

MANUELA BAGGIO LOBATO
MANUELA LACERDA PAIVA SAMPAIO

Avaliação da Inclusão de Variáveis Exógenas em Modelos SARIMA para
Previsão da Geração Fotovoltaica em Usinas Brasileiras

PROJETO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
APRESENTADO AO DEPARTAMENTO DE ENGENHARIA INDUSTRIAL
DA PUC-RIO, COMO PARTE DOS REQUISITOS PARA OBTENÇÃO
DO TÍTULO DE ENGENHEIRO DE PRODUÇÃO

Orientador: Prof. Fernando Luiz Cyrino Oliveira

Departamento de Engenharia Industrial
Rio de Janeiro, 15 de dezembro de 2025.

AGRADECIMENTOS

Agradecemos às nossas famílias e amigos pelo apoio ao longo da graduação.

Agradecemos também aos professores do Departamento de Engenharia Industrial da PUC-Rio, cuja formação recebida foi essencial para o desenvolvimento deste trabalho. Em especial, ao professor Fernando Cyrino, nosso orientador, pelo acompanhamento atento, pelas discussões técnicas que direcionaram a estrutura do estudo e pelas contribuições decisivas na interpretação dos resultados. Agradecemos ainda ao professor Cristiano Fernandes, cujas aulas de séries temporais forneceram a base necessária para a aplicação dos métodos utilizados neste trabalho.

RESUMO

A previsão precisa da geração solar fotovoltaica é um desafio crescente, especialmente porque séries de geração incorporam efeitos combinados de sazonalidade, variabilidade atmosférica e eventos operacionais difíceis de antecipar. Modelos estatísticos tradicionais, embora eficientes em contextos estáveis, podem apresentar desempenho limitado diante de mudanças abruptas no comportamento da série ou de condições climáticas atípicas. Uma alternativa é incorporar variáveis meteorológicas contemporâneas à modelagem, permitindo capturar variações do balanço radiativo e da cobertura atmosférica que influenciam diretamente a produção de energia. Nesse contexto, o presente estudo tem como objetivo avaliar em que medida variáveis exógenas provenientes da reanálise MERRA-2 — albedo da superfície, fração total da área coberta por nuvens, radiação de onda longa absorvida pela superfície, espessura óptica das nuvens, temperatura da pele da superfície — melhoram a previsão mensal da geração solar nas usinas de Ituverava (BA) e Pirapora (MG), comparando sistematicamente o desempenho de modelos SARIMA e SARIMAX. Para isso, desenvolveram-se as etapas da metodologia Box & Jenkins, abrangendo pré-processamento, identificação e estimação dos modelos, avaliação em múltiplas janelas temporais e projeção futura das séries. Os resultados mostraram que a inclusão de exógenas melhora de forma expressiva a capacidade preditiva dos modelos, com ganhos sistemáticos em Pirapora e efeitos dependentes do regime temporal em Ituverava. A principal contribuição do trabalho para a área de previsão energética está em demonstrar empiricamente que abordagens híbridas, que combinam modelos autorregressivos e informações meteorológicas, aumentam a robustez das previsões mensais de geração solar, oferecendo subsídios relevantes para o planejamento a operação e a tomada de decisão em usinas fotovoltaicas no Brasil, em um cenário de rápida expansão da energia solar.

PALAVRAS-CHAVE

Geração fotovoltaica; Variáveis meteorológicas; Parque Solar de Ituverava; Complexo Solar de Pirapora.

ABSTRACT

Accurately forecasting solar photovoltaic generation is an increasing challenge, as generation series incorporate combined effects of seasonality, atmospheric variability, and operational events that are difficult to anticipate. Traditional statistical models, although effective under stable conditions, may show limited performance when the series experiences abrupt changes or atypical climatic behavior. An alternative is to incorporate contemporary meteorological variables into the modeling process, allowing the model to capture variations in radiative balance and cloud cover that directly influence energy production. In this context, the present study aims to assess the extent to which exogenous variables from the MERRA-2 reanalysis — surface albedo, total cloud fraction, longwave radiation absorbed by the surface, cloud optical thickness, and surface skin temperature — enhance monthly solar generation forecasts for the Ituverava (BA) and Pirapora (MG) power plants, systematically comparing the performance of SARIMA and SARIMAX models. The study applies the Box & Jenkins methodology, including data preprocessing, model identification and estimation, evaluation across multiple temporal windows, and future projections of the series. The results show that incorporating exogenous variables significantly improves predictive accuracy, with consistent gains in Pirapora and regime-dependent effects in Ituverava. The main contribution of this work to the field of energy forecasting lies in empirically demonstrating that hybrid approaches — combining autoregressive models with meteorological information — increase the robustness of monthly solar generation forecasts, providing valuable support for planning, operation, and decision-making in photovoltaic plants in Brazil, in a context of rapid solar expansion.

KEYWORDS

Photovoltaic generation; Meteorological variables; Ituverava Solar Park; Pirapora Solar Complex.

SUMÁRIO

1.	<i>Introdução</i>	1
2.	<i>Referencial Teórico.....</i>	3
2.1.	<i>Usinas Fotovoltaicas</i>	3
2.2.	<i>Previsão de Geração de Energia Fotovoltaica.....</i>	4
2.3.	<i>Modelos ARIMA</i>	8
2.3.1.	Modelo Sazonal Autoregressivo Integrado de Médias Móveis (SARIMA)	9
2.3.2.	Modelo Sazonal Autoregressivo Integrado de Médias Móveis com Variáveis Exógenas (SARIMAX).....	10
3.	<i>Caracterização das Usinas e Variáveis do Estudo</i>	12
3.1.	<i>Usinas do Estudo</i>	12
3.1.1.	Parque Solar de Ituverava (BA)	12
3.1.2.	Complexo Solar Pirapora (MG).....	13
3.2.	<i>Variáveis Exógenas Climáticas.....</i>	14
3.2.1.	Albedo da superfície (ALBEDO)	14
3.2.2.	Fração total da área coberta por nuvens (CLDTOT)	15
3.2.3.	Radiação de onda longa absorvida pela superfície (LWGAB).....	15
3.2.4.	Espessura óptica das nuvens (TAUTOT)	16
3.2.5.	Temperatura da pele da superfície (TS)	17
4.	<i>Metodologia</i>	18
4.1.	<i>Bases de dados e pré-processamento.....</i>	18
4.2.	<i>Metodologia Box & Jenkins.....</i>	21
4.2.1.	Análise exploratória e transformação da série	21
4.2.2.	Estacionarização e diferenciação	22
4.2.3.	Identificação do modelo e estimação dos parâmetros	22
4.2.4.	Diagnóstico e validação do modelo.....	23
4.2.4.1	Diagnóstico dos resíduos	23
4.2.4.2	Validação fora da amostra	24
4.2.5.	Previsão e aplicação	25

5. Resultados e discussões.....	26
5.1. Análise exploratória e transformação da série.....	26
5.1.1. Geração de energia fotovoltaica	26
5.1.2. Variáveis meteorológicas exógenas	30
5.1.3. Relação entre a geração de energia e as variáveis exógenas	32
5.2. Estacionariedade e diferenciação	35
5.3. Identificação, Estimação e Diagnóstico dos modelos.....	36
5.4. Previsão e aplicação	42
5.4.1. Modelagem e projeção das variáveis exógenas.....	42
5.4.2. Previsão da geração de energia fotovoltaica	44
5.4.3. Validação com dados reais	47
6. Conclusões	49

Lista de Figuras

Figura 3.1: Localização das usinas	12
Figura 4.1: Fluxograma da Metodologia	18
Figura 4.2: Linha de corte dos dados de geração de energia	19
Figura 5.1: Série temporal mensal de geração de energia fotovoltaica	26
Figura 5.2: Histogramas da geração mensal das usinas.....	28
Figura 5.3: Boxplot comparativo das usinas Ituverava (BA) e Pirapora (MG).....	28
Figura 5.4: Decomposição STL das séries de geração fotovoltaica	29
Figura 5.5: Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) das séries de geração	30
Figura 5.6: Séries temporais das variáveis meteorológicas exógenas	31
Figura 5.7: Funções de Autocorrelação (FAC) das variáveis meteorológicas exógenas	32
Figura 5.8: Matriz de Correlação	34
Figura 5.9: Previsão da geração fotovoltaica nas janelas de teste em Ituverava	38
Figura 5.10: Previsão da geração fotovoltaica nas janelas de teste em Pirapora	40
Figura 5.11: Projeções futuras das variáveis exógenas (12 passos à frente)	44
Figura 5.12: Previsão futura da geração de energia fotovoltaica (12 passos à frente)	46
Figura 5.13: Comparação entre valores observados e previsões	48

Lista de Tabelas

Tabela 2.1: Principais métodos e horizontes de previsão da geração fotovoltaica.....	5
Tabela 4.1: Configuração das janelas de modelagem.....	25
Tabela 5.1: Estatísticas descritivas de geração de energia fotovoltaica	27
Tabela 5.2: Correspondência entre coordenadas das usinas e os pontos de grade do MERRA-2	30
Tabela 5.3: Correlação entre variáveis exógenas e geração na Usina de Ituverava (BA)	33
Tabela 5.4: Correlação entre variáveis exógenas e geração na Usina de Pirapora (MG).....	33
Tabela 5.5: Fator de Inflação da Variância (VIF) entre exógenas.....	35
Tabela 5.6: Resultados do Teste ADF para as Séries de Geração e Variáveis Exógenas	36
Tabela 5.7: Modelos selecionados para previsão da geração fotovoltaica em Ituverava	38
Tabela 5.8: Modelos selecionados para previsão da geração fotovoltaica em Pirapora.....	40
Tabela 5.9: Modelos selecionados para projeção das variáveis exógenas	43
Tabela 5.10: Modelos selecionados para previsão da geração fotovoltaica	45
Tabela 5.11: Equações dos modelos selecionados para previsão da geração fotovoltaica.....	45
Tabela 5.12: Desempenho das previsões em comparação com dados reais	48

1. INTRODUÇÃO

A transição energética é um dos maiores desafios contemporâneos, impulsionada pela necessidade de reduzir emissões de gases de efeito estufa e conter o avanço das mudanças climáticas (ONU, 2015; IRENA, 2025). Em 2024 a capacidade global de energia renovável alcançou 4,44 TeraWatts (TW), um crescimento de 15% em relação ao ano anterior, impulsionado principalmente pela expansão da energia solar fotovoltaica, que adicionou cerca de 580 GW e respondeu por quase dois terços das novas instalações no mundo (IRENA, 2025). Com isso, as fontes renováveis passaram a representar 43,5% da capacidade elétrica total mundial, consolidando-se como o principal vetor da transição energética global (IRENA, 2025).

No Brasil, cuja matriz elétrica é predominantemente limpa, 88% da oferta interna de eletricidade provém de fontes renováveis, sendo 56,8% de origem hídrica (EPE, 2025). O país lidera a América do Sul em capacidade renovável instalada, com aproximadamente 919 GW em 2024 (IRENA, 2025). Entre as fontes emergentes, a energia solar tem apresentado crescimento expressivo e já representa 9,3% da oferta interna (EPE, 2025).

Os estados da Bahia e de Minas Gerais destacam-se como polos da expansão solar, combinando alta insolação e investimentos robustos em infraestrutura. A Bahia abriga o segundo maior parque solar centralizado do Brasil e lidera a geração eólica, respondendo por mais de 30% da produção nacional (EPE, 2025). Minas Gerais, pioneiro na geração distribuída, totaliza 4,5 GW de potência solar instalada e se mantém entre os maiores produtores do país. Juntos, os dois estados concentram cerca de um terço da geração solar nacional — 23,3% em Minas Gerais e 8,7% na Bahia —, o que reforça a importância de estudos regionais que considerem seus padrões climáticos e operacionais (EPE, 2025).

Apesar do avanço das renováveis, a intermitência e a variabilidade climática ainda representam desafios significativos para a previsibilidade e a estabilidade do sistema elétrico. A previsão assertiva da geração solar é essencial para garantir confiabilidade e reduzir custos operacionais associados à variabilidade atmosférica, sendo um requisito central para a integração segura de níveis crescentes de energia solar (Inman et al., 2013; Breyer et al., 2017). O Plano da Operação Energética 2025–2029 prevê que fontes renováveis, além de micro e minigeração distribuídas representem cerca de um terço da capacidade total instalada do país até 2029, o que exigirá modelos de previsão mais robustos e flexíveis (ONS, 2025).

A diversidade climática brasileira intensifica esses desafios, pois padrões de radiação, nebulosidade e precipitação variam entre regiões e estações do ano. Fenômenos como a Zona de Convergência Intertropical (ZCIT) e a Zona de Convergência do Atlântico Sul (ZCAS) influenciam de forma distinta o Norte–Nordeste e o Sudeste, gerando regimes sazonais contrastantes (Melo et al., 2024). Como a geração fotovoltaica é altamente sensível a essas condições, modelos baseados apenas em séries históricas apresentam limitações. Torna-se, portanto, fundamental incorporar variáveis meteorológicas — como cobertura de nuvens, temperatura e umidade — para representar de forma mais realista os fatores climáticos que afetam a geração solar (Antonanzas et al., 2016).

Diante disso, este trabalho tem como objetivo propor e avaliar metodologias que aprimorem a previsão da geração de energia solar em usinas localizadas nos estados da Bahia e de Minas Gerais — especificamente no Parque Solar Ituverava (BA) e no Complexo Solar Pirapora (MG). Utilizam-se modelos da família SARIMA, que capturam padrões sazonais e autocorrelacionados, e modelos SARIMAX, que incorporam variáveis exógenas associadas a condições atmosféricas e de superfície, incluindo parâmetros de absorção da radiação, cobertura e espessura de nuvens, refletância e temperatura da superfície, buscando reduzir a incerteza e refletir com maior fidelidade as oscilações provocadas por fenômenos climáticos.

O estudo busca responder a três questões principais:

- (i) em que medida a inclusão de variáveis exógenas meteorológicas melhora o desempenho preditivo dos modelos univariados de séries temporais aplicados à geração solar;
- (ii) quais combinações de variáveis e configurações de modelos apresentam melhor ajuste para o contexto climático e operacional das usinas analisadas;
- (iii) quais são os limites dos modelos SARIMA e SARIMAX diante de rupturas operacionais e mudanças estruturais na série de geração

O trabalho está organizado em seis capítulos. O Capítulo 1 apresenta a introdução, o contexto e os objetivos do estudo. O Capítulo 2 reúne o referencial teórico sobre usinas fotovoltaicas, previsão de geração e modelos ARIMA/SARIMA/SARIMAX. O Capítulo 3 caracteriza as usinas analisadas e descreve as variáveis meteorológicas utilizadas. O Capítulo 4 detalha a metodologia e os procedimentos de modelagem. O Capítulo 5 apresenta os resultados, as comparações e as projeções. Por fim, o Capítulo 6 reúne as conclusões do trabalho.

2. REFERENCIAL TEÓRICO

2.1. Usinas Fotovoltaicas

As usinas fotovoltaicas convertem diretamente a radiação solar em energia elétrica por meio do efeito fotovoltaico, fenômeno em que fótons incidentes sobre materiais semicondutores excitam elétrons, gerando uma diferença de potencial capaz de produzir corrente elétrica contínua (Guerra et al., 2018). As células solares, compostas por junções p-n, são agrupadas em módulos que formam arranjos conectados a inversores e sistemas de controle, responsáveis por converter a corrente contínua (CC) em corrente alternada (CA) para injeção na rede elétrica. Em sistemas isolados, a energia pode ser armazenada em baterias para uso posterior. A eficiência desses sistemas depende de fatores físicos, como irradiação solar e temperatura das células, que afeta diretamente a tensão e o rendimento energético, além de perdas elétricas nos inversores e na infraestrutura de conexão (Singh, 2013).

Além desses aspectos intrínsecos ao processo de conversão, o desempenho das usinas também é condicionado por fatores operacionais associados à integração ao Sistema Interligado Nacional (SIN). Entre esses fatores, destaca-se o *curtailment*, entendido como a restrição temporária da geração elétrica mesmo quando há disponibilidade de recurso solar. Esse mecanismo é utilizado pelo Operador Nacional do Sistema Elétrico (ONS) para garantir o equilíbrio carga-geração e preservar a estabilidade eletroenergética, especialmente em um contexto de forte expansão das fontes renováveis variáveis (ONS, 2025).

Segundo o Relatório Técnico RT ONS DGL 0189/2025, o aumento do *curtailment* no Brasil decorre principalmente de dois fatores: (i) a redução da capacidade de escoamento do sistema, resultante da atualização dos modelos matemáticos das usinas após a perturbação de 15 de agosto de 2023, que revelou discrepâncias entre o comportamento real e o modelado; e (ii) o rápido crescimento da capacidade instalada de fontes renováveis, concentrado no Nordeste, em ritmo superior ao aumento da carga do sistema. Entre 2021 e 2025, a potência instalada conjunta de usinas eólicas, fotovoltaicas centralizadas e micro e minigeração distribuída (MMGD) passou de cerca de 29 GW para quase 90 GW, um aumento de aproximadamente 200%, sem expansão proporcional da infraestrutura de transmissão (ONS, 2025).

Dentro desse total, a MMGD já representa mais de 38 GW em 2025 e deve atingir 58 GW até 2029, ampliando a complexidade da operação por estar fora do controle direto do ONS.

Em períodos de sobreoferta, as restrições recaem sobre as usinas centralizadas, já que a geração distribuída não é sujeita aos cortes, o que acentua a assimetria entre agentes (ONS, 2025).

As restrições à geração são classificadas em três categorias: indisponibilidade externa, confiabilidade elétrica e razão energética. Nos últimos anos, os cortes por sobreoferta cresceram significativamente, totalizando 4.330 GWh em 2024, com picos de 22.766 MWmed entre 9h e 16h, período de maior geração solar (ONS, 2025). Para 2029, o ONS projeta que 96% das restrições serão motivadas por excesso de geração renovável, podendo levar à perda de até 20% da energia fotovoltaica potencial (ONS, 2025).

A mitigação do *curtailment* requer a expansão da infraestrutura de transmissão, maior integração entre os sistemas de transmissão e distribuição e aprimoramento regulatório para o controle dos Recursos Energéticos Distribuídos (REDs). O ONS também recomenda o uso de compensadores síncronos e políticas que promovam flexibilidade operativa e armazenamento de energia (ONS, 2025).

Além disso, o fenômeno do *curtailment* representa um desafio relevante para a operação e o planejamento do sistema elétrico, pois resulta em perdas de energia e redução da eficiência das usinas solares. A análise de tal fenômeno deve ser conduzida de forma específica, uma vez que seus eventos refletem restrições operativas e não a disponibilidade real de recurso solar. Por isso, é necessário interpretar suas ocorrências em paralelo às previsões de geração, já que os cortes introduzem desvios que distorcem o comportamento natural da série. Assim, previsões acuradas, combinadas ao monitoramento sistemático dos eventos de *curtailment*, permitem diagnósticos mais completos sobre a geração efetiva e seus condicionantes operacionais.

