mostrou temporária. Frequentemente, o modelo que inicialmente se destacava acabava sendo superado nas rodadas seguintes, indicando uma oscilação na performance. Essa variação contínua na posição dos modelos evidencia uma instabilidade inerente, dificultando a escolha de um único modelo como melhor solução em todos os contextos analisados.

Os resultados obtidos com o modelo combinado por meio de MOO demonstraram a eficácia dessa abordagem na criação de soluções mais estáveis e equilibradas em relação aos modelos individuais. Embora o modelo otimizado raramente tenha ocupado a primeira posição isolada nos rankings de desempenho por métrica (MAPE e RMSE), ele apresentou um comportamento consistente. Essa estabilidade é um indicativo da capacidade da MOO de combinar modelos trazendo robustez e estabilidade, mesmo sem alcançar o melhor desempenho absoluto em todos os cenários, ainda mantêm um nível de erro mais controlado e confiável ao longo das diferentes janelas e horizontes de previsão. A análise do ranking final, Tabela 2 com base na soma das pontuações atribuídas aos modelos, mostra que o modelo combinado via MOO alcançou a primeira colocação geral em quase todos os cenários. Na única situação em que não obteve essa posição, a diferença percentual em relação ao primeiro colocado foi mínima, o que permite, inclusive, considerar um empate técnico.

A partir dos resultados da Tabela 2, é visível que os modelos combinados MM1P, MM3P e MM7P ficaram em primeiro lugar no ranking geral, mostrando o melhor desempenho em todos os cenários avaliados. Em segundo lugar, ficou o Informer, com resultados consistentes e próximos ao modelo combinado. Na sequência, os modelos Transformer e CALF ficaram com desempenhos intermediários, alternando entre a terceira e a quarta posição. Já o Autoformer teve o pior desempenho geral, aparecendo com frequência na última posição.

Tabela 2 – Ranque dos modelos combinados.

Modelo combinado MOO							Modelo combinado Multicritério Janela de 1 passo - MM1P						
Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MOO	Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MM1P
SIM	10	3	2	5	4	1	SIM	10	3	2	5	4	1
NÃO	10	3	1	5	4	2	NÃO	10	4	2	5	3	1
SIM	25	3	2	5	4	1	SIM	25	3	2	5	4	1
NÃO	25	4	2	5	3	1	NÃO	25	4	3	5	2	1
SIM	50	3	2	4	5	1	SIM	50	3	2	4	5	1
NÃO	50	3	2	5	4	1	NÃO	50	4	2	5	3	1
Modelo combinado Multicritério Janela de 3 passos - MM3P							Modelo combinado Multicritério Janela de 7 passos - MM7P						
Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MM3P	Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MM7P
SIM	10	3	2	5	4	1	SIM	10	3	2	5	4	1
NÃO	10	4	2	5	3	1	NÃO	10	4	2	5	3	1
SIM	25	5	2	4	3	1	SIM	25	5	2	4	3	1
NÃO	25	4	2	5	3	1	NÃO	25	4	3	5	2	1
SIM	50	4	2	3	5	1	SIM	50	4	2	3	5	1
NÃO	50	4	2	5	3	1	NÃO	50	4	2	5	3	1

Fonte: Autor (2025)

Em contraste com o modelo baseado em otimização multiobjetivo, o modelo construído com base no método SAW apresentou um desempenho significativamente superior. A Tabela 3 mostra que, ao analisar a diferença percentual entre o primeiro e o segundo colocado em cada abordagem, os modelos baseados em decisão multicritério apresentaram, consistentemente, diferenças maiores do que aquelas observadas no modelo MOO. Isso evidencia uma vantagem clara dos modelos multicritério utilizados, em termos de precisão e capacidade de discriminar a melhor alternativa. Essa diferença ressalta o ganho da abordagem multicritério nesse cenário, que se propõe não a combinar modelos, mas a escolher adaptativamente aquele que apresenta o melhor desempenho local em função das condições recentes da série temporal. Essa estratégia, que pode ser descrita como “surfar a crista da onda”, busca continuamente identificar e utilizar o modelo mais promissor para cada instante de previsão. Como resultado, o método SAW consegue se ajustar rapidamente às mudanças no comportamento dos dados, o que se reflete em uma performance mais eficaz quando comparado a abordagem combinada com a MOO.

