

Previsão de Potência Solar Utilizando Métodos de Simulação de Monte Carlo, *K-Means* e SVM: Um Estudo de Caso na Região Metropolitana de Vitória/ES

Solar Power Forecasting Using Monte Carlo Simulation, K-Means, and SVM Methods: A Case Study in the Metropolitan Region of Vitória/ES

Caroline Tedesco Santos Mestre em Energia. Universidade Federal do Espírito Santo (UFES) – Brasil. tedescocarol@hotmail.com
<http://orcid.org/0000-0002-1453-2847>

Ivaniel Fôro Maia Especialista em Gestão de Emergências em Saúde Pública. Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER) – Brasil. ivanielforo@gmail.com
<http://orcid.org/0000-0002-1508-9074>

Augusto César Rueda Medina Doutor em Engenharia Elétrica. Universidade Federal do Espírito Santo (UFES) – Brasil. augusto.rueda@ufes.br
<http://orcid.org/0000-0002-4291-3153>

Jussara Farias Fardin Doutora em Engenharia Elétrica. Universidade Federal do Espírito Santo (UFES) – Brasil. jussara@ele.ufes.br
<http://orcid.org/0000-0003-4785-556X>

RESUMO

O estudo apresenta uma abordagem para prever a potência gerada por sistemas fotovoltaicos considerando a variabilidade da cobertura de nuvens. A partir da combinação da Simulação de Monte Carlo, do método de agrupamento *K-Means* e de previsões com a ferramenta SVM, foram analisados dados de radiação solar da região metropolitana de Vitória/ES. Os modelos mostraram-se eficazes, produzindo previsões precisas com baixos índices de erro, mesmo sob diferentes condições climáticas. Os resultados indicam que o uso de dados reais e a segmentação adequada dos cenários meteorológicos permitem compreender como a variação da radiação solar afeta a geração de energia. Essa abordagem contribui para o planejamento e a otimização da energia solar e evidencia a importância de estudos futuros que integrem novas técnicas de aprendizado de máquina para aperfeiçoar as previsões.

Palavras-chave: radiação solar; cenários meteorológicos; cobertura de nuvens.

ABSTRACT

This study presents an approach to forecasting the power generated by photovoltaic systems while considering cloud cover variability. By combining Monte Carlo Simulation, the *K-Means* clustering method, and predictions using the SVM tool, solar radiation data from the metropolitan region of Vitória-ES were analyzed. The models proved to be effective, delivering accurate forecasts with low error rates even under varying weather conditions. The results indicate that using real data and appropriately segmenting weather scenarios allows for a better understanding of how solar radiation variation affects power generation. This approach supports solar energy planning and

optimization and highlights the need for future studies that integrate new machine learning techniques to further enhance forecasting accuracy.

Keywords: solar radiation; weather scenarios; cloud cover.

Recebido em 16/04/2025. Aprovado em 02/09/2025. Avaliado pelo sistema *double blind peer review*. Publicado conforme normas da ABNT.

<https://doi.org/10.22279/navus.v16.2116>

1 INTRODUÇÃO

Nos últimos anos, os sistemas de distribuição de energia elétrica têm registrado um aumento expressivo na adoção de fontes de energia distribuída (DERs). Esse avanço é impulsionado pela busca por alternativas energéticas mais sustentáveis e menos poluentes, diante da comprovada insustentabilidade do modelo tradicional baseado em combustíveis fósseis. As DERs, especialmente aquelas provenientes de fontes renováveis, têm se mostrado eficientes na redução das emissões de dióxido de carbono (CO₂) e na mitigação dos impactos ambientais associados à geração centralizada de eletricidade (Alqurashi, 2021). A manutenção da dependência global de combustíveis fósseis é considerada um risco, uma vez que, sem políticas mitigadoras, estima-se que 80% da eletricidade mundial ainda será gerada por essas fontes em 2030, contribuindo para o agravamento das mudanças climáticas e para ameaças à saúde pública (Akorede *et al.*, 2010). Pesquisas indicam que a substituição gradual da geração centralizada por tecnologias renováveis distribuídas pode reduzir significativamente as emissões de gases de efeito estufa e aumentar a resiliência energética (Sharma *et al.*, 2016). Ademais, a crescente aceitação pública dessas tecnologias reforça a tendência global de transição para um sistema energético mais limpo e resiliente (Gardner *et al.*, 2008).

Como alternativa promissora aos combustíveis fósseis, a energia solar vem se destacando em pesquisas e no mercado por suas inúmeras vantagens. A geração fotovoltaica não emite poluentes, exige pouco espaço e apresenta baixos custos de manutenção, configurando-se como uma das opções mais viáveis e sustentáveis para o futuro energético (Ferraz; Ferraz & Rueda, 2024). O elevado potencial de irradiação solar no Brasil torna essa fonte uma alternativa estratégica na busca por um desenvolvimento nacional mais limpo e sustentável.

Para que a energia solar assuma o papel de principal fonte de geração elétrica no país e substitua gradualmente as fontes não renováveis, é imprescindível compreender o comportamento da irradiação solar ao longo do tempo. Isso inclui analisar fatores externos ao controle humano, como a presença de nuvens e as variações climáticas, que afetam diretamente a captação de energia solar. Esse conhecimento é essencial para o planejamento eficiente do uso da fonte renovável e para assegurar a confiabilidade do fornecimento de energia elétrica em escala nacional.

A radiação proveniente do Sol é responsável por abastecer a atmosfera da Terra e sustentar os processos vitais, sendo a principal fonte de energia do planeta (Paralič, 2020). As nuvens exercem papel determinante no sistema climático, pois interferem no balanço de radiação solar (Yuan *et al.*, 2016). Suas propriedades ópticas e o grau de cobertura (Zhang; Zhang, 2017) fazem com que a radiação varie rapidamente ao longo do dia, influenciando diretamente a geração de energia fotovoltaica. A presença e a densidade das nuvens afetam a quantidade de radiação disponível, impactando a eficiência e a previsibilidade da geração elétrica.

Este artigo tem como objetivo analisar a correlação entre a formação de diferentes tipos de nuvens e as variações na radiação solar, a fim de compreender o impacto dessas variáveis na geração de energia elétrica a partir da fonte solar. A investigação concentra-se na previsão da irradiação solar, considerando os efeitos da presença e da densidade das nuvens. Para isso, propõe-se o desenvolvimento de um método de previsão baseado em técnicas de *deep learning*, visando obter estimativas mais precisas e representativas da realidade. Essa abordagem é viável porque o banco de dados utilizado é

composto por medições reais da região metropolitana de Vitória, no Espírito Santo, um local relevante tanto por seu contexto climático quanto por seu elevado potencial para geração de energia solar.

2 ESTADO DA ARTE

A crescente integração de fontes renováveis de energia, especialmente a solar, aos sistemas de distribuição elétrica tem impulsionado o desenvolvimento de métodos para previsão precisa da irradiância solar. Essa previsão é fundamental para assegurar a confiabilidade e a eficiência da geração de energia, uma vez que a radiação solar é naturalmente variável e influenciada por fatores meteorológicos, como a cobertura de nuvens.

De acordo com Reikard (2009), métodos tradicionais de séries temporais já apresentavam bom desempenho na previsão da radiação solar em horizontes de curto prazo, sobretudo quando aplicados a dados com alta resolução temporal. Entretanto, o advento de técnicas baseadas em inteligência artificial, como as redes neurais artificiais (ANNs), trouxe avanços expressivos em termos de acurácia. Nesse contexto, Di Piazza *et al.* (2019) demonstraram que modelos de redes neurais profundas superam os métodos estatísticos clássicos, especialmente em aplicações que envolvem séries temporais de demanda e geração de energia solar, evidenciando o potencial dessas abordagens para o setor energético.