2.2. Previsão de Geração de Energia Fotovoltaica

A previsão da geração de energia solar fotovoltaica tem se consolidado como um componente essencial para a integração eficiente das fontes renováveis ao sistema elétrico. A variabilidade intrínseca da irradiância solar e as rápidas flutuações causadas pela passagem de nuvens podem comprometer o equilíbrio entre oferta e demanda, gerar custos adicionais de reserva e aumentar o risco de *curtailment* (Antonanzas et al., 2016; Inman et al., 2013). Dessa forma, a elaboração de previsões confiáveis da geração fotovoltaica é fundamental tanto para o planejamento do despacho energético quanto para a otimização econômica e operacional das usinas (Bird et. al. 2014)

As previsões podem ser caracterizadas por três dimensões principais: o horizonte de previsão, que representa o intervalo de tempo entre o momento presente e o instante futuro a ser previsto; a resolução temporal, que indica a frequência com que as previsões são emitidas; e o intervalo de previsão, que corresponde à duração do período futuro coberto pelas estimativas (Antonanzas et al., 2016). Os métodos empregados evoluíram significativamente nas últimas décadas, passando de modelos puramente estatísticos baseados em séries históricas para abordagens híbridas que integram dados meteorológicos, modelagem física e aprendizado de máquina (Sobri et al., 2018).

De forma geral, as técnicas utilizadas para previsão da geração fotovoltaica podem ser agrupadas em três categorias principais:

- (1) Modelos físicos, que utilizam dados de irradiação e temperatura para simular o desempenho dos módulos e prever a potência resultante;
- (2) Modelos estatísticos e de aprendizado de máquina, que estimam a geração com base em padrões históricos de comportamento da série e variáveis meteorológicas associadas; e
- (3) Modelos híbridos, que combinam as duas abordagens anteriores, incorporando elementos de física e estatística (Antonanzas et al., 2016).

A Tabela 2.1 resume os principais métodos, dados de entrada e horizontes de previsão descritos na literatura especializada.

Tabela 2.1: Principais métodos e horizontes de previsão da geração fotovoltaica.

Horizonte temporal	Modelos mais utilizados na literatura	Resumo
Intra-hora	Modelo de persistência; modelos autorregressivos (AR, ARMA); redes neurais (ANN; k-NN); técnicas baseadas em vetores de movimento de nuvens (<i>Cloud Motion Vectors – CMV</i>); imagens de céu (<i>sky imagers</i>).	A persistência é amplamente usada em horizontes muito curtos; métodos estatísticos e imagens de céu trazem ganhos modestos.
Intra-dia	Modelos baseados em satélite; vetores de movimento de nuvens (CMV); modelos de previsão numérica do tempo (<i>Numerical Weather Prediction – NWP</i>); modelos híbridos.	Modelos de satélite são eficazes até 5 h; NWP apresentam melhor desempenho acima de 4 h; abordagens híbridas também são comuns.
Dia seguinte	Modelos baseados em NWP; modelos físicos (ex.: Solis, Ineichen, ESRA); SARIMA/ARIMA; pós-processamento estatístico; previsões por ensemble.	Uso de NWP e modelos de céu claro, ajustados por técnicas estatísticas e combinações de previsões.

Dois dias ou mais	Modelos de previsão numérica do tempo (NWP); modelos de séries temporais (ARIMA, SARIMA); regressões com termos de Fourier.	Uso de NWP e modelos de séries temporais e regressões sazonais para capturar tendências de longo prazo e variabilidade climática
-------------------	---	--

Fonte: Adaptado de Antonanzas et. al. (2016)

A literatura converge em destacar que não existe um modelo universalmente superior: o desempenho depende fortemente do horizonte de previsão, da resolução temporal e da qualidade dos dados disponíveis. Para horizontes muito curtos, modelos de persistência continuam sendo a referência básica; já para horizontes de horas ou dias, as abordagens híbridas e probabilísticas tendem a apresentar maior precisão. De modo geral, a combinação entre modelos físicos e estatísticos tem se mostrado a estratégia mais robusta, especialmente quando se utilizam dados meteorológicos de alta frequência e algoritmos de aprendizado de máquina para correção de viés (Antonanzas et al., 2016; Sobri et al., 2018).

Entre os métodos estatísticos, os modelos da família SARIMA (*Seasonal Autoregressive Integrated Moving Average*) têm sido amplamente empregados em aplicações de curto prazo para previsão da geração fotovoltaica. Esses modelos capturam os padrões de autocorrelação e a sazonalidade presentes nas séries históricas de irradiância e potência, ajustando componentes autorregressivos, integrados e de média móvel que refletem a dinâmica temporal do sistema.

Os modelos SARIMA são especialmente úteis quando a única informação disponível é a série de potência da usina, permitindo gerar previsões rápidas e de fácil interpretação. Apesar da simplicidade, esses modelos servem como referência em estudos comparativos, sendo frequentemente utilizados como base para avaliar métodos mais complexos. Além disso, há evidências de que a inclusão de variáveis exógenas tende a melhorar o desempenho desses modelos, ampliando sua capacidade preditiva (Antonanzas et al., 2016; Bacher et al., 2009). Abordagens autorregressivas foram algumas das primeiras aplicadas à previsão solar de curto prazo, por modelarem adequadamente a persistência e a autocorrelação das séries de irradiância e geração. Contudo, o desempenho tende a decair em horizontes superiores a poucas horas, devido à dificuldade em representar variações abruptas causadas por nuvens e fenômenos atmosféricos não lineares (Inman et al., 2013).

De forma complementar, os modelos ARIMA e SARIMA são classificados entre os métodos estatísticos clássicos, adequados para previsões de curto prazo em contextos com disponibilidade limitada de dados exógenos. Embora não capturem plenamente a natureza

estocástica e não linear da geração solar, o SARIMA mantém relevância prática por incorporar padrões sazonais diários e anuais de radiação, mostrando-se particularmente útil em séries com periodicidade bem definida — como a geração mensal ou diária de usinas solares (Sobri et al., 2018). Evoluções desses modelos, como o SARIMAX, ampliaram sua aplicabilidade ao permitir a inclusão de variáveis meteorológicas externas — irradiação, nebulosidade, temperatura — como covariáveis no processo autorregressivo, o que melhora seu desempenho e aproxima o modelo estatístico da dinâmica física do sistema sem perda de interpretabilidade (Antonanzas et al., 2016; Sobri et al., 2018).

Entretanto, observa-se que a literatura se concentra majoritariamente em previsões intradiárias ou horárias, com uso de dados meteorológicos de alta frequência. Nesse sentido, o presente estudo amplia essa perspectiva ao avaliar a aplicabilidade dos modelos SARIMA e SARIMAX em uma escala temporal mensal, incorporando médias mensais de variáveis atmosféricas e de superfície como preditores para estimar a geração fotovoltaica nos meses seguintes. Essa abordagem busca investigar se, mesmo com dados agregados, é possível capturar padrões relevantes de variabilidade climática e melhorar a acurácia das previsões.

Além da relevância operacional para usinas individuais, as previsões de geração fotovoltaica vêm assumindo um papel cada vez mais central no planejamento energético de médio prazo no Brasil, especialmente no contexto do despacho integrado hidro-termo-eólico-solar que caracteriza o Sistema Interligado Nacional (SIN). A rápida expansão das fontes renováveis, associada às limitações da infraestrutura de transmissão e ao avanço da micro e minigeração distribuída, tem elevado a complexidade da operação, tornando a previsão solar um insumo estratégico para as análises conduzidas pelo Operador Nacional do Sistema Elétrico (ONS).

Em horizontes mensais ou sazonais, estimativas acuradas de geração fotovoltaica apoiam decisões relacionadas à política de operação dos reservatórios hidrelétricos, à programação do despacho térmico e à avaliação das condições de segurança operativa. A antecipação de padrões de disponibilidade solar permite ao ONS ajustar a quantidade de água turbinada ou armazenada, otimizar o acionamento das usinas termelétricas e identificar, com maior antecedência, possíveis riscos de sobreoferta e eventos de *curtailment*, especialmente em regiões com elevada concentração de empreendimentos eólicos e solares, como o Nordeste.

Nesse contexto, modelos estatísticos como SARIMA e SARIMAX, quando aplicados a séries mensais, apresentam potencial para complementar as ferramentas utilizadas no

planejamento energético de médio prazo, fornecendo estimativas que podem ser integradas aos cenários do Plano Mensal de Operação (PMO). Assim, a presente abordagem contribui para aproximar métodos de previsão baseados em séries temporais das necessidades práticas do planejamento brasileiro, fortalecendo a capacidade de antecipação das condições de operação do SIN em um contexto de crescente participação da energia solar.

2.3. Modelos ARIMA

Os modelos da classe ARIMA, desenvolvidos por Box & Jenkins, constituem uma das abordagens mais consolidadas para modelagem e previsão de séries temporais univariadas. Esses modelos utilizam apenas os valores passados da própria série para estimar seu comportamento futuro (Box et al., 2015).

A formulação básica se apoia em dois componentes fundamentais. O modelo autorregressivo de ordem p , denotado por $\text{AR}(p)$, pressupõe que o valor atual da série z_t pode ser expresso como uma combinação linear de seus p valores passados acrescida de um termo de erro aleatório a_t . Por outro lado, o modelo de médias móveis de ordem q , $\text{MA}(q)$, relaciona o valor presente com o erro atual e com q erros passados. A combinação dessas duas estruturas origina o modelo $\text{ARMA}(p, q)$, adequado para séries estacionárias que não apresentam tendência ou sazonalidade significativa. Para que o ARMA seja válido, devem ser atendidas simultaneamente (i) a condição de estacionariedade, vinculada ao componente AR, garantindo que choques transitórios não provoquem explosão da série e (ii) a condição de invertibilidade, associada ao componente MA, assegurando que os erros passados possam ser representados de forma única por meio de valores presentes.

Quando a série apresenta tendência, níveis mutáveis ou não estacionariedade de ordem baixa, aplica-se o modelo $\text{ARIMA}(p, d, q)$, no qual o componente “Integrado” (I) corresponde às diferenciações sucessivas necessárias para tornar a série estacionária. Assim, o ARIMA representa a dinâmica temporal a partir da série diferenciada, que então pode ser modelada como ARMA.

No entanto, diversas séries temporais exibem dependência em lags sazonais, isto é, correlação significativa em múltiplos inteiros de um período s . Em séries mensais, por exemplo, é comum observar correlação elevada nos lags 12, 24, 36 etc. Nessas situações, a estrutura ARIMA não é suficiente para capturar padrões de repetição anual, sendo necessário incorporar uma sazonalidade estocástica, segundo a qual uma observação em determinado mês se relaciona

com observações do mesmo mês em anos anteriores (Morettin; Toloi, 1981). Esse comportamento motivou a extensão da formulação ARIMA para o modelo sazonal SARIMA, que permite representar simultaneamente dependências de curto prazo, efeitos de tendência e periodicidade.

2.3.1. Modelo Sazonal Autoregressivo Integrado de Médias Móveis (SARIMA)

O modelo SARIMA é uma extensão do ARIMA que incorpora tanto componentes sazonais quanto não sazonais, permitindo representar séries com dependência temporal e padrões periódicos. Ele é indicado para séries estacionárias ou que podem ser tornadas estacionárias por diferenciação. O modelo é identificado pelos parâmetros $(p, d, q) \times (P, D, Q)_s$, em que:

- p, d, q são os parâmetros do componente não sazonal (AR, diferenciação e MA);
- P, D, Q são os parâmetros do componente sazonal (AR sazonal, diferenças sazonais e MA sazonal);
- s é a periodicidade da sazonalidade, ou seja, o número de períodos que define a repetição do padrão (por exemplo, $s = 12$ para séries mensais com padrão anual).

Formalmente, seja z_t a série original, e $W_t = \nabla^d \nabla_s^D z_t$ a série diferenciada d vezes não sazonalmente e D vezes sazonalmente, onde o operador de diferença sazonal é definido como:

$$\nabla_s z_t = z_t - z_{t-s} = (1 - B^s)z_t \quad (2.1)$$

Então, a série W_t pode ser modelada como ARMA com componentes sazonais:

$$\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)a_t \quad (2.2)$$

ou, substituindo $W_t = \nabla^d \nabla_s^D z_t$, obtém-se a forma completa do SARIMA:

$$\phi(B)\Phi(B^s)\nabla^d \nabla_s^D z_t = \theta(B)\Theta(B^s)a_t \quad (2.3)$$

onde:

- $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ é o operador autorregressivo não sazonal;
- $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ é o operador de médias móveis não sazonal;
- $\Phi(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})$ é o operador autorregressivo sazonal;

- $\Theta(B^S) = (1 - \Theta_1 B^S - \dots - \Theta_Q B^{Qs})$ é o operador de médias móveis sazonal;
- a_t é ruído branco, com média zero e variância constante.

A aplicação das diferenças d e D garante que W_t seja estacionária, permitindo modelar tanto os efeitos sazonais quanto os não sazonais de maneira consistente (Box et al., 2015).

Além disso, em várias aplicações práticas, parte da variação da série não é explicada apenas pela sua estrutura interna, mas também por fatores externos observáveis. Para esses casos, o modelo SARIMAX amplia o SARIMA ao incluir variáveis exógenas, permitindo capturar efeitos adicionais de caráter explicativo.

2.3.2. Modelo Sazonal Autoregressivo Integrado de Médias Móveis com Variáveis Exógenas (SARIMAX)

O SARIMAX é uma extensão do modelo SARIMA que permite incluir uma ou mais variáveis exógenas (x_{it}) para explicar parte da variação da série z_t . Em termos gerais, o modelo combina os componentes autorregressivos e de médias móveis, tanto sazonais quanto não sazonais, com termos adicionais que representam o impacto das variáveis externas sobre a variável dependente z_t . Sua formulação pode ser expressa como:

$$\phi(B) \Phi(B^S) \nabla^d \nabla_s^D z_t = \theta(B) \Theta(B^S) a_t + \sum_{i=1}^k \beta_i x_{it} \quad (2.4)$$

onde:

- x_{it} é a variável exógena i no tempo t , com $i = 1, \dots, k$;
- β_i é o coeficiente associado à variável exógena;
- s é a periodicidade da sazonal;
- $\phi(B)$ e $\theta(B)$ são os operadores AR e MA não sazonais;
- $\Phi(B^S)$ e $\Theta(B^S)$ são os operadores AR e MA sazonais;
- a_t é o termo de erro aleatório (ruído branco).

Diferentemente de uma regressão clássica, os coeficientes β_i não podem ser interpretados isoladamente, como o efeito direto de um aumento unitário em x_{it} , já que seu efeito depende da defasagem da série z_t (Hyndman, 2010). Assim, o impacto de uma variável exógena não é imediato nem constante ao longo do tempo, dependendo do comportamento dinâmico da série.

De forma alternativa, Hyndman (2010) propôs representar o modelo SARIMAX como uma regressão com erros SARIMA, onde os resíduos da regressão seguem uma estrutura autorregressiva e de médias móveis. Assim, tem-se:

$$z_t = \sum_{i=1}^k \beta_i x_{it} + \eta_t \quad (2.5)$$

$$\eta_t = \phi_1 \eta_{t-1} + \cdots + \phi_p \eta_{t-p} + \Phi_1 \eta_{t-s} + \cdots + \Phi_P \eta_{t-ps} - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} - \Theta_1 a_{t-s} - \cdots - \Theta_Q a_{t-Qs} \quad (2.6)$$

Nessa formulação, existem dois termos de erro distintos: o erro da regressão (η_t) e o erro do modelo SARIMA (a_t). Assume-se que apenas o termo a_t segue um processo de ruído branco.

Utilizando o operador de defasagem B , o modelo SARIMAX pode ser reescrito de forma compacta como:

$$z_t = \sum_{i=1}^k \beta_i x_{it} + \frac{\theta(B)\Theta(B^S)}{\phi(B)\Phi(B^S)\nabla^d\nabla_S^D} a_t \quad (2.7)$$

Por fim, é fundamental ressaltar que, antes de estimar o modelo, deve-se verificar a estacionaridade tanto da série dependente z_t quanto das variáveis exógenas x_{it} . Caso alguma delas não seja estacionária, os coeficientes estimados podem se tornar inconsistentes e comprometer a validade das inferências.

Dado que as séries analisadas neste estudo podem apresentar tanto efeitos sazonais quanto influências externas, as formulações SARIMA e SARIMAX constituem alternativas adequadas para a modelagem adotada.

3. CARACTERIZAÇÃO DAS USINAS E VARIÁVEIS DO ESTUDO

3.1. Usinas do Estudo

Para o presente estudo, foram selecionadas duas usinas solares localizadas em regiões distintas do Brasil: o Parque Solar de Ituverava, na Bahia, e o Complexo Solar Pirapora, em Minas Gerais (Figura 3.1). A escolha desses empreendimentos permite analisar padrões de geração em diferentes contextos geográficos e climáticos, considerando tanto a capacidade instalada quanto o perfil de radiação solar de cada local.



Figura 3.1: Localização das usinas

Fonte: Elaboração própria (2025)

3.1.1. Parque Solar de Ituverava (BA)

O Parque Solar de Ituverava está localizado no município de Tabocas do Brejo Velho, a cerca de 800 km de Salvador. Em operação desde setembro de 2017, é um dos maiores parques

solares da América do Sul, ocupando uma área de aproximadamente 579 hectares (equivalente a 700 campos de futebol). O empreendimento possui 850 mil módulos fotovoltaicos, com capacidade instalada de 254 MW, gerando cerca de 550 GWh por ano, energia suficiente para abastecer aproximadamente 268 mil famílias e evitar a emissão de 318 mil toneladas de CO₂ anualmente (Enel Green Power, 2025).

Além da geração de energia, o parque incorpora práticas de sustentabilidade e economia circular, como reaproveitamento de água da chuva e de unidades de resfriamento, compostagem de resíduos orgânicos e programas de restauração da vegetação nativa. A usina também desenvolve iniciativas socioambientais, incluindo a instalação de poleiros e ninhos artificiais para a fauna local e projetos sociais, como a doação de painéis solares obsoletos para fornecer energia a uma bomba de água utilizada por famílias da comunidade local (Enel Green Power, 2025).

3.1.2. Complexo Solar Pirapora (MG)

O Complexo Solar Pirapora, localizado no norte de Minas Gerais, é um dos maiores empreendimentos fotovoltaicos da América Latina. Instalado em uma área de cerca de 800 hectares, o complexo abriga mais de um milhão de painéis solares que acompanham o movimento do sol, maximizando a captação da irradiação ao longo do dia. O projeto é operado pela EDF Renewables em parceria com a Canadian Solar, que também fabrica os módulos fotovoltaicos utilizados (G1, 2017).