A aplicação do método SAW em janelas deslizantes de decisão revelou que a janela de 1 passo produziu os melhores resultados, seguida pela janela de 3 passos e, por fim, pela de 7 passos. Apesar da redução gradual na performance com o aumento do tamanho da janela, o uso de janelas como estratégia ainda apresentou vantagens claras sobre abordagens baseadas apenas na escolha do melhor modelo anterior.

Tabela 3 – Diferença percentual entre primeiros colocados e demais posições.

Modelo combinado MOO							Modelo combinado Multicritério Janela de 1 passo - MM1P						
Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MOO	Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MM1P
SIM	10	9.0%	2.5%	22.4%	11.9%	0.0%	SIM	10	19.5%	16.9%	32.5%	24.6%	0.0%
NÃO	10	4.4%	0.0%	19.5%	8.2%	0.5%	NÃO	10	18.6%	14.0%	27.2%	15.7%	0.0%
SIM	25	12.0%	0.5%	17.0%	13.9%	0.0%	SIM	25	35.8%	27.3%	38.0%	35.9%	0.0%
NÃO	25	11.1%	3.7%	16.9%	9.3%	0.0%	NÃO	25	37.7%	34.6%	42.4%	34.5%	0.0%
SIM	50	12.2%	5.2%	17.8%	22.2%	0.0%	SIM	50	40.7%	36.0%	42.7%	46.3%	0.0%
NÃO	50	8.3%	3.3%	18.7%	14.1%	0.0%	NÃO	50	37.5%	34.2%	44.3%	36.9%	0.0%
Modelo combinado Multicritério Janela de 3 passos - MM3P							Modelo combinado Multicritério Janela de 7 passos - MM7P						
Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MM3P	Denoising	Janela de previsão	Transf.	Inf.	Auto.	CALF	Combinado MM7P
SIM	10	22.2%	18.2%	35.1%	28.2%	0.0%	SIM	10	8.7%	5.9%	23.3%	19.3%	0.0%
NÃO	10	17.7%	10.6%	25.7%	15.1%	0.0%	NÃO	10	20.9%	12.4%	26.1%	17.3%	0.0%
SIM	25	34.3%	24.2%	34.1%	33.4%	0.0%	SIM	25	31.8%	19.4%	30.7%	25.9%	0.0%
NÃO	25	33.2%	26.7%	35.8%	28.4%	0.0%	NÃO	25	28.6%	19.8%	29.8%	17.6%	0.0%
SIM	50	40.3%	33.6%	38.4%	45.8%	0.0%	SIM	50	33.4%	25.2%	28.6%	38.4%	0.0%
NÃO	50	36.7%	32.9%	42.0%	35.6%	0.0%	NÃO	50	28.8%	18.5%	30.3%	23.2%	0.0%

Fonte: Autor (2025)

A motivação por trás da utilização de janelas, em vez de decisões baseadas em resultados pontuais, foi justamente a busca por maior estabilidade ao longo do tempo. A hipótese era de que decisões tomadas com base em um histórico mais amplo, mesmo que com menor precisão imediata, tenderiam a gerar previsões mais robustas e menos suscetíveis às flutuações momentâneas no desempenho dos modelos. Os resultados mostram que, no caso avaliado, modelos com janelas menores se mostraram mais eficazes, pois foram capazes de se adaptar com maior agilidade às rápidas mudanças no desempenho relativo entre os modelos individuais.

Tabela 4 – Tempo de treino médio em segundo dos modelos, comparativo entre modelos com e sem denoising.