Nos últimos anos, modelos baseados em *deep learning* têm se destacado na previsão da irradiância solar, devido à sua capacidade de lidar com dados intermitentes, instáveis e não estacionários. Conforme revisado por Kumari e Toshniwal (2021), técnicas como redes neurais convolucionais (CNN), redes neurais recorrentes do tipo LSTM, redes de crença profunda (DBN) e redes de estado de eco (ESN) apresentam desempenho superior aos métodos estatísticos e de *machine learning* tradicionais. Além disso, modelos híbridos têm se mostrado particularmente promissores. Um exemplo é o modelo CEEMDAN-CNN-LSTM, proposto por Gao *et al.* (2020), que integra a decomposição empírica adaptativa em modo completo (CEEMDAN) à arquitetura CNN-LSTM para melhorar a extração de características dos dados históricos e prever a irradiância solar horária. Os experimentos realizados com diferentes bases de dados reais, representando variados tipos climáticos, mostraram que esse modelo não apenas supera de forma significativa técnicas isoladas, mas também mantém desempenho estável, alcançando um erro quadrático médio (RMSE) de 38,49 W/m². Esses avanços confirmam a eficácia dos modelos híbridos de *deep learning* e reforçam sua importância para o planejamento e a operação eficiente de sistemas fotovoltaicos em redes inteligentes.

No que se refere à influência das nuvens na radiação solar, Wang *et al.* (2021) conduziram uma investigação abrangente utilizando dados do satélite CloudSat para analisar a contribuição radiativa de diferentes fases de nuvens (gelo, mista e água super-resfriada) na região do Planalto Tibetano. Os resultados indicaram que, além de apresentarem variações sazonais, essas fases impactam de maneira distinta a radiação incidente. As nuvens de gelo foram predominantes no inverno, enquanto nuvens mistas e super-resfriadas se destacaram no verão, afetando de forma significativa o balanço energético atmosférico. De forma complementar, Zhang *et al.* (2022) analisaram a interação entre aerossóis, nuvens e radiação solar por meio de simulações com o modelo WRF-Chem, evidenciando que propriedades radiativas das nuvens – como fração de cobertura e tipo – desempenham papel central na variabilidade da energia solar em diferentes regiões da China. Dentre os efeitos observados, o efeito

radiativo das nuvens (CRE) mostrou-se dominante em várias localidades, especialmente no Delta do Rio das Pérolas, onde contribuiu com até 78,4% para o aumento da energia solar disponível. Esses achados reforçam que propriedades ópticas das nuvens, como espessura, fase e distribuição espacial, são determinantes para a precisão na previsão da irradiância solar e, consequentemente, para o desempenho de sistemas fotovoltaicos.

No contexto nacional, estudos também têm avançado no uso de dados reais para previsão da potência solar. Um exemplo relevante é apresentado por Ramos *et al.* (2022), que avaliaram o desempenho do modelo WRF-Solar na previsão da irradiância global horizontal (GHI) com até 48 horas de antecedência. O estudo foi realizado no Centro de Referência em Energia Solar da Eletrobras Chesf (CRESP), em Petrolina/PE, durante os meses de julho e outubro de 2020, representando períodos de menor e maior incidência solar. Os resultados mostraram que o WRF-Solar superou o modelo de controle (WRF-CTL) em termos de precisão, sobretudo em condições de céu claro e nublado, resultado atribuído à parametrização mais avançada de nuvens. Contudo, ambos os modelos apresentaram limitações na representação do ciclo diurno da GHI, com maiores erros concentrados no período da manhã. Esses achados reforçam o potencial do WRF-Solar para previsões de curto prazo, ao mesmo tempo em que destacam a necessidade de calibrações adicionais e validações em diferentes contextos climáticos para ampliar sua aplicabilidade.

Dessa forma, evidencia-se a necessidade de abordagens que integrem dados reais de cobertura de nuvens, com maior granularidade e qualidade, a modelos de previsão baseados em *deep learning*. Além disso, há espaço para pesquisas que considerem as especificidades regionais do clima tropical, como as da região Sudeste do Brasil, cujo regime de nuvens influencia diretamente a geração solar fotovoltaica.

3 METODOLOGIA

3.1 Características gerais do clima local

O Estado do Espírito Santo está localizado entre as coordenadas 17°52' e 21°17' de latitude Sul e 41°54' e 39°38' de longitude Oeste. Seu território, com aproximadamente 46.184 km², situa-se na costa sudeste do Brasil e faz divisa com Minas Gerais a oeste, Bahia ao norte, Rio de Janeiro ao sul e o Oceano Atlântico a leste. O estado apresenta, predominantemente, dois padrões climáticos: o tropical chuvoso e o tropical de altitude (Seki *et al.*, 2021).

O clima tropical caracteriza-se por temperaturas elevadas e pela alternância entre uma estação seca e outra chuvosa, sendo típico de áreas de média e baixa latitude (Medeiros *et al.*, 2020). Nessa configuração, o verão apresenta altos índices pluviométricos, enquanto o inverno é marcadamente seco, resultando em uma distribuição não uniforme das chuvas ao longo do ano (INPE, 2023). Essa condição varia de acordo com a latitude: quanto maior a distância em relação à linha do Equador, maior é a duração da estação seca, que pode se estender por mais de oito meses sem registros de chuva (Mastella *et al.*, 2019).

Na faixa litorânea do Espírito Santo, a estação chuvosa ocorre, em média, entre outubro e abril, enquanto os meses mais secos vão de junho a agosto – sendo agosto o menos chuvoso do ano. Maio e setembro são considerados meses de transição (Incaper, 2023).

Para compor o banco de dados referente às condições de nebulosidade no local de estudo, delimitou-se o período chuvoso e o período seco dos últimos seis anos (2019-2023). Como o período seco se restringe a apenas três meses, o período chuvoso foi padronizado para os três meses de maior índice pluviométrico: outubro, novembro e dezembro (Incapar, 2023). Assim, o período de análise considerou os dados de outubro a dezembro de 2013 e de junho a agosto de 2023, organizados conforme apresentado no Quadro 1.

Quadro 1 – Divisão entre meses secos e chuvosos na região de estudo

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Período												
Seco						X	X	X				
Chuvoso										X	X	X

Fonte: Autores (2025).

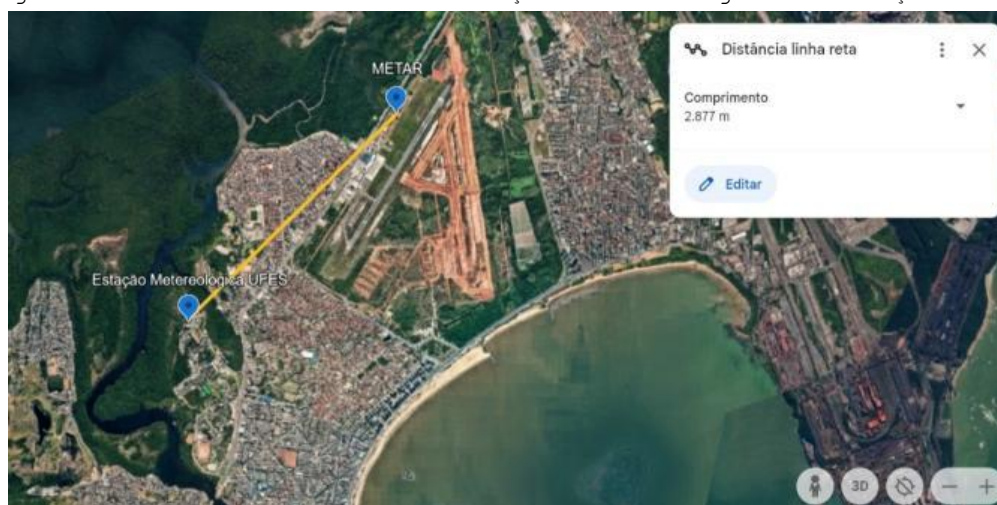
3.2 Coleta de dados

A disponibilidade dos dados de nebulosidade é fornecida pelo METAR (*METeorological Aerodrome Report*), boletim meteorológico regular emitido por aeródromos brasileiros. Esse relatório descreve de forma completa as condições meteorológicas observadas em um aeródromo específico. Cada METAR é gerado a cada 60 minutos, com base na observação meteorológica dos últimos 10 minutos (DECEA, 2019). Para este estudo, os dados foram obtidos na REDEMETS (2023), referentes ao Aeroporto de Vitória (SBVT), e decodificados por meio do software Metar Reader and Decoder v3.1, da *Flight Utilities* (2023). A estação de medição do METAR está localizada nas coordenadas 20°15'9.61" S e 40°17'11.87" O, com um raio de abrangência de 10 km.