Com capacidade instalada total de 321 MW, o complexo tem papel relevante na diversificação da matriz energética brasileira, sendo capaz de abastecer cerca de 420 mil residências. Em 2024, foi anunciada uma futura expansão de 100 MW, que elevará a capacidade total para 421 MW, com investimento de aproximadamente R\$ 1 bilhão, liderado pelas mesmas empresas (Cenário Energia, 2024). A nova fase prevê o uso de painéis bifaciais, capazes de captar luz em ambos os lados e aumentar a eficiência do sistema.

A operação do complexo evita a emissão de aproximadamente 350 mil toneladas de CO₂ por ano e gera impactos socioeconômicos positivos, com a criação de mais de 2 mil empregos diretos e indiretos durante o processo de expansão. Também são desenvolvidas ações de responsabilidade social voltadas à educação ambiental e ao fortalecimento da infraestrutura local, alinhando o empreendimento aos princípios de sustentabilidade e desenvolvimento regional (Cenário Energia, 2024).

3.2. Variáveis Exógenas Climáticas

A geração de energia em sistemas fotovoltaicos é fortemente condicionada por fatores atmosféricos e de superfície, como irradiação solar, cobertura de nuvens e temperatura da superfície terrestre. Por isso, o uso de variáveis climáticas como regressoras exógenas em modelos de previsão tem se mostrado essencial para capturar a variabilidade sazonal e as flutuações de curto prazo associadas à radiação solar.

De acordo com Stock e Watson (2020), uma variável exógena é aquela determinada fora do modelo, ou seja, não correlacionada com o termo de erro, ao contrário das variáveis endógenas, que são determinadas dentro da própria estrutura do modelo. A inclusão de variáveis exógenas permite que parte das variações históricas observadas na série seja explicada por fatores externos, levando a previsões mais precisas (Hyndman et al., 2021).

No contexto deste trabalho, foram selecionadas variáveis atmosféricas e de superfície diretamente relacionadas ao balanço radiativo e aos processos que determinam a geração fotovoltaica. Cada uma delas é detalhada a seguir, com ênfase em seus fundamentos físicos e na forma como influenciam a variabilidade da produção de energia.

3.2.1. Albedo da superfície (ALBEDO)

O albedo de superfície é a razão entre a irradiação solar refletida pela superfície terrestre e a irradiação total incidente, sendo uma das principais variáveis do balanço de energia da Terra (Lu et al., 2021). Essa propriedade óptica expressa a fração da radiação solar devolvida à atmosfera e controla a energia efetivamente absorvida pela superfície, influenciando processos como aquecimento do solo, evapotranspiração e balanço radiativo (Zhang et al., 2022).

O albedo apresenta forte variabilidade espacial e temporal em função de fatores como tipo de solo, umidade, vegetação e uso do solo (Lu et al., 2021). Alterações nesses componentes modificam o equilíbrio entre absorção e reflexão da radiação solar: albedos mais altos reduzem o aquecimento superficial, enquanto albedos mais baixos aumentam a absorção de energia (Zhang et al., 2022). Por sua relevância climática, é considerado pela Organização Meteorológica Mundial um parâmetro essencial de observação, amplamente monitorado por sensoriamento remoto (Lu et al., 2021).

No contexto brasileiro, o albedo é variável-chave em modelos de balanço de energia, como o SEBAL, utilizados para estimar radiação líquida e fluxos de calor em regiões tropicais

(Angelini et al., 2021). Pequenas variações nesse parâmetro podem gerar diferenças significativas em evapotranspiração e temperatura de superfície. Conceitualmente, áreas mais claras refletem mais radiação, parte da qual pode ser aproveitada pelos módulos fotovoltaicos. Assim, sob certas condições de inclinação e orientação, maiores valores de albedo podem favorecer a geração de energia solar, ceteris paribus (Zhang et al., 2022; Angelini et al., 2021). Além disso, albedos mais altos geralmente indicam maior incidência de radiação naquele instante, reforçando sua relação com períodos de maior irradiação.

3.2.2. Fração total da área coberta por nuvens (CLDTOT)

A variável CLDTOT (*total cloud area fraction*) representa a fração da área horizontal coberta por nuvens em determinado instante ou período, sendo definida como a razão entre a área coberta e a área total observada (Ding et al., 2023). É amplamente utilizada em produtos de reanálise e sensoriamento remoto para quantificar o grau de cobertura do céu e avaliar os efeitos das formações de nuvens sobre o balanço radiativo da atmosfera.

A cobertura total de nuvens exerce influência direta na radiação solar incidente, pois as nuvens refletem e absorvem parte da radiação de onda curta proveniente do sol, reduzindo a irradiação direta e modificando a componente difusa (Shuvalova, 2023). Assim, variações na fração de nuvens provocam oscilações significativas na energia solar disponível à superfície, afetando de forma imediata a geração fotovoltaica.

De forma empírica, Sarkar (2016) identificou uma relação não linear entre a fração de nuvens e a razão entre a radiação solar global e a extraterrestre, mostrando que o aumento da cobertura de nuvens reduz a irradiação incidente. O autor demonstrou ainda que, na ausência de medições diretas, é possível estimar a radiação solar a partir do grau de nebulosidade.

3.2.3. Radiação de onda longa absorvida pela superfície (LWGAB)

A radiação de onda longa absorvida pela superfície (LWGAB) é uma das componentes do balanço de radiação terrestre, representando o ganho líquido de energia térmica da superfície a partir da diferença entre a radiação emitida pela atmosfera e a emitida pela própria superfície (Wang & Dickinson, 2013). Essa variável reflete diretamente as trocas de energia entre a Terra e a atmosfera e influencia processos de evapotranspiração, convecção e aquecimento superficial (Zeng & Cheng, 2021).

A LWGAB é determinada pela temperatura da superfície, pela emissividade do solo e pelo conteúdo atmosférico de vapor d'água e nuvens, que modulam a emissão e absorção de

radiação infravermelha. Em regiões quentes e úmidas, grande parte dessa emissão provém das camadas atmosféricas mais próximas do solo, tornando a LWGAB especialmente sensível às condições locais (Wang & Dickinson, 2013). Produtos satelitais, como o *Global Land Surface Satellite* (GLASS), permitem estimar essa radiação com alta precisão e detalhamento espacial (Zeng & Cheng, 2021).

No contexto da geração solar, a LWGAB pode se associar ao desempenho fotovoltaico por mecanismos distintos. Por um lado, valores mais altos indicam maior energia térmica retida na superfície, o que tende a elevar a temperatura dos módulos e reduzir a eficiência das células solares (Wang & Dickinson, 2013; Zeng & Cheng, 2021). Por outro, fluxos elevados de radiação de onda longa absorvida também refletem maior carga radiativa atmosférica naquele instante, associando-se a condições de maior entrada de energia no sistema.

3.2.4. Espessura óptica das nuvens (TAUTOT)

A espessura óptica das nuvens (*cloud optical thickness*) é uma variável radiativa que expressa a capacidade de uma nuvem em atenuar ou dispersar a radiação solar incidente, sendo proporcional ao coeficiente de extinção integrado verticalmente. Valores mais altos indicam maior absorção e espalhamento da luz dentro da nuvem (Yi et al., 2025). A variável TAUTOT, disponível em produtos de reanálise como o MERRA-2, representa a espessura óptica total de todas as nuvens — líquidas e sólidas — em diferentes altitudes, refletindo de forma abrangente o impacto das formações de nuvens sobre a transmissão da radiação solar até a superfície (Barry et al., 2023).

Fisicamente, a espessura óptica influencia a razão entre radiação direta e difusa. Nuvens finas aumentam a radiação espalhada, enquanto nuvens espessas reduzem a irradiância global ao refletirem parte significativa da radiação para o espaço (Yi et al., 2025). Dessa forma, variações na TAUTOT são determinantes da atenuação da irradiância superficial e, consequentemente, do potencial de geração fotovoltaica.

Estudos recentes mostram que o aumento da espessura óptica das nuvens está associado à redução exponencial da radiação solar recebida, conforme a Lei de Beer–Lambert, confirmado seu valor como variável preditora robusta em modelos de previsão de irradiância e potência solar (Wandji Nyamsi & Lindfors, 2025). Assim, quanto maior o valor de TAUTOT, menor tende a ser a radiação solar incidente e, portanto, a geração fotovoltaica - *ceteris paribus* (Barry et al., 2023).

3.2.5. Temperatura da pele da superfície (TS)

A temperatura da pele da superfície (*surface skin temperature – TS*) representa a temperatura radiativa efetiva da interface entre a superfície terrestre e a atmosfera, sendo um dos principais parâmetros do balanço de energia da Terra. Essa variável reflete o estado térmico da camada mais externa do solo, vegetação ou infraestrutura urbana, resultante dos fluxos de energia absorvidos e emitidos pela superfície (Jin, 2010). A TS é obtida por sensores térmicos a bordo de satélites — como o MODIS — que captam a radiação infravermelha emitida pela superfície, permitindo estimativas contínuas da temperatura terrestre.

Além de sua importância climática, a TS é relevante para a análise de sistemas fotovoltaicos. Estudos indicam que a instalação de usinas solares altera o balanço energético local: ao comparar 23 usinas de grande escala, Zhang & Xu (2020) observaram uma redução média de 0,53 °C na temperatura superficial após a instalação dos painéis, atribuída à conversão de energia solar em eletricidade e ao sombreamento das placas. Esses resultados evidenciam a sensibilidade da TS às variações de absorção e emissão de energia em áreas fotovoltaicas.

No contexto da geração fotovoltaica, a TS também pode se relacionar ao desempenho dos sistemas por mecanismos distintos. Por um lado, temperaturas de pele mais altas indicam maior energia solar incidente e absorvida pela superfície naquele instante, o que costuma estar associado a condições de maior irradiância e, portanto, maior potencial de geração. Por outro, o aumento da temperatura da superfície tende a aquecer os módulos, reduzindo a eficiência das células solares e podendo limitar parte do ganho energético (Zhang & Xu, 2020).

4. METODOLOGIA

A metodologia adotada neste trabalho segue um fluxo estruturado que organiza, de forma clara, todas as fases da construção dos modelos e da geração das previsões. O fluxograma a seguir sintetiza esse percurso e orienta o desenvolvimento das próximas subseções.



Figura 4.1: Fluxograma da Metodologia

Fonte: Elaboração própria (2025)

A implementação computacional das etapas de obtenção, tratamento e agregação dos dados foi realizada no ambiente Python (versão 3.11.5). Nesse ambiente, foram desenvolvidas rotinas para automatizar o download das bases, a filtragem espacial dos dados, a agregação temporal para frequência mensal e a consolidação das séries utilizadas nas etapas posteriores de modelagem.

A etapa de modelagem estatística e avaliação dos modelos foi conduzida no ambiente RStudio, utilizando o R (versão 4.4.3), que oferece amplo suporte à estimação, diagnóstico e validação de modelos SARIMA e SARIMAX.

4.1. Bases de dados e pré-processamento

Os dados de geração de energia fotovoltaica foram obtidos junto ao Operador Nacional do Sistema Elétrico (ONS), por meio da plataforma pública de dados abertos, que disponibiliza informações de geração horária por usina (ONS, 2025). As séries horárias foram agregadas para frequência mensal por meio da soma dos valores de geração de cada mês. Essa escolha metodológica está associada ao objetivo de destacar padrões sazonais e tendências de médio e longo prazo, relevantes para fins de planejamento estratégico.

Além disso, a agregação mensal atenua as flutuações pontuais típicas da escala horária, influenciadas por variações meteorológicas transitórias ou indisponibilidades operacionais de curta duração, conferindo maior estabilidade estatística às séries. Por esse motivo, não foi necessário aplicar procedimentos adicionais de detecção e remoção de outliers, uma vez que a agregação mensal já reduz substancialmente a influência de valores extremos.

No caso das usinas analisadas, os primeiros registros datam de meados de 2017. Contudo, optou-se por iniciar a série em maio de 2018, uma vez que os meses iniciais

correspondem ao período de comissionamento das plantas, em que a geração ainda não refletia a plena utilização da capacidade instalada. A Figura 4.2 ilustra esse comportamento inicial tanto no Parque Solar de Ituverava quanto no Complexo Solar Pirapora, evidenciando a instabilidade da produção até a consolidação do regime normal de operação.

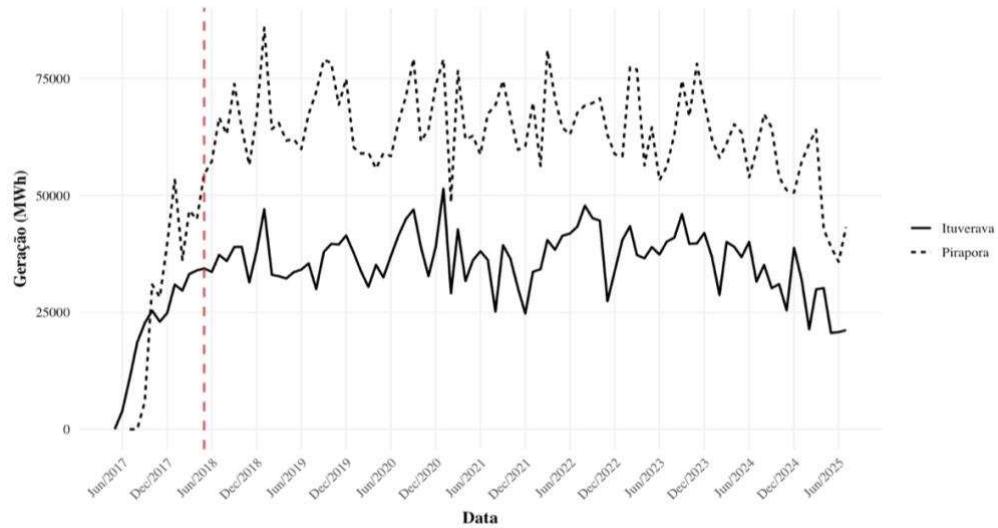


Figura 4.2: Linha de corte dos dados de geração de energia

Fonte: Elaboração própria (2025)

Para obter as variáveis exógenas utilizadas na modelagem, foi necessário recorrer a fontes climáticas que oferecessem séries históricas completas, consistentes e espacialmente contínuas. Diante da cobertura irregular de medições locais nas regiões das usinas, optou-se pelo uso de dados de reanálise, que combinam observações meteorológicas de satélites, estações de superfície e radares com modelos numéricos de previsão do tempo, produzindo estimativas contínuas, homogêneas e fisicamente coerentes ao longo do tempo (Araujo et al., 2024).

Entre as principais reanálises disponíveis, destacam-se o ERA5 e o ERA5-Land (do *European Centre for Medium-Range Weather Forecasts – ECMWF*), o CFSv2 (*Climate Forecast System*, da NOAA) e o MERRA-2 (*Modern-Era Retrospective analysis for Research and Applications*, da NASA). Conforme demonstrado por Araujo et al. (2024), em um estudo que comparou o desempenho dessas quatro bases em 35 localidades brasileiras, o MERRA-2 apresentou o menor erro médio quadrático e as maiores correlações com medições do INMET em 90% das estações analisadas. Além desse desempenho superior, o MERRA-2 oferece séries temporais contínuas desde 1980, com frequência horária e resolução espacial de aproximadamente $0,5^\circ$ latitude \times $0,625^\circ$ longitude, características que reforçam sua adequação

para estudos de previsão de geração fotovoltaica. Por esses motivos, o MERRA-2 foi adotado como fonte das variáveis exógenas neste trabalho.

A partir disso, foram selecionadas variáveis atmosféricas e de superfície consideradas relevantes para a modelagem da geração fotovoltaica: albedo da superfície (ALBEDO), fração total da área coberta por nuvens (CLDTOT), radiação de onda longa absorvida pela superfície (LWGAB), espessura óptica das nuvens (TAUTOT) e temperatura da pele da superfície (TS).

Para associar os dados meteorológicos às usinas analisadas, foram identificados os pontos de grade do MERRA-2 mais próximos às coordenadas geográficas de cada instalação. Essa correspondência foi realizada por meio do cálculo da distância geodésica entre cada célula da grade e as coordenadas reais das usinas, utilizando a fórmula de Haversine.

A fórmula de Haversine é amplamente empregada em aplicações de navegação e análises geoespaciais por estimar a menor distância entre dois pontos na superfície terrestre a partir de suas coordenadas geográficas (latitude e longitude). Segundo Mahmoud e Akkari (2016), essa equação calcula a distância de grande círculo, assumindo a Terra como uma esfera perfeita, aproximação que simplifica os cálculos e apresenta erro máximo inferior a 0,3%, considerado desprezível em escalas regionais. Embora a Terra seja, na realidade, um esferoide oblato, ou seja, ligeiramente achatado nos polos, essa representação esférica é suficientemente precisa e computacionalmente mais eficiente do que métodos mais complexos, como o de Vincenty, que utiliza um modelo elipsoidal. A equação geral da Haversine é expressa como:

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\varphi}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right) \quad (4.1)$$

em que:

- d é a distância entre dois pontos na superfície terrestre;
- r é o raio médio da Terra (≈ 6317 km);
- φ_1 e φ_2 são as latitudes dos pontos em radianos;
- $\Delta\varphi$ e $\Delta\lambda$ representam, respectivamente, as diferenças de latitude e longitude entre os pontos, em radianos.

Após o cálculo das distâncias, selecionou-se o ponto de grade com a menor distância para cada usina. Essa aproximação é considerada adequada para estudos desta natureza, considerando que a resolução espacial do MERRA-2 não permite uma correspondência exata entre coordenadas pontuais e os centros de grade, e que as distâncias envolvidas são

suficientemente pequenas para representar de forma confiável as condições meteorológicas locais.

Assim como os dados de geração do ONS, as séries meteorológicas também foram convertidas para frequência mensal, calculando-se a média dos valores horários para cada mês, de modo a suavizar oscilações de curto prazo e destacar a sazonalidade. Essa estratégia metodológica favorece a comparabilidade entre bases de dados distintas e contribui para análises mais robustas do ponto de vista estatístico.

4.2. Metodologia Box & Jenkins

Com as séries de geração e das variáveis meteorológicas já consolidadas em frequência mensal e submetidas ao pré-processamento necessário, procede-se a etapa da modelagem das séries. Para isso, adotou-se a metodologia Box & Jenkins, um procedimento empírico, sistemático e iterativo para construção de modelos SARIMA e variantes com variáveis exógenas, visando identificar um modelo parcimonioso e estatisticamente consistente, capaz de representar a dinâmica estocástica da série e produzir previsões confiáveis (Box et al., 2015).