Tempo médio em segundos gasto nos treinos dos modelos					
Denoising	Janela de previsão	Trans.	Inf.	Auto.	CALF
SIM	10	56	59	93.5	83
NÃO	10	62	63.5	106.5	109.5
Diferença percentual		9.7%	7.1%	12.2%	24.2%
SIM	25	59	67	91.5	120.5
NÃO	25	66.5	80	111.5	148.5
Diferença percentual		11.3%	16.3%	17.9%	18.9%
SIM	50	55	78	99	115
NÃO	50	71	92	116	151
Diferença percentual		22.5%	15.2%	14.7%	23.8%

Fonte: Autor (2025)

A análise dos resultados apresentados nas Tabelas 2 e 3, que mostram respectivamente as pontuações obtidas pelos modelos combinados e a variação percentual entre eles, indica que não há uma tendência clara quanto ao impacto da aplicação do *denoising* nos erros de previsão. Em

muitos cenários, os modelos com e sem *denoising* apresentaram desempenhos semelhantes, não sendo possível identificar uma influência consistente, seja positiva ou negativa, sobre as métricas MAPE e RMSE. No entanto, ao se observar os dados da Tabela 4, que detalha o tempo médio de treinamento dos modelos, torna-se evidente uma vantagem significativa da aplicação do *denoising* por EMD. Em todos os casos avaliados, a utilização do *denoising* resultou em redução do tempo de treinamento, o que sugere que a remoção de componentes ruidosos dos dados contribui para uma convergência mais rápida dos modelos. Portanto, ainda que o efeito sobre a acurácia não seja conclusivo, o uso do *denoising* se mostra vantajoso do ponto de vista computacional, especialmente em contextos em que eficiência de tempo é um fator relevante.

5. Conclusões:

Este trabalho teve como objetivo avaliar a eficácia de modelos combinados em tarefas de previsão de séries temporais, utilizando quatro arquiteturas baseadas em Transformers — Informer, Autoformer, Transformer e CALF — aplicadas com e sem técnicas de *denoising*. A partir desses modelos, foram construídas duas estratégias de combinação: uma baseada em MOO, com pesos ajustados para minimizar simultaneamente MAPE e RMSE, e outra baseada no método SAW, cuja lógica se fundamenta na escolha adaptativa do modelo mais adequado em cada instante de previsão.

Os resultados mostraram que modelos combinados podem superar os modelos puros, especialmente em contextos em que diferentes arquiteturas apresentam desempenhos complementares. O modelo MOO, apesar de raramente ocupar a primeira colocação em termos absolutos, demonstrou alta estabilidade ao evitar posições inferiores no ranking, sendo consistentemente competitivo ao longo das previsões. Essa característica o torna uma escolha segura, capaz de suavizar as flutuações associadas ao uso de modelos individuais.

Por outro lado, os modelos baseados no método SAW se destacaram por sua capacidade adaptativa, alcançando desempenho superior ao modelo MOO na maioria dos cenários. Essa vantagem se deve ao seu comportamento dinâmico, que "surfa a crista da onda" ao selecionar, a cada momento, o modelo que apresenta a menor soma dos erros normalizados recentes. Ainda que isso não comprove de forma definitiva a superioridade do SAW sobre o MOO em todos os contextos, os resultados observados neste cenário específico demonstram o potencial do uso de técnicas multicritério como alternativa viável e eficaz em ambientes com variação sequencial de desempenho entre modelos.

Adicionalmente, observou-se que a aplicação de *denoising* via EMD não apresentou impacto claro sobre a acurácia dos modelos, mas trouxe ganhos relevantes em termos de eficiência computacional, com redução significativa nos tempos de treinamento. Esses achados reforçam a importância de explorar estratégias híbridas na previsão de séries temporais, aproveitando a complementaridade entre modelos e otimizando recursos computacionais. Trabalhos futuros podem investigar a integração de ambas as abordagens (MOO e SAW), explorando cenários com maior diversidade de comportamento temporal e ampliando a robustez das decisões automatizadas em aplicações reais.

6. Agradecimentos:

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.157343/2025-00.