Como os dados de nebulosidade seriam correlacionados com os de irradiação solar, foram selecionadas apenas as observações realizadas durante o período de insolação. Dessa forma, o intervalo considerado foi das 07h às 17h no período seco e das 06h às 18h no período chuvoso. Os dados de radiação solar referentes aos mesmos intervalos foram cedidos pelo Incaper, provenientes de sua estação meteorológica de superfície localizada nas coordenadas 20°16'15.89" S e 40°18'21.77" O. Os registros são expressos em kJ/m² e possuem raio de abrangência de 25 km.

A distância entre a estação METAR e a estação meteorológica é de 2.877 m em linha reta, conforme ilustrado na Figura 1. Essa proximidade não compromete a correlação dos dados, uma vez que o raio de abrangência de ambas as estações cobre integralmente a área de estudo.

Figura 1 - Distância entre a estação meteorológica e estação METAR



Fonte: Autores (2025).

3.3 Análise de tendência

3.3.1 Simulação de Monte Carlo

A Simulação de Monte Carlo é uma técnica que utiliza simulações aleatórias repetidas para modelar o comportamento de sistemas complexos. Quando se pretende prever um cenário futuro com várias incertezas, a Simulação de Monte Carlo gera cenários variados, baseando-se em distribuições de probabilidade para essas incertezas, o que resulta em muitos resultados possíveis (Garg; Yadav & Aswal, 2019).

A fórmula básica utilizada pela Simulação de Monte Carlo é expressa conforme Equação 1.

$$\hat{E}(X) = \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (1)$$

Onde:

$\hat{E}(X)$ representa a estimativa esperada de X;

N é o número de simulações realizadas;

x_i é um valor amostrado a partir da distribuição de probabilidade especificada;

$f(x_i)$ é a função de interesse aplicada aos valores amostrados x_i .

Essa fórmula calcula a média dos resultados obtidos a partir de cada simulação, aproximando assim o valor esperado para a variável de interesse.

Neste estudo, a técnica de Monte Carlo foi utilizada para criar diferentes cenários de comportamento da variável de interesse ao longo do tempo. Por exemplo, ao analisar a radiação solar, a Simulação de Monte Carlo permite a geração de cenários variados com base em padrões climáticos distintos, a fim de proporcionar uma visão mais ampla dos possíveis comportamentos futuros dos dados.

Combinada a ela, utilizou-se o *K-Means*, isso se deu devido à necessidade de segmentar grandes volumes de dados em grupos com características semelhantes, permitindo identificar padrões distintos e facilitando a análise detalhada. A aplicação do *K-Means* possibilitou agrupar dados históricos em diferentes cenários, como condições climáticas ou operacionais, que apresentam comportamento homogêneo. A partir desses clusters, a simulação de

Monte Carlo foi implementada para modelar a variabilidade das variáveis de entrada em cada grupo, gerando múltiplas simulações que resultaram em distribuições probabilísticas específicas para cada cenário.

3.3.2 Agrupamento *K-Means*

O *K-Means* é uma técnica de agrupamento que organiza dados em diferentes grupos (ou clusters) com base nas semelhanças entre eles. Sua operação é bastante direta: ele busca padrões nos dados e agrupa os pontos que são mais semelhantes entre si em um mesmo grupo (Huang et al., 2021).

A fórmula básica utilizada pelo *K-Means* é a minimização da soma das distâncias quadradas entre cada ponto e o centro do cluster ao qual ele pertence. O objetivo é encontrar k clusters, minimizando a variabilidade interna dos clusters. A fórmula matemática do *K-Means* é expressa como na Equação 2.

$$J = \sum_{i=1}^k \sum_{j=1}^{n_i} \|x_j^{(i)} - c_i\|^2 \quad (2)$$

Onde:

J é a função objetivo que representa a soma das distâncias quadradas entre os pontos e os centroides dos clusters;

k é o número de clusters;

n_i é o número de pontos no cluster i .

$x_j^{(i)}$ representa o ponto j pertencente ao cluster i .

c_i é o centroide do cluster i .

$\|x_j^{(i)} - c_i\|^2$ representa a distância euclidiana entre o ponto $x_j^{(i)}$ e o centroide c_i elevado ao quadrado.

Neste contexto, o método *K-Means* é aplicado para identificar padrões e tendências ocultas nos dados. Ao analisar as variações de irradiação solar ao longo do ano, essa técnica auxilia na categorização de diferentes condições, como dias ensolarados, parcialmente nublados ou chuvosos. Tal segmentação é fundamental para que os dados possam ser analisados com maior nível de detalhe, possibilitando uma compreensão mais precisa das características de cada grupo identificado.

A definição do número de *clusters* foi realizada por meio da técnica do cotovelo, amplamente utilizada para determinar o ponto ótimo em que a adição de novos grupos deixa de trazer ganhos significativos na redução da variabilidade interna. Essa técnica consiste em calcular a soma dos erros quadráticos médios (ou variabilidade interna) para diferentes quantidades de *clusters* e, em seguida, representar esses valores em um gráfico. O ponto de inflexão da curva, que se assemelha a um "cotovelo", é adotado como o número ideal de *clusters* a ser utilizado (Fajar; Adi & Widodo, 2018), equilibrando a complexidade do modelo e a precisão da segmentação.

A técnica do cotovelo utiliza a fórmula da Soma das Distâncias Quadradas Intra-Clusters (*Within-Cluster Sum of Squares*, ou WCSS) para avaliar a variabilidade interna dos clusters. A fórmula do WCSS é dada pela Equação 3.

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - c_i\|^2 \quad (3)$$

Onde:

k é o número de clusters;

x representa um ponto de dado pertencente ao cluster C_i ;

c_i é o centroide do cluster C_i ;

$\|x - c_i\|^2$ é a distância euclidiana quadrada entre o ponto x e o centroide c_i .

3.3.3 Máquina de Vetores de Suporte (SVM)

A Máquina de Vetores de Suporte (SVM, Support Vector Machine) é uma técnica de aprendizado de máquina amplamente utilizada para tarefas de classificação e regressão. Seu objetivo principal é identificar uma linha – ou um hiperplano, no caso de dados em dimensões superiores – que separe de forma ótima as diferentes classes ou permita prever uma variável alvo (Cervantes *et al.*, 2020).

Nos problemas de classificação, o SVM procura encontrar o hiperplano que maximize a margem entre as classes de dados. Essa margem corresponde à distância entre o hiperplano e os pontos mais próximos de cada classe, denominados vetores de suporte. Dessa forma, o algoritmo seleciona o hiperplano que oferece a maior margem possível, garantindo uma separação mais robusta entre as classes (Tharwat, 2019).

Em situações envolvendo duas classes, o objetivo é determinar um hiperplano que satisfaça a condição de maximização da margem, conforme expresso na Equação 4.

$$w \cdot x - b = 0 \quad (4)$$

Onde:

w é o vetor de pesos que define a orientação do hiperplano.

x é o vetor de características de um ponto de dado.

b é o viés do hiperplano.

Em problemas de regressão, a SVM é adaptada para prever valores contínuos (Deo *et al.*, 2016). Nesse caso, a técnica tenta encontrar uma função que minimize o erro dentro de uma margem de tolerância (ϵ). Isso é feito de forma similar à classificação, mas o objetivo é ajustar uma linha (ou hiperplano) que passe pela maioria dos dados, mantendo os desvios dentro de uma faixa específica.