Os autores destacam que o processo é cíclico e envolve iterações até que um modelo adequado seja alcançado. Essa abordagem converge com o roteiro de Hyndman e Athanasopoulos (2021). Com base nesses referenciais, as etapas adotadas são as seguintes:

4.2.1. Análise exploratória e transformação da série

Inicialmente, a série temporal é plotada e inspecionada visualmente para identificar padrões sazonais, tendências, quebras estruturais e observações atípicas. Em seguida, avalia-se a necessidade de transformações estabilizadoras de variância, como Box-Cox ou logarítmica ($\ln y_t$), utilizadas quando há indícios de heterocedasticidade. No presente estudo, a decisão foi guiada pela inspeção visual da série e pela análise do comportamento da variância, não sendo identificada a necessidade de transformação adicional.

Nessa etapa também são tratados outliers e eventuais descontinuidades que possam distorcer o comportamento da série. Os dados originais de geração, disponibilizados em frequência horária, foram agregados para a frequência mensal por meio da soma. Esse procedimento atua como um mecanismo natural de suavização, reduzindo a influência de flutuações pontuais, falhas momentâneas de medição ou interrupções operacionais de curta duração. Dessa forma, não foi necessária a aplicação de métodos adicionais de correção ou exclusão de outliers.

4.2.2. Estacionarização e diferenciação

Em seguida, é preciso testar a estacionariedade da série para garantir que o modelo SARIMA seja aplicável. Neste trabalho, utilizou-se o teste de Dickey–Fuller Aumentado (ADF), com constante, baseado no modelo:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t \quad (4.2)$$

em que $\Delta y_t = y_t - y_{t-1}$, p corresponde ao número de defasagens adicionais incluídas para eliminar autocorrelação serial dos resíduos, e o termo constante α permite acomodar séries cuja média é diferente de zero.

As hipóteses do teste são:

$$\begin{cases} H_0: \gamma = 0 & (\text{a série possui raiz unitária e é não estacionária}) \\ H_1: \gamma < 0 & (\text{a série não possui raiz unitária e é estacionária}) \end{cases}$$

O coeficiente γ é estimado por mínimos quadrados ordinários, e a estatística de teste é calculada pela razão entre o estimador $\hat{\gamma}$ e seu erro padrão:

$$\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (4.3)$$

Se o teste indicar não estacionariedade, aplica-se a diferenciação regular ou sazonal da série até que suas propriedades estatísticas se tornem aproximadamente constantes ao longo do tempo. A diferenciação regular consiste na subtração do valor imediatamente anterior:

$$\Delta y_t = y_t - y_{t-1} \quad (4.4)$$

sendo utilizada para remover tendências de longo prazo. Já a diferenciação sazonal é aplicada quando há padrões que se repetem em intervalos fixos e consiste na subtração do valor observado em um período sazonal anterior:

$$\Delta y_t = y_t - y_{t-s} \quad (4.5)$$

em que s representa o período da sazonalidade. A escolha entre diferenciação regular, sazonal ou a combinação de ambas é guiada pelos resultados dos testes de estacionariedade e pela análise dos gráficos da série e de suas funções de autocorrelação.

4.2.3. Identificação do modelo e estimação dos parâmetros

Com a série estacionarizada, examinam-se as Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) para determinar a estrutura do modelo e identificar ordens

adequadas de componentes autorregressivos e de médias móveis — (p, d, q) para a parte regular e $(P, D, Q)_s$ para a parte sazonal.

Computacionalmente, embora a função `auto.arima()` do pacote `forecast` (Hyndman & Khandakar, 2008) seja amplamente utilizada para automatizar essa etapa, ela não garante a exploração completa do espaço paramétrico quando se deseja testar, de forma sistemática e controlada, todas as combinações plausíveis de modelos. Esse ponto é especialmente relevante em contextos que exigem comparar modelos sob o mesmo conjunto de restrições e assegurar a comparabilidade direta dos critérios de informação entre todas as alternativas consideradas. Por essa razão, este trabalho adotou uma busca sistemática controlada, aplicando manualmente uma grade de modelos definida com base na interpretação da FAC.

Com a grade definida, cada modelo foi estimado por máxima verossimilhança (Maximum Likelihood Estimation – MLE), método que busca os valores dos coeficientes que maximizam a probabilidade de observação dos dados sob o modelo proposto. A estimação por MLE fornece, para cada modelo, o valor da log-verossimilhança (ℓ), que representa o quanto bem aquela especificação explica a série observada.

A partir da log-verossimilhança e do número de parâmetros estimados (k), calculou-se o Akaike Information Criterion corrigido (AICc) como critério de seleção, recomendado em amostras finitas por penalizar modelos excessivamente complexos e favorecer estruturas mais parcimoniosas:

$$AIC = -2\ell + 2k \quad (4.6)$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (4.7)$$

em que n é o tamanho da amostra. Entre os modelos estimados, adotou-se sempre aquele que apresentou o menor valor de AICc, por oferecer a melhor relação entre qualidade de ajuste e complexidade do modelo.

4.2.4. Diagnóstico e validação do modelo

4.2.4.1 Diagnóstico dos resíduos

Após a estimação, é necessário verificar se os resíduos do modelo são compatíveis com o comportamento esperado em um processo bem especificado, isto é, ausência de autocorrelação e distribuição aproximadamente simétrica. Para isso, neste trabalho foram aplicados dois testes formais: Ljung–Box e Jarque–Bera.

O teste de Ljung–Box, proposto por *Ljung & Box (1978)*, avalia simultaneamente as autocorrelações dos resíduos até um número predefinido de defasagens m . Um modelo SARIMA adequadamente especificado deve produzir resíduos compatíveis com um ruído branco, isto é, sem autocorrelação significativa.

As hipóteses são dadas por:

$$\begin{cases} H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0 & (\text{não há autocorrelação nos resíduos}) \\ H_1: \exists k \leq m \text{ tal que } \rho_k \neq 0 & (\text{há autocorrelação remanescente}) \end{cases}$$

A estatística de teste é definida por:

$$Q = n(n+2) \sum_{k=1}^m \frac{|\hat{\rho}_k|^2}{n-k} \quad Q \sim \chi^2(m) \quad (4.8)$$

em que n é o tamanho da amostra, $\hat{\rho}_k$ é a autocorrelação amostral dos resíduos no lag k e m é o número de defasagens avaliadas

O teste de Jarque–Bera, proposto por *Jarque & Bera (1980)*, avalia se a assimetria (S) e a curtose (K) dos resíduos são compatíveis com os valores esperados sob normalidade ($S = 0$ e $K = 3$). Embora a normalidade não seja uma exigência estrita para fins de previsão, ela é importante no diagnóstico Box–Jenkins, pois garante que a estrutura do modelo foi adequadamente especificada e possibilita a comparação homogênea entre diferentes modelos.

As hipóteses do teste são:

$$\begin{cases} H_0: S = 0 \text{ e } K = 3 & (\text{os resíduos seguem distribuição normal}) \\ H_1: S \neq 0 \text{ ou } K \neq 3 & (\text{há violação da normalidade}) \end{cases}$$

A estatística de teste é dada por:

$$JB = \frac{n}{6} \left[S^2 + \frac{(K-3)^2}{4} \right] \quad JB \sim \chi^2(2) \quad (4.9)$$

Caso o modelo não seja aprovado em qualquer um dos testes ele é descartado. Dessa forma, o melhor modelo selecionado é aquele que combina menor AICc com aprovação simultânea nos testes de Ljung–Box e Jarque–Bera.

4.2.4.2 Validação fora da amostra

Para avaliar a capacidade preditiva dos modelos em períodos não utilizados na estimação, adotou-se uma estratégia de validação fora da amostra baseada em duas janelas temporais de teste, cada uma com duração de 12 meses, correspondendo a um ciclo sazonal completo (Tabela 4.1).

Para cada janela de teste, utilizou-se como conjunto de treinamento todo o histórico anterior. Essa estratégia garante que os modelos sejam estimados nas mesmas condições, permitindo uma comparação direta tanto no ajuste in-sample quanto na sua capacidade preditiva.

Tabela 4.1: Configuração das janelas de modelagem

<i>Janela</i>	<i>In-sample</i>	<i>Out-of-sample</i>
1	mai/2018-jul/2023	ago/2023-jul/2024
2	mai/2018-jul/2024	ago/2024-jul/2025

Fonte: Elaboração própria (2025)

A avaliação das previsões nas janelas de teste foi realizada por meio de duas métricas amplamente utilizadas na literatura de previsão de geração fotovoltaica: o Mean Absolute Percentage Error (MAPE) e o Root Mean Square Error (RMSE), definidas respectivamente por:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (4.10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.11)$$

em que y_i são os valores observados, \hat{y}_i os valores previstos pelo modelo e n o número de observações no período de teste.

4.2.5. Previsão e aplicação

Com o modelo validado, procede-se à geração de previsões. No contexto deste estudo, os modelos SARIMA e SARIMAX são aplicados à previsão da geração mensal de energia fotovoltaica nas usinas de Ituverava e Pirapora.

5. RESULTADOS E DISCUSSÕES

Nesta seção, são apresentados e discutidos os resultados obtidos a partir da aplicação da metodologia proposta às séries de geração das duas usinas analisadas. Cada etapa — desde a modelagem até a avaliação preditiva — foi conduzida separadamente para cada usina, permitindo comparar o desempenho dos modelos considerados e identificar aquele com melhor capacidade preditiva para as previsões futuras.

5.1. Análise exploratória e transformação da série

5.1.1. Geração de energia fotovoltaica

As séries históricas de geração mensal de energia fotovoltaica das usinas de Ituverava (BA) e Pirapora (MG) apresentam oscilações irregulares ao longo do tempo, com variações de amplitude entre meses e anos (Figura 5.1). Embora seja possível identificar padrões recorrentes de aumento e redução da produção, esses ciclos não seguem uma periodicidade rigidamente definida, indicando sazonalidade moderada ou variável. Esse comportamento é característico de usinas em regiões tropicais, onde a distribuição anual da radiação solar é relativamente uniforme, e as flutuações de geração são influenciadas principalmente por condições meteorológicas transitórias, como cobertura de nuvens, precipitação e temperatura ambiente, além de fatores operacionais, como manutenções e degradação dos módulos.

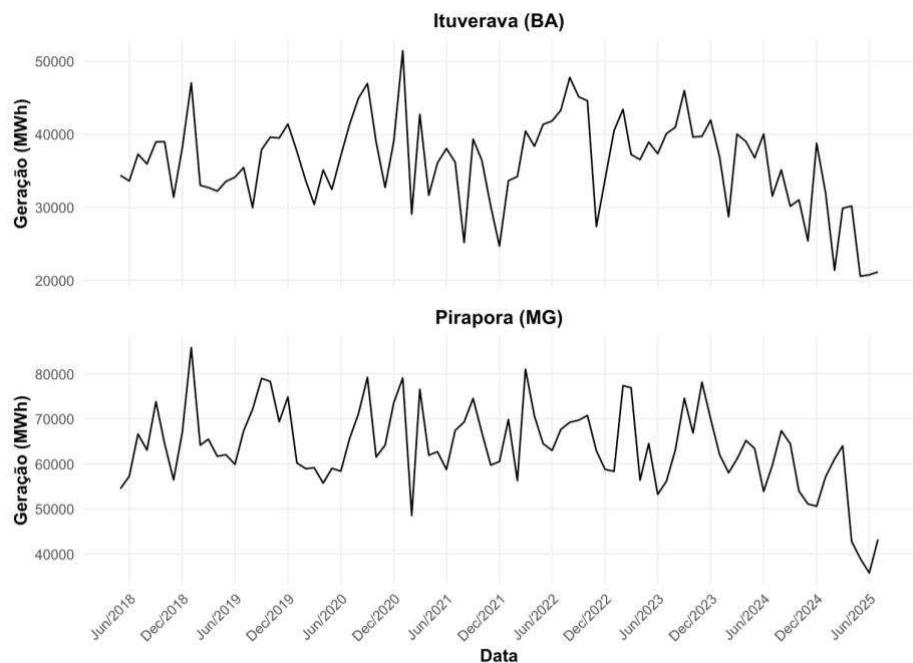


Figura 5.1: Série temporal mensal de geração de energia fotovoltaica

Fonte: Elaboração própria (2025)

Em relação às estatísticas descritivas apresentadas na Tabela 5.1, observa-se que a usina de Pirapora apresenta um nível médio de geração mensal superior ao de Ituverava (63.797 MWh contra 36.105 MWh). No entanto, Ituverava exibe maior variabilidade relativa, com coeficiente de variação (CV) de 17,3%, enquanto Pirapora apresenta CV de 14,6%, indicando maior estabilidade na produção ao longo do período analisado.

Tabela 5.1: Estatísticas descritivas de geração de energia fotovoltaica

Usina	Média	Desvio Padrão	Mínimo	Máximo	CV	Assimetria	Curtose
<i>Ituverava</i>	36.105,40	6.260,90	20.581,40	51.427,20	17,30	-0,37	3,22
<i>Pirapora</i>	63.797,30	9.286,70	35.793,50	85.841,90	14,60	-0,33	3,63

Fonte: Elaboração própria (2025)

Em ambas as séries, a assimetria ligeiramente negativa sugere uma leve concentração de valores acima da média, e a curtose próxima a três caracteriza distribuições aproximadamente mesocúrticas, sem ocorrência expressiva de valores extremos. Esses resultados indicam que as séries de geração das duas usinas apresentam comportamento estatístico estável e consistente, com variações atribuídas principalmente às condições atmosféricas locais, sem evidências de distorções significativas nos dados.

No contexto da transformação da série, verificou-se na análise exploratória que não havia sinais de heterocedasticidade, não sendo necessária a aplicação de técnicas de estabilização da variância, como Box–Cox ou transformação logarítmica. As séries de geração mantêm amplitude relativamente estável ao longo dos anos, não apresentam heterocedasticidade visível e exibem distribuição aproximadamente mesocúrtica. Dessa forma, optou-se por preservar a escala original dos dados, garantindo interpretação direta dos resultados e coerência física com a variável de interesse (MWh).

A distribuição dos valores pode ser observada nos histogramas da Figura 5.2, que confirmam a concentração de frequências em torno das médias, sem caudas alongadas. Em Ituverava (BA), observa-se uma maior ocorrência de valores de geração baixos, sugerindo episódios pontuais de limitação operacional ou redução temporária da produção. Já em Pirapora (MG), a distribuição é mais ampla, refletindo uma variação mais acentuada entre períodos de alta e baixa geração.

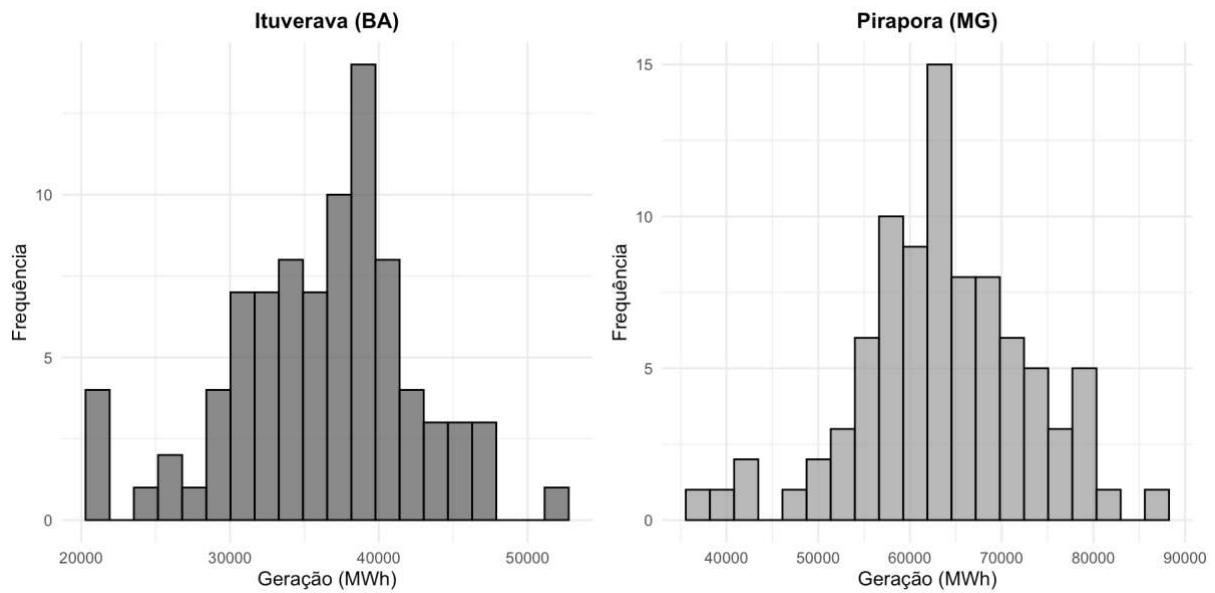


Figura 5.2: Histogramas da geração mensal das usinas

Fonte: Elaboração própria (2025)

O boxplot comparativo (Figura 5.3) complementa essa análise, mostrando que Pirapora apresenta níveis médios de geração mais elevados, porém com maior dispersão e presença de valores extremos, enquanto Ituverava mantém menor média, mas com comportamento mais concentrado e estável.

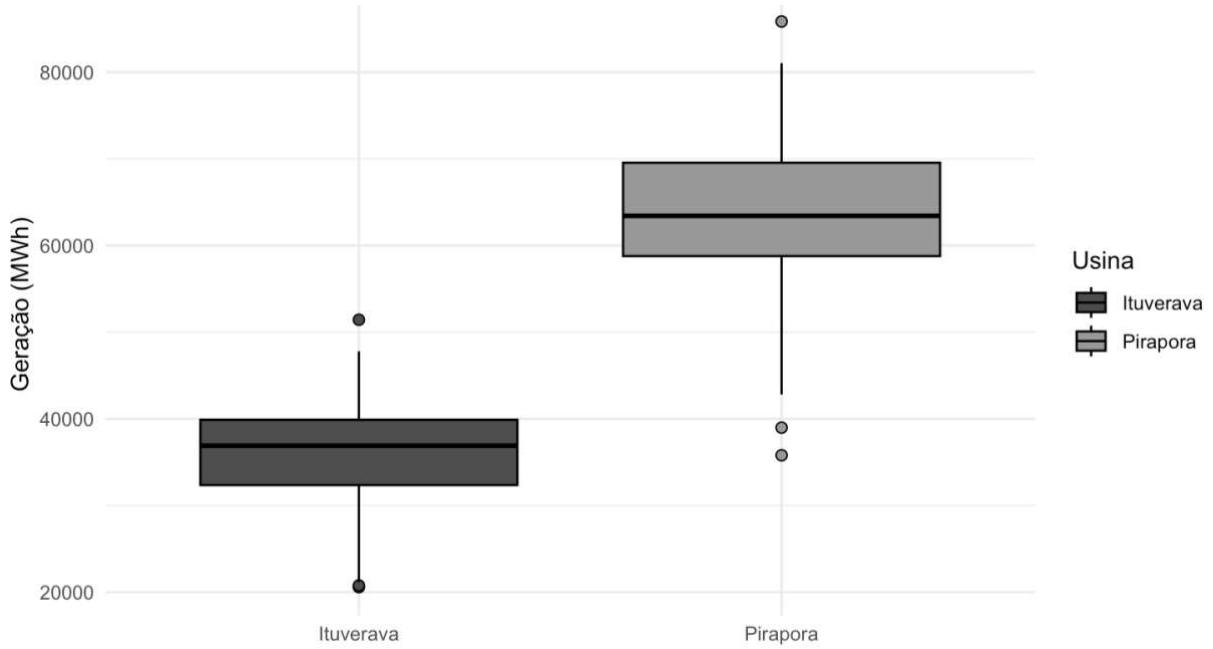


Figura 5.3: Boxplot comparativo das usinas Ituverava (BA) e Pirapora (MG)

Fonte: Elaboração própria (2025)

A decomposição STL das séries (Figura 5.4) permite identificar de forma clara os componentes de tendência, sazonalidade e ruído. Observa-se a existência de uma sazonalidade, com picos e vales que se repetem periodicamente, e uma tendência de queda mais pronunciada a partir de 2023, sobretudo em Ituverava. Essa redução pode estar relacionada a fatores operacionais que impactaram a produção efetiva. O componente residual mostra comportamento essencialmente aleatório, indicando boa separação das componentes da série.