Este Estudo foi financiado em partes pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) através do projeto PQ-2/310195/2022-5, Universal/408164/2021-2 & PDE [200014/2025-0].

Os autores agradecem à Companhia Paranaense de Energia Elétrica (COPEL) pelo suporte e financiamento na realização do projeto de pesquisa PD ANEEL 2866-0464/2017 (Metodologia para Análise, Monitoramento e Gerenciamento da Geração Distribuída por Fontes Incentivadas), e à Companhia de Saneamento do Paraná (Sanepar) pela cessão dos espaços e apoio no processo de instalação das torres.

7. Referências:

- Almeida, A. T. de. (2013). Processo de decisão nas organizações. Atlas.
- Alves Ribeiro, V. H., & Reynoso-Meza, G. (2019). A holistic multi-objective optimization design procedure for ensemble member generation and selection. *Applied Soft Computing*, 83, 105664.
- Anjos, D. Q. dos, & Reynoso-Meza, G. (2024). Modelo Combinado Utilizando Otimização Multi-Objetivo Para Previsão de Dados Eólicos. *Anais do 14 Congresso Brasileiro de Engenharia de Produção*, 1, 1–11.
- Blank, J. (2020). Part III: Multi-Criteria Decision Making. *Pymoo: Multi-Objective Optimization in Python*. https://pymoo.org/getting_started/part_3.html#Part-III:-Multi-Criteria-Decision-Making Acessado: 2025-04-01.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed). Springer.
- da Silva, R. G., Moreno, S. R., Ribeiro, M. H. D. M., Larcher, J. H. K., Mariani, V. C., & Coelho, L. dos S. (2022). Multi-step short-term wind speed forecasting based on multi-stage decomposition coupled with stacking-ensemble learning approach. *International Journal of Electrical Power & Energy Systems*, 143, 108504.
- Deb, K., & Jain, H. (2014). An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Transactions on Evolutionary Computation*, 18(4), 577–601.
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487.
- Kariniotakis, G. (2017). *Renewable energy forecasting: From models to applications*. Woodhead publishing.
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679.
- Liu, H., & Chen, C. (2019). Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Applied Energy*, 249, 392–408.
- Liu, P., Guo, H., Dai, T., Li, N., Bao, J., Ren, X., Jiang, Y., & Xia, S.-T. (2024). CALF: Aligning LLMs for Time Series Forecasting via Cross-modal Fine-Tuning (Versão 3). *arXiv*.
- Reynoso-Meza, G., Blasco, X., Sanchis, J., & Martínez, M. (2014). Controller tuning using evolutionary multi-objective optimisation: Current trends and applications. *Control Engineering Practice*, 28, 58–73.
- Ribeiro, M. H. D. M., Da Silva, R. G., Moreno, S. R., Canton, C., Larcher, J. H. K., Stefenon, S. F., Mariani, V. C., & Coelho, L. D. S. (2024). Variational mode decomposition and bagging extreme learning machine with multi-objective optimization for wind power forecasting. *Applied Intelligence*, 54(4), 3119–3134.
- Suresh, A., & Deb, K. (2024). On a Better Understanding of Unique Identifiers of Pareto Solutions for Multi-criterion Optimization, Visualization, and Decision-making. *2024 IEEE Congress on Evolutionary Computation (CEC)*, 1–8.
- Taherdoost, H. (2023). Analysis of Simple Additive Weighting Method (SAW) as a MultiAttribute Decision-Making Technique: A Step-by-Step Guide. *Journal of Management Science & Engineering Research*, 6(1), 21–24.
- Vieira, B. J. (2020). Metodologia para elaboração de mapa de potencial eólico de regiões metropolitanas [masterThesis, Universidade Tecnológica Federal do Paraná].
- Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2023). Transformers in Time Series: A Survey. *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 6778–6786.
- Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting (Versão 5).
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *The thirty-fifth AAAI conference on artificial intelligence, AAAI 2021, virtual conference*, 35, Artigo 12.