Para avaliar a eficácia de cada uma das ferramentas utilizadas, será calculado o Erro Médio Absoluto (MAE) de cada abordagem. O MAE é uma métrica que mede a média das diferenças absolutas entre os valores previstos e os valores reais, proporcionando uma visão clara da precisão de cada modelo, ele é dado pela Equação 5.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Onde:

n é o número total de observações;

y_i representa o valor real na observação i ;

$\hat{y}_i$ representa o valor previsto pelo modelo na observação i ;

$|y_i - \hat{y}_i|$ é o valor absoluto da diferença entre o valor real e o valor previsto.

A ferramenta que apresentar o menor MAE foi considerada a mais eficaz, indicando sua capacidade superior de previsão e modelagem dos dados analisados.

4 RESULTADOS E DISCUSSÃO

Os dados de irradiação solar utilizados neste estudo foram obtidos na estação meteorológica do Incaper, enquanto as informações sobre cobertura de nuvens foram extraídas dos relatórios METAR, que fornecem observações detalhadas e em tempo real das condições meteorológicas em aeródromos. Essas duas bases de dados foram fundamentais para assegurar a precisão das análises e previsões realizadas. Para o tratamento dos dados, todas as informações foram reunidas em uma planilha do Excel e organizadas linha a linha, de acordo com o dia e a hora de cada medição, garantindo o alinhamento entre os registros de irradiação e de cobertura de nuvens.

Nas medições de cobertura de nuvens, o céu é dividido em oitavos, denominados octas, que representam diferentes frações de ocupação do céu. Cada octa corresponde a um grau específico de cobertura de nuvens, afetando diretamente a quantidade de radiação solar que atinge a superfície. Por exemplo, um céu com 0 octas (céu limpo) permite a máxima incidência de radiação solar, enquanto o aumento do número de octas reduz progressivamente a radiação que atravessa a atmosfera. No entanto, essa redução não é linear, o que exige a utilização do Fator de Modificação de Nuvens (CMF, Cloud Modification Factor) para estimar a irradiação correspondente a cada condição de cobertura.

O CMF é uma métrica que quantifica o efeito das nuvens na atenuação da radiação solar incidente na superfície terrestre (Jesus, 2023). O cálculo da irradiação solar acima das nuvens possibilita estimar a quantidade de radiação que chega ao topo da camada de nuvens e, a partir disso, determinar o valor efetivo de irradiação em cada octa para cada medição.

A relação básica é expressa de acordo com a Equação 6.

$$I_{superior} = \frac{I_{superficie}}{CMF} \quad (6)$$

Onde:

$I_{superior}$ é a irradiação solar no topo das nuvens (ou irradiação sem a influência das nuvens);

$I_{superficie}$ é a irradiação solar medida na superfície;

CMF é o Fator de Modificação de Nuvens, que varia entre 0 e 1 e representa a fração da radiação que passa através das nuvens e atinge a superfície. Valores de CMF mais próximos de 1 indicam poucas nuvens ou céu claro, enquanto valores menores indicam maior cobertura de nuvens.

Em geral, o CMF é calculado da seguinte forma:

$$CMF = 1 - k \quad (7)$$

O índice k pode ser observado no Quadro 1.

Quadro 1 - Transmissão de UV-B para Tipos de Nuvens

Estágio	Tipo de nuvem	k
Alto (Acima de 8 km)	Cirrus (Ci)	≈ 0.2 0.3
	Cirrocumulus (Cc)	≈ 0.2 0.3
	Cirrostratus (Cs)	≈ 0.3 0.4
Médio (2 - 8 km)	Alto cumulus (Ac)	≈ 0.5 0.6
	Altostratus (As)	≈ 0.6 0.7
	Nimbostratus (Ns)	≈ 0.8 0.9
Baixo (Até 2 km)	Stratocumulus (Sc)	≈ 0.6 0.7
	Stratus (St)	≈ 0.7 0.8
	Cumulus (Cu)	≈ 0.5 0.6
	Cumulonimbus (Cb)	≈ 0.7 0.8

Fonte: Adaptado de Iqbal (1983), Kuchinke e Nunez (1999) e Jesus (2015).

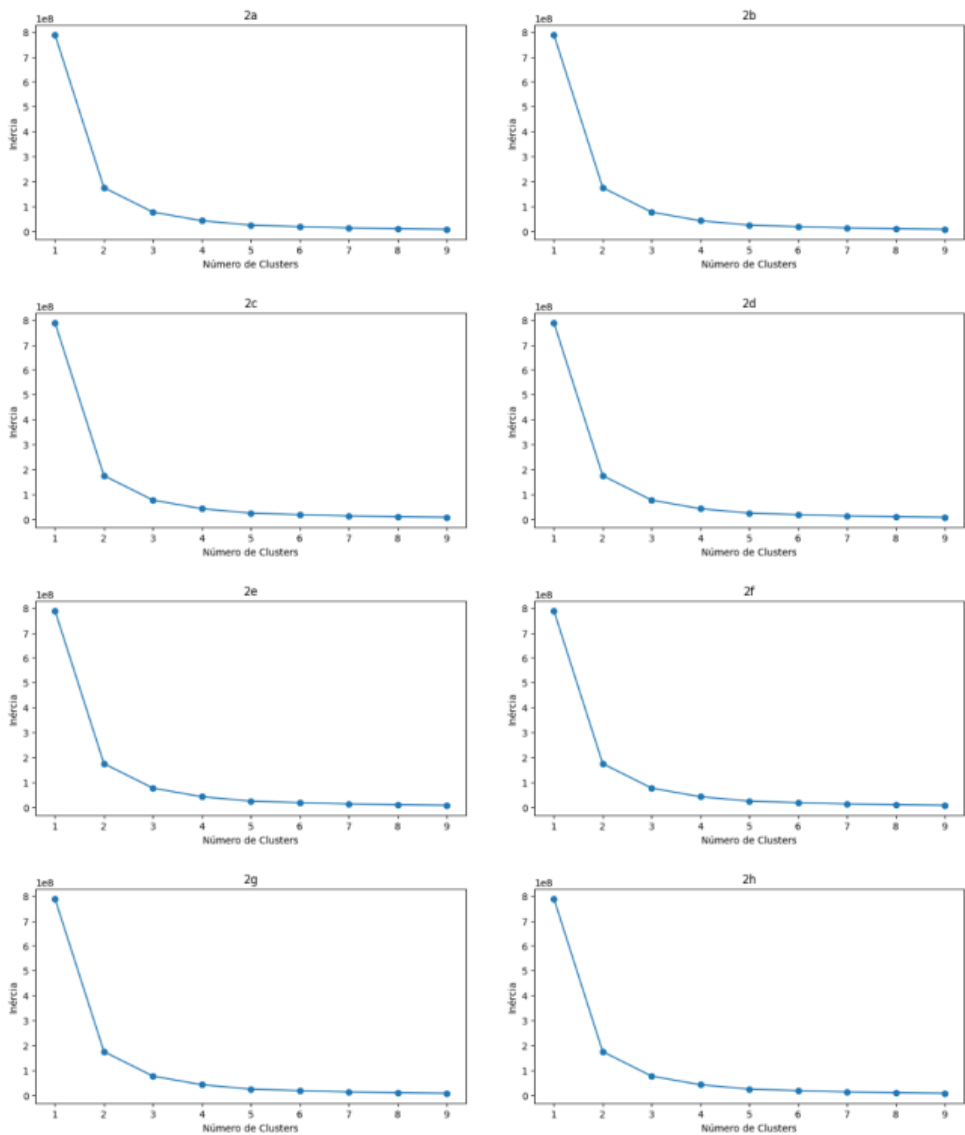
Após o cálculo do CMF para cada octa ao longo de seis anos, tanto para o período chuvoso quanto para o período seco, foi possível calcular a radiação solar e a potência disponível para cada octa, com base nos dados da radiação medida.