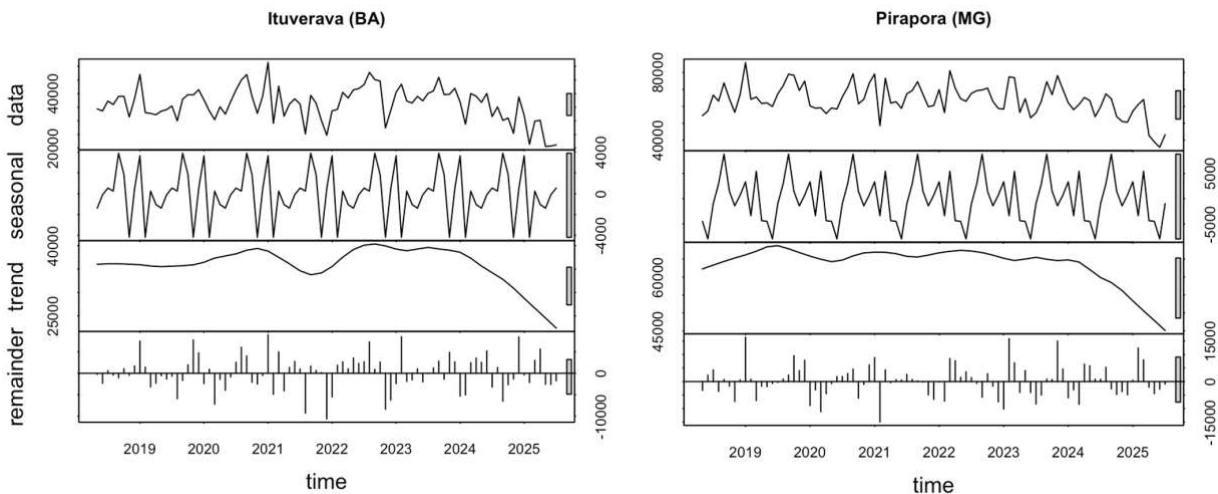


Figura 5.4: Decomposição STL das séries de geração fotovoltaica

Fonte: Elaboração própria (2025)

As Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) das séries de geração mensal (Figura 5.5) indicam dependência temporal de curta duração em ambas as usinas, compatível com processos estacionários. Em Ituverava (BA), observam-se autocorrelações positivas significativas apenas nos primeiros lags, seguidas de um rápido decaimento para valores próximos de zero, sem indicação de tendência persistente ou sazonalidade marcada. A FACP reforça esse padrão ao exibir um pico dominante no primeiro lag e valores pouco expressivos nas defasagens seguintes, sugerindo uma estrutura autorregressiva de baixa ordem. Em Pirapora (MG), a FAC também apresenta correlações iniciais moderadas e decaimento relativamente rápido, com oscilações de pequena magnitude em defasagens mais longas, possivelmente associadas a uma componente sazonal leve. A FACP mantém significância concentrada no primeiro lag, apontando novamente para uma dinâmica de memória curta.

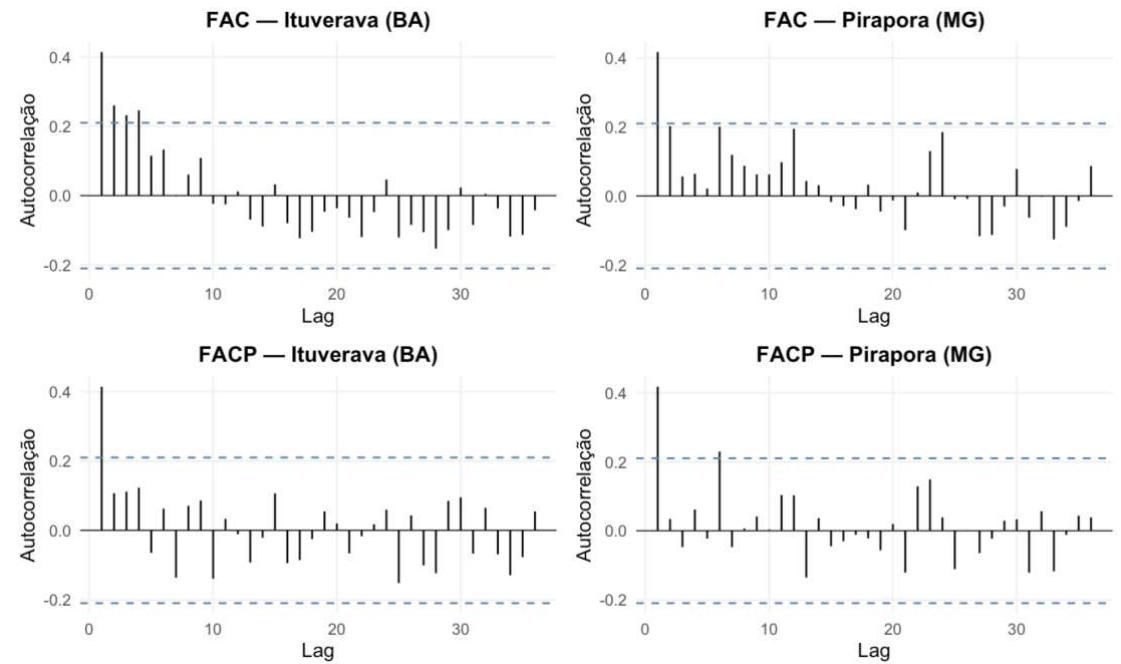


Figura 5.5: Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) das séries de geração
Fonte: Elaboração própria (2025)

5.1.2. Variáveis meteorológicas exógenas

As variáveis meteorológicas selecionadas como potenciais explicativas da geração fotovoltaica foram: albedo da superfície (ALBEDO), cobertura de nuvens (CLDTOT), radiação de onda longa absorvida pela superfície (LWGAB), espessura óptica das nuvens (TAUTOT) e temperatura da superfície (TS). Para cada usina, as séries mensais dessas variáveis foram extraídas do ponto de grade do MERRA-2 mais próximo às suas coordenadas geográficas, de modo a representar adequadamente as condições locais de radiação, temperatura e cobertura atmosférica. A proximidade entre cada usina e o respectivo ponto de grade foi obtida por meio do cálculo de distância baseado na fórmula de Haversine, conforme descrito na Seção 4.1. A Tabela 5.2 apresenta a correspondência entre as coordenadas reais das usinas, os pontos de grade utilizados e as distâncias calculadas.

Tabela 5.2: Correspondência entre coordenadas das usinas e os pontos de grade do MERRA-2

<i>Usina</i>	<i>Coordenadas Reais</i>	<i>Ponto de grade do MERRA 2</i>	<i>Distância (km)</i>
<i>Ituverava (BA)</i>	Lat. $-12,5996^{\circ}$, Lon. $-44,1114^{\circ}$	Lat. $-12,5^{\circ}$, Lon. $-44,375^{\circ}$	30,7
<i>Pirapora (MG)</i>	Lat. $-17,4066^{\circ}$, Lon. $-44,9219^{\circ}$	Lat. $-17,5^{\circ}$, Lon. $-45,0^{\circ}$	13,3

Fonte: Elaboração própria (2025)

Tendo estabelecida essa correspondência espacial, a Figura 5.6 apresenta as séries mensais das variáveis meteorológicas para Ituverava (BA) e Pirapora (MG).

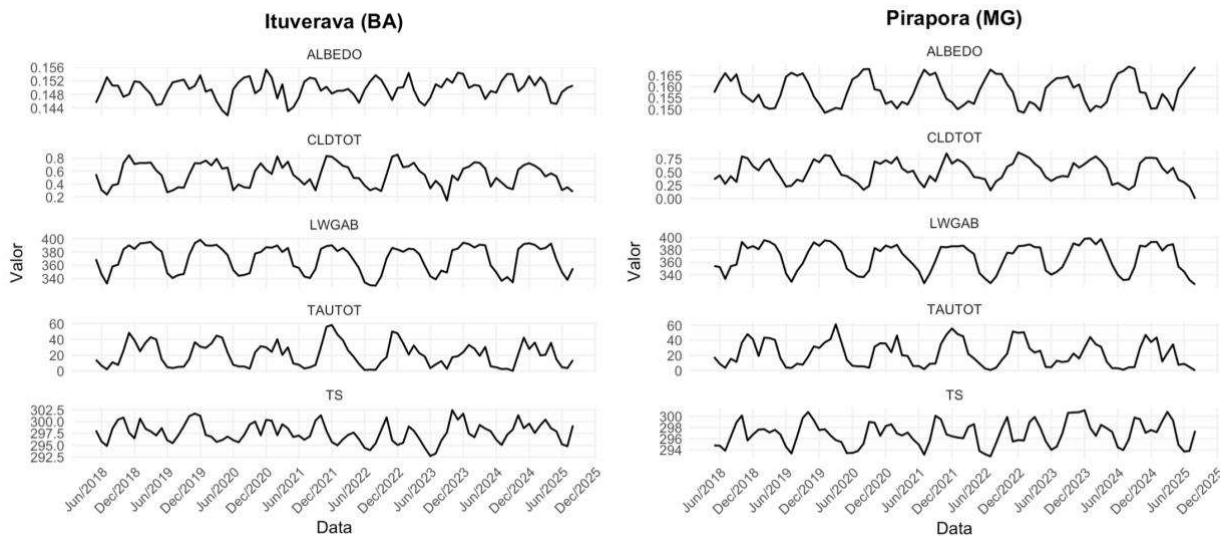


Figura 5.6: Sériess temporais das variáveis meteorológicas exógenas

Fonte: Elaboração própria (2025)

Observa-se que as variáveis exibem padrões sazonais anuais bem definidos, compatíveis com a alternância entre os períodos seco e chuvoso característicos do clima tropical. O albedo destaca-se pela diferença de comportamento entre as localidades: em Pirapora, a série apresenta sazonalidade mais regular e bem-comportada, refletindo a maior estabilidade climática e a presença de estações do ano mais definidas. Em Ituverava, o padrão é relativamente mais irregular, possivelmente associado à maior variabilidade atmosférica regional, como flutuações na umidade e na cobertura de nuvens, que influenciam a refletância média estimada pelo modelo de reanálise. Além disso, observa-se que em Pirapora há relação inversa entre ALBEDO e CLDTOT, coerente com o efeito físico da cobertura de nuvens na redução da radiação refletida pela superfície.

No que se refere à necessidade de transformação das séries exógenas, também não se observou qualquer indício de variância não constante ou assimetria acentuada que justificasse procedimentos como Box-Cox ou logaritmo natural. As variáveis meteorológicas apresentaram dinâmica estável, com amplitude aproximadamente constante e distribuição visualmente simétrica, além de sazonalidade anual bem definida, sem evidências de tendência ou heterocedasticidade.

A Figura 5.7 apresenta as Funções de Autocorrelação (FAC) das variáveis exógenas para Ituverava e Pirapora. Em ambas as localidades, observam-se picos em defasagens

múltiplas de 12 meses, confirmando a presença de um componente sazonal anual consistente com os padrões observados nas séries temporais.

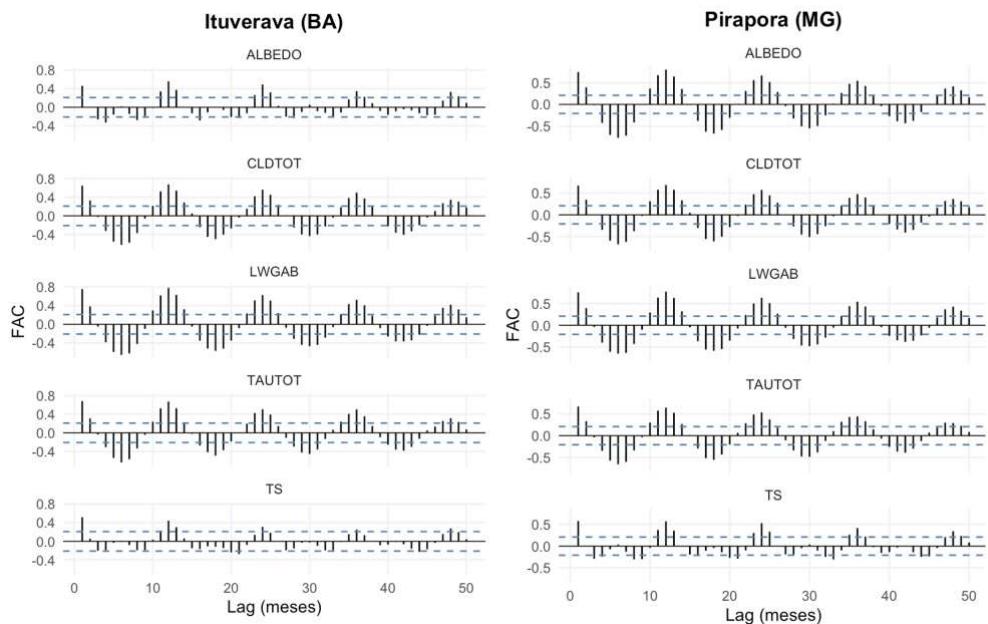


Figura 5.7: Funções de Autocorrelação (FAC) das variáveis meteorológicas exógenas

Fonte: Elaboração própria (2025)

Fora os picos sazonais, as autocorrelações decaem rapidamente e permanecem próximas de zero, indicando baixa persistência temporal e ausência de tendência de longo prazo. Esses resultados sugerem que as séries apresentam comportamento estatisticamente estável, com sazonalidade bem definida e estacionariedade em média e variância. Assim, mostram-se adequadas não apenas para uso como variáveis explicativas em modelos SARIMAX, mas também para previsões independentes com modelos da família SARIMA, permitindo projeções futuras das variáveis meteorológicas e, consequentemente, da geração fotovoltaica.

5.1.3. Relação entre a geração de energia e as variáveis exógenas

A fim de investigar a influência das variáveis meteorológicas sobre a geração mensal de energia fotovoltaica, realizou-se uma análise de correlação linear de Pearson, que mede a intensidade e a direção da associação linear entre duas variáveis contínuas (Benesty et al., 2009). Como já estabelecido na seção anterior, utilizaram-se as séries meteorológicas correspondentes a localização de cada usina. Considerando que os efeitos meteorológicos influenciam a geração de forma quase imediata, optou-se por analisar correlações contemporâneas (lag 0), sem defasagem temporal entre as séries de geração e as variáveis exógenas. (Jebli et al., 2021).

A significância estatística das correlações foi avaliada individualmente pelo teste t de Student, que testa a hipótese nula de ausência de correlação populacional ($H_0: \rho = 0$) (Pearson, 1932). Adotou-se nível de significância de 10% para identificar as relações estatisticamente relevantes.

Em Ituverava, o albedo, a espessura óptica das nuvens (TAUTOT), a cobertura de nuvens (CLDTOT) e a temperatura da superfície (TS) apresentaram correlações estatisticamente significativas com a geração (Tabela 5.3).

Tabela 5.3: Correlação entre variáveis exógenas e geração na Usina de Ituverava (BA)

Variável	Correlação	p-valor	Significativa
ALBEDO	0,322	0,002	*
TAUTOT	-0,267	0,013	*
CLDTOT	-0,22	0,041	*
TS	0,207	0,055	*
LWGAB	-0,135	0,212	

Fonte: Elaboração própria (2025)

O albedo e a temperatura da superfície (TS) apresentaram correlações positivas com a geração, sugerindo que condições típicas de maior irradiância — como céu mais limpo, maior refletância da superfície e temperaturas elevadas — tendem a coincidir com maior produção fotovoltaica. Em contraste, a espessura óptica das nuvens (TAUTOT) e a cobertura de nuvens (CLDTOT) exibiram correlações negativas, coerentes com o efeito de atenuação da radiação solar causado pela presença de nuvens. A variável radiação de onda longa absorvida pela superfície (LWGAB), embora também negativa, apresentou correlação fraca.

Em Pirapora, a temperatura da superfície (TS) apresentou a correlação mais elevada e estatisticamente significativa com a geração (Tabela 5.4), reforçando seu papel como indicador indireto de irradiância solar. As demais variáveis, individualmente, apresentaram correlações fracas e não significativas, mas com sinais coerentes com o comportamento físico esperado.

Tabela 5.4: Correlação entre variáveis exógenas e geração na Usina de Pirapora (MG)

Variável	Correlação	p-valor	Significativa
TS	0,424	< 0,001	*
ALBEDO	0,163	0,132	
TAUTOT	-0,154	0,155	
LWGAB	0,059	0,59	
CLDTOT	-0,041	0,708	

Fonte: Elaboração própria (2025)

As matrizes de correlação (Figura 5.8) evidenciam relações consistentes entre as variáveis atmosféricas analisadas. Observa-se uma forte interdependência entre CLDTOT, LWGAB e TAUTOT ($r > 0,85$), indicando que essas variáveis representam aspectos complementares de um mesmo fenômeno físico — a interação entre nuvens, vapor d'água e radiação incidente.

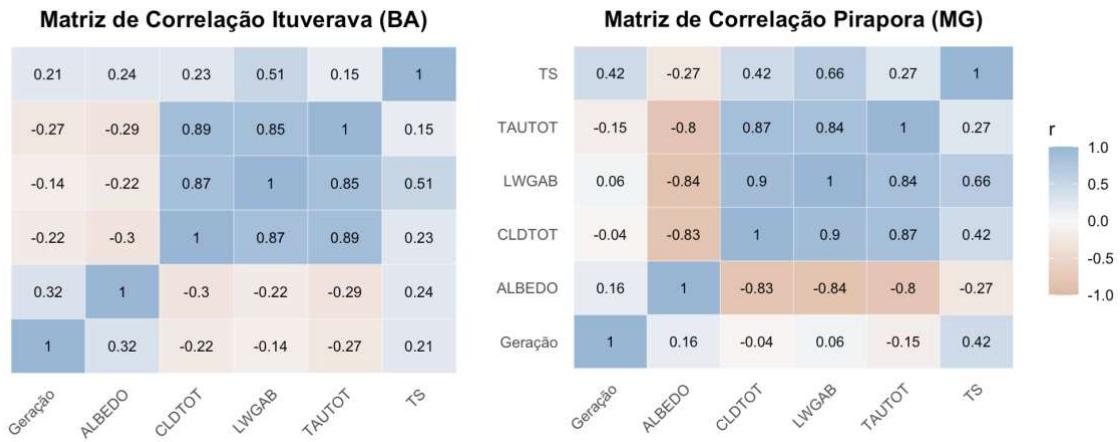


Figura 5.8: Matriz de Correlação

Fonte: Elaboração própria (2025)

As correlações elevadas observadas entre algumas dessas variáveis meteorológicas sugerem a presença de multicolinearidade, isto é, dependência linear entre os preditores. Esse fenômeno ocorre quando duas ou mais variáveis explicativas compartilham informações redundantes, o que pode gerar coeficientes com sinais/magnitudes pouco estáveis, aumentar erros-padrão e até mascarar significâncias individuais mesmo quando o R^2 do modelo é alto (Thompson et al., 2017).

Para mensurar o grau dessa redundância, aplicou-se o Fator de Inflação da Variância (Variance Inflation Factor – VIF), calculado como:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (5.1)$$

em que R_j^2 representa o coeficiente de determinação da regressão da variável j sobre as demais. O VIF quantifica o quanto a variância do coeficiente de j é inflada pela correlação com outros preditores. Quanto a pontos de corte, a literatura alerta que não há consenso rígido: valores maiores que 10 são por vezes usados como indicativo de multicolinearidade “severa”,

enquanto faixas intermediárias (5–10) sugerem atenção, mas a interpretação deve considerar tamanho amostral e R^2 do modelo, evitando decisões dicotômicas (Thompson et al., 2017).

Tabela 5.5: Fator de Inflação da Variância (VIF) entre exógenas

Variável	VIF Ituverava (BA)	Classificação	VIF Pirapora (MG)	Classificação
LWGAB	11,15	Alta	39,62	Alta
TAUTOT	7,39	Moderada	7,13	Moderada
CLDTOT	6,88	Moderada	8,17	Moderada
TS	2,78	Baixa	7,85	Moderada
ALBEDO	1,30	Baixa	8,17	Moderada

Fonte: Elaboração própria (2025)

Os resultados apresentados na Tabela 5.5 indicaram colinearidade moderada a alta nas variáveis CLDTOT, LWGAB e TAUTOT em ambas as usinas, mas sem colinearidade perfeita ($r < 1$), o que não inviabiliza a estimativa do modelo. Essa redundância parcial é esperada, uma vez que essas variáveis representam diferentes dimensões do balanço radiativo atmosférico.

Dessa forma, optou-se por testar todas as combinações possíveis de variáveis exógenas na modelagem SARIMAX, buscando capturar efeitos conjuntos e potenciais interações não lineares entre os fatores atmosféricos. Embora os valores de VIF indiquem certo grau de multicolinearidade, essa condição não inviabiliza o modelo, pois o foco deste estudo é previsão, e não inferência causal. De acordo com Thompson et al. (2017), a multicolinearidade pode aumentar a variância dos coeficientes, mas não compromete o desempenho preditivo se o modelo estiver corretamente especificado.

5.2. Estacionariedade e diferenciação

A análise exploratória das FAC e FACP na seção 5.1 sugeriu que tanto as séries de geração quanto as séries das variáveis exógenas apresentam comportamento estável ao longo do tempo, sem tendência determinística pronunciada e com sazonalidade bem delimitada nas variáveis exógenas, indicando um comportamento estacionário. Para verificar formalmente essa evidência, aplicou-se o teste Dickey–Fuller aumentado com constante, conforme descrito na seção 4.2.

Neste estudo, adotou-se nível de significância de 5%, que corresponde a um valor crítico de -2,89. Valores da estatística de teste mais negativos que os críticos levam à rejeição da hipótese nula de raiz unitária e, portanto, à conclusão de que a série é estacionária. A Tabela 5.6 apresenta os resultados para as séries analisadas.

Tabela 5.6: Resultados do Teste ADF para as Séries de Geração e Variáveis Exógenas

<i>Série</i>	<i>Estatística de Teste</i>	<i>Resultado</i>
<i>Ituverava (BA)</i>	-3,71	Estacionária
<i>Pirapora (MG)</i>	-4,86	Estacionária
<i>ALBEDO</i>	-6,24	Estacionária
<i>CLDTOT</i>	-4,49	Estacionária
<i>LWGAB</i>	-4,97	Estacionária
<i>TAUTOT</i>	-4,78	Estacionária
<i>TS</i>	-5,89	Estacionária

Fonte: Elaboração própria (2025)

Os resultados confirmam que nenhuma das séries apresenta raiz unitária, ou seja, todas são estacionárias em nível ao nível de significância adotado. Esse diagnóstico é consistente com o comportamento observado visualmente nas análises anteriores, em que não se verificaram tendências persistentes ou padrões de não estacionariedade. Dessa forma, a modelagem pode ser conduzida diretamente em nível, sem a necessidade de diferenciação regular adicional.

5.3.Identificação, Estimação e Diagnóstico dos modelos

Uma vez confirmada a estacionariedade das séries, procedeu-se à etapa de identificação e estimação dos modelos. Apesar de apresentarem comportamento estatisticamente estável ao longo da maior parte do período analisado, as séries de geração exibem, nos meses finais, reduções atípicas de geração que se desviam do padrão histórico dominante.

Diante desse comportamento, retoma-se a estratégia de validação fora da amostra definida na Seção 4.2.4.1, aplicando-se as duas janelas temporais de teste de 12 meses apresentadas na Tabela 4.1, permitindo avaliar o desempenho dos modelos em um período alinhado ao padrão histórico e no trecho recente de comportamento distinto.

No caso dos modelos SARIMAX, foi necessário, adicionalmente, especificar as variáveis climáticas correspondentes ao período de teste. Para cada usina, utilizaram-se os valores contemporâneos (lag 0) das variáveis exógenas obtidos no ponto de grade MERRA-2 mais próximo às coordenadas geográficas da usina, conforme Tabela 5.2.

Com as variáveis exógenas alinhadas, procedeu-se à identificação dos modelos SARIMA. Conforme descrito na seção 4.2, a identificação dos modelos SARIMA foi conduzida

por meio de uma varredura sistemática em uma grade controlada de combinações possíveis de $(p, d, q)(P, D, Q)_s$. A definição dessa grade baseou-se exclusivamente em evidências empíricas da série:

- Diferenciação regular (d) — o teste ADF indicou que a série de geração já era estacionária em nível, justificando a escolha de $d = 0$.
- Diferenciação sazonal (D) — a inspeção da FAC mostrou que os lags sazonais (12, 24, 36) permaneciam dentro das bandas de significância, sem picos representativos de raiz unitária sazonal. Assim, definiu-se $D = 0$, permitindo comparabilidade plena dos AICc entre os modelos.
- Ordens máximas (p, q, P, Q) — a análise conjunta das FAC e FACP indicou dependência temporal significativa apenas nos primeiros. Com base nisso, estabeleceram-se limites coerentes e parcimoniosos: $p_{\max} = 3$, $q_{\max} = 3$, $P_{\max} = 2$, $Q_{\max} = 2$.

Para cada modelo da grade foram calculados e armazenados o AICc e os resultados dos testes de diagnóstico (Ljung–Box e Jarque–Bera). A seleção final seguiu exatamente os critérios metodológicos da seção 4.2, considerando apenas os modelos estatisticamente válidos. Assim, o melhor modelo em cada janela foi definido entre aqueles que:

1. Apresentaram o menor valor de AICc dentre todas as especificações estimadas; e
2. Foram aprovados simultaneamente nos testes de diagnóstico, adotando-se:
 - p-valor $\geq 0,10$ no teste de Ljung–Box, assegurando maior rigor na detecção de autocorrelação residual;
 - p-valor $\geq 0,05$ no teste de Jarque–Bera, garantindo aderência adequada à suposição de normalidade dos resíduos.

O desempenho preditivo dos modelos selecionados nas janelas de teste foi então avaliado conforme o procedimento de validação fora da amostra descrito na Seção 4.2.1.4, utilizando as métricas de MAPE e RMSE definidas nas Equações (4.10) e (4.11).

O mesmo procedimento descrito ao longo desta seção foi integralmente aplicado aos modelos SARIMAX. Nesse caso, todas as combinações das variáveis exógenas contemporâneas foram cruzadas com a grade de ordens SARIMA definida anteriormente, resultando em uma varredura completa das especificações possíveis.

As Tabelas 5.7 e 5.8 apresentam, para cada usina e para cada janela, os melhores modelos SARIMA e SARIMAX identificados, juntamente com suas principais métricas: o

AICc obtido no período de treinamento, os p-valores dos testes de diagnóstico dos resíduos (Ljung–Box e Jarque–Bera) e os erros de previsão no período de teste.

Tabela 5.7: Modelos selecionados para previsão da geração fotovoltaica em Ituverava

	<i>Melhor Modelo</i>	<i>Exógenas</i>	<i>AICc</i>	<i>RMSE</i>	<i>MAPE</i>	<i>p-valor LB</i>	<i>p-valor JB</i>
Janela 1	SARIMA	SARIMA(2,0,1)(0,0,2)[12]	-	1273	6037	13,31%	0,25
	SARIMAX	SARIMA(1,0,0)(0,0,2)[12]	TAUTOT_10 + ALBEDO_10 + TS_10 + LWGAB_10	1225	7296	17,45%	0,12
Janela 2	SARIMA	SARIMA(2,0,1)(0,0,2)[12]	-	1512	11373	41,23%	0,37
	SARIMAX	SARIMA(1,0,1)(0,0,2)[12]	CLDTOT_10 + TAUTOT_10 + ALBEDO_10 + TS_10 + LWGAB_10	1472	10099	35,85%	0,10

Fonte: Elaboração própria (2025)

A Figura 5.9 apresenta, para as duas janelas de teste, a comparação visual entre a geração real e as previsões dos modelos SARIMA e SARIMAX em Ituverava, evidenciando as diferenças no nível previsto e na capacidade de cada modelo acompanhar as variações mensais da série.

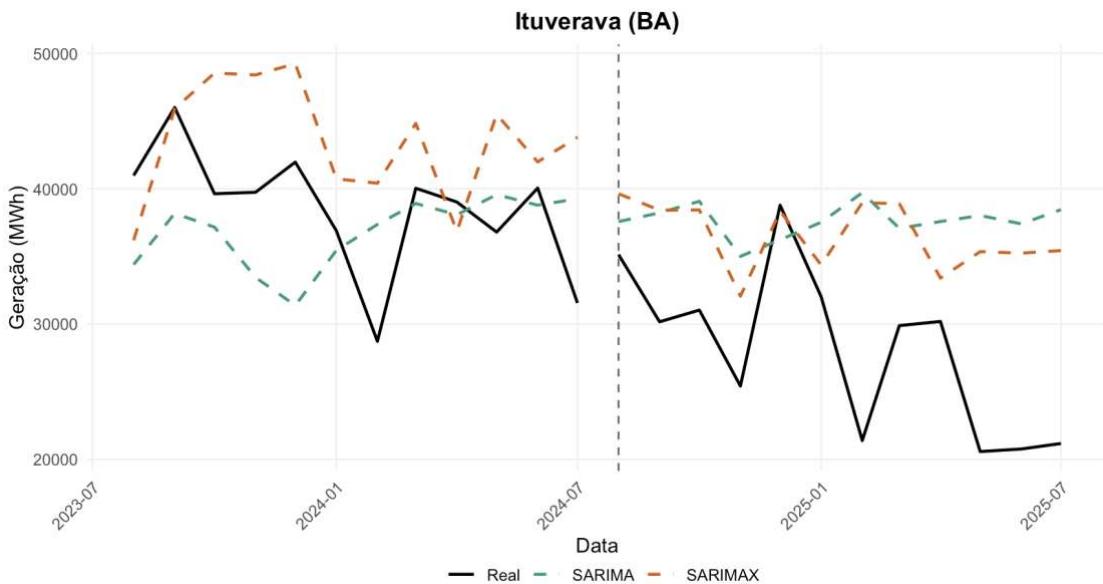


Figura 5.9: Previsão da geração fotovoltaica nas janelas de teste em Ituverava

Fonte: Elaboração própria (2025)

Os resultados de Ituverava mostram comportamentos distintos entre as duas janelas analisadas. Na Janela 1, embora o SARIMAX tenha apresentado um AICc aproximadamente 4% menor que o SARIMA — indicando melhor aderência ao conjunto de treinamento — essa vantagem não se manteve no período de teste. O erro percentual médio do SARIMAX foi cerca de 31% maior que o do SARIMA, além de apresentar maior magnitude média dos erros quadráticos. Nesse intervalo, em que a série ainda segue um padrão relativamente típico de variação, o modelo univariado mostrou-se mais eficaz, sugerindo que a inclusão das exógenas não trouxe ganhos preditivos. O modelo SARIMAX selecionado incorporou TAUTOT_10, ALBEDO_10, TS_10 e LWGAB_10. Entre essas variáveis, ALBEDO, TAUTOT e TS apresentaram correlações significativas com a geração na análise feita na seção 5.1.3, mas CLDTOT — também significativa — não foi incluída, ao passo que LWGAB foi selecionada apesar de sua correlação mais fraca.

Na Janela 2, o cenário se inverte. Assim como na primeira janela, o SARIMAX apresentou melhor ajuste em treinamento, reduzindo o AICc em aproximadamente 3%. No entanto, dessa vez, o ganho também ocorreu no teste: o erro percentual médio caiu cerca de 13% em relação ao SARIMA, e houve igualmente redução na magnitude média dos erros quadráticos. Essa janela coincide com o trecho de queda mais acentuada e comportamento atípico da série, em que a geração se afasta de forma intensa do padrão histórico. Nesses meses mais desafiadores, o SARIMAX acompanhou melhor a trajetória observada. O modelo selecionado incorporou todas as variáveis exógenas contemporâneas — ALBEDO_10, CLDTOT_10, TAUTOT_10, TS_10 e LWGAB_10. Essa especificação reflete de maneira direta o padrão de correlações estimado anteriormente, já que ALBEDO, TAUTOT, CLDTOT e TS estavam associadas de forma significativa ao comportamento da geração, enquanto LWGAB, embora menos correlacionada isoladamente, está fisicamente ligada ao regime de nuvens. A inclusão conjunta das variáveis torna o modelo mais responsável às mudanças atmosféricas que caracterizam esse período.

De forma geral, em Ituverava, os resultados indicam que o ganho proporcionado pelas variáveis exógenas depende do regime da série. Em períodos mais regulares, como na Janela 1, o SARIMA é suficiente e tende a generalizar melhor. Já em momentos de mais difícil previsão, como na Janela 2, o SARIMAX se mostra mais robusto ao incorporar informações meteorológicas adicionais, reduzindo os erros de previsão e capturando com maior sensibilidade as variações nas condições de operação da usina.

Tabela 5.8: Modelos selecionados para previsão da geração fotovoltaica em Pirapora

	<i>Melhor Modelo</i>	<i>Exógenas</i>	<i>AICc</i>	<i>RMSE</i>	<i>MAPE</i>	<i>p-valor LB</i>	<i>p-valor JB</i>	
Janela 1	SARIMA	SARIMA(1,0,2)(2,0,0)[12]	-	1319	6796	9,15%	0,82	0,68
	SARIMAX	SARIMA(2,0,3)(0,0,1)[12]	CLDTOT_10 + TAUTOT_10 + LWGAB_10	1265	4573	5,52%	0,26	0,75
Janela 2	SARIMA	SARIMA(1,0,1)(1,0,0)[12]	-	1566	16096	29,71%	0,62	0,59
	SARIMAX	SARIMA(1,0,1)(1,0,1)[12]	TAUTOT_10 + ALBEDO_10	1504	10370	18,92%	0,24	0,52

Fonte: Elaboração própria (2025)

A Figura 5.10 apresenta a comparação visual entre a geração real e as previsões dos modelos SARIMA e SARIMAX em Pirapora, permitindo observar como cada especificação se ajusta ao nível da série e à sua variabilidade mensal.

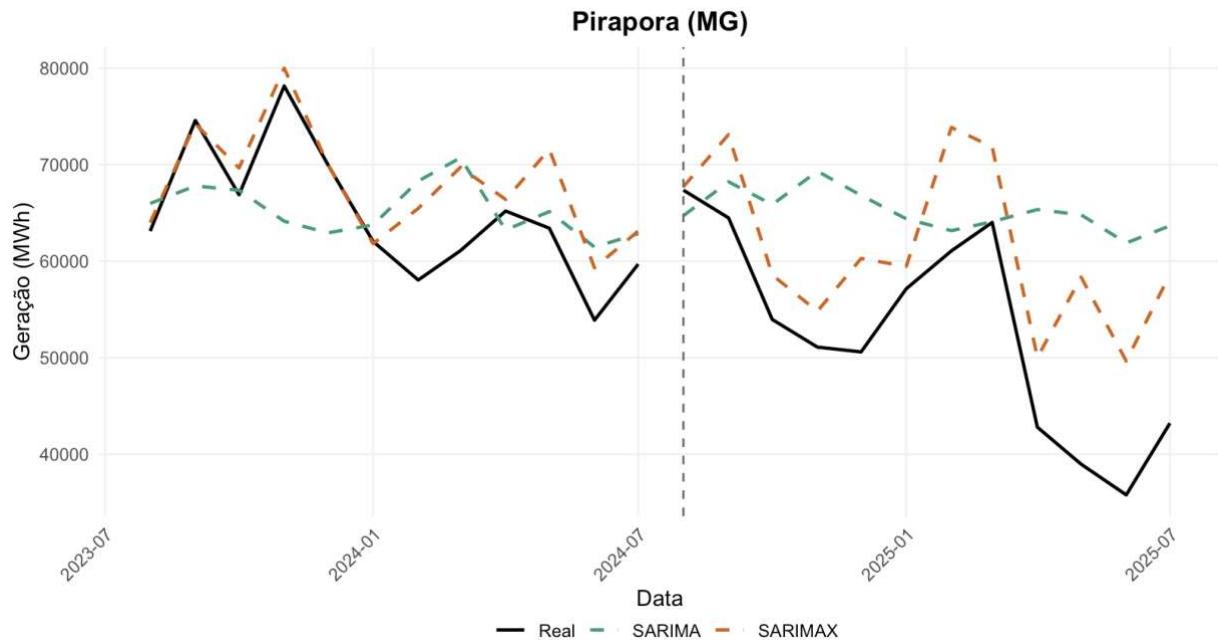


Figura 5.10: Previsão da geração fotovoltaica nas janelas de teste em Pirapora

Fonte: Elaboração própria (2025)

Os resultados de Pirapora mostram que a inclusão de variáveis exógenas trouxe ganhos substanciais. Na Janela 1, o SARIMAX reduziu o AICc em aproximadamente 4% em relação ao SARIMA, indicando melhor ajuste no treinamento. No período de teste, a melhora foi ainda

mais expressiva: o MAPE caiu cerca de 39,7% e o RMSE reduziu-se em aproximadamente 32,7%. A composição do modelo selecionado — CLDTOT, TAUTOT e LWGAB — reflete parcialmente as relações identificadas na análise de correlação da seção 5.1.3. Embora nenhuma dessas variáveis tenha se mostrado individualmente forte, dado que apenas TS apresentou correlação significativa com a geração, esse conjunto representa diferentes dimensões da cobertura de nuvens e do balanço radiativo atmosférico. Como essas variáveis são altamente correlacionadas entre si e estão diretamente relacionadas com a irradiância efetiva, sua inclusão é coerente com o comportamento físico da usina e ajuda a explicar por que o SARIMAX apresentou desempenho superior nesse intervalo.