Com os dados de potência e radiação, iniciaram-se as previsões, utilizando como variáveis preditoras a Radiação e o CMF, e como variável alvo a Potência, que é o parâmetro a ser previsto. Essa abordagem permitiu modelar a influência direta da radiação solar e da cobertura de nuvens sobre a geração de energia, possibilitando uma estimativa mais precisa do desempenho do sistema em diferentes cenários climáticos.

4.1 Simulação de Monte Carlo e agrupamento K-Means

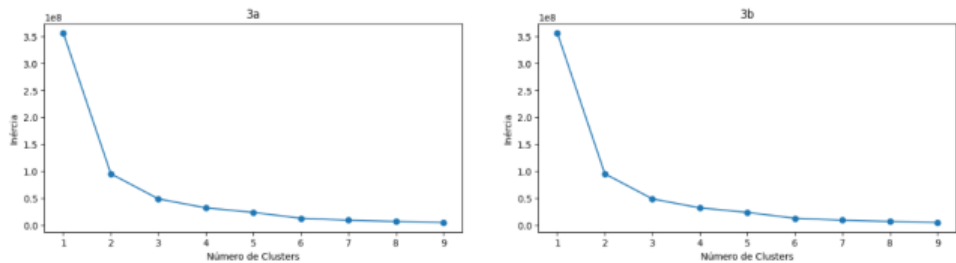
Os resultados gerados foram divididos por octa, ou seja, foram 8 previsões realizadas no período chuvoso e 8 previsões no período seco. O resultado da técnica do cotovelo pode ser visto na Figura 2 (2a à 2h) para o período chuvoso e Figura 3 (3a à 3h) para o período seco, o que resulta a escolha dos clusters.

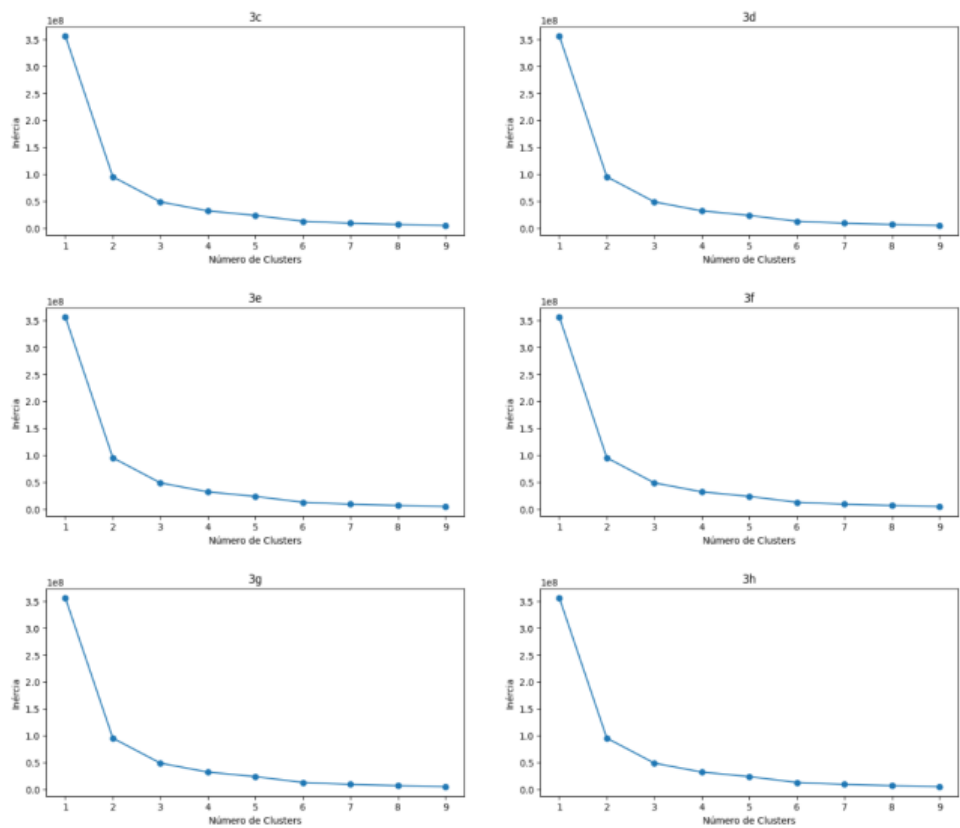
Figura 2 - Escolha dos clusters pela técnica do cotovelo período chuvoso



Fonte: Autores (2025).

Figura 3 - Escolha dos clusters pela técnica do cotovelo período seco





Fonte: Autores (2025).

Para o período chuvoso, definiu-se o número de *clusters* como cinco, enquanto para o período seco foram utilizados seis *clusters*. Essas quantidades foram determinadas a partir da identificação do ponto de inflexão (“cotovelo”) nos respectivos gráficos, que indicam o número ideal de *clusters* para cada regime climático.

Assim, os agrupamentos foram realizados utilizando o método *K-Means*, e as previsões subsequentes foram conduzidas por meio da Simulação de Monte Carlo. Os resultados do MAE obtidos para cada *octa* são apresentados na Tabela 1.

Tabela 1 – Resultado do MAE das simulações de Monte Carlo para o período chuvoso

Média de Potência: 418,41			
Mediana de Potência: 342,12			
Octa	MAE	Proporção do MAE em relação à Média	Proporção do MAE em relação à Mediana
1	1,5951219512195123	0,38%	0,47%
2	1,5863414634146342	0,38%	0,46%
3	1,6018118466898954	0,38%	0,47%
4	1,5887108013937283	0,38%	0,46%
5	1,5974912891986062	0,38%	0,47%
6	1,5689198606271777	0,37%	0,46%
7	1,5662717770034844	0,37%	0,46%
8	1,566829268292683	0,37%	0,46%

Fonte: Autores (2025).

A proporção do MAE em relação à média de potência manteve-se bastante baixa, variando entre 0,37% e 0,38%. Esse resultado indica que o erro absoluto médio representa menos de meio por cento da potência média durante o período chuvoso, evidenciando que o modelo apresenta alta precisão mesmo sob a influência da cobertura de nuvens em cada *octa*.

Quando comparado à mediana de potência, o MAE também se mostrou baixo, variando entre 0,46% e 0,47%. Essa proporção ligeiramente superior à da média sugere que o MAE exerce maior impacto relativo sobre a mediana, o que pode ser explicado pelo fato de a mediana ser um valor central inferior, aumentando a representatividade percentual do erro.

Observa-se ainda que os valores de MAE se mantiveram muito próximos entre os oito *octas*, variando em uma faixa estreita de 1,566 a 1,601. Essa consistência indica que o modelo não foi significativamente influenciado pela cobertura de nuvens em diferentes oitavos, sugerindo robustez e capacidade de capturar as variações de potência de forma consistente.

Os resultados referentes ao período seco são apresentados na Tabela 2.

Tabela 2 - Resultado do MAE das simulações de Monte Carlo para o período seco

Octa	Média de Potência: 340,06		
	Mediana de Potência: 319,07		
	MAE	Proporção do MAE em relação à Média	Proporção do MAE em relação à Mediana
1	1,81782243452479	0,54%	0,57%
2	1,8252347224509966	0,54%	0,57%
3	1,819799044638445	0,54%	0,57%
4	1,802997858672377	0,53%	0,57%
5	1,7692307692307692	0,52%	0,55%
6	1,8188107395816175	0,54%	0,57%
7	1,8059627738428594	0,53%	0,57%
8	1,8150222368637787	0,53%	0,57%

Fonte: Autores (2025).

Os valores de MAE para cada *octa* variaram entre 1,769 e 1,825, indicando apenas uma ligeira variação dos erros entre os diferentes oitavos. Esse resultado confirma que o modelo é consistente ao lidar com distintos padrões de cobertura de nuvens durante o período seco.