Na Janela 2, marcada por quedas acentuadas e comportamento atípico da série, os ganhos tornaram-se ainda mais relevantes. O SARIMAX voltou a reduzir o AICc em cerca de 4% e melhorou a previsão no teste, com redução aproximada de 36% no MAPE e de 35% no RMSE. Nesse caso, o modelo selecionado incorporou TAUTOT e ALBEDO, duas variáveis cuja interpretação física dialoga diretamente com condições atmosféricas instáveis: TAUTOT está associada à espessura óptica das nuvens, que atenua a irradiância, enquanto ALBEDO reflete mudanças no comportamento radiativo da superfície. Ainda que ALBEDO não tenha exibido correlação significativa isoladamente na análise preliminar, sua inclusão ao lado de TAUTOT é consistente com o alto grau de associação linear entre variáveis atmosféricas observado na matriz de correlação. Assim, em Pirapora, a escolha das variáveis exógenas nos modelos vencedores mantém coerência física com o fenômeno, contribuindo de forma clara para a melhora preditiva em ambos os regimes analisados.

De forma geral, em Pirapora a inclusão das variáveis exógenas trouxe ganho consistente em todas as janelas. Tanto no ajuste quanto no teste, o SARIMAX superou o SARIMA, indicando que a geração da usina responde de forma mais sensível às condições atmosféricas. Assim, mesmo variáveis com correlação individual moderada contribuíram para melhorar a previsão, tornando o SARIMAX mais eficaz em diferentes regimes de comportamento da série.

Em comparação direta entre as duas usinas, o ganho proporcionado pelo SARIMAX com relação ao SARIMA na Janela 2 foi significativamente maior em Pirapora — redução de cerca de 36% no MAPE — enquanto em Ituverava o ganho foi mais moderado, aproximando-se de 13%. Esse contraste torna-se particularmente relevante porque, em ambas as usinas, os modelos vencedores da Janela 2 incluíram a variável ALBEDO entre os preditores. Nesse contexto, a sazonalidade mais regular e bem definida do ALBEDO em Pirapora, conforme

discutido na Seção 5.1, pode ter contribuído para que o modelo com exógenas capturasse de forma mais eficiente as variações do período recente, marcado por quedas acentuadas e maior instabilidade. Já em Ituverava, onde o ALBEDO apresenta comportamento mais irregular, esse mesmo efeito tende a ser menos pronunciado, o que é compatível com o ganho preditivo mais modesto observado.

Por fim, tanto em Ituverava quanto em Pirapora observa-se um comportamento comum entre os modelos: os erros no período de teste aumentam de forma expressiva da Janela 1 para a Janela 2, independentemente da inclusão ou não de variáveis exógenas. Em Ituverava, o MAPE salta de aproximadamente 13 – 17% na Janela 1 para 36 – 41% na Janela 2. Já em Pirapora, cresce de cerca de 5 – 9% para 19 – 30% no mesmo intervalo. Esse padrão consistente nas duas usinas — mesmo estando localizadas em estados diferentes e expostas a condições meteorológicas distintas — sugere a ocorrência de um fenômeno externo à dinâmica usual das séries, possivelmente associado a *curtailment* ou limitações operacionais concentradas no período mais recente. A convergência desse aumento de erro entre usinas distintas reforça a interpretação de que houve um evento excepcional que nenhum modelo, univariado ou com exógenas, seria capaz de antecipar apenas com base no comportamento histórico.

5.4. Previsão e aplicação

Concluída a etapa de estimação dos modelos e a avaliação de seu desempenho nos períodos de treinamento e teste, procede-se à geração das previsões futuras. Como os modelos SARIMAX dependem de valores contemporâneos (lag 0) das variáveis exógenas, tornou-se necessário projetar também essas séries meteorológicas. Assim, o primeiro passo consistiu em ajustar modelos de previsão individuais para cada variável climática, possibilitando a construção dos cenários exógenos requeridos para as projeções de geração fotovoltaica. Assim, o primeiro passo consistiu em construir modelos de previsão individuais para cada variável climática.

5.4.1. Modelagem e projeção das variáveis exógenas

As séries exógenas apresentam comportamento sazonal e estatisticamente estável ao longo do tempo, com oscilações recorrentes em defasagens múltiplas de 12 meses, conforme discutido na Seção 5.1.2. Diante desse padrão, adotou-se a modelagem por meio da família SARIMA, adequada para capturar estruturas sazonais determinísticas e estocásticas. Antes da modelagem, a estacionariedade das séries foi avaliada por meio do teste de Dickey–Fuller

Aumentado (ADF), apresentado na Seção 5.2. Os resultados confirmaram a ausência de raiz unitária, permitindo o uso de modelos SARIMA sem transformações adicionais.

Em um primeiro momento, cada variável foi ajustada individualmente com o procedimento *auto.arima()*, do pacote *forecast* (Hyndman & Khandakar, 2008), que combina busca automatizada com otimização via MLE e seleção baseada no AICc, assegurando equilíbrio entre qualidade de ajuste e simplicidade estrutural. Essa estimação foi realizada utilizando janelas de teste para avaliar preliminarmente a capacidade dos modelos de reproduzir a dinâmica temporal das séries. Essa etapa teve caráter validatório e demonstrou que os modelos produziriam previsões aderentes, com resíduos sem autocorrelação significativa.

A partir dessa confirmação, os modelos foram reestimados utilizando toda a série disponível, uma vez que, na etapa de previsão futura, busca-se maximizar o uso de informação histórica para gerar projeções mais estáveis e estatisticamente robustas. Os modelos finais selecionados (Tabela 5.9) apresentam estruturas sazonais parcimoniosas e passam no teste de Ljung–Box, indicando ausência de autocorrelação residual. Embora alguns modelos não satisfaçam o teste de Jarque–Bera, essa violação não compromete o uso dos ajustes, pois a normalidade dos resíduos não é um requisito para previsões pontuais, que constituem o foco desta etapa.

Tabela 5.9: Modelos selecionados para projeção das variáveis exógenas

	<i>Variável</i>	<i>Modelo</i>	<i>AICc</i>	<i>p-valor LB</i>	<i>p-valor JB</i>
Ituverava	<i>ALBEDO</i>	SARIMA(1,0,0)(0,1,2)[12]	-712	0,30	0,01
	<i>CLDTOT</i>	SARIMA(0,0,0)(2,1,0)[12]	-140	0,26	0,00
	<i>TS</i>	SARIMA(0,0,1)(0,1,1)[12]	285	0,43	0,64
	<i>LWGAB</i>	SARIMA(1,0,1)(0,1,1)[12]	507	0,72	0,78
	<i>TAUTOT</i>	SARIMA(0,0,2)(2,1,0)[12]	524	0,86	0,04
Pirapora	<i>ALBEDO</i>	SARIMA(0,0,1)(0,1,2)[12]	-709	0,24	0,27
	<i>CLDTOT</i>	SARIMA(0,0,0)(0,1,2)[12]	-133	0,45	0,76
	<i>TS</i>	SARIMA(1,0,0)(0,1,2)[12]	248	0,65	0,01
	<i>LWGAB</i>	SARIMA(1,0,1)(0,1,1)[12]	523	0,82	0,85
	<i>TAUTOT</i>	SARIMA(1,0,0)(0,1,1)[12]	564	0,68	0,00

Fonte: Elaboração própria (2025)

As projeções das variáveis exógenas mantêm coerência com o comportamento histórico, preservando a sazonalidade anual característica e níveis médios compatíveis com a trajetória recente das séries. A Figura 5.11 mostra essa continuidade entre os valores observados e previstos, evidenciando uma transição suave, sem rupturas abruptas ou mudanças artificiais de padrão. Esse comportamento indica que os modelos SARIMA capturam adequadamente a estrutura temporal das séries, reforçando a confiabilidade das projeções geradas e sua adequação como insumo para os modelos SARIMAX utilizados na etapa subsequente.

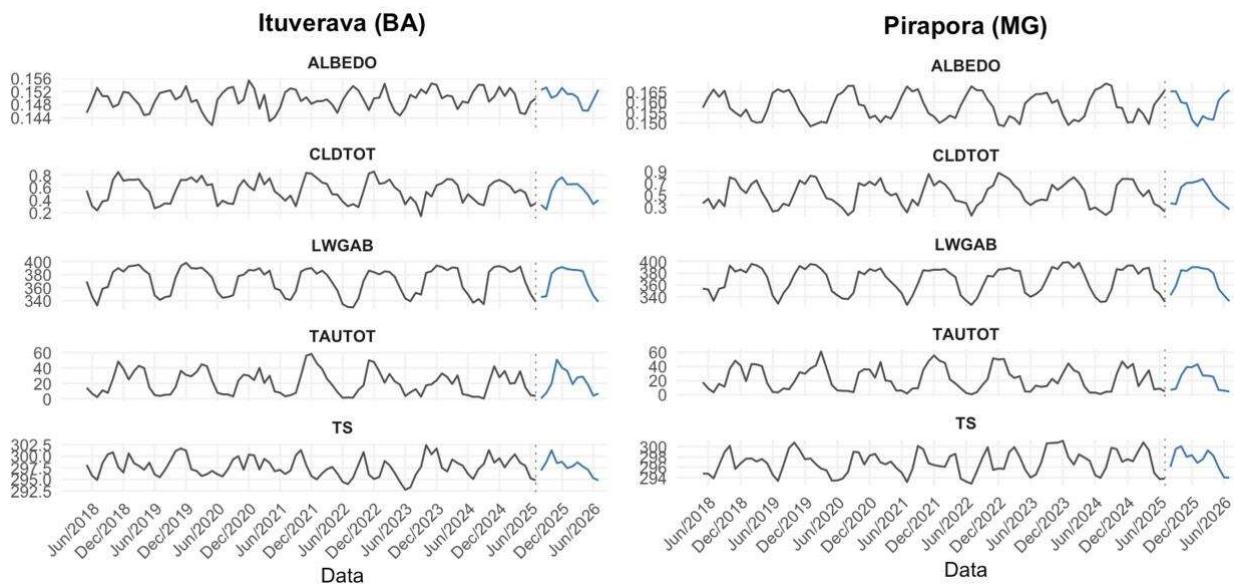


Figura 5.11: Projeções futuras das variáveis exógenas (12 passos à frente)

Fonte: Elaboração própria (2025)

5.4.2. Previsão da geração de energia fotovoltaica

Com as exógenas projetadas, procedeu-se à modelagem final da geração fotovoltaica para cada usina, estimando-se tanto modelos SARIMA quanto modelos SARIMAX. Essa etapa seguiu o procedimento metodológico previamente validado nas janelas de teste, mas agora utilizando toda a série histórica disponível, uma vez que o objetivo é gerar previsões futuras e, portanto, maximizar o uso de informação na estimação dos parâmetros.

A seleção dos modelos considerou dois critérios complementares. Primeiramente, adotou-se o menor valor de AICc entre as alternativas estimadas, assegurando parcimônia e adequação estatística. Em seguida, avaliou-se o diagnóstico residual, exigindo ausência de autocorrelação significativa conforme o teste de Ljung–Box e distribuição aproximadamente simétrica segundo o teste de Jarque–Bera. Essa combinação de critérios garante que os modelos

finais sejam estatisticamente consistentes, bem ajustados e apropriados para a tarefa de previsão.

A Tabela 5.10 apresenta os modelos selecionados para previsão da geração nas duas usinas. Em ambos os casos, os modelos SARIMAX exibiram valores de AICc inferiores aos dos modelos SARIMA, o que indica melhor ajuste estatístico quando as variáveis exógenas são incorporadas à dinâmica autoregressiva.

Tabela 5.10: Modelos selecionados para previsão da geração fotovoltaica

	<i>Modelo</i>	<i>Exógenas</i>	<i>AICc</i>	<i>p-valor LB</i>	<i>p-valor JB</i>
Ituverava	<i>SARIMA</i>	ARIMA(1,0,1)	-	1754	0,72
	<i>SARIMAX</i>	SARIMA(1,0,1)(0,0,2) _[12]	CLDTOT _{_10} + TAUTOT _{_10} + ALBEDO _{_10} + LWGAB _{_10}	1720	0,61
Pirapora	<i>SARIMA</i>	SARIMA(1,0,0)(1,0,1) _[12]	-	1819	0,56
	<i>SARIMAX</i>	SARIMA(3,0,1)(1,0,1) _[12]	TAUTOT _{_10} + ALBEDO _{_10} + TS _{_10}	1738	0,66

Fonte: Elaboração própria (2025)

A Tabela 5.11 apresenta as equações estimadas para cada usina, com os coeficientes ajustados pela modelagem final, seguindo a forma geral descrita nas Equações (2.5) e (2.6).

Tabela 5.11: Equações dos modelos selecionados para previsão da geração fotovoltaica

	<i>Equação do Modelo</i>
Ituverava	<i>SARIMA</i> $z_t = 0,9 z_{t-1} - 0,58 a_{t-1}$ (5.4)
	<i>SARIMAX</i> $z_t = 0,96 z_{t-1} - 0,55 a_{t-1} + 0,17 a_{t-12} + 0,33 a_{t-24} - 14.656 CLDTOT_t - 235 TAUTOT_t + 536.008 ALBEDO_t + 271 LWGAB_t$ (5.5)
Pirapora	<i>SARIMA</i> $z_t = 0,5 z_{t-1} + 0,85 z_{t-12} - 0,66 a_{t-12}$ (5.6)
	<i>SARIMAX</i> $z_t = 0,1 z_{t-1} + 0,38 z_{t-2} + 0,47 z_{t-3} + z_{t-12} + 0,23 a_{t-1} - 0,78 a_{t-12} - 344 TAUTOT_t + 999.827 ALBEDO_t + 1.543 TS_t$ (5.7)

Fonte: Elaboração própria (2025)

Os coeficientes estimados nas equações reforçam relações coerentes com os conceitos físicos discutidos na Seção 3.2 e com a análise de correlação apresentada na Seção 5.1.3. Em Ituverava, as variáveis fração total da área coberta por nuvens ($CLDTOT_t$) e espessura óptica das nuvens ($TAUTOT_t$) apresentam coeficientes negativos, em linha com o efeito de atenuação da radiação solar exercido pelas nuvens. Em contraste, o albedo da superfície ($ALBEDO_t$) e a radiação de onda longa absorvida pela superfície ($LWGAB_t$) apresentam coeficientes positivos, o que é compatível com a correlação positiva observada entre ALBEDO e geração, e com o fato de que maiores valores de LWGAB estão associados a condições atmosféricas que frequentemente ocorrem em dias de maior energia disponível.

Em Pirapora, a espessura óptica das nuvens ($TAUTOT_t$) mantém coeficiente negativo, alinhado à redução da irradiação sob nuvens mais espessas. Por outro lado, o albedo da superfície ($ALBEDO_t$) e a temperatura da pele da superfície (TS_t) exibem coeficientes positivos, consistentes com seus sinais na matriz de correlação e com sua interpretação física — TS, em particular, foi a variável mais correlacionada com a geração na usina, indicando sua forte associação com condições de alta irradiação.

Após a seleção das especificações, foram geradas as previsões 12 passos à frente para Ituverava e Pirapora, cujos resultados, apresentados na Figura 5.12, permitem comparar diretamente o comportamento dos modelos SARIMA e SARIMAX ao longo do horizonte futuro. Em ambas as usinas, observa-se continuidade entre os valores observados e projetados, sem rupturas abruptas ou mudanças artificiais de padrão.

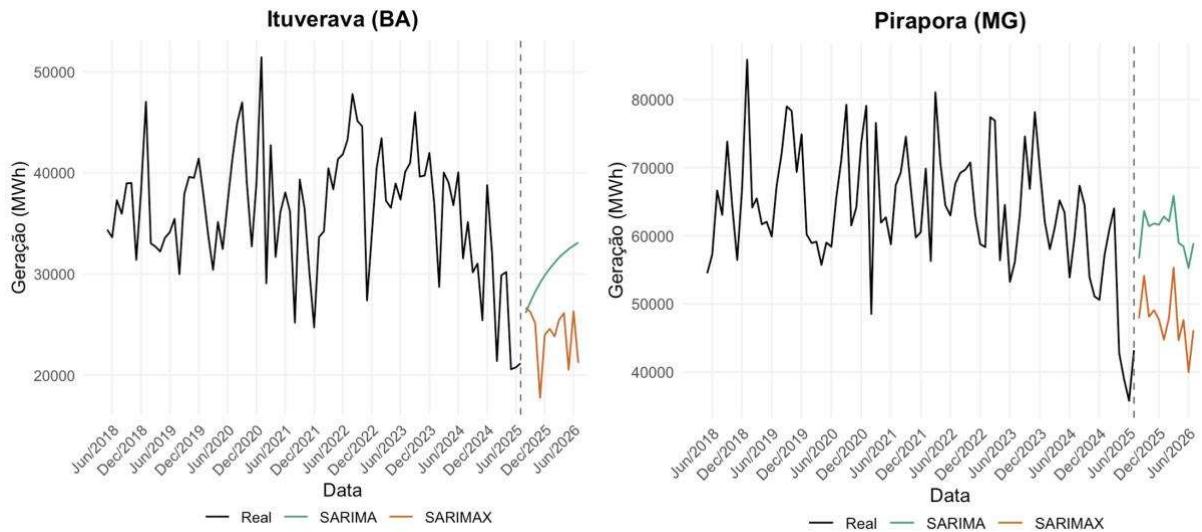


Figura 5.12: Previsão futura da geração de energia fotovoltaica (12 passos à frente)

Fonte: Elaboração própria (2025)

Em Ituverava, as previsões mostram trajetórias distintas entre os dois modelos. O SARIMA — que, no caso dessa usina, não possui componente sazonal explícita — produz uma recuperação mais pronunciada após o período de queda observado no final da amostra. Já o SARIMAX apresenta uma evolução mais moderada e contida, refletindo não apenas o efeito das variáveis exógenas projetadas, mas também a interação entre seus termos autoregressivos e de médias móveis, que tornam o modelo mais responsivo ao comportamento recente da série. Dessa forma, o SARIMAX tende a prolongar o nível reduzido de geração registrado nos últimos meses, produzindo previsões mais conservadoras no curto prazo.

Em Pirapora, a divergência entre os modelos é ainda mais evidente. O SARIMA sugere uma retomada em direção aos patamares historicamente observados, enquanto o SARIMAX mantém a série em níveis substancialmente mais baixos. Como as projeções das variáveis exógenas não indicam, por si só, deteriorações acentuadas, essa diferença parece decorrer da combinação entre a estrutura dinâmica do modelo e a queda expressiva observada nos meses finais da amostra, potencialmente associada a fatores operacionais como *curtailment* ou restrições temporárias de geração. Assim, o SARIMAX captura de forma mais forte essa tendência recente, propagando-a para o horizonte futuro e resultando em previsões mais conservadoras.

5.4.3. Validação com dados reais

Durante o desenvolvimento deste trabalho, novos dados de geração se tornaram disponíveis para os meses de agosto, setembro e outubro de 2025. Esses valores permitem realizar uma validação adicional dos modelos ajustados, comparando diretamente as previsões produzidas pelos modelos SARIMA e SARIMAX com as observações efetivamente registradas. Essa etapa complementa a avaliação realizada nas janelas históricas de teste, oferecendo evidências mais recentes sobre o desempenho preditivo dos modelos.

A Tabela 5.12 apresenta os resultados dessa comparação, com as métricas de erro RMSE e MAPE para cada usina e modelo. Em Ituverava, o SARIMAX reduziu o RMSE em cerca de 11% e o MAPE em aproximadamente 23% em relação ao SARIMA. Em Pirapora, os ganhos foram ainda mais expressivos, com quedas de cerca de 61% no RMSE e 68% no MAPE, mostrando superioridade em relação ao modelo sem variáveis exógenas. Embora a validação conte com apenas três meses, esses resultados indicam que, nas duas usinas, a inclusão das

variáveis climáticas melhorou de forma consistente a capacidade dos modelos de acompanhar o comportamento recente da geração.

Tabela 5.12: Desempenho das previsões em comparação com dados reais

	<i>Modelo</i>	<i>Exógenas</i>	<i>RMSE</i>	<i>MAPE</i>
Ituverava	SARIMA	ARIMA(1,0,1)	-	2370 9,47%
	SARIMAX	SARIMA(1,0,1)(0,0,2)[12]	CLDTOT_10 + TAUTOT_10 + ALBEDO_10 + LWGAB_10	2116 7,26%
Pirapora	SARIMA	SARIMA(1,0,0)(1,0,1)[12]	-	14404 30,25%
	SARIMAX	SARIMA(3,0,1)(1,0,1)[12]	TAUTOT_10 + ALBEDO_10 + TS_10	5634 9,67%

Fonte: Elaboração própria (2025)

A Figura 5.13 evidencia, para ambas as usinas, como os modelos se comportam em relação aos valores observados nos meses de agosto, setembro e outubro de 2025. Em Ituverava, o SARIMA projeta uma recuperação mais acentuada, enquanto o SARIMAX acompanha mais de perto o nível efetivamente registrado no início do horizonte, consistente com seus erros menores. Em Pirapora, o contraste é ainda mais claro, o SARIMA retorna rapidamente a patamares elevados, ao passo que o SARIMAX preserva o movimento de queda observado no final da amostra, alinhando-se de forma mais fiel à trajetória recente.

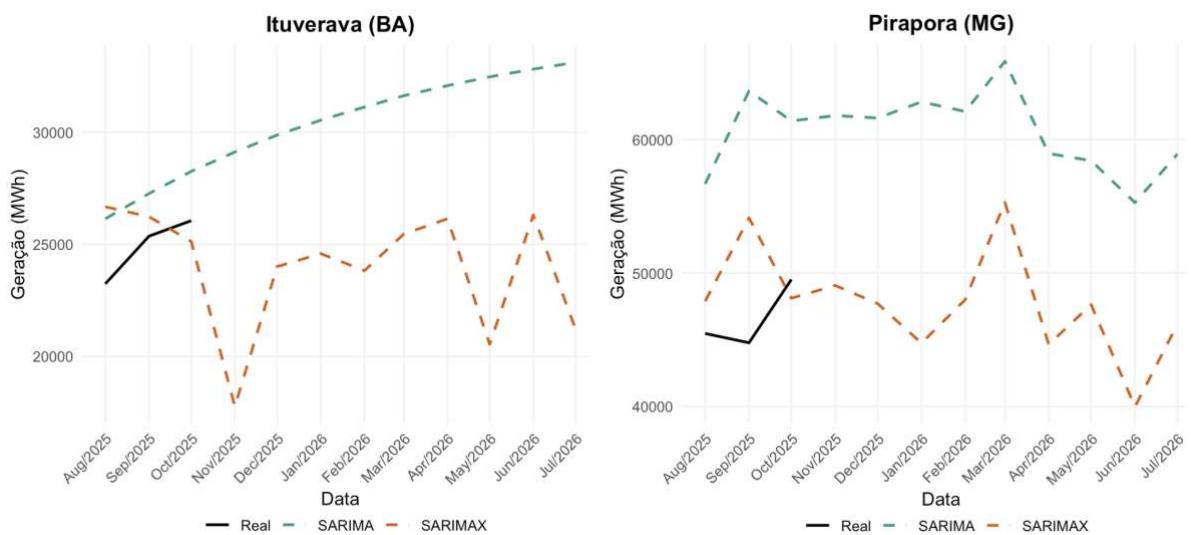


Figura 5.13: Comparaçao entre valores observados e previsões

Fonte: Elaboração própria (2025)

6. CONCLUSÕES

Este trabalho avaliou o impacto da inclusão de variáveis meteorológicas exógenas na previsão da geração fotovoltaica mensal nas usinas de Ituverava (BA) e Pirapora (MG), comparando sistematicamente modelos SARIMA e SARIMAX. A análise permitiu identificar de forma clara em quais condições as exógenas contribuem para a melhoria preditiva e quais combinações de variáveis e configurações de modelos se ajustam melhor a cada contexto operacional.

De forma geral, observou-se que as variáveis exógenas melhoraram o desempenho dos modelos sobretudo em períodos de maior instabilidade. Em momentos de comportamento mais regular — como em Ituverava na Janela 1 — o modelo univariado SARIMA capturou satisfatoriamente a sazonalidade e o nível médio da série, apresentando desempenho superior ao SARIMAX. Isso indica que, quando a dinâmica histórica é estável, a estrutura autorregressiva tende a generalizar bem, sem necessidade de covariáveis adicionais.

Entretanto, diante de mudanças abruptas — como reduções atípicas de geração ou desvios marcantes da sazonalidade — os modelos SARIMAX se mostraram mais eficientes. De maneira geral, todas as cinco variáveis meteorológicas analisadas (TAUTOT, ALBEDO, CLDTOT, TS e LWGAB) apareceram em alguma combinação vencedora, variando conforme a janela e o comportamento recente da série, o que reforça seu papel como mecanismos adicionais de ajuste às condições atmosféricas contemporâneas.

Em Pirapora, as combinações dessas variáveis se mostraram mais consistentes e recorrentes entre as janelas, com destaque para TAUTOT, CLDTOT, LWGAB e ALBEDO, além de TS no modelo final para previsão futura. Essa repetição sugere maior estabilidade na influência das exógenas e maior previsibilidade climática da usina, favorecendo o desempenho dos modelos SARIMAX. Já em Ituverava, embora as mesmas variáveis tenham sido selecionadas ao longo das janelas, o padrão foi menos uniforme: diferentes combinações foram escolhidas em cada período, refletindo maior sensibilidade às condições específicas de cada janela. Essa variação reforça que a influência das exógenas na usina é mais dependente do regime temporal e das mudanças recentes no comportamento da série.

O estudo também identificou um aumento expressivo dos erros de previsão na janela mais recente para ambas as usinas, independentemente do modelo. Esse padrão simultâneo sugere a ocorrência de fatores externos à dinâmica atmosférica — possivelmente *curtailment*, intervenções operacionais ou restrições temporárias — que não podem ser capturados por

variáveis meteorológicas, evidenciando limites inerentes aos modelos SARIMA e SARIMAX diante de rupturas estruturais.

As previsões futuras realizadas no tópico 5.4 mostraram trajetórias distintas entre os modelos: o SARIMA retorna mais rapidamente ao nível médio histórico, enquanto o SARIMAX projeta valores mais conservadores ao propagar a influência das variáveis exógenas. Ao comparar essas previsões com os dados reais de agosto a outubro de 2025, observou-se que o SARIMAX se manteve mais próximo da trajetória efetivamente registrada, apresentando reduções expressivas nos erros em ambas as usinas — especialmente em Pirapora, onde o ganho percentual foi substancialmente maior. Essa evidência adicional, obtida fora do período utilizado na modelagem, confirma a maior responsividade do SARIMAX às condições recentes da série e reforça o benefício da incorporação das variáveis meteorológicas no processo preditivo.

Em síntese, conclui-se que a inclusão de variáveis meteorológicas é uma ferramenta valiosa para aprimorar previsões de geração fotovoltaica, especialmente em contextos de instabilidade. Contudo, o desempenho ótimo depende do contexto específico de cada usina, da regularidade das séries meteorológicas e do regime observado no período analisado. Os resultados reforçam a relevância de abordagens híbridas no planejamento energético e mostram que os modelos SARIMAX capturam dimensões complementares da dinâmica de geração, ampliando a capacidade preditiva em diferentes cenários.

Apesar dos avanços obtidos, este estudo apresenta limitações que também apontam direções claras para pesquisas futuras. A ausência de informações operacionais detalhadas das usinas, como eventos de *curtailment* ou manutenções, limita a capacidade dos modelos de capturar integralmente rupturas estruturais observadas na série, especialmente nos períodos mais recentes. Ademais, a modelagem em frequência mensal reduz a sensibilidade a eventos intramensais, e a estrutura linear dos modelos SARIMAX pode não capturar relações atmosféricas mais complexas. Estudos futuros podem avançar ao incorporar dados operacionais reais das usinas, explorar frequências mais altas (diária ou horária), testar modelos híbridos e não lineares, expandir a análise para um maior número de usinas e regiões e incluir novas variáveis atmosféricas, ampliando a robustez e a generalização dos resultados.

BIBLIOGRAFIA

- AHMED, R.; SREERAM, V.; MISHRA, Y.; ARIF, M. D.; “A Review and Evaluation of the State-of-the-Art in PV Solar Power Forecasting: Techniques and Optimization”; *Renewable and Sustainable Energy Reviews*, v. 124; 2020.
- ANGELINI, L. P.; BIUDES, M. S.; MACHADO, N. G.; et al.; “Surface Albedo and Temperature Models for Surface Energy Balance Fluxes and Evapotranspiration Using SEBAL and Landsat 8 over Cerrado-Pantanal, Brazil”; *Sensors*; 2021.
- ANTONANZAS, J.; OSORIO, N.; ESCOLAR, D.; VIDAL, P. G.; HERNÁNDEZ, J. A.; BLANCO, J.; YANG, D.; “Review of Photovoltaic Power Forecasting”; *Solar Energy*, v. 136; 2016.
- ARAUJO, M. A. S. G.; AGUILAR, S.; SOUZA, R. C.; OLIVEIRA, F. L. C.; “Global Horizontal Irradiance in Brazil: A Comparative Study of Reanalysis Datasets with Ground-Based Data”; *Energies*, v. 17; 2024.
- BACHER, P.; MADSEN, H.; NIELSEN, H. A.; “Online Short-Term Solar Power Forecasting”; *Solar Energy*, v. 83, n. 10, p. 1772–1783; 2009.
- BARRY, J.; MEILINGER, S.; PFEILSTICKER, K.; et al.; “Irradiance and Cloud Optical Properties from Solar-PV Systems”; *Atmospheric Measurement Techniques*, v. 16, p. 4975–5000; 2023.
- BIRD, L.; COCHRAN, J.; WANG, X.; “Wind and Solar Energy Curtailment: Experience and Practices in the United States”; Golden, CO: National Renewable Energy Laboratory (NREL); 2014.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M.; “Time Series Analysis: Forecasting and Control”; Hoboken: John Wiley & Sons; 2015.
- BREYER, C.; et al.; “Solar Photovoltaics Demand for the Global Energy Transition in the Power Sector”; *Progress in Photovoltaics: Research and Applications*, v. 25, n. 7; 2017.

CHEN, C.; DUAN, S.; CAI, T.; LIU, B.; “Online 24-h Solar Power Forecasting Based on Weather Type Classification Using Artificial Neural Network”; *Solar Energy*, v. 85, n. 11, p. 2856–2870; 2011.

DING, Y.; LIU, Q.; LAO, P.; LI, M.; LI, Y.; ZHENG, Q.; PENG, Y.; “Spatial Distributions of Cloud Occurrences in Terms of Cloud Area Fraction and Volume Fraction”; *Remote Sensing*, v. 15, n. 16; 2023.

EPE – Empresa de Pesquisa Energética; “Balanço Energético Nacional 2025: ano-base 2024 – Relatório Final”; Rio de Janeiro: EPE; 2025.

H. MAHMOUD; N. AKKARI; “Shortest Path Calculation: A Comparative Study for Location-Based Recommender System”; *World Symposium on Computer Applications & Research (WSCAR)*; Cairo; 2016.

HYNDMAN, R. J.; “The ARIMAX Model Muddle”; 2010.

HYNDMAN, R. J.; ATHANASOPOULOS, G.; “Forecasting: Principles and Practice”; Melbourne: OTexts; 2021.

HYNDMAN, R. J.; KHANDAKAR, Y.; “Automatic Time Series Forecasting: The Forecast Package for R”; *Journal of Statistical Software*, v. 27, n. 3; 2008.

IEA – International Energy Agency; “Global Energy Review 2025”; Paris: IEA; 2025.

INMAN, R. H.; PEDRO, H. T. C.; COIMBRA, C. F. M.; “Solar Forecasting Methods for Renewable Energy Integration”; *Progress in Energy and Combustion Science*, v. 39, n. 6, p. 535–576; 2013.

IRENA – International Renewable Energy Agency; “Renewable Energy Statistics 2025”; Abu Dhabi: IRENA; 2025.

JARQUE, C. M.; BERA, A. K.; “Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals”; *Economics Letters*, v. 6, n. 3, p. 255–259; 1980.

JIN, M.; DICKINSON, R. E.; “Land Surface Skin Temperature Climatology: Benefiting from the Strengths of Satellite Observations”; *Environmental Research Letters*, v. 5, n. 4; 2010.

- LJUNG, G. M.; BOX, G. E. P.; “On a Measure of Lack of Fit in Time Series Models”; *Biometrika*, v. 65, n. 2, p. 297–303; 1978.
- LORENZ, E.; HURKA, J.; HEINEMANN, D.; BEYER, H. G.; “Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems”; *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 2, n. 1, p. 2–10; 2009.
- LU, Z.; et al.; “Evaluation of Satellite Land Surface Albedo Products over China Using Ground Measurements”; *Remote Sensing*; 2021.
- MAKRIDAKIS, S.; SPILIOOTIS, E.; ASSIMAKOPOULOS, V.; “Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward”; *PLOS ONE*, v. 13, n. 3; 2018.
- MELO, R. S.; et al.; “Renewable Energy Sources Spatio-Temporal Scenarios Simulation under Influence of Climatic Phenomena”; *Electric Power Systems Research*, v. 235; 2024.
- ONS – Operador Nacional do Sistema Elétrico; “Plano da Operação Energética 2025–2029 (PEN 2025)”; Rio de Janeiro: ONS; 2025.
- ONU – Organização das Nações Unidas; “Acordo de Paris”; Nova York: ONU; 2015.
- SARKAR, M. N. I.; “Estimation of Solar Radiation from Cloud Cover Data of Bangladesh”; *Renewables: Wind, Water, and Solar*, v. 3, n. 11; 2016.
- SOBRI, S.; KOON, L. W.; RAHIM, N. A.; MUSIRIN, I.; “A Review on Globally Implemented Solar Energy Prediction Methods, Models and Techniques”; *Renewable and Sustainable Energy Reviews*, v. 134; 2020.
- SOVACOOL, B. K.; “When Subterranean Worlds Collide: The Uncertain Future of Global Energy Transitions”; *Energy Research & Social Science*, v. 80; 2021.
- SHUVALOVA, J.; “Cloud Characteristics and Their Effects on Solar Irradiance”; *Atmosphere*, v. 14, n. 12; 2023.
- STOCK, J. H.; WATSON, M. W.; “Introduction to Econometrics”; Harlow: Pearson Education Limited; 2020.

THOMPSON, C. G.; KIM, R. S.; ALOE, A. M.; BECKER, B. J.; “Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results”; *Basic and Applied Social Psychology*, v. 39, n. 2, p. 81–90; 2017.

WANDJI NYAMSI, W.; LINDFORS, A. V.; MEYER, A.; LIPPONEN, A.; AROLA, A.; “A New Method for Estimating Cloud Optical Depth from PV Power Measurements”; *EGUsphere* [preprint]; 2025.

WANG, K.; DICKINSON, R. E.; “Global Atmospheric Downward Longwave Radiation at the Surface from Ground-Based Observations, Satellite Retrievals, and Reanalyses”; *Reviews of Geophysics*, v. 51; 2013.

YI, Y.; ZHENG, Z.; LV, T.; et al.; “Nowcasting of Surface Solar Irradiance Based on Cloud Optical Thickness from GOES-16”; *Remote Sensing*, v. 17, n. 16; 2025.

ZENG, Q.; CHENG, J.; “Estimating High-Spatial Resolution Surface Daily Longwave Radiation from the Instantaneous GLASS Longwave Radiation Product”; *International Journal of Digital Earth*, v. 14, n. 11, p. 1303–1318; 2021.

ZHANG, S.; et al.; “Review of Land Surface Albedo: Variance Characteristics, Climate Impacts and Management Strategies”; *Remote Sensing*; 2022.

ZHANG, X.; XU, M.; “Assessing the Effects of Photovoltaic Powerplants on Surface Temperature Using Remote Sensing Techniques”; *Remote Sensing*, v. 12, n. 11; 2020.