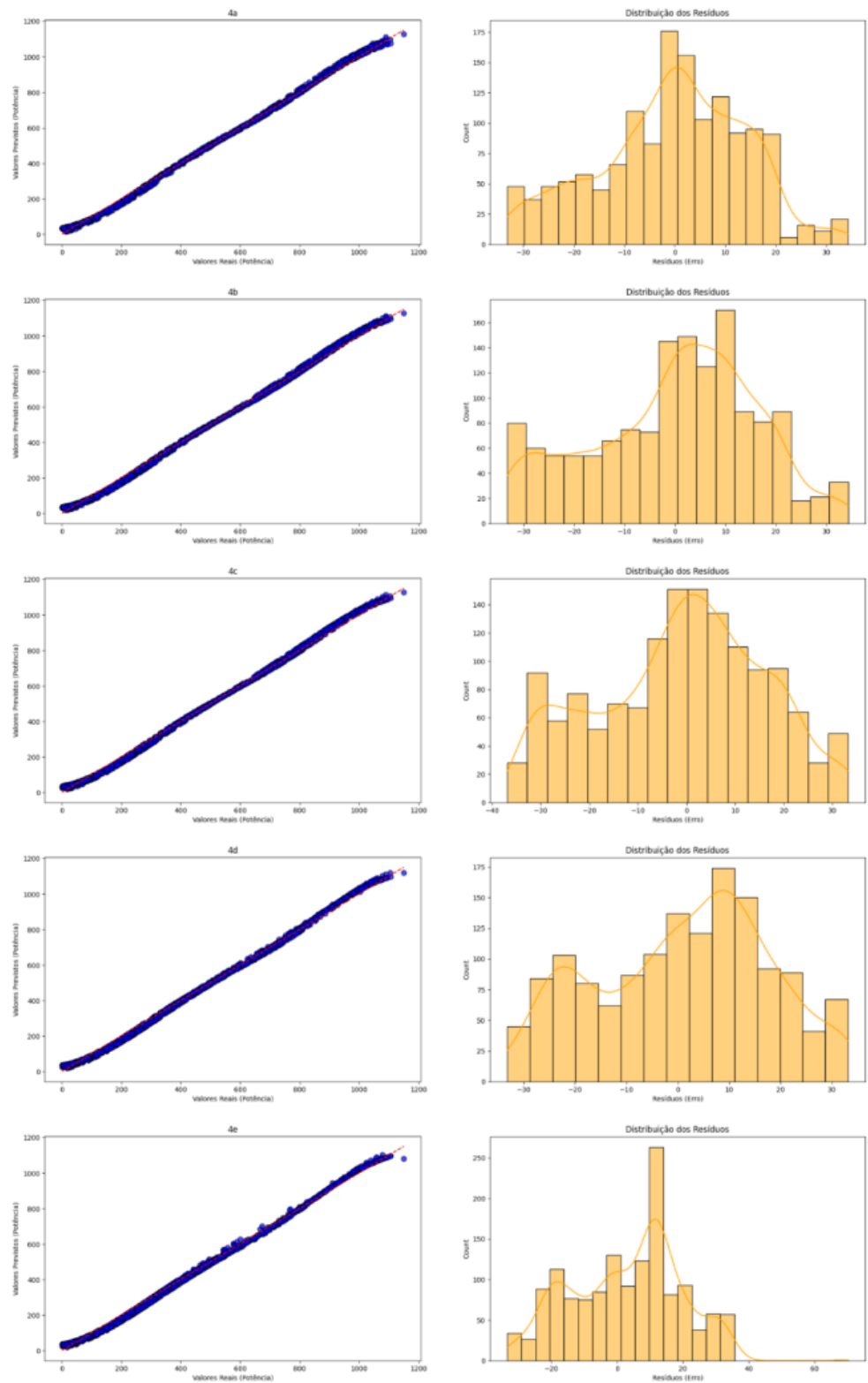
A proporção do MAE em relação à média de potência permaneceu baixa, entre 0,52% e 0,54%, enquanto em relação à mediana variou de 0,55% a 0,57%. Esses valores demonstram que o erro absoluto médio corresponde a pouco mais de meio por cento da potência média nesse período. Apesar de serem levemente superiores aos do período chuvoso, esses resultados ainda evidenciam elevada precisão do modelo.

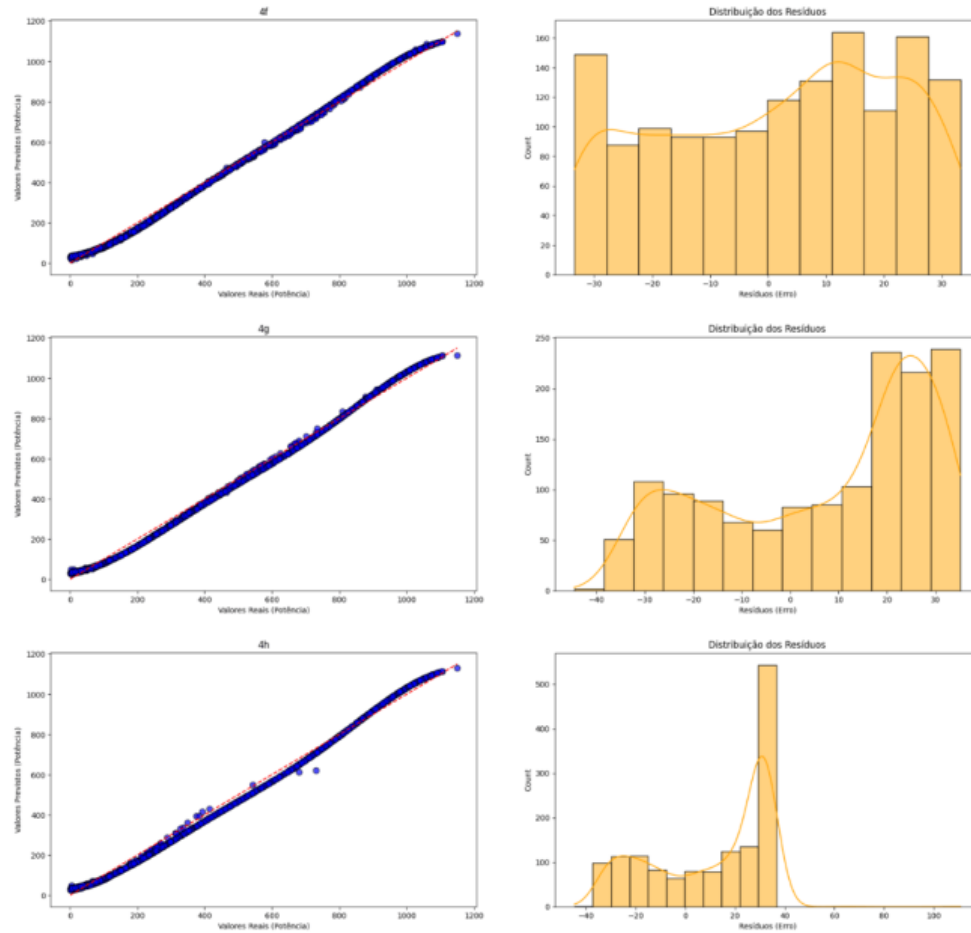
A pequena diferença observada entre os períodos chuvoso e seco reforça que o modelo é robusto e capaz de lidar adequadamente com diferentes condições de cobertura de nuvens e variações na irradiação solar. Dessa forma, confirma-se sua versatilidade para oferecer previsões confiáveis ao longo de diferentes cenários climáticos.

4.2 Máquina de Vetores de Suporte (SVM)

A Figura 4 (4a à 4h) expõe os resultados das previsões dos oito oitavos no período chuvoso e a distribuição dos resíduos dos erros.

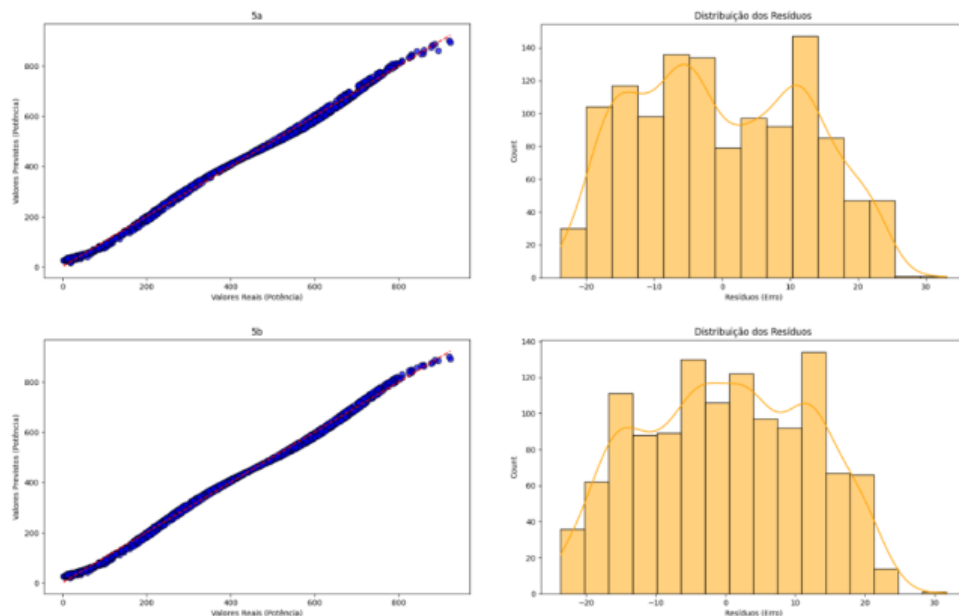
Figura 4 – Resultado SVM do período chuvoso

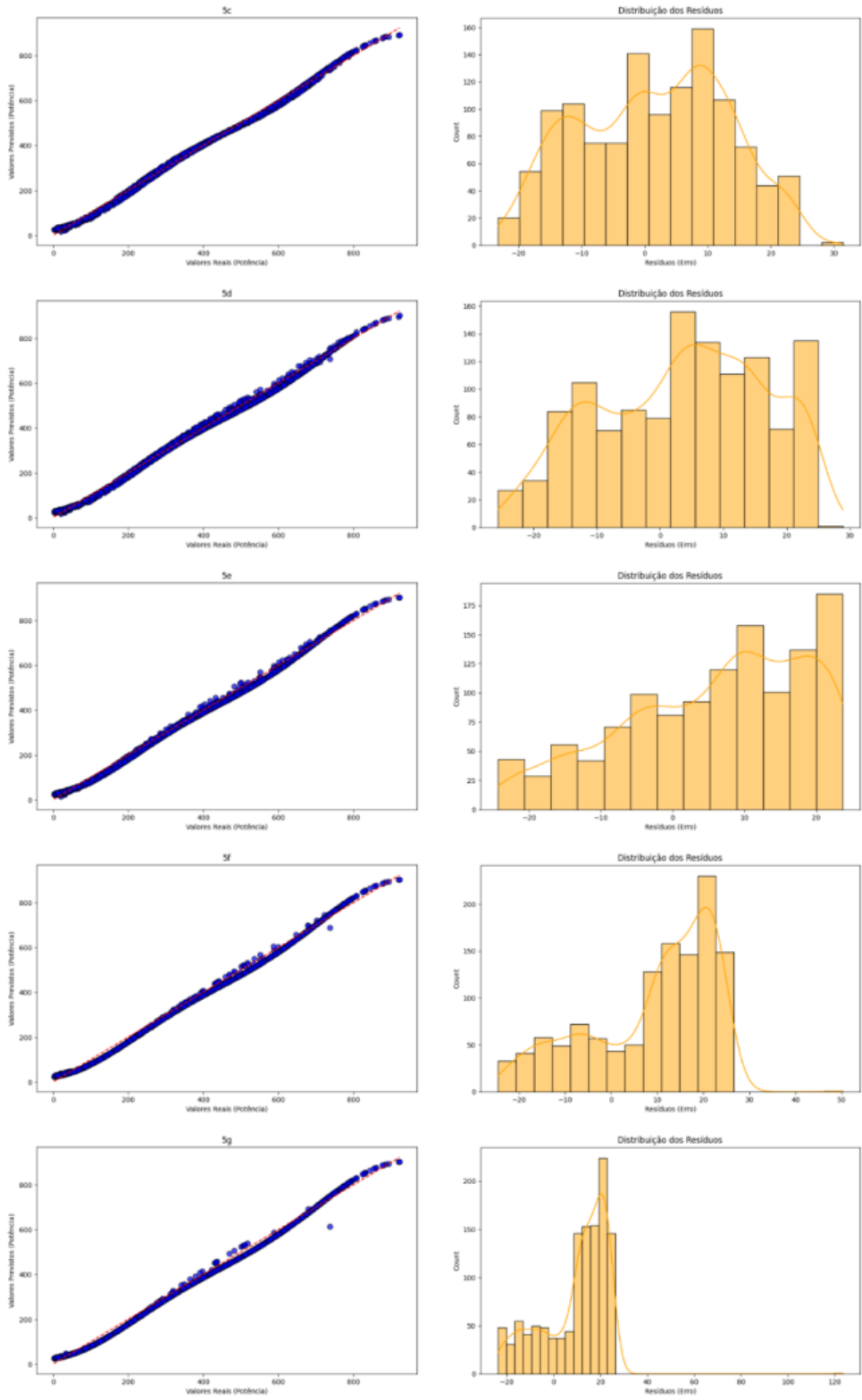


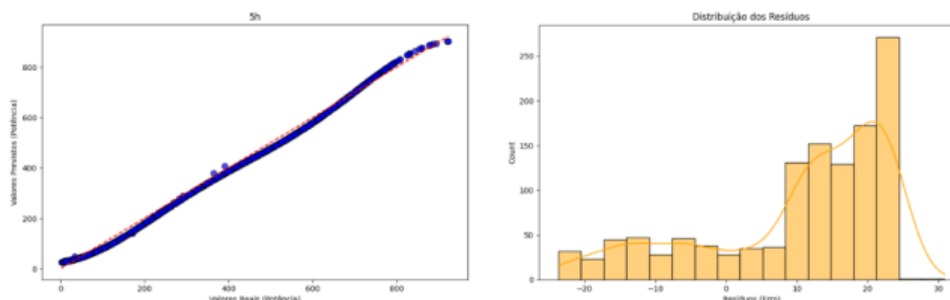


Fonte: Autores (2025).

Figura 5 – Resultado SVM do período seco







Fonte: Autores (2025).

Os valores de MAE variaram de 10,72 no primeiro octa até 15,36 no oitavo, apresentando uma progressão crescente. Esse comportamento sugere que os erros absolutos se tornam maiores à medida que a cobertura de nuvens aumenta ou as condições climáticas se tornam mais complexas ao longo dos oitavos.

Nos três primeiros octas, os MAEs permaneceram na faixa de 9,96 a 10,72, indicando maior precisão nessas condições. Já nos oitavos subsequentes, os erros se elevaram, evidenciando desafios adicionais na previsão em cenários de maior variabilidade climática.

A análise dos resíduos revelou uma distribuição com desvio positivo, indicando que o modelo tende a subestimar os valores reais em determinadas situações. Entretanto, a presença de picos próximos a zero na maioria dos gráficos é um indicativo favorável, pois demonstra que, na maior parte das previsões, os erros são reduzidos ou muito próximos de zero.

3 CONSIDERAÇÕES FINAIS

Conclui-se que a modelagem preditiva da potência disponível, desenvolvida a partir dos dados de radiação solar e utilizando as técnicas de Simulação de Monte Carlo e agrupamento *K-Means*, mostrou-se eficaz para a análise das variáveis de radiação solar e cobertura de nuvens. A abordagem proposta permitiu identificar padrões de variabilidade nos dados e gerar previsões precisas de potência, mesmo em cenários meteorológicos distintos. A aplicação da SVM evidenciou maior sensibilidade à variabilidade de potência quando considerada a cobertura de nuvens.

O uso de dados reais da região metropolitana de Vitória-ES, aliado à incorporação do CMF, contribuiu de forma significativa para o aumento da precisão das previsões, reforçando o potencial do modelo para apoiar a gestão eficiente da geração de energia solar. Contudo, persistem desafios relacionados à detecção de padrões mais complexos de cobertura de nuvens e às variações climáticas específicas da região. Nesse sentido, recomendam-se estudos futuros voltados à integração de novas técnicas de aprendizado de máquina, a fim de aprimorar ainda mais a qualidade das previsões.

REFERÊNCIAS

AKOREDE, M.; HIZAM, H.; POURESMANEIL, E. *Distributed energy resources and benefits to the environment*. **Renewable & Sustainable Energy Reviews**, [S.l.], v. 14, p. 724-734, 2010. DOI: 10.1016/j.rser.2009.10.025.

ALQURASHI, A. *Economic and environmental impacts of distributed energy resources*. **Journal of Clean Energy Technologies**, [S.l.], 2021. DOI: 10.18178/jocet.2021.9.2.527.

CERVANTES, J.; GARCÍA, F.; RODRÍGUEZ-MAZAHUA, L.; CHAU, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. **Neurocomputing**, v. 408, p. 189-215, 2020. Disponível em: <https://doi.org/10.1016/j.neucom.2019.10.118>. Acesso em: 10 abr. 2025.

DEO, R. C.; WEN, X.; QI, F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. **Applied Energy**, v. 168, p. 568-593, 2016.

DI PIAZZA, A.; DI PIAZZA, M. C.; LA TONA, G.; LUNA, M. An artificial neural network-based forecasting model of energy-related time series for electrical grid management. **Mathematics and Computers in Simulation**, v. 184, p. 294-305, 2021. Disponível em: <https://doi.org/10.1016/j.matcom.2020.05.010>. Acesso em: 10 abr. 2025.

FAJAR, M.; ADI, K.; WIDODO, C. An evaluation of educational process with K-means clustering for students grouping. **International Journal of Computer Applications**, 2018. Disponível em: <https://doi.org/10.5120/IJCA2018917858>. Acesso em: 10 abr. 2025.

FERRAZ, R. S. F.; FERREZ, R. S. F.; RUEDA MEDINA, A. C. A novel two-stage multi-objective optimization strategy for enhanced network planning and operation. **Electrical Engineering**, 2024. Disponível em: <https://doi.org/10.1007/s00202-024-02675-3>. Acesso em: 10 abr. 2025.

FLIGHT UTILITIES. Download Programs. 2023. Disponível em: http://www.flightutilities.com/download_prg.aspx. Acesso em: 14 set. 2023.

GAO, B.; HUANG, X.; SHI, J.; TAI, Y.; ZHANG, J. Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. **Renewable Energy**, v. 162, p. 1665-1683, 2020. Disponível em: <https://doi.org/10.1016/j.renene.2020.09.141>. Acesso em: 10 abr. 2025.

GARDNER, J.; CARR-CORNISH, S.; ASHWORTH, P. *Exploring the acceptance of a domestic distributed energy market in Australia*. **Australasian Journal of Environmental Management**, [S.l.], v. 15, p. 103-93, 2008. DOI: 10.1080/14486563.2008.9725189.

GARG, N.; YADAV, S.; ASWAL, D. Monte Carlo simulation in uncertainty evaluation: Strategy, implications and future prospects. **MAPAN**, v. 34, p. 299-304, 2019. Disponível em: <https://doi.org/10.1007/s12647-019-00345-5>. Acesso em: 10 abr. 2025.

HUANG, S.; KANG, Z.; XU, Z.; LIU, Q. Robust deep k-means: An effective and simple method for data clustering. **Pattern Recognition**, v. 117, 107996, 2021. Disponível em: <https://doi.org/10.1016/J.PATCOG.2021.107996>. Acesso em: 10 abr. 2025.

INSTITUTO CAPIXABA DE PESQUISA, ASSISTÊNCIA TÉCNICA E EXTENSÃO RURAL (INCAPER). Boletim agroclimático do Espírito Santo. Vitória, ES: Incaper, v. 10, n. 1, p. 5-25, jan./mar. 2023.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). Centro de Previsão de Tempo e Estudos Climáticos. [S. l.], 2023. Disponível em: <https://www.cptec.inpe.br/>. Acesso em: 19 set. 2023.

JESUS, H. S. de; COELHO COSTA, S. M. S.; CEBALLOS, J. C.; CORRÊA, M. P. Cloud modification factor parametrization for solar UV based on the GOES satellite: Validation using ground-based measurements in São Paulo city, Brazil. **Atmospheric Environment**, v. 309, 119942, 2023. Disponível em: <https://doi.org/10.1016/j.atmosenv.2023.119942>. Acesso em: 10 abr. 2025.

KUMARI, P.; TOSHNIWAL, D. Deep learning models for solar irradiance forecasting: A comprehensive review. **Journal of Cleaner Production**, v. 318, 2021, Art. 128566. Disponível em: <https://doi.org/10.1016/j.jclepro.2021.128566>. Acesso em: 10 abr. 2025.

PARALIČ, M. Statistical approach for sky clouds density classification. In: **New Trends in Signal Processing (NTSP)**, 2020, Demanovska dolina, Slovakia. Anais [...]. p. 1-5. Disponível em: <https://doi.org/10.1109/NTSP49686.2020.9229538>. Acesso em: 10 abr. 2025.

RAMOS, D. N. da S. et al. Previsão de curto prazo da irradiância global horizontal no semiárido brasileiro sob diferentes condições de nebulosidade usando o modelo WRF-Solar. In: **Congresso Brasileiro de Energia Solar - CBENS**, 2024. Disponível em: <https://anaiscbens.emnuvens.com.br/cbens/article/view/2393>. DOI: 10.59627/cbens.2024.2393. Acesso em: 10 abr. 2025.

REIKARD, G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. **Solar Energy**, v. 83, n. 3, p. 342-349, 2009. Disponível em: <https://doi.org/10.1016/j.solener.2008.08.007>. Acesso em: 10 abr. 2025.

SHARMA, A. et al. *Reduction in greenhouse gases emission using distributed energy resources (DER) in distribution network*. **International Journal of Engineering Research & Technology**, [S.l.], v. 3, n. 5, p. 37-41, 2016. (DOI não informado).

WANG, J. et al. Climatology of cloud phase, cloud radiative effects and precipitation properties over the Tibetan Plateau. **Remote Sensing**, Basel, v. 13, n. 3, p. 363, 2021. Disponível em: <https://doi.org/10.3390/rs13030363>. Acesso em: 10 abr. 2025.

YUAN, F. et al. Correlation between cloud optical thickness and solar radiation. In: **USNC-URSI Radio Science Meeting**, 2016, Fajardo, PR, USA. Anais [...]. p. 105-106. Disponível em: <https://doi.org/10.1109/USNC-URSI.2016.7588534>. Acesso em: 10 abr. 2025.

ZHANG, H.; ZHANG, CH. Short-term DNI forecasting using cloud history patterns matching. In: **7th International Conference on Information Science and Technology**, 2017, Da Nang, Vietnam. Anais [...]. p. 57-63.

ZHANG, Y. et al. Quantification of aerosol and cloud effects on solar energy over China using WRF-Chem. **Atmospheric Research**, v. 275, p. 106245, set. 2022. Disponível em: <https://doi.org/10.1016/j.atmosres.2022.106245>. Acesso em: 10 abr. 2025.